A

REPORT ON

A TOURIST'S NEIGHBOURHOOD BY: JAYA

PROBLEM & BACKGROUND:

Toronto and New York are the famous places in the world. They are diverse in many ways. Both are multicultural as well as the financial hubs of their respective countries. We want to explore how much they are similar or dissimilar in aspects from a tourist point of view regarding food, accommodation, beautiful places, and many more.

New York City received a ninth consecutive annual record of approximately 65.2 million tourists in 2018, Food tours are another option for visitors. New York is one of the top culinary destinations in the world. New York's food culture, influenced by the city's immigrants and large number of dining patrons, is diverse

Toronto is one of Canada's leading **tourism** destinations. In 2017, **Toronto** is one of Canada's leading **tourism** destinations.[1] In 2017.

Today Tourism is one of the pillars of the economy and the people most often visits those countries who are rich in heritage and developed enough from a foreign prospective, like friendly environment. Every city is unique in their own way and give something new. And now the information is so common regarding location of every place around the world on your fingertips which make it easier to explore. Therefore, tourists 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 assist to choose the specific places or according to their choice.

Business Problem:

Tourist may find it difficult to choose the right kind of place and in figuring out which city should they go for vacations or spending some good time Through data science, we can compare two cities (Here focused on visiting Manhattan and Downtown Torronto) and find out the most popular venues in the cities by analysing their neighbourhoods based on categories like Airport Lounge, Coffee Shop, Cafe, Restaurants & Grocery Store, commercial places, tourist places, Gastropubs etc.

Our Data Science Approach:

Data and our analysis that might contribute to determining most popular venues a tourist can visit in a country or to choose from a couple of city to travel on the basis of their choice .This project aims to predict how can a tourist or travel company always eager to find good places while checking out a new city and its different places can choose between the two cities and specific places according to their choice and on the basis of available information, and the comparison between the choices of his cities can be done.

Interest:

The Travel companies and tourist or bacpackers would be very interested in accurate prediction of cities popular places to decide which city they would like to travel and the travel companies can suggest cities to the tourist based on their choices and this will more economy and business values to the country to be visited and tourist can also get a good holiday package .

Data acquisition and cleaning

Data sources : For this problem, we will get the services of Foursquare AP to explore the data of two cities, in terms of their neighbourhoods. The data also include the information about the places around each neighbourhood like restaurants, hotels, coffee shops, parks, theaters, art galleries, museums and many more.

STEPS:

1. We selected one Borough from each city to analyse their neighbourhoods.
2. Manhattan from New York and Downtown Toronto from Toronto.
3. We will use machine learning technique, “Clustering” to segment the neighbourhoods with similar objects on the basis of each neighbourhood data.
4. These objects will be given priority on the basis of foot traffic (activity) in their respective neighbourhoods.
5. 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.¶

Data 1: New York Neighbourhood Data which will be used to get to know about the various neighbourhoods which are taken into consideration for our Project.(*Data source : https://cocl.us/new_york_dataset)*

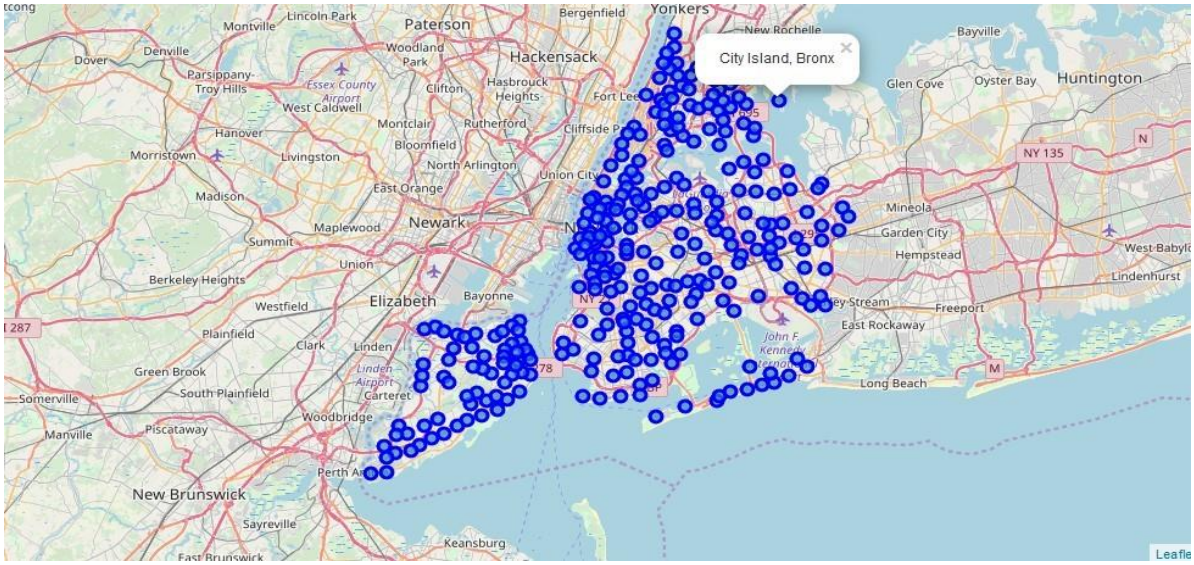
Neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighbourhood.

New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City

Exploratory Data Analysis :

Data 1- New york city Geographical Coordinates Data.

1. In this we load the data and explore data from newyork_data.json file.
2. Transform the data of nested python dictionaries into a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York city neighborhoods.
4. This data will used to get Venues data from Fouresquare.
5. We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.



New York neighbourhood visualization

Data2 : For Downtown Toronto case, we have extracted table of Toronto's Borough from Wikipedia page. Then we arrange the data according to our requirements. In the arrangement phase, which applied multiple steps including but not limited to, eliminating "Not assigned" values, combine neighbourhoods which have same geographical coordinates at each borough and sorted against the concerned borough. For data verification and further exploration, we use Foursquare API to get the coordinates of Downtown Toronto and explore its neighbourhoods. The neighbourhoods are further characterized as venues and venue categories.

To solve this problem, I will need below data:
● List of neighbourhoods in Toronto, Canada.
● Latitude and Longitude of these neighbourhoods
● Top 10 Venue data related.
This will help us find the neighborhoods that are most suitable to travel .

Extracting the Data :
● Scrapping of Toronto neighborhoods via Wikipedia
● Getting Latitude and Longitude data of these neighborhoods via Geocoder package.
Using Foursquare API to get venue data related to these neighborhoods

DATA ANALYIS :
As we have selected two cities Borough to explore their neighbourhoods. The data exploration, analysis and visualization for both boroughs are done in the same way but separately.

Now we will move towards New York Boroughs. We select "Manhattan" as a Borough and anylze its neighbourhoods as shown in below image

:

```
#reating new Dataframe manhattan_data
manhattan_data = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)
manhattan_data.head()
```

	Unnamed: 0	Borough	Neighborhood	Latitude	Longitude
0	6	Manhattan	Marble Hill	40.876551	-73.910660
1	100	Manhattan	Chinatown	40.715618	-73.994279
2	101	Manhattan	Washington Heights	40.851903	-73.936900
3	102	Manhattan	Inwood	40.867684	-73.921210
4	103	Manhattan	Hamilton Heights	40.823604	-73.949688

Using the foursquare api we can get the geographical coordinates of MANHATTAN

```
#Let's get the geographical coordinates of Manhattan.
address = 'Manhattan, NY'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Manhattan are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Manhattan are 40.7896239, -73.9598939.
```

Using the foursquare api we can get the geographical coordinates of DOWNTOWN TORRONTO

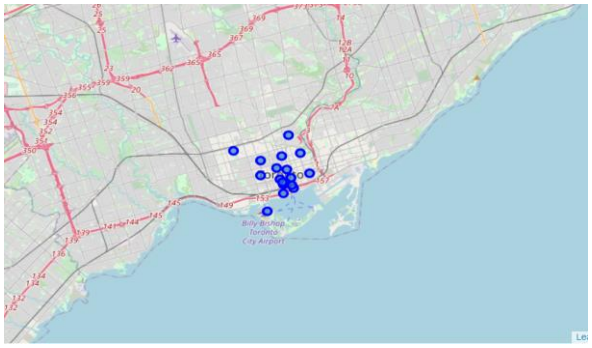
```
# get the geographical coordinates of Downtown Toronto
address = 'Downtown Toronto, ON, Canada'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude_downtown_toronto = location.latitude
longitude_downtown_toronto = location.longitude
print("Downtown Toronto","latitude",latitude_downtown_toronto, "& " "longitude" ,longitude_downtown_toronto)
```

```
Downtown Toronto latitude 43.6563221 & longitude -79.3809161
```

VISUALIZATION

We visualize the data many times at different stages. In the beginning, we visualize the selected borough neighbourhoods so that we can get an idea or confirmation regarding the coordinates of that Borough. The second time after clustered the neighbourhoods, we visualize the clusters to name them. Assigning the names are very important because it can identify the areas or specific places in each cluster.
Before clustering:

Map of Downtown Torronto neighbourhoods:



Map of Manhattan neighbourhoods:

ANALYSIS:

We analyze both boroughs neighborhoods through one hot encoding (giving ‘1’ if a venue category is there, and ‘0’ in case of venue category is not there). On the basis of one hot encoding, we calculate mean of the frequency of occurrence of each category and picked top ten venues on that basis for each neighborhood. It means the top venues are showing the foot traffic or the more visited places.

Resulting neighbourhoods For Torronto :

```
Harbourfront,Regent Park
Ryerson,Garden District
St. James Town
Berczy Park
Central Bay Street
Christie
Adelaide,King,Richmond
Harbourfront East,Toronto Islands,Union Station
Design Exchange,Toronto Dominion Centre
Commerce Court,Victoria Hotel
Harbord,University of Toronto
Chinatown,Grange Park,Kensington Market
CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South
Niagara
Rosedale
Stn A PO Boxes 25 The Esplanade
Cabbagetown,St. James Town
First Canadian Place,Underground city
Church and Wellesley
```

venues returned for each neighborhood:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Adelaide,King,Richmond	20	20	20	20	20	20
Berczy Park	20	20	20	20	20	20
CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South Niagara	15	15	15	15	15	15
Cabbagetown,St. James Town	20	20	20	20	20	20
Central Bay Street	20	20	20	20	20	20
Chinatown,Grange Park,Kensington Market	20	20	20	20	20	20
Christie	16	16	16	16	16	16
Church and Wellesley	20	20	20	20	20	20
Commerce Court,Victoria Hotel	20	20	20	20	20	20
Design Exchange,Toronto Dominion Centre	20	20	20	20	20	20
First Canadian Place,Underground city	20	20	20	20	20	20
Harbord,University of Toronto	20	20	20	20	20	20
Harbourfront East,Toronto Islands,Union Station	20	20	20	20	20	20
Harbourfront,Regent Park	20	20	20	20	20	20
Rosedale	4	4	4	4	4	4
Ryerson,Garden District	20	20	20	20	20	20
St. James Town	20	20	20	20	20	20
Stn A PO Boxes 25 The Esplanade	20	20	20	20	20	20

Applying one hot encoding :

	Vietnamese Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0

Display the top 10 venues for each neighborhood:

	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
0	Harbourfront,Regent Park	Coffee Shop	Park	Breakfast Spot	Spa	Performing Arts Venue	Chocolate Shop	Pub	Mexican Restauran
1	Ryerson,Garden District	Café	Steakhouse	Burrito Place	Burger Joint	Movie Theater	Pizza Place	Beer Bar	Plaza
7	Harbourfront East,Toronto Islands,Union Station	Café	Park	Hotel	New American Restaurant	Sports Bar	Salad Place	Bakery	Ice Cream Shop
11	Chinatown,Grange Park,Kensington Market	Café	Vietnamese Restaurant	Caribbean Restaurant	Organic Grocery	Arts & Crafts Store	Bakery	Bar	Cheese Shop
12	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Service	Airport Lounge	Airport Terminal	Plane	Airport	Airport Food Court	Airport Gate	Boutique
15	Cabbagetown,St. James Town	Restaurant	Café	Butcher	Bakery	General Entertainment	Gift Shop	Diner	Deli / Bodega
17	Church and Wellesley	Breakfast Spot	Hobby Shop	Burger Joint	Bubble Tea Shop	Mexican Restaurant	Park	Bookstore	Pizza Place

Examine Clusters For Downtown Torronto

Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster. Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster.

Cluster 1 (Airport Lounge, Coffee Shop, Cafe, Restaurants & Grocery Store

Cluster 2 (Gastropubs)

Cluster 3 (Cafes)

Cluster 4 (Coffee Shop, Cafe, Park & Japanese Restaurant)

Cluster 5 (Seafood, steakhouse, Hotel & Cafe)

Exploring Neighborhoods in Manhattan¶

Applying one hot encoding :

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaur
0	Marble Hill	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	0

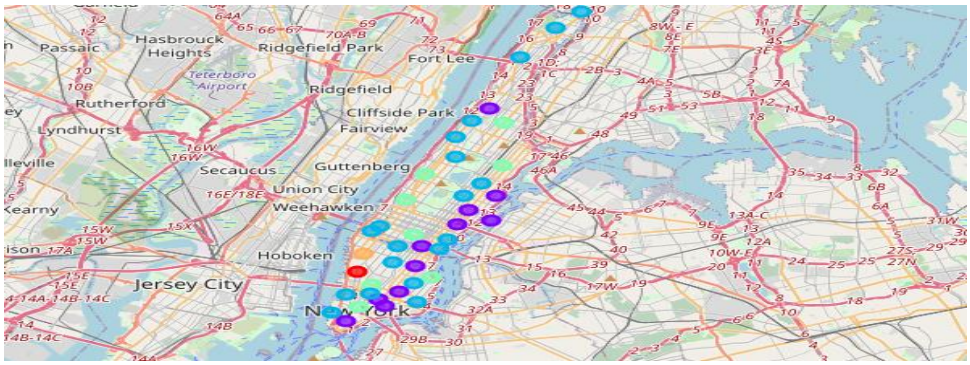
Analysing Top-10 Venues in Manhattan

```
There are 205 uniques categories.
```

	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	2	Discount Store	Coffee Shop	Yoga Studio	Diner	Seafood Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	1	Noodle House	Chinese Restaurant	Sandwich Place	Hotel	Greek Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	2	Wine Shop	Park	Café	Breakfast Spot	Bakery
3	Manhattan	Inwood	40.867684	-73.921210	2	Wine Bar	Park	Café	Yoga Studio	Coffee Shop
4	Manhattan	Hamilton Heights	40.823604	-73.949688	1	Yoga Studio	Mexican Restaurant	Caribbean Restaurant	Cocktail Bar	Japanese Restaurant

Clustering Neighbourhoods:

Now we applied Machine Learning Technique “Clustering” to segment the neighbourhoods in similar objects cluster. This will help to analyze from Tourist perspective and we can easily extract the Tourist places which are present on one of the clusters.



Examine clusters for Manhattan

Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster.

CLUSTER 1: RESIDENTIAL PLACES

CLUSTER 2: COMMERCIAL PLACES

CLUSTER 3: TOURIST AREAS AND HUBS

CLUSTER 4: CENTER ACTIVITY

CLUSTER 5: CULTURAL AND GOING OUT PLACES

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.

Observations & Recommendations:

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 i**n Downtown Toronto while venues like Nightlife, Climbing gym and Museums are present in Manhattan.
As far as concern to recommendations, we recommend Downtown Toronto Neighborhoods will be considered first to visit. The tourists have an easily travelling access due to Airport facility, which not only saves time but also helps to save money. This saved money can be utilized to explore more, the attracting venues.

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 exten