The Battle of Neighborhoods - Problem Case:

Rodrigo Alencar

September 13, 2019

1) Introduction

The city of Toronto is the financial capital and the most populous city of Canada, with a population growing rapidly each year. The ethnicity of Toronto is very mixed, almost 12.5 per cent being Chinese, and the immigration of Asian country like China and Indian to Canada is growing each year.

Aladdin's is a well-known Indian food restaurant target medium income neighborhoods; they are interested to know which Toronto's neighborhoods are the best to install their food chain restaurant.

As a consulting firm, they requested to do these analyses and find out the best neighborhoods to present to the Aladdin's CEO.

2) Datas:

2.1) Data sources:

To find the best neighborhoods with Indian food we should get those datas:

1. Toronto list of boroughs and neighborhood with their postcode.

    This data set was acquired by Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. Toronto list of incomes per neighborhood.

    This data set was acquired by Wikipedia page https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

3. Neighborhood with the most Indian restaurant.

    This dataset is provided by the Foursquare API.

4. Geo location dataset to visualize the best data with folium library.

    This dataset is provided by http://cocl.us/Geospatial_data

2.2) Data acquisition and cleaning:

The data set with the list of neighborhoods in Toronto was downloaded by the Wikipedia page with the name of df_toronto.

First, there was some Boroughs without data, in example ''Not assigned', then we needed to remove all borough without data, as seen below:

	Postcode	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights

Table 01 – List of Neighborhood in Toronto

We needed to know if there was any neighborhood without name, in example 'Not Assigned', in this case we discovered only one borough with a neighborhood without name, Queen's Park borough, in this case we changed the name in the cell with the name of the borough.

```
df_toronto[df_toronto.Neighbourhood == 'Not assigned']
```

Image 01 – code to discover neighborhoods without name.

	Postcode	Borough	Neighbourhood
8	M7A	Queen's Park	Not assigned

Image 02 –Neighborhood without name.

	Postcode	Borough	Neighbourhood
8	M7A	Queen's Park	Queen's Park

Image 03 – Neighborhoods with same name as Borough.

Then we organized the data set grouping the Neighborhood with tame same postcode and putting the names of the neighborhoods separated by a comma.

	Postcode	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 He...

Table 02 – Organized list of neighborhoods grouped by postcode.

We now can import the data set with the coordinates of the post codes with the geospatial data and merge the two data sets into one that will be called df_toronto_coor.

	Postcode	Borough	Neighborhood	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
9	M1N	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...	43.757410	-79.273304

Table 03 – Data set with coordinates.

Now it is possible to create a map showing all the boroughs in Toronto with folium library. Each blue dot is a different borough.
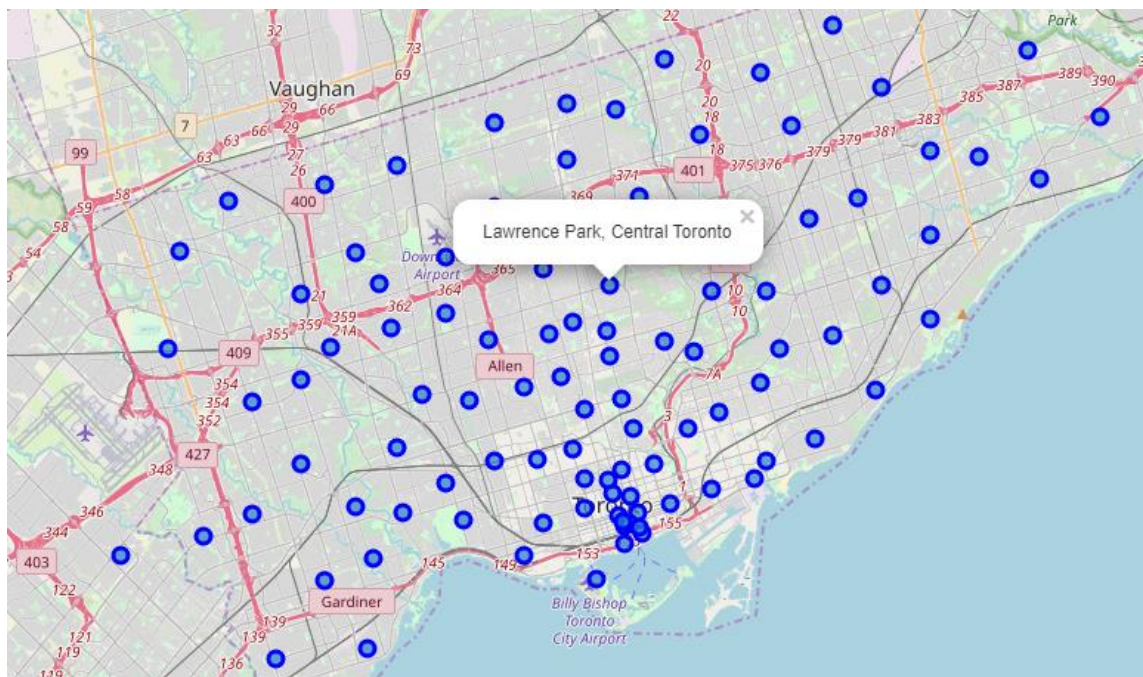

