Analysis of Boston Crime Incident Data: Exploring Crime Patterns and Trends

Addressed To: Boston Police Department

Co Authored By: Parag Khandelwal (02120540)
Akshit Bagga (02130081)

ISSUES:

This data, sourced from the Boston Police Department (Analyze Boston), covers the Crime Incidents reported from the year 2015 to the present date (November 2023), and includes details and parameters like incident number, offense code, offense code group, offense description, district code, reporting area, shooting status, date on which the crime occurred, year of occurrence, month of occurrence, day of occurrence, hour of the day when it occurred, universal crime reporting category classification, street on which the crime occurred and details of the longitude and latitude of the place where it occurred. Additionally, we have also referred to offense code data, which describes all the codes and other sources to know more details about the various neighborhoods in Boston.

The questions which we aim to address are as follows:

1. What crimes were more often reported to the police and frequency of these crimes in the last few years?
2. Which crime is more likely to be reported based on an hour, month and day-based analysis?
3. When and where do most shootings occur and what type of offenses are more likely to involve shootings?
4. Which street is more likely to see a crime happening and what type of crime is more common over there?
5. Considering the Uniform Crime Reporting standards, which crimes are committed as per the UCR levels?
6. Which crimes do police spend most of their time on and what percentage does that constitute?

FINDINGS:

From the data about the crime incident report and offense codes which were registered from the year 2015, we were able to conclude the following:

1. The maximum number of calls, i.e., 53466 were registered for the crime code 3115 (inspecting/reporting a person), which was followed by 3831 (leaving property after causing damage from a motor vehicle) and then by 3006 (reporting a sick person).
2. It is more likely that more reports are received on Friday in the month of September for the crime code 3115 (inspecting/reporting a person). Considering all year round, crime code 3410 (towing for a motor vehicle) is more likely to be reported at 8 AM on a Saturday.
3. Most shooting related crimes take place on Washington St., with a notable concentration of them occurring in the month of June, particularly more on Saturday midnights. The predominant offense associated with these shootings is aggravated assault.
4. Almost 40% of the crimes from the year 2015 until now were reported in Washington St., where larceny alone contributed to 16% of all crimes.
5. Part 1 (severe offense) crimes comprise nearly 20% of the reported crimes. Part 2 (less severe) crimes constitute of 30% of the reported crimes, while Part 3 (medical, towing related incidents) comprise of the remaining 50%.
6. Starting from the year 2015, 3115 (inspecting/reporting a person) constituted 7.37% of all the reports. Only 10 offense codes constituted 40.7% of all the reports and 38 offense codes out of 576 codes contributed to 80.37% of 725,013 crime incidents registered in the city of Boston.

DISCUSSIONS:

1. Starting from the year 2015 and until November 2023, offense code 3115 (inspecting/reporting a person) is the code for which 53466 people called the police. This was followed by the offense code 3831, where a motor vehicle damages property, and 39708 reports were registered with this offense code. 33819 people called the first responders to report a case under the offense code 3006 (to report a case which requires medical/injury/sick assistance). This was followed by people calling in for code 1402 (reporting vandalism) and then by 3114 (investigating a property). Crimes like manslaughter by a train (122) or robbery using knife (315) have only one case reported from the year 2015 until November 2023.
2. The highest number of calls to report under the offense code 3115 (investigating/reporting a person) were received on the Friday in the month of September. This constituted 859 calls in the last eight years with this particular combination. When analyzed in an hourly manner, 680 calls were addressed towards offense code 3410 (to report a missing vehicle) at 8 AM on a Saturday. To mention, this trend to report with offense code 3410 was more prevalent in the month of August.
3. A total of 5583 shooting-based crimes occurred on three streets, which are Washington St., Blue Hill Avenue and Columbia Rd. The fewest shooting-based crimes were reported on Menlo St. And Menton St, however there would be streets where no shooting would have occurred. The primary offenses linked to shootings were Aggravated Assault (413,423), Firearm violation and Homicides (111).
4. Continuing upon the trend of Washington St., Boylston St. contributed to 20% of all the crimes, while Tremont St., Centre St. and Commonwealth Avenue each represented nearly 13% of the incidents. Notable occurrences over here were under the categories of motor vehicle accident response, drug violation and medical assistance.
5. Considering the crimes with the Uniform Crime Reporting (UCR) categories, larceny-based crimes were the most common ones reported under Part 1 (severe offense) crimes. This was followed by drug violation crimes, which were reported under the Part 2 UCR categories and then majorly by motor vehicle accident response related crimes categorized under the Part 3 UCR categories. Washington St., Boylston St. and Blue Hill Avenue were some focal points of incidents across all three UCR categories. District D4 had the highest Part 1 crimes, while District B2 accounted for all Part 2 and Part 3 offenses.
6. The offense code 3115 (investigating/reporting a person) constituted 7.37%, when considering all the 725,013 crime incidents starting from the year 2015 until November 2023. Considering the cumulative percentages, this was followed by offense code 3831 (property damage by a motor vehicle) adding up to 12.85% of all the crimes and then by 3006 (reporting a sick person) contributing up to 17.51%. The top ten offense codes (3115, 3831, 3006, 1402, 3114, 340, 3301, 3201, 802, 613) added to 41.7% of all the reports. Out of all the 576 offense codes under which the Boston Police Department registers complaints, only thirty-eight offense codes contribute to 80.37% of all the reports, starting from the year 2015.

APPENDIX A:

We began with downloading the dataset from the Boston Analyze website, where this data was listed as Crime Incident Report (2015-Present) and listed all the crimes which were reported from the year 2015 until now. We initiated the analysis by loading the crime dataset from an Excel file, meticulously addressing data integrity concerns such as missing values and potential errors. The dataset encompassed crucial columns, including "SHOOTING," "STREET," "OFFENSE_CODE_GROUP," and temporal indicators like "OCCURRED_ON_DATE," "MONTH," "DAY_OF_WEEK," and "HOUR."

All these parameters provide the below mentioned details,

- Incident Number: Lists down the incident number for BPD's reference
- Offense Codes: Lists down the offense code number under which the case was reported
- Offense Category: Provides a categorization as per the offense code
- Offense Description: Provides a description of the offense being committed
- District: Lists down the district code
- Reporting Area: Lists the reporting area, block number with which the response is associated
- Shooting: Tells if shooting was involved or not
- Occurred on Date: Lists the date, time when crime occurred
- UCR Part Classification: Provides severity level details
- Street: Street on which it occurred
- Latitude: Lists down exact latitude of the crime
- Longitude: Lists down exact longitude of the crime

Starting the analysis, we first checked the frequencies of the offense codes and got an initial idea of how often a crime occurs. Subsequently, we worked towards visualizing the same using different methodologies. Further, we also did a date, year, month, hour-based analysis considering the offense codes, which helped us to list which offenses were more often committed and the days where offenses or crimes had high prevalence. We identified streets with the highest occurrences of shootings, determining the top three and unveiling the associated predominant offense types. Further exploration delved into the temporal patterns of shootings, discerning the most common time for such incidents.

We explored the diversity of crimes on streets, identifying the top five streets with the highest counts of unique offense types and rendering the findings visually through informative bar graphs. The methodology embraced a systematic and comprehensive approach, unraveling patterns and relationships within the crime dataset. Conducting a Uniform Crime Reporting (UCR) analysis constituted the subsequent step, enabling an assessment of crime distribution across distinct UCR parts. This involved calculating percentages and visualizing the results using bar graphs. To ascertain the nuanced relationship between streets and UCR categories, we scrutinized the top streets associated with each UCR part, emphasizing geographical insights. The analysis extended to district-level examination, revealing districts with the highest occurrences for each UCR part. We did a Pareto Curve analysis to analyze and get an insight of the common crimes where BPD is majorly involved. Other than this, we also plotted the major crimes occurring in the city of Boston.

APPENDIX B:

Each year the Boston Police Department receives numerous calls from people reporting incidents and crimes. Out of the 725,013 crime incident reports which the BPD received, starting from the year 2015 until November 2023, we were able to see multiple parameters available such as incident number, which is the internal BPD reporting code, offense code, which is the numerical code describing an offense, offense code group, which is the categorization of various offense codes under a single category of crime, offense description which is a primary description of a offense, district in which the crime is reported, reporting area code where the crime was reported from, details if the crime involved shooting, date, month, hour in which it occurred, details about the street where the crime occurred, latitude and longitude of the place where the crime occurred. While for each year we had about 50,000+ entries, we combined all the entries for the last eight years in a custom dataset and used it for the purpose of analysis.

Firstly, we did an initial analysis to understand the crimes which were reported under various offense codes. Overall BPD reports crime incidents under 576 officially declared offense codes, but starting from the year 2015, until November 2023, only 254 offense codes were used to register incidents, which directly indicates that no crime took place which could pertain to the left over 322 offense codes. Upon checking the frequency of the crimes and the reports for offense codes where the highest number of incident reports are registered, we were able to see that offense code 3115, which is to investigate or report a person, to mention the exact numbers, 53466 incidents were reported under the offense code 3115. This was followed by the offense code 3831, which is to report property damage by motor vehicles, where 39708 incidents were reported with this particular offense code in the last eight years. Offense code 3006, stood at number three, where 33819 people reported a person as sick or injured in the last eight years. When we talk about the least reported codes, offense codes 122 (manslaughter by a train), 315 (robbery using a knife) had only one case each registered in the last eight years. However, it should be mentioned that there were offense codes which did not have even a single case reported in the premises of Boston city.

	count
3115	53466
3831	39708
3006	33819
1402	32171
3114	29385
3410	27484
3301	24565
3201	21326
802	20264
613	20201

Fig: Count of Offense Codes

Next up, we did a date, month, hour-based analysis, wherein we considered multiple factors and combinations to find out the time and dates when the greatest number of cases were reported to the Boston Police Department. Upon extracting the date from the date and time column of the overall dataset, we arranged the dates in the required format and then did a frequency check of the dates. On doing the process, we were able to see that among all these eight years a day in 2023 had the highest number of cases and on the other hand, also one of the days had the lowest number of cases. It was on March 17th,

2023, that 274 cases were reported to the Boston Police Department. On April 9th, 2023, only 40 cases were reported to the Boston Police Department. It could be safe to say that the latter day in April would have been the safest to step out in the last eight years.

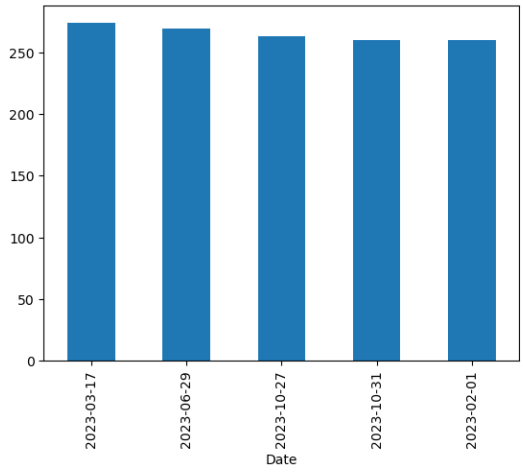

Fig: Date wise Analysis

On checking the overall data in crime specific manner, considering the offense codes and the days and hours, we were able to see that if it is the 8 AM in the morning on a Saturday, it is most likely that the police would get a call to report the offense code 3410, which is to report a towed motor vehicle. This trend can be justified with the number that 680 cases reported under such a combination. Upon checking this data in a month specific format, we were able to see that a total of 859 people called the BPD on the Friday of September to report under the offense code 3115, to investigate a person. Combining both these combinations, we were able to observe that on 8 AM of August and on a Saturday in all these years, people dialed BPD the most, to report a towed or a missing vehicle under the offense code 3410.

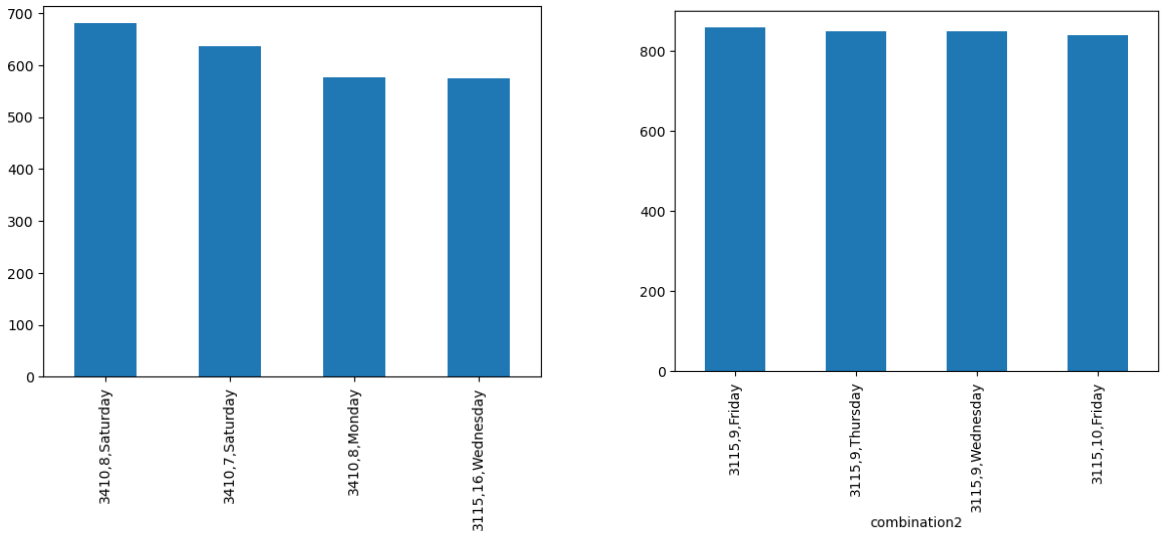

Fig: Graph for day, hour, offense code and day, month, offense code analysis

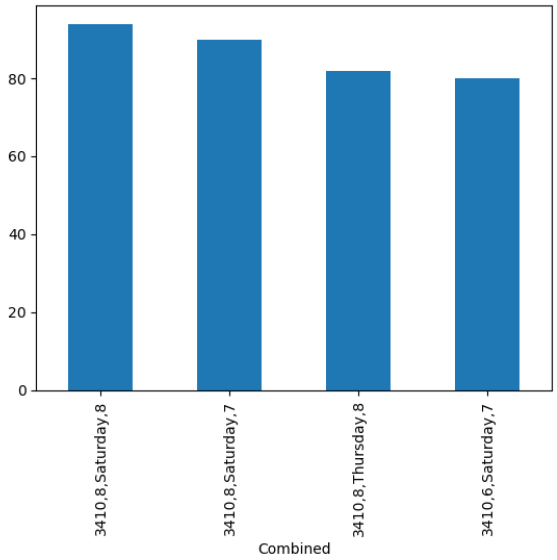

Fig: Graph for day, month, hour, offense code analysis

In the dataset, there is a parameter about shooting which does mentions if the incident involved use of weapons mainly guns and if shots were fired. We tried to relate this to the street parameter and check the streets which involved crimes related to shooting. As per our analysis, we were able to see that Washington St. Had more than 215 incidents which involved crimes related to shooting. This was followed by Blue Hill Avenue and then by Columbia Rd.

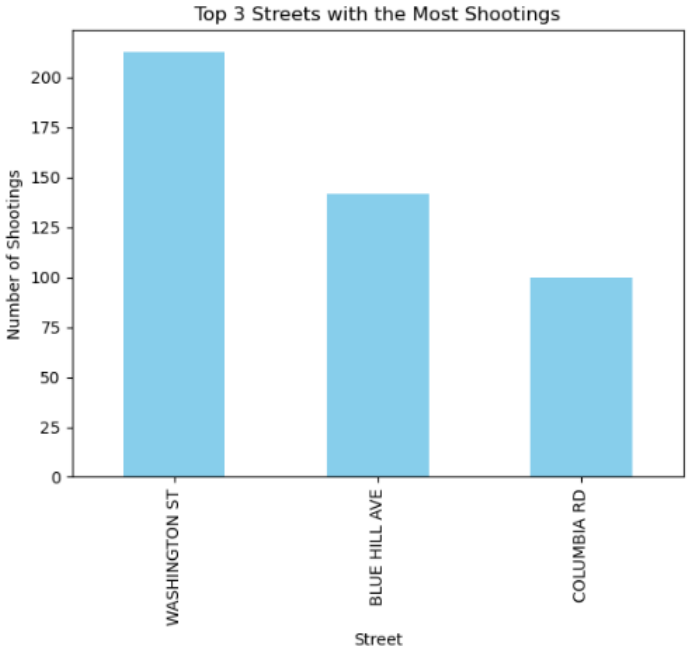

Fig: Graph for Top 3 streets with most shootings

Upon checking the data which involves shooting incidents, we were able to see that offense codes which are related to Aggravated Assault involve shooting related incidents. Upon checking the offense code table, we got to know that there are crimes under multiple categories which are classified as Aggravated

Assault. To mention in number, altogether 65 incidents were reported in the last eight years which involved shootings under Aggravated Assault category. This is followed by offenses registered under Firearm violation and then by Homicide related cases. As per our analysis most common time for shooting related crimes is in midnight of a Saturday in the month of June.

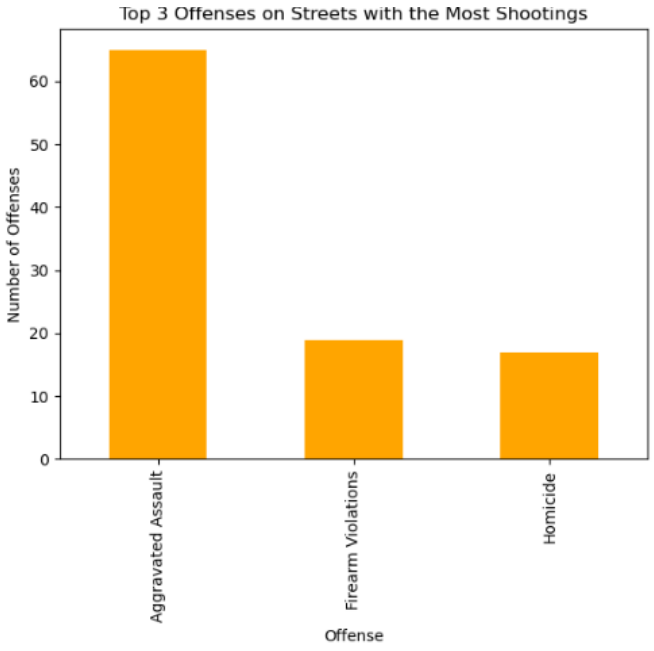

Fig: Graph for Top 3 offenses on streets based on shooting related crimes

Upon considering all the crimes which occur in the streets of Boston, we were able to see that Larceny related crimes are most likely to occur in all the offense categories and it is mostly that these crimes are reported in Washington St. As per the crime graph this is followed by Motor Vehicle Accident Response reports and then by Drug Violation. Upon seeing it in a top five street statistics manner, Washington St. is followed by Boylston St., Tremont St., Centre St., Commonwealth Avenue in that order.

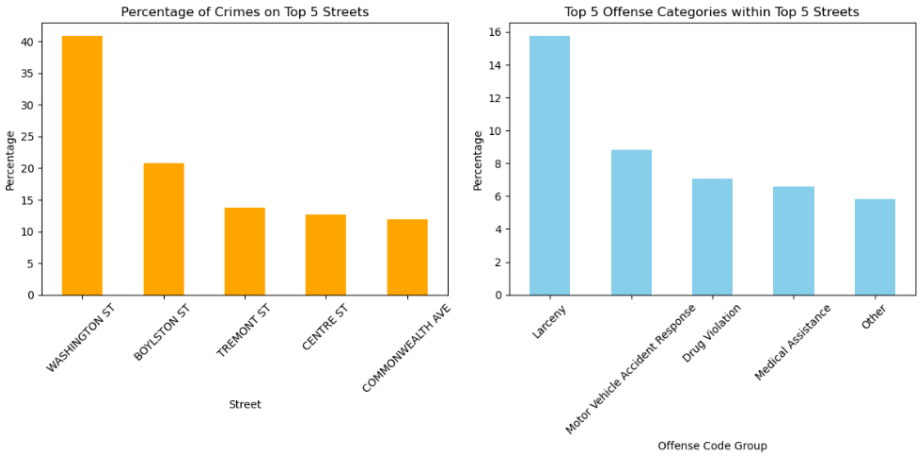

Fig: Graphs for Street and Crime Co-relation Analysis

All these crimes are also reported with a particular Universal Crime Reporting (UCR) code, which is used mainly to determine the severity of a crime. As per our understanding, if a crime has a high level of severity, it is most likely to be categorized as a Part 1 UCR level crime and so on, until the Part 3 UCR category. For the crimes which occurred from the year 2015, until November 2023, we were able to see that 49.84 % of the crimes were categorized in the Part 3 UCR code, 30.84% in the Part 2 UCR category and 19.27% of the overall crimes in the Part 1 UCR category. However, 0.39% of the crimes weren't classified in any of these categories.

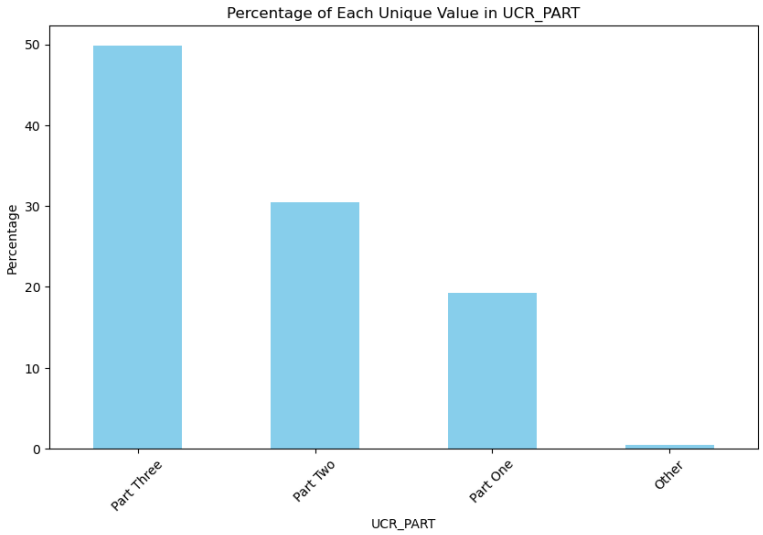

Fig: % based UCR crime plot

For Part 2 category UCR level crimes, it was mostly categorized in many offense codes, however mainly it was drug violation or a simple assault which made more sense in this category. For the Part 3 category UCR level crimes, it was mostly a call of Motor Vehicle Accident Response or a call for Medical Assistance or maybe to investigate a person on the streets of Boston or any premise in Boston.

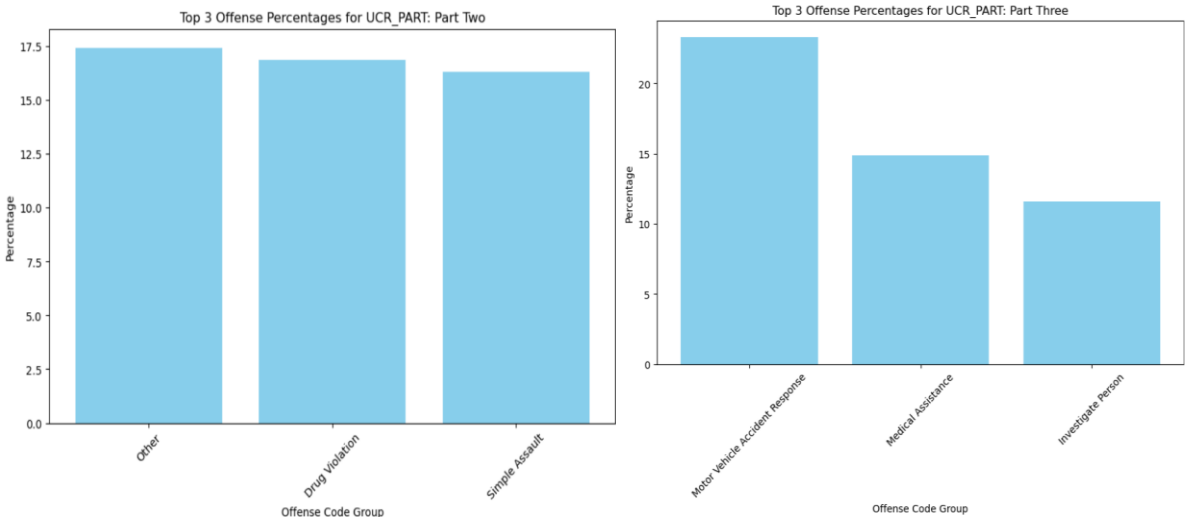

Fig: UCR Part 2 and Part 3 Crime wise Plot

When considering the Part 1 category UCR level classification, it was more evident that approximately 60% of the cases were for Larceny. It included both Larceny related cases (42.46%) and Larceny from Motor Vehicle related cases (17.46%). Aggravated Assault was also one of the most common Part 1 level crime, standing alone at 12.85%.

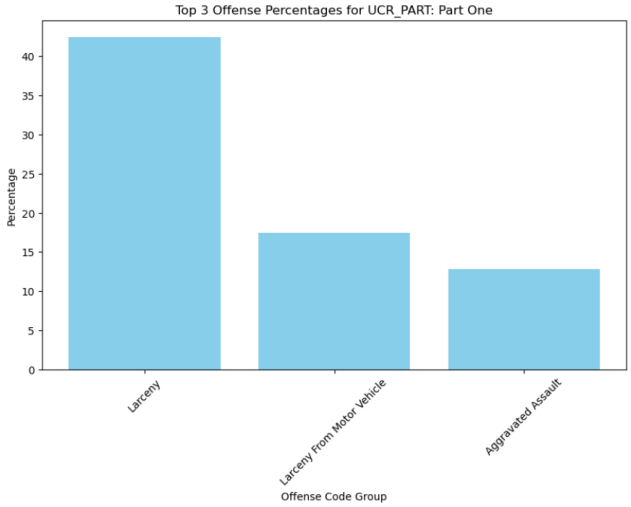

Fig: UCR Part 1 based crimes

In a street trend, it was more likely that a Part 2 or a Part 3 UCR crime occurred on Washington St. or Blue Hill Avenue. A Part 1 UCR-level crime occurred more likely on Washington St. or Boylston St. As per district related results, a Part 2 crime occurred more likely in B2 district, Part 3 in B2 district and a Part 1 related crime in the D4 district.

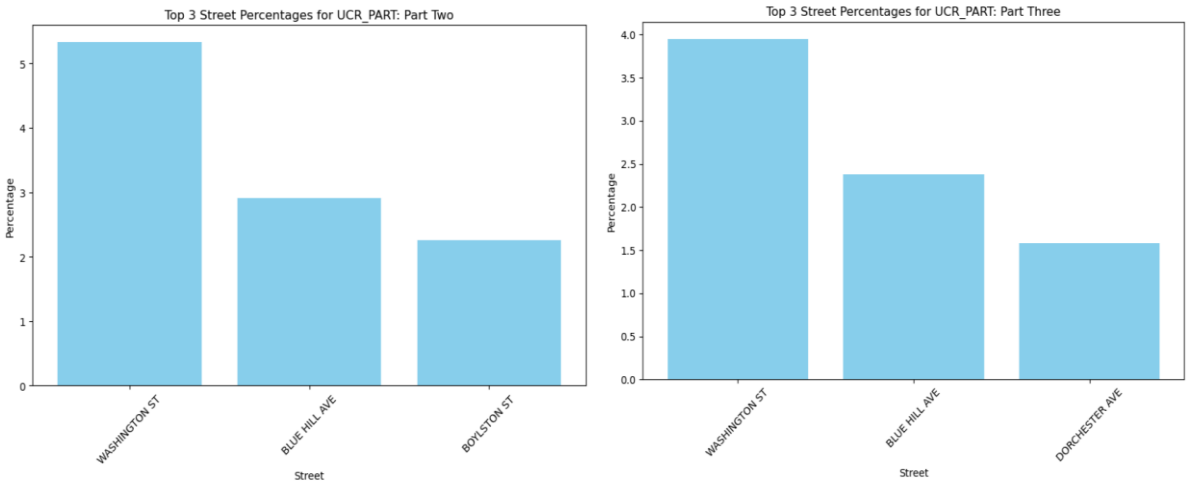

Fig: Part 2 and Part 3 UCR Based Categorization, considering street wise data

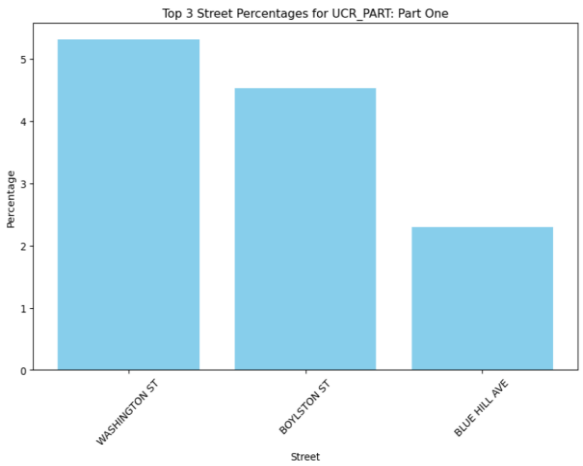

Fig: UCR Part 1 plot based on streets

We also did an overall crime analysis to see the crimes where the Boston Police Department might be spending most of their resources. To do this, we used the Pareto Curve method, which helped us to calculate the cumulative percentage of all the offense codes which were reported starting from the year 2015 until November 2023. As per the analysis, offense code 3115 (reporting/investigating a person) contributed to 7.37% of all the reported incidents. When added with the offense code 3831 (damage to property by a motor vehicle), this added to 12.85% and then offense code 3006 (reporting a sick/injured person) added to 17.51% of the reports. The top ten reported offense codes (3115, 3831, 3006, 1402, 3114, 3410, 3301, 3201, 802, 613) contributed to 41.7% of all the reported crimes in the city of Boston. Out of all the 576 offense codes under which the police department of Boston reports crime, 254 contributed to all crimes being reported in the last eight years. As per the Pareto curve, 80% of the cases are reported under only 38 offense codes. Equivalently, it would be safe to say that if a police officer in the city of Boston is rushing to attend a help call, 80% of the times it would be associated to any of these 38 offense codes.

	count	cumperc
3115	53466	7.374488
3831	39708	12.851356
3006	33819	17.515962
1402	32171	21.953262
3114	29385	26.006292
...	...	...
2609	1	99.999448
404	1	99.999586
2606	1	99.999724
432	1	99.999862
122	1	100.000000

Fig: Cumulative percentage result

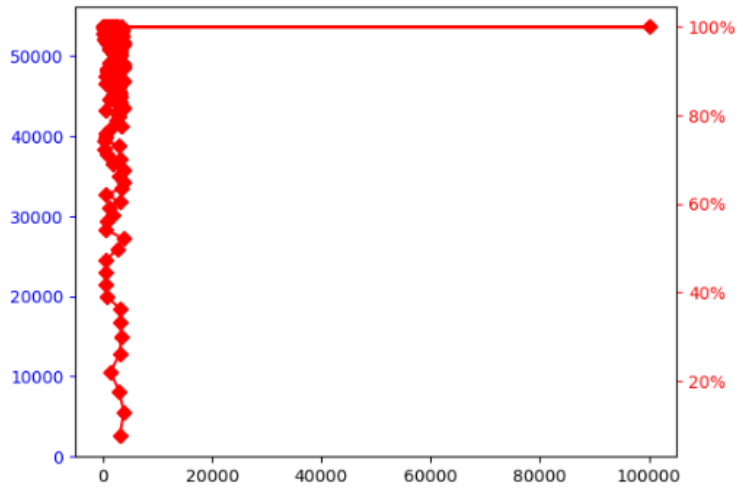

Fig: Crime plot on Pareto Curve

We also tried to make a plot of the top three offense codes (3115, 3831, 3006) on the city map of Boston to visualize where it is more likely that these crimes would occur. To do so, we extracted individual offense code related data and then used the subsequent latitude and longitude to make a plot on the map. The same can be referenced below.

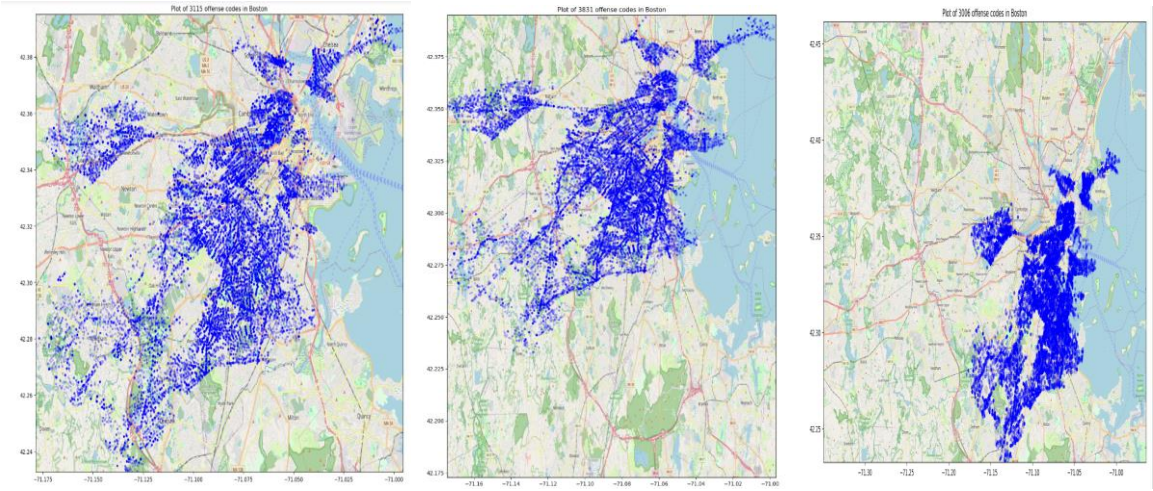

Fig: Boston Map Based plot for Offense codes 3115, 3831 and 3006 respectively

APPENDIX C:

We exclusively utilized the Python platform for all our coding tasks, primarily working within the Jupyter notebook or Anaconda Prompt environment. Our analysis heavily relied on several essential Python libraries, including pandas, numpy, matplotlib, sns, seaborn, sklearn, statsmodels.api, and scipy.stats.

It's important to note that although our initial code for data importation, extraction, and distribution remained consistent, we employed distinct code segments for various purposes. The individual code segments for all the analysis are listed below,

The codes in a sequential manner are:

- To import Data:

```
import pandas as pd
import numpy as np
file=pd.read_excel("D:\customdataset.xlsx")
file
```

Fig: To import the custom dataset

- To calculate frequency of the offense codes:

```
offcode=file["OFFENSE_CODE"]
```

```
frequency = {}

for i in offcode:
    if i in frequency:
        frequency[i] += 1
    else:
        frequency[i] = 1
print(frequency)

offensecount=frequency
```

```
offensecount.values()
```

```
df = pd.DataFrame({'count': offensecount.values()})
df.index = offensecount.keys()
```

```
df = df.sort_values(by='count', ascending=False)
```

Fig: To calculate frequencies and counts

- For Date, Hour, Month, Day, Offense code analysis:

```
data = {'offe': file['OFFENSE_CODE'],
        'month': file['MONTH'],
        'day':file['DAY_OF_WEEK'],
        'hour':file['HOUR'],
        'street':file['STREET']}

df = pd.DataFrame(data)
df["offe"] = df["offe"].astype(str)
```

```
data = {'offe': file['OFFENSE_CODE'],
        'month': file['MONTH'],
        'day':file['DAY_OF_WEEK'],
        'hour':file['HOUR'],
        'street':file['STREET']}

df = pd.DataFrame(data)

df["offe"] = df["offe"].astype(str)
df["month"] = df["month"].astype(str)
df["day"] = df["day"].astype(str)
df["hour"] = df["hour"].astype(str)
# Apply the custom function to create a new column 'Combined'
df['Combined'] = df[['offe', 'month', 'day', 'hour']].agg(','.join, axis=1)
```

```
df
```

```
az=df['Combined'].value_counts()
az
(az.head(4)).plot(kind='bar')
```

```
az.head()
```

```
df['combination2'] = df[['offe', 'month', 'day']].agg(','.join, axis=1)
```

```
count2=df['combination2'].value_counts()
count2
count2.head()
count2.head(4).plot(kind='bar')
```

```
count2.head(4)
```

```
df['combination3'] = df[['offe', 'hour', 'day']].agg(','.join, axis=1)
count3=df['combination3'].value_counts()
count3
count3.head(4).plot(kind='bar')
```

```
count3.head()
```

```
file
```

```
df2 = pd.DataFrame(data)
df2[['Date', 'Time']] = file['OCCURRED_ON_DATE'].str.split(' ', expand=True)
```

