

In [1]: `#import libraries`
`import pandas`

In [2]: `import numpy as np`

In [3]: `import pandas as pd`

In [6]: `import matplotlib.pyplot as plt`

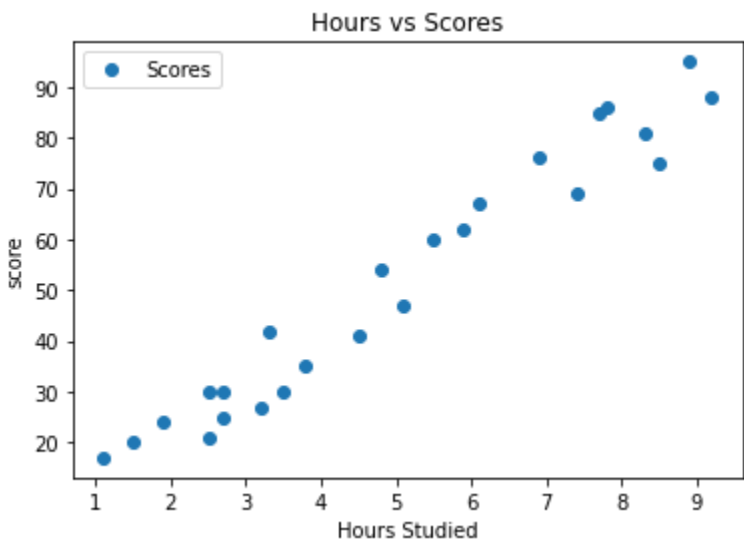
In [22]: `#read data`

`add= "http://bit.ly/w-data"`
`st_data=pd.read_csv(add)`
`st_data`

Out[22]:

	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

In [18]: `df1.plot(x='Hours', y='Scores', style='o')`
`plt.title('Hours vs Scores')`
`plt.xlabel('Hours Studied')`
`plt.ylabel('score')`
`plt.show()`



In [19]: `X=df1.iloc[:, :-1].values`
`Y=df1.iloc[:, 1].values`

In [23]: `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 [24]: `from sklearn.linear_model import LinearRegression`
`regressor=LinearRegression()`
`regressor.fit(X_train,Y_train)`

Out[24]: `LinearRegression()`

In [25]: `print(X_test)`

```
[[1.5]
 [3.2]
 [7.4]
 [2.5]
 [5.9]]
```

In [26]: `y_pred=regressor.predict(X_test)`
`print(y_pred)`

```
[16.88414476 33.73226078 75.357018    26.79480124 60.49103328]
```

In [27]: `df2=pd.DataFrame({'Actual': Y_test, 'Predicted': y_pred})`
`df2`

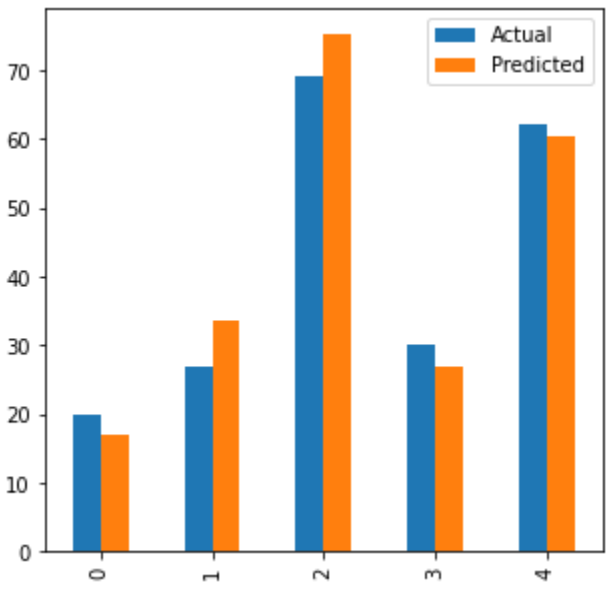
Out[27]:

	Actual	Predicted
0	20	16.884145
1	27	33.732261
2	69	75.357018
3	30	26.794801
4	62	60.491033

