

Los Angeles Crime Trends from 2020 to 2024

DATA 450 Capstone

Nikita Gerzhgorin

February 9, 2024

1 Introduction

Knowing the dynamics of urban safety and security requires an analysis of crime trends. One of the most populated and culturally diverse cities in the country, Los Angeles, has long struggled with different ranges of criminal activities. This project aims to delve into the crime trends of the city of Los Angeles from 2020 to 2024 for better understanding the city's security concerns and closely examining the growth of crime in the city over time.

2 Dataset

The dataset "Crime Data from 2020 to Present" was obtained from the [Los Angeles Open Data Website](#). Provided by the Los Angeles Police Department, it has over 890,000 rows of crime incidents that were reported in the city from 2020 till the present time. It is prepared by transcribing the crime reports typed on paper into a dataset format and is continuously being updated every two weeks by the city's police department.

The columns in the dataset that I plan to use are:

- Date Rptd: Date when the incident was reported
- DATE OCC: Date when the incident occurred
- TIME OCC: Time of occurrence
- AREA NAME: Name designation of the LAPD Geographic Area
- Crm Cd: The code that indicates the type of crime committed
- Crm Cd Desc: The definition of the Crime Code provided
- Vict Age: Age of the victim
- Vict Sex: Gender of the victim
- Victim Descent: Ethnicity of the victim
- Premise Cd: The code representing the type of location where the crime occurred

- Premise Desc: The definition of the Premise Code provided
- Weapon Used Cd: The code representing the type of weapon used in the crime
- Weapon Desc: The definition of the Weapon Used Code provided
- Status: Status of the case
- Latitude: Latitude coordinate of the crime location
- Longitude: Longitude coordinate of the crime location

3 Data Acquisition and Processing

Some of the columns like Victim's Descent are in a character format will be mapped to their actual values using the crime report pdf provided in the website. The date and time of occurrences will be combined into a single date time object and the date reported will be converted to a date object before analyzing the data. For the machine learning part, the categorial columns like victim's age and sex will be encoded into numbers and the description columns that have codes already associated with them will be dropped.

4 Research Questions and Methodology

1. How is the trend in criminal incidents in the city? A time series plot by month for each year will be plotted in a single line plot.
2. What is the distribution of crime based on geographical area within the city? A choropleth map will be plotted to visualize the distribution of crimes using the latitude and longitude columns for different areas of the city.
3. What is the demographic of the victims? A histogram will be plotted showing the distribution of victim age and two bar charts will be plotted for gender and ethnicity.
4. What kind of weapons are mostly used in the crimes? A bar chart displaying the frequency of different types of weapons used in crimes will be plotted.
5. What are the statuses of the crimes reported? A pie chart that shows the distribution of different statuses of the crimes will be plotted.
6. How has the frequency of crimes for different gender groups changed over time in different geographical area? A heatmap with a time slider will be plotted to show the change in frequency of crime reportings for different gender groups.
7. Are certain police regions more good in solving crimes? A stacked bar chart showing the resolved and unresolved crimes will be shown for the top 10 police geographical regions with the highest number of crime reportings.

8. What are the most common premise-related words? A word cloud showing the most common words in the premise description will be plotted.
9. Can the age of the victim be predicted based on the types of crime incidents and geographical regions? Three common machine learning algorithms (Linear Regression, Decision Trees, and Random Forest) will be used to identify the accuracies with the age of the victim can be predicted. A bar chart will be plotted showing the time and the accuracy of each machine learning algorithm.

5 Work plan

Week 4 (2/12 - 2/18):

- Preprocessing (3.5 hours).
- Question 1 (3.5 hours).

Week 5 (2/19 - 2/25):

- Question 2 (3.5 hours).
- Question 3 (3.5 hours).

Week 6 (2/26 - 3/3):

- Question 4 (3.5 hours).
- Question 5 (3.5 hours).

Week 7 (3/4 - 3/10):

- Presentation prep and practice (4 hours)
- Question 6 (3.5 hours).

Week 8 (3/11 - 3/17):

- Poster prep (4 hours)
- Presentation peer review (1.5 hours)
- Question 7 (1.5 hours).

Week 9 (3/25 - 3/31): *Final Poster due Sunday 3/31.*

- Peer feedback (3.5 hours)
- Poster revisions (3.5 hours)

Week 10 (4/1 - 4/7):

- Question 7 (1.5 hours).

- Question 8 (4.5 hours).

Week 11 (4/8 - 4/14):

- Question 9 (7 hours).

Week 12 (4/15 - 4/21):

- Revision of Question 1 - 8 (5 hours).
- Revision of Question 9 (2 hours).

Week 13 (4/22 - 4/28): *Blog post draft 1 due Sunday night 4/28.*

- Draft blog post (4 hours).
- Revision of Question 9 (3 hours).

Week 14 (4/29 - 5/5):

- Peer feedback (3 hours)
- Blog post revisions (4 hours)
- [Do not schedule any other tasks for this week.]

Week 15 (5/6 - 5/12): *Final blog post due Weds 5/8. Blog post read-throughs during final exam slot, Thursday May 9th, 8:00-11:20am.*

- Blog post revisions (2 hours)
- Peer feedback (2 hours)
- [Do not schedule any other tasks for this week.]

6 References

Los Angeles Crime Data from 2020 to Present: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data