Data

I am going to work on New York City neighborhood. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is: https://cocl.us/new_york_dataset

New York City geographical coordinates data will be utilized as input for the Foursquare API that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. I am going to find out the best location for starting a Thai restaurant so I will retrieve the venues for Thai category. Thai category Id 4bf58dd8d48988d149941735 is used for retrieving data from Foursquare API.

So the dataset having new York neighborhood names will be used for boroughs and the Thai restaurant venues are collected by foursquare API and both of them will be merged to get a complete dataset on which clustering can be done.