Introduction

Toronto is considered as one of the most populous cities in Canada and many immigrants choose Toronto as their number one destination when it comes to living due to its multi-cultural inclusiveness. As a result, different cuisines from all around the global can be found in Toronto as immigrants brought different food from their hometowns. No matter what you enjoy, sushi, Korean BBQ or Chinese hot pot etc., you can find them in Toronto.

All kinds of restaurants require servicing companies to provide some kind of support such as food supply and utensils providing etc. There is an issue, however, troubles many restaurant servicing companies and that is how to choose the location of the companies that enables the company serve as many restaurants as possible with relative low cost.

For this project, data from various sources were utilized and machine learning technique was applied to figure out the optimal location to build up a restaurant servicing company.

Data

Data for this project was obtained primarily from three sources:

1. Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. Foursquare API
3. http://cocl.us/Geospatial_data

Google map and Google search were also used as supplemental tools to explore areas and find geographical data.

Data from Wikipedia was scraped to data frame by using BeautifulSoup library, and it provides information of postal code, borough and neighborhood name. Foursquare data was obtained via API and provided more detailed information about restaurant name, restaurant type and restaurant geographical location. In addition to the data from Wikipedia and Foursquare, a online csv file from http://cocl.us/Geospatial_data was used to fetch the latitude and longitude information for each unique postal code.

Methodology

The data from Wikipedia was imported into the data frame and data wrangling was performed on the Wikipedia data. We will only focus on the data that has a valid borough, therefore rows that contain “Not Assigned” borough was removed from the data frame. Besides, neighborhoods from the same postal code was integrated together and separated by comma. In addition, “Not Assigned” neighborhood names were replaced by their corresponding borough name. The first five rows of the data frame are shown in table 1. The size of the data frame was observed with 3

columns and 103 rows. Each row shows a unique postcode and its corresponding borough name and neighborhoods.

Table 1 First five rows of the data frame after data wrangling

	Postcode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront,Regent Park
3	M6A	North York	Lawrence Heights,Lawrence Manor
4	M7A	Queen's Park	Queen's Park

In order to draw connections between data from Wikipedia and Foursquare, latitude and longitude was obtained from online source(http://cocl.us/Geospatial_data) and attached to the data frame that was previously created. The first five rows of this new data frame with latitude and longitude information incorporated can be found in table 2. The latitude and longitude for Toronto was determined to be 43.653963, -79.387207 by using geocoders.

Table 2 First five rows of the data frame with latitude and longitude

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636
3	M6A	North York	Lawrence Heights,Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494

After the data frame with latitude and longitude information incorporated, Foursquare API was accessed. By using latitude and longitude information of different neighborhoods, it is possible to explore venues near each neighborhood and venues information was extracted from Foursquare. The venues information including venue latitude and longitude and venue category was added to the data frame. First five rows of the data frame are given in table 3.

Table 3 First five rows of the data frame with venues information

	Postcode	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M3A	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	M3A	Parkwoods	43.753259	-79.329656	KFC	43.754387	-79.333021	Fast Food Restaurant
2	M3A	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
3	M4A	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	M4A	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

In order to apply machine learning technique and allow machine learning algorithm to perform effectively, one hot encoding was performed to transform categorical data variables. The resulting data frame after one hot encoding has 1338 rows and 238 columns, which means 1331 venues in Toronto belongs to 240 venues categories. After these venues were grouped by neighborhoods, we obtained new data frame with 99 rows and 238 columns.

For the next step, top revenues of each neighborhood were found and the first five rows of the finding is given in table 4.

