The Battle of the Neighborhoods (Toronto)

1. Introduction

1.1 Problem

The business problem that will be attempted to solve through this project will be: “Where is the best possible location to set up a tourism-related business, F&B outlet and an office in Toronto”. Starting a new business is a complicated process starting from brainstorming ideas, gathering capital, registering with the local government to setting up of establishment, any hiccups at the various stages could result in the failure of the business. Location is an important consideration when setting up and establishment. Different businesses have different location requirements based on its targeted audience. For example, an Asian restaurant would get better business when opened in an Asian community whereas a Michelin-star restaurant would do better when opened in the city-center as the spending abilities of the community is higher and thus more likely to visit. As for offices, accessibility is important to ensure the attractiveness of the company amongst employees.

1.2 Background

According to the 2019 Ease of Doing Business ranking by The World Bank, Canada was ranked 3rd in ease of starting business and 23rd overall, which meant that Canada is an attractive location for new entrepreneur to kick start their business. Home to more than 6 million people, Toronto among all the cities in Canada, is one of the greatest cities in the world to do business as it consistently ranked among the top for global competitiveness, innovation and quality of life. Its people is also highly skilled and multilingual as 64% of Toronto residents between 25 and 64 have finished their post-secondary education. Being within a 90-minute flight away from the USA also makes Toronto an accessible and attractive choice for many Multi-National Companies.

2. Data

2.1 Data Sources

The 3 main data sources are:

- Wikipedia Toronto Postal Code List (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Postal Code Corresponding Coordinates (http://cocl.us/Geospatial_data)
- Foursquare API

2.2 Data Extraction

Toronto Wikipedia page has a list of information of the neighborhoods and the corresponding boroughs and postal code. By making use of the BeautifulSoup library, the list can be extracted via the table HTML element and stored into a data frame. (Named as toronto_df) Due to the large number of unassigned boroughs and neighborhoods, cleaning will be required.

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
...	...	...	...
175	M5Z	Not assigned	Not assigned
176	M6Z	Not assigned	Not assigned
177	M7Z	Not assigned	Not assigned
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...
179	M9Z	Not assigned	Not assigned

Due to the large number of neighborhoods and them being mostly grouped by the postal code, the corresponding coordinates for each postal code will be required for the searching of venues. Coursera Applied Data Science Capstone Week 3 provided a .csv file for its final assignment which contains a list of longitudes and latitudes for each Postal Code. The read_csv is a function in the Pandas library which allows the values in the .csv file to be directly loaded into a data frame. The data frame can then be merged with the Toronto_df using the postal code as the joining key.

	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

Foursquare API is a gateway provided by Foursquare, a location technology platform, for developers to extract venue-related information. The API requires a client ID, client secret, version to access most of its search features and is only obtainable via opening a Foursquare developer account. As business proportion can provide an insight into the profile of the area, Foursquare API explore feature will be used to obtain a list of venues within a pre-defined radius of a given coordinate. For this analysis, the radius has been set to 2000m and the limit of venues set to 300.

	PostCode	PostCode Latitude	PostCode 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	Images Salon & Spa	43.802283	-79.198565	Spa
2	M1B	43.806686	-79.194353	Toronto Pan Am Sports Centre	43.790623	-79.193869	Athletics & Sports
3	M1B	43.806686	-79.194353	Toronto Zoo	43.820582	-79.181551	Zoo
4	M1B	43.806686	-79.194353	Gorilla Exhibit	43.819080	-79.184235	Zoo Exhibit

2.3 Data Cleaning

Cleaning of data is an important process as it prepares the data for analysis. For the data frame with the postal code, borough and neighborhoods, all rows with Borough and Neighborhood as ‘Not assigned’ were removed. To ensure that there are no duplicated postal codes, all rows are grouped by the Postal Code and Borough and neighborhood names were joined by ‘,’.

	Postal Code	Borough	Neighbourhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
...	...	...	...
98	M9N	York	Weston
99	M9P	Etobicoke	Westmount
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
102	M9W	Etobicoke	Northwest, West Humber - Clairville

As clustering requires numerical data, the venues data collected via the Foursquare API would need to be cleaned. Venue Category is a categorical variable which describes the category of a venue nearby. By using Pandas get_dummies, venue categories can be converted to variables so clustering can be done.

