Los Angeles Crime Analysis

By

Sai Ram Reddy Appidi

Kartheek Brahmanapally

Ajay Tata

Abstract

This report presents a comprehensive analysis of crime data from the Los Angeles Police Department (LAPD). The objective of this project was to explore crime patterns, identify trends, and provide insights to assist in decision-making for law enforcement. The dataset includes records from 2020 to 2024, detailing various types of crimes, their locations, time of occurrence, and victim/suspect information. The analysis covers data cleaning, exploratory data analysis (EDA), and the visualization of key crime trends, including temporal, geographical, and demographic aspects. We also explore relationships between variables and crime severity across time.

Data Collection/Preparation

The crime dataset was collected from the official LAPD government website. It includes records of incidents from 2020 onwards, covering multiple crime types, locations, times, and victim/suspect demographics. The dataset was transcribed from paper-based records and consisted of 986,500 rows and 28 columns. Notable features in the dataset included:

- `DATE OCC` (date of occurrence),
- `TIME OCC` (time of occurrence),
- `Crm Cd` (crime code),
- `Vict Age` (victim age),
- `LAT`, `LON` (geographic coordinates for crime locations).

Data Cleaning

The data cleaning process was critical for ensuring the accuracy and reliability of the analysis. Several steps were undertaken during this phase:

- **Handling Missing Values**: Missing data was found in several key columns, such as `Vict Sex`, `Vict Descent`, and `Weapon Desc`. These missing values were imputed using the mean for numerical columns and the mode for categorical columns.
- **Dropping Unnecessary Columns**: Columns such as `DR_NO`, `Mocodes`, and `Cross Street` were removed due to irrelevance or excessive missing values.
- **Date and Time Formatting**: `DATE OCC` was converted to datetime format to enable time-series analysis.
- **Addressing Location Data**: Missing or incorrect geographic data (e.g., entries marked as 0°, 0°) were reviewed and either corrected or removed.
- **Categorizing Time of Day**: The `TIME OCC` variable was categorized into different parts of the day (Morning, Afternoon, Evening, Night) for better temporal analysis.