Table 4 First five row of data frame of top 10 venues for each neighborhood

	Neighborhood	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
0	Adelaide\n,King\n,Richmond\n	Steakhouse	Pizza Place	Hotel	Asian Restaurant	Café	Bar	Speakeasy	Smoke Shop	Seafood Restaurant	Plaza
1	Agincourt	Lounge	Chinese Restaurant	Breakfast Spot	Sandwich Place	Curling Ice	Dumpling Restaurant	Drugstore	Dog Run	Discount Store	Diner
2	Agincourt North,L'Amoreaux East\n,Milliken,Ste...	Park	Playground	Asian Restaurant	College Arts Building	College Gym	Dumpling Restaurant	Drugstore	Dog Run	Discount Store	Diner
3	Albion Gardens\n,Beaumond Heights,Humbergate\n...	Grocery Store	Coffee Shop	Pizza Place	Beer Store	Sandwich Place	Fast Food Restaurant	Fried Chicken Joint	Liquor Store	Pharmacy	Video Store
4	Alderwood,Long Branch	Pizza Place	Pub	Coffee Shop	Skating Rink	Gym	Pharmacy	Dance Studio	Pool	Sandwich Place	Deli / Bodega

K-means clustering was applied to group similar neighborhoods. The result after clustering is shown in table 5 to table 9. (Only top results are shown in the table, please refer to code page for full result.)

Table 5 Cluster label 0 result

	Borough	Postcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
7	North York	M3B	North York	Don Mills North\n	43.745906	-79.352188	0.0	Gym / Fitness Center	Basketball Court	Japanese Restaurant
8	East York	M4B	East York	Woodbine Gardens,Parkview Hill	43.706397	-79.309937	0.0	Pizza Place	Fast Food Restaurant	Gym / Fitness Center
10	North York	M6B	North York	Glencairn\n	43.709577	-79.445073	0.0	Japanese Restaurant	Italian Restaurant	Asian Restaurant
11	Etobicoke	M9B	Etobicoke	Cloverdale\n,Islington,Martin Grove\n,Princess...	43.650943	-79.554724	0.0	Golf Course	Bank	Women's Store
12	Scarborough	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	0.0	Golf Course	Bar	Women's Store
16	York	M6C	York	Humewood-Cedarvale	43.693781	-79.428191	0.0	Field	Dog Run	Trail
18	Scarborough	M1E	Scarborough	Guildwood\n,Morningside,West Hill	43.763573	-79.188711	0.0	Electronics Store	Rental Car Location	Pizza Place
19	East Toronto	M4E	East Toronto	The Beaches	43.676357	-79.293031	0.0	Other Great Outdoors	Trail	Pub
26	Scarborough	M1H	Scarborough	Cedarbrae\n	43.773136	-79.239476	0.0	Fried Chicken Joint	Bank	Athletics & Sports
27	North York	M2H	North York	Hillcrest Village	43.803762	-79.363452	0.0	Golf Course	Dog Run	Fast Food Restaurant
31	West Toronto	M6H	West Toronto	Dovercourt Village,Dufferin\n	43.669005	-79.442259	0.0	Supermarket	Bakery	Pharmacy

Table 6 Cluster label 1 result

	Borough	Postcode	Borough	Neighborhood	Latitude	Longitude	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
6	Scarborough	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	1.0	Fast Food Restaurant	Women's Store	Event Space	Empanada Restaurant	Electronics Store	Eastern European Restaurant

Table 7 Cluster label 2 result

	Borough	Postcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
1	North York	M4A	North York	Victoria Village	43.725882	-79.315572	2.0	Coffee Shop	Portuguese Restaurant	Intersection	H
2	Downtown Toronto	M5A	Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636	2.0	Coffee Shop	Bakery	Park	C
3	North York	M6A	North York	Lawrence Heights,Lawrence Manor	43.718518	-79.464763	2.0	Furniture / Home Store	Clothing Store	Women's Store	M
4	Queen's Park	M7A	Queen's Park	Queen's Park	43.662301	-79.389494	2.0	Coffee Shop	Park	Gym	
9	Downtown Toronto	M5B	Downtown Toronto	Ryerson\n,Garden District\n	43.657162	-79.378937	2.0	Café	Clothing Store	Shopping Mall	
13	North York	M3C	North York	Flemingdon Park,Don Mills South\n	43.725900	-79.340923	2.0	Coffee Shop	Asian Restaurant	Gym	
14	East York	M4C	East York	Woodbine Heights	43.695344	-79.318389	2.0	Skating Rink	Park	Pharmacy	
15	Downtown Toronto	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	2.0	Coffee Shop	Gastropub	Restaurant	
17	Etobicoke	M9C	Etobicoke	Bloordale Gardens\n,Eringate\n,Markland Wood,O...	43.643515	-79.577201	2.0	Park	Coffee Shop	Beer Store	
20	Downtown Toronto	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	2.0	Café	Coffee Shop	Cocktail Bar	

Table 8 Cluster label 3 result

	Borough	Postcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	North York	M3A	North York	Parkwoods	43.753259	-79.329656	3.0	Park	Fast Food Restaurant	Food & Drink Shop	Women's Store
21	York	M6E	York	Caledonia-Fairbanks\n	43.689026	-79.453512	3.0	Park	Pharmacy	Fast Food Restaurant	Market
40	North York	M3K	North York	CFB Toronto,Downsview East\n	43.737473	-79.464763	3.0	Park	Airport	Curling Ice	Ethiopian Restaurant
46	North York	M3L	North York	Downsview West	43.739015	-79.506944	3.0	Park	Grocery Store	Shopping Mall	Bank
61	Central Toronto	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	3.0	Park	Swim School	Bus Line	Coffee Shop
64	York	M9N	York	Weston	43.706876	-79.518188	3.0	Park	Convenience Store	Dance Studio	Ethiopian Restaurant
66	North York	M2P	North York	York Mills West\n	43.752758	-79.400049	3.0	Park	Bank	Convenience Store	Dance Studio
85	Scarborough	M1V	Scarborough	Agincourt North,L'Amoreaux East\n,Milliken,Ste...	43.815252	-79.284577	3.0	Playground	Park	Women's Store	Cuban Restaurant
91	Downtown Toronto	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	3.0	Park	Playground	Trail	Building

Table 9 Cluster label 4 result

	Borough	Postcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
32	Scarborough	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	4.0	Pizza Place	Playground	Convenience Store	Women's Store
35	East York	M4J	East York	East Toronto	43.685347	-79.338106	4.0	Park	Coffee Shop	Pizza Place	Convenience Store
50	North York	M9L	North York	Humber Summit	43.756303	-79.565963	4.0	Empanada Restaurant	Pizza Place	Women's Store	Curling Ice
63	York	M6N	York	The Junction North\n,Runnymede\n	43.673185	-79.487262	4.0	Grocery Store	Bus Line	Pizza Place	Brewery
70	Etobicoke	M9P	Etobicoke	Westmount\n	43.696319	-79.532242	4.0	Pizza Place	Coffee Shop	Chinese Restaurant	Sandwich Place
72	North York	M2R	North York	Willowdale West	43.782736	-79.442259	4.0	Coffee Shop	Pharmacy	Butcher	Pizza Place
77	Etobicoke	M9R	Etobicoke	Kingsview Village,Martin Grove Gardens\n,Richv...	43.688905	-79.554724	4.0	Park	Bus Line	Pizza Place	Mobile Phone Shop
89	Etobicoke	M9V	Etobicoke	Albion Gardens\n,Beaumond Heights,Humbergate\n...	43.739416	-79.588437	4.0	Grocery Store	Pizza Place	Coffee Shop	Video Store
93	Etobicoke	M8W	Etobicoke	Alderwood,Long Branch	43.602414	-79.543484	4.0	Pizza Place	Gym	Skating Rink	Coffee Shop

Results

From the K-means clustering result, it can be observed that places given in cluster label 2 (table 7) have higher density in food servicing stores. Among all the boroughs shown in table 7, it is noticed that Downtown Toronto and North York have more restaurants than the other boroughs. It is reasonable that we choose somewhere around North York and Downtown Toronto for building the food servicing company.

Discussion

Visualization is essential in result displaying and there are various visualization tools to perform the duty. The easiest way is to use Google Map and pin point each neighborhood on the map. This result is given in figure 1.

