**Analyzing Neighborhoods of Scarborough and Brooklyn**

Volkan Arıkan

April 22, 2020

**1. INTRODUCTION/BUSINESS PROBLEM**

Two of most famous cities in the world are Toronto and New York. Both cities are very diverse, multicultural and are the financial capitals of their respective countries. I want to explore the similarity and dissimilarity of these cities about their touristic places.

Tourism is an important activity for economy of countries and people usually like to visit developed countries. It is easy to explore and analyze cities with location data. Tourists like to visit different places and analysis of two cities will help them to make their choices easier and more efficient.

**2. DATA DESCRIPTION**

I will use Foursquare API and get location data to explore the neighborhoods of two cities. This data gives information about different places such as coffee shops, restaurants, hotels, museums and many more around each neighborhood. I selected Scarborough from Toronto and Brooklyn from New York to analyze their neighborhoods. I will use k-means clustering technique to segment the neighborhoods regarding their places. This will help to understand their similarity and dissimilarity.

I found data of Toronto neighborhoods from Wikipedia. The table has all boroughs and neighborhoods with their postal codes. I cleaned the data, eliminated “Not assigned” values, combined neighborhoods which have same geographical coordinates at each borough and sorted against the concerned borough. Finally I reduced it to Scarborough.

For New York neighborhoods, i used a saved data file. The .json file has all boroughs and neighborhoods with their location data. I cleaned the data, eliminated “Not assigned” values, combined neighborhoods which have same geographical coordinates at each borough and sorted against the concerned borough. Finally I reduced it to Brooklyn.

I used Foursquare API to get the coordinates of Scarborough and Brooklyn and explore their neighborhoods.

**3. METHODOLOGY**

I selected one borough from each city to explore their neighborhoods. The data exploration, analysis and visualization for both boroughs are done in the same way but separately. As a database, I used GitHub repository in my study. My master data has the main components Borough, Neighborhood, Latitudeand Longitude informations of each borough. I cleaned and reduced them to Scarborough and Brooklyn.

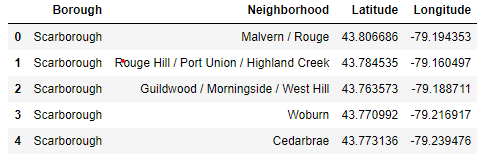


Figure 1. Table of Scarborough Neighborhoods

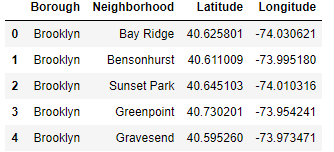


Figure 2. Table of Brooklyn Neighborhoods

I used python folium library to visualize geographic details of Scarborough, Brooklyn and their neighborhoods. I created a map of them with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:

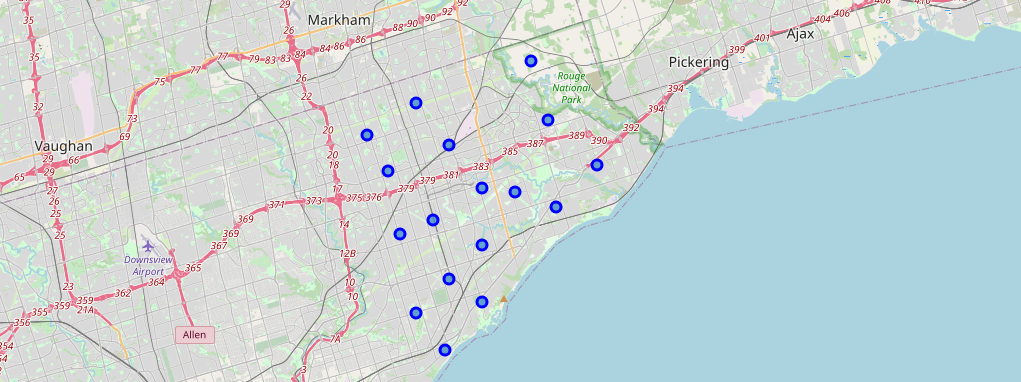


Figure 3. Map of Scarborough

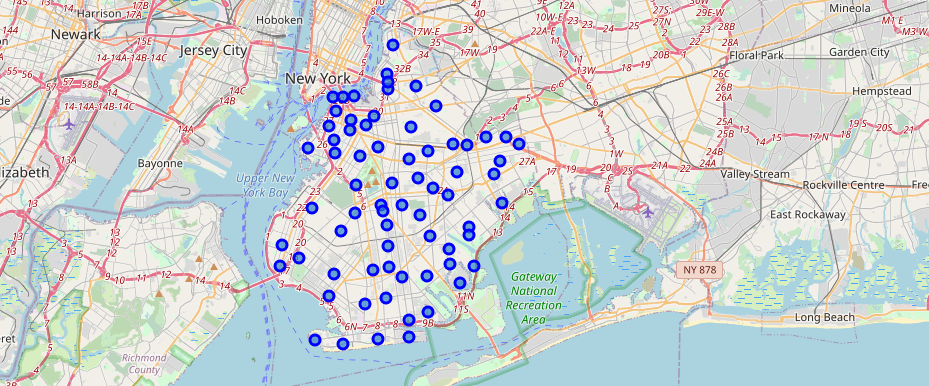


Figure 4. Map of Brooklyn

I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 20 venues and the radius 500 meter for each borough from their given latitude and longitude informations. Here are merged tables of neighborhoods and venues for both boroughs.

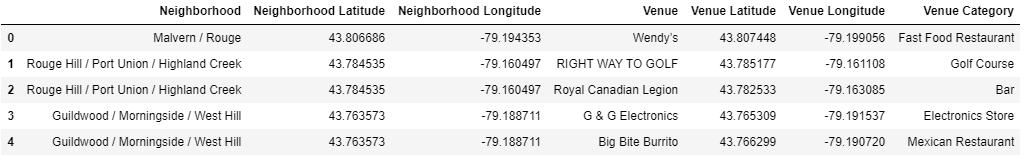


Figure 5. Neighborhoods and Venues of Scarborough



Figure 6. Neighborhoods and Venues of Brooklyn

In summary, 57 unique categories for Scarborough and 234 unique categories for Brooklyn were returned by Foursquare. Then i applied one hot encoding, measured the frequency of each venue and created a table which shows list of top 10 venue category for each neighborhood.



Figure 7. Table of Top 10 Venues of Scarborough Neighborhoods

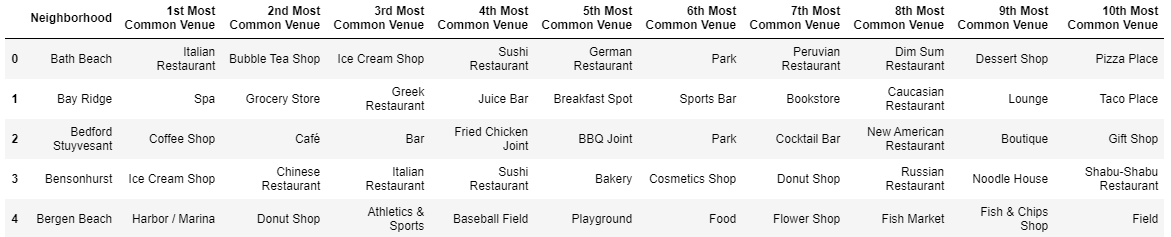


Figure 8. Table of Top 10 Venues of Brooklyn Neighborhoods

I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning. I run K-Means to cluster the neighborhoods into 3 clusters.

Here are my merged tables with cluster labels for each neighborhood of both boroughs.



Figure 9. Scarborough Merged Table with Labels



Figure 10. Brooklyn Merged Table with Labels

When we examine clusters, we can label each cluster as follows:

Scarborough:

* Cluster 1: Outdoor Activities
* Cluster 2: Commercial Places
* Cluster 3: Food Places
* Cluster 4: Residentials
* Cluster 5: Tourist Areas and Hubs

Brooklyn:

* Cluster 1: Multiple Social Venues
* Cluster 2: Commercial Places
* Cluster 3: Cultural and Going Out Places
* Cluster 4: Event Places
* Cluster 5: Marina and Seafood Markets

**RESULTS**

After clustering the data of the respective neighborhoods, it's clear that Brooklyn has more venues which can be explored and attract the tourists. The neighborhoods are much similar in features like clubs, coffee shops etc. As far as concern to dissimilarity, it differs in terms of some unique places like restaurants, sport places, historical places.

**DISCUSSION**

When we compare the tourist places, we observe that Brooklyn has more different venues. Different kind of stores, stadiums, coffee shops are available in Scarborough while venues like nightlife, restaurants, museums and theaters are present in Brooklyn.

As far as concern to recommendations, we recommend Brooklyn neighborhoods will be considered first to visit. The tourists will have the opportunity to explore different kind of venues. We recommend you visit Scarborough just in case of you explored Brooklyn completely and still have time to see new places.

**CONCLUSION**

Brooklyn is a more diverse borough in terms of venues compared to Scarborough. It’s obvious that Scarborough has way less venues and they are not so different in comparison with the venues of Brooklyn. But as we know that every place is unique in its own way, so that’s argument is present in both neighborhoods.