THE BATTLE OF NEIGHBORHOODS

1. INTRODUCTION

1.1 Background

In today's era, tourism is one of the pillars of the economy and people most often prefer to visit those countries which are rich in heritage and developed enough from a foreign perspective. Every city in this world is unique in its own way and gives something new. Due to the recent developments in technology, the information regarding the location of every place around the world is now available on your fingertips which makes it even easier to explore. Therefore, tourists are always eager to travel to different places on the basis of available information, and the comparison (the part of the information) between the two cities always assists them to choose the specific places according to their choice.

1.2 Problem

The city of Toronto in Canada and that of New York in the United States are one of the most famous places in the world. Not only are they multicultural, but also the financial hubs of their respective countries. So, the aim of this project is to explore the similarities and dissimilarities in these two diverse cities from the perspective of a tourist who wants to visit one of these two cities keeping in mind the areas of food, places to visit, culture, accomodation, etc.

1.3 Target Audience

Tourists who want to travel and explore one of the two selected cities will definitely want to know which city is the best to travel according to their taste. After selecting the place they want to visit based on our exploration, people can analyze what all places can be visited and what all activities can be done in their selected city.

2. DATA SECTION

For this project, the following data will be taken in account.

1) List of Boroughs and their corresponding Neighborhoods in Toronto, Canada which has been taken from the following Wikipedia page.
   https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2) Geographical coordinates of every postal code in Toronto, Canada which have been taken from the following csv file.
   https://cocl.us/Geospatial_data

3) Dataset containing the Boroughs and the Neighborhoods that exist in each borough along with the geographical coordinates of each neighborhood which has been taken from the following json file.
   https://cocl.us/new_york_dataset

Further, the services of the Foursquare API have been used to explore the data of the two cities, namely Toronto and New York, in terms of their neighborhoods.
The data also includes the information about the places that are present in each neighborhood like restaurants, hotels, coffee shops, parks, theaters, art galleries, museums and many more. One Borough is selected from each city to analyze their neighborhoods. For this project, Manhattan from New York and Downtown Toronto from Toronto have been selected. As two boroughs (one from each city) have been selected to explore their neighborhoods, the data exploration, analysis and visualization for both boroughs are done in the same way but separately. The neighborhoods are further characterized as venues and venue categories.

Then, the machine learning technique, "Clustering" is used to segment the neighborhoods with similar objects on the basis of each neighborhood data. These objects are given priority on the basis of foot traffic (activity) in their respective neighborhoods. This will help to locate the tourist's areas and hubs, and then we can judge the similarity or dissimilarity between two cities on that basis.
For the Downtown Toronto case, the table of Toronto's Borough from the Wikipedia page has been extracted. Then the data is arranged according to our requirements. Then, neighborhoods which have the same geographical coordinates at each borough have been combined and sorted against the concerned borough using the csv file. For

data verification and further exploration, Foursquare API is used to get the coordinates of Downtown Toronto and explore its neighborhoods.

For Manhattan, the dataset from the json file to do the same exploration and analysis as done for Downtown Toronto and then Foursquare API is used to do the exploration of its neighborhoods.

3. METHODOLOGY SECTION

The required data to be worked upon is taken from various resources mentioned in the data section and converted into the data frames to perform the exploration, visualization and analysis using machine learning techniques.

The final data frame of the Downtown Toronto Borough selected for the analysis of the city of Toronto looks as follows.

Out[12]:

	Borough	Neighborhood	Latitude	Longitude
0	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	Downtown Toronto	St. James Town	43.651494	-79.375418
4	Downtown Toronto	Berczy Park	43.644771	-79.373306

The final data frame of the Manhattan Borough selected for the analysis of the city of New York has the following form.

Out[18]:

|	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

After getting the much needed data frames of the two boroughs, the geographical coordinates of Downtown Toronto and Manhattan are determined using the **geolocator** package. The coordinates derived are as follows :-

