

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

```
In [2]: df=pd.read_csv(r"C:\Users\Sushma sree\Downloads\Heart Disease.csv")
df
```

Out[2]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
0	1	39	4.0	0	0.0	0.0	0	0
1	0	46	2.0	0	0.0	0.0	0	0
2	1	48	1.0	1	20.0	0.0	0	0
3	0	61	3.0	1	30.0	0.0	0	1
4	0	46	3.0	1	23.0	0.0	0	0
...
4233	1	50	1.0	1	1.0	0.0	0	1
4234	1	51	3.0	1	43.0	0.0	0	0
4235	0	48	2.0	1	20.0	NaN	0	0
4236	0	44	1.0	1	15.0	0.0	0	0
4237	0	52	2.0	0	0.0	0.0	0	0

4238 rows × 16 columns



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

Out[3]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	d
0	1	39	4.0	0	0.0	0.0	0	0	
1	0	46	2.0	0	0.0	0.0	0	0	
2	1	48	1.0	1	20.0	0.0	0	0	
3	0	61	3.0	1	30.0	0.0	0	1	
4	0	46	3.0	1	23.0	0.0	0	0	



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

```
Out[4]: <bound method NDFrame.describe of
cigsPerDay  BPMeds
0          1    39      4.0      0      0.0      0.0  \
1          0    46      2.0      0      0.0      0.0
2          1    48      1.0      1     20.0      0.0
3          0    61      3.0      1     30.0      0.0
4          0    46      3.0      1     23.0      0.0
...      ...    ...      ...      ...      ...      ...
4233       1    50      1.0      1      1.0      0.0
4234       1    51      3.0      1     43.0      0.0
4235       0    48      2.0      1     20.0      NaN
4236       0    44      1.0      1     15.0      0.0
4237       0    52      2.0      0      0.0      0.0

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP  BMI
0                   0              0         0    195.0  106.0   70.0  26.97
\
1                   0              0         0    250.0  121.0   81.0  28.73
2                   0              0         0    245.0  127.5   80.0  25.34
3                   0              1         0    225.0  150.0   95.0  28.58
4                   0              0         0    285.0  130.0   84.0  23.10
...      ...      ...      ...      ...      ...      ...      ...
4233              0              1         0    313.0  179.0   92.0  25.97
4234              0              0         0    207.0  126.5   80.0  19.71
4235              0              0         0    248.0  131.0   72.0  22.00
4236              0              0         0    210.0  126.5   87.0  19.16
4237              0              0         0    269.0  133.5   83.0  21.47

      heartRate  glucose  TenYearCHD
0         80.0     77.0           0
1         95.0     76.0           0
2         75.0     70.0           0
3         65.0    103.0           1
4         85.0     85.0           0
...      ...      ...      ...
4233        66.0     86.0           1
4234        65.0     68.0           0
4235        84.0     86.0           0
4236        86.0      NaN           0
4237        80.0    107.0           0

[4238 rows x 16 columns]>
```

```
In [5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4238 entries, 0 to 4237
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   male                  4238 non-null   int64
1   age                   4238 non-null   int64
2   education             4133 non-null   float64
3   currentSmoker         4238 non-null   int64
4   cigsPerDay            4209 non-null   float64
5   BPMeds                4185 non-null   float64
6   prevalentStroke       4238 non-null   int64
7   prevalentHyp          4238 non-null   int64
8   diabetes              4238 non-null   int64
9   totChol               4188 non-null   float64
10  sysBP                 4238 non-null   float64
11  diaBP                 4238 non-null   float64
12  BMI                   4219 non-null   float64
13  heartRate             4237 non-null   float64
14  glucose               3850 non-null   float64
15  TenYearCHD            4238 non-null   int64
dtypes: float64(9), int64(7)
memory usage: 529.9 KB
```

```
In [6]: df.isna().any()
```

```
Out[6]: male                False
age                False
education          True
currentSmoker      False
cigsPerDay         True
BPMeds             True
prevalentStroke    False
prevalentHyp       False
diabetes           False
totChol            True
sysBP              False
diaBP              False
BMI                True
heartRate          True
glucose            True
TenYearCHD         False
dtype: bool
```

```
In [7]: df.dropna(inplace=True)
```