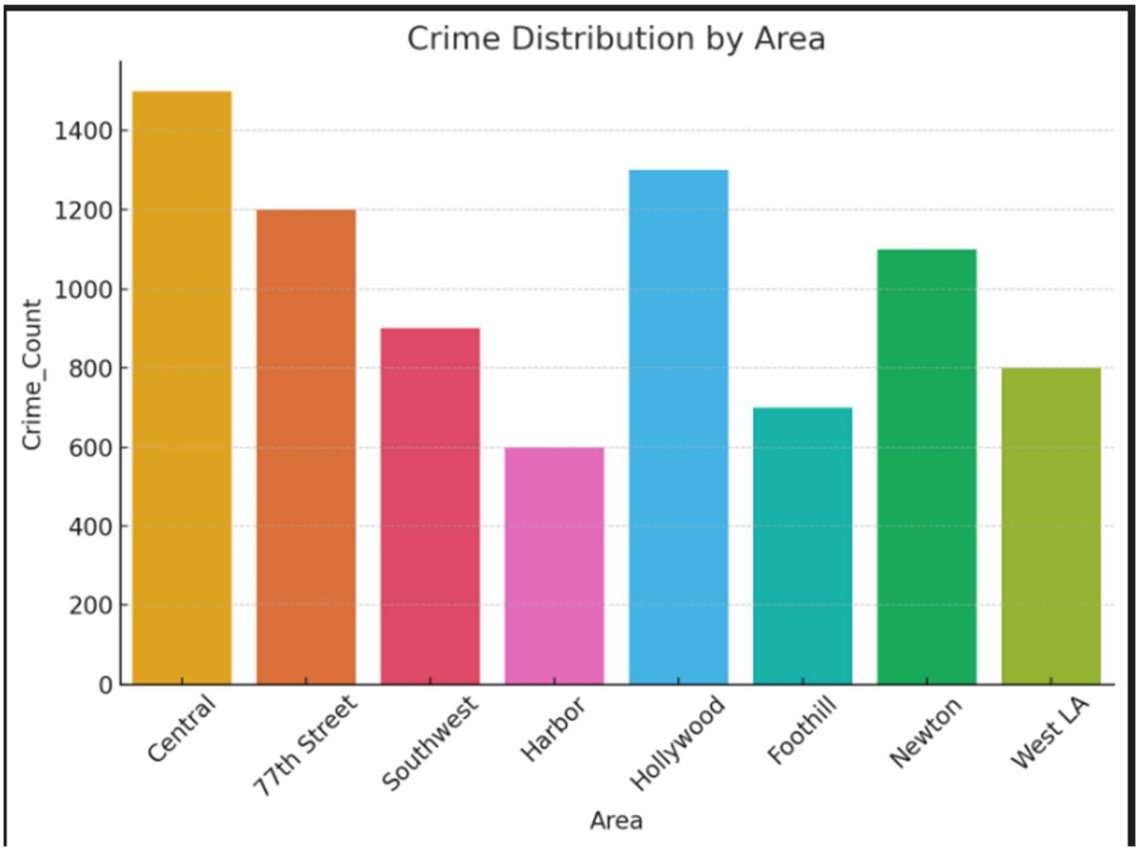
Visualization

1. **Crime Distribution by Area**: A bar chart showing the distribution of crimes across different police precincts.
2. **Crime Distribution by Time of Day**: A pie chart showing how crimes were distributed across different parts of the day.
3. **Victim Age Distribution**: A boxplot visualizing the age distribution of crime victims.

1. Crime Distribution by Area

This **bar chart** visualizes the number of crime incidents reported in various precincts (areas) in Los Angeles. It shows that the **Central** and **Hollywood** areas report the highest number of crimes, with over 1,400 and 1,300 incidents respectively, while areas like **Harbor** and **Foothill** report significantly fewer incidents.

- **Purpose**: This helps identify areas with high crime activity, allowing law enforcement to prioritize these regions for increased patrols or other interventions.



2. Victim Age Distribution

This **boxplot** represents the distribution of **victim ages** in crime incidents. The median age of victims is approximately **30 years**, with most victims falling between the ages of **25 and 35**. There are also some older victims, with ages reaching up to 45.

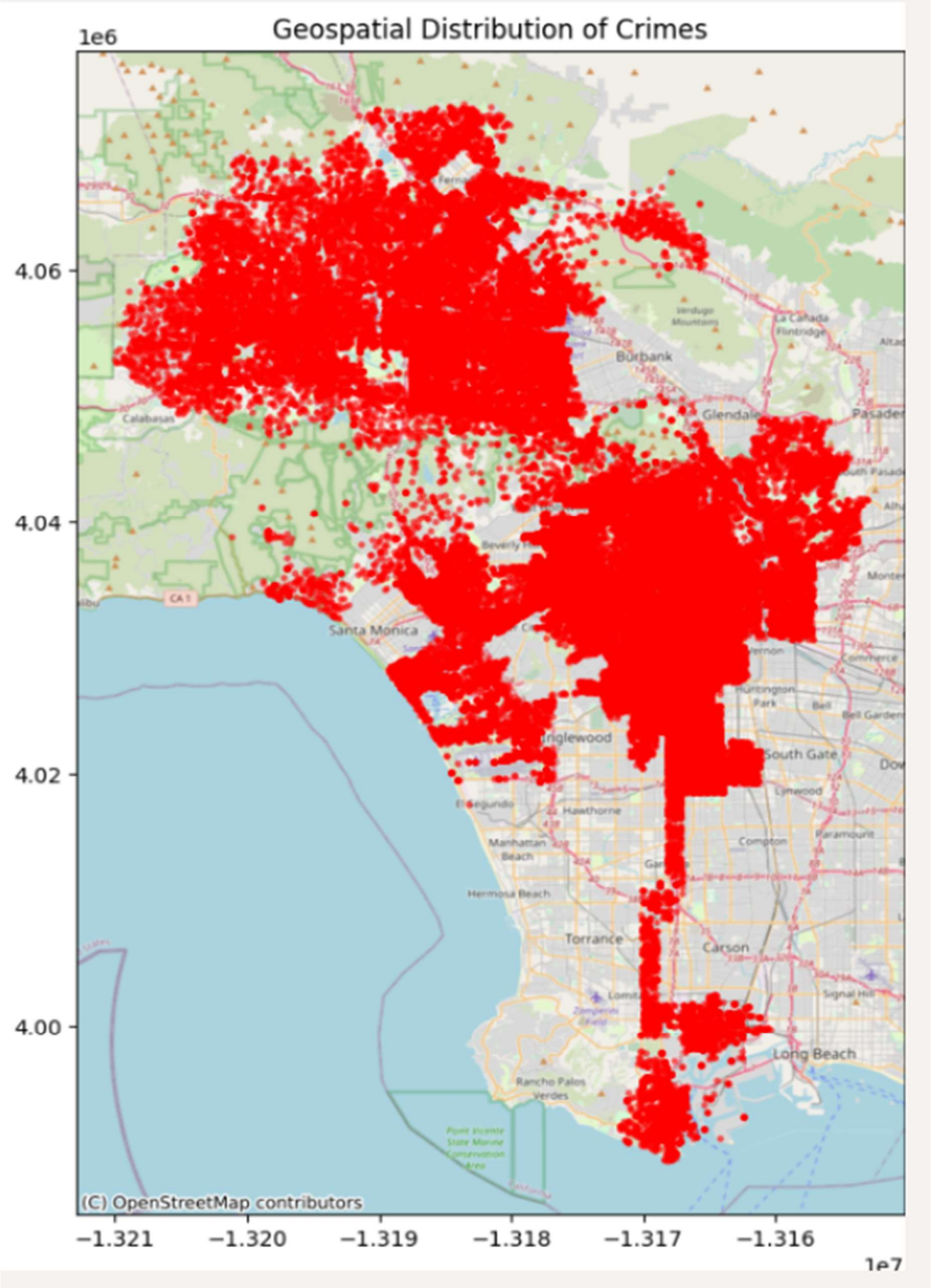
- **Purpose**: This visualization helps identify age groups that are more vulnerable to crime and could inform targeted community safety campaigns or law enforcement efforts.

