IE 6400 – FOUNDATIONS OF DATA ANAYTICS

Project 1
Crime Data Analysis

1. Introduction

Crime data analysis is essential for police departments to enhance public safety and reduce crime rates. Given the growing number of incidents, analysing crime trends allows for better decision-making and proactive measures.

In this project, we examine the **Los Angeles Police Department's Crime Dataset (2020–Present)**, which includes over **97,000 entries**. We selected the most relevant attributes for our analysis and visualization, aiming to derive insights that could help in the prevention and management of crime.

---

2. Project Workflow

2.1 Data Acquisition and Inspection

The initial step is to download the data from the publicly available source (Crime Data from 2020 to Present). The data is then loaded into the Jupyter notebook using Pandas Library. Pandas' library is used for facilitating easy manipulation and inspection of the data.

Steps followed:

1. pd.read was used to read the csv file.
2. .head() function was used to read the first few rows of the dataset.
3. .info() was used to check the information of the dataset such as which column has what data type.

2.2 Data Cleaning

Data Cleaning plays a crucial role in analysis. Here we addressed any inconsistencies that are present in the dataset and prepare the dataset for analysis.

1. Our first step of data cleaning involved which columns have how many NULL values in it.

a. From the above, we inferred that the Vict_Sex column had 32184 NULL values. Vict_Sex can have only three values, i.e., Male, Female or X(unknown).
b. We checked all the unique values in Vict_Sex to see the unique values in the dataset. Then we retained the information of only Male and Female present in the dataset.
c. Next we checked the values of the Vict_Age. It consisted of negative values. We know that Vict_Age cannot be negative since age cannot be negative.
d. We then plotted a graph to check if the data is normally distributed. Then we replaced the NULL Values with the mean value of the dataset.
e. We then dropped all the columns with higher number of NULL Values, and which were not relevant to our analysis ('Mocodes','Vict Descent', 'Premis Desc', 'Crm Cd 1','Crm Cd 2','Crm Cd 3','Crm Cd 4')

Now we have the cleaned dataset that is ready for the next step of Analysis.

2.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis holds a significant role in analysing the data. It helps us identify underlying patterns, spot anomalies etc before sending it to a formal modelling technique. This is often accompanied by visual methods.

2.4 Visualizations

Visualizations play a critical role in observing and understanding relationship between various attributes in our dataset.

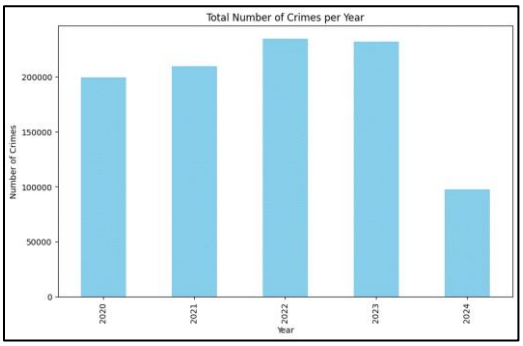
<table>
<tr>
<td> 

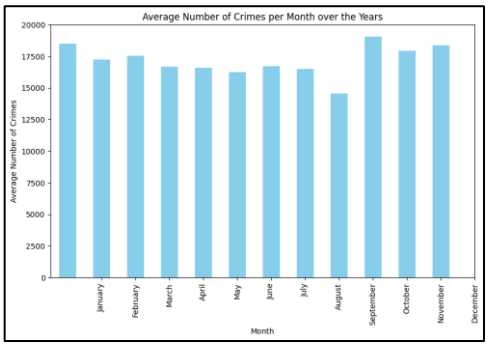
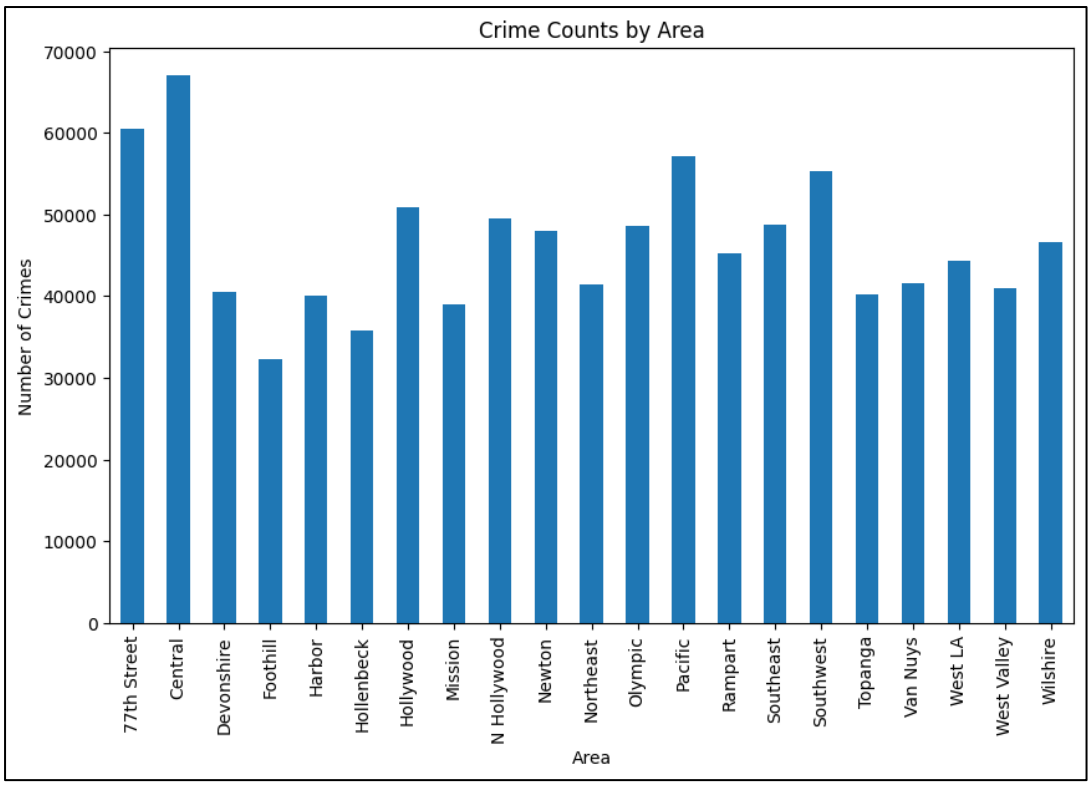
Chart 1</td>
<td> 

Chart 2</td>
</tr>
<tr>
<td>Visual 1 represents the total number of crimes from the year 2020 to 2023. The graph shows an upward trend from year 2020 – 2023 but has slightly declined in the year 2024. Also, from observation from the visual the highest number of crimes occurred in the years 2022 and 2023, with a slight decrease in 2023 compared to 2022. There is a significant drop in the total number of crimes for 2024 compared to previous years.</td>
<td>Visual 2 shows the average number of crimes per month. It remains comparatively stable, except for a visible drop in August. September and October show the highest monthly averages, indicating increased crime activity in these months. It is also observed that other months like January, February, and November also have high average crime rates, but they are slightly lower than in September.</td>
</tr>
</table>



The third chart, titled "Crime Counts by Area", shows the distribution of crime across various areas. **77th Street** and **Central** are clear outliers in terms of crime frequency, each having close to or more than 45,000 recorded crimes. In both areas, **vehicle theft** (blue bars) is the most common crime type.

- **77th Street** is known for its high volume of criminal activity. This could be due to its urban landscape, dense population, and socioeconomic factors contributing to higher levels of crime.
- **Central** also ranks high for vehicle thefts, reflecting the area's traffic density and the prevalence of public parking, which creates opportunities for car-related crimes.
- **Van Nuys** stands out with an exceedingly high incidence of **burglary** (red bar). This could indicate that this area is particularly susceptible to property crimes, due to more affluent neighborhoods, which are often targeted for home break-ins.
- **Devonshire** and **Harbor** areas, while showing substantial crime numbers, have **vehicle theft** as their most common crime. This suggests that the crime pattern in these regions may be like the 77th Street and Central areas, with vehicle theft being a widespread issue across various parts of the city.

- The areas like **West LA**, **Wilshire**, and **Topanga** show a more moderate crime rate, but the most prevalent crime type varies. For example:

- **West LA** has a prominent level of **burglary from vehicles** (orange bar), due to higher car ownership and accessible street parking.
- **Topanga** and **Wilshire** have lower crime counts overall, but each region shows different common crimes, which might reflect their unique socio-economic and demographic characteristics.

Key Insights

- **Vehicle theft** is a significant problem across many regions, making it the dominant crime type in most areas, particularly in the dense urban zones.
- **Burglary** and **burglary from vehicles** appear frequently in specific areas like **Van Nuys** and **West LA**, signalling potential targets of these crimes, related to higher incomes and wealthier neighborhoods.
- **Battery/simple assault** (green bar) has less dominance across most areas but is present in a few, indicating physical confrontations or violent interactions in smaller portions of the city.

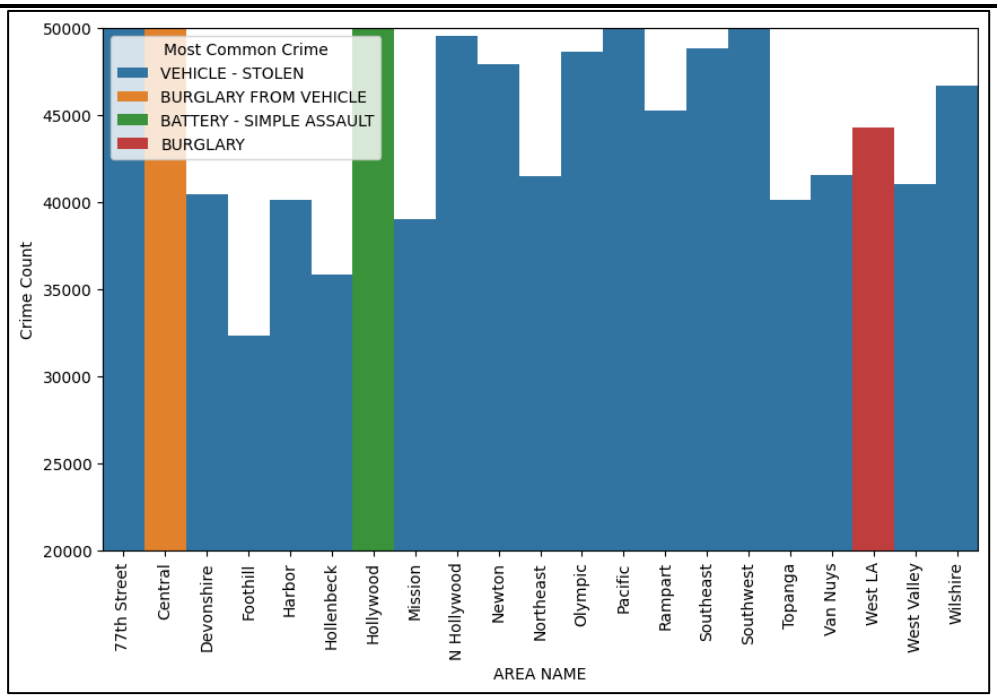
```
The most frequent crime type is 'VEHICLE - STOLEN' with 109087 occurrences.

Top 5 most common crimes:
'VEHICLE - STOLEN': 109087 occurrences
'BATTERY - SIMPLE ASSAULT': 74591 occurrences
'BURGLARY FROM VEHICLE': 60668 occurrences
'THEFT OF IDENTITY': 60258 occurrences
'VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)': 59088 occurrences
```

The table shows the count of different crimes that have occurred over the years, few of them being,

Vehicle Theft is the Most Prevalent Crime with crime frequency of 189,887 occurrences. This indicates a significant problem with car theft in the areas. The next most common crime type is Violent Crimes and Property Crimes, ranks second with 74,591 occurrences, suggesting a prevalence of minor physical altercations. "Burglary from Vehicle" and "Vandalism Felony" are also among the top 5, highlighting issues with property theft and damage. Other common kind of crime is Identity Theft, it is the fourth most common crime, with 68,258 occurrences. Overall, the data reveals a mix of violent and property crimes, with vehicle theft being the most significant issue in the area. Further analysis could explore the underlying causes of these crimes, potential prevention strategies, and the impact on the community.

Other Key Findings:

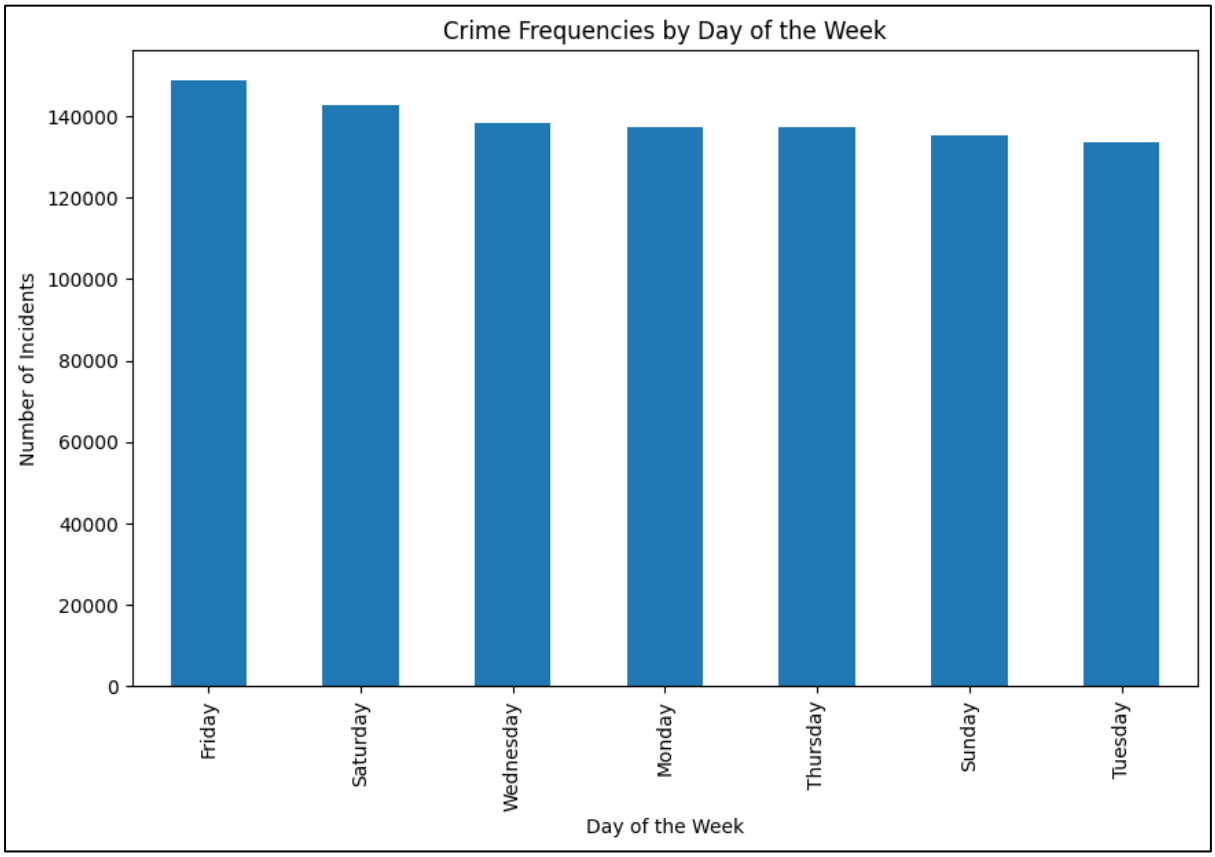
- **Area Specific Crime Patterns:** While vehicle theft is widespread, there are some variations in the most common crimes across different areas:
    - **Central:** Burglary from vehicles is the most common crime, suggesting a focus on targeting parked cars for theft.
    - **West LA:** Burglary is the most common crime, indicating a higher incidence of break ins into homes or businesses.
- **Consistent Crime Types:** Despite area specific variations, vehicle theft remains a consistently high-ranking crime type across most of the listed areas.



