ZOMATO RESTAURANT DATASET ANALYSIS USING PYTHON

A Comprehensive Study on Ratings, Pricing & Popularity Trends

Submitted By:
Mayank Kumar Jha
Bachelor of Computer Applications

Under the Guidance of:
Ms. Rajni

Department of Information, Communication & Technology

Tecnia Institute of Advance Studies
Guru Gobind Singh Indraprasth University

Abstract

This project focuses on analyzing the Zomato restaurant dataset using Python to extract meaningful insights about restaurant performance, customer ratings, cuisines, pricing, and delivery services. The objective of this analysis is to understand customer preferences and business trends in the food industry. Various data analysis techniques such as data cleaning, transformation, and visualization were applied using Python libraries including Pandas, NumPy, Matplotlib, and Seaborn. The project demonstrates how real-world datasets can be analyzed to support business decision-making and strategic planning.

Introduction

Zomato is one of the leading online food delivery and restaurant discovery platforms. It provides information about restaurants, customer reviews, ratings, cuisines, and pricing. With the growth of digital platforms, large volumes of customer data are generated daily.

Data analysis plays a crucial role in understanding customer behavior and improving business strategies. By analyzing the Zomato dataset, we can identify patterns such as:

- Most popular cuisines
- Restaurant rating trends
- Location-based restaurant distribution
- Price range comparison
- Impact of online delivery services

The main objective of this project is to perform exploratory data analysis (EDA) on the Zomato dataset and derive useful insights.

Problem Statement

The aim of this project is to analyze the Zomato restaurant dataset using Python and generate insights regarding customer ratings, pricing structure, restaurant locations, and cuisine popularity. The project also aims to identify business patterns that can help improve restaurant performance and customer satisfaction.

Dataset Description

The dataset used in this project contains information about various restaurants listed on Zomato. The dataset was obtained from Kaggle.

Key Features (Columns):

- Restaurant Name
- Location
- Cuisine
- Rating
- Votes
- Price Range
- Online Delivery (Yes/No)
- Table Booking (Yes/No)

Dataset Information:

- The dataset contains multiple rows representing restaurants.
- Each row contains details about one restaurant.
- Both numerical and categorical variables are present.

```
df= pd.read_csv('/kaggle/input/datasets/gauravkumar2525/zomato-restaurant-dataset/enhanced_zomato_dataset_clean.csv')
df.head()
```

	Restaurant_Name	Dining_Rating	Delivery_Rating	Dining_Votes	Delivery_Votes	Cuisine	Place_Name	City	Item_Name	Best_Seller	Votes	Prices	Average_Rating	Total_Votes	Price_per
0	Doner King	3.9	4.2	39	0	Fast Food	Malakpet	Hyderabad	Platter Kebab Combo	BESTSELLER	84	249.0	4.05	39	2.9
1	Doner King	3.9	4.2	39	0	Fast Food	Malakpet	Hyderabad	Chicken Rumali Shawarma	BESTSELLER	45	129.0	4.05	39	2.8
2	Doner King	3.9	4.2	39	0	Fast Food	Malakpet	Hyderabad	Chicken Tandoori Salad	NONE	39	189.0	4.05	39	4.7
3	Doner King	3.9	4.2	39	0	Fast Food	Malakpet	Hyderabad	Chicken BBQ Salad	BESTSELLER	43	189.0	4.05	39	4.2
4	Doner King	3.9	4.2	39	0	Fast Food	Malakpet	Hyderabad	Special Doner Wrap Combo	MUST TRY	31	205.0	4.05	39	6.4

Tools and Technologies Used

The following tools and libraries were used in this project:

- **Python** – Programming language
- **Kaggle Notebook** – Development environment
- **Pandas** – Data manipulation and analysis
- **NumPy** – Numerical computations
- **Matplotlib** – Data visualization
- **Seaborn** – Advanced visualization

These tools helped in cleaning, analyzing, and visualizing the dataset efficiently.

Data Preprocessing

Data preprocessing is an important step before analysis. The following steps were performed:

1. Handling Missing Values

- Missing values were identified using isnull() function.
- Rows with excessive null values were removed.
- Some missing values were filled using appropriate methods.

2. Removing Duplicates

- Duplicate entries were checked using duplicated() function.
- Duplicate rows were removed to avoid biased results.

3. Data Type Conversion

- Rating column was converted into numeric format.
- Categorical variables were properly formatted.

4. Cleaning Rating Column

- Some ratings were stored as strings.
- Special characters were removed.
- Converted ratings to float values for analysis.

Data preprocessing ensured the dataset was clean and ready for analysis.

Exploratory Data Analysis (EDA)

This section presents a detailed analysis of the Zomato dataset using various statistical and visualization techniques to extract meaningful business insights.

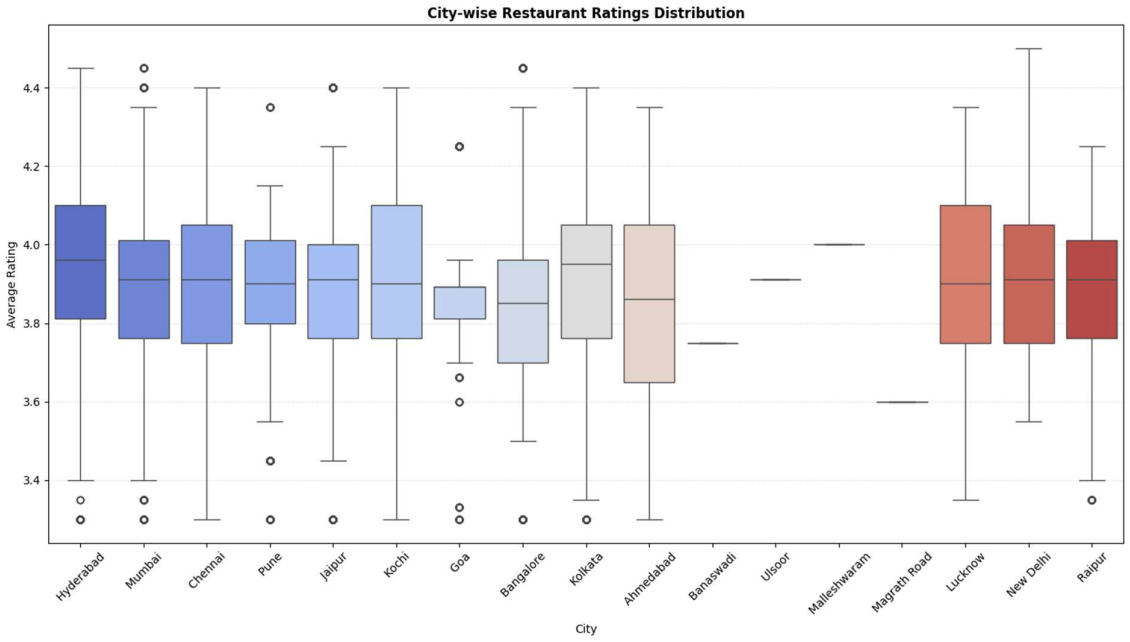
1. City-wise Restaurant Ratings Distribution

The rating distribution was analyzed across different cities to understand performance variation geographically.

Observation:

- Metropolitan cities show a higher concentration of restaurants with ratings above 4.0.
- Tier-2 cities show more variation in ratings.
- Some cities demonstrate consistent average ratings between 3.5–4.0, indicating stable service quality.

Insight:
Urban cities tend to maintain higher restaurant standards due to competition and customer expectations.



2. Most Popular Cuisines

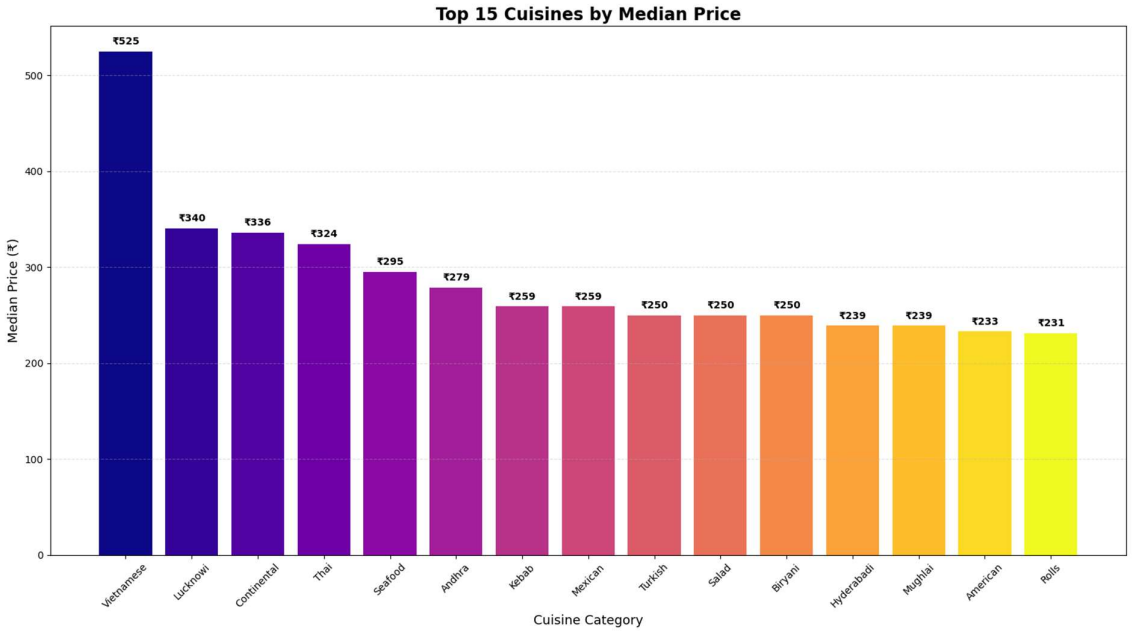
Cuisine frequency analysis was performed to identify customer preferences.

Observation:

- North Indian cuisine appears most frequently across cities.

- Chinese, Fast Food, and Continental cuisines follow closely.
- Multi-cuisine restaurants dominate the dataset.

Insight:
Traditional and fast-food cuisines drive the majority of restaurant business.



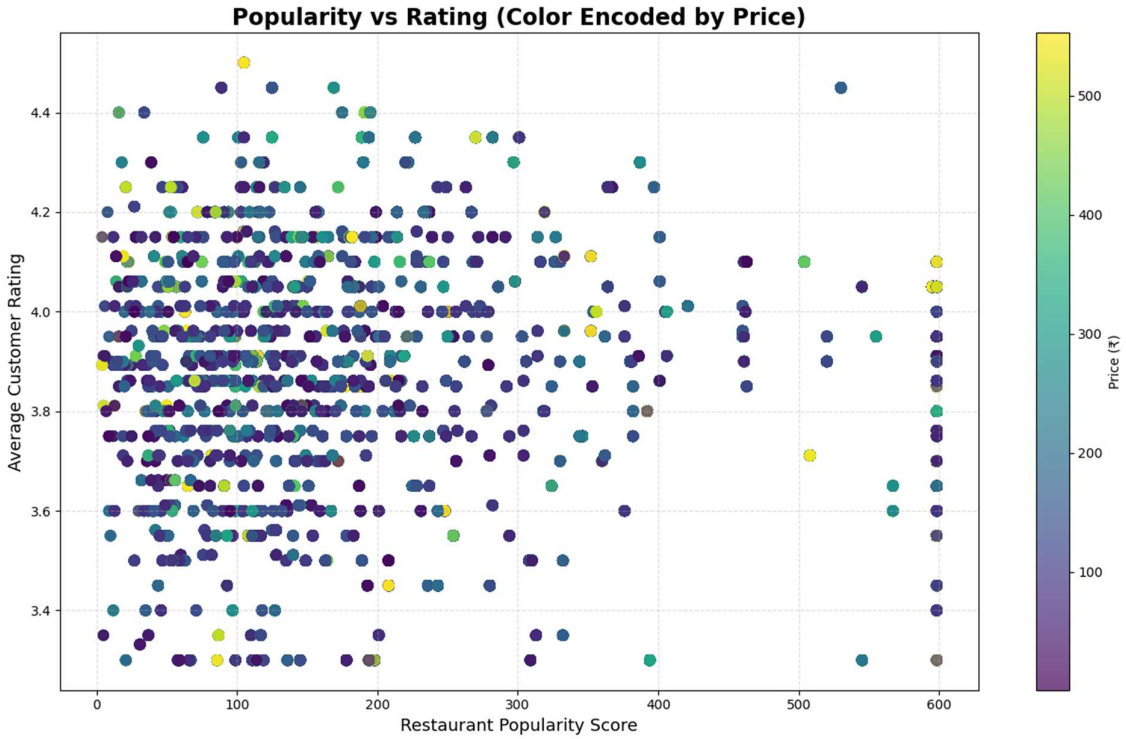
3. Popularity vs Rating

Popularity was measured using total votes and compared against ratings.

Observation:

- Restaurants with higher votes generally maintain ratings above 3.8.
- Some restaurants have high ratings but low votes (hidden gems).
- Popular restaurants tend to sustain quality service.

Insight:
Higher engagement often correlates with better service and consistent quality.



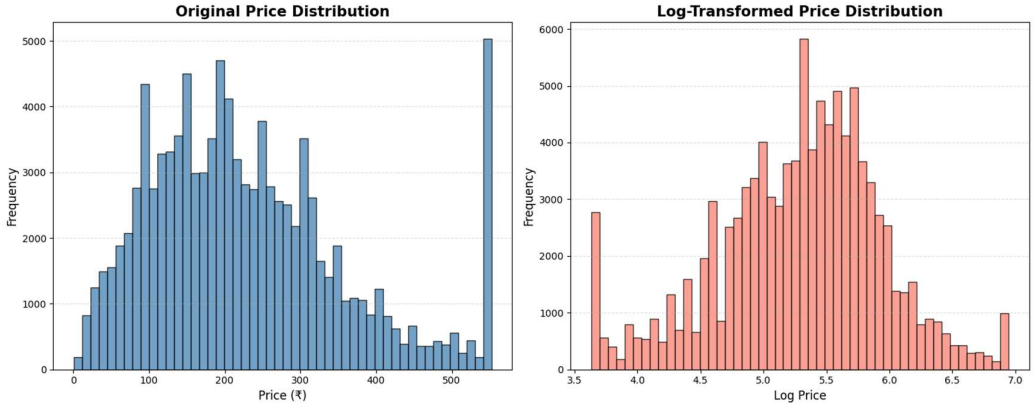
4. Original Price Distribution & Long-Term Price Distribution

Price data was analyzed to understand restaurant affordability.

Observation:

- Majority of restaurants fall in the medium price category.
- High-priced restaurants form a small segment.
- Long-term distribution suggests stable mid-range pricing dominance.

Insight:
Affordable and mid-range restaurants attract the highest customer base.



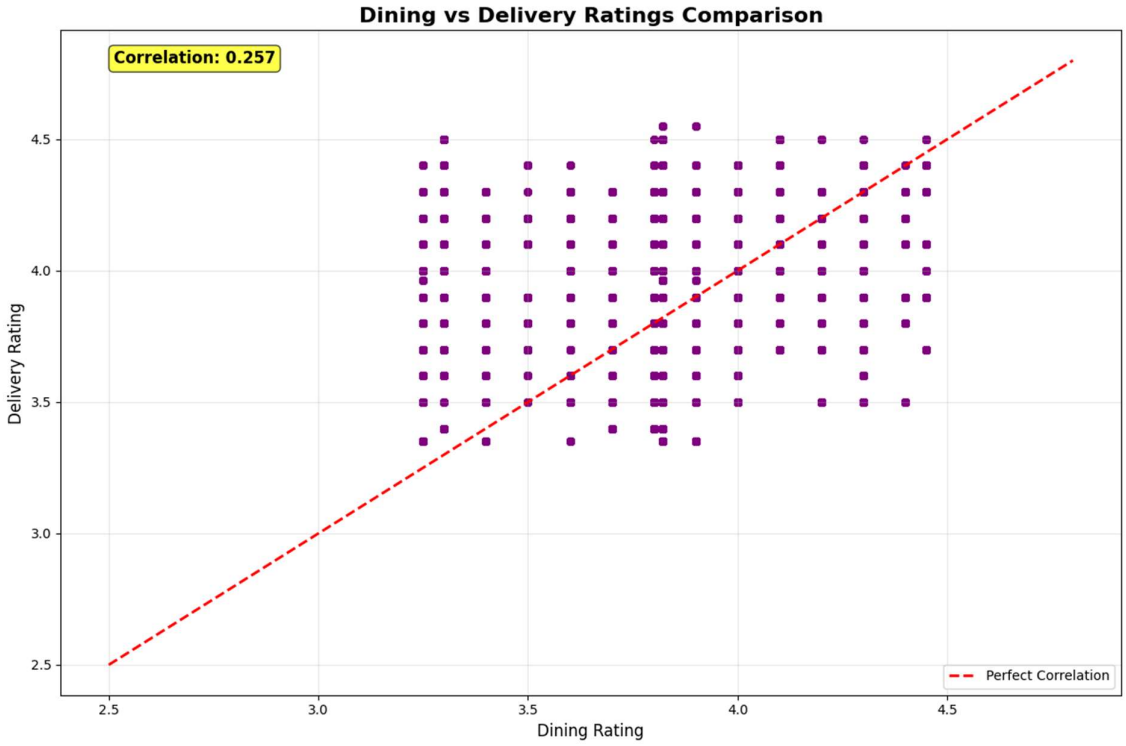
5. Dining vs Delivery Ratings Comparison

Ratings were compared between dining and delivery services.

Observation:

- Dining ratings are slightly higher than delivery ratings.
- Delivery ratings show more variability.
- Some restaurants perform equally well in both categories.

Insight:
Customer experience differs between dine-in and delivery services.



