Data

In this section, I will describe the data that I will be using to solve my problem and execute my idea. Recall that the broad goal of my project is to **Use location data to improve the overall health of the residents of New York City, NY**. I will follow the steps below to obtain the data that I need to carry out my analysis.

1. First, I will gather data about the neighborhoods of New York City (NYC). NYC 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 all the neighborhoods that exist in each borough as well as the coordinates of each neighborhood. Luckily, this dataset exists for free on the web. I will obtain it from the following link: https://geo.nyu.edu/catalog/nyu_2451_34572

2. I will use the Foursquare API location data to explore all neighborhoods of NYC. This will give me all the different types of venues (be it coffee shops, restaurants, bars, parks, yoga studios, hospitals, or schools …) in each neighborhood. I will put this data into a data frame. I will then create a map of NYC with all the neighborhoods.

3. Next, I will extract all those neighborhoods that lack hospitals/medical centers, put them into a data frame, and examine their distribution on a map of NYC (after some segmenting and/or clustering if needed). By comparing against the previous map, I can determine which areas need more hospitals.

4. I will repeat the above procedure for parks/gyms/fitness centers/yoga studios across the city.

This will then allow me to make recommendations to the Department of City Planning of NYC as to which areas of NYC need more health-improvement facilities.