The Battle of Neighborhoods: New York vs Toronto

Hanning Gu

1 January 2020

Introduction and Problem

As an international centre of finance, culture, arts and business and the most populous city of their respective countries, both New York and Toronto are recognized as one of the most cosmopolitan and multicultural cities in the world. Therefore, they have attracted tons of tourists and migrants every year. Despite their similarities, their differences have also been discussed. Which city would you consider as a **migrant**? Where city is a better choice to **open a restaurant** of your own cuisine? With the above questions in mind, I will conduct a comparison study of New York (Manhattan) and Toronto city using the techniques I learnt in the IBM data science course on Coursera. This report can also help **travellers** who are interested in visiting these two cities.

Data acquisition and pre-processing

Information of neighbourhoods of New York and Toronto were directly obtained from https://geo.nyu.edu/catalog/nyu_2451_34572 and https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. Then, data of venues of interests of these neighbourhood areas were scraped from Foursquare location data. The venue data not only contains information such as location (i.e. latitude, longitude) but also the category of the venue, which will provide the most important information on differentiating neighbourhoods.

Firstly, neighbourhood data of New York city was loaded to a table with information such as Borough name, Neighbourhood name and its location (Latitude and Longitude). For the sake of simplicity, I selected suburbs in Manhattan for further investigation. Next, for each neighbourhood in

Manhattan, I scraped data of up to 100 venues within a disk centred at its given coordinate with radius=500 from Foursquare. The data collected also tells which category every venue belongs to. Then I grouped all venue entries by their neighbourhoods to have aggregated information of each neighbourhood.

Secondly, neighbourhood data of City of Toronto were processed in a similar way except that I removed cells with missing 'Borough', merged neighbourhoods with same postcode and named neighbourhood as their boroughs if their names are missing. In the end, venues with missing category were removed.

Lastly, the two data frames of two cities were merged into one data frame for machine learning which will be discussed later.

Methodology

After data pre-processing, I have a data frame containing all neighbourhoods of interest and their numbers of different types of venues each of neighbourhood has. After standardising, this data frame is ready for machine learning. Same as the previous practices and assignment, k-means clustering was chosen to partition all the neighbourhoods into different clusters. K-means clustering was selected because of its simplicity and effectiveness to convert standardised numbers of venues into Euclidean distance for neighbourhood clustering. However, in my first attempt 90% of neighbourhoods from the two cities were classified into one single cluster. This partially explains the similarity between the two cities. However, in order to better understand their differences and features of each city, I increased the number of clusters from 5 to 10 to better differentiate the two cities.

Exploratory data analysis

After increasing the number of clusters, I performed an exploratory analysis in which the ten most common venues of every neighbourhood was given.
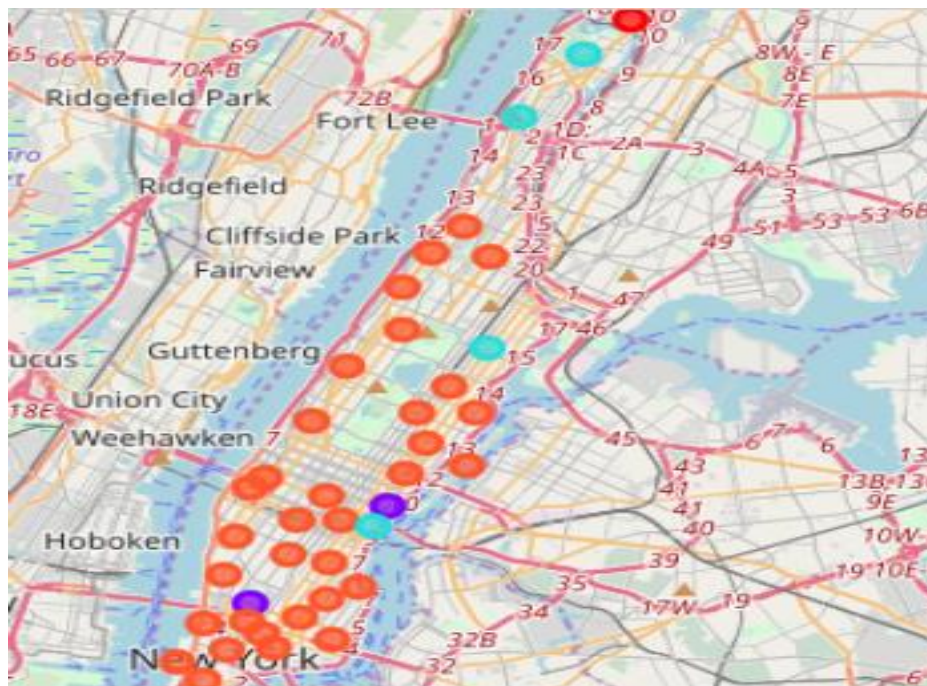
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
37	East Toronto	The Beaches	43.676357	-79.293031	6.0	Pub	Trail	Other Great Outdoors	Health Food Store	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dry Cleaner
41	East Toronto	The Danforth West Riverdale	43.679557	-79.352188	1.0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bookstore	Restaurant	Furniture / Home Store	Brewery	Bakery	Liquor Store
42	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	1.0	Pizza Place	Park	Fish & Chips Shop	Italian Restaurant	Brewery	Movie Theater	Steakhouse	Burger Joint	Burrito Place	Sushi Restaurant
43	East Toronto	Studio District	43.659526	-79.340923	0.0	Café	Coffee Shop	American Restaurant	Gastropub	Bakery	Italian Restaurant	Brewery	Bank	Bar	Latin American Restaurant
44	Central Toronto	Lawrence Park	43.728020	-79.388790	0.0	Photography Studio	Park	Swim School	Bus Line	Women's Store	Doner Restaurant	Donut Shop	Drugstore	Dry Cleaner	Dumpling Restaurant
45	Central Toronto	Davisville North	43.712751	-79.390197	5.0	Park	Hotel	Breakfast Spot	Food & Drink Shop	Sandwich Place	Asian Restaurant	Gym	Clothing Store	Women's Store	Eastern European Restaurant
46	Central Toronto	North Toronto West	43.715383	-79.405678	4.0	Sporting Goods Shop	Coffee Shop	Clothing Store	Spa	Mexican Restaurant	Burger Joint	Furniture / Home Store	Café	Chinese Restaurant	Salon / Barbershop
47	Central Toronto	Davisville	43.704324	-79.388790	1.0	Dessert Shop	Sandwich Place	Café	Gym	Italian Restaurant	Sushi Restaurant	Coffee Shop	Pizza Place	Diner	Fried Chicken Joint
48	Central Toronto	Moore Park,Summerhill East	43.689574	-79.383160	3.0	Park	Tennis Court	Playground	Trail	Women's Store	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore
49	Central Toronto	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	43.686412	-79.400049	0.0	Pub	Coffee Shop	Sports Bar	Fried Chicken Joint	Restaurant	Sushi Restaurant	Bagel Shop	American Restaurant	Liquor Store	Supermarket

