Exploring Crime Patterns in Los Angeles City: A Statistical Analysis of November 2023 Data

Authors

Saketh Saridena
Naryana Kaushik Mantha
Vineel Rayapati
Saketh Reddy Dodda
Navya Chalamalasetty
Kundan Sai Chowdary Sannapaneni

December 9, 2023

Table of Contents

Abstract

Our paper presents the dynamics of crime in Los Angeles City focusing specifically on the crimes that took place in November of 2023. The analysis involves detailed statistical analysis, statistical visualization, and hypothesis testing to identify data trends using a comprehensive dataset that was obtained from the Los Angeles City government. Understanding the categorical, temporal, and spatial dimensions of criminal activity is the main objective of the research. Our study uses a dataset with both quantitative and qualitative variables, and it applies a number of statistical techniques to glean insights. The results not only provide a detailed portrayal of crime dynamics but also highlight the significance of temporal variations and spatial concentrations. The findings have implications for resource allocation for focused crime prevention initiatives as well as law enforcement tactics.

1 Introduction

To create a safer neighborhood across the vast expanse of Los Angeles, it is essential to comprehend the complex patterns of criminal activity. Our research explores the complex field of criminal data analysis with a particular emphasis on November 2023's colorful cityscape. Recent research on crime trends in Los Angeles and California reveals concerning patterns. Violent crime has risen 12

Crime is a widespread social issue that affects people on a daily basis and extends beyond statistics. In order to understand the intricacies of crime dynamics, the central research issue of this study looks at the locations and times at which events are most likely to happen. The strong desire to improve the effectiveness of crime prevention initiatives and get a more sophisticated understanding of the variables influencing criminal activity is what motivates this investigation. In Los Angeles, certain offenses like homicide and motor vehicle theft are increasing amidst an overall downward crime rate [2].

Our study attempts to offer a current and pertinent snapshot that can assist in real-time decision-making by concentrating on data from November 2023. This research is conducted against the backdrop of a public domain dataset that painstakingly records criminal events in Los Angeles. This dataset, which includes both quantitative and qualitative factors, acts as the canvas on which the research creates a thorough image. The core of the inquiry is the complex interaction of variables including time, place, and types of crimes. It is important that we recognize the precedents that earlier studies in the field of crime analysis set before we begin our journey through the data. Our objective of this project is to enhance the existing conversation in the field and offer practical solutions for tackling current issues by building upon the collective expertise in the field.

In the upcoming sections, we will delve into crime patterns in Los Angeles. Through the analysis of statistics, the presentation of charts, and the testing of theories, we aim to identify significant trends. By thoroughly examining the data, we aim to enhance our understanding of crime in our city.

2 DATA

2.1 Data Description

According to the Los Angeles Police Department, biases can distort crime reporting and perceptions in different areas [3]. According to the UNODC, economic disparities often impact crime rates, so we must remain aware of potential data inaccuracies [4]. We analyze data on Los Angeles crime incidents to understand public safety issues. The November 2023 data covers many types of reported crimes based on official police records. Each crime has a "dr_no" number to uniquely label it in the reports. This lets us refer to specifics in tracking different criminal acts. In total, the dataset gives a broad picture of the current state of unlawful activity across the city.

2.2 Data Collection Methodology

As the LAPD open data team describes, organized capture of granular time and location details enables methodical analysis of crime patterns within LA communities [5]. Geo-tagged images further assist in monitoring geographic distributions. Key information about every crime incident is noted in the data. This comprises the time of day ("time occ"), the date and time of the incident ("date occ"), and the report date ("date rptd"). The location columns include the reporting district number ("rpt dist no"),

the area code ("area"), and the area name ("area name"). The geo-tagged photos aid in monitoring the locations of crimes in different parts and communities of Los Angeles. An organized method of capturing this data allows for the methodical examination of crime patterns according to certain time and place details. Tracking patterns inside particular districts as well as across areas yields granular information.

2.3 Bias Considerations

Despite the dataset's best efforts to provide complete coverage of crimes in Los Angeles, biases may still exist. The community's cooperation with the police, the goals of law enforcement, and regional economic disparities can all affect the amount of crime that is reported. Data completeness and accuracy may be impacted by this. Moreover, reporting practices probably vary by area, which might distort perceptions of crime everywhere. When examining trends, we should be aware of these potential data distortions.

