Capstone Project

Week 4 – part 1

Toronto neighborhoods: Where to live

1. Introduction/ Business Problem

Toronto is the most populous city in Canada. This city is an international center of business, finance, arts, and culture. Every year, there are thousands of people choose Toronto as their new destination for living and working. Before their departures, people want to get information about different neighborhoods in Toronto and then decide where they live.

The aim of this project is to describe different neighborhoods in Toronto to help people make choices of which places fit their interests. This project results can help people identify where is quiet, where is more crowded with lots of shopping centers, which places are suitable for cuisine, travelling, etc. From that, they can decide which places fits their lifestyle and interests.

Our target audience is people who is planning to move to Toronto. And people who want to travel to Toronto might care about this problem.

2. Data

2.1 Data Source

I use these **3 data sources** to create the data used for this project, as following:

1. a list of neighborhoods in Toronto is scraped from Wkipedia
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. a link to a csv file that has the geographical coordinates of each postal code
http://cocl.us/Geospatial_data

3. a list of venue categories is requested from Foursquare API
https://developer.foursquare.com/docs/build-with-foursquare/categories/.

The description of each dataset will be given in the 2.2 part.

2.2 Data preparation and data description

To describe the different between Toronto neighborhoods, the feature of neighborhood (from Toronto neighborhood data) and feature of Venues and Main venues category (from Foursquare dataset) will be extracted.

For the Toronto neighborhood data, I will prepare the data as followed:

+ Scrape the data from Wikipedia.

+ Drop cells with a borough that is Not assigned

+ Read it into a pandas dataframe. It is in a structured format with 3 columns: Postal Code, Borough, and Neighborhood; more than one neighborhood can exist in one postal code area.

After scraping from Wikipedia, the list is transformed into a pandas dataframe and merged with the geographical coordinates file. Finally, the data looks like this:

	Postal Code	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

This data contains 10 boroughs and 103 neighborhoods.

Using Foursquare API to explore the neighborhood and then segment them to get the data like this:

	name	categories	lat	lng
0	Downtown Toronto	Neighborhood	43.653232	-79.385296
1	Nathan Phillips Square	Plaza	43.652270	-79.383516
2	Indigo	Bookstore	43.653515	-79.380696
3	Chatime 日出茶太	Bubble Tea Shop	43.655542	-79.384684
4	Textile Museum of Canada	Art Museum	43.654396	-79.386500

After that, these two above data are merged into one as the following:

	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	Malvern, Rouge	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank

I also request from Foursquare API a list of venues and main venue category like this:

	Venue Category	Main Venues Category
0	Arts & Entertainment	Arts & Entertainment
1	Amphitheater	Arts & Entertainment
2	Aquarium	Arts & Entertainment
3	Arcade	Arts & Entertainment
4	Art Gallery	Arts & Entertainment
5	Bowling Alley	Arts & Entertainment
6	Casino	Arts & Entertainment
7	Circus	Arts & Entertainment
8	Comedy Club	Arts & Entertainment
9	Concert Hall	Arts & Entertainment
10	Country Dance Club	Arts & Entertainment
11	Disc Golf	Arts & Entertainment

The Foursquare venues data and merged Toronto data will merge to each other to get the final data:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Main Venues Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant	Food
1	Malvern, Rouge	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop	Shop & Service
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar	Nightlife Spot
3	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target	Travel & Transport
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank	Shop & Service

The feature of Neighborhood and Venue Category and Main Venues Category are extracted to describe the different between Toronto neighborhood.