

Toronto Neighborhood Clustering



CONTENT

CONTENT

Introduction

PART ONE

Data Processing

PART TWO

Result

PART THREE

Conclusion

PART FOUR



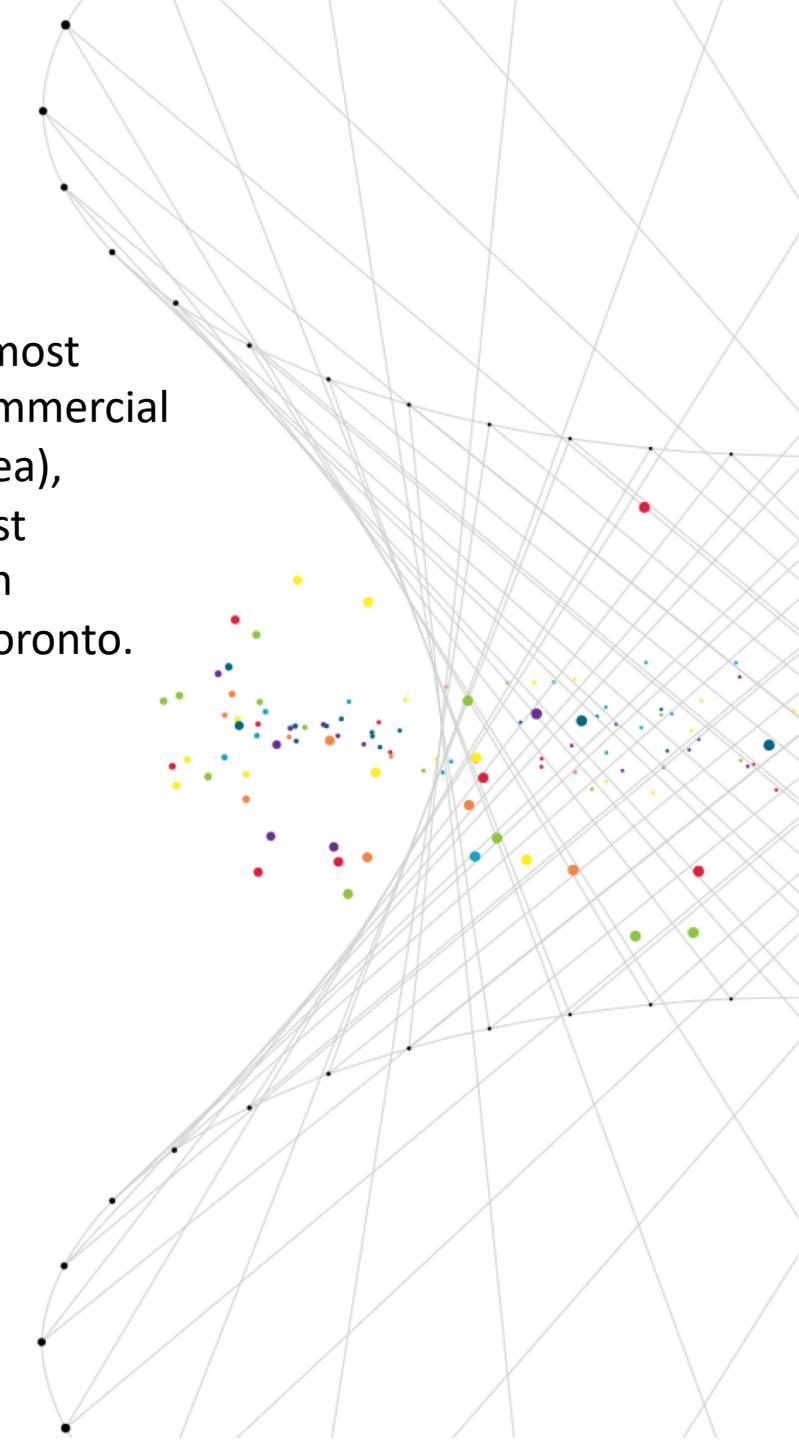


Introduction

PART ONE

Toronto

Toronto, city, capital of the province of Ontario, southeastern Canada. It is the most populous city in Canada, a multicultural city, and the country's financial and commercial center. Also, with many high-level universities located in GTA (Great Toronto Area), Toronto is also a technology center in North America. Most of the world's largest companies have branches in Toronto. With the fast development of Information Technology, more and more new graduated students are planning to settle in Toronto.