2.4 Key Features for Analysis

The analysis primarily relies on several key features within the dataset:

- Part 1-2 ("part 1-2"): This binary indicator distinguishes between different levels of severity in reported crimes, essential for understanding the gravity of incidents.
- Crime Code ("crm cd") and Crime Code Description ("crm cd desc"): These columns provide a categorization of the criminal activities reported, offering a detailed insight into the nature of each incident.
- Geographic Coordinates ("lat" and "lon"): Latitude and longitude values pinpoint the precise locations of reported crimes, facilitating spatial analysis and mapping.
- Time of Occurrence ("time occ"): This temporal feature allows for the exploration of patterns related to the timing of criminal activities.
- Additional Crime Codes ("crm cd 1-4"): These columns capture supplementary crime codes, potentially indicating multiple offenses within a single incident.
- Location Details ("location" and "cross street"): Specific information about where the crime occurred enhances the contextual understanding of each incident.

The collective utilization of these features empowers a holistic analysis, shedding light on temporal, spatial, and categorical dimensions of crime dynamics in Los Angeles.

3 METHODS

3.1 Exploratory Data Analysis (EDA)

Our team performed an exploratory data analysis (EDA) to understand the structure and properties of our dataset before beginning the analysis. This required examining summary statistics, identifying important features, and creating distributional visualizations. The discovery of patterns, trends, and possible outliers in the crime data was the main objective of the EDA phase.

3.1.1 Dealing with Missing Values

Taking care of missing values was essential to maintaining the accuracy of our dataset. Columns that had a large number of missing entries, like "Weapon Used CD," were carefully handled. We used different imputation techniques according to the type of missing data. For example, to preserve information about the lack of weapons, we substituted "No Weapon Used" for missing values in the "Weapon Used Cd" column.

3.1.2 Unwanted Column Removal

To make the dataset more manageable, we chose to eliminate any columns that didn't seem necessary or redundant for our analysis. The removal of columns such as "DR_NO" and "Crm Cd 2-4" allowed for a more targeted and effective analysis of our data.

3.1.3 Data Cleaning

1. **Handling Categorical Columns**

To handle missing values in our dataset, we processed categorical columns such as "Vict Descent" and "Vict Sex". We decided to use the mode, which represents the most common category, for imputation. We were able to reduce information loss and maintain the categorical nature of the data by using this method.

2. **Numeric Column Imputation**

We used the mean of non-zero values to perform imputation when working with numerical columns like "Vict Age" in our dataset. This tactic was used to preserve the age distribution in its entirety, thereby mitigating the bias caused by zero values.

3. **Text Values Replacement**

In our analysis, we standardized replacements for missing values in certain text columns, including "Weapon Desc" and "Premis Desc." This approach aimed to ensure consistency and facilitate a comprehensive exploration of the dataset.

3.2 Analysis

3.2.1 Correlational Analysis

Previous Los Angeles crime research has uncovered correlations between offense severity, weapon use, and the time of day [6]. Our analysis aims to extend these discoveries. We used correlation analysis in our analysis to find patterns among various variables. The strength and direction of associations were revealed by the correlation matrix. This helped us find possible dependencies, particularly with regard to tempo, severity indicators, and crime codes.

3.2.2 Distribution Comparison

In our analysis, we looked at how crime incidents were distributed throughout the day. We sought to determine whether crime frequency showed a uniform distribution over different time periods using chi-square goodness-of-fit tests.

3.2.3 Hypothesis Testing

Existing analysis has found significant differences in average victim age across crime categories (Phuong, 2023). We leverage hypothesis testing to validate whether similar age variations emerge in our November 2023 LA dataset. In our analysis, hypothesis testing was essential in addressing particular relevant questions. Notably, we investigated differences in the mean victim age across different crime types using a two-sample t-test. Chi-square tests were utilized to analyze variations in the frequency of a particular type of crime between different areas and to gauge how consistently crimes were distributed across time periods.

These techniques were chosen because they were pertinent to the study questions and could offer significant new perspectives on the dynamics of crime in Los Angeles. Exploratory data analysis, data cleansing, and different statistical tests combined to enable a thorough and sophisticated investigation of the dataset.

3.3 Statistical Tests

3.3.1 P- Value test:

The p-value is a statistical measure that helps to determine the significance of the results obtained from a test hypothesis. It calculates the probability of obtaining a result at least as extreme as the one observed, assuming the null hypothesis is true. We use it to decide whether to reject the null hypothesis. The formula is:

$$ p = P(T \geq t | H_0) $$

where T is the test statistic and t is the observed value.

3.3.2 T-test:

The t-test is used to determine if there is a significant difference between the means of two groups, which may be related in certain features. It is a hypothesis test that follows a Student's t-distribution under the null hypothesis. A t-test can tell whether any differences observed between groups in an experiment are likely to be attributable to chance or if they are likely due to the independent variable. The formula for a two-sample t-test is given in LaTeX as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where \(\bar{X}_1 \) and \(\bar{X}_2 are the sample means, \(s^2 \) is the pooled sample variance, and \(n_1 \) and \(n_2 \) are the sample sizes.

3.3.3 Chi-Squared Test

The chi-squared test is a statistical hypothesis test that measures how a model compares to actual observed data. It is used to determine whether there is a significant association between two categorical variables. The formula for the chi-squared test statistic in LaTeX is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where \(O_i \) represents the observed frequency, \(E_i \) represents the expected frequency under the null hypothesis, and the summation \(\sum \) is taken over all possible outcomes or categories. This test statistic follows a chi-squared distribution with \((r-1)(c-1) \) degrees of freedom, where \(r \) is the number of rows and \(c \) is the number of columns in the contingency table.

4 RESULTS

4.1 Visualizations

4.1.1 Top 5 Crime Codes with Descriptions

The top 5 crime codes along with their descriptions are highlighted in our visualization. The most common ones were "Assault With Deadly Weapon," "Burglary from Vehicle," "Theft From Motor Vehicle," and "Vehicle - Stolen."

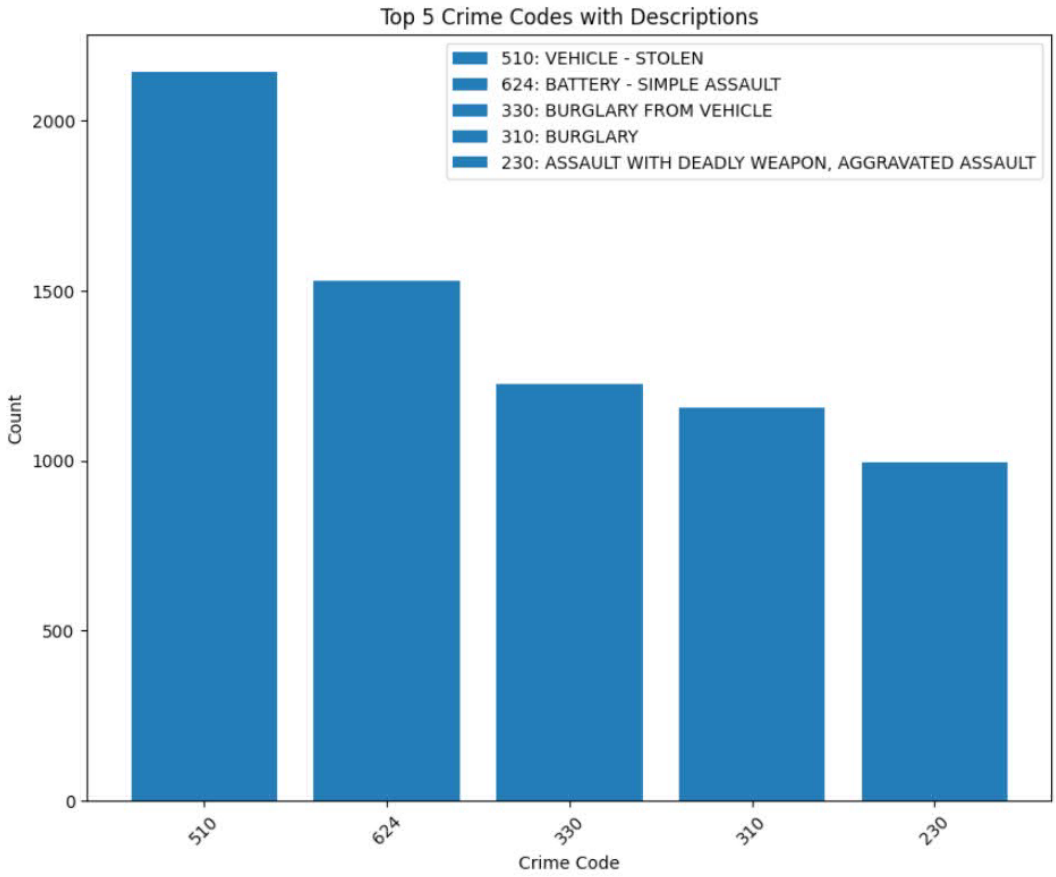

Figure 1: This bar plot shows top five crimes with crime codes

4.1.2 Crime Frequency by Date

The line plot in our analysis showed the frequency of crimes by date in November 2023. Interestingly, the dataset showed monthly variations in the number of crimes committed, offering a temporal perspective on criminal activity.

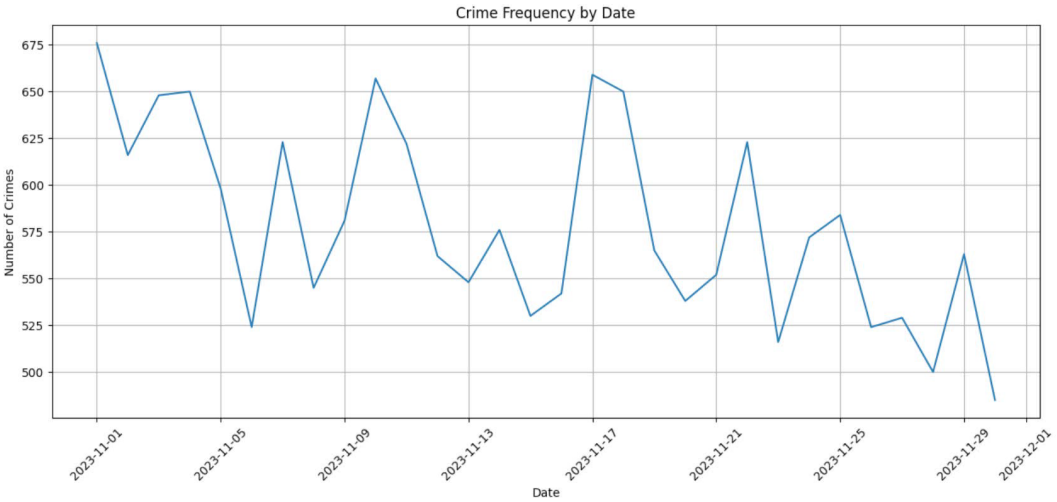

Figure 2: This line plot shows crime frequency by date

4.1.3 Crime Frequency by Time and Area

The distribution of crimes by area and time of day was depicted on the heatmap. Crime rates were

higher at night, especially in some areas, which suggests temporal and spatial patterns.

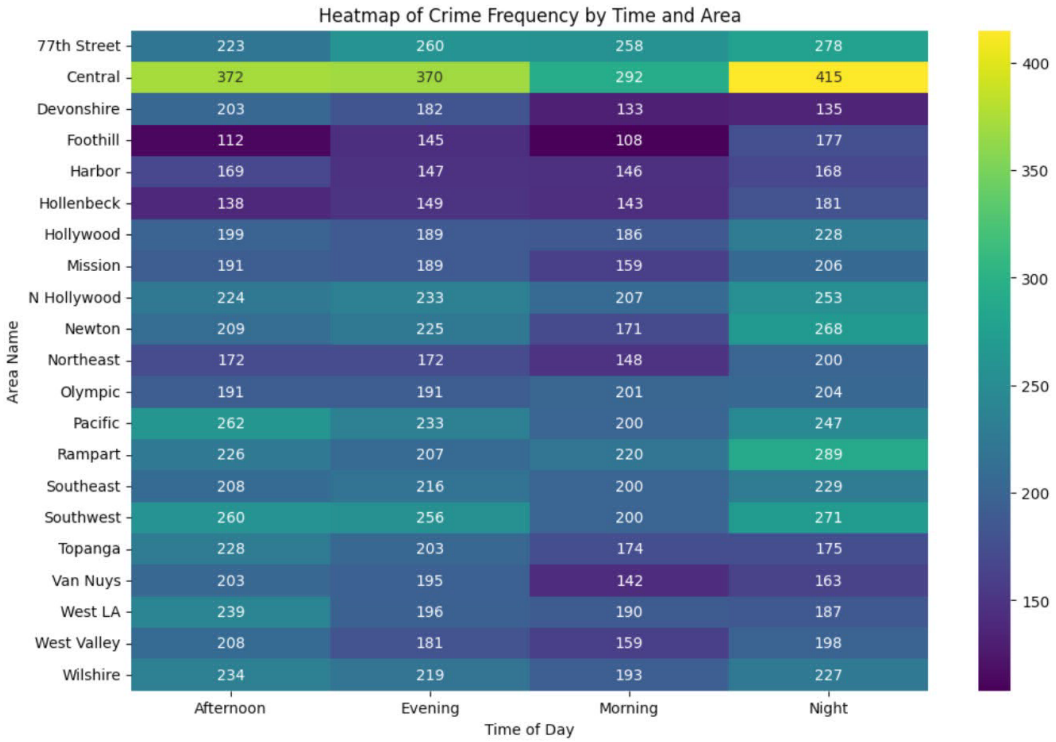

Figure 3: This heat map shows the Crime frequency by time and Area

4.1.4 Top 10 High Severity Crimes at Night

The pie chart showcased the top 10 high-severity crimes during the night. "Vehicle - Stolen" and "Assault With Deadly Weapon" dominated the night crime scene, emphasizing the prevalence of these crimes during this timeframe.

