REINFORCEMENT PROJECT

PYTHON – LAPTOP DATASET ARUNACHALAM N

22/07/2025

DATA ANALYTICS & DATA SCIENCE

MAY 25

**Introduction**

This project focuses on analysing a comprehensive laptop dataset that includes attributes like manufacturer, model type, screen dimensions, RAM, storage capacities, GPU, operating system, and price. The core aim is to thoroughly clean and process the data to ensure accuracy and consistency. By applying statistical and exploratory data analysis techniques, the project seeks to identify significant trends and relationships within the dataset. Ultimately, the analysis will uncover meaningful patterns and key factors that influence laptop pricing, providing actionable insights into what drives value in the laptop market.

**Aim**

The primary goal of this project is to analyse laptop specifications and pricing to generate valuable insights that can support purchasing decisions, marketing approaches, and product development. By conducting comprehensive Exploratory Data Analysis (EDA), the project aims to examine how critical features like RAM, processor type, storage, and brand affect laptop prices

**Business Problem / Problem Statement**

* In today’s competitive technology market, consumers have numerous laptop options across various brands, configurations, and prices.
* Manufacturers and retailers need to understand which specifications most influence customer preferences and pricing to boost sales and market share.
* Common business challenges include uncertainty about important features, difficulty in competitive pricing without profit loss, and limited visibility into shifting market trends.
* This project tackles these issues by analysing real-world laptop data.
* It explores how attributes like brand, processor, RAM, storage, and screen size impact sales and pricing.
* Key questions addressed include: which brands and specs are most popular, how features affect pricing, and what defines top-selling laptops.

**Project Workflow**

**Data Collection**

I have gathered a dataset from an online laptop retailer that lists various specs for each laptop—like brand, model type, screen size, processor details, RAM, storage, GPU, operating system, weight, and price.

**Getting to Know the Data**

We started by exploring the structure and size of our dataset, checking out the data types and spotting which fields are most crucial for analysis.

**Cleaning and Pre-processing**

To ensure the data’s accuracy, we removed rows with missing information and got rid of any duplicates. We also converted text fields such as RAM and storage into numeric formats, and broke down more complex columns—like Weight to **Weight\_in\_Kg** and Ram to **Ram in GB** into easier-to-use pieces

**Feature Engineering**

I have created new columns, like numeric category preference, and OS\_Type into categories to enrich our analysis and make comparisons simpler

**Filtering and Focusing**

To keep our analysis relevant, we filtered out extreme outliers in Weight and Inches in laptops that didn’t fit our study. We also targeted on the improvisation of the Total Sales by the current Trends.

**Exploratory Data Analysis (EDA)**

I have done detailed analysis using a mix of visualizations—histograms, bar charts, box plots, and scatter plots will be use to uncover important trends and relationships, especially around what features affect pricing and which laptops are most popular.

**Sharing Insights**

Finally, I have summed up the key findings in an easy-to-understand format, highlighting practical takeaways for the business. This included which laptop configurations are trending, how pricing varies, and which brands are performing best.

**Data Understanding**

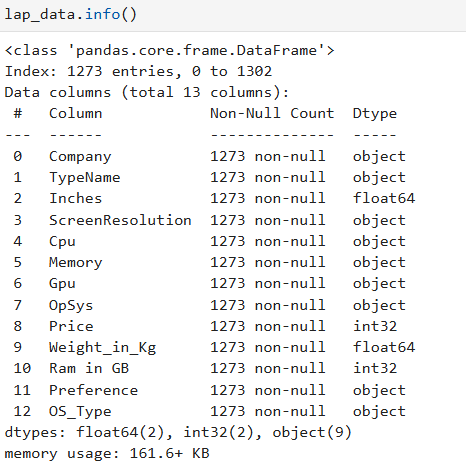
The dataset contains information about Laptop sales including various features like RAM, Processor, memory and sale price. It explains about the detailed overview over the laptop project.

**Dataset Overview**

**Format:** Tabular (structured data)

* **Dimensions:**
  + **Rows:** 1,303 (each row represents a laptop)
  + **Columns:** 12 (each column represents a specification or attribute)

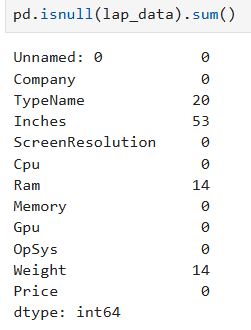
**Structure and Data Types**



**Data Cleaning**

**First, we will be finding out the null values by below.**

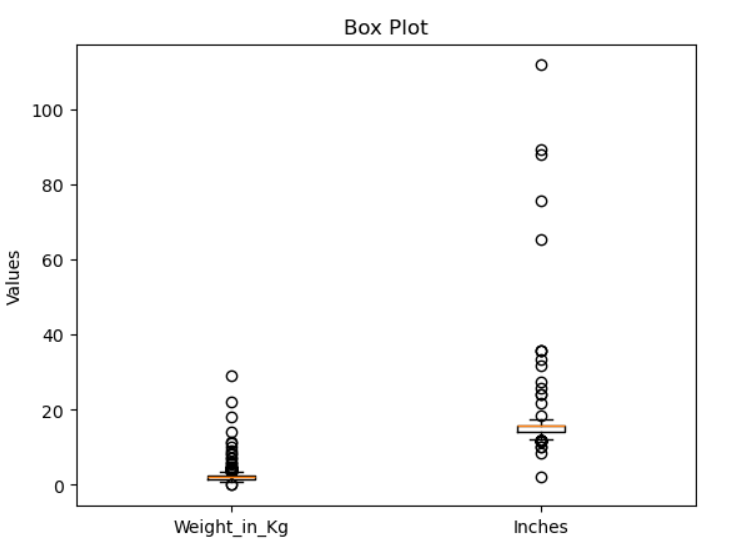
* Used .isnull().sum()



**Outlier Detection**

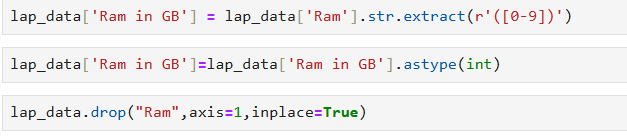
**Detection:**

* I have used Visual plots like (box plots, histograms) to spot unusually high or low values in numeric fields such as:
* Weight
* Inches



**Handling Inconsistent Values**

* Handling inconsistent values is an important step when getting the laptop dataset ready for analysis.
* Data changes may come from things like typos, different formatting styles, or combining data from several sources.
* And I have converted measurements and numeric data into the same units and formats across the dataset.
* We also converted measurements and numeric data into the same units and formats across the dataset.





**Obtaining Derived Metrics**

This laptop dataset already had clear and useful columns like RAM, CPU, and Price, which were enough for my analysis so I have created a new column **category preference** and **OS\_Type** for explaining the Range of the Laptop category the by Tier 1,2 and 3 also adding Low End Model , Mid-End Model and Premium-End.

Then I have added OS Type for shortlisting the name of the OpSys for quick understanding about the laptops versions easily.

**Filtering Data for Analysis**

I cleaned the dataset by getting rid of any duplicate entries and filling in or removing missing values. I also made sure that columns like “Weight” and “Inches” were converted into the right numeric formats so they could be easily analysed. To keep things simple and focused, I dropped any columns that weren’t needed for the analysis.

**Statistical Analysis**

**Descriptive Statistics**

Descriptive Statistics are numerical summaries of data sets that help make data understandable and interpretable.

