

# Visualization of Crime Distribution in Los Angeles (2020-2024) Using Power BI

*This project investigates the spatial and temporal distribution of crimes in Los Angeles from 2020 to 2024 using Power BI. By analyzing this data, we aim to identify crime patterns and hotspots, assisting the Los Angeles Police Department (LAPD) and city planners in resource allocation and policy formulation to enhance public safety.*

## 1 Introduction

Urban safety is a critical concern for both residents and authorities in large cities like Los Angeles. A strong sense of security significantly impacts the quality of life, economic stability, and overall well-being of urban populations. When urban safety is robust, residents feel secure in their daily activities, businesses thrive, and communities flourish. High levels of public safety lead to increased investment, tourism, and a higher standard of living as people are more willing to engage in social, recreational, and economic activities without fear.

Conversely, weak urban safety can have severe repercussions. High crime rates lead to fear and anxiety among residents, deterring investment and stifling economic growth. Communities with persistent safety issues often experience social fragmentation, where mistrust and fear prevent cohesive community development. This can result in deteriorating neighborhoods, reduced property values, and increased costs for law enforcement and healthcare due to crime-related injuries and stress.

Understanding crime distribution helps in making informed decisions about policing and community interventions. Effective crime prevention strategies rely on accurate data and analysis to identify patterns and hotspots, allowing for efficient resource allocation and proactive measures. This project leverages crime data from Los Angeles over a five-year period (2020-2024) to create a comprehensive map visualization using Power BI, showcasing the geographical and temporal trends of criminal activities. The insights derived from this project will aid the LAPD in optimizing patrol routes, deploying resources effectively, and implementing preventive measures to improve public safety in Los Angeles.

## 2 Data Preparation

**2.1 Data Overview.** The dataset includes detailed records of crimes reported in Los Angeles from 2020 to 2024. Each entry in the dataset contains information such as

- **DR\_NO (Division of Records Number):** This is an official file number that consists of a 2-digit year, an area ID, and 5 digits. It uniquely identifies each record in the dataset.
- **Date Rptd (Date Reported):** The date when a crime was reported to the police.
- **Date Occ (Date Occurred):** The date when the actual incident or event occurred.
- **Time Occ (Time Occurred):** The time when the incident occurred, represented in 24-hour military time.

- **Area:** Refers to the LAPD's 21 Community Police Stations, which are sequentially numbered from 1 to 21. Each area covers specific neighborhoods or regions within Los Angeles.
- **Area Name:** The name designation given to each Geographic Area or Patrol Division within the LAPD. These names often reference landmarks or the communities they serve.
- **Rpt Dist No (Report District Number):** A four-digit code representing a sub-area within a Geographic Area. It provides more granular information about the location of the crime within a larger police division.
- **Crn Cd (Crime Code):** Indicates the specific crime committed. Each crime is assigned a code for categorization and tracking purposes.
- **Crn Cd Desc (Crime Code Description):** Provides a description or definition of the crime based on its code.
- **Moscodes (Modus Operandi):** Describes the activities associated with the suspect during the commission of the crime, such as methods or behaviors.
- **Vict Age (Victim Age):** The age of the victim involved in the crime.
- **Vict Sex (Victim Sex):** Indicates the gender of the victim (F for Female, M for Male, X for Unknown).
- **Vict Descent (Victim Descent):** Descent codes categorize the ethnicity or descent of the victim.
- **Permis Cd (Premise Code):** Describes the type of structure, vehicle, or location where the crime occurred.
- **Permis Desc (Premise Description):** Provides a description corresponding to the Premise Code, explaining the type of location or structure involved.
- **Weapon Used Cd (Weapon Used Code):** Indicates the type of weapon (if any) used during the commission of the crime.
- **Weapon Desc (Weapon Description):** Describes the type of weapon specified by the Weapon Used Code.
- **Status:** Indicates the current status of the case.
- **Status Desc (Status Description):** Provides a description corresponding to the Status Code, explaining the current status of the case.

- **Crm Cd 1, Crm Cd 2, Crm Cd 3, Crm Cd 4:** These fields may contain additional crime codes if multiple crimes were committed in a single incident. Crm Cd 1 is typically the primary and most serious offense, while the others are less serious offenses.
- **LOCATION:** Provides the street address of the crime incident, rounded to the nearest hundred block to maintain anonymity.
- **Cross Street:** Indicates the cross street associated with the rounded address of the crime incident.
- **LAT (Latitude):** The geographical latitude coordinate of the crime incident location.
- **LON (Longitude):** The geographical longitude coordinate of the crime incident location.

**2.2 Data Cleaning & Preprocessing.** The data cleaning and preprocessing steps were crucial to ensuring the accuracy and reliability of our analysis.

