Decomposition

Hypothesis:

American fast-food chains (e.g., Domino's, McDonald's, Burger King, Subway) are expected to be among the most popular restaurants in India, based on total sales and number of orders, when compared to local Indian chains and independently owned restaurants.

Restaurant Analysis Questions, Metrics, & Visuals:

1. **Which restaurants are the most popular based on the number of orders they receive?**
   (This shows which restaurants receive the highest volume of orders.)

   Bar Chart showing Sum of sales_qty(orders) per id

   Label: Restaurants with Highest Order Volume

2. **Which restaurants generate the highest overall revenue?**
   (Highlights which locations make the most money overall.)

   Bar Chart showing Sum of sales_amount per id

   Label: Highest Earning Restaurant Locations

3. **Which restaurants have the highest average order value?**
   (This tells us how much customers typically spend in one visit.)

   Bar Chart showing restaurant locations are most popular based on average order value: sales_amount/sales_qty per id
   Label: Where Customers Spend the Most per Order

4. **Which restaurants are attracting the most new customers?**

   (This reveals which places are growing in popularity and reaching new people.)

   Bar Chart Sum of unique user_id per id

   Label: Locations Attracting the Most New Customers

5. **Which restaurant locations have the most customers visit overall?**

   (This reveals which restaurants have the highest number of total customers, including repeat visitors.)

   Bar Chart displaying highest sum user_id per id?

   Label: Total Visits per Location

Software used for analysis: Power BI

Sheets used & Relationship: The sheets that will be used in the analysis will be 'restaurants' and 'orders' with a Many to One relationship. The two sheets will be joined by id from the restaurant table and r_id from the orders table.

Methodology & Data Preparation:

The original dataset included sales figures (sales_amount) recorded in multiple currencies. To enable accurate comparison and aggregation of sales across all restaurants, all values in the sales_amount column will be converted to a common currency: Indian Rupees (INR).

As part of the data cleaning process, a new column named INR_sales will be created. This column will contain the standardized sales figures in INR, calculated by converting each value from sales_amount based on the appropriate exchange rate. This can be accomplished using the Googlefinance syntax in Google Sheets with the data set provided. This conversion ensures consistency in financial data across restaurants

operating in different regions and currencies, supporting more meaningful and reliable analysis.

Assumptions:

Missing or inconsistent values are assumed to have minimal impact on the analysis results. Any missing data will be handled appropriately to ensure accuracy, whether through imputation or exclusion, depending on the nature of the data.

The orders and restaurant tables are assumed to provide correct, accurate, and complete information for the analysis. A preliminary data quality check has been conducted to confirm this assumption.

Findings:

Discuss issues and inaccuracies with the hypothesis, such as data consideration errors, analysis bias, data discrepancies (e.g., multiple locations with the same name).

Discuss the findings of each analysis question, highlighting key insights such as differences in highest sales across locations and highest orders.

Possible Follow-up questions & analysis:

Where are the most popular restaurant locations across India?
Metric: Map(filled) of Top 10 restaurants from question 1 (Highest number of sales_qty & their city locations)
Purpose: Helps identify which areas have the highest-performing restaurants.

Methodology & Data Preparation:

The original dataset contains a location field where some values combined city and suburb, while others included only the city name. To ensure consistent analysis, the data will be standardized by extracting only the city portion.
This was achieved by splitting the location string at the comma delimiter and retaining the first element as the primary city identifier. This approach enables consistent grouping and comparison across all restaurant entries, regardless of the formatting differences in the original data.