**Title : Restaurant recommender system in Bangalore**

**Problem background:**

Bangalore is the capital and largest city of the Indian state of Karnataka. With a population of over 15 million (as of January 2016), Bangalore is the third largest city in India and 27th largest city in the world.

The diversity of the cuisine available is reflective of the social and economic diversity of Bangalore. Roadside vendors, tea stalls, South Indian, North Indian, Muslim food, Chinese and Western fast food are all very popular in the city. Udupi restaurants are very popular and serve predominantly vegetarian cuisine. The Chinese food and the Thai food served in most of the restaurants are can be customized to cater to the tastes of the Indian population. Bangalore can also be called a foodie's paradise because of its vast variety of foods and edibles with a touch of Bangalore's uniqueness and tradition.

**Problem description:**

Suppose i travel and keep changing places very frequently. This is very hectic and plus i get to experience very different types of environment, of which i donot have much knowledge about. In such situation, food can be an important factor for decided how you rate your trips and plus also recommmending it to the people. Food can also attract people around to world to try it out if it were to be the best. In such scenarios, we need to find the right place, at reasonable cost, to serve us the best possible way. So there are few questions that must be addresed, such as :

1. How many types of foods are available in the restaurant ?
2. which is the most nearest to me with good rating ?
3. How many "similar" restaurants are available near by me ?
4. Do the "similar" restaurants cost more ? if so, what speciality do that have ?

To address such question, XXYZ company's manager decides to allocate this project to me not just to find out solutions to the questions but also build a system that can help in recommending new places based on their rankings compared to the previously visited by me.

Expectations from this recommender system is to get answer for the questions, and in such a way that it uncovers all the perspective of managing recommendations. It is sighted to show :

1. What types of restaurants are present in a paeticular area ?
2. where are the similar restaurant present based on a preference to particular food ?
3. How do different restaurants rank with respect to my preferences ?

**Data :**

**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 analyze such large dataset and produce analytics to find a solutions.

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 neighborhood where the restaurant is located.

3. Average income of neighborhood to know how much is the restaurant 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 neighborhood is very important factor in determining a restaurant's growth and amount of customers who turn up to eat. Logically, the more the population of a neighborhood, 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 neighborhood is also very important factor as population was. Income is directly proportional to richness of a neighborhood. If people in a neighborhood 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 neighborhood.

**Data collection:**

1. Collecting geographical coordinates is not difficult but after googling for more than 2 days, it was not available on open source data websites such as Wikipedia, India gov website, census report websites etc. So I decided to use Google maps API to fetch latitude and longitude but google API has limited number of calls that I could make with my free account. So it would take around 15 - 20 days to fetch location of all the neighborhoods in Bangalore.

Initially I scrapped list of neighbor's using beautifulSoup4 from [wikipedia](https://en.wikipedia.org/wiki/List\_of\_neighbourhoods\_in\_Bangalore). The table headings becoming the boroughs and data becoming the neighborhoods. Bangalore has 8 boroughs and 64 neighborhoods. So i manually googled each neighborhood to find its corresponding latitude and longitude. After doing so, I produced the following data frame.



2. Population by neighborhood 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 neighborhood population is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The data frame for Bangalore neighborhood population looks like:



3. Income by neighborhood 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. (https://en.wikipedia.org/wiki/List\_of\_Indian\_cities\_by\_GDP\_per\_capita). Neighborhood Income is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The data frame for Bangalore neighborhood population looks like:



4. 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. After calling, the following data frame is created:

