Prediction using Supervised Machine Learning using Simple Linear Regression

In this task we have to find the students scores based on their study hours. This is a simple Regression problem type because it has only two variables.

```
import pandas as pd
```

```
data = pd.read_csv("/content/StudentHoursScores.csv")
```

```
data.head()
```

	Hours	Scores
0	7.7	79
1	5.9	60
2	4.5	45
3	3.3	33
4	1.1	12

```
data.tail()
```

	Hours	Scores
18	9.6	96
19	4.3	42
20	4.1	40
21	3.0	30
22	2.6	25

```
data.dtypes
```

```
Hours     float64
Scores      int64
dtype: object
```

```
data.shape
```

```
(23, 2)
```

```
data.columns
```

```
Index(['Hours', 'Scores'], dtype='object')
```

```
data.corr()
```

	Hours	Scores
Hours	1.000000	0.997656
Scores	0.997656	1.000000

```
data.describe()
```

	Hours	Scores
count	23.000000	23.000000
mean	4.817391	47.695652
std	2.709688	27.103228
min	1.100000	12.000000
25%	2.650000	27.000000
50%	4.100000	40.000000
75%	7.100000	72.500000
max	9.600000	96.000000

```
type(data)
```

```
pandas.core.frame.DataFrame
```

```
x = data.iloc[:,0:-1]
y = data.iloc[:,-1]
```

```
x.head()
```

	Hours
0	7.7
1	5.9
2	4.5
3	3.3
4	1.1

```
y.head()
```

```
0    79
1    60
2    45
3    33
4    12
Name: Scores, dtype: int64
```

```
x.shape
```

```
(23, 1)
```

```
y.shape
```

```
(23,)
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=.2,random_state=1)
```

```
x_train.shape
```

```
(18, 1)
```

```
y_train.shape
```

```
(18,)
```

```
x_test.shape
```

```
(5, 1)
```

```
y_test.shape
```

```
(5,)
```

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

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

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

```
y_pred
```

```
array([40.87711348, 25.025345  , 32.95122924, 34.9327003 , 42.85858454])
```

```
y_test
```

```
  File "<ipython-input-33-9d23855172e3>", line 1
    y_test"
          ^
SyntaxError: EOL while scanning string literal
```

SEARCH STACK OVERFLOW

```
x_test
```

	Hours
20	4.1
17	2.5
3	3.3
13	3.5
19	4.3

```
from sklearn.metrics import r2_score, mean_squared_error
print("The coefficient is: ", std_model.coef_)
print("The intercept is: ", std_model.intercept_)
print("The Accuracy is: ", r2_score(y_test,y_pred))
print("The Mean Square Error: ", mean_squared_error(y_test,y_pred))
```

```
The coefficient is:  [9.9073553]
The intercept is:  0.2569567372371395
The Accuracy is:  0.9311770139374417
The Mean Square Error:  3.7164412473781487
```

```
vif = 1/1-(r2_score(y_test,y_pred))
vif
```

```
0.06882298606255832
```