

DATA

Based on definition of our problem, the factor that will influence our decision is the number of existing Indian restaurants in the neighborhood. We will select the Cluster of neighborhood which has the lowest density of Indian restaurants.

To Analyse the neighborhoods of Manhattan, we will have to first analyse the New York City data.

New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 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 at : https://geo.nyu.edu/catalog/nyu_2451_34572

Following data sources will be needed to extract/generate the required information:

- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of New York City/Manhattan neighborhoods will be obtained using **geocoder**