Analyzing neighborhoods of Toronto city

Introduction

Introduction

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 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.

Toronto is a prominent centre for music, theatre,motion picture production,and television production,and is home to the headquarters of Canada's major national broadcast networks and media outlets.Its varied cultural institutions,which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities,attract over 43 million tourists each year.

As you can see from the figures, Toronto is a city with a high population and population density. Being such a crowded city leads the owners of shops and social sharing places in the city where the population is dense. When we think of it by the investor, we expect from them to prefer the areas where a particular business has less competition and the type of business they want to install is less intense. If we think of the city residents, they may want to choose the district according to the social place's density. However, it is difficult to obtain information that will guide investors in this direction, nowadays. Hence, our main aim in this project has been to analyze the venues on each neighborhood which can help both residents while choosing their housing based on housing or investors while choosing where and which company/firm to invest on.

Data description:

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

I scrapped the table of postal codes of Canada from the Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,) and transformed the data into a *pandas* dataframe. The dataframe consists of three columns: PostalCode, Borough, and Neighborhood. Then I cleaned and wrangled the data to remove the cells with a borough that is **Not assigned** and where more than one neighborhood exists in one postal code area in separate rows are combined into a single row separated by commas. Secondly,If a cell has a borough but a **Not assigned** neighborhood, then the neighborhood will be the same as the borough. The final dataframe looks like this:

	Postal Code	Neighbourhood	Borough
0	M1B	Malvern, Rouge	Scarborough
1	M1C	Port Union, Highland Creek, Rouge Hill	Scarborough
2	M1E	West Hill, Guildwood, Morningside	Scarborough
3	M1G	Woburn	Scarborough
4	M1H	Cedarbrae	Scarborough
...	...	...	...
98	M9N	Weston	York
99	M9P	Westmount	Etobicoke
100	M9R	Martin Grove Gardens, Richview Gardens, Kingsv...	Etobicoke
101	M9V	Humbergate, Mount Olive, Silverstone, Jamestow...	Etobicoke
102	M9W	Northwest	Etobicoke

There are total 103 neighborhood areas with unique postal codes in entire Canada.

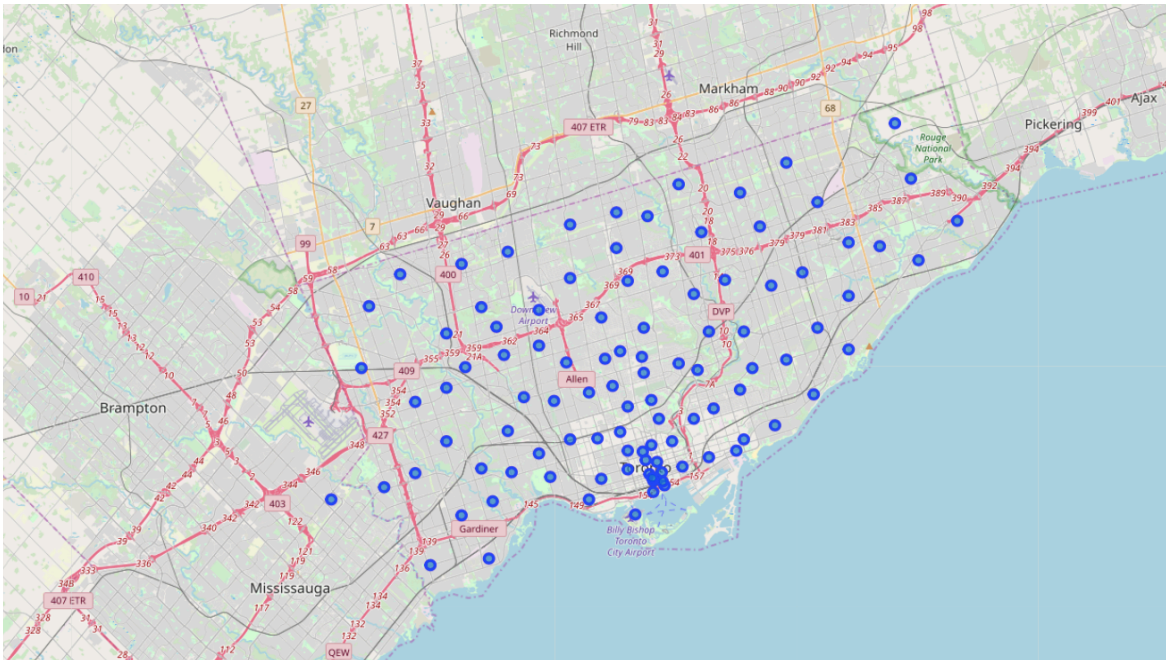
From the link to a csv file: http://cocl.us/Geospatial_data, I downloaded the dataset to obtain the geographical coordinates of each postal code. I utilized the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood of Toronto an merged the data. Then I cleaned the data to include only those Boroughs which has Toronto city and reduced the dataset to work only on neighborhoods of Toronto city.

Method:

As a database, I used GitHub repository in my study. My master data which has the main components Postal codes, *Neighborhood, Borough, Latitude* and *Longitude* informations of the city. There were 87 Neighborhood areas with unique postal codes in Toronto city

	Postal Code	Neighbourhood	Borough	Latitude	Longitude
	output; double click to hide	The Beaches	East Toronto	43.676357	-79.293031
41	M4K	Riverdale, The Danforth West	East Toronto	43.679557	-79.352188
42	M4L	The Beaches West, India Bazaar	East Toronto	43.668999	-79.315572
43	M4M	Studio District	East Toronto	43.659526	-79.340923
44	M4N	Lawrence Park	Central Toronto	43.728020	-79.388790
45	M4P	Davisville North	Central Toronto	43.712751	-79.390197
46	M4R	North Toronto West	Central Toronto	43.715383	-79.405678
47	M4S	Davisville	Central Toronto	43.704324	-79.388790
48	M4T	Summerhill East, Moore Park	Central Toronto	43.689574	-79.383160
49	M4V	Summerhill West, South Hill, Rathnelly, Forest...	Central Toronto	43.686412	-79.400049
50	M4W	Rosedale	Downtown Toronto	43.679563	-79.377529
51	M4X	St. James Town, Cabbagetown	Downtown Toronto	43.667967	-79.367675
52	M4Y	Church and Wellesley	Downtown Toronto	43.665860	-79.383160
53	M5A	Harbourfront, Regent Park	Downtown Toronto	43.654260	-79.360636
54	M5B	Ryerson, Garden District	Downtown Toronto	43.657162	-79.378937
55	M5C	St. James Town	Downtown Toronto	43.651494	-79.375418
56	M5E	Berczy Park	Downtown Toronto	43.644771	-79.373306

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



I utilized the Foursquare API to explore the boroughs neighborhoods 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.

In summary of this data, **38** unique venues were returned by Foursquare. Here is a merged table of Neighborhoods and venues. I created a table which shows list of top 10 venue category for each borough in below table.

	Neighbourhood	1st Most Common Venues	2nd Most Common Venues	3rd Most Common Venues	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
0	Berczy Park	Farmers Market	Beer Bar	Cocktail Bar	Seafood Restaurant	Breakfast Spot	Basketball Stadium	Italian Restaurant	Bistro	Fish Market	Clothing Store
1	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Restaurant	Burrito Place	Brewery	Fast Food Restaurant	Spa	Auto Workshop	Farmers Market	Pizza Place
2	Central Bay Street	Coffee Shop	Italian Restaurant	Spa	Café	Bubble Tea Shop	Ramen Restaurant	Japanese Restaurant	Seafood Restaurant	Sandwich Place	Portuguese Restaurant
3	Chinatown, Kensington Market, Grange Park	Café	Comfort Food Restaurant	Vietnamese Restaurant	Caribbean Restaurant	Mexican Restaurant	Coffee Shop	Snack Place	Cheese Shop	Cocktail Bar	Organic Grocery
4	Christie	Grocery Store	Café	Park	Baby Store	Coffee Shop	Convenience Store	Nightclub	Athletics & Sports	Diner	Restaurant

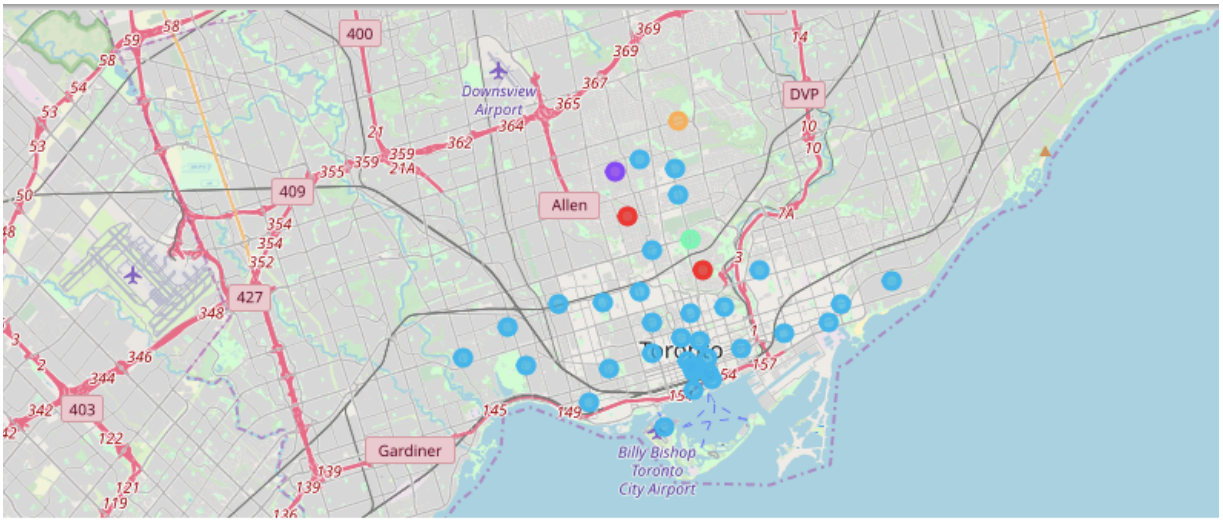
Clustering analysis

In this, I used unsupervised learning K-means algorithm to cluster the neighborhood based on venue categories. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I ran K-Means to cluster the boroughs into 5 clusters. The neighborhood clusters with labels are shown in the table below:

Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venues	Most Common Venues	3rd Most Common Venues	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
East Toronto	43.676357	-79.293031	2	Other Great Outdoors	Health Food Store	Trail	Pub	Neighborhood	Yoga Studio	Dance Studio	Event Space	Ethiopian Restaurant	Eastern European Restaurant
East Toronto	43.679557	-79.352188	2	Greek Restaurant	Ice Cream Shop	Italian Restaurant	Cosmetics Shop	Brewery	Bubble Tea Shop	Restaurant	Pub	Pizza Place	Coffee Shop
East Toronto	43.668999	-79.315572	2	Park	Brewery	Sandwich Place	Burger Joint	Burrito Place	Fast Food Restaurant	Fish & Chips Shop	Italian Restaurant	Steakhouse	Sushi Restaurant
East Toronto	43.659526	-79.340923	2	Café	Coffee Shop	Italian Restaurant	Bakery	Latin American Restaurant	Bar	Fish Market	Neighborhood	Seafood Restaurant	Sandwich Place
Central Toronto	43.728020	-79.388790	4	Park	Swim School	Bus Line	Fish & Chips Shop	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Eastern European Restaurant	Dog Run

Fig 2. Clustered neighborhood



Results

From the clusters, based on venues in each locality, we analyzed the neighborhood and tried to interpret which new firm could be a better investment in the neighboring areas of Toronto and which locality was best as residential area.

There are 191 different venues in the city of Toronto among which the top 5 venues as given by Foursquare API based on customer ratings and customer visits are recorded as:

```
----Victoria Hotel, Commerce Court-
                              venue
0                       Yoga Studio
1                          Wine Bar
2            Vietnamese Restaurant
3                 Video Game Store
4  Vegetarian / Vegan Restaurant
```

The top venues or invested firms around the neighborhood of Toronto area are as follows:

	Count
Café	61
Coffee Shop	55
Restaurant	26
Park	25
Italian Restaurant	25
...	...
Flea Market	1
Aquarium	1
Eastern European Restaurant	1
Market	1
Theme Restaurant	1

Based on the created clusters, It was observed that the cluster with label 2 had the maximum number of neighborhoods within their cluster.

The 5 clusters with labels 0,1,2,3 and 4 and corresponding venues are given below:

Cluster 1. Label 0: Venue in common- Park, Trail . No. of neighborhoods = 2

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 0, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.sha
```

	Neighbourhood	Cluster Labels	1st Most Common Venues	2nd Most Common Venues	3rd Most Common Venues	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
10	Rosedale	0	Park	Playground	Trail	Building	Deli / Bodega	Falafel Restaurant	Event Space	Ethiopian Restaurant	Eastern European Restaurant	Dog Run
23	Forest Hill West, Forest Hill North	0	Bus Line	Park	Trail	Jewelry Store	Sushi Restaurant	Yoga Studio	Falafel Restaurant	Event Space	Ethiopian Restaurant	Eastern European Restaurant

Cluster 2. Label 1: Venue in common-Garden . No. of neighborhoods = 1

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 1, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.sha
```

	Neighbourhood	Cluster Labels	1st Most Common Venues	2nd Most Common Venues	3rd Most Common Venues	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
22	Roselawn	1	Garden	Home Service	Deli / Bodega	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Eastern European Restaurant	Dog Run	Discount Store

Cluster 3. Label 2: Venue in common- Cafe, Coffee shop. No. of neighborhoods = 37

