The Battle of Neighborhoods

Project for Coursera course Applied Data Science Capstone

By Melody Zhang

8/21/2019

Introduction

Location for the restaurant is extremely important for a restaurant's success. The client want to open a Chinese restaurant, and they are looking for the best location to open it. To find this answer, we need to first consider whether that neighborhood has enough Asian population. To make good profit, we also need to consider the household income in that area, which cannot be too low. Else, other factors might also affect the choice of location. For example, not enough Chinese restaurant density, low crime rate, enough parking lot density, etc.

Data

To answer this question, following data will be used:

- List of Toronto neighborhood (Wikipedia :https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Toronto geospatial data (https://cocl.us/Geospatial_data)
- Population and Ethnic of each neighborhood (Toronto Census: https://open.toronto.ca/dataset/neighbourhood-profiles/)
- Household income of each neighborhood (Toronto Census: https://open.toronto.ca/)
- Number of Chinese restaurant (Foursquare API)
- Google

Methodology

All the data need to be combined to one data, which should include postal code, neighborhood, latitude, longitude, population, Income, Chinese restaurant count. When dealing with data from all places, some factors need to be checked carefully. For example, the neighborhood name is different from different datasets. The census data are plot on Toronto map, which can give a initial impression of areas having highest Chinese population, and highest average income.

To obtain the Chinese restaurant count, Foursquare API was used to explore each neighborhood and return 500 restaurants within 3000 meters of the longitude and latitude for each postal code. In the combined data frame, K-Mean clustering algorithms is used to analyze Asian origins, Chinese Population, restaurant count, and divide the data to 5 clusters. This result help narrowing down the target neighborhood which is good for opening a new Chinese restaurant.

Result

Below is a post code table scraped from Wikipedia. Drop the row if Borough has not assigned. Copy Borough name if neighborhood is not assigned.

Table 1: List of post code for neighborhoods in Toronto

	Postcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights
5	M6A	North York	Lawrence Manor
6	M7A	Queen's Park	Queen's Park
7	M9A	Etobicoke	Islington Avenue
8	M1B	Scarborough	Rouge
9	M1B	Scarborough	Malvern

Collect census data about Chinese population and average income of each neighborhood.

Table 2: Toronto neighborhood profiles sort by average income

index	Neighbourhood	Neighbourhood Number	Asian origins	Chinese	Total income: Average amount ($)
16	Bridle Path-Sunnybrook-York Mills	41.0	3005.0	1445.0	308010.0
104	Rosedale-Moore Park	98.0	3485.0	1300.0	207903.0
44	Forest Hill South	101.0	1880.0	560.0	204521.0
69	Lawrence Park South	103.0	2510.0	935.0	169203.0
21	Casa Loma	96.0	1480.0	450.0	165047.0
64	Kingsway South	15.0	1085.0	335.0	144642.0
70	Leaside-Bennington	56.0	2740.0	1125.0	125564.0
9	Bedford Park-Nortown	39.0	5210.0	1185.0	123077.0
137	Yonge-St.Clair	97.0	2330.0	645.0	114174.0
3	Annex	95.0	6485.0	2400.0	112766.0

Table 3 Toronto neighborhood profiles sort by Chinese population

index	Neighbourhood	Neighbourhood Number	Asian origins	Chinese	Total income: Average amount ($)
76	Milliken	130.0	23750.0	19140.0	28085.0
112	Steeles	116.0	21160.0	17835.0	31786.0
129	Willowdale East	51.0	36920.0	17240.0	45326.0
0	Agincourt North	129.0	24305.0	16950.0	30414.0
66	L'Amoreaux	117.0	30785.0	16745.0	31826.0
1	Agincourt South-Malvern West	128.0	17955.0	11455.0	31825.0
122	Waterfront Communities-The Island	77.0	24810.0	9790.0	70600.0
114	Tam O'Shanter-Sullivan	118.0	17925.0	9070.0	34200.0
52	Hillcrest Village	48.0	12030.0	8265.0	40442.0
6	Bay Street Corridor	76.0	15040.0	7585.0	56526.0

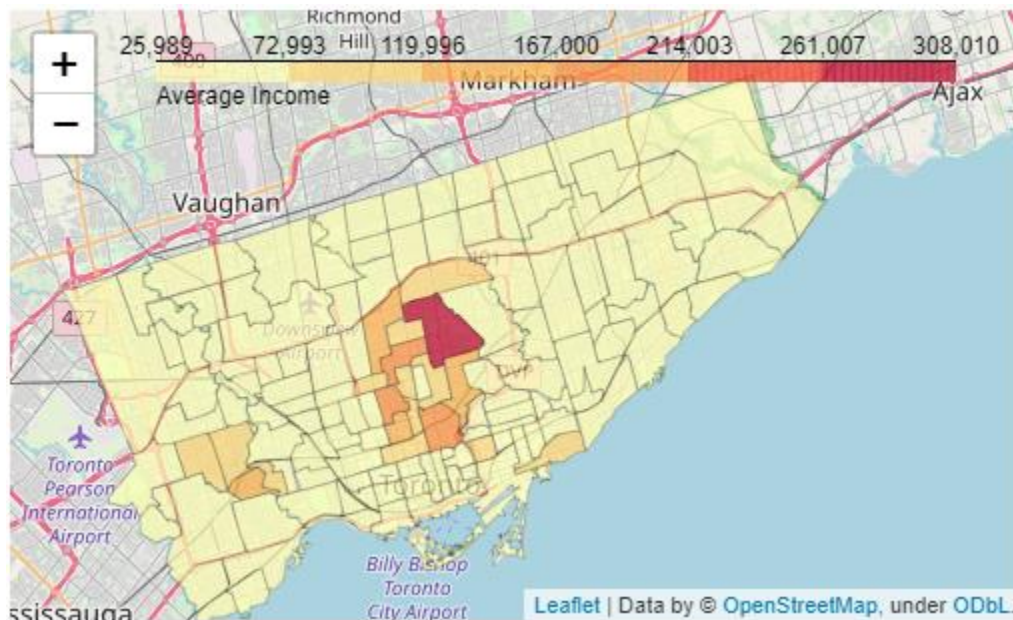Use maps to illustrate income and population distribution.

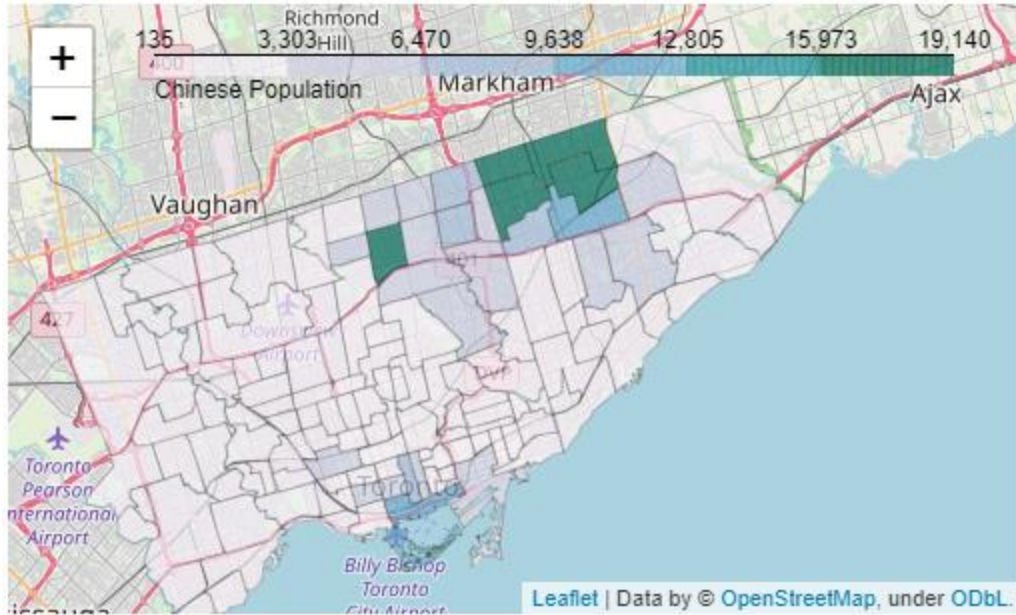Figure 1: Map of Toronto showing average income distribution

Figure 2: Map of Toronto showing Chinese Population distribution

To merge the census data with the post code data, data from Wikipedia are looked up using neighbourhood name from census table and inserted to census table. There are several names cannot be matched. The blank among the top 30 Chinese population neighbourhood are filled manually from googled data.

Table 4 Toronto Neighbourhood Population and Income with Postcode

index	Neighbourhood	Neighbourhood Number	Asian origins	Chinese	Total income: Average amount ($)	Postcode
76	Milliken	130.0	23750.0	19140.0	28085.0	M1V
112	Steeles	116.0	21160.0	17835.0	31786.0	M1V
129	Willowdale East	51.0	36920.0	17240.0	45326.0	M2M
0	Agincourt North	129.0	24305.0	16950.0	30414.0	M1V
66	L'Amoreaux	117.0	30785.0	16745.0	31826.0	M1V
1	Agincourt South|Malvern West	128.0	17955.0	11455.0	31825.0	M1S
122	Waterfront Communities|The Island	77.0	24810.0	9790.0	70600.0	M5P
114	Tam O'Shanter|Sullivan	118.0	17925.0	9070.0	34200.0	M1T
52	Hillcrest Village	48.0	12030.0	8265.0	40442.0	M2H
6	Bay Street Corridor	76.0	15040.0	7585.0	56526.0	M5G

Table 5: Grouped census data by Post code with population sum up and average income.

Postcode	Asian origins	Chinese	Total income: Average amount ($)
M1B	57160.0	7120.0	34564.500000
M1C	7175.0	1195.0	40972.000000
M1E	21485.0	2965.0	39605.666667
M1G	32850.0	4385.0	30878.000000
M1J	8120.0	665.0	32913.000000
M1K	29645.0	4535.0	32863.000000
M1L	7305.0	825.0	26793.000000
M1M	9895.0	2530.0	49539.000000
M1P	33855.0	7880.0	32474.000000
M1S	17955.0	11455.0	31825.000000

Use Foursquare API to get Chinese restaurant of each neighbourhoods.

Table 6: List of Chinese Restaurant with Neighbourhood

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	Lotus Pond Vegetarian Restaurant 蓮花素食	43.819421	-79.294682	Chinese Restaurant
1	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	Fishman Lobster Clubhouse Restaurant 魚樂軒	43.801909	-79.295409	Chinese Restaurant
2	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	Kenny's Noodles 聯記麵家	43.824484	-79.300542	Chinese Restaurant
3	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	Magic Noodle 大槐樹	43.814500	-79.294344	Chinese Restaurant
4	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	Alton Restaurant 益街坊	43.825582	-79.276038	Chinese Restaurant

Table 7: Toronto Neighbourhood sorted by Chinese restaurant count

	Neighbourhood	Chinese Restaurant Count
13	Church and Wellesley	100
50	The Annex, North Midtown, Yorkville	100
41	Little Portugal, Trinity	100
28	Harbourfront, Regent Park	100
9	Cabbagetown, St. James Town	100
11	Central Bay Street	100
12	Chinatown, Grange Park, Kensington Market	100
38	L'Amoreaux West	93
1	Agincourt North, L'Amoreaux East, Milliken, St...	89
8	CN Tower, Bathurst Quay, Island airport, Harbo...	85

Table 8: Full table of Toronto Neighbourhood with Postcode, census, and restaurant count

	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count
0	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	107920.0	75150.0	33871.4	89
1	M2M	North York	Newtonbrook, Willowdale	43.789053	-79.408493	62290.0	24790.0	41814.0	18
2	M1S	Scarborough	Agincourt	43.794200	-79.262029	17955.0	11455.0	31825.0	80
3	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	20830.0	10760.0	61487.2	25
4	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	28930.0	10755.0	36869.0	28
5	M5P	Central Toronto	Forest Hill North, Forest Hill West	43.696948	-79.411307	28080.0	10175.0	77849.5	12
6	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302	17925.0	9070.0	34200.0	68
7	M2H	North York	Hillcrest Village	43.803762	-79.363452	12030.0	8265.0	40442.0	26
8	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	33855.0	7880.0	32474.0	16
9	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	15040.0	7585.0	56526.0	100

Use K-Means clustering algorithm to analyze the neighbourhood, the clustering number is set to 5.

Table 9: Toronto Neighbourhood with cluster labels

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count
0	0	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	107920.0	75150.0	33871.4	89
1	0	M2M	North York	Newtonbrook, Willowdale	43.789053	-79.408493	62290.0	24790.0	41814.0	18
2	3	M1S	Scarborough	Agincourt	43.794200	-79.262029	17955.0	11455.0	31825.0	80
3	1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	20830.0	10760.0	61487.2	25
4	3	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	28930.0	10755.0	36869.0	28
5	1	M5P	Central Toronto	Forest Hill North, Forest Hill West	43.696948	-79.411307	28080.0	10175.0	77849.5	12
6	3	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302	17925.0	9070.0	34200.0	68
7	3	M2H	North York	Hillcrest Village	43.803762	-79.363452	12030.0	8265.0	40442.0	26
8	3	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	33855.0	7880.0	32474.0	16
9	1	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	15040.0	7585.0	56526.0	100

Discussion

Check 5 clusters independently. Cluster 0 has relatively low average income, however not bad Chinese population. Cluster 1 has middle range of income from 50k to 80k. Cluster 2 has highest average income and middle amount of Chinese population. Cluster 3 has lowest average income. Cluster 4 has middle high range of income, and middle range of Chinese population. Map the clusters with different color, which gives an idea of each cluster's location.

Table 10: 5 Clusters data separately

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count
0	0	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	107920.0	75150.0	33871.4	89
1	0	M2M	North York	Newtonbrook, Willowdale	43.789053	-79.408493	62290.0	24790.0	41814.0	18
10	0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	57160.0	7120.0	34564.5	2

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count
3	1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	20830.0	10760.0	61487.2	25
5	1	M5P	Central Toronto	Forest Hill North, Forest Hill West	43.696948	-79.411307	28080.0	10175.0	77849.5	12
9	1	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	15040.0	7585.0	56526.0	100
12	1	M2K	North York	Bayview Village	43.786947	-79.385975	13845.0	6065.0	52035.0	21
15	1	M3B	North York	Don Mills North	43.745906	-79.352188	12025.0	4850.0	67757.0	9
18	1	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	10740.0	4305.0	53583.0	100
24	1	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	8255.0	2720.0	70623.0	85
28	1	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242	11785.0	2250.0	52787.0	4
32	1	M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	43.650943	-79.554724	5765.0	1515.0	73028.5	5
42	1	M6P	West Toronto	High Park, The Junction South	43.661608	-79.464763	3400.0	920.0	71204.0	7

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count
33	2	M2L	North York	Silver Hills, York Mills	43.757490	-79.374714	3005.0	1445.0	308010.0	9
34	2	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	3485.0	1300.0	207903.0	23

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count
2	3	M1S	Scarborough	Agincourt	43.794200	-79.262029	17955.0	11455.0	31825.0	80
4	3	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	28930.0	10755.0	36869.0	28
6	3	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302	17925.0	9070.0	34200.0	68
7	3	M2H	North York	Hillcrest Village	43.803762	-79.363452	12030.0	8265.0	40442.0	26
8	3	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	33855.0	7880.0	32474.0	16
11	3	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	9140.0	6250.0	37422.0	100
13	3	M1W	Scarborough	L'Amoreaux West	43.799525	-79.318389	9515.0	5795.0	36346.0	93
14	3	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750	7770.0	4915.0	48215.5	100
16	3	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029	29645.0	4535.0	32863.0	11
17	3	M1G	Scarborough	Woburn	43.770992	-79.216917	32850.0	4385.0	30878.0	8

|	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count
20	4	M2P	North York	York Mills West	43.752758	-79.400049	8615.0	3890.0	100516.0	11
27	4	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678	6485.0	2400.0	112766.0	100
38	4	M5M	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750	5210.0	1185.0	123077.0	7
39	4	M4G	East York	Leaside	43.709060	-79.363452	2740.0	1125.0	125564.0	10
47	4	M4E	East Toronto	The Beaches	43.676357	-79.293031	2680.0	765.0	92580.0	9
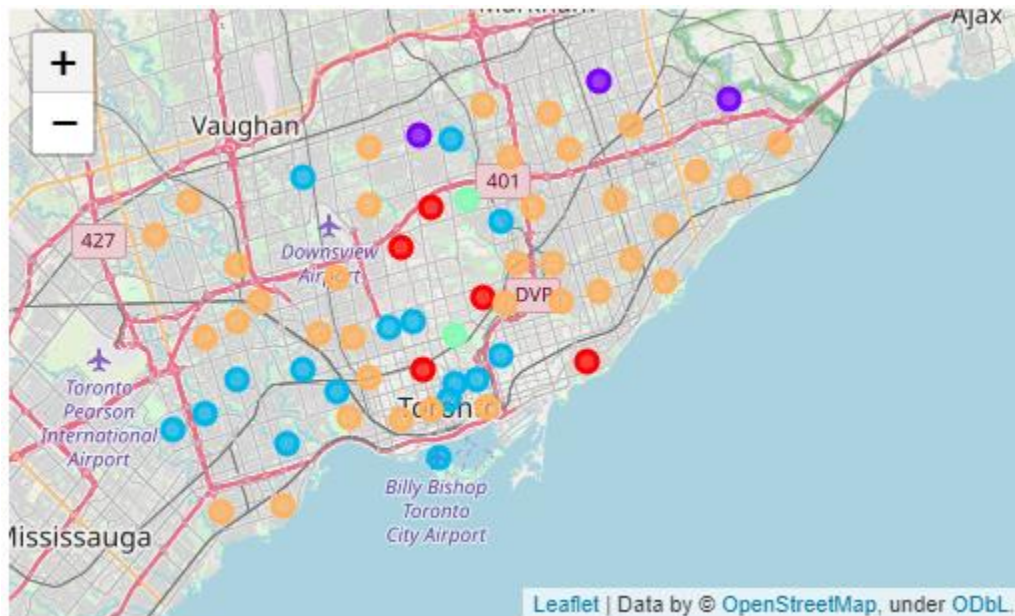
Figure 3: Divide Toronto neighbourhood to 5 clusters based on the population and income

Table 11: Explanation of Clusters

	Cluster	Meaning	Color
0	cluster 0	low income, high population	Purple
1	cluster 1	middle income	Blue
2	cluster 2	high income, middle low population	Turquoise
3	cluster 3	low income	Orange
4	cluster 4	middle high income, middle population	Red

To make a good profit, find the place has higher income is necessary. Thus cluster 2 and cluster 4 are reviewed further to narrow down the target location. Besides all the known data, a ratio of restaurant per person is calculated. Lower value means this area needs more restaurants.

Table 12: Cluster 2 and Cluster 4 data sorted by Restaurant per person ratio increasing

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count	Restaurant/Person
20	4	M2P	North York	York Mills West	43.752758	-79.400049	8615.0	3890.0	100516.0	11	0.002828
38	4	M5M	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750	5210.0	1185.0	123077.0	7	0.005907
33	2	M2L	North York	Silver Hills, York Mills	43.757490	-79.374714	3005.0	1445.0	308010.0	9	0.006228
39	4	M4G	East York	Leaside	43.709060	-79.363452	2740.0	1125.0	125564.0	10	0.008889
47	4	M4E	East Toronto	The Beaches	43.676357	-79.293031	2680.0	765.0	92580.0	9	0.011765
34	2	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	3485.0	1300.0	207903.0	23	0.017692
27	4	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678	6485.0	2400.0	112766.0	100	0.041667

From Table 12, it seems neighbourhood "York Mills West" with postcode "M2P" has the relatively low Chinese restaurant density and meanwhile has relatively high average income. The next two neighbourhoods are "Bedford Park, Lawrence Manor East" and "Silver Hills, York Mills". To confirm this result, all the neighbourhood is checked together with Restaurant per person ratio added.

Table 13: Full Toronto neighbourhood data with Restaurant per person

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count	Restaurant/Person
10	0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	57160.0	7120.0	34564.500000	2	0.000281
1	0	M2M	North York	Newtonbrook, Willowdale	43.789053	-79.408493	62290.0	24790.0	41814.000000	18	0.000726
23	3	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	21485.0	2965.0	39605.666667	3	0.001012
5	1	M5P	Central Toronto	Forest Hill North, Forest Hill West	43.696948	-79.411307	28080.0	10175.0	77849.500000	12	0.001179
0	0	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	107920.0	75150.0	33871.400000	89	0.001184

Narrow down data with some threshold setting. Eg: Restaurant/Person < 0.007 and Average income > 100k.

Table 14 Neighbourhood has low density of Chinese restaurant but high income

	Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Asian origins	Chinese	Total income: Average amount ($)	Chinese Restaurant Count	Restaurant/Person
20	4	M2P	North York	York Mills West	43.752758	-79.400049	8615.0	3890.0	100516.0	11	0.002828
38	4	M5M	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750	5210.0	1185.0	123077.0	7	0.005907
33	2	M2L	North York	Silver Hills, York Mills	43.757490	-79.374714	3005.0	1445.0	308010.0	9	0.006228

The result is same as what we investicated from Clusters. From Figure 4, we can see the location of these top three neighbourhoods.
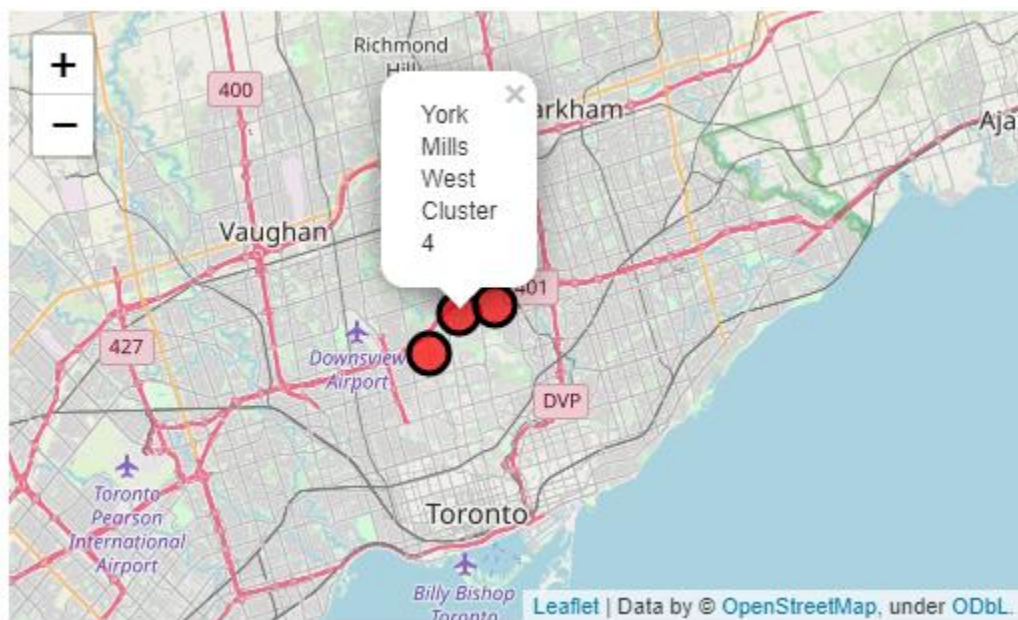

Figure 4: Neighbourhoods which are good to open a new high-class Chinese restaurant

Conclusion

Open a new Chinese restaurant could be very challenging. Picking a good location would greatly increase the possibility of succeeding. In this report, many data including Chinese population, average income, and current Chinese restaurant amount are investigated. The neighborhood "York Mills West" beat the other neighbourhoods and become a good candidate for a new high-class Chinese restaurant.