

# **Project 1 Report**

**TOPIC: CRIME DATA SET ANALYSIS**

**IE6400 19843 Foundations Data Analytics Engineering  
SEC 01 Fall 2023**

**Group Number 7**

Aishwariya Alagesan (002818973)

Arya Lokesh Gowda (002249418)

Keerthana Balswamy (002839068)

Deepana Dhakshinamurthy (002834708)

Suja Ganesh Murugan (002839290)

## **INTRODUCTION:**

This project focuses on examining real-world crime incidents in the City of Los Angeles, spanning from 2020 to the present. The report provides a comprehensive exploration of key elements, including data set acquisition, data refinement, an in-depth assessment of crime rates, forecasting future crime trends, and the utilization of visualizations to bolster the analysis.

## **DATA INSPECTION**

- **INITIAL DATA INSPECTION**

After loading the dataset, we performed an initial examination to grasp its format and attributes. The dataset encompasses information pertaining to crime particulars, including the location of the crime (comprising details such as city, latitude, and longitude), the weapons involved at the crime scene, the date of the incident, and victim information. We inspected few details such as column names, column data types and statistics report of the dataset.

- **DATA CLEANING**

During our data cleaning procedures, we identified and eliminated missing values from the dataset. Furthermore, we conducted the removal of several columns, specifically 'Crm Cd 2,' 'Crm Cd 3,' 'Crm Cd 4,' and 'Cross Street,' due to their high prevalence of null values, accounting for roughly 80% of the dataset. In situations where null values were encountered, we replaced them with '0' for integer columns and 'unknown' for string columns.

- **DATA TRANSFORMATION**

Transformed data types of certain columns as follows,

Date Rptd → datetime

DATE OCC → datetime

Premis Cd → Integer (int64)

Weapon Used Cd → Integer (int64)

- **OUTLIERS**

We tried observing outliers for the following columns crime rates against victim age (Fig 1) and Area (Fig 2).

Outlier was observed when we compare crime rates with victim age. (Age 120)

No Outliers were observed when we compare area with crime rates.

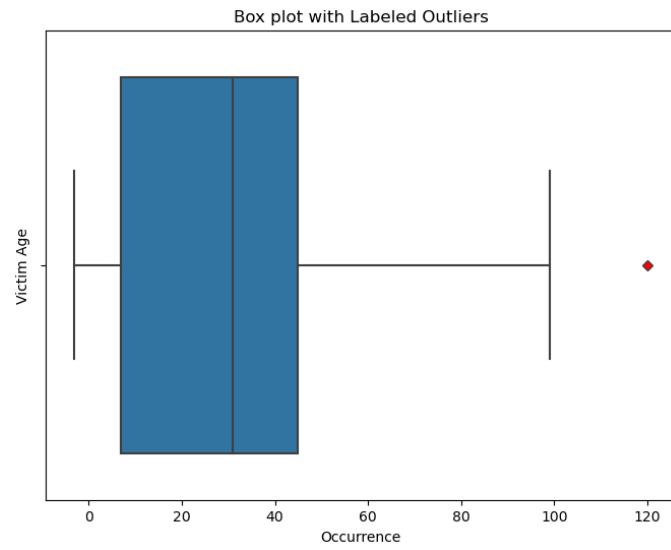


Fig (1)

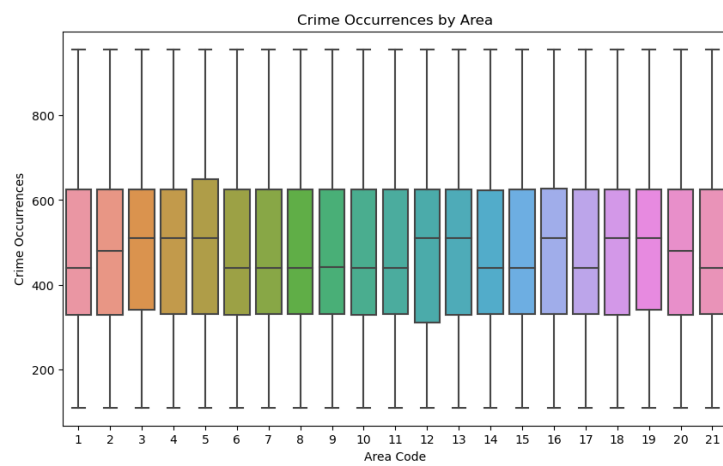
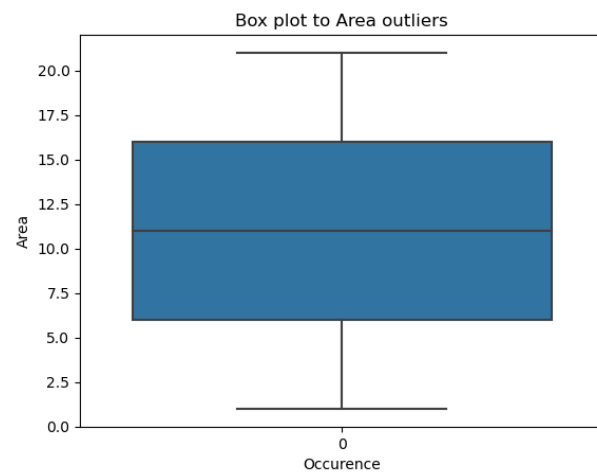