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 2, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.sha
```

	Neighbourhood	Cluster Labels	1st Most Common Venues	2nd Most Common Venues	3rd Most Common Venues	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
0	The Beaches	2	Other Great Outdoors	Health Food Store	Trail	Pub	Neighborhood	Yoga Studio	Dance Studio	Event Space	Ethiopian Restaurant	Eastern European Restaurant
1	Riverdale, The Danforth West	2	Greek Restaurant	Ice Cream Shop	Italian Restaurant	Cosmetics Shop	Brewery	Bubble Tea Shop	Restaurant	Pub	Pizza Place	Coffee Shop
2	The Beaches West, India Bazaar	2	Park	Brewery	Sandwich Place	Burger Joint	Burrito Place	Fast Food Restaurant	Fish & Chips Shop	Italian Restaurant	Steakhouse	Sushi Restaurant
3	Studio District	2	Café	Coffee Shop	Italian Restaurant	Bakery	Latin American Restaurant	Bar	Fish Market	Neighborhood	Seafood Restaurant	Sandwich Place
5	Davisville North	2	Park	Convenience Store	Food & Drink Shop	Sandwich Place	Dance Studio	Hotel	Breakfast Spot	Pizza Place	Gym	Clothing Store
6	North Toronto West	2	Coffee Shop	Sporting Goods Shop	Yoga Studio	Bagel Shop	Gym / Fitness Center	Diner	Dessert Shop	Metro Station	Mexican Restaurant	Park
7	Davisville	2	Dessert Shop	Coffee Shop	Sushi Restaurant	Pizza Place	Gym	Café	Sandwich Place	Italian Restaurant	Movie Theater	Pharmacy
9	Summerhill West, South Hill, Rathnelly, Forest...	2	Coffee Shop	Pub	American Restaurant	Sushi Restaurant	Fried Chicken Joint	Sports Bar	Pizza Place	Bagel Shop	Restaurant	Health & Beauty Service
11	St. James Town, Cabbagetown	2	Restaurant	Coffee Shop	Bakery	Italian Restaurant	Café	Butcher	Pub	Caribbean Restaurant	Playground	Pet Store
12	Church and Wellesley	2	Gay Bar	Burger Joint	Bookstore	Bubble Tea Shop	Salon / Barbershop	Restaurant	Ramen Restaurant	Pub	Pizza Place	General Entertainment
13	Harbourfront, Regent Park	2	Coffee Shop	Park	Bakery	Mexican Restaurant	Breakfast Spot	Restaurant	Café	Pub	Chocolate Shop	Performing Arts Venue
14	Ryerson,	2	Café	Clothing	Gastropub	Ramen	Spa	Hotel	Japanese	Sandwich	Restaurant	Diner

Cluster 4. Label 3: Venue in common- Gym. No. of neighborhoods = 1

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 3, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.sha
```

	Neighbourhood	Cluster Labels	1st Most Common Venues	2nd Most Common Venues	3rd Most Common Venues	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
8	Summerhill East, Moore Park	3	Gym	Playground	Tennis Court	Restaurant	Yoga Studio	Deli / Bodega	Falafel Restaurant	Event Space	Ethiopian Restaurant	Eastern European Restaurant

Cluster 5. Label 3: Venue in common- Park, Swim school. No. of neighborhoods = 1

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 4, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.sha
```

	Neighbourhood	Cluster Labels	1st Most Common Venues	2nd Most Common Venues	3rd Most Common Venues	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
4	Lawrence Park	4	Park	Swim School	Bus Line	Fish & Chips Shop	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Eastern European Restaurant	Dog Run

Discussion and Conclusion

The most common venue in Toronto city is Café/Coffee shop. Cluster 3 with 37 neighborhoods having café/coffee shop as their venue is clustered together. From the map (Figure 2), we see most of the coffee shops/café are located near the bay area away from the Downsview airport.

Most of the neighborhood proximal to the Downview airport and occurring as small individual clusters have Park, Trails, Garden etc, venues in common. This seems to be neighborhood areas with mainly residential spaces with quiet and spacious surroundings amidst trails and parks. However, nearer the bay area, it seems more commercialized centers.

An investor may target localities/neighborhoods such as The Beaches, Riverdale, Davisville, Studio District, North Toronto West etc near the bay areas which are heavily commercialized and frequent visiting spots.

However, since café/ coffee shops and also restaurants are already flooding these neighborhoods, the competition on opening a firm with the same product will face considerable competition and may take time to gain limelight. So, I would recommend some other commercial firms such as Metro Plaza, movie theatres or some public spots in these areas would be more economical. While opening a café/coffee shop near the Downview airport in localities such as Lawrence Park, Summer hill etc would have less competition but again these areas seem to be less socially populated, there will be less customer influx.

In conclusion, this project gives interesting insights an can be used as a navigating tool to recommend places to start new projects/firms in Toronto city but would have had better results if there were more data in terms of crime data within the area, traffic access and allowance of more venues exploration with the Foursquare (limited venues for free calls).