Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Chelsea Huang

Introduction

Canada has an open immigration policy and a diverse and inclusive culture. Its diverse dining culture makes many restaurants with different styles or exotic flavors emerge in Canada. Toronto, the capital of Ontario, is the largest city in Canada and also one of the most diverse. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada. But it is not easy for a restaurant to survive for a long time in a dynamic metropolis like Toronto. I am a Chinese who has lived in Toronto for four years and I found that many of my favorite restaurants have closed down after only a few years of opening. The location of the restaurant not only affects the market development capacity of the catering company, but also the size of its attractiveness to consumers, but more importantly, it has a strategic impact on the long-term benefits.

Here we are mainly trying to investigate where would be a good choice for opening a new Chinese restaurant in Toronto from the perspective of location selection.

Target audience:

Who would be interested in this project?

1. business person that are considering opening a restaurant and they wonder how to choose the best location for their business;
2. anyone looking for a nice place to get a meal in Toronto;
3. Data scientists that are interested in the same topic of using location data to explore a geographical location.

Data

Data that will be used to solve the problem:

1. Toronto neighborhood data scraped from a Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. a csv file that has the geographical coordinates of each postal code: https://cocl.us/Geospatial_data
3. Foursquare location data, with latitude and longitude coordinates of each neighborhood.
4. Foursquare API to explore neighborhoods in Toronto.

For the Toronto neighborhood data, the data that we use is collected from a Wikipedia page, which provides all the information we need to explore and cluster the neighborhoods in Toronto.

data wrangling

- The dataframe will consist of three columns: Postal Code, Borough, and Neighborhood.

- Only process the cells that have an assigned borough. Ignore cells with a borough that is **Not assigned**.
- More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that **M5A** is listed twice and has two neighborhoods: **Harbourfront** and **Regent Park**. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in row 11 in the above table.
- If a cell has a borough but a **Not assigned** neighborhood, then the neighborhood will be the same as the borough.

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...
11	M1R	Scarborough	Wexford, Maryvale

Get the geographical coordinates of the neighborhoods

First, we built a dataframe of the postal code of each neighborhood along with the borough name and neighborhood name. Then, we use a csv file that has the geographical coordinates of each postal code: https://cocl.us/Geospatial_data Finally, we merge these two dataframes together and get a full dataframe of Toronto with 103 rows and 5 columns. Below shows top 12 rows:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304
11	M1R	Scarborough	Wexford, Maryvale	43.750072	-79.295849

Also, we use the Foursquare API to explore neighborhoods in Toronto. And here we get all the Chinese restaurant in Toronto.

	Borough	Neighborhood	ID	Name
0	Scarborough	Malvern, Rouge	4c706524df6b8cfab244b84d	Charley's Exotic Cuisine
1	Scarborough	Guildwood, Morningside, West Hill	4c85aa5bee6fef3b1d1d3e5c	Peking Garden Restaurant
2	Scarborough	Woburn	4b64765ff964a52028b52ae3	Lucky Hakka
3	Scarborough	Cedarbrae	54a6ea76498ebc906c8c3652	Hakka Legend
4	Scarborough	Kennedy Park, Ionview, East Birchmount Park	4b6475def964a520f9b42ae3	Chung Moi

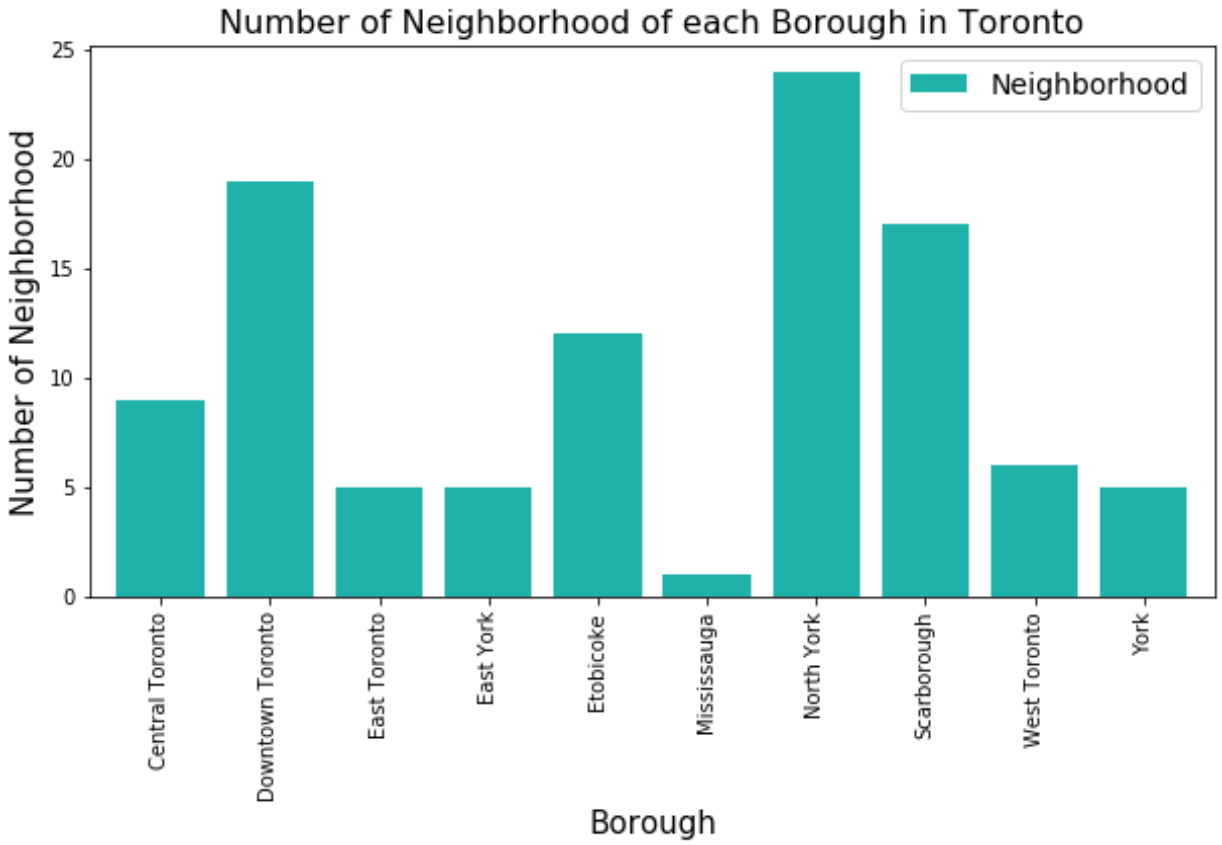
Methodology

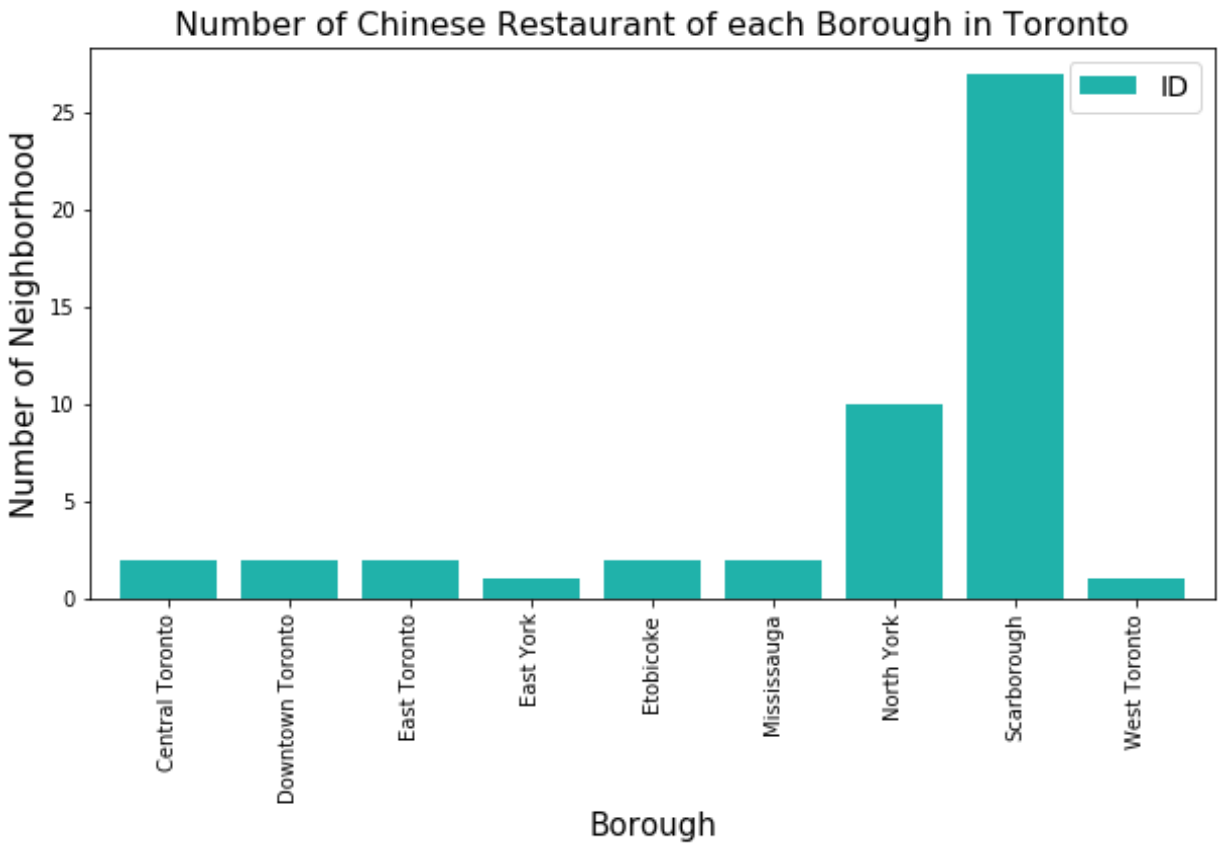
In this project we will direct our efforts on detecting areas of the Great Toronto Area (GTA) that have a high number of Chinese restaurants.

First, we have collected the required data: location and type (category) of every restaurant. We have also identified Chinese restaurants (according to Foursquare categorization).

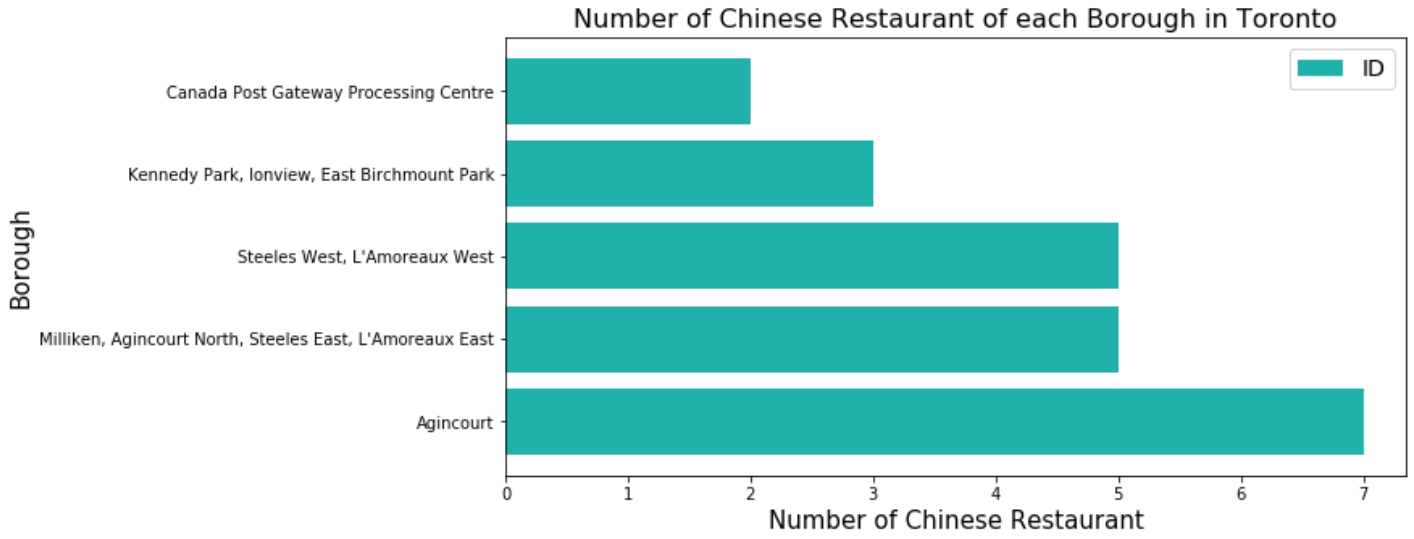
We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Exploratory Data Analysis





From the above barplot, we can see that Scarborough has the largest number of Chinese restaurants. And North York ranks the second place.



From above, we can see that Agincourt has the largest number of Chinese restaurants, which is a neighborhood in Scarborough. Next, we created a dataframe including top 100 nearby venues of each

neighborhood with a radius 500 meters.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant

Then, we create a dataframe using pandas one hot encoding for the venue categories and group rows by neighborhood and by taking the mean of the frequency of occurrence of each category.

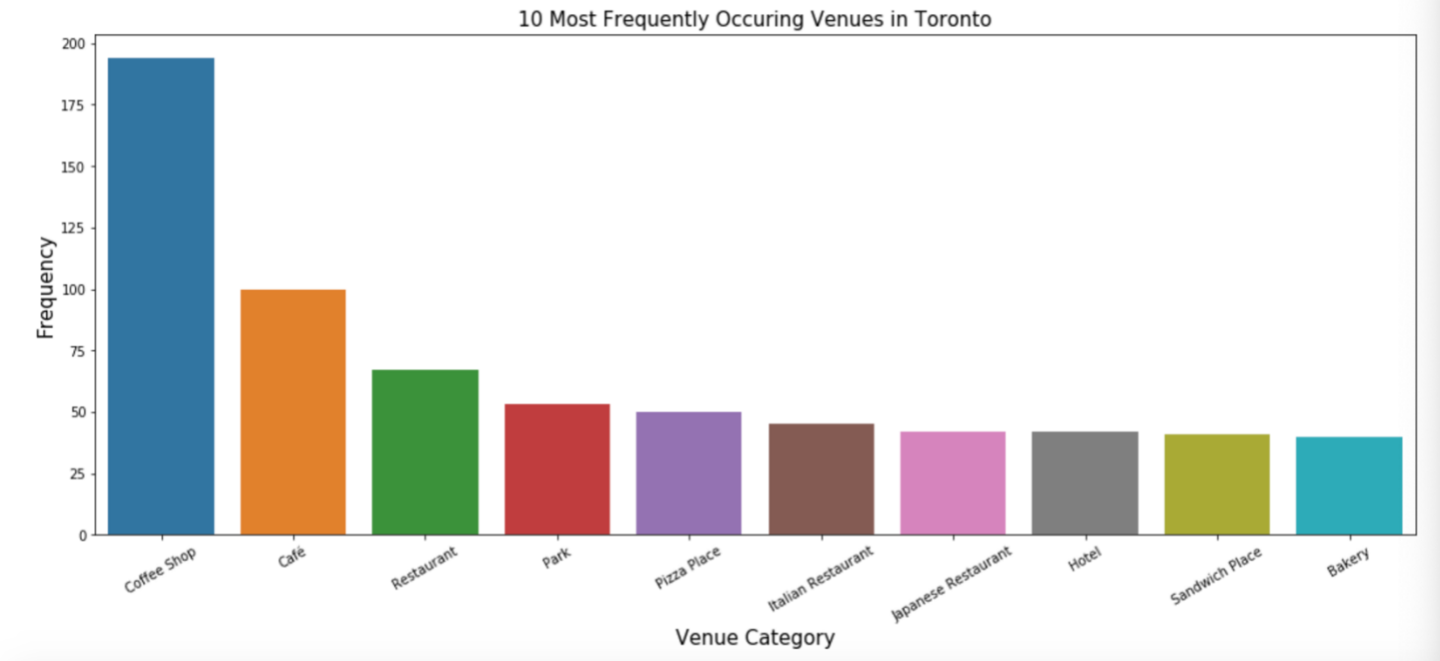
	Neighborhood	Yoga Studio	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Workshop	BBQ Joint	Baby Store	B
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.04	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Next, we create a dataframe with top 10 most common venues for each neighborhood to get an overview.

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berczy Park	Coffee Shop	Cocktail Bar	Farmers Market	Seafood Restaurant	Restaurant	Café	Cheese Shop	Bakery	Beer Bar	Hotel
1	Brockton, Parkdale Village, Exhibition Place	Café	Coffee Shop	Breakfast Spot	Nightclub	Bakery	Gym	Stadium	Burrito Place	Restaurant	Climbing Gym
2	Business reply mail Processing Centre, South C...	Light Rail Station	Yoga Studio	Burrito Place	Restaurant	Fast Food Restaurant	Brewery	Farmers Market	Auto Workshop	Spa	Butcher
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Sculpture Garden	Airport	Airport Food Court	Harbor / Marina	Bar	Boat or Ferry	Rental Car Location
4	Central Bay Street	Coffee Shop	Sandwich Place	Café	Italian Restaurant	Japanese Restaurant	Salad Place	Bubble Tea Shop	Department Store	Burger Joint	Korean Restaurant

With 267 unique venue categories, below it shows top 10 venue categories:

	Venue_Category	Frequency
0	Coffee Shop	194
1	Café	100
2	Restaurant	67
3	Park	53
4	Pizza Place	50
5	Italian Restaurant	45
6	Japanese Restaurant	42
7	Hotel	42
8	Sandwich Place	41
9	Bakery	40



It seems like coffee shops is the most frequent venue among all categories in Toronto according to Foursquare API. This makes sense since the coffee shop market is already booming in North America. For Chinese restaurants specifically, the frequency is quite low, maybe some of them have not been

verified and added to Foursquare API database yet. But it can be said that there is still a lot of room for expansion in the market for Chinese restaurant.

Results & Discussion

- Coffee shops, café, Restaurants, park, pizza place are the most common venues in Toronto.
- Coffee shops is the most frequent venue among all categories in Toronto according to Foursquare API. For Chinese restaurants specifically, the frequency is quite low, maybe some of them have not been verified and added to Foursquare API database yet. But it can be said that there is still a lot of room for expansion in the market for Chinese restaurant.
- Clustering neighborhoods based on their most popular venues grouped.
- Scarborough has the largest number of Chinese restaurants. And North York ranks the second place. And Agincourt has the largest number of Chinese restaurants, which is a neighborhood located in Scarborough.
- East York and West Toronto only have one Chinese restaurant in the database. Almost all Chinese restaurants are located in Scarborough. North York, East Toronto, Central Toronto, Downtown Toronto, Mississauga, Etobicoke have about the same number.

As outlined previously, we used Foursquare data so that we have first identified general boroughs that justify further analysis. We used K-means clustering algorithm to cluster those locations and then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

There are several limitations of our study. First, the findings of the present study were limited by the Foursquare API database. If we really want to open a Chinese restaurant, we haven't considered many other factors, such as rent, passenger flow, traffic and so on. And unfortunately, there is no information about rating, likes and tips for Chinese restaurants in Toronto through Foursquare API.

Conclusion

Depend on the results of our analysis, we conclude that final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like rent, passenger flow, traffic, etc.