Restaurant Recommender System

IBM Capstone Project

30-07-2020

By- Prapti Luthra

Recommender Engines or Systems are among the most popular applications of data science today. They are used to predict the “rating” or “preference” that a user would give to an item.

Recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries).

As the World Wide Web continues to grow at an exponential rate, the size and complexity of many web sites grow along with it. For the users of these web sites, it becomes increasingly difficult and time-consuming to find the information they are looking for. User interfaces could help users find the information that is in accordance with their interests by personalizing a web site.

Restaurant recommender system is a machine learning model

The overview of this project is to recommend restaurants based on user’s likes and dislikes and his previous interest data.

Table of Contents:

Introduction:

Known as both the "Garden City" and "The Silicon Valley of India," Bangalore (officially "Bengaluru") is a techie's paradise, Bengaluru, is the capital of the Indian state of Karnataka. It has a population of about 10 million and a metropolitan population of about 8.52 million, making it the third most populous city and fifth most populous urban agglomeration in India. Located in southern India on the Deccan Plateau, at a height of over 900 m (3,000 ft) above sea level, Bangalore is known for its pleasant climate throughout the year.

With a population of 8,443,675 in the city and 10,456,000 in the urban agglomeration, up from 8.5 million at the 2011 census, Bangalore is a megacity, and the third-most-populous city in India and the 18th-most-populous 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.

Target Audience:

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.

Problem Statement:

Today we all travel to new places all the time. But going to a new place can sometimes become a burden, especially if you do not know where to eat. Food is such a basic part of our lives that sometimes we forget how important it is to find the right connect.
In such situation, food can be an important factor for decided how you rate your trips and recommending it to people. Food can also attract people around to world to try it out if it is the best or has been made famous in the past. 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:

1. What type of foods is available in the restaurant and is their variety?
2. Which place has the best rating near me?
3. How many "similar" restaurants are available near me of my previous tastes and liking?
4. What speciality do these restaurants have as each place tends to have some?

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. Some of these questions are:

1. The types of restaurants in specific areas?
2. Which area can I target to find my “liking” restaurants?
3. Ranking the different restaurants in my preferences?

Data requirements:

To find a solution to these questions and build a recommender model, we need the right data. Data can answer question which are unimaginable and non-answerable by us. Having the right and clean data at our fingertips will help solve the problems that we have discussed earlier on.

Let's consider the current situation:

When we want to find a restaurant, we 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 is the restaurant worth.

Now, let’s take a closer look at each of these:

1. 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.
2. The walk-in customers determine the rating of the restaurant and the more the customers, the more welcoming it becomes to other people who might enter seeing a good crowd. Population of a neighbourhood is very important factor in determining a restaurant's growth and number of customers who visit the restaurant. Hence is a very important factor to be considered.
3. Income is in direct proportion to earning capacity 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:

Data is collected through various online sources like Wikipedia, Indian government website and has been complied into a data frame that looks like below:

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
Central	Pete area	12.962700	77.575800
Central	Rajajinagar	12.990100	77.552500
Central	Sadashivanagar	13.006800	77.581300
Central	Seshadripuram	12.993500	77.578700
Central	Shivajinagar	12.985700	77.605700

