

Exploring Crime Analysis with LAPD Leveraging Machine Learning for Public Safety

Project report submitted for
Industrial Project Based Learning in Data Science and Machine Learning

Under the Supervision of

P.Mohan
Director, Gyan Astra IT Solutions

&
Y.V.N Phani Kishore
Director, Gyan Astra IT Solutions

By

Team Members Name (Roll Number)

P Pranavi	21R11A6742
C Manognasri	21R11A6729
A. Sai Srujana	21R11A05A8
P Moulika	21R11A05E0

Department of Computer Science & Engineering

Nov 2023 - May 2024



Geethanjali College of Engineering & Technology

Accredited by NBA (UGC Autonomous)

(Affiliated to J.N.T.U.H, Approved by AICTE, New Delhi, NAAC - A+)

Geethanjali College of Engineering & Technology

(Affiliated to J.N.T.U.H, Approved by AICTE, NEWDELHI.)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that Mr./Ms. C.Manognasri, A.Sai Srujana, P Moulika , P Pranavi bearing roll no 21R11A6729,21R11A05A8,21R11A05E0, 21R11A6742 has successfully completed Industrial Project Based Learning in Data Science and Machine Learning held during November 2023 to May 2024 at Geethanjali college of Engineering and Technology.

Coordinator

Mr. E. Mahender

Assistant Professor

Dept of C.S.E

H.O.D - C.S.E

Dr. A. Sree Lakshmi

Professor & Head

Dept of C.S.E

ACKNOWLEDGEMENT

We are greatly indebted to the authorities of Geethanjali College of Engineering and Technology, Cheeryal, R.R Dist, for proving us the necessary facilities to successfully carry out this Industrial Project Based Learning titled “Exploring Crime Analysis with LAPD Leveraging Machine Learning for Public Safety” under Value added course.

Firstly, we would like to express our sincere gratitude to our **Principal Dr. S.Udaya Kumar** for providing the necessary infrastructure to complete our Value added course.

Secondly, we thank and express our solicit gratitude to **Dr. A. Sreelakshmi**, Prof & HOD, CSE department, Geethanjali College of Engineering, and technology, for her invaluable help and support which helped us a lot in successfully completing of our course.

We express our gratitude to **P.Mohan**, Director, Gyan Astra IT Solutions, **Y.V.N Phani Kishore** Director, Gyan Astra IT Solutions, **Dr Kamakshiah Kolli**, Associate Professor, Dept of CSE, GCET, **Mr. E.Mahendra**, Dept of CSE, GCET and **N.Madhavi**, **Assistant Professor**, Dept of CSE-Data Science, GCET for their valuable suggestions and encouragement which helped us in the successful completion of our course.

Finally, we would like to express our heartfelt thanks to our parents who were very supportive both financially and mentally and for their encouragement to achieve our set goals.

Team Names and roll Numbers.

P Pranavi	21R11A6742
C Manognasri	21R11A6729
A. Sai Srujana	21R11A05A8
P Moulika	21R11A05E0

INDEX TABLE

S.NO	LIST OF CONTENTS	PAGE NO
1.	ABSTRACT	5
2.	INTRODUCTION	6
3.	PROBLEM STATEMENT	7
4.	OBJECTIVES	8
5.	METHODOLOGY	9
6.	EXPLORATORY DATA ANALYSIS :	10
7.	CONCLUSION	20

1.ABSTRACT

This study explores crime analysis using a comprehensive dataset from the Los Angeles Police Department (LAPD) spanning 2012 to 2015, supplemented with data from 2020 to 2024. By applying machine learning techniques like predictive modeling and pattern recognition, the goal is to uncover hidden correlations and insights that can enhance public safety measures. The LAPD, as the main law enforcement agency for Los Angeles, deals with a wide range of criminal activities across different areas of the city. Advanced data analytics are used to help identify emerging trends, crime hotspots, and potential risk factors for criminal incidents. The methodology includes dataset preprocessing, feature engineering, model training, and evaluation to develop robust predictive models. The study aims to provide valuable insights to optimize resource allocation, inform law enforcement strategies, and reduce crime rates in Los Angeles. By integrating machine learning into crime analysis, this research supports the LAPD's mission of ensuring public safety and improving residents' quality of life in Los Angeles.

2.INTRODUCTION

This project aims to delve into crime analysis with the LAPD, utilizing machine learning to extract actionable insights from comprehensive crime datasets spanning from 2012 to 2015, supplemented with additional data from 2020 to 2024. By tapping into the rich information contained in these datasets, including detailed records of reported crimes, victim demographics, and incident characteristics, the goal is to uncover hidden patterns, correlations, and trends in criminal behavior.

The integration of machine learning methodologies offers a new approach to crime analysis, enabling the identification of predictive indicators and risk factors associated with various types of criminal activity. Through predictive modeling and pattern recognition algorithms, this study aims to assist the LAPD in proactively identifying emerging crime trends, allocating resources efficiently, and developing targeted interventions to mitigate crime rates across the city.

By leveraging data-driven insights derived from machine learning algorithms, this project seeks to contribute to the LAPD's overarching mission of ensuring public safety and enhancing the quality of life for all residents of Los Angeles. Through the development of predictive models and actionable recommendations, this research aims to empower law enforcement agencies with the tools and knowledge necessary to address the complex challenges posed by urban crime dynamics.

3.PROBLEM STATEMENT

Our study delves into the extensive dataset provided by the Los Angeles Police Department (LAPD), shedding light on diverse crime occurrences and victim demographics. The LAPD, formally known as the City of Los Angeles Police Department, serves as the law enforcement agency for Los Angeles, California, in the United States. It operates through four major bureaus – Central, Valley, South, and West Bureau – encompassing a total of approximately 21 divisions. Our analysis focuses on the LAPD crime dataset spanning from 2012 to 2015, which covers various crime categories occurring at different locations within the city.

