INTRODUCTION

Crime is something that many people worry about on a daily basis. Whether it's ensuring your door is locked when you leave the house, avoiding a rough Neighbourhood or be Vigiliant or installing a security alarm, prevention of crime takes up a significant part of our lives.

The Most Important aspect of our lives is 'Safety'. The Reason why i got my interest in this Dataset is because i want to analyze the Crime trends, victims, timings, and the trends per Time.

This dataset reflects incidents of crime in the City of Los Angeles dating back to 2020. This data is transcribed from original crime reports that are typed on paper. There are multiple attributes in this dataset with which we can analyze, manipulate and get insights from them.

Research Question

Before Starting our work it is Important to frame a question or a series of questions which you are willing to Analyze from the Dataset. Which could give us some Meaningful Insights from the Raw Data.

"How do demographic factors, including gender and age, interact with temporal trends in crime rates and the correlation between crime types and the time of day, contributing to a comprehensive understanding of crime patterns in the selected region, and what implications do these insights have for law enforcement and public safety policies?"

Data Summary

- This Dataset has been picked from DATA.GOV Organisation https://catalog.data.gov/dataset/crime-data-from-2020-to-present.

- This Dataset consists of 820599 Rows(Use Cases) and 28 Columns before cleaning.

- The Most important columns to be focused on are Time period, Age of Victim, Location ,Gender etc.

- By Looking at the Data we can see different types of crimes which were reported in the city at different locations and at different times.

- By Roughly Looking at the Data we will start to understand like which are the possible factors that help
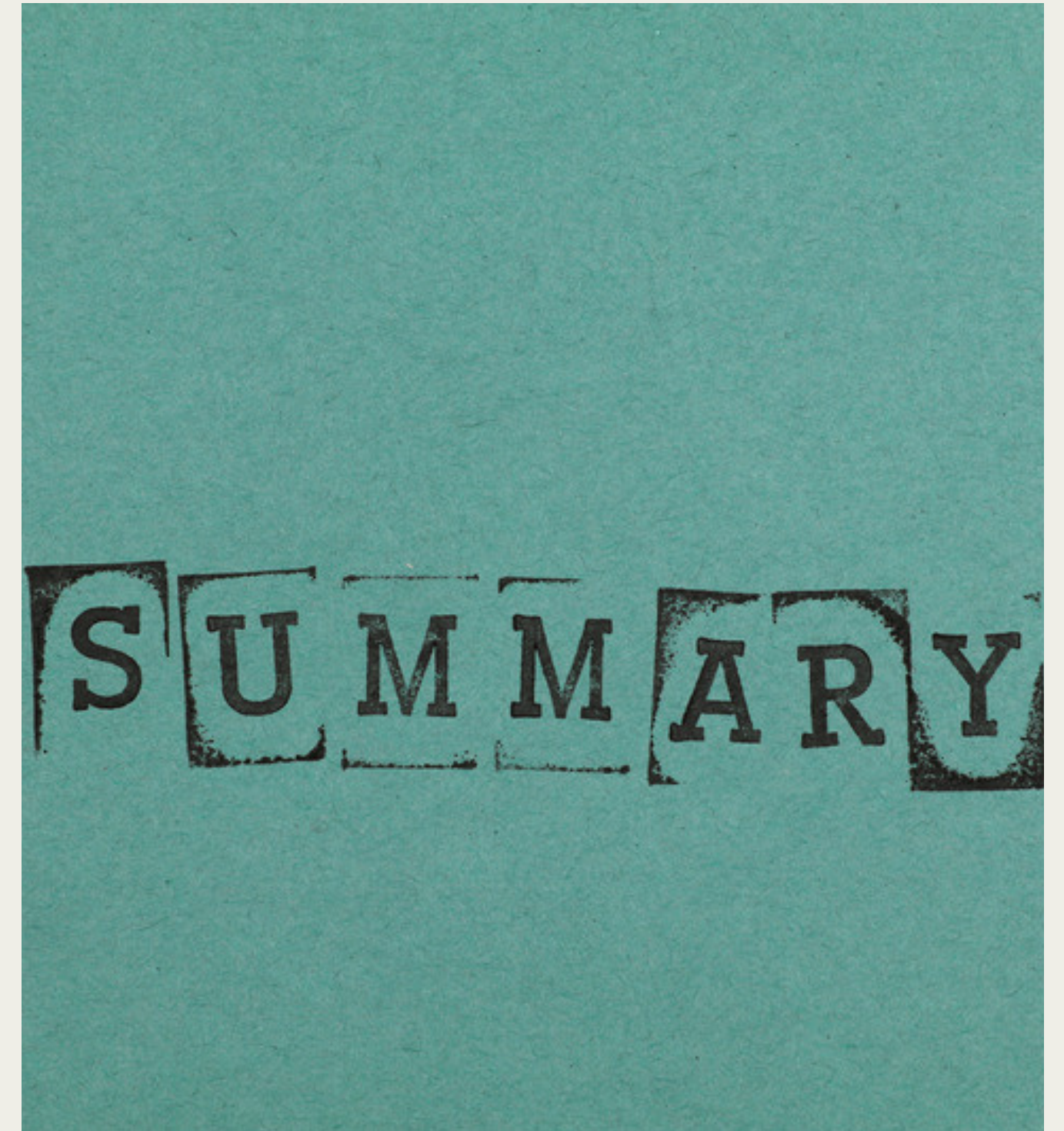
us Understand and Analyze the rough estimate and summary of the Data.

- This Dataset is categorized by 28 columns which specify each attribute of the dataset either categorical or Numeric based on the report lodged on the file.

- We can see some Missing Values in the dataset which should be cleaned before we use the Dataset as these may cause Misinterpretation of the Data.

- There are some Columns which are not relevant to our analysis which should also be ignored just to minimize the columns which can make our work more meaningful.

- And it is very important to understand the Column Names and what those identify as, because it will only make sense then on what we are trying to correlate.

Interpreting Columns

DR_NO: This column likely represents a unique identifier for each incident or report.

Date Rptd: This column appears to represent the date when the incident was reported.

DATE OCC: This column seems to represent the date when the incident occurred.

TIME OCC: This column appears to represent the time when the incident occurred.

AREA: This column might represent a numerical code or identifier for a specific area.

AREA NAME: This column lik represents the name or description of the area.

Rpt Dist No: This column might represent a numerical code or identifier for the reporting district.

Part 1-2: This column's meaning is not clear without additional context.

Crm Cd: This column represents a numerical code for the crime.

Crm Cd Desc: This column contains the description of the crime corresponding to the Crm Cd.

Mocodes: This column might contain additional codes or descriptions related to the mode of operation for the crime.

Premis Desc: This column contains the description of the premise corresponding to the Premis Cd.

Weapon Used Cd: This column represents a numerical code for the weapon used in the crime.

Weapon Desc: This column contains the description of the weapon corresponding to the Weapon Used Cd.

Status: This column likely represents the status of the incident or case.

Status Desc: This column contains the description of the status corresponding to the Status.

Crm Cd 1, Crm Cd 2, Crm Cd 3, Crm Cd 4: These columns might represent additional crime codes or identifiers.

LOCATION: This column contains the location or address of the incident.

Cross Street: This column might contain information about a nearby cross street or location.

LAT and LON: These columns likely represent latitude and longitude coordinates related to the incident location

Vict Age: This column represents the age of the victim.

Vict Sex: This column represents the gender of the victim.

Vict Descent: This column likely represents the ethnicity or descent of the victim.

Premis Cd: This column represents a numerical code for the premise where the crime occurred.

.

DATA CLEANING

Before the data is processed, it undergoes a "cleaning". This is when anomalies in the data are removed, it is transformed into a usable format, and organized to make it more useful. Without data cleaning, data analysts can have difficulty making sense of the data or interpreting the results correctly. It is important because it ensures consistency within your data set and helps you achieve reliable results from any analysis you perform on it. Additionally, regularly checking for inconsistencies allows you to identify problems in your data sets before they become bigger issues down the line.
Finally, when you clean your data properly, data analysis becomes smoother and More Efficient

Data identification - After loading the Data into the variable 'df',we identified the data's columns, rows and with shape functionality. After that we can look into the data columns by the functions named as info(),or Describe() to get familiarized with the Dataset.

- To Monitor the Data carefully we used the head() function and Tail() function to know the First5 and Last 5 Rows in the Dataset

- **Finding Null Values -** After we took the dataset we identified the Null values according to the percentages we were shown.

- **Dropping The Values -** After finding the Null Values we removed the columns which had more than 60%, We can either drop the rows where missing values are present or replace the missing values with some values like mean, median, or mode.

- **Finding The Missing Values -** After Removing the Null values from columns we looked for Missing values in the Rows

- **Removing The Missing Values -** After we identified the Missing values we made some operations to remove them and make data clean without any errors

- **Checking The Obtained Data -**After All the cleaning process it is important to check your Data if it is as desired.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an essential process in data analysis that involves examining and visualizing a dataset to summarize its main characteristics, often with the goal of gaining a deeper understanding of the data, identifying patterns, and formulating hypotheses.

Step 1- To Begin with we should be keen in Summarizing the data. We took Two columns which are Numeric to find if there are any outliers present in it. And we made a Box plot of Two columns which are 'TIME OCC', and 'Vict Age' using the Seaborn package The Box Plots determine the two columns vict age on the left and Time OCC on the Right.

- in the Vict Age we can see that there are victims who are aged between 20 and 50 and there are outliers too which is the old people who are aged above 90 and the Median is somewhere around 30.
- On Your Right hand side you have Time OCC which determines the time of crime happened which is somewhere between 10 am and 7 pm and the Median is somewhere around 2 pm it might be a bit confusing as the graph shows 24 hour Time.
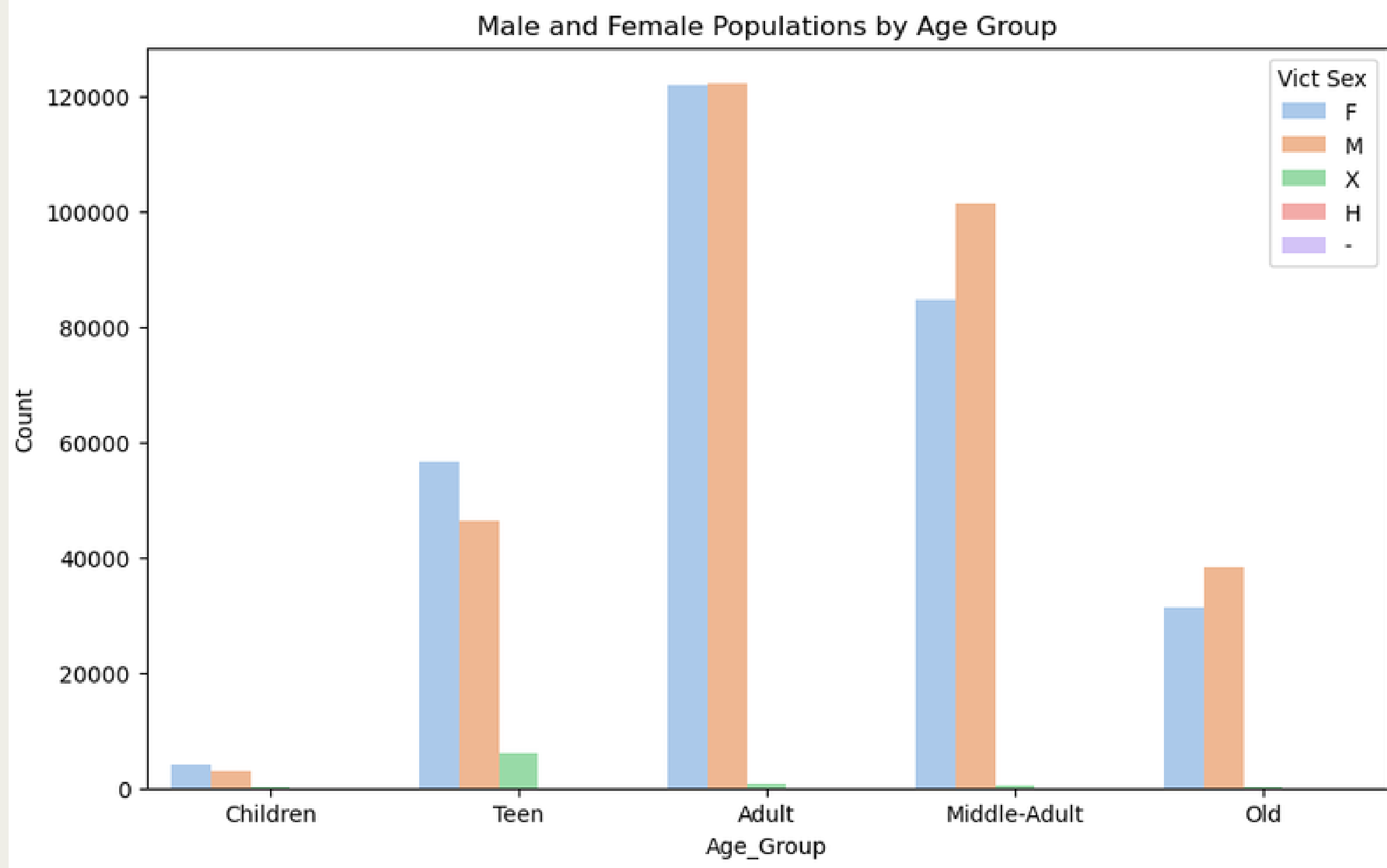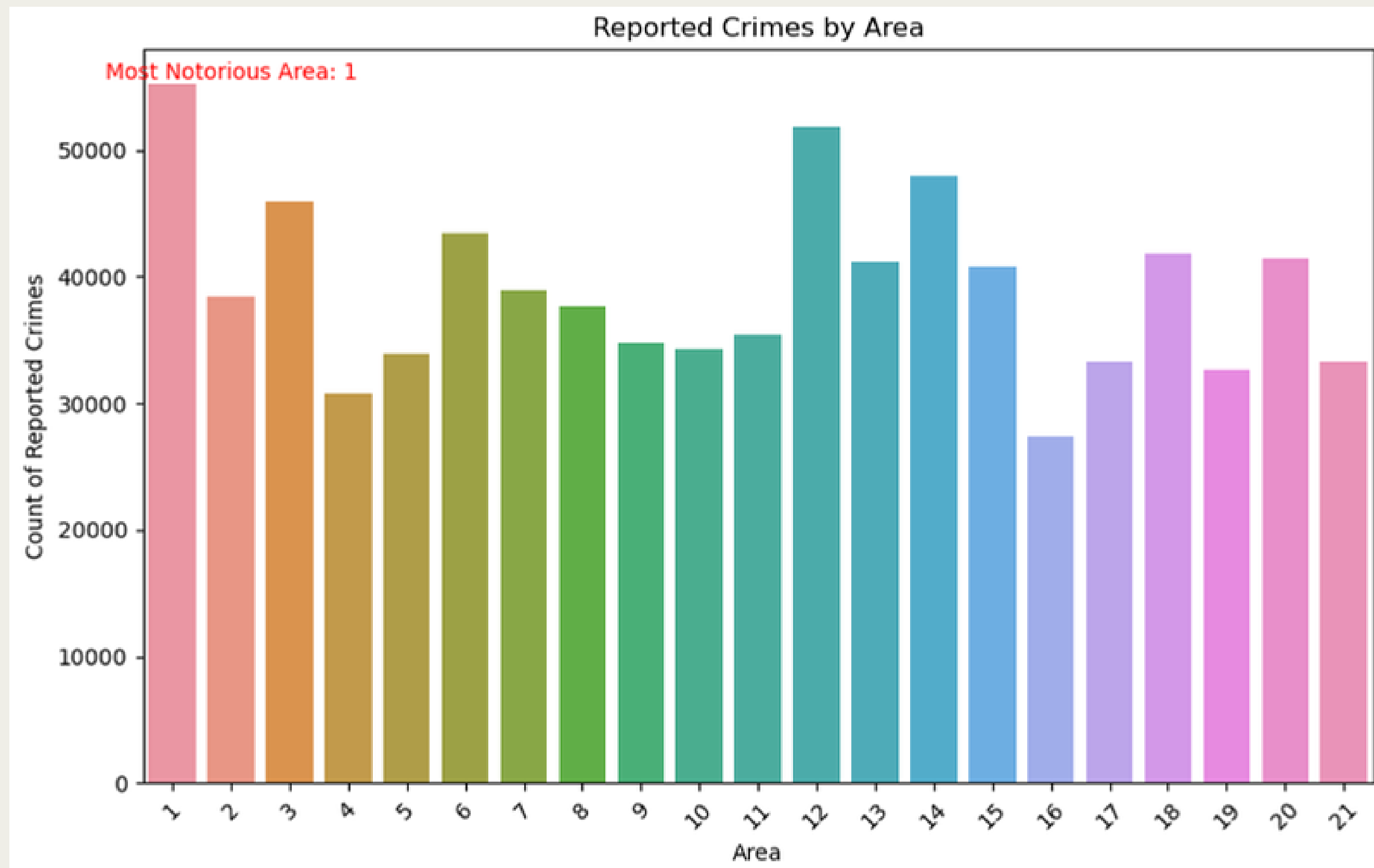
BarPlot

Plot A

Plot B
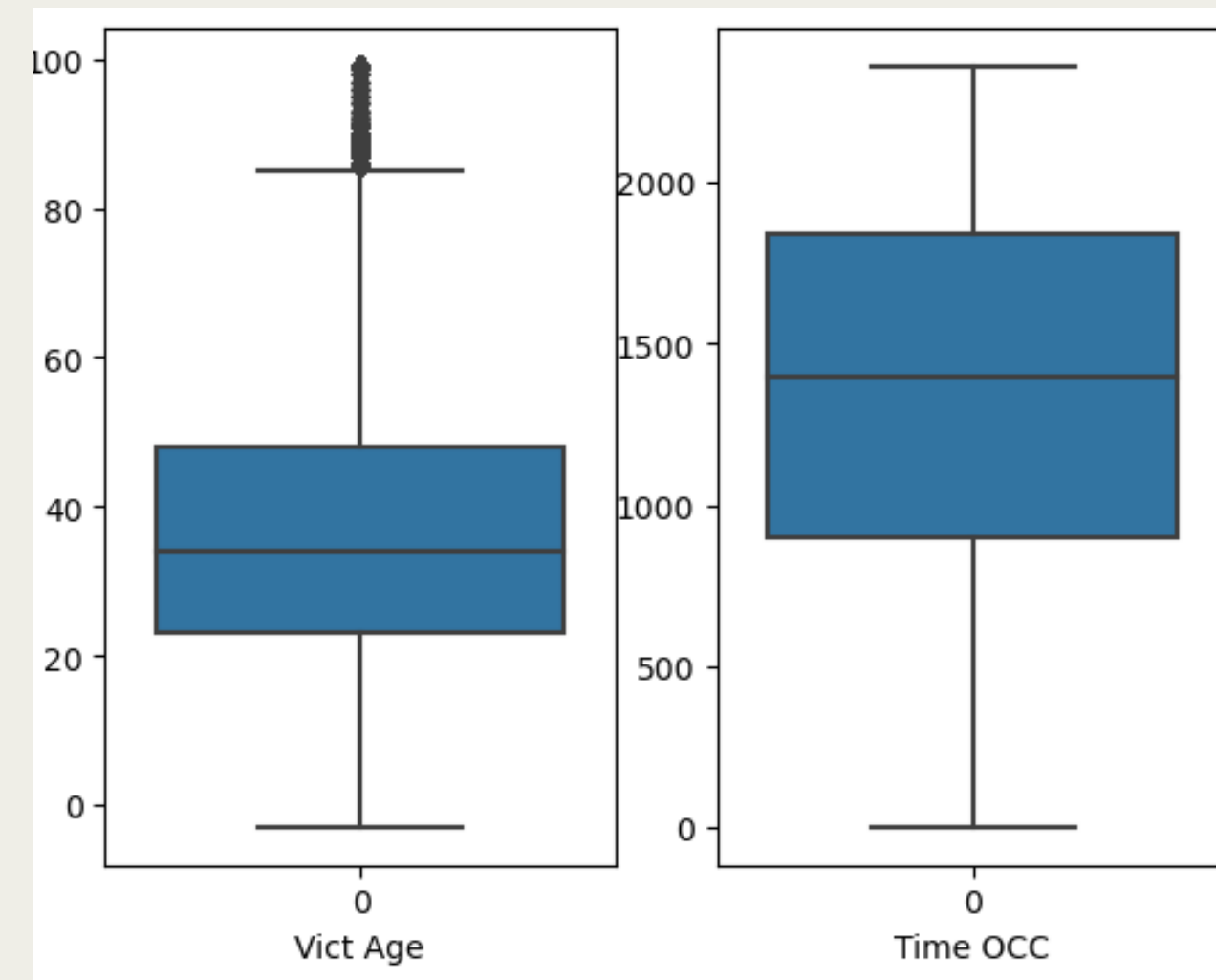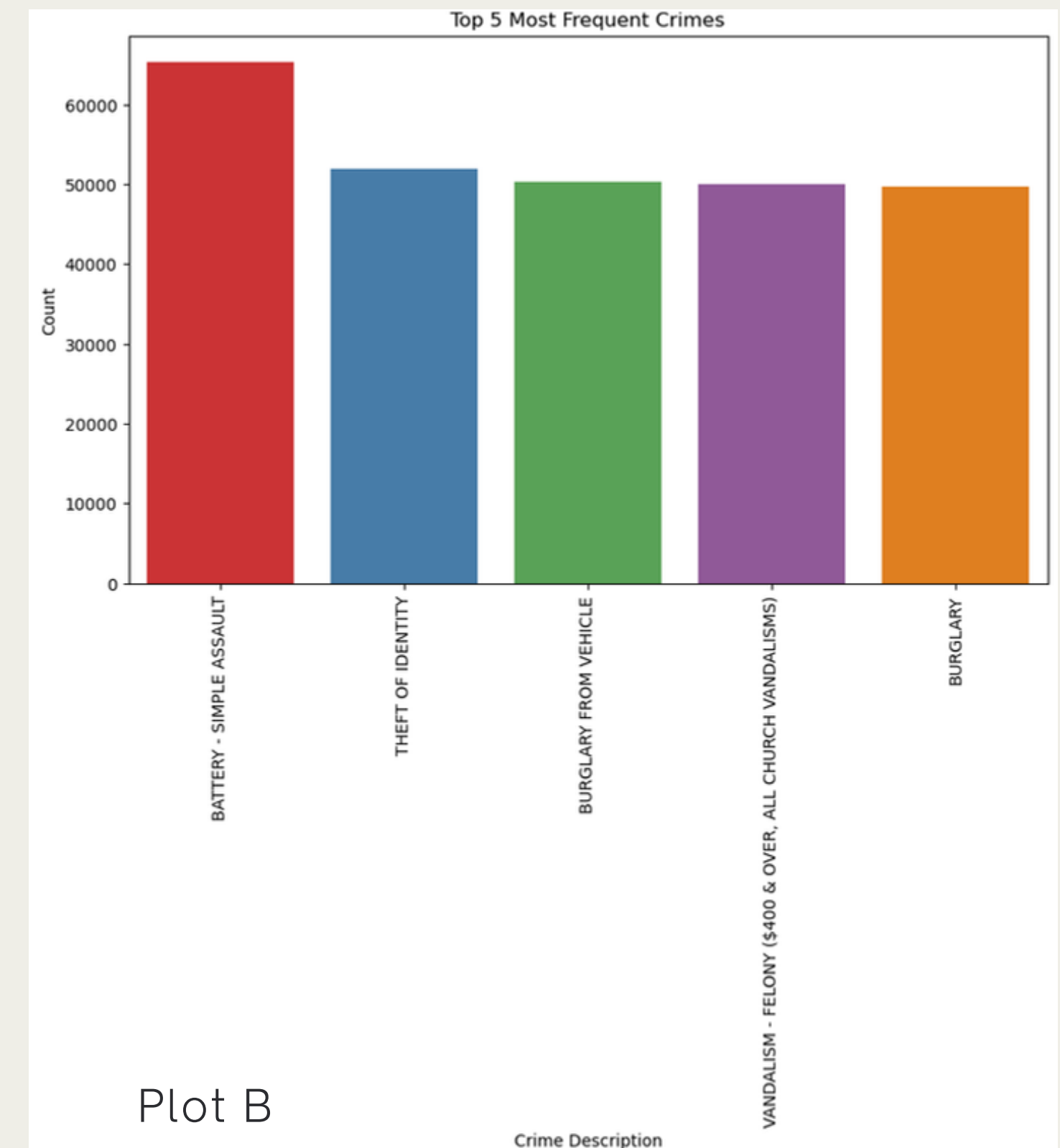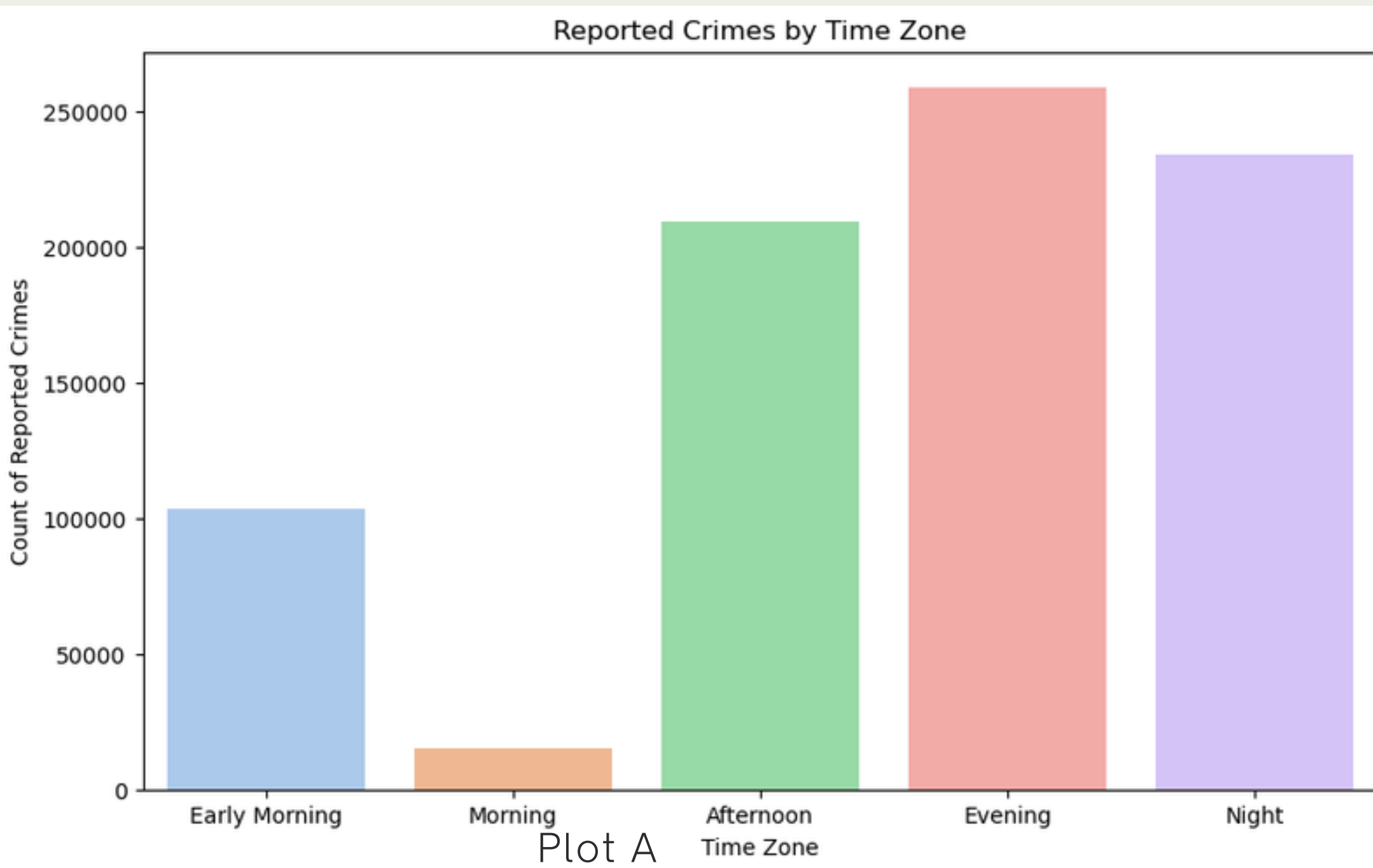
Plot A-The 24 hour time zone was made into 4 different timings named as Early Morning, Morning, Afternoon, Evening and Night by Binning the columns. The Above Table shows the Crime rates count corresponding to The Time of the Crime on a particular day.

