**Project Information**

|  |  |
| --- | --- |
| Title | Laptop dataset |
| Name | Suvadha M |
| DA/DS | May-2025 |
| Batch Number | RP-36 |
| Online/Offline | Offline |

**Table of Contents**

**1. Project Overview**

**2. Project Objective**

**3. Business Challenge**

**4. Process Flow**

5. **Initial Data Exploration**

**6. Data Preparation & Handling Missing Values**

**7. Summary Statistics**

**8. Hypothesis Validation**

**9. Single Variable Exploration**

**10. Two-variable Relationship Analysis**

**11. Multiple Variable Relationships**

**12. Key Takeaways from Findings**

**13. Potential Enhancements.**

**14. Closing Thoughts**

**1. Project Overview**

There are many laptop models with different features. This makes it challenging for consumers to choose the best option. In this project, we examine a dataset containing details such as brand, processor, RAM, storage, and operating system. We clean the data and use basic analysis to identify which features influence laptop prices and performance.

**2. Project Objective**

The goal is to explore and understand the laptop dataset. First, we clean the data. Then, we visualize and analyze how factors like brand, RAM, and screen size impact prices. This helps people compare laptops more easily. In the long term, this analysis can support price prediction models.

**3. Business Challenge**

Laptops differ widely in terms of brand, specifications, and price, making direct comparisons difficult. Businesses also need accurate insights to price products competitively and understand market positioning.

**4. Process Flow**

I start by loading the dataset and examining its structure. Next, we clean the data by handling missing values, outliers, and inconsistencies. After data preparation, we generate new metrics and filter the dataset. We then conduct descriptive analysis, hypothesis testing, and finally, exploratory data analysis (EDA) through various visualizations.

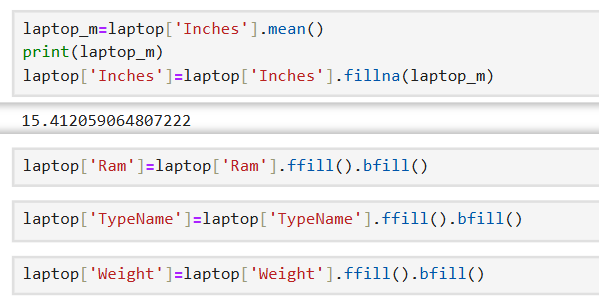
**5. Initial Data Exploration**

The dataset includes information on brand, laptop type, RAM, screen size, weight, and price. Using Python (pandas), we identified missing values and an unnecessary column, which we removed. We also corrected data formats in columns like RAM and Inches, preparing the dataset for deeper analysis.

**6. Data Preparation & Handling Missing Values**

Missing values appeared in fields like Inches, TypeName, and Ram. We applied forward-fill and backward-fill methods. For numeric fields like Inches, we replaced missing values with the mean.

Managing the missing values using mean, ffill, bfill.

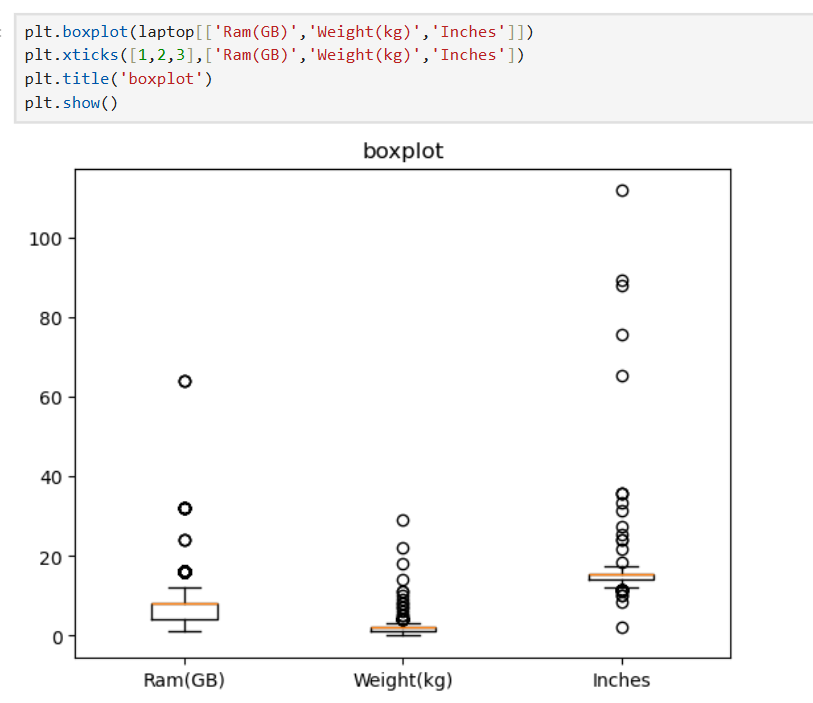


**7. Managing Outliers**

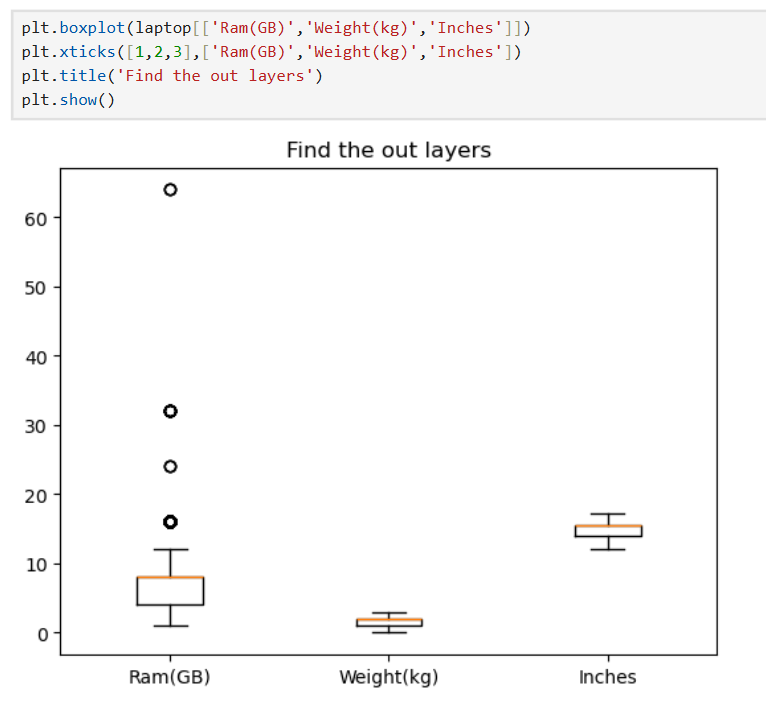
I detected extreme values in numeric columns like Price and Weight using box plots and visual checks. Some outliers were noted but kept if they seemed valid; others were adjusted or removed.



**CHART BEFORE REMOVING OUTLIERS:**

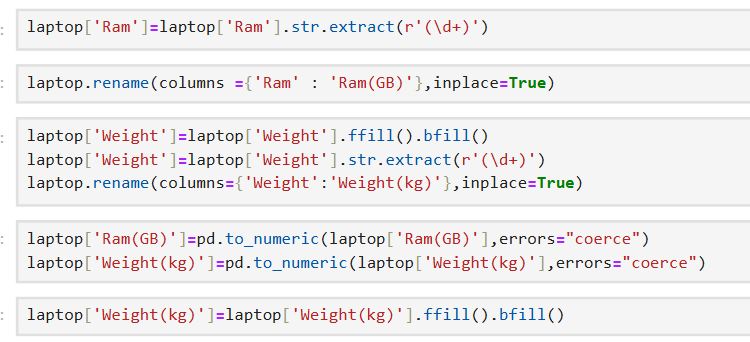
****

**CHART AFTER REMOVING OUTLIERS:**

****

**i. Standardizing Data Entries**

I cleaned columns like "Ram" by splitting values (e.g., "8GB") into numbers and units. We also removed empty rows and unified data formats to ensure consistency.



**ii. Creating New Features**

To enhance analysis, we derived new columns from existing ones—such as numeric RAM and cleaned Inches data. This transformation supported better visualizations and numerical analysis.

**iii. Data Selection for Insights**

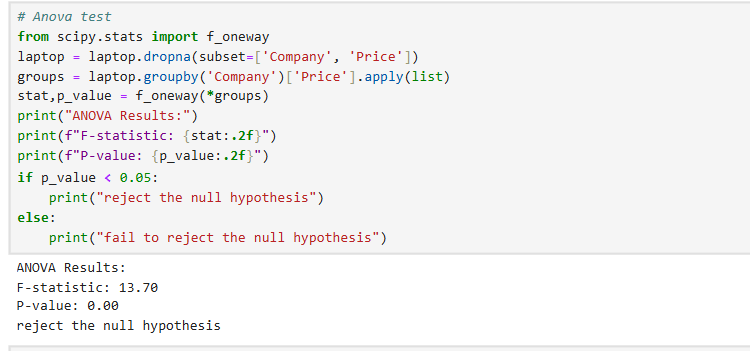
I filtered out invalid or duplicate records and removed rows with critical missing values to keep the dataset focused and relevant.

**iv. Summary Statistics**

I computed descriptive statistics (mean, median, standard deviation, range) to understand distributions of variables like Price, RAM, and Inches. The describe() function revealed patterns such as skewed price distribution.

**8. Hypothesis Validation**

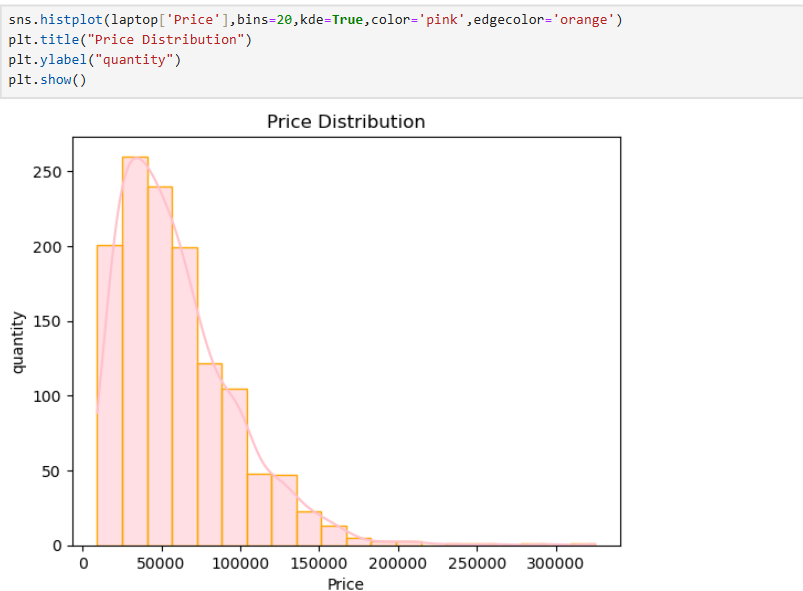
Though limited in this project, statistical tests like t-tests or ANOVA can check if average prices significantly differ between brands or laptop types, supporting data-driven decisions.



**9. Single Variable Exploration**

Iused histograms and bar plots to analyze distributions of individual variables:

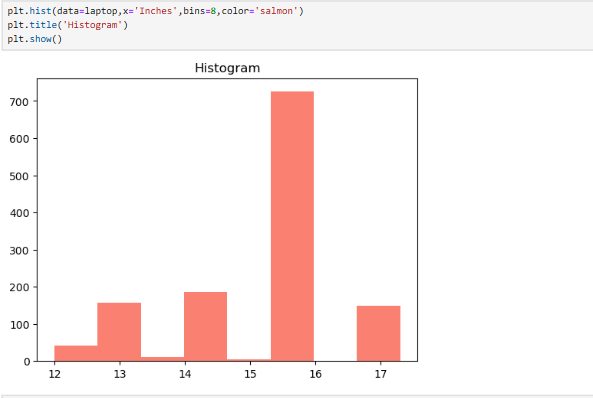
* Price mostly ranged between ₹30,000 to ₹80,000.
* "Notebook" and "Ultrabook" were common types.
* 4GB, 8GB, and 16GB RAM were the most frequent.



**10. Two-variable Relationship Analysis**

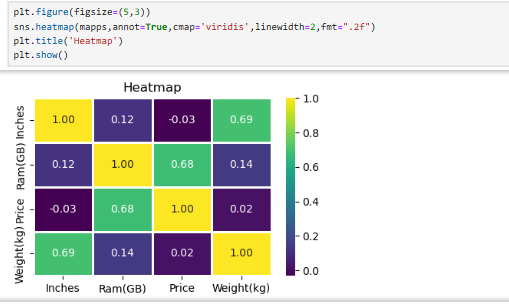
Scatter plots and box plots showed relationships like:

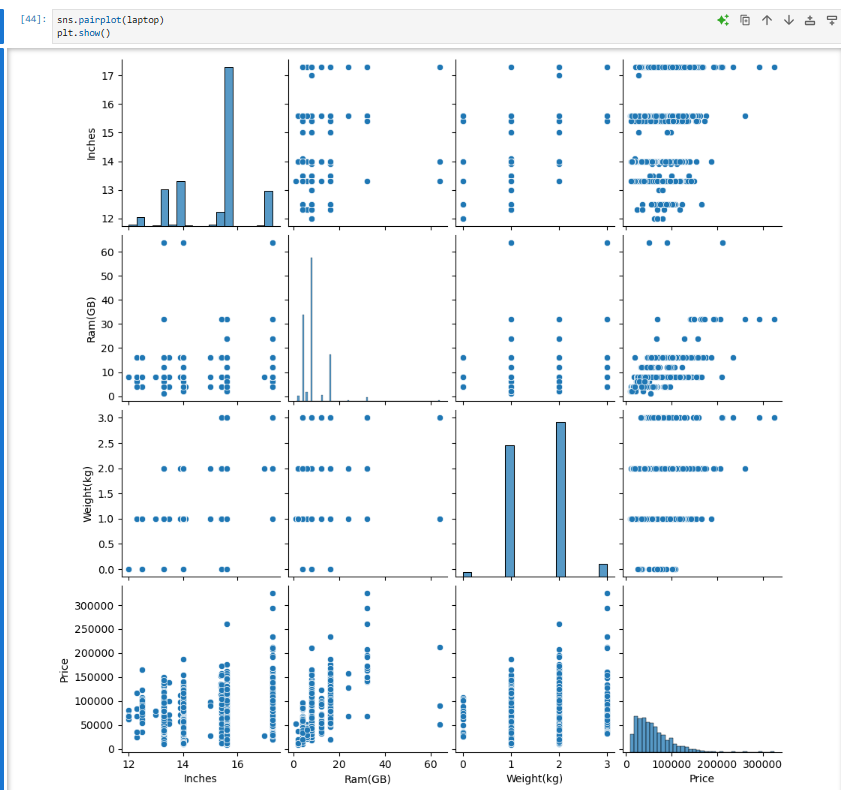
* Higher RAM typically meant higher prices.
* Gaming laptops were priced highest on average.
* Larger screen sizes correlated with higher prices, though not always directly.



**11. Multiple Variable Relationships**

Using pair plots and heatmaps, we examined complex interactions among variables. We observed moderate correlations between RAM and Price, and Inches and Price. Grouped comparisons showed that brands like Apple and MSI had higher prices even with similar specs.





**12. Key Takeaways from Findings**

* Mid-priced laptops dominate the market; premium models are less common.
* Gaming laptops and premium brands cost more.
* More RAM and larger screens generally increase price.
* Brand significantly influences pricing.
* Cleaned data provided clearer and more reliable insights.

**13. Potential Enhancements**

* Target budget models (2GB–4GB RAM) with free accessories to boost sales.
* Consider additional features (e.g., SSD, GPU) for deeper price analysis.

**14. Closing Thoughts**

This analysis helped identify how factors like RAM, brand, and screen size affect laptop prices. Data cleaning played a critical role in achieving accurate results. In the future, these insights could inform recommendation systems and price prediction tools to assist consumers and businesses alike.

Github Link : <https://github.com/SuvadhaM/pythonproject>