**EXPLORATORY DATA ANALYSIS USING PYTHON**

**Title:** Laptop Analysis using Statistical & EDA Techniques

**Name:** DAYANA J A

**Domain:** Data Analyst / Data Science

**Batch:** November Batch

**Mode:** Offline

**Roll Number:** 61024CBR016

**1. Introduction**

In the competitive personal computing market, understanding the factors influencing laptop prices is crucial for manufacturers, retailers, and consumers. This project delves into a comprehensive dataset of laptops, analyzing specifications such as brand, CPU, RAM, screen size, storage type, weight, and operating system. Utilizing Python's statistical and exploratory data analysis (EDA) techniques, the aim is to uncover patterns and relationships that drive price variations, aiding in informed decision-making for product development and marketing strategies.

**2. Aim**

The primary objective is to perform an in-depth exploratory and statistical analysis of a laptop dataset to determine how various features affect pricing. Through data preprocessing, feature engineering, visualization, and hypothesis testing, the project seeks to identify significant trends and dependencies. These insights can inform pricing strategies, guide customer recommendations, and enhance product feature planning in the electronics market.

**3. Problem Statement**

Accurately pricing laptops is challenging due to the multitude of models and features available. Mispricing can negatively impact sales and profitability. This project employs data analysis to identify which features—such as SSD presence, RAM size, processor brand, and screen resolution—significantly influence laptop pricing. By statistically validating these factors, the project aims to assist manufacturers and sellers in setting competitive and profitable prices.

**4. Project Workflow**

* **Data Loading**: Import the dataset using pandas.
* **Data Cleaning**: Handle missing values, correct errors, and remove outliers.
* **Feature Engineering**: Create new features from existing data for better analysis.
* **Statistical Analysis**: Apply descriptive and inferential statistics to validate assumptions.
* **Visualization**: Use plots and charts to explore data patterns.
* **Insights & Recommendations**: Draw conclusions and suggest practical applications for stakeholders.

**5. Data Understanding**

The dataset comprises various laptop models with attributes including:

* **Numerical Features**: Screen Size (inches), RAM (GB), Weight (kg), Price (₹)
* **Categorical Features**: Brand, Processor (CPU), Operating System (OS), Storage Type

Initial observations indicate significant price variations based on features and brand. Basic statistical measures provided insights into the distribution and spread of values, guiding subsequent data cleaning and analysis steps.

**6. Data Cleaning**

* **Missing Values**: Handled minimal missing data by removing affected rows or imputing with appropriate values (e.g., mode for categorical, median for numerical).
* **Outlier Detection**: Identified outliers in Price, Weight, and Screen Size using boxplots; extreme outliers were removed to prevent skewed analysis.
* **Inconsistent Values**: Standardized formats in Weight, Memory, and Screen Size using regular expressions and string manipulation to ensure uniformity.

**7. Feature Engineering**

To enhance analysis, new features were derived:

* **CPU\_Brand & CPU\_Series**: Extracted from the CPU column for better classification.
* **CPU\_GHz**: Parsed to obtain processor speed as a numerical value.
* **GPU\_Brand & GPU\_Series**: Extracted from the GPU column to separate brand and model.

These engineered features facilitated more meaningful groupings and comparisons.

**8. Data Filtering**

For focused analysis:

* **Operating Systems**: Retained only popular OS types like Windows, macOS, and Linux.
* **Storage Types**: Categorized into SSD, HDD, or Hybrid for clear comparisons.
* **CPU & GPU Brands**: Excluded less common brands to concentrate on well-known ones like Intel, AMD, and NVIDIA.
* These filters ensured clarity and relevance in the analysis.

**9. Statistical Analysis**

**Descriptive Statistics**:

* Calculated mean, median, mode, standard deviation, and range for key numerical variables.
* Observed significant price variations across different categories.

**Inferential Statistics**:

* **T-tests**:
  + Compared mean prices between different RAM sizes, revealing significant differences.
  + Assessed the impact of weight on pricing, finding notable variations.
  + Compared the average price of laptops , revealing significant differences
* **Chi-Square Tests**:
  + Analyzed the relationship between Operating System and Laptop Type, indicating dependency.

**10. Univariate Analysis**

Explored individual variables:

* **Price**: Most laptops priced between ₹25,000 and ₹70,000; premium models exceed ₹3,00,000.
* **RAM**: 8GB RAM is the most common, comprising nearly half of the dataset.
* **Screen Size**: 15.6 inches is the most prevalent, followed by 14.0 inches.
* **Distribution of Laptop Price:** laptop prices shows a right-skewed pattern with most laptops priced between ₹40,000 and ₹80,000

Visualizations like histograms, boxplots, and bar charts were employed for clarity.

**11. Bivariate Analysis**

Examined relationships between two variables:

* **Price vs. RAM**: Positive correlation; higher RAM generally leads to higher prices.
* **Screen Size vs. Weight**: Larger screens don't always mean heavier laptops; some are lightweight.
* **Price Distribution by OS**: Premium OS laptops (Windows 10, macOS) have varied pricing; Chrome OS or no OS laptops are more affordable.
* **Price vs. CPU**: Intel-based laptops have the highest average prices, followed by Samsung.
* **Price vs. GPU**: NVIDIA-equipped laptops are the most expensive, likely due to gaming and high-performance capabilities.

**12. Multivariate Analysis**

* **Correlation Heatmap**: Identified strong linear relationships among numerical variables.
* **Pair Plot**: Explored interactions between Price, RAM, Weight, and CPU\_GHz, revealing how these features collectively impact pricing.

**KEY INSIGHTS**

* SSDs make laptops faster to use, so laptops with SSDs usually cost more.
* Asus and Acer are more budget-friendly , Dell, HP, and Lenovo are mid-range.
* Lighter laptops are made for portability (travel, work).
* Heavier laptops are usually for gaming or heavy tasks.
* Laptops with 16GB+ are for gaming or advanced work.
* Higher resolution (Full HD or more) can increase the price, even if other specs are the same.
* A laptop with SSD + 16GB RAM + i7 CPU + GPU is always in the higher price range

**CONCLUSION**

This analysis effectively identified key factors influencing laptop pricing. Features such as SSD presence, RAM size, processor brand, and screen resolution significantly impact prices. These insights can guide manufacturers and sellers in developing accurate pricing strategies and enhancing competitive positioning in the market. Future work could involve building predictive models to forecast laptop prices based on specifications.