Plot B- In this we can see that the graph is between the Most frequent Crimes, this plot is against the count of crimes vs specific types of Crimes displayed

Above Bar-Plot is the Male and female populations by Age group. In which we can clearly see that the victims are mostly in the Adult age group who have been attacked the Most and children being the least attacked ones and the count of the Adult age group is somewhere above 12000 as per the data provided.

Above Bar-Plot shows the most Notorious Areas where the Occurence of crime is Higher this graph relates between the count of the reported crimes on the Y-Axis and the Area Codes corresponding on the X axis.
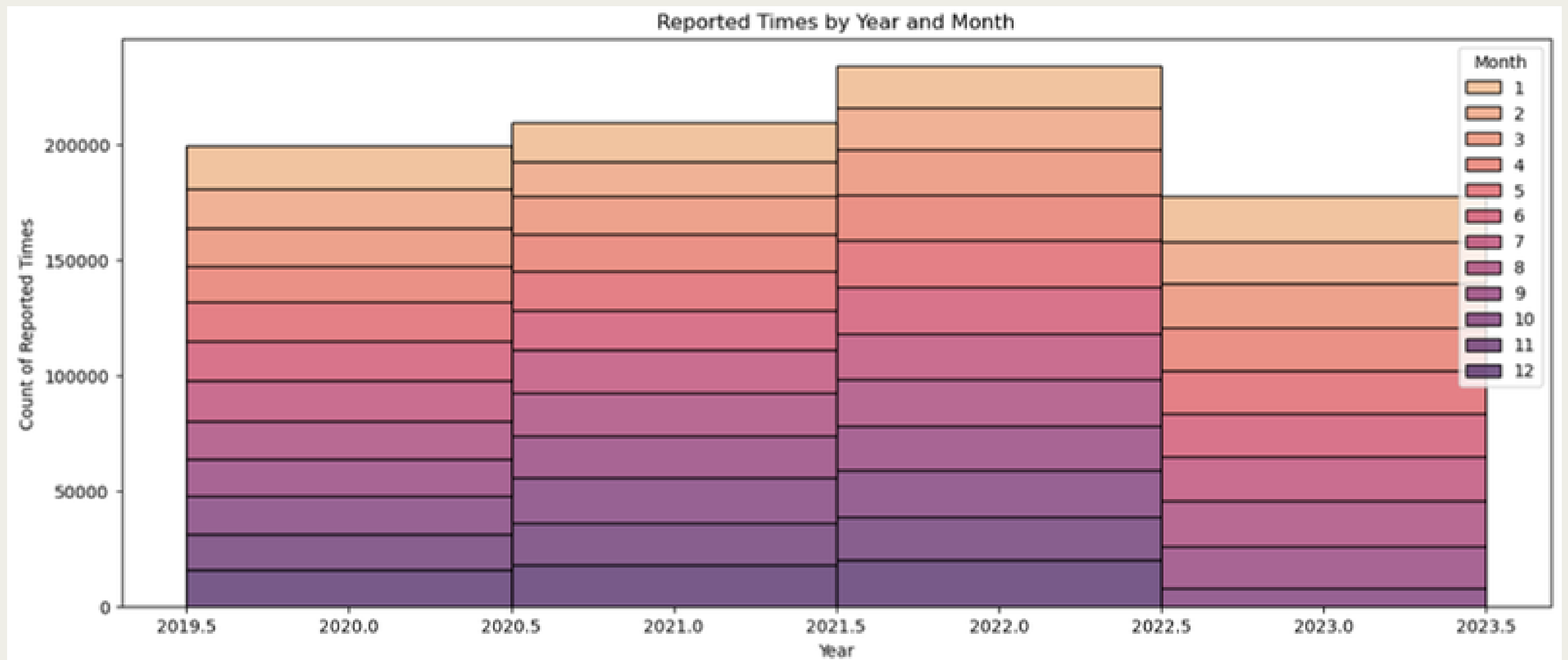
Above Plot shows the Reported Times by Year and Month , So the year 2022 shows the count of High reported crimes times
There is count of the reported times on the Y-Axis and the corresponding years on the X-Axis.

Reported Times by Year and Month (Line Plot)
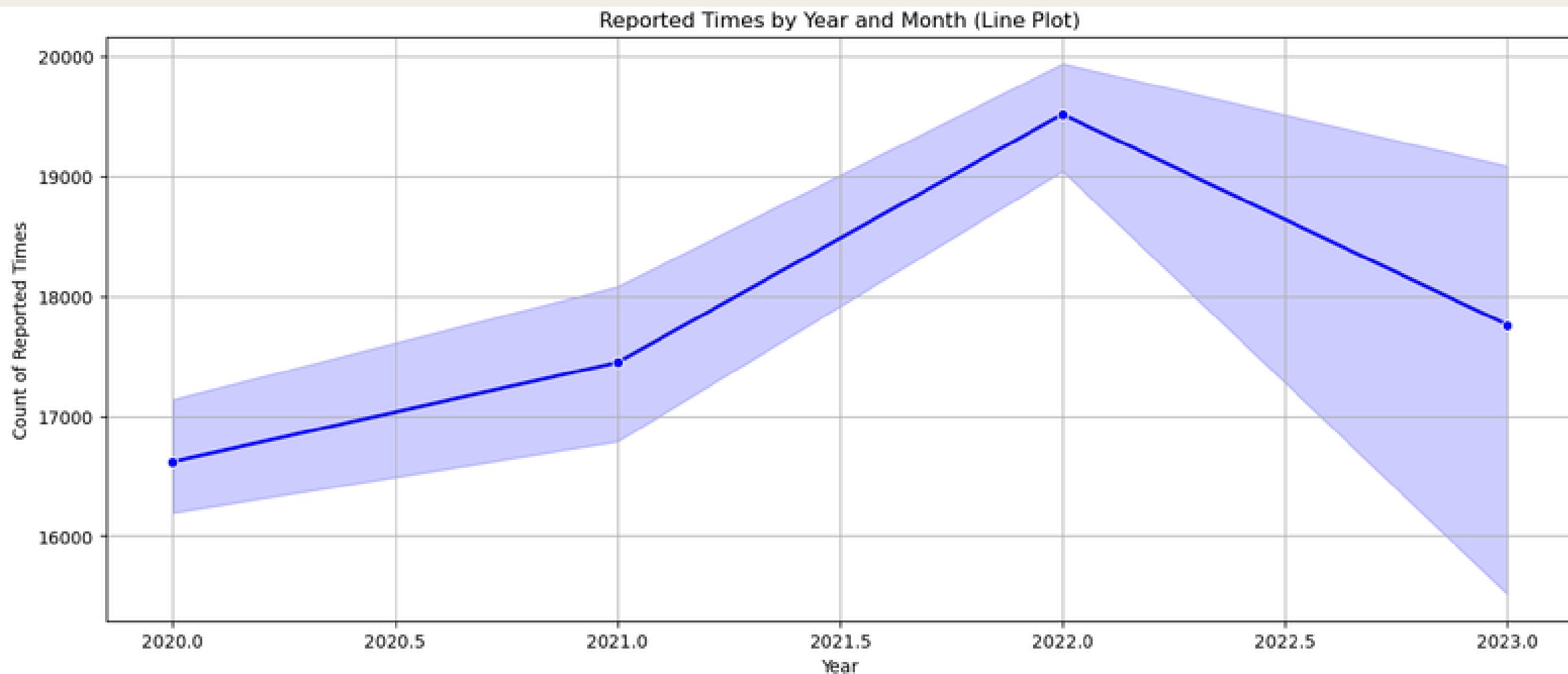
These two plots are the same. The corresponding plot is made by using the seaborn library . This is made against count of Crime reported and the Year it happened

