**Kevin Diaz Gochez**

Analysis of the possible effects of complaints on arrests in New York City

Everyone has made mistakes in their lifetime, but sometimes we don't realize the implications a simple complaint can have on someone's life. Therefore, it is imperative to understand whether a relationship exists between arrests and complaints filed.

The objective of my project is to examine two data sets released by the NYPD regarding complaints and arrests between the dates of '2019-01-01' and '2020-03-31 to better understand the relationship between these two variables. The data sets were obtained from the NYC open data website and I used an API for each data set to obtain fresh data from the URL. Within the API, I directly queried a count for both complaints and arrests, and I renamed the date columns which are already in "DateTime" format. However, for the sake of the project, I decided that instead of reading the CSV files created by the API, that I would instruct the reader to read an edited CSV file where I have reset the date format, changed the column name for the date, and I have removed data points randomly. After reading the edited CSV files into the notebook, I dropped entire instances that contained null data values because in the context of my problem it would not make sense to try and fill the null values with 0 or to try to get a predicted replacement value. Then I separately converted the dates in the two CSV's into "DateTime" values which I could merge on in the future. After merging the two data sets, I plotted a regression plot using seaborn which showed a linear relationship that was both positively correlated and showed a somewhat strong relationship. However, several plotted points show a large error term. Overall, the plot demonstrates that generally a larger amount of complaints in a day would also lead to a larger amount of arrests.
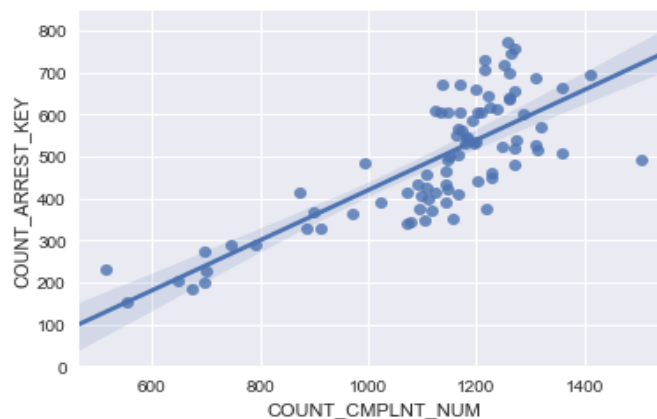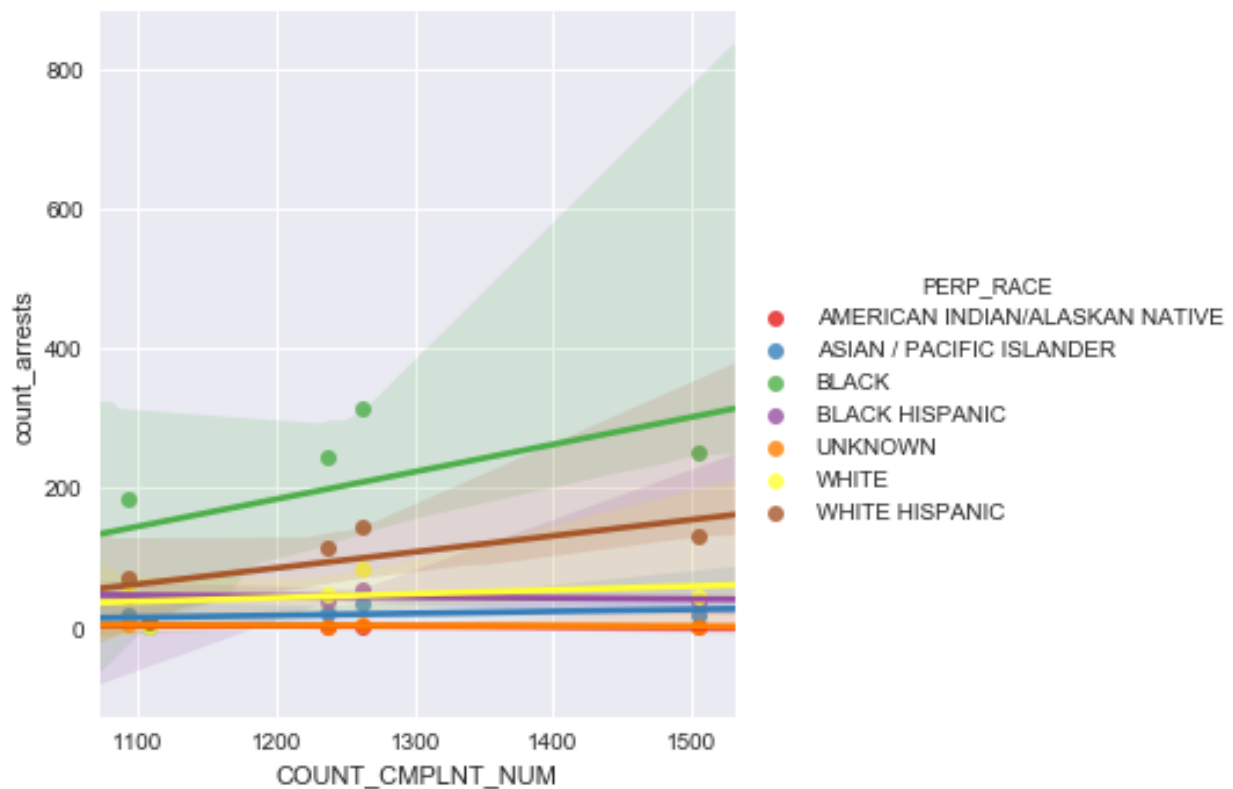
However, I was not satisfied with my answer and my new objective was to try and determine whether gender could play a role in the strength of the relationship between arrests and complaints filed. I created another API to have a field for gender, this time I did not require downloading any external CSV files because I had already demonstrated cleaning/manipulation of the data beforehand. After merging, I plotted the values with a moderating effect for gender and was surprised at my results. There were now two regression lines rather than one, both demonstrated positively correlated linear relationships. However, the slope for males was much steeper than for females but the strength of the correlation was stronger for the females. In context, this means that when complaints are made for females there is a higher likelihood that it will lead toward an arrest than if the complaint were made for a male. Perhaps there are a couple of reasons why this could be the case, there are generally more data points for the males within the two data sets which could explain some of the variations. On the other hand, a possible explanation may be that males are more likely to have complaints made on them despite the severity of the situation. Therefore, when a female has a complaint made on her it is likely to be for situations of higher severity.
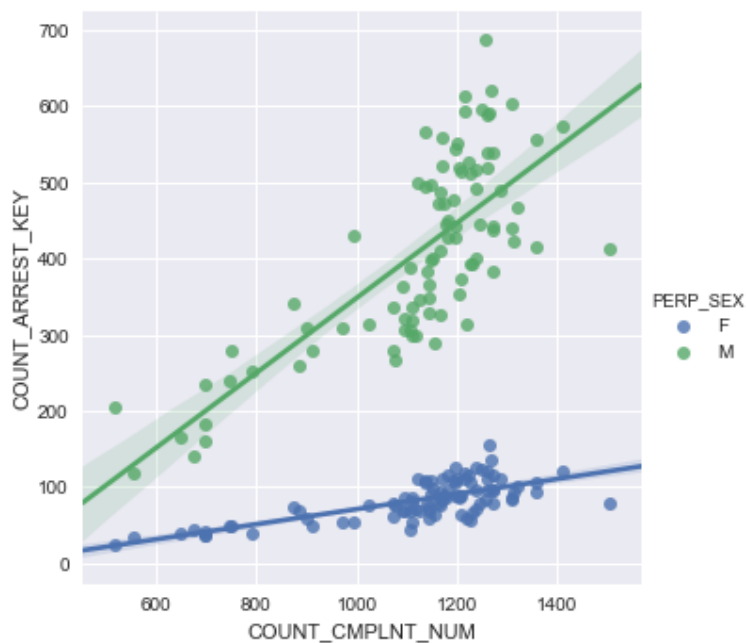
Lastly, I wanted to consider the implications that ethnicity could have on the relationship between complaints made and the number of arrests. I created a new API that would pull for new

criteria dependent upon race. However, I did not include an aggregate function for counting within the API as I did previously because I wanted to utilize the aggregation method demonstrated in class. I was surprised to find the disparity between ethnicity within my results, there is an obvious upward trend in the case of African Americans. My results show that they are more likely to be faced with complaints and arrests than other ethnicities represented in the sample. However, they are also the group with the highest room for error. The confidence bands surrounding the slope for African Americans are much more wide indicating that the correlation between arrests and complaints made is not as high as the other groups. These results could be skewed by the disproportionate amount of arrests made on the African American population in general, which in turn may be decreasing the strength of the relationship between complaints and arrests.

Works Cited:

https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc/data

https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243