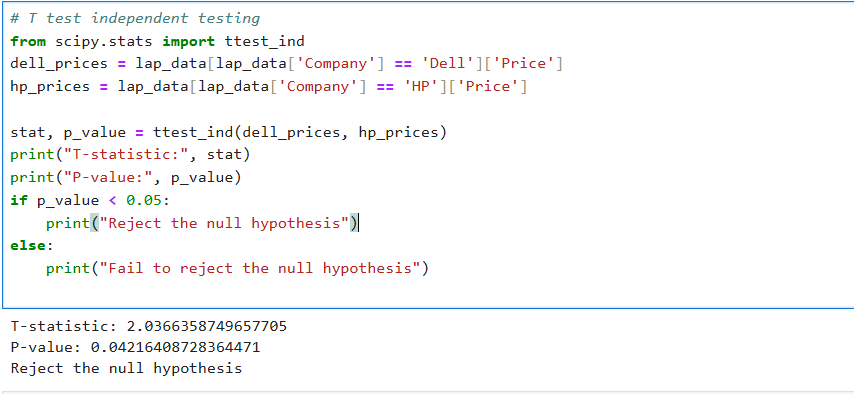


**Test statistics and Hypothesis testing:**

**Assumption:**

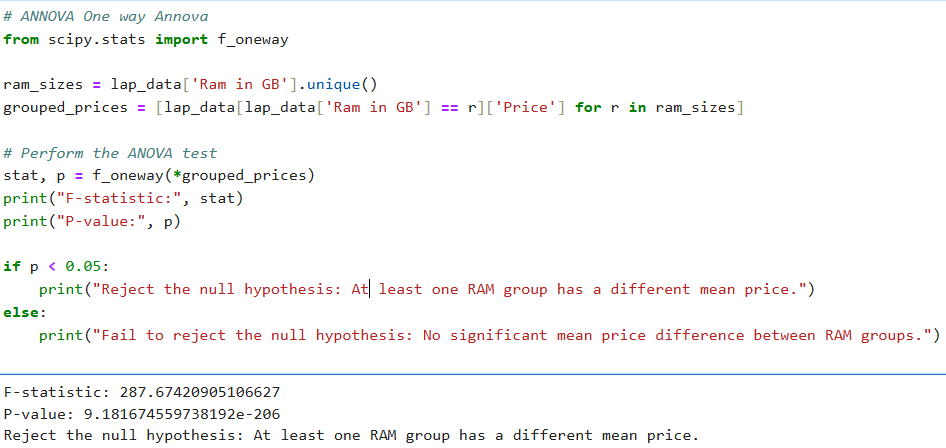
**t-test:**

* To **compare the means of two independent groups** and check if they are significantly different.
* Performed a **t-test** to check if the **average price** is different between **Dell** and **HP**.
* The t-test gave a **p-value of 0.0421**, which is **less than 0.05**. So, **rejected the null hypothesis**.
* HP and DELL laptops are **significantly more expensive**.



**One-way Anova:**

* One-Way ANOVA test is used to compare the average (mean) of a numerical variable across more than two groups.
* one-way ANOVA test has been used to check if all laptop brands have the same average RAM. The result showed a clear difference (F = 9.18-206, p < 0.05).
* This means some brands offer more RAM than others, and the average RAM is not the same aross all brands.

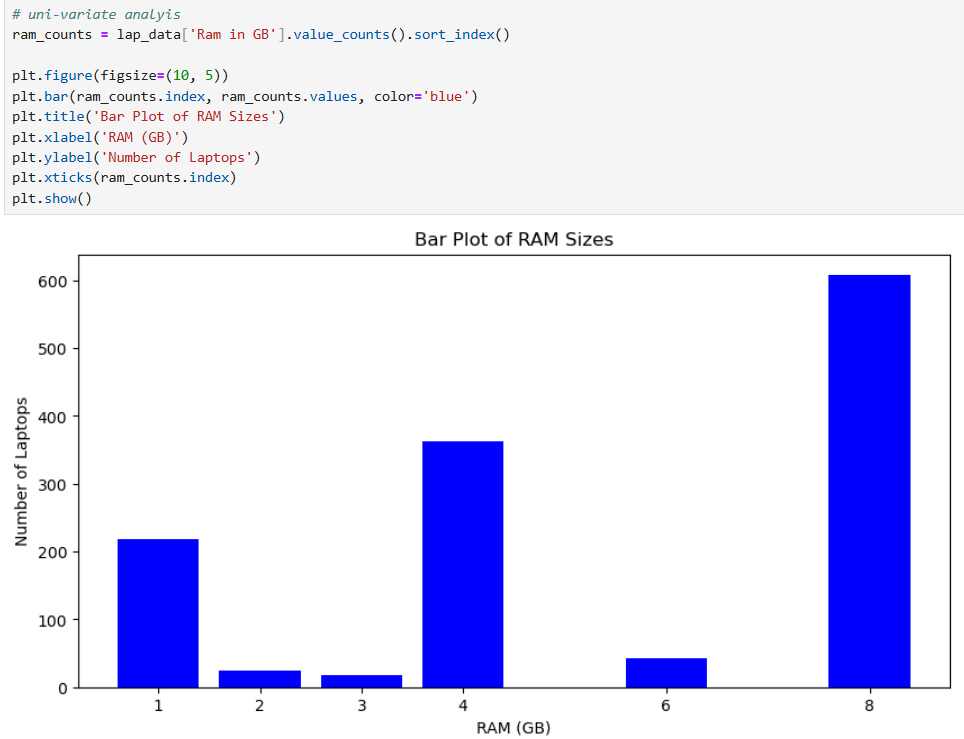


**EDA - Exploratory Data Analysis** is the process of analysing datasets to summarize their main characteristics often using visual methods

1. **Univariate Analysis:**

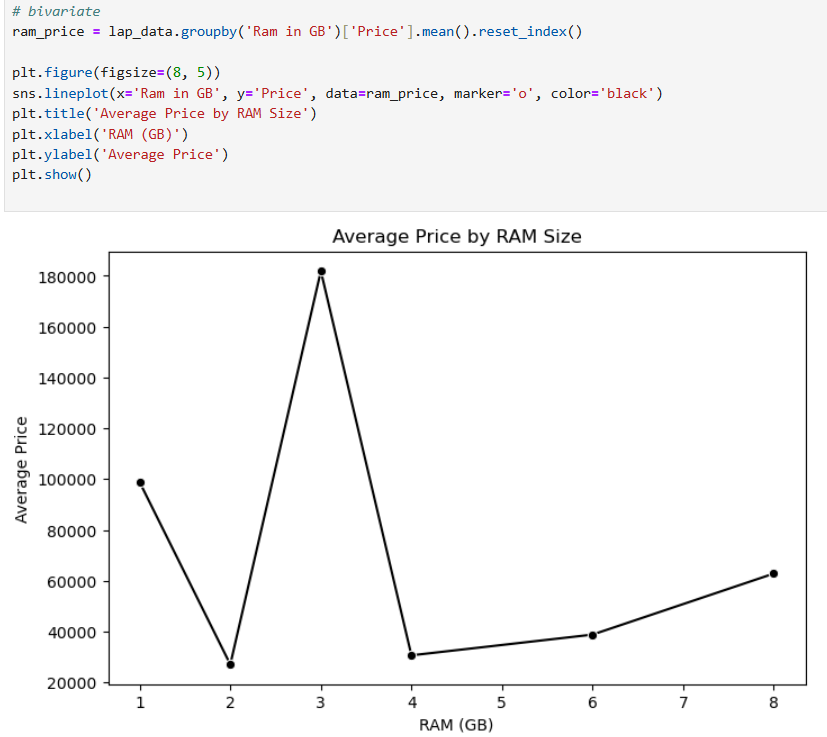
Univariate analysis is all about examining one variable at a time. In this project, I used visual tools like histograms and count plots to explore things like how laptop prices are distributed, how many laptops each company offers, and the breakdown of memory usage across models.

* First, I got familiar with the dataset’s overall structure and what information it holds.
* Then, I looked for patterns, trends, and any interesting relationships in the data.
* I also checked for missing values, unusual data points, or errors that could impact the results.
* Finally, I cleaned and prepared the data so it’s ready for deeper analysis or building machine learning models.



**Bivariate Analysis:**

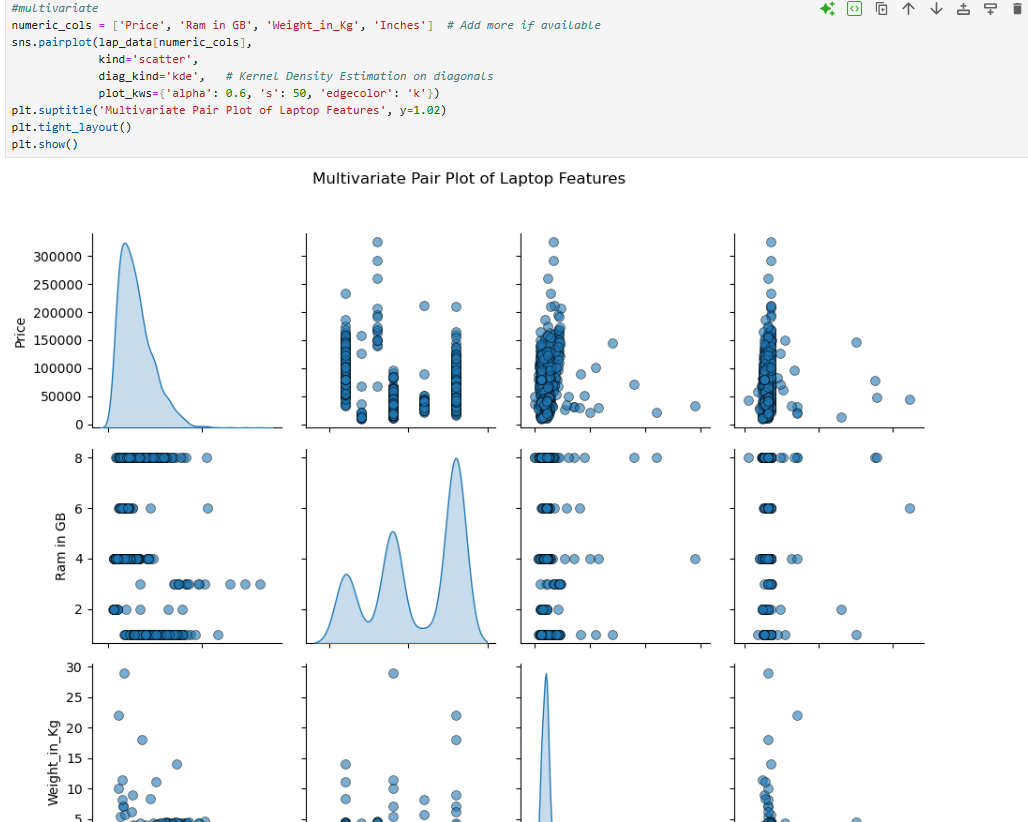
Bivariate analysis is about looking at two variables together to understand if and how they’re connected. In this project, I used bar plots and scatter plots to explore relationships like how laptop price varies with screen size, how RAM relates to price, and how different laptop types are distributed across operating systems. Essentially, it’s a way to study how pairs of features interact and influence each other.



**Multivariate Analysis**

Multivariate analysis involves examining more than two variables at once.

* It helps to explore relationships and patterns among three or more variables simultaneously to understand how they interact with each other.
* In simple terms, it means analyzing multiple factors together to see the combined effect and connections within the data.
* For this project, I used visual tools like Heatmaps, 3D Scatter Plots, and Pair Plots to study how variables such as screen size (Inches), weight (kg), price (₹), company and type, and RAM (GB) relate to each other.



**Overall Insights**

* Premium brands like Apple, Razer, and MSI generally have higher prices than brands like HP, Dell, and Lenovo.
* Gaming laptops are the most expensive, followed by Ultrabooks, while Notebooks and Netbooks tend to be more affordable.
* Laptops with 16GB RAM or more usually cost more and are mostly found in gaming or high-end models.
* SSD storage makes laptops faster but also more expensive compared to HDDs. Some laptops use both (hybrid) to balance speed and storage space.
* Gaming laptops are heavier (around 2.5 kg or more), whereas Ultrabooks and Notebooks are lighter and easier to carry.
* Most laptops run Windows OS, then Mac OS for Apple laptops, and a smaller number use Linux or come without an OS to keep costs low.
* Among all RAM options, laptops with 32GB RAM sell the most, while those with 2GB RAM have the lowest sales.
* The leading brand in laptop sales is typically HP, Dell, or Lenovo, showing strong customer preference.

**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 Tier 2 / Mid-End Model, with Lenovo and Dell leading sales, particularly in the budget and notebook categories. Gaming laptops, which often come with dedicated graphics cards, are both heavier and more expensive, whereas Notebooks and Ultrabook’s are lighter and ideal for everyday use. The analysis also found that factors like weight and screen size have less impact on price compared to others like processor or RAM. Additionally, Intel processors and integrated graphics are the most common across models, reflecting users’ preference for affordable solutions. Statistical tests confirm that laptops with dedicated GPUs are priced significantly higher than those with integrated graphics.