



CRIME TREND ANALYSIS

IE6400 FOUNDATIONS FOR DATA ANALYTICS

PROJECT 1 REPORT

GROUP NUMBER 26

VIKRAMADITHYA PABBA (002853240)

DHEERAJ KUMAR GOLI (002897655)

LAAWANYAA SAI THOTA (002208176)

ROHAN REDDY PATHI (002832891)

UDHAY CHITYALA (002830533)

Data Source:

We have downloaded the crime data from Jan'20 to Oct'23 from [here](#).

About the data:

The dataset has a total of 28 columns. Some of the important columns include:

DATE OCC: Date of crime occurrence

TIME OCC: Time of crime occurrence

AREA NAME: Name of the area where the crime happened

Crm Cd Desc: Description of the crime

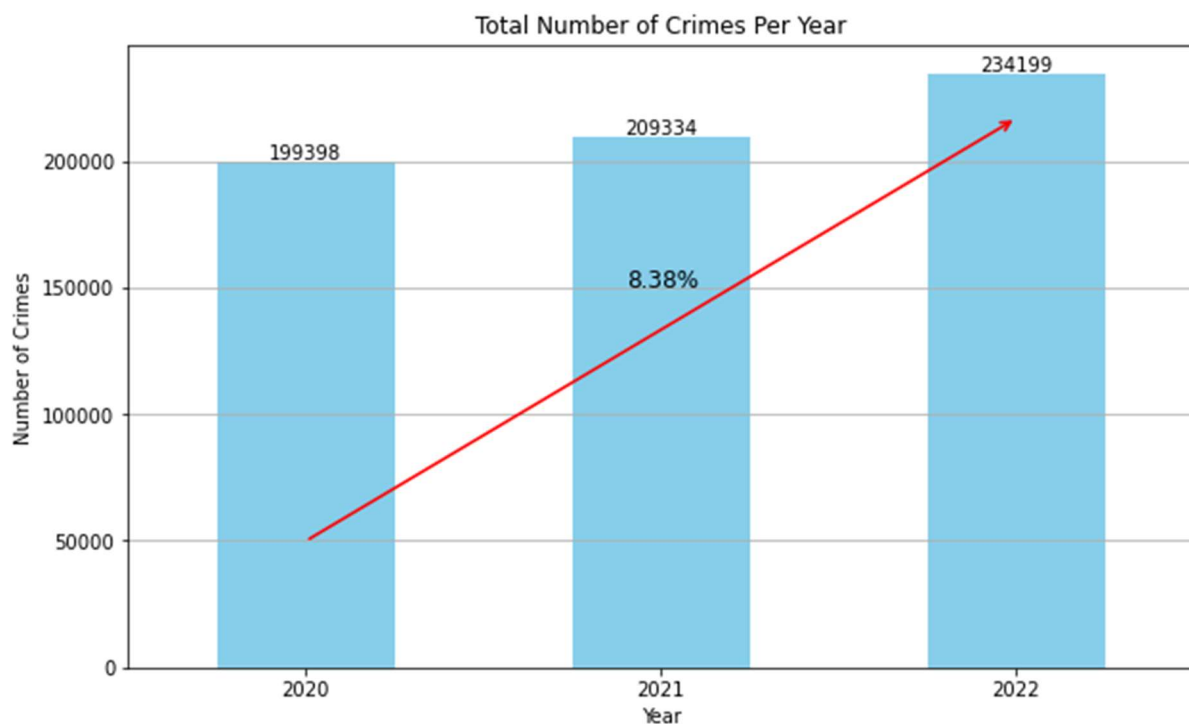
Vict Age: Age of the victim

Vict Sex: Gender of the victim

Data cleaning:

- We have considered years with complete data for the analysis. i.e., years having 12 months of data. 2023 has only 10 months of data (Jan'23 – Oct'23). So, we have excluded the year 2023 from the analysis
- We have replaced the null values in 'Vict Sex', 'Vict Descent', and 'Premis Desc' columns with 'Unknown'
- Replaced ['H', 'Unknown'] values in 'Vict Sex' column with 'X'
- We have checked for duplicate rows and found none
- We have converted date columns to datetime format.

