Exploring the Relationship Between Socioeconomic Factors and Crime in Chicago (Impact Investing Indicator)

Ya-Wei (Jeremy) Tsai

January 27, 2025

Abstract

This study examines the relationship between socioeconomic factors and crime rates across Chicago's community areas. Through exploratory data analysis (EDA) and statistical techniques, we identify key correlations and patterns that provide insights into the drivers of crime and their implications for investment and policy decisions.

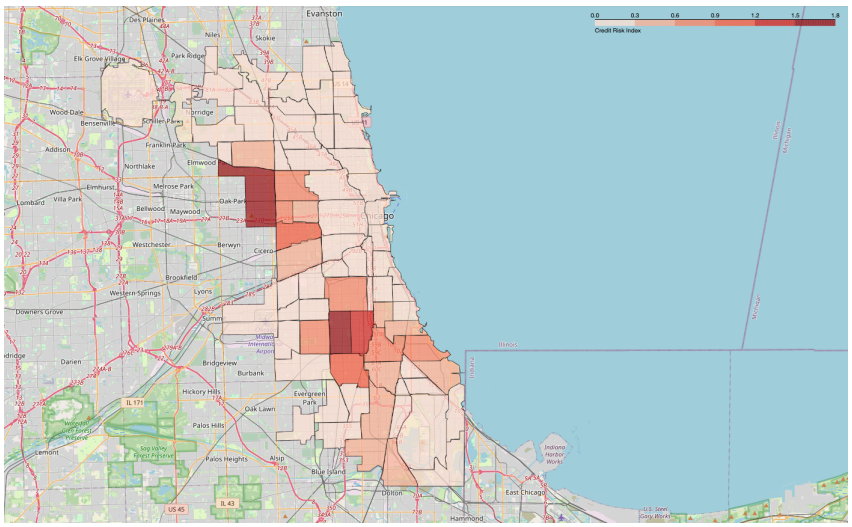

Figure 1: Credit Risk Index Distribution by Community Areas in Chicago

1 Introduction

Crime is a multifaceted issue influenced by various socioeconomic factors, including unemployment and education. In this study, we first propose a Credit Risk Indicator to evaluate community risks for investors and policymakers. Observing that this indicator correlates strongly with crime counts, we further analyze the factors associated with crime counts, uncovering potential drivers and their implications.

2 Credit Risk Indicator

To guide investment decisions, we developed a Credit Risk Indicator that combines socioeconomic and crime data. The formula is defined as:

$$\text{Credit Risk Index (CRI)} = \frac{\text{Crime Count} \times \text{Unemployment Rate (\%)}}{\text{Median Income}}$$

This indicator assumes that higher crime rates and unemployment contribute positively to credit risk, while higher median income mitigates risk.

Figure 2 visualizes the distribution of the Credit Risk Indicator across Chicago's community areas. Darker regions represent higher risk levels.

