Using Data Science Methods to Find Locations to Open a Chinese Restaurant in Toronto

Tran Viet Hung

1. Introduction

1.1. Background

Toronto, city, capital of the province of Ontario, southeastern Canada. It is the most populous city in Canada, a multicultural city, and the country's financial and commercial centre. Toronto is an international centre for business and finance. Generally considered the financial capital of Canada, Toronto has a high concentration of banks and brokerage firms on Bay Street, in the Financial District. The Toronto Stock Exchange is the world's seventh-largest stock exchange by market capitalization. The five largest financial institutions of Canada, collectively known as the Big Five, have national offices in Toronto. The city is an important centre for the media, publishing, telecommunication, information technology and film production industries; it is home to Bell Media, Rogers Communications, and Torstar. Other prominent Canadian corporations in the Greater Toronto Area include Magna International, Celestica, Manulife, Sun Life Financial, the Hudson's Bay Company, and major hotel companies and operators, such as Four Seasons Hotels and Fairmont Hotels and Resorts.

Toronto with a lot of people and a lot of culture is a wonderful land to start a new F&B business. In 2016, about 11.1% Toronto people are Chinese and it increased to about 12% in 2019. There may be a great chance to begin a new Chinese restaurant in Toronto. We think there are not enough Chinese restaurants in Toronto at the moment due to high rate of Chinese people. Moreover, the number of Chinese is increasing in Toronto year to year, they need more Chinese restaurants.

1.2. Business Problem

The object of my project is to find the most suitable location for someone who want to open new Chinese restaurant in Toronto. Using data science methods, data analysis method, machine learning tools such as

clustering, the project provides solutions to answer the question: If some people want to open a new Chinese restaurant in Toronto, where they can consider to open it?

1.3. Target Audience

Some people who wants to find the location to open authentic Burmese restaurant.

2. Data

To solve problem, I will need following data.

- List of neighborhoods in Toronto, Canada.
- Latitude and Longitude of these neighborhoods.
- Venue data related to Chinese restaurants. We can find the best location using venue data.

Getting data,

- List of neighborhoods in Toronto, Canada: Scrapping of Toronto neighborhoods via Wikipedia

	PostalCode	Borough	Neighborhood
9	M1B	Scarborough	Malvern, Rouge
18	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
27	M1E	Scarborough	Guildwood, Morningside, West Hill
36	M1G	Scarborough	Woburn
45	M1H	Scarborough	Cedarbrae

- Latitude and Longitude of these neighborhoods: Getting Latitude and Longitude data of these neighborhoods via Geocoder package

	PostalCode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

- Venue data related to Chinese restaurants: Using Foursquare API to get venue data related to these neighborhoods

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka Restaurant
1	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	Hakka No.1	43.727688	-79.266057	Chinese Restaurant
2	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304	Kim Kim restaurant	43.753833	-79.276611	Chinese Restaurant
3	Clarks Corners, Tam O'Shanter, Sullivan	43.781638	-79.304302	The Royal Chinese Restaurant 避風塘小炒	43.780505	-79.298844	Chinese Restaurant
4	Steeles West, L'Amoreaux West	43.799525	-79.318389	Mr Congee Chinese Cuisine 龍粥記	43.798879	-79.318335	Chinese Restaurant

3. Methodology

3.1. Analytic Approach

We use the position data and venue information, venue category data of Toronto neighborhood to predict the most suitable position to open a new Chinese restaurant in that city.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka Restaurant
1	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	Hakka No.1	43.727688	-79.266057	Chinese Restaurant
2	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304	Kim Kim restaurant	43.753833	-79.276611	Chinese Restaurant
3	Clarks Corners, Tam O'Shanter, Sullivan	43.781638	-79.304302	The Royal Chinese Restaurant 避風塘小炒	43.780505	-79.298844	Chinese Restaurant
4	Steeles West, L'Amoreaux West	43.799525	-79.318389	Mr Congee Chinese Cuisine 龍粥記	43.798879	-79.318335	Chinese Restaurant

The feature includes Neighborhood name (we do not use this feature for prediction), Neighborhood Latitude and Neighborhood Longitude (We use this for getting venue information from FourSquare and for visualization), venue name (We do not use), Venue Latitude and Venue Longitude (We use for visualization), Venue Category (We use this for my model, but we need to convert to one-hot since it is categorical feature now).

