Data

To do this project we will need the following data:

1. List of Neighborhoods in Mumbai. This will define the scope of the project – which is going to be limited to Mumbai, the financial capital of India.
2. The Latitude and Longitude co-ordinates of the neighborhoods. This will be very helpful in plotting the neighborhoods on the map and get the data of the venues.
3. Venue data, especially of shopping malls. This will be useful and important for the clustering part of the project.

Sources of Data and How to extract them

The Wikipedia page with the list of neighborhoods in Mumbai (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai) contains the list of the neighborhoods in Mumbai, nearly 95 of them. Since, the table of the longitude and latitude is already available, we will use geocoder to recheck the co-ordinates.

After that we will use the Foursquare API to get the venue data for those neighborhoods. Foursquare has a huge database of more than a 105 million places and is used by nearly 120,000 developers worldwide. While Foursquare provides a huge amount of venue data for neighborhoods, we will be specifically looking into the venue data for shopping malls in the neighborhoods of Mumbai. This project will be requiring some data science skills in terms of working with an API like Foursquare, data cleaning, data wrangling, machine learning in terms of K-Means Clustering and data visualization on maps using Folium.