**Laptop Price Prediction Project**

Name:

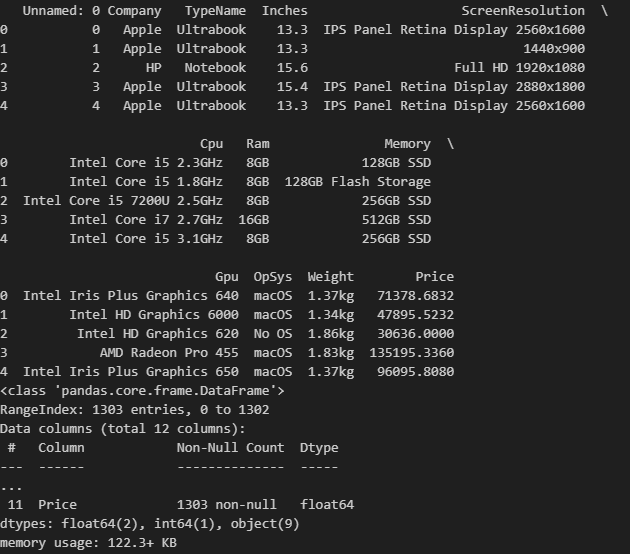
M.Araiz Abdullah

# 1. Introduction

The **Laptop Price Prediction Project** is designed to demonstrate how **Machine Learning** can be applied to real-world problems in the electronics market. Laptops come with various specifications such as processor type, RAM size, storage capacity, graphics card, and display features. These specifications directly influence the price of a laptop.

# 2. Dataset

The dataset used in this project contains detailed information about different **laptops and their specifications**, along with their respective prices. It was obtained from **Kaggle**, a popular platform for datasets and machine learning projects.

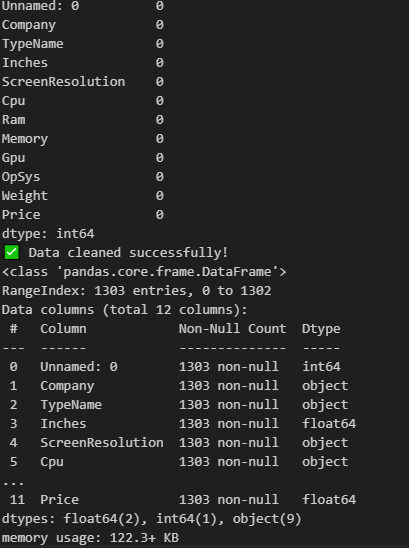


# 3. Data Preprocessing

Before training the machine learning model, the dataset underwent **preprocessing** to ensure the data was clean, consistent, and in a suitable format for analysis.

### Steps Performed:

1. **Handling Missing Values**
   * Checked for null values in the dataset.
   * Removed rows with missing data to avoid errors during training.
2. **Removing Duplicates**
   * Duplicate rows were identified and dropped to maintain dataset integrity.
3. **Data Type Conversion**
   * Converted RAM, Storage, and Weight columns into numerical values for better analysis.
   * Example: “8GB” → 8, “1TB” → 1024 (in GB).
4. **Encoding Categorical Variables**
   * Features such as Company, CPU, GPU, Operating System were categorical.
   * Used **One-Hot Encoding (pd.get\_dummies)** to convert them into numerical values.
5. **Feature Scaling (if required)**
   * Normalized numerical columns (e.g., Weight, Screen Size) to bring all values into a similar range.
6. **Splitting Features and Target**
   * Independent variables (**X**) = Laptop specifications (RAM, CPU, GPU, etc.).
   * Dependent variable (**y**) = Laptop Price.
7. **Train-Test Split**
   * Divided the dataset into **80% training data** and **20% testing data** using train\_test\_split.
   * Ensures the model is trained on one set of data and evaluated on unseen data.



**Data Analysis**

After preprocessing, **data analysis** was performed to understand the relationships between laptop features and their prices. This step helped in identifying the most important attributes that influence laptop pricing.

### 📊 **Exploratory Data Analysis (EDA)**

1. **Descriptive Statistics**
   * Calculated measures such as mean, median, and standard deviation for numerical features like RAM, Weight, and Screen Size.
   * Example: Laptops with higher RAM (16GB, 32GB) had significantly higher average prices compared to those with 4GB or 8GB.
2. **Price Distribution**
   * Created histograms of laptop prices.
   * Observation: Most laptops fall into the mid-price range, with fewer laptops in very high or very low price ranges.
3. **Correlation Analysis**
   * Generated a correlation heatmap to see how numerical features relate to price.
   * Key Insights:
     + **RAM** showed a strong positive correlation with Price.
     + **CPU performance** and **GPU type** also had significant impact.
     + **Weight** had a weaker correlation, showing price is not heavily influenced by laptop weight.
4. **Brand-wise Price Comparison**
   * Compared average prices across different companies (e.g., Apple, Dell, HP, Asus).
   * Observation: Apple laptops were priced significantly higher on average than other brands.
5. **Feature vs Price Visualization**
   * Scatter plots showed that higher RAM and SSD storage types generally lead to higher laptop prices.
   * Bar charts revealed that gaming laptops and ultrabooks are priced higher than notebooks.

****

# 5. Model Training

After preparing and analyzing the dataset, the next step was to **train a machine learning model** that can predict laptop prices based on their specifications.

### ⚙️ **Steps in Model Training**

1. **Feature Selection**
   * Independent Variables (**X**) included:  
     Company, TypeName, Inches, ScreenResolution, CPU, RAM, Memory, GPU, Operating System, Weight.
   * Dependent Variable (**y**) was:  
     Price of the laptop.
2. **Train-Test Split**
   * The dataset was divided into **80% training data** and **20% testing data** using train\_test\_split.
   * This ensures that the model is evaluated on unseen data for fair performance testing.
3. **Model Choice**
   * A **Linear Regression** model was selected as the baseline algorithm because it is simple and effective for predicting continuous values like price.
   * Future improvements may include advanced models such as Random Forest, Gradient Boosting, or XGBoost.

****

# 7. Application

After training and saving the model, a **Streamlit-based web application** was developed to make the Laptop Price Prediction system interactive and user-friendly. This application allows users to input laptop specifications and instantly receive an estimated price.

### ⚙️ **Application Features**

1. **User Input Form**
   * Users can enter details such as:
     + Brand (Company)
     + Laptop Type (Ultrabook, Gaming, Notebook, etc.)
     + RAM Size
     + Processor Type (CPU)
     + GPU (Graphics Card)
     + Storage (HDD/SSD)
     + Operating System
     + Screen Size and Resolution
     + Weight
2. **Model Integration**
   * The application loads the pre-trained model (laptop\_price\_model.pkl) using **Joblib**.
   * The user’s inputs are transformed into the same format as the training dataset.
   * The trained model then predicts the price in real-time.
3. **Output**
   * Displays the **predicted price of the laptop** with a success message.
   * Example: “💰 Estimated Laptop Price: $1200”
4. **User Interface**
   * Built with **Streamlit**, providing a clean and responsive interface.
   * Requires only a browser to run; no complex setup is needed.

