LapTop Price Prediction REPORT

Name: Ajay Kumar

Laptop Price Prediction Report

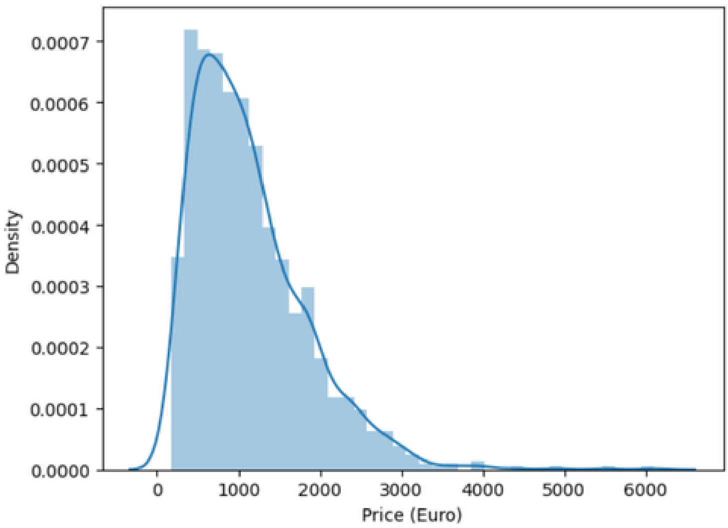
Overview
This project aims to predict laptop prices based on various laptop specifications such as brand, type, CPU, RAM, screen resolution, storage type, and GPU. The dataset consists of 1,275 records with 15 initial features, including price as the target variable.

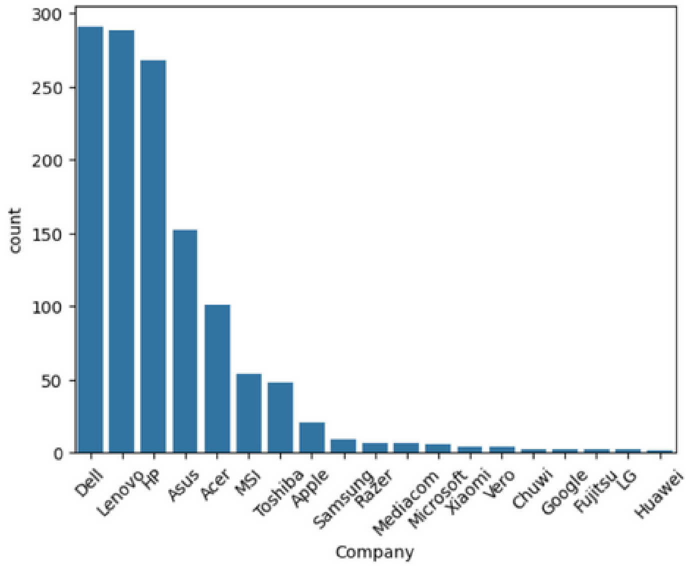
Data Analysis

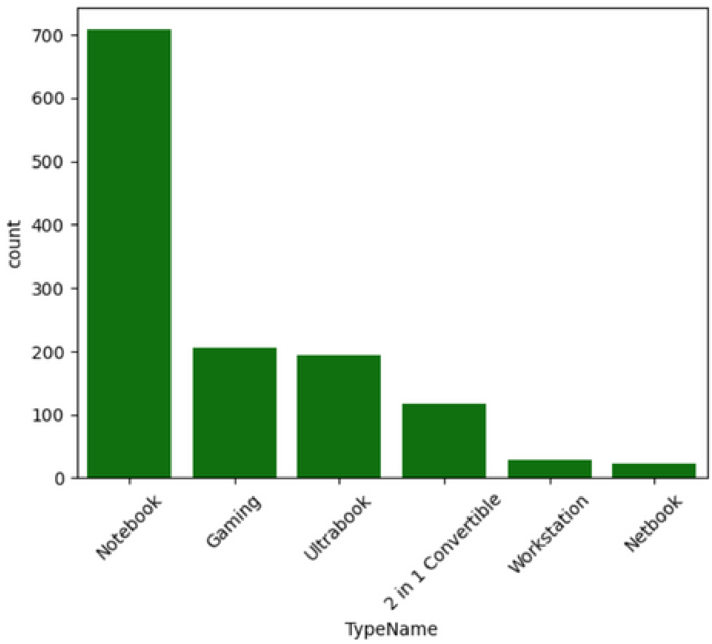
Data Cleaning and Preprocessing:
The dataset includes features such as laptop brand, screen size, processor, memory, and operating system.
The dataset was inspected for missing values, duplicates, and other inconsistencies. No missing values or duplicates were found.

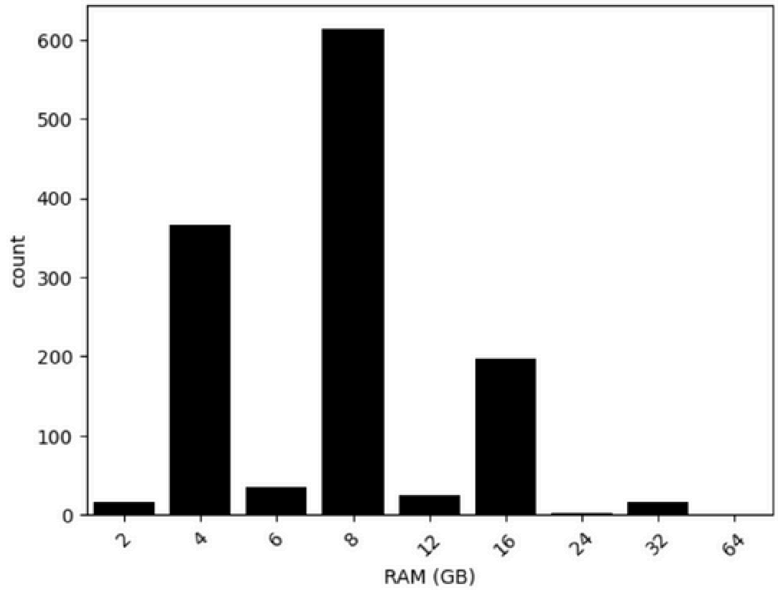
Exploratory Data Analysis (EDA):

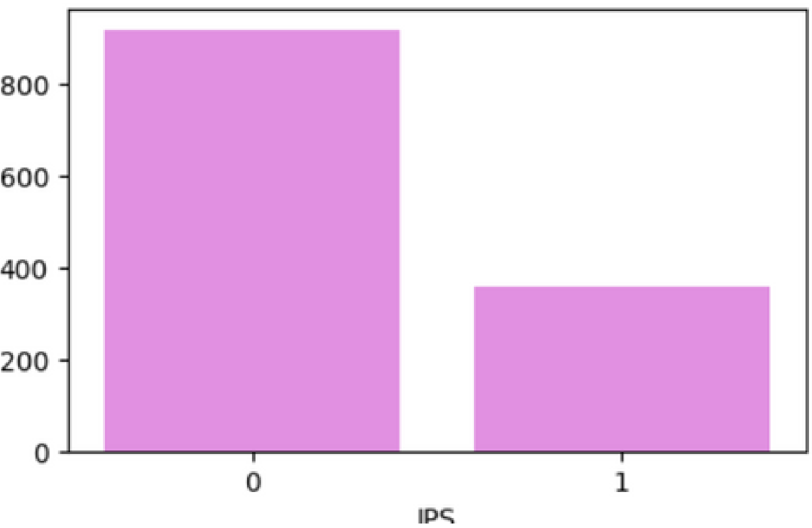
Price Distribution: The price distribution shows that most laptops are priced lower, while a few high-end models significantly increase the average.
Company Distribution: Brands like Dell, Lenovo, HP, Asus, and Acer are the most common manufacturers in the dataset.
Type Distribution: Notebooks make up the majority of laptops, followed by gaming and ultrabook models.
Screen Size: Laptops with 15.6-inch screens are the most common, followed by 14-inch and 17.3-inch models.
Touchscreen and IPS Panel: Only a minority of laptops include a touchscreen or an IPS panel, and these features often correspond to higher-priced models.
RAM and Storage: Laptops with 8GB of RAM are the most prevalent, and SSD storage is more common than HDD, reflecting modern trends in laptop design.
Operating System: Windows 10 is the dominant operating system, followed by macOS and Linux distributions.

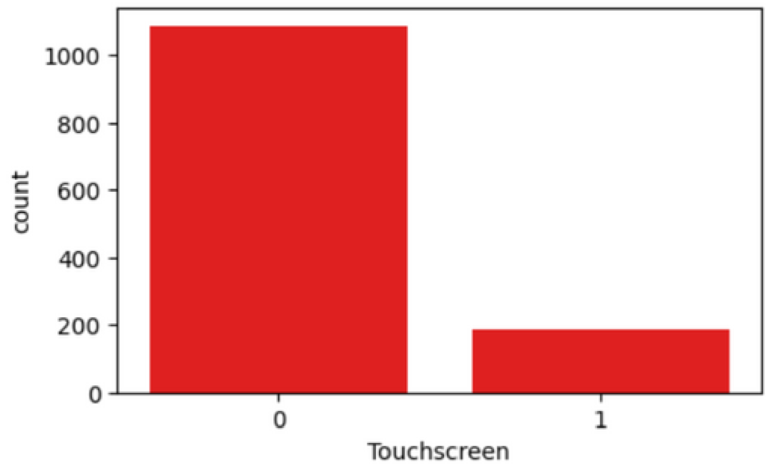

Density
0.0000
0.0001
0.0002
0.0003
0.0004
0.0005
0.0006
0.0007
0
1000
2000
3000
4000
5000
6000
Price (Euro)

