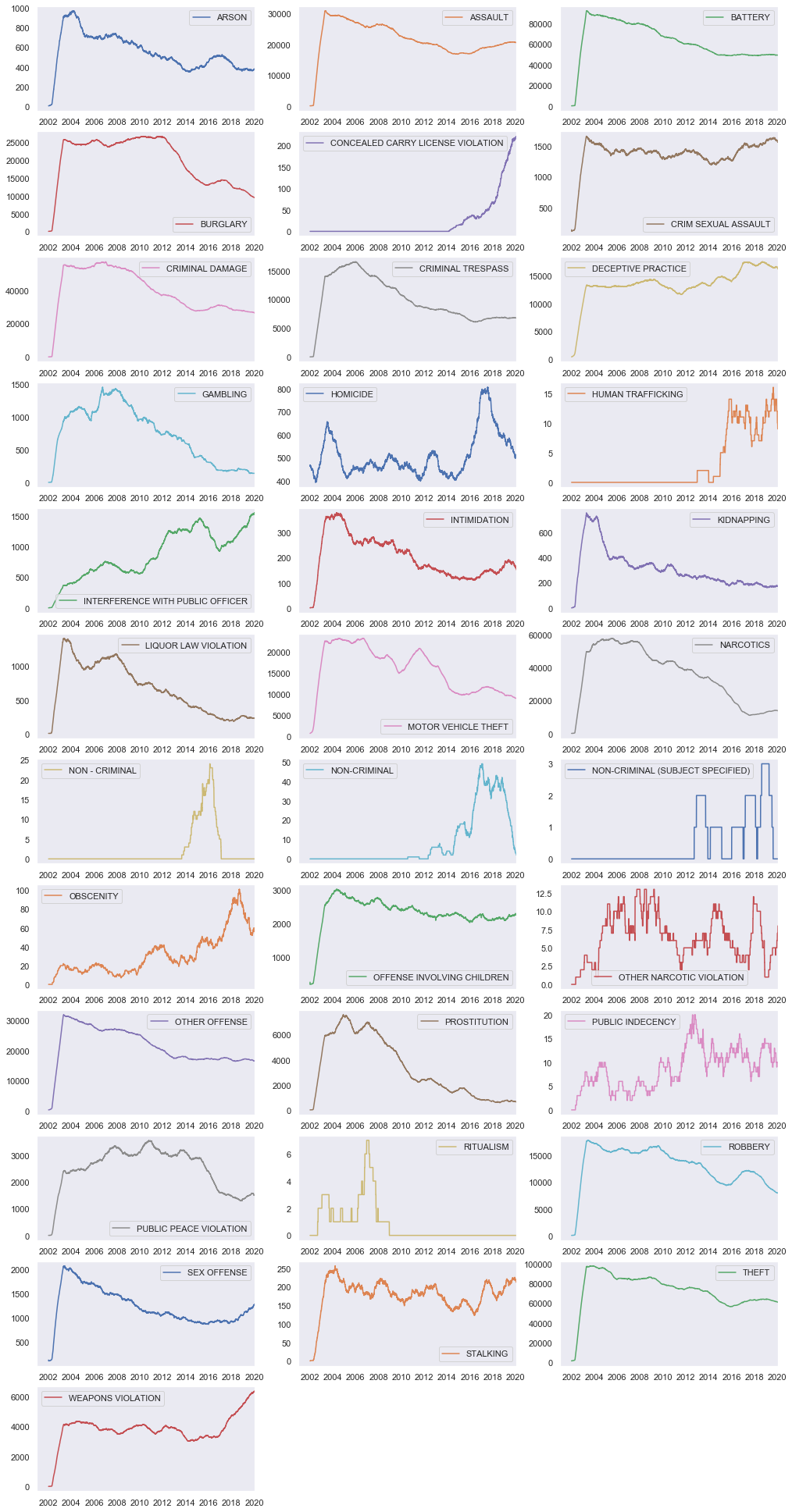
