Machine Learning Project

1. Project Description

In this project, you will explore and analyze a dataset containing information on various laptop models, including features such as brand, screen size, operating system, storage, and price. Your goal is to apply machine learning and data analysis techniques to gain insights into the factors that influence laptop prices, build predictive models, and perform data visualizations. This project will help you understand how to process real-world data and extract meaningful results through code.

2. Dataset Description:

The dataset contains information on different laptop models and their specifications, including the following columns:

- Company: The brand or manufacturer of the laptop.
- Product: The specific product model name.
- TypeName: The type of laptop (e.g., Notebook, Ultrabook, Gaming).
- Inches: The screen size of the laptop.
- Ram: Amount of RAM (in GB).
- OS: Operating system installed on the laptop.
- Weight: Weight of the laptop in kilograms.
- Price_euros: The price of the laptop in euros.
- Screen: Screen type (e.g., Standard, Full HD).
- ScreenW/ScreenH: The width and height of the screen resolution.
- RetinaDisplay: Whether the laptop has a Retina Display (Yes/No).
- CPU_company: The company that manufactures the CPU (e.g., Intel, AMD).
- CPU_freq: CPU frequency in GHz.
- CPU_model: The model of the CPU.
- PrimaryStorage: The primary storage capacity in GB.
- SecondaryStorage: The secondary storage capacity in GB.
- PrimaryStorageType: The type of primary storage (e.g., SSD, Flash Storage).
- SecondaryStorageType: The type of secondary storage (if available).
- GPU_company: The company manufacturing the GPU (e.g., Intel, Nvidia, AMD).
- GPU_model: The model of the GPU.

3. Project Steps

- Download laptop_prices.csv file
- Load the file in a notebook (Jupyter notebook or Google Colab)
- Share the code answering the questions in the next section

4. Your Goal

You submitted code should answer the following questions:

a. Identify the top 5 laptop brands by the number of products listed.
b. Find the average price of laptops for each brand. Identify which brand has the highest and lowest average price.
c. Present the correlation between Price_euros and other numeric features such as CPU_freq, Ram, Inches, and Weight. Identify which features are most correlated with price.
d. Create a new feature, StorageTotal, by summing PrimaryStorage and SecondaryStorage.
e. Build a regression model using features like Ram, Inches, CPU_freq, PrimaryStorage, and GPU_company to predict the Price_euros of a laptop. Suggest the best regression model
f. Create a classification model to predict the laptop type (TypeName) based on features such as Inches, Ram, PrimaryStorage, and Weight. Identify which features are the most important in your classification model. Suggest the best classification model

You don't need to do anything extra - simply extract the answers using the code and leave the code as is. I will run the code myself to verify if it produces the correct answers to the questions.

Happy Coding!