🍽️ Zomato Data Analysis

📌 Project Overview

This project is an in-depth analysis of **Zomato's restaurant dataset**, aimed at uncovering insights about the food and beverage industry. Zomato is one of the most popular restaurant discovery and food delivery platforms, and analyzing its data allows us to understand **customer preferences, pricing trends, cuisine popularity, and rating patterns**.

The project involves **data cleaning, exploratory data analysis (EDA), data visualization, and optional machine learning models**. It serves as a valuable reference for individuals or businesses in the food-tech sector who wish to leverage data-driven strategies.

📊 Dataset

- **Source:** Publicly available Zomato dataset (commonly found on Kaggle or Zomato's API dumps).
- **Attributes include:**
    - Restaurant ID / Name
    - Location (city, area)
    - Cuisines served
    - Average cost for two people
    - Aggregate rating and number of votes
    - Online delivery / Table booking availability

The dataset is first **cleaned and preprocessed** by handling missing values, removing duplicates, and standardizing categorical variables.

⚙️ Installation

To run this project locally:

1. Clone the repository:
2. git clone https://github.com/yourusername/zomato-analysis.git
3. cd zomato-analysis

4. Install required dependencies:
5. pip install -r requirements.txt
6. Run Jupyter Notebook or Python scripts:
7. jupyter notebook

🚀 Usage

- Open the provided notebook/script.
- Run all cells to perform:
    - **Data Cleaning** → Handle null values, duplicates, encoding categorical values.
    - **Exploratory Data Analysis (EDA)** → Summarize the dataset and discover trends.
    - **Visualization** → Graphical representation of insights.
    - (Optional) **Prediction/Recommendation** → Apply ML models for restaurant rating prediction or recommendation system.

🔎 Analysis & Insights

The project explores several key questions, such as:

1. **Restaurant Distribution** – Which cities or areas have the highest concentration of restaurants?
2. **Cuisines** – What cuisines are most popular among customers?
3. **Ratings** – Which factors influence restaurant ratings the most?
4. **Cost Analysis** – How does the cost for two vary across cuisines and locations?
5. **Delivery & Booking Trends** – What percentage of restaurants provide online delivery and table booking?

📈 Results

- Identified **top cuisines** preferred by customers.
- Discovered **affordable vs. premium dining zones** across different cities.

- Analyzed **customer rating distribution**, showing what drives high ratings.
- Visualized trends in **online delivery adoption and cost patterns**.
- (If ML included) Built a **restaurant rating prediction model** with good accuracy.

🔮 Future Scope

This project can be extended in several directions:

- **Machine Learning Models:**
  - Restaurant recommendation systems.
  - Predicting customer ratings based on features like location, cuisine, and cost.
- **Interactive Dashboards:**
  - Build real-time dashboards in **Power BI** or **Tableau**.
  - Enable users to explore restaurant insights interactively.
- **Integration with Zomato API:**
  - Pull live restaurant data instead of using static datasets.
  - Provide updated analytics for business insights.
- **Business Applications:**
  - Help restaurants optimize pricing strategies.
  - Identify under-served cuisines or locations for new restaurants.

🛠️ Tech Stack

- **Programming Language:** Python
- **Libraries/Packages:**
  - Data Processing → Pandas, NumPy
  - Visualization → Matplotlib, Seaborn, Plotly
  - Machine Learning → Scikit-learn
- **Optional Tools:** Tableau / Power BI for dashboards

- **Version Control:** Git & GitHub

📂 Project Structure

```
├── data/              # Dataset files
├── notebooks/         # Jupyter notebooks
├── scripts/           # Python scripts for cleaning/analysis
├── visuals/           # Saved graphs/plots
├── requirements.txt   # Dependencies
├── README.md          # Project documentation
```

🙌 Acknowledgements

- **Dataset Source:** Kaggle / Zomato
- Inspiration from real-world food-tech problems and data science case studies.