Business Problem

Jack is currently living in North York district in Toronto, with a postal code start from 'M2M'. Now he finds a new job and will be working in the Downtown area. Since the TTC of Toronto is quite slow and unreliable, he doesn't want to waste 2 hours on traveling every day. So he is planning to move to Downtown area.



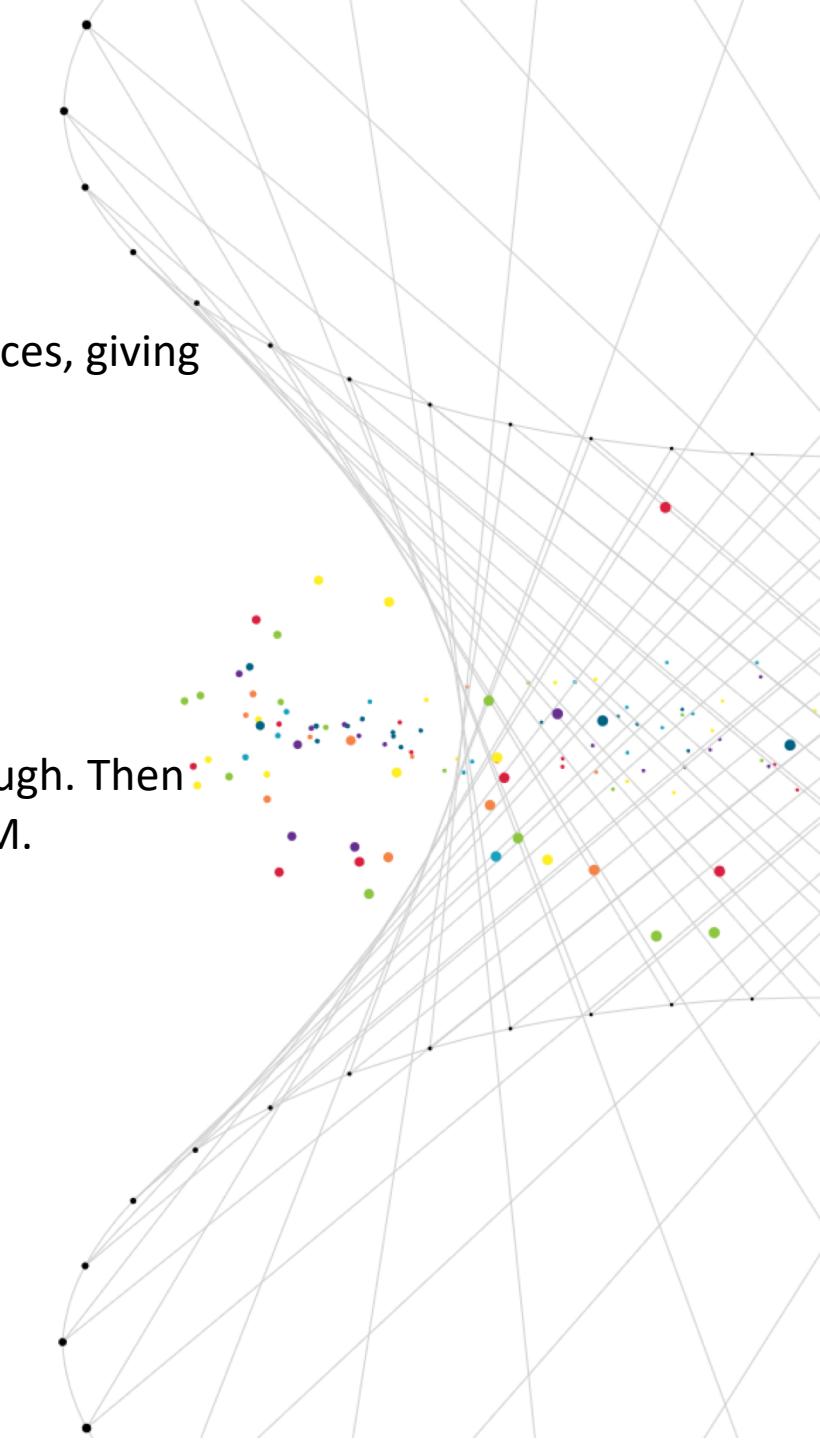
Data Description

The data for this project has been retrieved and processed from different sources, giving careful consideration to the accuracy of the methods used

The data consists of three parts:

- Borough data with postal code
- Geocoding data
- Venue data with borough details

By joining the first two dataset, we are able to find the centroids of each borough. Then use the data of venue to cluster the boroughs and find the similarity with M2M.





Data Processing

PART TWO

Data Importing

Import Neighborhood Data

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

Import Geocoding Data

	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

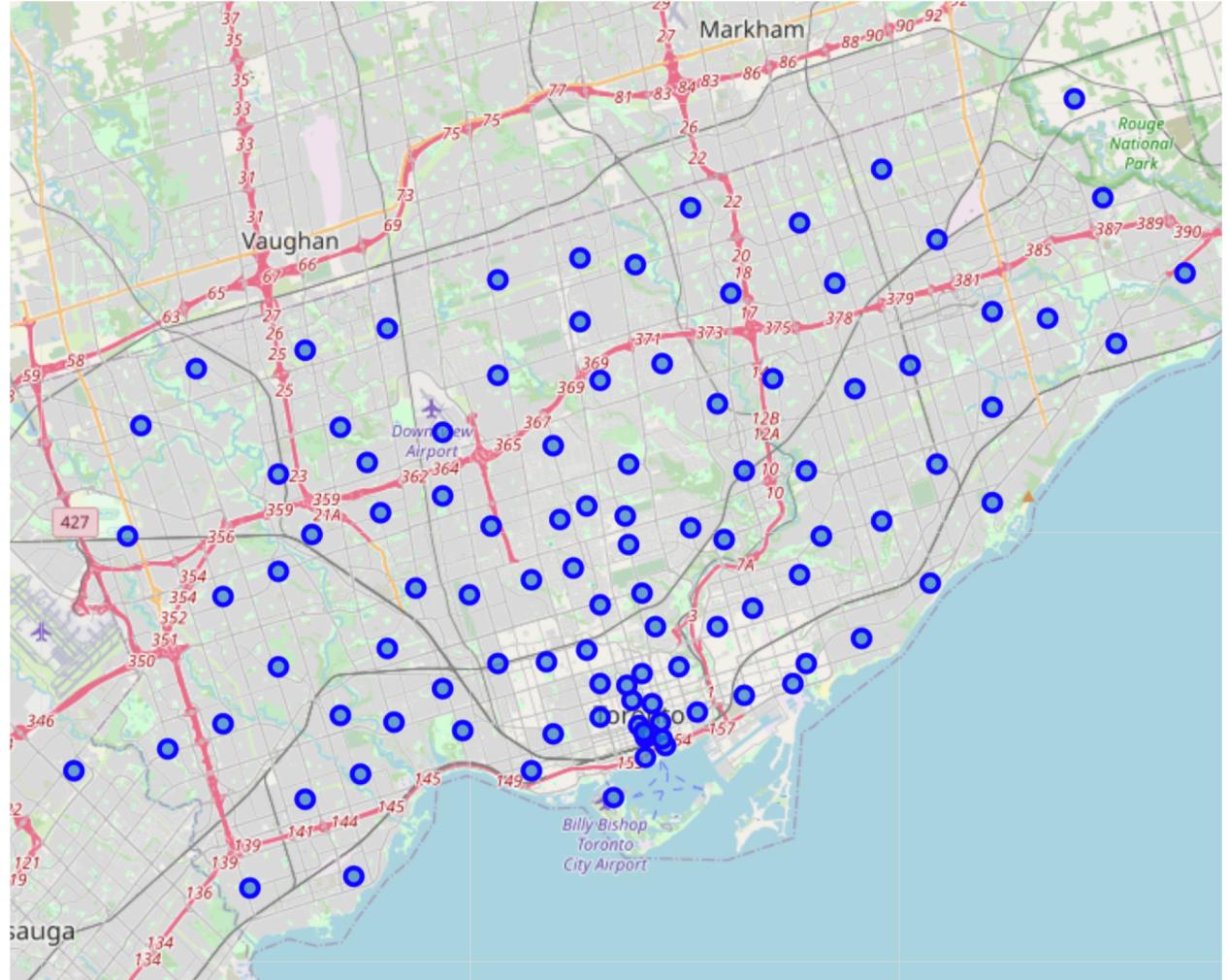
Data Merging and Cleaning

	Postal Code	Borough	Neighborhood	Latitude	Longitude
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
18	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
22	M1G	Scarborough	Woburn	43.770992	-79.216917
26	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Data Visualization

From the right graph.

