Analyzing Crime Patterns in Chicago

Introduction

The City of Chicago maintains a comprehensive database of criminal reports, including narcotics-related crimes such as marijuana possession (CANNABIS) and other offenses (NOTCANNABIS). This dataset, combined with demographic information from the 2010 US Census and the 2011 American Community Survey, offers a unique opportunity to analyze crime patterns across different neighborhoods. This report aims to explore key questions regarding the distribution of crime rates, the relationship between demographic factors and crime, and temporal trends.

Exploratory Data Analysis

Data Summary:
The dataset containscensus block groups in Chicago, with variables that include demographic information and crime data. It provides insights into how these factors vary across different neighborhoods and their potential relationship to crime rates. The crime data, spread over several years, offers an opportunity to explore patterns in crime, while demographic variables allow for the investigation of socio-economic and geographic factors that might influence crime trends.

- **Population Data**: `poptotal`, `popwhite`, `popblack`, `popasian`
- **Income Data**: `income.male`, `income.female`
- **Age Data**: `age.male`, `age.female`
- **Geographic Data**: `longitude`, `latitude`
- **Crime Data**: Crime counts for two types—`CANNABIS` and `NOTCANNABIS`—are recorded for 2010, 2011 and 2012.

Data Exploration:
NA values were removed from the dataset as they were sparse in number and did not contribute significantly to the overall analysis. Additionally, removing these NA values also eliminated rows where the population count was zero, which would have been irrelevant for future analysis. This cleaning step ensured that only meaningful data, with valid population values, was retained for further analysis.

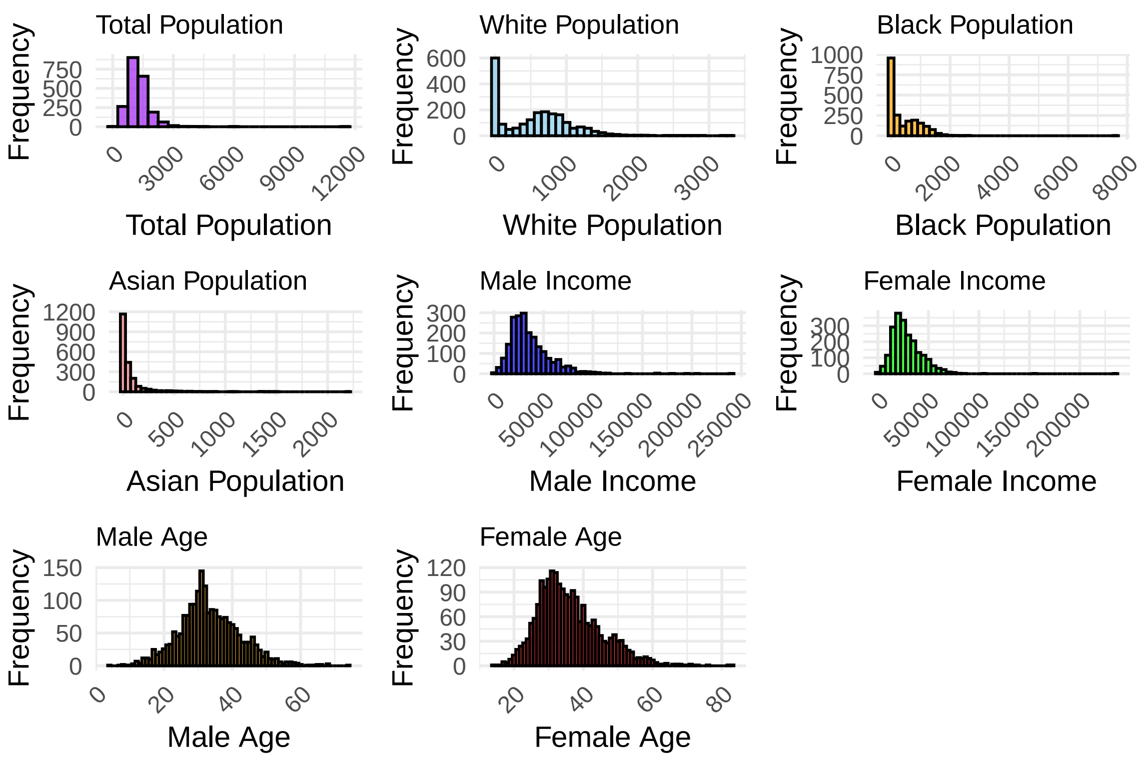

Figure 1: Distribution of Predictors

Geographic Distribution of Cannabis and Non-Cannabis Crime Rates in Chicago: Comparison to Mean Crime Rates:
In this analysis, we first derived the cannabis and non-cannabis crime rates per 100 population for each block group. We then classified each block group as either having "high" or "low" crime rates based on whether their crime rates were above or below the mean crime rate for each respective category. The comparison to the mean helped identify areas with significantly higher or lower crime rates for both cannabis-related and non-cannabis-related offenses.

