Alcohol/Weapon Venues and Crime in Atlanta

Jonathan Guy

March 28, 2019

1. Introduction

I. Background

In some of the previous labs and assignments, we had the opportunity to work with the location data of New York and Toronto to make clusters based on venue categories in neighborhoods. Also, in prior sections of the course, we were required to work with and analyze crime data of Chicago and San Francisco. This inspired me to combine the types of data and use them for a project.

II. Problem

Combining these two ideas together(above) and applying it to my home city of Atlanta, Georgia, United States, I am going to attempt to determine whether we can find any correlation between the alcohol/weapon venue categories in a neighborhood and the number of crimes or types of crimes in the neighborhood. I will also perform exploratory data analysis as I work through the project and ideas come to mind. This project would be of particular interest to anyone who is curious about the relation between alcohol/weapon venues-types of a location and crime of that location.

III. Interest

The police department, venue owners, or social workers may also be interested in the results of this study as it may provide insights relative to their fields. It will be interesting to have an "unbiased" look; for example, when we are in a particular location we automatically make assumptions based on the appearance of the buildings, infrastructure, people, etc. However, this study will not so much be looking at the perceived quality of the venues, but on the pattern of venue occurrence. I may also look in to venue ratings once clusters are made to explore whether a lower average rating of venues has any correlation to crime.

Previous studies have shown that higher density of bars/liquor stores have higher rates of violent crime, so I will investigate whether this holds true for Atlanta.
http://resources.prev.org/documents/alcoholviolencegruenewald.pdf

2. Data Acquisition and Cleaning

I. Data Sources

I will use Wikipedia to get data concerning Atlanta neighborhoods and NPUs(Neighborhood Planning Unit, geographically based). Conveniently, the crime data I will use includes the latitude and longitude of the crime in addition to the neighborhood name and NPU, so I will take the mean lat/long for crimes in a given neighborhood to determine its coordinates(to be used with Foursquare API). Finally, I will use the open data csv files from the Atlanta Police Department to acquire crime data; for simplicity, I will only use the crime data for 2018.

Links to the data below:
https://en.wikipedia.org/wiki/Neighborhood_planning_unit
http://opendata.atlantapd.org/Crimedata/Default.aspx

II. **Data Cleaning**

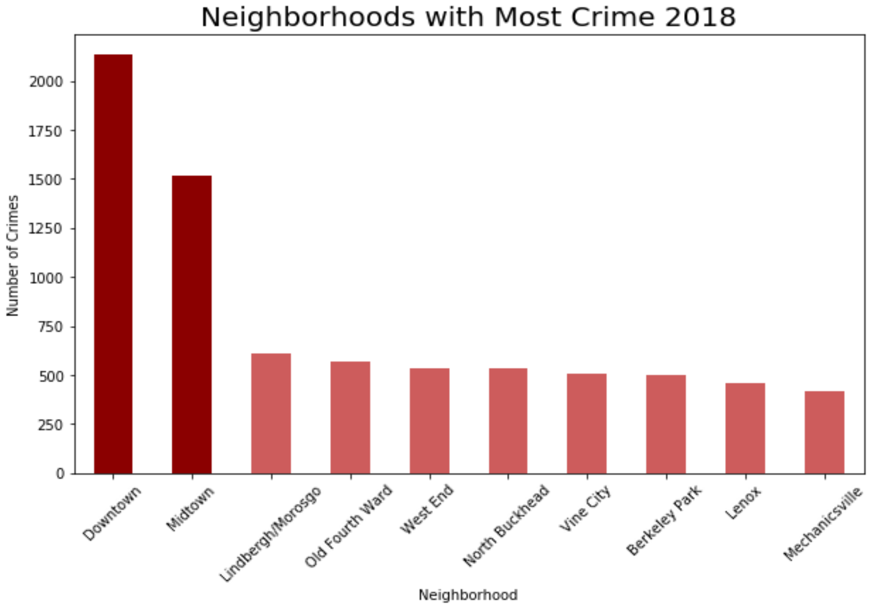
I manually created the CSV for NPU and neighborhood. Using 12 CSV's concerning 2018 crime, I first concatenated the data frames into a single data frame for all Atlanta crime in 2018. I decided I did not need the report number, date, location, beat, or unnamed: 9 columns for the study so I dropped those from the crime data frame. I then dropped any rows with NaN values. Finally, I searched for duplicate neighborhoods and dropped the neighborhoods with incorrect NPUs, and used this information to create an accurate neighborhood-information data frame with the name, NPU, latitude, and longitude. Since there were duplicates in the coordinate/neighborhood data frame, I repeated the process to remove incorrect duplicates from the crime data frame(crimes reported with incorrect neighborhood information). This gave me the crime data frame. Finally, using the Foursquare API and the latitude and longitude for each neighborhood, I fetched the location data for each neighborhood and created a data frame consisting of all the venues within a half-mile of the crime center of each neighborhood(latitude and longitude for neighborhoods determined by averaging the reported latitudes and longitudes in the crime report for each respective neighborhood).

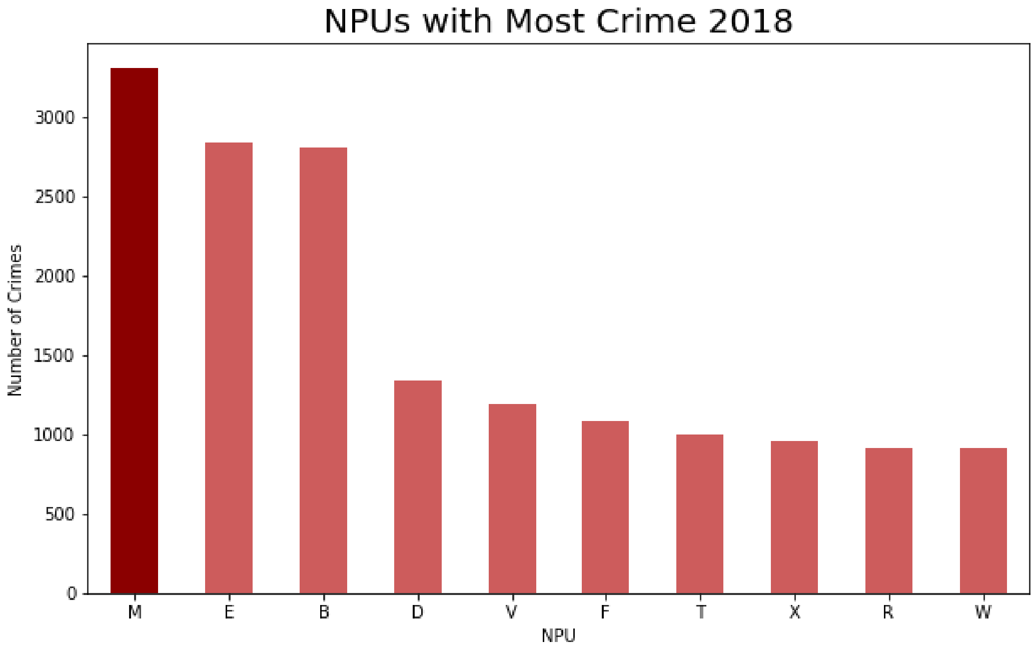
III. **Feature Selection**

To cluster the neighborhoods, I used a variety of combination of features including location, larceny crimes, violent crimes(homicide and aggravated assault), and robbery crimes.

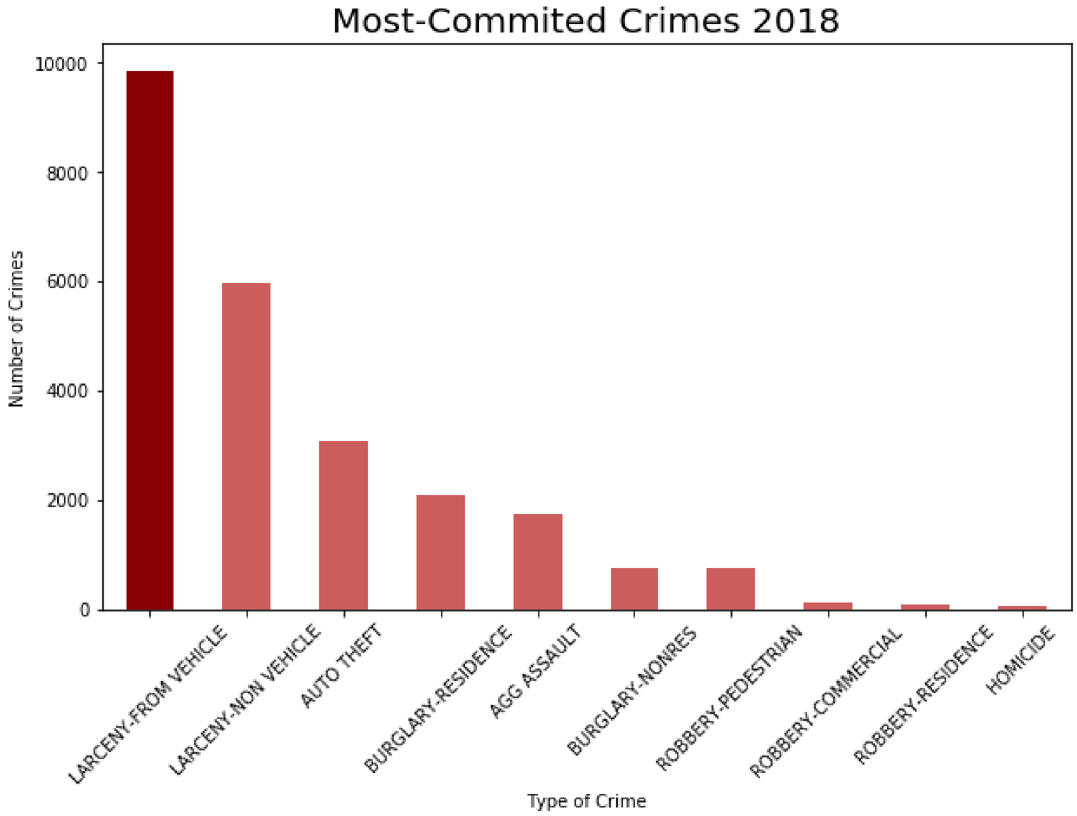
3. Exploratory Data Analysis

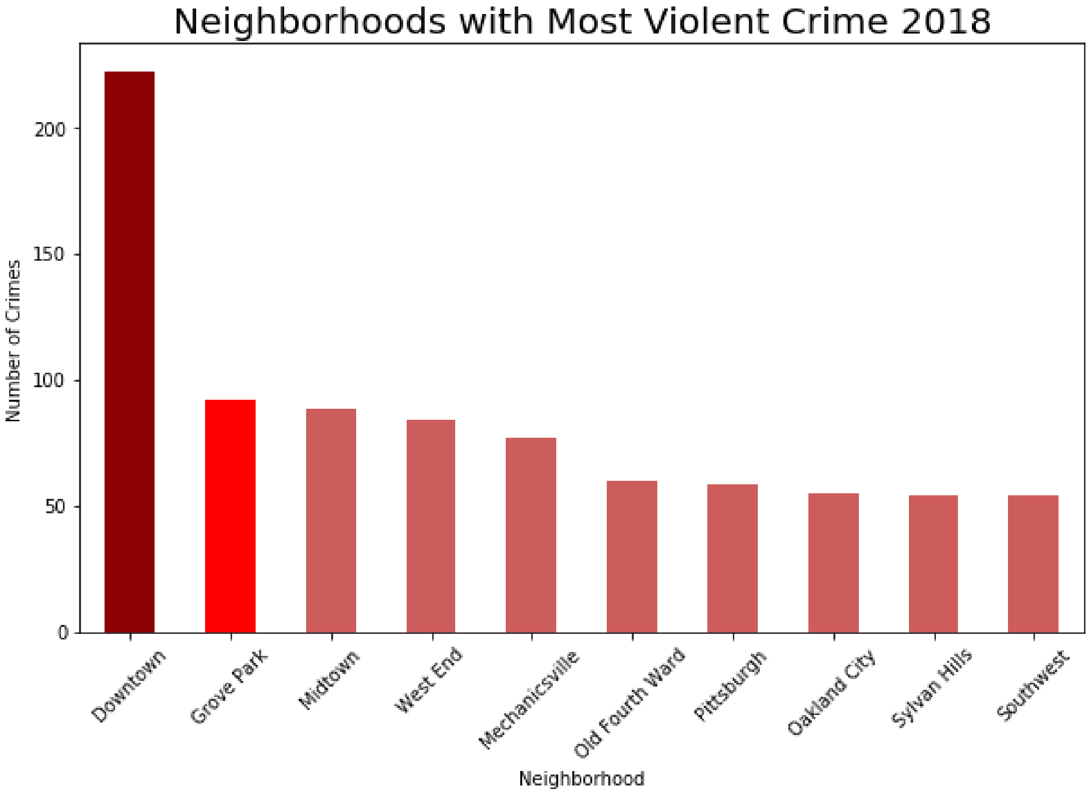
I explored and visualized many trends in the data, as seen below.



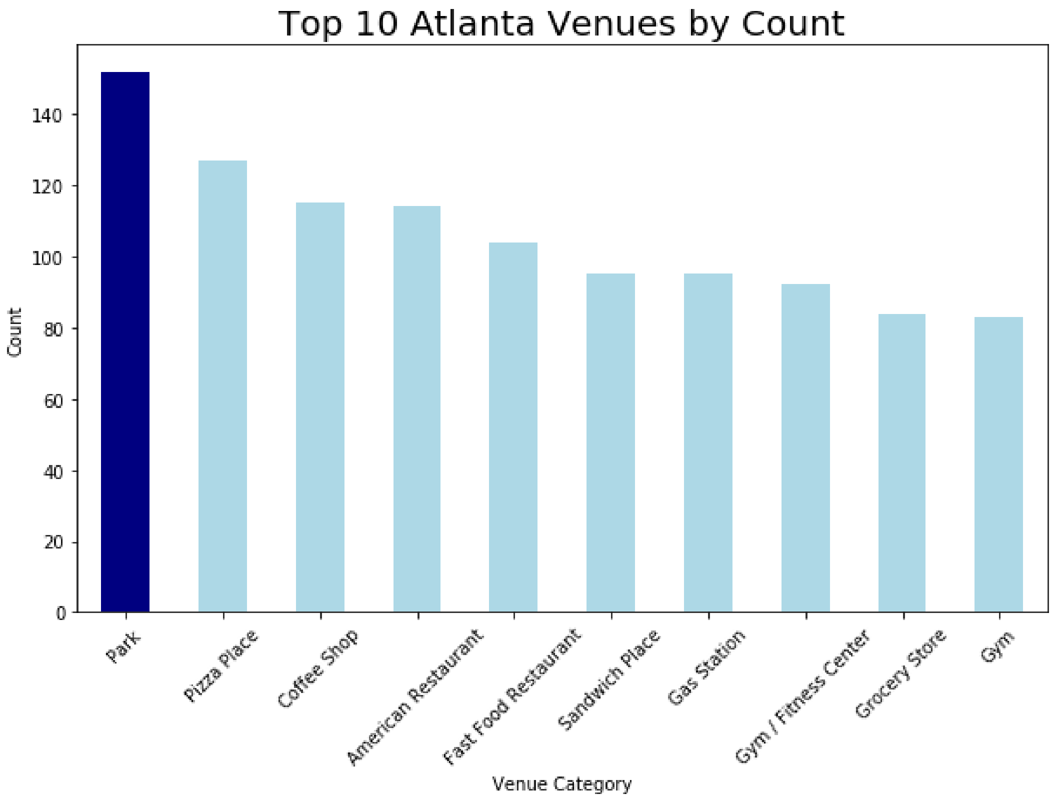


It is interesting to note that although 2 neighborhoods(Downtown, Midtown) distinctly separated themselves in number of crimes, the disparity between number of NPU(neighborhood conglomerates) crimes is not as great.





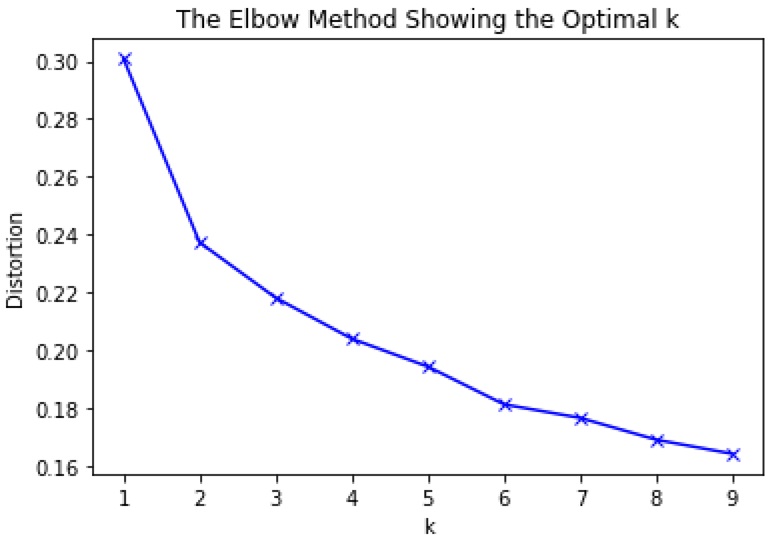
Grove Park, which was not even in the top 10 for total crimes, is the second-highest for violent crimes among neighborhoods.

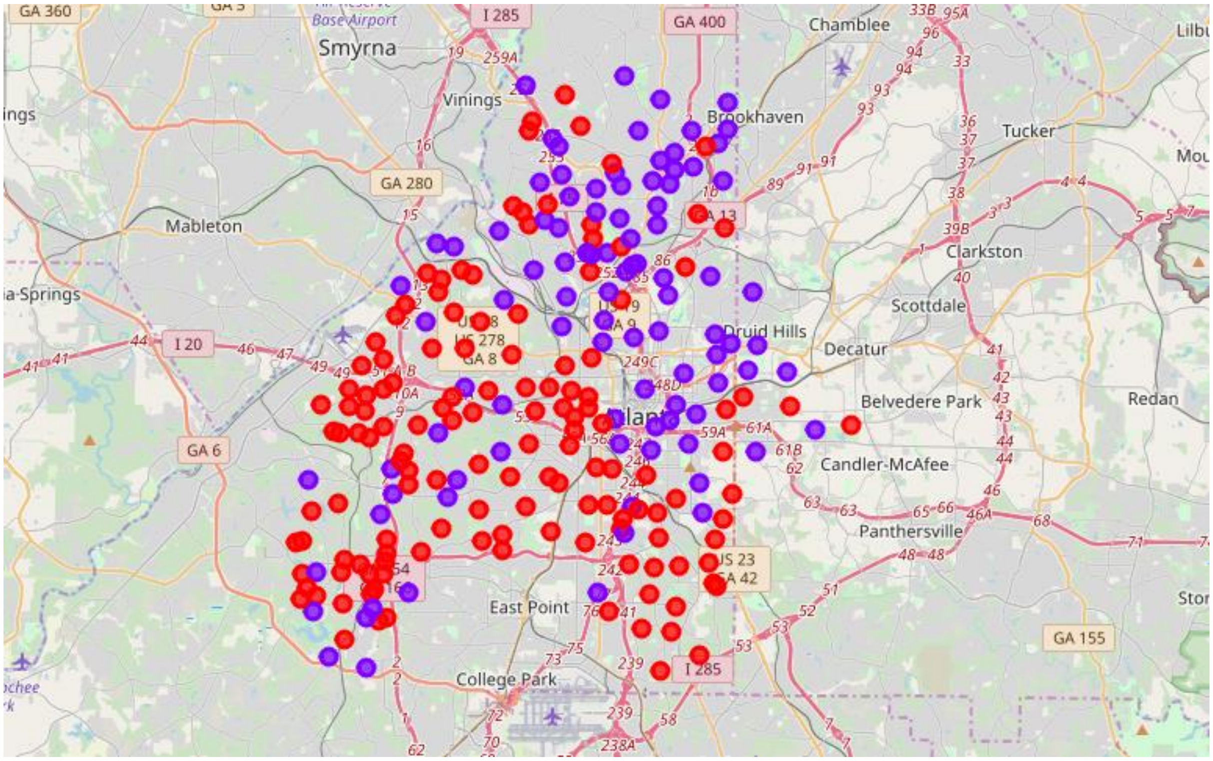


We can see that parks are the most common occurring venue in Atlanta neighborhoods, although there does not seem to be a huge disparity from one category to the next looking at the top 10. As someone from Atlanta, I did not realize how many parks there were! (Although there could be some double-counting in our data, as one park could be within half a mile of the crime-center of multiple neighborhoods, there are still more parks than I thought there were).

4. Cluster Modeling

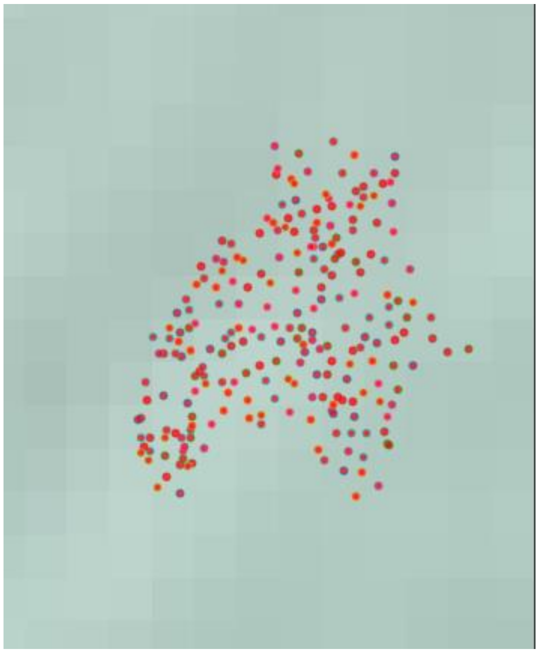
To experiment with different clustering methods, I used both KMEANS clustering and DBSCAN. I determined the best k value to be 2 and clustered the neighborhoods based on the mean frequency of each crime in the neighborhood. See KMEANS results below, showing the elbow method to determine best k value and the mapped representation of the clusters using folium.



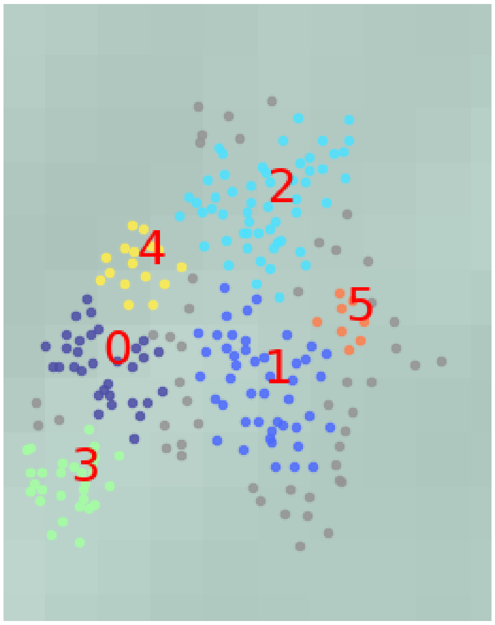


I also used DBSCAN to cluster neighborhoods by location, larceny crimes and location, violent crimes and location, and robbery crimes and location. Some iteration was required to determine the optimal epsilon and minimum samples, which differed for each variation of the DBSCAN clustering. This clustering method is what I ended up using to search for correlations between crime and venue types(specifically, alcohol and gun locations), and I used basemap to plot the clusters.

Original Plot Showing Neighborhood Distribution (note same shape of city although names, roads, etc do not appear).

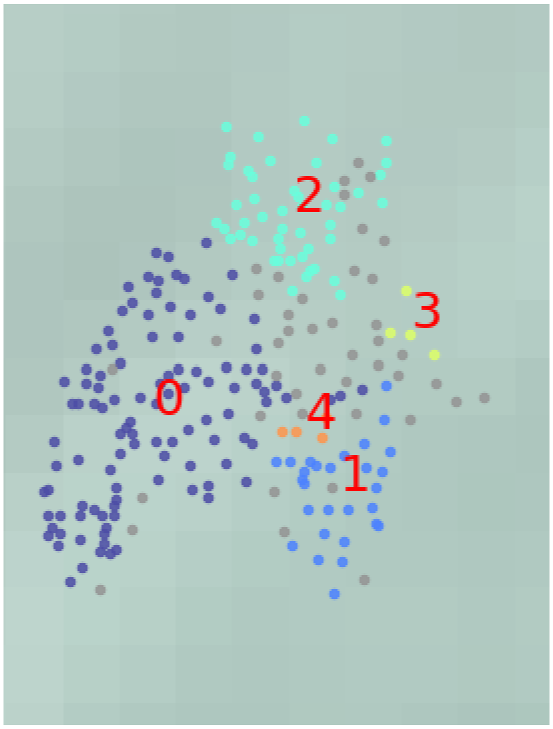


Clustering Based on Location (also outputting average number of total crimes per neighborhood in 2018 by cluster)



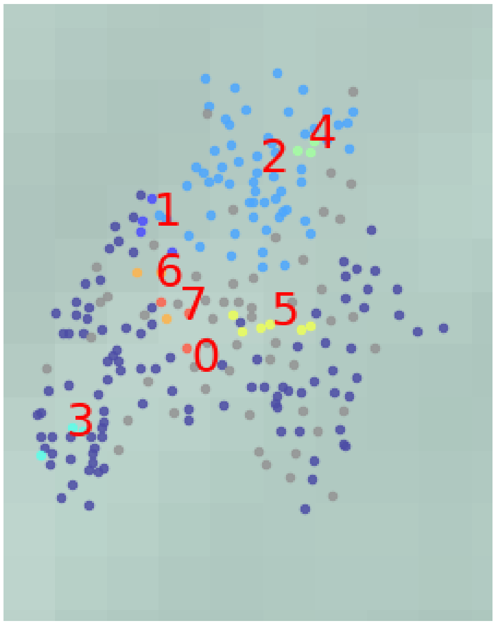
- Cluster 0, Avg Number of Crimes per Neighborhood: 31.75
- Cluster 1, Avg Number of Crimes per Neighborhood: 181.52173913043478
- Cluster 2, Avg Number of Crimes per Neighborhood: 100.41818181818182
- Cluster 3, Avg Number of Crimes per Neighborhood: 41.48275862068966
- Cluster 4, Avg Number of Crimes per Neighborhood: 52.375
- Cluster 5, Avg Number of Crimes per Neighborhood: 244.5

Clustering Based on Larceny Crimes (most common) and Location



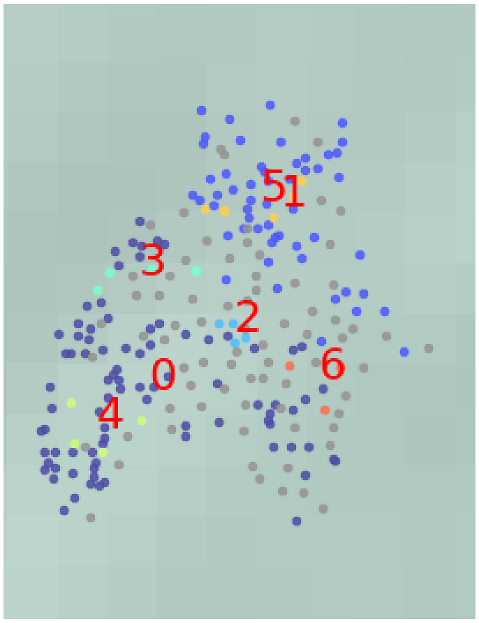
- Cluster 0, Avg Number of Larceny Crimes per Neighborhood: 40.83177570093458
- Cluster 1, Avg Number of Larceny Crimes per Neighborhood: 65.46428571428571
- Cluster 2, Avg Number of Larceny Crimes per Neighborhood: 34.833333333333336
- Cluster 3, Avg Number of Larceny Crimes per Neighborhood: 25.0
- Cluster 4, Avg Number of Larceny Crimes per Neighborhood: 185.5

Clustering Based on Violent Crimes (Aggravated Assault and Homicides) and Location



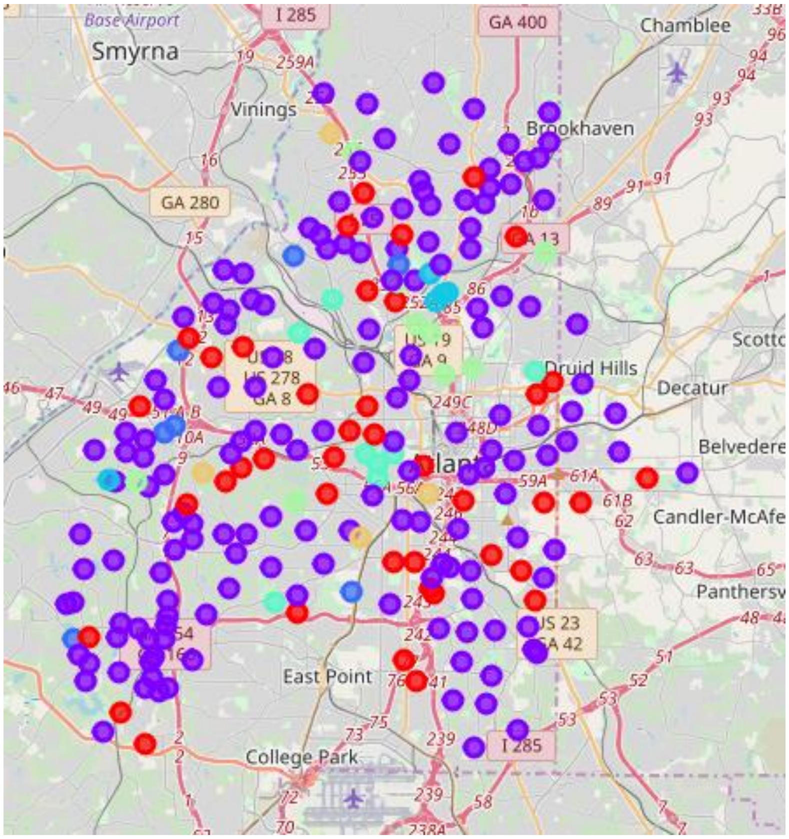
- Cluster 0, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 48.74
- Cluster 1, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 48.5
- Cluster 2, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 73.0
- Cluster 3, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 27.66
- Cluster 4, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 168.66
- Cluster 5, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 83.66
- Cluster 6, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 149.33
- Cluster 7, Avg Number of Homicide/Agg Assault Crimes per Neighborhood: 15.0

Clustering Based on Robbery Crimes and Location



- Cluster 0, Avg Number of Robbery Crimes per Neighborhood: 19.564705882352943
- Cluster 1, Avg Number of Robbery Crimes per Neighborhood: 59.98148148148148
- Cluster 2, Avg Number of Robbery Crimes per Neighborhood: 104.25
- Cluster 3, Avg Number of Robbery Crimes per Neighborhood: 37.0
- Cluster 4, Avg Number of Robbery Crimes per Neighborhood: 20.5
- Cluster 5, Avg Number of Robbery Crimes per Neighborhood: 42.0
- Cluster 6, Avg Number of Robbery Crimes per Neighborhood: 93.66666666666667

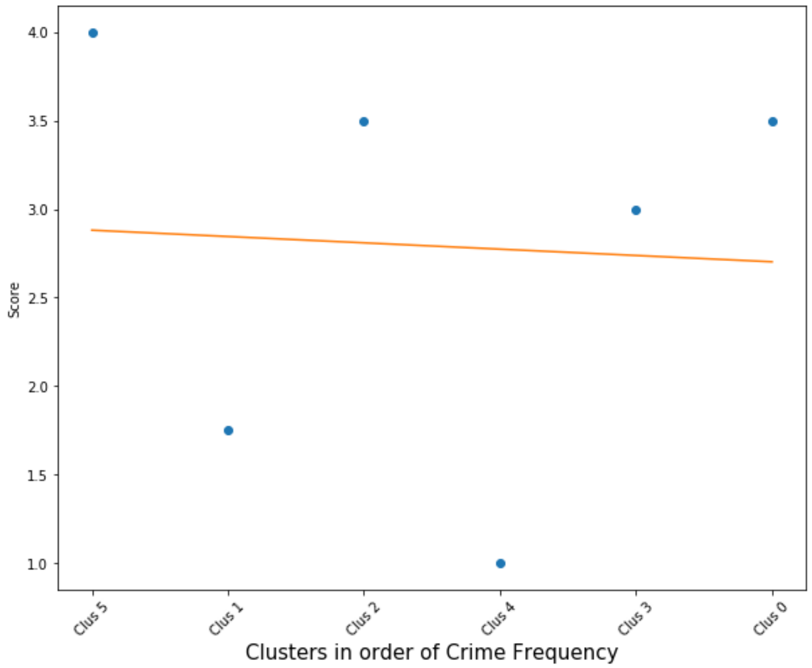
Experimental Clustering of Neighborhoods by Venues(did not use in results)



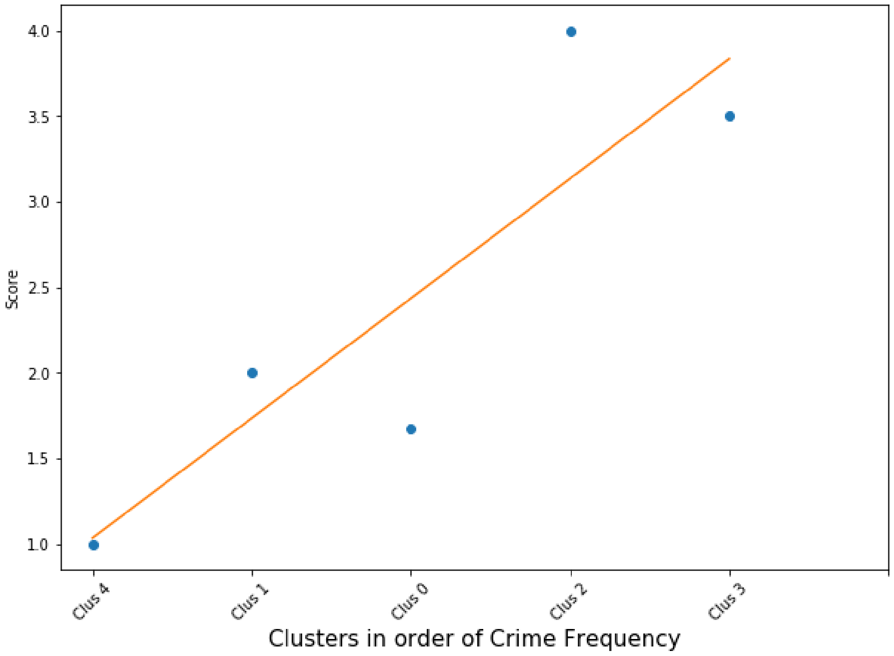
5. Conclusions

One of my goals was to search for correlations between venues traditionally associated with crime to see if it applied to Atlanta. For example, neighborhoods with more bars, liquor stores, etc would be expected to have higher crime rates by this logic. It is also important to remember that cluster sizes vary, so there may be variables that contribute to crime rates in specific areas that are unaccounted for in my results.

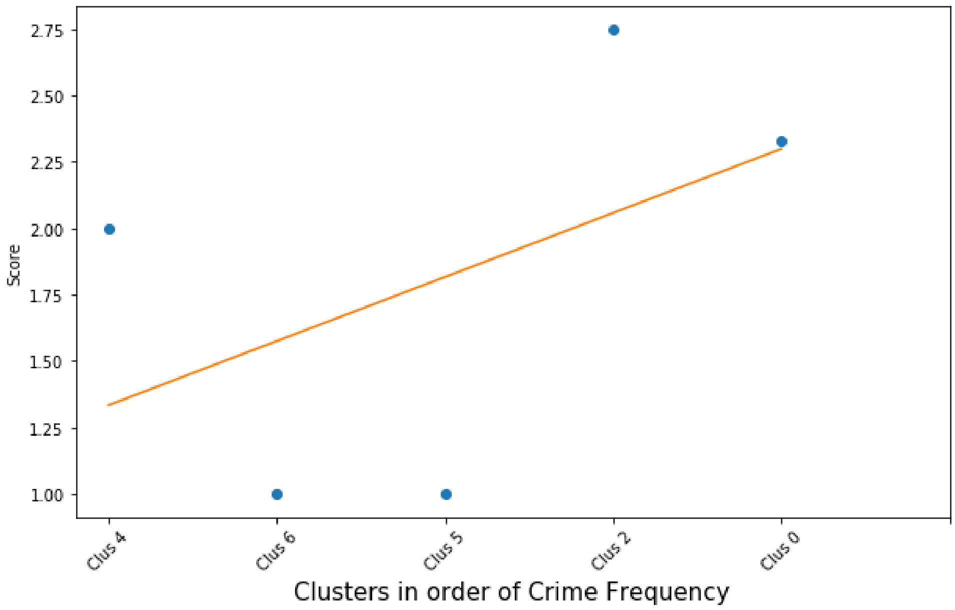
Using the crime rank and average scores given to each cluster I determined the correlations. For example, cluster with highest frequency of a particular venue received a 1, lowest a 6, and the applicable scores were averaged across. Only where a frequency greater than 0 occurred was taken into consideration. For example, a cluster with only 3 venues applicable would take the average rank only for 3 venues, not 5.



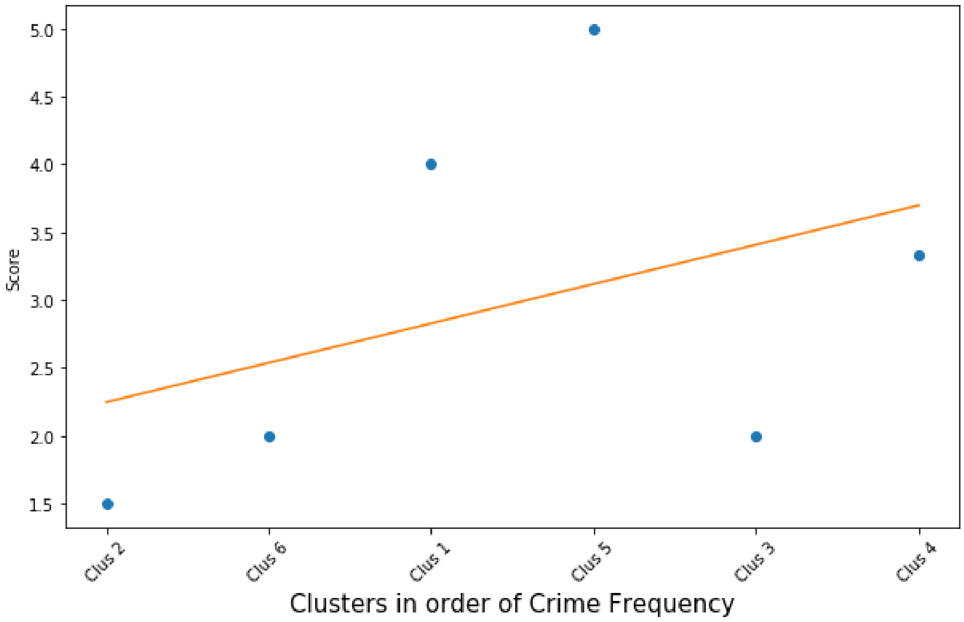
Overall Crime(Location) Appears to be a negative correlation surprisingly. We would expect the score average to increase(meaning lower frequency of the targeted venues) by my hypothesis.



Larceny Crimes The most common crime committed in Atlanta in 2018, there was a clear positive correlation between the frequency of the targeted venues and the average crime per neighborhood in the respective cluster.



Agg Assualt/Homicide Crimes Although to a lesser extent than larceny crimes, there is still a clear positive relationship between the frequency of the targeted venues and the average crime per neighborhood in the clusters.



Robbery Crimes Yet again, there is a positive correlation between the frequency of the targeted venues and the average number of robbery crimes per neighborhood committed in 2018.

6. Future Directions

Based on these findings, I can conclude that there appears to be a positive correlation between the frequency alcohol-selling venues and crime(gun shops and pawn shops were included in finding the results but their mean freqency was usually 0 among clusters). There may be other variables at play that also contribute to crime, so it is not definitive that reducing alcohol-selling venues would guarantee a decrease in crime. However, it is something to keep in mind for various groups of people such as police officers, venue owners, citizens, and more.

Future research could better implement the DBSCAN and KMEANS clustering methods in an attempt to have a better idea of crime clusters. Also, more variables or different groups of venues could be analyzed to search for any correlations between venue-types and crime rates.

Due to time constraints and my current knowledge with data science, not all methods were exhausted and more attention to detail could have been given, and better explanation could have been given in the report. For example, I could have included more variables associated with venues that serve/sell alcohol, among other things, but instead my results are derived from the description already given under the Results heading.