

Los Angeles Crime Data Analysis 2020 to Present



IE6400 – Foundations of Data Analytics Eng

Final Report

Group 2

Jeffrey Johnston

Sanidhya Karnik

Himanshu Randad

Digvijay Ashish Raut

John Ayush Xavier

Introduction

In an ever-evolving world, the analysis of crime data plays a pivotal role in understanding the complex dynamics of public safety. Criminal activity affects the lives of individuals, communities, and societies at large, making it imperative to investigate and decipher the underlying patterns, trends, and factors that contribute to crime rates. This report presents a comprehensive analysis of real-world crime data spanning from the year 2020 to the present, aimed at shedding light on the intricacies of criminal behaviour and the influences that shape our communities.

Crime data analysis has far-reaching implications, not only for law enforcement agencies but also for policymakers, community leaders, and citizens. By examining the intricate web of information within crime datasets, we seek to uncover valuable insights that can inform crime prevention strategies, resource allocation, and the promotion of safer environments for all. Moreover, crime data analysis benefits citizens directly by helping them make informed decisions about where to live, work, and raise their families. Accessible crime data empowers individuals to take precautions, advocate for safer neighborhoods, and engage in community initiatives aimed at reducing crime. It fosters transparency and accountability, enabling citizens to hold both law enforcement agencies and policymakers responsible for their commitment to public safety.

The importance of crime data analysis goes beyond law enforcement to include policymakers who look for empirically supported ways to improve public safety. Policymakers can create laws and policies that address the underlying causes of crime and its different manifestations by analyzing the data. Community leaders may also use this analysis to gain a deeper understanding of the difficulties that their neighborhoods face and collaborate with law enforcement and locals to create solutions that are tailored to their needs.

This report is structured to guide readers through a journey of data exploration, analysis, and interpretation. We will begin by addressing the fundamental task of data cleaning, ensuring that the dataset is free of errors and inconsistencies. Subsequently, we will embark on an exploratory data analysis (EDA) journey, using descriptive statistics and data visualization techniques to gain a deep understanding of the dataset.

Data Cleaning

Our dataset was downloaded from [Crime Data from 2020 to Present - Catalog](#), which gave us a .csv (comma separated values) file containing data representing crimes that took place in the City of Los Angeles from 2020 to present, which was 02 October 2023 at the time of download.

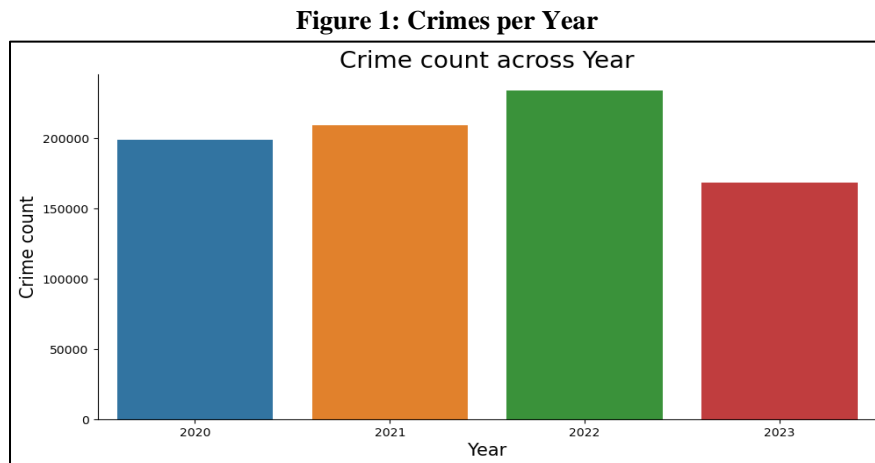
After displaying the data as a Dataframe in Python we identified our columns and data types, as well as identified that the data set contained many missing and unsorted values. To overcome any uncertainty and confusion with the missing values we began our data cleaning and used different Python functions like drop duplicates and sorted the missing values into the proper format. We standardized the Time Occurred column by converting all the values into a Time format. Next, we identified the Mocodes column was originally in space-separated values format, we converted them into a readable list of codes. We also identified that there were Crime Codes in multiple columns, so we collated all the crime codes into a list and deleted the redundant individual columns. We also joined the Location and Street columns together which gave a meaningful address and dropped the respective individual columns. There were many entries in the Victim Age column where the victim age was not specified so we found the median of the ages of the victims based on the type of crime that occurred to fill the empty values.

For the Premis Code column and Premis Description, we filled the empty values with the mode values based on the type of crime. We converted the Date Reported and Date Occurred columns to a standardized date-time format. Additionally, we created new columns for Day, Month, Year and Day of Week to help identify trends across them individually. Upon conclusion of the data cleaning, we were able to interpret and visualize the data in many ways allowing us to answer questions and analyse trends regarding crimes in Los Angeles.

Results and Findings

- **Overall Crime Trends:**

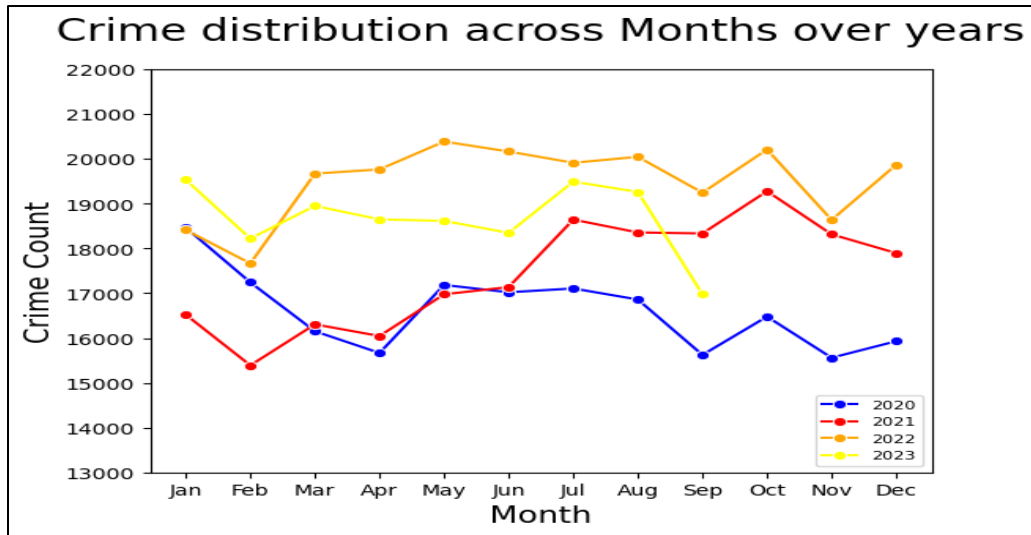
By plotting the total crimes over the years, we determined that the crime rate is increasing at a rate of 8.3% annually. The data for 2023 is skewed because the dataset is incomplete for the months of October, November and December.



- **Seasonal Patterns:**

To determine seasonal patterns, we grouped the data by month and plotted total crimes by month over the years as well as average crimes per month. We determined that in general, the crime rate drops in September after the summer months then increases slightly in October before it drops again in November prior to the winter holidays. After New Years the crime rate tends to increase in January and then drop in February before it begins to rise again for the summer.

