Laptop Price Analysis – Project Explanation

1️⃣ Project Objective

The goal of this project was to analyze the laptop market, identify trends in pricing, and understand how various features such as brand, RAM, and storage affect laptop prices. Additionally, a baseline machine learning model was built to predict laptop prices based on these features. The insights were presented via an interactive Power BI dashboard.

2️⃣ Tools Used

Tool	Purpose
Excel	Data cleaning, handling missing values, removing duplicates, and basic exploratory analysis.
Python (Pandas, Matplotlib, Seaborn, Scikit-Learn)	Exploratory data analysis (EDA), feature preparation, encoding categorical variables, building a Linear Regression model, and evaluating predictions.
Power BI	Visualizing insights interactively with charts and slicers, allowing users to explore pricing trends by brand, storage, RAM, and OS.

3️⃣ Steps Performed

Step 1: Data Cleaning (Excel)

- Removed duplicates and unnecessary columns.
- Corrected data types.
- Handled missing values.

- Extracted numeric features (e.g., `Primary_Storage_GB` for SSD capacity).
- Saved the cleaned dataset as `laptop_data_cleaned.csv`.

Step 2: Exploratory Data Analysis (Python)

- Loaded cleaned dataset.
- Checked for missing values and basic statistics.
- Visualized price distribution.
- Analyzed relationships:
    - Brand vs Price
    - RAM vs Price
    - Primary Storage vs Price
- Identified outliers (optional step); retained premium laptops as they were genuine data points.

Step 3: Machine Learning (Python)

- Defined target variable: `Price_EUR`.
- Selected features (numerical and categorical).
- Encoded categorical columns using LabelEncoder.
- Split dataset into **train (80%)** and **test (20%)**.
- Built **Linear Regression** model as baseline.
- Evaluated model performance using:

    - **MAE** – Mean Absolute Error
    - **RMSE** – Root Mean Squared Error
    - **R² Score** – Model explanation power
- Visualized actual vs predicted prices.

Step 4: Power BI Dashboard

- Created **3 interactive charts**:
    - **Average Laptop Price by Brand** – Clustered column chart.
    - **Price Distribution of Laptops** – Clustered column chart.
    - **Average Laptop Price by Primary Storage (GB)** – Clustered column chart.
- Added slicers for **Company**, **OS**, and **RAM**.
- Polished layout:
    - Clean dark theme
    - Consistent data labels
    - X-axis for storage set to **Categorical** for better readability
    - Removed unnecessary visuals

4️⃣ Key Insights from the Project

- Brands like **Razer** and **LG** have the highest average laptop prices.
- Majority of laptops are in the **€500–€1500** price range.
- Laptops with **higher RAM and SSD storage** tend to be more expensive.

- The dashboard allows **interactive exploration** to analyze any brand, OS, or RAM range.

5️⃣ Answers to Typical Project Questions

Q1: Why use Excel for cleaning?

Excel allows quick detection of missing values, duplicates, and incorrect data types before importing into Python.

Q2: Why Linear Regression?

Linear Regression provides a baseline model for predicting laptop prices. It is simple, interpretable, and suitable for small datasets.

Q3: Were outliers removed?

Outliers (high-priced premium laptops) were **retained** because they are genuine market data points.

Q4: How does Power BI help?

Power BI provides an **interactive, visual way** to explore insights, making it easy for non-technical users to analyze trends.

Q5: How can this project be improved?

- Advanced ML models (Random Forest, XGBoost)
- Hyperparameter tuning
- More feature engineering (GPU, screen resolution)

6️⃣ Project Deliverables

1. **Cleaned dataset**: `laptop_data_cleaned.csv`
2. **Python notebook**: EDA + ML + evaluation

3. **Power BI dashboard**: interactive insights
4. **Documentation / README**: project explanation