Introduction/Business Problem

Suppose there is a contractor Jim who is planning on opening his own mom and pop restaurant in Brooklyn, NY. He is not sure where to open the restaurant, but he knows he wants it to be in a place where there is not that many, so the competition is weaker. It is your job as a data scientist to find the best neighborhood in Brooklyn where Jim should open his restaurant. To do this, you will use Foursquare location data to organize all the neighborhoods of Brooklyn into five clusters based on what venues are the most common. For example, cluster 0 consists of two neighborhoods in Brooklyn that have spas as their most common venue.

Data

Before we use Foursquare, we must gather all the neighborhoods in Brooklyn with their respective latitude and longitude. We can do this by downloading the json file on the NYU website introduced in the lab from Week 3 in the capstone course. This file contains all the boroughs of New York City with their neighborhoods, latitudes, and longitudes. Now that we have the location data of New York City, we can use Foursquare to solve this problem. For this specific problem we will only focus on the borough Brooklyn. Foursquare API will allow us to find all the venues that appear in each neighborhood of Brooklyn by using each neighborhood's latitude and longitude. The data we are using includes:

- Borough (Brooklyn)
- Neighborhood
- Latitude
- Longitude
- Venue

Methodology

After splitting up Brooklyn into its 69 neighborhoods, we run Foursquare on the latitude and longitude of each neighborhood to return 100 venues that appear in each neighborhood. We then organize these neighborhoods into 5 clusters based on which venues are the most common or appear the most. Since Jim is opening a restaurant, we want to choose a cluster of neighborhoods that does not have many if any restaurants. This will ensure that Jim will not have too much competition when he opens his restaurant.

Results

Now that we have split up the 69 neighborhoods of Brooklyn into 5 clusters, we will analyze each cluster to see which one has the least number of restaurants. Cluster 0 contains neighborhoods Sea Gate and Madison. They seem to be clustered together based on their common frequency of spas. Cluster 1 contains lone neighborhood Paerdegat Basin which has Asian Restaurant and Child Care Service as its top 2 most common venues. Cluster 2 contains 56 neighborhoods which all have a pizza place, coffee shop, or type of restaurant in their top 2 most common venues. Cluster 3 contains lone neighborhood Mill Island which has Pool and Food as its most common venues. Cluster 4 contains 10 neighborhoods which either have a type of restaurant or pharmacy as one of their most common venues.

Discussion

We can see that Cluster 0 and Cluster 3 contain the least number of restaurants. However, Cluster 3's only neighborhood Mill Island contains Falafel and Fast Food Restaurants as its 6th and 9th most common venue. Whereas, Cluster 0 has a restaurant appearing once in its top 10 most common venues with neighborhood Madison having an Italian Restaurant as its 8th most common venue. Therefore, I will pick Cluster 0 as our cluster of choice for Jim to open his restaurant. Furthermore, **I will advise Jim opens a restaurant in the Sea Gate neighborhood**, from Cluster 0, since it has no restaurants in its list of top 10 most common venues.

Conclusion

We have successfully used Foursquare API and Python to suggest which neighborhood in Brooklyn Jim should open his restaurant. The Sea Gate neighborhood in Brooklyn is the best place for Jim to open his restaurant since it has very few restaurants. This will allow Jim to operate his restaurant with a lower level of competition than what he would face in other places in Brooklyn.