Introduction/Business Problem

New York City is known for having a plethora of dining options all over the city. With a population of 8.4 million people and over 65 million visitors a year, the demand for restaurants will remain constant. While many people enjoy the diversity of cuisine and the opportunity to get “authentic” foreign dishes, there is still a demand for authentic American dishes: residents who did not grow up in New York occasionally yearn for comfort food from their childhood, and visitors from overseas want to try traditional American food. One of the quintessential type of foods in this categories is barbecue (BBQ). And not just any BBQ – real, authentic southern style pit BBQ.

Many BBQ restaurants use cooking techniques and ingredients that are focused on cost and volume, not taste. Because of that, the taste and quality is ok but not great. Being a native Southerner I know the difference quality ingredients and more importantly proper cooking techniques. This provides an opportunity to open several BBQ restaurants in NYC that will stand out from the competition. Rather than open a single restaurant, opening a restaurant in each of the 5 boroughs will provide multiple locations that are far enough apart to not overlap.

For this project, I will segment and cluster each of the 5 boroughs in NYC and identify the best neighborhood to open an authentic BBQ restaurant. To predict the ideal location, I’ll identify locations with other restaurants nearby, which will show the area already supports food services. I’ll also find areas without other BBQ joints nearby to minimize direct competition.

Data

Similar to some of the labs, Foursquare will be used to identify the various venues and locations in a given neighborhood. This data is ideal to sort by venue type and frequency of location/how many of a given type are within a given distance to a neighborhood’s center. To provide a comprehensive list of all the neighborhoods in NYC, I’ll use the free data available from NYU that was used in one of the skills labs. I did look for similar data in other cities (specifically my hometown of Atlanta) however that particular information is not available except for a fee. This data will provide the latitude & longitude for each neighborhood. That information will be cross referenced and linked with the Foursquare data.

Once the dataframes are built and segmented into the 5 boroughs, I’ll then extract the top venues in each neighborhood. I’ll then separate using one-hot encoding and group by neighborhood and frequency. After this, I’ll search the results and segment any neighborhoods that already have a “BBQ Joint” listed (that is the Foursquare category listing). If the frequency of BBQ joints is high for a given neighborhood, I’ll remove that neighborhood from the dataframe. Alternatively, if the number of neighborhoods with BBQ is relatively high, I’ll keep them in the list as to not skew any subsequent filtering or sorting however I will not select them as the ideal location should the analyses indicate they are.

After all the neighborhoods are analyzed, I’ll use k-means clustering to cluster them appropriately. After the clustering is complete, I’ll look at the most frequent venues in the cluster and rank them from best to worst to open a BBQ restaurant, specifically naming each neighborhood in a given cluster. This entire process will be done for all 5 boroughs.