# "Exploratory Analysis of Crime Incidents: Unveiling Patterns, Trends, and Correlations in Urban Safety"

## The issues:

The dataset for crime-incident-reports raises the following questions:

1. **Temporal Patterns:**
    - Trends Over Years: How have crime incidents evolved over the years, and what are the trends in annual crime counts?
    - Seasonal Variations: Are there specific months or seasons when certain types of crimes are more prevalent?
    - Day and Shift Distribution: How does the distribution of incidents vary across different days of the week and shifts?
    - Shift-specific Crime Types: Do certain types of crimes occur more frequently during particular shifts (day, evening, night)?
2. **Geospatial Analysis:**
    - Crime Hotspots: What are the identified crime hotspots in terms of districts, reporting areas, or specific streets?
    - Correlation with Locations: Is there a correlation between certain types of crimes and specific geographical locations?
3. **Nature of Crimes:**
    - Distribution of Crime Types: What is the distribution of different crime types (e.g., residential burglary, aggravated assault, robbery)?
    - Common Crime Codes: Are there specific crime codes that are more common than others?
4. **Weapon Type Analysis:**
    - Frequency by Day: What is the frequency of incidents for each day of the week based on the weapon type used?
    - Unarmed Incidents: Why does the highest frequency of incidents occur when the weapon used is "Unarmed"? What are the potential explanations for this observation?
5. **Month-wise Analysis:**
    - Monthly Fluctuations: How do crime rates vary over the months, and what factors might contribute to these fluctuations?

- July and February Comparison: Why does July have the highest incidents, and why is February the lowest? What seasonal, weather-related, or cultural factors could be influencing these patterns?

6. **General Insights:**
    - Spatio-temporal Characteristics: How do spatio-temporal characteristics contribute to a comprehensive understanding of crime phenomena?
    - Policy and Enforcement Impact: How might changes in law enforcement strategies or policies influence crime rates over the years?

These questions cover a wide range of aspects, providing a holistic view of the dataset and enabling informed decision-making for law enforcement and public safety measures in Boston.

## Findings:

1. **Trends Over Years:**
    - The analysis reveals a moderate variability in annual crime counts, with a noticeable increase from 2012 to 2015. Possible explanations include changes in law enforcement strategies or policies influencing crime rates.
2. **Seasonal Variations:**
    - Fluctuations in crime rates over the months indicate seasonal patterns. July consistently has the highest incidents, possibly due to increased outdoor activities and gatherings. February records the lowest incidents, likely reflecting reduced overall activity during the winter.
3. **Day and Shift Distribution:**
    - Fridays witness the highest number of crimes, while Sundays have the lowest, suggesting a correlation between the day of the week and crime rates. Regarding shifts, the majority of crimes occur during the Day and First shifts, with significantly fewer incidents during the last shift.
4. **Shift-specific Crime Types:**
    - Further analysis can explore whether specific types of crimes are more prevalent during particular shifts, providing insights into the temporal dynamics of crime.
5. **Crime Hotspots:**
    - Geospatial mapping identifies areas with high crime density, indicating crime hotspots. Detailed analysis can pinpoint specific districts, reporting areas, or streets associated with elevated crime levels.

6. **Correlation with Locations:**

    - Geospatial analysis can uncover correlations between certain types of crimes and specific locations, contributing to targeted crime prevention strategies.

7. **Distribution of Crime Types:**

    - The bar chart visualizes the distribution of different crime types, offering a quick overview of prevalent categories within the dataset.

8. **Common Crime Codes:**

    - Additional exploration can identify specific crime codes that are more common, providing detailed insights into the nature of crimes in the dataset.

9. **Frequency by Day:**

    - The analysis of weapon types reveals that the highest frequency of incidents on every day of the week occurs when the weapon used is "Unarmed," suggesting a significant portion of reported incidents involves perpetrators without weapons.

10. **Unarmed Incidents:**

    - Possible explanations for the high frequency of incidents involving "Unarmed" weapons include non-violent encounters, verbal altercations, or incidents where weapon information is not accurately recorded.

11. **Monthly Fluctuations:**

    - The monthly analysis indicates fluctuations in crime rates, and factors such as seasonal changes, weather conditions, or cultural events may contribute to these variations.

