**Title**: Exploratory Data Analysis of Laptop Features and Pricing

**Name**: Nanthini H

**DA/DS**: May 2025

**Batch Number:**  RP-36

**Online/Offline**: Offline

**Roll Number**: 16525CBRE22

1. **Introduction:**

* This project focuses in performing an Exploratory Data Analysis of a laptop dataset containing specifications like Brand, RAM, Processor Type, Storage, Screen Resolution, Weight and Price.
* The goal is to uncover the hidden patterns and relationship between the price and performance of Laptops.
* By understanding this laptop data through summary statistics and visualizations, we aim to gain valuable insights that can assist customers, retailers and manufacturers in making data-driver decisions.

1. **Aim:**

* The aim of this project is to explore and analyse the various features of laptop and understand how these features affect their prices.
* Using Python and EDA techniques, we will explore the data to find important factors, spot unusual values and observe trends.
* This will help us to gain insights into what makes a laptop more popular and expensive.

1. **Problem Statement:**

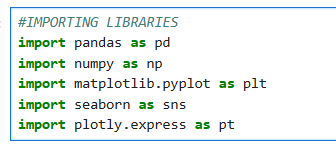
* There are many different laptops in the market, made by many brands with different features. Because of this, it can be confusing for people to choose the best laptop for the right price.
* At the same time companies that sell laptops need to understand which features (like RAM, Storage, Processor, Graphics card and Screen size) affect the price and what customers want.
* The main problem is to find out how these features change the cost of a laptop. This project will use data analysis to look at the information and find patterns and trends.

1. **Project Workflow:**

* This project follows a step-by-step process to explore the laptop data. First, we have to load the dataset and understand the basic details and important columns.
* Next, we need to clean the data by filling missing values and handling the outliers.
* After that, we need to analyse the data using statistics and create different types of charts -like Univariate, Bivariate and Multivariate. In the end, we have to explain what we found, what it means, and give helpful suggestions based on our results.

1. **Data Understanding:**

* We need to import the libraries for EDA process.

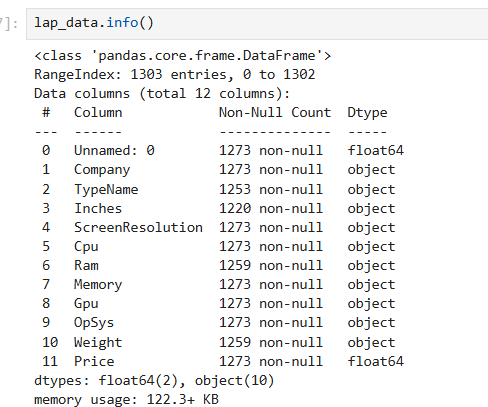


* First, We have to load the dataset 🡪 **laptop\_data = pd.read\_csv("laptopData.csv")**
* To understand the data, we can use below functions,

**laptop\_data.info():**

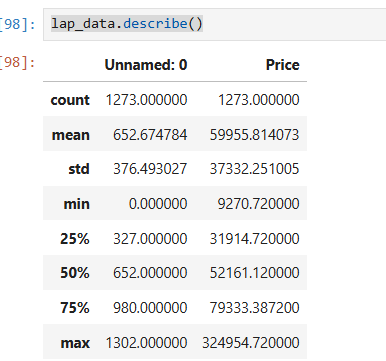
* This gives the information of the dataset, such as column names, total non-null values,

and the datatypes.



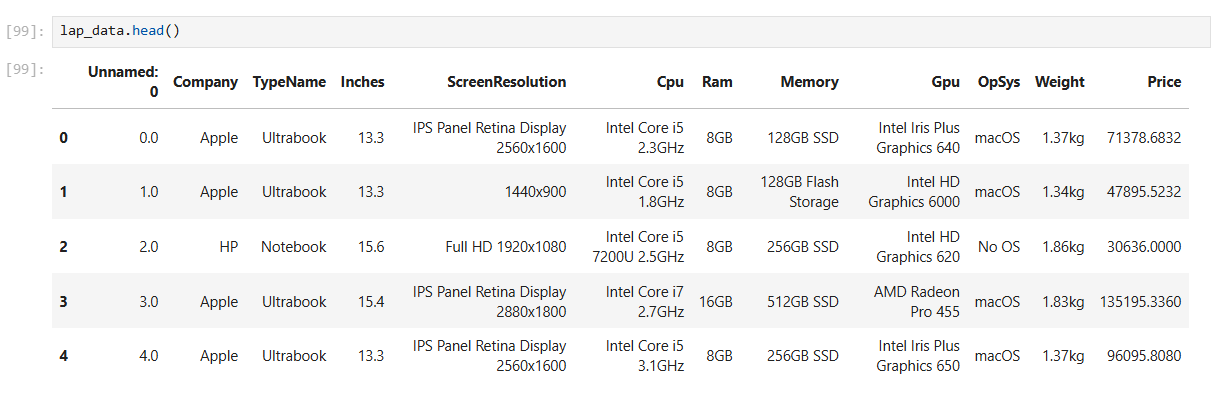
**laptop\_data.describe():**

* To get statistical information of the laptop dataset.



**laptop\_data.head():**

* This will give the first rows from the dataset.

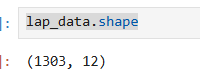


**laptop\_data.tail():**

* This will give the last rows from the dataset.

**lap\_data.shape :**

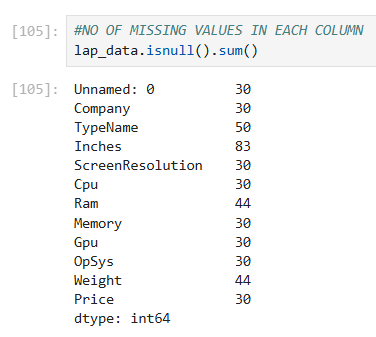
* This will give no. of rows and columns.



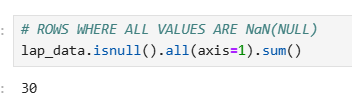
* In this laptop dataset, we have **1303** Rows and **12** Columns.

**isnull().sum:**

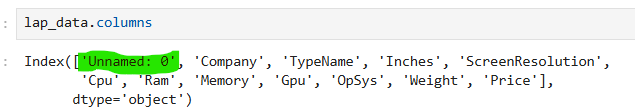
* This will give the null values count in each column.

****

And, we have 30 null rows in the dataset.

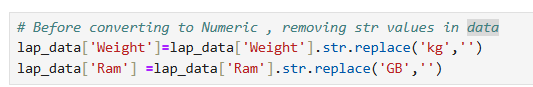
****

1. **Data Cleaning:**

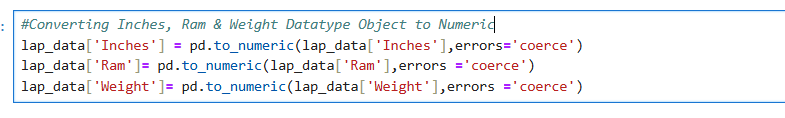
* In this dataset we have one unknown column named Unamed:0 .  
  
* First, I removed that unknown column by using **drop(column =[])** method.



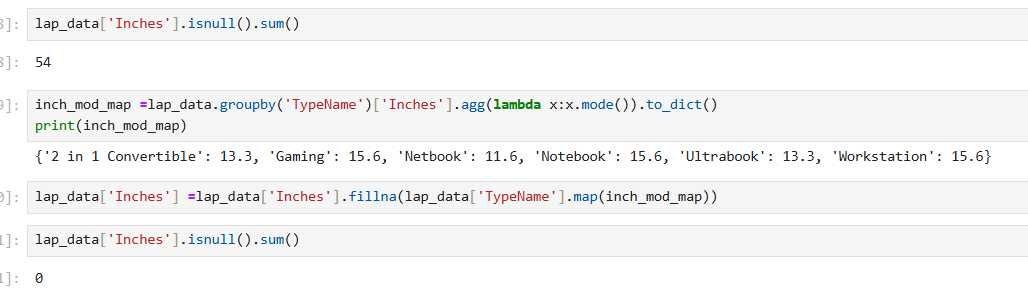
* And we have 30 null rows.
* Before converting the object datatype to numeric I removed string value in the data(like 1.2kg to 1.2, 8GB to 8)



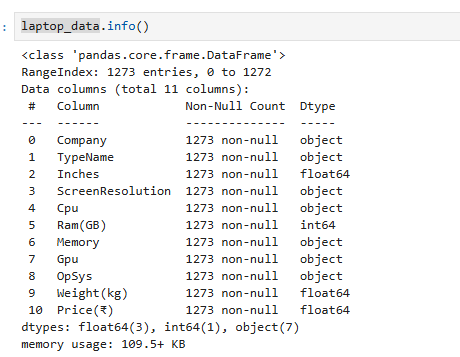
* After that I did **pd.to\_numeric**  function to convert the object datatype into numeric.



* Then, we have null values in Inches, Ram, and Weight columns. I filled the null values with mode value based on Type name.

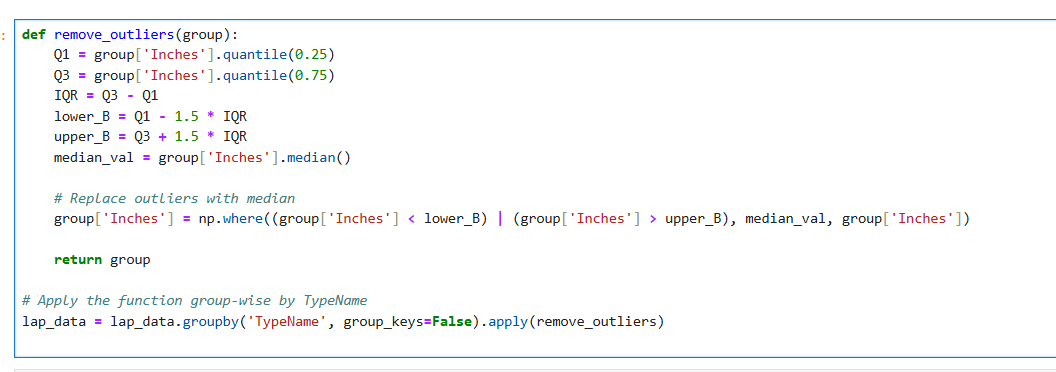


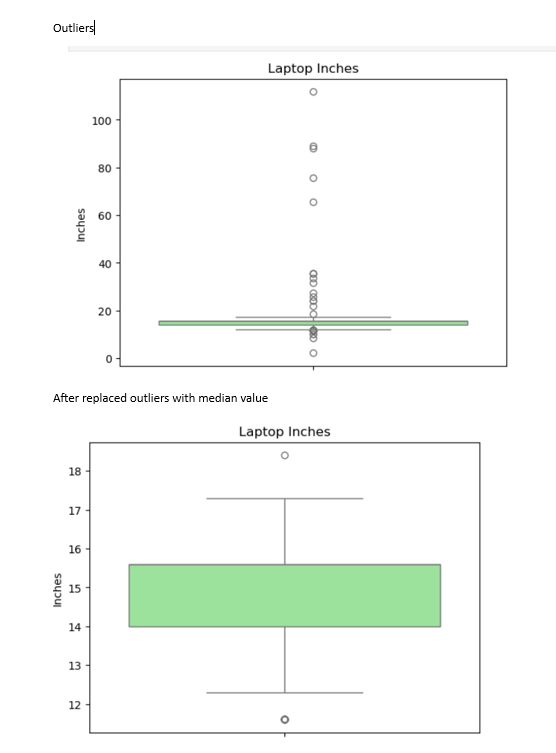
After replaced null values, changed data types



**Handling Outliers:**

* I found outliers in Price, Inches, Weight columns, I replaced mean value for Inches and Weight columns based on type name. I didn’t remove/replace outliers for price column because based on the laptop type the price might vary.





1. **Obtained Derived Metrics:**

* This laptop dataset already had clear and useful columns like RAM, CPU, and Price, which were enough for my analysis. So, I did not create any new metrics.