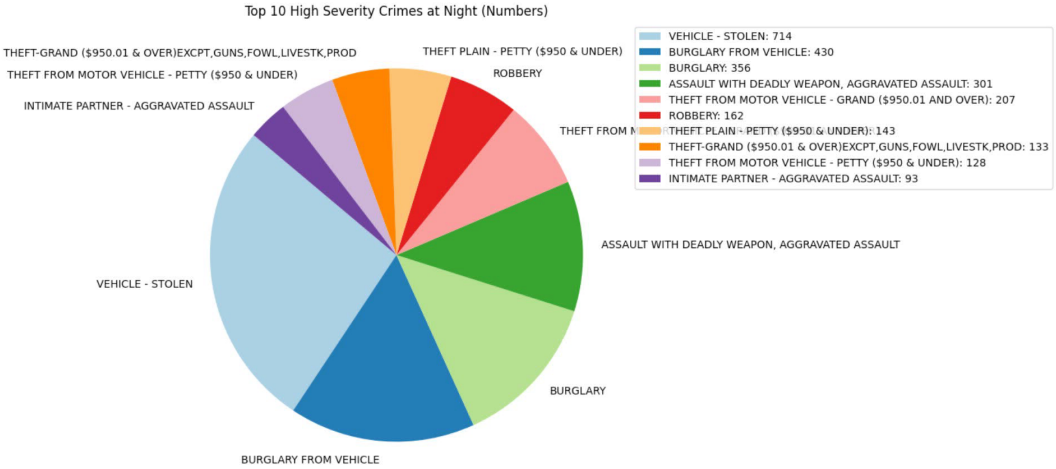

Figure 4: This pie chart shows top 10 high Serverity Crimes at night (Numbers)

4.1.5 Top 10 Crime Types Distribution

The top ten high-severity crimes that occurred at night were displayed in a pie chart. The words "vehicle - stolen" and "assault with deadly weapon" dominated the crime scene at night, highlighting how common these crimes were at this time.

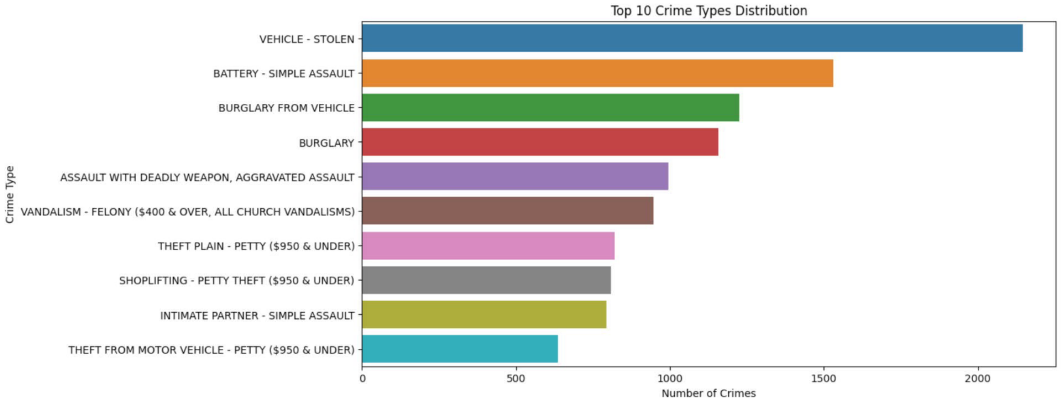

Figure 5: This chart shows top 10 crime types distribution

4.1.6 Pair Plots of the Whole Data

Pair plots helped us visualize the relationships between different numerical variables in our analysis. Although no clear patterns showed up, these plots added to our understanding of the interactions between the variables.

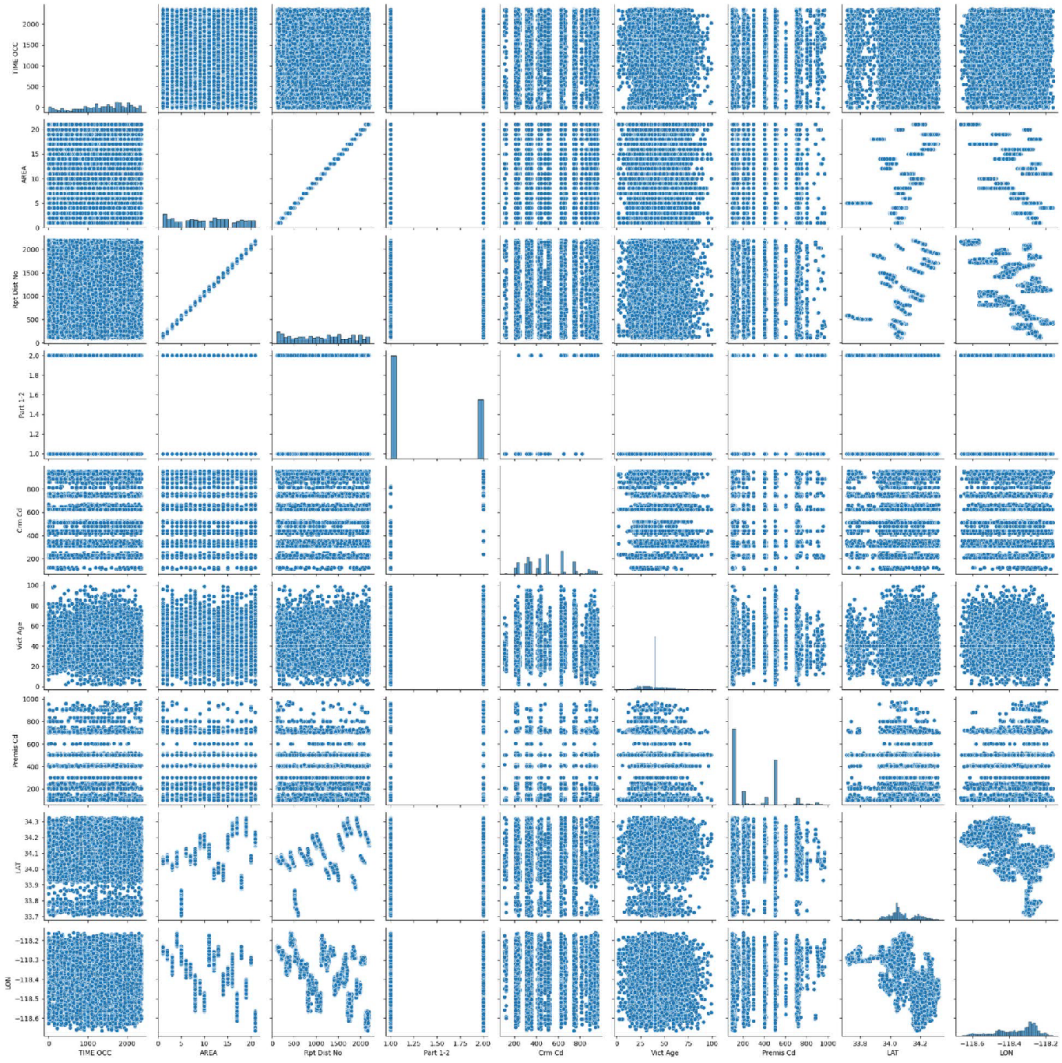

Figure 6: This pair plot shows relationships between numerical variables in our analysis

4.1.7 Crime Severity Analysis on the Los Angeles Map

The map's color-coding of incidents showed how serious crimes were. Red markers denoted high-severity incidents, and green markers represented less serious crimes. This geographical depiction provided a visual story of the severity of crime throughout Los Angeles.

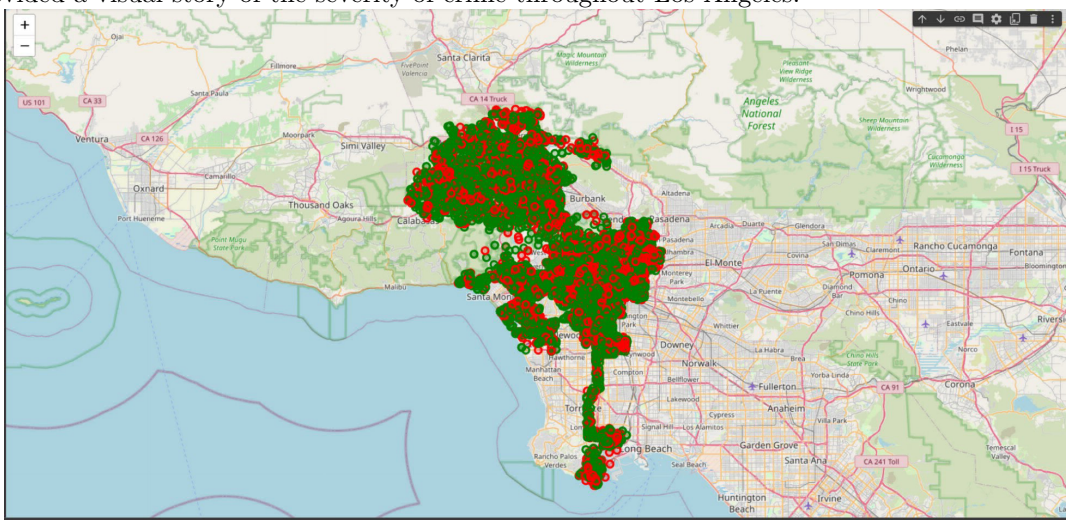

Figure 7: This Goe map shows the crime severity analysis on the Los Angeles Map

4.2 Correlation Analysis

The correlation matrix and heatmap helped us understand the relationships between the variables in our analysis. A moderately positive correlation between Part 1-2 crimes and specific crime codes was one of the main findings, highlighting the relationships between severity indicators and particular crime types.