6. Relationship Between Votes and Price (Top 8 Cuisines)

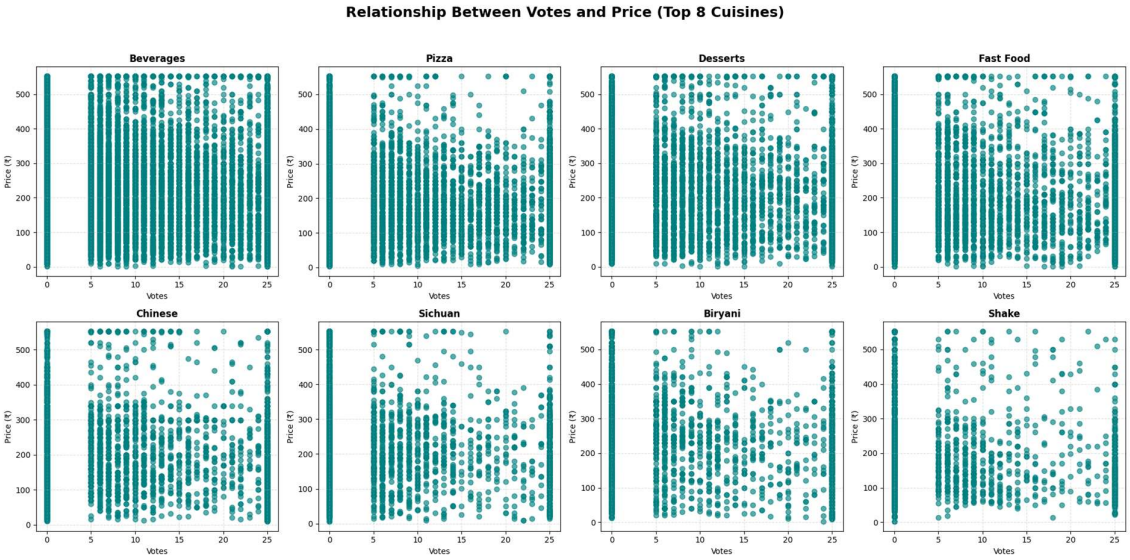
Votes and price relationship was studied for the top 8 cuisines.

Observation:

- Moderate price restaurants receive higher votes.
- Extremely high-priced restaurants receive fewer votes.
- Popular cuisines show balanced pricing and high engagement.

Insight:

Customers prefer value-for-money options.



7. Highly Rated vs Regular Restaurants – Performance Comparison

Restaurants were categorized based on rating threshold (e.g., above 4.2 as highly rated).

Observation:

- Highly rated restaurants attract significantly more votes.
- Regular restaurants dominate in number but not engagement.
- Highly rated restaurants show better price efficiency.

Insight:

High ratings directly impact popularity and customer trust.

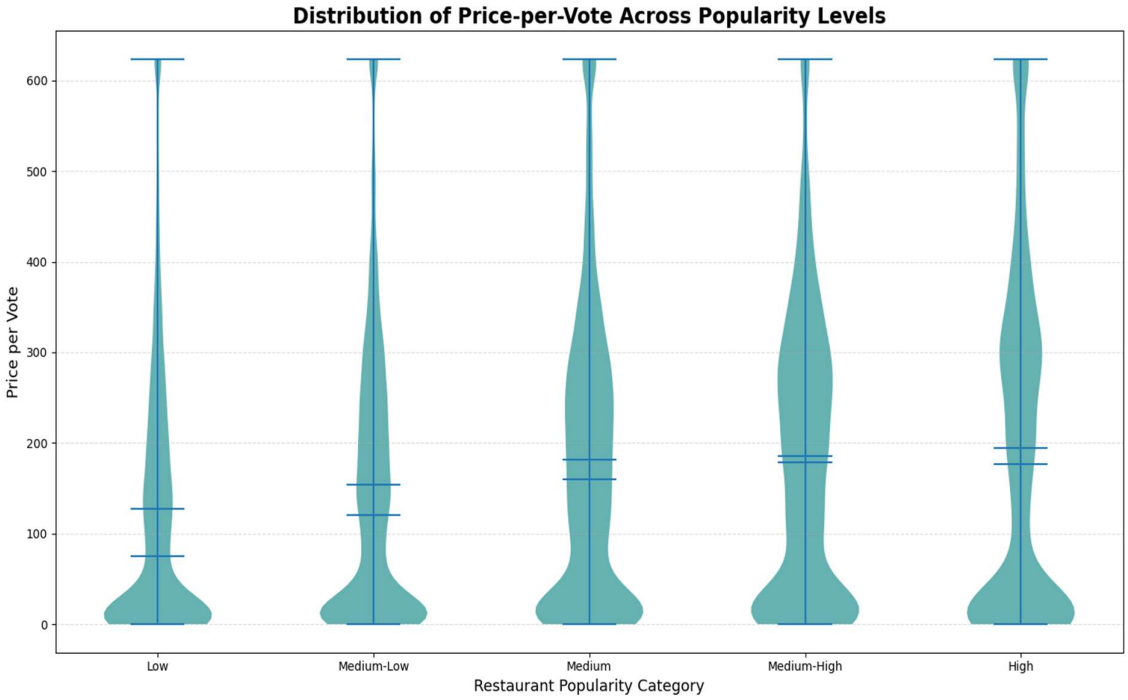
8. Distribution of Price-per-Vote Across Popularity Level

Price-per-vote metric was calculated to measure cost efficiency.

Observation:

- Highly popular restaurants have lower price-per-vote ratios.
- Less popular restaurants show inconsistent cost efficiency.

Insight:
Efficient pricing strategies increase engagement and customer satisfaction.



