Best Neighborhood to live!!! – Data Section

The purpose of the project is to analyze if there are any neighborhoods in a city which are suitable for living based on particular interests so that the expense of money, time and energy is minimum. **As an example, I have chosen the Mumbai metropolitan city to use for my project**. Firstly, we will need the information about the accurate neighborhoods in Mumbai city. **I decided to scrape the Wikipedia page named on Mumbai neighborhoods**. (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai) I have used Beautiful Soup API to extract the webpage and then get the names of relevant neighborhoods.

After extraction of neighborhood names, **we used the geolocator API to get the latitudes and longitudes of each neighborhoods**. We make a data frame containing the list of neighborhoods and their associated latitudes and longitudes.

Now the most important part is **searching for the venues. For this we use Foursquare API**. The first part is selecting a query which involves selecting the search query term such as research, education institutes, amusement parks, coffee shops or movie theaters etc. **Based on what interests you provide, we can analyze which neighborhood to live so the maximum access of interested places is provided**.

At first, I had decided to rank the places based of user's interests based on the tips or reviews of those places. As turns out, there are very few people in Mumbai who actually use Foursquare API and because of that there are no ratings or tips associated with any of the places. (I have demonstrated this in actual code file**) Because we could not use rating or tips system to rank the venues, we decided to demonstrate the project idea with the first limited venue results and based on those decide which neighborhood is best to live.**

Finally, **we use the foursquare API to generate a query which has the user's venue interest and that gives provides result in json format**. This json data has information of maximum 50 venues based on user's interest and we extract the relevant information from that json to convert in into a data frame. The 50 venues information is then parsed into proper latitude, longitude and venue name formats and used for further processing.