

## ASSIGNMENT SUBMISSION COVER SHEET

Student Id: bida23-211

Student name: Junior Mosimanenkwe

Student email: [bida23-211@thuto.bac.ac.bw](mailto:bida23-211@thuto.bac.ac.bw)

Cohort: November 2024

Assignment title: LA Crime Data 2020-2024

Date of submission: 11 November 2024

Programme of Study: BSC BUSINESS INTELLIGENCE & DATA  
ANALYTICS

Year of Study: YEAR 2

Intellectual property statement

By checking the box below, I certify that this assignment is my own work and is free from plagiarism. I understand that the assignment may be checked for plagiarism by electronic or other means and may be transferred and stored in a database for the purposes of data-matching to help detect plagiarism. The assignment has not previously been submitted for assessment in any other unit or to any other institution. I have read and understood the Botswana Accountancy College plagiarism guidelines policy.

☒ Agree

Signature: j.mosimanenkwe

Date: 11 November 2024

# Los Angeles Police Department Crime Analysis Report

## Overview

The purpose of this research is to give the Los Angeles Police Department (LAPD) a thorough understanding of the patterns and trends in crime in Los Angeles from 2020 to the present as stated by (Thompson, M., 2023). The main goal is finding the most prevalent crime categories, crime hotspots, and trends that might support focused interventions. According to Ooijen, et al., (2019), the aim is to provide the department with data-driven insights for better resource allocation and enhanced public safety.

## Dataset Description

Saurabh Badole generated and oversaw the "Los Angeles Crime Incidents: Incidents from 2020 to Present" dataset, which is accessible on Kaggle and contains the data used in this research. This dataset offers a comprehensive summary of crime incidences in the city during the past several years as suggested by (Wang, et al., 2016), including LAPD crime reports with information on victim demographics, crime type, location, date, and time.

## Importance of Analyzing the Data

Finding patterns and trends in this information is crucial for strategic planning aimed at preventing and combating crime.

## Key Variables Extracted from the Dataset

### 1. Crime Description (Crm Cd Desc)

- **Relevance:** This variable indicates the type of crime committed, such as violence, robbery, etc. This helps determine which crimes are more prevalent and which ones require focused response, as well as understanding the distribution and frequency of various crimes, as stated by Paynich, et al., 2013.
- **Implications:** Understanding the different types of crime is essential for resource allocation as stated by Farmer, (2013). For example, policing violent crimes may need a different strategy compared to policing property crimes. Knowing these distinctions allows the police

to prioritize higher-risk crime categories, assign specialist teams, and predict crime patterns using past data.

## **2. AREA NAME (Neighborhood or Area Name)**

- **Relevance:** This variable indicates the location of a crime, according to Aksoy (2017). It is crucial for identifying hotspots and trends in different communities. This variable is helpful for tracking changes in crime density across Los Angeles and identifying high-crime areas.
- **Implications:** Geographic analysis enables departments to develop preventative strategies, launch community initiatives targeted at high-crime areas, and concentrate resources where crime rates are higher. This information is particularly valuable for fostering community partnerships and proactive policing in affected neighborhoods.

## **3. Time of Occurrence (TIME OCC)**

- **Pertinence:** TIME OCC records the specific hour a crime was committed. This is crucial for studying trends in criminal activity over time. By identifying the times of day and night when crime is most frequent, police departments can adjust patrol schedules accordingly as stated by (Srinivasan, et al., 2013).
- **Significance:** Time-based data is vital for optimizing police operations, increasing visibility, and improving response times. Understanding peak crime times facilitates the reduction of crime during vulnerable periods and enhances service performance.

## **4. Victim Origin (Ethnic Heritage of the Victim)**

- **Relevance:** This variable records the victim's ethnicity, classified as O (Other), W (White), B (Black), H (Hispanic), etc. It helps identify whether certain demographic groups are disproportionately affected by crime and aids in designing targeted and equitable prevention strategies.
- **Significance:** Analyzing victim demographics allows agencies to pinpoint which populations are at risk and whether a specific community is more impacted by a type of crime. This is essential for preventing socioeconomic disparities in victimization, customizing outreach, and promoting fair resource allocation as stated by Hobbs

(2010). It also helps build trust, address community-specific concerns, and strengthen relationships with various groups.

## 5. Weapon Description

- **Significance:** This variable provides details about the type of weapon used in the crime, such as GUN, KNIFE, etc. Understanding weapon usage is crucial for analyzing the nature and severity of crimes, particularly violent crimes. Data on weapon use informs the assessment of public safety risks, especially related to violent crime, as indicated by Donohue, et al., (2019).
- **Implications:** High-crime areas may need increased police presence, community safety programs, or support for stricter gun control measures. This data also informs training on handling firearm-related incidents and enhances risk assessments for responding officers.

## Data Errors Identified and Cleaning Methods

### 1. Missing Data

- **Description:** Key columns such as Victim Age, Victim Descendants, AREA NAME, and Weapon Description had missing data. Missing values can distort analysis outcomes and lead to inaccurate insights if the data is predominantly incomplete, according to Enders (2022).
- **Method of Cleaning:**
  - **Identification:** For variables like victim age, averaging was used to replace missing values.
  - **Mode Imputation:** The most common parent category was used to fill in gaps for Victim Descendants.
  - **Excluded Rows:** Rows with excessive missing values were removed, and the Weapon Description column was dropped to maintain data integrity.

