Analyzing the Impact of Demographic Factors on Crime Patterns in Los Angeles:

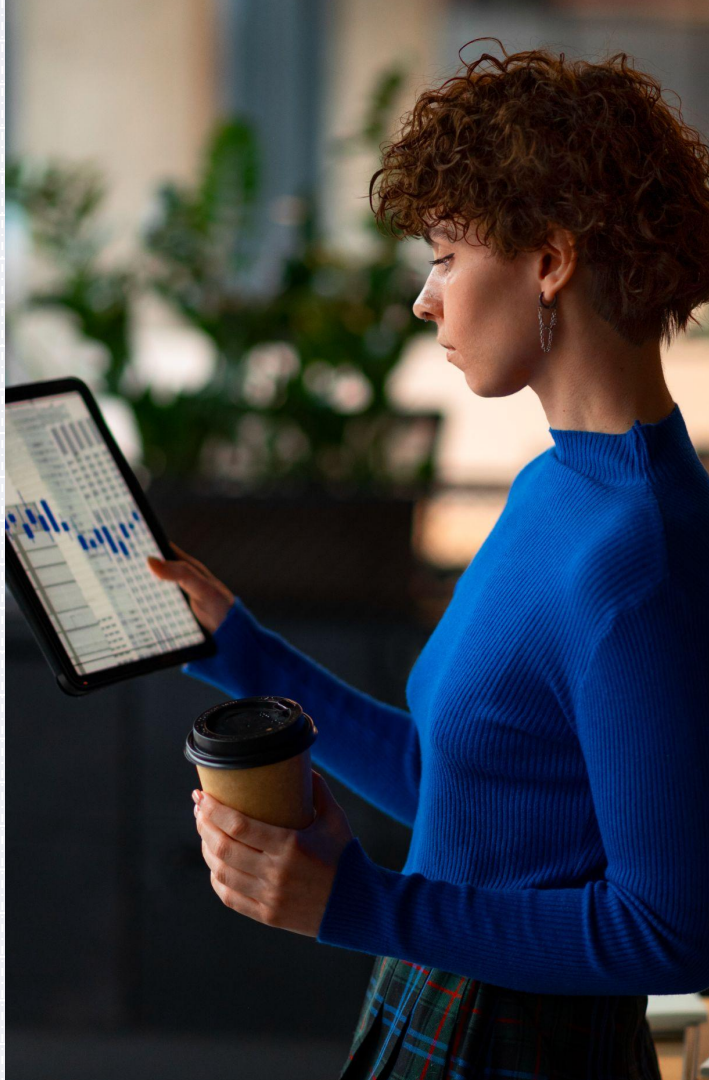
A Comparative Study of LAPD Strategies and Community Dynamics

Dewayne Barnes, Caia Gelli, and Jocelyn Niemiec

Introduction

Investigating crime patterns and LAPD effectiveness

- ➔ Objective: Investigate crime patterns and evaluate LAPD effectiveness by area, race, and crime type.
- ➔ Context: Addressing heightened scrutiny of policing institutions post-BLM movement.
- ➔ Analytical Approach: Utilizing data-driven techniques to uncover LAPD strategy shortcomings and assess Hollywood's influence on LA crime culture.
- ➔ Significance: Confronting public safety concerns and questioning the assumed effectiveness of current policing practices in Los Angeles.



Dataset Overview

Scope

This dataset captures crime incidents within Los Angeles from 2020 - 2024. Details including date, time, location, crime type, and more

Source

Kaggle Dataset “Crime Data from 2020 to Present” provided by Chaitanya K. C. K

Size and Format

Rows: 918443
Columns: 28
Format: CSV

Collection Methods

Transcribed from original paper crime reports

Content

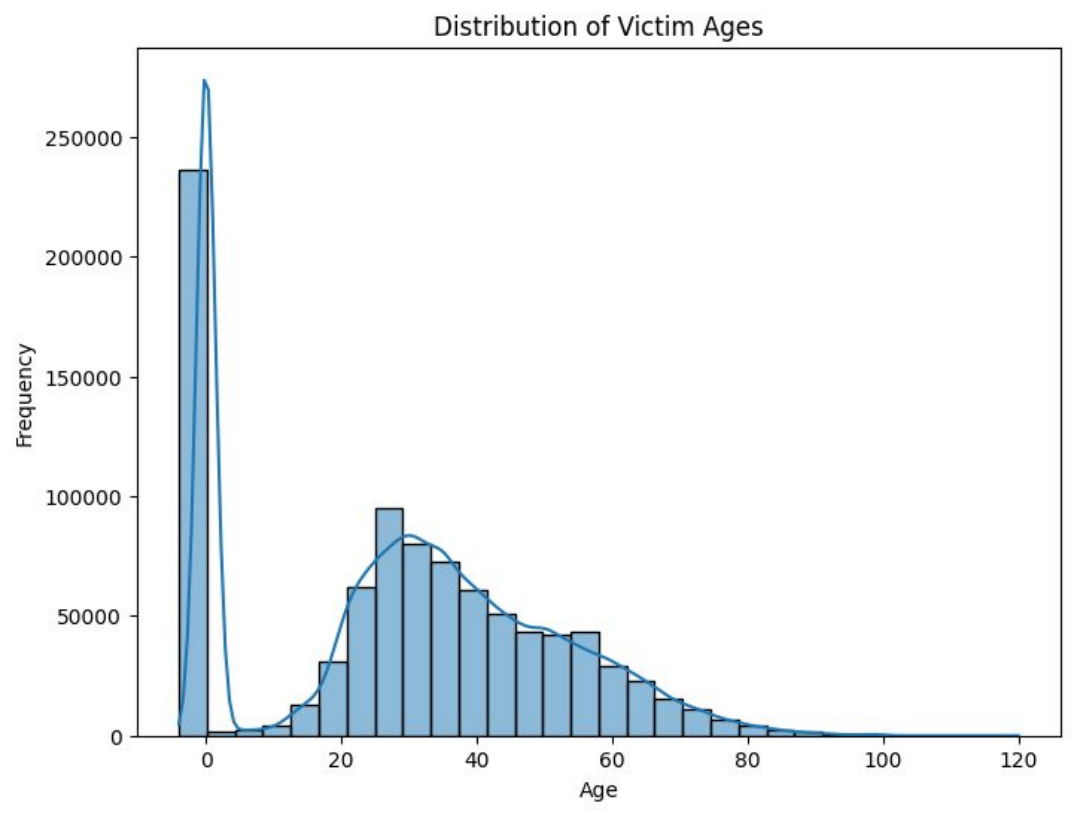
Details including date, time, location, crime type, and more

