# CRIME DATA ANALYSIS

## Project – I

Group 8

Yash Harale

Sakshi Aryal

Shruthi Kashetty

harale.y@northeastern.edu

aryal.sa@northeastern.edu

kashetty.s@northeastern.edu

# Introduction

In this project we will explore the domains of public safety and data analysis. A real-world dataset of crime rate from 2020 till now will be reviewed. Our primary goal is to properly clean and pre-process this data set so, that it can be effectively used for data analysis and research. Our compass will lead to through this dataset into a number of objectives. First, we will continue the series of vital actions for cleaning data so that our resulting dataset does not contain such mistakes and peculiarities. This comprises of managing data types, data duplication, managing outliers and missing numbers respectively.

# Objectives

Data Cleaning and Preparation:

Used data pre-processing tasks to clean the dataset and make it free from any defects and anomalies and includes operations like handling missing values, removing the duplicate values, controlling outliers and ensuring correct data types are used for the data set.

Exploratory Data Analysis (EDA):

To study the temporal characteristics of the crimes and spatial characteristics through monthly, seasonal or annual rates and geographical properties of the crimes. To get CORE-derivable insights that point to intervention solutions for managing crime rates, it is necessary to research the causes of crime rates.

Trend Analysis:

To compare statistics of crime incidence over the years in an attempt to explain variations in the extent of crime, determine areas of probable emergence of crime, and measure efficiency of policies. Discern broader forces of society which may explain high level of crime in various societies.

Predictive Modelling:

Employed methods of forecast analysis including the use of time series in an effort to forecast on the trends in crime. Explain how Increase or decrease in crimes or trends are affected by important events or policy reforms.

Advanced Analysis:

Investigate additional questions or hypotheses related to the dataset, such as seasonal trends in crime rates, correlation between crime occurrence and location, or the relationship between crime type and weapon used.

