In [2]:

```
import pandas as pd
df=pd.read_csv("file.csv")
```

In [3]:

```
df
```

Out[3]:

	number_courses	time_study	Marks
0	3	4.508	19.202
1	4	0.096	7.734
2	4	3.133	13.811
3	6	7.909	53.018
4	8	7.811	55.299
...	...	...	...
95	6	3.561	19.128
96	3	0.301	5.609
97	4	7.163	41.444
98	7	0.309	12.027
99	3	6.335	32.357

100 rows × 3 columns

In [4]:

```
df.value_counts()
```

Out[4]:

```
number_courses  time_study  Marks
3               0.301       5.609     1
6               6.594       39.965    1
7               2.913       18.238    1
                0.508       12.647    1
                0.423       12.132    1
                                     ..
4               3.133       13.811    1
```

```
                2.966        13.119     1
                2.438        10.844     1
                1.954        9.742      1
8               7.811        55.299     1
Length: 100, dtype: int64
```

In [5]:

```
df.keys()
```

Out[5]:

```
Index(['number_courses', 'time_study', 'Marks'], dtype='object')
```

In [6]:

```
df.head
```

Out[6]:

```
<bound method NDFrame.head of     number_courses  time_study   Marks
0                 3       4.508  19.202
1                 4       0.096   7.734
2                 4       3.133  13.811
3                 6       7.909  53.018
4                 8       7.811  55.299
..              ...         ...     ...
95                6       3.561  19.128
96                3       0.301   5.609
97                4       7.163  41.444
98                7       0.309  12.027
99                3       6.335  32.357

[100 rows x 3 columns]>
```

In [7]:

```
df.head()
```

Out[7]:

	number_courses	time_study	Marks
0	3	4.508	19.202
1	4	0.096	7.734
2	4	3.133	13.811
3	6	7.909	53.018
4	8	7.811	55.299

In [8]:

```
df.sample(10)
```

Out[8]:

	number_courses	time_study	Marks
40	4	0.140	7.336
47	4	4.779	22.701
35	3	7.543	43.978
38	7	6.533	41.358
2	4	3.133	13.811
12	7	4.218	24.318
4	8	7.811	55.299
71	5	2.518	13.416
95	6	3.561	19.128
17	8	6.080	38.490

In [9]:

```
df.describe()
```

Out[9]:

	number_courses	time_study	Marks
count	100.000000	100.000000	100.000000
mean	5.290000	4.077140	24.417690
std	1.799523	2.372914	14.326199
min	3.000000	0.096000	5.609000
25%	4.000000	2.058500	12.633000
50%	5.000000	4.022000	20.059500
75%	7.000000	6.179250	36.676250
max	8.000000	7.957000	55.299000

In [10]:
```
df.shape
```

Out[10]:
```
(100, 3)
```

In [11]:
```
df.values
```

Out[11]:
```
array([[ 3.   ,  4.508, 19.202],
       [ 4.   ,  0.096,  7.734],
       [ 4.   ,  3.133, 13.811],
       [ 6.   ,  7.909, 53.018],
       [ 8.   ,  7.811, 55.299],
       [ 6.   ,  3.211, 17.822],
       [ 3.   ,  6.063, 29.889],
       [ 5.   ,  3.413, 17.264],
       [ 4.   ,  4.41 , 20.348],
       [ 3.   ,  6.173, 30.862],
       [ 3.   ,  7.353, 42.036],
       [ 7.   ,  0.423, 12.132],
       [ 7.   ,  4.218, 24.318],
       [ 3.   ,  4.274, 17.672],
       [ 3.   ,  2.908, 11.397],
       [ 4.   ,  4.26 , 19.466],
       [ 5.   ,  5.719, 30.548],
       [ 8.   ,  6.08 , 38.49 ],
       [ 6.   ,  7.711, 50.986],
       [ 8.   ,  3.977, 25.133],
       [ 4.   ,  4.733, 22.073],
       [ 6.   ,  6.126, 35.939],
       [ 5.   ,  2.051, 12.209],
       [ 7.   ,  4.875, 28.043],
       [ 4.   ,  3.635, 16.517],
       [ 3.   ,  1.407,  6.623],
       [ 7.   ,  0.508, 12.647],
       [ 8.   ,  4.378, 26.532],
       [ 5.   ,  0.156,  9.333],
       [ 4.   ,  1.299,  8.837],
       [ 8.   ,  3.864, 24.172],
       [ 3.   ,  1.923,  8.1  ],
       [ 8.   ,  0.932, 15.038],
       [ 6.   ,  6.594, 39.965],
       [ 3.   ,  4.083, 17.171],
       [ 3.   ,  7.543, 43.978],
       [ 4.   ,  2.966, 13.119],
       [ 6.   ,  7.283, 46.453],
       [ 7.   ,  6.533, 41.358],
       [ 6.   ,  7.775, 51.142],
       [ 4.   ,  0.14 ,  7.336],
       [ 6.   ,  2.754, 15.725],
       [ 6.   ,  3.591, 19.771],
       [ 5.   ,  1.557, 10.429],
       [ 4.   ,  1.954,  9.742],
       [ 3.   ,  2.061,  8.924],
       [ 4.   ,  3.797, 16.703],
       [ 4.   ,  4.779, 22.701],
       [ 3.   ,  5.635, 26.882],
       [ 5.   ,  3.913, 19.106],
       [ 6.   ,  6.703, 40.602],
       [ 6.   ,  4.13 , 22.184],
       [ 4.   ,  0.771,  7.892],
       [ 7.   ,  6.049, 36.653],
       [ 8.   ,  7.591, 53.158],
       [ 7.   ,  2.913, 18.238],
       [ 8.   ,  7.641, 53.359],
       [ 7.   ,  7.649, 51.583],
       [ 3.   ,  6.198, 31.236],
       [ 8.   ,  7.468, 51.343],
       [ 6.   ,  0.376, 10.522],
       [ 4.   ,  2.438, 10.844],
       [ 6.   ,  3.606, 19.59 ],
       [ 3.   ,  4.869, 21.379],
       [ 7.   ,  0.13 , 12.591],
       [ 6.   ,  2.142, 13.562],
       [ 4.   ,  5.473, 27.569],
       [ 3.   ,  0.55 ,  6.185],
       [ 4.   ,  1.395,  8.92 ],
       [ 6.   ,  3.948, 21.4  ],
       [ 4.   ,  3.736, 16.606],
       [ 5.   ,  2.518, 13.416],
       [ 3.   ,  4.633, 20.398],
       [ 3.   ,  1.629,  7.014],
       [ 4.   ,  6.954, 39.952],
       [ 3.   ,  0.803,  6.217],
       [ 5.   ,  6.379, 36.746],
       [ 8.   ,  5.985, 38.278],
       [ 7.   ,  7.451, 49.544],
       [ 3.   ,  0.805,  6.349],
       [ 7.   ,  7.957, 54.321],
       [ 8.   ,  2.262, 17.705],
       [ 4.   ,  7.41 , 44.099],
       [ 5.   ,  3.197, 16.106],
       [ 8.   ,  1.982, 16.461],
       [ 8.   ,  6.201, 39.957],
       [ 7.   ,  4.067, 23.149],
       [ 3.   ,  1.033,  6.053],
       [ 5.   ,  1.803, 11.253],
       [ 7.   ,  6.376, 40.024],
       [ 7.   ,  4.182, 24.394],
       [ 8.   ,  2.73 , 19.564],
       [ 4.   ,  5.027, 23.916],
       [ 8.   ,  6.471, 42.426],
       [ 8.   ,  3.919, 24.451],
       [ 6.   ,  3.561, 19.128],
       [ 3.   ,  0.301,  5.609],
       [ 4.   ,  7.163, 41.444],
       [ 7.   ,  0.309, 12.027],
       [ 3.   ,  6.335, 32.357]])
```

In [12]:

```
df.isnull().sum()
```

Out[12]:

```
number_courses    0
time_study        0
Marks             0
dtype: int64
```

In [13]:

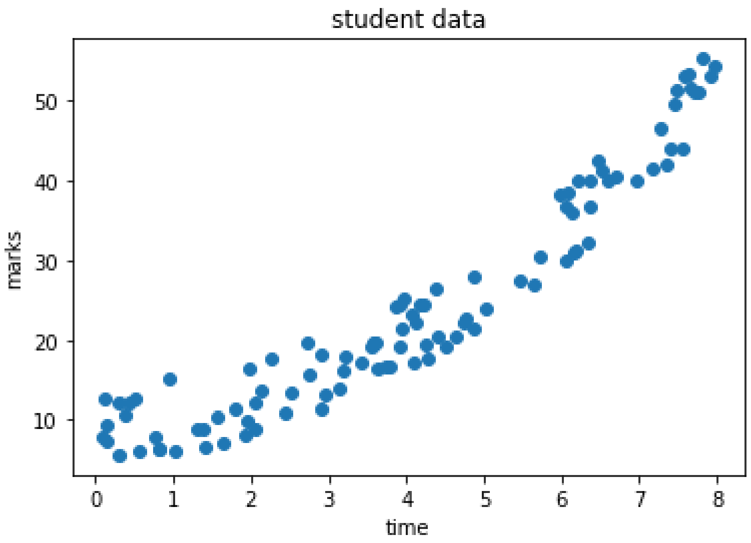
```
import numpy as np
import matplotlib.pyplot as plt
```