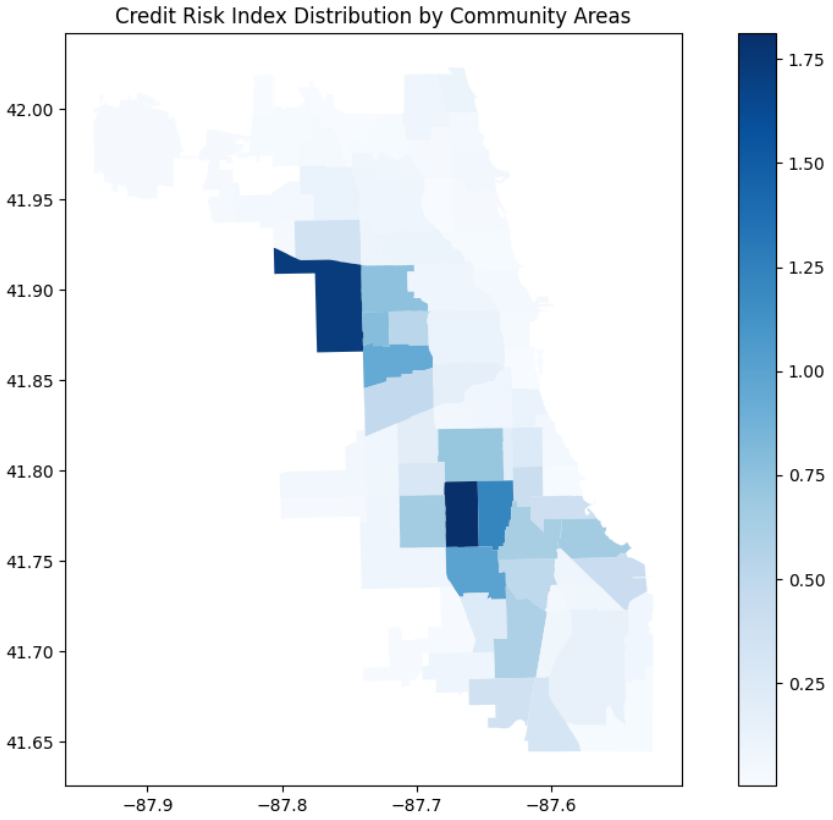

Figure 2: Credit Risk Index Distribution by Community Areas in Chicago

We found that the Credit Risk Indicator correlates strongly with crime counts ($r = 0.77$). Therefore, we focus our subsequent analysis on Crime Count to explore its relationships with other socioeconomic factors.

3 Exploratory Data Analysis (EDA)

To begin, we calculated the Crime Count Distribution and its relationship with other socioeconomic variables. Figures 3 to 4 illustrate these patterns.

3.1 Crime Count Distribution

Figure 3 visualizes the spatial distribution of crime counts across Chicago's community areas. Darker regions represent areas with higher reported crime counts.

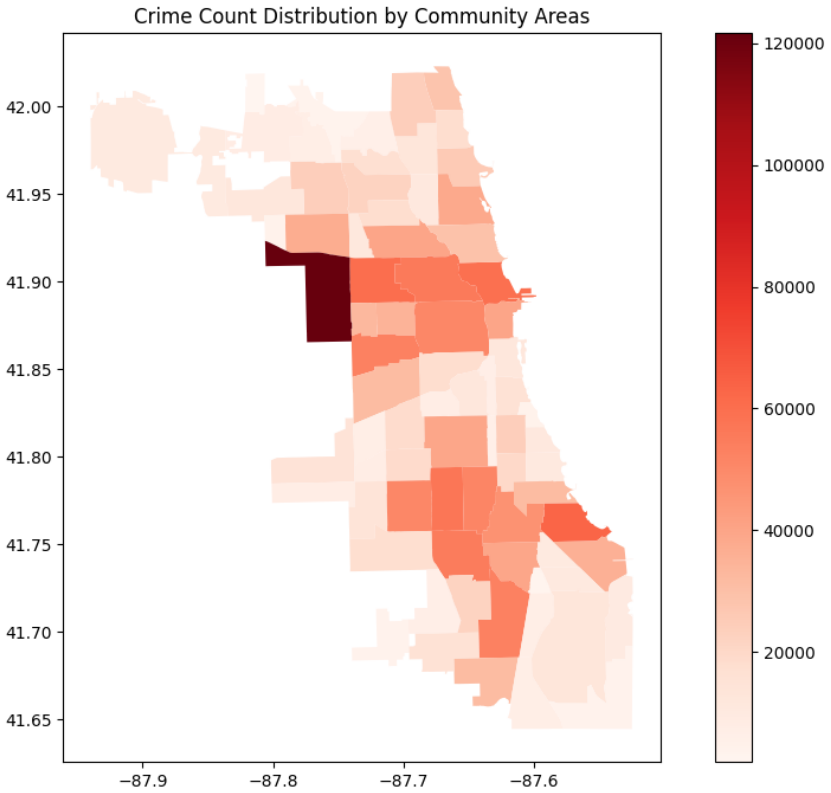

Figure 3: Crime Count Distribution by Community Areas in Chicago

3.2 Correlation Analysis

We examined the correlations between key socioeconomic indicators and crime count. The correlation matrix (Figure 4) highlights the following relationships:

- A moderate positive correlation ($r = 0.25$) between **Unemployment Rate** and **Crime Count**.
- A negligible correlation ($r = 0.06$) between **Education Level** (percent without a high school diploma) and **Crime Count**.

