PART -2 (A PERFECT TOURISTS NEIGHBOURHOODS)

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.
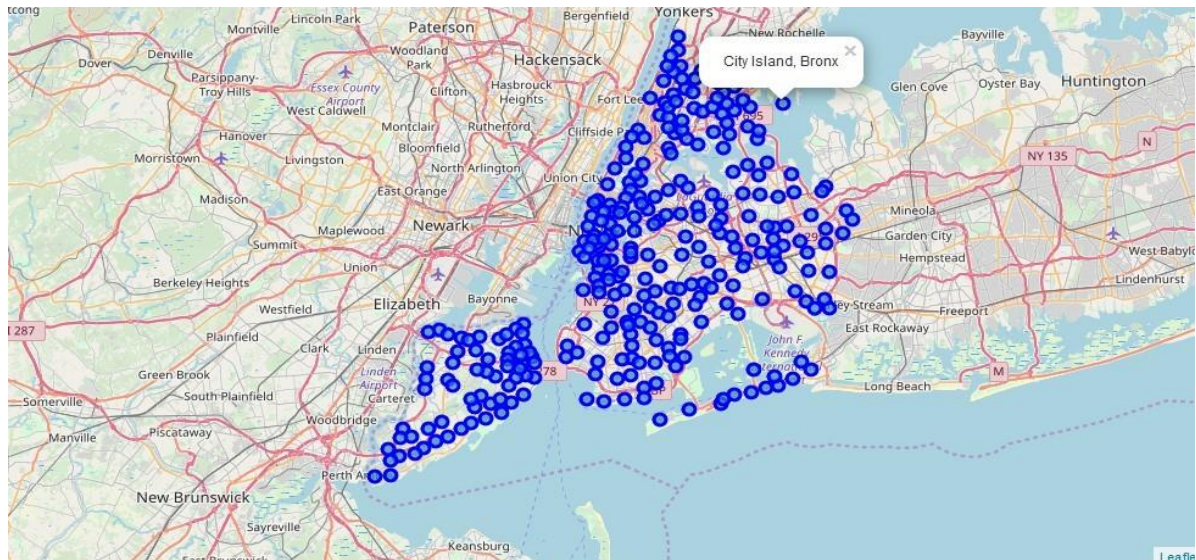
New York neighbourhood visualization

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: