Open a Chinese Restaurant in Toronto

Capstone Project - The Battle of Neighborhoods

Wenlong Shen

Contents

1. Introduction

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario, while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

In 2016, Toronto's city proper had a population of 2,731,571; the urban area had a population of 5,429,524; the census metropolitan area had a population of 5,928,040; and the Greater Toronto Area metropolitan area had a population of 6,417,516. The city's foreign-born persons made up 47 per cent of the population, compared to 49.9 per cent in 2006. According to the United Nations Development Program, Toronto has the second-highest percentage of constant foreign-born population among world cities, after Miami, Florida. While Miami's foreign-born population has traditionally consisted primarily of Cubans and other Latin Americans, no single nationality or culture dominates Toronto's immigrant population, placing it among the most diverse cities in the world. In 2010, it was estimated over 100,000 immigrants arrive in the Greater Toronto Area each year.

1.1 Project Background

Toronto has a very strong restaurant industry. There are more than 8,100 restaurants, caterers, and bars and they generate $5.8 billion in revenues, which represents 6.5% of all businesses in Toronto. Torontonians also love to dine out. They dine out at an average of 3.1 times per week putting the city in the same rank as Chicago (2.8), New York City (3.0), and Las Vegas (3.3). The large population of immigrants from all over the globe has also made Toronto one of the most multicultural cities in the world. According to the 2016 census,51.5% of Toronto’s population is composed of visible minorities, compared to 49.1% in 2011, and 13.6% in 1981. In 2016, the most commonly reported ethnic origins of immigrants in Toronto overall were Chinese (332,830 or 12.5 percent).

1.2 Business Problem

With the information given in the project background, one promising business opportunity is to open a Chinese restaurant in Toronto. Thus, the aim of this report is to find ideal neighborhoods in Toronto to open a Chinese restaurant. There are many important factors when choosing a location to open a restaurant. This report has considered location, population, popularity of Chinese restaurant, income, proportion of Chinese, competitor, crime rate of each neighborhood.

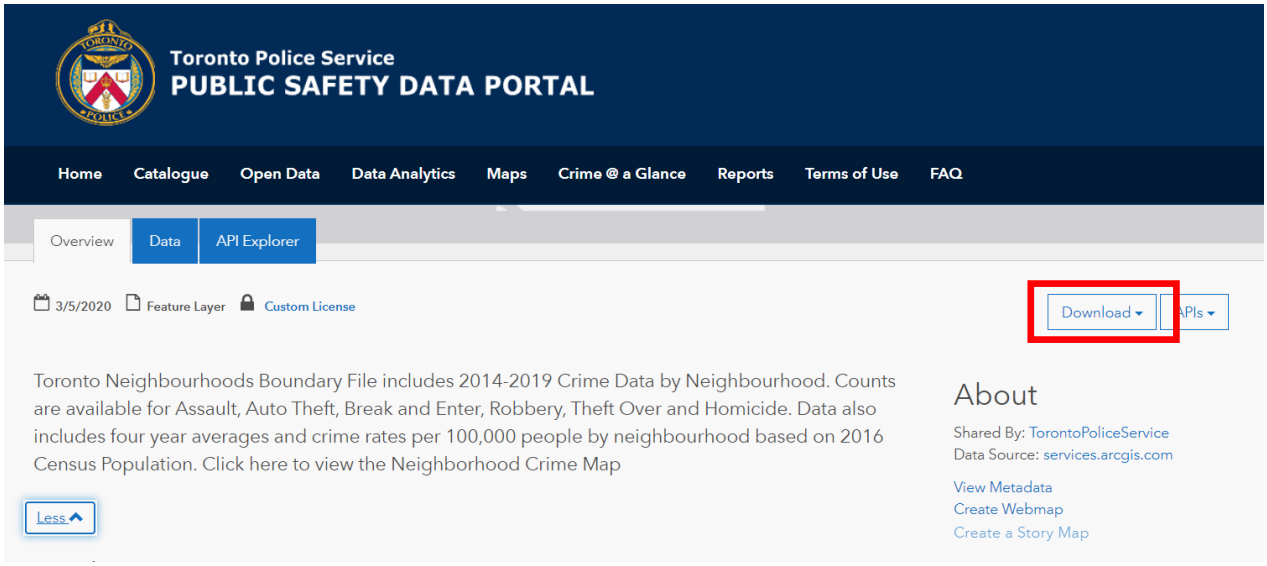
2. Data Sources

This report used five following data sets. The first two data sets are extremely large and cover almost all demographic information required for this report. The third data source provides information for nearby competitive restaurants. The fourth data set provides the coordinates of each neighborhood in Toronto. The fifth data set provides the neighborhood list of Toronto.

2.1 Neighborhood Crime Rate, Toronto Police Service

Toronto neighborhood crime rate data can be downloaded in Toronto police service website.

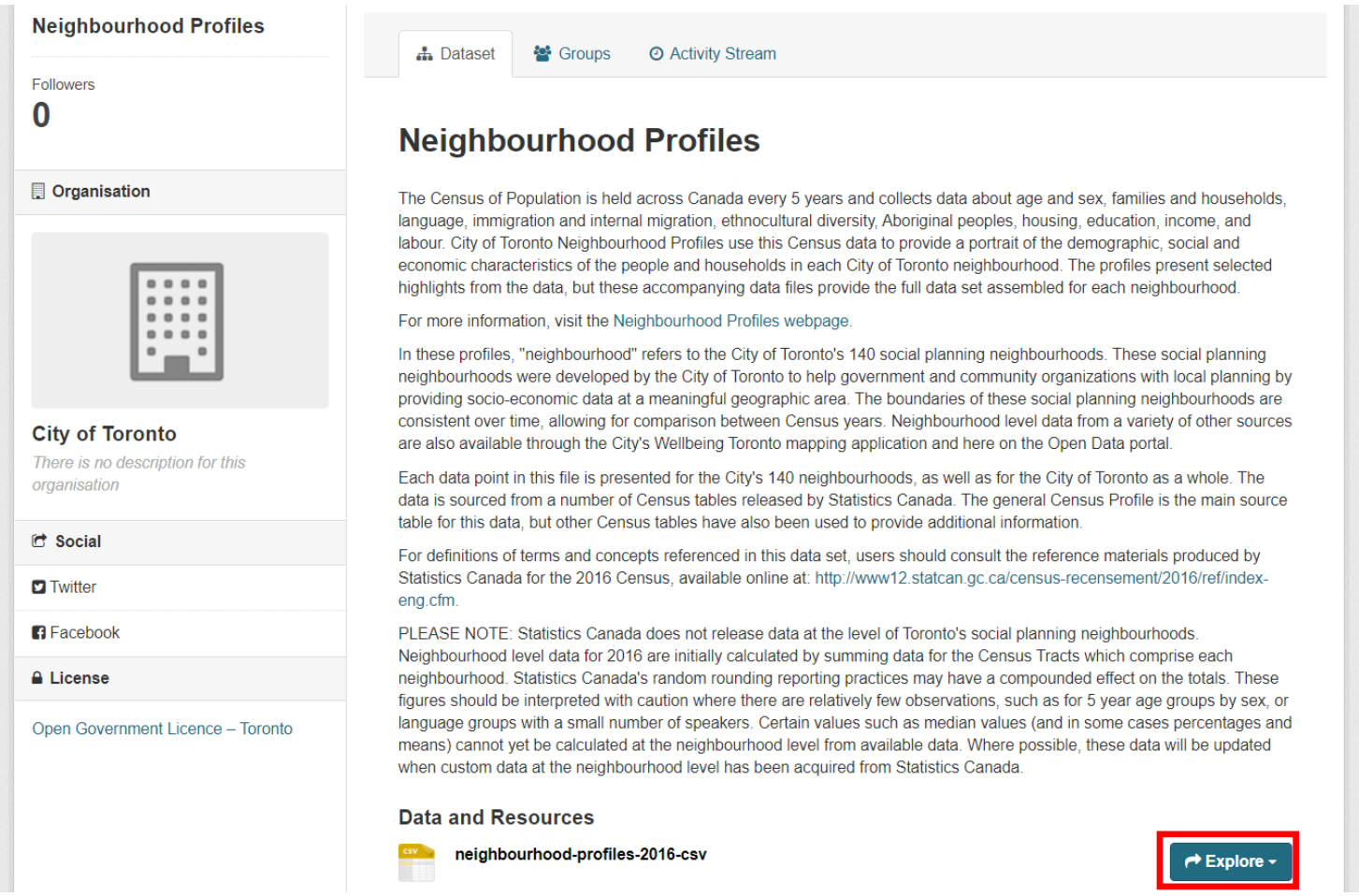
[Neighborhood Crime Rates (Boundary File) | Toronto Police Service Public Safety Data Portal]



2.2 2016 Neighborhood Profiles, Statistics Canada

Toronto 2016 Neighborhood Profiles data can be downloaded in CKAN Prod Toronto website.

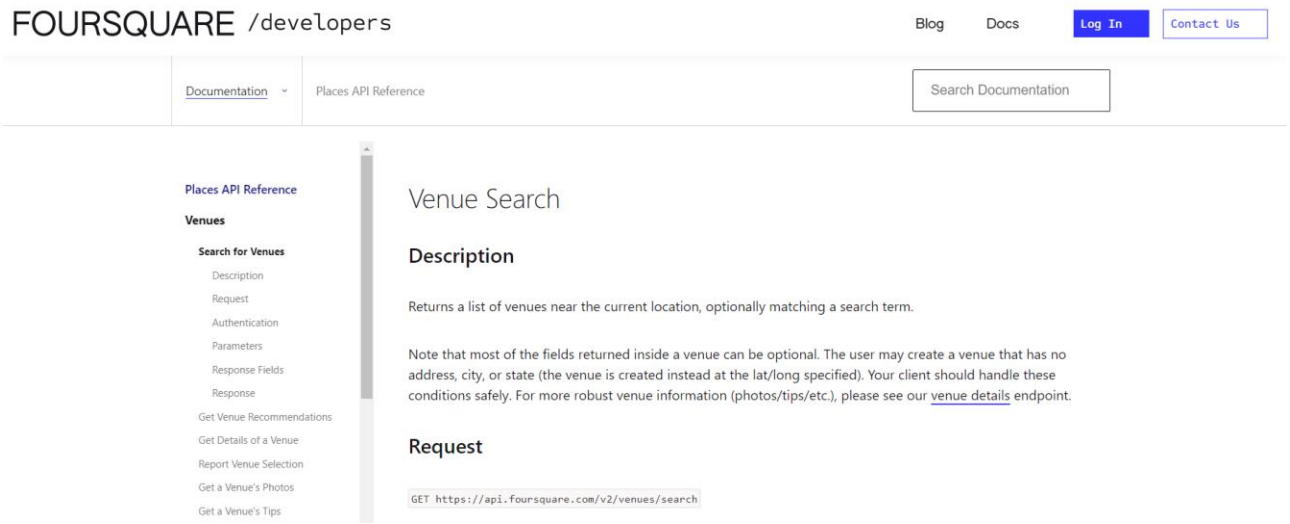
[Neighborhood Profiles - Datasets - CKAN (prod-toronto.ca)]



2.3 Foursquare Developer API

I used Foursquare developer API to search nearby venues. Below is the document for Foursquare developer.

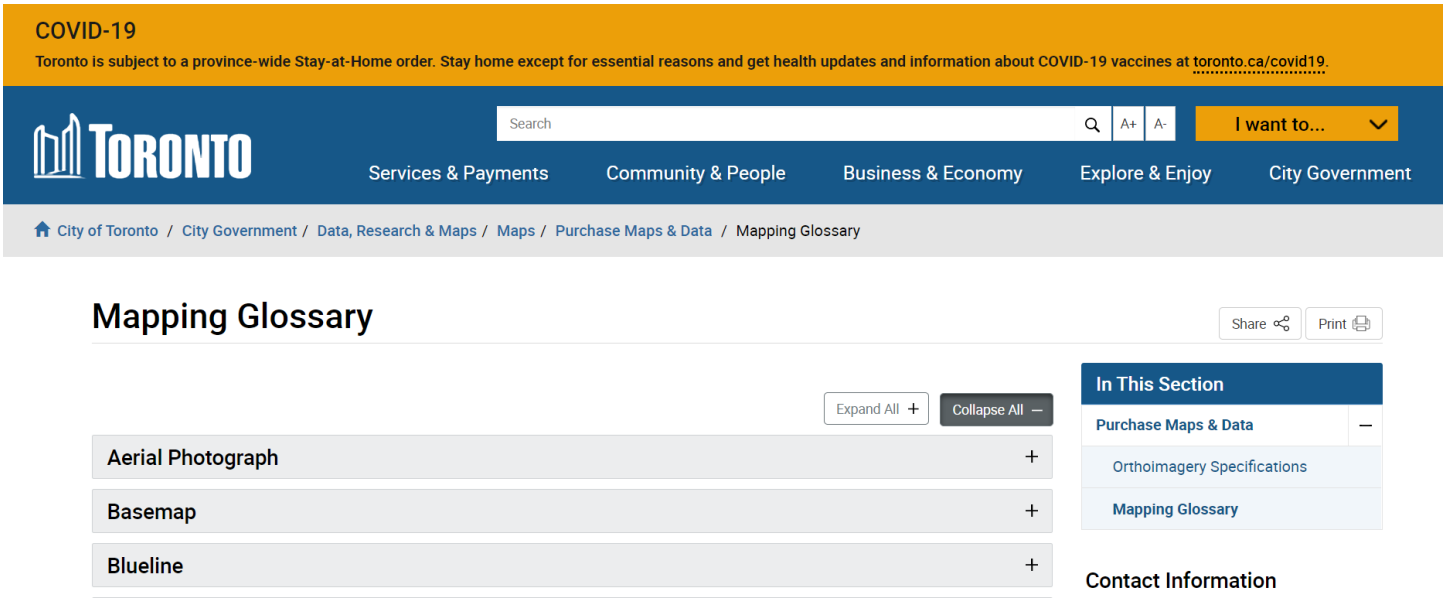
[Search for Venues | Places API Reference (foursquare.com)]



2.4 Geo-spatial coordinates of neighborhoods in Toronto

Geo-spatial coordinates of neighborhoods in Toronto can be downloaded in Mapping Glossary – City of Toronto website or Coursera Peer-graded Assignment: Segmenting and Clustering Neighborhoods in Toronto.

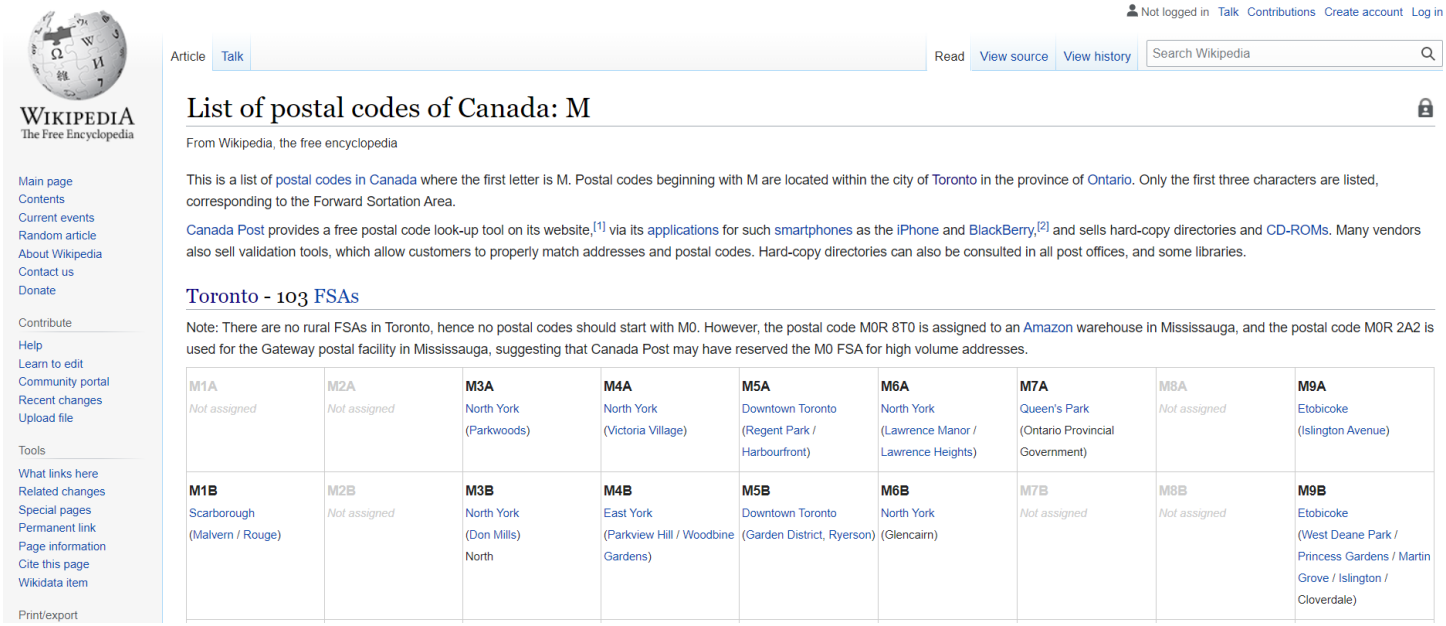
[Mapping Glossary – City of Toronto]



2.5 Wikipedia-List of postal codes of Canada: M

Toronto neighborhood data can be crawled in Wikipedia-List of postal codes of Canada: M.

[List of postal codes of Canada: M - Wikipedia]



3. Feature Engineering

3.1 Data Collection and Cleaning

Toronto neighborhood crime rate data, Toronto 2016 Neighborhood Profiles data and Geo-spatial coordinates of neighborhoods in Toronto are downloaded from csv files. Nearby venues data are acquired through Foursquare developer API and Toronto neighborhood data are crawled in Wikipedia-List of postal codes of Canada: M. All data mentioned above are well present in Python Pandas Data Frames.

1. Toronto 2016 Neighborhood Profiles contains massive amount of detailed information for each neighborhood in respect of population, families, language, labor, housing, income, ethnic origin, education, and mobility. However, this report only requires information in respect of population, income, ethnic, and ethnic origin. Thus, certain data are extracted by Python Pandas to pull out useful information.
2. Neighborhood Crime Rate data also contains massive amount of detailed information for each neighborhood. Similarly, only crime types that frequently happen in restaurants are considered. These crimes types are assault, auto theft, break and enter, robbery and theft over. These data are extracted and calculated by Python Pandas.
3. Foursquare API also return massive unique categories of nearby venues. However, not all these categories can be considered. As the objective is to open a Chinese restaurant, only categories related to Chinese restaurant and similar competitors are considered. Python Pandas are used to extract these certain data.

3.2 Feature Selection

In this report, below features are considered to decide where to open a restaurant in Toronto.

1. Population
2. Income
3. Crime rate
4. Proportion of Chinese
5. Competitor
6. Popularity of Chinese restaurant

4. Methodology and Analysis

The first part of this section presents visualizations of all considered features. The second part of this section combines all considered features together into one data frame and use K-Means clustering methodology to categorize neighborhoods.

4.1 Data Analysis and Visualization

4.1.1 Location

The Figure 1 Toronto Overview briefly shows the neighborhoods distribution in Toronto. Totally there are 103 neighborhoods.

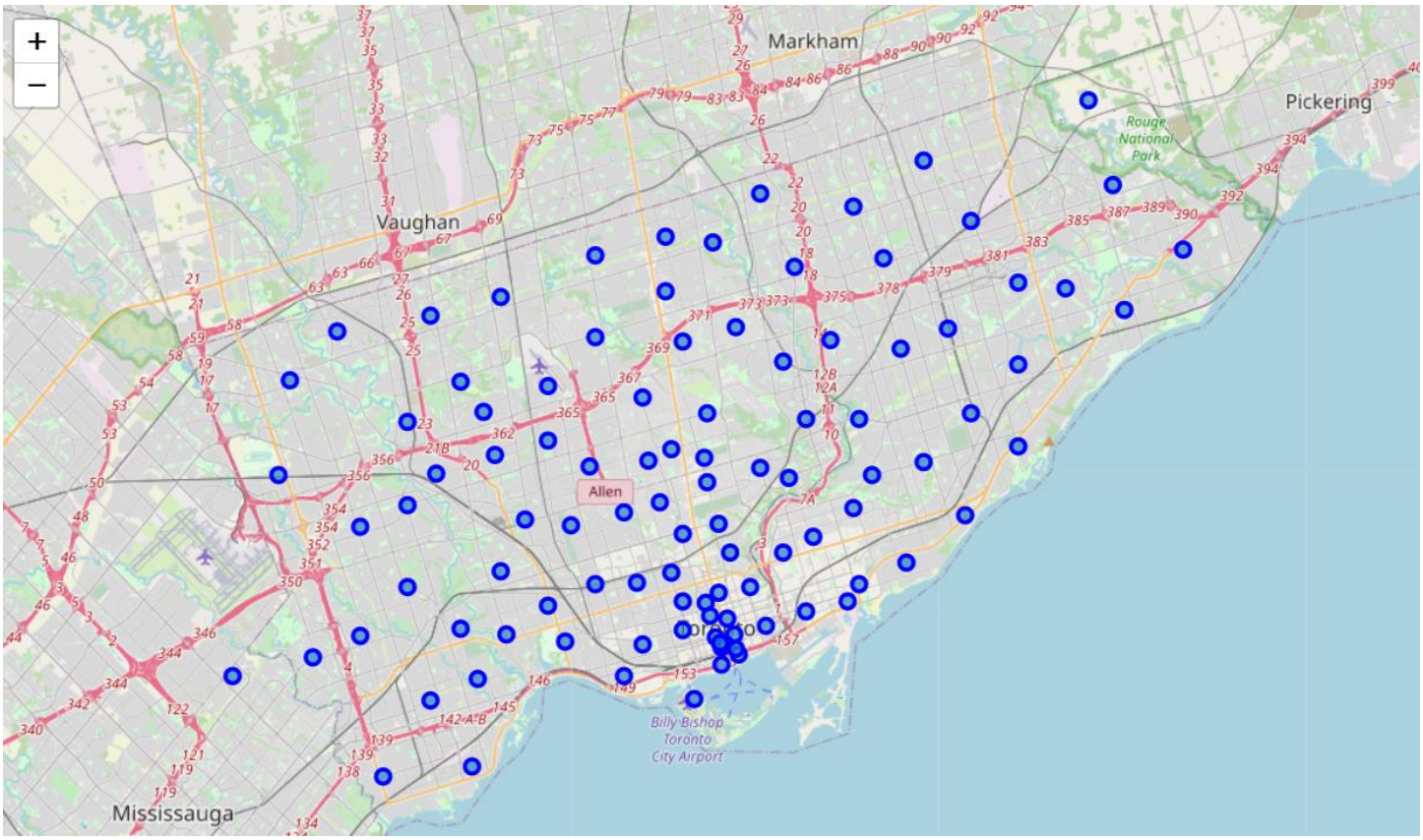
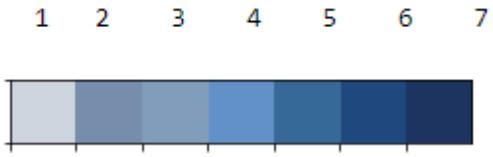