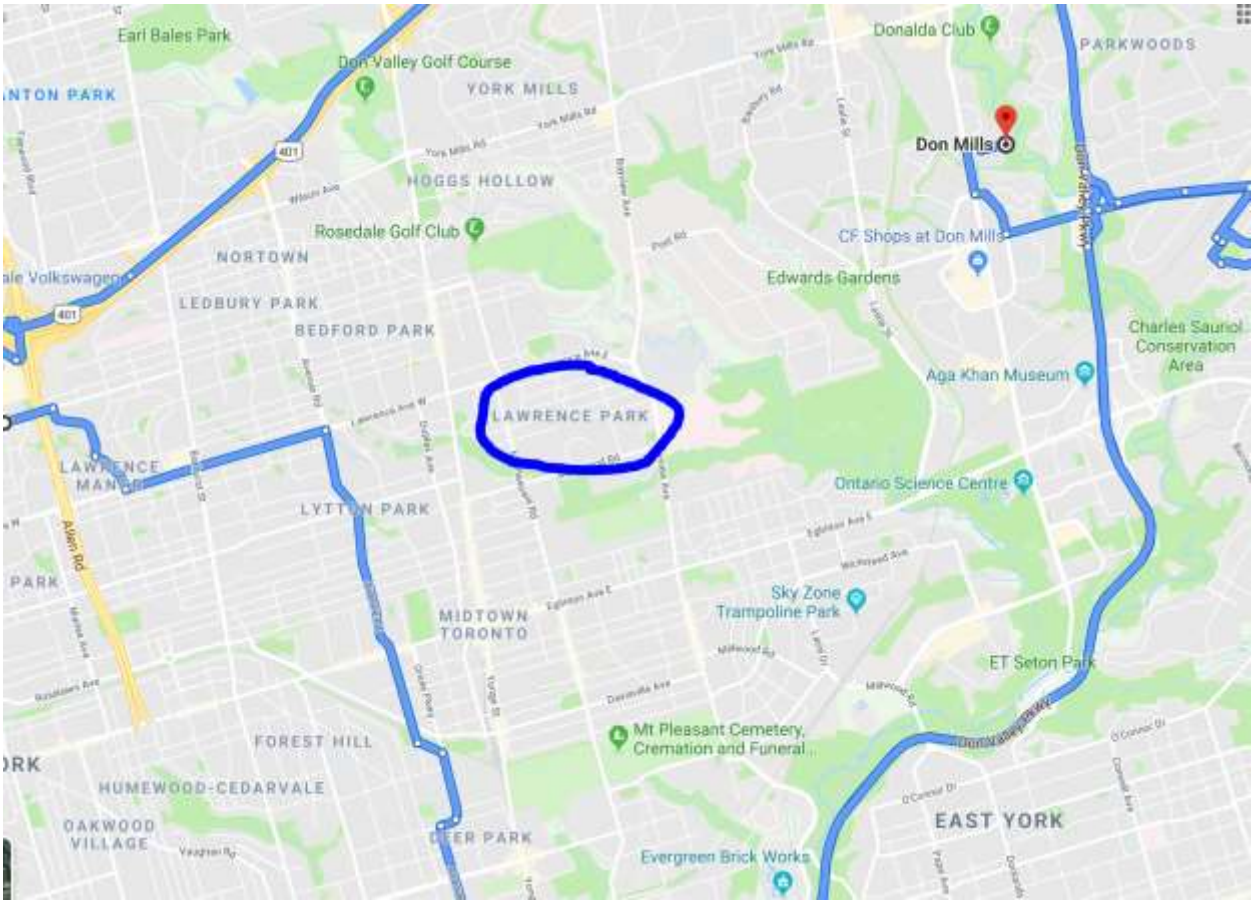

Figure 1 Result for company location selection

It is reasonable that we choose the neighborhood to build the company that has the shortest sum difference to all other neighborhoods. It can be observed that Lawrence Park is in the middle of all other neighborhoods in cluster 2 result and Lawrence Park can be a great spot to build a food servicing company.

Conclusion

This report studied different neighborhoods in Toronto and determined the best location for building up food servicing company by using various data and machine learning techniques. Data was obtained from various sources, wrangled and examined before applying K-means clustering. The Google Map was used to show the result and it shows that Lawrence Park is the best neighborhood for building up the food servicing company due to its closeness to all other neighborhoods that have high density food servicing stores.