Battle of Neighborhoods – New York City vs Toronto

Marco Custódio

December 8th, 2020

1. Introduction

New York City and Toronto are two important financial capitals of their countries, the USA and Canada respectively, and are very diverse in terms of their inhabitants and their venues. Throughout the years, the neighborhood from each city have evolved and transformed, offering a variety of places for entertainment, social gathering, nourishment, relaxation, etc. These unique characteristics of each neighborhood is also taken into consideration by people when they decide where to live. As a many number of people also move between both cities, sometimes temporarily, sometimes definitely, it would be advantageous if they could have an idea of the similarities between neighborhoods from one city to other so that they could us that information when choosing where to move. Some people will prefer to move into a new neighborhood in the new city with similar characteristics, in terms of venues, to the city previously inhabited. Some people might be interested in finding something rather different. We hope this analysis can help them make a more informed decision.

2. Data

Two datasets were collected from web:

1. New York City data was obtained from a JSON file available at https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json' with its Boroughs, Neighborhoods, and their geographical location (latitude and longitude). The dataset was restricted to Manhattan neighborhoods for simplicity and a total of 40 neighborhoods populated the final dataset.

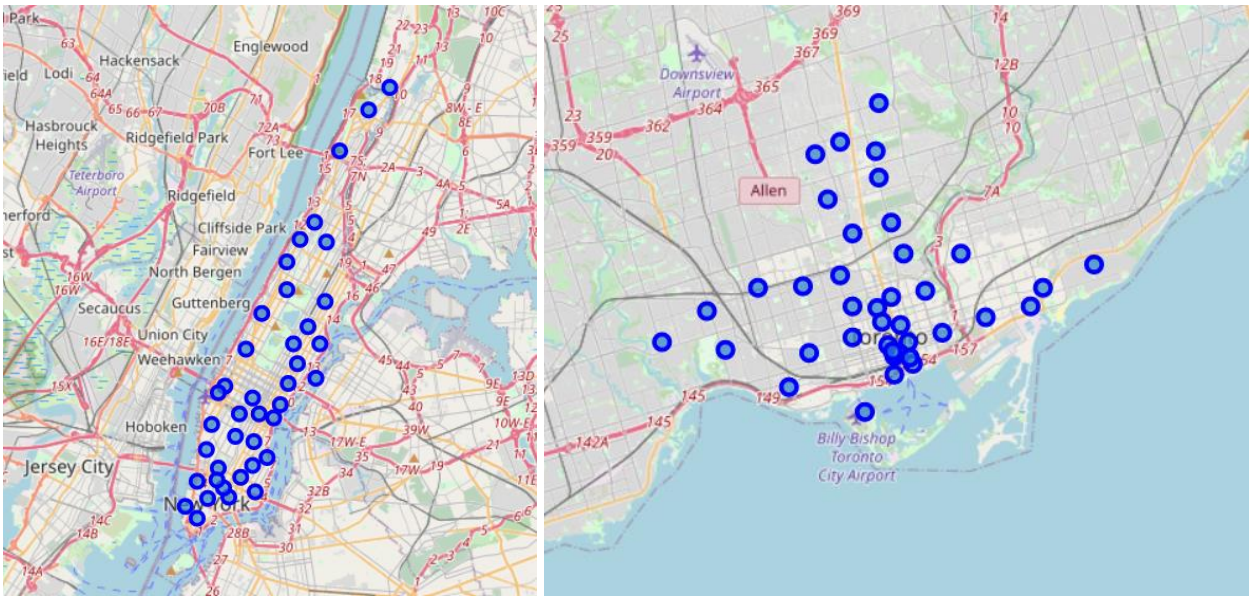
2. Toronto data was scrapped from a Wikipedia table available at https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M using the BeautifulSoup package. With the Postal Code data, we were able to merge a geographical dataset with latitude and longitude information available at https://cocl.us/Geospatial_data. The dataset was restricted to Boroughs with Toronto in the description (more central neighborhoods) and a total of 39 neighborhoods

Both datasets were than merged and here is a print screen of the final dataset of NY and Toronto neighborhoods:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688
...	...	...	...	...
34	Downtown Toronto	Stn A PO Boxes	43.646435	-79.374846
35	Downtown Toronto	St. James Town, Cabbagetown	43.667967	-79.367675
36	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280
37	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
38	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558

79 rows × 4 columns

Using Folium package, we can also visualize the geographical locations of the neighborhoods. The image on the left shows Manhattan and the second image displays the Toronto neighborhoods to be analyzed in this study.



With the geographical data of each neighborhood, we will use Foursquare API to explore the venues in each neighborhood and then cluster them to check how similar they are and make recommendations to anyone moving between both cities.

3. Methodology

As mentioned above, the Foursquare API will be used to explore the venues in each of the 79 neighborhoods in our dataset, belonging to the cities of New York and Toronto. The top 100 venues in each neighborhood within a radius of 500 meters of their geographical reference were retrieved from the API and added to a new data frame.

