

RESTAURANT SUGGESTION SYSTEM

14/11/2020

IBM CAPSTONE PROJECT

Restaurant suggestion system is an ML model developed to demonstrate as a capstone project to IBM via Coursera.

It recommends restaurants based on user preferences.

TABLE OF CONTENTS

TOPIC	PAGE
Introduction section	3
Data section	5
Methodology	8
Result section	10
Conclusion section	10

1. Introduction :

Problem background:

Chennai formerly known as Madras, is the capital of the Indian state of Tamil Nadu. Located on the Coromandel Coast of the Bay of Bengal, it is one of the largest cultural, economic and educational centres of south India. According to the 2011 Indian census, it is the sixth-most populous city and fourth-most populous urban agglomeration in India.

The first thing that comes to our mind while talking about Chennai cuisine is the early morning special 'Filter Coffee' and the all-time favourite 'Thayir Sadam' (curd rice). Other than that, the city has a variety of cuisine to offer food lovers – from the typical Idly-Sambhar to the spicy Chettinad dishes. While the authentic multi-cuisine restaurants are famous for their specific delicacies, carts selling steaming hot Idly-Chutney-Sambhar to office goers, small stalls selling Poli and Mini-Samosa at street corners and the Sundal, Bajji and Chat stalls at the beaches still remain the people's choice.

Now most of the people in Chennai can experience a variety of cuisine like South Indian, North Indian, Muslim food, Chinese and Western fast food, etc which are a few of the very popular ones.

Problem description:

As a person who loves to travel and visit so many places, one might particularly be interested in that place's cuisines too. Food and Travel are the best combination to ensure a happy and satisfied trip. If a person he/she loved the food at a place they're more likely to suggest that place to friend, colleague or another person which benefits the customer by giving them satisfaction and the profits to the restaurant. In such scenarios, we need to find restaurants at place and reasonable cost.

Few important points that must be addressed:

1. Nearest restaurants having the best feedbacks and reviews.
2. The different types of foods are available in the restaurant.
3. How many similar restaurants are available near me?
4. Comparing similar restaurants by price and their uniqueness/speciality.

Expectations from this system is to get answers for the previous requirements in such a way that it handles all the work of suggesting restaurants based on certain preferences like:

Types of restaurants present in a particular area. Location/Address of the similar restaurants present based on preference to particular food. How do different restaurants rank with respect to my preference.

Target Audience

The major target audience for this project would absolutely anybody who are willing to try out new food, explore or you can say foodies who love food. It can also benefit people who are looking forward to try a particular type of food, and there are few people who like to eat at restaurants based on their rank or stars. Hence, we can conclude this project is for everyone who are exploring different places or similar places looking for restaurants.

Success Rate

There are many people who love to travel and take a break from their stressed routine lives, and the best option is travelling and having the best food. With the generation of technology beside us now, one can easily get his/her smartphone to ask suggestions about where to go for a meal. Computers can store data for long and with the developement of ML and AI algorithms data keeps updating and evaluates to much better results as days keep passing by.

2. Data:

Data requirements :

For finding the requirements for this project model we will primarily need a lot of data classified into :

1. geographical coordinates of the restaurant.
2. Population of the neighborhood where the restaurant is located.
3. Average income of neighborhood to know how much is the restaurant worth.

With more detailed explanation, the requirements are as follows:

1. Latitude and Longitude are necessary to point the restaurants's coordinates and generate a map displaying all the restaurants pin-point with markers respectively.
2. The population of a neighbourhood is a crucial factor in determining a restaurant's growth and amount of customers who turn up to eat. Logically, the more the people of a neighbourhood, the more people will be interested to walk openly into a restaurant and less the population, less number of people frequently visit a restaurant.
3. 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 :

There are many people who love to travel and take a break from their stressed routine lives, and the best option is travelling and having the best food. With the generation of technology beside us now, one can easily get his/her smartphone to ask suggestions about where to go for a meal. Computers can store data for long and with the development of ML and AI algorithms data keeps updating and evaluates to much better results as days keep passing by.

Taking the list by copying the neighbourhood areas into an excel file, Chennai has 4 Direction classifications and 82 neighbourhoods. After this i have used the python geolocator library to extract data for every neighbourhood in the dataframe and displayed them as follows :

	Borough	Neighbourhood	Longitude	Latitude
0	North	Red Hills	80.183766	13.192279
1	North	Royapuram	80.294028	13.114619
2	North	Korukkupet	80.283791	13.121534
3	North	Vyasarpadi	80.257124	13.116973
4	North	Tondiarpet	80.289585	13.127767
...
143	South & East	Palavakkam	80.256027	12.959990
144	South & East	Varadharajapuram	80.205747	12.927599
145	South & East	Medavakkam	80.188290	12.922993
146	South & East	West Mambalam	80.227078	13.042732
147	South & East	Kottivakkam	80.261503	12.970505

Population is another aspect where we can determine the number of people that might visit a restaurant in that particular area. Finding the population was convenient as a basic setup from this link. This may not be accurate but can conveniently estimate as close prediction.

	Directions	Neighbourhoods	Population	Normalized_population
0	South & East	Adyar	7034	0.001222
1	South & East	Alandur	642237	0.111595
2	West	Alapakkam	9404	0.001634
3	West	Alwarthirunagar	8876	0.001542
4	West	Ambattur	466205	0.081007

Income by neighborhood is again easy to find out given that it's readily available. But in case of Chennai it is not the case. The given value may not be accurate but can suffice for our demonstrating project.

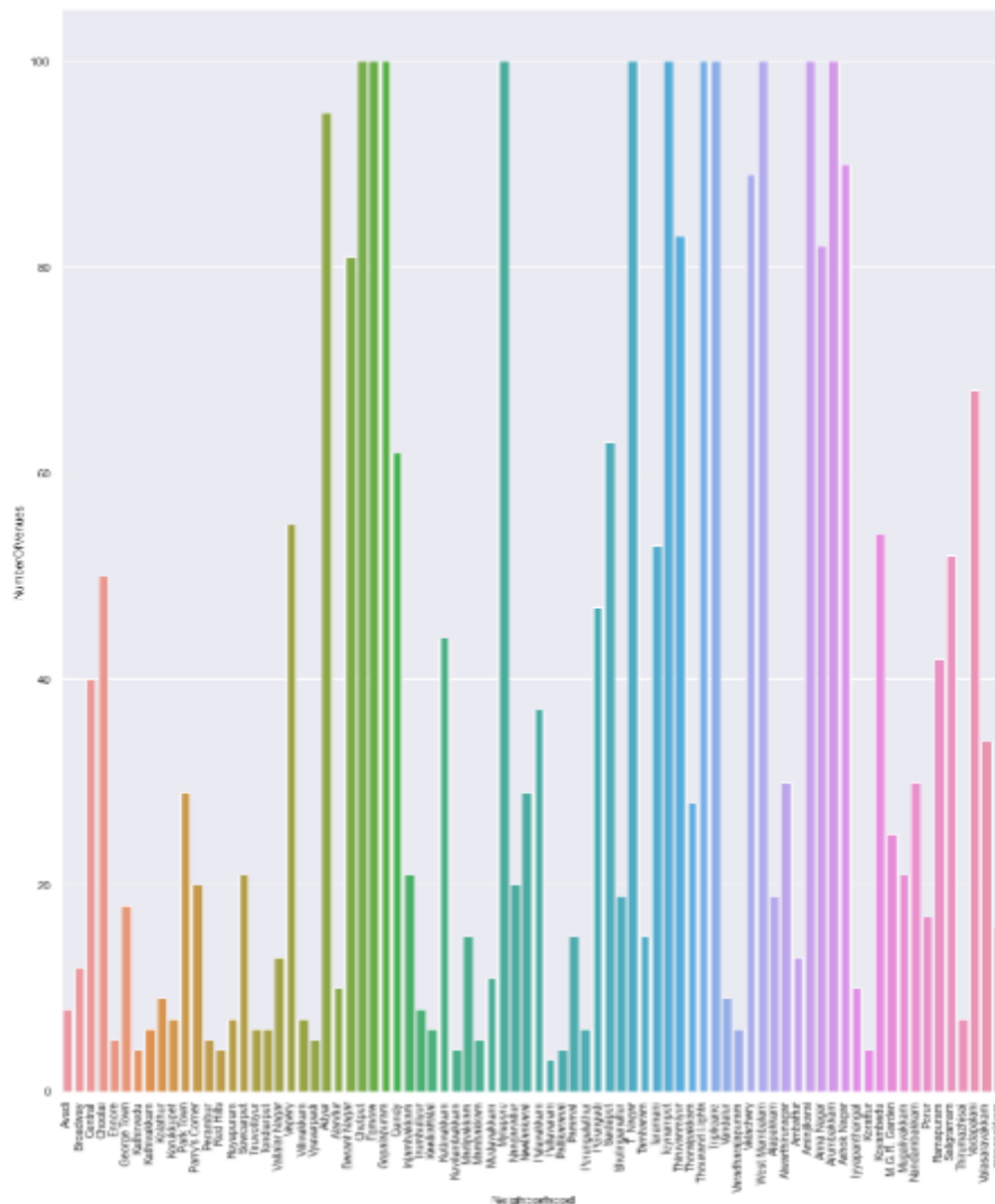
	Borough	Neighborhoods	AverageIncome	Normalized_income
0	Central	Cantonment area	18944.099792	0.293051
1	Central	Domlur	56837.022198	0.879225
2	Central	Indiranagar	41991.817435	0.649581
3	Central	Jeevanbheemanagar	6667.447632	0.103140
4	Central	Malleswaram	53270.063892	0.824047

4. Foursquare API:

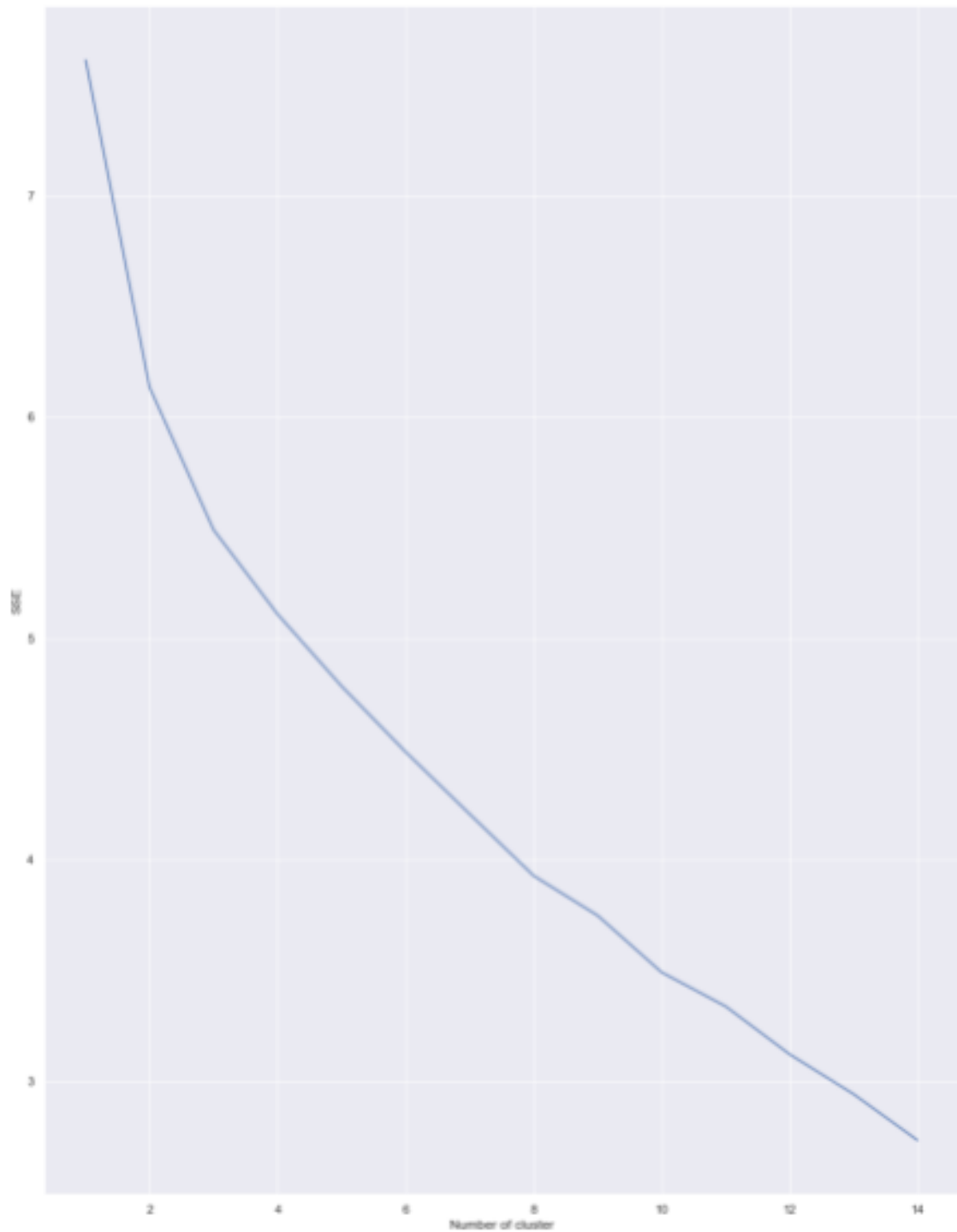
Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius (in my case: 2000mts) and also corresponding coordinates, venue location and names. After calling, the following data frame is created:

	Neighborhood	Directions	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adyar	South & East	13.00645	80.257779	Bombay Brassiere	13.006961	80.256419	North Indian Restaurant
1	Adyar	South & East	13.00645	80.257779	Prem's Graama Bhojanam	13.006345	80.253995	Vegetarian / Vegan Restaurant
2	Adyar	South & East	13.00645	80.257779	Cool Biz	13.006632	80.257275	Juice Bar
3	Adyar	South & East	13.00645	80.257779	Adyar Ananda Bhavan	13.005824	80.257368	Indian Restaurant
4	Adyar	South & East	13.00645	80.257779	That Madras Place	13.005848	80.250726	Café

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,



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 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 Page 9 neighbourhood will also increase. It is true to most of the case but also many cases differ to follow this trend. Similarly, a neighbourhood 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.

5. Result :

The result of the suggestion 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 'Mylapore' as the neighbourhood and then processed through our model so that it could recommend neighbourhoods with similar characters as that of 'Mylapore'. The following image shows the result:

	Neighbourhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Ranking
0	Adyar	Venue Category_Indian Restaurant	Venue Category_Café	Venue Category_Pizza Place	[0.220313523084876]
1	Alandur	Venue Category_Café	Venue Category_Hotel	Venue Category_Indian Restaurant	[0.4730221934446657]
2	Aminjikarai	Venue Category_Indian Restaurant	Venue Category_Fast Food Restaurant	Venue Category_Vegetarian / Vegan Restaurant	[0.2952657933829124]

6. Conclusion :

The suggestion 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.