## 📝 Zomato Data Analysis – Project Report

### 📌 Title:

**Exploratory Data Analysis on Zomato Bangalore Restaurant Data**

---

### 📖 1. Introduction

This project aims to perform **Exploratory Data Analysis (EDA)** on the Zomato Bangalore restaurant dataset. The primary objective is to uncover patterns and derive insights regarding customer preferences, restaurant types, cost distributions, online order trends, and customer ratings.

---

### 🧰 2. Libraries Used

| Library | Purpose |
|---|---|
| pandas | For data manipulation and handling tabular data (DataFrame). |
| numpy | For numerical operations and handling missing values. |
| matplotlib.pyplot | For basic visualizations like histograms and line plots. |
| seaborn | For advanced, aesthetic visualizations like boxplots and count plots. |

---

### 🗃️ 3. Dataset Description

- **Dataset Name**: Zomato Bangalore Restaurant Data
- **Loaded using**: pd.read_csv("Zomato data .csv")
- **Main Columns**:
    - rate: Average rating
    - votes: Number of people who voted
    - listed_in(type): Type/category of restaurant
    - approx_cost(for two people): Cost estimate
    - online_order: Whether the restaurant supports online ordering

---

### 🔄 4. Data Cleaning & Preprocessing

1. **Handled Rating Field**:
    - The rating values like "4.1/5" were converted to float using:

python

CopyEdit

```
def handleRate(value):
    value = str(value).split('/')
    return float(value[0])
df['rate'] = df['rate'].apply(handleRate)
```

2. **Verified Data Structure**:
    - Used df.info() and df.head() to understand the data types and preview the dataset.

---

### 📊 5. Visual Explorations & Analysis

#### ✅ Restaurant Type Distribution

- Visualized using:

```
python
CopyEdit
sns.countplot(x=df['listed_in(type)'])
```

- **Insight**: Helps identify the most common types of restaurants.

#### ✅ Total Votes by Type

- Calculated with:

```
python
CopyEdit
df.groupby('listed_in(type)')['votes'].sum()
```

- Visualized with a **line plot** showing which restaurant types receive the most votes.

#### ✅ Ratings Distribution

- Displayed using a histogram:

```
python
CopyEdit
plt.hist(df['rate'], bins=5)
```

#### ✅ Cost for Two People

- Used sns.countplot to understand common cost brackets.

#### ✅ Online Order vs Rating

- Boxplot:

```
python
CopyEdit
```

sns.boxplot(x='online_order', y='rate', data=df)

- **Insight**: Determines if online order availability affects ratings.

---

### 📈 6. Pivot Table Analysis

- A pivot table was created to explore relationships between restaurant types and cost/votes (though the code was cut off).

---

### 🧠 7. Conclusion & Insights

- **Most popular restaurant types** can be identified by vote counts.
- **Online ordering restaurants** may show variation in ratings.
- **Rating distributions** show customer satisfaction trends.
- **Cost analysis** provides insight into affordability.