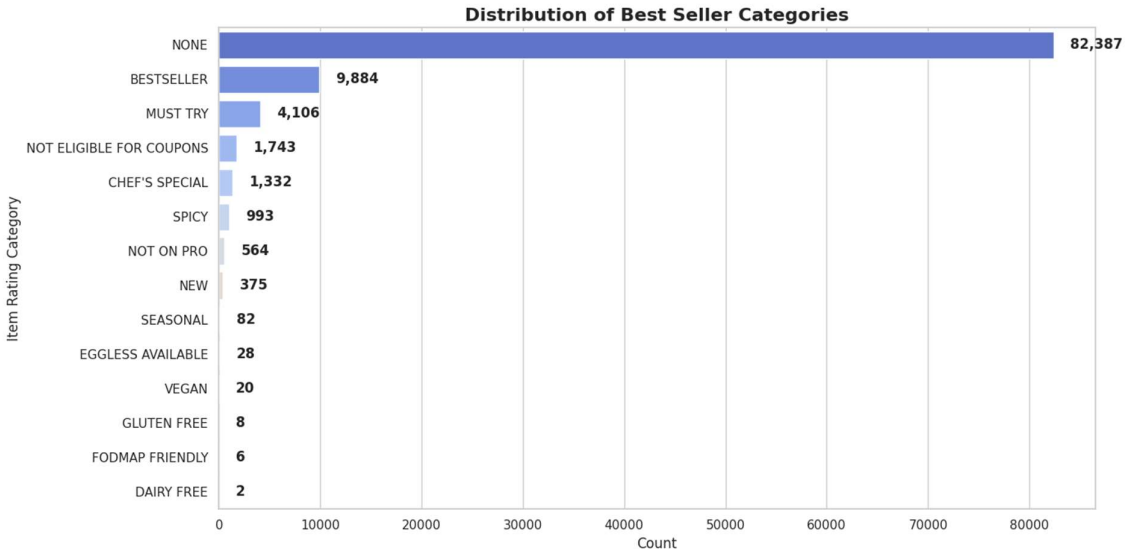
9. Distribution of Best Seller Categories

Best seller categories were analyzed to determine high-demand items.

Observation:

- Fast food and North Indian items dominate best seller categories.
- Premium cuisine categories appear less frequently.

Insight:
Mass-market cuisines outperform niche segments.



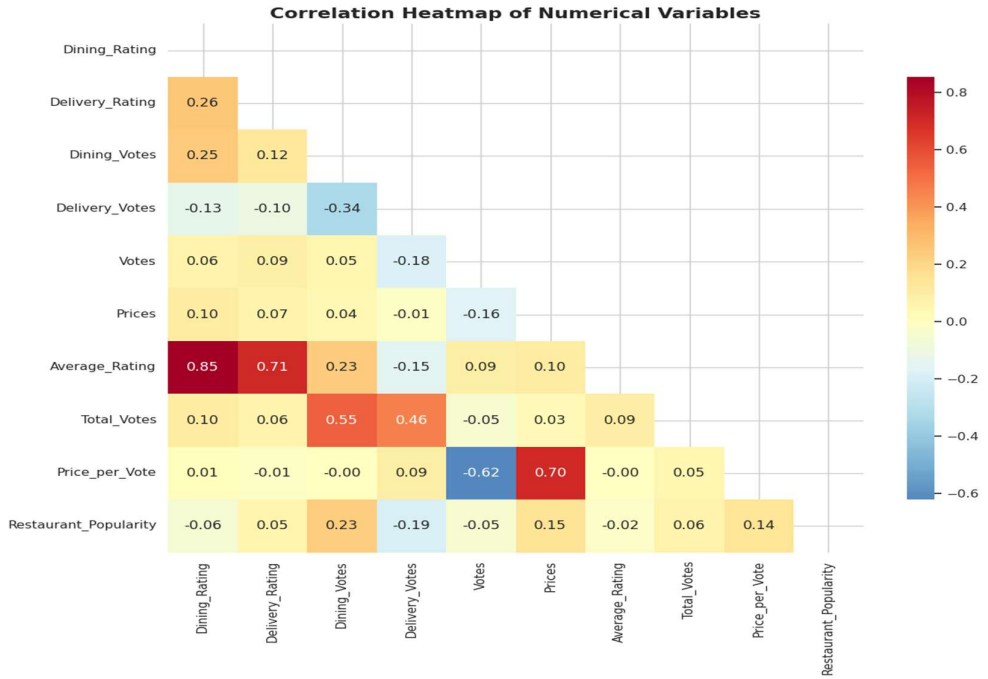
10. Correlation Heatmap of Numerical Variables

Correlation matrix was plotted among numerical variables.

Observation:

- Strong positive correlation between votes and popularity.
- Moderate correlation between price and rating.
- Weak correlation between rating and price-per-vote.

Insight:
Votes play a significant role in overall restaurant performance.



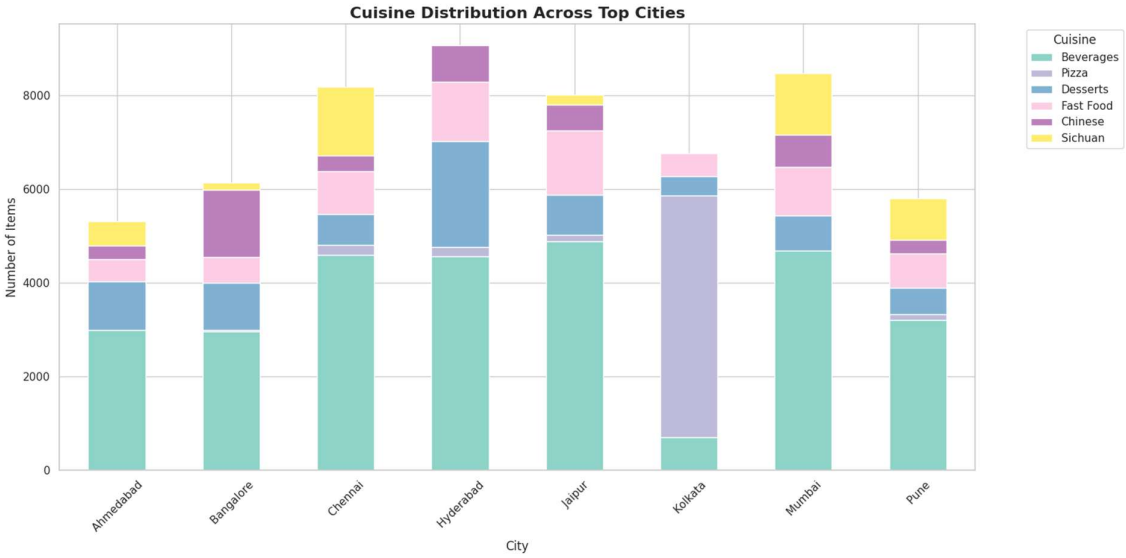
11. Cuisine Distribution Across Top Cities

Cuisine patterns were analyzed city-wise.

Observation:

- Major cities show diversified cuisine options.
- Smaller cities focus mainly on traditional cuisines.
- Multi-cuisine restaurants are common in metropolitan regions.

Insight:
Urban markets are more diversified in food preferences.



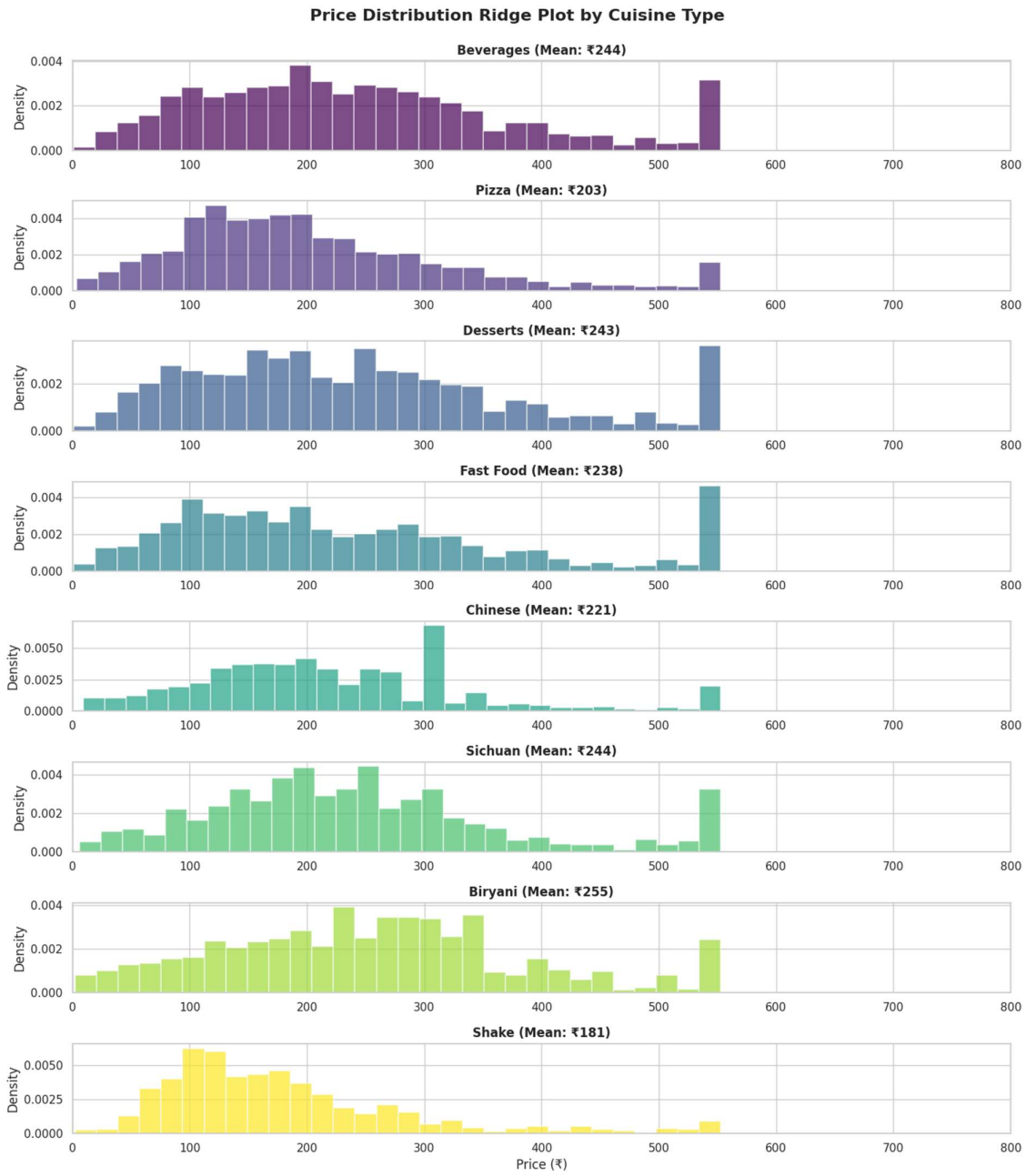
12. Price Distribution Ridge Plot by Cuisine Type

Ridge plot was used to compare price distribution across cuisines.

Observation:

- Continental cuisine shows higher price range.
- Fast food maintains lower price consistency.
- Indian cuisine spans across all price segments.

Insight:
Cuisine type significantly influences pricing strategy.



13. Restaurant Performance: Rating vs Price (Bubble Size = Votes)

Bubble chart was created to show multidimensional performance.

Observation:

- Most high-vote restaurants lie in mid-price, high-rating region.
- Very high-priced restaurants have smaller vote bubbles.
- Optimal performance lies in balanced pricing.

Insight:

Best-performing restaurants maintain quality with affordable pricing.

14. Rating Distribution Across Cities

City-level rating boxplots were analyzed.

Observation:

- Some cities show narrow rating ranges (consistent quality).
- Others show wide spread (mixed performance).

Insight:

Consistency varies significantly across regions.

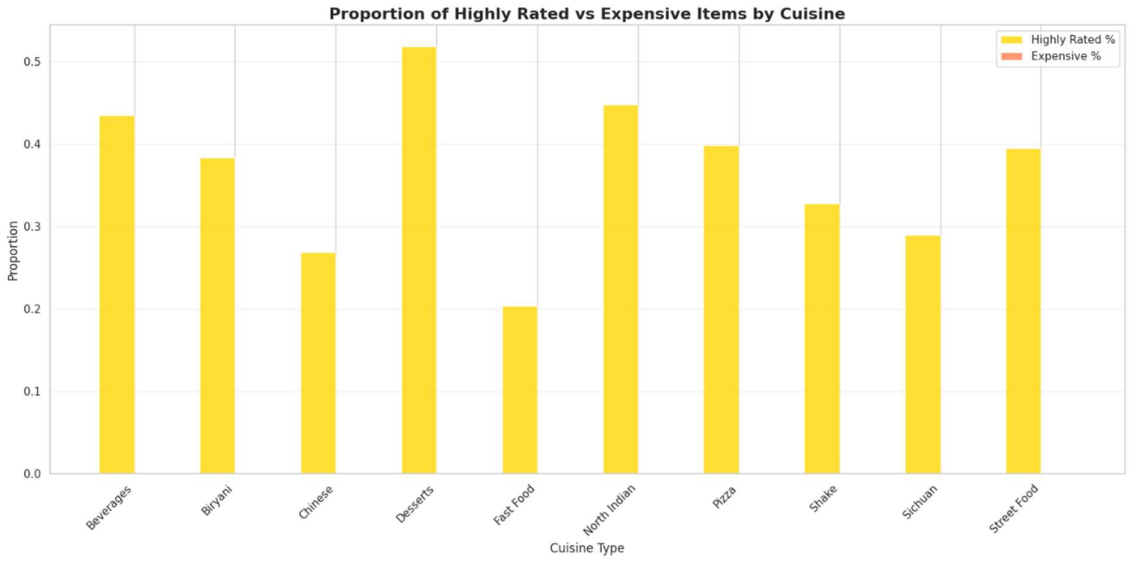
15. Proportion of Highly Rated vs Expensive Items by Cuisine

Cuisine categories were analyzed based on rating and pricing.

Observation:

- High-end cuisines tend to have higher ratings.
- Budget cuisines dominate in number but moderate ratings.

Insight:
Premium segments often focus on service quality.



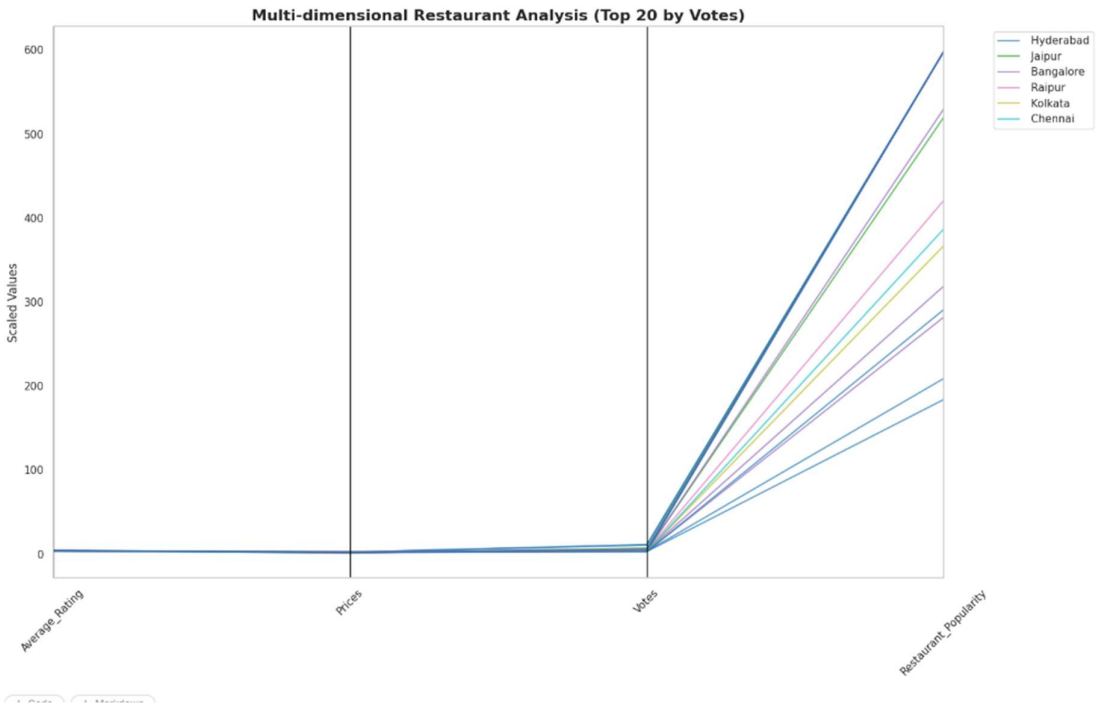
16. Multi-dimensional Restaurant Analysis (Top 20 by Votes)

Top 20 restaurants were analyzed based on votes, rating, and price.

Observation:

- Most top-voted restaurants maintain rating above 4.0.
- Price remains in mid to upper range.
- These restaurants show balanced performance metrics.

Insight:
Successful restaurants balance price, quality, and popularity.



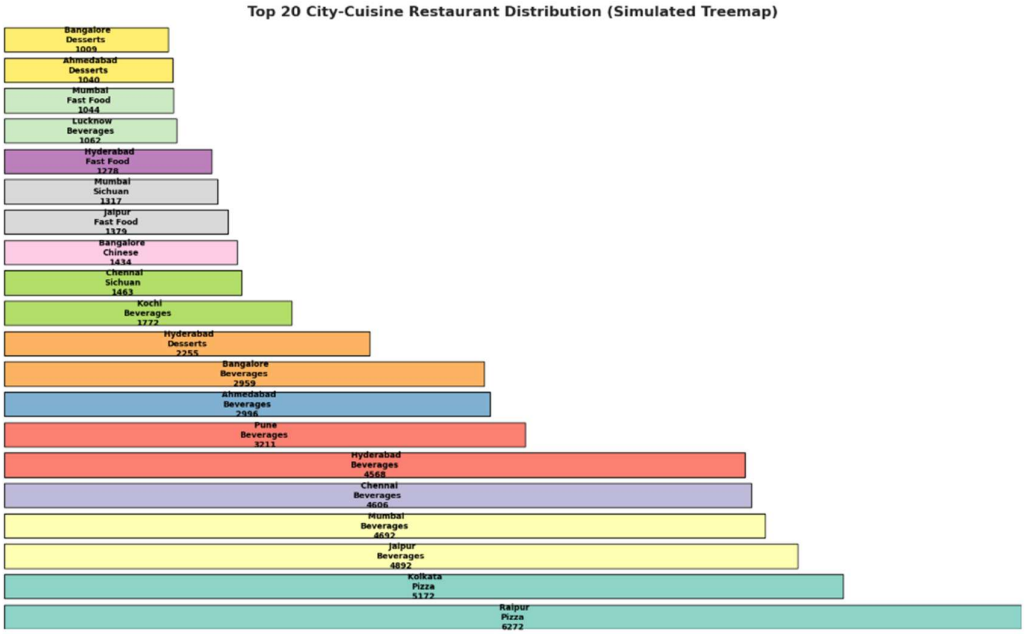
17. Top 20 City-Cuisine Restaurant Distribution (Simulated Treemap)

Treemap analysis shows distribution across city-cuisine combinations.

Observation:

- Certain city-cuisine combinations dominate the dataset.
- Major cities show heavy clustering of specific cuisines.

Insight:
Cuisine specialization varies by region.



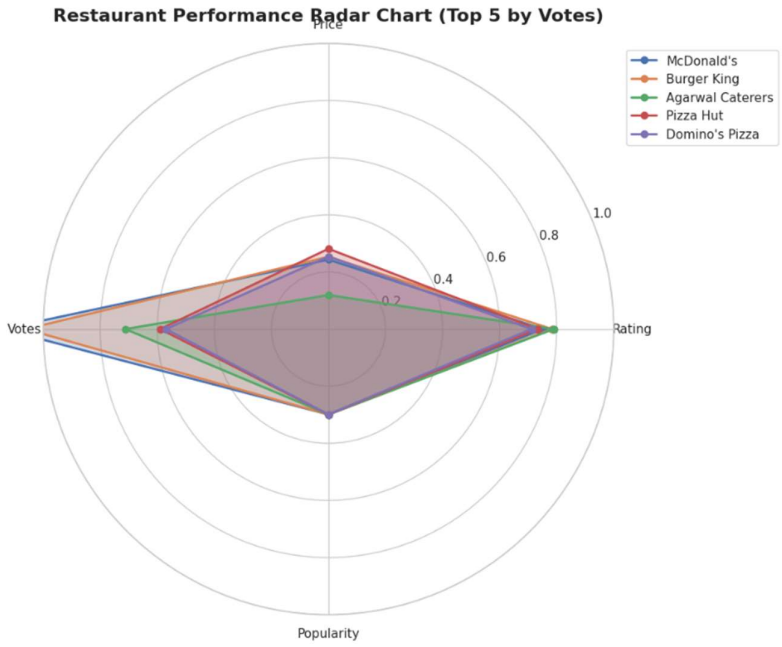
18. Restaurant Performance Radar Chart (Top 5 by Votes)

Radar chart compared top 5 restaurants across multiple metrics.

Observation:

- Top restaurants perform consistently across rating, votes, and pricing.
- Some excel in popularity but not in rating.

Insight:
Balanced performance leads to long-term success.



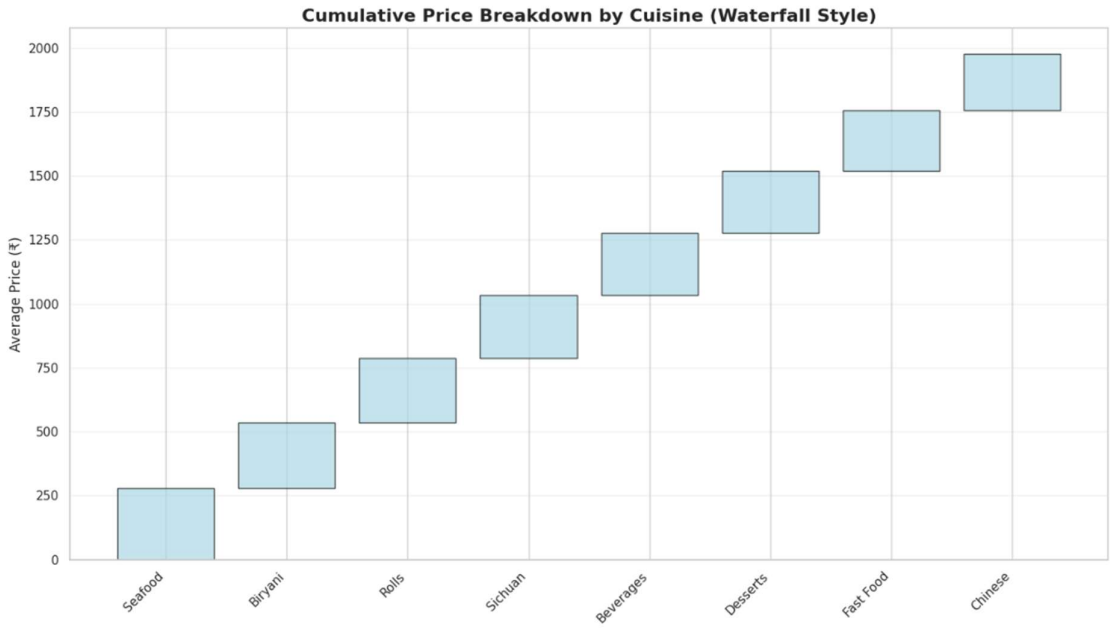
19. Cumulative Price Breakdown by Cuisine (Waterfall Style)

Waterfall visualization shows price contribution by cuisine.

Observation:

- Major cuisines contribute highest cumulative price volume.
- Premium cuisines contribute smaller but high-value segments.

Insight:
Revenue is volume-driven rather than high-price driven.



Key Insights and Findings

Based on the analysis, the following insights were derived:

- Most restaurants maintain ratings above 3.5, indicating overall good service quality.
- North Indian cuisine is the most popular among customers.
- Urban areas have a higher concentration of restaurants.
- Online delivery services significantly impact business growth.
- Medium-priced restaurants attract the majority of customers.

These insights can help businesses optimize pricing, service quality, and delivery strategies.

Conclusion

The Zomato dataset analysis provided valuable insights into restaurant trends, customer preferences, and pricing patterns. Through data cleaning and visualization techniques, meaningful patterns were identified that can help improve business strategies in the food industry.

This project demonstrates the practical application of Python in real-world data analysis tasks. It highlights how data-driven decisions can enhance business growth and customer satisfaction.

Future Scope

This project can be extended further by:

- Applying Machine Learning models to predict restaurant ratings.
- Building a restaurant recommendation system.
- Performing sentiment analysis on customer reviews.
- Creating an interactive dashboard using Power BI or Tableau.
- Developing a web-based analytics application.

References

- Kaggle Zomato Dataset
- Python Official Documentation
- Pandas Documentation
- Matplotlib & Seaborn Documentation

Contact Information

Mayank Jha
Email: mayankjha2525@gmail.com
LinkedIn: linkedin.com/in/mayank-jha-046085330
GitHub: github.com/mayank18-jha