**Aim:** To implement and evaluate Linear Regression for predictive modeling using Python, and analyze the relationship between independent and dependent variables.

**Software Tools:** Google Collaboratory, Python Libraries

**Theory:**

Linear regression is a type of [supervised machine-learning algorithm](https://www.geeksforgeeks.org/machine-learning/supervised-machine-learning/) that learns from the labelled datasets and maps the data points with most optimized linear functions which can be used for prediction on new datasets. It assumes that there is a linear relationship between the input and output, meaning the output changes at a constant rate as the input changes. This relationship is represented by a straight line.

**For example** we want to predict a student's exam score based on how many hours they studied. We observe that as students study more hours, their scores go up. In the example of predicting exam scores based on hours studied. Here

* **Independent variable (input):** Hours studied because it's the factor we control or observe.
* **Dependent variable (output):** Exam score because it depends on how many hours were studied.

**Why is Linear Regression Important?**

* **Simplicity and Interpretability:** It’s easy to understand and interpret, making it a starting point for learning about machine learning.
* **Predictive Ability:** Helps predict future outcomes based on past data, making it useful in various fields like finance, healthcare and marketing.
* **Basis for Other Models:** Many advanced algorithms, like logistic regression or neural networks, build on the concepts of linear regression.
* **Efficiency:** It’s computationally efficient and works well for problems with a linear relationship.
* **Widely Used:** It’s one of the most widely used techniques in both statistics and machine learning for regression tasks.
* **Analysis:** It provides insights into relationships between variables (e.g., how much one variable influences another).

**Hypothesis function in Linear Regression**

In **linear regression**, the hypothesis function is the equation used to make predictions about the dependent variable based on the independent variables. It represents the relationship between the input features and the target output.

For a simple case with one independent variable, the hypothesis function is:

**h(x)=β₀+β₁x**

Where:

* h(x)(ory^) is the predicted value of the dependent variable (y).
* x- xis the independent variable.
* Β₀- β₀ is the intercept, representing the value of y when x is 0.
* Β₁- β₁ is the slope, indicating how much y changes for each unit change in x.

For **multiple linear regressio**n (with more than one independent variable), the hypothesis function expands to:

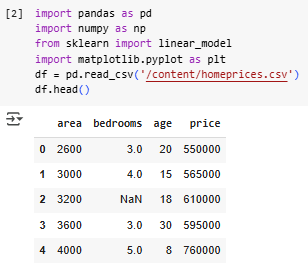
**h(x₁,x₂,...,xₖ)=β₀+β₁x₁+β₂x₂+...+βₖxₖ**

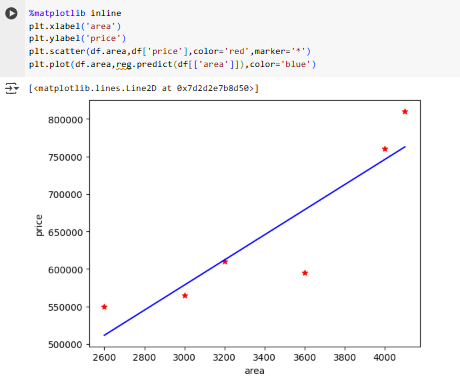
Where:

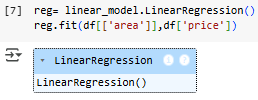
* x₁,x₂,...,xₖ
* x₁,x₂,...,xₖ are the independent variables.
* β₀
* β₀ is the intercept.
* β₁,β₂,...,βₖ
* β₁,β₂,...,βₖ are the coefficients, representing the influence of each respective independent variable on the predicted output.

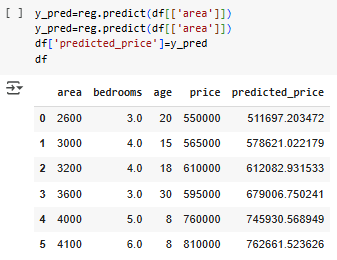
**Output:**

**1)House Pricing:**

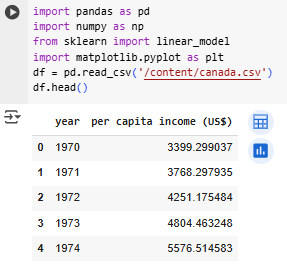
****

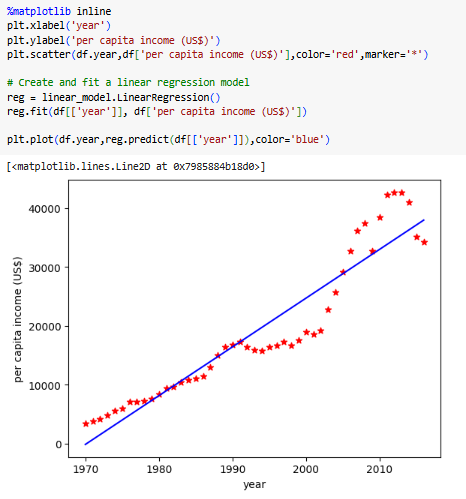
****

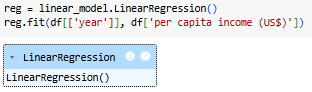
****

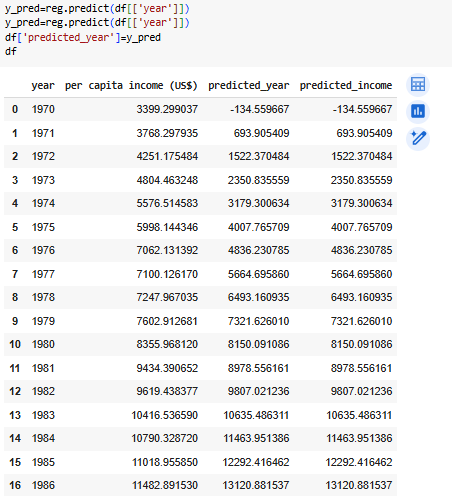
****

**2)Canada**

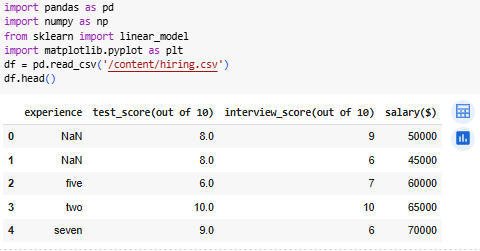


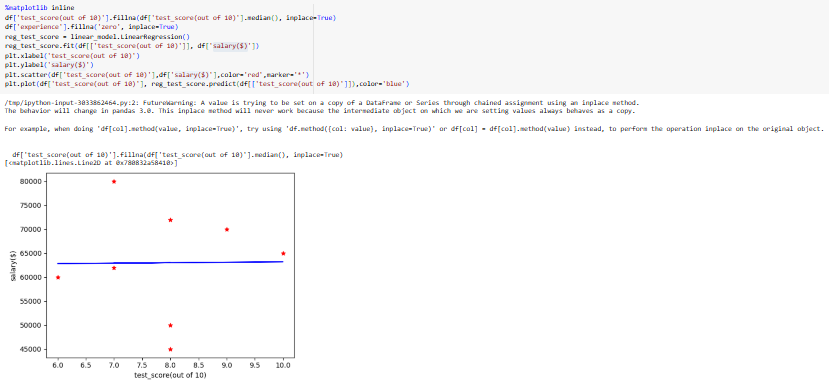


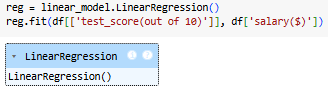


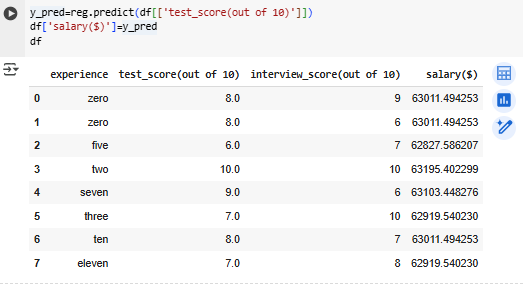


**3) Hiring**

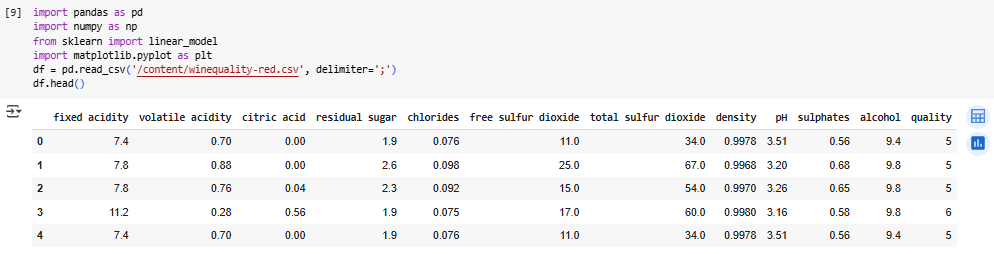