In [28]: `print("Training Score:", regressor.score(X_train,Y_train))`
`print("Testing Score: ", regressor.score(X_test,Y_test))`

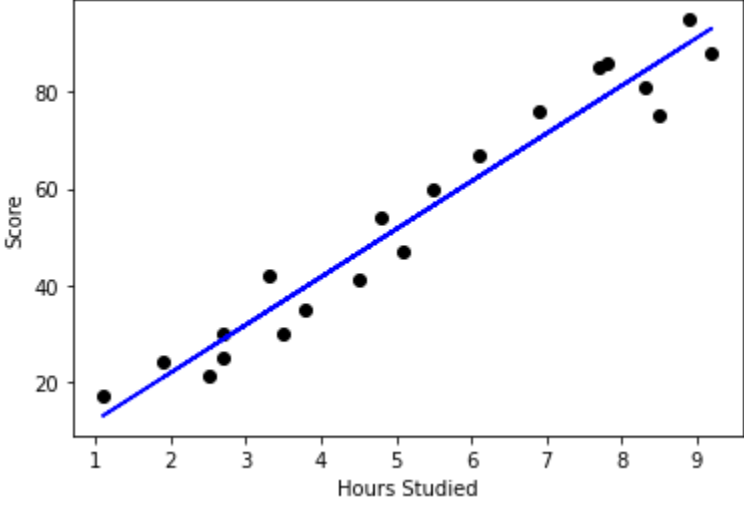
```
Training Score: 0.9515510725211552
Testing Score: 0.9454906892105355
```

In [29]: `df2.plot(kind='bar',figsize=(5,5))`
`plt.show()`



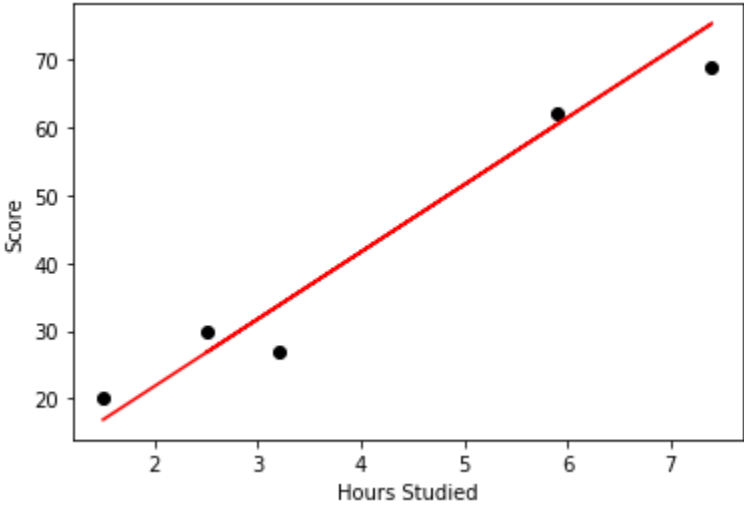
In [32]: `plt.scatter(X_train,Y_train,color='black')`
`plt.plot(X_train,regressor.predict(X_train),color='blue')`
`plt.xlabel('Hours Studied')`
`plt.ylabel('Score')`
`plt.show`

Out[32]: `<function matplotlib.pyplot.show(close=None, block=None)>`



In [33]: `plt.scatter(X_test,Y_test,color='black')`
`plt.plot(X_test,y_pred,color='red')`
`plt.xlabel('Hours Studied')`
`plt.ylabel('Score')`
`plt.show`

Out[33]: `<function matplotlib.pyplot.show(close=None, block=None)>`



In [34]: `hours =np.array(9.25)`
`hours=hours.reshape(1,-1)`
`own_pred = regressor.predict(hours)`
`print("No of Hours = {}".format(hours))`
`print("Predicted Score = {}".format(own_pred[0]))`

`No of Hours = [[9.25]]`
`Predicted Score = 93.69173248737535`

In [36]: `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: 4.183859899002975
Mean Squared Error: 21.598769307217406
Root Mean Squared Error: 4.647447612100367
```

In []: