Starting a New Business: Comparing New York and Toronto Neighborhoods

Waad Kahouli
May 2020

I. Introduction

When it comes to the best cities to live in North America, NYC and Toronto top the list. As the vibrant multicultural melting pots they are, both cities are considered some of the most attractive cities to live, especially for young millennials. In this project, we will explore the differences and similarities in both cities for someone who is interested in starting a new business. At the end of the report, we will recommend what would be the best business of each city.

NYC

Toronto

Problem

In this project, we will answer the question of what would be the best type business to open in each city. In order to do this, we will first look at the most common businesses in each city. Second, we will also look at the unique businesses combination in each city and finally we will compare both cities for a given type of business of interest.

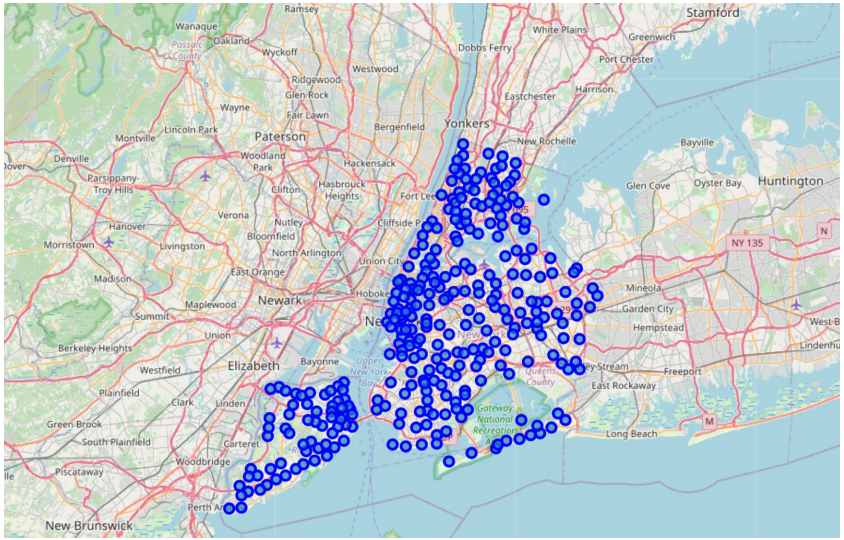
II. Data

For this project we will use two types of accessible data.

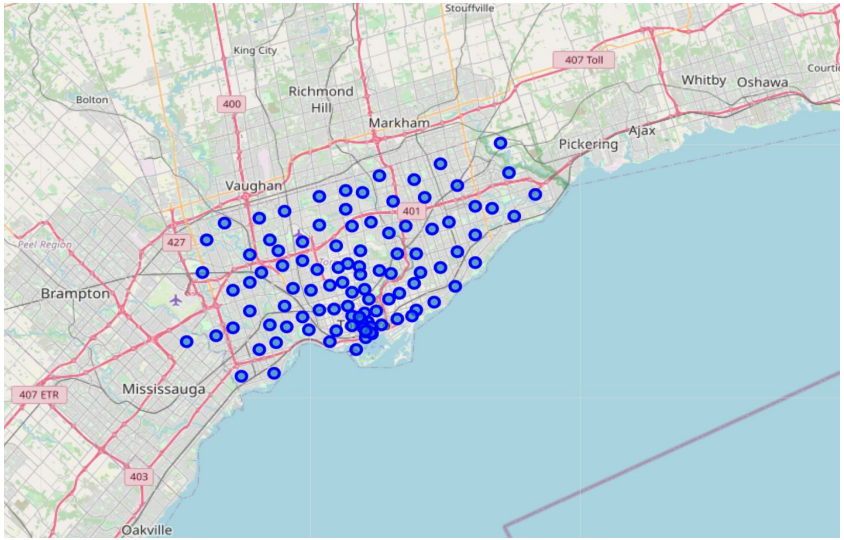
1. First we will use geographical data for both cities. This data encompases data about the boroughs, the neighborhoods and location data (longitude and latitude).
2. Second, we will rely on venue data which is data that will inform us about the type of businesses or attractions available for a given neighborhood.

The combination of these two sets of data will allow us to perform the analysis necessary to determine the recommendation at the end of the project.