The project teams are tasked with supporting the LAPD by analyzing crime data to discern patterns in criminal behavior. Their objective is to utilize these insights to allocate resources effectively for addressing different types of crimes across various areas. The dataset comprises comprehensive records of reported crimes in Los Angeles from 2020 to 2024. Each entry contains vital information such as the Department Report Number, date and time of occurrence, location details, crime type, victim demographics, and more. This dataset presents a valuable opportunity to investigate patterns, trends, and correlations in crime data, ultimately informing crime prevention efforts and law enforcement strategies.

4.OBJECTIVES

- Reveal insights into crime patterns, victim demographics, spatial-temporal trends, and factors influencing crime severity.
- Provide actionable insights to support law enforcement agencies and policymakers in strengthening crime prevention strategies and improving public safety across Los Angeles.
- Examine crime trends over various timeframes: Identify seasonal fluctuations (e.g., potential increases in property crimes during summer) or recurring daily/weekly patterns.
- Explore spatial distributions: Investigate areas with heightened crime rates and identify potential contributing factors (e.g., demographics, socio-economic indicators).
- Analyze relationships among crime types: Investigate potential connections or co-occurrences between different categories of criminal activities.

5.METHODOLOGY

1. Data Collection and Preprocessing:

- Collect comprehensive crime datasets provided by the LAPD, spanning from 2012 to 2015, and supplemented with data from 2020 to 2024.
- Preprocess the raw data to handle missing values, outliers, and inconsistencies. Standardize data formats and ensure data quality for accurate analysis.

2. Exploratory Data Analysis (EDA):

- Perform exploratory data analysis to gain insights into crime patterns, temporal trends, spatial distributions, and relationships between different crime types.
- Visualize key findings using graphs, charts, and maps to facilitate understanding and interpretation of the data.

3. Machine Learning Model Selection:

- Identify appropriate machine learning algorithms for crime analysis tasks, including classification, clustering, regression, and anomaly detection.
- Evaluate different models based on their performance metrics, scalability, interpretability, and suitability for addressing specific objectives.

4. Model Training and Validation:

- Split the dataset into training and validation sets to train machine learning models on historical data and assess their performance on unseen data.
- Utilize techniques such as cross-validation, hyperparameter tuning, and model selection to optimize model performance and generalize well to new data.

5. Predictive Modeling:

- Develop predictive models to forecast future crime trends, identify high-risk areas, and prioritize resource allocation for crime prevention efforts.
- Incorporate advanced machine learning techniques, such as ensemble learning, time series analysis, and spatial modeling, to capture complex relationships within the data.

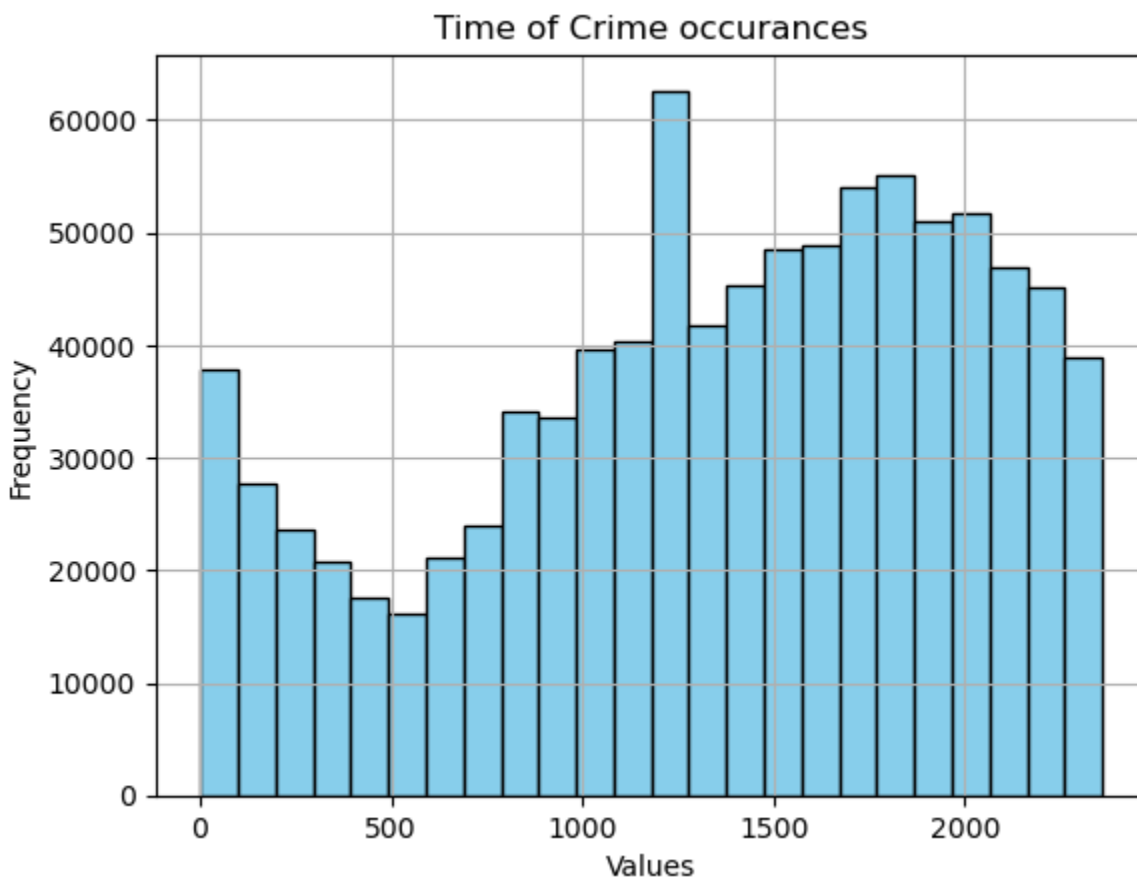
6. Integration and Deployment:

- Integrate machine learning models into existing crime analysis systems or develop standalone applications for real-time monitoring and decision support.
- Collaborate with stakeholders to deploy and evaluate the effectiveness of machine learning-based solutions in practice, iteratively refining the methodology based on feedback and outcomes.

7. Documentation and Knowledge Sharing:

- Document the methodology, codebase, and findings to facilitate reproducibility, transparency, and knowledge sharing within the research community and with relevant stakeholders.
- Publish research findings in academic journals, present at conferences, and engage in public outreach to disseminate insights and contribute to advancements in crime analysis and public safety.

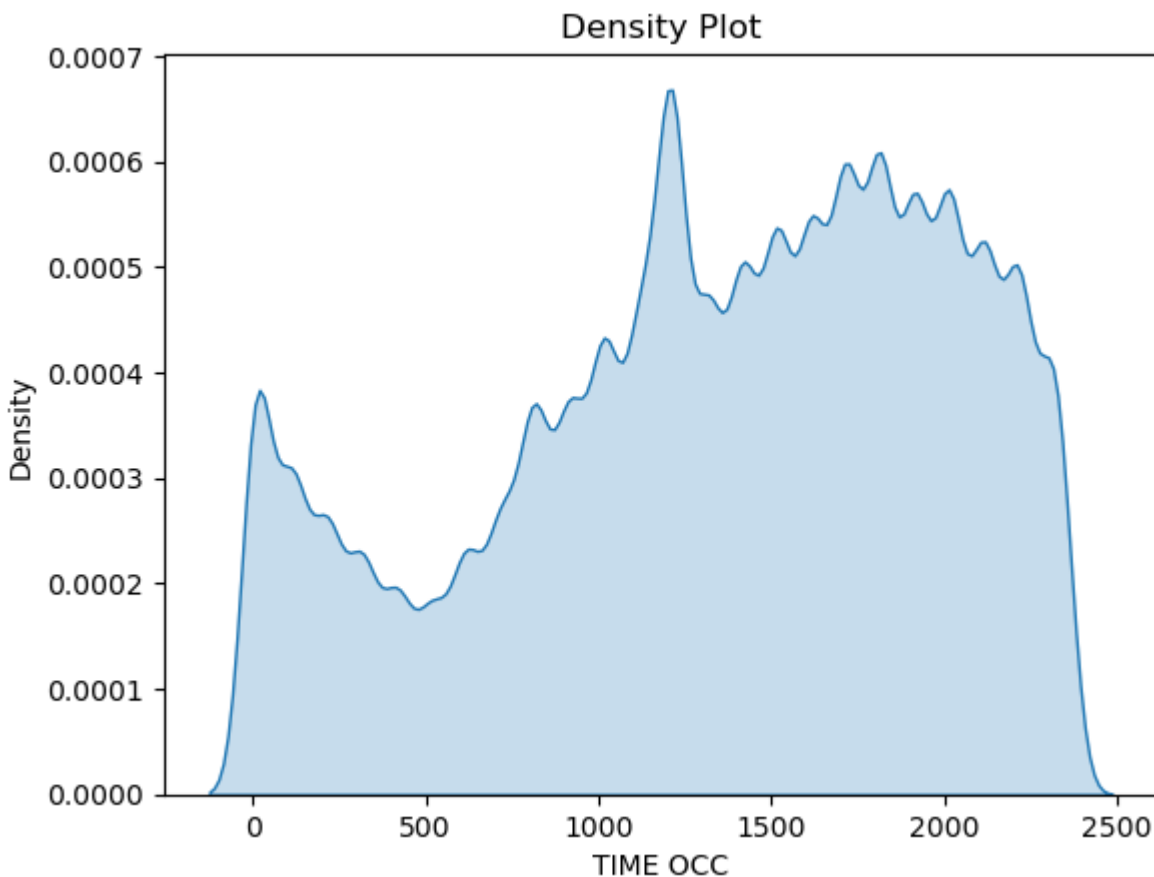
6.EXPLORATORY DATA ANALYSIS :



OBSERVATION:

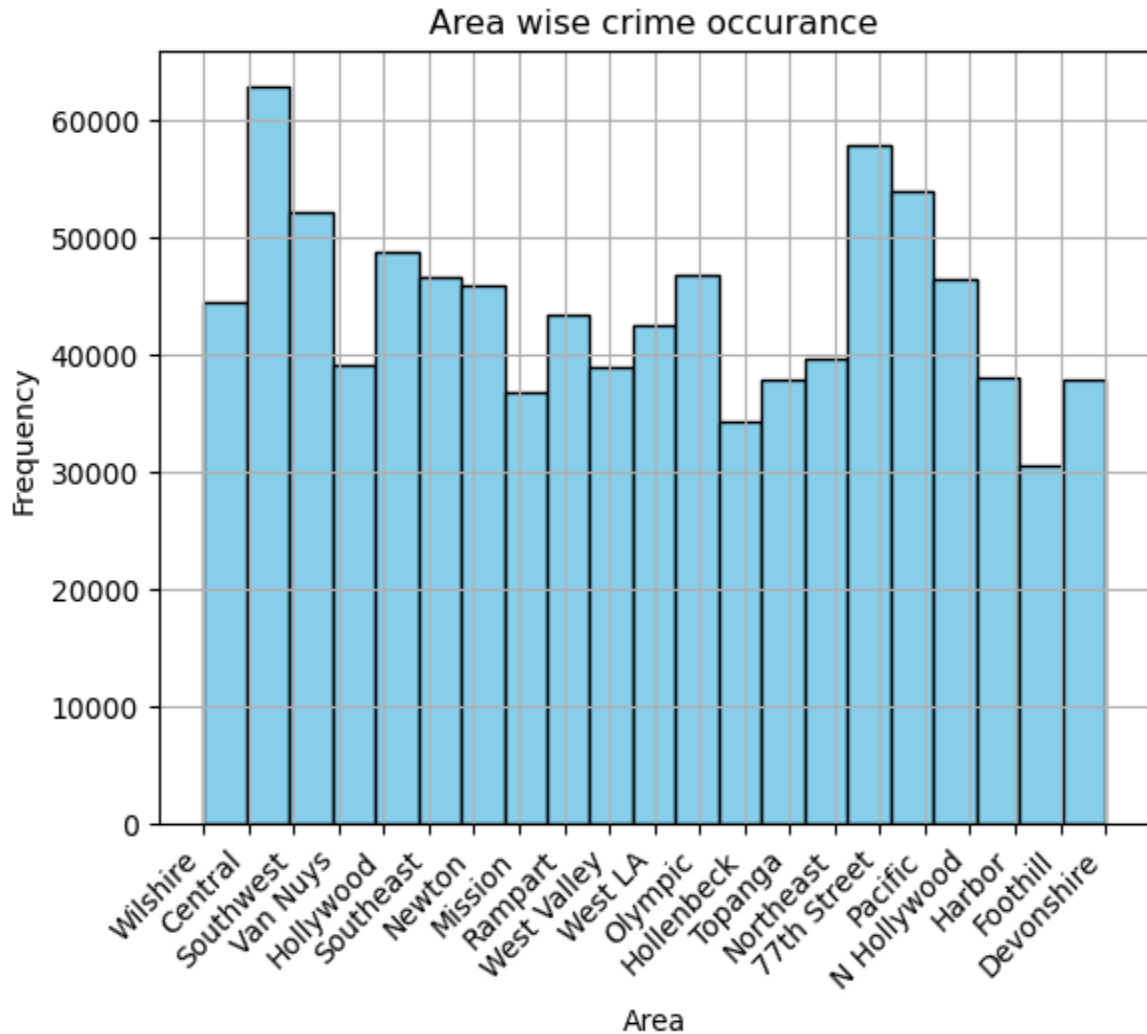
The bar graph depicts the distribution of crime incidents across a 24-hour cycle, with each hour of the day represented. Interestingly, the data highlights a notable spike in criminal occurrences

at 1300 hours, or 1:00 PM. This peak is particularly striking, with reported crimes exceeding 600,000 incidents. The significant concentration of criminal activity during this specific time period emphasizes the need for comprehensive analysis and heightened awareness from both law enforcement agencies and community stakeholders.



OBSERVATION:

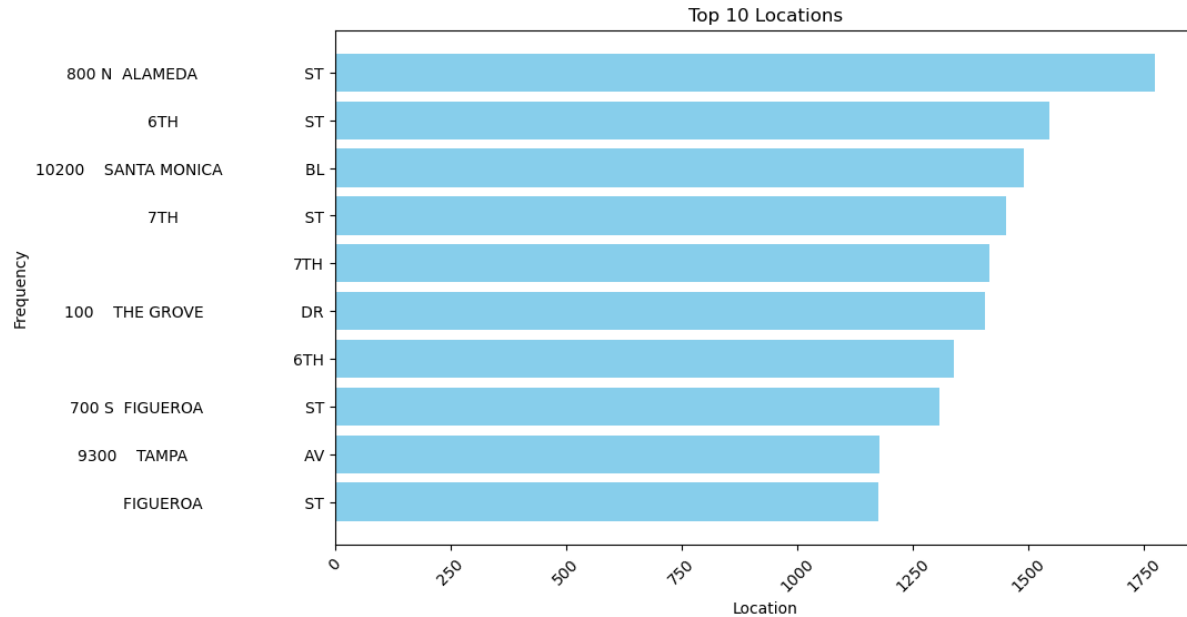
The graph provides a visual representation of the fluctuation in crime occurrences over the course of a day. Upon analysis, it's clear that the frequency of crimes varies significantly, with some periods experiencing notably higher rates than others. In particular, there is a distinct peak around noon, especially at 1:00 PM, indicating a heightened concentration of criminal activity during this time. These insights from the graph emphasize the significance of understanding temporal patterns in crime occurrence and may prompt further exploration into the underlying factors contributing to increased criminal incidents during specific hours of the



day.