Image 04 – Geospatial map with all boroughs of Toronto.

2.3) Foursquare API:

Now is time to get the datas from the foursquare api, we want to know how many Indian Restaurants there is by neighborhood and create a knew data set with a list of restaurants and merge it with the df_toronto_coor data set.

For that, we crate a function that calls all the venues separated by neighborhood, this gave a data set of over two thousand rows, and then divided this data set to get the unique venues for each neighborhood using the get_dummies function. Finally, we had a data

set sat of over two thousand rows and over two hundred and seventy columns (different category).

   With this big data set, we only needed the column of India Restaurant venues per neighborhood, then we just pop this column from the rest, and called toronto_neighborhood_indianfood. With that data set we grouped by neighborhood and counted to get the final data set with the sum of all Indian Restaurants per neighborhood.

	Neighborhood	Indian Restaurant
0	Adelaide,King,Richmond	100
20	Chinatown,Grange Park,Kensington Market	100
81	St. James Town	100
78	Ryerson,Garden District	100
44	First Canadian Place,Underground city	100
32	Design Exchange,Toronto Dominion Centre	100
27	Commerce Court,Victoria Hotel	100
50	Harbourfront East,Toronto Islands,Union Station	100
82	Stn A PO Boxes 25 The Esplanade	97
22	Church and Wellesley	88
19	Central Bay Street	86
43	Fairview,Henry Farm,Oriole	66
65	Little Portugal,Trinity	64
8	Berczy Park	57
51	Harbourfront,Regent Park	51
87	The Danforth West,Riverdale	44
15	Cabbagetown,St. James Town	44

Table 04 – Amount of Indian foods per neighborhood.

   Now, we merger this data set with the df_toronto to create the df_toronto_indian with the locations and amount of Indian Restaurants.

	Postcode	Borough	Neighborhood	Latitude	Longitude	Indian Restaurant
57	M5J	Downtown Toronto	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752	100
59	M5L	Downtown Toronto	Commerce Court,Victoria Hotel	43.648198	-79.379817	100
52	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937	100
53	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	100
56	M5H	Downtown Toronto	Adelaide,King,Richmond	43.650571	-79.384568	100
68	M5X	Downtown Toronto	First Canadian Place,Underground city	43.648429	-79.382280	100

Table 05 – df_toronto_indian data set.

We just need the average income per neighborhood column to start creating the model and then recommend the best neighborhood for the clients.

The data set with the list of income of Toronto's neighborhoods was downloaded by the Wikipedia page with the name of df_toronto_post.

	Neighborhood	Population	Average Income
0	Toronto CMA Average	5113149	40704
1	Agincourt	44577	25750
2	Alderwood	11656	35239
3	Alexandra Park	4355	19687
4	Allenby	2513	245592
5	Amesbury	17318	27546
6	Armour Heights	4384	116651
7	Banbury	6641	92319
8	Bathurst Manor	14945	34169
9	Bay Street Corridor	4787	40598
10	Bayview Village	12280	46752

Table 06 – df_toronto_post data set.

Then, we merged the last two columns with the df_toronto_indian to create the final data set called df_toronto_indian_demo, with all the data necessary to create a good model and start the analyses.

	Postcode	Borough	Neighborhood	Latitude	Longitude	Indian Restaurant	Population	Average Income
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	1	33524.000000	27453.500000
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	1	12156.666667	38205.000000
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	7	16641.333333	31960.333333
3	M1G	Scarborough	Woburn	43.770992	-79.216917	3	48507.000000	26190.000000
4	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	1	12796.000000	24413.000000
5	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029	7	13025.000000	25078.000000
6	M1L	Scarborough	Clairlea,Golden Mile,Oakridge	43.711112	-79.284577	10	12236.000000	27273.500000
7	M1M	Scarborough	Cliffcrest,Cliffside,Scarborough Village West	43.716316	-79.239476	4	11958.500000	35441.500000
8	M1N	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848	4	12266.000000	48965.000000
9	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...	43.757410	-79.273304	7	14189.000000	26525.000000
10	M1R	Scarborough	Maryvale,Wexford	43.750072	-79.295849	7	13322.000000	29750.000000

Table 07 – df_toronto_indian_demo data set.

Methodology:

3.1) Exploratory Data Analysis:

1. Get the neighborhoods and locations with df_toronto_coor.
2. Get the average income per neighborhood
3. Get the venues from Foursquare API for all neighborhoods.
4. Filter the category of Indian Restaurants.
5. Get the amount of Indian restaurants per neighborhood.
6. Group the amount of Indian restaurants per neighborhood with the average of income
7. Determine which neighborhood has the most and least Indian restaurants.
8. Determine by Kmeans cluster algorithms which neighborhood has the bigger average price and the medium price.
9. Decide which locations is the best fit for the Indian Restaurant Chain to place.

3.2) Regression model:

To make this model, it was used the Kmeans algorithms with the scikit learn package, it is an unsupervised machine learning algorithm that permits to create a model to segregate different types of data in clusters that shares the most similarity.

For this case, we want to select neighborhood with the most Indian Restaurant and the least average income to present for the CEO of Aladdin's,

for this we create there cluster (0, 1 and 2) and chose the neighborhoods from the 0 cluster, and picked the top 3 borough with the most Indian Restaurant.

	Postcode	Borough	Neighborhood	Latitude	Longitude	Indian Restaurant	Population	Average Income	Clusters
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	1	33524.000000	27453.500000	2
14	M2K	North York	Bayview Village	43.786947	-79.385975	4	12280.000000	46752.000000	2
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	1	12156.666667	38205.000000	2
25	M4S	Central Toronto	Davisville	43.704324	-79.388790	39	23727.000000	55735.000000	2
23	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	44	31007.000000	40139.000000	2
22	M4H	East York	Thorncliffe Park	43.705369	-79.349372	19	17949.000000	25340.000000	2
20	M4E	East Toronto	The Beaches	43.676357	-79.293031	4	20416.000000	67536.000000	2
40	M6L	North York	Downsview,North Park,Upwood Park	43.713756	-79.490074	4	36613.000000	26751.000000	2
18	M3H	North York	Bathurst Manor,Downsview North,Wilson Heights	43.754328	-79.442259	19	14338.500000	36073.500000	2
17	M3C	North York	Flemingdon Park,Don Mills South	43.725900	-79.340923	24	21287.000000	23471.000000	2
16	M3A	North York	Parkwoods	43.753259	-79.329656	3	26533.000000	34811.000000	2
15	M2L	North York	Silver Hills,York Mills	43.757490	-79.374714	1	17564.000000	92099.000000	2
19	M4A	North York	Victoria Village	43.725882	-79.315572	6	17047.000000	29657.000000	2
5	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029	7	13025.000000	25078.000000	2
12	M1V	Scarborough	Agincourt North,L'Amoreaux East,Milliken,Steel...	43.815252	-79.284577	2	26272.000000	25243.000000	2
11	M1S	Scarborough	Agincourt	43.794200	-79.262029	4	44577.000000	25750.000000	2
10	M1R	Scarborough	Maryvale,Wexford	43.750072	-79.295849	7	13322.000000	29750.000000	2

Table 08 – df_toronto_indian_demo data set with the cluster column.

