Introduction/Business Problem

Suppose there is a contractor Jim who is planning on opening his own pizza place in Brooklyn, NY. He is not sure where to open the pizza place, 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 pizza place. 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 1 consists of two neighborhoods in Brooklyn that have Asian and Chinese restaurants as their most common venues.

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 pizza place, we want to choose a cluster of neighborhoods that does not have many if any pizza places. This will ensure that Jim will not have too much competition when he opens his pizza place.

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 pizza places. Cluster 0 contains 40 neighborhoods which all seem to have a restaurant, food place, or coffee shop in their top 2 most common venue. Cluster 1

contains neighborhoods Canarsie and Paerdegat Basin. They seem to be clustered together based on their common frequency of restaurants. Cluster 2 contains 10 neighborhoods which all have a pizza place or food place 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 17 neighborhoods which either have a type of restaurant or store as one of their most common venues.

Discussion

We can see that Cluster 1 and Cluster 3 don't have a pizza place in their top 10 list of most common venues. However, Cluster 1's two neighborhoods Canarsie and Paerdegat Basin each contain a total of 5 restaurants and food places in their top 10 most common venues. Whereas, Cluster 3 only has two restaurants appearing in its top 10 most common venues. Therefore, Cluster 3 is the cluster of choice for Jim to open his pizza place since it has the least amount of food places and restaurants out of the clusters that don't have many if any pizza places (Cluster 1 and 3). Thus, **I will advise Jim to open a pizza place in the Mill Island neighborhood** from Cluster 3. This neighborhood will have the least amount of competition for pizza places and thus will allow Jim to sell his pizza to a fresh audience.

Conclusion

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