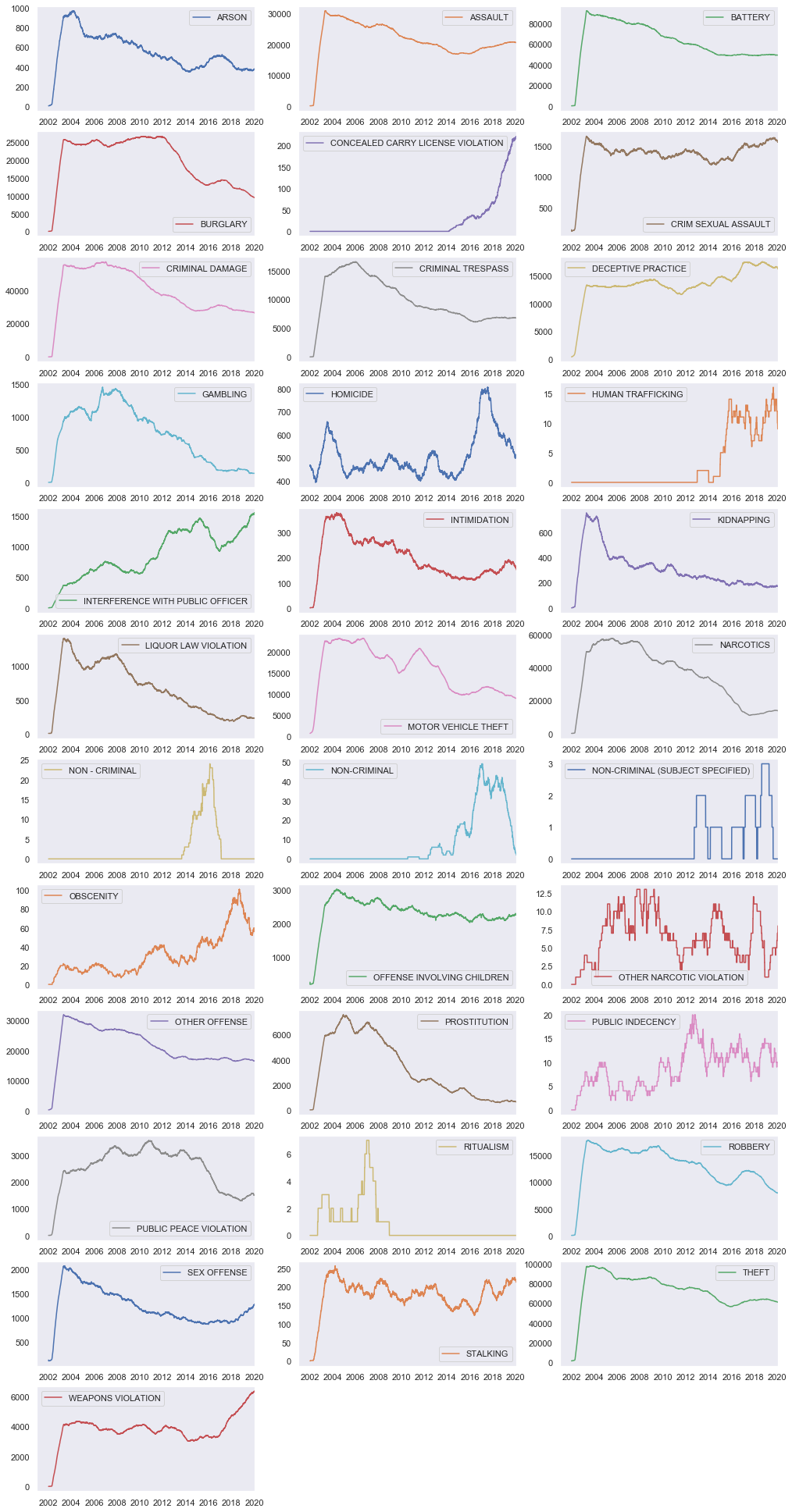
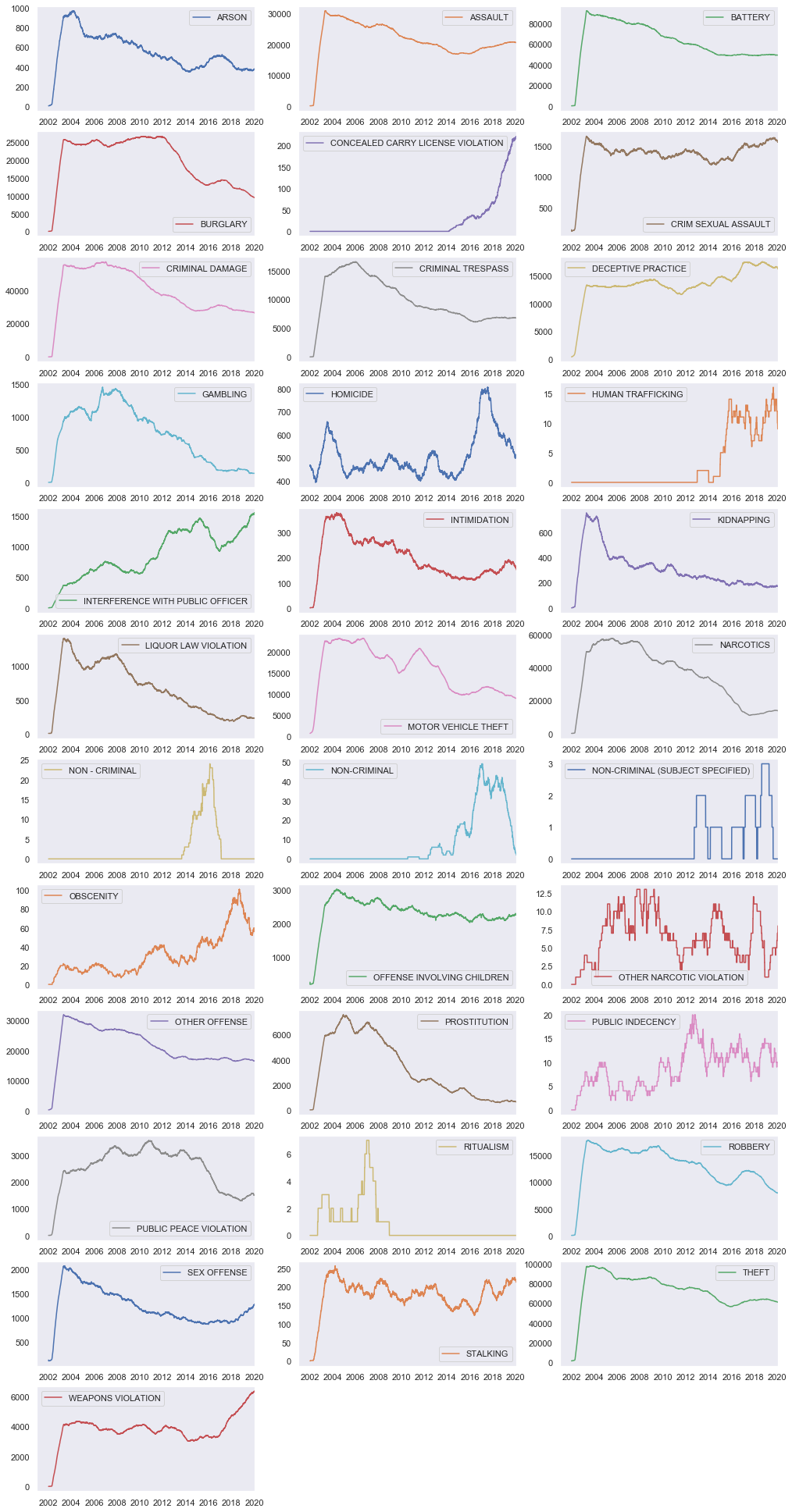
DATA 606 Delivery 3