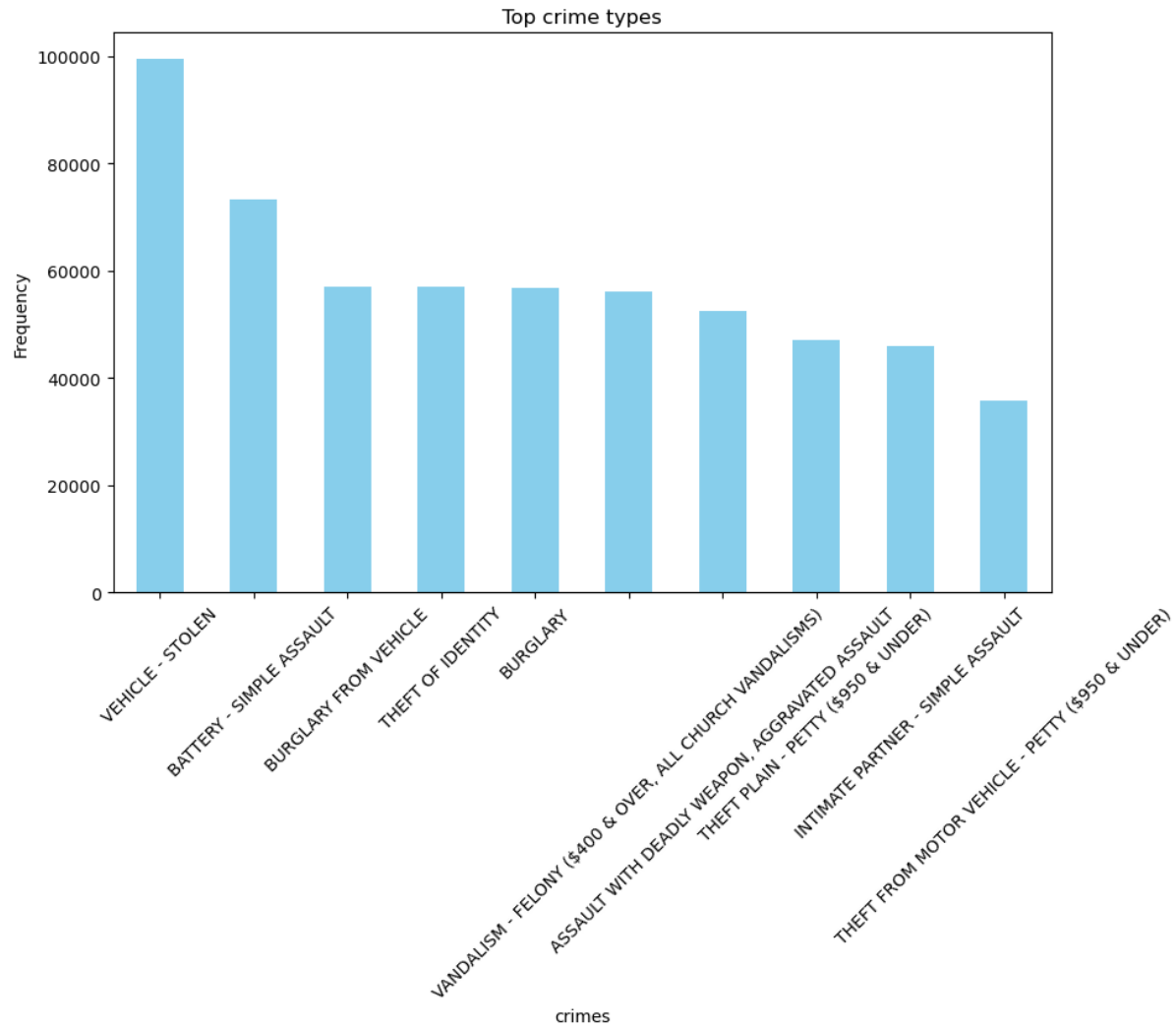
OBSERVATION:

The graph displays the distribution of crime rates across various areas. Upon examination, it's evident that the central area experiences the highest crime rate, with approximately 63,000 incidents recorded. Following closely is the 7th street area, reporting around 58,000 crimes. The Pacific, Southwest, and Hollywood areas exhibit moderate crime rates, while the foothill area records the lowest number of crimes, with approximately 31,000 incidents. This representation of area-wise crime rates underscores the disparities in crime prevalence across different regions and prompts further exploration into the factors influencing these variations.



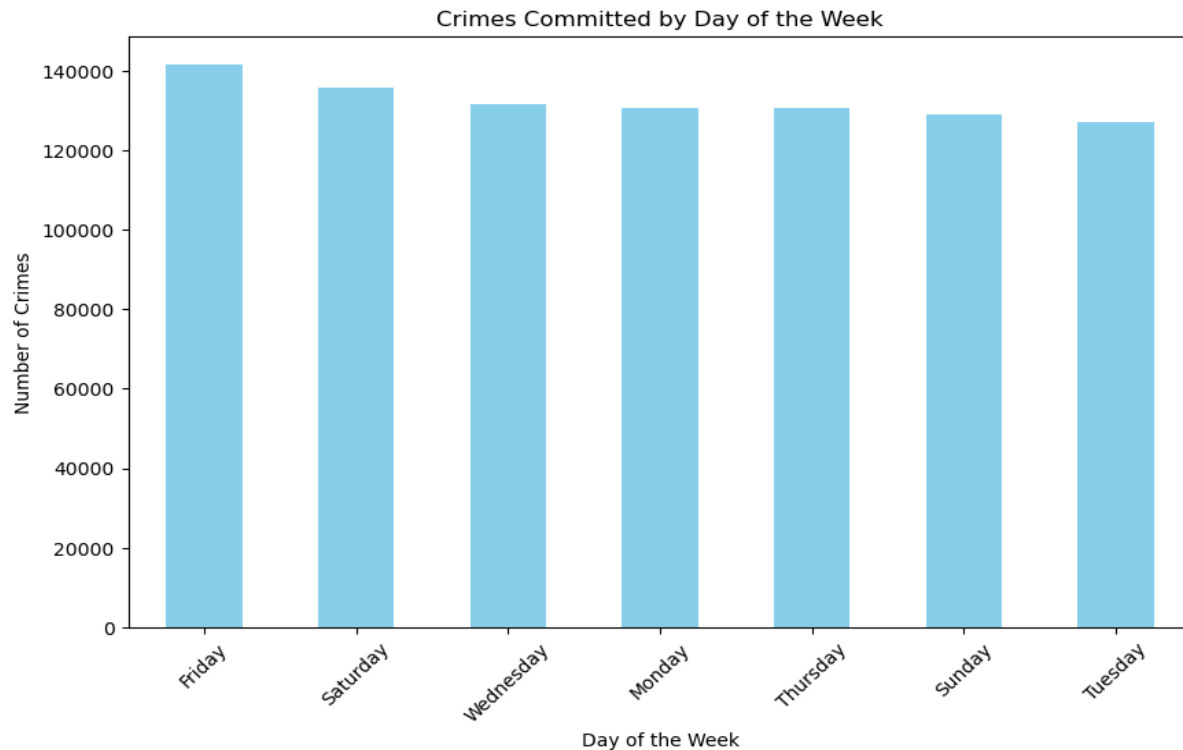
OBSERVATION:

The bar graph offers a straightforward depiction of incident frequencies across the top 10 locations. Remarkably, 800 N ALAMEDA stands out as the most frequently reported location, underscoring its significance within the dataset. This suggests a notable concentration of incidents at this specific location, surpassing others in the top 10 list. Such prominence indicates that 800 N ALAMEDA may serve as a hotspot for criminal activities or incidents warranting attention from law enforcement.



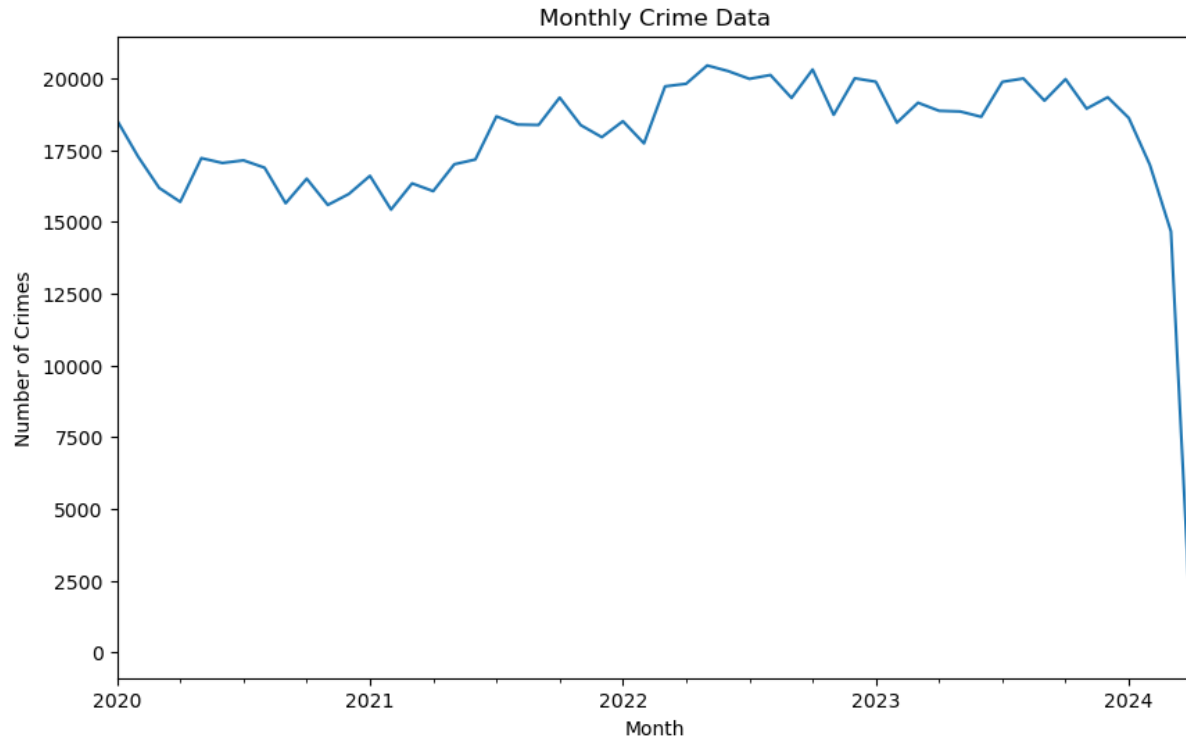
OBSERVATION:

Upon reviewing the provided data, it's clear that reported crimes encompass a diverse range of offenses. Particularly striking is the prevalence of "VEHICLE-STOLEN" incidents, which stands out among the rest. This category has garnered considerable attention due to its high frequency, with approximately 100,000 occurrences. Further analysis of the dataset underscores the significant concern posed by vehicle theft within law enforcement circles. The prevalence of such incidents underscores the urgent need for proactive measures to address this persistent challenge. Understanding the magnitude of this crime type is crucial for devising effective strategies to combat automotive theft and enhance community safety.



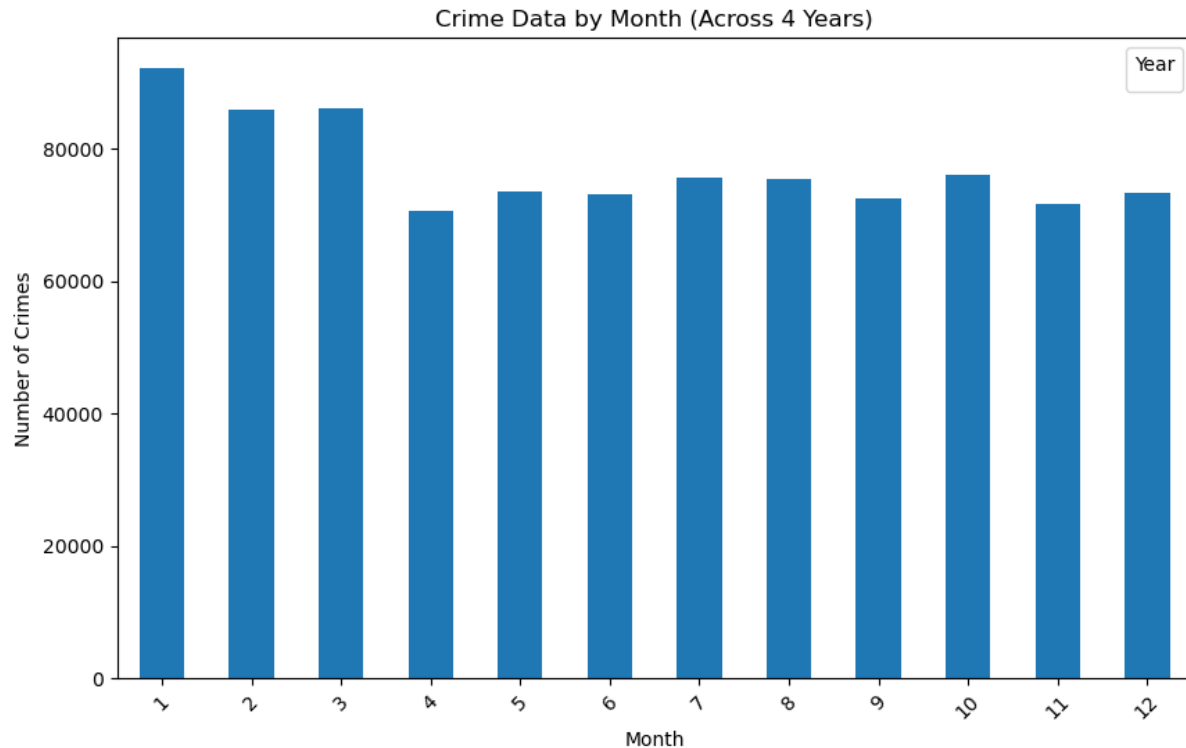
OBSERVATION:

The graph depicts the distribution of crimes across the days of the week. It's clear that the highest number of crimes occurs on Fridays, closely followed by Saturdays. Wednesdays and Mondays also show relatively high crime rates. In contrast, Tuesdays and Sundays experience lower numbers of reported crimes. This pattern suggests that the frequency of criminal incidents tends to peak towards the end of the week, gradually decreasing towards the beginning and middle. Insights into weekly crime patterns can inform law enforcement agencies in allocating resources and implementing targeted strategies to address fluctuations in crime rates across different days of the week.



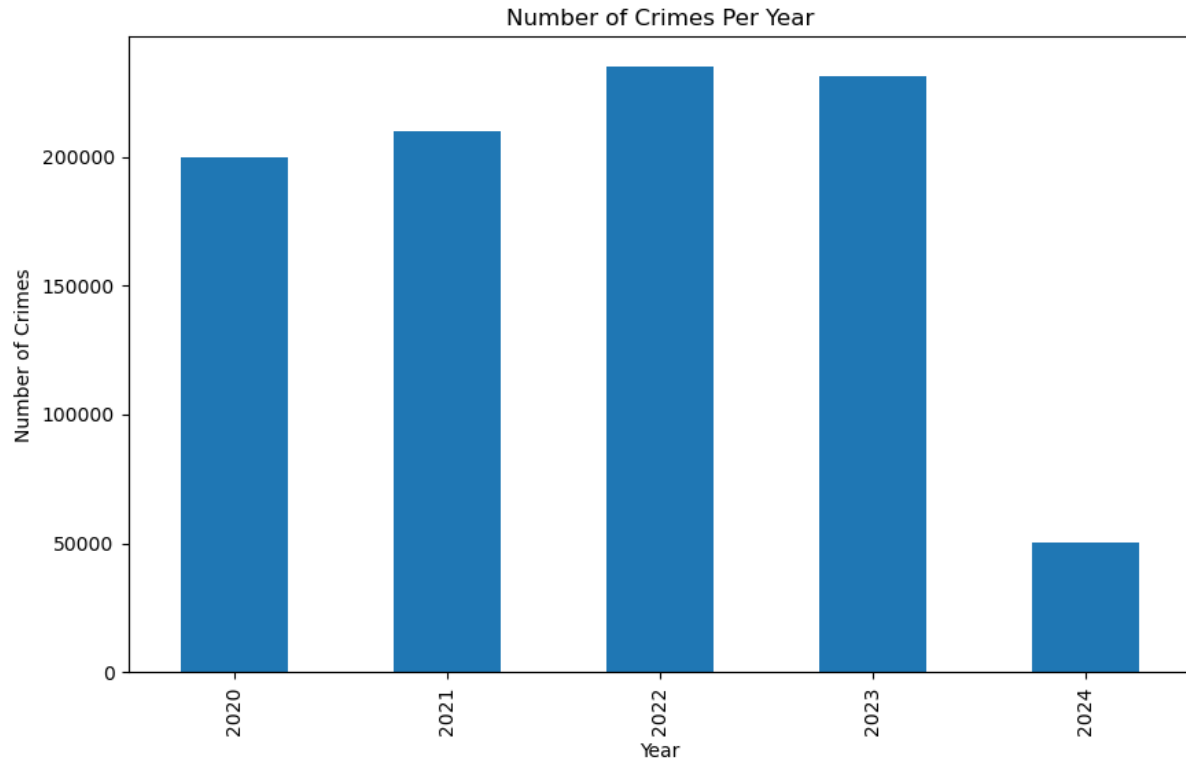
OBSERVATION:

Upon examining monthly crime data from 2020 to 2024, a distinct trend emerges: a surge in criminal incidents, peaking in May. This period experiences a notable increase, with approximately 200,000 reported crimes occurring within this timeframe. The observed pattern underscores the significance of seasonal variations in criminal activity, suggesting a potential link between environmental factors or societal influences and the occurrence of crimes during specific months.