### 2. Outliers in Victim Age

- **Description:** Outliers included exceptionally young or extremely old ages (e.g., less than five years or more than 100 years) as indicated by Aguinis, et al., (2013), which skew age-related analysis.
- **Method of Cleanup:**

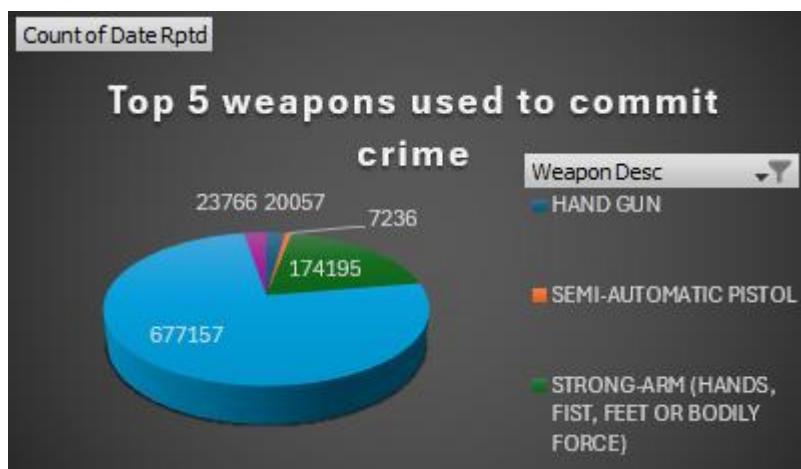
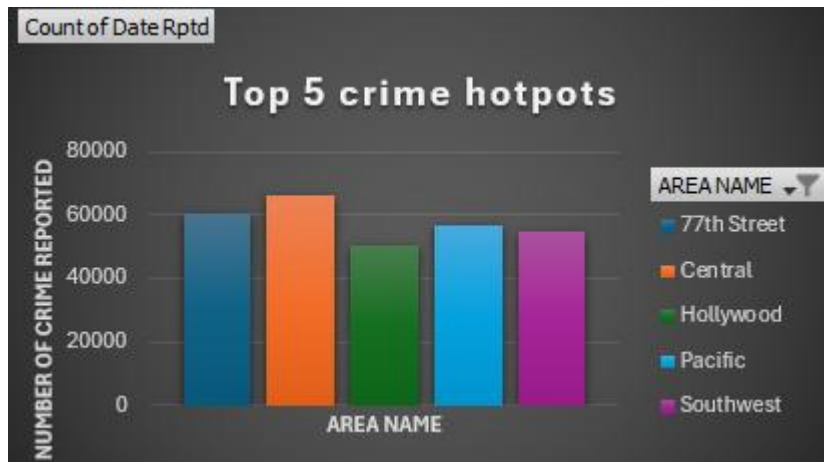
- **Filtering Values:** Ages not within a reasonable range (e.g., under five years or over 100 years) were excluded.
- **Conditional Filtering:** Used to limit the data to a realistic range.

### 3. Placeholder Coordinates

- **Description:** Some rows had placeholder values for location data, with LAT and LON set to 0, as stated by McGlinn, et al., (2021), indicating incorrect or unavailable data. These placeholder values create false hotspots and skew spatial analysis.
- **Method of Cleanup:**
  - **Location Imputation:** The average coordinates for the relevant region (34.0589, -118.3225) were used to replace placeholder values.

### c) Visualisation charts





#### d) Dashboard



## REFERENCES

- Aguinis, H., Gottfredson, R.K. and Joo, H., 2013. Best-practice recommendations for defining, identifying, and handling outliers. *Organizational research methods*, 16(2), pp.270-301.
- Aksoy, E., 2017, October. Geography of crime and its relation to location: the city of Balıkesir (Turkey). In *IOP Conference Series: Materials Science and Engineering* (Vol. 245, No. 7, p. 072012). IOP Publishing.
- Donohue, J.J., Aneja, A. and Weber, K.D., 2019. Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis. *Journal of Empirical Legal Studies*, 16(2), pp.198-247.
- Enders, C.K., 2022. *Applied missing data analysis*. Guilford Publications.
- Farmer, D.J., 2013. *Crime control: The use and misuse of police resources*. Springer Science & Business Media.
- Hobbs, R., 2010. *Digital and Media Literacy: A Plan of Action. A White Paper on the Digital and Media Literacy Recommendations of the Knight Commission on the Information Needs of Communities in a Democracy*. Aspen Institute. 1 Dupont Circle NW Suite 700, Washington, DC 20036.
- McGlinn, K., Brennan, R., Debruyne, C., Meehan, A., McNerney, L., Clinton, E., Kelly, P. and O'Sullivan, D., 2021. Publishing authoritative geospatial data to support interlinking of building information models. *Automation in Construction*, 124, p.103534.
- Paynich, R. and Hill, B., 2013. *Fundamentals of crime mapping: Principles and practice*. Jones & Bartlett Publishers.
- Srinivasan, S., P. Sorrell, T., Paul Brooks, J., J. Edwards, D. and Diehl McDougale, R., 2013. Workforce assessment method for an urban police department: Using analytics to estimate patrol staffing. *Policing: An International Journal of Police Strategies & Management*, 36(4), pp.702-718.
- Thompson, M., 2023. *AN EXPLORATORY ANALYSIS OF LOS ANGELES POLICE TRAFFIC STOPS* (Doctoral dissertation, California State Polytechnic University, Pomona).

van Ooijen, C., Ubaldi, B. and Welby, B., 2019. A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance.

Wang, H., Kifer, D., Graif, C. and Li, Z., 2016, August. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 635-644).