This new data frame has 7 columns and 4826 rows, each corresponding to a venue. Se the first 5 rows of the dataframe below, which includes the venue names, their geographical location, and their category.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

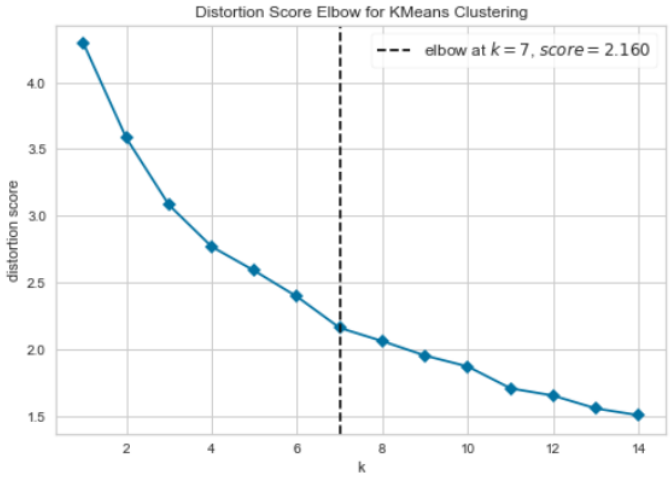
There was a total of 369 unique categories returned by Foursquare.

Afterwards, a 'one-hot encoding' was performed to get dummy variables of 'Venue Category' in each neighborhood and then grouped the rows by neighborhood and calculated the mean of the frequency of occurrence of each category. This allowed the creation of a table listing the top 5 venue categories in each neighborhood, as seen in the image below.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Park	Hotel	Coffee Shop	Gym	Memorial Site
1	Berczy Park	Coffee Shop	Seafood Restaurant	Cheese Shop	Cocktail Bar	Beer Bar
2	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Coffee Shop	Nightclub	Climbing Gym
3	Business reply mail Processing Centre, South C...	Fast Food Restaurant	Auto Workshop	Gym / Fitness Center	Restaurant	Skate Park
4	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Airport Terminal	Harbor / Marina	Sculpture Garden
...	...	...	...	...	...	...
74	Upper East Side	Exhibit	Italian Restaurant	Coffee Shop	Bakery	Gym / Fitness Center
75	Upper West Side	Italian Restaurant	Bar	Café	Coffee Shop	Indian Restaurant
76	Washington Heights	Café	Bakery	Grocery Store	Coffee Shop	Donut Shop
77	West Village	Italian Restaurant	New American Restaurant	American Restaurant	Park	Cocktail Bar
78	Yorkville	Italian Restaurant	Gym	Bar	Coffee Shop	Sushi Restaurant

79 rows × 6 columns

There are several common venue categories in neighborhoods. I used unsupervised machine learning algorithm known as **K-means clustering** to cluster them regarding their venue composition. The elbow method (also known as WCSS - within-cluster sums of squares) helped me pick the number of clusters, k.



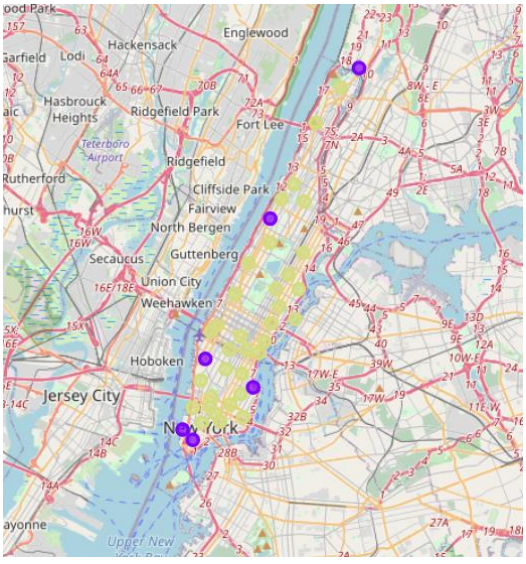
From the elbow method we can see that it is not clear what is the best k, but I made a practical choice based on several runs with different k's and decided to use a k=7 which seemed the most informative for the purpose of this analysis.

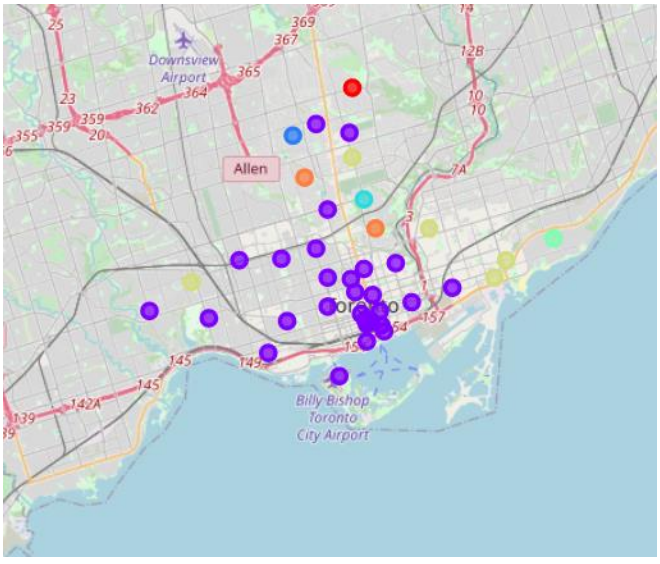
Here is the merged table with the cluster labels associated with each neighborhood. Only the top 5 rows are visible.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	1	Coffee Shop	Sandwich Place	Discount Store	Gym	Yoga Studio
1	Manhattan	Chinatown	40.715618	-73.994279	5	Chinese Restaurant	Bakery	Cocktail Bar	American Restaurant	Dessert Shop
2	Manhattan	Washington Heights	40.851903	-73.936900	5	Café	Bakery	Grocery Store	Coffee Shop	Donut Shop
3	Manhattan	Inwood	40.867684	-73.921210	5	Mexican Restaurant	Lounge	Restaurant	Café	Deli / Bodega
4	Manhattan	Hamilton Heights	40.823604	-73.949688	5	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Indian Restaurant

And below you can see again the map of each city (New York on the left and Toronto on the right), now with each neighborhood colored, each color representing a cluster as follows:

Cluster 0, Cluster 1, Cluster 2, Cluster 3, Cluster 4, Cluster 5, and Cluster 6.





4. Results

From the maps created above, it is possible to see that only 2 clusters overlap with the two cities. These are Clusters 1 and Cluster 5. The other clusters only comprehend neighborhoods in Toronto. Moreover, these Toronto-only clusters are only composed of one (4 clusters) or two (1 cluster) and are mostly at the peripheries of central Toronto.

Let us have a closer look into Clusters 1 and 5, since these are the ones that have neighborhoods from both cities. Cluster 1 has mostly Toronto neighborhoods (28) and a few from Manhattan (6) and Cluster 5 has mostly neighborhoods from Manhattan (34) and a few from Toronto (5). We can further look into the cluster, for example by analyzing which venues are the 1st most common in those neighborhoods.

Cluster 1 - 1st most common venues:

```
: Coffee Shop                    20
  Park                            3
  Café                            3
  Gift Shop                       1
  Clothing Store                  1
  Grocery Store                   1
  Vegetarian / Vegan Restaurant   1
  Bar                             1
  Hotel                           1
  Airport Lounge                  1
  Bakery                          1
```

Cluster 5 - 1st most common venues:

```
: Italian Restaurant      8
  Coffee Shop             4
  Café                    3
  Bar                     3
  Chinese Restaurant      2
  Park                    2
  Mexican Restaurant      2
  Hotel                   2
  African Restaurant      1
  Theater                 1
  Clothing Store          1
  Dessert Shop            1
  Exhibit                 1
  Japanese Restaurant     1
  Greek Restaurant        1
  Fast Food Restaurant    1
  Plaza                   1
  Sushi Restaurant        1
  Deli / Bodega           1
  Pizza Place             1
  Korean Restaurant       1
```

We see that in Cluster 1, the most common venues are by far Coffee Shops, which are 1st in 20 neighborhoods, followed by Parks (3 neighborhoods) and Cafés (3 neighborhoods). In Cluster 5, the most common venues are Italian Restaurants, which are 1st in 8 neighborhoods, followed by Coffee Shops (4 neighborhoods), Cafés (3) and Bars (3). Moreover, Cluster 5 has much more diversity of venues that come in 1st place than Cluster 1.

5. Discussion

Most people living in Manhattan that are looking to move into central Toronto, would feel mostly ‘at home´ by choosing most peripheric neighborhoods rather than most central ones. The Toronto neighborhoods most similar to Manhattan ones are: The Danforth West & Riverdale, India Bazaar & The Beaches West, High Park & The Junction South, Davisville and Business reply mail Processing Centre & South Central Letter Processing Plant in Toronto.

On the other hand, most people from Toronto, especially those living in more central neighborhoods, would find the following Manhattan neighborhoods more familiar to their former town: Marble Hill, Chelsea, Morningside Heights, Battery Park City, Financial District and Stuyvesant Town.

Toronto also has a few neighborhoods at its peripheries that are rather unique, since they clustered alone, and could be interesting for those wanting to move to a different neighborhood, whether they are coming from Manhattan or from central Toronto. Some of the most common venues in those different neighborhoods are Parks, Trails, Yoga Studios and Escape Rooms, which would be more appealing to particular types of people.

6. Conclusion

New York City and Toronto are two amazing cities, with a huge offer of venues and neighborhoods with unique characteristics. By exploring the venues available in each of their neighborhoods it became clear that both cities are more similar within themselves than between themselves, which was expected, but there is some overlap. People moving between those places can choose if they want to stay in a neighborhood with a more familiar feeling or to one rather different or not similar at all. By using Data Science tools and harnessing the capabilities of the Foursquare API it is possible for anyone (with some patience and will to learn) to make comparisons between different places, explore what they have to offer and make informed decisions before going on holidays or moving into a new city.