The Battle of Neighborhoods

EDUARDO DÍAZ

PROBLEM & BACKGROUND

Toronto is an immense city, with a population of 2,731,571 as of 2016. this fact is enough for consider Toronto a city full of business opportunities. Some people need a little help getting started on the tough but rewarding journey of starting their own business.

Starting a business is a difficult feat, so if using the knowledge obtained in this IBM course, we can help an entrepreneur fulfill his objective with less risk. This type of work can be a really good job opportunity in the future, there's always people with ideas of new projects and can be a good source of experience and employment.

My idea for this project consist in find the feasibility of a new gymnasium in a specific Neighborhood in the city of Toronto, using the data of the city we can find a neighborhood o a cluster of them with the best combinations for this project like gymnasiums in similar venues.

We want the least amount of competitors for a solid and more easy start, and lastly if we find more than one valid option is even better because the client can choose based in other parameters like the building, the rent of this one and so on.

DATA DESCRIPTION

the basis of our data is the foursquare API to explore data from the city of Toronto. The data also include the information about the places around each neighborhood like restaurants of any kind, coffee shops, parks, theaters, art galleries, museums gymnasiums and a lot more.

A single Borough was selected to analyze their neighborhoods, this borough can be easily changed based on the necessities of a client, the same goes for the city.

METODOLOGY

We start by extracting the data from a table on a Wikipedia page using the Beautiful Soup library.

Postal code	Borough	Neighborhood
M1A	Not assigned	
M2A	Not assigned	
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park / Harbourfront
M6A	North York	Lawrence Manor / Lawrence Heights
M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government
M8A	Not assigned	
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Malvern / Rouge

Figure 1. list of postal codes in Toronto

Then the data is arranged according to our requirements. In the arrangement phase we eliminate the “Not assigned” values, combine neighborhoods which have same geographical coordinates at each borough and sorted against the concerned borough. For data verification and further exploration, we use Foursquare API to get the coordinates of the desired borough and explore its neighborhoods. The neighborhoods are further characterized as venues and venue categories.

	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
8	M1M	Scarborough	Cliffside / Cliffcrest / Scarborough Village West	43.716316	-79.239476

Figure 2. boroughs and their neighborhood with their corresponding ubication.

```
Rosedale
St. James Town / Cabbagetown
Church and Wellesley
Regent Park / Harbourfront
Garden District, Ryerson
St. James Town
Berczy Park
Central Bay Street
Richmond / Adelaide / King
Harbourfront East / Union Station / Toronto Islands
Toronto Dominion Centre / Design Exchange
Commerce Court / Victoria Hotel
University of Toronto / Harbord
Kensington Market / Chinatown / Grange Park
CN Tower / King and Spadina / Railway Lands / Harbourfront West / Bathurst Quay / South Niagara / Island airport
Stn A PO Boxes
First Canadian Place / Underground city
Christie
Queen's Park / Ontario Provincial Government
```

Figure 3. all the neighborhood from Downtown Toronto

We analyze the borough through one hot encoding (giving ‘1’ if a venue category is there, and ‘0’ in case of venue category is not there). Based on one hot encoding, we calculate mean of the frequency of occurrence of each category and picked top ten venues on that basis for each neighborhood. with this we are showing the more popular categories in a neighborhood.

We then use k means to divide the neighborhood into clusters and determine the discriminating venue categories that distinguish each cluster, if we find a cluster that have more amount of gyms for example, then the neighborhoods that are part of it aren't good options for opening a new gymnasium.

RESULTS

After analyzing the result of the clustering, is incredible the amount of information we can extract an utilize for starting a business. Doing the same exercise but with a different borough, we can probably find the best location for a gymnasium or any business in general.

in the assignment, the cluster number 5 was the one who possess a greater number of gymnasiums and yoga studios, we can say that this cluster already have gymnasiums so does not need more. but we can say that this cluster is the ideal for gymnasiums because people already tried to start a business there and succeeded, and in our cluster 5 or "cluster gym" there is a neighborhood out of the 4 who compose the "cluster gym" who doesn't have gyms in the 10th common venues, so this can be an opportunity for an entrepreneur.

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 5, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	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
2	Church and Wellesley	Sushi Restaurant	Coffee Shop	Japanese Restaurant	Restaurant	Yoga Studio	Men's Store	Hotel	Gay Bar	Gastropub	Pub
8	Richmond / Adelaide / King	Coffee Shop	Café	Restaurant	Gym	Thai Restaurant	Hotel	Clothing Store	Deli / Bodega	American Restaurant	Bookstore
10	Toronto Dominion Centre / Design Exchange	Coffee Shop	Hotel	Café	Restaurant	Japanese Restaurant	Salad Place	Italian Restaurant	Seafood Restaurant	American Restaurant	Sporting Goods Shop
11	Commerce Court / Victoria Hotel	Coffee Shop	Restaurant	Café	Hotel	Gym	American Restaurant	Seafood Restaurant	Deli / Bodega	Italian Restaurant	Japanese Restaurant

Figure 4. cluster 5 or "gym cluster"

DISCUSSION

There´s a lot of ground to improve in this kind of "starting business" type of work, we can, for example improve in the database section, depending on a Wikipedia page is not the best, we can also improve in the method, maybe another type of machine learning can do a better job, and lastly we can do a better cleaning of the data to improve the final output.

CONCLUSION

As a conclusion, the result was satisfactory, we find hints in where we can start our gymnasium business, but a decision as difficult as this one need more than hints to be taken. Overall is a good way to have an idea of a place to start looking for the ideal neighborhood to start our business.