

Capstone Project-The Battle of Neighborhoods

(Choosing the best neighborhood for a Japanese customer)

1. Introduction

1.1 Background

I am working in a real estate company in Toronto which provides property consulting services. Toronto has wide range of neighborhoods with different life styles. The company uses machine learning and data analytics techniques to provide customers who intends to buy or rent a property with the best choice based on their needs.

1.2 Business Problem

A Japanese customer decided to move from North York to Downtown Toronto due to change in his work location. He was living in Glencairn neighborhood in North York borough. He has the following requirements for his new neighborhood:

- The life style in this neighborhood is similar to the life style in Glencairn.
- The new neighborhood contains Japanese restaurants.

2. Data Acquisition and Cleaning

2.1 Data sources

The following data sources were used to obtain the necessary information for the analysis:

- The information of Toronto neighborhoods and boroughs were obtained from Wikipedia https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- The coordinates of Toronto neighborhoods were obtained using Geospatial dataset which can be obtained from http://cocl.us/Geospatial_data.

- Foursquare API was used to explore the available venues in each neighborhood in Toronto.

2.2 Data cleaning

Data was downloaded and scrapped from multiple sources to obtain the required information for the analysis and applying the clustering algorithm.

Beautiful Soap library was used for performing web scraping for Wikipedia site to obtain Toronto neighborhood and borough information. Data cleaning was done to convert the scrapped data into tabular DataFrame format. The neighborhoods having missing boroughs name were excluded from the analysis. The neighborhoods having the same postal code were combined together. Some neighborhoods were having missing values (unassigned), these neighborhoods were assigned the same name as the borough.

The below shows the Toronto borough and neighborhood information after cleaning

	Postalcode	Borough	Neighbourhood
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

	Postalcode	Borough	Neighbourhood	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

The venues within 2 KM radius in each neighborhood are obtained using Foursquare API

[33]:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood 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	KFC	43.754387	-79.333021	Fast Food Restaurant
2	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
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

The most 10 venue categories for each neighborhood are obtained to be able to cluster neighborhoods based on the similarities in the venue categories

	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Steakhouse	Thai Restaurant	Asian Restaurant	Burger Joint	Restaurant	Breakfast Spot	Hotel
1	Agincourt	Lounge	Furniture / Home Store	Breakfast Spot	Skating Rink	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Dog Run
2	Agincourt North, L'Amoreaux, East Milliken, St...	Park	Playground	Dog Run	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Doner Restaurant
3	Albion Gardens, Beaumont Heights, Humbergate, ...	Grocery Store	Fried Chicken Joint	Pharmacy	Coffee Shop	Pizza Place	Fast Food Restaurant	Sandwich Place	Beer Store	Dive Bar	Dessert Shop
4	Alderwood, Long Branch	Pizza Place	Coffee Shop	Skating Rink	Pharmacy	Pool	Pub	Sandwich Place	Gym	Airport Terminal	Doner Restaurant

Cluster Neighborhoods

3. Methodology

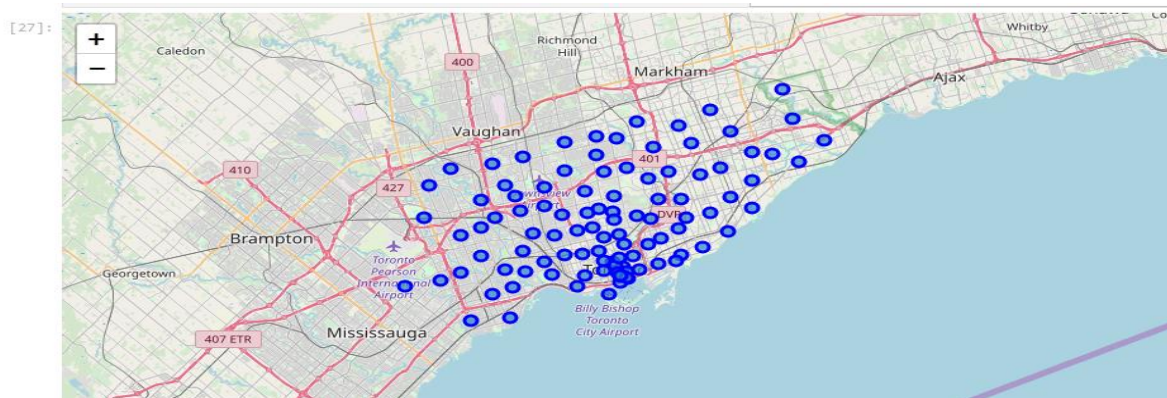
3.1 Summary

The first requirement of the Japanese customer is to live in a neighborhood in Downtown Toronto borough which has similar life style as Glencairn neighborhood in North York borough. This will be obtained by applying k-Means Clustering algorithm on Toronto neighborhoods to obtain the neighborhoods in Downtown Toronto which are in the same cluster as Glencairn.

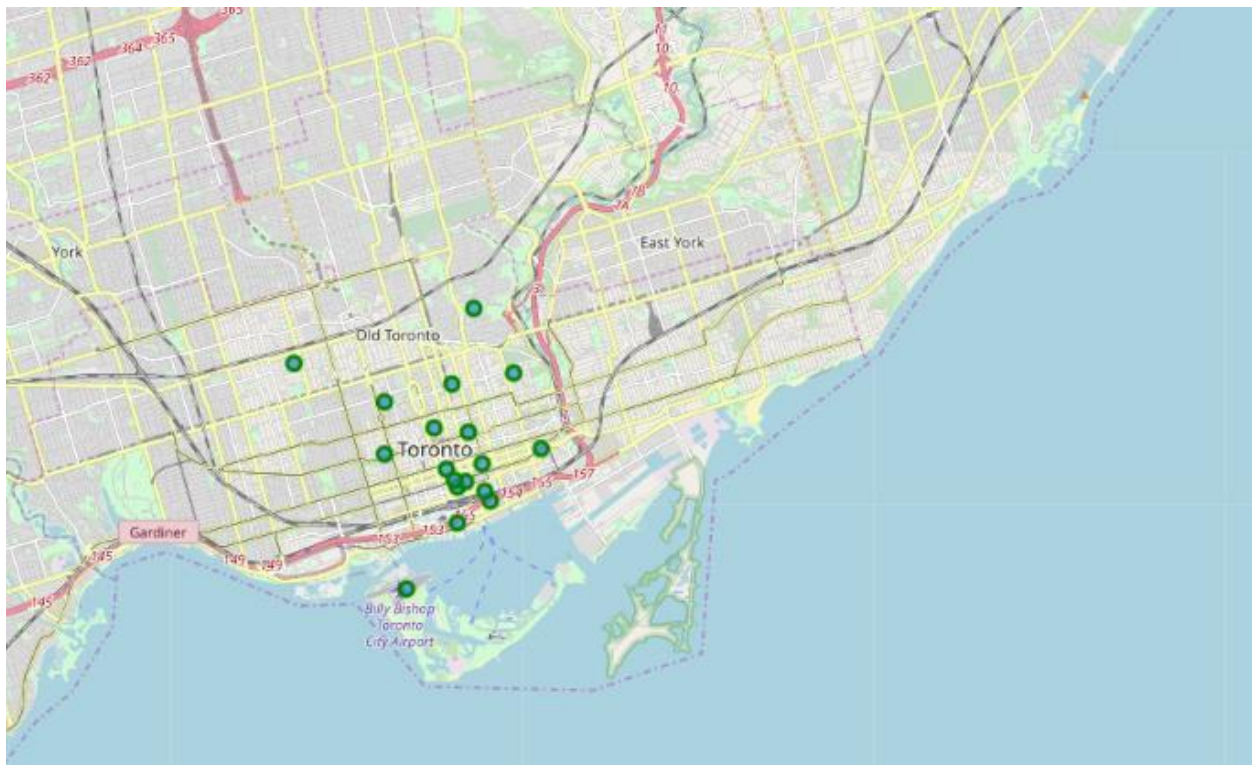
The second requirement of the Japanese customer is to live in a neighborhood which has Japanese restaurants. This will be obtained by filtering on the neighborhoods obtained from the first step which have Japanese restaurants and recommending the one with more Japanese restaurants to the customer.

3.2 Procedure

After obtaining Toronto neighborhood information, it was found that Toronto has 103 neighborhoods. These neighborhoods are plotted as shown below using Folium library.



Down Town Toronto is the target borough for the customer. It has 18 neighborhoods



Foursquare API was used to obtain the venues available for each neighborhood within 2 KM radius.

[33]:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood 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	KFC	43.754387	-79.333021	Fast Food Restaurant
2	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
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

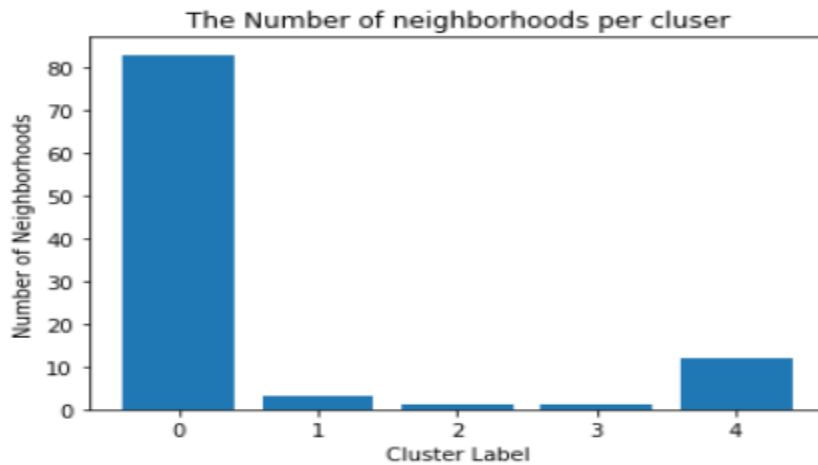
Using one hot encoding technique, the top 10 venue categories for each neighborhood was obtained to be used as an input for the clustering algorithm.

[48]:

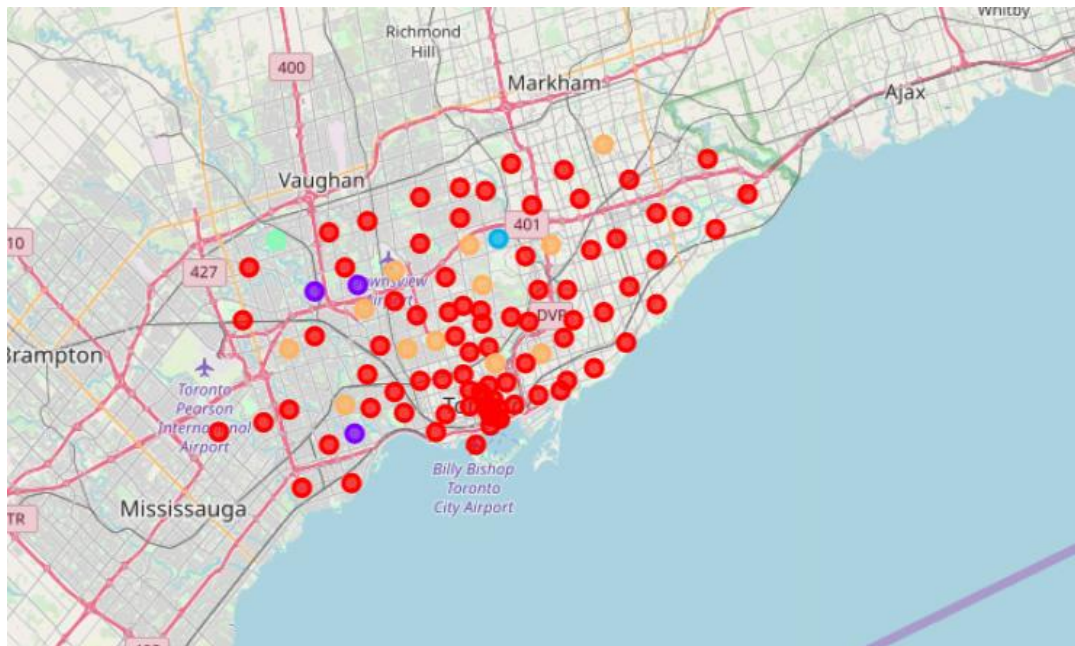
	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Steakhouse	Thai Restaurant	Asian Restaurant	Burger Joint	Restaurant	Breakfast Spot	Hotel
1	Agincourt	Lounge	Furniture / Home Store	Breakfast Spot	Skating Rink	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store	Olive Bar	Dog Run
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Dog Run	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Olive Bar	Doner Restaurant
3	Albion Gardens, Beaumont Heights, Humbergate, ...	Grocery Store	Fried Chicken Joint	Pharmacy	Coffee Shop	Pizza Place	Fast Food Restaurant	Sandwich Place	Beer Store	Olive Bar	Dessert Shop
4	Alderwood, Long Branch	Pizza Place	Coffee Shop	Skating Rink	Pharmacy	Pool	Pub	Sandwich Place	Gym	Airport Terminal	Doner Restaurant

K-means clustering is applied to divide Toronto neighborhoods into 5 clusters based on the venue categories.

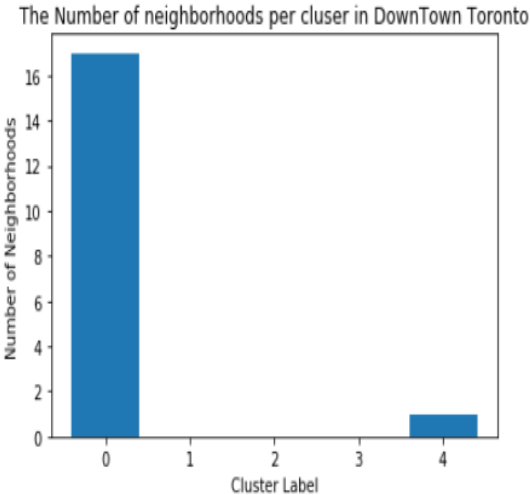
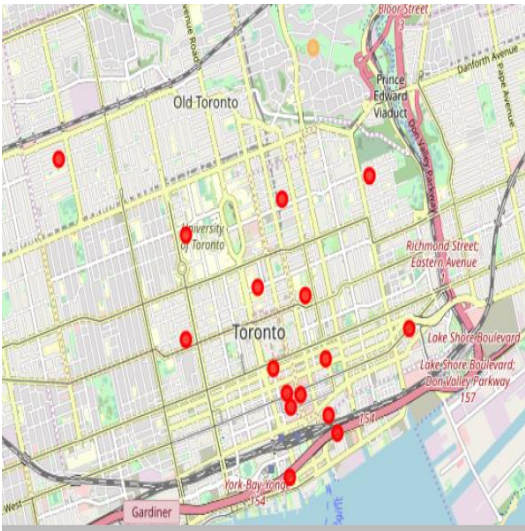
The below charts shows the number of neighborhoods for each cluster for Toronto neighborhoods.



The below map shows cluster for each Toronto neighborhoods.



When filtering on Downtown Toronto borough, it is found that 17 neighborhoods belong to cluster 0 while 1 cluster only belongs to cluster 4



4. Results

The cluster of Glencairn neighborhood is found to be cluster 0 and 17 clusters in Downtown Toronto are also in cluster 0, therefore 17 neighborhoods are candidate for the customer

	Postalcode	Borough	Neighbourhood	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
10	M6B	North York	Glencairn	43.709577	-79.445073	0.0	Park	Pub	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Dumpling Restaurant

Filtering is done to check which of the 17 candidate neighborhoods has Japanese restaurants in its common venues to achieve the customer requirement.

After performing filtering, it was found that 2 neighborhoods contain Japanese restaurants, Church and Wellesley neighborhood and Harbord, University of Toronto neighborhood.

[132]:

	Postalcode	Borough	Neighbourhood	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	7th Most Common Venue	8th Most Common Venue
80	M5S	Downtown Toronto	Harbord, University of Toronto	43.662696	-79.400049	0.0	Café	Bookstore	Bakery	Restaurant	Bar	Japanese Restaurant	Beer Bar	Beer
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	0.0	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Pub	Men's Store	Mediterranean Restaurant

Both neighborhoods are having almost same frequency of Japanese restaurants, therefore both can be good candidates for the customer.

----Church and Wellesley----

	venue	freq
0	Coffee Shop	0.07
1	Japanese Restaurant	0.06
2	Sushi Restaurant	0.05
3	Gay Bar	0.05
4	Restaurant	0.04
5	Pub	0.02
6	Fast Food Restaurant	0.02
7	Men's Store	0.02
8	Mediterranean Restaurant	0.02
9	Gastropub	0.02

----Harbord, University of Toronto----

	venue	freq
0	Café	0.09
1	Bar	0.06
2	Japanese Restaurant	0.06
3	Bookstore	0.06
4	Restaurant	0.06
5	Bakery	0.06
6	French Restaurant	0.03
7	Sandwich Place	0.03
8	Chinese Restaurant	0.03
9	Beer Store	0.03

5. Discussion

Toronto is a big city with 103 neighborhoods having different life styles and different venue categories.

Analysis was done to choose the best neighborhood for a Japanese customer moving to Downtown Toronto from North York using k-means clustering and data processing techniques. The Analysis included various data sources such Foursquare and Wikipedia to identify the best neighborhoods for the customer having same life style as his old home location Glencairn.

Focus was done on the top 10 venue categories in each neighborhood within 2 KM radius and clustering was done based on that

Map visualization techniques along charts were used for analysis of neighborhoods

6. Conclusion

In conclusion, Church and Wellesley neighborhood and Harbord, University of Toronto neighborhood are the best neighborhoods for the customer to relocate satisfying his requirements including the same life style as Glencairn and containing Japanese restaurants.

2 KM Radius and Top 10 venue category are considered as limitations for the analysis and could be reviewed in future studies.