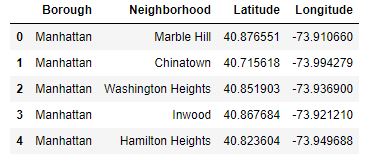
**DATA DESCRIPTION**

For this problem, we will get the services of Foursquare API to explore the data of two cities, in terms of their neighborhoods. The data also include the information about the places around each neighborhood like restaurants, hotels, coffee shops, parks, theaters, art galleries, museums and many more. We selected two cities to analyze their neighborhoods; Manhattan from New York and Central Toronto from Toronto. We will use machine learning technique, “Clustering” to segment the neighborhoods with similar objects on the basis of each neighborhood data. These objects will be 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 Toronto case, we have extracted table of Toronto’s Borough (Central Toronto) from Wikipedia page (<https://www.en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>), where a table containing the list of postal codes, cities, boroughs and neighborhoods of canada. 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, we also get the list of coordinates for each of the locations from a csv file, from this page, (<http://cocl.us/Geospatial_data>) after which we would sort each coordinates with their respective locations from the previous data. The resulting data frame is shown below.



Also, for Manhattan, being a city in New York, we extracted its data from a JSON file, (newyork\_data.json) after which we filtered out the details of Neighborhoods under Manhattan and converted it into a data frame as shown below:



For data verification and further exploration, we use Foursquare API to get the coordinates of Toronto and explore its neighborhoods (for both cities). The neighborhoods are further characterized as venues and venue categories as shown below. This would be further processed to be used in the machine learning algorithm, K-means Clustering in order to segment them and the comparison would be made between the two cities.