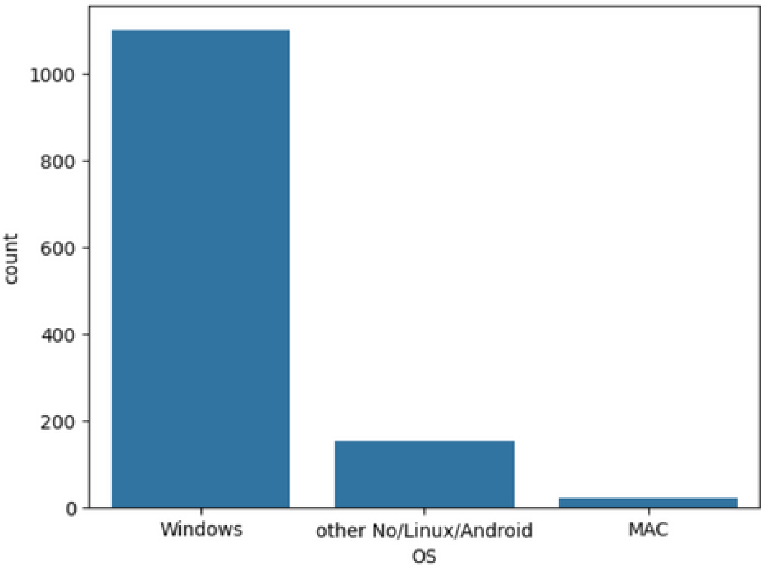

count
0
50
100
150
200
250
300
Dell
Lenovo
HP
Asus
Acer
MSI
Toshiba
Apple
Samsung
Razer
Mediacom
Microsoft
Xiaomi
Vero
Chuwi
Google
Fujitsu
LG
Huawei
Company

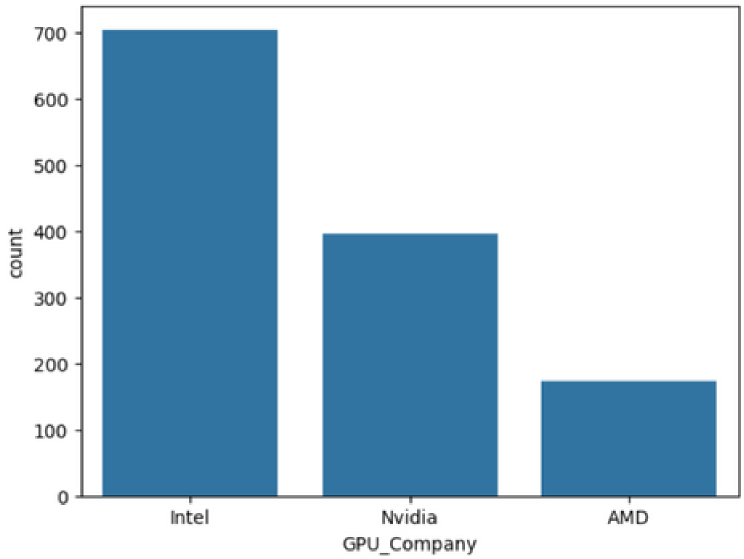

count
0
100
200
300
400
500
600
700
Notebook
Gaming
Ultrabook
2 in 1 Convertible
Workstation
Netbook
TypeName

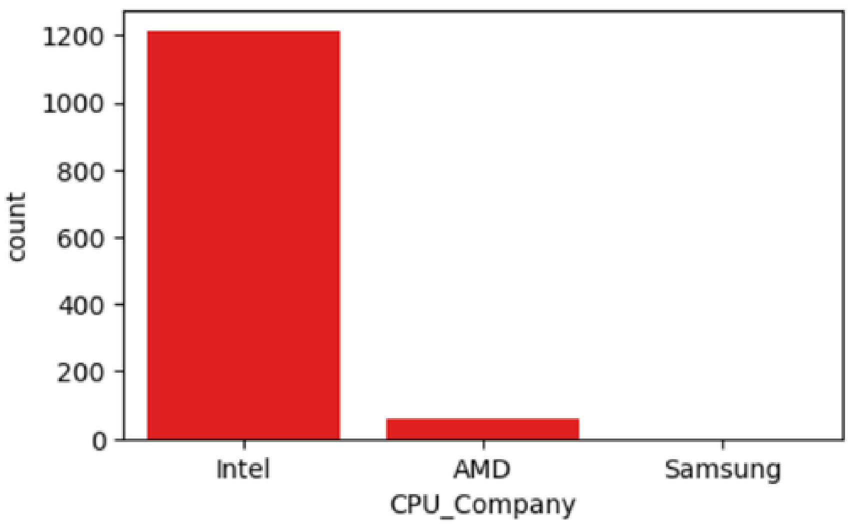

count
0
100
200
300
400
500
600
2
4
6
8
12
16
24
32
64
RAM (GB)

