Crime in Los Angeles

An Analysis from 2010 to 2023

Matthew Biller
University of Colorado

Problem Statement

With a population of 3.849 million people, Los Angeles, California is a vibrant and unique city. Home to numerous communities and neighborhoods, LA offers a home for anyone, bringing communities and nationalities together from across the world.

However, the City of Angels also has a dark side. With a crime rate of 36 per one thousand residents, Los Angeles has the highest crime rate in the United States compared to communities of any size [1].

This research project, covering the timeframe of 2010 to 2023, will be an analysis conducted to analyze all documented crimes that occurred. The intent is to uncover knowledge on the types, frequency, and location of crimes occurring within this 13-year timeframe and determine if any interesting information can be concluded from the research.

What knowledge is being looked for?

To generate any interesting conclusions from the data mined on historic crime in LA, interesting questions must be asked first:

- What crimes occurred the most between 2010 and 2023? The least?
- Was there a change in the type of crime that occurred the most/least frequently over a period of time? Every 5 years?
- What areas in Los Angeles had the most criminal activity?
    - What were those crimes?
    - What was the demographic of these areas?
- What factors may have contributed to certain crimes in certain areas?

Answering these questions will require determining if any interesting patterns can be found in the crime data, which may lead to a greater understanding of crime patterns as a whole.

Another interesting topic that will be explored is COVID-19 and the impact of a global pandemic on the crime in Los Angeles:

- Did COVID-19 change the frequency of any crimes?
- Did the crime rate in Los Angeles become higher? Lower?
- Did certain neighborhoods see a change in criminal activity with COVID-19?

While this data will be analyzed as a whole, there will also be a focus on the timeframe of 2020–2022 to determine if COVID-19 had any significant impacts on crime in Los Angeles.

Key Takeaways

The most important takeaways for this project will be:

- To determine if any interesting information, patterns, statistical evaluations, or connections can be concluded from the Los Angeles crime data.
- Evaluate crime in neighborhoods
- Understand if COVID-19 impacted crime and created any unique findings
- To develop a greater understanding of criminal patterns in major cities.

Literature Survey

DataLA:

An interesting study from the data team of Karen Bass, Mayor of Los Angeles, was found using the same dataset to explore crime trends in Los Angeles from 2010 to 2019. The study takes a dive into the crime rates per year for the specified timeframe, focusing on crimes per 10,000 residents.

An interesting thing to note is that this study has a primary focus on the demographics of victims, focusing on the change in crime rates with respect to the ethnicity of victims, and the patterns that emerged over the nine-year period.

This study also mentions COVID-19 and a general acknowledgment that COVID-19 may have impacted the crime rates of 2020 and 2021, however COVID-19 was only briefly mentioned in this regard. [2]

CAP Index:

CRIMECAST, a company that has developed a crime index scoring system called CAP, published their findings for the city of Los Angeles on the changes of crime rate at the beginning of COVID-19.

While the scope of their study is limited to only retail theft and to March of 2020, detailed percentages of various crime types in relation to retail theft are noted, with a week-by-week breakdown of percentage change. A general summary is also included as to why COVID-19 may have caused these changes in crime rates related to retail theft. [3]

Proposed Work

Data Cleaning:

The first step of the data cleaning process will be the merging of two datasets into one. The datasets for this project are currently divided into 2010-2019, and 2020-2023. The two sets will be merged together and presented as one set in assorted order from 2010 to 2023.

It has been noted in the dataset that null locational attributes are labeled as "0°, 0°". These values will be replaced with locations from the same neighborhood, block, street, cross street, etc. if possible, otherwise all null locations will be evaluated separately from the crime entries with locations, but will still be considered in the overall analysis.

Data Preprocessing:

Part of the data preprocessing that will occur is determining which attributes of the dataset are necessary. Several "code" attributes (city codes or codes from the Los Angeles Police Department) will not be relevant to this study and will be ignored.

Another import preprocessing step will be to ensure that the provided Lat/Lon are able to be used in reference for mapping and visual representation. The locations provided are only to the block-level to maintain privacy.

Data Grouping:

Determining groupings for crime entries will be a very important part of the data preprocessing and overall processing. Data entries may be grouped by year, type of crime, location, severity, etc., most likely several groupings will emerge to allow for a deeper evaluation of the data for finding patterns and interesting information.

Differences in Research:

While the DataLA study does use the same dataset, key differences include only using the 2010–2019 dataset with no use of the newer 2020–2023 dataset. Some comparisons to the changes in crime rates will be made between this research project and my own, mainly to cross-reference percent changes year-over-year.

This research also includes demographics of victims, which will not be a factor in my research project. I will only be focusing on the "neighborhood-level" or higher.

While the CRIMECAST analysis does look into COVID-19 and its' impact on crime rates in Los Angeles, they focus their findings to only retail crime and within the first four weeks of COIVD-19, March

2020. My research will involve various crime types and span a thirteen-year period.

Data Set:

Title: "LA Crime Data"

Type: CSV

Provided By: Kaggle (link)

With ~2.7 million entries, this dataset provides incidents of crime between 2010 to 2023 in Los Angeles, CA. All data entries were transcribed from original crime reports.

The dataset provides 28 different attributes that include items such as date, location, type of crime, type of weapon, severity, victim information, etc.

Evaluation Methods:

To fully evaluate the data for interesting patterns and relationships between attributes, several evaluation methods will be used:

Metrics:

Percentage change will be one of the most important metrics to calculate. Crime types will be viewed at various temporal levels (month-to-month, year-to-year, and year binning) to gain insight into the occurrences of crimes, their changes over time, and patterns that may emerge from that.

Clustering:

Clustering will be used to determine if similar crimes occurred in certain neighborhoods or areas around Los Angeles. Identifying hotspots of criminal activity could return interesting results, those hotspots will then be analyzed over different periods of time to look for patterns or changes in behavior of the types of crimes being committed throughout the city.

Data Visualization:

An overlay of world imagery will be used to demonstrate visually where crimes, or groupings of crimes, are occurring throughout Los Angeles. Imagery will be provided by Maxar/Google Earth and will have points, as well as hotspot areas overlaid as a layer to show where crimes occurred.

Tools:

Python will be used as the codebase for the data preprocessing and data analysis, packages Numpy and Panads will be used for statistical evaluations.

Mapbox will be used to visually represent the crime data, overlaid with satellite imagery. Data points and heat maps will be built with Mapbox's API and included in the final report to visualize emphasize the occurrence of crimes.

Milestones:

In three weeks: I would like to have the majority of my data analysis complete. This will involve organizing crime types, performing statistical analyses, and determining any clustering in the data.

In four weeks: I intend to have the majority of my analysis complete, I will spend this week looking for any/all interesting informational and connections that are a result from my research.

In five weeks: I would like to have the data visually represented to be included in my final report.

In six weeks: I intend to have my final report through the rough draft stage and nearing completion. I will continue refining my report, building my final presentation, and prepare my final submission for the project.

REFERENCES

[1] Neighborhood Scout, Los Angeles, CA Crime Rate (link)

[2] DataLA, 2022 A Data-Driven Exploration of Crime Trends in Los Angeles. Medium.com. Link

[3] CAP Index, 2020. COVID-19 & Crime - CAP's Perspective on crime & Loss in the Age of COVID-19. CAPINDEX. Link

Link to GitHub