

Introduction

The `LaptopFeatures` Database in this dataset provides a comprehensive overview of **823 different laptop models**, including their **technical specifications, features, and market data**. This dataset is structured with **19 attributes**, covering aspects such as **brand, hardware configuration, software details, pricing, and user reviews**.

This dataset is valuable for **price prediction models, market analysis, and feature-based comparisons** of laptops. Below is a breakdown of the key attributes:

1. Brand & Processor Specifications

- **Brand**: Identifies the manufacturer (e.g., ASUS, Lenovo, HP, Dell, etc.).
- **Processor Brand**: Specifies the CPU manufacturer (Intel or AMD).
- **Processor Name**: Describes the processor series (e.g., Core i3, Core i5, Ryzen 5).
- **Processor Generation**: Indicates the CPU generation (e.g., 10th Gen, 11th Gen).

These attributes help in understanding **brand popularity, processor trends, and performance benchmarks**.

2. Memory & Storage Configuration

- **RAM (Size & Type)**: Specifies the RAM capacity (e.g., 4GB, 8GB) and memory type (DDR4, LPDDR4).
- **Storage Options**: Includes both **SSD and HDD capacity** to analyze storage preferences.

This section is useful for **understanding the impact of RAM and storage on pricing and performance**.

3. Operating System & Architecture

- **Operating System**: Specifies the OS installed (Windows, macOS, Linux, etc.).
- **OS Architecture**: Indicates whether the OS is **32-bit or 64-bit**.

These attributes help in **analyzing OS trends and compatibility with modern software**.

4. Graphics & Display Features

- **Graphic Card (Dedicated vs. Integrated)**: Specifies the **GPU memory (in GB)**, indicating whether the laptop has dedicated graphics for gaming, content creation, or professional workloads.
- **Touchscreen Support**: Identifies whether the laptop has a touchscreen display.

These features are crucial for **gaming laptops, creative professionals, and general usability analysis**.

5. Additional Features & Build Quality

- **Weight**: Helps in differentiating between **ultrabooks, standard laptops, and gaming laptops**.
- **Warranty**: Indicates the presence of a manufacturer warranty.
- **Microsoft Office Pre-installed**: Shows whether the laptop comes with **MS Office included**.

These attributes contribute to **user convenience and purchasing decisions**.

6. Pricing & Market Data

- **Price**: Represents the laptop's selling price in the market.
- **User Rating**: Captures **customer satisfaction** through a star rating system.
- **Number of Ratings & Reviews**: Provides insights into **popularity and consumer trust**.

This section is highly valuable for **price trend analysis, demand forecasting, and sentiment analysis**.

Use Cases & Applications

This dataset is ideal for:

- ✓ **Market research** – Identifying popular brands and configurations.
- ✓ **Consumer behaviour analysis** – Understanding how specifications influence ratings and sales.
- ✓ **Feature impact assessment** – Evaluating how different attributes (RAM, SSD, GPU) affect pricing.

Data Cleaning Process

To ensure accuracy and consistency in the dataset, a structured data cleaning process was applied. The following steps were undertaken to refine the data and remove inconsistencies:

1. Handling Missing Values

- ✓ **Identification of Null Values** – Checked for missing values in each column.
 - ✓ **Removal of Incomplete Records** – Entries with missing values in critical fields (**brand**, **OS**, **processor details**, **RAM**, **price**) were **removed** to maintain data integrity.
-

2. Removing Duplicates

- ✓ Checked for **duplicate laptop models** based on brand, processor, RAM, and storage attributes.
 - ✓ Removed redundant entries to **avoid data bias** in analysis.
-

3. Handling Inconsistent Data Entries

- ✓ **Storage & RAM Mismatches** – Ensured that SSD and HDD sizes were correctly summed up for total storage.
 - ✓ **Unrealistic Values** – Removed outliers such as "500GB RAM" or "1TB RAM" which were likely errors.
-

4. Creating Dimension Tables (Dim*)

- ✓ Extracted categorical attributes like **brand**, **OS**, **processor**, and **laptop type** into separate dimension tables.
 - ✓ Assigned unique IDs for **lookup tables**, ensuring **data normalization and reduced redundancy**.
-

5. Creating the Fact Table Using VLOOKUP

After organizing the dataset into dimension tables, a Fact Table (**FactLaptop**) was created using VLOOKUP to replace categorical values with their respective IDs. This step:

- ✓ Replaced text-based categories (Brand, OS, Processor, etc.) with foreign keys from the respective dimension tables.

- ✓ Used VLOOKUP to map and retrieve the correct IDs from DimBrand, DimOS, DimProcessor, etc.
- ✓ Structured the Fact Table (FactLaptop) to store only numerical IDs and measurable attributes (Price, RAM, Storage, Ratings), improving query performance and data integrity.

Example formula used:

```
=VLOOKUP(WorkingData!B3,DimProcessor!$D$1:$E$29,2,0)
```

Benefits of Cleaning the Data

- ✓ Improved Accuracy – Reducing incorrect or missing information enhances analysis reliability.
- ✓ Better Query Performance – Normalized data structure helps in faster lookups and optimized data retrieval.
- ✓ Scalability – New categories can be added to dimension tables without modifying the Fact Table.

This cleaning process ensures that the dataset is ready for regression analysis, market research, and feature impact assessment.

Schema Design

The dataset includes several **dimension tables** that provide structured reference data for key laptop attributes. These tables serve as lookup tables for efficient data organization and analysis.

1. Brand Dimension Table (DimBrand)

- **brand_id** – Unique identifier for each brand.
- **brand** – Name of the laptop manufacturer (e.g., ASUS, HP, Dell, Lenovo).
- **Size:** 8 entries

This table helps standardize **brand names** across the dataset for consistency in analysis.

 DimBrand	...
brand	
brand_id	
Collapse ^	

2. Operating System Dimension Table (DimOS)

- **os_id** – Unique identifier for each OS.
- **os** – Name of the operating system (e.g., Windows, macOS, Linux).
- **Size:** 3 entries

This table allows easy **categorization of laptops based on their OS**.

 DimOS	...
os	
os_id	
Collapse ^	

3. Architecture Dimension Table (DimArchitecture)

- **architecture_id** – Unique identifier for each OS architecture.
- **architecture** – Specifies whether the system is **32-bit or 64-bit**.
- **Size:** 2 entries

Useful for analyzing **software compatibility and performance differences**.

DimArchitecture	...
architecture	
architecture_id	
Collapse ^	

4. Laptop Type Dimension Table (DimType)

- **type_id** – Unique identifier for each type.
- **type** – Defines the **category of the laptop**, such as **Ultrabook, Gaming, Standard**.
- **Size:** 3 entries

Helps segment **laptops based on their use case**.

DimType	...
type	
type_id	
Collapse ^	

5. Processor Dimension Table (DimProcessor)

- **processor_id** – Unique identifier for each processor type.
- **processor_brand** – Manufacturer of the processor (Intel or AMD).
- **processor_name** – Name of the processor (e.g., Core i5, Ryzen 7).
- **processor_gnrt** – Processor generation (e.g., 10th Gen, 12th Gen).
- **processor_details** – Consolidated data of brand, name, generation.
- **Size:** 28 entries

This table is crucial for **performance benchmarking, price comparison, and trend analysis**.

DimProcessor	...
processor_brand	
processor_details	
processor_gnrt	
processor_id	
processor_name	
Collapse ^	

Purpose of These Dimension Tables

- ✓ Ensures **data consistency** by standardizing key attributes.
 - ✓ Supports **faster lookups** and **efficient filtering** in database queries.
 - ✓ Facilitates **categorization and grouping** in **data analysis and visualization**.
-

Fact Table (FactTable)

- **brand_id** – Foreign key referencing DimBrand, representing the laptop's manufacturer.
- **processor_id** – Foreign key referencing DimProcessor, detailing the laptop's processor.
- **os_id** – Foreign key referencing DimOS, indicating the operating system.
- **architecture_id** – Foreign key referencing DimArchitecture, specifying the system architecture.
- **type_id** – Foreign key referencing DimType, categorizing the laptop type (e.g., Gaming, Ultrabook).
- **ram_gb** – The amount of system memory (e.g., 8GB, 16GB).
- **ram_type** – Type of RAM used (e.g., DDR4, LPDDR5).
- **ssd** – Storage capacity of the SSD in GB or TB.
- **hdd** – Storage capacity of the HDD in GB or TB.

- **graphic_card_gb** – Size of the dedicated GPU memory, if available.
- **warranty** – Duration of the warranty period in years.
- **Touchscreen** – Indicates whether the laptop has a touchscreen feature (Yes/No).
- **msoffice** – Indicates whether Microsoft Office is pre-installed (Yes/No).
- **Price** – The market price of the laptop.
- **rating** – Customer rating of the laptop.
- **Number of Ratings** – Total number of ratings received.
- **Number of Reviews** – Total number of customer reviews.

Facilitates price prediction, trend analysis, and performance evaluation by consolidating key attributes.

FactTable		...
	architecture_id	
	brand_id	
	graphic_card_gb	
	hdd	
	msoffice	
Σ	Number of Ratings	
Σ	Number of Reviews	
	os_id	
Σ	Price	
Collapse ^		

Star Schema for Laptop Dataset

The **Star Schema** is a widely used data modeling approach for organizing data in a relational database to support efficient querying and analysis. It consists of a **central Fact Table** surrounded by **multiple Dimension Tables**, forming a star-like structure.

Explanation of Schema Components

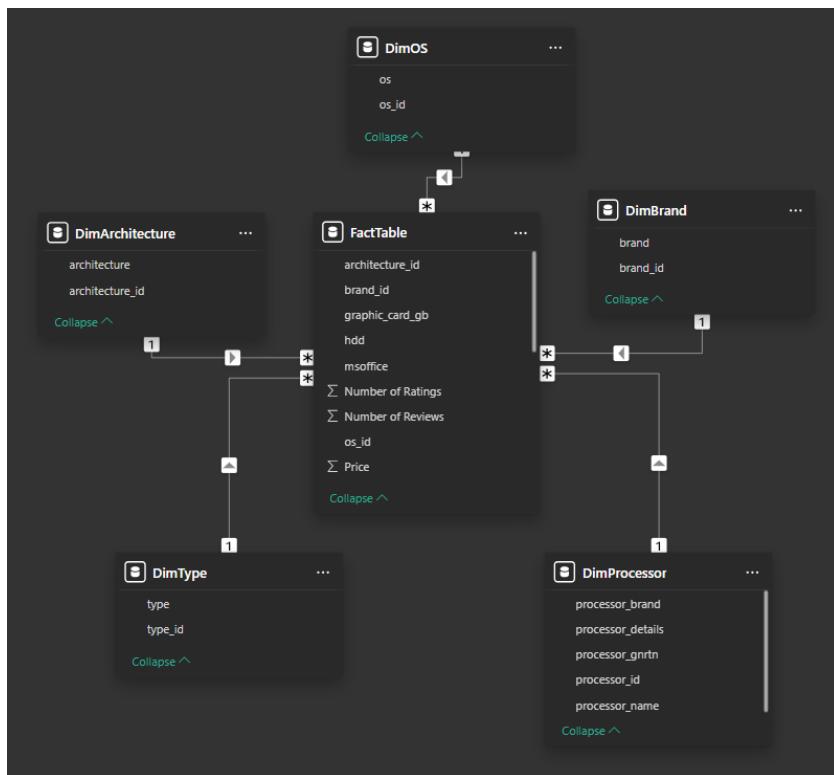
A. Fact Table: FactLaptop

- Stores measurable and transactional data.
- Contains foreign keys referencing dimension tables.
- Includes attributes required for price prediction and laptop analysis.

B. Dimension Tables

1. **DimBrand** – Stores laptop brand details (e.g., Dell, HP, Apple).
2. **DimProcessor** – Contains processor-related attributes (e.g., Intel i5, Ryzen 7).
3. **DimOS** – Stores details about the operating system (e.g., Windows, macOS).
4. **DimArchitecture** – Defines system architecture (e.g., 32-bit, 64-bit).
5. **DimType** – Categorizes laptops into different types (e.g., Gaming, Ultrabook).
6. **DimGraphics** – Contains information about the GPU and its memory.

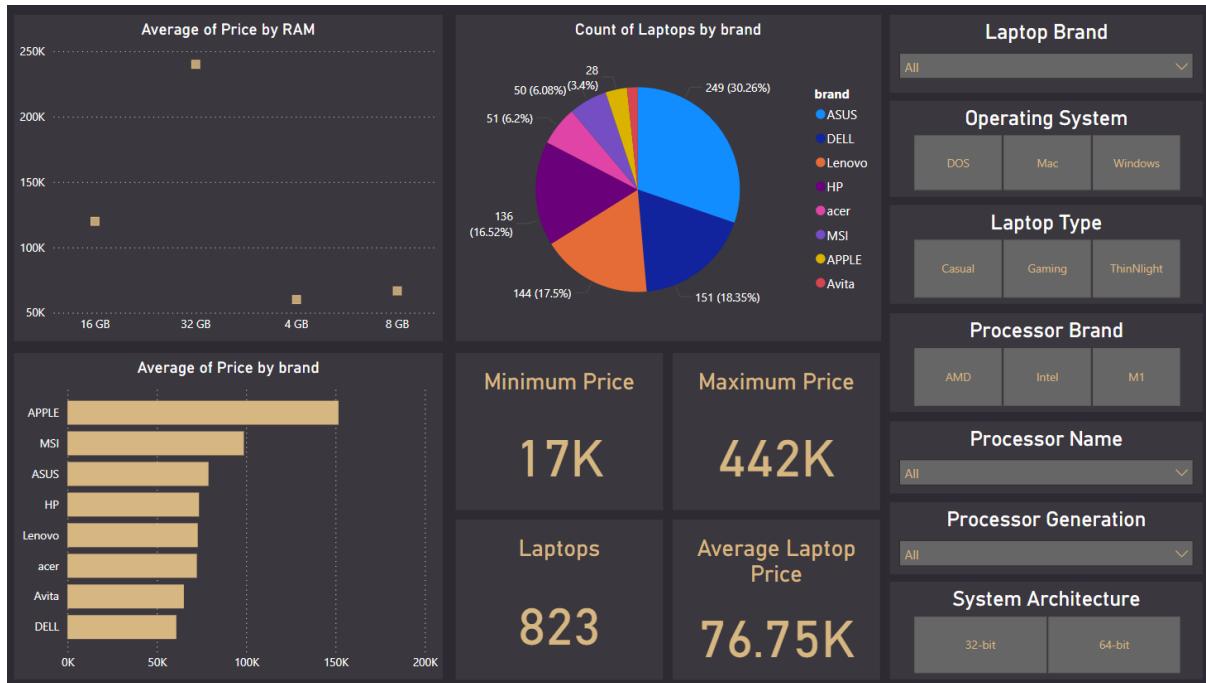
Diagram:



Benefits of Using Star Schema

- ✓ **Efficient Query Performance** – Joins are optimized due to the denormalized structure.
- ✓ **Easy to Understand & Maintain** – Simple relationships between fact and dimension tables.
- ✓ **Scalability** – New dimensions or facts can be added without disrupting existing structures.
- ✓ **Fast Aggregation & Reporting** – Designed for analytical queries in BI tools.

Dashboard



This **Power BI Dashboard** provides an **analytical overview of laptop pricing** based on various factors, such as **brand**, **RAM**, **operating system**, **processor**, and **system architecture**. It is designed to help users gain insights into laptop trends and pricing distribution.

1. Key Metrics and Insights

- Minimum Price: 17K** – The least expensive laptop in the dataset.
- Maximum Price: 442K** – The most expensive laptop in the dataset.
- Total Laptops Count: 823** – Total number of laptops analyzed.
- Average Laptop Price: 76.75K** – The mean price across all laptop brands.

2. Data Visualizations

A. Count of Laptops by Brand (Pie Chart - Center)

- Shows the **distribution of laptops** among different brands.
- ASUS (30.26%)** has the highest share, followed by **Lenovo (17.5%)**, **HP (18.35%)**, and others.

B. Average Price by RAM (Top Left - Scatter Plot)

- Displays the **average price of laptops based on RAM capacity** (4GB, 8GB, 16GB, 32GB).
- Higher RAM configurations tend to have **higher average prices**.

C. Average Price by Brand (Bottom Left - Bar Chart)

- **Apple has the highest average price**, followed by **MSI and ASUS**.
 - **Dell has the lowest average laptop price** among the brands shown.
-

3. Interactive Filters (Right Panel)

Users can **filter the data dynamically** based on multiple parameters:

- **Laptop Brand** – Select a specific laptop brand.
 - **Operating System** – Choose from **DOS, Mac, Windows**.
 - **Laptop Type** – Filter by **Casual, Gaming, Thin & Light**.
 - **Processor Brand** – Select between **AMD, Intel, and M1**.
 - **Processor Name & Generation** – Further refine based on specific processors.
 - **System Architecture** – Filter between **32-bit and 64-bit** laptops.
-

4. Business Use Cases

- ✓ **Price Analysis** – Identify trends in pricing based on brand, RAM, and other factors.
- ✓ **Brand Performance** – Evaluate which brands dominate the market and their pricing strategies.
- ✓ **Customer Preference Insights** – Understand how price correlates with laptop specifications.
- ✓ **Decision Support for Buyers** – Helps businesses and consumers choose laptops based on budget and specifications.

Conclusion

Data warehousing enables businesses and analysts to make **data-driven decisions** by consolidating and structuring large datasets for efficient querying and analysis. This laptop price dataset, structured in a **Star Schema**, allows for deeper insights into laptop pricing trends, brand performance, and market segmentation.

Key Decision-Making Areas

Pricing Strategy Optimization

- Identify the **average price range** of different laptop brands and types.
- Adjust pricing based on market trends and competitors' pricing strategies.

Market Segmentation & Customer Targeting

- Segment laptops based on **brand, RAM, processor, operating system, and type** to target specific customer groups.
- Develop marketing strategies based on customer preferences for gaming, casual, or high-end laptops.

Product Portfolio Management

- Recognize which laptop brands and specifications are most **popular** in the market.
- Make informed decisions about **stocking inventory** based on demand and pricing trends.

Sales Performance & Brand Analysis

- Evaluate **which brands dominate the market** and have the highest average price.
- Identify **gaps in the market** where new brands or product types could be introduced.

Data-Driven Business Intelligence

- The integration of a **Fact Table with Dimension Tables** ensures efficient data retrieval for advanced analysis.
- Users can leverage **Power BI dashboards** for **interactive and real-time decision-making**.

By utilizing **data warehousing techniques**, businesses can enhance their **strategic decision-making, market competitiveness, and operational efficiency** in the laptop industry.