Predicting the Best Location to Open a New Restaurant

Mizanur Rahman

December 10, 2019

1. Introduction

1.1 Background

Toronto is one of the diverse and popular cities in Canada. It is also being recognized as one of the financial capitals of Canada. There are several restaurants available in Toronto because of the cultural diversity which leads people trying out different restaurants and testing their cuisines. It is difficult to find the best location to open a new restaurant in the city of Toronto.

1.2 Problem

The goal of this project is to help finding the best location to open a new restaurant in order to gain maximum profit and engagement. In short, the problem is "which is a good location in Toronto to open up a new restaurant?" There are several other factors need to be considered to choose the best place such as public transport facilities, the security of the area, the rental price of the location. This project gives people a brief idea for opening new restaurants in the city of Toronto.

1.3 Interest

It is challenging to find the best location to open a new restaurant. Stakeholders who are interested to open a new restaurant are interested to find an optimum location to get maximum profit.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

For this project, I acquired the data from two different sources. The first data set is acquired through scrapping from this link- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

The dataset contains the following columns;

- Postcode
- Borough and
- Neighborhood

The latitude and longitude of these borough and neighborhood are acquired from this link: https://cocl.us/Geospatial_data.

2.2 Data Cleaning

Both data sets need to be merged to get the complete data set. Columns are renamed and then merged.

```
df_neigh = df.merge(geo,left_on=['PostalCode'],right_on=['PostalCode'], how='inner')
df_neigh.head()
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Heights	43.718518	-79.464763
4	M6A	North York	Lawrence Manor	43.718518	-79.464763

Then I sliced the original data frame and create a new data frame of the Toronto data

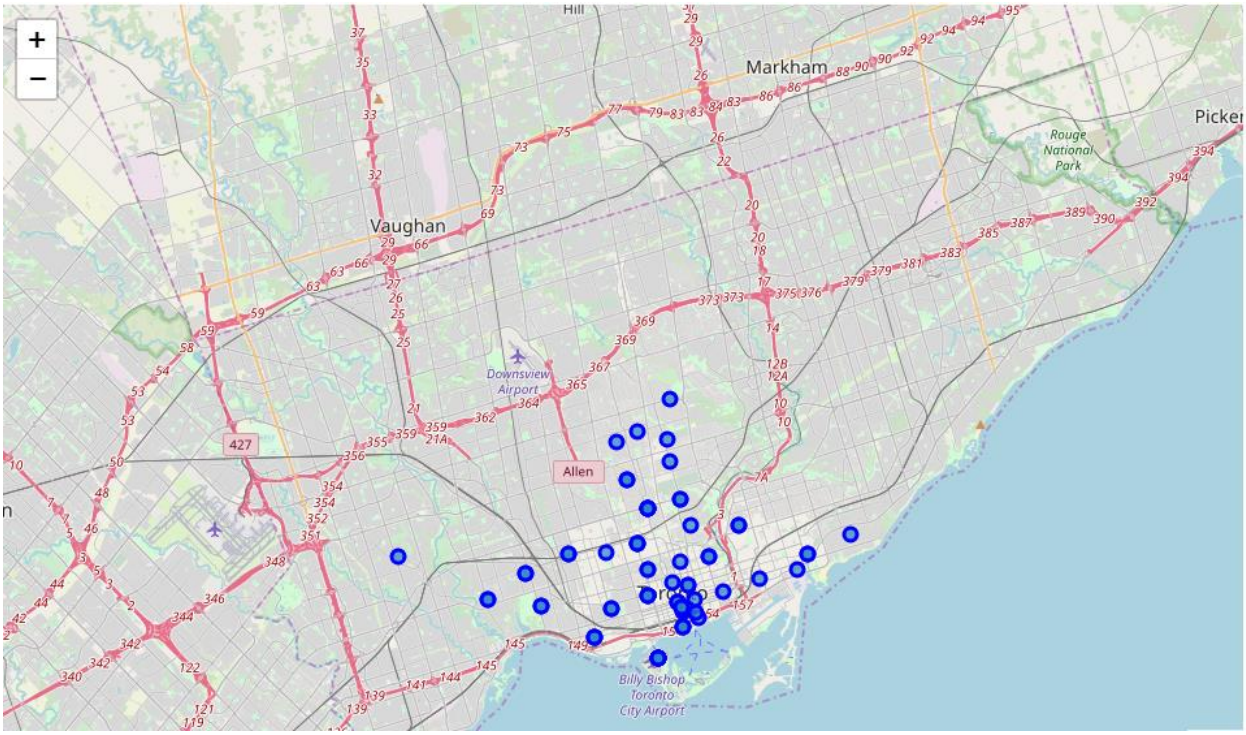
```
toronto_df = df_neigh[df_neigh['Borough'].str.contains('Toronto', na = False)]
toronto_df.reset_index(drop = True, inplace = True)
toronto_df.head()
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
1	M9A	Downtown Toronto	Queen's Park	43.667856	-79.532242
2	M5B	Downtown Toronto	Ryerson	43.657162	-79.378937
3	M5B	Downtown Toronto	Garden District	43.657162	-79.378937
4	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418

