FDA Project Report

Members – Group 8

Sneha Amin

Sai Dheeraj Malkar

Sharanaya Chikkegowda

Rakshit Dharmappa

Ankur Pandey

Contents:

Data Cleaning Steps

1. We loaded our CSV data into Pandas data frame.

2. Then we checked for the columns which might contain null values irrespective of datatype.

   a. Crime_data.isnull().any()

3. Filling the Null values according to the provided string for Unknown values on the Crime Data website.

   a. Crime_data['Vict Sex'].fillna('X', inplace = True)

   b. Crime_data['Vict Descent'].fillna('X', inplace = True)

4. Dropping records which have Premise code as Null because the small number (5 columns) will not affect our analysis.

   a. Crime_data.dropna(subset=['Premis Cd'])

   b. Crime_data = Crime_data.reset_index(drop=True)

5. No need to clean premis desc as we don't need it for our analysis. Also, no available data for it.

   a. Crime_data['Premis Desc'].fillna('Unknown', inplace = True)

6. For each blank weapon code, we do not have corresponding value of weapon desc

   a. Crime_data['Weapon Used Cd'].fillna('0', inplace = True)

   b. Crime_data['Weapon Desc'].fillna('Not available', inplace = True)

7. No need to clean Crime code 1 and 2 as they have corresponding values for the crimes committed. Also, crime code 3 and 4 are not required.

8. Not cleaning Cross Street as all the values are empty.

NOTE: We got more details regarding the data from the provided link for the data set https://catalog.data.gov/dataset/crime-data-from-2020-to-present and also, researched further on the internet.

Normalization steps:

1. Replacing Age 0yrs with mean of the Vict Age column

   No_null = pd.read_csv(r"C:\Users\ankur\Downloads\Cleaned_Crime_data.xls")

   No_null['Vict Age'] = No_null['Vict Age'].replace(0, No_null['Vict Age'].mean())
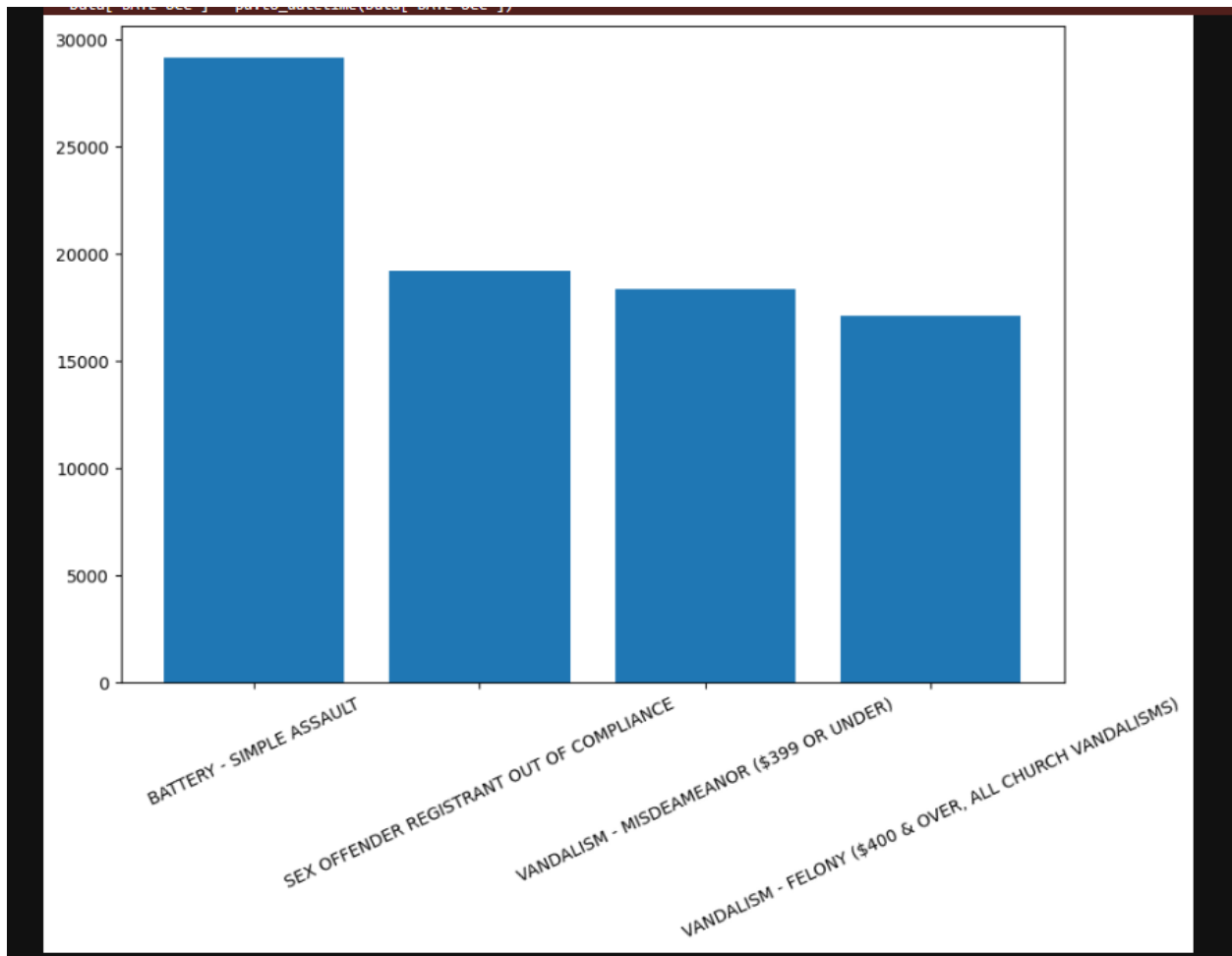
Duplicate data Check:

1. We do not have any duplicates present in the dataset.
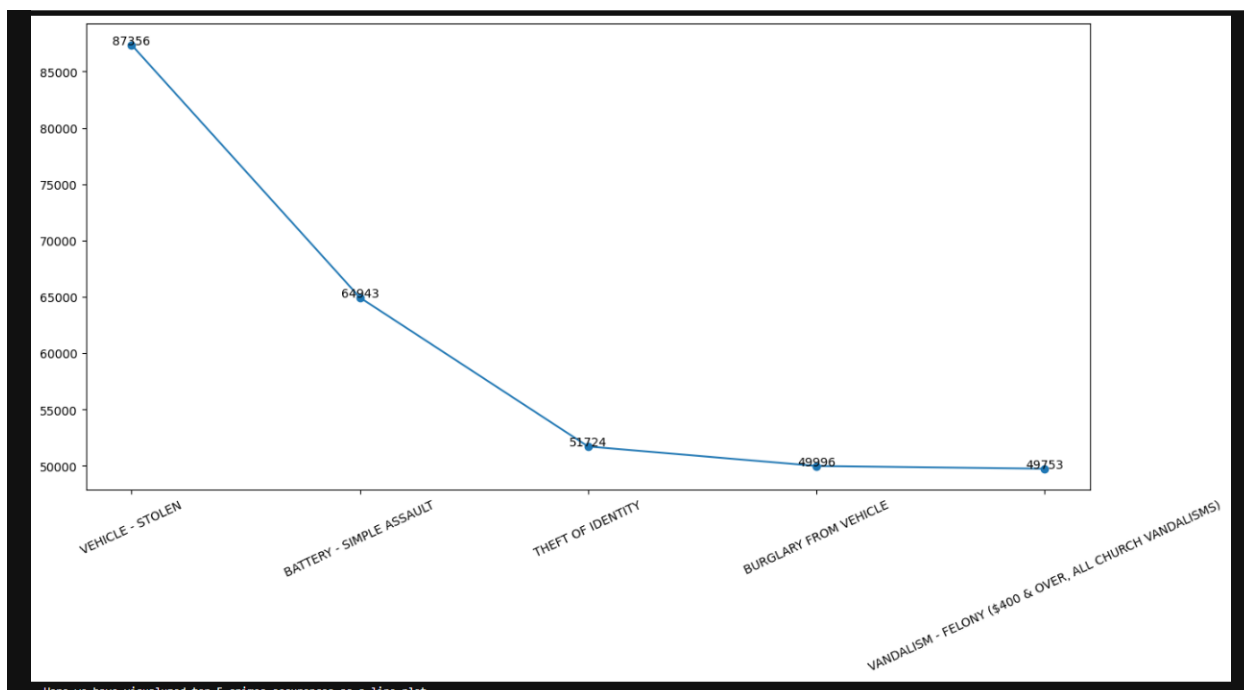
   No_null.duplicated().any()

