Project Information :

Title: Laptop Price Analysis

Name: K. Sri Viswanath

DA/DS: May-2025

Batch Number: RP-36

Online/Offline: Offline

Roll Number: B2025047210

Table of Contents :

Introduction :

Selecting a suitable laptop can be challenging due to the variety of models and specifications available. This project explores a laptop dataset containing details like manufacturer, CPU, RAM size, storage type, and OS. The data is processed and analyzed to uncover which factors mainly influence laptop pricing and specs performance.

Objective :

The goal is to explore the laptop dataset to identify key features impacting cost and usability. We begin with data cleaning and then use straightforward visualizations. This highlights how specifications like RAM, brand, and display size impact the price. The findings can guide buyers in comparing models and may assist in price prediction models in the future.

Problem Statement :

Comparing laptops is difficult due to diverse configurations in terms of brand, CPU, RAM, storage, display, and OS. Companies also require trustworthy insights to stay competitive in the laptop market.

Project Workflow :

The project workflow includes importing the dataset, examining its structure, and performing data cleaning such as handling nulls, outliers, and format issues. After cleaning, new variables are derived, and filtering ensures only valid data remains. Next, descriptive stats and hypothesis checks are done. Finally, Exploratory Data Analysis (EDA) covers univariate, bivariate, and multivariate visual reviews.

Data Familiarization :

The dataset provides specifications like make, model type, RAM, display dimensions, weight, and price. Using Python and pandas, the data was checked for nulls and redundant columns. Incorrect formats were corrected for smooth analysis.

Data Cleaning and Missing Data Handling :

Some fields like display size, model type, and RAM had missing entries. We used forward-fill and backward-fill to address this. For numerical columns, missing values were filled using mean imputation.

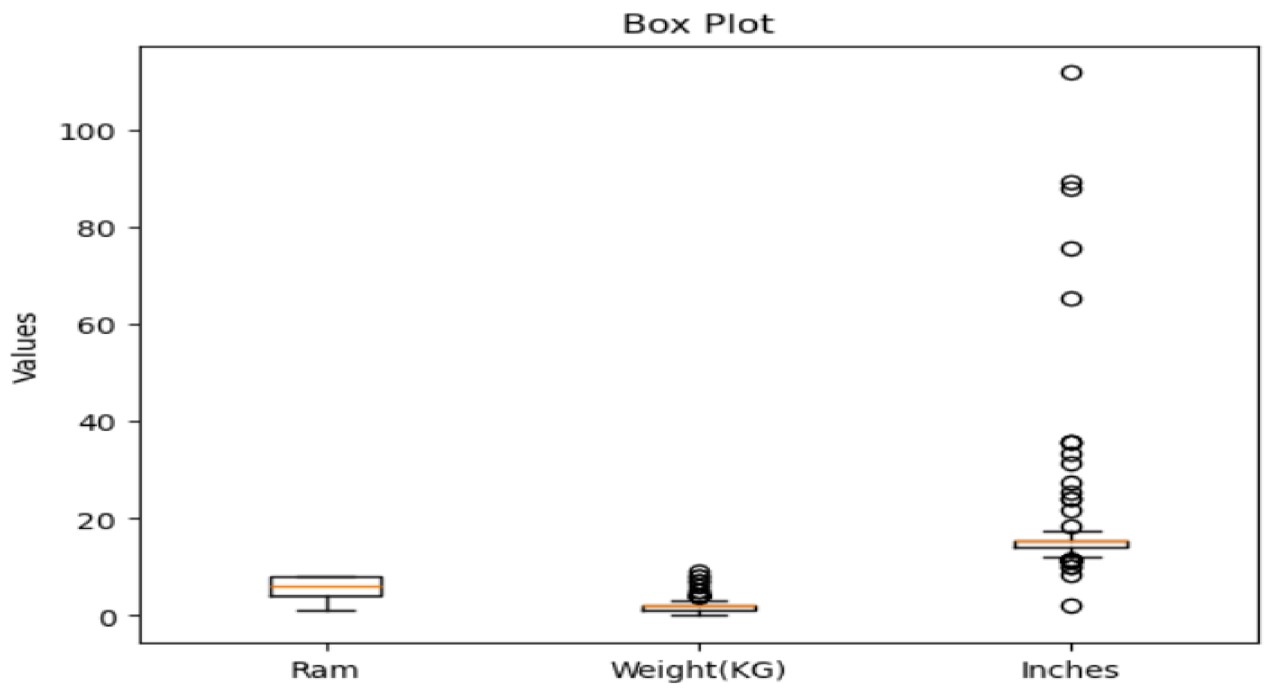
```
6]: mean_inches = vl['Weight(KG)'].mean()
    print(mean_inches)
    vl['Weight(KG)']=vl['Weight(KG)'].fillna(mean_inches)
```

```
1.6653418124006358
```