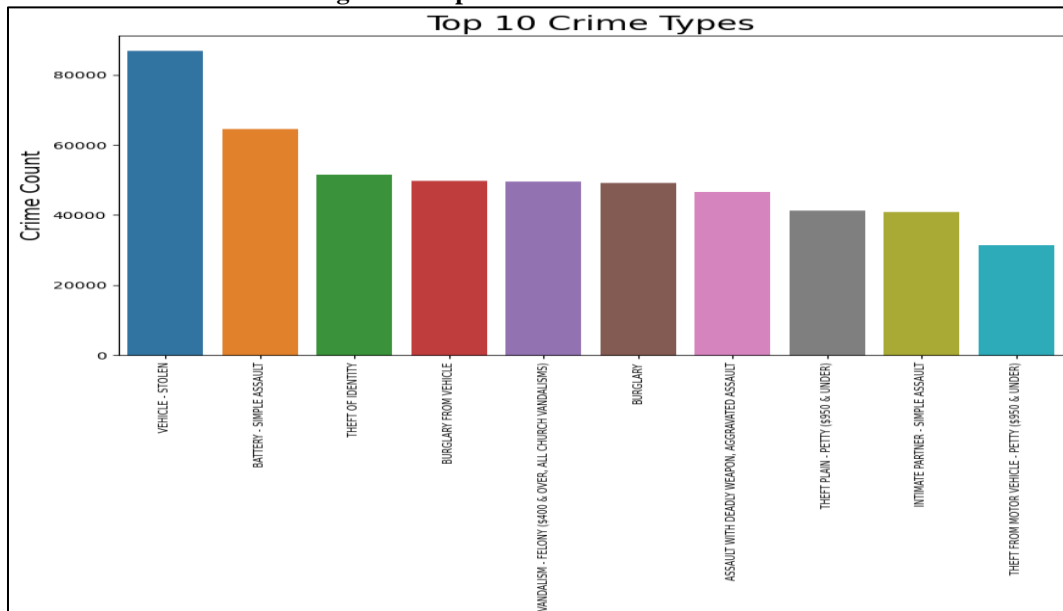
Figure 2: Crimes by Month per Year



- **Most Common Crime Type:**

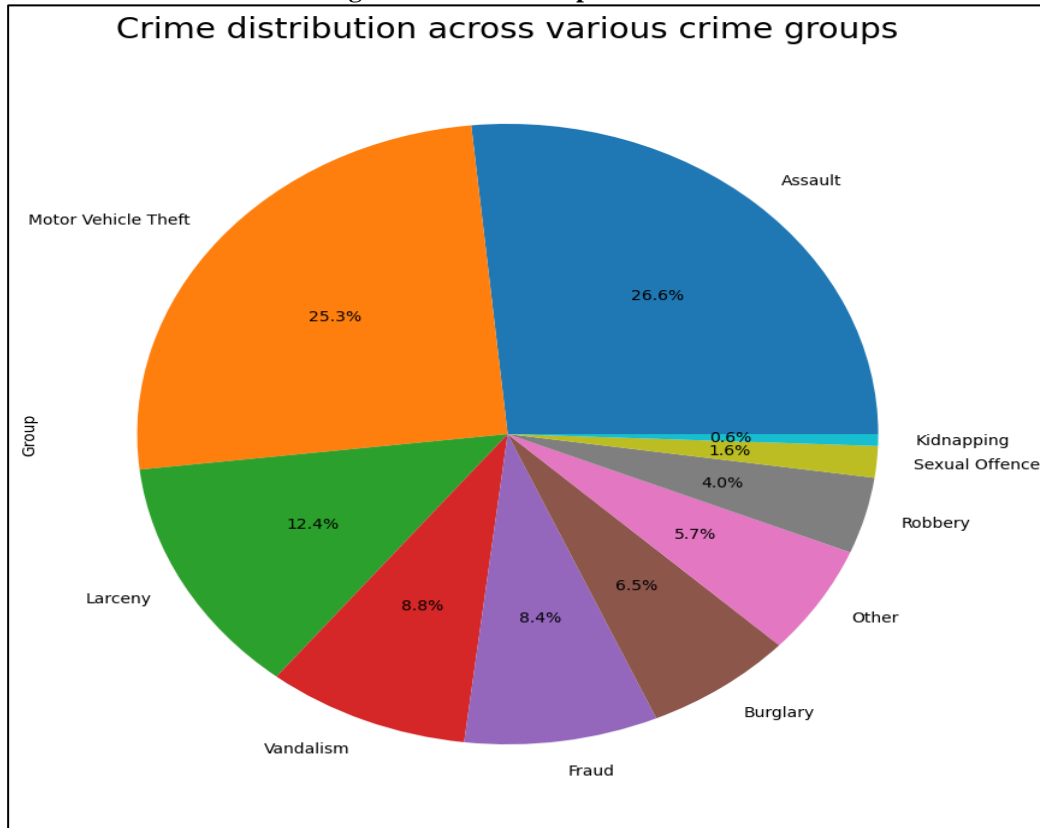
After analysing crime types, we were able to determine that the most common crime is Stolen vehicle with 86,335 occurrences followed by Battery simple assault with 64,205 occurrences and Identity theft with 51,327 occurrences. The least occurring crime was Inciting a riot with only one occurrence.

Figure 3: Top 10 Most Common Crimes



We decided to further analyse the crime types and grouped them from the 138 original crime types into 13 overall categories. The grouping hierarchy can be seen in Table 1, showing how each crime type was ordered into each category. From this, we could see that Assaults and Motor Vehicle Thefts were the most common occurring 26.6% and 25.3% of the time respectively with Larceny being the next common type of crime occurring 12.4% of the time.

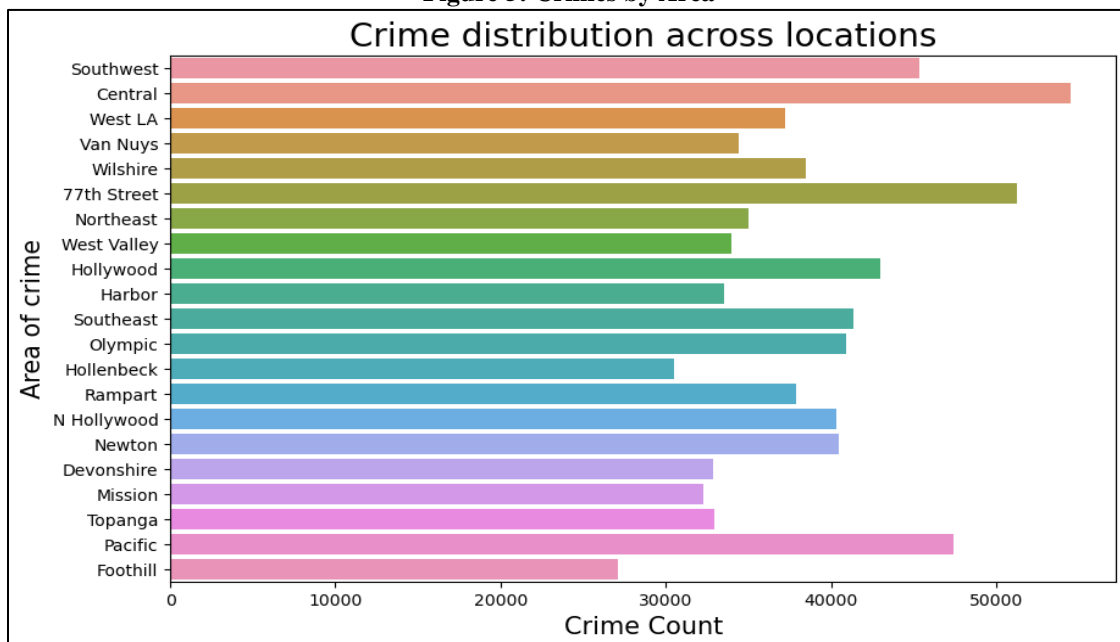
Figure 4: Crime Group Distribution



- Regional Differences:**

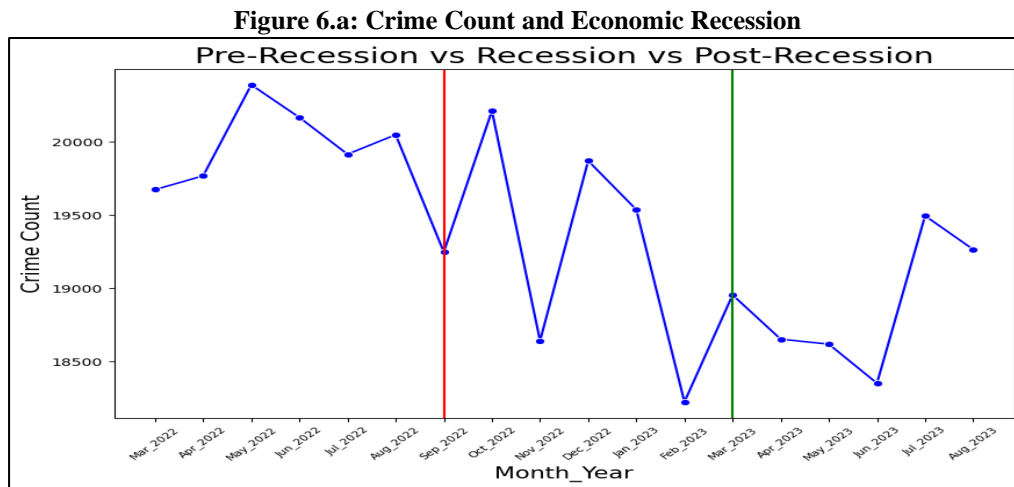
To determine regional differences among the crime data we grouped the crimes by Area name and created a plot of the crime count by area. From this, we determined that the area with the highest count was the Central area followed by 77th Street and Pacific. The area with the lowest crime was the Foothill area.

