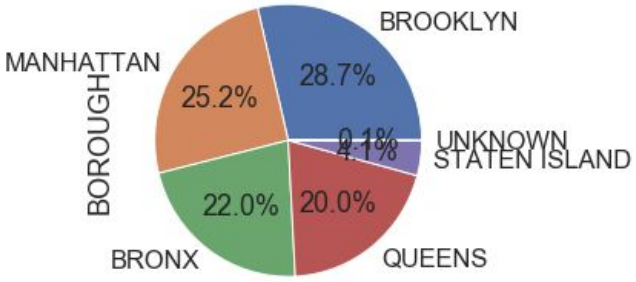
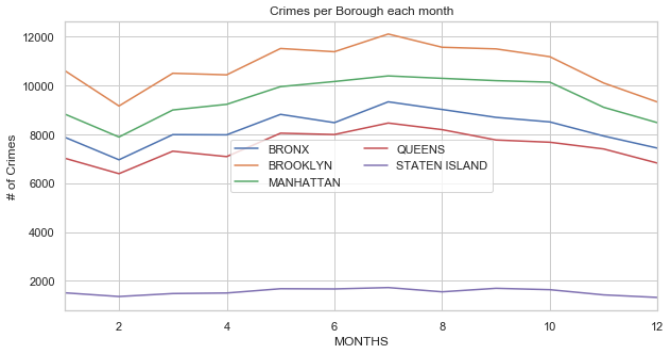
Data Storytelling
Capstone 1 Project: Crime Rates in NYC
Github Repository:
https://github.com/elizabeamedalla/Capstone-1/blob/master/Data%20Analysis.ipynb

1. Can you count something interesting?

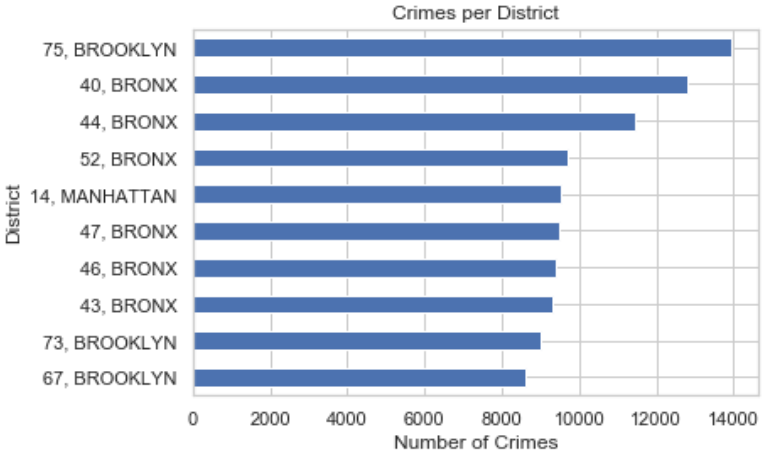


One of the initial approaches to see the value count of crimes in each borough. Next, using the pie chart, we see the ration of crime in each borough to overall crime reports. We counted the ratio of crimes in each borough to the total crimes reported and did a pie plot. Brooklyn had the highest reported crimes but it was not too far of a difference from Manhattan, Bronx, and Queens.

Next, we need to inspect the trend of crimes reported every month in each borough. Is there a peak season for crimes? When is the crime reported lowest? To see the bigger picture, I made a time series plot for all the boroughs in each month

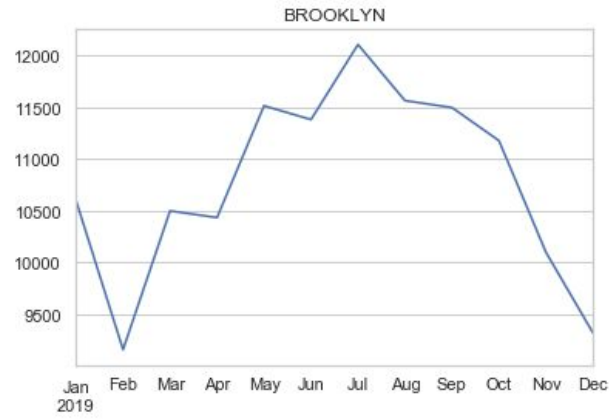


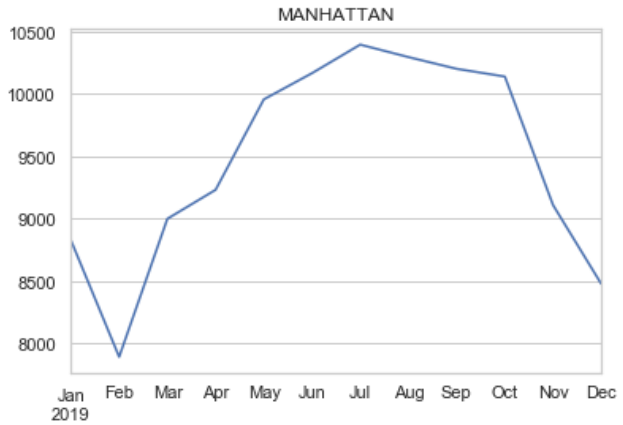
NYC is one of the most densely populated cities in the US so no doubt, crimes are reported in all parts of NYC. Now that we have the crime data for each borough, I want to break it down and see the top 10 districts with the most crimes reported.

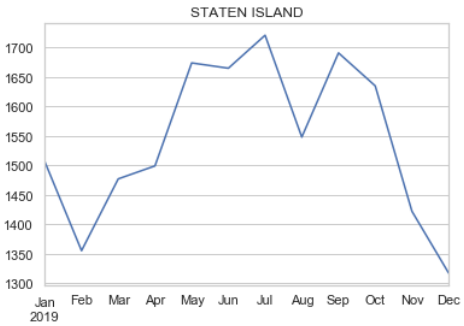


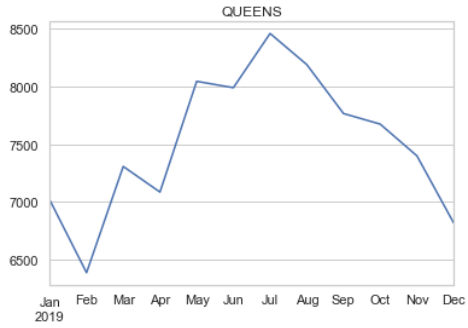
2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?

I wanted to check the crimes reported in each borough per month. I made a time series in each borough. All boroughs have a noticeable drop rate in February. Summer seems to be the peak season for crimes, starting in May, and the highest and dropping starts in October. The highest crime reported happened in July and the lowest in February and December.

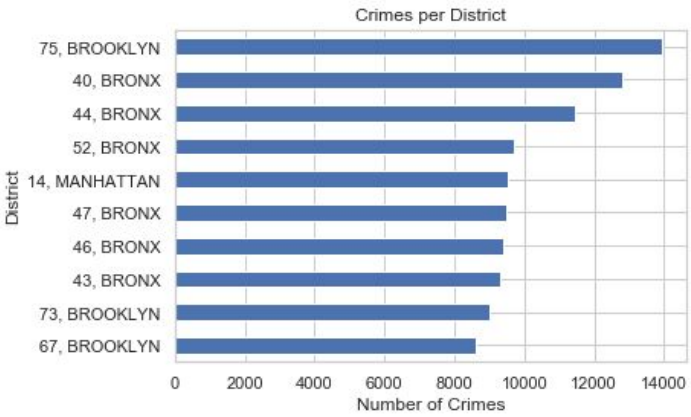








3. Can you make a bar plot or a histogram?



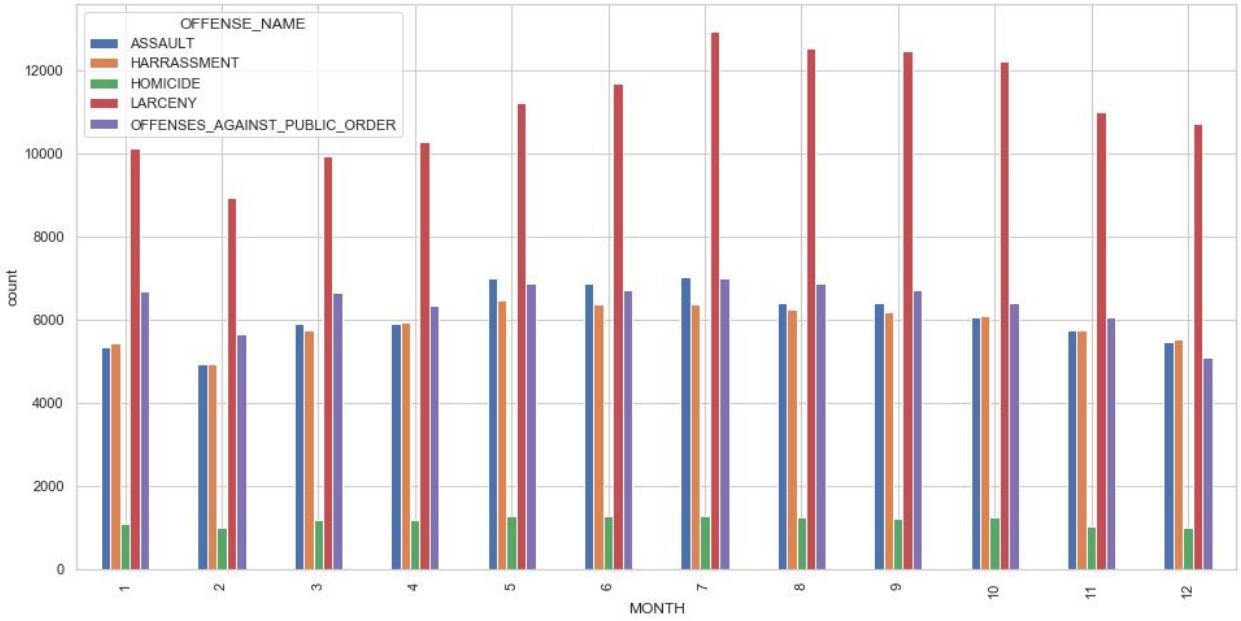
NYC is one of the most densely populated cities in the US so no doubt, crimes are reported in all parts of NYC. Now that we have the crime data for each borough, I want to break it down and see the top 10 districts with the most crimes reported. I created the bar graph in each district with most crimes and to plot the top 10 most reported crimes. To get this, I filtered the dataframe in each district and concatenated the borough in each district. Then, did the horizontal bar graph.

This is when it is concluded that the majority of the crimes reported are violent crimes in districts of Brooklyn and the Bronx.

4. Can you compare two related quantities?

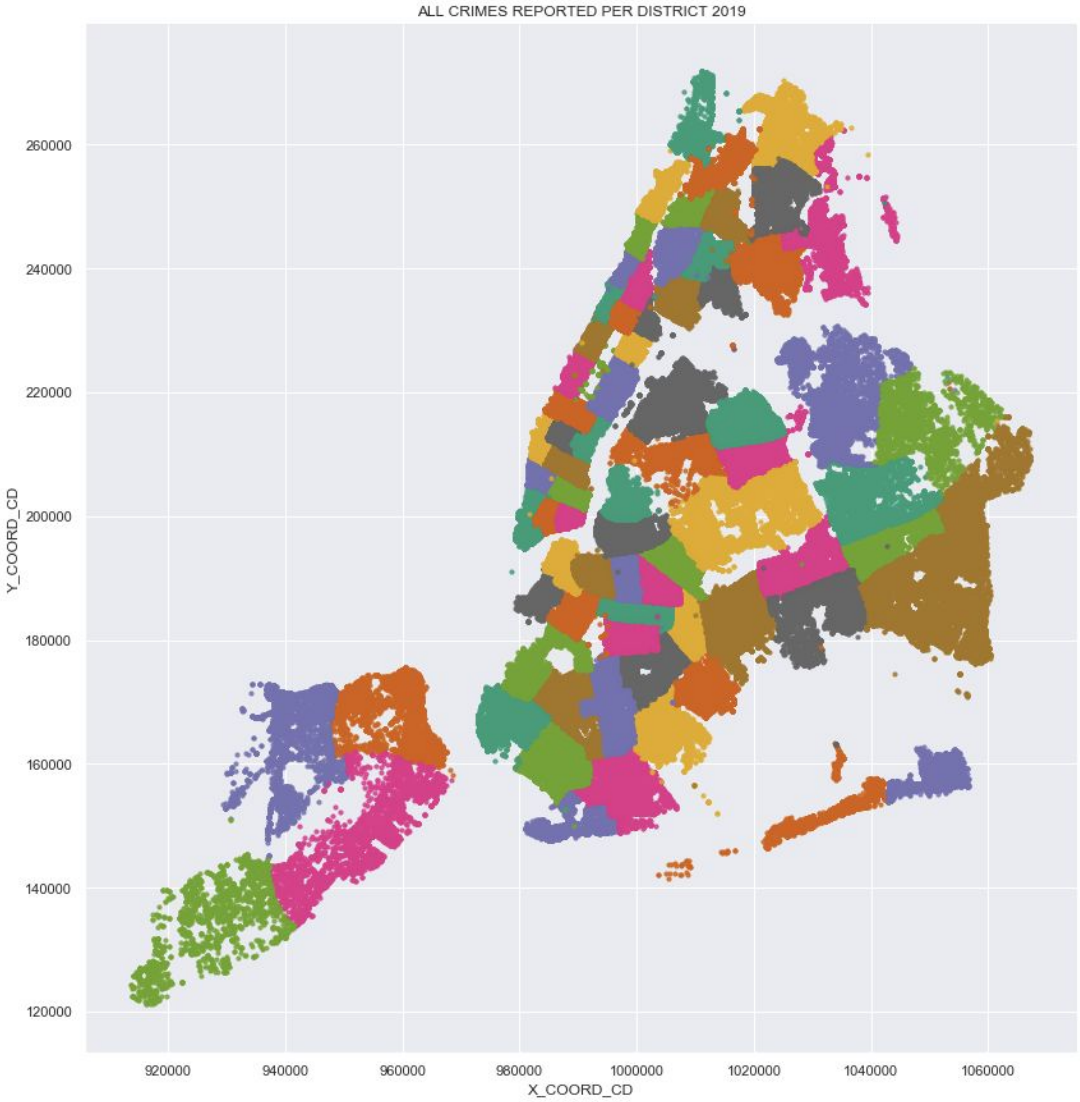
The comparison that was made is the level of offense committed. Misdemeanor is the top offense level of the crimes reported, around 20,000 crime reports.

I would like to see the distribution of crimes reported each month but due to various of crimes being reported, it will be huge data to plot. We can see this minute scale by plotting only the top 5 most crimes reported each month.



3. Can you make a scatterplot?

To visualize the crime data per district, we can use sns.lmplot to make the scatter plot of crimes reported per district. Just to clarify, this is not the exact map of NYC. This is a scatter plot based on the exact location of the occurrence of the crime reported per X and Y coordinates of the crime data. There might be parts of NYC that are not dotted here since no crime was reported.



4. Can you make a time-series plot?

Though I already made a time series of the frequency of crime reports in each borough per month in question #1, I also plot the frequency of all reports each month in one graph.