**Laptop Dataset Analysis**

**SAKTHI ARAVIND   
DA&DS**

**15-04-2025**

**Objective:**

1. To perform exploratory data analysis (EDA) on a dataset of laptops and uncover insights about pricing, specifications, and consumer trends.
2. The goal is to help manufacturers and marketers understand which features impact pricing and how they can optimize their strategies.

**Aim:**

**​**The aim of this project is to perform a comprehensive Exploratory Data Analysis (EDA) on a laptop dataset to uncover patterns, relationships, and key factors influencing laptop pricing and consumer preferences. By employing statistical techniques and data visualization methods, the analysis seeks to understand the distribution of individual features such as screen size, RAM, and weight, as well as their interdependencies. This process will facilitate the identification of trends and anomalies, providing valuable insights for stakeholders to make informed decisions regarding product development, marketing strategies, and pricing models in the competitive laptop market.

**Business Problem:**

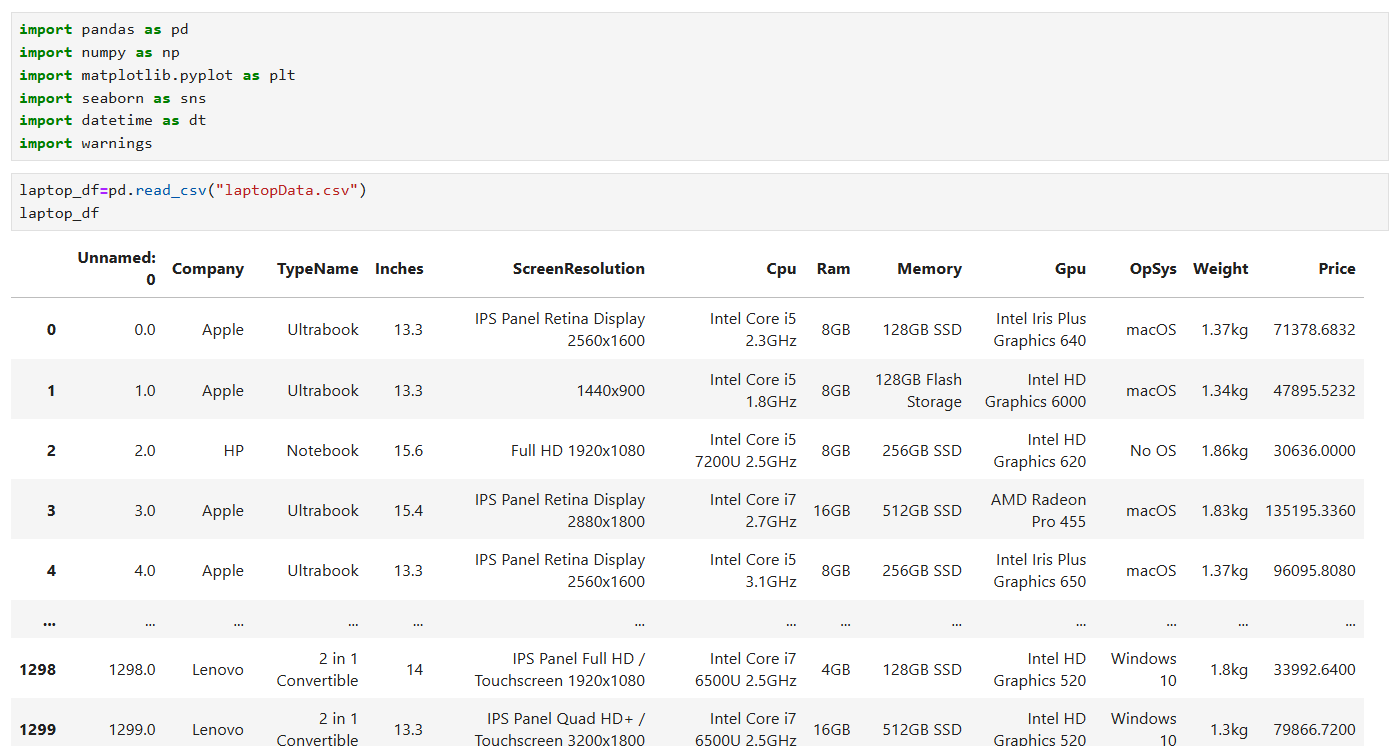
**​**The laptop market is experiencing rapid growth, driven by technological advancements, evolving consumer preferences, and increased demand for portability and performance. Consumers face challenges in selecting laptops that balance specifications, brand reputation, and budget constraints. Manufacturers and retailers need insights into factors influencing pricing and consumer choices to remain competitive. Therefore, this project aims to analyze a comprehensive laptop dataset to identify key determinants of pricing and consumer preferences. By employing exploratory data analysis and visualization techniques, the study seeks to uncover patterns and correlations among variables such as RAM, CPU, screen size, weight, and brand, providing actionable insights for stakeholders in the laptop industry.

**Dataset Description:**

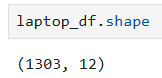
The dataset contains specifications and pricing details of 1273 laptops from multiple brands like Apple, HP, Dell, and Lenovo**.**

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **Company** | Brand name of the laptop |
| **TypeName** | Category/type of laptop (e.g., Ultrabook, Notebook) |
| **Inches** | Screen size in inches |
| **Screen Resolution** | |  | | --- | |  |  |  | | --- | | Display resolution and panel type | |
| **Cpu** | Processor model and frequency |
| **Ram** | RAM size (e.g., 8GB, 16GB) |
| **Memory** | Storage configuration (e.g., SSD, HDD, hybrid) |
| **Gpu** | |  | | --- | |  |  |  | | --- | | Graphics processing unit | |
| **OpSys** | Operating System (e.g., Windows, macOS) |
| **Weight** | Laptop weight in kilograms |
| **Price** | Selling price of the laptop (target variable) |

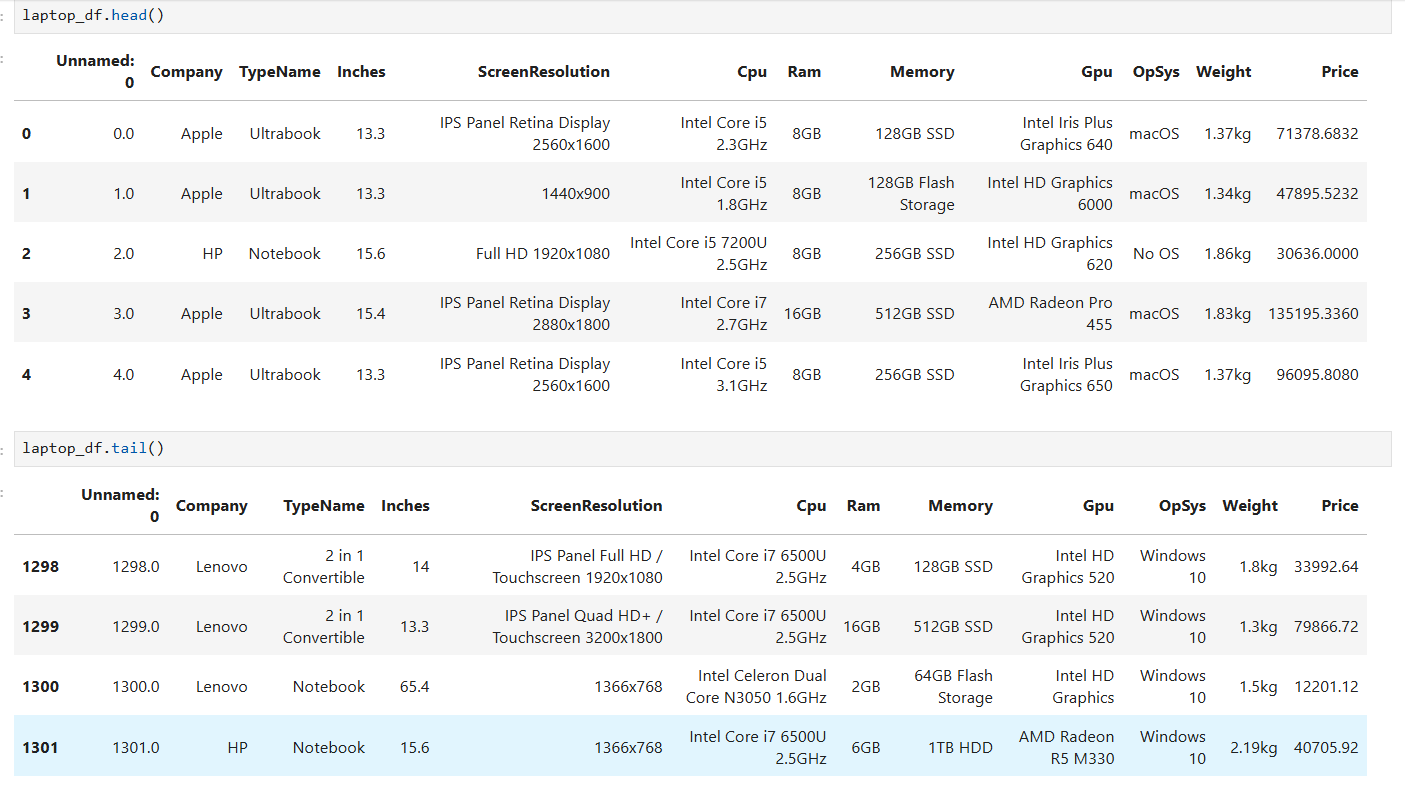
**Data Overview & Initial Exploration:**

