

The battle of neighborhood: choosing right location for new Chinese restaurant in Toronto

Yafen Huang

Oct 2019

1. Introduction

1.1 Background

Restaurant industry is one of the most competitive businesses in the world. When opening new restaurant in a big city like Toronto, not only planning for the budget, menu and operation are important for the business, but also picking up right location is very crucial to ensure thriving guest visit and bypass competitiveness among restaurants in the future.

The success of a restaurant is usually be influenced by a variety of social factors. Toronto is the capital of Ontario in Canada. Being a Centre of finance, arts and culture, it is recognized as one of the most international, multicultural and dynamic cities in the world. Therefore analyzing the demographic, social and economic characteristics as well as restaurant categories and distribution in each City of Toronto neighborhood will bring insights to understand the patterns and factors that associated with restaurant location, which eventually can be used as guidance for new restaurants to choose right location that fit their business model.

1.2 Interests of this project

The social status of Toronto can be analyzed by looking at the demographics, economic, income characteristics. And then by getting better understanding of the social factors that impact location of Chinese restaurant in Toronto, we could use it as guidance to choose right location when opening a Chinese restaurant in Toronto.

2. Data Preparation

2.1 Data source

Neighborhood profile data such as demographic, income, economic and geo data would be downloaded from Toronto open data website. Here are the links for each datasets:

Neighborhood Profiles dataset:

https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/ef0239b1-832b-4d0b-a1f3-4153e53b189e?format=csv#xd_co_f=NTFiZjBmMTQtOTk0My00ZDdlWE5YjctN2JiNjZiY2Q5Zjgx~

Neighborhood geography coordinates dataset:

https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=csv&projection=4326#xd_co_f=NTFiZjBmMTQtOTk0My00ZDdlWE5YjctN2JiNjZiY2Q5Zjgx~

Wellbeing Toronto – Economics:

< https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/78cf2678-8cd9-442b-a8e7-ed9ae6e82e01#xd_co_f=NTFiZjBmMTQtOTk0My00ZDdlWE5YjctN2JiNjZiY2Q5Zjgx~ >

Restaurant category and location data of each neighborhood are queried from Foursquare API by using the latitude and longitude in the neighborhood geography coordinates dataset.

2.2 Data transformation and feature selection

2.2.1 Neighborhood ethnic origin data

Original dataset contains totally 279 ethnic origins and the value of each category is the population per each neighborhood. A new column of percentage of each ethnic origin is generated to evaluate the contribution of each ethnic origin to the total population in that neighborhood. By looking at rate, a lot of origins contribute little to the overall with less than 5%. None of the neighborhood has more than 5 ethnic origins with percentage all greater than 5%. Therefore only top 5 most popular ethnic origins of each neighborhood are kept to represent the major composition and varieties. Data is then pivoted to get the income range variables as separate columns in aligned with each neighborhood.

2.2.3 Neighborhood income data

Original dataset contains totally 59 income ranges and the value of each category is the population per each neighborhood. A new column of percentage of each income range is generated to evaluate the contribution of each category to the total population in that neighborhood. Only top 5 most popular income ranges of each neighborhood are kept to represent the major composition and varieties. Data is then pivoted to get the income range variable as separate columns in aligned with each neighborhood.

2.2.4 Neighborhood restaurant venue data

Neighborhood altitude and longitude data are used to search the venue data for each neighborhood from Foursquare API. Columns of total venue or total restaurant count per neighborhood are generated by aggregating data. For statistical analysis, only neighborhoods with greater than 3 restaurants are kept. Restaurants categories such as "Sushi Restaurant", "Japanese Restaurant", "Ramen Restaurant" are combined as one category "Japanese Restaurant"; "Chinese Restaurant", "Indian Chinese Restaurant", "Xinjiang Restaurant", "Cantonese Restaurant", "Dim Sum Restaurant", "Dumpling Restaurant", "Taiwanese Restaurant" are combined as one category "Chinese Restaurant".

2.2.5 Neighborhood Chinese restaurant venue subset

Neighborhood Chinese restaurant subset is based on neighborhood restaurant venue data filtered by each neighborhood should at least have one Chinese restaurant.