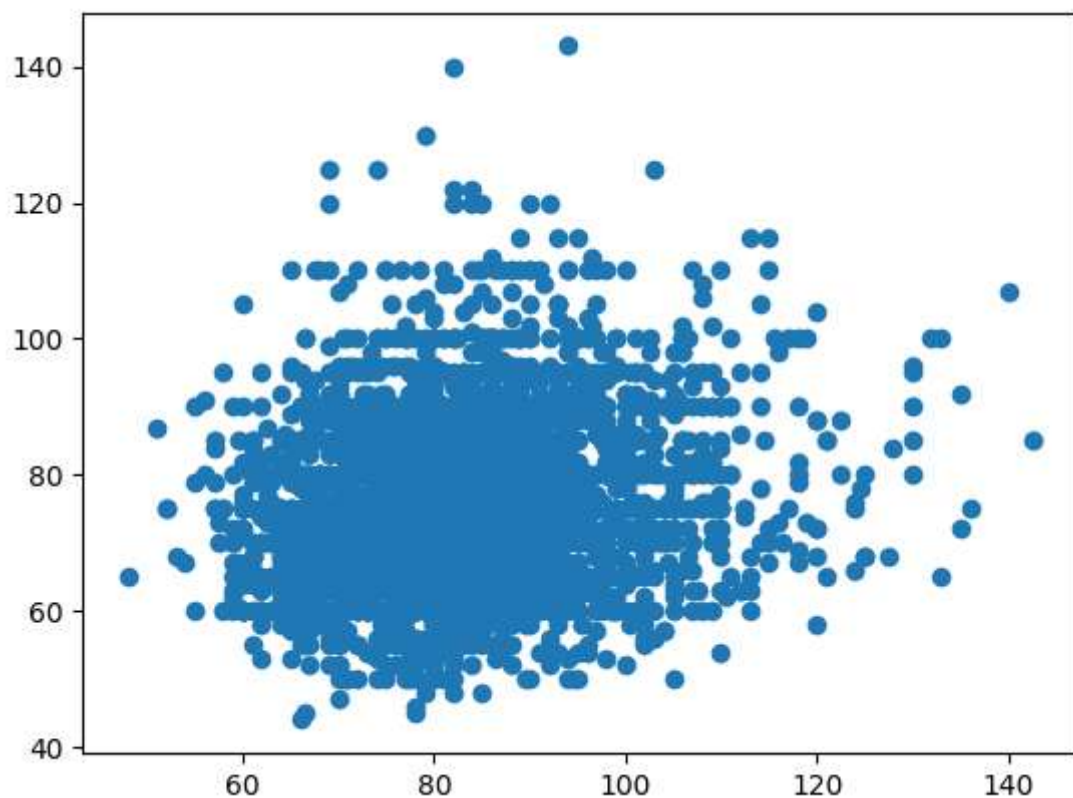
```
In [8]: df.isna().any()
```

```
Out[8]: male           False
age           False
education     False
currentSmoker False
cigsPerDay    False
BPMeds        False
prevalentStroke False
prevalentHyp  False
diabetes       False
totChol        False
sysBP          False
diaBP          False
BMI            False
heartRate      False
glucose        False
TenYearCHD     False
dtype: bool
```

```
In [9]: from matplotlib import pyplot as plt
```

```
In [10]: plt.scatter(df['diaBP'],df['heartRate'])
```

```
Out[10]: <matplotlib.collections.PathCollection at 0x21aef162ad0>
```



```
In [11]: x=df[['diaBP']]
y=df['heartRate']
x.head()
```

```
Out[11]:
```

	diaBP
0	70.0
1	81.0
2	80.0
3	95.0
4	84.0

```
In [12]: from sklearn.model_selection import train_test_split
```

```
In [13]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
```

```
In [14]: from sklearn.linear_model import LinearRegression
lr=LinearRegression()
```

