PREDICTING CRIME THROUGHOUT CHICAGO

Katie Zink

Northwestern University

May 30, 2021

Abstract

This research predicts total crime, non-violent crime, and violent crime throughout Chicago using historical, publicly available crime data from the Chicago Police Department. It uses Facebook Prophet to predict future crime and uses Choropleth maps to visualize the crime by location. The research shows that as crime increases throughout Chicago there is a need for alternative resources. These alternative resources can combat crime and determine the locations where resources are needed most.

Keywords: Crime, Chicago, Facebook Prophet, Choropleth, Predictive Policing, Time-Series.

Introduction

This research predicts total crime, non-violent crime, and violent crime throughout Chicago using past crime data. The predictions give the City of Chicago information about where crime is currently occurring and if it will continue to occur throughout the city. The city can then allocate alternative resources to the most affected neighborhoods. The research also shows the types of prevalent crime. All data was collected from public datasets available through the Chicago Police Department and the City of Chicago.

The research consists of the following steps. Beginning with exploratory data analysis, the researcher will determine the ground truth of the data provided. Then the data will be utilized to predict future total crime, non-violent crime, and violent crime. Crime is then visualized via Choropleth maps to find location and count of crime by location. The methods used to conduct the research are Facebook Prophet and Folium Choropleth. Prophet is used for crime prediction, while Choropleth is used to visualize crime.

Literature Review

Predictive modeling and forecasting are not new topics and are used in many different disciplines. One of the more recent ways in which predictive modeling is used is in police departments. Departments across the nation have access to past crime data and information about people who commit crimes. Therefore, they are attempting to predict future crime based on past crime.

Predictive Policing

Predictive Policing uses past crime data to predict what type of crime and where it will be committed in the future. A great example is Chicago's pilot program for predictive policing that launched in 2012 (Lau 2020). The City of Chicago attempted to create a list of all individuals they thought would be most likely to commit a crime or be victims of crime. Their dataset was anyone arrested or fingerprinted since 2013. Due to bias and ineffective data, the program was shelved in January 2020.

Since predictive policing is widely tested, it is extremely important to highlight that forecasting is based on patterns found within the data (Kaufmann, Egbert, and Leese 2018). However, when predicting crime, it is not so black and white as previous patterns. Kaufmann, Egbert, and Leese clearly state in their introduction that, "In this article, we want to expand on the idea that information does not only inform, but also creates form."

The researchers reiterate that we cannot solely decide based on the data; we need to combine it with other data to look past the patterns.

Recently a new technique has been introduced: the use of social network analysis for predictive policing (Tayebi and Glässer 2016). This new technique represents relationships among people within communities. It doesn't focus on social media networks, but on the relationships and interactions a person has. Using a person's network is another way to detect patterns to predict crime.

Time Series Forecasting

Time series forecasts use historical information and patterns to predict future information. More recently there have been studies combining time series forecasting with neural networks to improve results. In particular, this research (Lim and Zohren 2020) survey's encoder and decoder designs to determine the optimal neural network for forecasting. They review one-step-ahead and multi-horizon time series forecasting throughout their experiments.

There are many types of neural networks that can be used to conduct time series forecasting. A particular experiment compared multilayer perceptron (MLP), Elman recurrent neural network (ERNN), long-short term memory (LSTM), gated recurrent unit (GRU), echo state network (ESN), convolutional neural network (CNN), and temporal convolutional network (TCN) to determine the optimal predictions (Lara-Benítez, Carranza-García, and Riquelme 2021). The researchers found that convolutional architectures are easier to parameterize than the recurrent models. Therefore, LSTM performed the best in their experiments. Another commonly used time series forecasting Python package is Facebook Prophet.

Facebook Prophet

Facebook Prophet was open-sourced in February of 2017 (Taylor and Letham 2017). Prophet is an additive regression model that automatically detects changes in trends by selecting changepoints from the data (Taylor and Letham 2017). It consists of three main components: trend, seasonality, and holidays (Choudhary 2018). They are combined using the below equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

- **g(t)**: linear or logistic growth curve for modelling non-periodic changes
- **s(t)**: seasonality
- **h(t)**: effects of holidays with irregular schedules
- **εt**: error term accounts for any unusual changes not included by the model

Prophet was created to simplify forecasting for companies and break the data down into hourly, daily, weekly, etc. predictions. It works best with data that are complex and have a strong sense of seasonality. These items are essential for crime data as crime levels vary significantly by the weather, time of day, and year. Another advantage of Prophet is that it allows researchers to look at nonlinear trends. On its release date Facebook stated, “The important idea in Prophet is that by doing a better job of fitting the trend component very flexibly, we more accurately model seasonality, and the result is a more accurate forecast“ (Taylor and Letham 2017).

These researchers (Feng et al. 2019) compared the accuracy of Facebook Prophet, a convolutional neural network (CNN), and long-short term memory (LSTM) when predicting crime for three cities. They created this project to better understand historic patterns compared to future predictions to allow municipalities to plan resources better and find alternative solutions to policing. They determined that the LSTM and Prophet performed better than the CNN. Interestingly, they also found that the optimal time for sampling the training data is three years.

In time series forecasting there are always concerns about the challenges associated with producing reliable and high-quality forecasting (Taylor and Letham 2017). The challenges identified stem from a lack of expertise in analysts and difficulty producing effective forecasts. Taylor and Letham chose Facebook Prophet because of its flexibility, speed, easily interpretable parameters, and ability to run with missing values and outliers. Prophet allowed them to create a new forecasting model that helped them produce reliable and high-quality forecasting.

Choropleth Maps

Choropleth Maps are maps that use color and shading to indicate dimensions and measures for a specified area. They are very important tools when visualizing spatial data and are widely used. A particular use for them is locating a specific address. One such example of that is when this researcher took the spatial measures of a specific crime and located the criminal's residence (Rossmo 1995). The research looked specifically at serial criminals who committed violent crime to determine their future violent crimes. Then Choropleth maps were used to display the results of the research. This research paper was published in 1995, so there have been many advances since in Choropleth maps, but it shows just how useful they are.

Another important part of using Choropleth maps is understanding the basics. These researchers get back to the basics of Choropleth maps and determine the best ways to decide on geographic units (Boscoe and Pickle 2013). As they state, most of these decisions are made on an ad hoc basis without much thought. However, they are critical because if the researchers make the wrong decision in choosing geographic units, it may lead to identifying non-meaningful patterns within the data. They suggest the best way to add value to a Choropleth map is to aggregate geographic units.

Data

This project is built using the publicly available crime data provided by the Chicago Police Department (CPD). CPD includes data on all crimes since 2001. The crime data includes Case Number, Date, Block, Crime Code, Primary Type of Crime, Description of Crime, Location, Arrests, Community Area in which it occurred, and specific coordinates. However, the crime is reported and recorded by CPD, so it doesn't include unreported crime and may be misrepresented or include errors due to human involvement and interpretation. For purposes of this project, the CPD crime data used is starting on January 1, 2016, until April 19, 2021.

Preprocessing for Facebook Prophet is limited in comparison to Random Forests or neural networks. To run the final forecast, two variables must be provided: ds and y. For the research data, ds is Date, and y is the Count of Crime. The two variables are run through Prophet as a data frame. After running the first forecast on total crime, the data is split into non-violent and violent crimes. For this research, non-violent crimes by primary type are: narcotics, deceptive practice, offense involving children, criminal damage, other offense, criminal trespass, intimidation, obscenity, stalking, public peace violation, public indecency, prostitution, interference with public officer, gambling, other narcotic violation, and non-criminal. For this research, violent crimes by primary type are: criminal sexual assault, theft, burglary, weapons violation, assault, robbery, sex offense, battery, homicide, motor vehicle theft, kidnapping, human trafficking, crim sexual assault, concealed carry license violation, liquor law violation, and arson. Those primary types were put into separate data frames and run through Prophet.

Preprocessing is different for Choropleth maps as the program only needs a location to start mapping. Therefore, the coordinates of Chicago were input into Choropleth to start. Then the location of crimes were mapped based on the coordinates from the total crime, non-violent crime, and violent crime data frames.

Research Design and Modeling Methods

All experiments were conducted in Jupyter notebook using Python. There are many different packages used to conduct this research. These packages include but are not limited to: Pandas version 1.2.4, Matplotlib version 3.1.1, NumPy version 1.16.5, Folium version 0.11.0, and Prophet version 1.0. However, the main packages used in this research are Facebook Prophet and Folium Choropleth. Prophet will predict future crime in the City of Chicago and Choropleth maps will visualize crime in the City of Chicago.

Before conducting analysis, it is critical to understand the data and the composition of the dataset. Therefore, the first step is viewing the summary statistics of the dataset. Then, to better understand the Primary Type column, the number of times each Primary Type appeared in the data is counted. This visualizes what type of crime was committed and recorded the most. Lastly, the data was visualized as the top 15 Primary Types by bar chart, Figure 1 below, and top 15 locations of the crimes, Appendix

A. Also, since Prophet predicts based on a date range, the data was viewed by years, quarters, and months to see its current line graph. These views show the seasonal nature of the crime data.

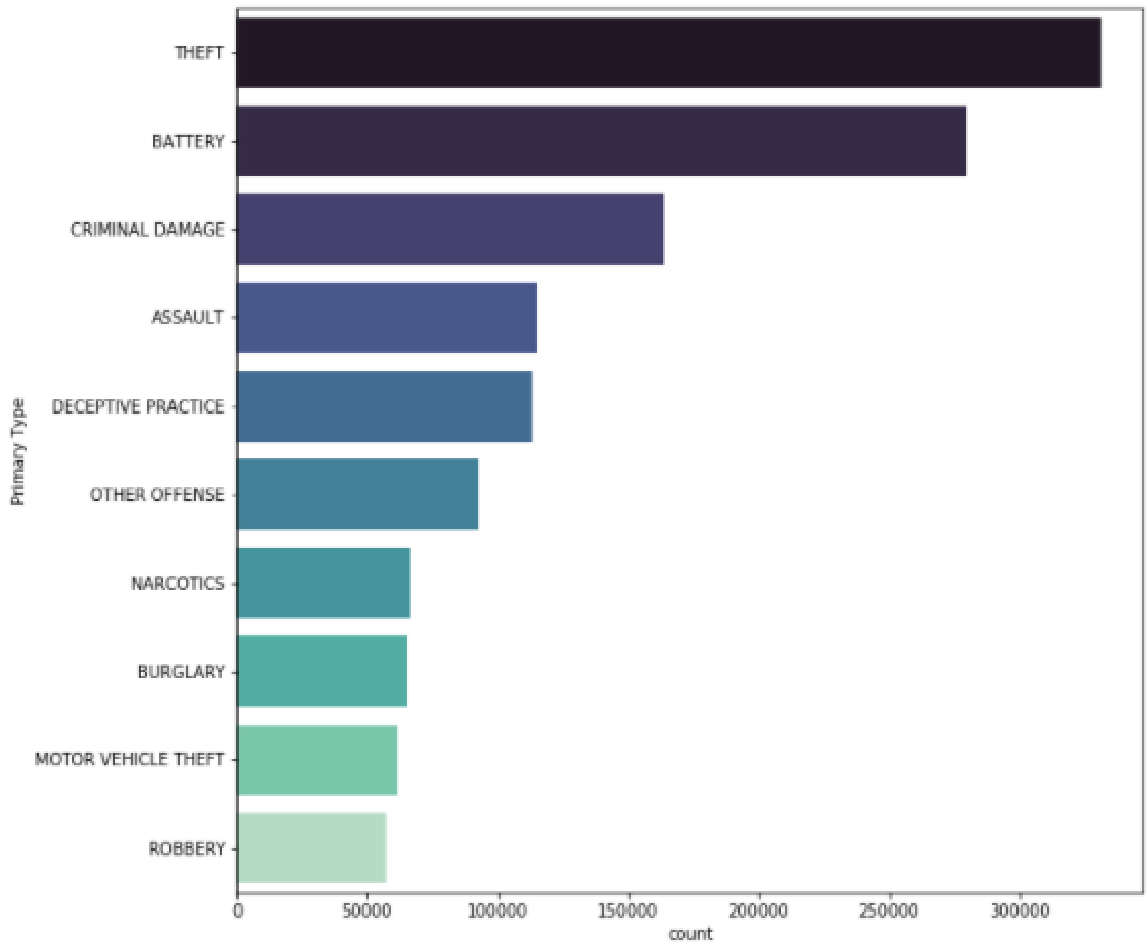

Figure 1: This is a representation of the most common primary crime type by count of occurrences in the data for Chicago crime from January 1, 2016 until April 19, 2021. The top 10 primary crime types are shown above in descending order of count.

After data exploration, the data was run through Prophet to predict total crime, non-violent, and violent crime. Prophet forecasts two years out from mid-April 2021. After forecasting, total crime, non-violent, and violent crime were plotted in Choropleth maps using coordinates provided from the data.

Results

For this research, three Facebook Prophet models were run to predict total crime, non-violent crime, and violent crime. These three models predicted crime until mid-April of 2023. Seasonality and holidays were taken into account for these models. After prediction by Facebook Prophet, the crime latitude and longitude were visualized by Choropleth maps. These maps showed the amount of crime by location throughout Chicago. There are three visualizations showing total crime, non-violent crime, and violent crime.

The results from this research show that total crime, non-violent crime, and violent crime will all steadily increase through 2023. Please find the results for total crime prediction below in Figure 2. Please find the results for non-violent crime and violent crime predictions in Appendix B. The seasonality of the crime data will remain the same with more crime during the summer and less during the winter. Also, crime will remain consistently high during summer holidays such as Memorial Day and Fourth of July. Also, crime steadily decreases from January until mid-May each year. Then it spikes in mid-May and consistently fluctuates for the rest of the year.

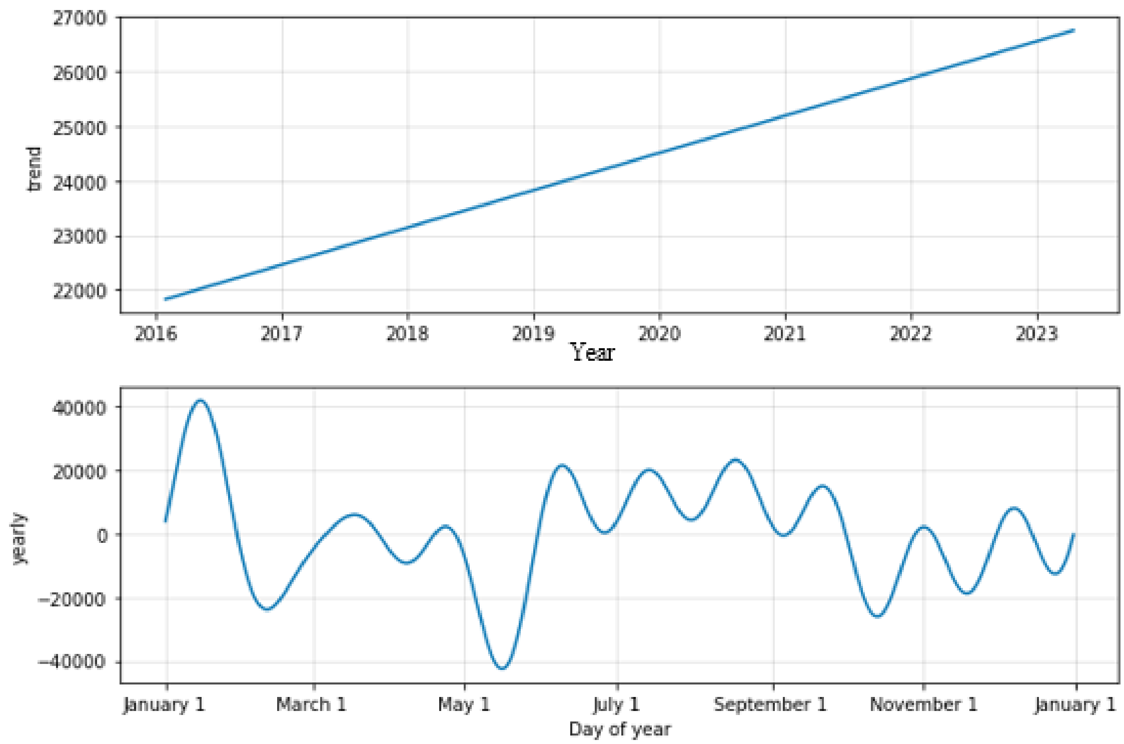

Figure 2: This is the output of the Facebook Prophet model predicting future Total Crime in Chicago. The top line graph shows the trend of crime by year through 2023. The bottom line graph shows the prediction of seasonality for crime yearly by day of year.

