Capstone Project - The Battle of Neighborhoods (Week 2)

Introduction

Toronto, the most populous city in Canada, is an international center of business, finance, arts, and culture. Its economy is highly diversified with strengths, such as technology, financial services, education, art, and tourism. [1] In the city of Toronto, booksellers could enjoy being part of a community, encouraging the pleasure of reading in adults, and helping to make lifelong readers out of children. For someone who is looking to open a bookstore, it is vital to choose the neighborhood and retail location. The goal of this project is to figure out where a bookstore should be set up for success with data analysis.

Data Acquisition

1. Neighborhoods in Toronto -- Wikipedia[2].
2. Using Geopy to get geological location by address name.
3. Using Foursquare API to get the most common venues of given Borough of Toronto.
4. Using Foursquare API to get the venues' record of given venues of Toronto.

Methodology

1. Scrape the Wikipedia page [2] and transform it into a pandas dataframe. Postal codes beginning with M are located within the city of Toronto. The dataframe consists of three columns: ‘Postcode’, ‘Borough’, ‘Neighborhood’. Since some cells are ‘Not assigned’, I drop the cells with a borough that is ‘Not assigned’, and change the cells’ neighborhood to borough for cells having a ‘Not assigned’ neighborhood. I also combine Neighborhoods with same postal code. The dataframe cleaned has three columns and 103 rows. The first five rows are shown below.

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

2. Acquire the data of latitude and the longitude coordinates in Toronto using Geopy. And combine the geological location of each neighborhood with the dataframe above. Here I have the dataset that contains Postcode, Borough, Neighborhood, Latitude, Longitude of each neighborhood. The table below shows the first rows of the new dataframe. To brief the project, I choose only to analyze the boroughs that contain the word 'Toronto'.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

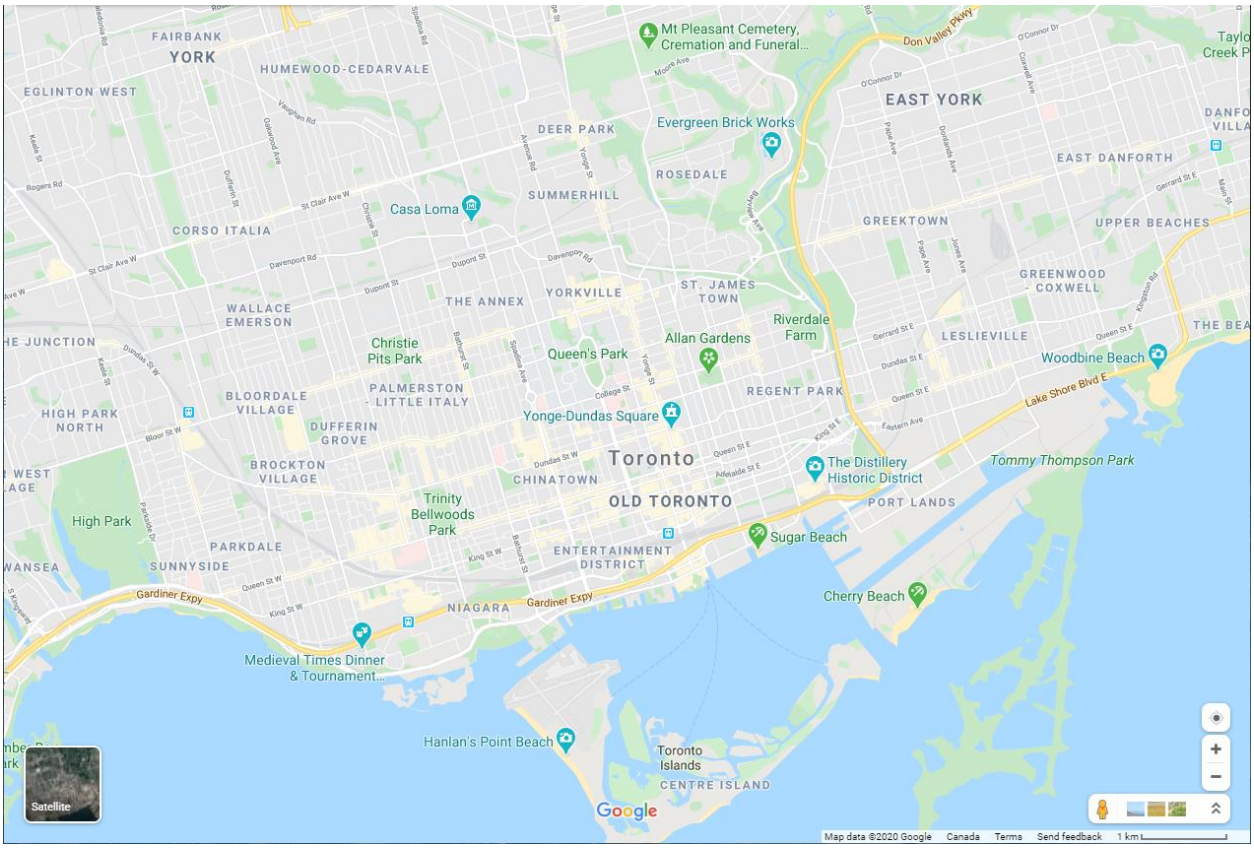
3. Explore the boroughs and neighborhood in Toronto using Foursquare API. The limit is set as 100 venues, and the radius is set as 500 meters for each borough from their given latitude and longitude. As the header of the dataset shows below, the dataset contains Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, Venue Category.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

4. Explore the bookstores in Toronto using Foursquare API. I got the neighborhood and bookstore information, sorted by the number of bookstores in a neighborhood. The first five rows are shown below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Bookstore
0	Adelaide, King, Richmond	43.650571	-79.384568	200
1	Ryerson, Garden District	43.657162	-79.378937	200
2	St. James Town	43.651494	-79.375418	100
3	First Canadian Place, Underground city	43.648429	-79.382280	100
4	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576	100

5. Use *k-means* to cluster and segment the bookstores in Toronto. And use *folium* to visualize the distribution of these bookstores in Toronto.

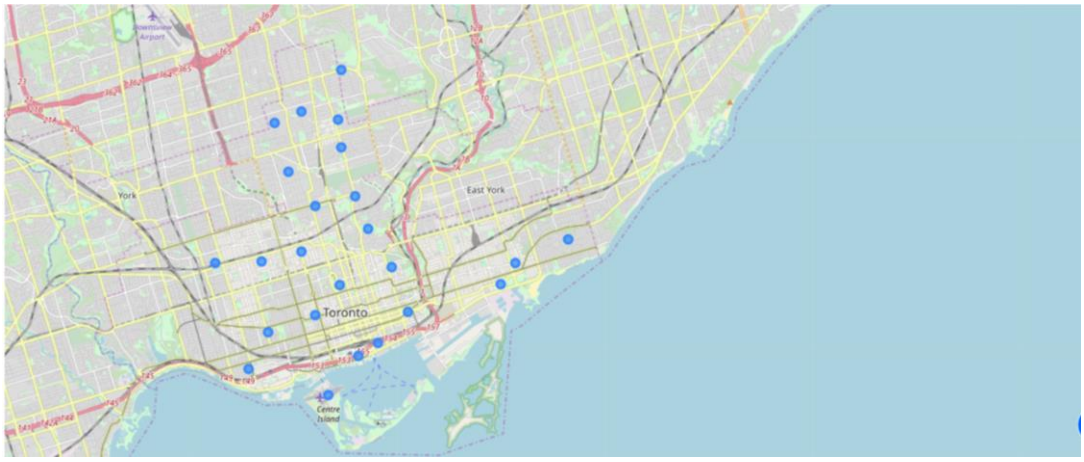
6. 

Conclusions

1. We can extract meaningful information from the dataset explored. The dataframe was sorted by the number of bookstores in a neighborhood, as shown below. Here I try to find optimal neighborhoods to start up a bookstore. Since there are lots of bookstores in Toronto, I would not recommend neighborhoods that are already crowded with bookstores. I find the neighborhoods with no bookstore, which are:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Bookstore
15	Christie	43.669542	-79.422564	0
16	The Beaches West, India Bazaar	43.668999	-79.315572	0
17	The Beaches	43.676357	-79.293031	0
18	The Annex, North Midtown, Yorkville	43.672710	-79.405678	0
19	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	0
20	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	0
21	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	0
22	Roselawn	43.711695	-79.416936	0
23	Rosedale	43.679563	-79.377529	0
24	Queen's Park	43.662301	-79.389494	0
25	Moore Park, Summerhill East	43.689574	-79.383160	0
26	North Toronto West	43.715383	-79.405678	0
27	Davisville	43.704324	-79.388790	0
28	Little Portugal, Trinity	43.647927	-79.419750	0
29	Lawrence Park	43.728020	-79.388790	0
30	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752	0
31	Berczy Park	43.644771	-79.373306	0
32	Forest Hill North, Forest Hill West	43.696948	-79.411307	0
33	Cabbagetown, St. James Town	43.667967	-79.367675	0
34	Dovercourt Village, Dufferin	43.669005	-79.442259	0
35	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	0
36	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049	0
37	Davisville North	43.712751	-79.390197	0
38	Harbourfront	43.654260	-79.360636	0

2. Showing in the map above, we get the location of the neighborhoods without a bookstore. These neighborhoods will have no competition for a bookstore owner. Other than that, this map also gives location information, for example, whether the location is close to city center.



Discussion

This report gives a recommendation of neighborhood and location to those who plan to open a bookstore. In real world, there must be more factors to consider, such as the cost of the location. What's more, the analysis of this report cannot solve the problem of how many customers will visit every day. With the data analysis above, the report will be constructive to open a bookstore in Toronto.

Reference

[1] Toronto -Wikipedia, https://en.wikipedia.org/wiki/Toronto

[2] List of postal codes of Canada: M -Wikipedia,

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M