Quality Assurance

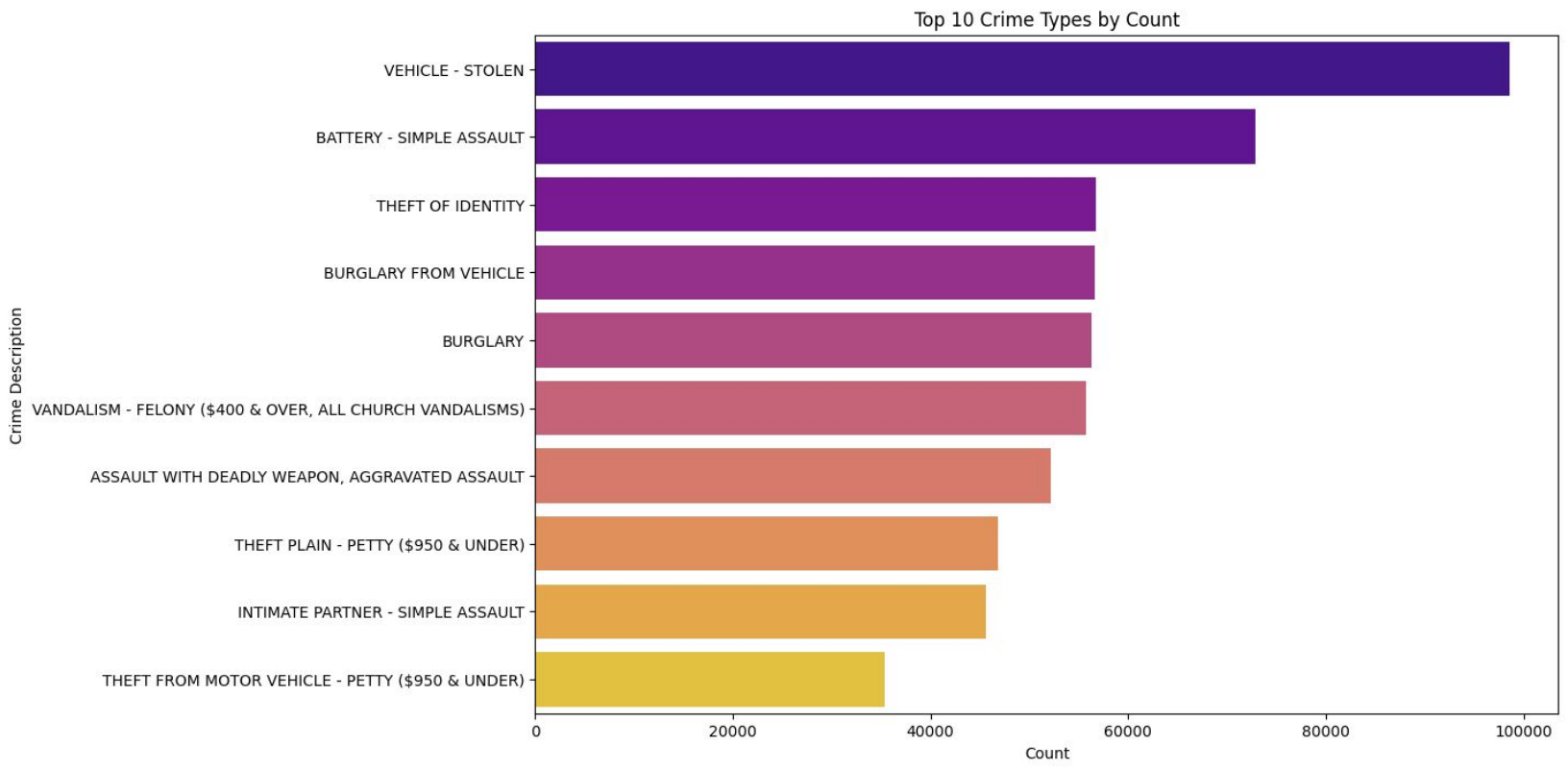
Potential inaccuracies inherent in large-scale data collection

