THE BATTLE OF NEIGHBORHOODS

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.