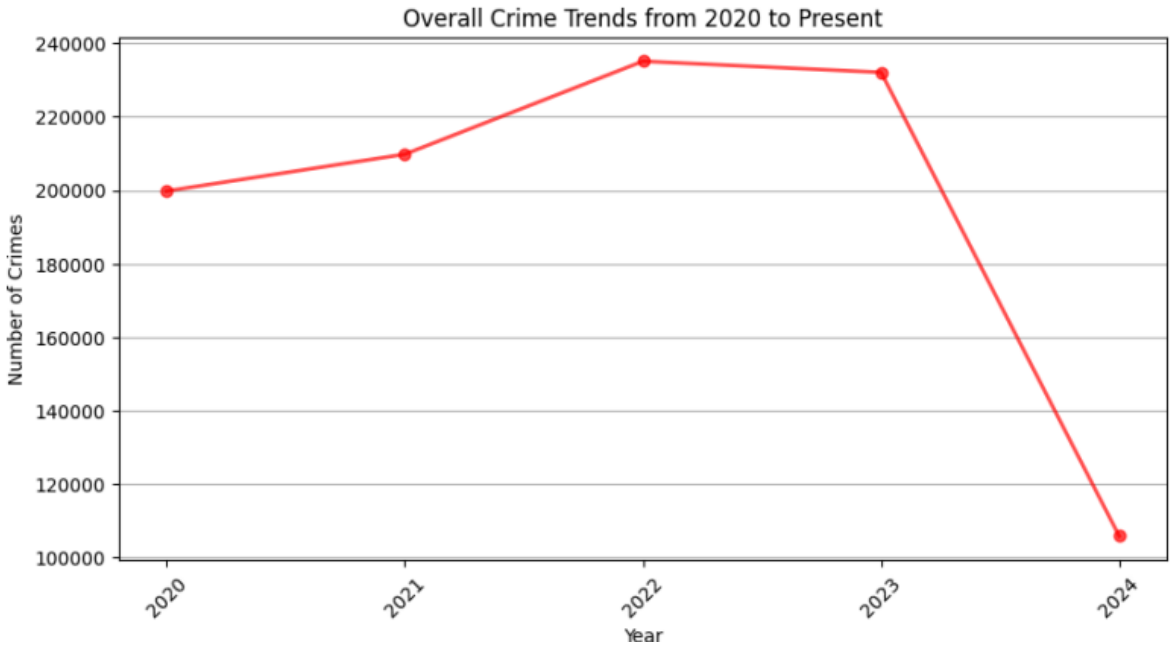
# Data Cleaning

Data cleaning is an important step in this project as a method for the analysis of crime records of the years 2020 to the present. The quality of the data and format in which it needs to be maintained is most critical in order that results turn out accurate and hit the mark. In this process, there is a way of dealing with missing data and outliers which may distort the findings of an analysis. These values are systematically dealt by replacing them through imputation or exclusion in order to keep the quality of the dataset high. At the same time, it is necessary to work with both types of values because if one

of them is much greater than the others, it can skew the trends and patterns observed. Data cleaning also includes issues with dates, handling of extra columns which can be of little or no use, and bringing consistency in the categories as well also is considered under data cleaning process. When executing these cleaning activities, the pathways are set for additional exploratory data analysis and accurate assumptions and findings about crime patterns and factors influencing such occurrences.

# Exploratory Data Analysis

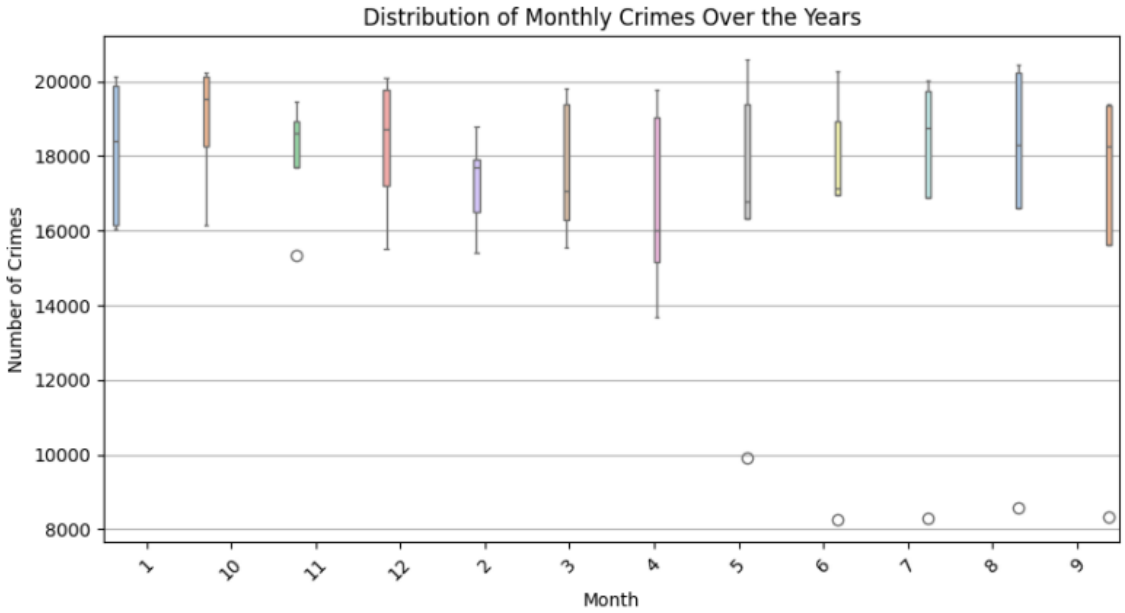
1. *Visualization of Crime trends from 2020 to the present year*



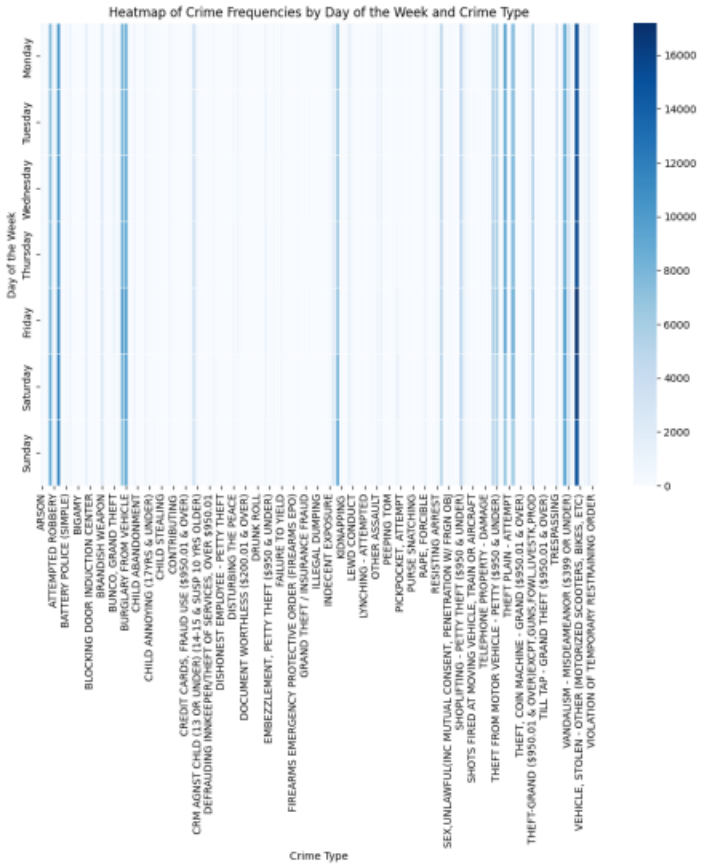

Overall crime patterns from 2020 to the present are depicted in this line graph. Following a period of relative stability, crime rates show a

noticeable decline, beginning in early 2024. This suggests that the number of offenses reported during that period decreased significantly.

