CS5803 2023/4 Data Visualisation - Coursework Assignment

Student ID: 2306691

Student Name: Mihir Gujarathi

1. Introduction

This dataset contains information about criminal incidents that occurred in Los Angeles. The addresses in the dataset are only provided up to the nearest hundred block to safeguard privacy. The data is owned by Los Angeles Police Department (LAPD). The dataset is sourced from Crime Data from 2020 to Present | Los Angeles - Open Data Portal (lacity.org). Since the original dataset was too big in size, we filtered out Date Reported from 2022 to 2023 to decrease the dataset size.

Column Name	Description	Type
DR_NO	Division of Records Number: Official file number made up of a 2 digit year, area ID, and 5 digits	Plain Text
Date Rptd	Date reported in MM/DD/YYYY format	Date & Time
DATE OCC	Date of occurrence in MM/DD/YYYY format	Date & Time
TIME OCC	Time of occurrence in 24-hour military time format	Plain Text
AREA	LAPD geographic area code from 1 to 21 representing Community Police Stations	Plain Text
AREA NAME	Name designation of LAPD geographic areas or Patrol Divisions	Plain Text
Rpt Dist No	Four-digit code representing a sub-area within a Geographic Area	Plain Text
Part 1-2	Number indicating whether the crime is Part 1 or Part 2	Number
Crm Cd	Crime code indicating the type of crime committed (same as Crime Code 1)	Plain Text
Crm Cd Desc	Description of the crime code	Plain Text
Mocodes	Modus Operandi: Activities associated with the suspect in commission of the crime	Plain Text
Vict Age	Age of the victim in two characters	Plain Text
Vict Sex	Gender of the victim (F - Female, M - Male, X - Unknown)	Plain Text
Vict Descent	Descent code indicating the ethnicity or race of the victim	Plain Text
Premis Cd	Code indicating the type of structure, vehicle, or location where the crime took place	Number
Premis Desc	Description of the premise code	Plain Text
Weapon Used Cd	Code indicating the type of weapon used in the crime	Plain Text
Weapon Desc	Description of the weapon used code	Plain Text
Status	Status of the case (IC is the default)	Plain Text
Status Desc	Description of the status code	Plain Text
Crm Cd 1	Primary crime code indicating the most serious offense	Plain Text
Crm Cd 2	Additional crime code, less serious than Crime Code 1	Plain Text
Crm Cd 3	Additional crime code, less serious than Crime Code 1	Plain Text
Crm Cd 4	Additional crime code, less serious than Crime Code 1	Plain Text
LOCATION	Street address of the crime incident rounded to the nearest hundred block to maintain anonymity	Plain Text
Cross Street	Cross street of the rounded address	Plain Text
LAT	Latitude of the location of the crime	Number
LON	Longitude of the location of the crime	Number

My persona is a **Crime Journalist** who analysis crime in the city.

Questions and Requirements

SQ1: Which gender (male or female) is involved in more reported crimes in Los Angeles?

R: To determine which gender (male or female) is involved in more reported crimes in Los Angeles, provide a visual comparison using a bar chart. The chart should have gender categories on the x-axis and the count of reported crimes on the y-axis. A simple grouped bar chart will effectively display the comparison between male and female involvement in reported crimes.

SQ2: What is the distribution of reported crimes across different weekdays in Los Angeles?

R: To analyze the distribution of reported crimes across different weekdays in Los Angeles, present a line chart with weekdays on the x-axis and the count of reported crimes on the y-axis. Each line on the chart represents a different weekday, allowing users to observe the variations in crime distribution throughout the week.

SQ3: How does the distribution of reported crimes vary across different geographic areas in Los Angeles?
R: To understand how the distribution of reported crimes varies across different geographic areas in Los Angeles, use a choropleth map. The map should visualize crime counts per geographic area, with colors representing varying levels of crime density. This will allow users to identify areas with higher and lower concentrations of reported crimes.

CQ1: Are there any specific geographic areas in Los Angeles where the distribution of reported crimes differs significantly between genders, and how does this pattern vary over different weekdays?
R: To identify specific geographic areas in Los Angeles where the distribution of reported crimes differs significantly between genders and analyze how this pattern varies over different weekdays, provide an interactive dashboard with linked views. Include a choropleth map showing crime distribution by geographic areas, a grouped bar chart comparing crime involvement by gender, and a line chart displaying crime trends over weekdays. Enable users to select specific geographic areas and weekdays, with dynamic updates across all views to facilitate comparative analysis.