Insights from EDA

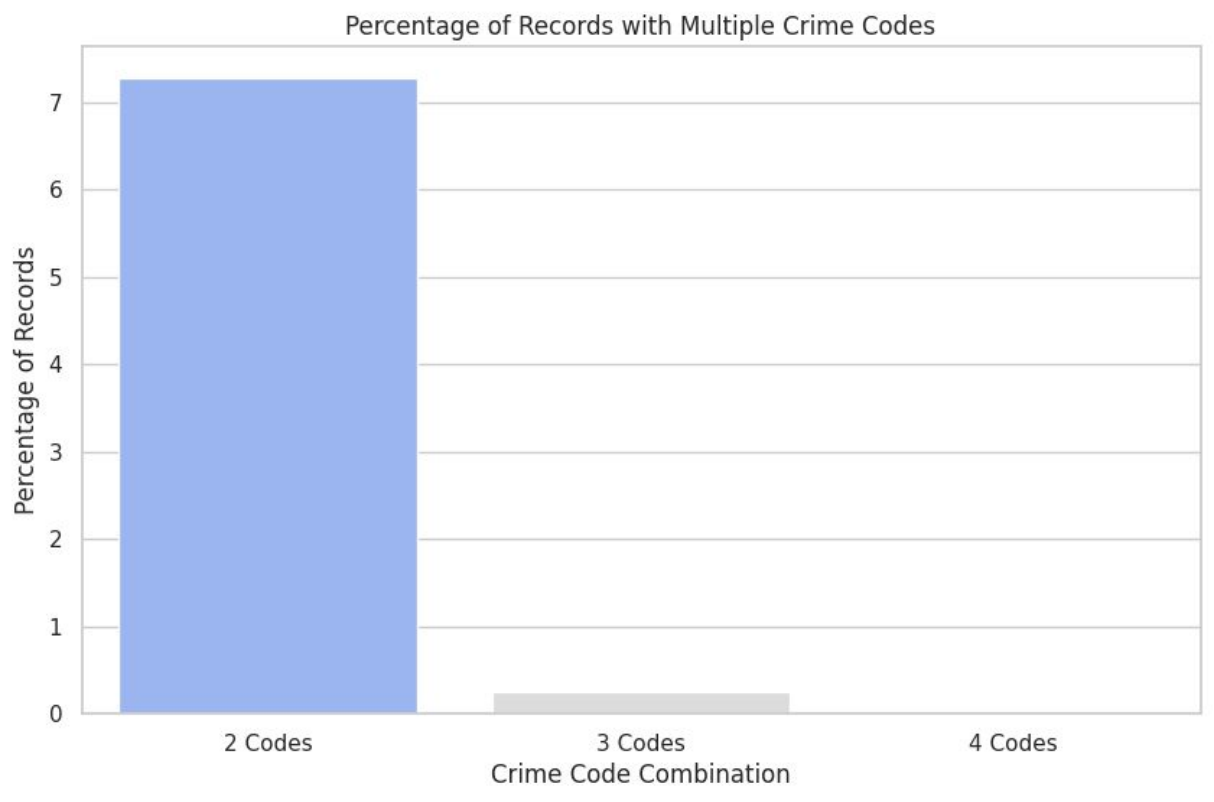
Age vs Frequency of Violent Attack



Top 10 Crimes by Count

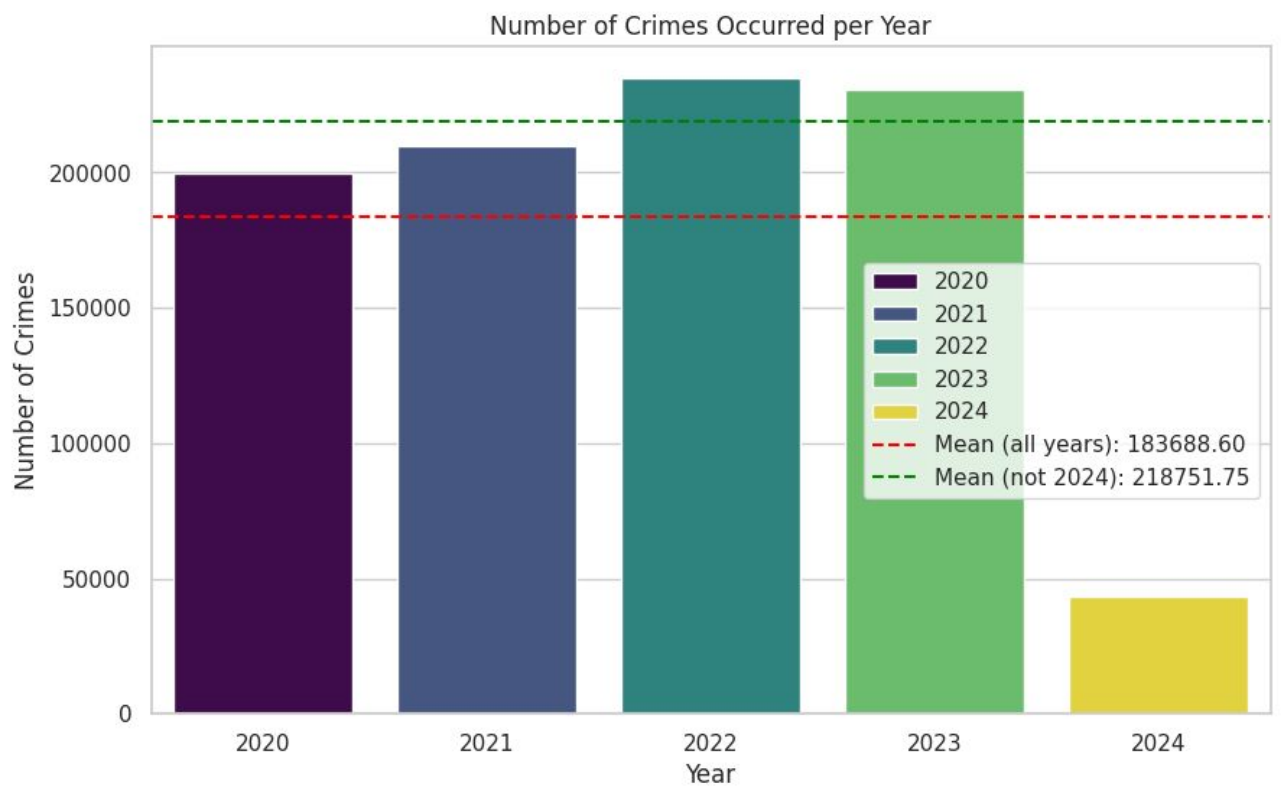


