Data Science

Capstone Project

Optimal location for opening Thai restaurant in Toronto Canada

October, 2019

Introduction/Business Problem section in final report

Project Name: Finding location to open new Thai Restuarant in Toronto, Canada

The project of finding location to open Thai restuarant in Toronto, Cananda. This project uses the lacation data tools to explore a geographical location, especially the Foursquare location data to compare neighborhoods in Toronto.
Introduction:

Not every available space is right for a restaurant. A good restaurant location is harder to find than some people think. Of course, food and service are important to the success of a restaurant, but the location can be just as crucial, especially in the early years. Know Your Neighbors When looking for a restaurant location, consider who else is doing business in the neighborhood. Are there already half a dozen restaurants with the same concept as yours?

Business Problem :

Finding the suitable location for the entrepreneur to starting the new Thai Restaurant in Toronto is quite difficult in the past, especially because of the big area of this city and lack of data about geographical location, neighborhood, and people's life style.

Now that we have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of machine learning, we have the opportunity to be as creative and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods of our choice or to come up with a problem that we can use the Foursquare location data to solve.

Data : To accomplish this project, we will use the data as below;

1. list of neighborhoods in Toronto
2. Latitude and Longitude of the neighborhoods
3. Venue data related to Total Resturants in Toronto ,especially Thai Restuarants located in this city

Methodology

1. Download Dataset

1.1 We need the data set of list of postal codes of Canada, which we can download via link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
Français
Edit links

List of postal codes of Canada: M

From Wikipedia, the free encyclopedia

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website,[1] via its applications for such smartphones as the iPhone and BlackBerry,[2] and sells hard-copy directories and CD-ROMs. Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Toronto - FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0, however, the postal code M0R 8T0 is assigned to an amazon warehouse in Mississauga, suggesting that Canada Post may be allocating the M0 FSA for high volume addresses.

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Rouge
M1B	Scarborough	Malvern
M2B	Not assigned	Not assigned

1.2 We use the BeautifulSoup to extract the data table that contains 3 columns: Postcode, Borough Neighbourhood. Then we save in the file name "toronto.csv". We will transfer the data in csv file to be dataframe by using pasdas. The result will be shown in dataframe table as below;

	Postalcode	Borough	Neighborhood
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	Harbourfront

1.3 We have to clean the dataset in dataframe that has the "Not assigned" data and combine the same postal code into one row with the neighborhood separated with comma.

	Postalcode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights

1.4 Furthermore, we need a csv file that has the geographical coordinated of each postal code via Link: http://cocl.us/Geospatial_data and name this file "Torontolonglatdata.csv"

	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

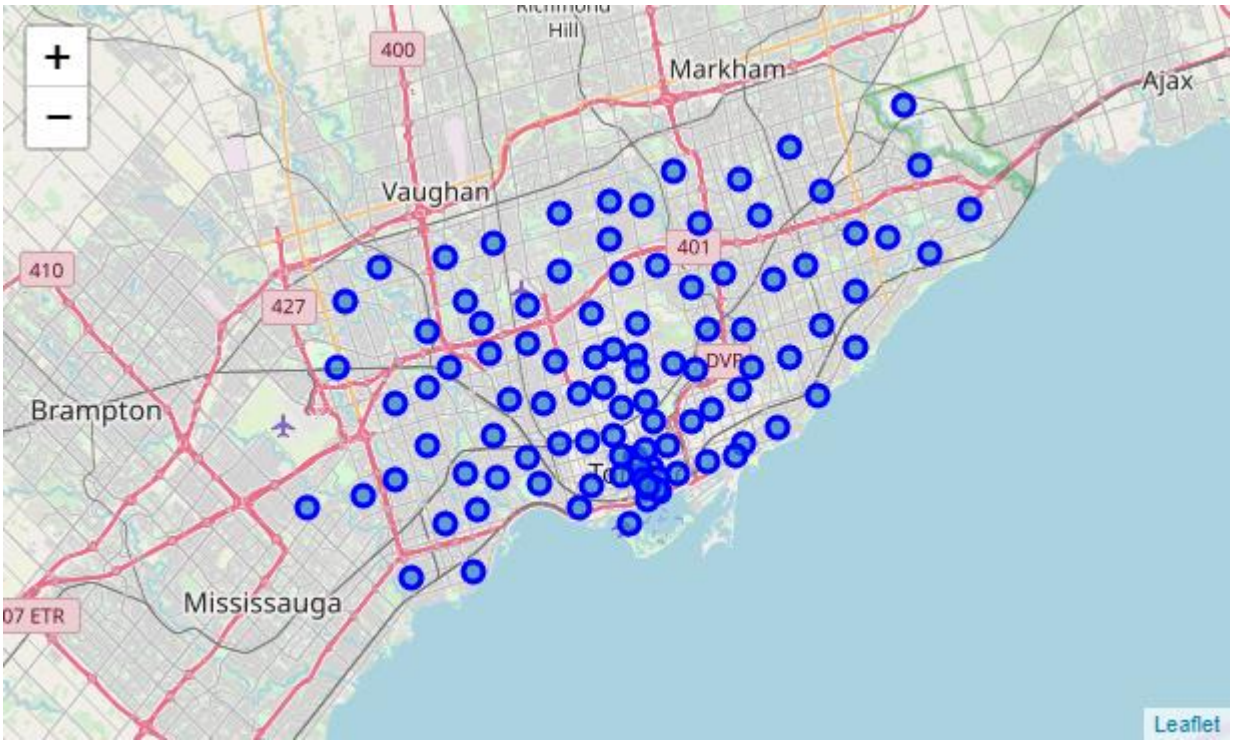
1.5 In "Torontolonglatdata.csv" file, we have 3 columns that is Postal Code, Latitude, and Longitude.
1.6 We merge the borough and neighborhood with the latitude and longitude data and get the new table which has 5 columns: Postal code, Borough, Neighborhood, Latitude, and Longitude with 103 rows × 5 columns.

	Postalcode	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
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
7	M3B	North York	Don Mills North	43.745906	-79.352188
8	M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937
9	M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	43.650943	-79.554724
12	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
13	M3C	North York	Flemingdon Park, Don Mills South	43.725900	-79.340923
14	M4C	East York	Woodbine Heights	43.695344	-79.318389
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
16	M6C	York	Humewood-Cedarvale	43.693781	-79.428191
17	M9C	Etobicoke	Bloordale Gardens, Eringate, Markland Wood, Ol...	43.643515	-79.577201

2. Expore and Cluster the dataset

2.1 Explore and cluster the neighborhoods in Toronto by using geopy, json, pandas, matplotlib folium library to get the latitude and longitude values of Toronto.

2.2 We create a map of Toronto with neighborhoods superimposed on top.



2.3 We expore the dataset of 200 venues within a radius of 2000 meters by connect foursquare API and then we can get the nearby venues returned for each neighborhoods.

	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	Parkwoods	43.753259	-79.329656	TTC stop - 44 Valley Woods	43.755402	-79.333741	Bus Stop
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
5	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
6	Victoria Village	43.725882	-79.315572	The Frig	43.727051	-79.317418	French Restaurant
7	Victoria Village	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
8	Harbourfront, Regent Park	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
9	Harbourfront, Regent Park	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop

2.4 We group rows by neighborhood and calculate the number of Thai restaurant and total number of restaurants in each region.