**Summary Report**

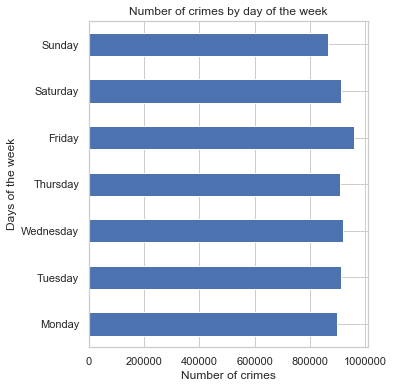
**Delivery 3**

**Exploratory Data Analysis:**

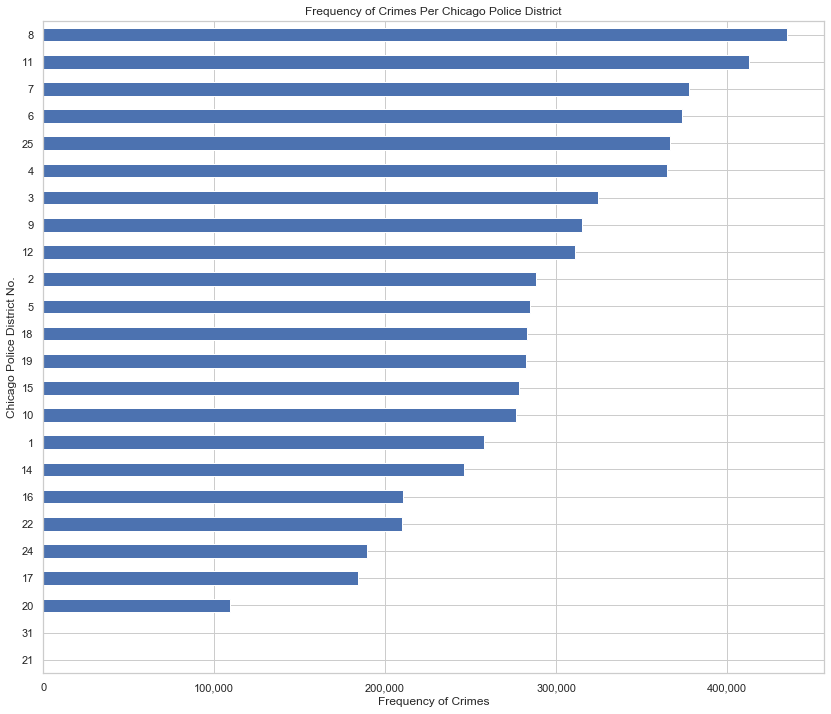
I did an extended exploratory data analysis which is a continuation to second deliverable’s EDA. My new extended EDA has produced five new visualizations, Firstly I plotted a graph to see how each crime has evolved over the years from 2001 to 2020. Out of all the sub graphs that I observed in that graph I noticed that Weapons violation, Concealed carry license violation and Interference with public officer are the crimes that got increased over the years linearly and reached there maximum in recent years. You can observe that in the below picture.



I continued my analysis by plotting a graph to answer this question: On which day of the week are more most no. of crimes committed in Chicago?. And the results were as follows:



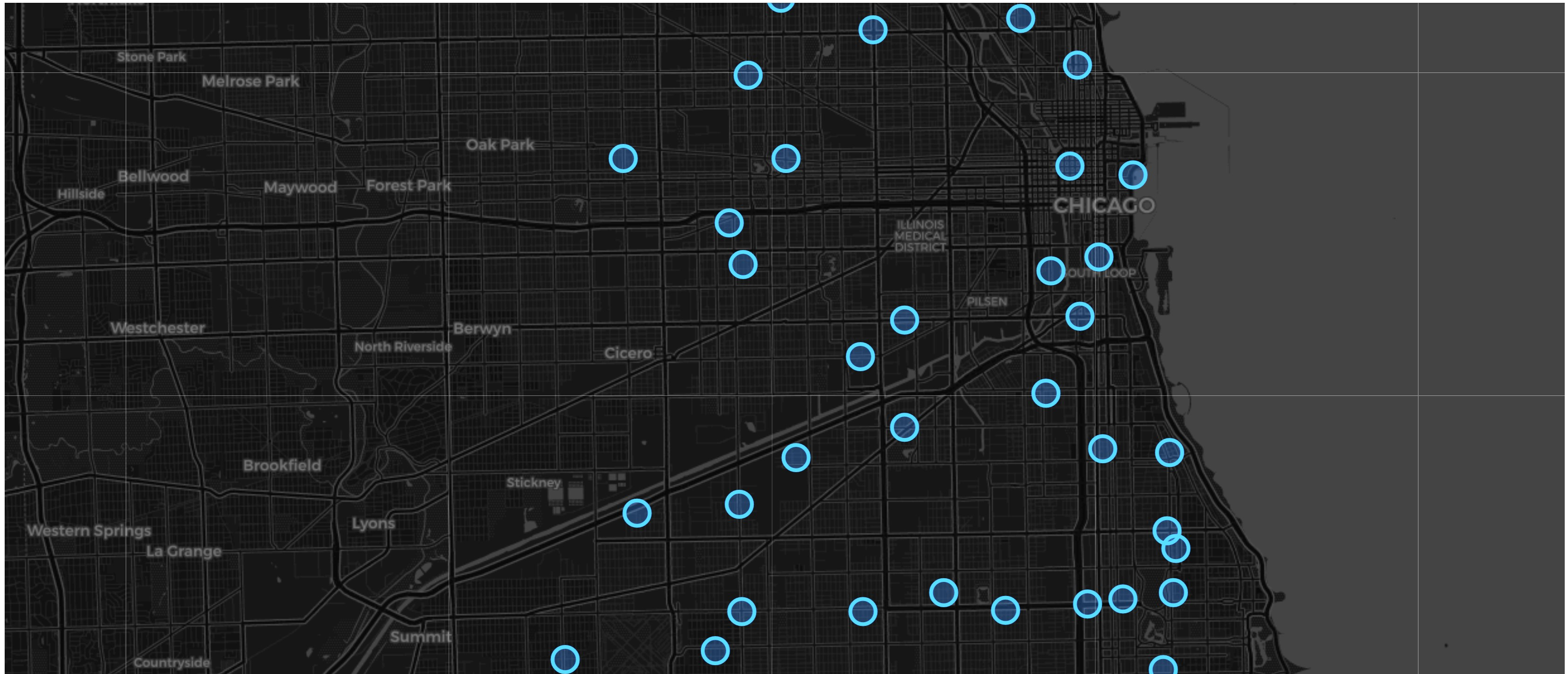
As per the graph we can say that number of crimes committed are distributed evenly through out the week but On Friday more crimes are committed when compared to other days in a week. I also plotted a graph to see how the crimes are distributed across different districts in Chicago and it came out like this:



From the above graph we can infer that District no.8 has more crime frequency than all other districts in Chicago which makes it the dangerous place to live in terms of crime in Chicago. Districts- 20,21,31 are the safer places to live in Chicago according to the graph as they are having lower crime frequency.

**Crime Mapping:**

Using the longitudinal and latitudinal location data I have implemented crime mapping using a package called folium. It gives us a interactive window as output in which we can zoom in and zoom out to see the crime spots that are plotted on the map. The image looks like following:



The blue circles in the map image above show us crime spots which are plotted on the Chicago map using the longitudinal and latitudinal data.

**Data Cleaning, Feature Engineering and Data Wrangling:**

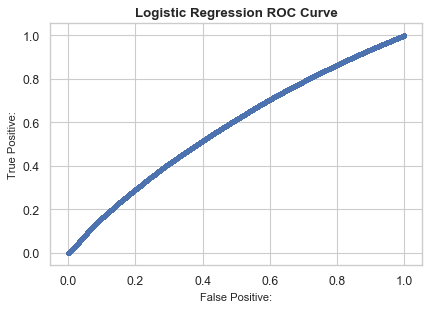
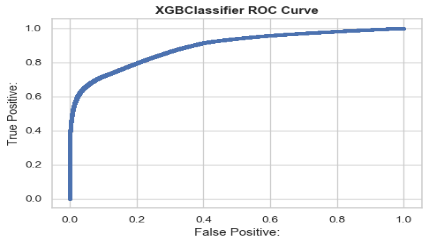
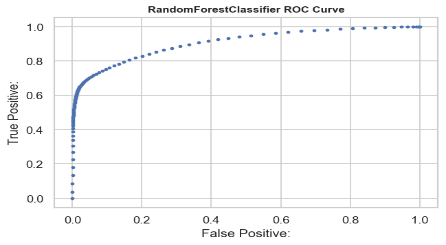
I made sure that the data is not having any inconsistencies, missing values and null values before starting the Feature Engineering. I then decided to focus on the required features, so dropped the columns which are not required for modelling. The columns that I have decided to include are Date, Primary Type, Arrest, Domestic, Beat, District, Ward, FBI Code, Latitude, Longitude. I felt that there is some scope for data wrangling for Date feature, so I have converted it from the given raw form to required form for analysis.

The next step that I took is Target selection, I initially chose Arrest feature as my target variable for this deliverable. But in next phase I am planning to work with two target variables by using multiclass multilabel classification algorithms to predict Primary Type of the Crime and Location of the based on other features.

After target selection I had split the data into train data and test data using train\_test\_split from scikit

**Modelling:**

In the final phase of this deliverable I used three relevant models and compared their results to find out which model is the best fit for predicting the bivariant target variable. Out of all three models **RandomForestClassifier performed well with an accuracy of 88.9 percentage. Logistic regression got an accuracy of 72.6% and XGBoost Classifier was almost at par with RandomForestClassifier model with an accuracy of 87.3%. Following are the ROC curves for three different models:**



**References:**

\* X. Zhang, Z. Hu, R. Li and Z. Zheng, "Detecting and mapping crime hot spots based on improved attribute oriented induce clustering," 2010 18th International Conference on Geoinformatics, Beijing, 2010

\* https://blog.dominodatalab.com/creating-interactive-crime-maps-with-folium/

\* McClendon, Lawrence & Meghanathan, Natarajan. (2015). Using Machine Learning Algorithms to Analyze Crime Data. Machine Learning and Applications: An International Journal. 2. 1-12. 10.5121/mlaij.2015.2101.

**Github and Video links:** [Video](https://youtu.be/2BUs4_5OsPk) & [Github Repository](https://github.com/YM53858/ChicagoCrimes/tree/master/Delivery%203)