12. **July and February Comparison:**

    - July's higher incidents may be due to increased outdoor activities, while February's lower count could result from reduced overall activity during winter. Other factors like holidays and local events could also influence varying crime rates.

13. **Spatio-temporal Characteristics:**

    - The dataset's spatio-temporal characteristics allow for a comprehensive understanding of crime phenomena, unveiling patterns, trends, and relationships over both space and time.

14. **Policy and Enforcement Impact:**

    - Changes in law enforcement strategies or policies implemented over the years may contribute to the observed trends in annual crime counts. Detailed policy analysis can provide a deeper understanding of their impact.

## Discussions:

We picked the dataset "crime-incident-reports-july-2012-august-2015-source-legacy-system" from 'Analyse Boston.' This dataset stands out for having both spatial (location-based) and temporal (time-based) information, allowing us to explore how things unfold over space and time. Each event in this data is tied to a specific location and time, offering a deep dive into patterns and trends. Analyzing this data can provide valuable insights into dynamic processes for informed decision-making.

The dataset includes features like Incident Type, Weapon Type, Shift, Year, Month, Day of the week, and Location. Our first analysis focused on crime incidents by the day of the week, showing that Fridays had the highest and Sundays the lowest occurrences. We then mapped crime locations on a Boston map, revealing areas with high crime density.

Exploring crime patterns throughout the day, we found that the Day shift and the First shifts experienced more incidents compared to the last shift. This suggests a potential link between the time of day and crime rates, with higher activity during daylight hours.

Examining the "Number of Crimes per Year," we noted a minimum in 2012 (around 43,186 crimes) and a maximum in 2015 (about 88,058 crimes). Changes in law enforcement strategies or policies over the years could explain this variation. The analysis indicated moderate variability in annual crime counts, with a noticeable increase from 2012 to 2015.

A bar chart illustrated the prevalence of different crime types, with "VAL" (Validation/Field Interview) being the most frequent and "GAMBLING OFFENSE" the least. The high frequency of "Unarmed" incidents on all days suggests a significant portion of reported incidents involves unarmed individuals.

Finally, analysing monthly incidents showed fluctuations, with July having the highest and February the lowest. This could be influenced by seasonal changes, weather conditions, or cultural events, impacting crime rates. The higher incidents in July might be due to increased outdoor activities, while the lower count in February could result from reduced overall activity during winter. Factors like holidays and local events might also contribute to varying crime rates across months.

## Appendix A: Method

We chose the dataset “crime-incident-reports-july-2012-august-2015-source-legacy-system” from ‘Analyse Boston’ for our analysis. Upon our first review, we observed that the dataset possesses spatio-temporal characteristics, indicating it contains information involving both spatial (location-based) and temporal (time-based) components. This dual aspect allows for a thorough exploration of how phenomena develop across both space and time. In spatio-temporal data, each event or observation is tied to a particular geographical location and timestamp. Analysing such data provides the opportunity to unveil patterns, trends, and relationships that unfold over space and time. This approach yields valuable insights into dynamic processes, supporting more informed decision-making.

The dataset comprised of features like Incident type Description, Weapon Type, Shift, Year, Month, Day of the week, Location and more that we used for our analysis. Upon careful examination of the data, we uncovered intriguing aspects to investigate.

We initiated our analysis by examining crime incidents according to the day of the week, utilizing the "DAY_WEEK" column in the dataset. Tallying the occurrences for each day, we generated a bar graph to depict the distribution, revealing that Fridays witnessed the highest number of crimes, while Sundays had the lowest.

Our next step involved geospatial mapping of the data. On mapping the location coordinates of the crime incidents onto the Boston map, we observed that majority of the areas exhibited high crime density. As a next step, we delved further into the data to examine the prevalence of crimes during various time intervals throughout the day. Our investigation into the distribution of data across different shifts uncovered a notable pattern: the majority of crimes occurred during the Day shift and the First shifts, whereas the last shift exhibited considerably fewer incidents. This finding hints at a possible connection between the time of day and crime rates, indicating higher activity during daylight hours in contrast to nighttime.

We then analysed the data based on the "Number of Crimes per Year" to understand how the data is distributed over the years which showed that the minimum crime count occurred in 2012 with approximately 43,186 crimes, while the maximum was recorded in 2015 with around 88,058 crimes. One possible explanation of this could be changes in law enforcement strategies or policies implemented over the years, influencing crime rates. The analysis suggests a moderate variability in annual crime counts, with a noticeable increase from 2012 to 2015. The interquartile range (IQR) of approximately 39,187 demonstrates the spread of crime counts within the middle 50% of the data.

To understand the prevalence of different types of crime incidents, we generated a simple bar chart depicting the frequency of each incident category. This visual representation allows for a quick overview of the most common types of crimes occurring in the dataset. Analysing the bar chart provides valuable insights into the distribution of crime occurrences and highlights the predominant categories within the dataset. The most frequent incident type, "VAL" (Validation/Field Interview), suggests common interactions for identity verification or investigation. Conversely, the least frequent type, "GAMBLING OFFENSE," indicates rare occurrences, possibly reflecting lower incidence or less law enforcement focus on gambling-related issues in the dataset.

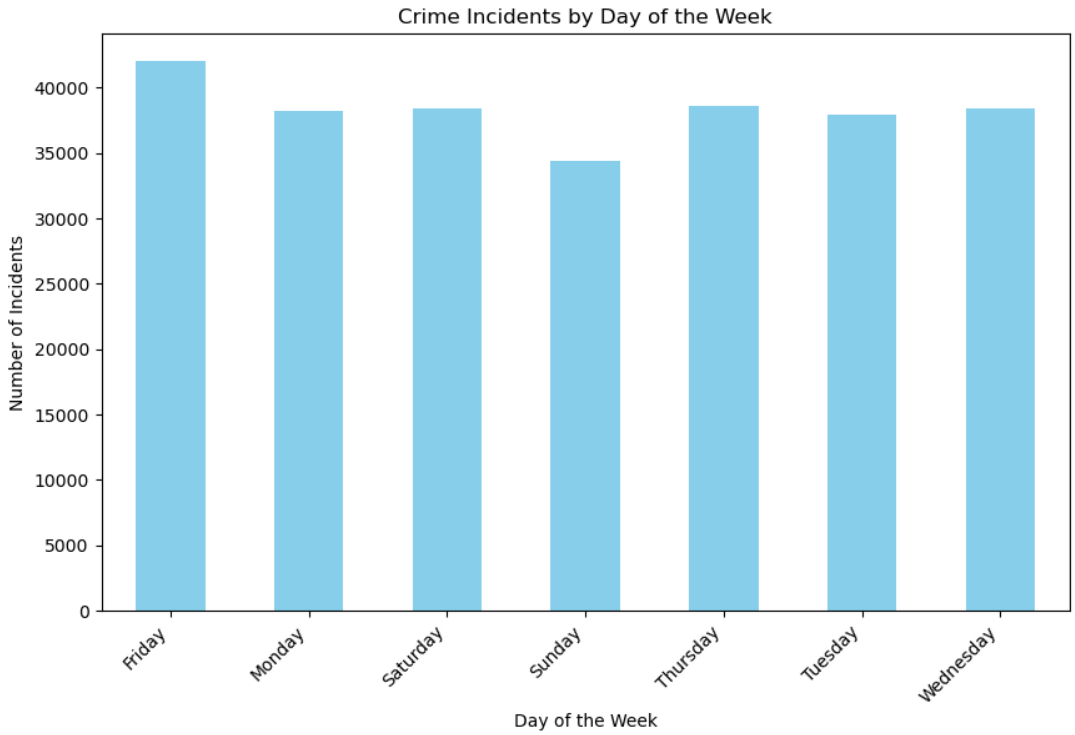
The next part of our analysis involved the Frequency of weapon type for each day of the week which reveals that the highest frequency of incidents on every day of the week occurs when the weapon used is "Unarmed." This observation may suggest that a significant portion of reported incidents involves perpetrators without weapons or those classified as "Unarmed." Possible explanations include a variety of non-violent or non-lethal encounters, such as verbal altercations, physical confrontations without weapons, or incidents where the weapon information is not available or recorded accurately in the dataset.

Lastly, we analysed the number of incidents based on the various months of the year and found that the data shows fluctuations in crime rates over the months, with July having the highest incidents and February the lowest. This pattern may be influenced by factors like seasonal changes, weather conditions, or cultural events. For instance, the higher number in July could

be due to increased outdoor activities and gatherings, while the lower count in February might result from reduced overall activity during the winter. Other elements like holidays and local events could also play a role in the varying crime rates across months.

