The Battle of Neighborhoods

Jennifer Ng
April 2020

1. Introduction

1.1. Background

For many first-time restaurateurs, it all begins with an idea: A lively restaurant inspired by food eaten on a recent trip or a secret family recipe. But having a good idea or cooking skills are just a small part of surviving and thriving in an industry notorious for high competition and failure rate. Among the many considerations, location can make or break a restaurant. Being able to recommend good locations for opening a certain type of restaurant is a functionality that would be very useful for new restaurant owners. In this paper, I aim to build a model that recommends locations to a restaurateur looking to open a Japanese sushi restaurant in the Toronto area.

1.2. Problem Description

The objective of this Capstone project is to provide a 'Yes/No' recommendation on Toronto neighborhoods based on whether they are ideal for a Japanese sushi restaurant. By using data science and machine learning methods such as clustering, this project aims at providing solutions to answer the business question: in Toronto, if an entrepreneur wants to open a Japanese sushi restaurant, where should he/she consider?

Several factors need to be studied to recommend a location:

- Whether location is a prime location.
- Restaurant competition in the neighborhoods – are there any existing Japanese or Sushi restaurants nearby?
- Would a Japanese restaurant "fit in" with the location – e.g. we wouldn't want to open a Japanese restaurant next to a freeway, rail, airport or around fast food restaurants.
- Nearby attractions – how accessible is it to potential customers.

The problem is modeled to predict a yes/no recommendation based on the above factors.

1.3. Target Audience

Entrepreneurs looking to open a Japanese sushi restaurant in Toronto, Canada

1.4. Success Criteria

The goal of the project is to create an observation table and a visualization map showing all Toronto neighborhoods together with a score and 'Recommend: Yes/No' next to each neighborhood. The success criteria are to provide a good recommendation on neighbourhood choices for opening a Japanese sushi restaurant.

2. Data Acquisition and Cleaning

2.1. Data sources

The data used in this project was obtained from the below sources-

- List of postal codes, boroughs and neighborhoods in Toronto, Canada obtained by scraping **Wikipedia** at the below URL: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Latitude and Longitude of these neighbourhoods from **Geospacial_Coordinates.csv**: http://cocl.us/Geospatial_data/Geospatial_Coordinates.csv
- Venues data of these neighborhoods using the **Foursquare API**. Foursquare API provides access to an enormous database consisting of venues from all around the world including their categories, addresses, tips, photos and comments. List of end-points: https://developer.foursquare.com/docs/places-api/endpoints/.

2.2. Data cleaning

Postal code table data was scraped from Wikipedia to create a dataframe consisting of Postal Code, Borough and Neighborhood.

However, some data cleansing was required as it had a few values under the column 'Borough' that were 'Not Assigned'. To correct this, I dropped any rows with a 'Not Assigned' value in the 'Borough' column. Another problem was that there were a few rows that had a borough but a 'Not Assigned' value in the "Neighborhood" column. For these cases, I set 'Neighborhood' to be the same as 'Borough'. To ensure there are no duplicate rows, I combined rows with the same 'PostalCode' by concatenating neighborhoods by a comma. There were also new line characters in the Postal Code column that had to be removed.

After fixing these problems, I verified the data to ensure that all the cells had correct values and formats.

2.3. Feature selection

After data cleansing, I combined the neighborhoods dataframe (PostalCode, Borough, Neighborhood) with the geospatial dataframe to add Latitude and Longitude columns. Since the primary focus was Toronto city, I filtered the data to work with only boroughs containing the word 'Toronto'.

The next step was to get venues information for all the neighborhoods using the Foursquare API. After signing up for a Foursquare developer account, using the Client ID and Client Secret, I made API requests to retrieve venue information. From the output received, I selected Venue name, Venue Latitude, Venue Longitude and Venue Category.

I then performed a One-Hot encoding on ‘Venue Category’ and grouped the rows by ‘Neighborhood’. The result was a dataframe that had 235 features (Neighborhood, Venue Categories as columns)

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Arts Cra Sto

3. Methodology

3.1. Exploratory Data Analysis

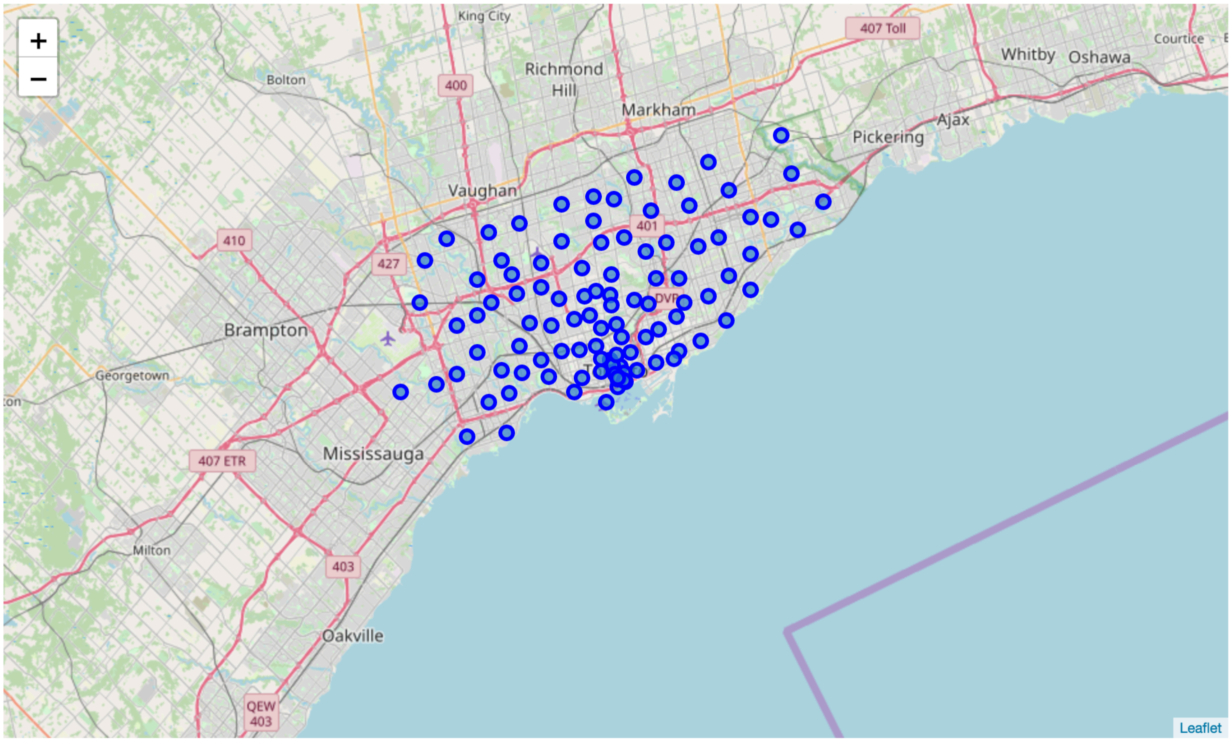
First, I needed to get a list of neighborhoods in Toronto, Canada. I did this by scraping Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) using the **BeautifulSoup** library. I then transformed the data into a *pandas* dataframe with 3 columns – PostalCode, Borough and Neighborhood. After cleansing the data, I obtained a dataframe of 103 rows. Below is a sample showing the first 5 rows-

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

The next step was to get geospatial data. Initially, I tried using Python **geocoder** library and built a function that returns latitude and longitude, given a postal code. I then ran a loop on all postal codes to call the function. However, I hit some issues as the Python geocoder library is unreliable – this is a known issue. So, I used an alternate data source, which is the Geospacial_Coordinates.csv. I joined the neighborhoods dataframe with the geoCoordinates dataframe. First 5 rows are shown below-

	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	Regent Park / Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494

I used Python **folium** library to create a map and visualize all neighborhoods in Toronto.



I decided to work with only boroughs containing the word Toronto, to reduce the dataset and focus on Toronto city. The filtering resulted in **39 postal codes**. Let's name this dataframe **toronto_data**. Below are the first 5 rows-

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

Now that I had all Toronto neighborhoods and their geo coordinates, the next step was to explore the neighborhoods. I used the **Foursquare API** to get venues information. I created a function that takes neighborhood names, latitudes, longitudes, radius and returns a dataframe of nearby venues. The function loops through all neighborhoods and makes a call to the foursquare */venues/explore* end-point, passing client ID, client secret, version, latitude, longitude, radius (default 500 metres) and limit (100 venues). The resulting dataframe after calling the function-

```
print(toronto_venues.shape)
toronto_venues.head()
```

```
(1661, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park / Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park / Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park / Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park / Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park / Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot

Running some analysis on the data to better a better understanding of the venues counts per neighborhood-

```
toronto_venues.groupby(['Neighborhood']).count()
```

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Berczy Park	56	56	56	56	56	56
Brockton / Parkdale Village / Exhibition Place	22	22	22	22	22	22
Business reply mail Processing CentrE	17	17	17	17	17	17
CN Tower / King and Spadina / Railway Lands / Harbourfront West / Bathurst	16	16	16	16	16	16
Central Bay Street	73	73	73	73	73	73
Christie	17	17	17	17	17	17
Church and Wellesley	80	80	80	80	80	80
Commerce Court / Victoria Hotel	100	100	100	100	100	100
Davisville	35	35	35	35	35	35
Davisville North	9	9	9	9	9	9
Dufferin / Dovercourt Village	16	16	16	16	16	16
First Canadian Place / Underground city	100	100	100	100	100	100
Forest Hill North & West	4	4	4	4	4	4
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East / Union Station / Toronto Islands	100	100	100	100	100	100
High Park / The Junction South	24	24	24	24	24	24

```
print('There are {} uniques categories.'.format(len(toronto_venues['Venue Category'].unique())))
```

```
There are 235 uniques categories.
```

I then performed a One-Hot encoding on ‘Venue Category’. Since we have 1661 rows and 235 venue categories, the shape of the dataframe was (1661, 235). I then grouped the rows by ‘Neighborhood’ by doing a mean(), which resulted in a dataframe of **38 rows x 235 columns** *(Note – Out of the 39 postal codes, one of them did not have venues information so, got dropped)*. Below are the first 5 rows-

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Arts Cra Sto
0	Berczy Park	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.017857	0.0(
1	Brockton / Parkdale Village / Exhibition Place	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.000000	0.0(
2	Business reply mail Processing CentrE	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.000000	0.0(
3	CN Tower / King and Spadina / Railway Lands / ...	0.0000	0.0625	0.0625	0.125	0.1875	0.125	0.000000	0.000000	0.00	0.000000	0.0(
4	Central Bay Street	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.013699	0.000000	0.00	0.000000	0.0(

```
toronto_grouped.shape
```

```
(38, 235)
```

The resulting dataframe was then used to perform the clustering operation.

3.2. Predictive Modeling

Given that neighborhoods share similarities, I used the **K-means algorithm** to segment and cluster the neighborhoods. K-Means is one of the most popular unsupervised machine learning algorithms.

First, I set the number of centroids k = 5. I dropped the Neighborhood column, then ran KMeans with a random_state of 0 to fit our dataframe (last dataframe obtained in section 4.1)

The result of KMeans is an array of labels generated for each neighborhood-

```python
# set number of clusters
kclusters = 5

toronto_grouped_clustering = toronto_grouped.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_[0:10]
```

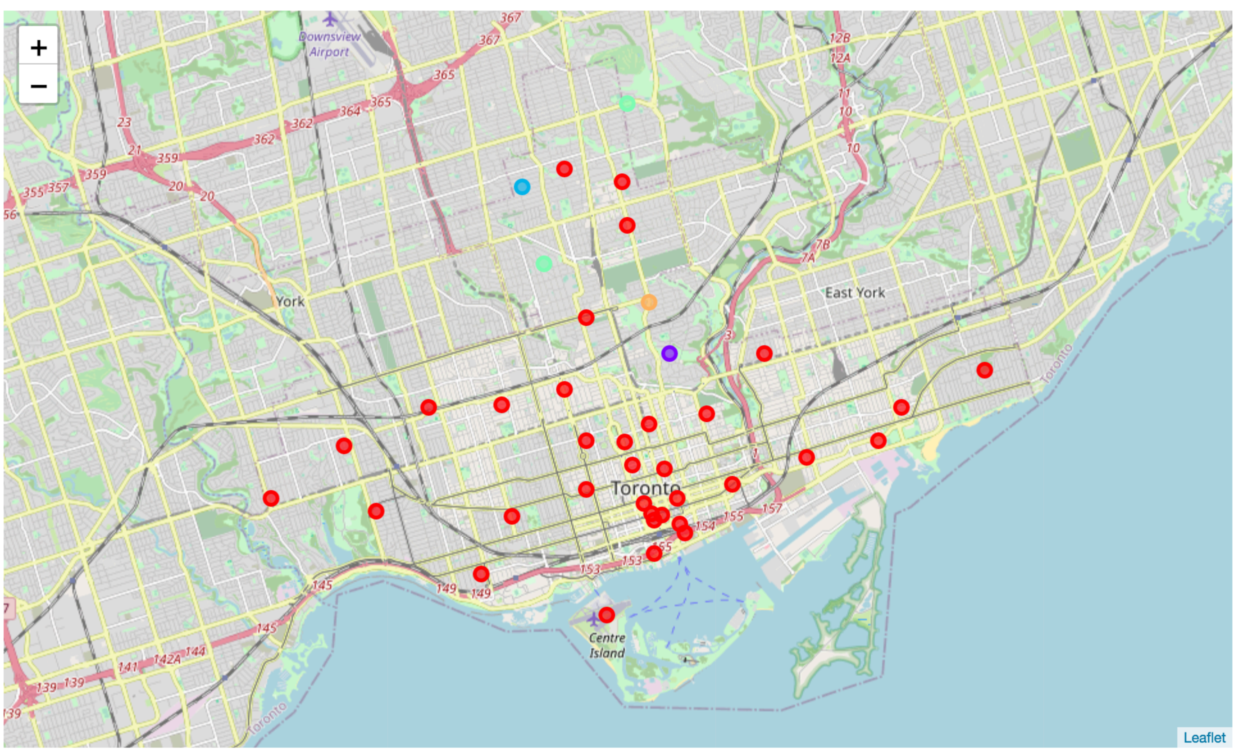
```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int32)
```

4. Results

After we obtained the cluster labels, we start preparing a result or observation table. We merged the toronto_data with venues data (apply sorting to get top 10 venues), and inserted Cluster Label as column-

	PostalCode	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636	0	Coffee Shop	Park	Bakery	Pub	Mexican Restaurant	Breakfast Spot	Café	Restaurant	Theater	Shoe Store
1	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Diner	Fried Chicken Joint	Burrito Place	Juice Bar	Sandwich Place	Café	Italian Restaurant	Beer Bar	Mexican Restaurant
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Japanese Restaurant	Italian Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Bookstore	Restaurant
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Café	Cocktail Bar	Restaurant	Beer Bar	American Restaurant	Hotel	Italian Restaurant	Seafood Restaurant	Park
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	Pub	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

We then used folium map to visualize the results-



Now, in the end result, we needed to provide a 'Recommend: Yes/No' next to each neighborhood. To do so, we applied a few criteria or factors (mentioned in section 1.2 of the report) to further refine the results.

Criteria 1 – Determine whether location is a prime location

Here, we identified the cluster with the maximum number of venues – this was cluster 0. If a neighborhood belonged to cluster 0, we populated the column '**PrimeLocation'** with a score of 1. Otherwise, we gave it a score of 0. Result after applying criteria 1 (first 5 rows displayed)-

	Neighborhood	Cluster Labels	PrimeLocation
0	Regent Park / Harbourfront	0	1
1	Queen's Park / Ontario Provincial Government	0	1
2	Garden District, Ryerson	0	1
3	St. James Town	0	1
4	The Beaches	0	1
5	Berczy Park	0	1
6	Central Bay Street	0	1
7	Christie	0	1
8	Richmond / Adelaide / King	0	1
9	Dufferin / Dovercourt Village	0	1
10	Harbourfront East / Union Station / Toronto Is...	0	1
11	Little Portugal / Trinity	0	1
12	The Danforth West / Riverdale	0	1
13	Toronto Dominion Centre / Design Exchange	0	1
14	Brockton / Parkdale Village / Exhibition Place	0	1
15	India Bazaar / The Beaches West	0	1
16	Commerce Court / Victoria Hotel	0	1
17	Studio District	0	1
18	Lawrence Park	3	0
19	Roselawn	2	0

Criteria 2 – Check Competition in the neighborhoods

We then checked whether there were any existing 'Japanese Restaurant' or 'Sushi Restaurant' in the neighborhoods. We wouldn't want to open our restaurant in a neighborhood that already has a Japanese or sushi restaurant due to the competition. We populated a column '**LowCompetition'** with a score of 0 or 1 (0 – if there is competition; 1 – if low competition). Result after applying criteria 2 (first 5 rows displayed)-

	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	LowCompetition