Then, using the four-square API, I explored the city of Toronto. I sent a call with requesting with a limit of 200 venues that are in Harbourfront and within a radius of 500 m. Finally, I sent a call with requesting with a limit of 200 venues that are in Toronto and within a radius of 1000 m.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653191	-79.357947	Gym / Fitness Center
3	Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
4	Harbourfront	43.65426	-79.360636	The Distillery Historic District	43.650244	-79.359323	Historic Site

I created a map of Toronto with neighborhoods superimposed on top.



I created a dataset called Toronto restaurants where type is Restaurant.

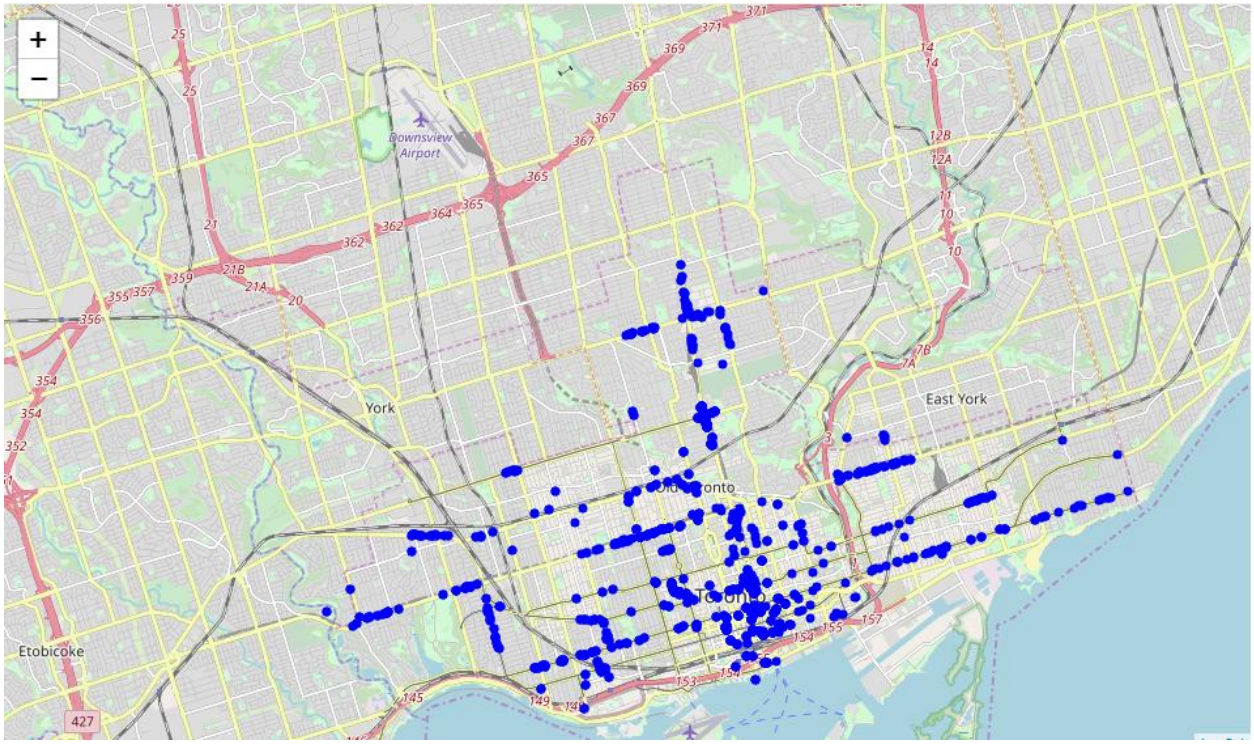
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
1	Harbourfront	43.654260	-79.360636	Souk Tabule	43.653756	-79.354390	Mediterranean Restaurant
2	Harbourfront	43.654260	-79.360636	Mangia and Bevi Resto-Bar	43.652250	-79.366355	Italian Restaurant
3	Harbourfront	43.654260	-79.360636	Cluny Bistro & Boulangerie	43.650565	-79.357843	French Restaurant
4	Harbourfront	43.654260	-79.360636	El Catrin	43.650601	-79.358920	Mexican Restaurant
5	Harbourfront	43.654260	-79.360636	Cocina Economica	43.654959	-79.365657	Mexican Restaurant
6	Harbourfront	43.654260	-79.360636	Sukhothai	43.658444	-79.365681	Thai Restaurant
7	Harbourfront	43.654260	-79.360636	Souvlaki Express	43.655584	-79.364438	Greek Restaurant
8	Harbourfront	43.654260	-79.360636	Fusaro's	43.653347	-79.369517	Italian Restaurant
9	Harbourfront	43.654260	-79.360636	Archeo	43.650667	-79.359431	Italian Restaurant

