

IBM Coursera Capstone Project Report

December 22, 2020

Introduction

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.

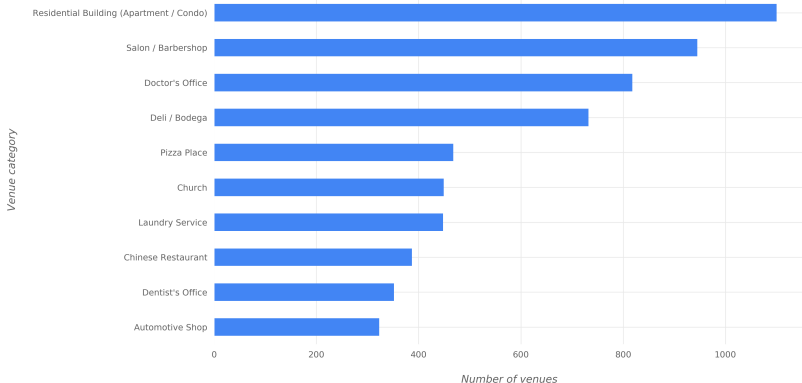
Data

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

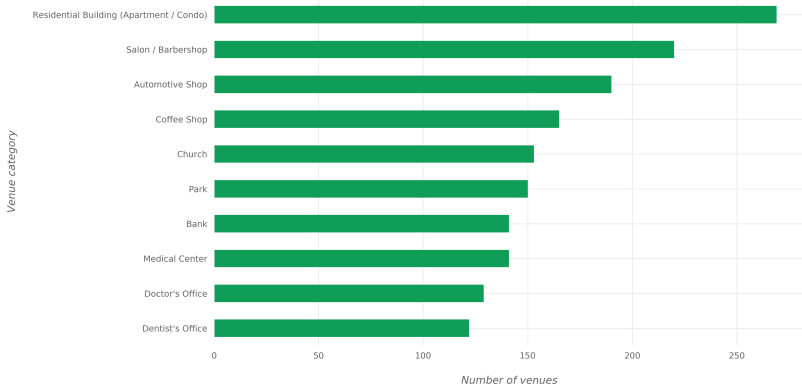
- Neighborhood coordinates for New York from IBM Developer Skills Network
- Neighborhood coordinates for Toronto from Wikipedia and Geospatial data
- Venue data for both New York and Toronto from FOURSQUIRE

Methodology and Results

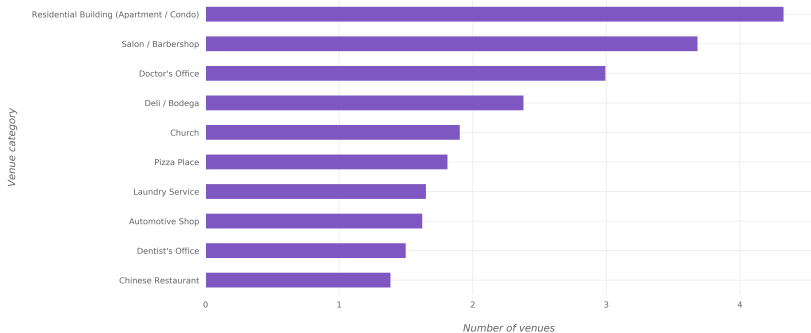
Top 10 Common Venue Categories in NYC



Top 10 Common Venue Categories in Toronto



Top 10 Common Venue Categories for NYC-Toronto



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

Count		Count		Count	
Venue Category		Venue Category		Venue Category	
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

K-means clustering

Cluster 1:

% of venues	
Category	
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:

% of venues	
Category	
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

K-means clustering

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

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

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.

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.