

Restaurant recommender system

2020

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.

IBM Capstone
project

Table of contents

Topic	Page number
Introduction section	2
Data section	4
Methodology section	7
Result section	10
Discussion section	11
Conclusion section	12

1. Introduction :

Problem background:

Guwahati (formerly known as Gauhati) is the largest city in the Indian state of Assam and also the largest metropolis in northeastern India. A major riverine port city along with hills, and one of the fastest growing cities in India, Guwahati is situated on the south bank of the Brahmaputra. It is called the 'Gateway to North East India'.

Guwahati is where much of Assam's city glitter resides. So, while most of the other towns in the state will offer you local preparations in local food joints and a few restaurants, Guwahati sparkles with choices only a commercially developed city can offer. Guwahati sustains the local Assamese foods in its menu such as Nga-Thongba which is a fish and ginger curry, Erongba, a jackfruit-kernel chutney, Laksa and a lot more, and at the same time offers everything from Indian, Mughlai, Chinese to Continental cuisines and a few more. The number of restaurants available here may be less compared to other big cities but the variety in the dishes is no less here.

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 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 near by me ? 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 ?

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.

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.

2. Data :

Data requirements:

To find a solution to the questions and build a recommender model, we need data for the same. Data can answer questions which are unimaginable and non-answerable by humans because humans do not have the tendency to analyse such large datasets and produce analytics to find solutions.

We would require 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.

Now let's take a look how these data would help in this project:

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. Population of a neighborhood is a 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 a very important factor.
3. Income of a neighborhood is also a very important factor as population was. Income is directly proportional to richness of a neighborhood. If people in a neighborhood earn 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. The neighborhood names with their geographical coordinates were not available in the internet directly. I got the names of the neighborhoods from the Guwahati Municipal Council, Govt. of Assam website. These neighborhoods are divided into 6 boroughs. The latitudes and longitudes of the neighborhoods were not directly available so I searched each neighborhood's coordinates in Google individually for each neighborhood. The following dataframe was created:-

	Borough	Neighborhoods	Latitude	Longitude	Population	City	AverageIncome
0	West	Jalukbari	26.1426	91.6435	20555	Guwahati	26872
1	West	Ferryghat Colony	26.1859	91.7216	72045	Guwahati	10123
2	West	Maligaon	26.1504	91.6960	26456	Guwahati	32343
3	West	Adabari	26.1686	91.6779	21421	Guwahati	29123
4	West	Kamakhya Town	26.1659	91.7074	18623	Guwahati	30345

2. The Population by neighborhoods were again not directly available in the net. The govt of assam website had a map of guwahati displaying the population of different wards of guwahati in range. So i matched those wards with the neighborhoods and assumed the population of the neighborhoods from that given range.

	Borough	Neighborhoods	Population
0	West	Jalukbari	20555
1	West	Ferryghat Colony	72045
2	West	Maligaon	26456
3	West	Adabari	21421
4	West	Kamakhya Town	18623

3. In case of guwahati, again the income by neighborhoods were not directly available. I 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.

	Borough	Neighborhoods	AverageIncome
0	West	Jalukbari	26872
1	West	Ferryghat Colony	10123
2	West	Maligaon	32343
3	West	Adabari	29123
4	West	Kamakhya Town	30345

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. Using all these a map of Guwahati was created.

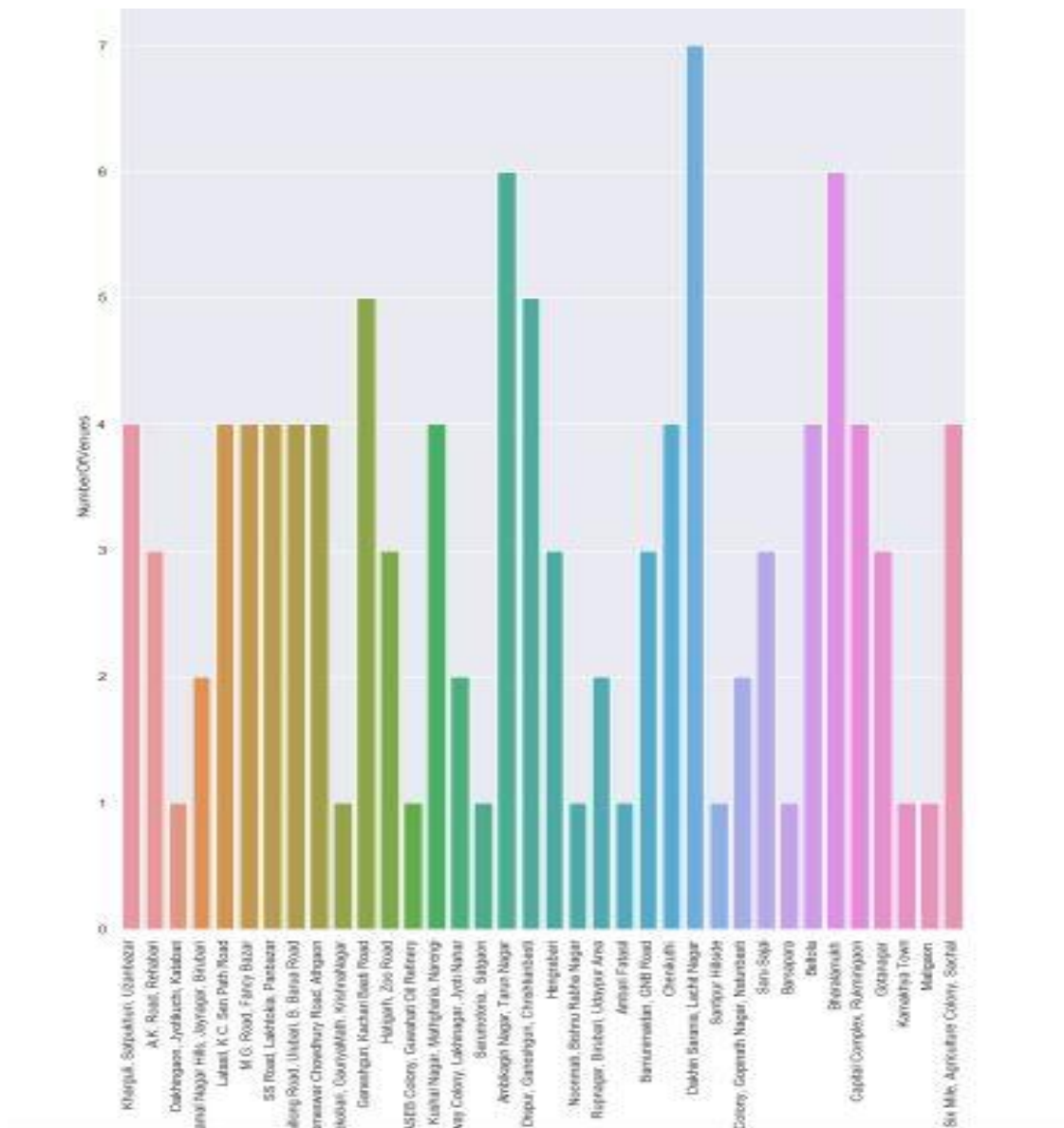
	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Maligaon	West	26.1504	91.6960	Goshala Bazaar Maligaon	26.150406	91.695621	Flea Market
1	Kamakhya Town	West	26.1659	91.7074	Kamakhya View Point	26.162372	91.705112	Scenic Lookout
2	Gotanagar	West	26.1420	91.6773	Radisson Blu	26.141747	91.673764	Hotel
3	Gotanagar	West	26.1420	91.6773	Hemalata Dhaba	26.140271	91.675350	Diner
4	Gotanagar	West	26.1420	91.6773	Cafe b you	26.141835	91.673881	Buffet

3. Methodology :

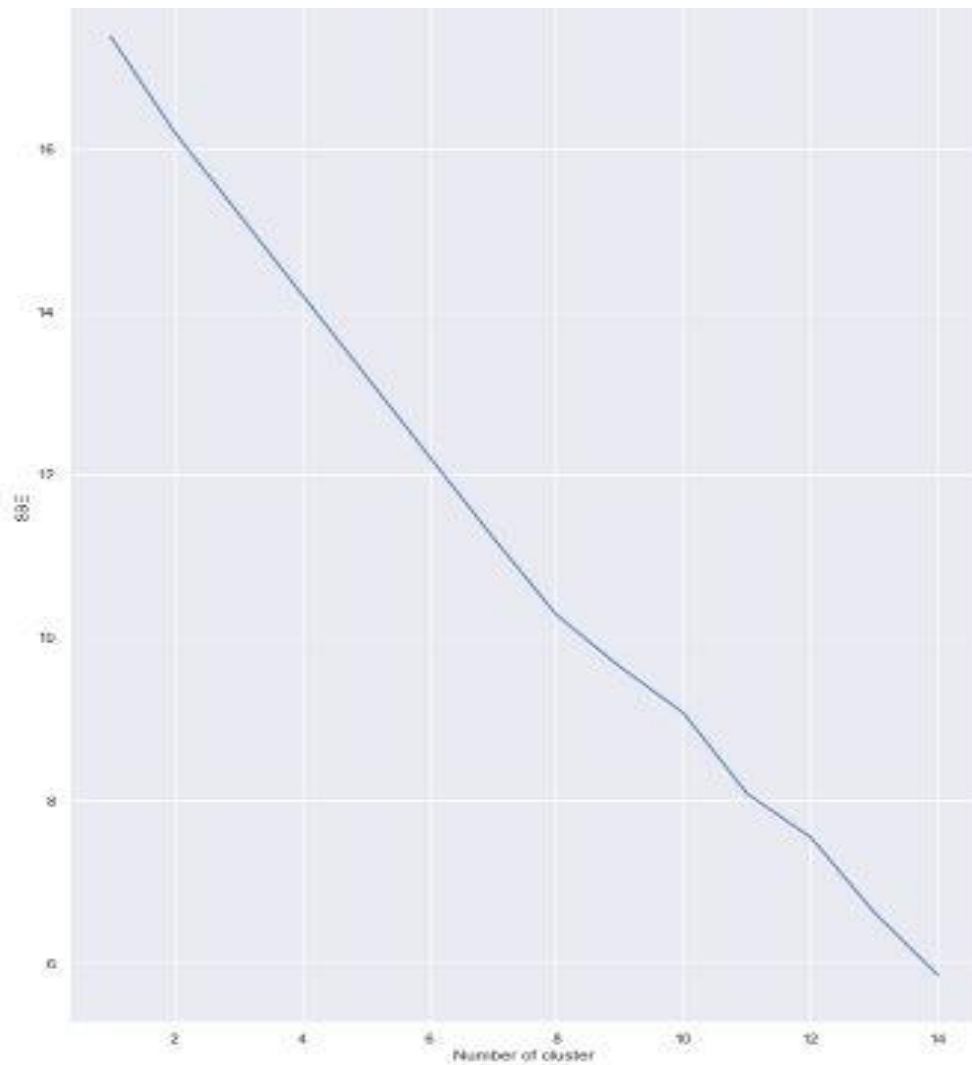
Exploratory analysis:

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 Lachit Nagar has most number of venues followed by Ambikagiri Nagar and Bharalamukh.



Below is 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 important 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 ‘Dispur, Ganeshguri, Chrishtianbasti’ as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of ‘Dispur, Ganeshguri, Chrishtianbasti’.

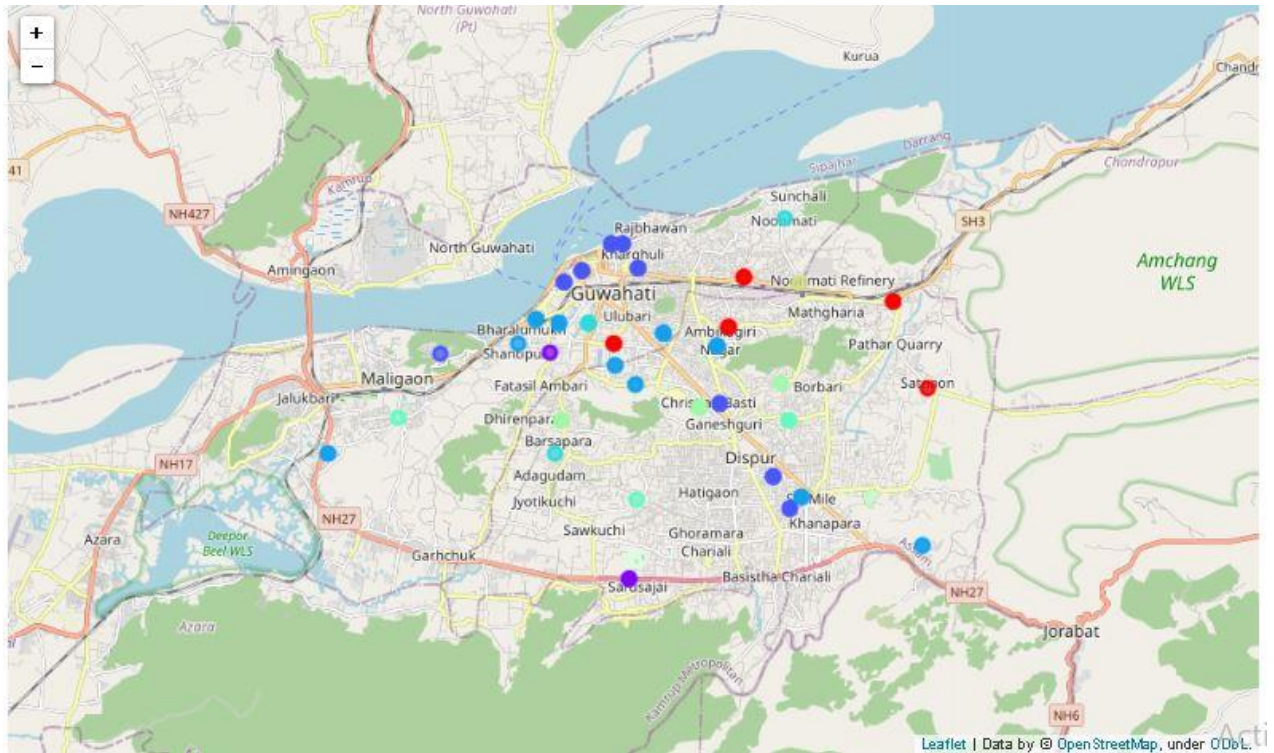
The following image shows the result:

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Ranking
0	Bishnupur Colony, Gopinath Nagar, Natunbasti	Venue Category_Hotel	Venue Category_Clothing Store	Venue Category_Zoo	[0.3206676818096672]
1	Dispur, Ganeshguri, Chrishtianbasti	Venue Category_Clothing Store	Venue Category_Shopping Mall	Venue Category_Fast Food Restaurant	[0.3733661012747169]
2	Tokobari, GauriyaMath, KrishnaNagar	Venue Category_Clothing Store	Venue Category_Zoo	Venue Category_Diner	[0.24348350807199406]

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.