

FROM DATA TO INSIGHTS: UNDERSTANDING CRIME PATTERNS ACROSS THREE AMERICAN URBAN LANDSCAPES

Vineet Jaywant (017408638)
Shreyas Kulkarni (017102878)
Rajath Nattoji Rajaram (017426097)
Maheedhar Reddy Bukkasamudram (017046939)

CMPE 255 Group 9

Under the guidance of, Dr. Taehee Jeong

TABLE OF CONTENTS:

1. ABSTRACT.....	3
2. INTRODUCTION.....	3
3. RELATED WORKS.....	4
4. DATA ANALYSIS.....	5
4.1. Trends, Patterns, and Anomalies in Urban Crime.....	6
A. Distribution of Crimes by Hour of the Day.....	6
B. Total Number of Crimes Over Time (Jan 2013-Dec 2023, Quarterly).....	7
C. Number of Crimes by Day of the Week.....	7
D. Trends of Top 10 Crime Types Over Time.....	8
E. Mapping Crimes by Area.....	8
F. Frequency of Top 10 Crime Types by Area.....	10
G. Top 10 Most Common Crime Types.....	10
5. COMMUNITY CONTRIBUTION.....	11
6. CONCLUSION.....	12

1. ABSTRACT

Crime rates continue to climb across countries, jeopardizing their prosperity, stability, and growth. Law enforcement agencies are looking for intelligent technologies that can efficiently evaluate crime data and target illegal activity. In this study, we utilize comprehensive crime datasets from three major American cities – San Francisco, Los Angeles, and Chicago – to obtain insights about patterns and trends in criminal activity. Our analysis focuses on the crime data, which includes detailed information on various aspects of reported crimes, such as the date and time of occurrence, location, crime type, victim demographics, and weapon usage.

We investigate several dimensions of crime, including spatial distribution, temporal trends, and criminal typologies, using thorough data preparation and visualization approaches. In particular, we examine how crimes are distributed geographically. We also investigate the temporal patterns of criminal activity, looking at differences in crime rates according to the day of the week, the time of day, and seasonal patterns. We also explore the many categories of crime, including property crimes and violent crimes, as well as the relative frequency of each in the city's various regions.

Through exploratory data analysis and predictive modeling, our goal is to provide evidence-based tactics for law enforcement agencies and crime prevention efforts by uncovering underlying patterns and trends in urban crime dynamics. By combining the findings from these analyses, we aim to further our understanding of the complexities surrounding criminal activity and contribute to the development of more effective strategies for maintaining public safety and security.

2. INTRODUCTION

The issue of crime is a recurring concern in urban settings, affecting the security and welfare of inhabitants and societies. Law enforcement agencies in major cities deal with a wide range of criminal activity, from small-time thefts to severe offenses. Data-driven methods present a viable way to address this complex problem by comprehending the dynamics of crime and developing successful intervention plans. In this work, we conduct a thorough examination of crime datasets obtained from San Francisco, Los Angeles, and Chicago. Using meticulous preprocessing and data visualization techniques, our goal is to discover meaningful insights regarding crime patterns and trends.

Preprocessing raw crime datasets with the goal of guaranteeing data reliability and integrity is essential to our study. The datasets are modified by a variety of preprocessing methods such as duplicate removal, outlier removal, and imputing missing values to enable insightful analysis. We look for intrinsic patterns and connections in the data by using visual aids. In particular, our methodology includes determining the most frequent crime types, tracking trends in crime occurrence over time, and looking at the frequency of the top crime types by area. Furthermore, we look into how crimes are distributed throughout the week and during the day to find trends in the timing of criminal activity.

We hope to further our understanding of the dynamics of urban crime and offer useful information to urban planners, law enforcement, and politicians by combining the findings from our investigation. Our objective is to provide light on the fundamental causes of criminal behavior and to assist law

enforcement authorities in making evidence-based decisions by methodically analyzing the temporal and spatial aspects of crime. In the end, we want our study to contribute to evidence-based plans for community safety, resource allocation, and crime prevention in urban settings

3. RELATED WORKS

“Crime Data Analysis in San Francisco, Seattle, and India using OpenStreetMap”

This study uses OpenStreetMap (OSM) to examine how significant violations, age trends, and seasonal fluctuations affect crime rates in India, Seattle, and San Francisco. It does data preparation, temporal trend analysis, correlation evaluation, and visualization using R software. Unfortunately, the study does not go into great detail when describing analytic methodologies, offers scant findings from visualizations, does not conduct comparative research between locations, and just briefly touches on issues like data biases and analysis limits.

