Text

Description automatically generated with medium confidence

**MACHINE LEARNING PROJECT REPORT**

on

**Laptop Price Prediction**

Submitted by

**Jaismeen**

**Registration Number: 12015083**

**Program Name: B. Tech Data Science (ML and AI)**

Under the Guidance of

**Mr. Ved Prakash Chaubey**

**School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

I**,** Jaismeen, certify that this project is my own work, based on my personal study and/or research and that I have acknowledged all material and sources used in its preparation, whether they be books, articles, reports, lecture notes, and any other kind of document, electronic or personal communication. I also certify that this project has not previously been submitted for assessment in any academic capacity, and that I have not copied in part or whole or otherwise plagiarised the work of other persons. I confirm that I have identified and declared all possible conflicts that I may have.

Signature: Jaismeen

**Acknowledgement**

I would like to express my gratitude to **Mr. Ved Prakash Chaubey** my project supervisor, for their guidance and support throughout the project. I would also like to thank **Lovely Professional University** for providing me with the necessary resources and infrastructure to complete this project.

**INTRODUCTION:**

A laptop price prediction dataset typically contains a variety of features and attributes of laptops that can be used to train machine learning models for predicting laptop prices. The laptop price prediction dataset is a real-world dataset that contains information about various laptops and their specifications.

The dataset includes features such as brand, model, display size, processor type, RAM, storage type, graphics card, weight, price, etc. This dataset can be used for predicting the price of a laptop based on its features. The dataset can be useful for individuals, retailers, and manufacturers who are interested in understanding the factors that influence laptop prices. It can also be used in data analysis and machine learning.

**WHY THIS DATASET?**

The purpose of choosing the laptop price prediction dataset for performing machine learning is to develop a model that can accurately predict the prices of laptops based on various features such as brand, processor, memory, screen size, etc. This can be useful for both consumers and sellers in the laptop market. Consumers can use the model to determine if they are getting a fair price for a laptop, while sellers can use it to set competitive prices for their products. Additionally, this dataset provides a good opportunity to explore various machine learning techniques such as linear regression, lasso regression, ridge regression, SVM, decision trees and random forest.

**DOMAIN:**

**Computer Hardware and Electronics:**

Computer hardware and electronics domain deals with the design, development, manufacturing, and maintenance of various computer components and electronic devices.

This dataset belongs to the domain of computer hardware and electronics. The dataset contains information about various features of laptops such as brand, screen size, RAM, processor, graphics card, operating system, weight and price. These features are important factors that determine the performance and price of a laptop.

**DESCRIPTION OF DATASET:**

This Dataset consists of 1303 rows and 12 columns:

1. Company: The name of the laptop brand, such as Acer, Apple, Asus, Dell, HP, Lenovo or MSI.

2. Product: The specific model of the laptop, such as Aspire 3, MacBook Air, ROG Strix G15 or ThinkPad E14.

3. TypeName: The general category of the laptop, such as Ultrabook, Notebook, Gaming, or Workstation.

4. Inches: The size of the laptop screen, measured diagonally in inches.

5. Screen Resolution: The screen resolution of the laptop, such as 1920x1080, 1366x768, or 3840x2160.

6. CPU: The processor type and speed, such as Intel Core i5-7200U or AMD Ryzen 5 2500U.

7. RAM: The amount of random-access memory (RAM) in the laptop, measured in gigabytes (GB).

8. Memory: The storage capacity of the laptop, measured in gigabytes (GB).

9. GPU: The graphics card type and memory, such as NVIDIA GeForce GTX 1650 with 4GB GDDR5.

10. OpSys: The operating system installed on the laptop, such as Windows 10, macOS or Linux.

11. Weight: The weight of the laptop, measured in kilograms (kg).

12. Price: The price of the laptop in Euros (€) at the time of data collection.

This dataset can be used to train machine learning models for predicting laptop prices based on the given features. The target variable in this case is the laptop price and the other features can be used as predictors.

