The Battle of Neighborhoods - Scarborough, Toronto

1. Introduction

1.1 Background

Scarborough is located in the eastern part of Toronto and situated atop the Scarborough Bluffs. The vast majority of Scarborough's population is composed of immigrants. According to the census in 2016, the population is more than 630,000 and 57% of residents were foreign-born. 55.9% of population is from Asia, including south Asian (22%), Chinese (19%), Filipino (8.4%), west Asian (1.6%), Southeast Asian (0.8%), Korean (0.4%) and Japanese (0.3). The immigrant population has created vibrant multicultural locales in various areas of Scarborough. The heavy concentration of Chinese businesses and restaurants in the Agrincourt neighborhood is one of the most notable among these. Besides that, Caribbean, African and Indian restaurants and shops are also everywhere in this neighborhood. Scarborough is one of the most diverse and multicultural area in the Greater Toronto Area and is a popular destination for new immigrants in Canada to reside.

1.2 Business Problem

A Chinese investor would like to open a restaurant in Scarborough. Before the investment, he/she should understand a few problem,

1. Target consumers: identify the key demographics, such as average age, marital status, average disposable income, etc.
2. Market situation: if the local market is overly saturated more restaurants
3. Competing restaurants: for example, do they offer similar cuisine? Are they easy to find and access?

2. Data

The data in this project consists of two parts.

2.1 Explore the data in Scarborough

Extract Scarborough from the following link:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, which contains the region information, including Postal Code, borough and neighbourhood in Toronto. In this table, we only need extract Scarborough information for further analysis.

In addition, latitude and longitude will be needed by using geopy library for a map of Scarborough with neighborhoods superimposed on top.

2.2 Venue data in Scarborough

Venue data is used to perform clustering on Scarborough.

The Foursquare API will be used to explore neighborhoods in Scarborough. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, latitudes, longitudes, etc. For this project, we'll choose the radius to be 1000 meters for this area. We'll use the captured data for exploration, analysis and clustering.

3 Methodology

3.1 Data cleansing

The data collected from Wiki webpage includes all of neighbourhoods in Toronto. Before we start the process, the data need to be cleansed and merged for further analysis. The following rules will be applied for data cleansing,

1. The dataframe consists of three columns: Postalcode, Borough, and Neighborhood. Only Scarborough is chosen for this project.
2. More than one neighborhood can exist in Scarborough. All listed neighborhoods with same Postalcode will be combined.

The data extracted for Scarborough will be grouped as following,

	Postalcode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Figure 1 Rows grouped for Scarborough

Next, make a call to get the latitude and longitude coordinates of Scarborough by using geopy library,

	Postalcode	Borough	Neighborhood	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

Figure 2 Merged latitude and longitude into the table

Then, utilize the Foursquare API to explore the neighborhoods. In this project, within 1000-meter radius for the local venue will be explored. There are 89 venues in this area in total. Below is the first 27 venues shown here,

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant
5	Guildwood, Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
6	Guildwood, Morningside, West Hill	43.763573	-79.188711	Enterprise Rent-A-Car	43.764076	-79.193406	Rental Car Location
7	Guildwood, Morningside, West Hill	43.763573	-79.188711	Woburn Medical Centre	43.766631	-79.192286	Medical Center
8	Guildwood, Morningside, West Hill	43.763573	-79.188711	Lawrence Ave E & Kingston Rd	43.767704	-79.189490	Intersection
9	Guildwood, Morningside, West Hill	43.763573	-79.188711	Eggsmart	43.767800	-79.190466	Breakfast Spot
10	Woburn	43.770992	-79.216917	Starbucks	43.770037	-79.221156	Coffee Shop
11	Woburn	43.770992	-79.216917	Tim Hortons	43.770827	-79.223078	Coffee Shop
12	Woburn	43.770992	-79.216917	Korean Grill House	43.770812	-79.214502	Korean Restaurant
13	Cedarbrae	43.773136	-79.239476	Drupati's Roti & Doubles	43.775222	-79.241678	Caribbean Restaurant
14	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka Restaurant
15	Cedarbrae	43.773136	-79.239476	Thai One On	43.774468	-79.241268	Thai Restaurant
16	Cedarbrae	43.773136	-79.239476	Centennial Recreation Centre	43.774593	-79.236500	Athletics & Sports
17	Cedarbrae	43.773136	-79.239476	TD Canada Trust	43.774830	-79.241251	Bank
18	Cedarbrae	43.773136	-79.239476	B&A Bakery	43.774391	-79.243877	Bakery
19	Cedarbrae	43.773136	-79.239476	Petro-Canada	43.774106	-79.243097	Gas Station
20	Cedarbrae	43.773136	-79.239476	The Avenue	43.772437	-79.241962	Lounge
21	Cedarbrae	43.773136	-79.239476	Popeyes Louisiana Kitchen	43.775930	-79.235328	Fried Chicken Joint
22	Scarborough Village	43.744734	-79.239476	McCowan Park	43.745089	-79.239336	Playground
23	Scarborough Village	43.744734	-79.239476	Primo Pizza	43.744759	-79.244751	Pizza Place
24	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	Giant Tiger	43.727447	-79.266240	Department Store
25	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	Tim Hortons	43.726895	-79.266157	Coffee Shop
26	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	Dollarama	43.727092	-79.265784	Discount Store

Figure 3 Venue data pulled from Foursquare API in Scarborough

The venues number could be returned for each neighborhood,

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	4	4	4	4	4	4
Birch Cliff, Cliffside West	4	4	4	4	4	4
Cedarbrae	9	9	9	9	9	9
Clarks Corners, Tam O'Shanter, Sullivan	12	12	12	12	12	12
Cliffside, Cliffcrest, Scarborough Village West	2	2	2	2	2	2
Dorset Park, Wexford Heights, Scarborough Town Centre	7	7	7	7	7	7
Golden Mile, Clairlea, Oakridge	10	10	10	10	10	10
Guildwood, Morningside, West Hill	8	8	8	8	8	8
Kennedy Park, Ionview, East Birchmount Park	6	6	6	6	6	6
Malvern, Rouge	1	1	1	1	1	1
Milliken, Agincourt North, Steeles East, L'Amoreaux East	3	3	3	3	3	3
Rouge Hill, Port Union, Highland Creek	1	1	1	1	1	1
Scarborough Village	2	2	2	2	2	2
Steeles West, L'Amoreaux West	13	13	13	13	13	13
Wexford, Maryvale	4	4	4	4	4	4
Woburn	3	3	3	3	3	3

Figure 4 Venue numbers

From all the returned venues, there are 54 unique categories in Scarborough.

3.2 Data exploration

By using Folium, create map of Scarborough using latitude and longitude values.
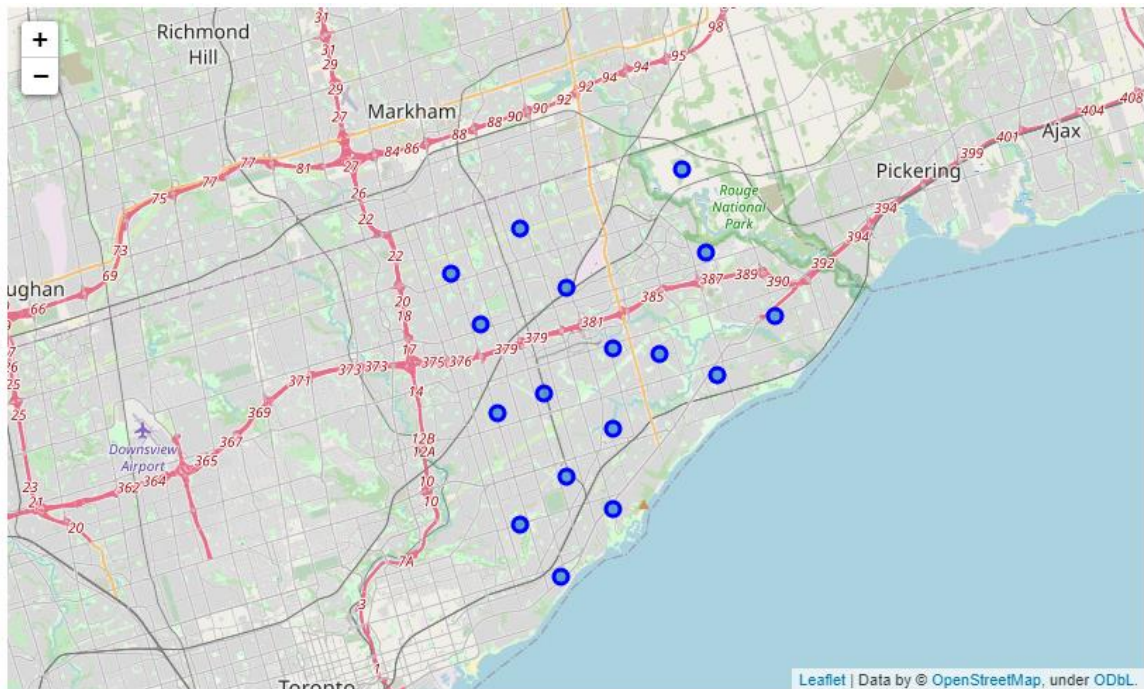

Figure 5 Map of Scarborough with neighborhoods superimposed on top

3.3 Machine learning

3.3.1 Categorize data

One-hot encoding is used on categorical data before fitting a machine learning model. One-hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction. Here is categorized data by using one-hot encoding,

	Neighborhood	American Restaurant	Asian Restaurant	Athletics & Sports	Bakery	Bank	Bar	Breakfast Spot	Bus Line	Bus Station	Café	Caribbean Restaurant	Chinese Restaurant	Coffee Shop	College Stadium	Convenience Store	Department Store	Discount Store
0	Malvern, Rouge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Rouge Hill, Port Union, Highland Creek	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	Guildwood, Morningside, West Hill	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 6 One-hot encoding

Then I grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category,

	Neighborhood	American Restaurant	Asian Restaurant	Athletics & Sports	Bakery	Bank	Bar	Breakfast Spot	Bus Line
0	Agincourt	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.250000	0.0
1	Birch Cliff, Cliffside West	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0
2	Cedarbrae	0.0	0.000000	0.111111	0.111111	0.111111	0.0	0.000000	0.0
3	Clarks Corners, Tam O'Shanter, Sullivan	0.0	0.000000	0.000000	0.000000	0.083333	0.0	0.000000	0.0
4	Cliffside, Cliffcrest, Scarborough Village West	0.5	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0

Figure 7 Grouped rows by neighborhoods for the frequency of occurrence

Top 10 venues were sorted for each neighborhood,

	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
0	Agincourt	Skating Rink	Breakfast Spot	Latin American Restaurant	Lounge	Vietnamese Restaurant	College Stadium	General Entertainment	Gas Station	Furniture / Home Store	Fried Chicken Joint
1	Birch Cliff, Cliffside West	College Stadium	General Entertainment	Skating Rink	Café	Vietnamese Restaurant	Grocery Store	Gas Station	Furniture / Home Store	Fried Chicken Joint	Fast Food Restaurant
2	Cedarbrae	Thai Restaurant	Athletics & Sports	Bakery	Bank	Gas Station	Fried Chicken Joint	Caribbean Restaurant	Lounge	Hakka Restaurant	Discount Store
3	Clarks Corners, Tam O'Shanter, Sullivan	Pizza Place	Fast Food Restaurant	Italian Restaurant	Chinese Restaurant	Noodle House	Pharmacy	Convenience Store	Fried Chicken Joint	Bank	Gas Station
4	Cliffside, Cliffcrest, Scarborough Village West	American Restaurant	Motel	Grocery Store	General Entertainment	Gas Station	Furniture / Home Store	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store

Figure 8 Top 10 venue example

3.3.2 K-Means clustering

K-mean clustering is used to cluster the neighborhoods based on the similarity of venues to each other. Let me simply explain this method. K-means is a type of partitioning clustering. It divides the data into k non-overlapping subset or clusters without any cluster internal structure or labels. Objects within a cluster are very similar, and objects across different clusters are very different or dissimilar.

Figure 9 Map with different clusters

3.4 Exploratory data analysis

There are total 5 clusters in Scarborough,

Cluster 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
5	Scarborough	0.0	Spa	Playground	Vietnamese Restaurant	College Stadium	Grocery Store	General Entertainment	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
14	Scarborough	0.0	Playground	Park	Vietnamese Restaurant	College Stadium	Grocery Store	General Entertainment	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant

Cluster 2:

	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
2	Scarborough	1.0	Intersection	Bank	Restaurant	Rental Car Location	Breakfast Spot	Electronics Store	Medical Center	Mexican Restaurant	Vietnamese Restaurant	Convenience Store
3	Scarborough	1.0	Coffee Shop	Soccer Field	Korean Restaurant	Indian Restaurant	Auto Garage	Department Store	Grocery Store	General Entertainment	Athletics & Sports	Gas Station
4	Scarborough	1.0	Thai Restaurant	Athletics & Sports	Hakka Restaurant	Bakery	Bank	Gas Station	Caribbean Restaurant	Lounge	Fried Chicken Joint	Discount Store
6	Scarborough	1.0	Discount Store	Chinese Restaurant	Coffee Shop	Department Store	Vietnamese Restaurant	Convenience Store	Hakka Restaurant	Grocery Store	General Entertainment	Gas Station
7	Scarborough	1.0	Bakery	Ice Cream Shop	Soccer Field	Intersection	Bus Station	Metro Station	Department Store	Grocery Store	General Entertainment	Gas Station
9	Scarborough	1.0	College Stadium	General Entertainment	Skating Rink	Café	Vietnamese Restaurant	Grocery Store	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
10	Scarborough	1.0	Indian Restaurant	Pet Store	Chinese Restaurant	Gaming Cafe	Light Rail Station	Vietnamese Restaurant	Restaurant	Coffee Shop	Gas Station	Soccer Field
11	Scarborough	1.0	Middle Eastern Restaurant	Auto Garage	Bakery	Sandwich Place	Breakfast Spot	Vietnamese Restaurant	Grocery Store	General Entertainment	Gas Station	Gaming Cafe
12	Scarborough	1.0	Skating Rink	Latin American Restaurant	Breakfast Spot	Lounge	Vietnamese Restaurant	Convenience Store	General Entertainment	Gas Station	Gaming Cafe	Fried Chicken Joint
13	Scarborough	1.0	Pizza Place	Convenience Store	Thai Restaurant	Italian Restaurant	Bank	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Chinese Restaurant	Noodle House
15	Scarborough	1.0	Fast Food Restaurant	Coffee Shop	Chinese Restaurant	Indian Restaurant	Breakfast Spot	Discount Store	Pharmacy	Pizza Place	Grocery Store	Sandwich Place

Cluster 3:

	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	2.0	Fast Food Restaurant	Vietnamese Restaurant	Thai Restaurant	Hakka Restaurant	Grocery Store	General Entertainment	Gas Station	Gaming Cafe	Fried Chicken Joint	Electronics Store

Cluster 4:

	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
8	Scarborough	3.0	American Restaurant	Intersection	Motel	Convenience Store	Grocery Store	General Entertainment	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant

Cluster 5:

	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
1	Scarborough	4.0	Moving Target	Bar	Vietnamese Restaurant	Grocery Store	General Entertainment	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store

4 Results

In those 5 clusters, cluster 2 has the most neighborhood and cluster 3, 4 and 5 has only 1 neighborhood. Based on the top 10 venues in cluster 2, most of the restaurants are Asian cuisine. There are 6 Vietnamese restaurants, 4 Chinese restaurants, 3 Indian restaurants, and 2 Thai restaurants. In other 4 clusters, there are also Asian restaurants.

5 Discussion

As mentioned in the beginning, over half of population in Scarborough is from Asia. According to the data analysis, most of restaurants are Asian food. In neighborhood 7 (Golden Mile, Clairlea, Oakridge), there is no any Asian restaurants. For neighborhood 7, since there are bus station and metro station, there are potential customers for a new restaurant. In addition, it could eliminate any competition because there is not any restaurants. Another neighborhood 9, if choose Asian cuisine, potentially, only one Vietnamese restaurant could be a competitor. Thus, it's still another option to open a new restaurant.

However, I only analyzed the venues in Scarborough. In order to make the research more complete, other factors, such as income, age, marital status, could be also considered.

6 Conclusion

To summarize, I utilized the data from different sources, such as Wikipedia and Foursquare API, and analyzed them by using different Python libraries. The analysis result could help deal with the business problem mention at the beginning.

To answer the questions from the Chinese investor, it still has potential to open a new Asian restaurant in Scarborough. Based on the above analysis, the first location could be in Golden Mile, Clairlea and Oakridge. The second option could be in Birch Cliff and Cliffside West.