Tamilaruvi, D., & Manjula, K. R. (2021). Crime Data Analysis in San Francisco, Seattle, and India using OpenStreetMap. International Journal of Research in Advanced Science, Engineering and Technology (IJRASET), 9(6), 1. ISSN: 2321-9653.

“Survey on Crime Analysis and Prediction Using Data Mining Techniques”

A detailed overview of data mining approaches for crime pattern analysis, communication analysis, spatial/geolocation algorithms, text/NLP techniques, and prisoner categorization systems is given in this work. However, deficiencies include a lack of holistic approaches, a focus on certain crime types, parameter optimization issues, and the need to integrate several data sources to improve accuracy.

H. Benjamin Fredrick David and A. Suruliandi, "Survey on Crime Analysis and Prediction Using Data Mining Techniques" ICTACT Journal on Soft Computing, vol. 7, no. 3, pp. 1459, April 2017. DOI: 10.21917/ijsc.2017.0202. ISSN: 2229-6959.

“Crime Analysis using Open Source Information”

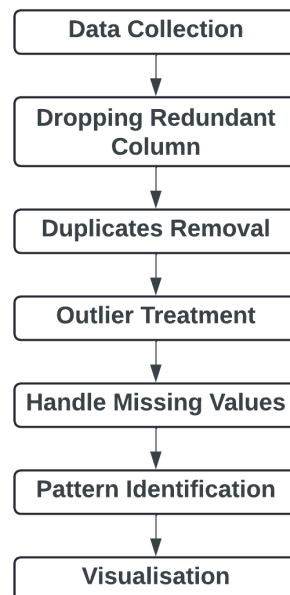
The project offers an open-source information-based method for analyzing crime that is based on FBI news releases and makes use of unsupervised data mining methods like clustering and association. Retrieving data, pre-processing, statistical analysis, visualization, clustering, association analysis, and manual evaluation are all part of the methodology. Its shortcomings include a limited scope, the absence of real-time analysis, a lack of methodology for social media analysis, and the necessity for external validation against existing tools or processes.

Nizamani, Sarwat et al. “Crime Analysis using Open Source Information.” ArXiv abs/1902.05684 (2015).

4. DATA ANALYSIS

In the course of our data analysis, we observed several significant characteristics and patterns that are crucial for understanding the dataset and subsequently making informed decisions for any predictive modeling or statistical analysis.

There are 7 steps involved in Exploratory Data Analysis are shown in the below flow diagram.



Data Collection: The datasets consist of incidents of crime reported in the cities of Chicago, San Francisco, and Los Angeles. There are three datasets used for this project and were acquired from the official websites of the police departments in San Francisco, Chicago, and Los Angeles. The crime datasets have been collected by the respective Police Departments and are updated daily. The data includes enough information about the Date, Type, Description, location, etc of the crime for our analysis. The combined data contains nearly 7 million records of the crime.

Dropping Redundant Column: The redundant columns not only introduce noise but also amplify computational complexity. By discarding such redundant features, analysts can streamline the analysis process, thereby enhancing the performance and interpretability of their models. In our dataset, we have dropped around 10 insignificant columns.

Duplicates Removal: Data is meticulously examined to identify and eliminate redundant entries. The presence of duplicates can significantly skew the results of data mining algorithms. More than 50000 duplicate entries were deleted from the combined data.

Outlier Treatment: Data outliers, which are data points that deviate significantly from the expected range for a specific variable, are identified using Box-Plots and removed. In the context of crime analysis, an outlier might be due to data entry errors. In our analysis, the outliers were deleted from the combined data.

Handling Missing Values: Missing Values are handled by either dropping the entire row or by imputing the values with a median by taking its subclasses or features into consideration. In our analysis, we have imputed values for more than 5 columns.

Pattern Identification: Following data cleaning, data mining techniques are employed to uncover patterns and relationships within the dataset. In our crime analysis, we have identified trends in specific crime types or pinpoint areas with high crime occurrences.

Visualization: Data visualization tools are leveraged to create informative charts and graphs that effectively communicate the discovered patterns and trends within the data.

4.1. Trends, Patterns, and Anomalies in Urban Crime

A. Distribution of Crimes by Hour of the Day

An analysis across Chicago, San Francisco, and Los Angeles reveals a consistent pattern in crime occurrence by hour, with peaks in the late evening and early morning hours, aligning with routine activity theory. This pattern underscores the relationship between human activity schedules and crime opportunities. The anomalous spike observed at 12 PM could indicate data inconsistencies, such as a clustering of unreported crimes getting logged at a common time due to administrative processes. Such anomalies warrant a review of data collection methods.

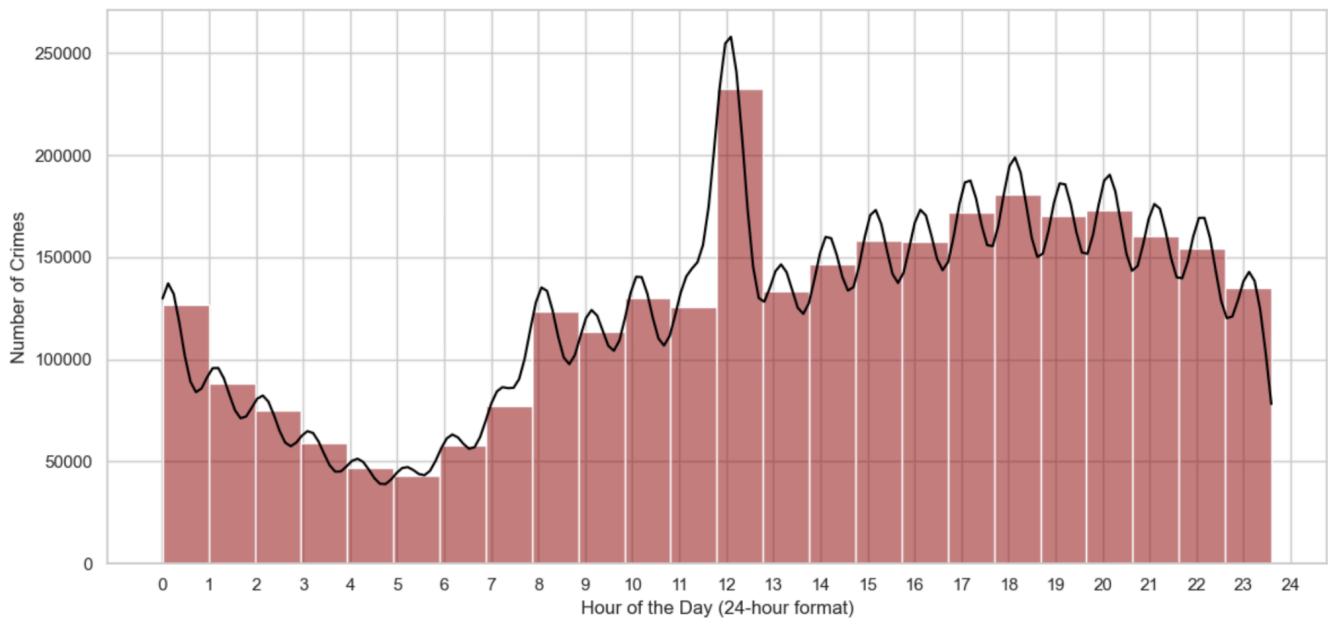


Fig. 1 Hourly Analysis of Number of crimes

B. Total Number of Crimes Over Time (Jan 2013-Dec 2023, Quarterly)

The quarterly crime data from 2013 to 2024 for the three cities exhibit cyclical trends with seasonal peaks, consistent with social behavior and environmental factors affecting crime rates. An unusual drop in crime rates in 2020 across all cities highlights the impact of external factors, such as the COVID-19 pandemic, demonstrating how societal shocks can disrupt typical crime trends.

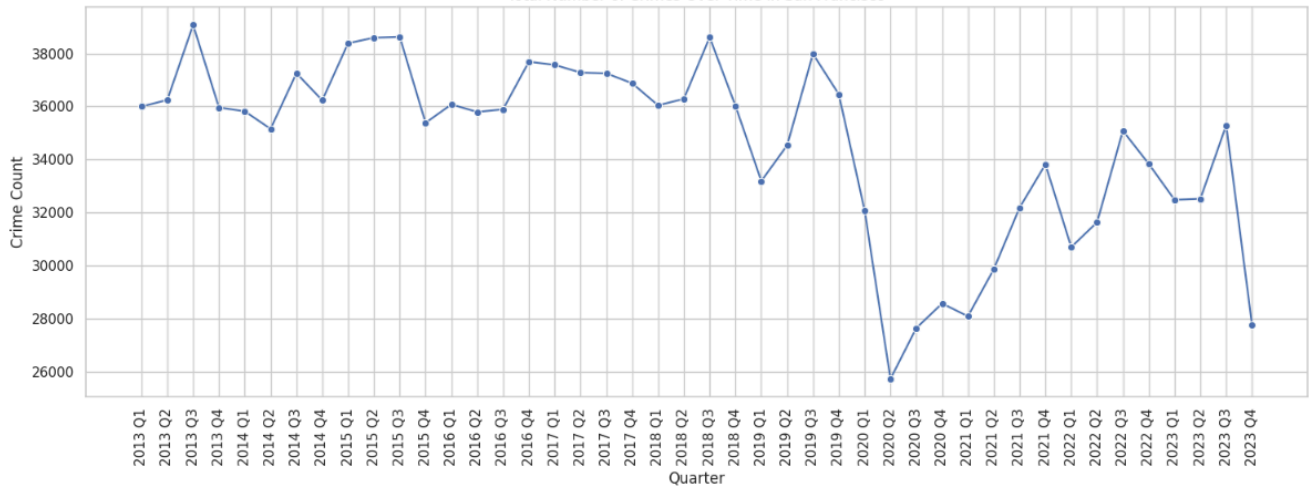


Fig. 2 Variation in the number of crimes through the years (Quarterly)

C. Number of Crimes by Day of the Week

Contrary to traditional beliefs that weekends or specific days witness higher crime rates, the analysis does not show significant variations in crime rates by day of the week across the three cities. This finding suggests that urban crime may be more influenced by other factors than weekly social cycles.

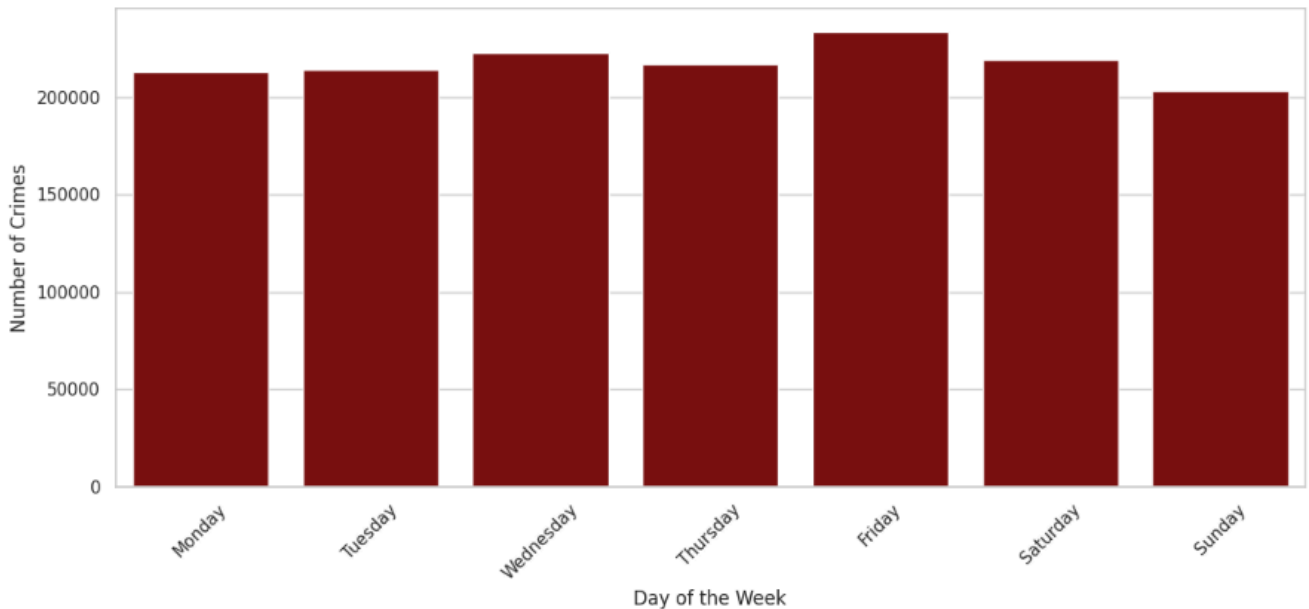


Fig. 3 Day-wise representation of the number of crimes

D. Trends of Top 10 Crime Types Over Time

A decrease in narcotics offenses and the consistency of theft as the most prevalent crime type are observed across all cities. These trends suggest shifts in law enforcement focus and societal changes. The analysis reveals both expected and unexpected patterns in the prevalence of crime types over time.

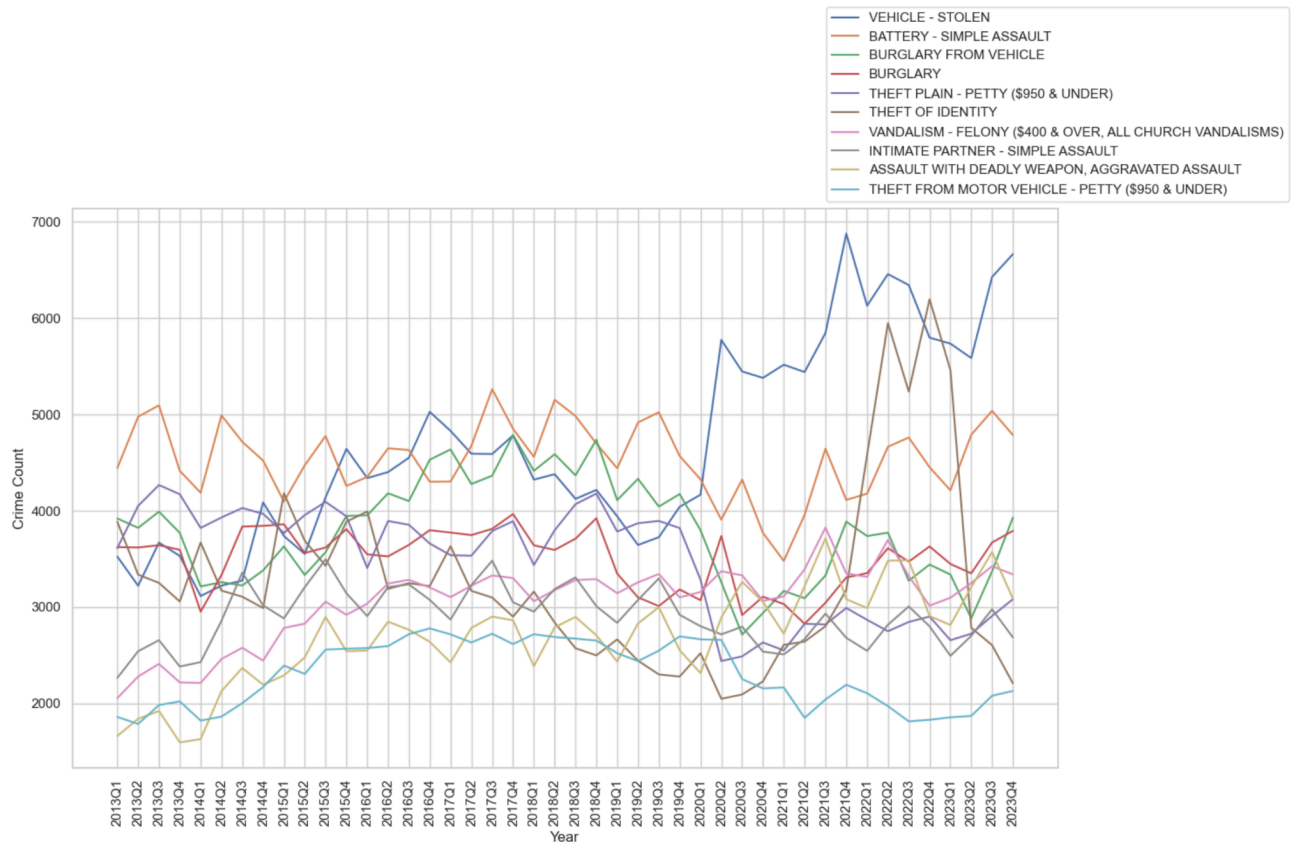


Fig. 4 Variation in different types of crimes

E. Mapping Crimes by Area

Geographical analysis of crime distribution highlights that high-crime areas often correlate with commercial activity and dense population centers. This supports the crime place theory, which posits that crimes are more likely to occur in areas providing more opportunities. Unexpectedly, some neighborhoods defy socio-economic expectations, indicating the complex interplay of factors, including community dynamics and policing strategies, in influencing crime rates.

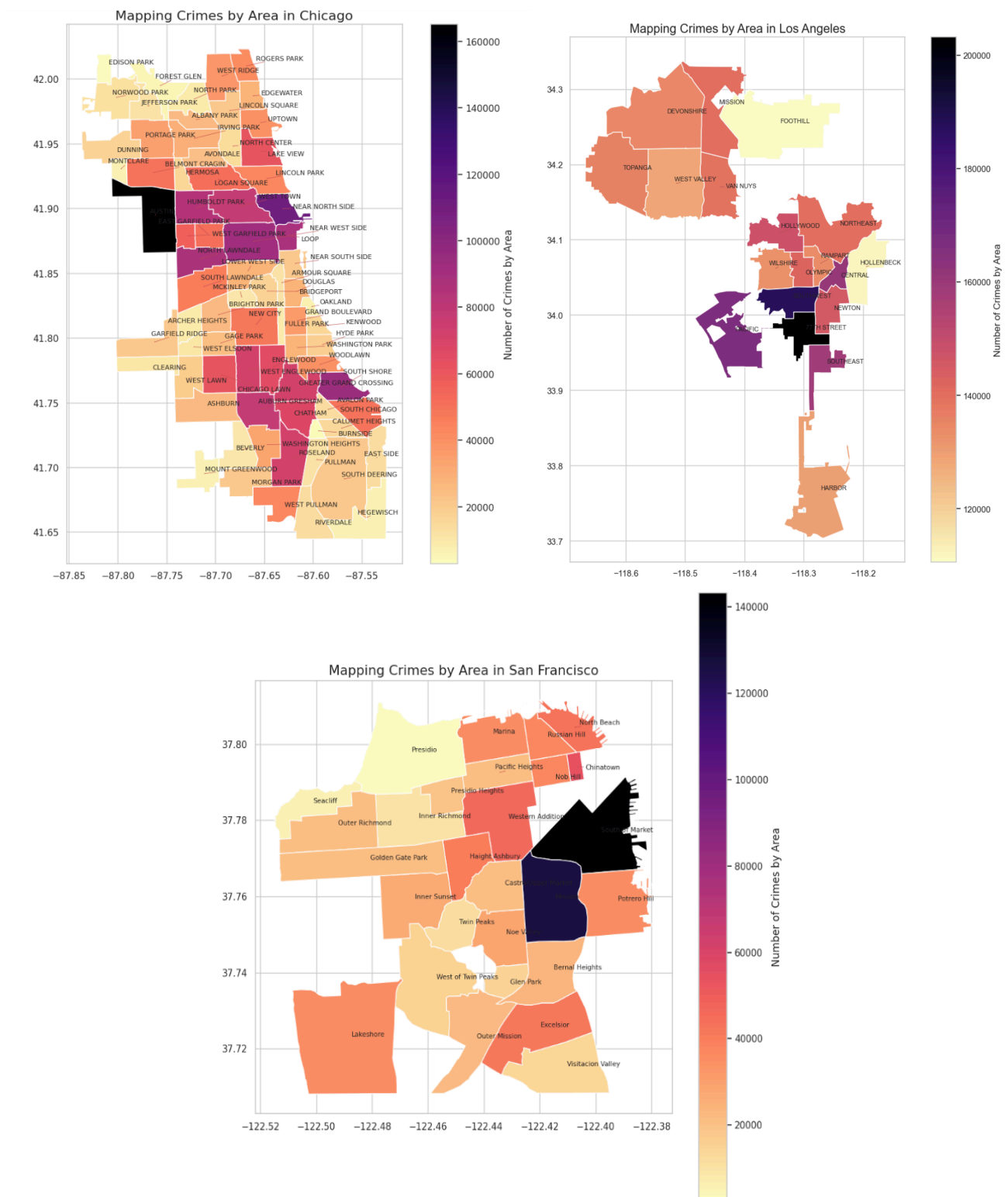


Fig. 5a Density graphs for the number of crimes in Chicago
 Fig. 5b Density graphs for the number of crimes in Los Angeles
 Fig. 5c Density graphs for the number of crimes in San Francisco

F. Frequency of Top 10 Crime Types by Area

The prevalence of specific crime types varies significantly across different urban areas, reflecting the influence of local socioeconomic conditions, law enforcement strategies, and community engagement. This variation underscores the necessity for targeted interventions and resources in high-crime areas.

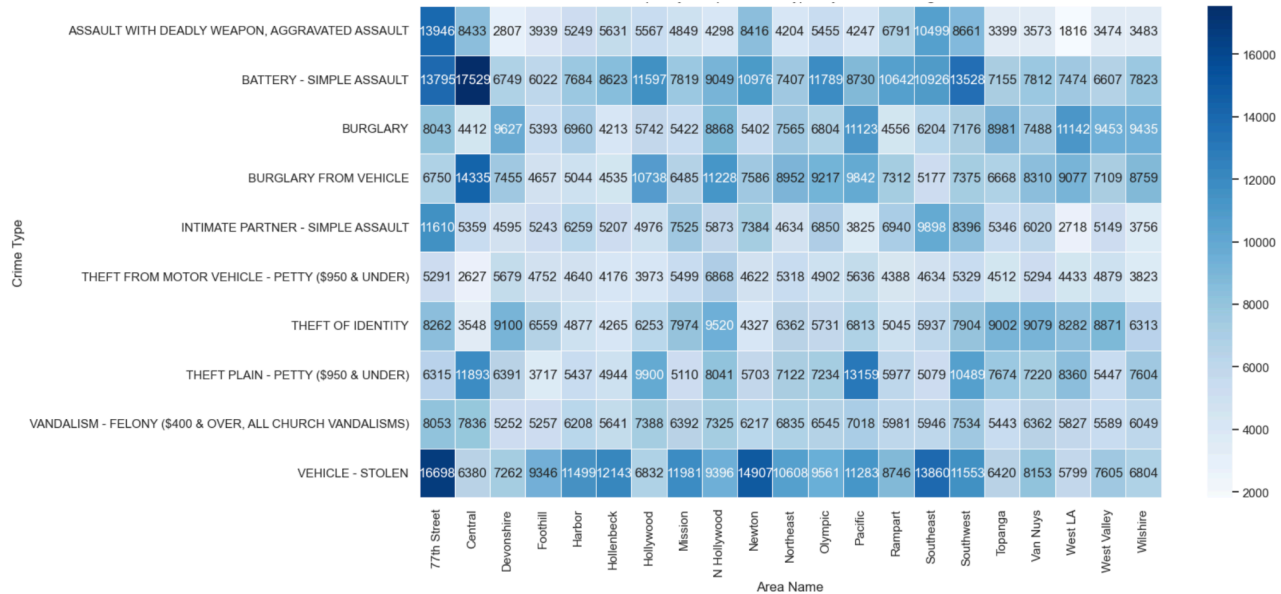


Fig. 6 Distribution of crime types across areas in Los Angeles

G. Top 10 Most Common Crime Types

Theft, assault, and narcotics offenses emerge as the most common crimes across Chicago, San Francisco, and Los Angeles. Despite the diverse geographic and socio-economic contexts of these cities, the similarity in top crime types points to underlying patterns of criminal behavior and opportunity structures in urban environments.