Importance of Multiple Crime Codes

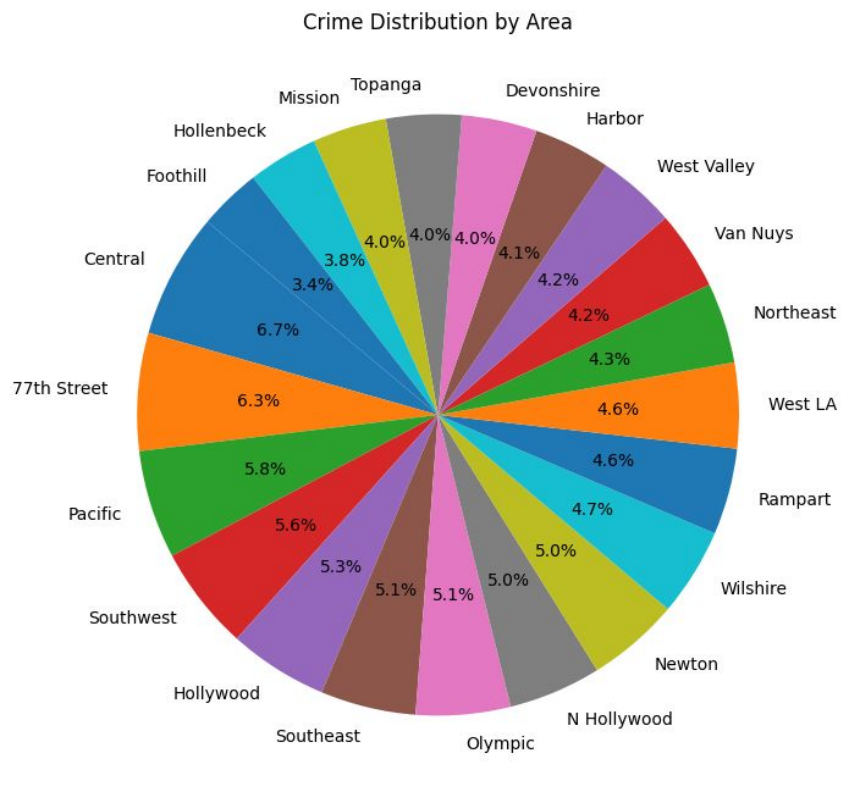


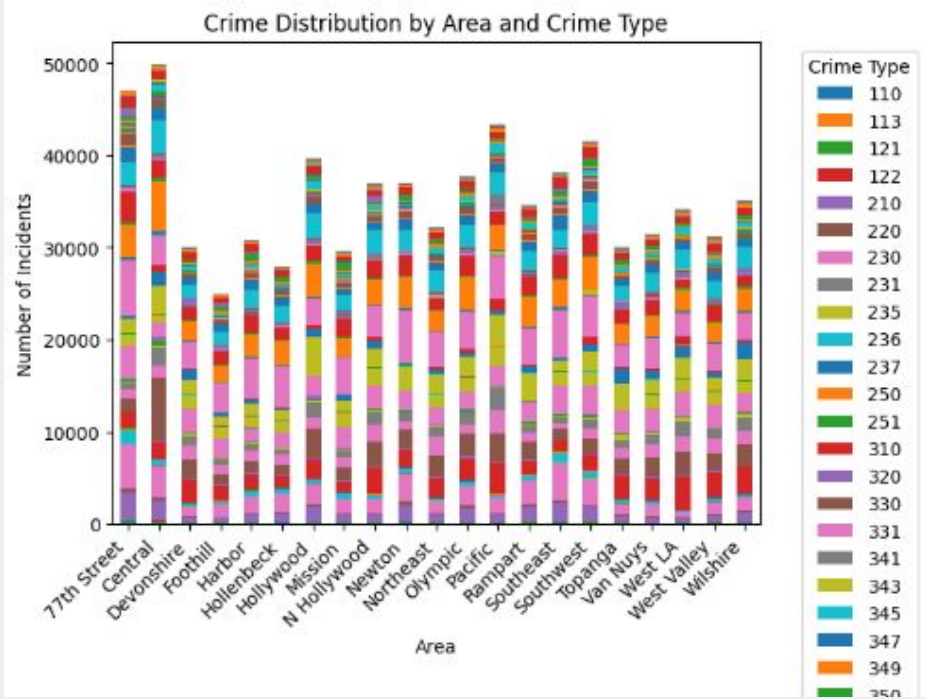
Category	Percentage
2 Codes	7.27
3 Codes	0.25
4 Codes	0.01

