Capstone Project:

Laptop Price Prediction with Machine Learning

- ML project using Python + Power BI

By Glanet Castelino

ID_S12125

Contents

Project Overview

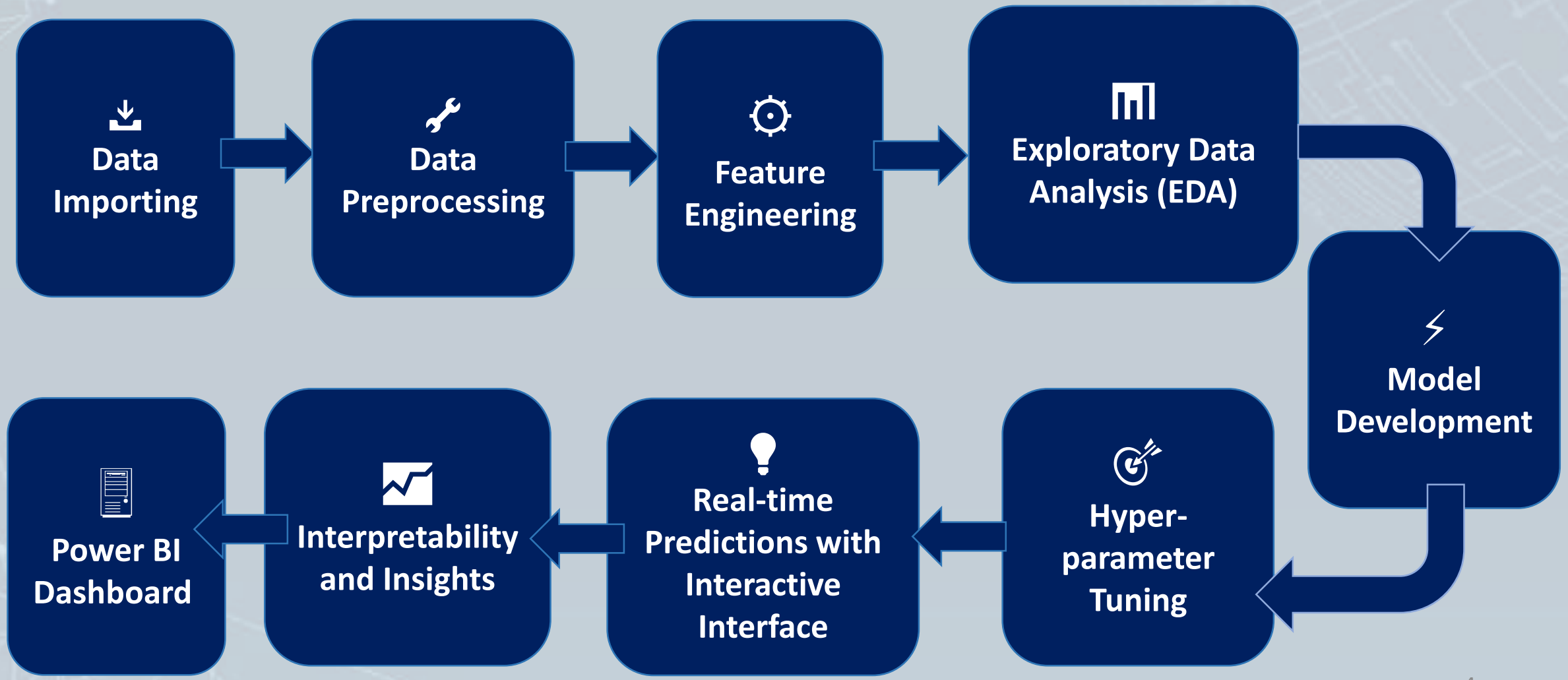
The Project Overview

- To develop a robust machine learning model that predicts laptop prices accurately for our client SmartTech Co.
- As the market for laptops continues to expand with a myriad of brands and specifications, , a precise pricing model is vital for both **consumers** and **manufacturers** to make informed decisions.

Business Objectives

- **Accurate Pricing:** Develop a model that can accurately predict laptop prices based on various features, helping our clients stay competitive in the market.
- **Market Positioning:** Understand how different features contribute to pricing, enabling SmartTech Co. to strategically position its laptops in the market.
- **Brand Influence:** Assess the impact of brand reputation on pricing, providing insights into brand perception and market demand.

Project Architecture



Data Preparation & ML Modeling:

Data Overview:

- 1303 entries, 13 columns

Target Feature (Dependent Variable):

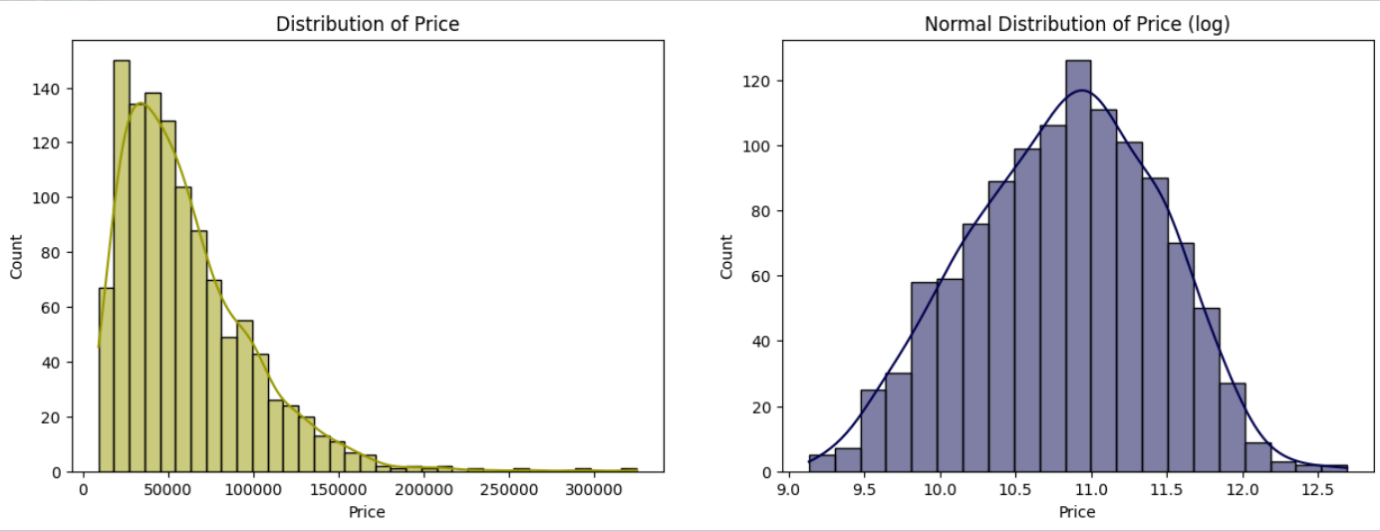
- **Price:** Laptop selling price, to be predicted by the ML model.

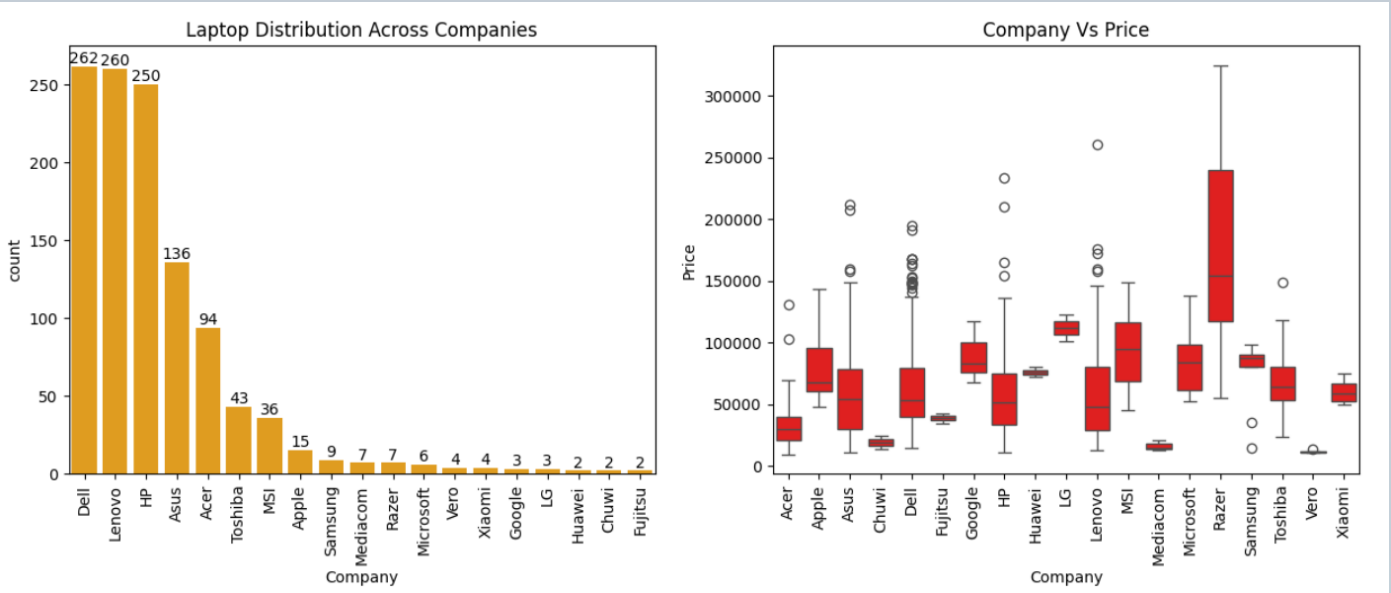
Input Features (Independent Variables):

- **Brand/Company:** Laptop manufacturer (Dell, HP, Apple, etc.)
- **TypeName:** Laptop category (Gaming, Ultrabook, 2-in-1, Notebook)
- **CPU:** Processor brand, series, and generation
- **RAM:** Memory in GB
- **GPU:** Dedicated graphics card
- **Storage:** Type and capacity (HDD/SSD/Hybrid)
- **Inches: & Resolution:** Screen size and quality
- **Weight:** Portability factor
- **Operating System:** Windows, macOS, Linux
- **Features/Connectivity:** Touchscreen, IPS display

	Unnamed: 0.1	Unnamed: 0	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price
0	0	0.0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	71378.6832
1	1	1.0	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	47895.5232
2	2	2.0	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	30636.0000
3	3	3.0	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	135195.3360
4	4	4.0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	96095.8080

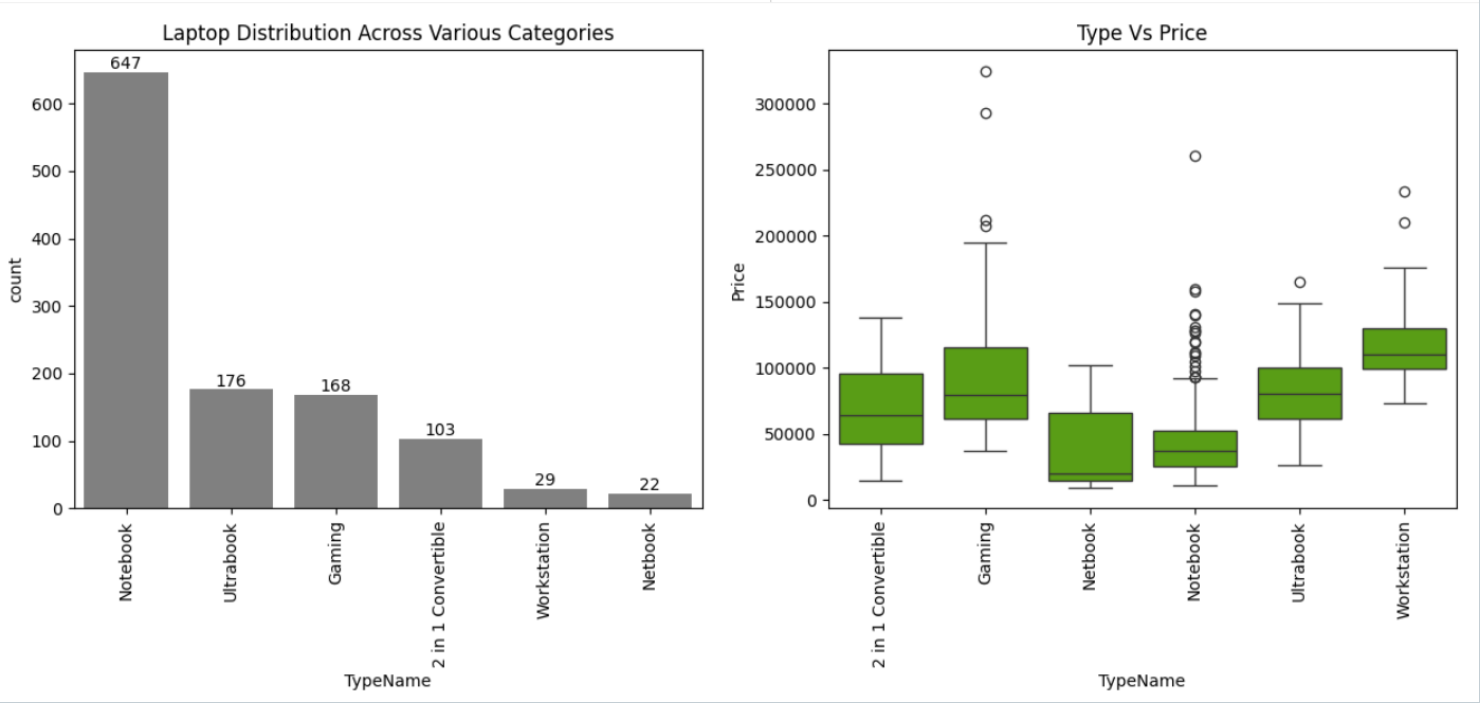
Exploratory Data Analysis (EDA)

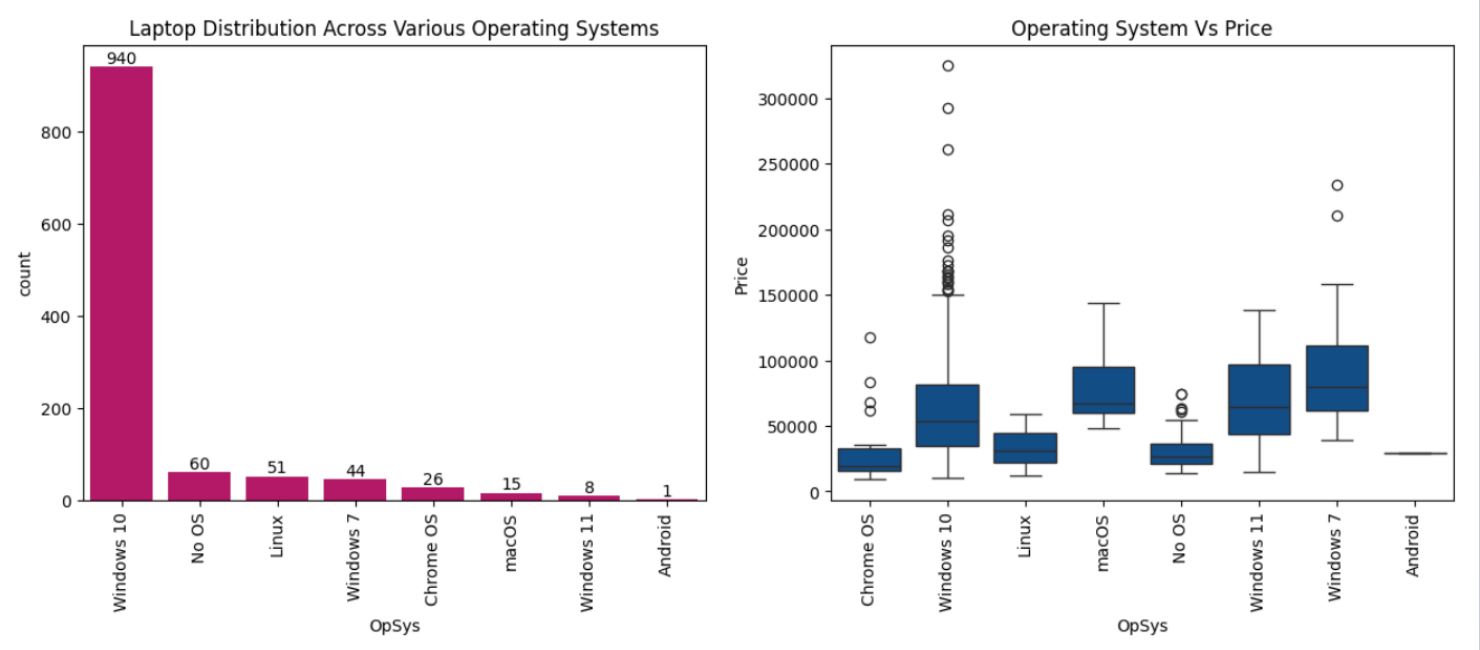




- ✓ **The distribution of the target variable price is skewed and it is obvious that commodities with low prices are sold and purchased more than the branded ones.**
- ✓ **To Normalize distribution, we will be log transforming Price Column values**

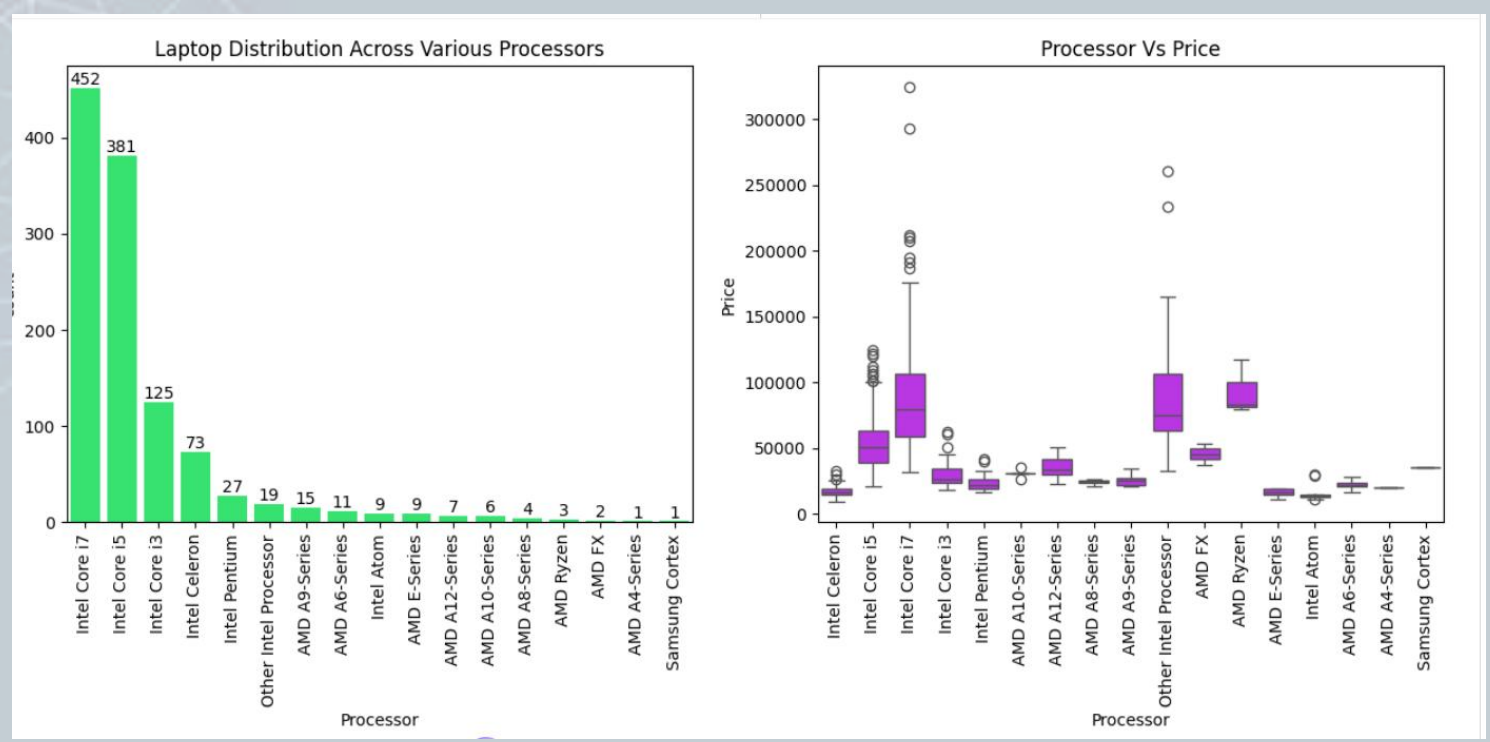
- ✓ **As per data, most purchased laptops are Dell followed by Lenovo and HP**
- ✓ **As per price, premium laptops are Razer, LG and MSI**



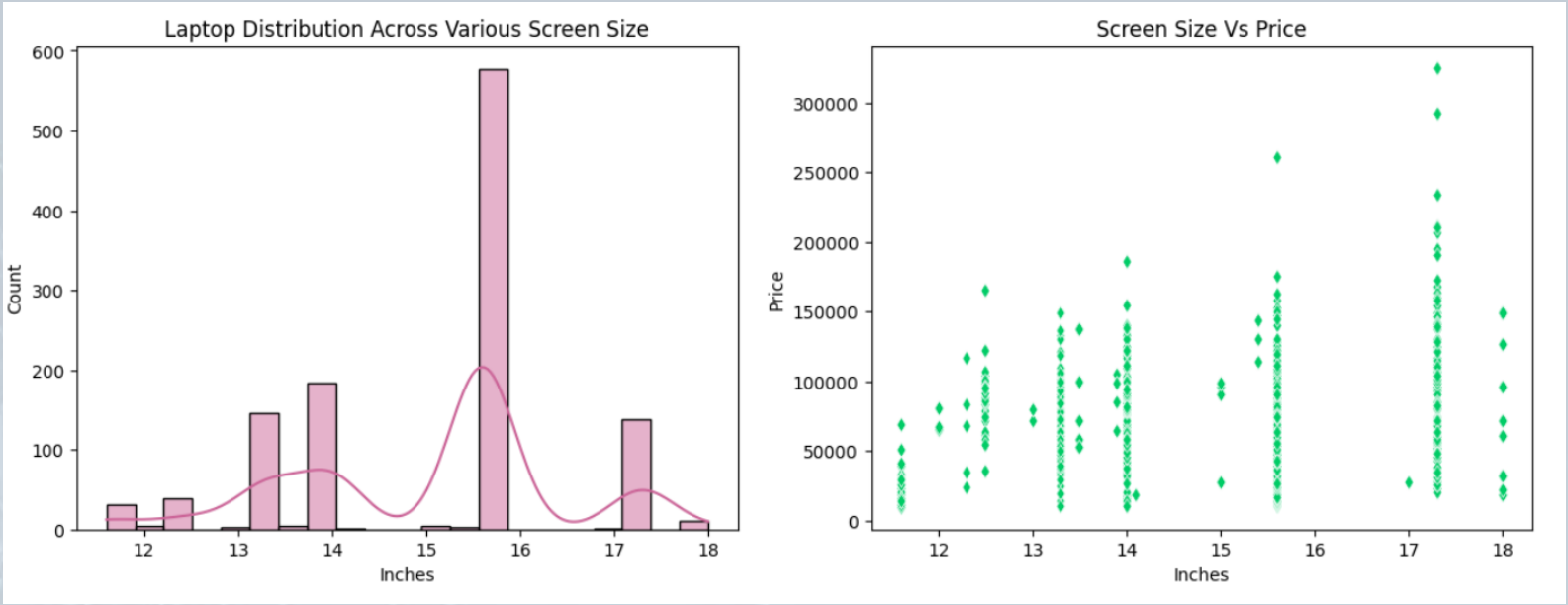


- ✓ **Most purchased TypeNames are Notebook, Ultrabook and Gaming**
- ✓ **Also the premium ones are Workstation, Ultrabook and Gaming (which have the most specs and configurations so tend to be more expensive)**

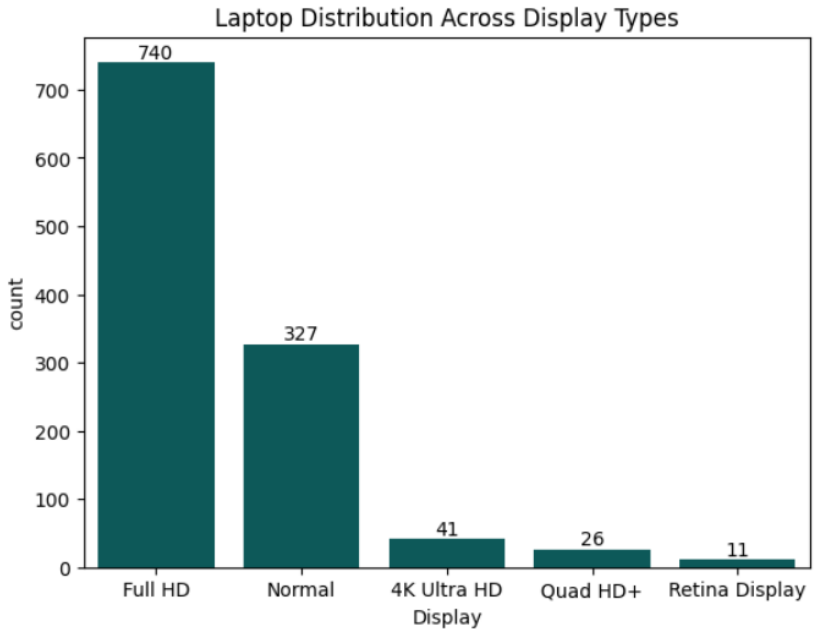
- ✓ **Most commonly purchased laptops have OS windows 10, other OS followed by Linux**
- ✓ **From price point, windows 7, 10 and MacOS are premium (however it is variable with the each laptops provided specs)**

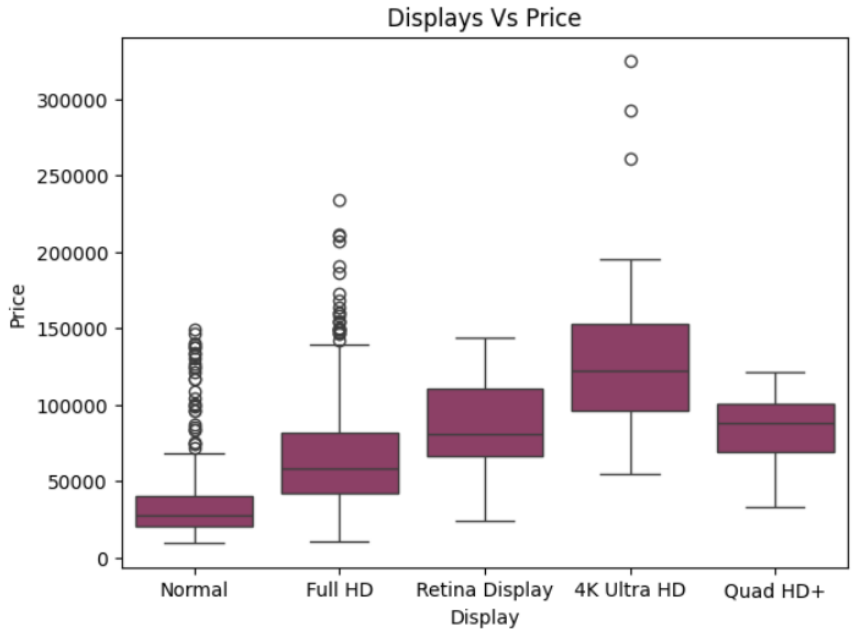


- ✓ **Common bought laptops have processors Intel Core i7,i5,i3**
- ✓ **Higher price point highlight processors such as Intel Core i7, other Intel processors and AMD Ryzen**

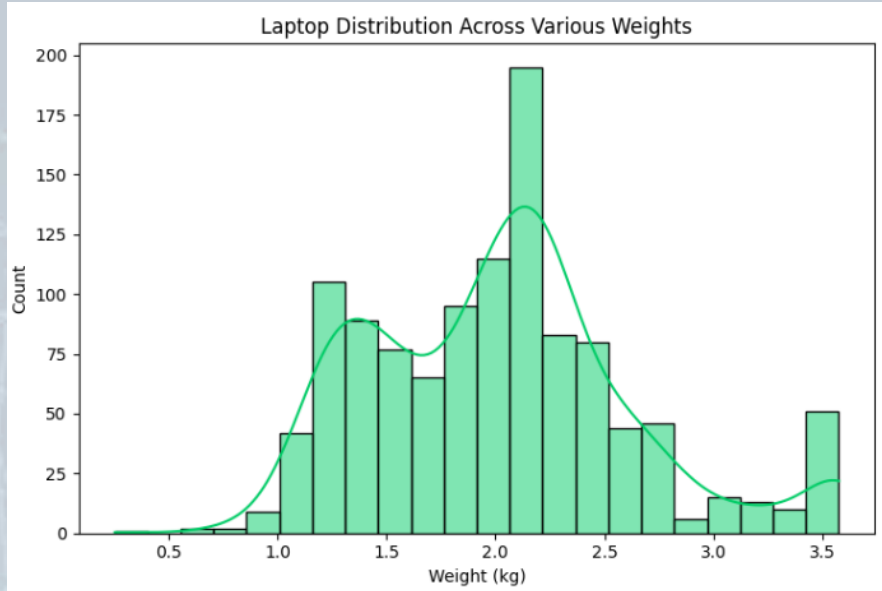


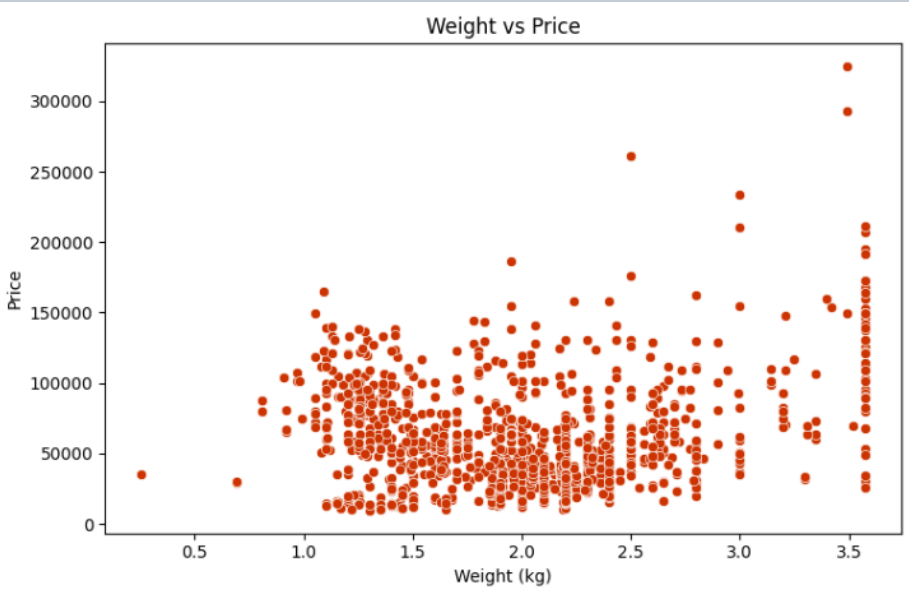
- ✓**Most commonly purchased laptops have OS windows 10, other OS followed by Linux**
- ✓**From price point, windows 7, 10 and MacOS are premium (however it is variable with the each laptops provided specs)**



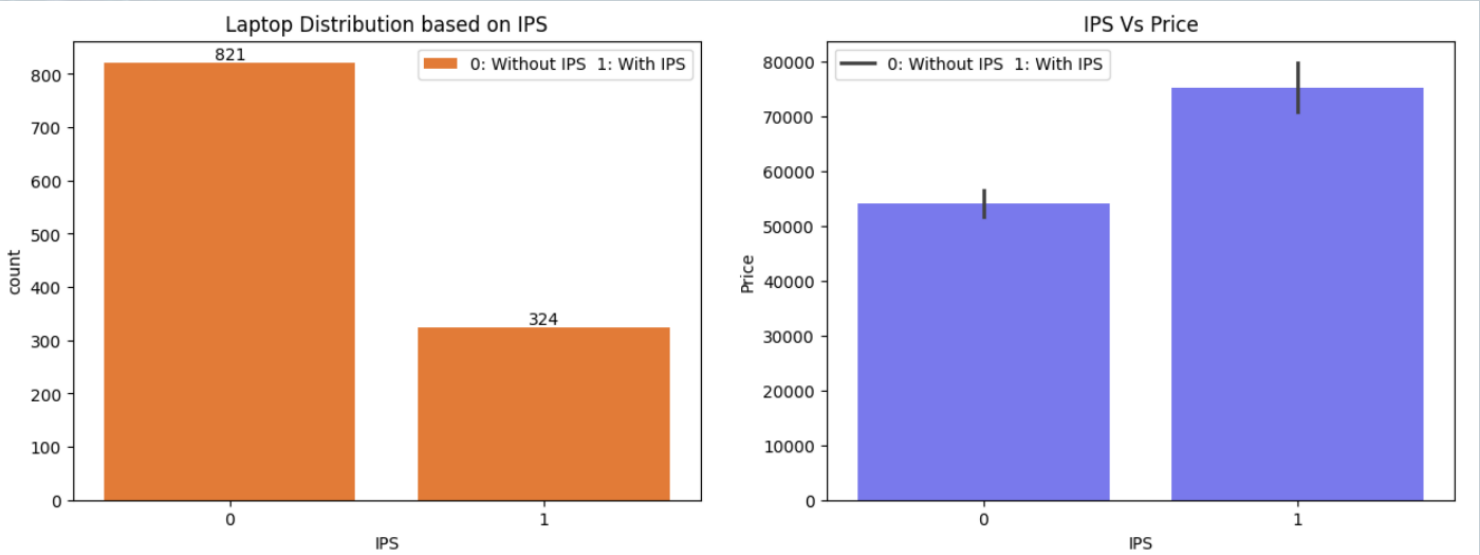


- ✓ **Highly purchased laptops have Full HD, Normal and 4k Ultra HD displays**
- ✓ **Laptops with 4k Ultra HDRetina Display and Quad HD+ are premium (as display quality is higher)**



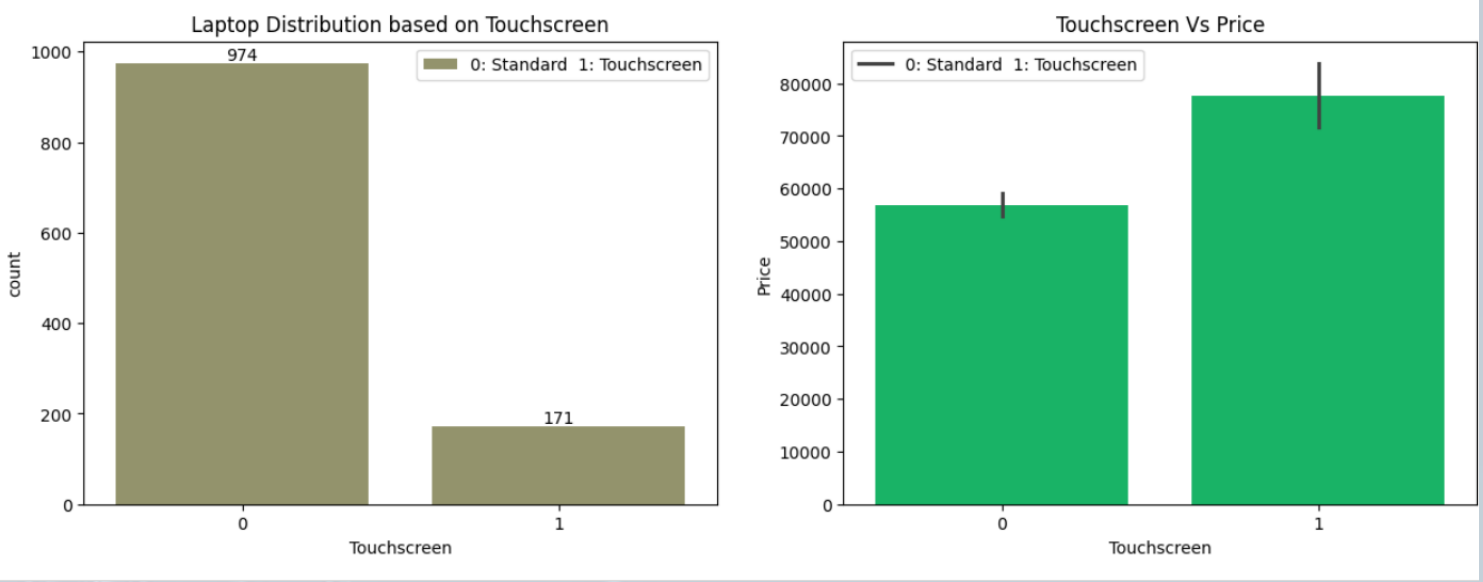


- ✓**There doesn't seem to be any correlation between the weight and price points**
- ✓**However, most of the high specs laptops tend to be heavier in weight as seen from laptops with weight above 3.0kg**



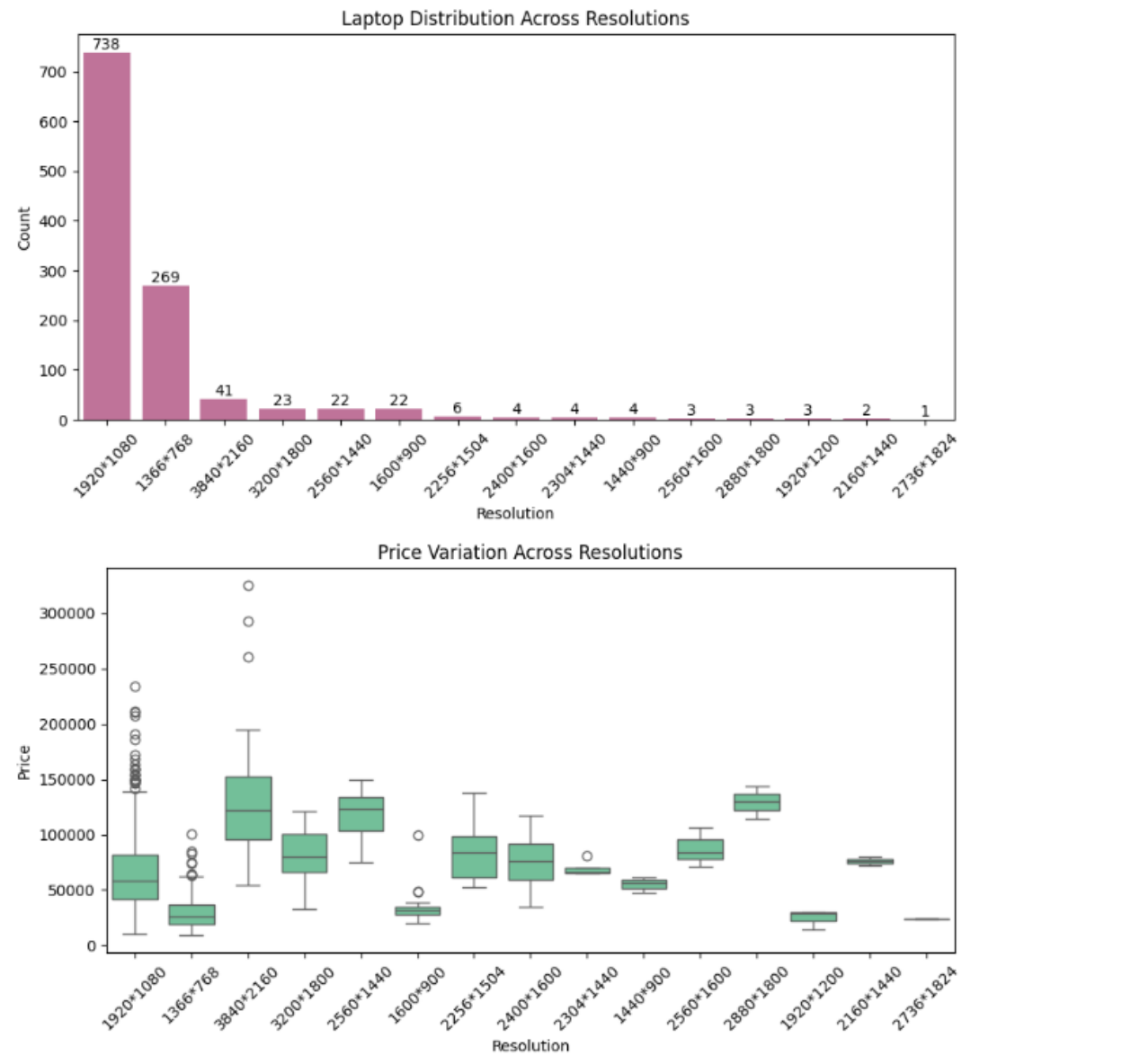
IPS (In plane switching) which is a type of LCD panel technology known for producing accurate colors and wide viewing angles

- ✓ **Most laptops bought have IPS tech**
- ✓ **Laptops with IPS tend to be on a higher price point**

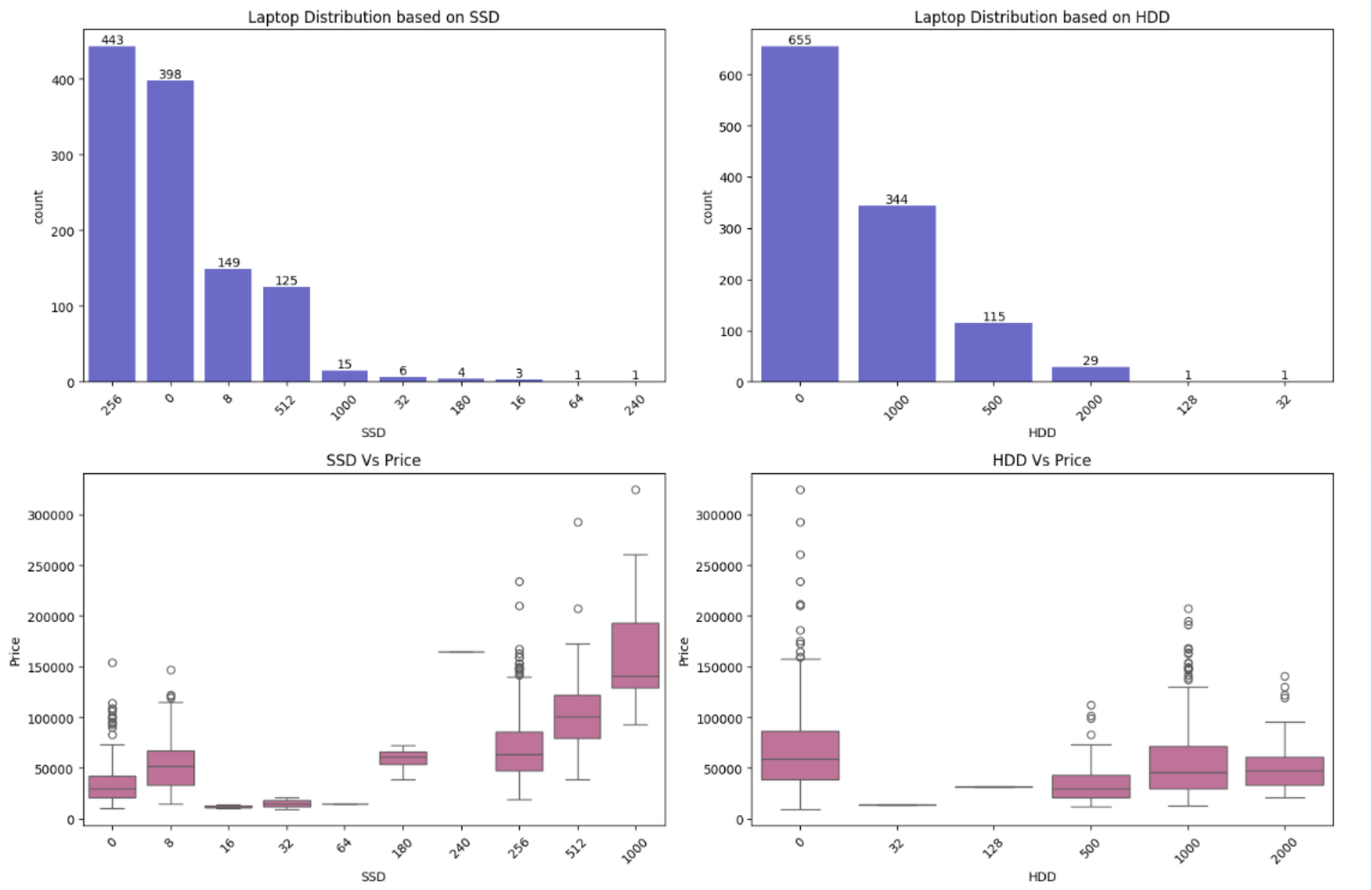


✓Most laptops bought had touchscreen feature

✓Laptops with touchscreen tend to be on a higher price point



- ✓ **Most laptops purchased were preferred with resolutions 19201080, 1366768, 3840*2160**
- ✓ **Laptops with higher resolutions were on the premium price point - 28801880, 38402160 , 2560*1440**

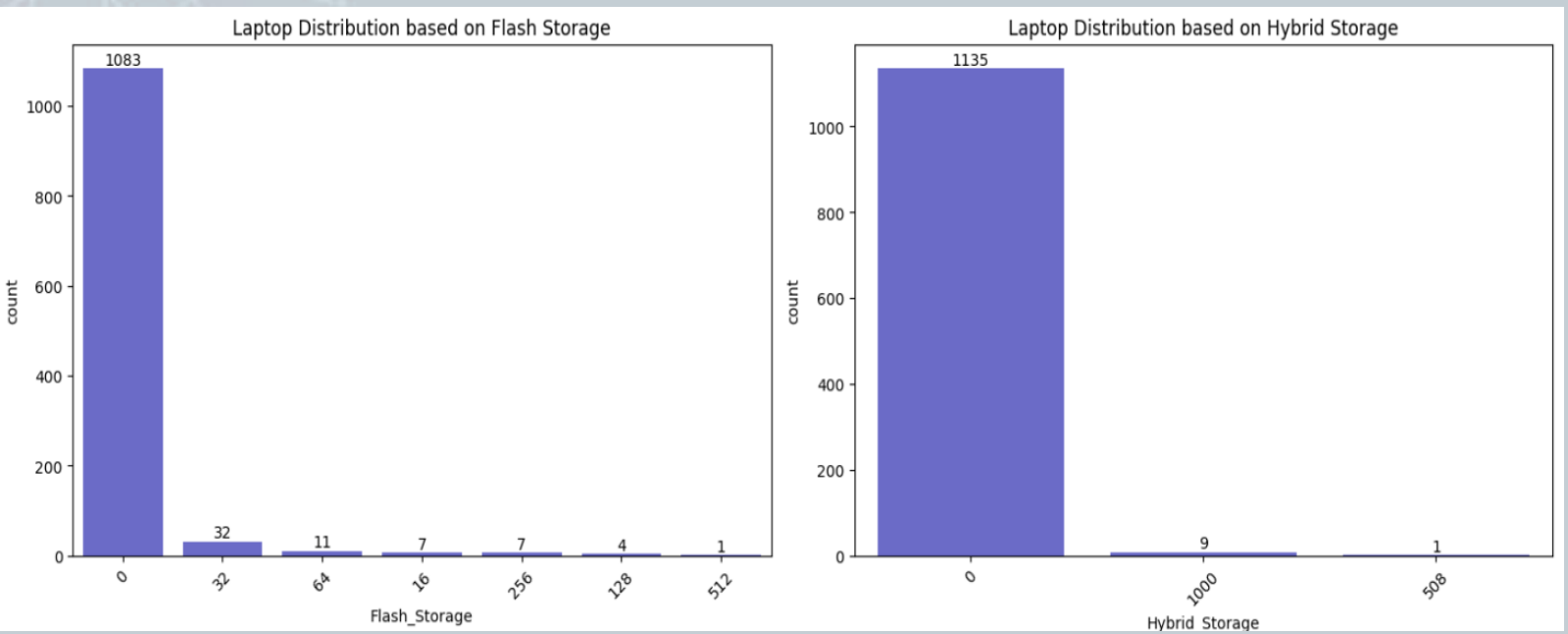


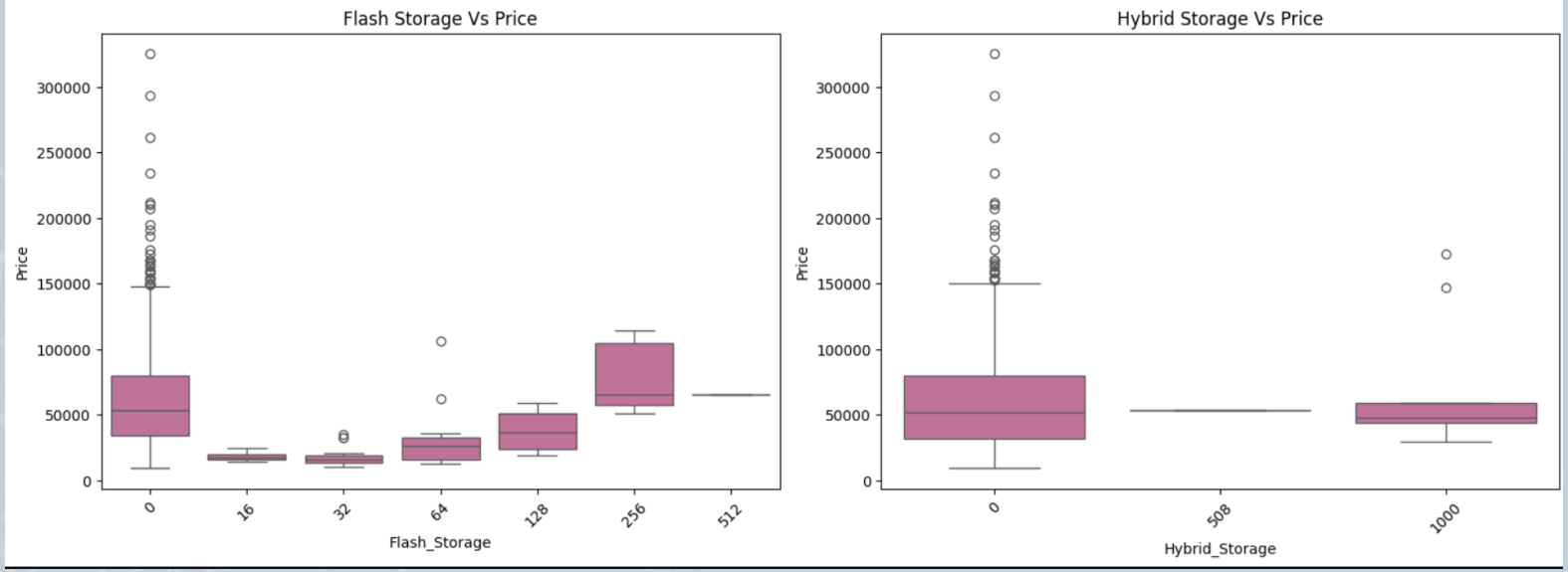
SSD

- ✓ Most purchased laptops have SSD 256, 0 , 8
- ✓ SSD correlates to price point so higher price point --> SSDs 1000,512,256

HDD

- ✓Most purchased laptops have HDD 0,1000,500
- ✓HDD doesn't correlate to price point --> so higher price point HDDs are 0 (very variable), 1000 and 2000 - most seem to be around same price point give or take



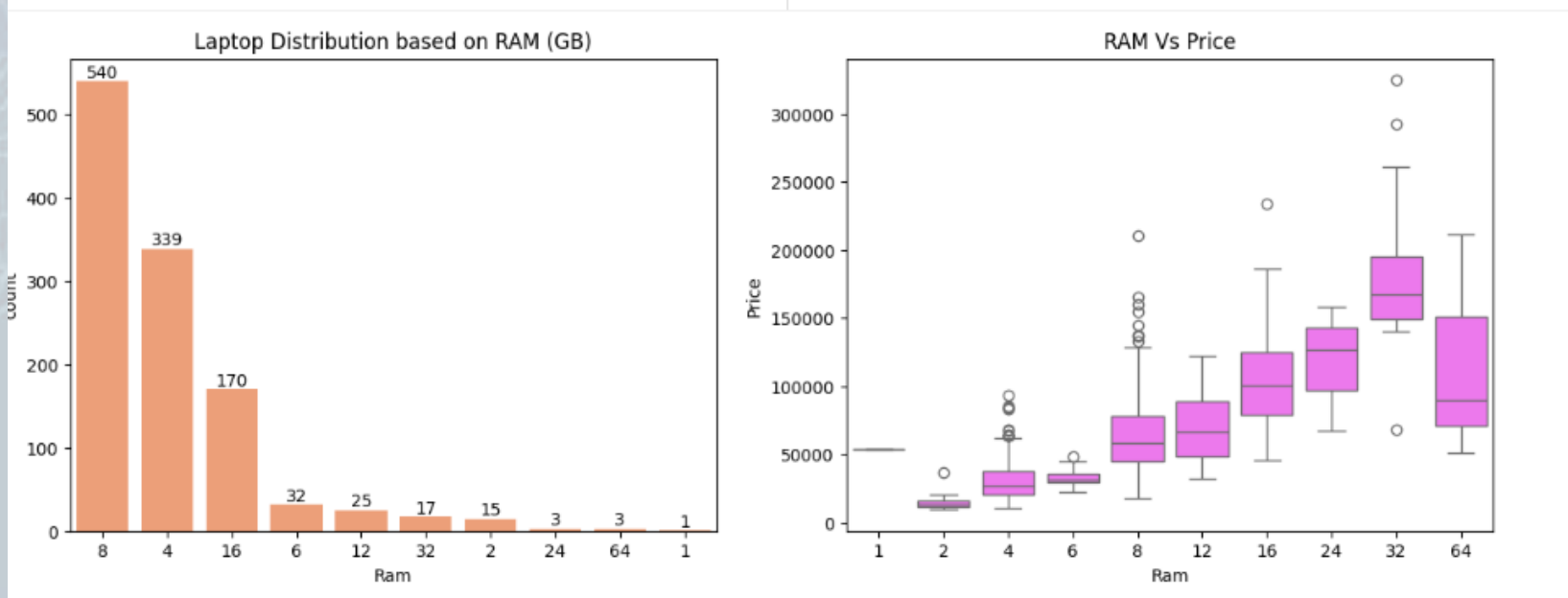


Flash Storage

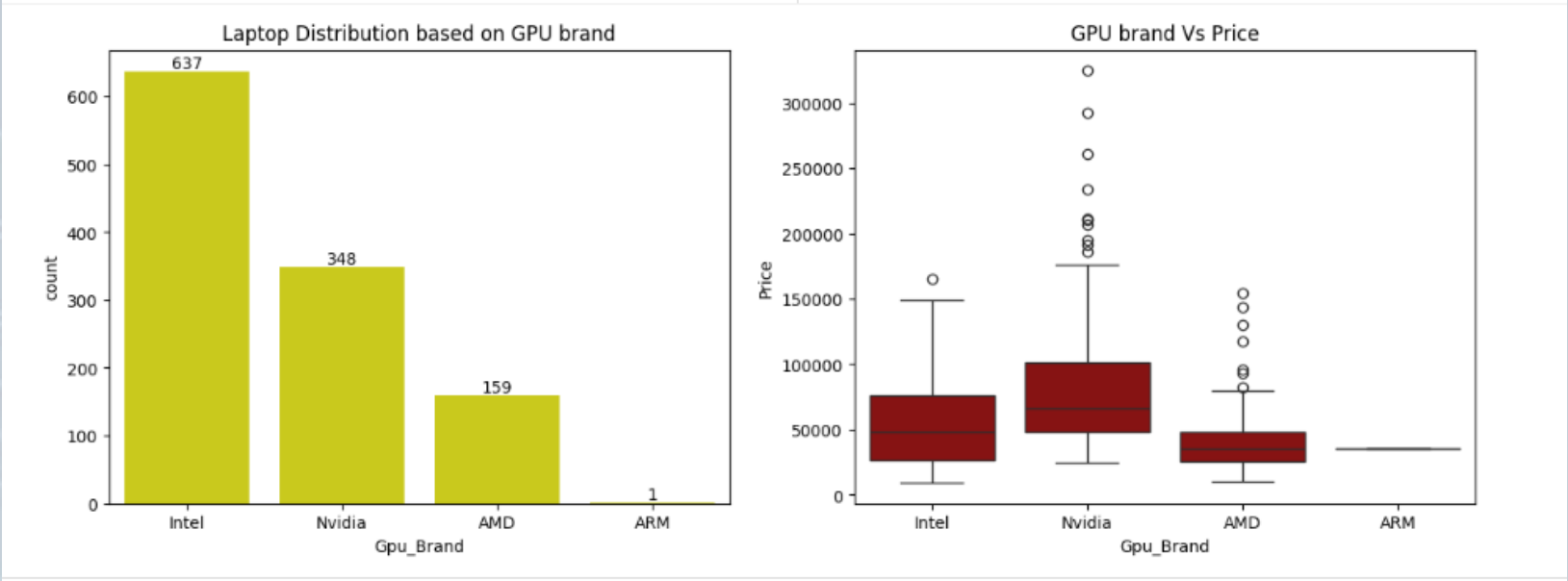
- ✓ Most purchased laptops have FS 0,32,64
- ✓ FS doesn't correlate to price point --> 0(very variable),256, 512

Hybrid storage

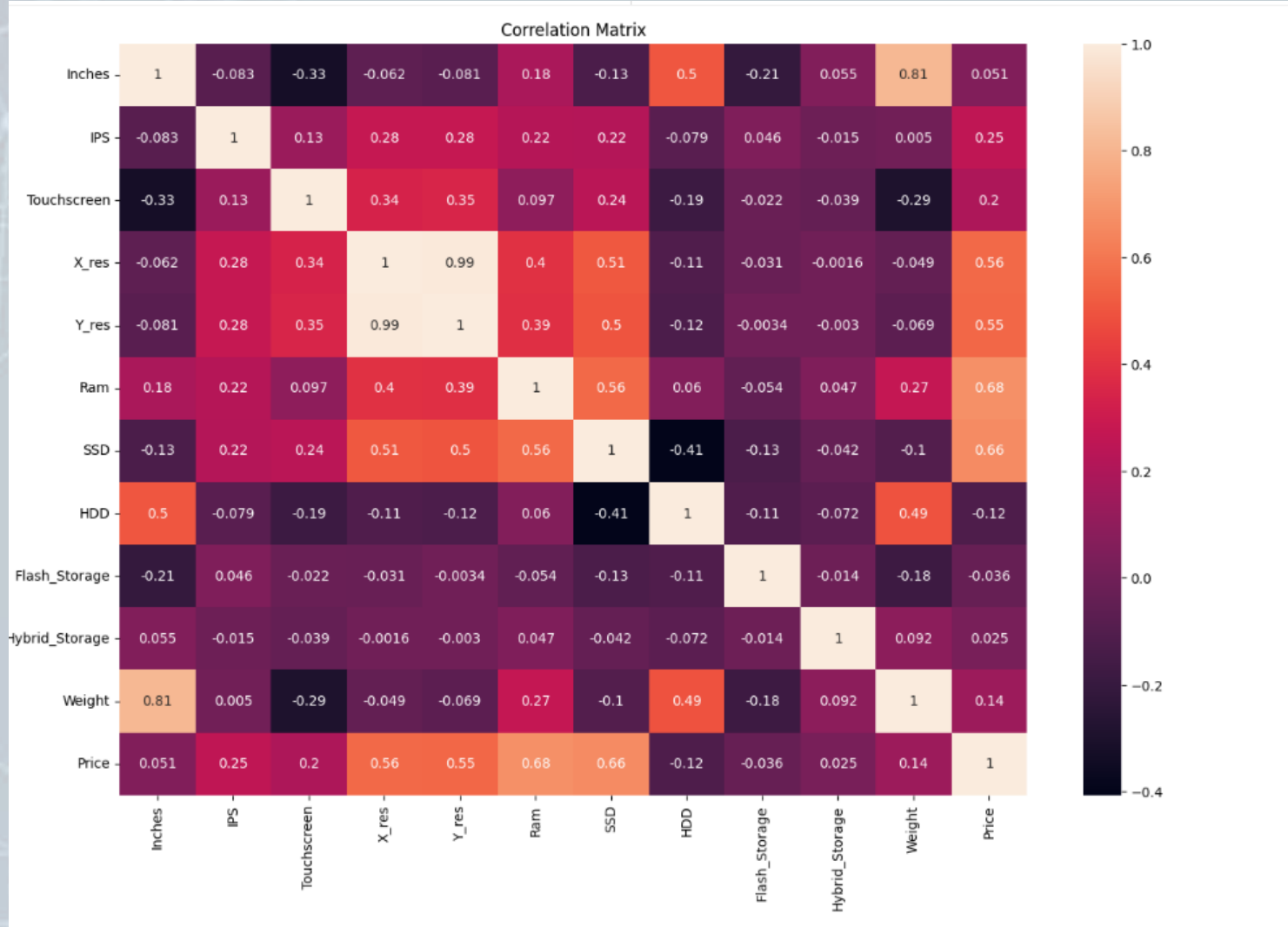
- ✓Most purchased laptops have HS have other types of HS, followed by 1.0TB and 508GB
- ✓HS with price point- while median is around same price point there are variability's with other types of HS
- ✓others -0
- ✓508 -'508GB Hybrid'
- ✓q1000- -'1.0TB Hybrid'

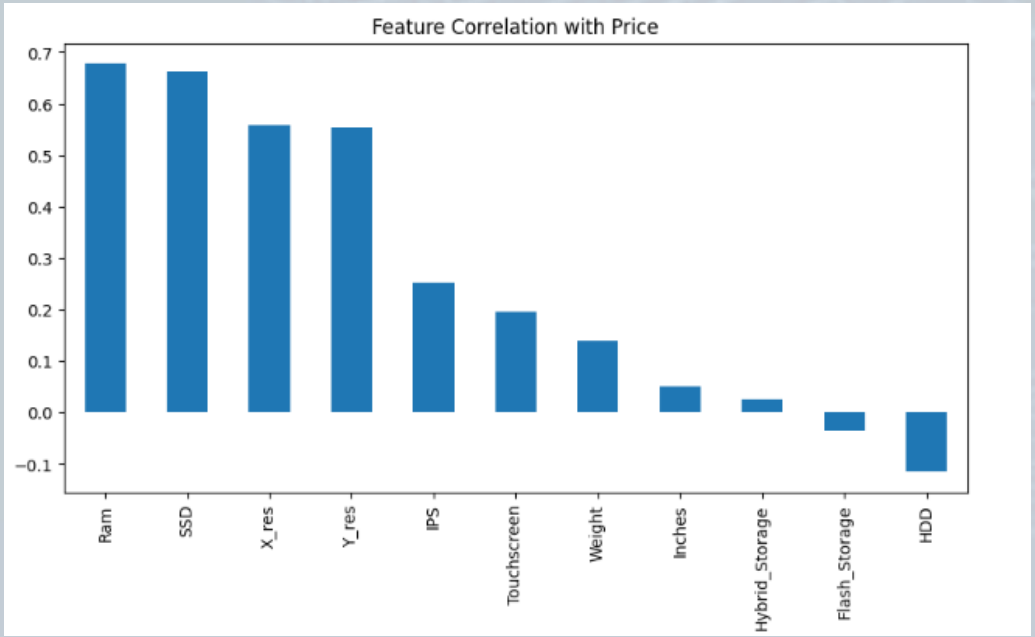


- ✓ RAM noticed in most purchased laptops were 8,4,16
- ✓ This is because higher RAM tends to correlate to higher price point (some variability noticed with 64GB RAM)



- ✓Mostly bought laptops have brand Intel followed by Nvidia and AMD
- ✓Nvidia is higher on premium side followed by Intel - this also accounts to the higher specs and GPU





From the numeric columns,

✓highly correlated: RAM, SSD and Screen resolution(X and Y res)

✓Least correlated - HDD and Flash storage

Data Preparation & ML Modeling:

Preprocessing:

- Dropped irrelevant columns (Unnamed: 0, Unnamed: 0.1)
- Extracted features from ScreenResolution, CPU, GPU, Memory
- Cleaned numeric columns: removed “GB” from RAM and “kg” from Weight
- Checked for duplicates, missing values (?), empty rows, and inconsistencies in OpSys

Feature Selection

- Dropped columns HDD and Flash storage as were least correlated to Target feature
- Split resolutions to X_res and y_res and Dropped Resolutions

ML modeling

- **Models tested:** Linear Regression, Random Forest, Gradient Boosting (GBR).
- The Gradient Boosting Regressor outperformed others and on further hyperparameter tuning improved its accuracy and generalization. → (R2= 0.8983, Mae = 0.1497 , RMSE =0.192)
- **Hyper parameter tuning used:** RandomizedSearchCV

Results and Performance

	Model	R2_log	MAE_log	RMSE_log	R2_Actual	MAE_Actual	RMSE_Actual
0	Linear Regression	0.852500	0.183779	0.231432	0.727466	11238.582591	17083.021794
1	Random Forest Regressor	0.875827	0.163167	0.212345	0.783916	9952.402942	15211.300318
2	Gradient Boost Regressor	0.895261	0.151192	0.195022	0.798517	9452.124601	14688.369669
3	Tuned Gradient Boost	0.898336	0.149796	0.192138	0.816508	9335.142181	14017.261812

```
## TRIAL 1

# UNTUNED MODEL
#HP Workstation Windows 7 Intel Core i7 15.6  Full HD 0 0 1920  1080  8 256 0 Nvidia  2.0 101178.72

# Estimated Laptop Price: ₹ 104808.37 TUNED MODEL  -- very close prediction
# Estimated Laptop Price: ₹ 100477.57  untuned model
# correct price: 101178.72
```

```
## TRIAL 2

# Apple Ultrabook macOS Intel Core i5 13.3  Retina Display  1 0 2560  1600  8 256 0 Intel 1.37

# Estimated Laptop Price: ₹ 75482.64 TUNED MODEL  -- close prediction
# Estimated Laptop Price: ₹ 73068.28  untuned model
# correct price: 84202.824
```

```
##  --------------------TRIAL 3---------with a latest model from Amazon but a lesser know brand
# https://www.amazon.in/Refurbished-Huawei-Ultrabook-i5-1135G7-KelvinD-WDH9A/dp/B0CGXB89MR/ref=sr_1_3?

# Huawei Ultrabook Windows 10 Intel Core i5 14 Full HD 0 0 1920 1080 8 512 0 Intel 1.5

# Estimated Laptop Price: ₹ 59391.87 tuned model  -- closer prediction
# Estimated Laptop Price: ₹ 73935.72  untuned model
# correct price: 54,730 refurbished - ₹1,09,990
```

- ❑ **Out of the 3 models, GBR performed the best and was taken for further tuning to improve its performance (R2 ~ 0.90)**
- ❑ **The Gradient Boosting tuned model performed well with most of the popular brands but is less reliable for brands with fewer examples in the dataset(due to underrepresentation in actual prices).**
- ❑ **This suggests that the model's accuracy for lesser-known brands depends heavily on the availability and diversity of data.**

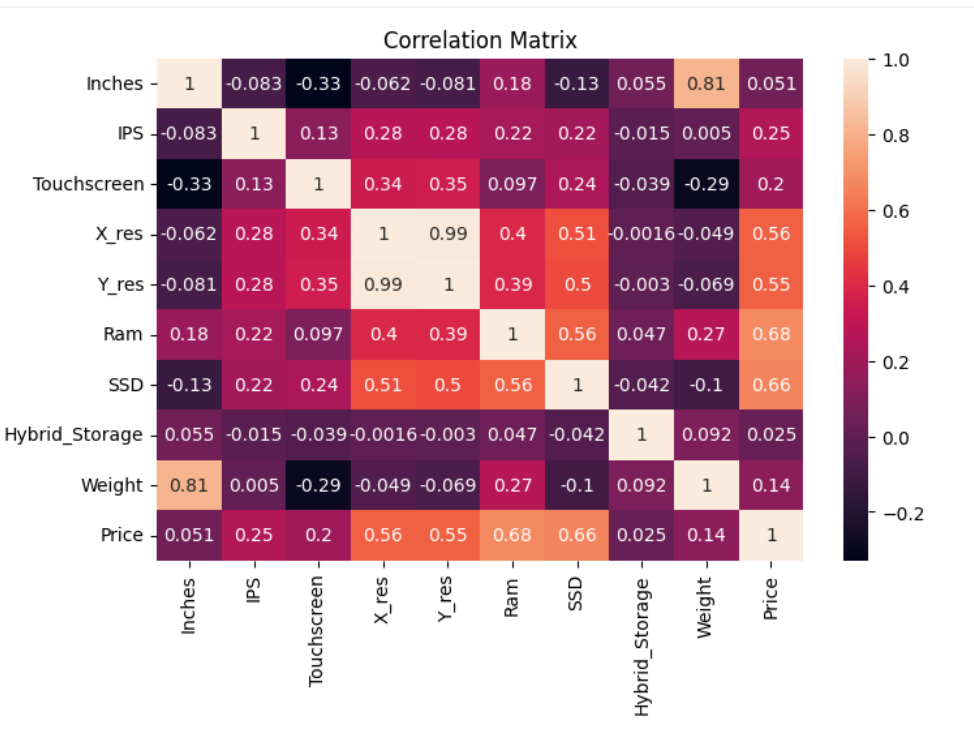
- For training prices were log transformed . The actual values are post conversions

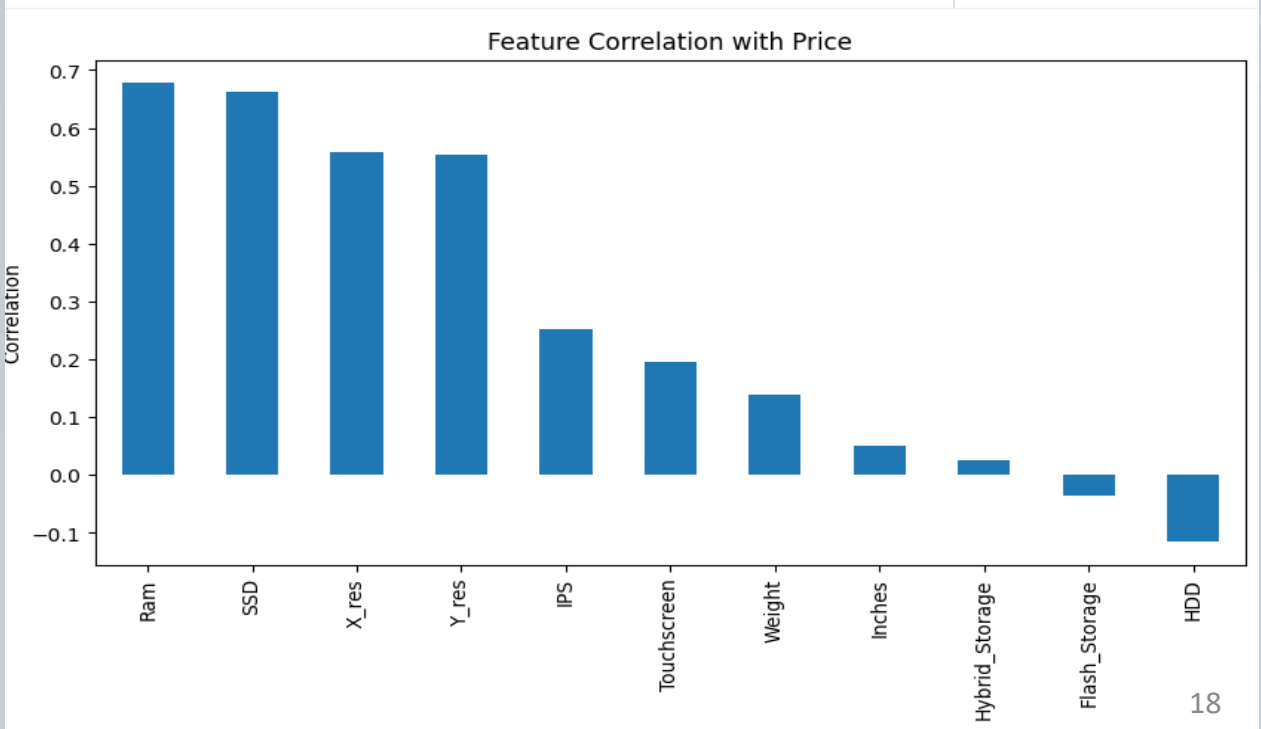
- ❑ **Also of the 3 laptop models taken for prediction (with 3rd data taken from amazon) the tuned model performed quite well**

Key Data Insights and Questions

1. Which features have the most significant impact on laptop prices?

- **In terms of Price point, RAM, SSD and Screen resolution have the biggest impact on the laptop prices. Also categorical columns - processors/CPU, IPS and touchscreen impact the price point**





2. Can the model accurately predict the prices of laptops from lesser-known brands?

1) **Fujitsu :**

- **Estimated Laptop Price: ₹ 39091.63 (correct price: ₹ 42570.72)**

2) **LG : → ₹ 121690.96 untuned model**

- **Estimated Laptop Price: ₹ 84032.25 (correct price: 122490.72)**

3) **Huawei- From Amazon details :**

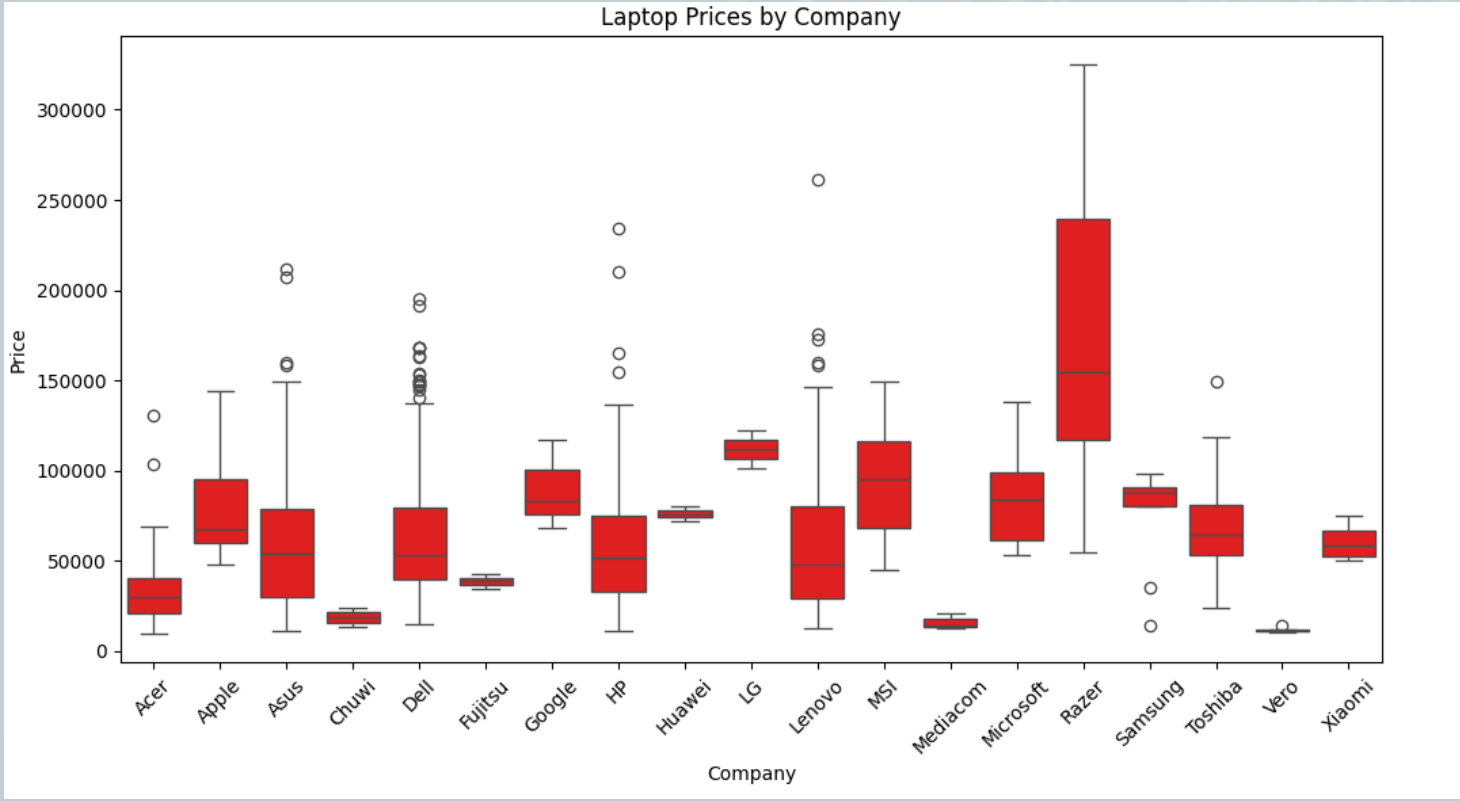
- **Estimated Laptop Price: ₹ 59391.87 (correct price: 54,730 refurbished - ₹1,09,990)**

- **The Gradient Boosting tuned model performs well with most of the popular brands but is less reliable for brands with fewer examples in the dataset.**
- **For instance, Fujitsu and Huawei was predicted closely accurate , while LG laptops showed higher errors due to underrepresentation in actual prices (but worked best with untuned GB model).**
- **This suggests that the model's accuracy for lesser-known brands depends heavily on the availability and diversity of data.**

	Company	Count	Category
0	Dell	262	Popular
1	Lenovo	260	Popular
2	HP	250	Popular
3	Asus	136	Popular
4	Acer	94	Popular
5	Toshiba	43	Lesser-known
6	MSI	36	Lesser-known
7	Apple	15	Lesser-known
8	Samsung	9	Lesser-known
9	Mediacom	7	Lesser-known
10	Razer	7	Lesser-known
11	Microsoft	6	Lesser-known
12	Vero	4	Lesser-known
13	Xiaomi	4	Lesser-known
14	Google	3	Lesser-known
15	LG	3	Lesser-known
16	Huawei	2	Lesser-known
17	Chuwi	2	Lesser-known
18	Fujitsu	2	Lesser-known

3. Does the brand of the laptop significantly influence its price?

- High-end brands: Apple, Razer, Microsoft → higher median prices.
- Budget brands: Acer, Chuwi, Vero → lower median prices.
- Some brands have wide spread (e.g., Asus, Lenovo, MSI), meaning their laptops range from budget to premium.
- Outliers are visible, e.g., very expensive laptops for Asus or Razer. Some brands are consistently more expensive. Both the boxplot and ANOVA analysis show that laptop brand significantly affects price.
- ANOVA confirms this statistically, with an F-statistic of 12.42 and a p-value < 0.0001, indicating strong evidence that brand influences price



Company	Count	Average_Price
Razer	7	178282.491429
LG	3	111834.720000
MSI	36	94130.503667
Google	3	89386.080000
Microsoft	6	85903.788000
Apple	15	80050.459040
Huawei	2	75870.720000
Samsung	9	75308.320000
Toshiba	43	67842.167442
Dell	262	64229.220989
Xiaomi	4	60390.882000
Asus	136	59016.904412
Lenovo	260	58397.605818
HP	250	57616.675162
Fujitsu	2	38574.720000
Acer	94	32785.910349
Chuwi	2	18539.974800
Mediacom	7	15717.600000
Vero	4	11584.404000

```
10 # Interpretation:
11 # p-value < 0.05 → Brand significantly affects price.
12 # p-value > 0.05 → No significant effect detected.
```

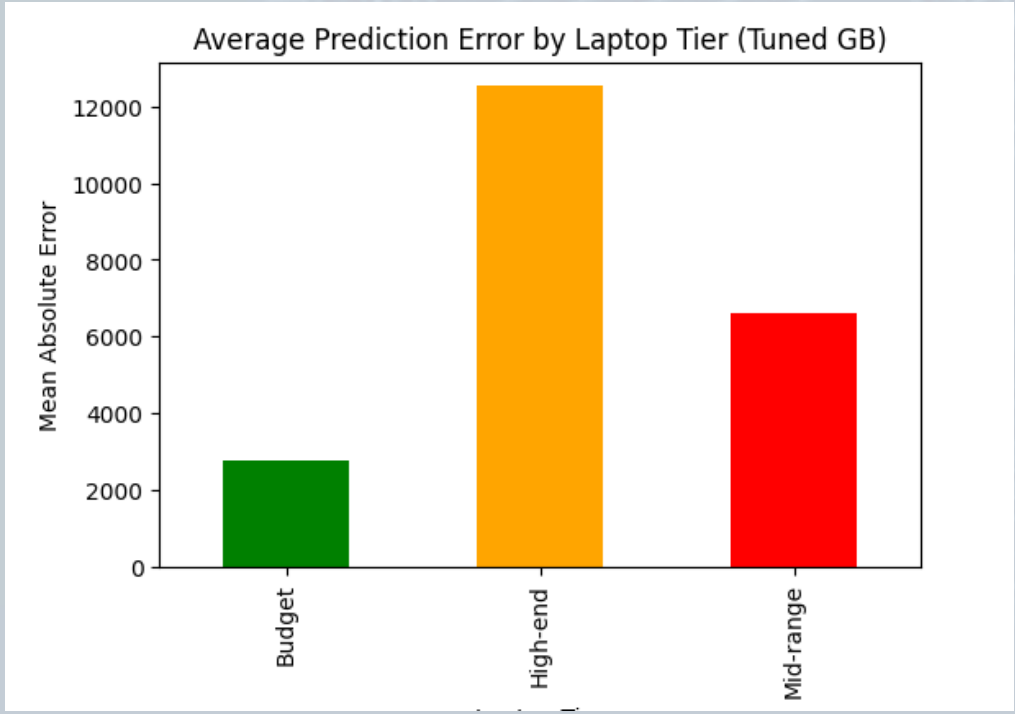
```
F-statistic: 12.423788658640548
p-value: 8.281165468007725e-34
```

4. How well does the model perform on laptops with high-end specifications compared to budget laptops?

- **Budget laptops (green bar):**
  - ➢ **Lowest MAE (~₹30,000): predicts budget laptop prices most accurately.**
- **Mid-range laptops (red bar):**
  - ➢ **Higher MAE (~₹70,000): less accurate here compared to budget laptops.**
- **High-end laptops (orange bar):**
  - ➢ **Highest MAE (~₹1,30,000): struggles as predictions deviate significantly from actual prices.**

✓ **High-end laptops often have more variability in features (gaming GPUs, premium displays, unusual specs), making them harder for the model to predict.**

✓ **Budget laptops are more standardized, so predictions are closer to the true prices.**



- **Budget:** Up to ₹50,000
- **Mid-range:** ₹50,001 – ₹1,00,000
- **High-end:** Above ₹1,00,000

5. What are the limitations and challenges in predicting laptop prices accurately?

Several limitations and challenges account in predicting laptop prices accurately:

- **Data quality issues such as limited, outdated or biased or imbalanced data affect prediction accuracy**
- **Price fluctuations that occur frequently due to new release, market changes**
- **Complex feature relationships needing careful engineering and encoding.**
- **Hyperparameter tuning complexity, generalization and risk of overfitting.**
- **Handling new or rare laptop features not in training data.**
- **External factions like discounts, vendor strategies, and geography not captured.**

6. How does the model perform when predicting the prices of newly released laptops not present in the training dataset?

Tested the GB model with a laptop model which isn't on our dataset(amazon) Huawei- From Amazon details : The model overestimated the price by : Overestimation: Untuned → ~₹19.2k, Tuned → ~₹4.6k

Estimated Price: ₹ 73935.72 (untuned model) and ₹ 59391.87 (tuned model)

Actual price: ₹ 54,730

Conclusion: Here Huawei laptops showed higher errors due to underrepresentation in actual prices(model may underperform on laptops with rare or unseen configurations, as these were underrepresented in the training data)

```
1 ##  --------------------TRIAL 3---------with a latest model from Amazon but a lesser know brand
2 # https://www.amazon.in/Refurbished-Huawei-Ultrabook-i5-1135G7-KelvinD-WDH9A/dp/B0CGXB89MR/ref=sr_1_3?dib=eyJ2I
3
4 # Huawei Ultrabook Windows 10 Intel Core i5 14 Full HD 0 0 1920 1080 8 512 0 Intel 1.5
5
6 # Estimated Laptop Price: ₹ 59391.87 tuned model
7 # Estimated Laptop Price: ₹ 73935.72  untuned model
8 # correct price: 54,730 refurbished - ₹1,09,990
```

Power BI Visualization

Laptop Price Prediction Dashboard

Total Laptops in Dataset
1145

Total Laptop Brands
19

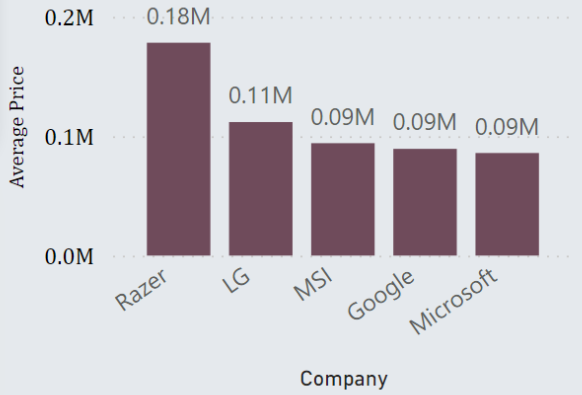
Model Types
6

Average Laptop Price(₹)
60.03K

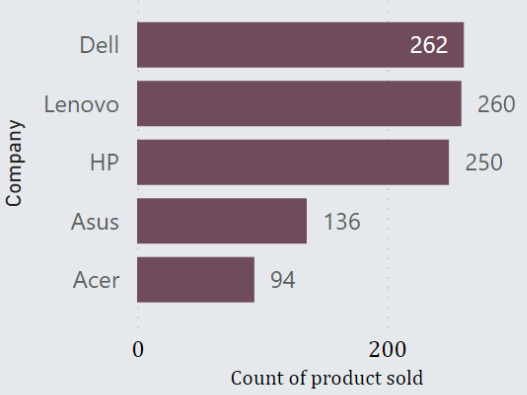
Total Laptops in Dataset
1145

Price Range of Laptops
9,270.72 324,954.72

Top Premium Laptop Brands



Top 5 Companies by laptops Sold



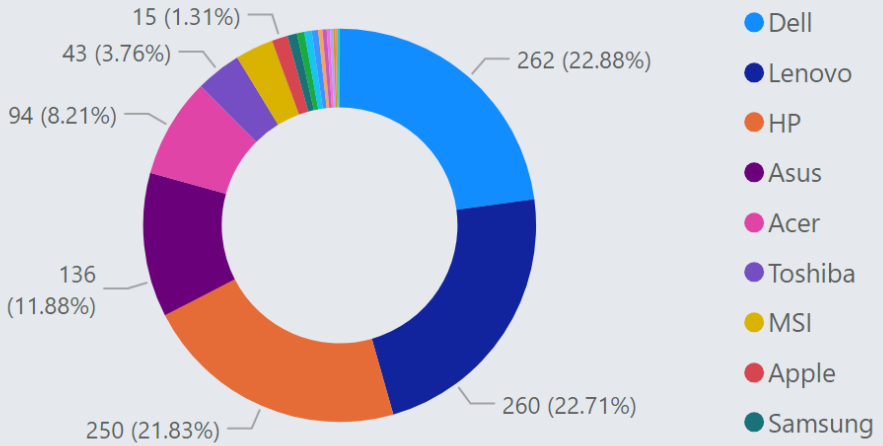
Brands
All

Model Type
All

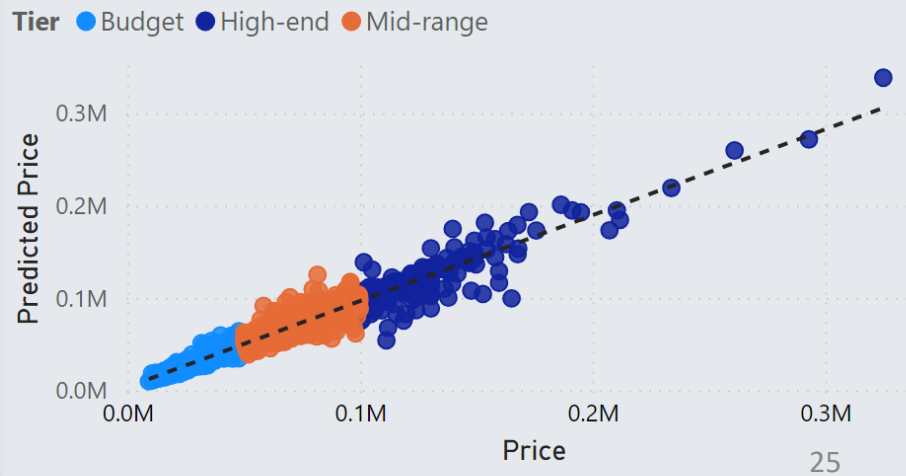
Operating System
All

Price Range Tier
All

Market Share



Actual Price Vs Predicted Price (Tuned Gradient Boost Regressor)



Top Brand	Top Processor	Top SSD	Top RAM
Dell (262 laptops)	Intel Core i7 (452 laptops)	256 (443 laptops)	8 (540 laptops)

Top GPU Brand	Top Resolution	Top Operating System
Intel (637 laptops)	1920x1080 (738 laptops)	Windows 10 (940 laptops)

Top Display	Top Model Type	Top Screen Size	Top Laptop Weight
Full HD (740 laptops)	Notebook (647 laptops)	15.6 (577 laptops)	2.2 (95 laptops)

Price Tier Count	IPS (In plane switching) Available	Touchscreen Feature Available	Hybrid Storage Available
550 | 438 | 157	324 | 821	171 | 974	1,135 | 1 | 9
Budget | Mid-Range | High-End	Yes | No	Yes | No	Other | 508GB | 1.0TB

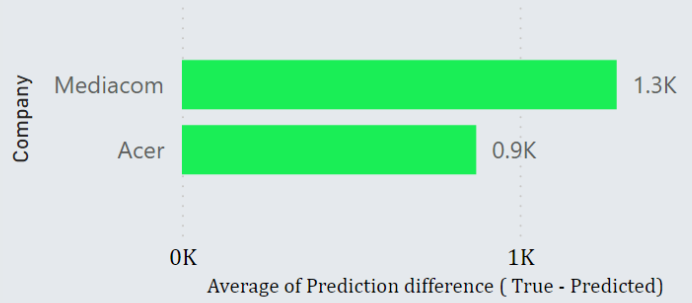
Performance Metrics (Log-Transformed price)

Model	R2_log	MAE_log	RMSE_log
Tuned Gradient Boost	0.90	0.15	0.19
Gradient Boost Regressor	0.90	0.15	0.20
Random Forest Regressor	0.88	0.16	0.21
Linear Regression	0.85	0.18	0.23

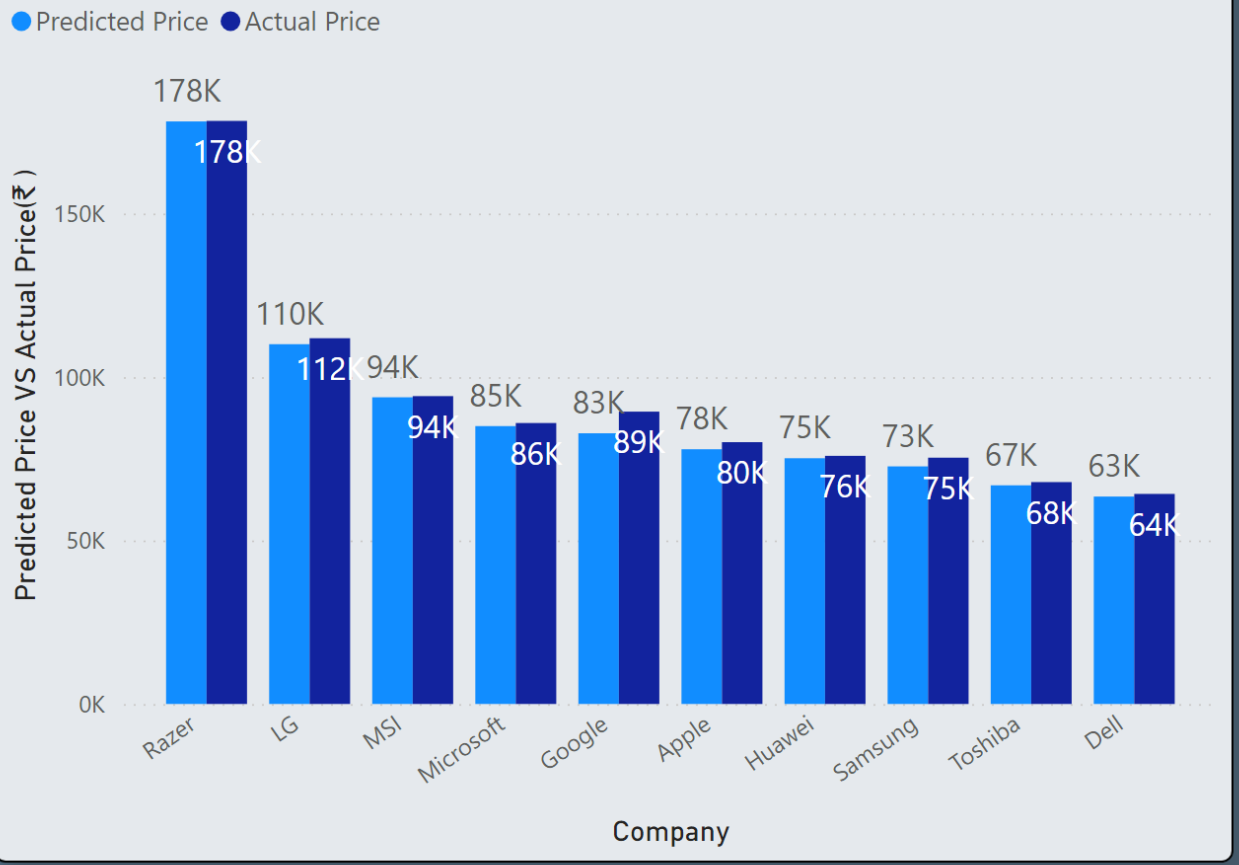
Performance Metrics (Converted to the Actual price)

Model	R2_Actual ₹	MAE_Actual ₹	RMSE_Actual ₹
Tuned Gradient Boost	0.82	9,335.14	14,017.26
Gradient Boost Regressor	0.80	9,452.13	14,688.37
Random Forest Regressor	0.78	9,952.40	15,211.30
Linear Regression	0.73	11,238.58	17,083.02

Top 2 Over-predicted Companies



Comparison of Predicted vs Actual Price by Model



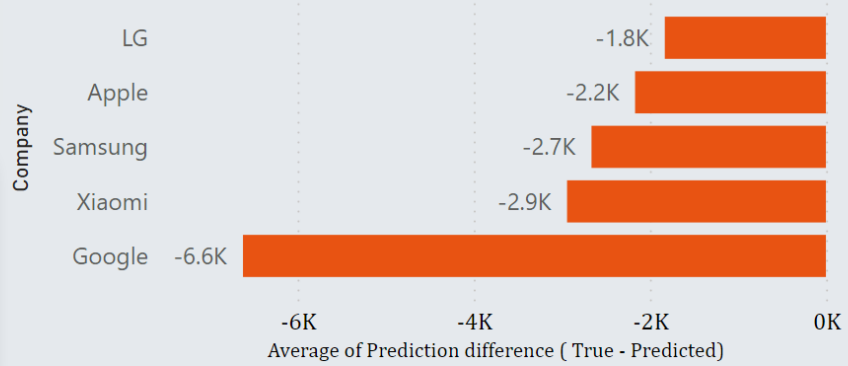
Price Tier

All

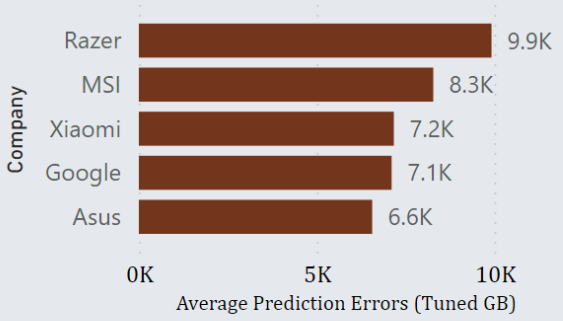
Company

All

Top 5 Under-predicted Companies



Top 5 companies with Largest Prediction Errors



Tuned GB Model Prediction Accuracy

Conclusion

- **Brand Influence**: Dell, Lenovo, and HP dominate in purchase volume, while Razer, LG, and MSI produce the most premium laptops with higher price points.
- **Type of Laptop**: Notebooks dominate sales, followed by Ultrabooks and Gaming laptops. Gaming laptops and workstations are the most expensive due to higher specs.
- **Operating System**: Windows 10 is the most popular OS but has minimal impact on price; pricing is driven mainly by hardware specs.
- **Processor:** Intel Core i7 laptops occupy the premium range, while models with Intel Core i3 or AMD processors are among the most affordable.
- **Display Quality:** Display quality strongly impacts price. 4K Ultra HD Retina and Quad HD+ displays are premium, while Full HD and lower resolutions are more common and more affordable.
- **Screen Technology**: IPS displays and touchscreen features are popular and associated with higher price points.
- **Weight and Price**: Weight does not directly correlate with price; however, higher-spec models tend to weigh more (typically >3.0 kg).
- **Resolution and Size:** Higher screen resolutions and larger screen sizes correlate positively with price. Most buyers prefer around a 14-inch medium sized screen as well as medium resolutions considering the affordability.
- **RAM**: Strong positive correlation between RAM size and price; 8 GB RAM is the most common among buyers. This is a major driver of price.
- **GPU:** Laptops mostly feature Intel GPUs; Nvidia GPUs are premium and increase cost due to better performance.
- **Storage:** SSD and hybrid storage moderately impact price, while HDD and flash storage have minimal influence.


Thank you