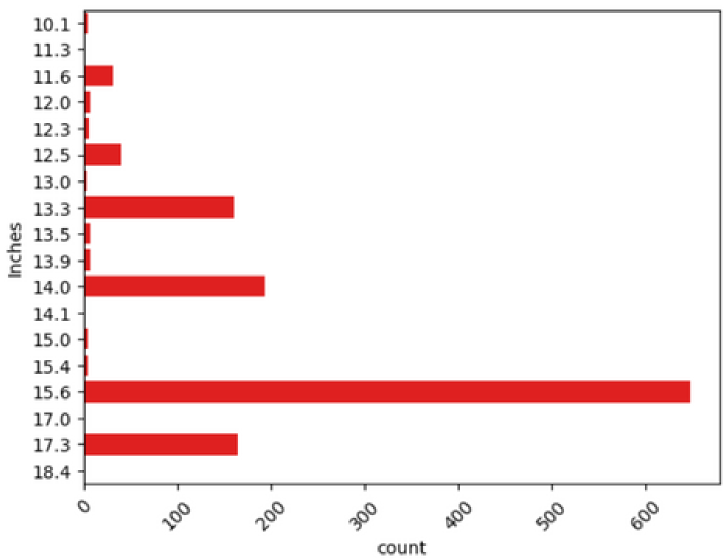

0
200
400
600
800
0
1
IPS

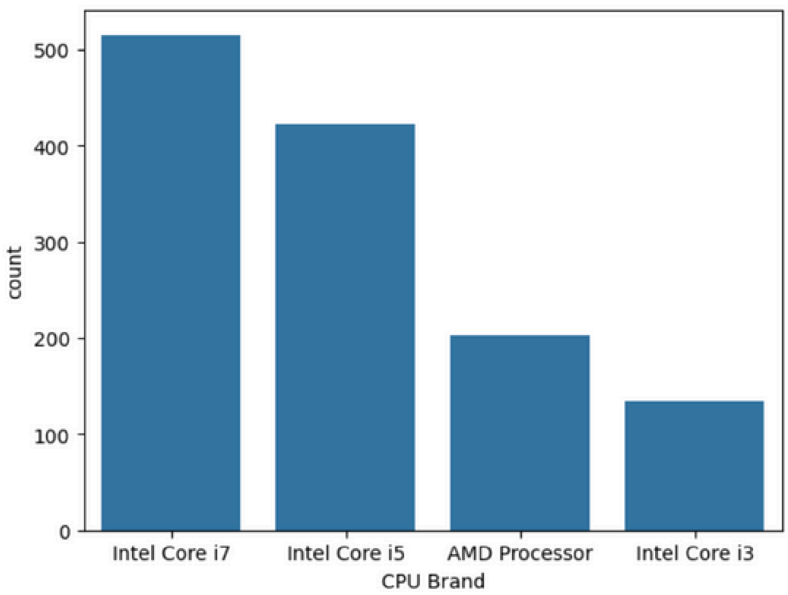

count
0
200
400
600
800
1000
0
1
Touchscreen


count
0
200
400
600
800
1000
Windows
other No/Linux/Android
MAC
OS


count
0
100
200
300
400
500
600
700
Intel
Nvidia
AMD
GPU_Company

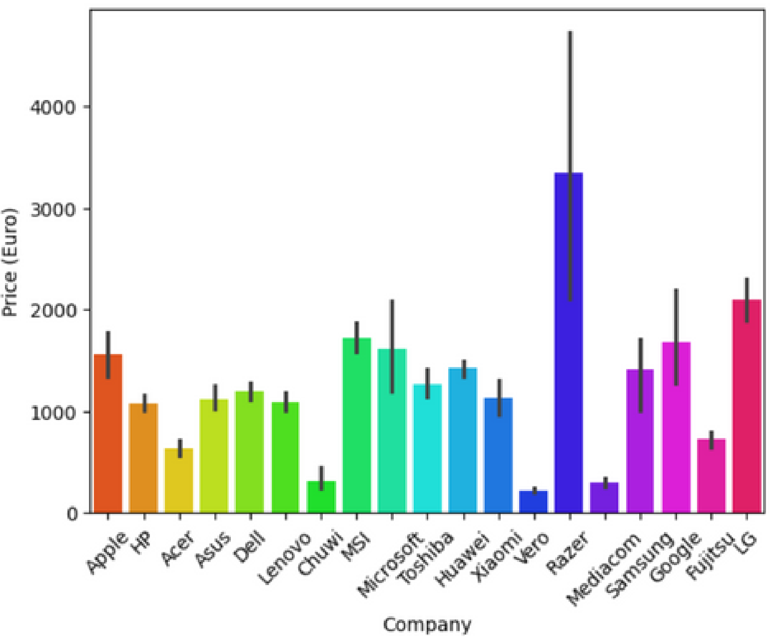

count
0
200
400
600
800
1000
1200
Intel
AMD
Samsung
CPU_Company

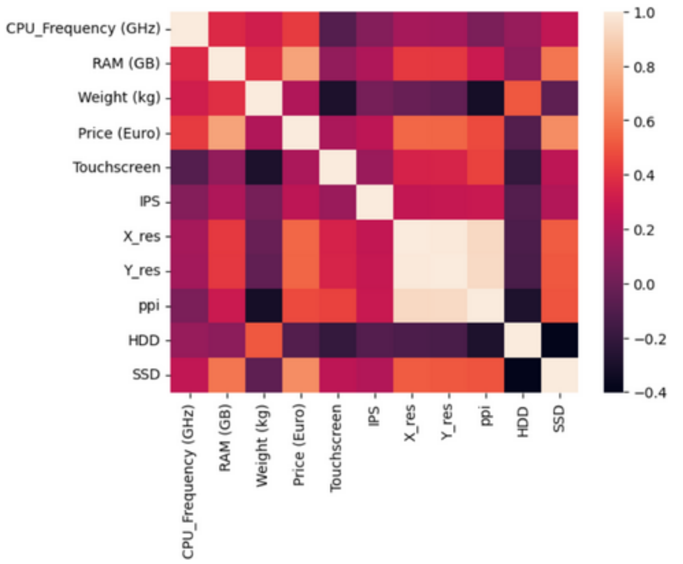

Inches
10.1
11.3
11.6
12.0
12.3
12.5
13.0
13.3
13.5
13.9
14.0
14.1
15.0
15.4
15.6
17.0
17.3
18.4
0
100
200
300
400
500
600
count

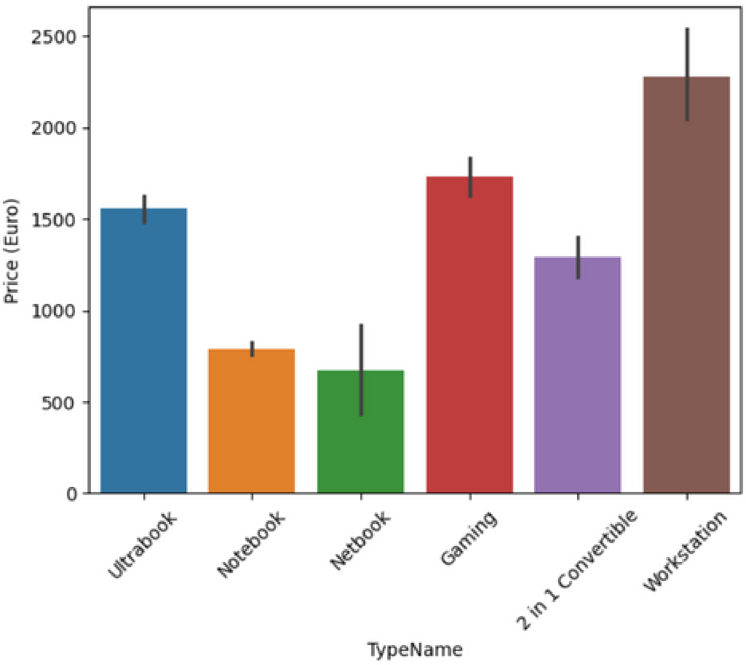

count
0
100
200
300
400
500
Intel Core i7
Intel Core i5
AMD Processor
Intel Core i3
CPU Brand

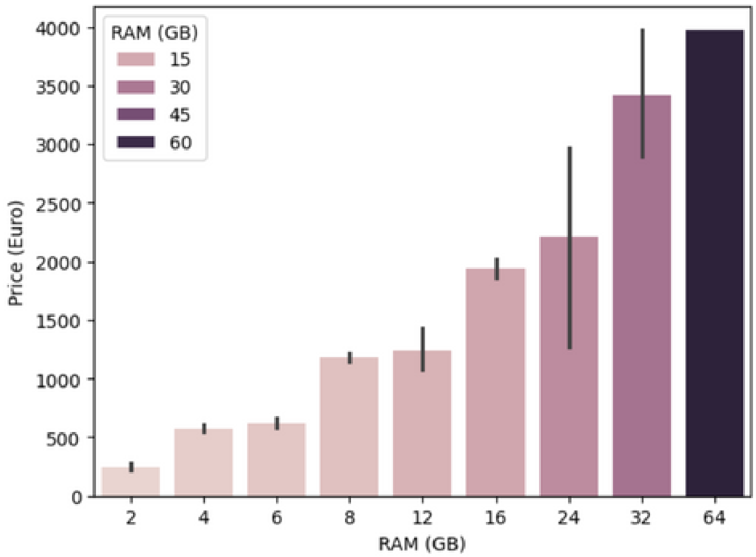
Correlations with Price:

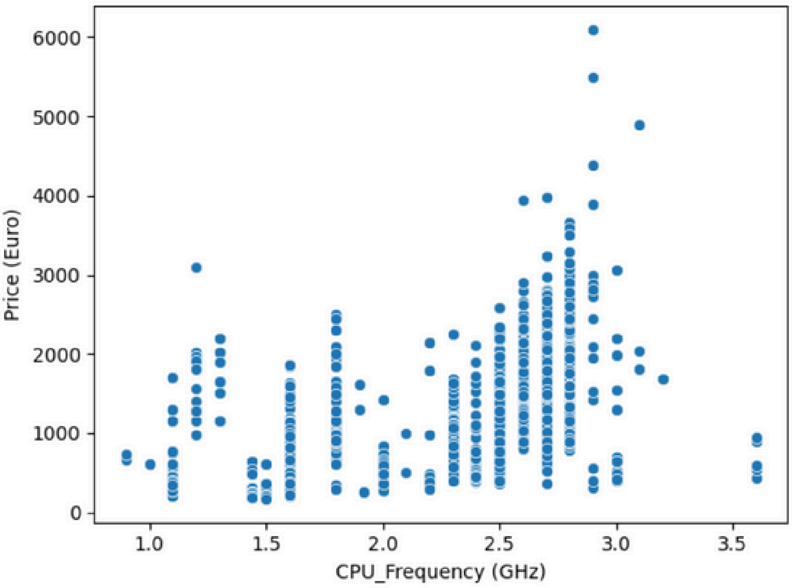
- RAM: The most significant correlation with price was observed for the amount of RAM, with higher RAM leading to higher prices.
- SSD: Laptops with SSD storage have higher prices compared to those with HDDs.
- Screen Resolution and PPI: Higher screen resolutions (X_res, Y_res) and pixel density (PPI) are positively correlated with price.
- CPU Frequency: Laptops with higher CPU clock speeds also tend to be more expensive.

