Exploratory Data Analysis: Zomato Dataset

Praveen Kumar Chandaliya

August 7, 2025

Zomato Dataset Overview

- Source: Zomato Restaurant Data[1] (via API or Kaggle)
- Total Entries: ~9500+ restaurants
- Contains restaurant-level metadata for various cities
- Useful for urban food service analytics and recommendation systems

Column Name	Description
Restaurant ID	Unique identifier for each restaurant
Restaurant Name	Name of the restaurant
City	Location city of the restaurant
Cuisines	List of cuisines offered
Average Cost for two	Estimated cost for two people
Currency	Currency in which cost is reported
Has Table booking	Yes/No – table reservation available
Has Online delivery	Yes/No – delivery option available
Aggregate rating	Overall rating (0.0–5.0)
Rating color	Visual category of rating (e.g., green, red)
Rating text	Textual sentiment (e.g., Excellent, Poor)
Votes	Number of user ratings received
Geo location	Latitude and Longitude (Geo info)

[1]https:
//github.com/krishnaik06/5-Days-Live-EDA-and-Feature-Engineering

Step 1: Data Cleaning

- Drop duplicates and unnecessary columns
- Handle missing values in key columns like `Cuisines`
- Standardize column names and formats
- Convert data types (e.g., `Votes` to int)

Step 2: Country and City Mapping

- Use mapping files to translate country codes into names
- Identify top cities with highest restaurant counts
- Filter for country-specific analysis (e.g., India)

Step 3: Univariate Analysis

- Distribution of Aggregate Ratings
- Count of each Rating Text and Rating Color
- Histogram of Votes
- Distribution of Average Cost for Two

Step 4: Bivariate Analysis

- Rating vs Votes: Are higher-rated restaurants more popular?
- Rating vs Cost: Are expensive restaurants better rated?
- Rating vs Online Delivery availability
- Rating vs Table Booking availability

Step 5: Multivariate and City-Level Insights

- Average Rating by City
- Top Cities by Restaurant Count
- Locality-based popularity and delivery trends

Step 6: Cuisine-Based Analysis

- Top 10 Most Common Cuisines
- Cuisine vs Rating: Which cuisines are most appreciated?
- Cuisine vs Cost: Which cuisines are more expensive?

Step 7: Price vs Rating Analysis

- Scatter plot of Average Cost vs Aggregate Rating
- Identify expensive but poorly rated restaurants
- Are budget-friendly options well rated?

Step 8: Feature Engineering Ideas

- Flags for: high rated, expensive, delivery available
- Cuisine count: number of cuisines per restaurant
- Multi-label encoding for cuisines
- Clustering by locality and ratings

Step 9: Correlation Analysis

- Correlation among: Rating, Votes, Price Range, Cost
- Heatmap for numeric feature relationships
- Identify multicollinearity or trends

Step 10: Geospatial Mapping

- Use Latitude and Longitude for plotting
- Visualize popular food areas on map (e.g., with Folium)
- Compare city clusters based on rating and delivery

Key Questions Answered

- Top cities: Delhi NCR, Bangalore, Mumbai
- Popular cuisines: North Indian, Chinese, Fast Food
- Online delivery is more common in metro cities
- Expensive restaurants aren't always higher-rated
- Highly rated restaurants attract more votes
- Which cities have the most restaurants listed?
- Are expensive restaurants better rated?
- Does online delivery impact rating?
- What cuisines are most popular or best rated?
- Which locality has the best-rated or most voted restaurants?