Description of Data

a. Data Source

The data for neighborhoods of New York City was obtained from the Spatial Data Repository of NYU. The .json file obtained has the coordinates for all the neighborhoods of New York City.

b. Data Cleaning

I cleaned and reduced the data to the Boroughs Manhattan and Brooklyn. For that I first transformed the data of NYC to pandas data frame and then sliced it into two new data frames for Manhattan and Brooklyn. Then used geopy library for obtaining the latitude and longitude values of Manhattan and Brooklyn. Then created a map of Manhattan and Brooklyn using Folium.

c. Explore neighborhoods using Foursquare data

After that I used Foursquare API for exploring the venues and Neighborhoods in Manhattan and Brooklyn.

d. Analyses the neighborhoods data

Further analyze each neighborhood by seeing how many venues are there in each neighborhood.

e. Cluster the neighborhoods and examine the clusters for comparing both boroughs and solving the problem of where to start a new business.

Then cluster the neighborhoods and examine the clusters formed to determine discriminating venue categories that distinguish each cluster. Using this cluster information, we can identify how similar or dissimilar are both boroughs and also one can decide where to open a new restaurant or shop.