In [1]:
```
# Write program for linear regression and find parameters like Sum of Squared Errors
#(SSE), Total Sum of Squares (SST), R2, Adjusted R2etc.
# Class : MCA-I                                CA LAB-VII(A): LAB on Machine Learning
```

In [2]:
```
import pandas as pd
import numpy as np

```

In [3]:
```
dataset = pd.read_csv('student_scores.csv')
```

In [4]:
```
dataset.describe()
```

Out[4]:

	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

In [5]:
```
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values
```

In [6]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

In [7]:
```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)

```

Out[7]:
```
LinearRegression()
```

In [8]:
```
X_train.shape
```

Out[8]:
```
(20, 1)
```

In [9]:
```
X_test

```

Out[9]:
```
array([[1.5],
       [3.2],
       [7.4],
       [2.5],
       [5.9]])
```

In [10]:
```
y_pred = regressor.predict(X_test)
```

In [11]:
```
#calculate sse
sse = np.sum((y_test - y_pred)**2)
print(sse)
```

```
107.99384653608699
```

In [12]:
```
#calculate ssr
ssr = np.sum((y_pred - y.mean())**2)
print(ssr)
```

```
2772.5251939399445
```

In [13]:
```
#calculate sst
sst = ssr + sse
print(sst)

```

```
2880.5190404760315
```

In [ ]:
```

```