Data Analysis Log

# Analysis Steps and Reasons

1. Suggesting target column based on existing columns: To identify the appropriate target for modeling.

2. Transforming data by imputing and normalizing: To handle missing values and ensure data consistency.

3. Evaluating models based on the target column type: To assess performance using suitable algorithms.

4. Generating and saving plots for data analysis: To visualize data distributions and relationships.

# Data Processing

The dataset was first checked for symbols and spaces in column names, and appropriate replacements were made to ensure compatibility.

# Data Transformation

The data was transformed using various steps including missing value imputation (using the mean for numerical columns), normalization of numerical values, and label encoding for categorical variables.

# Model Evaluation

## Model: Linear Regression

Metrics:  
{'MAE': np.float64(0.03258865123017764), 'MSE': np.float64(0.0024763437918032326), 'R²': 0.9049581999818268}

## Model: Random Forest Regressor

Metrics:  
{'MAE': np.float64(0.009248703790643093), 'MSE': np.float64(0.00025371049758191424), 'R²': 0.990262619247979}

## Model: Gradient Boosting Regressor

Metrics:  
{'MAE': np.float64(0.013693524049428647), 'MSE': np.float64(0.00034394767617120387), 'R²': 0.9867993263441119}

## Model: K-Neighbors Regressor

Metrics:  
{'MAE': np.float64(0.06505981510818594), 'MSE': np.float64(0.007524187104382528), 'R²': 0.7112225336235287}

## Model: Support Vector Regressor

Metrics:  
{'MAE': np.float64(0.10166016465998176), 'MSE': np.float64(0.016373194538686128), 'R²': 0.3715986099525793}

## Model: Decision Tree Regressor

Metrics:  
{'MAE': np.float64(0.010625175803544478), 'MSE': np.float64(0.0004395336155653471), 'R²': 0.9831307485939151}

# Best Model

The best-performing model was Support Vector Regressor based on the evaluation metrics.

## Model Performance

The Support Vector Regressor outperformed other models in terms of evaluation metrics, which include Mean Absolute Error, Mean Squared Error, R² (for regression), or confusion matrix and classification report (for classification).

Click here to view the Correlation Matrix

## Scatter Plots for Price\_euros vs Features

# Data Transformation Details

Imputation method used: Mean imputation for numerical columns.

Normalization: Applied to all numerical columns to scale the values.

Label Encoding: Applied to categorical columns for conversion into numerical form.