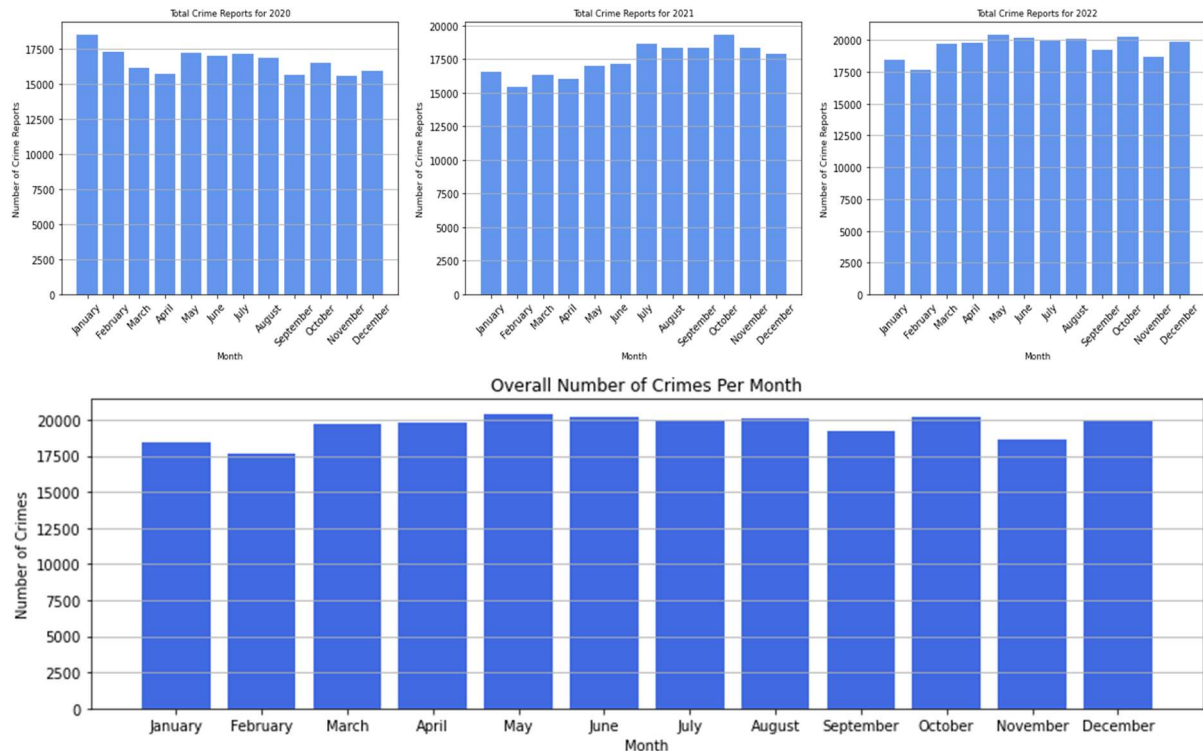
Overall crime trend:



Visualized the number of crimes across years using a bar chart (using matplotlib library)

- Number of crimes has been increasing at a CARG of 8.38% over the years (~200k in 2020 vs. 234k in 2022)

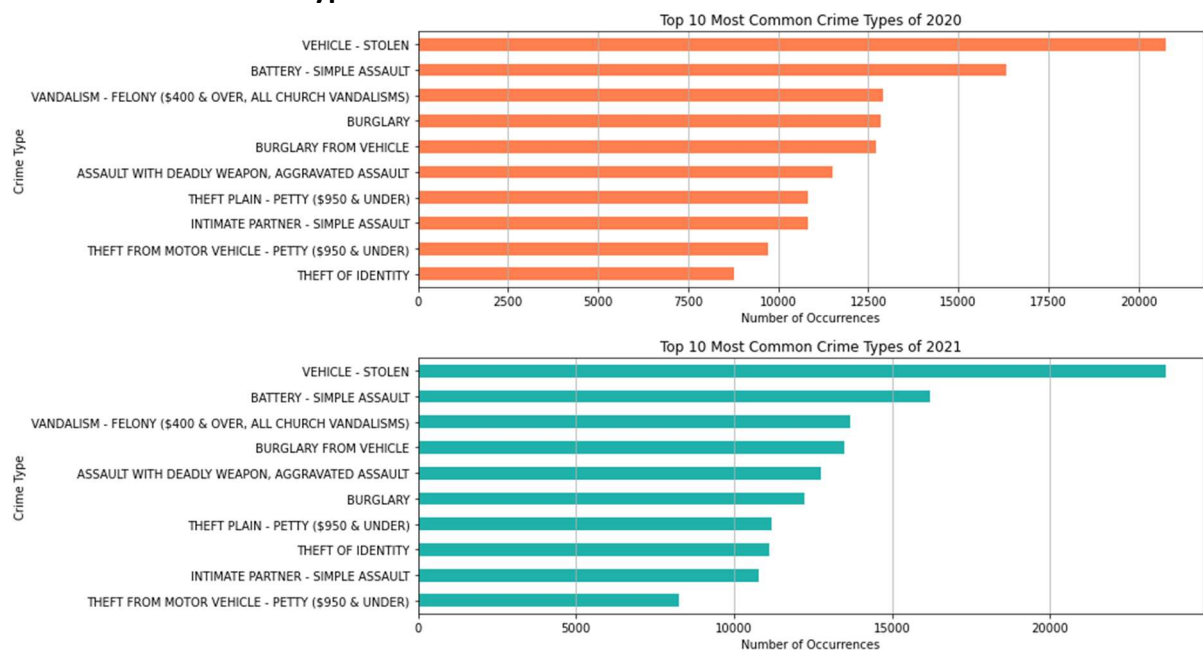
Seasonal patterns:

