Ideal Location for Chinese Restaurant in Toronto

Capstone Report

IBM Data Science

Mun Yee

INTRODUCTION

Background

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Toronto encompasses a geographical area formerly administered by many separate municipalities. Throughout the city there exist hundreds of small neighbourhoods and some larger neighbourhoods covering a few square kilometres. The three most commonly reported ethnic origins overall were Chinese (332,830 or 12.5%), English (331,890 or 12.3%) and Canadian (323,175 or 12.0%).

Opening a new restaurant can be a lucrative business, a lack of demands can lead to restaurant closing down within first year of business. There are many factors that can affect the success rate of a restaurant business which include the location of restaurant, competition, target customers and quality of food. It is important for the restaurant to be in a strategic location. To demonstrate the process of choosing the right location for a client who is opening a new restaurant in Toronto, the focus of this project will be the analysis of the geographical aspect of the area, the population, spending power and restaurants that exist in the area to guide the client in choosing the best location for his/her new Chinese restaurant.

Data

In order to solve the business problem, various data is collected for analysis purpose. The data required:

- Population and ethnic distribution of each neighbourhood
- Income distribution of each neighbourhood
- Number of restaurants in each neighbourhood
- Number of Chinese restaurants in each neighbourhood
- Latitude and Longitude of each neighbourhood

Methodology

The first of the project is to combine the Toronto dataset, containing neighbourhood name, borough, latitude and longitude for each postal code in Toronto. The combined data is the first four columns in table 4. Using income distribution and population data from Toronto census, the median spending power of each area is calculated using median of each category weighed by the number of people in the same income category. The calculated spending power represents the overall capital of each area. The spending power for each area is then standardised due to large variation and relative strength is difficult to visualise.

The next step is to visualise the location of various postal code within Toronto using Foursquare API to obtain a general overview of the location. Based on the map (figure 1), the postcodes are densely clustering around Toronto downtime area. The population in downtime area are considerably larger compare to other regions and the postal codes tend to spread out as we move away from Toronto downtown. This is important because some postal codes might have many restaurants, if the area is located near downtown, adjacent regions can heavily impact the profitability of the restaurant.

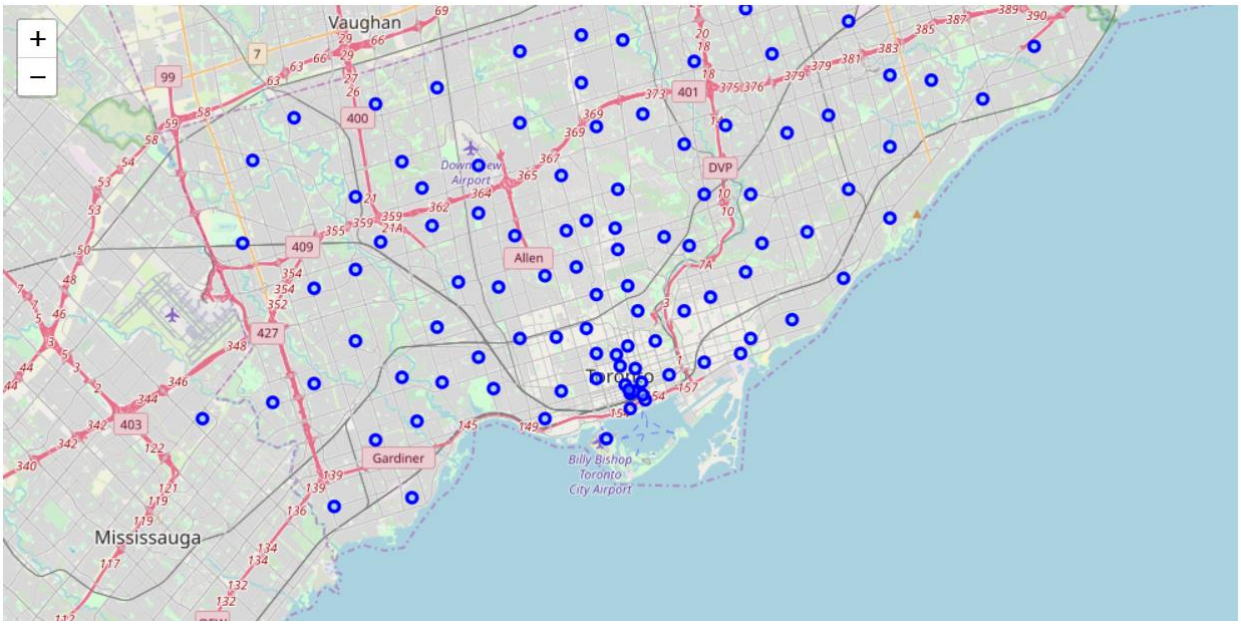

Figure 1: Overview of Toronto postcode locations

The top 200 venues within 2000 meters of the longitude and latitude for each neighbourhood is then obtained using Foursquare. The extracted venue categories were encoded using one-hot encoding and total number of restaurant and Chinese restaurants in each region were calculated as shown in Table 1.

	Neighborhood	Total Restaurants	Chinese Restaurants
0	Adelaide, King, Richmond	33	0
1	Agincourt	48	12
2	Agincourt North, L'Amoreaux East, Milliken, St...	47	16
3	Albion Gardens, Beaumond Heights, Humbergate, ...	20	1
4	Alderwood, Long Branch	38	0
5	Bathurst Manor, Downsview North, Wilson Heights	25	1
6	Bayview Village	15	5
7	Bedford Park, Lawrence Manor East	43	1
8	Birch Cliff, Cliffside West	15	1
9	Bloordale Gardens, Eringate, Markland Wood, Ol...	5	1
10	Brockton, Exhibition Place, Parkdale Village	34	0

Table 1: Total no. of restaurants and no. of Chinese restaurant in each neighbourhood

With the resulting data, the postal code, borough name, latitude, longitude and density of columns of each region were dropped from dataframe. The initial income dataset for income distribution contains the borough, neighbourhood name, latitude, longitude, population and income distribution in Canadian dollar (Table 2).

	PostCode	Borough	Neighborhood	Latitude	Longitude	Population	Density	Area	< 5k	5k - 10k	10k - 15k	15k - 20k	20k - 25k	25k - 30k	30k - 35k
0	M3A	North York	Parkwoods	43.753259	-79.329656	34805.0	4691.0	7.42	345.0	185.0	315.0	520.0	590.0	600.0	615.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	17510.0	3710.0	4.72	180.0	205.0	280.0	540.0	420.0	385.0	400.0
2	M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636	76716.0	25823.0	8.01	1975.0	1180.0	1535.0	1695.0	1520.0	1440.0	1370.0
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763	6577.0	3614.0	1.82	50.0	60.0	160.0	190.0	185.0	170.0	190.0
4	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	90290.0	6208.0	45.74	290.0	240.0	420.0	720.0	730.0	925.0	955.0

Tablet 2(i): Income data per household for each neighbourhood

35k - 40k	40k - 45k	45k - 50k	50k - 60k	60k - 70k	70k - 80k	80k - 90k	90k - 100k	100k - 125k	125k - 150k	150k - 200k	> 200k	South Asian	Chinese	Black	Filipino	Latin American	Arab
575.0	620.0	620.0	1200.0	1025.0	880.0	790.0	650.0	1220.0	790.0	890.0	890.0	10.37	8.26	11.23	7.00	2.94	4.35
370.0	370.0	425.0	640.0	570.0	490.0	385.0	290.0	565.0	315.0	325.0	240.0	17.05	4.17	11.85	7.08	2.46	6.11
1425.0	1485.0	1460.0	3085.0	3305.0	3150.0	2805.0	2580.0	4825.0	3200.0	3695.0	4000.0	17.93	13.09	9.55	2.46	2.39	2.25
170.0	155.0	125.0	205.0	200.0	160.0	110.0	105.0	165.0	95.0	100.0	60.0	5.02	1.67	31.47	6.01	5.09	1.06
1090.0	1055.0	1110.0	2330.0	2150.0	1930.0	1845.0	1640.0	3355.0	2315.0	2390.0	1300.0	41.64	6.00	16.49	9.92	1.41	0.84

Tablet 2 (ii): Income data per household for each neighbourhood(cont.)

Southeast Asian	West Asian	Korean	Japanese	White	Spending Power
0.56	4.60	1.67	0.65	44.98	0.061812
0.54	4.11	0.91	0.43	40.18	-0.693676
2.32	1.80	1.15	0.56	42.69	3.838132
3.27	0.15	0.23	0.00	39.68	-1.120335
0.55	1.32	0.16	0.15	14.64	1.756524

Tablet 2 (iii): Income data per household for each neighbourhood(cont.)

The population, area, spending power, total number of restaurants and the number of Chinese restaurants were used to train a K-Means clustering algorithm with five clusters. The characteristics of the resulting clusters are listed in Table 3

Result

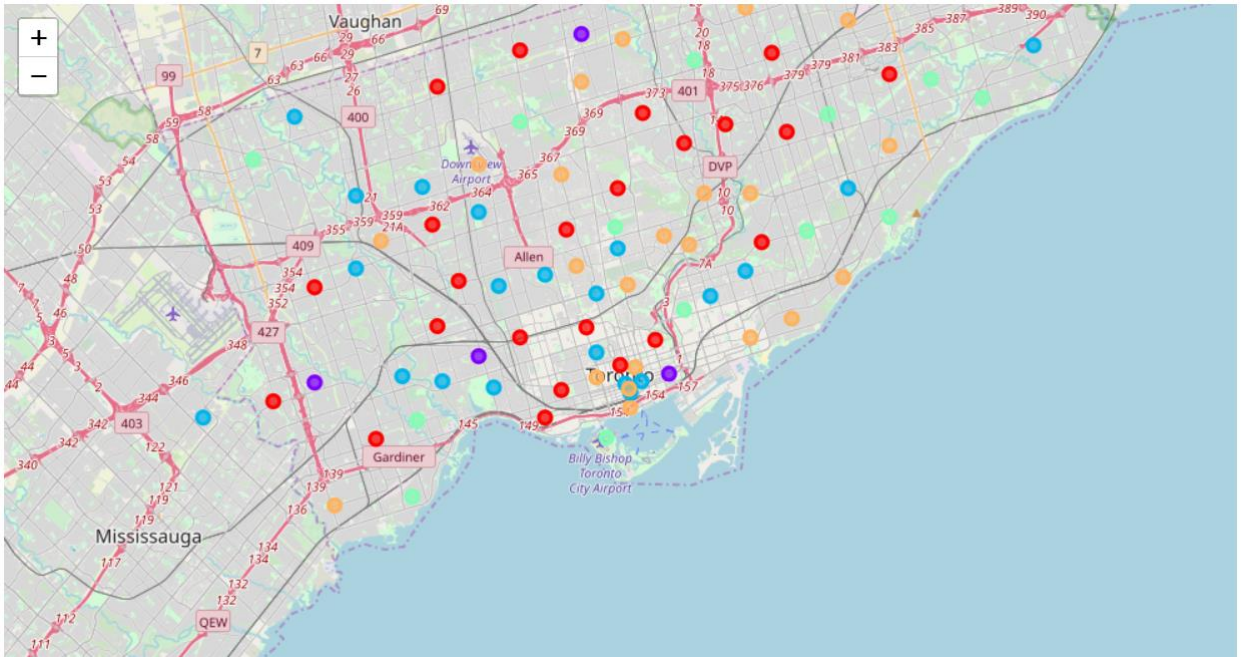

Figure 2: Clustering regions based on population, area, spending power, total no. of restaurant and no. of Chinese restaurant (Red= Cluster 0, Purple = Cluster 1, Blue = Cluster 2, Turquoise = Cluster 3, Orange= Cluster 4)

Cluster Label	Characteristics
0	Low population, low spending power (-0.4 to 0.6), various % of restaurant, area and Chinese
1	Generally large population, high spending power (1.7 to 3.8), various % of area and Chinese. Scarborough has the highest no. of Chinese restaurant in this cluster
2	Low population and extremely low spending power (negative)
3	Average size of population and average spending power (0.2- 1.7)
4	Low population and low spending power (negative to 0.2)

Table 3: Characteristic of each cluster

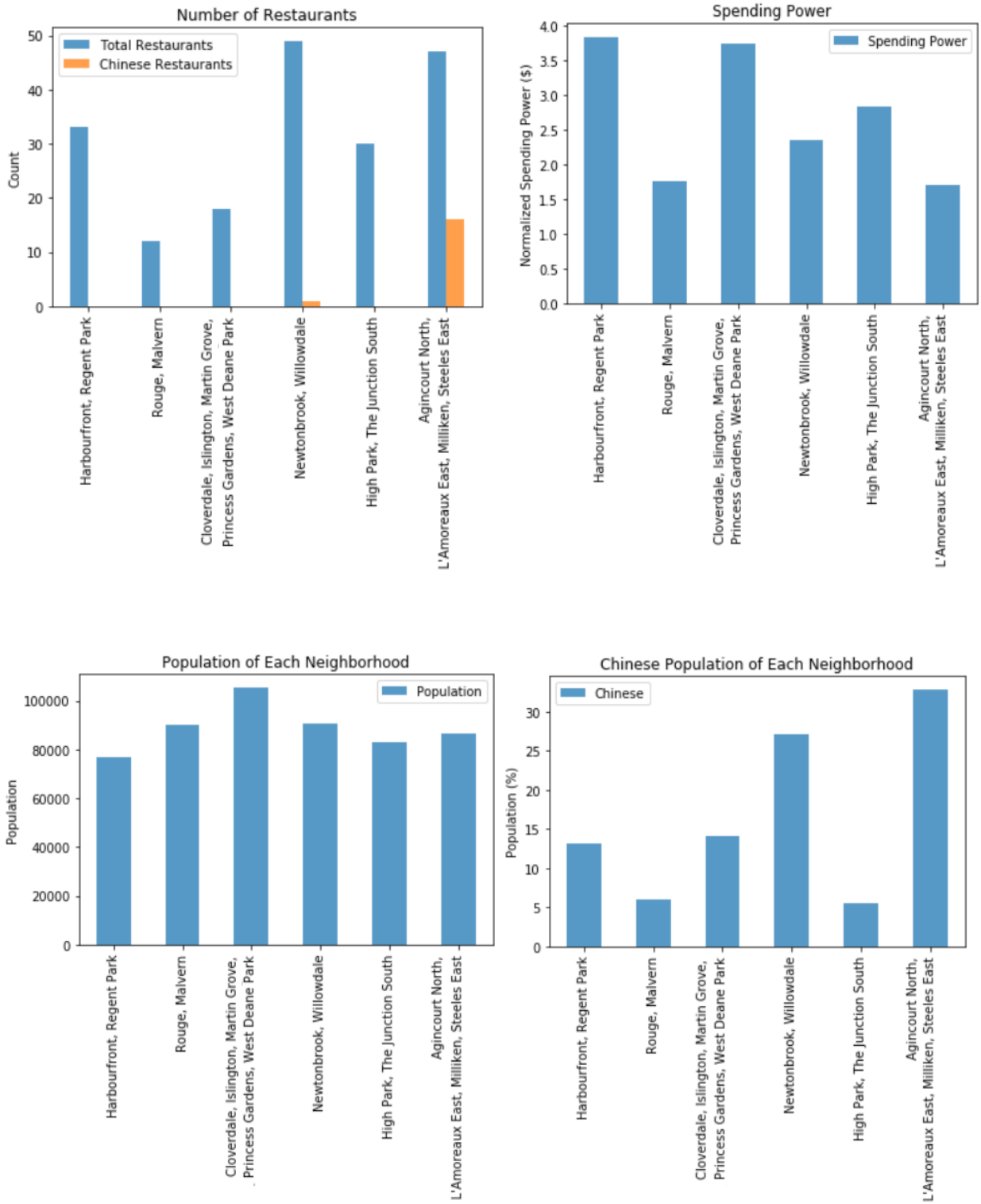

Figure 3: Plots that show characteristic of neighbourhoods belong to cluster label=1

Discussion

Based on the clustering algorithm results, cluster label= 1 has the largest population and highest normalised spending power compare to other regions. Cluster region 1 found to be the most appropriate region to open a new Chinese restaurant.

	Borough	Cluster Labels	Population	Area	Chinese	Spending Power	Total Restaurants	Chinese Restaurants
2	Downtown Toronto	1	76716.0	8.01	13.09	3.838132	33	0
4	Scarborough	1	90290.0	45.74	6.00	1.756524	12	0
8	Etobicoke	1	105450.0	26.38	14.15	3.748670	18	0
45	North York	1	90362.0	13.80	27.12	2.350813	49	1
59	West Toronto	1	82712.0	10.51	5.56	2.841538	30	0
74	Scarborough	1	86468.0	19.96	32.86	1.712083	47	16

Table 4: Details of neighbourhood in Cluster=1

There are total of six boroughs that are categorised under Cluster 1. Out of the six boroughs, Etobicoke found to have the largest population in all regions. The spending power of 3.75 for people in this borough are among few of the highest in Toronto. On top of that, the number of restaurants in this area is 18 and there is no Chinese restaurant currently present in this region. The neighbourhoods in this borough are Cloverdale, Islington, Martin Grove, Princess Gardens and West Deane Park. The % of Chinese population in this region (14.15%) is higher than the Scarborough. (4) which have lower % of Chinese but fewer restaurant compare to Etobicoke. Based on the results obtained, Etobicoke (Cloverdale, Islington, Martin Grove, Princess Gardens and West Deane Park) appear to be the most approach region to open a Chinese restaurant due to its high spending power and large population. This area stands a higher chance of success due to its relative low number of restaurant and no Chinese restaurant at present.

Conclusion

Opening a new restaurant in one of the big cities in Canada is not an easy task and can lead of monetary loss if not planned well. Extensive research and analysis of the area, population as well as the spending pattern in the area could help in choosing the right location and increase chances of success. Through this project, I manage to do a thorough research analysis to help client with the location choice for the new restaurant. Etobicoke region (Cloverdale, Islington, Martin Grove, Princess Gardens and West Deane Park) is the recommended region for the new Chinese restaurant.