Data Analysis on the Food Places of Zomato

[[1]Prashun Pareek, [2]Nidhi Kumari, [3]Dr. Stuti Saxena, [4]Dr Manish Kumar]

[1,2,3,4, Department of Computer Science & Engineering, Echelon Institute of Technology, Faculty, Echelon Institute of Technology]
[stutisaxena@eitfaridabad.co.in, drmanishkumar@eitfaridabad.co.in]

Abstract

The survey report is conducted on food places in India about how they are influencing more customers to the cafes, restaurants that are listed on zomato website. The main purpose behind this survey and observation was to find the influencing factors behind popularity of food places, with their working areas, prices, online availability and many more other things. The entire data analysis consists of graphs and figures that helps us understand the rating based popularity of the restaurants in a specific region.The main focus of the analysis is to identify the distribution of cuisines across different restaurants from different parts of India and their popularity by aggregating user ratings. Statistical techniques and other data visualization tools are utilized to explore patterns and relationships within the dataset. Machine learning algorithms such as clustering or classification are also used to understand the trend of popularity to the type of cuisines provided by the restaurants.

Introduction

This data analysis project was conducted in order to get an Idea on the popularity of cuisines on Zomato based on user ratings, extracting the valuable information about the restaurant industry and consumer behavior. For further analysis the graphs have been plotted in order to understand more detailed information about the influence of different factors affecting the price and availability of cuisines in the area. The factors are all listed in the dataset, furthermore, we need to select them explicitly. The Project is divided into 5 major steps:

1. Extraction of data

2. Data cleaning and mining
3. Exploratory Data analysis
4. Data visualization
5. Implementation of Machine learning algorithms.

Exploring the Data

We are fetching the country code and matching the country code with the data aggregated by the zomato. We are using Pandas and using a filter code to match the country code with the code in data. After getting some insights such as description of the data, sample of the data and all the unique features and number of cuisines provided by the restaurants. We will try to find out what factors influence the ratings of the restaurants. The ability to book a table, for example what might influence how a potential customer leaves a rating. Another important factor would be the authenticity of the cuisine being served. As a general rule of thumb, one might say that a multi-cuisine restaurant might be less authentic than one that serves only one type of cuisine.

Various Features such as:

1. Restaurants having table booking:

    It seems that for restaurants that are more affordable (has a lower price range) have higher ratings if they provide table booking services.For pricier restaurants table booking option does not seem to affect the ratings much.

2. Restaurants Having Online Delivery:
    Again, restaurants that are in the lower price range (1 and 2) seem to have better average ratings if they provide online delivery services. This is not so significant for the restaurants falling in the pricier category.

3. Restaurants that are providing delivery and can deliver now:
    Again the trend continues and it shows the delivery feature does not have any significant impact on the rating for pricier restaurants while the lower price restaurants have comparatively lower ratings than them.

4. Switch to order services:
    The 'Switch to order menu' service is not provided by any of the restaurants that have been reviewed in Zomato. But this interestingly shows the general trend of ratings provided by the users. The lower price restaurants in general have worser ratings compared to the pricier ones.
    There is essentially little difference in the ratings of price categories 1 and 2.

Cuisines Provided by the Restaurants

The above restaurants are a mix of single cuisine and multi-cuisine restaurant. The rating of restaurants will be affected by the authenticity of the cuisine they are serving.

It might be possible to predict the ratings of those restaurants based on the characteristics of the restaurants that have already been rated.

This is a very interesting situation. It seems that contrary to our own expectations, it seems that multi-cuisine restaurants in India seem to have a little better aggregate rating.

It would be interesting to see the distribution divided between the four price ranges
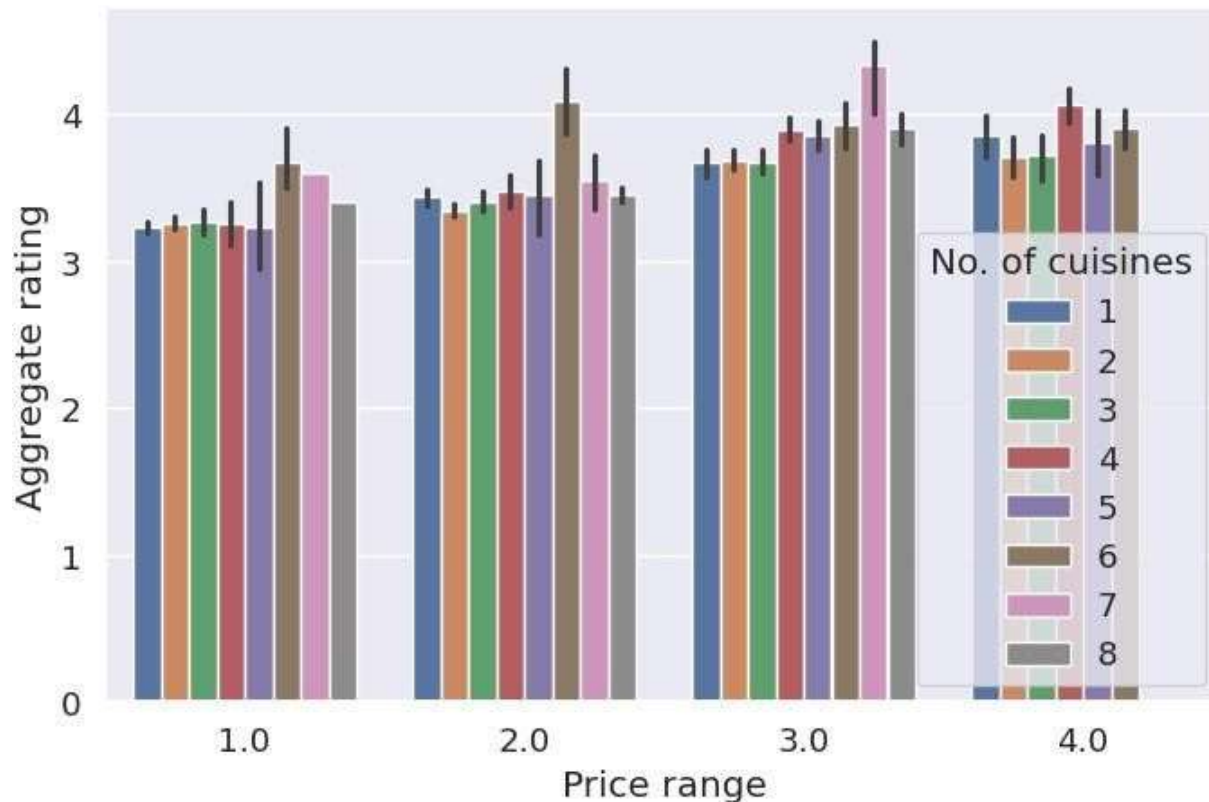

The trend of the aggregate rating being better for multi cuisine restaurants seems to be valid for all the price range of restaurants. Thus in India it is likely that if you have a multi-cuisine restaurant it would probably get better ratings. Additionally, the highest price category restaurants provide no more than 7 cuisines.
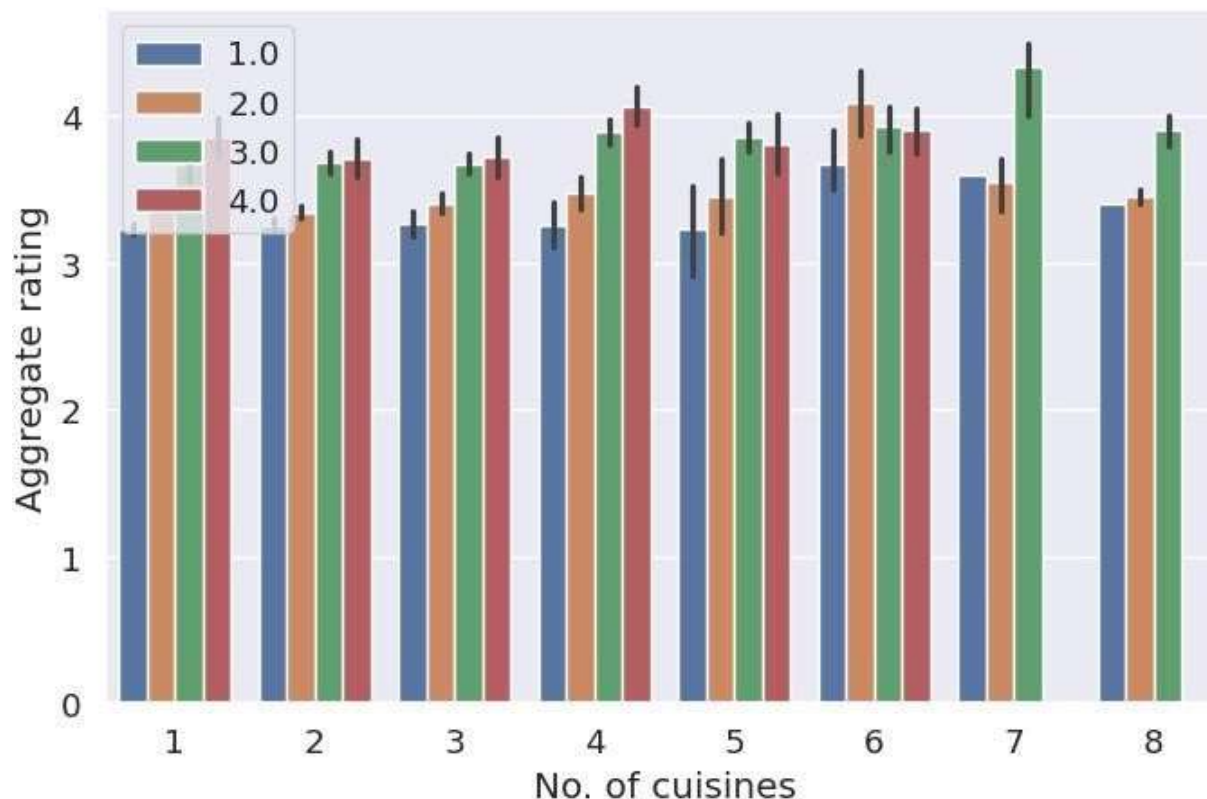
We can infer from above that the aggregate rating of the restaurants tend to increase with the price range irrespective of the number of cuisines that they serve.

Voting Trends

Another important parameter to know is the number of footfalls of each restaurant. However what we have here is the number of reviewers which is only proportional to the total number of footfalls over a given duration. Since the duration during which the reviews were obtained are not given, we will assume this to be the same for all restaurants as of now to have a general idea about the number of people who visit the restaurant. The number of footfalls per restaurant should also depend on the population of the city in which the restaurant is located, the location itself, cuisine etc. Curiously enough a better aggregate rating should also contribute to the number of footfalls.
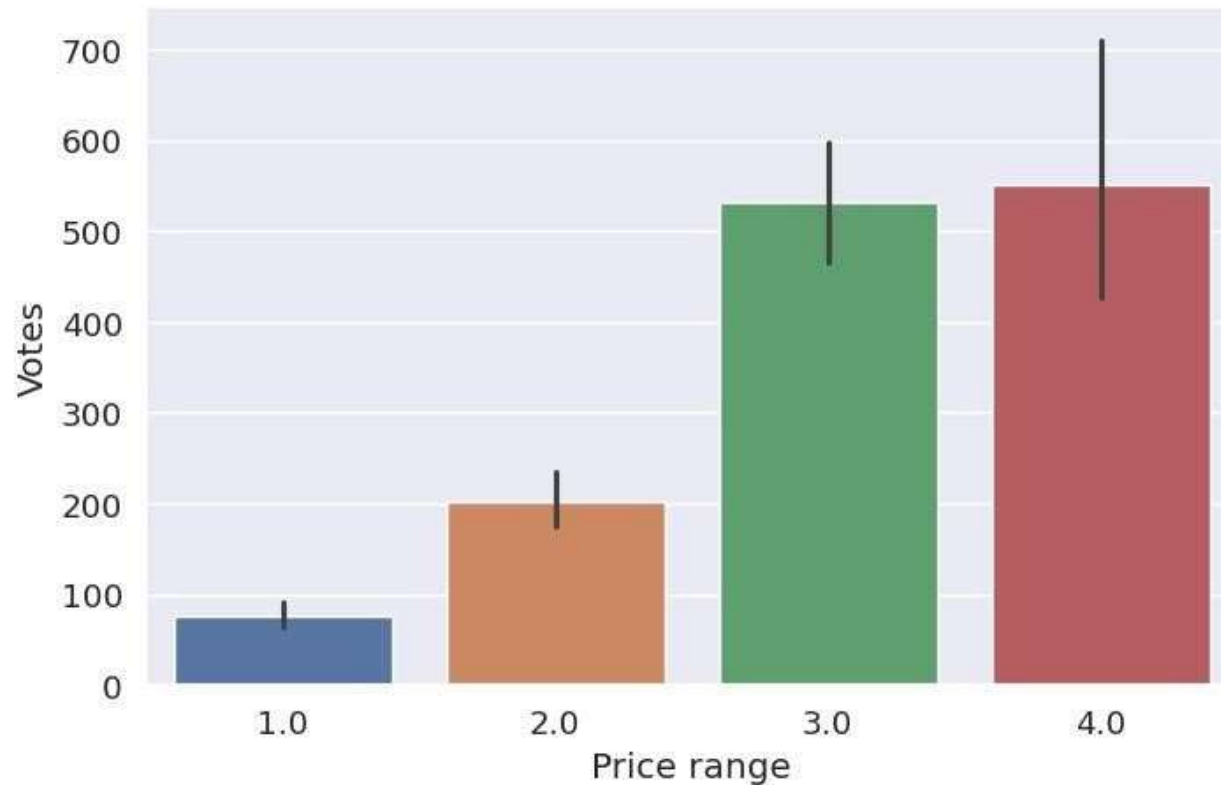
The more pricier restaurants seem to have a larger number of reviews which implies the pricier restaurants seem to have a better footfall. It might be also possible that people who visit these restaurants are more internet savvy and likely to leave a review.

Let us look at the relation between average rating and number of votes:

the variability of footfalls is also large for the highly rated restaurants. This might be because there might be new restaurants which have better aggregate ratings. However since the data does not mention the period during which the restaurants were obtained this is just a speculation. The location of the restaurant, how easily it is accessible by public transport, the population density in the region where it is located, whether it is closer to the center of the city etc.

15
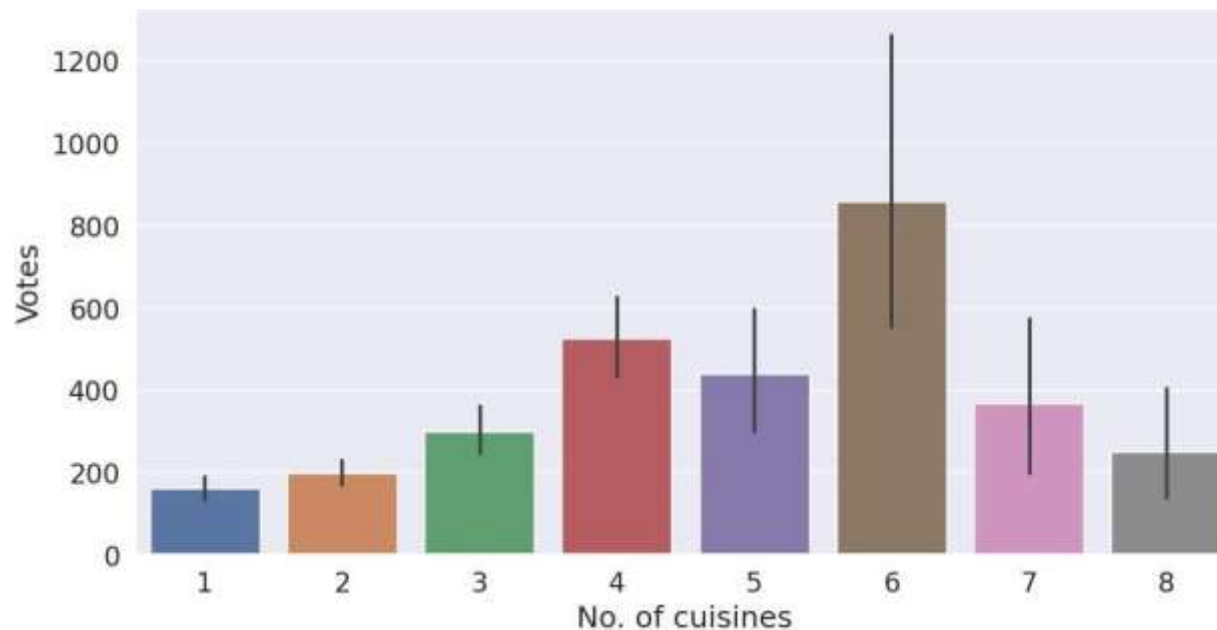
The above data shows if the restaurant consists of more than 5 cuisines are likely to receive more votes.

Cities

It would also be interesting to know the cities in India where these restaurants are located and to see in which city people use Zomato to rate the restaurants. The total number of votes in each city could also be a measure of the popularity of Zomato in that city (or state in which the city is located).

Most Popular Cuisines

In the above restaurants we have both single cuisine and multi-cuisine restaurants. We first need to make a list of all available cuisines across all restaurants. I will try to generate the list of cuisines that are provided by all the restaurants above.

['North Indian', 'Mughlai', 'Rajasthani', 'Chinese', 'European', 'Gujarati', 'Continental', 'South Indian', 'Desserts', 'Cafe', 'Italian', 'Mexican', 'Pizza', 'Fast Food', 'Mediterranean', 'Thai', 'Ice Cream', 'Beverages', 'Asian', 'Street Food', 'Sandwich', 'Burger', 'Healthy Food', 'American', 'Armenian', 'Salad',

'Bakery', 'Mithai', 'Biryani', 'Juices', 'Maharashtrian', 'Hyderabadi', 'Modern Indian', 'Finger Food', 'Tex-Mex',

'Arabian', 'Charcoal Grill', 'Steak', 'Seafood', 'Tea', 'Japanese', 'Malaysian', 'Burmese', 'Chettinad', 'Spanish',
'Greek', 'Indian', 'Parsi', 'Tibetan', 'Raw Meats', 'French', 'Goan', 'German',
'Kerala', 'Lebanese', 'Belgian', 'Kashmiri', 'Sushi', 'South American', 'Persian', 'Bengali', 'Portuguese',
'African', 'Iranian', 'Vietnamese', 'Lucknowi', 'Korean', 'Awadhi', 'Nepalese', 'Drinks Only', 'Pakistani',
'North Eastern', 'Oriya', 'Bihari', 'Afghani', 'Middle Eastern', 'Indonesian', 'Assamese', 'Andhra',
'Mangalorean', 'British', 'Malwani', 'Cuisine Varies', 'Turkish', 'Moroccan', 'Naga'] total no. of cuisines 86 Above you can see the list of all cuisines that are served by the restaurants that have been rated by Zomato. There are in total 86 types of cuisines available in total. It would be really interesting to see which cuisines are the most popular among restaurant goers according to Zomato. This is a little challenging as seen from the analysis before as the most popular restaurants are most likely a multi-cuisine restaurant. It is most likely that a combination of particular cuisines seems to be the key to a successful rating. It is also possible that the popular multi-cuisine restaurants have one or two common cuisines. Cuisines served can definitely provide a clue to understanding the popularity of a restaurant.

Average Ratings for Each Cuisine

Popularity for Each Cuisine
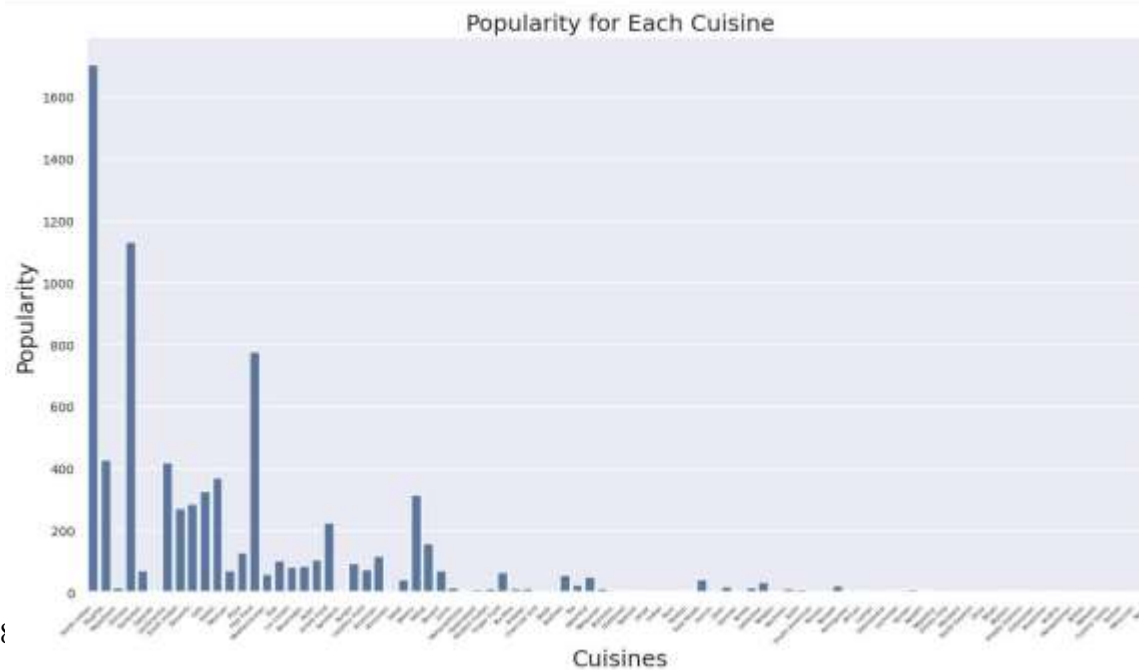
The above bar plot gives an idea about the ratings per cuisine. But it would be more interesting to know which type of cuisine is most likely consumed by the people rating on Zomato. But with this data we have to conclude that North Indian and Chinese cuisines are way more popular than other cuisines in India.

Machine Learning Model to get recommendations for the top 10 best restaurants.

The model uses random forest regressor to get the recommended data and trained on the final dataset that contains data with ratings of the restaurants with cuisines. It fetches 10 top restaurants with data provided. The demonstration and model is provided below:

```python
import pandas as pd
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split


# Assuming you have a new dataset called 'new_data' with columns ['Locality', 'Cuisines']
# Prepare the new data similar to the training data

new_x = data_rated_restaurants[['Restaurant Name','Locality', 'Cuisines']]
y = data_rated_restaurants['Aggregate rating']
transformed_new_x = transformer.transform(new_x)

# Predict ratings for new data
model = RandomForestRegressor(n_estimators = 50, random_state = 42)
x_train, x_test, y_train, y_test = train_test_split(transformed_new_x, y, test_size = 0.2)
model.fit(x_train, y_train)
predicted_ratings = model.predict(transformed_new_x)

# Combine predicted ratings with restaurant information
recommendations = pd.DataFrame({
    'Restaurant Name': new_x['Restaurant Name'],
    'Predicted_Rating': predicted_ratings
})

# Sort the recommendations by predicted rating
recommendations_sorted = recommendations.sort_values(by='Predicted_Rating', ascending=False)

# Print top recommendations
top_recommendations = recommendations_sorted.head(10)  # Change 10 to the number of recommendations you
print("Top Recommended Restaurants:")
print(top_recommendations)
```

Output:

```
Top Recommended Restaurants:
                 Restaurant Name  Predicted_Rating
3601   Indian Accent - The Manor             4.720
2307   AB's - Absolute Barbecues             4.704
822    AB's - Absolute Barbecues             4.696
2301   AB's - Absolute Barbecues             4.690
4298               Masala Library            4.606
809                       Chili's            4.576
1839                     Prankster           4.574
2495             Mirchi And Mime             4.510
2302                      Chili's            4.490
3732                       Tashan            4.474
```

Conclusion

With all the data analysis performed in this project leads to a better understanding of patterns and trends of popularity of restaurants in the country with different factors that are mentioned above, this data analysis shows how the popularity of north indian cuisine in undisputedly on top regarding other cuisines and the votes are significantly higher for the same. The data analysis depends on the current status of popular restaurants which offer various facilities to their consumers. Moreover, the data analysis also provides recommendations on the basis of predicted ratings, cuisines and locality of the restaurant. We have put a lot of effort into building this entire project of data analysis and machine learning and we hope this can bring growth to the sales of people who are looking to establish a new restaurant, cafe or food providing services in the current market situations.

Future Scope

We have analyzed the data as much as we could but this can be improved with better data provided and for future reference we can also implement more distinct data prediction and pattern analyzing models with neural networks and tensorflow modules in the Python Language.

Also we can use more features and factors that are not included in this project such as different prices for dishes and similar dishes but different prices of them. Location with nearby facilities such as metro, bus stands, auto

stands etc. We can include much more to analyze more deeply into the dataset. But with all our capabilities and time we did the best we could. We left some more room for the future and also we are working on the fine tuning of this project.

References

Books:

1. "Python for Data Analysis" by Wes McKinney - This book provides a comprehensive guide to data analysis using Python, including techniques for cleaning, manipulating, and analyzing data.
2. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron - This book covers essential machine learning algorithms and techniques using popular Python libraries like Scikit-Learn, Keras, and TensorFlow.
3. "Building Machine Learning Powered Applications: Going from Idea to Product" by Emmanuel Ameisen - This book focuses on the practical aspects of building machine learning applications, including data preparation, model deployment, and iteration.

Websites:

1. Kaggle (https://www.kaggle.com/) - Kaggle hosts various datasets and competitions related to data analysis and machine learning. You can find datasets related to restaurant reviews and recommendations on Zomato and other platforms.
2. Towards Data Science (https://towardsdatascience.com/) - This website offers a plethora of articles, tutorials, and case studies on data science and machine learning. You can find articles related to restaurant recommendation systems and data analysis projects.
3. Zomato API Documentation (https://developers.zomato.com/documentation) - Zomato provides an API that allows developers to access restaurant data, reviews, and other information. Understanding the API documentation can help you retrieve relevant data for your analysis.
4. GitHub (https://github.com/) - Many open-source projects related to data analysis and machine learning are hosted on GitHub. You can explore repositories containing code for building recommendation systems and analyzing restaurant data.

Academic Papers:

1. "Collaborative Filtering Recommender Systems" by Michael D. Ekstrand, John T. Riedl, and Joseph A. Konstan - This paper provides an overview of collaborative filtering techniques for building recommendation systems.
2. "Deep Learning for Recommender Systems: A Survey and New Perspectives" by Xavier Amatriain, Alejandro Jaimes - This survey paper explores the application of deep learning techniques in building recommender systems.