The Twentieth Annual Ted Winnowski ‘63 Student Conference in Business

TWSCB 2026

Vol. 19 (1)

Editors:
Kaitlyn Hathaway (‘26)

Cheryl Buff, Ph.D.
Professor of Marketing

Ashley Provencher, Ph.D.
Professor of Economics

Kimberley Preiksaitis, Ph.D.
Associate Professor of Marketing

Ted Winnowski '63

Ted Winnowski, a financial and corporate consultant, is also a strategic member of Columbia Business Advisors, a private consulting firm in the Northwest, plus Veber partners an investment banking firm. Previously, he served as president and chief executive officer of Centennial Bank and Centennial Bancorp, Portland. He has more than 30 years of financial and banking experience, having begun his career at KeyCorp's flagship bank in Albany, New York. He served as president and CEO of KeyBank of Oregon, KeyBank of Washington, and EVP of KeyCorp's Northwest Region in Seattle. Prior to banking assignments, he served with the IRS in New York City and the U.S. Army in France.

He graduated with a bachelor's degree in economics from Siena College and did post-graduate work in business and management at Rensselaer Polytechnic Institute and at the National Graduate Trust School of Northwestern University. He recently served as a board member for Regence BlueCross and BlueShield of Oregon, is the former Honorary Consul of the Republic of Poland in Oregon, and is honorary board chair for the Portland Chamber of Commerce (currently Portland Business Alliance).

His former community activities include the chairmanship of the Oregon Bankers Association, and membership in the Mayor's Business Round Table, the Association for Portland Progress, the World Affairs Council of Oregon, and the United Cerebral Palsy. He currently serves on the Board of Regents, University of Portland and as past chair for six years.

TABLE OF CONTENTS

SOCIETAL IMPACT FELLOWS

OPTIMIZING HEALTH INSURANCE COSTS THROUGH QUANTITATIVE DATA ANALYSIS FOR A NONPROFIT ORGANIZATION……………………….………… 99

Author: Ava Greer and Alyssa Ramundo
Faculty Mentor: Dr. John O'Neill and Dr. Shahryar Gheibi

ENHANCING SCAM PREVENTION MESSAGING: A STUDY ON SOURCE CREDIBILITY AND AUTHENTICITY…………………………………………………. 60

Author: AKM Alam and Olivia Botelho
Faculty Mentor: Dr. Marie Rice

ACCOUNTING

PROFESSIONALS' PERCEPTIONS OF RISKS IN AUDITING CRYPTO CURRENCY…………………………………………………………………………. 102

Author: Caden Kramer
Faculty Mentor: Dr. Marie Rice

BUSINESS ANALYTICS

INDIVIDUAL HEALTH EXPENDITURES: KEY DETERMINATS…………………………………………………………………………… 80

Authors: McKenna Rowley, Ava Greer, Barbara Borgonhone, and Matthew Witkop
Faculty Mentor: Dr. Manimoy Paul

UNCONVENTIONAL FACTORS THAT AFFECT MLB WINNING PERCENTAGE………………………………………………………………………… 123

Authors: Samuel Desany, Maxfield Taylor, and Alexander Eberlein
Faculty Mentor: Dr. Manimoy Paul

UNDERSTANDING THE DRIVERS OF BINGE DRINKING: AN EMPIRICAL ANALYSIS USING U.S. SURVEY DATA…………………………………………………………………………………. 138

Authors: Rory McNerney, Kolawole Olawole, and Tristan Terry
Faculty Mentor: Dr. Manimoy Paul

USING ANALYTICS TO IDENTIFY THE KEY FACTORS INFLUENCING JUVENILE DELINQUENCY………………………………………………………………………… 153

Authors: Lee Mara, Kara Stallmer, and Ty Strong
Faculty Mentor: Dr. Manimoy Paul

ECONOMICS

AGRICULTURE AND CRIME IN MEXICO: A 2023 ANALYSIS…………………………………………………………..…….…………………… 5

Authors: Sidney Tovar
Faculty Mentor: Dr. Scott Trees

ANALYSIS OF CRIME AND INEQUALITY IN LATIN AMERICA…………………………………………………………..…..…………….. 26

Authors: Sidney Tovar
Faculty Mentor: Dr. Elias Shukralla

FEDERAL RESERVE CHALLENGE 2025 ………………………………………………... 72

Presentation Team: Kylie Castellana, Bradley Fraser, Andrew Kirk, Giovanna Perez, Kerry Wise
Research Team: Emily Film, Josh Florence, Brendan Kasuba, Elizabeth Lankford, Micah Moore, Christopher Nass, Jack Reilly, Sophia Tagliafierro, Natalie Zhu
Faculty Mentor: Dr. Aaron Pacitti

ENTREPRENEURSHIP

BLAKE BRAKES EXECUTIVE SUMMARY……………………………………………… 47

Authors: Logan Allen, Blake Demianenko, Ian Downer, Declan Coletti, and Mark Grosky
Faculty Mentor: Professor Natasha Adams

CULTURE SHOCK EXECUTIVE SUMMARY: TAST THE WORLD ONE PLATE AT A TIME……………………………………………………………………………………… 58

Authors: Madison Milton, Leo Alvarez- Guzman, Tehillah Mbango, Ameer Lester, and Abigail Trinkle
Faculty Mentor: Professor Natasha Adams

SACRED PAWS CO. EXECUTIVE SUMMARY…………………………………………. 108

Authors: Madison Bray, Evan Parker, Annabella Rivero, and Megan Ferreira
Faculty Mentor: Professor Natasha Adams

TEMPO: FITTING HEALTH & WELLNESS INTO THE RHYTHM OF YOUR DAY……………………………………………………………………………………… 111

Authors: Josh Kotlow, Evan Williams, Matt Witkop, Nolan Towers, Lucas Centanni, and Joe Gilbert
Faculty Mentor: Professor Natasha Adams

MANAGEMENT

HANNAFORD PRELIMINARY HR AUDIT……………………………………………… 79

Authors: Ashley Wilkinson and Sylvan Obando
Faculty Mentor: Professor Deborah Kelly

INTRODUCTION TO A PRELIMINARY HR AUDIT: COLUMBIA LAND CONSERVANCY…………………………………………………………………….............. 100

Authors: Esther Ma and Elizabeth Lankford
Faculty Mentor: Professor Deborah Kelly

PEDIATRIC VACCINATION RATES FOR RURAL CHILDREN: A COMPARATIVE ANALYSIS………………………………………………………………………………… 101

Authors: Mea Amedio and Joshua Hunter
Faculty Mentor: Dr. Andrea Smith-Hunter and Dr. James Nolan

MARKETING

CONSUMER PERCEPTIONS OF ARTIFICIAL INTELLIGENCE IN ADVERTISING…………………………………………………………………………. 49

Authors: Angus Grout, Carson Hines, Cayden Catalina, and Joshua Miller
Faculty Mentor: Dr. Cheryl Buff

FOOD WASTE IN THE CAPITAL REGION: DISTRIBUTION NETWORKS & REGULATORY CONSTRAINTS………………………………………………………… 76

Authors: Tehillah Mbango
Faculty Mentor: Dr. Cheryl Buff

THE WILLINGNESS TO PURCHASE CANNABIS/ THC……………………………… 114

Authors: Corrinne Blond, Ethan Hunt, Ethan Linton, and Hailey Westbrook
Faculty Mentor: Dr. Cheryl Buff

WASTED POTENTIAL: UNDERSTANDING FOOD WASTE ON COLLEGE CAMPUSES………………………………………………………………………………. 164

Authors: Stephanie Davis, Lydia Griffen, and Rachel Drew
Faculty Mentor: Dr. Cheryl Buff

AGRICULTURE AND CRIME IN MEXICO: A 2023 ANALYSIS

Sidney Tovar, Siena University
Dr. Scott Trees, Siena University - Faculty Mentor

ABSTRACT

This study examines whether rising crime rates contributed to a decline in agricultural employment across Mexico's 32 states in 2023. This year was marked by escalating cartel violence following the arrest of Ovidio Guzmán, a 20% minimum wage increase, and mounting environmental pressures. Using cross-sectional regression analysis, agricultural workforce growth is modeled as a function of state-level crime rates, labor informality, wage growth, access to financing, and environmental staffing activity. The results indicate that while the overall model is statistically significant and explains roughly 55% of the variation in agricultural employment growth, no individual variable, including crime rate, achieves statistical significance on its own. Robustness checks incorporating GDP distribution, income inequality, and homicide rates preserve the model's overall significance with a marginal gain in explanatory power. Estimation of non-linear functional forms reveals that cubic specifications substantially improve model fit, raising the R-squared to 0.924, suggesting that the effects of crime and other structural factors on agricultural employment are non-linear in nature. These findings indicate that agricultural employment in Mexico is shaped by a complex interaction of violence, economic modernization, labor informality, and structural transformation rather than by crime alone. Policy responses aimed solely at reducing insecurity are unlikely to be sufficient. Comprehensive strategies addressing rural development, technological change, and economic diversification are needed to stabilize Mexico's agricultural workforce.

INTRODUCTION

The agricultural sector plays a vital and economically significant role in Mexico, accounting for approximately 4% of the nation's Gross Domestic Product (GDP). While official government data suggests a gradual decrease in overall unemployment across Mexico (Trading Economics, 2026), the country has simultaneously struggled with a substantial and concerning rise in crime rates in recent years (Valle-Jones, 2026). This surge in criminal activity has disproportionately and severely impacted agricultural workers and the industry.

The year 2023 saw a dramatic and widespread decline in harvest levels across various regions (OECD, 2023). This decline was not attributable to a single cause but was aggravated by a confluence of factors, most notably the intense escalation of violence and organized criminal activity following the high-profile capture of Ovidio Guzmán (Stevenson et al, 2023). In the wake of this event, criminal groups aggressively intensified their tactics, primarily through extortions. These groups began demanding steep "protection payments" from farmers, often under the explicit and credible threat of violence or death (DeHaro, 2025). Furthermore, these illicit organizations frequently resorted to burning existing, legitimate crops to forcibly clear the land and replace them with lucrative, illicit plantations, often referred to as "green gold."

Criminal violence has profoundly destabilized Mexico's agricultural communities, generating widespread fear and disruption. In response to the continuous threats, extortions, and destruction of livelihoods, agricultural workers have organized and conducted various strikes. These labor actions have not only stalled domestic operations but have also significantly disrupted critical supply chains that connect Mexican produce to international markets (Mexico Solidarity Media, 2025). The ongoing loss of productive farmland to illicit activities represents a critical challenge for Mexico's agricultural future, as it is projected to cause an inflationary effect on the global prices of natural products originating from Mexico. This crisis forms the foundation of a specific econometric investigation: The central hypothesis being tested is that the measurable increase in the crime rate within Mexico acts as a direct cause, resulting in a demonstrable decline in agricultural employment throughout the country.

BACKGROUND

Why 2023? Rationale for Year Selection

The combination of social, economic, and environmental factors in 2023 established a pivotal year for analyzing the resilience and long-term stability of Mexico’s agricultural sector. This period, situated two years post-COVID-19 pandemic disruptions and nearing the conclusion of the first six-year term of the MORENA administration, amplified existing vulnerabilities within the nation's rural economy.

Socially, the year was marked by a severe escalation in organized crime violence, directly impacting agricultural regions. The January 2023 arrest of Ovidio Guzmán, a high-profile drug trafficker, triggered immediate and intense retaliatory violence. This culminated in events such as the devastating "second Black Thursday" in Sinaloa, where blockades, widespread arsons, and confrontations severely disrupted supply chains, imperiling the transport and sale of agricultural products (DeHaro, 2025). The pervasive threat of extortion and land seizure by criminal groups further undermined the safety and livelihood of rural producers across key growing states.

Economically, producers faced a compounding set of pressures. On one hand, the government implemented a significant 20% increase in the national minimum wage (Creel, 2022). This policy was intended to alleviate poverty but which simultaneously raised labor costs for agricultural operations. This increased pressure on expenses coincided with a troubling decline in land availability and utilization. The total number of hectares dedicated to agriculture continued a downward trend (Trejo, 2024), showing a complex interplay of urbanization, land tenure issues, and decreased profitability, collectively intensifying the economic hardship on small and medium-sized rural producers.

Environmentally, the sector worsened with climate-related challenges. Widespread and prolonged droughts became a major factor in 2023, severely limiting water availability for irrigation and consequently reducing overall agricultural productivity (Voiland, 2024). This environmental stressor not only lowered crop yields but also put immense strain on livestock farming, forcing difficult decisions regarding resource allocation. The combined effects of violence, rising costs, land pressure, and droughts, underscore the urgent necessity for a comprehensive assessment of the structural challenges facing Mexico's vital agricultural sector.

Data and Variable Selection

This study analyzes cross-sectional data across Mexico's 32 states for the year 2023. All variables are measured at the state level, reflecting the administrative unit at which most labor market and crime data are consistently reported in Mexico. The dependent variable and all independent variables were collected from two primary sources: Data México, and the Instituto Nacional de Estadística y Geografía (INEGI).

The dependent variable is the annual workforce growth in the agricultural sector by state in 2023, measured as the change in the proportion of individuals employed in agriculture over the course of the year. This variable was sourced from Data México and captures short-run fluctuations in agricultural labor demand at the state level. Workforce growth was selected as the outcome variable as the study is interested in how changing conditions in 2023 (particularly rising crime) affected the trajectory of agricultural employment.

The primary independent variable is the crime rate by state of occurrence, measured as the total number of crimes per 100,000 inhabitants in 2023, sourced from INEGI. This variable serves as the central proxy for violence, cartel activity, and rural insecurity in the regression model. From a theoretical standpoint, higher crime rates are expected to reduce agricultural employment through several channels. Farmers and agricultural workers facing threats of extortion, crop destruction, or physical violence may abandon their land or exit the labor force entirely. Additionally, crime-driven uncertainty raises the cost of agricultural production, potentially leading employers to reduce hiring. Based on this reasoning, a negative relationship between crime rate and agricultural workforce growth is anticipated.

Four control variables are included to isolate the effect of crime and account for other structural factors that influence agricultural employment. The first is labor informality in the agricultural sector, measured as the share of agricultural workers without formal contracts or social protections in 2023. Informality is included because informal workers lack the institutional protections that might otherwise protect them against external shocks, making them more likely to exit the labor force when conditions deteriorate. A negative relationship between informality and workforce growth is expected, as higher informality may reflect weaker labor market institutions that are less capable of retaining workers under adverse conditions.

The second control variable is annual wage growth in the agricultural sector, measured as the average monthly change in earnings for agricultural workers in 2023 by state. Wage growth is included as a measure of the relative attractiveness of agricultural employment. Standard labor supply theory predicts that higher wages incentivize workers to remain in or enter the agricultural sector, and therefore a positive relationship between wage growth and workforce growth is anticipated. However, the practical significance of this variable is expected to be low, given that agricultural wages in Mexico remain modest even after the 2023 minimum wage increase.

The third control variable is the distribution of economic units that received financing by state, measured as the percentage of agricultural businesses that successfully obtained financing in 2023. Access to financing is included as a proxy for the financial capacity of agricultural producers to sustain operations and maintain their workforce. While a positive relationship is theoretically expected, the possibility of capital-labor substitution means that financing could also facilitate the adoption of technology that reduces labor demand, which may complicate interpretation of this variable's coefficient.

The fourth control variable is the share of large economic units with staff dedicated to environmental protection activities by state. This variable captures the degree of institutional formalization and regulatory compliance within the agricultural sector. States where a higher proportion of large agricultural firms invest in environmental staffing may reflect broader organizational capacity and stability, which could support employment retention. A positive relationship is anticipated, though its practical significance is expected to be limited.

METHODOLOGY

This study employs Ordinary Least Squares (OLS) regression to examine the relationship between crime rates and agricultural employment growth across Mexico's 32 states in 2023. The analysis uses cross-sectional data due to the availability of consistent state-level observations for the year 2023, which was selected as a critical inflection point for Mexico's agricultural sector, as discussed in the Background section. A full summary of all variables, including definitions, units, and descriptive statistics, is presented in Table 1.

The baseline model is as follows,

$$EmpAgriculture \; = \beta_1 + x_2 CrimeRate + x_3 LaborInfAgr + x_4 GrowthWageAgr \; +$$

$$x_5 DistEconUnits \; + \; x_6 EnvAct + \varepsilon$$

Where i indexes each of the 32 Mexican states, β_1 is the intercept, β_2 through β_6 are the estimated slope coefficients for each independent variable, and ε represents the error term.

To assess the robustness of the baseline results, a second regression model is estimated incorporating three additional control variables: the distribution of GDP by federal entity, income inequality measured by the Gini coefficient, and homicide rates by state. The inclusion of these variables tests whether the main findings hold after accounting for broader macroeconomic and institutional conditions. Several diagnostic tests are conducted to validate the regression model. Multicollinearity among independent variables is assessed using a correlation matrix, with coefficients below 0.6 considered acceptable. Heteroscedasticity is evaluated using the eyeball test on residual plots and formally tested using the White test. Outliers are identified using the three standard deviation threshold, with a secondary check at two standard deviations. Autocorrelation is not tested, as it is not applicable to cross-sectional data.

It is important to acknowledge two key limitations of this methodology. First, cross-sectional analysis establishes correlation rather than causation. The results indicate associations between variables at a single point in time and cannot confirm the direction of causality. Second, potential endogeneity exists between crime and agricultural employment, as declining agricultural activity may itself contribute to higher crime rates rather than the reverse. While this study does not employ instrumental variables or other techniques to formally address endogeneity, this limitation is considered when interpreting the findings. Additionally, the use of state-level aggregates may mask significant within-state variation in both crime exposure and agricultural labor market conditions.

FINDINGS

Regression output

The regression was initially conducted including all variables previously proposed. Workforce Growth in the Agricultural Sector in Mexico acted as the dependable variable, while Crime Rate by State remained the main independent variable. All data was divided by the 32 states in Mexico, in the year 2023. The results of the regression were as shown in table 3.

Table 2. Results of initial regression.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.745[a]	.555	.462	.18129557975

a. Predictors: (Constant), EconUnitsInEnv, CrimeRate, WageGrowth, DistributionFinUnits, LaborInformality

The result of the regression reveals a significant relationship between the variables, where 55 percent of the changes in Growth in Employment in the Agricultural Sector in Mexico can be explained by all independent variables, as noted by the R squared. The regression had a high statistical significance, as the p-value was less than .001, meaning that there is a high probability that the true slope of the function is not zero. Nonetheless, as seen on the preceding results of the regression, most of the variables did not have a high statistical significance by themselves. In other words, the individual independent variables did not exhibit statistical significance on their own. Specifically, Crime Rate, Labor Informality, Wage Growth, and Distribution of Financial Units have p-values greater than 0.05, indicating that they do not significantly affect agricultural employment growth at the individual level.

All the estimated coefficients (table 3) align with theoretical expectations, with the exception of the Distribution of Financial Units, which has an unexpected sign. The coefficient suggests that lower financial distribution is associated with higher agricultural employment in Mexico. This result may reflect increased adoption of technology in the sector, where greater access to financial resources allows producers to substitute capital for labor, reducing employment demand. Alternatively, this relationship may be due to reverse causality, as increased financial resources can facilitate internal migration by incentivizing farmers to seek higher-paying opportunities outside the agricultural sector, thereby reducing agricultural employment growth.

Table 3. Coefficients of Initial Regression

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.010	.131		.076	.940
	CrimeRate	-2.748E-7	.000	-.010	-.065	.949
	LaborInformality	-.007	.007	-.177	-1.060	.300
	WageGrowth	.152	.102	.224	1.488	.150
	DistributionFinUnits	-.011	.006	-.301	-1.910	.068
	EconUnitsInEnv	.055	.012	.686	4.412	<.001

a. Dependent Variable: WorkforceGrowth

The regression model was later run with all variables that were initially proposed in the project. The main purpose in running the regression model with all variables was to test that the main findings are not sensitive to the inclusion of additional controls. This regression reveals the main relationships are reliable, rather than driven by omitted variables or model specification choices. The regression now included the following variables,

- Distribution of GDP by Federal Entity: measures the economic contribution of each state to the national GDP, as a percentage in 2023. This variable acts as a proxy to measure the economic output of agricultural production by state.It is expected to have a positive coefficient.
- Income distribution inequality based on Gini coefficient, by state: Measures inequality and institutional strength, governance effectiveness, and regulatory quality that influence rural labor markets in 2022, by state. It is expected to have a positive coefficient.
- Homicides By Federal Entity, According to year Registration: Rate of homicides, broken down by which federal agency was responsible in 2023, by state. It is expected to have a negative coefficient.

The results of the regression that included all variables resulted in a slight higher r-squared. The results (table 4) suggest that 56 percent of the changes in Growth in Employment in the Agricultural Sector in Mexico can be explained by all independent variables, as noted by the R squared. The regression still had statistically significant, shown by the p-value being lower than .05, meaning that there is a high probability that the true slope of the function is not zero. The overall new results confirm the significance of the main findings, where the results hold even after accounting for all relevant factors, increasing confidence that the observed effects are real and not due to model misspecification.

Table 4. Second Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.750[a]	.562	.395	.19234211124

a. Predictors: (Constant), Homocides, CrimeRate, EconUnitsInEnv, WageGrowth, Gini, DistributionFinUnits, LaborInformality, GDPperState

All the new estimated variables (table 5) showed no statistical significance, as the p-value was greater than 5 percent. Their coefficients were mostly as expected, with the exception of the share of GDP per state. Its negative relationship suggests that the more distribution of GDP by the Federal State reduces the growth of workers in the agricultural sector. This interpretation can be due to an increase in technology, and innovation, where workers are not needed in the sector. Its coefficient could also indicate an increase in the economic diversification and structural transformation, whereby states with a higher share of GDP are shifting away from labor-intensive agriculture toward more productive sectors such as manufacturing and services, as portrayed in the Lewis Model (Lewis, 1954).

Table 5. Coefficients of Second Regression

Coefficients[a]

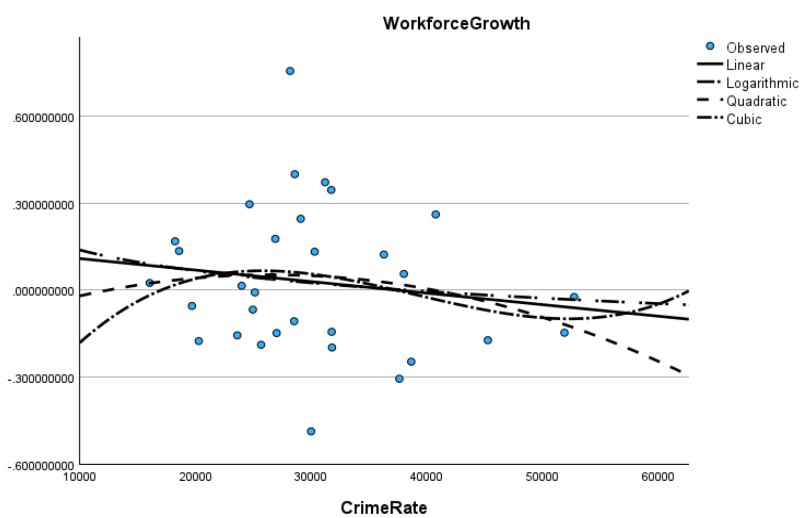
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.370	.820		-.452	.656
	CrimeRate	8.651E-7	.000	.031	.157	.877
	LaborInformality	-.005	.010	-.128	-.516	.611
	WageGrowth	.163	.128	.241	1.276	.216
	DistributionFinUnits	-.012	.006	-.318	-1.857	.077
	EconUnitsInEnv	.060	.019	.757	3.247	.004
	Gini	.858	1.813	.101	.473	.641
	GDPperState	-.671	2.070	-.083	-.324	.749
	Homocides	5.247E-6	.000	.019	.088	.931

a. Dependent Variable: WorkforceGrowth

Estimation

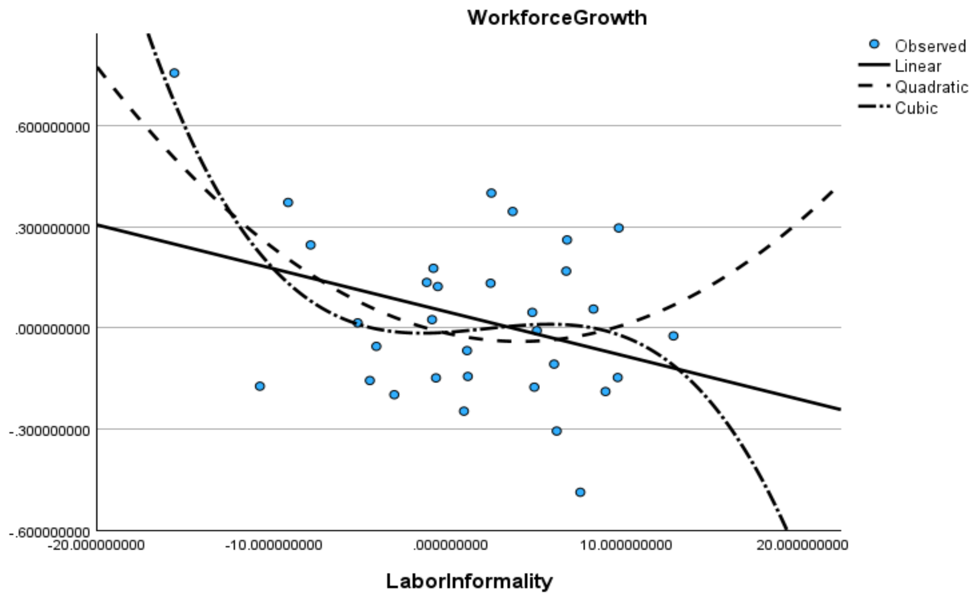
The appropriate curve for each variable was determined, after conducting the initial regression, which determined the overall shape of the curve on the analysis. The results were as shown in the following graphs,

Graph 1. Crime Rate vs Workforce Growth



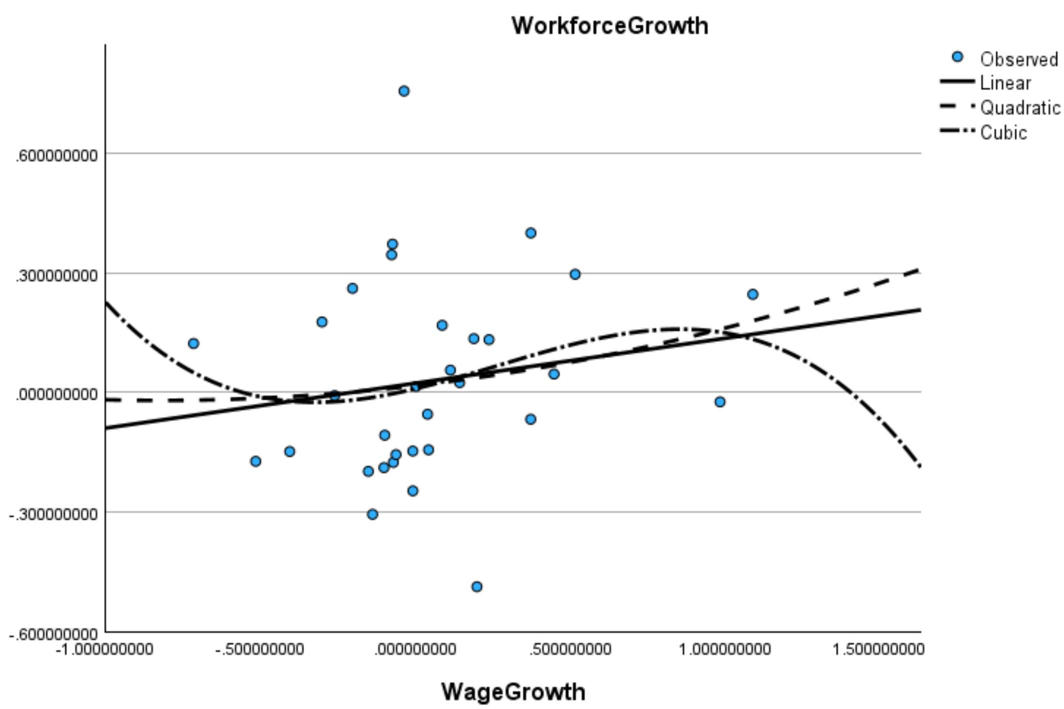
**Crime rate: Highest R-squared with Cubic shape, with a total of 0.37.

Graph 2. Labor informality vs Workforce Growth



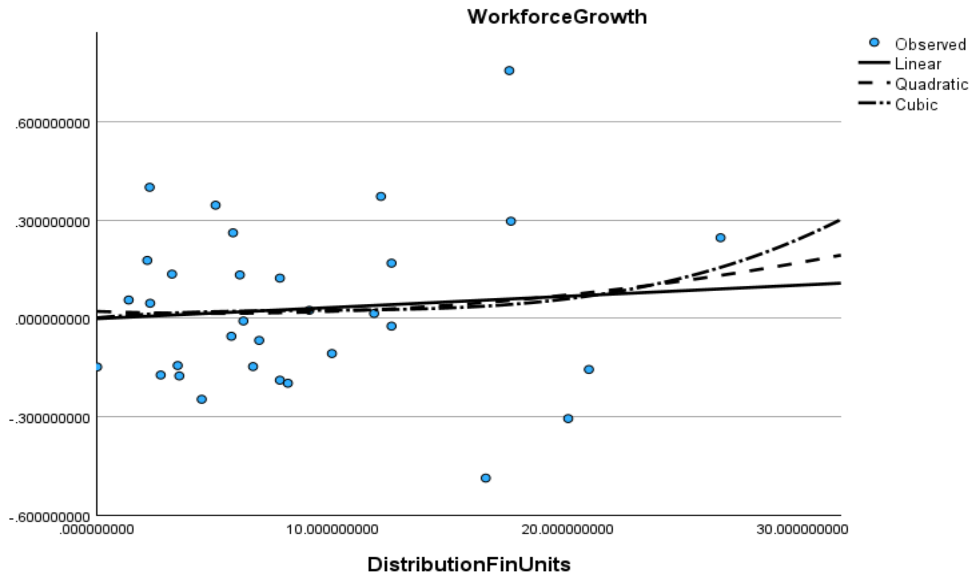
**Labor Informality: Highest R-squared with Cubic shape, with a total of 0.264

Graph 3. Wage Growth vs Workforce Growth



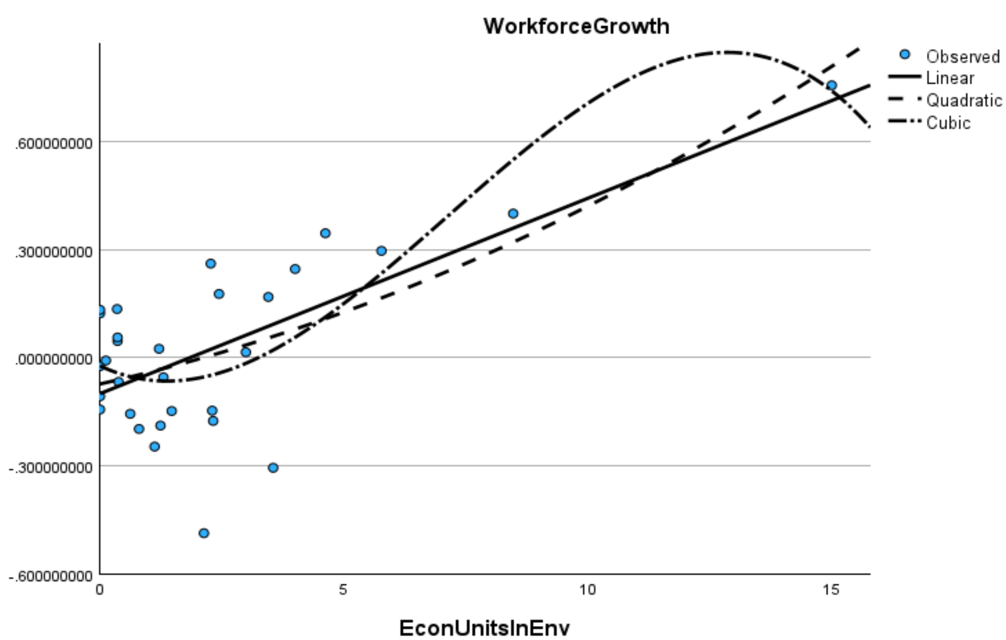
**Wage Growth: Highest R-squared with Cubic shape, with a total of 0.39.

Graph 4. Distribution of Financial Units vs Workforce Growth



**Distribution of Financial Units: Highest R-squared with Cubic shape, with a total of 0.011.

Graph 5. Economic Units in Environment vs Workforce Growth



**Economic Units in Environmental Staff: Highest R-squared with Cubic shape, with a total of 0.538.

The regression was run again after all variables were cubed. As a result, the r-squared appeared significantly higher, with a value of 0.924 (table 6). Meaning that 92 percent of the changes in Growth in Employment in the Agricultural Sector in Mexico can be explained by all independent variables. The regression had a significant statistical significance, where the p-value was less than .001. The shape of the regression further indicates that the relationship between the independent variables and growth in agricultural employment is non-linear, meaning that changes in the explanatory variables do not affect employment growth at a constant rate. Instead, their effects vary at different levels, suggesting the presence of increasing or diminishing marginal effects. This implies that small changes in the variables may have limited impact at low levels but much stronger (or weaker) effects as their values increase, capturing more complex dynamics in the agricultural labor market.

Table 6. Third Regression Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.961[a]	.924	.909	.02447163245

a. Predictors: (Constant), Distribution of Large Economic Units that Had Staff in Environmental Protection Activities by state, MonthlyGrowthWageAgriculture , CrimeRate per 100,000 inhabitants, DistributonFinancingUnits, LaborInformality (growth) Agricultue

The new coefficients of the regression showed higher statistical significance (table 7). However, they were still insufficient to be considered fully statistically significant, as the p-values remained greater than 0.05. These results suggest that it may be worth considering a one-tailed test for some variables, particularly those for which the theoretical framework predicts a positive relationship, such as wages, financing, and environmental investment. These variables are expected to increase agricultural employment growth or have no effect. However, for other variables, especially crime, it still makes sense to use a two-tailed test, since the expected effect could be negative. Moreover, the p-value for crime remains non-significant even after cubing the variable, indicating that testing both tails is appropriate in this case.

Table 7. Coefficients of Third Regression

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.003	.007		.403	.690
	CrimeRate per 100,000 inhabitants	-1.561E-17	.000	-.007	-.110	.914
	LaborInformality (growth) Agricultue	-1.694E-5	.000	-.182	-1.827	.079
	MonthlyGrowthWageAgriculture	.043	.023	.156	1.869	.073
	DistributonFinancingUnits	-3.312E-6	.000	-.157	-1.950	.062
	Distribution of Large Economic Units that Had Staff in Environmental Protection Activities by state	.000	.000	.840	9.236	<.001

a. Dependent Variable: Workforce Growth 2023(Agriculture)

ROBUSTNESS CHECK

Outlier Check

There were no cases of outliers, when tested by 3 standard deviations, suggesting that the dataset is free from extreme observations that could disproportionately influence the regression results. However, there was an outlier found, when testing for 2 standard deviations. The variable found was the dependent variable, Work Growth Force in the Agricultural Sector, which is shown in table 8.

Table 8. Fourth Regression Output.

Casewise Diagnostics[a]

Case Number	Std. Residual	WorkforceGrowth	Predicted Value	Residual
25	-2.184	-.487273526	-.09129260336	-.3959809226

a. Dependent Variable: WorkforceGrowth

These results might signal that another type of measure for the agricultural sector could be used in the regression analysis. However, 2 standard deviations difference, as an outlier, do not seem to have a great significance in the outcomes of the analysis. Thus, the variable is appropriately used.

Multicollinearity

No multicollinearity was detected among the independent variables. All correlation coefficients were below 0.6, indicating that multicollinearity is not a significant concern in the regression (Table 9).

Table 9. Correlation Matrix

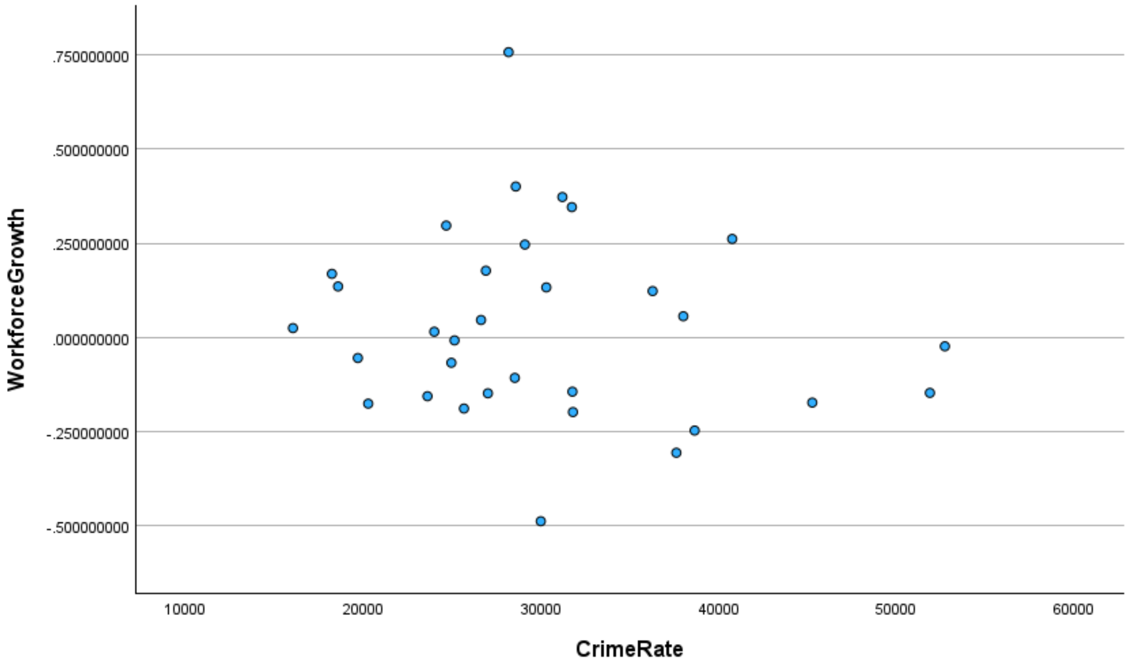
Correlations

		WorkforceGrowth	CrimeRate	LaborInformality	WageGrowth	DistributionFinUnits	EconUnitsInEnv
Pearson Correlation	WorkforceGrowth	1.000	-.115	-.370	.156	.045	.681
	CrimeRate	-.115	1.000	.400	.085	-.001	-.078
	LaborInformality	-.370	.400	1.000	.116	-.213	-.406
	WageGrowth	.156	.085	.116	1.000	.375	.095
	DistributionFinUnits	.045	-.001	-.213	.375	1.000	.327
	EconUnitsInEnv	.681	-.078	-.406	.095	.327	1.000

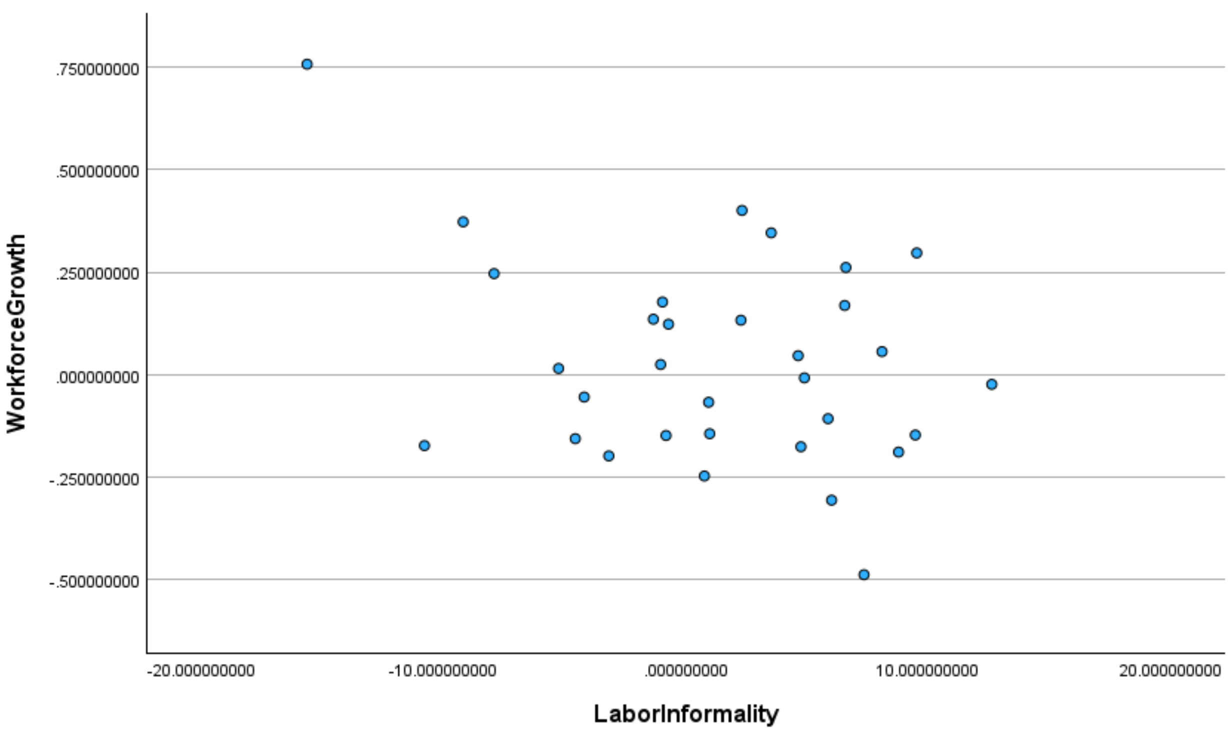
Heteroscedasticity

The tests of Heteroscedasticity were negative. It was first tested by the eyeball test, and the results were as shown in the following graphs,

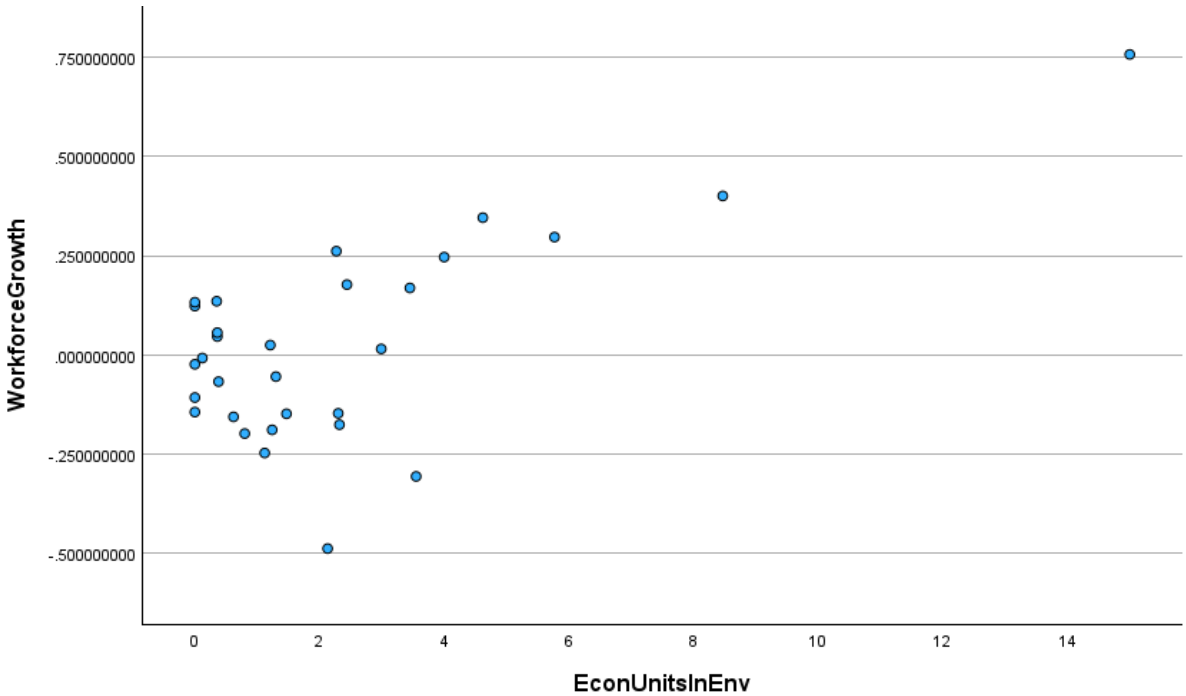
Graph 6. Workforce Growth vs Crime Rate



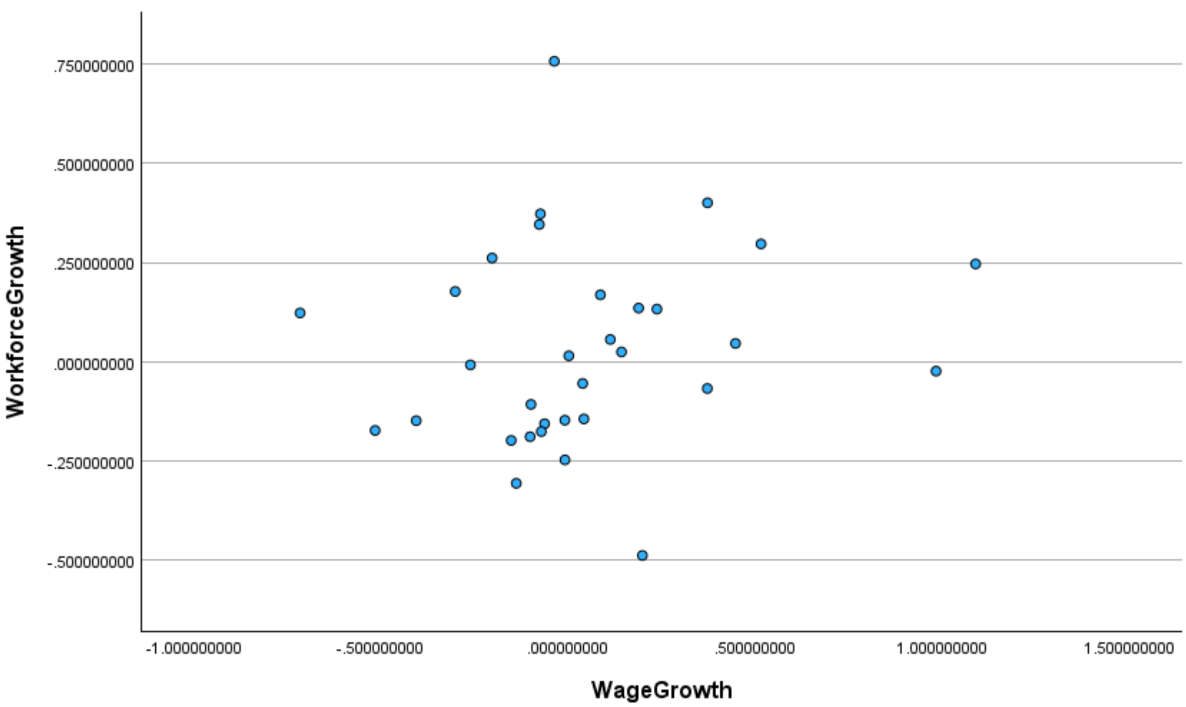
Graph 7. Workforce Growth vs Labor Informality



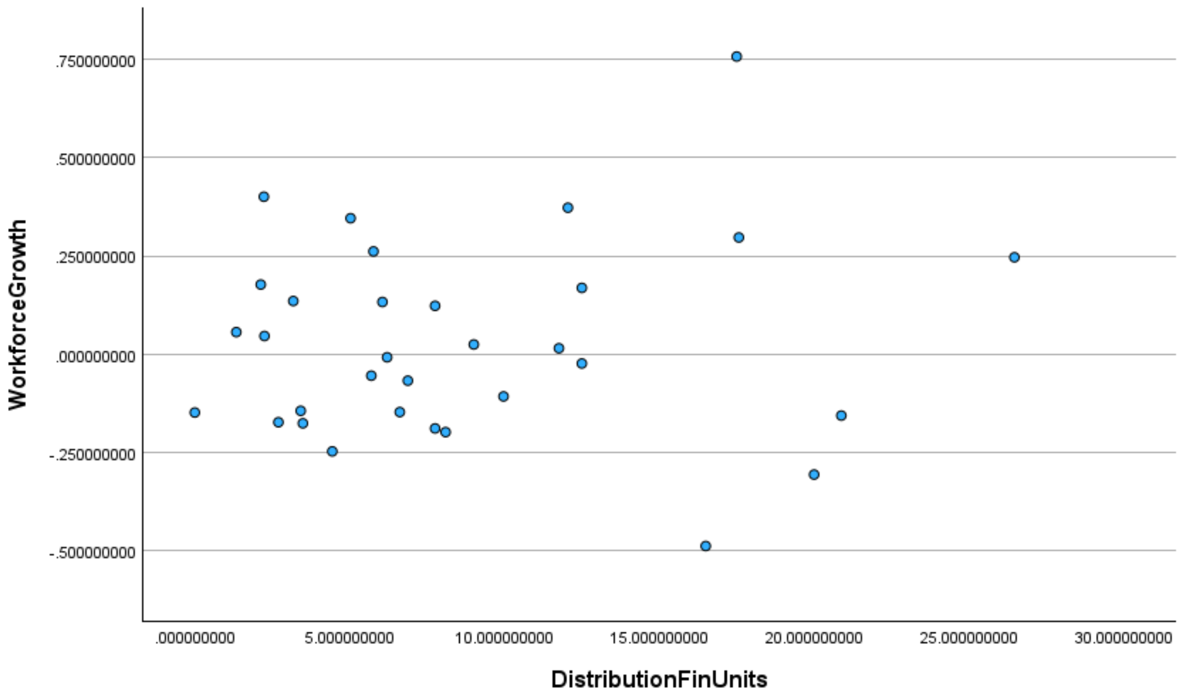
Graph 8. Workforce Growth Economic Units in Environment



Graph 9. Workforce Growth vs Wage Growth



Graph 10. Workforce Growth vs Distribution of Financial Units



The eyeball test suggests that the dispersion of the data points is uneven and exhibits some irregular patterns across the range of the explanatory variables. This indicates the possible presence of heteroscedasticity. To formally assess this issue, the White test was conducted. The results further showed that there was no heteroscedasticity, as shown in table 10, and table 11.

Table 10. White Test Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.496[a]	.246	.089	.09858753353

a. Predictors: (Constant), Distribution of Large Economic Units that Had Staff in Environmental Protection Activities by state, CrimeRate per 100,000 inhabitants, MonthlyGrowthWageAgriculture , DistributonFinancingUnits, LaborInformality (growth) Agricultue

Table 11. Coefficients of White Test Regression

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.064	.071		-.894	.380
	CrimeRate per 100,000 inhabitants	3.903E-6	.000	.331	1.699	.102
	LaborInformality (growth) Agricultue	-.003	.004	-.163	-.749	.461
	MonthlyGrowthWageAgriculture	-.104	.055	-.367	-1.872	.073
	DistributonFinancingUnits	.001	.003	.084	.410	.685
	Distribution of Large Economic Units that Had Staff in Environmental Protection Activities by state	-.007	.007	-.224	-1.106	.280

a. Dependent Variable: ABS(PRE) Squared

Heteroscedasticity was initially tested using the White Test approach, in which the squared residuals were regressed on the explanatory variables. The results did not provide strong evidence of heteroskedasticity, as none of the coefficients were statistically significant at conventional levels, and the r-squared seemed low. Thus, no action was taken.

Autocorrelation

Autocorrelation was not applicable, as it was a cross-sectional regression.

DISCUSSION AND RESULTS

The results of this study provide insight into the complex relationship between crime and agricultural employment growth across Mexico's 32 states in 2023. While the overall regression model is statistically significant, the findings indicate that no single independent variable, including the crime rate, independently drives changes in agricultural workforce growth. Instead, the results indicated there is an existence of a multifaceted set of structural, institutional, and economic forces shaping rural labor markets.

The baseline regression model explains approximately 55 percent of the variation in agricultural workforce growth, with an F-statistic significant at the $p < .001$ level. Despite the model's overall significance, each individual independent variable — crime rate, labor informality, wage growth, and distribution of financial units — failed to achieve statistical significance at the conventional 0.05 threshold. The crime rate coefficient was negative, consistent with the theoretical expectation that violence and insecurity reduce agricultural employment by increasing production risks, encouraging land abandonment, and reducing labor force participation. However, the lack of individual statistical significance indicates that crime does not operate alone. This finding is consistent with the broader literature on rural labor markets, where violence tends to interact with pre-existing structural vulnerabilities rather than being an independent driver of employment decline (Lazarte, 2017). Moreover, the results do not confirm the hypothesis that rising crime rates independently caused a decline in agricultural employment across Mexican states in 2023. However, they do not rule out that crime contributes as part of a broader set of pressures.

The labor informality variable also produced a negative coefficient, in line with expectations. States with higher shares of informal agricultural workers may face weaker institutional protection against external shocks, making their labor markets more susceptible to disruption. Thus, the absence of statistical significance at the individual level suggests that informality's effect may be conditioned by other state-level characteristics, such as access to financing or the degree of economic diversification.

Wage growth produced a positive coefficient, consistent with standard labor supply theory, which predicts that higher earnings attract and retain workers in a given sector (Blundell 1999). However, the coefficient's lack of significance likely reflects the low wages earned by agricultural workers in Mexico, even following the 2023 minimum wage increase. In a sector where earnings remain comparatively low, marginal changes in wages may not be sufficient to meaningfully alter employment trajectories at the state level.

The distribution of financial units produced an unexpected negative coefficient, suggesting that greater access to financing is associated with lower agricultural employment growth. Two interpretations are plausible. First, access to capital may facilitate the adoption of labor-saving technologies, enabling producers to substitute capital for workers and thus reduce employment demand. Second, improved financing may encourage labor reallocation, as workers with greater economic mobility are incentivized to exit agriculture in favor of higher-paying opportunities in other sectors.

The environmental staffing variable was the only individual predictor that achieved statistical significance in the cubic model, producing a large positive coefficient. This finding suggests that states where larger agricultural firms invest in environmental compliance staff tend to have stronger workforce growth. This may reflect that such firms have greater organizational capacity and institutional stability, enabling them to sustain or expand their workforces under adverse conditions.

The robustness check incorporating GDP distribution, the Gini coefficient, and homicide rates increased the model's explanatory power to 56 percent while preserving overall statistical significance. The negative coefficient on GDP share per state is consistent with the Lewis dual-sector model of structural transformation, in which economic development leads wealthier states to shift labor away from agriculture toward manufacturing and services (Lewis, 1954). Homicide rates and income inequality did not achieve individual significance, though their inclusion did not alter the direction or magnitude of the core coefficients, supporting the reliability of the baseline findings.

The most substantive improvement in model fit came from estimating cubic functional forms for all independent variables. This specification raised the R-squared to 0.924, indicating that the relationships between the independent variables and agricultural employment growth are highly non-linear. The effects of crime, informality, wages, and institutional factors are not constant across their ranges; rather, they intensify or diminish as their values change. While individual coefficients in the cubic model remained below the conventional significance level, several approached significance. The overall model performance provides strong evidence that linear specifications underrepresent the true complexity of these relationships.

No multicollinearity was detected, as all correlation coefficients among the independent variables remained below 0.6. Heteroscedasticity was measured via the eyeball test and formally confirmed using the White test; neither approach yielded evidence of non-constant variance in the residuals. A single outlier was identified at the two standard deviation threshold, corresponding to the dependent variable in state 25, but this observation did not meet the three standard deviation criterion and was retained in the analysis without materially affecting the results.

Taken together, these findings suggest that Mexico's agricultural labor market in 2023 was shaped by a combined structural factors, rather than a single driver. Violence, economic modernization, labor informality, and ongoing structural transformation appear to be simultaneously reshaping rural employment. Policy interventions targeting crime reduction alone are not sufficient to stabilize agricultural employment. Broader strategies addressing rural development, labor market formalization, and economic diversification are necessary to support the long-term resilience of Mexico's agricultural workforce.

CONCLUSION

The main objective of this study was to test whether increasing crime rates across Mexican states in 2023 contributed to a decline in employment within the agricultural sector. To do so, a cross-sectional regression model was estimated using data from all 32 states, including a range of economic, labor, and institutional control variables to isolate the effect of crime. The results of the main regression indicate that, while the model as a whole is statistically significant and significantly explains the variation in agricultural employment growth, the crime rate variable itself is not statistically significant. This suggests that, once other relevant factors are accounted for, crime alone does not have an independent effect on short-run changes in agricultural employment across states in 2023. Similarly, other independent variables, such as labor informality, wage growth, and access to financing, were not individually statistically significant, despite their theoretical relevance. Overall, these findings indicate that agricultural employment is likely influenced by a complex interaction of factors rather than by crime alone.

Although most individual coefficients lacked statistical significance, their estimated signs largely aligned with economic theory. In particular, the negative coefficient on crime is consistent with the hypothesis that violence and insecurity negatively impact agricultural labor by discouraging employment and increasing production risks. The unexpected negative coefficient on the distribution of financial units suggests that greater access to financing may facilitate capital–labor substitution or encourage labor reallocation away from agriculture toward non-agricultural sectors. This interpretation is consistent with the processes of structural transformation, where economic development leads to declining agricultural employment even as productivity increases.

To measure the significance of these findings, additional control variables were added, including GDP distribution, income inequality, and homicides rates. The inclusion of these variables slightly increased the r-squared of the regression (by 1%), while preserving the overall significance of the model. The negative coefficient on GDP share further supports that states with higher economic output may be transitioning away from labor-intensive agriculture toward more diversified economic structures, which is similar to the Lewis dual-sector model.

Further analysis revealed that the relationships between the independent variables and agricultural employment growth are likely non-linear. When cubic functional forms were applied, the model's explanatory power increased significantly. This suggests that the effects of crime, wages, informality, and institutional factors vary at different levels rather than affecting employment growth at a constant rate.

Overall, while this study does not show strong statistical evidence that crime alone directly reduces agricultural employment growth at the state level in 2023, the results underscore the structural pressures Mexico's agricultural sector faces. Violence, economic modernization, financial development, and institutional change appear to reshape rural labor markets in the country. These findings suggest that policy responses aimed at reducing crime alone may not be enough to stabilize agricultural employment. Extensive strategies to increase agricultural employment in Mexico need to also address rural development, labor informality, technological change, and economic diversification. As Mexico continues to face rising insecurity and structural transformation, understanding these interactions will be critical for ensuring the long-term stability and sustainability of its agricultural workforce.

REFERENCES

Blundell, R., & MaCurdy, T. (1999). Labor supply: A review of alternative approaches. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, pp. 1559–1695). Elsevier. https://public.econ.duke.edu/~vjh3/e262p/readings/Blundell_MaCurdy.pdf

Creel, García-Cuéllar, Aiza y Enríquez. (2022, December 13). Minimum wage 2023. https://www.creel.mx/en/publicaciones/client-alerts/minimum-wage-2023/

De Haro, I. (2025). Avocados: Mexico's green gold, drug cartel violence and the U.S. opioid crisis. World Development, 191, Article 106942. https://doi.org/10.1016/j.worlddev.2025.106942

Data México. (2023). *Data México: Economic and labor market data platform*. Secretaría de Economía. https://www.economia.gob.mx/datamexico/

Instituto Nacional de Estadística y Geografía. (2024, September 26). *Crime rate in Mexico in 2023, by state (per 100,000 inhabitants)* [Graph]. In Statista. https://www.statista.com/statistics/984146/mexico-crime-rate/

Instituto Nacional de Estadística y Geografía. (2024). *Crime incidence rate by state of occurrence per 100,000 inhabitants, 2023*. INEGI. https://en.www.inegi.org.mx/temas/incidencia/

Instituto Nacional de Estadística y Geografía. (2023). *Estadísticas de seguridad pública: Cifras de criminalidad por entidad federativa* [Crime statistics by state]. INEGI. https://www.inegi.org.mx/temas/incidencia/

Lazarte, A. (2017). U*nderstanding the drivers of rural vulnerability: Towards building resilience, promoting socio-economic empowerment and enhancing the socio-economic inclusion of vulnerable, disadvantaged and marginalized populations for an effective promotion of Decent Work in rural economie*s (Employment Working Paper No. 214). International Labour Office. https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_policy/documents/publication/wcms_568736.pdf

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, *22*(2), 139–191. https://la.utexas.edu/users/hcleaver/368/368lewistable.pdf

Mexico Solidarity Media. (2025, November 25). Mexican farmers occupy US border crossing, protesting agricultural policy. https://mexicosolidarity.com/mexican-farmers-occupy-us-border-crossing-protesting-agricultural-policy/

OECD. (2023). Mexico. In Agricultural policy monitoring and evaluation 2023: Adapting agriculture to climate change (pp. not paginated). OECD Publishing. https://doi.org/10.1787/b14de474-en

Secretaría de Economía. (2023). *Workers in agriculture: Wages, diversity, industries and labor informality* [Data set]. Data México. https://www.economia.gob.mx/datamexico/en/profile/occupation/trabajadores-en-actividades-agricolas?quarters5=20231&quarters6=20234

Secretaría de Economía. (2023). *Agriculture, animal production, forestry, fishing and hunting: Wages, production, investment, opportunities and complexity* [Data set]. Data México. https://www.economia.gob.mx/datamexico/en/profile/industry/agriculture-animal-production-forestry-fishing-and-hunting?optionsSelector1=Total+Investment#environmental-standard

Stevenson, M., & Verza, M. (2023, January 6). Violence breaks out after 'El Chapo' son detained, Mexican authorities say. PBS NewsHour. https://www.pbs.org/newshour/politics/violence-breaks-out-after-el-chapo-son-detained-mexican-authorities-say

Trading Economics. (2026). *Mexico unemployment rate*. Trading Economics. https://tradingeconomics.com/mexico/unemployment-rate

Trejo, R. (2024, September 24). Grain and feed update: Mexico (Report No. MX2024-0044). U.S. Department of Agriculture, Foreign Agricultural Service. https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Grain%20and%20Feed%20Update_Mexico%20City_Mexico_MX2024-0044.pdf

Valle-Jones, D. (2026, March). March 2026 Mexico crime report. El Crimen. https://elcri.men/en/

Voiland, A. (2024, June 10). *Drought parches Mexico*. NASA Earth Observatory. https://science.nasa.gov/earth/earth-observatory/drought-parches-mexico-152908/

APPENDIX

Table 12. Data used for regression analysis

State	Workforce Growth 2023(Agriculture)	CrimeRate per 100,000 inhabitants	LaborInformality (growth) Agriculture	MonthlyGrowthWageAgriculture	DistributionFinancingUnits	DistributingUnitsOfEnvironmentalStaff
Aguascalientes	-0.172683414	45,262	-10.78543483	-0.515543198	2.700000048	
Baja California	0.372338619	31,198	-9.187904526	-0.073918981	12.05000019	
Baja California Sur	-0.188461538	25,661	8.779273649	-0.101546203	7.760000229	1.24000001
Campeche	-0.197796097	31,793	-3.177726966	-0.151948745	8.100000381	0.800000012
Chiapas	0.02484105	16,038	-1.039265023	0.142766392	9.010000229	1.210000038
Chihuahua	0.246325031	29,090	-7.912816047	1.089083279	26.46999931	4
Ciudad de Mexico	-0.023466361	52,723	12.62676302	0.983275212	12.5	0
Coahuila de Zaragoza	0.296688544	24,661	9.529852085	0.515628021	17.56999969	5.769999981
Colima	-0.107118851	28,522	5.869511634	-0.099090623	9.970000267	0
Durango	0.168672028	18,230	6.556359359	0.086346343	12.5	3.450000048
Estado de Mexico	0.046334219	26,620	4.635679257	0.447363837	2.25	0.360000014
Guanajuato	0.015238237	23,989	-5.251310781	0.002007103	11.76000023	2.99000001
Guerrero	0.177060081	26,894	-0.962377765	-0.301554835	2.130000114	2.440000057
Hidalgo	0.34561389	31,731	3.523354483	-0.076829691	5.03000021	4.619999886
Jalisco	-0.146905552	51,881	9.470631511	-0.008532772	6.619999886	2.299999952

Michoacan de Ocampo	-0.054469807	19,688	-4.193762079	0.039077668	5.699999809	1.299999952
Morelos	0.123015099	36,278	-0.714860072	-0.716141616	7.760000229	0
Nayarit	-0.155784979	23,607	-4.561271399	-0.062710521	20.87999916	0.620000005
Nuevo Leon	0.756434331	28,171	-15.62853111	-0.036711425	17.5	15
Oaxaca	0.135224074	18,575	-1.339456271	0.188505299	3.180000067	0.349999994
Puebla	-0.24675997	38,642	0.765495144	-0.008283992	4.440000057	1.120000005
Querataro	0.261295144	40,755	6.603968309	-0.202996074	5.769999981	2.269999981
Quintana Roo	0.132768855	30,294	2.274195068	0.237427546	6.059999943	0
San Luis Potosa	-0.143724075	31,764	0.985570488	0.042348033	3.420000076	0
Sinaloa	-0.487273526	29,984	7.353467303	0.198569713	16.5	2.130000114
Sonora	-0.305677007	37,613	6.016741162	-0.138388282	20	3.549999952
Tabasco	0.056240051	38,004	8.097096756	0.112907695	1.340000033	0.360000014
Tamaulipas	-0.007752964	25,129	4.893708258	-0.260894001	6.210000038	0.119999997
Tlaxcala	0.400232658	28,576	2.319702446	0.372641011	2.230000019	8.470000267
Veracruz de Ignacio de la Llave	-0.175488988	20,272	4.750219651	-0.071419738	3.49000001	2.319999933
Yucatan	-0.067083298	24,952	0.94175506	0.372073849	6.880000114	0.379999995
Zacatecas	-0.148042377	27,007	-0.817666304	-0.405765571	0	1.470000029

Table 1. Summary of Variables used

Variable	Description	Source	Anticipated Slope	Anticipated Statistical Significance	Anticipated Practical Significance
Y Workforce Growth in Agricultural Sector in the Mexico, by state	Change in the number or proportion of individuals employed within the agriculture industry over the year 2023, by state.	DATA MEXICO	-	-	-
β_2 Crime Rate by State of occurrence per 100,000 habitants	Proxy for violence, cartel activity, and rural insecurity affecting farming labor in 2023, by state.	INEGI	Negative	Yes	High
β_3 Labor Informality of Agricultural Sector	Share of agricultural workers lacking formal contracts or social protections in 2023, by state.	DATA MEXICO	Negative	Yes	High
β_4 Annual Growth Wage in the Agricultural Sector	Measurement of the average change in earnings for workers in the agriculture industry over a one-month period in 2023, by state.	DATA MEXICO	Positive	Yes	Low
β_5 Distribution of Economic Units that Received Financing (by State)	Measures the percentage breakdown of businesses (economic units) that successfully obtained financing in 2023, by state.	DATA MEXICO	Positive	Yes	High
β_6 Distribution of Large Economic Units that Had Staff in Environmental	Percentage distribution of economic units in the private and parastatal sectors of	DATA MEXICO	Positive	No	-

Protection Activities by state	Agriculture, Animal Production, Forestry, Fishing and Hunting that had staff in environmental protection activities by state.				

ANALYSIS OF CRIME AND INEQUALITY IN LATIN AMERICA

Sidney Tovar, Siena University
Dr. Elias Shukralla, Siena University – Faculty Mentor

ABSTRACT

Latin America remains one of the most violent regions in the world, often cited as having the highest homicide rate globally. The regional homicide rate is roughly three times higher than the global average, dominating the list of the world's most dangerous areas. Of the top 50 most violent cities in the world, 42 are in Latin America (World Bank, 2025). The high crime rates in the region impose severe social and economic costs on its populations, aggravating the urgency to identify their structural dependency for change within their policy. This study examines the relationship between income inequality and crime in Latin America from 1990 to 2025, using panel data and Ordinary Least Squares regression with multiple robustness checks. Income inequality, measured through the Gini coefficient, serves as the primary independent variable, while intentional homicide rate per 100,000 people serves as the dependent variable. Drawing on relative deprivation theory, social resistance frameworks, and rational choice theory, this study argues that income inequality is the primary structural driver of crime in the region. While the baseline model produces an insignificant Gini coefficient due to data limitations and simultaneity concerns, curve estimation reveals a non-linear, self-reinforcing relationship between inequality and crime. Across robustness checks, the interaction term between inequality and corruption perception emerges as the most consistent predictor of homicide rates, confirming that inequality is especially criminogenic when compounded by institutional illegitimacy. Enrollment in secondary school consistently suppresses homicide rates across all specifications. These findings suggest that reducing crime in Latin America requires simultaneous interventions targeting income inequality, institutional corruption, and educational access, rather than conventional law enforcement approaches alone.

INTRODUCTION

Security in Latin America represents a persistent and multifaceted crisis that has gained international attention following the 2024 United States presidential election. While the region's rich culture, customs, and traditions represent an invaluable part of the world's heritage, the safety of its citizens remains a challenge the rest of the world can no longer afford to ignore. According to the *Legal Information Institute,* crime is a "behavior, either by act or omission, defined by statutory or common law as deserving of punishment or penalty" (Legal Information Institution, 2025). Crime is considered to be of significant importance as it deteriorates quality of life, by directly causing physical, psychological, and imposing significant economic costs to businesses, and society. Due to significant data limitations, this analysis implements crime through the metric of intentional homicide. The *World Bank* defines intentional homicide as an "unlawful death inflicted upon a person with the intent to cause death or serious injury" (World Bank, 2023). Despite representing only 8% of the global population (WorldMeters), Latin America continues to bear a disproportionate burden of violence, accounting for one third of the world's homicides (World Bank). This pattern of violence does not exist on its own. Scholars and policymakers have long identified economic inequality as a structural driver of crime. Latin America consistently ranks among the most unequal regions globally. Income inequality, commonly measured through the Gini coefficient, highlights the deep economic divides that pervade the region and have long been tied to its persistently high levels of violent crime. This study examines how income inequality drives crime in Latin America, drawing on relative deprivation theory, social resistance frameworks, and rational choice perspectives to assess the mechanisms through which economic disparity causes criminal behavior. Besides highlighting the nexus between inequality and crime, this research also adds value to the existing literature by introducing social media as a theoretical variable that amplifies the inequality-crime relationship, by increasing exposure to visible income gaps and perceptions of relative deprivation among marginalized groups. While data limitations existed for formal inclusion of the variable in regression models, the theoretical framework in this study provides a foundation for future empirical investigation. Overall, this review supports income inequality as the primary structural drive of crime in Latin America, reinforcing the urgency for structural policy intervention.

BACKGROUND

Security Crisis in Latin America

Latin America's relationship with violence is disproportionate to its population size. Despite accounting for approximately 8% of the world's population, the region bears responsibility for roughly one third of all global homicides (World Bank, 2025). This disparity positions Latin America as a persistent outlier in global security analyses, and signals that the drivers of violence in the region are not incidental but deeply structural. Understanding the scale of this problem is a crucial starting point for examining how deeply rooted economic and social conditions drive violence across the region.

Measuring crime in Latin America presents significant methodological challenges. One of the most pressing obstacles is underreporting data, driven by inconsistencies in data collection across countries. Not all countries collect the same data, and what is collected is often incomplete or inaccurate due to the limitations of self-reporting. Homicides are similarly distorted, as a death is often not officially classified as an intentional killing until a legal process confirms it. This process in many Latin American countries is delayed or never completed at all. More broadly, Perez-Vincent et al. (2024) find that approximately three out of four crimes go unreported across the region, including homicides, which skews official statistics and hinders effective law enforcement responses. Their study *Crime Underreporting in Latin America and the Caribbean* identifies several structural factors that discourage reporting, including material costs, social and psychological barriers, such as shame and social conformity, personal safety concerns, legal risks, and low confidence in the efficacy of law enforcement and judicial institutions. Moreover, sociodemographic factors such as ethnicity and immigration status also influence the likelihood of reporting. Immigrants and minority groups are particularly less likely to report crimes due to fears of legal repercussions or discrimination, further distorting available data. Together, these factors significantly limit the reliability of available crime data, making it difficult to conduct consistent cross-national comparisons.

Violence in Latin America has undergone a significant transformation over the past several decades, yet the crisis remains unresolved. During the mid-to-late twentieth century, much of the region's instability was rooted in state-sponsored conflict, including civil wars, military dictatorships, and guerrilla movements (Brands, 2010). Once these political conflicts began to end, organized criminal groups seized control of territories abandoned by the state, embedding themselves into local economies, and governance structures. Today, this violence has filtered into everyday life across the region. In Mexico, clashes between security forces and armed groups doubled in the first eleven months of 2025 compared to the same period in 2024 (ACLED, 2025), driven in large part by cartel fragmentation and territorial disputes. In Colombia, an offensive launched by the National Liberation Army against Revolutionary Armed Forces of Colombia-People's Army dissident factions resulted in at least 52 confirmed deaths and forced more than 52,000 people to flee their homes (ACLED, 2025). This marked the deadliest month in the Norte de Santander region since records began. In Ecuador, clashes between rival gangs inside a prison in El Oro province left 32 gang members dead in a single day (ACLED, 2025), underscoring the extent to which organized crime has penetrated even state institutions. Throughout the region, the expansion of armed groups has increasingly undermined governmental authority and destabilized local communities (ACLED, 2025), with drug trafficking, corruption, and related migration crises remaining persistent threats to all citizens and institutions (OAS Secretariat, 2026). The 2024 United States presidential election brought these issues into international focus, as immigration, drug trafficking, and border security – all directly tied to Latin American instability – became central themes of the campaign (Americas Quarterly, 2024). The election's outcome raised significant concerns across the region, as proposals for mass deportations and trade tariffs threatened to reshape U.S. relations with Latin American governments (Inter-American Dialogue, 2024), making it increasingly urgent for scholars and policymakers to understand the structural roots of the region's ongoing violence.

<u>Income Inequality in Latin America</u>

Latin America consistently ranks among the most unequal regions in the world (Inter-American Development Bank, 2024), This distinction that has persisted across decades despite periods of economic growth. As the *Inter-American Development Bank* (2024) notes, inequality in the region has fluctuated considerably over time, rising rapidly in the 1970s, peaking in the 1990s, and then gradually declining. However, since 2014 income inequality has plateaued, leaving the region still among the most unequal globally. This pattern suggests that while some progress has been made, the structural conditions driving economic inequality remain deeply rooted in the system, and difficult to change.

Income inequality in the region is most commonly measured through the Gini coefficient, a statistical tool that captures the degree of income distribution across a population. The *World Bank* defines it as the "extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution "(World Bank, 2025), where a score of zero represents perfect equality and a score of one represents maximum inequality. Therefore, a higher Gini coefficient signals a wider gap between the wealthiest and poorest segments of society. When applied to Latin America, this measure reveals an economic divide that persists across nearly every country in the region. Countries such as Brazil, Colombia, Honduras, and Mexico consistently record some of the highest Gini coefficients globally, reflecting how capital accumulation is unequal, rewarding only top earners while reducing the purchasing power of lower- and middle-income households (World Bank, 2025). Despite periods of modest economic progress in the early 2000s, reductions in poverty have not yet narrowed the income gap between the rich and the poor. As of 2024, 25.5% of Latin America's population continued to live in conditions of income poverty (ECLAC, 2025), underscoring that economic gains have remained concentrated at the top rather than reaching those most in need, as seen on table 1.

Inequality in Latin America has a deep historical context. The majority of the region's economic disparity can be traced with the conquest of European powers during the colonial period, in which the economic system concentrated land, wealth, and political power in the hands of a small elite, leaving the natives with small gains in income, and almost null social mobility. As Montgomery (2024) notes, the initial conditions established during the colonial period – including labor exploitation, land concentration, and legal inequality – had a lasting impact on the distribution of resources and economic opportunities, as these colonial legacies continue to shape patterns of economic development and social inequality across the region. Similarly, Acemoglu and Robinson (2012) argue that colonialism created extractive institutions in Latin America that produced poor long-run development outcomes, which persists today. Land concentration in particular has remained a defining and destabilizing feature of Latin American economies. The original distribution of territories in colonial times followed practices such as the encomienda system, a concentration of land in very few hands, which established an unequal distribution of resources that has carried forward into the modern era (World Bank, 2023).

This historic legacy of economic disparity endures today in Latin America. Oxfam (2016) found that just one percent of large landholdings in Latin America now control more productive land than the remaining 99 percent combined. Beyond land, labor market segmentation has further entrenched inequality, as formal and informal employment sectors operate similarly. According to the ILO, approximately 53% of Latin American workers were in informal employment in 2016, a situation that exposes workers to severe vulnerability in terms of income, working conditions, and access to labor rights and social protection (ILO, 2018). Marginalized groups bear a disproportionate share of this burden. The *Inter-American Development Bank* reports that on average 43% of the indigenous population and 25% of Afro-descendant people in the region live in poverty, with Afro-descendants earning wages approximately 17% lower than the rest of the population even after adjusting for education (IDB, 2020). Together, these structural conditions have created a self-reinforcing cycle in which economic disadvantage is passed down across generations, making inequality in the region particularly difficult to dismantle through short-term policy measures alone.

BRIEF LITERATURE REVIEW

The Inequality-Crime Nexus

The relationship between income inequality and crime requires a theoretical framework to understand the mechanism through which economic disparity transforms into criminal behavior. Scholars across disciplines have proposed various frameworks to explain this relationship, each illustrating a different dimension of how inequality shapes individual motivations, social dynamics, and institutional trust. This section examines four theoretical perspectives that together provide a comprehensive account of the inequality-crime nexus, including relative deprivation theory, social resistance frameworks, and rational choice theory. While each framework approaches the relationship from a different angle, they collectively converge on a shared conclusion: income inequality is not merely correlated with crime, but it is its structural driver. In the Latin American context, where economic disparities are among the highest in the world, these frameworks are practical tools for understanding this ongoing crisis.

a. *Relative Depravation Theory*

Relative Deprivation refers to the perceived gap between one's own economic position and the others in society. In other words, relative deprivation is the perception of being worse-off compared to a reference group or expectation, leading to resentment and discontent. This theoretic framework offers an important explanation for the relationship between income inequality and crime in Latin America. As Kim et al note in *A Systematic Review and Meta-analysis of Income Inequality and Crime in Europe: Do Places Matter?*, "high levels of relative deprivation produced by income inequality are ultimately responsible for high crime rates" (Kim et al., 2020), as the perception of an unjust system creates the feeling of taking fairness into one's own hands. This dynamic is particularly pronounced in the Latin American context, where extreme wealth concentration exists alongside visible poverty, making social comparisons and amplifying feelings of injustice and exclusion among disadvantaged populations. As a result, it is the perception of unjust inequality, rather than poverty alone, that emerges as the more powerful driver of criminal behavior in the region.

Furthermore, the economist Kelly (2000) establishes a foundational cause chain in this regard, arguing that "inequality is associated with crime because it is linked to poverty: areas with high inequality tend to have high poverty rates." However, the mechanism extends beyond material hardship alone. Kelly further contends that "unsuccessful individuals become alienated from society and commit crime in response" (Kelly, 2000), suggesting that it is the psychological experience of exclusion, and economic deprivation, that translates inequality into criminal behavior. This alienation is particularly important in highly unequal societies, where individuals who perceive themselves as structurally excluded from legitimate pathways to prosperity become disconnected from institutional norms and social bonds.

Itskovich deepens this analysis by highlighting the subjective dimension of inequality, arguing that "economic inequality has a subjective essence, which is tied with how people experience, perceive, and react to observable disparities in wealth" (Itskovich, 2023). Individuals who perceive the wealth gap as unjust or illegitimate are more likely to feel alienated from social institutions, regardless of their absolute income level. In Latin America, where daily exposure to economic disparities is pervasive, this subjective experience of inequality is especially pronounced. Furthermore, Kim et al. find that "income inequality is a more critical determinant of homicide than poverty" (Kim et al., 2020), directly reinforcing the argument that it is the structure of income distribution that drives the region's elevated homicide rates. Overall, these findings suggest that the perception of income inequality leads to higher criminal behaviors, underscoring the urgency to implement policies that address deep structural inequalities to lower relative deprivation, rather than relying only on anti-poverty interventions.

b. *Social Resistance Frameworks*

While relative deprivation theory focuses on the psychological experience of inequality, social resistance frameworks offer a complementary sociological explanation, positioning crime not merely as a symptom of individual frustration but also as a form of collective resistance against a perceived unjust social order. Itskovich (2023) illustrates this argument by contending that "economic inequality alienates individuals from the institutions and values of society, and this, in turn, leads them to resist these institutions and values by, among other things, engaging in criminal behavior." Within this framework, criminal behavior is not understood as a random or purely an individual act, but instead as a rational response to a system experienced as exclusionary and corrupt.

This process of alienation does not remain at the individual level. Over time, Itskovich (2023) argues that "alienation from society promotes the development of a collective identity that contrasts with the dominant group's values and institutions," suggesting that long-term structural inequality can give rise to oppositional group norms that institutionalize immoral behavior within marginalized communities. In the Latin American context, where inequality often overlaps with racial, ethnic, and class-based hierarchies, this dynamic is particularly relevant. Itskovich (2023) further notes that "scholars have argued that economic inequality inherently entails discrimination," and that when marginalized groups perceive their exclusion as rooted in systemic bias, their willingness to abide by institutional rules weakens considerably.

c. *Rational Choice Theory*

Rational choice theory further adds complexity to the explanation of how income inequality translates into criminal behavior, by examining the framework in which individuals always make logical, self-interested decisions that maximize their utility, while minimizing costs from structural conditions. While social resistance frameworks highlight collective and institutional dimensions of the income-crime nexus, rational choice theory explains how inequality systematically skews the cost-benefit calculations of economically excluded individuals. As Sugiharti et al. (2023) argue in *The Nexus between Crime Rates, Poverty, and Income Inequality: A Case Study of Indonesia,* "individuals disadvantaged by income inequality often face choices between staying in the legal domain or venturing into the illegal one," framing criminal activity not as moral failure, but as a rational response to the constrained opportunity structures that continuous inequality creates.

When legitimate pathways to economic advancement are systematically blocked, illegal alternatives become comparatively more rational. This logic is further reinforced by the structure of inequality itself, where "income inequality may trigger crimes because there are more upper-income level individuals who can become targets," (Sugiharti et al, 2023) meaning that as wealth concentrates, the expected returns from criminal activity increase alongside the perceived costs of remaining in poverty. This thinking is overall reinforced in the Latin American context, where the gini coefficient rank among the highest in the world. Schargrodsky and Freira (2023) find that homicides in the region "are concentrated among the poor," portraying how the most extreme criminal violence is concentrated in areas where high inequality undertakes avenues for legitimate economic participation for citizens. Taken together, rational choice theory reveals that income inequality does not merely correlate with crime, but actively restructures the decision-making environment of marginalized individuals, making criminal behavior a predictable outcome of persistent structural disparity.

<u>The Role of Social Media</u>

After the emergence of Social Media in the world, there was a social transformation, where an instant global connectivity was introduced. This event reshaped communication norms, as "access to the internet and social media has made people in their country more informed about domestic current events," (Wike et al, 2022) breaking all geographical barriers. This shift is particularly relevant to the study of crime, as social media has introduced new channels through which criminal norms, group identities, and perceptions of inequality increase. Curiel et al. (2020) find in *Crime and its fear in social media,* that "traditional media shows a strong bias towards violent crime," and that social media displays a similar bias, disproportionately circulating content related to violence. This framing effect matters, as repeated exposure to crime-related content normalizes violence and may reinforce the oppositional group identities discussed in the social resistance framework, where marginalized individuals already alienated from institutions encounter social media environments that further validate criminal behavior as a rational response to inequality. In Latin America, where inequality is already a structural driver of crime, social media functions as an accelerant exposure to criminal networks, increasing relative deprivation through visible wealth comparisons, shaping how individuals weigh the costs and benefits of unlawful behavior. Ideally the effect of social media would be captured through an interaction term between income inequality and social media usage. However, while cross-national data on social media usage is available, it only extends back to 2021, providing insufficient variation to construct a meaningful panel regression. Future research exploring this interaction term could show the extent to which social media accelerates the inequality-crime relationship, particularly in Latin America.

METHODOLOGY

This study employs Ordinary Least Squares (OLS) to pooled data, combining cross-sectional and time series observations to examine the relationship between crime rates and Income Inequality across Latin America, from 1990 to 2025.. Data from multiple Latin American countries observed over time are stacked together and the OLS estimator then applied to estimate the impact of inequality on crime, controlling for other determinants of crime.

The baseline model is as follows,

$$IntentHomicide_{it} = \beta_1 + \beta_2 Gini_{it} + \beta_3 Unemployment_{it} + \beta_4 IncomePerCapita_{it} + \beta_5 ProgSecSchool_{it} + \beta_6 RuleOfLaw_{it} + \beta_7 Inflation_{it} + \varepsilon_{it} \quad (1)$$

Where i indexes each of country and t indexes time (year), β_1 is the intercept, β_2 through β_7 are the estimated slope coefficients for each independent variable, and ε represents the error term.

To assess the robustness of the baseline results, a second regression model is estimated controlling for three additional variables: Corruption Perception, Government Effectiveness, and Enrollment to Secondary School by country, from 1990 to 2025. A few control variables were additionally added, such as the strength of legal institutions affecting income distribution – by combining rule of law and the gini index – and the interaction term between gini index and corruption perception, to measure crime under economic disparity given an illegitimate system perceived by the population. The inclusion of these variables tests whether the main findings hold after accounting for broader macroeconomic and institutional conditions.

It is important to acknowledge some key limitations of this methodology. First, while the use of panel data strengthens the analysis by controlling for both country and time specific effects, the causality between income inequality and crime may not be fully established. Secondly, data restrictions do not allow to fully assess the true impact of income inequality on the dependent variable, as crime is assessed with a different crime indicator. Further studies should analyze the different variations of crime, such as property crime, robbery, aggravated assault, cyber-robbery, or crime severity index.

Data and Variable Selection

This study aims to demonstrate the impact of income inequality into crime rates, especially in Latin America. The study further analyzes panel data across countries in Latin America from 1990 to 2025. All the initial data is collected by the World Bank Group, which is publicly available. The summary of the data selection is registered in table 2.

The dependent variable is the annual rate of intentional homicides per 100,000 people by country, measured as the number of unlawful, intentional killings per 100,000 inhabitants in a given country and year. This variable serves as the primary proxy for crime, capturing lethal violence as the most consistently reported and comparable indicator of criminal activity across countries and over time. The intentional homicide rate was chosen as the crime rate indicator because data constraints make other crime indices less suitable for cross-country analysis.

The primary independent variable is the Gini Index (β_2) by country, which measures the gap between the actual income distribution and a perfectly equal one. This variable serves as the central proxy for income inequality in the regression model. From a theoretical perspective, higher inequality is expected to increase homicide rates, as societies with greater income disparities tend to exhibit weaker social cohesion, higher relative deprivation, and stronger incentives for criminal behavior among those excluded from economic opportunity. Based on this reasoning, a positive relationship between the Gini Index and intentional homicides is anticipated, with both high statistical and high practical significance.

Five control variables are included in the model to isolate the effect of inequality on crime, and account for the other structural factors that influence the dependent variable. The first was the unemployment rate (β_3), which measures the percentage of the labor force that has actively sought work in the past four weeks, but found no success. This variable is included in the baseline as individuals without stable income face an increased economic pressure, and social marginalization, which can act as an incentive to turn into criminal activity. Thus, their relationship is expected to be positive, with high statistical and practical significance.

The next three variables are income per capita (β_4), progression rate to secondary school (β_5), and rule of law (β_6). Income per capita is measured by dividing the total income of GDP by the total population, and is adjusted for purchasing power parity. It has an important relation to crime, as lower income increases economic desperation, lowering the opportunity cost of committing offenses. Schooling is measured as the percentage of students who advance to secondary school after completing primary school. This has a direct impact on crime, as greater education is associated with improved economic prospects, and reduced likelihood of criminal behavior. The rule of law refers to the extent to which agents trust and abide by society's rules, mostly focusing on property rights, contract enforcement, police, and courts. This last variable ensures that juridical systems are equally applied, helping prevent crime through reliable institutions. All three variables are expected to have a negative relationship with crime, with a high statistical and practical significance.

Finally, the fifth variable is inflation (β_7), which represents the annual rate at which consumer prices increase. Inflation influences crime because rising prices diminish real purchasing power. This increasing economic pressure creates an environment where lower-income households may be more likely to commit criminal offenses as an attempt to secure basic goods and services. This variable is thus expected to have a positive relationship with the dependent variable, with a high statistical and practical significance.

The time frame for measurement is 1990 to 2025, which was dictated by data availability. Most Latin American countries began consistently tracking the necessary data around 1990, establishing this as the initial intersection point for analysis. Future research should consider narrowing or expanding time periods, to isolate specific historic periods, as heterogeneity of political and economic conditions across Latin America might introduce noise into cross-national comparisons.

EMPIRICAL FINDINGS

After running the initial linear regression, the model had an R-squared of .415, indicating that the model explains approximately 41.5% of the variance in homicide rates, across the sample (table 3). The results further suggested that income inequality had a negative relationship with crime (table 4), with no statistical significance (P-value > 0.05). This result might hold true due to data limitations for the gini values, where not all information is complete. It could also be due to a lagging effect, where inequality affects crime in a lagging manner. Another justification could be a simultaneous or reverse causation, where crime also reproduces inequality, distorting the coefficient direction. Data limitations can be an explanation for low statistical significance, where not all data was complete, or continuous.

Income per capita shows a strong and statistically significant negative relationship with homicides ($\beta = -0.452$, $p < 0.001$), suggesting that higher income levels are linked to lower crime rates. Similarly, rule of law is negatively and significantly associated with homicides ($\beta = -0.263$, $p = 0.016$), indicating that stronger institutions contribute to reduced crime. Inflation also exhibits a significant negative effect ($\beta = -0.245$, $p = 0.002$), while unemployment is positively and significantly related to homicide rates ($\beta = 0.204$, $p = 0.020$), implying that higher unemployment may increase crime. In contrast, the Gini coefficient and secondary school enrollment are not statistically significant predictors in this model ($p > 0.05$), suggesting that their independent effects on homicide rates are not clearly supported in this specification.

Table 4. Coefficients of First Model

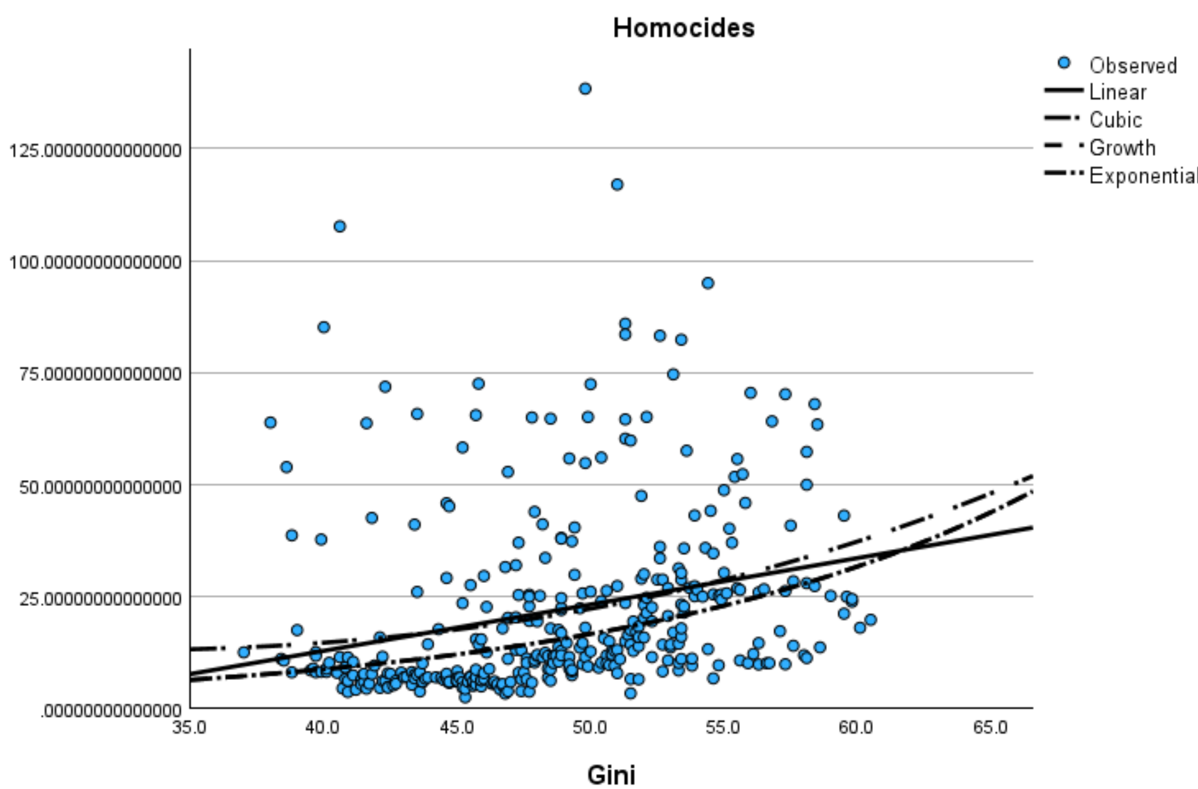
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	86.747	27.072		3.204	.002
	Gini	-.430	.407	-.083	-1.056	.293
	Unemployment	1.424	.601	.204	2.367	.020
	IncomePerCapita	-.002	.000	-.452	-4.562	<.001
	ProgSecSchool	-.283	.218	-.104	-1.298	.197
	RuleLaw	-9.548	3.899	-.263	-2.449	.016
	Inflation	-.547	.176	-.245	-3.102	.002

a. Dependent Variable: Homocides

To further explore the relationship between income inequality and crime, a curve estimation was conducted to assess whether the association is better characterized by a linear or non-linear functional form. The results indicated that a growth and exponential curve better fit the crime-inequality relationship, with an R-squared value of .167 ($F = 71.788$, $p < .001$), approximately .10 higher than the linear curve. This suggests that a meaningful portion of the variance in crime is explained by the non-linear model that a simple linear regression fails to capture. The result highlights that small increases in inequality at high Gini levels produce disproportionately large increases in crime (graph 1). As the relative deprivation theory suggests, as inequality widens, social comparisons become more extreme, so the psychological damage compounds rather than grows steadily. Similarly, by the rational choice theory, the curve supports that as wealth concentrates, the returns from targeting wealthy individuals grow exponentially, increasing criminal incentives at an accelerating rate. Overall, the curve estimation reinforces that the relationship between income inequality and crime in Latin America is self-reinforcing, growing more severely as inequality deepens.

Graph 1. Curve estimation for crime-inequality relation



Nonetheless, while the curve estimation suggests a non-linear functional form may better capture the inequality-crime relationship, the subsequent panel data analysis retains the linear specification to maintain methodological consistency and comparability across models

ROBUSTNESS CHECK

Extended Control Variables

To assess the robustness of the baseline model, several additional specifications were estimated. First, three additional control variables were introduced: Enrollment in Secondary School, Government efficacy, and Corruption Perception (table 6). Furthermore, an interaction term was constructed by multiplying the Gini index and the Corruption Perception index (Control Variable), with the purpose of capturing whether the effect of income inequality on crime is conditioned by the level of perceived institutional corruption. Income inequality is hypothesized to have a greater impact on crime in environments characterized by high corruption. Additionally, an attempt was made to substitute the dependent variable with an alternative crime measure to reduce reliance on a single proxy; however, due to data limitations, intentional homicide was retained as the primary dependent variable across all specifications.

The robustness model is specified as following,

$$IntentHomicide_{it} = \beta_1 + \beta_2 Gini_{it} + \beta_3 Unemployment_{it} + \beta_4 IncomePerCapita_{it} + \beta_5 Inflation_{it} + \beta_6 EnrSecSchool_{it} + \beta_7 CorrPercept_{it} + \beta_8 ControlVar_{it} + \beta_9 GovEffectivness_{it} + \varepsilon_{it}$$

The robustness check results largely support the central argument of this paper. The model demonstrates strong overall fit, with an R-squared of .979 and high statistical significance ($p < .001$), indicating that the specified variables collectively explain approximately 97.9% of the variance in homicide rates across the sample (Table 7).

The Gini index remains positive and statistically significant (β = .179, p = .010), confirming that income inequality is a robust determinant of homicide rates even after the introduction of additional control variables (Table 8). Inflation also shows a positive and significant effect (β = .046, p = .011), suggesting that macroeconomic instability compounds the criminal incentives already produced by structural inequality. Notably, Enrollment in Secondary School (β = -.121, $p < .001$) and Corruption Perception (β = -.168, $p < .001$) both exhibit negative and highly significant coefficients, indicating that higher educational attainment and lower perceived corruption are associated with reduced homicide rates, which remained consistent with the paper's argument that structural investments in education and institutional quality are effective tools for crime reduction. The interaction term (β = .027, $p < .001$) is positive and highly significant, suggesting that the combined presence of income inequality and corruption amplifies the effect on homicide rates, meaning inequality affects crime, especially in contexts where institutional trust is low. Unemployment (β = -.007, p = .950) and Government Effectiveness (β = .027, p = .756) were not statistically significant, suggesting their independent effects on homicide are insignificant once other structural factors are accounted for. All together, these results strengthen the paper's central claim that income inequality is a persistent and robust driver of crime in Latin America.

Table 8. Coefficients Robustness Check

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	10.212	3.410		2.995	.003
	Gini	.179	.069	.044	2.614	.010
	Unemployment	-.007	.112	-.001	-.063	.950
	IncomePerCapita	.000	.000	-.042	-1.733	.086
	Inflation	.046	.018	.037	2.567	.011
	EnrSecSchool	-.121	.020	-.125	-6.136	<.001
	Corruption Perception	-.168	.045	-.126	-3.735	<.001
	ControlVar	.027	.000	.899	59.086	<.001
	Government Effectivness	.027	.086	.012	.311	.756

a. Dependent Variable: Homocides

Baseline with Interaction Terms

The second robustness check expands the baseline model by introducing another interaction term that measures the strength of institutions by combining the Gini index and Rule of Law. The base model is specified as follows:

$$IntentHomicide_{it} = \beta_1 + \beta_2 Gini_{it} + \beta_3 Unemployment_{it} + \beta IncomePerCapita_{it} + \beta_5 Inflation_{it} + \beta_6 CorrupPercept_{it} + \beta_7 GovEffectivness_{it} + \beta_8 ProgSecSchool_{it} + \beta_9 RuleOfLaw_{it} + \beta_{10} InstStrength_{it} + \beta_{11} ControlVar_{it} + \varepsilon_{it}$$

This second robustness check further strengthens the model by introducing the effect of institutional strength in crime. The model achieved an R-squared of .990, indicating that the model explains approximately 99% of the variance in homicide rates across the sample (table 9). Nonetheless, in this specification the Gini index returns a negative and statistically significant coefficient (β = -.834, p = .028). Rather than contradicting the paper's central argument, this result likely reflects the moderating role of the newly introduced institutional variables, particularly Rule of Law (β = 60.498, p = .006) and Institutional Strength (β = -1.393, p = .005), which captures how the quality and legitimacy of institutions shapes the inequality-crime relationship. When institutional variables are held constant, the direct linear effect of inequality on homicide shifts, suggesting that inequality's impact on crime is heavily conditioned by the institutional environment in which it operates. This is consistent with the social resistance framework, which argues that inequality drives crime most powerfully when institutional trust is low. Enrollment in Secondary School remains negative and highly significant (β = -.290, p = .002), reinforcing the protective role of education against criminal behavior. The interaction term remains positive and highly significant (β = .028, p < .001), confirming that the combined effect of inequality and corruption perception consistently amplifies homicide rates across specifications. Unemployment, Income Per Capita, Inflation, Corruption Perception, and Government Effectiveness were not statistically significant in this specification, suggesting their effects are absorbed by the institutional variables introduced in this model (table 10).

Table 10. Coefficients of Robustness check model 2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	56.226	16.331		3.443	.002
	Gini	-.834	.363	-.114	-2.297	.028
	Unemployment	.509	.479	.042	1.064	.295
	IncomePerCapita	.000	.000	.031	.734	.468
	Inflation	.583	.400	.035	1.459	.154
	Corruption Perception	-.244	.142	-.102	-1.719	.095
	Government Effectivness	.237	.218	.072	1.085	.285
	ProgSecSchool	-.290	.084	-.095	-3.451	.002
	RuleLaw	60.498	20.493	1.459	2.952	.006
	InstStrength	-1.393	.459	-1.572	-3.036	.005
	ControlVar	.028	.001	.948	37.249	<.001

a. Dependent Variable: Homocides

Business Cycle Adjustment

The third robustness check accounts for the business cycle fluctuations, by using a 3-year average. Data restriction did not permit using either a 4 or 5 year average. The baseline of the model was the following,

$$\text{Avg}IntentHomicide_{it} = \beta_1 + \beta_2 AvgGini_{it} + \beta_3 AvgUnemployment_{it} + \beta_4 AvgIncomePerCapita_{it} + \beta_5 AvgProgSecSchool_{it} + \beta_6 AvgRuleOfLaw_{it} + \beta_7 AvgInflation_{it} + \varepsilon_{it}$$

The results of the model were notably different, as it had an R-squared of .422 (table 11), indicating a significant decline in explanatory power compared to previous models. This reduction is expected, as averaging across three-year periods smooths out short-term variation that the model previously captured, and the reduced sample size introduced by the averaging process limits statistical power. In this specification, the Gini index is negative but statistically insignificant (β = -.126, p = .781), suggesting that over medium-term cycles, the direct linear effect of inequality on homicide is not robustly detected (table 12). This is likely due to the data constraints mentioned above rather than an absence of the crime-inequality relationship. Unemployment has a positive and significant output (β = 2.320, p = .001), indicating that over 3-year business cycles, labor market exclusion is a meaningful driver of homicide rates, consistent with the rational choice framework's emphasis on constrained economic opportunity. Income Per Capita shows a negative and significant effect (β = -.001, p = .023), suggesting that higher average incomes are associated with reduced homicide rates over time. Rule of Law is negative and significant (β = -13.934, p = .003), reinforcing the argument that stronger institutional frameworks suppress criminal behavior independently of inequality levels. Inflation is close on having statistical significance, but is off by .05 (β = -.494, p = .052), indicating that the test would fit the variable better by having one tail. And Enrollment in Secondary School is not significant in this specification (β = -.077, p = .757).

Table 12. Coefficients for Business Cycle Model

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	32.761	30.511		1.074	.287
	AVERAGE of Gini	-.126	.452	-.029	-.279	.781
	AVERAGE of Unemployment	2.320	.689	.368	3.368	.001
	AVERAGE of IncomePerCapita	-.001	.001	-.302	-2.331	.023
	AVERAGE of ProgSecSchool	-.077	.247	-.033	-.310	.757
	AVERAGE of RuleLaw	-13.934	4.459	-.421	-3.125	.003
	AVERAGE of Inflation	-.494	.249	-.208	-1.980	.052

a. Dependent Variable: AVERAGE of Homocides

Business Cycle Adjustment with Interaction Term

The fourth and final robustness check, reintroduces the interaction term and control variables into the 3-year average model, having the following baseline,

$$\text{Avg}IntentHomicide_{it} = \beta_1 + \beta_2 AvgGini_{it} + \beta_3 AvgUnemployment_{it} + \beta_4 AvgIncomePerCapita_{it} +$$

$$\beta_5 AvgProgSecSchool_{it} + \beta_6 AvgRuleOfLaw_{it} + \beta_7 AvgInflation_{it} + \beta_8 AvgGovEffectivness_{it} + \beta_8 AvgControlVar_{it} + \beta_8 AvgCorrPercept_{it} + \varepsilon_{it}$$

The results yield an R-squared of .984, (table 13). This showed a significant improvement in explanatory power over the third model, suggesting that institutional and interaction variables recover some of the variance lost in the business cycle. The average Gini index, is negative (β = -.055) does not again reach statistical significance (p=.852), likely reflecting the reduced statistical power introduced by the 3-year average process. Enrollment in secondary school is negative and highly significant (β=-.366, p =0.006), reinforcing the consistent finding that educational attainment is a structural component of homicide rates. The interaction term is positive and highly significant (β =.028, p <.001), confirming that the effect of the combined presence of inequality and corruption perception on homicide rates is robust even when accounting for business cycles. The rest of the variables are not significant in this model, possibly due to multicollinearity among institutional variables after averaging.

Table 14. Coefficients for Business Cycle model with interaction term

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	37.112	20.465		1.813	.086
	AVERAGE of Gini	-.055	.289	-.009	-.189	.852
	AVERAGE of Unemployment	-.331	.500	-.031	-.662	.516
	AVERAGE of IncomePerCapita	.000	.000	-.057	-.740	.468
	AVERAGE of ProgSecSchool	-.366	.119	-.126	-3.071	.006
	AVERAGE of RuleLaw	.234	4.731	.007	.049	.961
	AVERAGE of Inflation	.832	.626	.055	1.329	.199
	AVERAGE of Government Effectivness	.341	.272	.136	1.254	.225
	AVERAGE of ControlVar	.028	.001	.904	22.814	<.001
	AVERAGE of Corruption Perception	-.365	.229	-.191	-1.592	.128

a. Dependent Variable: AVERAGE of Homocides

DISCUSSION AND RESULTS

The results of this study are presented across multiple specifications, beginning with the baseline OLS regression, followed by a curve estimation to assess functional form, and conclusion with four robustness checks designed to test the consistency of the findings across alternative model configurations.

The baseline regression model (table 3), yields and R-squared of .415, indicating that the model explains approximately 41.5% of the variance in homicide rates, across the sample. The Gini index is negative and statistically insignificant (table 4), which as discussed in the methodology section, is likely attributable to data limitations in Gini measurements across Latin America, and the potential presence of simultaneity – where crime itself reproduces inequality, distorting the direct coefficient. This interpretation is supported by the curve estimation, which demonstrates that an exponential model better fits the inequality-crime relationship, suggesting that the effects of inequality on crime compound at higher Gini levels, instead of growing at a fixed rate. This is consistent with relative deprivation theory's argument, where social comparisons become increasingly extreme as inequality widens.

Despite the insignificant Gini coefficient, several control variables perform as theoretically expected. The positive and significant effect of unemployment (β = 1.424, p = .020) aligns with the rational choice framework, where labor market exclusion narrows legitimate economic opportunities and makes criminal activity comparatively more attractive. The negative and significant effect of Income Per Capita (β = -.002, p < .001) suggests that aggregate economic growth is associated with reduced homicide rates, though as Kim et al. (2020) caution, the distribution of that growth is ultimately a more critical determinant of violent crime. The significant negative effect of Rule of Law (β = -9.548, p = .016) is consistent with the social resistance framework, where institutional legitimacy suppresses the alienation and oppositional identities that drive criminal behavior in highly unequal societies.

Altogether, the four robustness checks demonstrate that the main findings of this paper are consistent across alternative models. The interaction term between inequality and corruption perception is the most consistent and robust predictor of homicide rates, remaining positive and highly significant across all annual specifications and the business cycle adjusted model. This confirms that income inequality leads to crime when compounded by institutional illegitimacy, which is consistent with Itskovich's (2023) social resistance framework. When individuals are simultaneously excluded from economic opportunity and perceive their institutions are corrupt, their rationale for following social norms becomes irrational, making crime a rational and collective response. This dynamic is particularly pronounced in Latin America, where institutional distrust and economic disparity have historically coexisted, suggesting that anti-crime interventions cannot address inequality or institutional reform alone, interventions across different sectors must be implemented to reduce crime.

Moreover, Enrollment in Secondary School is found as the most consistent suppressant of homicide rates across all four checks, reinforcing Sugiharti et al.'s (2023) argument that structural investment in education reduces the conditions that make criminal behavior rational for economically excluded individuals. This finding is particularly significant as it holds across both annual and business cycle adjusted models, suggesting that the effect of education is not a short-term phenomenon but a structural suppressor of criminal behavior. Education reduces crime not only by expanding legitimate economic opportunities, but also by strengthening individuals' attachment to institutional norms and social values, counteracting the alienation and social resistance that Itskovich (2023) identifies as the primary pathway in which inequality transforms into criminal behavior.

While the Gini index does not reach significance in every specification (particularly in the business cycle adjusted models) its positive and significant effect in the first robustness check (β = .179, p = .010), combined with the exponential curve estimation results, provides sufficient empirical grounding to support that income inequality is a persistent structural determinant of crime in Latin America, whose effect is conditioned by institutional quality and mediated by access to education.

CONCLUSION

This study examines the structural relationship between income inequality and crime in Latin America, arguing that income inequality is not merely correlated with criminal behavior but is its primary structural driver. Drawing on relative deprivation theory, social resistance frameworks, and rational choice theory, the analysis demonstrates that the mechanisms through which inequality translates into crime are multidimensional.

While the baseline OLS regression does not produce a statistically significant Gini coefficient, the curve estimation reveals that the inequality-crime relationship is non-linear and self-reinforcing, compounding at higher levels of disparity in a manner consistent with relative deprivation theory. Across the robustness checks, the interaction term between inequality and corruption perception appears as the most consistent and robust predictor of homicide rates, confirming that inequality leads to crime when compounded by institutional illegitimacy. Enrollment in Secondary School is the most stable suppressant of homicide rates across all models, underscoring that structural investment in education reduces the conditions that make criminal behavior rational for economically excluded individuals.

These findings carry important policy implications. Reducing crime in Latin America requires interventions that go beyond conventional law enforcement approaches. Policymakers must also target income inequality, institutional corruption, and educational access to reduce crime. Anti-poverty measures alone are insufficient if the underlying structure of income distribution remains significantly unequal and institutional trust remains low. Future research should expand this analysis by incorporating alternative crime indicators, longer time horizons, and qualitative assessments of how social media amplifies relative deprivation and criminal behavior diffusion across the region. Ultimately, addressing Latin America's crime crisis demands a structural reckoning with the economic and institutional inequalities that have defined the region since the colonial period.

REFERENCES

Acemoglu, D., & Robinson, J. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Publishers.

ACLED. (2025a, February). *Latin America and the Caribbean overview: February 2025*. https://acleddata.com/update/latin-america-and-caribbean-overview-february-2025

ACLED. (2025b, December). *Latin America and the Caribbean overview: December 2025*. https://acleddata.com/update/latin-america-and-caribbean-overview-december-2025

ACLED. (2025c). *Conflict index & 2026 watchlist*. https://acleddata.com/conflict-index-2026-watchlist

ACLED. (2026, March). *Latin America and the Caribbean overview: March 2026*. https://acleddata.com/update/latin-america-and-caribbean-overview-march-2026

Americas Quarterly. (2024). *What the U.S. election means for Latin America*. https://americasquarterly.org/article/new-aq-what-the-u-s-election-means-for-latin-america

Curiel, R. P., Cresci, S., Muntean, C. I., & Bishop, S. R. (2020). Crime and its fear in social media. *Palgrave Communications, 6*, Article 57. https://doi.org/10.1057/s41599-020-0430-7

ECLAC. (2025). *Income concentration still extreme in Latin America: Wealthiest 10% obtains 34.2% of total income*. https://www.cepal.org/en/pressreleases/income-concentration-still-extreme-latin-america-wealthiest-10-obtains-342-total

ECLAC/ILO. (2018). *Policies to address the challenges of existing and new forms of informality in Latin America*. https://repositorio.cepal.org/server/api/core/bitstreams/431aed2d-2e7a-4f05-a942-b2d12db14bb5/content

InSight Crime. (2025). *InSight Crime 2025 homicide round-up*. https://insightcrime.org/news/insight-crime-2025-homicide-round-up/

Inter-American Development Bank. (2021). *Combating crime in Latin America and the Caribbean: What public policies do citizens want?* https://doi.org/10.18235/0003306

Itskovich, E., & Factor, R. (2023). Economic inequality and crime: The role of social resistance. *Journal of Criminal Justice, 86*, Article 102065. https://doi.org/10.1016/j.jcrimjus.2023.102065

Kelly, M. (2000). Inequality and crime. *The Review of Economics and Statistics, 82*(4), 530–539. https://www.jstor.org/stable/2646649

Kim, B., Seo, C., & Hong, Y. (2022). A systematic review and meta-analysis of income inequality and crime in Europe: Do places matter? *European Journal on Criminal Policy and Research, 28*, 573–596. https://doi.org/10.1007/s10610-020-09450-7

Legal Information Institute. (n.d.). *Crime*. Cornell Law School. https://www.law.cornell.edu/wex/crime

Montgomery, R. (2024). *Colonial legacy and economic persistence in the Americas*. Preprints.org. https://www.preprints.org/manuscript/202410.0396

Organization of American States. (2026). *Drugs*. https://www.oas.org/en/topics/drugs.asp

Perez-Vincent, S. M., & Schargrodsky, E. (2024). *Crime underreporting in Latin America and the Caribbean*. Inter-American Development Bank. https://publications.iadb.org/en/publications/english/viewer/Crime-Underreporting-in-Latin-America-and-the-Caribbean.pdf

Pew Research Center. (2022, December 6). *Views of social media and its impacts on society in advanced economies 2022*. https://www.pewresearch.org/global/2022/12/06/views-of-social-media-and-its-impacts-on-society-in-advanced-economies-2022/

Schargrodsky, E., & Freira, L. (2023). Inequality and crime in Latin America and the Caribbean: New data for an old question. *Economía LACEA Journal, 22*(1), 175–202. https://doi.org/10.31389/eco.413

Sugiharti, L., Purwono, R., Esquivias, M. A., & Rohmawati, H. (2023). The nexus between crime rates, poverty, and income inequality: A case study of Indonesia. *Economies, 11*(2), Article 62. https://doi.org/10.3390/economies11020062

Walker, I., & Smith, H. J. (2002). *Relative deprivation: Specification, development, and integration*. Cambridge University Press. https://pubmed.ncbi.nlm.nih.gov/22194251/

World Bank. (2010). *Crime, violence, and development: Trends, costs, and policy options in the Caribbean*. http://documents.worldbank.org/curated/en/537301468231536879

World Bank. (2023). *Origins of Latin American inequality*. https://thedocs.worldbank.org/en/doc/6af04888efedf21dcd63d721444f9a85-0370012023/original/Paper-Origins-of-Latin-America-Inequality.pdf

World Bank. (2025a). *Gini index*. https://data.worldbank.org/indicator/SI.POV.GINI

World Bank. (2025b). *Gini index — metadata*. https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SI.POV.GINI

World Bank. (2025c). *Intentional homicides, male (per 100,000 male) — metadata*. https://databank.worldbank.org/metadataglossary/world-development-indicators/series/VC.IHR.PSRC.MA.P5

World Bank. (2025d). *Latin America and the Caribbean overview*. https://www.worldbank.org/en/region/lac/overview

World Bank. (2025e). *Violence without borders*. https://www.worldbank.org/en/research/publication/violence-without-borders

APPENDIX

Table 1. Gini Coefficient in Latin America (2026)

Source: World Bank Poverty and Inequality Platform (via World Population Review, 2026)

High (>50) Elevated (43–50) Moderate (<43)

Country	Gini coefficient	Inequality level
Colombia (2023)	53.9	High
Brazil (2023)	51.6	High
Panama (2024)	49.7	Elevated
Costa Rica (2024)	45.8	Elevated
Guatemala (2023)	45.2	Elevated
Ecuador (2024)	45.2	Elevated
Paraguay (2024)	44.2	Elevated
Mexico (2022)	43.5	Elevated
Chile (2022)	43.0	Elevated
Argentina (2024)	42.4	Moderate
Bolivia (2023)	42.1	Moderate
Peru (2024)	40.1	Moderate
Uruguay (2024)	40.0	Moderate
El Salvador (2023)	39.8	Moderate
Dominican Republic (2024)	39.0	Moderate

Note: Gini scores reported as index (0–100). Year of most recent World Bank data shown in parentheses.

Table 2. Summary of Variables used

Variable	Description	Source	Anticipated Slope	Anticipated Statistical Significance	Anticipated Practical Significance
Y Intentional Homicides (pero 100,000 people)	Proxy for crime, measuring the number of unlawful, intentional killings per 100,000 people in a given area. From 1990 to 2025, by country.	World Bank Group	-	-	-
β_2 Gini Index	The gap between the actual income distribution and a perfectly equal one. From 1990 to 2025, by country.	World Bank Group	Positive	Yes	High
β_3 Unemployment (% of total labor force)	Percentage of the labor force that is unemployed(a person must have actively looked for work in the past four weeks). From 1990 to 2025, by country.	World Bank Group	Positive	Yes	High
β_4 Income Per Capita (PPP)	Primary indicator of standard of living, economic well-being and development. Measured by dividing the total income (or GDP) by the total population. From 1990 to 2025, by country.	World Bank Group	Negative	Yes	High
β_5 Progression to Secondary School	Percentage of students who move from the final grade of primary school to the first grade of secondary school in	World Bank Group	Negative	Yes	High

	the following year. From 1990 to 2025, by country.				
β_6 Rule of Law	The extent to which agents have confidence in and abide by society's rules, specifically contract enforcement, property rights, police quality, and court impartiality. From 1990 to 2025, by country.	World Bank Group	Negative	Yes	High
β_7 Inflation, Consumer princes (%annual)	Year-over-year percentage change in the cost of a "basket" of goods and services. From 1990 to 2025, by country.	World Bank Group	Positive	Yes	High

Table 3. Model Summary of First Model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.645[a]	.415	.385	17.147603267

a. Predictors: (Constant), Inflation, RuleLaw, Gini, ProgSecSchool, Unemployment, IncomePerCapita

Table 5. Parameter Estimate

Model Summary and Parameter Estimates

Dependent Variable: Homocides

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.067	25.957	1	359	<.001	-28.762	1.037		
Cubic	.070	13.546	2	358	<.001	28.794	-.753	.000	.000
Growth	.167	71.788	1	359	<.001	-.446	.065		
Exponential	.167	71.788	1	359	<.001	.640	.065		

The independent variable is Gini.

Table 6. Robustness Check Summary

Variable	Description	Source	Anticipated Slope	Anticipated Statistical Significance	Anticipated Practical Significance
β_8 Enrollment to Secondary School	The registration of students into high school or junior high. From 1990 to 2025, by country.	World Bank Group	Negative	Yes	High
β_9 Corruption Perception	How corrupt a nation's public sector is perceived to be by experts and business executives. From 1990 to 2025, by country.	Transparency International	Positive	Yes	High
β_{10} Government effectiveness	Measures the quality of public services, independence from political pressure, policy implementation, and the credibility of government commitments. From 1990 to 2025, by country.	World Bank Group	Negative	Yes	High

Table 7. Robustness check model 1 summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.990[a]	.979	.978	2.6970568094

a. Predictors: (Constant), Government Effectivness, Gini, Inflation, ControlVar, Unemployment, EnrSecSchool, IncomePerCapita, Corruption Perception

Table 9. Robustness check model 2 summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.995[a]	.990	.987	2.8588783991

a. Predictors: (Constant), ControlVar, Unemployment, Inflation, Corruption Perception, ProgSecSchool, Gini, IncomePerCapita, Government Effectivness, RuleLaw, InstStrength

Table 11. Robustness check (Business Cycle)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.650[a]	.422	.373	15.956220229

a. Predictors: (Constant), AVERAGE of Inflation, AVERAGE of RuleLaw, AVERAGE of Gini, AVERAGE of ProgSecSchool, AVERAGE of Unemployment, AVERAGE of IncomePerCapita

Table 11. Robustness check (Business Cycle with interaction terms)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.992[a]	.984	.977	3.4284736272

a. Predictors: (Constant), AVERAGE of Corruption Perception, AVERAGE of Gini, AVERAGE of ControlVar, AVERAGE of ProgSecSchool, AVERAGE of Inflation, AVERAGE of Unemployment, AVERAGE of IncomePerCapita, AVERAGE of Government Effectivness, AVERAGE of RuleLaw

BLAKES BRAKES

Executive Summary

Contact Information

Logan Allen

Siena University

ls16alle@siena.edu

Blake Demianenko

Siena University

bk24demi@siena.edu

Ian Downer

Siena University

im15down@siena.edu

Declan Coletti

Siena University

d31stir@siena.edu

Mark Grosky

Siena University

mb03gros@siena.edu

Industry: Electric Bikes

Development Stage: Startup

Year Founded:2026
Employees:5

Funding Opportunity:50,000

Use of Fund:82% Variable Cost 10% marketing 5% equipment(fixed) 3% other

Monthly Burn Rate: $52,083.33

Average Estimated Year 1 Monthly: $ 66,913.33

THE GRAB

Blake's Brakes provides a lightweight integrated energy recovery system for your bicycle that turns wasted braking energy into stored electrical power. By turning every stop into a performance advantage, this system enhances efficiency and provides on demand power, without compromising the natural riding experience.

PROBLEM/OPPORTUNITY

Everyday Cyclists lack an affordable way to recover lost energy and reuse it to power your bicycle so that it will be easier to go from stopping to your average biking speed. There is an open space in the market for our product due to E-bikes being a luxury item and only the top-notch bikes have this feature. Another problem we noticed is that food delivery service people in the city require quick and efficient riding to ensure maximum profit, and due to that our product comes in since in that position they don't have the financials to acquire an E-bike but need to have a bicycle that is working at full capacity.

SOLUTION/PRODUCT

This product is a compact energy recovery and assist module designed to integrate seamlessly with almost any traditional mechanical bike. During deceleration the system captures kinetic energy that would be otherwise lost and converts it into stored electrical power through regenerative braking. The stored energy is then redeployed from a high efficiency supercapacitor to the motor to provide a power boost during hill climbs or acceleration. A handlebar mounted control allows riders to activate the power boost on demand, delivering performance when it is needed.

POTENTIAL RETURN/REVENUE MODEL

We estimate that it will cost us 250 to produce one product and will be sold for 285 and the estimated revenue for the first 6 months is $239,350. The total cost of goods sold for the first 6 months is $227,500. The total return for 5 years is $216,432, also our company breaks through at the end of year 3.

COMPETITION

Our current main competitors within this demographic include Trek, Rad Power Bikes, Juiced Bikes, and Lectric eBikes. These are the current brands that control the market, by offering complete e-bike systems. These companies mainly focus on their products' design off power, and convenience for everyday riders/commuters. However, there is a downfall with this; fully integrated e-bike systems come at extremely high costs and are usually way heavier than a normal pedal bike. Also, there are nationwide bans on these systems across major cities and transportation limitations due to safety regulations. This creates a clear gap within the market, there aren't many brands addressing riders who are looking for performance enhancements, while not wanting to purchase a full system e-bike. Our product designed by Blakes Brakes targets riders who are looking for the benefits of relative braking and efficiency, without having to replace their current bicycle. Instead of competing head-to-head with the full-fledged E-bike manufactures, we are positioning our brand as a modular add-on solution, which then allows consumers to have a more cost-effective and effective way to upgrade their riding experience, wherever desired.

EXECUTION PLAN/GO TO MARKET STRATEGY

We are going to target Urban delivery drivers and commuters, and our secondary market is Eco- friendly consumers. To attract attention, we are going to advertise mainly in social media online platforms such as TikTok, Instagram, and influencers that work in the same market as our product. We are going to price this low and reliable so that people in the working class will be able to easily afford this product. All of our marketing and advertising will be focused on the people in the city.

FINANCIALS

We came up with the revenue from looking at the survey and looking at our variable costs. Our survey results indicate that our target customer wants the price to be between 100 and 250. However, our variable costs came in at 250 per product so we had to bring up the costs to about 285. This is a lot lower for the other products in the market that are similar to our competitors that also sell a hybrid bike add on.

We can have low costs for putting the parts together and we have given it a generous estimate that it will be lower.

THE TEAM/RELEVANT EXPERIENCE

Our team is made up of five members who are all Business Majors: Blake Demianenko, Logan Allen, Ian Downer, Marky Grosky, and Declan Coletti. Blake is the founder of this and created this idea last year thinking of different ways to become more eco-friendly. Logan and Declan have taken on most of the financials using prior experience from their past Spark Tank Presentations. Ian has been taking on Survey and Marketing details making sure Blakes Brakes align with the consumer values. Mark has been tasked with Market Growth and Competitive Landscape to make sure we have a great niche in a very competitive market.

CONSUMER PERCEPTIONS OF ARTIFICIAL INTELLIGENCE IN ADVERTISING

Angus Grout, Siena University
Carson Hines, Siena University
Cayden Catalina, Siena University
Joshua Miller, Siena University
Dr. Cheryl Buff, Siena University – Faculty Mentor

ABSTRACT

As technology continues to advance, artificial intelligence is continuing to become more prevalent, with brands relying heavily on AI to generate imagery and scripts, despite an actor's strike in response to the AI takeover. Overall objectives of the research assess the attitudes and trust of consumers when viewing advertisements generated with AI rather than humans.

Our team chose non-probability sampling methods such as convenience and snowball sampling, which are most efficient when studying a specific group. Choosing participants who are suitable to the study and are referred to our survey through shared links builds trust, familiarity, and grows participation naturally. Our IRB approved survey, designed through Qualtrics, consists mainly of interval questions shared by team members on platforms such as Instagram, LinkedIn, Snapchat, and Facebook. From the 135 total responses, with 84% of participants between the ages of 18-24, and 65% of participants being women, our target towards young adult consumers was successful. Some key findings include that AI usage transparency leads to a more positive perception of the company, consumers are attracted more to companies that use high-quality AI within their ads, and when an ad is credible and believable, the viewer is accepting of the use of AI in it. Additionally, if the AI within the ad is easily noticeable and identifiable, then the viewer will have a negative perception of it. Although there is no significance in age and perception, we did find that our older group has a higher perception of AI on average. These findings can bring insight into how brands choose to advertise their brands with the use of AI, while maintaining consumer loyalty and ethical usage.

Methodological constraints for our research include the sample size failing to reach a wider population, artificial intelligence remaining an emerging threat, and time constraints preventing a more in-depth analysis from being conducted. Recommendations for future research would include removing "young adult consumer perceptions" from our recruitment language, recruiting more participants in-person, and creating a section in the survey ensuring that respondents are near completion, overall increasing the survey completion rate. In conclusion, our analysis directly answered our research question of "What are young adult consumers' perceptions of advertisements that utilize AI?" and showed strong statistical significance across most of our main hypotheses.

LITERATURE REVIEW

Reactions and Responses to the Growing use of AI in Advertising

The rapid progression of artificial intelligence has raised the immediate need for research regarding its impact on society. In the marketing world, one pressing topic is how consumers perceive and respond to the presence of AI in advertising. More and more examples of AI generated advertisements are being created with mixed reactions from consumers. The need for more information regarding this development has been answered by a few research studies. For example, an experiment conducted with Gen Z in Vietnam revealed that AI voices can be just as effective as human voices in driving advertisement recall (Lu 2025). This raises important questions by consumers about transparency and the ethics of replacing humans. Another research study focused on consumer perceptions of AI in advertising, measured emotional reactions to the usage of "digital resurrections" of deceased celebrities (Aboulnasr 2025). These emotions ranged from uncomfortable to positive, depending on the viewer's attachment to the brand and celebrity (Aboulnasr 2025). This usage of the technology once again raises ethical concerns amongst consumers, and the varying responses emphasize the need for more research into consumer perceptions of AI.

The development of AI-powered creative systems also raises questions about the authenticity of ad messaging. One article proposes the creation of a Creative Advertising System (CAS) that generates ideas based on past ads, classifying styles, generating new concepts, and evaluating novelty and value (Vakratsas 2021). This is a prime example of how advertisers are using the technology to make advertisement creation more efficient. However, in a similar capacity to the AI audio advertisements, the technology could be eliminating jobs.

Another study we looked at aims to understand whether biases exist between human and AI advertisements (Zhang 2023). While looking into reactions and responses to AI advertisements, we thought it would be essential to include studies that compare human advertisements to AI ads. In this study, participants were either informed, slightly informed, or not informed about the source of the ad (Human or AI), and their responses were measured based on satisfaction, willingness to pay, interest, and persuasion (Zhang 2023). This research reveals that while people do not dislike AI-generated content, they exhibit bias in favor of content they know to be created by humans, especially when it's explicitly told to them.

Building Frameworks & Constructs

One of the studies we looked at develops a framework that identifies four core components of understanding responses to manipulated advertising: manipulation sophistication, perceived verisimilitude, perceived creativity, and awareness of ad falsity, all of which interact to influence persuasion outcomes (Campbell 2022). The framework developed here is highly useful to our understanding of consumer perceptions when related to AI marketing. It outlines how certain marketing strategies affect consumers' perceptions of the ad and the likelihood that the ad will be effective. Some of these factors align with the same factors outlined in the next section, which we added to our research design and are stated in the next section.

Another developed framework we found to support our research was in an article describing what goes into consumer trust in AI personalized advertising (Feng 2025). Authors Feng, Y. & Kim, J in *Decoding the Trust Matrix: Unraveling Key Predictors of Consumer Trust in AI-Generated Personalized Advertising* (2025), conceptualize trust as a construct made up of five factors: Ad transparency, Ad verisimilitude, personalization need, AI familiarity, and AI policy uncertainty. The researchers stated that Ad transparency, Ad verisimilitude, personalization need, and AI familiarity would have a positive effect on AI-generated personalized advertising, which they found and proved to be true (Feng 2025). The hypothesis that was not supported was that AI policy uncertainty negatively affects AIGPA. They had assumed that with an uncertain policy, distrust would follow. However, a positive relationship was found instead. We decided to take a few of these factors: Ad transparency, AI familiarity, and Ad verisimilitude (which we reworded), and implemented them into our own research and hypothesis.

Consumer Trust, Authenticity, and Risk Perception in AI-Driven Advertising

Across all three articles, Consumer Responses to Ai-generated Charitable Giving Ads by Arango, L., Singaraju, S. P., & Niininen, O. (2023), Programmatic advertising in online retailing: Consumer Perceptions and Future Avenues by Ciuchita, R., Gummerus, J. K., Holmlund, M., & Linhart, E. L. (2022), and Investigating the impact of artificial intelligence on consumer's purchase intention in e-retailing by Bhagat, R., Chauhan, V., & Bhagat, P. (2022), a consistent theme emerges, which is that consumers' trust in AI-driven advertising is highly dependent on how authentic and ethical the content feels. While AI can enhance relevance and efficiency in digital marketing, it also carries reputation risks when it feels unauthentic and when it is visually jarring. Arango et al. (2023) show this clearly in the charitable realm, where consumer trust and willingness to donate drop when people realize imagery is AI-generated. Their findings highlight a core issue in AI advertising which is when the technology becomes too visible or feels emotionally insincere, it can reduce credibility and damage the organization's perceived authenticity (Arango et al. 2023). Similarly, Ciuchita et al. (2022) find that although AI can improve ad relevance and strengthen attitudes toward online retailers, this only works when consumers do not feel that their data is being misused or that the technology is manipulating them. Data-privacy concerns and perceived risk hurt the positive effects of AI, illustrating how fragile consumer trust can be when AI becomes too intrusive. Together, these articles reinforce that authenticity and transparency play a huge role in shaping how consumers interpret AI-generated or AI-delivered content.

At the same time, Bhagat et al. (2022) add additional narrative by showing that AI can also increase trust, satisfaction, and purchase intention when consumers perceive it as useful, efficient, and aligned with their expectations. Their results demonstrate that not all AI leads to having skepticism attached to it. When AI reduces effort and improves the experience, consumers respond positively (Bhagat et al. 2022). Combined with the findings of the other two studies, this suggests that the impact of AI on consumer perception is highly dependent on the context in which it is being used in. AI helps when it enhances relevance and convenience but hurts when it compromises emotional realism, privacy, or transparency. For our project, these articles collectively point to the idea that AI in advertising is not inherently good or bad. Instead, consumer perceptions depend on whether the AI maintains authenticity, respects boundaries, and supports the consumer experience without trying to replace or fabricate emotional truth. It can be a powerful tool when used in a proper fashion.

Consumer Satisfaction and Purchase Intention

Studies surrounding consumers' opinions on AI when encountering a chatbot found an increase in both satisfaction and purchase intention. The article *Consumers' Perception on Artificial Intelligence Applications in Marketing Communication* by Chen , H., Chan-Olmsted, S., Kim, J., & Mayor Sanabria, I. (2021) examines how people perceive AI, especially voice-enabled technology. Out of 20 qualitative interviews, the study found that consumers evaluate AI both practically and emotionally, relating it to human communication very often. They enjoy its efficacy, adaptability, and communication abilities but are concerned with privacy and trust (OpenAI, 2023). Although most view AI marketing as inevitable and in large part acceptable, they believe it does not affect brand attitudes or purchasing behavior. Overall, the article is helpful in that it identifies the emotional and relational elements of AI that are ignored in most quantitative research. Its qualitative nature allows for richness, though, although the limited sample compromises generalizability. The study advises marketers to have less persuasion and more trust, privacy protection, and improved customer experience. Despite its flaws, it is an influential addition to the body of knowledge regarding how consumers perceive and respond to AI marketing communication (OpenAI, 2023).

Additional research conducted in this study and explained throughout the article relates to the research topics because the idea of the AI "chatbot" is explored and tested (Chen et al. 2021). This advancement studied the engagements made by consumers and the levels of satisfaction that came from the level of chatbot interaction. The study found that the AI chatbot's qualities and abilities to communicate heavily shaped the results and purchases made by consumers (Chen et al. 2021). Whether the number of things purchased increased, or the AI technique fully persuaded customers to purchase in the final decision-making process, the involvement of AI in marketing campaigns had a positive effect on the engagement and overall perception of products from specific brands for consumers.

RESEARCH DESIGN AND SAMPLING

Research Design

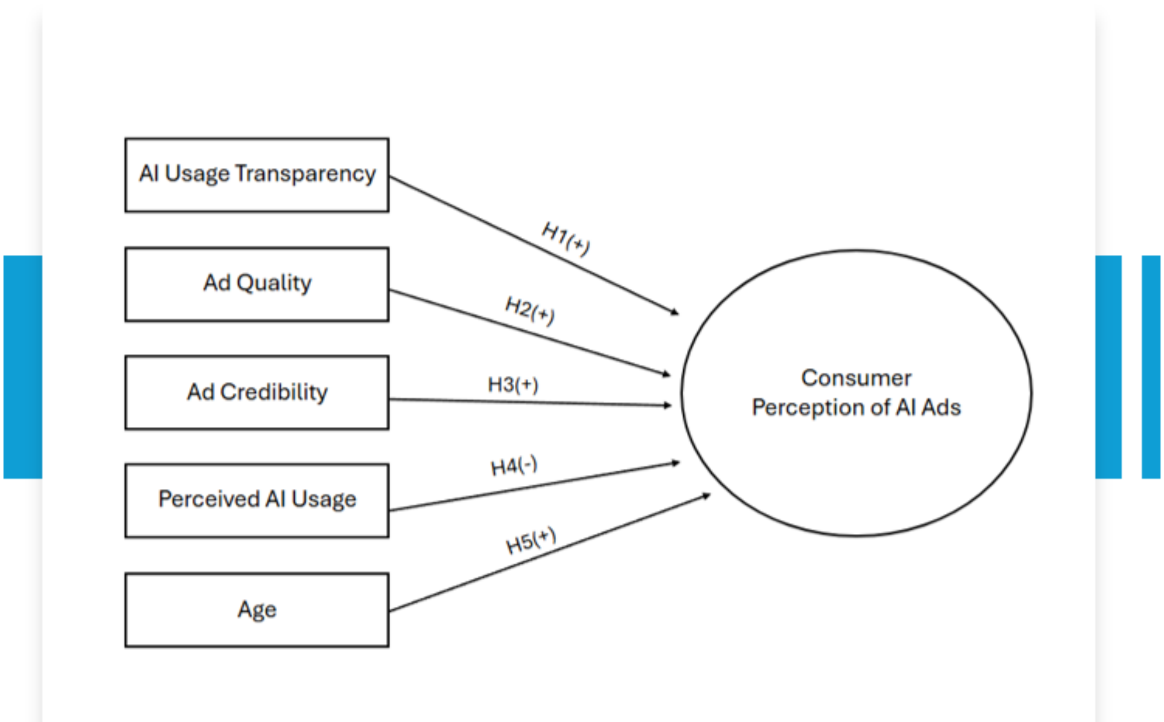
To learn more about consumer perceptions of AI in advertisements, a self-administered online survey was created through Qualtrics as a form of quantitative descriptive research. The questionnaire was intended to measure the relationship between several variables through statistical analysis. To perform precise numerical comparisons, sixteen of the survey's twenty total questions were created with interval measurement. The remaining four questions were split evenly between nominal and ordinal questions. This questionnaire served as our primary source in researching consumer perceptions of AI usage in advertisements, and ran from October 14 to November 6, 2025.

The peer-reviewed journal articles covered in the literature review were used to build our background knowledge in AI usage for marketing and advertisement purposes. These articles served as secondary sources in our research and helped form several hypotheses—found in the conceptual model below—and related questions in the survey.

The respondents of our online survey were collected through non-probability sampling methods—convenience and snowball sampling to be particular. The questionnaire recruitment language was posted by each member of our team on social media platforms such as Instagram, Facebook, and LinkedIn. The result of this sampling method was that the majority of respondents were within the age range of 18 - 24 because that is the most common audience interacting with our social media accounts.

To measure the credibility of advertisements that utilize AI, we employed the "Attitude Toward the Ad (Credibility) scale. This scale contained five 7-point Likert scale items. These questions measured the believability, quality, trustworthiness, bias, and overall credibility of these advertisements.

Conceptual Model



<u>Hypotheses</u>

<u>First Hypothesis</u>: *The more transparent companies are about the usage of AI in their ad, the more positive the consumer perception is of the usage of AI in the ad*. We believe that even if a consumer doesn't agree with the use of AI in the creation of advertisements, they will still have a more positive perception of the ad if they are explicitly notified about the usage of AI.

Second Hypothesis: *The higher the quality of an AI ad is, the more positive the consumer perception is of the usage of AI in the ad.* If the viewer of an advertisement that uses AI considers the ad to be of high quality, then they will have a more favorable perception of the ad's usage of AI.

Third Hypothesis: *The more credible the AI ad is perceived to be by the consumer, the more positive the consumer perception is of the usage of AI in the ad*. Whether it be the ad message itself or the organization behind the advertisement, if the consumer finds the content to be credible, they will favor the usage of AI more.

Fourth Hypothesis: *The more the consumer believes the AI ad is generated by AI. The more negative the consumer perception is of the usage of AI in the ad.* If the consumer believes that an advertisement uses AI without actually knowing for sure, they will have a more negative perception of the ad. In this hypothesis, the independent variable has an inverse relationship with the dependent variable.

Fifth Hypothesis: T*he older the consumer is, the more positive the consumer perception is of the usage of AI in the ad.* This hypothesis may be surprising to some who follow the common trend that younger individuals will be more accepting of new technology. However, through research and an understanding of our peers, older individuals seem to understand less about AI, resulting in more neutral or positive views. On the other hand, young adults have a better grasp of the technology and many have formed negative outlooks on what it could mean for their own futures.

Scales and Validation

The scale we decided to use for our research was Attitude Toward the Ad (Credibility). This scale was composed of 5 interval questions with seven-point responses. The scale measures the degree to which a person believes that a particular advertisement is trustworthy or unbiased. We decided to use this scale because it was very clearly aligned with our research question centered around attitudes and perceptions in advertising. The alphas reported for the scale were .78 (Study 1), .71 (Study 4), and .81 (Supplemental Study), and the validity was not discussed by Jung and Critcher. The questions in the scale were: How credible is the ad? (Not at all credible / Very Credible). How believable is the ad? (Not at all Believable / Very Believable). What is your overall evaluation of the ad? (Very Negative / Very Positive). How trustworthy or untrustworthy is the ad? (Not at all Trustworthy / Very Trustworthy). How biased or unbiased is the ad? (Very Biased / Not at all Biased). To test how well our scale performed, we ran an Internal Reliability test and found our Cronbach's Alpha to be .878. This means that our scale shows very high internal reliability with the scale questions, and we can accurately rely on it to perform.

Survey Design and Implementation

Our survey consisted of 21 questions designed to answer each of our hypotheses and our overall question of: "What are consumer perceptions of AI in advertising?" We had skip logic at the very beginning of our survey that would push anyone who is under 18 or does not want to participate to the end of the survey. The survey consisted of 21 total questions, 16 interval questions, including 2 nominal questions, consisting of our skip logic and a question asking how respondents identify. There were also 2 ordinal questions within the survey, the first asked how many times a day respondents watch advertisements of any kind, and the second asked to specify their age. After the skip logic, we started by asking a few questions to understand our respondents' familiarity with advertisements in general and AI within advertising. We were looking for answers to factors like how comfortable respondents are with AI in marketing, and how well consumers thought they were at recognizing and identifying ads that had AI. On top of including questions that we felt would provide sufficient information for our hypotheses, we also included our scale questions of Attitude Toward the Ad (Credibility). This was a set of 5 questions in which we asked respondents to put themselves "in the mindset of watching an advertisement generated with the help of AI" before responding. This scale asked consumers questions about credibility, believability, trustworthiness, bias, and their overall evaluation of AI advertisements. The end of the survey included our demographic questions, in which we asked respondents to specify their age and gender.

For this survey, we used non-probability sampling methods of convenience and snowball sampling. We distributed this survey through several different platforms, trying to reach as many people as possible. We chose the most convenient means of communication to reach as many of our peers, family, and friends as possible. We posted to Instagram, Snapchat, Facebook, LinkedIn, sent the survey directly via email and instant messaging, and used word-of-mouth to get our survey out. We had also asked friends and family members to continue to send the survey to other people they knew who would be willing and able to participate in our survey. In doing this, we were able to gather 135 total responses, and after editing and culling, we had 108 responses to gather data and run statistical tests on.

Sample Profile

In terms of the sample profile, findings revealed that a majority of respondents (65%) identified as women, and most respondents are within the 18-24 age range (84%). Most of the respondents reported that they watch 11-50 advertisements a day, and are average or better at identifying AI advertisements. This would mean it is safe to assume that our respondents are familiar with both advertisements and AI in general. We wished to know if AI also affected the companies that produced the ads, so we asked a question regarding the consumer's opinion on certain brands and the usage of AI. The answers reported in our survey suggested that advertisements may not have a large effect on viewers' opinions of those companies that produce the AI ad. As previously mentioned, our team used both convenience and snowball sampling. These non-probability sampling methods allowed us to reach a defined target population and gather a large number of respondents in a relatively short time, given the time constraints of our survey being active.

DATA ANALYSIS AND FINDINGS

Dependent Variable for all Tests:

Survey Question #5: On a scale from 0-10 how favorable is your perception of AI use in marketing? (0 = Very unfavorable, 10 = Very favorable)

Hypothesis	Statistical Test	Question #	R Square	F Stat	P Value
1 The more transparent companies are about the usage of AI in their ad, the more positive the consumer perception is of the usage of AI in the ad.	Bivariate Regression	13. To what degree do you agree with this statement? Advertisements that utilize AI should have a disclaimer.	.042	4.62	.034 Hypothesis Supported
2 The higher the quality of an AI ad is, the more positive the consumer perception is of the usage of AI in the ad.	Bivariate Regression	7. Do you feel advertisements that use AI are higher or lower in quality?	.190	24.71	<.001 Hypothesis Supported
3 The more credible the AI ad is perceived to be by the consumer, the more positive the consumer perception is of	Bivariate Regression	Credibility Scale	.334	50.702	<.001 Hypothesis Supported

the usage of AI in the ad.					
4 The more the consumer believes the AI ad is generated by AI. The more negative the consumer perception is of the usage of AI in the ad.	Bivariate Regression	8. When you notice an ad using AI, how does it make you feel?	.296	44.635	<.001 Hypothesis Supported
Hypothesis	Statistical Test	Question #	Means	T-stat	P Value
5 The older the consumer is, the more positive the consumer perception is of the usage of AI in the ad.	Independent Samples T-Test	21. What age range do you fall in?	18-24: 3.736 25+ 4.529	T= -1.167	.246 Hypothesis Not Supported

We found that four out of five of our proposed hypotheses had statistical significance. We felt that with a larger sample size of an older demographic, this last hypothesis would have been proven to be statistically significant because there is something to be said about the difference of means between the younger and older sections. It is clear that the younger population, on average, has a lower perception of AI in advertising based on the means. We also understand that our non-probability-based sampling limits the generalizability of these results. A different sample using probability-based methods could bring different results. Looking at our first hypothesis, we can see that it is statistically significant as it is less than the alpha of .05. However, a deeper story is revealed by our R-squared and F-stat values; they are small values, so while the model will predict, it will not do so very well. We rethought our earlier thinking that, if consumers are explicitly told that what they are about to see has AI, then they would be more receptive to the ad. Instead, we now think that even when you tell consumers that AI will be in the ad, there is a certain number of respondents and consumers who have an immediate negative connotation with AI, no matter what. Our 2nd hypothesis found that producing an advertisement that is of higher quality is more likely to produce positive consumer perceptions of AI in advertising. Being statistically significant with a P value of <.001 with the supportive F-stat of 24.71, this model shows that our assumption is correct and accurately predictable. Our credibility scale was our most supported hypothesis. With a p-value of <.001 and an F-stat of 50.7, this hypothesis may be the best to take action upon. Our study concluded that creating a credible AI ad is the best way to increase consumer perceptions. Moving on to our 4th hypothesis, a p-value of <.001 shows us this test is statistically significant with a f-stat of 44.6 and an R-squared of 29.6%, which supports that this hypothesis is a good predictor model.

CONCLUSIONS AND RECOMMENDATIONS

Overall, our findings in this research study show that consumers are not as skeptical of AI in advertising as companies may assume. In fact, people are more than willing to accept it as long as it is used thoughtfully, honestly, and at a high level of quality. Transparency stood out as one of the most important factors when it comes to using AI. When companies are open in acknowledging that AI played a role in creating an advertisement, viewers actually accept that honesty with a more positive perception of the brand. It makes the company seem modern, trustworthy, and confident in its creativity. That perception changes quickly, though, when the AI itself looks rushed or obvious to spot out. High-quality AI enhances credibility and makes brands appear more innovative, but low-quality AI hurts the overall immersion and gives the impression that the company might be cutting corners or relying too heavily on shortcuts. Essentially, it shows the viewers that the company took the easy and lazy way out. At the end of the day, believability is what holds everything together and makes an AI ad feel authentic. If an ad feels real, relatable, and emotionally rooted, the presence of AI doesn't bother the consumers at all. The only time AI becomes a problem is when the viewer can clearly see it, and that is when perceptions of it turn negative.

One surprising detail was the role of age in the perception of artificial intelligence use. We expected younger audiences to be more critical, yet open to AI, but our results showed that older respondents actually had slightly more positive perceptions overall. Age didn't significantly affect attitudes statistically, but the trend still showed that the assumption that AI advertising only resonates with younger consumers could possibly be false. This opens the door for marketers to confidently use AI in campaigns targeted at older demographics as well. This is as long as the content remains in good quality and is relatable and realistic.

From a strategic standpoint, these findings suggest that brands should not avoid AI use. Instead, they should commit to using it well and in a strategic fashion. If AI is going to be part of an ad, it needs to be high-quality, intentional, and invisible. Companies should be upfront about using AI, but the creative execution has to feel natural and credible. AI should support the story, not distract from it. Believability and emotional truth still matter more than any additional tool behind the scenes.

On the recommendations side, we also learned a lot from our own research process when looking back at it. The recruitment language we used, particularly phrases like "young adult consumer perceptions," unintentionally discouraged older audiences from participating. Removing that language would help expand the sample and get a more inclusive view of different age groups. We also feel that other sampling techniques should have been used to ensure those older age groups are more represented, in terms of number of respondents. We had a lack of older respondents and that possibly led to there not being statistical significance in our age hypothesis. Adding a short progress indicator within the survey could also help increase completion rates by reassuring respondents that they are close to finishing as they are taking it. Finally, it's important to acknowledge that AI is still an emerging field with a limited amount of prior research, and our own time constraints kept us from diving deeper into the existing research of AI perception. A larger sample and more time would allow for stronger statistical testing and potentially a more in-depth analysis.

This study shows that AI in advertising is not something consumers are going to reject when they see it. Customers and viewers of these ads just expect that brands will use AI responsibly, ethically, and at a high standard. When those expectations are met, AI becomes an asset instead of a liability and with continued research like this study, brands can leverage AI to strengthen trust, creativity, and audience connection.

RFERENCES CITED

Aboulnasr, K., and Y. S. Song. 2025. "Emotional Responses to Digital Resurrections: Consumer Reactions to AI-Generated Dead Celebrity Endorsements." *Journal of Consumer Behaviour* 1– 22. https://doi-org.ezproxy.siena.edu/10.1002/cb.70029.

Arango, L., Singaraju, S. P., & Niininen, O. (2023). Consumer responses to ai-generated charitable giving ads. *Journal of Advertising*, *52*(4), 486–503. https://doi.org/10.1080/00913367.2023.2183285

Bhagat, R., Chauhan, V., & Bhagat, P. (2022). Investigating the impact of artificial intelligence on consumer's purchase intention in e-retailing. *Foresight*, *25*(2), 249–263. https://doi.org/10.1108/fs-10-2021-0218

Campbell, C., Plangger, K., Sands, S., Kietzmann, J. (2022). Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising. *Journal of Advertising, 51*(1), 22-38. https://research.ebsco.com/c/chbvno/viewer/pdf/gmtkpjyhpv?route=details

Chen , H., Chan-Olmsted, S., Kim, J., & Mayor Sanabria, I. (2021). *Consumers' perception on* *artificial intelligence applications in marketing communication*. ProQuest. *Romanian Journal of Human-Computer Interaction. 11(3), 220-231.* https://www.proquest.com/scholarly-journals/artificial-intelligence-perception-expectations/docview/2213034601/se-2

Ciuchita, R., Gummerus, J. K., Holmlund, M., & Linhart, E. L. (2022). Programmatic advertising in online retailing: Consumer Perceptions and Future Avenues. *Journal of Service Management*, *34*(2), 231–255. https://doi.org/10.1108/josm-06-2021-0238

Feng, Y. & Kim, J. (2025). Decoding the Trust Matrix: Unraveling Key Predictors of Consumer Trust in AI-Generated Personalized Advertising. *Journal of Interactive Advertising 25*(2), 123-138. https://research-ebsco-com.ezproxy.siena.edu/

Jiang, H., Cheng, Y., Yang, J., & Gao, S. (2022, September). *AI-powered chatbot communication with customers: Dialogic interactions, satisfaction, engagement, and customer behavior*. ScienceDirect.https://www.sciencedirect.com/science/article/abs/pii/S0747563222001510

Lu, SH., Tran, H.T.T. & Ngo, TS. Are we ready for artificial intelligence voice advertising? Comparing human and artificial intelligence voices in audio advertising in a multitasking context. Qual Quant 59 (Suppl 1), 1–22 (2025). https://doi-org.ezproxy.siena.edu/10.1007/s11135-024-01967-x

Minton, E. A., Kaplan, B., & Cabano, F. G. (2022, November). *The influence of religiosity on consumers' evaluations of brands using artificial intelligence.* EBSCO Information Services, Inc. https://about.ebsco.com/products/research-databases

Vakratsas, D., & Wang, X. (Shane). (2021). Artificial Intelligence in Advertising Creativity. Journal of Advertising, 50(1), 39–51. https://doi-org.ezproxy.siena.edu/10.1080/00913367.2020.1843090

Zhang, Y. & Gosline, R. (2023) Human Favoritism, Not AI Aversion: People's Perceptions (and bias) Toward Generative AI, Human Experts, and Human-GAI Collaboration in Persuasive Content Generation. *Cambridge University Press. https://www.cambridge.org/core/journals/judgment-and-decision-making/article/*

CULTURE SHOCK

Executive Summary

Taste the World One Plate at a Time

Contact Information
Madison Milton
515 Loudon Road, Loudonville, NY 12211
Ms28milt@siena.edu

Industry
Hospitality
Meal Delivery

Development stage
Startup/Revenue/Profitable

Year founded
2026

Number of Employees
7

Funding Opportunity
$100,000

Use of Fund
30% Product Development
30% Marketing/Sales
25% Operation/Inventory
0% Existing Debt
15% Legal/Other ...

Current Monthly Revenue
$42,375

Existing Debt
$0

Existing Investors
Madison Milton [⅕; $20,000]

Leo Alvarez-Guzman [⅕; $20,000]

Tehillah Mbango [⅕; $20,000]

Ameer Lester [⅕; $20,000]

Abigail Trinkle [⅕; $20,000]

THE GRAB

Culture Shock is a ghost kitchen concept in Williamsburg, Brooklyn, that provides convenient access to diverse, multicultural cuisine through delivery, take-out, and street food while supporting a local nonprofit that promotes cultural unity. We aim to connect community members through delicious food and celebrations of heritage.

PROBLEM/OPPORTUNITY

50% of Gen Z and 48% of Millennials eat culturally influenced foods other than Mexican, Chinese, and Italian each week (Nation's Restaurant News, 2023). However, there are not currently delivery options for nontraditional international cuisines, despite the fact that these demographics order delivery 1.8 times per week on average (National Restaurant Association, 2025).

SOLUTION/PRODUCT

Culture Shock is a convenience-focused dining establishment, with menu offerings provided in a street food format for ease of travel and consumption. Culture Shock reaches our customers through three channels: third-party delivery apps DoorDash and Uber Eats, a takeout window, and a street food booth, all supported by our ghost kitchen. Our menu rotates to a new international cuisine and features a visiting chef every four months, sharing their story and spreading cultural heritage through our menu and social media.

POTENTIAL RETURN/REVENUE MODEL

The online food delivery market in the US was estimated at $34.9 billion in 2025, and is expected to grow 9.6% year over year from 2025-2030 (imarc, 2026; Grand View Research, 2026). This trend is even more pronounced in New York, with the industry growing 14.6% annually over the last five years (Lee, 2026). The rapid growth in the online food delivery and street food markets offers an exceptional opportunity for Culture Shock to develop a strong presence in the New York City area, which is experiencing higher growth rates than the US as a whole. We estimate earning a net profit of $400,000 by the end of our fifth year.

COMPETITION

Culture Shock faces 3 main strong competitors with around 100-150 different fusion restaurants within the area. The 3 strongest competitors we have in the neighborhood are Fushimi, Patio Tropical, and Pokito.(Lewin, n.d.),(Orlow, 2025). Important success factors are managerial experience, employee training, product quality, convenience, and menu variety. With all our competitors facing conflict in either one or two areas making their brand weak, which strengthens our competitive advantage.

EXECUTION PLAN/GO TO MARKET STRATEGY

Culture Shock will open our ghost kitchen simultaneously with our mobile street food booth, which will help to build awareness in the local community to drive traffic to our delivery and takeout options. Because we do not support in-person dining, our overhead and operating costs are reduced. This allows our menu's 300% markup to deliver a higher profit margin than traditional restaurants. Culture Shock's promotion strategy will focus on utilizing social media story-telling to draw consumers into our brand, educate them on international cuisine, and share information about our visiting chefs and menu changes. We want to ensure that we demonstrate to customers our reliability, authenticity, and social mission in all of our promotional materials. In pursual of social responsibility, we will partner with the United Cultural Association based in Brooklyn. We seek to support the UCA in achieving their mission by raising money and awareness for their causes and participating in their events.

FINANCIALS

We project deficits for the first two years of operations, with all income reinvested for development. We anticipate that Culture Shock will become profitable by the third year, and our goal is to expand into more locations after year five. We aim to reach a daily sales level of 250 meals per day by year five, supporting our projected net profit of around $400,000 in the fifth year.

THE TEAM/RELEVANT EXPERIENCE

Our team is composed of five members, Abigail Trinkle, Madison Milton, Leo Alvarez-Guzman, Ameer Lester, and Tehillah Mbango. Abigail has worked in non-profit settings, particularly focused on financial management.Madison has worked in the marketing department for Siena University's Damietta Cross-Cultural Center and the Black Student Union on campus. Leo has worked in various different industries, ranging from Customer Service to a bit of Sales and restaurant experiences**.** Ameer has worked in a digital marketing position for culinary foods and recipes that will prepare him for building awareness for Culture Chock. Tehillah has worked primarily in hospitality and non-profit settings, bringing real-world knowledge to apply within multiple management facets.

ENHANCING SCAM PREVENTION MESSAGING: A STUDY ON SOURCE CREDIBILITY AND AUTHENTICITY

AKM Alam, Siena University
Olivia Botelho, Siena University
Dr. Marie Rice, Siena University - Faculty Mentor

ABSTRACT

This study answers the 2024 call from the Federal Trade Commission's Scams Against Older Adults Advisory Group to examine senders of scam prevention messages. Scams against older adults are growing exponentially each year and, as a result, financial institutions, government agencies, advocacy groups and others exert significant effort and expenditure to prevent and detect losses from scams. Although the problem is widely known and garners headlines, little research has been done into the effectiveness of scam prevention messages. Using focus groups of older Americans, the study finds that middle-aged senders are more likely to be perceived as credible when communicating scam prevention messages. This study fails to meet saturation and will therefore continue in future months.

A STUDY ON SOURCE CREDIBILITY AND AUTHENTICITY

Based on a random survey of over 9,000 Americans conducted by the Pew Research Center, 73% of U.S. adults have been scammed (Gottfried, Park, and Anderson, 2025) and losses from scams topped $12.5 billion in 2024, a 25% increase over 2023 (FTC, 2025). As a result, financial institutions, governmental agencies, advocacy groups, and others post scam prevention messages, hold community events, provide training sessions and work diligently to raise awareness about various scam trends. These efforts have become commonplace in the U.S. and organizations have put forth significant effort to reduce the incidence of and losses from scams. In 2022, the Federal Trade Commission formed the Scams Against Older Adults Advisory Group to assist in the fight against scams. The Advisory Group generated a report in 2024, titled "A Review of Scam Prevention Messaging Research: Takeaways and Recommendations," in which they called for additional research into scam prevention messages, including the message sender's credibility (SAOA, 2024). Citing Bristol (1996), the Advisory Group specifically noted older Americans' preference for brands represented by middle-aged endorsers, as compared to younger or older endorsers, and suggested that a similar preference may exist with scam prevention messages (SAOA, 2024, p.8). They further suggest research into the effectiveness of scam prevention messages from "the government or a corporation" (SAOA, 2024, p.11).

Effective scam prevention messages may help reduce the incidence of victimization with older Americans and/or may reduce the losses incurred by these victims. Scam prevention messages are similar to social responsibility messages because of their shared goals of trust, education, and protection, which are driven by the message sender's credibility and authenticity. Joo, Miller and Fink (2019) explain that authenticity is comprised of seven dimensions, including benevolence, community link, transparency, reliability, commitment, congruence, and broad impact. Authenticity is further a dimension of credibility, along with accuracy and believability (Appelman and Sundar, 2016). This study examines scam prevention message authenticity and credibility by answering the

Advisory Group's (SAOA, 2024) research calls. Specifically, the study examines older Americans' perception of a scam prevention message sender's credibility and authenticity depending upon the message sender's age, and whether the message sender represents a local bank or a local government.

Older Americans often rely on advocacy groups, like the Association of American Retired Persons (AARP), as a reliable source of authentic and credible messages. This study solicits older Americans' perceptions about scam prevention messages using focus group participants from the AARP. Participants attend Zoom sessions where they are presented with three images generated with Microsoft Copilot that include women at three ages (e.g., 20, 40 and 70) from two local community organizations – either a local bank or a local government. Focus group participants quantitatively rate the credibility and authenticity of each image before engaging in guided discussions that provide deeper, qualitative information about their perceptions of the images. Following Eutsler and Lang (2015), quantitative data is collected using seven-point, labeled responses and measured using analysis of variance. Qualitative responses are evaluated following Braun and Clarke's (2006) six-step thematic analysis, in conjunction with Malsch and Salterio's (2016) commentary on the need for a theoretical framework, generating an abductive methodology (Timmermans and Tavory, 2012) that allows for both organic generation of data from participant responses and validation against a theoretical framework.

Preliminary results with a small number of focus group participants support Bristol (1996). Specifically, six focus group participants show preference for a middle-aged (40-year-old) scam prevention message sender, as compared to an older (70-year-old) or a younger (20-year-old) message sender.

---See Figures 1 through 3 --

However, participants did not favor one organization type (i.e., local bank or local government) over the other. While Malterud, Siersma, and Guassora (2016) explain that information power can be derived in narrow settings with dense data with as few as six focus group participants, this study will benefit from a greater number of participants. Therefore, the study will continue, soliciting additional participants with a goal of achieving saturation, as well as information power.

LITERATURE REVIEW

Scammers employ tactics and strategies that are similar to corporate social responsibility. Corporate social responsibility (CSR) considers and maintains the environmental and social responsibilities in an organization's operations and its stakeholder communications. Companies employing CSR do so at great personal cost to themselves (Brown & Dacin,1997) while scammers utilize these tactics to generate profit or monetary gain. Vulnerable elderly are susceptible to these tactics because of the rapid evolution of technology, which makes determining the authenticity of information they receive difficult. This difficulty increases as a result of the loneliness elderly individuals experience, particularly since the COVID-19 pandemic

CSR and Older Generations Susceptibility to Fraud

So there have been many cases where scammers and fraudulent groups will target someone who is remarkably susceptible to fraud and scamming and there are instances of what the scammer would look for in a target. Scammers will come up with fake stories and manipulate the target to build a sense of trust between them and the scammer wants this relationship to remain a secret and will tell the target to enforce secrecy. This secrecy would often be enforced with lies crafted should the target be questioned for financial transactions to keep the target under control. According to the Safeguarding retirement in the age of scams written by the TIAA Institute has stated that this is one of the many aspects that secrecy is used to make sure all the financial transactions are confidential and private (p.3, Safeguarding retirement in the age of scams). This secrecy is a key element in performing phantom fixation which is a scam that tailors the specific scam to the specific person and while it does take up time and resources the target will start to deepen the scam's premise and meet the target's unmet needs. In the Safeguarding retirement in the age of scams they label phantom fixation where this scam meets the targets unmet needs from financial independence to romantic relationships or relating to a sense of purpose (p. 8, Safeguarding retirement in the age of scams). These aspects and elements in the fraudulent behaviour and scandals being demonstrated could be correlated with CSR and how people define authenticity.

To look at authenticity and how it can be correlated back to CSR and the greater understanding of its relation to fraud, there needs to be a model of how people classify authenticity and how authenticity is being framed as. Authenticity could be defined as how genuine something is based on multiple assessments and what categories are checked off to formally classify it as authentic. There has been various research in many areas of business considering what authenticity is but there will be the usage of the seven dimensions of authenticity according to (Joo, Miller, & Fink 2019). The first and most important of the seven dimensions would be community link where the CSR's activities are linked to the social aspect of the organisation where it qualifies as the dimensions of authenticity and consistent with these ideas (Mazutis and Slawinski 2014). The second dimension is reliability in how the CSR activity is consistent with the business and the connection between what it says it does and what it actually does. There seems to be a reflection of whether the program is consistent and claims of what it means to do while giving no proof that it actually does (Browns et al., 2003; Grayson and Martinec, 2004).

The third dimension is commitment where it defines the degree to which the business will believe it's CSR that it will start to accrue even personal loss due to it supporting CSR. Beckman et al. (2009) says that commitment isn't just a claim that the company can make, the CSR activity is so embedded into the company's core values itself and the belief it has for the core values. The fourth dimension is that congruence is defined as the alignment of how organizations CSR and the business core values and that there is also an alignment of the brand's concept or an action. (Alhouti et al. 2016; Mazutis & Slawinski, 2014). The fifth dimension is where there are usually the second most important is benevolence and how stakeholders perceive CSR strategies are altruistic as opposed to only concerning profits (Alhout et al., 2016; Leigh et al., 2006; Spiggle et al., 2012). The sixth dimension is transparency, how open and available is this information regarding the CSR to the wider public and allowing stakeholders to understand what is happening within the organization (Joo, Miller, & Fink 2019). Finally the last dimension is broad impact which is similar to community link but it shows that the CSR can be used for multiple purposes and not something too specific according to Joo, Miller, & Fink (2019). There are seven dimensions of authenticity of the CSR activity that allows the stakeholders to believe them to be authentic and trustworthy that can be connected to the elderly generation and fraud.

When scammers utilize phantom fixation scams and other scams they want to make these scams more believable and seem to feign authenticity and trustworthiness. Utilizing these seven dimensions the scammers will make organizations and activities that align with each other that makes it more believable to a potential target audience to believe in it and start a connection and conversation with other scammers. Also, the stakeholders that were mentioned in the dimensions of authenticity could be loosely applied to the older generations who the scammers are targeting as depending on what the target values the scammer will tailor a scam that responds with the target. The scammer would use attributes such as benevolence and link two of the most important attributes where the scammer will start to appeal to the target's good nature and altruism to build a sense of trust so the scammer can exploit the target. Though there are many ways that scammers can utilize to get monetary gains through the seven dimensions of CSR activities, there are other aspects of authenticity that older generations could look into such as transparency and reliability of how much information is available to the wide public and how reliable the system is. By looking at these details specifically and then having the other aspects the target could determine if the scam was inauthentic and help with scam prevention.

Fraud v. Types of Authenticity and Communication

There are many dimensions of what determines an object or activity as authentic, there are also different types of authenticity where scammers could utilize these types of scams with these types of authenticity in mind. This also forces the target to understand that these different types of authenticity could be referred to as objective authenticity, existentialist authenticity, and symbolic authenticity. This is a strong connection of how the scammer might use these types of authenticity to feign a sense of trustworthiness and believability of these scams similar to how the seven dimensions of authenticity are used to feign trustworthiness as well. Though the objectivist, constructivist and existentialist perspectives are still intertwined they all still share definitions of authenticity and interplay amongst each other as well. This also builds on communication models of CSR where it also builds on messaging and communicating authenticity to the target individual.

As stated above there the different types of authenticity are enhanced by the three models of communicating CSR authenticity to these people. There are three different types of authenticity that study will be focusing on such as existential authenticity where existentialist motives are connected to the subject's authenticity similar to self-actualization schemes that scammers use to target (Molleda, 2010). There is also symbolic authenticity where there is subjective mental association with an object's authenticity though it would come as romance scams that scammers might use to feign a sense of belonging and trust (Bigne et al., 2009). Finally, an objective authenticity is authenticity arising from an interplay of other authentic messaging (Kolar and Zabkar, 2010). An example would be a message that has many credible organizations and facts to distract the target individual from the original intent the scammer has in mind. These types of authenticity are also being enhanced by other methods of communicating authenticity and how it can be received by the target individual.

There are methods of communicating authenticity and how it can be received by the target individual. There are mainly three methods of communication that will be focused which is the HSM process, attribution theory, and identity-brand management model. The identity-brand management model is when the internal system of stakeholders such as employees and executives are classified as the identity and the brand being seen as an exterior with perspective of customers (Meffert et al., 2012). The attribution theory where people start to attribute a certain feeling depending on whether it's positive or negative to the object's authenticity likening it to altruistic or egotistic motives (Becker-Olsen et al., 2006). And, finally there is the HSM process where the heuristic mode of processing information is using prior or subsidiary information that are somewhat guided by information that the targeted individual would know. Then there is systematic processing which is defined as an analytic orientation in which consumers access and scrutinise all information for the task and use all information in forming judgements (Chaeken et al., 1989). These methods are now discussed to affect how scammers might utilize these methods of communication and the different types of authenticity to exploit the targeted individual.

The scammers might utilize these methods of communication and enhance the message of the authenticity via targeting to enhance their schemes. Utilizing these ideas of identity brand would be commonly used for job scams and other professional emails regarding employment. HSM uses primarily lottery and sweepstakes scams which forces the targeted individual to use heuristic mode processing using prior information and biases to quickly make a decision regarding the scam. And, attribution theory is more or less how the target is customized to attribute or assimilate this particular concept regarding either positive or negative attributes. These attributes could be seen as either altruistic and/or egotistical and would be mostly seen in romance scams.

Social Awareness and Content of Information

Later (Perez, Garcia, & Liu 2019) confirms that relationships between the CSR company and consumer, coupled with the trust in the company and attitude towards the company shows a positive correlation towards trust and attitudes towards the company. The research presented shows that should the company devote more attention and time to both information specificity and social topic awareness are qualities of the CSR message that significantly improve consumer perceptions of message authenticity. This is usually in regards to government scams where the consumer would have a perception of the government trying to do the most amount of good helping the target and some fraudsters would use this information to target people susceptible to these. Also the lack of social topic awareness forces the consumer into doing more research and to verify that the CSR company is being transparent and providing reliable information that would improve consumer responses and build on trust and have a more favorable attitude towards the company. The more information the companies provide the more there is in transparency which would in turn allow for customer and consumers to positively respond to the message and increase the consumer knowledge regarding the social topic and improving social awareness and would require less effort to understand the CSR activity. Though there is a countereffect which simply states that because there is a large amount of information that there is a strong usage of HSM using short cuts and prior information to help analyze the information given. This could be used to confuse the target individual to make it seem much more likely for them to fall into the scam.

When CSR companies utilize this they seem more trustworthy and this is where scammers yet succeed to convey through their messaging elderly victims. Due to the mental health of certain elderly adults, sometimes overwhelming information can be hard to process, especially with scammers using emotional arousal to cloud the part of the brain that processes information. Along with other age-associated risk factors such as cognitive impairment leading to a great scam susceptibility and impacts skills like critical thinking, working memory and other vital cognitive skills to recognize scams. Where scammers use this information to their advantage by making their messages more simple and easy to understand even though it might seem inauthentic, elderly victims can't help but fall for it. This along with the state of isolation and loneliness since COVID-19 makes it harder for them to talk to others about the information being presented to the senior citizens including other family members and other community people trying to help the senior citizen in processing the information.

RESEARCH DESIGN AND METHODOLOGY

Qualitative research methods, such as focus groups, allow participants to elaborate on topics in ways that generate richly detailed responses. However, the group setting can inhibit candid responses, particularly when sensitive or personal issues are involved. For this reason, focus groups are most effective in studies that explore perceptions, opinions, and contextual factors areas where the subject matter is less personal and unlikely to trigger social desirability bias. This study aligns well with focus group design as participants' perceptions of images and responses to non-personal questions are collected.

The study solicits focus group participants from volunteers of the AARP. Participants were invited to participate in the study during their virtual monthly meeting and by email invitation from a state-level Executive Director. Those interested in participating were provided with a link to Qualtrics where they acknowledged the Informed Consent, were asked to select a day and time for participating in a focus group, and they were asked to provide an email address to receive the Zoom link for the focus group meeting. Participants were reminded of the focus group day and time and received the Zoom link one week prior to the session.

AARP volunteers meet virtually each month to share information, training, and discuss advocacy for older Americans. Because of participants' familiarity with Zoom, focus groups were held on Zoom, utilizing a waiting room. To protect anonymity, participants were instructed to keep their cameras off; however, they were advised that their voices might still be recognizable to others. Upon entering the session, participants were placed in the waiting room and assigned a participant number in place of their names. All participants were reminded to maintain confidentiality regarding the responses and any personally identifiable information shared during the discussion. A total of six participants engaged in focus group sessions that lasted approximately one hour.

RESULTS AND DISCUSSION

Two focus group sessions of six participants generated twenty-six pages of transcribed text, one hundred-forty-eight quotes, 7,173 words included in 1,071 groups of words. Participants responses were coded manually and using Atlas.ti, generating eighty-one codes that were summarized in six themes. Themes included authenticity, credibility attributes, engagement factors, message elements, sender age, and sender attributes. These aligned with source credibility attributes (Pornpitakpan, 2004).

Focus group participants' responses showed that they perceived the image of the middle aged woman to be the most authentic and credible, as compared to the images of the older and younger women. These perceptions were not influenced by whether the image was associated with a local bank or a local government. Specifically, participants perceived the image of the older woman to not be credible, but, rather, questioned her cognitive abilities, by claiming:

> *"She looks like she could have a dementia. I mean, I'm not saying she does, but she looks a little confused."*

They also explained that they felt sorry for the image of the older woman, and that "...she looks kind of unhappy", "She seems to be someone who potentially was a victim", and "...she's got that victim look on her face, so… you know, I feel sorry for her…" These statements show that participants found the image of the older woman to lack credibility and to be an ineffective scam prevention message sender.

They similarly felt that the image of the younger woman lacked credibility because she lacked experience. They explained that her naivete impedes her ability to be effective at scam prevention messaging, making comments such as:

> *"Maybe it's my own bias, but with a younger person, I would find that without a caveat, I would have think* [sic] *that the person wouldn't have been exposed to these scams," and*

"She's younger, and I don't mean to be biased about that, but I don't know if it would make me even read the thing."

However, one participant explained that the image of the younger woman might be perceived as more credible if she were aligned with a role in the organization. They stated, *"Maybe if it were pitched as a volunteer for an organization, let's say a high school or university that was assisting seniors. That would be credible."*

Conversely, participants felt that the image of the middle-aged woman *"Appears authoritative…"* They explained that the credibility of the middle-aged woman would further be enhanced if her clothing were modified to indicate a role of authority with the organization she represented. One participant explained that,

"...some further images, information in the photo or something, or giving a name and saying that this is the manager or someone who has a title which correlates with a level of trust."

While participants expressed direct statements about the images of the women in the artificially-generated scam prevention messages in this study, they were less convicted with regard to the organization type. For example, when the organization listed on the message was a "local government," participants explained that "*…there's a lot of cynicism with the government in general*" and alternatively that, "*I'm kind of the generation where government can be progressive and can do good things, especially the local government.*" Likewise, participants' opinions regarding the credibility of a local bank delivering scam prevention messages were conflicted by making statements such as:

"I'm concerned with the commercial bent of that organization. And I think that they need to do a better job in separating the advocacy from the money making parts of their business,"

"I think the bank has a greater reason to be concerned. It has more risk and more opportunity to do good," and

"Well, for me, the big thing is the local bank and my mother had experienced a scam. And it was a local banker who spent the most time."

Thus, focus group participants' perceptions of the credibility of either a local government or local bank sending scam prevention messages depended on their firsthand experiences with each organizational type.

Preliminary results from our study show consistency with Bristol's (1996) findings, that older Americans prefer to receive messages from middle-aged adults. However, the number of participants in the focus groups does not yield saturation. While the number of participants provides sufficient information power (Malterud, et. al, 2016), additional participants are needed to achieve saturation and provide greater reliability.

LIMITATIONS AND CONTRIBUTIONS

Like all studies, this study suffers from limitations related to the research design and execution. The greatest limitation to qualitative research is its lack of generalizability. Further, employing focus groups could result in participant identification, if participants recognize each other's voices. While we take measures to protect participant anonymity by changing participant's screen name to their participant number and keeping participant cameras off, it is possible that participants might recognize each other's voices. The use of focus groups generates results that may not be extended to other settings and also suffer from self-desirability bias.
Furthermore, participants in this study were recruited from AARP volunteers, a group that is knowledgeable

of scams and insular. The study subsequently suffers from self-selection bias and may include omitted correlated variables that influence participants' perceptions of the scam prevention messages evaluated during the focus group sessions. Finally, only six individuals participated in our focus group sessions, resulting in a lack of saturation.

This study contributes to source credibility messaging, fraud, and anti-scam literature. Specifically, it answers a call for additional research into scam prevention message senders. The study expands upon Bristol (1996) by showing that older Americans' perceptions about message sender age generalizes to scam prevention messages. Specifically, like Bristol (1996), this study finds that older Americans prefer to receive scam prevention messages from middle-aged adults, as compared to older or younger adults. Future studies may evaluate these findings in other contexts or domains, or they may examine different attributes of scam prevention messages.

REFERENCES

Alhouti, S., Johnson, C. M., & Holloway, B. B, (2016). Corporate social responsibility authenticity: investigating its antecedents and outcomes. *Journal of Business Research,* 69, 1242-1249

Appelman, A. and Sundar, S.S., 2016. Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59-79.

Becker-Olsen, K.L., Cudmore, B.A. and Hill, R.P. (2006), "The impact of perceived corporate social responsibility on consumer behavior", *Journal of Business Research,* Vol. 59 No. 1, pp. 46-53.

Beckman, T., Colwell, A., & Cunningham, P. H. (2009). The emergence of corporate social responsibility in Chile: The importance of authenticity and social networks. *Journal of Business Ethics, 86,* 191-206

Bilz, A., Shepherd, L.A. and Johnson, G.I., 2023. Tainted love: A systematic literature review of online romance scam research. *Interacting with Computers*, *35*(6), 773-788. Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Bigne, E., Curras R. and Sanchez, I. (2009), "Brand credibility in cause related marketing: the moderating role of consumer values", *Journal of Product and Brand Management*, Vol. 18 No. 6, pp. 437-447.

Bristol, Terry. "Persuading Senior Adults: The Influence of Endorser Age on Brand Attitudes." *Journal of Current Issues and Research in Advertising*, vol. 18, no. 2, 1996, 59–67, https://doi.org/10.1080/10641734.1996.10505052.

Brown, T., & Dacin, P.A. (1997). The company and the product: Corporate associations and consumer product responses. *Journal of Marketing*, *61*(1), 68-84.

Burnes, D., Henderson Jr, C.R., Sheppard, C., Zhao, R., Pillemer, K. and Lachs, M.S., 2017. Prevalence of financial fraud and scams among older adults in the United States: A systematic review and meta-analysis. *American journal of public health*, *107*(8), e13-e21.

Button, M., Nicholls, C.M., Kerr, J. and Owen, R., 2014. Online frauds: Learning from victims why they fall for these scams. *Australian & New Zealand journal of criminology*, *47*(3), 391-408.

Button, M., Shepherd, D., Hawkins, C. and Tapley, J., 2024. Disseminating fraud awareness and prevention advice to older adults: perspectives on the most effective means of delivery. *Crime Prevention and Community Safety*, *26*(4), 385-400.

Chaiken, S., Liberman, A. and Eagly, A.H. (1989), "Heuristic and systematic information processing within and beyond the persuasion context", in Uleman, J.S. and Bargh, J.A. (Eds), *Unintended Thought,* The Guilford Press, New York, NY, pp. 212-252.

Coluccia, A., Pozza, A., Ferretti, F., Carabellese, F., Masti, A. and Gualtieri, G., 2020. Online romance scams: relational dynamics and psychological characteristics of the victims and scammers. A scoping review. *Clinical practice and epidemiology in mental health: CP & EMH*, *16*, 24.

Coombs, W.T. and Holladay, J.S., 2012. The paracrisis: The challenges created by publicly managing crisis prevention. *Public relations review*, *38*(3), 408-415.

Cosenza, T.R., Solomon, M.R. and Kwon, W.S., 2015. Credibility in the blogosphere: A study of measurement and influence of wine blogs as an information source. *Journal of consumer behaviour*, *14*(2), 71-91.

Creswell, J.W. and Creswell, J.D., 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

DeLiema, M., Li, Y. and Mottola, G., 2023. Correlates of responding to and becoming victimized by fraud: Examining risk factors by scam type. *International Journal of Consumer Studies*, *47*(3), 1042-1059.

Eutsler, J. and Lang, B., 2015. Rating scales in accounting research: The impact of scale points and labels. *Behavioral Research in Accounting*, *27*(2), 35-51.

Federal Trade Commission Consumer Advice (FTCCA). 2025. *Addressing Scams Affecting Older Adults*. Accessed November 15, 2025. Available from
https://consumer.ftc.gov/features/addressing-scams-affecting-older-adults.

Federal Trade Commission (FTC), 2025. Explore Data with the FTC.

Fugard, A.J. and Potts, H.W., 2015. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. *International journal of social research methodology*, *18*(6), 669-684. Gottfried, J., Eugenie Park, and Monica Anderson. 2025, July 31. Online Scams and Attacks in America Today: Pew Research Center.

Houtti, M., Roy, A., Gangula, V.N.R. and Walker, A.M., 2024. A survey of scam exposure, victimization, types, vectors, and reporting in 12 countries. *arXiv preprint arXiv:2407.12896*.

Jampen, D., Gür, G., Sutter, T. and Tellenbach, B., 2020. Don't click: towards an effective anti phishing training. A comparative literature review. *Human-centric Computing and Information Sciences*, *10*(1), 33.

Joo, S., Miller, E.G. and Fink, J.S., 2019. Consumer evaluations of CSR authenticity: Development and validation of a multidimensional CSR authenticity scale. *Journal of business research*, *98*, 236-249.

Kharraz, A., Robertson, W. and Kirda, E., 2018, May. Surveylance: Automatically detecting online survey scams. In *2018 IEEE Symposium on Security and Privacy (SP)* (70-86). IEEE.

Kim, S., 2019. The process model of corporate social responsibility (CSR) communication: CSR communication and its relationship with consumers' CSR knowledge, trust, and corporate reputation perception. *Journal of business ethics*, *154*(4), 1143-1159.

Kim, S.B. and Kim, D.Y., 2014. The effects of message framing and source credibility on green messages in hotels. *Cornell Hospitality Quarterly*, *55*(1), 64-75.

Kolar, T. and Zabkar, V. (2010), "A consumer-based model of authenticity: an oxymoron or the foundation of cultural heritage marketing?", *Tourism Management,* Vol. 31 No. 5, pp. 652-664.

Lane, P., McKenna, H., Ryan, A.A. and Fleming, P., 2001. Focus group methodology. *Nurse Researcher (through 2013)*, *8*(3), 45.

Leigh, T. W., Peters, C., & Shelton, J. (2006). The consumer quest for authenticity: The multiplicity of meanings within the MG subculture of consumption. *Journal of the Academy of Marketing Science, 34,* 481-493.

Levi, M., 2012. States, frauds, and the threat of transnational organized crime. *Journal of International Affairs*, 39-50.

Malsch, B. and Salterio, S.E., 2016. "Doing good field research": Assessing the quality of audit field research. *Auditing: A Journal of Practice & Theory*, *35*(1), 1-22.

Malterud, K., Siersma, V.D. and Guassora, A.D., 2016. Sample size in qualitative interview studies: guided by information power. *Qualitative health research*, *26*(13), 1753-1760. Morhart, F., Malär, L., Guèvremont, A., Girardin, F. and Grohmann, B., 2015. Brand authenticity: An integrative framework and measurement scale. *Journal of consumer psychology*, *25*(2), 200-218.

Mazutis, D. D., & Slawinski, N. (2014). Reconnecting business and society: Perceptions of authenticity in corporate social responsibility. *Journal of Business Ethics, 131,* 1-14.

Meffert, H., Burmann, C. and Jaworski, B.J. (2012), *Marketing - Grundlagen Marktorientierter Unternehmensführung,* Gabler, Wiesbaden.

Molleda, J. C. (2010). Authenticity and the construct's dimensions in public relations and communication research. *Journal of Communication Management, 14,* 223-236

Mouncey, E. and Ciobotaru, S., 2025. Phishing scams on social media: An evaluation of cyber awareness education on impact and effectiveness. *Journal of Economic Criminology*, *7*, 100125.

Norris, G., Brookes, A. and Dowell, D., 2019. The psychology of internet fraud victimisation: A systematic review. *Journal of Police and Criminal Psychology*, *34*(3), 231-245. O. Nyumba, T., Wilson, K., Derrick, C.J. and Mukherjee, N., 2018. The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and evolution*, *9*(1), 20-32.

Park, D.H., Kim, H.K., Choi, I.Y. and Kim, J.K., 2012. A literature review and classification of recommender systems research. *Expert systems with applications*, *39*(11), 10059-10072. Pinjarkar, L., Hete, P. R., Mattada, M., Nejakar, S., Agrawal, P., & Kaur, G. (2024, July). An Examination of Prevalent Online

Scams: Phishing Attacks, Banking Frauds, and E Commerce Deceptions. In *2024 Second International Conference on Advances in Information Technology (ICAIT)* (Vol. 1, 1-6). IEEE.
Perez A. (2019). Building a theoretical framework of message authenticity in CSR communication. *An International Journal,* Vol. 24 No. 2 pp. 334-350.
Pornpitakpan, C., 2004. The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of applied social psychology*, *34*(2), 243-281.
Scams Against Older Americans Advisory Group (SAOA). 2024. A Review of Scam Prevention Messaging Research: Takeaways and Recommendations: Federal Trade Commission. Sim, J. and Waterfield, J., 2019. Focus group methodology: some ethical challenges. *Quality & quantity*, *53*(6), 3003-3022.
Spiggle, S., Nguyen, H. T. & Caravella, M. (2012). More than fit: Brand extension authenticity. *Journal of Marketing Research, 49,* 967-983
Timmermans, S. and Tavory, I., 2012. Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological theory*, *30*(3), 167-186.
Wang, S.Y.K. and Huang, W., 2011. The evolutional view of the types of identity thefts and online frauds in the era of the internet. *Internet Journal of Criminology*, *12*, 1-21.
Whitty, M.T., 2025. A systematic literature review of profiling victims of cyber scams: setting up a framework for future research. *Cogent Social Sciences*, *11*(1), p.2563781. Zhou, S., Liu, X.F., Nah, F.F.H., Harrison, S., Zhang, X., Zhen, S., Yeung, D., Hsiao, J.H.W., Lc, R., Chan, A.B. and Wang, X., 2024, June. Understanding and fighting scams: Media, language, appeals and effects. In *International Conference on Human-Computer Interaction* (392-408). Cham: Springer Nature Switzerland.Zabkar

FIGURES AND APPENDICES

Figures

Figure 1 Older Message Sender, 70-years-old

Figure 2 Middle-Age Message Sender, 40-years-old

Figure 3 Younger Message Sender, 20-years-old

Figure 4 Word Cloud from Focus Group Responses

Appendix A: Guided Discussion Questions for Scam Prevention Messaging Focus Groups

Effectiveness of Anti-Scam Messages

1) Clarity and Simplicity:

a) How easy was it to understand the message? Were there any parts that were very clear or very unclear?

b) How does the sender's credibility affect your perception of the message's clarity and actionable advice?

2) Visual Elements:

a) What did you think of the visuals used in the message (e.g., infographics, warning icons)? In what ways did they help you understand the message better or did they make the message more confusing?

b) Are there any additional visual elements you think would improve the message? Should some of the visual elements included be removed?

Attributes of the Message Sender

1) Credibility and Reputation:

a) On a 1 to 7 scale, where 1 is not familiar and 7 is very familiar, how familiar are you with the sender of the message? (If 5 or above, how are they familiar to you?)

b) How credible did you find the sender of the message? What made you trust or distrust the sender?

c) Would you be more likely to follow the advice in the message if it came from this sender, or if it came from a different sender? If so, who?

2) Empathy and Understanding:

a) Did the sender seem to understand your concerns and fears about scams? How did they convey this understanding?

b) What would make you feel more or less reassured by the sender? How could they do this?

3) Consistency and Action:

a) Does consistency in messaging from the same sender make you more or less likely to trust or distrust the message?

b) On a 1 to 7 scale, where 1 is not likely and 7 is very likely, how likely are you to act on the message if you trust the sender?

Post Hoc Questions

1) In what ways does the sender's specific knowledge of preventing scams influence your perception of the message?

2) In what ways does the sender's specific knowledge of fighting financial crimes affect your perception of the message?

3) In what ways does the sender's organization influence your perception of the message? 4) Can you think of any senders whose messages you would automatically trust or distrust? Why?

FEDERAL RESERVE CHALLENGE 2025

OCTOBER 2025

Presentation Team: Kylie Castellana, Bradley Fraser, Andrew Kirk, Giovanna Perez, Kerry Wise
Research Team: Emily Film, Josh Florence, Brendan Kasuba, Elizabeth Lankford, Micah Moore, Christopher Nass, Jack Reilly, Sophia Tagliafierro, Natalie Zhu
Dr. Aaron Pacitti, Siena University – Faculty Mentor

ABSTRACT

The Federal Reserve Challenge team analyzes current macroeconomic conditions and presents a monetary policy recommendation based on these findings. Given on the state of the economy in the fall of 2025, the team argued that the Fed should lower interest rates by 25 basis points given the evidence of a softening labor market, policy uncertainty, and the non-trivial risk to inflation.

This analysis and monetary policy recommendations were made based on the prevailing macroeconomic conditions in October of 2025 as part of the ECON 460 Federal Reserve Challenge course.

Thank you for joining us today to discuss the current state of the economy and our monetary policy recommendation. This recommendation will be supported by a more rigorous discussion of labor markets, inflation, and financial conditions.

MONETARY POLICY RECOMMENDATION SLIDE:

Based on our reading of macroeconomic dynamics, we recommend lowering the Federal Funds Rate by 25 basis points to a target range of 3.75 - 4 percent. We believe at this rate, we can successfully balance both sides of our dual mandate. Current economic indicators suggest the economy is moving towards stagflation, with inflation slightly above target and a weakening labor market. Based on our analysis of the data, the softening labor market presents a greater downside risk to economic stability than the moderate increase in inflation.

Our recommendation is entirely data-driven and our sole focus is on achieving full employment and price stability. However, there is a large amount of policy uncertainty -- tariffs, deportations, the fiscal impacts of the One Big Beautiful Bill Act, and the recent government shutdown, all of which will likely have contractionary effects.

The government shutdown could act as a drag on growth because of the potential for the aggressive downsizing of government employment and reduced spending for furloughed government workers. Additionally, important government data releases could be delayed, complicating our data-driven approach. These forces could further reduce consumer, business, and investor confidence going forward.

This collective uncertainty has the potential to increase inflation along with unemployment, while slowing growth. This will require us to carefully monitor changing economic conditions and respond appropriately.

CHANGE IN PAYROLL EMPLOYMENT: LAST 4 MONTHS OF PAYROLL SLOWDOWN

I firmly believe that we should be cutting rates more aggressively. It is evident that there has been a meaningful and significant reduction in the pace of job growth. Starting in January of 2025, there has been a noticeable slowdown in the change of payroll employment. Payroll employment experienced a modest positive change from January-April, showing that the labor market was expanding at a steady pace. However, this growth began plummeting in May of 2025. The labor market experienced the first dip into negative payroll employment change before rebounding into positive gains in July. To be exact, job growth has slowed to 29,000 per month on average over the past three months. This slowdown occurred after the tariffs were announced in April of 2025, which suggests that companies were beginning to hold off hiring due to heightened policy uncertainty. **Although the September employment report was delayed due to the government shutdown, private estimates from ADP show that the economy lost 32,000 jobs last month. I view this downward trend in employment as the most significant issue facing the economy, as it reflects emerging signs of fragility in the labor market.**

UNEMPLOYMENT RATE:

I agree with my colleague, Although the unemployment rate is relatively inline with our full employment goal, it has ticked up from 4% to 4.3% since early January. Despite this low level, the increase is concerning. This upward trend in unemployment levels, combined with inflation still above target, raises concerns of possible stagflation. Additionally, due to lag effects when adjustments are made to the federal funds rate, it's important to cut rates aggressively to get ahead of the softening labor market. This reinforces our belief that maintaining full employment is our top priority, and cutting the federal funds rate aligns with this goal.

PCE & CORE PCE:

While I acknowledge the softening of the labor market, we can't lose sight of the other part of our dual mandate, price stability. Headline PCE inflation which is shown in black, and Core PCE inflation, shown in red, have been above our 2% target since March 2021. More concerning is that both measures have been increasing since April of this year. These trends suggest that we should maintain a more restrictive monetary policy stance to ensure price stability.

EFFECT OF 2025 TARIFFS:

Tariffs will put upward pressure on inflation, and based on forecasts, will have an adverse effect on economic growth. Estimates from the Yale Budget Lab suggest that tariffs are likely to lower GDP by up to 1 percentage point in the coming year, with a long-term effect of ONLY half a percentage point. Given this, I believe that inflation poses a bigger risk, meaning maintaining price stability should be the primary objective.

SAHM RULE:

I also support a more restrictive approach to monetary policy, due to the economic trends shown by the Sahm rule. When using the Sahm rule to check for indications of the start of a recession, by using the three-month moving average of unemployment, we are currently below its 0.5 threshold. Over the past year, the Sahm rule has been steadily decreasing, currently sitting at 0.15 as of August 2025. This reading suggests that earlier concerns regarding labor market weakness may have been overstated. This also indicates that the rise in unemployment rates may be due to an expanding labor market rather than an increase in layoffs. As the Sahm Rule does not currently signal an impending recession, it indicates that labor market conditions may be more resilient than previously anticipated. This relative stability provides policymakers with greater flexibility to concentrate efforts on addressing inflationary pressures.

EQUITY MARKET VOLATILITY TRACKER:

I acknowledge the competing views of my dovish and hawkish colleagues. Because of this, we are actively monitoring business sentiment as we contemplate the appropriate path for policy. Our monetary policy decision is a measured response, not a reaction to short-term volatility.

In our analysis of the equity market volatility tracker, our attention was caught by the notable spike of 1.9 index units in April 2025, which is the highest level year-to-date.

The initial uncertainty did not translate into a sustained shift in outlook. However, in July and August, we observed a gradual uptick, indicating that market sensitivity may be returning, and the recent government shutdown has the potential to accelerate this.

CONSUMER SENTIMENT:

While achieving price stability is important and business confidence seems stable, consumer sentiment tells a different story, suggesting larger rate cuts are needed. Consumer sentiment has decreased significantly from its peak in late 2024, and bottomed out in April 2025. This decline could have been caused by the April tariff announcement, the subsequent cancellation of the tariffs, and the following reintroduction of the tariffs, to say nothing of the government shutdown. The fall might also be attributed to more intensive immigration raids and deportations. While consumer sentiment has since seen a slight increase, it still hasn't fully recovered to its 2024 levels. Lower levels of consumer confidence may lead to lower investment and spending, in turn, forcing GDP growth to stall. This potential decline further raises concerns of higher unemployment.

10-YEAR INFLATION EXPECTATIONS:

10-year inflation expectations remain well anchored. The average expected inflation rate has stayed consistently between 2 and 2.5%. While this is still a tick above our 2% target, expectations have remained anchored. Despite inflation concerns, this indicator shows no heightened concern of potential surges along with a fairly stable inflation rate. This is an indicator that market participants are not overly concerned with inflation going forward, therefore, we have a buffer to conduct further rate cuts to meet our full employment obligation.

LONG-TERM UNEMPLOYMENT RATE:

Although the overall unemployment rate remains low, the percentage of unemployed workers who have been jobless for 27 weeks or more has risen significantly, reaching 25.7 percent, which is the highest year-to-date. This increase indicates that individuals who have lost their jobs are facing greater challenges in finding new employment.

A rise in long-term unemployment often signals that other labor market indicators may not fully capture the underlying softening in the employment landscape. This development points to potential structural issues in the labor market, which could impede workers' ability to regain employment and may weigh on the broader economic recovery. Increases in long-term unemployment, even with a low overall unemployment rate, can not just act as a drag on GDP growth, but also put downward pressure on wage growth. Both of these forces would serve to moderate inflation and provide more space for accommodative monetary policy. The weakness in the labor market showcases the importance of a larger rate cut.

WAGE GROWTH

Wage growth has slowed over the past year, suggesting a softening labor market. However, wage growth continues to outpace the rate of inflation, which indicates that the labor market, while exhibiting more slack, still remains on firm footing. Because of this, we need to be measured in our monetary policy response. Cutting rates too aggressively could lead to a resurgence in wage growth, causing an increase in inflation. But cutting rates too timidly could allow the labor market to contract to such a degree that it adversely impacts our full employment objective.

MARKET VOLATILITY TRACKER:

The announcement of tariffs in early April triggered a sharp increase in market volatility. The index surged to 52.33 in early April, the highest level seen since the COVID-19 pandemic. However, this spike proved short-lived, with volatility steadily declining and stabilizing by June. As of September, the volatility index has lowered to pre-shock levels, indicating a return to more typical market conditions. This quick stabilization shows a resilient economy despite events of high uncertainty. Cuts to the federal funds rate may be a premature overreaction that will adversely impact market conditions. However, we will aggressively monitor this in light of the government shutdown.

TREASURY YIELDS:

Building on my colleagues previous statements, we should be more mindful of cutting rates too soon. To support this claim we can look at the 10 and 2-year U.S. Treasury yields that have stabilized following their spike in April 2025. Both Treasury yields have moved broadly in the same direction, with the red measurements showing the 10-year and the black measurements the 2-year. As yield rates are largely dependent on expected inflation, this demonstrates that bond investors aren’t overly concerned about the potential of rising inflation coming from tariffs and rising deficits. Higher inflation rates are unlikely to materialize, meaning an additional rate cut is unlikely to cause upward inflationary pressure.

FINAL SLIDE:

After a thorough discussion of current economic conditions and incoming data, we recommend a 25 basis point cut to the federal funds rate, which will stimulate job growth while maintaining price stability. Taking into account policy uncertainty, we believe that this change will best fulfill our dual mandate. Our mandate has been complicated by recent indications that the U.S. economy might be entering a period of stagflation. There are indications of tariff policy, deportations, and the government shutdown affecting the economy, and we will monitor their full effects going forward.

We are aware that our decisions affect all American consumers and businesses, and we take this responsibility seriously. Monetary policy that creates the economic conditions that foster maximum employment and price stability is our one and only concern.

Thank you and we are happy to answer any questions you might have.

FOOD WASTE IN THE CAPITAL REGION: DISTRIBUTION NETWORKS & REGULATORY CONSTRAINTS

Tehillah Mbango, Siena University
Faculty Mentor: Dr. Cheryl Buff, Siena University

ABSTRACT

Food waste remains a persistent global issue with significant economic, environmental, and social implications. As the global population is projected to reach 9.8 billion by 2050, reducing food loss and waste is an increasingly important strategy in addressing rising food demand. Existing research has extensively quantified food waste across household, retail, and food service sectors and has identified a range of behavioral and psychological drivers influencing waste generation. At the household level, which is widely recognized as the largest contributor to food waste, studies emphasize the role of consumer habits, planning behaviors, and decision-making processes. Frameworks such as the Theory of Planned Behavior and the "squander sequence" highlight how heuristics, routines, and competing goals shape waste behavior (Block et al., 2016; Barone et al., 2019; Stancu et al., 2016). Additional research suggests that improving planning and food management skills can reduce waste at the household level (Romani et al., 2018).

Despite this, food waste continues to persist at high levels across sectors. This suggests that individual attitudes and intentions alone are not enough to explain why wasteful behavior remains consistent. While existing studies have successfully quantified waste volumes and identified consumer motivations, there is still limited understanding of the structural mechanisms that determine where surplus food ultimately ends up after it leaves the point of generation. Research examining redistribution systems highlights inefficiencies such as weak communication between suppliers and recipient organizations, unclear demand signaling, and regulatory constraints that limit effective redistribution (Sundgren, 2022). This points to a broader systemic issue that extends beyond individual behavior.

This study builds on prior survey-based research conducted on a college campus to further examine the relationship between behavioral and structural factors in food waste. The original study surveyed 72 students and tested multiple hypotheses related to behavioral predictors of waste, including attitudes, awareness, and perceived responsibility. None of the hypotheses were statistically supported, reinforcing the idea that psychological factors alone are not sufficient to explain food waste behavior.

Additional analyses conducted on the same dataset further support this conclusion. A majority of students (71.4%) reported discarding leftover food in dining halls, while significantly fewer students compost (13%) or save food (5.2%). Students reported a moderate level of agreement with feeling less responsible when food is served by staff ($M = 3.23$, $SD = 1.09$), but this perception showed only weak relationships with other behavioral and situational factors (all $r < |0.2|$) (Davis, Griffin, Drew, & Mbango, 2025). Regression analysis examining predictors of composting intention—including responsibility perception, perceived initiative effectiveness, and structural factors—explained only 7.6% of the variance ($R^2 = .076$), with no statistically significant predictors (Davis, Griffin, Drew, & Mbango, 2025).

At the same time, several structural factors demonstrated stronger relationships with one another. Portion size and food taste were positively correlated ($r = .338$, $p < .01$), as were dining hall hours and peer influence ($r = .413$, $p < .01$), and peer influence and lack of storage options ($r = .454$, $p < .01$) (Davis, Griffin, Drew, & Mbango, 2025). Students also rated food taste ($M = 4.15$) and portion sizes ($M = 3.51$) as the most important

factors influencing whether they discard food, suggesting that operational conditions may play a more significant role than individual attitudes.

Taken together, these findings reinforce a key takeaway: food waste is not solely a behavioral issue, but a structural one. Institutional practices such as portion sizing, service models, and infrastructure availability may constrain individual behavior, making waste difficult to avoid regardless of awareness or intent.

Moving forward, this research highlights the need to shift focus toward understanding how surplus food is managed beyond the point of consumption. Future research should examine how local redistribution networks function in practice, including how supply and demand are coordinated and how policy and infrastructure shape food disposition pathways. Addressing these structural gaps may be critical in developing more effective, system-level approaches to reducing food waste.

REFERENCES

Agmapisarn, C., Seetha, P., & Fakfare, P. (2025). Optimizing surplus food redistribution: Leveraging digital platforms to advance sustainable consumption. *Sustainable Development, 33*, 3831–3846. https://doi.org/10.1002/sd.3322

Armington, W. R., & Chen, R. B. (2018). Household food waste collection: Building service networks through neighborhood expansion. *Waste Management, 79*, 471–482. https://doi.org/10.1016/j.wasman.2018.04.012

Barone, A. M., Grappi, S., & Romani, S. (2019). “The road to food waste is paved with good intentions”: When consumers' goals inhibit the minimization of household food waste. *Resources, Conservation & Recycling, 149*, 97–105. https://doi.org/10.1016/j.resconrec.2019.05.037

Beretta, C., & Hellweg, S. (2019). Potential environmental benefits from food waste prevention in the food service sector. *Resources, Conservation & Recycling, 147*, 169–178. https://doi.org/10.1016/j.resconrec.2019.03.023

Block, L. G., Keller, P. A., Vallen, B., Williamson, S., Birau, M. M., Grinstein, A., Haws, K. L., LaBarge, M. C., Lamberton, C., Moore, E. S., Moscato, E. M., Reczek, R. W., & Tangari, A. H. (2016). The squander sequence: Understanding food waste at each stage of the consumer decision-making process. *Journal of Public Policy & Marketing, 35*(2), 292–304. https://doi.org/10.1509/jppm.15.132

Davis, S., Griffin, L., Drew, R., & Mbango, T. (2025). *Wasted potential: Understanding food waste on college campuses* (Unpublished manuscript). School of Business, Siena University.

Diaz-Ruiz, R., Costa-Font, M., & Gil, J. M. (2018). Moving ahead from food-related behaviours: An alternative approach to understand household food waste generation. *Journal of Cleaner Production, 172*, 1140–1151. https://doi.org/10.1016/j.jclepro.2017.10.148

Elimelech, E., Ayalon, O., & Ert, E. (2018). What gets measured gets managed: A new method of measuring household food waste. *Waste Management, 76*, 68–81. https://doi.org/10.1016/j.wasman.2018.03.031

Firmansyah, F., Park, I., Corona, M., Aphale, O., Ahuja, A., Johnston, M., Thyberg, K. L., Hewitt, E., & Tonjes, D. J. (2024). Variation in municipal solid waste generation and management across time and space. *Resources, Conservation & Recycling, 202*, 107472. https://doi.org/10.1016/j.resconrec.2024.107472

Hennchen, B. (2019). Knowing the kitchen: Applying practice theory to issues of food waste in the food service sector. *Journal of Cleaner Production, 225*, 675–683. https://doi.org/10.1016/j.jclepro.2019.03.293

Martin-Rios, C., Rogenhofer, J., Ruiz-Molina, M.-E., & Chan, E. K. (2025). Young professionals' attitudes toward food waste: A global study on awareness, behavior, and regional differences. *International Journal of Gastronomy and Food Science, 31*, 101110. https://doi.org/10.1016/j.ijgfs.2025.101110

Riesenegger, L., Santos, M. J., Ostermeier, M., Martins, S., Amorim, P., & Hübner, A. (2023). Minimizing food waste in grocery store operations: Literature review and research agenda. *Sustainable and Modern Distribution, 1*, 100023. https://doi.org/10.1016/j.samod.2023.100023

Romani, S., Grappi, S., Bagozzi, R. P., & Barone, A. M. (2018). Domestic food practices: A study of food management behaviors and the role of food preparation planning in reducing waste. *Appetite, 121*, 215–227. https://doi.org/10.1016/j.appet.2017.11.093

Sanders, R. E. (2023). Dynamic pricing and organic waste bans: A study of grocery retailers' incentives to reduce food waste. *Marketing Science, 43*(2), 289–316. https://doi.org/10.1287/mksc.2020.0214

Stancu, V., Haugaard, P., & Lähteenmäki, L. (2016). Determinants of consumer food waste behaviour: Two routes to food waste. *Appetite, 96*, 7–17. https://doi.org/10.1016/j.appet.2015.08.025

Sundgren, C. (2022). Circular supply chain relationships for food redistribution. *Journal of Cleaner Production, 351*, 130393. https://doi.org/10.1016/j.jclepro.2022.130393

Zhang, H. (2023). Reasonable motivations and barriers for college students to effectively sort household waste. *Journal of Environmental and Public Policy Research, 5*, 20220474. https://doi.org/10.54254/2753-7048/5/20220474

HANNAFORD PRELIMINARY HR AUDIT

Ashley Wilkinson, Siena University
Sylvan Obando, Siena University
Deborah Kelly, Siena University – Faculty Mentor

ABSTRACT

An HR audit is used to review Human Resource activities in a business in order to identify issues related to policies, practices, regulations, compliance, employee issues, and other HR focus areas. Students in the MGMT432 Strategic Human Resource Management capstone course have an opportunity to conduct a preliminary HR Audit of the business of their choice by collecting information on its HR practices and procedures from various sources. For this project, our group selected Hannaford as the organization to evaluate. Hannaford is a for-profit supermarket company operating within the retail grocery industry, making it an effective subject for analyzing HR practices in a large, customer-facing business environment. The HR Audit presentation slides include background information about the company, outlines of its current HR policies, procedures and practices, a SWOT analysis of the company's HR practices, and a preliminary reflection on possible strategic recommendations. This presentation will describe the project, the research process, limitations experienced, and observations.

INDIVIDUAL HEALTH EXPENDITURES: KEY DETERMINANTS

McKenna Rowley, Siena University
Ava Geer, Siena University
Barbara Borgonhone, Siena University
Matthew Witkop, Siena University
Dr. Manimoy Paul, Siena University – Faculty Mentor

ABSTRACT

This research investigates the primary factors influencing individual annual total direct healthcare expenditures in the United States, utilizing data from the 2023 Integrated Public Use Microdata Series Medical Expenditure Panel Survey (IPUMS MEPS). The study emphasizes the role of demographic, socioeconomic, insurance-related, and health-status elements in explaining variations in healthcare spending. A total of eighteen independent variables were analyzed, including age, sex, race, education, income, health status, usual source of care, and insurance coverage. Following dataset cleaning and recording categorical variables for analysis, various statistical methods were employed, including linear regression, K-nearest neighbors clustering, and principal component analysis. The results from the regression analysis indicated that demographic variables alone provided minimal explanatory power, accounting for merely 1.6% of the variation in healthcare expenditures. Socioeconomic factors yielded slightly better results, explaining only 7.1% of the variation, even when combined with demographic variables. Nevertheless, several variables consistently emerged as significant predictors, notably health status, usual source of care, private insurance, and Medicare coverage. The clustering analysis offered further insights by identifying distinct population segments. One cluster predominantly consisted of younger, lower-income individuals lacking private insurance, while another represented middle-aged, higher-income individuals with private insurance. A second clustering model revealed that younger, higher-income individuals generally reported better health status, whereas middle- to older-aged, lower-income individuals reported poorer health status. In summary, the findings indicate that healthcare expenditure is shaped by a complex interplay of access, coverage, income, and health condition, suggesting that traditional regression models may not adequately capture this complexity. These results contribute to a broader understanding of healthcare spending patterns and highlight the importance of combining predictive and exploratory analytic methods in business analytics research.

INTRODUCTION

Health insurance is a critical component of the United States healthcare system and plays a crucial role in the economy and individual financial stability. This also determines an individual's overall healthcare access. Understanding the factors that influence health insurance costs is important for understanding access and care for individuals. High health insurance costs can limit individuals' access to necessary medical services, and lower costs can improve overall health outcomes as well as financial outcomes (Centers for Disease Control and Prevention, 2024). Healthcare expenditure is also very important as healthcare policy makers view this spending to evaluate program effectiveness, such as Medicare and other insurance markets (U.S. Census Bureau, 2024). We have analyzed socioeconomic and demographic characteristics such as income, education, age, health status, etc., to view the effects they have had on individual annual healthcare expenditure.

To analyze health expenditure, this study utilizes data from the Integrated Public Use Microdata Series Medical Expenditure Panel Survey Data (IPUMS MEPS), which has provided national health insurance data on demographic, socioeconomic, and health-related variables in the United States. The dataset also includes factors such as income, education, insurance coverage, health status, lifestyle behaviors, and more, which allow for a more comprehensive data analysis of healthcare spending patterns, leading to the overall pricing. This data gives a strong foundation for how different factors and characteristics influence health insurance costs.

This study aims to examine how demographic, social, economic, and health factors contribute to health care spending. Individuals with a worse health status, limited medical care access, and lifestyle choices such as smoking, drinking, etc., are more likely to have higher healthcare expenses. This shows that all factors must be taken into consideration when analyzing data on health care expenses. By applying statistical models to this dataset as well as clustering models, this study seeks to find the most significant factors that drive health insurance costs and give insights into the patterns that affect these prices across different populations.

LITERATURE REVIEW

Research shows that demographic, economic, and social factors significantly influence an individual's annual total of direct healthcare payments in the United States. The paper "Direct Cost of COPD in the US: An Analysis of Medical Expenditure Panel Survey" shows that, while these categories of determinants are closely interconnected, they operate through distinct mechanisms. Demographic and economic factors primarily shape access to healthcare services and an individual's financial capacity to obtain care. In contrast, social and lifestyle factors more directly affect patterns of healthcare utilization and overall costs, as they influence health behaviors and the progression of disease.

Demographic factors such as income, education, and age are consistently identified as key predictors of healthcare expenditures. Higher levels of income and educational attainment are generally associated with increased spending on both health insurance premiums and out-of-pocket medical expenses, largely due to improved access to private insurance and healthcare services (DeNavas-Walt et al., 2012; Cutler & Lleras-Muney, 2010). In contrast, individuals from low-income backgrounds and minority populations are more likely to experience gaps in insurance coverage, which may reduce short-term healthcare spending but often lead to delayed treatment and higher long-term costs (Artiga & Hinton, 2018). This dynamic reflects the income–health gradient, whereby lower-income individuals tend to face poorer health outcomes and consequently incur higher average medical expenditures over time (AHRQ, 2020). Additionally, disparities related to race and geographic location further exacerbate inequalities in healthcare access and outcomes, with minority groups often experiencing reduced access to services and lower life expectancy (U.S. Census Bureau, 2023).

Health status and lifestyle behaviors further influence healthcare utilization and costs. Health status, particularly the presence of chronic conditions, is one of the most significant drivers of direct healthcare costs. Individuals with chronic illnesses require ongoing medical care, resulting in substantially higher expenditures compared to healthier individuals. Lifestyle behaviors, including body mass index (BMI), smoking status, and other health-related habits, further contribute to differences in healthcare costs by influencing an individual's risk profile and likelihood of requiring medical services (Cawley & Meyerhoefer, 2012).

Institutional and behavioral factors also contribute to variations in healthcare payments by influencing insurance enrollment and retention. Research indicates that individuals' decisions to enroll in health insurance are shaped by their perceptions of healthcare quality, trust in providers, and understanding of available insurance options (Berk & Schur, 1998). Policy interventions under the Affordable Care Act have demonstrated that targeted strategies, such as community outreach and personalized enrollment assistance, can effectively increase insurance coverage among previously uninsured populations (Sommers et al., 2015). Furthermore, while initial enrollment decisions are influenced by access and awareness, the renewal of insurance coverage is more strongly associated with customer satisfaction and perceived value of services rather than prior claims experience (Bhat & Jain, 2007)

In summary, the literature indicates that demographic factors primarily determine an individual's access to healthcare and financial ability to pay, whereas social and lifestyle factors exert a more immediate influence on healthcare utilization and direct costs. Understanding the relative impact of these factors is essential for developing policies aimed at reducing disparities in healthcare spending and improving overall health outcomes.

DATA

To accurately forecast a person's annual total of direct health care payments, it is important to consider demographic and social factors that could influence the total of payments. We analyzed data collected from the IPUMS MEPS from 2023. We identified 18 independent variables that could potentially affect the forecast. To determine the significance of these variables to their ability to forecast, we developed the following research questions:

1. What demographic and social variables heavily influence a person's annual total of direct health care payments?
2. What role do demographic variables play in influencing social variables in relation to a person's annual total of direct health care payments?
3. How do low-income individuals and high-income individuals compared to their health insurance plans affect the forecast of their annual total of direct health care payments?

Description Tables

Dependent Variable

LABEL	VARIABLE NAME	VARIABLE DESCRIPTION
EXPTOT	Annual Total of Direct Health Care Payments	Captures the sum of direct payments for care provided during the year, including out-of-pocket payments and payments by private insurance, Medicaid, Medicare, and other sources. Payments for over-the-counter drugs and indirect payments not related to specific medical events, such as Medicaid Disproportionate Share and Medicare Direct Medical Education subsidies, are not included in this amount.

Independent Variables

LABEL	VARIABLE NAME	VARIABLE DESCRIPTION
AGE	Age	Indicated the age of the person
SEX	Sex	Indicated whether the person was male or female
RACE	Main Racial Background	Main racial background, self-reported or interviewer-reported
US-BORN	Born in the United States	Whether or not the person was born in the US
SPK-OTH-LANG	Speaks a language other than English at Home	Indicates if they speak a language other than English at home.
EDU-C	The highest level of schooling an individual has completed	The highest level of schooling an individual has completed, in terms of completed grades for persons with less than a high school degree, in terms of degrees attained for high school graduates, and those with higher education
HI-DEG	Highest Degree Completed	Indicated the highest academic degree attained
INC-TOT	Total Personal Income	The sum of all person-level income for the current calendar year, excluding income from tax refunds and capital gains.

HEALTH	Health Status	Rates an individual's general health on a five-point Likert scale, ranging from "excellent" to "poor."
USUAL-PL	Has Usual Place for Medical Care	Indicates where they have a particular doctor's office, clinic, health center, or other place they usually go if they are sick or need advice about their health
HI-NOT-COV	Has No Health Insurance	Indicated whether the individual was not covered by health insurance at any time in the current calendar year
HI-PRIVATE	Has Any Private Health Insurance	Indicates whether the person had private health insurance coverage from any source for at least one day during the calendar year
HI-MCARE	Has Medicare Insurance	Indicates whether the person has Medicare coverage for at least one day of the calendar year
HI-OTH-GOV	Covered by Other Public Healthcare Coverage	Indicates whether the person for at least one day of the current calendar year, the person had other public healthcare coverage
MSDMQ-ADV	Health Provider Advised to Quit Smoking, Past 12 Months	For adults aged 18 and older who currently smoke and have visited a doctor in the past 12 months, a doctor advised the respondent to quit smoking
CANCER-EV	Ever Told Had Cancer	For adults aged 18 and older, it identifies respondents who have been diagnosed with cancer

SMOKE-NOW	Smokes Cigarettes Now	For adults aged 18 and older, indicates if the respondent currently smokes
EXER-MOD-VIG-5D	Spends 30+ minutes in physical activity	Indicated whether the respondent currently spends 30+ minutes in moderate or vigorous physical activity at least 5x per week

Some notes about this data set: data cleaning steps were done to ensure consistency and usability for modeling. Any response variables that were “unknown,” “not reported,” “no response,” etc., were removed from the data set. This was necessary to maintain the quality of the data and minimize ambiguity in analysis.

The continuous variables within our data set, Income and Age were also restructured into more meaningful categories. The income variable was grouped into three categories: low income (less than $56,000), moderate income ($56,000 to $167,000), and high income (above $167,000). These groups were based on “the income you need to fall in America’s lower, middle, and upper

classes – find out where you rank and how these social levels are defined” (Warren 2024). The age variable was grouped into three categories: young (18-34), middle-aged (35-64), and older (65 and above).

To have a complete understanding of the type of demographic captured within the data set we analyzed descriptive statistics for the continuous variables. The average age of participants is 54 years old and the average income of participants was $34,686. This allows us to understand throughout the analysis, their demographic is middle aged adults that have a low to moderate income.

Age	
Average	54
Minimum	18
Maximum	85

Income	
Average	$34,686
Minimum	$0
Maximum	$330,734

Additionally, our dependent variable, Direct Health Care Expenditure, is continuous. To better understand the volatility of our dependent variable the average, minimum, and maximum are stated below. These descriptive statistics allow us to understand a monetary range of expenditure, as well as, contribute to the understanding of how income and these payments are related.

Direct Health Care Payments	
Average	$12,315
Minimum	$0
Maximum	$275,269

Categorical variables were transformed into dummy (binary) variables to make them suitable for regression modeling. This allowed each categorical variable to be represented numerically, enabling the model to interpret qualitative information effectively.

Race was simplified into two categories: White and Other to ensure sufficient representation. Health status was categorized into three levels – good, moderate, and bad—based on the range of responses available in the survey. The highest degree completed variable was categorized into no degree, GED or diploma, bachelor's, master's or doctorate, and other degrees.

However, for clustering analysis and principal component analysis (PCA), the income and age variables were retained as continuous variables rather than being categorized. This decision was made to preserve the full variability of the data, allowing these methods to more effectively capture underlying patterns and relationships.

These transformations improved the interpretability of the dataset and ensured that all variables were in a consistent format for regression modeling.

RESULTS/ANALYSIS

Initially our literature review led us to perform two separate regression analyses on our data set, separating demographic variables from socioeconomic variables. However, when these regressions were utilized the variables independently had little explanatory power. Therefore, we performed a regression analysis combining both demographic and socioeconomic variables. From the regression output it is evident that demographic variables were outweighed by socioeconomic variables.

However, the evidence from the literature review suggests that demographic factors do influence expenditure. Therefore we conducted further regression analysis, binary logistic regression, with the three most significant socioeconomic factors that determined expenditure and as our dependent variables. This binary logistic regression aids in determining the relationship between demographic factors and influential socioeconomic factors.

Linear Regression

Model 1: Regression Analysis of Demographic and Socioeconomic Variables The model indicates that certain socioeconomic variables, such as health status, usual place of care, private insurance, and Medicare insurance, significantly influence healthcare

8

expenditure. The model only explains 7.1% of the variance in healthcare expenditures, which is not high explanatory power, however in the context of our separated regression, the combined model had the highest adjusted r-squared value.

The model summary demonstrates that Good Health Status, Poor Health Status, and the presence of Medicare Insurance were the most significant variables in determining an individual's health care expenditure, however no demographic variables represented any significance in the model.

Model Summary: Predictors (Constant), MedicareInsurance, GoodHealthStatus, UsualPlaceForCare, PoorHealthStatus, PrivateInsurance

Model	R	R Square	Adjusted R-Square	Std. Error of the Estimate
1	.277	.077	.071	25571.488

Coefficients Summary

Variables	Sig.	Statistics VIF
(Constant)	.083	
UsualPlaceForCare	.003	1.048
GoodHealth Status	<.001	1.050
PoorHealthStatus	<.001	1.060
PrivateInsurance	.016	1.119
MedicareInsurance	<.001	1.136

Dependent Variable: DIRECTPAY

<u>Binary Logistic Regressions</u>

Model 2: Binary Logistic Regression of Medicare

Due to the demographics lack of explanatory power with expenditure as the dependent variable, we analyzed demographic and socioeconomic variables relationship with the powerful socioeconomic variables through binary logistic regression.

The model found that two demographic variables; no language other than English spoken at home and age (young, middle, old) were significant in determining if a respondent has Medicare insurance. Additionally, socioeconomic variables; low income, poor health status, and private insurance were also significant in determining the dependent variable. The combination of both demographic and socioeconomic factors accounted for a 90% predictive accuracy.

Medicare Insurance				
		Predicted		**Percentage Correct**
Observed		0	1	
	0	497	10	98.0%
	1	78	251	76.3%
	Overall Percentage			**89.5%**

Variable	Significance
LowIncome	0.020
NoOtherLanguages	0.051
PoorHealthStatus	0.024
PrivateInsurance	<0.001
Young	<0.001
MiddleAge	<0.001

OldAge <0.001

Constant <0.001

We performed this model two more times using good health status and poor health status as the independent variables. These models found two significant independent variables each. Model 3 holds good health status as the dependent variable with private insurance and old age as the independent variables. Model 4 holds poor health status as the dependent variable with private insurance and Medicare insurance as the independent variables.

Model 3: Binary Logistic Regression

Good Health Status				
		Predicted		**Percentage Correct**
Observed		0	1	
	0	23	163	12.4%
	1	14	636	97.8%
	Overall Percentage			**78.8%**

P-Values	
Variable	**Significance**
Private Insurance	<0.001
Old Age	<0.001
Constant	<0.001

Model 4: Binary Logistic Regression

Poor Health Status				
		Predicted		**Percentage Correct**
Observed		0	1	
	0	786	0	100%
	1	50	0	0%
	Overall Percentage			**94%**

P-Values	
Variable	**Significance**
Private Insurance	<0.001
Medicare	0.018
Constant	<0.001

Model 1 demonstrated that Medicare, poor, and moderate health status greatly influence expenditure. Models 2-4 demonstrate that private insurance and age greatly influence the most important factors in determining expenditure.

The binary logistic regression models demonstrate a deeper understanding of where an individual can look to understand how to account for their personal health care expenditure. To illustrate, a young individual who recently graduated from a higher education institution acquires a lot of recurring expenses they have never previously experienced; expenditures such as rent, car payments, cell phone payments, subscriptions, etc. Another expenditure they must consider is healthcare. Through the result of this analysis the individual can weigh the influence of a public health insurance at a younger age to decrease expenditure, but consider shifting to private insurance with older age.

These regression analyses provide throughout insight for individuals to better understand what factors influence their healthcare expenditure.

Model 5: K-Nearest Neighbor Clustering for Private Insurance

In order to analyze how low-income individuals and high-income individuals, compared to their health insurance plans, affect the forecast of their annual total of direct health care payments, we utilized machine learning K-Nearest Neighbor Clustering (K-NN). To perform this analysis, the continuous variables, income, age, and healthcare expenditure were standardized to perform a more accurate analysis in comparison to dummy variables, private health insurance. The K-NN consisted of income, age, and whether or not the individual had private health insurance, classified by healthcare expenditure.

This model determined after 16 iterations two significantly different clusters. Cluster one consisted of young individuals who do not have private health insurance and are low-income. Cluster two consisted of middle-aged individuals that did have private health insurance and higher income. Cluster one captures 83% of the dataset, indicating the majority of the dataset consists of young adults with no private healthcare insurance and low income, leading to higher healthcare expenditures.

This is important because there has been a consistent rise in healthcare costs, especially in the private sector. The number of private health insurance companies has decreased, as markets become more concentrated, premiums increase (U.S. Government Accountability Office 2024).

Number of Cases in each Cluster

Cluster 1	693.0
Cluster 2	143.0
Valid	836.0
Missing	0

Final Cluster Centers

	Cluster 1	Cluster 2
Age (Z-score)	0.05174	-0.25072
Private Insurance	1	2
Income (Z-score)	-0.35761	1.73304

The graph below provides a visual representation of the two clusters, with standardized income as the x-axis and standardized age as the y-axis. The graph's markers are set by the cluster number, and the cases are labeled by standardized healthcare expenditure.

13

Model 6: K-Nearest Neighbor Clustering for Health Status

To investigate the increasing trend in health expenditure over the last decade, considering the previous clustering model that grouped age, income, and private insurance, this model clusters based on age, income, and health status. The health status variable is categorized as follows: 1–Excellent, 2–Very Good, 3–Good, 4–Fair, 5–Poor.

The clustering analysis found two clusters after 22 iterations. Cluster one captures the low end of the health status scale (Very Good-Excellent), high income, and young age. Cluster two captures the high end of the health status scale (Good-Poor), low income, and middle to older aged.

Number of Cases in each Cluster

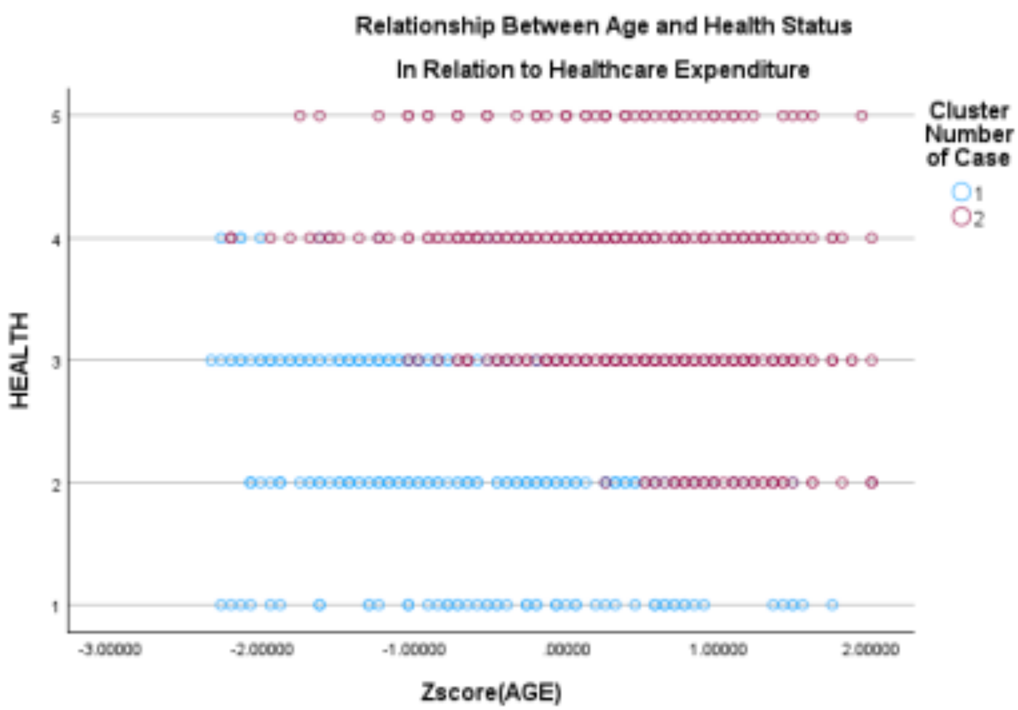
Cluster 1	354.0
Cluster 2	482.0
Valid	836.0
Missing	0

Final Cluster Centers

	Cluster 1	Cluster 2
Health Status	2	3
Income (Z-score)	0.54395	-0.39950
Age (Z-score)	-0.59906	0.43997

These clusters demonstrate that young individuals have a better health status, as observed from the previous cluster analysis, and pay significantly higher healthcare expenditures. Middle-aged and older individuals have a poor health status and lower healthcare expenditures.

The graph below provides a visual representation of these clusters with standardized age as the x-axis and health status as the y-axis. The graph's markers are set by the cluster number, and the cases are labeled by standardized healthcare expenditures.

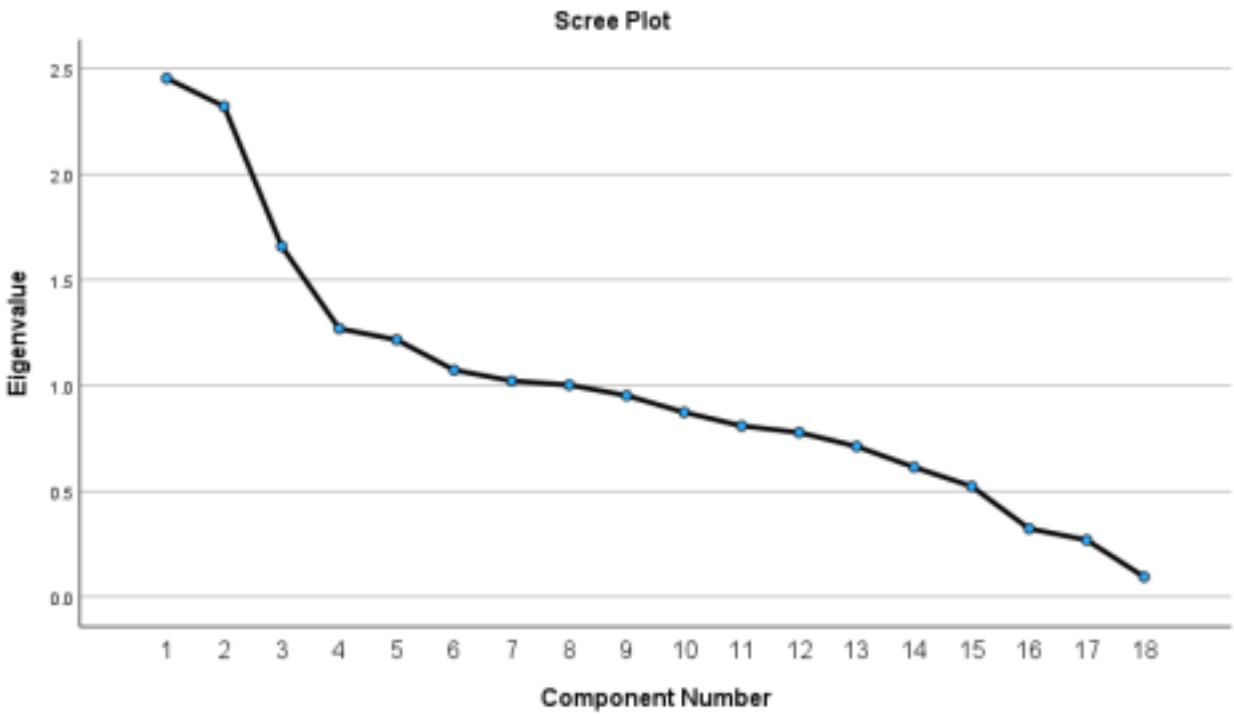


The combination of these two clustering analyses supports the idea that while healthcare expenditures are increasing, especially in the private health insurance sector, the quality of health is not improving. This can be observed because Model 4 identified that young individuals with low income can not afford private insurance and have lower expenditures, while middle to older-aged adults who can afford private insurance have higher expenditures but poor health status.

Model 7: Principal Component Analysis (PCA)

We conducted principal component analysis (PCA) to lower the dimensions of this data set and find a better understanding of the relationships among demographic, socioeconomic, and health-related variables. These findings reveal distinct latent structures that clarify the complex interplay between demographic, socioeconomic, and healthcare access factors.

The scree plot demonstrates a sharp decline in eigenvalues from the first to the third component, indicating that these components explain a substantial proportion of the total variance. An evident "elbow" appears around Component 3, after which the slope levels off. This suggests that additional components contribute relatively limited explanatory power. The pattern supports the retention of the first few principal components as the most meaningful dimensions. Overall, the scree plot confirms a strong underlying structure dominated by a small number of factors.



The rotated component matrix reveals a structured set of latent dimensions underlying the relationships among the variables. The first component represents Education Level, with extremely high loadings, indicating that education is a well-defined and independent structural factor within the dataset. The second component reflects Aging Risk factors, as variables such as age, Medicare insurance, and cancer history are strongly associated, reinforcing their combined influence on healthcare costs.

The third component captures Cultural Identity, suggesting that these factors do not directly influence healthcare spending but instead operate as a separate dimension of individual characteristics. The fourth component represents healthcare access and engagement, combining a usual source of care and provider advice with a negative loading for being uninsured, highlighting the importance of system access for preventive care.

The fifth component reflects a Socioeconomic Status, linking income and private insurance, with gender adding variation. The sixth component captures health behavior and demographics, combining race and exercise patterns, although it is less clearly defined. The seventh component represents general health and public coverage, suggesting a link between government assistance and perceived health status. Finally, the eighth component reflects preventive behavior, with smoking emerging as an independent and distinct factor.

Rotated Component Matrix[a]

	Component							
	1	2	3	4	5	6	7	8
EDU-C	0.939							
HI-DEG	0.934							
AGE		0.850						
HI-MCARE		0.830						
CANCER-EV		0.549						
US-BORN			0.903					
SPK-OTH-LANG			0.903					
USUAL-PL				0.666				
MSDMQ-ADV				0.626				
HI-NOT-COV				-0.547				
HI-PRIVATE					0.776			
INC-TOT					0.668			
SEX					-0.413			
RACE						0.690		
EXER-MOD-VIG-5D						-0.620		
HEALTH						0.453		
HI-OTH-GOV							0.860	
SMOKE-NOW								0.955

Extraction Method: Principal Component Analysis.
a. Rotation converged in 22 iterations.

The PCA results highlight that healthcare-related outcomes are multidimensional, with distinct and relatively independent components capturing education, lifecycle risk, cultural identity, access to care, economic status, and health behaviors.

REGRESSION ANALYSIS

Model Summary

In our regression model using PCA, we were able to explain 10.1% of the total variation in healthcare expenditures, indicating a relatively low explanatory power. Despite the low R-squared, the model is statistically significant (Sig. < .001), suggesting that the combination of demographic, socioeconomic, and lifestyle factors has a non-random and meaningful relationship with healthcare spending. 17

Model	R	R Square	Adjusted R-Square	Std. Error of the Estimate
1	0.317	0.101	0.092	25288.446

Coefficients Summary

Variables	Significance	Statistics VIF
(Constant)	<0.001	1.00
Education Level (Factor 1)	.273	1.00
Aging Risk (Factor 2)	<0.001	1.00
Cultural Identity (Factor 3)	.352	1.00
Healthcare Access (Factor 4)	<0.001	1.00
Socioeconomic Status (Factor 5)	.206	1.00
Demographic Behavior (Factor 6)	<0.001	1.00
Health Status (Factor 7)	.647	1.00
Smoking Behavior (Factor 8)	.109	1.00

Among the factors, Components 2, 4, and 6 are statistically significant predictors ($p < .001$), indicating that lifecycle/health risk, healthcare access, and behavioral/demographic factors play a meaningful role in influencing expenditures. In contrast, other components are not statistically significant, reinforcing that not all latent dimensions identified by PCA directly translate into spending variation.

Importantly, the low R^2 confirms the earlier finding that healthcare expenditures are not easily predictable, even after dimensionality reduction. This suggests that unobserved factors likely play a substantial role.

Overall, the PCA regression improves interpretability by reducing multicollinearity and identifying key underlying dimensions, but it also highlights the inherent complexity and limited predictability of health.

CONCLUSION

This research aimed to determine the primary factors influencing annual direct healthcare costs by examining demographic, socioeconomic, and health-related variables utilizing the 2023 IPUMS MEPS data. The results indicate that demographic characteristics alone do not sufficiently account for healthcare expenditures. Factors such as age, gender, ethnicity, and educational attainment exhibited limited predictive capability, with the demographic regression model accounting for merely 1.6% of the variance in spending. Similarly, although the socioeconomic model revealed statistically significant factors including health status, usual source of care, private insurance, and Medicare coverage, its overall explanatory power was still relatively low at 7.1%. Even when these groups of variables were integrated, the model continued to elucidate only a minor fraction of the variation in healthcare expenditures.

These findings are somewhat consistent with existing literature, yet they also highlight significant discrepancies. Previous studies indicate that demographic and economic variables significantly influence healthcare access, usage, and financial strain. However, in this analysis, these factors did not yield strong predictive capabilities in the regression models. This discrepancy may illustrate the intricate nature of healthcare spending itself. Annual expenditures are likely influenced not only by fixed attributes such as age or income but also by medical occurrences, the severity of chronic conditions, the frequency of healthcare utilization, benefit structures, and other unmeasured elements that are not adequately represented by the chosen variables. In essence, while the literature provides directional support, the current analysis implies that these elements function in a more intricate and less straightforward manner than previously anticipated.

The clustering analyses provided a deeper understanding of these relationships. One model distinguished younger, lower-income folks without private insurance from middle-aged, higher-income individuals who had private insurance, with the first group making up the bulk of the sample. A second clustering model revealed that younger, higher-income individuals generally reported better health, while middle-aged to older, lower-income individuals were more prone to report poorer health. All in all, these results indicate that healthcare spending isn't linked to just one factor at a time, but rather to a mix of age, income, insurance access, and health status that leads to different spending patterns among various groups.

A significant takeaway from this study is that increasing healthcare spending does not always lead to improved health outcomes. The clustering analysis indicates that certain individuals who have better access to private insurance and higher spending still report worse health conditions, whereas younger individuals with limited financial means and less private insurance may incur lower costs but encounter barriers to access. This raises a larger issue within the U.S. healthcare system: elevated expenditures do not inherently result in better health, particularly when expenses are influenced by market concentration, the structure of insurance, and disparities in access to care.

In conclusion, this project shows how valuable it is to mix regression methods with machine learning and exploratory analytics in the field of business analytics research. Even though the regression models had some limitations in their predictive power, the clustering analysis uncovered significant patterns at the population level that clarify the differences in healthcare spending among various groups. For future studies, enhancing this analysis by adding

more detailed information on chronic illnesses, healthcare usage, geographic variations, and plan design features could be beneficial. By broadening the model in these ways, we might achieve a more thorough understanding of healthcare costs and provide better insights for policymakers, insurers, and healthcare administrators who are looking to minimize disparities and enhance cost efficiency.

REFERENCES

Centers for Disease Control and Prevention. (2025). *Health insurance coverage*. National Center for Health Statistics. https://www.cdc.gov/nchs/fastats/health-insurance.htm

U.S. Census Bureau. (2024). *Health insurance coverage in the United States*. https://www.census.gov/library/stories/2024/09/health-insurance-coverage.html

U.S. Government Accountability Office. (2024). Health insurance costs are increasing as markets become more concentrated with fewer insurance companies.

https://www.gao.gov/blog/health-insurance-costs-are-increasing-markets-become-more-conc entrated-fewer-insurance-companies-interactive-map

Sommers, B. D., Gunja, M. Z., Finegold, K., & Musco, T. (2015). Changes in self-reported insurance coverage, access to care, and health under the Affordable Care Act. *JAMA, 314*(4), 366–374. https://doi.org/10.1001/jama.2015.8421

Miller, J. D., et al. (2005). Direct costs of COPD in the U.S.: An analysis of Medical Expenditure Panel Survey (MEPS) data. *COPD: Journal of Chronic Obstructive Pulmonary Disease, 2*(3), 311–318. https://doi.org/10.1080/15412550500218221

Bhat, R., & Jain, N. (2007). A study of factors affecting the renewal of health insurance policy. Indian Institute of Management Ahmedabad.

Artiga, S., & Hinton, E. (2018). Beyond health care: The role of social determinants in promoting health and health equity.

Morrisey, M. A. (2025). *Health insurance* (4th ed.). ACHE Learn.

Warren, D. (2024, February 3). The income you need to fall in America's lower, middle, and upper classes. *Yahoo Finance.* https://finance.yahoo.com/news/income-fall-americas-lower-middle-122100515.html 21

OPTIMIZING HEALTH INSURANCE COSTS THROUGH QUANTITATIVE DATA ANALYSIS FOR A NONPROFIT ORGANIZATION

Ava Geer, Siena University
Alyssa Ramundo, Siena University
Dr. John O'Neill and Dr. Shahryar Gheibi, Siena University – Faculty Mentor

ABSTRACT

Identifying and understanding the underlying factors that drive rising health insurance costs is essential for effective cost management, particularly within the nonprofit sector. Many nonprofit organizations lack the capacity to conduct in-depth analyses of health insurance claims data, which can hinder their ability to make informed decisions. This research, conducted collaboratively within a data analytics team, sought to determine the most significant cost drivers within a comprehensive insurance cost claims dataset through a structured process of data cleaning, data mining, and cost driver assessment.

The analysis began with an extensive data cleaning phase to remove duplicate entries and claims containing missing or inconsistent values. Addressing these inaccuracies was critical, as such discrepancies could distort mean cost estimates, trend analyses, and subsequent interpretations. After establishing a reliable dataset, data mining techniques were applied using both Excel and RStudio to perform exploratory data analysis from multiple perspectives. Data mining served as a foundational step in uncovering hidden patterns, relationships, and anomalies that could suggest potential cost drivers.

To further refine these insights, regression tree modeling—a recursively partitioning statistical method—was employed to identify and visualize the key variables most strongly associated with insurance claim costs. This model provided a clear structure of influential factors, enabling a deeper understanding of how specific variables interact to shape overall cost trends.

The combined approach of data cleaning, exploratory data mining, and recursive partitioning produced data-driven insights highlighting the primary contributors to the nonprofit's health insurance expenditures. By integrating these findings into strategic decision-making, nonprofit organizations can more effectively allocate resources, negotiate insurance plans, and enhance long-term financial sustainability.

INTRODUCTION TO A PRELIMINARY HR AUDIT: COLUMBIA LAND CONSERVANCY

Esther Ma, Siena University
Elizabeth Lankford, Siena University
Professor Deborah Kelly - Faculty Mentor

ABSTRACT

An HR audit is used to review Human Resource activities in a business in order to identify issues related to policies, practices, regulations, compliance, employee issues, and other HR focus areas. Students in the MGMT432 Strategic Human Resource Management capstone course have an opportunity to conduct a preliminary HR Audit of the business of their choice by collecting information on its HR practices and procedures from various sources, including communication with an HR professional within the organization.

For this project, our group is conducting a preliminary HR audit of the Columbia Land Conservancy, a nonprofit organization focused on land conservation and community engagement. The HR Audit presentation slides include background information about the company, outlines of its current HR policies, procedures, and practices, and an analysis of key areas such as recruitment, training, performance management, and employee relations. The project also includes a SWOT analysis of the organization's HR practices, along with a preliminary reflection on potential strategic recommendations.

This presentation will describe the project, the research process, limitations experienced, and key observations related to the organization's HR function.

PEDIATRIC VACCINATION RATES FOR RURAL CHILDREN: A COMPARATIVE ANALYSIS

Mea Amedio, Siena University
Joshua Hunter, Dartmouth College
Dr. Andrea Smith-Hunter, Professor of Management, Siena University – Faculty Mentor
Dr. James Nolan, Professor of Business Analytics, Siena University – Faculty Mentor

ABSTRACT

This study looked at the vaccination rates of two and five year old children in rural areas in specific locales in Upstate New York and compared those rates with those in corresponding and closely located urban and suburban areas. The paper went further by looking within those zip codes to see if any disparities existed across race and across gender using summary data from the New York State Department of Health. The results showed that overall the five year old children were more likely to be vaccinated when compared to the two year old children. It further showed that children in urban and suburban areas were more likely to be vaccinated versus their rural counterparts, regardless of age, gender or race. Finally, the results showed that minority children in both age groups, regardless of zip codes were less likely to be vaccinated versus their White counterparts and that males were slightly less likely to be vaccinated regardless of locale versus their female counterparts. It is advised that various community organizations work with health authorities and work with parents and care givers to provide access and information on vaccinations for children.

Key Words: Vaccination, Vaccination Rates, Children, MMR, Poliovirus, DTap, HepB, HepA, Rotavirus, Influenza, Combined 7-Vaccine Series, Rural, Urban, Suburban

PROFESSIONALS’ PERCEPTIONS OF RISKS IN AUDITING CRYPTO CURRENCY

Caden Kramer, Siena University
Dr. Marie Rice, Siena University – Faculty Mentor

ABSTRACT

Corporations are continuing to expand their use of crypto assets. This leaves auditors exposed to the unique risks and challenges of the crypto-assets landscape. Recent changes to U.S. GAAP, including FASB’s shift to fair-value measurement, have improved financial reporting but have not resolved persistent challenges related to valuation, ownership, existence, and completeness. At the same time PCAOB inspection findings highlight audit deficiencies in engagement involving crypto assets. This signals the gap between regulatory expectations and audit practice. Despite growing guidance, limited research captures how practicing auditors perceive these risks working on real engagements. This study aims to address the gap by surveying audit professionals with direct experience auditing crypto assets. This study will examine perceived challenges, the clarity and effectiveness of future guidance, and the expectations for future audit work in rapidly developing crypto asset landscape. Insights from this research contribute to ongoing discussions surrounding valuation risk, audit quality, evolving standards, and other crypto-asset related risks.

Keywords: crypto assets, auditing, digital assets, audit risk, valuation risk, auditor perceptions

INTRODUCTION

According to an estimate in early 2024, more than 6,000 companies accept bitcoin as a means of payment (Deloitte 2025). Another survey of senior executives showed that “merchants are embracing digital currency payments with the hope of gaining a competitive advantage in the marketplace” (Deloitte 2025). Corporations are increasingly adopting digital assets. Investments in digital assets require corporations and accounting firms to address accounting, reporting, regulatory, and control risks (Deloitte, 2025). Digital assets represent a new frontier for many corporations, and evolving rules and standards require regulators, corporations, accounting firms and standard setters to listen to one another to gain an understanding of operational and technical difficulties to tackle the challenges together (Deloitte 2025).

As crypto assets become more prominent on corporation’s balance sheet and accounting standards continue to change, concerns about audit deficiencies regarding crypto assets have been raised. It is important to consider the challenges auditor face when working with these types of assets. Current research focuses on technical issues or regulatory commentary rather than working auditors’ perceptions of these risks. To address this gap this study conducts a survey of audit professionals who have worked on an engagement involving crypto assets to answer the question:

What do auditors perceive as the main challenges when working with crypto assets?

BACKGROUND AND LITERATURE REVIEW

<u>Standard Setting Bodies</u>

Public Company Accounting Oversight Board

In 2023, the PCAOB highlighted inspection observations relating to public company audits involving crypto assets. In this spotlight, the PCAOB noted five common audit deficiencies relating to crypto assets. These deficiencies included fraud and significant unusual transactions, ownership of crypto assets, relevance and reliability of information used as audit evidence revenue recognition in crypto asset transfers and arrangements with mining pool operators (PCAOB 2023). During the inspection, the PCAOB observed good practices they believe can enhance audit quality. These good practices included utilizing consultations, subject matter specialists, and technology-based tools.

Financial Accounting Standards Board

In December 2023, FASB issued an update of accounting standards regarding crypto assets. This update included changes to how corporations measure, present, and disclose crypto assets. The most significant change in this update was on measurement, "An entity shall measure crypto assets at fair value in the statement of financial position and gains or losses from remeasurement of crypto assets should be included in net income" (FASB 2023). This differed from the previous cost-less-impairment method. FASB decided to make these changes due to stakeholder feedback. These stakeholders felt specifically accounting for only the decreases, but not increases, in the value of crypto assets in the financial statements until these assets were sold affects the underlying economics of those assets as well as an entity's financial position (FASB 2023). These changes went into effect for all entities for fiscal years beginning after December 15, 2024. This includes interim periods within those fiscal years.

Today, these changes are reflected in the current accounting standards codification. According to ASC 350-60-15-1, crypto assets are a subset of digital assets that meet all the following criteria,

a. Meet the definition of intangible assets as defined in the Codification

b. Do not provide the asset holder with enforceable rights to or claims on underlying goods, services, or other assets

c. Are created or reside on a distributed ledger based on blockchain or similar technology

d. Are secured through cryptography

e. Are fungible

f. Are not created or issued by the reporting entity or its related parties

Crypto assets therefore do not include other types of digital assets like non-fungible-tokens, or asset backed tokens. If the digital asset meets the criteria of a digital asset according to ASC 350-60-35-1, "An entity shall measure crypto assets at fair value in the statement of financial position. Gains and losses from the remeasurement of crypto assets shall be included in net income".

While this update shows improvement aligning accounting standards to the nature of crypto assets, it does not eliminate the broader challenges faced by auditors. Risks remain for auditors in assessing the valuation of these assets, proving ownership and existence as well as completeness concerns.

Stakeholders Perceptions

Research shows that accounting for crypto assets should depend on multiple factors. These factors should include, but are not limited to, the intention of the holder, business model, and arrangements behind the transactions, and the nature of the transaction (Chou et al. 2022). Further, this research found that stakeholders agreed that crypto should not be accounted for as cash and cash equivalents due to their unregulated and speculative nature (Chou et al. 2022). Stakeholders also agreed that measuring crypto at fair value provides "relevant and useful information" (Chou et al. 2022). These preferences are reflected in the current FASB codification discussed in the introduction. Stakeholders also mentioned the importance of full disclosure of the risks associated with these assets (Chou et al. 2022). Because of the quickly evolving nature of crypto assets stakeholders also noted that it is important that standard setters continuously monitor the issue (Chou et al. 2022).

Standard setters have attempted to keep pace with evolutions and evolving challenges in the crypto assets landscape. Standard setters have listened to stakeholders, but there is little research specifically looking at auditors' perspectives. This led to the research question, "to what extent is current guidance perceived by auditors to be clear and effective?

Issues, Risks, and Challenges for Auditing Crypto Assets

Crypto Asset Transactions

Sheng-Feng Hsieh and Gerard Brennan (2022) reviewed non-authoritative literature guidance from professional organizations to summarize issues that should be considered when auditing crypto asset transactions. Hsieh and Brennan (2022) cited a publication issued by the PCAOB that listed responsibilities under the existing standards that auditors should comply with when their clients had material crypto asset holdings. This includes a firm level quality control system, planning and risk assessment (Hsieh and Brennan 2022). The PCAOB listed some essential factors that auditors may pay attention to, this included, custody of crypto assets, how many types are traded or held and the number of customers and suppliers of the client accepting and trading crypto. Auditors should also pay attention to the types of crypto assets being traded or held, and the number of customers or suppliers. Further the auditors should understand the client's primary business model and strategies. The PCAOB also noted that the auditor should look at how the clients comply with know-your customer and anti-money-laundering laws and identifying related party transactions.

Crypto Asset Audit Risks

Lazea et al (2025) explored the inherit, control, valuation and risks associated with crypto assets, "auditing crypto assets is challenging due to unique risks, control issues, valuation complexities, and rapidly advancing technology" (Lazea (Trifa) et al. 2025, 204). Some of the unique risks that come with crypto assets is high price volatility. Frequent price swings can complicate getting an accurate valuation of the asset (Lazea (Trifa) et al. 2025). Further, blockchain is inherently vulnerable to manipulation by a majority holder. This manipulation can lead to significant financial losses and fraudulent transactions. Those who engage in the crypto asset markets are often risk-tolerant, which can lower auditor confidence (Lazea (Trifa) et al. 2025).

Broby and Paul give some reasons that the core audit assertions of existence, ownership and rights and obligations and completeness are at risk when working with crypto assets. First, existence and ownership are at risk because having a key does not necessarily establish ownership, "the private keys used to access a wallet may simply be transferred between parties … This makes it difficult to determine the identity of the party operating an address" (Broby and Paul 2017, 83). This shows that having a key does not mean being the legal owner of the asset. Therefore, the auditor cannot rely on key possessions alone for existence or ownership of the asset. Completeness is at risk because "funds within online exchanges and wallets are often interchanged between accounts without any blockchain-based audit trail" (Broby and Paul 2017). Because not all crypto transactions appear on the blockchain, the ledger is incomplete, and it becomes difficult for the auditor to verify the true value of the assets in the exchange.

Challenges of Auditing Blockchain-Based Assets

While the market for crypto assets is growing, it is primarily composed of start-ups that lack the "financial sophistication and maturity of similarly valued traditional firms, and that rely on outside funding to develop and grow" (Pimentel et al. 2021). For these start-ups to generate traditional forms of credit like bank loans or going public, they will need to have audited financial statements. Public accounting firms have business risk concerns associated with these clients (Pimentel et al. 2021). A challenge for any firm planning to take on crypto asset work is demonstrating sufficient competence to address the relevant risks (Pimentel et al. 2021). The authors of this study found that the three areas that generate the most challenge include existence, ownership and valuation of crypto assets (Pimentel et al. 2021). While crypto assets are complex, auditing them is not impossible. Auditors typically proceed with caution when working with crypto assets, but audit risk can be reduced by properly utilizing experts and properly vetting potential clients (Pimentel et al. 2021).

Existing research identifies issues and challenges with auditing crypto assets but rarely looks at auditor's perceptions of these challenges. This led to the research question, "What do auditors perceive as the main challenges when working with crypto assets?"

RESEARCH DESIGN AND METHODOLOGY

This study will quantitatively research design through an online survey to gather auditors' perceptions of the challenges associated with auditing crypto assets and the clarity of current guidance. A structured close-end survey was used to ensure consistency across responses and to allow comparison across participants. Substantially all the survey questions were designed using a 7-point Likert scale.

Participants are required to have worked or are currently working on an engagement with material crypto asset holdings. To ensure participants meet this criterion, a screening question at the beginning of the survey asked participants if they have worked on an engagement involving crypto assets. Participants were recruited through professional accounting networks, LinkedIn, alumni connections and referrals. Because auditors with experience relevant to this survey are a small and specialized group, the anticipated sample size was approximately 30 participants.

The survey was created and administered using Qualtrics. Most questions were close ended to support a quantitative analysis of the response. The instrument included four sections, demographics, perceived challenges, perceptions about guidance clarity and effectiveness and questions about the future of crypto asset engagements.

Participants receive a link to the survey via email. The survey began with an informed consent stating the purpose of the study, voluntary participation, anonymity and the option to withdrawal at any time. The survey required about 10-15 minutes to be completed.

Approval from the institutional review board (IRB) was obtained prior to data collection. Participation was voluntary, and no personally identifying information was collected. All data was stored securely and reported in aggregate to protect confidentiality.

CONTRIBUTIONS AND LIMITATIONS

This study contributes to the growing body of research that examines the risks and rewards associated with cryptocurrencies, as well as the literature on valuation, audit quality, and auditors' perceptions of audit risks. While current studies have examined the risks of cryptocurrency from various perspectives, only one that we are aware of (Chou, et al., 2022) includes professionals' perceptions of the topic. Chou et al., (2022) employed interviews of academics, standard-setters, and accounting practitioners. This study extends Chou, et al. (2022) by eliciting audit and valuation professionals; perceptions about the risk of cryptocurrency using quantitative survey methodology.

Like all studies employing survey data, this study suffers from limitations. Specifically, the results of this study may only be used to ascertain correlations within the data obtained, but causality may not be inferred from the data collected. In addition, this study seeks to solicit perceptions from working professionals who are notoriously busy and yield low response rates for survey data. It is possible that sufficient power may not be obtained to yield reliable results. Future studies may examine the risks of auditing cryptocurrency and its valuation using other methods.

REFERENCES

Akter, M., Kummer, T.-F., & Yigitbasioglu, O. (2024). Looking beyond the hype: The challenges of blockchain adoption in accounting. *International Journal of Accounting Information Systems*, 53, 100681. https://doi.org/10.1016/j.accinf.2024.100681

Botte, A., & Nigro, M. (2021). Risk analysis of crypto assets. *TwoSigma, July*.

Broby, D., & Paul, G. (2017). The financial auditing of distributed ledgers, blockchain and cryptocurrencies. *Journal of financial transformation*, *46*, 76-88.

Chou, J. H., Agrawal, P., & Birt, J. (2022). Accounting for crypto-assets: Stakeholders' perceptions. *Studies in Economics and Finance*, 39(3), 471–489. https://doi.org/10.1108/SEF-10-2021-0469

Cryptocurrency Benefits for Corporations | Deloitte US. (2025). Deloitte. https://www.deloitte.com/us/en/services/audit-assurance/articles/corporates-using-crypto.html

Cryptocurrency Regulation: A Guide to U.S. & Global Policies | Britannica Money. (n.d.). Retrieved November 12, 2025, from https://www.britannica.com/money/cryptocurrency-regulation

Deloitte. (2025, March 18). The Rise of Digital Assets: Considerations for Corporate Balance Sheets—WSJ. https://deloitte.wsj.com/riskandcompliance/the-rise-of-digital-assets-considerations-for-corporate-balance-sheets-c90586ac?mod=Deloitte_riskcompliance_wsjsf_title&_gl=1*1puiq0*_gcl_au*MTM1NDAwMDQwNi4xNzQxNjQ3NDYx*_ga*OTU3MjI0NTMzLjE3NDE2NDc0NTg.*_ga_K2H7B9JRSS*MTc0MjQ4MjkyOC40LjEuMTc0MjQ4MzE3Ni4yMS4wLjA

Dunn, R. T., Jenkins, J. G., & Sheldon, M. D. (2021). Bitcoin and blockchain: Audit implications of the killer Bs. *Issues in Accounting Education*, *36*(1), 43-56.

Dupuis, D., Smith, D., Gleason, K., & Kannan, Y. (2023). Bitcoin and beyond: Crypto asset considerations for auditors/forensic accountants. *Journal of Forensic and Investigative Accounting*, *15*(3), 1-29.

Dyball, M. C., & Seethamraju, R. (2021). The impact of client use of blockchain technology on audit risk and audit approach—An exploratory study. *International Journal of Auditing*, *25*(2), 602-615.

Eutsler, J., & Lang, B. (2015). Rating scales in accounting research: The impact of scale points and labels. *Behavioral Research in Accounting*, *27*(2), 35-51.

FASB. (2023, December 8). Intangibles—Goodwill and Other—Crypto Assets (Subtopic 350-60). FASB.

https://fasb.org/page/ShowPdf?path=ASU%202023-08.pdf&title=ACCOUNTING%20STANDARDS%20UPDATE%202023-08%E2%80%94Intangibles%E2%80%94Goodwill%20and%20Other%E2%80%94Crypto%20Assets%20%28Subtopic%20350-60%29:

Georgiou, I., Sapuric, S., Lois, P., & Thrassou, A. (2024). Blockchain for Accounting and Auditing—Accounting and Auditing for Cryptocurrencies: A Systematic Literature Review and Future Research Directions. *Journal of Risk and Financial Management*, 17(7), 276. https://doi.org/10.3390/jrfm17070276

Garnett, A. Grace (2025, April 16). Cryptocurrency regulation: Rules are in development. Encyclopedia Britannica. https://www.britannica.com/money/cryptocurrency-regulation

Harrast, S. A., McGilsky, D., & Sun, Y. T. (2022). Determining the inherent risks of cryptocurrency: A survey analysis. *Current Issues in Auditing*, *16*(2), A10-A17.

Hsieh, S. F., & Brennan, G. (2022). Issues, risks, and challenges for auditing crypto asset transactions. *International Journal of Accounting Information Systems*, *46*, 100569.

Huang, F., Wang, T., & Yen, J. C. (2024). Opportunities or challenges? Audit risk and blockchain disclosures in 10-K filings. *Auditing: A Journal of Practice & Theory*, *43*(2), 131-158.

Juma'h, A. H., & Li, Y. (2023). The effects of auditors' knowledge, professional skepticism, and perceived adequacy of accounting standards on their intention to use blockchain. *International Journal of Accounting Information Systems*, 51, 100650. https://doi.org/10.1016/j.accinf.2023.100650

Lazea (Trifa), G. – I., Balan, A. D., Solovastru, M. S., & Bucharest Stock Exchange. (2025). Navigating Auditing Risks in the Crypto Asset Landscape. *Audit Financiar*, 23(177), 197–209. https://doi.org/10.20869/AUDITF/2025/177/006

Malsch, B., & Salterio, S. E. (2016). "Doing Good Field Research": Assessing the Quality of Audit Field Research. *AUDITING: A Journal of Practice & Theory,* 35(1), 1–22. https://doi.org/10.2308/ajpt-51170

Morozova, T., Akhmadeev, R., Lehoux, L., Yumashev, A. V., Meshkova, G. V., & Lukiyanova, M. (2020). Crypto asset assessment models in financial reporting content typologies. *Entrepreneurship and Sustainability Issues*, *7*(3), 2196.

PCAOB. (n.d.). 350-60 Crypto Assets. https://asc.fasb.org/350/60/showallinonepage

PCAOB. (2023, June). Inspection Observations Related to Public Company Audit Involving Crypto Assets. PCAOB. https://assets.pcaobus.org/pcaob-dev/docs/default-source/documents/crypto-assets-spotlight.pdf?sfvrsn=f0e1d51c_4

Pimentel, E., Boulianne, E., Eskandari, S., & Clark, J. (2021). Systemizing the Challenges of Auditing Blockchain-Based Assets. *Journal of Information Systems*, 35(2), 61–75. https://doi.org/10.2308/ISYS-19-007

Smith, S. S. (2020). Blockchains impact on risk assessment procedures. *Journal of Forensic and Investigative Accounting*, 12(1), 55-65.

Tan, B. S., & Low, K. Y. (2019). Blockchain as the database engine in the accounting system. *Australian Accounting Review*, 29(2), 312-318.

Weigand, H., Blums, I., & de Kruijff, J. (2020). Shared ledger accounting—implementing the economic exchange pattern. *Information Systems*, 90, 101437.

SACRED PAWS CO. - EXECUTIVE SUMMARY

"Let us honor the love that walks beside us."

Madison Bray, Evan Parker, Annabella Rivero, Megan Ferreira
Professor Natasha Adams - ENTR410.04
Siena University

CONTACT INFORMATION

Madison Bray
mp20bray@siena.edu
Evan Parker
e18park@siena.edu
Annabella Rivero
am12rive@siena.edu
Megan Ferreira
me24ferr@siena.edu

ABSTRACT

Sacred Paws Co. creates handcrafted, spiritually inclusive cards that celebrate the full journey of pet companionship from adoption and birthdays to blessings, wellness, and remembrance. The company addresses a clear gap in the pet specialty consumer goods market, where families and veterinary clinics lack high-quality, emotionally resonant keepsakes. With strong early unit economics, scalable production, and a growing demand for personalized pet-life products, Sacred Paws Co. is positioned for sustainable growth through direct-to-consumer sales and clinic partnerships. The venture combines creative design, operational discipline, and a deep understanding of the human–animal bond.

COMPANY SNAPSHOT

Industry: Pet Products

Stage: Startup

Founded: 2026

Employees: 4

Funding Opportunity: $15,000

Use of Funds:

Inventory & Materials: $5,000

Equipment Upgrades: $4,000

Marketing & Branding: $3,000

Operational Runway: $3,000

Monthly Burn Rate: ~$800/month

Avg. Year 1 Monthly Revenue: ~$1,980/month

OUR WHY

Sacred Paws Co. was created because each member of the team are lifelong pet owners who understand how deeply animals shape our lives. Each of us have felt the excitement of bringing home a new companion, the comfort of everyday routines, and the heartbreak of saying goodbye. Those experiences inspired us to design handcrafted prayer cards that honor every part of the journey. Our goal is to celebrate the entire pet journey from joyful beginnings to hard goodbyes, something that reflects the love we all share for the animals who walk through life with us. Families want meaningful ways to honor their pets through adoption, birthdays, blessings, health and recovery and remembrance. Sacred Paws Co. provides handcrafted, emotionally resonant cards that help people mark every moment of love, connection and companionship.

PROBLEM/OPPORTUNITY

Most veterinary clinics and pet families rely on generic, mass-produced cards that fail to capture the emotional depth of the human-animal bond. The market lacks high-quality, spiritually inclusive, customizable keepsakes for both celebration and remembrance. With over 90 million United States pet households and rising demand for personalized pet products, the opportunity is significant and underserved.

SOLUTION/PRODUCT

Sacred Paws Co. offers handcrafted 5-pack card sets in four sizes, featuring adoption messages, birthday greetings, blessings, interfaith prayers, health and wellness wishes and remembrance phrases. Each design is crafted with premium materials and a warm, spiritual tone. Products are available for both direct purchasing and clinic bulk ordering.

MARKET/IMPACT

The U.S. pet market includes over 90 million pet-owning households, with spending on personalized and emotionally meaningful pet products increasing each year. Families seek ways to celebrate milestones such as adoption days, birthdays, and recovery, while also wanting more heartfelt options for remembrance. Veterinary clinics and shelters lack premium, customizable keepsakes to offer their clients, creating a strong entry point for wholesale partnerships. Sacred Paws Co. meets this demand with handcrafted, spiritually inclusive products that elevate the emotional experience of pet ownership and strengthen clinic–client relationships.

REVENUE MODEL

Revenue is generated through four product lines including standard size, modern size, index size and postcard size. The standard size is 2.5’’ x 4.25’’ priced at $18.00, the modern size is 2.75’’ x 4.75’’ priced at $20.00, the index size is 3''x 5'' priced at $22.00 and the postcard size is 4'' x 6'' priced at $25.00. Growth is driven by direct-to-consumer online sales through the company's website, veterinary clinic partnerships, repeat purchases for celebrations, milestone and remembrance and expansion into blessing cards, seasonal collections and wellness-focused offerings. Projected revenue grows from $23,760 in Year 1 to $108,900 by Year 5.

COMPETITION

Competitors include Etsy sellers, mass-market card companies and generic sympathy suppliers. Sacred Paws Co. differentiates through handcrafted, premium materials, emotional depth and spiritual inclusivity, multiple categories beyond loss (adoption, birthdays, blessings, wellness), clinic-friendly wholesale options and consistent, elevated design aesthetic.

EXECUTION PLAN/GO TO MARKET STRATEGY

The execution plan is to launch a direct-to-consumer website with customizable options, build partnerships with veterinary clinics, shelters and pet providers to not only sell our product but to also market our company through their business as well. We will also offer sample kits and wholesale pricing to clinics, use targeted social media advertising and grief support/celebration communities and expand product lines into seasonal and wellness-focused collections. Finally, we will scale production through workflow optimization and equipment upgrades.

FINANCIALS

Sacred Paws Co.'s financial outlook demonstrates a strong early startup supported by low material costs and scalable production. In Year 1, the company is projected to generate $23,760 in revenue, with $12,045 in cost of goods sold, resulting in a gross profit of $11,715 and a net income of $7,715. As brand awareness grows and veterinary clinic partnerships expand, revenue is expected to increase steadily, reaching $108,900 by Year 5. The model reflects healthy margins, predictable unit economics, and a clear path to sustainable growth as the product line broadens across celebration, wellness, and remembrance categories.

TEAM & RELEVANT EXPERIENCE

The Sacred Paws Co. team brings a balanced blend of creative, operational, and strategic experience. Madison Bray, Co-CEO, was born into a business-focused environment and has four years of formal business education with concentrations in project management and entrepreneurship, providing a strong foundation in leadership, organization, and early-stage venture development. Evan Parker also brings four years of business education with concentrations in project management and entrepreneurship, contributing strengths in financial modeling, forecasting, and operational efficiency. Co-CEO Annabella Rivero leads creative direction and product design, shaping the emotional and aesthetic identity of the brand, as well as a background in strategic Human Resource Management to ensure employee satisfaction among the group. Megan Ferreira drives marketing and partnership development, leveraging social media strategy and clinic outreach to expand brand awareness and distribution. Together, the team combines creativity, operational discipline, and a deep understanding of the emotional needs of pet-loving families.

TEMPO

"Fitting health & wellness into the rhythm of your day."

Team Contact Information
Josh Kotlow
Siena University
je04kotl@siena.edu

Evan Williams
Siena University
ef14will@siena.edu

Matt Witkop
Siena University
mr09witk@siena.edu

Nolan Towers
Siena University
nm13towe@siena.edu

Lucas Centanni
Siena University
ll10cent@siena.edu

Joe Gilbert
Siena University
jh06gilb@siena.edu

Faculty Mentor: Professor Natasha Adams
Siena University
nadams@siena.edu

ABSTRACT

Tempo is an AI-driven fitness application to simplify and personalize the workout experience for busy individuals. The app combines customized workout plans as well as personalized meal plans and recommendations, all in one application. By applying AI technology as well as well-picked fitness content, Tempo delivers a convenient yet efficient solution, helping users stay consistent and healthy.

INDUSTRY/CONTEXT

Tempo operates in a rapidly growing industry of digital health & fitness, as well as the health technology industry. As awareness of health and wellness has increased, this has created a position for apps such as Tempo to succeed. Advancements in AI technology have also created room for AI-based applications that are more tailored to user-friendly experiences. This makes it an ideal time for Tempo to enter the industry with a solution that focuses on personalization and convenience.

PROBLEM/OPPORTUNITY

Conventional fitness apps lack a personal user interface and engaging experience. Most of the apps use very basic workout plans that do not align with the consumers' goals, but are more of a broad, basic male/female workout plan. They do not show you how to use proper form or movement control, which can lead to injury or not be an effective workout. There are many apps for workouts, meal plans, and videos, but not all in the same one, streamlining the experience. Over time, usually within a week, users lose motivation because using these apps becomes a chore, because they don't see results. After all, there is a need for something better, a more complete solution. That's where we come in, Tempo.

COMPETITION

Tempo operates in a highly competitive fitness market. Major apps such as MyFitnessPal, Peloton, and Apple Fitness+ are serious threats because they already have large and well-established user bases. MyFitnessPal is an AI-powered health and nutrition tracker specifically designed to change food and fitness habits and track them all in one place. They charge $79.99/yr after a one-month free trial. Peloton offers a variety of services focusing on connected fitness, but its fitness app delivers highly personalized workout plans designed for users to lock in and level up their workouts. Peloton prices are $129/yr and $289/yr for the respective tiers. Apple Fitness+ offers custom plans personalized to users' needs, a variety of 12 different workout types, and real-time metrics. Apple Fitness+ offers multiple deals for bundling services or free trials for new devices, but a standard membership for the service is $79.99/yr.

SOLUTION/PRODUCT OR RESEARCH FOCUS

Tempo provides a comprehensive, AI-driven fitness solution that simplifies how users approach their health and wellness. The app generates personalized workout plans by curating high-quality fitness videos from trusted online sources and organizing them into customized workout “albums” based on each user’s goals, fitness level, and preferences. This allows users to follow structured routines without spending time searching for content. In addition to workouts, Tempo offers customized meal plans and nutritional recommendations tailored to individual needs, creating a complete, all-in-one fitness platform. By acting as a “middleman,” Tempo avoids the high costs of producing original content while still delivering a high-quality, personalized experience. The platform continuously adapts using AI, updating recommendations based on user progress and feedback. This focus on personalization, convenience, and cost-efficiency differentiates Tempo from competitors and provides users with a streamlined and effective path toward achieving their fitness goals.

MARKET/IMPACT

Tempo targets health-conscious individuals looking to improve their health and nutrition, all while being time-friendly. More particularly for students and young professionals who need convenient and personalized fitness solutions. As digital fitness platforms continue to grow, Tempo is positioned to find users who are not content with generic workout plans and uncoordinated fitness tools. By offering an all-in-one platform combining curated workouts and personalized meal plans, Tempo improves users' experience by saving time, creating more consistency, and achieving better results. The AI-driven personalization builds a better experience for the user and will make them more committed to their goals. The impact of Tempo is more than just convenience; it promotes a healthier lifestyle by making fitness and nutrition advice more accessible and organized. Tempo looks to build a strong, loyal user base and become a competitive company in the digital fitness market.

FINANCIALS OR IMPLICATIONS

Revenue for TEMPO is created through a subscription model that functions every month priced at $40/month with a 30-day free trial for new users, meaning new users only contribute to 11 months of revenue in their first year while returning users contribute to the full 12 months, with user growth monitored using a cohort approach that separates new and returning users while incorporating a monthly 5% churn rate, about 54% annual retention, which is reflective of industry standards. User growth is projected as Year 1: 2000, Year 2: 4000, Year 3: 7500, Year 4: 12000, Year 5: 18000, showing an early startup growth stage with a gradual decline in growth rate (100% to 50%) consistent with industry standards. Revenue is calculated using Paid Months = (returning users x 12) + (new users x 11) and Revenue = Paid Months x $40. Cost of goods sold (COGS) includes all direct costs needed to deliver the service including payment processing, AI usage, cloud infrastructure and reserves, calculated based on active months meaning users still incur costs during the free trial period, with payment processing at 2.9% + $0.30 per transaction (about $1.46 per user per month), AI costs decline from $4.00 per user per month in Year 1 to $3.00 by Year 5, cloud infrastructure costs decline from $2.50 to $1.75, and a refund reserve of 0.5% of revenue. Operating expenses include 1 software developer at $100,000, 1 support staff at $45,000, 20% additional burden, marketing set at 35% of revenue, legal & accounting fixed at $10,000 annually, and office costs at $0 in Year 1 and 2 then $10,000 annually moving forward, with margins showing strong unit economics due to a scalable subscription model with Year 1 gross margin ~78% and Year 5 projected ~82%, while key considerations include free trial impact compressing early margins, churn effects impacting long-term revenue, and scalability from declining per user costs improving profitability over time.

TEAM/AUTHORS

Our team is made up of six dedicated young professionals with diverse experience ranging from entrepreneurship, marketing, Operations, finance, and technology. Josh Kotlow, the founder, created the concept and continues to help run the business smoothly. Evan Williams oversees operations on a day-to-day basis, ensuring efficiency and keeping consistency between multiple aspects of the company. Matt Witkop is the lead of Human Resources, focusing on employee training and development while keeping a positive work environment. Nolan Towers drives our marketing team, using different strategies to help raise brand awareness and reach new customers. Lucas Centanni manages the finances for the company, helping keep strong financial outcomes and long-term profits. Joe Gilbert is the lead of the technology division, helping develop and maintain our app. Together, our team is able to combine different skills to support the company's success and growth long-term.

THE WILLINGNESS TO PURCHASE CANNABIS/ THC

Corrinne Blond, Siena University
Ethan Hunt, Siena University
Ethan Linton, Siena University
Hailey Westbrook, Siena University
Dr. Cheryl Buff, Siena University – Faculty Mentor

ABSTRACT

This study explores and provides a deeper understanding of a consumer's willingness to purchase Cannabis/ THC products amid the growing legalization and normalization. A self-administered survey, distributed through various social media platforms using snowball and convenience sampling, yielded 147 responses, with 128 valid responses included in our analysis. Guided by our research, we developed four hypotheses that were tested: effects of THC level, perceived health benefits, dependency, and age on willingness to purchase. Quantitative methods, including linear regression, independent t-tests, ANOVA, and multiple regression, revealed that higher THC levels, greater perceived health benefits, and a stronger dependency significantly increase the willingness to purchase; however, age showed no statistical significance. The scale used, Willingness to Purchase, demonstrated excellent reliability with a Cronbach's alpha score of 0.975. Based on our research results, marketers should make a concerted effort to communicate the potency of their products, highlight health-related benefits, and target consumers who exhibit higher dependency rather than focusing on demographic ages. Although limitations were faced by sample diversity and size, this research provides meaningful insights into consumer behavior and offers practical implications for cannabis marketers.

INTRODUCTION

Consumption of cannabis/ THC products has been a growing trend since the 1970s, gaining increasing popularity in recent years. Not until recently has cannabis been legalized at the state level; however, it is still illegal at the federal level. This shift rapidly created a booming industry filled with questions and uncertainties regarding regulations and consumer perceptions. As cannabis products became more accessible, understanding the factors that influence consumer willingness to purchase is essential for marketers seeking to capitalize on a new, growing market.

Consumer attitudes toward cannabis are shaped by a combination of product attributes, the perceived health benefits, societal norms, and individual usage behavior. Prior research suggests that the intrinsic positive factors, such as THC potency, perceived effectiveness, and level of dependency, may play a stronger role in purchase behavior than demographic characteristics alone (Zhu et al., 2021). At the same time, lingering stigma, health concerns, and generational differences continue to drive a wedge into the industry, influencing how consumers evaluate cannabis-related products. These dynamics make the cannabis industry intrigue consumer behavior studies, identifying the decision-making of consumers.

The purpose of this study is to examine the major factors that influence an individual's willingness to purchase cannabis/THC products. Using primary data collected through a structured survey, we study the relationship between THC levels, perceived health benefits, dependency, age, and willingness to purchase. By identifying which factors significantly affect consumer behavior, this study becomes a resource for insights for cannabis businesses in terms of best marketing practices to maximize sales and company growth.

LITERATURE REVIEW

The legalization and normalization of cannabis have prompted extensive research on consumer behavior, health perceptions, and market trends surrounding THC products, more specifically, edibles. Existing literature indicates that cannabis purchasing behavior is influenced by attributes, neurological and physiological effects of THC, perceived health risks, demographic shifts, and cultural normalization. Collectively, these studies support the hypothesis that prior experience with THC increases consumers' willingness to purchase THC-infused products, especially when the products are perceived as regulated, familiar, and aligned with wellness values (Zhu et al., 2021; Grewal et al., 2020; Reboussin et al., 2019).

Research on cannabis purchasing decisions consistently emphasizes the dominance of intrinsic attributes. (Zhu et al.,2021) found that "quality, strain type, price, and regulated THC levels were significantly more influential than extrinsic factors such as packaging." A systematic review by Donnan et al. (2022) similarly concluded that "potency, quality, consistency, and labeling are central drivers of consumer choice." However, secondary packaging and marketing cues remain influential. Experimental studies have shown that medical or nature-themed packaging, health claims, and warning styles can reduce perceived harm and increase the appeal of products, particularly for edibles (Kowitt et al., 2022).

Neuroscientific research provides insight into repeated cannabis consumption. (Moreno-Rius, 2019) "demonstrated that THC affects cerebellar systems involved in reward processing and emotional regulation, suggesting that repeated exposure may reinforce habitual use and continued purchasing behavior". This mechanism helps explain why prior THC experience may increase consumer openness to new THC products, including food-based products.

Consumption trends reveal a shift away from smoking toward edibles, driven largely by perceived health benefits. Longitudinal and survey research show increasing edible use alongside declining smoking rates, despite concerns about delayed onset, overconsumption, and accidental exposure (Reboussin et al., 2019; Grewal et al., 2020). Clear THC per-serving labeling has been shown to improve customer understanding and confidence in edible products (Goodman et al., 2022). Media and public health sources further note that edibles are often perceived as healthier alternatives to smoking, though evidence remains limited (Blum, 2024).

Broader social and demographic factors also shape consumer behavior. Younger adults and women are increasingly supportive of cannabis use and are more likely to consume edibles, beverages, and wellness-oriented products, often citing sleep, anxiety, and self-care motivations (McEvoy, 2025; MMW News, 2025; Schaeffer, 2024). Cultural normalization through food marketing and events such as 4/20, alongside sustained industry growth, further reinforces acceptance of THC-infused food products (Fabricant, 2024; IBISWorld, 2024).

RESEARCH DESIGN AND SAMPLING

Research Design

To confirm our research targets and obtain valuable insights, we designed a survey of 25 thoughtfully developed questions. The survey process was designed to be simple, straightforward, and rapidly completed. The survey only took an average of five minutes for users to complete, leaving room for little drop-off and a maximum of valuable responses.

We utilized a range of scale types within the survey to provide both quantitative and qualitative data. These consisted of multiple-choice questions, Likert scales, and open-ended questions providing the opportunity for extended responses. To streamline the workflow of the survey as well as the applicability of each question, skip logic was also utilized. This made the survey adapt dynamically based on the respondent's answer, in a way that each respondent only responded to questions specific to their background or experience. This not only assisted the respondents but also provided us with accurate data.

We also added demographic questions so that we could have some idea of who took the survey. These questions told us about the respondents' backgrounds, identities, and age ranges, allowing us to identify patterns of use of cannabis in specific groups—i.e., young women, our main areas of focus. Our survey in total found a balance between being concise while being complete enough for us to have quality responses, with limited time for completion.

Sampling Approach

We used convenience snowballing sampling and non-probability methods. Our convenience sample consisted of our social networks: family, friends, and acquaintances reached through Instagram, Facebook, Email, and LinkedIn. These key persons forwarded the survey to their acquaintances, friends, and peers, allowing the survey to spread organically.

This was an ideal method for our study because we were researching a specialty group of a young generation who consume marijuana. Snowball sampling helped us interview that particular group of individuals with greater ease than would have been the case with a random sample approach. It helped us tap individuals who would not have been reached using the normal dissemination avenues. We surveyed from March 13th until April 3rd and received 128 usable responses. The snowball sampling technique of our approach not only provided us with an abundance of responses but also assured that our results reflected existing experiences being conveyed within respective peer groups. It was convenient, as well as efficient, and combined to work for our purpose**.**

Hypotheses

Based on our literature research, we examined what attributes could have a great influence on a consumer's choice to purchase cannabis (THC) products before formulating our hypotheses. Research and analysis of our survey results led us to four key areas that appeared to play a role in purchasing cannabis (THC): the amount of cannabis (THC) in the product, whether individuals believe it has health benefits, how dependent one may be on cannabis (THC), and their age. We developed our hypotheses using these and conducted experiments to determine whether a significant correlation exists between each one and a consumer's readiness to buy. We applied techniques including linear regression and T-tests to determine this, which allowed us to determine whether the outcomes were statistically significant or merely random. The four hypotheses include THC Levels, Health Benefits, Dependency, and Age.

To test our first hypothesis, we evaluated how important the level of cannabis (THC) is when deciding to purchase a product by crafting a question that asks, "How important is a product's Cannabis (THC) level?" We created a new variable called "Scale Variable", which we used to run tests against. For our first hypothesis, we ran a linear regression test to uncover the correlation between the level of cannabis (THC) and the willingness to purchase.

To test our second hypothesis, we evaluated how perceived health benefits that come with consuming a product that contains cannabis (THC) will affect the willingness to purchase. We created a question that states "To what degree do you believe that Cannabis (THC) products can provide potential health benefits, for example, pain relief, reducing anxiety, and or better-quality sleep?" We ran a linear regression to understand the correlation with the scale variable that we created.

To test our third hypothesis, we evaluated how the level of dependency of a person correlates with the willingness to purchase. We formed a question that states, "How dependent are you on consuming Cannabis (THC) products?" A linear regression test was conducted to test the relationship between the new scale variable and the question regarding dependency.

To test your fourth hypothesis, we evaluated how age affects the willingness to purchase Cannabis (THC) products. We formed a simple question that states, "Are you over 40 years of age?" We conducted an independent sample T-test to test the relationship between age and willingness to purchase.

Conceptual Model

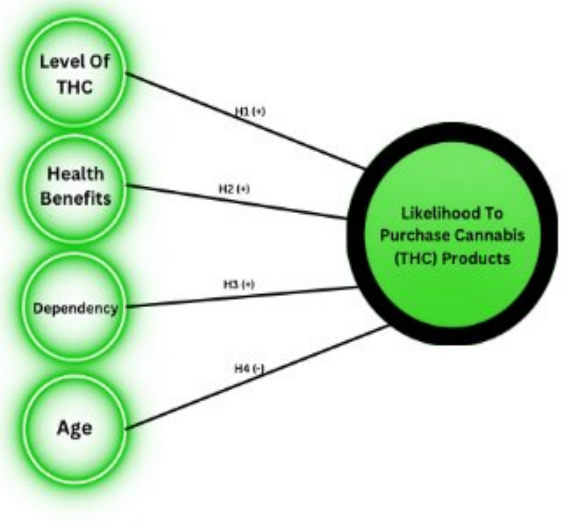
The conceptual model better illustrates the relationship between our independent and dependent variables, which is the willingness to purchase cannabis (THC). This conceptual model was formulated based on literature and exploratory research. We have developed 4 independent variables— THC levels, Health benefits, Dependency, and Age, which we believe will influence willingness to purchase.

The first hypothesis we developed was (THC) levels and the influence they have on the willingness to purchase cannabis (THC). THC levels are typically associated with users who may have a higher cannabis tolerance and seek to have a stronger, more effective experience. We have hypothesized that as the (THC) level increases, the likelihood of purchasing Cannabis (THC) also increases, indicating a positive relationship.

The second hypothesis we developed was based on health benefits and the influence they have on willingness to purchase cannabis (THC). We discovered that if Cannabis (THC) is perceived to provide health benefits that include pain relief, better quality sleep, stress reduction, and appetite improvement, this may influence consumers' interest in purchasing cannabis (THC). We believe that the more health benefits people perceive in consuming cannabis (THC) products, the greater the likelihood of consumers purchasing cannabis (THC) products. Thus, indicating a positive relationship.

The third hypothesis we developed is dependency and the willingness to purchase cannabis (THC). Dependency is often related to consumers who perceive themselves as needing or habitually using Cannabis (THC). We decided to include dependency because people who use cannabis (THC) more often are probably going to buy it, too. If a person habitually relies on it daily, whether for stress or sleep, they're more likely to keep purchasing. These kinds of users also tend to know what products they like, which allows marketers to better understand the consumers' behavioral patterns. We believe that consumers who perceive themselves with a Cannabis (THC) dependency will have a greater likelihood of purchasing Cannabis (THC).

The last hypothesis we developed is that age and the willingness to purchase cannabis (THC). We decided to include age because we were curious about how different age groups might feel about Cannabis (THC). We predicted that people older than 40 might be less likely to purchase due to how cannabis (THC) was viewed during their generation. However, people under 40 might be more open to purchasing cannabis (THC) due to recent legalization across many states and people becoming more open to it since it is more accepted now. We thought this difference could affect whether someone is likely to buy a cannabis (THC) product or not. However, when analyzing the data collected from our survey, it was found to have no statistical significance between the two variable relationships, indicating a negative relationship.



Scales And Validation

The scale that was utilized was the Willingness to Purchase scale, from the Marketing Book of Research (Bruner, 2019).

Firstly, the scale used was tested for internal consistency using the Willingness to Purchase, which had a very good internal reliability score. The Cronbach's alpha for the Willingness to Purchase scale was 0.975, with a sample size of 5 items.

Sample Profile

In terms of the sample profile, the findings revealed that most respondents identified as Female (57.9%), 41.3% were male, and 0.8% preferred not to say. Most of our respondents were younger than 40 years old. 62.2% reported that they were either 40 years old or under, while the other 37.8% of respondents were over the age of 40.

Table 1. Sample Profile Demographics, Frequencies, and Valid Percentages

Demographic Variables	**Demographics**	**Frequency**	**Valid Percentage**
Gender	Female	73	57.9%
	Male	52	41.3%
	Prefer not to say	1	0.8%
Age	40 years old or younger	79	62.2%
	Over the age of 40 years old	48	37.8%

Note: Research sample n = 127

Table 2. Sample profile, Ranges, Frequency, Valid percentage

Variables to help better understand the sample	**Scale Points**	**Frequency**	**Valid Percentage**
Amount of money spent per month on Cannabis products	$0	69	54.3%
	$1-$20	17	13.4%
	$21-$50	12	9.4%
	$51-100	15	11.8%
	$100 +	14	11.0%
		N=127	
Importance of taste in Cannabis products.	1 Not important	45	35.7%
		11	8.7%
	2	6	4.8%
	3	17	13.5%
	4	18	14.3%
	5	9	7.1%
	6	20	15.9%
	7 Extremely Important	N=126	
The primary source of purchasing Cannabis.	Dispensary	39	58.2%
	Online Store	1	1.5%
	Friends or	23	34.3%
	family	4	6.0%
	Other	N=67	
Consumption rate of Cannabis products.	Never	58	46.0%
	Occasionally	33	26.2%
	Once a week	5	4,0%

	Multiple times a week Daily	15 15 N=126	11.9% 11.9%

DATA ANALYSIS AND FINDINGS

To test our first hypothesis, we evaluated how important the level of THC is when deciding to purchase a product by crafting a question that asks, "How important is a product's Cannabis (THC) level?" We created a new variable called "willingness to purchase ", which represented each respondent's willingness to purchase cannabis products. For our first hypothesis, we ran a linear regression test to uncover the correlation between the level of THC and the willingness to purchase. The findings of the test were that there is a significant relationship between the willingness to purchase and the level of THC. Therefore, it is safe to say that the higher the level of THC that a product contains, the higher the probability that a consumer will be more likely to purchase it and vice versa. These findings inform marketers to improve the level of THC that their product contains due to a direct correlation with a higher chance of purchasing. With these findings, we can confidently say that our first hypothesis is supported.

To test our second hypothesis, we evaluated how perceived health benefits that come with consuming a product that contains THC will affect the willingness to purchase. We created a question that states "To what degree do you believe that Cannabis (THC) products can provide potential health benefits, for example, pain relief, reducing anxiety, and or better-quality sleep?" We ran a linear regression to understand the correlation with the scale variable that we created. The findings of the test showed a significant relationship between perceived health benefits and the willingness to purchase. Therefore, it is safe to say that the more perceived health benefits a product has, the higher the probability of purchase. The findings did support our hypothesis. These findings are very beneficial for markets, as they can promote the health benefits that the product may contain. If companies take this approach to promote the health benefits, they should see an increase in sales.

To test our third hypothesis, we evaluated how the level of dependency of a person correlates with the willingness to purchase. We formed a question that states, "How dependent are you on consuming Cannabis (THC) products?" A linear regression test was conducted to test the relationship between the new scale variable and the question regarding dependency. The results of the test were that there is a statistical significance between dependency and the willingness to purchase. Therefore, it is reasonable to state that a person who has a higher dependency on THC will be more likely to purchase THC products. These results support our hypothesis. This information can be useful to marketing firms, in that this shows they should be marketing to consumers who have a higher dependency and customers who are frequent purchasers.

To test your fourth hypothesis, we evaluated how age affects the willingness to purchase Cannabis (THC) products. We formed a simple question that states, "Are you over 40 years of age?" We conducted an independent sample T-test to test the relationship between age and willingness to purchase. The results indicated that there is no statistical significance between age and the willingness to purchase. Therefore, it is reasonable to say that age does not impact the willingness to purchase THC products. The results from this test do not support our initial hypothesis. This information is incredibly valuable to sellers, in that they should focus on the level of THC that the product contains, promote the health benefits of the product, and that the more someone is dependent on THC, the higher the chance they are willing to purchase, but not market to a certain age group.

Table 3. Description of Hypotheses, Results, and Explanation (Support or Not Support)

Hypotheses	**Test Results (Statistics)**	**Hypothesis supported or not**
H1: A higher level of THC increases the likelihood of purchasing Cannabis (THC) products.	Linear Regression: P< 0.001, F statistic 62.543, Beta coefficient of 0.579, R^2 value of 0.335.	The hypothesis is supported due to a p-value that is less than 0.001.
H2: The more health benefits there are while consuming Cannabis (THC) products, the higher the	Linear Regression: P< 0.001, F Statistic of 60.414, Beta Coefficient of 0.572, R^2 value of 0.328.	The hypothesis is supported due to a p-value that is less than 0.001.

likelihood of people purchasing Cannabis (THC) products.		
H3: Consumers with a higher dependency on Cannabis (THC) are more likely to purchase Cannabis (THC) products.	Linear Regression: $P< 0.001$, F Statistic of 84.233, Beta Coefficient of 0.636, and an R^2 value of .405.	The hypothesis is supported due to a p-value that is less than 0.001.
H4: The older an individual is, the less likely they are to purchase Cannabis (THC) products.	Independent Samples T-test- μ of 17.2766 for individuals over the age of 40 and 18.9241 for individuals under the age of 40. The Standard Deviation was 11.72085 for people over the age of 40 and 12.54200 for individuals under the age of 40, t value of -730 and -743, P value of .467 and .459.	The hypothesis is not supported due to a p-value that is greater than 0.05

Table 4. Statistical test to better understand the willingness to purchase based on several factors.

Variable Being Tested	**Test Results (Statistics)**	**Result**
Money spent on Cannabis monthly	One-Way ANOVA- Statistical significance due to the F-statistic of 99.620 and a P-value of >0.05	There is a statistical significance between the amount spent on Cannabis/THC products and the willingness to purchase.
Money spent on Cannabis monthly	Post Hoc Test (Scheffe)- <.001 significance at the $0.00 spent.	There is statistical significance in that the number of people who spent $0.00 was significantly higher than that of all other groups.
- Rating of knowledge about Cannabis/ THC products. - How Important is taste in Cannabis/ THC products? - Agreement with the statement that Cannabis/ THC products are socially acceptable. - Degree of perceived health benefits. - Concerned about potential health risks associated with consuming or inhaling THC - Importance of Products, THC level	- Bivariate Regression Test- Beta coefficient of .190, significance level of .007. - Multiple Regression Test- Beta coefficient of .073, a significance level of .308. - Multiple Regression Test- Beta coefficient of .108 and a significance level of .219. - Multiple Regression Test- Beta coefficient of .269 and a significance level of .005. - Multiple Regression Analysis- Beta coefficient of -.125, and a .063 level of significance. - Multiple Regression Test- Beta coefficient of .345 and a significance level of <.001.	- There is a statistical significance due to a p-value of <.05. - There is no statistical significance due to a p-value that is greater than .05. - There is no statistical significance due to a p-value that is >.05. - There is a statistical significance due to a p-value that is less than the .05 threshold. - There is no statistical significance due to a p-value that is greater than .05. - There is a statistical significance due to a p-value <.05.

CONCLUSION AND RECOMMENDATIONS

Even with limitations, there are a few key takeaways that marketers can use regarding the willingness to purchase. The level of THC in a product is crucial for an individual's willingness to purchase the product. Therefore, a company should market products that contain a higher THC level due to the demand of the consumer. Also, if a marketer can show health benefits for the product, they have a higher chance of selling more THC products. Marketing to people who have a higher dependency level of THC is extremely beneficial because that is the consumer market that is more willing to purchase. Lastly, there is no need to market to a specific age group based on the results of our study. Therefore, a company should focus on the level of THC in products, the health benefits that the product may have, and consumers who have a higher dependency on THC products.

After analyzing the data collected, we found there were a few limitations in our research design and overall execution of the survey. When the survey was first created, our knowledge of how to develop and administer a survey was minimal; we were new to using scales as a team for a foundation for a project, and our ability to run statistical tests was limited. To better understand the strength of our statistical results, it would have been preferred to have collected a wider data set than we received. Also, answering some of our questions would have allowed us to possibly better determine relationships regarding the willingness to purchase cannabis infused products. The last major limit was who we sent the survey to and collected responses from. The majority of our respondents being female may have impacted the results in some manner due to a sampling error. Also, using snowball sampling may lead to an inaccurate sample of the overall population, leading to possible skewed results. Recommendations for further research are to collect a larger sample size to better help determine an accurate representation of the population. This would lead to a stronger and more accurate statistical output.

REFERENCE CITED

Blum, D. (2024, January 29). *The health effects of cannabis edibles versus smoking*. *The New York Times*. https://www.nytimes.com/2024/01/29/well/cannabis-edibles-smoking.html

Bruner, G. C., II. (2019). *Marketing scales handbook: Multi-item measures for consumer insight research* (Vol. 10). GCBII Productions.

Donnan, J., Shogan, O., Bishop, L., Swain, L., Najafizada, M., & Smith, G. (2022). Characteristics that influence purchase choice for cannabis products: A systematic review. *Journal of Cannabis Research, 4*(17). https://doi.org/10.1186/s42238-022-00135-4

Fabricant, F. (2024, April 20). *The best foods to enjoy on 4/20*. *The New York Times*. https://www.nytimes.com/2024/04/20/dining/420-food.html

Goodman, S., Wadsworth, E., Leos-Toro, C., & Hammond, D. (2022). THC labeling on cannabis products: An experimental study of approaches for labeling THC servings on cannabis edibles. *International Journal of Drug Policy, 102*, 103597. https://doi.org/10.1016/j.drugpo.2022.103597

Grewal, J. K., Loh, L. C., & others. (2020). Health considerations of the legalization of cannabis edibles. *Canadian Journal of Public Health, 111*(5), 791–798. https://doi.org/10.17269/s41997-020-00382-0

IBISWorld. (2024). *Industrial hemp production in the United States* (IBISWorld Industry Report OD3315). https://www.ibisworld.com/united-states/industry/industrial-hemp-production/3315/

Kowitt, S. D., Yockey, R. A., Lee, J. G. L., Jarman, K. L., Gourdet, C. K., & Ranney, L. M. (2022). The impact of cannabis packaging characteristics on perceptions and intentions. *American Journal of Preventive Medicine, 63*(5), 751–759. https://doi.org/10.1016/j.amepre.2022.04.030

McEvoy, E. (2025, January 17). *Survey finds more adults turn to cannabis than prescription sleep aids to help them sleep*. *MMW News*. https://mmw.news/2025/01/17/survey-finds-more-adults-turn-to-cannabis-than-prescription-sleep-aids/

Moreno-Rius, J. (2019). The cerebellum, THC, and cannabis addiction: Findings from animal and human studies. *Neuroscience & Biobehavioral Reviews, 98*, 1–11. https://doi.org/10.1016/j.neubiorev.2019.01.002

Reboussin, B. A., Wagoner, K. G., Sutfin, E. L., Suerken, C. K., Ross, J. C., Egan, K. L., Walker, S., & Johnson, R. M. (2019). Trends in marijuana edible consumption and perceptions of harm in a cohort of young adults. *Drug and Alcohol Dependence, 205*, 107660. https://doi.org/10.1016/j.drugalcdep.2019.107660

Schaeffer, K. (2024, April 10). *9 facts about Americans and marijuana*. *Pew Research Center*. https://www.pewresearch.org/short-reads/2024/04/10/9-facts-about-americans-and-marijuana/

Zhu, B., Guo, H., Cao, Y., An, R., & Shi, Y. (2021). Perceived importance of factors in cannabis purchase decisions: A best–worst scaling experiment. *International Journal of Drug Policy, 91*, 102793. https://doi.org/10.1016/j.drugpo.2020.102793

UNCONVENTIONAL FACTORS THAT AFFECT MLB WINNING PERCENTAGE

Samuel Desany, Siena University
Maxfield Taylor, Siena University
Alexander Eberlein, Siena University
Dr. Manimoy Paul, Siena University – Faculty Mentor

ABSTRACT

Baseball teams are often evaluated using traditional statistics such as batting average, home runs, and earned run average, but these measures do not fully explain why some teams are more successful than others in key scoring situations. This study examines unconventional factors that contribute to offensive inefficiency in Major League Baseball, with a particular focus on wasted scoring opportunities measured through left-on-base (LOB) efficiency. Using a team-level panel dataset covering all 30 MLB teams across approximately 2014–2025, this paper analyzes offensive, baserunning, defensive, and contextual variables drawn from sources such as Baseball-Reference, ESPN, and Baseball Savant. Multiple analytical methods were used, including clustering, principal component analysis, multiple regression, and stepwise regression, to identify the factors most strongly associated with differences in scoring efficiency. The regression results showed that runs per game were the strongest positive predictor of LOB efficiency, while the full regression model explained about 49.9% of the variation in the dependent variable. The stepwise regression model provided stronger explanatory power, explaining about 85% of the variation in LOB efficiency, and identified runs per game, run differential, grounded into double plays, and caught stealing as significant predictors. Home runs per game showed a significant negative relationship with efficiency, suggesting that power hitting alone does not guarantee effective conversion of baserunners into runs. Overall, the findings show that offensive success in Major League Baseball depends not only on generating opportunities but on executing efficiently in high-leverage situations.

INTRODUCTION

Baseball is a game that not only considers performance in terms of standard statistics like batting averages, home runs, and earned run averages but also incorporates more complex forms of statistics. With the development of modern-day Major League Baseball, it is becoming increasingly popular for baseball clubs to use analytics to gain insight into the reasons behind the consistent performance of certain clubs while others underperform. Consequently, inefficiencies in team performance have been recognized as a valuable research topic. A clear example of inefficiency involves situations where teams create opportunities to score but fail to turn those chances into runs.

The consequences of failure to score runs when given such an opportunity could be highly influential in determining the outcome of baseball games, as baseball is highly dependent on the quality of performance during high-pressure moments. While teams are able to create runners on base, if they do not capitalize on their opportunities to score, then the inability to score will ultimately cost them a win against another team. In the span of a 162-game season, small inefficiencies could add up to have an impact on team results.

This topic would be significant to baseball fans as well as MLB executives because it may be that a team appears very powerful on all fronts, including hitting, pitching, and defense, yet cannot perform well enough due to their inability to score runs during the scoring opportunities presented to them. Analysis of such missed opportunities provides an opportunity to examine the performance of such a team in greater depth and detail, rather than from the basic statistics. This topic would also be relevant to the study of this dataset since the data comprises offensive, defensive, baserunning, and environmental variables, including but not limited to left on base and travel miles.

The purpose of this study is to examine which atypical factors lead to wasted scoring opportunities in Major League Baseball. Using team-level data from sources such as Baseball Reference, ESPN, and Baseball Savant, this paper analyzes whether unconventional in-game and environmental variables help explain differences in offensive inefficiency across teams. More specifically, the study focuses on left-on-base totals as a measure of wasted scoring opportunities and tests whether situational, environmental, and team-efficiency variables help explain why some teams consistently strand more runners than others over the course of a season.

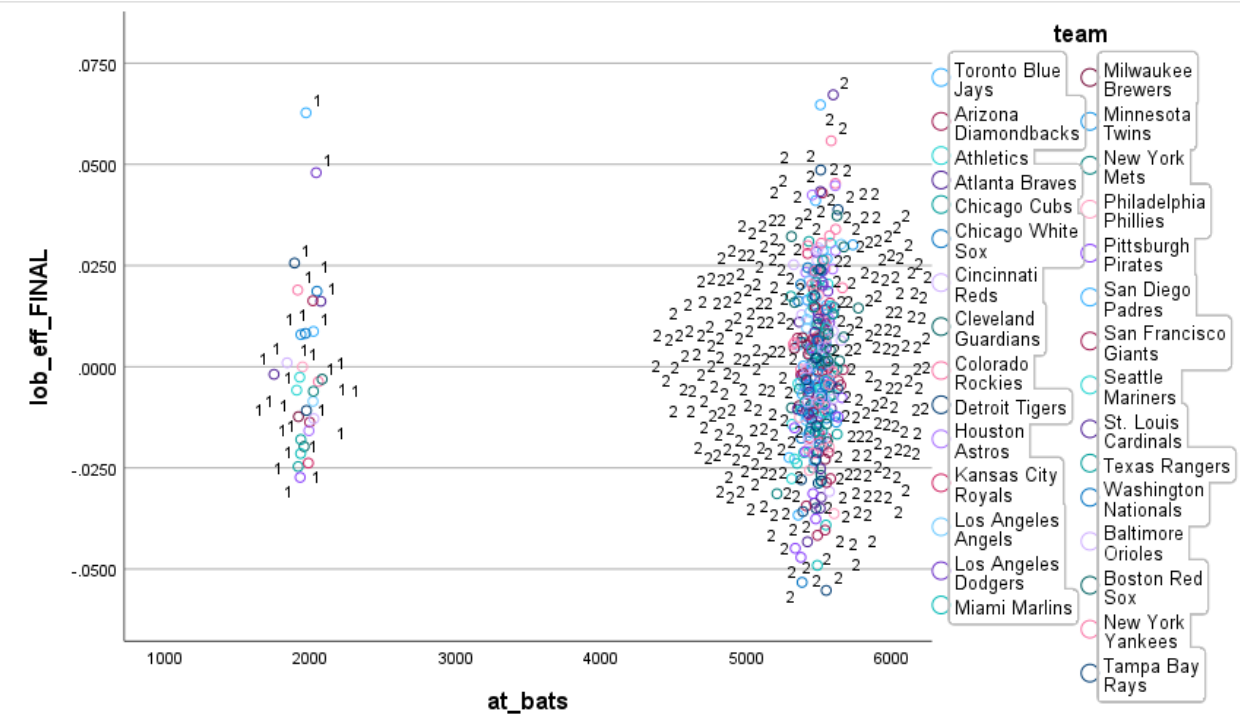
Prior research shows that success in Major League Baseball is closely linked to offensive performance, scoring prevention, and efficiency. According to Fuller (2023), run differential was among the best predictors of winning percentage, with an R^2 value of 0.846, indicating that a significant amount of variance in team success is attributed to the number of runs scored minus the number of runs allowed. This conclusion is further supported by the findings of another study conducted by Ballou-Crawford (2023). The study discovered that the use of a five-variable regression yielded a coefficient of determination of 0.882, implying that 88.2% of wins recorded during the regular season could be predicted using just five statistical factors. In the process, the study pointed out the significance of power hitting, on-base performance, and pitching in determining team success.

Even more recently, studies have reiterated the significance of pitching and batting efficiency. In a study done by Choi & Ji in 2025, data from the MLB for the years between 1901 and 2019 were subjected to factor analysis for 68 variables. They indicated that the value of the Kaiser-Meyer-Olkin statistic was 0.73 and the value of Bartlett's test p-value was less than 0.0001, which made the factor analysis appropriate for use. The results indicated that the pitcher's on-base percentage allowed and the batter's on-base percentage had the most significant impact on the winning percentage. In addition, it is also revealed that hits, home runs, baserunner control, and ERA of pitchers are important factors when the winning percentage becomes high. These observations further support the notion that the success of a team does not only depend on their overall offensive or defensive ability, but also on how they play in specific situations.

Other studies indicate the significance of team performance for organizational performance as a whole. According to Booth (2024), the increase in the proportion of foreign-born players in the team positively impacts their ERA and leads to a higher winning percentage. These results are observed regardless of payroll, population size, stadium capacity, number of All-Stars, and competing teams in other professional leagues. The same pattern has been revealed by Bradbury (2016), who found a positive correlation between revenue and winning percentage in MLB. Thus, the impact of team efficiency is not limited to its performance and may also affect the business side of baseball as well.

Despite the numerous significant determinants of winning percentage highlighted by prior studies, very little is known about the extent to which offensive inefficiency, measured by scoring opportunities that were not utilized, affects performance. Prior research largely concentrates on team results like wins, runs scored/allowed difference, offensive output, or pitching performance. Two teams might produce equal amounts of base-runners but differ greatly in their ability to convert such opportunities into runs. Therefore, this area needs additional research since left on base (LOB) totals and other situational measures become highly valuable. The present research expands on the available body of literature, providing new insights into this topic while paying special attention to the inefficiencies within offensive performance.

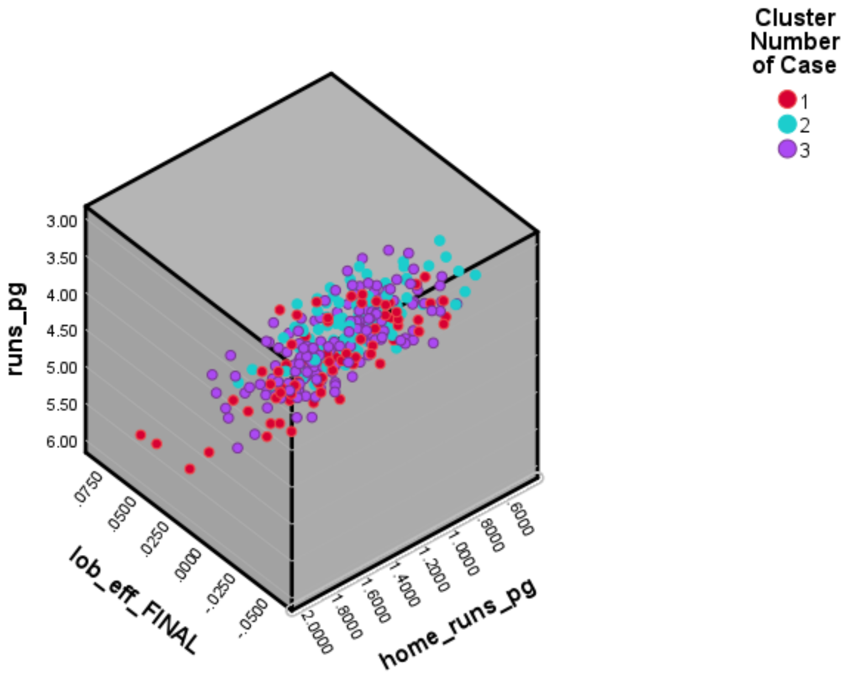
DATA



This dataset includes all 30 Major League Baseball teams across multiple seasons, including the COVID-affected 2020 season. Each observation represents a team-season combination, allowing for both cross-sectional and time-series analysis of team performance. Including the 2020 season introduces a unique variation in the dataset due to the shortened schedule and unusual playing conditions, which may slightly distort trends compared to full 162 game seasons. However, retaining this year strengthens the dataset by preserving continuity and allowing for a more comprehensive analysis of team efficiency across different contexts.

From a statistical perspective, this full dataset provides a strong foundation for identifying patterns in offensive inefficiency, particularly through the key dependent variable, LOB Efficiency. Because the dataset spans approximately 20 years, it captures long-term trends in how teams generate and convert scoring opportunities. This broad scope is critical for minimizing the impact of outliers and ensuring that any relationships observed, such as between offensive production and stranded runners, are consistent and not driven by short-term fluctuations.

Most importantly, this dataset supports the central research question: why do some teams consistently strand more baserunners than others? By including a wide range of variables such as hits, runs, plate appearances, travel distance, and situational metrics, the dataset allows for a deeper exploration of both traditional and non-traditional factors that may influence offensive inefficiency. This comprehensive structure ensures that the analysis is not limited to basic performance metrics but instead captures the complexity of team success in Major League Baseball.



The 3D cluster analysis visually groups team-seasons into three distinct clusters based on selected performance variables. Clustering is particularly useful in this context because it allows us to identify natural groupings of teams with similar statistical profiles without imposing predefined categories. In this case, the three clusters likely represent tiers of team performance, such as high-efficiency teams, average teams, and low-efficiency teams, based on their offensive production and LOB efficiency.

From an analytical standpoint, the separation (or lack of separation) between clusters provides insight into how distinct these performance groups actually are. If the clusters appear tightly grouped and clearly separated, it suggests that the variables chosen-such as runs per game, home runs, or LOB efficiency-are strong differentiators of team performance. However, if the clusters overlap, this indicates that inefficiency in scoring opportunities is influenced by multiple interacting factors rather than a single dominant variable. This aligns with the broader theme of the project, which emphasizes that offensive inefficiency is not driven by one metric alone but by a combination of situational and contextual variables.

Additionally, the inclusion of LOB Efficiency as either a clustering variable or a comparative metric is crucial. It allows us to observe whether teams with similar offensive output differ significantly in how efficiently they convert baserunners into runs. If teams with similar hits or runs fall into different clusters due to differences in LOB efficiency, this provides strong evidence that situational performance plays a key role in overall team success. This finding directly supports the hypothesis that wasted scoring opportunities are a meaningful differentiator in Major League Baseball performance.

The dataset used in this study is a team-level panel dataset constructed from publicly available data on Major League Baseball (MLB) teams across multiple seasons (approximately 2014–2025). Each observation represents a single team in a single season, producing a dataset that captures both cross-sectional and time-series variation in team performance.

Raw data were compiled primarily from Baseball-Reference and supplemented with additional calculated variables. These data include traditional offensive statistics (e.g., hits, runs, home runs), situational metrics (e.g., left on base), and environmental or contextual variables (e.g., travel distance, ballpark characteristics).

To align with the research objective, understanding inefficiency in scoring opportunities, several variables were engineered. Most importantly, a new metric called Left On Base (LOB) Efficiency was created to quantify how effectively teams convert baserunners into runs.

Key Variable: LOB Efficiency and Wasted Opportunity Rate:

LOB Efficiency = Runs / Total Baserunners

Total baserunners are approximated as the sum of hits, walks, and hit-by-pitches.

Wasted Opportunity Rate = 1 - LOB Efficiency

This variable directly captures the proportion of baserunners that fail to score, making it a more intuitive measure of offensive inefficiency.

Preliminary analysis of the dataset indicates that LOB Efficiency Final values are tightly distributed, generally ranging from approximately -0.050 to 0.050, with the majority of team-season observations concentrated near zero. This narrow range highlights that differences in offensive efficiency across MLB teams are often subtle; however, even small deviations can accumulate over the course of a full season and meaningfully impact total run production and team success.

From an interpretive standpoint, values near -0.050 represent teams with low offensive efficiency, where a large proportion of baserunners are left stranded, and scoring opportunities are frequently wasted. Values between approximately 0.000 and 0.025 reflect average performance, while values exceeding 0.040 indicate highly efficient teams that consistently convert baserunners into runs through strong situational execution. The clustering of observations around zero suggests that most teams generate similar offensive opportunities, but differ in how effectively they capitalize on them.

These descriptive patterns are strongly supported by the regression and stepwise regression results. In both models, runs per game and run differential emerged as the most significant positive predictors of LOB efficiency ($p < 0.001$), indicating that teams that score more frequently and outperform opponents tend to fall toward the higher end of the efficiency distribution. At the same time, situational variables such as grounded into double plays and caught stealing were also highly significant, reinforcing that inefficiency is driven not only by overall offensive production but by mistakes in key moments that prevent baserunners from scoring.

Interestingly, the stepwise regression results revealed that home runs per game had a statistically significant negative relationship with LOB efficiency ($p < 0.01$). This suggests that teams relying heavily on power hitting may not necessarily improve their ability to convert baserunners, as home runs often produce isolated scoring events rather than sustained offensive sequences. This finding reinforces the idea that teams positioned near the upper end of the efficiency range are not simply those with the highest offensive output, but those that demonstrate consistent situational execution and sequencing of events.

Overall, the descriptive distribution of LOB efficiency, combined with regression-based evidence, supports the central conclusion of this study: offensive efficiency in Major League Baseball is not driven solely by volume-based statistics, but by a combination of situational performance, decision-making, and the ability to capitalize on scoring opportunities.

The dataset incorporates a comprehensive set of team-level variables designed to capture multiple dimensions of performance, including offensive production, situational execution, baserunning behavior, defensive effectiveness, and contextual factors. Offensive production variables include measures such as runs per game (runs_pg), hits per game (hits_pg), home runs per game (home_runs_pg), total plate appearances, at-bats, and runs batted in (RBI), which collectively describe a team's ability to generate scoring opportunities.

Plate discipline and approach at the plate are represented through variables such as walks per game (walks_pg) and strikeouts by batters (strikeouts_bat), which reflect a team's ability to control at-bats and avoid unproductive outcomes. Baserunning metrics, including stolen bases per game (stolen_bases_pg) and caught stealing, capture aggressiveness and efficiency on the basepaths, while situational performance variables such as left on base (LOB), grounded into double plays (grounded_into_dp), and sacrifice flies provide direct measures of how effectively teams convert scoring opportunities into runs.

In addition to offensive and situational factors, the dataset includes defensive and advanced performance metrics such as defensive efficiency, assists, double plays, total defensive runs (rtot), and defensive runs saved (rdrs and rdrs_per_year). These variables help capture how defensive performance may indirectly influence offensive efficiency through game context and run prevention. Pitching-related impact is incorporated through variables such as runs allowed per game (runs_allowed_pg), which reflects overall defensive and pitching effectiveness.

Finally, the dataset includes contextual and environmental variables such as team travel distance (travel_miles) and average batter age (bat_age), which are intended to capture external influences like fatigue, experience, and roster composition. By combining traditional statistics with situational and advanced metrics, the dataset is structured to provide a multidimensional view of team performance, allowing for a deeper analysis of the factors that contribute to inefficiency in converting baserunners into runs.

HYPOTHESIS

H1: Teams with higher runs per game are more likely to have higher LOB efficiency.

H2: Teams with a higher run differential are more likely to convert baserunners into runs efficiently.

H3: Teams with more hits per game are not necessarily more efficient at scoring.

H4: Teams with higher home runs per game are not more likely to have higher LOBE efficiency.

H5: Teams that ground into more double plays are more likely to have lower LOB efficiency.

H6: Teams with more caught stealing are more likely to waste scoring opportunities.

H7: Teams with better plate discipline (more walks, fewer strikeouts) are more likely to improve efficiency.

H8: Teams with higher LOB totals are more likely to have lower offensive efficiency.

H9: Teams with stronger overall offensive production are not always the most efficient in scoring situations.

H10: Teams that minimize unproductive outs are more likely to convert scoring opportunities.

H11: Teams with better situational execution are more likely to strand fewer baserunners.

H12: Offensive efficiency in MLB is driven more by situational performance than by total offensive output.

H13: Offensive efficiency in MLB is driven more by situational performance than by total offensive output.

H14: Teams with longer travel distances are more likely to experience reduced efficiency due to fatigue.

H15: Teams with more consistent offensive sequencing are more likely to convert opportunities into runs

H16: Teams with higher LOB totals are more likely to have lower offensive efficiency.

H17: Teams with balanced offensive approaches (not relying solely on power) are more efficient in scoring situations.

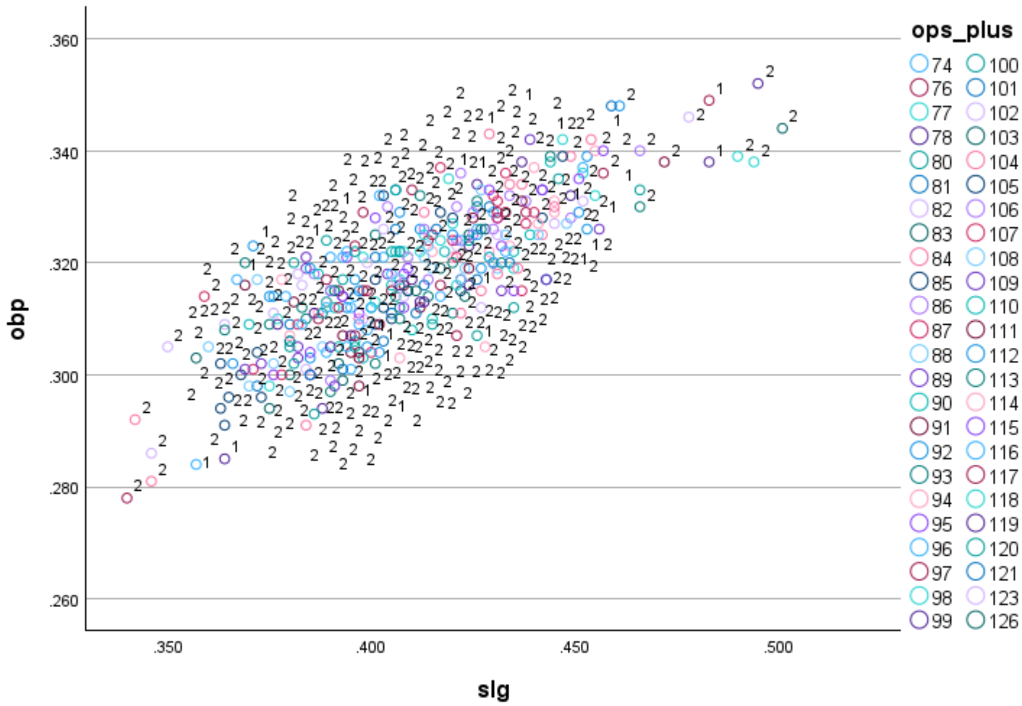
METHODOLOGY AND RESULTS

The central idea of this dataset is that not all offenses are created equal. Two teams may generate the same number of baserunners but differ significantly in how efficiently they convert those opportunities into runs. Traditional metrics such as batting average or OPS capture overall offensive ability but fail to isolate situational performance. This dataset instead focuses on how often teams waste scoring opportunities, using LOB efficiency as a key measure of offensive effectiveness under pressure. By combining standard offensive metrics with contextual variables like travel fatigue and roster composition, the dataset allows for a deeper investigation into why some teams consistently strand runners while others capitalize on similar opportunities. The dataset is structured to answer the question: Which factors contribute to wasted scoring opportunities in Major League Baseball?

A critical implication of this framework is that offensive success is not purely a function of volume, but of execution. While metrics such as hits per game or home runs per game measure how frequently teams reach base or generate power, they do not account for sequencing, which is when those events occur within an inning. The regression and stepwise results support this distinction, showing that variables tied to overall scoring outcomes, such as runs per game and run differential, are far more influential than raw offensive totals. Additionally, situational variables such as grounded into double plays and caught stealing emerged as significant contributors to inefficiency, indicating that poor decision-making and unproductive outs can negate otherwise strong offensive performance.

Furthermore, the dataset highlights the importance of multidimensional analysis in understanding team performance. Offensive inefficiency is not driven by a single factor, but rather by an interaction of offensive production, situational execution, and contextual influences. For example, the presence of both offensive and defensive variables in the final models suggests that overall team quality and game context play a role in shaping scoring outcomes. This reinforces the idea that teams positioned at the higher end of LOB efficiency are not simply those with the most talent, but those that demonstrate consistency, discipline, and strategic effectiveness in high-leverage situations. Ultimately, this dataset provides a more nuanced lens through which to evaluate performance, moving beyond traditional statistics to better explain the underlying causes of success and inefficiency in Major League Baseball.

Clustering:



The scatterplot visualization provides a two-dimensional representation of the relationships between key variables, often with clusters or grouping variables overlaid. Typically, one axis represents an offensive production metric (such as runs per game or hits per game), while the other may represent LOB Efficiency or another efficiency-based variable. This type of chart is particularly effective for identifying trends, correlations, and potential outliers within the dataset.

One key insight from this visualization is the relationship between offensive production and efficiency. While one might expect that teams with higher offensive output would naturally have higher efficiency, the scatterplot often reveals that this is not always the case. Some teams may generate a high number of baserunners but still exhibit low LOB efficiency, indicating that they struggle in high-pressure or situational hitting scenarios. Conversely, other teams may have moderate offensive output but higher efficiency, suggesting better performance in key moments.

The clustering or grouping within the scatterplot further reinforces the idea that team performance is multi-dimensional. If distinct clusters appear, they highlight that teams can be categorized not just by how much offense they produce, but by how effectively they convert opportunities into runs. This directly ties back to the central narrative of the dataset: that traditional metrics alone do not fully explain team success. Instead, situational efficiency-captured through LOB Efficiency-is a critical factor in understanding why some teams outperform others despite having similar overall statistics.

PCA Test:

The next analytic approach that we took when looking at this data set was to run a PCA test on the different variables and see if there was any relationship between them that affected teams. A PCA test is a technique used on large datasets to reduce the size of them and their variables inside. These fewer uncorrelated variables are called the PCA's, which means principal components. This test makes sure to keep the maximum variance even after clustering variables into these groups.

The initial PCA test that we ran was in R-studio. When running this test, we found that the first 4 PCA's had taken up over 70% of the variance, and these variables included in each PC1, PC2, PC3, and PC4 were the most important when looking at all of the variables. When looking at each PC we were able to label what they are and what variables belong to each one. The label we came up with for PC1 is the most important variable, taking up 35.5% of the variance, is titled Offensive Volume Variables. This PC includes the variables hits, runs, plate_appearances, at_bats, RBI, and home_runs. The reason all of these variables are important is that they have a high positive loading on PC1. Since these variables can work together in a specific PC, they help define the overall offensive production in MLB teams. When interpreting this PC1 we can conclude that the teams that have more hits, get more at-bats, and score more runs will have the highest ranks when looking at the performance dimensions.

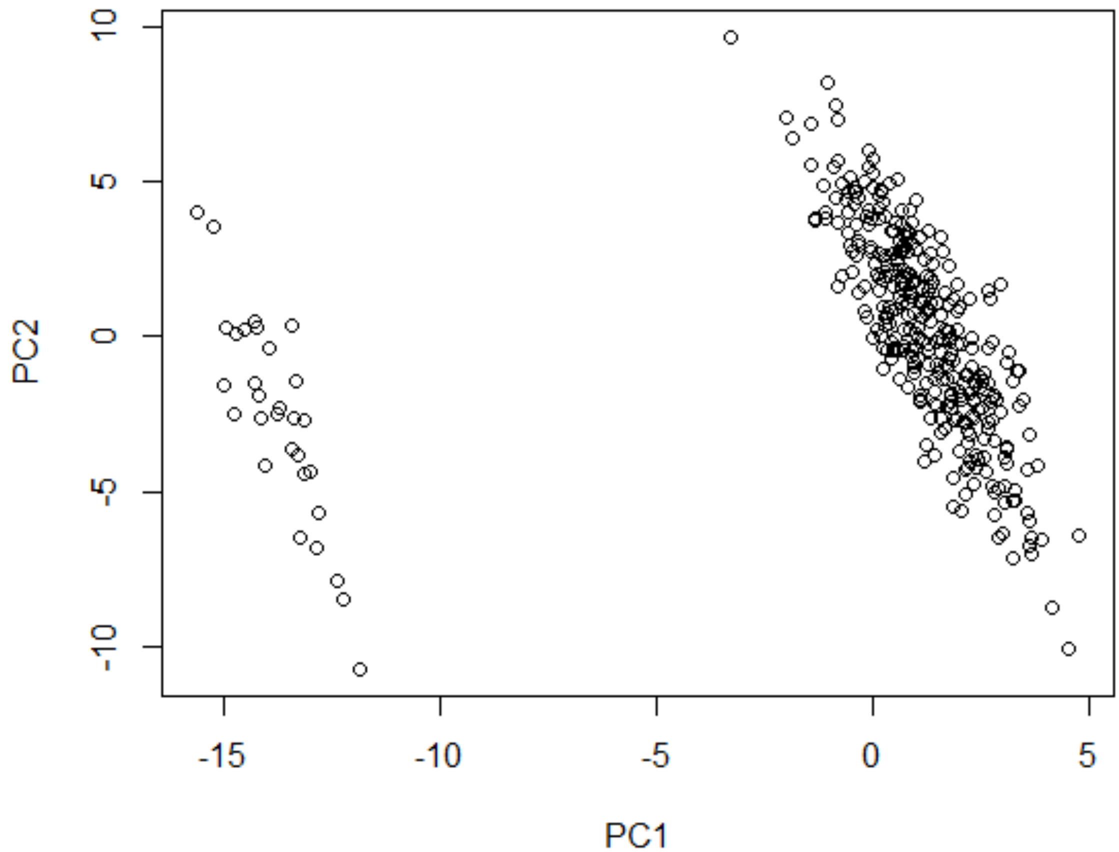
The next PC that we talked about, PC2 takes up around 20.7% of the variance is labeled Efficiency / Situational Performance Variables. The different variables that fall under this PC are runs_pg (runs per game), runs_diff (runs scored - runs allowed), lob_efficiency (how often are players left on base), and runs_allowed_pg. These all matter because of their strong correlation to PC2. PC2 is also the second largest source of the variance, making it useful to look at. This PC specifically separates teams into categories of those that score efficiently and those that waste opportunities. Overall, PC2 describes how well teams can convert their chances into runs.

The Third PCA being PC3 is what we look at and takes up around 8.7% of the variance, making it pretty useful. The title we gave this PCA is Pitching / Defense Variable. The key variables when looking at the PC test results are runs_allowed_pg, and we also took some influence from the run_diff. These variables matter to PC3 because of how strongly they show up and they look at preventing runs from happening, as opposed to scoring runs. This PC3 really shows us a different angle by saying that a team's success is not just the team's offense, but the defense as well as the pitching will explain to us a different chunk of the variance. Overall, this PCA is saying how well a sports team stops its opponents.

The final PC that we took into consideration was PC4, which took up 6.1% of the variance. Although this has the least amount of variance out of the 4, it is still very important to our research because when adding the first 4 PC variances together, we get over 70% of the variance. The name of the PC4 that we found is Context / Secondary Variables. The variables included in this are travel_miles, bat_age, triples, and situational stats such as hits_pg.

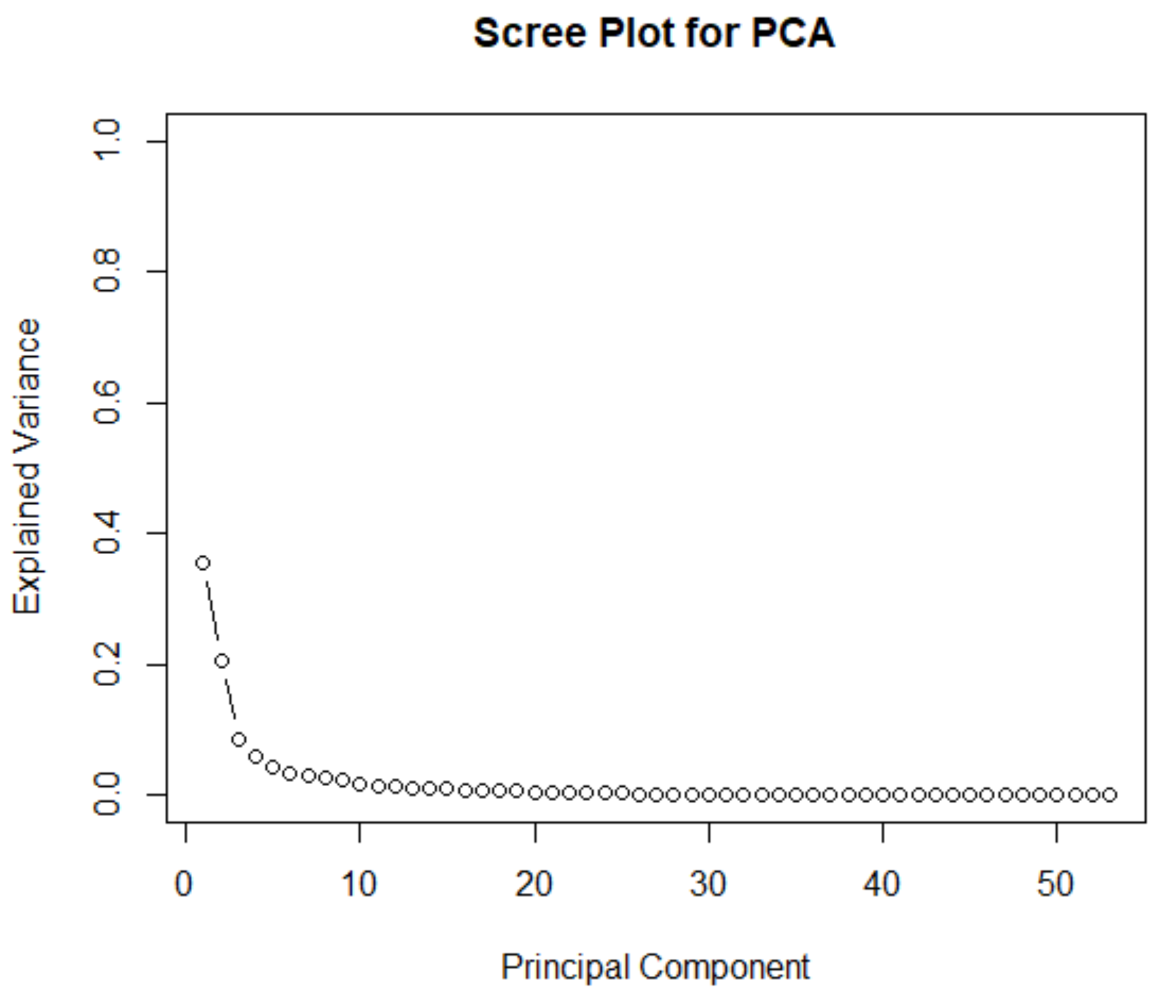
These factors are good to look at because they are outside factors that don't all relate to the actual gameplay. These are important because they are smaller influences, but also some contextual factors.

PC1 vs PC2:



The PC1 vs PC2 scatter plot compares the two most important principal components because they explain the largest share of variation in the dataset. By focusing on these two dimensions, the plot allows us to visualize the most meaningful patterns in team performance without the complexity of all 53 variables. PC1 primarily represents overall offensive production, while PC2 captures efficiency in converting scoring opportunities. The downward trend in the plot suggests a negative relationship between these two components, indicating that teams with higher offensive output do not always perform efficiently in scoring situations. Additionally, the spread of points along the vertical axis shows that teams with similar offensive production can differ significantly in efficiency. The presence of clustering also highlights that most teams fall within a similar performance range, while a smaller group of teams stands out with notably lower offensive production. Overall, the plot provides a clear visual that inefficiency is a distinct and important factor in explaining differences in team performance.

Scree Plot:



The scree plot illustrates how much variance each principal component explains and helps determine how many components should be retained for analysis. The plot shows a steep decline from the first principal component to the third or fourth, followed by a noticeable flattening of the curve. This "elbow" point indicates that the first few components capture the majority of the meaningful variation in the dataset, while the remaining components contribute relatively little additional information. In this case, the first four principal components account for approximately 71% of the total variance, suggesting that most of the structure in the data can be summarized using only a small number of underlying factors. The gradual flattening after this point reflects diminishing returns, meaning that additional components mainly capture noise or very specific patterns rather than broad trends. This supports the decision to focus on the first few principal components when interpreting the results and analyzing team performance.

Regression Analysis:

The multiple linear regression model produced statistically significant results, with an R^2 value of 0.499, indicating that approximately 49.9% of the variation in LOB efficiency is explained by the included variables. The overall model was highly significant (F-statistic $p < 2.2e-16$), suggesting strong explanatory power.

Among the independent variables, runs per game emerged as the most significant predictor of LOB efficiency ($p < 0.001$), with a positive relationship indicating that teams scoring more runs tend to convert baserunners more effectively. Run differential was marginally significant ($p \approx 0.074$), suggesting that teams with stronger overall performance may also exhibit improved efficiency, although this effect is weaker when controlling for other variables.

Interestingly, several traditional offensive metrics, including hits per game and home runs per game, were not statistically significant predictors. This suggests that simply generating offensive output does not guarantee efficiency in converting scoring opportunities. Additionally, travel distance was not significant, indicating that environmental factors such as travel may have a limited direct effect on offensive efficiency. Overall, these findings support the idea that situational performance, rather than raw offensive volume, plays a key role in determining LOB efficiency.

Stepwise Regression:

Stepwise regression was conducted to identify the most important predictors of LOB efficiency while reducing redundancy among variables. The final model demonstrated strong explanatory power, with an R^2 of 0.8495 and an adjusted R^2 of 0.8402, indicating that approximately 85% of the variation in offensive efficiency is explained by the selected variables. The overall model was highly statistically significant (F-statistic $p < 2.2e-16$).

Among the variables retained in the final model, runs per game and run differential emerged as the strongest positive predictors of LOB efficiency ($p < 0.001$), suggesting that overall offensive production and team performance are critical drivers of efficiency. Additionally, situational variables such as grounded into double plays and caught stealing were highly significant, indicating that poor baserunning decisions and double plays contribute to wasted scoring opportunities. Interestingly, home runs per game showed a statistically significant negative relationship with efficiency, suggesting that power hitting alone does not necessarily improve the ability to convert baserunners into runs.

It is important to note that left_on_base was also highly significant; however, this variable is closely related to the dependent variable and may inflate the model's explanatory power. Despite this, the model highlights that offensive inefficiency is driven not only by overall production but also by situational execution and decision-making. Overall, the stepwise regression results reinforce the central hypothesis that unconventional and situational factors play a critical role in determining scoring efficiency in Major League Baseball.

Stepwise Regression Results

Stepwise Regression Results

Variable	Step 1 (b)	Step 1 (p-value)	Step 2 (b)	Step 2 (p-value)
runs_pg	0.02314	1.49e-08	0.02052	8.24e-07
runs_diff	0.00245	0.0736	0.00864	2.43e-13
home_runs_pg	0.00326	0.5616	-0.01057	0.006533
hits_pg	0.00260	0.2050	—	—
travel_miles	1.13e-07	0.1470	—	—
grounded_into_dp	—	—	0.000239	1.93e-13
caught_stealing	—	—	0.000373	1.51e-10

Models excluding variables directly related to LOB were also considered to ensure robustness.

Chart:

lob_eff_FINAL	Final left-on-base efficiency metric used as the primary outcome (Y variable). Represents how effectively a team converts baserunners into runs after adjustments or transformations applied for analysis.
lob_efficiency	Proportion of baserunners that are left on base (LOB ÷ total baserunners). Measures how efficiently a team capitalizes on scoring opportunities—lower values indicate better run conversion, while higher values suggest more missed chances.
travel_miles	Total miles the team traveled during the season. Used to measure potential fatigue or schedule difficulty.
bat_age	Average age of the team's hitters. Gives context on experience vs youth.
runs_pg	Runs scored per game. Measures offensive scoring ability.
runs_allowed_pg	Runs allowed per game. Measures pitching + defense effectiveness.
runs_diff	Runs scored minus runs allowed. A key indicator of overall team performance.
games	Total games played (usually 162). Used to standardize per-game stats.
plate_appearances	Total batting opportunities. Includes all times a batter comes to the plate.
at_bats	Official batting attempts (excludes walks, HBP, etc.). Used for batting average.
runs	Total runs scored by the team. Core offensive output.
hits	Total number of hits. Shows how often the team gets on base via hits.
hits_pg	Hits per game. Standardizes hitting across seasons.
doubles	Number of 2-base hits. Indicates gap power.
triples	Number of 3-base hits. Usually reflects speed and ballpark factors.
home_runs	Total home runs hit. Measures power hitting.
home_runs_pg	Home runs per game. Normalizes power across teams.
rbi	Runs batted in. Shows how often players drive in runs.
stolen_bases	Total stolen bases. Reflects team speed and aggressiveness.
stolen_bases_pg	Stolen bases per game. Standardized speed metric.

caught_stealing	Times runners were thrown out stealing. Measures inefficiency on bases.
walks	Total walks (BB). Indicates plate discipline.
walks_pg	Walks per game. Shows how often the team gets on base via patience.
strikeouts_bat	Total strikeouts by hitters. Indicates contact issues.
strikeouts_bat_pg	Strikeouts per game. Standardized measure of strikeouts.
batting_avg	Hits divided by at-bats. Basic measure of hitting success.
obp	On-base percentage. Measures how often players reach base.
slg	Slugging percentage. Measures power (extra-base hit ability).
ops	OBP + SLG combined. Overall offensive performance metric.
ops_plus	OPS adjusted for league and ballpark (100 = average). Lets you compare across seasons.
total_bases	Total bases from hits (1B=1, 2B=2, etc.). Measures total offensive production.
grounded_into_dp	Double plays hit into. Negative offensive stat that kills rallies.
hit_by_pitch	Times batters were hit by pitches. Another way to reach base.
sacrifice_hits	Bunts that advance runners. Reflects small-ball strategy.
sacrifice_flies	Fly balls that score a run. Shows situational hitting.
intentional_walks	Walks given on purpose. Indicates respect for hitters.
left_on_base	Total runners left stranded. Measures missed scoring opportunities.
defensive_efficiency	Percent of balls in play turned into outs. Key team defense metric.
innings_fielded	Total innings played on defense. Usually ~1458 for full season.
chances	Total defensive opportunities (putouts + assists + errors). Measures workload.
putouts	Outs recorded directly by fielders. Core defensive stat.
assists	Plays where a fielder helps make an out (e.g., throwing to first). Measures defensive involvement.
errors	Mistakes that allow runners to advance. Indicates defensive weaknesses.
double_plays	Number of double plays turned. Shows defensive efficiency under pressure.
fielding_pct	(Fielding %): (Putouts + Assists) / Chances. Measures defensive reliability.
rtot	Total runs saved or allowed by defense compared to average. Advanced defensive metric.
rtot_per_year	Rtot adjusted per season. Standardizes defensive impact.
rdrs	Defensive Runs Saved. Estimates how many runs a defense prevented.
rdrs_per_year	RDRS adjusted per season. Allows fair comparisons.
rgood	Number of "good" defensive plays. Tracks positive defensive contributions.
pred_runs_diff	Predicted run differential from your regression model. Model's expected performance.
model_residual	Actual runs_diff minus predicted runs_diff. Shows model error.
model_score	Overall model evaluation score for the team. Higher means better fit or performance in your model.

CONCLUSION

In conclusion, this study provides strong evidence that offensive success in Major League Baseball is not determined solely by how much offense a team generates but by how efficiently it executes in critical scoring situations. While traditional statistics such as hits and home runs are commonly used to evaluate performance, the results of this analysis demonstrate that these measures alone fail to capture a key dimension of team success. Instead, variables such as runs per game and run differential emerged as the most consistent predictors of LOB efficiency, highlighting that overall scoring ability and team quality are more closely tied to efficiency than raw offensive totals.

Across all analytical methods used in this study, including clustering, principal component analysis, and regression modeling, a consistent pattern emerged. PCA revealed that offensive volume and situational efficiency are distinct dimensions of team performance, while clustering showed that teams with similar levels of offensive production can differ significantly in their ability to convert scoring opportunities. Regression and stepwise results further reinforced these findings by identifying situational variables, such as grounded into double plays and caught stealing, as key contributors to inefficiency. These results suggest that wasted scoring opportunities are not random but are influenced by specific, measurable aspects of team behavior and decision-making.

One of the most compelling findings of this study is that certain commonly valued offensive metrics may not contribute to efficiency in the way that is often assumed. In particular, the negative relationship between home runs per game and LOB efficiency suggests that power hitting alone does not guarantee success in converting baserunners into runs. This highlights the importance of sequencing, discipline, and situational execution, as teams that rely heavily on isolated scoring events may struggle to capitalize on broader offensive opportunities. In contrast, teams that minimize mistakes-such as avoiding double plays and baserunning errors-are more likely to consistently convert scoring chances into runs.

Ultimately, this study reframes how offensive performance should be evaluated in Major League Baseball. Rather than focusing solely on how much offense a team produces, it is equally important to consider how effectively that offense is utilized. Small inefficiencies, when repeated over the course of a long season, can have a significant impact on team outcomes. From both a competitive and strategic perspective, these findings suggest that teams, analysts, and front offices should place greater emphasis on situational performance and efficiency-based metrics when evaluating players and constructing rosters. In doing so, they can gain a more complete understanding of what truly drives success in baseball. Taken together, these findings demonstrate that in Major League Baseball, success is not simply about generating opportunities, but about capitalizing on them-ultimately reinforcing that efficiency, not volume, is the true driver of performance.

REFERENCES

Ballou-Crawford, C. (2023). Which statistics in baseball are most important for winning? *Fort Hays State University*. https://scholars.fhsu.edu/cgi/viewcontent.cgi?article=1303&context=sacad

Baseball Reference. (n.d.). *Baseball statistics and history*. Retrieved April 12, 2026, from https://www.baseball-reference.com/

Booth, A. S. (2024). The impact of international baseball players on MLB organizations: An analysis of performance, attendance, revenue, and valuation, 2012–2022 (Bachelor’s thesis, Claremont McKenna College). *Scholarship@Claremont*. https://scholarship.claremont.edu/cmc_theses/3416

Bourne, C. (2025). Strategic management and profit optimization in Major League Baseball: A framework for sustainable franchise success (Undergraduate honors thesis, Brigham Young University). *BYU ScholarsArchive*. https://scholarsarchive.byu.edu/studentpub_uht/473

Bradbury, J. C. (2016). Winning and other determinants of revenue in North America’s major professional sports leagues. *Kennesaw State University*. https://digitalcommons.kennesaw.edu/facpubs/4011/

Bronshteyn, M. (2025). The impact of MVP seasons on MLB team financial performance: Evidence from gate receipts and attendance. *International Journal of High School Research*, 7(8). https://terra-docs.s3.us-east-2.amazonaws.com/IJHSR/Articles/volume7-issue8/IJHSR_2025_78_36.pdf

Choi, S. H., & Ji, S.-K. (2025). A study of winning percentage in the MLB using fuzzy Markov regression. Mathematics, 13(6), 1008. https://doi.org/10.3390/math13061008

Denaux, Z. S., Denaux, D. A., & Yalcin, Y. (2011). Factors affecting attendance of Major League Baseball: Revisited. *Atlantic Economic Journal,* 39(2), 117–127. https://doi.org/10.1007/s11293-011-9274-2

ESPN. (n.d.). *Sports news, scores, and statistics*. Retrieved April 12, 2026, from https://www.espn.com/

FanGraphs. (n.d.). *Baseball statistics and analysis*. Retrieved April 12, 2026, from https://www.fangraphs.com/

Fuller, Z. O. (2023). An analysis of Major League Baseball sabermetrics. *Winona State University.* https://openriver.winona.edu/cgi/viewcontent.cgi?article=1036&context=rca

Lim, N., & Pedersen, P. M. (2018). Examining determinants of sport event attendance: A multilevel analysis of a Major League Baseball season. *Journal of Global Sport Management,* 7(1), 181–198. https://doi.org/10.1080/24704067.2018.1537675

Ormiston, R. (2014). Attendance effects of star pitchers in Major League Baseball. Journal of Sports Economics, 15(4), 338–364. https://doi.org/10.1177/1527002512461155

Ormiston, R. (2015). The evolution of competitive balance and its effects on attendance in Major League Baseball. *Journal of Sports Economics,* 16(5), 451–477. https://doi.org/10.1177/1527002514535171

Park, K. W. (Ken), Lee, S., & Miller, P. (2013). Ticket pricing per team: The case of Major League Baseball (MLB). *Journal of Economics & Economic Education Research,* 14(3), 89–105.

Schmidt, M. B., & Berri, D. J. (2001). Competitive balance and attendance: The case of Major League Baseball. Journal of Sports Economics, 2(2), 145–167. https://doi.org/10.1177/152700250100200204

Scully, G. W. (1974). Pay and performance in Major League Baseball. *American Economic Review,* 64(6), 915–930. https://www.jstor.org/stable/1815242

UNDERSTANDING THE DRIVERS OF BINGE DRINKING: AN EMPIRICAL ANALYSIS USING U.S. SURVEY DATA

Rory McNerney, Siena University
Kolawole Olawole, Siena University
Tristan Terry, Siena University
Dr. Manimoy Paul. Siena University – Faculty Mentor

ABSTRACT

This study investigates the primary determinants of binge drinking behavior within the United States, utilizing individual-level data derived from the Behavioral Risk Factor Surveillance System (BRFSS). A multi-stage empirical methodology was employed, starting with logistic regression and subsequently incorporating clustering and principal component analysis to address issues of heterogeneity and multicollinearity among the predictor variables. The results demonstrate that demographic characteristics, health status, and economic conditions all exert significant influence upon binge drinking behavior. Specifically, sex emerged as a dominant predictor, while individuals reporting chronic health conditions and physical limitations exhibited a significantly lower propensity to engage in binge drinking. The final predictive model achieved robust performance, attaining an overall accuracy of 94.8% and substantially enhancing the identification of high-risk individuals relative to the initial regression analysis. Collectively, the findings underscore the multidimensional nature of binge drinking and emphasize the critical necessity of integrating behavioral, health, and structural factors in the formulation of targeted public health interventions. The hypotheses guiding this analysis are detailed in Appendix B.

INTRODUCTION

Binge drinking remains a persistent and significant public health concern in the United States, contributing to a wide range of adverse outcomes that span chronic physical and mental disease, acute injury, violence, and premature death. This pattern of consumption, fundamentally defined by the intake of an excessive amount of alcohol over a short duration, is recognized as a high-risk behavior that is intricately linked to alcohol-related harm. While distinct from a formal clinical diagnosis of alcohol use disorder (AUD), binge drinking is widely regarded by public health professionals as a critical behavioral marker for problematic and dangerous alcohol use, serving as a primary contributor to alcohol-related illness and mortality statistics. Despite substantial and ongoing investment in public health campaigns, legislative changes, and educational initiatives, binge drinking stubbornly persists across the spectrum of American society, affecting diverse age cohorts, genders, and socioeconomic strata. This persistence underscores a critical need for more nuanced, deeper research into the complex and interacting factors that determine its occurrence. The challenge of effectively reducing binge drinking stems from its multifactorial causation; it is not attributable to a singular cause but rather emerges from a complex interaction of individual behaviors, immediate social environments, and overarching structural and environmental conditions.

Recognizing this complexity, this study undertakes an empirical examination of the determinants of binge drinking using a robust source of data: individual-level responses from the Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS offers a massive, nationally representative dataset, providing an opportunity to analyze the interplay between an extensive array of behavioral, demographic, and socioeconomic variables and the propensity for heavy episodic drinking. The analysis is specifically geared toward outlining the relationship between high-risk alcohol consumption and a defined set of variables, focusing particularly on variables related to demographics, health status, and health access. The central research inquiry guiding this analysis is: Which specific determinants—encompassing individual behaviors, demographic characteristics, and access to health resources—most significantly contribute to the likelihood of individual-level binge drinking within the United States? By addressing this question, this study endeavors to generate concrete, actionable insights into the core drivers of high-risk alcohol consumption. These findings are intended to directly inform the development of more strategically targeted and effective public health interventions, ultimately enhancing the understanding of the complex risk landscape associated with binge drinking and contributing to improved population health outcomes.

LITERATURE REVIEW

The phenomenon of binge drinking and its underlying determinants have been the subject of extensive investigation across a wide array of academic disciplines. These fields include, but are not limited to, economics, which often examines the role of pricing and availability; public health, which focuses on population-level interventions and risk factors; and psychology, which delves into individual-level motivations, personality traits, and coping mechanisms. The established literature across these disciplines consistently demonstrates that the consumption of alcohol, particularly in patterns that constitute binge drinking, is not a simple, monocausal behavior. Instead, it is demonstrably influenced by a complex confluence of interwoven factors. These influential factors fall into three main categories: demographic and socioeconomic (age, sex/gender, race/ethnicity, location, finances, education, employment), behavioral and psychological (substance use, academics, risky behaviors, social networks), and policy and environmental (legal drinking age, alcohol taxes and pricing, outlet density/location restrictions, and drunk driving law enforcement). In essence, binge drinking is not driven by a single, isolated cause but rather reflects a dynamic and complex interaction between inherent individual characteristics. A comprehensive understanding requires moving beyond a search for a solitary driver to an appreciation of the multifaceted, ecological nature of this public health challenge.

Demographic and Socioeconomic Factors

A substantial body of research finds that demographic characteristics are pivotal in shaping patterns of binge drinking behavior across populations. Among these, age stands out as one of the most consistently identified and robust predictors. Specifically, a clear developmental pattern is observed: younger individuals, particularly adolescents and young adults, are statistically more likely to engage in risky behaviors such as heavy episodic or binge drinking. Conversely, older adults tend to exhibit a different pattern of consumption, characterized by more frequent alcohol intake but generally at lower quantities per occasion. This age-related distinction is well-supported by developmental psychology and behavioral research, which posits that risk-taking is intrinsically linked to age-related neurobiological changes, social maturation, and major life transitions, such as entering college or the workforce (Schulte et al., 2019). The prefrontal cortex, responsible for impulse control, continues to develop into the mid-twenties, which partially accounts for the heightened risk-taking observed in younger age groups.

The influence of socioeconomic status on alcohol use is equally significant, although its effects are often nuanced and complex. Research indicates a differential pattern of consumption and consequence based on economic standing. Individuals from higher-income brackets may consume alcohol more frequently, often as part of social or professional activities, and may be able to mitigate some of the immediate negative repercussions due to better access to healthcare and stable social support systems. In stark contrast, lower-income populations, while not necessarily drinking more frequently or in higher quantities than their affluent counterparts, often experience more severe negative consequences from their alcohol use. These consequences can manifest as increased health problems, greater financial strain, and higher rates of alcohol-related legal issues, highlighting a disparity in the protective factors available across the socioeconomic status spectrum.

Beyond individual and economic factors, the immediate and broader social environment, including family background, cultural norms, and peer dynamics, exerts a powerful influence on drinking behavior. Cultural acceptance or prohibition of alcohol, the prevalence of alcohol use among peer groups, and specific household characteristics all interact to shape an individual's decisions regarding alcohol consumption. The family environment is particularly critical during formative years. Longitudinal research consistently suggests that early exposure to alcohol within the home environment, coupled with factors like family instability (e.g., parental divorce, conflict, or mental health issues), is strongly associated with elevated levels of alcohol use, and potentially alcohol use disorders, later in life (Khamis et al., 2022). This evidence underscores the essential idea that alcohol use is not a purely isolated individual choice but is deeply embedded within a wider, influential social and familial context, where norms, access, and protective or risk factors are established. Therefore, effective interventions must consider this layered interplay of demographic, economic, and social determinants.

Behavioral and Psychological Factors

High-risk alcohol consumption, such as heavy episodic drinking or binge drinking, is a complex phenomenon driven by a confluence of interconnected psychological, behavioral, and biological factors. Behavioral elements, especially the co-use of other psychoactive substances, are strongly correlated with elevated risk of substance misuse, including binge drinking. A well-documented example is the relationship between smoking and alcohol consumption, which are frequently observed as complementary behaviors. Individuals who smoke are statistically more likely to engage in heavy episodic drinking or binge drinking. This pattern is not merely a coincidence but often reflects broader underlying tendencies toward risk-taking, impulsivity, and addictive behavior across different domains. The simultaneous consumption of tobacco and alcohol can also pharmacologically enhance the rewarding effects of both substances, creating a potent synergistic reinforcement loop that drives continued high-risk use.

Psychological distress is a significant determinant of binge drinking. A robust body of evidence indicates that individuals grappling with mental health issues such as chronic stress, generalized anxiety, or major depressive disorder are substantially more prone to heavy episodic drinking. For these individuals, alcohol often serves as an immediate, albeit maladaptive, coping mechanism to temporarily mitigate or numb unpleasant emotional states. The research conducted during the COVID-19 pandemic provides compelling evidence: studies, such as the one by Schmits & Glowacz (2022), empirically demonstrated a clear association between heightened levels of anxiety and depression during the global crisis and a corresponding increase in alcohol consumption. This finding underscores the critical importance of evaluating and addressing an individual's mental health status as a foundational step in understanding and preventing binge drinking. Moreover, individuals with poor emotion regulation skills may be particularly susceptible to using alcohol to manage intense feelings, further solidifying the link between psychological vulnerability and high-risk drinking patterns.

On the biological front, genetic vulnerability plays a critical role, influencing both an individual's sensitivity to alcohol and their propensity for developing dependence. Differences in enzyme activity (e.g., alcohol dehydrogenase) and neurotransmitter systems (e.g., dopamine, GABA) can modulate the experience of intoxication and withdrawal, thus affecting drinking patterns. The complexity of these interactions has led to the adoption of advanced analytical techniques in the field. For example, Lee et al. (2019) utilized machine learning approaches to sift through large datasets and successfully identified key predictors of negative alcohol-related outcomes. These predictors included specific drinking behaviors, the presence of depression, and a history of substance dependence. The success of these multidimensional models in forecasting alcohol-related issues emphatically highlights the necessity of considering the full spectrum of psychological, behavioral, and biological elements to gain a comprehensive understanding of and to effectively intervene against high-risk alcohol consumption.

Policy and Environmental Factors

Beyond individual characteristics, policy and environmental factors are critical determinants of binge drinking and alcohol misuse, representing powerful levers for public health intervention. Economic theory robustly predicts that increasing the price of a commodity will reduce demand, a relationship consistently and strongly supported in the alcohol literature. Alcohol taxation is the most direct and effective pricing policy. A comprehensive review of alcohol tax studies found compelling evidence that a 10% increase in alcohol prices leads to approximately a 5% reduction in overall alcohol consumption (Wagenaar et al., 2010). This price elasticity is a crucial finding, indicating that consumers are sensitive to price changes. Furthermore, the impact of higher alcohol taxes extends far beyond simple consumption levels. They have been directly linked to significant reductions in alcohol-related morbidity and mortality. This includes decreases in acute outcomes such as traffic fatalities and alcohol-involved violence, as well as reductions in chronic disease outcomes such as liver cirrhosis and certain cancers (Wagenaar et al., 2010). These findings suggest that taxation and pricing policies are not merely punitive measures but highly effective, evidence-based tools for mitigating the substantial societal and health burdens associated with excessive alcohol use. Adjusting tax structures and implementing minimum unit pricing are contemporary policy discussions aimed at maximizing this public health benefit, particularly by targeting cheaper forms of alcohol favored by heavy drinkers.

Alcohol availability plays another key role in shaping drinking patterns. The ease with which individuals can access alcohol—determined by factors like the density of retail outlets (e.g., bars, liquor stores, etc.) or fewer restrictions on days and hours of sale—has been consistently associated with increased consumption and higher rates of alcohol-related harm. This harm affects not only the drinker but also others in the community (e.g., through alcohol-involved crashes or assaults). The regulatory environment, particularly stronger state-level alcohol policies, acts as a protective factor. Research by Fairman et al. (2019) and Greenfield, et al. (2020) demonstrated that states with more robust alcohol control policies—which may include restrictions on hours of sale, stricter server training requirements, or limitations on outlet density—are associated with lower odds of aggression-related and driving-related harms. These policies manage the environment of consumption and moderate the risk, illustrating that legislative action can directly influence public safety and health outcomes. Specific regulatory interventions, such as privatization versus state control of alcohol sales, the use of dram shop liability laws, and bans on specific marketing practices, also contribute significantly to the overall policy landscape

Geographic variation is an important determinant of binge drinking and alcohol misuse, highlighting the influence of broad, contextual factors. Research consistently shows that drinking patterns differ significantly across states and regions, reflecting a complex interplay of variations in culture and societal norms (e.g., local acceptance of heavy drinking, celebratory traditions, and community-level attitudes toward intoxication), policy environments (e.g., state-level regulatory differences), and demographics and economic structure (e.g., regional economic conditions, population density, and demographic composition). For instance, research has mapped significant differences in binge drinking prevalence across the United States. These regional differences powerfully suggest that alcohol consumption is not solely an individual choice or a manifestation of individual pathology, but is profoundly shaped by the broader social and policy environment in which individuals live (Kerr, 2010). Understanding this geographic heterogeneity is crucial for developing targeted, locally appropriate prevention and treatment strategies that address the specific environmental risks present in a given community.

Gaps in the Literature

While existing research has undoubtedly shed light on the complex phenomenon of binge drinking, a significant gap remains in the literature due to several methodological limitations. Many seminal studies, while foundational, often rely heavily on aggregate-level data, which obscures the crucial heterogeneity present at the individual level. This reliance on broad statistics, such as county or state averages, inherently limits the precision with which researchers can identify the specific, nuanced risk and protective factors for any given person. Consequently, the ability to tailor effective intervention strategies based on these findings is diminished. Furthermore, much of the existing work tends to focus on specific, often easily accessible, subpopulations (e.g., college students, military personnel, or clinical samples), which limits the generalizability of the findings to the wider adult population. A more critical constraint in the current body of literature is the frequent compartmentalization of explanatory factors. Few studies employ a comprehensive, multi-factorial framework that simultaneously examines the interplay of crucial determinant categories. Specifically, research often isolates one set of variables—such as focusing primarily on behavioral risks (e.g., smoking status, physical activity), demographic characteristics (e.g., age, race, gender), or socioeconomic status (e.g., income, education, employment)—while excluding others. This isolated approach fails to capture the intricate, synergistic relationships among these factors. For instance, the effect of low educational attainment on binge drinking may be significantly amplified or mitigated depending on a person's employment status and their community's social environment.

To address these critical limitations, this study makes a robust contribution to the public health and behavioral economics literature by adopting a highly detailed, individual-level empirical strategy. The research utilizes data from the Behavioral Risk Factor Surveillance System (BRFSS), a large-scale, nationwide health survey. The BRFSS provides the granular, self-reported health and behavioral information necessary to move beyond aggregate analysis. This foundation allows for the construction and testing of a unified empirical model. This unified framework is designed specifically to analyze multiple determinants of binge drinking simultaneously. By incorporating a wide and diverse range of variables—spanning self-reported health metrics, detailed demographic data, indicators of socioeconomic stability, and other relevant behavioral risk factors—the analysis aims to provide a far more holistic and comprehensive understanding of the full constellation of factors that influence and predict heavy episodic drinking behavior among the adult population. This comprehensive approach is essential for developing evidence-based prevention programs and resource allocation strategies that target the most salient risk combinations.

DATA

The core of this research lies in an empirical analysis, conducted using individual-level survey data. This robust dataset is particularly comprehensive, capturing an abundance of data about each participant, specifically encompassing detailed demographic characteristics, socioeconomic circumstances, a spectrum of health-related behaviors, and critical data regarding access to and utilization of healthcare services. The inherent structure and scope of the dataset ensure substantial variation across individuals, which is a crucial prerequisite for facilitating a thorough, methodologically sound, and structured examination of the multitude of factors potentially associated with engagement in binge drinking behavior. The methodological framework employs a clearly defined dependent variable, which is a precise measure of an individual's reported engagement in binge drinking. Conversely, the selection of independent variables has been a deliberate and carefully considered process, aiming to reflect the multiple, distinct dimensions of potential influence on this behavior. These explanatory variables are not chosen arbitrarily; rather, their selection is grounded firmly in the existing body of scholarly literature and is specifically intended to approximate several primary, overarching categories of factors consistently associated with binge drinking. These categories include, but are not limited to, demographic variables (such as age, gender, and race/ethnicity), socioeconomic indicators (including income level, employment status, and educational attainment), behavioral factors (such as smoking status and physical activity), and various health-related influences (like mental health status and chronic conditions). It is important to acknowledge that the comprehensive scope of these categories is necessarily subject to the specific limitations and parameters of the available survey data. The complete and detailed list of all variables, including their operational definitions, is presented below for reference.

Variables

Socioeconomic and Demographic Factors

● AGE_G - Imputed age in six groups (1 = Age 18 to 24, 2 = Age 25 to 34, 3 = Age 35 to 44, 4 = Age 45 to 54, 5 = Age 55 to 64, 6 = Age 65 or older)

● SEX - Calculated sex variable (1 = Male, 0 = Female)

● RACE_W - White or Not (0 = No, 1 = Yes, NA = Don’t Know/Not Sure/Refused)

● RACE_B - Black or Not (0 = No, 1 = Yes, NA= Don’t Know/Not Sure/Refused)

● RACE_OPOC - POC Other Than Black or Not (0 = No, 1 = Yes, NA = Don’t Know/Not Sure/Refused) ● INCOMG1 - Computed income categories (1 = Less than $15,000, 2 = $15,000 to $25,000, 3 = $25,000 to $35,000, 4 = $35,000 to $50,000, 5 = $50,000 to $100,000, 6 = $100,000 to $200,000, 7 = $200,000 or more, NA = Don’t know/Not sure/Missing)

● EDUCAG - Computed level of education completed categories (1 = Did not graduate High School, 2 = Graduated High School, 3 = Attended College or Technical School, 4 = Graduated from College or

Technical School, NA = Don't know/Not sure/Missing)

● MARITAL - Marital Status (0 = Not Married, 1 = Married, NA = Refused)

● RENTHOM1 - Own or Rent Home (0 = No, 1 = Own/Rent, NA = Don’t know/Not Sure/Refused) ● VETERAN3 - Are You A Veteran (0 = No, 1 = Yes, NA = Don’t know/Not Sure/Refused)

● EMPLOY1 - Employment Status (0 = Unemployed, 1 = Employed, NA = Refused)

● CHLDCNT - Computed number of children in household (1 = No children in household, 2 = One child in household, 3 = Two children in household, 4 = Three children in household, 5 = Four children in

household, 6 = Five or more children in household, NA = Don't know/Not sure/Missing)

● METSTAT - Metropolitan Status (0 = Nonmetropolitan counties, 1 = Metropolitan counties)

Health Factors

● PHYSHLTH - Number of Days Physical Health Not Good (1 - 30 = Number of days, NA = Don't know/Not sure/Refused)

● BMI5CAT - Computed body mass index (1 = Underweight, 2 = Normal Weight, 3 = Overweight, 4 = Obese)

● DIFFWALK - Difficulty Walking or Climbing Stairs (0 = No, 1 = Yes, NA = Don't know/Not Sure/Refused)

● CHCKDNY2 - Ever told you have kidney disease? (0 = No, 1 = Yes, NA = Don't know / Not sure/ Refused)

● DIABETE4 - Ever told you had diabetes (0 = No, 1 = Yes, NA = Don't know/Not Sure/Refused)

● CVDSTRK3 - Ever Diagnosed with a Stroke (0 = No, 1 = Yes, NA = Don't know/Not sure/ Refused) ● SMOKER3 - Computed Smoking Status (1 = Current smoker - now smokes every day, 2 = Current smoker - now smokes some days, 3 = Former smoker, 4 = Never smoked, NA = Don't know/Refused/Missing) ● TOTINDA - Leisure Time Physical Activity Calculated Variable (0 = No physical activity or exercise in last 30 days, 1 = Had physical activity or exercise, NA = Don't know/Refused/Missing)

● MENTHLTH - Number of Days Mental Health Not Good (0 = None, 1 - 30 = Number of days, NA = Don't know/Not sure/Refused)

● GENHLTH - General Health (1 = Excellent, 2 = Very good, 3 = Good, 4 = Fair, 5 = Poor, NA = Don't know/Not Sure/Refused)

Health Access Factors

● CHECKUP1 - Length of time since last routine checkup (1 = Within past year (anytime less than 12 months ago), 2 = Within past 2 years (1 year but less than 2 years ago), 3 = Within past 5 years (2 years but less than 5 years ago), 4 = 5 or more years ago, NA = Don't know/Not sure/Never/Refused)

● PRIMINS1 - Is Your Current Primary Source of Health Care Coverage Private? (0 = No, 1 = Yes, NA = No coverage of any type/Don't know/Not Sure/Refused)

● PERSDOC3 - Have Personal Health Care Provider? (No = 0, 1 = Yes, only one, 2 = More than one, NA = Don't know/Not Sure/ Refused)

● MEDCOST1 - Could Not Afford To See Doctor (0 = No, 1 = Yes, NA = Don't know/Not sure/ Refused)

Descriptive Statistics

Binary Variables	Meaning	N	Yes (1)	No (2)	Total Check	Not Known
MEDCOST1	Could Not Afford To See Doctor	144,283	12,941	130,795	143,736	547
CVDSTRK3	Ever Diagnosed with a Stroke	144,283	6,191	137,626	143,817	466
CHCKDNY2	Ever told you have kidney disease?	144,283	7,427	136,147	143,574	709

DIABETE4	Ever told you had diabetes	144,283	20,631	123,322	143,953	330
MARITAL	Marital Status	144,283	72,644	70,333	142,977	1,306
RENTHOM1	Own or Rent Home	144,283	98,467	44,564	143,031	1,252
VETERAN3	Are You A Veteran?	144,283	14,952	128,511	143,463	820
EMPLOY1	Employment Status	144,283	72,425	69,202	141,627	2,656
DIFFWALK	Difficulty Walking or Climbing Stairs	144,283	22,125	115,665	137,790	6,493
METSTAT	Metropolitan Status	144,283	101,776	42,507	144,283	0
TOTINDA	Physical Activity Calculated Variable	144,283	109,852	33,984	143,836	447
SEX	Calculated sex variable	144,283	68,235	76,048	144,283	0
V_RACE_W	White or Not	144,283	107,406	33,943	141,349	2,934
PRIMINS1	Is Your Current Primary Source of Health Care Coverage Private?	144,283	61,833	68,992	130,825	13,458
V_RACE_B	Black or Not	144,283	9,991	134,292	144,283	0
V_RACE_OPOC	POC Other Than Black or Not	144,283	26,886	117,397	144,283	0

Non-Binary Variables	Meaning	N	Mean	Median	Mode
GENHLTH	General Health on 1-5 scale	144,283	Good (3)	Good (3)	Good (3)
PHYSHLTH	Number of Days Physical Health Not Good	144,283	4 days	0 days	0 days
MENTHLTH	Number of Days Mental Health Not Good	144,283	4 days	0 days	0 days
PERSDOC3	Have Personal Health Care Provider?	144,283	More than one	Only one	Only one
CHECKUP1	Length of time since last routine checkup	144,283	Within past 2 years	Within past year	Within past year
AGE_G	Imputed age in six groups	144,283	Age 45 to 54	Age 55 to 64	Age 65 or older
BMI5CAT	Computed body mass index	144,283	Overweight	Overweig ht	Overweig ht
CHLDCNT	Computed number of children in household	144,283	None	None	None
EDUCAG	Computed level of education completed categories	144,283	Attended College or Technical School	Attended College or Technical School	Graduate d from College or Technical School
INCOMG1	Computed income categories	144,283			$50,000 to $100,000 $50,000 to $100,000 $50,000 to $100,000

SMOKER3	Computed Smoking Status	144,283	Former smoker	Never smoked	Never smoked

Methodology and Analysis

This analysis aims to isolate the most influential predictors of individual-level binge drinking by examining a comprehensive set of behavioral, demographic, and healthcare access variables. The full set of hypotheses tested in this study is provided in Appendix B. The analysis of binge drinking behavior employed a multi-stage methodology, beginning with the estimation of a binary logistic regression model. The dependent variable, a binary indicator for binge drinking (_BFBING6), was derived from survey responses. To create a clear and precise model from the approximately 300 variables in the original dataset, a vital initial step was variable exclusion. This was done to minimize "noise" and enhance model precision. Preliminary assessment, specifically a Chi-Square test, determined that all categorical variables were statistically significant predictors of binge drinking.

Subsequently, all variables were subjected to the binary logistic regression model to systematically ascertain the most influential predictors of binge drinking behavior. This process successfully reduced the initial approximately 300 variables to a final selection of twenty-seven. This selection was predicated upon statistical significance and theoretical relevance, yielding three distinct categories of predictors: demographic and socioeconomic factors, health factors, and health access factors. Prior to the final model execution, all selected variables underwent stringent checks for collinearity, missing values, and outliers. The definitive regression model was executed in SPSS utilizing a stepwise procedure, which systematically evaluated and removed variables based on their statistical significance, ultimately clarifying the key determinants of binge drinking behavior.

To further enhance the analysis and identify underlying population segments, k-means clustering was applied using the selected variables. This approach grouped individuals into distinct clusters based on similarities across demographic, socioeconomic, and health-related characteristics. The resulting clusters were interpreted as representing low-risk and high-risk population groups, with one cluster characterized by more favorable health outcomes and the other by poorer physical and mental health indicators. To address potential multicollinearity and reduce redundancy among the explanatory variables, principal component analysis (PCA) was conducted. PCA serves as a dimensionality reduction technique, allowing correlated variables to be grouped into underlying components based on shared variance. This step improves model efficiency and ensures that the regression results are not distorted by overlapping predictors. The analysis was performed using the principal components extraction method, followed by Varimax rotation to enhance interpretability. The rotation procedure redistributes variance across components, producing a clearer structure in which variables load more strongly onto a single component. Components were retained based on eigenvalues and overall interpretability of the factor structure. The rotated component matrix revealed a set of six distinct groupings among the variables, indicating the presence of underlying dimensions within the data. Variables were assigned to components based on their highest factor loadings, with values above conventional thresholds considered meaningful contributors to each component. This process enabled the identification of clusters of related variables, which informed subsequent model refinement. Based on the PCA results, highly correlated variables were evaluated and selectively removed to reduce multicollinearity while preserving the conceptual integrity of the model. This step ensured that the final regression analysis was both efficient and statistically robust.

Following these refinements, another logistic regression model was estimated using cluster membership as the dependent variable. This model was designed to identify the key factors distinguishing high-risk and low-risk population segments. The resulting classification model achieved near perfect predictive accuracy. This outcome is consistent with the structure of the analysis, as cluster membership was derived from the same set of explanatory variables used in the regression model.

RESULTS

The initial logistic regression model was estimated using all twenty-seven selected variables to evaluate their influence on binge drinking behavior. Model performance was assessed using classification accuracy and statistical significance measures. The results indicated that the model correctly classified 99.5% of non-binge drinkers, but only 3% of binge drinkers, resulting in an overall accuracy of 85.3%. This imbalance in predictive performance is attributable to the distribution of the dependent variable within the dataset. Specifically, only 17,643 individuals in the sample of 144,283 were classified as binge drinkers, leading to a substantial class imbalance. As a result, the model is heavily biased toward predicting the majority class (non-binge drinkers), limiting its effectiveness in identifying binge drinking behavior.Following the initial model estimation, the statistical significance of each variable was evaluated. Variables with high p-values were systematically removed through an iterative refinement process. The first variables excluded were MEDCOST1 and EDUCAG, which did not demonstrate statistical significance.

After re-estimating the model, additional variables were removed due to lack of statistical significance, including PERSDOC3 and V_RACE_OPOC. Subsequent iterations resulted in the further exclusion of GENHLTH and RENTHOM1. At each stage, the model was re-estimated to assess the impact of variable removal on overall significance and model stability. The final refined model, seen below, retained only variables that were statistically significant predictors of binge drinking. This iterative selection process ensured that the model remained both parsimonious and robust, isolating the variables that provide the strongest explanatory power.

Variable B Sig.

-0.622
-0.509
-0.429
-0.398
-0.293
-0.167
-0.162
-0.129
-0.115
-0.096
-0.081
-0.072
-0.004
0.011

0.053
0.055
0.119
0.125
0.206
0.26
0.333

Constant <.001 CHCKDNY2 <.001 DIABETE4 <.001 V_SMOKER3 <.001 V_AGE_G <.001 V_RACE_B 0.001 VETERAN3 <.001 MARITAL <.001 DIFFWALK 0.005 CVDSTRK3 0.015 V_CHLDCNT <.001 V_METSTAT 0.001 PHYSHLTH 0.005 MENTHLTH <.001 CHECKUP1 <.001 V_BMI5CAT <.001 V_INCOMG1 <.001 V_RACE_W <.001 V_TOTINDA <.001 EMPLOY1 <.001 PRIMINS1 <.001 V_SEX 0.519 <.001

The table above presents the estimated coefficients and statistical significance levels from the initial logistic regression model. The results indicate that all retained variables are statistically significant predictors of binge drinking, with the majority exhibiting p-values below 0.001. The direction of the coefficients provides insight into how individual characteristics are associated with binge drinking behavior. Males are significantly more likely to binge drink, aligning with existing research. Conversely, negative coefficients for chronic conditions like kidney disease and diabetes, as well as physical limitations and poor general health, suggest these individuals are less likely to binge drink, consistent with medical expectations. Socioeconomic factors like income, employment, and physical activity are positively associated, potentially reflecting greater social engagement. Overall, the results highlight the multidimensional nature of binge drinking, with both demographic and health-related factors playing significant roles in shaping behavior. While the initial regression model provides valuable insight into individual predictors, it is limited by potential multicollinearity among variables and the presence of overlapping explanatory effects. To address these limitations and further refine the analysis, additional techniques were employed. Specifically, k-means clustering was used to identify distinct population segments based on shared characteristics, while principal component analysis (PCA) was conducted to reduce dimensionality and uncover underlying structures within the data. Building on these approaches, a final logistic regression model was estimated using cluster membership as the dependent variable. This framework allows for a more structured examination of the factors that distinguish high-risk and low-risk groups, providing deeper insight into the determinants of binge drinking behavior beyond the initial variable-level analysis.

To further examine heterogeneity within the data, k-means clustering was conducted using the refined set of variables. The algorithm converged after multiple iterations, indicating stable cluster formation. Two distinct clusters were identified, representing groups with markedly different health, behavioral, and demographic profiles. The final cluster distribution shows that Cluster 1 contains 80,906 individuals, while Cluster 2 includes 12,539 individuals, indicating a substantial imbalance between groups. This suggests that the majority of the population falls into a lower-risk profile, while a smaller subset exhibits characteristics associated with higher health risk. Examination of the final cluster centers reveals clear differences between the two groups. Cluster 1 is characterized by significantly lower values for both physical health limitations (PHYSHLTH) and mental health distress (MENTHLTH), indicating better overall health status. In contrast, Cluster 2 exhibits substantially higher values for these variables, suggesting greater physical and mental health burden. For example, the average number of poor physical health days is considerably higher in Cluster 2 compared to Cluster 1, with a similar pattern observed for mental health. Additional differences are observed in functional limitations and health conditions. Cluster 2 shows higher levels of difficulty walking (DIFFWALK) and greater prevalence of chronic conditions, indicating a population with reduced physical capacity and elevated health risk. Socioeconomic and behavioral differences are also evident, with variations in employment status, income levels, and lifestyle indicators further distinguishing the two groups. The ANOVA results confirm that these differences are statistically significant across nearly all variables included in the clustering process ($p < .001$). Variables such as physical health, mental health, difficulty walking, and income exhibit particularly large F-statistics, indicating strong discriminatory power between clusters. This suggests that health-related and socioeconomic variables play a central role in defining the separation between groups.

Based on these patterns, Cluster 1 can be interpreted as a lower-risk, healthier population, characterized by better physical and mental health, fewer functional limitations, and more stable socioeconomic conditions. In contrast, Cluster 2 represents a higher-risk population, defined by poorer health outcomes, greater functional impairment, and more constrained socioeconomic conditions.Overall, the clustering results reveal meaningful segmentation within the dataset, highlighting the presence of distinct population groups with different risk profiles. These findings provide a structured foundation for subsequent analysis, allowing for a more targeted examination of the factors that differentiate high-risk and low-risk individuals in the final regression model.

To address multicollinearity and reduce dimensional complexity, Principal Component Analysis (PCA) was conducted on the selected variables. Based on eigenvalues greater than one and supported by the scree plot, six principal components were retained. Together, these components explain approximately 51.6% of the total variance, indicating that a substantial portion of the information in the dataset is preserved while reducing redundancy across variables. Using a Varimax rotation to improve interpretability, each component was examined based on its strongest factor loadings and assigned a conceptual label:

Component 1: Life Stage & Household Stability

● Captures variation in age, employment, and household structure, reflecting an individual's position within the life cycle and domestic environment.

Component 2: Economic & Insurance Status

● Represents financial resources and access to care, driven by income, marital status, and healthcare-related variables.

Component 3: Physical & Mental Health Burden

● Dominated by physical health, mental health, and functional limitations, indicating overall health strain. Component 4: Population & Geographic Context

● Reflects environmental and contextual influences, including race and metropolitan status. Component 5: Individual Identity Profile

● Includes relatively fixed characteristics such as sex and veteran status.

Component 6: Chronic Health Risk

● Captures the presence of long-term conditions such as diabetes and other chronic disease indicators. The communalities suggest that most variables are well represented within this reduced structure, confirming that the PCA successfully captures the shared variance across predictors while minimizing overlap. These components were subsequently used as independent variables in a final logistic regression model, with cluster membership serving as the dependent variable. The model demonstrated strong predictive performance, achieving an overall classification accuracy of 94.8%. Notably, the model correctly classified 97.9% of individuals in the low-risk (non-binge) cluster and 74.6% of individuals in the high-risk (binge) cluster. This represents a substantial improvement in identifying high-risk individuals compared to the initial regression model, which struggled with class imbalance. Overall, the PCA results not only reduced dimensionality but also revealed meaningful underlying structures in the data, enabling a more stable and interpretable final model. The strong classification performance further supports the relevance of these components in explaining patterns of alcohol-related risk behavior. ***Conclusion***

This study provides a comprehensive empirical assessment of the determinants of binge drinking in the United States by integrating multiple analytical approaches within a unified framework. Beginning with a traditional logistic regression model, the analysis identified several statistically significant predictors, including sex, income, employment status, and various health-related factors. However, the initial model was limited by class imbalance and potential multicollinearity, restricting its ability to accurately identify high-risk individuals. To overcome these limitations, the analysis incorporated clustering and dimensionality reduction techniques. The k-means clustering results revealed two distinct population segments: a larger, lower-risk group characterized by better physical and mental health outcomes, and a smaller, higher-risk group exhibiting greater health burdens and functional limitations. These findings underscore the importance of heterogeneity in understanding alcohol-related behaviors, as individuals do not respond uniformly to the same risk factors.

The application of Principal Component Analysis further refined the model by identifying six underlying dimensions within the data: Life Stage & Household Stability, Economic & Insurance Status, Physical & Mental Health Burden, Population & Geographic Context, Individual Identity Profile, and Chronic Health Risk. These components captured the shared variance across variables while reducing redundancy, enabling a more efficient and interpretable modeling approach.The final logistic regression model, utilizing PCA-derived factor scores, demonstrated substantial improvement in predictive performance, achieving 94.8% overall accuracy and significantly enhancing the identification of binge drinkers compared to the initial model. The results confirm that binge drinking behavior is not driven by a single factor but rather emerges from the interaction of demographic characteristics, health conditions, and socioeconomic circumstances. In particular, gender remains a dominant predictor, while individuals with chronic health conditions or physical limitations are less likely to engage in binge drinking, likely due to medical constraints or lifestyle adjustments.

Overall, this study successfully addresses the central research objective by identifying the most influential determinants of binge drinking at the individual level. The findings support the hypotheses outlined in Appendix B and highlight the value of combining statistical modeling with clustering and dimensionality reduction techniques. From a policy perspective, these results suggest that effective interventions should be multidimensional, targeting not only behavioral risk factors but also broader socioeconomic and health-related conditions. By providing a more structured and data-driven understanding of high-risk populations, this research contributes to the development of more precise and impactful public health strategies aimed at reducing binge drinking and its associated harms.

BIBLIOGRAPHY

Centers for Disease Control and Prevention (CDC). (2025). *Behavioral Risk Factor Surveillance System Survey Data: 2024* (Combined Landline and Cellular Telephone Survey, Multiple Questionnaire Versions). U.S. Department of Health and Human Service, Centers for Disease Control and Prevention.

https://www.cdc.gov/brfss/annual_data/annual_2024.html

Fairman, B. J., Simons-Morton, B. G., Haynie, D. L., Liu, D., Goldstein, R. B., Hingson, R. W., & Gilman, S. E. (2019). *State alcohol policies, taxes, and availability as predictors of adolescent binge drinking trajectories into early adulthood*. Addiction, 114(7), 1173–1182. https://doi.org/10.1111/add.14600

Greenfield, T. K., Cook, W. K., Karriker-Jaffe, K. J., Patterson, D., Kerr, W. C., Xuan, Z., & Naimi, T. S. (2019). *The relationship between the U.S. state alcohol policy environment and individuals' experience of secondhand ef ects: Alcohol harms due to others' drinking*. Alcoholism: Clinical and Experimental Research, 43(8), 1764–1773. https://doi.org/10.1111/acer.14105

Kerr, W. C. (2010). *Categorizing U.S. state drinking practices and consumption trends*. International Journal of Environmental Research and Public Health, 7(1), 269–283.https://doi.org/10.3390/ijerph7010269

Khamis, A. A., Salleh, S. Z., Karim, M. S. A., Mohd Rom, N. A., Janasekaran, S., Idris, A., & Abd Rashid, R. B. (2022). *Alcohol consumption patterns: A systematic review of demographic and sociocultural influencing* *factors*. International Journal of Environmental Research and Public Health, 19(13), 7869. https://doi.org/10.3390/ijerph19137869

Lee, M. R., Sankar, V., Hammer, A., Kennedy, W. G., Barb, J. J., McQueen, P. G., & Leggio, L. (2019). *Using machine learning to classify individuals with alcohol use disorder based on treatment seeking status*. EClinicalMedicine, 12, 70–78. https://doi.org/10.1016/j.eclinm.2019.05.008

Schmits, E., & Glowacz, F. (2022). *Changes in alcohol use during the COVID-19 pandemic: Impact of the lockdown conditions and mental health factors*. International Journal of Mental Health and Addiction, 20(2), 1147–1158. https://doi.org/10.1007/s11469-020-00432-8

Schulte, M. T., Ramo, D., & Brown, S. A. (2009). *Gender dif erences in factors influencing alcohol use and drinking progression among adolescents*. Clinical Psychology Review, 29(6), 535–547.

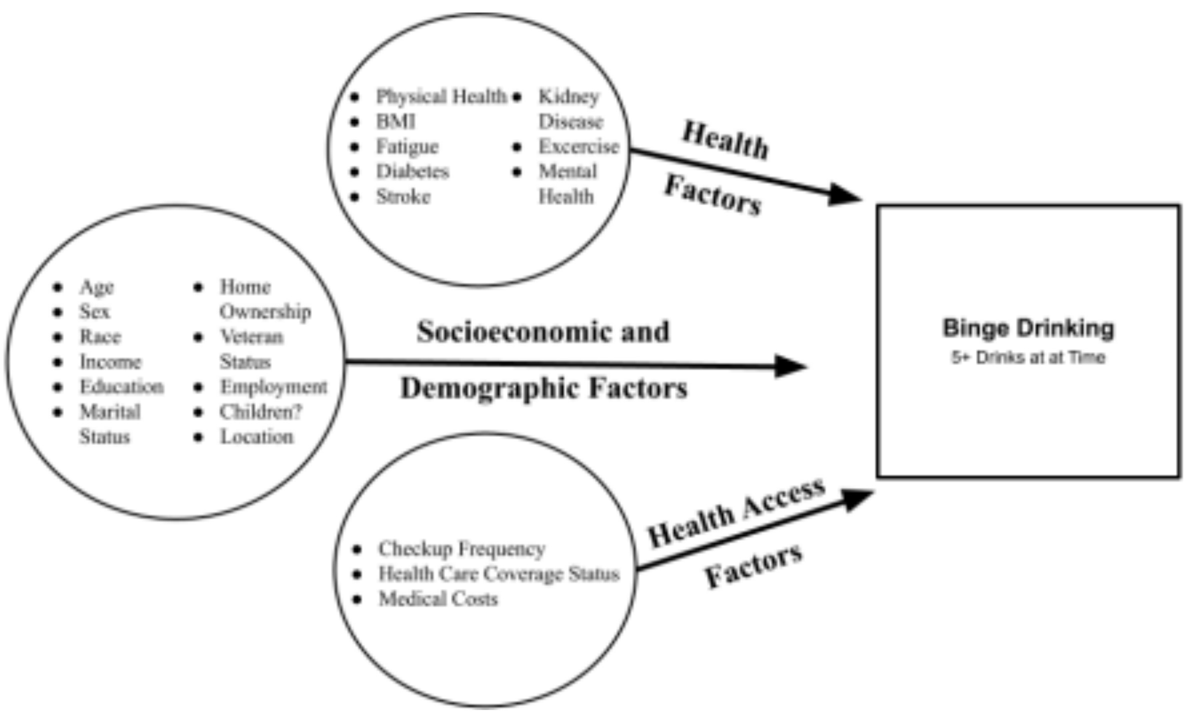
https://doi.org/10.1016/j.cpr.2009.06.003

Trangenstein, P. J., Subbaraman, M. S., Greenfield, T. K., Mulia, N., Kerr, W. C., & Karriker-Jaffe, K. J. (2020). *Association between state-level alcohol availability and taxation policies on the prevalence of alcohol-related harms to persons other than the drinker in the USA, 2000–2015*. Drug and Alcohol Review, 39(3), 255–266. https://doi.org/10.1111/dar.13041

Wagenaar, A. C., Tobler, A. L., & Komro, K. A. (2010). *Ef ects of alcohol tax and price policies on morbidity and mortality: A systematic review*. American Journal of Public Health, 100(11), 2270–2278. https://doi.org/10.2105/AJPH.2009.186007

APPENDICES

<u>Appendix A</u>



Appendix B

H1:	An individual aged 25 to 34 is more likely to binge drink.
H2:	An individual that is a man is more likely to binge drink.
H3:	An individual with an income of less than $50,000 is more likely to binge drink.
H4:	An individual that did not graduate high school is more likely to binge drink.
H5:	An individual that is not married is more likely to binge drink.
H6:	An individual that rents opposed to owning a home is more likely to binge drink.
H7:	An individual that is a veteran is more likely to binge drink.
H8:	An individual that is unemployed is more likely to binge drink.
H9:	An individual who has no children in the household is more likely to drink.
H10:	An individual who lives in a metropolitan area is more likely to binge drink.
H11:	An individual that has more than 10 days where their physical health is not good is more likely to binge drink.
H12:	An individual that is overweight is more likely to binge drink.
H13:	An individual that has difficulty walking or climbing the stairs is more likely to binge drink.
H14:	An individual that has been told they have kidney disease is more likely to binge drink.
H15:	An individual that has never had diabetes is more likely to binge drink.
H16:	An individual that has been diagnosed with a stroke is more likely to binge drink.
H17:	An individual that is a current smoker is more likely to binge drink.
H18:	An individual that spends no time doing a physical activity or exercise in the last 30 days is more likely to binge drink.
H19:	An individual that has had more than 10 days of not good mental health is more likely to binge drink.
H20:	An individual that has poor general health is more likely to binge drink.
H21:	An individual that has been more than 5 years since then routine checkup is more likely to binge drink.
H22:	An individual that does not have private healthcare is more likely to binge drink.
H23:	An individual that has only one health care provider is more likely to binge drink.
H24:	An individual that cannot afford to see a doctor is more likely to binge drink.
H25:	An individual that is white only is more likely to binge drink.

USING ANALYTICS TO IDENTIFY THE KEY FACTORS INFLUENCING JUVENILE DELINQUENCY

Lee Mara, Siena University
Kara Stallmer, Siena University
Ty Strong, Siena University
Dr. Manimoy Paul - Faculty Mentor

ABSTRACT

This study examines the primary determinants associated with juvenile crime across all 50 U.S. states using data sourced from the Office of Juvenile Justice and Delinquency Prevention (OJJDP) and other related sources. Analyzing 20 demographic, social, familial, economic, and health-related factors, we identified the most significant predictors of juvenile delinquency. The analysis used a backward stepwise multiple linear regression model, starting with all potential predictors and iteratively removing the least significant variables. The final model explains a substantial proportion of the variation in juvenile crime rates ($R^2 = 0.876$), indicating strong explanatory power. Several variables emerged as statistically significant predictors, including larceny arrest rates, drug arrest rates, assault arrest rates, divorce rates, and the percentage of Hispanic youth. In particular, larceny and drug-related arrest rates show the strongest relationships with overall juvenile crime, while the other 3 factors demonstrate meaningful associations. In contrast, we also performed a cluster analysis, which groups the states with similar characteristics. The 3 variables that had good clusters were the median household income, teen birth rate, and child poverty rate. These variables revealed underlying structures in our data, specifically in variables that were not significant when linear regression was formed. These findings highlight the importance of both behavioral indicators (such as prior arrest patterns) and broader social factors (such as family structure and demographics) in understanding juvenile delinquency. The results suggest that effective policy interventions should take a complex approach, addressing not only criminal behavior but also underlying social and economic conditions that contribute to youth involvement in the justice system.

INTRODUCTION

Over the last decade, there has been a notable downward trend in youth offending rates. Still, juvenile delinquency continues to be a concern for communities, policymakers, and researchers, since early criminal activity significantly alters the life trajectories of young individuals. The consequences of delinquency expand beyond the offender, with impacts also reaching the offender's family and the surrounding community. Understanding the multidimensional factors that contribute to the onset of juvenile delinquency is essential in developing data-driven intervention and prevention strategies.

Juvenile delinquency is influenced by a complex interaction of social, economic, and psychological factors. Existing literature emphasizes the role that family stability and socioeconomic status have in relation to juvenile delinquency. These factors often compound one another, creating environments where delinquent behaviors are more likely. Given the diverse nature of youth crime, our study seeks to identify which variables have the most predictive power on juvenile delinquency by analyzing relevant state-level data. It is important to note that some states may exhibit higher crime rates due to factors such as population size, geographic location, and other underlying social and demographic differences. We will be applying statistical methods such as multiple linear regression and clustering analysis in order to uncover patterns and relationships between the key variables identified. Ultimately, our goal is to contribute to a better understanding of the key drivers of juvenile delinquency and provide insights that may inform more targeted and effective policy responses.

LITERATURE REVIEW

Before forming our own hypothesis, we reviewed existing literature from past studies found in scholarly articles, journals, and websites. All the sources we examined focus on key factors contributing to juvenile delinquency across different social contexts. Our findings help guide this study by identifying the most important factors included in our analysis.

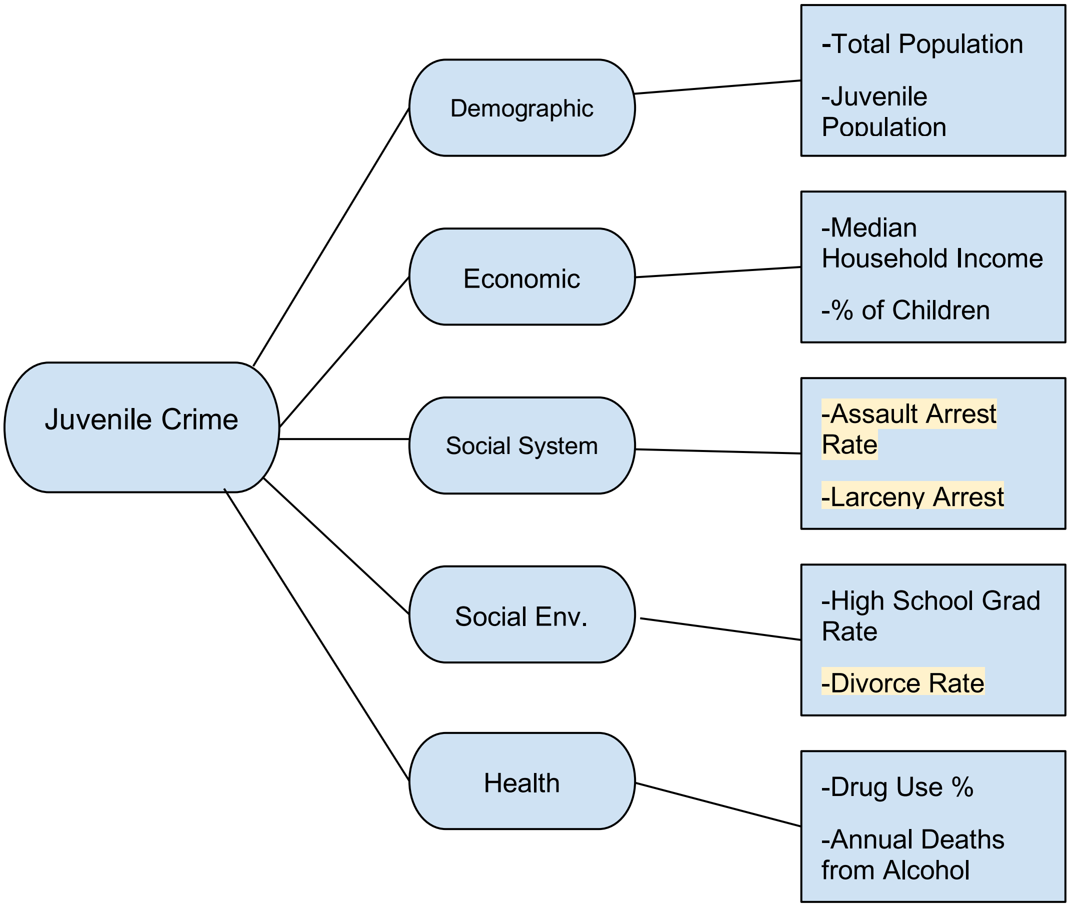
One area we researched extensively was the social-environmental factors that contribute to juvenile delinquency. We found an article from *Old Dominion University* highlighting the effect divorce rate has on juvenile crime rates. Conn conducted a macro-level analysis examining the relationship between divorce rates and juvenile violent crime and drug abuse across 135 cities and counties in Virginia, drawing on three theoretical frameworks: social control, differential association, and social disorganization. Using FBI arrest statistics, Conn utilized an OLS regression model to test whether divorce rates across the cities were a significant predictor of violent crime: "broken homes may facilitate crime by decreasing community networks of informal social control" (Conn, 2006). The author argues that family disruption at the community level weakens informal social controls and increases adolescent delinquency.

Another area we focused on closely was the demographic factors that contribute to juvenile delinquency. Race was a demographic factor that we wanted to include in our study. We found an article from the *U.S. Department of Justice* on delinquency cases involving Hispanic youth in 2013, revealing significant racial disparities at multiple decision points within the juvenile justice system. Hockenberry discovered a rapid upward trend of juvenile delinquency associated with the Hispanic youth from 1990 to 2013: "Hispanic youth accounted for 23% of the juvenile population ages 10 to 17 in 2013. Between 1990 and 2013, the Hispanic youth population increased 134%, from 3.2 million in 1990 to 7.5 million in 2013. As a result, the Hispanic proportion of the youth population grew 11 percentage points between 1990 and 2013" (Hockenberry et al., 2016). Despite the increase in the Hispanic population, a regression analysis revealed that Hispanic youth are referred to juvenile court at a rate 20% higher than white youth, and once adjudicated, are 30% more likely to be ordered to out-of-home placement. Therefore, we can assume that there are systemic issues of differential processing.