0	Regent Park / Harbourfront	Coffee Shop	Park	Bakery	Pub	Mexican Restaurant	Breakfast Spot	Café	Restaurant	Theater	Shoe Store	1
1	Queen's Park / Ontario Provincial Government	Coffee Shop	Diner	Fried Chicken Joint	Burrito Place	Juice Bar	Sandwich Place	Café	Italian Restaurant	Beer Bar	Mexican Restaurant	1
2	Garden District, Ryerson	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Japanese Restaurant	Italian Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Bookstore	Restaurant	0
3	St. James Town	Coffee Shop	Café	Cocktail Bar	Restaurant	Beer Bar	American Restaurant	Hotel	Italian Restaurant	Seafood Restaurant	Park	1
4	The Beaches	Health Food Store	Trail	Pub	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	1
5	Berczy Park	Coffee Shop	Farmers Market	Cheese Shop	Beer Bar	Bakery	Café	Restaurant	Italian Restaurant	Seafood Restaurant	Cocktail Bar	1
6	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Spa	Ice Cream Shop	Japanese Restaurant	Sushi Restaurant	Bubble Tea Shop	Salad Place	0

Note – As seen in the above table, Neighborhood 6 – Central Bay Street has a Japanese Restaurant already (also has a Sushi restaurant), so we gave it a score of 0.

<u>Criteria 3 – Would a Japanese restaurant "fit in" with the location?</u>

Our Japanese restaurant requires a more upscale restaurant, and requires time for dine-ins customers. Hence, a location next to a freeway, railway station, airport (where people are in a rush) or around fast food restaurants would not be appropriate, So, we populated a column called '**LocationFit**' by giving a score of 0 or 1 (0 – if there are unwanted venues; 1 – otherwise). Result after applying criteria 3 (first 5 rows displayed)-

	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	LocationFit
0	Regent Park / Harbourfront	Coffee Shop	Park	Bakery	Pub	Mexican Restaurant	Breakfast Spot	Café	Restaurant	Theater	Shoe Store	1
1	Queen's Park / Ontario Provincial Government	Coffee Shop	Diner	Fried Chicken Joint	Burrito Place	Juice Bar	Sandwich Place	Café	Italian Restaurant	Beer Bar	Mexican Restaurant	1
2	Garden District, Ryerson	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Japanese Restaurant	Italian Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Bookstore	Restaurant	1
3	St. James Town	Coffee Shop	Café	Cocktail Bar	Restaurant	Beer Bar	American Restaurant	Hotel	Italian Restaurant	Seafood Restaurant	Park	1
4	The Beaches	Health Food Store	Trail	Pub	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	1
5	Berczy Park	Coffee Shop	Farmers Market	Cheese Shop	Beer Bar	Bakery	Café	Restaurant	Italian Restaurant	Seafood Restaurant	Cocktail Bar	1
6	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Spa	Ice Cream Shop	Japanese Restaurant	Sushi Restaurant	Bubble Tea Shop	Salad Place	1

For example, the below neighborhood has a Fast Food Restaurant in one of its top 10 venues – this wouldn't be an appropriate location so we gave it a score of 0-

15	India Bazaar / The Beaches West	Park	Fast Food Restaurant	Sushi Restaurant	Pizza Place	Pub	Restaurant	Italian Restaurant	Ice Cream Shop	Movie Theater	Sandwich Place	0

<u>Criteria 4 – Nearby attractions – How accessible is it to potential customers</u>