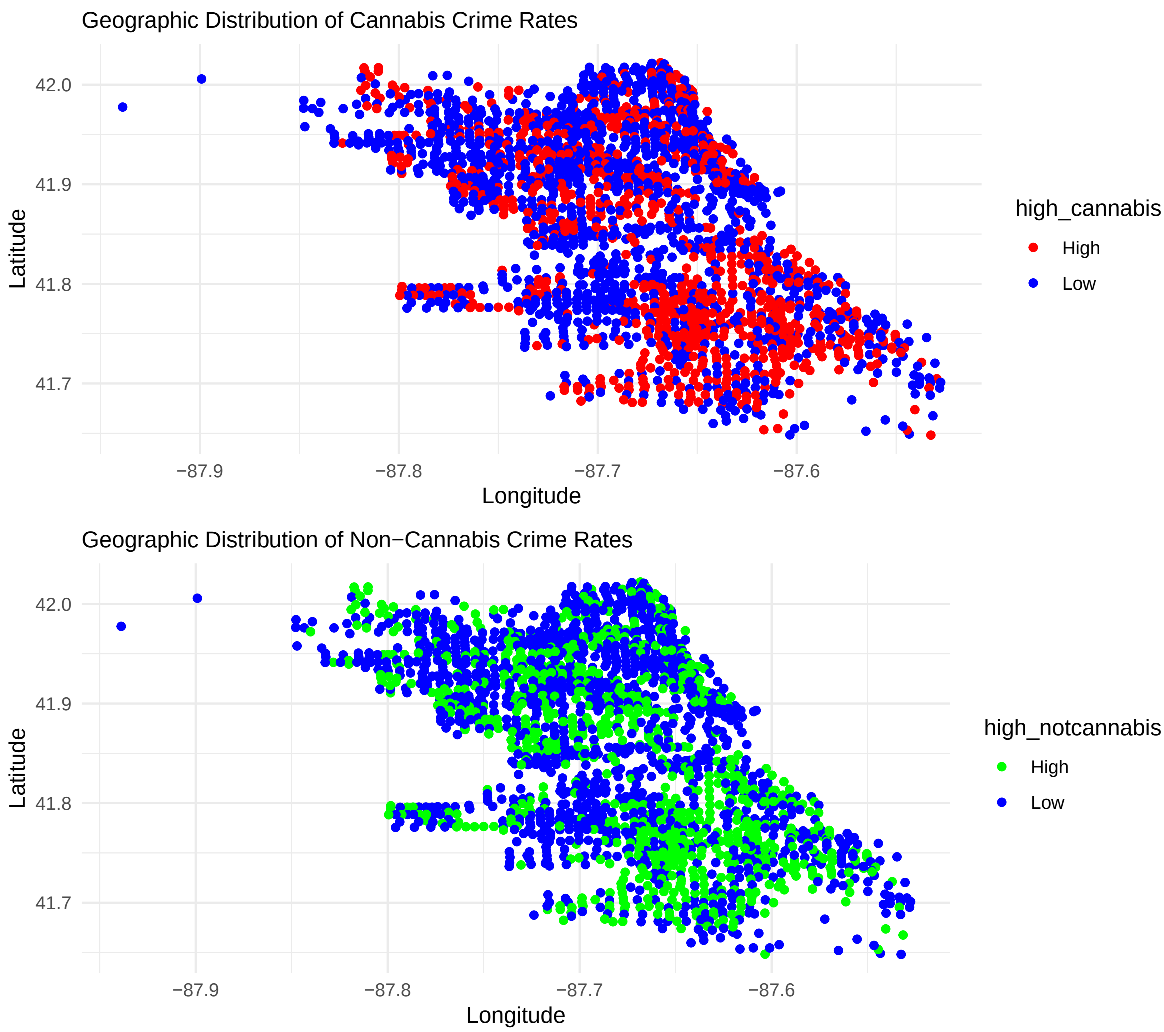

Figure 2: Geographic Distribution of Crime Rates in Chicago

Given that both cannabis-related and non-cannabis-related crime rates are evenly distributed across the city, as observed through their geographic spread along longitude and latitude, there appears to be no clear geographic concentration or clusters of block groups with significantly higher or lower crime rates. This suggests that crime patterns for both types of offenses are relatively uniform across the city, rather than being concentrated in specific neighborhoods.

Crime Rate Variations and the Relationship Between Cannabis and Non-Cannabis Crimes Across Wards and Community Areas

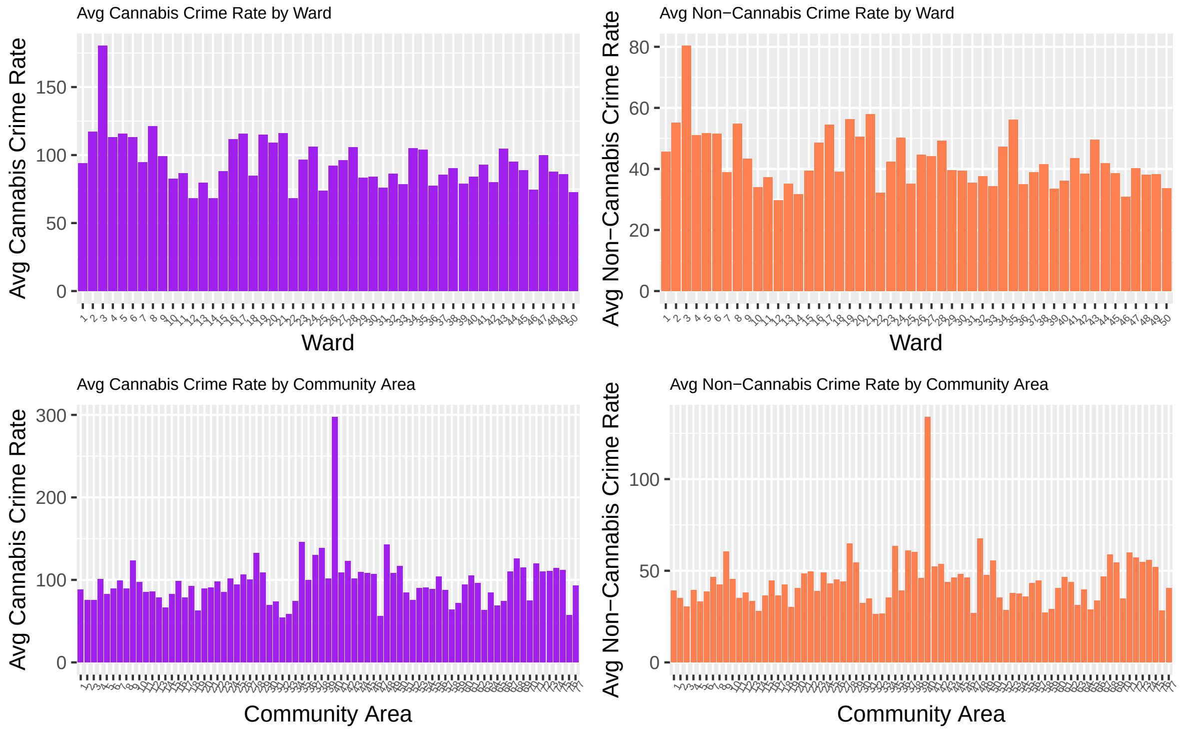

Figure 3: Crime Rate Variations by Ward and Community Area

The crime rate patterns across Wards and Community Areas appear quite similar. Both levels show comparable trends in cannabis and non-cannabis crime rates, with some geographic areas exhibiting consistently higher crime rates.