NOTE: No need to encode the data as we already have corresponding numerical data present for categorical columns

Exploratory Data Analysis (EDA):
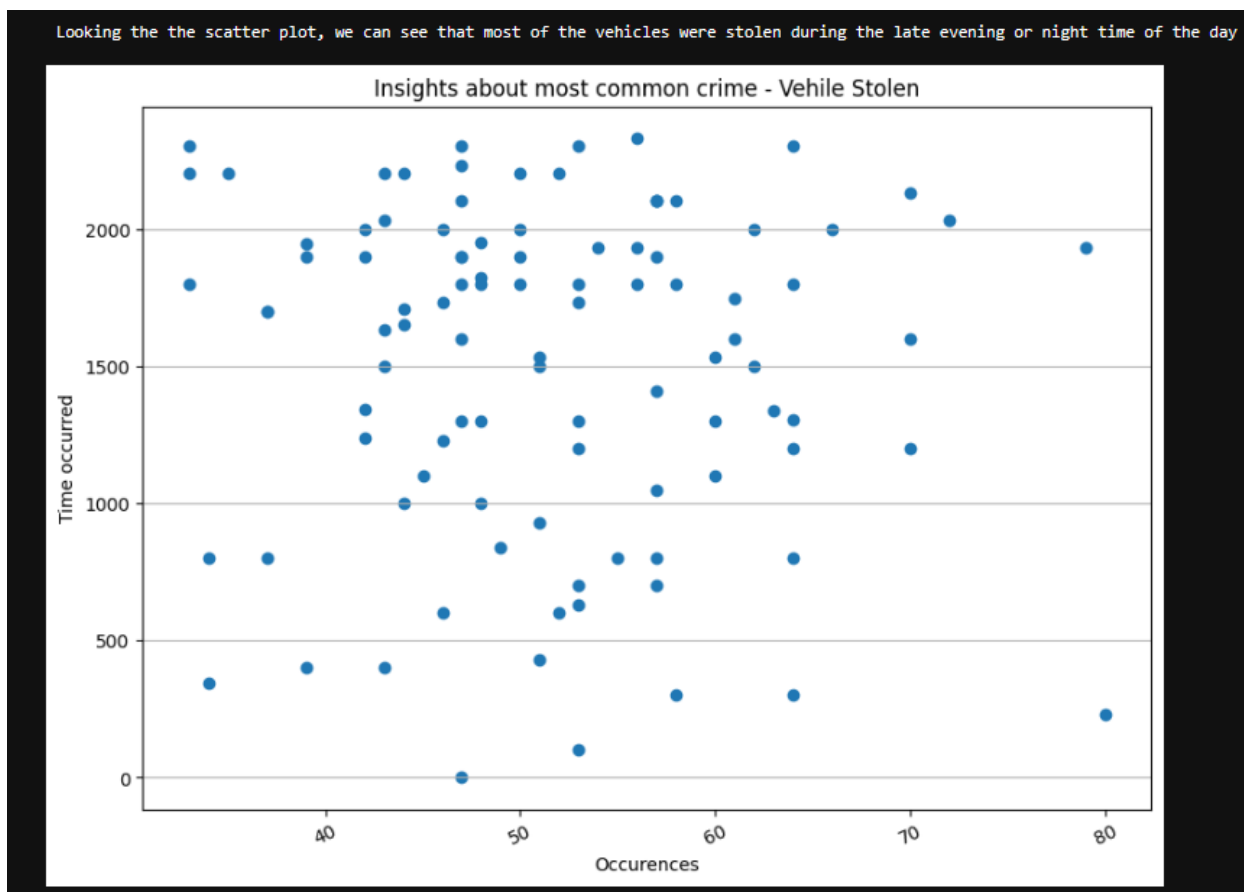
o Analyze and visualize seasonal patterns in crime data.

o Identify the most common type of crime and its trends over time.

Here we have visualuzed top 5 crimes occurences as a line plot

Looking the the scatter plot, we can see that most of the vehicles were stolen during the late evening or night time of the day

o Investigate if there are any notable differences in crime rates between regions or cities.

The below plot shows us that the crime occurrences in the Central, 77th Street, Pacific, Southwest and Hollywood are far more than other areas.
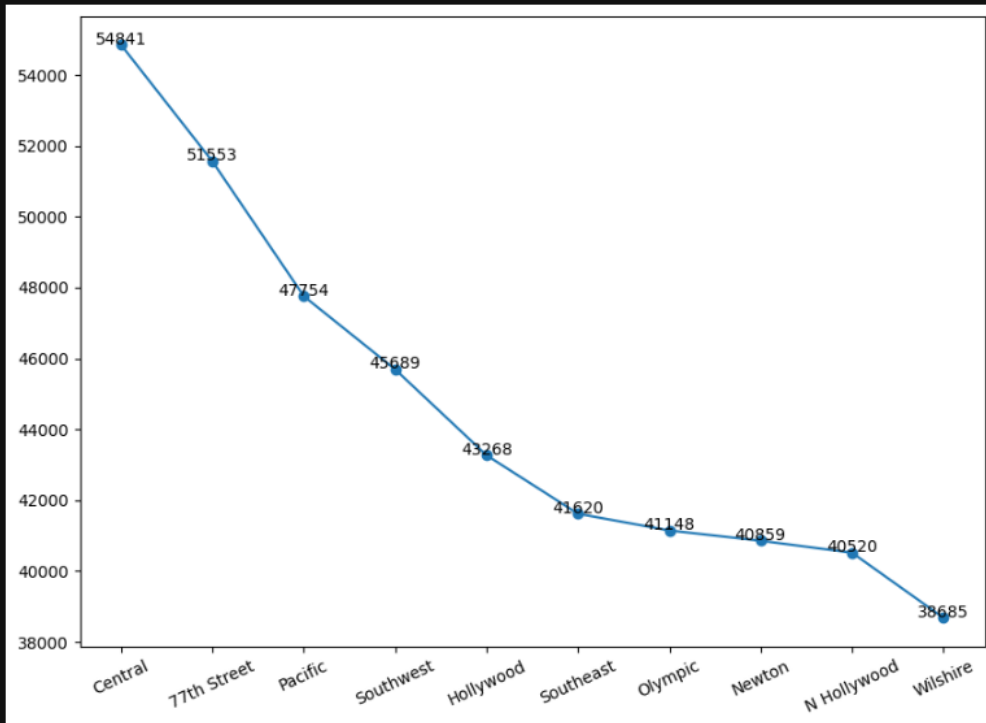
o Explore correlations between economic factors (if available) and crime rates.

Ans: We do not have any data to derive the correlation between economic factor and crime rates.

o Analyze the relationship between the day of the week and the frequency of certain types of crimes.

