Crime Patterns and Victim Demographics in Los Angeles

Emmanuel Mokua

@03085667

Howard University

MATH-014: Introduction to Data Science

Dr. Meenakshi Nerolu

April 18, 2025

## **Objective:**

This project aims to explore, clean, and visualize a crime dataset from Los Angeles in order to identify key trends, victim demographics, crime hotspots, and relationships between variables. Through thoughtful data wrangling and storytelling via visualizations, we uncover meaningful insights to better understand public safety patterns in the city.

## **Introduction & Dataset Reference:**

The dataset used in this project is sourced from the Los Angeles Open Data Portal and contains detailed records of reported crimes in Los Angeles, including fields such as victim age, sex, descent, crime description, weapon used, location coordinates, time of occurrence, and arrest status. This analysis uses a cleaned subset of the data focused on core features relevant to visual exploration and storytelling.

## **Method – Data Wrangling & Handling Missing Values:**

1. **Invalid Entries Removed:**
   * Victim age values equal to **0** were dropped, as they likely represented missing or placeholder values.
   * Null values in columns such as Weapon Desc were also filtered out when analyzing weapon usage.
2. **Redundant Columns Identified:**
   * Highly correlated fields such as Crm Cd and Crm Cd 1, and AREA vs. Rpt Dist No, were flagged as duplicates for potential removal.
   * Redundant columns such as Crm Cd 1, Crm Cd 2, Crm Cd 3, Crm Cd 4 were dropped.
3. **Data Type Verification:**
   * Numerical and categorical fields were reviewed to ensure appropriate usage for correlation analysis and grouping.

## **Storytelling – Data Visualization & Interpretation:**

* **Victim Age Distribution (Histogram & Boxplot):** Most crime victims fall between the ages of **20 and 40**, with a noticeable right-skew in the distribution. Outliers over age 70 were observed, possibly related to elder-targeted fraud or abuse.
* **Top Crimes (Bar Chart):** Theft, battery, and vandalism emerged as the most frequently reported crimes. A small number of crime types accounted for a large portion of all incidents.
* **Victim Demographics (Count Plot – Victim Sex):** The majority of victims were **male**, followed by females. A small number of entries marked as “X” indicated non-binary or unknown identities.
* **Average Victim Age by District (Grouped Bar Plot):** Differences in average victim age across districts revealed that **central and downtown areas** had younger victims, while **West LA** and **Pacific** districts had older averages.
* **Geographic Crime Distribution (Scatter Plot):** Incidents were densely clustered around **Downtown Los Angeles (LAT ~34.05, LON ~–118.25)**, confirming it as a crime hotspot.
* **Weapon Usage (Pie Chart):** A surprising 64% of incidents involved **no weapon**, while 25% involved **bodily force** (“Strong-Arm”). Weaponized crimes (handguns, knives) were rare but likely more severe.
* **Correlation Heatmap:** Most numeric variables showed **weak or no correlation**, except for perfect or near-perfect correlations among redundant fields. This highlights the importance of feature reduction and the potential value of categorical/non-linear analysis.

## **Conclusion / Summary of Insights:**

* Crimes are **most concentrated during evening and early afternoon** hours.
* **Downtown LA** is the central hotspot for criminal activity, based on geographic clustering.
* **Young to middle-aged adults** (20–40) are most frequently affected.
* **Male victims are more common**, and males are especially more prone to battery/assault cases.
* **Most crimes involve no weapon or bodily force**, not firearms or knives.
* **High arrest rates are more likely in violent crimes** such as manslaughter or homicide.
* Redundant features such as AREA and Rpt Dist No were identified through correlation analysis and should be dropped in modeling workflows.

This project demonstrates how visual analysis can reveal both expected and surprising patterns in real-world crime data, supporting more focused public safety strategies.

## **References:**

* Los Angeles Open Data Portal: <https://data.lacity.org/>
* <https://seaborn.pydata.org/api.html>
* <https://matplotlib.org/stable/index.html>
* Howard University – MATH 014 Course Resources

## **Acknowledgements:**

* Sophia Hurtado-Hernandez
* Saiida Simone Webb

Your feedback and collaboration made this project fun and insightful.