**Project Information**

* **Title**: laptop dataset
* **Name**: Rahul Nikshai P
* **DA/DS**: Data Analyst
* **Batch Number**: RB-36
* **Online/Offline**: Offline

**Table of Contents**

|  |  |
| --- | --- |
| **S.No** | **INDEX** |
| **1** | Introduction |
| **2** | Aim |
| **3** | Business Problem / Problem Statement |
| **4** | Project Workflow |
| **5** | Data Understanding |
| **6** | Data Cleaning |
| **7** | Obtaining Derived Metrics |
| **8** | Filtering Data for Analysis |
| **9** | Statistical Analysis   * Test Statistics and Hypothesis Testing |
| **10** | Exploratory Data Analysis (EDA) |
| **11** | Univariate Analysis |
| **12** | Bivariate Analysis |
| **13** | Multivariate Analysis |
| **14** | Overall Insights from Analysis |
| **15** | Future Engineering (Key Findings & Strategy) |
| **16** | Conclusion |

1. **Introduction**

There are many laptop models with different features. This makes it hard for people to choose the right one. In this project, we study a dataset that has details like brand, processor, RAM , storage, and operating system. We clean the data and use simple data analysis to understand which features affect the laptop’s price and performance. This project uses Python tools to explore a dataset related to laptop specifications and sales. The goal is to clean, analyze, and visualize the data to identify trends and insights that can help businesses make smarter inventory and marketing decisions.

1. **Aim**

We want to understand what laptop Dataset. first we clean the data. Then we use charts and simple way. This helps us see how things like RAM, brand and size change the price. its to help people choose the right laptop by comparing features and prices. In the future, this work can also help to predicts laptop prices.

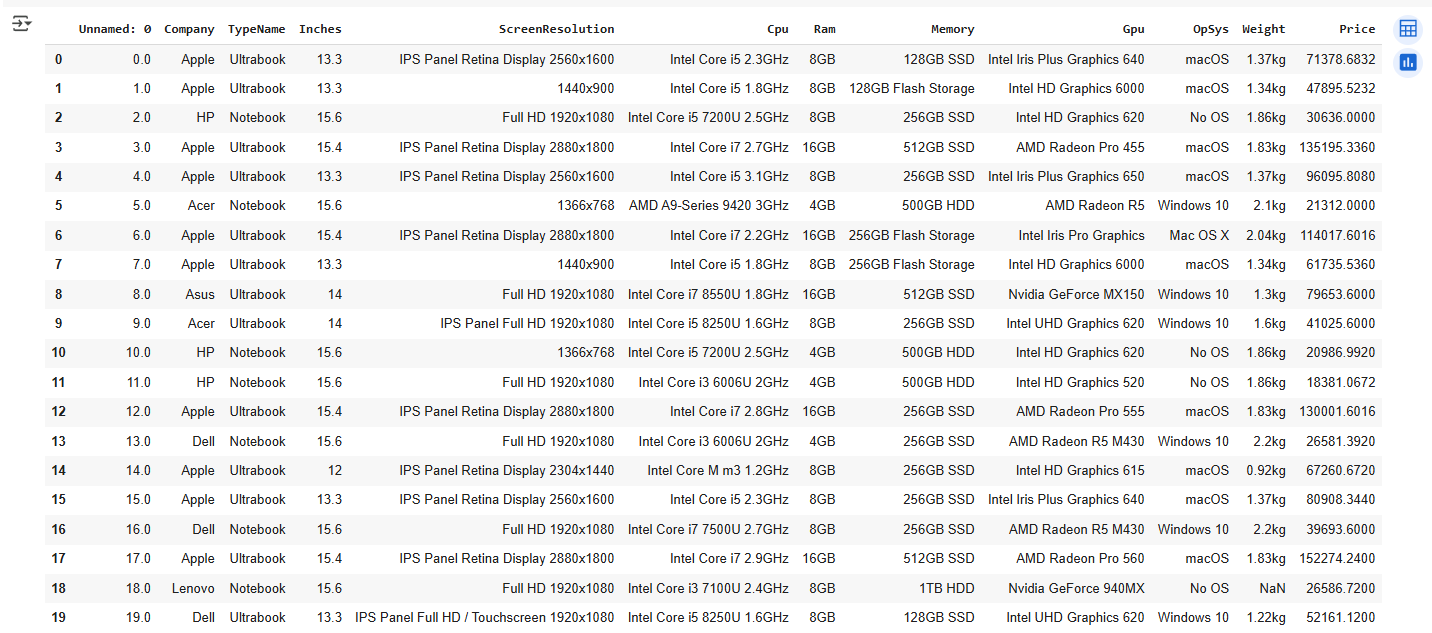
1. **Business Problem / Problem Statement**

Each laptop differs in brand, processor, RAM, storage type, display size, operating system, and other specifications, making direct comparisons complex. Businesses, too, need reliable insights to price their products competitively

1. **Project Workflow**

The data preparation phase involved cleaning the dataset by handling missing values, addressing outliers, and resolving inconsistencies in formatting or data entries. After cleaning, new features were engineered where relevant, and the dataset was filtered to retain meaningful records for analysis.

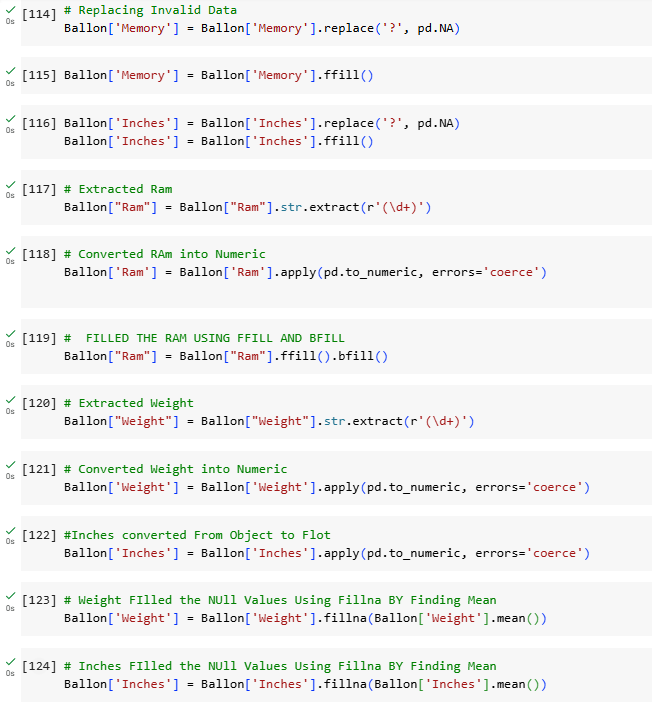
The dataset has details about laptops like brand, type, RAM, screen size, weight, and price.We used Python (pandas) to load the data and found some missing values and an extra column, which we removed. Some data like RAM and Inches were in the wrong format, so we fixed them. This helped us get the data ready for analysis.



1. **Data Cleaning Missing Values Imputation** 
   * Missing Values Imputation

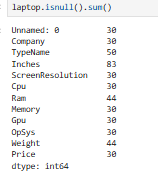
Missing values were found in fields like Inches, TypeName, and Ram. These were filled using forward-fill (ffill) and backward-fill (bfill) methods. For numeric fields like Inches, the mean value was calculated and used to replace missing entries.

Ex: using ffill(),bfill(),fillna(),like mean,medien,mode



* + Outlier Treatment

Extreme values in numeric fields like Price or Weight can distort analysis. Visual inspection using box plots or value counts helped identify and understand outliers. Based on domain knowledge and visual analysis, extreme or unrealistic entries were noted, though not all were removed if they were valid.



1. **Obtaining Derived Metrics**

To make the analysis better, we created new columns from existing ones. For example, we split the "Ram" column into number and unit. We also cleaned "Inches" and "Weight" to use them in comparisons. Changing text to numbers helped us make charts and do math easily. This made the data ready for deeper study.

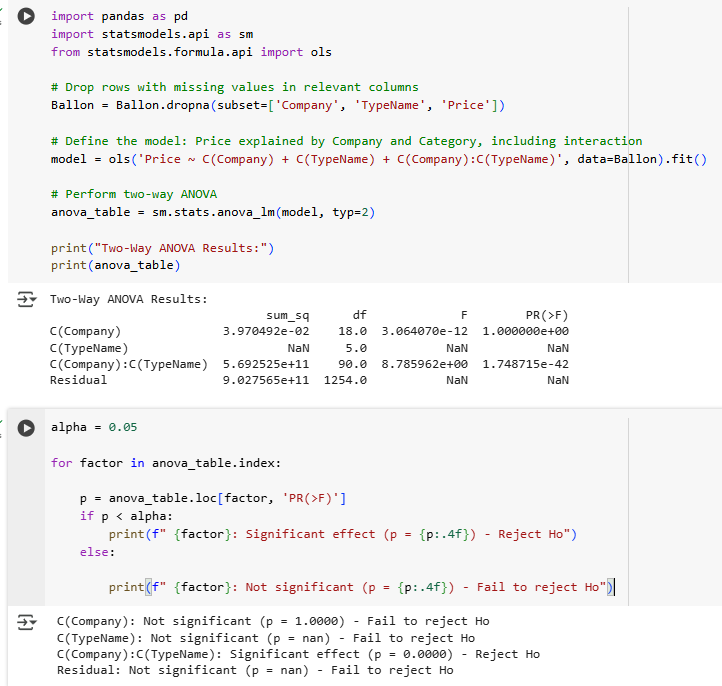
1. **Filtering Data for Analysis**

To keep the data clean and useful, we removed rows with missing or wrong values. We also dropped empty entries using simple filters. Some laptops with very odd or repeated details were removed too. This helped us keep only real and useful laptop data for better analysis

Eg: dropna()

1. **Statistical Analysis** 
   * Test statistics and hypothesis testing

statistical tests like ANOVA could be used to examine whether price differences between brands or types (e.g., Gaming vs Notebook) are statistically significant. For example, we could test if the average price of “Gaming” laptops differs significantly from “Notebook” types.



1. **Exploratory Data Analysis (EDA)**

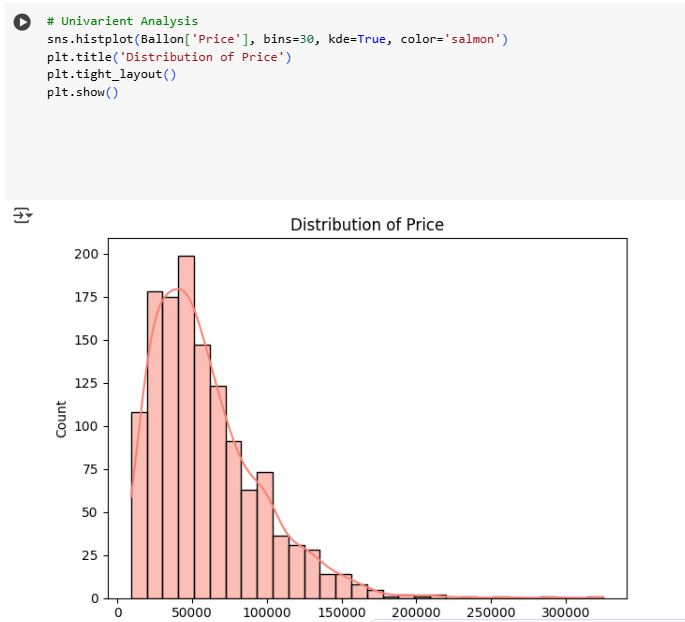
**There are 3 types**

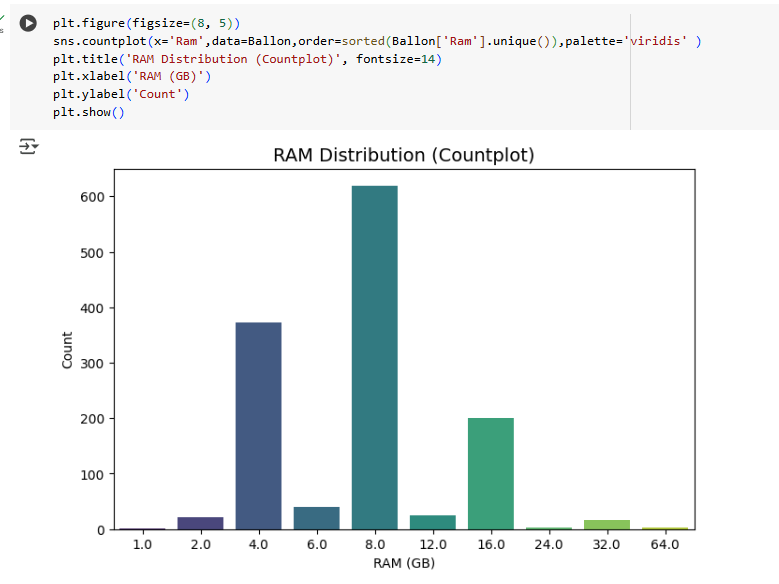
* Univariate analysis
* Bivariate analysis
* Multivariate analysis

1. **Univariate Analysis**

Used histograms and countplot to explore individual feature distributions:

* + Price: Most laptops priced between ₹30,000–₹80,000; fewer beyond ₹100,000.
  + RAM: 4GB, 8GB, and 16GB were the most common configurations.

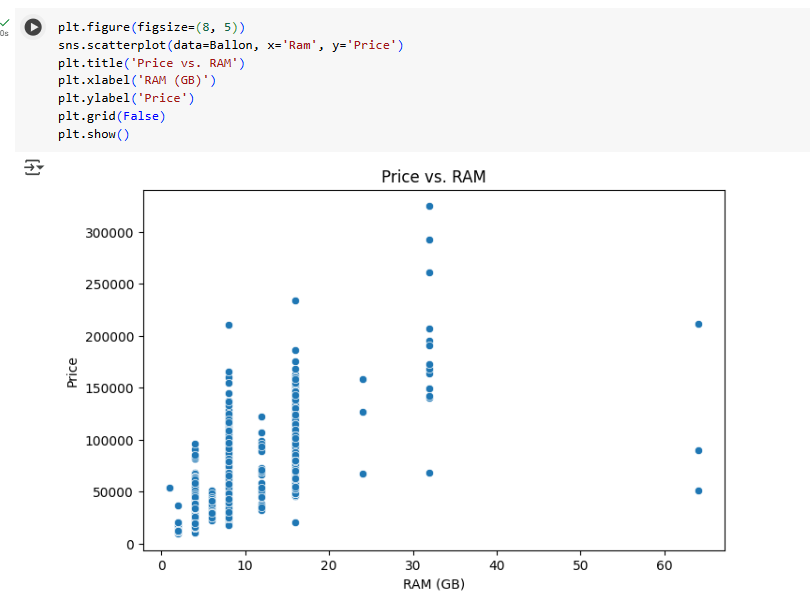


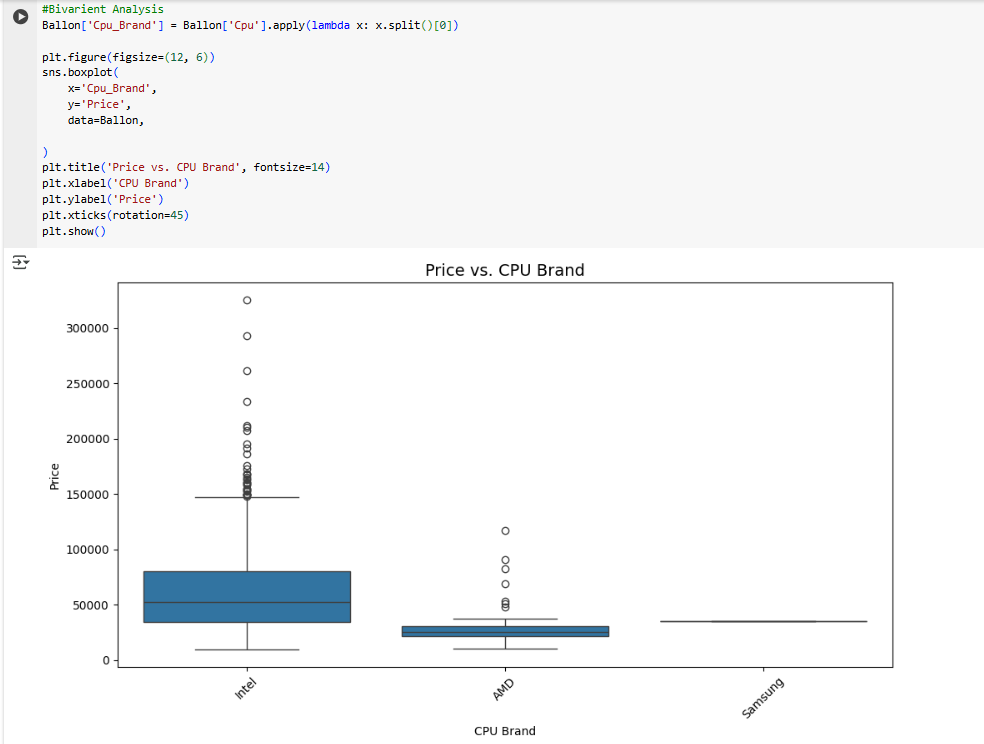


1. **Bivariate Analysis**

Explored relationships between pairs of variables using visualizations like scatter plots, box plots, and bar plots:

* + **Price vs. RAM**: Higher RAM generally correlated with higher prices.
  + **Price vs. TypeName**: Gaming laptops were the most expensive, followed by Workstations and Ultrabooks.
  + **Price vs. Screen Size**: Larger screens tended to cost more, though with notable variation.

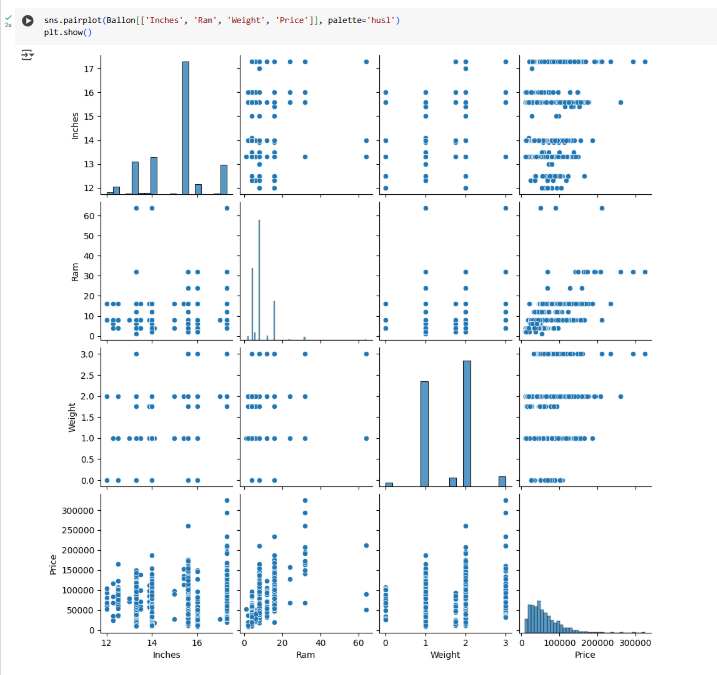
****

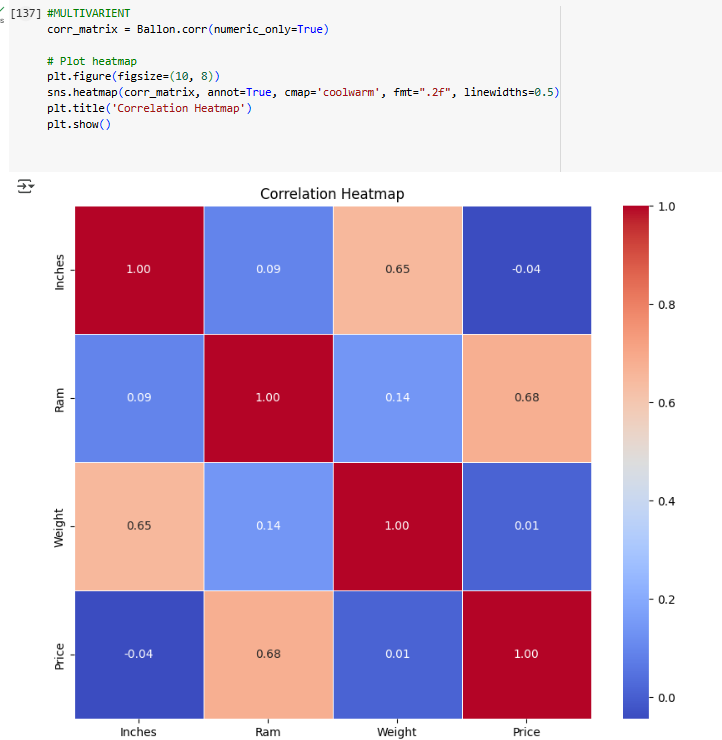
****

1. **Multivariate Analysis**

Explored interactions between multiple variables using pair plots and heatmaps:

* **Pair Plot**: Visualized combinations like Price vs. RAM vs. Screen Size.
* **Heatmap**: Highlighted correlations—moderate between RAM & Price, and Inches & Price.





1. **Overall Insights from Analysis**
2.  **Price**: Most laptops are mid-range; high-end models are fewer and brand-driven.
3.  **Type**: Gaming laptops are costliest; Notebooks and Ultrabooks are more common and affordable.
4.  **RAM & Screen**: Higher specs generally increase price, but gains flatten beyond 16GB RAM or 15.6" screens.
5.  **Brand Impact**: Premium brands (e.g., Apple, MSI) charge more for similar specs.
6.  **Storage & GPU**: SSDs and dedicated GPUs influence price, though not deeply analyzed.
7.  **Data Quality**: Cleaning improved accuracy and clarity of insights.
8. **Future engineering**

**Key Findings**:

* Laptops with **2GB and 4GB RAM** show **higher sales** compared to other configurations.
* These models are likely **affordable** and attract **budget-conscious customers**.

**Boosting Sales Strategy**:

* Offer **complementary items** like:
  + - Free **keyboard and mouse**
    - Optional upgrade to **SSD storage**
* These extras **increase perceived value**, encouraging more purchases without raising the laptop price.

1. **Conclusion**

EDA provided valuable insights to improve laptop sales through smarter pricing, bundling, and inventory strategies. Python proved effective in turning raw data into actionable business understanding.