COURSERA CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS

DATA SOURCE AND EXPLORATION :

The data that will be used in this analysis is combination of a json file (newyork_data.json) which is downloaded from the link- https://cocl.us/new_york_dataset and the information of venues in Foursquare. The required data from newyork_data.json , converted into a pandas dataframe as needed. This dataframe will contains Boroughs , neighborhoods in Boroughs and their latitudes and longitudes.

Borough	Neighborhood	Latitude	Longitude
Bronx	Wakefield	40.894705	-73.847201
Bronx	Co-op City	40.874294	-73.829939
Bronx	Eastchester	40.887556	-73.827806
Bronx	Fieldston	40.895437	-73.905643
Bronx	Riverdale	40.890834	-73.912585

Then comes the Foursquare data about neighborhood and venues(restaurants). The venues of different neighborhoods needs to

be retrieved and should be stored in a pandas dataframe. Then the total number of restaurants are considered along with the type of restaurants(cuisine type) in different neighborhoods.

This can be achieved by grouping the data according to neighborhoods. The next step is to categorize the restaurants according to their types(cuisine) and find their frequency in each neighborhood. This will help to determine the most in-demand cuisine in different neighborhoods.

After that, using the clustering techniques, it can be easily determined that which type of restaurant(cuisine) is in demand in which neighborhoods, as it has greater number of restaurants of that specific category.