2. *Monthly Crimes over the years*

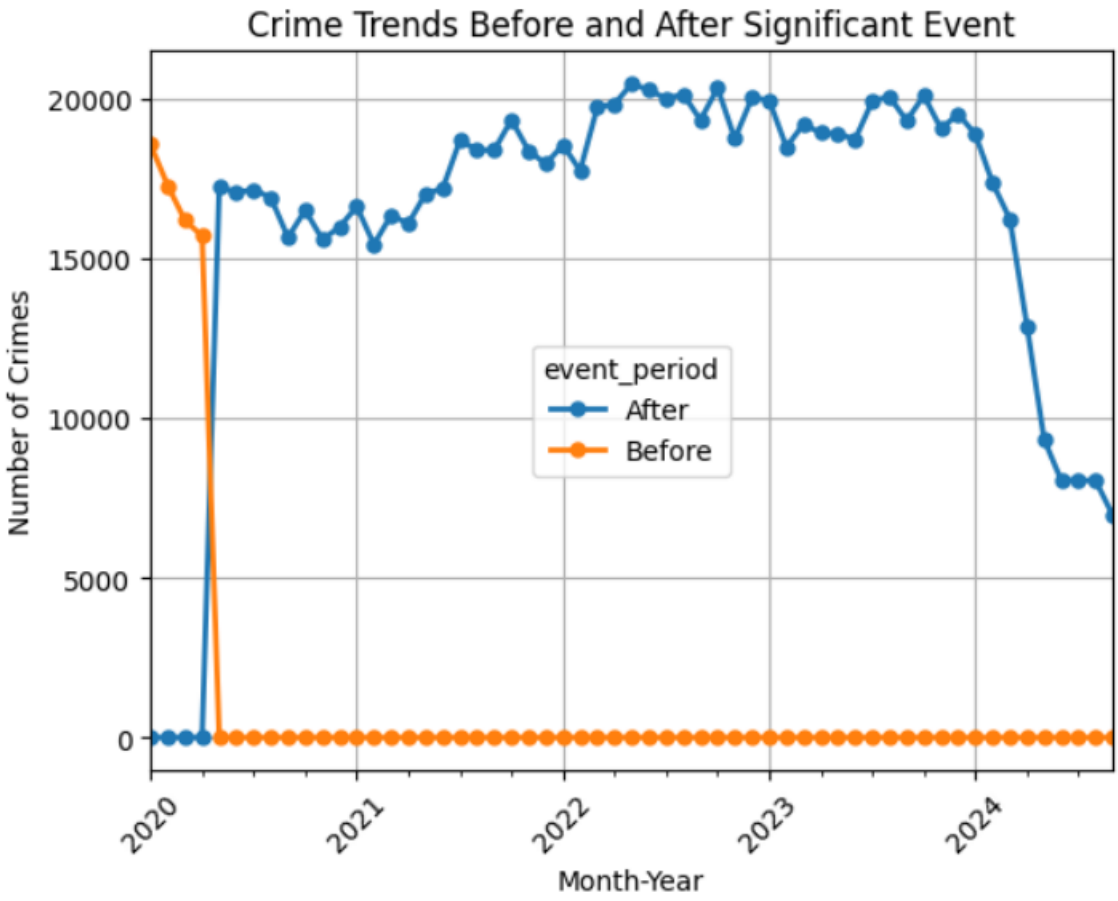



3. *Crime frequencies by Day of the Week*



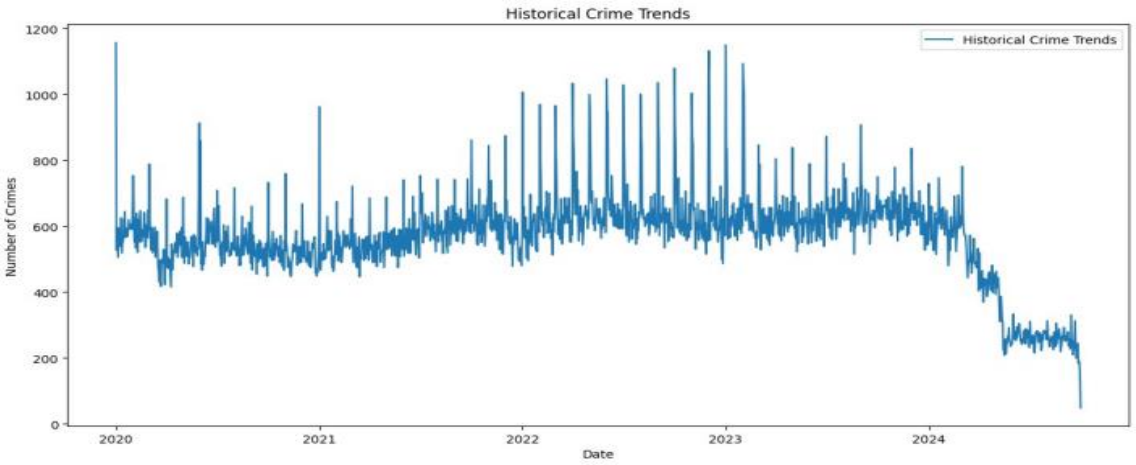

The heatmap makes it easy to review the days and types of crimes as shown by the colors on the figure. The cell color intensity can give an inference about how frequently which crimes are committed on which day of the week. For instance, the heatmap may reveal that "Theft from Motor Vehicle Petty" occurs frequently on the weekend, but other crimes probably, such as "Battery Police (Simple)" and "Disturbing the Peace," are almost equal any other day of the week. Moreover, when the total number of different types of crimes is analyzed to determine which crimes are more common than others at the moment. This visualization allows for the best sense-making regarding the distribution and fluctuations in crime incidents, which is useful for police and other people involved in urban planning.