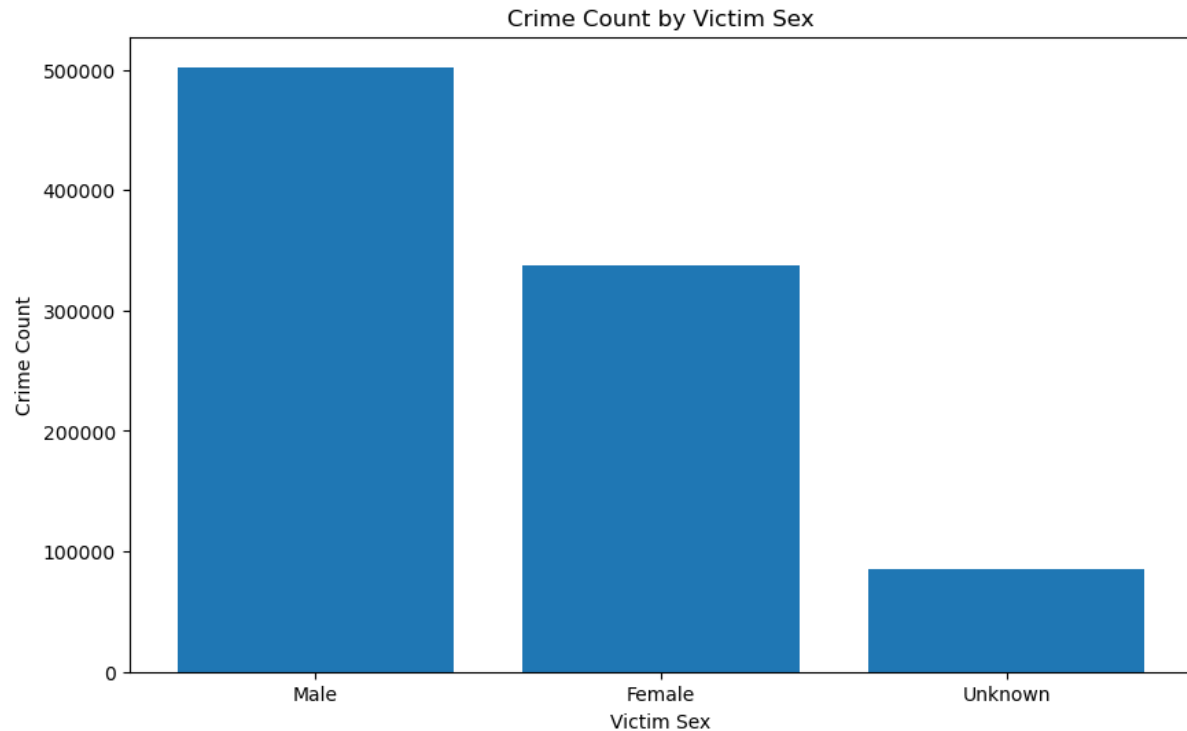
OBSERVATION:

The graph visually displays the frequency of crime incidents over four years, categorized by month. A notable observation is the prominence of January, which shows a significantly higher number of reported crimes compared to other months. Specifically, the data indicates that January accounts for over 800,000 recorded incidents. This substantial increase in crime occurrences during January hints at potential seasonal influences or other factors contributing to heightened criminal activity during this period. Further analysis and contextual understanding of these trends could provide valuable insights for law enforcement agencies and policymakers in addressing and mitigating crime rates during specific months.



OBSERVATION:

Since 2020, the LAPD has grappled with managing over 200,000 crime cases, highlighting the substantial challenges in upholding public safety. Moreover, there was a noticeable surge in crime rates during the transition from 2022 to 2023, prompting concerns among city residents about their safety and security. This increase in criminal activity during that period emphasizes the necessity for proactive measures and strategic interventions by law enforcement agencies to address emerging crime trends and safeguard the well-being of the community.



OBSERVATION:

The provided graph illustrates the crime count by victim sex. The bar graph distinctly shows the distribution of victims by gender, with a significant difference evident, notably higher numbers of male victims compared to females and other genders. This stark contrast underscores a prevailing trend of male victimization within the dataset, emphasizing the disparity in victim demographics.

7.CONCLUSION

In summary, our study delved into the extensive dataset provided by the Los Angeles Police Department (LAPD) covering the years 2012 to 2015, providing comprehensive insights into reported crimes in Los Angeles. Through thorough analysis, we uncovered significant patterns and trends in criminal behavior, illuminating various crime categories and geographical hotspots across the city.

The primary aim of our project is to support the LAPD in optimizing resource allocation for crime prevention efforts. By leveraging the insights derived from our analysis, law enforcement agencies can strategically allocate resources to areas with higher crime rates, thereby maximizing the effectiveness of their interventions. Moreover, our findings lay a valuable groundwork for the development of proactive law enforcement strategies aimed at reducing crime rates and enhancing overall public safety throughout Los Angeles.

Our study highlights the importance of data-driven approaches in guiding decision-making within law enforcement agencies. By harnessing the power of data analytics, we aim to equip the LAPD with actionable insights that enable them to effectively address the complex challenges posed by urban crime dynamics. Through collaborative efforts and continuous analysis, we are dedicated to contributing to the LAPD's mission of ensuring the safety and well-being of the residents of Los Angeles.