Relationship Between Cannabis and Non-Cannabis Crimes at these levels:

$$\text{Avg Cannabis Crime Rate}_{\text{Ward}} = \beta_0 + \beta_1 \cdot \text{Avg Non-Cannabis Crime Rate}_{\text{Ward}} + \epsilon$$

$$\text{Avg Cannabis Crime Rate}_{\text{Community Area}} = \beta_0 + \beta_1 \cdot \text{Avg Non-Cannabis Crime Rate}_{\text{Community Area}} + \epsilon$$

There is a strong and statistically significant relationship between cannabis and non-cannabis crime rates at both the Ward and Community Area levels. In the Ward data, non-cannabis crime rates explained 93.47% of the variance in cannabis crime rates ($R^2 = 0.9347$), with cannabis crime rates increasing by 1.99 units for each unit increase in non-cannabis crime rates ($F(1, 48) = 687.3$, $p < 2.2 \times 10^{1}$). Similarly, the Community Area model explained 95.68% of the variance ($R^2 = 0.9568$), with cannabis crime rates increasing by 2.08 units for each unit increase in non-cannabis crime rates ($F(1, 75) = 1662$, $p < 2.2 \times 10^{1}$).

Correlation between cannabis- and non-cannabis-related police reports in each block group

There is a strong positive correlation (0.9700) between cannabis crime rates and non-cannabis crime rates across each block groups in the data, meaning higher cannabis crimes tend to be associated with higher non-cannabis crimes. This suggests that the two types of crime often occur together in the same block groups.

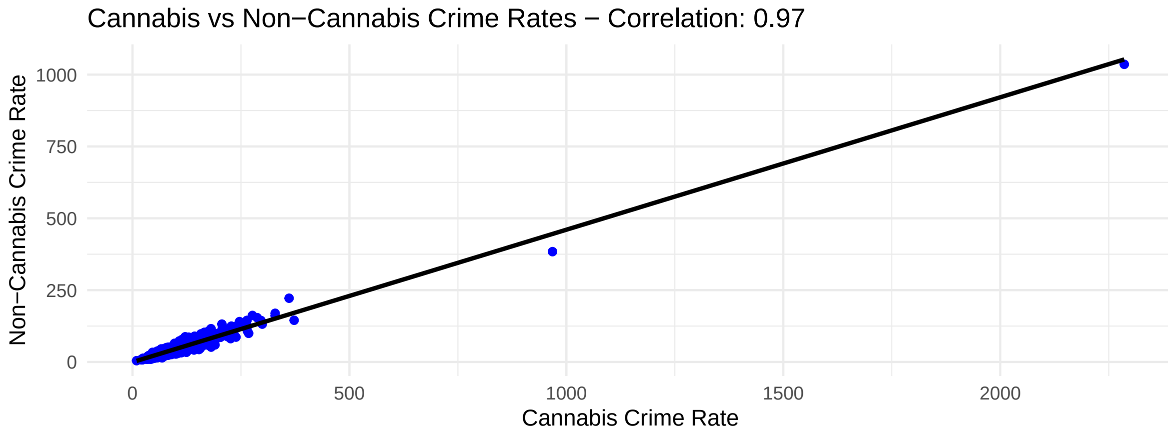

Figure 4: Correlation between cannabis and non-cannabis

When controlling for race:

$$\text{Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Non-Cannabis Crime Rate}) + \beta_2(\text{Pop White}) \\ + \beta_3(\text{Pop Black}) + \beta_4(\text{Pop Asian}) + \epsilon$$

$$\text{Non-Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Cannabis Crime Rate}) + \beta_2(\text{Pop White}) \\ + \beta_3(\text{Pop Black}) + \beta_4(\text{Pop Asian}) + \epsilon$$

