Beginner ML
GDGoC UIN JAKARTA

ICN

Machine Learning Final Project

Laptop Price Prediction

our Team

AHMAD IHSAN
TEAM MODEL

KEINO AFLAH ZAHIRY
TEAM DATA

NAZWA RAHMA SUGETHA
TEAM DATA

ZAINUDDIANSAH
TEAM MODEL

Project Overview

Problems:

- Students face a significant challenge in selecting a suitable laptop that fits their needs and budget due to the overwhelming number of options available on the market.
- Comparing technical specifications between different models is often confusing, leading to uncertainty and a lack of confidence in making the best choice.
- The absence of a data-driven tool to predict laptop prices based on their specifications makes the entire decision-making process inefficient and prone to guesswork.

Goals:

- To build a machine learning model that accurately predicts laptop prices based on key specifications such as RAM, storage, screen size, and processor.
- To deliver trustworthy price estimations that empower students to easily select the right laptop that aligns with their specific needs and budget.
- To attain a robust model performance with an R-squared (R^2) score of approximately 85% by implementing and evaluating algorithms such as Linear Regression, Decision Tree, and Random Forest.

Key Components

Strategy to Reach our Goals

Describing Data

- Understanding the dataset's structure (columns, data types, data noise).
- Identifying initial relationships between features and price.
- Ensuring all essential laptop specification information is available for analysis.

Exploratory Data Analysis

- Visualizing data to understand distributions, outliers, and feature correlations.
- Handling outliers in the target variable.
- Using correlation analysis to find features relevant to the target.

Feature Engineering

- Creating new, more informative features.
- Encoding categorical features.
- Standardizing numerical features.

Modelling + Evaluation

- Building several models for price prediction.
- Using Cross-Validation to evaluate model performance.
- Measuring accuracy using MSE, RMSE, MAE, & R2 Score metrics.
- Selecting the best model based on combined evaluation scores.

Describing Data

- Data Quality & Noise: The dataset is relatively clean, however, several outliers were identified in the price and feature variables.
- Data Types: The dataset contains a mix of numerical features (e.g., RAM, Weight) and categorical features (e.g., Company, Operating System).
- Data Volume: The dataset consists of 1303 rows (records) and 13 columns (features).
- Target Variable: The target variable for prediction is Price_euros, representing the laptop's price in Euros.

	laptop_ID	Company	Product	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price_euros
0	1	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	1339.69
1	2	Apple	Macbook Air	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	898.94
2	3	HP	250 G6	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	575.00
3	4	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	2537.45
4	5	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	1803.60

```
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   laptop_ID         1303 non-null   int64
 1   Company           1303 non-null   object
 2   Product           1303 non-null   object
 3   TypeName          1303 non-null   object
 4   Inches            1303 non-null   float64
 5   ScreenResolution  1303 non-null   object
 6   Cpu               1303 non-null   object
 7   Ram               1303 non-null   object
 8   Memory            1303 non-null   object
 9   Gpu               1303 non-null   object
 10  OpSys             1303 non-null   object
 11  Weight            1303 non-null   object
 12  Price_euros       1303 non-null   float64
dtypes: float64(2), int64(1), object(10)
```

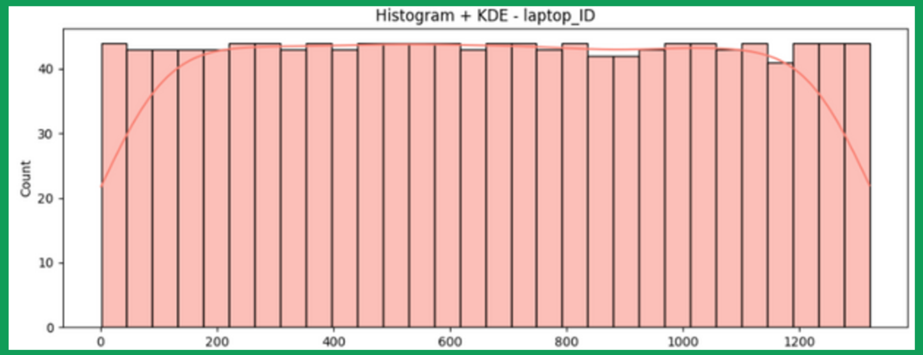
```
Jumlah nilai kosong per kolom:
laptop_ID           0
Company             0
Product             0
TypeName            0
Inches              0
ScreenResolution    0
Cpu                 0
Ram                 0
Memory              0
Gpu                 0
OpSys               0
Weight              0
Price_euros         0
dtype: int64

Jumlah data duplikat: 0
```

```
Jumlah data: (1303, 13)
```

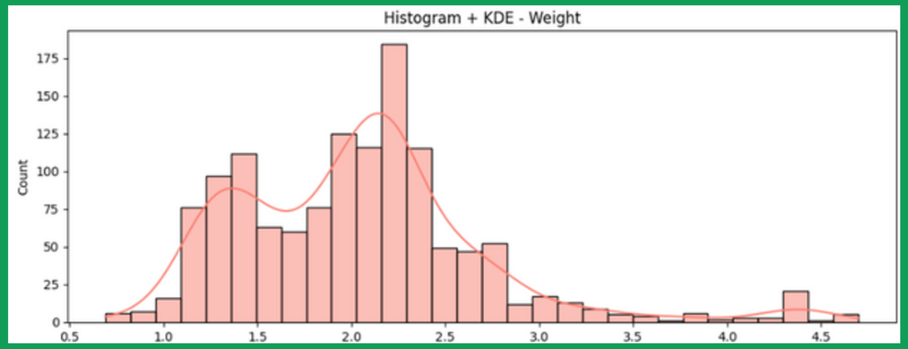
EDA

Univariate - numeric feature

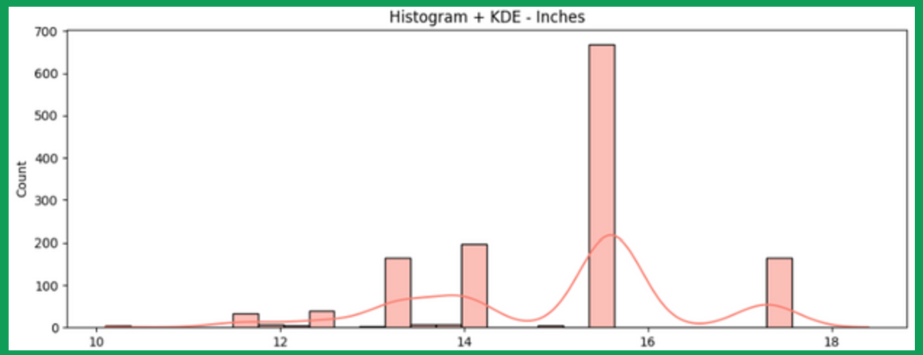
A clear correlation was found between a laptop's price and its specifications, where premium features like larger screens, increased RAM, lighter weight, and high-capacity SSDs consistently lead to a higher cost.



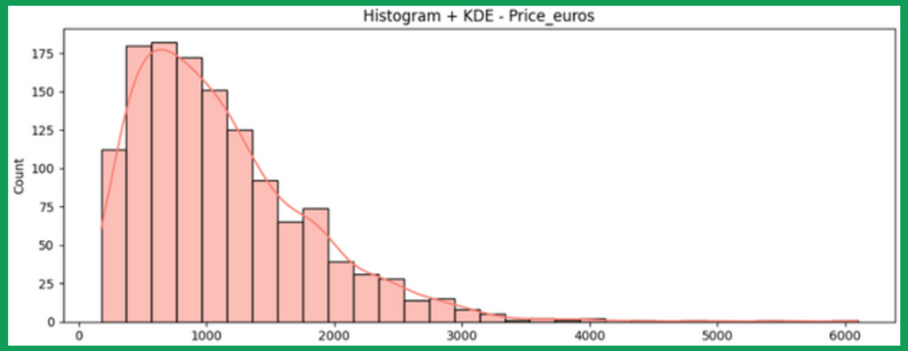
Laptop_ID



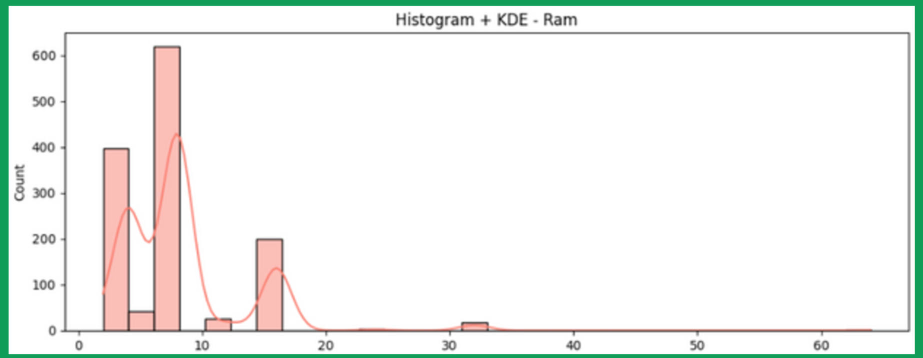
Weight



Inches



Price in euro



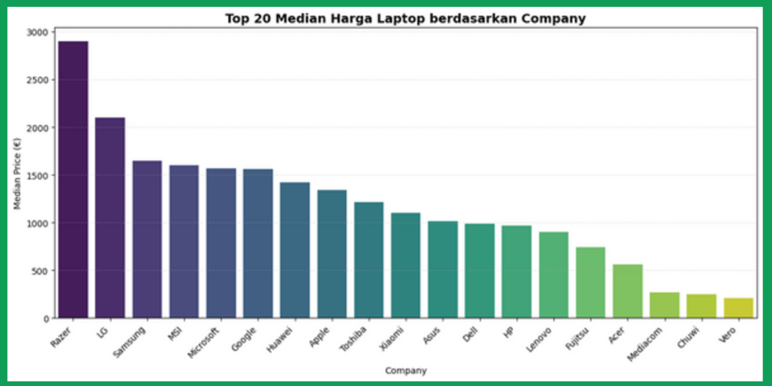
RAM

*median value

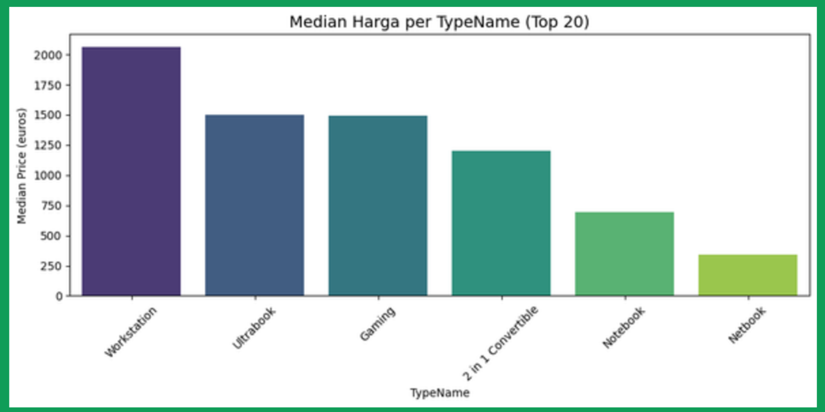
EDA

Univariate - categorical feature

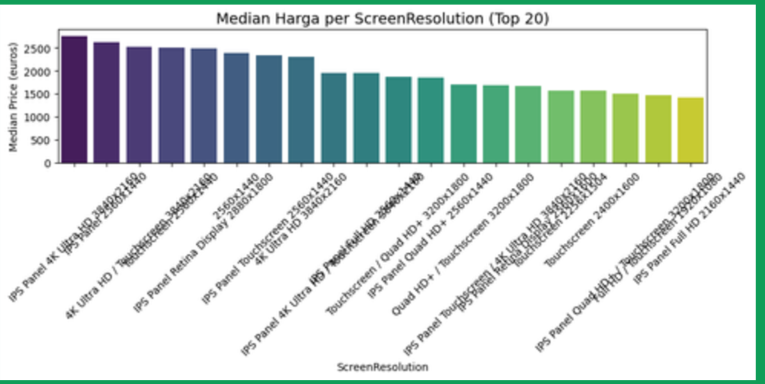
- Premium Components Drive Higher Prices: Laptop prices are heavily influenced by their type (e.g., Workstation/Gaming), premium components like Nvidia Quadro/RTX GPUs, hybrid SSD+HDD storage, and high-quality 4K/IPS displays.
- Brands Define Market Segments: Company branding clearly dictates market position; for instance, Razer targets the premium segment, while Vero and Chuwi focus on the budget market.
- Operating System as a Price Differentiator: macOS consistently positions devices in the upper price range, whereas Chrome OS is a key feature of more affordable laptops.
- Main Market Concentration: The most competitive and active segment of the laptop market is concentrated within the €700–€800 price range.