The below plot shows us that Most of the crimes occurred during the weekends

## Insights about Crime occurrences according to weekdays
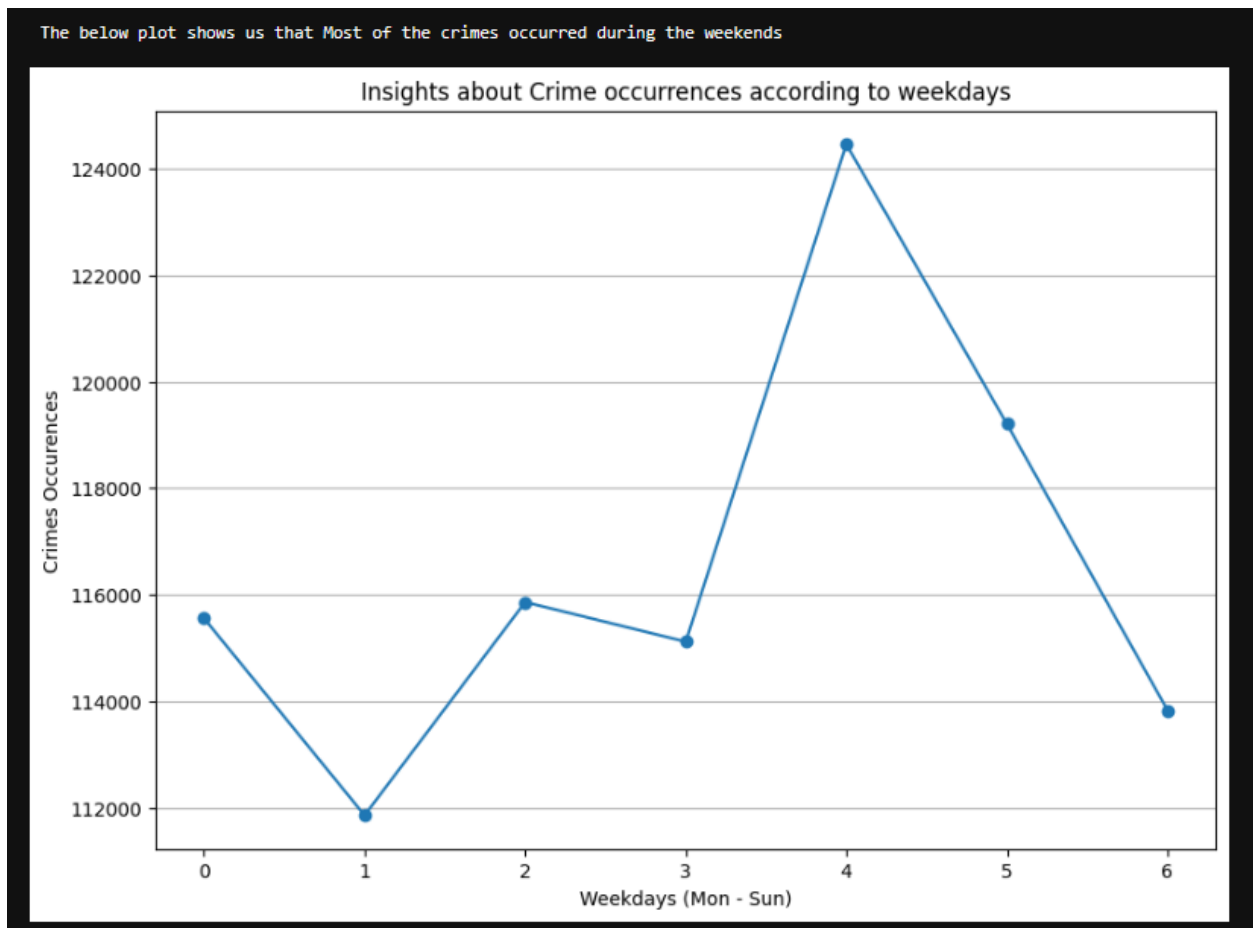
In this graph, we have visualized crimes according to the weeks. We have observed that most of the crimes occurred during the weekends.

- **Advanced Analysis (Optional):**

  o Use predictive modeling techniques (e.g., time series forecasting) to predict future
  crime trends.
 o Explore additional questions or hypotheses related to the dataset.

We used ARIMA model to predict future Crime trends. Below are the visualized future trends of the Crime as predicted by our ARIMA Model.
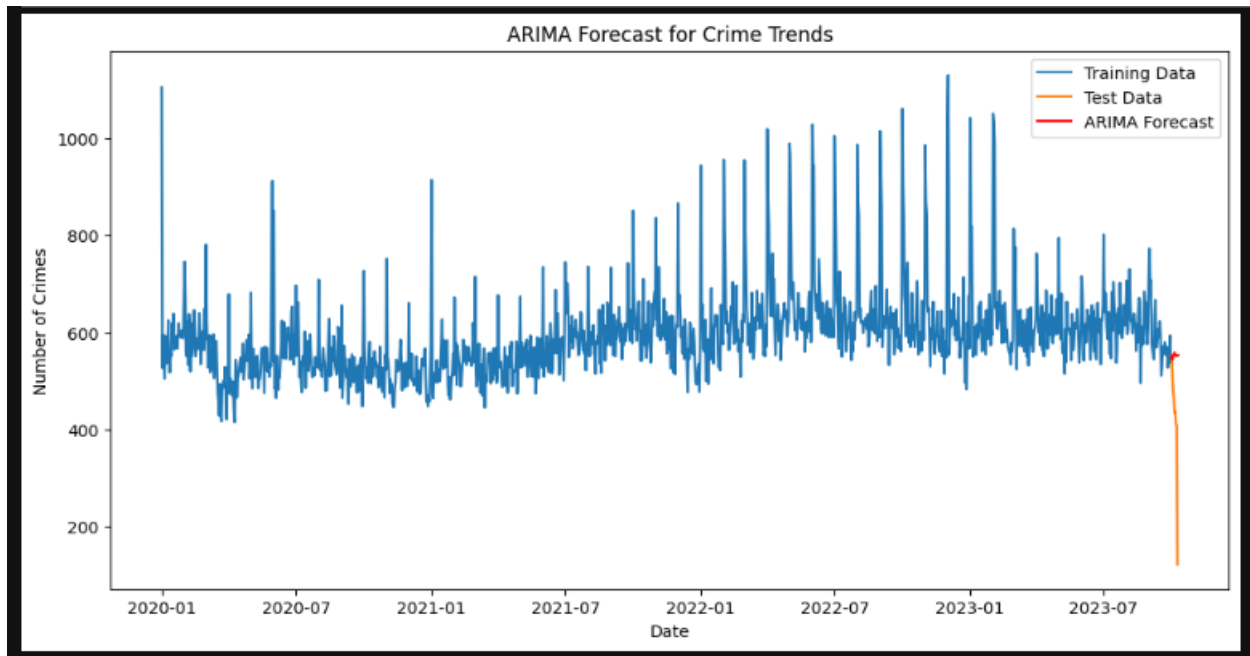
The model predicts that the crime rate will drop further as the time passes. This is true as we have observed that the crime rate dropped significantly from year 2022 to 2023 due to LAPDs additional forces.

Find the solutions to these questions:

1**. Overall Crime Trends**: Calculate and plot the total number of crimes per year to visualize the trends.
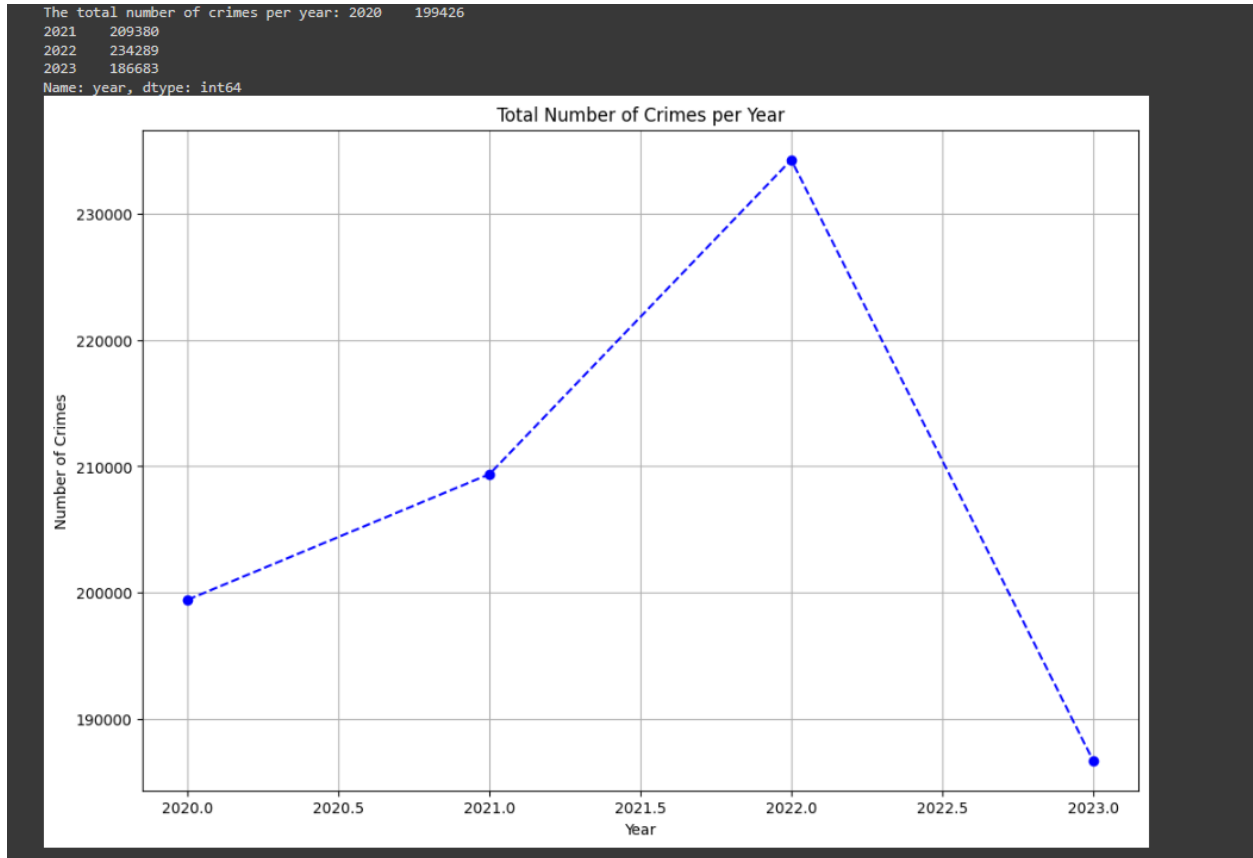
In this question we've visualized the total number of crimes per year.
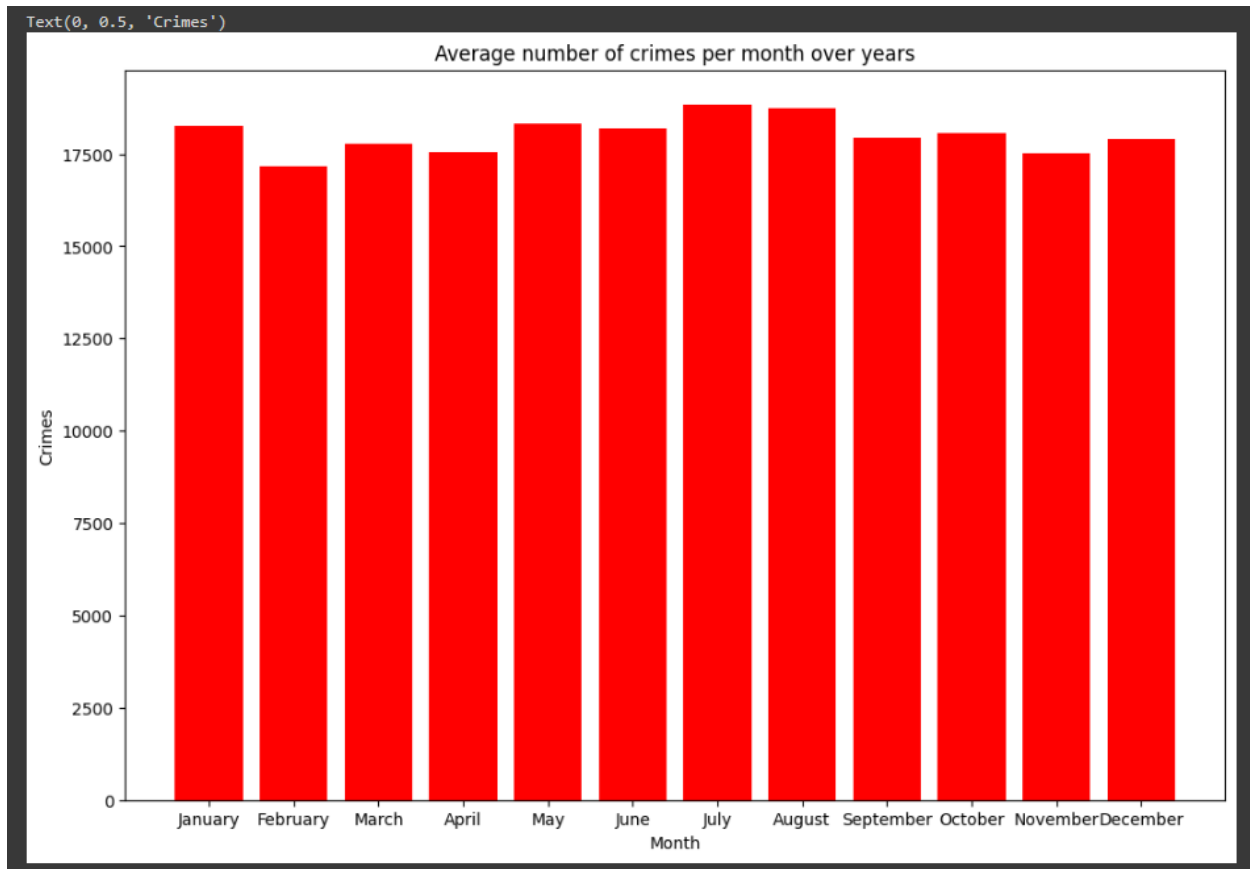
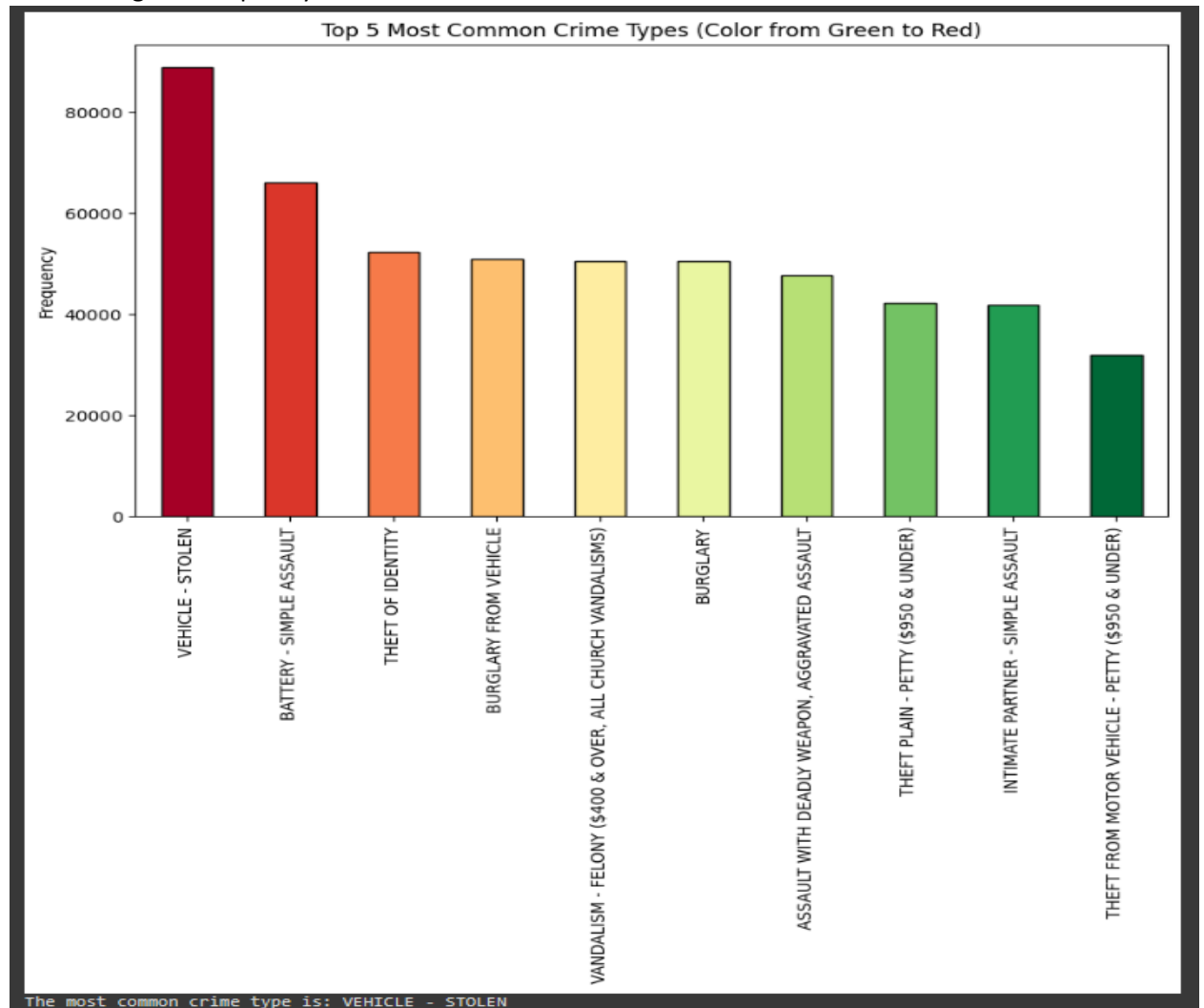The total number of crimes per year:

