Vivian Shu Yi Li
DS 3000: Foundation of Data Science
Prof. John Rachlin
May 31, 2023

<div align="center">**Crime in Boston**</div>

## Introduction

In the past 4-5 years, the number of crime cases in Boston has fluctuated. According to the Covid-19 impact report by the Council on Criminal Justice, homicide rates increased by 30% in 2020, while aggravated assault and gun assault increased by 6% and 8%, respectively. However, robbery rates declined by 9%, along with property and drug crimes such as burglary, larceny, and drug offenses. This short project aims to investigate the impact of the Covid-19 pandemic on crime trends over the past few years, analyzing crime datasets from 2018, 2020, and 2022 (provided by Professor Rachlin). The 2018 dataset represents the pre-Covid period, 2020 represents the Covid-19 era, and 2022 represents the post-Covid period. Each dataset contains information such as incident number, offense code, description, date, month, year, location, and more.

## Methods

To analyze crime trends over time, I will utilize line graphs, heatmaps, and histograms. Prior to incorporating interpolation into the dataset, I cleaned each dataset by eliminating unnecessary columns like offense description, district, reporting area, ucr_part, and location. Using the .drop function, I removed any duplicate entries based on incident number, assuming a one-to-one relationship between incident numbers and offense cases. Subsequently, I sorted the columns based on latitude and longitude, followed by street name, while excluding rows with missing values in these three columns. By arranging known coordinates and street names, I can employ .interpolate() to estimate and fill in missing coordinates based on neighboring point estimations. Finally, I merged the three datasets into one for easier analysis.

Firstly, I generated a line plot using the matplotlib library to compare the monthly crime case counts across different years. To calculate the monthly case counts for each year, I iterated through the data and counted the number of cases per month, storing the results in a dictionary for future plotting.

To examine the overall distribution of cases in Boston over the years, I utilized the folium library to create an interactive heatmap, showcasing the density of cases across various regions. The heatmap used a color gradient, with green indicating low values and yellow and red indicating high values. To generate the heatmap, I extracted the latitude and longitude coordinates for the specified year and organized them into a list for visualization plotting.

Lastly, to analyze the temporal pattern of cases within each year, I utilized matplotlib to construct a histogram illustrating the frequency of crime occurrences during each military hour. Each bar in the histogram represents the total number of cases reported within that specific time frame of a given year.

## Analysis

As depicted in the line plot (Figure 1), the highest number of reported cases was observed from May to August 2018. The summer of 2022 had the second highest number of reported cases, while 2020 had the lowest. Notably, there was a significant decline in cases from March to April 2020, followed by a quick increase from April to May 2020. This pattern can be attributed to the COVID-19 lockdown measures when the virus was highly infectious, resulting in people staying indoors. In addition,
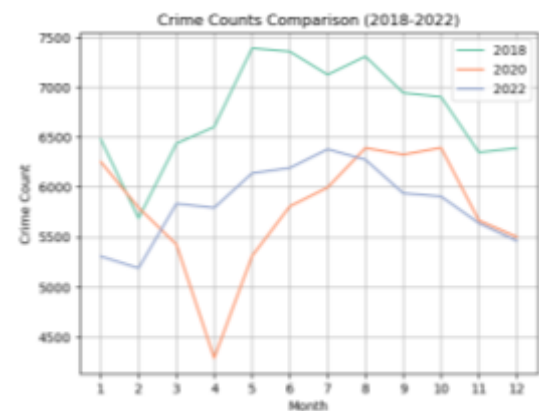


Figure 1: Line plot comparing 2018, 2020, and 2022 crime counts per month. As shown, 2018 has the highest number of cases overall, while 2020 has the lowest number of cases overall.

Government-provided stipends due to unemployment may have contributed to the lower crime rates compared to 2018 and 2020.

In the heat map (Figure 2) from 2018 to 2022, there are minimal changes in the number of cases observed. However, upon closer examination, the 2018 crime map shows slightly higher density in red, while 2020 and 2022 indicate more orange and yellow shades in terms of density.



Figure 2.1: 2018 Crime Heatmap. The color represents the level of density. Majority of the map is covered in red-orange and yellow-orange, which indicate a high number of crime cases in this year.

Figure 2.2: 2020 Crime Heatmap. Majority of the map is covered in red-orange and yellow-orange, which indicate a high number of crime cases. However, this heatmap is more orange and yellow than 2018 Heatmap, which means there is a lower number of crime cases in this year.
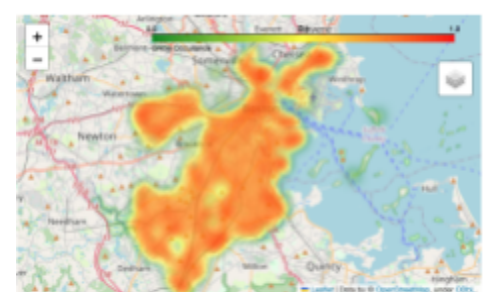
Figure 2.3: 2022 Crime Heatmap. The color density is very similar to the 2020 Crime Heatmap.

