Final Report
Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 2)

Introduction/Business Problem-

Given the COVID-19 outbreak and the subsequent quarantines and shutdowns, individuals and businesses have begun to have to abide with social distancing and quarantining rules in order to flatten the curve and reduce the number of infections at any given time. In order to stay informed, individuals, businesses, or even the government may wish to use data science in order to determine the most densely populated zones in order to better allocate resources towards the area for COVID treatment or support for at risk individuals who would be more likely to contract COVID if they visited or lived in those zones.

Data science may be used to identify and visualize the major areas of a city, in this project I will be using New York, and determining the density of residential and commercial venues that have the highest probability of either inhabitants or visitors. Determining where these venues are most likely to be clumped together may give insight into which zones require more support or should be avoided by at risk individuals when businesses begin to re-open.

Most informed individuals or corporations would benefit from knowing which portions of the city are most densely populated, thus giving insight into which zones will require the most support and observation in order to minimize the spread and damage caused by the pandemic. At risk individuals may use this information to avoid densely populated zones, businesses may decide whether or not to reopen given the density of other venues in the area to protect their customers or employees, and the city or healthcare system may be able to observe and provide support proportional to the amount of risk that will occur with higher density locations reopening.

Data-

There are three sets of data that will be used to analyze the problem: the previously given New York City data that contains lists of the Boroughs and Neighborhoods with their respective latitudes and longitudes, the GeoSpace data to draw the borders on the maps from NYC OpenData, and the Foursquare venues to determine the density of venues in an area.

The Borough and Neighborhood data will come from a source we previously used in this course: https://cocl.us/new_york_dataset. It will consist of a list of Boroughs, Neighborhoods, and their coordinates in order to generate a list of zones.

The Foursquare data will come from accessing the Foursquare API to generate the number of venues in each zone.

The GeoSpace data will come from https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm and it will provide the geo spatial data to generate a map visualization of the data analysis.

Methodology-

First, I retrieved the data regarding the coordinates of the borough as well as which boroughs exist in New York City from the "https://cocl.us/new_york_dataset" dataset. Using this dataset the size of a borough is easily visualized. As shown in the generated graph, Queens has by far the largest number of neighborhoods, with Manhattan at the lowest.
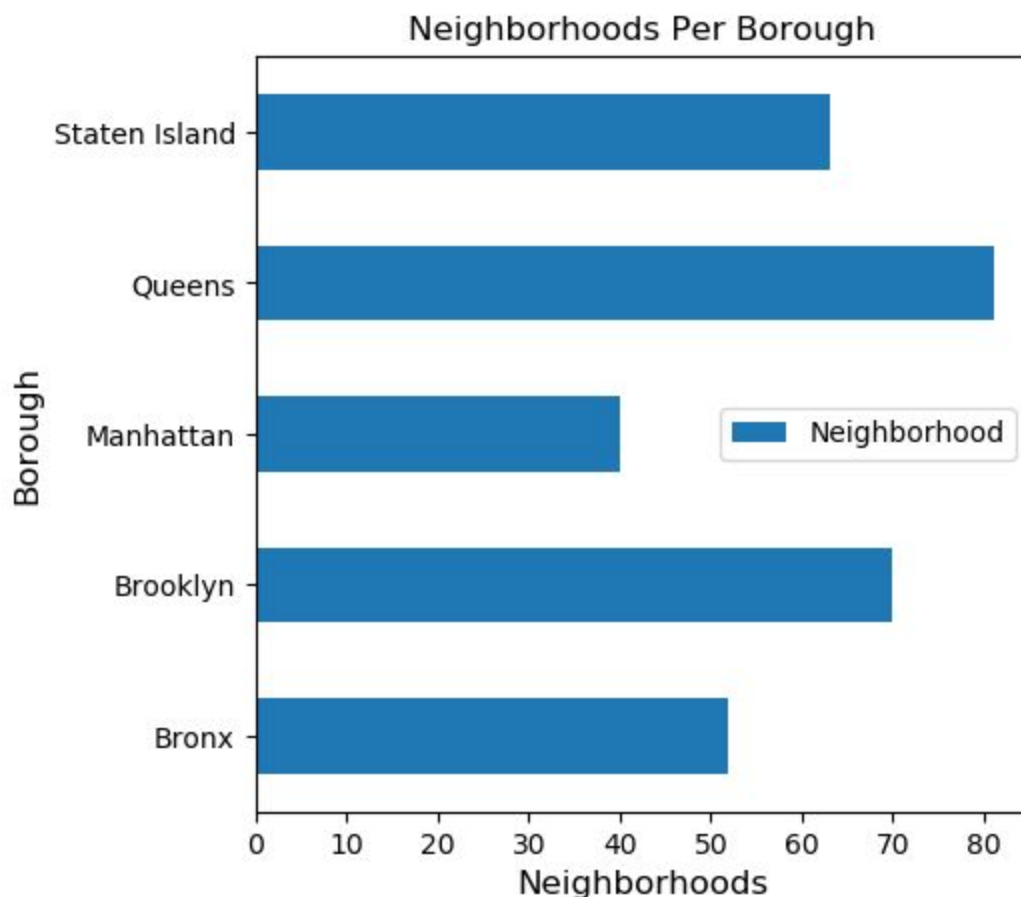
Secondly, I used the Foursquare API to pull every "food" section venue in each borough in order to represent which businesses will be the most at risk for both customers and employees due to the nature of the business itself as well as the high density and close quarters people find themselves in at these businesses, typically. As evidenced by the graph below, there are thousands of venues in total with most being situated in Manhattan and the least number in Staten Island.
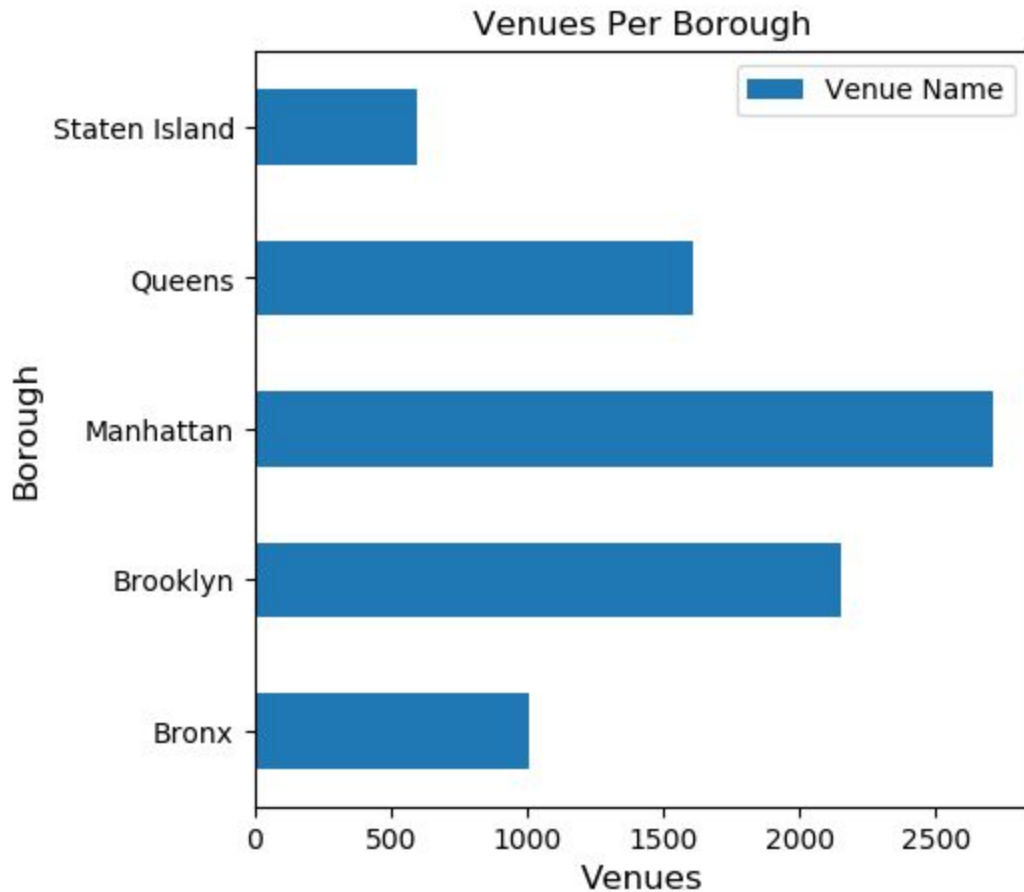
Finally, I used the geojson provided by the NYC OpenData project to draw a map of the boroughs and highlighted them according to the number of venues present in the borough, representing the density of close quarters venues in the area.
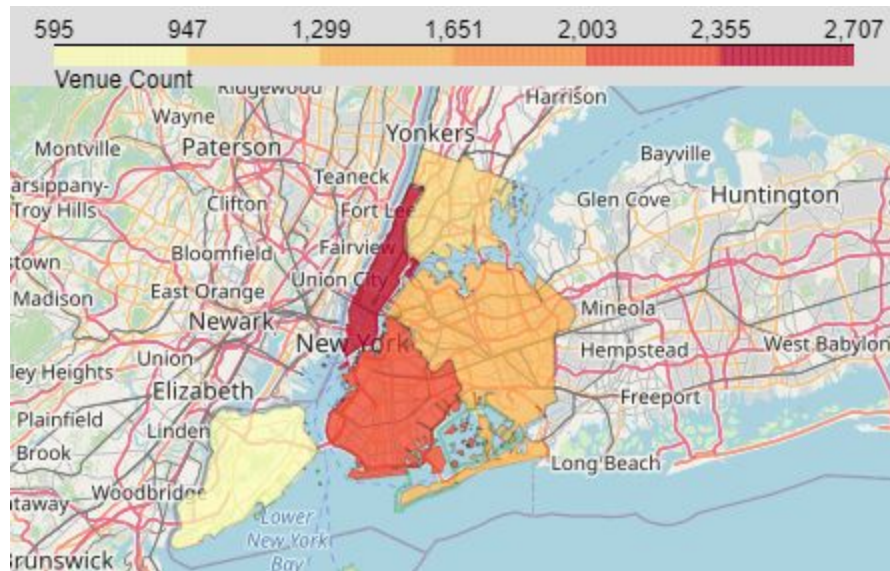
Results-

From the data it is easily observable that the probable reopening density of each borough varies greatly, and each business and individual must assess the risk of contact with the pandemic while keeping the population and venue density in mind.

For Queens, despite having the largest number of neighborhoods at roughly 80, the number of venues lays near the middle as it maintains only about 1,600 venues that were examined in this analysis. Its large size from the neighborhoods and middle of the pack number of venues puts it at a middling density for venues and potential risk.

On the other hand, Manhattan, with the lowest number of neighborhoods, maintains about 2,700 examined venues with a much lower number of neighborhoods, at roughly 40 neighborhoods. The low neighborhood count and high number of venues places Manhattan at the absolute top of the density and risk.

Staten Island sits at around 65 neighborhoods, the third largest of the boroughs, but has the lowest number of venues at just under 600 examined venues. The incredibly low venues alone suggest a low density, but coupled with it being on the larger side of neighborhoods suggests that the probably reopening density will be the lowest here than any other borough.

Brooklyn has the second highest count of examined venues, at just about 2,150, as well as having the second highest count of neighborhoods, at around 70. This puts it as a proportionally middling density as its large size is proportional to its number of venues.

The Bronx maintains the second lowest number of venues as well as second lowest number of neighborhoods, at roughly 1,000 and 50, respectively. This mimics the trend from Brooklyn which suggests that it will have a proportionally middling density as its size is proportional to its number of venues.

Discussion-

Per the analysis above it is my recommendation that any at risk individuals or businesses seeking to minimize exposure to or risk of infection during this pandemic avoid high density boroughs, such as Manhattan, and instead aim to only go out into the public if it is possible to maintain a safe distance from others in sparser density areas.

Conclusion-

In conclusion, if an at risk individual or business seeks to reduce risk and exposure, it is wisest to remain closed or indoors if you are in a high population density environment and it is possible to determine these population density environments by analyzing the number of venues that may return to activity as businesses reopen, causing a spike in population density in that and nearby locations.