

Laptop Price Analysis & Machine Learning

Exploratory Data Analysis |
Regression & Classification | ML
Model

Objective

- Perform exploratory data analysis (EDA) on laptop dataset
- Understand how features like RAM, Storage, and Display influence pricing
- Build machine learning models to predict price and classify price categories
- Visualize key insights using graphs and statistical tests

Dataset Overview

Loaded using pandas

```
# Reading the dataset
#reading data in pandas
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
from wordcloud import WordCloud
from sklearn.metrics import mean_squared_error
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
warnings.filterwarnings('ignore')
sns.set_style('darkgrid')
df=pd.read_csv("C:\Users\hp\OneDrive - subho\Desktop\UNIFIED\Laptop Price Analysis_ML_FA_DA Project\laptop_prices.csv")
print(df)
type(df)
```

Missing values checked

```
# Check for missing values
print(df.isnull().sum())
```

Company	0
Product	0
TypeName	0
Inches	0
Ram	0
OS	0
Weight	0
Price_euros	0
Screen	0
ScreenW	0
ScreenH	0
Touchscreen	0
IPspanel	0
RetinaDisplay	0
CPU_company	0
CPU_freq	0
CPU_model	0
PrimaryStorage	0
SecondaryStorage	0
PrimaryStorageType	0
SecondaryStorageType	0
GPU_company	0
GPU_model	0

Duplicates removed

```
# Checking duplicates rows and Removing duplicate rows if any
print(df.duplicated().value_counts())
df.drop_duplicates(inplace = True)
print('Shape After deleting duplicate values:', df.shape)
```

```
False    1275
Name: count, dtype: int64
Shape After deleting duplicate values: (1275, 23)
```

Number of rows and columns present

```
#number of rows and columns present
df.shape
```

```
(1275, 23)
```

head(5)

```
#top 5 row
df.head()
```

	Company	Product	TypeName	Inches	Ram	OS	Weight	Price_euros	Screen	ScreenW	...	RetinaDisplay	CPU_company	CPU_freq	CPU_model	
0	Apple	MacBook Pro	Ultrabook	13.3	8	macOS	1.37	1339.99	Standard	2560	...	Yes	Intel	2.3	Core i5	
1	Apple	MacBook Air	Ultrabook	13.3	8	macOS	1.34	899.94	Standard	1440	...	No	Intel	1.8	Core i5	
2	HP	250 G8	Notebook	15.6	8	No OS	1.88	575.00	Full HD	1620	...	No	Intel	2.5	Core i5 7200U	
3	Apple	MacBook Pro	Ultrabook	15.4	16	macOS	1.83	2537.45	Standard	2880	...	Yes	Intel	2.7	Core i7	
4	Apple	MacBook Pro	Ultrabook	13.3	8	macOS	1.37	1883.65	Standard	2560	...	Yes	Intel	3.1	Core i5	

5 rows x 23 columns

Column Classification

Categorical Columns: Company, Product, TypeName, Ram, OS, Screen, ScreenW, ScreenH, Touchscreen, IPSpanel, RetinaDisplay, CPU_company, CPU_model, PrimaryStorage, SecondaryStorage, PrimaryStorageType, SecondaryStorageType, GPU_company

Continuous Columns: Inches, Weight, Price_euros, CPU_freq

```
categorical_columns = []
continuous_columns = []

for column in df.columns[1:-1]:
    dtype = df[column].dtype
    unique_values = df[column].nunique()

    if dtype == 'object' or unique_values < 15:
        categorical_columns.append(column)
    elif pd.api.types.is_numeric_dtype(df[column]):
        continuous_columns.append(column)
```

categorical_columns

```
['Company',
 'Product',
 'TypeName',
 'Ram',
 'OS',
 'Screen',
 'ScreenW',
 'ScreenH',
 'Touchscreen',
 'IPSPanel',
 'RetinaDisplay',
 'CPU_company',
 'CPU_model',
 'PrimaryStorage',
 'SecondaryStorage',
 'PrimaryStorageType',
 'SecondaryStorageType',
 'GPU_company']
```

continuous_columns

```
['Inches', 'Weight', 'Price_euros', 'CPU_freq']
```

Average Laptop Price (€) by Company

- Observation: Razer laptops have the highest average price (€3346.14), indicating a premium positioning. In contrast, brands like Mediacom, Vero, and Chuwi offer the most budget-friendly options, averaging below €320.

Distribution of laptop company by average price in euros:

Company

Razer 3346.14

LG 2099.00

MSI 1728.91

Google 1677.67

Microsoft 1612.31

Apple 1564.20

Huawei 1424.00

Samsung 1413.44

Toshiba 1267.81

Dell 1199.23

Xiaomi 1133.46

Asus 1123.83

Lenovo 1093.86

HP 1080.31

Fujitsu 729.00

Acer 633.46

Chuwi 314.30

Mediacom 295.00

Vero 217.42

Name: Price_euros, dtype: float64

Average Price (€) by Laptop Type

- Observation: Workstations and Gaming laptops command the highest average prices, reflecting their high-performance specifications. In contrast, Netbooks and Notebooks are the most affordable, catering to basic usage needs.

```
Distribution of type by average price in euros:  
TypeName  
Workstation          2280.36  
Gaming               1731.38  
Ultrabook            1556.68  
2 in 1 Convertible   1289.71  
Notebook              788.74  
Netbook              673.38  
Name: Price_euros, dtype: float64
```

Combined Distribution: Type & Brand vs. Average Price (€)

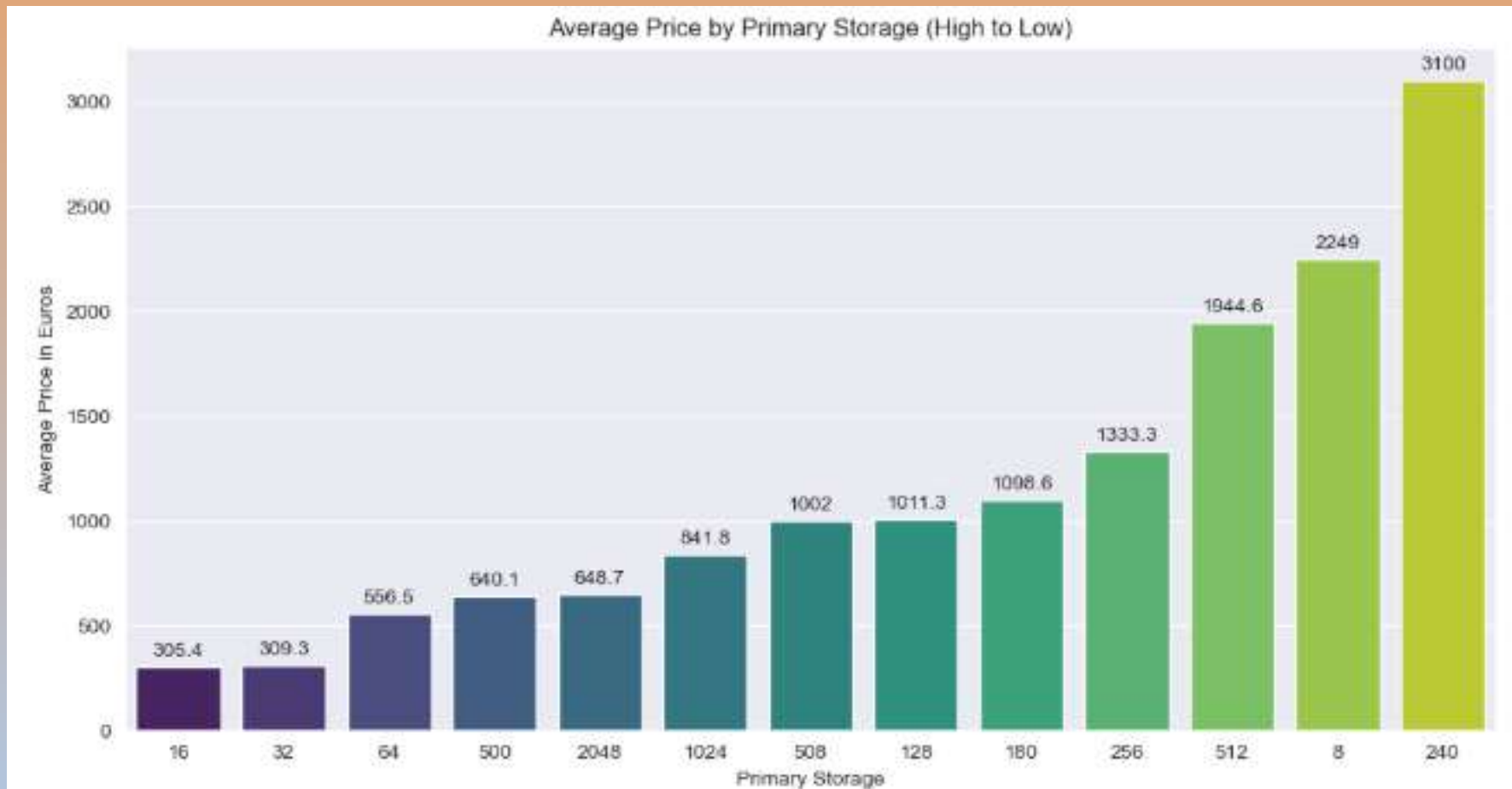
- Observation: 2 in 1 Convertibles and Netbooks offer more affordable options, while Gaming laptops, especially from Razer, command significantly higher prices. Ultrabooks and Workstations consistently show premium pricing across brands, indicating their high-end features.

