Capstone Project - The Battle of Neighborhoods (Week 2)

Table of Content:

1. Introduction/Business Problem

Chicago has often been in national headlines for fluctuations in violent crime. The south side of Chicago is notoriously known for being riddled with high violent crime rates. In particular, Chicago was in the nation's spotlight in 2016 for raising the nation's crime rate overall because of the large uptick in crimes from 2015 to 2016. There is a speculation that the significant jump in violent crime circled around the case involving Laquan Mcdonald and the subsequent release of the video that recorded the incident.

By analyzing the data, I am interested in finding out any possible correlation between certain key variables that lead to predictive behaviors. Criminology theories suggest that certain types of sociological parameters such as income or presence of authority can play a role to determine and predispose people to commit crimes. In addition, I will visualize and confirm the uptick in crime rate and where/when crime is happening in Chicago.

2. Data

I used the Chicago crime dataset from Kaggle spanning from 2012 - 2017. The data included information such as date/time when the crime happened, block where the crime occurred, type of crime, location description, whether there was an arrest, and location coordinates.

I removed data from 2017 as this data was not complete for the whole year so in any year over year analysis, this would have made for an inaccurate comparison. Some additional columns I added were for the type of charge (felony or misdemeanor) the crime fell under in Chicago, general location buckets based off the location description, and also some parts of date and time parsed out in order to view the data by different aspects of time such as year, month, hour, etc.

In addition, I generated the heatmap that visualizes density of crime in different areas filtering by charge (misdemeanor or felony), location, and year range for analysis.

3. Methodology

In this project, I used python's framework to analyze the data. The different frameworks and packages are installed below. I extracted the data into Pandas dataframe and applied the format to set time as hour, month, and year. I also excluded the data in 2017 because it was only available for one month.

Then, I analyzed the general trend of the crime data from 2012 to 2016. To reduce a processing time, I used a resampling time-series method by months for the number of crimes. The resampling method in pandas is similar to the groupby method for a certain time span.

Graph 1: Number of crimes per month 2012-2017
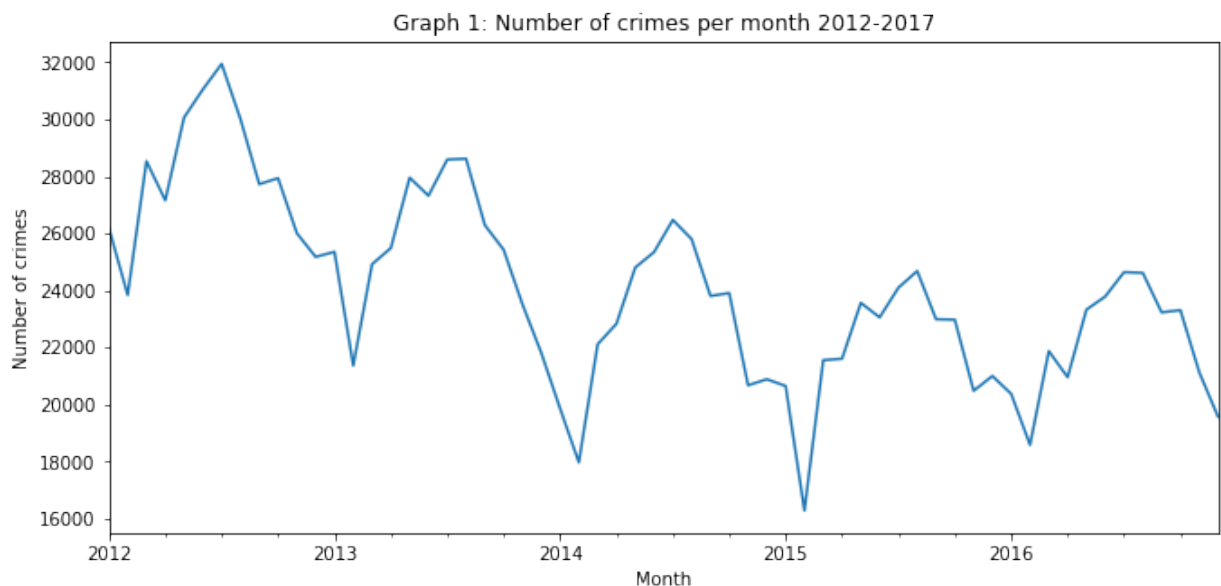
In Graph 1, it is shown that the frequency of crimes in Chicago is decreasing, with seasonal peaks between June to August and drops between January to March.
