Crime Data from 2020 to present analysis

Abstract- **Crime data analysis is analysing crime data to identify patterns, trends, and relationships that can be used to inform crime prevention and intervention strategies. My project is to analyse the data of crime in Los Angeles from 2020 to present and identify the crime pattern before COVID-19 and after COVID-19. Besides the patterns, the most practiced crime with age distribution, and the area with highest crime took place can be identified. This data helps in taking required measures to prevent crime or take necessary steps to prevent crime.**

keywords—Analysis, pandas, matplotlib, Hexbin plot, Ordinal, Nominal, CMR, CMS.

I. Introduction

Now a days crime is one of the major problems in every place. People need some awareness on how, when and where the crime is being done. From the dataset I selected, which is *crime data from 2020 to present* in Los Angeles, helps to check in which place more crime is taking place. The data is obtained from Los Angeles police authority. This data is accurate and more reliable to use for analysis. The main source of this dataset is catalog data gov website [2]. There is no unwanted data or false information in the data, it is easy to use and good to visualize.

The dataset I choose has all types of NOIR data which are Nominal, ordinal, interval and ratio data. As Nominal data types, the dataset has area of the crime, location name, longitude and latitude columns. The column 'Premises description' as ordinal data. Date column time columns are interval column. The columns' part 1-2 and status description are as the Ratio data. All 4 kinds of data are included in this data.

Research problems:

1. What temporal patterns and trends in crime rates emerged in the City of Los Angeles from 2020 to the present, and were these patterns affected by significant events, including the COVID-19 pandemic?

2. How did specific crime categories (e.g., property crimes, violent crimes) in Los Angeles change during the COVID-19 pandemic, and what factors might explain variations in these changes?

3. To what extent did geographical factors impact crime patterns in Los Angeles during the COVID-19 pandemic, and are there specific neighbourhoods or regions that experienced distinctive shifts in crime rates?

The potential benefits of answering the research questions related to the "Crime Data from 2020 to Present" dataset are significant and multifaceted. Police officers and legislators in the area can benefit greatly from knowing how the COVID-19 pandemic has affected Los Angeles' crime trends. They can use data-driven decision-making to efficiently allocate resources and create crime prevention plans that are customised for unique situations.

Public safety and reallocation of resources: By identifying changes in crime patterns during the pandemic, it is possible to enhance public safety measures and allocate resources to areas with the greatest need. This can help law enforcement agencies proactively respond to emerging challenges. A comprehensive analysis of crime data can assist in the efficient allocation of resources, as it provides a basis for determining which types of crime and areas may require more attention and resources, especially in times of crisis.

Urban Planning, the findings can be useful for urban planners and city officials in designing and improving public spaces, transportation, and infrastructure, with the goal of reducing opportunities for criminal activity in vulnerable areas. Public Health, Understanding the relationship between the pandemic and crime rates can provide insights into the broader public health implications. It can help health authorities and policymakers address the social determinants of health, which may include crime as a factor.

Research and Academia, the dataset can be valuable for researchers and academics studying the intersection of public health, social dynamics, and crime. It can contribute to a deeper understanding of how external events impact urban crime patterns. Community Engagement, Knowledge of how specific neighbourhoods or communities were affected by changes in crime rates during the pandemic can facilitate community engagement and partnerships with local authorities to address local concerns and improve community safety. Preventative Measures, the findings can guide the development of targeted preventative measures and outreach programs to address the root causes of crime in the context of major events like pandemics.

In conclusion, examining the "Crime Data from 2020 to Present" dataset and responding to the research questions can yield useful information with real-world applications for urban planning, public safety, and policymaking in addition to being intellectually stimulating. It may result in more

potent deterrent and mitigating tactics against crime in Los Angeles and elsewhere.

II. *Literature review*

This paper makes the case that the COVID-19 pandemic caused alterations in human behavior, which in turn caused an uncommon divergence between crime rates and victimization risk in US communities. During the epidemic, most violent crimes decreased in number. Nevertheless, an examination utilizing activity data indicates that there was an increased probability of becoming a victim of street crime in 2020. Individuals who were in public areas had a 15–30% higher risk of being robbed or attacked. Changes in crime reporting or potential victims choosing to engage in outdoor activities are unlikely to account for this increase. Conventional crime figures could give an inaccurate impression of the most recent improvements in public safety [1].

Massive changes to social and economic life were brought about by the COVID-19 epidemic, including an unparalleled rise in killings (1). However, overall crime was down, supporting the conventional narrative that, contrary to popular belief, the epidemic has not had a negative influence on public safety and that gun violence has been an exception (2, 3). This knowledge, which is based on examinations of conventional crime data, contributes to a crucial conversation concerning recent shifts in crime among academics, public safety advocates, and journalists.

The author demonstrated how the COVID-19 pandemic caused alterations in human behavior that resulted in an uncommon discrepancy between official crime rates and victimization risk [3]. We pay particular attention to public violence, which is defined as robberies and assaults that take place in public or commercial settings. These crimes are frequently committed by strangers and cause a disproportionate amount of fear in the general public (7–10). Starting in March 2020, there was a roughly 35% decrease in public violence as individuals stayed home more in response to lockdowns and disease risks. Throughout the summer, these offences were 10 to 15% below 2019 levels; in the fall, they resumed their 2019 levels. What became of the threat of violent crime that people encountered while they were in public areas?

III. *Strategy & methods & tools*

Tools used:

Python, Python is a general-purpose, high-level programming language that is renowned for its readability, adaptability, and simple syntax. Its emphasis on code clarity in its architecture makes it a great option for both novice and seasoned developers. Python has several uses, such as web development, data research, machine learning, and scripting, which contribute to its popularity. Python known for its simplicity, versatility, rich standard library, extensive third-party libraries, interpreted language and cross-platform. I used juyter lab to access python which is used to perform the code work online and very easy to make changes as per the required kernel.

Libraries used:

1. Pandas: Pandas is an open-source, robust Python library that is frequently used for analysis and data processing. For jobs like data wrangling, data aggregation, data cleansing, and data visualization, it offers a wide range of data structures and functions. Pandas is regarded as a necessary tool for any Python-using data scientist or analyst.

2. Matplotlib: A Python charting library is called Matplotlib. It is an extensive Python visualization toolkit that can be used to create static, animated, and interactive visualizations. It is one of the most widely used data visualization libraries in the Python community and is frequently applied to data science, engineering, and scientific applications.

3. Seaborn: A Python module called Seaborn is used to create statistical visuals. It offers a higher-level interface for making eye-catching and educational statistical graphics, and it is developed on top of Matplotlib. When it comes to displaying data with a category structure, like that seen in surveys or trials, Seaborn excels.

R language, R is a software environment and free and open-source programming language for statistical computing and graphics. Researchers, statisticians, and data scientists utilize it extensively for statistical modelling, data analysis, and visualization. R is renowned for its broad user base, adaptability, and extensibility. R language is known for its powerful statistical tools, flexible data handling, extensive graphical capabilities, and large community of users. The main applications of R are data analysis, data visualizations, statistical modeling, machine learning, scientific research. I used R Studios for utilizing the R language as it is freely available and easy to use.

Libraries used:

1. Ggplot2: With the help of the well-known R data visualization package ggplot2, users may easily construct intricate and visually beautiful

visuals. Its foundation is the Grammar of Graphics (GG) framework, which offers an organized and standardized method for creating visualizations.

2. R's tidyr package is an effective tool for cleaning and rearranging data. With its range of operations, it may turn data into a standard format that is easier to visualize and analyze, known as a "tidy" format.

3. An effective tool for data manipulation in R is the dplyr library. It facilitates the process of translating ideas into code by offering a collection of verbs (functions) that match typical data manipulation tasks. The tidy verse, a group of R packages that offer a standardized and simple-to-understand syntax for data science activities, includes dplyr.

MySQL, one of the most often used relational database management systems (RDBMS) for online applications is MySQL, an open-source option. It is renowned for being quick, dependable, and simple to operate. Millions of websites, including some of the most well-known ones like Facebook, Twitter, and YouTube, use MySQL. Numerous companies, associations, and people utilize it for a wide range of purposes, such as e-commerce, customer relationship management (CRM) systems, and content management systems (CMS). MySQL is known for its reliability, easy to use and scalable.

All the three programming languages are used for various purposes. Python is used for data cleaning and performing changes in data and visualizing a few graphs. R language is used to plot charts and graphs. Whereas, MySQL is used to make tables and data. Every language has its own use in the project work.

The main aim of using these different languages is to show the visualization that I learnt in the whole course. The python work is to clean the data and plot the main figures, the r studios is to plot the figures with greater number. R language is more reliable in plotting high quality figures and more accurate pics. MySQL is used to get the table contents.

In Python, the libraries mentioned above are used and used accordingly to plot, access data and more. In R studios, I used ggplot2 library to plot the figures. In MySQL I imported the dataset with the required columns and used them to create tables to observe the information.

IV. Results

I used three languages to visualize the graphs and tables. All the graphs, plots, tables are plotted as per the research questions.

The first research question is, what temporal patterns and trends in crime rates emerged in the City of Los Angeles from 2020 to the present, and were these patterns affected by significant events, including the COVID-19 pandemic?
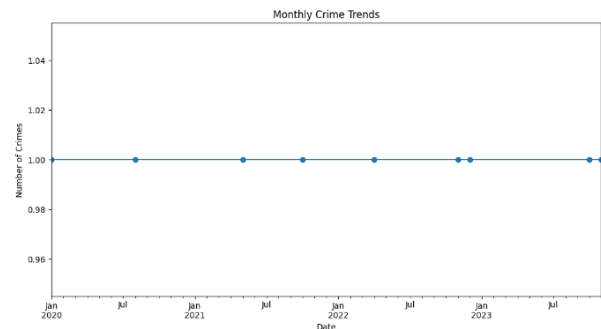

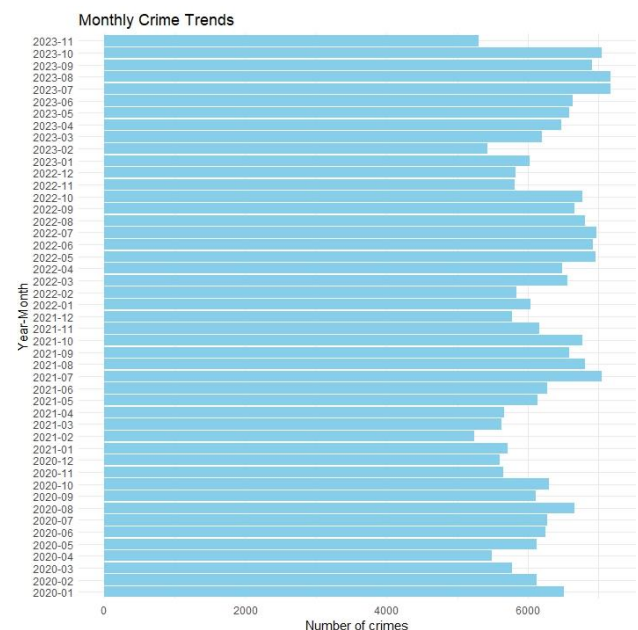Figure 1: Monthly crime trends

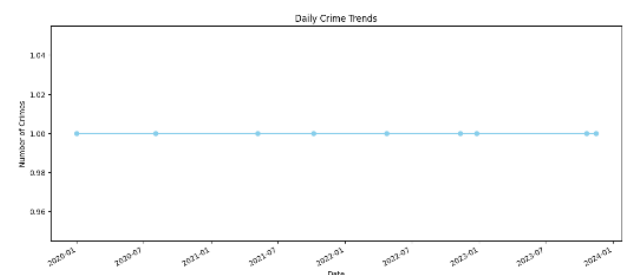

Figure 2: Monthly Crime Trend in bar graph

Figure 3: Daily Crime Trends
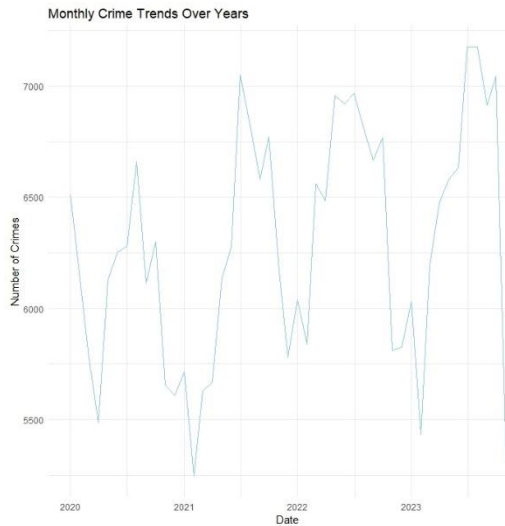
Figure 4: Monthly Crime Trends Over Years
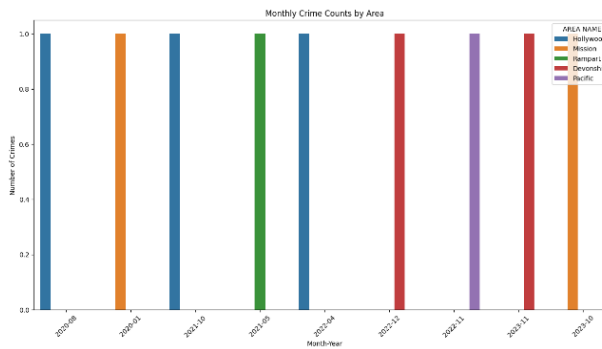

Figure 5: Monthly Crime Count by Area

City	No of incidents	City	No of incidents	City	No of incidents
17	276	02	4348	09	4010
03	5153	12	329	10	3756
01	5779	04	3791	11	3534
07	4410	05	5202	13	264
15	438	20	346	16	187
19	229	14	496	18	237
06	4732	08	4435	21	278

Table 1: A table between city and number of incidents.

From the table 1, I can say that city number *03* has highest number of incidents. The table is created in MySQL. In the same way all the 21 cities entries are listed, and their total entries of crime incidents are also mentioned in it.

Second research question is, how did specific crime categories (e.g., property crimes, violent crimes) in Los Angeles change during the COVID-19 pandemic, and what factors might explain variations in these changes?
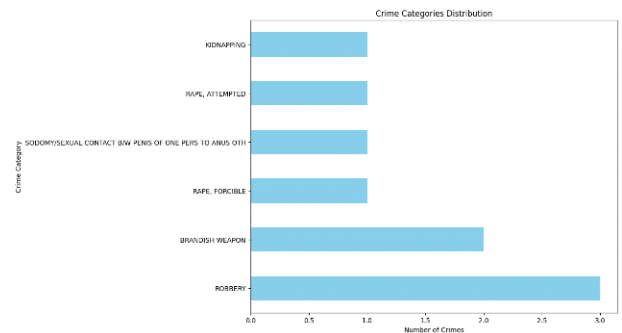

Figure 6: Crime Categories distribution

From the figures above it is easy to answer the research question. From figure 1, I can conclude that in which months the crime trend changes over 2020 to present. In Figure 2, it is shown how many crimes are done in every month from 2020 to present. By considering the graph, I can conclude that crime was little less at the time of COVID-19. After 2022, the crime rate is high again. The changes are not significantly more, but still the changes can be seen.

From the figure 3, the graph shows the crime trend on particular date throughout 2020 to present. The dates are also mentioned below them to observe. From figure 4, we can see how the crime rate is changing over time. I can observe a drop in 2021, which is the main time of COVID-19. And a high peak in the middle of 2023. The COVID-19 effected the crime a little bit. Many people stay at home because of the lockdown, which may be the cause of the fall of the crime at that particular situation. Figure 5 shows the highest crimes done in the particular area of Los Angeles according to its month.

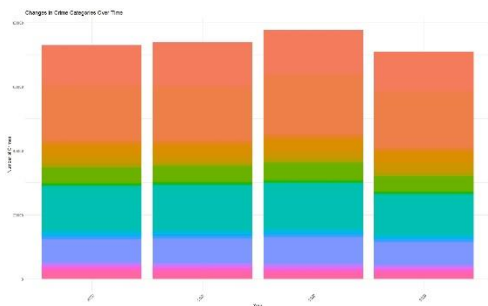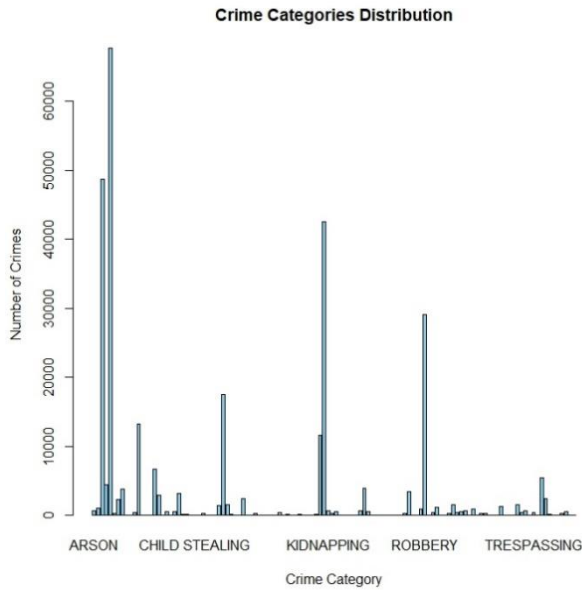Figure 7: Crime Categories Distribution

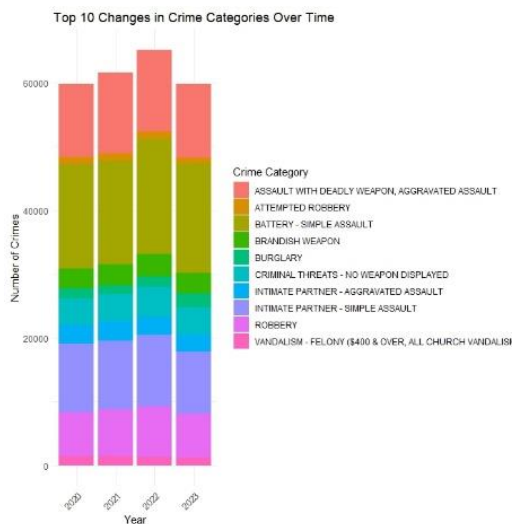Figure 8: Change in Crime categories over time

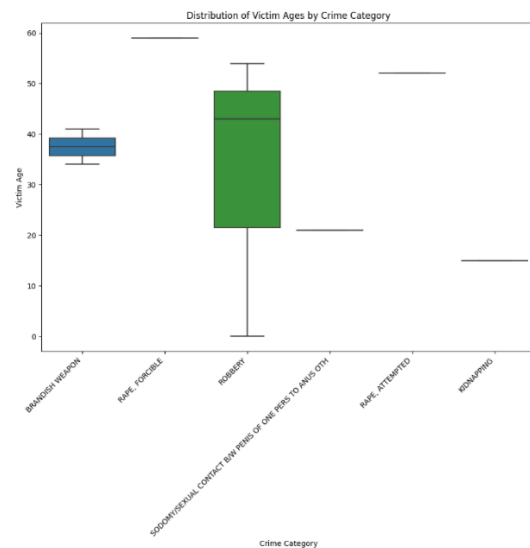Figure 9: Top 10 changes in Crime categories over time

Figure 10: Distribution of victim Ages by Crime Category

From the above figures I can answer the research question. From figures 6 and 7 we can see two types of crime category distributions. In both the graphs,

Arson, robbery, kidnap are common and have high distribution rate. Figure 8 and 9 shows the correct change of crime category over time. Over years Assault with deadly weapons is on the top throughout the years. From figure 8 we can see the change in crime over time for all kinds of crimes. But it is hard to read the data. There are over 60 types of crimes in that graph which changed over time 2020 to present (2023). To simplify I plotted the graph for the top 10 crimes. From the graph the top 10 crimes count is less in 2020 and 2021 as compared to 2022, which means the crime in 2020 and 2021 is less and COVID-19 had an impact in crime rate.

Figure 10 shows the distribution of victim age over the crimes. The robbery has the highest distribution of age, which means a greater number of age group people are committing robbery over other crimes.

	CrimeType	Frequency
	05/30/2020 12:00:00 AM	396
	01/01/2020 12:00:00 AM	350
	05/29/2020 12:00:00 AM	309
	02/14/2020 12:00:00 AM	260
	03/02/2020 12:00:00 AM	255
	06/01/2020 12:00:00 AM	255
	02/01/2020 12:00:00 AM	253
	01/24/2020 12:00:00 AM	252
	03/01/2020 12:00:00 AM	250
	02/21/2020 12:00:00 AM	250
	01/17/2020 12:00:00 AM	249

Figure 10(a): table between date and frequency of the crime

From the above figure, it is a table from SQL query. As the table is too long to show here a screenshot of the table is shown. In this table it is shown that on 0/30/2020 at 12AM a greater number of crimes are being recorded. From the table I also observed that a greater number of crimes are done in 2020 daily, which means the crime rate is high before COVID-19 on daily basis. This is not the highest number of crimes in a month or day, but it's the highest number of crimes in a day, and they are listed in descending order.

Third research question is, to what extent did geographical factors impact crime patterns in Los Angeles during the COVID-19 pandemic, and are there specific neighbourhoods or regions that experienced distinctive shifts in crime rates?
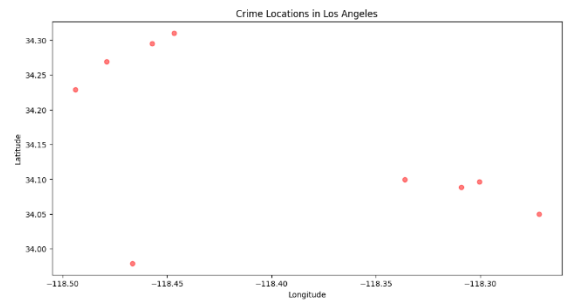

Figure 11: Crime Locations in Los Angeles as per co-ordinates
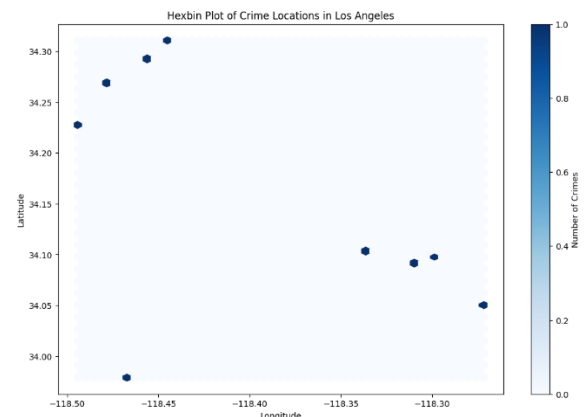

Figure 12: Hexbin plot of crime locations in Los Angeles.

From both the figures (figure 11 and 12), the co-ordinates are plotted to identify the most crime zone areas. In those locations the crime rate is high as compared to other locations. The Hexbin plot shows that all the crimes in that co-ordinates are high, none of them are low in number.

Arrested area	Incident count	percentage	Arrested area	Incident count	percentage
Devonshire	276	0.52843	West LA	4435	8.49129
South west	5153	9.86598	Van Nuys	4010	7.67758
Central	5779	11.06452	West Valley	3756	7.19127
Wilshire	4410	8.44342	Northeast	3534	6.76623
N Hollywood	438	0.83860	Newton	264	0.50546

Missio n	229	0.43 845	Foot hill	187	0.35 803
Holly wood	473 2	9.05 993	Sout heast	237	0.45 376
Ramp art	434 8	8.32 472	Topa nga	278	0.53 226
77th Street	329	0.62 991	Olym pic	346	0.66 245
Hollen beck	379 1	7.25 828	Pacifi c	496	0.94 965
Harbo r	520 2	9.95 979			

Table 2

From table 2, I can see the area name of the incident happened. The highest number of crimes took place in the Central area of Los Angeles.

Future scope and limitations:

The figures, plots and tables I extracted from the dataset are very good to analyze the research questions. This dataset can be used further to develop and answer more research questions. This dataset can be used to analyze which weapons are being used in the crime incident. Male and female ratio can be studied in the crime scenes. The dataset can be improved for other regions and the same type of analysis can be done for more detailed review of crime around the world [3].

Even though there are many uses, and the data is provided by the police authorities, there are some limitations in this dataset. The main limitation is the data is dynamic. The website always keeps updating the data monthly. So, the older visualizations are no use to analyze. The graphs and plots need to be updated frequently to get proper understanding. As the data is too big which has almost 800k+ entries, the process time is more, and the graphs are too hard to understand and study. The data needs to be shortened to get proper analysis.

V.   Reference

[1] Massenkoff, M., & Chalfin, A. (2022). Activity-adjusted crime rates show that public safety worsened in 2020. Proceedings of the National Academy of Sciences, 119(46), e2208598119. https://doi.org/10.1073/pnas.2208598119

[2] Publisher data.lacity.org. (2023, December 2). *Crime data from 2020 to present*. Catalog. https://catalog.data.gov/dataset/crime-data-from-2020-to-present

[3] Hou, M., Zeng, Z., Hu, X., & Hu, J. (2022). Investigating the impact of the COVID-19 pandemic on crime incidents number in different cities. Journal of Safety Science and Resilience, 3(4), 340-352. https://doi.org/10.1016/j.jnlssr.2021.10.008