Figure 1 Toronto Overview

4.1.2 Population

The Figure 2 Population in each neighborhood shows the population in each neighborhood. There is one neighborhood that has a population more than 76000 and more than half neighborhoods have a population that less than 18000. The population density in Toronto is highly uneven.

```
1: 7049 => 18654
2: 18655 => 30179
3: 30180 => 41705
4: 41706 => 53230
5: 53231 => 64756
6: 64757 => 76281
7: 76282 => 87807
```



<Figure size 792x648 with 0 Axes>

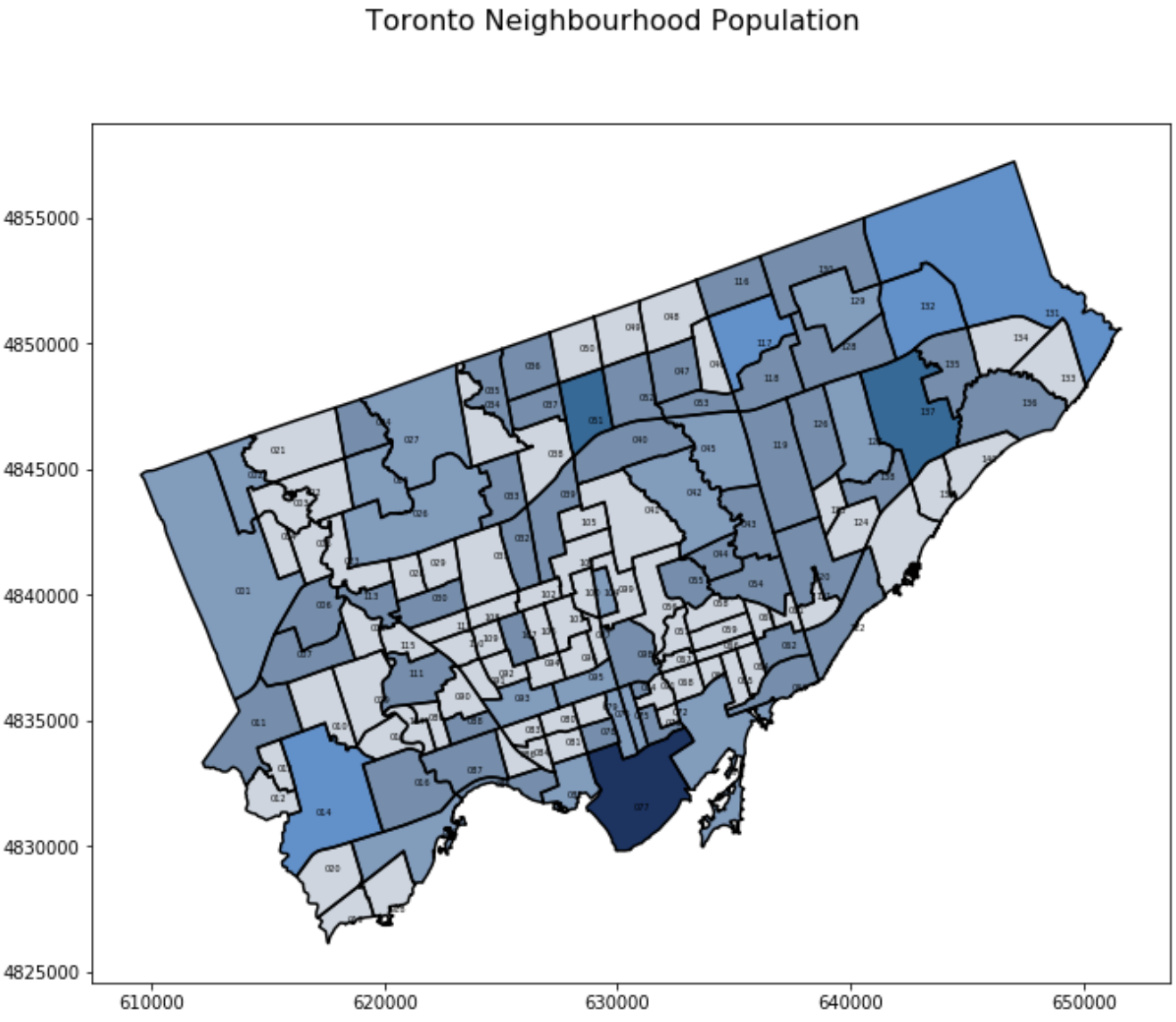


Figure 2 Population in each neighborhood

This report also investigate population in each age group and gender. Consider Figure 3, Figure 4 and Table 1.

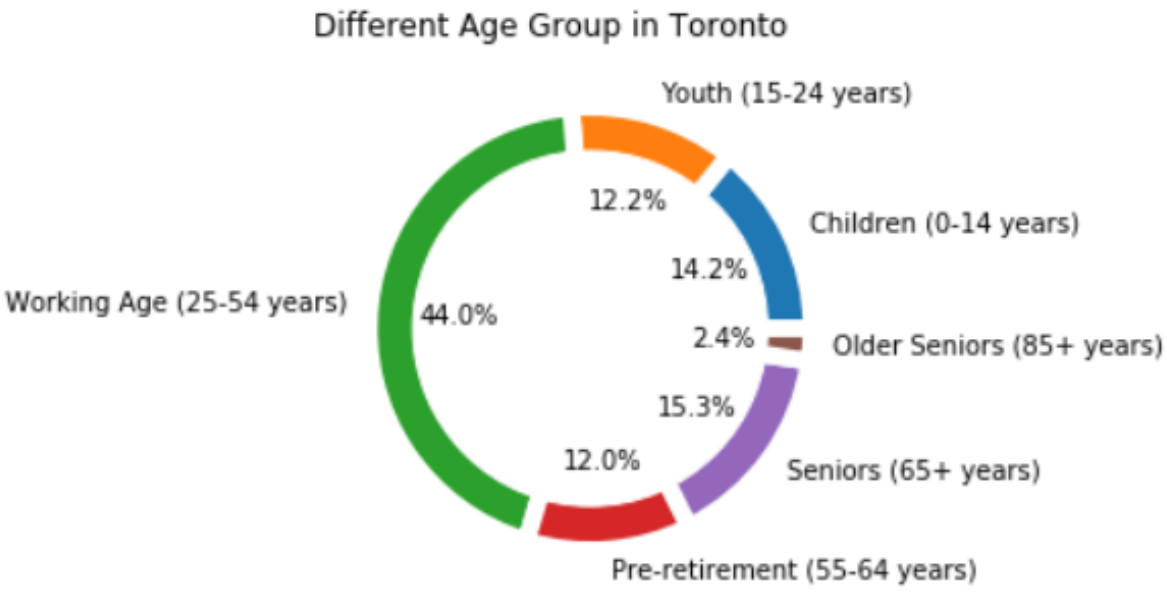


Figure 3 Different Age Group

In figure 4, it shows the most population in Toronto are in the working age group. It accounts for 44% of the total population. Toronto does not need to worry about population ageing for now. However, Toronto has 3.1% seniors than youth which is a sign for population ageing.

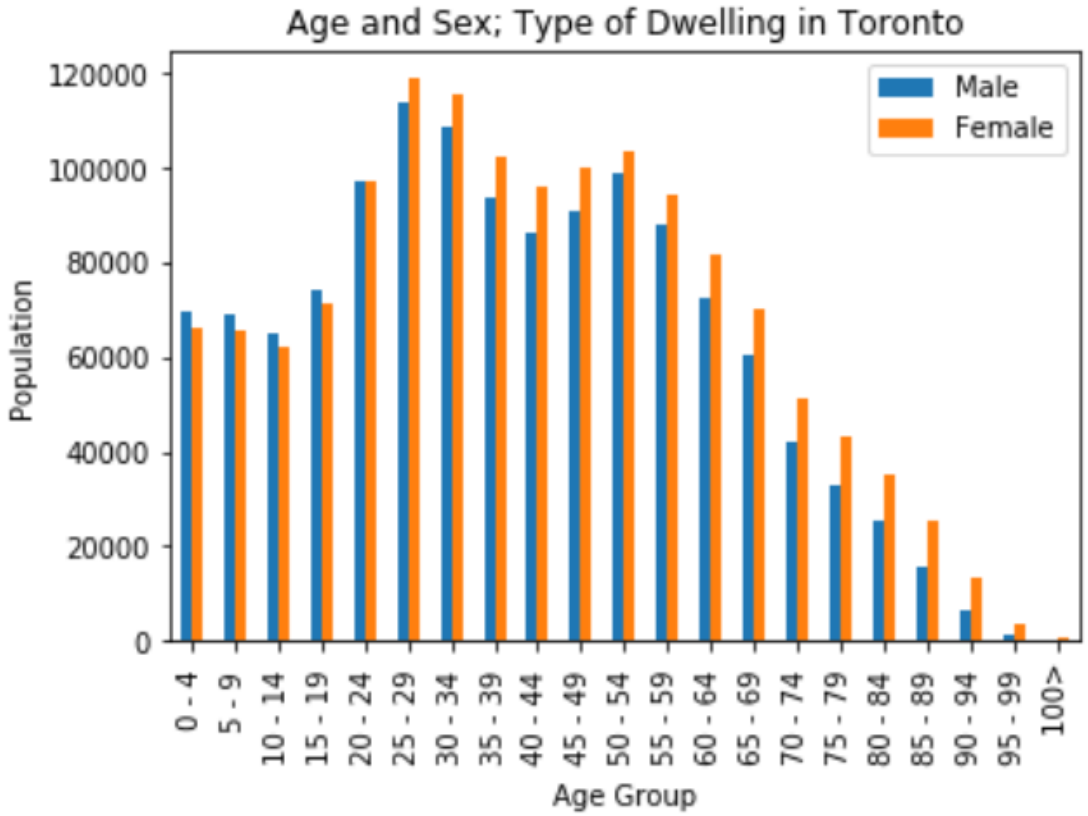


Figure 4 Age and Sex

Table 1 shows the first five row and the data frame that describes the male and female population in each age group and Figure 4 is the visualization of Table 1. Female population have exceeded the male population from age group 25-29.

