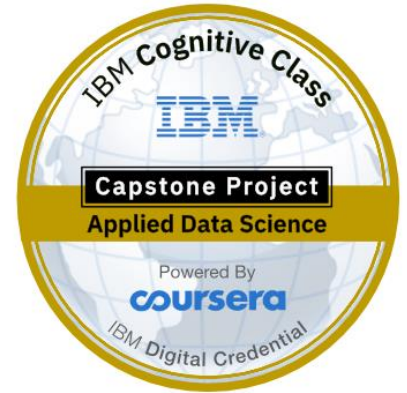


Chennai Attractions Recommender System



IBM-Applied Data Science Capstone Project

By

Dhinakar Babu

Table of contents

Topic	Page number
Introduction	2
Data	3
Methodology	5
Result	6
Discussion	7
Conclusion	7

Introduction

Chennai, on the Bay of Bengal in eastern India, is the capital of the state of Tamil Nadu. The city is home to Fort St. George, built in 1644 and now a museum showcasing the city's roots as a British military garrison and East India Company trading outpost, when it was called Madras. Religious sites include Kapaleeshwarar Temple, adorned with carved and painted gods, and St. Mary's, a 17th-century Anglican church. For tourists or new comers it is difficult to get to know about the nearby attractions. The diversity of the places available is reflective of the social and economic diversity of Chennai. Temples, Malls, Shopping sites are all very popular in the city. Chennai can also be called a Tourist favorites because of its vast variety of attractions and edibles with a touch of Tamil Nadu's uniqueness and tradition.

We have to build recommender system which recommends tourist travel locations based on his nearby venues. Recommended engine is built on an observation that tourist always try to explore places which are nearby first. We will be using location data to get best spots in neighborhood. For getting location data we will use Foursquare API.

Problem description:

Recommending nearby attractions to the users based on their preference and tastes, we would help customers to find their most related attractions in the nearby areas. This helps the problem of finding and exploring inappropriate places.

Business Questions

- How many types of attractions are available?
- Which is the nearest to me with good rating?
- How many "similar" places are available nearby me ?
- What types of places are present in a particular area?
- Where are the similar places present based on a preference to particular place ?
- How do different places 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 place all the time because they are addicted to a specific category of places. The target for this project is basically everyone who is exploring different places or similar places.

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 questions 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 solution.

Geographical coordinates (latitude and longitude) to find out where exactly it is located. To access location of attractions, its Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the attractions with its labels respectively.

- Taken the longitudes and latitudes from the google map and prepared the dataset. The data set contains Place with its Latitudes and Longitudes.
- To access location, 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.

Data collection:

Collected the coordinates of the geographical location in google.

It looks like the below table

BOROUGH	NEIGHBORHOODS	LATITUDE	LONGITUDE
1	Adyar Bus Debot.	12.5950	80.1525
1	Adyar Signal	13.0023	80.1527
4	Alandur	13.0028	80.1235
2	Ambattur	13.0636	80.1012
2	Anna Arch	13.0428	80.1306
4	Anna Nagar Roundana	13.0504	80.1305
3	Anna Nagar West Terminus	13.0535	80.1155
2	Anna Statue	13.0405	80.1619
1	Anna University Entrance	13.0029	80.1406
1	Avadi	13.0713	80.0636
1	AVM Studio	13.0252	80.1218
3	Ayyappa Temple	13.0323	80.1354
4	Basin Bridge	13.0608	80.1617

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:

	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Anna Arch	2	13.0428	80.1306	Ies Croissants	13.039185	80.131532	Bakery
1	Anna Arch	2	13.0428	80.1306	Aiyanar Bhavan Veg Hotel	13.040989	80.127192	Indian Restaurant
2	Anna Arch	2	13.0428	80.1306	Pizza Corner	13.045141	80.133925	Pizza Place
3	Anna Arch	2	13.0428	80.1306	Akshaya Bhavan	13.039774	80.132018	Indian Restaurant
4	Anna Nagar Roundana	4	13.0504	80.1305	The Colloseum	13.050660	80.131263	Sports Club

Methodology

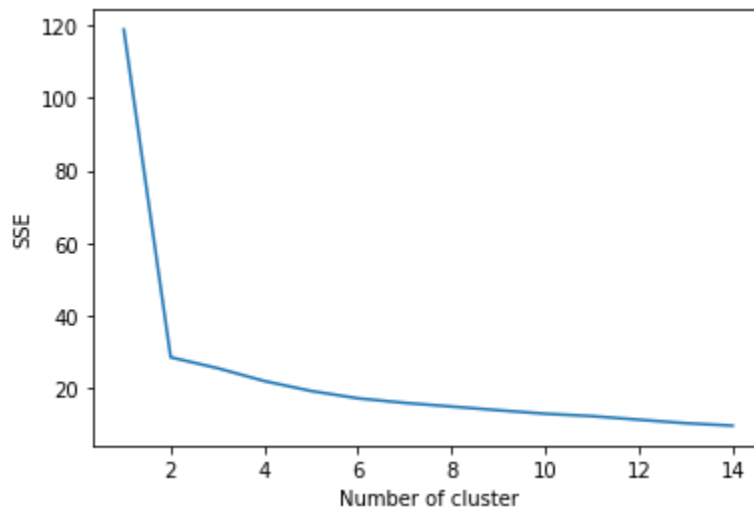
Exploratory analysis:

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.

- Done Data preprocessing to get clear data and one-hot encoded the data to model the data.

	Neighborhood	Venue Category_ATM	Venue Category_Afghan Restaurant	Venue Category_Asian Restaurant	Venue Category_Bakery	Venue Category_Bank	Venue Category_Bar
0	Anna Arch	0.000000	0.0	0.00	0.250000	0.000000	0.00
1	Anna Nagar Roundana	0.000000	0.0	0.00	0.000000	0.000000	0.00
2	Anna Nagar West Terminus	0.000000	0.0	0.00	0.000000	0.000000	0.00
3	Anna Statue	0.000000	0.0	0.00	0.000000	0.000000	0.25
4	Ayyappa Temple	0.000000	0.0	0.00	0.000000	0.000000	0.00
5	Basin Bridge	0.000000	0.0	0.00	0.000000	0.000000	0.00
6	CIT Nagar Roundana	0.000000	0.0	0.00	0.000000	0.000000	0.00

- 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.



Result

The result of the recommender system is that it produces a list of top attractions and the most common venue item that the user can enjoy. During the runtime of the model, a simulation was done by taking

- Anna Arch,
- Anna Nagar Roundana
- Anna Nagar West Terminus
- Anna Statue

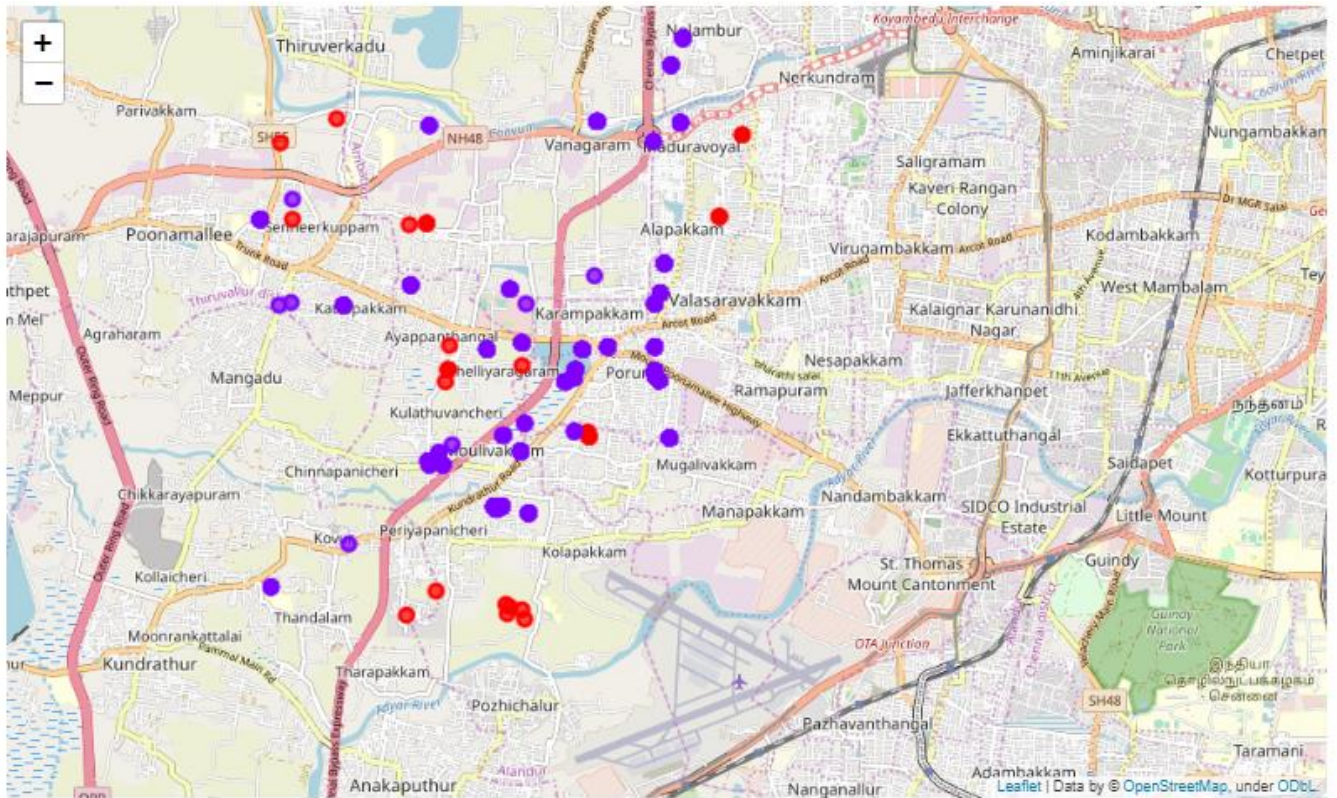
as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of it.

The following image shows the result:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7
0	Anna Arch	Cluster	Venue Category_Indian Restaurant	Venue Category_Pizza Place	Venue Category_Bakery	Venue Category_Food Service	Venue Category_Food Court	7
1	Anna Nagar Roundana	Cluster	Venue Category_Sports Club	Venue Category_Clothing Store	Venue Category_Food Service	Venue Category_Food Court	Venue Category_Food & Drink Shop	8
2	Anna Nagar West Terminus	Cluster	Venue Category_Indian Restaurant	Venue Category_Coffee Shop	Venue Category_Food Service	Venue Category_Food Court	Venue Category_Food & Drink Shop	9
3	Anna Statue	Cluster	Venue Category_Bar	Venue Category_Hotel	Venue Category_Electronics Store	Venue Category_Clothing Store	Venue Category_Bank	10
					Venue	Venue	Venue	11

Discussion

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.

Conclusion

The recommender system is a system that 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. Thus we have developed the Chennai based Attraction recommendation System using FourSquare API.