Fig: For Date, month, hour analysis

- For street-based analysis of the shootings:

```python
import pandas as pd
import matplotlib.pyplot as plt

def find_shooting_and_offense_stats(file_path):
    try:
        # Read the Excel file into a pandas DataFrame
        df = pd.read_excel(file_path)
    except FileNotFoundError:
        return f"File not found at path: {file_path}"
    except pd.errors.EmptyDataError:
        return f"The file at path {file_path} is empty."

    # Check if the required columns exist in the DataFrame
    required_columns = ["SHOOTING", "STREET", "OFFENSE_CODE_GROUP"]
    for col in required_columns:
        if col not in df.columns:
            return f"Column '{col}' not found in the dataset."

    # Filter out rows with zero values in the "SHOOTING" column
    non_zero_shootings_df = df[df["SHOOTING"] != 0]

    # Group by "STREET" and calculate the total number of shootings in each street
    street_shooting_counts = non_zero_shootings_df.groupby("STREET")["SHOOTING"].count()

    # Get the top 3 streets with the most and least shootings
    top3_most_shootings = street_shooting_counts.sort_values(ascending=False).head(3)
    top3_least_shootings = street_shooting_counts.sort_values(ascending=True).head(3)

    # Filter the DataFrame to include only the top 3 streets
    top3_most_df = non_zero_shootings_df[non_zero_shootings_df["STREET"].isin(top3_most_shootings.index)]
    top3_least_df = non_zero_shootings_df[non_zero_shootings_df["STREET"].isin(top3_least_shootings.index)]

    # Group by "OFFENSE_CODE_GROUP" and calculate the total number of offenses in each category
    top3_most_offense_counts = top3_most_df.groupby("OFFENSE_CODE_GROUP")["SHOOTING"].count()
    top3_least_offense_counts = top3_least_df.groupby("OFFENSE_CODE_GROUP")["SHOOTING"].count()

    # Get the top 3 offenses for both most and least shootings
    top3_most_offenses = top3_most_offense_counts.sort_values(ascending=False).head(3)
    top3_least_offenses = top3_least_offense_counts.sort_values(ascending=False).head(3)

    # Plot bar graphs
    plt.figure(figsize=(12, 6))

    # Bar graph for streets with the most shootings
    plt.subplot(1, 2, 1)
    top3_most_shootings.plot(kind='bar', color='skyblue')
    plt.title('Top 3 Streets with the Most Shootings')
    plt.xlabel('Street')
    plt.ylabel('Number of Shootings')

    # Bar graph for offenses on streets with the most shootings
    plt.subplot(1, 2, 2)
    top3_most_offenses.plot(kind='bar', color='orange')
    plt.title('Top 3 Offenses on Streets with the Most Shootings')
    plt.xlabel('Offense')
    plt.ylabel('Number of Offenses')

    plt.tight_layout()
    plt.show()

    return top3_most_shootings, top3_most_offenses, top3_least_shootings, top3_least_offenses

# Example usage:
excel_file_path = r"D:\General\UMass Dartmouth\Subjects\Fall 2023 - MTH 522 - Mathematical Statistics\Project 3\customdataset.xl
top3_most_shootings, top3_most_offenses, top3_least_shootings, top3_least_offenses = find_shooting_and_offense_stats(excel_file_
```

Fig: For Street based analysis

- To co-relate street and offenses altogether:

```python
import pandas as pd
import matplotlib.pyplot as plt

# Read the data from an Excel file
df = pd.read_excel(r'D:\General\UMass Dartmouth\Subjects\Fall 2023 - MTH 522 - Mathematical Statistics\Project 3\customdataset.x

# Remove rows where either 'OFFENSE_CODE_GROUP' or 'STREET' contains integers
# Also, drop rows with missing values in 'OFFENSE_CODE_GROUP' or 'STREET' columns
df_cleaned = df[
    ~df.applymap(lambda x: isinstance(x, (int, float)))['OFFENSE_CODE_GROUP'] &
    ~df.applymap(lambda x: isinstance(x, (int, float)))['STREET']
].dropna(subset=['OFFENSE_CODE_GROUP', 'STREET'])

# Group by street and count unique types of crimes
street_crime_counts = df_cleaned.groupby('STREET')['OFFENSE_CODE_GROUP'].nunique().sort_values(ascending=False)

# Get the top 5 streets
top5_streets = street_crime_counts.head(5).index

# Filter the DataFrame to include only the top 5 streets
df_top5_streets = df_cleaned[df_cleaned['STREET'].isin(top5_streets)]

# Calculate the percentage of each offense category within the top 5 streets
offense_percentage = df_top5_streets['OFFENSE_CODE_GROUP'].value_counts(normalize=True) * 100

# Get the top 5 offense categories
top5_offenses = offense_percentage.head(5)

# Print the result, including the highest to the lowest offenses for the top 5 streets
print(df_top5_streets.groupby('STREET')['OFFENSE_CODE_GROUP'].nunique().sort_values(ascending=False).to_frame().reset_index().to

# Print the top 5 offense categories
print("\nTop 5 Offense Categories:")
print(top5_offenses)

# Visualize the percentage of crimes on the top 5 streets
plt.figure(figsize=(12, 6))

# Bar graph for the percentage of crimes on the top 5 streets
plt.subplot(1, 2, 1)
street_percentage = df_top5_streets['STREET'].value_counts(normalize=True) * 100
street_percentage.plot(kind='bar', color='orange')
plt.title('Percentage of Crimes on Top 5 Streets')
plt.xlabel('Street')
plt.ylabel('Percentage')
plt.xticks(rotation=45)

# Bar graph for the top 5 offense categories
plt.subplot(1, 2, 2)
top5_offenses.plot(kind='bar', color='skyblue')
plt.title('Top 5 Offense Categories within Top 5 Streets')
plt.xlabel('Offense Code Group')
plt.ylabel('Percentage')
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```

Fig: For Street and Offenses co-relation

- To do analysis based on UCR categorization:

```
df = pd.read_excel(file_path)

# Specify the column name for which you want to find unique values
column_name = 'UCR_PART'

# Get unique values in the specified column
unique_values = df[column_name].unique()

# Calculate the percentage of each unique value
value_counts = df[column_name].value_counts(normalize=True) * 100

# Print the unique values and their corresponding percentages
print("Unique values and their percentages in the column '{}':".format(column_name))
print(value_counts)

# Visualize the percentages on a bar graph
plt.figure(figsize=(10, 6))
value_counts.plot(kind='bar', color='skyblue')
plt.title('Percentage of Each Unique Value in {}'.format(column_name))
plt.xlabel(column_name)
plt.ylabel('Percentage')
plt.xticks(rotation=45)
plt.show()
```

```
# IN THIS GRAPH THAT YOU NEED TO INLCUDE IS OF UCR PART 1 AND PAR 2


import pandas as pd
import matplotlib.pyplot as plt

# Path to the Excel file
file_path = r'D:\General\UMass Dartmouth\Subjects\Fall 2023 - MTH 522 - Mathematical Statistics\Project 3\customdataset.xlsx'

# Load the Excel file into a DataFrame
df = pd.read_excel(file_path)

# Specify the columns for analysis
ucr_part_column = 'UCR_PART'
offense_code_group_column = 'OFFENSE_CODE_GROUP'

# Get unique values in the 'UCR_PART' column
unique_values = df[ucr_part_column].unique()

# Iterate through unique values in the 'UCR_PART' column
for ucr_part_value in unique_values:
    # Filter the DataFrame for the current 'UCR_PART' value
    ucr_part_df = df[df[ucr_part_column] == ucr_part_value]

    # Calculate the percentage of each offense within the current 'UCR_PART' value
    offense_percentage = ucr_part_df[offense_code_group_column].value_counts(normalize=True) * 100

    # Get the top 3 offenses
    top3_offenses = offense_percentage.head(3)

    # Print the results
    print("UCR_PART: {}".format(ucr_part_value))
    print("Top 3 offense percentages:")
    print(top3_offenses)
    print("\n")

    # Visualize the top 3 offenses using a bar graph
    plt.figure(figsize=(10, 6))
    plt.bar(top3_offenses.index, top3_offenses, color='skyblue')
    plt.title('Top 3 Offense Percentages for UCR_PART: {}'.format(ucr_part_value))
    plt.xlabel('Offense Code Group')
    plt.ylabel('Percentage')
    plt.xticks(rotation=45)
    plt.show()
```

```
ucr_part_column = 'UCR_PART'
street_column = 'STREET'

# Get unique values in the 'UCR_PART' column
unique_values = df[ucr_part_column].unique()

# Iterate through unique values in the 'UCR_PART' column
for ucr_part_value in unique_values:
    # Filter the DataFrame for the current 'UCR_PART' value
    ucr_part_df = df[df[ucr_part_column] == ucr_part_value]

    # Calculate the percentage of each street within the current 'UCR_PART' value
    street_percentage = ucr_part_df[street_column].value_counts(normalize=True) * 100

    # Get the top 3 streets
    top3_streets = street_percentage.head(3)

    # Print the results
    print("UCR_PART: {}".format(ucr_part_value))
    print("Top 3 street percentages:")
    print(top3_streets)
    print("\n")

    # Visualize the top 3 streets using a bar graph
    plt.figure(figsize=(10, 6))
    plt.bar(top3_streets.index, top3_streets, color='skyblue')
    plt.title('Top 3 Street Percentages for UCR_PART: {}'.format(ucr_part_value))
    plt.xlabel('Street')
    plt.ylabel('Percentage')
    plt.xticks(rotation=45)
    plt.show()
```

```
import pandas as pd

# Define the path to the Excel file
file_path = r'D:\General\UMass Dartmouth\Subjects\Fall 2023 - MTH 522 - Mathematical Statistics\Project 3\customdataset.xlsx'

# Read the Excel file into a DataFrame
df = pd.read_excel(file_path)

# Specify the target column for unique values
target_column = 'UCR_PART'

# Extract unique values from the specified column
unique_values = df[target_column].unique()

# Display unique values in the chosen column
print(f"Unique values in the column '{target_column}':")
print(unique_values)

# Specify the column for finding the district with the highest occurrence
district_column = 'DISTRICT'

# Iterate over each unique value in the "UCR_PART" column
for part_value in unique_values:
    # Filter the DataFrame for the current UCR_PART value
    part_df = df[df[target_column] == part_value]

    # Check if the DataFrame is not empty before finding the mode
    if not part_df.empty:
        # Find the district with the highest occurrence for the current UCR_PART value
        most_common_district = part_df[district_column].mode().iloc[0]

        # Display the result for the current UCR_PART value
        print(f"\nThe district with the highest occurrence in the column '{district_column}' for UCR_PART '{part_value}': {most_
    else:
        print(f"\nNo data available for UCR_PART '{part_value}' in the DataFrame.")
```

Fig: For UCR based analysis

- To plot a Pareto curve:

```
import matplotlib.pyplot as plt
from matplotlib.ticker import PercentFormatter
newdf=df.head(25)
fig, ax = plt.subplots()
ax.bar(newdf.index,newdf['count'],color=color1)
```

```
color1 = 'blue'
color2 = 'red'
fig, ax = plt.subplots()
ax.bar(newdf['cumperc'], newdf['count'], color=color1)
ax2 = ax.twinx()
ax2.plot(newdf['cumperc'], newdf['cumperc'], color=color2, marker="D")
ax2.yaxis.set_major_formatter(PercentFormatter())
ax.tick_params(axis='y', colors=color1)
ax2.tick_params(axis='y', colors=color2)
```

Fig: Code for Pareto Curve plot

- To plot data on Boston City Map:

```
import matplotlib.pyplot as plt
df=file
df.head()
a=df['Lat']
b=df['Long']
a,b
```

```
a=df['Long'].min()
b=df['Long'].max()
c=df['Lat'].min()
d=df['Lat'].max()
a,b,c,d
```

```
ruh_m = plt.imread('D:\map (9).png')
```

```
BBox = (a,b,c,d)
fig, ax = plt.subplots(figsize = (15,15))
ax.scatter(df.Long, df.Lat, zorder=1, alpha= 0.2, c='b', s=10)
ax.set_title('Plot of 3115 offense codes in Boston')
ax.set_xlim(BBox[0],BBox[1])
ax.set_ylim(BBox[2],BBox[3])
ax.imshow(ruh_m, zorder=0, extent = BBox, aspect= 'equal')
```

```
file
```

Fig: To plot data on Boston city map

CONTRIBUTIONS:

PARAG KHANDELWAL: In my analysis of a crime dataset, I initially identified the top three streets with the highest number of shootings, including "WASHINGTON ST," "BOYLSTON ST," and "BLUE HILL AVE," along with the most prevalent offenses in these areas. I then determined the most common time for shootings, finding that incidents were most frequent in June, on Saturdays, and at midnight. Further investigation into UCR categories revealed that "Part Three" crimes were predominant, with variations in top offenses across UCR parts. Examining streets associated with UCR parts, "WASHINGTON ST" consistently appeared prominently. Additionally, I explored district-level data, highlighting the districts with the highest occurrences for different UCR parts. Finally, I identified the top five streets with the most diverse range of crimes, such as "CENTRE ST" and "WASHINGTON ST," and visualized the findings through insightful bar graphs. Overall, the analysis provided a comprehensive understanding of the dataset's crime patterns, street occurrences, and UCR categories.

AKSHIT BAGGA: I worked over analyzing individual offense codes, where I generated an insight of how the offense codes prevail in the last eight years and their frequencies. Further, considering other parameters, I generated an analysis from the date, hour, month, and what days were prone to more crime and which ones saw lesser number of crimes. Other than this, I also worked towards the cumulative analysis of all the crime codes. With the cumulative percentages of all the offense codes, I worked with the Pareto curve to generate an analysis about the crimes which are more likely contributing to the work of Boston Police Department and how do other flare when compared. Also, considering the frequencies of some of the offense codes, I worked towards plotting those on the map of Boston City. Above all this, I worked on the report writing and different sections for the same.