Distribution of TypeName and Company by average price in euros:

	TypeName	Company	Price_euros
5	2 in 1 Convertible	Mediacom	299.00
0	2 in 1 Convertible	Acer	698.75
1	2 in 1 Convertible	Asus	976.39
2	2 in 1 Convertible	Dell	1131.67
6	2 in 1 Convertible	Samsung	1229.00
3	2 in 1 Convertible	HP	1506.81
4	2 in 1 Convertible	Lenovo	1529.02
11	Gaming	Lenovo	1328.97
10	Gaming	HP	1424.08
7	Gaming	Acer	1468.12
8	Gaming	Asus	1720.01
12	Gaming	MSI	1728.91
9	Gaming	Dell	1918.39
13	Gaming	Razer	4119.00
15	Netbook	Asus	266.20
19	Netbook	Samsung	269.00
14	Netbook	Acer	306.00
16	Netbook	Dell	519.50
18	Netbook	Lenovo	735.75
17	Netbook	HP	1234.57
30	Notebook	Vero	217.42
27	Notebook	Mediacom	294.33
22	Notebook	Chuwi	314.30
20	Notebook	Acer	546.75
21	Notebook	Asus	632.34
24	Notebook	Fujitsu	729.00
26	Notebook	Lenovo	794.33
25	Notebook	HP	820.17
23	Notebook	Dell	883.15
29	Notebook	Toshiba	1109.00
31	Notebook	Xiaomi	1299.47
28	Notebook	Samsung	1699.00
32	Ultrabook	Acer	890.50
45	Ultrabook	Xiaomi	967.45
34	Ultrabook	Asus	1352.60
42	Ultrabook	Razer	1414.00
38	Ultrabook	Huawei	1424.00
35	Ultrabook	Dell	1477.92
37	Ultrabook	HP	1537.14

Primary Storage Option Offering the Best Value

- Chart: Column chart of Average Laptop Price by Primary Storage Size (in Euros)
- Observation: Primary storage options with lower capacities (16–64 GB) offer the best value, having significantly lower average prices. Conversely, higher storage options like 240 GB and 512 GB are priced much higher, with 240 GB showing an unusually steep price spike.



Impact of Retina Display on Laptop Pricing

- Observation: Laptops with Retina Displays have a significantly higher average price (€1657.85) compared to non-Retina models (€1127.90). Statistical testing confirms this difference is significant ($t = 3.569$, $p = 0.0024$).

T-Statistic: 3.569

P-Value: 0.0024

Mean price for retina display laptop: 1657.85

Mean price for non retina display laptop: 1127.9

Reject the null hypothesis: Laptops with Retina Display have significantly different prices. retina display laptop tend to have a higher overall price.

Top GPU Brands in High-End Laptops

- Observation: Intel and Nvidia dominate the GPU market in high-end laptops, together accounting for over 94% of units. AMD holds a smaller share at just over 5%.

GPU Companies Dominating in High-End Laptops:

	Count	Percentage
GPU_company		
Intel	326	51.34
Nvidia	275	43.31
AMD	34	5.35

CPU Brand Distribution and Its Impact on Price

- Observation: Intel dominates the CPU market with over 95% share and commands the highest average laptop price (€1163.73). AMD, while more affordable (€560.99), holds a much smaller market share.

```
Distribution of CPU_company (Counts):  
Count Percentage  
CPU_company  
Intel      1214      95.22  
AMD         60       4.71  
Samsung     1       0.08  
Average Price by CPU_company in euros:  
CPU_company  
Intel      1163.73  
Samsung     659.00  
AMD         560.99  
Name: Price_euros, dtype: float64
```

Correlation Between Screen Size and Price

- Chart: Scatter plot with regression line showing correlation between Screen Size and Price
- Observation : There is a weak positive correlation between screen size and price — larger screens tend to cost slightly more, but screen size alone doesn't strongly explain price variation.

Regression Equation: $\text{Price} = 32.65 \times \text{Inches} + 644.43$
R-squared: 0.084
p-value: 0.0174



The correlation is weak - Laptop screen Inches does not explain much of the variation in Price.
The relationship between screen size and price is statistically significant — there is sufficient evidence to conclude the relationship is real.

Relationship Between Laptop Weight and Cost

- Chart: Scatter plot with regression line showing correlation between Laptop Weight and Price
- Observation : There is a weak positive correlation between laptop weight and price — heavier laptops tend to cost slightly more, but weight alone does not significantly explain price variation.

Regression Equation: $\text{Price} = 221.87 \times \text{Weight} + 682.23$
R-squared: 0.045
p-value: 0.8



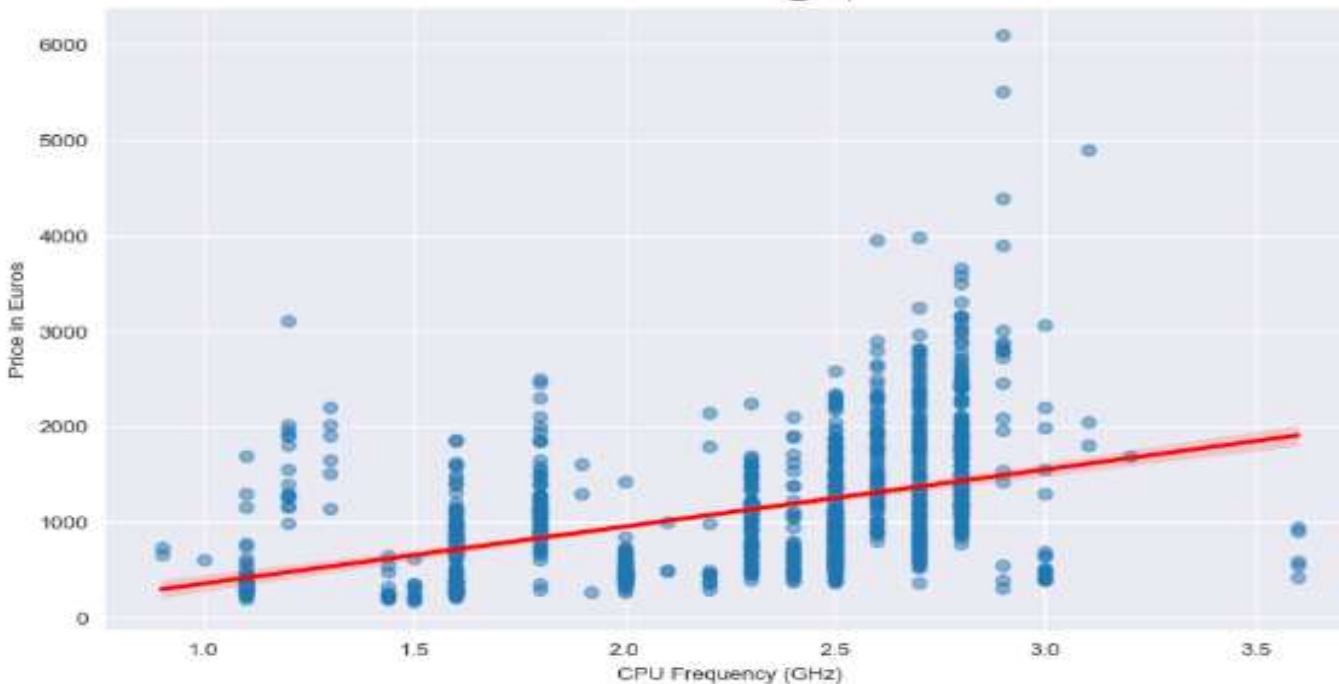
The correlation is weak - Weight does not explain much of the variation in Price.
The relationship between weight and price is highly significant - the relationship is very unlikely to be due to chance.

How CPU Frequency Affects Laptop Price

- Chart: Scatter plot with regression line showing correlation between CPU Frequency and Price
- Observation: There is a moderate positive correlation between CPU frequency (GHz) and laptop price. As CPU frequency increases, price tends to rise, suggesting that higher-performing processors are associated with more expensive laptops. However, CPU frequency alone does not fully explain price variation.

Regression Equation: $\text{Price} = 596.44 \times \text{CPU_freq} + -238.63$
R-squared: 0.184
p-value: 0.0

Correlation between CPU_freq and Price



The correlation is weak - CPU_freq does not explain much of the variation in Price.
The relationship between CPU_freq and price is highly significant - the relationship is very unlikely to be due to chance.

Effect of RAM and Storage on Price

- Observation: RAM has a strong positive impact on laptop price, while Primary Storage has a slight negative effect. The model explains about 56.7% of the price variation, indicating a moderate correlation.

