How similar is Toronto and New York City?

Introduction

If we plan to visit or invest in some places in North America, New York City and Toronto are two good choices. Toronto is the largest city with the population of 6,471,850 in Canada while New York City is the largest city in the United States with the population of 8,836,817 as recorded in 2019. First we need to explore by visualizing the neighborhoods of each city and then check the neighborhoods that meet our criterias for example having hotels and restaurants nearby.

Data Description

Since each city has a lot of neighborhoods and In order to compare two cities, we only select the borough containing Toronto and borough of Manhattan to compare. The essential data we need is the borough, latitude, longitude and the venues' categories for each neighborhood of the two cities which can be obtained from the website and Foursquare API. In this project, we are interested in the neighborhoods that have hotels, Chinese restaurants and parks.

Methodology

I used the GitHub repository in this project. I downloaded the borough, latitude and longitude of each neighborhood for New York City from a new_york_dataset and then transformed it into a pandas dataframe. I sliced the data to get the Manhattan borough which has 40 neighborhoods. The first five rows of the dataframe is displayed as follows.

	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

Fig.1 The first five row of the dataframe of Manhattan

I scraped the data for boroughs that contain Toronto from a Wikipedia page and transformed it into a pandas dataframe using the BeautifulSoup package. The

dataframe has 7 boroughs and 39 neighborhoods. The first five rows of the dataframe is displayed as follows.

	Borough	Neighborhood	Latitude	Longitude
0	Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636
1	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
2	Downtown Toronto	St. James Town	43.651494	-79.375418
3	East Toronto	The Beaches	43.676357	-79.293031
4	Downtown Toronto	Berczy Park	43.644771	-79.373306

Fig.2 The first five row of the dataframe of Toronto

In order to visualize geographic details and boroughs of Manhattan and Toronto I created the map of both cities with neighborhoods superimposed on top using python folium library.

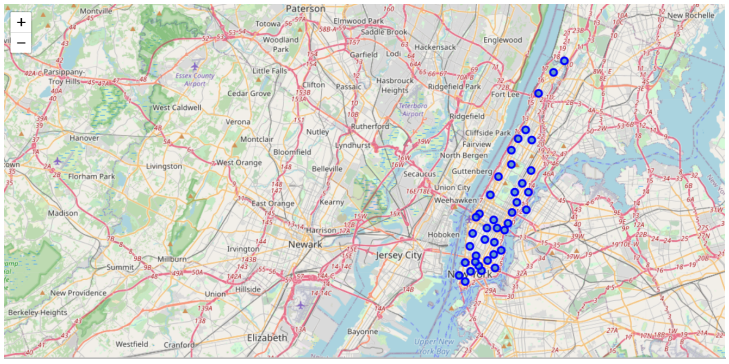

Fig 3. The map of Manhattan with neighborhoods superimposed on top

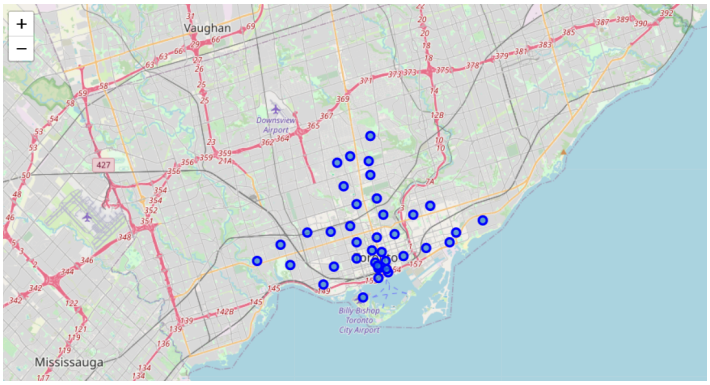

Fig4. The map of Toronto with neighborhoods superimposed on top

I utilized the Foursquare API to explore the venues of each neighborhood. I set the limit as 100 venues and the radius 500 meter for each neighborhood from their given latitude and longitude information. Manhattan has 329 unique venue categories and Toronto has 234 unique venue categories. After I put the venues into a pandas dataframe, I run k-means to cluster the neighborhood into 5 clusters.