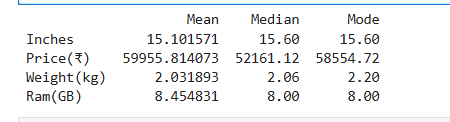
1. **Filtering Data for Analysis:**

* I cleaned the data by removing duplicates and handling missing values. I also converted columns like “Weight” and “Inches” into proper numeric formats.
* Unnecessary column was dropped to keep only the useful ones for analysis. This helped make the dataset clean and ready for better analysis**.**

1. **Statistical Analysis:**

**Descriptive Statistics:**

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



* + - 1. **Test statistics and Hypothesis testing:**

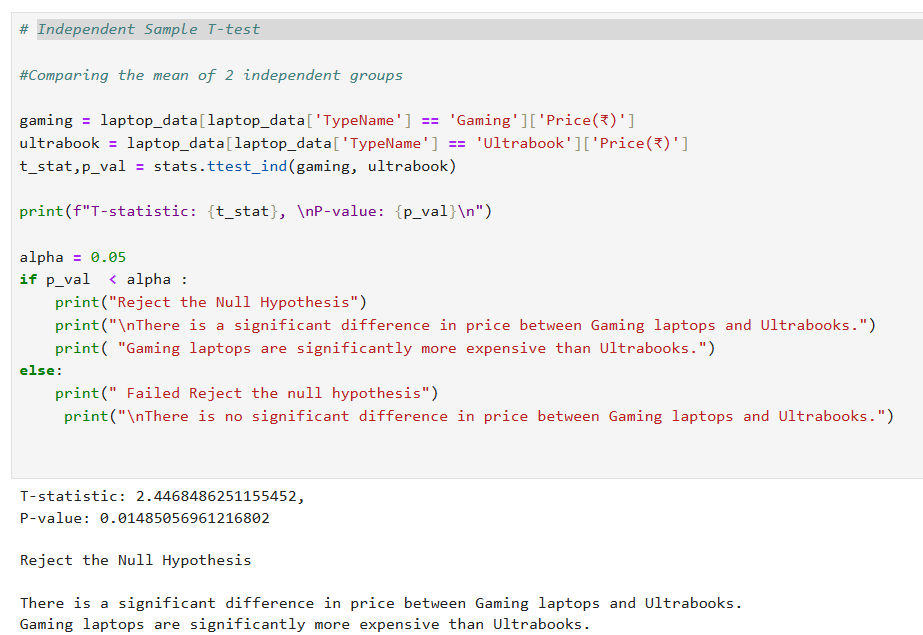
**Hypothesis testing is a method used in statistics to make decisions based on data.  
It helps us check if a claim or assumption about the data is true.**

1. **Assumption:**

* Whether the Gaming laptops and Ultrabooks are same in Price or any difference is there between each other.

**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 **Gaming laptops** and **Ultrabooks**.
* The test gave a **p-value of 0.0148**, which is **less than 0.05**. So, **rejected the null hypothesis**.
* Gaming laptops are **significantly more expensive** than Ultrabooks.

****

1. **Assumption:**

• Whether the RAM size is the same across all laptop brands, or if there is a significant difference.

**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 = 8.15, p < 0.05).
* This means some brands offer more RAM than others, and the average RAM is not the same across all brands.

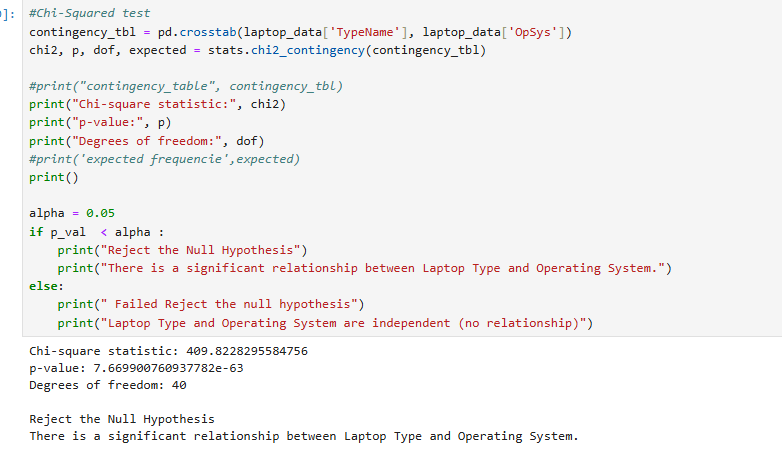


1. **Assumption:**

Whether Laptop Type and Operating System are related or independent.

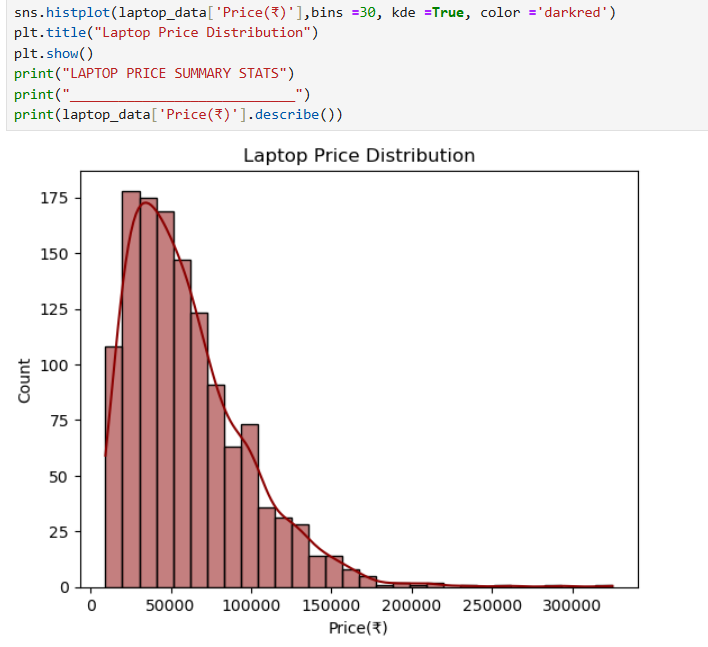
**Chi- Squared Test:**

* **Chi-Squared** Test is used to check if two categorical variables are related to each other.
* This test has been used to check if laptop type is related to the operating system.
* The result was significant (p < 0.05), so reject the null hypothesis.
* This means laptop type and operating system are related certain laptop types tend to come with specific OS.
* For example, Gaming laptops mostly use Windows, while MacBooks use macOS.

