Exploring Start-up Ideas in Toronto

Sourabh Surage

May 6, 2020

Introduction

In this project our main focus is to find a suitable and popular business that promises to be successful in the city of Toronto. This project will be particularly beneficial for entrepreneurs who want to start a business but don't have any ideas in particular.

Plan is to find popular business in multiple neighborhoods of Toronto and see how many of them are there in that particular area, we'll also do a frequency count of popular venues in Toronto (to see the choice of people)

After that we'll identify neighborhoods in Toronto where those popular business are either absent or the frequency is very less, thus we'll choose a neighborhood which is farthest from the popular venues (and don't have one) so that we can set-up our business idea in that area with high chances of getting success.

This will also be beneficial for the community of that neighborhood since they'll not have to go to long places to get the same thing which is now in their own neighborhood.

Data Acquisition and Cleaning

Data Sources

We first get our primary data from Wikipedia where we get all details of all neighborhoods with their respective boroughs and postal code. Then we combine this data with the latitude and longitude data table, we perform a left merge to get the following table.

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

The table above shows the raw data that we get from the Wikipedia page, since that data has missing neighborhoods and missing boroughs, it requires to be cleaned which we are going to discuss in the next sub-section of this section.

There was also a little issue of backslashes that we have in the neighborhoods columns, which is pretty visible in the table shown above, which we'll replace with commas so that the data looks better.

Next, we use Foursqare to get the list of venues (100 at maximum) in the radius of 500m from all the neighborhoods, the table looks like following:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lawrence Park	43.728020	-79.388790	Lawrence Park Ravine	43.726963	-79.394382	Park
1	Lawrence Park	43.728020	-79.388790	Zodiac Swim School	43.728532	-79.382860	Swim School
2	Lawrence Park	43.728020	-79.388790	TTC Bus #162 - Lawrence-Donway	43.728026	-79.382805	Bus Line
3	Roselawn	43.711695	-79.416936	Rosalind's Garden Oasis	43.712189	-79.411978	Garden
4	Davisville North	43.712751	-79.390197	Homeway Restaurant & Brunch	43.712641	-79.391557	Breakfast Spot

This completes the use of sources to get the data, next we'll look at how we cleaned it to make sense out of the data and get important inferences.

Data Cleaning

The data that we got from the Wikipedia page had some missing neighborhoods which we removed, also it contained backslashes which we replaced with commas.

After that we formed a table where we can see the frequency of a particular venue category corresponding to that particular neighborhood, the table looked like following:

	Neighborhood	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	A
0	Agincourt	0	0	0	0	0	0	0	0	0	
1	Alderwood, Long Branch	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	
3	Bayview Village	0	0	0	0	0	0	0	0	0	
4	Bedford Park, Lawrence Manor East	0	0	0	0	0	0	0	1	0	

Here, on we have assigned columns for particular venue category and the values represent the number of those venue category corresponding to that neighborhood.

Getting the number of venue categories in whole Toronto was also necessary to identify the businesses, for that we created a frequency table which shows the number of venues for that particular category in Toronto. The table looks like follows:

	Venue Category	frequency
0	Accessories Store	2
1	Airport	2
2	Airport Food Court	1
3	Airport Gate	1
4	Airport Lounge	2

This completes our data gathering and cleaning of the data for further analysis.

Exploratory Data Analysis

Frequency of venue categories

Our first and foremost analysis that needed to be performed was getting the total number of occurrences of all venue categories in whole Toronto so that we can identify those businesses which have a good frequency and are popular in the city of Toronto, we sorted a following table which helps us in the analysis:

	Venue Category	frequency
0	Coffee Shop	182
1	Café	101
2	Restaurant	68
3	Park	51
4	Pizza Place	46

Frequency cut-off

After looking at the table, we realised that it might not be a good idea to go ahead with the business with occurrences higher than the total number of neighborhoods (which is 103 in our case), so we decided to drop the venues which have a frequency of more than 100 (so coffee shop and Café goes out from out list).

The idea behind that was that if anything is equivalent to the number of neighborhood than it is more likely that the distribution of those venues are spread across all the neighborhoods, which might not be true (one can say that) but this is a fair assumption that we made.

It was also decided to drop the restaurants because there were lot of categories of those e.g. Japanese, American, Mexican etc. which were classifies as different venue category in our table but if we sum that up, then it goes way above the mark of 100 that we did set formerly.

Now the idea was to get top 3 businesses which are feasible and different from restaurants so we analysed the following table to find out the latter:

	Venue Category	frequency
0	Restaurant	68
1	Park	51
2	Pizza Place	46
3	Italian Restaurant	46
4	Hotel	43
5	Japanese Restaurant	42
6	Sandwich Place	39
7	Bakery	38
8	Clothing Store	37
9	Gym	34

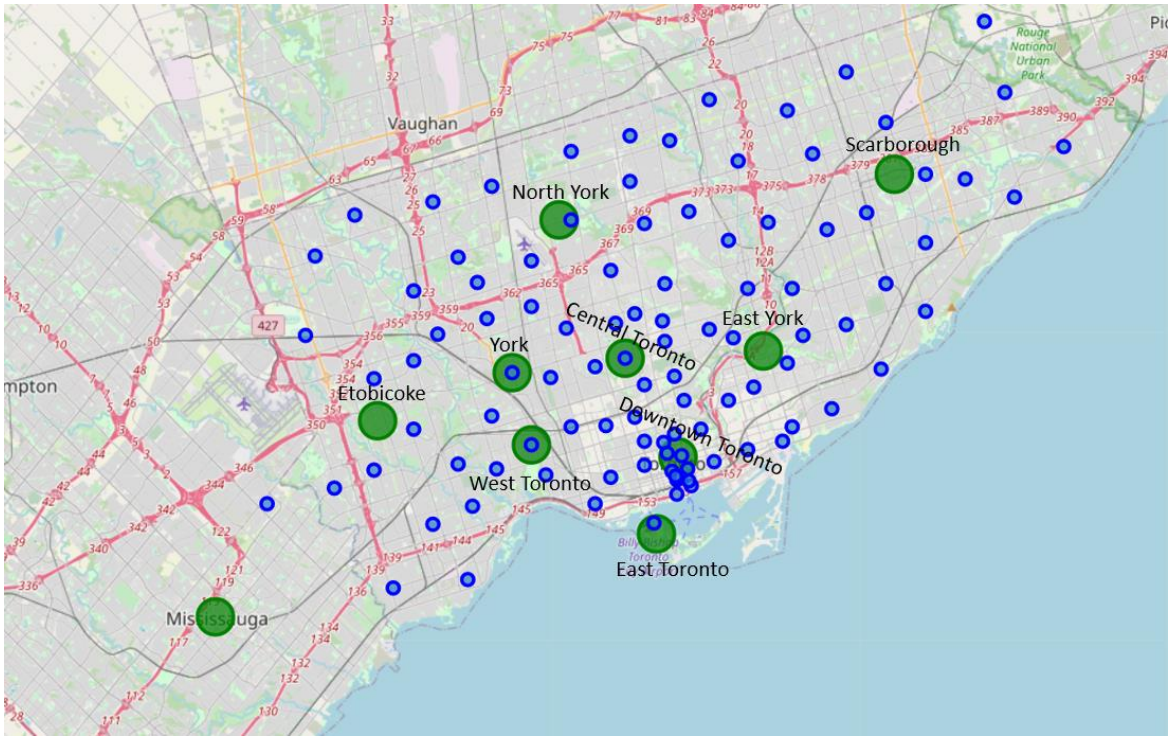
From this table we picked the 3 businesses as: BAKERY, CLOTHING STORE and GYM

Visualizing the Boroughs

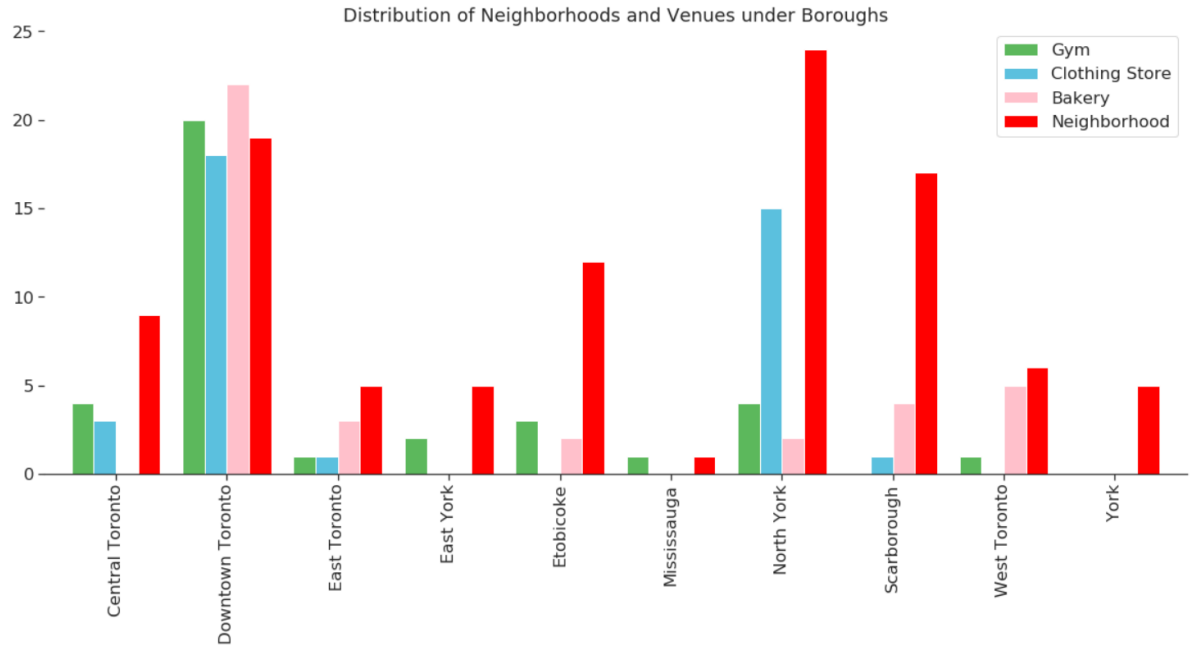
Now that we are clear which businesses, we are interest in investing our time, we continued the analysis by getting the count of the business of our interest with the corresponding boroughs, which looked like following:

	Borough	Gym	Clothing Store	Bakery
0	Central Toronto	4	3	0
1	Downtown Toronto	20	18	22
2	East Toronto	1	1	3
3	East York	2	0	0
4	Etobicoke	3	0	2
5	Mississauga	1	0	0
6	North York	4	15	2
7	Scarborough	0	1	4
8	West Toronto	1	0	5
9	York	0	0	0

Now, let's understand the positioning of the boroughs on the map that we generated using folium (green circles)



Now, we know the positioning of the boroughs the analysis is continued by plotting a bar graph to better visualize and pick boroughs and which business to do at which borough.



Let's filter out those Boroughs with high number of neighborhoods

- North York
- Downtown Toronto
- Scarborough
- Etobicoke

Getting the Business Ideas

All of the above boroughs have > 10 neighborhoods, now let's analyse each of them individually

1. North York: Contains very high number of clothing stores with some amount of gym and bakery – so setting-up gym or bakery would be feasible
2. Downtown Toronto: All three venues are in plenty and equivalent to the number of neighborhood – let's not have any business here
3. Scarborough: There are very few bakery and clothing store, but we notice that there is not gym here – Opening-up a gym would really be great in this borough
4. Etobicoke: Condition is somewhat similar to Scarborough but here there is no Clothing Store – Opening-up a Clothing Store would be a good idea here

Results and Discussion

After doing a thorough analysis by we can suggest two ideas confidently which promise to get successful in the suggested boroughs, the ideas that are being suggested are as follows:

- Gym in Scarborough
- Clothing Store in Etobicoke

We saw that these are the boroughs that have a good number of overall neighborhoods (>10) but don't have the category of venue that is suggested by us, which we believe is going to be a good hit since it does not exist in that borough at all.

Since, we are suggesting our idea to an entrepreneur, we must also see if we can suggest the rates or the amount of money that'll be required to open-up those businesses and could perform a economic analysis for the same.

One should also note that, we have given the above ideas considering non-corona situation as the cases in Canada are rising and nobody would certainly go to a gym or go out to buy clothes. This is a limitation which we can understand that is out of scope for this project as our main focus was to use the Foursqare data and make sense out of that.

So, when things get normal and the COVID situation goes away, then this is something we can start for sure.

Conclusion

We can conclude that opening-up a Gym at Scarborough and a Clothing Store at Etobicoke in Toronto can be a great business idea to do when things are according to BAU (Business As Usual) and we don't have any situation like of the current COVID-19.

Also finished our main objective of the project by giving business ideas to those start-ups who want to start something and have the money but don't know what they can do.

Provided start-ups with 2 ideas with the location where they can set-up their business as decided in the definition of the problem statement at the start.