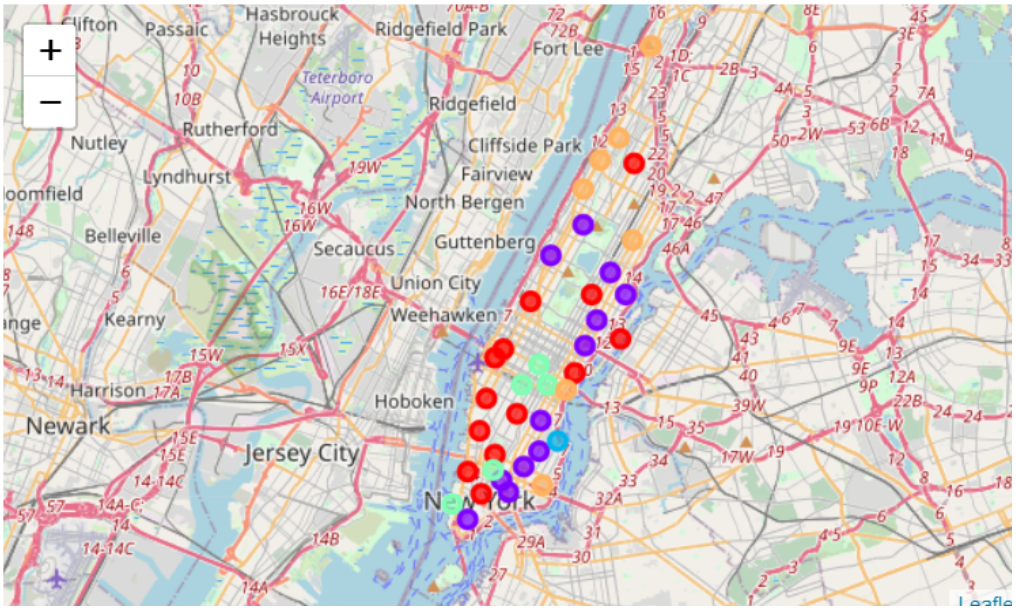

Fig.5 K-means with 5 clusters map for Manhattan

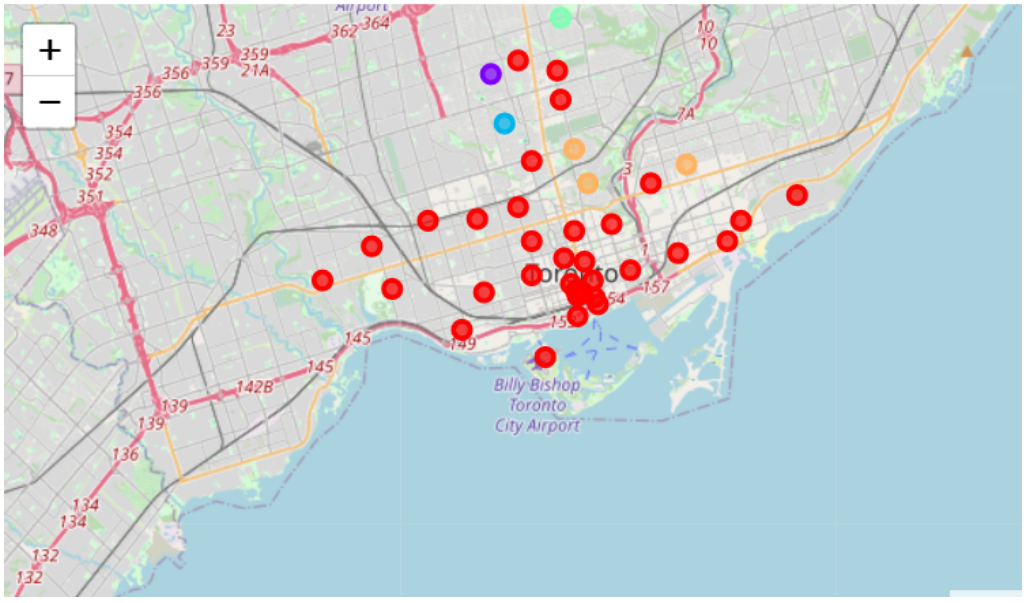

Fig.6 K-means with 5 clusters map for Toronto

I create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

	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	4	Gym	Sandwich Place	Yoga Studio	Pizza Place	Steakhouse	Shopping Mall	Seafood Restaurant	Clothing Store	Coffee Shop	Deli / Bodega
1	Manhattan	Chinatown	40.715618	-73.994279	1	Bakery	Chinese Restaurant	Cocktail Bar	American Restaurant	Hotpot Restaurant	Ice Cream Shop	Dessert Shop	Spa	Salon / Barbershop	Asian Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Deli / Bodega	Bank	Mobile Phone Shop	Grocery Store	Spanish Restaurant	Pizza Place	Tapas Restaurant	Sandwich Place
3	Manhattan	Inwood	40.867684	-73.921210	4	Mexican Restaurant	Café	Restaurant	Chinese Restaurant	Lounge	Spanish Restaurant	Bakery	Caribbean Restaurant	Pizza Place	Park
4	Manhattan	Hamilton Heights	40.823604	-73.949688	4	Pizza Place	Coffee Shop	Café	Deli / Bodega	Mexican Restaurant	Bakery	Park	Cocktail Bar	Sandwich Place	Chinese Restaurant

Fig7. Dataframe that includes the cluster as well as top 10 venues for neighborhoods in Manhattan

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
Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636	0	Coffee Shop	Bakery	Café	Pub	Park	Breakfast Spot	Theater	Yoga Studio	Dessert Shop	Shoe Store
Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Coffee Shop	Clothing Store	Café	Bubble Tea Shop	Middle Eastern Restaurant	Japanese Restaurant	Cosmetics Shop	Hotel	Pizza Place	Burger Joint
Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Café	Beer Bar	Gastropub	Cosmetics Shop	Cocktail Bar	Gym	Farmers Market	Hotel	Italian Restaurant
East Toronto	The Beaches	43.676357	-79.293031	0	Asian Restaurant	Health Food Store	Trail	Pub	Wings Joint	Donut Shop	Distribution Center	Dog Run	Doner Restaurant	Eastern European Restaurant
Downtown Toronto	Berczy Park	43.644771	-79.373306	0	Coffee Shop	Cocktail Bar	Farmers Market	Seafood Restaurant	Bakery	Restaurant	Pharmacy	Cheese Shop	Beer Bar	Japanese Restaurant

Fig8. Dataframe that includes the cluster as well as top 10 venues for neighborhoods in Toronto