2.2.6 Final data set for exploratory analysis or clustering modeling

Restaurant datasets are merged with other ethnic, income, population, economic dataset to generate final table for downstream analysis.

3. Results and Discussion

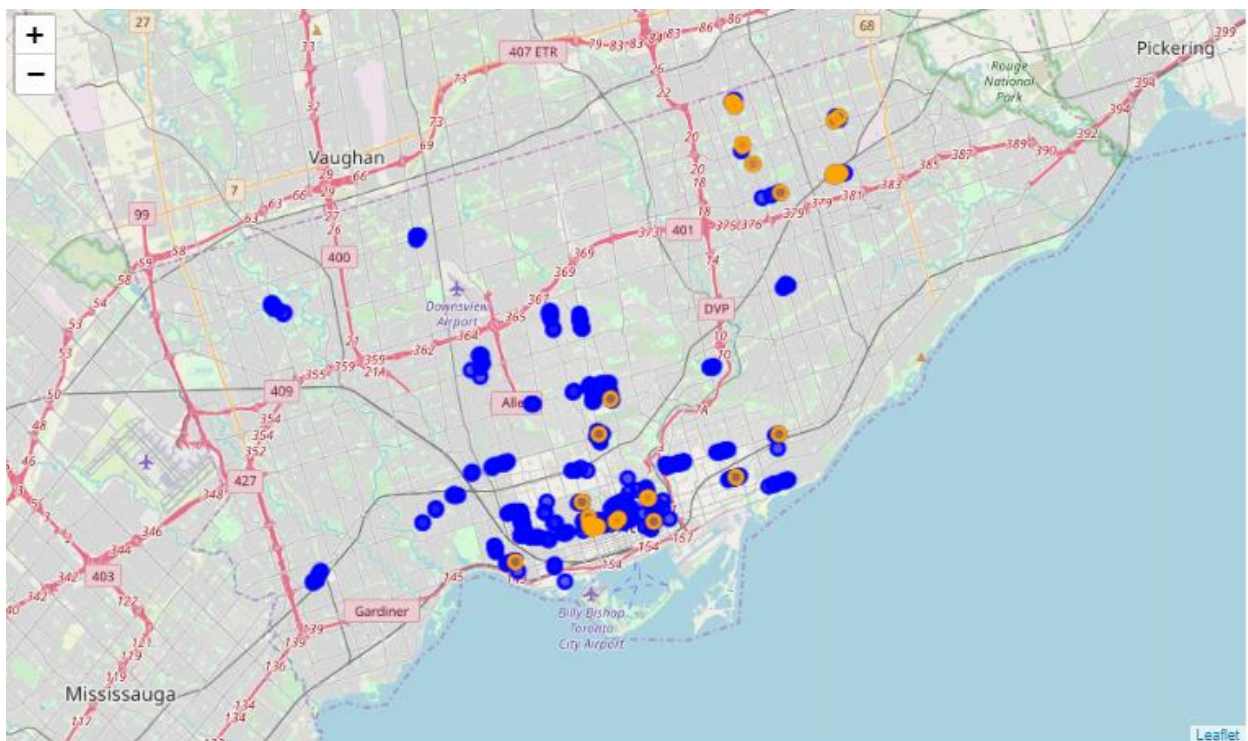
3.1 Data Exploratory Analysis

3.1.1 Analyze the pattern of location of all restaurant and Chinese restaurant in Toronto

The city of Toronto has overall 397 restaurants. The most popular ethnic cuisines are Japanese, Italian, Chinese, fast food and Indian. There are 38 Chinese restaurants in total and distributed in 14 different neighborhoods.



By analyzing the geographic distribution of all the restaurant (in blue) and Chinese restaurant (in yellow) as shown below, we can see that density of restaurant is relatively higher in the neighborhoods near the coast. The distribution of Chinese restaurant follows similar pattern in these neighborhoods near coast. However, there are also neighborhoods that have only higher number of Chinese restaurant, suggesting a correlation of Chinese Restaurant with all restaurants in some neighborhoods but the others.



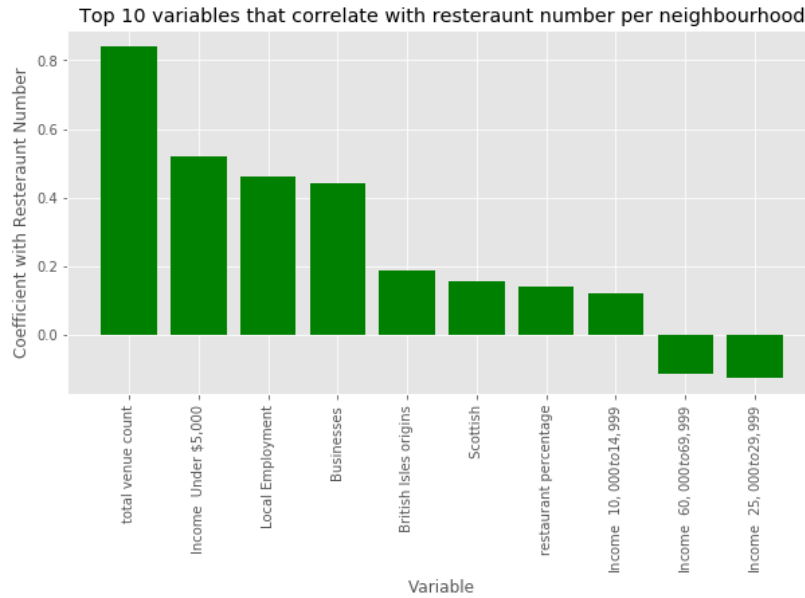
Blue: All restaurants in Toronto

Orange: Chinese restaurants in Toronto

3.1.2 Analyze the correlation of restaurant rate per neighborhoods with social factors.

To check the social factors that may impact total number of restaurant per neighborhood, the correlation matrix is calculated and coefficient greater than 0.1 are listed below. Restaurant rate shows high positive correlation with venue rate, business, local employment, certain income ranges and ethnic origins, suggesting that neighborhoods that have more business, people visits and traffic have more restaurants.

variable	Coefficient
total restaurant count	1
total venue count	0.8289
Businesses	0.501082
Income Under \$5,000	0.48339
Local Employment	0.455862
South Asian origins	-0.22008
Income \$100,000 and over	-0.21525
Income \$80,000 to \$89,999	-0.21133
Irish	-0.18406
Polish	-0.15936
British Isles origins	0.14454
Income \$200,000 and over	-0.12788
Eastern European origins	-0.12458
Southern European origins	-0.12446
Home Prices	-0.12437
Income \$20,000 to \$24,999	-0.1167
Income \$25,000 to \$29,999	-0.10821
East Indian	-0.1001

