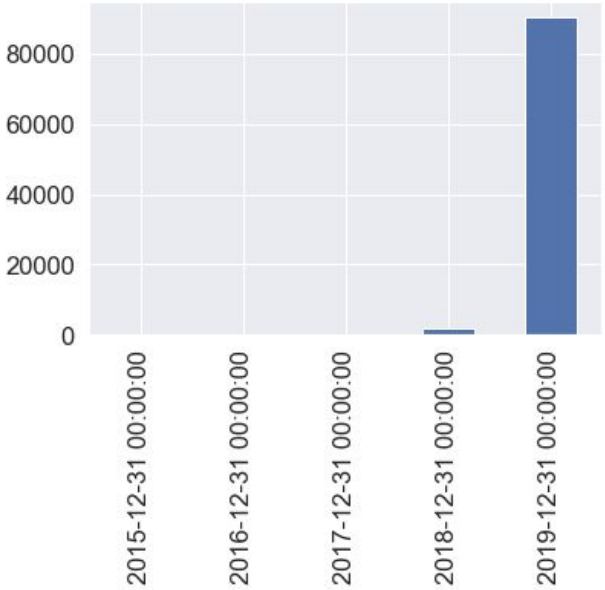
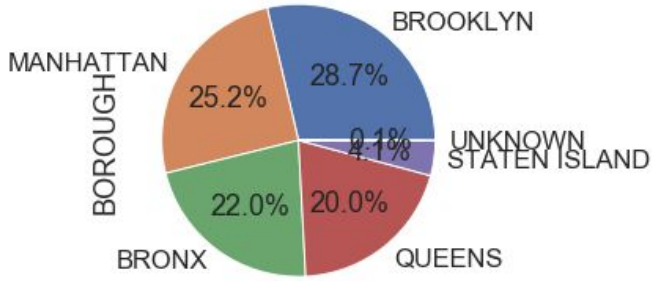
Data Storytelling
Capstone 1 Project: Crime Rates in NYC
Github Repository:
https://github.com/elizabeamedalla/Capstone-1/blob/master/Capstone%201.ipynb

1. Can you count something interesting?

   One of the initial approaches to see the dataset is to plot the total crimes each year. It is then figured out that the only sufficient data is 2019. The rest of the project is now focused on data exploration in crime records in 2019.

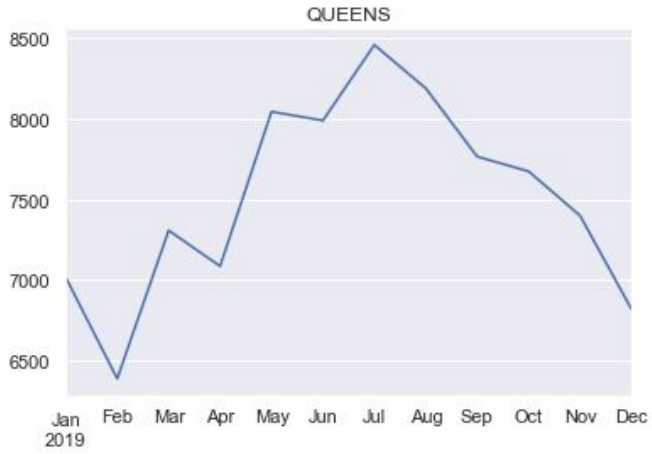


Then, we counted the ratio of crime in each borough to the total crimes reported and did a pie plot. The highest crime frequency happened in Brooklyn but not too far from a difference to Manhattan, the Bronx, and Queens.



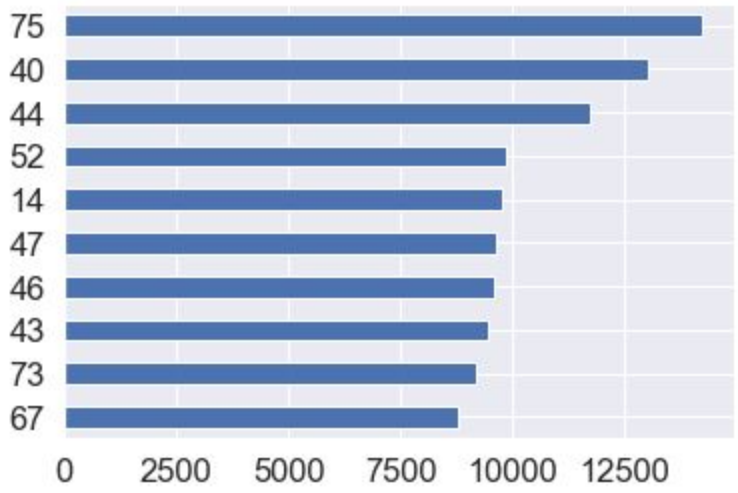
2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?

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.



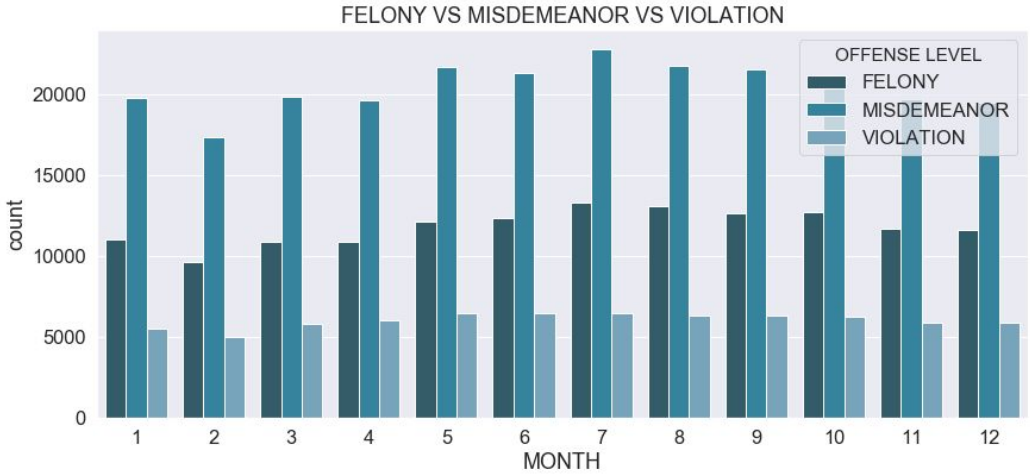
3. Can you make a bar plot or a histogram?

The majority of the plots made are bar graphs. 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 borough and did the df.value_counts(). 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 Manhattan and Brooklyn.

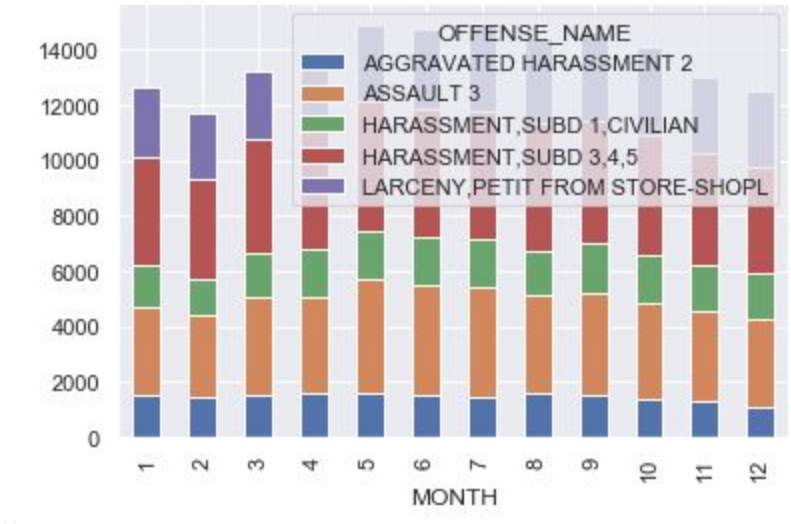


4. Can you compare two related quantities?

The comparison that was made is the level of offense committed. Misdemeanor is the top offense level of the crimes reported.



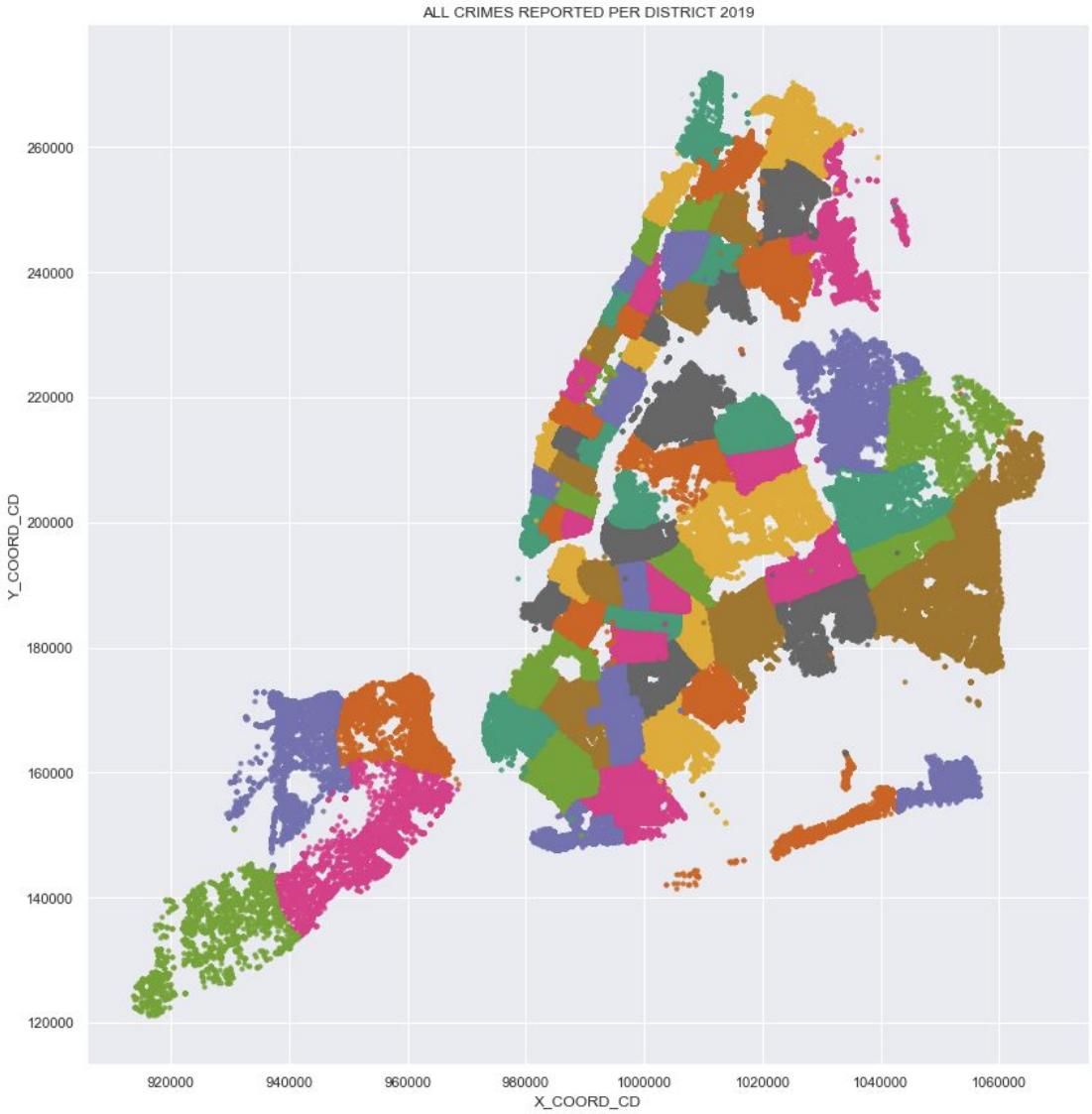
I used stacked bar graph to plot the distribution of the top 5 crimes each month. Harassment 3, 4, 5 is still leading.



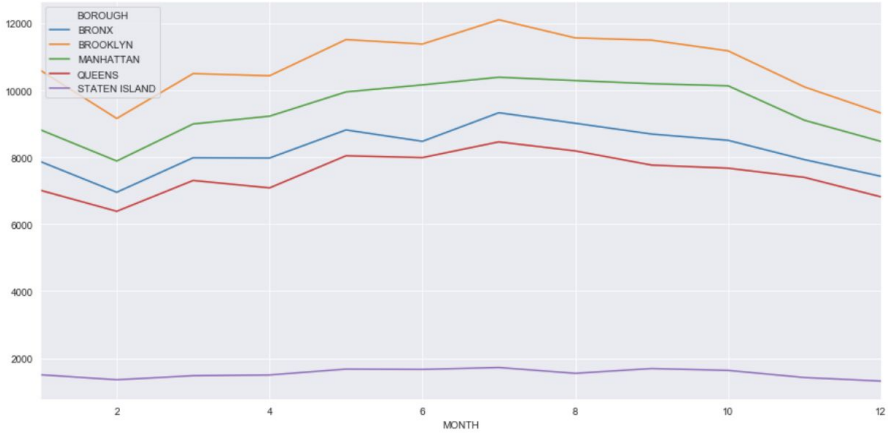
5. Can you make a scatterplot?

   A Scatter plot was made for the location of the reported crime using the X coordinates and Y coordinates columns of the crime data. I used sns.lmplot to

plot the columns.



6. 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, I also plot the frequency of all reports each month in one graph.