**LAB # 03**

**“ IMPLEMENTATION LINEAR REGRESSION USING**

**SCIKIT-LEARN ”**

* **OBJECTIVE:**

To understand the fundamental concepts of linear regression and implement it using

Python. This lab will lay the groundwork for understanding more complex deep learning models.

* **LAB TASKS:**

**TASK 1:**

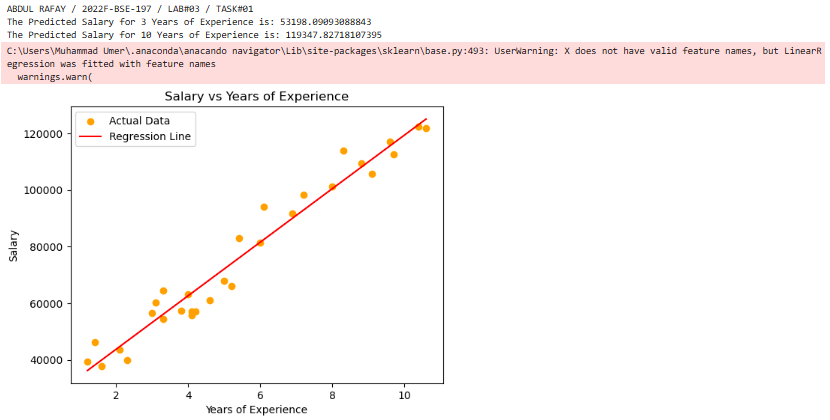
Modify the code in Step 5 to predict the salary for someone with 3 years of experience and

someone with 10 years of experience. Run the code and note down the predicted salaries.

* **CODE:**



* **OUTPUT:**



**TASK 2:**

Investigating the Impact of Outliers on Linear Regression

1. Open the CSV File:

* Open the Salary\_dataset.csv file using a Microsoft Excel.

2. Add the Outlier Data:

* Go to the end of the file (add a new row).
* Enter the following values in the appropriate columns:
* YearsExperience: 1.5
* Salary: 150000

3. Save the Changes:

* Save the modified Salary\_dataset.csv file.

4. Run Your Python Script:

* Now, run your original Python script (from Step 1 onwards). The script will load the modifiedCSV
* file, which now includes the outlier data point.

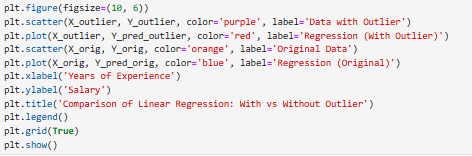
**Analyze the Output:**

Compare the results you obtained in this run (with the outlier) to the results you got when you ran the code with the original dataset (without the outlier). Consider the following questions:

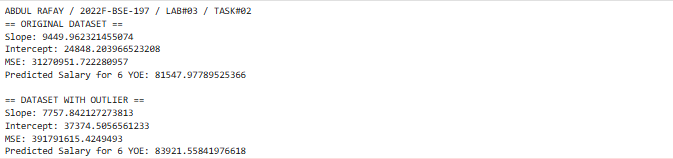
1. How did the regression line on the plot change? Did it tilt more or less? Did it shift up or down?
2. How did the value of the slope (coefficient) change? Did it increase or decrease?
3. How did the value of the intercept (bias) change? Did it increase or decrease?
4. How did the predicted salary for 6 years of experience change? Was it higher or lower?
5. Does the regression line seem to fit the majority of the original data points as well as it did before you added the outlier?
6. What does this experiment demonstrate about the influence of outliers on a linear regression model?

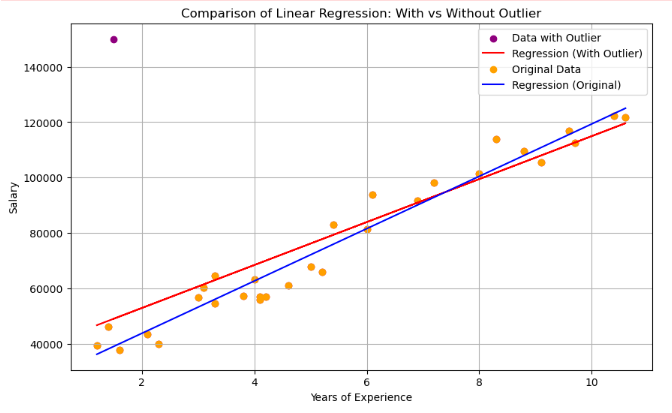
* **CODE:**





* **OUTPUT:**





* **POST ANALYZED OUTPUTS AND REVIEWS:**

1. The regression line with the outlier (red line) has a steeper slope and lower y-intercept compared to the original regression line (blue line). It tilted more and shifted downward at the lower end of the x-axis.
2. The slope increased from 7757.84 in the original dataset to 9449.96 with the outlier, showing a significant increase.
3. The intercept decreased from 37374.59 in the original dataset to 24848.28 with the outlier, a substantial reduction.
4. The predicted salary for 6 years of experience decreased from 83921.56 in the original dataset to 81547.98 with the outlier.
5. No, the regression line with the outlier doesn't fit the majority of data points as well as the original line. It appears to underestimate salaries at lower experience levels and overestimate at higher experience levels.
6. This experiment demonstrates that outliers can significantly influence linear regression models by changing the slope and intercept, which affects predictions. Even a single outlier can pull the regression line toward itself, reducing the model's ability to represent the majority of the data accurately, as evidenced by the increase in MSE from 39179161 to 31270951.

**TASK 3:**

Find the Slope and Intercept of the above code.

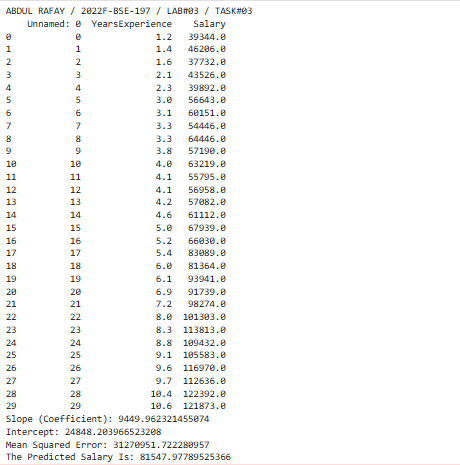
Hint: Slope→ model.coef\_

Intercept→ model.Intercept\_.

* **CODE:**



* **OUTPUT:**





**TASK 4:**

Calculate the Error using Mean Square Error

**Hint:**

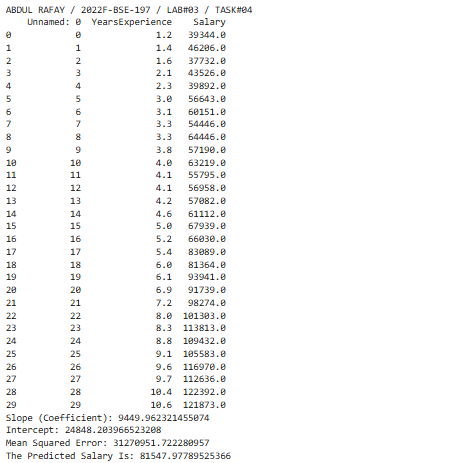
from sklearn.metrics import mean\_squared\_error

mean\_squared\_error(Actual Y, Predicted Y).

* **CODE:**



* **OUTPUT:**





* **GITHUB UPLOAD:**
* **KAGGLE UPLOAD:**