      Downtown Toronto - latitude 43.6563221 & longitude -79.3809161
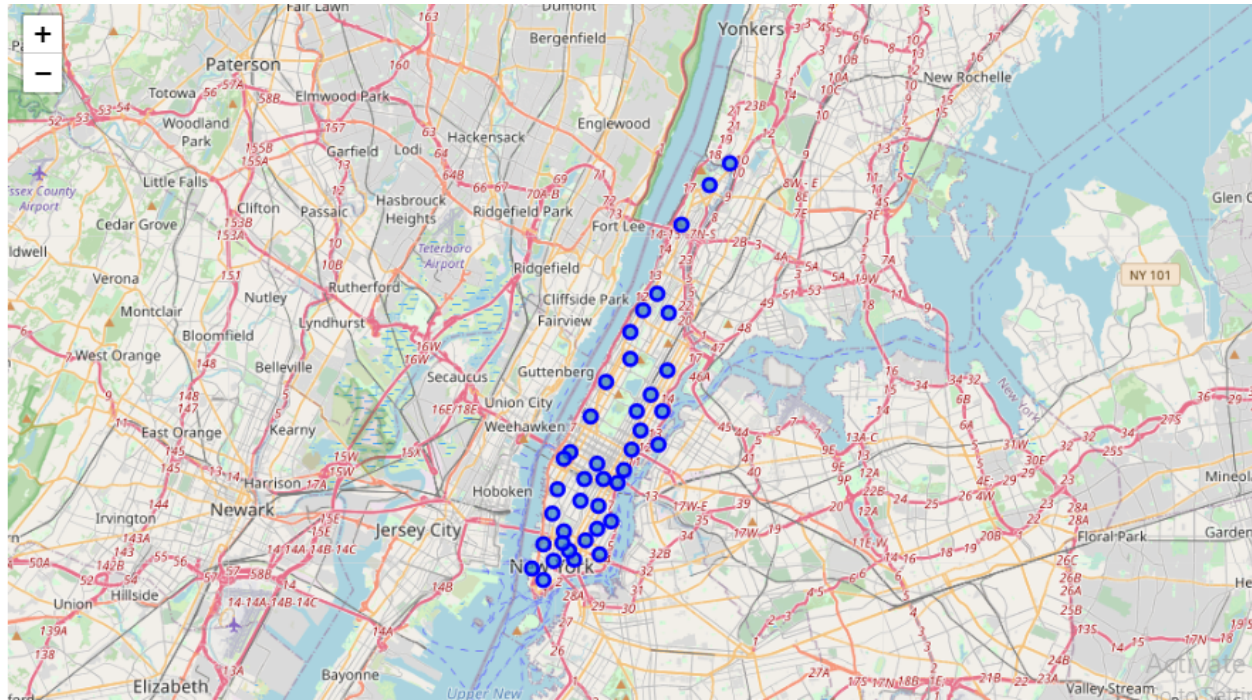   Manhattan - latitude 40.7896239 & longitude -73.9598939

Using the geographical coordinates of both the boroughs, we then visualize the data using the **folium** package before and after completing the step of analysis.

Before clustering, the visualizations of Downtown Toronto and Manhattan have the following forms respectively.
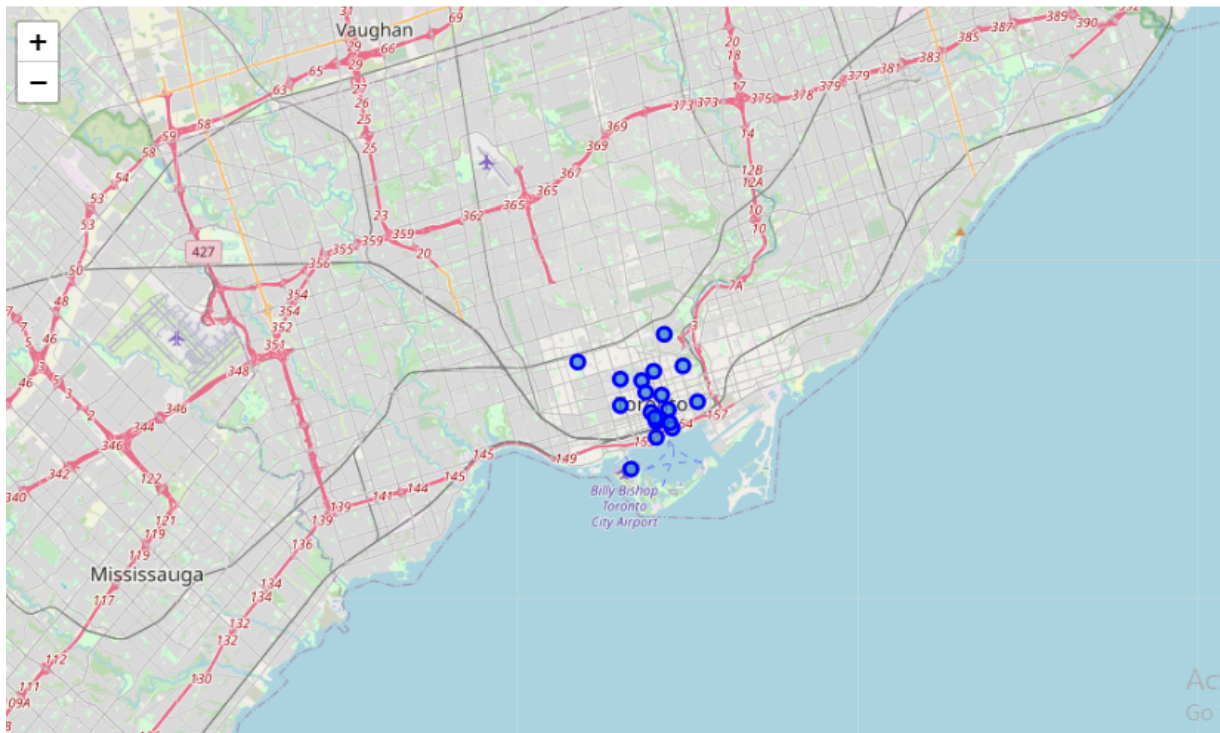
(DOWNTOWN TORONTO)

(MANHATTAN)

Now, we start doing the analysis of the data frames using the one hot encoding in which '1' is assigned if a venue category exists or else '0' is assigned. Once the encoding has been done, the mean of the frequency of occurrence of each category is calculated and top ten venues are picked for each neighborhood on that basis using the data in Foursquare API.

The top 10 venues for each of the neighborhoods in Downtown Toronto borough is as follows.

Out[35]:

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berczy Park	Seafood Restaurant	Cheese Shop	Farmers Market	Beer Bar	Basketball Stadium	Bakery	Restaurant	Concert Hall	Breakfast Spot	Bistro
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Plane	Harbor / Marina	Rental Car Location	Bar	Boat or Ferry	Sculpture Garden	Coffee Shop
2	Central Bay Street	Coffee Shop	Café	Sushi Restaurant	Modern European Restaurant	Bubble Tea Shop	Ramen Restaurant	Middle Eastern Restaurant	Sandwich Place	Bar	Spa
3	Christie	Grocery Store	Café	Park	Italian Restaurant	Candy Store	Coffee Shop	Nightclub	Restaurant	Baby Store	Athletics & Sports
4	Church and Wellesley	Pizza Place	Park	Restaurant	Bookstore	Ramen Restaurant	Breakfast Spot	Bubble Tea Shop	Burger Joint	Mexican Restaurant	Beer Bar

The top 10 venues for each of the neighborhoods in Manhattan borough is of the following form.

Out[50]:

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Memorial Site	Park	Food Court	Plaza	Gym	Smoke Shop	Shopping Mall	Scenic Lookout	Sandwich Place	Monument / Landmark
1	Carnegie Hill	Italian Restaurant	Gym / Fitness Center	Gym	Wine Bar	Dance Studio	Pizza Place	American Restaurant	Spa	Bookstore	Shoe Store
2	Central Harlem	French Restaurant	African Restaurant	Juice Bar	Café	Music Venue	Boutique	Cocktail Bar	Ethiopian Restaurant	Library	Beer Bar
3	Chelsea	Seafood Restaurant	Ice Cream Shop	Theater	Market	Taco Place	Chinese Restaurant	Office	Fish Market	Coffee Shop	Scenic Lookout
4	Chinatown	Sandwich Place	Chinese Restaurant	Indie Movie Theater	Tea Room	Hotpot Restaurant	Ice Cream Shop	Greek Restaurant	Museum	English Restaurant	New American Restaurant

Finally, to compare the neighborhoods of both the boroughs, namely Downtown Toronto and Manhattan, the neighborhoods of the two boroughs are clustered using the machine learning technique, **k-means clustering** using the **KMeans** function available in the **scikit-cluster** package.

In this case, the number of clusters considered are 5 for each of the boroughs that means the neighborhoods of both boroughs are clustered in 5 clusters.

Henceforth, a data frame which consists of name of the borough, name of the neighborhood, geographical coordinates of each neighborhood, cluster labels and the top 10 venues for each of the neighborhood is constructed for the two boroughs selected for the problem, i.e. Downtown Toronto and Manhattan.

The data frame for the Downtown Toronto looks like:-

Out[37]:

|	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 Ve
0	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1	Coffee Shop	Park	Breakfast Spot	Farmers Market	Chocolate Shop	Pub	Restaurant	Perform Arts Ve
1	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	4	Coffee Shop	Sushi Restaurant	Wings Joint	Park	Arts & Crafts Store	Beer Bar	Burrito Place	Creperi
2	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Café	Mexican Restaurant	Shopping Mall	Restaurant	Ramen Restaurant	Coffee Shop	Plaza	Steakh
3	Downtown Toronto	St. James Town	43.651494	-79.375418	3	Gastropub	Coffee Shop	Café	Creperie	Art Gallery	BBQ Joint	Italian Restaurant	Cosme Shop
4	Downtown Toronto	Berczy Park	43.644771	-79.373306	1	Seafood Restaurant	Cheese Shop	Farmers Market	Beer Bar	Basketball Stadium	Bakery	Restaurant	Concer Hall

The data frame for the Manhattan borough is of the form :-

|	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	8 C
0	Manhattan	Marble Hill	40.876551	-73.910660	1	Coffee Shop	Gym	Yoga Studio	Pharmacy	Seafood Restaurant	Steakhouse	Supplement Shop	Sandw Place
1	Manhattan	Chinatown	40.715618	-73.994279	0	Sandwich Place	Chinese Restaurant	Indie Movie Theater	Tea Room	Hotpot Restaurant	Ice Cream Shop	Greek Restaurant	Museu
2	Manhattan	Washington Heights	40.851903	-73.936900	3	Wine Shop	Park	Café	Breakfast Spot	Bakery	Pet Café	Pizza Place	Rame Resta
3	Manhattan	Inwood	40.867684	-73.921210	0	Wine Bar	Café	Park	Yoga Studio	Diner	Spanish Restaurant	Farmers Market	Frozer Shop
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Yoga Studio	Mexican Restaurant	Cocktail Bar	Bakery	Historic Site	Italian Restaurant	Japanese Restaurant	Medite Resta

After the neighborhoods have been clustered, the step of visualization is again repeated to visualize the clusters on the map of Downtown Toronto and Manhattan.

The map of Downtown Toronto along with the clustered neighborhoods is as follows.
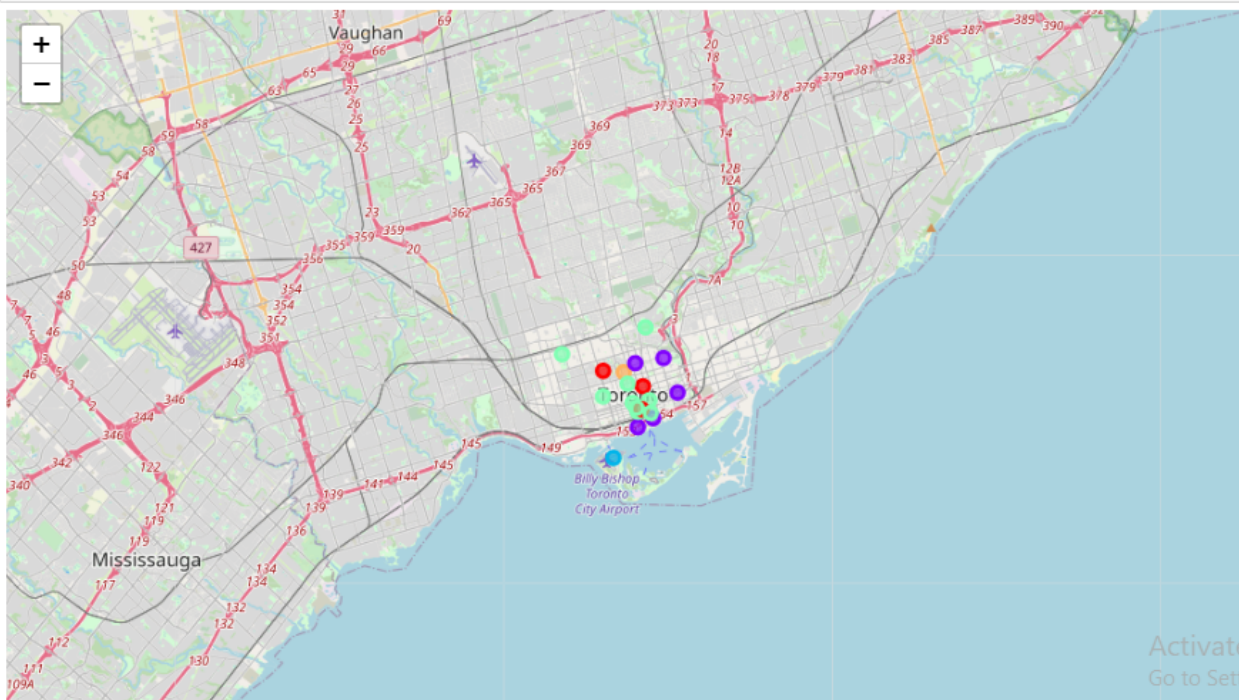
The clustered neighborhoods of Manhattan are represented on the map as.
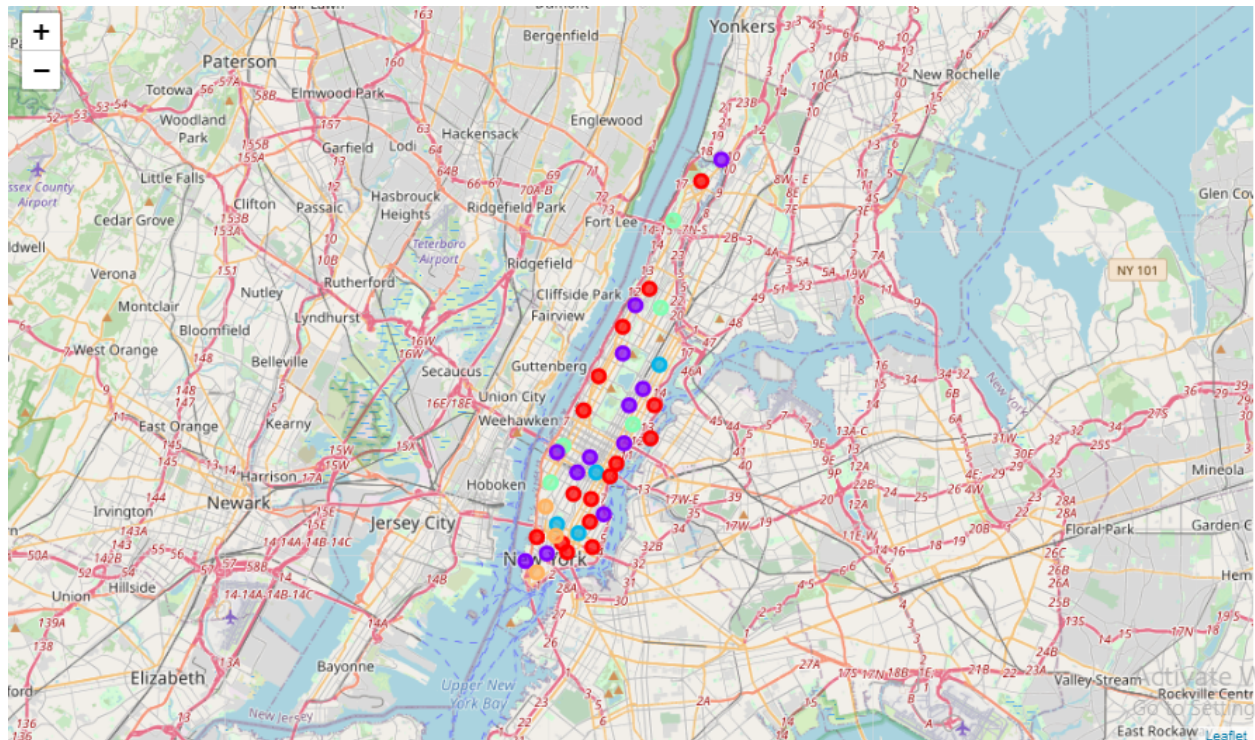
Now, since the clusters of the neighborhoods have been defined and visualized on the maps of their respective boroughs, they are finally examined individually in order to determine the discriminating venue categories that distinguish each cluster.

On examining the clusters of neighborhoods in both boroughs, Downtown Toronto and Manhattan, the results, observations, recommendations and conclusions are made which have been described in the following section of the report.

4. RESULTS SECTION

After clustering the data of the respective neighborhoods, both cities or boroughs, namely Downtown Toronto (Toronto, Canada) and Manhattan (New York, United States) have venues which can be explored and attract the tourists all over the world. The neighborhoods are much similar in features like theaters, opera houses, food places, clubs, museums, parks etc. As far as dissimilarity is concerned, it differs in terms of some unique places like historical places and monuments.

5. OBSERVATIONS AND RECOMMENDATIONS

When the tourist places in both the boroughs are compared, it can be observed 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.

As far as recommendations are concerned, Downtown Toronto Neighborhoods will be recommended to visit first. The tourists have easy travelling access due to the Airport facility, which not only saves time but also helps to save money. This saved money can be utilized to explore more attracting venues.

6. CONCLUSION SECTION

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 argument is present in both neighborhoods. The dissimilarity exists in terms of some different venues and facilities but not on a larger extent.