The corresponding plot is made by using the MatplotLib library . This is made against count of Crime reported and the Year it happened

Reported Times by Year and Month (Line Plot)
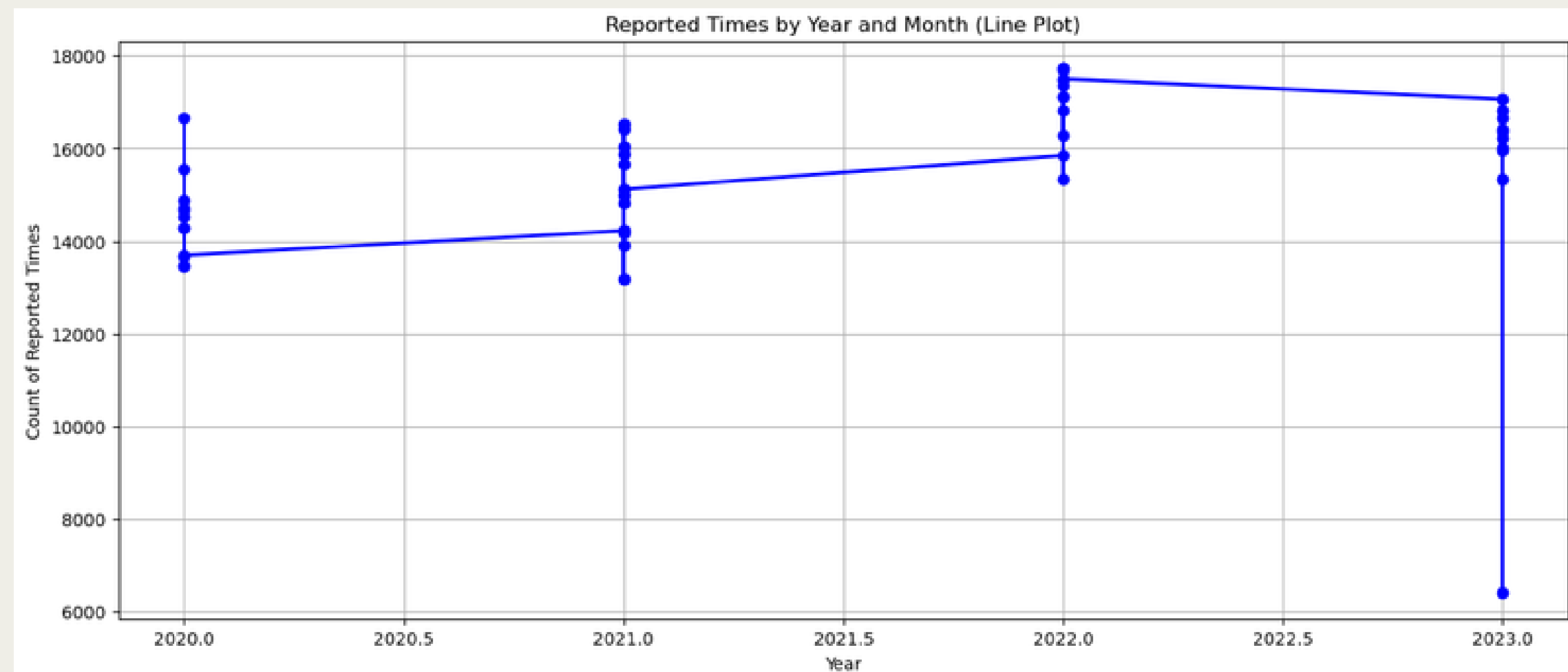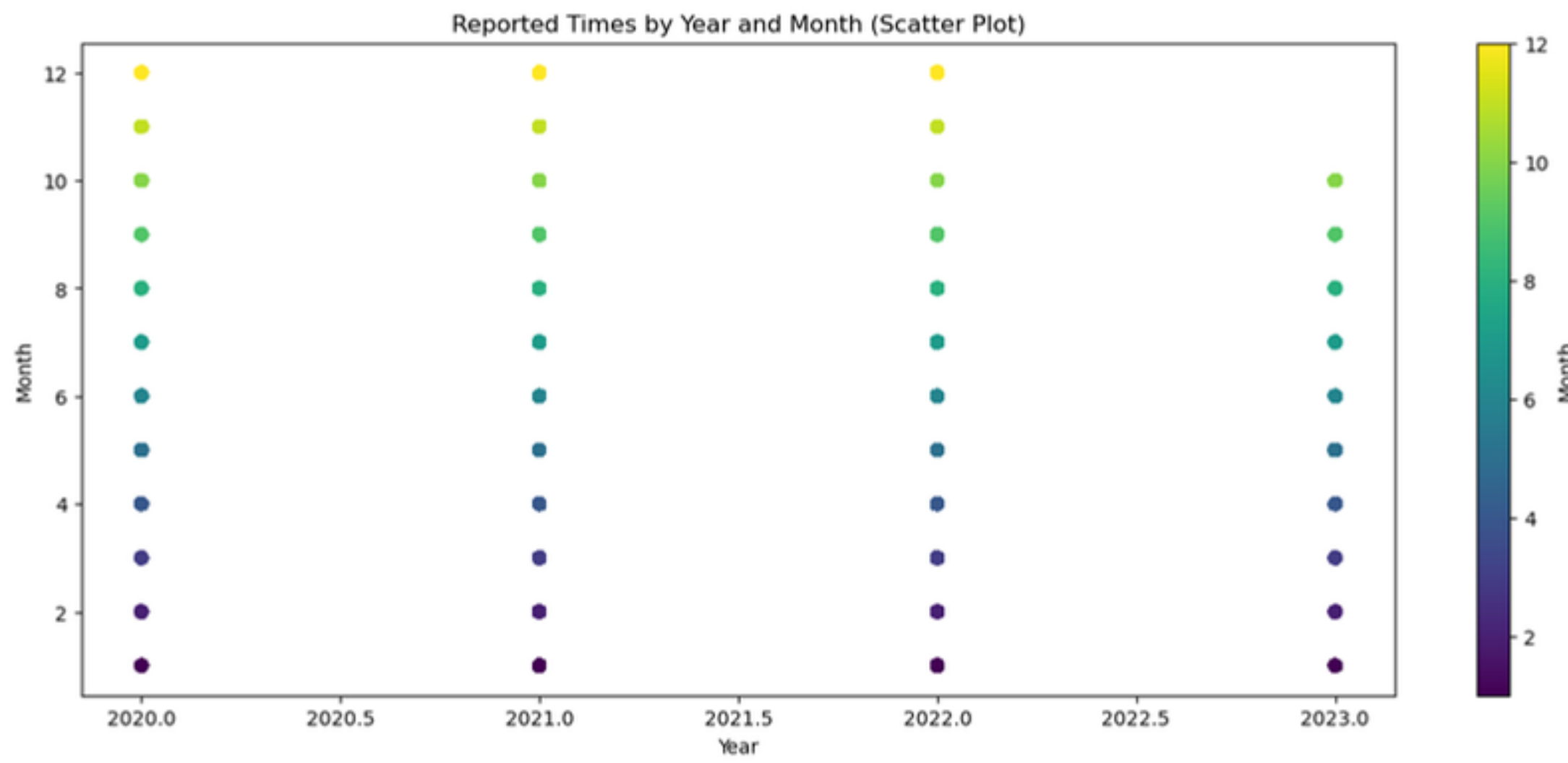
These two Scatter plots are the same. The corresponding plot is made by using the seaborn library on the Top and Matplotlib in the bottom .

