Capstone Project - The Battle of Neighborhoods

Selecting the best location to open an SUSHI BAR IN Toronto, Canada

Ying

Coursera applied Data Science Capstone

CONTENTS

Introduction

- Toronto is the provincial capital of Ontario and the most populous city in Canada. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk of starting a new business. And the Return on Investment will be reasonable.

Business Problem

- The cuisine of Toronto reflects Toronto's size and multicultural diversity. Let’s say one of my friends, Lily, wants to open a sushi restaurant in Toronto. Starting a sushi restaurant can be a great business opportunity.
- New sushi bar should be open in an area that inadequate neighborhood in this way the bar can attract more customers. Therefore, this analysis necessary to ensure that we have enough customers and that we are not so close to other sushi places.

Data

- Neighborhood has a total of 10 boroughs and 103 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. The data is from our week3 assignment, I will export the data as csv file for future analysis.

```
[5]: data = pd.read_csv("Toront data.csv",header=0,index_col=0)

     print('Data downloaded!')

     data.head(5)
```

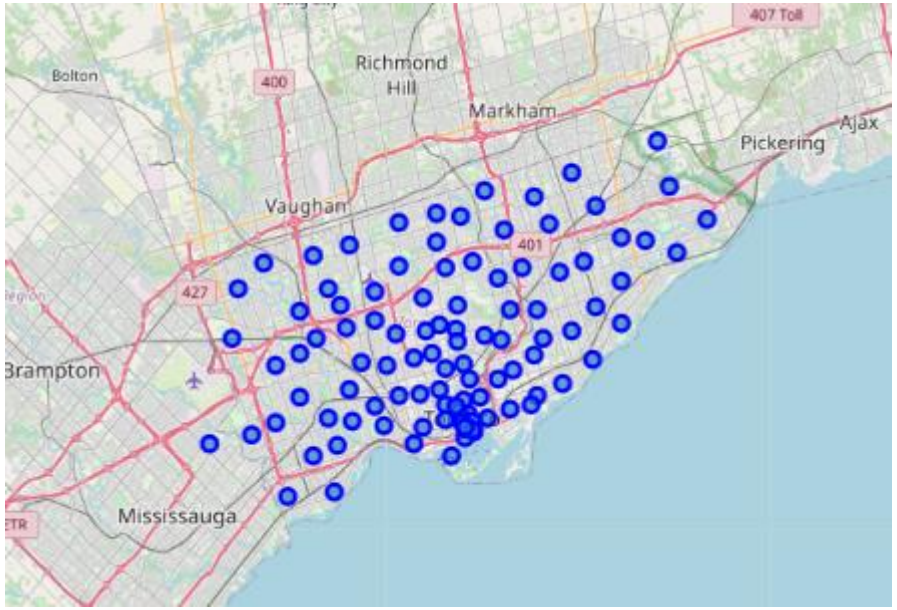
```
Data downloaded!
```

[5]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

```
[74]: data.shape
```

```
[74]: (103, 5)
```



Data

- Now once we got the neighborhood's latitude and longitude, let's use Foursquare Location to get the Mission District Neighborhood's all restaurant details. The restaurant details can be retrieved using search endpoint. Sushi category Id 4bf58dd8d48988d1d2941735 is used for retrieving data from Foursquare API.

```
[10]: #https://developer.foursquare.com/docs/resources/categories
      #Sushi = 4bf58dd8d48988d1d2941735
      Toronto_venues_sushi = getNearbyVenues(names=data['Neighbourhood'], latitudes=data['Latitude'], longitudes=data['Longitude'], radius=1000, categoryIds='4bf5
      Toronto_venues_sushi.head()
```

[10]:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cedarbrae	43.773136	-79.239476	HO Sushi Express	43.780965	-79.247445	Sushi Restaurant
1	Cedarbrae	43.773136	-79.239476	Fukuoka Sushi	43.780679	-79.247320	Sushi Restaurant
2	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	Ikki sushi	43.733489	-79.257644	Sushi Restaurant
3	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304	Shiro Sushi	43.756228	-79.266965	Japanese Restaurant
4	Agincourt	43.794200	-79.262029	Sushi Legend	43.796602	-79.270292	Sushi Restaurant

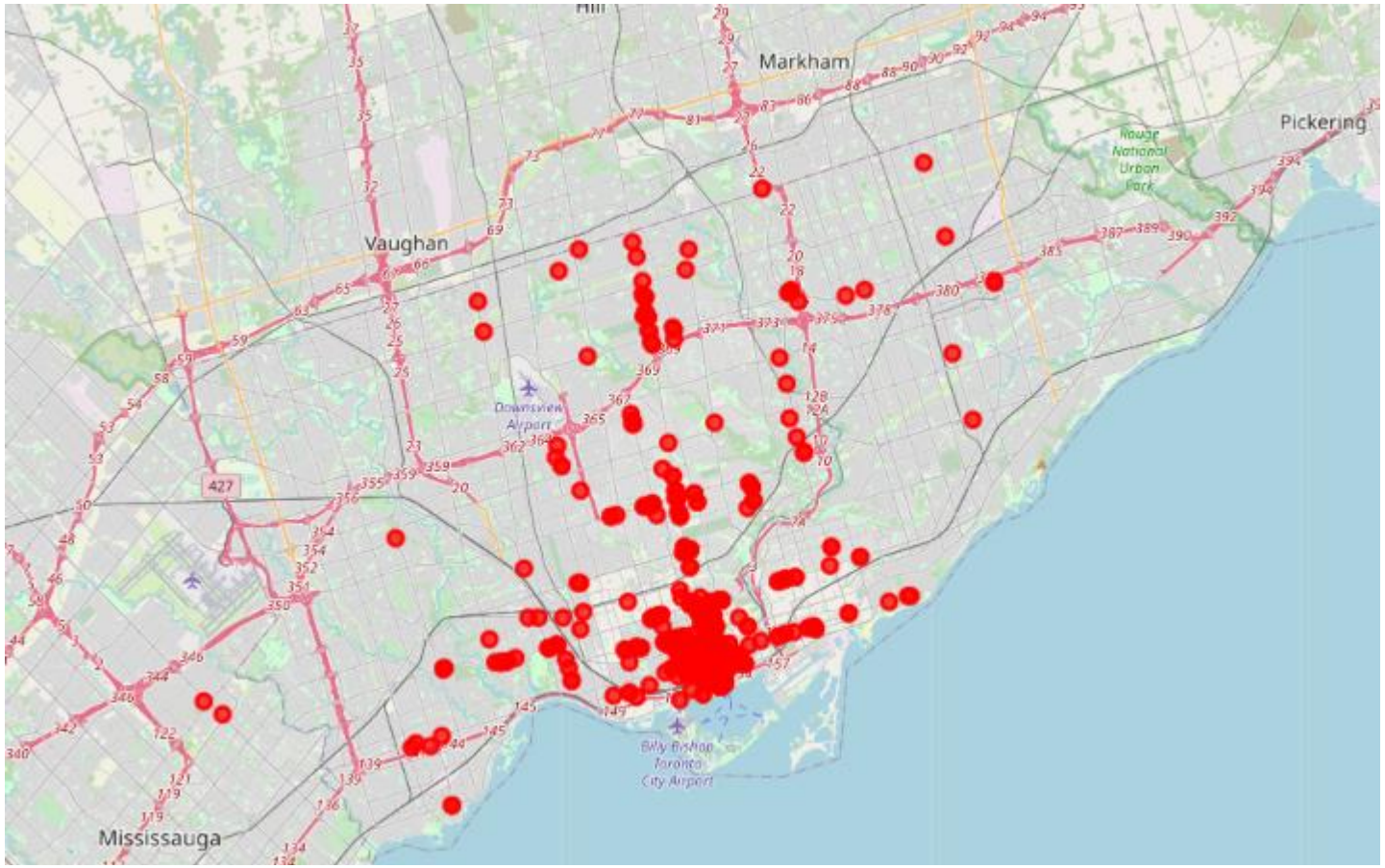
```
[11]: Toronto_venues_sushi.to_csv("Toronto_venues_sushi.csv")
```

```
[12]: Toronto_venues_sushi.shape
```

```
[12]: (778, 7)
```