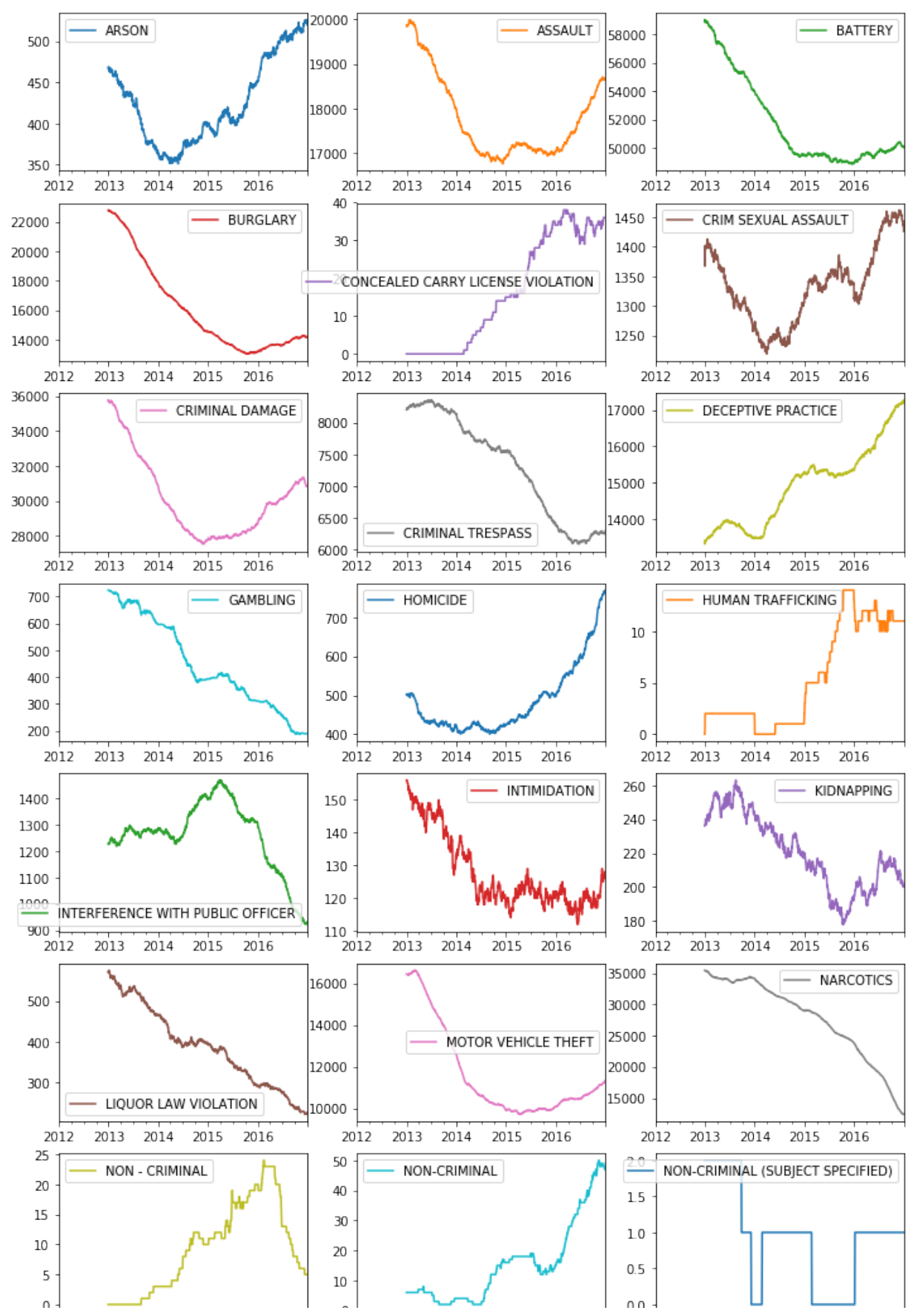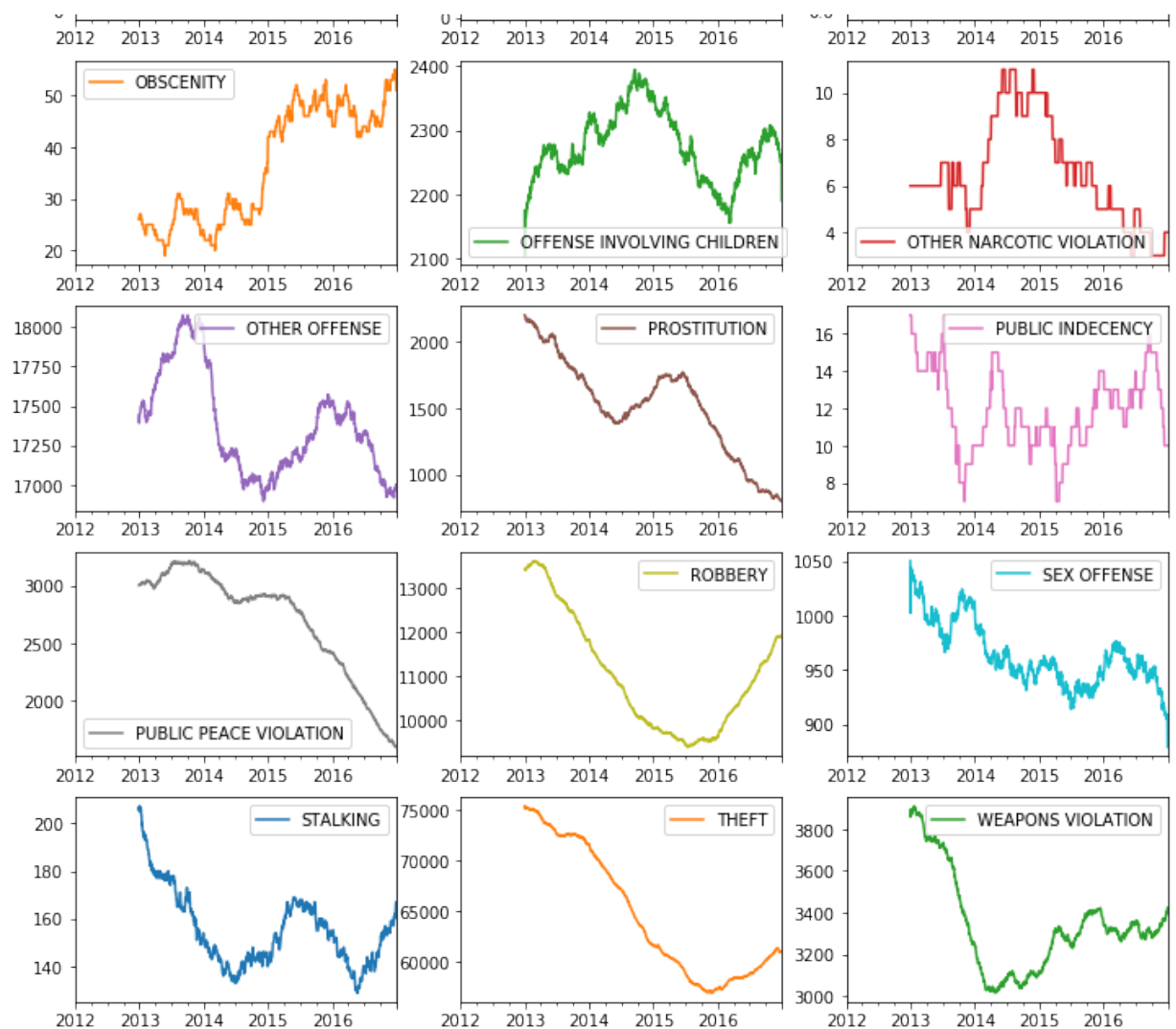
In the graph of the trend frequency for each kind of crime, it is shown that despite the decrease in the general trend of crimes, the following crimes have an increasing trend: Arson, Concealed Carry, Sexual Assault, Obscenity, Deceptive Practice, Homicide, Non-criminal, Robbery, Theft, Weapons Violation, Criminal Damage, Motor Vehicle Theft, Human Trafficking, and Stalking.

From the data above, I would like to focus on the high-frequency crimes with a rising trend such as Ignoring Stalking, Obscenity, Human Trafficking, and Non-criminal offenses. I first filtered the dataset of crimes, grouped them by the hour of the day, and subsequently plotted them based on the frequency of occurrence.

Graph 2: Total crimes by hourly frequency

In Graph 2, it is shown that crime happens mostly during daytime and dips in the early morning. It would probably be better to do a comparison of the ratio of people who are awake and people who are committing the crime. The ratio is expected be higher in the early morning; however, I do not have enough data to prove this assumption.

Nevertheless, it mentions the specific time period when citizens and policemen need to be more careful throughout the day. I can also further narrow down each crime based on crime type.

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd45fafa2e8>
```
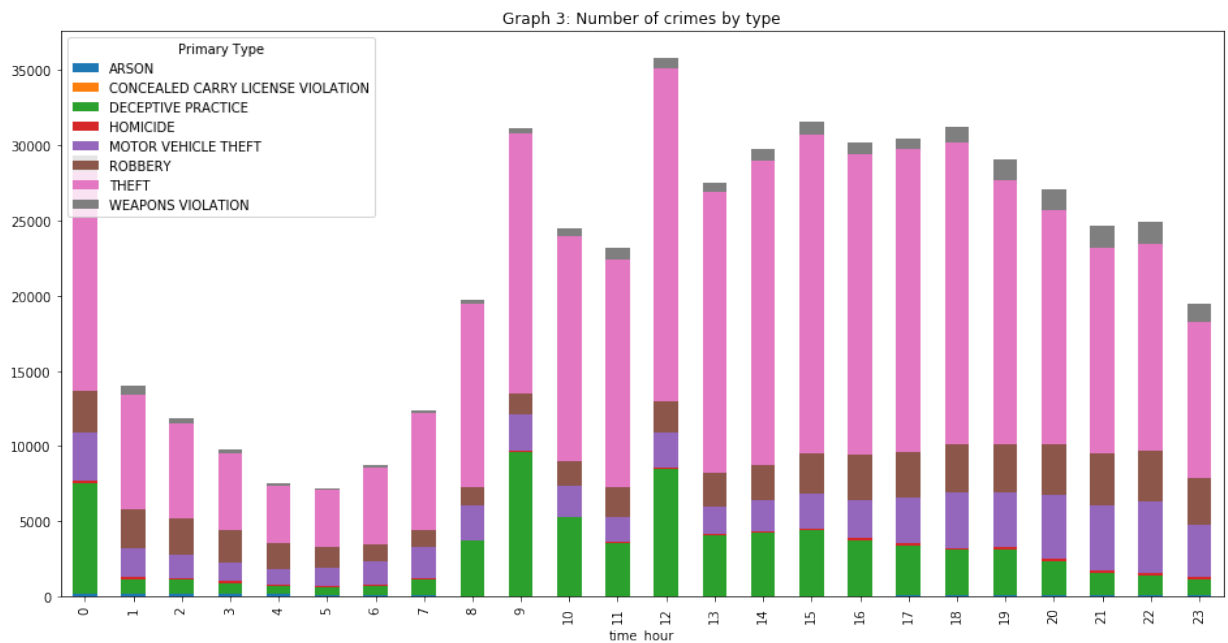

Looking at the trends across the time periods in graph 3, each crime type seems to have similar proportions within every hour. Theft and Criminal Damage form a significant portion of crimes committed. Deceptive practices happen more often during day-time, perhaps being linked to white-collar crimes with arrests that happen during office hours.

Then, I am going to analyze the trend for monthly crimes by grouping the same data by month and plot a heatmap based on top crime description and months.

Graph 4: Average number of crimes by month
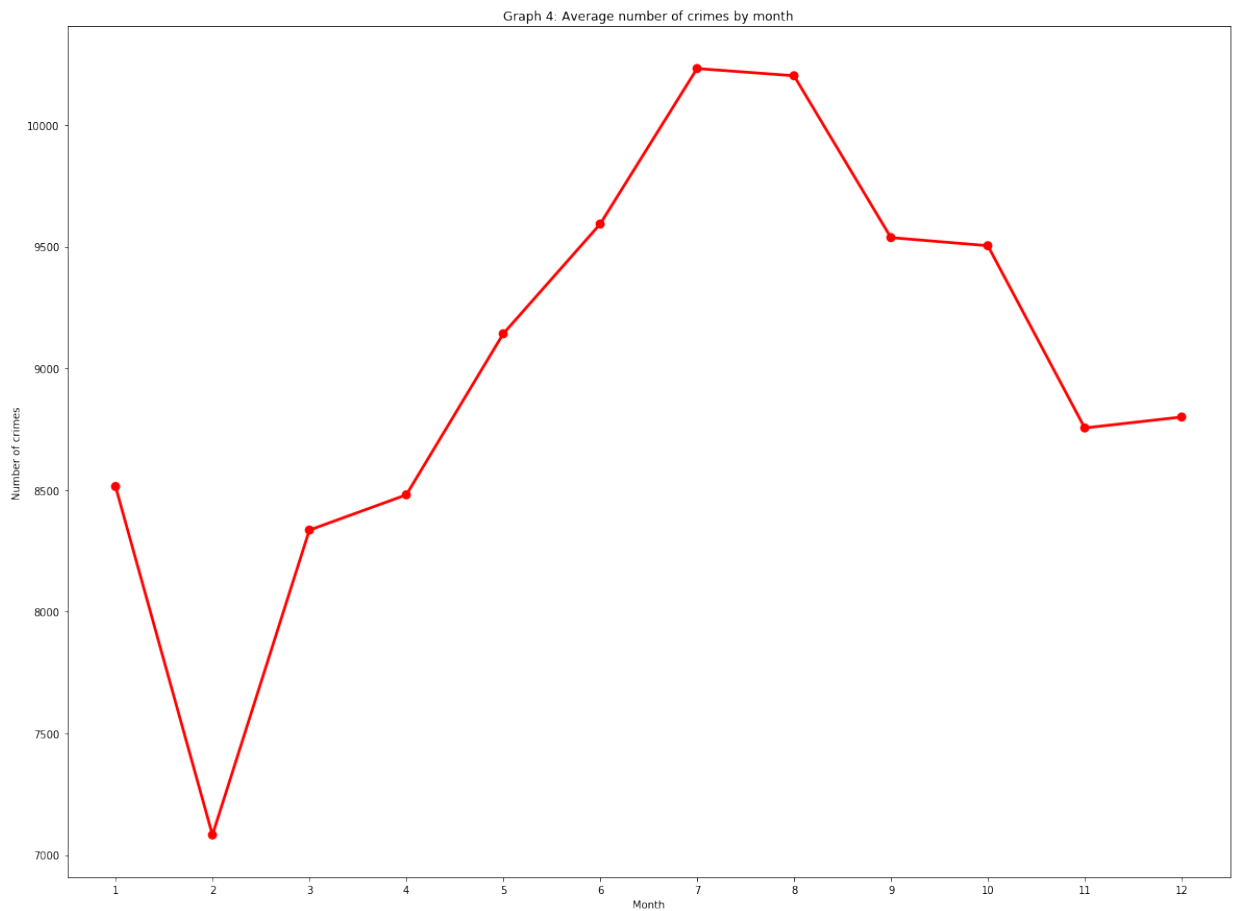
According to the monthly crime rates in graph 4, it is stated that crimes dip in February and spike in July. For tourists, this could be useful for travel information, and for the police department, the data suggests which months to be on high alert.

Using the same data of high-frequency crimes with rising trends, I analyzed the proportion of arrest for each crime. In particular, I grouped the data by crime type and arrest, then I retrieved the ratio between true and false arrest.

<matplotlib.axes._subplots.AxesSubplot at 0x7fd40fcdf160>

Graph 5: Crimes arrest ratio for the 10 different crime types
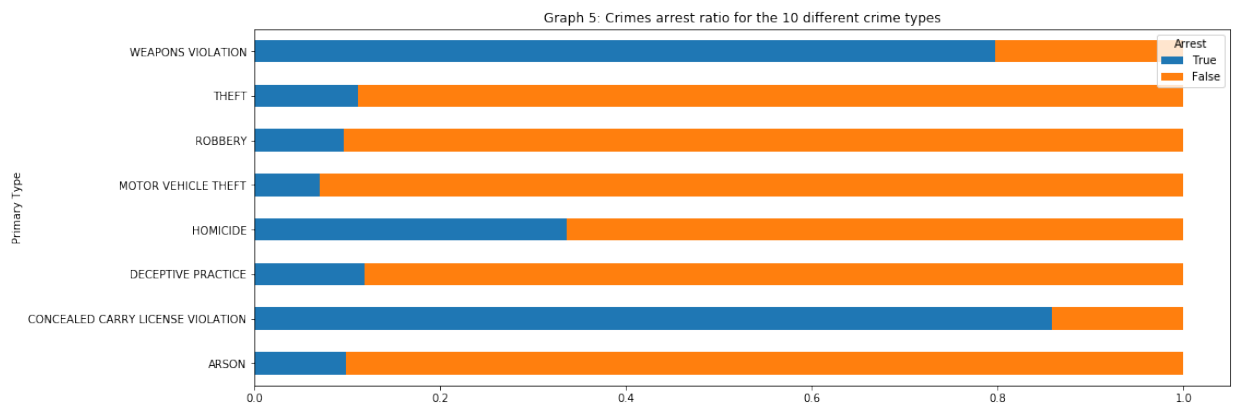
In Graph 5, it shows us that crimes involving "Weapons Violation" & "Concealed Carry License Violation" have a high arrest count. However, for the other crime types, the Chicago police department might want to look into better ways of tackling investigations. There is also a possibility of false reports and allegations.

