REINFORCEMENT TEST

PYTHON – LAPTOP DATASET  
KISHORE VASUDEVAN  
08/06/2025  
DATA ANALYTICS & DATA SCIENCE  
FEB 25’

**INTRODUCTION:**

This project analyses a laptop dataset containing detailed specifications such as company, type, screen size, RAM, storage, GPU, OS, and price. The primary objective is to clean and analyse this dataset to extract meaningful insights and understand key drivers of laptop pricing.

**OBJECTIVE:**

To analyze and understand the key specifications and factors that influence laptop pricing by cleaning, transforming, and exploring a real-world dataset.

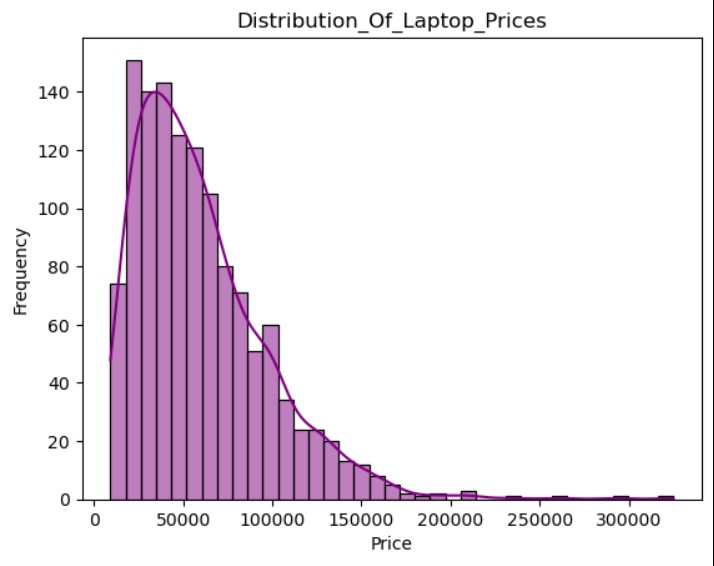
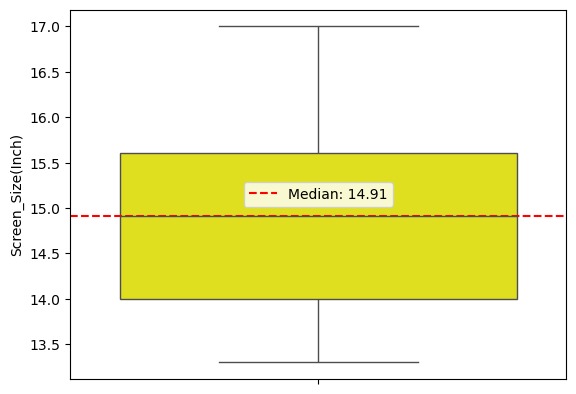
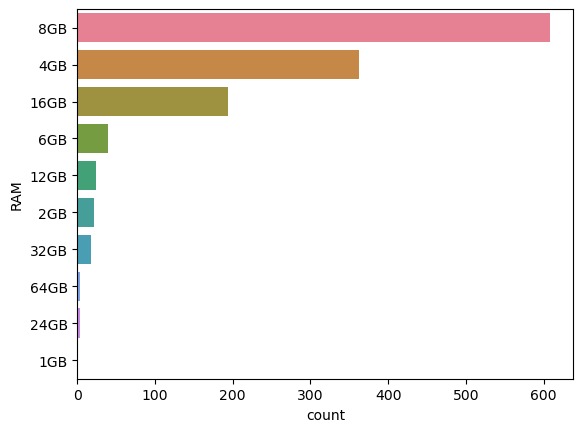
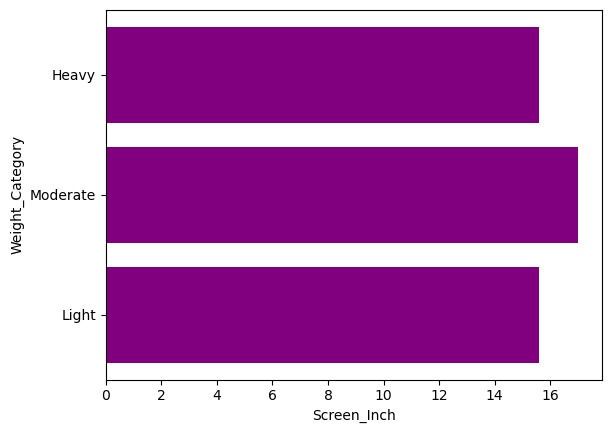
**TOOL USED:**