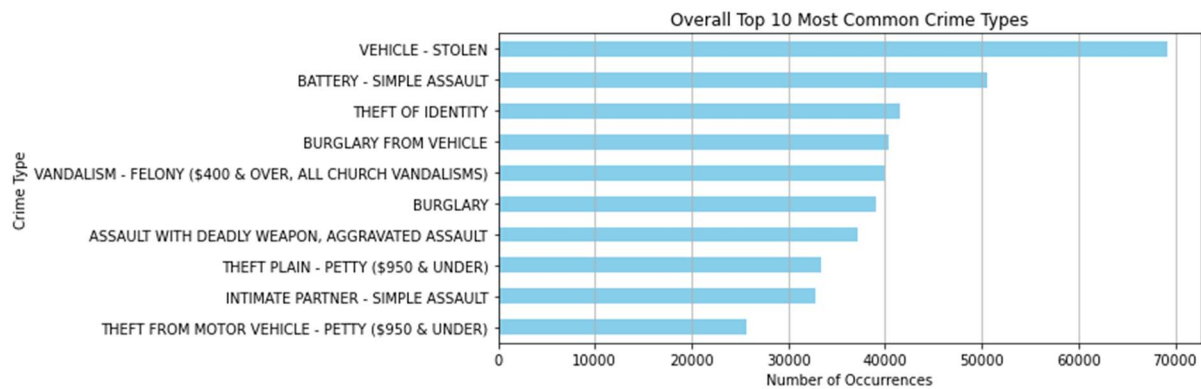
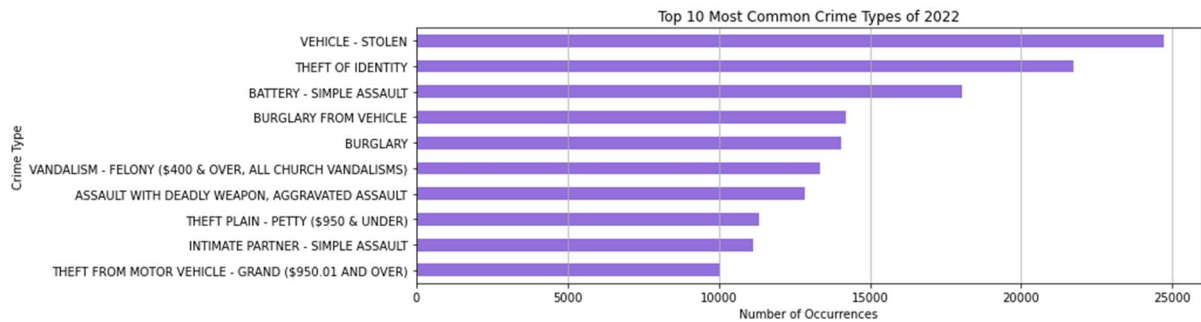


To check for seasonality, we have plotted a bar chart of number of crimes across months (using matplotlib library)

- From the chart it seems like there isn't any seasonality as such in the given crime data.
- However, there is a slight dip in the number of crimes in February which might be because of lower number of days in February compared to other months.

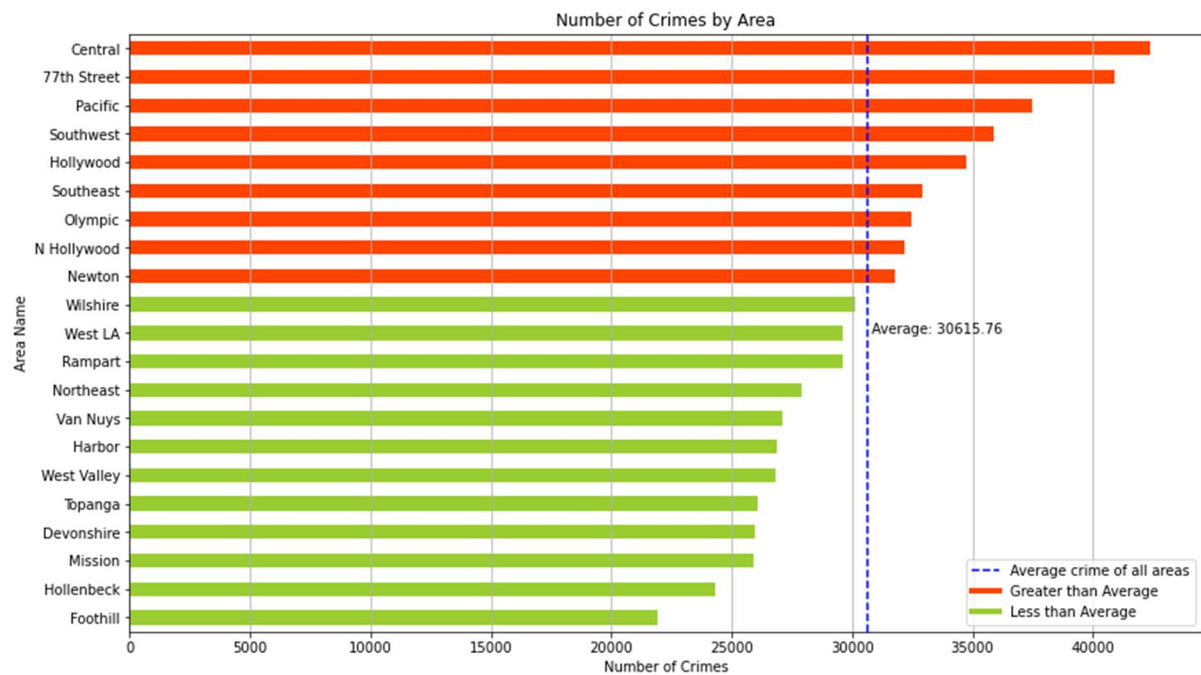
Most Common Crime Type:





- On an overall level
 - 'VEHICLE – STOLEN' is the most common type of crime (~70k)
 - 'BATTERY – SIMPLE ASSAULT' is the 2nd most common type of crime (~50k)
- In 2020
 - 'VEHICLE – STOLEN' is the most common type of crime (~21k)
 - 'BATTERY – SIMPLE ASSAULT' is the 2nd most common type of crime (~16k)
- In 2021
 - 'VEHICLE – STOLEN' is the most common type of crime (~24k)
 - 'BATTERY – SIMPLE ASSAULT' is the 2nd most common type of crime (~16k)
- In 2022
 - 'VEHICLE – STOLEN' is the most common type of crime (~25k)
 - 'THEFT OF IDENTITY' is the 2nd most common type of crime (~22k)
- 'THEFT OF IDENTITY' has been growing rapidly over the years (~8.5k in 2020 (10th position) vs. 22k in 2022 (2nd position))

Number of crimes across regions:



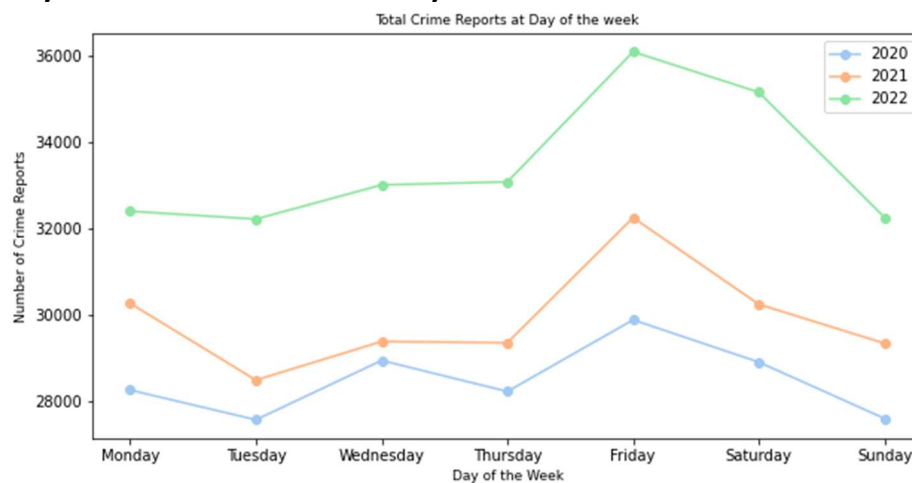
Illustrated the number of crimes across areas using horizontal bar chart (using matplotlib library)

