**SIMPLE LINEAR RIGRESSION**

**Business Problem** = Build a prediction model for Salary hike.

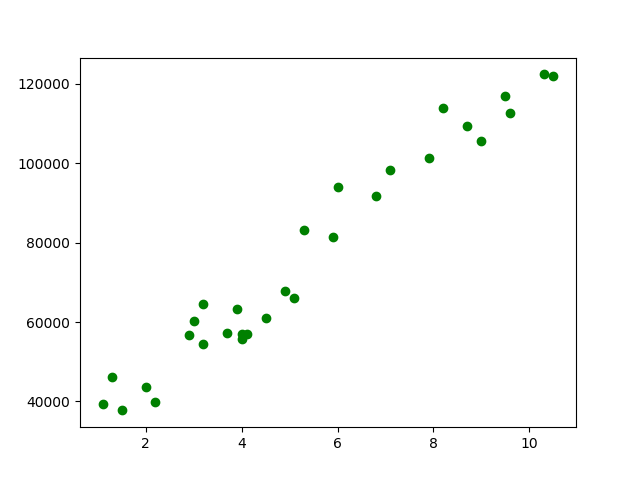
* **Name of the File: -** Salary\_Data.csv
* **Size of the File: -** 454 bytes
* **Data: -** 30 Observation, 2 Variable
* **Column Name: -** YearsExperience, Salary

**Exploratory data Analysis** =

* **Skewness: -** Both variable have Right Skewed data
* **Outliers: -**  Data don’t have Outliers
* **Missing Value: -** Data don’t have Missing Values
* **Normality: -** Data are not normal
* **Transformation: -**  Required

**Liner Regression Model =**

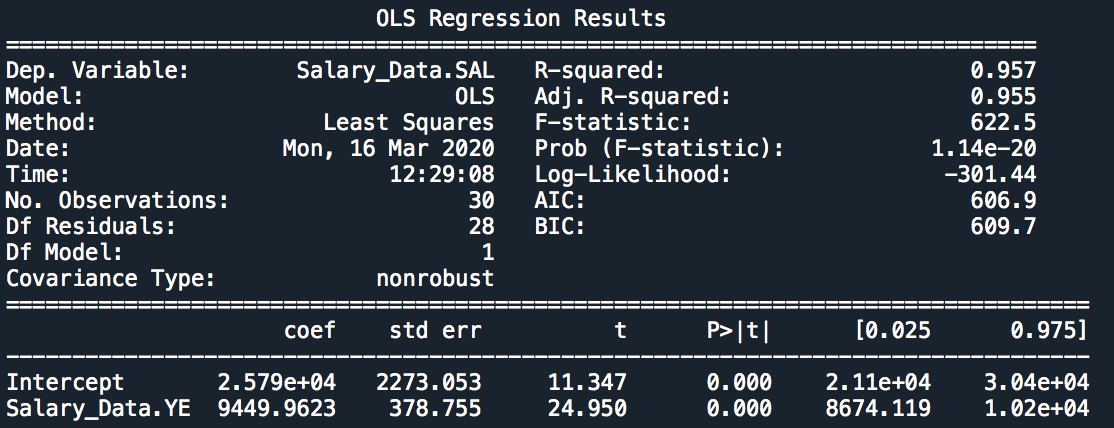
**Scatter plot =** From below scatter plot we can assume that the direction of correlation of data is positive and linear in nature.



**Correlation Coefficient (r) =** Correlation Coefficient In between Weight gain and Calories consumed is equal to 0.98 this means the data have strong positive correlation.

**Model Building =**

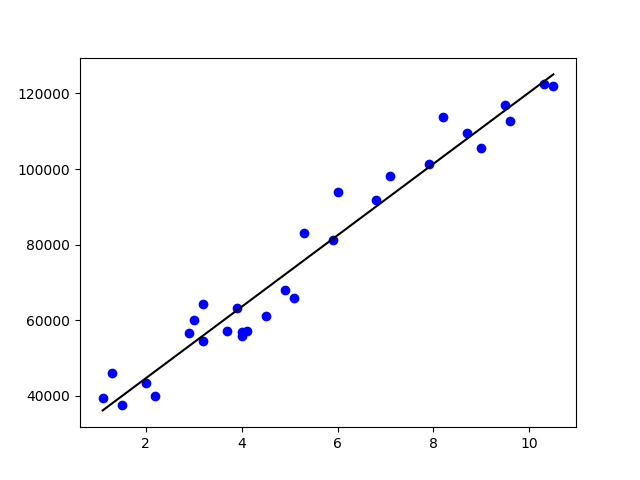
* **Summary: -**

****

* **Intersect(B0): -** 25790
* **Slope(B1): -** 9449.96
* **Coefficient of**

**Determination (R2): -** 0.96 - Strong Correlation

* **Pvalue: -** P value less than 0.05 so we can use B1 and B0 for model building.
* **Model: - Salary\_hike** = 9449.96(YearsExperience) – 25792
* **RMSE:** - 5592

**Final Scatter Plot with Best Fit line**: -

After testing all other transformation we can say that this model give us accurate predication as P value is significant, the Determination of coefficient is high and distribution of data is homoscedasticity .

**Final Models with 95% confidence interval: -**

* **Salary\_hike** = 8674.11(YearsExperience) – 21100
* **Salary\_hike** = 10200(YearsExperience) –30400

**Python code file**: - [Salary Data Analysis .py](https://github.com/nilaydeshmukh0/Simple-Linear-Regression-With-EDA/blob/master/Salary%20Data%20Analysis/Salary%20Data%20Analysis%20.py)