Capstone Project:
Opening a Chinese Restaurant in Toronto

Yimin Gan

July 28, 2019

Problem Description

- Usually, we need to consider a lot of factors when opening a restaurant:
  - Location: where is a good place to open a restaurant in a city
  - Style of restaurant: it depends on what kind of restaurant we are good at
  - Return of investment: we need to know the investment of renting a house, hiring people, etc. as well as a prediction of revenue/profit.
- In this project, we will focus on addressing the problem of where to open a Chinese restaurant in Toronto.
- Suppose the good place to open a Chinese restaurant is the neighborhoods with a large number of restaurants but very few Chinese restaurant.
- People who want to start a restaurant business in Toronto will have interest to this project

Data Source and Description

- a Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M that has all the information we need to explore and cluster the postal codes, boroughs and neighborhoods in Toronto.
- a csv file that has the geographical coordinates of each postal code of Toronto city: http://cocl.us/Geospatial_data
- Foursquare API to get the most common venues of given neighborhoods of Toronto city.
- Based on the venue data, we can pull out the restaurant information in each neighborhood of Toronto, then make statistics how many restaurants in each neighborhood and the style of each restaurant.

Data Acquisition and Cleaning

- Get the neighborhood info of Toronto city from the Wikipedia page, clean the data, and combine it with the geographical coordinates into a *pandas* dataframe.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	East Toronto,Riverdale	43.679557	-79.352188
2	M4L	East Toronto	East Toronto,India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	East Toronto	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
5	M4P	Central Toronto	Central Toronto	43.712751	-79.390197
6	M4R	Central Toronto	Central Toronto	43.715383	-79.405678
7	M4S	Central Toronto	Central Toronto	43.704324	-79.388790
8	M4T	Central Toronto	Moore Park,Central Toronto	43.689574	-79.383160
9	M4V	Central Toronto	Deer Park,Central Toronto,Rathnelly,South Hill...	43.686412	-79.400049