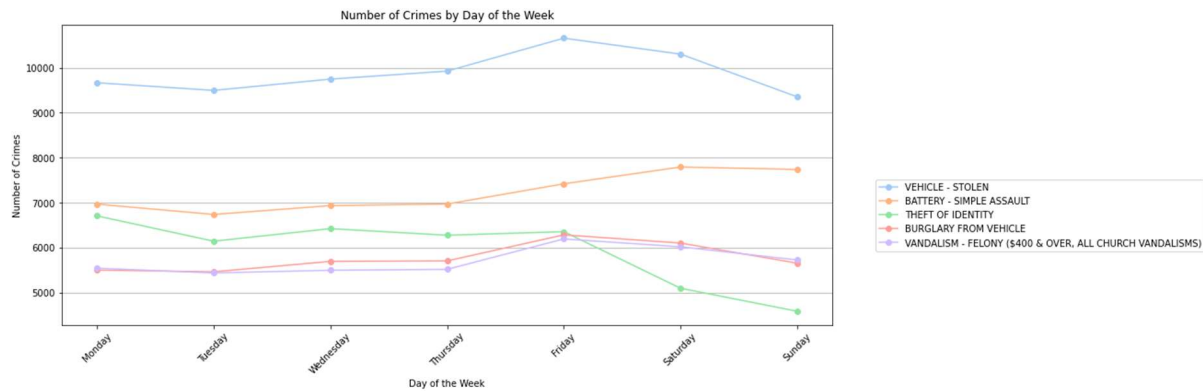
Represented the regions with # crimes higher than the average number of crimes in red color

Represented the regions with # crimes lower than the average number of crimes in green color

- On an average ~31k crimes were reported per area from 2020 to 2022
- 'Central' area has the highest number of crimes (~42k) while 'Foothill' has the least number of crimes (~22k)

Day of the Week vs. crimes analysis:



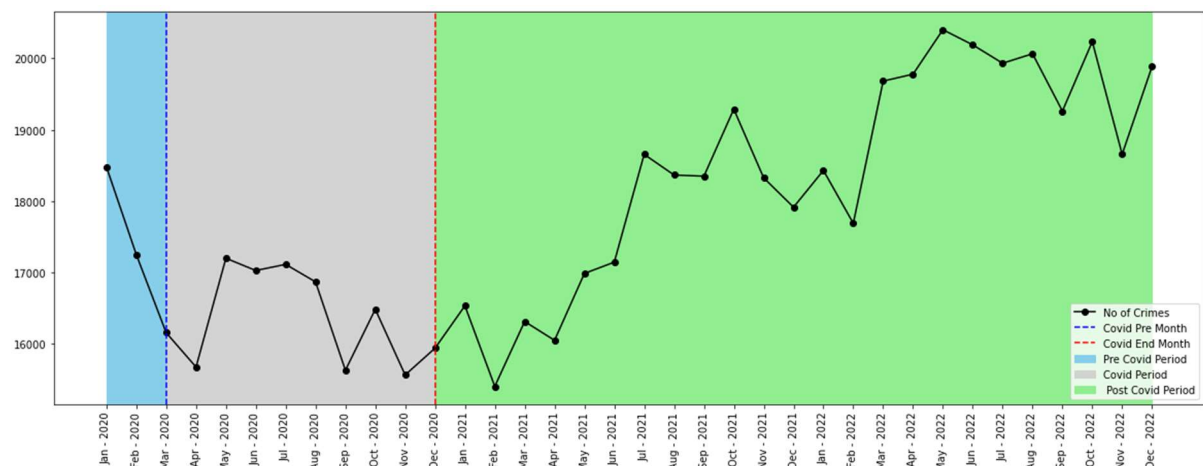


Illustrated the number of crimes by years based on day of the week using line chart (using matplotlib library)

- By year:
 - Friday and Saturday are witnessing higher number of crimes compared to other days
 - Number of Crimes on Tuesdays are lower compared to other days
- By crime type
 - 'THEFT OF IDENTITY' is lower on weekends (Saturday and Sunday)
 - All the other types of crimes are higher on Friday and Saturday

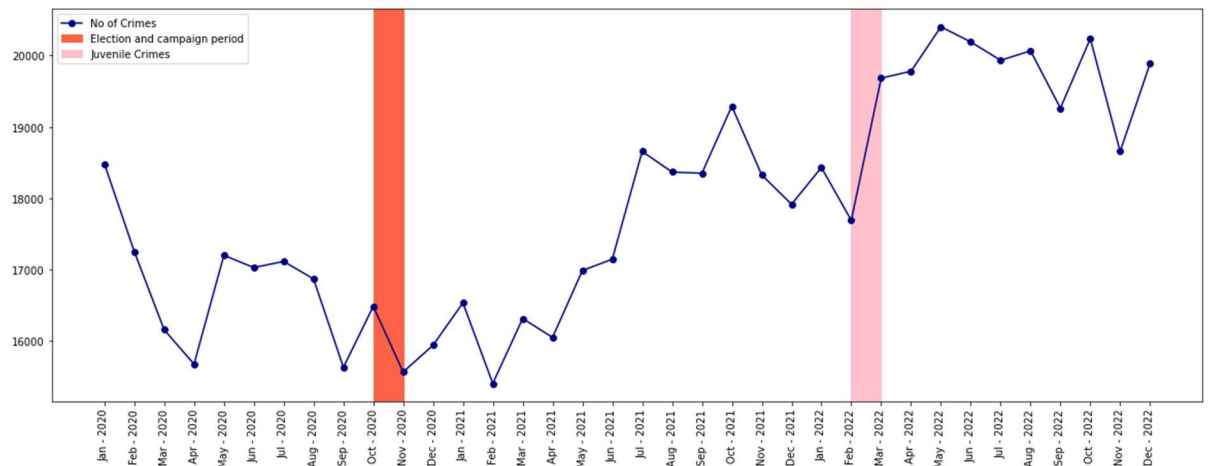
Impact of Major Events:

- **COVID-19**



- Before COVID (Jan 2020 - Mar 2020): Crime rates were relatively stable
- COVID-19 Impact (Mar 2020 - Dec 2020): When the pandemic started, there was a big drop in reported crimes. This is because of lockdowns and restrictions that kept people at home
- After COVID (Dec 2020 - Dec 2022): Crime reports gradually increased as restrictions eased and life returned to normal
- Special Months: The blue and red lines mark important points in time - one just before the pandemic and the other when it was considered over

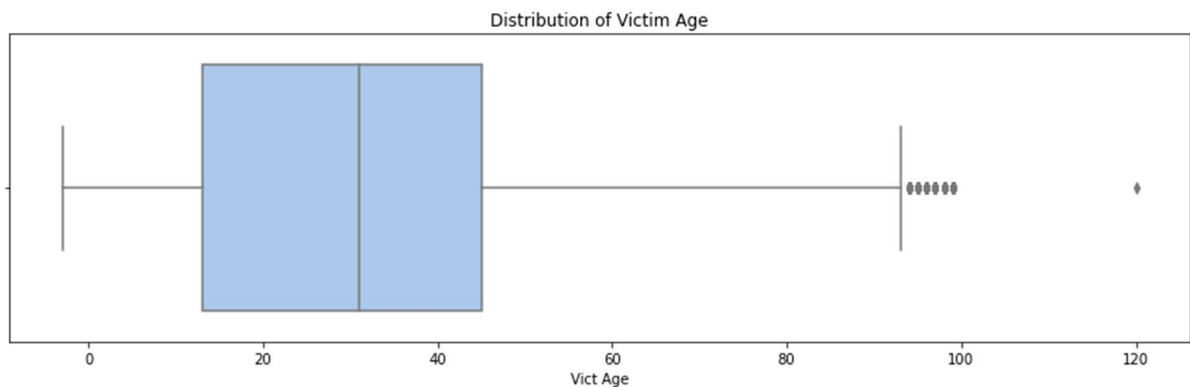
- **US Presidential elections and Juvenile Crimes**



- There seems to be a dip in number of crimes during the election period. This might be due to high security during election campaigns
- There is a significant uptick in number of crimes from March 2022. This might be due to gangs sending younger criminals to carry out their requests in a response to L.A. District Attorney not punishing minors

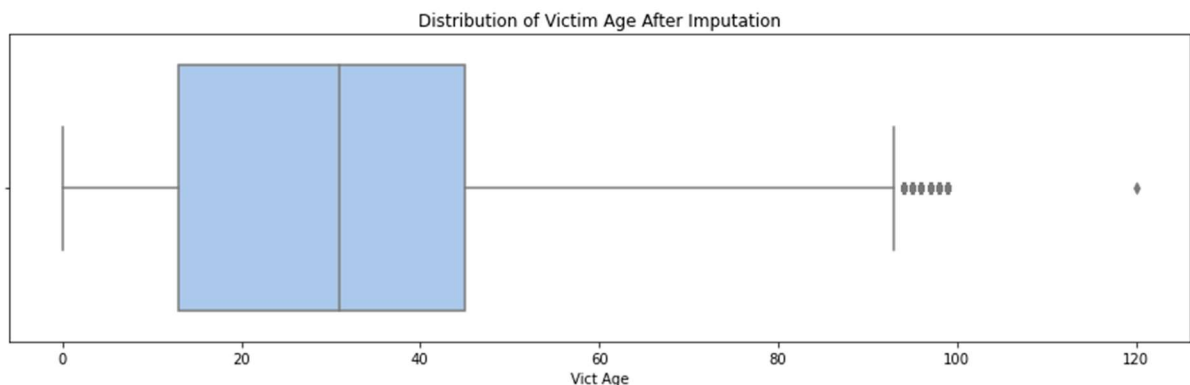
Outliers and Anomalies:

Victim age distribution

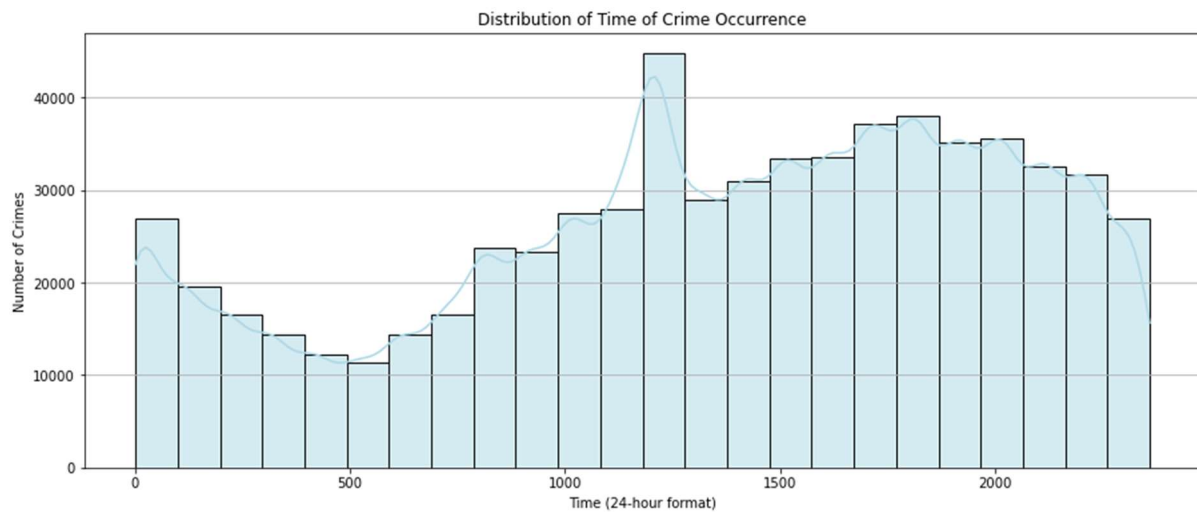


The boxplot visualizes the distribution of the "Vict Age" column. We can observe a few potential outliers, specifically some age values that seem unrealistically high or low.

For the purpose of this analysis, let's consider ages below 0 as outliers and handle them. One approach would be to replace such values with the median age.



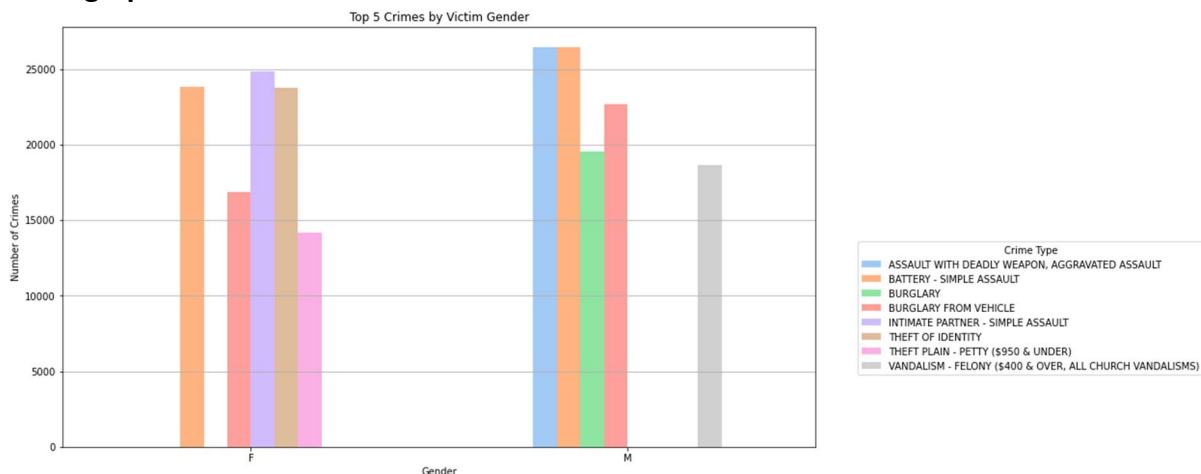
Crime occurrence by time:



The histogram showcases the distribution of crime occurrences throughout the day

- There seems to be a noticeable dip in crimes during the early morning hours, around 5-6 AM
- Crime occurrences start to increase from mid-morning and peak during the afternoon and early evening hours
- Late evening and nighttime (around 12 AM) also witness a significant number of crimes
- This distribution is fairly expected, as certain times of the day, like early mornings, are typically quieter with fewer people outdoors, potentially leading to fewer reported crimes

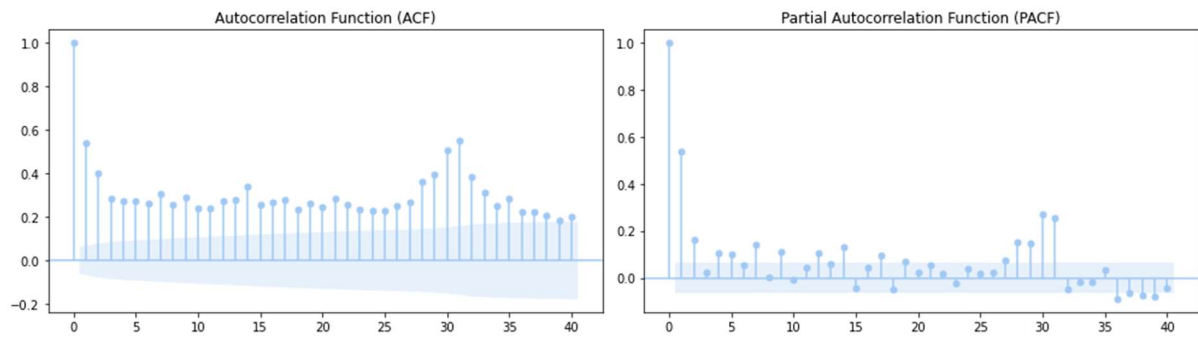
Demographic Factors:



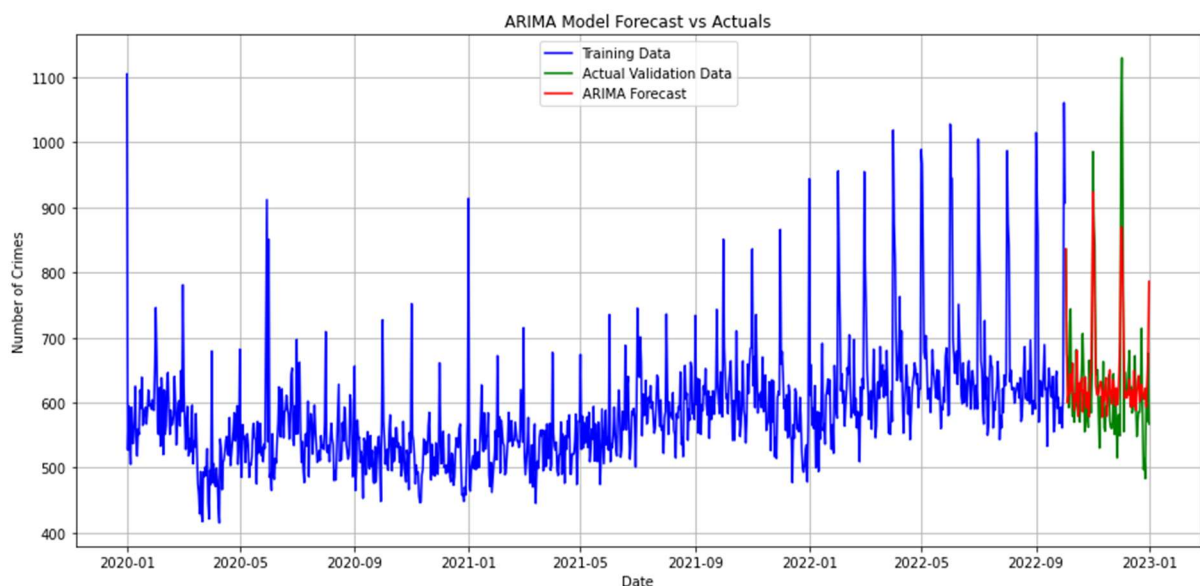
The bar chart presents the top 5 crimes experienced by victims of different genders:

- 'ASSAULT WITH DEADLY WEAPON, AGGREVATED ASSAULT' is the highest frequency Crime Type which Male victims faced
- 'INTIMATE PARTNER - SIMPLE ASSAULT' is the highest frequency Crime Type which Female victims faced
- 'BATTERY - SIMPLE ASSAULT' is common among both genders
- There are some differences in the ranking of these top crimes between genders, but the overall types of crimes experienced are relatively similar

Predicting Future Trends:



- The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots provide insights into the order of the ARIMA model:
- The ACF plot shows a gradual decline, suggesting a moving average (MA) component in the data.
- The PACF plot shows a sharp drop after the thirty lag, suggesting that the autoregressive (AR) component is of order 30.
- Based on these observations, we can start with an initial ARIMA model with parameters $p=30$, $d=0$ (since the series is already stationary), and $q=30$



- The code fits an ARIMA model with a complex order of (30, 0, 30) to a time series dataset of crimes.
- It forecasts the next 90 days of crime data and calculates the Mean Squared Error (MSE) on the validation set, resulting in an MSE of approximately 5034.67.
- The code generates a plot that shows the training data in blue, the actual validation data in green, and the ARIMA model's forecast in red.

R-squared (R^2) score of the model is 55.73% which means that the ARIMA model can explain or predict about 55.73% of the differences or changes that appear in the 'y' variable. So, it's doing a decent job