**Project Information**

Title : Exploratory Data Analysis on Laptop Sales

Dataset

Name : Meganathan A

DA/DS : May-2025

Batch Number : RP-36

Online/Offline : Offline

Roll Number : B2025057352

**Table of Contents**

1. **Introduction**

There are many laptop models with different features. This makes it hard for people to choose the right one. In this project, we study a dataset that has details like brand, processor, RAM , storage, and operating system. We clean the data and use simple data analysis to understand which features affect the laptop’s price and performance.

1. **Aim**

We want to understand what laptop Dataset. first we clean the data. Then we use charts and simple way. This helps us see how things like RAM, brand and size change the price. its to help people choose the right laptop by comparing features and prices. In the future, this work can also help to predicts laptop prices.

1. **Business Problem / Problem Statement**

Each laptop differs in brand, processor, RAM, storage type, display size, operating system, and other specifications, making direct comparisons complex. Businesses, too, need reliable insights to price their products competitively

1. **Project Workflow**

It starts with loading the dataset and understanding its structure. The next step is data cleaning, which includes handling missing values, treating outliers, and resolving inconsistencies. Once the data is clean, we generate new metrics and filter the dataset for analysis. This is followed by descriptive statistical analysis and hypothesis testing to verify assumptions. We then perform Exploratory Data Analysis (EDA), including univariate, bivariate, and multivariate analysis using visualizations

1. **Data Understanding**

The dataset has details about laptops like brand, type, RAM, screen size, weight, and price. We used Python (pandas) to load the data and found some missing values and an extra column, which we removed. Some data like RAM and Inches were in the wrong format, so we fixed them. This helped us get the data ready for analysis.

1. **Data Cleaning Missing Values Imputation**

* + Missing Values Imputation

Missing values were found in fields like Inches, TypeName, and Ram. These were filled using forward-fill (ffill) and backward-fill (bfill) methods. For numeric fields like Inches, the mean value was calculated and used to replace missing entries.

* + Outlier Treatment

Extreme values in numeric fields like Price or Weight can distort analysis. Visual inspection using box plots or value counts helped identify and understand outliers. Based on domain knowledge and visual analysis, extreme or unrealistic entries were noted, though not all were removed if they were valid.

* + Handling Inconsistent Values

The "Ram" column had values like "8GB" and "16GB". We split it into two parts: one for the number and one for the unit. We also removed empty rows and fixed data formats. This made the data clean and easy to use.

1. **Obtaining Derived Metrics**

To make the analysis better, we created new columns from existing ones. For example, we split the "Ram" column into number and unit. We also cleaned "Inches" and "Weight" to use them in comparisons. Changing text to numbers helped us make charts and do math easily. This made the data ready for deeper study.

1. **Filtering Data for Analysis**

To keep the data clean and useful, we removed rows with missing or wrong values. We also dropped empty entries using simple filters. Some laptops with very odd or repeated details were removed too. This helped us keep only real and useful laptop data for better analysis

Eg: dropna()

1. **Statistical Analysis** 
   * Descriptive analysis

Descriptive statistics such as mean, median, standard deviation, and range were calculated for numeric variables like Price, Inches, and Ram\_Number. The describe() function showed that price distribution was skewed, and RAM had common values like 4GB, 8GB, and 16GB.

* + Test statistics and hypothesis testing

Although not deeply covered in the notebook, statistical tests like t-tests or ANOVA could be used to examine whether price differences between brands or types (e.g., Gaming vs Notebook) are statistically significant. For example, we could test if the average price of “Gaming” laptops differs significantly from “Notebook” types. This approach helps validate observed patterns and provides confidence in decision-making based on data.

1. **Exploratory Data Analysis (EDA) - Univariate Analysis**

Univariate analysis involves analyse one variable at a time. Histograms, bar plots, and count plots were used to understand the distribution of individual features. For instance:

* Price: A histogram showed that most laptops are priced between ₹30,000 to ₹80,000, with fewer high-end models exceeding ₹100,000.
* TypeName: Count plots revealed that "Notebook" and "Ultrabook" are the most common types, while “2 in 1 Convertible” is less frequent.
* Ram Number: Bar plots showed the popularity of 4GB, 8GB, and 16GB RAM.

1. **Bivariate Analysis**

Bivariate analysis explores the relationship between two variables. In this project, several bivariate relationships were visualized using scatter plots, box plots, and grouped bar plots.

* Price vs. Ram\_Number: A clear trend was observed — laptops with higher RAM typically have higher prices.
* Price vs. TypeName: Gaming laptops had the highest prices on average, followed by Workstation and Ultrabook types.
* Price vs. Inches: As screen size increased, there was a general increase in price, although with more variation**.**

1. **Multivariate Analysis**

Multivariate analysis examines interactions between three or more variables simultaneously. Techniques like pair plots and heatmaps were used to understand complex relationships.

* Pair Plot: Helped visualize multiple combinations such as Price vs. Ram\_Number vs. Inches.
* Heatmap: Showed correlation coefficients between numerical variables. A moderate correlation was seen between RAM and Price, and also between Inches and Price.
* Group-wise comparisons: Using grouped bar plots, it was observed that brands like Apple and MSI had higher price ranges even with similar specs compared to other brands, indicating brand influence.

1. **Overall Insights from Analysis**

* **Price Range**: Most laptops cost a medium amount. Very costly ones are fewer and usually have special features or top brands.
* **Laptop Type**: Gaming laptops are the most expensive. Notebooks and Ultrabooks are cheaper and more common.
* **RAM and Screen Size**: More RAM and bigger screens usually mean higher prices. But after 16GB RAM or 15.6 inches, the price doesn’t go up much.
* **Brand Effect**: Brands like Apple and MSI cost more, even with the same features as others.
* **Storage and Graphics**: SSDs and dedicated GPUs also affect the price, though we didn’t explore them deeply here.
* **Clean Data**: Cleaning the data helped us get clear and useful results.

1. **Future engineering**

**Key Findings**:

* Laptops with **2GB and 4GB RAM** show **higher sales** compared to other configurations.
* These models are likely **affordable** and attract **budget-conscious customers**.

**Boosting Sales Strategy**:

* Offer **complementary items** like:
  + - Free **keyboard and mouse**
    - Optional upgrade to **SSD storage**
* These extras **increase perceived value**, encouraging more purchases without raising the laptop price.

1. **Conclusion**

This project studied laptop data to find out what affects the price. We cleaned the data and looked at different features like RAM, screen size, type, and brand. These things were found to be important for pricing. In the future, this work can help build tools to predict laptop prices or give buying suggestions. Overall, the project shows how data analysis can help both customers and companies make better choices.