Crime Over Time



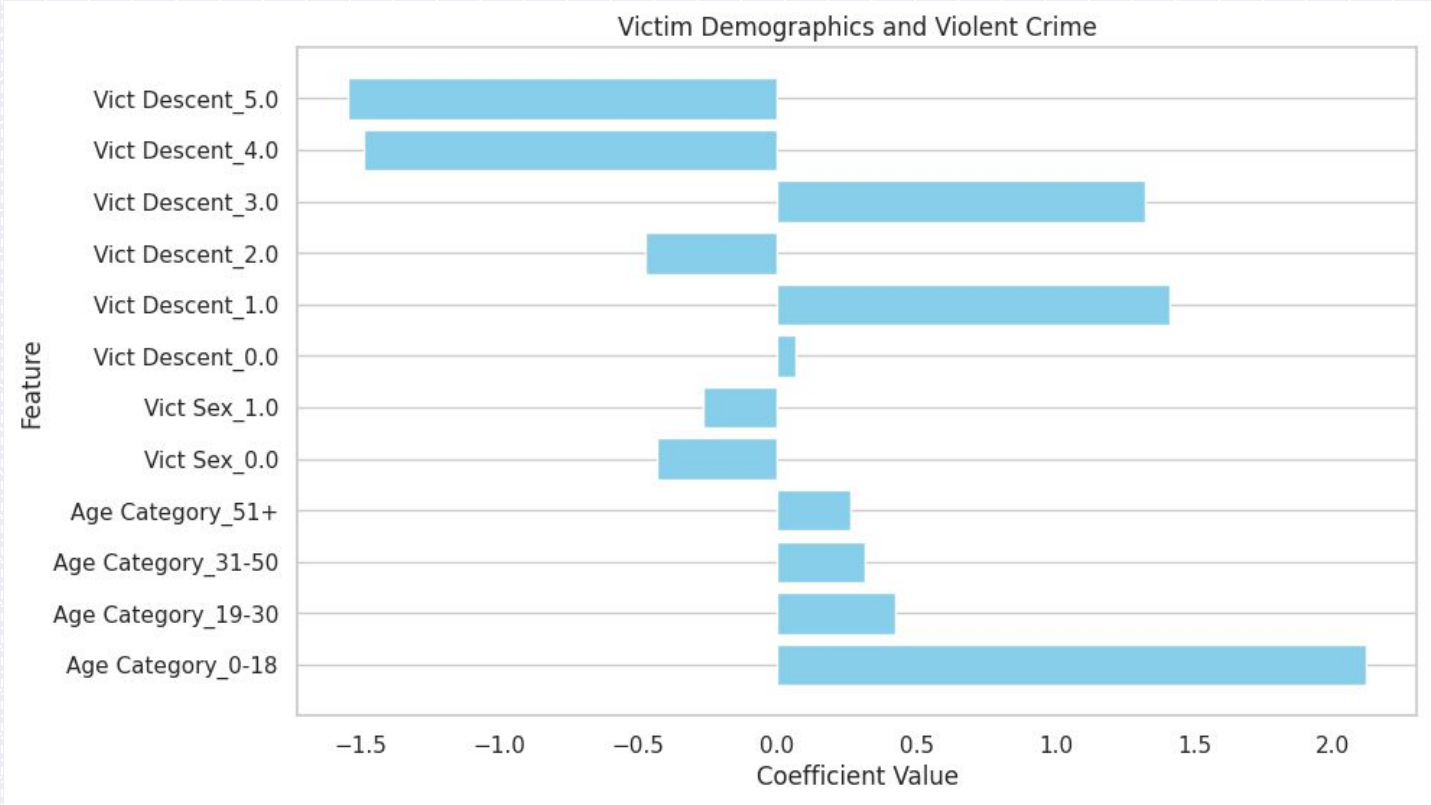
Crime Distribution by Area





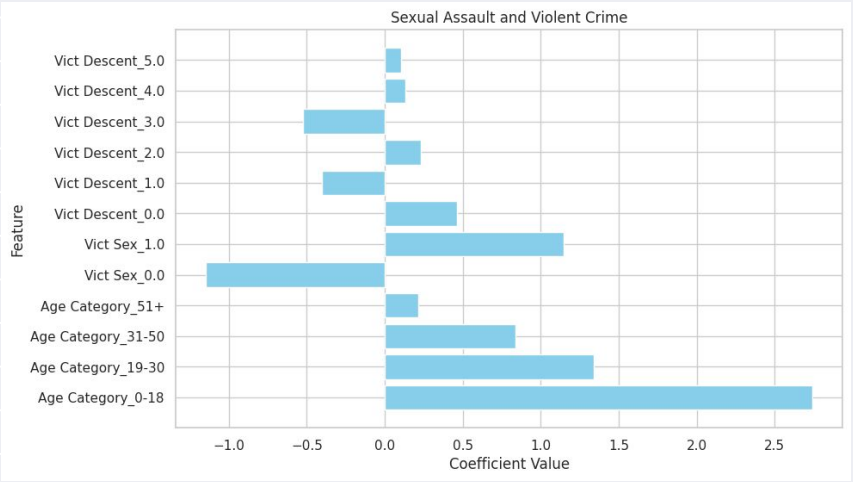
Modeling Results

Victim Demographics of Violent Crime

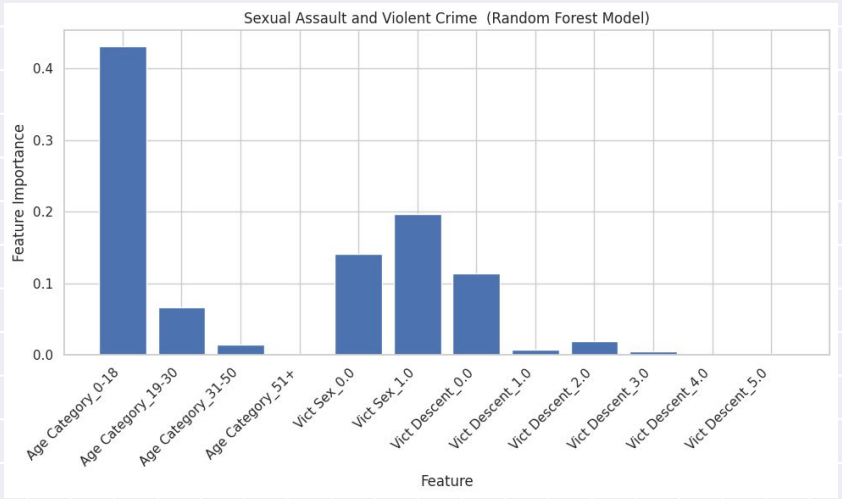


Victim Demographics of Sexual Assault

Log Reg
Accuracy: 0.97

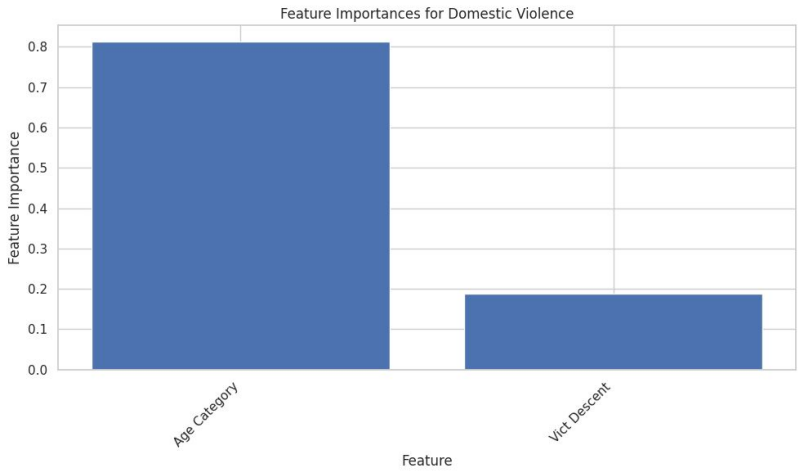


