Zomato Data Analysis Project Report

Valadasu Gnaneswar

Data Analyst | July 2025 | Email: gnaneswargnana@gmail.com

GitHub: github.com/ValadasuGnaneswar

✅ Introduction

This project focuses on performing an in-depth Exploratory Data Analysis (EDA) on the Zomato dataset using Python. The goal is to uncover trends, patterns, and actionable business insights related to restaurants, customer preferences, cuisine popularity, and online delivery trends.

🗁 Dataset Overview

The dataset includes the following key features:

- Restaurant names
- Locations and city areas
- Ratings and number of votes
- Average cost for two people
- Online delivery availability
- Cuisines offered

Sample Data

```
import pandas as pd

df = pd.read_csv("zomato.csv")

df.head()
```

□ Data Cleaning

Key Steps:

- Removed duplicate entries using df.drop_duplicates()
- Filled or dropped null values based on context (e.g., missing cuisines)
- Converted cost column to numeric
- Standardized text columns (e.g., lowercased and stripped whitespace)
- Encoded binary features like Online Delivery (Yes/No)

Example

```
df['cost'] = df['cost'].str.replace(',', '').astype(float)

df['online_order'] = df['online_order'].map({'Yes': 1, 'No': 0})
```

🔍 Exploratory Data Analysis (EDA)

The dataset was analyzed to understand restaurant types, distribution by city, cost ratings, and popular cuisines.

Rating Distribution

```
df['rate'].value_counts().plot(kind='bar')
```

City-wise Restaurant Count

```
df['location'].value_counts().head(10).plot(kind='barh')
```

Top Cuisines

```
df['cuisines'].value_counts().head(10).plot(kind='bar')
```

Heatmap of Feature Correlation

```
import seaborn as sns

sns.heatmap(df.corr(), annot=True)
```

📉 Key Insights

- Online delivery restaurants generally receive higher ratings.
- North Indian and Chinese cuisines are most popular.
- Bangalore and Delhi NCR lead in restaurant counts.
- Mid-cost restaurants are rated better than extreme low/high cost ones.

📊 Business Recommendations

- Prioritize online ordering and delivery optimizations.
- Focus on popular cuisines like North Indian and Chinese.
- Offer mid-priced menu combos to appeal to broader audiences.
- Promote in cities like Bangalore where competition is high but profitable.

🧠 Code Improvements

- Refactored repetitive code into functions
- Used vectorized Pandas operations for performance
- Structured notebook with markdown and sectioning for clarity

---

🚀 Future Work

- Build an interactive dashboard using Streamlit or Power BI
- Train ML models to predict restaurant ratings based on features