* Python

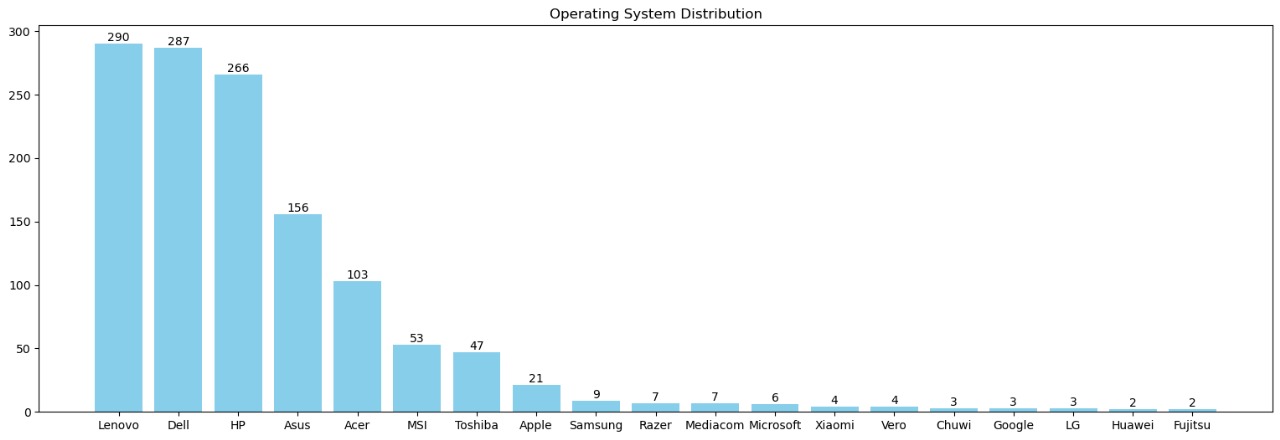
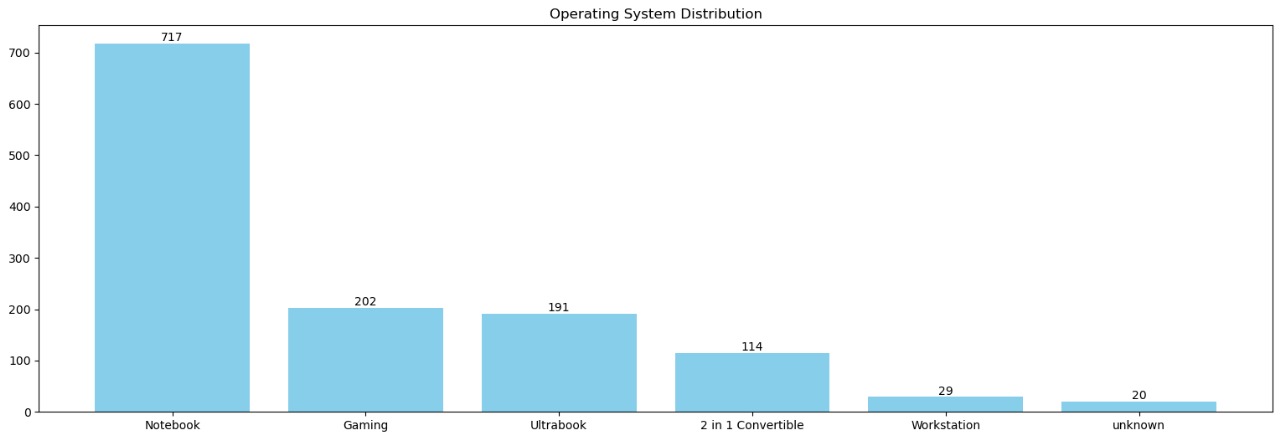
**DATA CLEANING AND PREPARATION:**

* The dataset was first loaded and explored to understand its structure, data types, and general statistics. This initial inspection helped in identifying potential issues such as missing values, duplicates, and formatting inconsistencies.
* The presence of missing values in each column was analysed both in count and percentage. This helped determine which columns required cleaning actions like filling or replacing values.
* Duplicate rows in the dataset were identified and removed using the drop\_duplicates() function. This ensured that repeated data entries did not skew the analysis results.
* Rows that contained null values in all columns were dropped from the dataset. These rows offered no useful information and would negatively impact data quality.
* An unnecessary column named "Unnamed: 0" was dropped because it did not contain any relevant or unique information. It was likely an index column from a previous CSV export.
* Missing entries in the TypeName column were filled with the label "unknown" to retain those records in the dataset. This avoids discarding rows that may still hold useful data in other columns.
* In the TypeName column, the category "Netbook" was replaced with "Notebook" to maintain consistency in laptop type classification. This standardization is important for accurate grouping and filtering.
* The Inches column contained invalid entries like "?" which were replaced with 0, and all null values were filled similarly. This was necessary to prepare the data for numerical operations.
* After replacing invalid symbols, the Inches column was converted from a string to a float data type. This allowed proper numerical analysis such as averaging or filtering.
* Screen sizes outside the normal range of 13 to 17.3 inches were considered unrealistic and replaced with the mean of valid screen sizes. This helped smooth out outliers that could distort analysis.
* The Weight column also had invalid "?" values, which were replaced with 0 for consistency. This step was a prerequisite for converting the values into a proper numeric format.
* Units like "kg" or "Kg" were removed from the Weight column to leave only numeric values. This ensured that the column could be correctly interpreted as numerical data.
* The cleaned Weight column values were converted from string to float type. This transformation enabled further statistical computations and data visualization.
* Values in the Weight column falling outside a reasonable range (1–4.5 kg) were replaced with the mean of valid weights. This treatment of outliers improved the accuracy of the analysis.
* Final Weight values were rounded to one decimal place to maintain consistency and improve readability. This step also ensures cleaner display in charts or tables.
* After all these transformations, data types of the cleaned columns were verified to ensure correctness. The dataset was now clean, standardized, and ready for further analysis or visualization.

**INSIGHTS:**

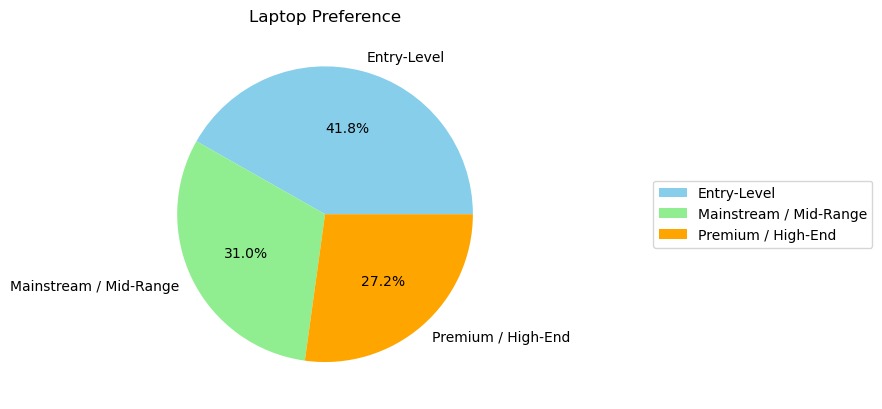
* The Price feature is a significant factor and shows positive skewness. Most laptops are priced around ₹52,000, falling within the mid-range segment. About 99% of prices fall below ₹167,785.
* The maximum laptop price is ₹324,954, which is not considered an outlier as it includes a dedicated GPU, indicating it’s a gaming laptop, thereby justifying the high price.
* The Screen Size has a median of 14.91 inches with no outliers detected, meaning 50% of the screen sizes lie close to this value, showing a consistent trend in laptop displays.
* The RAM configuration shows that 8GB is the most common with 608 laptops, followed by 4GB RAM at 362 units, highlighting a user preference for performance within budget.
* When comparing Price vs RAM, Processor Type, and GPU Type, although some points appear as outliers, they are mostly high-end laptops with dedicated GPUs, commonly gaming models, which justifies the elevated prices.
* By Weight category, moderate-weight laptops (≤ 3.5kg) are the most preferred, offering a balance between portability and performance.

***In the correlation analysis:***

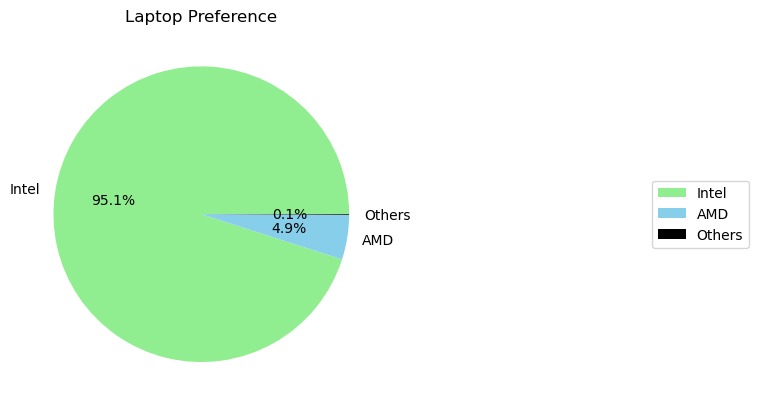
* Price is strongly correlated with RAM, likely due to the influence of high-tech processors.
* There’s a moderate correlation with Weight.
* Negligible correlation is observed with Screen Size, indicating it doesn’t heavily impact pricing.
* Lenovo (290 units) and Dell (287 units) are the top-selling brands, likely due to their wide range of offerings and availability.
* Lenovo dominates the entry-level segment, making it more attractive for budget-conscious buyers.
* Notebook-type laptops are the most popular, with 717 units, far ahead of Gaming laptops at 202 units.
* Specifically, Lenovo contributes 174 Notebook-type laptops, reinforcing its budget-friendly branding.

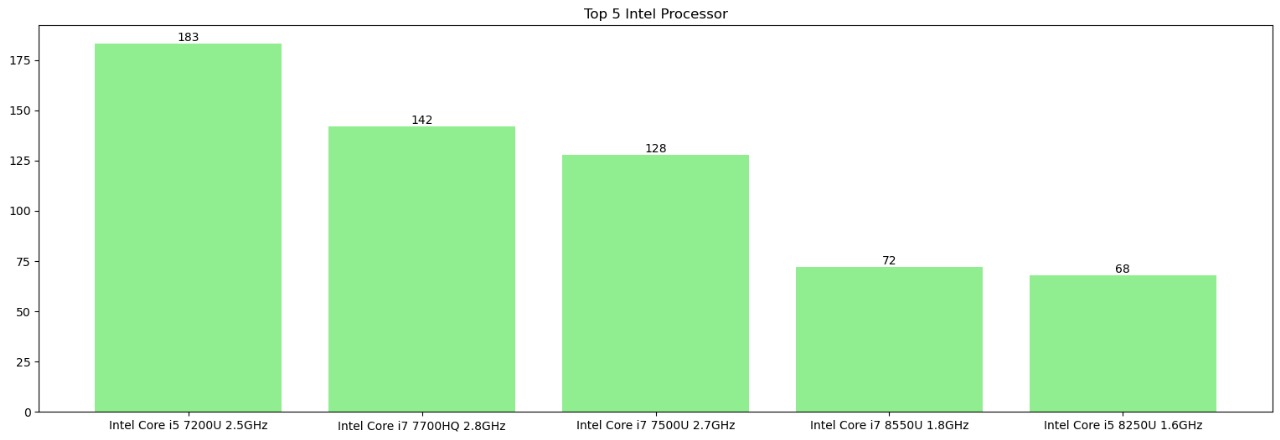
***In market segmentation:***

* Entry-level laptops account for 41.8% of sales,
* Followed by Mainstream/Mid-range at 31%, and
* Premium/High-end at 27.2%.
* This clearly shows a strong price sensitivity among users.



***For Processor Type:***

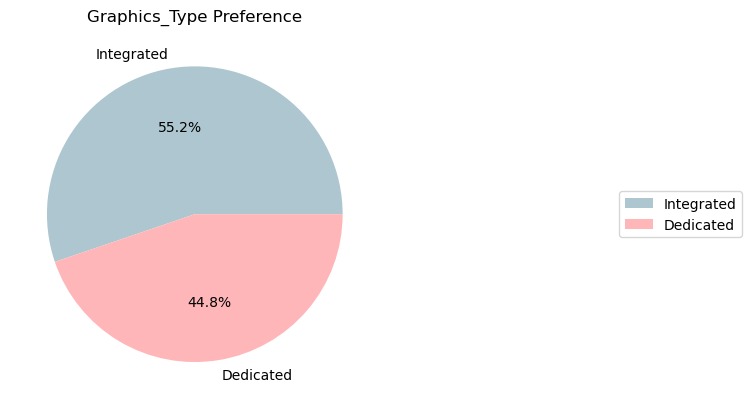
* Intel dominates with a 95.1% market share, likely due to its better perceived performance.

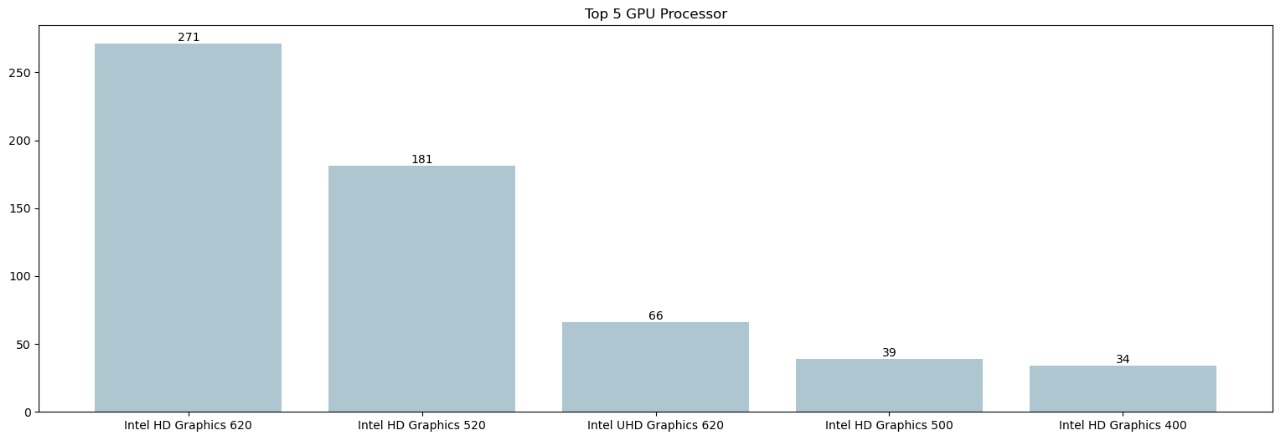


* Among Intel models, Intel Core i5 U GHz is the most selected, with 183 units.

***For GPU Type:***

* Integrated GPUs are preferred due to affordability and basic usage needs.
* The most common GPU is Intel HD Graphics 620, present in 271 laptops.





* An Independent T-test on GPU Type vs Price reveals a statistically significant price difference between Integrated and Dedicated GPUs, leading to rejection of the null hypothesis.

**CONCLUSION:**

The analysis of the laptop dataset reveals that price is strongly influenced by RAM and processor type, with gaming laptops and dedicated GPUs commanding higher prices. Most laptops fall within the mid-range price segment, with Lenovo and Dell leading sales, particularly in the budget and notebook categories. Weight and screen size have less impact on pricing. Intel dominates the processor market, and integrated GPUs are more common due to affordability. Market segmentation shows a strong user price sensitivity favouring entry-level and mid-range laptops. Statistical testing confirms a significant price difference between laptops with integrated versus dedicated GPUs. Overall, the dataset provides valuable insights into key factors driving laptop pricing and consumer preferences.