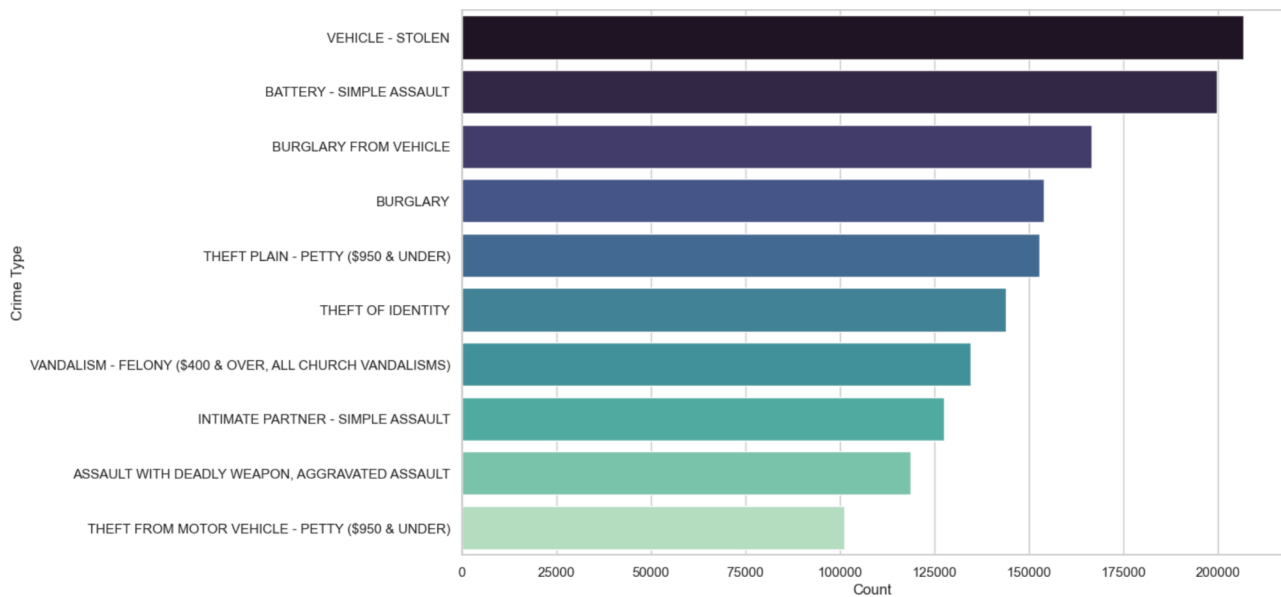


Fig. 7 Top 10 crime types and the number of crimes

5. COMMUNITY CONTRIBUTION

An analysis of crime statistics in major cities such as San Francisco, Los Angeles, and Chicago can yield a plethora of information that can substantially impact the safety and well-being of local government, residents, and tourists. By analyzing crime types, locations, timings, and distribution over a week and comparing trends spanning ten years, we arm ourselves with the knowledge necessary to make well-informed decisions and carry out focused interventions.

This analysis provides a guide for safer travel experiences for tourists. Travelers can plan their trips to reduce risks by learning which areas are more likely to experience specific types of crime and what times these crimes are most likely to occur. They can now confidently and peacefully explore the cities, which not only increases their safety but also their overall experience. In addition, knowing the patterns of crime over the previous ten years gives travelers a better idea of the safety situation in a city, enabling them to plan their travels with knowledge.

This research will be very beneficial to the residents. They are better equipped to take preventative action in their daily lives when they are aware of how crimes are distributed by time and place. Because neighbors can set up neighborhood watches or work with local law enforcement to address specific issues identified in the analysis, it promotes a sense of community. Furthermore, being aware of long-term trends in crime enables locals to support policies or initiatives that may address recurring or new crime problems, as well as to recognize the success of previous interventions.

For local governments, the detailed crime analysis serves as a critical tool for planning and resource allocation. Law enforcement organizations can more effectively deploy staff and technology where

they are most needed by identifying hotspots for different kinds of crime. This tactical approach guarantees the prudent use of public funds while optimizing the impact of crime prevention measures. Government officials can also assess the success of previous policies and initiatives by looking at crime trends over the last ten years. This information helps them make decisions in the future that will further lower crime rates.

In addition to the obvious advantages, this thorough analysis of crime promotes a more general climate of openness and responsibility. Governments show their dedication to public safety and their readiness to collaborate with the community to address issues by providing access to such information. Consequently, this fosters public trust in law enforcement, which is critical for successful crime prevention and community policing.

6. CONCLUSION

In conclusion, our project embarked on a comprehensive analysis of crime statistics spanning from January 2013 to December 2023 across three major cities: San Francisco, Los Angeles, and Chicago. With meticulous attention to detail, our team diligently processed each dataset, addressing missing values, duplicate entries, and erroneous data points. Through individual preprocessing tailored to the unique characteristics of each dataset, we ensured the integrity and accuracy of our analyses.

A notable accomplishment of our preprocessing efforts was the development of a method to estimate the approximate geographical location based on latitude and longitude coordinates. This innovation allowed us to enhance the granularity of our analyses and provided valuable insights into spatial patterns of crime.

The heart of our analysis lay in the visualization of data through meticulously crafted graphs, enabling us to discern temporal trends and identify anomalies. Notably, our quarterly analysis of crime rates from 2013 to 2023 illuminated subtle shifts in patterns over time, empowering us to delve deeper into the underlying factors driving these changes.

Throughout our project, community engagement remained a core principle. By sharing our findings with stakeholders and fostering dialogue around crime prevention and intervention strategies, we endeavored to contribute meaningfully to the well-being of our communities.

In summation, our project stands as a testament to the power of data-driven insights in addressing complex societal challenges. Through rigorous analysis and a commitment to transparency and community involvement, we strive to pave the way for safer and more resilient cities.