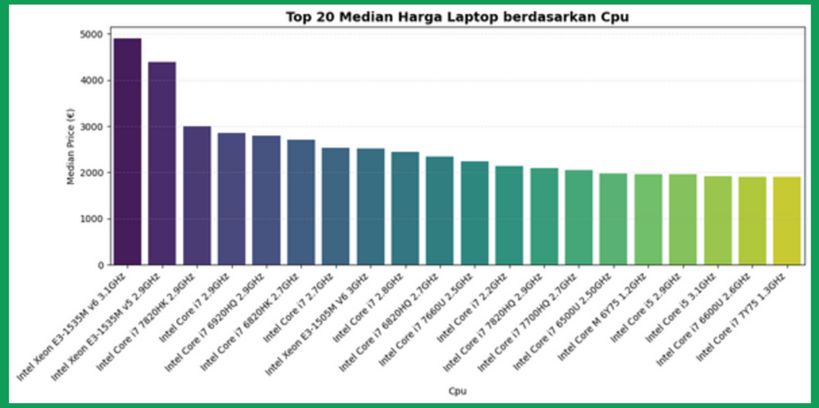
Company



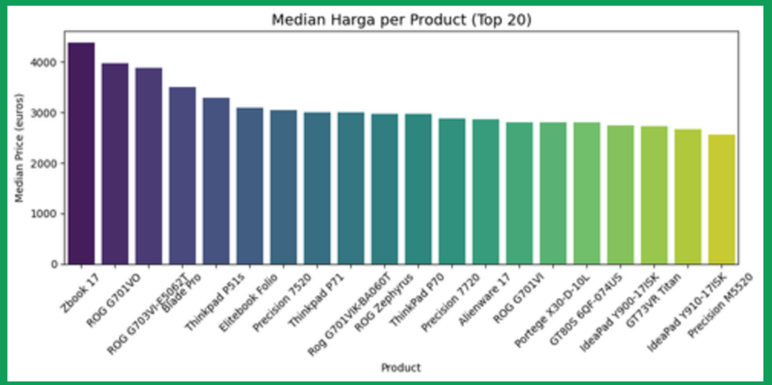
TypeName



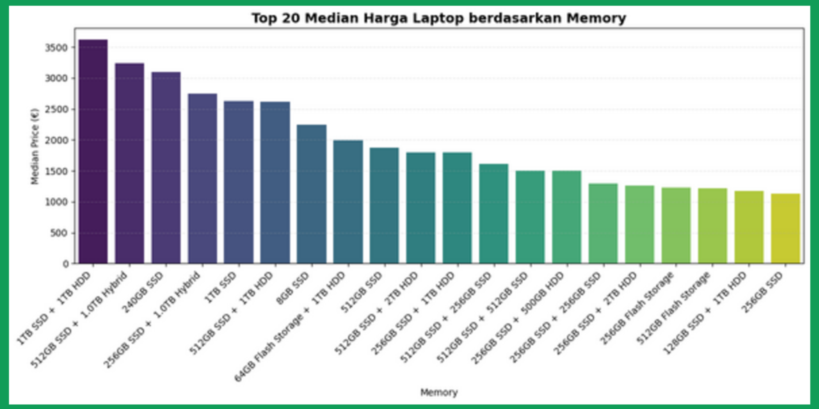
ScreenResolution



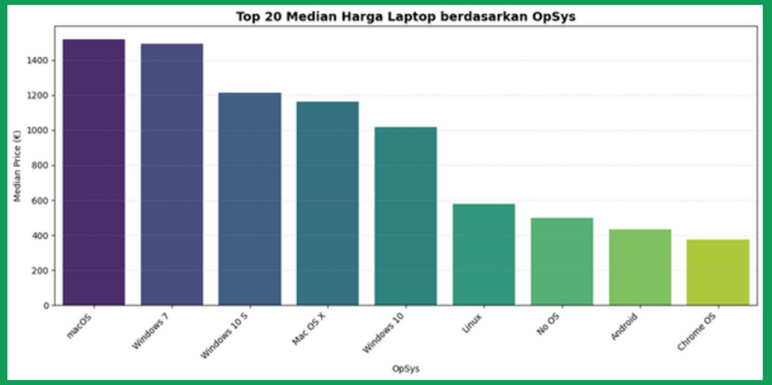
CPU



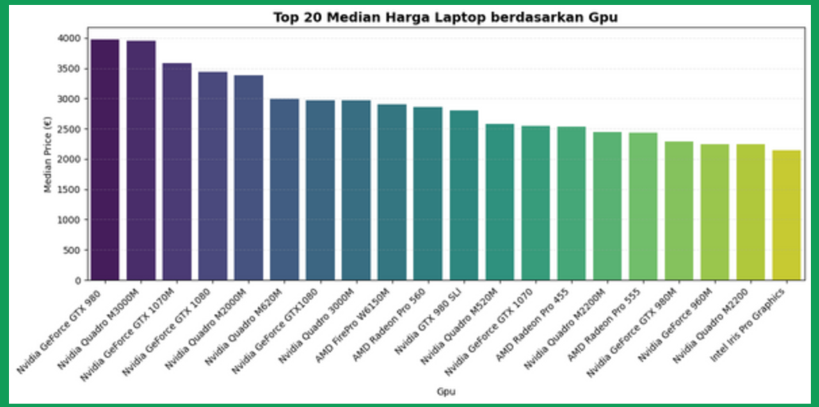
Top Product



Memory



OS

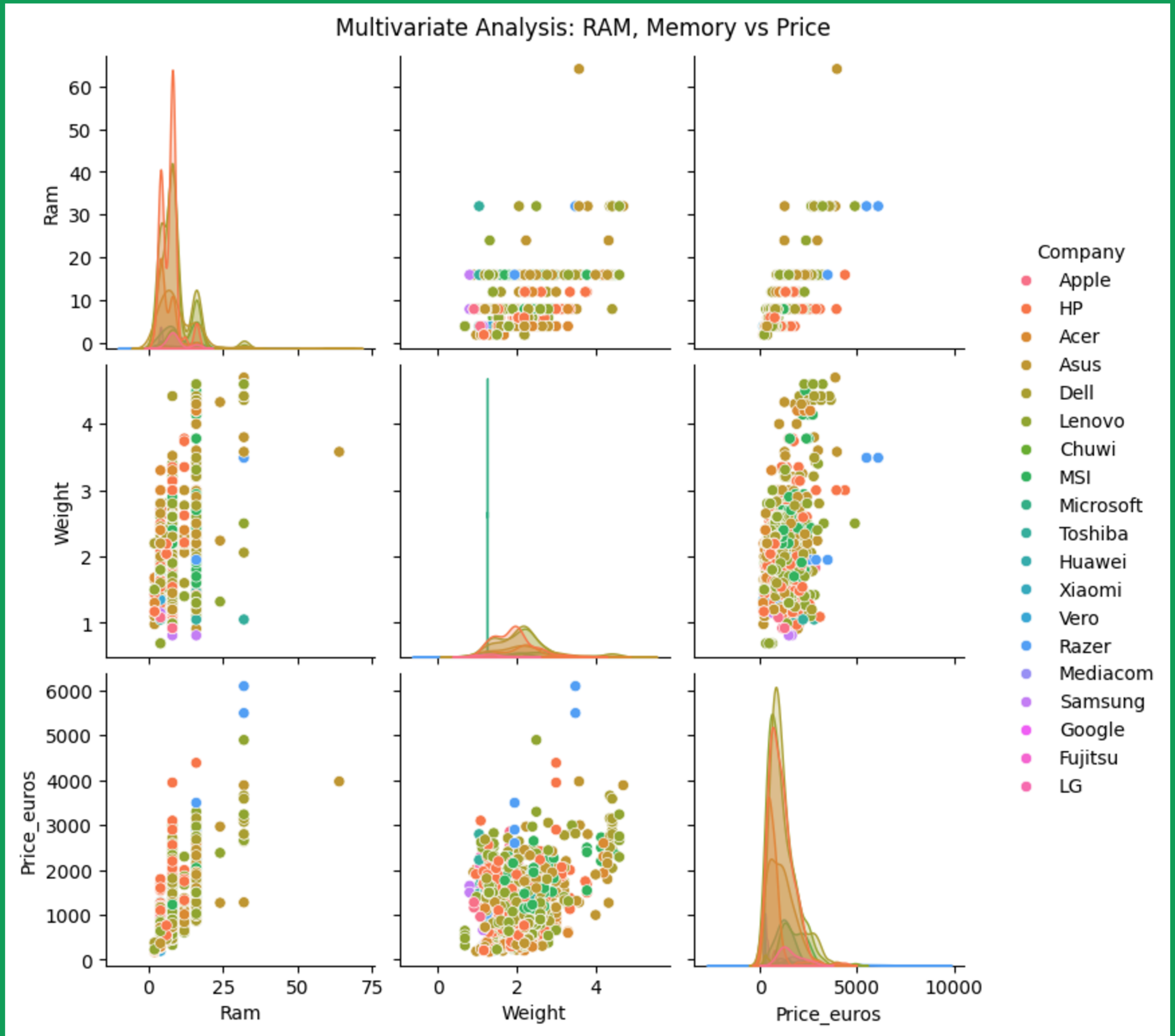


GPU

EDA

Multivariate Analysis

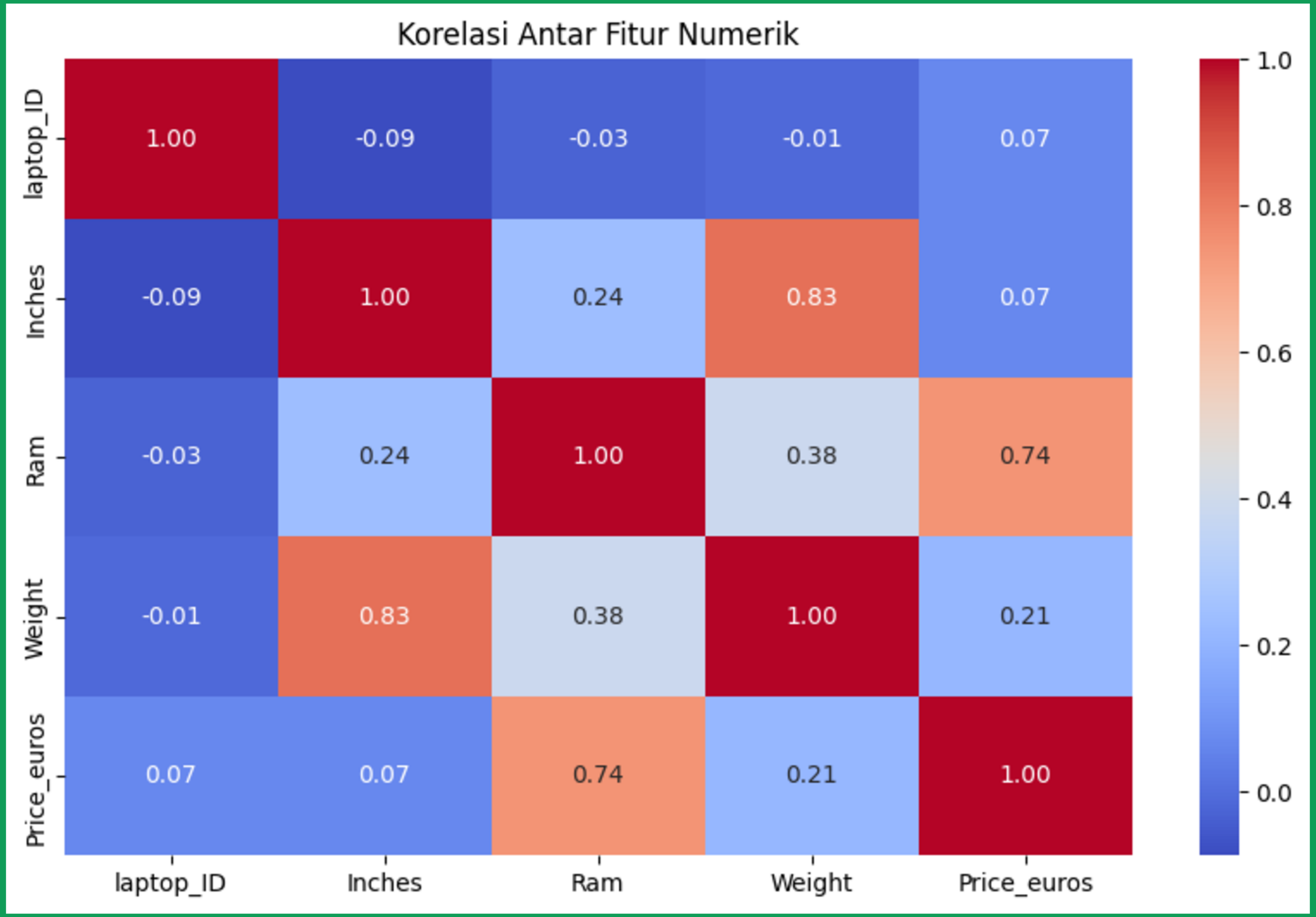
While powerful specifications like ample RAM and storage are key price drivers, brand positioning is a significant factor, enabling premium brands like Apple to command higher prices than competitors with comparable hardware.



EDA

Correlation Features

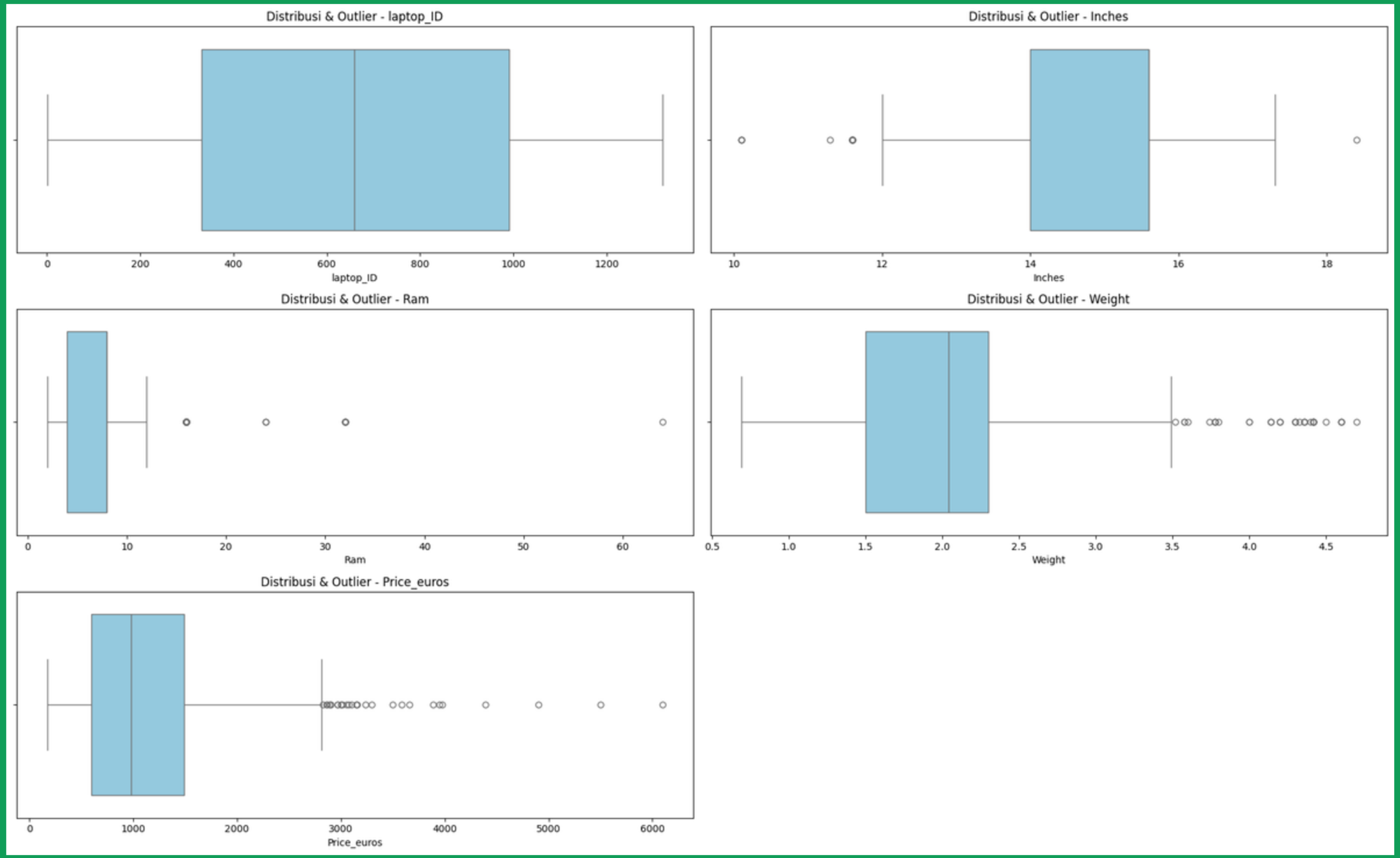
- A strong positive correlation (0.83) exists between the laptop's screen size and its weight, indicating that larger screens typically mean heavier devices.
- RAM is a significant price driver, showing a strong positive correlation of 0.74 with the laptop's price.
- A moderate positive correlation (0.38) was also observed between the laptop's weight and its RAM capacity.



EDA

Outlier Check - Numerical Features

- Inches Features (2.99%)
- Ram Features (16.96%)
- Weight Features (3.53%)
- Price_euros Features (2.23%)



Beginner ML
GDGoC UIN JAKARTA

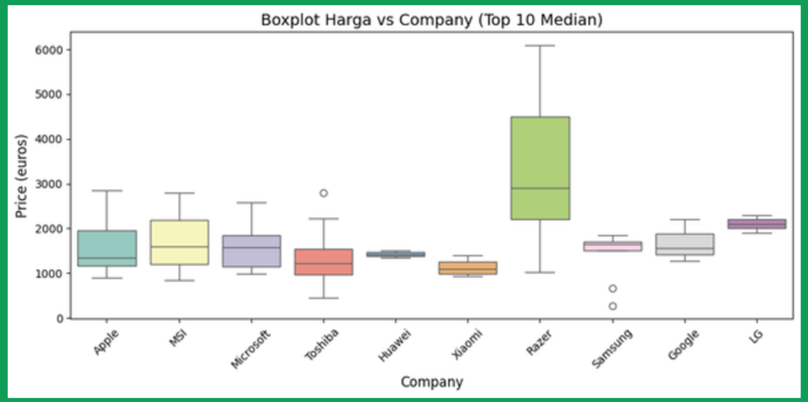
Group 4

EDA

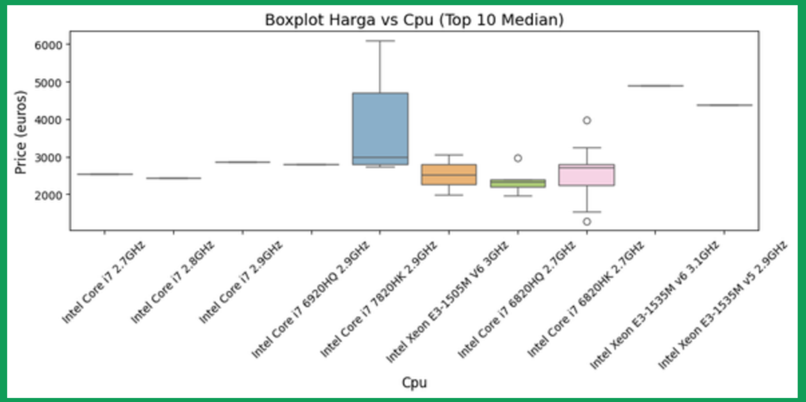
Outlier Check - Categorical Features

- Company features
- Product features
- ScreenResolution features
- CPU features
- Memory features
- GPU features
- OpSys features

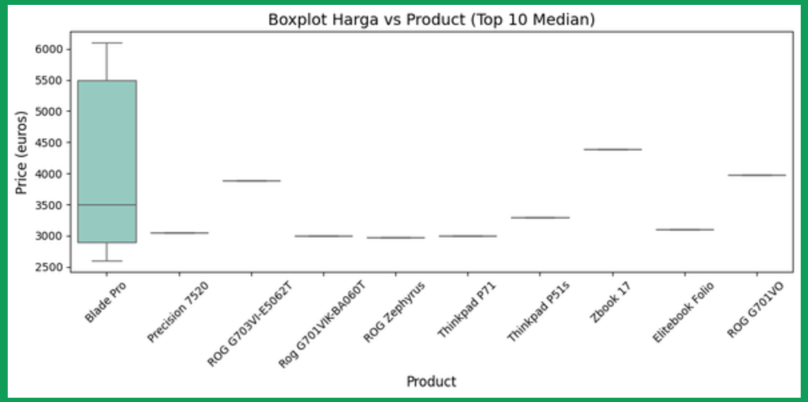
*median value



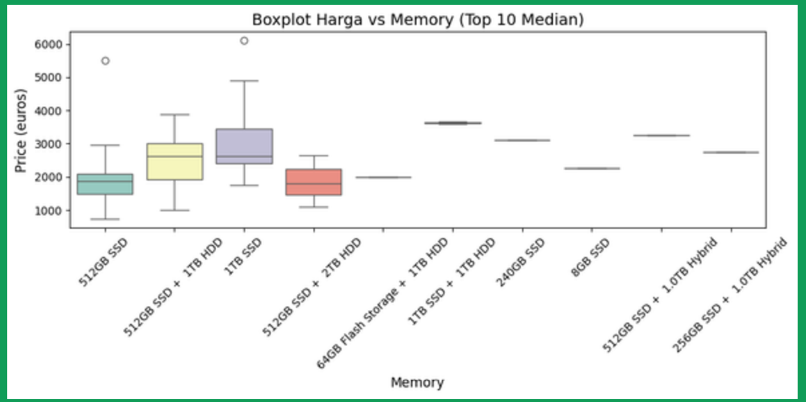
Company



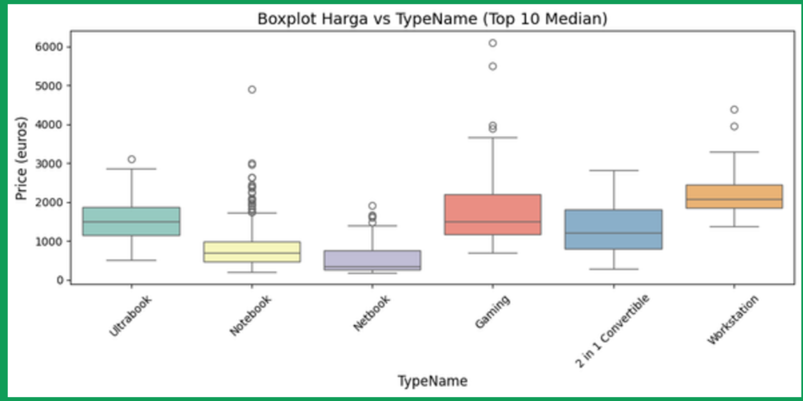
TypeName



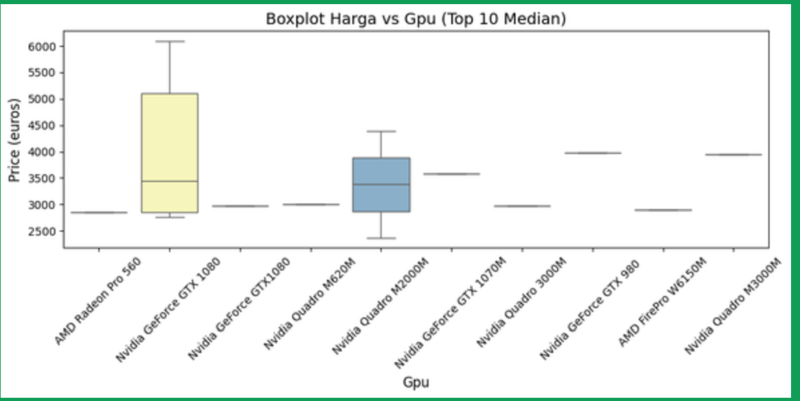
ScreenResolution



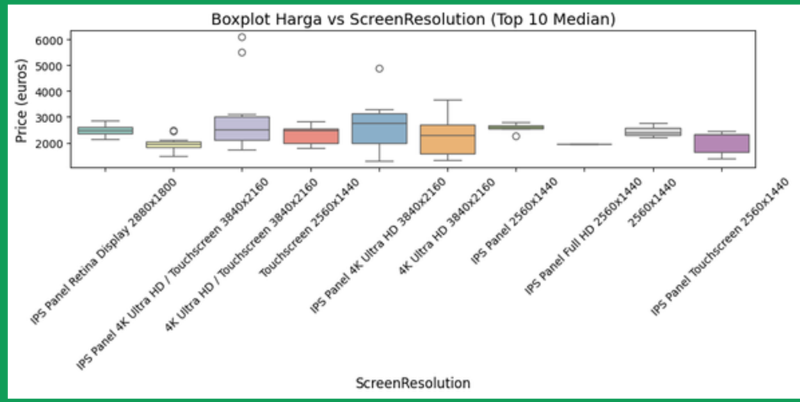
CPU



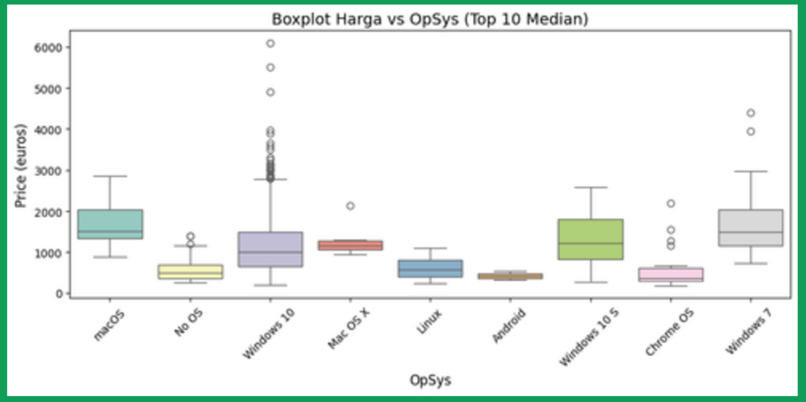
Top Product



Memory



OS



GPU

Feature Engineering

Feature Extraction

GPU_Vendor
CPU_Speed_GHz
Storage_Size
Resolution_Size
Price_Category

Scaling

Features
(StandardScaler)
Target
(MinMaxScaling)

Encoding

One hot encoding
(categorical_features)

Model Building

Challenges - 40%

- The mix of categorical and numerical features complicates the preprocessing stage.
- The presence of outliers in the numerical data could potentially disrupt model accuracy.

Strategy - 30%

- Handling outliers in the target variable using the IQR method.
- Using a Random Forest model, which is robust to outliers in the predictor features.
- Applying One-Hot Encoding for categorical features and StandardScaler for numerical features.

Solution - 30%

- Building and comparing several models: Linear Regression, Decision Tree, and Random Forest.
- Performing Cross-Validation to evaluate model stability and performance.

Model Evaluation

Linear Regression

Training set

Mean Squared Error (MSE)
0.004542147548224353

Root Mean Squared Error (RMSE)
0.06739545643605623

Mean Absolute Error (MAE)
0.05144124566315576

R-squared (R2)
0.9088029221203751

Testing set

Mean Squared Error (MSE)
0.006226653519643523

Root Mean Squared Error (RMSE)
0.07890914724950158

Mean Absolute Error (MAE)
0.061491443767882695

R-squared (R2)
0.8860064481893432

Model Evaluation

Decision Tree

Training set

Mean Squared Error (MSE)
0.0001270762912712981

Root Mean Squared Error (RMSE)
0.011272812039207346

Mean Absolute Error (MAE)
0.0026315968553114024

R-squared (R2)
0.9974485667168049

Testing set

Mean Squared Error (MSE)
0.007464913308670685

Root Mean Squared Error (RMSE)
0.08639972979512543

Mean Absolute Error (MAE)
0.06152064310902882

R-squared (R2)
0.8633371875712251

Model Evaluation

Random Forest

Training set

Mean Squared Error (MSE)
0.0007004405585763358

Root Mean Squared Error (RMSE)
0.026465837575567787

Mean Absolute Error (MAE)
0.018714008439553714

R-squared (R2)
0.9859365792299051

Testing set

Mean Squared Error (MSE)
0.005146084728274344

Root Mean Squared Error (RMSE)
0.0717362162946607

Mean Absolute Error (MAE)
0.05344908861344202

R-squared (R2)
0.9057888038504259

Model Conclusion

The Random Forest model was selected as the final model due to its superior and most stable predictive performance, achieving an R-squared (R^2) score of 0.90. This score indicates that our model can explain 90% of the price variance in laptops based on the selected features. This success was supported by the effective implementation of scaling, encoding, and feature engineering, and was validated using cross-validation to ensure the model does not overfit.

Model Usability

1.For the End-User

The model acts as a smart shopping assistant, providing objective and transparent price estimations. This empowers students to make more confident and well-informed purchasing decisions that align with their budget.

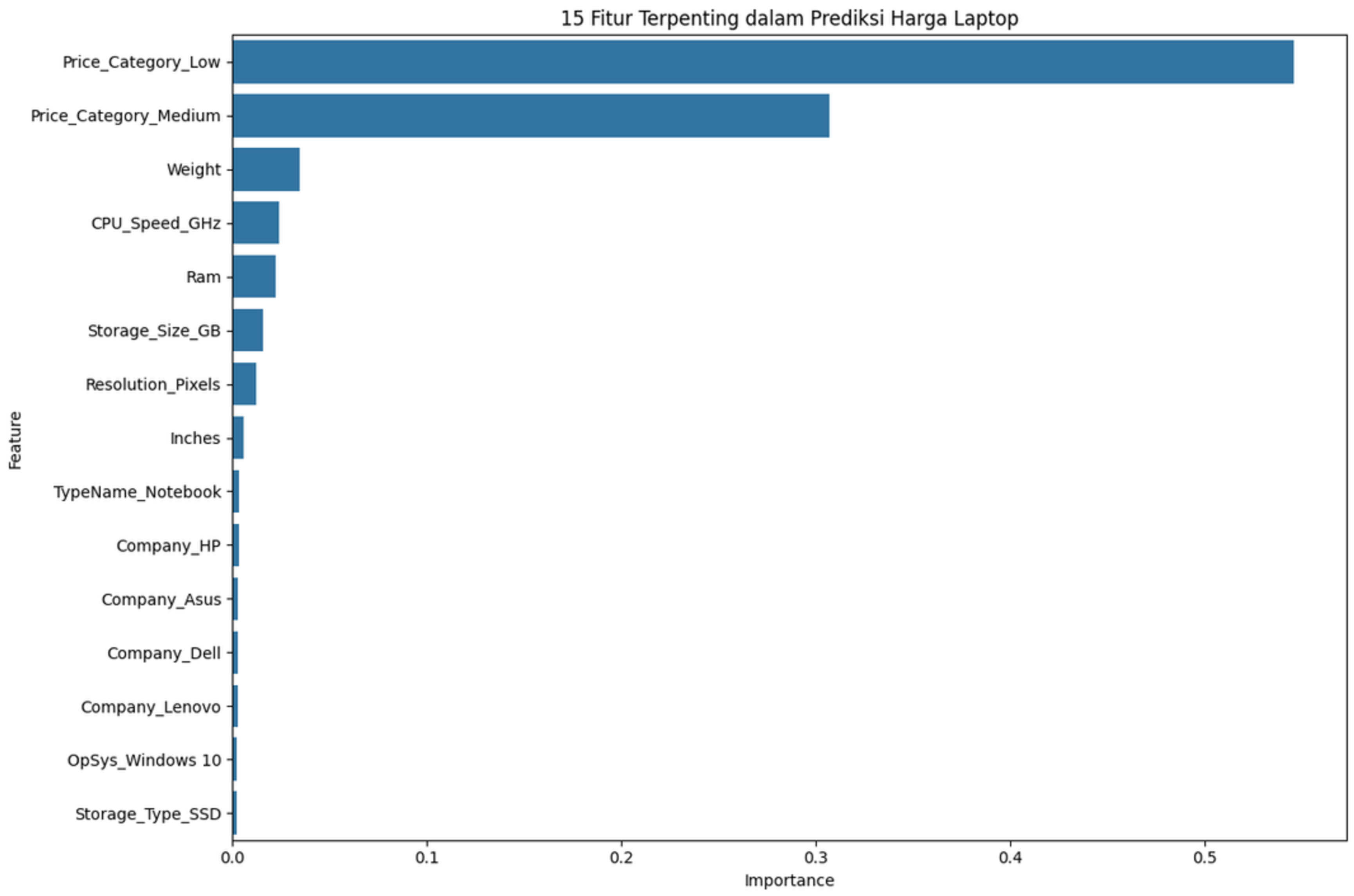
2.Bisnis Usability

This model serves as a foundation for future innovation. It can be integrated into a web application, developed into a public API, or enriched with real-time data to create an even more advanced laptop comparison tool.

3.For Developers

Serves as a business strategy optimization tool. Companies can leverage this model to set competitive pricing, develop sharper product segmentation, and manage inventory more efficiently based on trends in high-demand specifications.

Data Analysis



- Most Impactful Feature Engineering: The engineered feature, Price_Category, emerged as the strongest predictor, proving that market segmentation is a fundamental factor in price determination.
- Core Specifications Remain Dominant: RAM, CPU Speed, and Storage Size consistently form the most influential trio of technical specifications driving the price.
- Consumers Prioritize Specs Over Brand: Interestingly, the influence of the Brand and Operating System was significantly lower than that of hardware specifications, indicating a highly value-oriented market.

Recomendation

1.For the End-User (Students)

Focus your budget on the core specification trio—RAM, CPU, and Storage—to get the best performance for your money.

2.For Producers & Retailers

Use these core specifications as the foundation for your pricing and product segmentation strategies to attract value-oriented consumers.

3.For Developers

Continue to invest in feature engineering, as it has proven to be highly effective in boosting the model's accuracy.

ICN

Thank you

Our model successfully predicts laptop prices with 90% accuracy, empowering students to make smarter purchasing decisions.

Let's Connect:

ahmadihsan506@gmail.com

LinkedIn

GitHub