EATERY RECOMMENDER FRAMEWORK IN BANGALORE

Author : Raunak Kumar Singh

27/05/2021

IBM Coursera Capstone project

Eatery recommender framework is an AI model, created to exhibit as a capstone undertaking to IBM through coursera. It suggests cafés dependent on client's preferences and his past interest information.

Table of contents

1. Introduction :

Problem background:

Bangalore is the capital and largest city of the Indian state of Karnataka. With a population of over 12 million (as of May 2021), Bangalore is the third largest city in India and 27th largest city in the world.

The variety of the food accessible is intelligent of the social and financial variety of Bangalore. Side of the road merchants, tea slows down, South Indian, North Indian, Muslim food, Chinese and Western cheap food are generally mainstream around there. Udupi eateries, are extremely famous and serve transcendently vegan food. The Chinese food and the Thai food served in a large portion of the cafés are can be modified to take into account the inclinations of the Indian populace. Bangalore can likewise be known as a foodie's heaven on account of its huge assortment of food sources and edibles with a dash of Bangalore's uniqueness and custom.

Problem description:

Assume I travel and continue to change puts habitually. This is exceptionally chaotic and in addition to I will encounter altogether different sorts of climate, of which I don't have a lot of information about. In such circumstance, food can be a significant factor for chose how you rate your excursions and in addition to likewise prescribing it to individuals. Food can likewise draw in individuals around to world to give it a shot if it somehow managed to be the awesome. In such situations, we need to track down the correct spot, at sensible expense, to serve us the most ideal way. So there are not many inquiries that should be tended to, for example,

1. How many types of foods are available in the restaurant?
2. Which is the most nearest to me with good rating?
3. How many "similar" restaurants are available nearby me?
4. Do the "similar" restaurants cost more? If so, what specialty do that have?

To resolve such inquiry, ABCD Company's director chooses to assign this task to me not simply to discover answers for the inquiries yet in addition construct a framework that can help in suggesting new places dependent on their rankings contrasted with the recently 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 particular 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 crowds for this undertaking doesn't restrict to an individual who continues to travel yet everybody. Individuals could essentially choose to search for a comparable café all the time since they are dependent on a particular class of food. Individuals who once in a while use eateries would like to have the most appraised cafés close by them and this could be handily taken care of by our recommender framework. So focus for this undertaking is essentially every individual who is investigating better places or comparative spots.

Success rate:

With eateries developing, new food classifications arise, cross breed food begins to be more mainstream, we need a framework that could help us access tremendous number of food assortments. It is outlandish for an individual to get some information about their visit to a specific spot and furthermore not every person remembers everything. Then again, Computers are acceptable at recalling things, and with Machine figuring out how to its pinnacle, it high time innovation will by our own direction and help us by and by dependent on our preferences. So individuals would think often about this venture as their own help and achievement rate could unquestionably increment with time.

2. 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.

Let's take a closer look at each of these:
1. To access location of a restaurant, it's 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 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 is very important factor.
3. Pay of an area is additionally vital factor as populace was. Pay is straightforwardly relative to lavishness of an area. Assuming individuals in an area acquires in excess of a normal pay, it is a lot of conceivable that they will spend all the more anyway not in every case valid with less likelihood. So an eatery evaluation is relative to pay of an area.

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.

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. Bangalore has 8 boroughs and 64 neighborhoods. So i manually googled each neighborhood to find its corresponding latitude and longitude. After doing so, I produced the following data frame.

[92]:

	Unnamed: 0	Borough	Neighborhoods	Latitude	Longitude	Population	City	AverageIncome
0	0	Central	Cantonment area	12.972442	77.580643	866377	Bangalore	18944.099792
1	1	Central	Domlur	12.960992	77.638726	743186	Bangalore	56837.022198
2	2	Central	Indiranagar	12.971891	77.641151	474289	Bangalore	41991.817435
3	3	Central	Jeevanbheemanagar	12.962900	77.659500	527874	Bangalore	6667.447632
4	4	Central	Malleswaram	13.003100	77.564300	893629	Bangalore	53270.063892
5	5	Central	Pete area	12.962700	77.575800	730999	Bangalore	50712.430215
6	6	Central	Rajajinagar	12.990100	77.552500	981362	Bangalore	60967.535874
7	7	Central	Sadashivanagar	13.006800	77.581300	662625	Bangalore	59943.541564
8	8	Central	Seshadripuram	12.993500	77.578700	396862	Bangalore	58407.090338
9	9	Central	Shivajinagar	12.985700	77.605700	77836	Bangalore	55850.962099