The histogram represents Most Common Crimes in the city of Los Angeles. The insights show:

- **Depicts the distribution of vehicle theft**: The tallest bars, which stand for "VEHICLE STOLEN," are constantly high in most of the areas, suggesting that this is the most common sort of crime in Los Angeles.
- **Area-Specific Crime Patterns**: The histogram reveals that varied locations have varied heights of bars for certain crimes (assault, burglary, etc.), indicating that the city's crime patterns are not all the same. Certain crimes are more common in some places than others, and their distributions may be more balanced in others.

- **Outliers**: Some neighbourhoods, including Newton and 77th Street, have noticeably taller bars overall, indicating that they may be crime hotspots.



- **Friday** stands out as having the highest number of crimes, closely followed by **Saturday**. These two days are significantly more crime-prone than the other days of the week. This trend can often be attributed to increased social activities and public gatherings during weekends. Bars, nightclubs, and other social venues are more active during these days, which often leads to more opportunities for crimes like theft, assault, and disturbances.

Additionally, alcohol consumption, which tends to be higher on weekends, could contribute to the spike in criminal activities on Fridays and Saturdays.

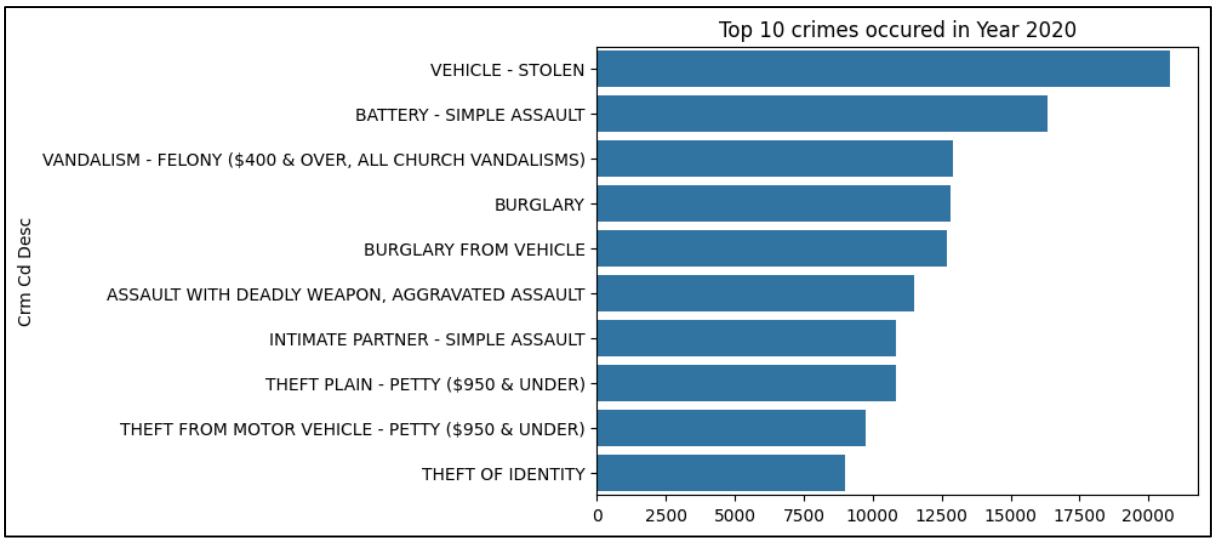
- **Wednesday**, **Thursday**, and **Monday** have consistent crime frequencies, though these days still see more incidents than **Sunday** and **Tuesday**. This middle-of-the-week steadiness could be due to routine working days, where crime opportunities remain moderate.
- **Sunday** and **Tuesday** exhibit the lowest crime rates. Sunday, being a day when many people stay home or engage in family or religious activities, sees fewer

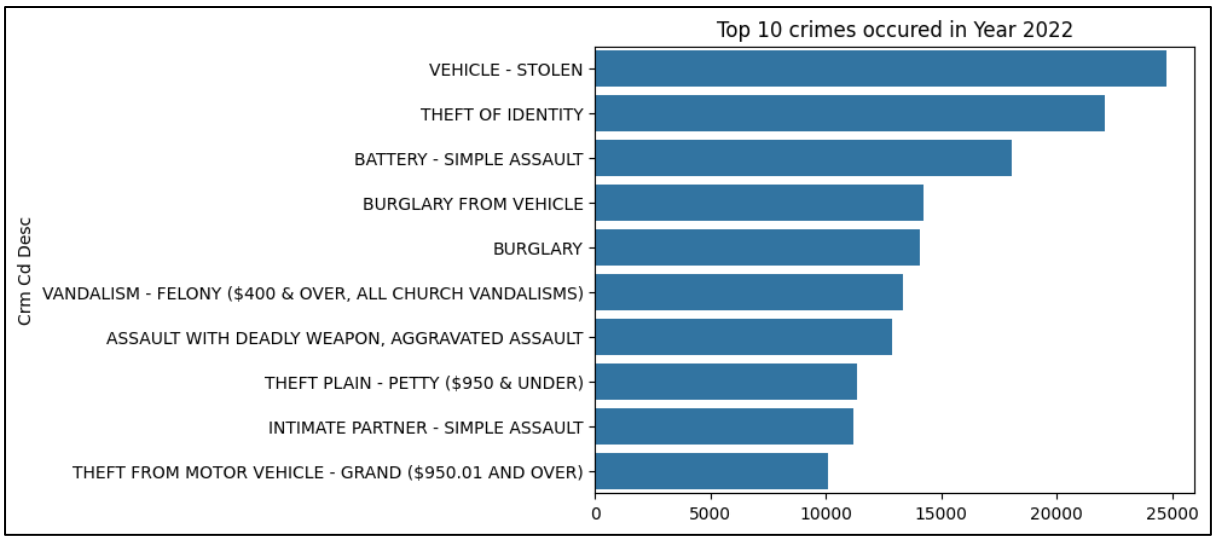
opportunities for crimes, particularly those involving large gatherings or public spaces.

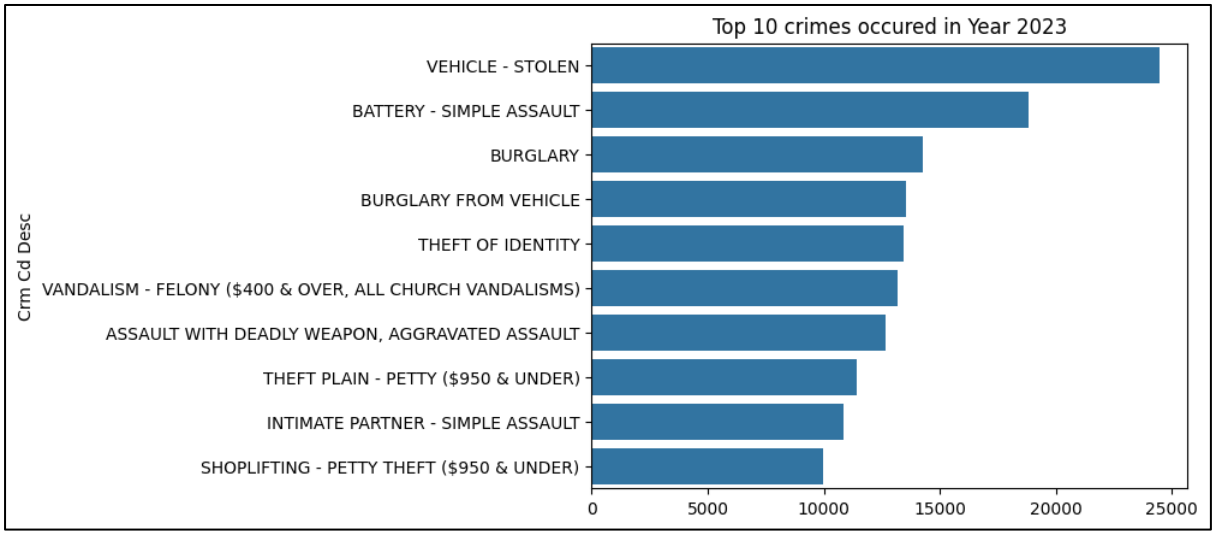
Key Insights

- The significant uptick in criminal incidents on **Friday** and **Saturday** suggests a need for increased law enforcement presence and resources during these days. Patrols around nightlife districts or public spaces could be strengthened to pre-empt criminal behaviour.
- **Monday**, **Wednesday**, and **Thursday** see steady crime levels, indicating that crime is not confined to weekends but occurs consistently throughout the week. However, the type of crime might differ from what is observed over the weekends, potentially reflecting crimes related to work environments or routine weekday activities.
- **Sunday** is the least crime-ridden day, offering a respite from the spike seen on Fridays and Saturdays. Law enforcement could leverage this trend to allocate resources more effectively, scaling back on Sundays and deploying more officers during the high-risk periods.

This temporal analysis is invaluable in understanding when crimes are most likely to occur, allowing for optimized scheduling of law enforcement efforts and public safety campaigns. Knowing that Fridays and Saturdays are peak times, authorities can adjust their strategies, accordingly, potentially reducing crime by intervening proactively.







Year 2020

- **Most Frequent Crime:** Vehicle theft ranked as the highest-reported crime, with close to 20,000 cases.
- **Battery - Simple Assault** was the second most frequent, with significant reporting.
- **Vandalism (Felony)** ranked third, indicating a notable issue with property damage.
- **Burglary** and **Burglary from Vehicle** were also highly prevalent.
- **Theft of Identity** and **Theft from Motor Vehicle (Petty)** appeared at lower ranks, but still represent key concerns.

3.2 Year 2022

- **Vehicle Stolen** remained the most frequently reported crime, with an increase in incidents compared to 2020, reaching 25,000 cases.

- **Theft of Identity** saw a significant rise, ranking second in 2022, overtaking **Battery - Simple Assault**.
- **Burglary and Burglary from Vehicle** continued to be persistent crimes.
- **Vandalism (Felony)** remained prevalent but showed no notable change in ranking.
- **Theft from Motor Vehicle (Grand)** appeared in the top 10, indicating an increase in higher-value vehicle thefts.

3.3 Year 2023

- **Vehicle Stolen** continued to dominate crime reports with a similar frequency as in 2022, remaining the most reported crime.
- **Battery - Simple Assault** regained its position as the second-highest crime.
- **Burglary** and **Burglary from Vehicle** maintained their high frequency, indicating ongoing challenges with property-related crimes.
- **Theft of Identity** saw a drop in ranking compared to 2022, suggesting a potential improvement in identity theft prevention or reporting.
- **Shoplifting (Petty Theft)** emerged in 2023, showing a notable rise in petty theft, while **Theft from Motor Vehicle (Grand)**, which was significant in 2022, was no longer in the top 10.

4. Trend Analysis

4.1 Consistent Trends

- **Vehicle Stolen** consistently ranked as the most frequent crime across all three years. This indicates a persistent issue with motor vehicle theft, which remains a critical concern for law enforcement.
- **Battery - Simple Assault** also remained in the top three throughout the period, underscoring the continued presence of interpersonal violence or physical altercations.
- **Burglary and Burglary from Vehicle** were consistent across the years, reflecting an ongoing challenge with property crime.

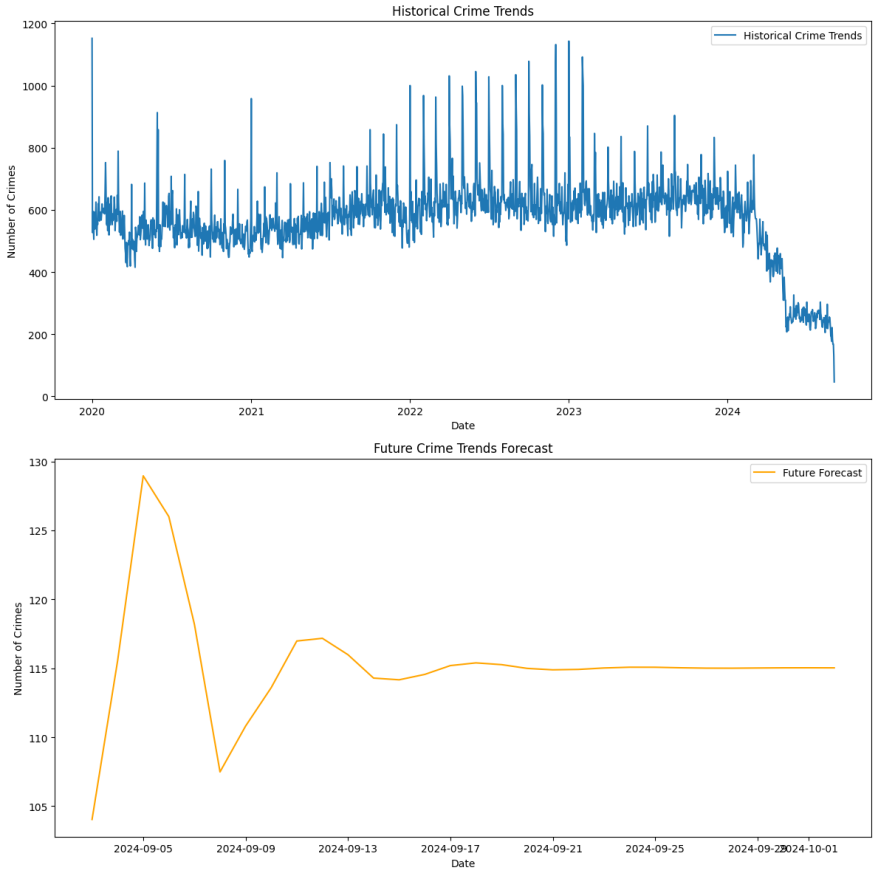
4.2 Emerging Patterns

- **Theft of Identity** peaked in 2022 but dropped significantly by 2023, which may indicate changes in either prevention measures or reporting practices.

- **Shoplifting (Petty Theft)** entered the top 10 in 2023, reflecting a shift towards an increase in minor thefts.

4.3 Changes in Crime Distribution

- **Theft from Motor Vehicle (Grand)** was notable in 2022 but dropped out of the top 10 in 2023, indicating a reduction in more severe vehicle-related thefts.
- The rise of **Shoplifting (Petty Theft)** in 2023 suggests a shift towards smaller, more frequent thefts, which may indicate economic factors influencing crime patterns.



The graphs show charts representing Historical Crime Trends and Future Crime Trends Forecast.

1. Historical Crime Trends:

Time Period: The chart covers from 2020 to 2024, displaying the number of crimes over time.

Crime Patterns:

There is some fluctuation in the number of crimes over the years, with occasional spikes.
Steady crime rate: From 2020 to early 2023, crime counts seem stable, though there are minor peaks and troughs.

Decline in 2024: After 2023, the chart shows a sharp decline in the number of crimes, particularly as it nears mid-2024.

Key Insights:

The significant drop in crime towards the end of the historical data (2024) could indicate a change in policy, enforcement, or other socioeconomic factors.

2. Future Crime Trends Forecast:

Forecast Period: This chart predicts future crime trends from early September to October 2024.

Pattern:

The forecast begins with a sharp increase in crimes, reaching around 130 by early September. After this spike, the forecast stabilizes around 110120 crimes for the remainder of September. The forecast flattens towards the end, suggesting a period of relative stability.

Key Insights:

There is an expected short-term rise in crime, followed by stabilization in the number of crimes over the forecast period.

Overall Summary:

Historical trends show fluctuating crime rates with a sharp drop towards mid2024.

Future forecasts suggest a brief rise in crime followed by stable levels.

This could indicate some intervention or policy change affecting crime rates in 2024 and beyond.

Conclusion

This project successfully implemented a detailed data cleaning and exploratory analysis of crime data from 2020 onwards. By visualizing trends, identifying common crime types, and comparing regional differences, the project delivered actionable insights that can be used for further investigation or to prevent future crimes and deploy strategies.