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

```
Out[15]:
```

▼ LinearRegression

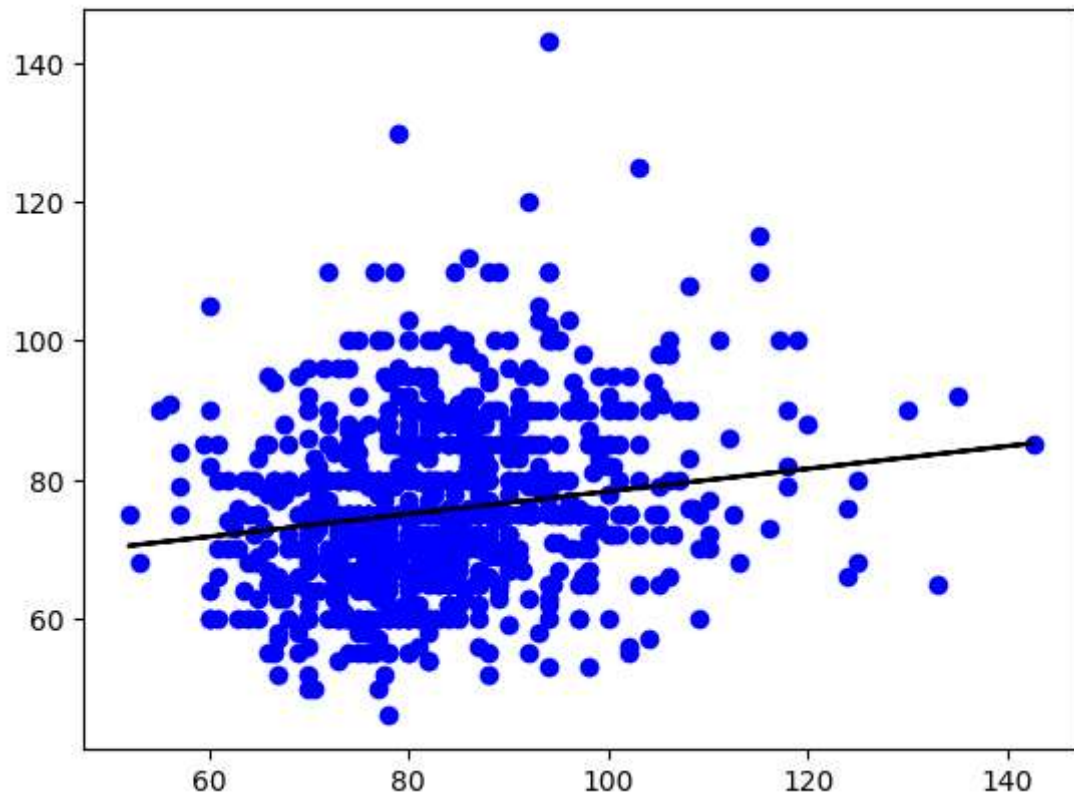
LinearRegression()

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

```
In [17]: lr.score(x_test,y_test)
```

```
Out[17]: 0.04117527794230713
```

```
In [18]: y_pred=lr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```



```
In [19]: df1=df[:2000]
df1
```

Out[19]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
0	1	39	4.0	0	0.0	0.0	0	0
1	0	46	2.0	0	0.0	0.0	0	0
2	1	48	1.0	1	20.0	0.0	0	0
3	0	61	3.0	1	30.0	0.0	0	1
4	0	46	3.0	1	23.0	0.0	0	0
...
2328	1	66	4.0	1	20.0	0.0	0	1
2329	1	40	2.0	1	20.0	0.0	0	0
2330	0	42	1.0	0	0.0	0.0	0	0
2331	0	44	1.0	1	10.0	0.0	0	1
2332	1	46	2.0	1	20.0	0.0	0	1

2000 rows × 16 columns



```
In [20]: df1.head()
```

```
Out[20]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	d
0	1	39	4.0	0	0.0	0.0	0	0	
1	0	46	2.0	0	0.0	0.0	0	0	
2	1	48	1.0	1	20.0	0.0	0	0	
3	0	61	3.0	1	30.0	0.0	0	1	
4	0	46	3.0	1	23.0	0.0	0	0	

```
In [21]: df1.tail()
```

```
Out[21]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	d
2328	1	66	4.0	1	20.0	0.0	0	1	
2329	1	40	2.0	1	20.0	0.0	0	0	
2330	0	42	1.0	0	0.0	0.0	0	0	
2331	0	44	1.0	1	10.0	0.0	0	1	
2332	1	46	2.0	1	20.0	0.0	0	1	

```
In [22]: df1.isnull().sum()
```

```
Out[22]: male          0
age          0
education    0
currentSmoker 0
cigsPerDay   0
BPMeds       0
prevalentStroke 0
prevalentHyp  0
diabetes     0
totChol      0
sysBP        0
diaBP        0
BMI          0
heartRate    0
glucose      0
TenYearCHD   0
dtype: int64
```

```
In [23]: x=df1[['diaBP']]
y=df1['heartRate']
```

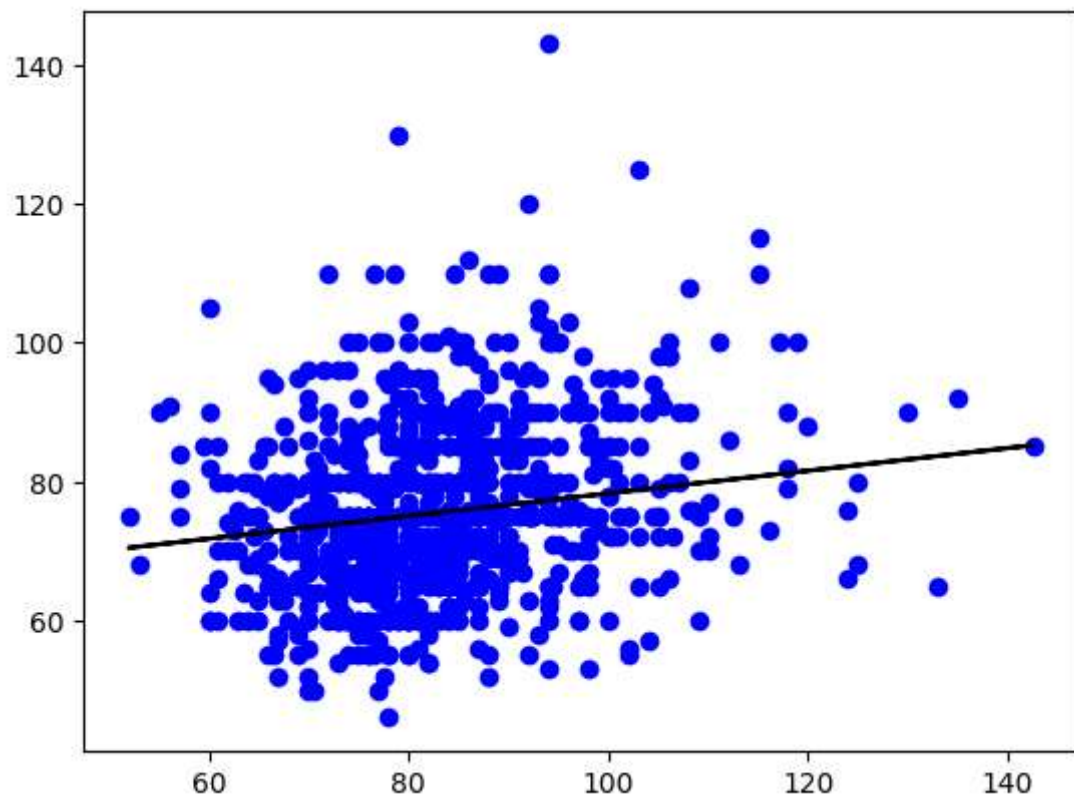
```
In [24]: lr=LinearRegression()  
lr.fit(x_train,y_train)
```

```
Out[24]: ▾ LinearRegression  
LinearRegression()
```

```
In [25]: lr.score(x_test,y_test)
```

```
Out[25]: 0.04117527794230713
```

```
In [26]: y_pred=lr.predict(x_test)  
plt.scatter(x_test,y_test,color='b')  
plt.plot(x_test,y_pred,color='k')  
plt.show()
```



```
In [ ]:
```