IBM Coursera Capstone Project Report

December 21, 2020

Contents

List of Figures

1 Introduction: Business Problem

In this project, we are going to compare the neighborhoods of the New York City in the US and the city of Toronto in Canada, and determine how similar they are in terms of businesses. We will analyze for both cities:

- what types of businesses are more likely to thrive;
- what neighborhoods are suitable for each type of business;
- what types of businesses are less desirable.

The results of the project will enable more effective decisions making for business people who want to start their business in these two big cities.

2 Data

Based on definition of our problem, following data sources will be needed to extract/generate the required information:

- Neighborhood coordinates for New York from IBM Developer Skills Network
- Neighborhood coordinates for Toronto from Wikipedia and Geospatial data
- Venue data for both New York and Toronto from FOURSQURE

We load the downloaded neighborhood data for New York into a pandas dataframe (see Figure 2.1).

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 2.1: Head of the neighborhood data table for New York City

We download the neighborhood names for Toronto from Wikipeda and the coordinate data from http://cocl.us/Geospatial_data (see Figure 2.2).

	PostalCode	Borough	Neighborhood		Postal Code	Latitude	Longitude
0	M1A	Not assigned	Not assigned	0	M1B	43.806686	-79.194353
1	M2A	Not assigned	Not assigned	1	M1C	43.784535	-79.160497
2	M3A	North York	Parkwoods	2	M1E	43.763573	-79.188711
3	M4A	North York	Victoria Village	3	M1G	43.770992	-79.216917
4	M5A	Downtown Toronto	Regent Park, Harbourfront	4	M1H	43.773136	-79.239476

Figure 2.2: Head of neighborhood names table and the coordinates table for Toronto

We then combine these two tables for Toronto into one table (see Figure 2.3).

	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 2.3: Head of the combined table for Toronto

3 Methodology and Results

We use pandas.DataFrame.value_counts method to find Top 10 Common Venue Categories in NYC and Toronto.

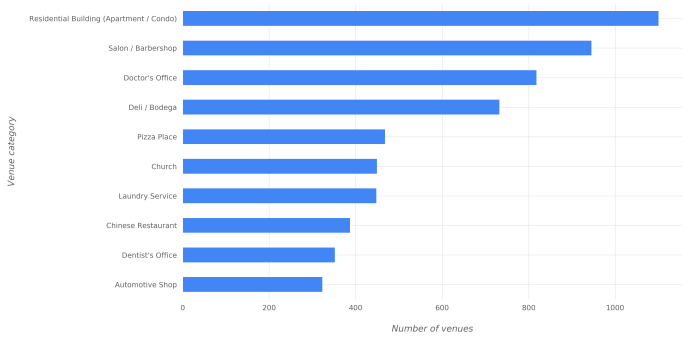

Figure 3.1: Top 10 Common Venue Categories in NYC

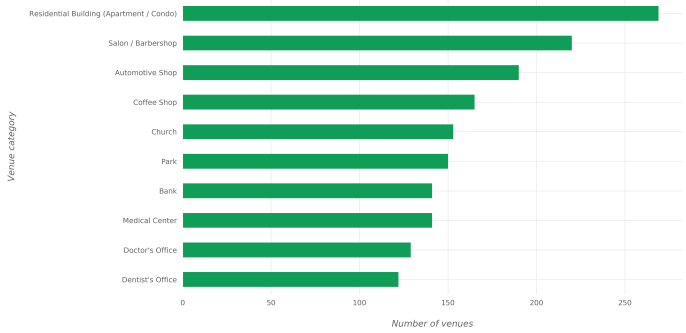

Figure 3.2: Top 10 Common Venue Categories in Toronto

We also find the Top 10 Common Venue Categories for NYC and Toronto together:

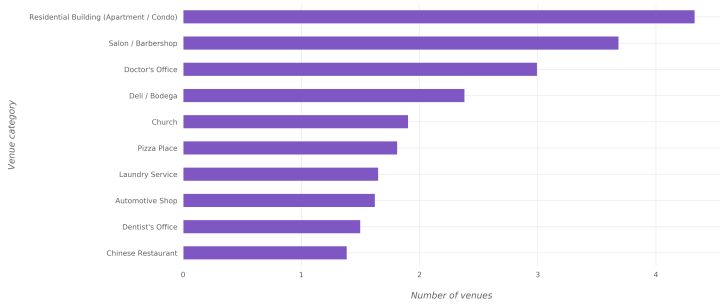

Figure 3.3: Top 10 Common Venue Categories for NYC-Toronto

With the same method, we also find the Rare Venue Categories for NYC and Toronto and NYC-Toronto.

Venue Category	Count	Venue Category	Count	Venue Category	Count
Colombian Restaurant	1	Cable Car	1	Aquarium	1
Light Rail Station	1	Language School	1	Swim School	1
Line / Queue	1	Roof Deck	1	Roller Rink	1
Hunan Restaurant	1	Baggage Claim	1	Australian Restaurant	1
Costume Shop	1	Carpet Store	1	Newsagent	1
Roller Rink	1	Ski Chalet	1	Airport Food Court	1
Cable Car	1	Pastry Shop	1	Toll Booth	1
Caucasian Restaurant	1	College Math Building	1	Piercing Parlor	1
Forest	1	Stables	1	Stadium	1
Adult Boutique	1	Buffet	1	Rugby Pitch	1

Figure 3.4: Rare Venue Categories for NYC (left), Toronto (middle), and NYC-Toronto (right)

We apply K-means clustering on the dataframe stored in `nyc_tor_grouped_clustering` variable which includes the relative frequency of each venue-category for each neighborhood.

Neighborhood	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
Wingate_NYC	3	Salon / Barbershop	School	Deli / Bodega	Caribbean Restaurant	Event Space	Food	Residential Building (Apartment / Condo)
Woodhaven_NYC	3	Deli / Bodega	Salon / Barbershop	Laundry Service	Miscellaneous Shop	Liquor Store	Doctor's Office	Chinese Restaurant
Woodlawn_NYC	3	Bar	Deli / Bodega	Salon / Barbershop	Pub	Church	Food & Drink Shop	Pizza Place
Woodrow_NYC	4	Pool	Grocery Store	School	Physical Therapist	Dentist's Office	Salon / Barbershop	Bar
Woodside_NYC	3	Bar	Salon / Barbershop	Mexican Restaurant	Thai Restaurant	Platform	Deli / Bodega	Miscellaneous Shop
Yorkville_NYC	2	Residential Building (Apartment / Condo)	Laundry Service	Spa	Pharmacy	Flower Shop	Gym	Salon / Barbershop
Agincourt_Toronto	0	Automotive Shop	Storage Facility	Post Office	Church	Coffee Shop	Gas Station	Business Service
Alderwood, Long Branch_Toronto	4	Convenience Store	Salon / Barbershop	Spa	Dentist's Office	Gas Station	Daycare	Bank
Bathurst Manor, Wilson Heights, Downsview North_Toronto	1	Doctor's Office	Residential Building (Apartment / Condo)	Medical Center	Synagogue	Bank	Spa	Ice Cream Shop
Bayview Village_Toronto	4	Salon / Barbershop	Doctor's Office	Church	Dog Run	Bank	Grocery Store	Japanese Restaurant

Figure 3.5: K-means clustering

For each cluster, we list the top 7 venues:

Cluster 1:

Category	% of venues
Automotive Shop	13.9286
Church	2.78571
Gas Station	2.5
Salon / Barbershop	2.21429
Factory	2
Deli / Bodega	1.92857
Pizza Place	1.85714

Cluster 2:

Category	% of venues
Doctor's Office	13.5384
Residential Building (Apartment / Condo)	5.23145
Dentist's Office	3.8364
Medical Center	2.88523
Salon / Barbershop	2.40964
Deli / Bodega	2.34623
Laundry Service	1.6487

Cluster 3:

Category	% of venues
Residential Building (Apartment / Condo)	14.2916
Salon / Barbershop	3.28209
Deli / Bodega	2.61737
Laundry Service	2.30577
Doctor's Office	2.285
Church	1.72414
Park	1.51641

Figure 3.6: Top 7 venues in Cluster 1–3

Cluster 4:

Category	% of venues
Salon / Barbershop	7.343
Deli / Bodega	3.98551
Pizza Place	2.47585
Church	2.35507
Laundry Service	2.29469
Chinese Restaurant	2.18599
Residential Building (Apartment / Condo)	2.04106

Cluster 5:

Category	% of venues
Residential Building (Apartment / Condo)	2.3309
Salon / Barbershop	2.0878
Church	1.7446
Doctor's Office	1.7303
Park	1.6016
Pizza Place	1.5873
Deli / Bodega	1.4014

Figure 3.7: Top 7 venues in Cluster 4–5

4 Discussion

Our analysis shows that NYC and Toronto have the same Top 2 common venue categories: apartments and barbershops.

The Top 10 Common Venue Categories suggest what types of businesses are more likely to thrive.

The Rare Categories suggest what types of businesses are less desirable.

5 Conclusion

The neighborhoods of New York City and Toronto were clustered into multiple groups based on the categories of the venues in these neighborhoods. The results show that there are venue categories that are more common in some cluster than the others.