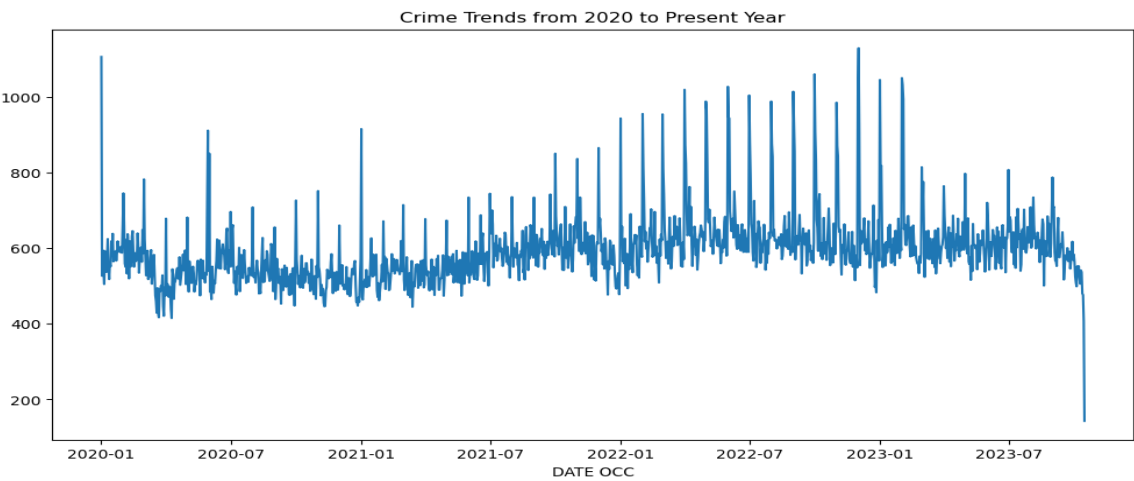
Fig (2)

- **STANDARDIZING NUMERICAL VALUES**

We applied a standardization process to the Latitude and Longitude columns, rescaling their values to fall within the range of 0 to 1.

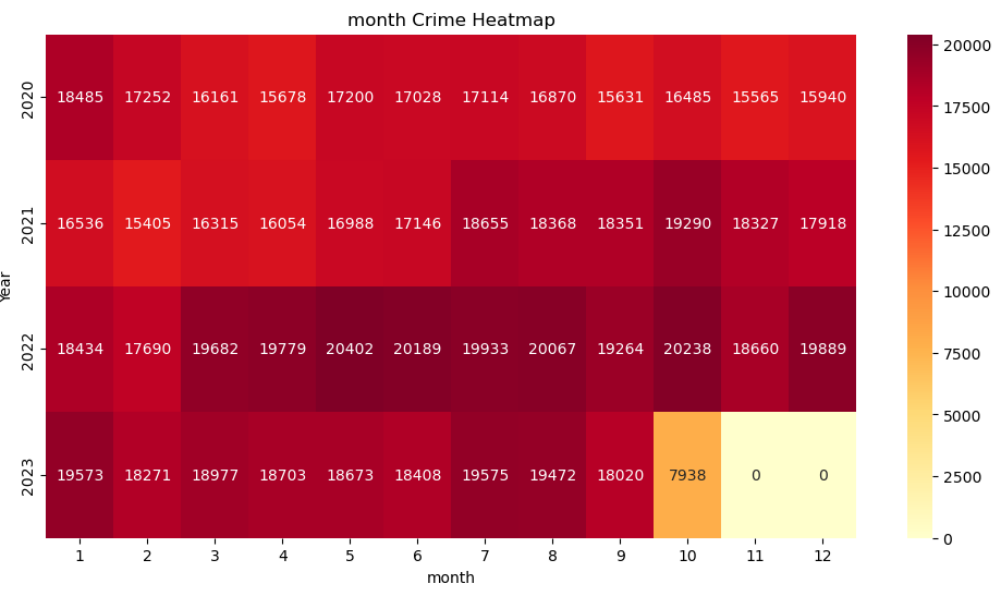
**EXPLORATORY DATA ANALYSIS:**

**CRIME TRENDS VISUALIZATION (2020 TO PRESENT)**



The line graph reveals that crime rates experienced their highest peaks in the initial months (January 2020) and during the period from May 2022 to Oct 2022. Subsequently, crime counts steadily decreased to below 200 by the end of the observation period, starting in Oct 2023.

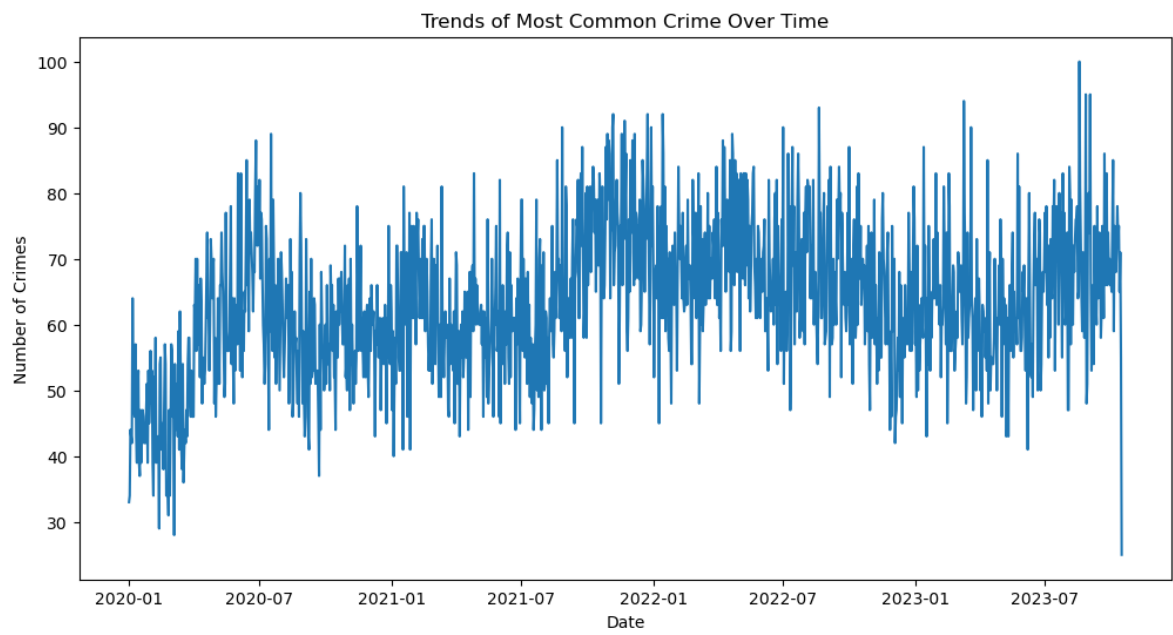
**SEASONAL PATTERNS IN CRIME RATE**



We employed a heatmap to visually represent the distribution of crime counts across monthly intervals, spanning from January to December, covering the years 2020 through 2023.

