|  |
| --- |
| Restaurant recommender system |

Subhrajyoti patra

Restaurant recommender system is a machine learning model, developed to demonstrate as a capstone project. It recommends restaurants based on user’s likes and dislikes and his previous interest data. I have used the city of ‘Bangalore’ to demonstrate the methodology and extensive analysis done by the ML code. The user can use this for any city of his/her choice.

**Problem description:**

A frequent traveler gets to experience different types of environment, of which he does not have much knowledge about. Food becomes an important factor in deciding how one rates his/her trips.

Expectations from this recommender system is develop a solution, in such a way that it uncovers the perspective of managing recommendations. It is sighted to show:

1. What types of restaurants are present in a particular area?

2. Where are similar restaurants located based on preference to a particular cuisine?

3. How do different restaurants rank with respect to user’s 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 my recommender system.

**Success rate:**

With restaurants evolving, new food categories emerge, and 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 everyone about their visit to a particular place and also not everyone remembers everything. But, computers are good at remembering things, and with Machine learning to its peak, it is high time that technology be our personal guidance and help us personally based on our likes and dislikes. So people would care about this project as a personal assistance and the success rate could certainly increase with time.

1. **Data :**

**Data requirements:**

To find a solution to the questions and build a recommender model, we need a ton 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 datasets and produce analytics to find 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 the restaurant is 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, the less number of people 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 significant factor. Income is directly proportional to the richness of a neighborhood. If people in a neighborhood earn more than the average income, then it is very much possible that they will spend more however this is not always true. So a restaurant assessment is proportional to income of a neighborhood.

**Data collection:**

1. I decided to use Google maps API to fetch latitude and longitude, but google API has limited number of calls that I could make. So it took around 15 - 20 days to fetch location of all the neighborhoods in Bangalore.

Initially I scrapped the 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.



2. I was able to find population data for a 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 is to get the working model. The data frame for Bangalore neighborhood population looks like:



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. [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:



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:



1. **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 Indiranagar has most number of venues while Varthur has the least.



Also while producing the graph for number of clusters, I produced a graph to explore all the values for n\_clusters and then found the best by exploring the elbow graph.



**Inferential analysis:**

Population and income were the most significant factors while building the recommender system. They have a nonlinear relationship according to our dataset.

Inferential analysis was necessary to understand this nonlinear relationship. Population increase does not always imply an increase in average income. Similarly, a decrease in population does not necessarily mean a low average income. It is possible to low population and high average income and vice versa. This can be inferred from the following graph:



1. **Discussion :**

We must always perform inferential approach to find relationship among different set of features since there may be non-linear relationships. During clustering, similar neighborhoods must be dumped into the right cluster.

The following graph shows the clusters:



Different values for the number of clusters could produce diverse results. They may be over fitted or under fitted.

**4.Result :**

The result of the recommender system is that recommends the user the top restaurants and the most common venue item that the user can enjoy. “Whitefield” was taken as the neighborhood during the runtime of the program. It was then processed through the model so that it could recommend neighborhoods with similar characters as that of ‘Whitefield’.

The following image shows the result:

 **5. Conclusion :**

The recommender system 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.