****

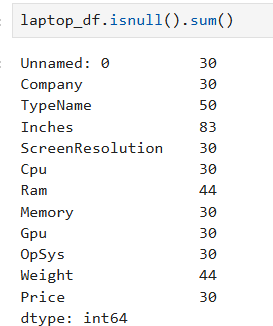
**Dataset Shape and Columns:**

****

**Initial Data Preview :**

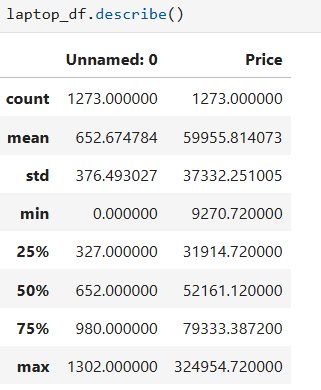
****

**Missing Values Check:**

****

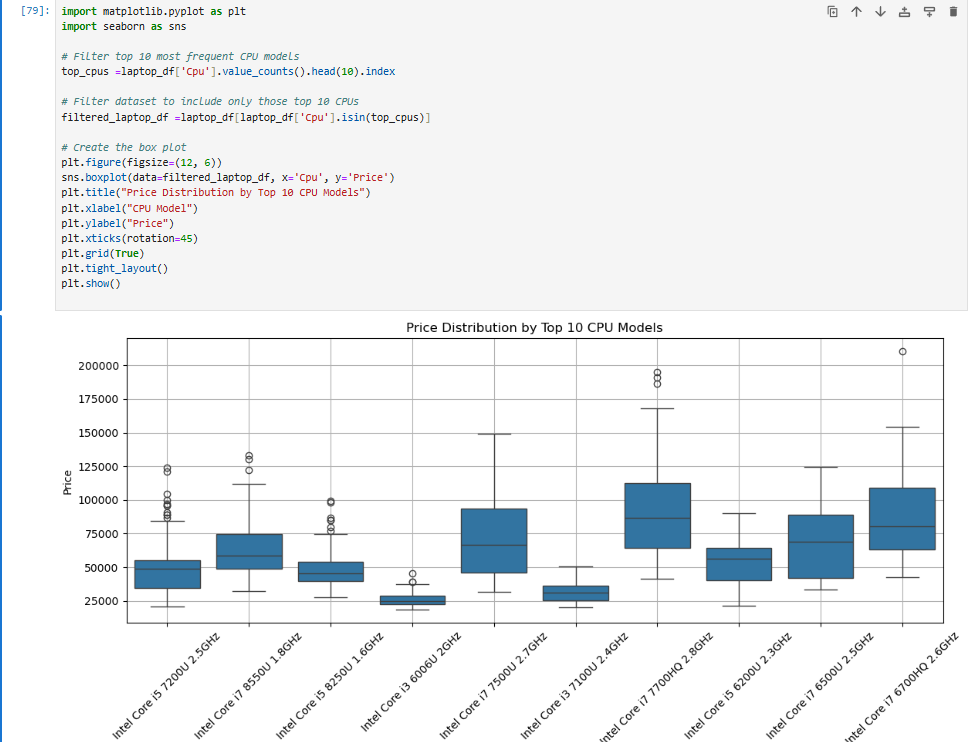
**Data Cleaning & Preparation:**

1. **Dropped** Unnamed: 0 column as it was just an index**.**
2. **Removed rows with missing or invalid entries** in key fields like Inches, Ram, Weight, Price**.**
3. **Transformed** Ram from '8GB' to numeric 8 (integer)**.**
4. **Cleaned** Weight by removing units like 'kg'/'Kg' and converting to float**.**
5. **Ensured** Inches **and** Price are stored as proper floats**.**

****

**Univariate Analysis:**

**Price Distribution by Top 10 CPU Modals:**

****

**Highest Priced CPU Model:**

* Intel Core i7 7700HQ 2.8GHz has the highest overall price range.
* This CPU shows both a high median price and many upper outliers, indicating premium-priced laptops.

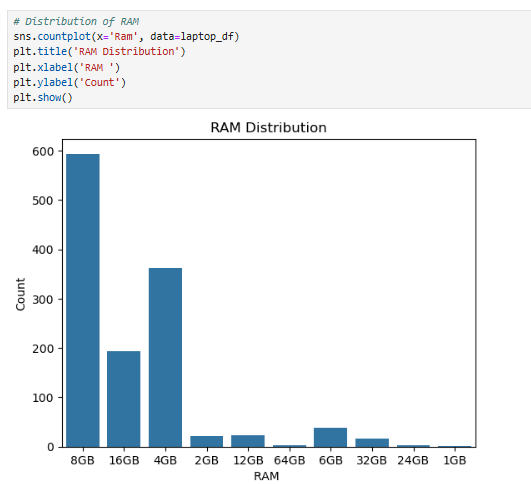
**Widest Price Spread:**

* Intel Core i7 6700HQ 2.6GHz also shows a wide interquartile range and high outliers, making it one of the more expensive models overall.

**Consistently Expensive Models:**

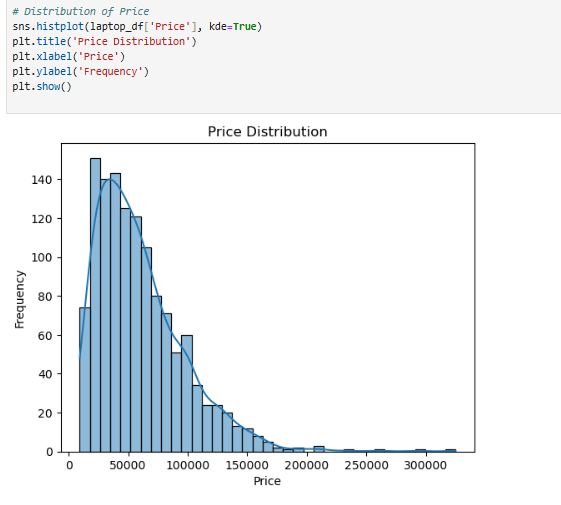
* Intel Core i7 7500U 2.7GHz and Intel Core i7 7700HQ 2.8GHz show higher medians compared to others, suggesting they are generally found in higher-priced laptops.

**RAM Distribution:**

****

* 8GB RAM is the most common configuration.
* 4GB and 16GB are also popular, while 32GB+ are rare and found mostly in gaming/premium devices.

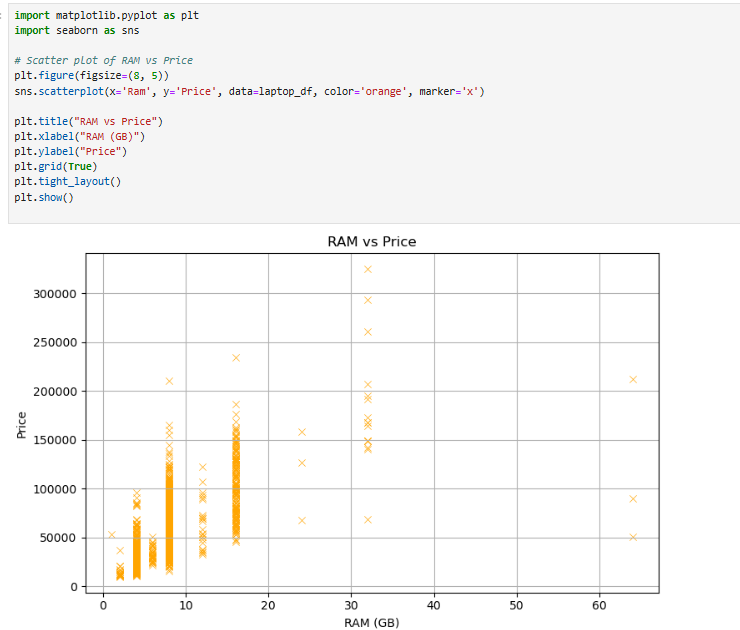
**Price Distribution:**



* Most laptops are priced between ₹30,000 to ₹80,000.
* The distribution is slightly right-skewed, with a few high-end models above ₹1,00,000.

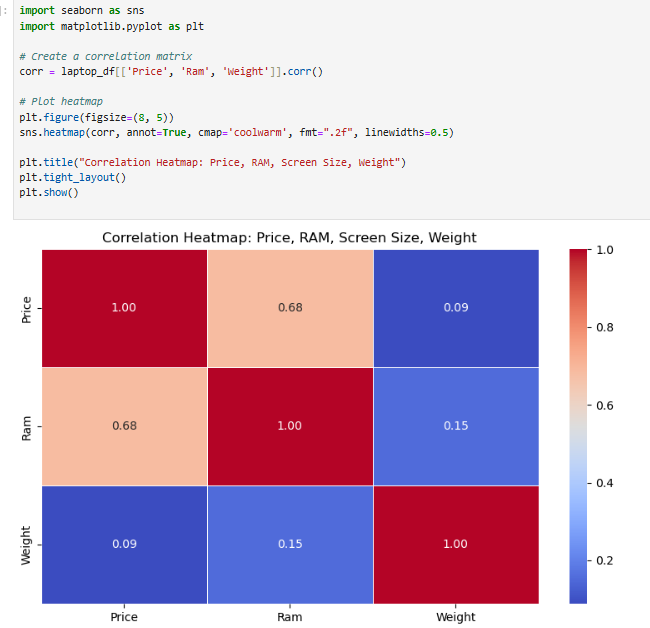
**Bivariate Analysis:**

**Price vs RAM:**

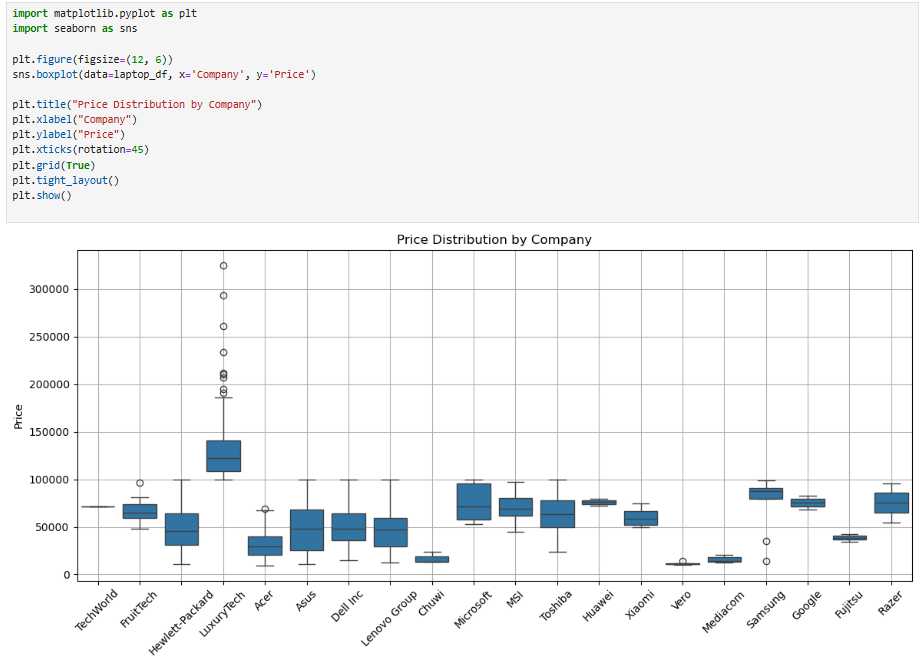
****

* Laptops with **higher RAM** are generally more expensive.
* **8GB and 16GB** configurations show wide price variation, indicating impact of other factors like CPU and storage type.

**Screen Size vs Weight:**

****

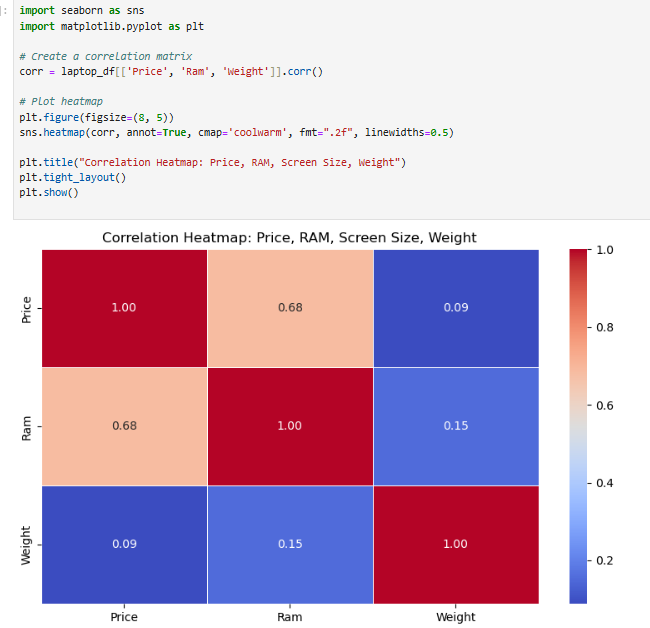
* There's **a positive correlation**: larger screens generally mean heavier laptops.
* A few outliers show that **premium ultrabooks** manage to stay lightweight even with larger displays.

**Company vs Price:** ****

* **Apple, Dell, and Microsoft** dominate the **high-end pricing tier.**
* **HP, Lenovo, and Acer** offer a wider price range with many budget models.
* **Asus** shows balanced pricing across segments**.**

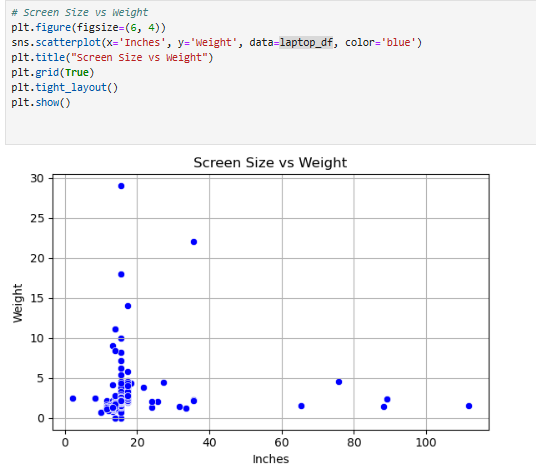
**Multivariate Analysis:**

**Correlation Heatmap:**

****

* Price has the strongest positive correlation with RAM and moderate correlation with Screen Size (Inches).
* Weight and Inches also show a moderate positive correlation, which is expected — bigger screens tend to weigh more.
* Most variables have low multicollinearity, which is good for further analysis or modeling.

**Pair plot: Interactions Between Variables:**

****

* **RAM and Price** show noticeable upward trends.
* **Inches and Weight** demonstrate a linear pattern.
* Useful for observing clusters or outliers in the dataset.

**Feature Engineering:**

Extracted the processor tier (e.g., i3, i5, i7, Ryzen 5) from the CPU string.

|  |  |
| --- | --- |
| **Sample CPU** | **Extracted (CPU Types)** |
| Intel Core i5 2.3GHz | i5 |
| Intel Core i7 2.7GHz | i7 |
| AMD Ryzen 5 2500U | Ryzen 5 |
| Intel Celeron Dual Core N4000 | Other |

**Total Storage:**

Parsed hybrid memory configurations (like 256GB SSD + 1TB HDD) and converted them into total numeric storage in GB.

|  |  |
| --- | --- |
| Memory | Total Storage (GB) |
| 128GB SSD | 128 |
| 512GB SSD | 512 |
| 1TB HDD + 256GB SSD | 1256 |

**Brand Tier:**

Classified laptop brands into **High-end**, **Mid-range**, and **Budget**, based on average pricing:

|  |  |
| --- | --- |
| **Company** | **Brand Tier** |
| Apple | High-end |
| HP | Mid-range |
| Acer | Budget |
| Dell | High-end |
| Asus | Mid-range |

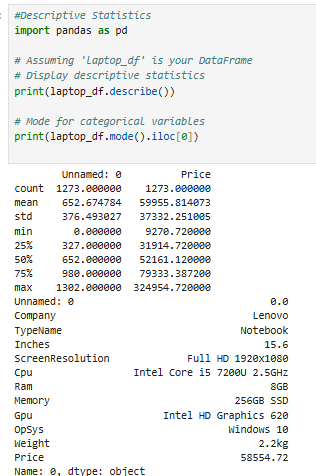
**Statistical Analysis:**

**Descriptive Statistics**

Descriptive statistics provide a summary of your dataset's central tendency, dispersion, and shape.

* **Mean (Average)**: Provides the average value for features like Price, RAM, Screen Size, and Weight.​
* **Median**: Identifies the middle value, offering insight into the distribution's symmetry.​
* **Mode**: Highlights the most frequently occurring value, useful for categorical features.​
* **Standard Deviation**: Measures the amount of variation or dispersion from the mean.​
* **Minimum and Maximum**: Indicate the range of values present in the dataset.​

These statistics help in understanding the general distribution and variability of the data, which is crucial for further analysis and modelling.​



**Test Statistics and Hypothesis Testing**

To explore relationships and differences within the data, we conducted the following statistical tests:​

* **Independent Two-Sample t-Test**: Assessed whether there is a statistically significant difference in the average price between two groups, such as laptops with different RAM sizes. The test assumes normal distribution and equal variances.​
* **Chi-Square Test of Independence**: Evaluated the association between two categorical variables, for example, Brand and Operating System. This test determines whether distributions of categorical variables differ from each other.



**Dimensionality Reduction (PCA):**

To simplify the dataset by reducing it to two principal components that capture the maximum variance, enabling better visualization and analysis.

**Process:**

* Selected numeric features: Price, Ram, Inches, Weight, Total\_Storage
* Standardized the data using StandardScaler
* Applied Principal Component Analysis (PCA) to reduce dimensionality to 2 components.

**PCA Result:**

* PCA1 and PCA2 capture the major variance in specifications across laptops.
* High-end brands like Apple and Dell are clustered distinctly from budget brands, indicating unique feature combinations.
* Overlapping clusters show competition in the mid-range segment (e.g., Asus, HP).

**Insights & Interpretation:**

**Based on our analysis and recent market trends, several factors significantly impact laptop pricing:**

* Specifications: Higher RAM, advanced CPUs (e.g., Intel i7, Ryzen 7), and larger SSD storage correlate with increased prices.
* Brand Reputation: Premium brands like Apple and Dell often command higher prices due to perceived quality and brand value.
* Design & Portability: Lightweight and slim designs, especially in ultrabooks, add to the cost.
* Technological Advancements: Integration of AI features and high-performance GPUs (e.g., NVIDIA RTX 50 series) contribute to higher pricing.

**Emerging Market Trends (2025)**

* AI Integration: AI-enabled PCs are projected to constitute 44% of notebook shipments in 2025, temporarily boosting average selling prices .
* Windows 10 End-of-Life: The end of support for Windows 10 is driving a significant upgrade cycle, leading to increased demand for new laptops .
* Tariff Impacts: U.S. tariffs on Chinese imports have led to price increases and supply chain adjustments, affecting global laptop pricing .

**Recommendations**

* **For Manufacturers:**
  + Focus on integrating AI capabilities to meet emerging consumer demands.
  + Consider diversifying manufacturing locations to mitigate tariff impacts**.**
* **For Marketers:**
  + Highlight performance and portability in marketing campaigns.
  + Leverage brand reputation to justify premium pricing.
* **For Consumers:**
  + Evaluate specifications relative to personal needs to ensure cost-effectiveness.
  + Consider emerging brands offering competitive features at lower prices.

**Summary of Findings**

Our comprehensive analysis of the laptop dataset has revealed several key insights:​

* **Specifications Influence Pricing**: Higher RAM, advanced CPUs (e.g., Intel i7, Ryzen 7), and larger SSD storage correlate with increased laptop prices.​
* **Brand Reputation Matters**: Premium brands like Apple and Dell command higher prices due to perceived quality and brand value.​
* **Design & Portability**: Lightweight and slim designs, especially in ultrabooks, add to the cost.​
* **Technological Advancements**: Integration of AI features and high-performance GPUs (e.g., NVIDIA RTX 50 series) contribute to higher pricing.​