

Project Requirement Document

Project Title

Restaurant Orders Data Warehouse & Business Analytics (Swiggy Dataset)

Objective

Design and implement an end-to-end data warehousing and analytics solution for restaurant order data. The goal is to ingest raw transactional data, perform validation and cleaning, model it using a star schema, load dimension and fact tables, and generate business-ready KPIs and analytical insights.

Data Source

- **Source Type:** CSV file
 - **Dataset:** Swiggy restaurant orders data
 - **Import Method:** PostgreSQL `COPY` command (UI-assisted import)
 - **Key Attributes:** State, City, Order Date, Restaurant, Location, Category, Dish, Price, Rating, Rating Count
-

Scope of Work

The project covers the full data lifecycle: 1. Raw data ingestion 2. Data validation and cleaning 3. Dimensional data modeling (Star Schema) 4. Data loading (ETL) 5. KPI computation 6. Business analysis and insights

Tasks Performed

1. Raw Table Creation

- Created a staging table `restaurant_orders` to store raw transactional data.
- Defined appropriate data types for dates, numeric values, and textual dimensions.

Purpose: - Act as a landing table for raw, unprocessed data. - Preserve original data structure before transformations.

2. Data Ingestion

- Imported CSV data into the staging table using the PostgreSQL `COPY` command.
- Ensured correct column mapping and delimiter handling.

Outcome: - Successful population of raw order data into the database.

3. Data Validation & Cleaning

3.1 Null Checks

- Identified missing values across all critical columns such as state, city, order date, restaurant name, price, ratings, and rating count.

3.2 Blank / Empty String Checks

- Detected empty strings in dimensional attributes (state, city, restaurant name, location, dish name).

3.3 Duplicate Detection

- Checked for duplicate records based on a composite of all business-relevant columns.

Purpose: - Ensure data quality before loading into dimension and fact tables.

4. Dimensional Modeling (Star Schema)

Designed a **Star Schema** to support analytical queries efficiently.

4.1 Dimension Tables

Date Dimension (`dim_date`) - Attributes: Full date, year, month, month name, quarter, day, week - Enables time-based analysis (daily, monthly, quarterly, yearly trends)

Location Dimension (`dim_location`) - Attributes: State, city, location - Supports geographic analysis

Restaurant Dimension (`dim_restaurant`) - Attribute: Restaurant name - Enables restaurant-level performance analysis

Category Dimension (`dim_category`) - Attribute: Food category / cuisine - Used for cuisine-level insights

Dish Dimension (`dim_dish`) - Attribute: Dish name - Supports dish-level performance tracking

5. Fact Table Creation

Fact Table: `fact_restaurant_orders`

- Grain: One record per dish order
- Measures:
 - Price (INR)
 - Ratings
 - Rating count
- Foreign keys linking to all dimension tables

Purpose: - Central transactional table for analytics and KPI computation

6. Data Transformation & Loading (ETL)

- Loaded **distinct values** into each dimension table from the staging table.
- Generated surrogate keys for dimensions.
- Populated the fact table by joining staging data with all dimension tables.
- Fixed join and naming issues (e.g., column mismatches, typos) during loading.

Outcome: - Fully populated star schema with referential integrity.

7. Integrated Data View

- Created multi-table joins between fact and dimension tables.
 - Enabled a consolidated view for end-to-end validation and analysis.
-

KPI Definitions & Computation

Core KPIs

- **Total Orders:** Count of order records in fact table
 - **Total Revenue:** Sum of price values (reported in millions)
 - **Average Dish Price:** Mean price per order
 - **Average Rating:** Mean customer rating across all orders
-

Business Analysis Performed

1. Time-Based Analysis

- Monthly order trends
- Monthly revenue trends
- Quarterly order distribution
- Yearly order growth
- Orders by day of the week

2. Location-Based Analysis

- Top 10 cities by order volume
- Revenue contribution by state

3. Food & Restaurant Performance

- Top 10 restaurants by order count
- Top food categories by demand
- Most ordered dishes
- Cuisine performance based on total orders and average ratings

4. Price Sensitivity Analysis

- Order distribution across price ranges:
- Under 100
- 100–200
- 200–300
- 300–500
- 500+

5. Ratings Distribution

- Distribution of orders across different rating values
-

Deliverables

- Cleaned and validated dataset
 - Star schema (dimensions + fact table)
 - SQL scripts for table creation, loading, and analysis
 - Business-ready KPIs and insights
-

Conclusion

This project demonstrates a complete SQL-based data warehousing and analytics workflow—from raw data ingestion to advanced business analysis. The implemented star schema ensures scalability, query efficiency, and clarity, making the dataset suitable for dashboards, reporting, and decision-making use cases.