SWIGGY SALES ANALYSIS

BUSINESS REQUIREMENTS

Data Cleaning & Validation

The raw table swiggy_data contains food delivery records across states, cities, restaurants, categories, and dishes.

The first goal is to ensure data quality by performing:

Null Check

Identify missing values in:

- State
- City
- Order_Date
- Restaurant_Name
- Location
- Category
- Dish_Name
- Price_INR
- Rating
- Rating_Count

Blank/Empty String Check

Detect fields containing blank values that may cause inaccurate analysis.

Duplicate Detection

Find duplicate rows using grouping on all business-critical columns.

Duplicate Removal

Use ROW_NUMBER() to delete surplus duplicate rows while retaining one clean copy for each unique order.

Dimensional Modelling (Star Schema)

Dimensional modelling helps organize data in a way that makes analysis simple, consistent, and extremely efficient. Instead of keeping all information mixed in one large table, the Star Schema separates descriptive information into dimensions and keeps all measurable values in a central fact table. This structure reduces data duplication, improves clarity, and makes the entire dataset easier to understand. It also supports faster reporting because queries can access small, focused tables rather than scanning a single bulky dataset. Most analytics and BI tools are designed to work best with star schemas, so this approach ensures smooth dashboard creation, accurate aggregations, and reliable insights. Overall, dimensional modelling provides a clean, scalable, and performance-optimized foundation for any analytical system, regardless of the industry or data source.

To optimize analytics and reporting, build a **Star Schema** with the following dimension tables:

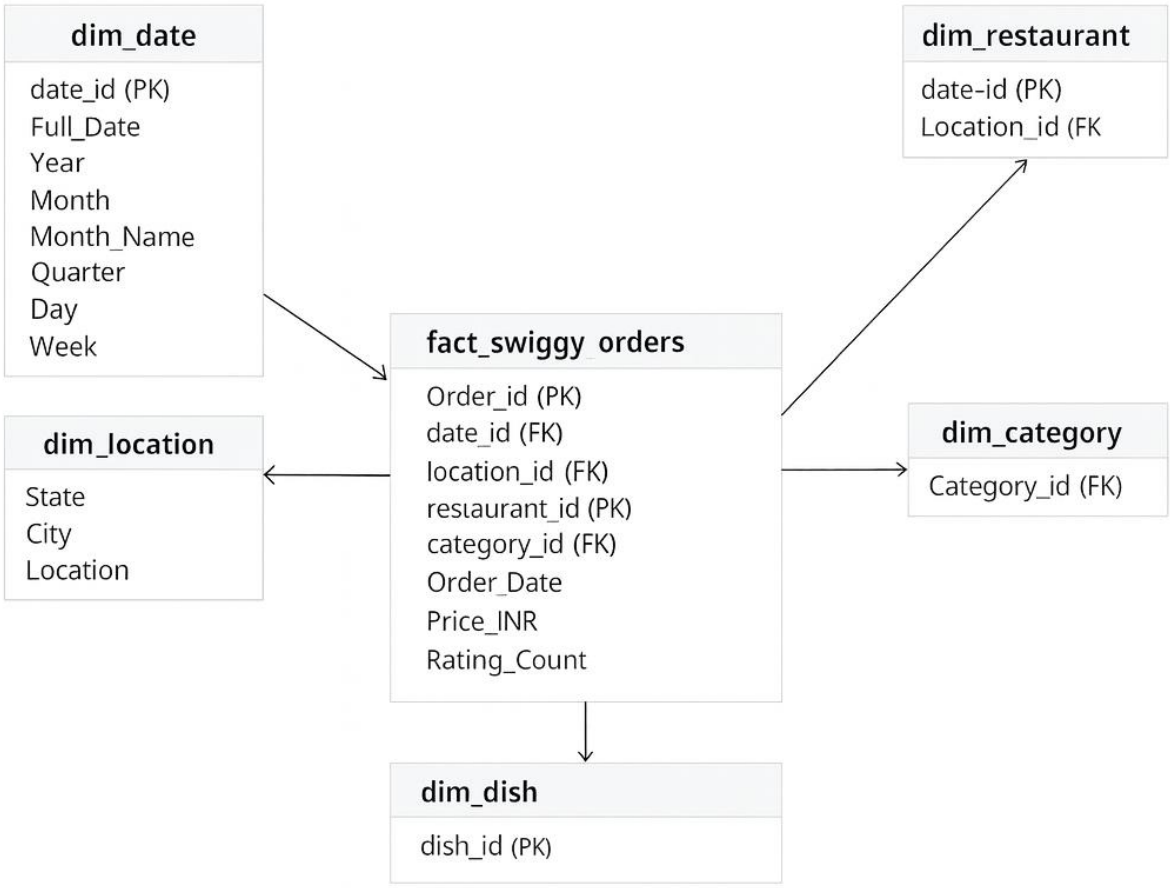
- **dim_date** → Year, Month, Quarter, Week
- **dim_location** → State, City, Location
- **dim_restaurant** → Restaurant_Name
- **dim_category** → Cuisine/Category
- **dim_dish** → Dish_Name

Central fact table:

- **fact_swiggy_orders** → Price_INR, Rating, Rating_Count, foreign keys to all dimensions

Populate each dimension with **distinct** data from the cleaned source and load fact table with all keys resolved.

ERD DIAGRAM: STAR SCHEMA



Star Schema for Swothy dataset (fact_swchy_orders wth dim_date, dim_location, dim_category)

KPI Development

Once the schema is built, compute core performance indicators:

Basic KPIs

- Total Orders
- Total Revenue (INR Million)
- Average Dish Price
- Average Rating

Deep-Dive Business Analysis

Date-Based Analysis

- Monthly order trends
- Quarterly order trends
- Year-wise growth
- Day-of-week patterns

Location-Based Analysis

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

Food Performance

- Top 10 restaurants by orders
- Top categories (Indian, Chinese, etc.)
- Most ordered dishes
- Cuisine performance → Orders + Avg Rating

Customer Spending Insights

Buckets of customer spend:

- Under 100
- 100–199
- 200–299
- 300–499
- 500+

With total order distribution across these ranges.

Ratings Analysis

Distribution of dish ratings from 1–5.