1. Outlier Detection

Outliers in numerical fields such as price or weight can mislead analysis. We used box plots and value checks to detect and understand anomalies. Unrealistic outliers were flagged but only removed if proven invalid

```
plt.boxplot(vl[['Ram', 'Weight(KG)','Inches']])
plt.xticks([1,2,3,],['Ram', 'Weight(KG)','Inches'])
plt.ylabel('Values')
plt.title('Box Plot')
plt.show()
```




```
q3 = vl['Inches'].quantile(0.75)
q1 = vl['Inches'].quantile(0.25)
iqr = q3 - q1
lower_bound=q1-1.5*iqr
upper_bound=q3+1.5*iqr

med_val=vl['Inches'].median()

vl.loc[(vl['Inches']<lower_bound) | (vl['Inches']>upper_bound),'Inches']=round(med_val,0)
```

```
q3 = vl['Weight(KG)'].quantile(0.75)
q1 = vl['Weight(KG)'].quantile(0.25)
iqr = q3 - q1
lower_bound=q1-1.5*iqr
upper_bound=q3+1.5*iqr

med_val=vl['Weight(KG)'].median()

vl.loc[(vl['Weight(KG)']<lower_bound) | (vl['Weight(KG)']>upper_bound),'Weight(KG)']=round(med_val,0)
```

2.Resolving Data Inconsistencies

Inconsistencies like '8GB' RAM were standardized by splitting values into numeric and unit parts. Empty rows and format mismatches were resolved to keep the dataset uniform.

3.Creating New Variables

To improve insights, extra columns were generated, for example, splitting RAM into numbers and units, or cleaning weight and display size fields. This allowed easier calculations and comparisons.

4.Filtering the Dataset

Data rows with faulty or duplicate entries were removed. This step ensured that the final dataset only included credible and relevant laptop data.

Example: dropna()

```
vl = vl.drop("Unnamed: 0", axis=1)
```

Statistical Summary :

1.Descriptive Statistics

Basic statistics like mean, median, and standard deviation were calculated for numeric data, e.g., price, RAM size, and display size. The describe() function showed skewness in pricing and frequent RAM values at 4GB, 8GB, and 16GB.

2.Hypothesis and Significance Testing

Though limited in scope, tests like ANOVA could verify price differences across brands or laptop categories.

For example, checking if gaming laptops are statistically costlier than standard notebooks.

```
from scipy.stats import f_oneway
vl = vl.dropna(subset=['Company', 'Price'])

groups = vl.groupby('Company')['Price'].apply(list)

stat,p_value = f_oneway(*groups)

print("ANOVA Results:")
print(f"F-statistic: {stat:.2f}")
print(f"P-value: {p_value:.2f}")

if p_value < 0.05:
    print("reject the null hypothesis.")
else:
    print("fail to reject the null hypothesis.")
```

```
ANOVA Results:
F-statistic: 14.07
P-value: 0.00
reject the null hypothesis.
```

Exploratory Data Analysis (EDA)

