Zomato Power BI Dashboard Project – Documentation

---

1. Data Source

The primary data source for this project was a **single dataset table named Zomato**, which contains restaurant-level information focused on Bangalore city. This dataset included key attributes such as:

- Restaurant name
- Online ordering availability
- Table booking option
- Customer rating
- Number of votes
- Location
- Type of restaurant
- Popular dishes
- Cuisines offered
- Approximate cost for two people
- Listing category and city

To support deeper analysis and enable better dashboard insights, additional tables were created **through transformation logic** using Power BI's data modelling features also I have created an index column in Zomato data so that I can perform Data modelling easily. These include:

- **Cuisines Table:** Extracted from the Zomato table by splitting and normalizing the 'Cuisines' column to analyse cuisine preferences independently.
- **Most_Liked_Dish Table:** Created by separating and deduplicating values from the 'dish_liked' column to evaluate dish-level popularity.
- **Calendar_Lookup Table:** A synthetic date dimension created by randomly generating dates for each restaurant, allowing for time-based trend analysis like month-over-month performance, seasonal cuisine shifts, and restaurant growth.

This modular approach helped in building a **normalized, scalable data model** essential for effective visual storytelling and performance.

---

2. Data Cleaning and Preparation

To prepare the data for analysis, several **data cleaning and preprocessing** steps were performed:

a. Handling Missing & Invalid Data:

- Replaced or removed nulls and inconsistent entries in key fields such as 'rate', 'votes', and 'location'.
- Converted ratings from string format (e.g., "4.2/5") to numeric format (e.g., 4.2).
- Ensured 'votes' are integers and missing values were treated as 0 for aggregations.

b. Standardizing and Splitting Columns:

- Removed special characters, trimmed extra spaces, and standardized naming conventions across fields.
- 'Cuisines' and 'Dish_Liked' columns were split and converted into separate normalized tables for granular insights.

c. Date Table Creation:

- Since the original dataset lacked timestamps, a **randomized date column** was added using MS Excel. This allowed creation of a proper Calendar table (Calendar_Lookup) which was then related to restaurant data using the index column.

d. Data Validation:

- Checked for duplicate rows and inconsistent mappings across new tables (e.g. ensuring a dish exists only for valid restaurants).
- Verified integrity of relationships to maintain accurate one-to-many connections in the model.

---

3. Data Transformations and Modelling

To support rich insights and KPI-driven storytelling, several **transformations and DAX measures** were implemented:

a. Measures (KPIs):

- `Avg_Cost(For_Two) = CALCULATE(AVERAGE(zomato[approx_cost(for two people)]))`
- `Avg_Rating = AVERAGE(zomato[rate(out of 5)])`
- `Avg_Votes = AVERAGE(zomato[votes])`
- `Cuisine_Growth = CALCULATE([Total_Cuisine_Count],DATESMTD('Calendar Lookup'[Order_date]))`
- `Dining_In = CALCULATE(COUNT(zomato[book_table]),zomato[book_table] = "Yes")`
- `Dining_Out_Count = CALCULATE([Total Restaurants],zomato[listed_in(type)]= "Dine-Out")`
- `Food_Delivery_Count = CALCULATE([Total Restaurants],zomato[listed_in(type)]="Delivery")`
- `Monthly_New_Restaurants = CALCULATE(DISTINCTCOUNT(zomato[name]),MONTH('Calendar Lookup'[Order_date]))`
- `Online_Orders = CALCULATE(COUNT(zomato[online_order]),zomato[online_order]="Yes")`
- `Total Restaurants = COUNT(zomato[name])`
- `Total_Cuisine_Count = COUNT(Cuisine[cuisines])`
- `Total_Dish_Count = CALCULATE(COUNT(Most_Liked_Dish[dish_liked]))`
- `Total_Rating = SUM(zomato[rate(out of 5)])`
- `Total_Restaurant_Type = COUNT(zomato[rest_type])`
- `Total_Votes = CALCULATE(SUM('Zomato'[votes]))`

b. Model Relationships:

- Established one-to-many relationships between the Zomato table and the other derived tables (Cuisines, Dish_Liked, Calendar_lookup).

---

4. Assumptions Made

To simplify the analysis and manage missing elements in the dataset, several **practical assumptions** were made:

- **Missing values in 'rate' or 'votes'** were assumed to indicate no customer engagement and were converted to 0.
- The **'Calendar_Lookup'** table, though not present in original data, was manually generated using random but realistic dates for analysis purposes.
- Each row in the Zomato table is assumed to represent a unique index.

---

5. Business Objective and Alignment

The objective of this Power BI project is to provide **actionable insights into restaurant trends, customer preferences, and business opportunities** using data-driven visualizations. Specifically:

- Understand online vs. offline customer engagement
- Identify popular cuisines and dishes
- Track restaurant growth trends over time
- Recommend business strategies to improve satisfaction and market positioning

Every step, from data modelling to dashboard design, was executed with this objective in mind — ensuring the final outcome delivers meaningful, real-world impact for stakeholders and decision-makers.