Figure 5: Crimes by Area



- **Correlation with Economic Factors:**

To visualize the effects of economic factors on the crime rate we used data from the most recent economic recession in the US. As displayed on the graph below, the start of the recession is depicted by the red vertical line (September 2022) and the end of the recession by the green vertical line (March 2023). The trends show that crimes generally decreased during the recession and continued to decrease for three months post-recession before it began to increase again.



We also analysed the correlation between unemployment rate and crimes. As Fig 6.b shows, there is a huge spike in unemployment during covid and a few bumps during recession from May 2022. The graph is plotted against unemployment rate and month-year. Also, a correlation matrix is demonstrated on unemployment rate, crime rate, years and months, as shown in Fig 6.c. It can be observed that there is a strong negative correlation between unemployment rate and crime rate.

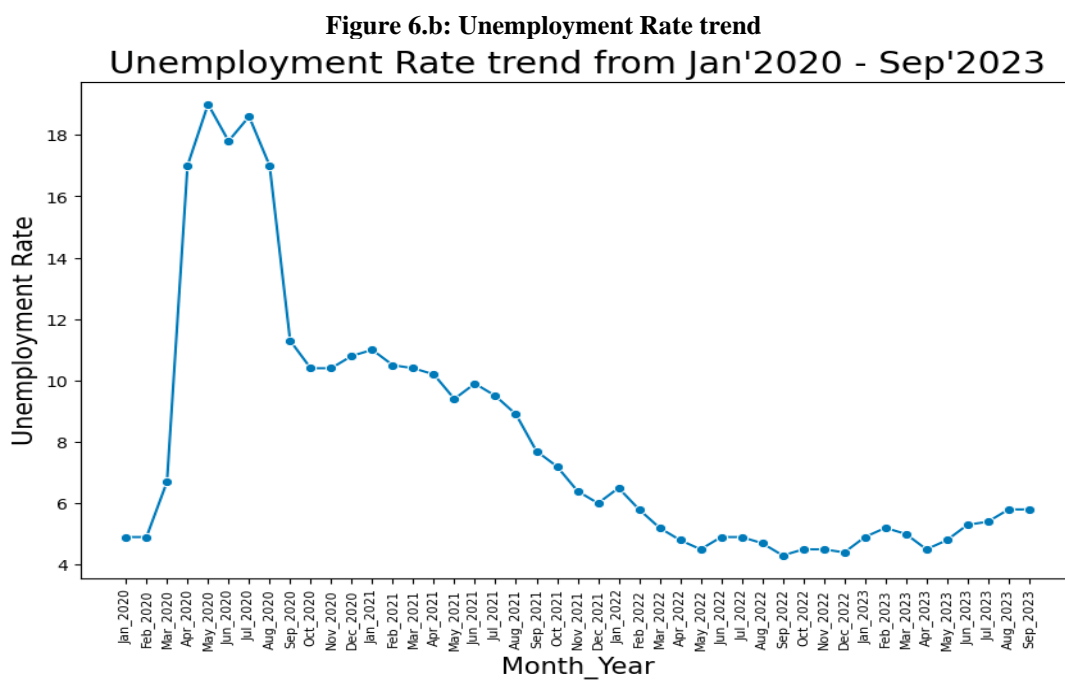
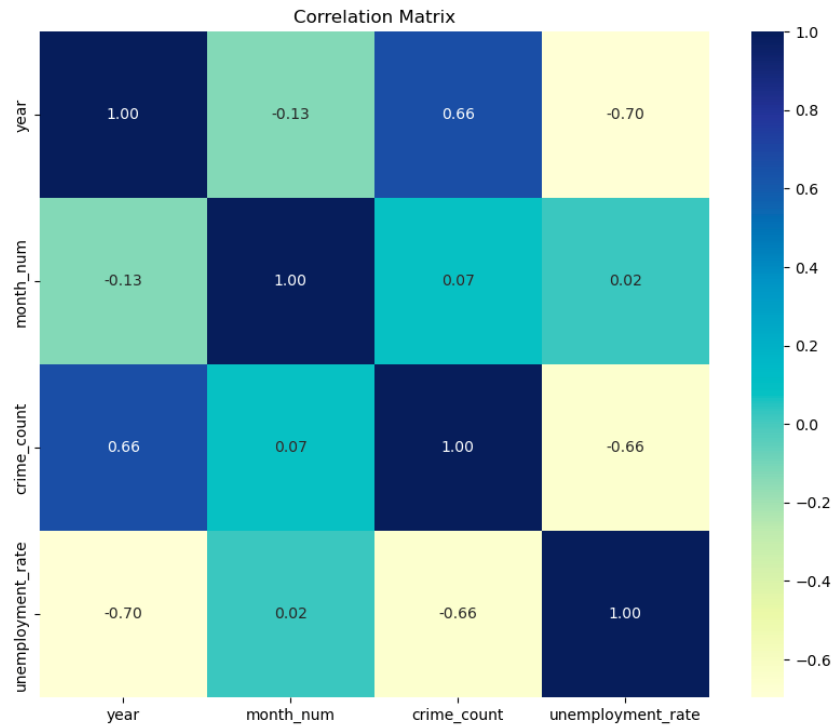


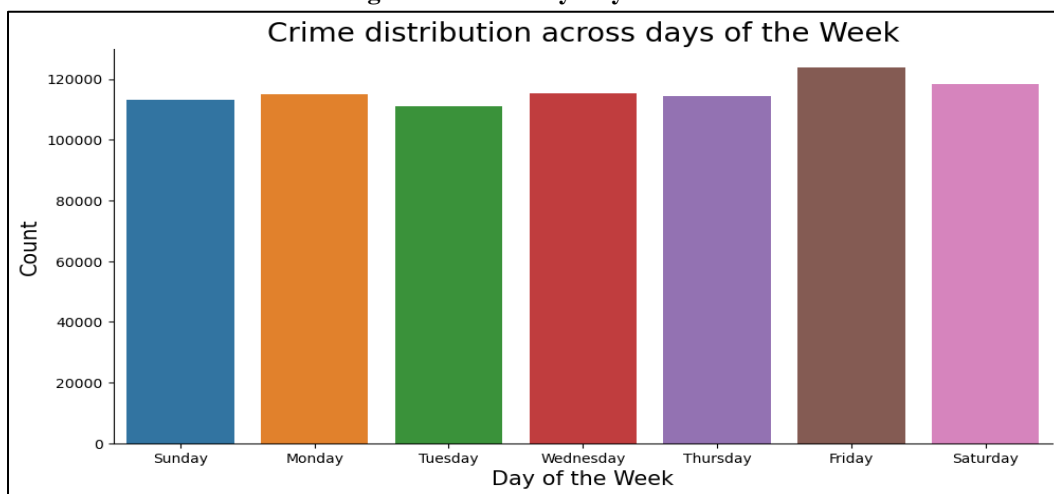
Figure 6.c: Correlation Matrix



- **Day of the Week Analysis:**

We have also done the analysis based on days on which most of the crimes occurred as shown in the above bar graph where we group the count and days of the week and we concluded that most of the crimes took place at the start of the weekend, specifically on Friday.

Figure 7: Crimes by Day of Week

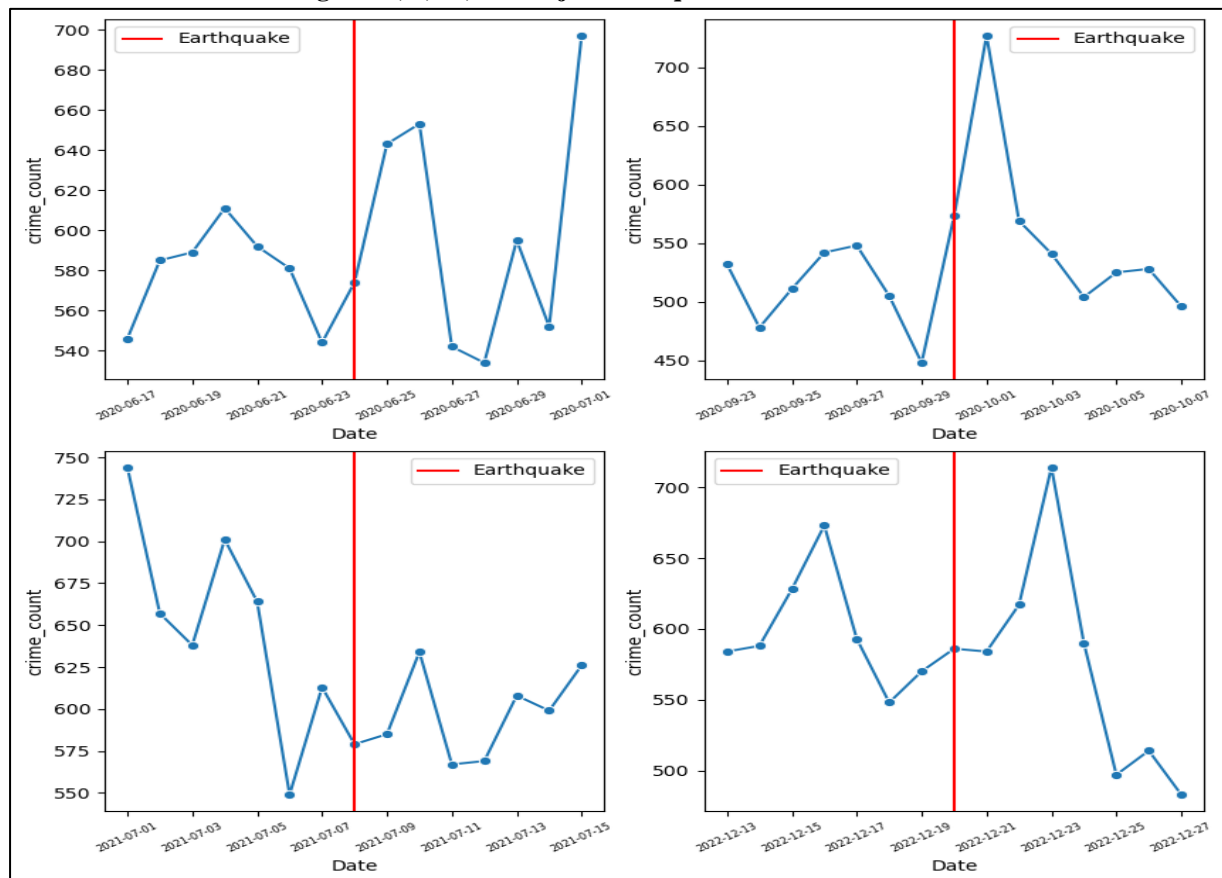


- **Impact of Major Events:**

Earthquakes

To analyse the impact of major events on the crime rate we looked at both major earthquakes data as well as Covid waves data. The below graphs show crime data with earthquake occurrences represented by red vertical lines. The major earthquakes that affected Los Angeles occurred on 24 June 2020, 30 September 2020, 08 July 2021, and 20 December 2022. After each of these earthquakes, the crime counts increased during the following days. This is potentially due to the opportunity that people have, to cause disorder following a chaos and damage producing event such as an earthquake.

Figures 8, 9, 10, 11: Major Earthquakes in LA Since 2020



George Floyd:

On 25 May 2020 George Floyd was wrongfully murdered by Police Officers in Minneapolis, Minnesota. Following this were a wave of riots and protests throughout the country, especially Los Angeles County. During this time, we can see a sharp increase in crimes from April to May 2020 and remaining high for the next few months. Many of the riots in LA were violent and destructive which included an increase in theft and looting.

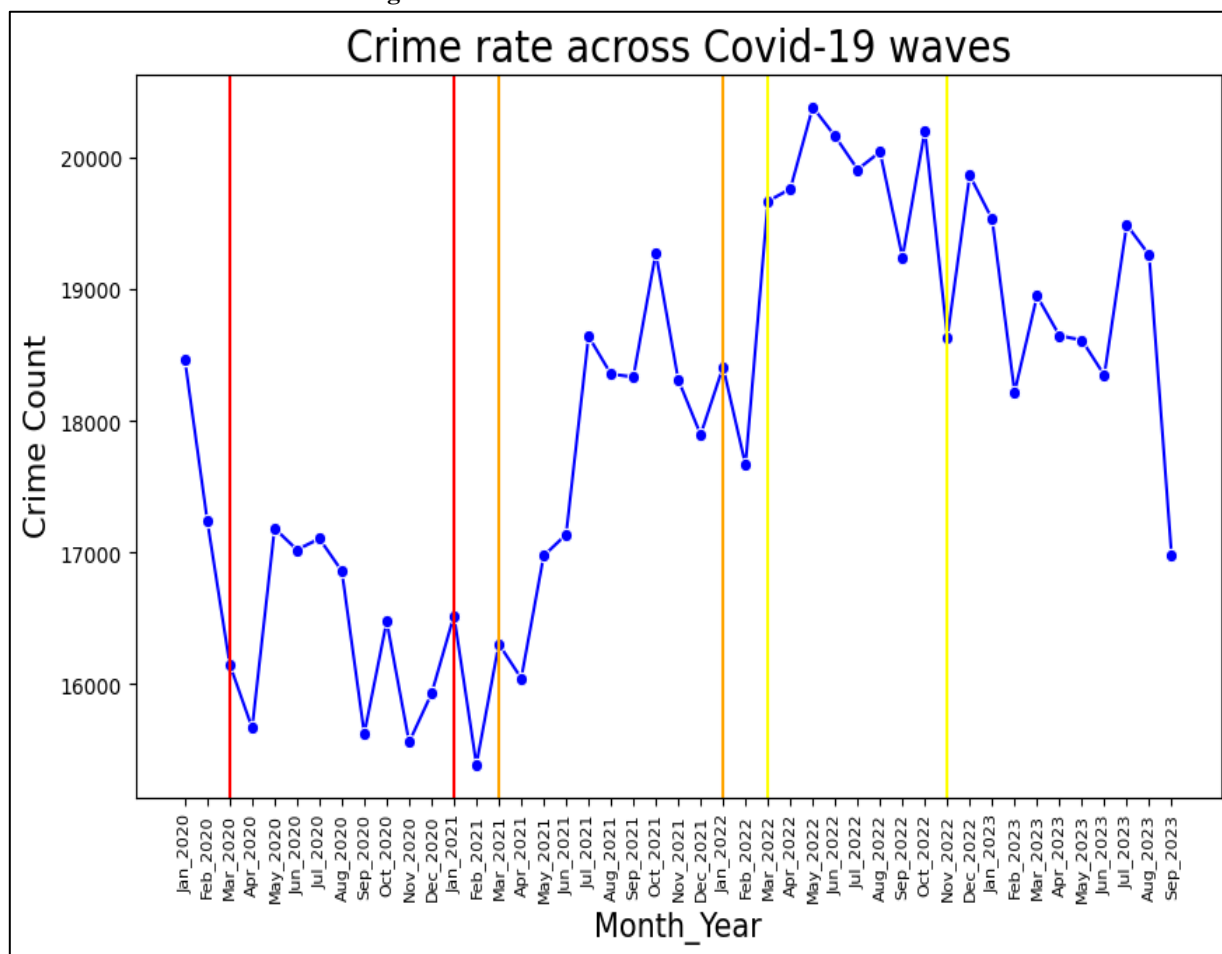
Covid-19:

The below figure shows crime count per month from 2020 to September 2023 overlaid with vertical lines depicting the start and end of Covid-19 waves in the US. The vertical red lines depict the first wave of Covid-19 from March 2020 to January 2021, the start and end of the second wave is depicted by the vertical orange lines from March 2021 to January 2022 and the third wave is depicted by the vertical yellow lines from March 2022 to November 2022.

During the first wave of Covid-19, the crime rates were relatively low. This is potentially due to the initial fear of the virus causing people to isolate and distance themselves from others or the restrictions that were put in place by the government causing many businesses to close and people to stay at home and off the streets.

During the second wave, the crime rate began to rise. This is most likely caused by society's declining fear of Covid-19. After the production of the vaccine and the decline of Covid-19 related deaths people were less likely to remain at home and isolate. Similarly, for the third wave the fear of the virus was almost completely lost and most people returned to normal daily life, businesses were open and government restrictions were removed. With this the crime rate rose again.

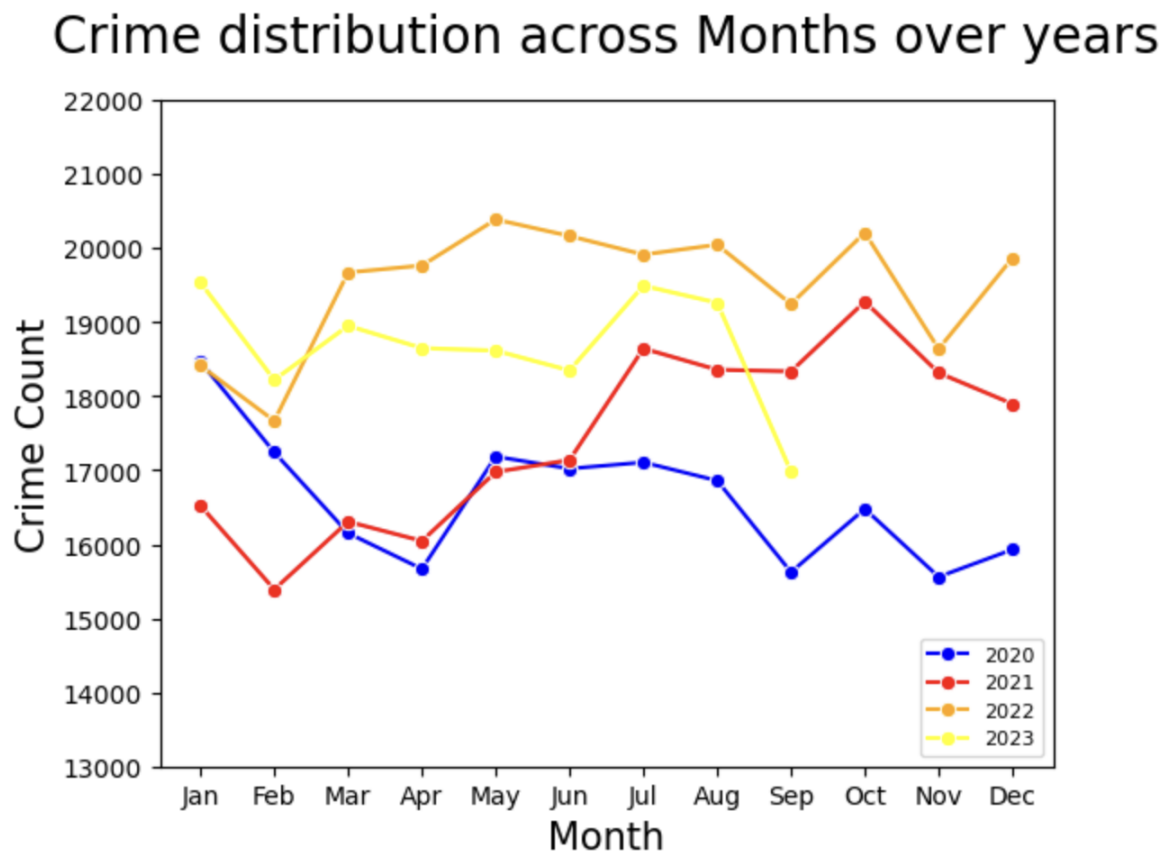
Figure 12: Crime Count and Covid-19 Waves



- **Outliers and Anomalies:**

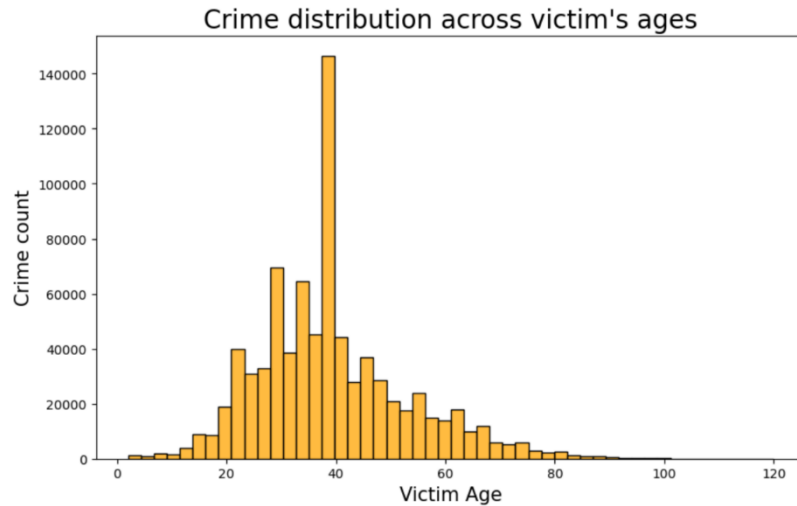
The graph below (Figure 13) shows the distribution of crime across the months over the years 2020-present. We find one anomaly in the crime pattern. We see that from August of 2023, there has been an extremely steep decline in the crime rate. This could be caused due to multiple factors. One of them being the introduction of new laws that started taking effect from July of 2023 such as the ability to file restraining orders online. Another important law that could be the reason for this anomaly is the new Firearm Civil Suits where the firearm industry is now required to take necessary precautions to prevent the sale, distribution or provision of firearms and the related products to distributors. Furthermore, they now cannot manufacture, market, advertise, import or provide wholesale sales of products that are deemed ‘abnormally dangerous.’

Figure 13: Crime distribution across the months over years.



The graph below (Figure 14) shows the crime distribution across the months over the years 2020-2023(present). One outlier or anomaly we observe here is that the crimes against victims in their late 30s to early 40s are extremely high, alarmingly higher than any other age group. This could be attributed to the general pattern of middle-aged people being financially stable, usually a little isolated socially and their tendency to live in urban areas putting them at a higher risk of being victimised in a crime.

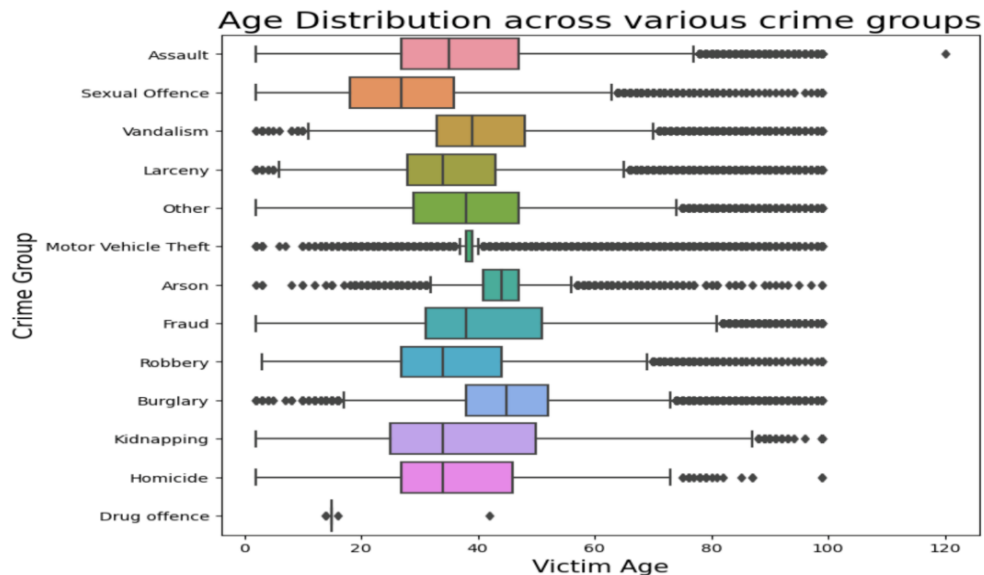
Figure 14: Crime distribution across the victim's ages.



Figures 15,16 and 17 below are Box Plots we plotted to analyse the outliers in age distribution across the crime groups, victim ages and victim sexes.

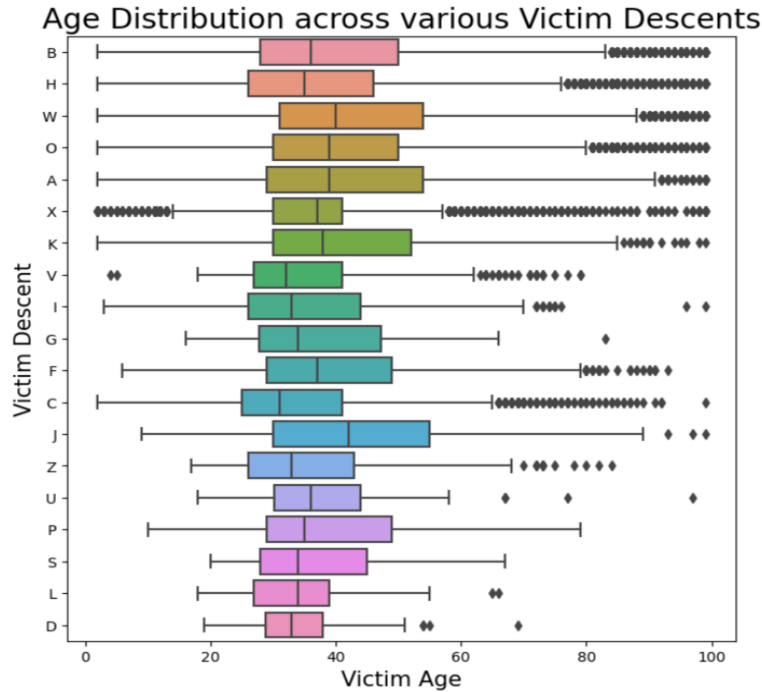
In figure 15, we observe that in every type of crime, the middle-aged crime groups are most frequent. There is a slight anomaly in this table as we can see that in the crime type of Motor Vehicle Theft, there is a surprisingly small crime group belonging to the crime group of people in their late 30s however this crime has the greatest number of outliers when compared to other types of crime. The crime type - Drug offence also has a very small age group of mostly teenagers with extremely few outliers.

Figure 15: Boxplot representing Age Distribution across crime groups



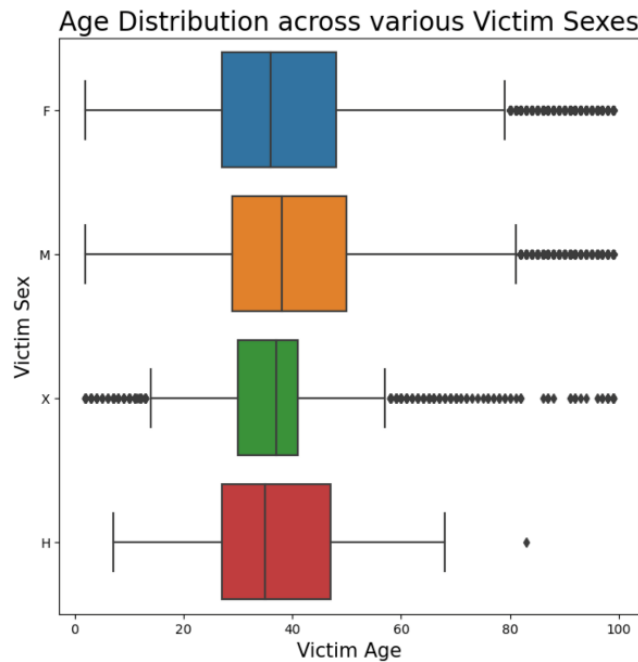
In figure 16 we observe a lot of outliers in the Latinx group and again even in this distribution, the age groups of victims are most frequently middle-aged individuals.

Figure 16: Boxplot representing Age distribution across various victim descents



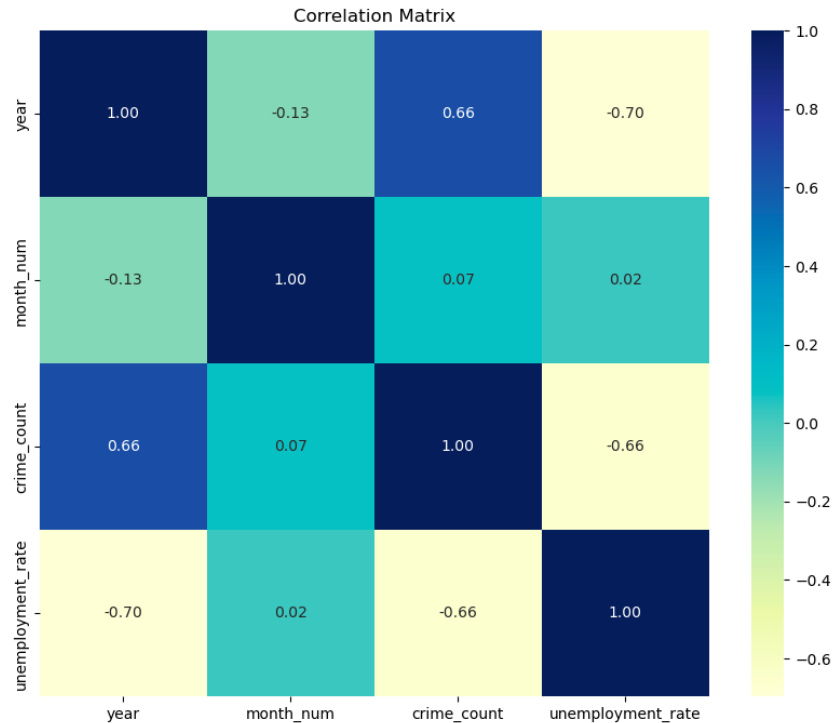
In figure 17 we observe that the Victim sex X has the most outliers whereas H has the least.

Figure 17: Boxplot representing Age distribution across victim sexes.



In the figure 18 we can observe that the unemployment rate is inversely proportional to crime rate, which is an anomaly, but this anomaly can be attributed to covid.

Figure 18: Correlation Matrix



- **Demographic Factors:**

To visualise the effects of Demographic Factors on the crime rate we used data from the Crime distribution in the US. Displayed on the pie graph below, all the analysis for this part is done on age, race and sex. To start with, Hispanic are more than a quarter of all people. After Hispanics, white and black people have the most victim count.

Pie chart for crime against sex shows that around 50% of victims are male, 42% for females. As the chart shows, the remaining 10% victims are of other sex.

For crime rate count against victim age, we plotted a bar graph depicting what age group gets more abused. From the graph it can be observed that people from age 35-50 are victims of most crime with count reaching 160000.

