

# Report for the Business and location selection in Toronto

## 1. Introduction

This report is for the business location selection in Toronto in order to help business to figure out the place that would be the most profitable for each cluster and segment.

## 2. Data

The datasets are collected from the internet and Coursera IBM Certificate course. Both data include the postal code of different location, borough, and latitude and longitude. This would help us into forming cluster and segment and doing the further analysis.

## 3. Methodology

The methodology that I used is based on Clustering and Segmentation. Also, I used folium to create a map around Toronto to do the visualization. Plotting on the map can do the clustering and centering to help us find the location to open the business. Also, by analyzing the characteristics in the specific cluster we can find some patterns such as the popularity, the size of the cluster, population around the same cluster, and it would be helpful in making decision whether we should open a business around there, depends on the kind of the business.

## 4. Results and discussion:

Here are some screenshots that we can look deeper into the data itself.

```

df = df.groupby('Borough').agg({'Postcode':'first',
                                'Neighbourhood': ', '.join
                               }).reset_index()
df

```

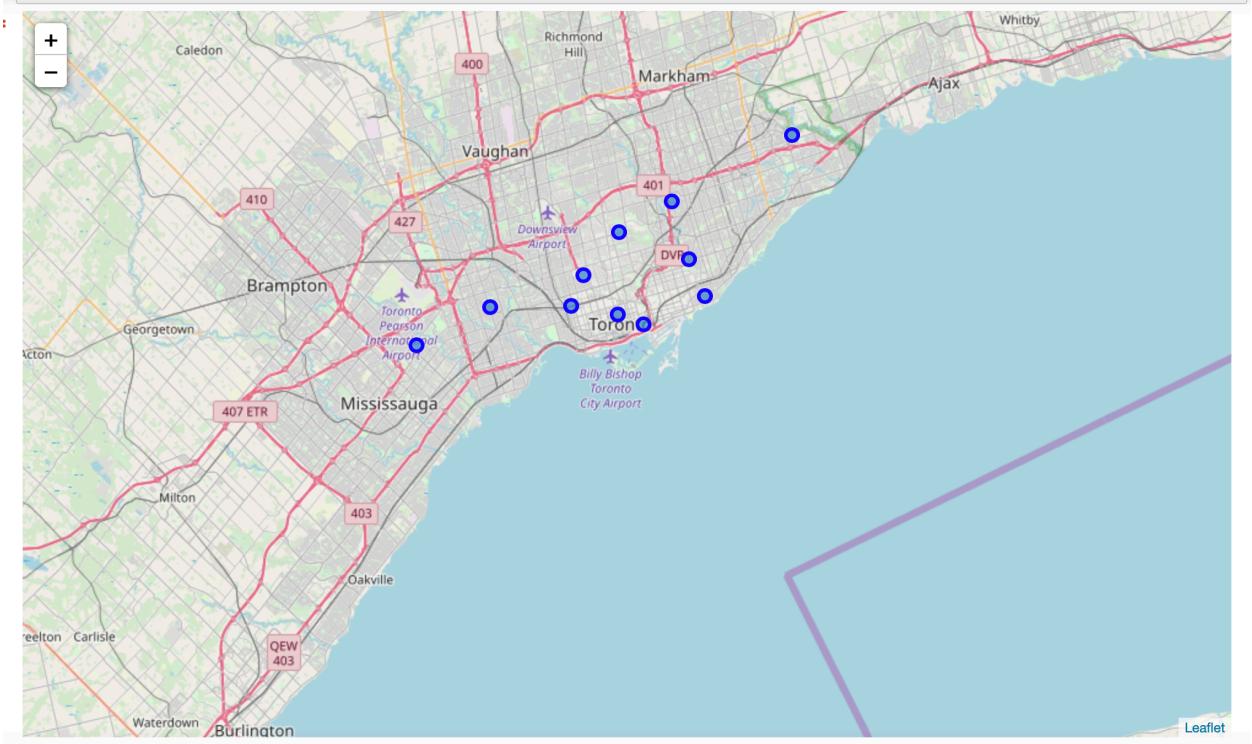
	Borough	Postcode	Neighbourhood
0	Central Toronto	M4N	Lawrence Park, Roselawn, Davisville North, For...
1	Downtown Toronto	M5A	Harbourfront, Regent Park, Ryerson, Garden Dis...
2	East Toronto	M4E	The Beaches, The Danforth West, Riverdale, The...
3	East York	M4B	Woodbine Gardens, Parkview Hill, Woodbine Heig...
4	Etobicoke	M9A	Islington Avenue, Cloverdale, Islington, Marti...
5	Mississauga	M7R	Canada Post Gateway Processing Centre
6	North York	M3A	Parkwoods, Victoria Village, Lawrence Heights,...
7	Queen's Park	M7A	Queen's Park
8	Scarborough	M1B	Rouge, Malvern, Highland Creek, Rouge Hill, Po...
9	West Toronto	M6H	Dovercourt Village, Dufferin, Little Portugal,...
10	York	M6C	Humewood-Cedarvale, Caledonia-Fairbanks, Del R...

