Capstone Project - The Battle of Neighbourhoods (Part-1)

Data Section

New York City's demographics show that it is a large and ethnically diverse metropolis. With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, and French etc.

For this project we need the following data:
- New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.
  - Data source : https://cocl.us/new_york_dataset
  - Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of New York City

- Indian restaurants in each neighbourhood of New York City.
  - Data source : Foursquare API
  - Description: By using this API we will get all the venues in each neighbourhood. We can filter these venues to get only Indian restaurants.

- GeoSpace data
  - Data source : https://data.cityofnewyork.us/City-Government/BoroughBoundaries/tqmj-j8zm
  - Description: By using this geo space data we will get the New York Borough boundaries that will help us visualize choropleth map.

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

Libraries Which are Used to Develope the Project:

- Pandas: For creating and manipulating dataframes.

- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

- Scikit Learn: For importing k-means clustering.

- JSON: Library to handle JSON files.

- XML: To separate data from presentation and XML stores data in plain text format.

- Geocoder: To retrieve Location Data.

- Beautiful Soup and Requests: To scrap and library to handle http requests.

- Matplotlib: Python Plotting Module.