

Swiggy Sales Data Analysis: A SQL Case Study

Transforming 197,000+ raw records into actionable business intelligence using Microsoft SQL Server.

Tech Stack: Microsoft SQL Server (SSMS)

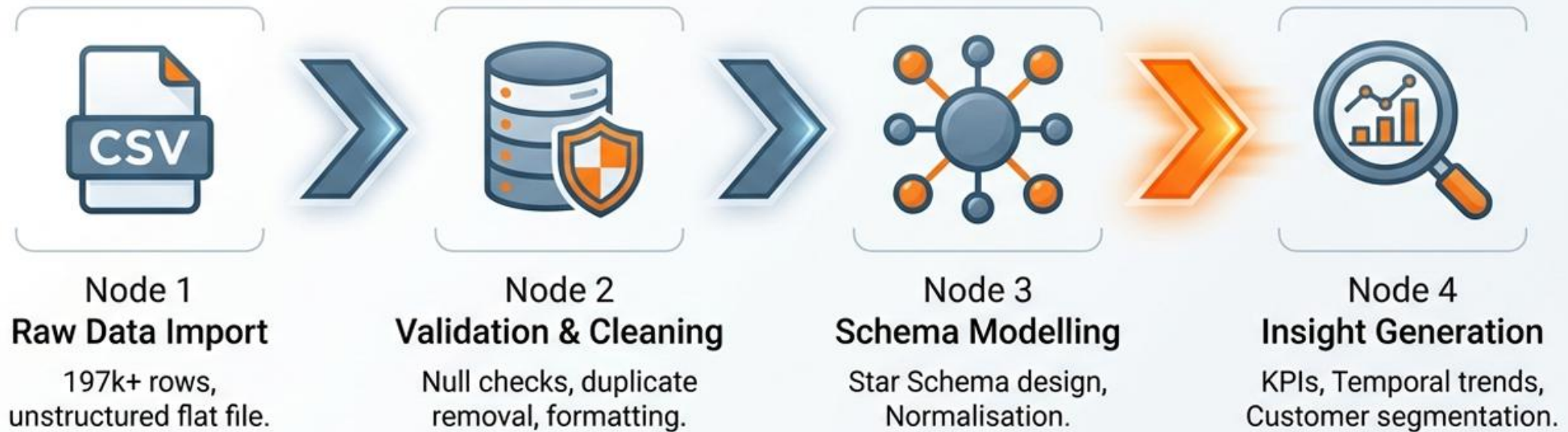
Scope: End-to-End ETL (Extract, Transform, Load) & Analysis

Dataset: Real-time sales data spanning three quarters



Project Overview: From Raw Chaos to Structured Insight

The objective was to simulate a real-world data analysis lifecycle, converting bulky, unstructured sales data into a reliable source of truth.



Data Inc: Data Import and Initial Inspection

Resolving Schema Mismatches in the Raw Flat File

The Challenge

The raw source was a single CSV file containing ~197,430 rows. Initial import attempts failed due to rigid default constraints.

- Dish Names exceeded standard character limits.
- Dates risked being interpreted as text strings.



Error: Import Failed

String or binary data would be truncated.
Statement terminated.

The Fix

Adjusting the schema definition to accommodate real-world variability.

```
-- Schema Definition Update  
DISH_NAME: VARCHAR(200) -- Expanded from default 50  
ORDER_DATE: DATE -- Enforced for time-series analysis  
STATE: VARCHAR(150)
```


Data Validation: The Null Value Check

*"Garbage In, Garbage Out —
Validation is the safety net
of analysis."*

The Logic

```
SELECT
  SUM(CASE WHEN State IS NULL THEN 1
        ELSE 0 END) as Null_State,
  SUM(CASE WHEN City IS NULL THEN 1 ELSE
        0 END) as Null_City,
  SUM(CASE WHEN Restaurant_Name IS NULL
        THEN 1 ELSE 0 END) as Null_Restaurant
FROM swiggy_data;
```