## Appendix B: Results

**Plot on Incidents by Days of the week:**

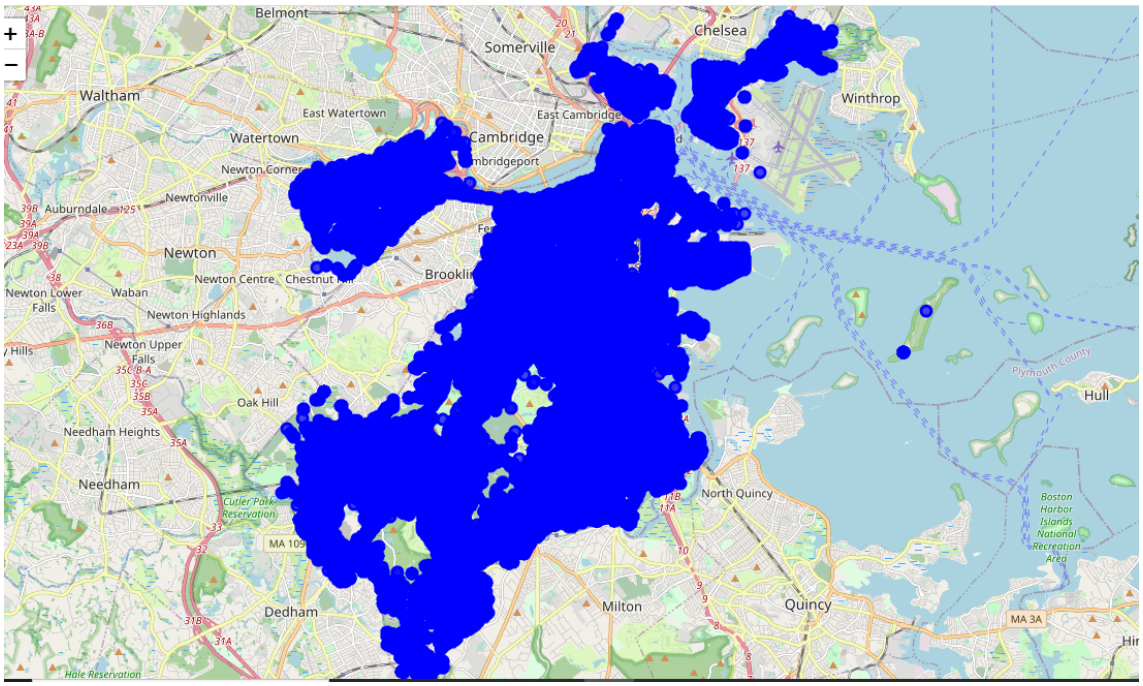



```
Number of Crimes per Week:
Friday       42031
Monday       38256
Saturday     38446
Sunday       34393
Thursday     38627
Tuesday      37911
Wednesday    38392
Name: DAY_WEEK, dtype: int64

Day with Maximum Crimes: Friday
Day with Minimum Crimes: Sunday
```

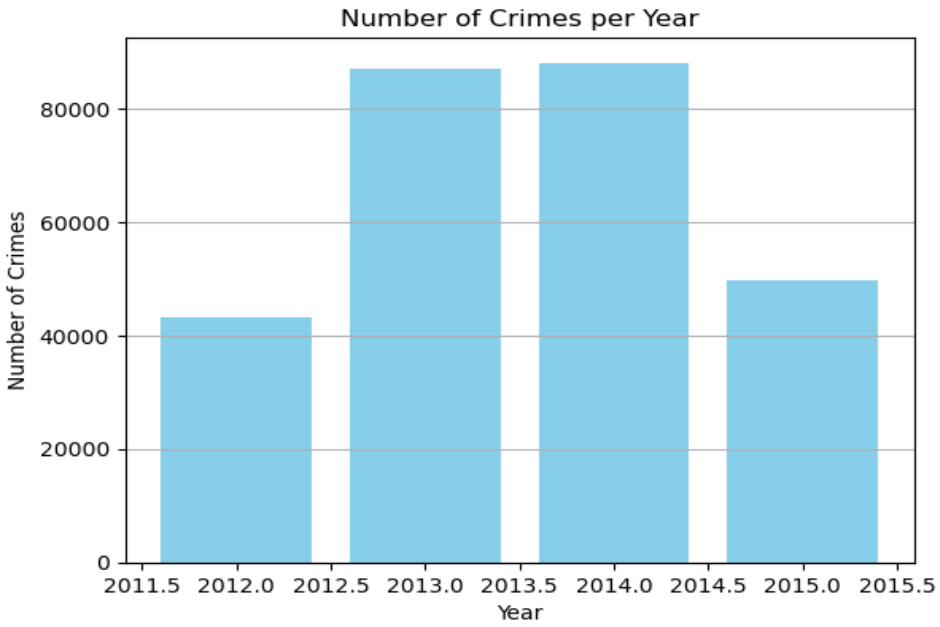
The heightened criminal activity on Fridays may be influenced by factors such as increased social activities, gatherings, or nightlife, warranting targeted interventions on specific days. Conversely, the lower incidence on Sundays might be attributed to reduced public events or heightened law enforcement presence. Further exploration and consideration of socio-economic factors could provide a more comprehensive understanding of these patterns.

## Geospatial plot of the data using Folium:




The observation that majority of areas in Boston exhibit high crime density based on the spatial distribution of crimes suggests potential hotspots or regions with elevated criminal activity. This information can be crucial for law enforcement and urban planning initiatives. Identifying areas with higher crime rates can aid in resource allocation, strategic policing efforts, and the development of targeted intervention programs to enhance public safety and reduce criminal incidents in those specific areas. It highlights the importance of understanding the geographic patterns of crime to implement effective crime prevention and control measures.

## Distribution of Number of Crimes per Year:

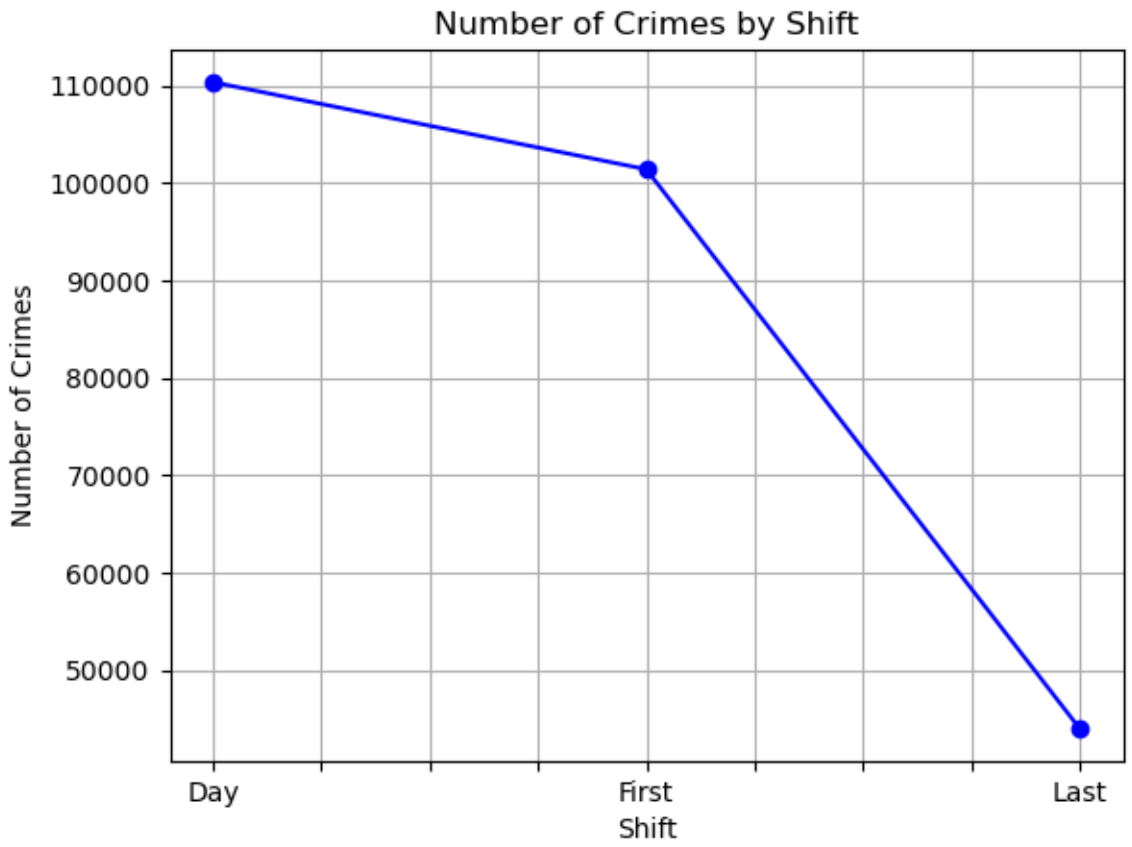


Statistics:

| | Year | Crime Count |
|---|---|---|
| count | 4.000000 | 4.00000 |
| mean | 2013.500000 | 67014.00000 |
| std | 1.290994 | 23873.59406 |
| min | 2012.000000 | 43186.00000 |
| 25% | 2012.750000 | 48116.50000 |
| 50% | 2013.500000 | 68406.00000 |
| 75% | 2014.250000 | 87303.50000 |
| max | 2015.000000 | 88058.00000 |

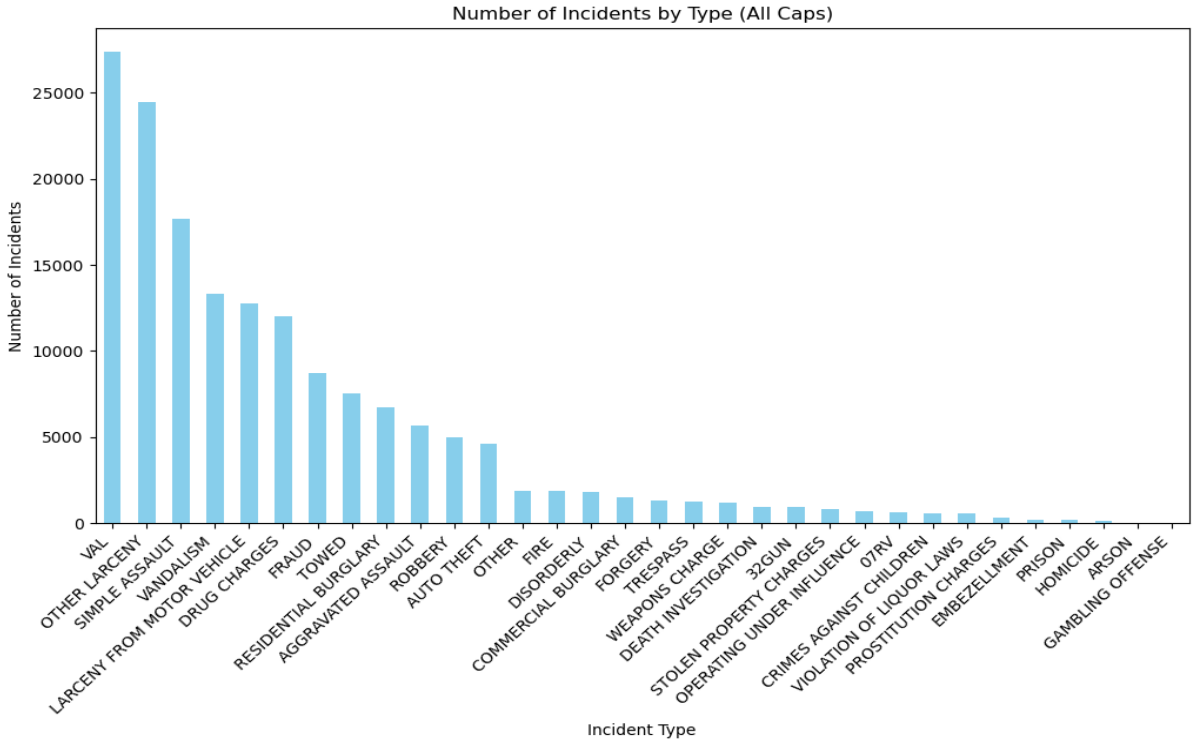
There are four data points corresponding to the years 2012, 2013, 2014, and 2015. The average number of crimes per year is approximately 67,014. This provides a central measure around which the annual crime counts revolve. The standard deviation of about 23,874 indicates a moderate level of variability or dispersion in the annual crime counts. This suggests that crime rates exhibit some fluctuation from year to year. The minimum crime count occurred in 2012 with approximately 43,186 crimes, while the maximum was recorded in 2015 with around 88,058 crimes. The quartiles provide insights into the distribution of crime counts. The first quartile (Q1) at 48,116.5 indicates that 25% of the years had crime counts below this threshold. The median (Q2) or the middle value is 68,406, and the third quartile (Q3) at 87,303.5 represents the point below which 75% of the years fall.