In [14]:

```
plt.scatter(x=df.time_study, y=df.Marks)
plt.title("student data")
plt.xlabel("time")
plt.ylabel("marks")
```

Out[14]:

```
Text(0, 0.5, 'marks')
```



In [15]:

```
df.mean()
```

Out[15]:

```
number_courses     5.29000
time_study         4.07714
Marks             24.41769
dtype: float64
```

In [16]:

```
df=df.fillna(df.mean())
df.isnull().sum()
```

Out[16]:

```
number_courses    0
time_study        0
Marks             0
dtype: int64
```

In [17]:

```
df1 = pd.DataFrame(df)
df2=df1.drop(['number_courses','Marks'], axis=1)
df2
```

Out[17]:

	time_study
0	4.508

	time_study
1	0.096
2	3.133
3	7.909
4	7.811
...	...
95	3.561
96	0.301
97	7.163
98	0.309
99	6.335

100 rows × 1 columns

In [18]:

```
df1 = pd.DataFrame(df)
df3=df1.drop(['number_courses','time_study'], axis=1)
df3
```

Out[18]:

	Marks
0	19.202
1	7.734
2	13.811
3	53.018
4	55.299
...	...
95	19.128
96	5.609

	Marks
97	41.444
98	12.027
99	32.357

100 rows × 1 columns

In [19]:
```
x=df2
y=df3
```

In [20]:
```
from sklearn.model_selection import train_test_split
```

In [21]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=51,test_size=0.2)
```

In [22]:
```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

Out[22]:
```
((80, 1), (20, 1), (80, 1), (20, 1))
```

In [23]:
```
from sklearn.linear_model import LinearRegression
```

In [24]:
```
lr=LinearRegression()
lr.fit(x_train,y_train)
lr.score(x_test,y_test)
```

Out[24]:
```
0.8826200571575015
```

In [25]:
```
lr.intercept_
```

Out[25]:
```
array([2.02911007])
```

In [26]:

```
pred=lr.predict(x_test)
```

In [27]:

```
pred
```

Out[27]:

```
array([[15.85824516],
       [10.58032664],
       [ 4.35227296],
       [19.58738841],
       [42.95082791],
       [23.87123486],
       [45.72983599],
       [44.92798781],
       [ 6.4502593 ],
       [ 6.43927507],
       [27.47405957],
       [12.76069465],
       [22.88265491],
       [24.99711759],
       [43.99432897],
       [19.23589332],
       [28.27590775],
       [24.71152782],
       [26.78754572],
       [41.36909999]])
```

In [28]:

```
lr.score(x_test,y_test)
```

Out[28]:

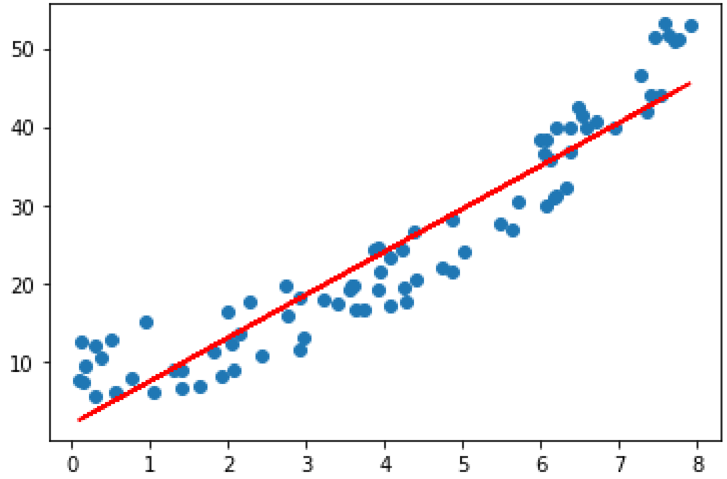
```
0.8826200571575015
```

In [29]:

```
plt.scatter(x_train,y_train)
plt.plot(x_train,lr.predict(x_train),color='r')
```

Out[29]:

```
[<matplotlib.lines.Line2D at 0x2811e07ad30>]
```



In [31]:

```
import joblib
joblib.dump(lr,'stu_mar_pre.pk1')
model=joblib.load('stu_mar_pre.pk1')
model.predict([[1]])[0][0]
```

Out[31]:

```
7.52122090838123
```