Decoding Los Angeles Crime Patterns

Authors

Rita Chau, Bonnie Gu, Yuki Kitamura
Kaylin Lee, William Pan, Ellen Wei

UCLA College | Physical Sciences
Statistics & Data Science

Introduction

The data we analyzed came from the Los Angeles Crime dataset, obtained from the Los Angeles Police Department. It originates from transcribed police reports and constitutes a comprehensive repository of information encapsulating diverse criminal incidents from 2020 to the present. We aim to dissect the complexities of crime patterns in Los Angeles using details including crime types, severity, temporal nuances, and geographical specifics. As we investigate, we will employ various statistical methods to find patterns and get insights that not only enhance our understanding of crime in Los Angeles but also contribute to the development of effective crime prevention strategies. For more detail, please scan the QR code for a more detailed report.

Research Question

Can we predict the type of crime based on when it occurred, the area it occurs in, and who it occurred to?

Background Information

The full dataset has 28 columns and 853 thousand rows where each row is a crime incident. The columns can be divided into 6 categories: location, crime type, time, demographic, crime details, and documentation.To explore our research question, we selected 3 specific features to investigate in relation to crime type:

- Crime Location
- Time of Occurence
- Victim Sex

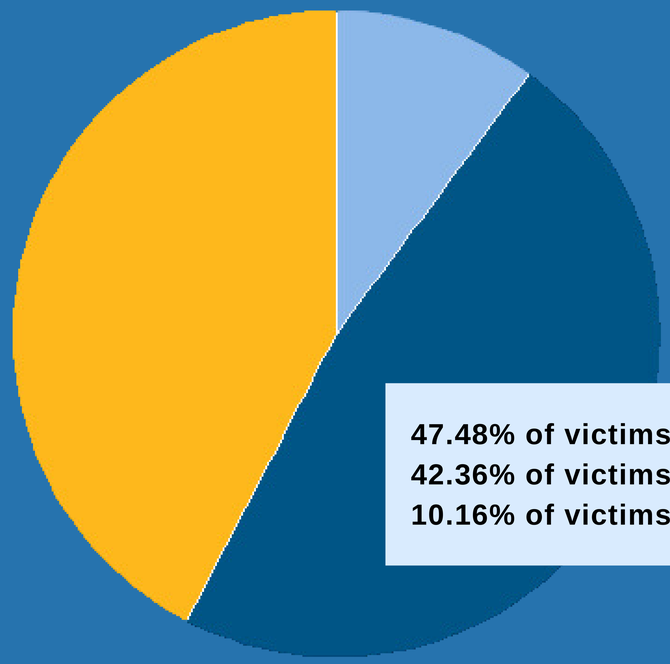
Results

The results of the Chi-Square tests indicate a statistically significant association between the type of crime and all the following variables: the geographical area, the time at which it occurs, and victim sex. This emphasizes the importance of considering the geographical context, time of crime occurrence, and victim sex when predicting and understanding criminal activities. A further investigation of this association through fitting a multinomial logistic regression model reinforces our findings from Chi-Square testing. The testing accuracy indicates an improvement over naive guesses, but shows room for improvement through additional data and expert consultation.

Conclusion

This study showed that crime location, time of occurence, and victim sex were all significant predictors of crime type, with victim sex being the best predictor of the variables we investigated. We hope that this study serves as an exploratory look into possible crime prevention strategies and future research.

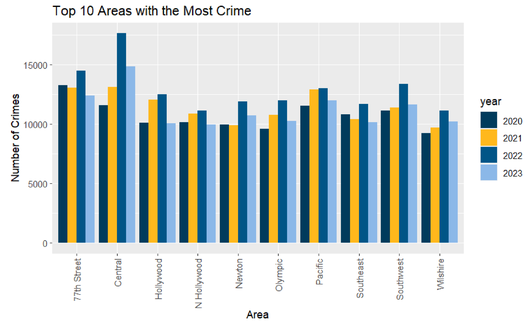
Smarter Crime Intervention: Victim sex, location, and time are meaningful keys to predicting crime type.

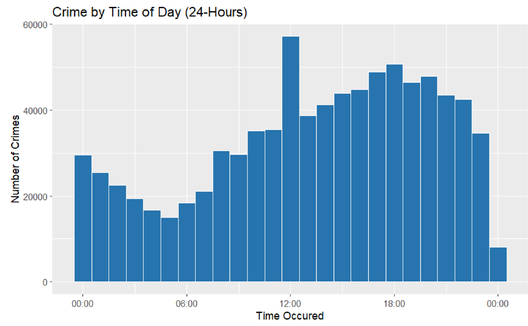


47.48% of victims are male
42.36% of victims are female
10.16% of victims are of unknown sex

Exploratory Data Analysis

We used exploratory data analysis to better understand the distributions of individual variables and possible interactions between them. The graphs included in this poster were relevant to our findings.





Analysis

Our research question aims to explore the factors that can be utilized to predict the type of crime. Initially, we hypothesized a correlation between the time a crime occurs, the specific area in which it takes place, and the sex of victim. In commencing our research, our primary objective was to investigate whether these three factors are strongly associated with the type of crime observed.

Hypothesis Testing

We conducted hypothesis testing using Chi-squared testing for each of the following associations,

- Type of Crime vs. Area It Occurs In
- Type of Crime vs. Time It Occurs
- Type of Crime vs. Victim Sex

with the null and alternative hypothesis
H_0 : There is no association between the two variables.
H_a : There is an association between the two variables.
For all three pairs of features, Pearson's Chi-squared test produced a p-value of less than 0.05, indicating a statistically significant association between type of crime and area, time, and victim sex.

Classification Model

We constructed a classification model with types of crime as target and area, time and victim sex as predictors. For the predictor time, we converted the variable into a factor based on the hour of occurence. We split the data randomly with 80% training and 20% testing. Then, we fitted a multinomial logistic regression model and the model achieved an accuracy around 24% on both training and testing sets, which was considerably better than a naive baseline that would make classification based on the most frequent crime type (around 10%). The contribution of each predictor was also evaulated through the caret package in R, where the sex of victim was shown to be the most important.