Next, we checked whether there are parks, hotels and tourist attractions – these are signs that the neighborhood is busy and accessible. We populated a column called '**LocationAccessible**' by giving a score of 0 or 1 (1 - if there are desirable venues; 0 otherwise). Result after applying criteria 4 (first 5 rows displayed)-

	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	LocationAccessible
0	Regent Park / Harbourfront	Coffee Shop	Park	Bakery	Pub	Mexican Restaurant	Breakfast Spot	Café	Restaurant	Theater	Shoe Store	1
1	Queen's Park / Ontario Provincial Government	Coffee Shop	Diner	Fried Chicken Joint	Burrito Place	Juice Bar	Sandwich Place	Café	Italian Restaurant	Beer Bar	Mexican Restaurant	0
2	Garden District, Ryerson	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Japanese Restaurant	Italian Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Bookstore	Restaurant	0
3	St. James Town	Coffee Shop	Café	Cocktail Bar	Restaurant	Beer Bar	American Restaurant	Hotel	Italian Restaurant	Seafood Restaurant	Park	1
4	The Beaches	Health Food Store	Trail	Pub	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	1
5	Berczy Park	Coffee Shop	Farmers Market	Cheese Shop	Beer Bar	Bakery	Café	Restaurant	Italian Restaurant	Seafood Restaurant	Cocktail Bar	0
6	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Spa	Ice Cream Shop	Japanese Restaurant	Sushi Restaurant	Bubble Tea Shop	Salad Place	0

FINAL RESULT – Putting it all together

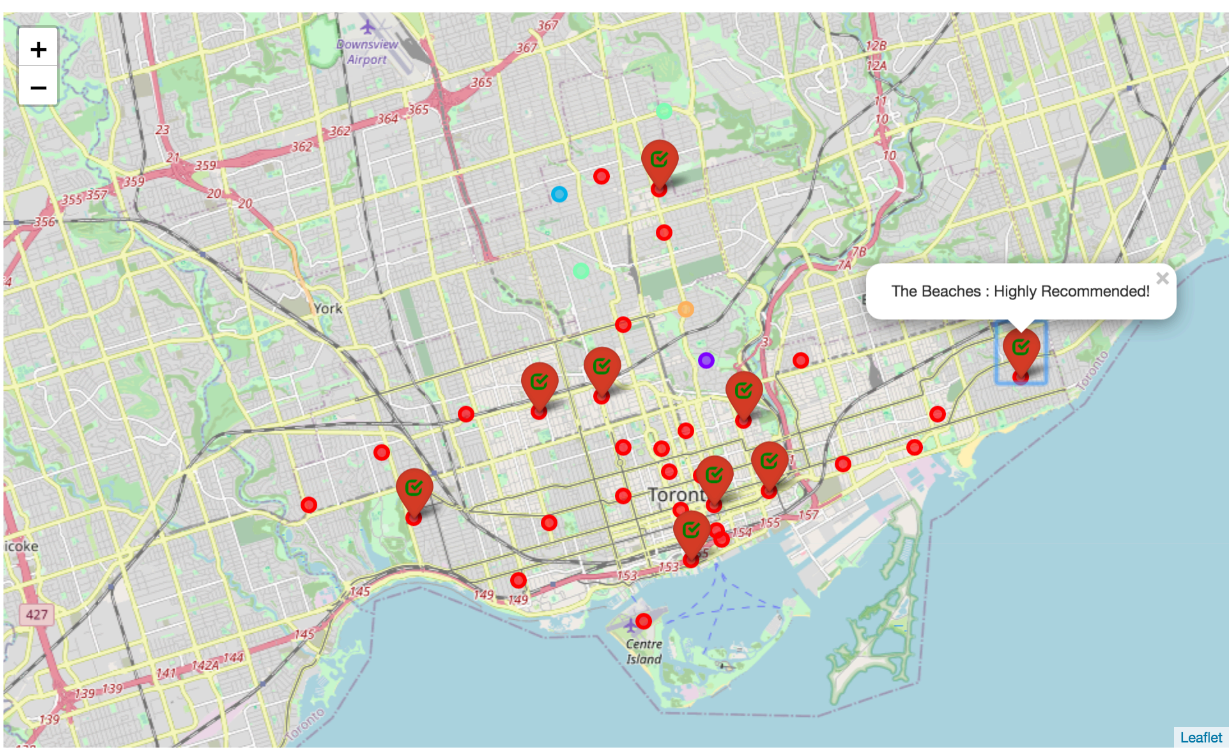
Finally, we added up the scores of the four columns. We populated the 'Recommend: Yes/No' column with a 'Yes' where the total score is 4. If score is lower than 4, we gave a 'No' recommendation. Result of final observation table (first 5 rows displayed)-

	PostalCode	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	PrimeLocation	LowCompetition	LocationFit	LocationAccessible	Score	Recommend: Yes/No
0	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636	0	Coffee Shop	Park	Bakery	Pub	Mexican Restaurant	Breakfast Spot	Café	Restaurant	Theater	Shoe Store	1	1	1	1	4	Yes
1	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Diner	Fried Chicken Joint	Burrito Place	Juice Bar	Sandwich Place	Café	Italian Restaurant	Beer Bar	Mexican Restaurant	1	1	1	0	3	No
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Japanese Restaurant	Italian Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Bookstore	Restaurant	1	0	1	0	2	No
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Café	Cocktail Bar	Restaurant	Beer Bar	American Restaurant	Hotel	Italian Restaurant	Seafood Restaurant	Park	1	1	1	1	4	Yes
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	Pub	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	1	1	1	1	4	Yes

Filtering by Recommended: Yes/No equal to 'Yes', we ended up with 9 neighborhoods with a strong recommendation-

	PostalCode	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	PrimeLocation	LowCompetition	LocationFit	LocationAccessible	Score	Recommend: Yes/No
0	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636	0	Coffee Shop	Park	Bakery	Pub	Mexican Restaurant	Breakfast Spot	Café	Restaurant	Theater	Shoe Store	1	1	1	1	4	Yes
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Café	Cocktail Bar	Restaurant	Beer Bar	American Restaurant	Hotel	Italian Restaurant	Seafood Restaurant	Park	1	1	1	1	4	Yes
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	Pub	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	1	1	1	1	4	Yes
7	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0	Grocery Store	Café	Park	Coffee Shop	Nightclub	Italian Restaurant	Candy Store	Baby Store	Athletics & Sports	Restaurant	1	1	1	1	4	Yes
10	M5J	Downtown Toronto	Harbourfront East / Union Station / Toronto Is...	43.640816	-79.381752	0	Coffee Shop	Aquarium	Hotel	Restaurant	Italian Restaurant	Café	Brewery	Sporting Goods Shop	Scenic Lookout	Fried Chicken Joint	1	1	1	1	4	Yes
20	M4P	Central Toronto	Davisville North	43.712751	-79.390197	0	Park	Breakfast Spot	Gym	Food & Drink Shop	Sandwich Place	Department Store	Hotel	Discount Store	Ethiopian Restaurant	Electronics Store	1	1	1	1	4	Yes
24	M5R	Central Toronto	The Annex / North Midtown / Yorkville	43.672710	-79.405678	0	Café	Sandwich Place	Coffee Shop	Pharmacy	Indian Restaurant	BBQ Joint	History Museum	Pizza Place	Pub	American Restaurant	1	1	1	1	4	Yes
25	M6R	West Toronto	Parkdale / Roncesvalles	43.648960	-79.456325	0	Gift Shop	Bookstore	Movie Theater	Eastern European Restaurant	Italian Restaurant	Dog Run	Bar	Bank	Restaurant	Dessert Shop	1	1	1	1	4	Yes
35	M4X	Downtown Toronto	St. James Town / Cabbagetown	43.667967	-79.367675	0	Coffee Shop	Café	Restaurant	Italian Restaurant	Pub	Bakery	Pizza Place	Park	Sandwich Place	Butcher	1	1	1	1	4	Yes

We used folium map again to visualize all the neighborhoods and clusters, but adding an additional green check marker to highlight neighborhoods that received a 'Yes' recommendation.



5. Discussion

I used the KMeans algorithm as part of this clustering study. Since this is an unsupervised clustering case, many different approaches can be adopted in order to achieve better results. One improvement area is to try different unsupervised clustering algorithms to determine which one gives the best accuracy.

There is also scope to explore the demographics of Toronto, to find details about population count, income or ethnicity by neighborhood. Unfortunately, I found challenges in getting this data as Toronto census data is not only old (dates back from 2016), but is not available by postal code or neighborhood. Data is the key challenge in all data science projects.

The project was done using postal codes of Toronto each having 235 features. Having more samples may result in a better clustering.

6. Conclusion

The restaurant business is a highly competitive and difficult business, with location being a key aspect of restaurant success. Being able to recommend locations for opening a Japanese restaurant can be a very useful tool for first time restaurateurs. The use of geospatial data and application of neighborhood segmentation and clustering lies beyond this application. This project can serve as a tool to recommend locations for any type of business. Furthermore, the approaches and methodologies used can be applied to build other recommender systems – for example, to match venues with customers when combined with customer preferences and venue ratings data.