

Open a Chinese restaurant in Toronto

IBM Data Science Capstone Project

Zhiao He

Introduction

As more and more different cultural people immigrant from all over the world to Canada, people need more and more different cultural things in their life. Restaurant could represent for one aspect in a culture. Toronto, the capital of the province of Ontario, is the most populous Canadian city. In Toronto, there are more and more Chinese immigrants especially young people. They really want to have a taste of Chinese original food instead of revised Chinese food. Therefore there is a big opportunity for a business of opening a Chinese restaurant. While opening a restaurant can be a very lucrative business, a lack of demand caused many restaurants to close within the first year of opening. There are many different factors that can account for a restaurant's success such as price, location, quality of the food.

The goal of this project is to use the Foursquare API to determine a good location in Toronto to open a Chinese restaurant. For Chinese Restaurant specifically, location and competition are both determined by where the restaurant will be opened.

Data

To answer the business problem, the following factors to be extracted from various data sources:

- Population & Ethnic Distribution of Each Neighborhood (Toronto Census)
- Income Distribution of Each Neighborhood (Toronto Census)
- Number of Restaurants in Each Neighborhood (Foursquare API)
- Number of Indian Restaurants in Each Neighborhood (Foursquare API)

The Toronto Census data was extracted from <https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a>.

Methodology

The first step of the project was to combine the Toronto dataset, containing the postal code, borough, neighborhood name, latitude and longitude for each postal code in Toronto, and the census dataset.

Using the income distribution for each neighborhood, the spending power of each area was calculated using the median of each category weighted by the number of people in that income category. Thus, the spending power represents the overall capital of each area (i.e. total income of the inhabitants). Since the spending power for each area is considerably large and the relative strength is difficult to visualize, the spending power for each area was standardized.

The next step was to visualize the location of the various postal codes within Toronto to obtain a general understanding the location (Figure 1). As seen from the map, the postal codes are densely clustered near downtown Toronto and spread out as the distance from downtown increases. This is important because while some postal codes might not have many restaurants, if the area is located near downtown, adjacent regions can heavily impact the profitability of the restaurant.

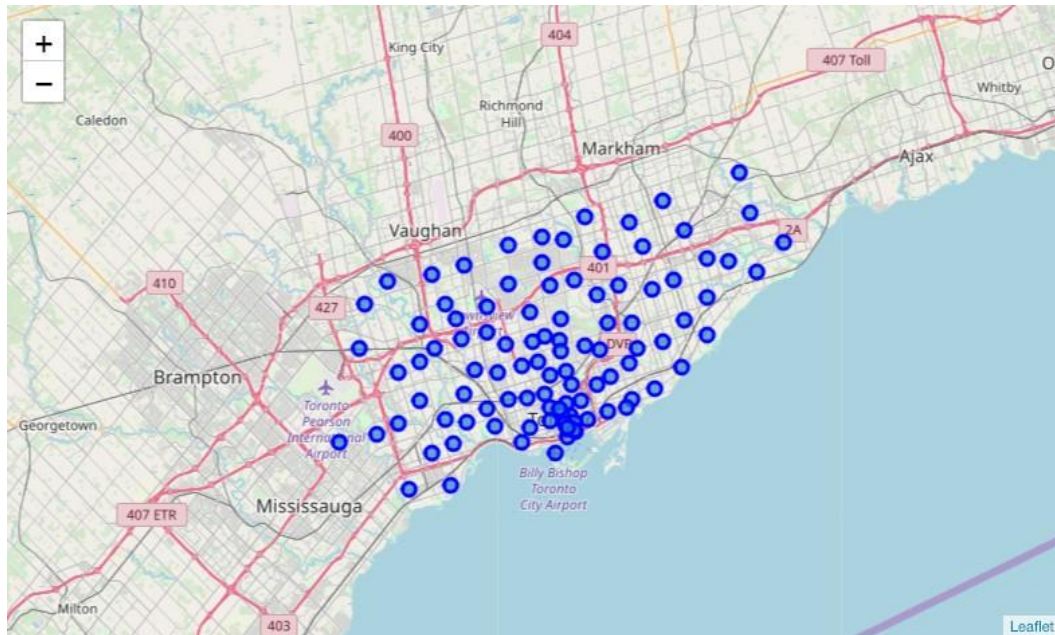


Figure 1: Location of each postal code within Toronto, Canada.

Now that the region has been clearly visualized, the Foursquare API was used to explore each neighborhood and return the top 200 venues within 2,000 meters (1.2 miles) of the longitude and latitude for each postal code. The extracted venue categories were encoded using one-hot encoding and the total restaurants and Chinese restaurants in each region were calculated (Figure 2).

	Neighborhood	Total Restaurants	Chinese Restaurants
0	Agincourt	51	12
1	Alderwood, Long Branch	40	0
2	Bathurst Manor, Wilson Heights, Downsview North	24	0
3	Bayview Village	16	5
4	Bedford Park, Lawrence Manor East	48	1

Figure 2: Result of calculating the number of restaurants in every region(first 5 rows).

With the resulting data, the Postal Code, Borough name, Latitude, Longitude and Density columns of each region were dropped from the DataFrame. Then, the population, area, spending

power, total number of restaurants and the number of Indian restaurants were used to train a k-Means clustering algorithm with 5 clusters (Figure 3). The characteristics of the resulting clusters can be found in Table 1.

Results

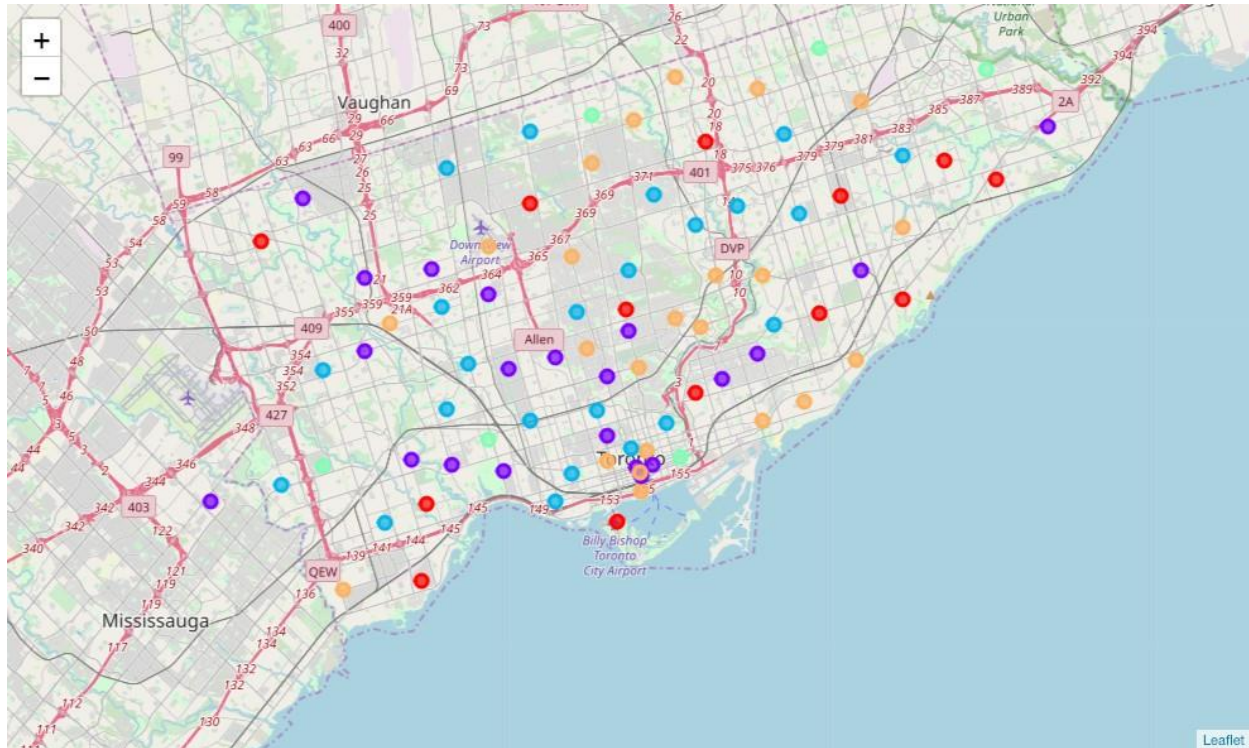


Figure 3: Result of the clustering algorithm. Cluster 0 = Red Cluster 1 = Purple Cluster 2 = Blue Cluster 3 = Turquoise Cluster 4 = Orange

Cluster	Characteristics
Cluster 0	Positive Spending Power (0.3 – 1.8)
Cluster 1	Negative Spending Power (-1.2 -- -0.8)
Cluster 2	Near Zero Spending Power (-0.5 – 0.5)
Cluster 3	High Positive Spending Power (1.7+)
Cluster 4	Negative Spending Power (-0.8 – 0) With Large Number of Restaurants

Table 1: Characteristics of the clusters resulting from k-Means clustering algorithm

	Neighborhood	Population, 2016	Land area in square kilometres	Spend Power	Total Restaurants	Chinese Restaurants
0	Parkwoods	34805.0	7.42	0.039008	41	3
1	Victoria Village	17510.0	4.72	-0.724715	30	1
2	Regent Park, Harbourfront	76716.0	8.01	3.841548	27	0
3	Lawrence Manor, Lawrence Heights	6577.0	1.82	-1.117428	33	1
4	Malvern, Rouge	90290.0	45.74	1.707959	12	0

Figure 4: Data from neighborhoods belonging to Cluster 3(first 5 rows)

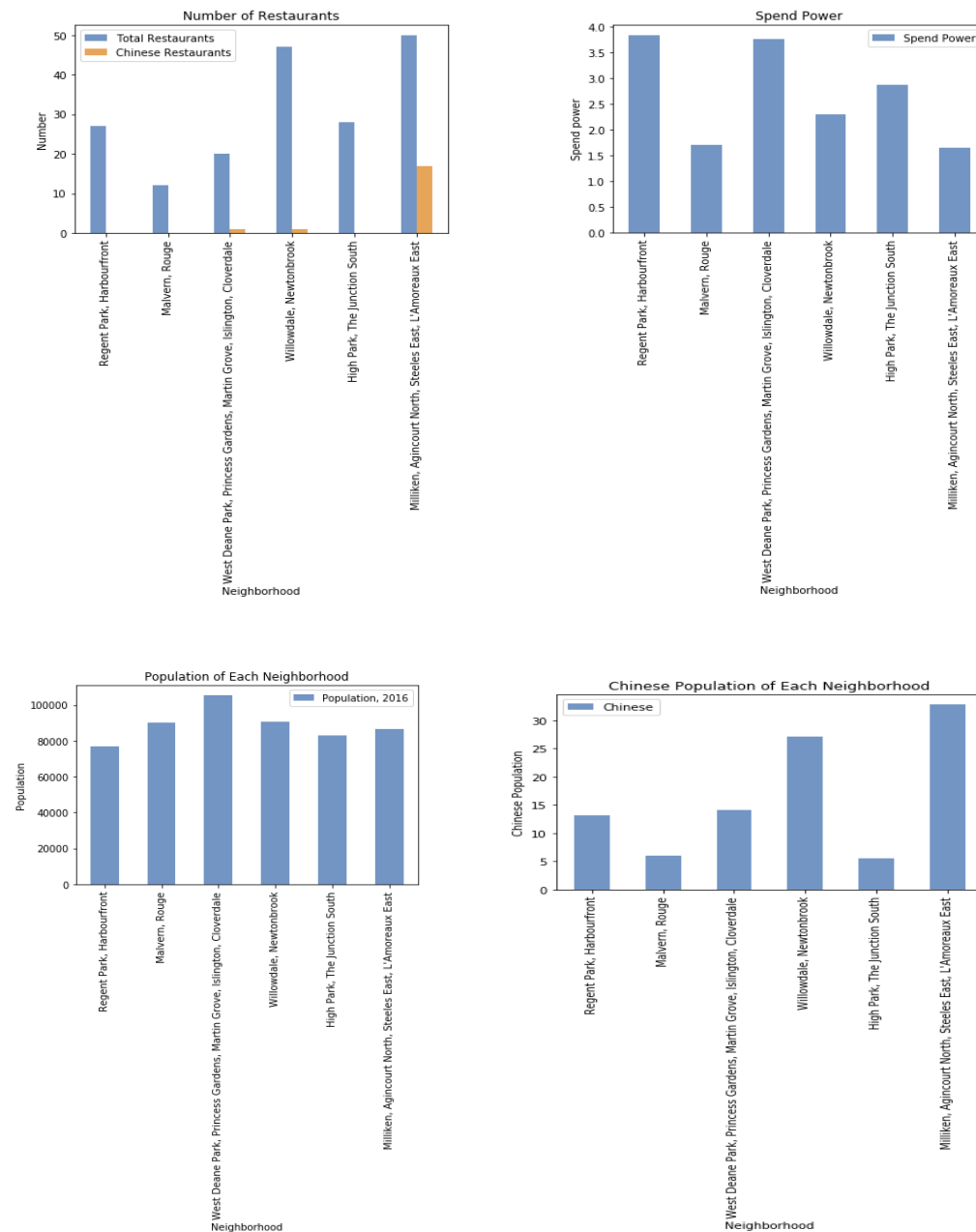
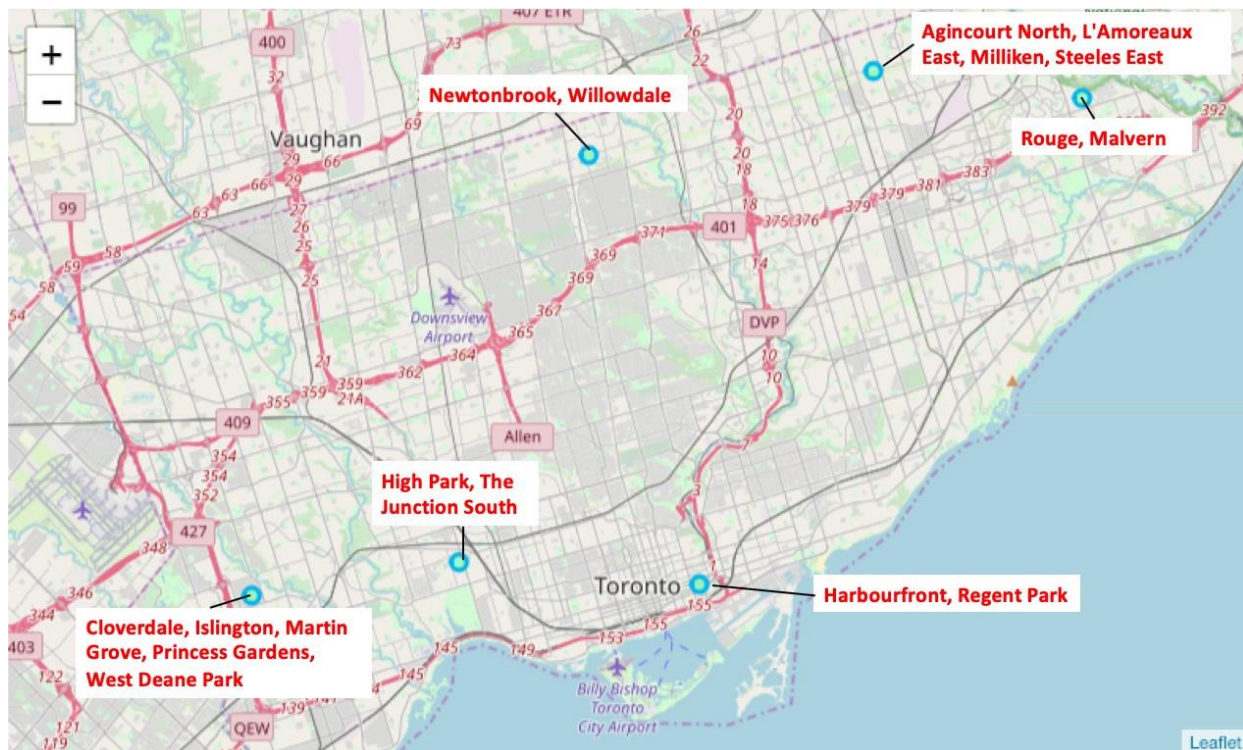