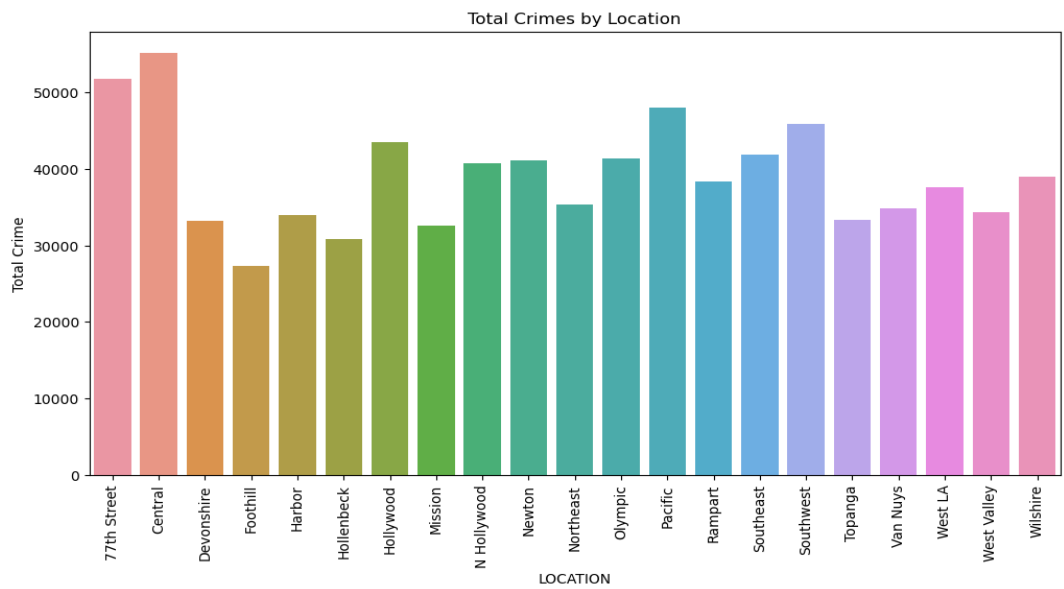
**MOST COMMON TYPE OF CRIMES OVER TIME**

The most common type of crime that we observed from dataset is “Vehicle – stolen”



After visualizing the data spanning from 2020 to the present time, we observed a significant increase in crime counts related to 'vehicle theft' after July 2023.

**NOTABLE DIFFERENCE IN CRIME RATES BETWEEN REGIONS**



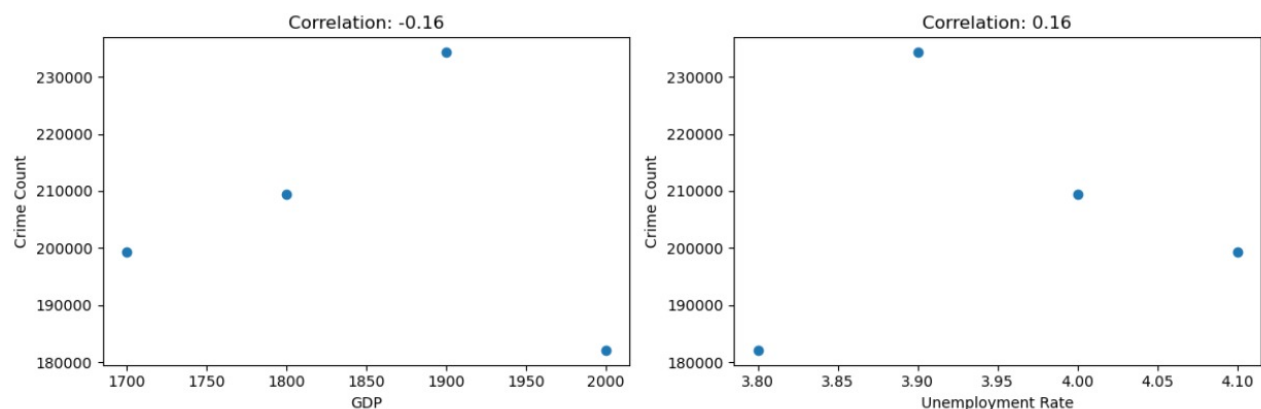
The highest crime rates were recorded in the 'Central' location, exceeding 50,000 incidents, while the lowest were documented in the 'Foothill' area, with fewer than 30,000 incidents.

## CORRELATIONS BETWEEN ECONOMIC FACTORS AND CRIME RATE

To assess the correlation between economic indicators and crime rates, we have consulted external data sources, encompassing unemployment rates and GDP statistics for the years 2020 through 2024.

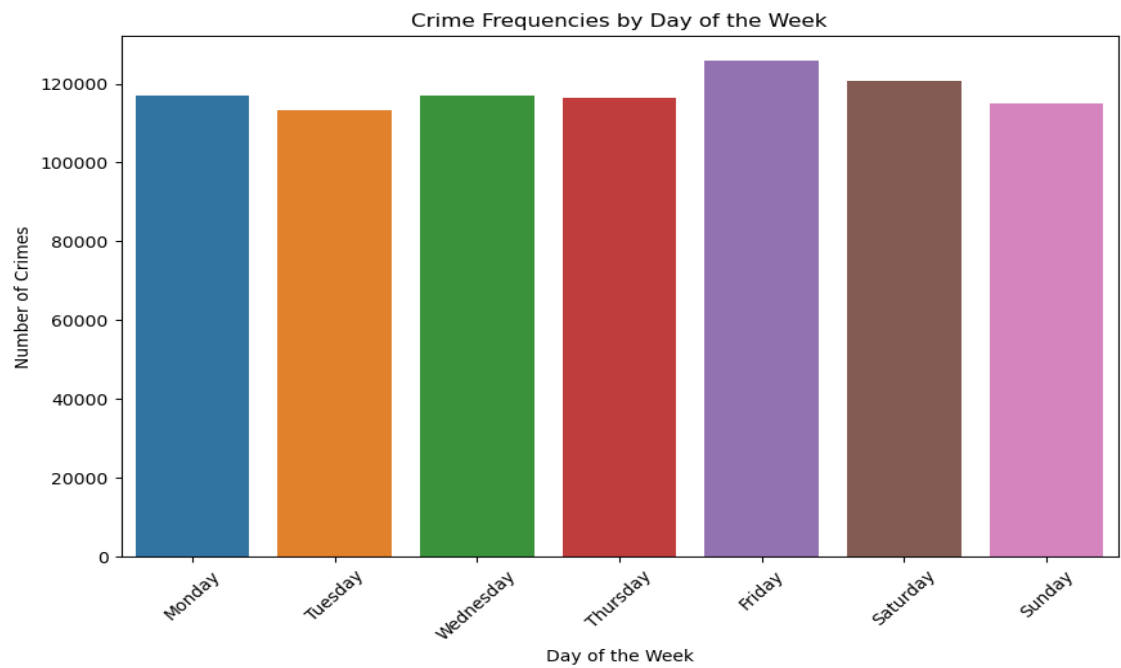
```
data = {  
    'Year': range(2020, 2024),  
    'GDP': [1700, 1800, 1900, 2000],  
    'Unemployment_Rate': [4.1, 4.0, 3.9, 3.8],  
}
```

- 1) A correlation of -0.16 between crime rate and GDP suggests a weak, negative relationship, but it is only a small piece of the puzzle when it comes to understanding the complex factors affecting crime rates.
- 2) The negative sign (-) of the correlation coefficient indicates an inverse relationship. In other words, as one variable (GDP) increases, the other variable (crime rate) tends to decrease, and vice versa. However, the correlation is weak, so other factors likely have a more substantial impact on crime rates.
- 3) A correlation coefficient of 0.16 indicates a weak positive correlation. This means that as the unemployment rate increases, the crime rate tends to increase slightly, but the relationship is not very strong.
- 4) The correlation coefficient of 0.16 suggests that unemployment rate alone is not a strong predictor of crime rates. While there is a slight tendency for crime rates to increase with higher unemployment, many other factors, such as economic conditions, social factors, and law enforcement policies, also influence crime rates.

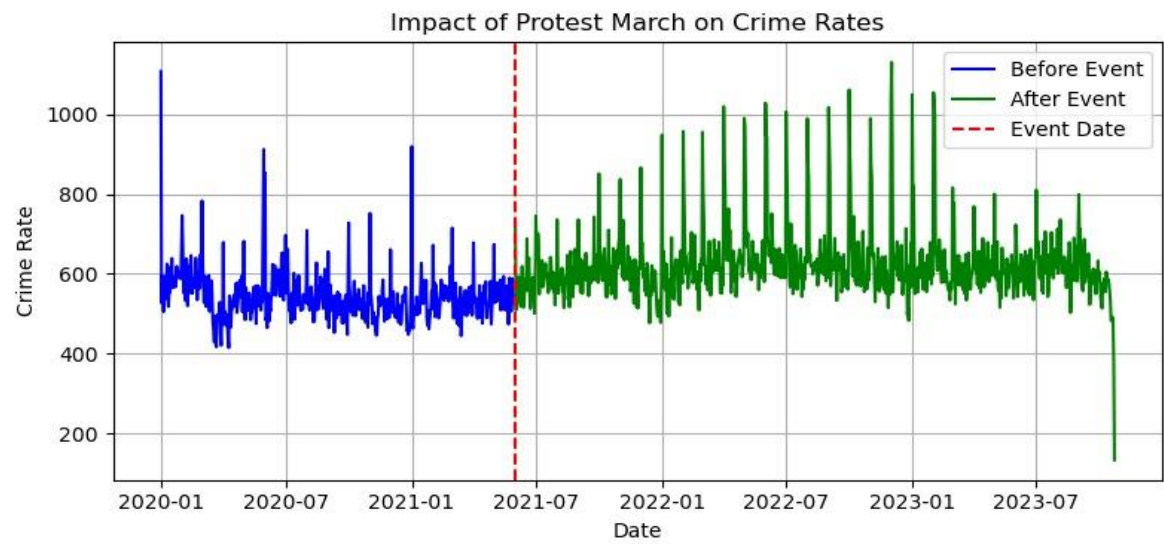


**RELATIONSHIP BETWEEN WEEK AND CRIME RATES**

The highest crime rates (>120000 incidents) were recorded on Friday as per the bar chart represented below.



**IMPACT OF MAJOR EVENTS ON CRIME RATES**



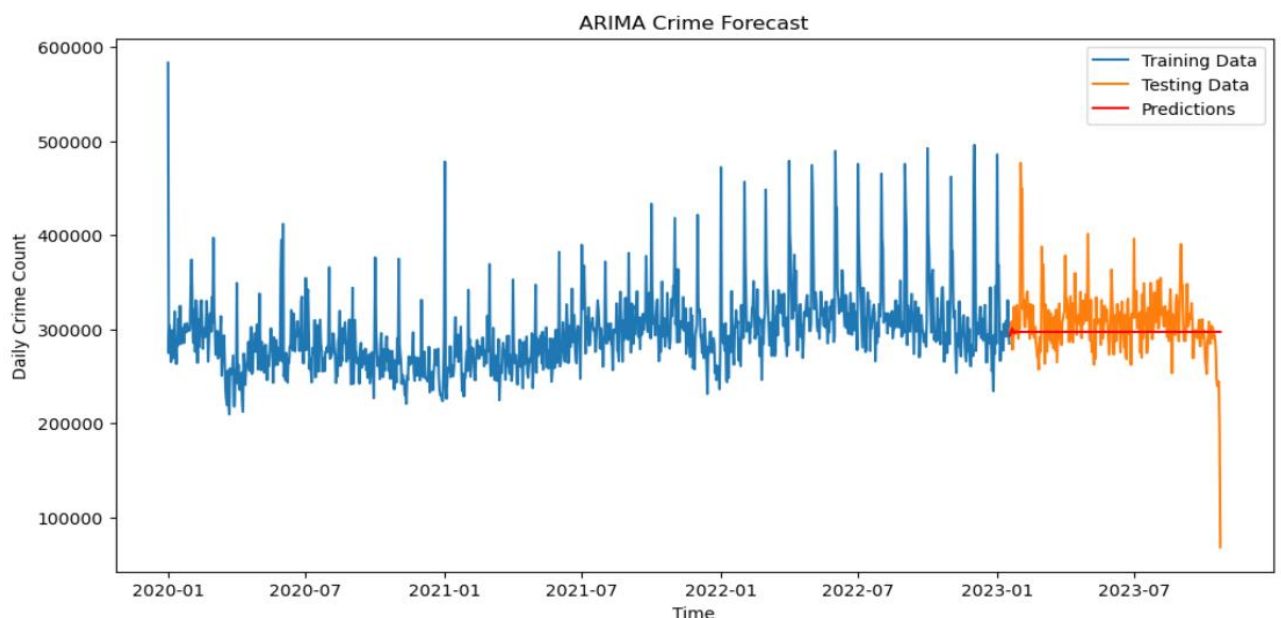
In conducting this Exploratory Data Analysis, we generated a synthetic dataset as follows,

```
events_data = pd.DataFrame({
    'Event_Date': ['2021-06-01'],
    'Event_Description': ['Policy Change']
})
```

Following the integration of this newly generated dataset with the existing crime dataset, we observed that there was no discernible impact on the crime rate counts.

## PREDICTING FUTURE CRIME TRENDS

We have split the dataset into train and test datasets and used ARIMA model to forecast the future crime trends based on historical data. Future trend crime rates remain the same as that of current rate.

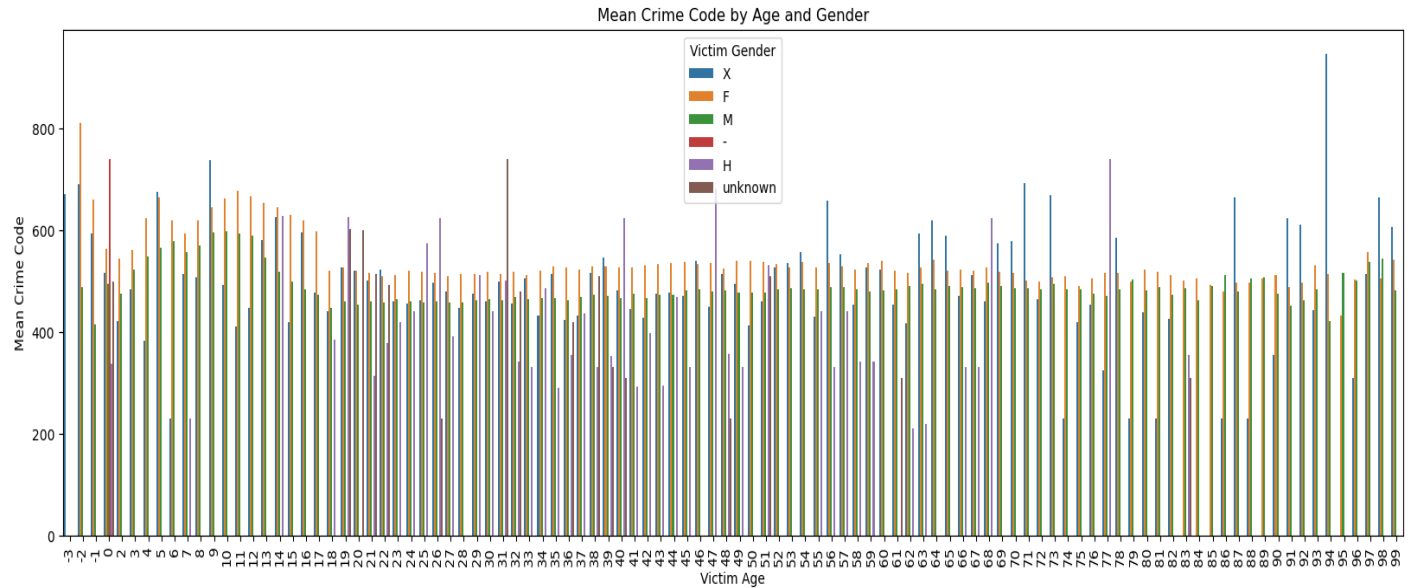


## Correlation between demographic factors (e.g., age, sex) and specific types of crimes

Observations after visualizing correlation between crime rate and ages using seaborn bar.

- The dataset does not provide information on certain gender codes.
- In general, crime rates are observed to be higher for females when compared to males.
- The highest recorded crime count, exceeding 800 incidents, is associated with individuals aged 94. However, the gender details for this age group are undisclosed, indicated as 'X'.





## CONCLUSION

In the course of this project, we undertook a comprehensive examination of a crime dataset, encompassing meticulous data inspection, cleansing, and an extensive exploratory data analysis (EDA).

Our initial steps involved scrutinizing the dataset's structural elements and attributes, where we diligently identified potential issues, including missing values and inconsistencies. Additionally, we conducted an outlier analysis specifically focusing on the "Age" field.

During the Exploratory Data Analysis (EDA) phase, we crafted a wide array of visual representations, including bar plots, line plots, and heat maps. These visualizations were instrumental in uncovering valuable insights into diverse aspects, such as the evolution of crime trends across different years, the ebb and flow of crime rates throughout the seasons, and the geographical distribution of crime rates in specific areas.

Furthermore, we simulated specific datasets and conducted EDA on several scenarios, including an examination of correlations between economic factors and their potential influence on crime rates. We also explored the impact of policy changes on crime rates, contributing to a well-rounded analysis of the dataset.

## Key Findings

- The crime rates reached their zenith in January 2020 and during the period between May 2022 and October 2022.
- The prevailing type of crime within the dataset is "Vehicle theft."

- 'Central' emerged as the area with the highest recorded crime rates, surpassing 50,000 incidents, while 'Foothill' reported the lowest, with fewer than 30,000 incidents.

### **Limitations**

- This analysis is based on the available dataset, which may not capture all relevant factors influencing crime rates
- Correlation does not imply causation, and further research is required to understand the complex dynamics of crime