II. Data



NYC



Toronto

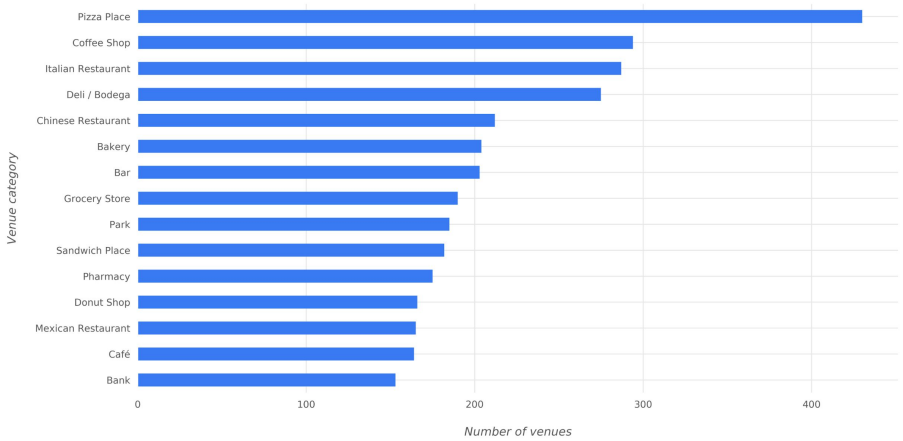
II. Data

Example of the dataframe organizing the geographical and venue data for the neighborhoods of the city of Toronto.

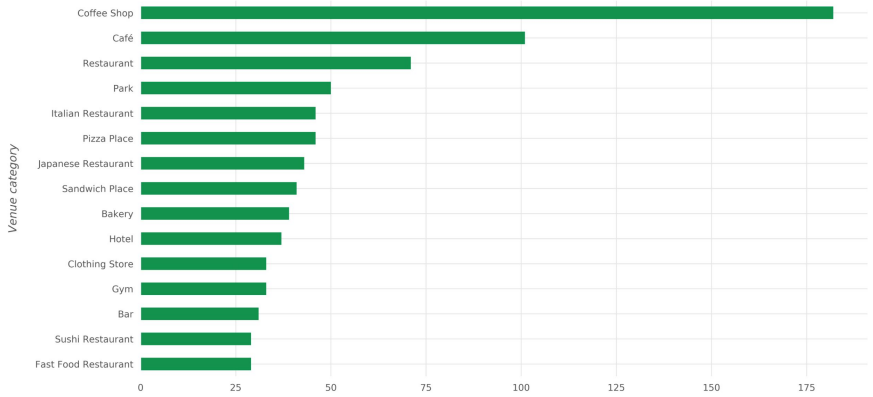
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

III. Methodology

Most common venues for both cities



NYC



Toronto

III. Methodology: Clustering

Using one-hot encoding the data was merged and presented in a dataframe similar to the one presented below

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Animal Shelter	Antique Shop	Aquarium	Arcade
0	Allerton_NYC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
1	Annadale_NYC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.125000	0.0	0.0	0.0	0.0
2	Arden Heights_NYC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
3	Arlington_NYC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.333333	0.0	0.0	0.0	0.0
4	Arrochar_NYC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0