	TIME OCC	AREA	Rpt Dist No	Part 1-2	Crm Cd	Vict Age	Premis Cd	Lat	LON
TIME OCC	1.0000	-0.0080	-0.0082	-0.0379	0.0090	-0.0076	-0.0247	0.0058	-0.0096
AREA	-0.0080	1.0000	0.9990	-0.0155	-0.0245	0.0353	-0.0274	0.3467	-0.4814
Rpt Dist No	-0.0082	0.9990	1.0000	-0.0155	-0.0238	0.0353	-0.0269	0.3441	-0.4827
Part 1-2	-0.0379	-0.0155	-0.0156	1.0000	0.7577	-0.0308	0.2732	-0.0092	0.3855
Crm Cd	0.0090	-0.0245	-0.0238	0.7577	1.0000	-0.0135	0.1751	-0.0096	0.0348
Vict Age	-0.0076	0.0353	0.0354	-0.0308	-0.0137	1.0000	0.0212	0.0314	-0.0645
Premis cd	-0.0247	-0.0274	-0.0269	0.2732	0.1751	0.0212	1.0000	0.0489	-0.0389
LAT	0.0058	0.3467	0.3441	-0.0092	-0.0096	0.0314	0.0489	1.0000	-0.5803
LON	-0.0096	-0.4814	-0.4827	0.0385	0.0348	-0.0645	-0.0389	-0.5803	1.0000

Figure 8: Tabular representation of the matrix

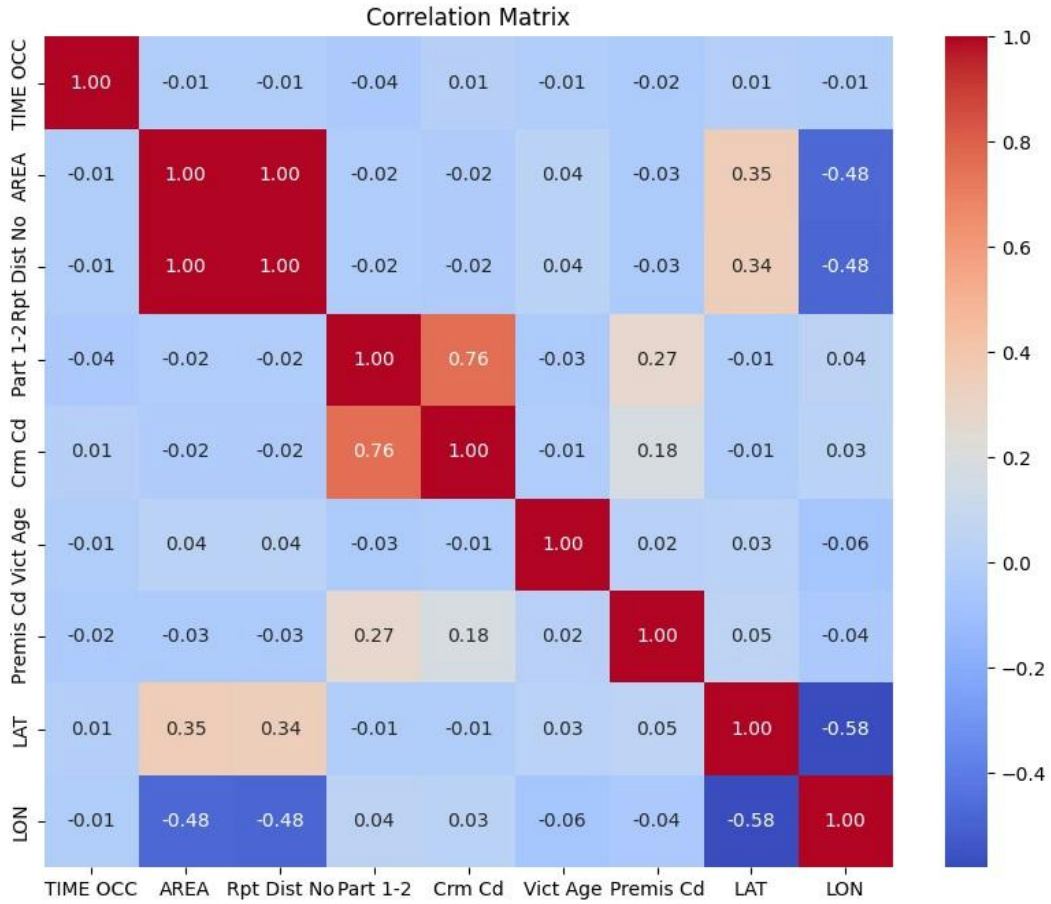

Figure 9: Heat map of the correlation matrix

4.3 Hypothesis Testing

First hypothesis test results from our analysis showed that there was no statistically significant difference in the mean victim age between "Vehicle - Stolen" and "Battery - Simple Assault." The results of the second hypothesis test showed that the distribution of crimes throughout the day was not uniform. The results of the third hypothesis test indicated that there was no discernible difference between the two different areas' "Vehicle - Stolen" frequencies. We have performed several tests on the data: P-value test, T- test, Chi-squared Statistic test either to accept or reject the null hypothesis statements.

Hypothesis Test-1: The examination of mean victim age between 'VEHICLE - STOLEN' and 'BATTERY - SIMPLE ASSAULT' crime types reveals no significant difference. The statistical analysis, employing a two-sample t-test, yields a p-value of 0.15, surpassing common significance levels. Therefore, we fail to reject the null hypothesis, indicating that the mean victim age for the specified crime types is nearly equal.

Hypothesis Test-2: Investigating the uniform distribution of crime occurrences throughout the day using a Chi-square goodness-of-fit test shows a significant departure from uniformity. The extremely low p-value of 0 suggests rejecting the null hypothesis, emphasizing a non-uniform distribution of crimes across different times. This implies distinct patterns in crime occurrences throughout various time periods of the day.

Hypothesis Test-3: The analysis of the occurrence frequency of 'VEHICLE - STOLEN' in two distinct areas, employing a Chi-square test for independence, results in a p-value of 1.0. This high p-value leads to accepting the null hypothesis, indicating that the frequency of this specific crime type occurs at a similar rate in both areas. The observed differences are not statistically significant, suggesting consistency in the occurrence of 'VEHICLE - STOLEN' across the designated areas.

5 CONCLUSION

This research explored Los Angeles crime dynamics, leveraging November 2023 data to uncover insights. Our analysis revealed key trends in high prevalence crimes, time and location patterns, and variable correlations. Key findings demonstrate nighttime being especially high-risk for severe criminal acts. While we did not find age differences between crime types, distinct distributions over the course of a day emerged. Spatial analysis also flagged areas experiencing varying degrees of crime intensity.

These revelations carry real-world impact for law enforcement agencies and community safety. The data-driven insights highlight the need for heightened vigilance and proactive deployment of resources during nighttime hours. They also emphasize focused prevention efforts in identified hotspots based on crime severity. Low income areas surfaced as disproportionately affected, signaling socio-economic disparities influencing crime rates.

As next steps, incorporating demographic and census data could shed more light on how economic factors shape criminal behavior. Additionally, future work should investigate dynamics within specific crime categories as patterns may differ. For instance, areas with higher gang activity may exhibit unique trends. Continual analysis should incorporate new data to promptly identify emerging developments. In summary, this study demonstrates the value of in-depth crime data mining to guide targeted policies for enhancing community welfare.

article hyperref

References

[1] Lofstrom, M., Martin, B. (2018). *Crime Trends in California - Public Policy Institute of California*. Public Policy Institute of California. https://www.ppic.org/publication/crime-trends-in-california/

[2] Crime Reporting, F. (2023). *Federal Bureau of Investigation Crime Data Explorer*. Cjis.gov. CDE UCJIS. https://cde.ucr.cjis.gov/LATEST/webapp/\#/pages/home

[3] Nguyen, V. N., Nguyen, A., Nguyen, D. (2023, April 18). *RPubs - Criminal Activities - Jan 2023 - Los Angeles*. RPubs. https://rpubs.com/andrewnguyen/crimela

[4] Harrendorf, S., Heiskanen, M., Malby, S. (2010). *INTERNATIONAL STATISTICS on CRIME AND JUSTICE*. UNODC. https://www.unodc.org/documents/data-and-analysis/Crime-statistics/International_Statistics_on_Crime_and_Justice.pdf

[5] *Crime Data from 2020 to Present*. (2022, November 16). Data.gov; data.lacity.org. Data.gov. https://catalog.data.gov/dataset/crime-data-from-2020-to-present

[6] DataLA. (2022, October 4). *A Data-Driven Exploration of Crime Trends in Los Angeles*. DataLA. https://medium.com/datala/a-data-driven-exploration-of-crime-trends-in-los-angeles-6124c2980eda

[7] Phuong, C. (2023, November 30). *Crime in Los Angeles: An Analysis of Crime Patterns and Victim Demographics*. Medium. Medium Article. https://medium.com/@chloephuong09/crime-in-los-angeles-an-analysis-of-crime-patterns-and-victim-demographics-780ddd90363c