Random Forest
MSE: 0.023

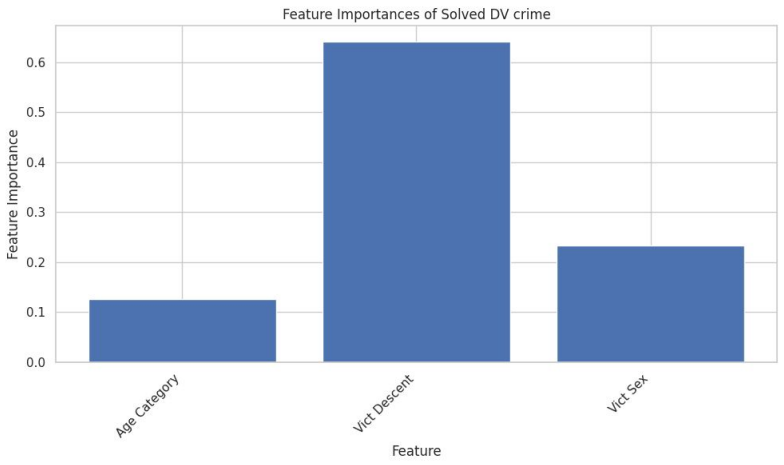


Victim Demographics of Domestic Violence

Random Forest
MSE: 0.18

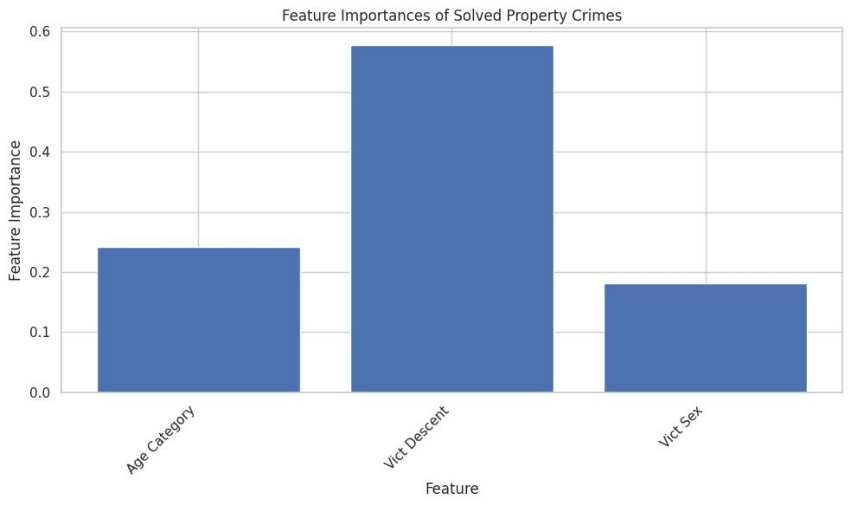


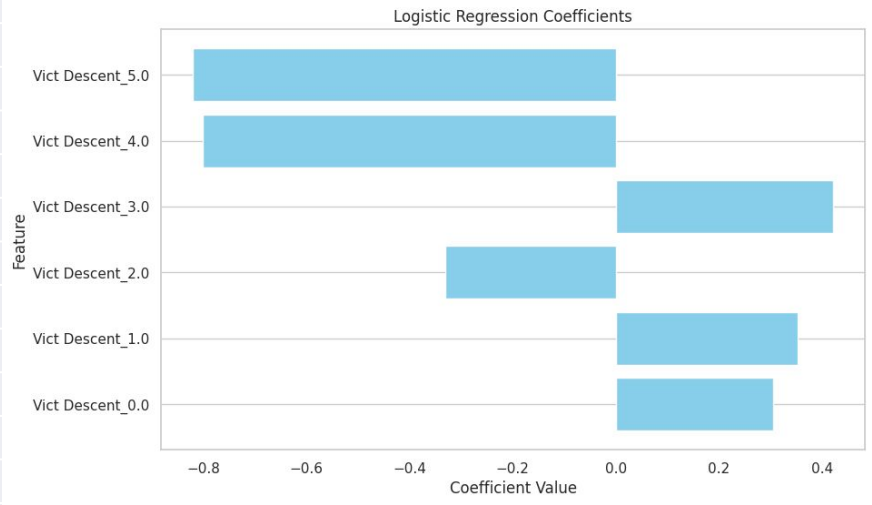
Random Forest
MSE: 0.22



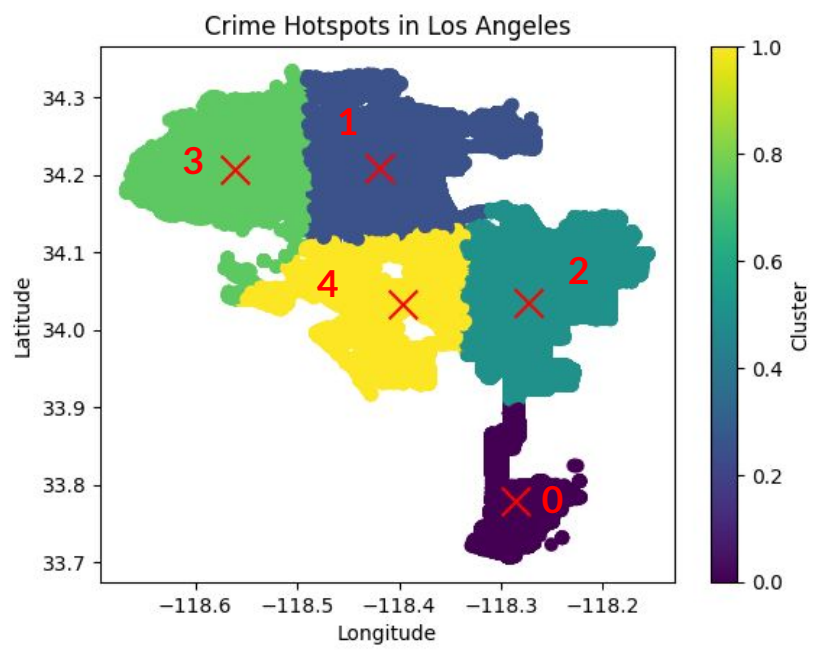
Victim Demographics of Solved Property Crimes

