

Predicting used Laptop Prices using Machine Learning

Group 2:
Aden Ali

Muhammad Ammar Faisal
Syed Muhammad Abdullah
Raahim Poonawala
Affan Zaman

Initials:

- This project will develop on a machine learning model to predict the price of used laptops and uncover key market insights.

Research Questions:

- Can a predictive model be created to estimate laptop prices based on given parameters?
- Which parameters are most crucial in determining laptop prices?

Data:

- We used data of 300 (rows) laptops with 19 key variables (columns)

Data Cleaning:

- Data Structuring: Columns are renamed for clarity, and unnecessary columns are dropped.
- Handling Duplicates and Null Values: Duplicate records are removed, and missing values are filled based on logical assumptions or public information

Exploratory Data Analysis (EDA)

01.

- Brand Distribution: The dataset includes laptops from 16 different brands, with HP, Dell, Apple, Lenovo, and Acer being the most common.

02.

- Processor Generation: The 8th generation is the most common, with distinct patterns observed for Apple and non-Apple laptops.

03.

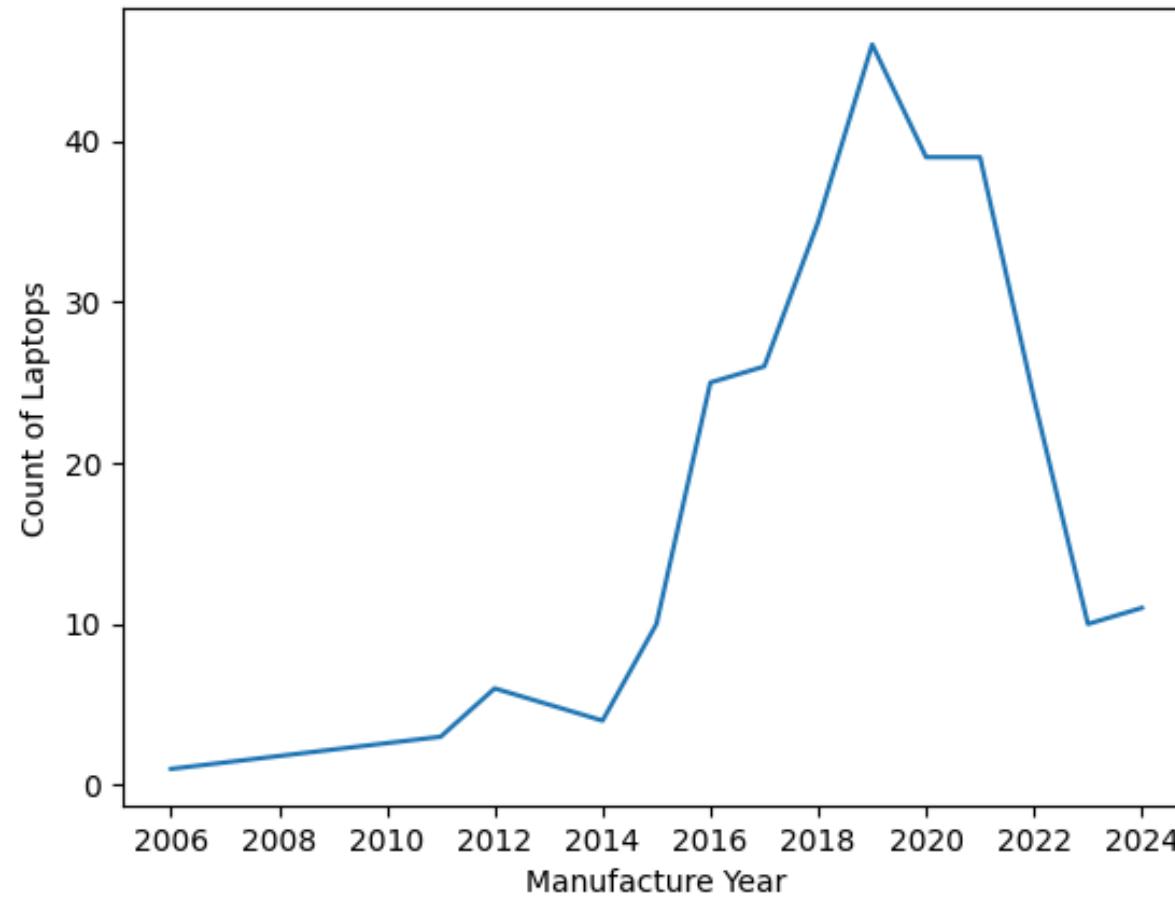
- Processor Type: Intel i7 and i5 processors dominate the used laptop market.
- Processor Speed: Most laptops have processor speeds between 1.5 to 2.7 GHz.

04.

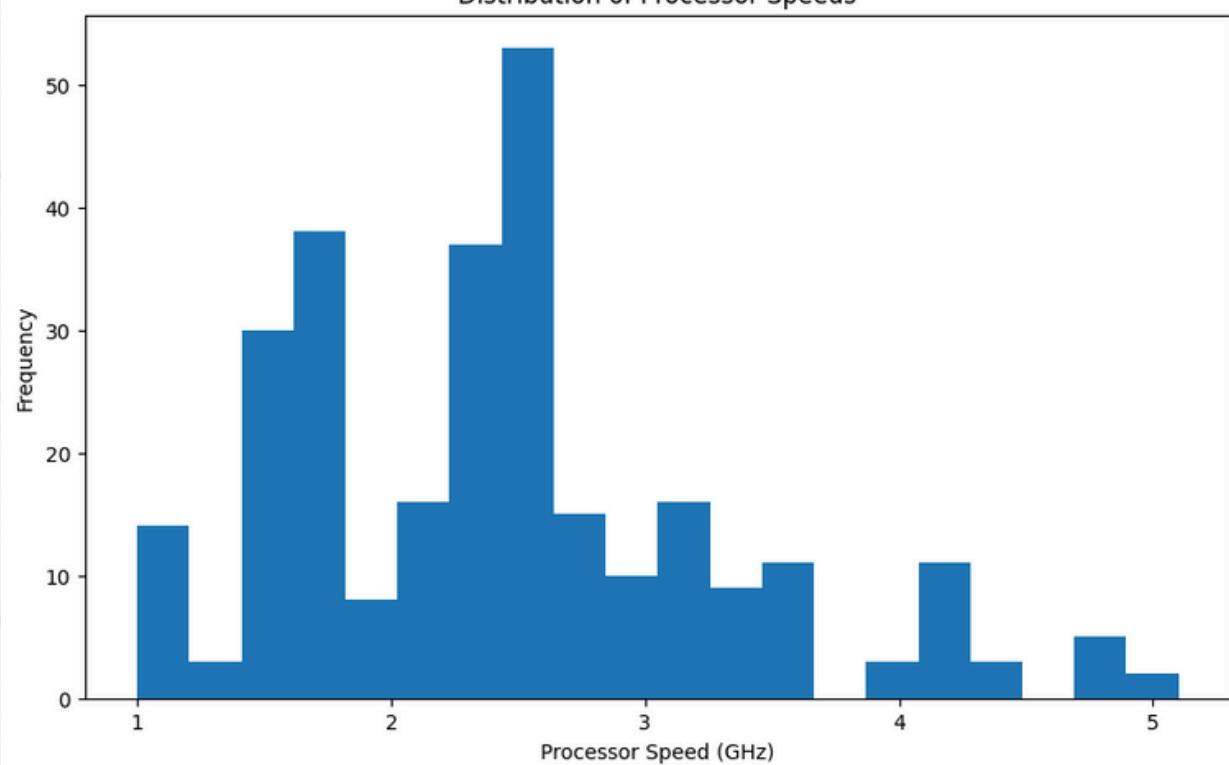
- Manufacture Year: Most laptops were manufactured between 2016 and 2022.
- Other Features: Analysis includes operating systems, colors, SSD and RAM sizes, screen resolutions, and battery capacities.

EDA Results:

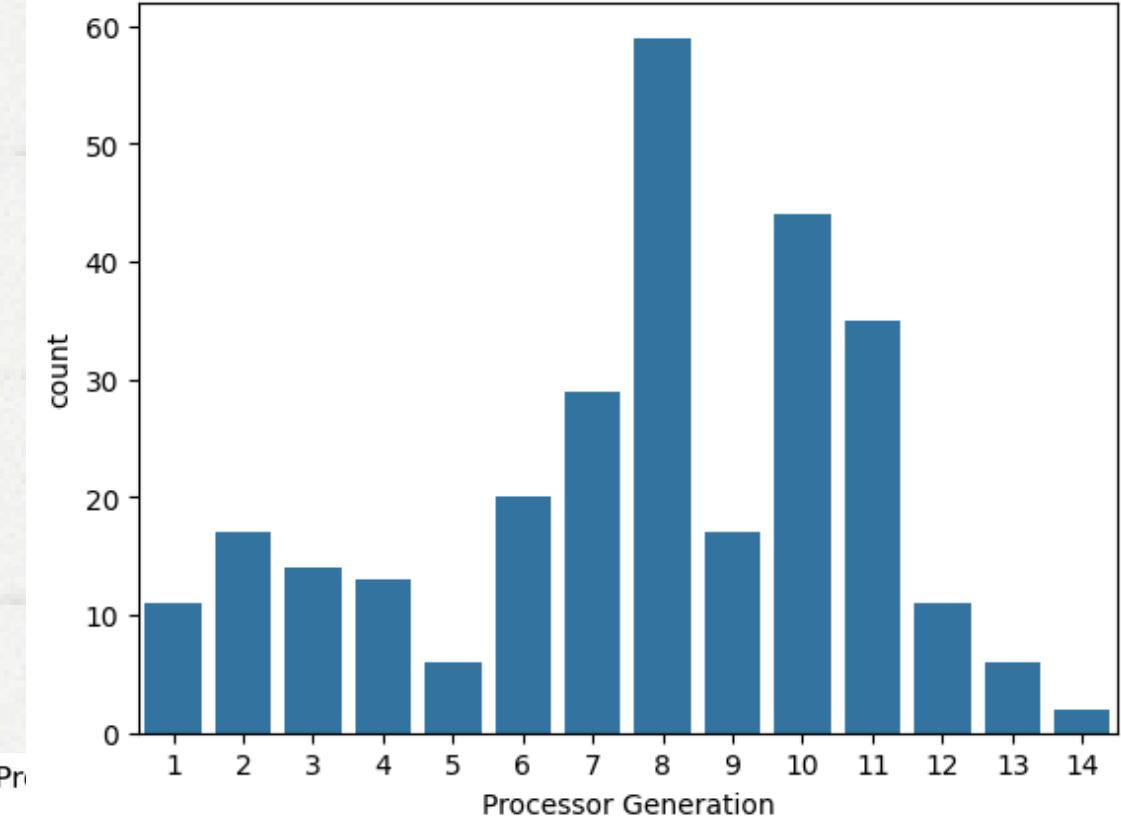
Laptops Count by Manufacture Year



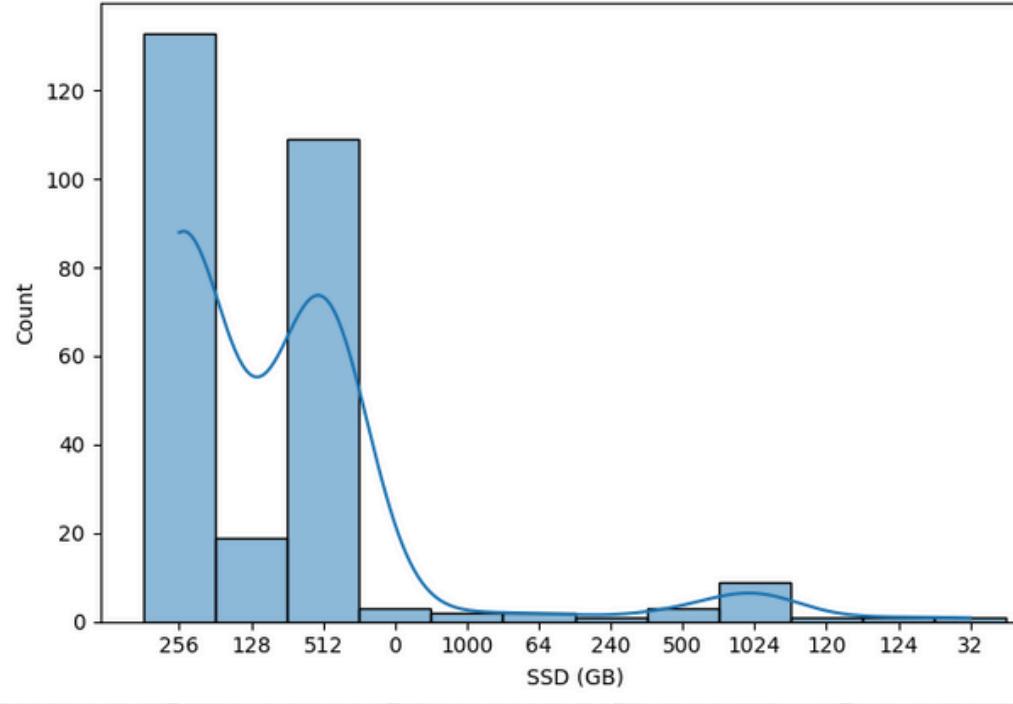
Distribution of Processor Speeds



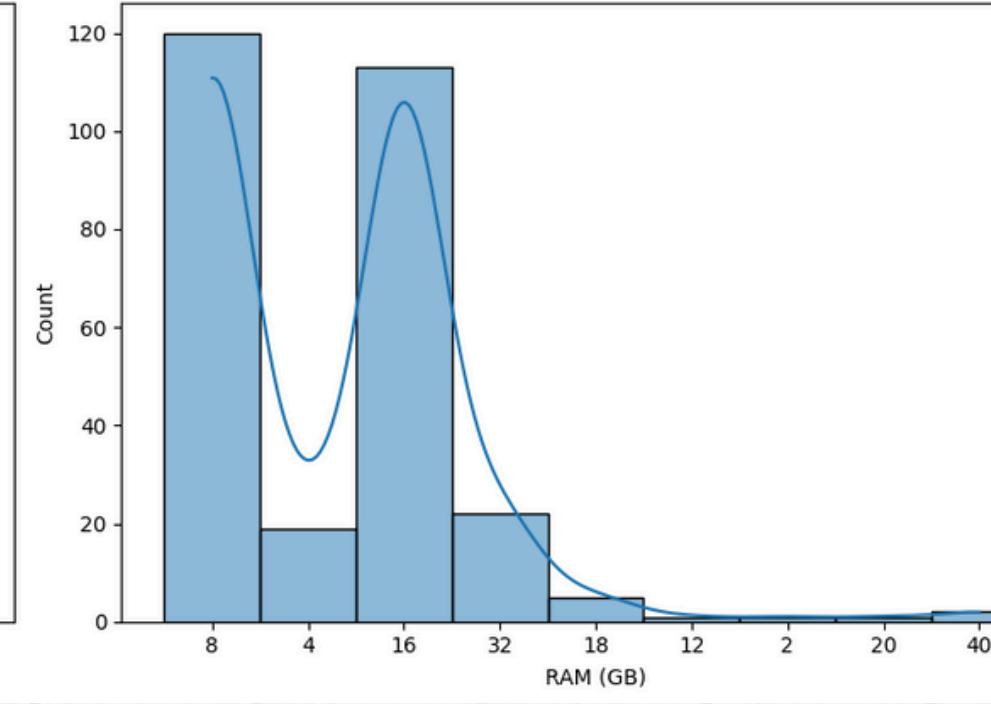
Processor Generation Distribution



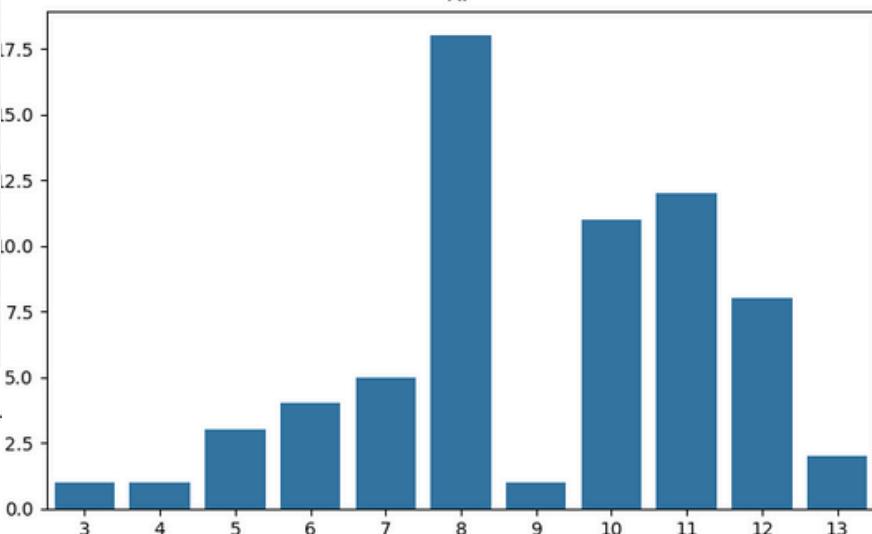
SSD Distribution



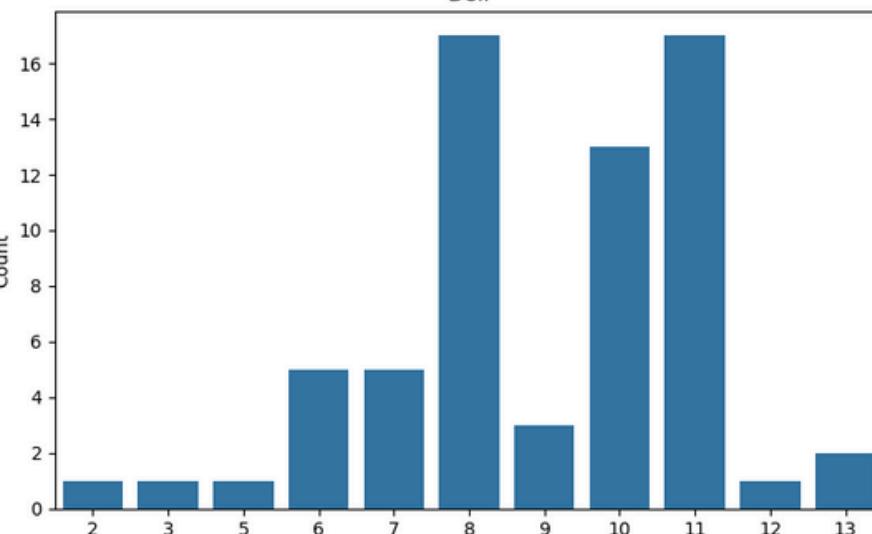
RAM Distribution



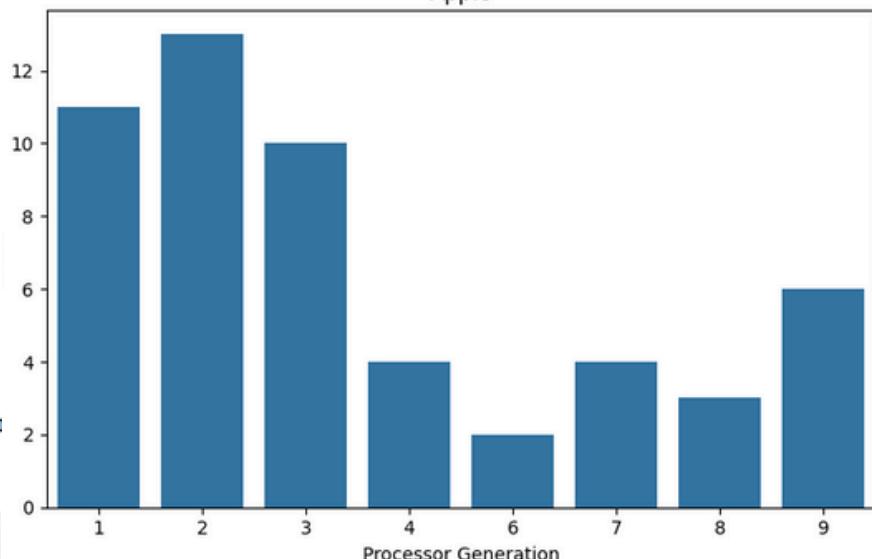
HP



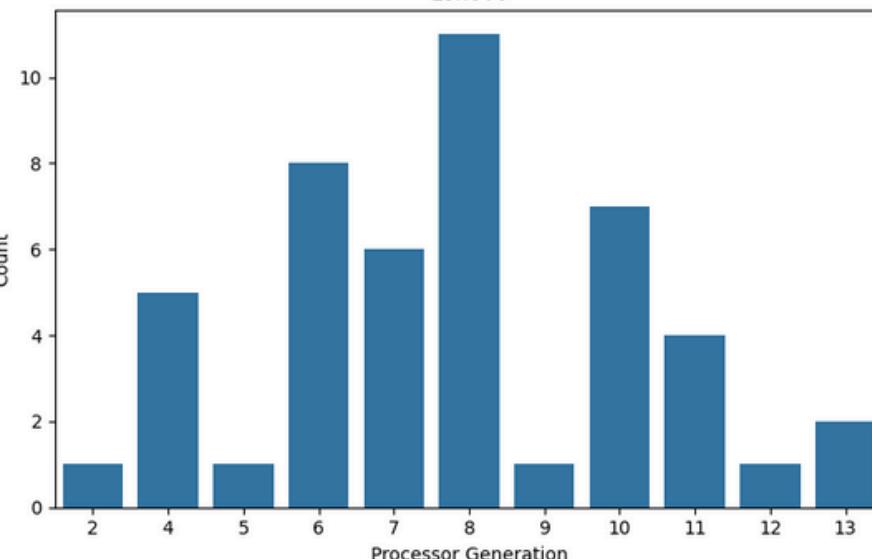
Dell



Apple

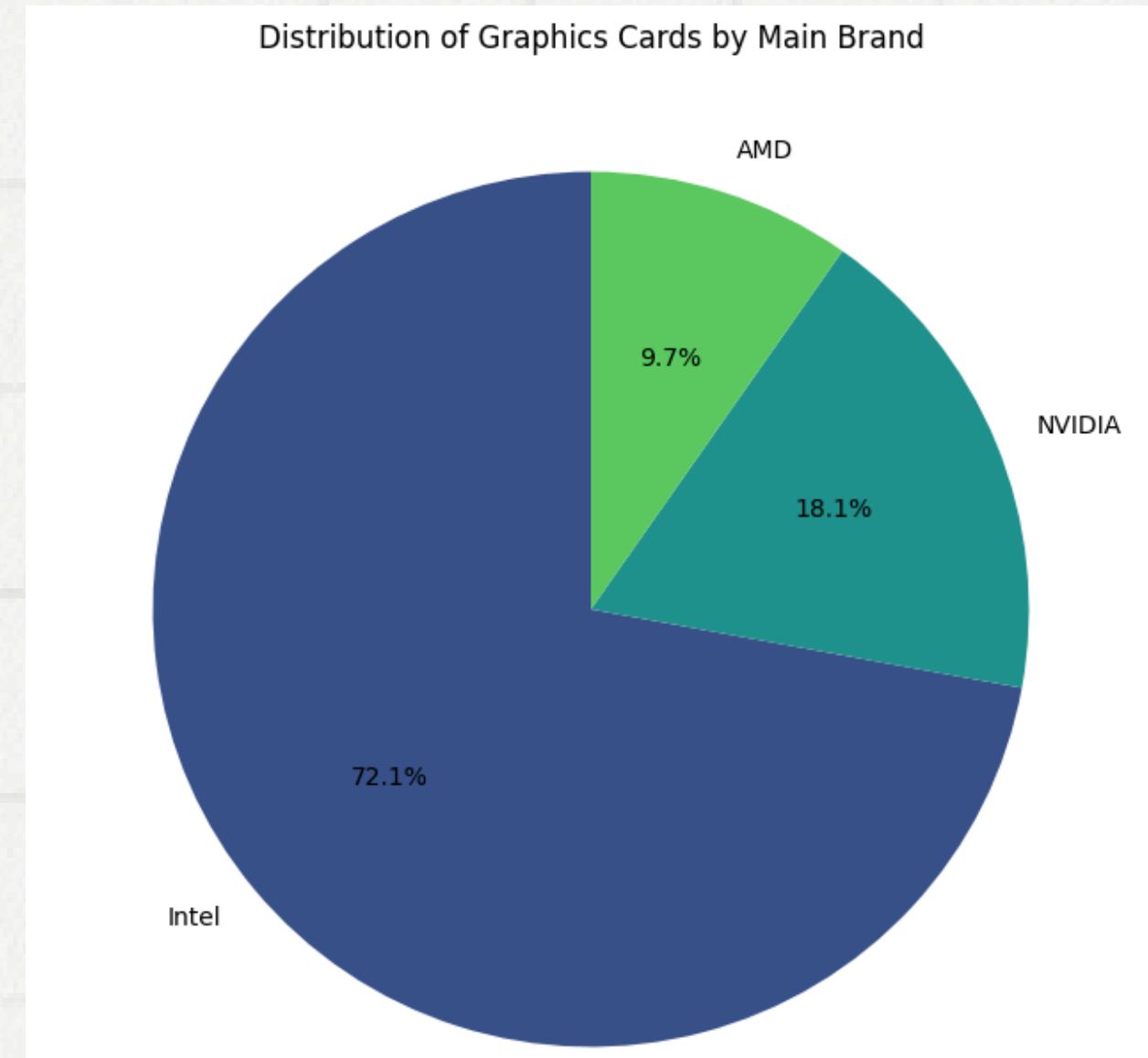
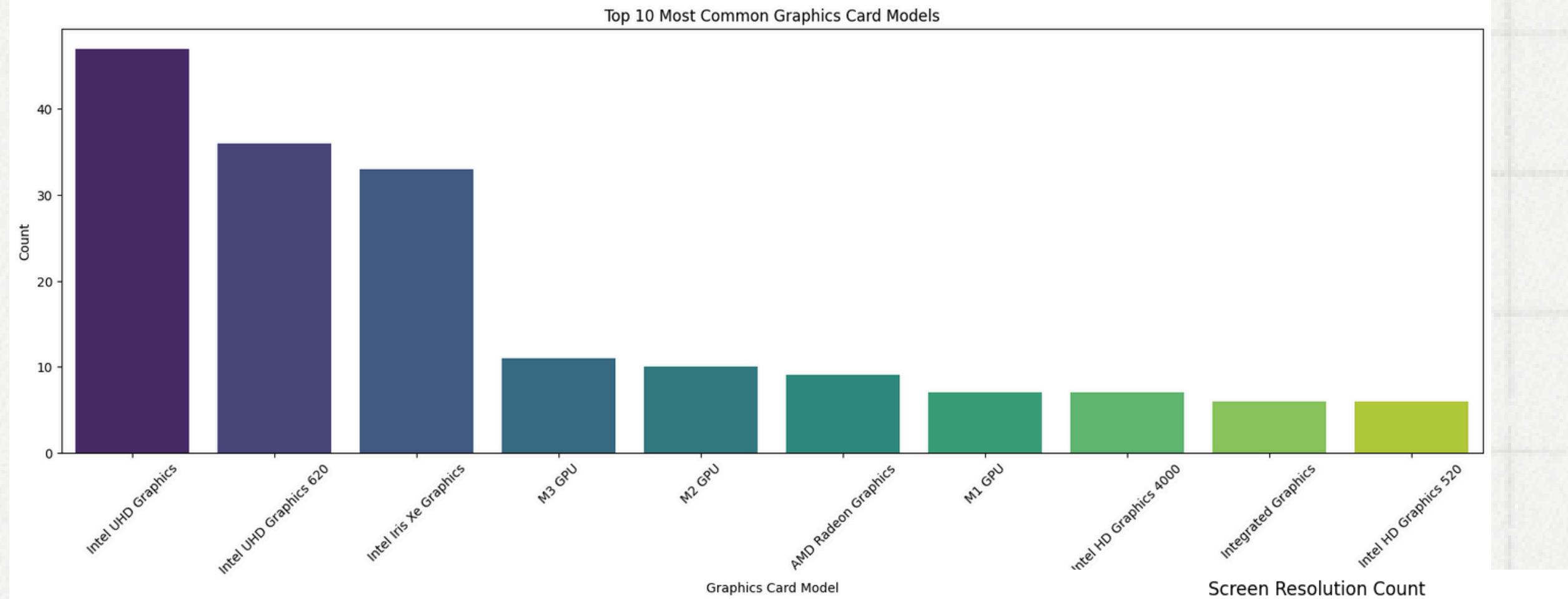


Lenovo

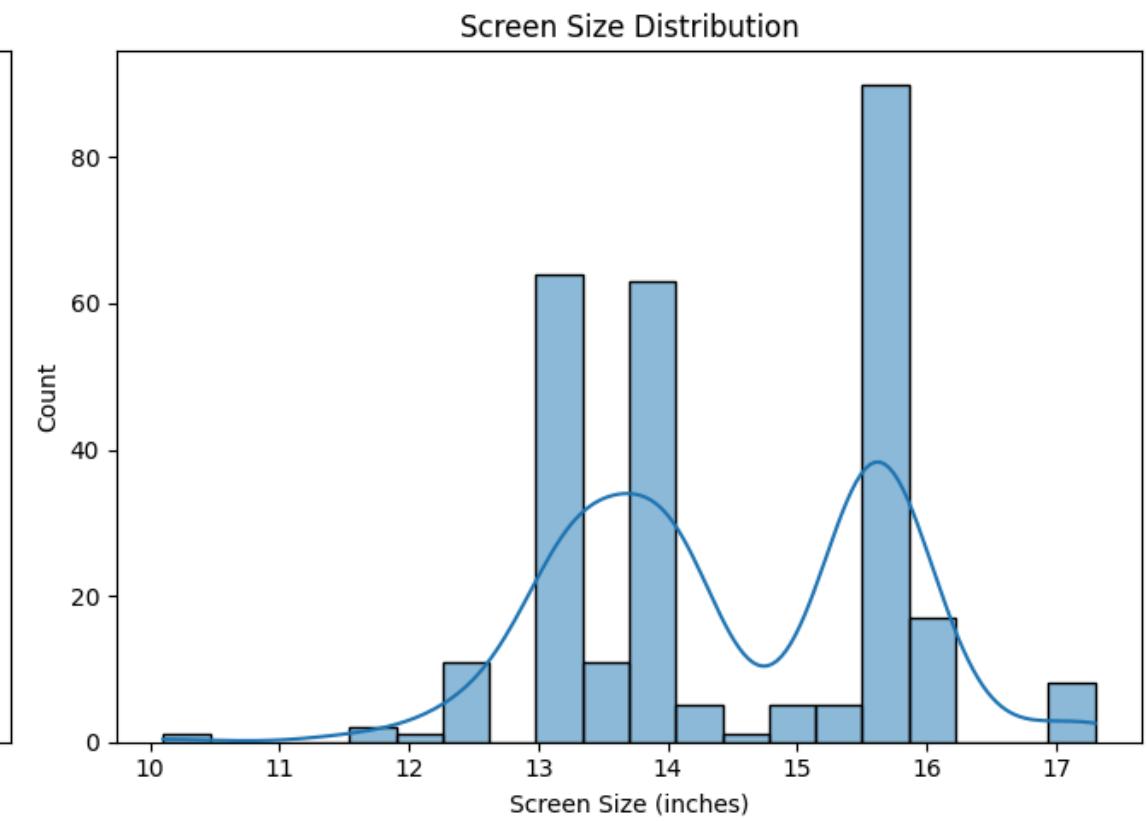
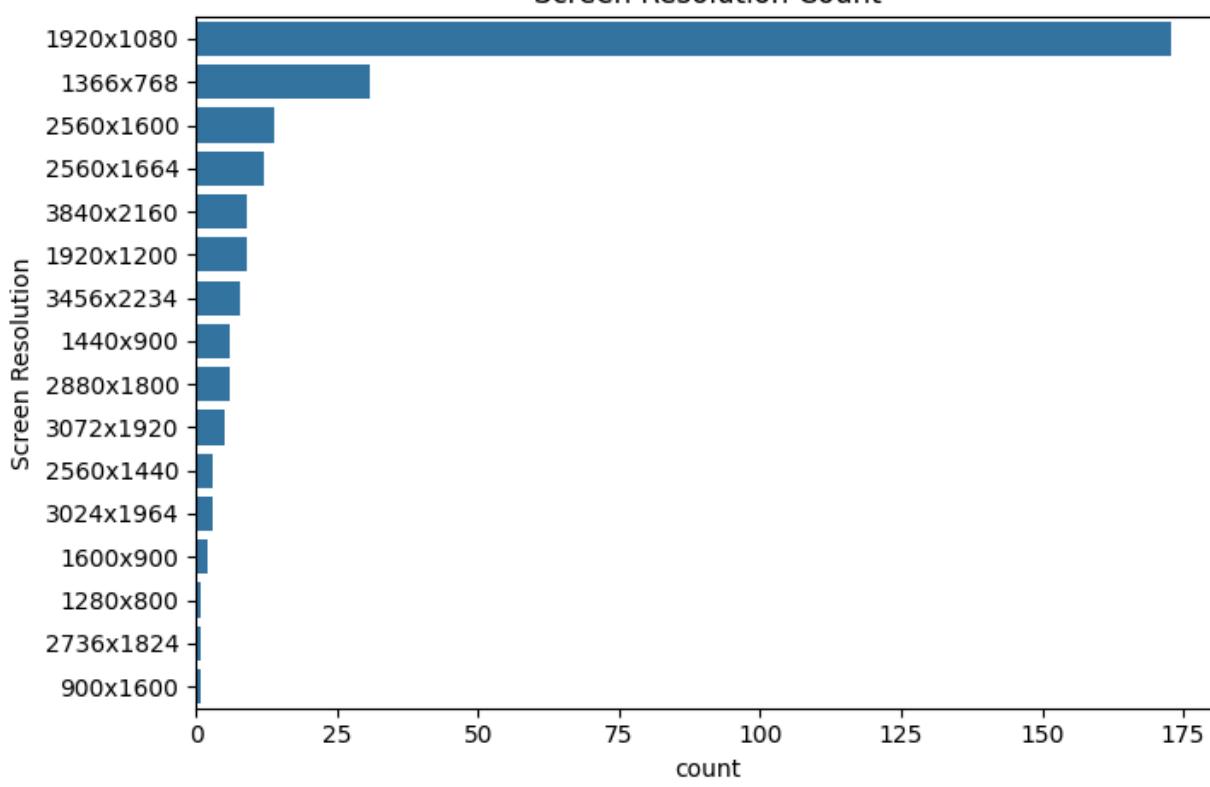


EDA Continued:

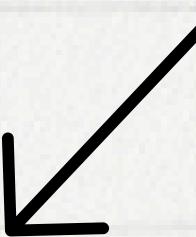
- Most Common Laptop GPU Brands
- Common Screen Size & Resolutions



- Most Laptops use Intel integrated graphics of UHD series while for dedicated chips NVIDIA is popular
- Screen Size that's common among these laptops turns out to be around 15-16 inches
- Meanwhile, the resolution of most laptops is 1920x1080 (1080p FULL HD)



Regression Analysis:



We narrow down the most significant variables effect on our laptop prices. Our R^2 goes down a bit to 0.863 however, mostly because we might have still removed a few loosely related significant variables.

OLS Regression Results						
Dep. Variable:	Price (PKR)	R-squared:	0.863			
Model:	OLS	Adj. R-squared:	0.853			
Method:	Least Squares	F-statistic:	82.99			
Date:	Wed, 04 Dec 2024	Prob (F-statistic):	2.97e-101			
Time:	10:41:31	Log-Likelihood:	-3418.7			
No. Observations:	284	AIC:	6879.			
Df Residuals:	263	BIC:	6956.			
Df Model:	20					
Covariance Type:	nonrobust					
		coef	std err	t	P> t	[0.025 0.975]
const		-1.861e+07	2.3e+06	-8.099	0.000	-2.31e+07 -1.41e+07
Manufacture Year		9238.1972	1139.574	8.107	0.000	6994.348 1.15e+04
SSD (GB)		87.5854	16.072	5.449	0.000	55.938 119.232
RAM (GB)		3566.5966	444.045	8.032	0.000	2692.261 4440.932
Brand_Microsoft		5.985e+04	2.54e+04	2.356	0.019	9840.549 1.1e+05
Laptop Type_Ideapad		1.806e+05	4.29e+04	4.205	0.000	9.6e+04 2.65e+05
Laptop Type_OMEN		1.109e+05	2.28e+04	4.870	0.000	6.6e+04 1.56e+05
Laptop Type_Tiger_Lake		5.945e+04	1.85e+04	3.211	0.001	2.3e+04 9.59e+04
Processor Type_M1_Pro		1.207e+05	3e+04	4.030	0.000	6.17e+04 1.8e+05
Processor Type_M2_Pro		1.81e+05	2.95e+04	6.127	0.000	1.23e+05 2.39e+05
Processor Type_M3		1.898e+05	2.25e+04	8.447	0.000	1.46e+05 2.34e+05
Processor Type_M3_Pro		3.305e+05	2.2e+04	15.028	0.000	2.87e+05 3.74e+05
Processor Type_i5		-5678.0736	9170.174	-0.619	0.536	-2.37e+04 1.24e+04
Processor Type_i7		3623.5714	9302.420	0.390	0.697	-1.47e+04 2.19e+04
Processor Type_intel (R) Core (TM) ultra 7 155H 3.80GHz		2.089e+05	4.41e+04	4.737	0.000	1.22e+05 2.96e+05
Operating System_Windows		-3237.1399	7471.972	-0.433	0.665	-1.79e+04 1.15e+04
Operating System_Windows_10_Pro		-2903.4037	6786.742	-0.428	0.669	-1.63e+04 1.05e+04
Colour_Platinum		-3.012e+04	3.92e+04	-0.769	0.443	-1.07e+05 4.7e+04
GraphicsCardBrand_Intel		-2.158e+04	6099.628	-3.539	0.000	-3.36e+04 -9574.153
GraphicsCardBrand_M1		7.773e+04	1.86e+04	4.174	0.000	4.11e+04 1.14e+05
GraphicsCardBrand_M2		1.422e+05	1.92e+04	7.415	0.000	1.04e+05 1.8e+05
Omnibus:	92.668	Durbin-Watson:	1.656			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	392.900			
Skew:	1.303	Prob(JB):	4.82e-86			
Kurtosis:	8.139	Cond. No.	1.87e+06			

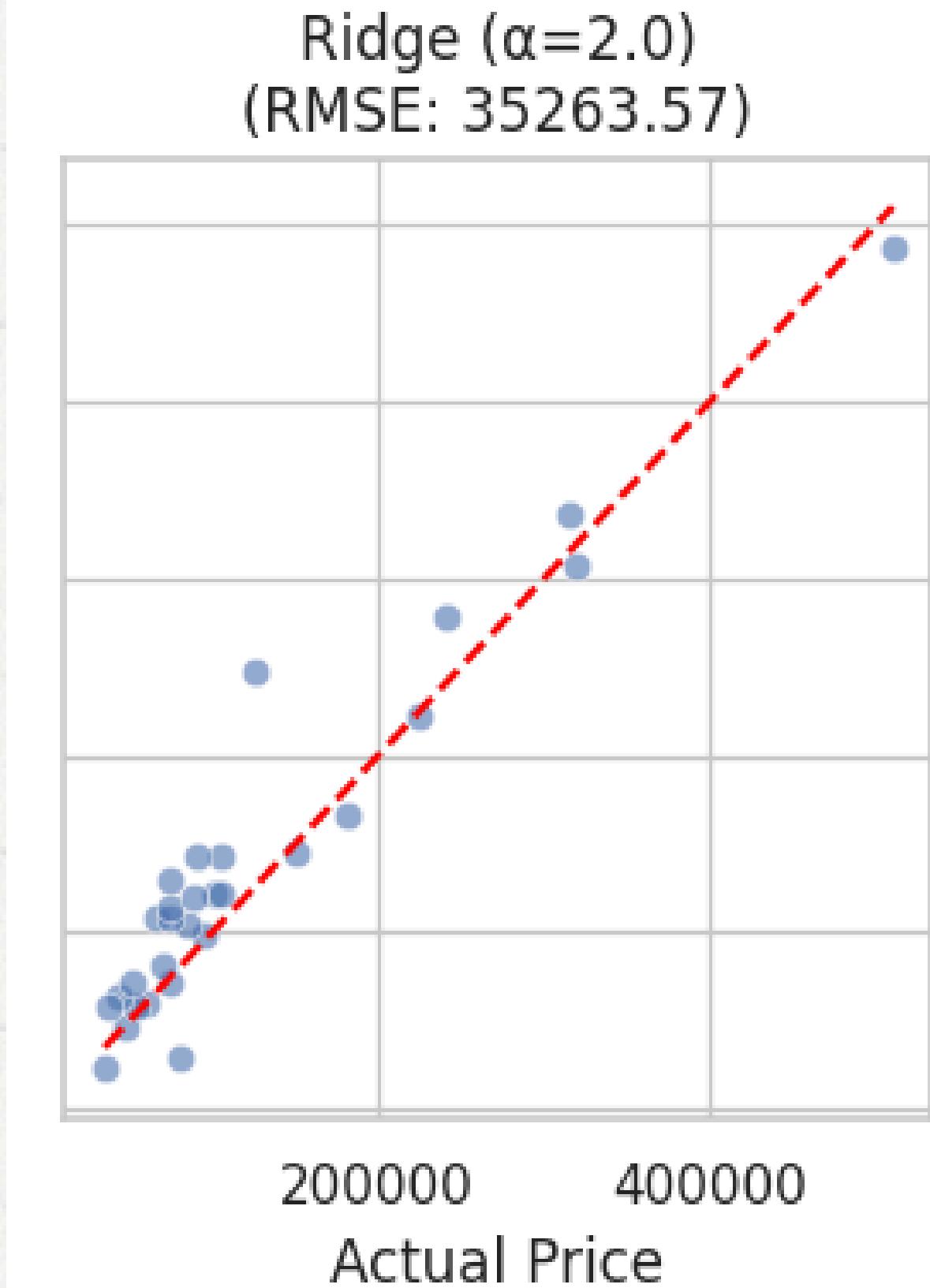
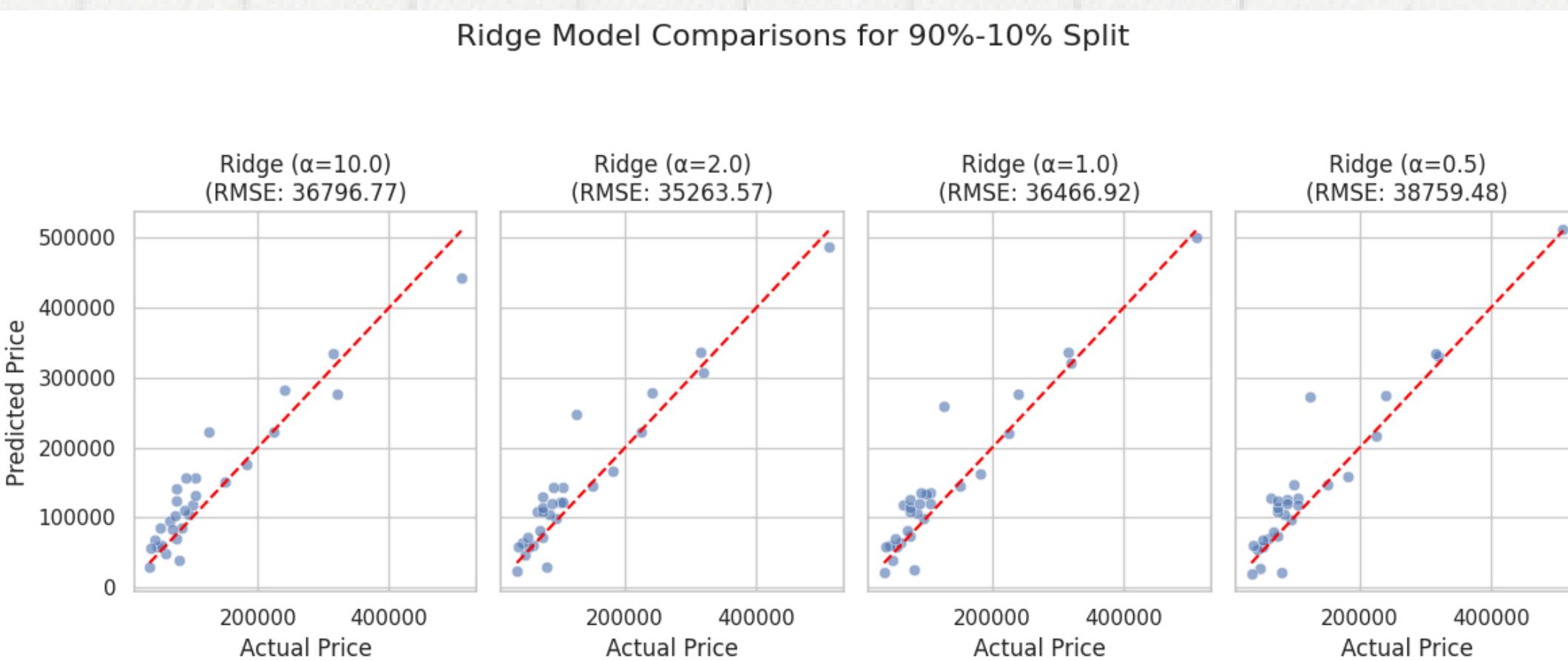
OLS Regression Price Prediction:

- **User Input:** A function is provided to predict laptop price based on user input for various features.
 - **Depreciation Model:** An analysis of value decrease per year for different brands.

- R² score is decent. Close to 1. This means that the regression prediction perfectly fits the data!!

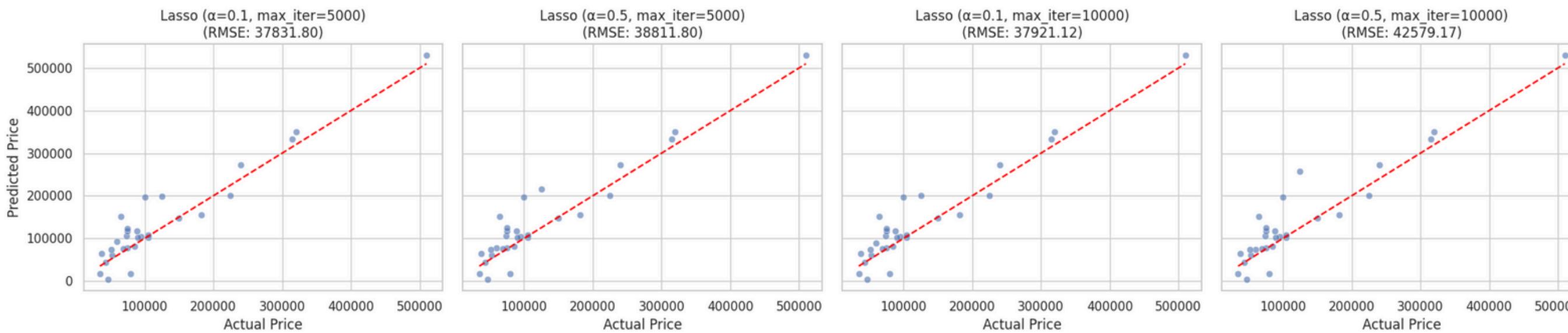
Model Training:

- Models Used: Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, and Random Forest.
- Best Model: Ridge Regression was found to be the best model with an RMSE of ~35,000, using the 90-10% split while keeping the $\alpha = 2.0$ balancing performance and simplicity.
- Due to the very nature of the data, the models showed arbitrary results commonly found due to the element of variation



Lasso, Decision Tree, and Forestry

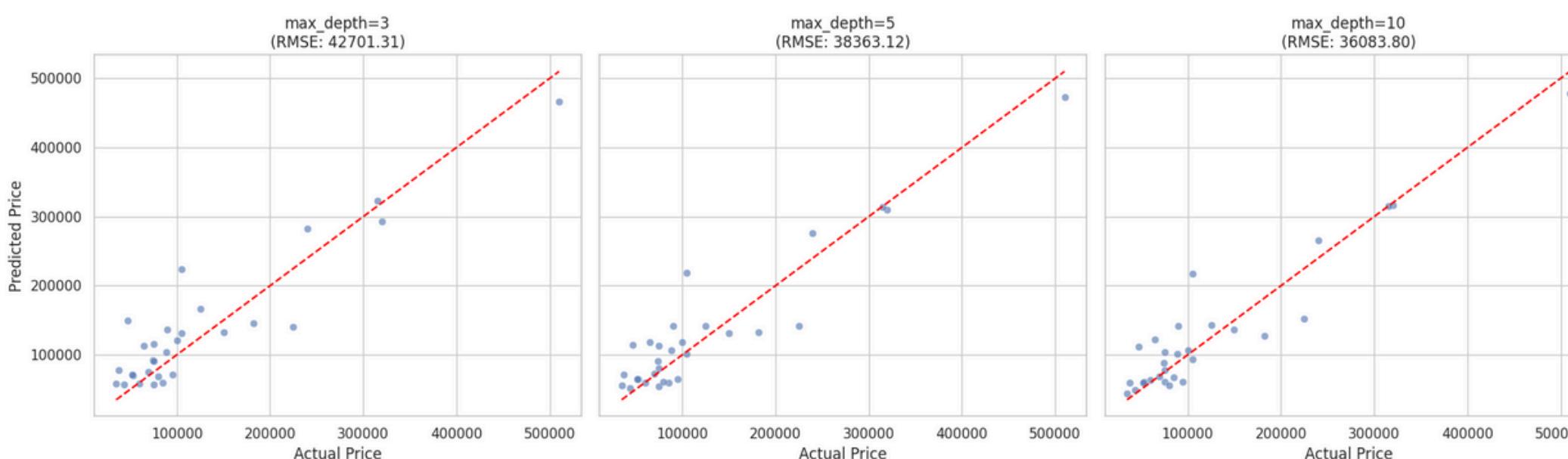
Lasso Model Comparisons for 90%-10% Split



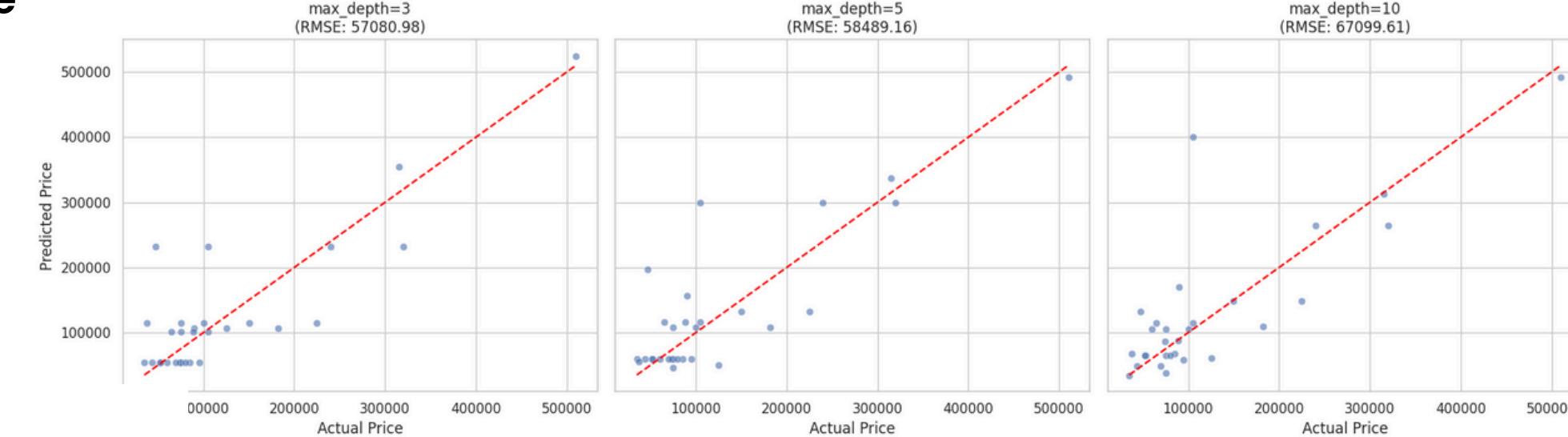
Decision Tree Performance (Train-Test Split: 90%-10%)

- We get the best results during model training by using the 90-10% split.
- The best Lasso Model turns out to be around $\alpha=0.1$
- Similarly, the best model for Decision Tree is also around $\alpha=0.1$
- Meanwhile the Random Forest model performs its best at maximum depth of 10.

Random Forest Performance (Train-Test Split: 90%-10%)

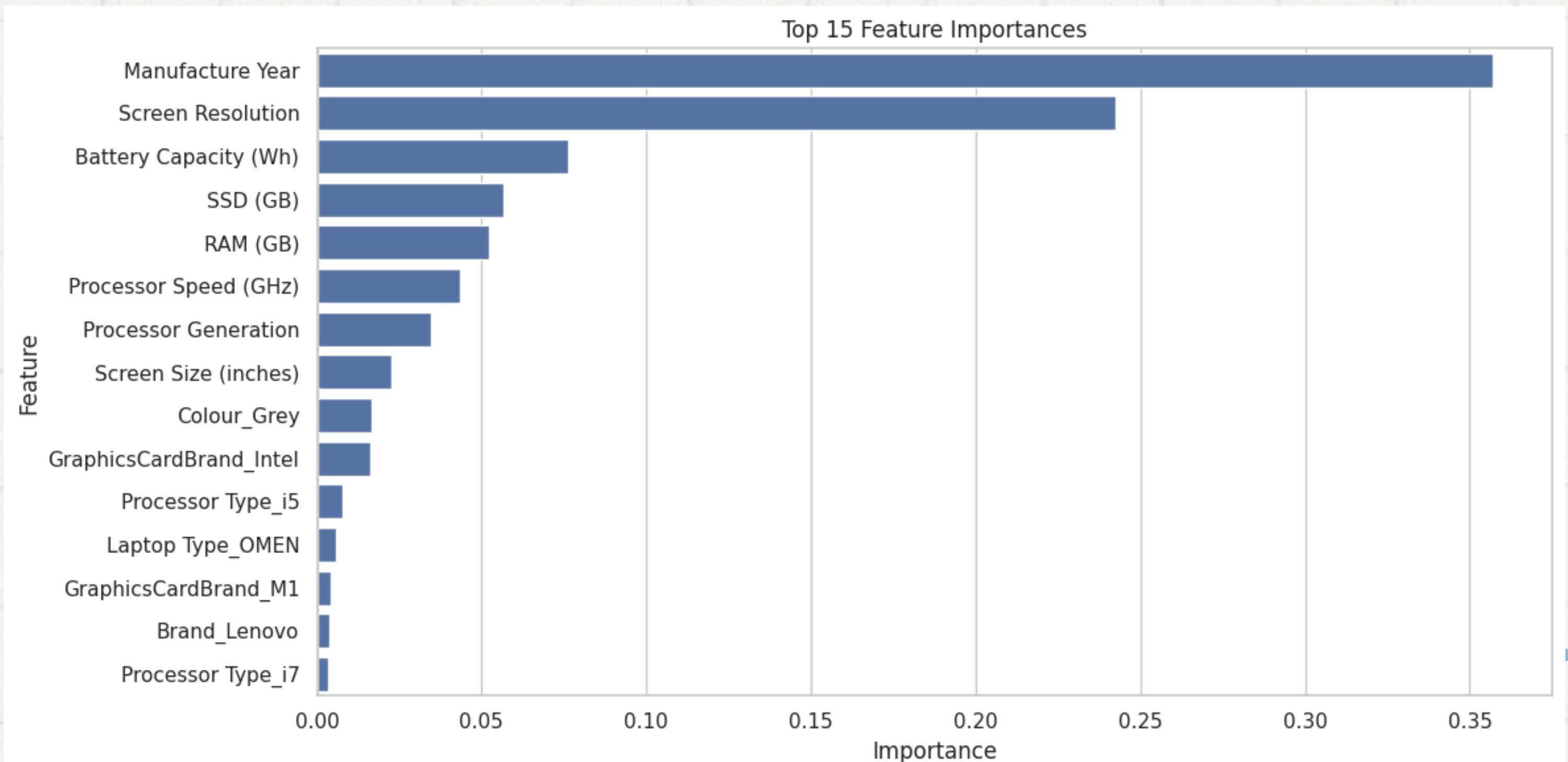


- When compared to Lasso, Forestry, and Ridge's model, Decision trees give poor results as the standard best RMSE value is higher than all the other training models.
- However, among these three, random forest performs better as it extracts more depth, making it a better alternative.



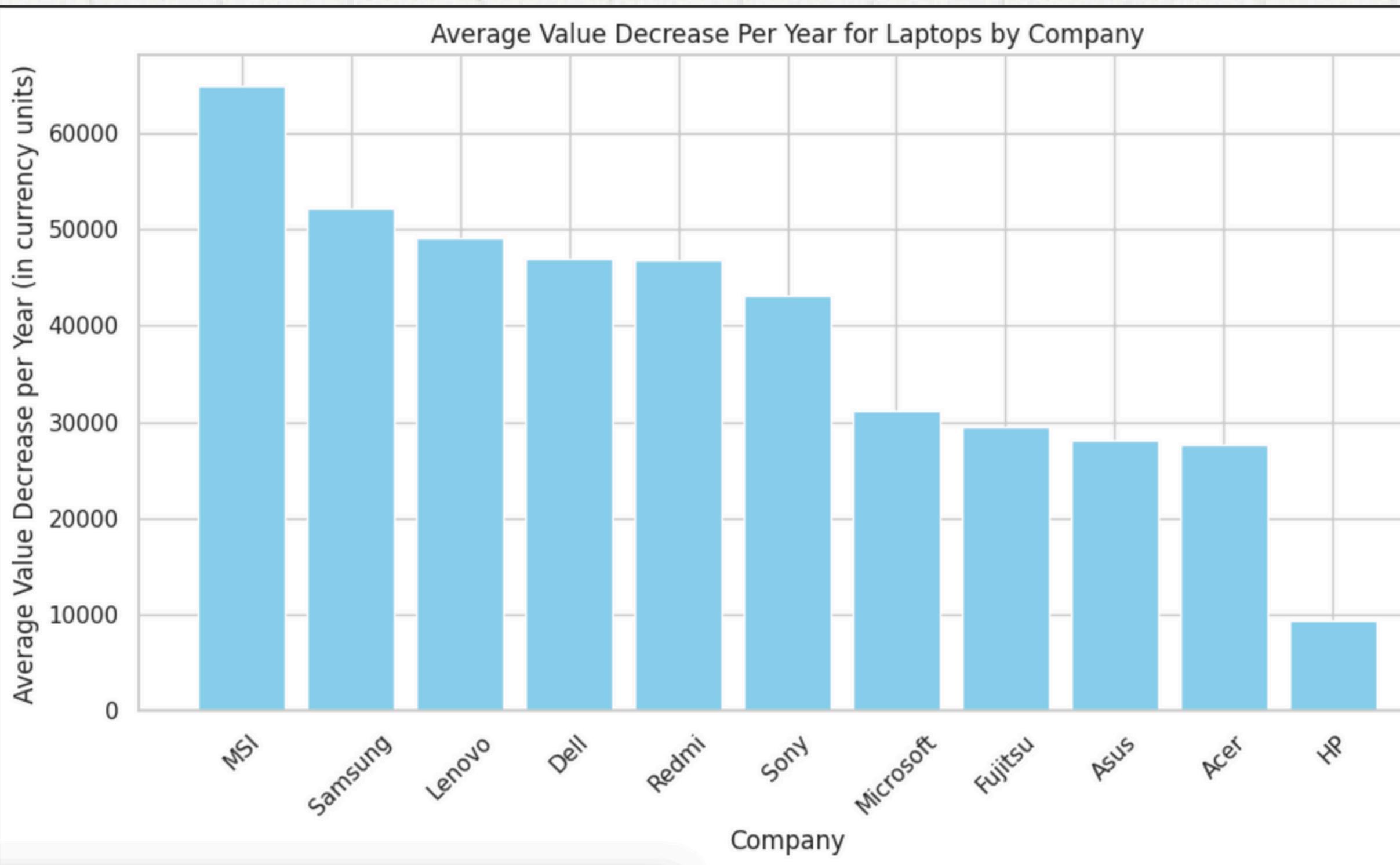
Most Important Features

- Furthermore, as the feature importance graph suggests above, the most important features seem to be Manufacture_Year, Screen_Resolution and Battery Capacity with SSD and RAM having an effect as well on the price of used Laptops.



Depreciation Predictor

We were able to successfully plot a Depreciation Prediction Graph for each Laptop Brand. If a user ever wants to sell their used laptop for the highest price, they should opt for HP!! And the data and ML support it!!



Final Conclusion:

- Overall, we have successfully made a Machine Learning model that predicts the price of a used laptop trained on the data from Pakistani Laptop Market
- Currently, we are facing some feasibility constraints regarding the pragmatic implementation of our model.
- Our main goal was to make a form in which we could ask a user about the specs of their desired laptop, but the coding environment in which we're making our model doesn't allow a user to explicitly enter a plethora of values.
- This issue is particularly due to the categorical variables which need to be translated again by the user which is out of scope for now.

**Thank you
very much!**