Business Inteligence and Data Warehouse Mobile Dataset

Group Member :

* Harold Hamdani - 2702342501
* Juvensen Ivanlie - 2702226312
* Lie Richard - 2702276241
* Johansenn Wesley Chu - 2702267565
* Timothy Rafael Darmadi - 2702328951

1. Introduction

In the digital age, global smartphone markets generate vast amounts of data across regions, brands, and product specifications. Businesses must analyze this data efficiently to make informed decisions on pricing strategies, market expansion, and product development. The dataset used in this project contains transactional-level information about smartphone models, their technical specifications, and their corresponding prices across five countries USA, India, China, Dubai, and Pakistan. Each price point is recorded separately per country, while the products are associated with specific companies and technical features such as RAM, camera quality, and battery capacity. To transform this raw data into meaningful insights, a structured approach using data warehousing is required.

Data warehousing serves as a foundational technology to consolidate, clean, and organize large, diverse datasets into a central repository optimized for analytical querying. By transforming this multi-file, multi-dimensional transactional data into a star schema-based data warehouse, we enable better data integration, consistency, and accessibility. This not only simplifies analytical processes like trend analysis and price comparison but also supports strategic decision-making by offering a unified view of product pricing across international markets.

1. Objective

The objective of building a data warehouse from this dataset is :

1. To Identifying key smartphone specifications that correlate with market demand and pricing tiers, enabling informed R&D investments and product feature prioritization
2. To inform dynamic pricing adjustments, localized promotions, and tailored product offerings that account for regional market sensitivities and competition
3. To help businesses strategically position their offerings volume vs. value and optimize resource allocation across different product lines.
4. To direct targeted marketing, sales promotions, and inventory management, especially during peak demand periods.
5. Methodology
6. Transform the Dataset

The initial step in the methodology involved transforming the raw dataset into a structured format suitable for data warehousing. The original dataset was gotten from Kaggle by Abdul Malik ([Kaggle link](https://www.kaggle.com/datasets/abdulmalik1518/mobiles-dataset-2025)) , the original dataset contained the following attributes:

* Company Name
* Model Name
* Mobile Weight
* RAM
* Front Camera
* Back Camera
* Processor
* Battery Capacity
* Screen Size
* Launched Price (Pakistan)
* Launched Price (India)
* Launched Price (China)
* Launched Price (USA)
* Launched Price (Dubai)
* Launched Year

To prepare the data for analytical purposes, a normalization was developed to understand entity relationships and their multiplicities. Subsequently, normalization was applied up to the **Third Normal Form (3NF)** to remove redundancy and ensure data integrity ([Link of Normalization Process](https://docs.google.com/spreadsheets/d/17-9T59tjfwcbO_DK4OLsXgdYpZtrpKZJlbu1OeLcGQE/edit?usp=sharing)). This process allowed the creation of a clean and logical **Entity-Relationship Diagram (ERD)**. ERD helps establish the foundational data model, focusing on entities and their relationships, which reflects the final schema structure presented in the attached diagram. As we see in Fig. 1 is the result of the final ERD.

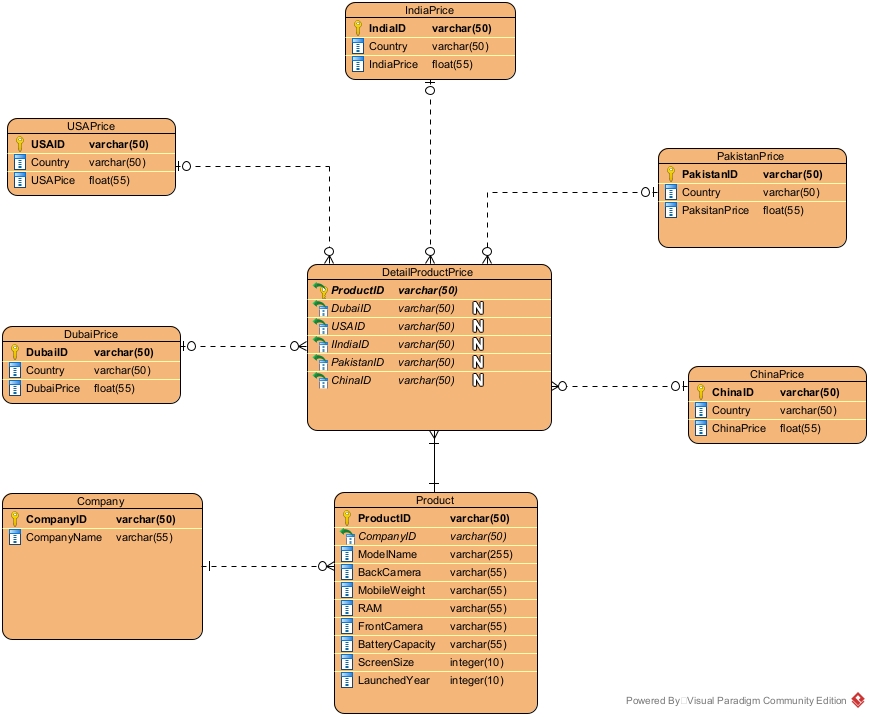


Fig. 1 ERD MobileDataset

1. Design the Data Warehouse
2. Schema Choice:

The schema design adopted for this data warehouse is the **Star Schema**, chosen for its simplicity, ease of understanding, and optimal performance in read-heavy analytics environments such as Business Intelligence (BI) projects. Compared to the **Snowflake Schema**, which normalizes dimension tables into multiple related tables, the Star Schema reduces query complexity by denormalizing dimensional data that enhances performance in visualization tools like Tableau. Since the ERD already looks like Star Schema we don't have to design the data warehouse.

1. Define Tables:

Define Fact Table

1. DetailProductPrice

This table serves as the central fact table, linking each product to its corresponding price across different countries. The primary measures or quantitative metrics are the product prices in various regions (e.g., USAPrice, IndiaPrice, etc.). It also contains foreign keys to the dimension tables, forming the "star" around which the schema is designed.

Define Dimension Tables

1. Product

This dimension describes detailed product specifications such as Model Name, RAM, Camera, Battery, etc. These attributes support slicing and dicing during multidimensional analysis.

1. Company

This table stores information about the manufacturing companies for each product.

1. DubaiPrice, USDPrice, IndiaPrice, PakistanPrice, ChinaPrice

Although these tables are labeled as "Price", they serve as **dimension tables** because they provide contextual information about product pricing within each geographic location. These dimensions allow analysis from a geographic pricing perspective. Optionally, they can be denormalized into a single DimCountry table for simplification.

1. Database Tools:

* MySQL

The data warehouse was implemented using **MySQL** as the database management system. MySQL was selected due to its wide adoption, open-source licensing, strong community support, ease of integration with BI tools like **Tableau**, and robust support for data manipulation and querying via **SQL**. Its reliability, performance, and accessibility make it an ideal choice for academic and medium-scale professional projectbi.

1. Implement ETL Process
2. Extract:

The dataset was separated into multiple CSV files, where each file represents a class (or table) derived from the normalized ERD structure. These files were then imported into MySQL under a unified database named BIProject.

1. Transform:

As shown in Fig. 2, Fig. 3, and Fig.4 during the transformation phase, several issues were encountered:

1. **Data Type Mismatches:** Many price columns were incorrectly read as strings due to inconsistent formatting (e.g., currency symbols, commas). These were corrected by cleaning and converting them into appropriate float data types.
2. **Attribute Renaming (Schema Alignment):** To maintain consistency with the finalized ERD, original column names from the source dataset were renamed using ALTER TABLE CHANGE COLUMN statements.
3. **Standardizing Data Formats:** Columns such as LaunchedYear were explicitly cast into appropriate types (INTEGER) using ALTER TABLE to enable consistent filtering, grouping, and temporal analysis.

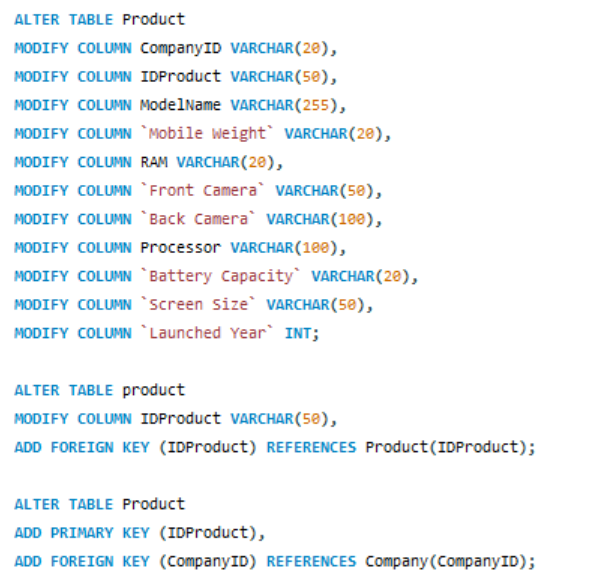


Fig.2 Attribute Renaming and Standardizing Data Formats

1. **Foreign Key and Primary Key Constraints:** Defined with ADD FOREIGN KEY and ADD PRIMARY KEY statements to ensure referential integrity across tables.

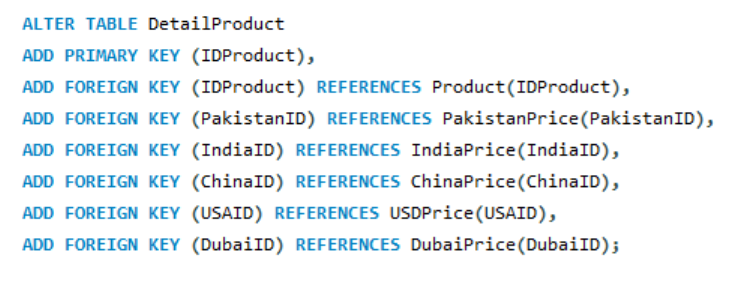


Fig. 3 Alter Table For Foreign Key

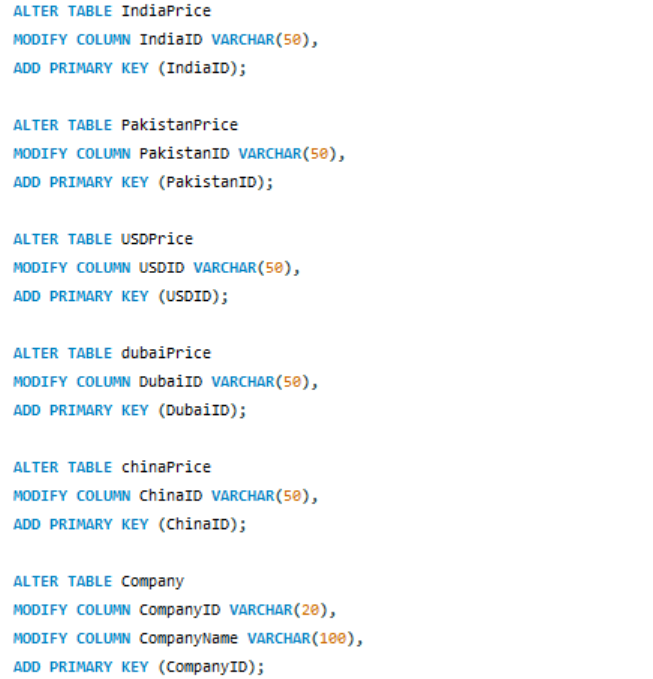


Fig. 4 Alter Table For Primary Key

1. **Cleaning and Validation:** Duplicates and null entries were identified and removed. String-to-float conversion errors and other data inconsistencies were addressed through validation scripts.
2. **Data Enrichment and Aggregation:** Although minimal, some data was aggregated (e.g., average prices per country) and enriched with additional flags or codes as needed for improved analysis during visualization.
3. Load:

Although no restructuring was needed between the ERD and the data warehouse schema due to the direct mapping via Star Schema, the Load phase was completed using INSERT INTO SELECT MySQL queries to move data into the final tables.

1. ETL Tools:

SQL scripts

The ETL process in this project was executed using SQL scripts to provide full control over data transformation and ensure compatibility with MySQL and Tableau. This approach allowed direct manipulation of data through commands like ALTER, MODIFY, and INSERT, making the process transparent, easy to debug, and cost-effective. Using SQL was ideal for this scale of project, offering a simple yet powerful solution without the need for complex or external ETL tools.

1. Findings

Business Intelligence & Visualization

1. Key Dashboards:
2. Product Launch Price Dashboard: Metrics: Launched Price of Each Product in Different Countries, Heat Map Based on Company Product Launch Price.
3. Product Launched by Each Company Dashboard: Metrics: Products Made by Company, Product Launched Every Year by Every Company.
4. Phone Specifications Dashboard: Metrics: Phone Specifications.
5. Product Price Insight Dashboard: Metrics: Highest Launch Price Product, Lowest Launch Price Product
6. Visualization Tools:

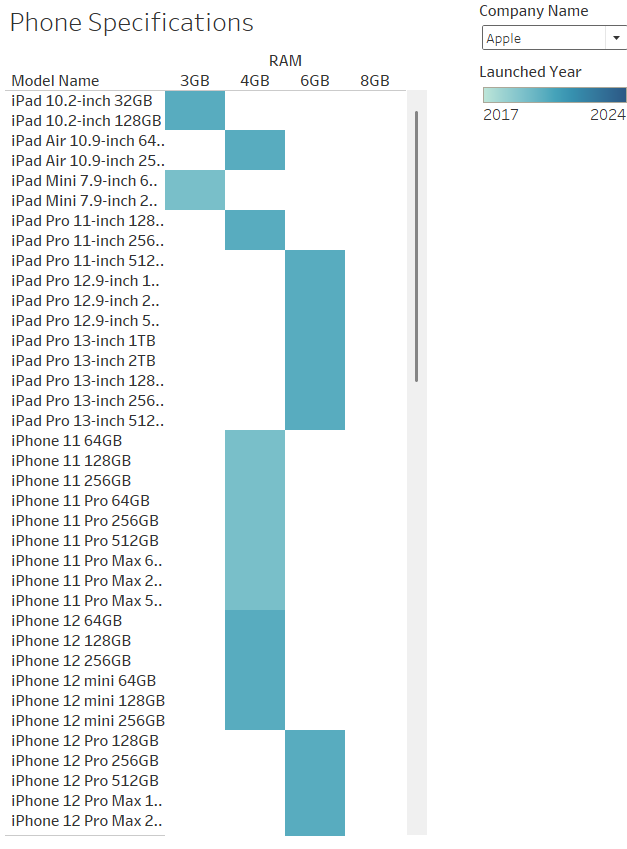
Tableau

For analysis and visualization, Tableau was used. Its powerful drag-and-drop interface, seamless integration with MySQL, and robust support for analytical operations made it ideal for this project. Tableau enables rapid prototyping of dashboards and clear visual representation of data such as price comparisons, market trends, and product feature distributions across countries.

Here the link of the visualization in tableu :

[Visualization Link](https://public.tableau.com/app/profile/lie.richard/viz/VisualisasiBID/PhoneSpecificationsDashboard)

1. Insights & Recommendations
2. **Phone Specifications Dashboard**

****

**Business Trends:**

* The trend shows that products with higher specifications (RAM, storage, battery) are increasingly more common.
* Mid-range phones are offering flagship-level specs, increasing competition in that segment.

**Product Performance:**

* Devices with 6GB+ RAM, >4000mAh battery, and large displays are more frequently launched.
* These specs correlate with moderate-to-high price points and strong consumer demand.

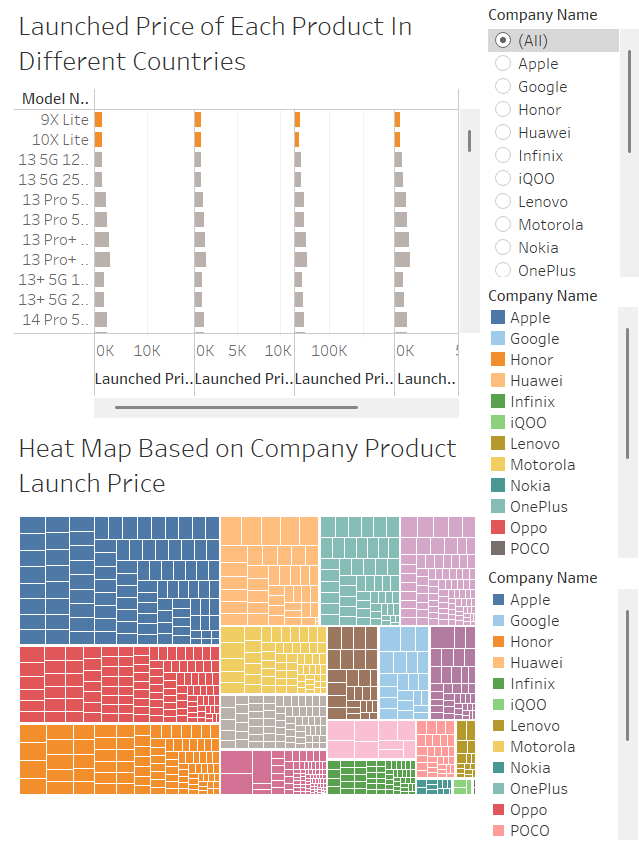
**Recommendations:**

"Focus product development and marketing on phones with higher RAM and battery, especially in the mid-range price bracket. These specs are now a baseline expectation, and meeting them boosts competitiveness."

**Impact on Company:**

* Helps align R&D investment with market expectations.
* Improves product-market fit, increasing launch success rate.

1. **Product Launch Price Comparison Dashboard**

****

**Business Trends:**

* Significant variation in launch pricing across regions.
* Some companies adjust pricing more aggressively depending on the market.

**Product Performance:**

* Products launched in certain countries (e.g., India or Indonesia) are priced lower, potentially due to price sensitivity or local competition.

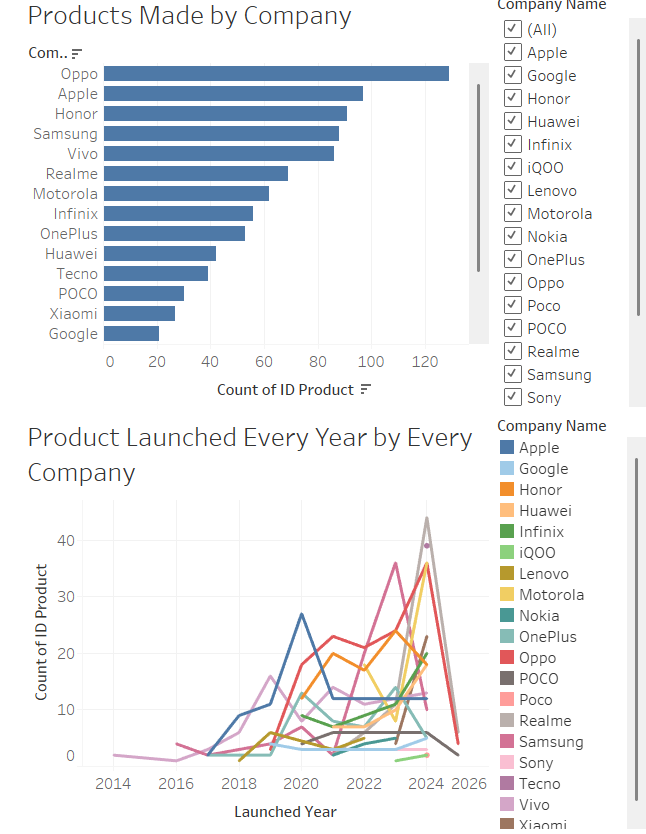
**Recommendations:**

"Apply price optimization strategies for each market. Consider bundling, promotions, or trimmed-down specs to cater to price-sensitive regions without affecting profitability."

**Impact on Company:**

* Enables competitive pricing.
* Helps avoid underpricing in premium markets or overpricing in emerging markets.

1. **Product Launched by Each Company Dashboard**

**Business Trends:**

* Certain brands (likely Xiaomi, Samsung, etc.) dominate the number of product launches. Other companies release fewer but more premium models.

**Product Performance:**

* Frequent launchers may saturate the market, while premium-focused companies retain niche strength.

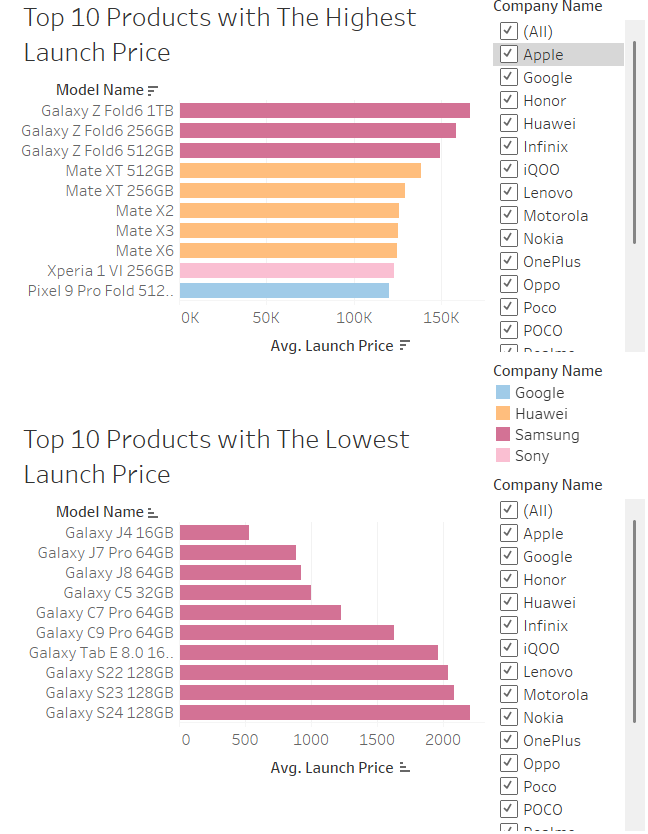
**Recommendations:**

"For companies launching fewer products, emphasize quality and innovation in marketing. For high-frequency launchers, streamline product lines to avoid cannibalization and focus on standout features."

**Impact on Company:**

* Helps companies align product strategy (volume vs. value).
* Reduces resource waste and clarifies branding.

1. **Product Price Insight Dashboard**



**Business Trends:**

* Launch prices follow a bell curve—most products are clustered in the mid-range.
* Premium devices are fewer but command much higher margins.

**Product Performance:**

* Top-performing models are not always the most expensive value-for-money models that often lead sales.

**Recommendations:**

"Increase marketing investment for high-margin, high-performing models. Consider launching special editions or bundles for top sellers to maximize Q4 revenue."

**Impact on Company:**

* Boosts profit by doubling down on successful products.
* Strengthens brand loyalty by promoting high-value models.

**Overall Strategy:**

To maximize revenue and competitiveness, companies should focus on launching mid- to high-spec devices within competitive price ranges, tailor pricing strategies based on each region’s purchasing power, and leverage data on top-performing products to drive targeted promotions—especially during high-demand periods like Q4.

1. Conclusion

This project successfully built a star schema data warehouse for global smartphone data, transforming raw information into an organized, analytical tool. Through a careful ETL process, we ensured data was clean and ready for analysis using Tableau. This enabled us to uncover crucial insights: that high-spec phones (more RAM, bigger batteries) are in high demand, signaling where product development should focus; that pricing needs to be tailored for each country to maximize sales and avoid missed opportunities; and that understanding which companies launch how many products helps in shaping overall market strategy. Ultimately, this data warehouse provides a powerful, unified view of the international smartphone market, allowing businesses to make proactive, data-driven decisions that improve product fit, optimize pricing, and strengthen their competitive position for sustained growth.