	Age Group	Male	Female
0	0 - 4	69895.0	66105.0
1	5 - 9	69350.0	65680.0
2	10 - 14	64945.0	62165.0
3	15 - 19	74240.0	71285.0
4	20 - 24	97415.0	97330.0
5	25 - 29	113905.0	119040.0
6	30 - 34	108895.0	115675.0
7	35 - 39	94070.0	102240.0
8	40 - 44	86535.0	95860.0
9	45 - 49	90860.0	100065.0
10	50 - 54	98735.0	103670.0
11	55 - 59	88145.0	94660.0
12	60 - 64	72270.0	81600.0
13	65 - 69	60360.0	70180.0
14	70 - 74	42320.0	51285.0
15	75 - 79	32730.0	43430.0
16	80 - 84	25670.0	34965.0
17	85 - 89	15665.0	25135.0
18	90 - 94	6185.0	13500.0
19	95 - 99	1280.0	3475.0

Table 1 Male and Female

4.1.3 Crime Rate

This report also consider crime as a factor when choosing a location to open a restaurant. From the neighborhoods' crime dataset, this report has considered crime types of assault, auto theft,

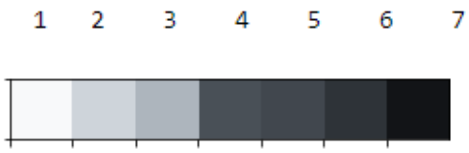
break and enter, robbery and theft over. The following Table 2 shows the average crime cases of the mentioned crime types and their percentage at each neighborhood.

	Neighbourh	crimes	average_crime_percentage
0	Yonge-St.Clair	11.000000	0.001917
1	York University Heights	120.000000	0.020917
2	Lansing-Westgate	33.333333	0.005810
3	Yorkdale-Glen Park	63.833333	0.011127
4	Stonegate-Queensway	34.833333	0.006072
5	Tam O'Shanter-Sullivan	40.666667	0.007088
6	The Beaches	31.500000	0.005491
7	Thistletown-Beaumond Heights	20.500000	0.003573
8	Thorncliffe Park	27.166667	0.004735
9	Danforth East York	19.833333	0.003457

Table 2 Crime at each neighborhood

The Figure 5 shows the Choropleth Map of Table 2. From the choropleth map, it is can be seen that the neighbourhood "Waterfront Communities-The Island" with most popluation also have the most number of crime cases. The good news is more than half of the neighbourhoods have less than 38 crime cases per year.

```
1: 8 => 38
2: 39 => 68
3: 69 => 98
4: 99 => 128
5: 129 => 158
6: 159 => 188
7: 189 => 218
```




```
<Figure size 792x648 with 0 Axes>
```

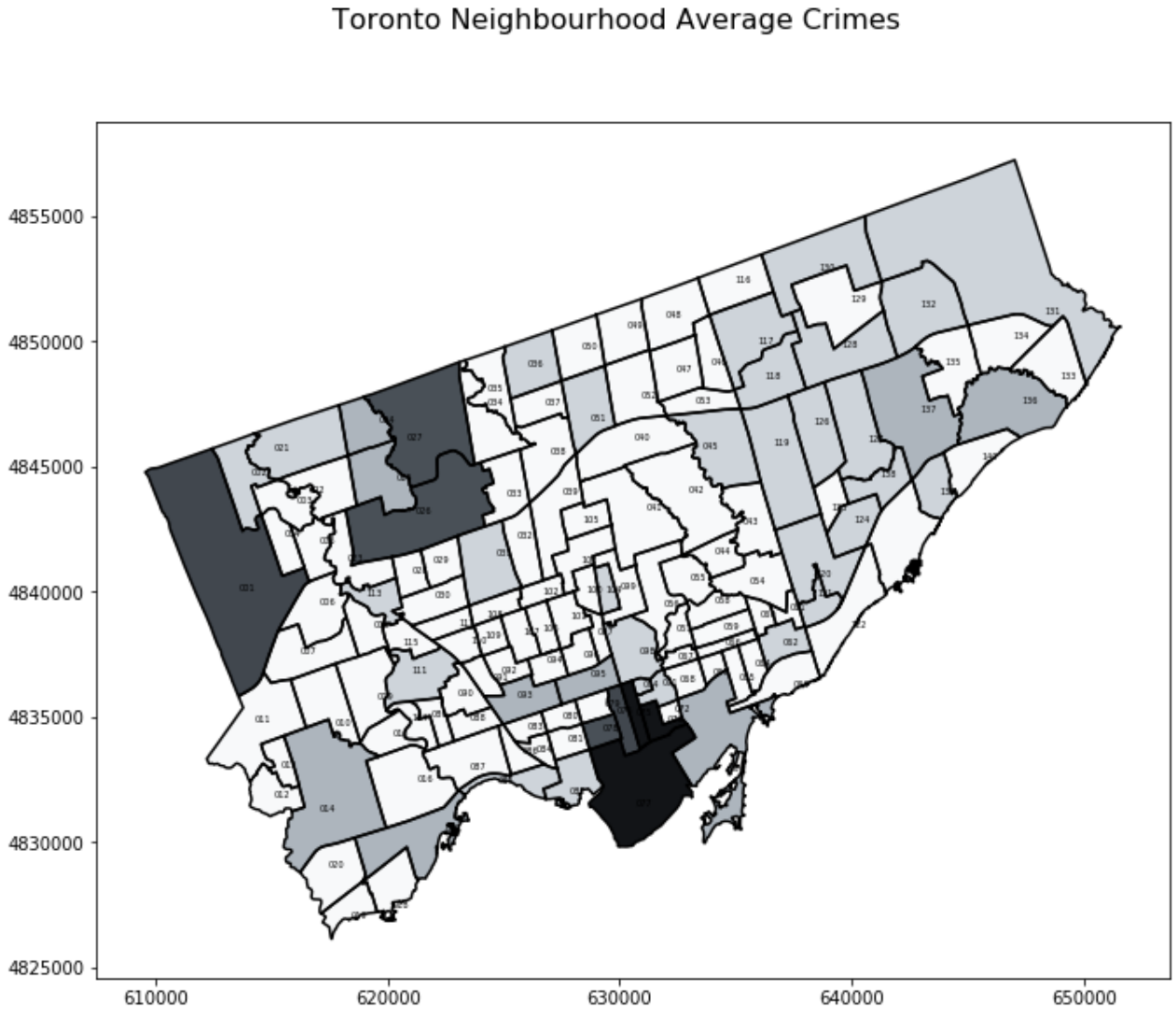


Figure 4 Neighbourhood Average Crime

4.1.4 Income

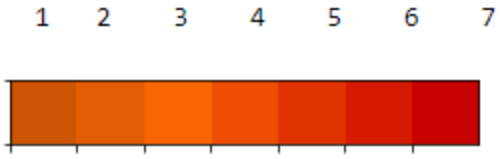
The report also investigates the average income of each neighborhood. Consider the Table 3 and Figure 6 below. Table 3 shows the income percentage of each neighborhood and Figure 6 shows the choropleth map of average income.

	income
Agincourt North	0.00390571
Agincourt South-Malvern West	0.00408691
Alderwood	0.0061267
Annex	0.0144812
Banbury-Don Mills	0.00870123
Bathurst Manor	0.00589902
Bay Street Corridor	0.00725897
Bayview Village	0.00668224
Bayview Woods-Steeles	0.00606725
Bedford Park-Nortown	0.0158053

Table 3 Income percentage of each neighborhood

One interesting discover is that “St.Andrew-Windfields” has the highest average income more than 270000. It's higher than any other neighborhoods. The average income of each neighborhood in Toronto is highly uneven.

```
1: 25706 => 66276
2: 66277 => 106565
3: 106566 => 146854
4: 146855 => 187142
5: 187143 => 227431
6: 227432 => 267720
7: 267721 => 308009
```




```
<Figure size 792x648 with 0 Axes>
```

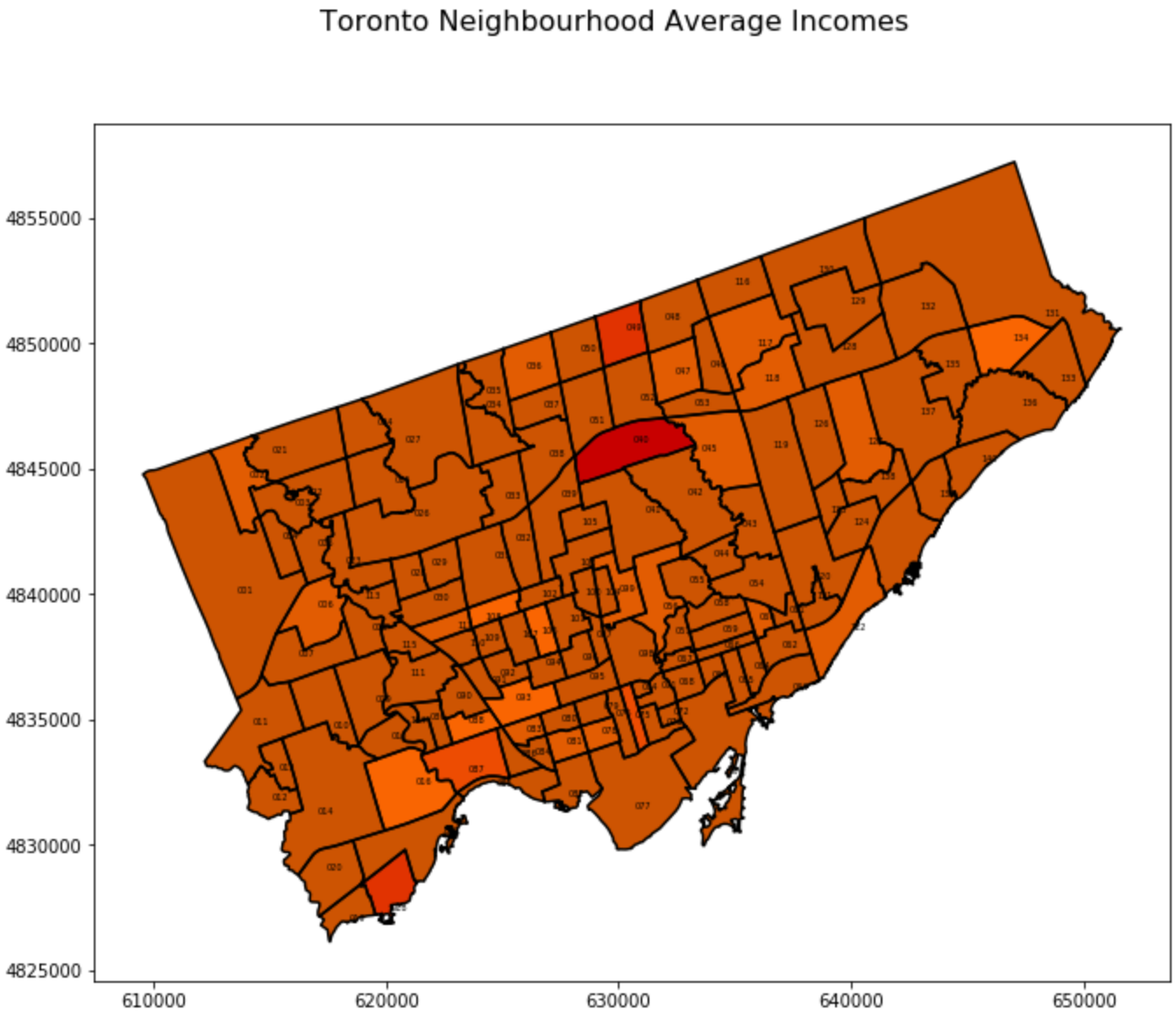


Figure 6 Neighbourhood Income

4.1.5 Proportion of Chinese

As the objective is to open a Chinses restaurant, so this report especially considers the population of Chinese. As the Figure 7 is too large, you can also see it on the Jupiter notebook or below link. From the Figure 7, through all neighborhoods, Chinese population almost doubled or tripled the population of other immigrants.