Analysis

- Sushi restaurant in Toront



Analysis

- All sushi restaurants are grouped by neighborhoods and analyzed based on categories.

```
neighborhoods_venues_sorted.head()
```

[36]:

	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	Agincourt	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
1	Bathurst Manor, Wilson Heights, Downsview North	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
2	Bayview Village	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
3	Bedford Park, Lawrence Manor East	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
4	Berczy Park	Sushi Restaurant	Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant

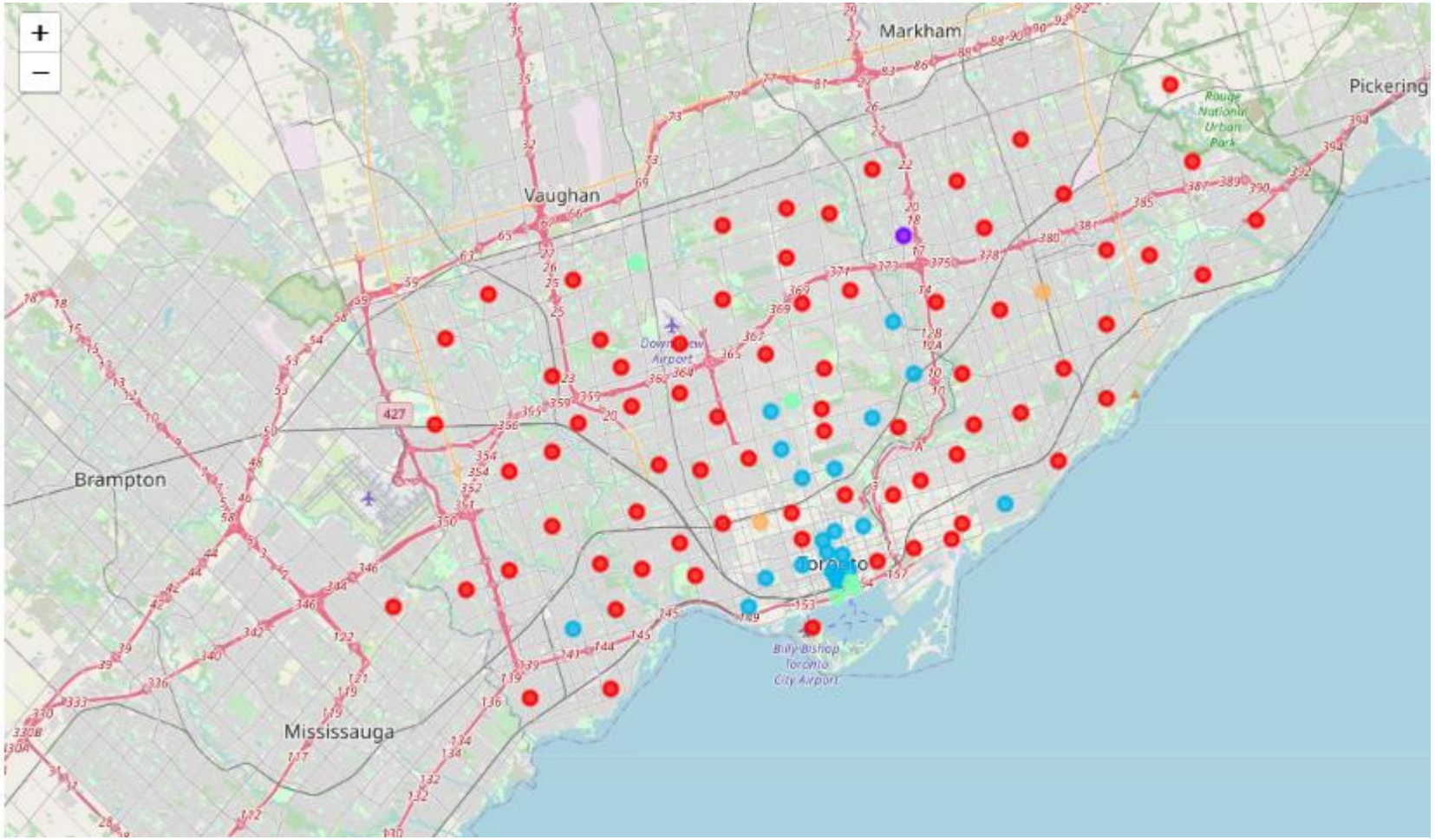
Analysis

- The neighborhoods are grouped into clusters using 1st common venue. K-means clustering algorithm will be use to complete this task.