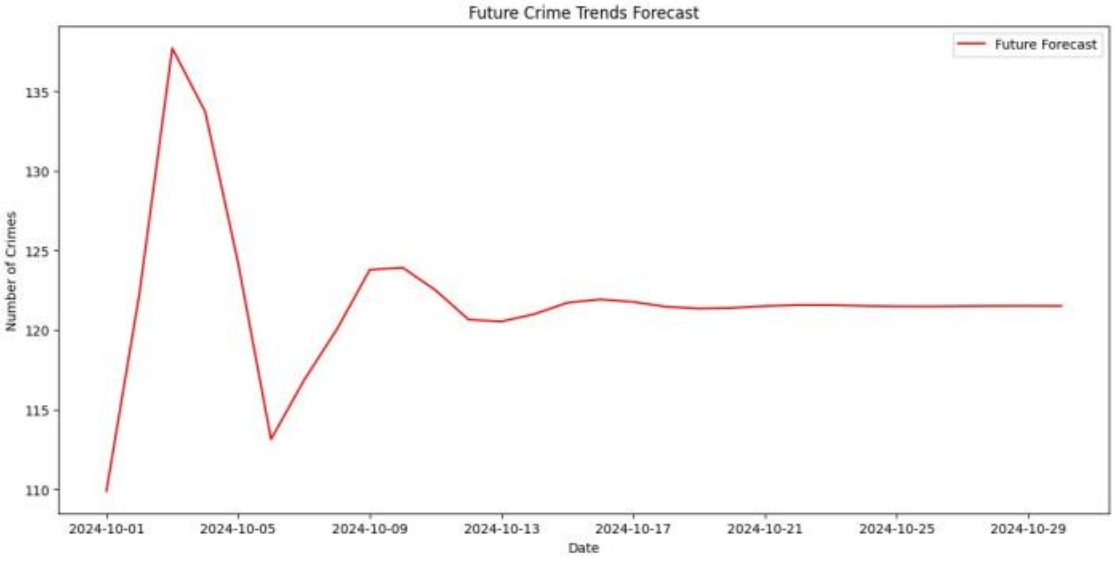
4. *Visualization representing crime trends before and after significant events*



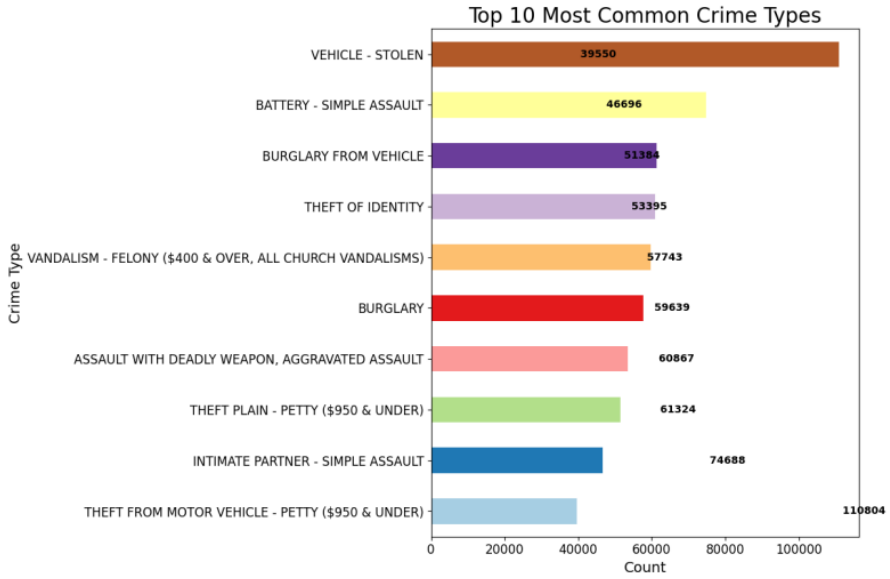

# Advanced Analysis

Used predictive modeling techniques (e.g., time series forecasting) to predict future crime trends.

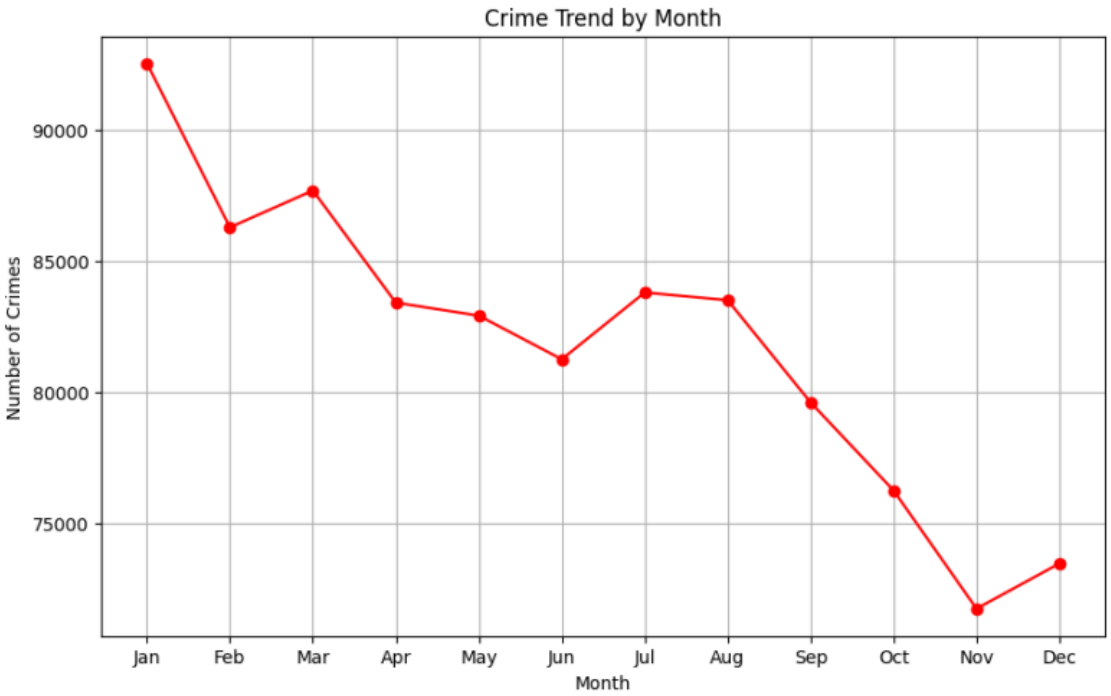


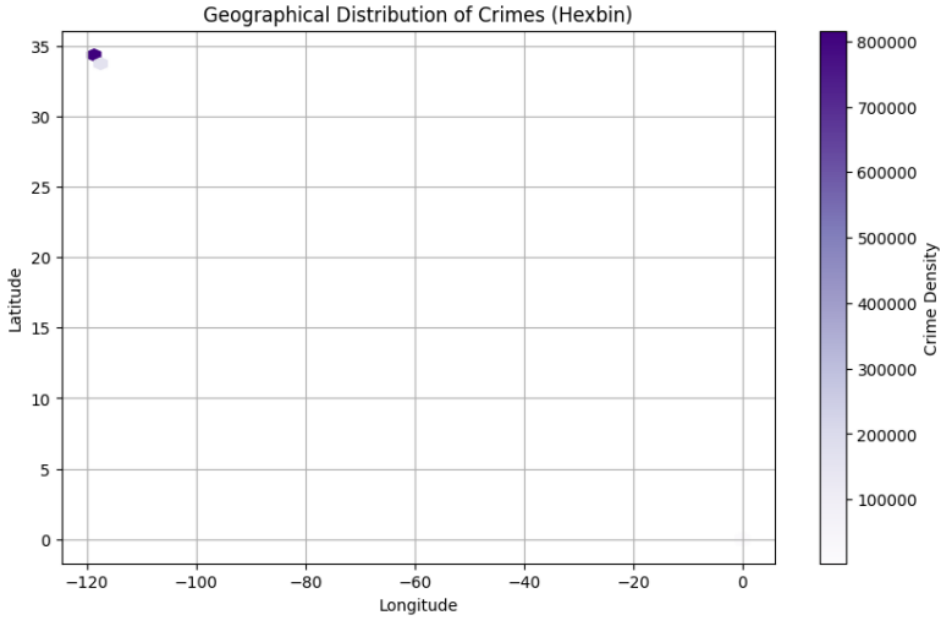
## *Common Types of Crimes*



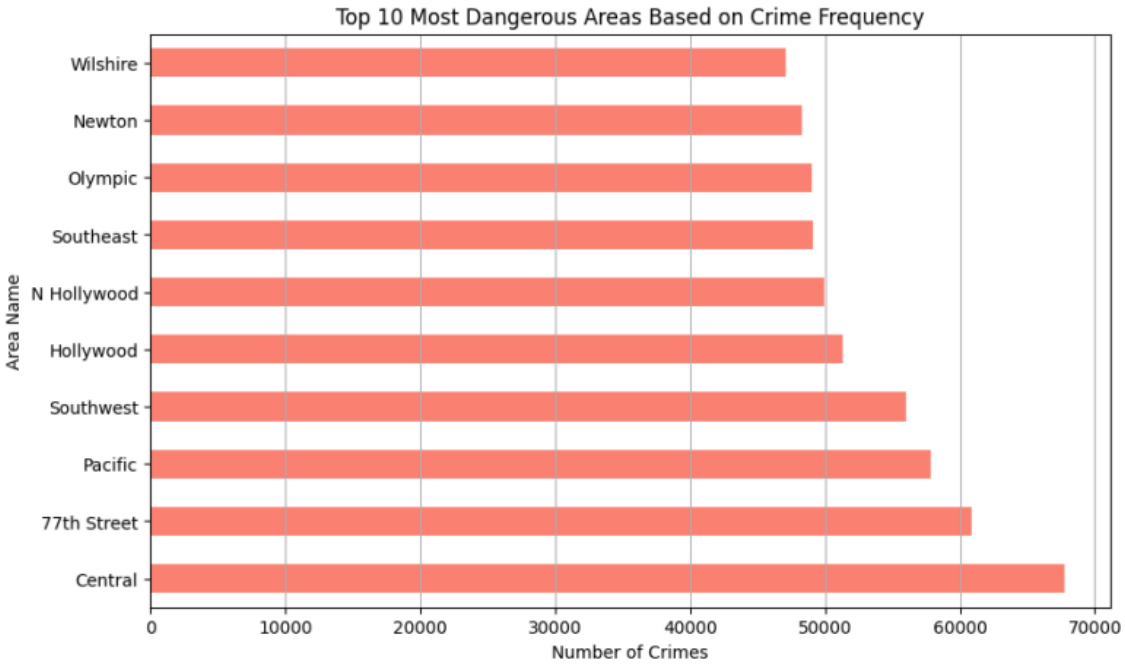

## *Crime Trends by Month*



## *Geographical Distribution of Crimes*




## *Dangerous Areas based on crime frequency*



# Summary of the Findings

***Most common crime:*** Vehicle Theft was the most common kind of the crime and it accounted for over 110,804 reported cases.

***Predictive Policing Potential:*** The application of time series forecasting shows the feasibility of predictive policing. A two way help is gained since from the crime analyses prediction of the crime trends in the future is helpful in giving an early hint towards crime areas and particular time in the day that is most affected.

***Geographical Insights:*** Geographical distribution analysis helped to determine crime hot spots. The layouts also showed such ten areas with the highest crime frequencies to be as follows.

***Temporal Patterns:***

Seasonal Trends: The results identified the existence of seasonality effect as the crime rates per month differed from each other.

Time of Day: Four different crimes were committed which gave the analysis of different types of crimes according to the different times of the day, such as Night, Morning, Afternoon, and Evening.

Day of Week: In the following a heatmap displayed the occurrence of specific crime types depending on days of the week, in order to identify potential patterns.