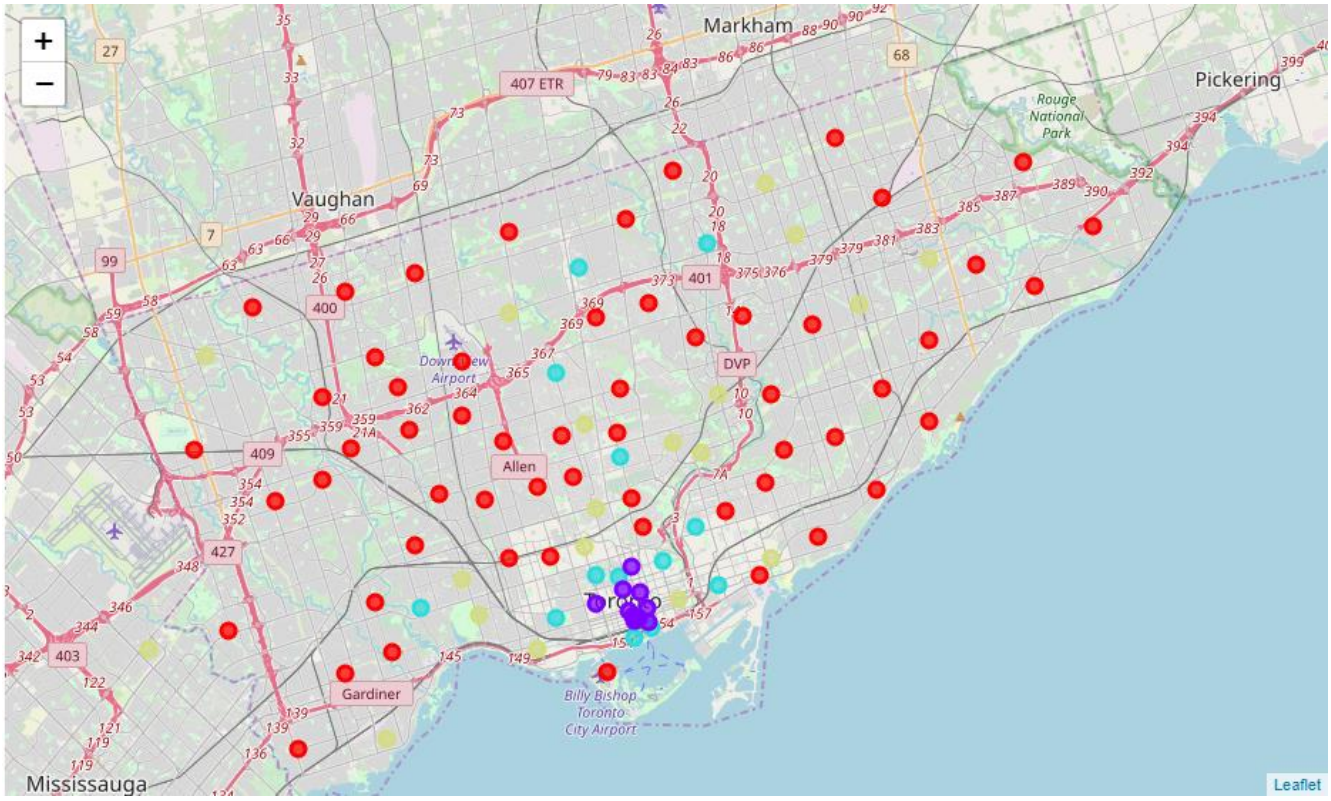
	Neighborhood	Total Restaurants	Thai Restaurants
0	Adelaide, King, Richmond	39	4
1	Agincourt	2	0
2	Agincourt North, L'Amoreaux East, Milliken, St...	0	0
3	Albion Gardens, Beaumond Heights, Humbergate, ...	3	0
4	Alderwood, Long Branch	1	0
5	Bathurst Manor, Downsview North, Wilson Heights	8	0
6	Bayview Village	2	0
7	Bedford Park, Lawrence Manor East	12	1
8	Berczy Park	18	1
9	Birch Cliff, Cliffside West	0	0
10	Bloordale Gardens, Eringate, Markland Wood, Ol...	0	0
11	Brockton, Exhibition Place, Parkdale Village	6	0
12	Business Reply Mail Processing Centre 969 Eastern	2	0
13	CFB Toronto, Downsview East	0	0
14	CN Tower, Bathurst Quay, Island airport, Harbo...	0	0
15	Cabbagetown, St. James Town	14	1
16	Caledonia-Fairbanks	1	0
17	Canada Post Gateway Processing Centre	5	0
18	Cedarbrae	4	1
19	Central Bay Street	35	1
20	Chinatown, Grange Park, Kensington Market	33	1
21	Christie	3	0
22	Church and Wellesley	35	1
23	Clairlea, Golden Mile, Oakridge	1	0
24	Clarks Corners, Sullivan, Tam O'Shanter	6	1
25	Cliffcrest, Cliffside, Scarborough Village West	1	0
26	Cloverdale, Islington, Martin Grove, Princess	0	0