2020   199426

2021   209380

2022   234289

2023   186683

2. Seasonal Patterns: Group the data by month and analyze the average number of crimes per month over the years.

Average number of crimes per month over years

3.Most Common Crime Type: Count the occurrences of each crime type and identify the one with the highest frequency.


The most common crime type is: Crm Cd Desc

| Crime Type	Frequency
VEHICLE - STOLEN	87356
BATTERY - SIMPLE ASSAULT	64943
THEFT OF IDENTITY	51724
BURGLARY FROM VEHICLE	49996
VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)	49753
BURGLARY	49658
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	46986
THEFT PLAIN - PETTY ($950 & UNDER)	41463
INTIMATE PARTNER - SIMPLE ASSAULT	41142
THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)	31537

4. **Regional Differences**: Group the data by region or city and compare crime rates between them using descriptive statistics or visualizations
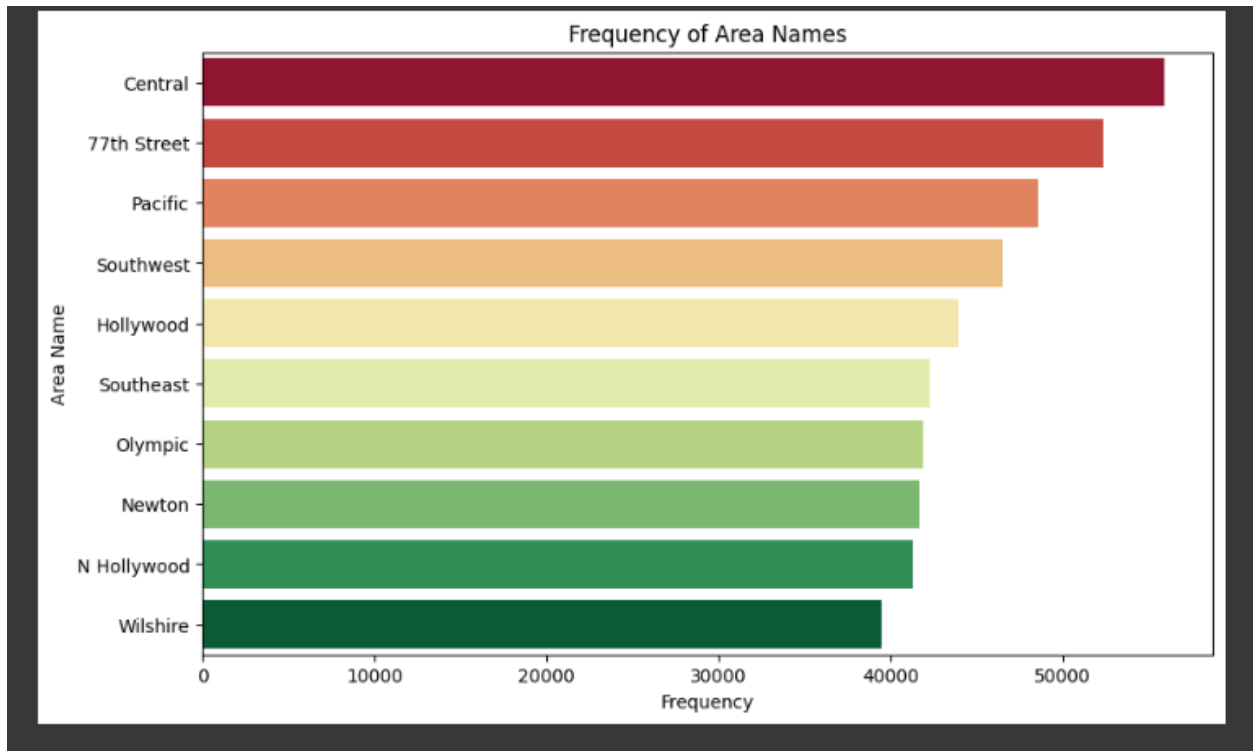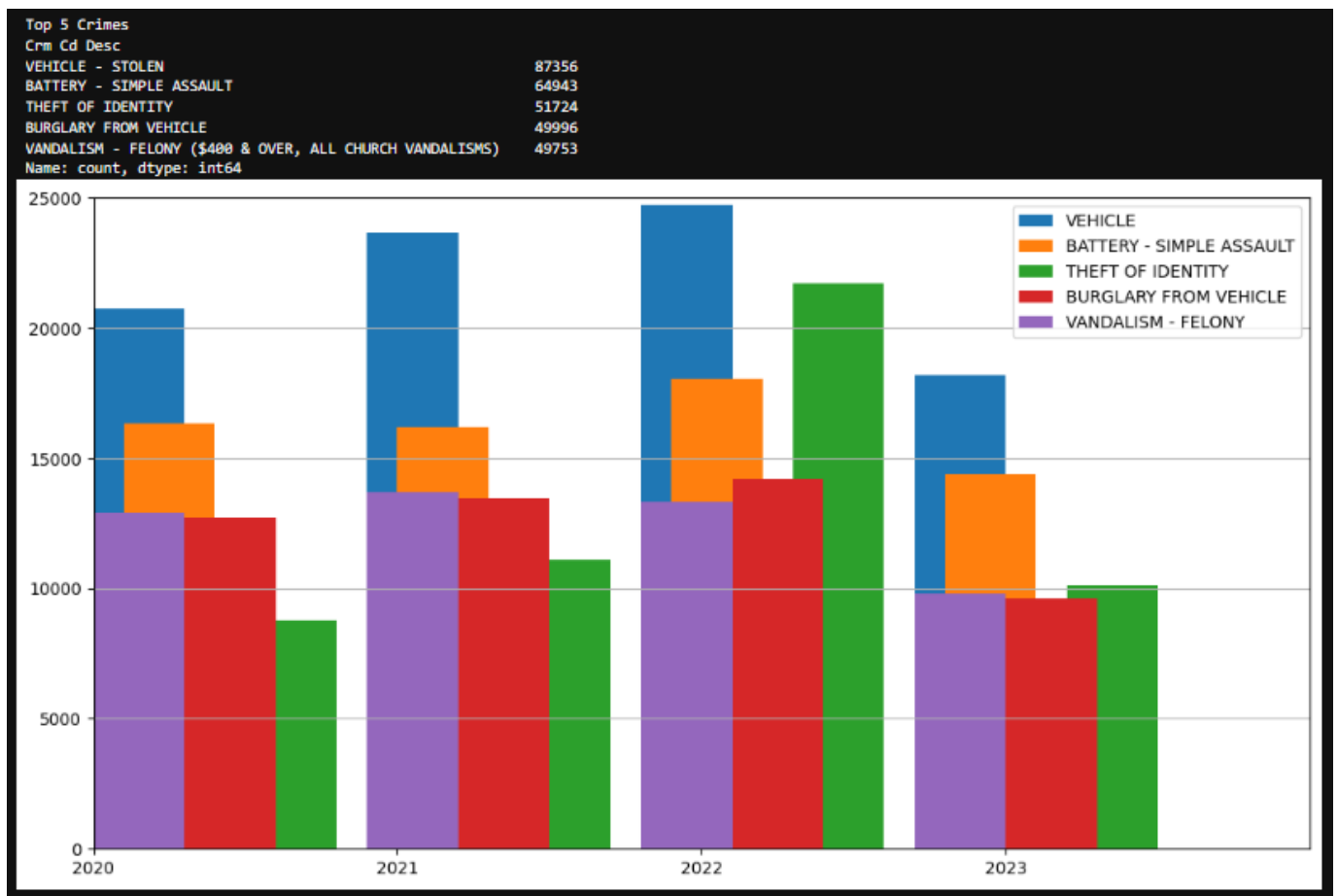
5. Correlation with Economic Factors: Collect economic data for the same time frame and use statistical methods like correlation analysis to assess the relationship between economic factors and crime rates.
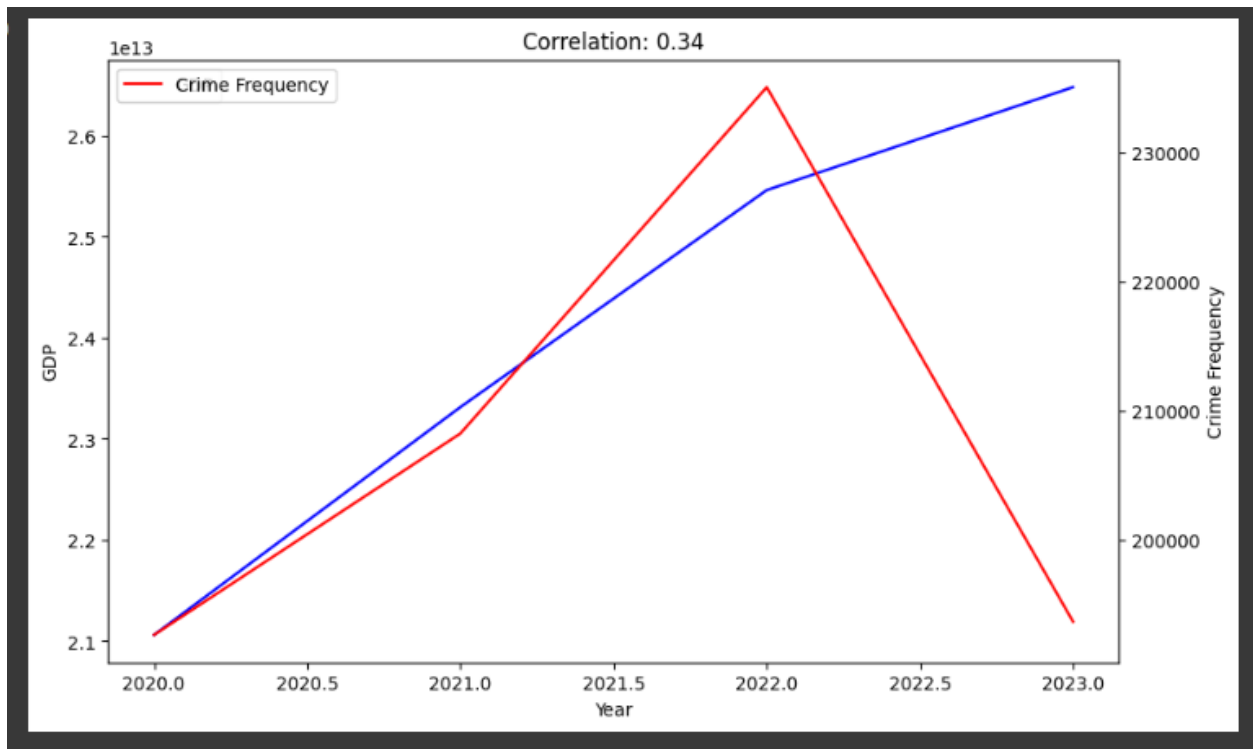
6. **Day of the Week Analysis**: Group the data by day of the week and analyze crime frequencies for each day.

