Zomato Bengaluru Restaurants Analysis and Recommender System

Revanth Babu P N
Computer Science and Engineering
PES University
Bengaluru, India
putturevanth@gmail.com

Srinath M P
Computer Science and Engineering
PES University
Bengaluru, India
srinathmp77@gmail.com

Makrand Kulkarni
Computer Science and Engineering
PES University
Bengaluru, India
makrandkool@gmail.com

***Abstract*— Using the Zomato Bengaluru Restaurant dataset we analysed, a recommender system was built using the reviews and user preferences. The recommendation considers cuisines and user reviews that are similar to their restaurant entries.**

Keywords*— *recommender systems, preferred, amenities, ratings

I. Introduction

The principal goal of the Zomato Bengaluru dataset analysis is to get a decent understanding of the factors influencing the establishment of different types of restaurants in Bengaluru. Bengaluru being the silicon valley of India has seen a surge in people hailing from different parts of the world and most of them being dependent on restaurants for food .The overwhelming demand for restaurants and unending appetite of people has led to establishment of more than 12,000 of them, and still counting. Factors such as demography of location, income of people, popularity of certain cuisines in localities become key influential factors in deciding the success of a restaurant. A modern age of knowledge has spawned the exponential growth of data aggregation. To build more effective processes, data is continually used as fuel and this is where Recommendation Systems come into play. Recommender Systems are a type of knowledge filtering systems as they improve upon the accuracy of the results of the search and include recommendations that are more applicable to the search object or in line with the user's search history. They are proactive information filtering systems which personalize the content presented to a user based on their interests, relevance,preferences etc. Recommender systems are prominent in technology, retail, marketing ,education, entertainment,hospitality etc. We aim to discover further insights and unearth underlying patterns from the data through various data analytics techniques.

Further we propose a recommender system to make restaurant recommendations according to the users' favoured cuisines, prices and other amenities like home delivery, location proximity and others, as specified by the user.

II. Problem Statement

Analysis of the Zomato Bengaluru Dataset to find out the relations such as what kind of cuisine is more popular in a locality based on recommendation techniques. Recommending restaurants based on the user ratings of that specific geographic region. We aim to find the insights and unearth underlying patterns using the Zomato Bangalore restaurants data through data analytics and design a recommender system that provides personalized restaurant recommendations to the users based on their food preferences like favoured cuisines, restaurant type and their location. The input to the recommender system will be the user profile which contains the preferred food items, cuisine of the user, and the recommendation generated will be based on restaurants nearby the user which have a reviewed acclaim of that item.

III. Literature Survey

The paper[1], by Fu et al proposed a method to restaurant recommendations that makes use of multiple aspect ratings including geographic information, and profile information of restaurants and users. Their proposed model requires the use of multi-aspect ratings. The key takeaway from this paper is that with the incorporation of profile similarity, multi aspect rating and geographical influence, we can observe significant improvements in the performance of the model.

Gomathi et al [2] proposed a model which works on the assumption that users get attracted to praiseworthy and good feedback. Restaurants which have positive feedback are more likely to be preferred by the user. Using the tripadvisor search data, they proposed a machine learning algorithm to provide personalized Restaurant recommendation. The various existing techniques such as BPN, NLP, PNN and SVM were compared against the sentiment analysis techniques for restaurant recommendation. Then, the necessary details were obtained by parsing the user reviews. It was found out that NLP (Natural Language Processing), a machine learning technique is more efficient than other sentiment analysis techniques such as Back Propagation neural network Support Vector Machine(SVM) and Probabilistic neural network and hence these techniques can be used for improved results in recommending restaurants.

The paper by Habib et al uses LSBN and the geographic factors for predicting the users rating instead of user reviews which is more practical for real time use. The recommendation was based on the following four key factors, which are, user preference, the time of a day, the distance of the restaurants, and the ratings of the restaurants. Firstly it discovers users' preference trend using the users' historical data. Then their model includes the timing of the restaurant operation in order to identify the type of meals served by the restaurant. Finally, it considers the distance of the restaurant for modelling the willingness of a user going to a restaurant, as distance works inversely with the willingness of going to a restaurant.

The papers described above propose various novel approaches to recommending restaurants such as multi-aspect ratings, geographic information, user profiles, restaurant profiles and review sentiments but none consider the similarity of user reviews. We also believe ratings alone are not a true representative of people's opinion towards a particular restaurant. In addition to these factors, in a place like Bengaluru, proximity plays a major hand in deciding whether a customer visits a restaurant or not and insights to how far a restaurant is from current location while recommending a restaurant will help them choose decisively. The solutions to these are described in the following section.

IV. Proposed Solution

The dataset chosen is Zomato Bangalore Restaurants which is hosted on Kaggle. The chosen dataset contains a plethora of information about restaurants in Bangalore. The dataset contains information about restaurants and their associated features and not user profiles. The absence of user profiles limits the application of recommender systems such as collaborative filtering that are dependent on similarity of user ratings. This limitation also inhibits the use of validation metrics such as RMSE, MSE to evaluate the recommendation as there are no actual values to compare against and evaluate.

The proposed solution is a hybrid approach to recommendation that mainly considers the similarity of the user reviews of restaurants with the one given by the user.

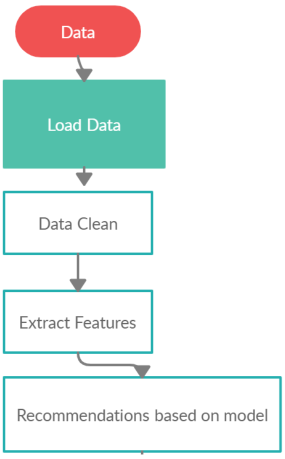

Figure 1

Pre Processing:

The dataset was cleaned to remove NaN values, transformations were applied on columns like cost, rate which is required for further analysis. We have used MinMaxScaler to scale the average ratings to range between 0 and 5.

The user reviews are cleaned and processed for stop words removal using the nltk python module, punctuation removal and also universal resource locators.

Model Building:

The user review for the given restaurant is transformed using Term Frequency - Inverse Document Frequency (TF - IDF) returning a vector. The relevance of a word in a document among a group of documents is commonly evaluated by using a statistical measure called TF-IDF. The term frequency corresponds to how many times a given term appears in a document and the Inverse document frequency (idf) measures the weight of the word in the document, i.e. if the word is common or rare in the entire document. The metric can be computed by calculating the logarithm of the total number of documents, dividing it by the number of documents that contain a word.

The vector returned is compared with the vectors of restaurants from the database represented as a matrix based on cosine similarity and then further processed.This procedure of comparing review vectors returns restaurants that have been reviewed similarly by customers.Thus a user can find restaurants with matching features in reviews such as food quality, ambience, location aesthetic, hospitality, culinary expertise, hygiene maintenance and other qualities that are preferable and present only in the reviews and nowhere else. The equations for term frequency and inverse document frequency are given below.

The term tf_{ij} represents the term frequency for the i *th* word in the j *th* document, n_{ij} is the number of occurrences of word i in document j and the denominator is the total number of words in the document.

$$tf_{i,j} = \frac{n_{i,j}}{\Sigma_k n_{i,j}} \tag{1}$$

The term $idf(w)$ represents the inverse document frequency of the word w, N is the total number of documents and df_t is the number of documents that contain the word .

$$idf(w) = log\left(\frac{N}{df_t}\right) \tag{2}$$

The cuisines are also similarly processed to match similarity to the user given restaurant . The results of the two methods described above are combined based on the weighted preference and the result is presented to the user after sorting based on ratings . These methods implement a case-based reasoning form of content-based recommendation .

The user may not always be familiar with the restaurants in the city or may be new to the place (cold start problem). In these instances the user is given a choice to enter features such as cuisines, restaurant types and location based on which the recommendation is made thus offering a knowledge based approach to recommendation. The similarity metric considered for this purpose is cosine similarity. The two proposed approaches are shown in figure 2 and figure 3.

The case where the user wants a recommendation for a restaurant similar to his choice is represented by figure 2.

Here the user gives a restaurant as input and the corresponding cuisines, reviews and other features of the user specified restaurant are obtained from the dataset . The term-frequency inverse-document-frequency vectors are computed for these and also the restaurants in the dataset. The similarity is found between each of these pairs of vectors and the most similar are selected for further processing. The results of this approach is shown in figure 4.

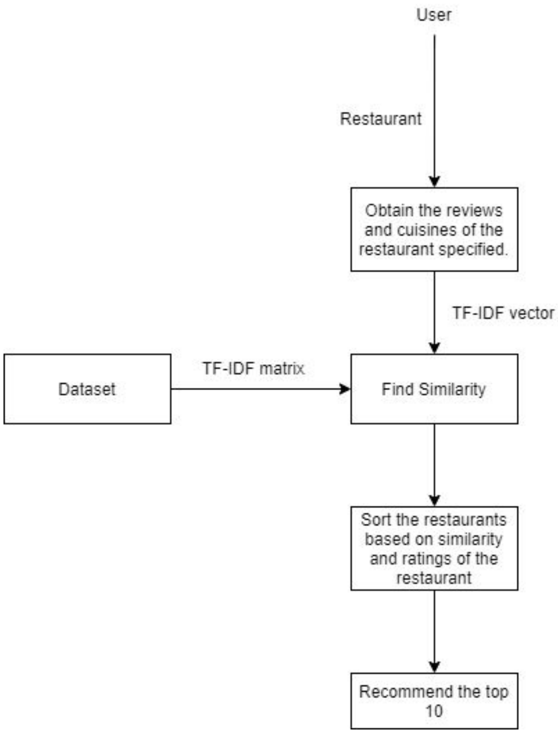

Figure 2

Figure3 represents the second approach, this is the case where the user is not aware of the local restaurants and wishes to get recommendations based on features such as cuisines, location. Here feature columns of the various restaurants are compared to match the user specified constraints based on the cosine similarity metric and most similar are sent for further processing. The results of this approach is shown in figure 5.

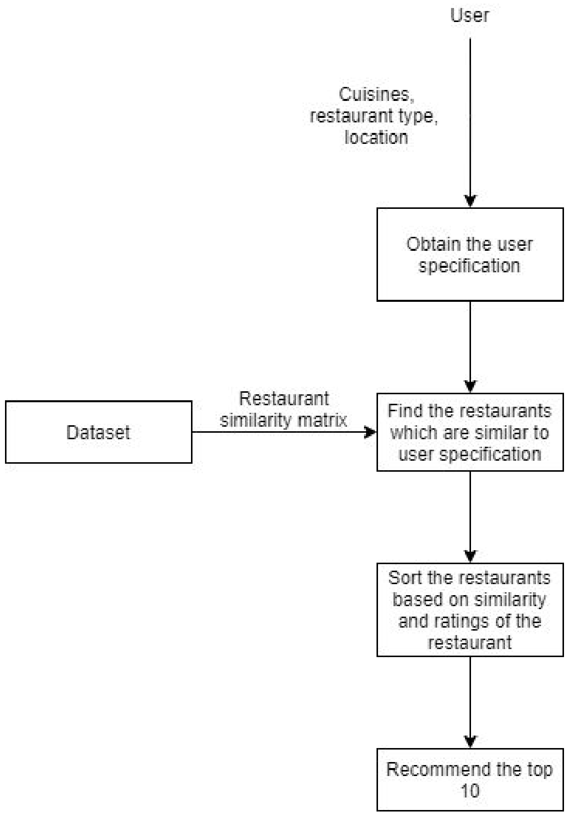

Figure 3

The recommendations also give the distance of a restaurant from the current location , which helps the user in picking the closest one that is viable . As described by Habib et al in [3], the willingness of going to a restaurant works inversely with the distance one has to travel especially in a city like Bengaluru with widespread and heavy traffic congestion throughout the year.

V. Experimental Results

The recommender system in action is described in this section , demonstrating the techniques used , described above .

```
Welcome to Yottabytes Restaurant Recommender!

Please enter your location
BTM Layout

Would you like to search by restaurant or cuisine? press 1 for Restaurant 2 for cuisine
1
Enter Restaurant Name
Punjab Bistro

Restaurants with similar reviews and cuisines to this are:
TOP 9 RESTAURANTS LIKE Punjab Bistro:
|    | name                                    | cuisines                                  |   Distance(km) |   Mean Rating |   cost | rest_type           | location          |
|----|-----------------------------------------|-------------------------------------------|----------------|---------------|--------|---------------------|-------------------
|
|  0 | Punjab Grill                            | North Indian                              |           16.3 |          4.96 |   2000 | Casual Dining       | Whitefield        |
|  1 | Brew And Barbeque - A Microbrewery Pub  | Continental, North Indian, BBQ, Steak     |           10.5 |          4.64 |   1400 | Microbrewery, Pub   | Marathahalli      |
|  2 | The Globe Grub                          | Continental, North Indian, Asian, Italian |            0.8 |          4.48 |   1300 | Casual Dining       | BTM               |
|  3 | Delhi Highway                           | North Indian, Mughlai                     |            7.2 |          4.41 |   1500 | Casual Dining       | Indiranagar       |
|  5 | Mumbai Tiffin                           | ['North Indian']                          |            3.5 |          4.35 |    400 | Quick Bites         | HSR               |
|  4 | Dhaba Estd 1986 Delhi                   | North Indian                              |           10.5 |          4.29 |   1100 | Casual Dining, Bar  | Marathahalli      |
|  6 | The Paratha Company                     | ['North Indian']                          |           16.3 |          4.19 |    400 | Delivery            | Whitefield        |
|  7 | Mooch Marod                             | ['North Indian']                          |           16.3 |          4.17 |    350 | Quick Bites         | Whitefield        |
|  8 | Punjabi Times                           | ['North Indian']                          |            2.4 |          3.97 |   1000 | Casual Dining       | Bannerghatta Road |
```

Figure 4

The recommendation for a user whose location and preferred restaurant is given is depicted in figure 4 . The final recommendation has the name , available cuisines distance from user location, mean rating of the restaurant, approximate cost for two people , the type of restaurant and the physical location. The recommendation returns similar restaurants that serve the same kind of food and have similar customer reviews and geographically close to the user.

```
Welcome to Yottabytes Restaurant Recommender!

Please enter your location
BTM Layout

Would you like to search by restaurant or cuisine? press 1 for Restaurant 2 for cuisine
2

The available cuisines across all the restaurants are
:
0   Biryani
1   South Indian
2   North Indian
3   Chinese
```

```
Enter number of cuisines
1
Enter cuisines of your choice:
Street Food
Restaurants with similar cuisines are:
|    | name                   | location          |   Approx distance in km | rest_type   | cuisines    |   cost |   Mean Rating |
|----|------------------------|-------------------|-------------------------|-------------|-------------|--------|---------------|
|  0 | Sri Sai Ram'S          | Malleshwaram      |                    10.7 | Quick Bites | Street Food |    200 |          4.1  |
|  1 | Gappe                  | Bannerghatta Road |                     2.4 | Kiosk       | Street Food |    200 |          3.58 |
|  2 | Samosa Corner          | Bellandur         |                     6.5 | Quick Bites | Street Food |    100 |          3.45 |
|  3 | Royal Chowpatty        | Electronic City   |                     8.7 | Quick Bites | Street Food |    100 |          3.45 |
|  4 | Chintu Chit Chat Centre | Banaswadi        |                    11.9 | Quick Bites | Street Food |    100 |          3.28 |
|  5 | Goli Vada Pav No. 1    | Electronic City   |                     8.7 | Quick Bites | Street Food |    150 |          3.24 |
|  6 | Ganga Sagar            | Rajajinagar       |                    10.1 | Quick Bites | Street Food |    150 |          3.06 |
|  7 | Thalassery Cafe        | BTM               |                     0.8 | Quick Bites | Street Food |    150 |          3.06 |
|  8 | Samosa King            | BTM               |                     0.8 | Quick Bites | Street Food |    200 |          3.06 |
|  9 | Cafe Bollywood         | JP Nagar          |                     2   | Quick Bites | Street Food |    200 |          3.06 |
```

Figure 5

The recommendation for a user whose location and preferred cuisines is given is depicted in figure 5. The final recommendation here also has the name , available cuisines distance from user location, mean rating of the restaurant, approximate cost for two people , the type of restaurant and the physical location. Here the restaurants with similar cuisines are returned that are close to the user's location.

The recommendation relies highly on the quality of reviews given by the user. The performance is better for instances with longer and more detailed explanatory reviews and performs poorly for brief reviews .

VI. Conclusion

In this paper, we provided a recommender selection process system to make restaurant recommendations according to the users' favoured cuisines, prices and location proximity and others, as specified by the user.

The analysis of factors such as demography of location, income of people, popularity of certain cuisines in localities become key influential factors in deciding the success of a restaurant. Since ratings alone are not a true representative of people's opinion towards a particular restaurant we factor in the reviews that possess many characteristics such as food quality, ambience, location aesthetic, hospitality, hygiene maintenance and other qualities.

The experimental results on real-world restaurant rating data verify the performance of the method. The future prospect is that as the number of users increases, user profiles can be stored in the database. Once the number reaches a significant value, the user profile similarity can be used with the current model making use of novel techniques such as collaborative filtering, thus functioning as a hybrid approach to the recommendation process.

Individual Contributions **:**

Revanth Babu P N: Contributed to data preprocessing exploratory analysis, model design first and second approach.

Srinath M P: Contributed to exploratory data analysis, literature survey, model design and implementation.

Makrand Kulkarni: Contributed to literature survey, data preprocessing, exploratory data analysis, and report documentation.

REFERENCES

[1] Fu, Y., Liu, B., Ge, Y., Yao, Z., & Xiong, H. (2014, April). User preference learning with multiple information fusion for restaurant recommendation. In Proceedings of the 2014 SIAM international conference on data mining (pp. 470-478). Society for Industrial and Applied Mathematics.

[2] Gomathi, R. M., Ajitha, P., Krishna, G. H. S., & Pranay, I. H. (2019, February). Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities. In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) (pp. 1-6).

IEEE.

[3] Habib, M. A., Rakib, M. A., & Hasan, M. A. (2016, December). Location, time, and preference aware restaurant recommendation method. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 315-320). IEEE.

[4] Business Analytics, The Science of Data-Driven Decision Making,U Dinesh Kumar, Wiley 2017.

[5] Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber and Jian Pei

[6]https://towardsdatascience.com/location-based-recommendations-bb70af7b1538

VII. Appendix

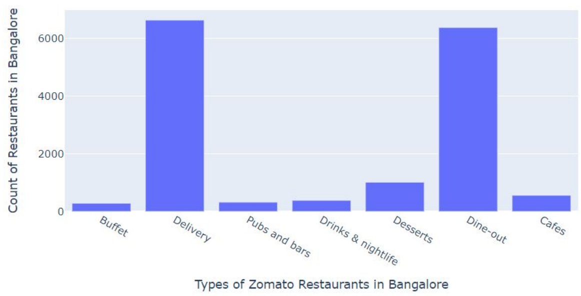

Figure 6

Figure 6 shows the major types of restaurants that are prevalent in Bangalore.

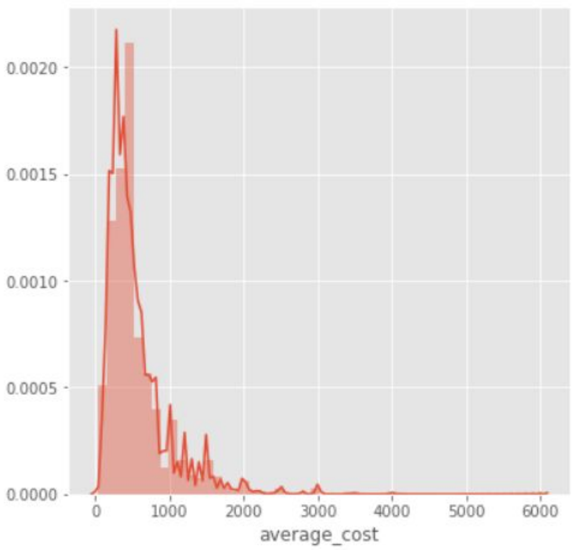

Figure 7

We can see in figure 7 that the distribution is left skewed. This implies most restaurants serve food for a budget less than 1000 INR. Low cost and street food make up for the majority of the restaurant types in Bangalore.

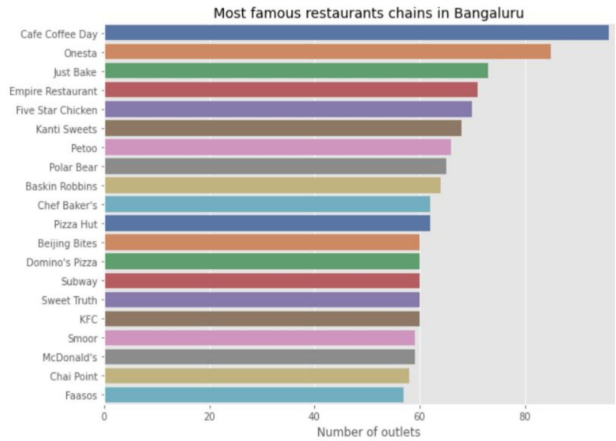

Figure 8

The graph shown in figure 8 depicts the restaurant chains with a number of sub branches that it has deployed . Our model's recommendation doesn't depend on the number of restaurant chains as it is based on geographical locations and restaurant ratings by the user.

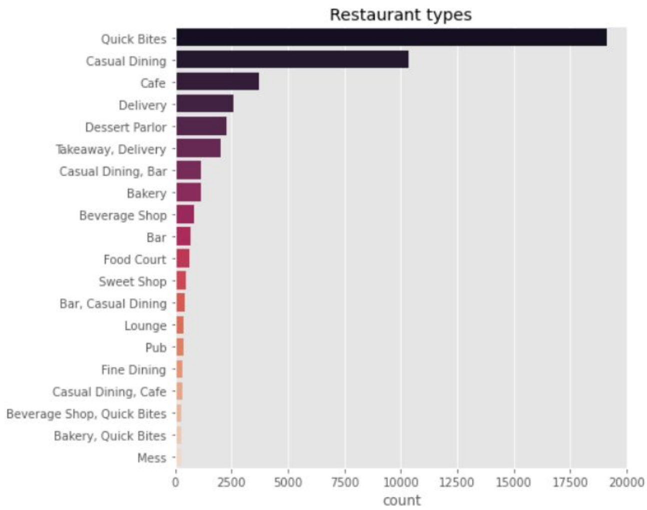

Figure 9

Figure 9 clearly shows the domination of quick bytes type of restaurants in food orders.

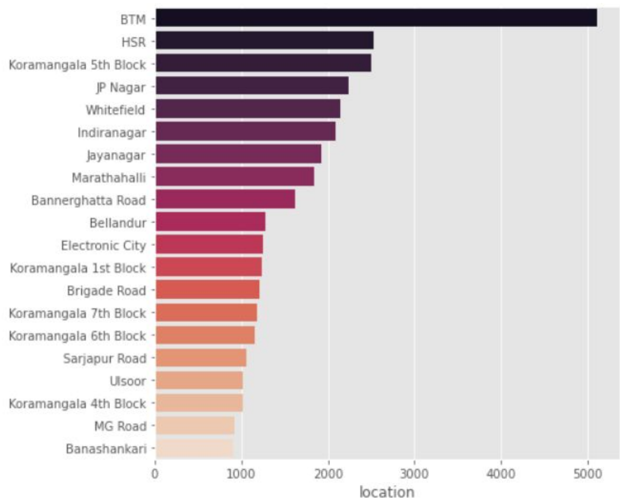

Figure 10

In BTM layout in the city maximum number of restaurants are established.

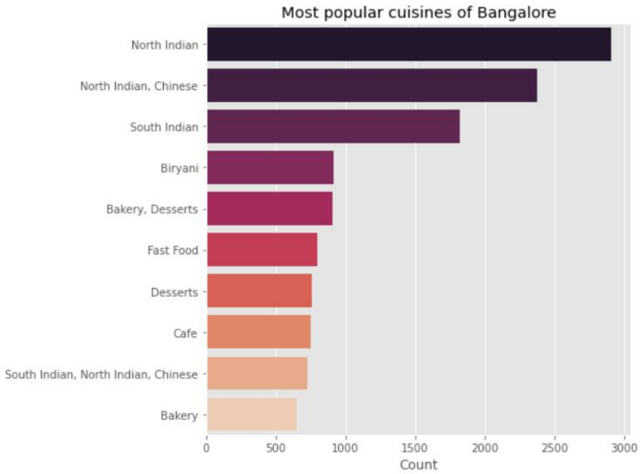

Figure 11

The North Indian Cuisine is a hot favorite of the bangaloreans.

Keeping the various factors of the items that are favored by the users it is likely that the users might be recommended on the basis of those above said factors.

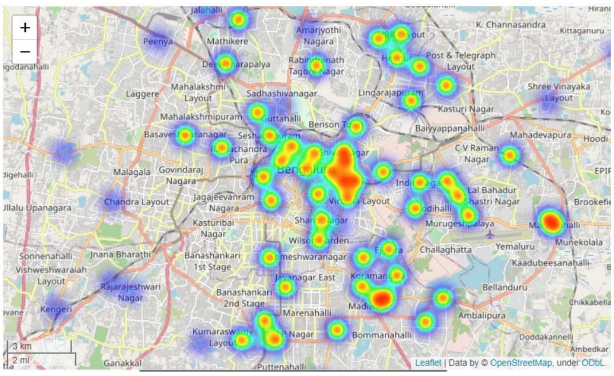

Figure 12

We clearly see from figure 12 that most of the restaurants are concentrated in the central part of Bangalore. As we move away from the central region the concentration tends to decrease. Entrepreneurs need to consider this pattern while searching for potential restaurant locations.