# RESTAURANT RECOMMENDATION SYSTEM

# TABLE OF CONTENTS

# INTRODUCTION

## Problem Background:

Bangalore officially known as Bengaluru, is the capital of Indian state of Karnataka. It has a population of over ten million, making it a megacity and the third-most populous city and fifthmost populous urban agglomeration in India.

Bangalore is famous for different types of cuisines from different places which reflects social and economic diversity of the city. In Bangalore tea stalls, South Indian, North Indian, Chinese and Western fast food are all very popular in the city. Some restaurants are very popular and serve predominantly vegetarian cuisine. The Chinese food and the Thai food are famous and served in most of the restaurants. It is important to know what kind of food is available in a particular location, and to which cuisine a location or place is famous for.

## Business Problem:

There are so many people in Bangalore majorly from different places(non-Bangaloreans) stay in Bangalore having different cultures and likes different types of cuisines. Bangalore is also a tourist place where people from other places visit. Food is an important factor for everyone. Finding the right place, at reasonable cost, to serve us with the cuisine is very important. There are many other problems associated in recommending restaurants to people from different places 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?

**Problem Statement:** The problem here is to find the best restaurant nearby for the cuisine we like with restaurant having good rating.

## Target audience:

Target audiences for this Recommendation system does not limit to a person it is for everyone who search for particular type of cuisine in a particular location. 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 the recommendation system. So target for this project is basically everyone who loves different types of cuisines.

# DATA SECTION

## Data requirements:

The Data Requirements for the restaurants recommendation systems include list of restaurants in Bangalore. To access location of a restaurant, its 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.

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 it is very important factor.

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.

## Data Collection:

Collecting geographical coordinates is not difficult. It is possible to with Google maps API to fetch latitude and longitude. Initially scrape the list of neighbor's using beautifulSoup4. The table headings becoming the boroughs and data becoming the neighborhoods. Bangalore has 8 boroughs and 64 neighborhoods. Find its corresponding latitude and longitude. After doing that produced the following dataframe.

| Borough | Neighborhoods | Latitude | Longitude |
|---------|---------------|----------|-----------|
| Central | Cantonment area | 12.972442 | 77.580643 |
| Central | Domlur | 12.960992 | 77.638726 |
| Central | Indiranagar | 12.971891 | 77.641151 |
| Central | Jeevanbheemanagar | 12.962900 | 77.659500 |
| Central | Malleswaram | 13.003100 | 77.564300 |

Population by neighborhood is again easy to find out given that its readily available. But incase of bangalore, it is again not the case. It was able to find population data for few cities. 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 dataframe for bangalore neighborhood population looks like as following.

|   | Borough | Neighborhoods | Population |
|---|---------|---------------|------------|
| 0 | Central | Cantonment area | 866377 |
| 1 | Central | Domlur | 743186 |
| 2 | Central | Indiranagar | 474289 |
| 3 | Central | Jeevanbheemanagar | 527874 |
| 4 | Central | Malleswaram | 893629 |

Income by neighborhood is again easy to find out given that its readily available. But in case of bangalore, it is again not the case. It was able to find Income data for main city. Neighborhood Income is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The dataframe for bangalore neighborhood population looks like the following.

| | Borough | Neighborhoods | AverageIncome |
|---|---|---|---|
| 0 | Central | Cantonment area | 18944.099792 |
| 1 | Central | Domlur | 56837.022198 |
| 2 | Central | Indiranagar | 41991.817435 |
| 3 | Central | Jeevanbheemanagar | 6667.447632 |
| 4 | Central | Malleswaram | 53270.063892 |

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, and also corresponding coordinates, venue location and names. After calling, the following dataframe is created:

| | Neighborhood | Borough | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|---|
| 0 | Cantonment area | Central | 12.972442 | 77.580643 | Hotel Fishland | 12.975569 | 77.578592 | Seafood Restaurant |
| 1 | Cantonment area | Central | 12.972442 | 77.580643 | Vasudev Adigas | 12.973707 | 77.579257 | Indian Restaurant |
| 2 | Cantonment area | Central | 12.972442 | 77.580643 | Adigas Hotel | 12.973554 | 77.579161 | Restaurant |
| 3 | Cantonment area | Central | 12.972442 | 77.580643 | Sapna Book House | 12.976355 | 77.578461 | Bookstore |
| 4 | Cantonment area | Central | 12.972442 | 77.580643 | Kamat Yatrinivas | 12.975985 | 77.578125 | Indian Restaurant |

# METHODOLOGY

## Exploratory data 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 Indira nagar has most number of venues while Varthur has the least.
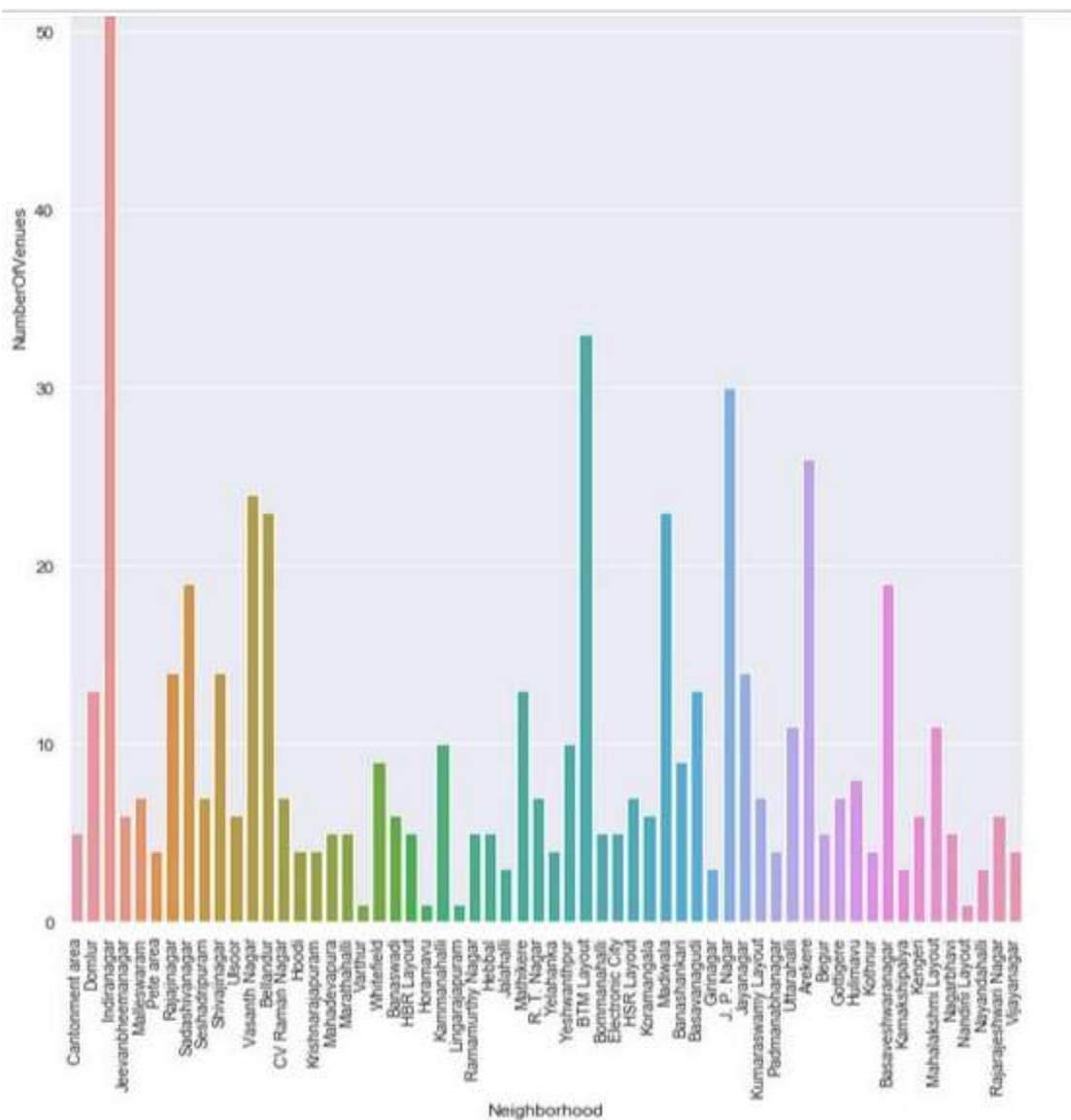
Also while producing graph for number of cluster, I produced a graph to explore all the values for n_clusters and then finding the best by exploring the elbow graph.
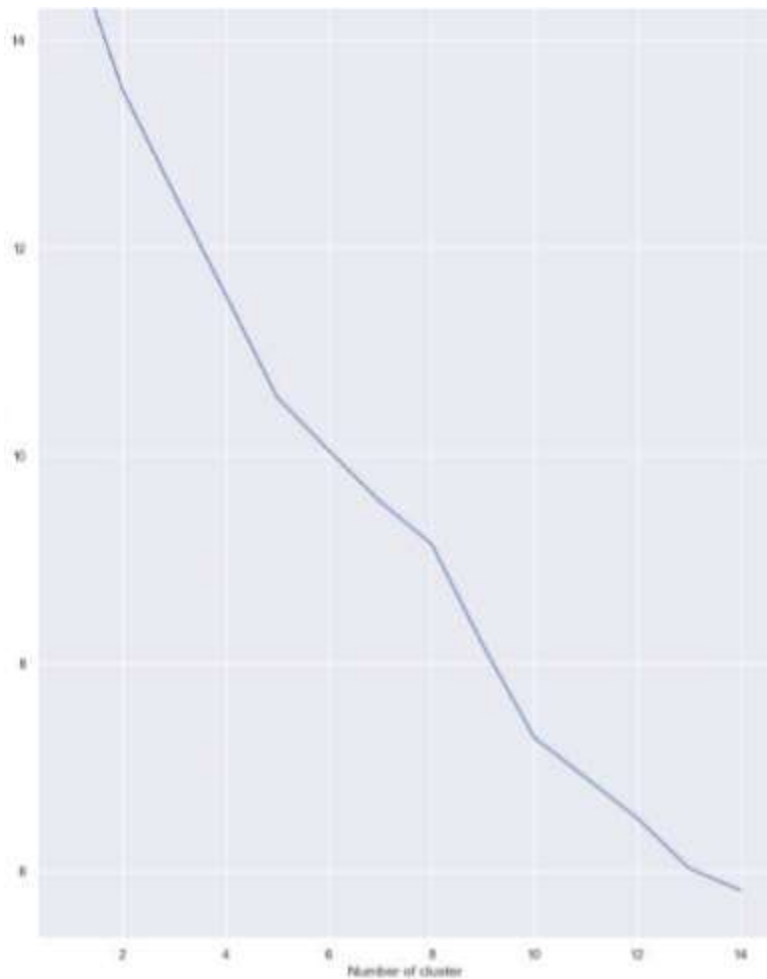


## Inferential analysis:

Most important factors while building the Restaurant recommendation 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. This can be inferred from the following graph.
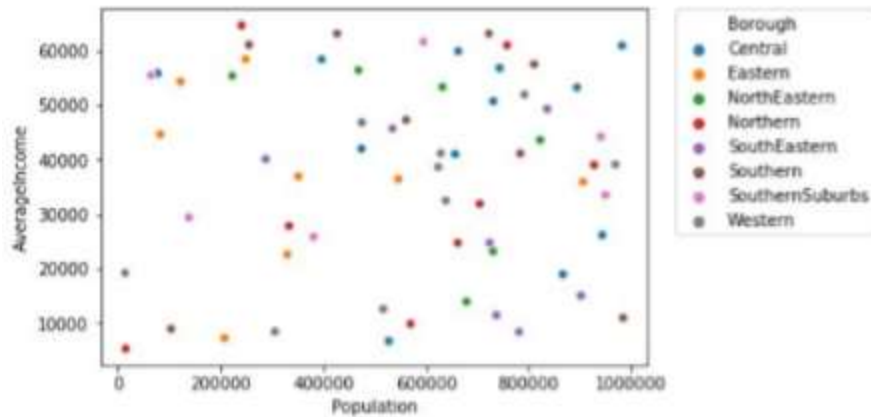
## Clustering:

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). Clustering is an unsupervised machine leaning algorithm which is used in the restaurant recommendation system. We used the clustering algorithm to group the restaurant which serve similar type of cuisines and recommend that to the people.
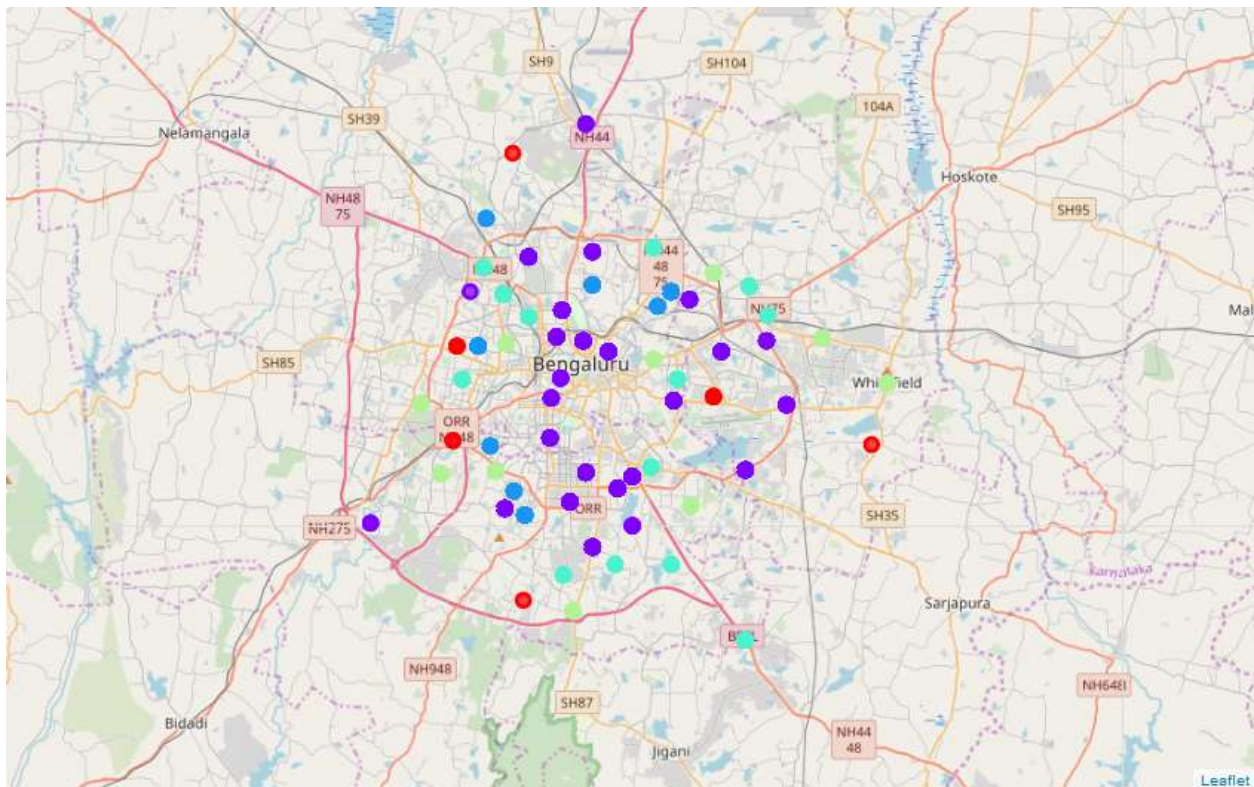
# RESULTS

The result of the Restaurant recommendation 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:

| | Neighborhoods | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | Ranking |
|---|---|---|---|---|---|
| 0 | Banashankari | Venue Category_Clothing Store | Venue Category_Café | Venue Category_Accessories Store | [0.8234029969357847] |
| 1 | Gottigere | Venue Category_Indian Restaurant | Venue Category_Pharmacy | Venue Category_Department Store | [0.43352975051479636] |
| 2 | HSR Layout | Venue Category_Badminton Court | Venue Category_Pizza Place | Venue Category_Café | [0.6193945016263558] |

# DISCUSSION

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.

# CONCLUSION

The Restaurant recommendation 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.