First, I got the data clean without any not assigned value for Borough and group the neighborhoods together for the same Borough. As shown in the data, there are at least 3 neighborhoods for each Borough except Queen's Park. This is the first step to figure out the cluster for the different neighborhood.

```
df2
```

	Borough	Postcode	Neighbourhood	Latitude	Longitude
0	Central Toronto	M4N	Lawrence Park, Roselawn, Davisville North, For...	43.728020	-79.388790
1	Downtown Toronto	M5A	Harbourfront, Regent Park, Ryerson, Garden Dis...	43.654260	-79.360636
2	East Toronto	M4E	The Beaches, The Danforth West, Riverdale, The...	43.676357	-79.293031
3	East York	M4B	Woodbine Gardens, Parkview Hill, Woodbine Heig...	43.706397	-79.309937
4	Etobicoke	M9A	Islington Avenue, Cloverdale, Islington, Marti...	43.667856	-79.532242
5	Mississauga	M7R	Canada Post Gateway Processing Centre	43.636966	-79.615819
6	North York	M3A	Parkwoods, Victoria Village, Lawrence Heights,...	43.753259	-79.329656
7	Queen's Park	M7A	Queen's Park	43.662301	-79.389494
8	Scarborough	M1B	Rouge, Malvern, Highland Creek, Rouge Hill, Po...	43.806686	-79.194353
9	West Toronto	M6H	Dovercourt Village, Dufferin, Little Portugal,...	43.669005	-79.442259
10	York	M6C	Humewood-Cedarvale, Caledonia-Fairbanks, Del R...	43.693781	-79.428191

For this part I did a mapping and merging from the two datasets to involve latitude and longitude since I want to do the visualization on the map to see how they are closed to each other. The map of those Boroughs is shown below:



As shown in the graph, there are 11 different Borough on the map, pinned for each blue dot. The next step would be doing the clustering for those dots to some specific areas. We are not caring about the area where it is far away from Toronto such as Mississauga or Markham. Therefore, we need to create a subset of data just focus on the area near Toronto, for this time I chose Downtown Toronto, North York, East York, and Central Toronto. First, I would like to look at the venues for different areas. The result is shown below:

```
tor_grouped = tor_onehot.groupby('Neighborhood').mean().reset_index()
```