**QUESTIONS THAT CAN BE SOLVED USING THIS DATA:**

* Which features have the most significant impact on the price of a laptop?
* Are customers willing to pay a premium for laptops with better graphics cards or processors?
* Is there a particular brand that tends to be priced higher or lower than its competitors?
* Are laptops with larger screen sizes generally more expensive than those with smaller screens?
* Can we predict the price range for a laptop based on its features and if so, how accurate are our predictions?
* Are there any combinations of features that result in a laptop being priced significantly higher or lower than expected?
* Can we identify any trends in the laptop market, such as the increasing popularity of certain features or brands?
* How well do different machine learning algorithms perform in predicting laptop prices and which one is the most accurate?
* Can we use the insights gained from this project to inform pricing strategies for laptop manufacturers and retailers?
* What additional data could we collect to improve the accuracy of our price predictions and how feasible would it be to collect this data?

**OBJECTIVE:**

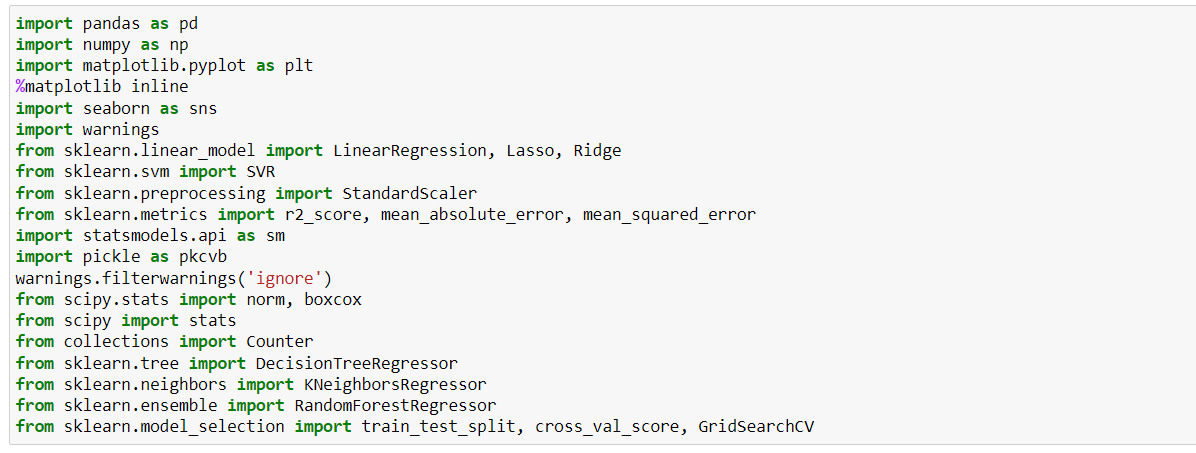
Given the various specifications and features of laptops, the task is to predict the price of a laptop accurately. This can help customers in making informed decisions while purchasing a laptop and can also assist companies in setting prices for their products based on their features and specifications. Additionally, this can aid in identifying the important features that contribute the most towards the price of a laptop.

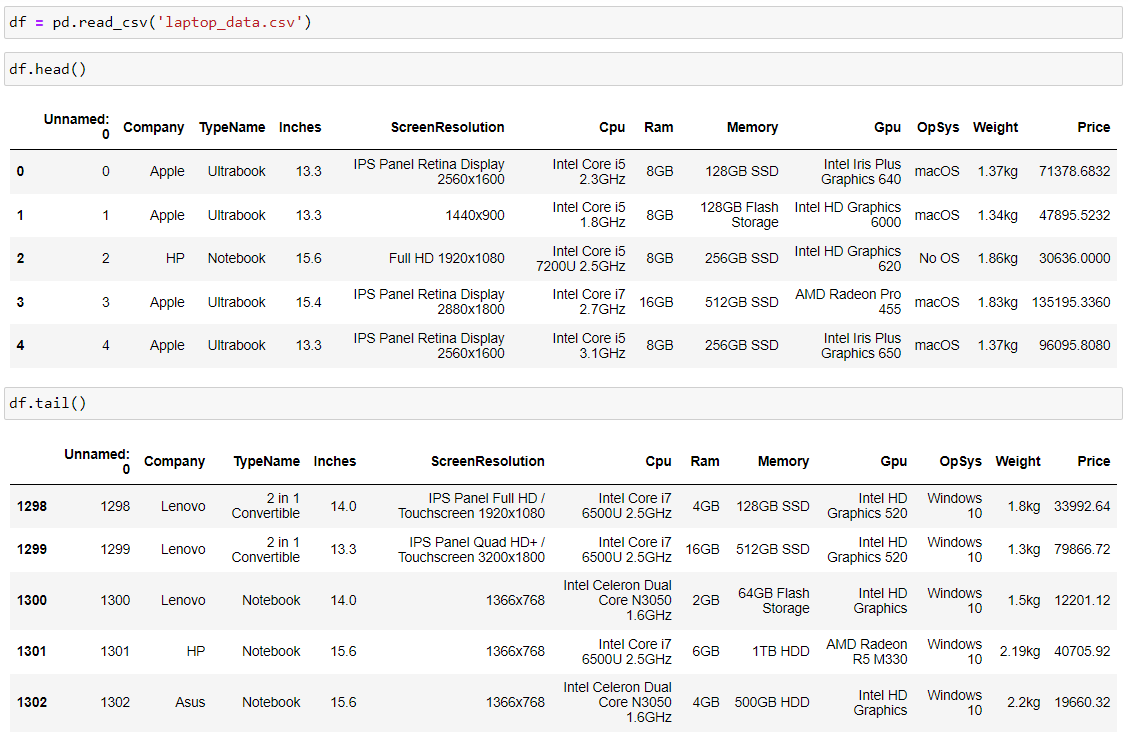
The objective of this project is to develop a machine learning model that can accurately predict the prices of laptops based on their hardware specifications and other features.

This will enable businesses in the computer hardware and electronics domain to offer competitive prices to their customers and optimize their inventory management. By accurately predicting the prices of laptops, businesses can ensure that they are not overpricing or under-pricing their products, which can lead to loss of sales or profit.

By understanding which features have the greatest impact on laptop prices, businesses can make informed decisions about which products to stock and how to price them. The goal is to provide better customer satisfaction and increase revenue for the business.

**Importing the Libraries:**

****

**Reading and Understanding Data:**

**Checking the Datatypes of Columns:**

Graphical user interface, text, application

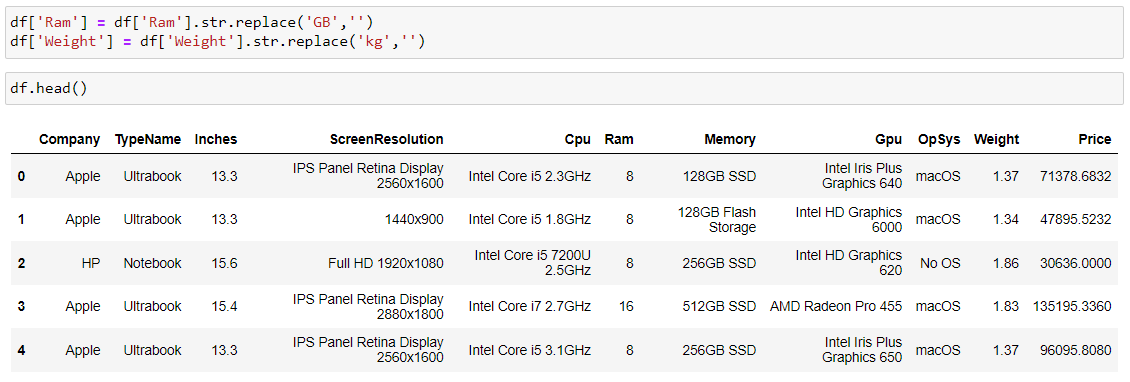
Description automatically generated

**Checking the Null – Values:**

**Graphical user interface, application

Description automatically generated**

**Removing Categorical features from Numerical Columns:**

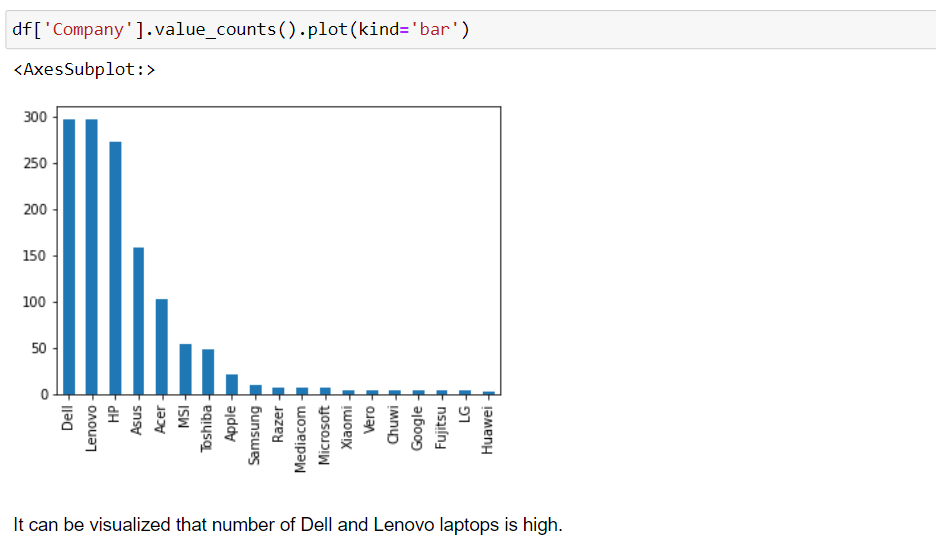
****

**Plotting the Prices of Laptops on a Distplot:**

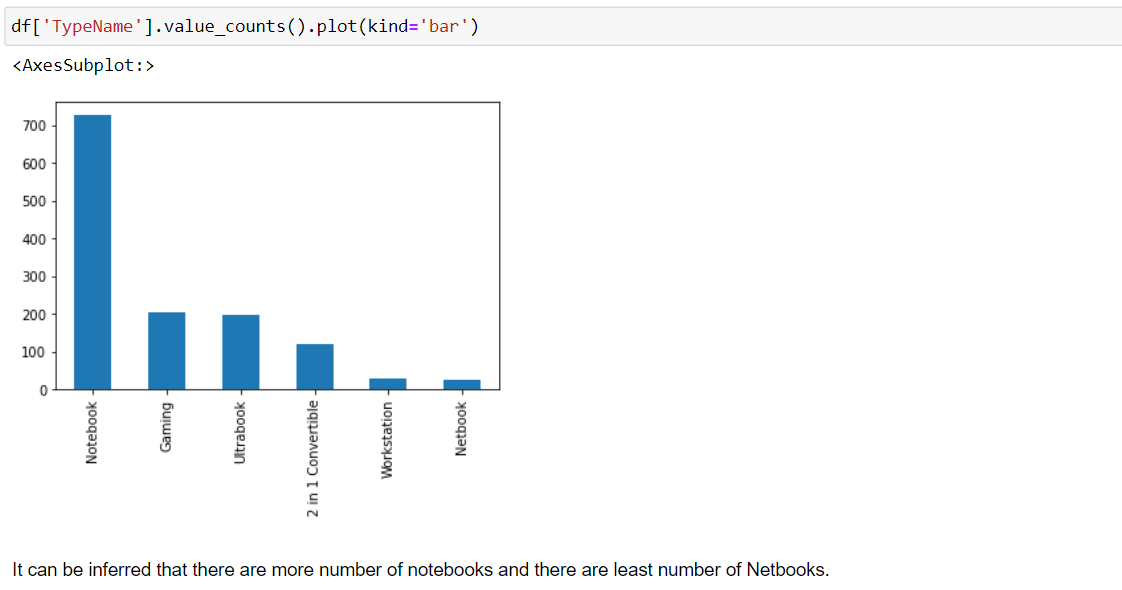
**Graphical user interface, application

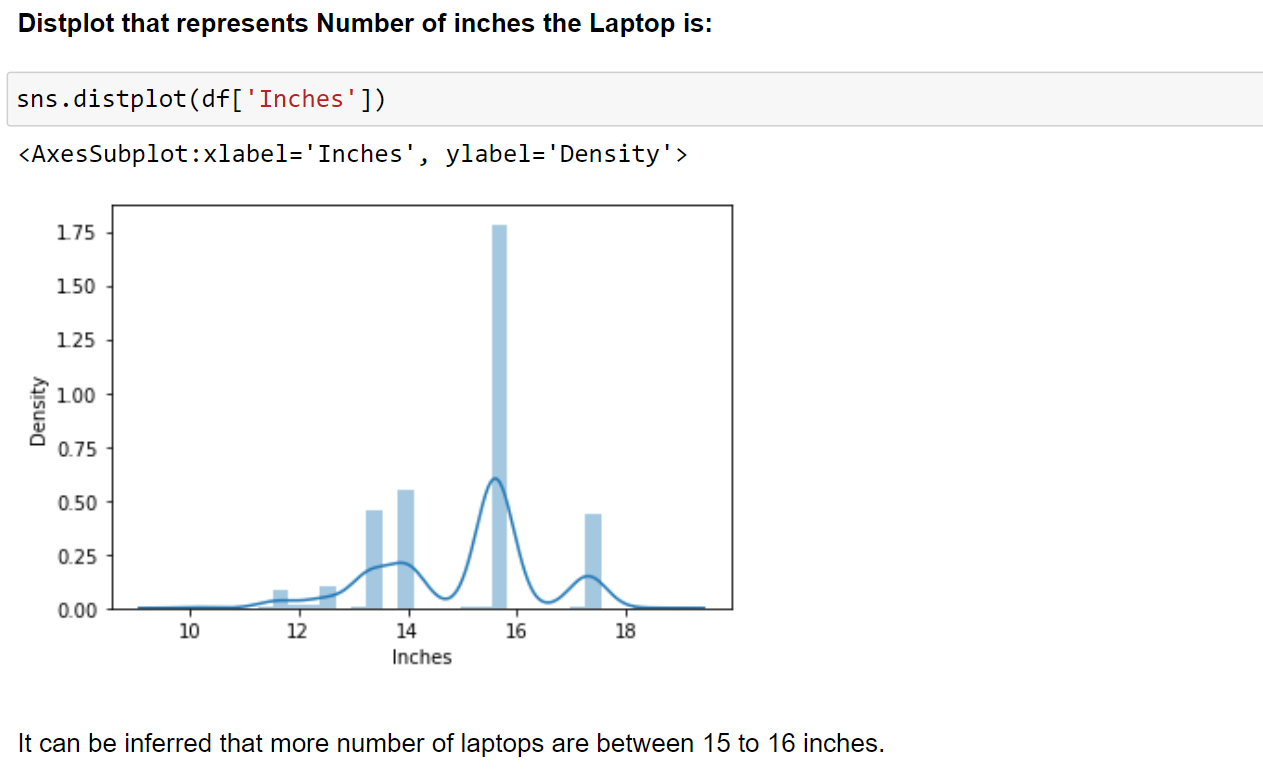
Description automatically generated**

**Visualizing the Laptop Companies on a Bar plot:**

****

**Plotting the Type of Laptop on a Bar Plot:**

****

****

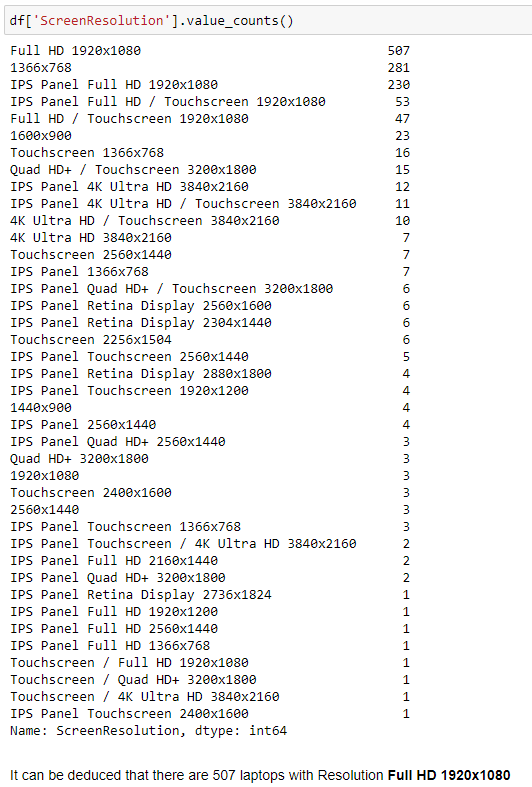
**Chart, scatter chart

Description automatically generated**

**Plotting the Type of Laptop and its price on a Bar Plot:**

**Chart, bar chart

Description automatically generated**

****

**Text

Description automatically generated with low confidence**

**Chart, bar chart

Description automatically generated**

**Text

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**Chart, bar chart, waterfall chart

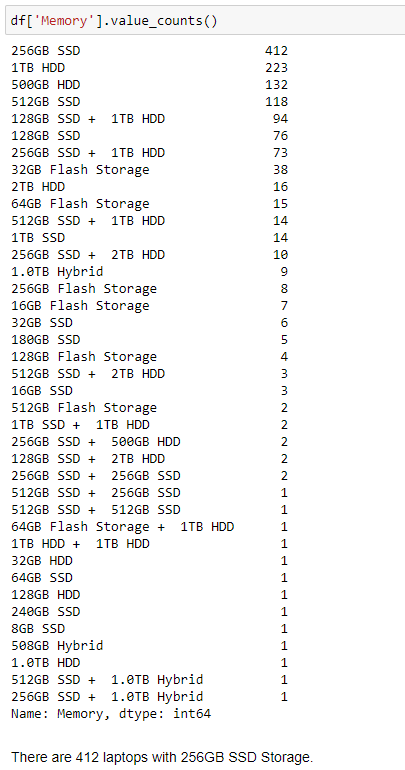
Description automatically generated**

**Graphical user interface, application

Description automatically generated**

**Chart, bar chart

Description automatically generated**

****

**Graphical user interface, text, application, email

Description automatically generated**

**Dropping the Redundant Columns:**

**Graphical user interface, text, application

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**A picture containing chart

Description automatically generated**

**Heatmap Representation:**

**Chart

Description automatically generated**

**Pairplot Representation:**

**Diagram

Description automatically generated**

This pairplot shows that there is a strong positive correlation between the laptop’s price and its specifications such as the processor speed, RAM and hard disk capacity. This suggests that laptops with higher specifications tend to have a higher price. Additionally, we observe a positive correlation between the price and the weight of the laptop, indicating that lighter laptops are typically more expensive.