2.5 We run K-means to cluster the neighborhoods into 4 clusters and show all clusters in the map.

```
# set number of clusters
kclusters = 4

toronto_grouped_clustering = df_restaurants.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_
# to change use .astype()
```

```
array([1, 0, 0, 0, 0, 3, 0, 2, 2, 0, 0, 3, 0, 0, 0, 2, 0, 3, 3, 1, 1, 0,
       1, 0, 3, 0, 0, 1, 2, 0, 3, 3, 1, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 2,
       1, 3, 0, 0, 0, 2, 2, 3, 3, 0, 0, 3, 0, 0, 0, 0, 0, 3, 0, 0, 3, 2,
       0, 0, 3, 0, 0, 3, 0, 2, 0, 0, 0, 2, 1, 0, 1, 1, 2, 3, 0, 3, 2, 0,
       0, 3, 0, 0, 0, 2, 0, 0, 0, 0, 0], dtype=int32)
```

	Postalcode	Borough	Neighborhood	Latitude	Longitude	ClusterLabel	Total Restaurants	Thai Restaurants
0	M3A	North York	Parkwoods	43.753259	-79.329656	0.0	2.0	0.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	0.0	2.0	0.0
2	M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636	3.0	10.0	0.0
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763	0.0	0.0	0.0
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494	2.0	13.0	0.0
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242	NaN	NaN	NaN
6	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	0.0	1.0	0.0
7	M3B	North York	Don Mills North	43.745906	-79.352188	0.0	2.0	0.0
8	M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	0.0	2.0	0.0
9	M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937	1.0	28.0	1.0



3. Examine Cluster

3.1. We explore each cluster.

Cluster1:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	ClusterLabel	Total Restaurants	Thai Restaurants
0	M3A	North York	Parkwoods	43.753259	-79.329656	0	2.0	0.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	0	2.0	0.0
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763	0	0.0	0.0
6	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	0	1.0	0.0
7	M3B	North York	Don Mills North	43.745906	-79.352188	0	2.0	0.0
8	M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	0	2.0	0.0
10	M6B	North York	Glencairn	43.709577	-79.445073	0	1.0	0.0
12	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0	0.0	0.0
14	M4C	East York	Woodbine Heights	43.695344	-79.318389	0	1.0	0.0
16	M6C	York	Humewood-Cedarvale	43.693781	-79.428191	0	0.0	0.0
17	M9C	Etobicoke	Bloordale Gardens, Eringate, Markland Wood, Ol...	43.643515	-79.577201	0	0.0	0.0
18	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	0	2.0	0.0
19	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	0.0	0.0
21	M6E	York	Caledonia-Fairbanks	43.689026	-79.453512	0	1.0	0.0
22	M1G	Scarborough	Woburn	43.770992	-79.216917	0	1.0	0.0
25	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0	3.0	0.0
27	M2H	North York	Hillcrest Village	43.803762	-79.363452	0	1.0	0.0
31	M6H	West Toronto	Dovercourt Village, Dufferin	43.669005	-79.442259	0	1.0	0.0
32	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0	0.0	0.0
34	M3J	North York	Northwood Park, York University	43.767980	-79.487262	0	2.0	0.0
35	M4J	East York	East Toronto	43.685347	-79.338106	0	0.0	0.0

Cluster2:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	ClusterLabel	Total Restaurants	Thai Restaurants
9	M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937	1	28.0	1.0
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	1	35.0	1.0
24	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	1	36.0	1.0
30	M5H	Downtown Toronto	Adelaide, King, Richmond	43.650571	-79.384568	1	39.0	4.0
42	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576	1	35.0	1.0
48	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	1	38.0	2.0
84	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	1	31.0	0.0
92	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846	1	31.0	1.0
97	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	1	38.0	2.0
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	1	35.0	1.0

Cluster3:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	ClusterLabel	Total Restaurants	Thai Restaurants
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494	2	13.0	0.0
20	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	2	18.0	1.0
33	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	2	15.0	0.0
36	M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752	2	20.0	0.0
37	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750	2	20.0	0.0
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	2	16.0	0.0
54	M4M	East Toronto	Studio District	43.659526	-79.340923	2	12.0	1.0
55	M5M	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750	2	11.0	1.0
59	M2N	North York	Willowdale South	43.770120	-79.408493	2	14.0	0.0
79	M4S	Central Toronto	Davisville	43.704324	-79.388790	2	16.0	2.0
80	M5S	Downtown Toronto	Harbord, University of Toronto	43.662696	-79.400049	2	12.0	0.0
81	M6S	West Toronto	Runnymede, Swansea	43.651571	-79.484450	2	12.0	0.0
96	M4X	Downtown Toronto	Cabbagetown, St. James Town	43.667967	-79.367675	2	15.0	1.0

Cluster 4:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	ClusterLabel	Total Restaurants	Thai Restaurants
2	M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636	3	10.0	0.0
13	M3C	North York	Flemingdon Park, Don Mills South	43.725900	-79.340923	3	9.0	0.0
23	M4G	East York	Leaside	43.709060	-79.363452	3	8.0	0.0
26	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	3	4.0	1.0
28	M3H	North York	Bathurst Manor, Downsview North, Wilson Heights	43.754328	-79.442259	3	8.0	0.0
29	M4H	East York	Thorncliffe Park	43.705369	-79.349372	3	6.0	0.0
43	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	3	6.0	0.0
47	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572	3	7.0	0.0
65	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	3	4.0	0.0
69	M6P	West Toronto	High Park, The Junction South	43.661608	-79.464763	3	9.0	2.0
73	M4R	Central Toronto	North Toronto West	43.715383	-79.405678	3	5.0	0.0
74	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678	3	8.0	0.0
75	M6R	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325	3	6.0	0.0
76	M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-79.615819	3	5.0	0.0
82	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302	3	6.0	1.0
86	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049	3	4.0	0.0
88	M8V	Etobicoke	Humber Bay Shores, Mimico South, New Toronto	43.605647	-79.501321	3	5.0	0.0
89	M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	3	4.0	0.0

3.2. The results from K-means clustering show that we can group Toronto neighborhoods into 4 clusters based on the number of Thai Restaurant and the total amount of restaurant in Toronto as following detail;

- Cluster 1 : Neighborhoods with small amount of restaurant and no Thai Restaurant (mean = 0.87, 0.0)
- Cluster 2: Neighborhoods with medium amount of restaurant and little Thai Restaurant (mean = 34.6, 1.4)
- Cluster 3: Neighborhoods with small amount of restaurant and little Thai Restaurant (mean = 14.92, 0.46)
- Cluster 4: Neighborhoods with small amount of restaurant and little Thai Restaurant (mean = 6.36, 0.21)

Conclusion :

We choose the suitable cluster which has small amount and no Thai Restaurant lacated in its area that is cluster No. 1. We create a map that shows the name of Neighborhood, cluster label , total number of restuarant, and total number of Thai Restaurant in cluster No.1