The Battle of Neighbourhoods

Haripriya Subramanian

21 November 2019

1. Introduction

1.1 Background

Toronto is one of the most populous cities and the largest city located in the Ontario province which is the capital of Canada. This place seems to be the favourite location for tourists visiting Canada and for the foreigners who wish to settle in Canada. As there are many immigrants like Employees, Students to this location, it is very important for the people to get to know about the neighbourhood in this city. For the people who wish to start their new business, it is necessary in which neighbour hood is the best for that. This project is based on the analysis of data obtained using foursquare API which provides the answers to the all these questions.

1.2 Business problem

According to the data from this website: https://en.wikipedia.org/wiki/Demographics_of_Toronto , it is evident that there is a gradual increase in the number of immigrants to Toronto every year. If a new immigrant wishes to start a new business in Toronto which is the best neighbourhood for this? For the tourist in who come to visit Toronto how could they find which is the best venue in the neighbourhoods. They might also be interested in a particular type of restaurant and how could they know easily in which neighbour is that restaurant available?

2. Data

The initial data is obtained by web scraping the data from the Wikipedia website https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The data is prepared by removing Borough which is not assigned. The Neighbourhoods are grouped according to the postal codes. The data set obtained is as follows:

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Highland Creek, Port Union, Rouge Hill
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...
11	M1R	Scarborough	Maryvale, Wexford
12	M1S	Scarborough	Agincourt
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...
15	M1W	Scarborough	L'Amoreaux West
16	M1X	Scarborough	Upper Rouge
17	M2H	North York	Hillcrest Village
18	M2J	North York	Fairview, Henry Farm, Oriole

Then the longitudes and latitudes for the postalcodea are obtained from 'https://cocl.us/Geospatial_data'.

The top 5 rows for the data are:

Postal Code	Latitude	Longitude
M1B	43.806686	-79.194353
M1C	43.784535	-79.160497
M1E	43.763573	-79.188711
M1G	43.770992	-79.216917
M1H	43.773136	-79.239476

The Geospatial data is then merged with the prepared Wikipedia data. This dataset is as follows:
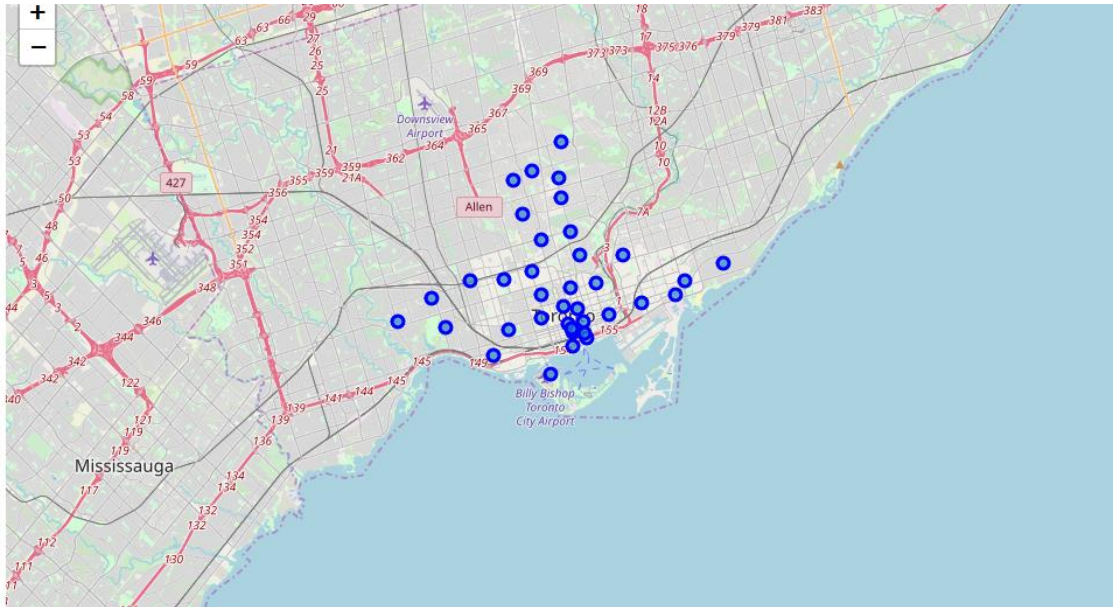
Postal Code	Borough	Neighbourhood	Latitude	Longitude
M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
M1C	Scarborough	Highland Creek, Port Union, Rouge Hill	43.784535	-79.160497
M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
M1G	Scarborough	Woburn	43.770992	-79.216917
M1H	Scarborough	Cedarbrae	43.773136	-79.239476

From the above dataset the Borough which contain the word Toronto is filtered out:

The glimpse of the dataset is shown below:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
45	M4P	Central Toronto	Davisville North	43.712751	-79.390197
46	M4R	Central Toronto	North Toronto West	43.715383	-79.405678
47	M4S	Central Toronto	Davisville	43.704324	-79.388790
48	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160
49	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049

From the geocoder API the coordinates(latitude and longitude) of Toronto is obtained. From the data, a folium map is created.

Using the foursquare API data, we obtain the venues for the neighbourhoods: The following dataset set shows the venues along with their latitude, longitude and the venue type.

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	Riverdale, The Danforth West	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

The top five venues for the neighbour hoods are found out:

```
----Adelaide, King, Richmond----
               venue  freq
0          Steakhouse  0.10
1     Sushi Restaurant  0.07
2               Hotel  0.07
3         Pizza Place  0.07
4     Asian Restaurant  0.07


----Bathurst Quay, CN Tower, Harbourfront West, Island airport, King and Spadina, Railway Lands, South Niagara----
               venue  freq
0      Airport Service  0.19
1      Airport Lounge  0.12
2             Airport  0.06
3  Airport Food Court  0.06
4                 Bar  0.06


----Berczy Park----
               venue  freq
0         Coffee Shop  0.07
1       Farmers Market  0.07
2   Seafood Restaurant  0.07
3        Cocktail Bar  0.07
4            Beer Bar  0.07


----Brockton, Exhibition Place, Parkdale Village----
               venue  freq
0      Breakfast Spot  0.10
1                Café  0.10
2          Coffee Shop  0.10
```

Then by using k-means clustering 10 clusters are created based on the most common venues from each neighbour hood. The following table shows in which cluster the neighbourhood belongs to and the most common venues in the neighbourhood:

	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	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 Ven
0	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	Pub	Neighborhood	Yoga Studio	Cuban Restaurant	Ethiopian Restaurant	Easte Europe Restaura
1	East Toronto	Riverdale, The Danforth West	43.679557	-79.352188	8	Greek Restaurant	Ice Cream Shop	Italian Restaurant	Yoga Studio	Pizza Place	Cosmetics Shop	Pub	Desse Sh
2	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	0	Park	Sandwich Place	Gym	Sushi Restaurant	Pet Store	Pizza Place	Movie Theater	P
3	East Toronto	Studio District	43.659526	-79.340923	8	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Park	Bar	Booksto
4	Central Toronto	Lawrence Park	43.728020	-79.388790	0	Bus Line	Park	Swim School	Yoga Studio	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant	Dog R

Then a folium map is created with the clusters: