Data Description
🍽 Zomato Restaurant Dataset
Divya Ramanujam

This dataset is all about restaurants listed on Zomato in India. 🇮🇳 Each row represents a restaurant, and we get to see its location, cost, delivery options, user ratings, and more. Our goal? To figure out what makes a restaurant *successful* — based on the ratings it receives!

📋 What's in the Dataset?

There are thousands of restaurants across different Indian cities, and for each one, we have:

Feature	What it Means
Longitude	East-west GPS coordinate of the restaurant
Latitude	North-south GPS coordinate of the restaurant
Average Cost for two	Average meal price for two people
Has Table booking	Whether it accepts table bookings (Yes/No → 1/0)
Has Online delivery	Whether it offers online food delivery (Yes/No → 1/0)
Price range	Pricing tier assigned by Zomato (1 = low, 4 = high)

`Aggregate rating`	Average user rating of the restaurant (0.0 to 5.0)
`Votes`	Number of user votes submitted on Zomato
`Successful`	Target variable — 1 if rating > 4.0, else 0 (indicates popularity/success)
`alcohol`	Alcohol percentage
`quality`	A score from 0 to 10 given to each wine (this is what we want to predict!)

🧼 What Cleaning Did We Do?

Filtered to Indian Restaurants:
We only kept rows where `Country Code == 1`, so we're looking at Indian data.

Dropped Unnecessary Columns:
Removed IDs, verbose addresses, and features not useful for modeling.

Handled Missing Data:
Dropped rows where the rating (`Aggregate rating`) was missing.

Encoded Categorical Variables:
Converted "Yes"/"No" columns like `Has Table booking` and `Has Online delivery` into 1s and 0s.

Outlier Treatment:
For numeric features like cost and votes, we used **IQR** (interquartile range) to remove extreme outliers.

Fixed Skewed Data:

We applied **log transformation** to highly skewed features — this makes models work better.

🎯 **What Are We Trying to Predict?**

We introduced a new column: **Successful**

- 1 → The restaurant is successful (rating > 4.0)

- 0 → Otherwise

This turns our task into a **classification problem** — predicting which restaurants are likely to be successful on Zomato.

📈 **What Did We Learn from the Data?**

- Restaurants with **more votes and higher price ranges** often have better ratings ⭐
- **Online delivery** and **table booking** aren't always linked with higher success
- Some locations (by Latitude/Longitude) show clusters of higher ratings
- We used a **heatmap** to visualize how all features relate to each other