- **Removing Irrelevant Columns:** Columns such as 'DR\_NO', 'AREA', 'AREA NAME', 'Part 1-2', 'Crm Cd', 'Mocodes', 'Premis Cd', 'Weapon Used Cd', 'Crm Cd 1', 'Crm Cd 2', 'Crm Cd 3', 'Crm Cd 4', 'Cross Street', 'LAT', 'LON' were removed as they did not contribute significantly to our analysis of crime distribution.
- **Data standardization:** The 'Date Rptd', 'DATE OCC', and 'TIME OCC' columns were standardized to ensure consistency and facilitate accurate time series analysis.
- **Validating Location Data:** The 'Lat' and 'Lon' columns were validated for accuracy, ensuring that all crimes were correctly mapped to their respective coordinates.
- **Handling Missing Values:** Any missing or null values within essential columns were addressed through appropriate imputation techniques or data exclusion where necessary to maintain data integrity.

### 3 Visualization in Power BI

**3.1 Importing Data.** The cleaned dataset was imported into Power BI. Proper data modeling was ensured by establishing accurate relationships between tables.

**3.2 Designing the Dashboard.** A user-friendly dashboard was designed to present the data interactively.

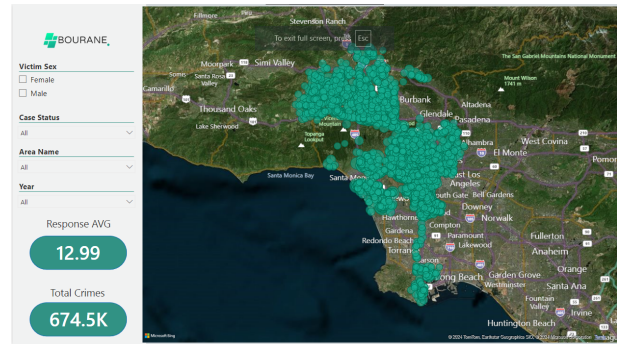


Fig. 1 Main Dash



Fig. 2 Details Dash

The "Main Dash" shows a map of Los Angeles, showcasing the distribution of crimes across the city. Blue circles of varying sizes overlay the map, representing the total crimes in different areas. The left panel offers filters for "Victim Sex," "Descent," "Case Status," and "Area Name." Notably, the visualization allows for interactive analysis based on demographics and geographic regions, providing valuable insights for policy planning and resource allocation.

The "Details Dash" becomes visible when stakeholders seek specific information about each bubble on the map, providing comprehensive insights. The top left section presents a line graph depicting the total number of crimes recorded each month. Meanwhile, the top right section showcases a donut chart categorizing crimes by victim sex, with segments representing "Male" and "Female," aiding in the analysis of gender-specific crime patterns. On the left side, a bar graph visually represents different crime statuses, offering a clear overview. Additionally, the right side features a horizontal bar graph listing various weapons associated with crimes, enhancing the understanding of crime dynamics and weapon usage trends. Furthermore, a horizontal bar graph appears when stakeholders click on the "Premise" button, providing insights into various premises associated with the crimes.

## 4 Conclusions

This project successfully demonstrated the use of Power BI for crime data visualization, providing valuable insights into the distribution of crimes in Los Angeles from 2020 to 2024. The interactive dashboard allows for continuous exploration and analysis, assisting city planners and policymakers in making informed decisions to enhance public safety.

By understanding the distribution and patterns of crimes through effective visualization, stakeholders can implement targeted interventions and allocate resources efficiently, ultimately improving the safety and quality of life for all residents in Los Angeles.

## References

- [1] <https://medium.com/towards-data-science/predicting-airbnb-prices-with-machine-learning-and-location-data-5c1e033d0a5a>
- [2] [https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about\\_data](https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data)
- [3] <https://courses.dataschool.io/view/courses/pandas-in-30-days/>
- [4] [https://www.youtube.com/watch?v=OQuKwy6eEDkab\\_channel=Curbal](https://www.youtube.com/watch?v=OQuKwy6eEDkab_channel=Curbal)
- [5] <https://www.youtube.com/watch?v=XZOPriif6Pwt=1021s>
- [6] <https://www.youtube.com/watch?v=BbplhqDCWOMt=392s>
- [7] <https://www.youtube.com/watch?v=A95nqPLdmRUt=87s>
- [8] [https://www.youtube.com/watch?v=CynnJ04xy8klist=PLGZUzt4E4O2J3Aoetrs\\_N2Za1b3a8JJCDab](https://www.youtube.com/watch?v=CynnJ04xy8klist=PLGZUzt4E4O2J3Aoetrs_N2Za1b3a8JJCDab)