Figure 19: Distribution of Victims by Descent

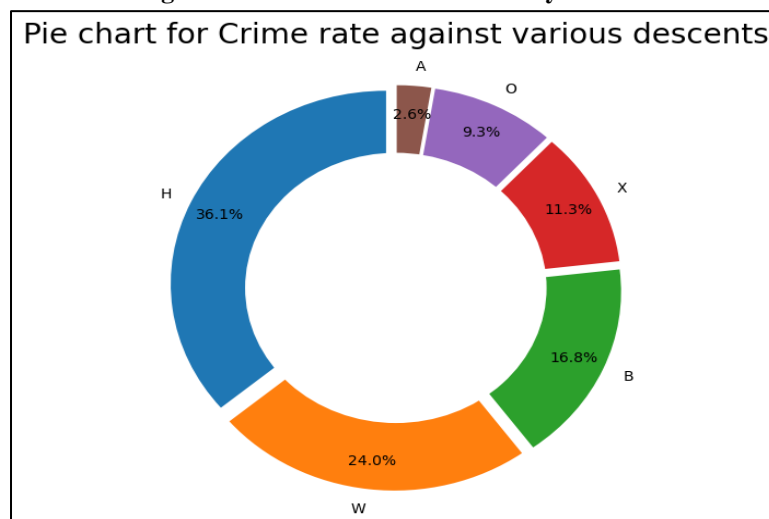


Figure 20: Distribution of Victims by Sex

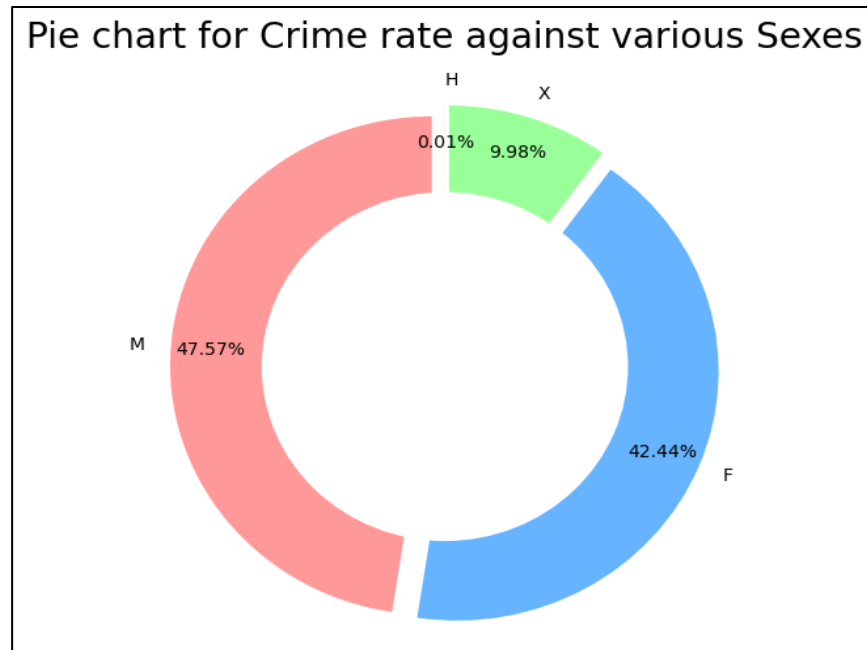
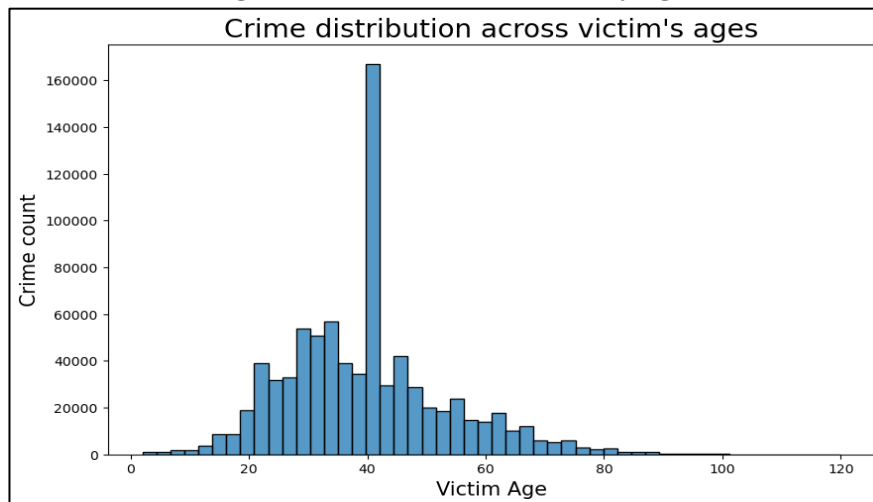


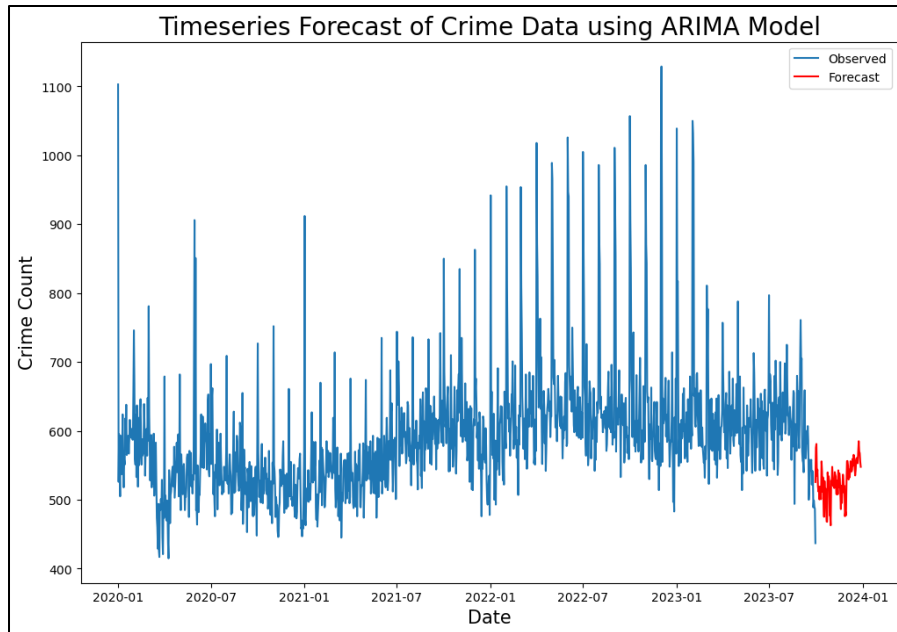
Figure 21: Distribution of Victims by Age



- **Predicting Future Trends:**

Conducting a Time Series Analysis using the ARIMA (Autoregressive Integrated Moving Average) model we predicted future values based on the past values. We created a data frame with dates before 01 Oct 2023. The set lag value is 1 for autoregression, a difference order of 0 is used to make the time series stationary and we used a moving average model of 100. We are forecasting 90 days (about 3 months) of data as shown in the figure. The data in red color is the future predicted (forecasted) data by model. This model predicts that the crime rate will drop over the next few months, which is in line with the general trend of the crime rate dropping in the last few months of the year for the previous years.

Figure 22: Timeseries forecast using ARIMA



Other Findings and Analysis:

- We observed that the crime count is increasing year by year gradually from 2020 till date
- Most of the crime that occurred were during Friday.
- The time during which most of the crime that took place were 10 AM to 10 PM
- Assaults and the motor vehicle theft are the categories that occurred and registered till date
- Most of the victims are of the age group of 30 to 45.

Conclusion

This report presents a comprehensive analysis of real-world crime data from 2020 to present in Los Angeles, aiming to shed light on the intricacies of criminal behaviour and the influences that shape our communities. Crime data analysis has far-reaching implications for law enforcement agencies, policymakers, community leaders, and citizens. By examining the intricate web of information within crime datasets, valuable insights can be uncovered that can inform crime prevention strategies, resource allocation, and the promotion of safer environments for all. Accessible crime data empowers individuals to take precautions, advocate for safer neighborhoods, and engage in community initiatives aimed at reducing crime. It fosters transparency and accountability, enabling citizens to hold both law enforcement agencies and policymakers responsible for their commitment to public safety. The importance of crime data analysis goes beyond law enforcement to include policymakers who look for empirically supported ways to improve public safety. Community leaders may use this analysis to gain a deeper understanding of the difficulties their neighbourhoods face and collaborate with law enforcement and locals to create tailored solutions.

The data analysis revealed an 8.3% annual increase in crime rates in Los Angeles, with the most common crime being stolen vehicles, followed by battery simple assault and identity theft. The least common crime was inciting

a riot. The most common crime type was assaults and motor vehicle thefts, with larceny being the next most common. The highest crime count was found in the Central area, followed by 77th Street and Pacific, while the lowest was in the Foothill area. The effects of economic factors on crime rates were visualized using data from the US economic recession, showing a decrease in crimes during the recession and a subsequent increase post-recession. The most frequent crimes occurred on Fridays. Major events such as earthquakes and Covid wave also impacted crime rates, with major earthquakes causing increased crime counts. Overall, the data analysis provides valuable insights into the crime patterns and trends in Los Angeles.

Appendix

Table 1. Crime Grouping Hierarchy

Crime Category	Crime Type
Arson	ARSON
Assault	BATTERY - SIMPLE ASSAULT CRIMINAL THREATS - NO WEAPON DISPLAYED INTIMATE PARTNER - SIMPLE ASSAULT ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT BRANDISH WEAPON INTIMATE PARTNER - AGGRAVATED ASSAULT BATTERY POLICE (SIMPLE) LETTERS, LEWD - TELEPHONE CALLS, LEWD THROWING OBJECT AT MOVING VEHICLE OTHER ASSAULT SHOTS FIRED AT INHABITED DWELLING DISCHARGE FIREARMS/SHOTS FIRED

	THREATENING PHONE CALLS/LETTERS CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT SHOTS FIRED AT MOVING VEHICLE, TRAIN OR AIRCRAFT ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER BATTERY ON A FIREFIGHTER CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT LYNCHING - ATTEMPTED LYNCHING
Burglary	BURGLARY BURGLARY, ATTEMPTED
Drug offence	DRUGS, TO A MINOR
Fraud	THEFT OF IDENTITY BUNCO, GRAND THEFT EXTORTION DOCUMENT FORGERY / STOLEN FELONY EMBEZZLEMENT, GRAND THEFT (\$950.01 & OVER) BUNCO, PETTY THEFT UNAUTHORIZED COMPUTER ACCESS BUNCO, ATTEMPT EMBEZZLEMENT, PETTY THEFT (\$950 & UNDER) DEFRAUDING INNKEEPER/THEFT OF SERVICES, \$950 & UNDER COUNTERFEIT CREDIT CARDS, FRAUD USE (\$950 & UNDER DOCUMENT WORTHLESS (\$200 & UNDER) DEFRAUDING INNKEEPER/THEFT OF SERVICES, OVER \$950.01 DOCUMENT WORTHLESS (\$200.01 & OVER) CREDIT CARDS, FRAUD USE (\$950.01 & OVER) FALSE POLICE REPORT BRIBERY GRAND THEFT / INSURANCE FRAUD
Homicide	CRIMINAL HOMICIDE
Kidnapping	THEFT, PERSON CHILD STEALING KIDNAPPING - GRAND ATTEMPT KIDNAPPING
Larceny	SHOPLIFTING - PETTY THEFT (\$950 & UNDER) THEFT-GRAND (\$950.01 & OVER) EXCPT, GUNS, FOWL, LIVESTK, PROD THEFT PLAIN - PETTY (\$950 & UNDER)

	BIKE - STOLEN PURSE SNATCHING SHOPLIFTING-GRAND THEFT (\$950.01 & OVER) THEFT PLAIN - ATTEMPT SHOPLIFTING - ATTEMPT THEFT FROM PERSON - ATTEMPT PICKPOCKET DISHONEST EMPLOYEE - GRAND THEFT TILL TAP - PETTY (\$950 & UNDER) TILL TAP - GRAND THEFT (\$950.01 & OVER) PETTY THEFT - AUTO REPAIR PURSE SNATCHING - ATTEMPT BIKE - ATTEMPTED STOLEN THEFT, COIN MACHINE - GRAND (\$950.01 & OVER) DISHONEST EMPLOYEE - PETTY THEFT THEFT, COIN MACHINE - ATTEMPT THEFT, COIN MACHINE - PETTY (\$950 & UNDER) PICKPOCKET, ATTEMPT DISHONEST EMPLOYEE ATTEMPTED THEFT
Motor Vehicle Theft	BURGLARY FROM VEHICLE VEHICLE - STOLEN THEFT FROM MOTOR VEHICLE - PETTY (\$950 & UNDER) THEFT FROM MOTOR VEHICLE - GRAND (\$950.01 AND OVER) THEFT FROM MOTOR VEHICLE - ATTEMPT VEHICLE - ATTEMPT STOLEN BURGLARY FROM VEHICLE, ATTEMPTED DRIVING WITHOUT OWNER CONSENT (DWOC) BOAT - STOLEN GRAND THEFT / AUTO REPAIR VEHICLE, STOLEN - OTHER (MOTORIZED SCOOTERS, BIKES, ETC)
Other	OTHER MISCELLANEOUS CRIME VIOLATION OF COURT ORDER TRESPASSING VIOLATION OF RESTRAINING ORDER DISTURBING THE PEACE CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER) BOMB SCARE FAILURE TO YIELD VIOLATION OF TEMPORARY RESTRAINING ORDER RESISTING ARREST CHILD NEGLECT (SEE 300 W.I.C.) CONTEMPT OF COURT CHILD ANNOYING (17YRS & UNDER)

	STALKING FALSE IMPRISONMENT PROWLER HUMAN TRAFFICKING - COMMERCIAL SEX ACTS HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE FIREARMS RESTRAINING ORDER (FIREARMS RO) CRUELTY TO ANIMALS ILLEGAL DUMPING MANSLAUGHTER, NEGLIGENT RECKLESS DRIVING CONSPIRACY CONTRIBUTING WEAPONS POSSESSION/BOMBING DISRUPT SCHOOL REPLICA FIREARMS (SALE, DISPLAY, MANUFACTURE OR DISTRIBUTE) DRUNK ROLL CHILD ABANDONMENT TELEPHONE PROPERTY - DAMAGE BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT WITH ANIM BIGAMY FAILURE TO DISPERSE FIREARMS EMERGENCY PROTECTIVE ORDER (FIREARMS EPO) BLOCKING DOOR INDUCTION CENTER INCITING A RIOT
Robbery	ROBBERY ATTEMPTED ROBBERY
Sexual Offence	SEX OFFENDER REGISTRANT OUT OF COMPLIANCE RAPE, FORCIBLE BATTERY WITH SEXUAL CONTACT SEX, UNLAWFUL (INC MUTUAL CONSENT, PENETRATION W/ FRGN OBJ SEXUAL PENETRATION W/FOREIGN OBJECT INDECENT EXPOSURE ORAL COPULATION LEWD/LASCIVIOUS ACTS WITH CHILD LEWD CONDUCT SODOMY/SEXUAL CONTACT B/W PENIS OF ONE PERS TO ANUS OTH PIMPING RAPE, ATTEMPTED CHILD PORNOGRAPHY PEEPING TOM PANDERING INCEST (SEXUAL ACTS BETWEEN BLOOD RELATIVES)

Vandalism	VANDALISM - MISDEAMEANOR (\$399 OR UNDER) VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VANDALISMS)
-----------	---

References

- “Crime Data from 2020 to Present.” *Data.Gov*, catalog.data.gov/dataset/crime-data-from-2020-to-present. Accessed 2 Oct. 2023.
- “List of Notable and Major California Earthquakes.” *CEA / California Earthquake Authority*, www.earthquakeauthority.com/California-Earthquake-Risk/California-Earthquake-History-Timeline. Accessed 3 Nov. 2023.
- “Local Area Unemployment Statistics.” *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, data.bls.gov/timeseries/LAUDV0631084000000003?amp%253bdata_tool=XGtable&output_view=data&include_graphs=true. Accessed 3 Nov. 2023.