Because the data does not include labels, we decide to use unsupervised learning to solve problem. We will cluster all the neighborhood based on their venue categories rates. It is clear that venue does not have hierarchical structure, so k-means clustering is a good choice.

The tools we will use: Pandas for cleaning and understanding data, BeautifulSoup for crawling data, geocoder package for getting location information, folium for visualizing map, scikit-learn for machine learning algorithm.

3.2. Data Requirements

The feature includes Neighborhood name (we do not use this feature for prediction), Neighborhood Latitude and Neighborhood Longitude (We use this for getting venue information from FourSquare and for visualization), venue name (We do not use), Venue Latitude and Venue Longitude (We use for visualization), Venue Category (We use this for my model, but we need to convert to one-hot since it is categorical feature now).

3.3. Data Collections

- List of neighborhoods in Toronto, Canada: Scrapping of Toronto neighborhoods via Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Latitude and Longitude of these neighborhoods: Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Venue data related to Chinese restaurants: Using Foursquare API to get venue data related to these neighborhoods

3.4. Data Understanding and Preparation

- Crawl Toronto neighborhood from wikipedia data and push to pandas dataframe.

	Postal code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park / Harbourfront

- Cleaning Data: change cloumns name, drop not assign rows, regex job.

	PostalCode	Borough	Neighborhood
9	M1B	Scarborough	Malvern, Rouge
18	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
27	M1E	Scarborough	Guildwood, Morningside, West Hill
36	M1G	Scarborough	Woburn
45	M1H	Scarborough	Cedarbrae

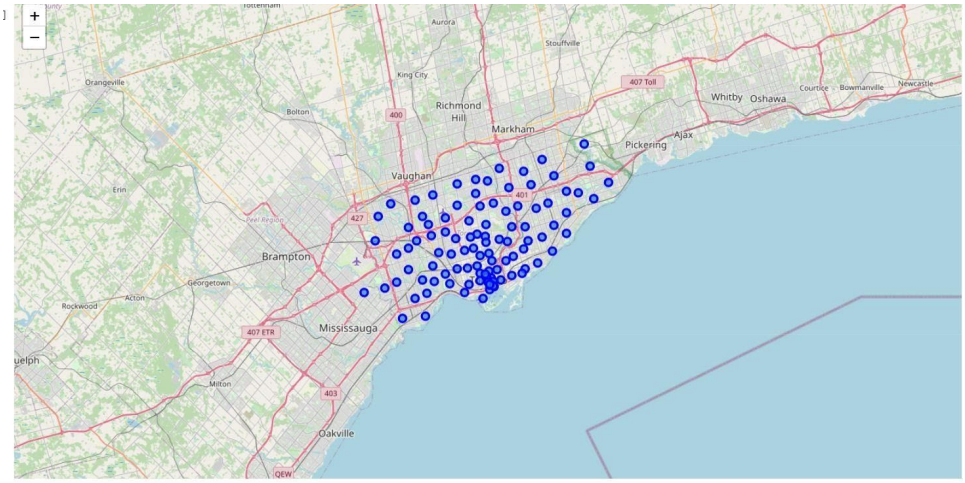
- Get Location data of Toronto, push to dataframe and clean data.

	PostalCode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

- Merge Neighborhood data and location data

	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

- Get location of Toronto
- Visualize Toronto neighborhood



- Define Foursquare Credentials and Version
- Get Venue Information of Toronto from FourQuare, we only get information about neighborhood that has Venue Category is Chinese restaurant (id of Chinese restaurant category is '4bf58dd8d48988d145941735')

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka Restaurant
1	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	Hakka No.1	43.727688	-79.266057	Chinese Restaurant
2	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304	Kim Kim restaurant	43.753833	-79.276611	Chinese Restaurant
3	Clarks Corners, Tam O'Shanter, Sullivan	43.781638	-79.304302	The Royal Chinese Restaurant 避風塘小炒	43.780505	-79.298844	Chinese Restaurant
4	Steeles West, L'Amoreaux West	43.799525	-79.318389	Mr Congee Chinese Cuisine 龍粥記	43.798879	-79.318335	Chinese Restaurant