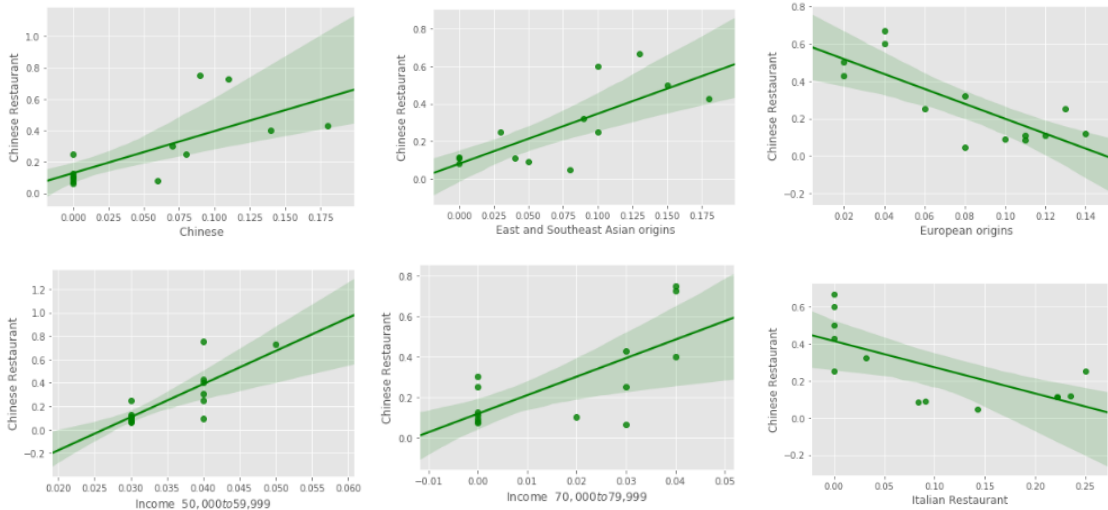
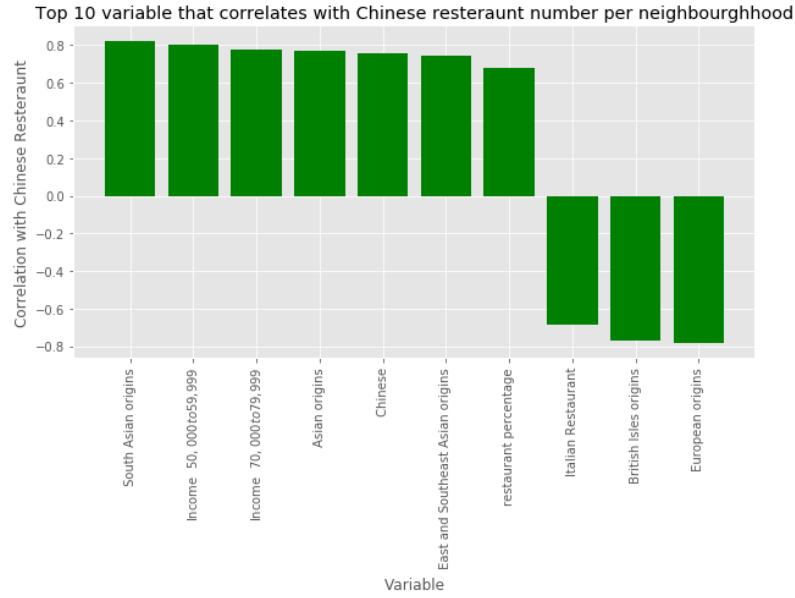


3.1.3 Analyze the correlation of Chinese restaurant rate per neighborhoods with social factors

When looking at social factors that correlate with Chinese restaurant rate, the pattern and dependency are very different from restaurant rate, suggesting that the context of Chinese cuisine popularity is different. Chinese restaurant rate has stronger correlation with ethnic origins and certain income ranges, and other restaurants categories as shown below. The top variables on the list are mainly ethnic origins and income ranges. Chinese restaurant rate is strongly positively associated with eastern ethnic origins including 'South Asian origins', 'Asian origins', 'Chinese' and 'East and Southeast Asian origins' and negatively associated with western origins 'European origins', 'English' and 'British Isles origins', suggesting that Chinese cuisine is much more popular in neighborhood with eastern ethnic origins. Secondly, looking at income ranges that correlate with Chinese restaurant, Income \$50,000 to \$59,999, \$60,000 to \$69,999, \$70,000 to \$79,999, and \$80,000 to \$89,999 show positive correlation, while \$200,000 and over shows negative correlation, implying that Chinese food is more popular in the middle class.

variable	coefficient
South Asian origins	0.824487059
Income \$50,000 to \$59,999	0.800598272
European origins	-0.782042888
Income \$70,000 to \$79,999	0.778692694
British Isles origins	-0.773185252
Asian origins	0.768252098
Chinese	0.755291086

East and Southeast Asian origins	0.743540024
Italian Restaurant	-0.688561277
restaurant percentage	0.677491243
English	-0.643240958
Income \$60,000 to \$69,999	0.575452777
Income \$200,000 and over	-0.536678046
Income \$80,000 to \$89,999	0.506825631
total venue count	-0.490734991
Home Prices	-0.470517623
Debt Risk Score	0.438265667
American Restaurant	0.39817434
Asian Restaurant	-0.393912428
Other North American origins	-0.37320343
Restaurant	-0.335533732
Mediterranean Restaurant	0.33414609
Falafel Restaurant	-0.323289178
Income \$25,000 to \$29,999	0.318915537
Income \$20,000 to \$24,999	-0.261834182
Mexican Restaurant	-0.242922964
Caribbean Restaurant	-0.239507004
Thai Restaurant	-0.234856428
Greek Restaurant	-0.233155183
French Restaurant	-0.233155183
Scottish	-0.223876684
Japanese Restaurant	-0.213051568
Social Assistance Recipients	0.172819176
Income \$100,000 and over	-0.156704335
Eastern European origins	-0.147721029



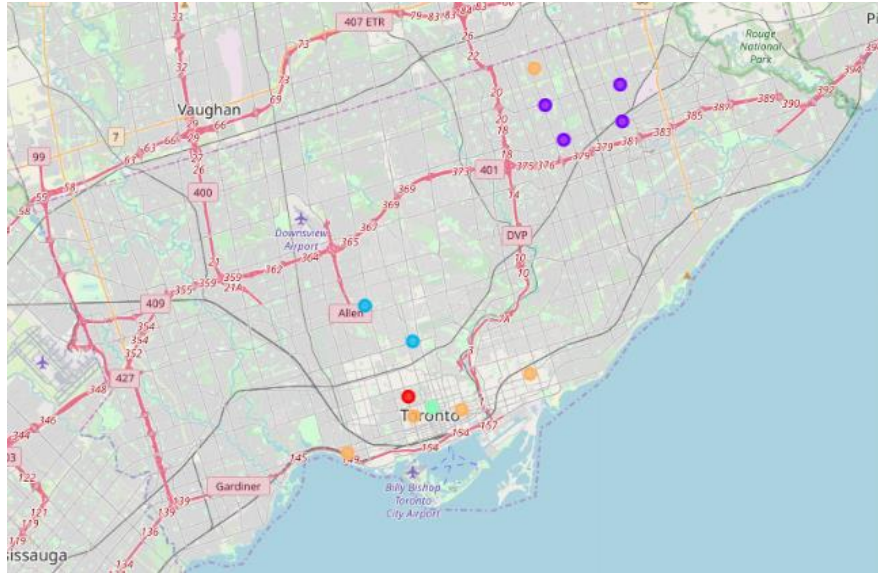
3.2 K mean clustering of neighborhoods with Chinese restaruants

To cluster the neighborhoods with Chinese restaurants, K mean clustering method is used. The neighborhoods are clustered into 5 groups. Cluster 1 and 4 are two biggest clusters that contain 4 and 5 neighborhoods, while clusters 0, 2, 3 only have 1 or 2 neighborhood. Neighborhoods in cluster 4 are geographically close with each other and located near coast area. The characteristics of each cluster could be used as guidance for the business owners to decide best location that suits their business models. For example, Cluster 1 have higher Chinese restaurant rate, people with Asian ethnic origins and home price average around 368k. Neighborhoods such as 'Tam O'Shanter-Sullivan', and 'Forest Hill North' with relatively

low rate of Chinese restaurant and more Asian people may be less competitive and thus considered as good options for new Chinese restaurants. Neighborhoods in the Cluster 4 have more venues and relatively lower Chinese restaurants. They are located in the center of city where daily people visits and traffic is heavy. They could be good choice if the theme of the new Chinese restaurant is tailored to people with quick pace or tourists seeking for something special. Restaurants with South East recipes can specifically choose neighborhoods with higher population in East and Southern Asian origins.