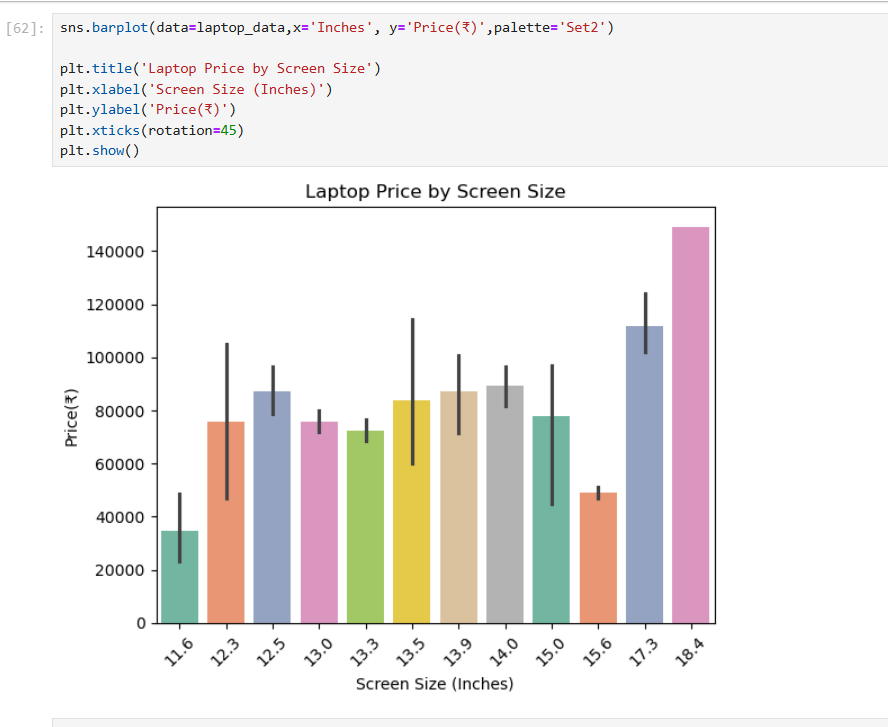
****

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

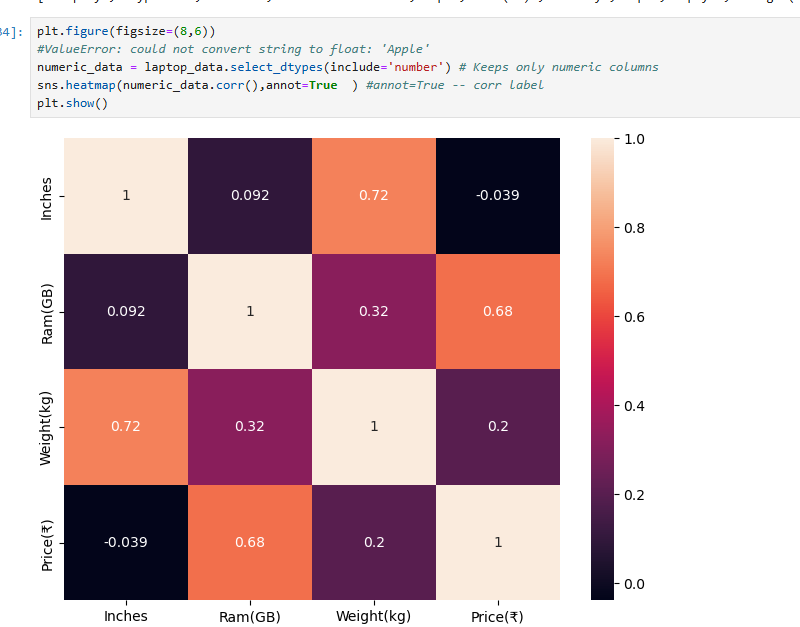
1. **Univariate Analysis:**
   * Univariate Analysis means analyzing one variable at a time. Here I used Histogram plot, Count plot and analysed Price Distribution of laptops, laptop count by company, and Memory usage distribution.



1. **Bivariate Analysis:**
   * Bivariate Analysis means analyzing two variables at the same time to see if there’s a relationship between them. In this project I used Bar Plot, Scatter plot and did for Laptop price with Screen size, Ram with price, Type with OS distribution.

****

1. **Multivariate Analysis:**
   * Multivariate Analysis means analyzing more than 2 variables at the same time to understand the relationship between all of them together.
   * In this project by using Heatmap, 3D scatter Plot, Pair Plot I analyzed the relationship between 'Inches', 'Weight(kg)' , 'Price(₹)', 'Company\_Type', 'Ram(GB)'.

****

1. **Overall Insights:**

**Based on the data analysis, observed the following key insights:**

* **Company vs. Price**Premium brands like Apple, Razer, and MSI have higher average prices compared to brands like HP, Dell, and Lenovo.
* **Laptop Type vs. Price**Gaming laptops are the most expensive type, followed by Ultrabooks. Notebooks and Netbooks are usually more budget-friendly.
* **RAM vs. Performance & Price**Laptops with higher RAM (16GB or more) are generally priced higher and are found mostly in Gaming and high-end laptops.
* **Storage Type**Laptops with SSD are faster and more expensive compared to HDD-based laptops. Some models offer hybrid storage (SSD + HDD) for balance.
* **Weight vs. Portability**Gaming laptops are heavier (~2.5kg or more), while Ultrabooks and Notebooks are lightweight and more portable.
* **Operating System**Most laptops come with Windows OS, followed by Mac OS (Apple) and a few with Linux/No OS (lower price segment).

1. **Conclusion:**

* From this laptop data analysis, we learned that the price of a laptop depends on many factors like brand, RAM, storage type, screen quality, processor, and laptop type. Expensive laptops usually have better features like more RAM, faster SSD storage, high-resolution screens, and powerful processors.
* Brands like Apple and Razer are costlier, while brands like HP and Dell offer more budget-friendly options. Gaming laptops are heavier and more expensive, while Notebooks and Ultrabooks are lighter and better for daily use.
* This analysis helps us to understand what features affect the price and performance of laptops. It can guide buyers in choosing the right laptop and help sellers to understand customer needs better.