- Get category information of neighborhood from FourSquare, we only get information about neighborhood that has Venue Category is Chinese restaurant (id of Chinese restaurant category is '4bf58dd8d48988d145941735')
- Convert venue categories to one-hot

	NEIGHBORHOOD	Asian Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	Hakka Restaurant	Hong Kong Restaurant	Hotpot Restaurant	Noodle House	Peking Duck Restaurant
0	Cedarbrae	0	0	0	0	0	0	0	0	0	1	0	0	0	0
1	Kennedy Park, Ionview, East Birchmount Park	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	Dorset Park, Wexford Heights, Scarborough Town...	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3	Clarks Corners, Tam O'Shanter, Sullivan	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	Steeles West, L'Amoreaux West	0	0	0	0	1	0	0	0	0	0	0	0	0	0

- Sum up categories for each neighborhood and normalize data

	NEIGHBORHOOD	Asian Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	Hakka Restaurant	Hong Kong Restaurant	Hotpot Restaurant	Noodle House
0	Alderwood, Long Branch	0.000000	0.000000	0.000000	0.000000	1.000000	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
1	Bathurst Manor, Wilson Heights, Downsview North	0.000000	0.000000	0.000000	0.000000	1.000000	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
2	Bayview Village	0.000000	0.000000	0.000000	0.000000	1.000000	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
3	Berczy Park	0.000000	0.000000	0.000000	0.000000	1.000000	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
4	Cedarbrae	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	1.0	0.000000	0.000000	0.000000
5	Central Bay Street	0.000000	0.000000	0.052632	0.000000	0.789474	0.0	0.000000	0.000000	0.052632	0.0	0.000000	0.000000	0.000000

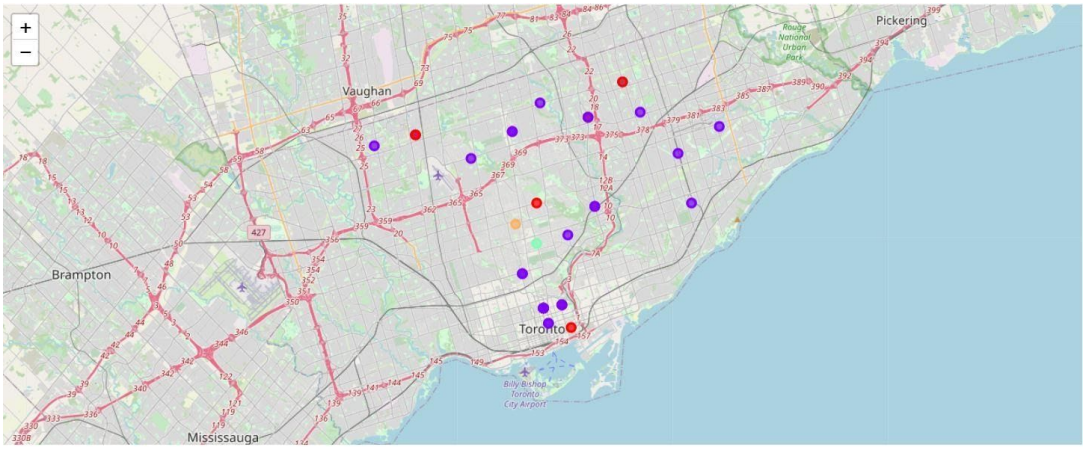
- Get information about top 10 categories for each neighborhood

	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	Neighborhood Latitude	Neighborhood Longitude
0	Alderwood, Long Branch	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.773136	-79.239476
1	Bathurst Manor, Wilson Heights, Downsview North	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.727929	-79.262029
2	Bayview Village	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.757410	-79.273304
3	Berczy Park	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.781638	-79.304302
4	Cedarbrae	Hakka Restaurant	Thai Restaurant	Dumpling Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fried Chicken Joint	43.799525	-79.318389

3.5. Modeling and Evaluation

We use Kmean Clustering Algorithm to cluster all neighborhood that has venue category Chinese restaurant to 5 clusters and visualize to the map using folium.

	Cluster Labels	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	Neighborhood Latitude	Neighborhood Longitude
0	1	Alderwood, Long Branch	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.773136	-79.239476
1	1	Bathurst Manor, Wilson Heights, Downsview North	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.727929	-79.262029
2	1	Bayview Village	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.757410	-79.273304
3	1	Berczy Park	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint	43.781638	-79.304302
4	2	Cedarbrae	Hakka Restaurant	Thai Restaurant	Dumpling Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fried Chicken Joint	43.799525	-79.318389



4. Results

We find the kmean clusters with less Chinese restaurants.

- Cluster 0

	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
5	Central Bay Street	Chinese Restaurant	Thai Restaurant	Bubble Tea Shop	Fried Chicken Joint	Sushi Restaurant	Peking Duck Restaurant	Noodle House	Hotpot Restaurant	Hong Kong Restaurant	Hakka Restaurant
6	Church and Wellesley	Chinese Restaurant	Asian Restaurant	Taiwanese Restaurant	Bubble Tea Shop	Bar	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Thai Restaurant
16	Garden District, Ryerson	Chinese Restaurant	Thai Restaurant	Fried Chicken Joint	Sushi Restaurant	Peking Duck Restaurant	Noodle House	Hotpot Restaurant	Hong Kong Restaurant	Hakka Restaurant	Taiwanese Restaurant
19	Kensington Market, Chinatown, Grange Park	Chinese Restaurant	Dumpling Restaurant	Fried Chicken Joint	Bar	Bubble Tea Shop	Cantonese Restaurant	Dim Sum Restaurant	Asian Restaurant	Hong Kong Restaurant	Hotpot Restaurant
24	Queen's Park, Ontario Provincial Government	Chinese Restaurant	Taiwanese Restaurant	Thai Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
29	St. James Town, Cabbagetown	Chinese Restaurant	Taiwanese Restaurant	Thai Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
35	University of Toronto, Harbord	Chinese Restaurant	Comfort Food Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint

- Cluster 1

	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	Alderwood, Long Branch	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
1	Bathurst Manor, Wilson Heights, Downsview North	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
2	Bayview Village	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
3	Berczy Park	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
7	Clarks Corners, Tam O'Shanter, Sullivan	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
8	Commerce Court, Victoria Hotel	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
9	Davisville	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
10	Del Ray, Mount Dennis, Keelsdale and Silverthorn	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
12	Dorset Park, Wexford Heights, Scarborough Town...	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
13	Downsview	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
14	Fairview, Henry Farm, Oriole	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint
15	First Canadian Place, Underground city	Chinese Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Fried Chicken Joint

- Cluster 2

	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
4	Cedarbrae	Hakka Restaurant	Thai Restaurant	Dumpling Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fried Chicken Joint

- Cluster 3

	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
21	Leaside	Peking Duck Restaurant	Thai Restaurant	Dumpling Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fried Chicken Joint

- Cluster 4

	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
11	Don Mills	Asian Restaurant	Chinese Restaurant	Dim Sum Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dumpling Restaurant	Thai Restaurant
20	Lawrence Park	Dim Sum Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dumpling Restaurant	Fried Chicken Joint
27	Runnymede, Swansea	Chinese Restaurant	Dim Sum Restaurant	Thai Restaurant	Taiwanese Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Comfort Food Restaurant	Dumpling Restaurant	Fried Chicken Joint

Based on the results, we suggest people should open Chinese restaurant in cluster 1, 3, or 4 in the following neighborhood: Cedarbrae, Leaside, Don Mills, Lawrence Park, Runnymede, Swansea.

5. Discussion

We analyzed data, cluster to choose suitable location for opening new Chinese restaurant. In this section, we will discuss more about the results and method.

Firstly, we think my method only work for a limited data and a simple feature, we only choose locations based on that location categories (is Chinese restaurant) and number of Chinese business at that location (less is better). However, sometimes there are only a few restaurants in that location that because not very much Chinese live there. Therefore, we think we need more data about inhabitant, transportation, number of Chinese live in each neighborhood, income of Chinese restaurant in each venue.

Secondly, we think FourSquare data maybe outdated after several months, so we can write some scripts to automatically update the data and run machine learning algorithm again.

6. Conclusion

Although our model needs a lot of improvements and refactor, we think we have a clear initial methodology and results. The goals of project were satisfied and our algorithm can be easily improved by collecting more data and update outdated data, it will improve a lot after some circles of data science application product developing. Our model can be a very good application providing suggestion for any people who want start new business of Chinese restaurant.