Data Requirements**:**

To find a solution to the questions and build a recommender model, we need data and lots of data. Data can answer question which are unimaginable and non-answerable by humans because humans do not have the tendency to analyse such large dataset and produce analytics to find a solution.

Let's consider the base scenario

Suppose I want to find a restaurant, then logically, I need 3 things:

1. Its geographical coordinates (latitude and longitude) to find out where exactly it is located.
2. Population of the neighbourhood where the restaurant is located.
3. Average income of neighbourhood to know how much the restaurant would worth.

Let’s take a closer look at each of these

1. To access location of a restaurant, it’s Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the restaurants with its labels respectively.
2. Population of a neighbourhood is very important factor in determining a restaurant's growth and the amount of customers who turn up to eat. Logically, the more the population of a neighbourhood, the more people will be interested to walk openly into a restaurant and less the population, less number of people frequently visit a restaurant. Also if more people visit, better the restaurant is rated because it is accessed by different people with different taste. Hence is very important factor.
3. Income of a neighbourhood is also very important factor as population was. Income is directly proportional to richness of a neighbourhood. If people in a neighbourhood earns more than an average income, then it is very much possible that they will spend more however not always true with very less probability. So a restaurant assessment is proportional to income of a neighbourhood.

Data Collection:

1. Collecting geographical coordinates is not difficult job, as I could fetch them directly from foursquare API by giving the location / neighbourhood details to fetch latitude and longitude.
2. Initially I scrapped list of neighbours’ using beautifulSoup4 from [wikipedia] (<https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore>). The table headings becoming the boroughs and data becoming the neighbourhoods. Bangalore has 8 boroughs and 64 neighbourhoods.
3. Population by neighbourhood is again easy to find out given that it’s readily available. But in case of Bangalore, it is again not the case. I was able to find population data for few cities. [Here is the link] (<https://indikosh.com/dist/655489/bangalore>). Rest other neighbourhood population is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model.
4. Income by neighbourhood is again easy to find out given that it’s readily available. But in case of Bangalore, it is again not the case. I was able to find Income data for main city. [Here is the link] (<https://en.wikipedia.org/wiki/List_of_Indian_cities_by_GDP_per_capita>). Neighbourhood Income is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model.
5. Foursquare API: Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius (in my case: 500mts) and also corresponding coordinates, venue location and names.