

# Hi This is Sathwika Madarapu

This is my First task as Data Science & Business Analytics Intern in its Gradual Rotational Internship Program(GRIP)

Prediction using Supervised ML(Simple Linear Regression)

In this regression task we will predict the percentage of marks that a student is expected to score based upon the number of hours they studied.This is a simple linear regression task as it involves just two variables.

```
In [7]: import pandas as pd
import matplotlib.pyplot as plt
```

```
In [8]: dataset=pd.read_csv("https://raw.githubusercontent.com/AdiPersonalWorks/Random/master/student_scores%20-%20student_scores.csv")
```

```
In [9]: dataset
```

Out[9]:

	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 [54]: dataset.head()
```

Out[54]:

	Hours	Scores	Predicted Values
0	2.5	21	27.542383
1	5.1	47	51.855494
2	3.2	27	34.088221
3	8.5	75	83.649562
4	3.5	30	36.893579

```
In [10]: x=dataset[['Hours']]
y=dataset['Scores']
```

```
In [11]: plt.scatter(x,y)
plt.xlabel('Hours Studied')
plt.ylabel('Percentage Score')
```



```
In [13]: from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
```

lets split our data , I will take 90% data for tranning and rest 10% for testing

```
In [14]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.1)
```

```
In [15]: ##making object of Linear regression class
model=LinearRegression()
```

```
In [16]: model.fit(x_train,y_train)
```

```
Out[16]: LinearRegression()
```

```
In [17]: predicted_values=model.predict(x_test)
```

```
In [18]: predicted_values
```

```
Out[18]: array([27.54238303, 77.10372411, 76.16860447])
```

```
In [19]: y_test
```

```
Out[19]: 0      21
24      86
10      85
Name: Scores, dtype: int64
```

```
In [20]: model.score(x_test,y_test)
```

```
Out[20]: 0.9279235589816689
```

```
In [22]: plt.scatter(x,y,marker="+",color="green")
plt.plot(x,model.predict(x))
```



```
In [23]: #predicted values of all data in original dataset
y_predicted=model.predict(x)
```

```
In [24]: dataset[["Predicted Values"]]=y_predicted
```

```
In [25]: dataset
```

Out[25]:

	Hours	Scores	Predicted Values
0	2.5	21	27.542383
1	5.1	47	51.855494
2	3.2	27	34.088221
3	8.5	75	83.649562
4	3.5	30	36.893579
5	1.5	20	18.191187
6	9.2	88	90.195399
7	5.5	60	55.595972
8	8.3	81	81.779322
9	2.7	25	29.412622
10	7.7	85	76.168604
11	5.9	62	59.336451
12	4.5	41	46.244776
13	3.3	42	35.023340
14	1.1	17	14.450708
15	8.9	95	87.390040
16	2.5	30	27.542383
17	1.9	24	21.931665
18	6.1	67	61.206690
19	7.4	69	73.363246
20	2.7	30	29.412622
21	4.8	54	49.050135
22	3.8	35	39.698938
23	6.9	76	68.687647
24	7.8	86	77.103724

What will be the predicted score if a student studies for 9.25 hrs/ day?

```
In [52]: own_pred= model.predict([[9.25]])
print("Predicted Score={}".format(own_pred[0]))

Predicted Score=90.66295893749069
```

```
In [58]: #Evaluating the model
from sklearn import metrics
print("Mean Absolute Error:",metrics.mean_absolute_error(y_test,model.predict(x_test)))

Mean Absolute Error: 8.090018149239139
```

Task Completed