**Line Graph for Number of Crimes per Shift:**



The line chart illustrates that the highest number of crimes occurred during the Day shift and the First shifts of the day, while the last shift recorded significantly fewer incidents. This observation suggests a potential correlation between the time of day and crime rates, with increased activity during daylight hours compared to nighttime. Further analysis could delve into the factors contributing to this temporal variation, such as increased visibility and human activity during the day, and explore strategies to address and mitigate crime during specific shifts.

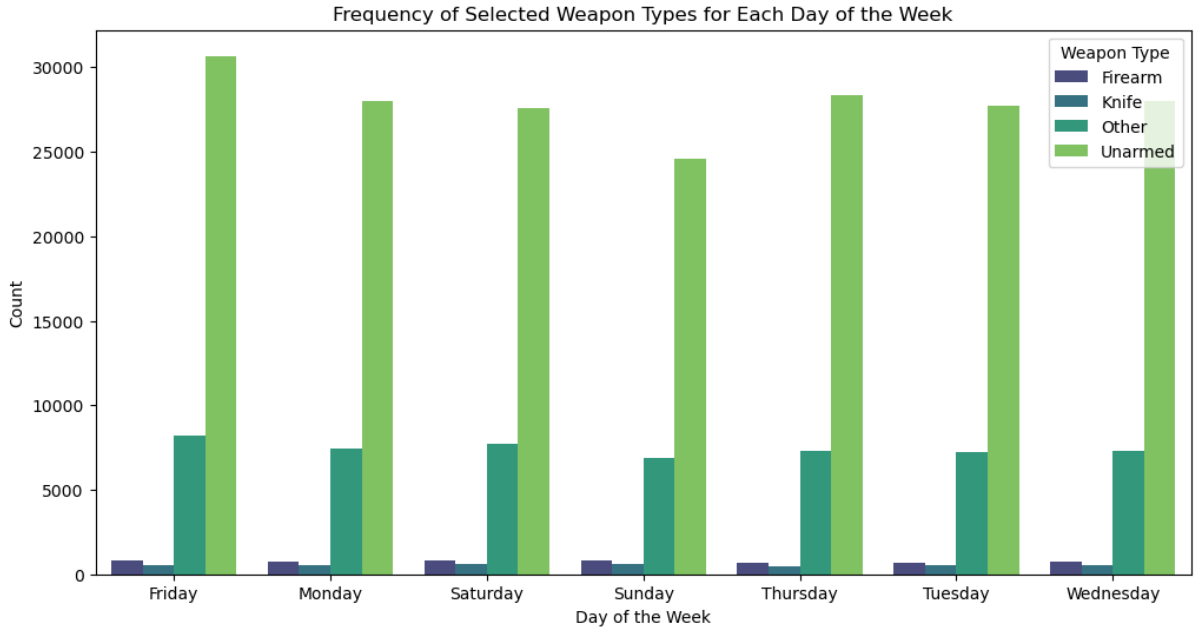
**Graph for Frequency of Incidents by Type:**




```
Incident Type with the Highest Frequency: VAL
Incident Type with the Lowest Frequency: GAMBLING OFFENSE
```

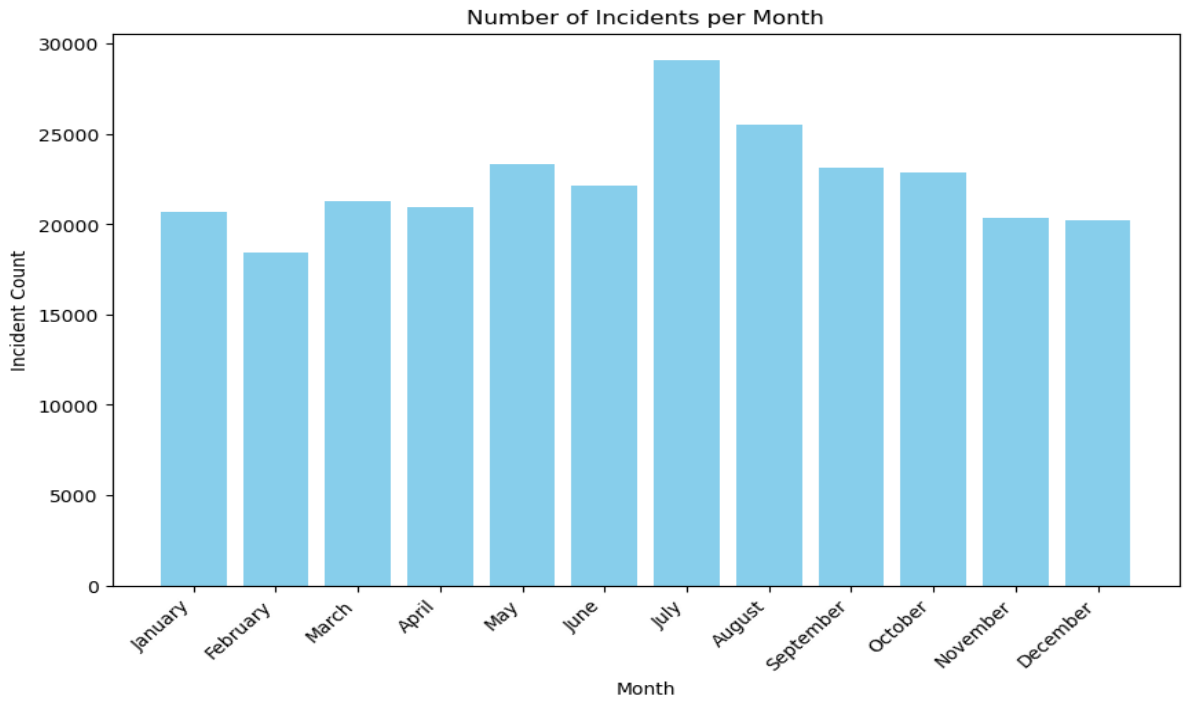
The incident type with the highest frequency, "VAL" (which stands for Validation/Field Interview), could indicate that these incidents are relatively common and occur frequently in the reported data. Validation/Field Interviews are interactions between law enforcement and individuals for various reasons, such as identity verification or investigation. On the other hand, the incident type with the lowest frequency, "GAMBLING OFFENSE," suggests that offenses related to gambling are less common in the dataset. Gambling offenses may be relatively rare compared to other types of incidents, reflecting a lower occurrence of such activities or law enforcement attention to gambling-related issues.

## Graph for Frequency of Weapon Types per Day of the Week:




The graph shows that most incidents reported each day involve individuals labelled as "Unarmed," implying that a substantial number of incidents occur without weapons. This could be due to non-violent situations like arguments or physical confrontations where weapons are not involved. Additionally, it might happen because information about weapons is sometimes missing or not accurately recorded in the dataset, making "Unarmed" a common category.

## Bar Chart for Number of Incidents per Month:



```
Incident Counts per Month:
         Month  Incident Count
0      January           20697
1     February           18461
2        March           21295
3        April           20930
4          May           23294
5         June           22132
6         July           29113
7       August           25525
8    September           23159
9      October           22836
10    November           20379
11    December           20235

Month with the highest number of incidents: July
Month with the lowest number of incidents: February
```

The output displays the number of incidents per month, revealing variations in crime rates throughout the year. July has the highest number of incidents, while February has the lowest. This pattern could be influenced by various factors, such as seasonal variations, weather conditions, or cultural events. For example, July might see increased outdoor activities and gatherings, potentially leading to more reported incidents, while February, being a winter month, may experience lower overall activity and, consequently, fewer incidents.

## Appendix C: code

Github Link for codes: https://github.com/Tiyasa-Saha/MTH522-Project-3/blob/main/Project%203.ipynb

## Contribution:

**Tiyasa Saha:** Worked on the Issues, Findings, Discussion, Methods, Code and Results sections. Also self-plotted the graphs to analyse the data using the various methods discussed in the report.

**Kanishka Patre:** Worked on the Issues, Findings, Methods, Code and Result sections. Plotted graphs and used various statistical models and tests to analyse the data.

**Srikanth Koncherry:** Worked on identifying issues, writing code for and the analysis models and producing the graphs for them.

**Gautam Marathe:** Worked on initial analyses, cleaning data, analysing, and looking for different models to fit the data on, testing various fits for their errors, testing non-linear models on the data to describe trends between predictors.