```
import numpy as np
import pandas as pd
import matplotlib.pyplot as mtp
%matplotlib inline
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
dataset=pd.read_csv("/content/drive/MyDrive/Personal/Studies/MSC Data Science Material/SEM2/ML/Practical/data_set/student_scores.csv")
```

```
dataset
```

	Hours	Scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30
5	1.5	20
6	9.2	88
7	5.5	60
8	8.3	81
9	2.7	25
10	7.7	85
11	5.9	62
12	4.5	41
13	3.3	42
14	1.1	17
15	8.9	95
16	2.5	30
17	1.9	24
18	6.1	67
19	7.4	69
20	2.7	30
21	4.8	54
22	3.8	35
23	6.9	76
24	7.8	86

```
x=dataset.iloc[:,:1].values
x
```

```
array([[2.5],
       [5.1],
       [3.2],
       [8.5],
       [3.5],
       [1.5],
       [9.2],
       [5.5],
       [8.3],
       [2.7],
       [7.7],
       [5.9],
       [4.5],
       [3.3],
       [1.1],
       [8.9],
       [2.5],
       [1.9],
       [6.1],
```

```
            [7.4],
            [2.7],
            [4.8],
            [3.8],
            [6.9],
            [7.8]])
```

```
y=dataset.iloc[:,1].values
y
```

```
array([21, 47, 27, 75, 30, 20, 88, 60, 81, 25, 85, 62, 41, 42, 17, 95, 30,
       24, 67, 69, 30, 54, 35, 76, 86])
```

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test= train_test_split(x,y, test_size=1/3, random_state=0)
```

```
x_train
```

```
array([[1.1],
       [5.1],
       [7.7],
       [3.3],
       [8.3],
       [9.2],
       [6.1],
       [3.5],
       [2.7],
       [5.5],
       [2.7],
       [8.5],
       [2.5],
       [4.8],
       [8.9],
       [4.5]])
```

```
x_test
```

```
array([[1.5],
       [3.2],
       [7.4],
       [2.5],
       [5.9],
       [3.8],
       [1.9],
       [7.8],
       [6.9]])
```

```
y_train
```

```
array([17, 47, 85, 42, 81, 88, 67, 30, 25, 60, 30, 75, 21, 54, 95, 41])
```

```
y_test
```

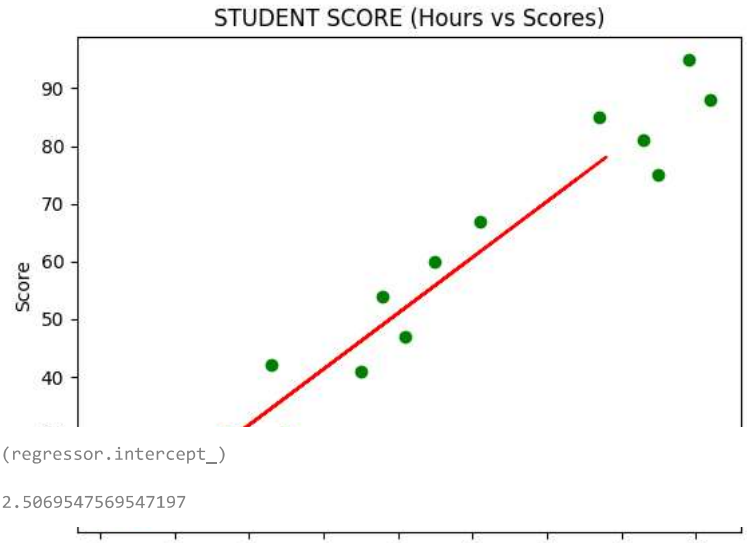
```
array([20, 27, 69, 30, 62, 35, 24, 86, 76])
```

```
from sklearn.linear_model import LinearRegression
regressor=LinearRegression()
regressor.fit(x_train,y_train)
```

```
▾ LinearRegression
LinearRegression()
```

```
y_pred=regressor.predict(x_test)
```

```
mtp.scatter(x_train,y_train, color="green")
mtp.plot(x_test,y_pred, color="red")
mtp.title("STUDENT SCORE (Hours vs Scores)")
mtp.xlabel("Hours")
mtp.ylabel("Score")
mtp.show()
```




```
print(regressor.intercept_)
```

```
2.5069547569547197
```

```
print(regressor.coef_)
```

```
[9.69062469]
```

```
from sklearn import metrics
print("Mean Absolute Error:", metrics.mean_absolute_error(y_test,y_pred))
print("Mean Squared Error:", metrics.mean_squared_error(y_test,y_pred))
print("Root Mean Squared Error:", np.sqrt(metrics.mean_squared_error(y_test,y_pred)))
```

```
Mean Absolute Error: 4.691397441397446
Mean Squared Error: 25.463280738222593
Root Mean Squared Error: 5.046115410711748
```

```
dataset.shape
```

```
(25, 2)
```

```
dataset.describe()
```

	Hours	Scores
count	25.000000	25.000000
mean	5.012000	51.480000
std	2.525094	25.286887
min	1.100000	17.000000
25%	2.700000	30.000000
50%	4.800000	47.000000
75%	7.400000	75.000000
max	9.200000	95.000000

✓ 0s completed at 11:11 PM