2. Design

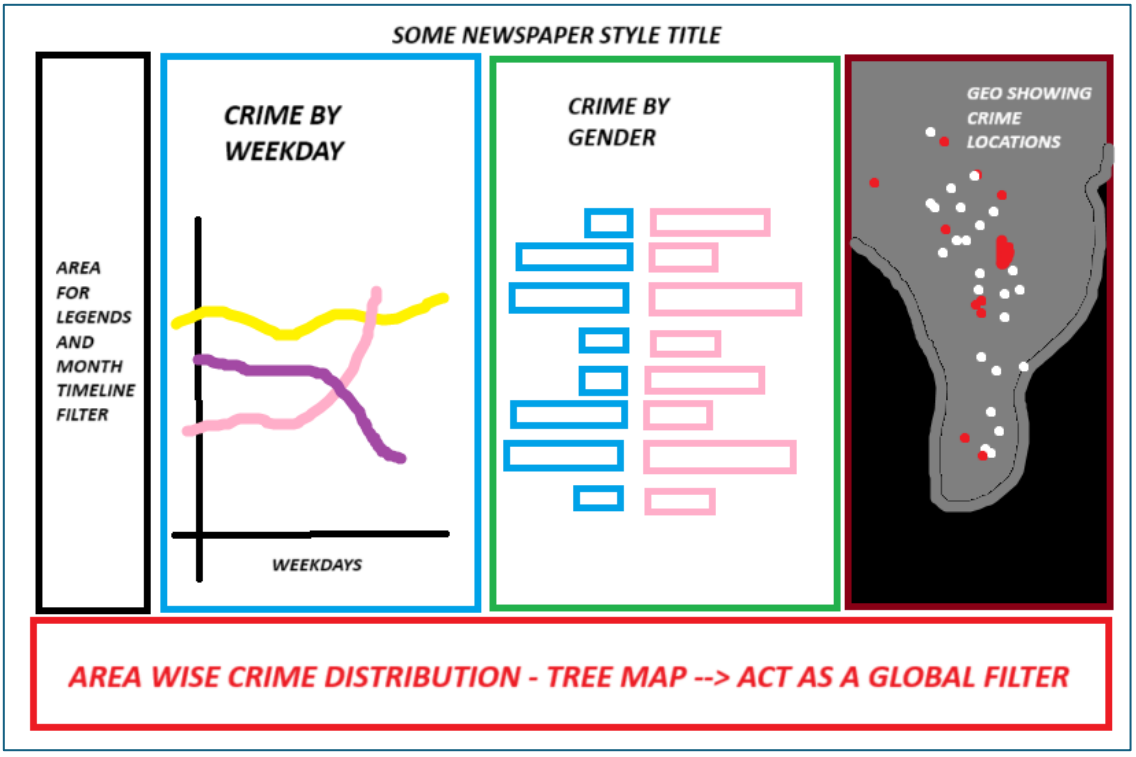


Figure 1: Paper Landscape

a) Area Wise Crime Distribution: As the name suggests, it will be a tree map which shows crime distribution by Area. I intend to keep it as a filter so that we can see area specific details for other graphs. (Q2 and Q4)

b) Crime by Weekday: This will be a line chart with Weekdays on x-axis and number of crimes on y-axis. There will be up to 5 crimes (5 unique lines showing different crimes) to analyze the trend over the week. (Q4)

c) Crime by Gender: This will be a comparison bar chart but not in the usual way. This will be a butterfly chart which will compare victim gender (only male vs female) opposite to each other. On the y-axis, there will be type of crime. (Q1 and Q4)

d) Geo Map showing Crime Location: This is a geo map which shows area of where the crime took place for in depth knowledge of crime. As crime trend changes every month, there will be a month timeline filter to see crime hotspots in different time periods. (Q3)

