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

In [2]:
```
df= pd.read_csv('weight-height.csv')
```

In [5]:
```
df
```

Out[5]:

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801
...	...	...	...
9995	Female	66.172652	136.777454
9996	Female	67.067155	170.867906
9997	Female	63.867992	128.475319
9998	Female	69.034243	163.852461
9999	Female	61.944246	113.649103

10000 rows × 3 columns

In [7]:
```
df.shape
```

Out[7]:
```
(10000, 3)
```

In [9]:
```
df.isna().sum()
```

Out[9]:
```
Gender    0
Height    0
Weight    0
dtype: int64
```

In [11]:
```
df.duplicated().sum()
```

Out[11]:
```
0
```

In [13]:
```
df.dtypes
```

Out[13]:
```
Gender     object
Height    float64
Weight    float64
dtype: object
```

In [15]:
```
df=df.drop("Gender",axis="columns")
```

In [17]:
```
df
```

Out[17]:

	Height	Weight
0	73.847017	241.893563
1	68.781904	162.310473
2	74.110105	212.740856
3	71.730978	220.042470
4	69.881796	206.349801
...	...	...
9995	66.172652	136.777454
9996	67.067155	170.867906
9997	63.867992	128.475319
9998	69.034243	163.852461
9999	61.944246	113.649103

10000 rows × 2 columns

In [19]:
```
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

In [20]:
```
X = df.iloc[:,0:1]
X
```

Out[20]:

	Height
0	73.847017
1	68.781904
2	74.110105
3	71.730978
4	69.881796
...	...
9995	66.172652
9996	67.067155
9997	63.867992
9998	69.034243
9999	61.944246

10000 rows × 1 columns

In [23]:
```
Y = df.iloc[:,1:2]
Y
```

Out[23]:

	Weight
0	241.893563
1	162.310473
2	212.740856
3	220.042470
4	206.349801
...	...
9995	136.777454
9996	170.867906
9997	128.475319
9998	163.852461
9999	113.649103

10000 rows × 1 columns

In [25]:
```
X_train, X_test, Y_train, Y_test = train_test_split(X,Y,test_size = 0.2)
```

In [27]:
```
Reg = LinearRegression()
```

In [29]:
```
Reg.fit(X_train,Y_train)
```

Out[29]:
```
LinearRegression()
```

In [31]:
```
Y_predict = Reg.predict(X_test)
```

In [33]:
```
print(Reg.coef_)
```

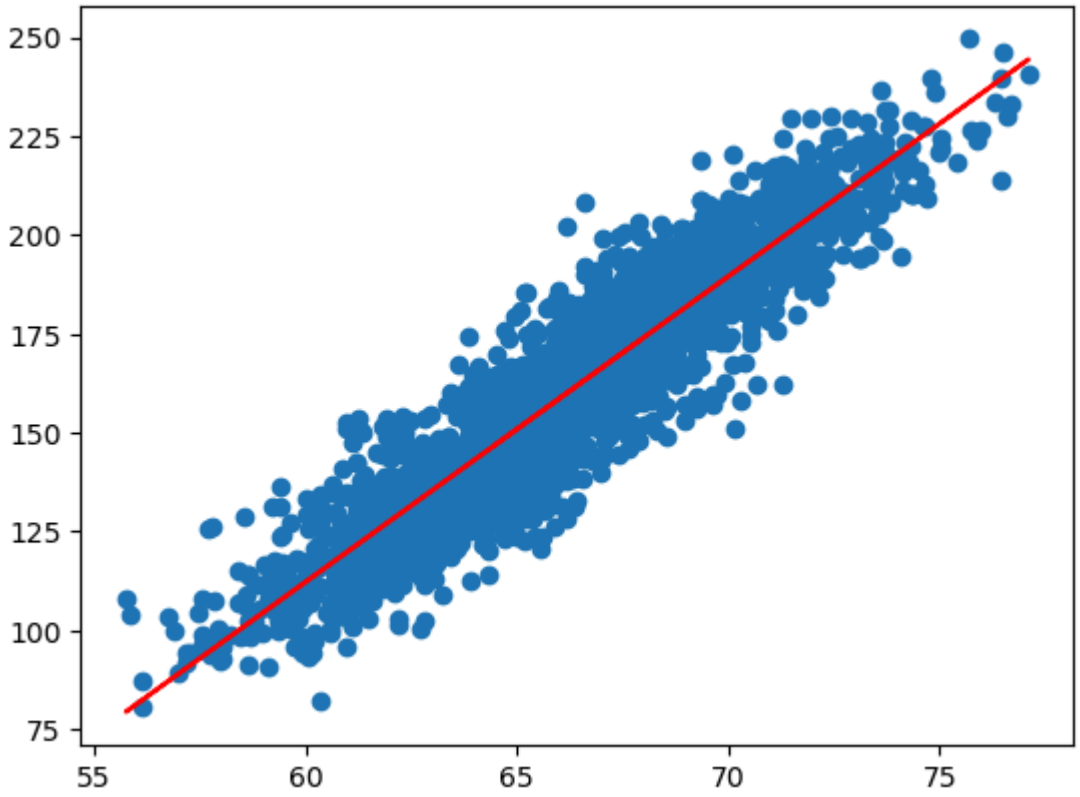
```
[[7.72633084]]
```

In [35]:
```
print(Reg.intercept_)
```

```
[-351.28031984]
```

In [37]:
```
plt.scatter(X_test,Y_test)
plt.plot(X_test,Y_predict,color = "red")
```

Out[37]:
```
[<matplotlib.lines.Line2D at 0x1c44ed29b90>]
```



In [39]:
```
print('meansqaureerror',metrics.mean_squared_error(Y_test,Y_predict))
```

```
meansqaureerror 144.0251904853563
```

In [41]:
```
print("meanabsoluteerror",metrics.mean_absolute_error(Y_test,Y_predict))
```

```
meanabsoluteerror 9.484232633357667
```

In [43]:
```
Rsquare = Reg.score(X_train,Y_train)
```

In [45]:
```
print(Rsquare)
```

```
0.8546781617760344
```

In [ ]:
```

```