Kumar, Ashish

RESTAURANT RECOMMENDER SYSTEM

**RESTAURANT RECOMMENDER SYSTEM**

**IBM Capstone project**



Restaurant recommender system is a machine learning model, developed to demonstrate as a capstone project to IBM through coursera. It recommends restaurants based on user’s likes and dislikes and his previous interest data.

Table of contents

* Introduction section
* Data section
* Methodology section
* Result section
* Discussion section
* Conclusion section

**1.Introduction :**

**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.

**Success rate:**

With restaurants evolving, new food categories emerge, hybrid food starts to be more popular, we need a system that could help us access vast number of food varieties. It is impossible for a person to ask each and every one about their visit to a particular place and also not everyone remembers everything. On the other hand, Computers are good at remembering things, and with Machine learning to its peak, it high time technology will by our personal guidance and help us personally based on our likes and dislikes. So people would care about this project as their personal assistance and success rate could certainly increase with time.

**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.

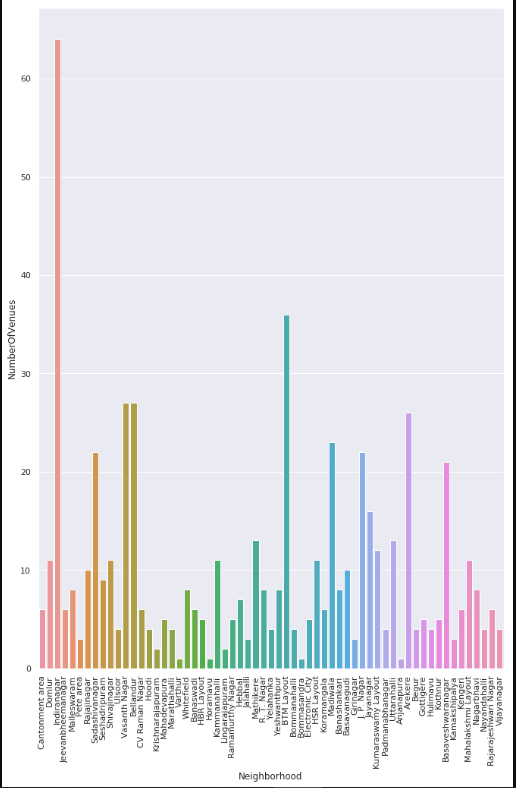
**Data collection:**

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.
2. 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.
3. 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>).
4. 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. [Here is the link](<https://en.wikipedia.org/wiki/List_of_Indian_cities_by_GDP_per_capita>).
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.

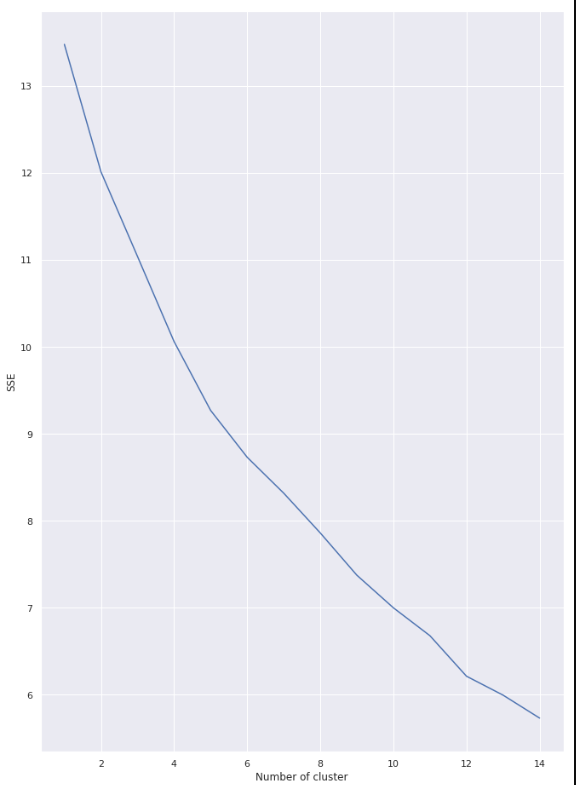
3.**Methodology :**

**Exploratory analysis:**

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched. Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data. While exploring the dataset, I found out that Inderanagar has most number of venues while Varthur has the least.



Also while producing graph for number of cluster, I produced a graph to explore all the values for n\_clusters and then finding the best by exploring the elbow graph.



**Inferential analysis:**

Most important factors while building the recommender system were population and income. They are the most import factor because they have a nonlinear relationship according to our dataset. It needed to make some inferential analysis to understand this nonlinear relationship. As the amount of population increases, it does not necessarily mean that average income of a neighborhood will also increase. It is true to most of the case but also many cases differ to follow this trend. Similarly, a neighborhood with less number of people may not necessarily have less average income. It is possible to have less number of people and more income and vice versa.

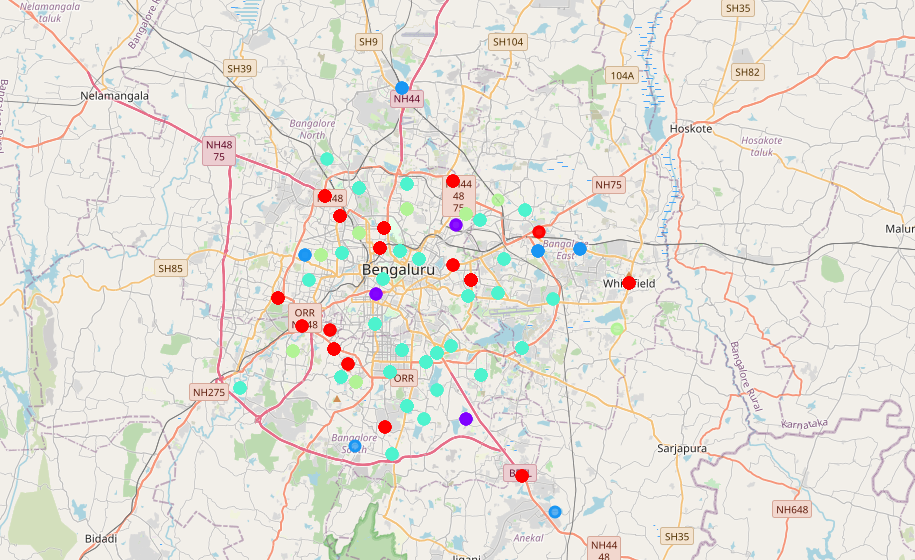
**4. Result :**

The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy. During the runtime of the model, a simulation was done by taking ‘Whitefield’ as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of ‘Whitefield’. The following image shows the result:



**5.Discussion :**

Since there was a nonlinear relationship between income and population, it can be concluded that we must always perform inferential approach to find relationship among different set of features. Also during clustering, similar neighborhoods must be dumped into the right cluster. The following graph shows the clusters:



Another observation that we can make is that choosing number of clustering could produce very diverse results. Some may be over fitted or some may be under fitted. Hence analysis of number of clusters must be done. Ref elbow\_graph in the Methodology section.

**6.Conclusion :**

The recommender system is a system that considers factors such as population, income and makes use of Foursquare API to determine nearby venues. It is a powerful data driven model whose efficiency may decrease with more data but accuracy will increase. It will help users to finish their hunger by providing the best recommendation to fulfil all their needs.