Since we are mostly interested in the neighborhoods with Chinese Restaurants, parks and hotels, we selected the neighborhoods containing all three venues. There are 8 out of 40 neighorboods in Manhattan meet the criteria.

	Neighborhood	Chinese Restaurant	Hotel	Park
0	Battery Park City	0.012658	0.050633	0.101266
3	Chelsea	0.010000	0.020000	0.020000
6	Clinton	0.010000	0.030000	0.020000
10	Flatiron	0.010000	0.010000	0.020000
12	Greenwich Village	0.020000	0.010000	0.010000
17	Lincoln Square	0.010526	0.021053	0.010526
23	Midtown	0.010000	0.100000	0.010000
31	Sutton Place	0.020000	0.020000	0.030000

Fig9. The Neighborhoods in Manhattan meet the criteria

There are 2 out of 39 Neighborhoods meet the criteria in Toronto.

	Neighborhood	Chinese Restaurant	Hotel	Park
14	Garden District, Ryerson	0.01	0.02	0.01
15	Harbourfront East , Union Station , Toronto Is...	0.01	0.04	0.02

Fig10. The neighborhoods in Toronto meet the criteria

Results

From the above analysis, we can visualize each neighborhood on the map from Fig3 and Fig4. After clustering the neighborhoods into five clusters we can see that it has less unique venue categories than New York City. And from Fig5 and Fig6, we can see that 32 Toronto's neighborhoods are in one cluster and the neighborhoods in Manhattan are more evenly distributed. From Fig9 and Fig10 we can see that Manhattan has more neighborhoods that meet our criterias.

Discussion

Based on the neighborhood and venue analysis we would say that although New York City and Toronto have a similar population, New York City is more diverse, convenient and has more choices for tourism and investments. However, we did not compare other facts such as price, security, environment and traffic. In the future work, we can take more facts into consideration and create an user interface for the users to view the comparison by selecting the facts they are more interested in.

Conclusion

In this project, we obtained the neighborhood data for New York City and Toronto and transformed them into pandas dataframes. We used the folium library to visualize the neighborhood information on the map. Then we obtained venue categories from Foursquare API and filtered the neighborhoods that meet our criterias. Although the analysis suggests New York City is more diverse, other facts need to take into consideration in the future work.