```
NAME: VIBHASREE G M
TASK 1
```

PREDICTION USING SUPERVISED ML

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 [1]:

```
# Importing all libraries required in this notebook
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from scipy import stats
```

In [14]:

```
# Reading data from remote link
url = "http://bit.ly/w-data"
data = pd.read_csv(url)
data.head()
```

Out[14]:

	Hours	Scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30

In [3]:

```
data.shape
```

Out[3]:

```
(25, 2)
```

we can see there are 25 rows and 2 columns in the given dataset

In [4]:

```
data.corr()
```

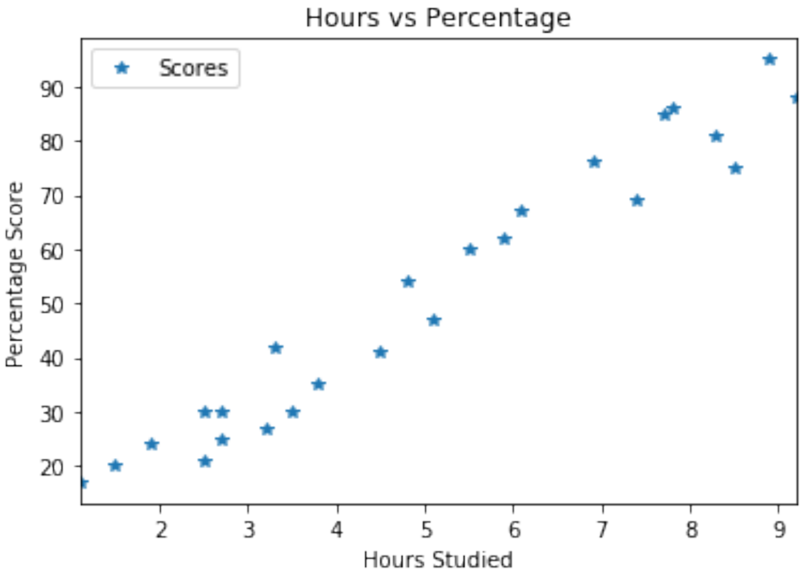
Out[4]:

	Hours	Scores
Hours	1.000000	0.976191
Scores	0.976191	1.000000

the correlation between number of hours of study and scores are 0.976 so can say there is a positive correlation between these two variables.

In [5]:

```
# Plotting the distribution of scores
data.plot(x='Hours', y='Scores', style='*')
plt.title('Hours vs Percentage')
plt.xlabel('Hours Studied')
plt.ylabel('Percentage Score')
plt.show()
```



From the graph above, we can clearly see that there is a positive linear relation between the number of hours studied and percentage of score.

MODELLING

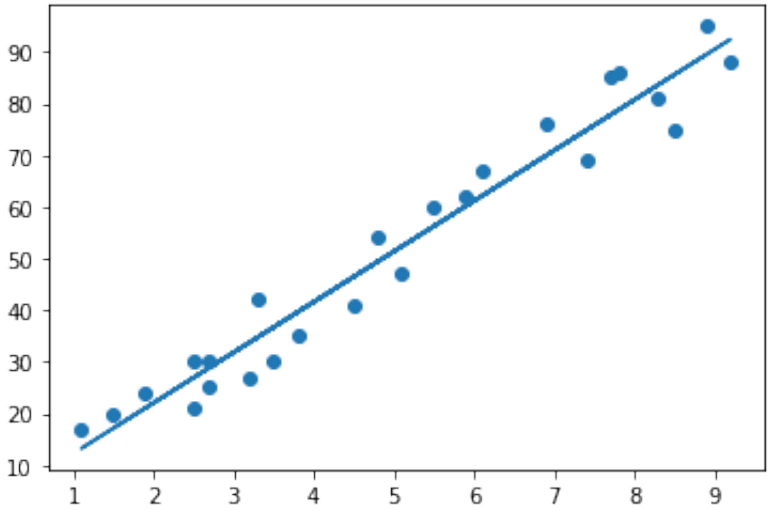
In [7]:

```
x=data["Hours"]
y=data["Scores"]
slope, intercept, r, p, std_err = stats.linregress(x, y)
print("regression coefficient",r)
res = stats.linregress(x, y)
print("Regression result: \n", res)
```

```
regression coefficient 0.9761906560220887
Regression result:
 LinregressResult(slope=9.775803390787473, intercept=2.48367340537
31817, rvalue=0.9761906560220887, pvalue=9.132624467459331e-17, st
derr=0.45294141201804755)
```

score =2.4836734053731817 + 9.775803390787473 * Hours Of Study We,can also observe that the R value is 0.9761 ,this infers that the above model is 97% adequete

In [11]:
```
#Implymenting it WRT to a Function
def myfunc(x):
    return slope * x + intercept
mymodel = list(map(myfunc, x))
plt.scatter(x, y)
plt.plot(x, mymodel)
plt.show()
```



From the above graph,It can be observed that by fitting a abline(Fitted Line) an increasing Trend can be observed ,implying that as the Number of Studing Hours increases the Score gradually Increasing,Interpretina a strong Assosiation Between the above two Variables.

Predicting For Particular Hour 9.25

In [13]:
```
mymodel_Predicted = list(map(myfunc, [9.25]))
print("Predicted score", mymodel_Predicted)
```

```
Predicted score [92.90985477015731]
```

The predicted score if a student studies for 9 Hours 25 Minutes is 92.91 %