	cluster	Neighbourhood	Chinese Restaurant	South Asian origins	Income \$50,000 to \$59,999	European origins	Income \$70,000 to \$79,999	British Isles origins	Asian origins	Chinese	East and Southeast Asian origins	Italian Restaurant	restaurant percentage	English	Income \$60,000 to \$69,999	Income \$200,000 and over	Income \$80,000 to \$89,999	total venue count	Home Prices
34	0	University	0.111111111	0	0.03	0.12	0	0.07	0.06	0	0.04	0.222222	0.22	0.03	0.03	0.05	0	40	773920
1	1	Agincourt South-Malvern V	0.666666667	0.04	0.05	0.04	0.04	0	0.18	0.11	0.13	0	0.78	0	0.04	0	0.03	18	332710
15	1	L'Amoreaux	0.6	0.04	0.04	0.04	0.04	0	0.16	0.09	0.1	0	0.36	0	0.04	0	0.03	14	355438
0	1	Agincourt North	0.5	0.04	0.04	0.02	0.04	0	0.2	0.14	0.15	0	0.33	0	0.04	0	0	18	375307
30	1	Tam O'Shanter-Sullivan	0.25	0.04	0.04	0.06	0.03	0	0.15	0.08	0.1	0.25	0.27	0	0.04	0	0.03	15	411143
10	2	Forest Hill North	0.25	0	0.03	0.13	0.03	0	0.05	0	0.03	0	0.22	0	0.03	0.09	0	18	957688
37	2	Yonge-St.Clair	0.117647059	0	0.03	0.14	0.03	0.08	0	0	0	0.235294	0.33	0.05	0.03	0.09	0	54	995616
3	3	Bay Street Corridor	0.047619048	0	0.03	0.08	0	0.04	0.12	0.06	0.08	0.142857	0.24	0	0.03	0.04	0	92	457787
29	4	Steeles	0.428571429	0.02	0.04	0.02	0.03	0	0.21	0.18	0.18	0	0.58	0	0.04	0	0.03	12	440688
14	4	Kensington-Chinatown	0.322580645	0	0.04	0.08	0	0.05	0.11	0.07	0.09	0.032258	0.34	0	0.03	0	0	93	477989
19	4	Moss Park	0.111111111	0	0.03	0.11	0	0.07	0.05	0	0	0.222222	0.28	0.03	0	0.03	0	36	457524
28	4	South Parkdale	0.090909091	0	0.04	0.1	0	0.06	0.07	0	0.05	0.090909	0.34	0.03	0.03	0	0	35	500054
11	4	Greenwood-Coxwell	0.083333333	0	0.03	0.11	0	0.08	0.06	0	0	0.083333	0.35	0.04	0.03	0.04	0.03	37	490920

cluster	Chinese Restaurant	South Asian origins	Asian origins	Chinese	East and Southeast Asian origins	Italian Restaurant	restaurant percentage	Income \$60,000 to \$69,999	Income \$70,000 to \$79,999	Income \$80,000 to \$89,999	Income \$200,000 and over	total venue count	Home Prices
1	0.50416667	0.04	0.1725	0.105	0.12	0.0625	0.435	0.04	0.0375	0.0225	0	16.25	368649.5
4	0.20730112	0.004	0.1	0.05	0.064	0.085744542	0.378	0.026	0.006	0.012	0.014	42.6	473435
2	0.18382353	0	0.025	0	0.015	0.117647059	0.275	0.03	0.03	0	0.09	36	976652
0	0.11111111	0	0.06	0	0.04	0.222222222	0.22	0.03	0	0	0.05	40	773920
3	0.04761905	0	0.12	0.06	0.08	0.142857143	0.24	0.03	0	0	0.04	92	457787



- Cluster 0: red;
- Cluster 1: purple
- Cluster 2: blue
- Cluster 3: light green
- Cluster 4: orange

4 Conclusion

In this project, the distribution of Chinese restaurants and correlation with social factors in the city of Toronto was studied. Chinese restaurant is shown to be top 3 popular ethnic cuisines in the city. Correlation analysis suggests that Chinese restaurant is strongly associated with Asian ethnic origins, people with income range representative of the middle class. Finally K mean clustering method was used to take all the features such as ethnicity, purchasing ability, housing price, competitiveness within or between each restaurant categories etc. into consideration and cluster the neighborhoods with Chinese restaurant. Therefore it can be used as guidance for choosing right location for new Chinese restaurant and fit specific business needs.