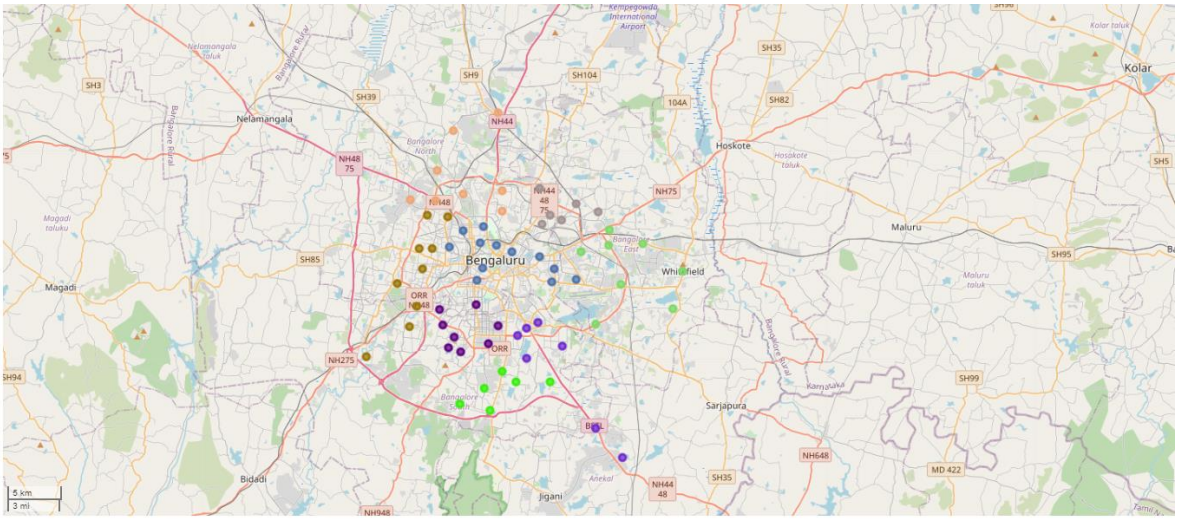
The population of Bengaluru is shown as the below data frame:

	Borough	Neighborhoods	Population	Normalized_population
0	Central	Cantonment area	866377	0.880810
1	Central	Domlur	743186	0.755567
2	Central	Indiranagar	474289	0.482190
3	Central	Jeevanbheemanagar	527874	0.536668
4	Central	Malleswaram	893629	0.908516

The income of Bengaluru is shown as the below data frame:

	Borough	Neighborhoods	AverageIncome	Normalized_income
0	Central	Cantonment area	18944.099792	0.293051
1	Central	Domlur	56837.022198	0.879225
2	Central	Indiranagar	41991.817435	0.649581
3	Central	Jeevanbheemanagar	6667.447632	0.103140
4	Central	Malleswaram	53270.063892	0.824047

The following map is exploring Bangalore Neighbours using Four Square API. As it is seen the different regions of Bangalore are colour coded and that gives a clear picture of the different areas.



Exploring bangalore Neighborhoods using FourSquare API

Methodology:

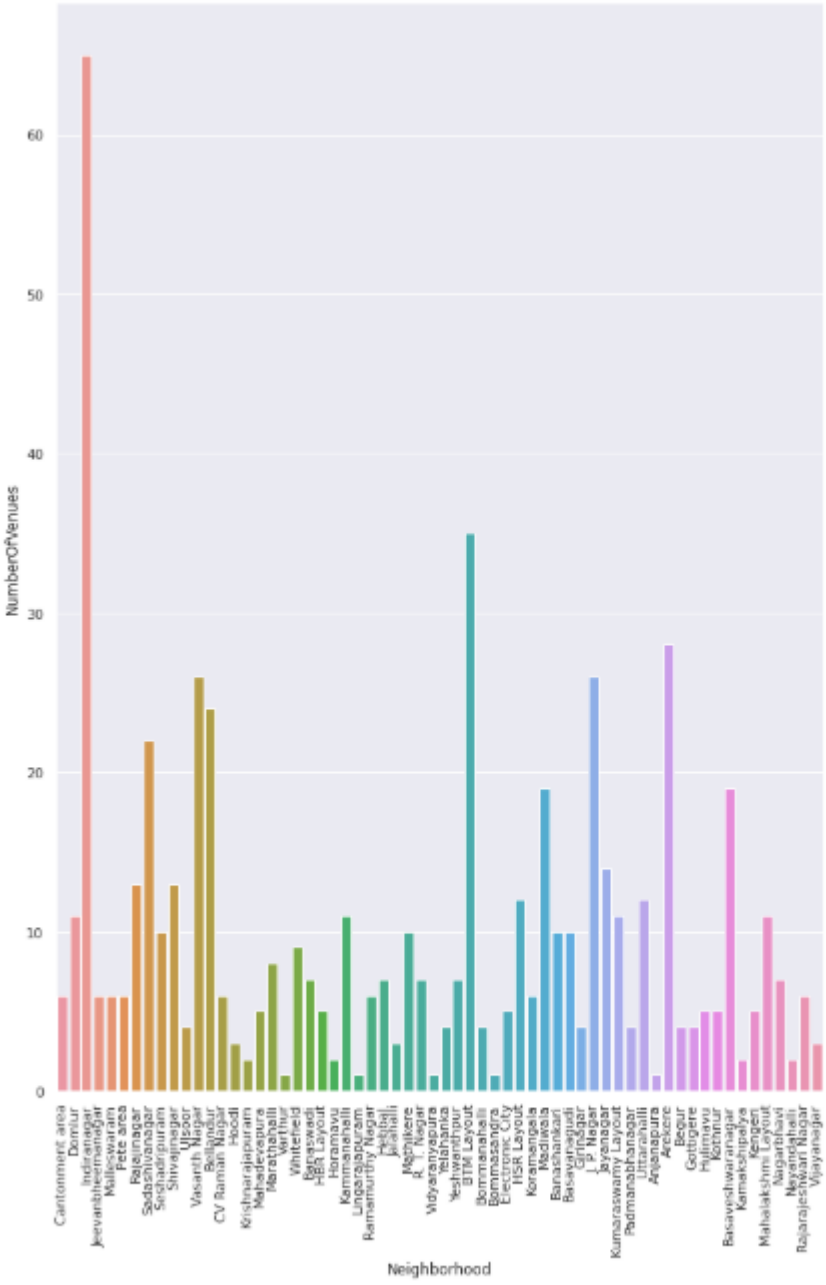
Exploratory analysis:

In statistics, exploratory data analysis (EDA) is an approach to analysing data sets to summarize their main characteristics, often with visual methods.

The purpose of exploratory analysis is to "get to know" the dataset.

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 Indiranagar has the greatest number of venues while Varthur has the least.

This graph shows us the number of venues in each neighbourhood in a simple yet impactful manner.



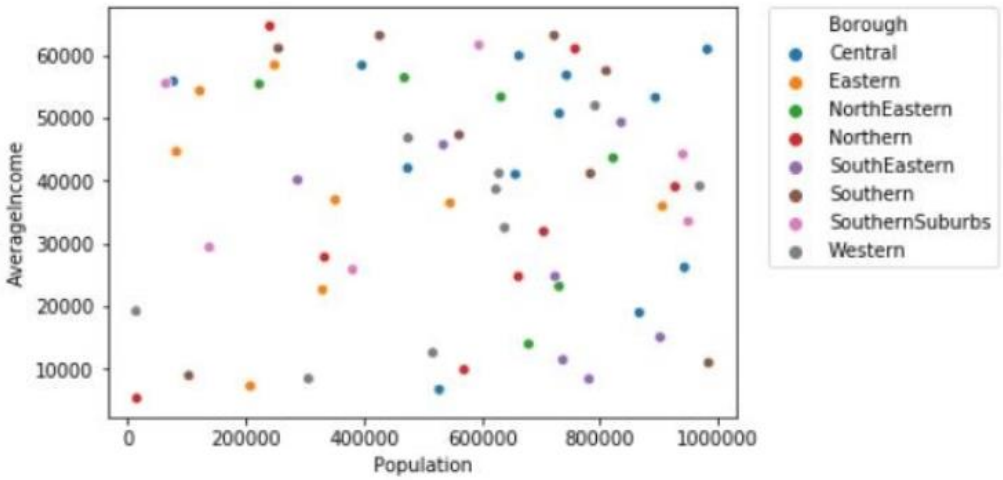
From the above graph we can see that inderanagar has most number of venues followed by BTM layout and further more

Inferential analysis:

Inferential statistical analysis involves objectively and quantitatively summarizing the data, determining which data patterns are significant, and making inferential statements about system performance.

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 is

possible to have a smaller number of people and more income and vice versa. This can be inferred from the following graph:

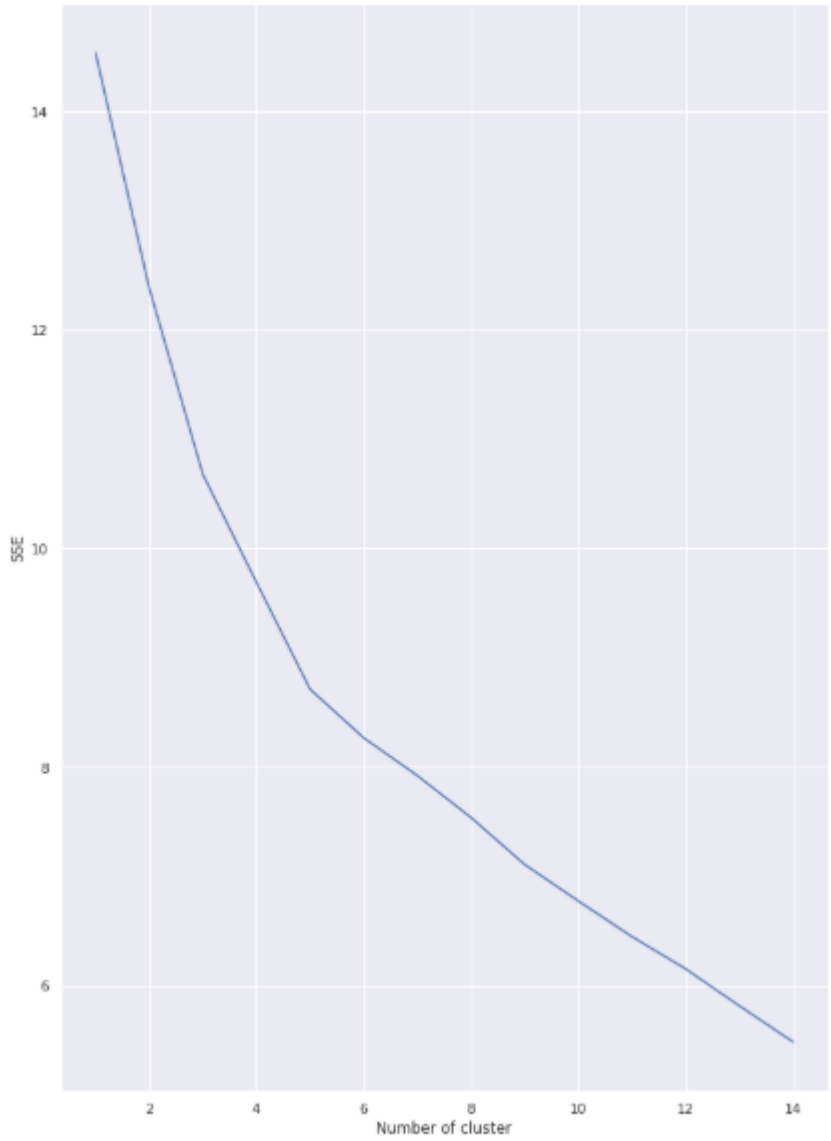


The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy.

Elbow graph:

In cluster analysis, the elbow method is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters, and picking the elbow of the curve as the number of clusters to use.

Using the "elbow" or "knee of a curve" as a cut-off point is a common heuristic in mathematical optimization to choose a point where diminishing returns are no longer worth the additional cost. In clustering, this means one should choose several clusters so that adding another cluster doesn't give much better modelling of the data.




```
From the above graph, we can see the optimal value for cluster is 5.
```

In our case, the optimal value of “k” or the number of clusters is 5.

Result:

The result of the recommendation system is that it produces a list of top 3 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 neighbourhood and then processed through our model so that it can recommend neighbourhoods with similar characters as that of “Whitefield”.

The following image shows the result:

[91]:

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Ranking
0	Begur	Venue Category_ATM	Venue Category_Clothing Store	Venue Category_Supermarket	[0.7361321887298261]
1	Electronic City	Venue Category_Outlet Store	Venue Category_Furniture / Home Store	Venue Category_Auto Garage	[0.5423513638776338]
2	Kamakshipalya	Venue Category_ATM	Venue Category_South Indian Restaurant	Venue Category_Hyderabadi Restaurant	[0.8041873599273973]

Here, according the data, we see that 'Begur,Electronic City and kamakshipalya' are top 3 neighborhoods to find similar food to whitefield restaurant's veg food ! Here our model will recommend these neighborneeds and top 3 common venues to visit.

The recommendation system is a system that consider factors like population, income and makes use of Foursquare API to determine nearby venues.

It is a powerful data driver 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.