Correlation

Reported Crimes by Time Zone and Crime Description
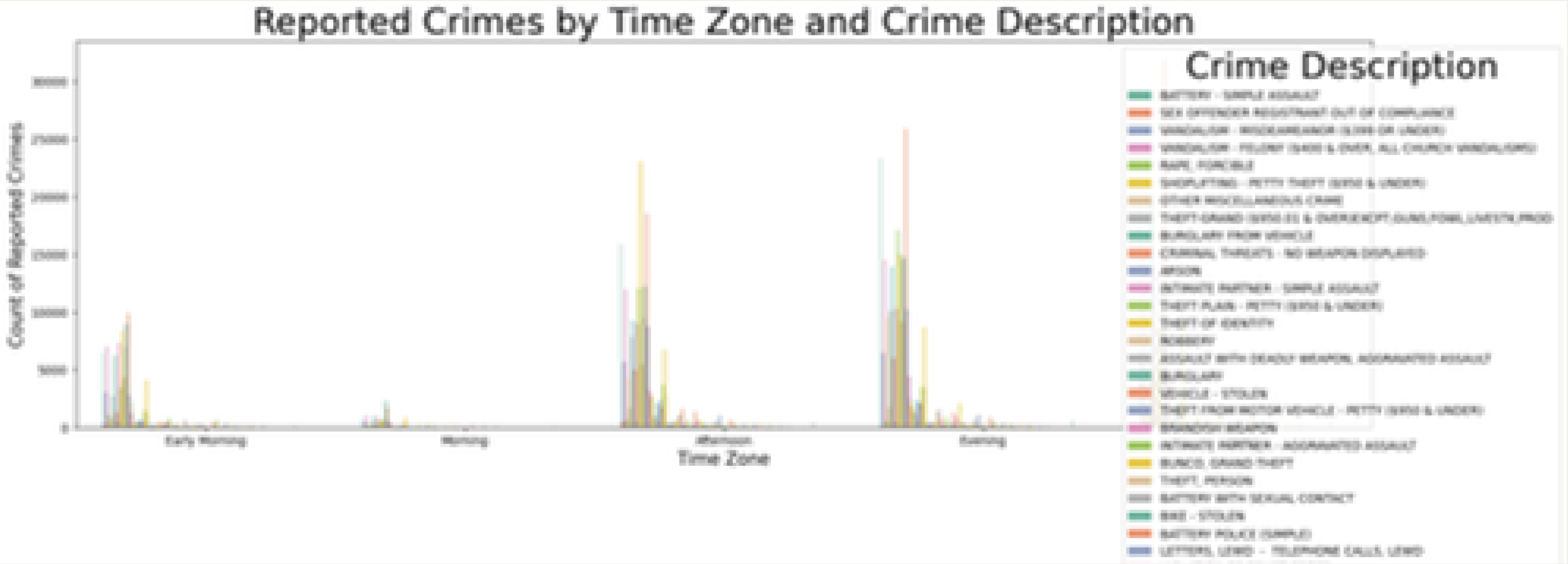
The Above plot is a graphical representation of the Crimes reported by Time zone and Crime description ,This is done to find out which type of crime has occured the most and likely to happen in the part of the day.

Correlation
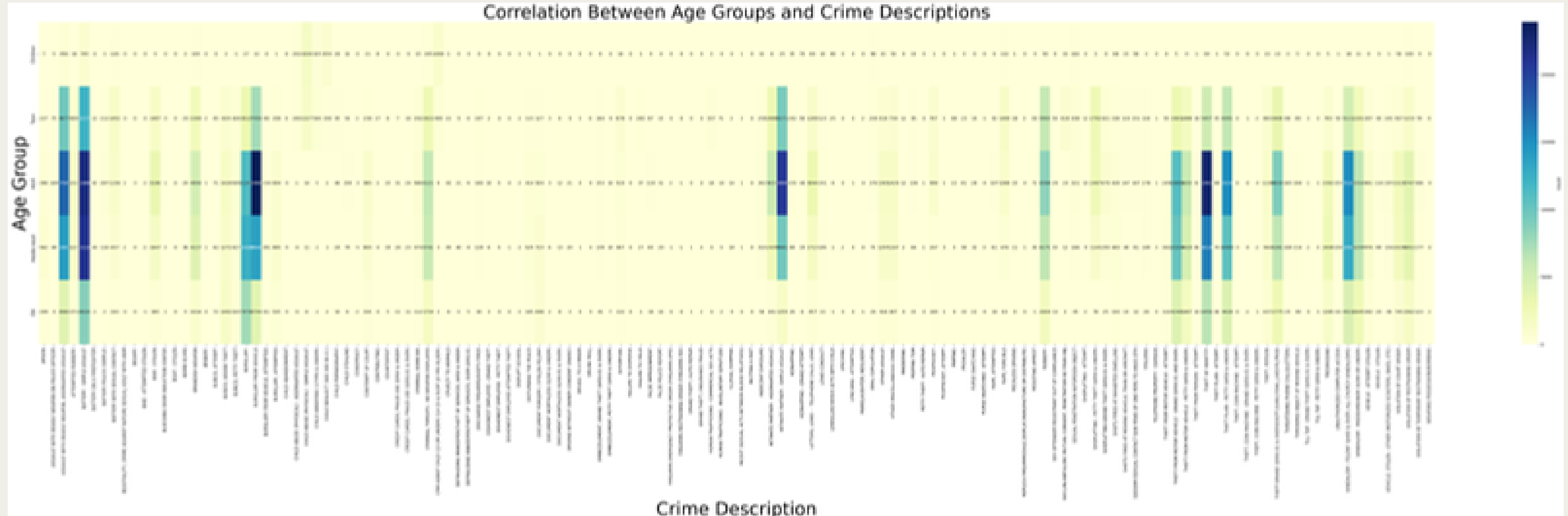
The Above plot is a graphical representation of the Crimes reported as a correlation between Age groups and Crime Descriptions,With this we can learn the correlation between the Age_Group and Crime Description .