When viewing the Choropleth maps and location of crime there are differences in where violent and non-violent crime are occurring in Chicago. The Choropleth maps show that violent crime occurred 30 times more in downtown Chicago than non-violent crime. The specific location of this increase is in the Loop near Millennium Park in Chicago. Whereas, non-violent crime occurred 6 times more in the North Lawndale neighborhood of Chicago than violent crime. Also, non-violent crime occurs at a more similar rate across all of Chicago than violent crime. Please see Appendix C to view the Choropleth maps indicating location of crime.

Analysis and Interpretation

The results clearly show that crime will continue to increase within Chicago for the foreseeable future. This means that the city needs to invest in alternative resources to combat crime within Chicago. The CPD budget has increased every year and crime is still increasing. Therefore, the City of Chicago needs to make a new plan for combating crime. In addition to this result, there are many interesting insights that came from the Facebook Prophet results.

The first interesting insights come from the predictions of seasonality. More specifically, the large dip at the beginning of the year and the large spike that occurs in mid-May. The large dip at the beginning of this year may be due to the increase in crime last year. When reviewing the past crime data from the beginning of 2016 until 2019, there were, on average, 265,000 crimes reported each year. However, in 2020 there were over 361,000 crimes reported. This spike in 2020 crime was factored into the predictions and can be viewed in Appendix D. Also, the large spike in crime in mid-May is also very interesting. It is proven every year that crime increases in the summer and decreases in the winter. Therefore, the spike in mid-May is due to the weather finally getting nicer.

Another interesting insight is the location of violent crime versus non-violent crime. The most violent crime happened in the Loop in downtown Chicago. This is the opposite of where most people think crime is located in Chicago. However, there are a few reasons why this may happen. The first is the population density in downtown Chicago versus other neighborhoods of Chicago. Second, there are more tourists downtown during the summer than other neighborhoods of Chicago. Therefore, it is important to understand these factors when reviewing location of crime in Chicago.

Conclusions

This research predicted future crime in Chicago by total crime, non-violent crime, and violent crime. The research shows that crime will continue to increase throughout the city, whether it is violent or non-violent crime. The research also shows that violent crime occurs most often in downtown Chicago. Whereas, non-violent crime occurs most often on the Southwest side of Chicago. These results clearly show that the current way of combating crime within Chicago isn't working. Therefore, the City of Chicago needs to find alternative ways to decrease crime. They also need to direct those alternative resources to the neighborhoods with the most crime. It is imperative that the City of Chicago does this to decrease crime throughout Chicago, so all residents of Chicago can live safely.

Directions for Future Work

This research can be expanded in many ways for future work. The first additional experiment can be varying the amount of input data. The crime data goes all the way back to 2001, but this research only included data to 2016. Therefore, a researcher could include more data to see if the predictions are more accurate. Future researchers may also provide less data, as other research (Feng et al. 2019) has proven the Facebook Prophet runs best with just three years of past data.

Another option would be to tie the crime data to socioeconomic data for the City of Chicago. This will be useful in determining other factors that could lead to crime in certain neighborhoods. Then, this information could be used to determine what kind of alternative resources are needed by location. Adding socioeconomic data is very important because crime prediction shouldn’t happen in a bubble. There are many other factors that effect crime rates.

Another direction for future work is predicting Chicago crime by neighborhood and street name. Similar to including socioeconomic data, this will allow the city to determine which locations need alternative resources. It may also help in determining locations for facilities to help with mental health, homelessness, gun violence, domestic violence, etc.

Author Information

Corresponding authors

Katie Zink (KZ), zink.kath@gmail.com

Author contributions

The whole paper and code was written by the author, Katie Zink.

Acknowledgment

This paper is the result of course work produced during the Master's course in the Artificial Intelligence capstone at Northwestern University.

Conflict of interest statement

The author reports no conflict of interest.

Abbreviations

KZ, Katie Zink

References

Boscoe, Francis P, and Linda W Pickle. "Choosing Geographic Units for Choropleth Rate Maps, with an Emphasis on Public Health Applications." *Cartography and Geographic Information Science* 30, no. 3 (2003): 237–48. https://doi.org/10.1559/152304003100011171.

Choudhary, Ankit. "Generate Quick and Accurate Time Series Forecasts using Facebook's Prophet (with Python & R codes)." *Analytics Vidhya* (blog). May 10, 2018. https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/ .

Feng, Mingchen, Jiangbin Zheng, Jinchang Ren, Amir Hussain, Xiuxiu Li, Yue Xi, and Qiaoyuan Liu. "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data." *IEEE Access* 7 (2019): 106111–23. https://doi.org/10.1109/ACCESS.2019.2930410.

Kaufmann, Mareile, Simon Egbert, and Matthias Leese. "Predictive Policing and the Politics of Patterns." *British Journal of Criminology 59,* no. 3 (2019): 674–92. https://doi.org/10.1093/bjc/azy060.

Lara-Benítez, Pedro, Manuel Carranza-García, and José C Riquelme. "An Experimental Review on Deep Learning Architectures for Time Series Forecasting," 2021. https://doi.org/10.1142/S0129065721300011.

Lau, Tim. "Predictive Policing Explained." *Brennan Center for Justice.* Published April 1, 2020. https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained#:~:text=The%20Chicago%20Police%20Department%20ran,be%20a%20victim%20of%20it.

Lim, Bryan, and Stefan Zohren. "Time-Series Forecasting with Deep Learning: a Survey." *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences* 379, no. 2194 (2021): 20200209–20200209. https://doi.org/10.1098/rsta.2020.0209.

Rossmo, D. K. (1995). "Place, space, and police investigations: Hunting serial violent criminals." *In Eck, J. E., and Weisburd, D. (eds.), Crime and Place, Criminal Justice Press,* New York, pp. 217–235.

Tayebi, Mohammad A., and Uwe. Glässer. "Social Network Analysis in Predictive Policing : Concepts, Models and Methods." *1st ed. 2016*. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-41492-8.

Taylor, Sean J, and Benjamin Letham. “Forecasting at Scale.” *The American Statistician* 72, no. 1 (2018): 37–45. https://doi.org/10.1080/00031305.2017.1380080.

Taylor, Sean J., and Ben Letham. “Prophet: forecasting at scale.” *Facebook Research* (blog). February 23, 2017. https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/.

Appendix A

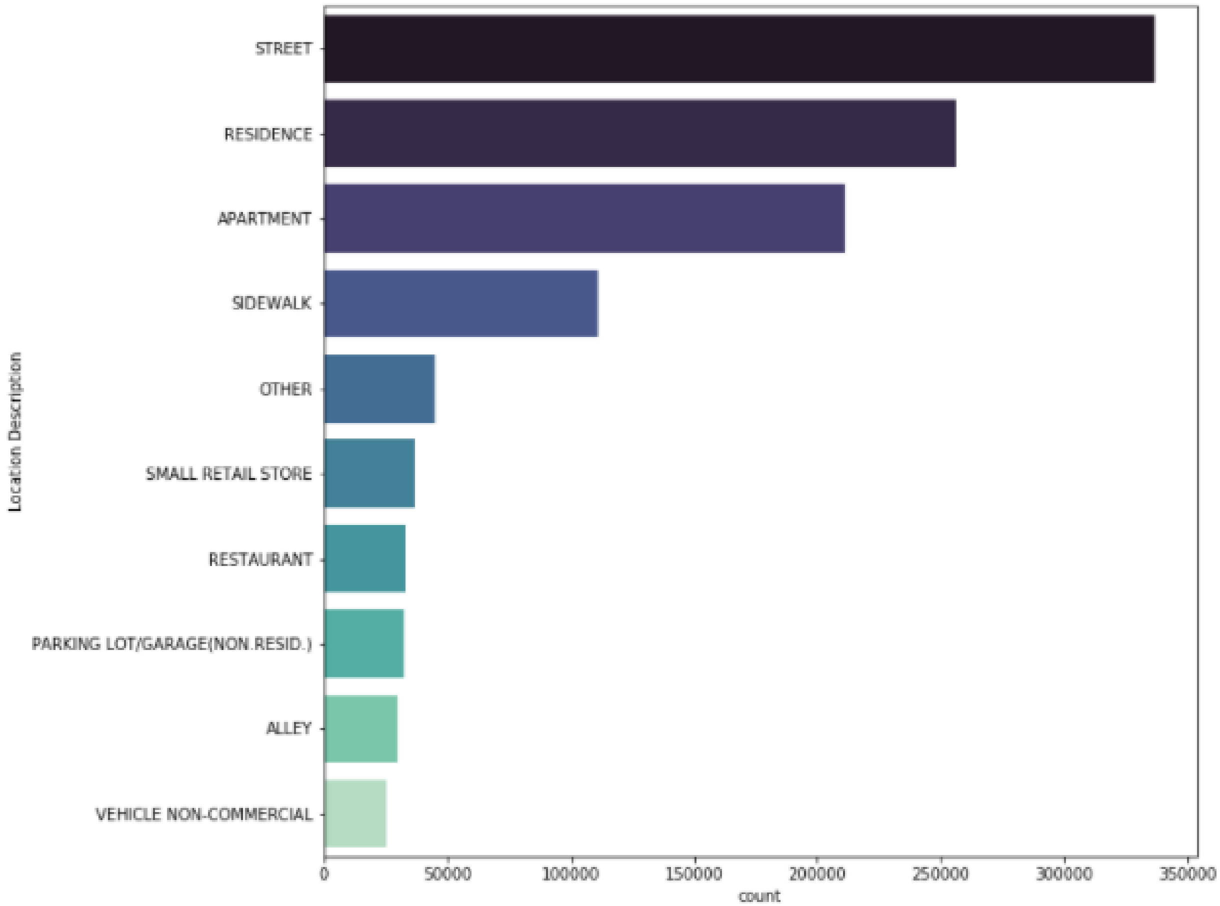

Figure 3: This is a representation of the most common location description by count of occurrences in the data for Chicago crime from January 1, 2016 until April 19, 2021. The top 10 location descriptions are shown above in descending order of count.

Appendix B

Non-violent crime predictions

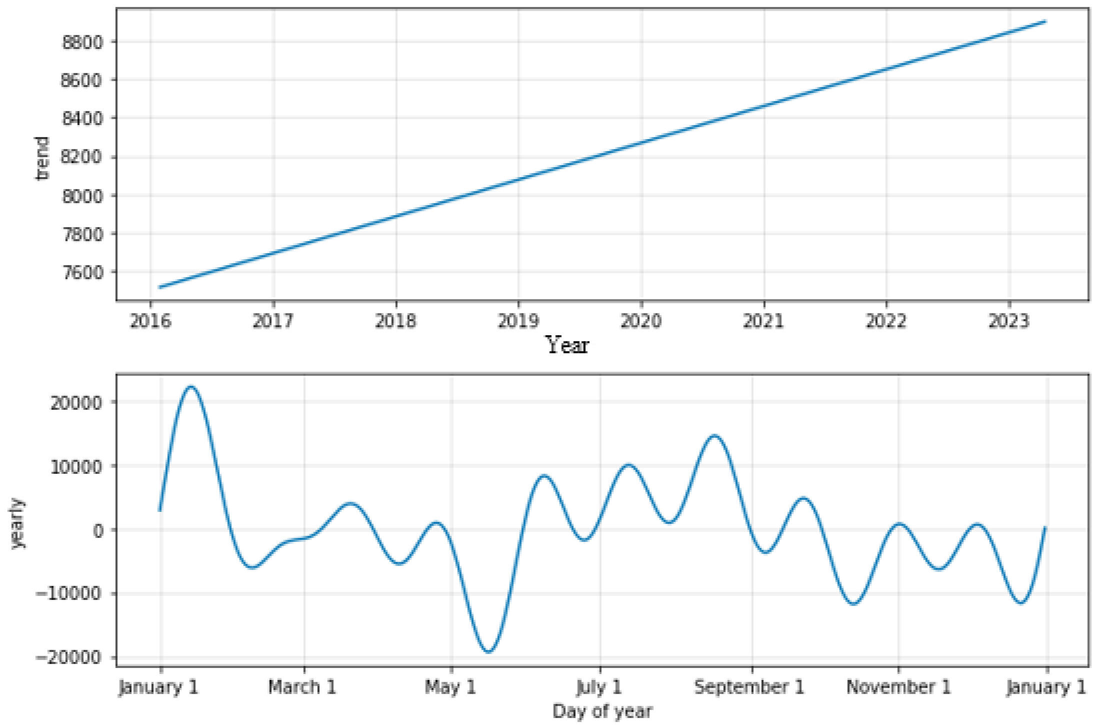

Figure 4: This is the output of the Facebook Prophet model predicting future non-violent Crime in Chicago. The top line graph shows the trend of crime by year through 2023. The bottom line graph shows the prediction of seasonality for crime yearly by day of year.

Violent crime predictions

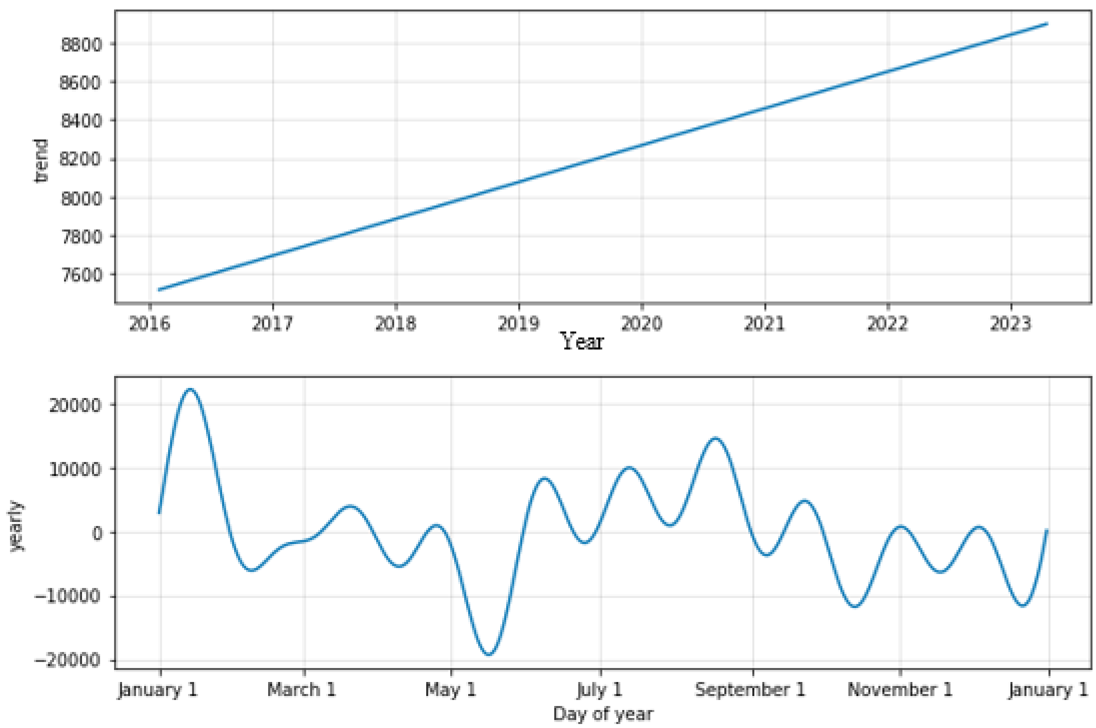

Figure 5: This is the output of the Facebook Prophet model predicting future violent Crime in Chicago. The top line graph shows the trend of crime by year through 2023. The bottom line graph shows the prediction of seasonality for crime yearly by day of year.

Appendix C

Total Crime Map

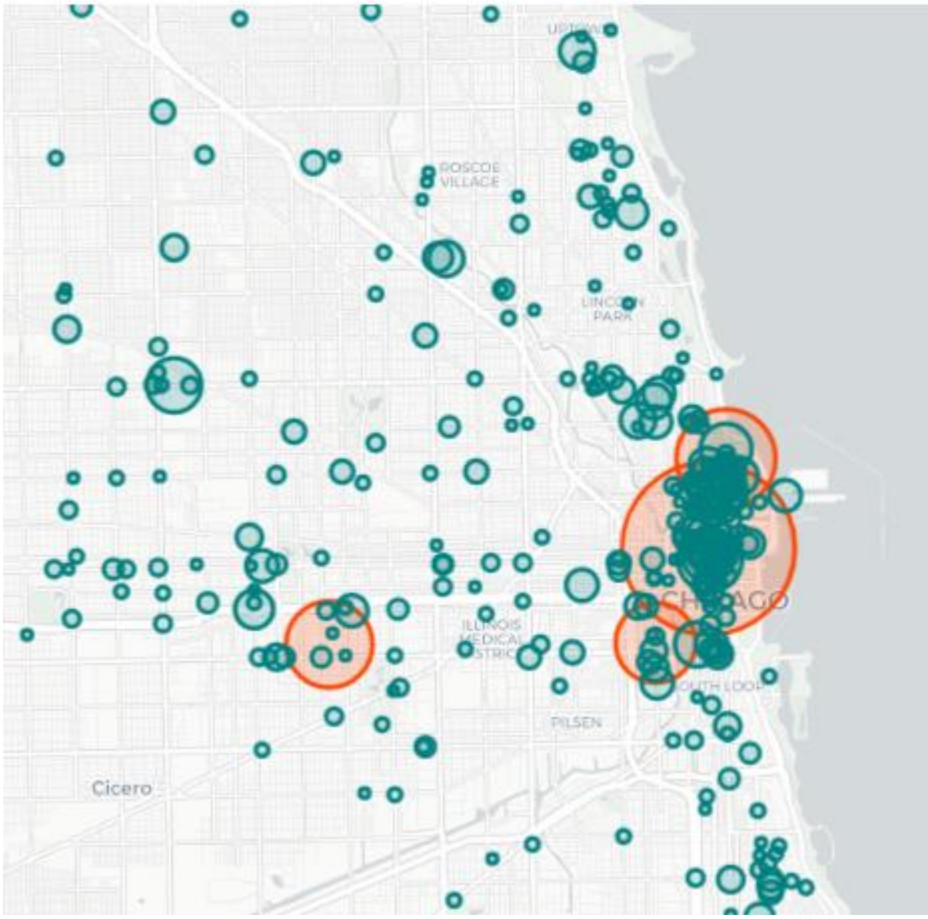

Figure 6: This is a Choropleth map of total crime within Chicago from January 1, 2016 until April 19, 2021. The size of each bubble is determined by how much crime is in each location. The orange bubbles indicate the highest level of crime in that location. The most total crime occurs in Downtown Chicago.

Non-violent Crime Map

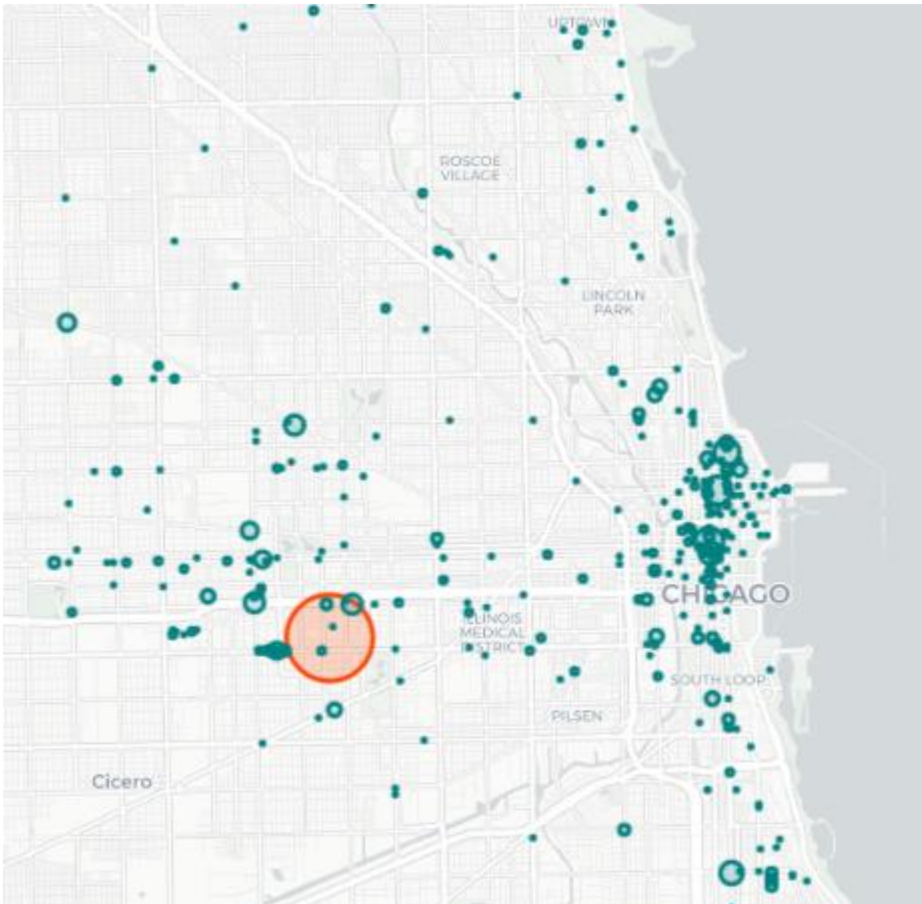

Figure 7: This is a Choropleth map of non-violent crime within Chicago from January 1, 2016 until April 19, 2021. The size of each bubble is determined by how much crime is in each location. The orange bubbles indicate the highest level of crime in that location. The most non-violent crime occurs in the Southwest neighborhood of North Lawndale.