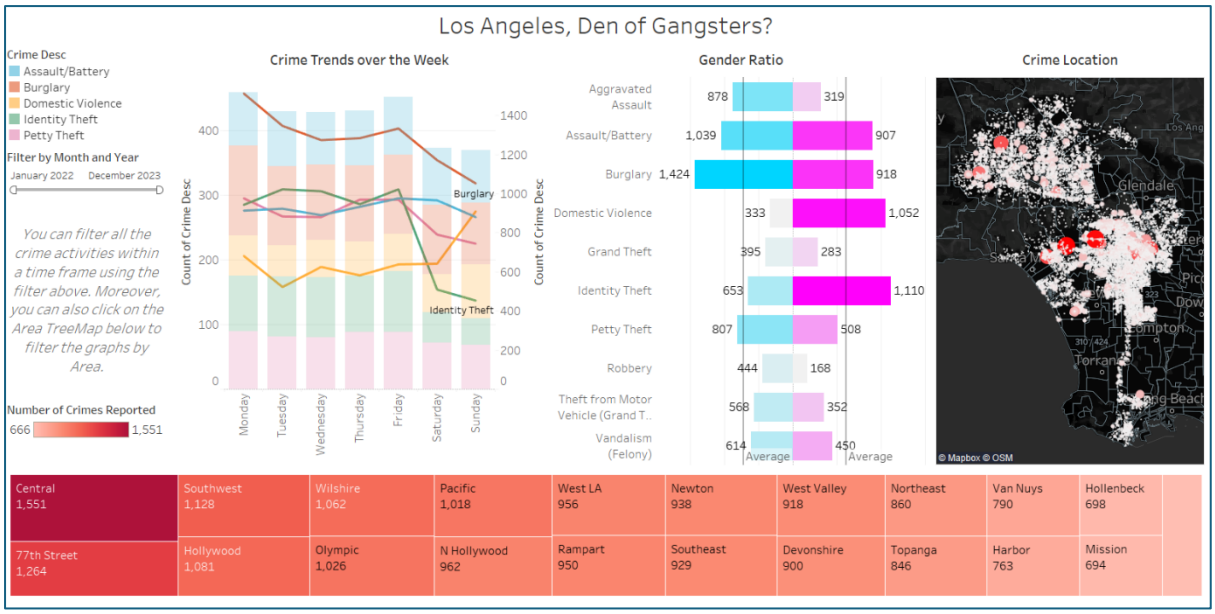


Figure 2: Tableau Dashboard

There is a slight difference between the original paper landscape and the Tableau Dashboard. The difference is that I added stacked bar chart as background for Crime Trends on Weekdays for better understanding of overall crime distribution. Also, as I had extra space in the Legends section, I added simple instruction for user as they might not know that the Area Tree Map acts as a filter.

Furthermore, I also attempted to recreate the dashboard in PowerBI. Due to some graphical formatting limitations, I could not recreate the dashboard exactly but it looks similar to what I had planned in Paper Landscape.

