Restaurant Rating Prediction Project

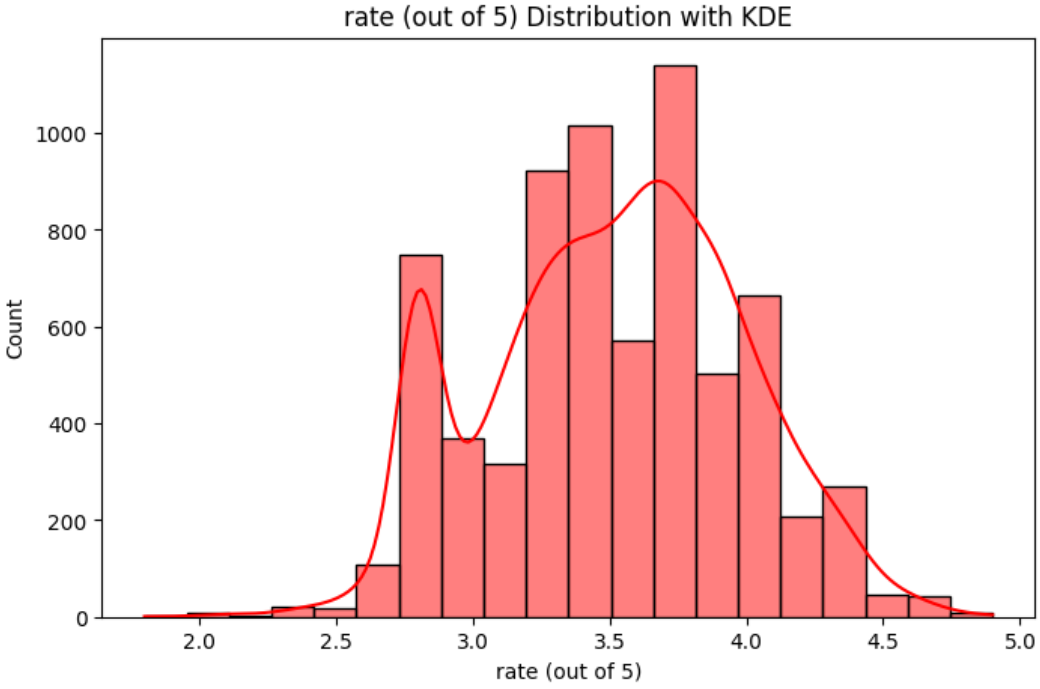
Name-Sunita Rai

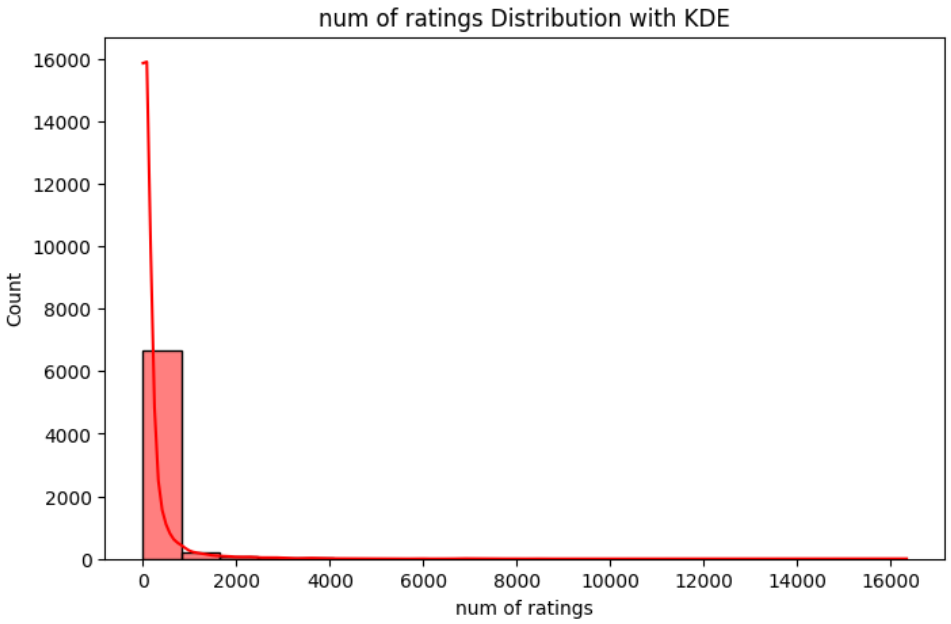
1.Dataset Description-The dataset contains information about the restaurant ,their features ,and customers rating. It includes details such as : Number of ratings, Average cost for two people, online ordering services and table booking, cuisines type. Here we analyze how features like price ,online services affect the ratings. The main goal is to predict a restaurant's rating (out of 5) based on these features.

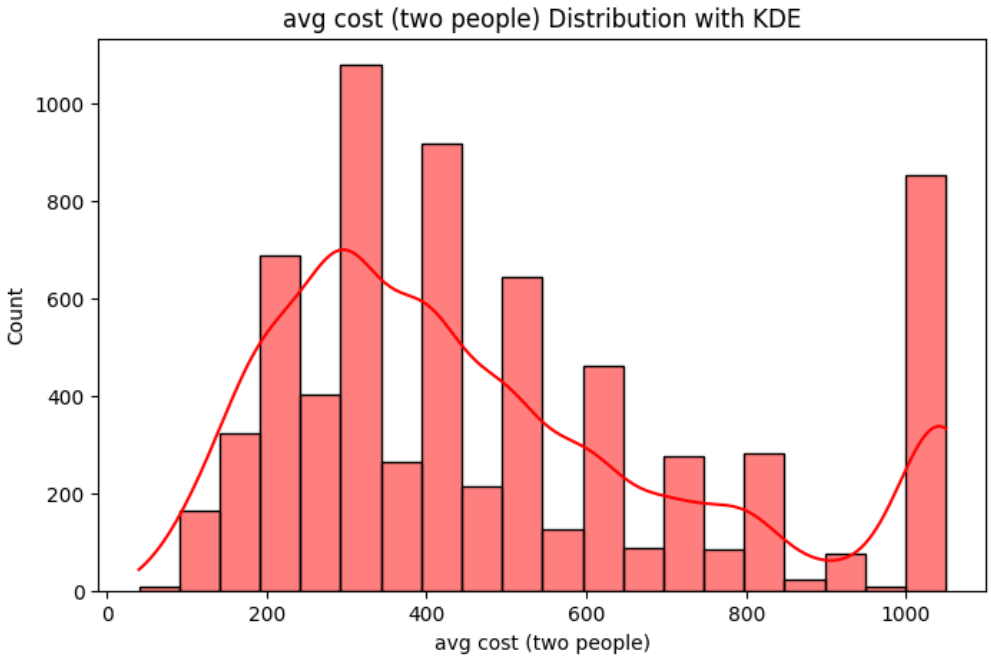
2.Data Cleaning and Transformation-.Rows with missing values are removed using dropna() inorder to improve the data quality.Unnecessary columns such as restaurant name and local address were removed.Categorical variables were encoded using OneHotEncoding. Feature engineering was performed by creating a popularity score and cost categories.

3.Exploratory Data Analysis(EDA)-EDA was performed using visualization such as histogram,boxplots,scatterplots, and heatmaps.

- Histogram shows that most of the ratings is in between 3.5 to 4. The tail on the left shows only a small number of poorly rated restaurants.
- Most of the restaurants are rated by very few people.
- Most restaurants cost 200–500 for two people i.e restaurants are affordable.Very few exceeds 800.

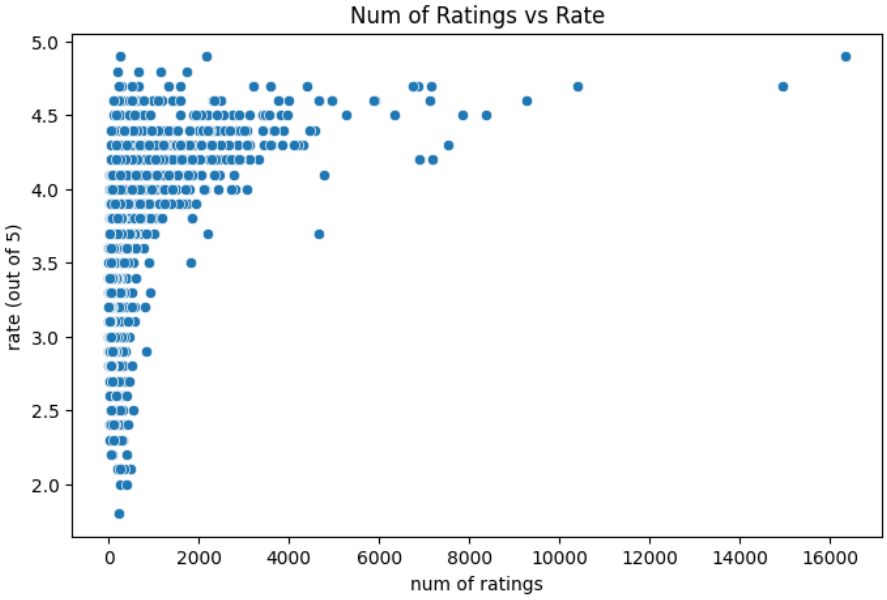


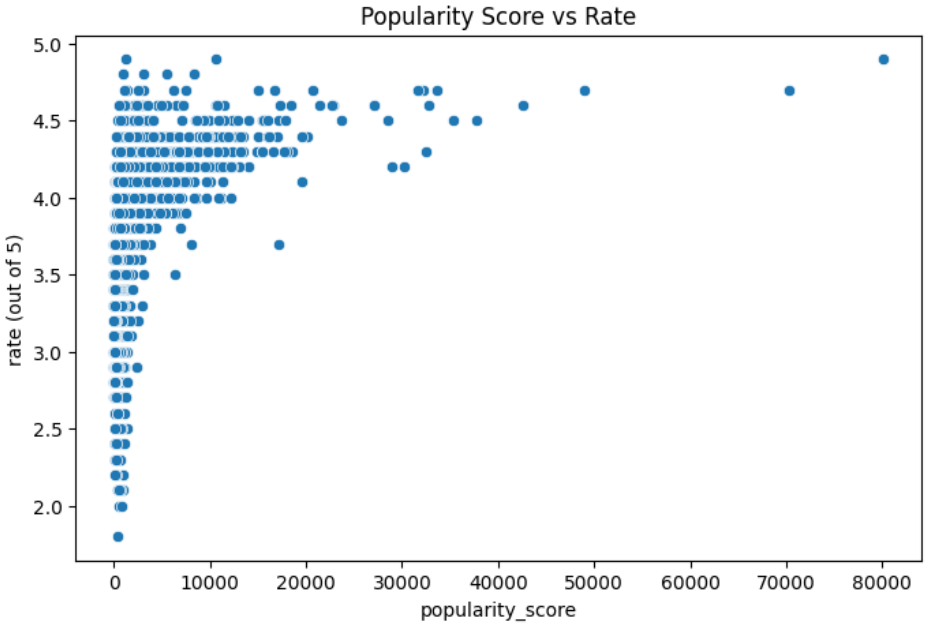


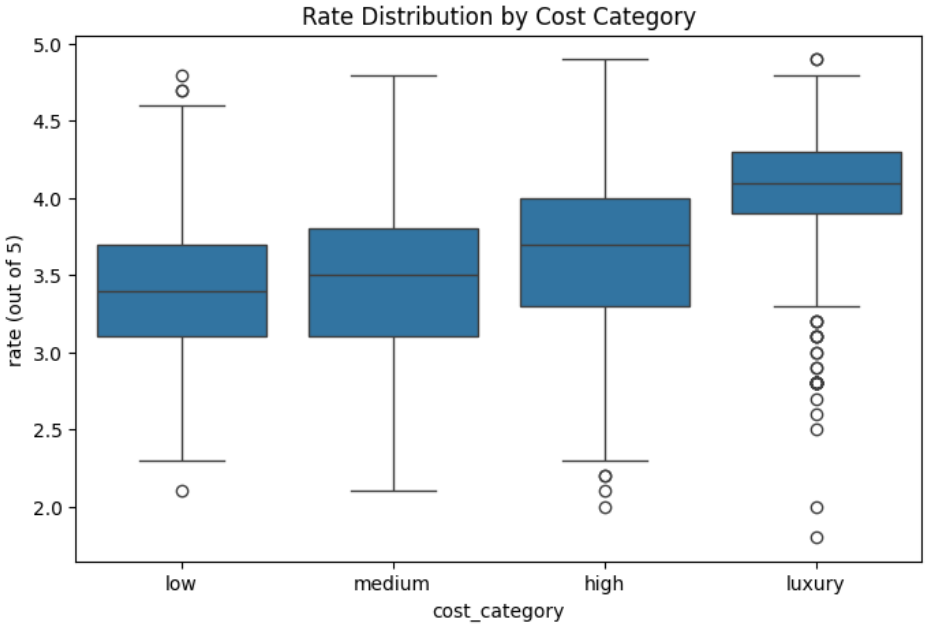


- num of ratings and popularity_score have a positive correlation (1), indicating that restaurants with more reviews are more popular.

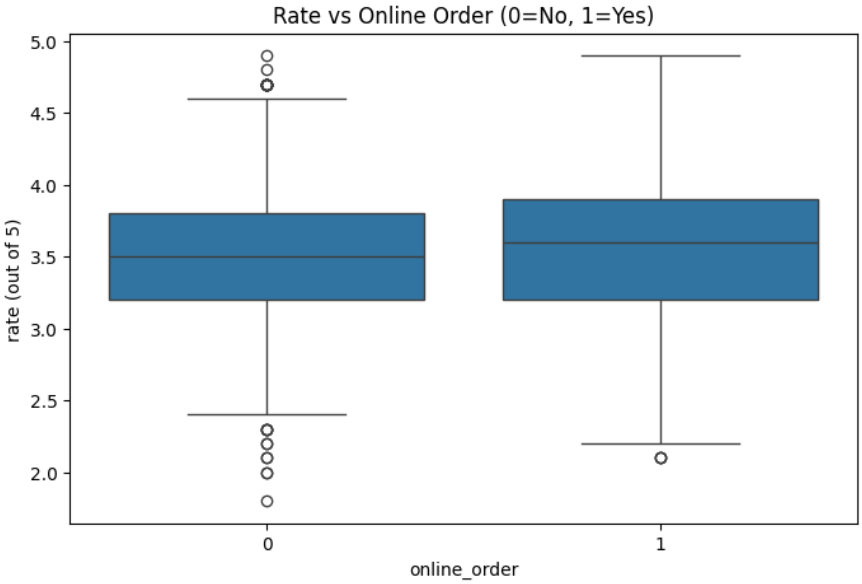
- we can observe that as popularity score increases ,the ratings are mostly high and consistent while less popular items have ratings vary more.
- Second graph shows that items with higher number of ratings usually have higher ratings while items with fewer ratings shows more variation in their ratings.



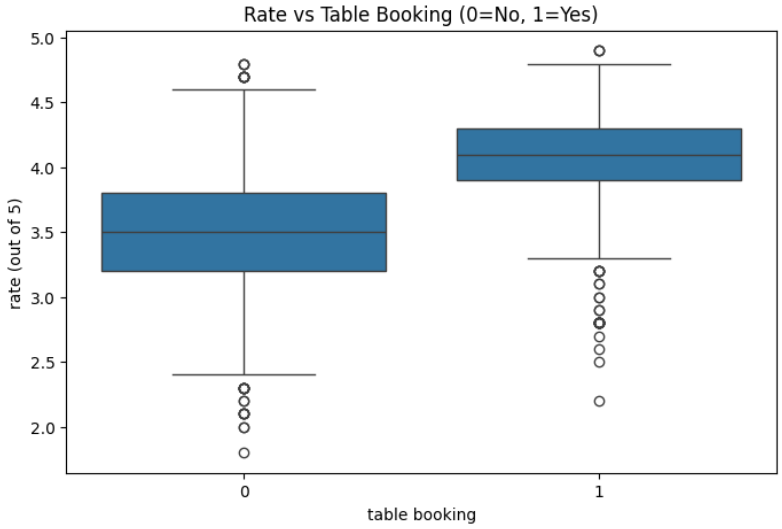




- This graph compares ratings of restaurant based on price categories;The line inside the box is average ratingThe box is where most ratings lies and The dots outside is unusual ratings which is outliers.It shows as cost increases the average rating also increases.



- It compares rating of restaurants that do not offer online order vs those that offer online order.
- 0-doesnot offer online order(rating is around 3.5)
- 1-does offer online order(rating is around 3.7)
- It shows restaurants that offer online ordering usually have better ratings.



- It compares ratings of restaurants that do not allow table booking vs those that allow.
  0-no table book(ratings are around 3.5)
  1-table book(ratings are around 4)
  It shows restaurants with table booking are rated much better.

4.Feature selection-The final features includes; 'num of ratings','avg cost (two people)', 'popularity_score', 'online_order', 'table booking', 'cost_category'.

5. Model Development-The dataset was split into training and testing sets using an 80-20 ratio. Linear Regression was used to predict restaurant ratings because the target variable is continuous.

6. Model Evaluation and Validation -Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared score. Cross-validation with 5 folds was used to ensure model stability and generalization.

7. Model Deployment-The trained model was deployed using Streamlit as a web application running on localhost. Users can input restaurant details and obtain real-time rating predictions through an interactive interface

8. Conclusion-This project demonstrates a complete restaurant rating prediction project that includes data preprocessing, exploratory analysis, feature engineering, model training, evaluation, and deployment.