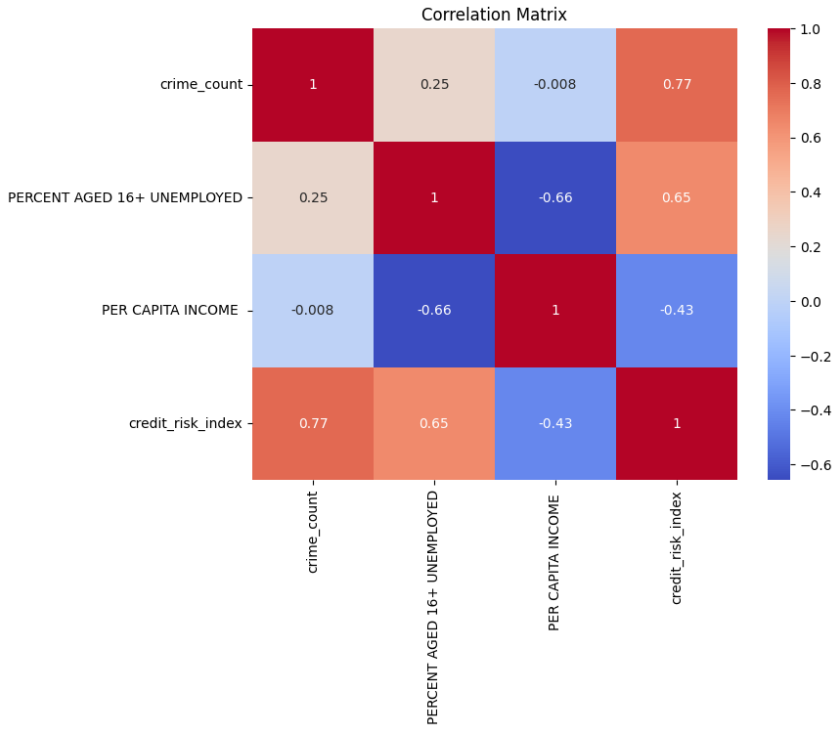

Figure 4: Correlation Matrix Between Crime Count and Socioeconomic Factors

3.3 Crime Count vs. Education Level

To further explore the weak correlation between education and crime count, we categorized community areas into three groups based on education levels:

- High Education: Less than 15% without a high school diploma.
- Medium Education: Between 15% and 30%.
- Low Education: Greater than 30%.

The boxplot in Figure 5 shows that areas with lower educational attainment tend to have slightly higher crime counts, but the difference is not statistically significant.

The average crime counts for each group are visualized in Figure 6. Medium-education areas exhibit the highest average crime count.

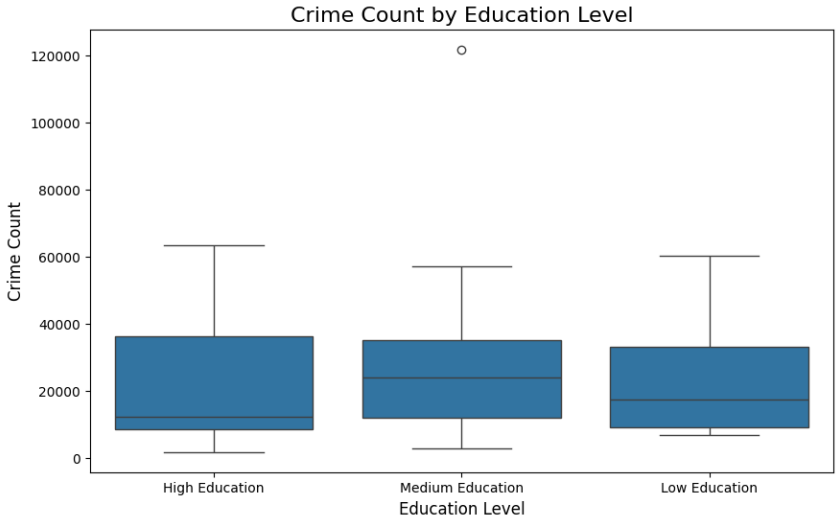

Figure 5: Crime Count Distribution by Education Level

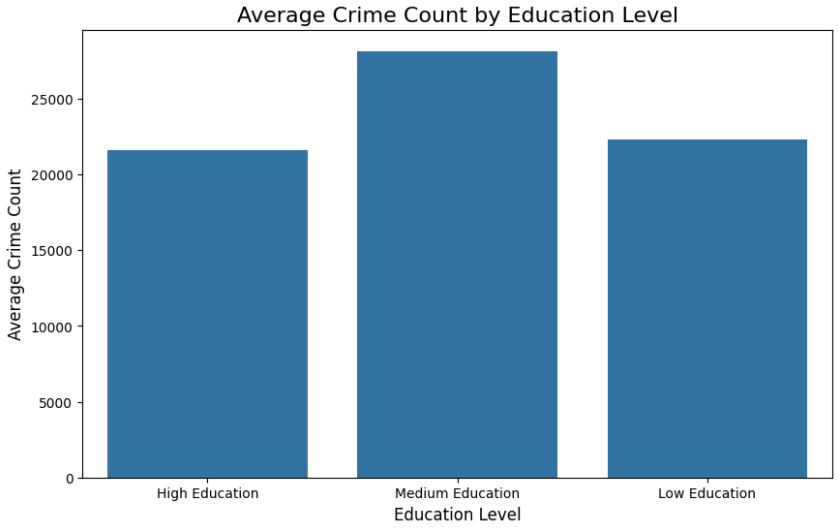

Figure 6: Average Crime Count by Education Level

3.4 Crime Count vs. Unemployment Rate

Figure 7 illustrates the relationship between unemployment rate and crime count. The scatter plot suggests a moderate positive relationship, with higher unemployment rates generally associated with higher crime counts.

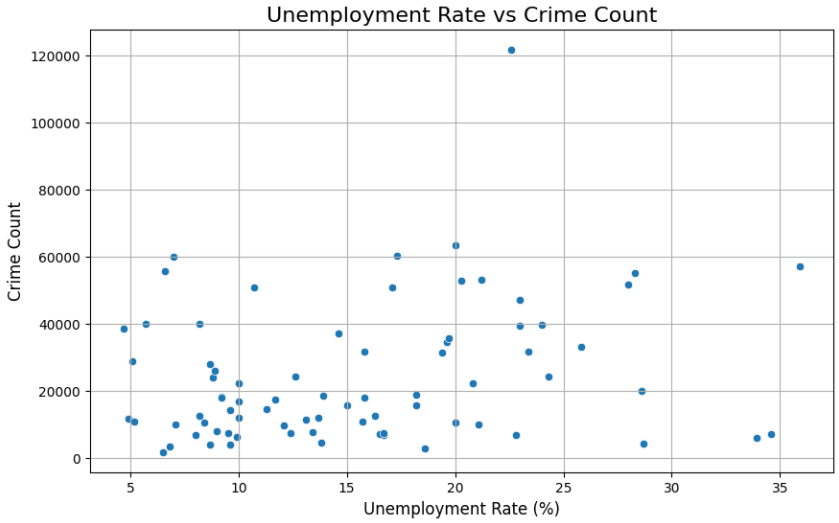

Figure 7: Unemployment Rate vs Crime Count

3.5 Crime Count vs. Unemployment Rate: Statistical Comparison

To analyze whether high and low unemployment areas differ significantly in crime counts, we divided community areas into two groups:

- **High Unemployment**: Unemployment rate $> 10\%$.
- **Low Unemployment**: Unemployment rate $\leq 10\%$.

Using an independent-samples t-test, we found the following:

- **T-statistic:** 1.64
- **P-value:** 0.106

The results indicate that there is no statistically significant difference in crime counts between high and low unemployment areas at the 0.05 level. However, the trend suggests that higher unemployment rates may be associated with higher crime counts, warranting further investigation.

3.6 Crime Count vs. Education Level: Statistical Comparison

We categorized community areas based on education levels as follows:

- Group 1: High Education (Less than 15% without a high school diploma)
- Group 2: Medium Education (15% to 30%)
- Group 3: Low Education (Greater than 30%)

Using one-way ANOVA to test the differences in crime counts across these groups, we found:

- **F-statistic:** 0.88

- **P-value:** 0.421

The results suggest that there is no statistically significant difference in crime counts among the three education groups. Group sizes were as follows:

- Group 1 Size: 30
- Group 2 Size: 31
- Group 3 Size: 16

The findings indicate that education level, as measured by the percentage of residents without a high school diploma, is not a strong determinant of crime counts.

3.7 Regression Analysis

To further quantify the relationship between socioeconomic factors and crime counts, we conducted a multiple linear regression analysis using the following independent variables:

- **Unemployment Rate (%)**: Percentage of residents aged 16+ unemployed.
- **Per Capita Income**: Median per capita income in each community area.
- **Education Level (%)**: Percentage of residents aged 25+ without a high school diploma.

The dependent variable is the total **Crime Count** for each community area. Table 1 summarizes the regression results.

Table 1: OLS Regression Results

Variable	Coefficient	Std. Error	t-Statistic	P-value
Intercept (Constant)	-18300	15500	-1.184	0.240
Unemployment Rate (%)	1272.28	404.67	3.144	0.002**
Per Capita Income	0.61	0.27	2.290	0.025*
Education Level (%)	374.23	276.49	1.354	0.180
Model Statistics				
R-squared	0.125			
Adjusted R-squared	0.089			
F-statistic	3.467 (p = 0.0204)			
Observations (n)	77			

Notes: **p ¡ 0.01, *p ¡ 0.05. Standard Errors assume the covariance matrix of errors is correctly specified. The data used are from the Chicago Data Portal (2008–2012), including crime and socioeconomic indicators from census data.

Findings:

- The regression model explains 12.5% of the variation in crime counts ($R^2 = 0.125$).
- **Unemployment Rate (%)**: A significant positive relationship with crime counts ($p = 0.002$). For each 1% increase in unemployment, crime counts increase by approximately 1272 incidents.

- **Per Capita Income**: A small but significant positive association with crime counts ($p = 0.025$). This could indicate that higher-income areas report more crimes, potentially due to better reporting mechanisms.
- **Education Level (%)**: The relationship is not statistically significant ($p = 0.180$), suggesting education has a weaker impact on crime counts.

While the model reveals important trends, the relatively low R^2 value suggests that additional factors, such as housing density or policing policies, may also contribute to crime rates and warrant further investigation.

4 Data and Variables

4.1 Data Sources

The data used in this study come from two main sources:

- **Crime Data:** Obtained from the Chicago Data Portal (2008–2012), this dataset contains detailed records of reported crimes across Chicago's community areas, including the type of crime, date, and location.
- **Census Data:** Socioeconomic indicators were sourced from the Census Data provided by the City of Chicago. Key indicators include unemployment rates, education levels, and per capita income for each community area.

4.2 Variables

The analysis involves the following variables:

- **Dependent Variable: Crime Count** – Total number of reported crimes in each community area during the period 2008–2012.
- **Independent Variables:**
  - **Unemployment Rate (%):** Percentage of residents aged 16+ unemployed.
  - **Per Capita Income:** Median per capita income in each community area.
  - **Education Level (%):** Percentage of residents aged 25+ without a high school diploma.
- **Derived Variable: Credit Risk Index (CRI)** – A composite index defined as:

$$\text{CRI} = \frac{\text{Crime Count} \times \text{Unemployment Rate (\%)}}{\text{Median Income}}$$

4.3 Data Coverage and Limitations

The crime data covers 77 community areas in Chicago for the years 2008 to 2012. Census data reflects socioeconomic indicators aggregated over the same period. While this dataset provides a comprehensive view of crime and socioeconomic factors, it does not include variables such as policing strategies, housing density, or healthcare access, which may also influence crime rates.

5 Conclusion

This study demonstrates the strong correlation between socioeconomic factors and crime in Chicago's community areas. The key findings include:

- The **Credit Risk Indicator (CRI)**, combining unemployment rates, median income, and crime counts, provides a useful proxy for evaluating community risk levels.
- A strong correlation ($r = 0.77$) between CRI and crime counts suggests that crime is a significant driver of community risk.
- **Unemployment Rate** is moderately associated with higher crime counts ($p = 0.002$), indicating that addressing unemployment may have a positive impact on reducing crime.
- **Education Level** shows no statistically significant relationship with crime counts ($p = 0.180$), suggesting its influence may be less direct.

These insights are critical for policymakers and impact investors. For example:

- Policies targeting employment opportunities may yield more significant reductions in crime rates compared to initiatives solely focused on education.
- Impact investors can use the CRI to identify high-risk areas that may benefit from targeted social or financial investments.

5.1 Future Work

The relatively low R^2 value of the regression model indicates the need for further research into additional factors influencing crime, such as:

- Housing conditions and population density.
- Accessibility of healthcare and public services.
- Community policing strategies and their effectiveness.

Incorporating these variables could provide a more comprehensive understanding of the drivers of crime and inform more effective policies and investment strategies.