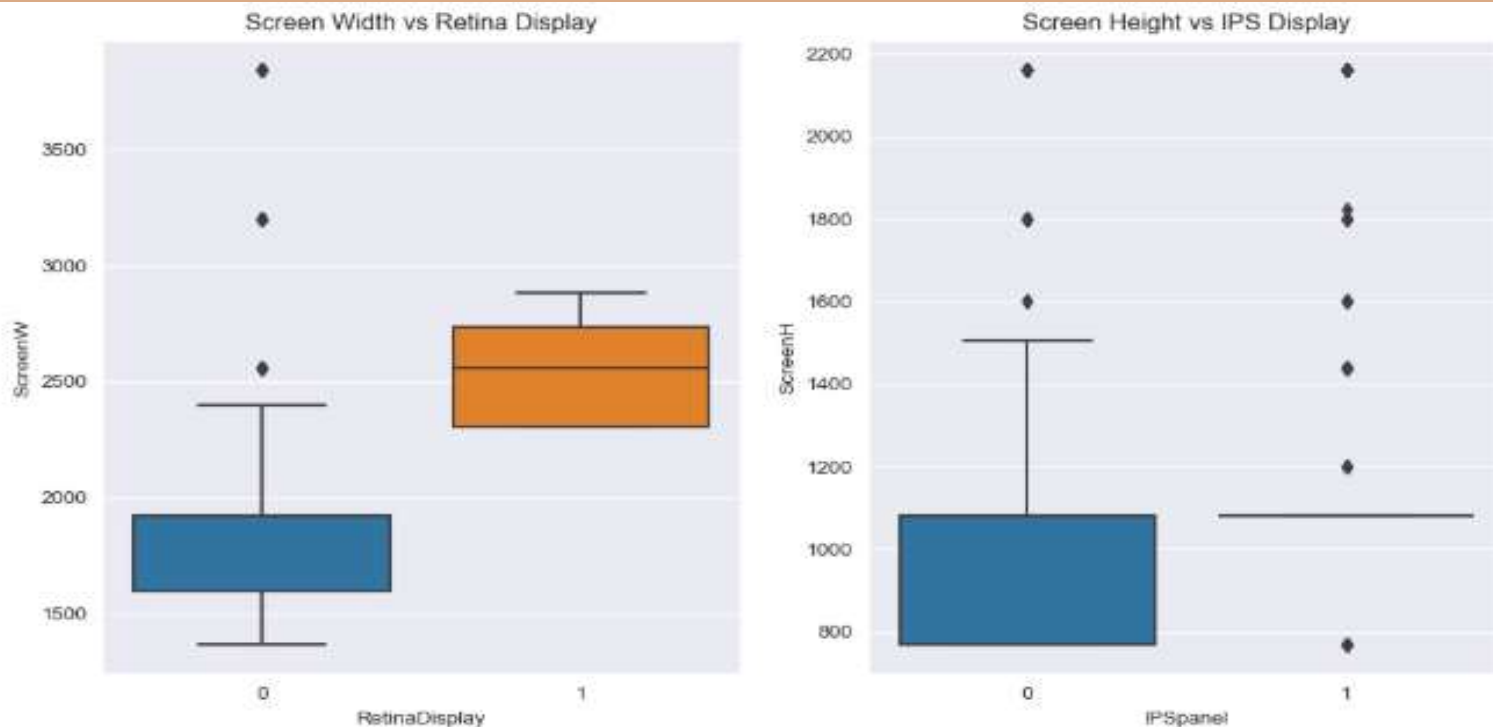
Regression Equation: $\text{Price} = 102.05 \times \text{RAM} - 0.26 \times \text{PrimaryStorage} + 389.65$

R-squared: 0.567

The correlation is moderate – RAM and PrimaryStorage together explain a moderate amount of the variation in price.

High-Resolution Screens with Retina and IPS Displays

- Chart: The box plots and correlation analysis reveal insights into how Retina and IPS display technologies relate to screen resolution dimensions.
- Observation : Laptops with Retina and IPS displays generally have higher screen resolutions, as shown by wider screen dimensions in box plots. However, the correlation between screen width and height is not consistently very high, indicating variability in aspect ratios.



Screen Resolution Correlation for Retina Laptops: nan

Screen Resolution Correlation for IPS Laptops: nan

The correlation between Screen Width and Screen Height is not very high for both Retina and IPS displays.

Top OS & Type Combinations in Premium Laptops

- Observation : Premium laptops are mostly Ultrabooks, Workstations, and Gaming types paired with Windows 7/10 or macOS. Workstations with Windows and Ultrabooks with macOS or Windows have the highest average prices.

Combinations of OS and TypeName that appear most in high-priced laptops:

	OS	TypeName	Price_euros
19	Windows 7	Workstation	2534.55
13	Windows 10	Workstation	2183.52
15	Windows 7	2 in 1 Convertible	2050.38
9	Windows 10	Gaming	1860.08
20	macOS	Ultrabook	1820.52
14	Windows 10 5	Ultrabook	1736.97
18	Windows 7	Ultrabook	1715.88
8	Windows 10	2 in 1 Convertible	1678.94
1	Chrome OS	Ultrabook	1677.67
12	Windows 10	Ultrabook	1657.96
10	Windows 10	Netbook	1619.80
16	Windows 7	Netbook	1599.00
17	Windows 7	Notebook	1535.56
11	Windows 10	Notebook	1365.33
4	Mac OS X	Ultrabook	1306.28
6	No OS	Notebook	1299.47
0	Chrome OS	2 in 1 Convertible	1159.00
5	No OS	Gaming	1146.20
3	Linux	Ultrabook	1099.00
2	Linux	Notebook	1079.00
7	No OS	Ultrabook	999.90

Price Impact of Touchscreen Laptops

- Observation : Touchscreen laptops have a significantly higher average price (€1453.12) compared to non-touchscreen models (€1079.94). Statistical evidence ($t = 6.149$, $p = 0.0$) confirms this price difference is highly significant.

T-statistic: 6.149

P-value: 0.0

Mean price for touchscreen laptop: 1453.12

Mean price for non touchscreen laptop: 1079.94

Reject the null hypothesis: Touchscreen laptops have significantly different prices.
touchscreen laptop tend to have a higher overall price.

Best RAM and Storage Combinations for Value per Euro

- Observation : Laptops with 8GB RAM and 2048GB (2TB) storage offer the best value per euro (value score: 3.81), followed by 6GB + 2TB (3.23) and 12GB + 2TB (2.37). Higher RAM and storage do not always guarantee better value — balance matters more than raw specs for price efficiency.

Average value score for combinations of RAM and Primary Storage:

	Ram	PrimaryStorage	value_score
27	8	2048	3.81
17	6	2048	3.23
32	12	2048	2.37
12	4	1024	2.24
37	16	2048	2.08
16	6	1024	1.89
26	8	1024	1.43
3	2	500	1.32
31	12	1024	1.17
10	4	500	1.02
36	16	1024	0.65
11	4	500	0.51
24	8	500	0.49
30	12	512	0.42
15	6	256	0.41
9	4	256	0.39
25	8	512	0.33
2	2	64	0.29
35	16	512	0.29
42	32	1024	0.27
43	64	1024	0.27
23	8	256	0.23
7	4	128	0.23
38	24	256	0.22
8	4	180	0.21
39	24	512	0.20
6	4	64	0.20
14	6	128	0.19
41	32	512	0.17
29	12	256	0.16
13	6	64	0.16
21	8	180	0.16
34	16	256	0.15
1	2	32	0.14
20	8	128	0.14
40	32	256	0.13
5	4	32	0.12
28	12	128	0.11

Machine Learning – Feature engineering

```
# Machine Learning

# Predicting the app rating based on the features in the dataset

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

# Label Encoding categorical features
label_encoder = LabelEncoder()

df['Company'] = label_encoder.fit_transform(df['Company'])
df['TypeNane'] = label_encoder.fit_transform(df['TypeNane'])
df['OS'] = label_encoder.fit_transform(df['OS'])
df['Touchscreen'] = label_encoder.fit_transform(df['Touchscreen'])
df['IPspanel'] = label_encoder.fit_transform(df['IPspanel'])
df['RetinaDisplay'] = label_encoder.fit_transform(df['RetinaDisplay'])
df['CPU_company'] = label_encoder.fit_transform(df['CPU_company'])
df['PrimaryStorageType'] = label_encoder.fit_transform(df['PrimaryStorageType'])
df['SecondaryStorageType'] = label_encoder.fit_transform(df['SecondaryStorageType'])
df['GPU_company'] = label_encoder.fit_transform(df['GPU_company'])

# Define features and target variable
X = df[['Company', 'TypeNane', 'Inches', 'Ram', 'OS', 'Weight', 'ScreenW', 'ScreenH',
        'Touchscreen', 'IPspanel', 'RetinaDisplay', 'CPU_company', 'CPU_freq',
        'PrimaryStorage', 'SecondaryStorage', 'PrimaryStorageType', 'SecondaryStorageType', 'GPU_company']]
y = df['Price_euros']

# Handle missing values
X.fillna(X.median(), inplace=True)
y.fillna(y.median(), inplace=True)

# Split the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Random Forest Regression: Laptop Price Prediction

- Observation: Random Forest Regression achieved high accuracy with $R^2 = 0.85$, explaining 85% of price variation. Mean Squared Error of ~78,798 indicates a reasonably good fit for predicting laptop prices.

```
#Random Forest Model (regression model)

from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score

# Train the model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

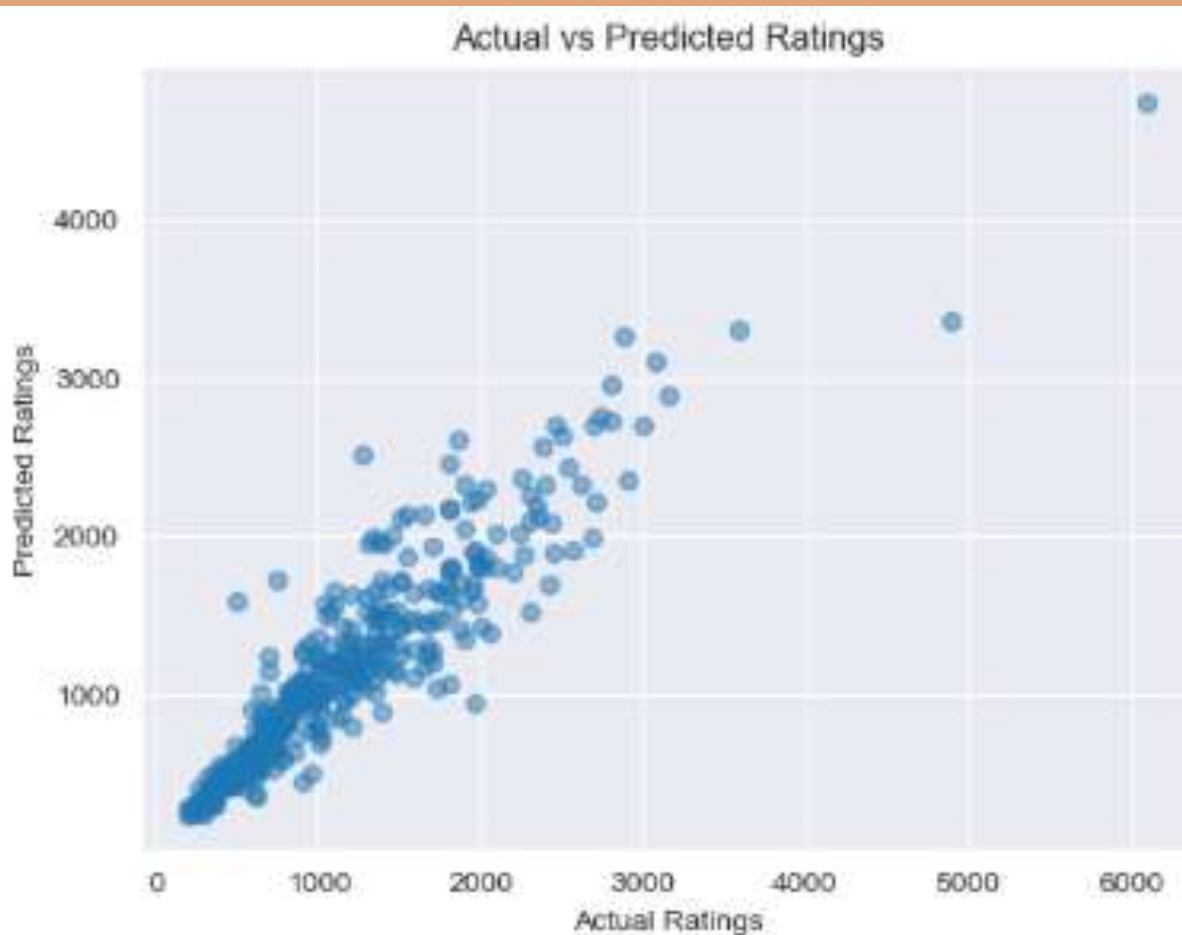
# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print('Mean Squared Error:', mse)
print('R-squared:', r2)
```

```
Mean Squared Error: 78798.24166806889
R-squared: 0.8470537166870762
```

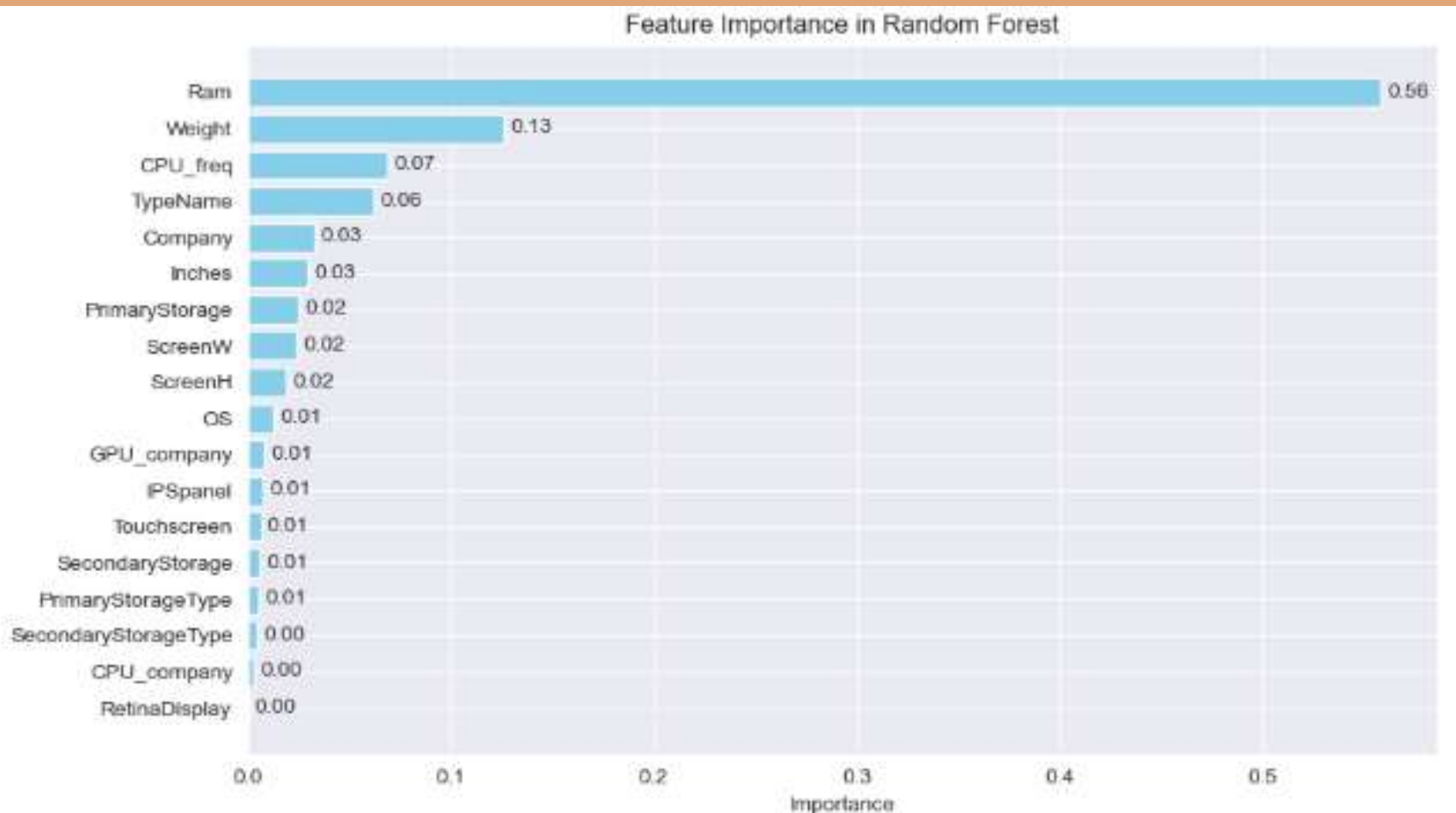
Actual vs. Predicted Laptop Prices

- Observation: The scatter plot shows how closely the predicted prices match the actual laptop prices. Ideally, points should lie along the diagonal, indicating accurate predictions.



Most Influential Features in Random Forest Model

- Observation: RAM is the most influential feature in the model with an importance of 56%, followed by Weight (13%) and CPU Frequency (7%). These features play a key role in predicting the target variable accurately.



Random Forest Classification: Price Category Prediction

- Observation: Random Forest Classification predicted laptop price categories with 82.25% accuracy and 0.82 macro F1-score. Performance was balanced across all three price classes with minimal misclassification.

```
Class distribution: {0: 424, 1: 425, 2: 426}
```

```
Confusion Matrix:
```

```
[[114   0  13]
 [  0 108  20]
 [ 24  11  93]]
```

```
Accuracy: 82.25%
```

```
Balanced Accuracy: 0.8226500984251969
```

```
F1 Score (Macro): 0.8223849167276671
```

```
Classification Report:
```

	precision	recall	f1-score	support
0	0.83	0.90	0.86	127
1	0.91	0.84	0.87	128
2	0.74	0.73	0.73	128
accuracy			0.82	383
macro avg	0.82	0.82	0.82	383
weighted avg	0.82	0.82	0.82	383

Final Observations & Storyline (left)

- Razer laptops average €3346.14, positioning them at the top of the premium market. In contrast, Mediacom, Vero, and Chuwi offer entry-level models under €320. Gaming and Workstation laptops are the priciest types, while Netbooks and Notebooks remain the most affordable.
- RAM is the strongest price driver, with a 56% feature importance in prediction models. Laptops with 8GB RAM + 2TB storage offer the best value (score: 3.81), outperforming higher-spec options. Surprisingly, 240 GB storage shows a price spike despite mid-tier capacity.
- Retina displays boost average prices to €1657.85 vs. €1127.90 for non-Retina models ($p = 0.0024$). Touchscreen laptops also cost more on average (€1453.12 vs. €1079.94; $p = 0.0$). These premium features drive up manufacturing costs and consumer price points.
- Intel holds over 95% CPU share and dominates GPU use alongside Nvidia (combined 94%+). AMD trails in both markets, reflected in its lower average laptop price of €560.99. This shows a clear divide between high-performance and cost-effective computing segments.
- Random Forest Regression achieved $R^2 = 0.85$, with RAM, CPU frequency, and weight as top predictors. Screen size and weight show weak price correlation, highlighting their limited impact. Overall, balanced mid-range specs deliver the best value in a price-sensitive market.