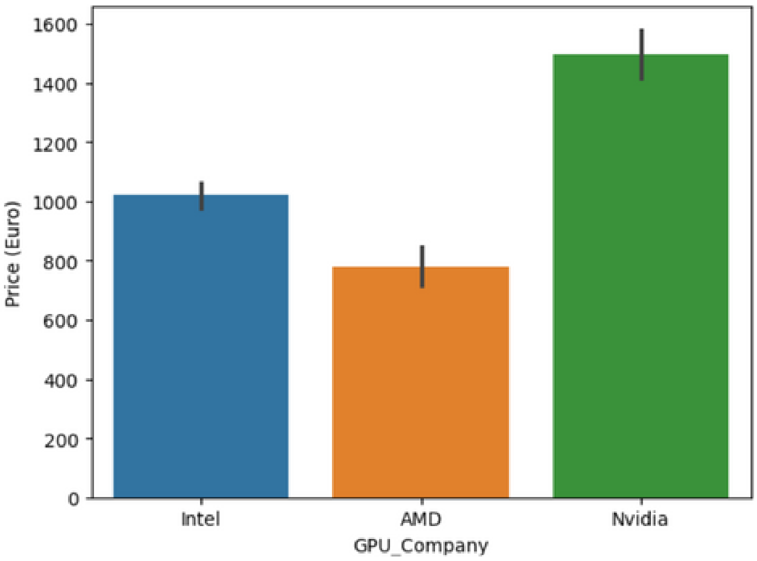


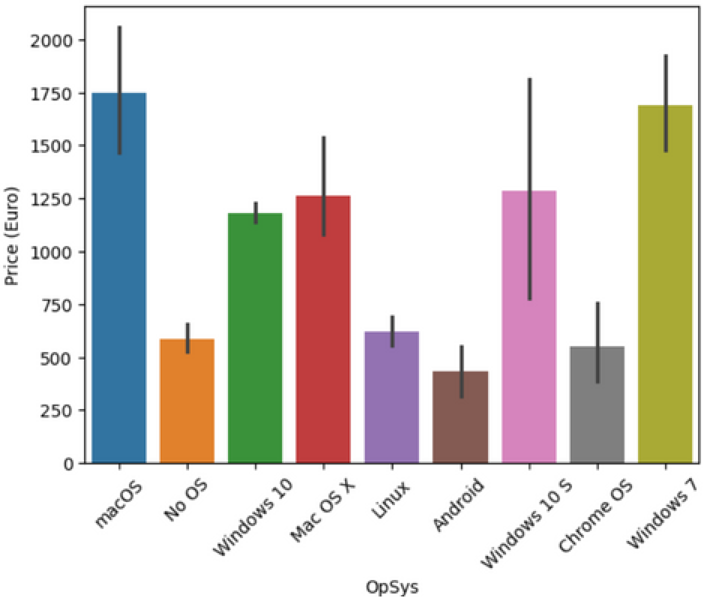


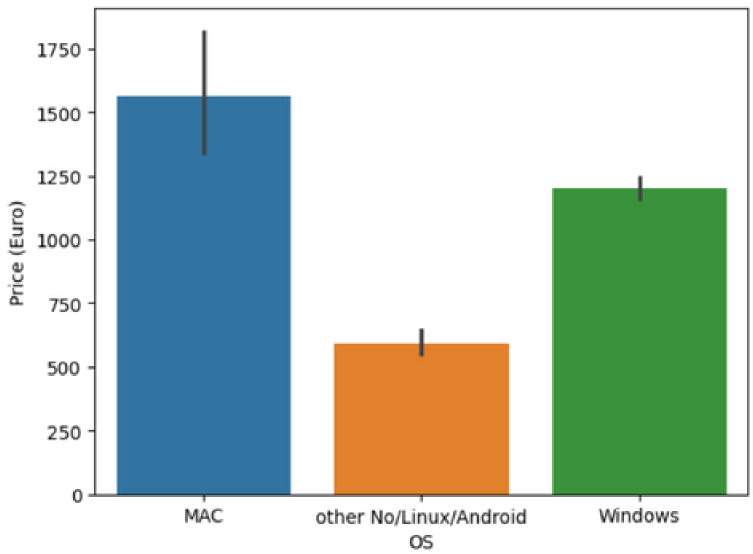


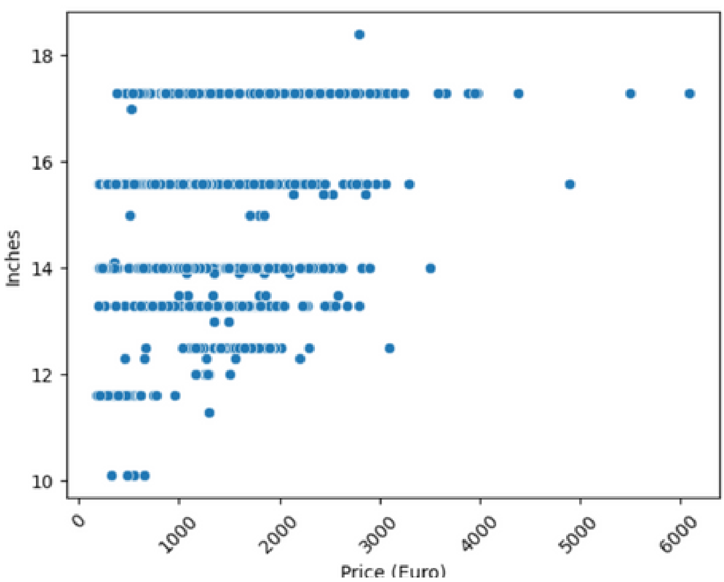


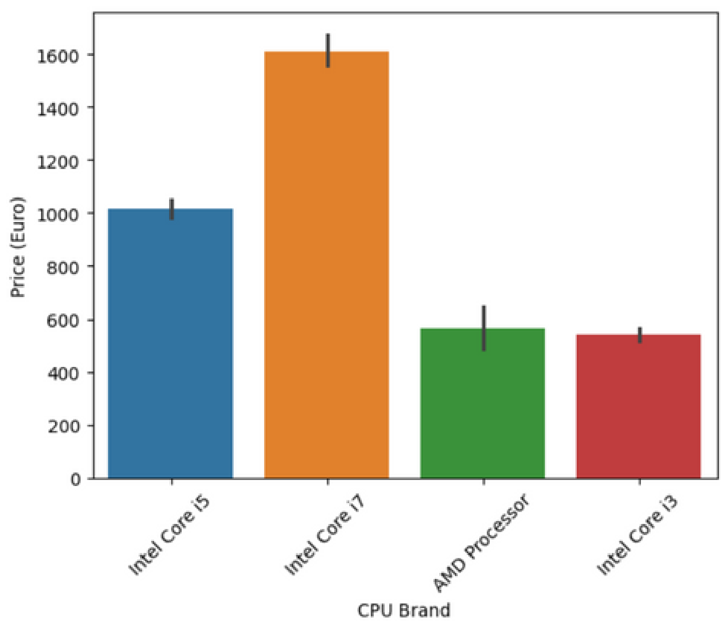

Price (Euro)
0
1000
2000
3000
4000
5000
6000
CPU_Frequency (GHz)
1.0
1.5
2.0
2.5
3.0
3.5


Price (Euro)
0
200
400
600
800
1000
1200
1400
1600
Intel
AMD
Nvidia
GPU_Company


Price (Euro)
0
250
500
750
1000
1250
1500
1750
2000
macOS
No OS
Windows 10
Mac OS X
Linux
Android
Windows 10 S
Chrome OS
Windows 7
OpSys


Price (Euro)
0
250
500
750
1000
1250
1500
1750
MAC
other No/Linux/Android
Windows
OS


Inches
10
12
14
16
18
0
1000
2000
3000
4000
5000
6000
Price (Euro)


Price (Euro)
0
200
400
600
800
1000
1200
1400
1600
Intel Core i5
Intel Core i7
AMD Processor
Intel Core i3
CPU Brand

Feature Engineering:

- A new feature, PPI (Pixels per Inch), was derived from the screen resolution and screen size, offering a better understanding of screen quality in relation to price.
- Storage types were broken down into separate categories for SSD, HDD, Flash Storage, and Hybrid drives.

Model Building and Prediction

For the prediction phase:

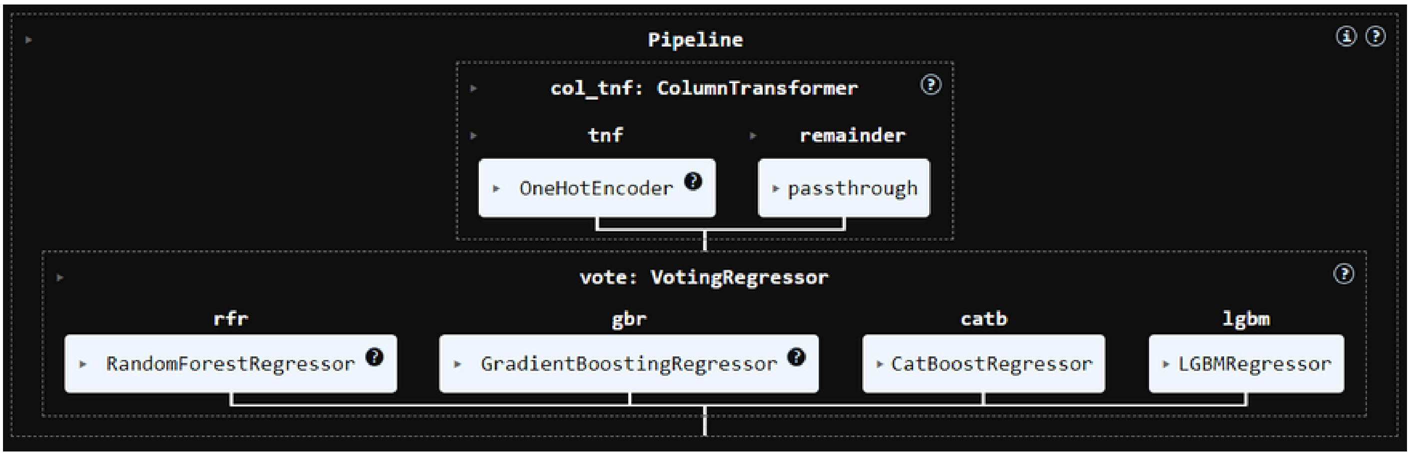
- The price was log-transformed to normalize its distribution and improve the model's performance.
- The dataset was split into training and test sets for model training using various regression models.

Pipeline Overview:

- The model pipeline consists of a ColumnTransformer and a Voting Regressor.

- **ColumnTransformer:**
  - It uses OneHotEncoder to transform categorical features while passing other columns through.

- **Voting Regressor:**
  - Four base models are included:
    - ***RandomForestRegressor (rfr)***
    - ***GradientBoostingRegressor (gbr)***
    - ***CatBoostRegressor (catb)***
    - ***LGBMRegressor (lgbm)***



Model Performance:

- The R^2 score achieved on the test set is 0.9002, indicating that the model explains approximately 90% of the variance in laptop prices.
- The Mean Absolute Error (MAE) is 0.138, which is a reasonable margin, reflecting how far the predictions deviate from actual values on average.

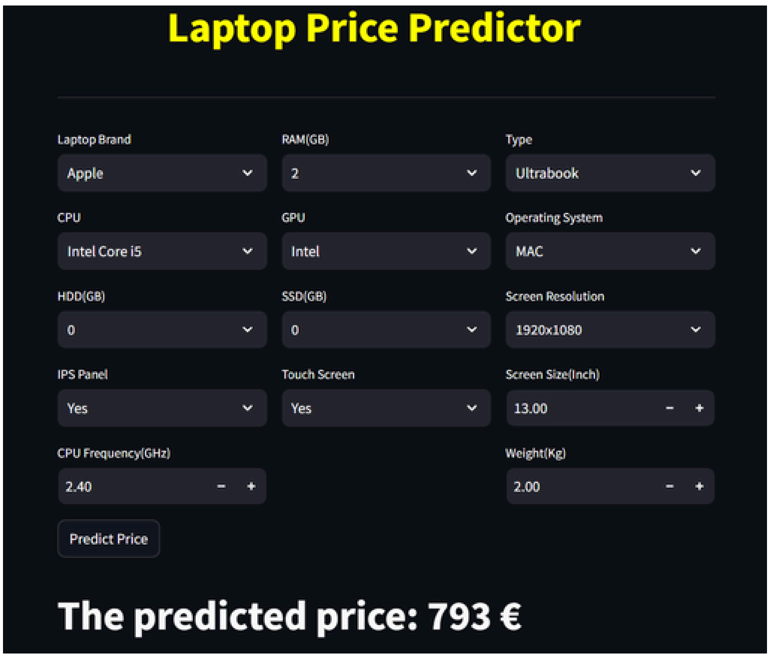
```
y_pred = pipe.predict(X_test)

print('R2 score',np.round(r2_score(y_test,y_pred),5))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.9002
MAE 0.13819942117623965
```

Deployment:

The trained model has been deployed as a web application using Streamlit and is hosted on Hugging Face Spaces. Users can input the laptop specifications, such as brand, RAM, storage type, and screen size, to receive an estimated price prediction instantly.



Conclusion

The Laptop Price Prediction project successfully predicts laptop prices based on various specifications, including brand, type, RAM, storage, screen size, and CPU. By leveraging a robust machine learning pipeline that integrates a Voting Regressor with models like RandomForestRegressor, GradientBoostingRegressor, CatBoostRegressor, and LGBMRegressor, the model achieved an impressive R^2 score of 0.90, explaining 90% of the variance in prices. The Mean Absolute Error (MAE) of 0.138 reflects accurate and reliable predictions.

Through feature engineering and careful preprocessing, important factors like RAM, storage type, screen resolution, and processor type were identified as key contributors to laptop prices. This project demonstrates the power of ensemble learning for handling complex regression tasks with heterogeneous data.

The model is deployed on Hugging Face Spaces using Streamlit, allowing users to input laptop specifications and receive real-time price predictions, providing practical utility in evaluating laptop prices for various configurations.

Thank You