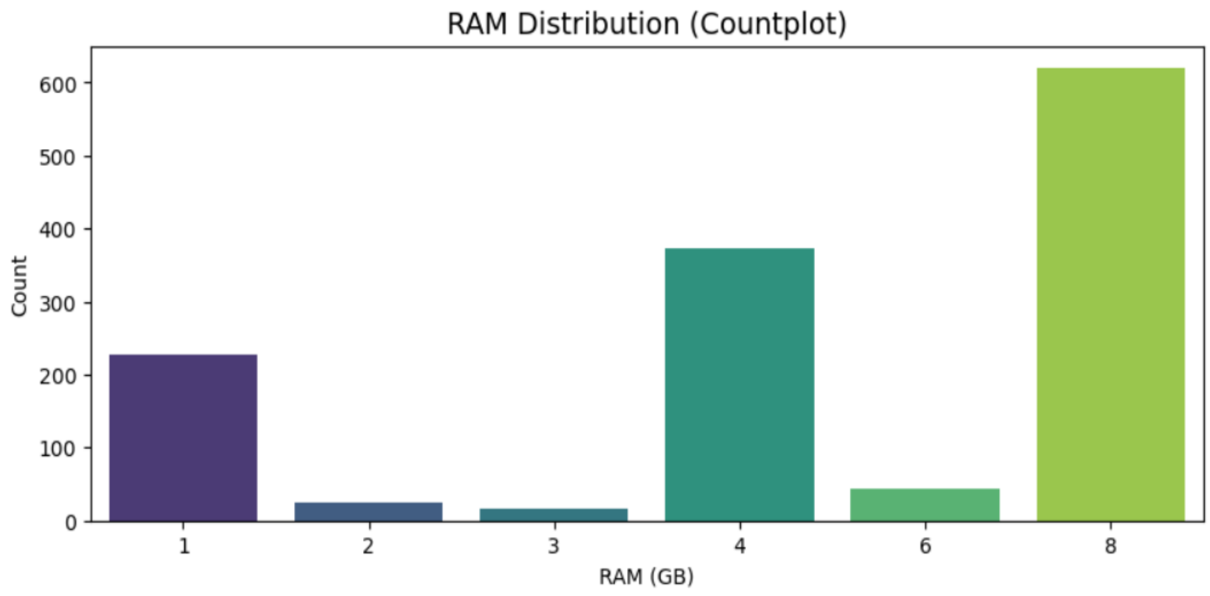
1.Univariate Review

Examined one variable at a time with histograms and count plots.

- Price: The bulk of laptops cost between '30,000'80,000.

- Model Type: 'Notebook' and 'Ultrabook' dominate the dataset.

RAM: 4GB, 8GB, and 16GB configurations are most frequent.



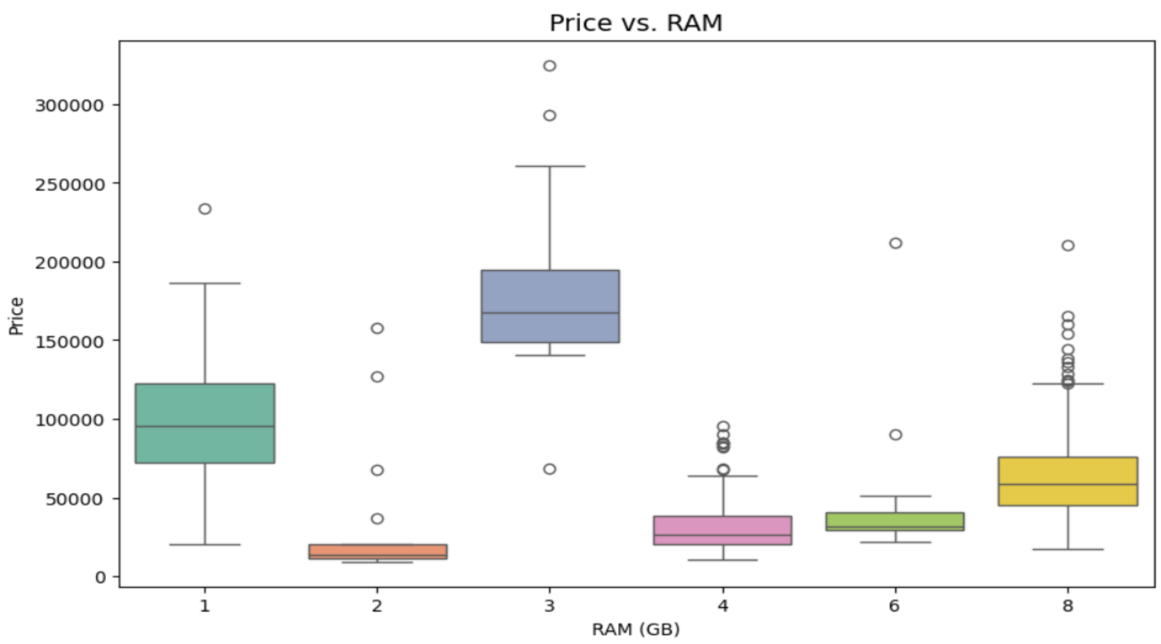
2.Bivariate Review

Looked at relationships between two variables.

- Price vs. RAM: More RAM generally equates to higher price.

- Price vs. Model Type: Gaming models have the highest average price.

- Price vs. Display: Larger screens trend towards higher pricing but vary**.**



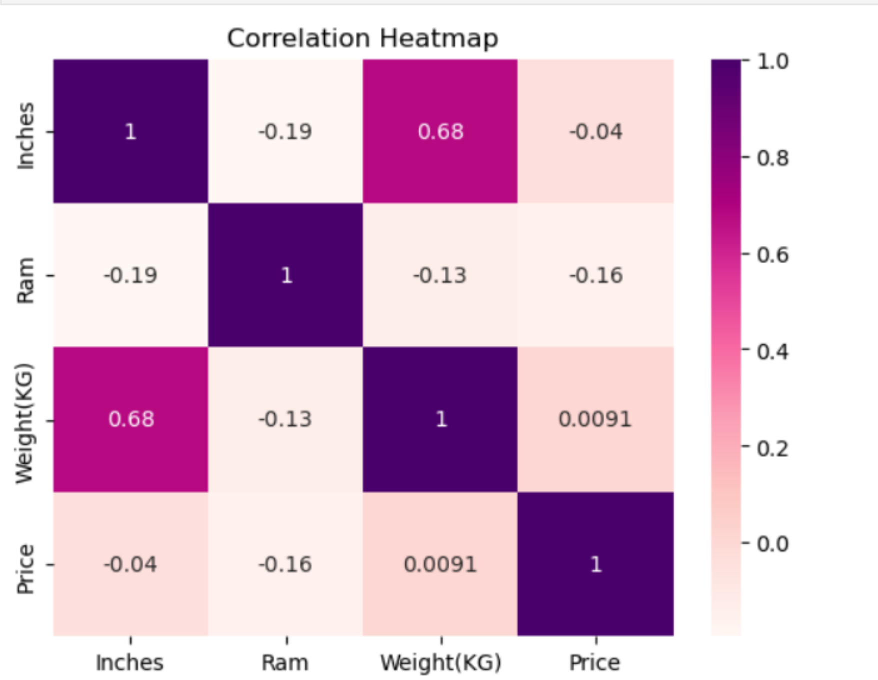
3.Multivariate Review :

Relationships among multiple factors were studied using pair plots and correlation heatmaps.

- Pair Plot: Showed interactions among price, RAM, and display size.

- Heatmap: Displayed correlations ' RAM and display size moderately correlate with price.

- Group Analysis: Brands like Apple and MSI maintain higher price points despite similar specs.



-

Key Takeaways

- Cost Spectrum: Majority of laptops fall in mid-price ranges.

- Model Type: Gaming laptops top the price range; Notebooks are more affordable.

- Specs: Higher RAM and larger screens usually increase cost but plateau after certain thresholds.

- Brand: Premium brands command higher prices.

- Other Factors: SSDs and GPUs could further influence pricing.

- Data Quality: Proper cleaning produced reliable insights**.**

Feature Engineering

- RAM column split into number (e.g., 8) and unit (GB) for analysis.

- Display size (*Inches*) and *Weight* cleaned to numeric format.

- Created new metrics: *Price per Inch* and *RAM per Rupee*.

- Added columns to mark if laptops have SSD storage or dedicated GPU.

- Grouped *Weight* and *Price* into categories for better comparison.

Conclusion

The analysis revealed that RAM, screen size, model type, and brand significantly impact laptop pricing. The project emphasizes how data analytics can help consumers choose wisely and enable sellers to refine pricing strategies. Future work could expand to building prediction models for laptop prices or recommendation systems.

Githuib Link : https://github.com/viswanath2525/pythonproject