

INTERNSHIP PROJECT REPORT ON

**Laptop Price Analysis**

Submitted By:

KENNETH AMAN (3PD21AI012)

BACHELOR OF ENGINEERING IN

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

POOJYA DODDAPPA APPA COLLEGE OF ENGINEERING KALABURAGI

Unified Mentors Internship :

Data Analytics 16 weeks 10th Oct to 10th Feb(online) Unified Mentors is a professional training platform specializing in data analytics, AI, ML, DATA Science and more. The organization focuses on equipping aspiring interns with real-world experience, hands on projects, and industry relevant skills under expert mentorship

Address:

Plot No 247B/7,Dhola Kuva Road Hansi City, Hansi Hisar, Haryana 125033 India

Contact Information:

• Phone: +91 6283 800330

• Email: hello@unifiedmentor.com

**Laptop Price Analysis**

**Project Overview**

The project titled "Laptop Price Analysis" aims to predict the price of laptops based on their specifications using machine learning techniques. The dataset contains various features of laptops, including brand, processor, RAM, storage, screen size, GPU, weight, and price. The project involves data preprocessing, feature selection, model training, and evaluation.

**Problem Statement**

The primary objective of this project is to develop a machine learning model that can accurately predict the price of laptops based on their specifications. This involves analyzing the dataset, preprocessing the data, selecting relevant features, training a regression model, and evaluating its performance.

**Tools Used**

* **Programming Language:** Python
* **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
* **Development Environment:** VS Code, Jupyter Notebook

**Data Sources**

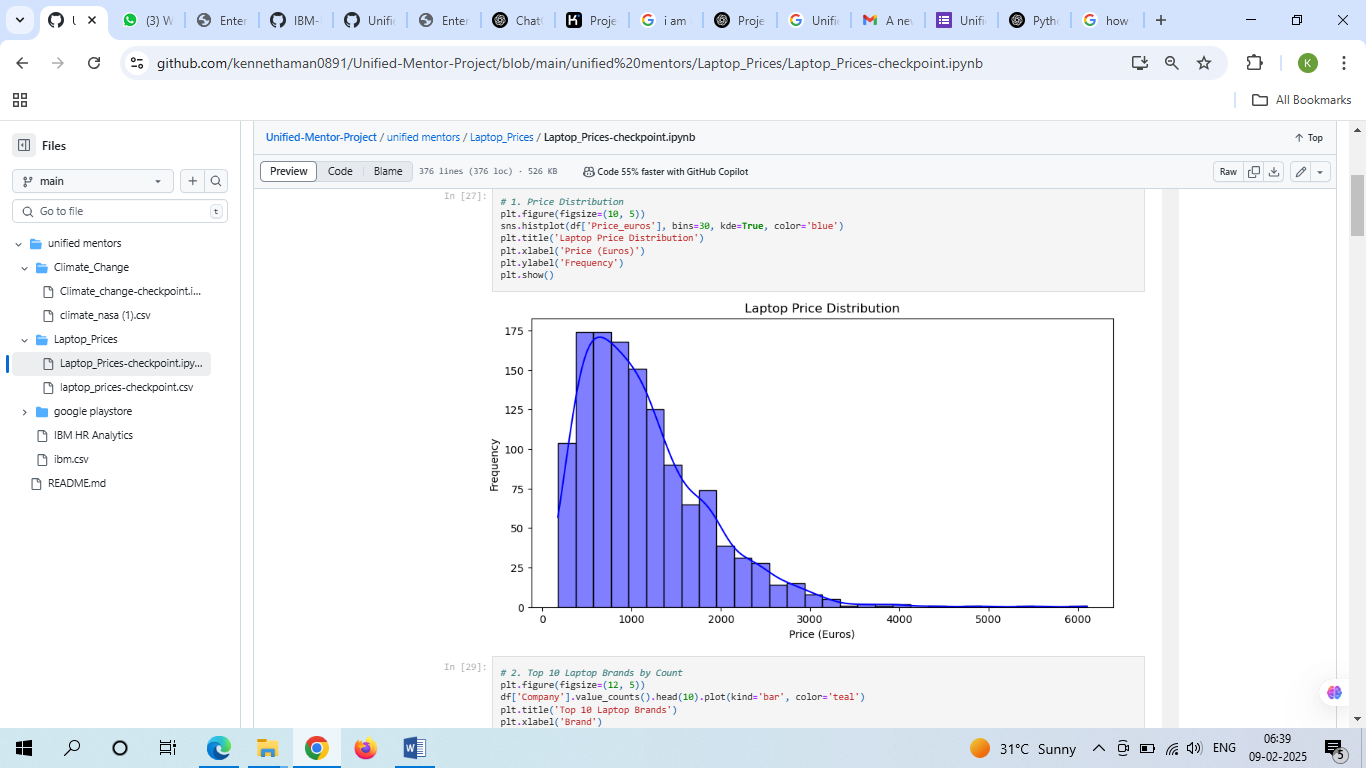
The dataset used in this project is available through a provided link. It contains detailed information about various laptops, including their specifications and prices. The dataset is in CSV format and includes the following columns:

* Company
* Product
* TypeName
* Inches
* Ram
* OS
* Weight
* Price\_euros
* Screen
* ScreenW
* ScreenH
* Touchscreen
* IPSpanel
* RetinaDisplay
* CPU\_company
* CPU\_freq
* CPU\_model
* PrimaryStorage
* PrimaryStorageType
* SecondaryStorage
* SecondaryStorageType
* GPU\_company
* GPU\_model

**Data Analysis and Cleaning**

**Initial Data Inspection**

The dataset was initially inspected to identify any formatting issues or missing values. The dataset was found to be well-formatted with no missing values, making it suitable for analysis.

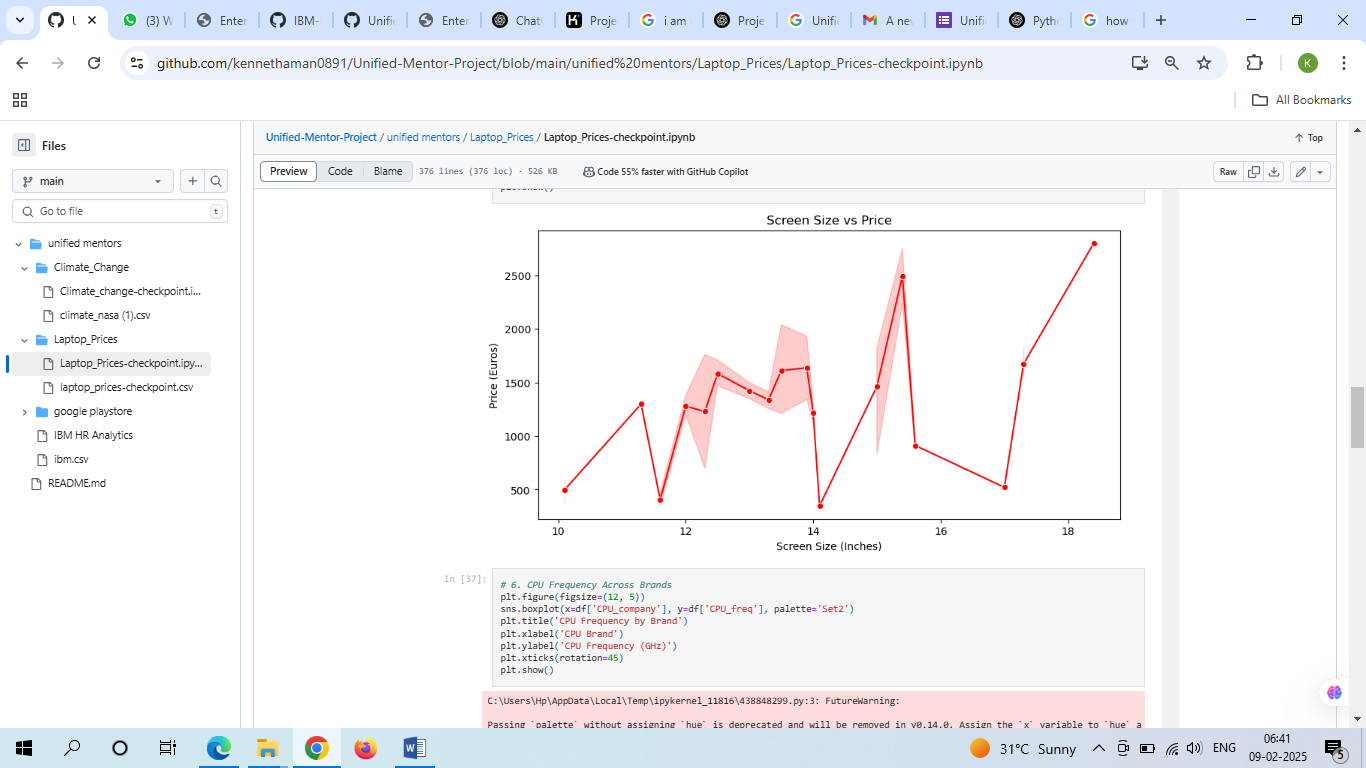


**Data Preprocessing**

1. **Handling Missing Values:** The dataset had no missing values, so no imputation or removal was necessary.
2. **Converting Categorical Data to Numerical:** Categorical columns such as Brand, Processor, and GPU were converted to numerical format using One-Hot Encoding.
3. **Feature Selection:** The features and target variable were selected. The target variable was 'Price\_euros', and the features included all other columns except 'Product' and 'Price\_euros'.

**Train-Test Split**

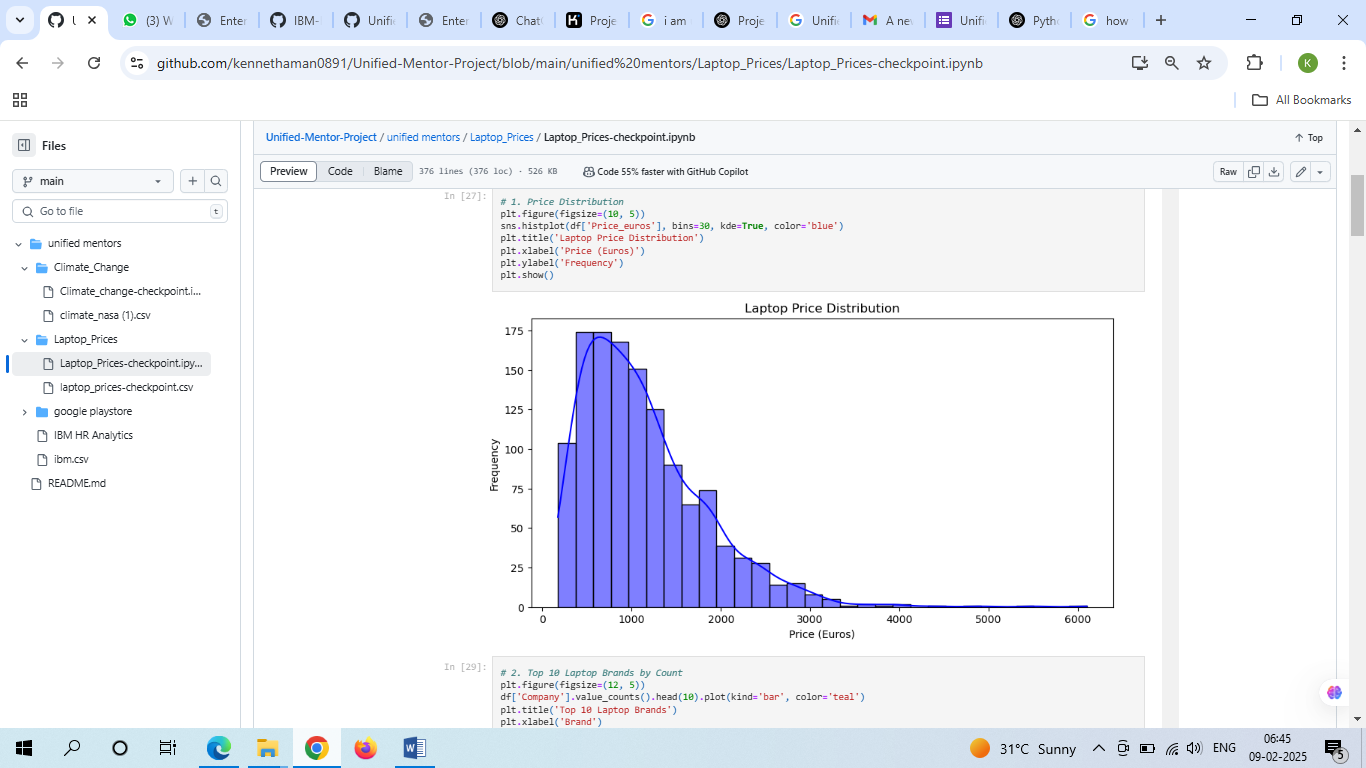
The dataset was split into training and testing sets using an 80-20 split ratio to evaluate the model's performance on unseen data.



**Visualization**

**Exploratory Data Analysis (EDA)**

1. **Univariate Analysis:**
   * Bar plots and pie charts were used to visualize the distribution of categorical variables such as Company, OS, Touchscreen, CPU\_company, GPU\_company, IPSpanel, PrimaryStorageType, and Screen.
   * Histograms and box plots were used to visualize the distribution and outliers in numerical variables such as Inches, Ram, Weight, and Price\_euros.
2. **Bivariate Analysis:**
   * Box plots and bar plots were used to analyze the relationship between categorical variables and the target variable (Price\_euros).
   * Scatter plots were used to visualize the relationship between numerical variables such as CPU\_freq, Ram, and Price\_euros.



**Results**

**Model Training and Evaluation**

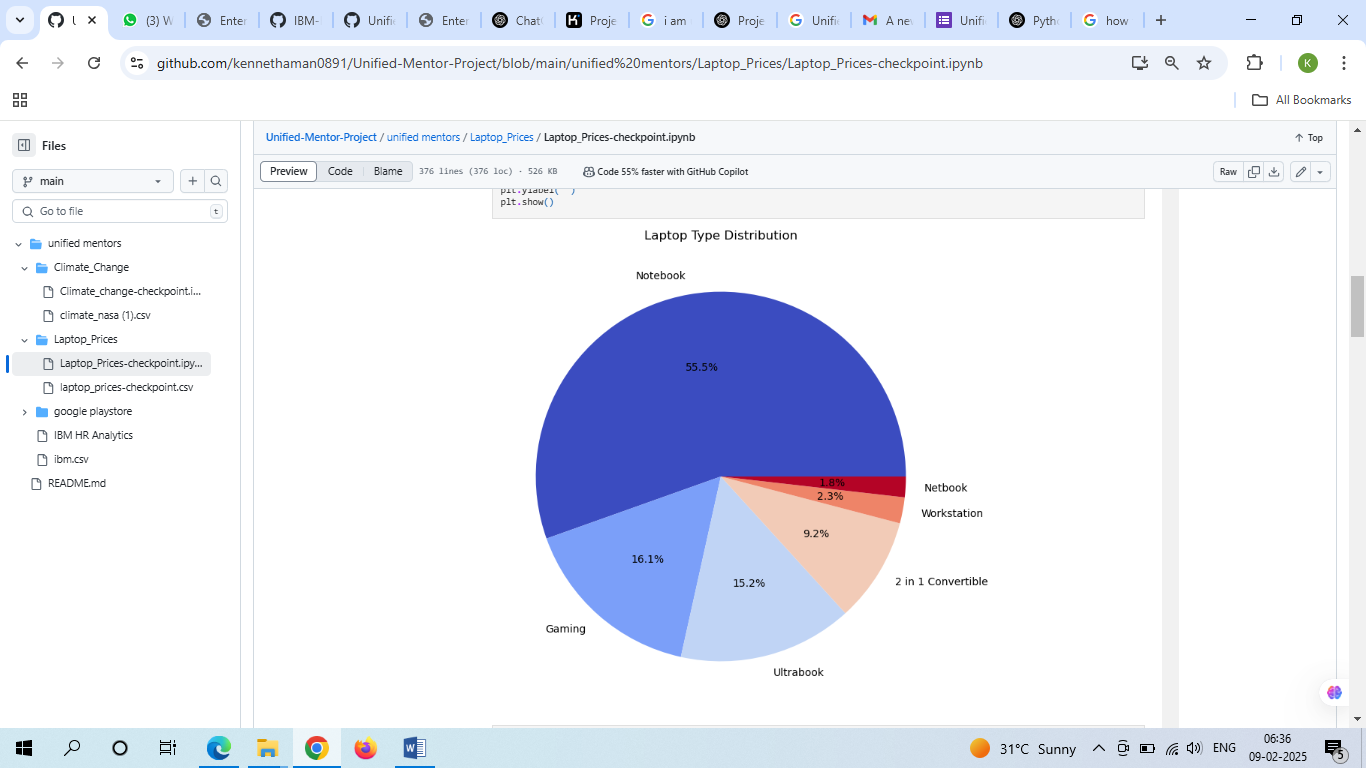
A Linear Regression model was trained on the dataset. The model's performance was evaluated using Mean Squared Error (MSE) and R-squared (R²) metrics.

* **Mean Squared Error (MSE):** The MSE was calculated to measure the average squared difference between the predicted and actual prices.
* **R-squared (R²):** The R² metric was used to determine the proportion of variance in the target variable that is predictable from the features.

**Visualization of Results**

The actual vs predicted prices were visualized using a scatter plot to assess the model's accuracy.

**Recommendations**

1. **Feature Engineering:** Additional features such as brand reputation, user reviews, and market trends could be incorporated to improve model accuracy.
2. **Advanced Models:** Experiment with more advanced regression models such as Random Forest, Gradient Boosting, or XGBoost to potentially improve performance.
3. **Hyperparameter Tuning:** Perform hyperparameter tuning to optimize the model's performance.

**Conclusions**

The project successfully demonstrated the process of predicting laptop prices using machine learning techniques. The Linear Regression model provided a baseline for further improvements. The project highlighted the importance of data preprocessing, feature selection, and model evaluation in building accurate predictive models. Future work could focus on incorporating additional features and experimenting with more advanced models to enhance prediction accuracy.

This comprehensive report provides an overview of the project, the steps taken, the results obtained, and recommendations for future improvements.