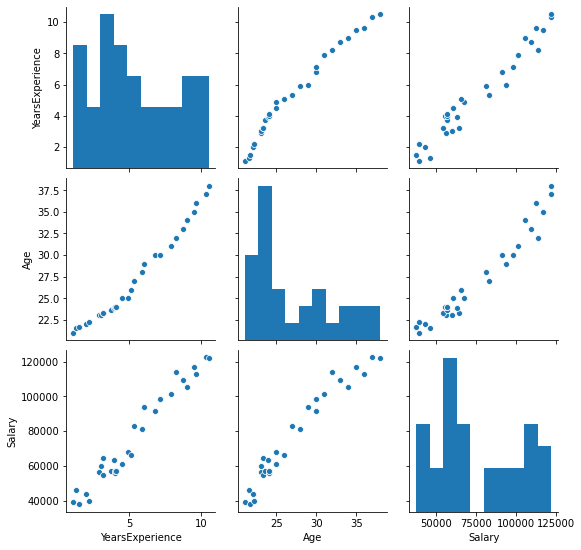
**Salary Data**

X= “Years Of Experience” , “Age”

y= “Salary”



Cor plot.

YearsExperience Age Salary

YearsExperience 1.000000 0.987258 0.978242

Age 0.987258 1.000000 0.974530

Salary 0.978242 0.974530 1.000000

From the above pair-plot and co-relation matrix, it can be said that there is very good correlation between dependent variable and each of the independent variables.

But it is also shown that there is high co-relation of 0.974530 between two independent variables which should not be there and it effects performance.

YearsExperience Age

YearsExperience 1.000000 0.987258

Age 0.987258 1.000000

**Linear Regression Model Summary**

**<class 'statsmodels.iolib.summary.Summary'>**

**""" OLS Regression Results**

**==============================================================================**

**Dep. Variable: Salary R-squared: 0.960**

**Model: OLS Adj. R-squared: 0.957**

**Method: Least Squares F-statistic: 323.9**

**Date: Sun, 28 Mar 2021 Prob (F-statistic): 1.35e-19**

**Time: 20:19:01 Log-Likelihood: -300.35**

**No. Observations: 30 AIC: 606.7**

**Df Residuals: 27 BIC: 610.9**

**Df Model: 2**

**Covariance Type: nonrobust**

**===================================================================================**

**coef std err t P>|t| [0.025 0.975]**

**-----------------------------------------------------------------------------------**

**const -6661.9872 2.28e+04 -0.292 0.773 -5.35e+04 4.02e+04**

**YearsExperience 6153.3533 2337.092 2.633 0.014 1358.037 1.09e+04**

**Age 1836.0136 1285.034 1.429 0.165 -800.659 4472.686**

**==============================================================================**

**Omnibus: 2.695 Durbin-Watson: 1.711**

**Prob(Omnibus): 0.260 Jarque-Bera (JB): 1.975**

**Skew: 0.456 Prob(JB): 0.372**

**Kurtosis: 2.135 Cond. No. 626.**

**==============================================================================**

**Warnings:**

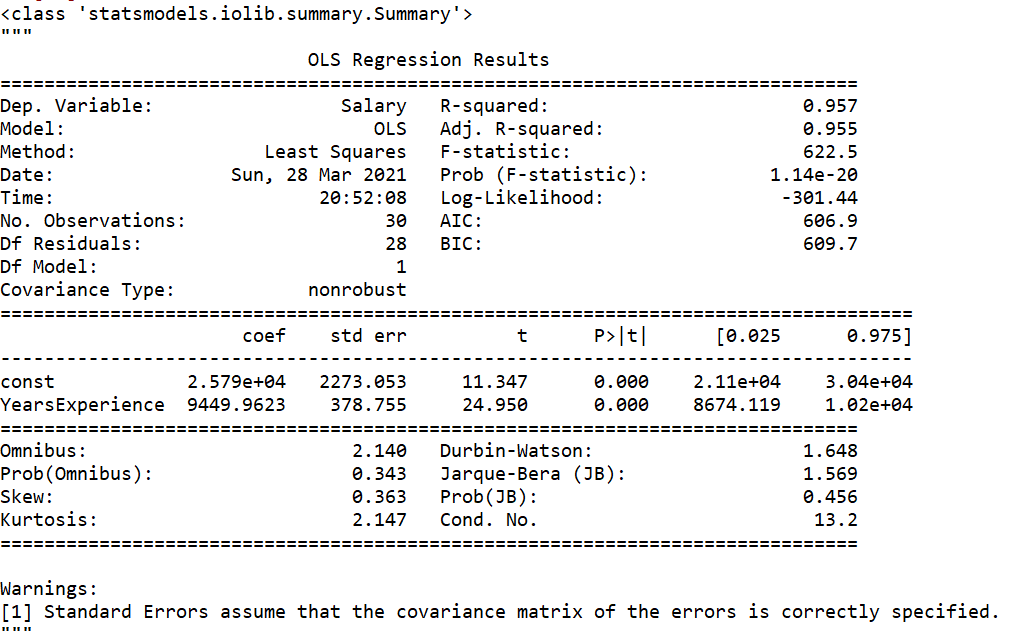
**[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.**

"""

From the above model Summary, if we observe the standard error, it can be said that this is very high standard error which signifies that the model is affected by multi-Collinearity which exists among the independent variables. This multi-collinearity problem should be addressed in order to get best model.

As the dataset is small, we can remove one variable from that as both the variables have similar impact on y variable. So p-value will decide which variable to remove. The variable having high p-value will be dropped. Age have p-value = 0.165 more than 0.05 can be removed as it is not having much significant in predicting the dependent variable.

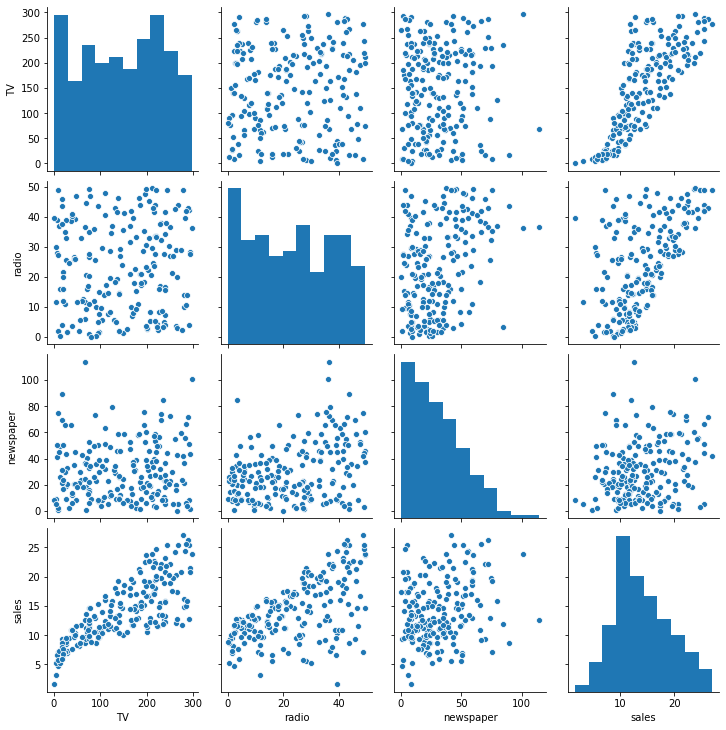
**Linear Model Summary After droping “Age ” variable**

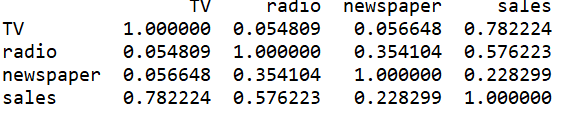


**2. Advertisement Dataset**

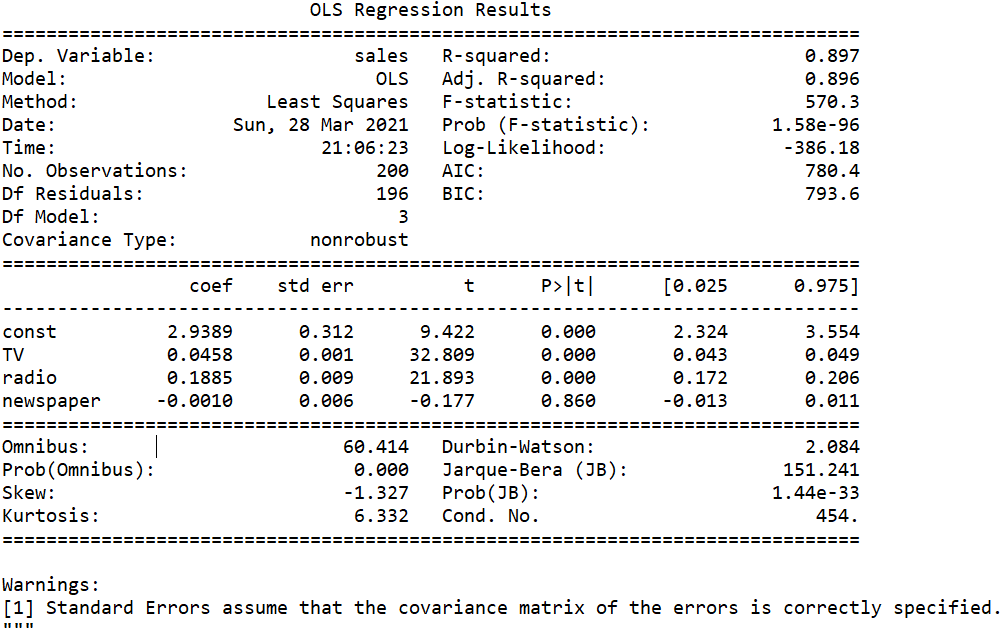
X = “TV”, “radio”, “newspaper”

y = "sales”





**Model Summary**



Dropping the newspaper as its having very high p-value indicating not significant to y-variable and training the model again.

**Model Summary after Dropping the “newspaper”**