IBM-Data-Science-Capstone-Project/fig7.pdf at main · shenwl1014/IBM-Data-Science-Capstone-Project (github.com)

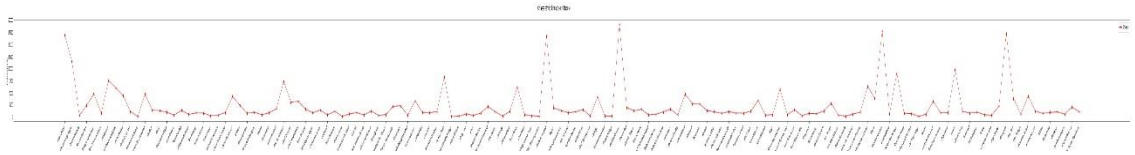


Figure 7 Chinese population in each neighbourhood

4.1.6 Popularity of Chinese Restaurant

To find a location for open a Chinese restaurant, the popularity of Chinese Restaurant must be considered. Thus, this report also uses Foursquare API to extract Chinese Restaurant and competitive restaurants. By giving the coordinates of all neighborhoods, the Foursquare API return the nearby venues in the radius of 2 kilometers. Then, this report selects Chinese restaurant and out competitive restaurants such as Asian Restaurant, Doner Restaurant, Hakka Restaurant, Cantonese Restaurant, Hotpot Restaurant, Szechuan Restaurant, Tibetan Restaurant, Xinjiang Restaurant and so on. After I selected Chinese Restaurant and competitive restaurants, then I ranked them to see the most popular restaurant in each neighborhood. According to Table 4, for example, Chinese restaurant has the most popularity in “Clarks Corners, Tam O'Shanter, Sullivan” and “Clairville, Humberwood, Woodbine Downs, West Humber”.

	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
10	Cedarbrae	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant
11	Central Bay Street	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant
12	Christie	Asian Restaurant	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant
13	Church and Wellesley	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant
14	Clairville, Humberwood, Woodbine Downs, West H...	Chinese Restaurant	Asian Restaurant	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Cantonese Restaurant
15	Clarks Corners, Tam O'Shanter, Sullivan	Chinese Restaurant	Asian Restaurant	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Cantonese Restaurant
16	Cliffside, Cliffcrest, Scarborough Village West	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant
17	Commerce Court, Victoria Hotel	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant
18	Davisville	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant
19	Davisville North	Xinjiang Restaurant	Szechuan Restaurant	Hakka Restaurant	Doner Restaurant	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant

Table 4 Popularity of restaurant in each neighborhood

4.1.7 Competitive Restaurants

In the meantime, I calculated the 'visit frequency' of selected restaurant in each neighborhood. See the Table 5 below as an example. In "York Mills, Silver Hills", only there is no restaurant get positive 'visit frequency', all restaurants have zero frequency. This indicates that rest competitive restaurants are not so popular in "York Mills, Silver Hills" or there is no such restaurant in "York Mills, Silver Hills".

York Mills, Silver Hills

	venue	freq
0	Asian Restaurant	0.0
1	Cantonese Restaurant	0.0
2	Chinese Restaurant	0.0
3	Doner Restaurant	0.0
4	Hakka Restaurant	0.0
5	Szechuan Restaurant	0.0
6	Xinjiang Restaurant	0.0

Table 5 Visit frequency of restaurant in each neighborhood

4.2 K-Means Clustering

After collecting features includes population, crime rate, income, proportion of Chinese and popularity of Chinese restaurant and competitive restaurants, this report formed the following data frame (Table 6) for K-mean Clustering.

	Borough	Neighborhood	PostalCode	Latitude	Longitude	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
0	North York	Parkwoods	M3A	43.753259	-79.329656	0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	37931	0.005739	0.009294	0.010023
1	North York	Victoria Village	M4A	43.725882	-79.315572	0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	38981	0.003425	0.005111	0.019726
2	Downtown Toronto	Regent Park, Harbourfront	M5A	43.654260	-79.360636	1	0.0	0.00	0.00	0.0	0.0	0.0	0.0	7130	0.004176	0.014348	0.002324
3	North York	Lawrence Manor, Lawrence Heights	M6A	43.718518	-79.464763	0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	21196	0.003590	0.005262	0.020888
4	Queen's Park	Ontario Provincial Government	M7A	43.662301	-79.389494	1	0.0	0.00	0.00	0.0	0.0	0.0	0.0	17619	0.001007	0.018575	0.002934
5	Etobicoke	Islington Avenue	M9A	43.667856	-79.532242	3	0.0	0.00	0.05	0.0	0.0	0.0	0.0	12050	0.001262	0.006832	0.004067
6	Scarborough	Malvern, Rouge	M1B	43.806686	-79.194353	1	0.0	0.00	0.00	0.0	0.0	0.0	0.0	22545	0.001968	0.003792	0.004358
7	North York	Don Mills North	M3B	43.745906	-79.352188	5	0.0	0.05	0.00	0.0	0.0	0.0	0.0	14672	0.001352	0.021195	0.001975
8	East York	Parkview Hill, Woodbine Gardens	M4B	43.706397	-79.309937	0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	26115	0.050311	0.004087	0.008919
9	Downtown Toronto	Garden District, Ryerson	M5B	43.657162	-79.378937	1	0.0	0.00	0.00	0.0	0.0	0.0	0.0	10679	0.001412	0.013041	0.001569

Table 6 neighborhood profile mergered

4.2.1 Elbow method

Firstly, we use the elbow method to find the optimal number of K. From Figure 7, it can be seen that the optimal number of K should be around 6 - 10.

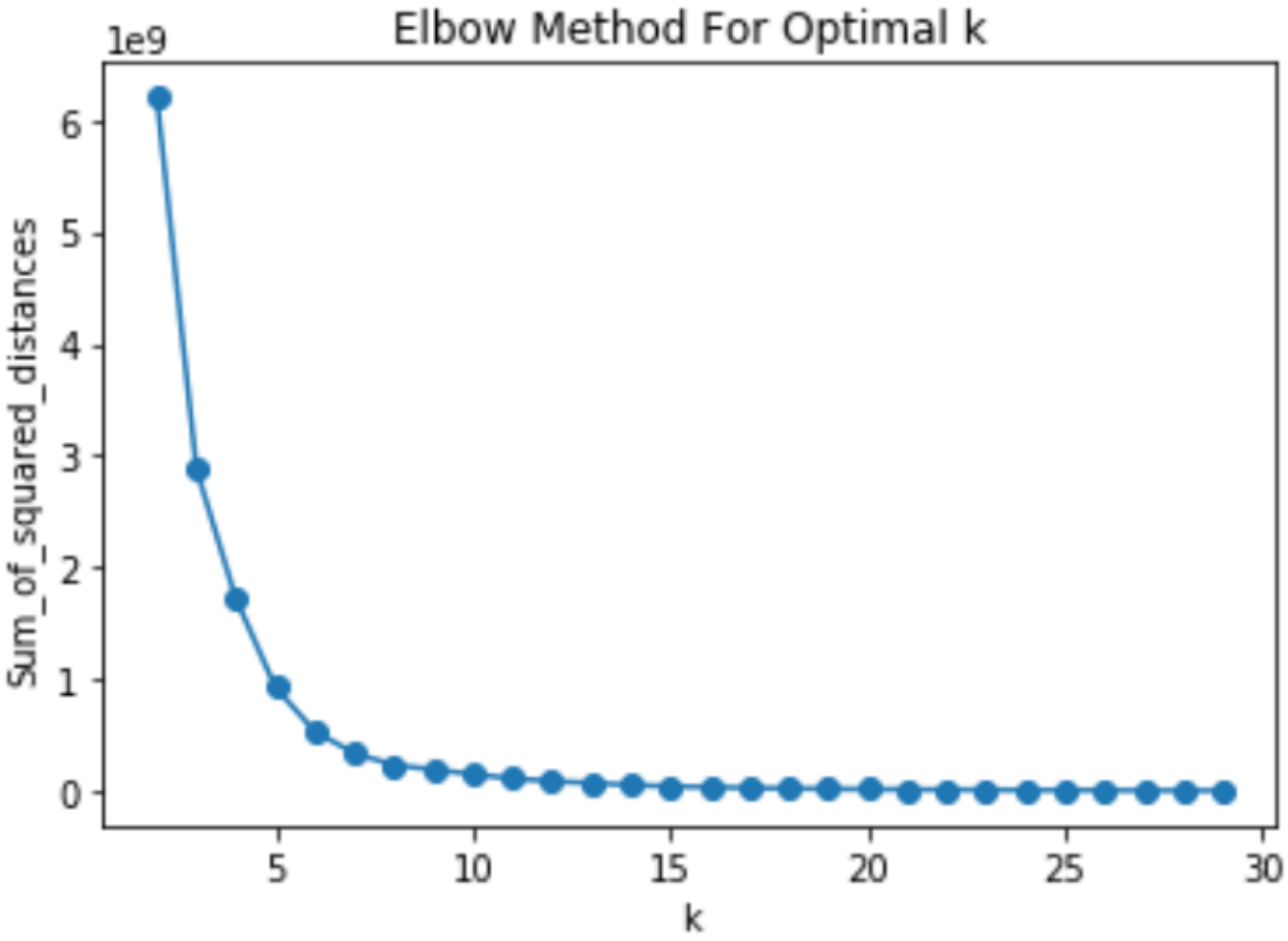


Figure 7 Elbow Method for optimal K

4.2.2 Kelbow Visualizer

In Figure 8, we have applied a Yellowbrick Kelbow Visualizer tool to help determine the optimal K. The cross point in Figure 8 shows that the optimal number of K is 6 – 8.