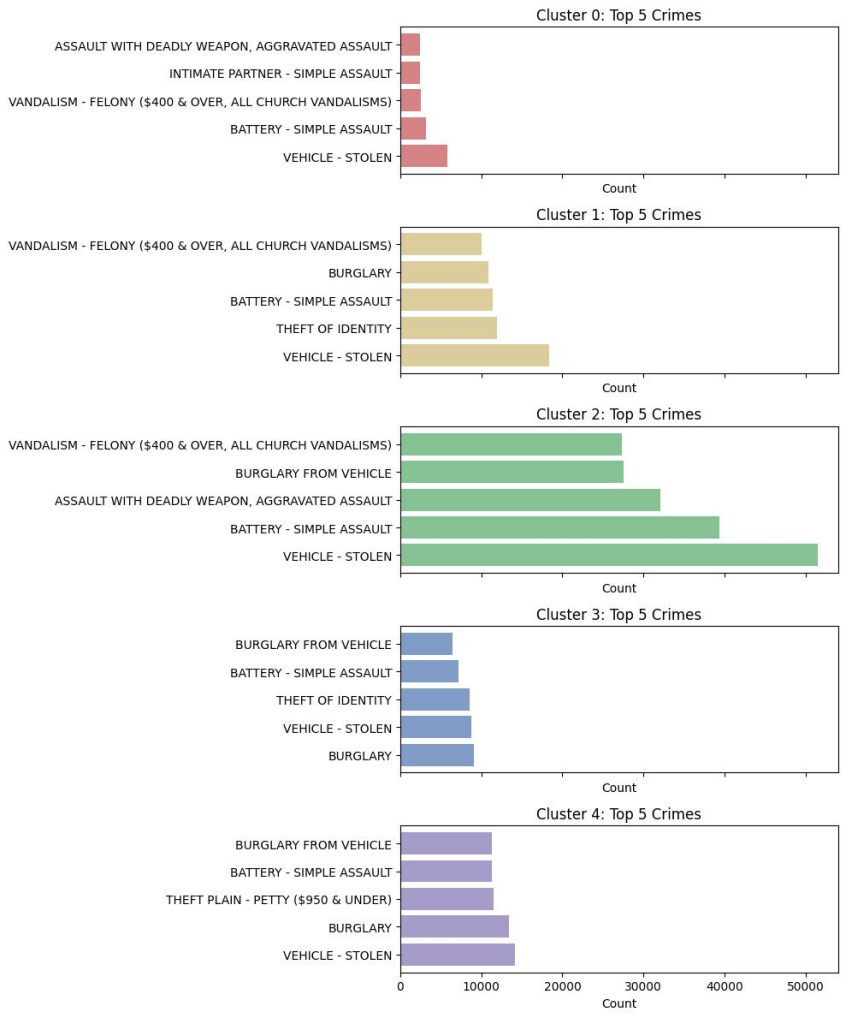
Random Forest
MSE: 0.049

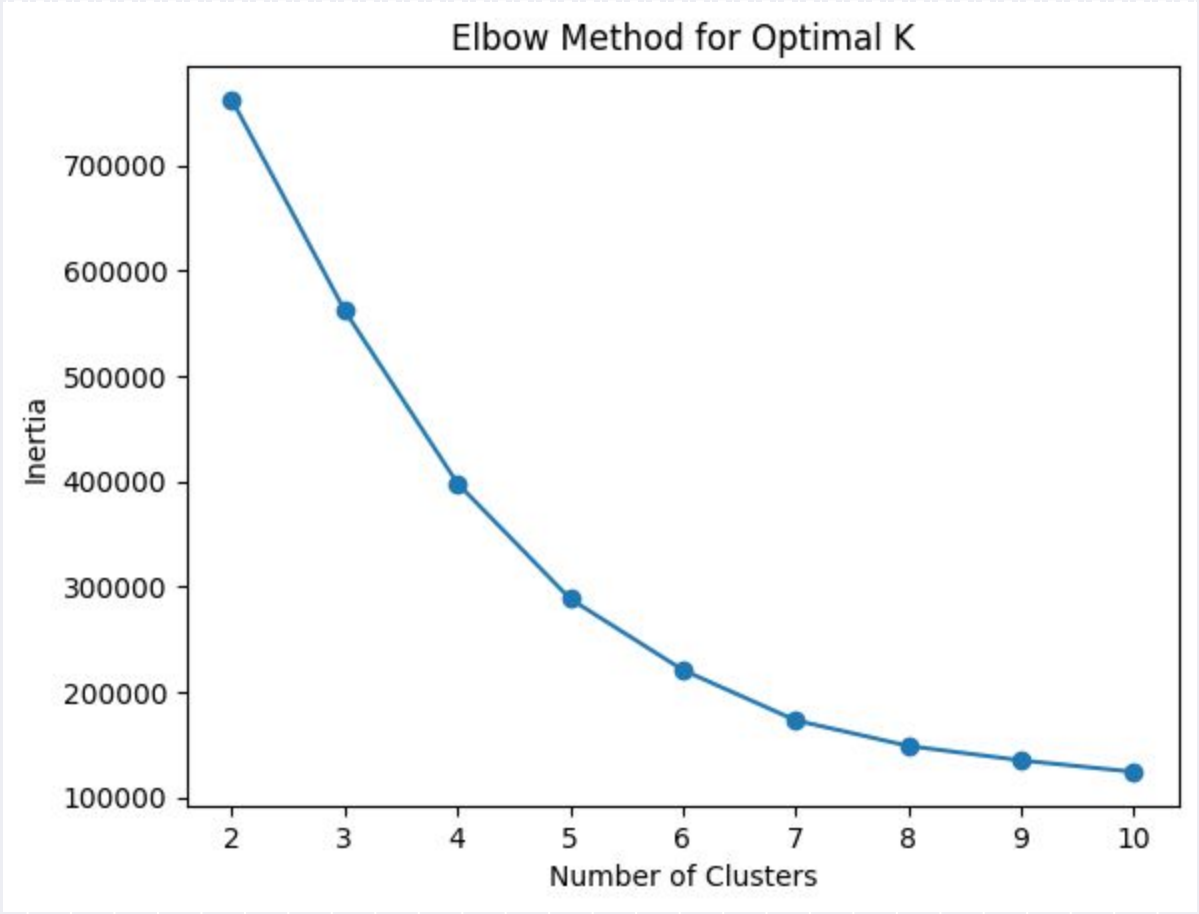




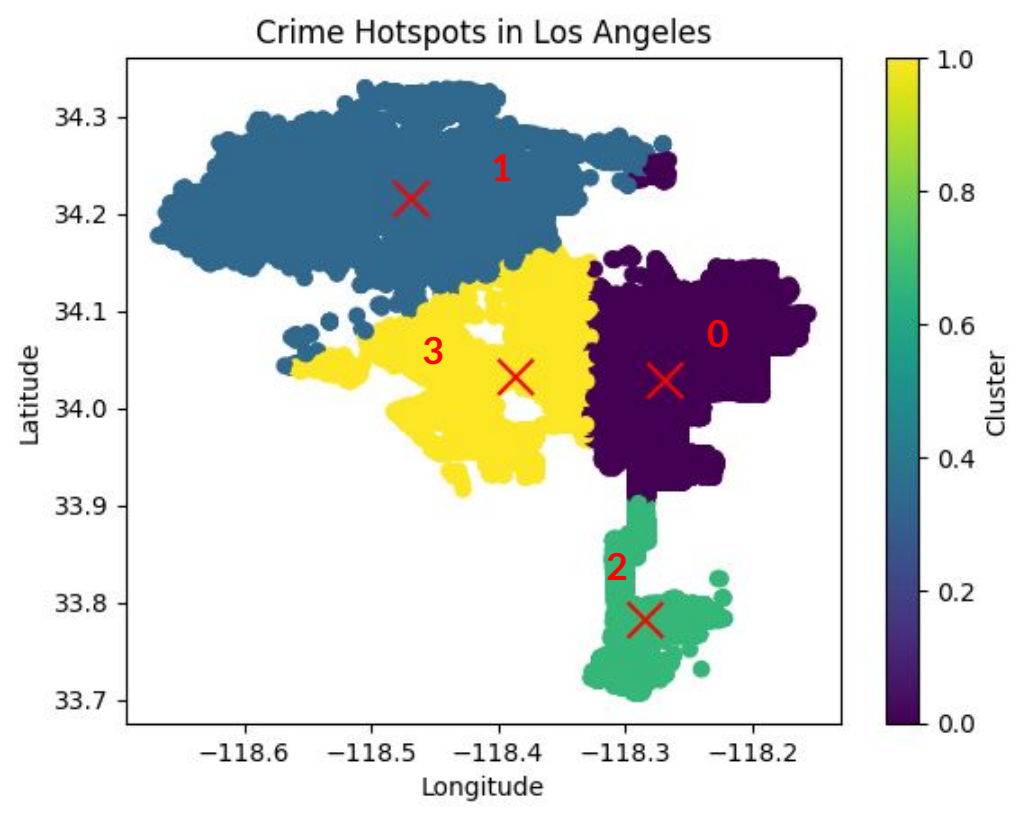
Geographic Crime Distribution





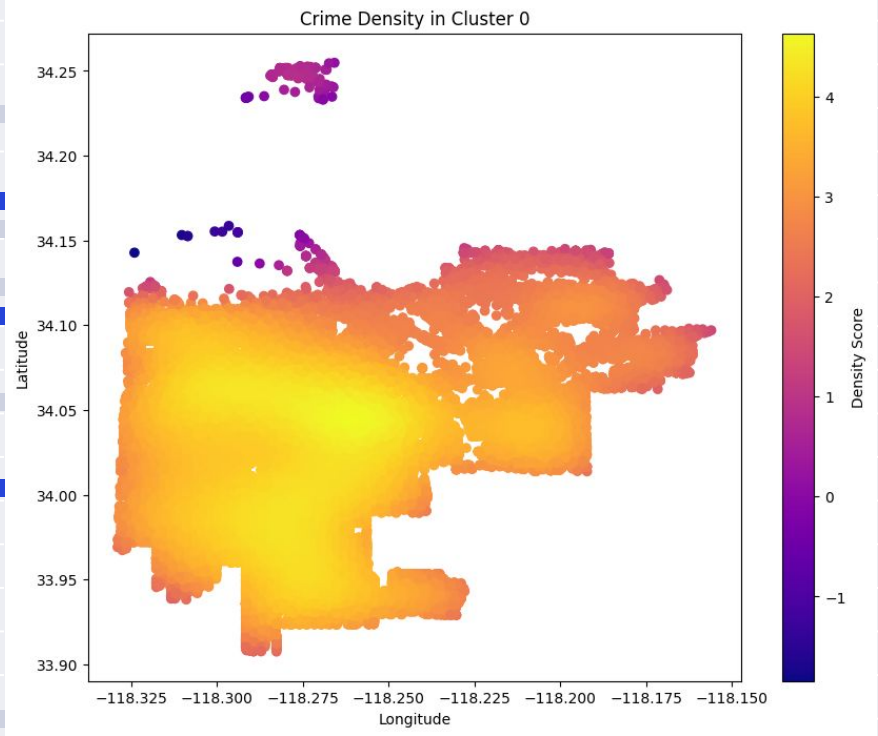

Elbow Method for Optimal K
Inertia
700000
600000
500000
400000
300000
200000
100000
2
3
4
5
6
7
8
9
10
Number of Clusters

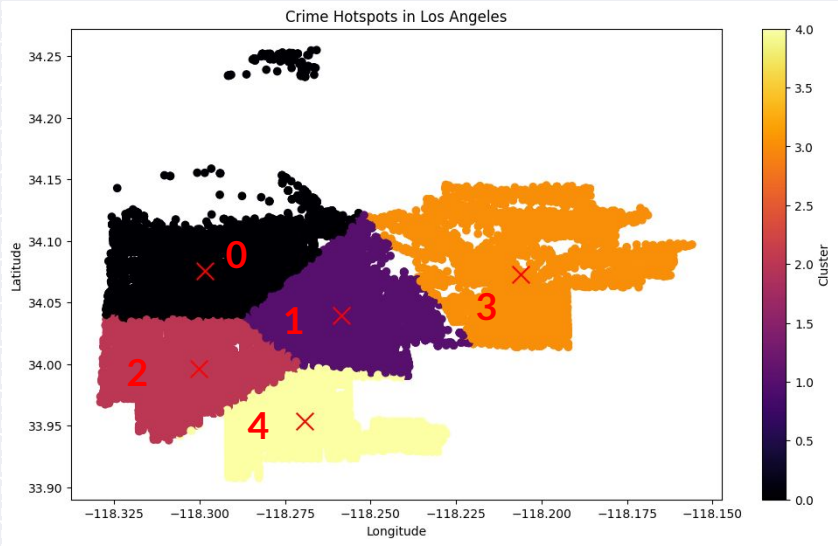
Geographic Distribution of Vehicle Theft



Cluster	Crime Count
0	49816
1	26248
2	5773
3	16711

Geographic Distribution of Vehicle Theft - Cluster 0





Cluster	Crime Count
0	10535
1	12918
2	9322
3	7990
4	9051

Implications and Insights

Insights

We analysed the LA CRIME dataset from 2020-2024:

-We saw a dip in criminal activity during covid however a subsequent rebound in crime reports after the BLM protests, then a flattening

-We found the most influential factor in predicting whether or not someone would be a victim of a crime was their race and tangentially that the younger you were the more likely you were to be a victim of a violent crime.

-For non personal crime we discovered that vehicle theft was the most overwhelming crime committed. Se we delved further into that.

-We discovered that the map of vehicles thefts correlates strongly with LA's local poverty rates, the highest of each occurring the most in West Hollywood.

Implications

After the peak of the BLM protests in 2020, there was a notable increase in crime rates. This escalation was particularly pronounced in the most economically disadvantaged areas of LA, with a significant portion of the crimes attributed to youth. Based on these observations, we propose the following hypotheses:

1. The intensified policing of inner city communities post-protests may have contributed to a surge in reported crimes, thereby inflating the overall crime statistics.
2. The aftermath of the protests might have fueled heightened civil unrest, leading to increased criminal activities, community instability, and a rise in theft incidents.
3. The current spike in crime could be a transient phenomenon linked to challenging economic circumstances experienced by many, suggesting a temporary uptick rather than a sustained trend.

Overall there is a disturbing trend the youth in west LA (near Hollywood) and being a victim of criminal activity

Areas of Improvement

Hyperparameter tuning

We could have incorporated data from different sources to enrich the diversity of the dataset, potentially reducing biases and improving model generalization.

Additional dataset

When conducting our analysis on crime density, we had to use a small sample size (<50,000) or else the notebook crash.

K means regions

Only looked into the subregion that had the most crimes, could have done a more thorough analysis on more of the subregions, and then compared the results to the districts designated by the dataset.

Limitations

Contains only Victims

The model does not contain any information on perpetrators of crimes only the victims so we cannot analyze any inter-sexual or inter-racial dynamicas at play.

Time Frame too short

We had to limit ourselves because the dataset is only four years old thus we cannot properly analyze seasonal or yearly patterns with in various crimes

Differences in area

Our geographic representation of LA does not account for each district whose officers may treat and report various crimes at different rates.

Potential future work

If This dataset continues to be contributed to we could:

- Analyze the trend of crimes over years instead of months or days to see if it is cyclic
- Use outside datasets to determine to what degree LA as a microcosm is representative of California or even the entire US in terms of crime.
- Use the longitude and latitude parameters to determine the most common areas for various crimes (burglary, assault, GTA) and determine why they are so prevalent in those areas and not the others.
- Determine whether the current distribution of LAPD officers is most effective in combating crime where it's located.

Challenges and Limitations

Underfitting

We experienced underfitting when looking at the correlation between violent crimes and victim age

Random sampling

When conducting our analysis on crime density, we had to use a small sample size (<1000) or else the notebook crash.

Overrepresentation

The analysis of smaller population sizes, such as those of Native Americans and Pacific Islanders, often suffers from inaccuracies because their limited numbers can disproportionately impact statistical outcomes.

Resources

Dataset:

- https://catalog.data.gov/dataset/crime-data-from-2020-to-present