

```
In [1]: import pandas as pd
import numpy as np
```

```
In [3]: data=pd.read_csv(r"C:\Users\Aspire\Desktop\data\Student_Marks.csv")
data
```

Out[3]:

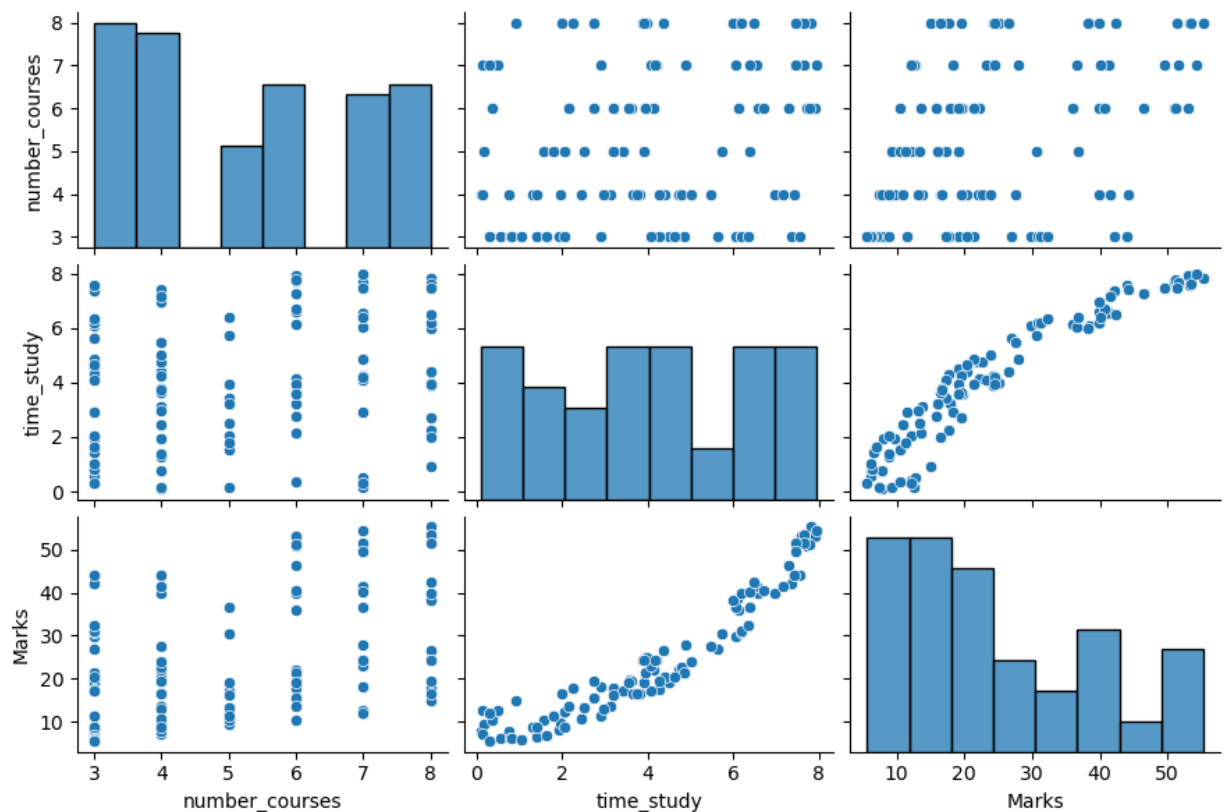
	number_courses	time_study	Marks
0	3	4.508	19.202
1	4	0.096	7.734
2	4	3.133	13.811
3	6	7.909	53.018
4	8	7.811	55.299
...
95	6	3.561	19.128
96	3	0.301	5.609
97	4	7.163	41.444
98	7	0.309	12.027
99	3	6.335	32.357

100 rows × 3 columns

```
In [4]: import seaborn as sns
sns.pairplot(data,height=2,aspect=1.5)
```

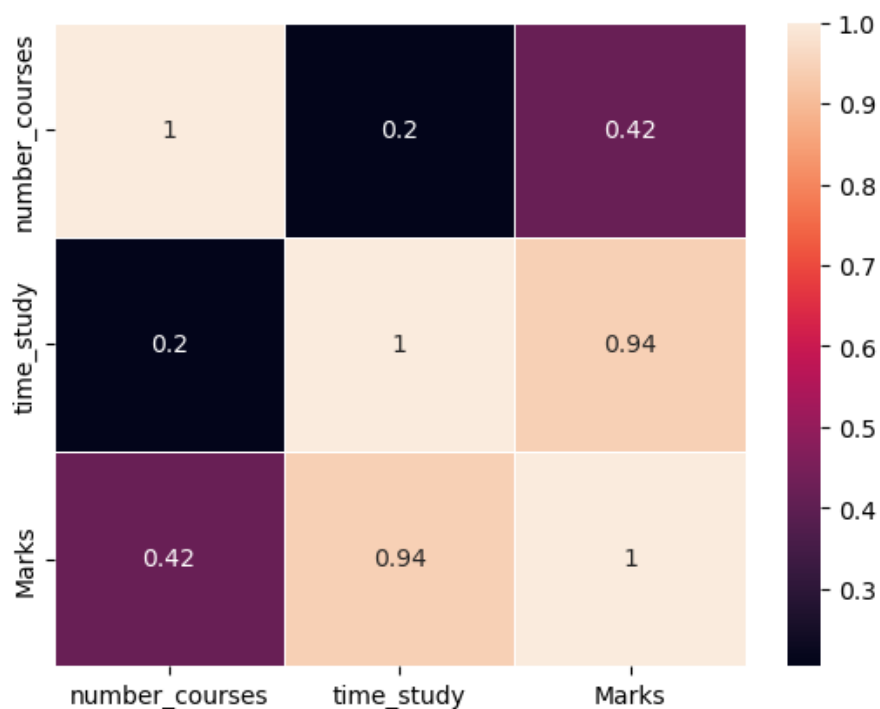
C:\Users\Aspire\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)

Out[4]: <seaborn.axisgrid.PairGrid at 0x1e6e8b75fd0>



```
In [5]: sns.heatmap(data.corr(),annot=True,linewidth=0.6)
```

```
Out[5]: <Axes: >
```



```
In [8]: X=data.drop('Marks',axis='columns')
print(X)
Y=data.drop(['number_courses','time_study'],axis='columns')
print(Y)
```

```
   number_courses  time_study
0                3      4.508
1                4      0.096
2                4      3.133
3                6      7.909
4                8      7.811
..            ...      ...
95               6      3.561
96               3      0.301
97               4      7.163
98               7      0.309
99               3      6.335
```

```
[100 rows x 2 columns]
```

```
   Marks
0  19.202
1   7.734
2  13.811
3  53.018
4  55.299
..      ...
95  19.128
96   5.609
97  41.444
98  12.027
99  32.357
```

```
[100 rows x 1 columns]
```

```
In [9]: from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.4,random_state=2)
print(X_train)
print(X_test)
print(Y_train)
print(Y_test)
```

