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 consist of the following columns:

Postal Code, Borough, Neighbourhood.

Then the longitudes and latitudes for the postalcodea are obtained from 'https://cocl.us/Geospatial_data'. The data frame obtained by reading the csv file in the url consists of the Postal code, longitude and latitude.

The Geospatial data is then merged with the prepared Wikipedia data on the postal code as the key. The merged data set consists of postal Code , Borough, Neighbour, Longitude and Latitude.

From the merged dataset the Borough with the words 'Toronto' is filtered.

The four square uses these latitudes and longitudes or the neighbours to provide information about the venues of the neighbours, top venues of the neighbours based on the frequency of visits to this venue etc.

By using the four square API, the venues for the neighbour are found. The information is printed in a data set consists of the venue list for each of the neighbours along with their latitude and longitude.

Based on the latitude and longitude location given the foursquare API generates the venues present for a particular neighbourhood.

A one hot encoding of the data is done for each of the venues that belong to the neighbour hoods. The venues are then grouped by their mean.

The top 5 venues are found for each of the neighbours. This would help to new tourists to find out the venues for a particular neighbours easily.

Based on the top 10 venues, 5 different clusters of the neighbours are created using the k means clustering algorithm.

3. Exploratory Data Analysis:

The data set from the Wikipedia and after performing the steps of data preparation is obtained is as follows

	Postal Code	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
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

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

The top 5 rows for the data are:

The Geospatial data is then merged with the prepared Wikipedia data. This dataset is as follows:

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

After merging the data from the 2 data frames, the resultant data set is:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-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 category.

	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 count of the number of venues is found for each neighbour and a data frame is created out of that:

	Venue Category
Neighbourhood	
Adelaide, King, Richmond	100
Berczy Park	57
Brockton, Exhibition Place, Parkdale Village	21
Business Reply Mail Processing Centre 969 Eastern	18
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	16
Cabbagetown, St. James Town	43
Central Bay Street	82
Chinatown, Grange Park, Kensington Market	96
Christie	17
Church and Wellesley	89
Commerce Court, Victoria Hotel	100
Davisville	38
Davisville North	7
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	16
Design Exchange, Toronto Dominion Centre	100
Dovercourt Village, Dufferin	18
First Canadian Place, Underground city	100
Forest Hill North, Forest Hill West	4

The top five venues for the neighbour hoods based on the frequency of visits 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
```

One hot encoding created based on the category type of the venues for each of the neighbour hoods:

	Neighbourhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant
0	The Beaches	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0
1	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
4	The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

By using k-means clustering 5 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 10 most common venues in the neighbourhood:

	Postal Code	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 M Com Ve
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Neighborhood	Health Food Store	Pub	Trail	Event Space	Ethiopian Restaurant	Electronics Store	Eas Europ Restau
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Furniture / Home Store	Cosmetics Shop	Brewery	Bu Tea S
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572	0	Park	Sandwich Place	Brewery	Steakhouse	Sushi Restaurant	Food & Drink Shop	Fish & Chips Shop	Ice Cr S
43	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Yoga Studio	Fish Market	Cowor S
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	4	Park	Swim School	Bus Line	Yoga Studio	Diner	Falafel Restaurant	Event Space	Ethio Restau

Then a folium map is created to visualize the clusters.
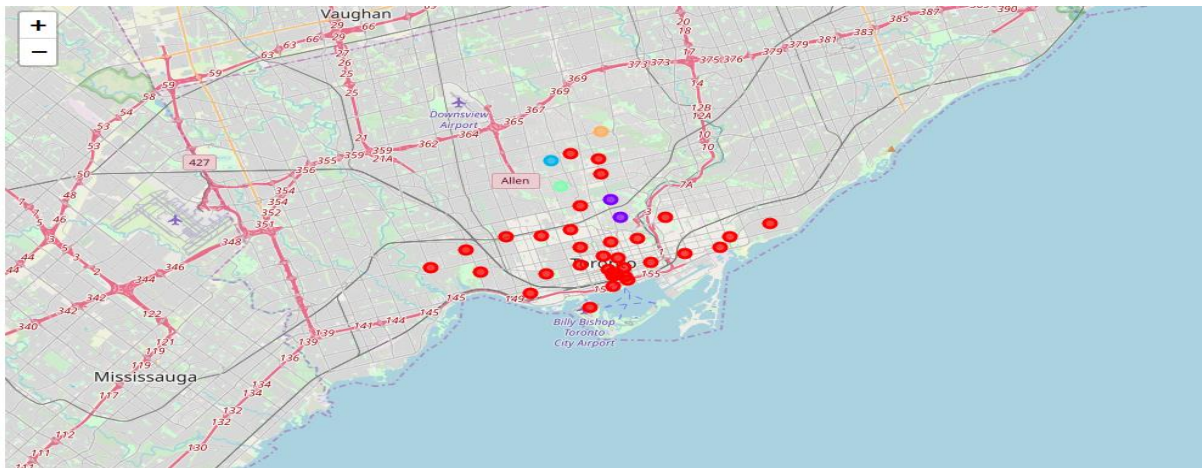
 Cluster 0 data:

This cluster mainly consists of the most common venues like café and restaurants as the first or second most common venues.

	Borough	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 Venue	9th Most Common Venue	10th Most Common Venue
37	East Toronto	0	Neighborhood	Health Food Store	Pub	Trail	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dessert Shop	Falafel Restaurant
41	East Toronto	0	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Furniture / Home Store	Cosmetics Shop	Brewery	Bubble Tea Shop	Restaurant	Caribbean Restaurant
42	East Toronto	0	Park	Sandwich Place	Brewery	Steakhouse	Sushi Restaurant	Food & Drink Shop	Fish & Chips Shop	Ice Cream Shop	Pub	Fast Food Restaurant
43	East Toronto	0	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Yoga Studio	Fish Market	Coworking Space	Latin American Restaurant	Brewery
45	Central Toronto	0	Gym	Breakfast Spot	Food & Drink Shop	Hotel	Clothing Store	Sandwich Place	Park	Electronics Store	Ethiopian Restaurant	Diner
46	Central Toronto	0	Clothing Store	Coffee Shop	Sporting Goods Shop	Gym / Fitness Center	Mexican Restaurant	Diner	Dessert Shop	Park	Chinese Restaurant	Rental Car Location
47	Central Toronto	0	Sandwich Place	Gym	Dessert Shop	Pizza Place	Café	Italian Restaurant	Coffee Shop	Sushi Restaurant	Farmers Market	Fried Chicken Joint
49	Central Toronto	0	Pub	Coffee Shop	Pizza Place	Liquor Store	Sports Bar	Restaurant	Supermarket	Sushi Restaurant	Bagel Shop	Fried Chicken Joint
51	Downtown Toronto	0	Restaurant	Coffee Shop	Pub	Italian Restaurant	Park	Bakery	Pizza Place	Café	Snack Place	Japanese Restaurant
52	Downtown Toronto	0	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Restaurant	Gay Bar	Gastropub	Hotel	Fast Food Restaurant	Italian Restaurant	Mediterranean Restaurant

Cluster 2 data:

	Borough	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 Venue	9th Most Common Venue	10th Most Common Venue
48	Central Toronto	1	Trail	Playground	Park	Tennis Court	Donut Shop	Diner	Discount Store	Dog Run	Doner Restaurant	Yoga Studio
50	Downtown Toronto	1	Park	Playground	Trail	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

4. Result:

The neighbourhood of Adelaide, King, Richmond has more 100 venues. The most common venue in this neighbour hood is the coffee shop. The venue information present in this neighbour hood are found and printed as a dataframe.

	name	categories	lat	lng
0	Four Seasons Centre for the Performing Arts	Concert Hall	43.650592	-79.385806
1	Nathan Phillips Square	Plaza	43.652270	-79.383516
2	The Keg Steakhouse & Bar	Steakhouse	43.649937	-79.384196
3	Rosalinda	Vegetarian / Vegan Restaurant	43.650252	-79.385156
4	Shangri-La Toronto	Hotel	43.649129	-79.386557
5	Noodle King	Asian Restaurant	43.651706	-79.383046
6	Canadian Opera Company	Opera House	43.650660	-79.386242
7	Soho House Toronto	Speakeasy	43.648734	-79.386541
8	M Square Coffee Co	Coffee Shop	43.651218	-79.383555
9	Bosk at Shangri-La	Asian Restaurant	43.649023	-79.385826
10	Cafe Landwer	Café	43.648753	-79.385367
11	Hy's Steakhouse	Steakhouse	43.649505	-79.382919
12	Pizzeria Libretto	Pizza Place	43.648334	-79.385111

For a tourist who wishes to visit a concert hall in this neighbourhood, then the Four seasons Centre for performing Arts would be one of his choice.

The neighbourhood that has the least number of venues is Roselawn with 2 venues.

	name	categories	lat	lng
0	Ceiling Champions	Home Service	43.713891	-79.420702
1	Rosalind's Garden Oasis	Garden	43.712189	-79.411978

As this place does not have much venues, it would be good to start a business like for example coffee shop or a restaurant as the tourists visiting the garden in this neighbourhood would like to have them.

5. Conclusion:

Thus the Toronto location data is analysed. One of the best neighbourhood for tourists is Adelaide, King, Richmond as it has more 100 venues. For the persons who wish to start a new business Roselawn area would be of a choice as it has just 2 venues and the people visiting this neighbour hood would like to see more venues such as coffee shop, restaurants etc.