Zomato Recommendation and Price Prediction System

Table of Contents

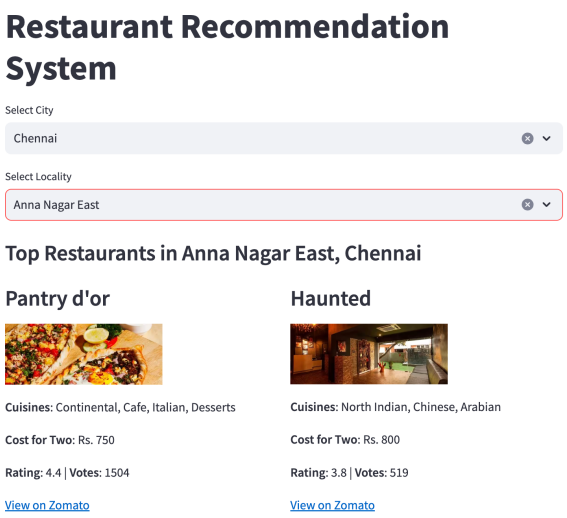
1. Introduction

This project aims to create a sophisticated Zomato-like restaurant recommendation and price prediction system. Leveraging machine learning algorithms and data visualization, the system empowers users to explore dining options based on location and cuisine while accurately predicting the cost for two people. Additionally, sentiment analysis provides insights into customer satisfaction through user ratings and reviews.

2. Project Overview

The project comprises three critical components:

- **Restaurant Recommendations:** An intelligent recommendation engine suggesting restaurants based on user preferences such as location and cuisine.



- **Cost Prediction:** A machine learning model that forecasts the average cost for two people, offering a glimpse into the affordability of dining experiences.

- Sentiment Analysis: Sentiment classification based on user-generated ratings and votes, providing valuable insights into customer satisfaction trends.

These components are presented in an interactive Streamlit web application, seamlessly deployed on AWS EC2, & RDS making it accessible to a wide audience.

3. Data Preprocessing

Initial Dataset Insights

The original dataset consisted of 29,753 records, capturing global restaurant information. The dataset included key features like restaurant names, location (city, locality), cuisines offered, ratings, votes, and cost for two. However, initial exploration revealed several inconsistencies and missing data points, necessitating a comprehensive preprocessing pipeline.

Rationale for Data Reduction

To build a reliable model, we chose to focus on regions with robust data. The global data was highly fragmented, with certain countries having only a single entry, making them unsuitable for modeling. After careful consideration, India was selected as the focal region due to its rich data, contributing the majority of records. This reduced the dataset to 8,672 records, allowing us to focus on data with high integrity and consistency.

Preprocessing Techniques

The following steps were employed to ensure the data was ready for analysis:

1. Duplicate Records: Identified and removed to avoid redundancy and ensure data integrity.

2. Handling Missing Values: Critical missing values in essential fields like cost, rating, and location were addressed by either removal or imputation.

3. Outlier Detection: Outliers in numeric fields, such as extremely high or low values for cost or votes, were examined and treated based on domain knowledge.

4. Data Filtering: Only records from India were retained to ensure the data was comprehensive enough for accurate model development.

5. Feature Selection: Non-essential fields were discarded, and key features like name, location, rating, and cost were retained for building models.

4. Model Development

Predictive Modeling for Cost Estimation

Goal: To estimate the average cost for two people at a restaurant.

Algorithm: A Random Forest Regressor was employed to harness the power of ensemble learning, ensuring robust predictions even with non-linear feature relationships.

Key Features: Location, Cuisines, Facilities, Ratings, and Votes.

Performance Metrics: RMSE and MAE were used to gauge the model's predictive accuracy, helping refine it for optimal performance.

Model Hypertuning: To enhance model performance, we utilized

Grid Search CV, a powerful technique for hyperparameter tuning. The grid search tested different combinations of parameters, such as `n_estimators`, `max_depth`, and `min_samples_split`. This ensured that the final model was optimally configured for high accuracy.

Sentiment Analysis

Objective: To categorize restaurants based on user sentiment derived from ratings and votes.

Methodology: Restaurants were classified into five distinct categories (Exceptional, Excellent, Good, Average, Poor) based on aggregate ratings, giving users a clear sense of customer satisfaction levels.

Visual Representations: Interactive charts and graphs were incorporated to illustrate trends in restaurant sentiment across various cities and localities.

5. Web Application Implementation

The core functionalities of this system were deployed in a sleek, user-friendly Streamlit web application hosted on AWS EC2 & RDS, ensuring accessibility and scalability.

Key components include:

- Dynamic Restaurant Recommendations
- Cost Prediction Interface
- Sentiment Analysis Dashboard.

6. Results and Performance Evaluation

Model Accuracy and Evaluation: The Random Forest Regressor achieved near-target performance, with **GridSearchCV** improved the model's accuracy, with the best hyperparameters leading an accuracy level of around 99% based on RMSE and MAE evaluations. The model demonstrated its ability to predict restaurant costs effectively, even in diverse localities with varying cuisines and services.

Application Usability and Engagement: The Streamlit application was designed with a focus on user engagement, providing seamless interactions for users seeking restaurant recommendations and cost predictions.

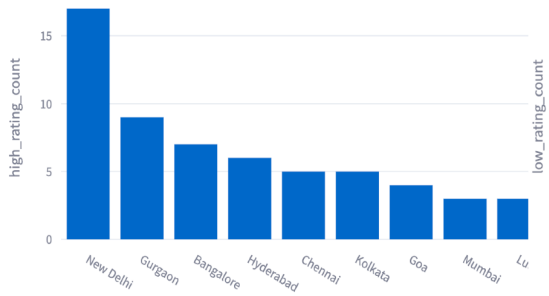
7. Conclusion and Future Scope

The Zomato Recommendation and Price Prediction System serves as a powerful tool for users looking to explore restaurants and estimate dining costs with precision. By leveraging machine learning and data analytics, the system offers a modern solution for both restaurant-goers and owners seeking operational insights.

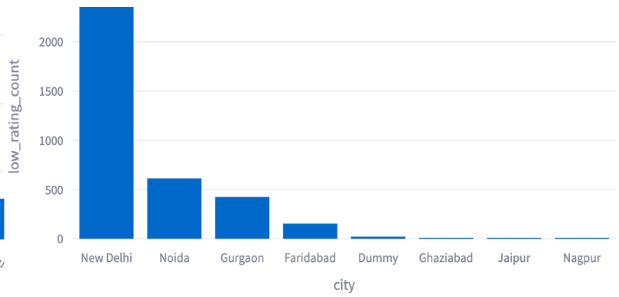
Future Enhancements could include expanding the dataset, NLP integration, and personalized recommendations.

8. Snippets

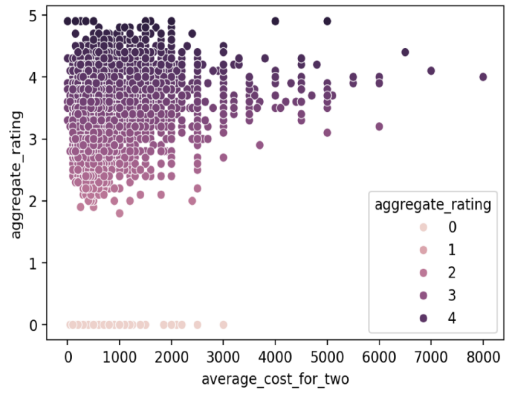
Cities with Most Highly Rated Restaurants



Cities with Most Low Rated Restaurants

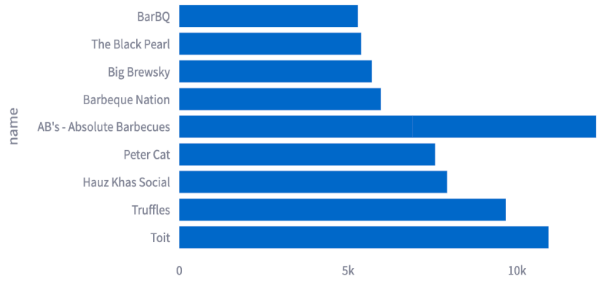


3. Average Cost for Two vs Aggregate Rating

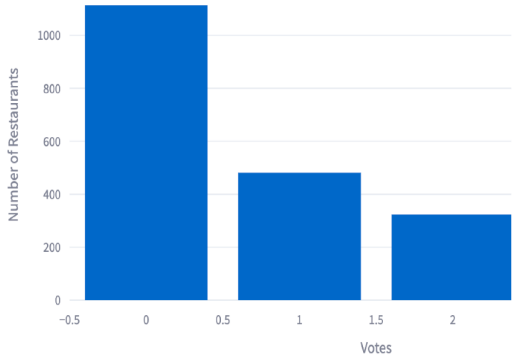


4. Top 10 Restaurants by Votes

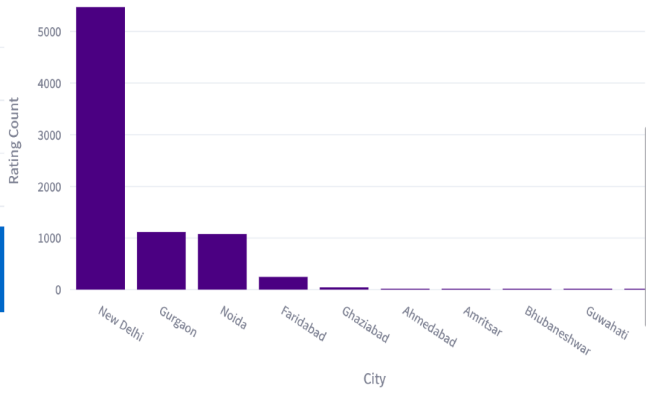
Top 10 Restaurants by Votes



Restaurants with Votes Between 0 to 3



Top 10 Cities with Most Ratings



Top 5 Restaurants with Rating 4.0 by City

	name	city	average_cost_for_two	aggregate_rating
0	Taj Bano - ITC Mughal	Agra	2,500	4
1	MoMo Cafe	Agra	2,000	4
2	Chokho Jeeman Marwari Jain Bhojanalya	Agra	400	4
3	Turquoise Villa	Ahmedabad	1,200	4
4	Cafe Alfresco	Ahmedabad	700	4
5	Ahuja Milk Bhandar	Amritsar	100	4
6	Glen's Bakehouse	Bangalore	800	4
7	10 Downing Street	Bhopal	1,000	4
8	Chi Kitchen	Bhopal	1,000	4
9	Aangan Horizon	Bhubaneshwa	900	4