

# Report



**Project Title:** Exploratory Data Analysis

**Date:** 24/12/2025

**Team Leader :** Vaishnavi / Samruddhi

**Deadline:** 24/12/2026

## Project Overview

This project involves conducting Exploratory Data Analysis on the Zomato dataset to analyze country-wise restaurant distribution, rating patterns, customer engagement, and currency usage. The goal is to extract meaningful insights that can support data-driven decision-making and dashboard creation.

## Objectives

- Understand the structure and key attributes of the dataset
- Analyze country-wise restaurant distribution
- Examine customer rating patterns and engagement
- Identify countries with zero ratings
- Study currency usage across different countries
- Derive actionable business insights

## Dataset Overview

- The dataset includes restaurant-level information such as:
- Country
  - Aggregate Rating
  - Rating Count
  - Rating Color
  - Currency

## Visualization Insights

### Visualization Insights

- Bar plots were used to analyze rating count against aggregate rating
- Higher ratings correspond to higher rating counts

### Observation:

- Customer engagement increases with better ratings.

## Key Insights Summary

- India dominates the dataset in terms of records
- Some countries show zero ratings indicating low engagement
- Currency usage is country-specific
- Higher ratings lead to higher customer participation

## Conclusion

- The Exploratory Data Analysis provided valuable insights into Zomato's global presence, rating behavior, and country-wise characteristics.
- These insights form a strong foundation for dashboard creation, advanced analytics, and business decision-making.

## Next Steps

- Develop interactive dashboards using Power BI or Tableau
- Perform deeper country-specific analysis
- Analyze relationships between ratings and restaurant attributes
- Create executive-level summary visuals



+810-465-6237



aman.data.acc@gmail.com



# Project Overview

This project involves conducting Exploratory Data Analysis (EDA) on the Zomato dataset to gain a comprehensive understanding of restaurant distribution and customer behavior across different countries. The analysis focuses on examining country-wise restaurant presence, customer rating patterns, levels of user engagement, and currency usage associated with each country.

The primary objective of this project is to explore the dataset systematically, identify key trends and patterns, and uncover meaningful insights that reflect customer preferences and platform performance across regions. By analyzing aggregate ratings and rating counts, the project evaluates how customer feedback varies geographically and how engagement levels are influenced by rating quality.

Additionally, the project studies currency distribution to understand how Zomato operates across multiple markets with localized pricing systems. Data cleaning, aggregation, and visualization techniques are applied to ensure accuracy and clarity in analysis.

The insights derived from this exploratory analysis provide a strong foundation for data-driven decision-making and serve as a base for building interactive dashboards and performing further advanced analytical tasks.



# INDEX

Sr. No.	Title	Description
1	Report Overview	Overview of the report structure and purpose
2	Project Overview	Brief description of the project and scope
3	Objectives of the Analysis	Key goals and objectives of the EDA
4	Data Cleaning and Preprocessing	Steps performed to clean and prepare the dataset
5	Key Analytical Findings	Major analytical observations from the dataset
6	Visualization Insights	Graphical analysis with explanations
6.1	Country-wise Distribution Analysis	Country-level distribution of records
6.2	Rating Distribution & Customer Engagement	Analysis of ratings and engagement patterns
6.3	Rating Color Distribution	Distribution of restaurants by rating color
6.4	City-wise Distribution Analysis	City-level concentration of restaurants
7	Key Insights Summary	Consolidated insights derived from the analysis
8	Conclusion	Final conclusions and outcomes of the project



# Objectives

## 1. Understand the Structure and Key Attributes of the Dataset

The objective is to gain a clear understanding of the dataset by examining its structure, data types, and key attributes. This includes identifying numerical and categorical variables, checking for missing or inconsistent values, and understanding how different columns relate to each other.

## 2. Analyze Country-wise Restaurant Distribution

This objective focuses on analyzing how restaurants are distributed across different countries in the dataset. By grouping data at the country level, the analysis identifies regions with the highest and lowest number of restaurant listings.

## 3. Examine Customer Rating Patterns and Engagement

The goal of this objective is to study customer feedback by analyzing aggregate ratings and rating counts. This helps in understanding overall customer satisfaction, identifying common rating ranges, and evaluating how user engagement varies with different rating levels.

## 4. Derive Actionable Business Insights

The final objective is to consolidate findings from the analysis and translate them into actionable business insights. These insights support data-driven decision-making, help identify growth opportunities, and serve as a basis for dashboard creation, reporting, and further advanced analytics.



# Data Cleaning & Preprocessing

## Data Cleaning & Preprocessing :

Before performing the analysis, the dataset was carefully cleaned and preprocessed to ensure accuracy and consistency. The following steps were carried out as part of the preprocessing phase:

- Missing and zero values in rating-related columns were identified and reviewed to understand their impact on the analysis.
- Data types of numerical and categorical columns were verified and corrected where necessary to enable proper aggregation and visualization.
- The data was filtered and grouped appropriately to support accurate country-level and rating-level analysis.
- Consistency across categorical variables such as country names, rating categories, and currency values was ensured to avoid duplication and misinterpretation.

These preprocessing steps helped improve data quality and ensured that the analysis results are reliable and meaningful.



# Key Analytical Findings

The analysis of the Zomato dataset reveals important insights related to country-wise distribution, customer ratings, engagement levels, and currency usage. India contributes the highest number of records in the dataset, indicating that Zomato has its strongest market presence and user engagement in India. The United States and the United Kingdom follow India in terms of record count, reflecting Zomato's secondary markets.

Customer ratings were classified into categories ranging from Excellent to Average to better understand customer satisfaction levels. The majority of restaurants fall within the Good to Excellent rating ranges, suggesting generally positive customer feedback across the platform.

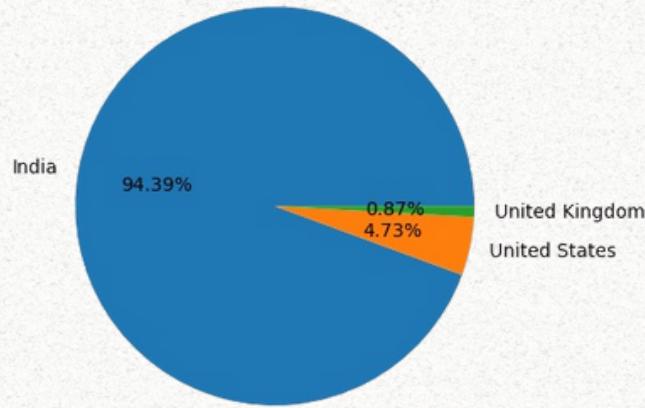
The analysis also identified countries with zero aggregate ratings. These cases may represent newly listed restaurants, limited customer engagement, or insufficient feedback data. Identifying such countries highlights regions that may require increased user participation or data validation.

Additionally, the currency analysis shows that each country primarily operates using its local currency, such as INR for India, USD for the United States, and GBP for the United Kingdom. This reflects Zomato's localized pricing and transaction systems across different regions.



# Visualization Insights

## A. Insight: Country-wise Distribution of Records



- The pie chart illustrates the distribution of records across countries in the Zomato dataset. India accounts for an overwhelming 94.39% of the total records, clearly indicating that Zomato's primary market presence and user activity are concentrated in India.
- This dominance highlights India as the core region driving platform engagement and data volume.
- In comparison, the United States contributes approximately 4.73% of the records, while the United Kingdom accounts for only 0.87%. The relatively small share of records from these countries suggests a limited dataset representation or lower platform penetration compared to India.

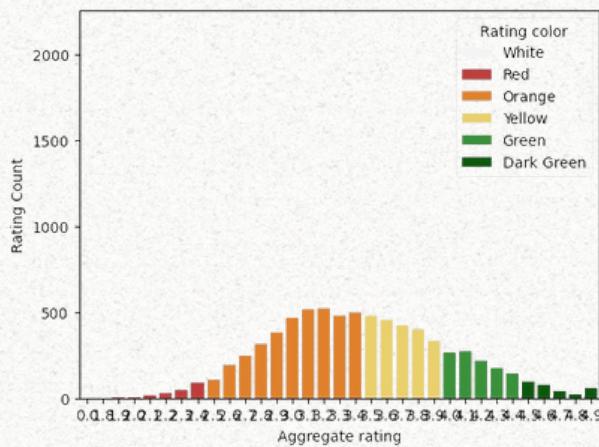
### Overall Interpretation:

The dataset is highly skewed toward India, implying that insights and trends derived from the analysis are largely influenced by the Indian market. This should be considered while generalizing findings across other countries.



# Visualization Insights

## B. Insight: Rating Distribution and Customer Engagement



- The bar chart illustrates the relationship between aggregate ratings and rating counts, segmented by rating color categories. The distribution shows that the majority of customer ratings are concentrated between 3.0 and 4.0, indicating that most restaurants receive average to good ratings.
- A gradual increase in rating count is observed as ratings move from lower values toward the mid-range, with peak engagement occurring around the 3.0–3.5 rating range. This suggests that restaurants with moderate ratings attract the highest number of customer interactions.
- As aggregate ratings increase beyond 4.0, the rating count begins to decline. Although higher-rated restaurants reflect better quality and customer satisfaction, they represent a smaller proportion of the dataset. Similarly, restaurants with very low ratings contribute minimally to overall engagement.

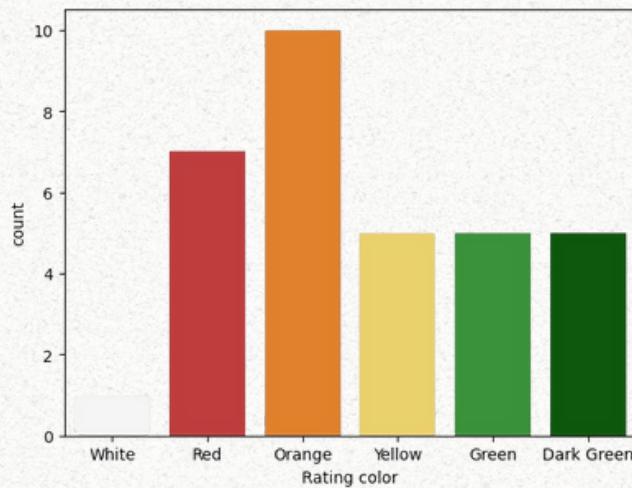
### Overall Interpretation:

Customer engagement is highest for restaurants with mid-range ratings, while extremely high or low ratings are less frequent. This highlights a balanced customer feedback pattern, with most restaurants clustered around average to good performance levels.



# Visualization Insights

## C.Insight: Distribution of Restaurants by Rating Color



- The bar chart represents the distribution of restaurants across different rating color categories. The Orange rating category has the highest count, indicating that a large number of restaurants fall within the corresponding rating range. This suggests that most restaurants are positioned in the average to good performance segment.
- The Red category follows next, reflecting a noticeable presence of restaurants with lower ratings. The Yellow, Green, and Dark Green categories show moderate and nearly similar counts, indicating a balanced distribution among higher-rated restaurants. The White category has the lowest count, suggesting a minimal number of unrated or newly listed restaurants.

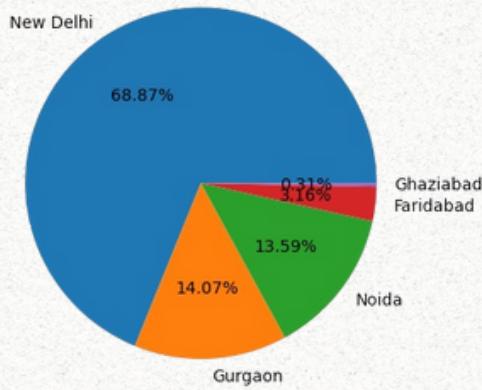
### Overall Interpretation:

The distribution shows that the majority of restaurants are clustered in mid-level rating categories, while fewer restaurants achieve very high ratings. This highlights a competitive landscape where consistent quality improvement is required to move into higher rating segments.



# Visualization Insights

## D.Insight: City-wise Distribution of Restaurants



- The pie chart shows the distribution of restaurants across major cities. New Delhi accounts for the largest share, contributing approximately 68.87% of the total records. This indicates that New Delhi is the primary hub of restaurant listings and customer activity in the dataset.
- Gurgaon and Noida follow with 14.07% and 13.59%, respectively, showing moderate presence and engagement. In comparison, Faridabad and Ghaziabad contribute a very small percentage of records, indicating relatively lower representation.

### Overall Interpretation:

The data is heavily concentrated in New Delhi, with significantly fewer records from surrounding cities, highlighting an uneven city-wise distribution.



# Key Insights – Detailed Summary

## Key Insights – Detailed Summary

### 1. Dominance of India in the Dataset

The analysis indicates that India has the highest number of restaurant records in the dataset. This highlights India as Zomato's strongest market in terms of platform presence and data volume.

### 2. Positive Customer Satisfaction Levels

A large proportion of restaurants fall within the Good to Excellent rating categories. This reflects overall positive customer experiences and satisfaction with restaurant services listed on the platform.

### 3. Presence of Zero-Rated Restaurants

The analysis identifies certain countries where restaurants have zero aggregate ratings. This may be due to newly listed outlets, limited customer engagement, or insufficient feedback data, pointing to potential areas for improvement.

### 4. Foundation for Data-Driven Decision-Making

Collectively, these insights provide a strong foundation for data-driven decision-making and support further analytical exploration, dashboard development, and strategic reporting.



# Conclusion

## 1. Strong Market Presence in India

The analysis shows that India contributes the highest number of restaurant records in the dataset. This indicates that Zomato has its strongest market presence and user engagement in India compared to other countries.

## 2. Overall Positive Customer Satisfaction

A significant proportion of restaurants fall under the Good to Excellent rating categories. This reflects generally positive customer experiences and a healthy level of satisfaction across the platform.

## 3. Existence of Zero-Rated Countries

Certain countries were identified where restaurants have zero aggregate ratings. This may indicate newly listed restaurants, limited customer interaction, or insufficient feedback data, highlighting areas that may require increased engagement or data validation.

## 4. Country-Specific Currency Usage

The currency analysis confirms that each country primarily uses its local currency. This reflects Zomato's localized pricing and transaction systems across different regions.

## 5. Support for Data-Driven Decision-Making

The insights derived from this Exploratory Data Analysis provide a strong foundation for building dashboards, performing advanced analytics, and supporting data-driven business decisions.