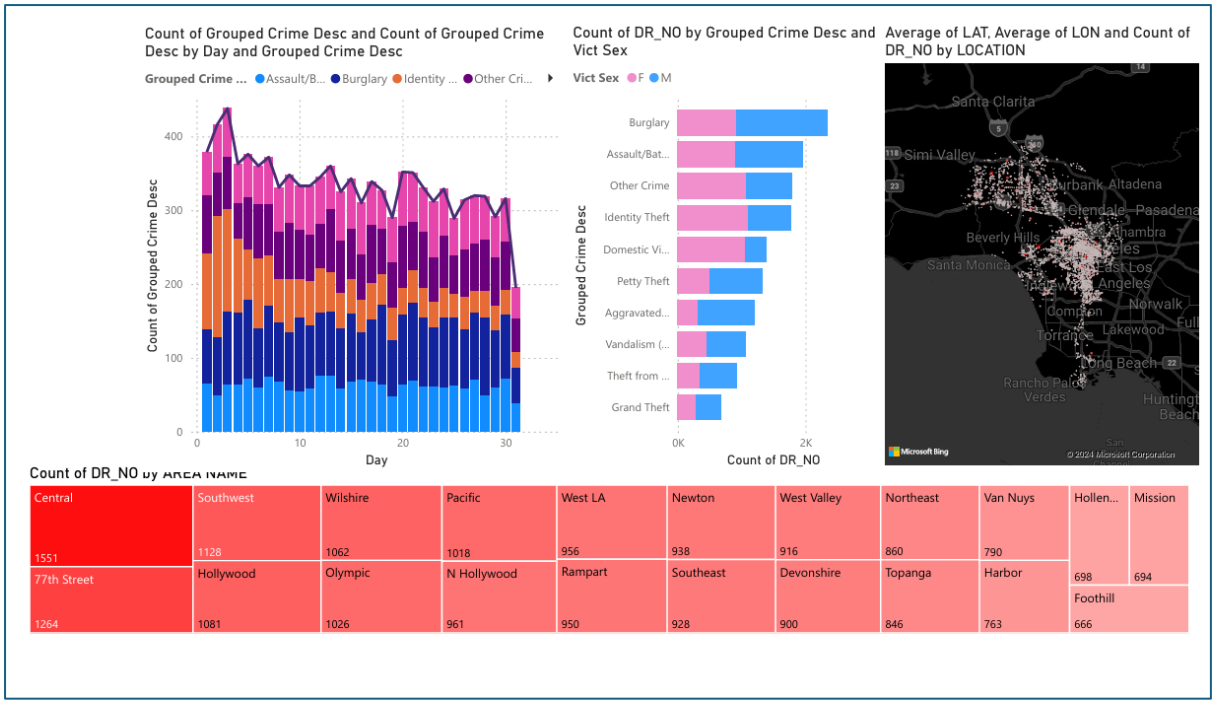


Figure 3: PowerBI Dashboard

3. Implementation

After importing the crime dataset, I made some calculated fields in the dataset. I did this to make my desired charts. There were a lot of similar crime descriptions, I made a new field and grouped some common crime types.

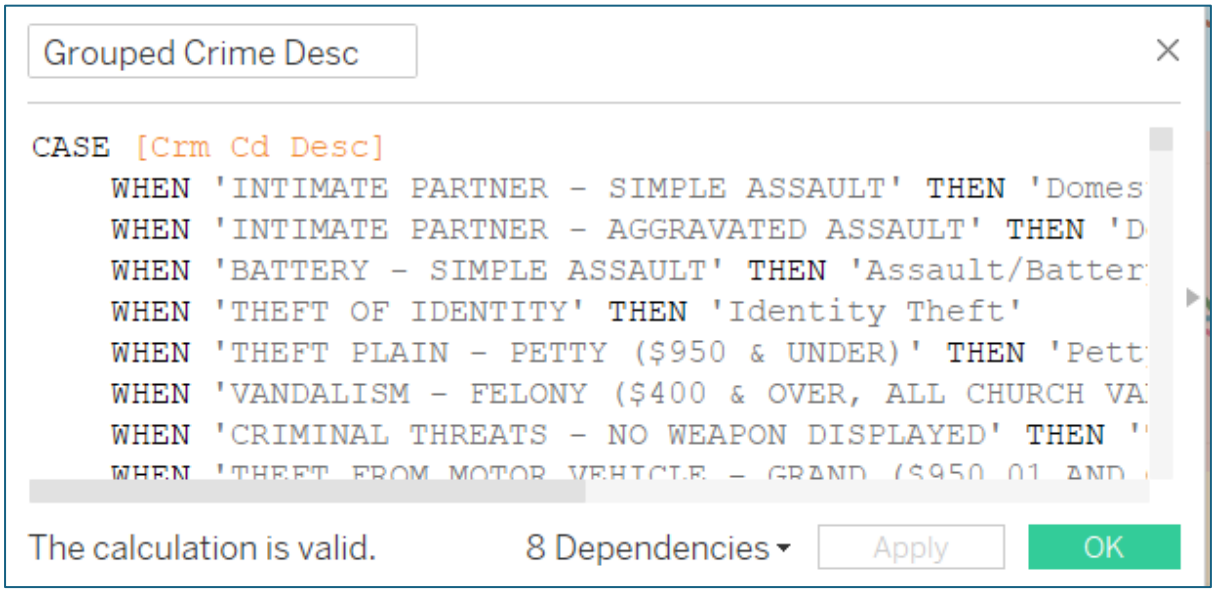


Figure 4: Grouped Crime Description - Calculated Field

After making all the necessary calculated fields for the graphs I planned from my Paper Landscape, I made new worksheet and started plotting the graphs.

a) Crime by Area:

This is the easiest graph of all.

- Add “Area Name” on any axis and “Dr. No” on the other axis. You must get a bar chart view.
- From the Show Me Panel, click on treemaps. You can add “Dr. No” to colors in marks for better visualization. Make sure that Dr.No is set to Measure (Count).