3. Methodology

3.1 Data Analysis and exploration

I performed some basic analysis and derived some additional info our raw data.

Each neighborhood is analyzed though one hot process. Then we found the top 5 restaurants in each neighborhood. Next, I visualized all the restaurants in Toronto.



K means is run to cluster the neighborhood into 5 clusters.

3.2 Clustering Results:

Cluster 1:

	Neighborhood	Restaurant	Hotel	Movie Theater	Park	University	Cluster Labels	PostalCode	Borough	Latitude	Longitude
7	Central Bay Street	0.010000	0.01	0.01	0.030000	0.01	0	M5G	Downtown Toronto	43.657952	-79.387383
8	Chinatown	0.010000	0.00	0.00	0.000000	0.00	0	M5T	Downtown Toronto	43.653206	-79.400049
9	Christie	0.010000	0.00	0.00	0.020000	0.00	0	M6G	Downtown Toronto	43.669542	-79.422564
12	Davisville	0.020000	0.00	0.02	0.010000	0.00	0	M4S	Central Toronto	43.704324	-79.388790
13	Davisville North	0.020000	0.01	0.02	0.010000	0.00	0	M4P	Central Toronto	43.712751	-79.390197
24	Grange Park	0.010000	0.00	0.00	0.000000	0.00	0	M5T	Downtown Toronto	43.653206	-79.400049
29	High Park	0.020000	0.00	0.00	0.030000	0.00	0	M6P	West Toronto	43.661608	-79.464763
32	Kensington Market	0.010000	0.00	0.00	0.000000	0.00	0	M5T	Downtown Toronto	43.653206	-79.400049
40	Parkdale	0.030000	0.00	0.01	0.010000	0.00	0	M6R	West Toronto	43.648960	-79.456325
46	Riverdale	0.030000	0.00	0.00	0.010000	0.00	0	M4K	East Toronto	43.679557	-79.352188
47	Roncesvalles	0.030000	0.00	0.01	0.010000	0.00	0	M6R	West Toronto	43.648960	-79.456325
49	Roselawn	0.000000	0.00	0.00	0.000000	0.00	0	M5N	Central Toronto	43.711695	-79.416936
50	Runnymede	0.027397	0.00	0.00	0.027397	0.00	0	M6S	West Toronto	43.651571	-79.484450
56	Studio District	0.010000	0.01	0.00	0.020000	0.00	0	M4M	East Toronto	43.659526	-79.340923
59	Swansea	0.027397	0.00	0.00	0.027397	0.00	0	M6S	West Toronto	43.651571	-79.484450
61	The Beaches	0.012195	0.00	0.00	0.036585	0.00	0	M4E	East Toronto	43.676357	-79.293031
63	The Danforth West	0.030000	0.00	0.00	0.010000	0.00	0	M4K	East Toronto	43.679557	-79.352188
64	The Junction South	0.020000	0.00	0.00	0.030000	0.00	0	M6P	West Toronto	43.661608	-79.464763

Cluster 2:

	Neighborhood	Restaurant	Hotel	Movie Theater	Park	University	Cluster Labels	PostalCode	Borough	Latitude	Longitude
6	Cabbagetown	0.027778	0.000000	0.000000	0.055556	0.0	1	M4X	Downtown Toronto	43.667967	-79.367675
10	Church and Wellesley	0.020000	0.020000	0.000000	0.040000	0.0	1	M4Y	Downtown Toronto	43.665860	-79.383160
14	Deer Park	0.025974	0.012987	0.012987	0.064935	0.0	1	M4V	Central Toronto	43.686412	-79.400049
21	Forest Hill SE	0.025974	0.012987	0.012987	0.064935	0.0	1	M4V	Central Toronto	43.686412	-79.400049
26	Harbourfront	0.040000	0.000000	0.000000	0.040000	0.0	1	M5A	Downtown Toronto	43.654260	-79.360636
30	India Bazaar	0.025000	0.000000	0.012500	0.050000	0.0	1	M4L	East Toronto	43.668999	-79.315572
37	Moore Park	0.033898	0.000000	0.016949	0.050847	0.0	1	M4T	Central Toronto	43.689574	-79.383160
39	North Toronto West	0.022727	0.000000	0.000000	0.068182	0.0	1	M4R	Central Toronto	43.715383	-79.405678
44	Rathnelly	0.025974	0.012987	0.012987	0.064935	0.0	1	M4V	Central Toronto	43.686412	-79.400049
52	South Hill	0.025974	0.012987	0.012987	0.064935	0.0	1	M4V	Central Toronto	43.686412	-79.400049
57	Summerhill East	0.033898	0.000000	0.016949	0.050847	0.0	1	M4T	Central Toronto	43.689574	-79.383160
58	Summerhill West	0.025974	0.012987	0.012987	0.064935	0.0	1	M4V	Central Toronto	43.686412	-79.400049
62	The Beaches West	0.025000	0.000000	0.012500	0.050000	0.0	1	M4L	East Toronto	43.668999	-79.315572

Cluster 3:

	Neighborhood	Restaurant	Hotel	Movie Theater	Park	University	Cluster Labels	PostalCode	Borough	Latitude	Longitude
0	Adelaide	0.020000	0.050000	0.020000	0.000000	0.01	2	M5H	Downtown Toronto	43.650571	-79.384568
2	Berczy Park	0.040000	0.050000	0.010000	0.020000	0.00	2	M5E	Downtown Toronto	43.644771	-79.373306
11	Commerce Court	0.030000	0.050000	0.000000	0.010000	0.00	2	M5L	Downtown Toronto	43.648198	-79.379817
15	Design Exchange	0.020000	0.070000	0.010000	0.020000	0.00	2	M5K	Downtown Toronto	43.647177	-79.381576
19	First Canadian Place	0.030000	0.070000	0.010000	0.010000	0.00	2	M5X	Downtown Toronto	43.648429	-79.382280
27	Harbourfront East	0.050000	0.060000	0.010000	0.030000	0.00	2	M5J	Downtown Toronto	43.640816	-79.381752
33	King	0.020000	0.050000	0.020000	0.000000	0.01	2	M5H	Downtown Toronto	43.650571	-79.384568
45	Richmond	0.020000	0.050000	0.020000	0.000000	0.01	2	M5H	Downtown Toronto	43.650571	-79.384568
54	St. James Town	0.036765	0.036765	0.007353	0.029412	0.00	2	M5C	Downtown Toronto	43.651494	-79.375418
54	St. James Town	0.036765	0.036765	0.007353	0.029412	0.00	2	M4X	Downtown Toronto	43.667967	-79.367675
55	Stn A PO Boxes 25 The Esplanade	0.060000	0.050000	0.010000	0.010000	0.00	2	M5W	Downtown Toronto	43.646435	-79.374846
65	Toronto Dominion Centre	0.020000	0.070000	0.010000	0.020000	0.00	2	M5K	Downtown Toronto	43.647177	-79.381576
66	Toronto Islands	0.050000	0.060000	0.010000	0.030000	0.00	2	M5J	Downtown Toronto	43.640816	-79.381752
68	Underground city	0.030000	0.070000	0.010000	0.010000	0.00	2	M5X	Downtown Toronto	43.648429	-79.382280
69	Union Station	0.050000	0.060000	0.010000	0.030000	0.00	2	M5J	Downtown Toronto	43.640816	-79.381752
71	Victoria Hotel	0.030000	0.050000	0.000000	0.010000	0.00	2	M5L	Downtown Toronto	43.648198	-79.379817

Cluster 4:

	Neighborhood	Restaurant	Hotel	Movie Theater	Park	University	Cluster Labels	PostalCode	Borough	Latitude	Longitude
3	Brockton	0.05	0.02	0.00	0.02	0.0	3	M6K	West Toronto	43.636847	-79.428191
18	Exhibition Place	0.05	0.02	0.00	0.02	0.0	3	M6K	West Toronto	43.636847	-79.428191
23	Garden District	0.03	0.02	0.01	0.02	0.0	3	M5B	Downtown Toronto	43.657162	-79.378937
25	Harbord	0.04	0.01	0.00	0.02	0.0	3	M5S	Downtown Toronto	43.662696	-79.400049
36	Little Portugal	0.05	0.01	0.00	0.02	0.0	3	M6J	West Toronto	43.647927	-79.419750
38	North Midtown	0.04	0.01	0.00	0.01	0.0	3	M5R	Central Toronto	43.672710	-79.405678
41	Parkdale Village	0.05	0.02	0.00	0.02	0.0	3	M6K	West Toronto	43.636847	-79.428191
51	Ryerson	0.03	0.02	0.01	0.02	0.0	3	M5B	Downtown Toronto	43.657162	-79.378937
60	The Annex	0.04	0.01	0.00	0.01	0.0	3	M5R	Central Toronto	43.672710	-79.405678
67	Trinity	0.05	0.01	0.00	0.02	0.0	3	M6J	West Toronto	43.647927	-79.419750
70	University of Toronto	0.04	0.01	0.00	0.02	0.0	3	M5S	Downtown Toronto	43.662696	-79.400049
72	Yorkville	0.04	0.01	0.00	0.01	0.0	3	M5R	Central Toronto	43.672710	-79.405678

Cluster 5:

	Neighborhood	Restaurant	Hotel	Movie Theater	Park	University	Cluster Labels	PostalCode	Borough	Latitude	Longitude
1	Bathurst Quay	0.000000	0.0	0.0	0.066667	0.0	4	M5V	Downtown Toronto	43.628947	-79.394420
4	Business Reply Mail Processing Centre 969 Eastern	0.020833	0.0	0.0	0.083333	0.0	4	M7Y	East Toronto	43.662744	-79.321558
5	CN Tower	0.000000	0.0	0.0	0.066667	0.0	4	M5V	Downtown Toronto	43.628947	-79.394420
16	Dovercourt Village	0.014493	0.0	0.0	0.086957	0.0	4	M6H	West Toronto	43.669005	-79.442259
17	Dufferin	0.014493	0.0	0.0	0.086957	0.0	4	M6H	West Toronto	43.669005	-79.442259
20	Forest Hill North	0.000000	0.0	0.0	0.086957	0.0	4	M5P	Central Toronto	43.696948	-79.411307
22	Forest Hill West	0.000000	0.0	0.0	0.086957	0.0	4	M5P	Central Toronto	43.696948	-79.411307
28	Harbourfront West	0.000000	0.0	0.0	0.066667	0.0	4	M5V	Downtown Toronto	43.628947	-79.394420
31	Island airport	0.000000	0.0	0.0	0.066667	0.0	4	M5V	Downtown Toronto	43.628947	-79.394420
34	King and Spadina	0.000000	0.0	0.0	0.066667	0.0	4	M5V	Downtown Toronto	43.628947	-79.394420
35	Lawrence Park	0.000000	0.0	0.0	0.125000	0.0	4	M4N	Central Toronto	43.728020	-79.388790
42	Queen's Park	0.000000	0.0	0.0	0.090909	0.0	4	M9A	Downtown Toronto	43.667856	-79.532242
43	Railway Lands	0.000000	0.0	0.0	0.066667	0.0	4	M5V	Downtown Toronto	43.628947	-79.394420
48	Rosedale	0.000000	0.0	0.0	0.090909	0.0	4	M4W	Downtown Toronto	43.679563	-79.377529
53	South Niagara	0.000000	0.0	0.0	0.066667	0.0	4	M5V	Downtown Toronto	43.628947	-79.394420

4. Results and Discussion:

The analysis shows that although there is a great number of restaurants in Toronto, there are areas of low restaurant density as well as other amenities like Movie theatre, Universities, Parks etc. Highest concentration of restaurants was detected on cluster 4 and lowest concentration was detected on cluster 5.

Final decision on finding the best location will be made by stakeholders based on specific characteristics of neighborhoods, and locations in every recommended zone, taking into consideration additional factors like rental prices, security, and other an attractiveness.

5. Conclusion:

This Project helps to better understand the venues in that neighborhood. After analyzing all the neighborhoods, I found that cluster 3 has a smaller number of restaurants and thus it would be best to consider open a restaurant there. For better understanding, other factors such as crime rates, public transport facilities, rental price should be added into consideration. The final decision on finding the best location will be made by Stakeholders.