Zomato Dataset Analysis Project

-Anubhav Singh Pathania (2024010015)
-Yash Arora(2024010120)
MCA-4th Group

About the Dataset

This dataset is a collection of restaurants registered on Zomato in Bengaluru City. It contains over 50,000 rows and 17 columns, making it fairly large and suitable for real-world data analysis practice.

The dataset provides details such as:

- Restaurant name
- Location and city
- Online delivery and table booking availability
- Cuisines offered
- Average cost for two people
- Ratings and votes
- Restaurant type and services

Dataset Tasks in Analysis

To make sense of the data and extract useful insights, the following steps are performed:

1. **Explore the Data**
    - Load and understand each feature
    - Separate categorical and numeric columns
    - View dataset info, descriptive stats
2. **Data Cleaning**
    - Handle missing values
    - Remove duplicates and redundant columns
    - Rename columns for clarity
    - Clean individual columns (e.g., cost formatting, converting to numeric)
    - Apply transformations if required
3. **Data Visualization & Analysis**
    - Online delivery vs no delivery restaurants
    - Restaurants allowing table booking vs not
    - Table booking vs rating relation
    - Best performing locations
    - Relation between location and ratings
    - Restaurant types and service types
    - Cost of restaurants (low, medium, high)
    - Distribution of restaurants by locality
    - Most popular restaurant chains in Bengaluru

Exploratory Data Analysis (EDA)

Key analyses performed:

- **Restaurant Distribution:** Count of restaurants city-wise and locality-wise

- **Online Ordering:** Identified proportion of restaurants offering delivery
- **Table Booking:** Compared restaurants with and without reservation services
- **Table Booking vs Rating:** Found whether bookings are correlated with better customer experience
- **Best Locations:** Identified localities with highest density and top ratings
- **Cuisine Analysis:** Most popular cuisines and combinations (e.g., North Indian + Chinese)
- **Restaurant Types:** Distribution across casual dining, cafes, delivery outlets, etc.
- **Cost Analysis:** Average price range for two people, affordability across areas
- **Popular Chains:** Top-performing restaurant chains such as Domino's, McDonald's

Advanced Customer-Centric Analysis

- **Cuisine Combination Analysis:** Popular pairings (North Indian + Chinese were most common).
- **Value-for-Money Restaurants:** Budget restaurants with high ratings were identified.
- **City Rating Patterns:** Some areas showed stricter ratings compared to others.
- **Best Affordable Options:** Highlighted hidden gems with high quality but low cost.

Restaurant Owner / Business Perspective

- **Service Impact:** Restaurants offering online delivery and table booking generally earned better ratings.
- **Top Chains:** Domino's, McDonald's, and Café Coffee Day showed consistent presence across Bengaluru.
- **Price vs Popularity:** Mid-priced restaurants earned more votes compared to luxury outlets.
- **Cuisine Demand:** North Indian, South Indian, and Chinese cuisines dominate customer preferences.

Geographic Insights

- **City-Wise Cost Trends:** Average cost for two is lower in tier-2 localities compared to metro hubs.
- **Locality Hotspots:** Koramangala and Indiranagar are popular for premium dining.
- **Cuisine by Area:** South Indian food dominates certain areas, while fast food is widespread across student-populated localities.
- **Interactive Maps:** Plotly/Folium visualizations were used to show restaurant density and performance.

Machine Learning Extension (Not confirmed)

As an additional step, the dataset can also be extended to include predictive modeling, where restaurant ratings are classified into categories based on features such as votes, cost, cuisines, and services. This helps in understanding which factors influence customer ratings the most. However, this step is optional, and the analysis remains complete even without it.

Key Insights

- Majority of restaurants are clustered in metropolitan areas.
- **Online ordering** is more popular than table booking.
- North Indian and Chinese cuisines dominate customer demand.

- Cost for two is mostly clustered between ₹300–₹800.
- High-rated restaurants generally attract more votes, proving credibility.
- Casual dining is the most common restaurant type.
- Service features like online delivery strongly influence ratings and popularity.

Conclusion

This project demonstrates how a real-world dataset can be cleaned, explored, and analyzed to extract actionable insights.

- **Customers** can use the insights to find affordable, highly-rated restaurants and explore cuisine combinations.
- **Restaurant owners** can design better menus, pricing strategies, and service offerings to improve customer satisfaction.
- **Market trends** highlight areas for new restaurant openings, cuisine demand, and price sensitivities.

Future Scope

- Build a **personalized restaurant recommendation system** based on customer history and preferences.
- Expand predictive modeling with advanced algorithms (XGBoost, Neural Networks).
- Integrate live Zomato API data for real-time insights.
- Merge with external datasets (population, income levels, tourism data) for market feasibility studies.

Dataset Link- https://www.kaggle.com/datasets/rajeshrampure/zomato-dataset?resource=download