Violent Crime Map

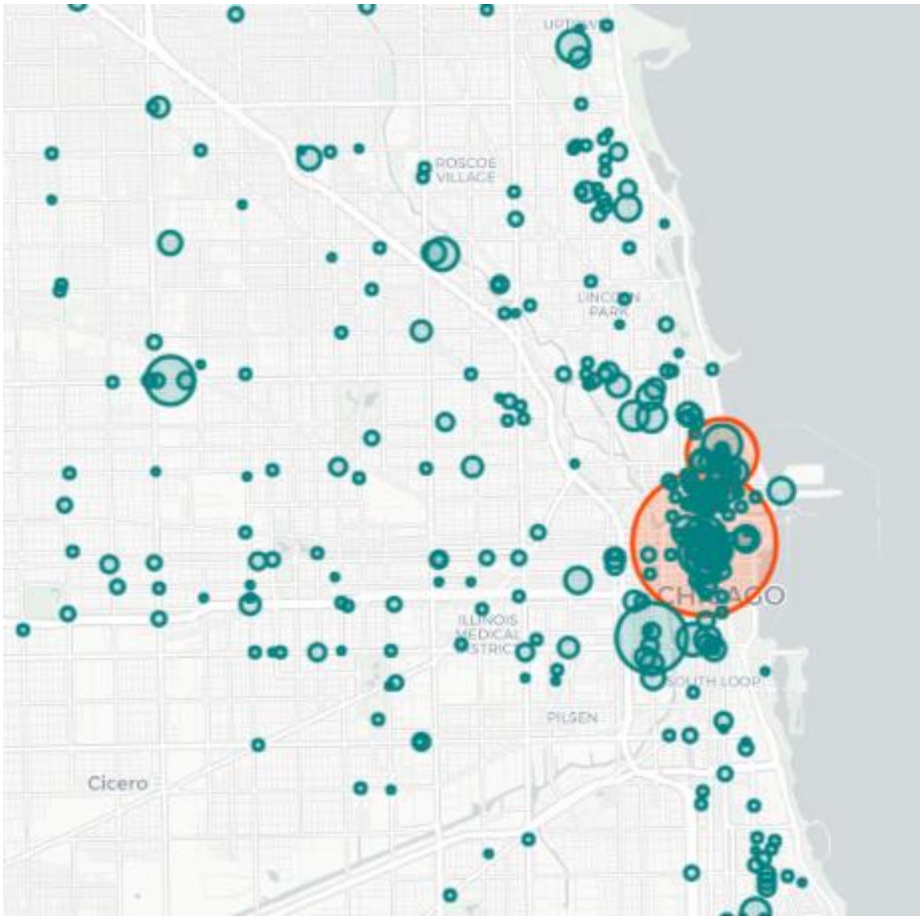

Figure 8: This is a Choropleth map of violent crime within Chicago from January 1, 2016 until April 19, 2021. The size of each bubble is determined by how much crime is in each location. The orange bubbles indicate the highest level of crime in that location. The most violent crime occurs in the Loop of downtown Chicago.

Appendix D

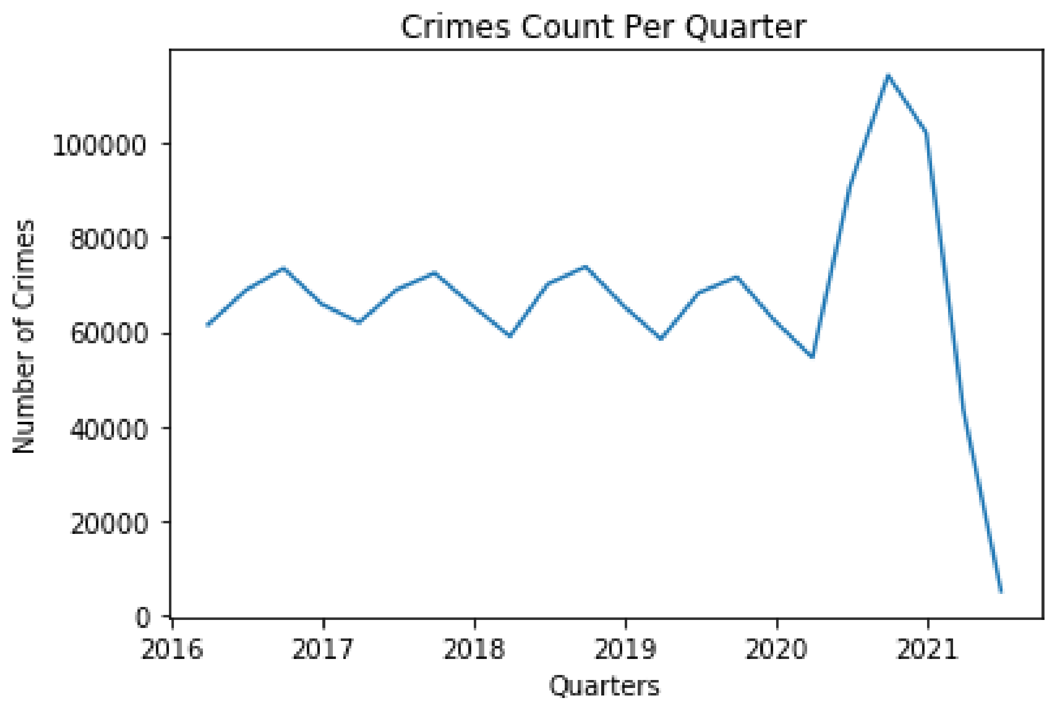

Figure 9: The line graph shows the historical number of crimes per quarter. The data is from January 1, 2016 until April 19, 2021. The data is availably publicly on the Chicago Data Portal. The line graph shows the seasonality of crime over the last five years.