A**. Description & Disscusion of the Background**

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 as of 2016.Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the

most multicultural and cosmopolitan cities in the world (source: wikipedia). These reasons above are my motivation to do this study for helping investors if they decide to invest in Toronto.

We can create a map where the real estate index is placed on Toronto and each Neighborhood is clustered according to the venue density.

Data Description

To consider the problem we can list the datas as below:

I build a code to scrape the following

link, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M in order to obtain the data that is in the table of postal codes and to transform the data into a pandas dataframe, also i use the geospatial data (here is the link: http://cocl.us/Geospatial_data). I join the both data to get a unique data frame cleaned reduced it to necessaries columns that i used to create a map of Toronto with superimposed neighborhood. I used Forsquare API to get the most common venues of given Borough of Toronto, and cluster it.

B. Methodology

As a database, I used GitHub repository in my study. My master data which has the main components *PostalCode, Borough, Neighborhood , Latitude and Longitude.*

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek	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
			Kennedy Park / Ionview /		

used python **folium** library to visualize geographic details of Toronto and its boroughs and I created a map of Toronto with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:

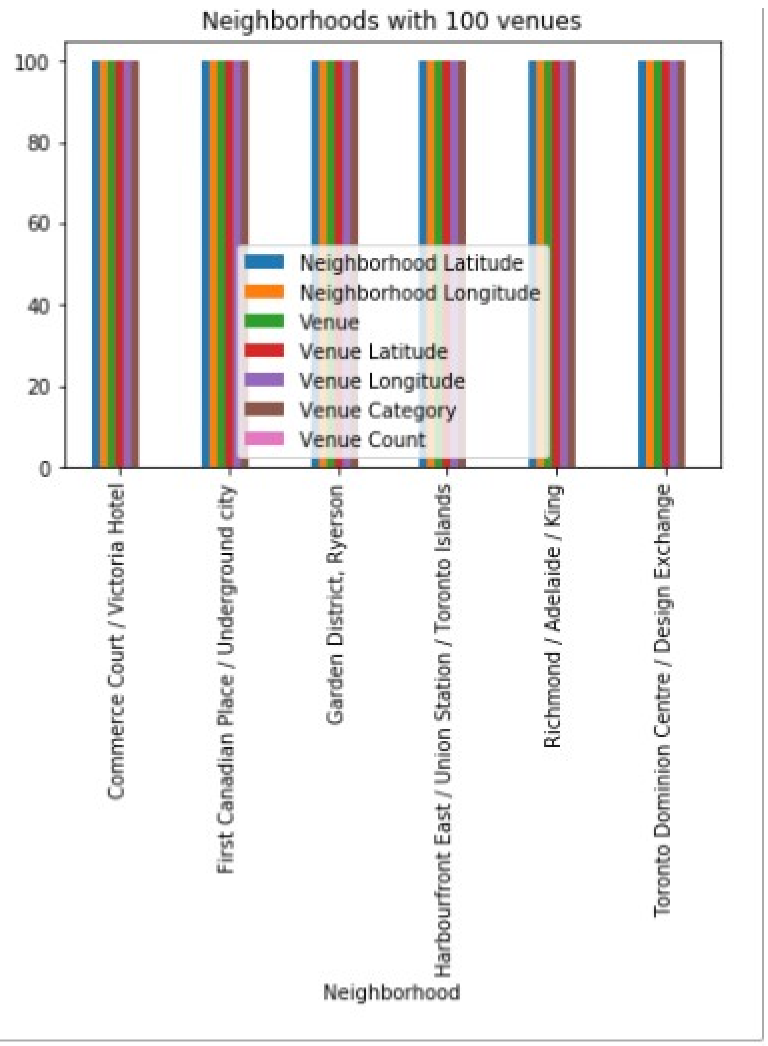
I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as **100 venue** and the radius **500 meter** for each borough from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API.