**Splitting Dataset into Train and Test Data:**

**A picture containing graphical user interface

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated Linear Regression:**

**Ridge Regression:**

**Graphical user interface, text, application, email

Description automatically generated**

**Lasso Regression:**

**Graphical user interface, text, application, email

Description automatically generated**

**Decision Tree:**

**Graphical user interface, text, application, email

Description automatically generated**

**SVM (Support Vector Machines):**

**Graphical user interface, text, application, email

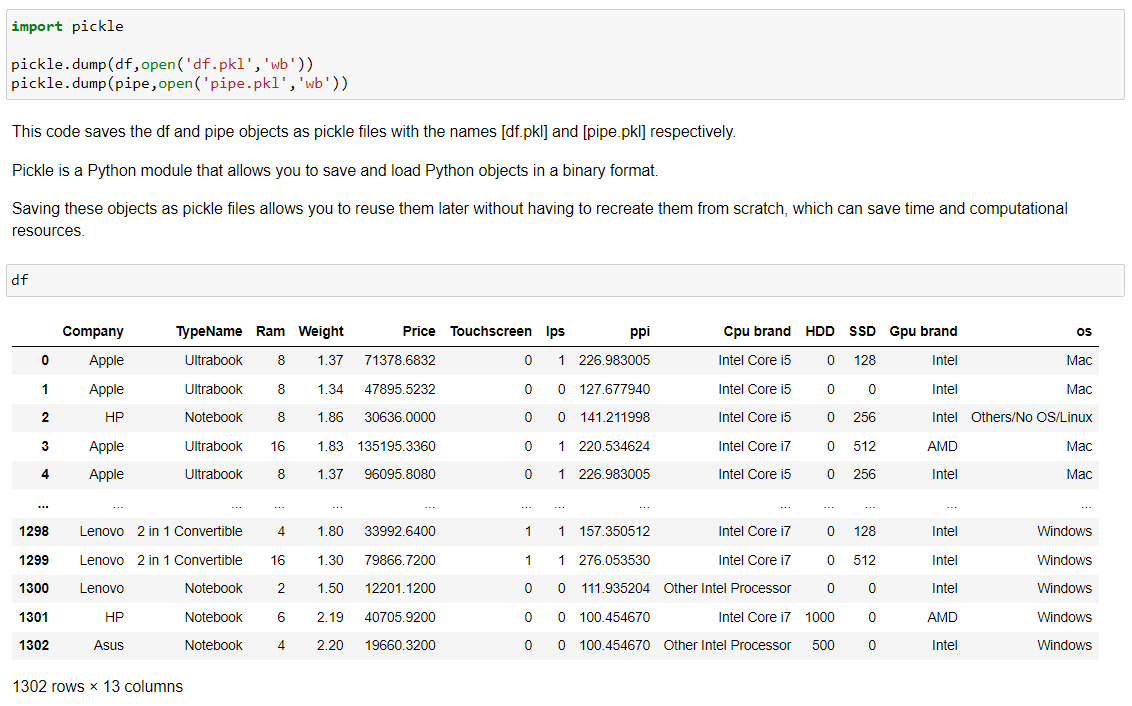
Description automatically generated**

**Random Forest:**

**Graphical user interface, text

Description automatically generated**

**Saving the Model:**

****

**Conclusion:**

The data contains various types of features, including categorical and numerical features. The data was cleaned by removing the duplicates and filling in the missing values using mean or median of the respective columns.

Exploratory Data Analysis (EDA) revealed several insights, such as HP and Dell are the top two brands in terms of the number of laptops sold, and the average price of laptops with discrete graphics cards is higher than that of laptops without them.

Machine learning models were built using various algorithms such as linear regression, lasso regression, ridge regression, SVM, decision trees and random forest to predict the laptop prices.

Among all the models, the random forest model performed the best, with an R2 score of 0.887 and MAE of 0.159.

After performing various regression algorithms on the laptop price prediction dataset, we can conclude that Random Forest Regression algorithm gave the best results.

* The Random Forest Regression model has an R2 score of 0.8873 and MAE of 0.1586 which means the model can explain almost 88.7% of the variance in the target variable and the average difference between the actual and predicted values is 0.1586.
* After Random Forest, Decision Tree gave a good R2 score of 0.8433 and MAE of 0.1809, and Lasso Regression gave an R2 score of 0.8071 and MAE of 0.211.
* Ridge Regression gave a slightly better result than Lasso Regression with an R2 score of 0.8127 and MAE of 0.209.
* Linear Regression and SVM gave R2 score of 0.8073 and 0.8083 respectively.
* The Decision Tree and Random Forest models gave lower R2 scores compared to other models with R2 scores of 0.18 and 0.20 respectively.

On a conclusive note, the Random Forest Regression model could be used for predicting laptop prices, when given relevant features. The accuracy of the model could be improved by gathering more data and fine-tuning hyperparameters.

The insights derived from the machine learning models can help businesses in making data-driven decisions, such as deciding on the price of new laptops or optimizing the inventory levels of different brands and configurations of laptops.

**Dataset Reference:**

https://www.kaggle.com/datasets/mohidabdulrehman/laptop-price-dataset