Battle of Neighborhood

A Coursera Capstone Project

By: *Daljeet Singh Gora*

1. Introduction

Background

Every individual in today's world wants to have best living standards and quality of life. We all are so busy now-a-days that we do have enough time to travel from one part of the city to another for our daily needs. Facilities like gymnasiums, nearby schools and colleges, grocery stores, coffee shops, etc. all these adds to our daily necessities. Everyday busy schedules and heavy traffics consumes so much of time that we struggle to achieve our daily needs.

One of the options is to select and reside in a neighborhood which have all or most of the facilities nearby.

One of my friend is looking to settle in a city which provides him better living standards and quality of life. He did his initial analysis and came up with 2 options, Toronto, Canada and New York City, USA. He has asked me for the help to decide one of them. I am doing this project which will analyze neighborhoods between Toronto and New York. It will provide detailed analysis and insight into the neighborhoods and local businesses so that it helps my friend make a decision to understand which city has better facilities and quality of life. This project will also explore the similarities and dissimilarities between neighborhoods in the two cities and help determine which neighborhoods is the best place to settle.

Target Audience

Although this project is to help my friend make a decision but similar analysis can also be done for any other city of the world to understand the neighborhoods. This can be helpful for any individual just like my friend, any company who is in search of a best location to set up an office so that all the employees working can be benefitted or anyone looking to setup a business related to daily necessity like gymnasium, coffee shop, grocery store, etc.

2. Data

City Data:

The data used for this project was acquired from the respective cities Wikipedia website pages. The datasets consist of the postal codes, neighborhood names, latitude, and longitude information for each neighborhood. The data from Wikipedia pages was scraped, wrangled, cleaned and then read into a pandas dataframe so that it is in a structured format.

• Toronto Neighborhoods - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

• New York City Neighborhoods - https://geo.nyu.edu/catalog/nyu_2451_34572

Neighborhood Venue Data:

Foursquare lets users search most common venues like restaurants, nightlife spots, shops and other places in a location. It provides personalized recommendations based on factors that include user experience and their venue ratings.

Foursquare API search feature was used to collect neighborhood venue information. Details about local venues and locality provides insight into the qualities of a neighborhood.

• Foursquare - https://foursquare.com/

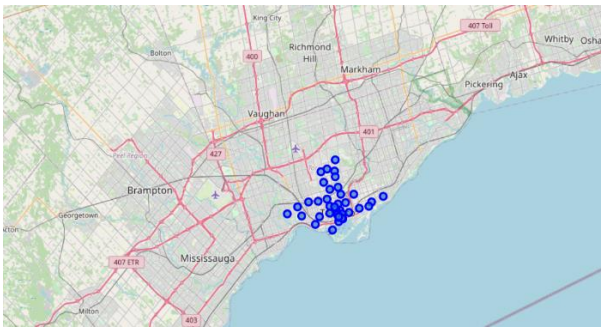
In addition to Foursquare, various python packages were used to create maps and machine learning models to further provide insights into the neighborhood battle project.

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

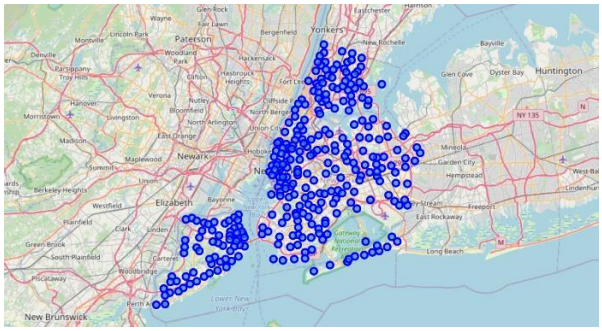
3. Methodology

Toronto and New York City data was acquired from the respective cities Wikipedia website pages to segment, cluster and explore neighborhoods. The dataset is a list of postal codes including boroughs and neighborhood names. As a part of data cleaning any incomplete data was excluded. Next step was to download longitude and latitude data for each neighborhood and combine with the city dataset. This became my master dataset.

Toronto Map



New York Map



Since both the cities are widely spread in terms of area, one of the most famous neighborhoods from each city was selected for exploration. Scarborough from Toronto and Queens from New York City.

Next step was to find out venue details, Foursquare API was used to fetch the venue information. Each neighborhood was further analyzed to find top 10 venues. Below is the sample data of the top venues of Scarborough and Queens.

Top Venues for Scarborough

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Co
0	Agincourt	Skating Rink	Breakfast Spot	Latin American Restaurant	Lounge	Clothing Store	Vietnamese Restaurant	Coffee Shop	Grocery Store	General Entertainment	Gas S
1	Birch Cliff, Cliffside West	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gym	Grocery Store	Gas Station	Fried Chicke Joint
2	Cedarbrae	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	Gas Station	Fried Chicken Joint	Caribbean Restaurant	College Stadium	Gym
3	Clarks Corners, Tam O'Shanter, Sullivan	Pizza Place	Chinese Restaurant	Noodle House	Thai Restaurant	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Intersection	Bank	Italian Resta
4	Cliffside, Cliffcrest, Scarborough Village West	Motel	American Restaurant	Vietnamese Restaurant	Gym	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electr Store

Top Venues for Queens

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Mos Commo Venu
0	Arverne	Surf Spot	Metro Station	Sandwich Place	Playground	Restaurant	Bus Stop	Café	Board Shop	Beach	Thai Restaurant
1	Astoria	Middle Eastern Restaurant	Bar	Seafood Restaurant	Indian Restaurant	Greek Restaurant	Hookah Bar	Bakery	Mediterranean Restaurant	Food Truck	Café
2	Astoria Heights	Plaza	Bakery	Burger Joint	Food	Supermarket	Bowling Alley	Chinese Restaurant	Laundromat	Bus Station	Playgrounc
3	Auburndale	Italian Restaurant	Comic Shop	Train	Supermarket	Furniture / Home Store	Mattress Store	Korean Restaurant	Bar	Fast Food Restaurant	Gymnastic Gym
4	Bay Terrace	Clothing Store	Shoe Store	Cosmetics Shop	American Restaurant	Donut Shop	Kids Store	Women's Store	Mobile Phone Shop	Furniture / Home Store	Gift Shop

It was observed that there are many common venue categories. Final step was to cluster the neighborhoods using K-means algorithm. It is one of the most common cluster methods of unsupervised learning. The objective of K-means is to group similar data points together and discover underlying patterns.

Cluster details are provided under Results section.

4. Results

Below are the results/observations from neighborhood clustering.

Scarborough, Toronto:

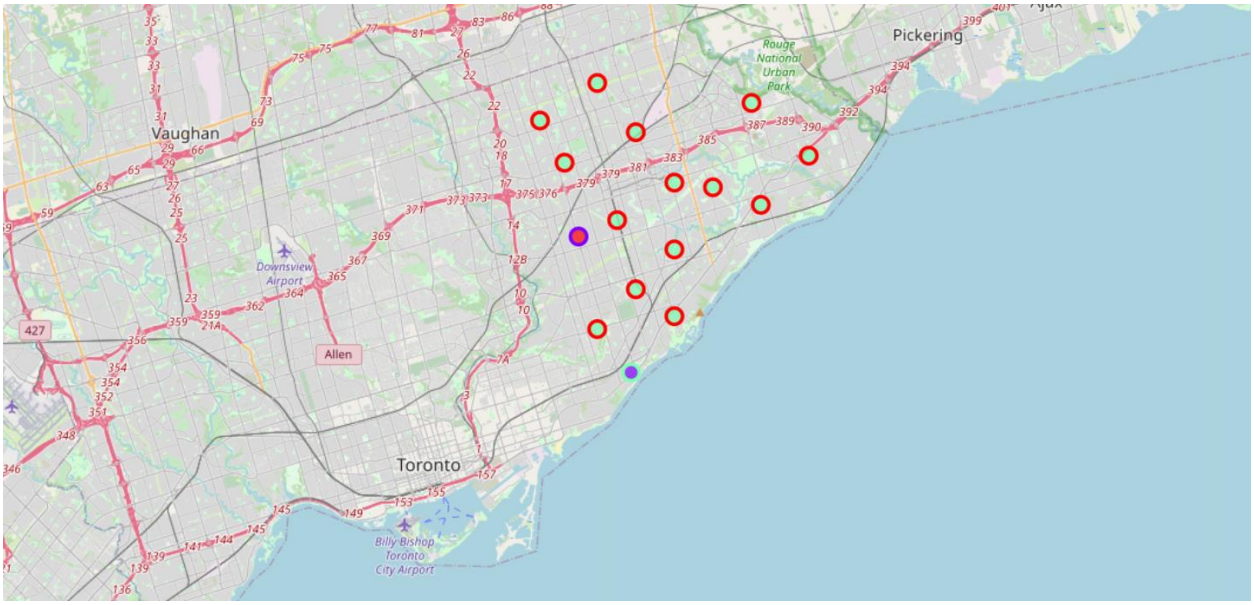
Scarborough neighborhood has:

- 89 venues in 17 neighborhoods
- 80 distinct venues in 55 categories

Scarborough neighborhoods were divided into 3 clusters.

- Cluster 1 has 1 neighborhood with most common venue as Accessories Store
- Cluster 2 has 1 neighborhood with most common venue as General Entertainment
- Cluster 3 has 15 neighborhoods with most common venue as Restaurants and Bakery.

Clusters of Scarborough



Queens, New York:

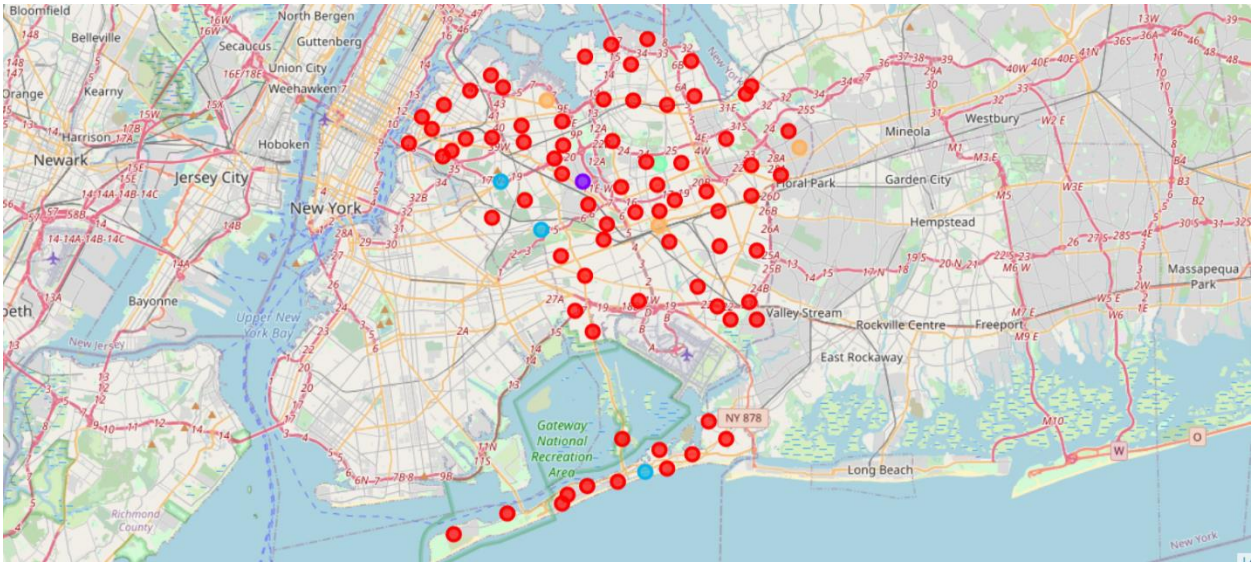
Queens neighborhood has:

- 2113 venues in 81 neighborhoods
- 1742 distinct venues in 278 categories

Neighborhoods were divided into 5 clusters.

- Cluster 1 has 1 neighborhood with most common venue as Gym/Fitness Center
- Cluster 2 has 3 neighborhoods with most common venue as Pizza Place, Deli/Bodega and Beach
- Cluster 3 has 1 neighborhood with most common venue as Deli/Bodega
- Cluster 4 has 3 neighborhoods with most common venue as Donut Shop, Pizza Place and Grocery Store.
- Cluster 5 has 73 neighborhoods with most common venue as Restaurants Deli/Bodega, Pizza Place and Bakery.

Clusters of Queens



5. Discussion

Queens borough of New York City is the largest borough geographically and is the fourth-most-densely populated with over 2.2 million population among New York City's boroughs.

Scarborough borough of Toronto has grown from a collection of small rural villages and developed as a suburb of Toronto. In 1998, Scarborough and the rest of Metropolitan Toronto were amalgamated into the present city of Toronto. With population of over 0.6 million, Scarborough is one of the most popular destination.

Project analysis shows Queens borough is having much more to offer in terms of venues and facilities as compared to Scarborough.

It is observed that many of the neighborhoods are homogenous and are very similar to each other. Both Scarborough and Queens borough consist of neighborhood cluster that contain majority of the neighborhoods and the remaining clusters have very less 1-3 neighborhoods.

Comparing between Queens and Scarborough, Queens borough has significant greater number of neighborhoods and venues than Scarborough. Queens has 2113 venues in 81 neighborhoods whereas Scarborough has 89 venues in 17 neighborhoods.

6. Conclusion

Based on the number and variety of venues, I would recommend Queens over Scarborough as a choice to relocate to my friend to enjoy the best living standards and quality of life. Queens offer way more in choices for restaurants, gymnasiums, grocery stores, and extracurricular activities for individuals and families.