- We see all the neighborhoods are in the Toronto area (Not include neighborhoods in nearby cities like Missisauga, Richmond Hill, Pickering etc). We don't need to do any other data filtering on the current data



Get Venues Data

To analyze the similarity of each neighborhood, we need to find the venues data in each neighborhood. For example, how many café shops are there in a neighborhood? How many categories of venues are there in a neighborhood? How many total venues are there in a neighborhood. To download the data, we use the API call from FourSquare.

	name	categories	lat	lng
0	B.B. Cafe	Café	43.791117	-79.418078
1	Starbucks	Coffee Shop	43.796409	-79.419653
2	Eat Bkk Thai Kitchen & Bar	Thai Restaurant	43.795696	-79.419205
3	Daldongnae	Korean Restaurant	43.789729	-79.418104
4	Galleria Supermarket	Supermarket	43.799003	-79.420931

Join venues data with neighborhood data

	PostalCode	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M1B	43.806686	-79.194353	African Rainforest Pavilion	43.817725	-79.183433	Zoo Exhibit
1	M1B	43.806686	-79.194353	Toronto Pan Am Sports Centre	43.790623	-79.193869	Athletics & Sports
2	M1B	43.806686	-79.194353	Polar Bear Exhibit	43.823372	-79.185145	Zoo
3	M1B	43.806686	-79.194353	Toronto Zoo	43.820582	-79.181551	Zoo
4	M1B	43.806686	-79.194353	Orangutan Exhibit	43.818413	-79.182548	Zoo Exhibit

Get Venues Data

By calculating the total number of venues of each neighborhood, we can get an overview of the business development of each neighborhood.

Neighborhood	Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
M1B		72		72	72	72
M1C		58		58	58	58
M1E		59		59	59	59
M1G		94		94	94	94
M1H		100		100	100	100
M1J		94		94	94	94
M1K		85		85	85	85

By calculating the mean value of venues group by neighborhoods, we can find the appearing frequency of venues, based on which we can perform the clustering model to make clusters of the neighborhoods.

	PostalCode	Accessories Store	Afghan Restaurant	Airport	American Restaurant	Amphitheater	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	...	Vietnamese Restaurant	Warehouse Store	Whisky Bar	Wine Bar	Wings Joint	Women's Store	Xinjiang Restaurant	Yoga Studio	Zoo	Zoo Exhibit
0	M1B	0.000000	0.00	0.0	0.000000	0.00	0.00	0.00	0.00	0.000000	...	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.027778	0.222222
1	M1C	0.000000	0.00	0.0	0.000000	0.00	0.00	0.00	0.00	0.000000	...	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
2	M1E	0.000000	0.00	0.0	0.000000	0.00	0.00	0.00	0.00	0.000000	...	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
3	M1G	0.000000	0.00	0.0	0.010638	0.00	0.00	0.00	0.00	0.000000	...	0.010638	0.000000	0.00	0.00	0.000000	0.00	0.010638	0.000000	0.000000	0.000000
4	M1H	0.000000	0.00	0.0	0.020000	0.00	0.00	0.00	0.00	0.000000	...	0.010000	0.000000	0.00	0.00	0.020000	0.00	0.010000	0.010000	0.000000	0.000000

Get Venues Data

Since we have 103 neighborhoods to cluster, we select 15 as the number of clusters. The method we use is KMeans Clustering. Based on the appearing frequency of each kind of venues, the model could calculate the distance by using the frequency. The frequency calculation process actually already finish the standard scaling process so we can directly perform the model.

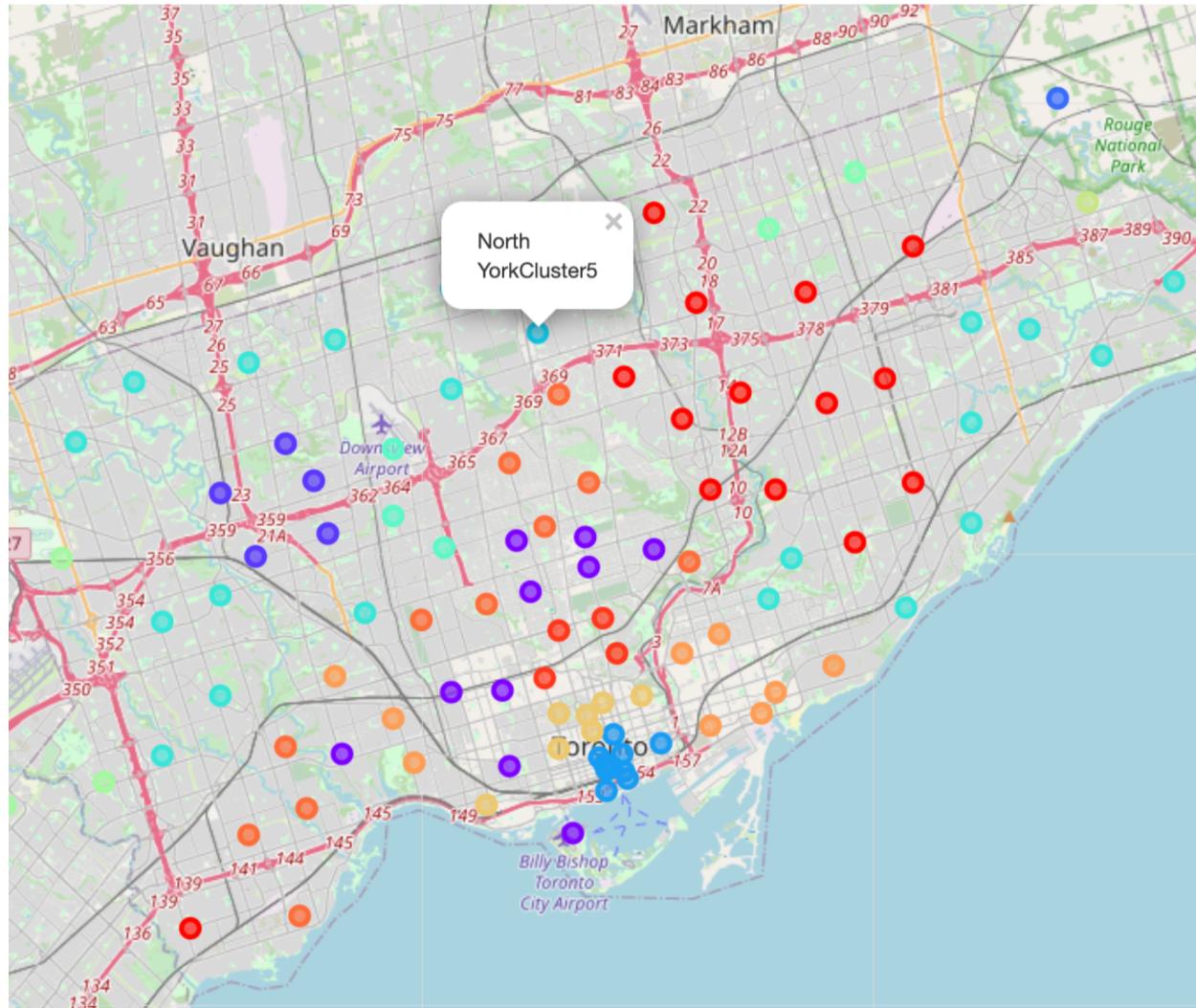
After modeling, each neighborhood is assigned a label, which is the cluster it belongs to. Then we can attach the label column into the result table, which shows the neighborhood details, labels and top 10 popular venues.

PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Label	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
M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	10	Zoo Exhibit	Fast Food Restaurant	Park	Pizza Place	Grocery Store	Gas Station	Restaurant	Pharmacy	Zoo	Burger Joint
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	6	Coffee Shop	Park	Breakfast Spot	Sandwich Place	Pet Store	Grocery Store	Bank	Burger Joint	Liquor Store	Fast Food Restaurant
M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	6	Pizza Place	Coffee Shop	Grocery Store	Hotel	Park	Sandwich Place	Bank	Juice Bar	Supermarket	Discount Store
M1G	Scarborough	Woburn	43.770992	-79.216917	6	Coffee Shop	Fast Food Restaurant	Bank	Pizza Place	Gas Station	Sandwich Place	Chinese Restaurant	Beer Store	Discount Store	Indian Restaurant
M1H	Scarborough	Cedarae	43.773136	-79.239476	6	Coffee Shop	Sandwich Place	Gas Station	Pharmacy	Clothing Store	Bank	Gym	Indian Restaurant	Pizza Place	Caribbean Restaurant



Result

PART THREE



Now we find the clusters distribution of each neighborhood. The one marked on map is the neighborhood Jack currently living in, and the neighborhoods marked deep blue are the ones he plans to move in

From the precious graph, we can clearly find that the label of neighborhood ‘M2M’ has no similar one in downtown area. To better evaluate the influence of difference, we find the details of neighborhood ‘M2M’ and cluster in downtown.

PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Label	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
52	M2M	North York	Willowdale, Newtonbrook	43.789053 -79.408493	5	Korean Restaurant	Coffee Shop	Café	Bubble Tea Shop	Middle Eastern Restaurant	Ramen Restaurant	Japanese Restaurant	Pizza Place	Seafood Restaurant	Dessert Shop
PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Label	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
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260 -79.360636	4	Coffee Shop	Park	Restaurant	Café	Neighborhood	Plaza	Ice Cream Shop	Mediterranean Restaurant	Gym / Fitness Center	Bakery
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162 -79.378937	4	Coffee Shop	Park	Restaurant	Yoga Studio	Café	Plaza	Pizza Place	Japanese Restaurant	Sporting Goods Shop	Sandwich Place
15	M5C	Downtown Toronto	St. James Town	43.651494 -79.375418	4	Coffee Shop	Café	Japanese Restaurant	Park	Plaza	Italian Restaurant	Thai Restaurant	Restaurant	Supermarket	Liquor Store
20	M5E	Downtown Toronto	Berczy Park	43.644771 -79.373306	4	Coffee Shop	Café	Plaza	Dessert Shop	Park	Italian Restaurant	Bakery	Liquor Store	Baseball Stadium	Neighborhood
30	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571 -79.384568	4	Coffee Shop	Café	Italian Restaurant	Yoga Studio	Japanese Restaurant	Plaza	Sandwich Place	Park	Pizza Place	Thai Restaurant
36	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816 -79.381752	4	Coffee Shop	Park	Café	Plaza	Hotel	Scenic Lookout	Art Gallery	Theater	Gym	Yoga Studio
42	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177 -79.381576	4	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Yoga Studio	Park	Plaza	Japanese Restaurant	Baseball Stadium	Hotel
48	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198 -79.379817	4	Coffee Shop	Café	Park	Sandwich Place	Plaza	Italian Restaurant	Yoga Studio	Cosmetics Shop	Sporting Goods Shop	Japanese Restaurant
92	M5W	Downtown Toronto	Stn A PO Boxes	43.646435 -79.374846	4	Coffee Shop	Café	Plaza	Dessert Shop	Park	Thai Restaurant	Italian Restaurant	Farmers Market	Baseball Stadium	Neighborhood
97	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429 -79.382280	4	Coffee Shop	Italian Restaurant	Café	Yoga Studio	Sandwich Place	Plaza	Park	Theater	Restaurant	Japanese Restaurant

From the above tables, it is clear to see the overall popular venues of the two different areas are quite similar. The difference is caused by the number of venues, downtown area all have much more venues than it in North York. Which result in the different clusters of neighborhoods. Since in FourSquare it will only return 100 results for all venues, we can not get more detailed data to perform a more accurate model.

The company location that Jack is going to work in is (43.645832, -79.383097). By calculating the distance from the centroid of each neighborhood, we can get an understanding of how far each neighborhood is to the company.

PostalCode	Borough	Neighborhood	Distance
M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	0.193271
M5X	Downtown Toronto	First Canadian Place, Underground city	0.296272
M5L	Downtown Toronto	Commerce Court, Victoria Hotel	0.372808
M5H	Downtown Toronto	Richmond, Adelaide, King	0.540263
M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	0.568363
M5W	Downtown Toronto	Stn A PO Boxes	0.667485
M5E	Downtown Toronto	Berczy Park	0.796825
M5C	Downtown Toronto	St. James Town	0.882381
M5B	Downtown Toronto	Garden District, Ryerson	1.303924
M5A	Downtown Toronto	Regent Park, Harbourfront	2.036333

By referring the TTC public transportation lines distribution, we find the transportation is way better than it in North York. Since every ride of TTC cost 3 CAD, it could save a lot for Jack if he decides to walk to company. From the table above, we can find the top 5 neighborhoods are not farer than 600 meters to the company, which is an appropriate distance for walking. Jack could select his new room in these neighborhoods: **M5K, M5X, M5L, M5H**.



Conclusion

PART FOUR

- Increase the venue limits by using premium account, and get more venues that could better cluster the neighborhoods.
- Get room for rental data, which could reflect how many available rooms are there in a neighborhood. Even though some neighborhoods have both great distance, traffic conditions, and public services, the neighborhood may not be designed for living and don't have many available rooms.



- Get public transportation data, like station location, station count in a neighborhood, total traffic volume.
- Get room price data. Room price is a very important point when people is looking for a room for rental. By analyzing this feature, we are able to find the most suitable neighborhood based on the budget.

The background features a large, light gray triangular grid centered on the slide. Scattered throughout the grid are numerous small, semi-transparent colored dots in various colors like green, yellow, red, blue, and purple.

THANK YOU FOR WATCHING