The Battle of Neighborhood

Introduction/Business Problem

Opening a business in Toronto is a target of this project. To choose the business type and the location of the business, the location data and characteristic of each region should be investigated. To achieve this goal, the borough and area data would be imported from the Foursquare as we have some experience on that. The below are the list to be done:

- Area name and Zip code.

- Information about the individual area such as median income, house price, public school score, or residential to commercial ratio.

- Data about venues already located in the region/area.

Data Acquisition

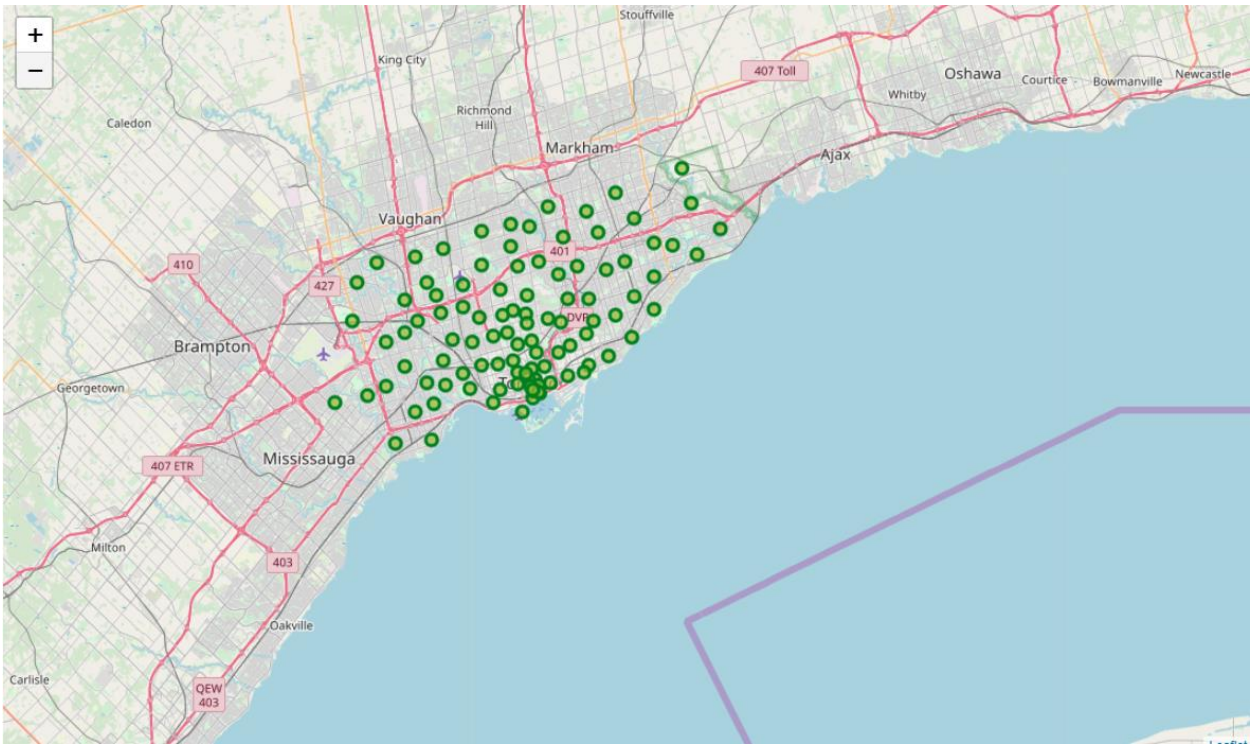
As we do for the previous project, the area/borough segmentation can be done according to the data in Wikipedia. Foursquare data will be added once needed to investigate the area venue information

Scraping the data from the Wikipedia and creating neighborhood table

Using postal code data from Wikipedia, assign the neighborhood into postal code and borough

