**Zomato Restaurant**

**Project Description**

Zomato Data Analysis is one of the most useful analysis for foodies who want to taste the best

cuisines of every part of the world which lies in their budget. This analysis is also for those who

want to find the value for money restaurants in various parts of the country for the cuisines.

Additionally, this analysis caters the needs of people who are striving to get the best cuisine of

the country and which locality of that country serves that cuisines with maximum number of

restaurants.

1 Problem statement

The quest begins with a clear understanding of the problem at hand: predicting average prices of restaurants listed on Zomato. This task is pivotal for both consumers seeking affordability and businesses aiming for competitive pricing strategies. By leveraging machine learning, we aim to develop models that provide accurate estimations of average prices based on various restaurant attributes.

2 Data Analysis

 It contains dimensional information of restaurants listed on the platform. The columns available are listed below.

1. Restaurant Id: Unique id of every restaurant across various cities of the world
2. Restaurant Name: Name of the restaurant
3. Country Code: Country in which restaurant is located
4. City: City in which restaurant is located
5. Address: Address of the restaurant
6. Locality: Location in the city
7. Locality Verbose: Detailed description of the locality
8. Longitude: Longitude coordinate of the restaurant’s location
9. Latitude: Latitude coordinate of the restaurant’s location
10. Cuisines: Cuisines offered by the restaurant
11. Average Cost for two: Cost for two people in different currencies (local currency)
12. Currency: Currency of the country
13. Has Table booking: yes/no
14. Has Online delivery: yes/ no
15. Is delivering: yes/ no
16. Switch to order menu: yes/no
17. Price range: range of price of food
18. Aggregate Rating: Average rating out of 5
19. Rating color: depending upon the average rating color
20. Rating text: text on the basis of rating of rating
21. Votes: Number of ratings given

## 3 EDA and Data visualization

From EDA we have obtained important insights as the current dataset has restaurants across 15 countries with about 90% observation for Indian and 5% USA respectively.

We observe that india has the largest Restaurants with 43 cities in India and 35 cities in USA and so on .

Also ,the most preferred Cuisine in India is North Indian cuisine preferred by people

USA restaurants have higher ratings compared with Indian restaurants, this could indicate a better service in the US. A major of Indian restaurant ratings are zero.Also, US customers have a higher vote count than Indians, reiterating the previous finding of ratings

Talking about Cuisines:

North Indian, Chinese, Fast food, Mughlai are a few popular cuisines in India, so new restaurants opening up in this space can face stiff competition from established restaurants. The potential for growth in south Indian cuisines seems good due to the lack of competition and variety of vegetarian dishes.

Americans love everything American – steak, seafood, burgers, BBQ apart from Mexican and Chinese food.

It seems that India has the lowest price for two amongst the 15 restaurants. 40% of restaurants across geographies fall into the cheaper price range of 1, with 32%, 14%, 6% of restaurants falling into 2,3,4 respectively, this is because Indian price range 1 restaurants are dominating the dataset. Similar distribution can be found for Indian restaurants as well.

For the US about 40% belong to price range 2, with 31, 25, and 5 belonging to 1,3,4 respectively.

Overall Philippines($233), the UK($138), and Singapore($206) have higher food prices(Price range 4) and Indonesia($5.3), Turkey($4.8) and Sri Lanka($6.3) (price range 2) have lower prices.

In India, for the Price ranges 1, 2, 3, 4 the average price is Rs 284, 620, 1258, 2583 respectively. The distribution is shifting from 200 on the left to about 2500 on the right.

The majority of ratings are between 2.5 to 3.7, and the rests are 0. As the rating increases, there is a slight shift in average cost for two as well. For ratings 0, the average cost is at about Rs 300-400 range but for rating 3.7 the average cost moves up to Rs 1000. There is a positive relationship between the two.

4 Pre processing Pipeline

The missing data was handled in dataset,while the categorical columns have been separated and further been encoded using Labelencoder technique for further model building.

Also,the outliers were been handledusing zscore method and the columns showing skewness >0.5 were handled using skewness removal technique.Standarisation techniques were applied to prepare data for Model Building.

5. Building machine learning Models

I have applied various machine learning algorithms such as linear regression, decision trees, random forests, and gradient boosting are employed to predict the Zomato average cost for Two. After cross-validation techniques optimize model performance., I choose the EXTRATREESEGRESSOR MODEL giving best accuracy with minimum error.

Also,hypertuned the model for best parameters to give best accuracy for prediction.

Additionally, ensemble methods are explored to enhance predictive accuracy.

6   Concluding Remarks

I performed data analysis and machine learning to predict two important characteristics of restaurants: the average cost for two and the price range. We explored the dataset, visualized key aspects, and built and evaluated regression and classification models. The EXTRA TREES Regressor model performed well in predicting the average cost for two, and the Random Forest Classifier showed good accuracy in predicting the price range. These models can be valuable for restaurant owners and customers in making informed decisions.