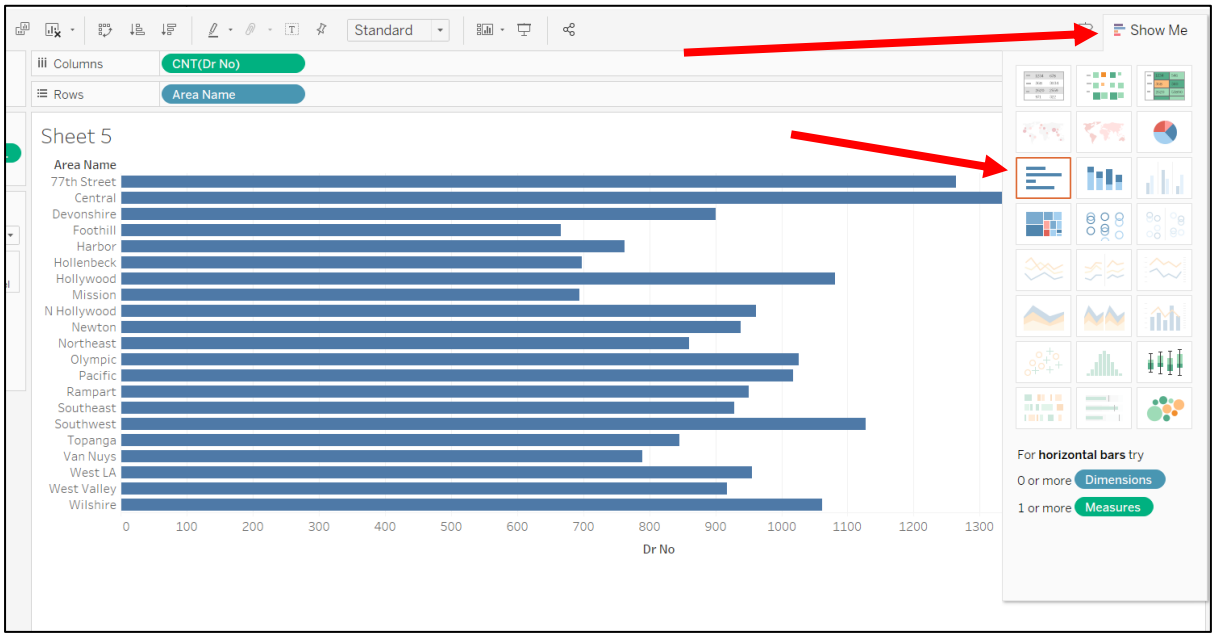


Figure 5: Creating Crime by Area Tree Map

To recreate the Crime by Area graph in PowerBI, simply:

- Drag and drop the "Area Name" field into the Axis area of a new visual.
- Drag and drop the "Dr. No" field into the Values area.

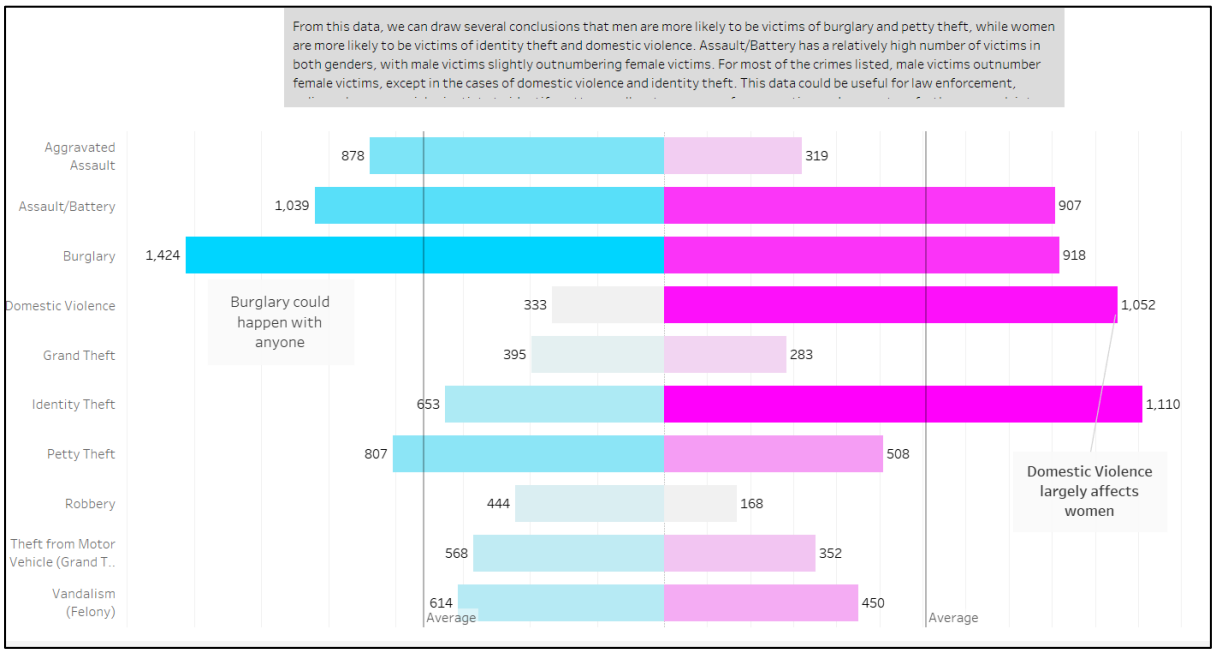
This will create a basic bar chart showing the count of crimes by area.

b) Crime by Location:

To create this graph, simply double click LAT and LONG. They will be assigned to x and y axis accordingly. Then, add “Location” to the detail section in marks. And finally, drag “Dr. No” to color as well as size in marks. Make sure that Dr.No is set to Measure (Count).

You can follow the exact same steps in PowerBI to recreate the geo map.

4. Walkthrough



To answer Q1, we can see that women are involved more towards specific crime types like Domestic Violence and Identity Theft. Men on the other hand are mostly a victim of Aggravated Assault. Burglary seems to be gender irrelevant which means that everyone should be alert.

Mondays have high chances for Burglary, although there is a decreasing trend over the week. Domestic Violence seems to be rising on the weekends suggesting the days of holidays. Identity Theft seems to drop remarkably during the weekends.

To answer my Q3, distribution of crime varies by a lot in Los Angeles. Crime is highly active in the Central and 77th Street, followed by Southwest. All these areas are located near the central. Hence, we can say that the central part of Los Angeles is not safe. On the contrary, areas like Foothill and Northeast are calmer areas.

5. Discussion

Reflecting on the project, it's evident that data visualization tools like Tableau and Power BI offer powerful capabilities for analyzing and presenting complex datasets. These tools enable users to gain insights from data quickly and effectively, allowing for informed decision-making. Throughout this project, I've learned how to manipulate and visualize data to extract meaningful insights, which is a valuable skill in various professional fields.

One strength of these visualization tools is their user-friendly interfaces, which make it easy to create visually appealing and interactive dashboards. Additionally, they offer a wide range of visualization options, allowing users to choose the most suitable representation for their data. However, one limitation is the learning curve associated with mastering these tools, especially for more advanced features and functionalities. Furthermore, while these tools are proficient in handling structured data, they may face challenges when dealing with unstructured or semi-structured data sources.

Looking forward, I aim to further enhance my skills in data visualization and analysis, exploring advanced techniques and methodologies to extract deeper insights from data. Additionally, I

intend to explore other data visualization tools and platforms to broaden my toolkit and adapt to different project requirements and scenarios. Overall, this project has been a valuable learning experience, highlighting the importance of effective data visualization in understanding complex datasets and driving informed decision-making processes.

6. Conclusion

In conclusion, this project aimed to analyze crime data in Los Angeles to uncover patterns and trends using Tableau where similar design is recreated in PowerBI. The solution approach involved creating visualizations for crime by gender, crime trends over weekdays, and crime distribution across geographic areas. Through this analysis, it was observed that women are more involved in specific crime types like Domestic Violence and Identity Theft, while men are often victims of Aggravated Assault. Furthermore, Mondays showed a higher likelihood of Burglary, with a decreasing trend over the week, while Domestic Violence appeared to rise on weekends. In terms of geographic distribution, areas like Central and 77th Street exhibited higher crime activity, highlighting the need for increased attention to safety measures in these regions. Conversely, areas like Foothill and Northeast were comparatively calmer. These insights provide valuable information for law enforcement agencies and policymakers to develop targeted crime prevention strategies and enhance public safety measures in Los Angeles.