In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df=pd.read_csv('IceCreamData.csv')
df
```

Out[2]:

	Temperature	Revenue
0	24.566884	534.799028
1	26.005191	625.190122
2	27.790554	660.632289
3	20.595335	487.706960
4	11.503498	316.240194
...	...	...
495	22.274899	524.746364
496	32.893092	755.818399
497	12.588157	306.090719
498	22.362402	566.217304
499	28.957736	655.660388

500 rows × 2 columns

In [3]:
```
df.corr()
```

Out[3]:

	Temperature	Revenue
Temperature	1.000000	0.989802
Revenue	0.989802	1.000000

In [4]:
```
df.info
```

Out[4]:
```
<bound method DataFrame.info of      Temperature     Revenue
0      24.566884  534.799028
1      26.005191  625.190122
2      27.790554  660.632289
3      20.595335  487.706960
4      11.503498  316.240194
..           ...         ...
495    22.274899  524.746364
496    32.893092  755.818399
497    12.588157  306.090719
498    22.362402  566.217304
499    28.957736  655.660388

[500 rows x 2 columns]>
```

In [5]:
```
df.isnull().sum()
```

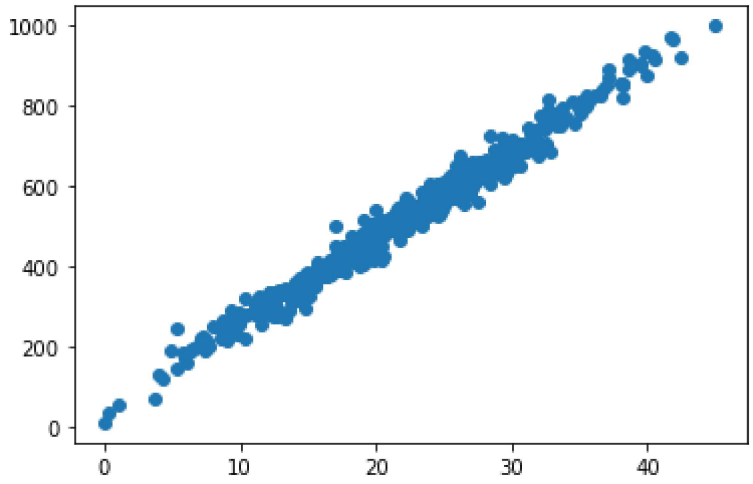
Out[5]:
```
Temperature    0
Revenue        0
dtype: int64
```

In [6]:
```
x=df['Temperature'].values.reshape(-1,1)
y=df['Revenue'].values.reshape(-1,1)
```

In [7]:
```
from matplotlib import pyplot as plt
```

In [8]:
```
plt.scatter(x,y)
```

Out[8]:
```
<matplotlib.collections.PathCollection at 0x20f24da1850>
```



In [9]:
```
df.shape
```

Out[9]:
```
(500, 2)
```

In [10]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,random_state=0,test_size=
```

In [11]:
```
x_train.shape
```

Out[11]:
```
(450, 1)
```

In [12]:
```
x_test.shape
```

Out[12]:
```
(50, 1)
```

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

In [14]:
```
lr = LinearRegression()
```

In [15]:
```
lr.fit(x_train,y_train)
```

Out[15]:
```
LinearRegression()
```

In [16]:
```
y_pred = lr.predict(x_test)
```

In [17]:
```
prediction = pd.DataFrame(y_test,columns=['Y'])
prediction['y_hat']=y_pred
prediction['residuals'] = y_test - y_pred
```

In [18]:
```
lr.predict(x_test)
```

Out[18]:
```
array([[697.70707182],
       [652.73904132],
       [664.13404026],
       [450.14772339],
       [664.8776824 ],
       [441.00665054],
       [583.55377641],
       [623.27199574],
       [666.88804887],
       [468.33368301],
       [546.35475862],
       [443.04781122],
       [622.39921265],
       [377.35127062],
       [366.77670698],
       [944.77968398],
       [892.95903261],
       [693.82704082],
       [545.57871799],
       [420.24507165],
       [390.77577899],
       [596.48894725],
       [283.03972028],
       [654.91399637],
       [380.68932769],
       [411.98660707],
       [370.76234541],
       [509.80490521],
       [479.3005344 ],
       [456.30404169],
       [639.54533326],
       [281.45779729],
       [313.96089451],
       [469.62163019],
       [559.23843057],
       [539.2857608 ],
       [307.50189059],
       [508.21927054],
       [570.93399312],
       [731.58893366],
       [440.07912022],
       [493.97664874],
       [567.07104388],
       [443.57715242],
       [913.60815774],
       [602.66172688],
       [541.36582095],
       [199.84105078],
       [693.4156069 ],
       [350.83232268]])
```

In [19]:
```
m=lr.coef_ #slope (m-value)
m
```

Out[19]:
```
array([[21.49082669]])
```

In [20]:
```
b=lr.intercept_ #y_intercept
b
```

Out[20]:
```
array([43.78867085])
```

In [21]:
```
x_test
```

Out[21]:
```
array([[30.42779184],
       [28.33536277],
       [28.86558895],
       [18.90848865],
       [28.90019172],
       [18.48314099],
       [25.11606991],
       [26.96421749],
       [28.99373705],
       [19.75470829],
       [23.38514451],
       [18.57811922],
       [26.9236056 ],
       [15.52116187],
       [15.02911176],
       [41.92444647],
       [39.5131548 ],
       [30.24724825],
       [23.34903419],
       [17.51707397],
       [16.14582413],
       [25.71796257],
       [11.13270573],
       [28.43656665],
       [15.67648661],
       [17.13279538],
       [15.21456942],
       [21.68442569],
       [20.26501213],
       [19.19495126],
       [27.72143999],
       [11.05909651],
       [12.57151377],
       [19.81463838],
       [23.98464085],
       [23.05621357],
       [12.27096675],
       [21.61064376],
       [24.5288527 ],
       [32.00436506],
       [18.43998163],
       [20.94791347],
       [24.34910395],
       [18.60275025],
       [40.47398918],
       [26.00519115],
       [23.15300185],
       [ 7.2613484 ],
       [30.22810362],
       [14.28719594]])
```

In [22]:
```
y_ans= m * x_test + b
```

In [23]:
```
y_ans
```

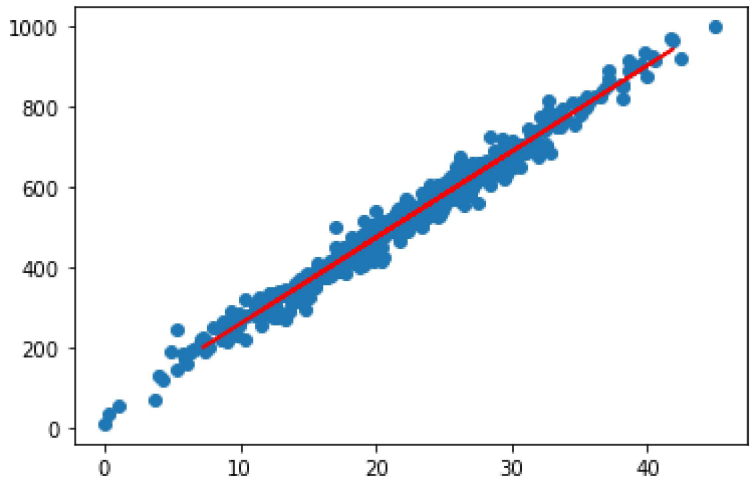
Out[23]:
```
array([[697.70707182],
       [652.73904132],
       [664.13404026],
       [450.14772339],
       [664.8776824 ],
       [441.00665054],
       [583.55377641],
       [623.27199574],
       [666.88804887],
       [468.33368301],
       [546.35475862],
       [443.04781122],
       [622.39921265],
       [377.35127062],
       [366.77670698],
       [944.77968398],
       [892.95903261],
       [693.82704082],
       [545.57871799],
       [420.24507165],
       [390.77577899],
       [596.48894725],
       [283.03972028],
       [654.91399637],
       [380.68932769],
       [411.98660707],
       [370.76234541],
       [509.80490521],
       [479.3005344 ],
       [456.30404169],
       [639.54533326],
       [281.45779729],
       [313.96089451],
       [469.62163019],
       [559.23843057],
       [539.2857608 ],
       [307.50189059],
       [508.21927054],
       [570.93399312],
       [731.58893366],
       [440.07912022],
       [493.97664874],
       [567.07104388],
       [443.57715242],
       [913.60815774],
       [602.66172688],
       [541.36582095],
       [199.84105078],
       [693.4156069 ],
       [350.83232268]])
```

In [24]:
```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test,y_pred)
mse
```

Out[24]:
```
510.36278285590174
```

In [25]:
```
plt.scatter(x,y)
plt.plot(x_test,y_pred , color = 'red')
```

Out[25]:
```
[<matplotlib.lines.Line2D at 0x20f27931c70>]
```



In [ ]: