



Opening a new Fitness Center in Mumbai,IN

IBM Applied Data Science Capstone

AUGUST 2020



Data

For this project, we will use the following data components:

- List of neighborhoods in Mumbai- This list will be extracted from Wikipedia and will contain a list of all the neighborhoods and their respective regions in the city of Mumbai
- Latitude and Longitude coordinates for these neighborhoods will be available from the above table as well as from the Geocoder library
- Venue and location data in and around points of interest- To identify regions of interest and shortlist clusters which can be potentially key to determining the new location



Data Sources

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai contains a list of all the neighborhoods in Mumbai along with their regions and latitude and longitude data. Using pandas library's "read_html" module, we will extract our desired table for further analysis. We will then use the geocoder package to verify and update the latitudes and longitudes.

Once that is done, we will use the Foursquare API to determine the data for venues within the neighborhoods. Foursquare also provides the category types for these venues, allowing for further analysis of the data. Once the data is cleaned and munging is done, we will begin exploring the data and how we can cluster locations of interest(particularly fitness centers) to develop insights for a new location for the fitness center.