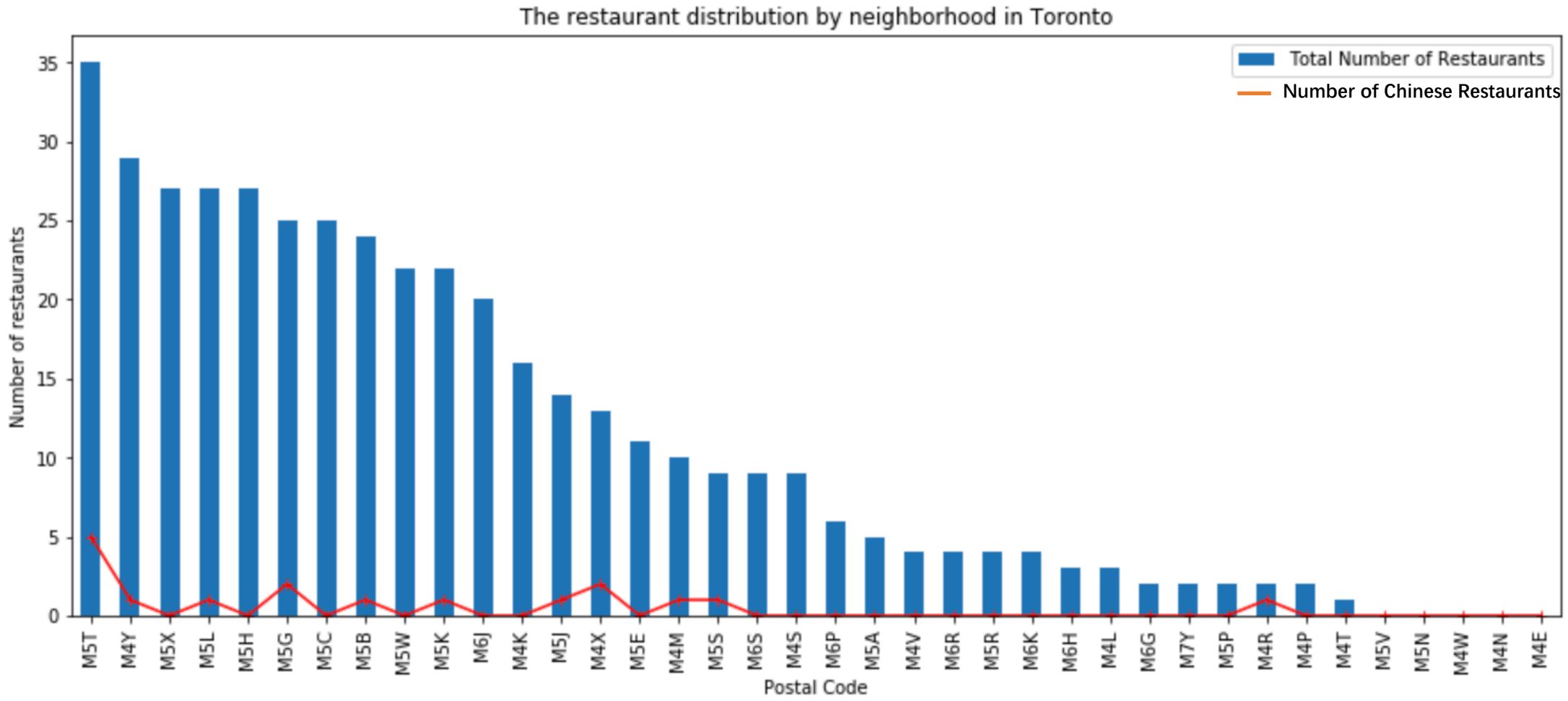
Restaurant Data Analysis

- Firstly, utilize Foursquare API to get venue info of each neighborhood, keep only the restaurant data (number and style of restaurant), then list the number of Chinese restaurants, and total number of restaurants in each neighborhood by descending.

	PostalCode	Neighborhood	Chinese Restaurant	Total Number of Restaurants
0	M5T	Chinatown,Grange Park,Kensington Market	5	35
1	M4Y	Church and Wellesley	1	29
2	M5X	First Canadian Place,Underground city	0	27
3	M5L	Commerce Court,Downtown Toronto	1	27
4	M5H	Downtown Toronto,Downtown Toronto,Downtown Tor...	0	27
5	M5G	Downtown Toronto	2	25
6	M5C	St. James Town	0	25
7	M5B	Downtown Toronto,Downtown Toronto	1	24
8	M5W	Downtown Toronto	0	22
9	M5K	Design Exchange,Toronto Dominion Centre	1	22

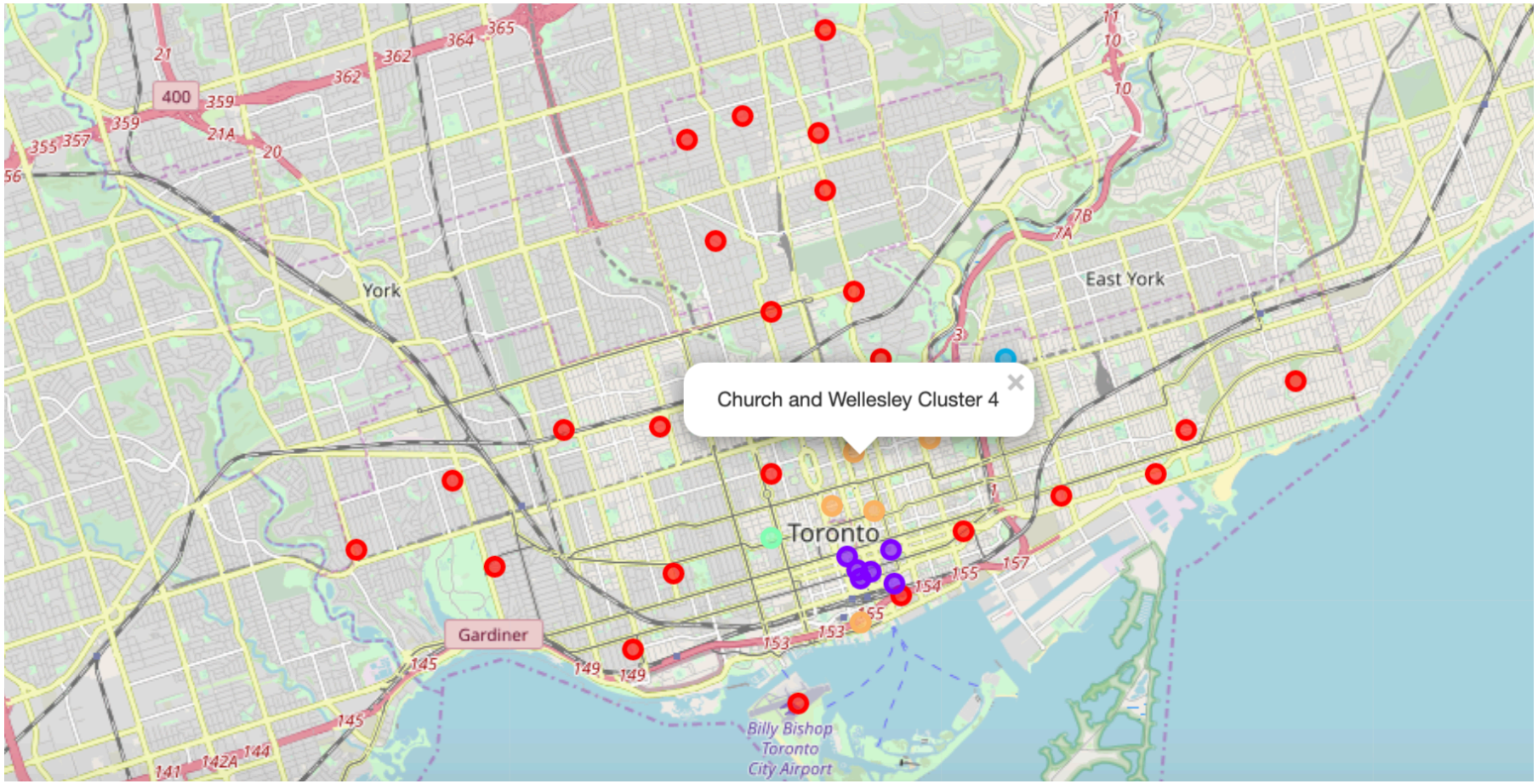
The restaurant distribution by neighborhood

- Use a bar chart to show the total number of restaurants, and a line to show the number of Chinese restaurants in each neighborhood.
- As examples, M5X and M5H are good places for opening a Chinese restaurant according to assumption.



Clustering the neighborhoods by restaurant data

- Use K-Means algorithm to cluster the neighborhoods into 5 clusters based on the restaurant data
- Both M5X and M5H are in cluster 1 (purple mark), so we can think all the neighborhoods in cluster 1 are good places for opening a Chinese restaurant.



Recommendation (1)

- According to our analysis, the first recommendation for opening a Chinese restaurant is the neighborhoods in cluster 1, it includes 6 neighborhoods listed below:

	PostalCode	Borough	Neighborhood	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
15	M5C	Downtown Toronto	St. James Town	1	Restaurant	Italian Restaurant	American Restaurant	Japanese Restaurant	Seafood Restaurant
18	M5H	Downtown Toronto	Downtown Toronto,Downtown Toronto,Downtown Tor...	1	Thai Restaurant	American Restaurant	Asian Restaurant	Restaurant	Sushi Restaurant
20	M5K	Downtown Toronto	Design Exchange,Toronto Dominion Centre	1	Italian Restaurant	Restaurant	American Restaurant	Japanese Restaurant	Seafood Restaurant
21	M5L	Downtown Toronto	Commerce Court,Downtown Toronto	1	Restaurant	American Restaurant	Seafood Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant
28	M5W	Downtown Toronto	Downtown Toronto	1	Restaurant	Fast Food Restaurant	Seafood Restaurant	Italian Restaurant	Japanese Restaurant
29	M5X	Downtown Toronto	First Canadian Place,Underground city	1	Restaurant	American Restaurant	Asian Restaurant	Seafood Restaurant	Thai Restaurant

Recommendation (2)

- When checking the restaurant data and map, we think the neighborhoods in cluster 4 will be our second recommendation for opening a Chinese restaurant, it includes 5 neighborhoods listed below:

	PostalCode	Borough	Neighborhood	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
11	M4X	Downtown Toronto	Cabbagetown,St. James Town	4	Restaurant	Italian Restaurant	Chinese Restaurant	Indian Restaurant	American Restaurant
12	M4Y	Downtown Toronto	Church and Wellesley	4	Japanese Restaurant	Sushi Restaurant	Restaurant	Fast Food Restaurant	Mediterranean Restaurant
14	M5B	Downtown Toronto	Downtown Toronto,Downtown Toronto	4	Fast Food Restaurant	Middle Eastern Restaurant	Ramen Restaurant	Italian Restaurant	Japanese Restaurant
17	M5G	Downtown Toronto	Downtown Toronto	4	Italian Restaurant	Indian Restaurant	Sushi Restaurant	Japanese Restaurant	Chinese Restaurant
19	M5J	Downtown Toronto	Downtown Toronto,Toronto Islands,Union Station	4	Italian Restaurant	Restaurant	Indian Restaurant	Japanese Restaurant	New American Restaurant

Conclusion and Future Directions

- Built a simple model to cluster the neighborhoods based on restaurant data, and provided recommendation for the business problem.
- Accuracy of the model has room for improvement, for example, giving more weight to the number of Chinese restaurants in K-Means clustering.
- Future Directions
  - Collect more restaurant data for the analysis, such as the visitors flow rate, operation cost, income, etc.
  - Develop machine learning models to predict the cost and revenue/profit for opening a restaurant.