Neighborhood	Antique Shop	Art Gallery	Athletics & Sports	Bakery	Bank	Beer Store	Breakfast Spot	Brewery	Bus Line	...	Pharmacy	Pizza Place	Pub	Restaurant
Harbourfront, Regent Park, Ryerson, Garden Dis...	0.019608	0.019608	0.000000	0.058824	0.019608	0.019608	0.039216	0.019608	0.000000	...	0.000000	0.000000	0.058824	0.039216
Humewood-Cedarvale, Caledonia-Fairbanks, Del R...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
Lawrence Park, Roselawn, Davisville North, For...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	...	0.000000	0.000000	0.000000	0.000000
Parkwoods, Victoria Village, Lawrence Heights,...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
Woodbine Gardens, Parkview Hill, Woodbine Heig...	0.000000	0.000000	0.090909	0.000000	0.090909	0.000000	0.000000	0.000000	0.000000	...	0.090909	0.181818	0.000000	0.000000

5 rows × 45 columns

For those values which is relatively high shown above, it means that probably there are more venues type in that neighborhood. For instance, the value for restaurant in the first neighborhood is relatively higher than any others. It means that there might be more competition to open a restaurant in that area. This is a good indicator for the business decision, but it is not necessarily right since we do not have any idea if it is because there are more population, or the size of that neighborhood is much larger than others. To figure out how we should make good business decision based on the data, we need to find more patterns for each neighborhood and how they look similar to each other based on that information. Since we need to know which position or location and what kind of business we shall open around that, we need to find the cluster results. Below is the cluster result:

----Harbourfront, Regent Park, Ryerson, Garden District, St. James Town, Berczy Park, Central Bay Street, Christie, Adelaide, King, Richmond, Harbourfront East, Toronto Islands, Union Station, Design Exchange, Toronto Dominion Centre, Commerce Court, Victoria Hotel, Harbord, University of Toronto, Chinatown, Grange Park, Kensington Market, CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara, Rosedale, Stn A PO Boxes 25 The Esplanade, Cabbagetown, St. James Town, First Canadian Place, Underground city, Church and Wellesley----

	venue	freq
0	Coffee Shop	0.18
1	Café	0.06
2	Bakery	0.06
3	Park	0.06
4	Pub	0.06

----Humewood-Cedarvale, Caledonia-Fairbanks, Del Ray, Keelesdale, Mount Dennis, Silverthorn, The Junction North, Runnymede, Weston----

	venue	freq
0	Trail	0.25
1	Hockey Arena	0.25
2	Field	0.25
3	Moving Target	0.25
4	Antique Shop	0.00

----Lawrence Park, Roselawn, Davisville North, Forest Hill North, Forest Hill West, North Toronto West, The Annex, North Midtown, Yorkville, Davisville, Moore Park, Summerhill East, Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West----

	venue	freq
0	Swim School	0.33
1	Bus Line	0.33
2	Park	0.33
3	Antique Shop	0.00
4	Pet Store	0.00

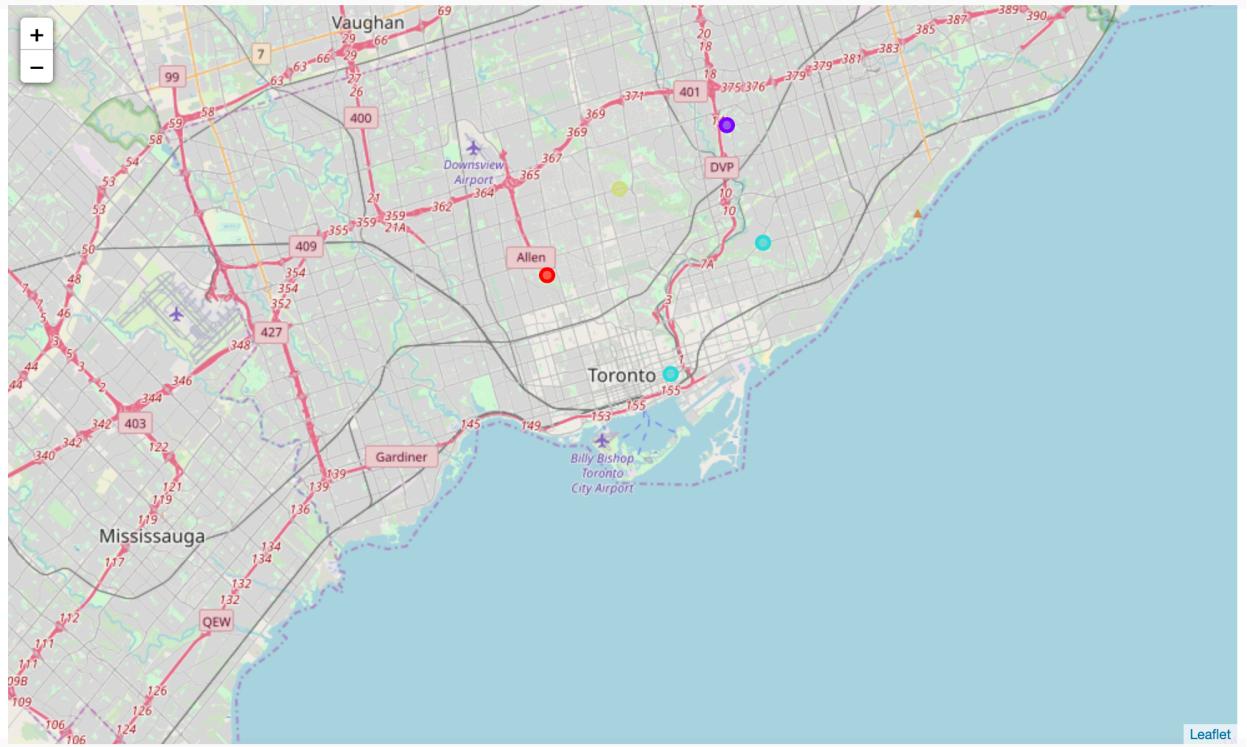
As shown above, it gives some business insight by ranking different venues based on the frequency.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Harbourfront, Regent Park, Ryerson, Garden Dis...	Coffee Shop	Café	Pub
1	Humewood-Cedarvale, Caledonia-Fairbanks, Del R...	Trail	Field	Hockey Arena
2	Lawrence Park, Roselawn, Davisville North, For...	Swim School	Bus Line	Park
3	Parkwoods, Victoria Village, Lawrence Heights,...	Food & Drink Shop	Fast Food Restaurant	Park
4	Woodbine Gardens, Parkview Hill, Woodbine Heig...	Fast Food Restaurant	Pizza Place	Gym / Fitness Center

For each neighborhood around Toronto, there is a brief summary of the 3 most common venues. Based on the data, we can get to know the preference of customer in that area.

For example, in neighborhood Lawrence Park, there are swim school, bus line, park. It means that this part has more space and maybe some sports related business can be developed well in this neighborhood.

Here we want to know the relation between different neighbor. Here is the visualization of the after clustering process.



As shown above, the downtown Toronto and East York area are in the same cluster, meaning that the characteristic for these areas are similar. Therefore, if we want to open a business around downtown Toronto, it is recommended that we choose some point in the central part of this cluster. It can be more beneficial because the rent can be lower if we open far away from Toronto, and it would not affect a lot since they are in the same cluster.

## 5. Conclusion:

1	Downtown Toronto	M5A	Harbourfront, Regent Park, Ryerson, Garden Dis...	43.654260	-79.360636	2	Coffee Shop	Café	Pub
2	East York	M4B	Woodbine Gardens, Parkview Hill, Woodbine Heig...	43.706397	-79.309937	2	Fast Food Restaurant	Pizza Place	Gym / Fitness Center

For this specific situation, if we focus on open a business around Toronto area, it is not recommended to open a coffee shop or fast food restaurant since there is huge competition around that neighbor. Since there are more population around that area, we can open some entertain business such as some bars, golfing center or something related to them. Also, the location of the business can be in the middle part of the Downtown Toronto and East York. We can find some cheaper place to rent to lower the

cost without have much decrease in the revenue, since the characteristics of two neighbors looks similar and they are in the same cluster.