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:

	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

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:

Out[95]:

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

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. After calling, the following data frame is created:

Out[29]:

	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

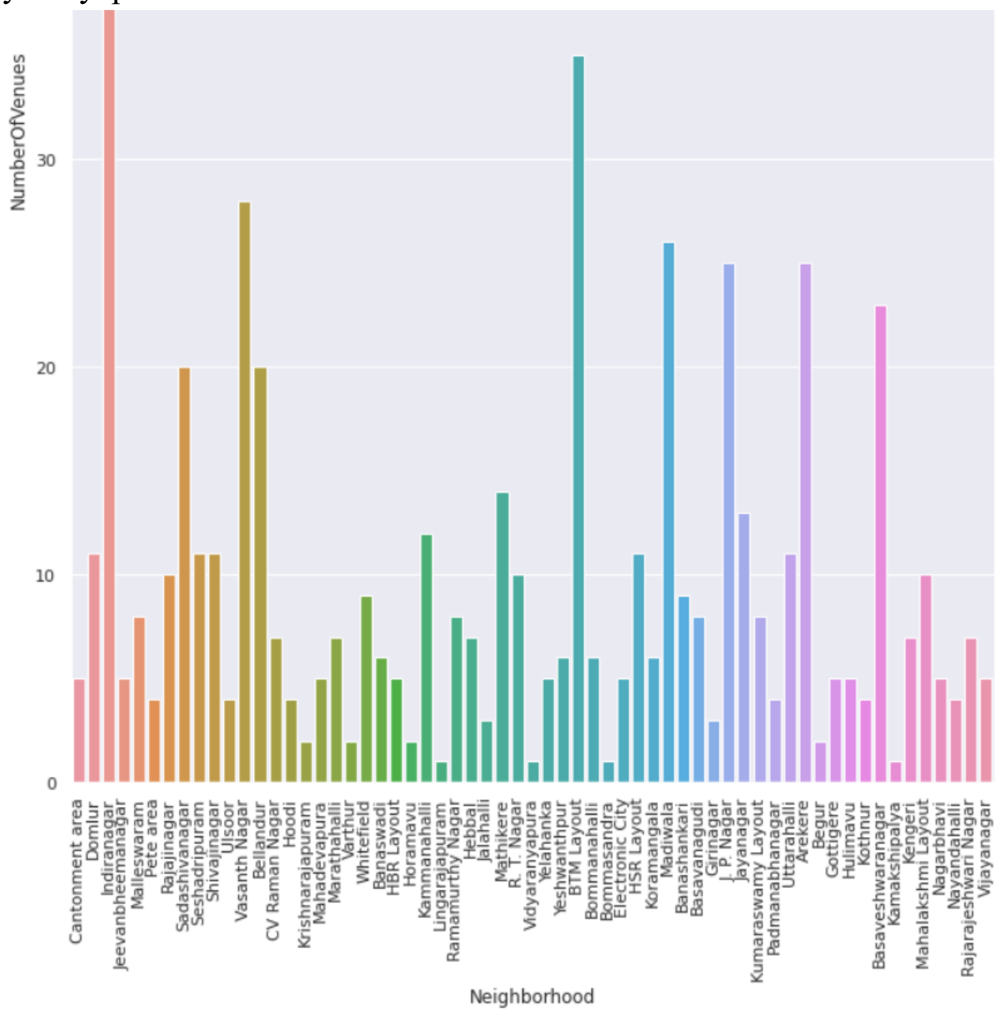
3. Methodology :

Exploratory analysis:

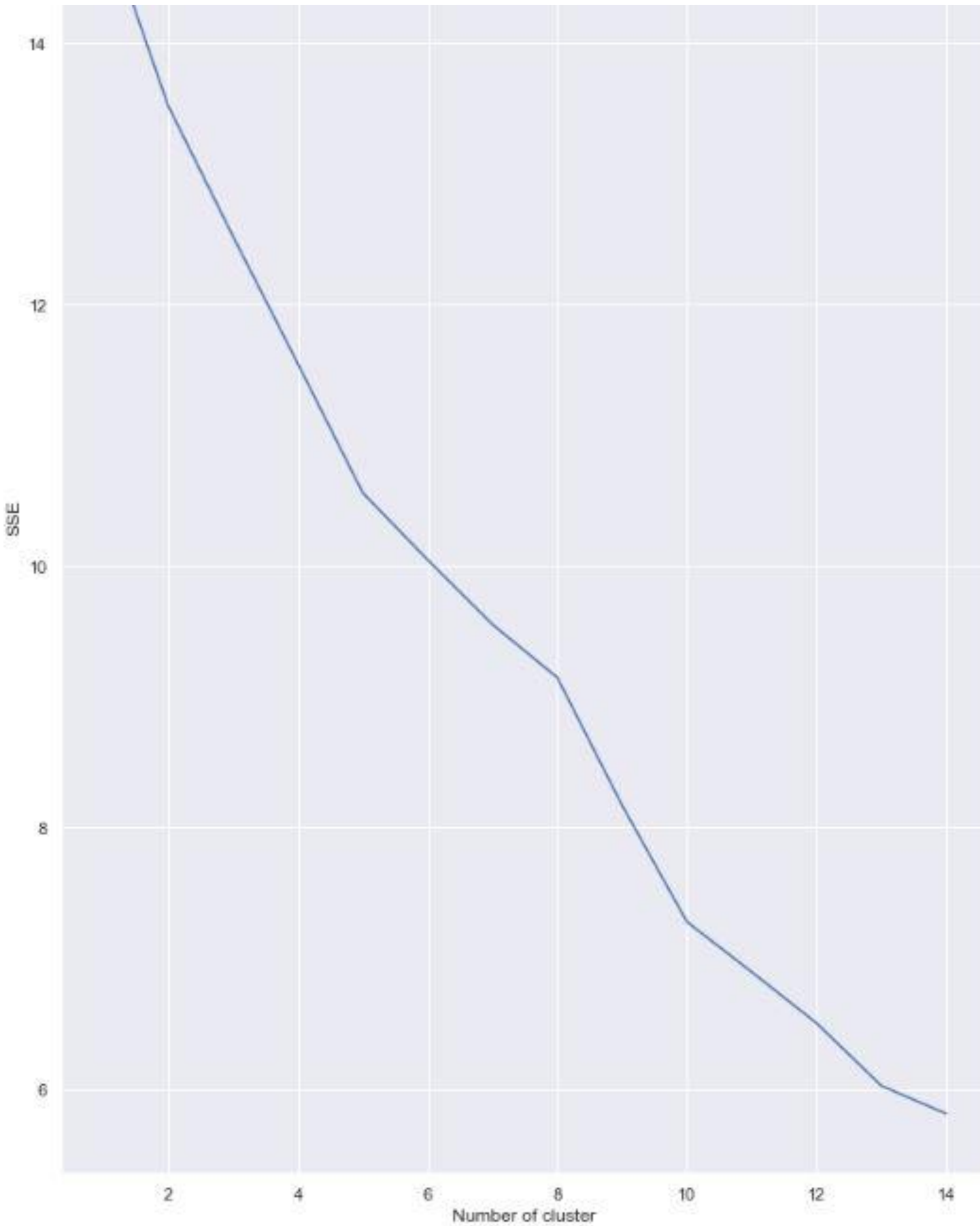
Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched.

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 Domlur has most number of venues while Vidyaranyapura 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.

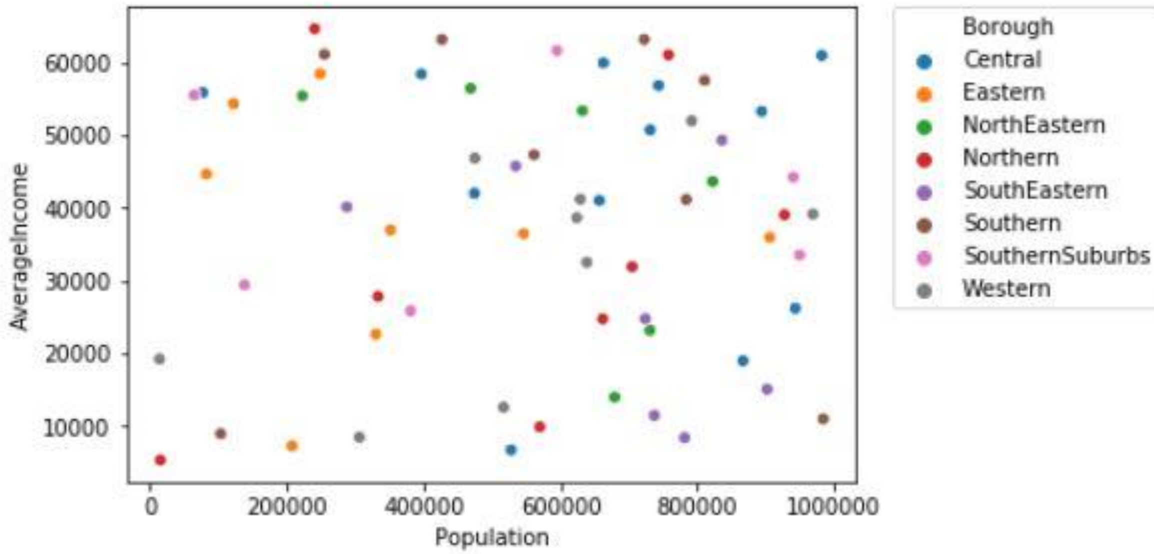


Inferential analysis:

Most significant elements while building the recommender framework were populace and pay. They are the most import factor since they have a nonlinear relationship as per our dataset.

It expected to make some inferential investigation to comprehend this nonlinear relationship. As the measure of populace builds, it doesn't really imply that normal pay of a local will likewise increment. It is consistent with the vast majority of the case yet in addition numerous cases vary to follow this pattern. Essentially, a neighborhood with less number of individuals may not really have less normal pay. It is feasible to have less number of individuals and more pay and the other way around.

This can be inferred from the following graph:



4. Result :

The consequence of the recommender framework is that it delivers a rundown of top cafés and the most well-known scene thing that the client can appreciate. During the runtime of the model, a recreation was finished by accepting 'Whitefield' as the area and afterward handled through our model so it could suggest neighborhoods with comparable characters as that of 'Whitefield'.

The following image shows the result:

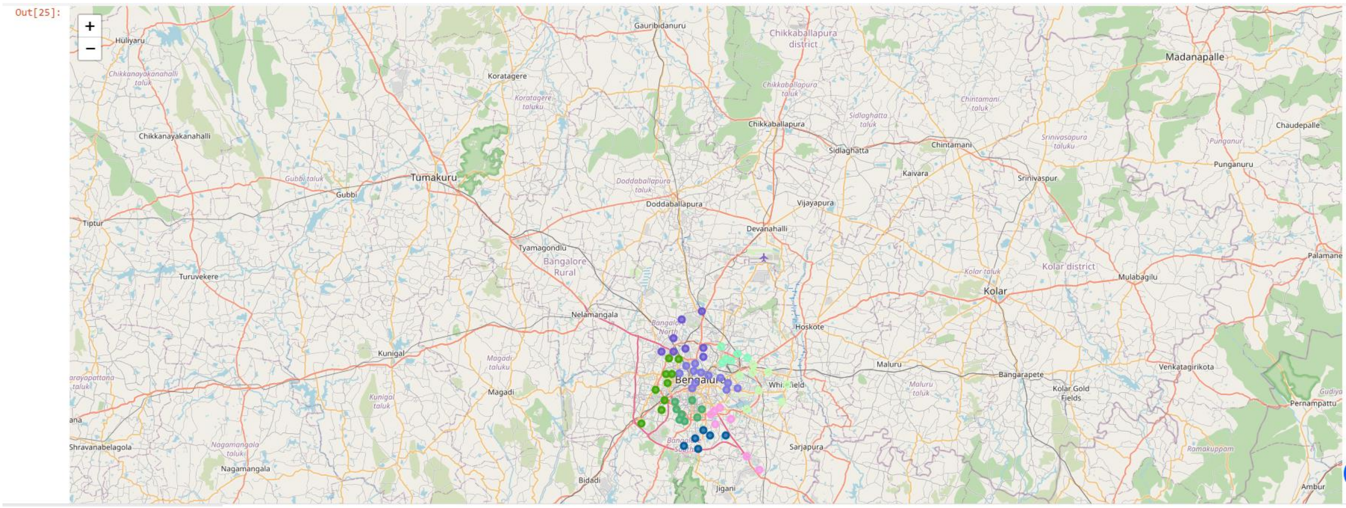
Out[89]:

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Ranking
0	Arekere	Venue Category_Sporting Goods Shop	Venue Category_Indian Restaurant	Venue Category_Pizza Place	[0.32959888840700646]
1	Electronic City	Venue Category_Furniture / Home Store	Venue Category_Bus Stop	Venue Category_Outlet Store	[0.5423513638809381]
2	HBR Layout	Venue Category_Road	Venue Category_Café	Venue Category_Coffee Shop	[0.7540959810582033]

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 framework is a framework that considers factors like populace, pay and utilizes Foursquare API to decide close by scenes. It is an amazing information driven model whose proficiency may diminish with more information however exactness will increment. It will assist clients with completing their craving by giving the best suggestion to satisfy every one of their necessities.