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 mst related attractions in the nearby areas. This helps the prblem of finding and exploring inappropriate places.

**Business Questions**

* How many types of attractions are available ?
* Which is the most nearest to me with good rating ?
* How many "similar" places are available near by me ?
* What types of places are present in a particular area ?
* Where are the similar places present based on a preference to particular food ?
* 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 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.

Geographical coordinates(latitide and longitude) to find our where exactly it is located.To access location of a 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 longitutes and latitudes from the gogle map and prepared the dataset. The data set contains Place with its Latitudes and Longitutes.
* To access location, 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.

**Data collection:**

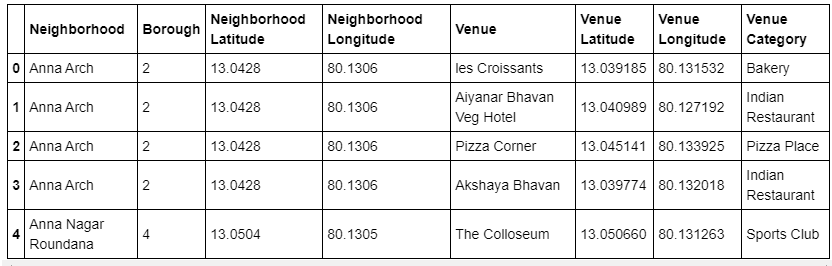
Collected the coordinated of the geographical location in google.

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:

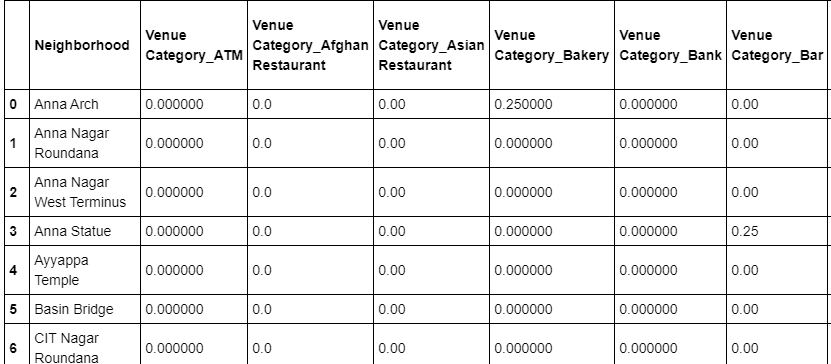


**Methodology**

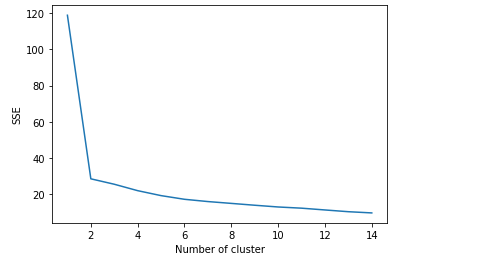
**Exploratory analysis:**

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.



* 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

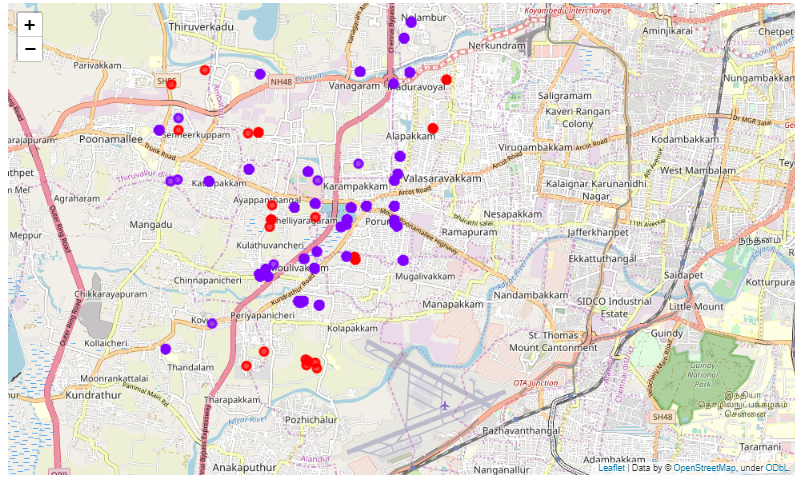
as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of it.

The following image shows the result:



**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 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.