Open a Chinese Restaurant in Toronto

Capstone Project - The Battle of Neighborhoods

Wenlong Shen

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). In 2016, the most commonly reported ethnic origins of immigrants in Toronto overall were Chinese (332,830 or 12.5 percent).

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.

CONTENT

PART 01
Data Sources

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. 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)

3. Foursquare Developer API

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

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

4. Geo-spatial coordinates of neighborhoods in Toronto

Geo-spatial coordinates of neighborhoods in Toronto can be downloaded in Mapping Glossary.

Mapping Glossary – City of Toronto

5. Wikipedia-List of postal codes of Canada: M

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

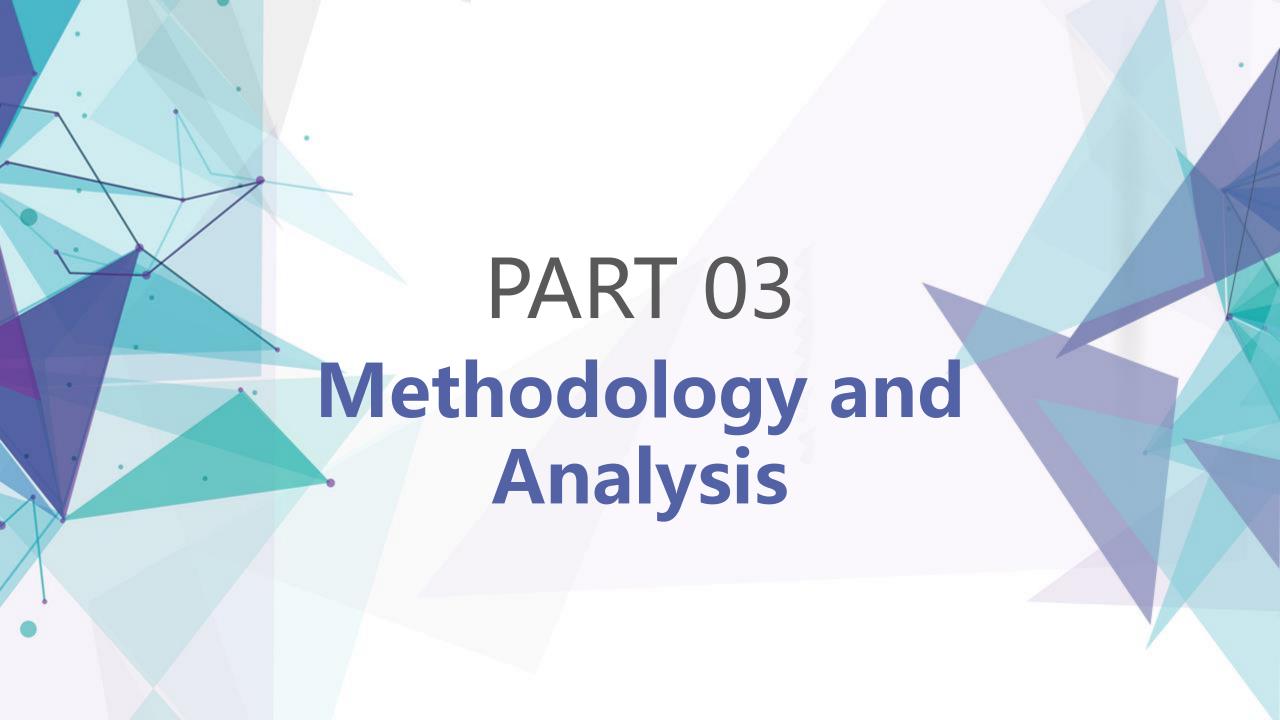List of postal codes of Canada: M - Wikipedia

PART 02
Feature Engineering

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.
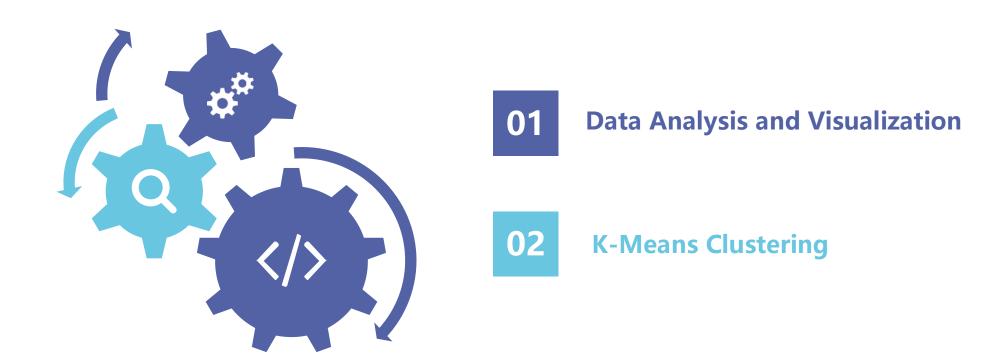
2. Feature Selection

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 Income
5. Competitor
6. Popularity of Chinese restaurant

PART 03
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.

01   **Data Analysis and Visualization**

02   **K-Means Clustering**

1. Location

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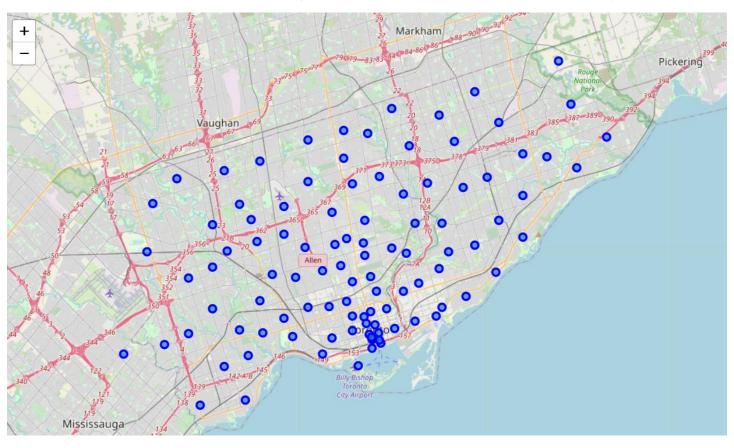

Figure 1 Toronto Overview

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.
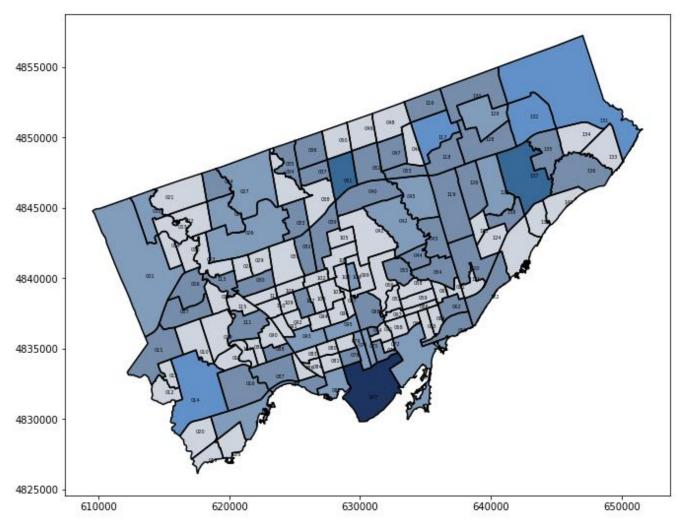

Figure 2 Population in each neighborhood

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. 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.