	Postcode	Borough	Neighborhood	Latitude	Longitude	Indian Restaurant	Population	Average Income	Clusters	Income
31	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937	100	8240.0	37614.0	0	37614.0
32	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	100	14666.0	22341.0	0	22341.0
33	M5J	Downtown Toronto	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752	100	627.0	43344.0	0	43344.0
36	M5T	Downtown Toronto	Chinatown,Grange Park,Kensington Market	43.653206	-79.400049	100	6373.5	29306.0	0	29306.0
30	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	87	13397.0	37653.0	0	37653.0
13	M2J	North York	Fairview,Henry Farm,Oriole	43.778517	-79.346556	64	2790.0	56395.0	0	56395.0
38	M6J	West Toronto	Little Portugal,Trinity	43.647927	-79.419750	64	5013.0	29224.0	0	29224.0
29	M4X	Downtown Toronto	Cabbagetown,St. James Town	43.667967	-79.367675	45	12893.0	36369.5	0	36369.5

Table 09 – df_toronto_indian_demo only 0 cluster neighborhoods.

Results:

The result was five neighborhood in Toronto in the same borough who are low average income, similar close to each other end have one hundred Indian Restaurants in it, they are Ryerson,Garden Distric, St. James Town, Harbourfront East,Toronto Islands,Union Station.

	Postcode	Borough	Neighborhood	Latitude	Longitude	Indian Restaurant	Population	Average Income	Clusters	Income
31	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937	100	8240.0	37614.0	0	37614.0
32	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	100	14666.0	22341.0	0	22341.0
33	M5J	Downtown Toronto	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752	100	627.0	43344.0	0	43344.0

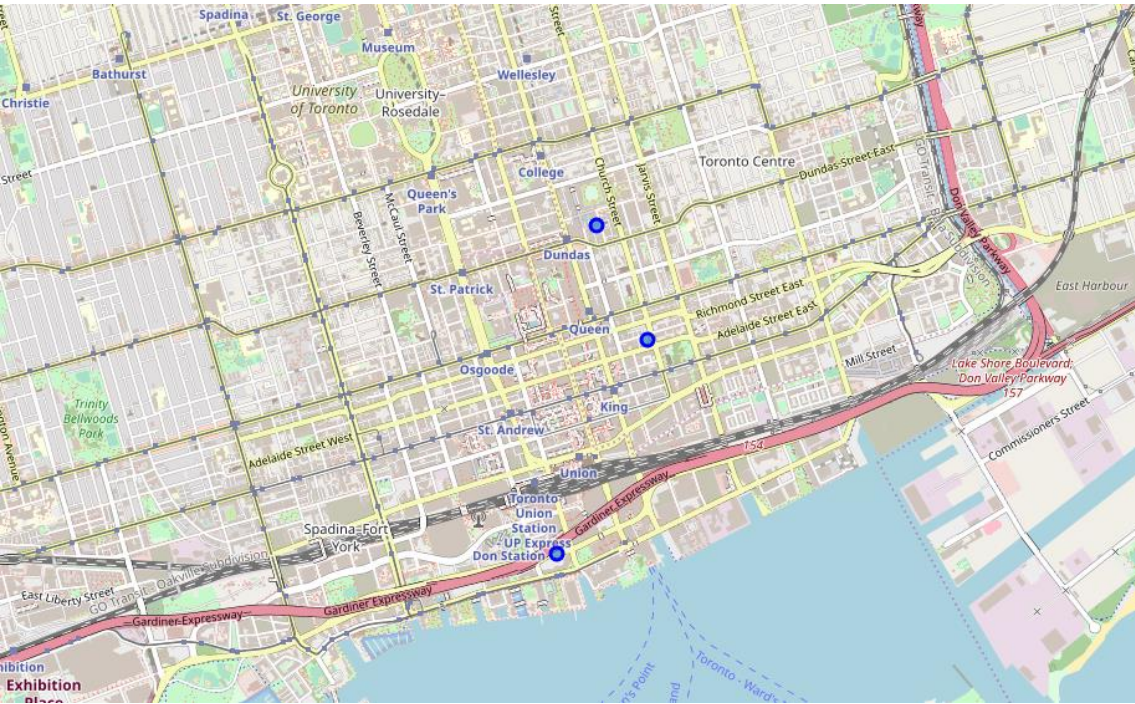Table 10 – df_toronto_indian_demo only 0 cluster neighborhoods top 3.

Image 05 – Geospatial map with there boroughs recommend to Aladdin's.