Comparison of Neighborhoods

Mohammad Iliyas

14 July 2020

1. Introduction

1.1 Background

Suppose we are shifting from one city to another and want to shift to a neighborhood similar to which we are residing or as tourists we want to compare two cities and visit the city which has popular places according to our tastes. Hence it is important for comparing neighborhoods which have all the available amenities we want. Having the information will help in determining the proper place to visit or reside.

1.2 Business problem

So in this project we will be comparing the places and this will help in selecting the best among them. This will be particularly helpful for tourism industry and people who are shifting due to various reasons. For tourists it will help them to properly plan their visit and for people who are shifting it will help them to decide the neighborhood in which they should shift.

2. Data Section

In this analysis we will be using geospatial data of New York and Toronto city along with their neighborhoods and boroughs. In particular Manhattan from New York and Downtown Toronto from Toronto will be narrowed for comparison of neighborhoods. Next with the help of folium we will visualize the neighborhoods of the city which will provide a better understanding of the places.

Example of the sample data is provided below.

	Borough	Neighborhood	Latitude	Longitude
35	Manhattan	Turtle Bay	40.752042	-73.967708
36	Manhattan	Tudor City	40.746917	-73.971219
37	Manhattan	Stuyvesant Town	40.731000	-73.974052
38	Manhattan	Flatiron	40.739673	-73.990947
39	Manhattan	Hudson Yards	40.756658	-74.000111

Next from the Foursquare API we will be getting the top locations for a particular borough. We will be using clustering by Kmeans to cluster similar venues and folium to visualize the data on maps.

3. METHODOLOGY

3.1 Data acquisition

Toronto city data was obtained from Wikipedia page i.e. 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M' and the New York city data was obtained from already created data frame in the labs section of the course.

3.2 Data cleaning

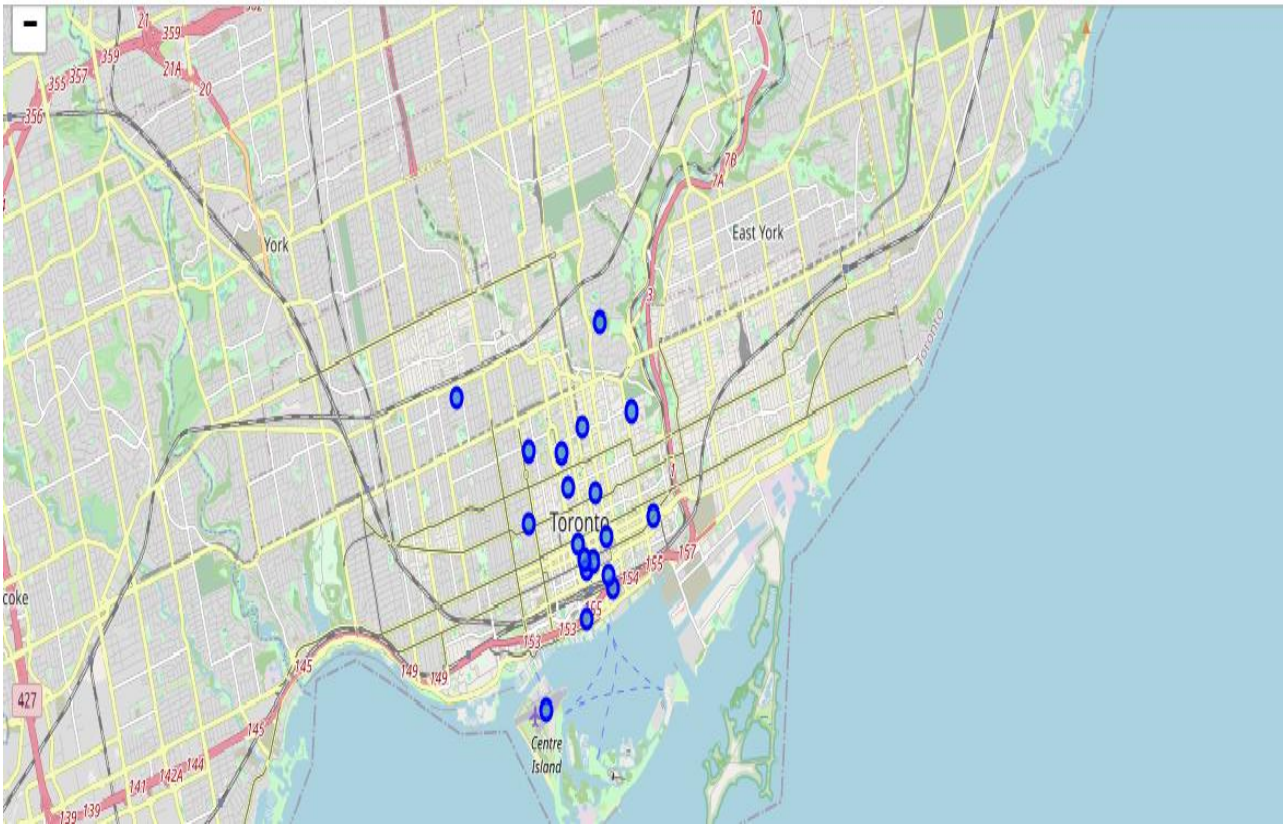
Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values especially in the Toronto city data obtained from the Wikipedia page. The Boroughs which were not available were dropped from the data, next the neighborhoods which had NA values were filled with the corresponding borough data. Some neighborhoods which had the same postal code were combined. After cleaning the data, it was concatenated with a data set of latitude and longitude.

The New York city data was read and required no cleaning.

3.3 VISUALIZING THE CITY

Folium is a powerful data visualization library in Python that was built primarily to help people visualize geospatial data. With Folium, one can create a map of any location in the world as long as its latitude and longitude values are known. Also, the maps created by Folium are interactive in nature, so one can zoom in and out after the map is rendered. Below is a figure which gives the details of the Downtown Toronto Neighborhood.

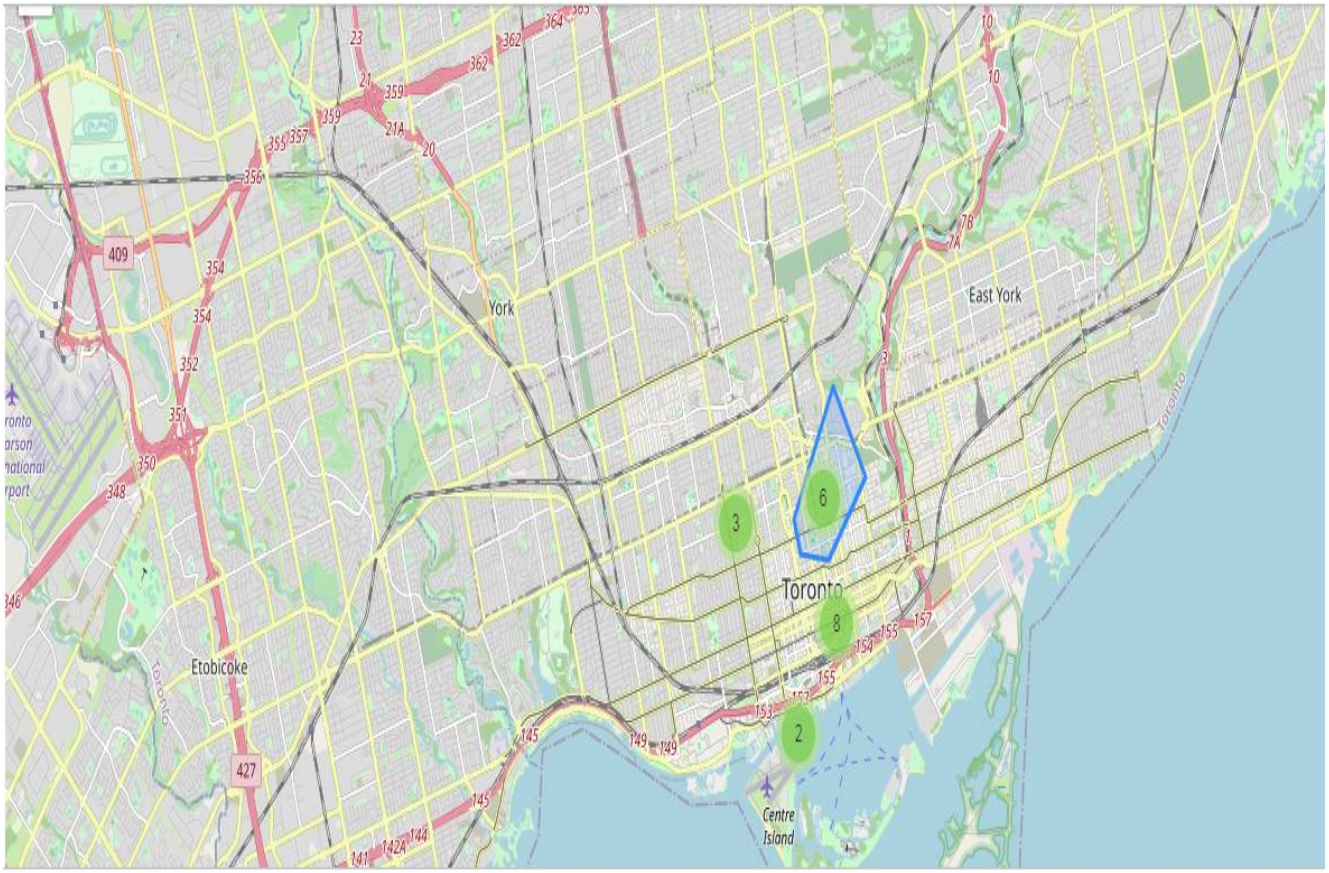
Toronto city Map



The above Toronto city map is created using folium giving inputs such as latitude and longitude of the city. Next markers are added to obtain the neighborhoods of Downtown Toronto. In similar way visualization of New York city Manhattan was done and neighborhoods were obtained as markers on the map.

3.4 Clustering of markers

Below figure shows clustering of neighborhoods of Downtown Toronto.



This is one of the important features of the folium maps. In this the markers which are close are grouped into clusters. This is particularly helpful for logistics or delivery industry as warehouse location can be determined easily. Also this helps us to visualize the concentration of venues or places. The above map shows the neighborhoods of Downtown Toronto clustered at 4 positions.

In similarly we can obtain the clustering of Manhattan neighborhoods.

3.5 Exploring and Analyzing the Neighborhoods

By using Foursquare API, we obtain the trending venues around each of the neighborhood. For further details, we obtain pandas data frame with top 10 venues for each of the neighborhood of Downtown Toronto and Manhattan. Below fig depicts the data obtained

Downtown Toronto

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Berczy Park	Farmers Market	Seafood Restaurant	Bakery	Jazz Club	Museum	Cocktail Bar
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Boat or Ferry	Airport	Airport Food Court
2	Central Bay Street	Coffee Shop	Gastropub	Spa	Park	Modern European Restaurant	Pizza Place
3	Christie	Grocery Store	Café	Park	Athletics & Sports	Candy Store	Nightclub

Manhattan

	Neighborhood	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
0	Battery Park City	Park	Food Court	Memorial Site	Sandwich Place	Cooking School	Burrito Place	Smoke Shop
1	Carnegie Hill	Gym / Fitness Center	Coffee Shop	Italian Restaurant	Gym	Bagel Shop	Shoe Store	French Restaurant
2	Central Harlem	French Restaurant	American Restaurant	Bar	Beer Bar	Café	Food Truck	Music Venue
3	Chelsea	Seafood Restaurant	Hotel	Theater	French Restaurant	Scenic Lookout	Fish Market	Speakeasy
4	Chinatown	Spa	Sandwich Place	Chinese Restaurant	Greek Restaurant	Pizza Place	Dessert Shop	Noodle House

We will be using Kmeans clustering to cluster the top venues. Both Manhattan and Downtown Toronto will be divided into 3 clusters using Kmeans clustering. Now the clusters obtained will be analyzed for similarities and differences.

Cluster 1 - Downtown Toronto

```
downtown_toronto_merged.loc[downtown_toronto_merged['Cluster Labels'] == 0,
    downtown_toronto_merged.columns[[2] + list(range(5, downtown_toronto_merged.shape[1]))]]
```

	Neighborhood	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
0	Rosedale	0	Park	Playground	Trail	Dance Studio	Cheese Shop	Chocolate Shop
1	St. James Town, Cabbagetown	0	Café	Restaurant	General Entertainment	Indian Restaurant	Park	Caribbean Restaurant
2	Church and Wellesley	0	General Entertainment	Creperie	Park	Coffee Shop	Mexican Restaurant	Burger Joint
4	Garden District, Ryerson	0	Café	Clothing Store	Sporting Goods Shop	Music Venue	Burrito Place	Burger Joint
10	Toronto Dominion Centre, Design Exchange	0	Coffee Shop	Café	Gym	Steakhouse	Pizza Place	Pub

Cluster 1- manhattan

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0,
       manhattan_merged.columns[[2] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	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
1	Chinatown	0	Spa	Sandwich Place	Chinese Restaurant	Greek Restaurant	Pizza Place	Dessert Shop
9	Yorkville	0	Deli / Bodega	Italian Restaurant	Coffee Shop	Wine Shop	Bagel Shop	Beer Store
11	Roosevelt Island	0	Food & Drink Shop	Gym	Monument / Landmark	Farmers Market	School	Sandwich Place
13	Lincoln Square	0	Theater	Concert Hall	Performing Arts Venue	Indie Movie Theater	Cycle Studio	Fountain
16	Murray Hill	0	Coffee Shop	Burger Joint	Japanese Restaurant	Hotel	Theater	Café
18	Greenwich Village	0	Italian Restaurant	Cosmetics Shop	Café	Sushi Restaurant	Yoga Studio	Beer Bar

Each of the clusters needs to be analyzed and also can be classified into categories.

4. RESULTS

After clustering the data of the respective neighborhoods, both cities (Boroughs) have venues which can be explored and attract the Tourists. The neighborhoods are much similar in features like Theaters, opera houses, food places, clubs, museums, parks etc. As far as concern to dissimilarity, it differs in terms of some unique places like historical places and monuments.

5. OBSERVATIONS

When we compare the tourist places, we observe that the historical place is only situated in Downtown Toronto and the Monument or landmark venue is in Manhattan neighborhoods. Similarly, Airport facility, Harbor, Sculpture garden and Boat or ferry services are also available in Downtown Toronto while venues like Nightlife, climbing gym and Museums are present in Manhattan.

Another observation found by clustering is that it will help in setting up of warehouses (For vegetables, fruits or regular items for bakery's, restaurants, cafes etc.) at suitable locations.

6. CONCLUSION

The downtown Toronto and Manhattan neighborhoods have more like similar venues. As we know that every place is unique in its own way, so that's argument is present in both neighborhoods. The dissimilarity exists in terms of some different venues and facilities but not on a larger extent.