Investigation of Crimes in NYC
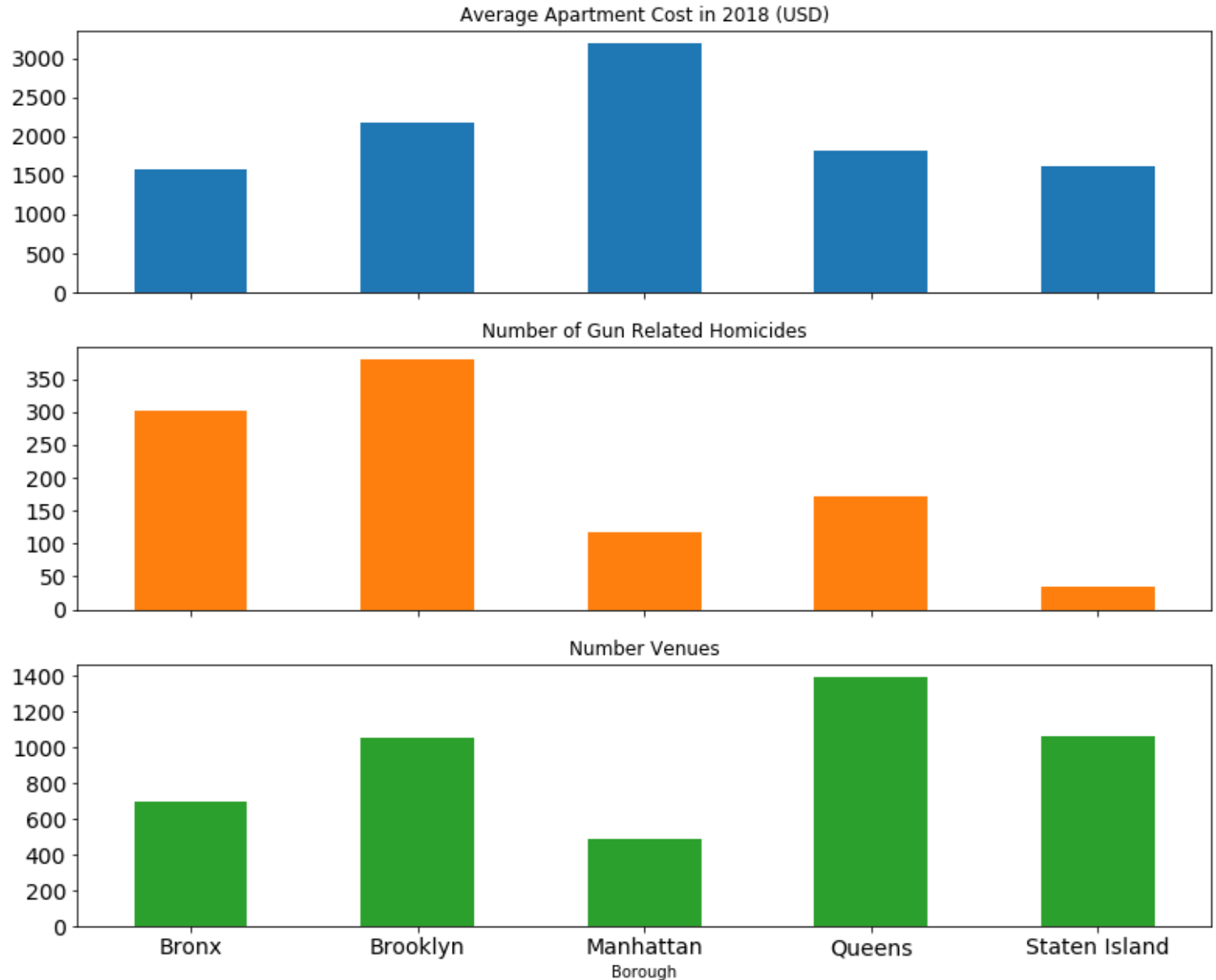
Predicting High Levels of Crime

- Crime is inevitable in larger cities due to the larger population
- Most people who are looking to live in a new city are unfamiliar with the location and aren't sure of where it is good to live
  - Most people are also looking for cheap locations to live which is most likely a place for higher activities of crime
- New Start-Ups also have interest on where it might be a good place to set-up shop
- Law Enforcement could also benefit by knowing different locations of high crime

Data Acquistion

- Shooting Homicides was found on public [Tableau](#) webpage

- Neighborhoods in NYC was found on [Wikipedia](#)

- Different Apartment Costs were found on [Streeteasy](#)

- In total, there were 144 different neighborhoods that were looked at with 3 main features investigated here
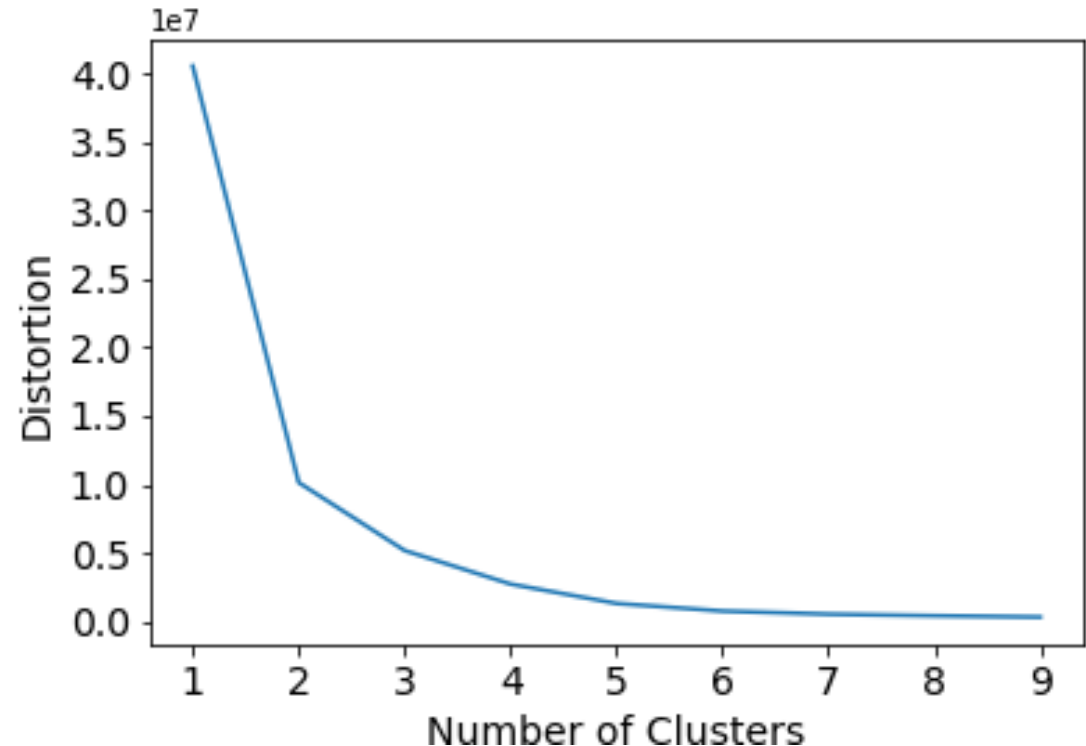
Investigation of Boroughs

- The cost of the apartment was seen to match somewhat the number of crimes committed
  - Lower apartment cost led to higher number of crimes
- It is seen that a higher number of venues leads is seen with a higher number of related homicides
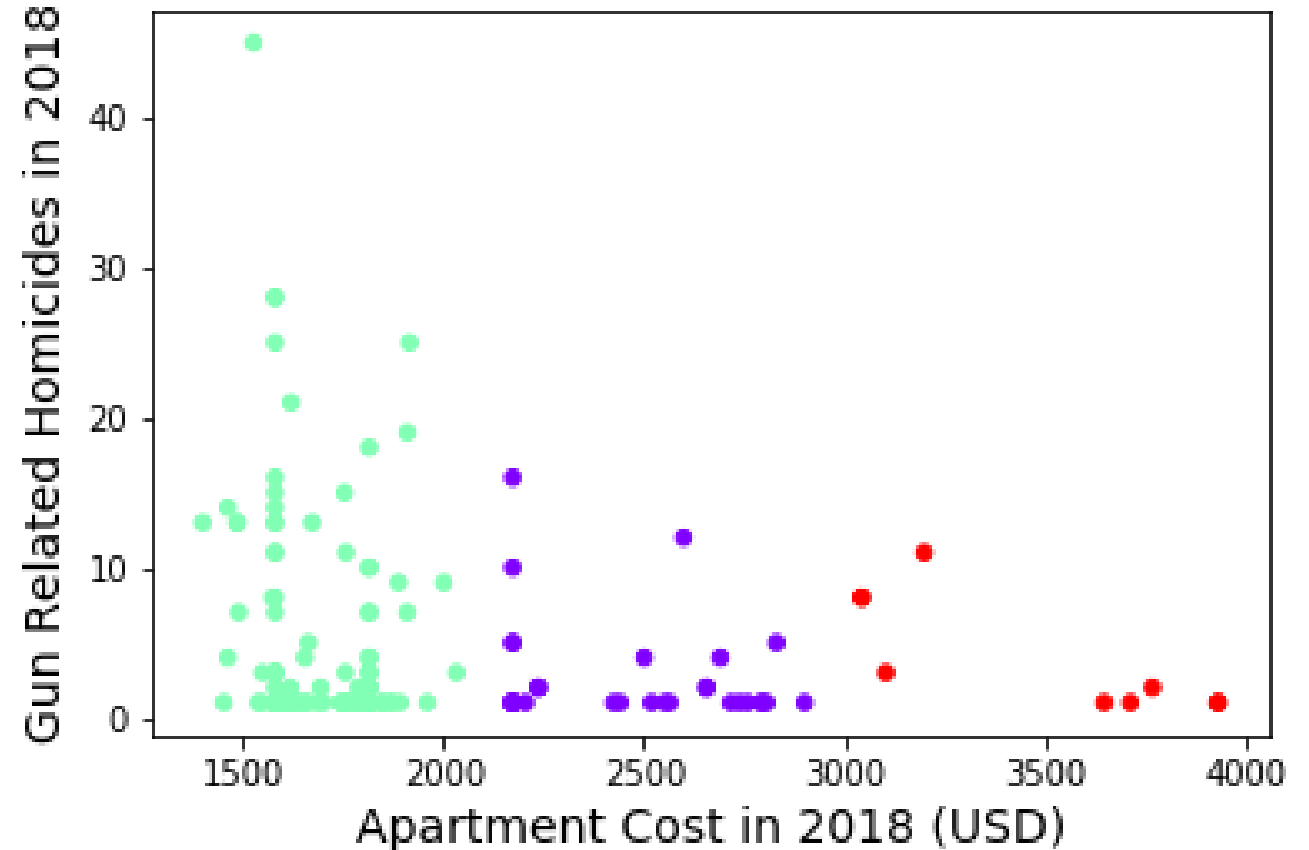
Kmeans Algorithm for Clustering

- KMeans Algorithm was used to determine different clusters of neighborhoods
  - Number of clusters are explicitly determined and fit to the data
- Optimal number of clusters are found using the elbow method
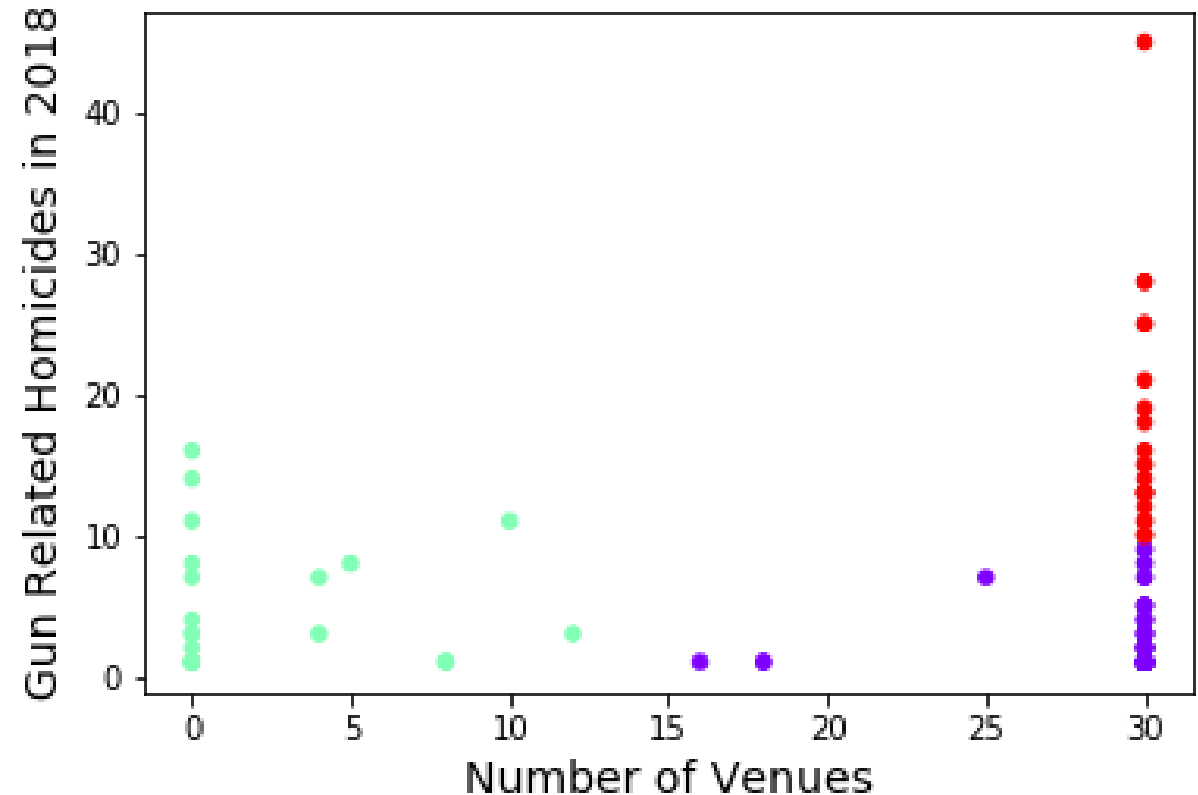 - Based on distortion of the data

Apartment Cost

- There are three separate groupings, that were found.

- The first cluster (red) where the apartment cost is very high and in turn the crime is low.

- There is a second cluster (purple) where the apartment cost is moderate, but the crimes are also moderate slightly more dangerous but safer.

- The last cluster is the lowest apartment cost. This is characterized with a large range of whether the area is dangerous.

Number of Venues

- It was optimally found that three clusters fit the data best.

- The red cluster shows neighborhoods that have both high numbers of venues and high number of homicides.

- The purple cluster shows neighborhoods where there are high numbers of venues but also having low numbers of crimes.

- The cluster in green shows neighborhoods with low numbers of venues and a moderate number of incidents.

Conclusion

- There is some correlation within each borough between both the cost of an apartment and the number of venues

- Looking more closely at each neighborhood, there was some correlation amongst the different variables
  - The higher the cost of the apartment the less likely there would be crime
  - The higher number of venues the more likely there is to be crime

- Future Directions:
  - Look at more variables such as population and population density in each of the neighborhoods
  - Look at multiple cities and see if the trends still hold true