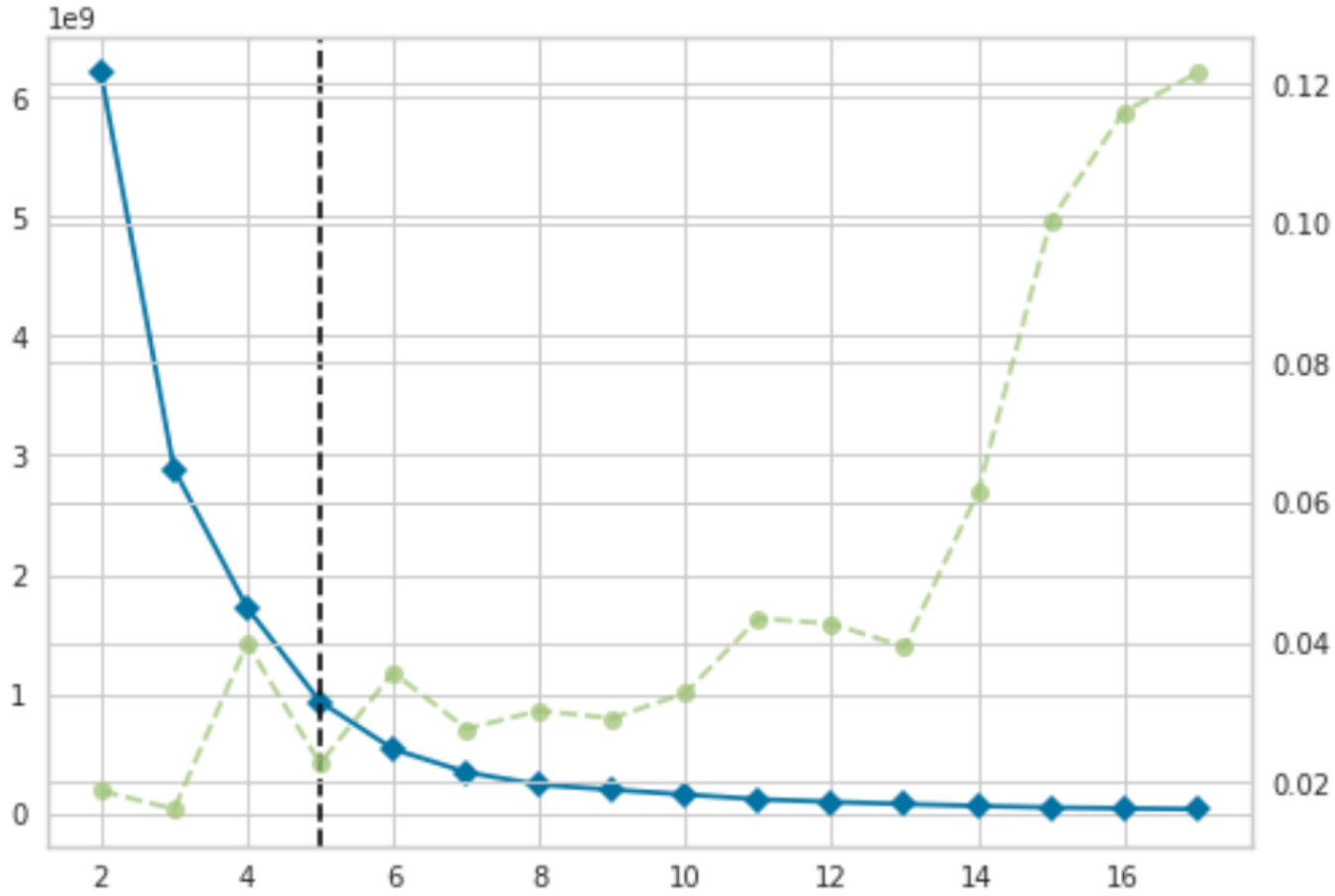


Figure 8 Yellowbrick Kelbow Visualizer for optimal K

4.2.3 Silhouette Score

In Figure 9, we have applied Silhouette score to further determine the optimal K. Silhouette score is highest when K is equal to 7. Hence, the optimal K is 7 and we are going to have 7 number of clusters.

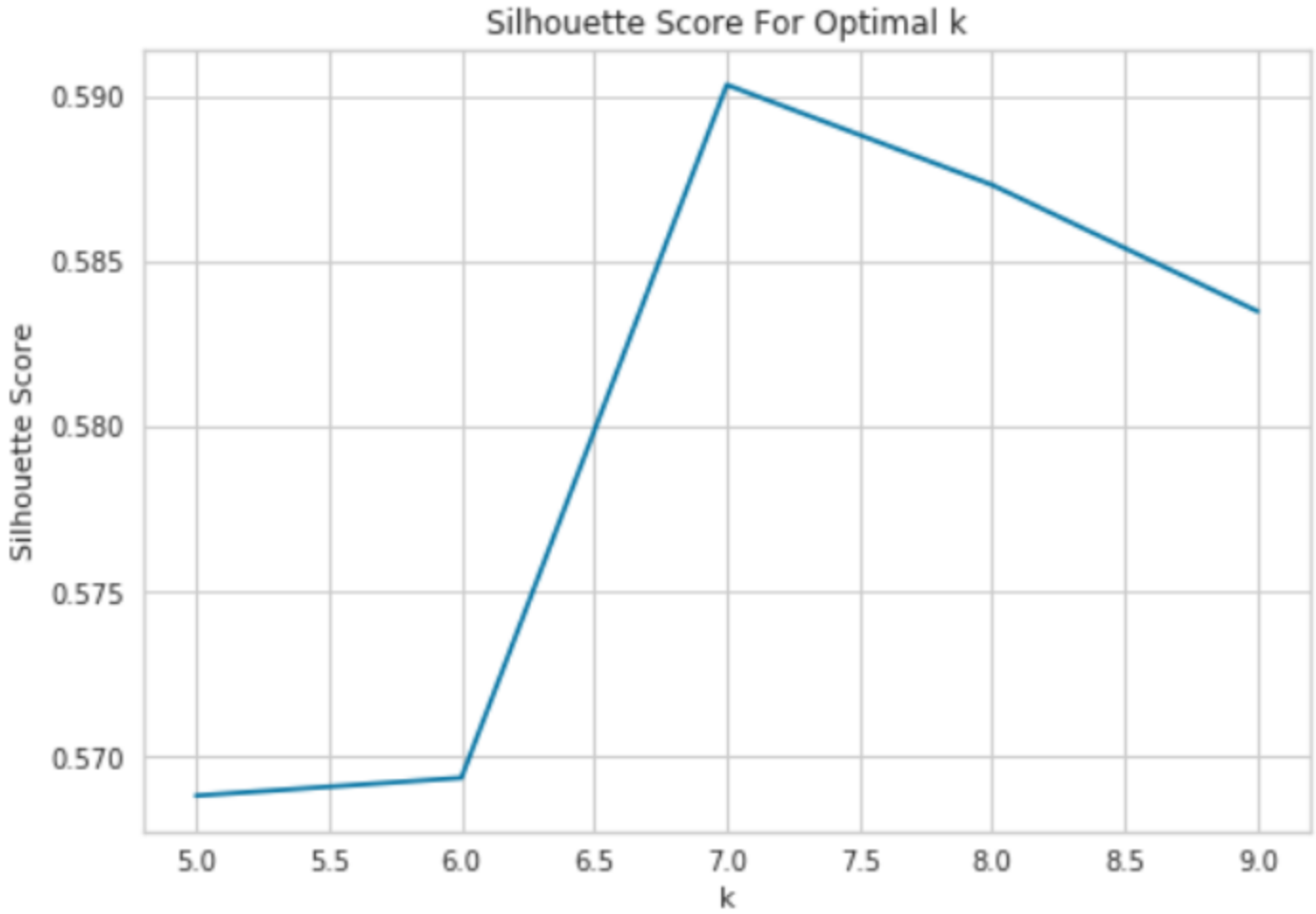


Figure 9 Silhouette score for optimal K

4.2.4 Silhouette Visualizer

Now, we have obtained the optimal K, we want to use Silhouette visualizer for verification. In Figure 10, when K is equal to 7, the Silhouette coefficient values is maximum. It verified that the optimal K is equal to 7.

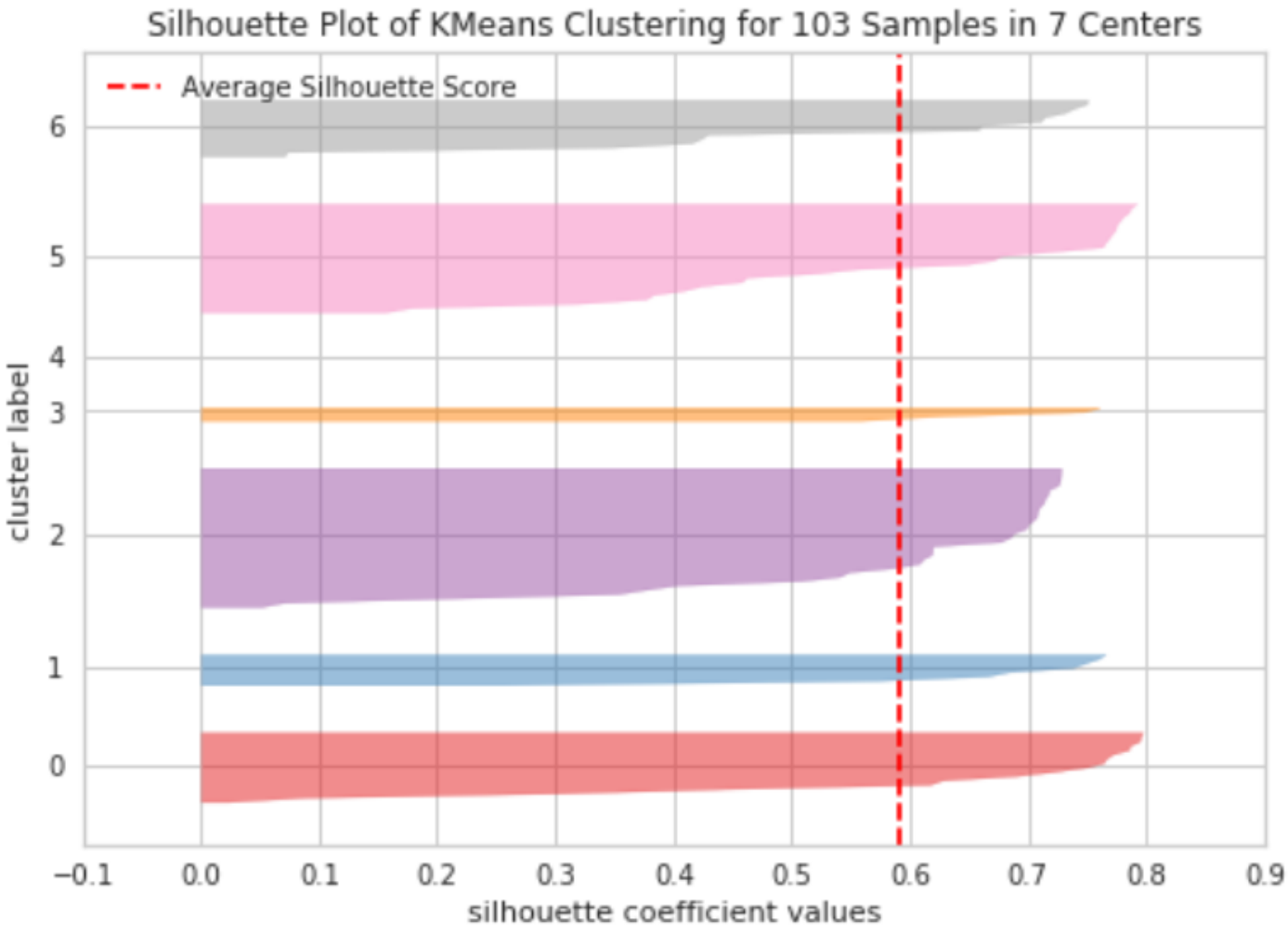


Figure 10 Silhouette Plot

5. Results

Currently, we have obtained the optimal K, which is equal to 7. We use K-Means Clustering to divide all neighborhoods into 7 clusters. The results are showed in below Table 7.

