Laptop Dataset Analysis

Objectives:

The objective of this project is to analyze a laptop dataset by first performing data cleaning and preprocessing to ensure accuracy. The analysis involves exploring the structure and distribution of key features such as price, RAM, weight, and CPU. Descriptive statistics and visualizations are used to identify trends, patterns, and outliers. Hypothesis testing is applied to understand relationships between variables like price and specifications. The ultimate goal is to extract meaningful insights that reflect market dynamics and consumer preferences.

Dataset Overview:

* Company: The manufacturer or brand of the laptop.
* Type Name: The type of the laptop (Gaming, Ultrabook, Notebook).
* Inches: The size of the laptop screen in inches.
* Screen Resolution: The resolution and display technology of the laptop screen.
* CPU: The processor model of the laptop.
* RAM: The amount of Random Access Memory in the laptop.
* Memory: The storage capacity (hard drive or SSD) of the laptop.
* GPU: The graphics processor (GPU) model of the laptop.
* OpSys: The operating system installed on the laptop.
* Weight: The weight of the laptop in kilograms.
* Price: The price of the laptop in the local currency.

Tools Used:

* Python

Steps Involved:

1. Data Cleaning and Preprocessing

* Handled Outliers

Outliers were detected using boxplots and the IQR method to identify extreme values. These outliers were removed to maintain data accuracy and improve analysis reliability.

* Handled null values

Null values were identified and handled by replacing them with appropriate measures such as mean, mode and median, depending on the data type and distribution.

* Removed Duplicates

Duplicate records were identified and removed to ensure data accuracy and prevent skewed analysis results.

* Identified Invalid Values

Invalid values such as placeholders ('?', 'NA') were identified and replaced or removed to maintain data consistency.

* Standardized Values

Standardized inconsistent formats across columns (e.g., 'GB' in RAM, 'kg' in Weight) to ensure uniform data types for analysis.

2.Feature Engineering

* Cleaned and standardized categorical values such as OpSys and Company to ensure consistency (e.g., grouped similar OS types like 'Windows 10' and 'Windows 11' under 'Windows').
* **Assigned a reputation score** to each brand based on their average laptop price, reflecting perceived market positioning (e.g., premium vs budget brands).
* Engineered features by cleaning and transforming data, handling missing values, and encoding categories. Assigned brand reputation scores based on average prices. Reduced dimensions by removing redundant features through correlation analysis for better interpretability.

3. Hypothesis Testing

* T - Test

Conducted to evaluate the impact of RAM affects prices.The results indicate that RAM significantly affect the price.

* One-Way ANOVA

Conducted to evaluate the impact of Operating System affects prices.The results indicate that Operating System significantly affects the price.

* One-Way ANOVA

Performed to test if Company (Brand) influence price variations. Findings show that Brand significant impact on laptop prices.

4. Exploratory Data Analysis

3.1. Univariate Analysis

3.1.1. Laptop Price Distribution

* Performed univariate analysis to understand the distribution of individual features.
* Used a histogram (histplot) to visualize the distribution of laptop prices, identifying price concentration ranges and detecting outliers.

4.1.2. RAM Distribution

 Analyzed the RAM (in GB) using a histplot to observe the frequency of different RAM sizes across laptops.

4.1.3. Laptop Weight Distribution

* Explored the Weight (in kg) of laptops using a boxplot to identify the spread and detect outliers.

4.1.4. Laptop Brand Distribution

* Analyzed the distribution of laptop brands using a box plot, highlighting the most common brands in the dataset like Dell, Asus, and Lenovo.

4.2. Bivariate Analysis

4.2.1. Price vs RAM

* Analyzed laptop price versus RAM using a box plot, which showed that 16 GB models are priced higher, 4 GB models are on the lower end, and 8 GB models vary widely due to other specifications.

4.2.2. Price vs Company

* The bivariate analysis of Price vs Company explores the relationship between laptop prices and their respective brands. This analysis provides insights into pricing patterns across different companies.

4.2.3. Price vs CPU

* The bivariate analysis of Price vs CPU explores the relationship between laptop prices and CPU types. It highlights how different CPUs impact laptop pricing.

4.3. Multivariate Analysis

4.3.1. Correlation Heatmap

* The multivariate analysis with a correlation heatmap shows the relationships between multiple variables. It highlights how factors like price, CPU, and RAM are interrelated.

4.3.2. Pairplot

* Generated a pairplot to visualize relationships between numerical features like Price, RAM, and Storage. It showed that higher RAM and OS generally correspond to higher laptop prices.

5.Insights

* Majority of laptop prices fall between ₹25,000 and ₹60,000, with fewer models priced above ₹1,00,000.
* Most laptops come with 8 GB RAM, followed by 4 GB and 16 GB configurations.
* Common laptop weights range from 1.8 kg to 2.2 kg, while heavier models (2.8 kg – 3.5 kg) are less frequent.
* Brands like Apple, Asus, Dell, HP, Lenovo, Acer, and Microsoft have the highest number of laptops in the dataset.
* Ultrabooks, Notebooks, and Gaming laptops are the most common types observed.
* Laptop price increases with higher RAM size and CPU type, with i7 models priced higher than i3.
* Devices with macOS and Windows tend to be more expensive compared to those with Android or Chrome OS.
* Premium brands like Razer, LG, Microsoft, and Apple are associated with higher laptop prices.
* Most laptops in the dataset run on Windows operating system.
* The most common screen size among laptops is 15.5 inches.

6.Recommendations

* Launch budget-friendly laptops with strong specifications to attract a wider customer base.
* Focus on reducing laptop weight for better portability and user convenience.
* Consider lowering prices on Apple laptops to increase macOS adoption.
* Enhance operating systems to deliver better performance and hardware optimization.
* Invest in improving battery life for longer and more reliable usage.

7.Conclusion

* Conducted detailed Exploratory Data Analysis (EDA) to explore laptop features and pricing.
* Visualized data using plots to identify patterns, relationships, and trends.
* Derived actionable insights to understand laptop market dynamics.
* Identified key price-influencing factors like RAM, brand, OS, and processor type.
* Provided strategic recommendations based on observed data-driven findings.
* Highlighted the dominance of mid-range laptops (₹25k–₹60k) in the market, with premium models being less common.
* Confirmed that 8 GB RAM and Windows OS are the most popular configurations among users.
* The analysis offers valuable insights for buyers, sellers, and manufacturers to make informed decisions.