The Battle of Toronto Neighborhoods

A S M Mahbub Afzal Khan

March 06, 2019

1. Introduction

1.1 Problem Description

I am living in Toronto. It is a multicultural place where everyday new immigrants are coming. Last few years the population of Toronto is increasing rapidly. When a new immigrant landed in Toronto, the first problem they face is to find a convenient place to live. Most of them have no personal vehicle so they want to live in a neighborhood where they can find all the services in a short distance.

Customer:

My customer is the settlement agencies who work on settlement issues for newcomers. They want to find the best neighbourhoods in Toronto based on available services in a short distance. Their customers are mostly interested on those neighborhoods.

Business Problem:

"Find best neighborhoods depending on the nearby available services"

1.2 Background

Toronto, the capital of the province of Ontario, is a major Canadian city. Nearly one million immigrants will be coming to Canada over the next three years, and tens of thousands of them will wind up in Toronto. Settlement services are one of the most important service provided by many government and non-profit organization. Everyday hundreds of the newcomers are coming

to get information about the neighborhoods where they can start their Canada life. Most of them want a place where the services are available within a short distance. The organizations are now trying to find the best neighborhoods based on the nearby services

2. Data Preparation

Toronto is a big city with a big number of neighbourhoods. To find the best neighborhood in Toronto based on the nearby services, different types of data are required. We need to find the name of the neighborhoods, their postal codes and geographical locations. We also need the services around Toronto and their geological locations so that we can determine the best neighbourhoods for different services.

Toronto neighborhoods are scraped from Wikipedia table. From this data, we'll find the names of the Neighborhoods, Postal codes and boroughs. The link of the Wikipedia is given below:

Toronto Neighbourhoods from Wikipedia

We also need the geological information like Latitude and Longitude for the postal areas to determine and show the places on the Map. Below is the link for the geographical data:

Geographical Information for Toronto

We use Foursquare API to get the venues inside Toronto and their geographical information. Please find the below link for foursquare.com:

Foursquare for Developers

We'll cleanup the data as per our requirements and use them to find the best neighborhoods based on nearby services.

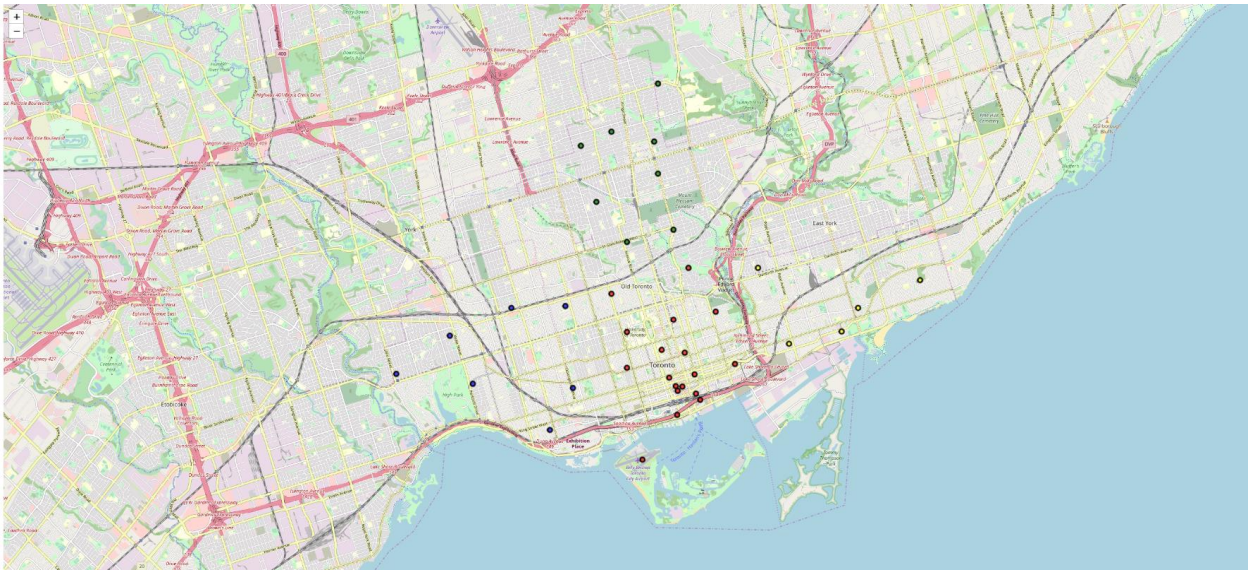
3. Data Cleaning and Preparation

Data downloaded or scraped from multiple sources. They are combined into one table. A lot of missing values were there. Toronto neighbourhood data was taken from Wikipedia and it contains three columns: ‘Postcode’, ’Borough’, ‘Neighbourhood’. I’ve decided to delete the records where there is no value or not assigned value for Borough. After deleting that, I check

the Neighbourhood column. There is also ‘Not Assigned’ value. I’ve replaced the ‘Not Assigned’ value with the value taken from ‘Borough’ column.

I’ve loaded geographical data loaded in a different data frame. It contains ‘Postal Code’, ‘Latitude’ and ‘Longitude’. I’ve combined the neighbourhood data with the geographical data based on the ‘Postal Code’ and assigned the data in a new data frame. The new data frame contains the columns: ‘Postcode’, ‘Borough’, ‘Neighbourhood’, ‘Latitude ‘, and ‘Longitude’.

I’ve further refined the data with the column ‘Borough’. I wanted the Toronto neighbourhood information, so I filter the data frame with the ‘Borough’ where the result contains ‘Toronto’.



4. Grabbing data from Foursquare

After collecting and processing data, I have the information about the Toronto neighbourhoods and their geographical data. Based on that information I’ve collected the Toronto neighbourhood data from foursquare. I’ve collected the popular places near Toronto neighbourhoods.

I’ve used my Foursquare developer account. Every time when I run the script, I collect data from the internet. I use ‘picking’ feature so that the downloaded data are saved and don’t need to access internet all time. I get a list of nearby places for all the Toronto neighbourhoods.

```
import pickle
with open("toronto_foursquare_dataset.txt", "wb") as fp:   #Pickling
    pickle.dump(toronto_foursquare_dataset, fp)
print('Received Data from Internet is Saved to Computer.')
```

```
Received Data from Internet is Saved to Computer.
```

```
with open("toronto_foursquare_dataset.txt", "rb") as fp:   # Unpickling
    toronto_foursquare_dataset = pickle.load(fp)
# print(type(Scarborough_foursquare_dataset))
toronto_foursquare_dataset
```

Head and Tail for the Toronto venues are attached with this report:

```
toronto_venues.head(15)
```

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Roselle Desserts	This spot is popular	Bakery	143
1	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Tandem Coffee	This spot is popular	Coffee Shop	122
2	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Toronto Cooper Koo Family Cherry St YMCA Centre	This spot is popular	Gym / Fitness Center	247
3	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Impact Kitchen	This spot is popular	Restaurant	376
4	M5A	Harbourfront,Regent Park	43.65426	-79.360636	The Distillery Historic District	This spot is popular	Historic Site	459
5	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Distillery Sunday Market	This spot is popular	Farmers Market	475
6	M5A	Harbourfront,Regent Park	43.65426	-79.360636	SOMA chocolatemaker	This spot is popular	Chocolate Shop	452
7	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Body Blitz Spa East	This spot is popular	Spa	80
8	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Rooster Coffee	This spot is popular	Coffee Shop	479
9	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Souk Tabule	This spot is popular	Mediterranean Restaurant	506
10	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Arvo	This spot is popular	Coffee Shop	482
11	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Figs Breakfast & Lunch	This spot is popular	Breakfast Spot	349
12	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Cacao 70	This spot is popular	Dessert Shop	466
13	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Corktown Common	This spot is popular	Park	387
14	M5A	Harbourfront,Regent Park	43.65426	-79.360636	Sumach Espresso	This spot is popular	Coffee Shop	440

```
toronto_venues.tail(15)
```

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
3064	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Athens Pastries Cafe	This spot is popular	Bakery	938
3065	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	YAYA	This spot is popular	Diner	890
3066	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Beach Volleyball	This spot is popular	Beach	955
3067	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Edward's 1290	This spot is popular	New American Restaurant	577
3068	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Burrito Bandidos	This spot is popular	Burrito Place	582
3069	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Murphy's Law	This spot is popular	Pub	879
3070	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Loblaws	This spot is popular	Grocery Store	598
3071	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Leslie Grove Park	This spot is popular	Park	957
3072	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Leslie Jones	This spot is popular	American Restaurant	827
3073	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Pet Valu	This spot is popular	Pet Store	728
3074	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Tim Hortons	This spot is popular	Coffee Shop	935
3075	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Measurement Park	This spot is popular	Park	855
3076	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Breakfast Club	This spot is popular	Breakfast Spot	916
3077	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Carters Landing	This spot is popular	Bistro	939
3078	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Ashbridges Bay Yacht Club (ABYC)	This spot is popular	Harbor / Marina	988

5. Categorized

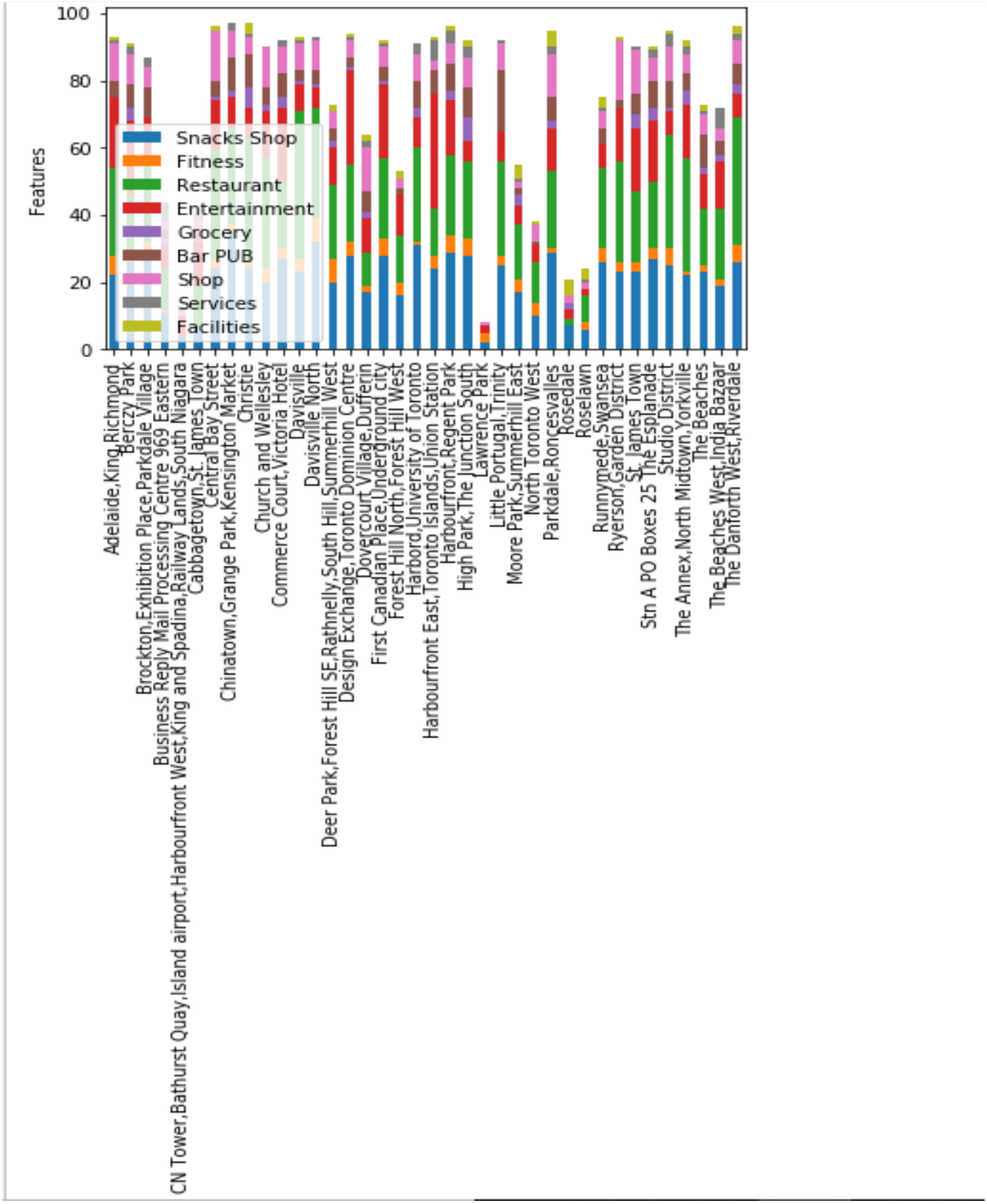
After collecting and processing data from Foursquare, I've found 272 unique categories.

```
print('There are {} uniques categories.'.format(len(toronto_venues['Venue Category'].unique())))

print('Here is the list of different categories:')
list(toronto_venues['Venue Category'].unique())
```

```
There are 272 uniques categories.
Here is the list of different categories:

']: ['Bakery',
 'Coffee Shop',
 'Gym / Fitness Center',
 'Restaurant',
 'Historic Site',
 'Farmers Market',
 'Chocolate Shop',
 'Spa',
 'Mediterranean Restaurant',
 'Breakfast Spot',
 'Dessert Shop',
 'Park',
 'Pub',
 'Performing Arts Venue',
 'Italian Restaurant',
 'Mexican Restaurant',
 'Liquor Store',
```

Based on that information, I've further categorized the features manually based on the similarity and importance. And finally end up with 9 categories.

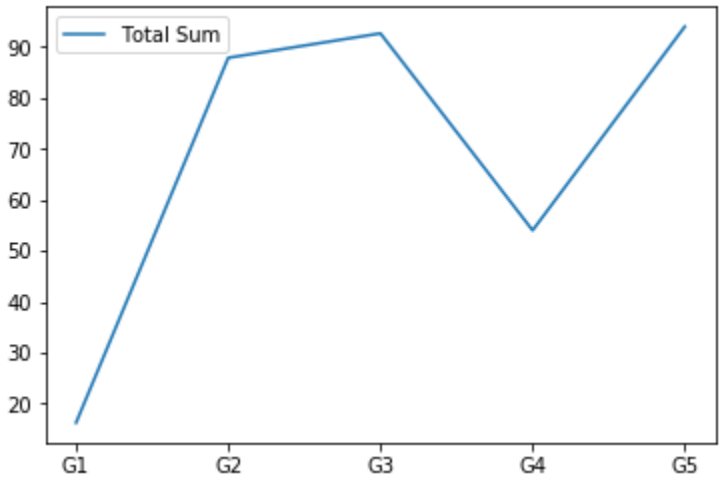
Neighborhood	Snacks Shop	Fitness	Restaurant	Entertainment	Grocery	Bar PUB	Shop	Services	Facilities
Adelaide,King,Richmond	22	6	26	21	0	5	11	1	1
Berczy Park	26	2	18	22	4	7	9	2	1
Brockton,Exhibition Place,Parkdale Village	27	5	23	14	0	9	6	3	0
Business Reply Mail Processing Centre 969 Eastern	11	1	12	7	2	3	2	6	0
CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South Niagara	4	0	0	6	0	0	1	2	0



6. Clustering

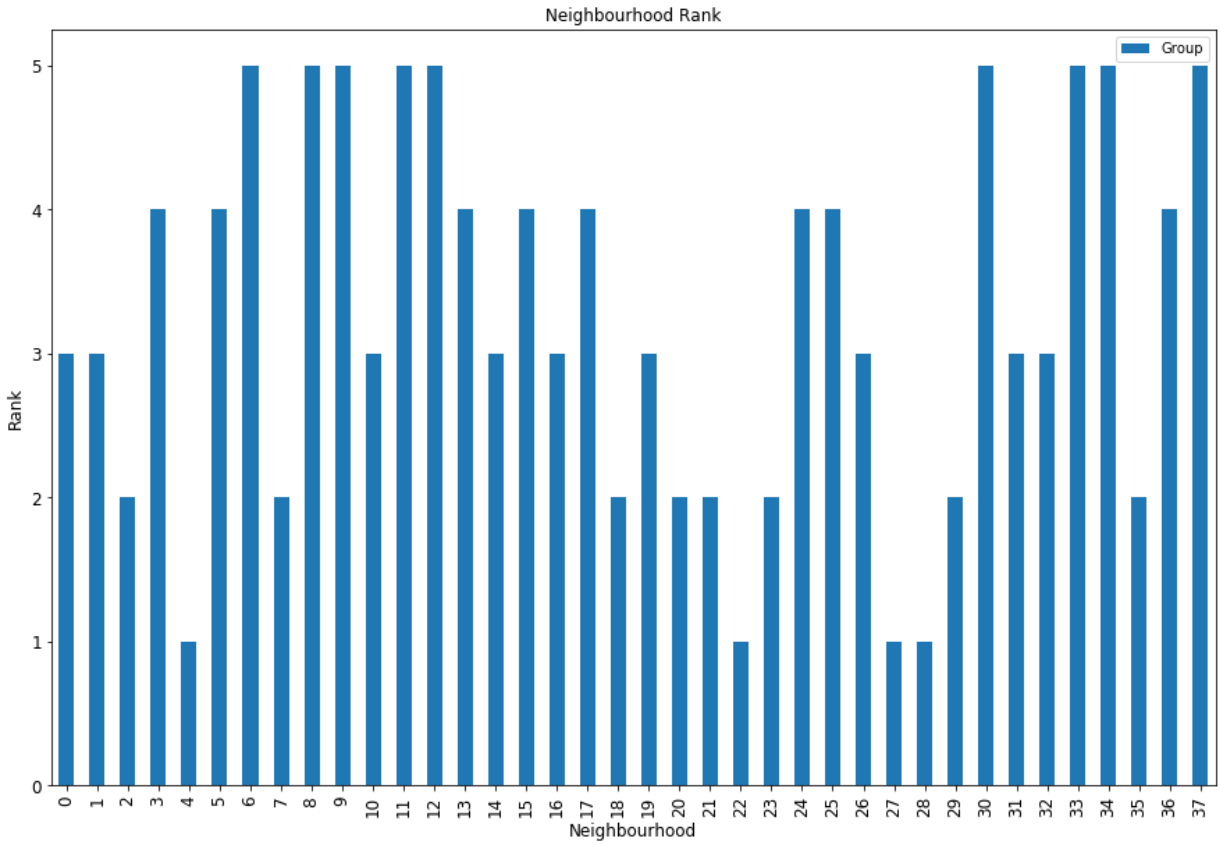
Using K-means clustering, data grouped by five groups. I've used 5 cluster to categized Neighbourhoods. I've calculated the means and the total score of the Neighbourhoods based on the near by facilities. Scored all the features selected previously and get their means and calculated the total sum. Based on the total sum, grouped the neighbourhoods. The higher the total sum, the best value of the neighbourhoods.