Finally, the histograms (Figure 3) exhibit consistent patterns across all three years. They reveal a higher frequency of crimes between 10 am and 8 pm, followed by a decrease before a subsequent increase in crime rates during midnight. The number of crimes during midnight surpasses those occurring during other hours for most years, indicating that the majority of crimes take place during this time.
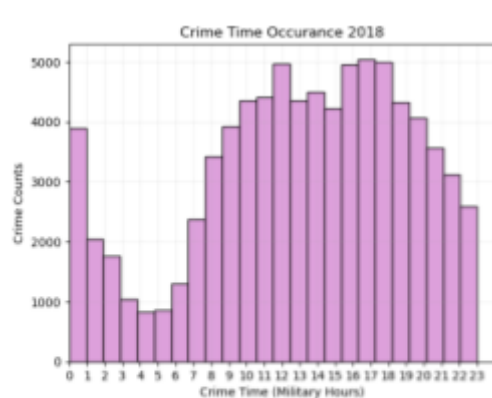


Figure 3.1: 2018 Crime Time. Each bar indicates the crime count frequency. Most crimes occur between 10 am to 7 pm. The number of frequency drops after 7pm and rise when it is midnight. However, it drops again between 1 am to 6 am.
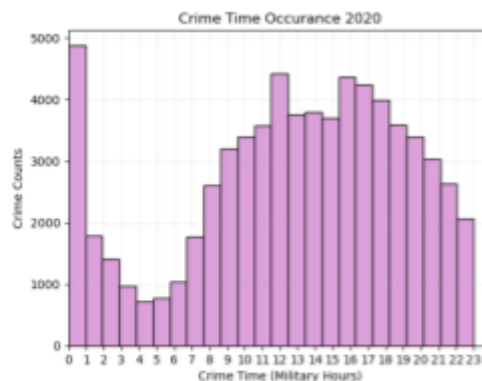
Figure 3.2: 2020 Crime Time. Most crimes occur between 10 am to 7 pm. The number of frequency drops after 7pm and rise when it is midnight. The number of crime occurrence is the highest during midnight. However, it drops again between 1 am to 6 am.
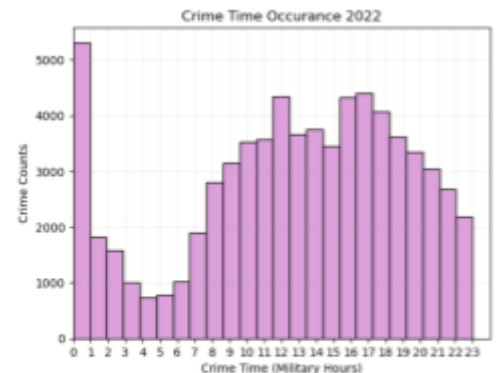
Figure 3.1: 2022 Crime Time. Most crimes occur between 10 am to 7 pm. However, the overall total of crime is the lowest compared to 2018 and 2020. Like 2020, the number of crime occurrence is the highest during midnight.

## Conclusion

In conclusion, COVID-19 pandemic appears to have little impact on the crime rate in Boston, despite the report made by the Council on Criminal Justice. The line plot demonstrates a difference in the number of crimes per year, with a noticeable deep downslope in April 2020 coinciding with the implementation of the COVID-19 lockdown measures. However, the heat map and histogram portray similarity in the crime patterns throughout each hour and in the location density. Overall, COVID-19 is not the only factor that has affected the crime rates in Boston. Other factors like socioeconomic factors, gang activity, substance abuse, and effectiveness of the law enforcement and the justice system can also contribute to the differences in crime counts.

# References

*Matplotlib — Visualization with Python*, https://matplotlib.org. Accessed 31 May 2023.

*Python Guides: Best Way To Learn Python And Machine Learning*, https://pythonguides.com/. Accessed 31 May 2023.

Breuss, Martin. "Python Folium: Create Web Maps From Your Data – Real Python." *Real Python*, 11 January 2023, https://realpython.com/python-folium-web-maps-from-data/. Accessed 31 May 2023.

"Impact Report: COVID-19 and Crime." *Council on Criminal Justice*, 31 January 2021, https://counciloncj.org/impact-report-covid-19-and-crime-6/. Accessed 31 May 2023.

"Matplotlib: Plotting Subplots in a Loop." *Engineering for Data Science*, 17 June 2021, https://engineeringfordatascience.com/posts/matplotlib_subplots/. Accessed 31 May 2023.

"matplotlib.pyplot.hist — Matplotlib 3.7.1 documentation." *Matplotlib*, https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.hist.html. Accessed 31 May 2023.

Norris, Will. "Plotting Heat Maps in Python using Bokeh, Folium, and hvPlot." *Towards Data Science*, 23 August 2022, https://towardsdatascience.com/plotting-heat-maps-in-python-using-bokeh-folium-and-hvplot-eb7c7f49dbc6. Accessed 31 May 2023.

"Plotly Python Graphing Library." *Plotly*, https://plotly.com/python/. Accessed 31 May 2023.

"Plotting Geographical Heatmaps with Python Folium." *wellsr.com*, 19 August 2022, https://wellsr.com/python/plotting-geographical-heatmaps-with-python-folium-module/. Accessed 31 May 2023.