# HOTEL RECOMMENDER

## Data Science Capstone Project

Hotel recommender framework could be a machine learning show, created to illustrate as a capstone extend to IBM through coursera. It prescribes eateries based on user's likes and loathes and his past intrigued information.

# Table of contents

# 1. Introduction:

## Issue Foundation:

Bengaluru is the capital and biggest city of the Indian state of Karnataka. With a populace of over 15 million (as of January 2016), Bengaluru is the 15th biggest city in India. The differences of the cooking accessible is intelligent of the social and financial differing qualities of Bengaluru. Roadside merchants, tea slows down, South Indian, North Indian, Muslim nourishment, Chinese and Western quick nourishment are all exceptionally prevalent within the city. Udupi eateries, are exceptionally well known and serve overwhelmingly vegan food. The Chinese nourishment and the Thai nourishment served in most of the eateries are can be tweaked to cater to the tastes of the Indian populace. Bangalore can too be called a foodie's heaven since of its endless assortment of nourishments and edibles with a touch of Bengaluru's uniqueness and convention*

## Problem description:

## Issue portrayal:

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 do not have much knowledge about. In such situation, food can be an important factor for decided how you rate your trips and plus also recommending 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 addressed, such as:  How many types of foods are available in the restaurant? which is the most nearest to me with good rating? How many "similar" restaurants are available nearby me? Do the "similar" restaurants cost more? if so, what specialty 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:

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

## Target People:

Target audiences for this project does not limit to a person who keeps travelling but everyone. People could simply decide to look for a similar restaurant all the time because they are addicted to a specific category of food. People who rarely use restaurants would prefer to have the most rated restaurants nearby them and all this could be easily handled by our recommender system. So target for this project is basically everyone who is exploring different places or similar places.

## Positive 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 everyone 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 guidence 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.

# 2. Data :

## Information prerequisites:

To discover a arrangement to the questions and construct a recommender show, we require information and parcels of information. Information can reply address which are unfathomable and non liable by people since people don't have the inclination to examine such huge dataset and deliver analytics to discover a solutions. Let's consider the base situation : Suppose i need to discover a eatery, at that point coherently, I require certain things such as ,Its geological coordinates(latitude and longitude) to discover our where precisely it is located. Population of the neighborhood where the eatery is located. Average salary of neighborhood to know how much is the eatery worth. Lets take a closer see at each of these : To get to area of a eatery, its Scope and Longitude is to be known so that ready to point at its facilitates and make a outline showing all the eateries with its names respectively. Population of a neighborhood is exceptionally imperative calculate in deciding a restaurant's development and sum of clients who turn up to eat

## Information collection

Collecting geological facilitates isn't troublesome but after googling for more than 2 days, it was not accessible on open-source information websites such as Wikipedia, India gov site, census report websites etc. So, I chose to utilize Google maps API to bring scope and longitude but google API has constrained number of calls that i seem make with my free account. At first i rejected list of neighbor's utilizing beautifulSoup4 from Wikipedia. The table headings getting to be the boroughs and information getting to be the neighborhoods. So i physically googled each neighborhood to discover its comparing scope and longitude.

### DATA LINK

HTTPS://EN.WIKIPEDIA.ORG/WIKI/LIST_OF_NEIGHBOURHOODS_IN_BANGALORE

We'll need information almost diverse settings in several neighborhoods of that particular borough. In order to pick up that data we'll utilize "Foursquare" locational data. Foursquare may be a area information supplier with data almost all way of settings and occasions inside an range of intrigued. Such data incorporates scene names, areas, menus and indeed photographs. As such, the foursquare area stage will be utilized as the sole information source since all the expressed required data can be gotten through the API. After finding the list of neighborhoods, we at that point interface to the Foursquare API to assemble data approximately settings interior each and each neighborhood. For each neighborhood, we have chosen the sweep to be 100 meters. The information recovered from Foursquare contained data of settings inside a indicated separate of the longitude and scope of the postcodes. The data will get as per setting are Nearest Places, Latitude, Longitude, Venue, title of a store or restaurant, Scene Latitude, Setting Longitude, Scene Category

| | Borough | Neighborhoods | Latitude | Longitude | Population | City | AverageIncome |
|---|---|---|---|---|---|---|---|
| 0 | Central | Cantonment area | 12.972442 | 77.580643 | 866377 | Bangalore | 18944.099792 |
| 1 | Central | Domlur | 12.960992 | 77.638726 | 743186 | Bangalore | 56837.022198 |
| 2 | Central | Indiranagar | 12.971891 | 77.641151 | 474289 | Bangalore | 41991.817435 |
| 3 | Central | Jeevanbheemanagar | 12.962900 | 77.659500 | 527874 | Bangalore | 6667.447632 |
| 4 | Central | Malleswaram | 13.003100 | 77.564300 | 893629 | Bangalore | 53270.063892 |

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. [Here is the link](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:

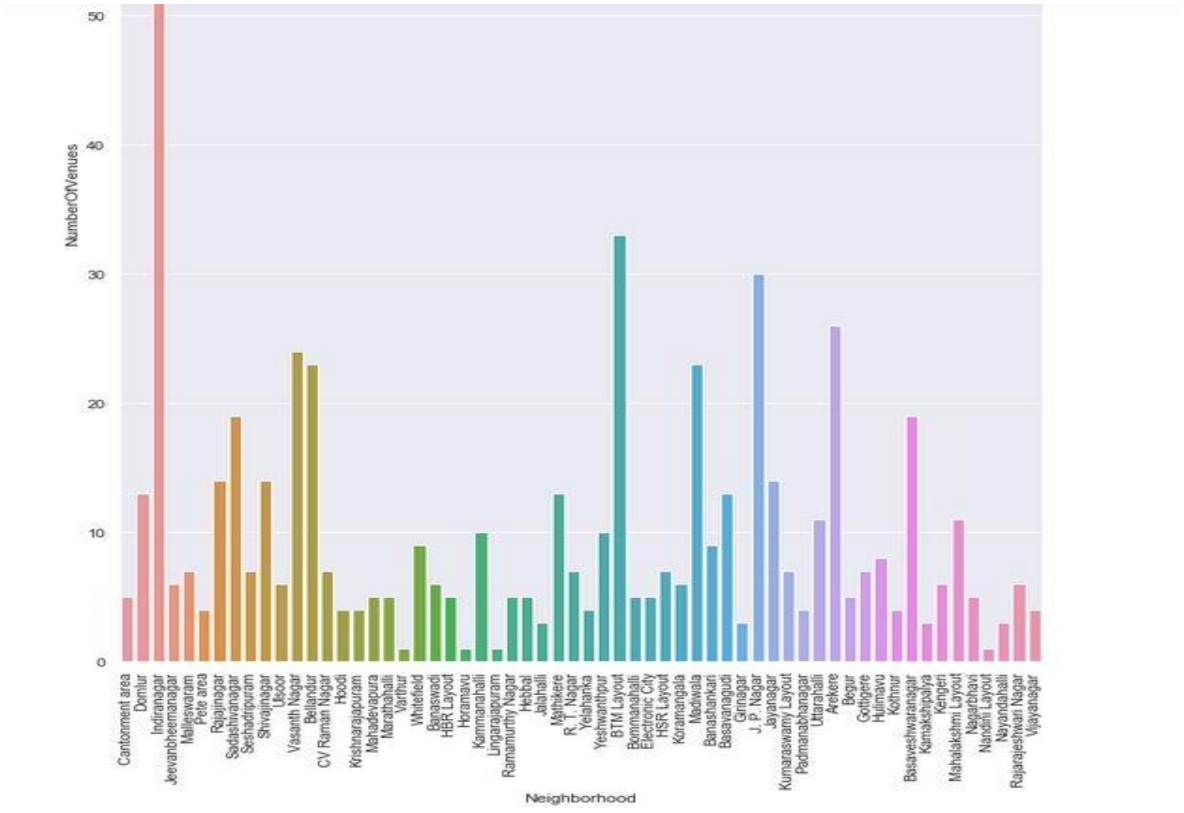
Utilize of foursquare is centered to bring closest setting areas so that we will utilize them to make a cluster. Foursquare API leverages the control of finding closest scenes in a sweep (in my case: 500mts) additionally comparing arranges, scene area and names. After calling, the taking after information outline is created:

| | Borough | Neighborhoods | Latitude | Longitude | Population | City | AverageIncome |
|---|---|---|---|---|---|---|---|
| **0** | Central | Cantonment area | 12.972442 | 77.580643 | 866377 | Bangalore | 18944.099792 |
| **1** | Central | Domlur | 12.960992 | 77.638726 | 743186 | Bangalore | 56837.022198 |
| **2** | Central | Indiranagar | 12.971891 | 77.641151 | 474289 | Bangalore | 41991.817435 |
| **3** | Central | Jeevanbheemanagar | 12.962900 | 77.659500 | 527874 | Bangalore | 6667.447632 |
| **4** | Central | Malleswaram | 13.003100 | 77.564300 | 893629 | Bangalore | 53270.063892 |

# 3. Methodology:

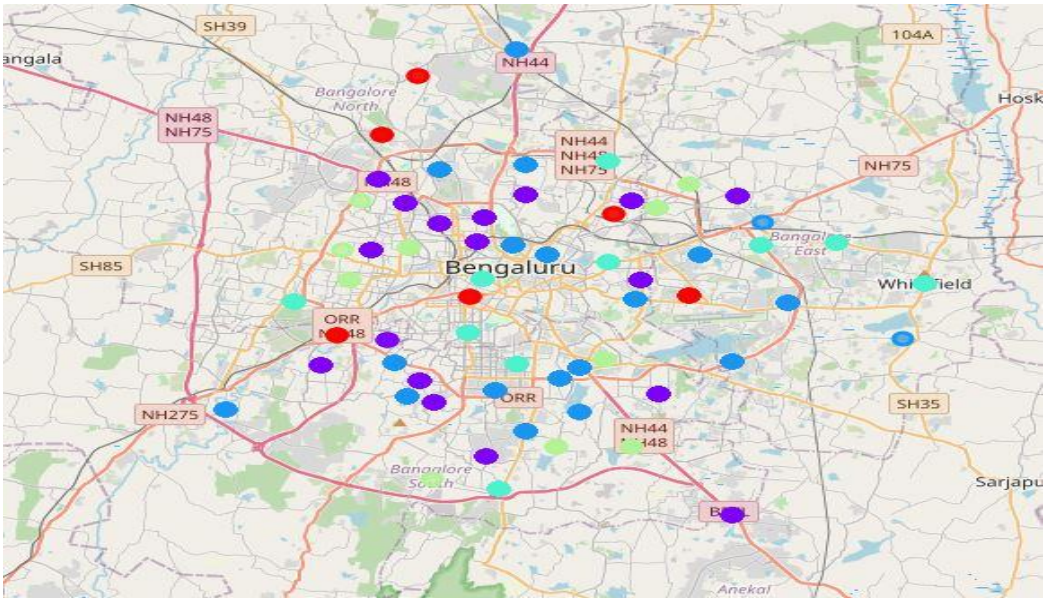
**Exploratory Investigation:**

Rejecting the information from distinctive sources and after that combining it to make a single-ton dataset may be a troublesome assignment. To do so, we got to investigate the current state of dataset and after that list up all the highlights required to be fetched. Exploring the dataset is vital since it gives you beginning bits of knowledge and may assist you to urge halfway thought of the answers that you just are looking to discover out from the data. While investigating the dataset, I found out that Indiranagar has most number of settings whereas Varthoor has the least.



Moreover whereas creating chart for number of cluster, I created a chart to investigate all the values for n clusters and after that finding the finest by investigating the elbow chart.

Most critical components whereas building the recommender framework were populace and pay. They are the foremost consequence calculate since they have a nonlinear relationship concurring to our dataset. It required to form a few inferential investigations to get it this nonlinear relationship. As the sum of populace increments, it does not fundamentally cruel that normal pay of a neighborhood will to increment. It is genuine to most of the case but moreover numerous cases vary to take after this drift. Essentially, a neighborhood with a smaller number of individuals may not essentially have less normal pay. It is conceivable to have a smaller number of individuals and more pay and bad habit versa. This may be induced from the taking after map




:

# OutCome:

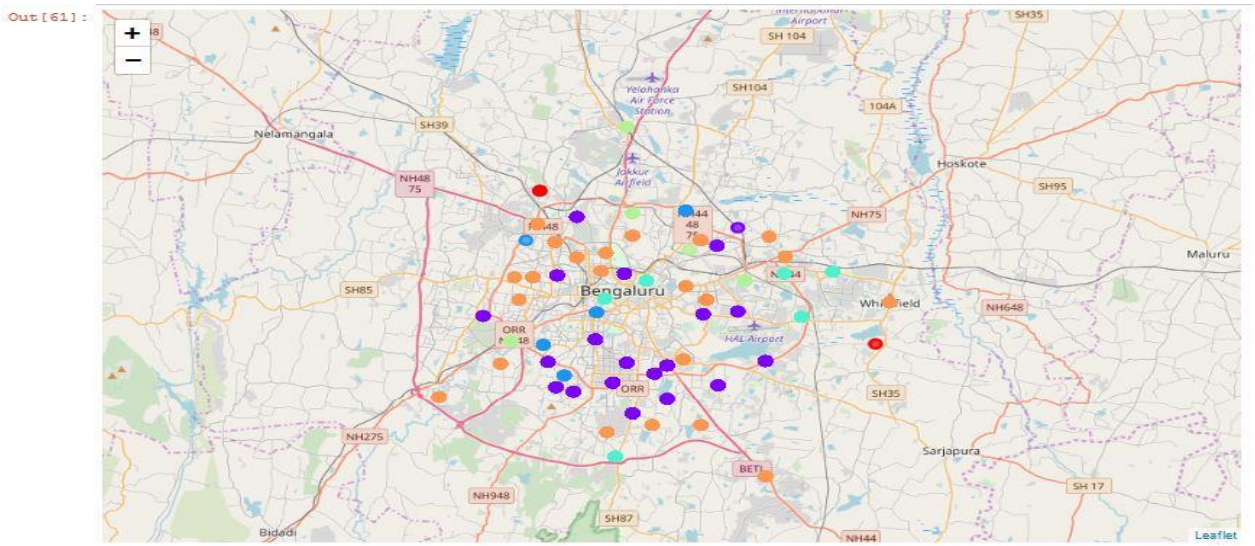
The result of the recommender framework is that it produces a list of beat eateries and the foremost common scene thing that the client can appreciate. Amid the runtime of the show, a recreation was done by taking 'Whitefield' as the neighborhood and after that prepared through our demonstrate so that it seem suggest neighborhoods with comparative characters as that of 'Whitefield'. The taking after picture appears the result

| | Neighborhoods | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | Ranking |
|---|---|---|---|---|---|
| 0 | Basavanagudi | Venue Category_Indian Restaurant | Venue Category_Metro Station | Venue Category_Restaurant | [0.6589806874118267] |
| 1 | Cantonment area | Venue Category_Indian Restaurant | Venue Category_Restaurant | Venue Category_Bookstore | [0.6429726634818886] |
| 2 | Gottigere | Venue Category_Indian Restaurant | Venue Category_Pharmacy | Venue Category_Department Store | [0.43352975051479636] |

# Dialog Section:

Since there was a nonlinear relationship between salary and populace, it can be concluded that we must continuously perform inferential approach to discover relationship among diverse set of highlights. Moreover, amid clustering, comparative neighborhoods must be dumped into the

correct cluster. The taking after chart appears the clusters correct cluster. The taking after chart appears the clusters



Another perception that we are able make is that choosing number of clustering seem deliver exceptionally assorted comes about. A few may be over fitted or a few may be beneath fitted. Consequently, examination of number of clusters must be done. Ref elbow graph within the Strategy segment.

# Conclusion:

The recommender framework could be a framework that considers variables such as populace, pay and makes utilize of Foursquare API to decide adjacent scenes. It may be a capable data driven show whose proficiency may diminish with more information but accuracy will increment. It'll offer assistance clients to wrap up their starvation by giving the leading proposal to satisfy all their needs.