During our research, we have gained an understanding of the social system factors that may have the most influence on juvenile delinquency in the United States. We believed that drug-related offenses would be among the strongest predictors of juvenile delinquency. We found an article from the *Child Crime Prevention and Safety Center* underlining the relationship between substance abuse and juvenile delinquency as one of the most well-documented behavioral pathways into the justice system: "Studies have shown that 80 percent of minors in state juvenile justice systems were under the influence of drugs or alcohol when committing their crimes, test positive for drugs, were arrested for committing an alcohol or drug offense, admitted to having substance abuse or addiction problems or shared some combination of these characteristics (*Juvenile Crime and Substance Abuse*, 2026). We found a strong link between drug-related offenses and larceny arrests. Many of the drug-related offenders were committing theft to fund drug purchases. This connection between drug use and property crime reinforces our regression findings, where both drug arrest rates and larceny arrest rates emerged as the strongest predictors of overall juvenile delinquency.

While we have created a strong base of much of our research on the most prevalent factors contributing to juvenile delinquency, we decided to investigate further into other variables that we thought could play a part in the story. We looked into economic and social-environment factors like median house income and the percentage of children living in poverty, teen birth rates per 1,000, and child abuse rates per 1,000. We also dissected health behavior factors of annual deaths from alcohol consumption and the percentage of alcohol related deaths under the age of 21. By examining these additional economic, health behavior, and social-environment variables, we aimed to gain a more comprehensive understanding of the broader social landscape that may influence the juvenile delinquency rate.

We have inserted a flow chart illustrating the variables we have chosen to use and placed them into five key categories: Demographic, Economic, Social System, Social Environment, and Health Behavior factors. We highlighted the variables that we found to be the most significant to the juvenile crime rate in the United States, based on our regression analysis.



Hypotheses

H1: States with a larger total population are more likely to have higher juvenile crime rates.

H2: States with a larger juvenile population are more likely to have higher juvenile crime rates.

H3: States with a higher proportion of White youth are more likely to have higher juvenile crime rates.

H4: States with a higher proportion of Black youth are more likely to have higher juvenile crime rates.

H5: States with a higher proportion of Hispanic youth are more likely to have higher juvenile crime rates.

H6: States with lower median household income are more likely to have higher juvenile crime rates.

H7: States with a higher percentage of children living in poverty are more likely to have higher juvenile crime rates.

H8: States with higher assault arrest rates are more likely to have higher juvenile crime rates.

H9: States with higher larceny arrest rates are more likely to have higher juvenile crime rates.

H10: States with higher drug arrest rates are more likely to have higher juvenile crime rates.

H11: States with higher weapons arrest rates are more likely to have higher juvenile crime rates.

H12: States with higher detention rates are more likely to have higher juvenile crime rates.

H13: States with higher commitment rates are more likely to have higher juvenile crime rates.

H14: States with lower high school graduation rates are more likely to have higher juvenile crime rates.

H15: States with higher divorce rates are more likely to have higher juvenile crime rates.

H16: States with higher teenage birth rates per 1,000 are more likely to have higher juvenile crime rates.

H17: States with higher child abuse rates per 1,000 are more likely to have higher juvenile crime rates.

H18: States with higher drug use percentages are more likely to have higher juvenile crime rates.

H19: States with higher annual deaths from alcohol consumption are more likely to have higher juvenile crime rates.

H20: States with a higher percentage of alcohol-related deaths under age 21 are more likely to have higher juvenile crime rates.

DATA

To accurately determine the factors that influence juvenile crime the most, it's important to draw data from the same year. For this research, we analyzed juvenile justice profile data from all 50 states for the year 2020 from the Office of Juvenile Justice and Delinquency Prevention (OJJDP). Washington DC is not included in this data. Our dependent variable in this case analysis is the juvenile crime rate per 100,000. There are 13 independent variables identified from the public data website directly, and after reading articles, we pinpointed 7 other variables that needed to be included in the data – divorce rate, teen birth rate, child abuse rate, median income household numbers, high school graduation rate, drug use percentage, annual deaths from alcohol consumption, and percentage of alcohol-related deaths under the age of 21. We concluded that all of these factors may play a big part in why juveniles decide to commit a crime.

Explanation of Variables

Label	Variable Name	Variable Description
TOTAL_POP	Total population	Total number of people in each state based on 2020 U.S. Census Data.
JUV_POP	Total juvenile population	Total number of individuals under 18 in each state, calculated using Census Data and age distribution.

W_YOUTH	Percentage of White youth	Percentage of youth population (under 18) identified as White (non-Hispanic).
B_YOUTH	Percentage of Black youth	Percentage of the youth population (under 18) identified as Black.
HIS_YOUTH	Percentage of Hispanic youth	Percentage of the youth population (under 18) identified as Hispanic.
CHILD_POV	Child poverty rate	Percentage of youth living in poverty within each state.
AA_RATE	Assault arrest rate	The number of juvenile arrests for assault per 100,000 youth (ages 10-17) measures how frequently youth are arrested for violent offenses.
LA_RATE	Larceny arrest rate	The number of juvenile arrests for theft per 100,000 youth (ages 10-17) reflects the prevalence of property-related offenses among juveniles.
DA_RATE	Drug arrest rate	The number of juvenile arrests for drug-related offenses per 100,000 youth (ages 10-17) captures youth involvement in drug habits.
WEAP_RATE	Weapon arrest rate	The number of juvenile arrests involving weapons per 100,000 youth (ages 10-17) indicates the extent of weapon-related offenses among youth.
DET_RATE	Detention rate	The number of juveniles held in detention per 100,000 youth (ages 10-17) refers to contemporary confinement before trial.
COMMIT_RATE	Commitment rate	The number of juveniles committed to correctional or residential facilities per 100,000 (ages 10-17) shows the youth's involvement in the system after adjudication.
DIV_RATE	Divorce rate	The number of divorces per 1,000 married women in each state.
TB_RATE	Teenage birth rate	The number of teenage births per 1,000 by state.
CHAB_RATE	Child abuse rate	The number of children abused per 1,000 by the state.
HIGH_RATE	High school graduation rate	The High school graduation rate by each state.
MH_INCOME	Median household income	The median household income for each state.
PER_DRUG	Drug use percentage	The State-Level annual estimates of drug use by youth (ages 12-17).

ANN_ALCD	Annual deaths from alcohol consumption	The number of alcohol-related deaths by state.
ALC_U21	Percentage of alcohol-related deaths under the age of 21	The rate of alcohol related deaths for those under the age of 21 in each state.

Descriptive Statistics

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
TOTAL_POP	50	38961372	576851	39538223	6593283.60	7448582.289	5.548E+13
Juvenile Crime per state per 100,000 (Y)	50	1167	87	1254	482.06	273.415	74755.976
JUV_POP	50	8778417	117683	8896100	1442956.42	1700313.992	2.891E+12
W_YOUTH	50	70.8	20.3	91.1	60.976	17.0937	292.194
B_YOUTH	50	41.3	1.6	42.9	13.066	10.1720	103.470
HIS_YOUTH	50	58.8	2.8	61.6	18.346	13.1489	172.894
CHILD_POV	50	15.0	7.1	22.1	15.212	4.1635	17.335
AA_RATE	50	200	12	212	84.30	49.368	2437.235
LA_RATE	50	786	21	807	327.14	164.781	27152.898
DA_RATE	50	880	21	901	303.28	197.636	39059.798
WEAP_RATE	50	158	2	160	49.14	31.314	980.572
DET_RATE	50	115	0	115	42.92	24.195	585.422
COMMIT_RATE	50	232	9	241	70.28	46.916	2201.144
DIV_RATE	50	27.2	4.7	31.9	14.242	4.2048	17.680
TB_RATE	50	9.8	2.2	12.0	6.186	2.3888	5.707
CHAB_RATE	50	21.3	1.6	22.9	9.642	5.0989	25.999
HIGH_RATE	50	17.5	73.9	91.4	85.324	3.9209	15.373
MHI_RATE	50	58100	52800	110900	80085.20	13701.207	187723074.45
PER_DRUG	50	16.40%	11.40%	27.80%	17.5912%	4.24612%	18.030
ANN_ALCD	50	19311	432	19743	3548.92	3684.232	13573566.238
ALC_U21	50	2.38%	1.19%	3.57%	2.2102%	0.58363%	.341
Valid N (listwise)	50						

There is a substantial variation across the states in many variables, including population and crime rate-related measures. The average crime rate is about 482 juvenile individuals who commit a crime out of 100,000. However, this number widely ranges from 87 to 1,254, showing large differences per state. On average, white youth make up about 61% of the population, black youth make up about 13%, and Hispanic youth make up around 18%. These percentages do not add up to 100% because we did not include all races, just the 3 most prominent. Further, the larceny and drug arrest rates seem to be the highest compared to assault and weapon rates. Although the means are much higher, it doesn't mean the other two variables are insignificant. Then, looking at the social and economic factors, divorce rates and child abuse rates vary significantly across the 50 states. High school graduation is around 85% and does not differ significantly. Median household income also shows a large variation, which indicates the economic differences between states. Overall, the descriptive statistics reveal significant variation across states in juvenile crime rates, demographic composition, and social conditions, suggesting that multiple factors may contribute to differences in juvenile delinquency.

METHODOLOGY AND ANALYSIS

Model 1: Linear Regression

Dependent Variable:	Reg - Step 1		Reg - Step 2		Reg - Step 3		Reg - Step 4		Reg - Step 5	
Amount of juvenile crime cases	Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value
TOTAL_POP	-1.57E-05	0.578	-2.30E-05	0.32	-1.07E-05	0.377				
JUV_POP	2.63E-05	0.736	3.97E-05	0.524						
W_YOUTH	-0.133	0.957								
B_YOUTH	-2.358	0.453	-1.531	0.426	-1.464	0.429				
HIS_YOUTH	-4.36	0.143	-4.113	0.006	-4.207	0.004	-4.056	0.003	-3.098	0.011
CHILD_POV	-7.105	0.4	-4.204	0.432	-4.235	0.396				
AA_RATE	1.044	0.05	1.071	0.009	1.005	0.008	0.875	0.011	0.746	0.026
LA_RATE	0.513	<.001	0.523	<.001	0.533	<.001	0.542	<.001	0.546	<.001
DA_RATE	0.854	<.001	0.892	<.001	0.895	<.001	0.934	<.001	0.904	<.001
WEAP_RATE	0.297	0.721								
DET_RATE	-0.432	0.662	-0.262	0.72						
COMMIT_RATE	0.054	0.915								
DIV_RATE	-8.482	0.133	-8.373	0.044	-8.017	0.042	-8.438	0.029	-7.709	0.044
TB_RATE	5.51	0.767								
CHAB_RATE	4.483	0.265	4.931	0.162	4.743	0.168	3.661	0.246		
HIGH_RATE	-0.455	0.947								
MHI_RATE	-0.001	0.768								
PER_DRUG	-9.865	0.16	-9.997	0.041	-9.407	0.04	-6.091	0.124		
ANN_ALCD	0.028	0.4	0.036	0.189	0.03	0.229	0.007	0.15		
ALC_U21	2.206	0.964								

- R^2 = 0.876

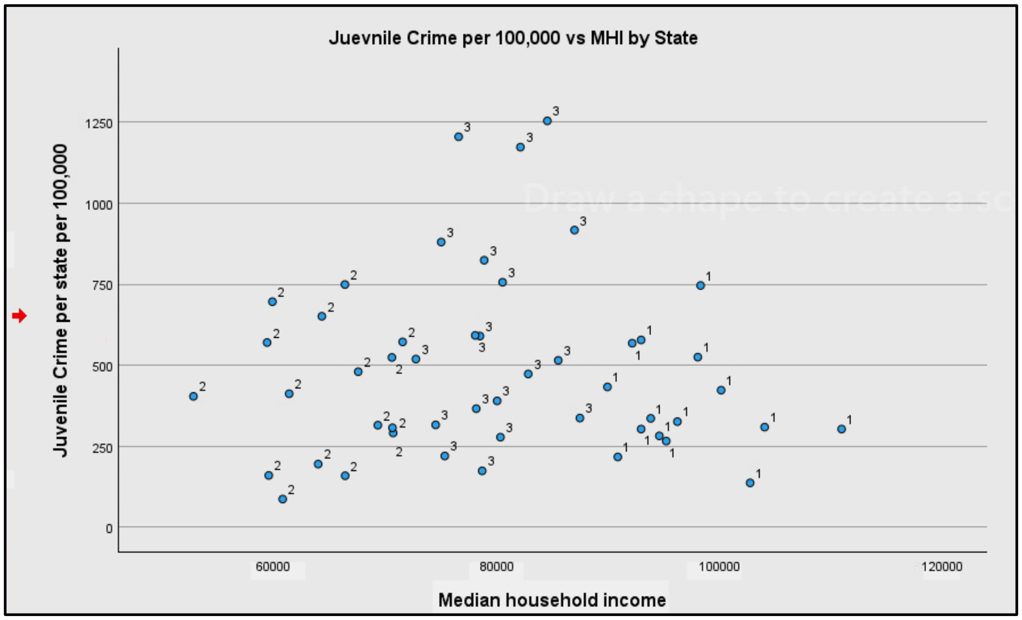
The first model we decided to use was linear regression. We did this in 5 steps. We first ran the regression on all of the 20 independent variables, and then we weeded out any of the variables with a p-value over 0.7. We then ran the regression again with 13 variables. After running the regression, we eliminated more variables with a p-value over 0.5. We repeated this process until we found all of the variables with a p-value of 0.1 or lower. Out of the 20 independent variables, 5 of the variables showed up as significant, such as: Hispanic youth, assault arrest rate, larceny rate, drug arrest rate, and divorce rate. We found that these 5 factors are most likely to convince juveniles to commit a crime.

Looking at the R^2, it is clear that our model fits really well, predicting almost 88% of cases.

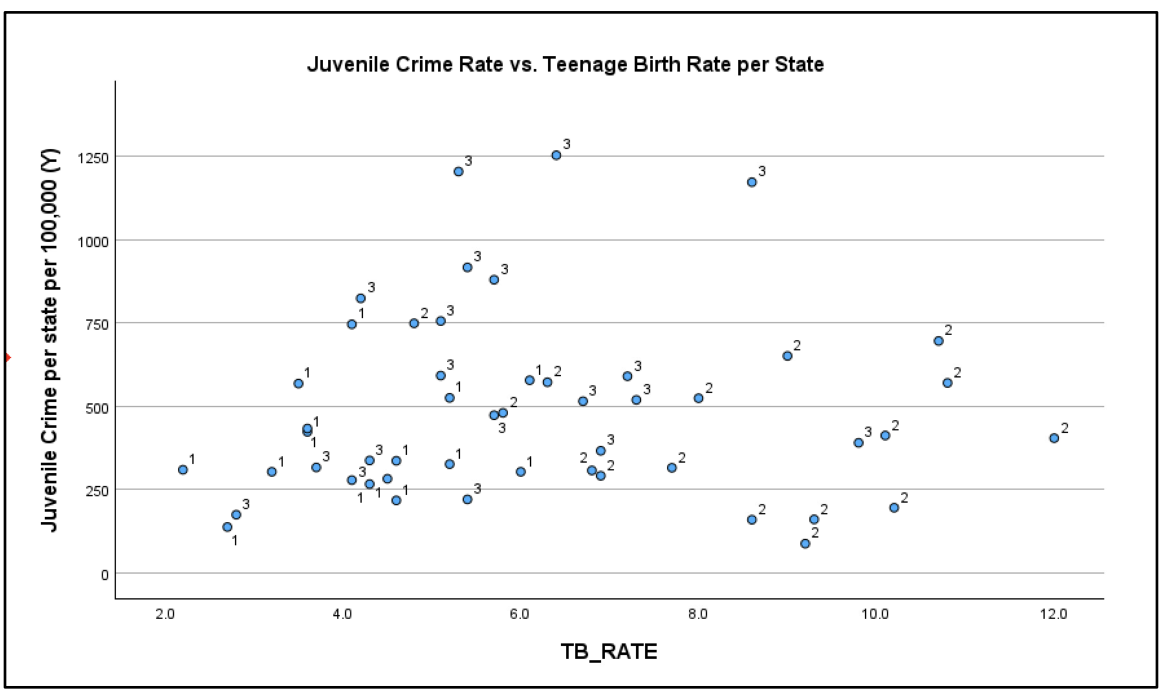
However, before running the linear regression, we believed that the child poverty rates and teenage birth rates would be significant factors as well. With the p-values of these two variables being large, we were shocked when we realized they were not highly related to juvenile crime.

Model 2: KNN Clustering

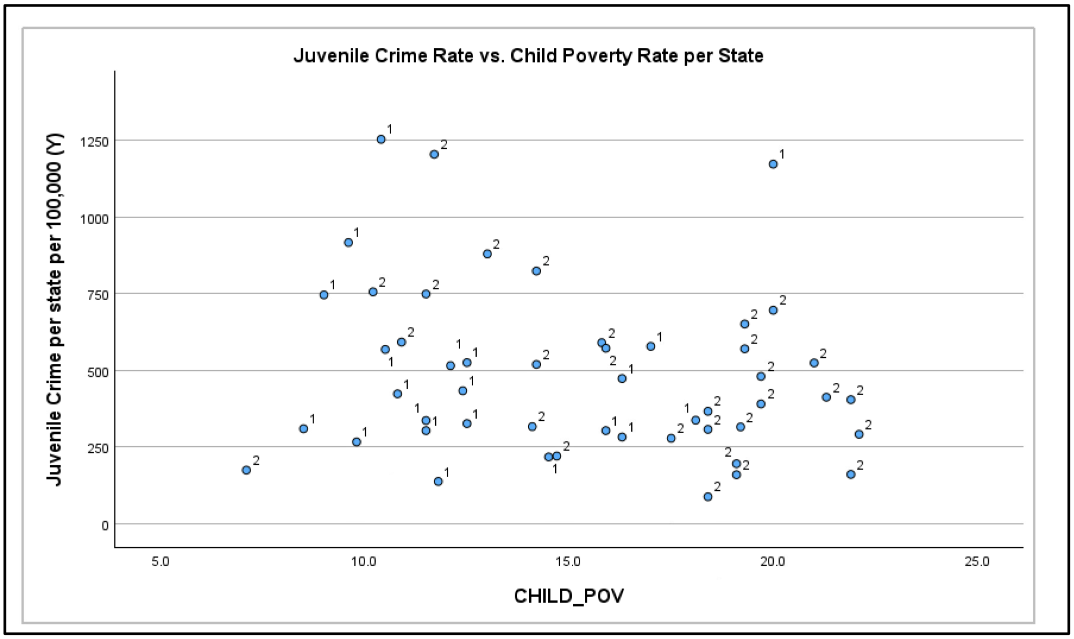
The KNN cluster analysis explores patterns across the United States using clustering techniques applied to three key social and economic variables: median household income, teenage birth rate, and child poverty rate. Clustering allows us to group the states with similar characteristics, revealing underlying structures in the data that may not be visible through simple comparisons. We decided to focus on these three variables due to their stronger ability to form meaningful clusters compared to the other variables we tested. In the linear regression model above, we found 5 of the variables significant. None of those 5 variables created strong clustering models. Since the beginning, we all thought that child poverty rates, teenage birth rates, and median household income would be variables that influence juvenile delinquency. Our approach provides insight into how combinations of social and economic factors lead to state differences.



The graph above shows the juvenile crime rate per state per 100,000 (y-axis) vs the median household income (x-axis), with numbers 1, 2, and 3 representing clusters. Based on the scatterplot, one can assume that higher-income states tend to have lower crime rates, while lower-income states are associated with moderate juvenile crime rates. The middle clustering is labeled "3," and represents middle-income states that display the highest spikes of juvenile crime rates and the most variation. Minors living in higher-income households usually have more access to better education and programs, better neighborhood safety and policy, and resources for supervision, which could explain Cluster 1. Despite the conclusion for Cluster 1, one would assume that lower-income households would be more likely to have higher juvenile crime rates, but that is not the case. Lower-income states tend to have less urban density, which could decrease the juvenile crime rate by a good margin.



The graph above shows the correlation of our dependent variable, juvenile crime per state per 100,000, vs the teen birth rate per state. We separated teen birth rate into 3 separate clusters, the cluster labeled 1 group in the bottom left of the graph, which includes states that have a low teen birth rate in their state and a low amount of juvenile crime. This cluster suggests that fewer social environmental challenges for the youth may result in less crime committed. Cluster 2 is grouped in the bottom right corner of the graph and includes states that have a low amount of juvenile crime and a high teen birth rate within that state. This is an unusual cluster, as higher teen birth rates are often associated with social environmental challenges such as lower income levels. This cluster suggests that factors like community support systems, education, and policing may help lower the crime rate in these areas. Cluster 3 is centered towards the upward middle of the graph and represents states that have moderate teen birth rates and higher juvenile crime rates. This indicates that this teen birth rate alone does not fully explain the juvenile crime levels in each state; other factors play a larger role.



The graph above shows the juvenile crime rate per state per 100,000 (y-axis) vs child poverty rate (x-axis), with numbers 1 and 2 representing clusters. Based on the scatterplot, it is clear that there is a weak relationship between the child poverty rate and juvenile crime per state. Many of the states grouped in cluster 1 have the highest juvenile crime rate, yet the lowest child poverty rate. Likewise, the states grouped in cluster 2 that have the highest child poverty rates have some of the lowest juvenile crime rates. The clusters show that juvenile crime is not solely driven by poverty, as states with similar poverty levels can have very different crime rates.

RESULTS

Using an alpha level of 1%, we concluded that only some of our hypotheses were supported by the data from our analysis. The results of our analysis show significant differences between the findings from our linear regression model and the clustering models. The linear regression identified five statistically significant predictors of juvenile crime rates: Hispanic youth population, assault arrest rate, larceny rate, drug arrest rate, and divorce rate. It also produced a strong model fit ($R^2 \approx 0.88$), suggesting that these variables collectively explain a large portion of the variation in juvenile crime across the United States. However, variables we initially expected to be the most influential, such as child poverty rate, teenage birth rate, and median household income, were not statistically significant when we ran the regression model. In contrast, the clustering models proved more subtle and less uniform patterns. When exploring median household income, states were grouped into three clusters, with higher-income states generally associated with lower juvenile crime rates, while middle-income states exhibited the greatest variability and highest crime spikes. The teenage birth rate clustering also produced three distinct groups, including one unusual cluster where states with high teen birth rates still had low juvenile crime, suggesting the likelihood of mitigating factors such as community support, education, or policy differences. Similarly, the child poverty clustering divided states into two groups but showed a weak relationship overall, with some low-poverty states experiencing high crime and some high-poverty states showing relatively low crime rates. With that being said, these results suggest that while the regression model captures strong overall statistical relationships, the clustering analysis reveals that socioeconomic variables like income, poverty, and teenage birth rates do not consistently determine juvenile crime outcomes across all states. Instead, the influence of these factors appears to differ by group, indicating that juvenile crime is shaped by a more complex combination of social, economic, and environmental conditions than any single variable can explain.

CONCLUSION

For this research study, we analyzed data from all 50 states to determine the primary demographic, economic, and social factors that influence juvenile delinquency. Evaluating 20 different variables, we developed a regression model that explains nearly 88% of the variation in state juvenile crime rates.

The regression analysis identified five statistically significant predictors, and our regression analysis shows that a young person's environment is an important factor. The statistical significance of divorce rates as a predictor of juvenile delinquency supports existing research that family stability plays a critical role in providing the structure, guidance, and oversight important to keeping youth out of the justice system. Our clustering analysis revealed the relationship between economic status and juvenile crime. While higher-income states typically have lower crime rates, poverty alone is not a definitive predictor. Several states with high child poverty rates have relatively low levels of juvenile delinquency. This can indicate that community support, education, and other support systems can be mitigators in economically disadvantaged areas.

The analysis shows that no single variable can explain existing trends in juvenile delinquency. Since delinquency is driven by a complex mix of factors that impact youth, improving intervention and prevention strategies for these at-risk youth can be targeted using data-driven insights.

REFERENCES

Conn, B. M. (2006). *A Macro-Level analysis of divorce rates and juvenile violent crime and drug abuse in Virginia*. ODU Digital Commons. https://digitalcommons.odu.edu/sociology_criminaljustice_etds/169/

Delinquency Cases involving Hispanic Youth, 2013 | Office of Justice Programs. (n.d.). Office of Justice Programs. https://www.ojp.gov/library/publications/delinquency-cases-involving-hispanic-youth-2013

Divorce rate in the U.S.: Geographic Variation, 2020. (n.d.). Bowling Green State University. https://www.bgsu.edu/ncfmr/resources/data/family-profiles/westrick-payne-divorce-rate-US-geographic-variation-fp-22-08.html

Hanson, M. (2026, January 5). *Alcohol Abuse Statistics [2026]: National + State Data -*

NCDAS. NCDAS. https://drugabusestatistics.org/alcohol-abuse-statistics/

Hockenberry, S., Puzzanchera, C., & U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention. (2016). Delinquency Cases involving Hispanic youth, 2013. https://ojjdp.ojp.gov/sites/g/files/xyckuh176/files/pubs/249915.pdf

Juvenile crime and Substance abuse. (2026). Child Crime Prevention & Safety Center. https://childsafety.losangelescriminallawyer.pro/juvenile-crime-and-substance-abuse.html

Juvenile Delinquency - causes, prevention, and the ways of rehabilitation - Penal Reform International. (2020, July 3). Penal Reform International. https://www.penalreform.org/blog/juvenile-delinquency-causes-prevention-and-the-ways-of/

NSDUH state releases. (n.d.). CBHSQ Data. https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/state-releases

Proportion of persons living in poverty by State, 2021 | Office of Juvenile Justice and Delinquency Prevention. (n.d.). Office of Juvenile Justice and Delinquency Prevention. https://ojjdp.ojp.gov/statistical-briefing-book/population/faqs/qa01403

State Juvenile Justice Profiles | Office of Juvenile Justice and Delinquency Prevention. (n.d.). Office of Juvenile Justice and Delinquency Prevention. https://ojjdp.ojp.gov/statistical-briefing-book/special_topics/state-juvenile-justice-profiles

U.S. Census Bureau. (2024). *Median household income, by state, 2017-2023*. https://ksdata.ku.edu/ksdata/ksah/income/9inc1.pdf

Youth arrest rates by State | Office of Juvenile Justice and Delinquency Prevention. (n.d.). Office of Juvenile Justice and Delinquency Prevention. https://ojjdp.ojp.gov/statistical-briefing-book/crime/faqs/qa05103

WASTED POTENTIAL: UNDERSTANDING FOOD WASTE ON COLLEGE CAMPUSES

Stephanie Davis, Siena University
Lydia Griffin, Siena University
Rachel Drew, Siena University
Dr. Cheryl Buff, Siena University – Faculty Mentor

INTRODUCTION

Food waste has been a growing sustainability issue for decades not only on college campuses, but across the planet. The causes of food waste vary between individual behaviors, dining structures, and people's awareness of available waste reduction initiatives. Our research project, "Wasted Potential: Understanding Food Waste on College Campuses", examines the psychological, situational, and structural factors that shape food waste behavior among Siena University students. More specifically, we observed how involvement (processing effort), guilt, attitudes toward food, and awareness of composting programs influence students' likelihood of wasting food with the help of insights from existing literature.

To address these questions, we designed a self-administered Qualtrics survey using an approved scale called "Involvement in the Task (Processing Effort), to measure how mindful students are during food-selection decisions. Our survey incorporated scenario-based involvement questions, measured guilt, ranked dining locations, and assessed awareness and effectiveness of Siena's composting initiatives. To administer the survey, we used convenience and snowball sampling through LinkedIn, Instagram, and Snapchat to produce a sample of Siena students, faculty, administrators, and alumni.

Our four hypotheses and post-hoc analyses were found using SPSS regressions, ANOVAs, and descriptives, highlighting that students' involvement scores were neutral, suggesting that there is minimal cognitive effort in food decisions. Neither guilt nor environmental concern strongly predicted lower wasteful behavior, and awareness of global or school composting initiatives did not significantly increase composting involvement. These results align with existing research indicating that food waste is less about personal intention and more about structural and habitual dining patterns. They found students want to waste less but default to existing routines shaped by portion sizes, dining formats, and convenience, which our results lined up with.

While some students reported greater motivation to reduce waste after taking the survey, overall awareness of our campus initiatives remained low. Our study was limited by a small sample size, incomplete responses, and a short collection window restricted to certain social platforms. Future research should strengthen in-person recruitment efforts and explore systematic or structural interventions such as flexible portioning, clearer signage, and educational programs that address the root causes of food waste on our college campus.

LITERATURE REVIEW

Behavioral & Psychological Factors

Analysis of factors influencing college students' food waste behavior and evaluation of labor education intervention: With several co-authors Wang's article examines the behavioral and demographic factors influencing food waste among Chinese college students, while also testing a solution (labor education in cafeterias) as a way to shift attitudes and perspective. Using a survey of 407 students across three universities and logistic regression analysis, the study found that food waste was higher among female students, upperclassmen, those with greater monthly consumption, and those susceptible to peer influence. A smaller group participated in a labor education activity which led to improved awareness and reduced indifference toward waste.

The Last Bite: Exploring behavioural and situational factors influencing leftover food waste in households: With several co-authors, Aloysius' article investigates the behavioral and situational factors influencing leftover food waste in Australian households. Drawing from a survey of over 1,000 households and applying structural equation modeling, the study found that motivation and ability were the strongest predictors of leftover food management. Factors like lifestyle and access to technology mattered less, while competing goals had a negative influence. Overall, better leftover management significantly reduced household food waste.

What influences students' food waste behaviour in campus canteens?: Wang, Ma, Cudjoe, Farrukh, and Bai (2023) explore the factors shaping food waste among college students in China. Using an extended Theory of Planned Behavior (TPB), the study adds moral norms, food taste, and campus food-saving climate to the traditional TPB model. Survey data from 513 students were analyzed with structural equation modeling. This article explains how household food waste is significantly different from college food waste, and points out several surprising patterns. Findings show that attitudes, subjective norms, perceived behavioral control, moral norms, and food taste all positively affect students' intentions to reduce food waste. One key finding is that attitudes are directly linked to the intention to reduce food waste, which contrasts with previous research. This shows that individuals who are more aware of and acknowledge the food waste issue are less likely to engage in wasteful behaviors. Moral norms also play a key role, as students who feel a responsibility are more driven to reduce waste, despite past studies. In the article, it states, "people who make food selections based on their taste have a higher inclination to minimize food waste." Ultimately, the existence of a food-saving environment on campus aids in strengthening this behavior by applying external pressure and narrowing in on the responsibility regarding food consumption.

Structural & Dining Environment Factors

Food Choice and Waste in University Dining Commons: This multi-campus study investigated how food type and personal factors influence food choice, consumption, and waste in all-you-care-to-eat university dining facilities. Conducted across five universities in 2019, the study used before-and-after photos of student plates alongside surveys to measure food taken, food wasted, and decision factors. Results showed that animal protein and mixed dishes occupied more plate space and were strongly associated with both hedonic appeal and pre-plating. Students who expressed more confidence in liking a food before choosing it tended to take larger portions but wasted less of it. Higher satisfaction with meals and more frequent visits to dining commons were correlated with lower food waste overall. The findings suggest that portion size, pre-plating, confidence, and satisfaction are key drivers of food waste in university dining contexts.

How to reduce college students' food waste behavior: From the perspective of college canteen catering modes: The article How to Reduce College Students' Food Waste Behavior explores how canteen catering modes affect food waste among students in China. Using a framework that combines the Theory of Planned Behavior, the Norm Activation Model, and the Attitude–Context–Behavior theory, the authors analyze 422 student surveys with structural equation modeling. They find that personal norms, attitudes, and social pressures shape students' intentions to save food, while herd mentality can encourage waste. Importantly, buffet and flexible portion-size options help students act on their intentions, whereas standard-quantity meals often lead to more waste.

Characterization of food waste in Grizzly Dining Hall at Georgia Gwinnett College: A critical step toward sustainable food waste management: The research indicates that, on average, every student produced 140 grams of food waste daily in Spring 2023, emphasizing the substantial chance for colleges to assess cafeteria waste and establish a more effective waste management system. The issue is not exclusive to a single institution, but around 3.6 million tons of food are wasted each year on the United States college and university campuses, with an average of 250 grams per individual. On a national scale, approximately 10 million kilograms of food are being thrown away annually in the United States, with the majority of universities disposing of their food in landfills.

Social, Cultural, & Habitual Influences

College students observe the waste of edible food and make recommendations: Connors' article explores how college students perceive and experience food waste, specifically focusing on their observations of food being discarded in environments such as grocery stores, restaurants, catered events, and their own homes. Using reflective essays from a non-major nutrition course, the study identified two key themes: how food waste happens and how to reduce it. Causes were identified as overproduction, misinterpretation of expiration labels, and overbuying. Students suggested solutions like reducing portion sizes, meal prepping, better education on storage and preparation, and greater use of donation and composting. Using reflective essays gives a different approach compared to surveys which we will be using.

Consumer food waste behaviour in universities: Sharing as a means of prevention.: This article examines food waste behavior among university students and staff, focusing on how sharing practices could prevent waste. Conducted at a UK university, the study used surveys, interviews, focus groups, and a social media intervention to analyze how routines and habits influence waste. Findings revealed that students often underestimate their own waste, attributing the problem to other parts of the supply chain. Economic factors, convenience, and time pressures shaped campus food waste behaviors differently from at-home habits. The study concludes that food waste is embedded in everyday routines, and overcoming cultural and structural barriers is key to prevention. Its reliability comes from the mixed-method design, though its single-site focus limits generalizability.

Institutional Programs & Intervention Strategies

The Relationship between a Campus Composting Program and Environmental Attitudes, Environmental Locus of Control, Compost Knowledge, and Compost Attitudes of College Students: This article examines the role of campus composting programs in shaping student knowledge, attitudes, and behaviors toward the environment. The researchers compared two universities: Texas State University, which has the Bobcat Blend composting program, and Farmingdale State College, which does not. A survey of 660 students measured environmental attitudes, environmental locus of control, compost knowledge, and compost attitudes. Results showed that students at Texas State had stronger environmental attitudes, higher compost knowledge, and a more internal locus of control. Compost attitudes were positively related to all three measures, suggesting that participation in composting fosters pro-environmental beliefs. The study concludes that composting programs not only reduce waste but also encourage environmental responsibility and empower students to believe their actions matter.

Evaluation of a sustainable student-led initiative on a college campus addressing food waste and food insecurity: The Farm to Fork (F2F) is linked with the Campus Kitchen (CK) model, which works to tackle both food waste and food insecurity within college campuses. CK seeks to enhance the community of food security, encourage healthy eating habits, and strengthen social connections by salvaging food that would have been thrown away, instead creating nutritious meals from these restored items, all while engaging with students and local community organizations in educational resources. Redirecting this food through reclamation provides an eco-friendly and financially viable alternative by reusing high-quality, surplus food and secondary products from farms, restaurants, and supermarkets. This shows that college campuses are affected by the linked issues of food insecurity and food waste, prompting the growing numbers of administrators, students, and researchers to address this dilemma within their specific environments.

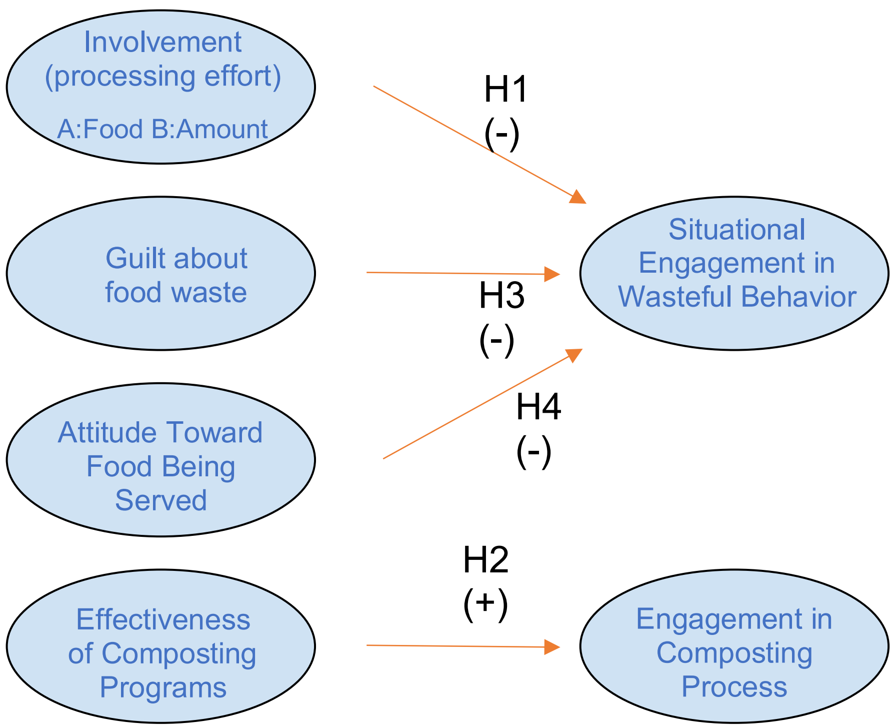
A research brief describing a logic model framework for planning a Food Recovery Network chapter at an undergraduate university.: The article by Altomare and Payton (2025) describes the development of a logic model framework for planning and sustaining a Food Recovery Network (FRN) chapter at Moravian University. FRN chapters are student-led organizations that reduce food waste and redistribute excess food to food-insecure populations. The authors outline the situation of food waste and insecurity in the Lehigh Valley and at Moravian University, then present a logic model that identifies priorities, inputs, outputs, outcomes, assumptions, and external factors.

RESEARCH DESIGN AND SAMPLING

Research Design

This research project was conducted to develop a deeper understanding of how students' environmental knowledge, beliefs, and emotions, specifically involvement and guilt, influence their food waste behaviors and composting participation. It also examines how awareness of campus composting programs and attitudes toward food interact with these factors to shape disposal choices and the amount of waste generated. To answer our research questions and test our hypotheses, we conducted an exploratory research study. We designed an electronic survey using Qualtrics, with the intent of collecting quantitative data for statistical analysis and insight generation once the survey closed. To distribute the survey, we used a combination of convenience sampling and snowball sampling. Following approval from our institution's IRB, we distributed the survey via each team member's personal profiles on the social media platforms LinkedIn, Snapchat, and Instagram. The survey was available to respondents from October 21, 2025, to November 6, 2025.

Conceptual Model



Hypothesis 1A: Students with higher involvement scores when processing their food will report lower situational engagement in wasteful behavior. The independent variable for Hypothesis 1 is involvement (processing effort). To improve clarity, Hypothesis 1 is divided into two related propositions. Hypothesis 1A states that higher involvement with the content of the food a person is eating will result in lower situational engagement in wasteful behavior. Hypothesis 1B states that higher involvement when considering the amount of food a person takes will result in lower situational engagement in wasteful behavior.

Hypothesis 1B: Students with higher involvement scores when processing the amount of food will report lower situational engagement in wasteful behavior.

Hypothesis 2: Students who are aware of composting programs at their schools are more likely to have greater involvement in the composting process. Hypothesis 2 examines the perceived effectiveness of campus composting programs as the independent variable. It states that a higher level of effectiveness, or perceived effectiveness, of a campus composting program will result in greater engagement in the composting process and environmentally friendly behavior.

Hypothesis 3: Students with a more guilt towards food waste will report lower situational engagement in wasteful behavior. Hypothesis 3 uses guilt as the independent variable. It states that a higher self-reported level of guilt toward food waste will result in lower situational engagement in wasteful behavior.

Hypothesis 4: More positive attitudes toward the food will predict lower situational self-reported waste. Hypothesis 4 uses attitude toward food as the independent variable. It states that a more positive attitude toward the food being served will result in lower situational engagement in wasteful behavior.

Scales and Validation

The primary scale used in this study is the “Involvement in the Task (Processing Effort)” scale from the *Marketing Scales Handbook* by Gordon C. Bruner II. This scale measures the cognitive effort a person believes they expend when processing a message or making a decision. In the context of this research, it was used to assess students’ involvement when deciding both what food to take (Hypothesis 1A) and how much food to take (Hypothesis 1B). The items were adapted directly from the handbook and include measures such as the extent to which participants deliberated, the time they spent thinking, the amount of attention they paid, and the extent to which they thought about the task. These items were measured using a 5-point Likert scale ranging from “Not at all” (1) to “A great deal” (5), consistent with prior applications of the scale. The Involvement (Processing Effort) scale has been validated in previous research using internal consistency reliability, typically assessed through Cronbach’s alpha, with reported values ranging from .79 to .91 (Bruner, 2023). In our study, reliability testing produced a Cronbach’s alpha of .86 for Hypothesis 1A (Content) and .913 for Hypothesis 1B (Amount), indicating strong internal consistency across both applications. However, because the scale relies on self-report measures, the accuracy of the data depends on participants’ ability to recall and reflect on their cognitive involvement during the dining experience, which may introduce some limitations to the precision of the results.

Survey Design and Implementation

The survey was constructed and organized systematically to evaluate each hypothesis, structured through utilizing the block feature in Qualtrics. This structure allowed us to isolate measures by hypothesis and minimize unintended priming effects across constructs. The survey consisted of a consent section, four hypothesis-specific blocks, and a final block containing demographic questions and a question assessing future behavioral intention. In total, the survey contained 20 questions, not including the consent and demographics sections.

The survey began with an informed consent section outlining the purpose of the research, participation requirements, and eligibility criteria, followed by a screening question assessing how many meals participants eat on campus per week. This question served to both understand respondents’ exposure to campus dining environments and anchor their experiences before answering engagement-related questions. The first hypothesis block measures involvement (processing effort), asking participants to imagine a realistic campus dining scenario and report the extent to which they consider both the type and amount of food they take. These items were presented as interval-level 5-point Likert scale questions.

The next sections focused on psychological and behavioral factors related to food waste. One block examined guilt and current waste behavior by measuring perceived environmental impact, typical disposal habits, frequency of over-portioning, and self-reported guilt. These questions used a combination of nominal and Likert-scale formats and were intentionally placed earlier in the survey to avoid influence from later questions. Another block assessed attitudes toward food, including dining location rankings, personal preferences, and the influence of others on waste-related behaviors, incorporating both ordinal and interval-level measures.

The final hypothesis block evaluated perceptions of campus composting programs, including awareness, perceived effectiveness, likelihood of use, and self-efficacy in reducing waste. These questions were primarily measured באמצעות Likert scales, along with one nominal yes/no item. Informational statistics about food waste were also introduced at this stage to assess their effect on participants’ reported intentions. The survey concluded with demographic questions and a final Likert-scale item designed to measure whether completing the survey influenced participants’ future intentions to reduce food waste.

Sample Profile

In looking at the sample profile of our respondents, our data showed that the majority of participants were seniors, making up 55.6% of valid responses. Juniors were our second largest portion, represented with 19.4% followed by the smaller groups of freshmen with 6.9%, sophomores with 5.6%, graduate students with 5.6%, and respondents who selected “other” at 6.9%. With 72 total valid responses, this distribution indicated that our sample was heavily skewed towards older students, many of whom may be approaching graduation and have gone through 3-4 years of college dining. Understanding this class year concentration helped provide us with insight into the maturity and academic experience of our surveyed population. This demographic allows for more accurate interpretations of data, and the responses may reflect the perspectives of students who have been significantly exposed to college-level coursework, campus life, and future-oriented decision-making.

Table 1: Sample Profile Demographics, Frequencies, and Valid Percentages

Demographic Variables	Demographics	Frequencies	Valid Percentages
Class Year	Freshman	5	6.9%
	Sophomore	4	5.1%
	Junior	14	17.9%
	Senior	40	51.3%
	Graduate Student	4	5.1%
	Other	5	6.4%
	Total	72	92.3%

DATA ANALYSIS AND FINDINGS

Results are summarized in Table 2 below.

Table 2: Hypotheses, Results, and if Supported

Hypothesis	Results	Supported or Not Supported
H1A: *Students with higher involvement scores when processing their food will report lower situational engagement in wasteful behavior.*	One-way ANOVA p<.05 level F-statistic: 0.066 p-value: 0.798	Not Supported
H1B: *Students with higher involvement scores when processing the amount of food will report lower situational engagement in wasteful behavior.*	One-way ANOVA p<.05 level F-statistic: 0.280 p-value: 0.598	Not Supported
H2: *Students who are aware of composting programs at their schools are more likely to have greater involvement in the composting process.*	One-way ANOVA p<.05 level F-statistic: 0.879 p-value: 0.352	Not Supported
H3: *Students with a more guilt towards food waste will report lower situational engagement in*	One-way ANOVA p<.05 level	Not Supported

wasteful behavior.	F-statistic: 1.460 p-value: 0.231	
H4: *More positive attitudes toward the food will predict lower situational self-reported waste.*	One-way ANOVA p<.05 level F-statistic: 0.006 p-value: 0.941	Not Supported

DISCUSSION

There are numerous important takeaways from the results of this project. A key takeaway to note is that the resulting lack of significance in the hypotheses of this project indicates that the research conducted aligns with many of the concepts that have already been reported in secondary data surrounding food waste and the psychological behavior that drives it. Our data support the findings in existing literature that state that waste is a structural issue.

The results of our project indicate and confirm that psychological factors, personal beliefs, and awareness alone are not enough to alter wasteful behavior. A structural issue means that the systems in place within not only college campuses but throughout the economy and government regulations are not effective at preventing large-scale food waste in the United States. This includes institutional policies, food service models, and operational processes that make waste easier to prevent than to treat.

Another thing to note is that the participants in our survey were reported to be mostly seniors on our college campus. This is something that can be further explored, as it may be an indicator that the results of our survey show that some waste habits may already be formed, regardless of attitude and motivation. This presents the possibility that a senior on a college campus may already have formed habitual routines that become second nature as they progress in their college career, and these routines may result in wasteful behavior. In contrast, a freshman or sophomore may still be forming said habits, and may still be more conscious or forming their views about how their behavior contributes to waste on their campus and in the environment.

Overall, the results show that in order to conduct further research and implement changes within certain structures to lower the amount of food waste in the United States, the questions should surround infrastructure, practical action, and habit formation. These questions may challenge how the habits that we have formed, while they may be convenient for us, are harmful to the world that we live in. It would be beneficial to expand this study with future research analyzing how institutional systems contribute to food waste and identify specific structural changes that could reduce waste at scale, as well as the psychology of forming habits. A discussion may include what methods would be effective in breaking habits surrounding food waste. Any interventions to make a change in what we see in our society today should focus on structural change instead of trying to have an emotional appeal. This is because our results show that while an emotional appeal may work in some scenarios, breaking habits and reforming structure require a more grounded approach. Our study demonstrates that one can have an awareness and a belief in a higher “moral good” surrounding food waste, and still engage in wasteful behavior.

Post-Hoc Analysis

Following our hypothesis testing, we conducted several post-hoc analyses to explore additional patterns in students’ attitudes and behaviors related to food waste. First, we examined the difference in guilt between students aware of Siena’s food waste initiatives such as the Roots Cafe or our Natural Upcycling programs and those who were not aware. We found a statistically significant difference (t-stat: -2.813, p-value: .006). Interestingly, students who were not aware of initiatives reported significantly higher self guilt (Mean: 4.10) than those who were aware (Mean: 3.39). These findings suggest that more people feel guilty than those who are actually aware of the help/services available. On the other hand, increased awareness may reduce personal guilt as students could feel the system is handling the issue through programs (like Roots Cafe or Natural Upcycling), therefore lessening their personal burden of responsibility.

A second analysis we explored was whether class year explains feelings of guilt toward food waste. We grouped respondents into three “life stages” as follows: Freshmen, Sophomores, and Juniors (Group 1), Seniors (Group 2), and Graduate Students/Others (Group 3). We conducted an ANOVA test that confirmed academic level is a significant predictor of guilt (F-stat: 3.957, p-value: .024). The results show a clear upward trend: Group 1 had the lowest guilt (3.48), followed by seniors (3.95), then finally graduate students and other adults reported the highest levels of guilt (4.56). Overall, this progression suggests that academic maturity and life experience intensify one’s moral concern or self awareness regarding sustainability or food waste in our case specifically.

Lastly we analyzed how likely one is to change their behavior after taking our survey. Please note that in addition to all the questions that we included in our survey, we also provided facts from the USDA around food waste statistics. A One-Sample T-test test demonstrates that this awareness is a significant predictor of future effort. These results suggest that while simple awareness of a national statistic might not be the sole driver of change, the act of engaging with the survey and its information served as a successful intervention. The high overall mean of 4.01 out of 5.00 indicates that students are highly receptive to reducing waste once the topic is brought to their attention, regardless of their prior knowledge. This demonstrates that students possess a latent willingness to improve their habits. The survey itself, and others like it, or even simple awareness building can act as a prompt to shift individuals’ from passive observers to individuals with a self-reported high intent to act.

CONCLUSIONS AND RECOMMENDATIONS

Our data collection and analysis aimed to gain a deeper understanding of food waste on college campuses. We looked at four different research questions about how college students feel about food waste on campus. We found that our sample size was too narrow; we started with 115 participants and could only end up using 79 of the participants' surveys. This left 36 surveys unusable because they were not fully completed. Our first recommendation is to use different means to circulate the survey around campus. Such as using QR codes and email, in addition to going into classes and asking students to complete the survey. This recommendation should be helpful because it would increase the number of people who would have access to the survey; the more people who have access, the better, because this is a completely voluntary survey. If we had been approved to use email, we would have been able to email this survey to other marketing professionals, and students would be more inclined to participate. Another recommendation would have been to go to classes where the student's majors align more with environmental awareness. Classes or majors in which we would focus are environmental studies, actuarial sciences, psychology, and biology. Overall, we feel that in order to increase the number of surveys completed, we need to increase the means by which people can complete the survey.

When cleaning our data, we noticed that 36 individuals did not want to rank the dining halls on campus where they eat from most often to least often on the Siena campus when provided with 6 options. Most people who did not finish the survey dropped off at around question 14 out of 20, which may be because they did not like the question design. The majority of the questions at the beginning of the survey were multiple choice, which promotes participants' engagement. When the questions switched to ranking or scale questions, students stopped engaging and decided to drop out of the survey. After considering these factors, we generated our second recommendation. While we had 22 questions, including the two demographic questions, the 22 questions took up a lot of time, which is why we saw people drop off on question 14. To avoid survey fatigue, in the future, we would shorten the survey. Survey fatigue is when an individual becomes tired or overwhelmed with too many survey questions and ends up skipping them or clicking random answers to get it over with (Ghafourifard, 2024). Reducing the number of questions in the survey would increase the likelihood of the survey being fully completed and our group meeting the required 220 respondents.

People who participated in the study had two weeks to complete the survey, from October 21st, 2025, to November 6th, 2025. My team believes that if we were given more time to reach the 220 responses that were required, we would be able to complete the assignment. As for a recommendation of extending the data collection period from two weeks to four weeks to boost participation in subsequent surveys, this would provide us more time to reach participants and raise the total response rate.

The Institutional Review Board (IRB) authorized our group to disperse the survey via LinkedIn, Snapchat, Instagram, and Facebook. However, we did not use Facebook during our research. Looking back, we feel that Facebook would have been a good avenue to receive more responses, such as the Facebook Local Mothers page. Making one of our mom's posts with the link to our survey would have made us receive more responses. When submitting to the Institutional Review Board, my group should have also received an email and QR codes approved, as that would have helped us spread the survey around campus. We advise expanding the distribution methods for future surveys by distributing them via email, flyers, QR codes, and personal interaction. For instance, we could send the survey to previous teachers and other students through email. For the future, we would create an engaging flyer and put it all around campus with a QR code that people can scan and go to. In addition to the flyer going up to students all over campus with the QR code and asking if they would fill out a survey, then walk away and not know if they ever completed the fully voluntary survey. Having a human touch goes a long way with projects, especially when trying to engage a particular segment.

According to our research, structural systems and habits, rather than students' attitudes or guilt, are the main causes of food waste on college campuses. However, students have worries about sustainability; awareness by itself does not result in action. Practical system-level approaches that encourage students to make necessary and meaningful decisions when choosing their food. Campuses can better motivate students for long-lasting habit change by increasing outreach, increasing data gathering, and creating more easily available waste reduction choices.

While food waste is a habitual and structural issue, our project ends on a positive note. Students are not indifferent. They possess a latent willingness to do better and be better. By providing the "why" through awareness and the "how" through improved programs and infrastructure, Siena University and other colleges can bridge the gap between student guilt and meaningful environmental action.

REFERENCES

Aloysius, N., Ananda, J., Mitsis, A., & Pearson, D. (2025). The Last Bite: Exploring behavioural and situational factors influencing leftover food waste in households. *Food Quality and Preference*, 123, Article 105327. https://doi.org/10.1016/j.foodqual.2024.105327

Altomare, T. K., & Payton, C. (2025). A research brief describing a logic model framework for planning a Food Recovery Network chapter at an undergraduate university. *Journal of Agriculture, Food Systems, and Community Development, 14*(1), 487–501. https://doi.org/10.5304/jafscd.2024.141.027

Bruner, G. C. II. (2012). *Marketing Scales Handbook* (Vol. 9). GCBII Productions, LLC.Connors, P. L. (2024). College students observe waste of edible food, and make recommendations. *Journal of Food Composition and Analysis*, 126, Article 105833. https://doi.org/10.1016/j.jfca.2023.105833

Cisse, R. S., Aguilar, F. S., Ezenagu, S. B., Cisse, M. S., & Niang, A. (2025b). Characterization of food waste in Grizzly Dining Hall at Georgia Gwinnett College: A critical step toward sustainable food waste management. Heliyon, 11(2). https://doi.org/10.1016/j.heliyon.2025.e41750

Ghafourifard, M. (2024, January 14). Survey fatigue in questionnaire based research: The Issues and Solutions. Journal of caring sciences. https://pmc.ncbi.nlm.nih.gov/articles/PMC11833437/#:~:text=This%20type%20of%20fatigue%20is,terminate%20participation%20as%20a%20result.

Lazell, J. (2016). Consumer food waste behaviour in universities: Sharing as a means of prevention. Journal of Consumer Behaviour, 15(5), 430–439. https://doi.org/10.1002/cb.1581

Menus of Change University Research Collaborative. (2021). Food choice and waste in university dining commons—A Menus of Change University Research Collaborative study. Foods, 10(3), 577. https://doi.org/10.3390/foods10030577

OoNorasak, K., Barr, M., Pennell, M., Hinton, J., Garner, J., Kerber, C., Ritter, C., Dixon, L., Rohde, C., & Stephenson, T. (2022). Evaluation of a sustainable student-led initiative on a college campus addressing food waste and food insecurity. Journal of Agriculture, Food Systems, and Community Development, 11(4), 223–237. https://doi.org/10.5304/jafscd.2022.114.014

Waliczek, T., McFarland, A., & Holmes, M. (2016). The relationship between a campus composting program and environmental attitudes, environmental locus of control, compost knowledge, and compost attitudes of college students. HortTechnology, 26(5), 592–598. https://doi.org/10.21273/HORTTECH03320-16

Wang, A., Luo, X., Liu, X., & Sun, Y. (2024). How to reduce college students' food waste behavior: From the perspective of college canteen catering modes. *Sustainability, 16*(9), 3577. https://doi.org/10.3390/su16093577

Wang, D., Zhang, K., Lv, X., Xue, L., Yang, Z., & Li, P. (2024). Analysis of factors influencing college students' food waste behavior and evaluation of labor education intervention. *Frontiers in Public Health*, 12, 1372430. https://doi.org/10.3389/fpubh.2024.1372430

Wang, H., Ma, B., Cudjoe, D., Farrukh, M., & Bai, R. (2023). What influences students' food waste behaviour in campus canteens? *British Food Journal, 125*(2), 381–395. https://doi.org/10.1108/BFJ-10-2021-1103

The Mission of the School of Business

We empower students with the knowledge and values to lead, serve and succeed in their careers, communities and lives.

The Vision of the School of Business

To be the standard of educational excellence for a diverse learning community developing business minds for today and the future.

The Creed of the School of Business

Excellence in every interaction.

SIENA UNIVERSITY
School of Business Siena Hall 301
515 Loudon Road
Loudonville, NY 12211-1462

(518) 783-2321
Email: cbuff@siena.edu
Fax: (518) 786-5040