Crime Frequencies by Day of the Week
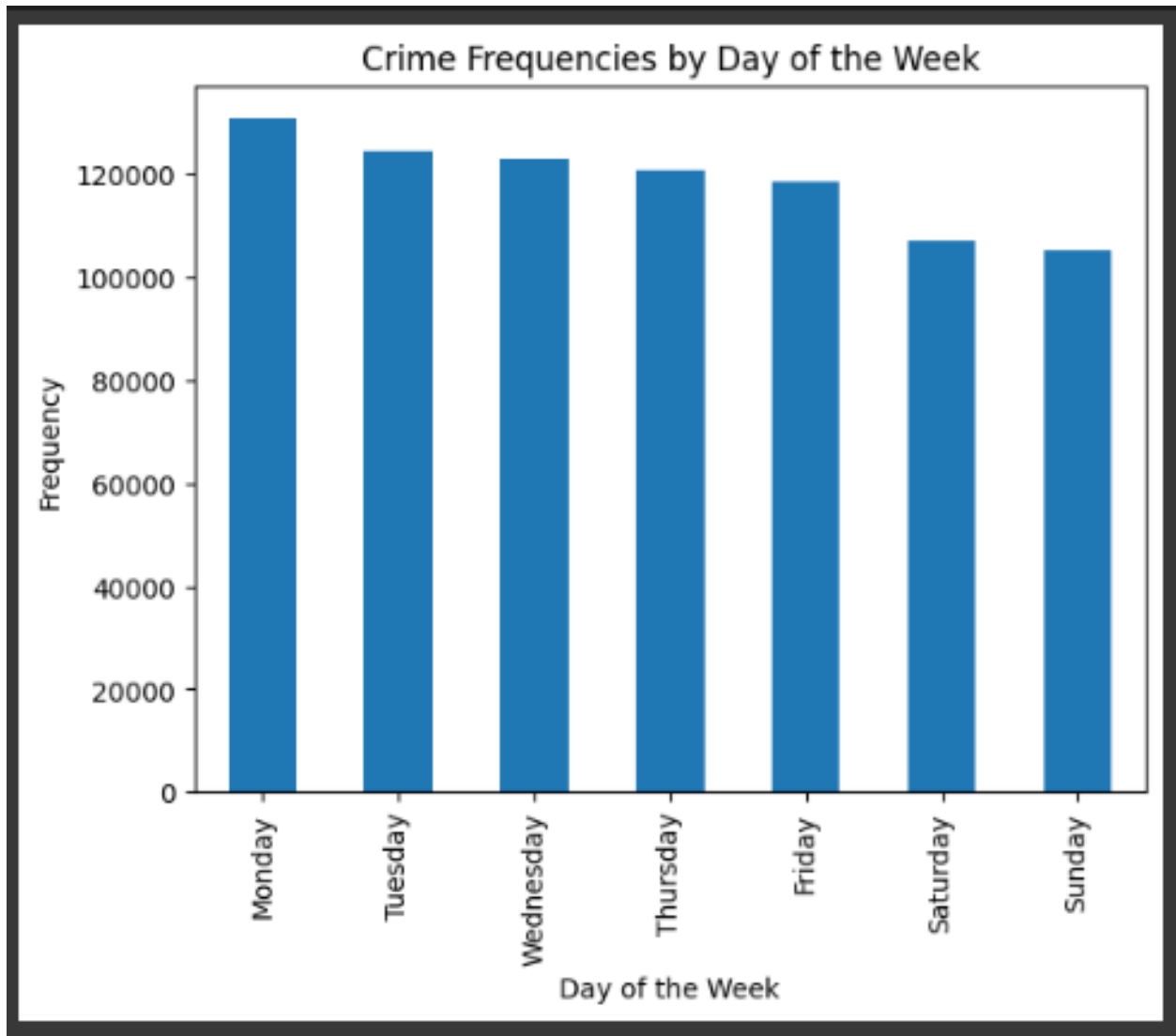
7. **Impact of Major Events**: Identify major events or policy changes during the dataset period and analyze crime rate changes before and after these events.

```
Top 5 Crimes
Crm Cd Desc
VEHICLE - STOLEN                                            87356
BATTERY - SIMPLE ASSAULT                                   64943
THEFT OF IDENTITY                                          51724
BURGLARY FROM VEHICLE                                      49996
VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)    49753
Name: count, dtype: int64
```


From the above plot, we noticed that the crime rates decreased after 2022. This is due to LAPD expanding their budget by $3.2B. The LAPD hired hundreds of more officers which ultimately resulted in a decrease of the crime rates.