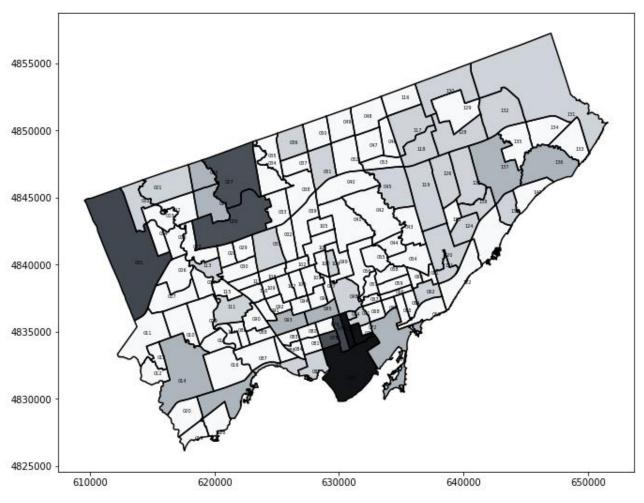*Figure 5 Neighbourhood Average Crime*

4. Income

The report also investigates the average income of each neighborhood. Consider the Figure below. 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.
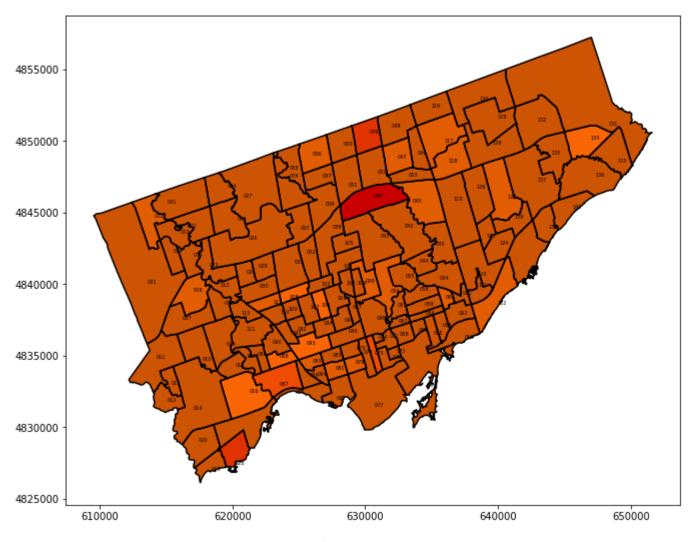
Figure 6 Neighbourhood Income

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

6. Popularity of Chinese Restaurant and Competitor

To find a location for open a Chinese restaurant, the popularity of Chinese Restaurant must be considered. Thus, this repor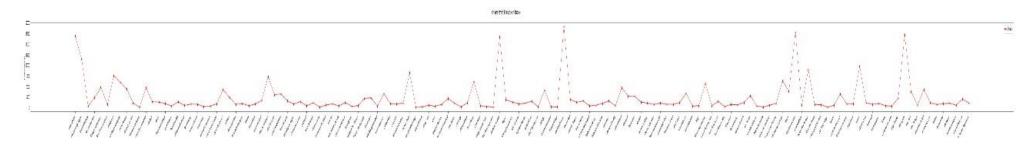t also uses Foursquare API to extract Chinese Restaurant and competitive restaurants. By giving the coordinates of all neighborhoods, the Foursquare API return the nearby venues in the radius of 2 kilometers. Then, this report selects Chinese restaurant and out competitive restaurants such as Asian Restaurant, Doner Restaurant, Hakka Restaurant, Cantonese Restaurant, Hotpot Restaurant, Szechuan Restaurant, Tibetan Restaurant, Xinjiang Restaurant and so on.

	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
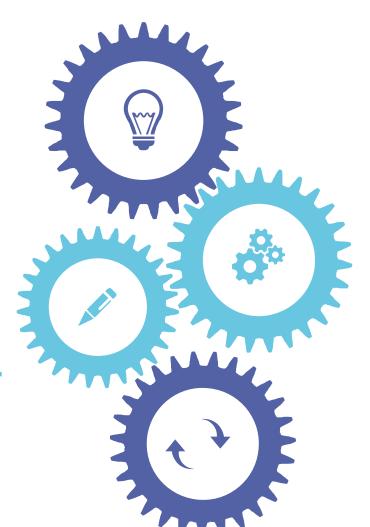
7. K-Means Clustering

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

Below methods are used to find the optimal number of K.

Elbow method

Silhouette Score

Kelbow Visualizer

Silhouette Visualizer

PART 04
Results

We have obtained the optimal K, which is equal to 7. We use K-Means Clustering to divide all neighborhoods into 7 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.
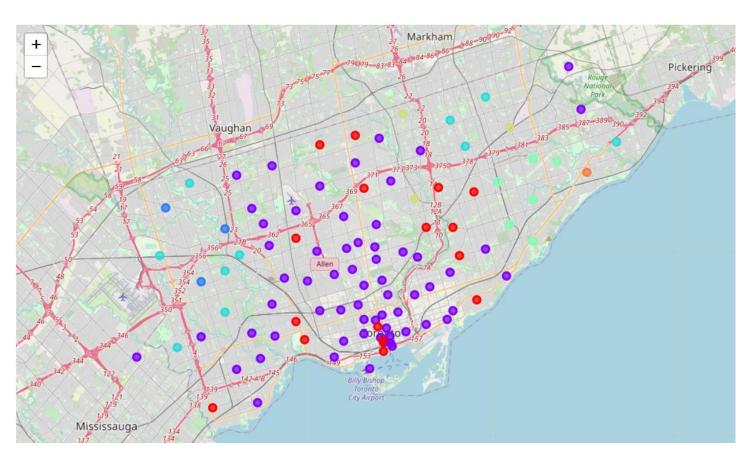

Figure 11 Neighborhood in each cluster

PART 05
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

PART 06
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 map of Toronto.

Figure 12 Recommended neighborhoods

THANKS FOR WATCHING

Wenlong Shen