	Neighborhood	klabels
0	Agincourt	5
1	Alderwood, Long Branch	0
2	Bathurst Manor, Wilson Heights, Downsview North	1
3	Bayview Village	1
4	Bedford Park, Lawrence Manor East	1
5	Berczy Park	1
6	Birch Cliff, Cliffside West	1
7	Brockton, Parkdale Village, Exhibition Place	1
8	CN Tower, King and Spadina, Railway Lands, Har...	1
9	Caledonia-Fairbanks	1
10	Cedarbrae	4
11	Central Bay Street	0
12	Christie	1
13	Church and Wellesley	1
14	Clairville, Humberwood, Woodbine Downs, West H...	3

Table 7 neighborhood in each clusters

The Figure 11 shows the different clusters of neighborhoods in Toronto. Each color in this figure represents each cluster. There are 103 neighborhoods total.

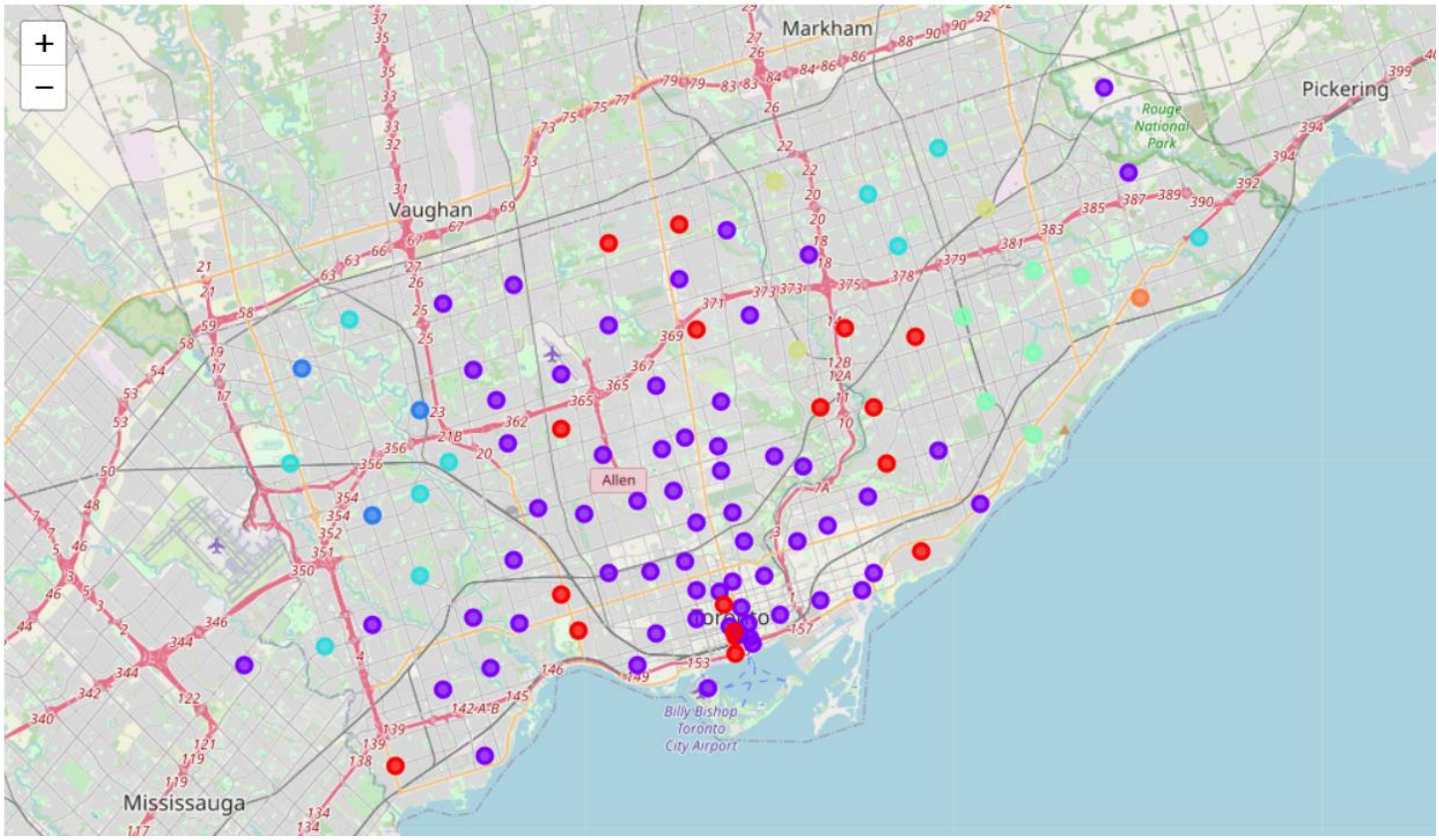


Figure 11 Neighborhood in each cluster

5.1 Cluster 0

In Table 8, it shows this cluster of neighborhoods has very high population, but it has no interest in Chinese restaurant. We do not recommend to open Chinese restaurant in these neighborhoods.

	Neighborhood	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
1	Alderwood, Long Branch	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	30277	0.034417	0.004087	0.020917
11	Central Bay Street	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	51481	0.014317	0.004271	0.015571
22	Don Mills South	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	39279	0.002929	0.006940	0.035297
34	First Canadian Place, Underground city	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	33031	0.002764	0.005212	0.015165
40	Harbourfront East, Union Station, Toronto Islands	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	29669	0.002914	0.006036	0.014351
41	High Park, The Junction South	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	37133	0.001427	0.004913	0.025681
51	Lawrence Manor, Lawrence Heights	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	21196	0.003590	0.005262	0.020888
65	Parkdale, Roncesvalles	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	40355	0.000661	0.009841	0.015310
66	Parkview Hill, Woodbine Gardens	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	26115	0.050311	0.004087	0.008919
67	Parkwoods	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	37931	0.005739	0.009294	0.010023
83	The Beaches	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	48519	0.005033	0.007325	0.008948
88	Toronto Dominion Centre, Design Exchange	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	32770	0.004161	0.009407	0.009703
91	Victoria Village	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	38981	0.003425	0.005111	0.019726
95	Wexford, Maryvale	0	0.05	0.0	0.0	0.0	0.0	0.0	0.0	57856	0.010441	0.005460	0.009965
97	Willowdale West	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	50790	0.001652	0.009096	0.008715
98	Willowdale, Newtonbrook	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	35772	0.017411	0.004667	0.009994
101	York Mills West	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	36333	0.001007	0.004420	0.010575

Table 8 Cluster 0 of neighborhoods

5.2 Cluster 1

In Table 9, it shows this cluster of neighborhoods has no interest in Chinese restaurant. We do not recommend to open Chinese restaurant in these neighborhoods.

	Neighborhood	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
2	Bathurst Manor, Wilson Heights, Downsview North	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	18146	0.001262	0.006127	0.005810
3	Bayview Village	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	17560	0.007211	0.014481	0.011127
4	Bedford Park, Lawrence Manor East	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	27410	0.014572	0.008701	0.006072
5	Berczy Park	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	29970	0.002479	0.005899	0.007088
6	Birch Cliff, Cliffside West	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	23364	0.022789	0.007259	0.005491
7	Brockton, Parkdale Village, Exhibition Place	1	0.05	0.0	0.0	0.00	0.0	0.0	0.0	10948	0.018223	0.006682	0.003573
8	CN Tower, King and Spadina, Railway Lands, Har...	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	23518	0.013460	0.006067	0.004735
9	Caledonia-Fairbanks	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	18427	0.003560	0.015805	0.003457
12	Christie	1	0.05	0.0	0.0	0.00	0.0	0.0	0.0	10353	0.004477	0.006981	0.003864
13	Church and Wellesley	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	10756	0.004131	0.003337	0.003167
17	Commerce Court, Victoria Hotel	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	20070	0.001758	0.005722	0.004154
18	Davisville	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	16651	0.002704	0.004171	0.005665
19	Davisville North	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	13458	0.002569	0.008092	0.008309
20	Del Ray, Mount Dennis, Keelsdale and Silverthorn	1	0.00	0.0	0.0	0.00	0.0	0.0	0.0	16794	0.000976	0.004509	0.006217

Table 9 Cluster 1 of neighborhoods

5.3 Cluster 2

In Table 10, it shows this cluster of neighborhoods has high interest on Chinese restaurant but also has some competitors, we can choose some neighborhoods to open a Chinese restaurant.

	Neighborhood	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
44	Humberlea, Emery	2	0.15	0.0	0.05	0.0	0.0	0.0	0.0	23318	0.001683	0.026264	0.005316
50	Kingsview Village, St. Phillips, Martin Grove ...	2	0.10	0.0	0.10	0.0	0.0	0.0	0.0	12334	0.002764	0.009144	0.003602
76	South Steeles, Silverstone, Humbergate, Jamest...	2	0.15	0.0	0.10	0.0	0.0	0.0	0.0	29864	0.057507	0.003607	0.011330

Table 10 Cluster 2 of neighborhoods

5.4 Cluster 3

In Table 11, it shows this cluster of neighborhoods has a certain level of interest on Chinese restaurant and less competitors, we can choose some neighborhoods to open a Chinese restaurant.

	Neighborhood	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
14	Clairville, Humberwood, Woodbine Downs, West H...	3	0.05	0.0	0.05	0.0	0.00	0.0	0.0	18116	0.003110	0.006230	0.007350
15	Clarks Corners, Tam O'Shanter, Sullivan	3	0.05	0.0	0.05	0.0	0.00	0.0	0.0	24212	0.001352	0.004465	0.010749
32	Eringate, Bloordale Gardens, Old Burnhamthorpe...	3	0.00	0.0	0.05	0.0	0.00	0.0	0.0	13072	0.009960	0.005103	0.003050
43	Humber Summit	3	0.05	0.0	0.05	0.0	0.00	0.0	0.0	18407	0.001157	0.010928	0.007931
47	Islington Avenue	3	0.00	0.0	0.05	0.0	0.00	0.0	0.0	12050	0.001262	0.006832	0.004067
56	Milliken, Agincourt North, Steeles East, L'Amo...	3	0.00	0.0	0.10	0.0	0.05	0.0	0.0	34680	0.001172	0.008382	0.017227
72	Rouge Hill, Port Union, Highland Creek	3	0.00	0.0	0.05	0.0	0.00	0.0	0.0	20595	0.000751	0.006085	0.010662
79	Steeles West, L'Amoreaux West	3	0.00	0.0	0.05	0.0	0.00	0.0	0.0	31618	0.004957	0.007566	0.004822
93	Westmount	3	0.10	0.0	0.05	0.0	0.00	0.0	0.0	32790	0.002524	0.006847	0.026059
94	Weston	3	0.05	0.0	0.05	0.0	0.00	0.0	0.0	23206	0.003786	0.007457	0.012463

Table 11 Cluster 3 of neighborhoods

5.5 Cluster 4

In Table 12, it shows this cluster of neighborhoods has high interest on Xinjiang restaurant but has less interest on Chinese Restaurant. We do not recommend to open Chinese restaurant in these neighborhoods.

	Neighborhood	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
10	Cedarbrae	4	0.00	0.0	0.00	0.0	0.0	0.0	0.05	15854	0.000511	0.004344	0.003660
16	Cliffside, Cliffcrest, Scarborough Village West	4	0.00	0.0	0.00	0.0	0.0	0.0	0.05	30225	0.004342	0.039554	0.013102
23	Dorset Park, Wexford Heights, Scarborough Town...	4	0.05	0.0	0.05	0.0	0.0	0.0	0.05	29302	0.012935	0.006881	0.009587
48	Kennedy Park, Ionview, East Birchmount Park	4	0.05	0.0	0.00	0.0	0.0	0.0	0.05	20156	0.010110	0.004669	0.005578
75	Scarborough Village	4	0.05	0.0	0.00	0.0	0.0	0.0	0.05	22588	0.000811	0.008010	0.005781
99	Woburn	4	0.00	0.0	0.00	0.0	0.0	0.0	0.05	12655	0.001637	0.012720	0.007059

Table 12 Cluster 4 of neighborhoods

5.6 Cluster 5

In Table 13, it shows this cluster of neighborhoods has a certain level of interest on Chinese restaurant but also has some competitors. We do not recommend to open Chinese restaurant in these neighborhoods.

	Neighborhood	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
0	Agincourt	5	0.05	0.05	0.05	0.0	0.0	0.00	0.0	14083	0.050927	0.003906	0.001917
21	Don Mills North	5	0.00	0.05	0.00	0.0	0.0	0.00	0.0	14672	0.001352	0.021195	0.001975
42	Hillcrest Village	5	0.00	0.05	0.05	0.0	0.0	0.05	0.0	28248	0.003861	0.003680	0.005142

Table 13 Cluster 5 of neighborhoods

5.7 Cluster 6

In Table 14, it shows this cluster of neighborhoods has very high population but has no interest in Chinese restaurant and has high crime rate. We do not recommend to open Chinese restaurant in these neighborhoods.

	Neighborhood	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
39	Guildwood, Morningside, West Hill	6	0.0	0.0	0.0	0.0	0.0	0.0	0.05	87808	0.002028	0.006181	0.038202

Table 14 Cluster 6 of neighborhoods

6. Discussion

According to K-Means clustering results we can find that cluster 2 and cluster 3 has a higher interest on Chinese restaurant and less competitors, so we can choose some neighborhoods to open a Chinese restaurant in these two clusters. Next, we can look at the remaining 4 features: population, crime rate, income and proportion of Chinese. By comparing 6 features together, we have tallied the scores of the neighborhoods in both cluster 2 and cluster 3 in Table 15. In Table 15, the higher population, income, popularity of Chinese restaurant and proportion of Chinese get the higher score, and the lower crime rate and less competitor get the higher score.

Index	Neighborhood	Popularity of Chinese Restaurant	Competitor	Population	Proportion of Chinese	Income	Crime Rate	Score
44	Humberlea, Emery	★★★	★★	★★★	★★	★★★★★	★★★★	19
50	Kingsview Village, St. Phillips, Martin Grove ...	★★★★★	★★★	★	★★	★★★	★★★★★	19
76	South Steeles, Silverstone, Humbergate, Jamest...	★★★★★	★★	★★★★	★★★★★	★	★★★	20
14	Clairville, Humberwood, Woodbine Downs, West H...	★★★	★★★★	★★	★★	★★	★★★★	17
15	Clarks Corners, Tam O'Shanter, Sullivan	★★★	★★★★	★★★	★★	★	★★★	16
32	Eringate, Bloordale Gardens, Old Burnhamthorpe...	★★★	★★★★	★	★★★	★★	★★★★★	18
43	Humber Summit	★★★	★★★★★	★★	★★	★★★	★★★★	19
47	Islington Avenue	★★★	★★★★★	★	★★	★★	★★★★★	18
56	Milliken, Agincourt North, Steeles East, L'Amo...	★★★★★	★★★★	★★★★★	★★	★★	★★★	21
72	Rouge Hill, Port Union, Highland Creek	★★★	★★★★★	★★	★	★★	★★★	16
79	Steeles West, L'Amoreaux West	★★★	★★★★★	★★★★★	★★	★★	★★★★★	22
93	Westmount	★★★	★★★	★★★★★	★★	★★	★★★	18
94	Weston	★★★	★★★★	★★★	★★	★★	★★★	17

Table 15 the scores of neighborhoods

7. Conclusion

In conclusion, based on the scores in Table 15, we selected the three neighborhoods with the highest scores as recommended locations to open a Chinese restaurant.

	Neighborhood	Borough	PostalCode	Latitude	Longitude	klables	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Doner Restaurant	Hakka Restaurant	Szechuan Restaurant	Xinjiang Restaurant	Population	Chinese	income	crime
56	Milliken, Agincourt North, Steeles East, L'Amo...	Scarborough	M1V	43.815252	-79.284577	3	0.00	0.0	0.10	0.0	0.05	0.0	0.0	34680	0.001172	0.008382	0.017227
76	South Steeles, Silverstone, Humbergate, Jamest...	Etobicoke	M9V	43.739416	-79.588437	2	0.15	0.0	0.10	0.0	0.00	0.0	0.0	29864	0.057507	0.003607	0.011330
79	Steeles West, L'Amoreaux West	Scarborough	M1W	43.799525	-79.318389	3	0.00	0.0	0.05	0.0	0.00	0.0	0.0	31618	0.004957	0.007566	0.004822

Table 16 Recommended neighborhoods

The Figure 12 shows the recommended neighborhoods to open a Chinese restaurant in Toronto.

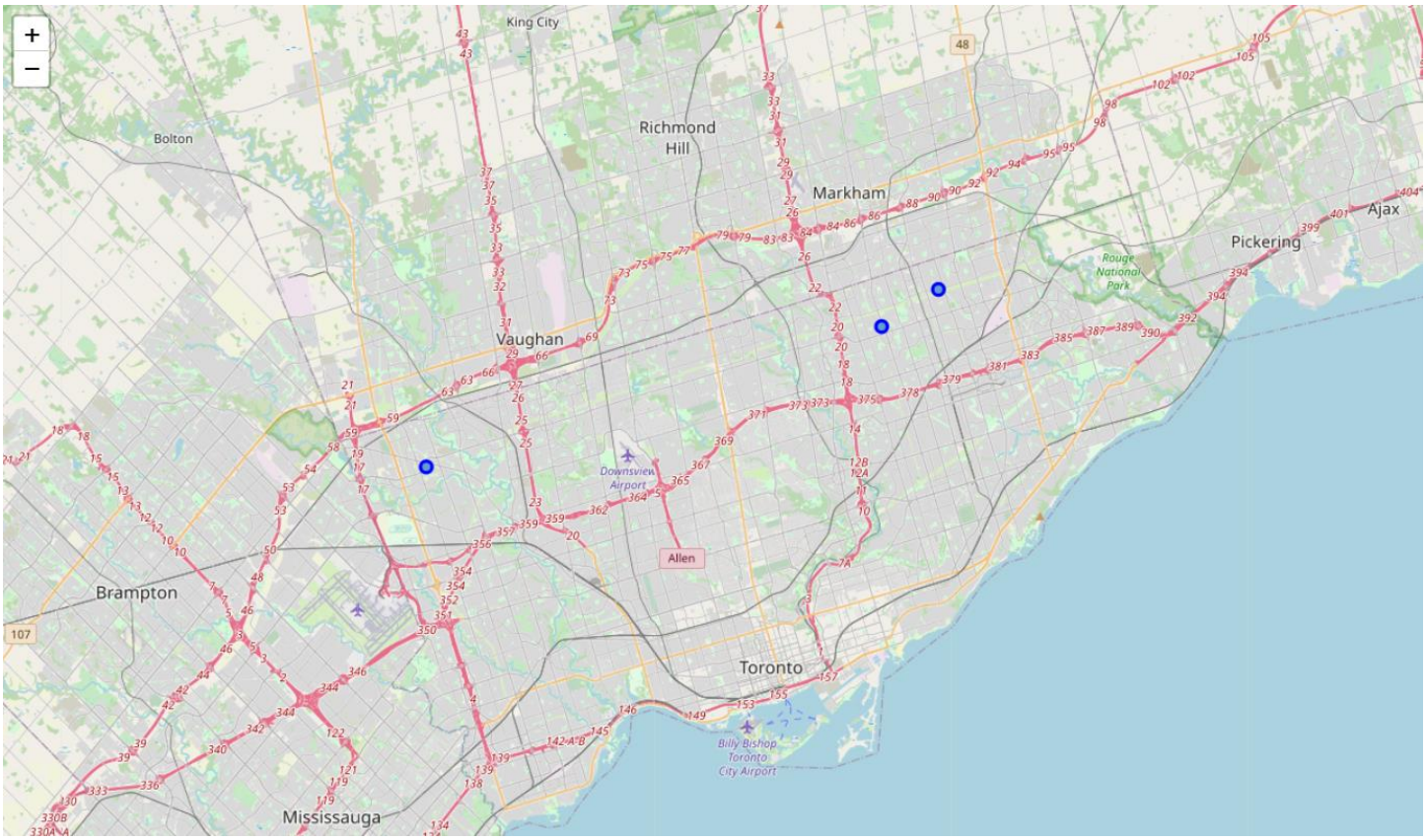


Figure 12 Recommended neighborhoods