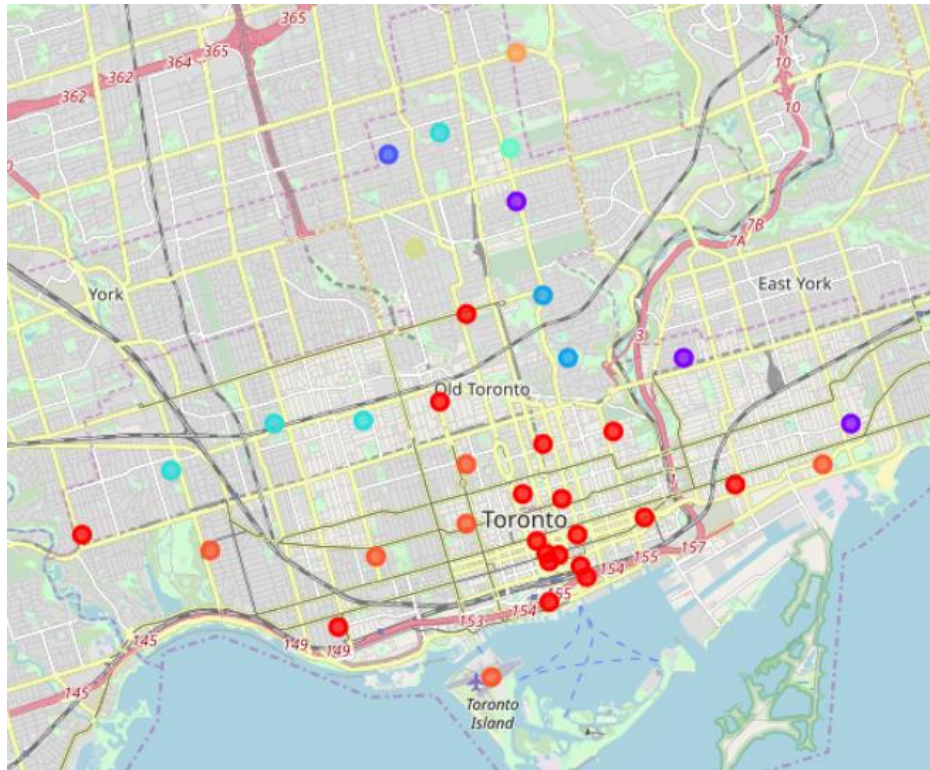|	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	0.0	Coffee Shop	Sandwich Place	Yoga Studio	Gym	Big Box Store	Steakhouse	Supplement Shop	Shopping Mall	Donut Shop	Seafood Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	9.0	Chinese Restaurant	Cocktail Bar	American Restaurant	Spa	Hotpot Restaurant	Vietnamese Restaurant	Bakery	Salon / Barbershop	Optical Shop	Furniture / Home Store
2	Manhattan	Washington Heights	40.851903	-73.936900	4.0	Café	Bakery	Grocery Store	Chinese Restaurant	Mobile Phone Shop	Spanish Restaurant	Pizza Place	Coffee Shop	Park	New American Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	4.0	Café	Mexican Restaurant	Bakery	Restaurant	Pizza Place	Lounge	American Restaurant	Chinese Restaurant	Pharmacy	Deli / Bodega
4	Manhattan	Hamilton Heights	40.823604	-73.949688	9.0	Pizza Place	Café	Deli / Bodega	Coffee Shop	Mexican Restaurant	Yoga Studio	Caribbean Restaurant	Bakery	Indian Restaurant	School
5	Manhattan	Manhattanville	40.816934	-73.957385	9.0	Coffee Shop	Mexican Restaurant	Italian Restaurant	Seafood Restaurant	Park	Deli / Bodega	Supermarket	BBQ Joint	Climbing Gym	Bakery
6	Manhattan	Central Harlem	40.815976	-73.943211	9.0	Chinese Restaurant	Cosmetics Shop	Seafood Restaurant	Bar	French Restaurant	American Restaurant	African Restaurant	Café	Fried Chicken Joint	Cocktail Bar
7	Manhattan	East Harlem	40.792249	-73.944182	4.0	Mexican Restaurant	Bakery	Thai Restaurant	Deli / Bodega	Latin American Restaurant	Historic Site	Park	French Restaurant	Taco Place	Beer Bar
8	Manhattan	Upper East Side	40.775639	-73.960508	9.0	Italian Restaurant	Exhibit	Coffee Shop	Bakery	Art Gallery	Gym / Fitness Center	Juice Bar	Hotel	French Restaurant	Sushi Restaurant
9	Manhattan	Yorkville	40.775930	-73.947118	9.0	Italian Restaurant	Bar	Gym	Coffee Shop	Pizza Place	Deli / Bodega	Bakery	Japanese Restaurant	Diner	Park

In the first table above, we can see the 10 selected neighborhoods in Toronto have a more diverse set of clusters than Manhantan. Furthermore, one can tell that Torontonians like indoor and outdoor activities more than New Yorkers as per the second table. In addition, in Manhantan restaurant is the most popular venue in most neighborhoods. An interesting oberservation is that coffee shops and cafes are popular in both cities.

The above exploratory data analysis confirms the existences of similarities and dissimilarities between New York and Toronto. In the next section I will show and discuss the K-means clustering results.

Results

I aggregated venue data frames of the two cities and performed K-means clustering, then plotted the clustering of all the neighbourhoods on separate maps as per below. The first figure shows the clusters of Manhattan and the second figure shows the results of City of Toronto. Each represents a cluster in both figures.

One can notice that the orange cluster (cluster 9) is the the most dominant cluster in Manhattan and it is also an important cluster in Toronto. Different from Manhattan, Toronto is more diverse and has more significant clusters than Manhattan. Other than the aforementioned orange cluster (cluster 9), the red (cluster 0), green and purple clusters are also an important part of the Toronto community witrh the red cluster being the largest cluster in terms of neighborhood quantity. Now let's the features of these clusters.

Cluster 9:

Restaurants are the most common and popular venues of neighbourhoods in cluster 9. New Yorkers and Torontonians all like food.

Cluster 0:

Being the largest cluster in the City of Toronto, cluster 0 has coffee shop and café as the most important venue followed by Bakery and Sandwich Place.

Other important clusters (1,3,4) in Toronto:

As expected, there are other crucial clusters that contribute to the multiculture of Toronto. For example, in cluster 1 there are many Greek and Italian restaurants while Mexico and Asian restaurants are also common in cluster 4. Another interesting oberservation is that cluster 3 has lots of places for indoor and outdoor activities.

Discussions and Recommendations

Toronto and New York have lots of things in common but are also quite different in some ways. For someone who is looking to start a business or considering travelling to either of the cities, it

will ultimately come down to what you want to do or what you like. If you want to open a restaurant or want to try a variety of cuisines then New York could be a better choice for you. If your business plan is to open a coffe shop or you want to have a relaxing morning sitting in a coffee shop reading your favourite then Toronto would be the right choice for you. If you love work-out then Toronto is also a nice choice. In summary, it is a life-style choice.

Conclusion:

In this report, I presented a comparison study of Manhattan and City of Toronto using data collected from NYU Spatial Data Repository, Wikipedia and Foursquare location data. K-means clustering was used to divide neighbourhoods of the two cities into different groups. Analysis of the similarities and dissimilarities of the two cities were conducted. In the end, business plan and travelling recommendations were given.