	name	categories	lat	lng
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Cooper Koo Family YMCA	Distribution Center	43.653249	-79.358008
3	Body Blitz Spa East	Spa	43.654735	-79.359874
4	Morning Glory Cafe	Breakfast Spot	43.653947	-79.361149

In summary of this data **2201** venues were returned by Foursquare. Here is a merged table of boroughs and venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

We can see that Commerce Court Victoria Hotel, First Canadian Place/Underground City, Garden District, Ryerson, Harbourfront East /Union Station / Toronto Islands, Richmond /Adelaide /King, Toronto Dominion Centre / Design Exchange have the top venues



In summary of this graph **256** unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

Out[40]:

	Borough	Cluster Labels	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
1	North York	1.0	Hockey Arena	Intersection	Pizza Place	Coffee Shop	Portuguese Restaurant	Donut Shop	Discount Store
2	Downtown Toronto	1.0	Coffee Shop	Pub	Bakery	Park	Mexican Restaurant	Breakfast Spot	Restauran
3	North York	1.0	Furniture / Home Store	Accessories Store	Clothing Store	Coffee Shop	Event Space	Miscellaneous Shop	Athletics & Sports
4	Downtown Toronto	1.0	Coffee Shop	Yoga Studio	Restaurant	Bar	Beer Bar	Italian Restaurant	Juice Bar
6	Scarborough	1.0	Fast Food Restaurant	Diner	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Empanada Restauran
7	North York	1.0	Gym	Coffee Shop	Restaurant	Beer Store	Japanese Restaurant	Supermarket	Clothing S
8	East York	1.0	Pizza Place	Gym / Fitness Center	Pharmacy	Intersection	Athletics & Sports	Gastropub	Bus Line
9	Downtown Toronto	1.0	Coffee Shop	Clothing Store	Middle Eastern Restaurant	Bubble Tea Shop	Japanese Restaurant	Cosmetics Shop	Café
10	North York	1.0	Park	Pizza Place	Pub	Japanese Restaurant	Doner Restaurant	Diner	Discount Store

We have some common venue categories in boroughs. In this reason I used unsupervised learning **K-means algorithm** to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning. First, I will run K-Means to cluster the boroughs into **5** clusters . Here is my merged table with cluster labels for each borough.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th M Comn Venu
0	M3A	North York	Parkwoods	43.753259	-79.329656	0.0	Park	Food & Drink Shop	Falafel Restaurant	Event Space	Ethio Resta
1	M4A	North York	Victoria Village	43.725882	-79.315572	1.0	Hockey Arena	Intersection	Pizza Place	Coffee Shop	Portu Resta
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1.0	Coffee Shop	Pub	Bakery	Park	Mexic Resta
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Furniture / Home Store	Accessories Store	Clothing Store	Coffee Shop	Event Space
4	M7A	Downtown	Queen's Park, Ontario	43.662301	-79.389494	1.0	Coffee	Yoga Studio	Restaurant	Bar	Beer I

1-The first cluster markers is colored with red in the map and they

moslty consist of parks.

2- The second cluster appears to be mostly consist of coffee shops and cafes and it's colored with purple.

3- The third cluster only shows one row and there are mostly reataurants, and colored with blue.

4- The fourth one has only two rows of data and it seems that playgrounds are popular in that area, and it's colored with mint green.

5- The fifth cluster which is also the last one has only two rows of data as well and it can be seen that there are sports or activity places at most and it's colored with orange color.

C. Results

I examine each cluster;

I created map which also has the below informations for each borough:

- Borough name,
- Cluster name,
- Top 10 number of venue

D. Discussion

As I mentioned before, Toronto is a big city with a high population density. Different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropol.

I used the Kmeans algorithm as part of this clustering study. I fixed the number of clusters to 5. I also performed data analysis through this information by adding the python notebook file on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

I ended the study by visualizing the data and clustering information on the Toronto map. In future studies, web or telephone applications can be carried out to direct investors.

F. Conclusion

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.