Data

The New York geographic data are the newyork_data.json provided in the Applied Data Science Capstone Class of the Coursera IBM Data Science Professional Certificate Specialization. The Toronto geographic borough and geographic data are from Wikipedia and Geospatial_Coordinates.csv file provided in class. The geographic data of both cities will be used to construct the dataset of neighborhoods in each borough and their coordinates and convert to two pandas dataframes for New York and Toronto. Each row of the data frame is one neighborhood with its coordinates. A function looping through each row of each dataframe sends requests to Foursquare API to request a maximum of 100 venues in a radius of 500 meters for each neighborhood coordinates. The returned venue data for every neighborhood in each city are converted to a data frame for each city. These data will be cleaned and pre-processed with on-hot encoding to get ready for Kmeans clustering analysis.