

Mobile Analytics Project

Project Requirements

1. Source Data

- The dataset named smartphones contains 980 smartphone records.
- It includes information on brand, model, specifications, performance, and pricing.
- Each row represents one smartphone model and its key attributes such as processor, memory, display, camera, battery, OS, and price.

2. Normalization & Data Warehouse Design

- The source table contains duplicate and redundant data (e.g., brands, operating systems, processors appear repeatedly).
- To ensure efficient querying and reduce redundancy, the dataset has been normalized using a Star Schema approach.
- The schema consists of multiple Dimension Tables and one Fact Table that links them together.

3. Dimension Tables

The following dimension tables store descriptive information:

Dimension Name	Key Columns	Description
Dim_Brand	brand_id, brand_name	Contains unique smartphone brands.
Dim_Model	model_id, model_name, brand_id	Stores smartphone model names and their brand relationships.
Dim_Processor	processor_id, processor_brand, num_cores, processor_speed	Details about smartphone processors.

Dimension Name	Key Columns	Description
Dim_Battery	battery_id, battery_capacity, fast_charging_available, fast_charging	Battery information including fast-charging capability.
Dim_Memory	memory_id, ram_capacity, internal_memory, extended_memory_available	Memory configuration of each smartphone.
Dim_Display	display_id, screen_size, refresh_rate, resolution_height, resolution_width	Screen and display attributes.
Dim_Camera	camera_id, num_rear_cameras, primary_camera_rear, primary_camera_front	Camera setup information.
Dim_OS	os_id, os	Operating system used by the device.

Each dimension table has a primary key that serves as a foreign key reference in the fact table.

4. Fact Table

- The fact_smartphones table stores measurable data and foreign key references to each dimension.
- **Columns include:**
 - smartphone_id (Primary Key)
 - brand_id, model_id, processor_id, battery_id, memory_id, display_id, camera_id, os_id (Foreign Keys)
 - price, avg_rating, is_5G (quantitative and categorical facts)

5. ETL Process

The ETL pipeline was implemented in MySQL using Stored Procedures:

- **Extract:** Source data from the raw smartphones table.
- **Transform:**
 - Remove duplicates using ROW_NUMBER() window function.

- Normalize the dataset and map each unique record to its corresponding dimension.
- **Load:**
 - Insert clean, structured data into the Fact and Dimension tables.

Procedures used:

- Dimensiontable() → Creates and loads all dimension tables.
- Facttable() → Creates and loads the fact table with deduplicated, normalized data.

6. Objectives

- Build a **Mobile Analytics Data Warehouse** using normalized schema design.
- Perform **descriptive and comparative analysis** using SQL, CTEs, and Window Functions.
- Enable **data-driven insights** for smartphone market trends, pricing, and performance.

7. Analytical Questions:

1. Total Sales and Average Price by Brand.
2. Top 5 Smartphones by Rating and Price.
3. Smartphone Price Distribution by Brand and OS.
4. Market Share by Brand and Processor Speed.
5. Number of Models and Avg Price by RAM Size and Brand.
6. Top 3 Fastest Charging Smartphones by Price.
7. Brand Performance by 5G Availability.
8. Correlation Between Processor Speed and Price by Brand.
9. Price-to-Performance Ratio by Brand.
10. Most Popular Display Features.
11. Multi-Feature Ranking (Composite Score).
12. Average Battery Capacity and Fast-Charging Comparison by OS.