Postcode	Borough	Neighbourhood	Borough
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
5	M6A	North York	Lawrence Heights
6	M6A	North York	Lawrence Manor
7	M7A	Downtown Toronto	Queen's Park
8	M8A	Not assigned	Not assigned
9	M9A	Queen's Park	Not assigned

Folium helps to visualize the map and mark the location of each neighborhood on the map



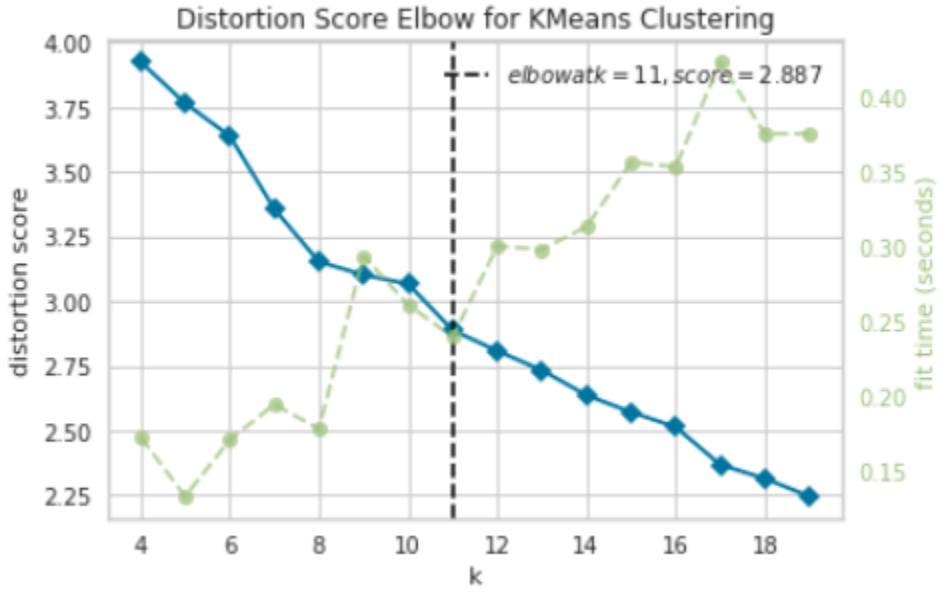
Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	100	100	100	100	100	100
Agincourt	46	46	46	46	46	46
Agincourt North, L'Amoreaux East, Milliken, Steeles East	29	29	29	29	29	29
Albion Gardens, Beaumond Heights, Humbergate, Jamestown, Mount Olive, Silverstone,	20	20	20	20	20	20

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
South Steeles, Thistletown						
Alderwood, Long Branch	26	26	26	26	26	26
Bathurst Manor, Downsview North, Wilson Heights	30	30	30	30	30	30
Bayview Village	14	14	14	14	14	14
Bedford Park, Lawrence Manor East	40	40	40	40	40	40
Berczy Park	100	100	100	100	100	100
Birch Cliff, Cliffside West	14	14	14	14	14	14
Bloordale Gardens, Eringate, Markland Wood, Old Burnhamthorpe	18	18	18	18	18	18
Brockton, Exhibition Place, Parkdale Village	100	100	100	100	100	100
Business Reply Mail Processing Centre 969 Eastern	50	50	50	50	50	50
CFB Toronto, Downsview East	21	21	21	21	21	21

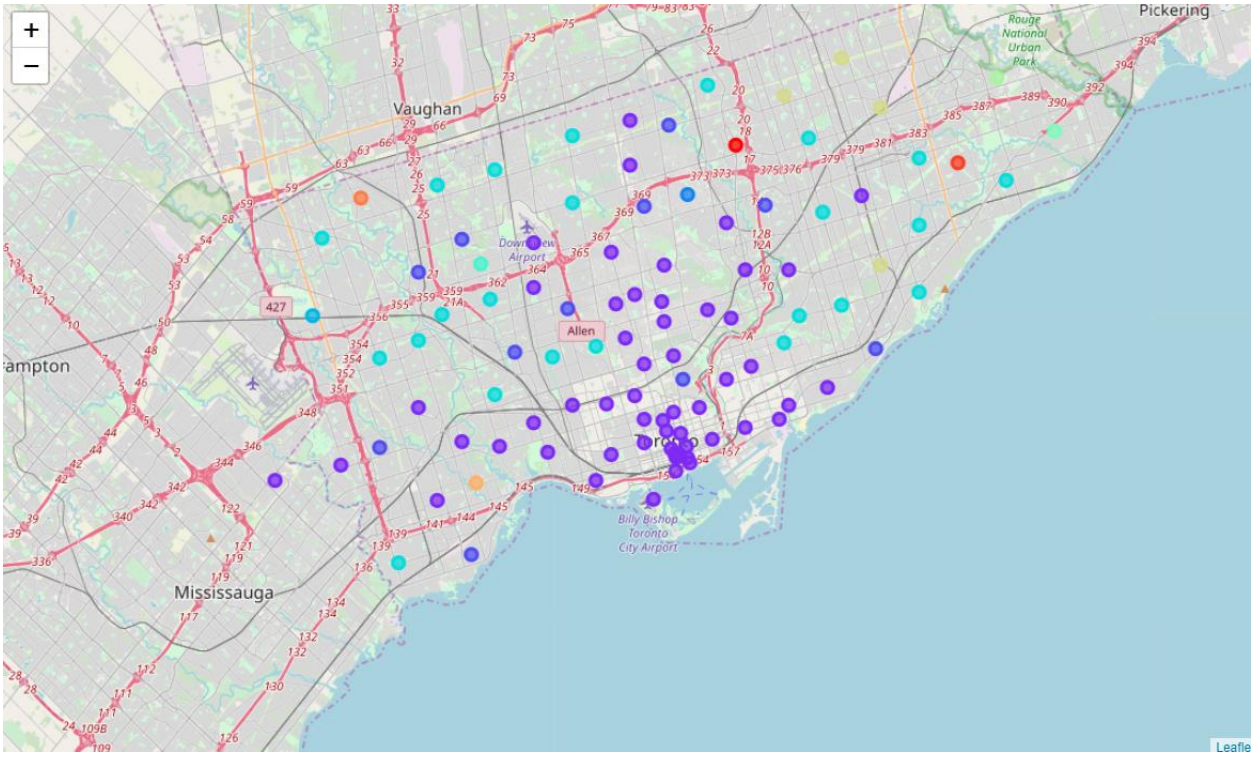
Querying Foursquare, the neighborhood data such as venue are collected and sort by common venues in the neighborhood

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Adelaide, King, Richmond	Café	Coffee Shop	Hotel	Theater	Restaurant	Sushi Restaurant
1	Agincourt	Chinese Restaurant	Shopping Mall	Caribbean Restaurant	Sandwich Place	Bakery	Shanghai Restaurant
2	Agincourt North, L'Amoreaux East, Milliken, St...	Chinese Restaurant	Bakery	Pizza Place	Noodle House	Park	Grocery Store
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Pizza Place	Grocery Store	Bus Line	Fried Chicken Joint	Beer Store	Sandwich Place
4	Alderwood, Long Branch	Discount Store	Pharmacy	Pizza Place	Sandwich Place	Skating Rink	Donut Shop
5		Pizza Place	Coffee Shop	Ski Area	Bathurst Manor, Downsview North, Wilson Heights	Sushi Restaurant	Fried Chicken Joint

Determine the best K value for the KMean Clustering method. Elbow method choose 11 as K.



The clusters by KMean method marked on the Toronto map.



Discussion

According to the K- mean Clustering and neighborhood analysis, there are some characteristic in some clusters.

For example, restaurants are easily found in Cluster 9 (Lime color)

Cluster 10 is featured with park and coffee shop and burger joints are located.

Conclusion

To achieve our goal to open new business, we can peak any cluster 9 for restaurant or cluster 10 for coffee shop.