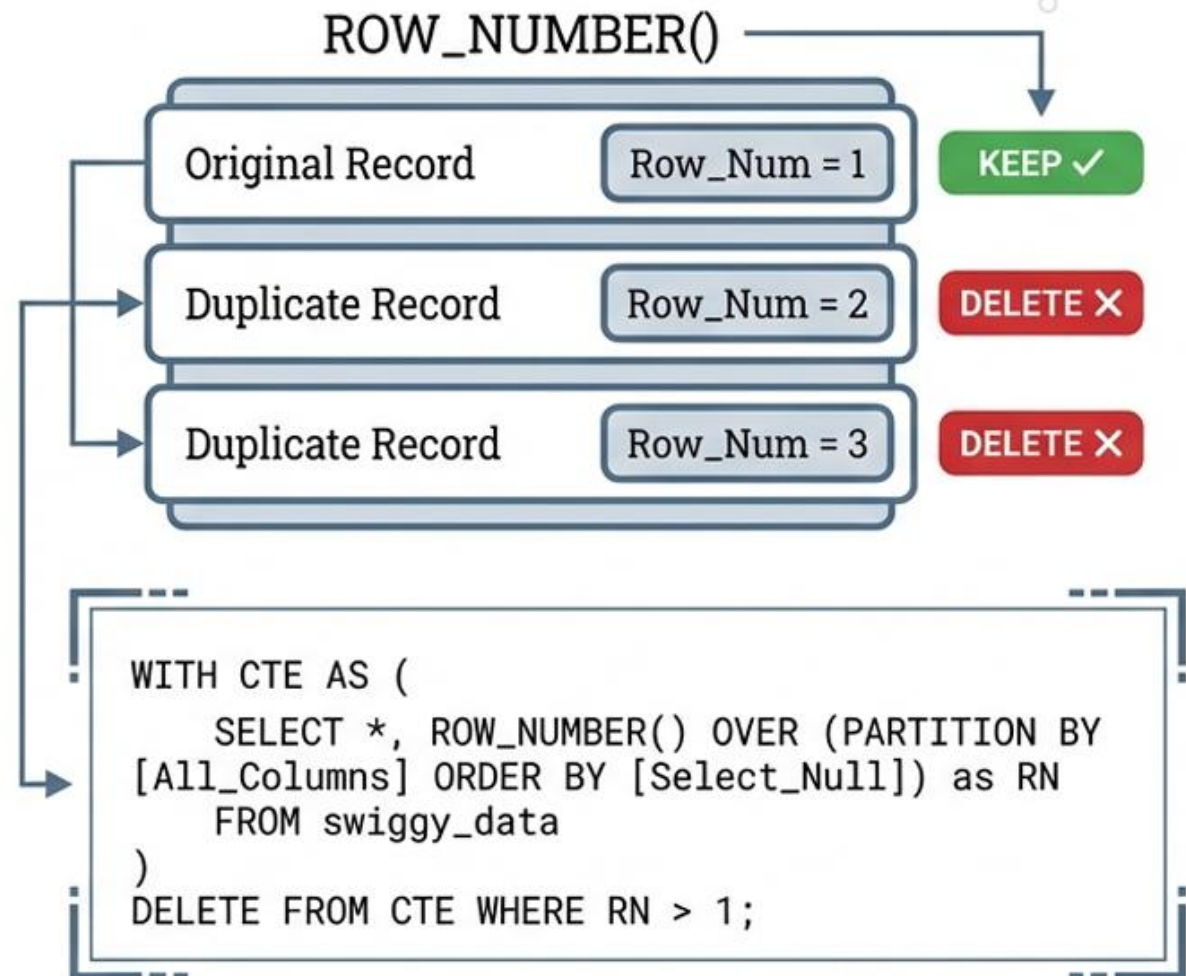
Output Result

Null_State	Null_City	Null_Restaurant	Null_Dish
0	0	0	0

Result: Zero null values confirmed across all dimension columns.

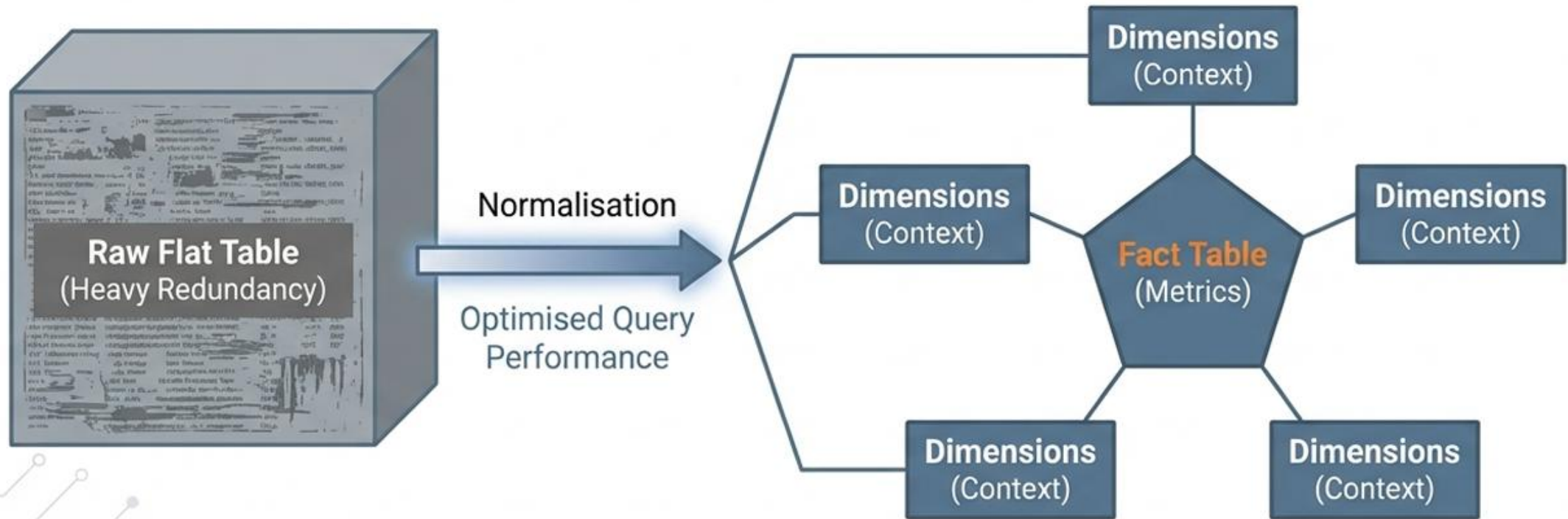
Advanced Cleaning: Eliminating Redundancy

A duplicate check revealed 29 records with identical transactional entries. A Common Table Expression (CTE) with a Window Function was deployed to isolate and remove them.



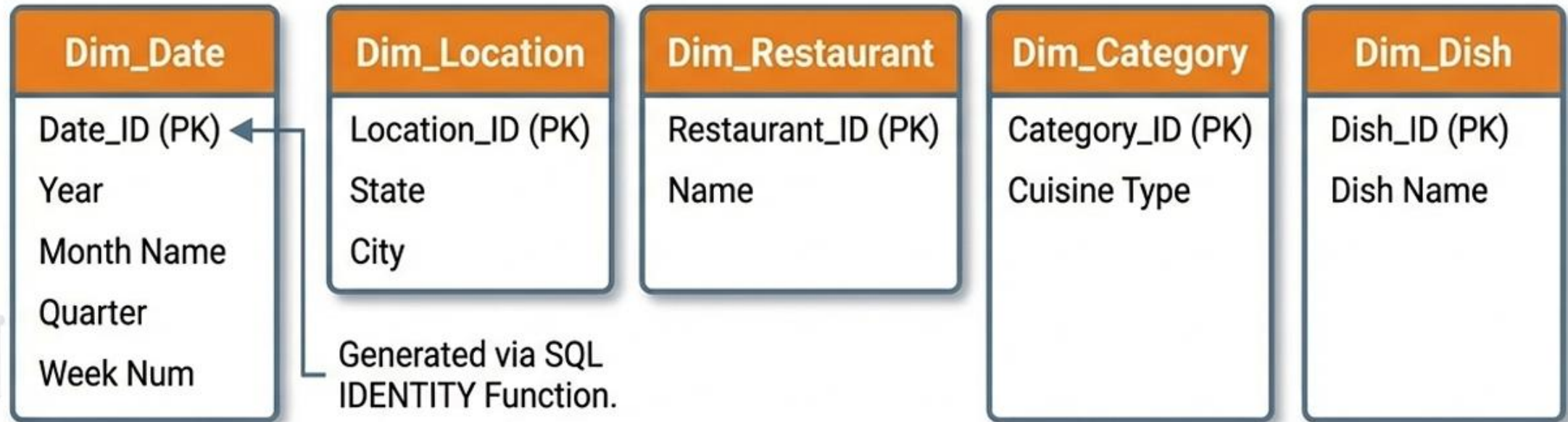
Strategic Data Modelling: The Star Schema

The raw data was a flat, bulky table with heavy text redundancy.
Normalisation splits this into a performant **"Star Schema"** structure.



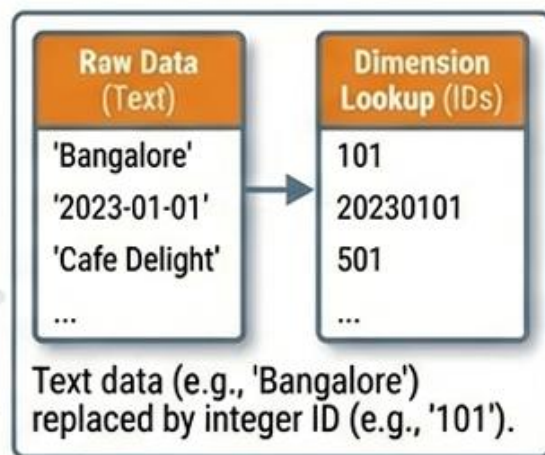
Building the Dimensions

Extracting categorical context into five distinct tables using **IDENTITY(1,1)** for automated Primary Key generation.






Constructing the Central Fact Table

The `Fact_Swiggy_Orders` table stores the transactional truth, linking foreign keys to measures.

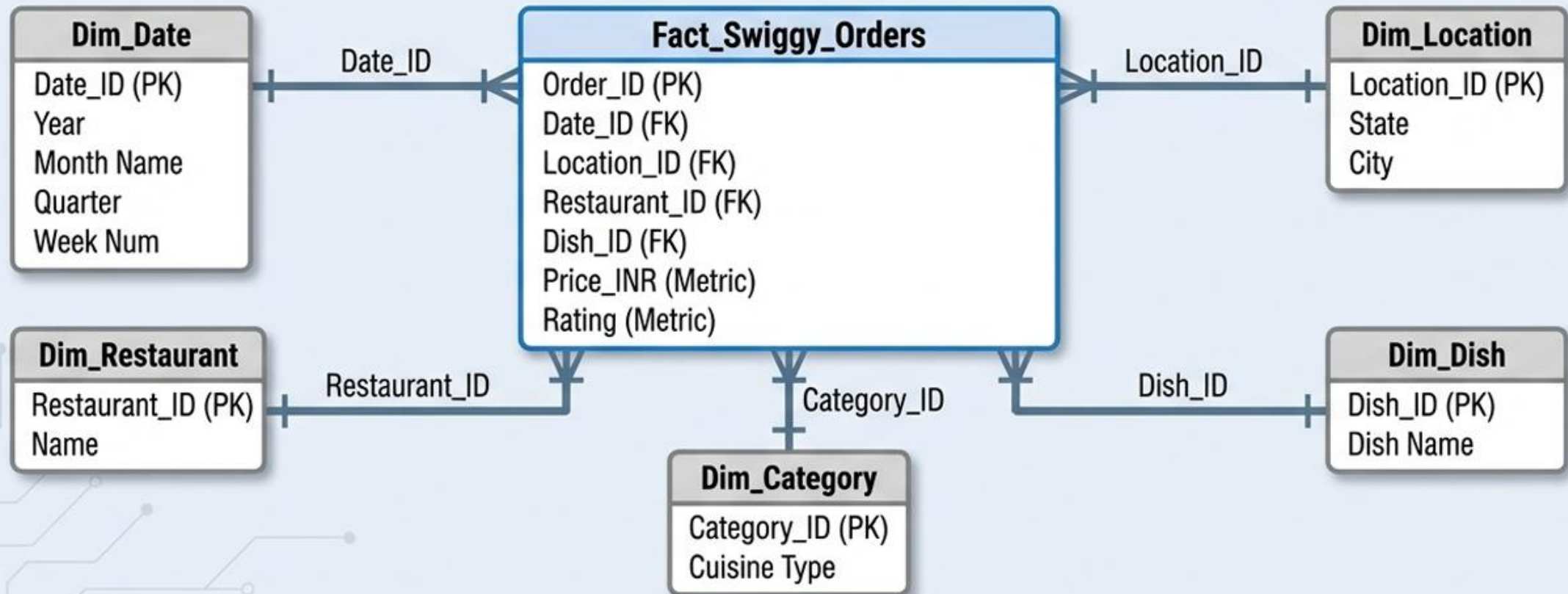


INSERT INTO...
SELECT...
JOIN dimensions
ON string_match

Fact_Swiggy_Orders					
Order_ID  (Primary Key)	Date_ID  (FK)	Location_ID  (FK)	Restaurant_ID  (FK)	₹ Price_INR (Metric)	★ Rating (Metric)
1001	20230101	101	501	350.00	4.5
1002	20230101	102	502	220.50	3.8
1003	20230102	101	501	410.00	5.0
1004	20230101	102	502	220.50	4.5
1005	20230102	101	501	410.00	3.9
1006	20230103	101	501	410.00	5.0
1007	20230103	102	502	410.00	4.7
...
...

The Final Architecture: Entity Relationship Diagram (ERD)

A fully relational database structure optimised for scalable analysis.



High-Level KPIs: The State of Business

Total Orders

197,401

Verified Transactional Volume



Total Revenue

53 Million INR

Aggregated Sales Volume



Avg Dish Price

268 INR

Average Ticket Size



Avg Rating

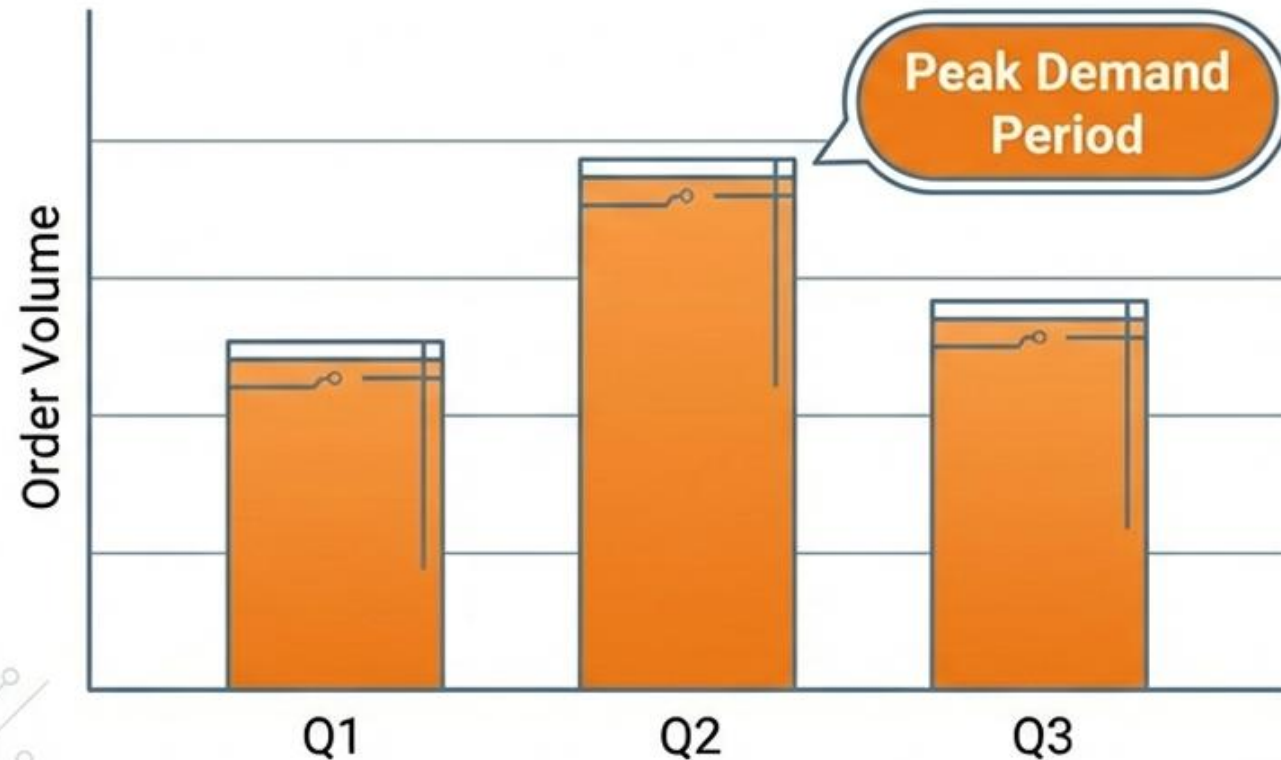
4.0 / 5.0

Customer Satisfaction Score



Temporal Trends: Monthly and Quarterly Growth

Analysis of order volume aggregated by **Dim_Date** attributes.



Monthly Insights:

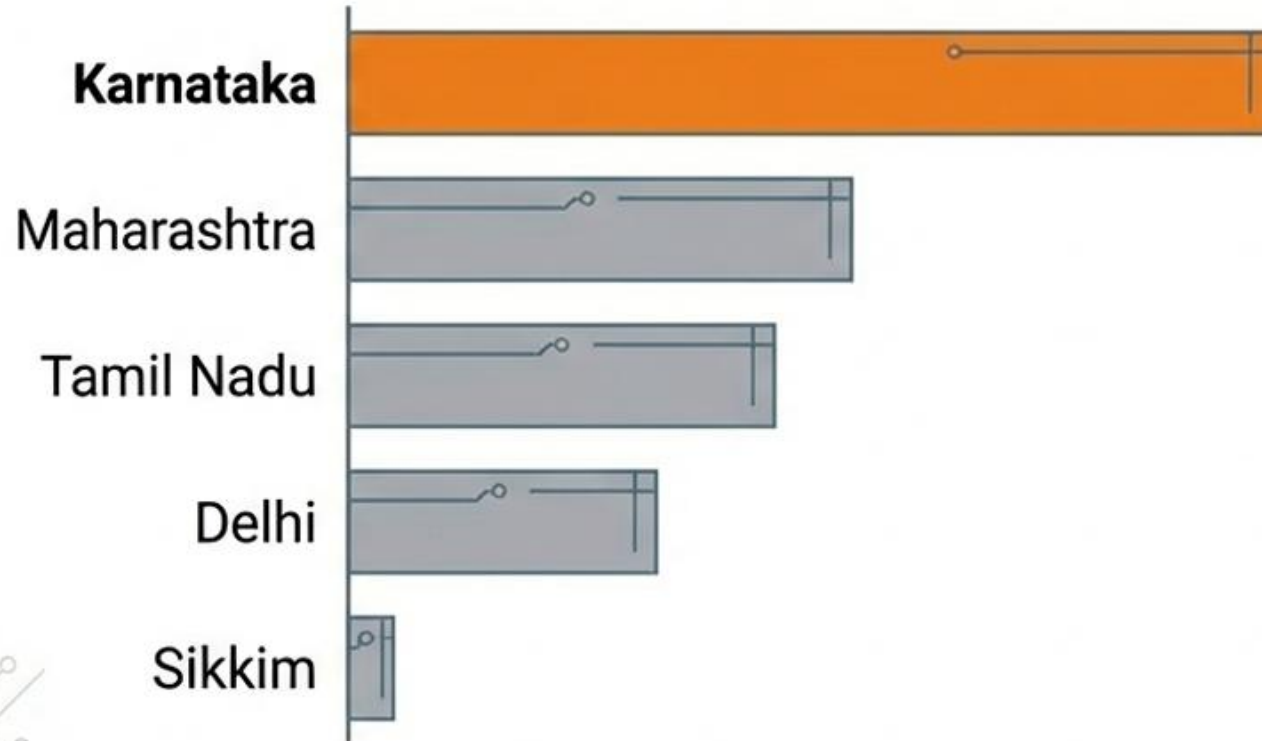
- Top Months: January, May, August.

SQL Technique:

JOIN Fact with Dim_Date,
GROUP BY Month/Quarter.

Geographic Insights: Location Performance

Revenue contribution and order volume by **`Dim_Location`**.



Top Cities (Order Volume)

1. Bangalore
2. Gurgaon
3. Hyderabad
4. Noida
5. New Delhi

Product & Partner Performance

Top Partner Brands

1.	 Domino's Pizza
2.	 KFC
3.	 Pizza Hut
4.	 Burger King
5.	The Good Bowl

Most Popular Items

Menu

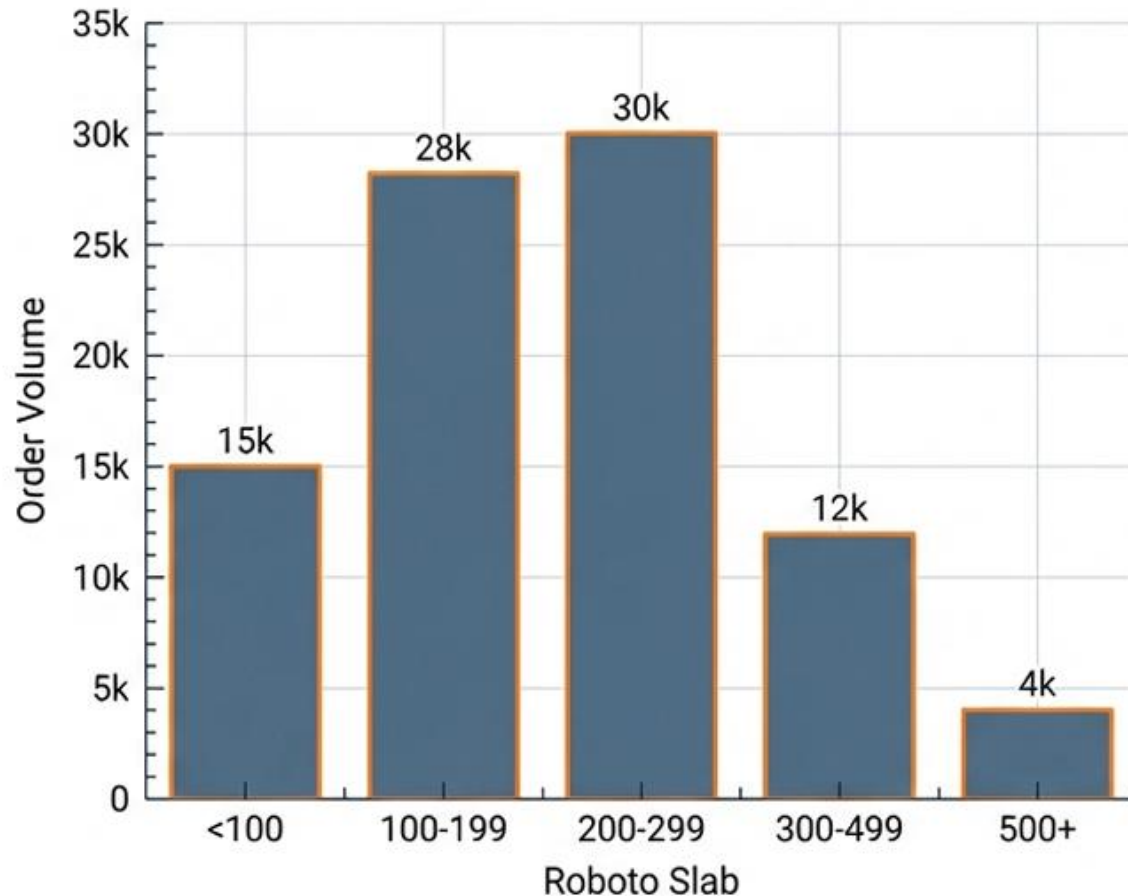
- ☒ Veg Fried Rice **Top Seller**
-  Choco Lava Cake
 - Paneer Butter Masala
 - Chicken Biryani

Insight: The 'Recommended' category in the app drives the highest volume of orders.

Customer Segmentation & Satisfaction

Roboto Slab: Bucketing order value and rating distribution using SQL `CASE` statements.

Order Volume by Price Bucket (INR)



Ratings Distribution



Conclusion: The Value of Structured Analysis



Data Integrity

Rigorous SQL validation ensured zero nulls and unique transactional records.



Scalability

Dimensional modeling (Star Schema) optimised the dataset for complex, high-performance querying.



Actionable Insight

Granular analysis identified Karnataka as a key market and the 100-300 INR price range as the volume driver.

“Data Analysis isn’t just about finding answers; it’s about building the structure that makes answers possible.”