Non-cannabis crime rates significantly predicted cannabis crime rates, explaining 94.2% of the variance ($R^2 = 0.942$), with a 2.01-unit increase in cannabis crimes for every 1-unit rise in non-cannabis crimes (= 2.01, $p < .001$). Conversely, cannabis crimes predicted non-cannabis crimes, explaining 94.1% of the variance ($R^2 = 0.941$), with a 0.46-unit increase (= 0.46, $p < .001$). While racial composition impacted cannabis crimes, it had no significant effect on non-cannabis crimes, and controlling for race did not alter the relationship between the two crime types.

When controlling for income:

$$\text{Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Non-Cannabis Crime Rate}) + \beta_2(\text{Income - Male}) + \beta_3(\text{Income - Female}) + \epsilon$$

$$\text{Non-Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Cannabis Crime Rate}) + \beta_2(\text{Income - Male}) + \beta_3(\text{Income - Female}) + \epsilon$$

Non-cannabis crime rates significantly predicted cannabis crime rates ($R^2 = 0.941$), with a 2.04-unit increase per 1-unit rise (= 2.04, $p < .001$), while income had no significant effect. Similarly, cannabis crime rates predicted non-cannabis crimes ($R^2 = 0.941$), with a 0.46-unit increase per 1-unit rise (= 0.46, $p < .001$). Controlling for income did not change the relationship between cannabis and non-cannabis crime rates.

The Impact of Income on Cannabis and Non-Cannabis Crime Rates:

$$\text{Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Income - Male}) + \beta_2(\text{Income - Female}) + \epsilon$$

$$\text{Non-Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Income - Male}) + \beta_2(\text{Income - Female}) + \epsilon$$

Income (male and female) did not significantly predict cannabis crime rates ($R^2 = 0.00016$, $F(2, 2099) = 0.17$, $p = 0.842$), with male ($= -0.000047$, $p = 0.563$) and female income ($= 0.000044$, $p = 0.699$) both non-significant. Similarly, income did not predict non-cannabis crime rates ($R^2 = 0.00031$, $F(2, 2099) = 0.32$, $p = 0.724$), with male ($= -0.000027$, $p = 0.491$) and female income ($= 0.000039$, $p = 0.475$) showing no effect. Thus, income is not a significant predictor of cannabis or non-cannabis crime rates.

The Role of Racial Distribution in Cannabis and Non-Cannabis Crime Rates

$$\text{Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Pop White}) + \beta_2(\text{Pop Black}) + \beta_3(\text{Pop Asian}) + \epsilon$$

$$\text{Non-Cannabis Crime Rate} = \beta_0 + \beta_1(\text{Pop White}) + \beta_2(\text{Pop Black}) + \beta_3(\text{Pop Asian}) + \epsilon$$

The multiple regression analyses revealed that higher proportions of white, black, and Asian populations in a block group were significantly associated with lower cannabis and non-cannabis crime rates. For both types of crime, the coefficients for all racial groups were negative, indicating that as the percentage of these populations increased, crime rates decreased. The models explained 18.34% of the variance in cannabis crime rates and 16.79% of the variance in non-cannabis crime rates, suggesting a notable, though modest, influence of demographic composition on crime rates.

Interaction between Income and Racial Composition on Crime Rates

$$\begin{aligned}
\text{Cannabis Crime Rate} = \beta_0 &+ \beta_1(\text{Income - Male}) + \beta_2(\text{Pop White}) + \beta_3(\text{Income - Male} \times \text{Pop White}) \\
&+ \beta_4(\text{Pop Black}) + \beta_5(\text{Income - Male} \times \text{Pop Black}) + \beta_6(\text{Pop Asian}) \\
&+ \beta_7(\text{Income - Male} \times \text{Pop Asian}) + \beta_8(\text{Income - Female}) + \beta_9(\text{Pop White}) \\
&+ \beta_{10}(\text{Income - Female} \times \text{Pop White}) + \beta_{11}(\text{Pop Black}) \\
&+ \beta_{12}(\text{Income - Female} \times \text{Pop Black}) + \beta_{13}(\text{Pop Asian}) \\
&+ \beta_{14}(\text{Income - Female} \times \text{Pop Asian}) + \epsilon
\end{aligned}$$

$$
\begin{aligned}
\text{Non-Cannabis Crime Rate} = {} & \beta_0 + \beta_1(\text{Income - Male}) + \beta_2(\text{Pop White}) + \beta_3(\text{Income - Male} \times \text{Pop White}) \\
& + \beta_4(\text{Pop Black}) + \beta_5(\text{Income - Male} \times \text{Pop Black}) + \beta_6(\text{Pop Asian}) \\
& + \beta_7(\text{Income - Male} \times \text{Pop Asian}) + \beta_8(\text{Income - Female}) + \beta_9(\text{Pop White}) \\
& + \beta_{10}(\text{Income - Female} \times \text{Pop White}) + \beta_{11}(\text{Pop Black}) \\
& + \beta_{12}(\text{Income - Female} \times \text{Pop Black}) + \beta_{13}(\text{Pop Asian}) \\
& + \beta_{14}(\text{Income - Female} \times \text{Pop Asian}) + \epsilon
\end{aligned}
$$

Racial composition (white, black, and Asian populations) is significantly linked to lower cannabis and non-cannabis crime rates. Female income has a positive effect on both crime types, while male income shows no significant impact. Female income interacts significantly with the proportion of Black residents, indicating a stronger effect in these areas. Overall, female income and racial composition, particularly Black residents, are key factors influencing crime rates.

Analysis of Crime Rate Trends Across Months

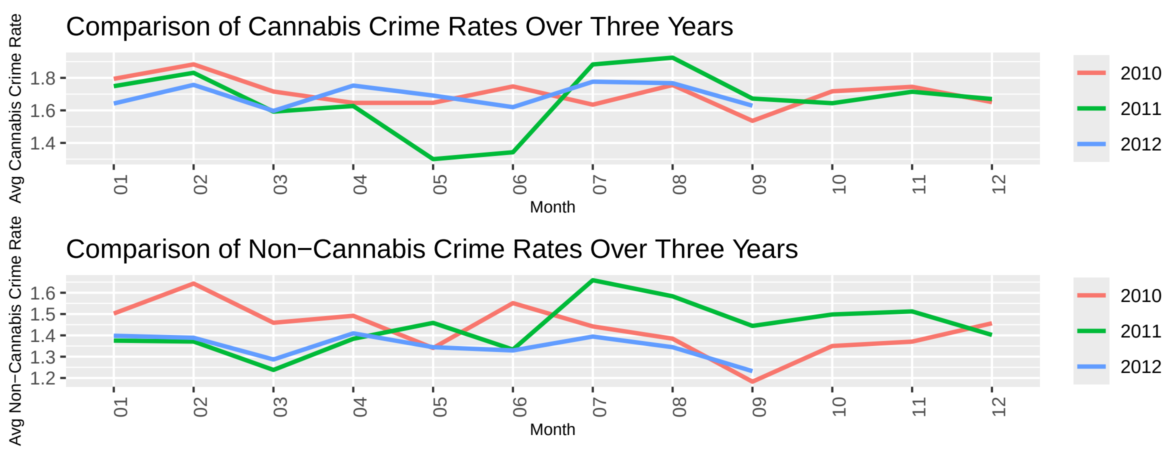

Figure 5: Comparison of Cannabis and Non-Cannabis Crime Rates Over Three Years

Both cannabis and non-cannabis crime rates rise during summer, peaking in June to August, with notable volatility in 2011 compared to stable trends in 2010 and 2012. Seasonal increases may be linked to warmer weather, outdoor activities, and social gatherings, showing parallel trends in crime activity.