****

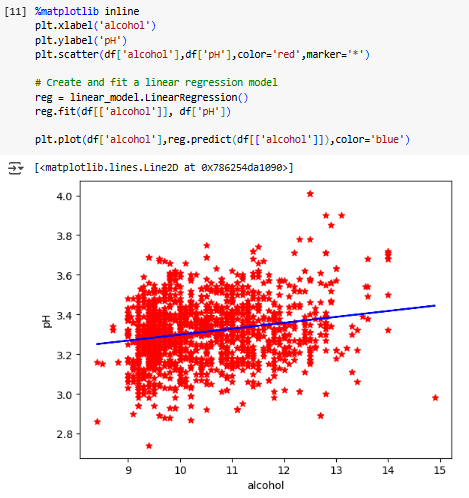
****

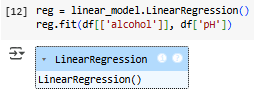
****

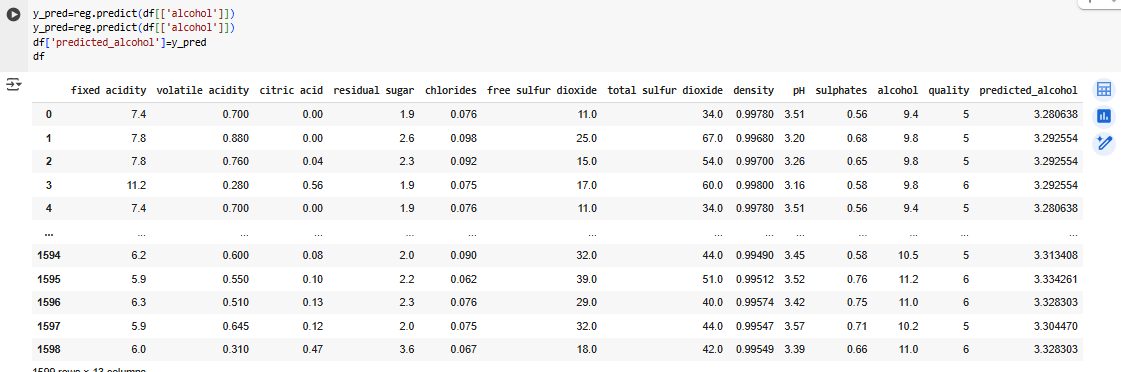
****

**4) Red Wine**

****

****

****

****

**Conclusion:**  Hence, through this experiment, we learned how to preprocess data using NumPy and Pandas to clean, transform, structure raw datasets, model the data and predict the value from actual data. This process is essential for improving the accuracy and effectiveness of data analysis and machine learning models.