**Exploratory Data Analysis using Python**

**Project Information:**

**Title:** Laptop Analysis using Statistical & EDA Techniques  
**Name:** SURYA  
**DA/DS:** Data Analysis and Data Science  
**Batch:** November Batch  
**Online/Offline:** Offline  
**Roll Number:** 61024CBR007

# **1. Introduction**

In the personal computing market, understanding the key factors that influence laptop prices is essential for manufacturers, consumers, and retailers. This project examines a laptop dataset using statistical analysis and exploratory data techniques with Python. The dataset contains detailed information on laptop specifications including brand, CPU, RAM, screen size, storage type, weight, and operating system. Through this analysis, the goal is to identify the patterns and relationships that drive price variation, ultimately supporting data-informed decisions in product development and marketing strategies.

**2. Aim**

The primary objective of this project is to perform a thorough exploratory and statistical analysis of a laptop dataset to determine how different features affect laptop pricing. By employing data preprocessing, feature engineering, visualization, and hypothesis testing, the project aims to uncover significant trends and dependencies. These findings can be utilized to optimize pricing strategies, guide customer recommendations, and improve product feature planning in the competitive electronics market.

**3. Business Problem / Problem Statement**

Accurately pricing laptops based on specifications is a complex challenge due to the wide variety of models and feature combinations. Mispricing can lead to reduced sales or missed profit opportunities. Manufacturers and sellers require a data-driven approach to understand which factors—such as SSD presence, RAM size, processor brand, or screen resolution—have a significant impact on pricing. This project addresses the business problem by statistically validating the contribution of each feature to the price, enabling more effective product positioning and competitive pricing.

**4. Project Workflow**

The project follows a step-by-step process to ensure a robust and reproducible analysis. It starts with data acquisition and inspection, followed by cleaning to handle missing, inconsistent, or outlier values. Next, feature engineering is applied to extract meaningful information from complex fields like CPU and GPU. Statistical analyses, including t-tests and descriptive statistics, are used to validate assumptions and assess significance. Exploratory Data Analysis (EDA) is then conducted using visualizations such as scatter plots, histograms, and heatmaps to uncover underlying trends. The project concludes by deriving insights and suggesting actionable strategies for stakeholders in the laptop market.

**5. Data Understanding**

The dataset contains information about different laptop models and their prices. It includes numbers like screen size, RAM, weight, and price as well as categories like brand, processor, operating system, and storage type. At first look, laptop prices varied a lot depending on their features and brand. Basic statistics gave a clear idea of how prices and other values are spread out. The data types and formats were also checked to make sure the data could be used correctly in the analysis. This early understanding helped guide the steps for cleaning and analysing the data.

**6. Data Cleaning**

Data cleaning was an essential step to prepare the dataset for analysis.

1. Missing Values: Minimal missing values were handled by removing rows or filling with appropriate values (like the most frequent category for mode or the median for numerical ones).
2. Outlier Detection: Outliers in variables like Price, Weight, and Inches were detected using boxplots. Extreme outliers were removed to avoid skewed analysis.
3. Inconsistent Values: Inconsistent formats in Weight, Memory, and Inches were standardized using regular expressions and string manipulation, ensuring uniformity in the dataset.

This data cleaning approach helps maintain a high-quality, reliable dataset for analysis and decision-making.

**7. Obtaining Derived Metrics**

To improve analysis, new features were created from existing data:

1. CPU \_Brand and CPU \_ Series were extracted from the CPU column for better classification.
2. CPU\_ GHz was extracted to get the processor speed as a number.
3. GPU\_ Brand and GPU\_ Series were extracted from the GPU column to separate brand and model.

These new features made it easier to group laptops and compare them based on clear categories.

**8. Filtering Data for Analysis**

To prepare the laptop dataset for analysis, filtering was used to focus on relevant and clean data. Laptops with uncommon operating systems were removed to keep only popular ones like Windows, macOS, and Linux. Storage types were separated into SSD, HDD, or hybrid to allow clear comparisons. CPUs and GPUs from less common brands were excluded to focus on well-known brands like Intel, AMD, and NVIDIA. These steps helped create clear and meaningful groups for accurate analysis.

**9. Statistical Analysis**

This step validated assumptions using both descriptive and inferential statistics:

* **Descriptive Analysis:** Mean, median, mode, standard deviation, and range were computed for key numerical variables. Price distributions varied significantly across categories.
* **Hypothesis Testing:**

1.T-tests were used to compare mean prices between two independent groups:

* **Weight vs Price**: Weight significantly difference in price.
* **Price vs RAM:** Price differs significantly based on RAM size

2.Chi-Square tests are used to assess whether there is a relationship between categorical variables.

* **Operating System vs TypeName:** Variables are dependent. Indicating that the choice of OS depends on the laptop type.

**10.Exploratory Data Analysis (EDA) – Univariate Analysis**

Univariate analysis helped in understanding the distribution of each variable:

* Price: Laptop prices range widely, with most falling between ₹25,000 and ₹70,000, and premium models exceeding ₹3,00,000.
* RAM: The proportion of Ram with 8GB is nearly half of the total.
* Screen size: The most popular screen size is 15.6 inches, followed by 14.0 inches
* Company: The top 3 companies are Lenovo, Dell, HP, they have similar proportions.

Visualizations such as Histograms, boxplots, and bar charts were used to clearly present and summarize the data.

**11.Bivariate Analysis:**

This step analysed the relationship between price and other variables:

* **Price vs. RAM**: Positive trend-higher RAM generally correlated with higher prices.
* **Screen Size vs Weight:** Larger screen sizes don’t always correlate with heavier laptops, as some larger screens are lightweight.
* **Price Distribution by Operating System:** Laptops with premium OS (like Windows 10 or macOS) vary in price, while those with Chrome OS or no OS are more affordable.
* **Price vs CPU:**Intel has the highest average price, followed by Samsung.
* **Price vs GPU**:Nvidia-based laptops are the most expensive, likely due to their use in gaming and high-performance devices.

**12. Multivariate Analysis**

Multivariate analysis considered the combined effect of multiple features:

* A correlation heatmap also revealed strong linear relationships among numerical variables.
* A pair plot was used to analyse the relationships between price, RAM, weight, and CPU GHz, revealing how these features impact each other and laptop pricing.

**13. Overall Insights from Analysis**

* Most laptops have 8GB of RAM and 256–512GB SSD storage, catering to general user needs.
* Mid-range laptops dominate, with a few high-end models priced over ₹3,00,000.
* HP, Dell, and Lenovo have the most models, offering a variety of options.
* Better specs (more RAM, stronger GPU, faster CPU) generally mean higher prices.
* Razer and Apple laptops are among the most expensive.
* Gaming laptops are heavier and cost more due to powerful components.
* 15.6-inch screens are the most common, offering a balanced display size for most users.

**14. Conclusion**

The laptop analysis project effectively identified key factors influencing laptop pricing. Key variables like SSD presence, RAM size, processor brand, and were found to have a significant impact on price. These insights can guide more accurate pricing strategies and improve competitive positioning in the market. Future work can build predictive models to forecast laptop prices based on specifications.