	PostCode	Accessories Store	Afghan Restaurant	African Restaurant	Airport	American Restaurant	Amphitheater	Antique Shop	Aquarium	Argentinian Restaurant	...	Volleyball Court	Warehouse Store	Whisky Bar
0	M1B	0.0	0.0	0.019608	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
1	M1C	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
2	M1E	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
3	M1G	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
4	M1H	0.0	0.0	0.000000	0.0	0.020000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
98	M9N	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
99	M9P	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
100	M9R	0.0	0.0	0.000000	0.0	0.014706	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
101	M9V	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
102	M9W	0.0	0.0	0.000000	0.0	0.078947	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0

By making use of the venues collected from the Foursquare API, clustering of the various postal codes using the venue categories will allow a better understanding of profile of the various postal codes which can enable identification of possible locations most suitable to the 3 types businesses: tourism-related business, F&B outlet and office.

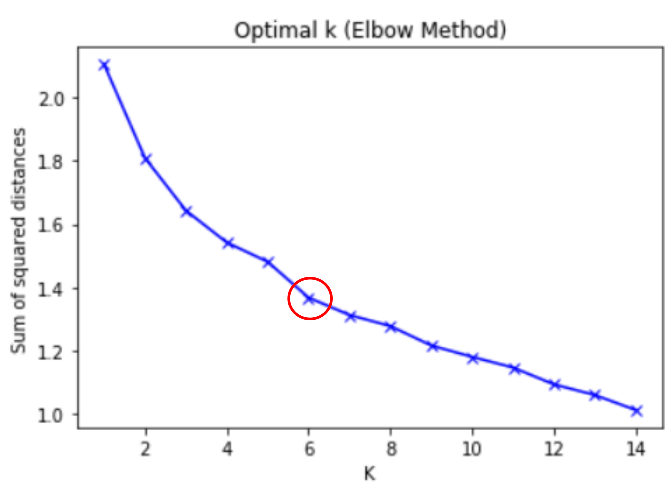
3. Methodology

Due to the nature of the venue data being unlabeled, k-means clustering will be the most appropriate algorithm to use to group similar postal codes. K-Means clustering is an unsupervised learning model which partitions data into non-overlapping *k* number of clusters where *k* is pre-defined by user. By employing K-Means algorithm, similar neighborhoods will be grouped together, and visualization of the various clusters will allow us to find out characteristics of each cluster and identify suitable location for the different businesses.

One of the major weakness of K-means is that the model is very sensitive to the choice of initial *k* centroids. When *k* is too small, the clusters tends to be too generalized and identification of cluster characteristics will be tough. When *k* is too large, the clusters would be overfitted by design and affect the quality of clusters.

3.1 Optimization

In order to find the optimal *k*, one method to validate the number of clusters is the elbow method. To execute this validation method, K-Means model must be built for a range of *k* values and the sum of squared error is plotted on a line graph. The optimal *k* will the point where the gradient of the graph decreases the most. The idea is that we want a small *k* value and at the same time also producing a small sum squared error. The range of *k* used to find the optimal *k* is 1 to 15. By examining the sum of squared error line graph, the "elbow point" identified is 6.



4. Results

After running the data frame through the K-Means (k = 6), an array of cluster number was returned. As the data frame has 317 columns, understanding the characteristics of each clusters will be difficult. Therefore, I decided to create a new data frame, featuring the top 10 most common venues, cluster and the postal code.

	PostCode	Cluster	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
1	M1C	1	Coffee Shop	Breakfast Spot	Pharmacy	Pet Store	Sandwich Place	Bank	Ice Cream Shop	Mexican Restaurant	Grocery Store	Beer Store
2	M1E	1	Pizza Place	Coffee Shop	Breakfast Spot	Park	Grocery Store	Fast Food Restaurant	Bank	Athletics & Sports	Sports Bar	Greek Restaurant
4	M1H	1	Coffee Shop	Clothing Store	Gas Station	Gym	Sandwich Place	Bank	Restaurant	Indian Restaurant	Pizza Place	Fast Food Restaurant
9	M1N	1	Coffee Shop	Grocery Store	Golf Course	Gas Station	Bank	Beer Store	Pizza Place	Ice Cream Shop	Convenience Store	Supermarket
11	M1R	1	Coffee Shop	Restaurant	Fast Food Restaurant	Pizza Place	Middle Eastern Restaurant	Burger Joint	Indian Restaurant	Vietnamese Restaurant	Breakfast Spot	Sandwich Place
12	M1S	1	Chinese Restaurant	Coffee Shop	Restaurant	Pharmacy	Sandwich Place	Discount Store	Indian Restaurant	Bank	Shopping Mall	Caribbean Restaurant

After observing some of the most common venues, I found out that there were numerous transport-related venues which could provide some meaningful insights for each cluster. By identifying the various transport-related venue categories, I grouped the original data by postal code and counted the of transport-related venues and stored them as 'No. of Transport Hub'. The column was then merged with the final data frame for analysis.

	PostCode	Cluster	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	No. of Transport Hub
0	M1B	6	Zoo Exhibit	Fast Food Restaurant	Other Great Outdoors	Zoo	Gas Station	Pizza Place	Curling Ice	Bus Station	Caribbean Restaurant	Skating Rink	1
1	M1C	1	Coffee Shop	Breakfast Spot	Pharmacy	Pet Store	Sandwich Place	Bank	Ice Cream Shop	Mexican Restaurant	Grocery Store	Beer Store	0
2	M1E	1	Pizza Place	Coffee Shop	Breakfast Spot	Park	Grocery Store	Fast Food Restaurant	Bank	Athletics & Sports	Sports Bar	Greek Restaurant	1
3	M1G	5	Coffee Shop	Fast Food Restaurant	Sandwich Place	Discount Store	Bank	Supermarket	Pharmacy	Chinese Restaurant	Pizza Place	Beer Store	0
4	M1H	1	Coffee Shop	Clothing Store	Gas Station	Gym	Sandwich Place	Bank	Restaurant	Indian Restaurant	Pizza Place	Fast Food Restaurant	0

4.1 Cluster Visualization

*Due to space constrains, displayed cluster data is not complete. For full cluster data, please refer to the codes.

Cluster 1

	PostCode	Cluster	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	No. of Transport Hub
1	M1C	1	Coffee Shop	Breakfast Spot	Pharmacy	Pet Store	Sandwich Place	Bank	Ice Cream Shop	Mexican Restaurant	Grocery Store	Beer Store	0
2	M1E	1	Pizza Place	Coffee Shop	Breakfast Spot	Park	Grocery Store	Fast Food Restaurant	Bank	Athletics & Sports	Sports Bar	Greek Restaurant	1
4	M1H	1	Coffee Shop	Clothing Store	Gas Station	Gym	Sandwich Place	Bank	Restaurant	Indian Restaurant	Pizza Place	Fast Food Restaurant	0
9	M1N	1	Coffee Shop	Grocery Store	Golf Course	Gas Station	Bank	Beer Store	Pizza Place	Ice Cream Shop	Convenience Store	Supermarket	0
11	M1R	1	Coffee Shop	Restaurant	Fast Food Restaurant	Pizza Place	Middle Eastern Restaurant	Burger Joint	Indian Restaurant	Vietnamese Restaurant	Breakfast Spot	Sandwich Place	1
12	M1S	1	Chinese Restaurant	Coffee Shop	Restaurant	Pharmacy	Sandwich Place	Discount Store	Indian Restaurant	Bank	Shopping Mall	Caribbean Restaurant	1
13	M1T	1	Bank	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Restaurant	Gas Station	Pharmacy	Pizza Place	Sandwich Place	Men's Store	1

Cluster 2

	PostCode	Cluster	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	No. of Transport Hub
22	M2N	2	Grocery Store	Korean Restaurant	Ramen Restaurant	Supermarket	Bubble Tea Shop	Café	Sushi Restaurant	Theater	Thai Restaurant	Coffee Shop	0
40	M4J	2	Café	Greek Restaurant	Coffee Shop	Bakery	Ice Cream Shop	Gastropub	American Restaurant	Ethiopian Restaurant	Park	Cocktail Bar	0
41	M4K	2	Café	Greek Restaurant	Park	Bakery	Coffee Shop	Ice Cream Shop	American Restaurant	Italian Restaurant	Vietnamese Restaurant	Pub	0
42	M4L	2	Park	Brewery	Café	Coffee Shop	Beach	Bar	BBQ Joint	Pizza Place	Bakery	American Restaurant	0
43	M4M	2	Coffee Shop	Park	Brewery	Bakery	Bar	Café	Vietnamese Restaurant	Diner	French Restaurant	Pet Store	0
45	M4P	2	Coffee Shop	Italian Restaurant	Park	Bakery	Café	Sushi Restaurant	Restaurant	Indian Restaurant	Japanese Restaurant	Pizza Place	0
46	M4R	2	Italian Restaurant	Coffee Shop	Sushi Restaurant	Bakery	Café	Fast Food Restaurant	Japanese Restaurant	Park	Burger Joint	Dessert Shop	0
47	M4S	2	Italian Restaurant	Park	Bakery	Coffee Shop	Café	Sushi Restaurant	Restaurant	Indian Restaurant	Yoga Studio	Bookstore	0
48	M4T	2	Park	Italian Restaurant	Sushi Restaurant	Bakery	Café	Coffee Shop	Dessert Shop	Restaurant	Grocery Store	Thai Restaurant	0
49	M4V	2	Italian Restaurant	Coffee Shop	Park	Café	Sushi Restaurant	French Restaurant	Ice Cream Shop	Grocery Store	Spa	Hotel	0
50	M4W	2	Coffee Shop	Italian Restaurant	Park	Café	Restaurant	Gastropub	Spa	Japanese Restaurant	Grocery Store	Juice Bar	0

Cluster 3

	PostCode	Cluster	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	No. of Transport Hub
16	M1X	3	Grocery Store	Golf Course	Playground	Sculpture Garden	Trail	Farm	Farmers Market	Ethiopian Restaurant	Event Service	Event Space	0

Cluster 4

	PostCode	Cluster	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	No. of Transport Hub
8	M1M	4	Harbor / Marina	Park	Coffee Shop	Sandwich Place	Beach	Pub	Grocery Store	Fast Food Restaurant	Pizza Place	Pharmacy	1
30	M3K	4	Clothing Store	Furniture / Home Store	Athletics & Sports	Coffee Shop	American Restaurant	Cosmetics Shop	Vietnamese Restaurant	Pizza Place	Restaurant	Pharmacy	0
35	M4B	4	Pizza Place	Park	Coffee Shop	Skating Rink	Sandwich Place	Ice Cream Shop	Athletics & Sports	Fast Food Restaurant	Beer Store	Convenience Store	0
36	M4C	4	Park	Coffee Shop	Café	Gastropub	Pizza Place	Skating Rink	Ethiopian Restaurant	Thai Restaurant	Ice Cream Shop	Breakfast Spot	2
37	M4E	4	Coffee Shop	Pub	Breakfast Spot	Beach	Japanese Restaurant	Bakery	BBQ Joint	Ice Cream Shop	Park	Grocery Store	0
38	M4G	4	Coffee Shop	Indian Restaurant	Park	Bakery	Grocery Store	Restaurant	Afghan Restaurant	Turkish Restaurant	Sporting Goods Shop	Sandwich Place	0
39	M4H	4	Coffee Shop	Park	Sandwich Place	Grocery Store	Restaurant	Pizza Place	Indian Restaurant	Pharmacy	Japanese Restaurant	Bakery	0
44	M4N	4	Coffee Shop	Park	Sushi Restaurant	Italian Restaurant	Bakery	Pharmacy	Pizza Place	Pub	Sandwich Place	Restaurant	1
62	M5M	4	Coffee Shop	Italian Restaurant	Sushi Restaurant	Bakery	Bagel Shop	Sandwich Place	Restaurant	Pharmacy	Pizza Place	Pub	2
71	M6A	4	Clothing Store	Coffee Shop	Fast Food Restaurant	Furniture / Home Store	Dessert Shop	Grocery Store	Vietnamese Restaurant	Bank	Pizza Place	Cosmetics Shop	0

Cluster 5

	PostCode	Cluster	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	No. of Transport Hub
3	M1G	5	Coffee Shop	Fast Food Restaurant	Sandwich Place	Discount Store	Bank	Supermarket	Pharmacy	Chinese Restaurant	Pizza Place	Beer Store	0
5	M1J	5	Coffee Shop	Fast Food Restaurant	Pizza Place	Sandwich Place	Pharmacy	Grocery Store	Liquor Store	Bank	Gym	Beer Store	0
6	M1K	5	Grocery Store	Pharmacy	Chinese Restaurant	Coffee Shop	Discount Store	Fast Food Restaurant	Beer Store	Bank	Sandwich Place	Pizza Place	3
7	M1L	5	Coffee Shop	Fast Food Restaurant	Sandwich Place	Pizza Place	Restaurant	Hardware Store	Clothing Store	Beer Store	Burger Joint	Burrito Place	0
10	M1P	5	Coffee Shop	Fast Food Restaurant	Pizza Place	Sandwich Place	Restaurant	Indian Restaurant	Breakfast Spot	Bank	Gas Station	Pharmacy	0
32	M3M	5	Pharmacy	Coffee Shop	Pizza Place	Bank	Fast Food Restaurant	Vietnamese Restaurant	Grocery Store	Gas Station	Supermarket	Sandwich Place	0
33	M3N	5	Coffee Shop	Hotel	Fast Food Restaurant	Gas Station	Pizza Place	Grocery Store	Kitchen Supply Store	Latin American Restaurant	Theater	Tennis Stadium	0
34	M4A	5	Coffee Shop	Fast Food Restaurant	Gym	Sandwich Place	Clothing Store	Japanese Restaurant	Grocery Store	Middle Eastern Restaurant	Department Store	Gas Station	0
79	M6L	5	Vietnamese Restaurant	Coffee Shop	Fast Food Restaurant	Gas Station	Pizza Place	Furniture / Home Store	Bank	Bakery	Supermarket	Grocery Store	0
80	M6M	5	Coffee Shop	Furniture / Home Store	Sandwich Place	Fast Food Restaurant	Pizza Place	Gas Station	Grocery Store	Bakery	Burger Joint	Italian Restaurant	0

Cluster 6

	PostCode	Cluster	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	No. of Transport Hub
0	M1B	6	Zoo Exhibit	Fast Food Restaurant	Other Great Outdoors	Zoo	Gas Station	Pizza Place	Curling Ice	Bus Station	Caribbean Restaurant	Skating Rink	1

5. Discussion

After analyzing the various clusters, I have managed to make a few observations for the various clusters. Clusters 3 and 6 will not be analyzed as both clusters only have 1 postal code, it is difficult to identify useful characteristics for neighborhood profiling. As for clusters 1, 2, 4 and 5, the following are some observation:

Cluster	Observations
1 (Central Business District)	• Bank is on the top 10 common location for majority of the postal codes • Most postal codes have transport hubs
2 (Tourist Site)	• Café appears on the top 10 common location for majority of the postal code • Most postal codes do not have transport hubs • Recreation F&B such as bars and dessert shops are common stores for postal codes in this cluster • Hotels more commonly appear in the top 10 common venues
4 (Upper Class District)	• Stores that offer consumer goods such as Furniture/Home, Electronics, Sports goods, Clothing and Cosmetics shop are among common venues • Recreation facilities such as Beach, Skating Ring, Golf Course and Trail are in common area of some postal codes in this cluster
5 (Lower Class District)	• Cheaper restaurants such as Fast Food Restaurants, Pizza Place and Sandwich Place are among the common venues • Small scale commodity store such as Grocery Store and Convenient Store are among the common venues

Through the observations, cluster 1 is being named as Central Business District as there is an abundance of banks. With good accessibility based on the transport hubs, cluster 1 is good for setting up of offices as employees will be able to get to work easily, making the working environment more favorable. Cluster 2 is being named as Tourist Site as there are many recreation F&B, Cafes and Hotels which is likely to be areas which is largely visited by tourist. Opening a tourism-related business in neighborhoods in cluster 2 will be appropriate as tourist are targeted. Cluster 4 is being named as Upper-Class District as there is evidently more consumer goods stores which can offer the residents with spending abilities to shop for goods. Recreation facilities which are common in cluster 4 as residents with higher standard of living is likely to visit such facilities more often thus, giving rise to more facilities in such area. F&B restaurants which are targeted at customers with higher spending capabilities should look to open their business in cluster 4. Cluster 5 is being named as Lower-Class District as there are cheaper F&B options such as Fast Food restaurants, Pizza Place and Sandwich Place which gives residents options which are more affordable. Small scale commodity store such as Grocery store and Convenient store provides residents with access to more basic commodities such as food, drink items and household items in comparison with supermarkets or malls which provides wider assortment of things for sale but also at a higher price. F&B outlets which aims to provide consumers with wider variety of food at more economic prices should look to start their business in cluster 5.

One interesting observation made was coffee shop is in the top few common venue in all the postal code. Coffee shop could possibly be a popular location for locals to visit for meals and drinks thus, there are many coffee shops co-existing despite competition. F&B operators can study reasons why locals always visit coffee shop to gain insights and further improve and customize their business to target the people of Toronto.

6. Conclusion

By making use of venue data provided by Foursquare API, K-Means was employed to cluster the various postal codes into 6 different clusters. Through the analysis of common venues in each cluster, I was able to identify characteristics in each cluster and profile them individually which is able to solve the problem of 'Where is the best possible location to set up a tourism-related business, F&B outlet and an office in Toronto?'. It is interesting to learn how the demographics of venues in a certain area helps to profile the nature of the neighborhoods and its residents. Other data such as population size and earnings per capital can be used in the future to improve the model for more obvious characteristics of each cluster.

One difficulty faced during this project is the inconsistent shape and size of each postal code. Due to the inconsistencies, it is inaccurate to run the Foursquare API explore feature with a static radius. Foursquare API also has unbalanced number of venues stored in its database thus, limiting countries with lesser venues from being chosen for this project.