	Postcode	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	9th Most Common Venue	Co
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	0	0	0	0	0	0	0	0	0	0	
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0	0	0	0	0	0	0	0	0	0	
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	0	0	0	0	0	0	0	0	0	0	
3	M1G	Scarborough	Woburn	43.770992	-79.216917	0	0	0	0	0	0	0	0	0	0	
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Res
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0	0	0	0	0	0	0	0	0	0	
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	0	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Res
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577	0	0	0	0	0	0	0	0	0	0	

Analysis

- The Folium library to visualize the neighborhoods in Torontos and its emerging clusters.



Results

- Using K-mean to clustering data area with less number of sushi bars.
  - Cluster 0

```
print('There are {} uniques cluster labels.'.format(len(toronto_merged['Cluster Labels'].unique())))
```

```
There are 6 uniques cluster labels.
```

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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	Scarborough	0	0	0	0	0	0	0	0	0	0	0
1	Scarborough	0	0	0	0	0	0	0	0	0	0	0
2	Scarborough	0	0	0	0	0	0	0	0	0	0	0
3	Scarborough	0	0	0	0	0	0	0	0	0	0	0
4	Scarborough	0	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
5	Scarborough	0	0	0	0	0	0	0	0	0	0	0
6	Scarborough	0	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
7	Scarborough	0	0	0	0	0	0	0	0	0	0	0
8	Scarborough	0	0	0	0	0	0	0	0	0	0	0
9	Scarborough	0	0	0	0	0	0	0	0	0	0	0
11	Scarborough	0	0	0	0	0	0	0	0	0	0	0
12	Scarborough	0	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant

Results

- Using K-mean to clustering data area with less number of sushi bars.
  - Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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
18	North York	1	Japanese Restaurant	Sushi Restaurant	Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant

Results

- Using K-mean to clustering data area with less number of sushi bars.
  - Cluster 2

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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
26	North York	2	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
27	North York	2	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
37	East Toronto	2	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
38	East York	2	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
48	Central Toronto	2	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
49	Central Toronto	2	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
51	Downtown Toronto	2	Sushi Restaurant	Korean Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
52	Downtown Toronto	2	Sushi Restaurant	Japanese Restaurant	Restaurant	Korean Restaurant	Fast Food Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Hawaiian Restaurant	Grocery Store
54	Downtown Toronto	2	Sushi Restaurant	Japanese Restaurant	Fast Food Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store
55	Downtown Toronto	2	Sushi Restaurant	Restaurant	Japanese Restaurant	Grocery Store	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Hawaiian Restaurant	Fast Food Restaurant
57	Downtown Toronto	2	Sushi Restaurant	Japanese Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant
58	Downtown Toronto	2	Sushi Restaurant	Japanese Restaurant	Restaurant	Grocery Store	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Hawaiian Restaurant	Fast Food Restaurant

Results

- Using K-mean to clustering data area with less number of sushi bars.
  - Cluster 3

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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
29	North York	3	Sushi Restaurant	Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
46	Central Toronto	3	Sushi Restaurant	Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
56	Downtown Toronto	3	Sushi Restaurant	Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant
59	Downtown Toronto	3	Sushi Restaurant	Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
69	Downtown Toronto	3	Sushi Restaurant	Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant

Results

- Using K-mean to clustering data area with less number of sushi bars.
  - Cluster 4

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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
10	Scarborough	4	Japanese Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant
75	Downtown Toronto	4	Japanese Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant	Restaurant	Korean Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant

  - Cluster 5

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 5, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

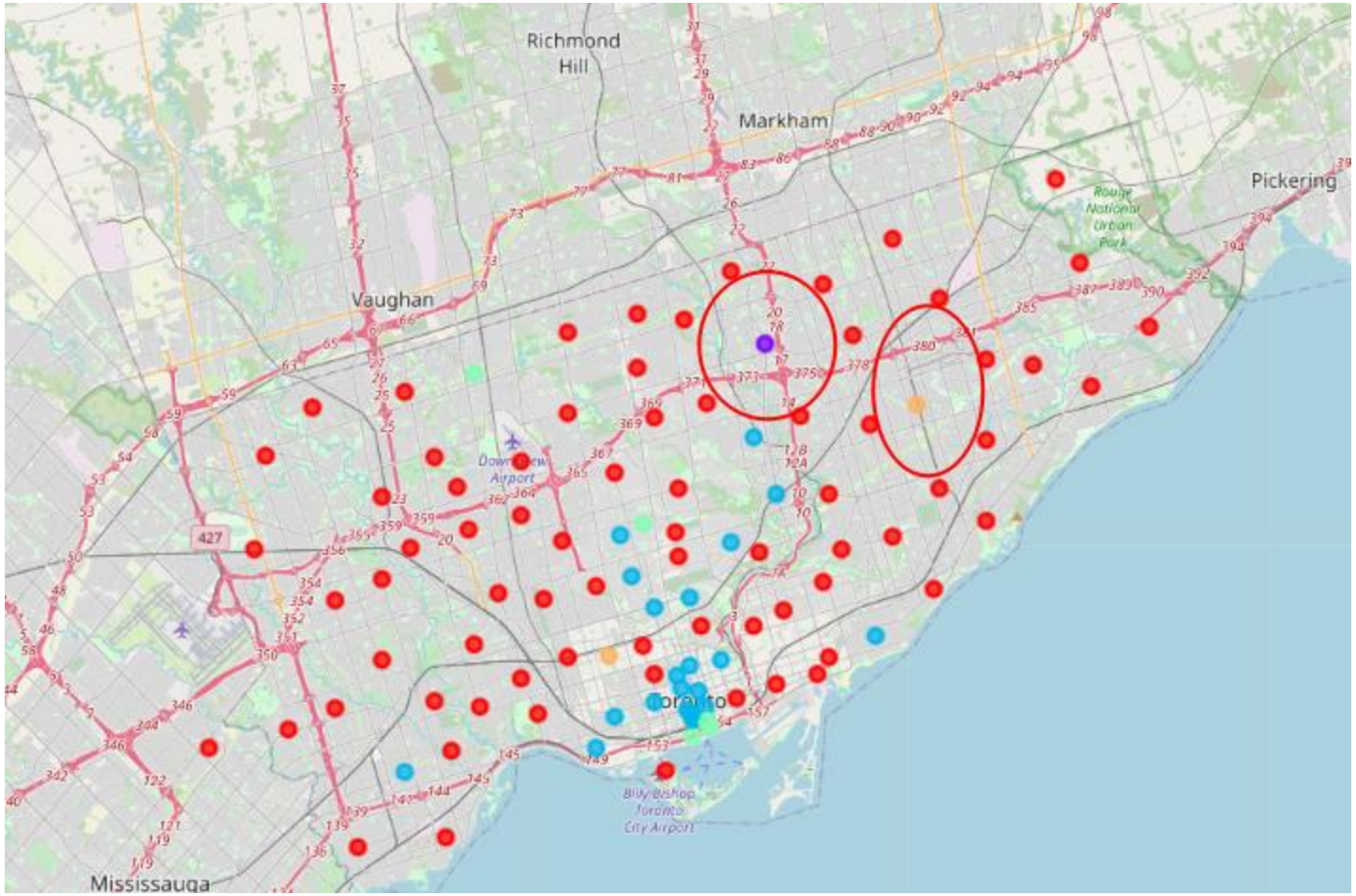
	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
50	Downtown Toronto	5	Sushi Restaurant	Korean Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Japanese Restaurant	Hawaiian Restaurant	Grocery Store	Fast Food Restaurant	Asian Restaurant

Discussion

- Our analysis shows that although there is a great number of Sushi restaurants in Toronto (~800 in our initial area of interest which was radius 1000 around each neighbourhood), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected in downtown Toronto area, so we focused our attention to areas north, north-east, far away from downtown area.

- Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Purpose of this analysis was to only provide info on areas in Toronto but not crowded with existing restaurants (particularly Sushi) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

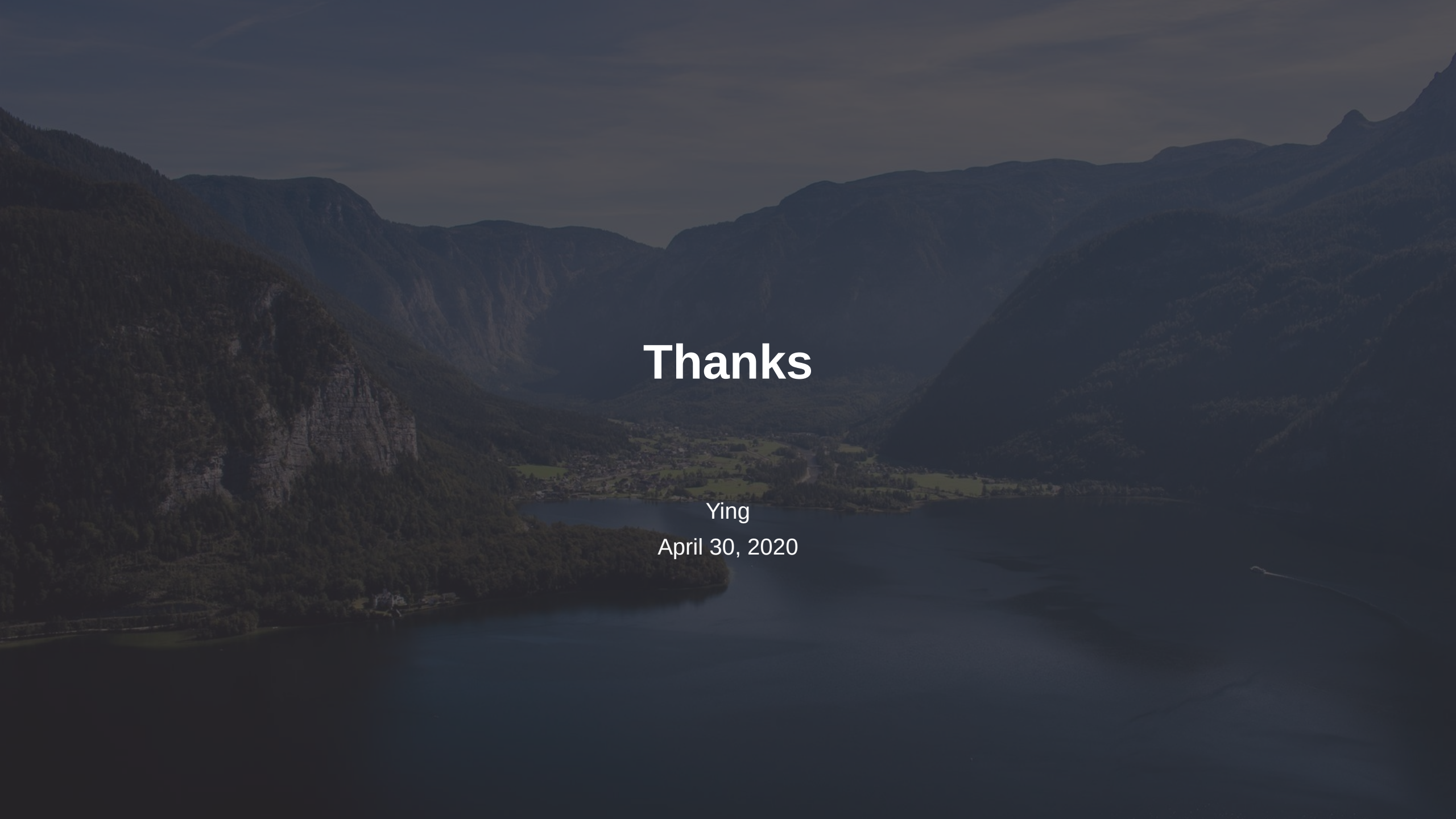
Conclusion



Based on dataframe analysis Cluster 2 (North York) and Cluster 4 (Scaborough Town) areas are good places to open a new sushi bar business.

Conclusion

- Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.



Thanks

Ying

April 30, 2020