https://www.latimes.com/california/story/2023-06-01/crime-is-down-in-la-as-city-plans-to-expand-lapd

8. **Outliers and Anomalies:** Use statistical methods or data visualization techniques to identify dataset outliers and investigate unusual patterns.

```
below is the list of outlier value along with counts in the age column:
Vict Age
99.0    305
94.0     92
96.0     84
95.0     76
98.0     64
97.0     61
120.0     1
Name: count, dtype: int64
Series([], Name: count, dtype: int64)
"
```
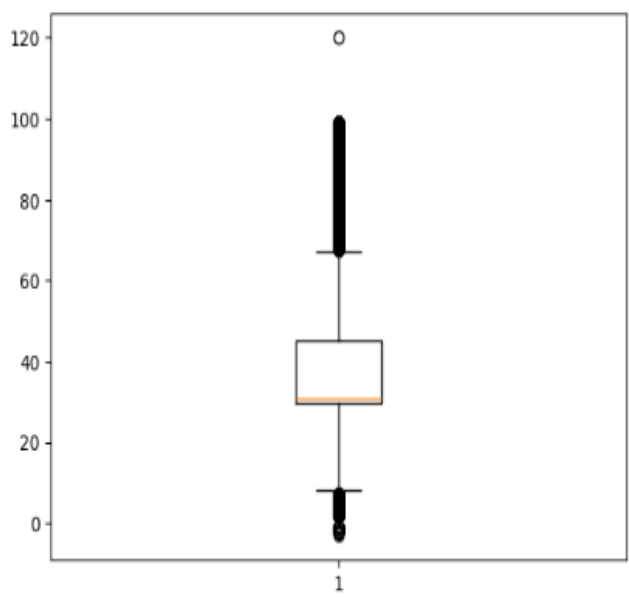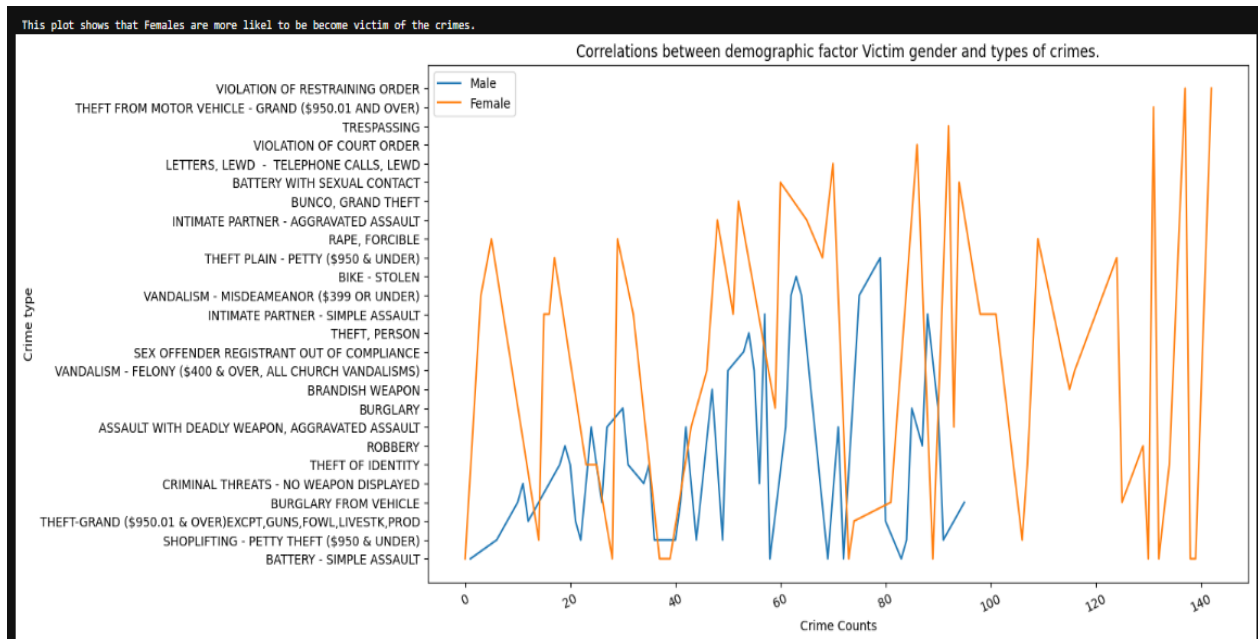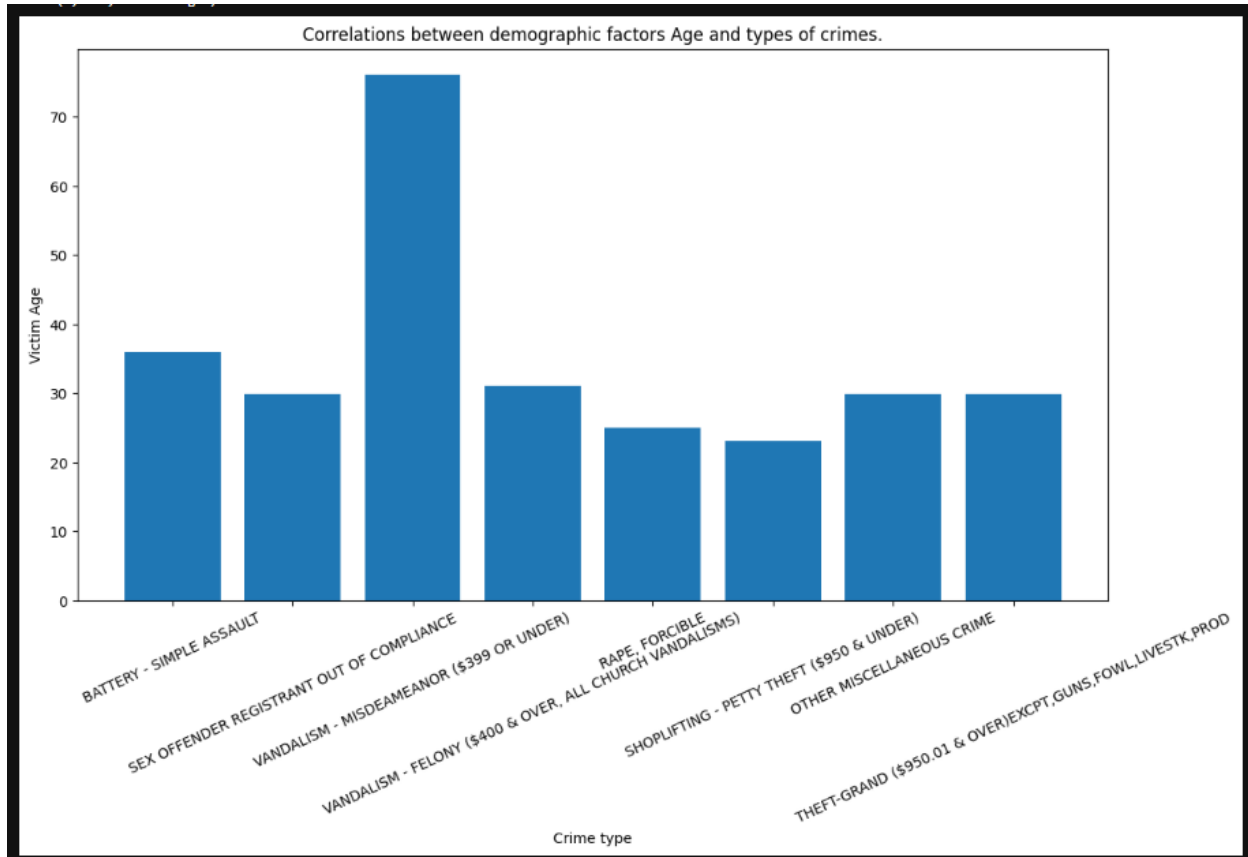
We analyzed only age column for outlier as other columns cannot have outlier values other than NULL or 0. In the age column we observed an unusual pattern that, Crime of all sorts were commited on the people elderly people(above 90yrs).


