Data and Artificial Intelligence
Cyber Shujaa Program

Week 7 Assignment
Regression Models

Student Name: Weldon Kipkoech

Student ID: CS-DA02-25064

Table of Contents

Introduction

This project demonstrates the application of **linear regression** to predict housing prices based on features such as area, number of bedrooms, and age. The analysis is divided into two parts:

- **Univariate Linear Regression** (using area to predict price)
- **Multivariate Linear Regression** (using area, bedrooms, and age)

The goal is to build a model that can accurately estimate house prices, which can be useful for real estate valuation.

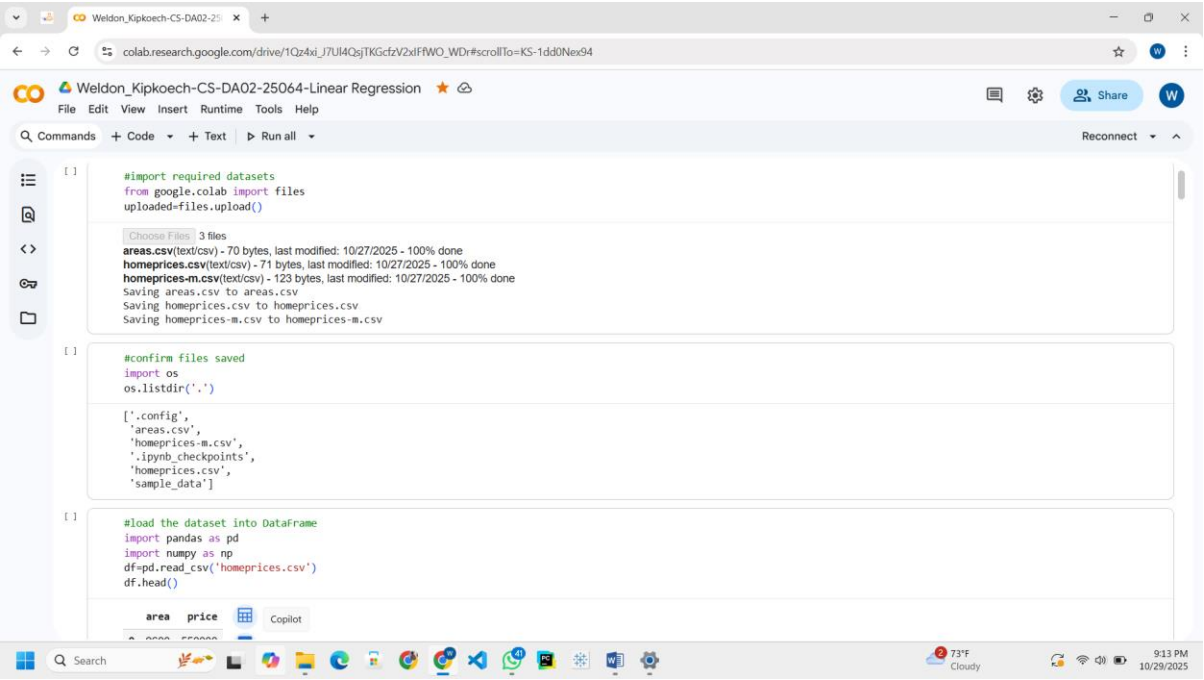
The purpose was to gain hands-on practice in:

- Exploring a real-world dataset
- Preparing and splitting data for training and testing
- Building a simple linear regression model
- Evaluating the model using key metrics
- Visualizing predictions and regression lines
- Publishing your project as part of your portfolio collection

Tasks completed

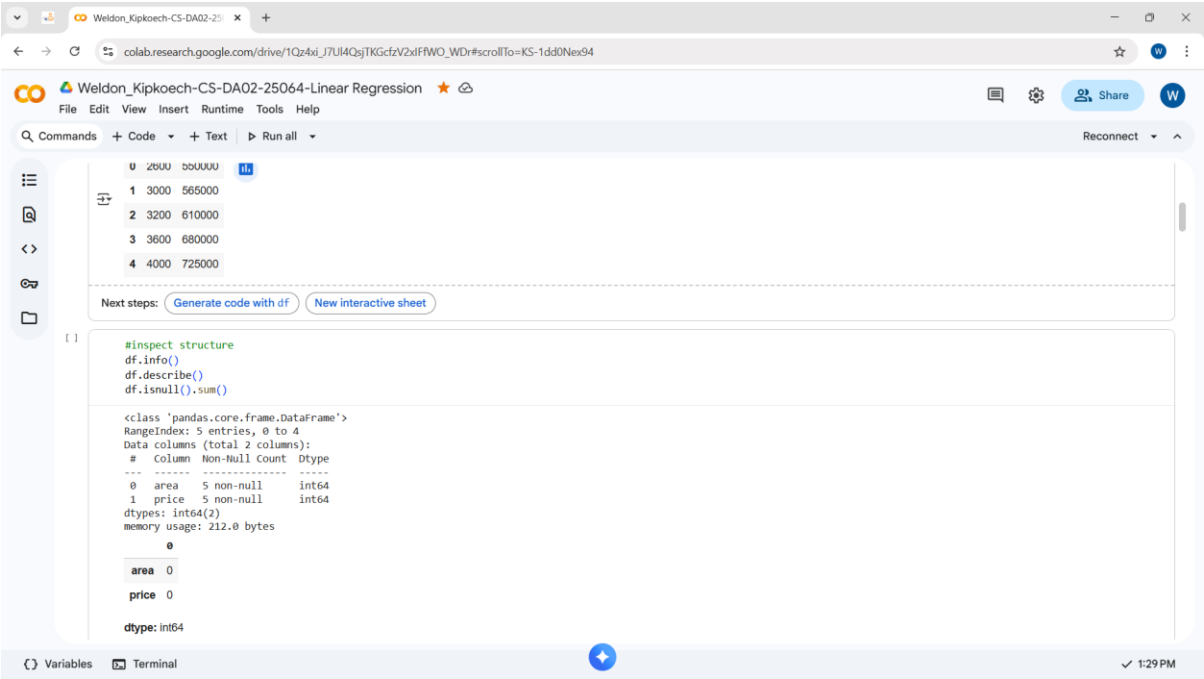
2.1 Data Loading and Exploration

- The dataset was loaded using the pandas library.

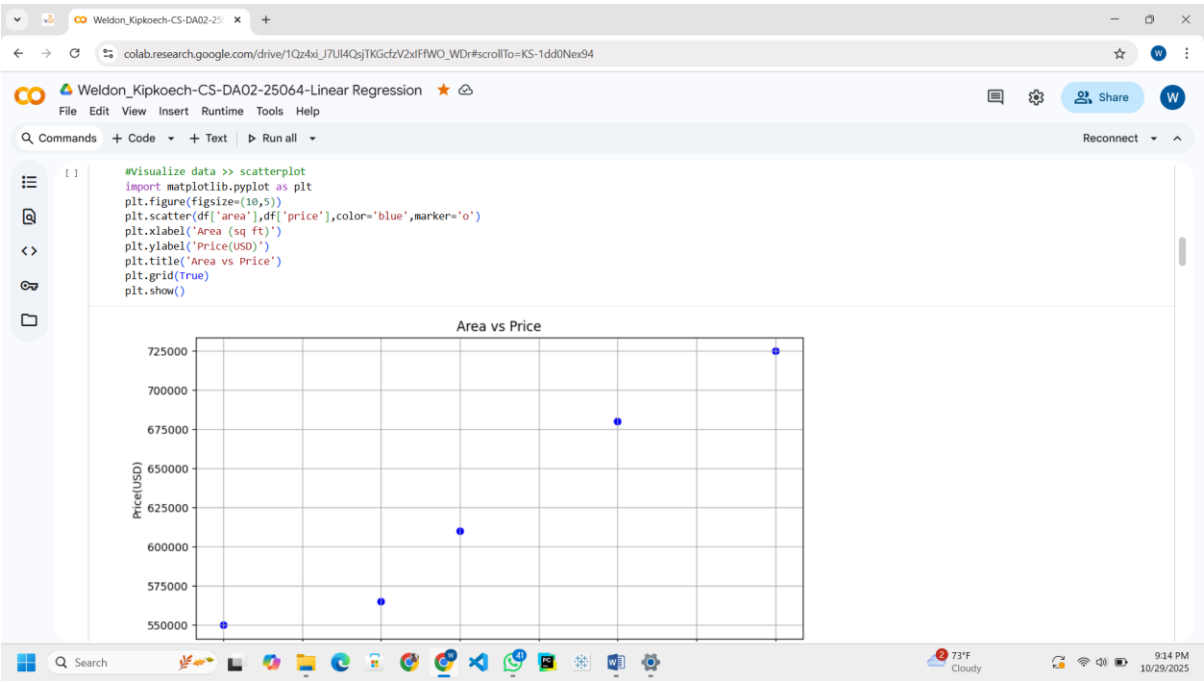


- Initial exploration (.head(), .info(), .describe()) helped understand data structure, feature types, and summary statistics.

- Missing values and outliers were checked and handled accordingly.

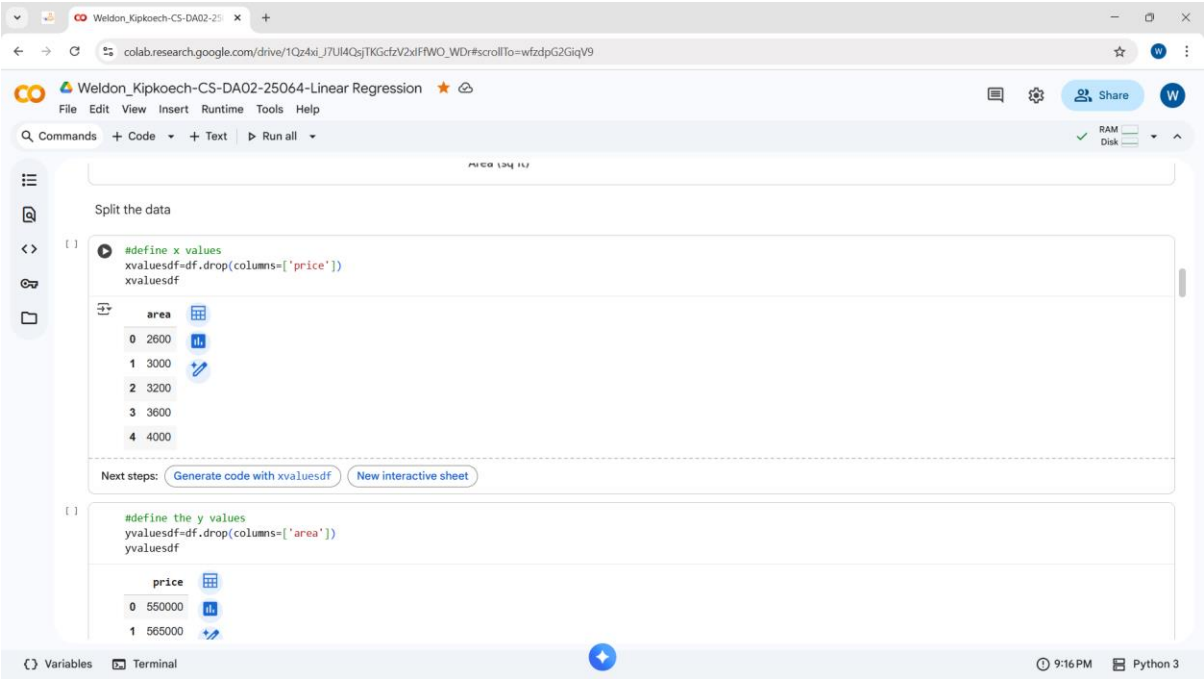


- A scatterplot were plotted to explore data distribution and detect any irregularities.

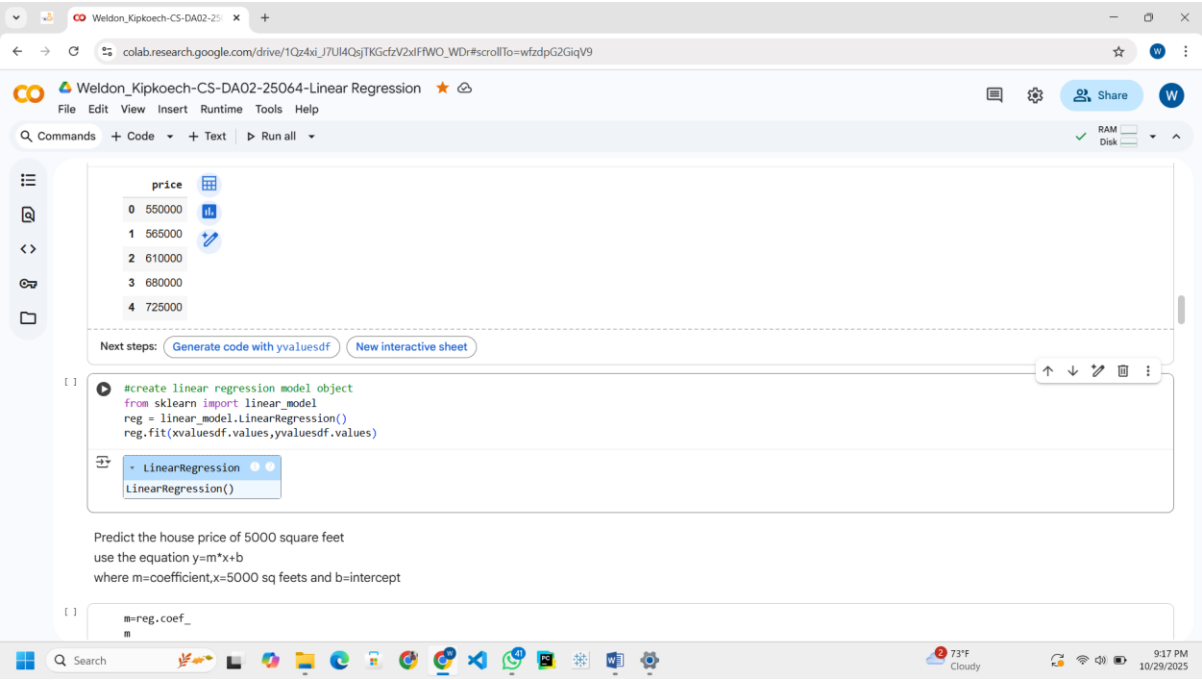


2.2 Model Training

- The dataset was split into training and testing sets using an 80/20 ratio.



- A **Linear Regression model** was trained using the sklearn.linear_model.LinearRegression class.
- After fitting the model, predictions were generated on the test data.

