Data

To solve the problem, we will need the following data :

- In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the boroughs and the neighborhoods exist in each borough.
- We will also need the latitude and longitude coordinates of each neighborhood. This is required to plot the map and to get data about the venues.
- Along with it we will require complete venue data, i.e. venue name, location and category in order to perform clustering of the neighborhoods.

Sources of data and method to extract them :

The data set for Toronto is available at wikipedia. It includes postal-code, boroughs and neighborhood names. Here is the link :
Toronto Data set with postal-codes, boroughs, neighborhoods and their locations
We will scrap this data from wikipedia into a dataframe. But we need to clean and prepare this data as it contains some 'Not assigned' boroughs and neighborhoods. Along with it we have to add location coordinates of each neighborhood, we found it on a coursera given link :
Geospatial data of Toronto

The data set for New York was given by coursera. It includes boroughs, neighborhood names and location of each neighborhood. Here is the link :
New York data set
This data set was already cleaned and formatted, so we just loaded it into pandas dataframe.

In between the process i reduced the list of boroughs by removing duplicate values so that there would only be one occurrence of each neighborhood and their corresponding data.

Here we used Foursquare API to analyze the nearby places to these neighborhoods and see the proximity of important places from the corresponding neighborhoods.