	Snacks Shop	Fitness	Restaurant	Entertainment	Grocery	Bar PUB	Shop	Services	Facilities	Total Sum
G4	24.333333	3.777778	35.000000	10.888889	2.111111	5.333333	10.000000	1.222222	1.222222	93.888889
G1	25.625000	3.750000	20.750000	23.250000	2.125000	5.875000	7.750000	2.000000	0.750000	91.875000
G5	28.111111	3.111111	24.333333	10.333333	2.111111	9.333333	7.666667	2.222222	1.444444	88.666667
G3	14.625000	3.000000	14.875000	9.750000	1.500000	3.000000	4.750000	2.000000	1.375000	54.875000
G2	4.750000	1.250000	2.500000	3.250000	0.500000	0.000000	1.500000	0.750000	2.000000	16.500000



7. Ranking

I've ranked the Neighbourhood based on the group. From the group ranking, I've found the Group 5 neighbourhoods are the best based on the nearby facilities. Group 3 is second and group 2 is number third. Now, from the group, I extract the neighbourhood.

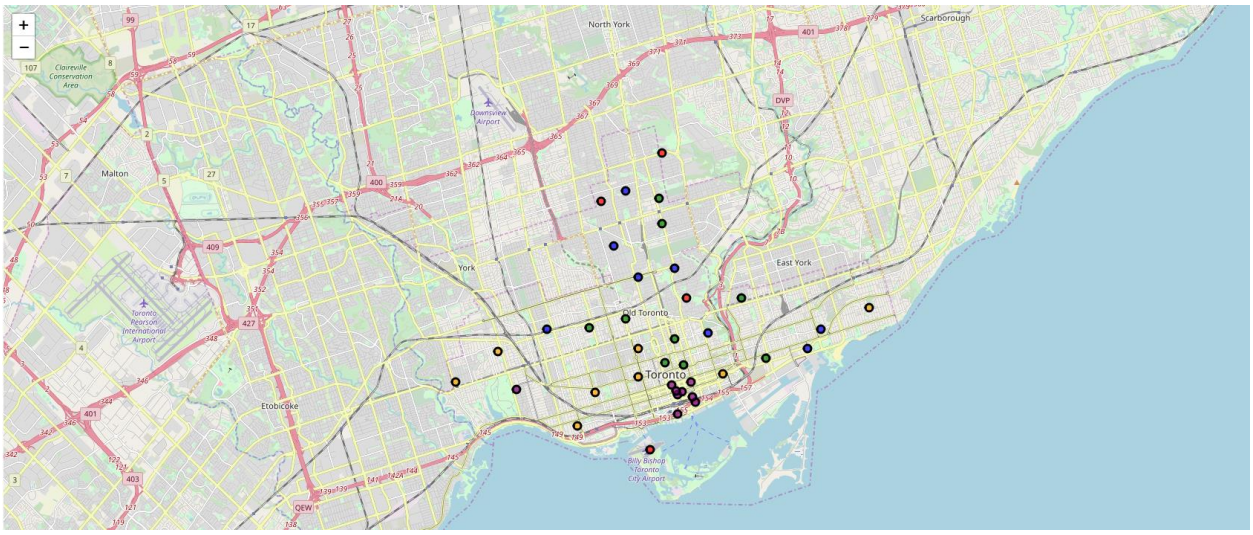


The best neighbourhoods are given below:

	Neighbourhood	Group
6	Central Bay Street	5
8	Christie	5
9	Church and Wellesley	5
11	Davisville	5
12	Davisville North	5
30	Ryerson,Garden District	5
33	Studio District	5
34	The Annex,North Midtown,Yorkville	5
37	The Danforth West,Riverdale	5

8. Best Neighbourhoods

After calculating the features of the neighbourhood, I've found 9 Neighbourhoods that's are the best scored. I've put them into the Toronto maps. Color 'Green' neighbourhoods are scored top in my analysis.



9. Conclusion

I've found the best Neighbourhoods based on the nearby services captured from Foursquare. It can be more refined by taking consideration of the Transportation, Subway, populations, educations etc. That will give more accurate ranking but, in this analysis, I've considered all the features and It'll also give approximately the best ranking.