We analyzed only the age column for outlier as other columns cannot have outlier values other than NULL or 0. In the age column we observed an unusual pattern that, Crime of all sorts were committed on the people elderly people (above 90yrs).

9. **Demographic Factors**: Analyze the dataset to identify any patterns or correlations between demographic factors (e.g., age, gender) and specific types of crimes.

Correlations between demographic factors Age and types of crimes.

This plot shows that Females are more likel to be become victim of the crimes.
Correlations between demographic factor Victim gender and types of crimes.

The above plot shows that Females are more likely to be become victim of the crimes.

10. **Predicting Future Trends:** Employ time series forecasting methods, such as ARIMA or Prophet, to predict future crime trends based on historical data. Consider incorporating relevant external factors into your models.
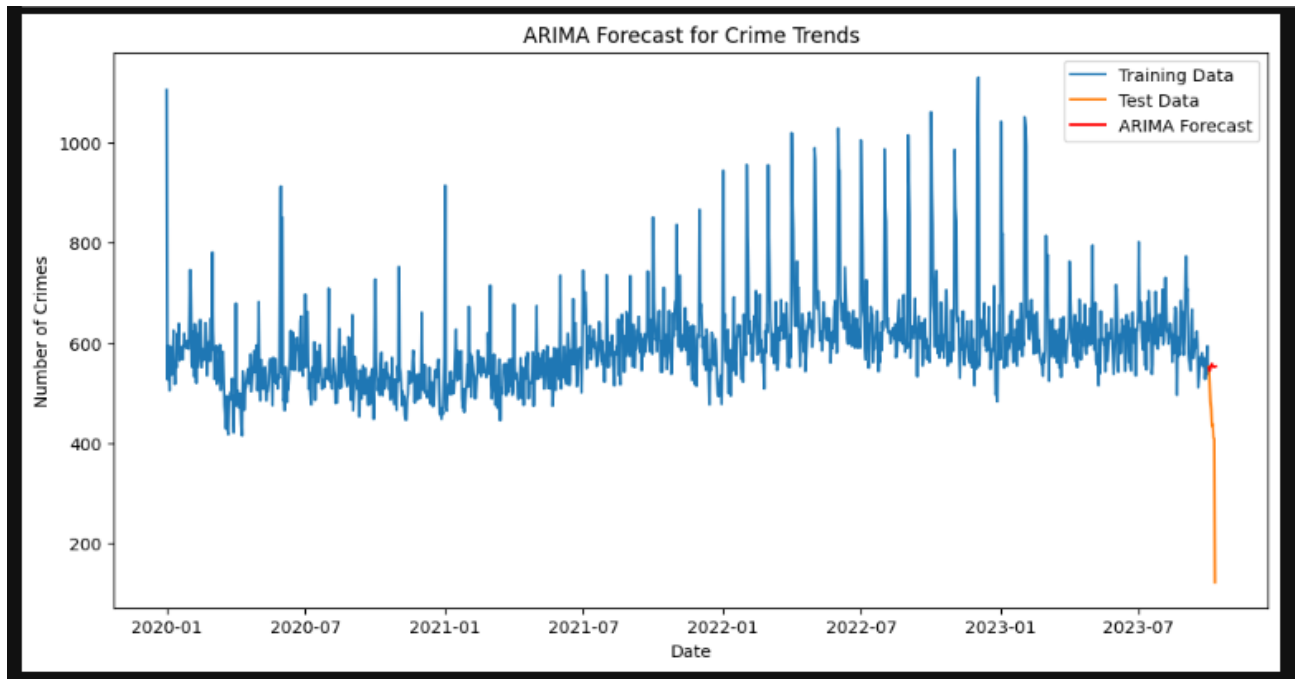
Our model predicts that the crime rate will drop further as the time passes. This is true as we have observed that the crime rate dropped significantly from year 2022 to 2023 due to LAPDs additional forces.

References:

- https://catalog.data.gov/dataset/crime-data-from-2020-to-present
- https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8
- https://data.lacity.org/api/views/2nrs-mtv8/files/4591b6bf-5846-4de0-9fb0-8780a77a036c?download=true&filename=MO_CODES_Numerical_20191119.pdf