A Poisson regression model predicts cannabis and non-cannabis crime rates for October to December 2012 using data from January 2011 to September 2012. The model includes demographic variables (population, race, income, age) and geographic factors (longitude, latitude, ward, community area). Population size (`poptotal`) is used as an offset to account for population differences across block groups. Separate models are fitted for each crime type, capturing the relationship between predictors and crime occurrences.

$$
\begin{aligned}
\text{Crime Rate} = {} & \beta_0 + \beta_1(\text{poptotal}) + \beta_2(\text{popwhite}) + \beta_3(\text{popblack}) + \beta_4(\text{popasian}) \\
& + \beta_5(\text{income.male}) + \beta_6(\text{income.female}) + \beta_7(\text{age.male}) + \beta_8(\text{age.female}) \\
& + \beta_9(\text{longitude}) + \beta_{10}(\text{latitude}) + \beta_{11}(\text{Ward}) + \beta_{12}(\text{Community.Area}) \\
& + \log(\text{poptotal}) + \epsilon
\end{aligned}
$$

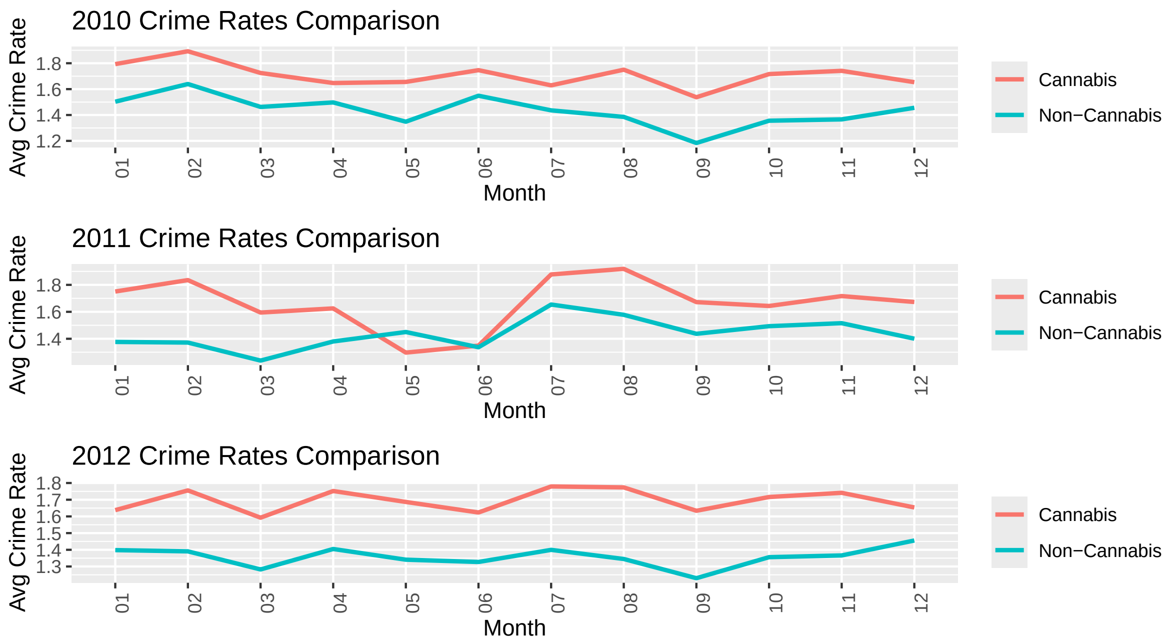

Figure 6: Monthly Comparison of Past and Predicted Cannabis and Non-Cannabis Crime Rates (2010–2012)

These line plots display the trends in average crime rates for both cannabis and non-cannabis offenses across all months from 2010 to 2012. The data compares past crime rates (actual values) with the predicted crime rates for October, November and December across all three years. They highlight fluctuations over time and the impact of various factors on crime during this period.

Discussion

Our analysis shows that cannabis and non-cannabis crimes are evenly distributed across Chicago, with a strong correlation between the two. Racial composition plays a significant role in influencing crime rates, while income appears to have minimal impact. Seasonal trends further reveal higher crime rates during the summer months. Using past crime data, our predictive models successfully estimated crime counts for late 2012. However, the analysis is limited by potential inaccuracies in crime reporting, gaps in socioeconomic data, and unaccounted factors such as variations in policing strategies.