III. Methodology: K-means method

Below is the code used to implement the K-means method

```
# the number of clusters
kclusters = 5

nyc_tor_grouped_clustering = nyc_tor_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(nyc_tor_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

IV. Results

Table 1. Cluster 1

Category	% of venues
Pizza places	7.2
Pharmacy	4.1
Chinese restaurant	3.6
Bank	3.6
Donut shop	3.4
Sandwich place	3.4
Deli/Bodega	3.1

Table 2. Cluster 2

Category	% of venues
Park	38.88
Playground	6.9
Construction/landscaping	4.2
Deli/Bodega	4.2
Bank	2.7
Pool	2.7
Coffee shop	2.7

Table 3. Cluster 3

Category	% of venues
Pool	37.5
Basketball field	37.5
Locksmith	12.5
Convenience store	12.5

IV. Results

Table 4. Cluster 4

Category	% of venues
Business service	100

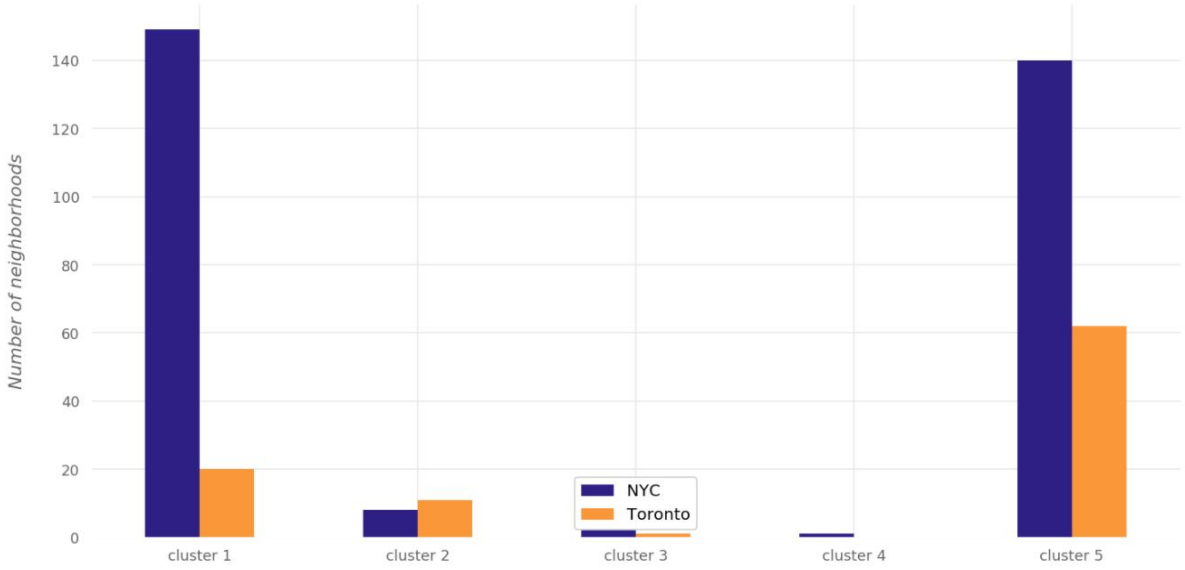
Table 5. Cluster 5

Category	% of venues
Coffee shop	5.1
Italian restaurant	3.4
Pizza place	2.8
Cafe	2.7
Bar	2.4
Bakery	2.2
Deli/Bodega	2.1

Each cluster in the tables presented has a different combination of venues. For example cluster 1 has a diverse set of businesses that range from Pizza places, Coffee shops and bodegas and sandwich places. It seems that the neighborhoods in this cluster have a commercial dynamism and plenty of places where people go to find a place to eat. Cluster 1 is very similar to cluster 5 where there are also many eateries and places to socialize. This contrasts with cluster 3 with most of the neighborhoods having parks (38%) and playgrounds (7%). These neighborhoods might be more residential and less commercial. Cluster 4 also seems to be less commercial and can possibly be a cluster of residential neighborhoods.

IV. Results

Figure 25 represents the number of neighborhoods in each cluster for both cities. Note that most neighborhoods in both cities fall in cluster 1 and 5.



V. Discussion

Based on the analysis presented in this report, we can issue the following recommendations:

1. If we are interested in opening a business, neighborhoods in cluster 1 and cluster 5 can be good candidates for a location of opening a business such as restaurant, cafe, coffee shop or bar. There is a lot of food traffic.
2. The neighborhoods that have amenities like in cluster 1 and cluster 5 are quite spread out and geographically diverse. Hence, it is possible to find more affordable neighborhoods within these clusters to rent a venue for a business
3. If the business of interest is a restaurant, it is recommended to avoid a pizza place, italian restaurant, american restaurant, mexican restaurant or chinese restaurant. There are plenty of these restaurants in both cities and it will be very difficult to compete with very established restaurants.
4. Thorough analysis of real estate value is needed in order to recommend one neighborhood versus the other.

VI. Conclusion

In this report, we present our work regarding clustering the neighborhoods from the city of New York, NY and the city of Toronto, ON. The clusters constitue groups that have a similar combination of businesses or venues. The project resulted in a list of neighborhoods where it might be appealing to open a certain business like a restaurant of a less common cuisine or even a new cuisine. Additional analysis of real estate data for both cities and ease of business indices, in addition to client specifications, can help issue a strong recommendation for one city versus the other.