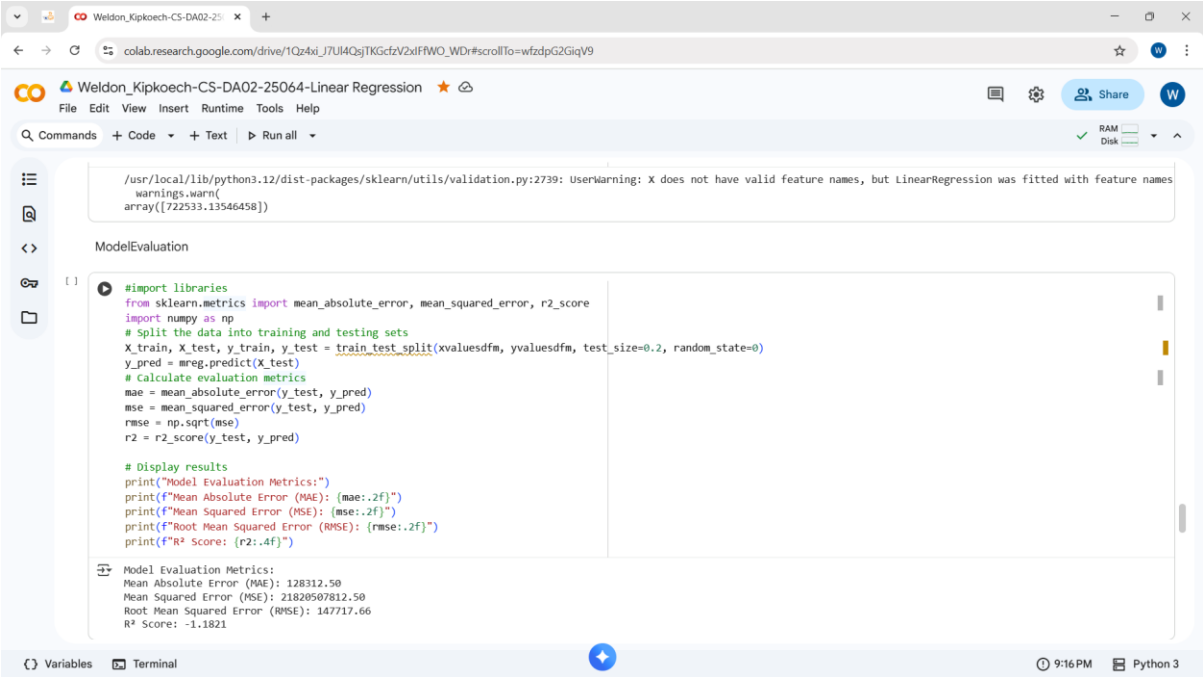


2.3 Model Evaluation

Model performance was evaluated using the following metrics:

- **Mean Absolute Error (MAE):** Measures average magnitude of errors.
- **Mean Squared Error (MSE):** Squares the errors, penalizing larger deviations.

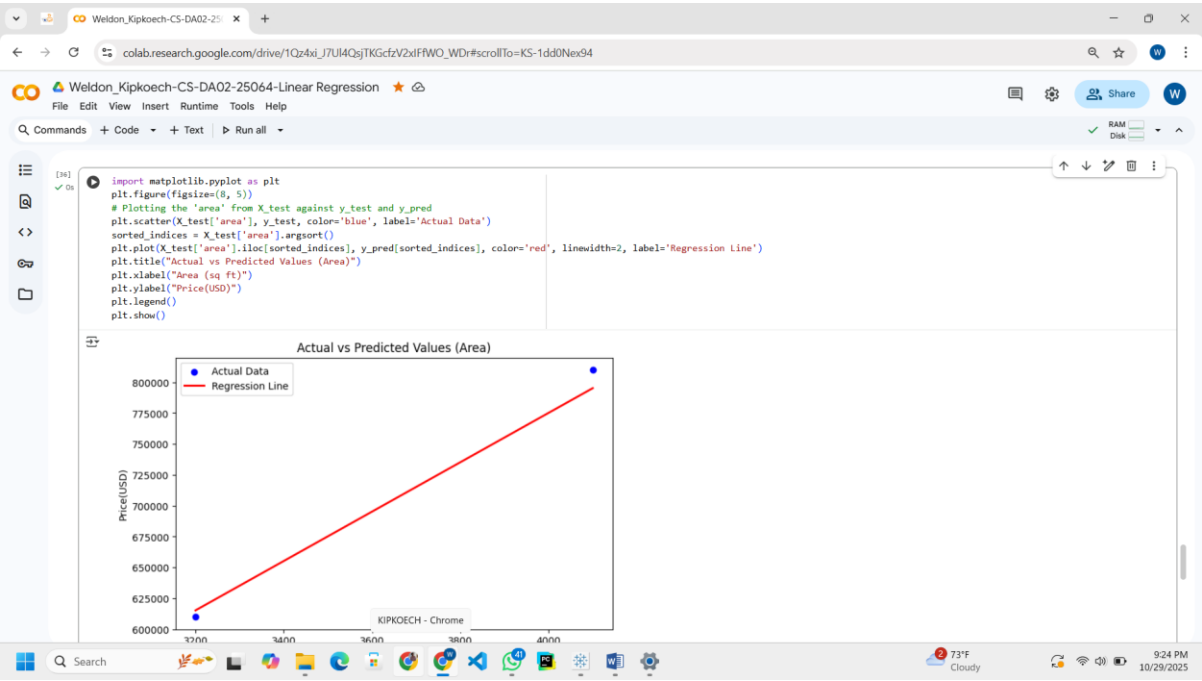
- **Root Mean Squared Error (RMSE):** Square root of MSE, interpretable in the same unit as the target.
- **R^2 Score:** Indicates how well the model explains variance in the target variable.

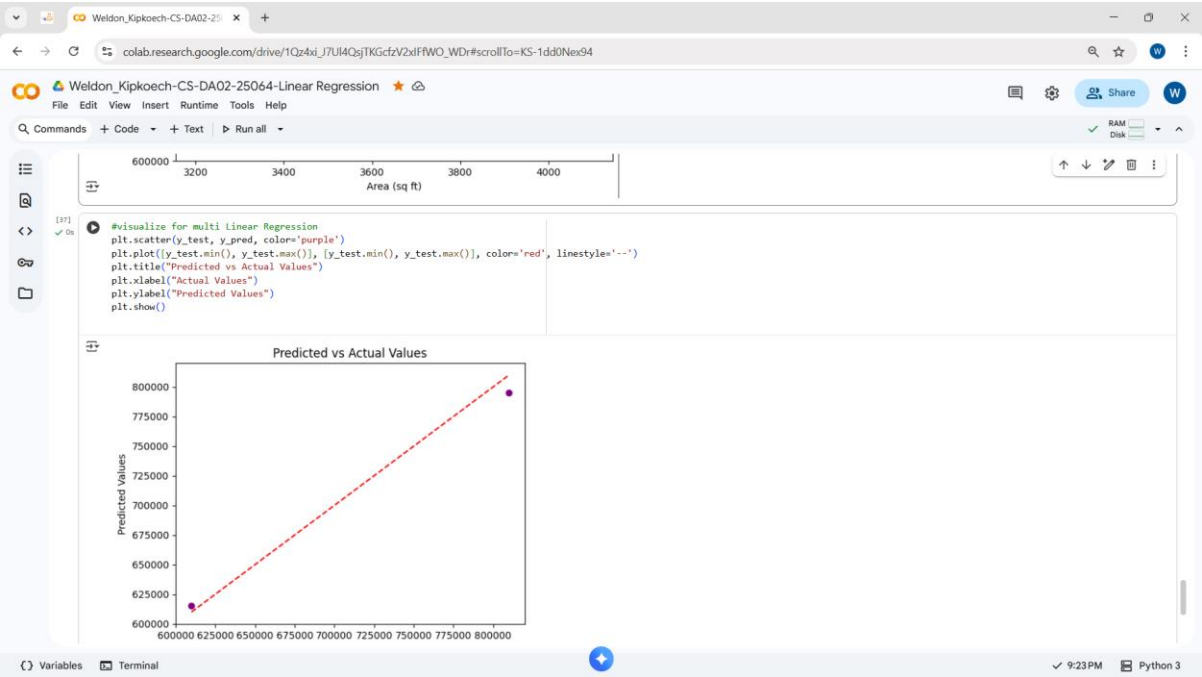


2.4 Graphs and Regression Plots

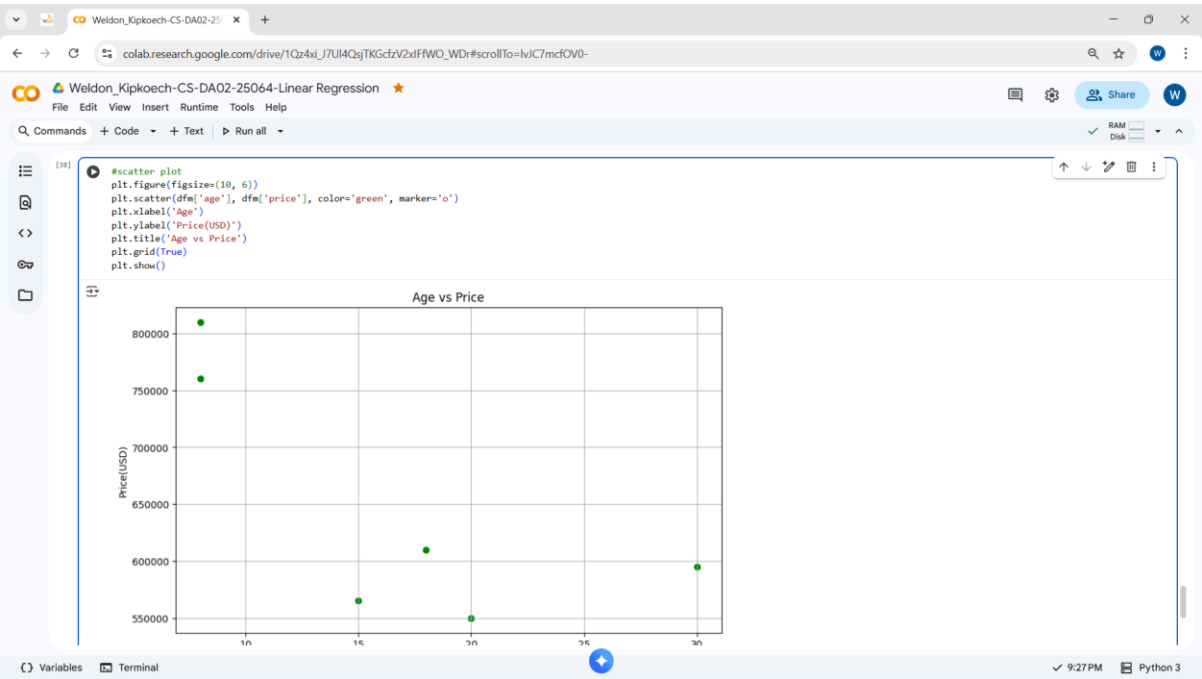
Two visualizations were included:

1. **Regression Line vs Actual Data**
   - Shows how well the predicted line fits actual points.





2. **Predicted vs Actual Values Scatter Plot**
   - Visualizes model accuracy and deviation from the ideal 1:1 line.



Link to Code:
https://colab.research.google.com/drive/1Qz4xi_J7Ul4QsjTKGcfzV2xIFfWO_WDr?usp=sharing

Conclusion

The Linear Regression model successfully learned the relationship between the independent and dependent variables.
From the evaluation results:

- A **low RMSE and MAE** indicated good model accuracy.
- A **high R^2 value** demonstrated that the model explains most of the variance in the dataset.
- Visualization confirmed that predictions closely follow the actual trend.

Key Insights:

- Understanding the mathematical intuition behind regression helps interpret model performance.
- Data cleaning and feature scaling play a crucial role in improving accuracy.
- Visual analysis complements numeric evaluation, ensuring transparency and explainability in machine learning models.