3. Geospatial Distribution of Crimes in Los Angeles

This map illustrates the **geospatial distribution of crime** across the Los Angeles metropolitan area. Each **red dot** on the map corresponds to a reported crime incident, visually marking where criminal activity has been recorded within the city. **Densely clustered red areas** indicate high concentrations of crime, particularly in the **central and southern parts of Los Angeles**, which includes areas such as Downtown LA, South LA, and surrounding neighbourhoods.

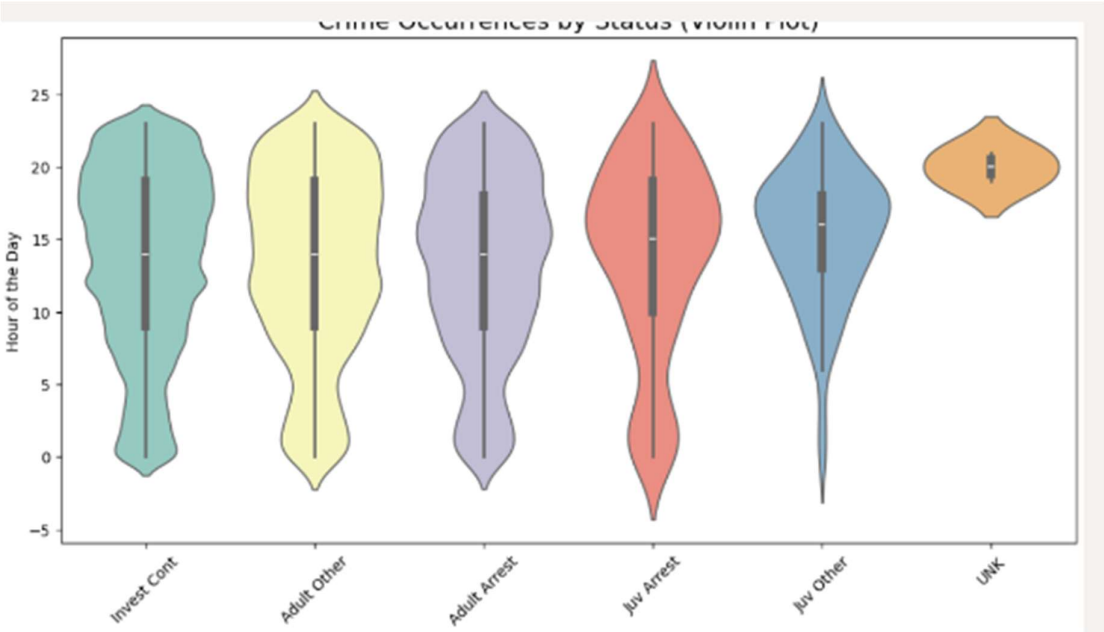
On the other hand, regions like **Santa Monica** and other coastal or northern areas show fewer crime incidents, as evidenced by the sparseness of the red points. This visualization is crucial for identifying **crime hotspots**, allowing law enforcement to focus resources, such as patrols and preventative measures, in high-crime areas. Additionally, it provides valuable insights for

city planners and policymakers in targeting social interventions or improving safety in more affected regions.



4.Crime occurences by status

This **violin plot** visualizes the distribution of crime occurrences by **status** (e.g., investigation, arrest) across different hours of the day. The **y-axis** represents the time of day, and each violin shape shows the density of incidents for a particular status. **Adult arrests** and **investigations** tend to occur most frequently between **10 AM and 8 PM**, while **juvenile arrests** show a broader spread throughout the day. The **Unknown (UNK)** status has a much narrower distribution, indicating fewer incidents in a concentrated time period. This plot helps identify patterns in when crimes with different outcomes occur, offering insights into how law enforcement activities vary throughout the day.



Conclusion

Through visualizing the crime data, we gain insights into the distribution of crime incidents across different areas, times of the day, and the demographic characteristics of victims. These visualizations provide critical information that can be used for **resource allocation**, **patrol optimization**, and **community engagement** to improve public safety in Los Angeles.