Inference Analysis

Gender and Age-Related Patterns in Victimization: The analysis revealed distinct gender-related patterns in victimization across different crime categories. While both males and females were victims of various crimes, there were variations in the prevalence and types of crimes experienced. Notably, the data indicated that the Females in Child and Teen group are the most victimized, In Old and Middle-Adult population the Male population got victimized the most However the Adult population group has victimized the most with equal Male and Female Populations to certain crimes.

A thorough examination of age-related patterns unveiled substantial variations in victimization by age group.Certain age groups were more susceptible to specific crime categories, suggesting that age plays a crucial role in influencing the likelihood of being a victim of particular crimes.

Temporal Trends in Crime Rates: Over the period from 2020 to the present, temporal trends in crime rates were discerned. These trends shed light on how crime rates have fluctuated over time. The analysis demonstrated that crime rates exhibited both short-term fluctuations and longer-term trends, emphasizing the significance of considering time when assessing crime patterns. In 2020 the count of crimes which were reported could be seen around 200000 which further kept on increasing until 2022 which recorded the highest crimes in the city that crossed above 200000 limit and were marked as the Highest, Later due to some protective measures the crime decreased the following Year.

Correlation Between Crime Types and Time of Day:
 An exploration of the relationship between crime types and the time of day they are most likely to occur unveiled valuable insights. Different crime categories exhibited varying temporal patterns, highlighting the influence of time on the commission of specific crimes. This finding is crucial for law enforcement and policymakers in optimizing resource allocation and crime prevention strategies.In our data, we can witness that the crime rate in the evening is higher than the crime rate at any time of the day. At Night and Evenings, the most prevalent CRIME is Vehicle Theft. We have gathered all the Data Regarding all the Thefts that are being reported all the time in the day wise.

Correlation Between Age Groups and Crime Description:
 Correlations between age groups and crime descriptions are essential for understanding patterns of victimization within different age groups and the types of crimes they are more likely to experience. This analysis can help law enforcement, policymakers, and researchers tailor crime prevention and intervention strategies to specific demographics. In our dataset, we can see that Most of the crimes committed are in the Age group of Adults. and crimes like Bulgary in the Office, Battery simple assault, and Theft of Identity are prevalent.
Demographic factors, such as age and gender, significantly influence crime patterns. The analysis highlights that specific crime types exhibit distinct temporal trends, with higher crime rates during the evening and night, especially in vehicle theft. Furthermore, age plays a pivotal role, as adults are more susceptible to crimes like burglary battery, and simple assault. Gender-related patterns show child and teen females as the most victimized, while middle and old adult males face higher victimization. Temporal trends underscore the impact of external factors on crime rates, with an increase in 2022 and a decrease in the following year. These insights are pivotal for law enforcement and public safety policies, aiding in resource allocation, tailored interventions, and enhanced community safety.

Research Result Analysis

Demographic factors, including age and gender, exhibit profound influences on the intricate tapestry of crime patterns in the selected region. Our analysis unraveled distinct gender-related patterns in victimization across a spectrum of crime categories. While both males and females encountered diverse criminal experiences, intriguing variations emerged in terms of prevalence and crime types. Notably, the data pinpointed that females in the child and teen age group were the most vulnerable, while middle and old adult males faced the highest victimization rates. Interestingly, adults, in particular, bore the brunt, with a near-equitable distribution of male and female populations falling victim to certain crimes.

Delving deeper into the examination of age-related patterns, we unveiled substantial disparities in victimization, underlining the pivotal role of age in influencing one's likelihood of being a crime victim. Specific age groups exhibited pronounced susceptibilities to particular crime categories, marking a significant departure from the uniformity one might expect.

Temporal trends in crime rates further enriched our understanding of the evolving landscape. Our analysis scrutinized the period spanning from 2020 to the present, revealing the intricate dynamics of crime rate fluctuations over time. The data painted a compelling picture, showing a surge in reported crimes reaching around 200,000 in 2020, and this crescendo continued, culminating in a historic high in 2022, exceeding the 200,000 threshold. Subsequently, the imposition of protective measures prompted a decline in crime rates in the following year.

The correlation between crime types and the time of day illuminated invaluable insights. Various crime categories demonstrated divergent temporal patterns, underscoring the temporal dimension's profound influence on specific criminal activities. Most notably, we discerned that the evening harbored a heightened crime rate, with nighttime being particularly conducive to the occurrence of vehicle thefts. Comprehensive data on thefts reported throughout the day further bolstered our findings, shedding light on these temporal nuances.

Overall, these revelations hold significant implications for law enforcement and public safety policies. They provide a robust foundation for optimizing resource allocation, tailoring interventions, and fostering enhanced community safety. The interplay of age, gender, and time in shaping crime patterns is a multidimensional puzzle, and our insights contribute to a more informed and strategic approach to crime prevention and management in the region

Conclusion

This comprehensive analysis explored various aspects of crime data spanning from 2020 to the present, focusing on gender-related patterns in victimization, age-related patterns in victimization, temporal trends in crime rates, and the correlation between crime types and the time of day. The investigation yielded significant insights, providing answers to the research question posed.

In conclusion, this analysis underscores the multifaceted nature of crime data and the importance of considering gender, age, temporal factors, and crime types when formulating crime prevention and intervention strategies. The findings from this study provide a foundation for informed decision-making, enabling law enforcement agencies and policymakers to better understand and address crime-related challenges in the selected region. Additionally, this analysis demonstrates the value of data-driven insights in enhancing public safety and Promoting effective Crime Management.

Reference and Citation

How To Conduct Exploratory Data Analysis in 6 Steps

Learn about what exploratory data analysis is, the steps you can follow to conduct it and some of the major...

indeed.com

Medium
Sign up

Medium

What is Data Cleaning? Unfold It's Process, Importance & Benefits
Even before getting to the analyzing stage, data cleaning takes precedence, & with good reason. Find

Aquibgitt/
A.P_Project_1

1	0	0	0
Contributor	Issues	Stars	Forks

Aquibgitt/A.P_Project_1

Contribute to Aquibgitt/A.P_Project_1 development by creating an account on GitHub.

GitHub

Github Repository Link.

https://cde.ucr.cjis.gov

All these are the websites which i refered and took help from to understand the relativity. ,Also Took some Help from AI chatbots for Complex Queries