Photo: Google

Introduction/Business Problem

A Chinese multinational company wants to open a branch in Toronto, ON. They are looking for a good neighborhood in Toronto. Even though this is a Chinese company people from different background will work here. Majority might be Canadian and then Chinese. Top executive from China will visit time to time.

Our goal is to find a suitable location/borough to open the office. Preferably around downtown Toronto. The company will prefer a location with lots of Chinese and local restaurant. Also, with some good touristy place nearby. Good access to gym and others.

Data

The data we will use in this project will be collected from Wikipedia. This data will be consisting of postal code, brough and neighborhood. Then another set of data will be collected from the website 'cocl.us'. This data will consist of longitude and latitude along with postal code. Then we will marge the two data set according to their postal code. A foursquare API will be used to find the certain venues in the neighborhood. And then by using python package folium will be used to find the map location.

The website link for the data:
Toronto Neighborhoods - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Toronto Latitude and Longitude - http://cocl.us/Geospatial_data

Data Processing

As expected, the data was not properly processed. After downloading the data from Wikipedia, column name of the Data Frame (DF) was changed to match the given condition by coursera. Some of the values in boroughs and neighborhood was given as ‘not assigned’, so the row that has this ‘not assigned’ value was dropped.

	PostalCode	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn

As we can see in this DF longitude and latitude was not given. So, it’s was downloaded from a different website as discussed in the data section. This longitude and latitude data were then combined with the ‘PostalCode’ column. Then we visualize the data entire neighborhood location on the map.

	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

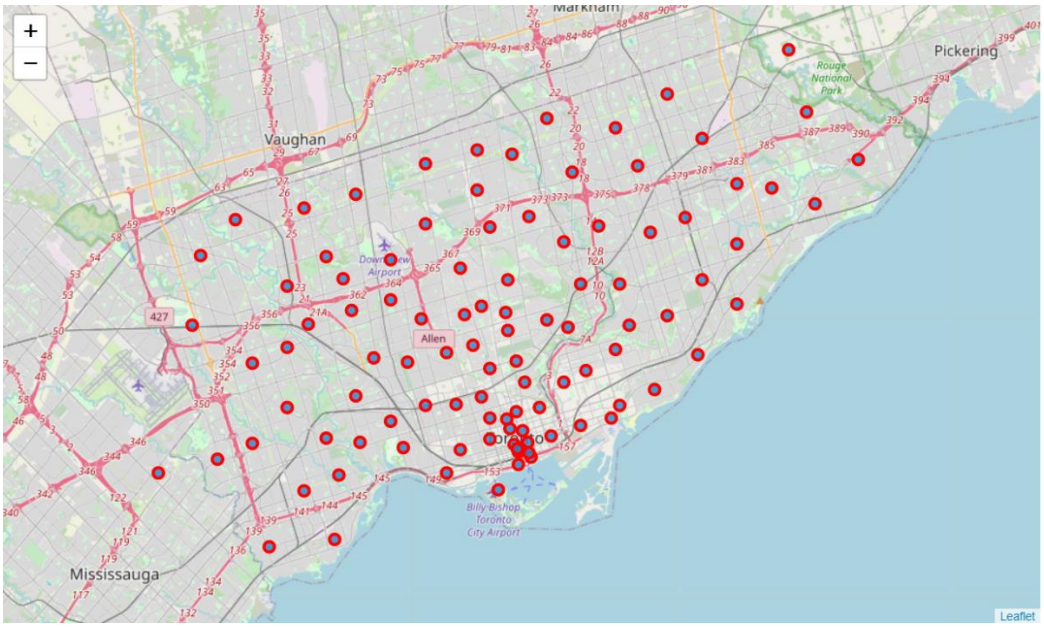

Figure 1: Toronto Neighborhood

Our main goal was to find a suitable neighborhood in *Downtown Toronto* borough. To do that data for *'Downtown Toronto'* was segmented from the total data and then visualized again.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529
1	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.667967	-79.367675
2	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
3	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
4	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937

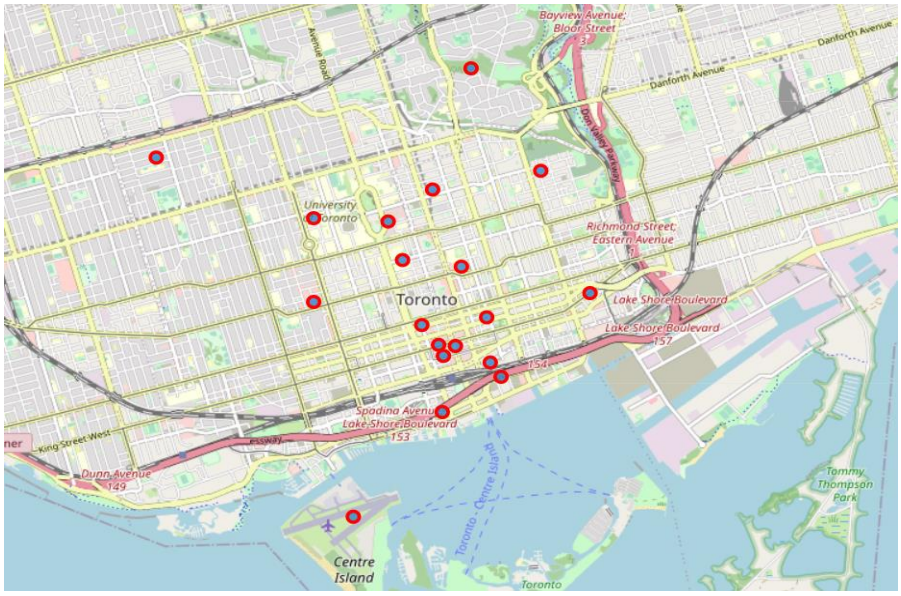

Figure 2: Downtown Toronto neighborhood

Foursquare API was used to collect the relevant information of *'Downtown Toronto'*, such as longitude and latitude venue category and the frequency. The venue (restaurant, park, yoga center etc.) geographical location was merge with the *'Downtown Toronto'*,

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rosedale	43.679563	-79.377529	Rosedale Park	43.682328	-79.378934	Playground
1	Rosedale	43.679563	-79.377529	Whitney Park	43.682036	-79.373788	Park
2	Rosedale	43.679563	-79.377529	Rosedale Tennis Club	43.683226	-79.378984	Tennis Court
3	Rosedale	43.679563	-79.377529	Alex Murray Parkette	43.678300	-79.382773	Park
4	Rosedale	43.679563	-79.377529	Milkman's Lane	43.676352	-79.373842	Trail

Next step was to find the top 10 venue category of each neighborhood. To do that first the neighborhood was grouped by the frequency of each venue category. Then we make a list containing the top 10 category of each neighborhood and then the data was made into a pd DF

	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	Bakery	Cocktail Bar	Seafood Restaurant	Cheese Shop	Beer Bar	Farmers Market	Restaurant	Sandwich Place	Breakfast Spot
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Bar	Harbor / Marina	Rental Car Location	Coffee Shop	Boat or Ferry	Historic Site	Sculpture Garden	Airport Terminal
2	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Japanese Restaurant	Thai Restaurant	Department Store	Salad Place	Burger Joint	Bubble Tea Shop
3	Christie	Grocery Store	Café	Park	Athletics & Sports	Baby Store	Coffee Shop	Nightclub	Candy Store	Italian Restaurant	Restaurant
4	Church and Wellesley	Coffee Shop	Gay Bar	Japanese Restaurant	Sushi Restaurant	Restaurant	Yoga Studio	Hotel	Men's Store	Café	Bubble Tea Shop

We then cluster the neighborhood data with k-means clustering. We got total 5 clustering. All the

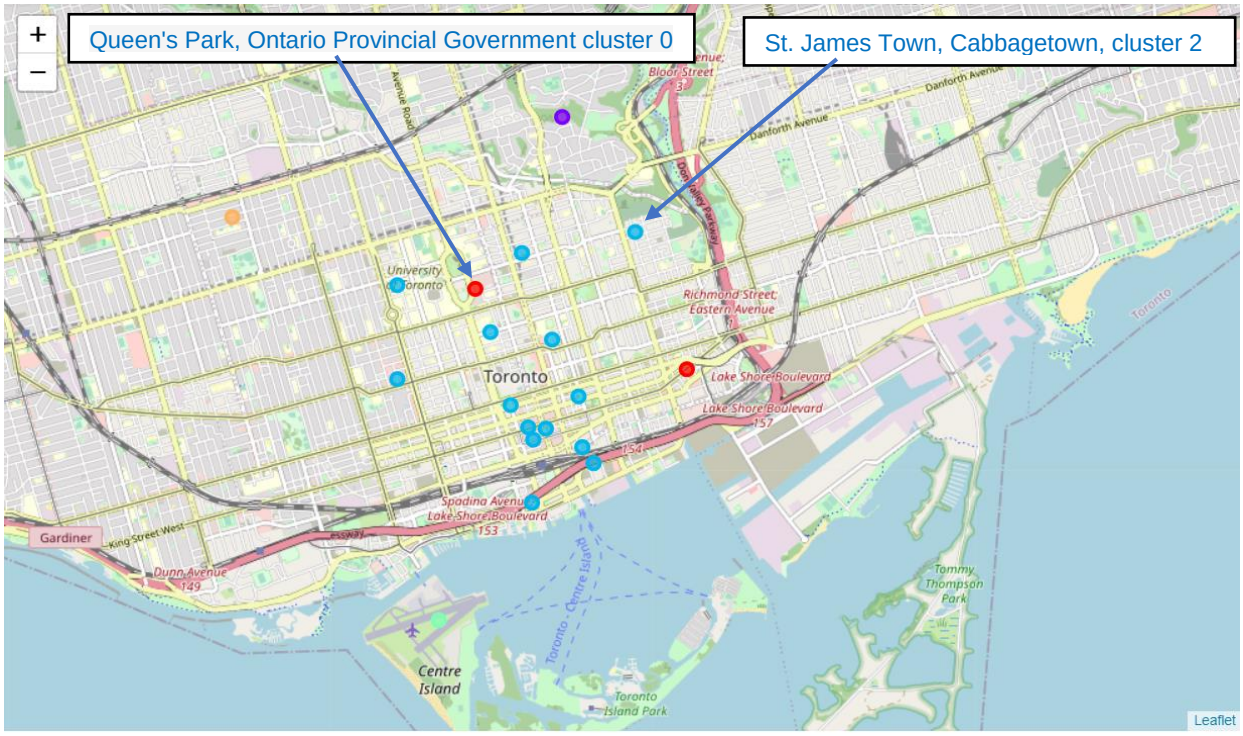

Figure 3: Neighborhood Clustering

the five cluster are shown in the above image.

Results

In *'Downtown Toronto'* borough only two neighborhood has Chinese restaurant at their top 10 frequent venues those are *'St. James Town, Cabbagetown (choice-1)'*, and *'Queen's Park, Ontario Provincial Government (choice-2)'*. Now, choice is in cluster two and choice-2 is in

Neighborhood	Latitude	Longitude	Cluster Labels	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
St. James Town, Cabbagetown	43.667967	-79.367675	2	Coffee Shop	Café	Pizza Place	Restaurant	Pub	Chinese Restaurant	Market	Bakery	Park	Italian Restaurant
Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Yoga Studio	Chinese Restaurant	Sandwich Place	Diner	Music Venue	Beer Bar	Italian Restaurant	Smoothie Shop	Portuguese Restaurant

Cluster-0. In choice-1 Chinese restaurant is 6th most frequent place and choice-2 has Chinese restaurant as the 3rd most frequent venue. Right of the bat this was the best place to open the Chinese office. Other thing is choice-2 is very close to another cluster as can be seen in the Figure-3. This is convenient for all the other activities the employees would like to do since other places are nearby. So the choice of the neighborhood in *'Downtown Toronto'* for the Chinese company would be **'*Queen's Park, Ontario Provincial Government*'**

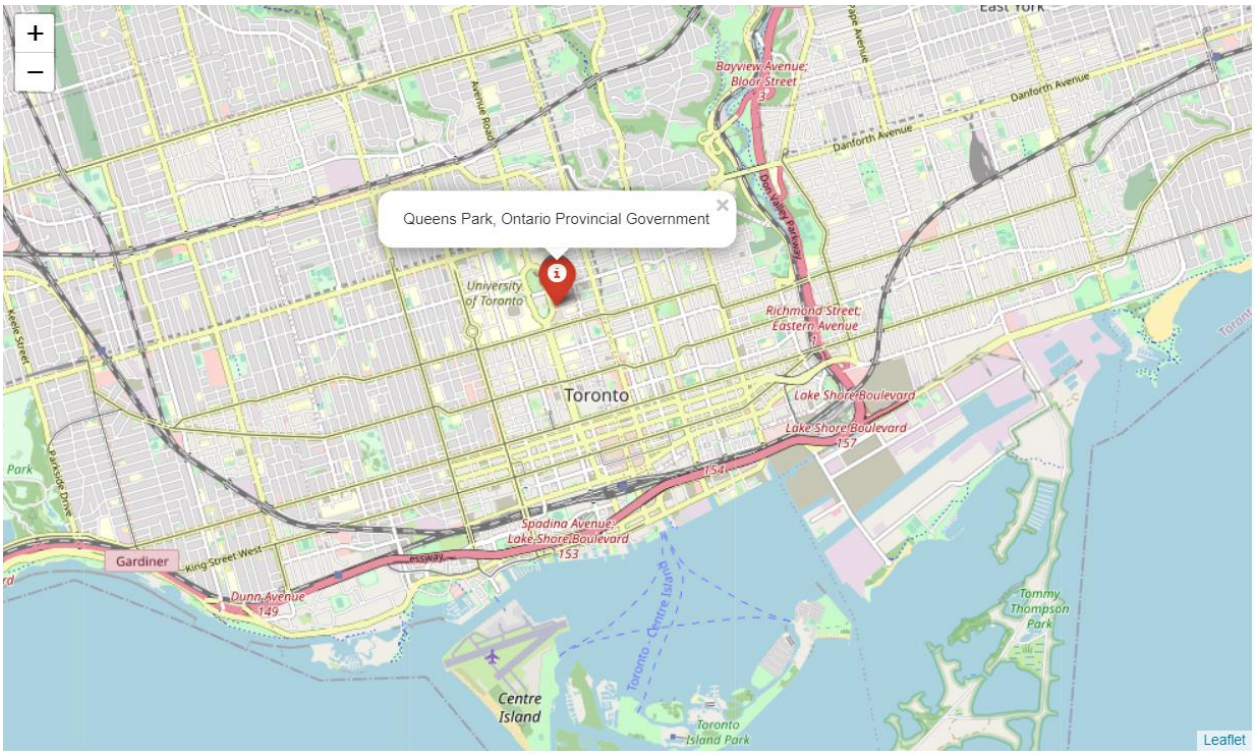

Figure 4: Office location for the Chinese Company