48	3	5.635
36	4	2.966
78	7	7.451
6	3	6.063
89	7	6.376
91	8	2.730
10	3	7.353

Marks

12	24.318
53	36.653
87	6.053
54	53.158
95	19.128
32	15.038
19	25.133
26	12.647
60	10.522
55	18.238
9	30.862
22	5.600

```
In [10]: from sklearn.linear_model import LinearRegression
eqn=LinearRegression()
```

```
In [11]: eqn.fit(X_train,Y_train)
```

```
Out[11]: ▾ LinearRegression
LinearRegression()
```

```
In [12]: eqn.intercept_
```

```
Out[12]: array([-7.66355012])
```

```
In [13]: eqn.coef_
```

```
Out[13]: array([[2.00809425, 5.26612959]])
```

```
In [16]: Y_test_predicted=eqn.predict(X_test)
Y_test_predicted
```

```
Out[16]: array([[19.21273742],
 [28.7495286 ],
 [48.63970008],
 [19.51120794],
 [32.49391626],
 [32.06549138],
 [16.86761088],
 [31.45631922],
 [ 3.19843733],
 [20.8681705 ],
 [31.72166359],
 [26.84166033],
 [35.96956179],
 [13.67463748],
 [22.10044482],
 [36.64532525],
 [46.03483431],
 [ 7.20952921],
 [13.20765082],
 [ 2.59996694],
 [38.08314814],
 [ 8.62068243],
 [18.83867271],
 [10.65884409],
 [ 6.93925773],
 [21.29455749],
 [ 5.77017696],
 [39.91898947],
 [36.98949206],
 [23.37467867],
 [15.66506495],
 [ 0.87437531],
 [44.99214065],
 [28.03537288],
 [15.98816724],
 [45.63104121],
 [30.28927634],
 [39.96995189],
 [22.77773765],
 [37.08258352]])
```

```
In [17]: from sklearn import metrics

MAE=metrics.mean_absolute_error(Y_test, Y_test_predicted)
print("MAE is",MAE)

MSE=metrics.mean_squared_error(Y_test,Y_test_predicted)
print("MSE is", MSE)

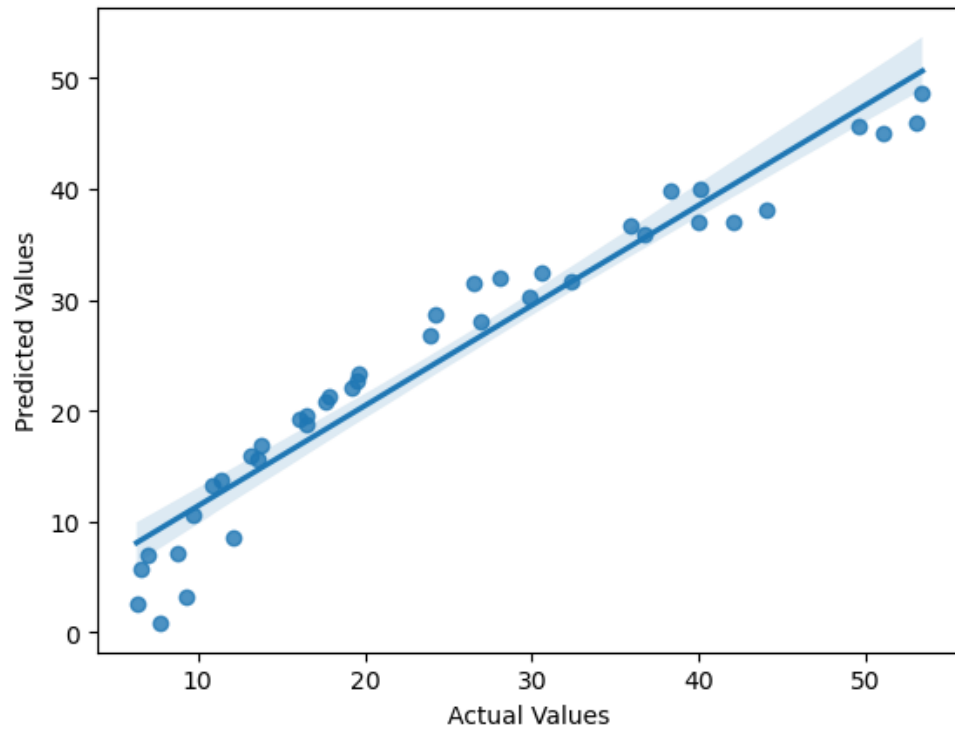
RMSE=np.sqrt(metrics.mean_squared_error(Y_test, Y_test_predicted))
print("RMSE is", RMSE)

r_squared=metrics.r2_score(Y_test, Y_test_predicted)
print("R2 is", r_squared)

MAE is 3.015587980997714
MSE is 12.435520333535068
RMSE is 3.526403314077258
R2 is 0.9372933020994655
```

```
In [18]: import matplotlib.pyplot as plt
sns.regplot(x=Y_test,y=Y_test_predicted)
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
```

Out[18]: Text(0, 0.5, 'Predicted Values')



In []: