

**Data Requirements:** To build the recommender system, we require population data regarding the city and areas in Bengaluru. The most important data that needs to be collected are:

- Geographic coordinates such as latitude and longitude of the areas to find out where they are located
- Population of the area where the venues are located
- Average income of population across the areas.

The following steps must be accomplished for data collection:

- To go to a restaurant's location, we need to know its Latitude and Longitude so that we can point to its coordinates and generate a map with all of the restaurants labelled appropriately.
- A neighborhood's population is a significant element in influencing a restaurant's growth and the number of customers that come in to dine. Logic dictates that the larger the population of an area, the more people will be interested in walking into a restaurant openly, and the smaller the population, the less people will frequent a restaurant. Also, the higher the number of visitors, the higher the restaurant's rating because it is visited by individuals of various tastes. As a result, it is a crucial element.
- A neighborhood's income is just as significant as its population. The wealth of a neighborhood is directly related to income. If residents in an area make more than the national average, it is highly likely that they will spend more, although this is not always the case. As a result, a restaurant's evaluation is proportionate to the neighborhood's revenue.

### **Data Collection:**

Gathering location position is not hard, but that was not accessible on open access websites such as Wikipedia, India's government website, census report websites, and so on after more than two days of searching. So I opted to utilize Google Maps API to get latitude and longitude, however the free account only allowed me to make a limited amount of requests. Firstly, I used BeautifulSoup4 to scrape a list of neighbors from Wikipedia<sup>1</sup>. The table headers serve as boroughs, while the data serve as neighborhoods. Bangalore is divided into eight boroughs and 64 neighborhoods. So, I did a Google search and manually looked up each neighborhood's coordinates. Following that, I created the data

---

<sup>1</sup> [https://en.wikipedia.org/wiki/List\\_of\\_neighbourhoods\\_in\\_Bangalore](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore)

frame below. Because the information is easily available, determining the population by neighborhood is very simple. In Bangalore, however, this is not the case. For a few areas, it was possible to find population figures. The population of the rest of the neighborhood is estimated and may be wrong, but because this is a demonstration project, the major goal is to get the model to operate. Neighborhood income is freely accessible through Wikipedia page. We next extract all of the columns from this Wikipedia page and convert them to a pandas dataframe using the BeautifulSoup4 package, a Python tool that helps you scrape data from web sites. The latitude and longitude of all of the regions in the dataframe are then obtained using Python's GeoPy module. The purpose of using Foursquare is to find the closest venue locations so that we may build a cluster. The Foursquare API makes use of the ability of Foursquare to identify nearby venues within a certain radius (in my instance, 500 meters) as well as matching coordinates, venue location, and names.