SUPPLEMENTAL: EXPANDED METHODOLOGY

DATA PREPARATION:

To analyze Zomato's customer base, I used Power Query in Excel to clean and transform the data before importing it into Tableau. The key steps included:

- Orders Table:
    - Loaded the raw data from a CSV file.
    - Standardized column names for consistency.
    - Converted currency values from INR to USD using an exchange rate of 0.01155 (as of November 31, 2024).
    - Removed unnecessary columns like currency type.
- Restaurant Table:
    - Loaded restaurant details from a CSV file.
    - Split the "cost" column into two parts to isolate numeric values.
    - Converted restaurant costs from INR to USD.
    - Removed columns not relevant to the analysis, such as restaurant name, city, and address.
- Users Table:
    - Imported customer data from a CSV file.
    - Removed personally identifiable information (names, emails, and passwords).
    - Kept key demographic information such as age, occupation, and monthly income.
    - Tables were then merged to create a structured dataset:
- The Orders Table was left-joined with the Restaurant Table to link order details with restaurant attributes.
- The Orders Table was fully joined with the Users Table to associate customer data with their purchase history.

SEGMENTATION & CALCULATED FIELDS

To better understand customer behavior, I created customer segments by using calculated fields in Tableau:

- Monthly Income Segments:
    Customers were grouped based on **income brackets** using a segmentation model from a published report by the **Insitute for Competitiveness** (Kapoor & Duggal, 2022). The Monthly Income field values were adjusted via calculated column for easier interpretation to:

    - No Income
    - (Below ₹10,000) → Low Income
    - (₹10,001 - ₹25,000) → Modest Income
    - (₹25,001 - ₹50,000) → Mid-Level Income
    - (More than ₹50,000) → High Income
- Age Segments:
    This segmentation was based on a study on **consumer spending habits** (Adams, Alldredge, & Kohli, 2024), suggesting that **ages 18-24** are associated with **exploratory purchasing behaviors**, while **ages 25+** show more **brand loyalty and stable spending habits** due to career establishment and income growth.
    - Customers under 25 were categorized as "**Younger Customers**".
    - Customers 25 and older were labeled "**Older Customers**".
- Occupation Segments:
    Occupation categories were adjusted for a more neutral and audience-friendly classification:
    - "Student" remained Student
    - "Employee" changed to Professional

- o "Self-Employed" changed to Entrepreneur
  - o "Housewife" changed to Homemaker
  - o Others remained as "Other"
- Restaurant Ratings:
  - Simplified restaurant rating categories for easier interpretation:
  - o 4.0+ → Highly Rated
  - o 3.0 - 3.99 → Moderately Rated
  - o Below 3.0 → Low Rated
  - o Missing values → No Rating
- Repeat Orders by Restaurant:
  - Categorized users based on how many orders they have placed:
  - o If a user has placed **exactly 1 order**, they fall into the **"1 Order"** category.
  - o If they have **2 or 3 orders**, they are in the **"2-3 Orders"** category.
  - o If they have **4 or 5 orders**, they belong to the **"4-5 Orders"** category.
  - o If they have **6 or 7 orders**, they are in the **"6-7 Orders"** category.
  - o If they have **8 or 9 orders**, they fit into the **"8-9 Orders"** category.
  - o If they have **10 or more orders**, they are placed in the **"10+ Orders"** category.
- Restaurant Cost Segments:
  - Categorized restaurants by cost using percentiles:

Segment	Percentile	Explanation
Low Cost	25th	Bottom 25% of costs
Moderate Cost	50th	Median range
High Cost	75th	Higher-end but not premium
Premium Cost	95th	Top 5% but not the most expensive
Luxury Cost	Above 95th	Highest-priced category

Table 1

- Rating Count Groups:
  - Combined original rating count segments into broader, more digestible groups:

Original Rating Count Segment	New Broader Group	Explanation
(NULL) or empty	no ratings	Accounts for missing or empty values.
Too Few Ratings	no ratings	Consolidates entries labeled as having too few ratings into "no ratings."
20+ ratings	under 100 ratings	Groups smaller rating counts (20+ and 50+) into "under 100 ratings."
50+ ratings	under 100 ratings	Same logic as above.
100+ ratings	over 100 ratings	Marks entries with at least 100 ratings as "over 100 ratings."
500+ ratings	over 500 ratings	Simplifies the original label into a broader "over 500 ratings" group.
1K+ ratings	over 1K ratings	Consolidates into "over 1K ratings."
5K+ ratings	over 5K ratings	Groups higher-rated establishments into "over 5K ratings."
10K+ ratings	over 10K ratings	Simplifies the highest rating count category into "over 10K ratings."

Table 2

- Order Value Metrics:
  - o Calculated median and average order values directly in Tableau.

Additional helper fields were created but not used in the final analysis—only for visualization enhancements.