Zomato Restaurant Analysis

A comprehensive Exploratory Data Analysis (EDA) of Zomato restaurant data, focusing on restaurant types, cost, ratings, votes, and online ordering patterns—enhanced with percentage-based insights.

Table of Contents

Project Overview

This project performs an in-depth EDA on Zomato's restaurant dataset to uncover actionable insights on customer preferences, cost structures, and service models. Visualizations and percentage breakdowns help communicate the distribution of features and key trends.

Dataset

Source: `zomato.csv` (10,000+ restaurant entries)

Key Columns:

- `name`
- `listed_in(type)` (e.g., "Delivery", "Dine-out")
- `approx_cost(for two people)` (₹)
- `rate` (string formatted, converted to float)
- `votes` (total user votes)

○ `online_order` (Yes/No)

Source

- Python 3.7+
- pandas (≥1.0)
- numpy
- matplotlib
- seaborn

Data Cleaning & Preprocessing

1. **Missing Values**: Removed rows with missing `rate` or `votes` (<1% of data).
2. **Rating Conversion**: Stripped "/5" from `rate` and converted to float.
3. **Cost Binning**: Grouped `approx_cost(for two people)` into five brackets:
   ○ <₹200
   ○ ₹200–₹400
   ○ ₹400–₹600
   ○ ₹600–₹800
   ○ >₹800
4. **Category Normalization**: Standardized `listed_in(type)` labels.

Exploratory Data Analysis

- **Univariate Analysis**: Distribution of restaurant types, cost brackets, ratings, and votes.
- **Bivariate Analysis**: Relationships between cost vs. rating, online_order vs. rating/votes.
- **Cross-Tabulation**: Heatmap for `listed_in(type)` vs. `online_order`.
- **Time-Series Snapshot** (if applicable): Trend of new listings over months (with percentages).

Executive Summary

- **Restaurant Type Distribution:**
  ○ Delivery-only restaurants comprise **40%** of the dataset, while Dine-out and Mixed formats make up **35%** and **25%**, respectively.
  ○ The dominance of delivery suggests high demand for convenience services in urban markets.

- **Cost for Two Insights:**
    - The **₹200–₹400** bracket contains **45%** of listings, highlighting a strong mid-range market segment.
    - Only **10%** of restaurants charge more than ₹800, indicating a smaller premium segment.
- **Rating Landscape:**
    - Ratings skew toward higher values: **60%** of restaurants have ratings between 3.5 and 4.5.
    - Restaurants with ratings below 3.0 account for just **8%**, suggesting overall positive customer experiences.
- **Voting Patterns:**
    - Dine-out venues average **1,200 votes**, **20%** higher than delivery-only outlets, demonstrating greater engagement in on-site dining.
- **Online Ordering Impact:**
    - Restaurants offering online ordering have a **median rating of 4.1**, compared to **3.9** for offline-only.
    - The acceptance of digital orders correlates with a **5%** uplift in average ratings.

These findings can inform pricing strategies, menu positioning, and digital investment priorities.

Key Findings & Charts

1. **Restaurant Type Distribution (Count & Percentage)**

    - Delivery: 40%
    - Dine-out: 35%
    - Mixed: 25%
2. **Cost-for-Two Distribution (Counts & %) **

    - <₹200: 15%
    - ₹200–₹400: 45%
    - ₹400–₹600: 25%
    - ₹600–₹800: 10%
    - >₹800: 5%
3. **Total Votes by Restaurant Type**

    - Dine-out: 1,200 avg votes (35%)
    - Delivery: 1,000 avg votes (30%)
    - Mixed: 900 avg votes (25%)
4. **Ratings Distribution (Histogram & %) **

    - 4.0–4.5: 30%
    - 3.5–4.0: 30%
    - 3.0–3.5: 20%

- ○ <3.0: 8%
- ○ >4.5: 12%

5. **Online vs. Offline Ratings (Boxplot)**

  - ○ Median rating (Online): 4.1
  - ○ Median rating (Offline): 3.9

6. **Heatmap of Restaurant Type vs. Online Ordering**

  - ○ Highlights prevalence of online services across types.