Figure 5: These plots shows the characteristics of neighborhoods belonging to cluster 2

Discussion

From the results of the clustering algorithm, it was determined that neighborhoods corresponding to cluster 2 were the best choice for opening a Chinese restaurant based on the normalized spending power and population. This narrowed down possible locations to 3 different areas. Using the results in Figure 5, L'Amoreaux East, Milliken, Steeles East region the Willowdale region and the Milliken region were eliminated due to the large number of restaurants in the area.

From the three remaining regions, I would recommend that the client open his/her restaurant in either the Rouge, Malvern region or the Cloverdale, Islington, Martin Grove, Princess Gardens, Willowdale, Nettonbrook. Both regions have very few restaurants and are farther away from the downtown area. Compared with Milliken area, Willowdale Nettonbrook area has almost the same spend power and Chinese population. But the total number of Chinese restaurants have a huge difference. Therefore there might be a big opportunity in this area.



Conclusion

Opening a restaurant is a complex task that can lead to a large monetary loss if not done properly. Thus, extensive research about the area would greatly increase the likelihood of the restaurant succeeding. From the project above, I demonstrated the workflow necessary for a client to determine what area the restaurant should open. For specifically, I determined that the optimal location to open a Chinese restaurant in Toronto should be in the Willowdale, Nettonbrook area.