<matplotlib.axes._subplots.AxesSubplot at 0x7fd3f91969e8>

Major crimes from January to December

Details	1	2	3	4	5	6	7	8	9	10	11	12
ASSAULT, SIMPLE	4269	3908	5186	5107	5982	6065	5872	5900	5492	5183	4268	4224
BATTERY, DOMESTIC BATTERY SIMPLE	9861	8718	10822	10822	12191	12586	12530	12010	10642	10177	9568	9731
BATTERY, SIMPLE	6045	5683	7561	7239	8798	8486	8188	8080	7878	7519	6284	5725
BURGLARY, FORCIBLE ENTRY	4620	3380	3843	3948	4354	4335	4708	4820	4645	4902	4671	4857
CRIMINAL DAMAGE, TO PROPERTY	5445	4288	5839	6039	6462	6373	6572	6462	6089	6110	5782	5576
CRIMINAL DAMAGE, TO VEHICLE	4985	4344	5787	6062	6584	6497	7177	6794	6324	6348	5736	5298
MOTOR VEHICLE THEFT, AUTOMOBILE	4845	4028	4289	3793	4035	4412	4338	4290	4065	4260	4220	4523
NARCOTICS, POSS. CANNABIS 30GMS OR LESS	5402	5770	6406	5637	5782	5279	5646	4562	4314	4217	3978	3517
THEFT, $500 AND UNDER	9691	7779	9762	10274	11662	12542	14177	13986	12725	11784	10364	10313
THEFT, OVER $500	4977	4229	5241	5683	6115	6799	7421	7449	7098	6984	6409	5912

month

From the heatmap, it is mentioned that theft under $500, domestic and simple battery are the most frequently occurring crimes in Chicago based on arrest counts. Each of these crimes has slightly different peak months, but they are all within the summer months.

4. Results and Discussion

A similar trend we notice across our analysis shows that crime rates spike from July to August during the summer period. Some researchers reason that increased temperatures from the increased daylight hours can lengthen the amount of time people spend away from their homes. It leads to the increase in the number of people in public and the amount of time that homes are left empty which creates a greater possibility of criminal activities.

Others point to the effect of students on summer vacation, who are otherwise occupied with schooling during other seasons; whereas others postulate that heat-induced discomfort could also make people more aggressive and more likely to act out.

Numerous studies have shown that rates of criminal behavior among teens and young adults drop when their communities provide them with other ways to spend their time and earn money. This was found to be true in Los Angeles during several time periods, where gang activity in poor communities was reduced when community centres were opened for teens. Similarly, a 2013 study conducted by the University of Chicago Crime Lab found that participation in a summer job program slashed the arrest rate for violent crimes by more than half among teens and young adults who were at a high risk for committing crime.

Taking these facts into consideration, it seems clear that the problem is not about more people being outside during the summer months, but about they being outside in unequal societies that are not providing for their needs. Crime might spike at a time when a greater amount of people being in public together simultaneously, and leaving their homes unattended, but that is not why crime exists.

Crime rates are falling over the year and police attributes the drop to putting more officers on the streets and investing in new technology and a smarter policing strategy.

"Chicago Police Department now has data-driven command hubs known as Strategic Decision Support Centres in 12 of the department's 22 districts. The centers link gunshot detection sensors and video to a system that can rapidly alert cops in the field to a shooting and identify potential suspects. The centers, operated by police and civilians from the University of Chicago Crime Lab, employ a predictive computer program called HunchLab to look at data such as crime incidents and proximity to liquor stores and gas stations. The software anticipates hot spots and suggests the deployment of officers to prevent retaliatory shootings."

Of course, this is not to mean all crime types have seen falling crime rates. Some crimes such as stalking, theft and sexual assault still have a rising trend. The Chicago Police Department needs to be aware of these incidents and take appropriate actions.

5. Conclusion

We believe this data analytic project give us a scientific view about the security status and crime rate of the Chicago city. According to the analysis result and visualization, we can view the most frequently occurring crimes and the frequent occurring months where crimes happened. We alsospecifically looked into certain crime types to view how they have changed over the years, such as theft, homicide, and sexual crimes. Even though there were a lot of reported crimes in Chicago each year, the arrest rate was not even as high as 50% for each year letting us believe that Chicago's police arrest or investigation methods were not effective enough. We believe if our data analytics can give us all these information about the security status of the Chicago city, a bigger data analytics project will provide much more valuable information which can be used as a powerful source for taking wise actions that increases the security status of our cities.