

In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
data=pd.read_csv(r"C:\Users\Prathyusha\Downloads\Heart Disease.csv")
data
```

Out[2]:

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

4238 rows × 16 columns

In [3]:

```
data.head()
```

Out[3]:

	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

In [4]:

data.tail()

Out[4]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalent
4233	1	50	1.0	1	1.0	0.0	0	
4234	1	51	3.0	1	43.0	0.0	0	
4235	0	48	2.0	1	20.0	NaN	0	
4236	0	44	1.0	1	15.0	0.0	0	
4237	0	52	2.0	0	0.0	0.0	0	

In [5]:

data.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]:

```
data.isna().sum()
```

Out[6]:

```
male          0
age           0
education     105
currentSmoker 0
cigsPerDay    29
BPMeds        53
prevalentStroke 0
prevalentHyp  0
diabetes      0
totChol       50
sysBP         0
diaBP         0
BMI           19
heartRate     1
glucose       388
TenYearCHD    0
dtype: int64
```

In [7]:

```
data.dropna(inplace=True)
```

In [8]:

```
data.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]:

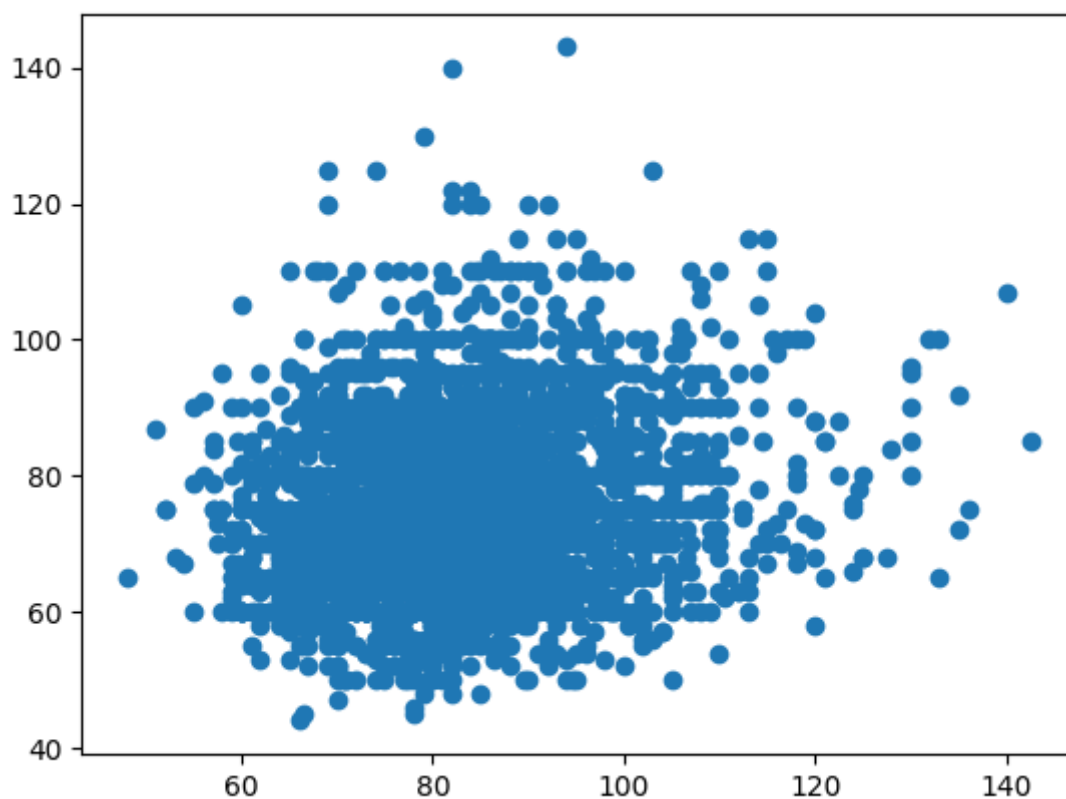
```
from matplotlib import pyplot as plt
```

In [10]:

```
plt.scatter(data['diaBP'],data['heartRate'])
```

Out[10]:

<matplotlib.collections.PathCollection at 0x24cacbcf160>



In [11]:

```
x=data[['diaBP']]  
y=data['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]:

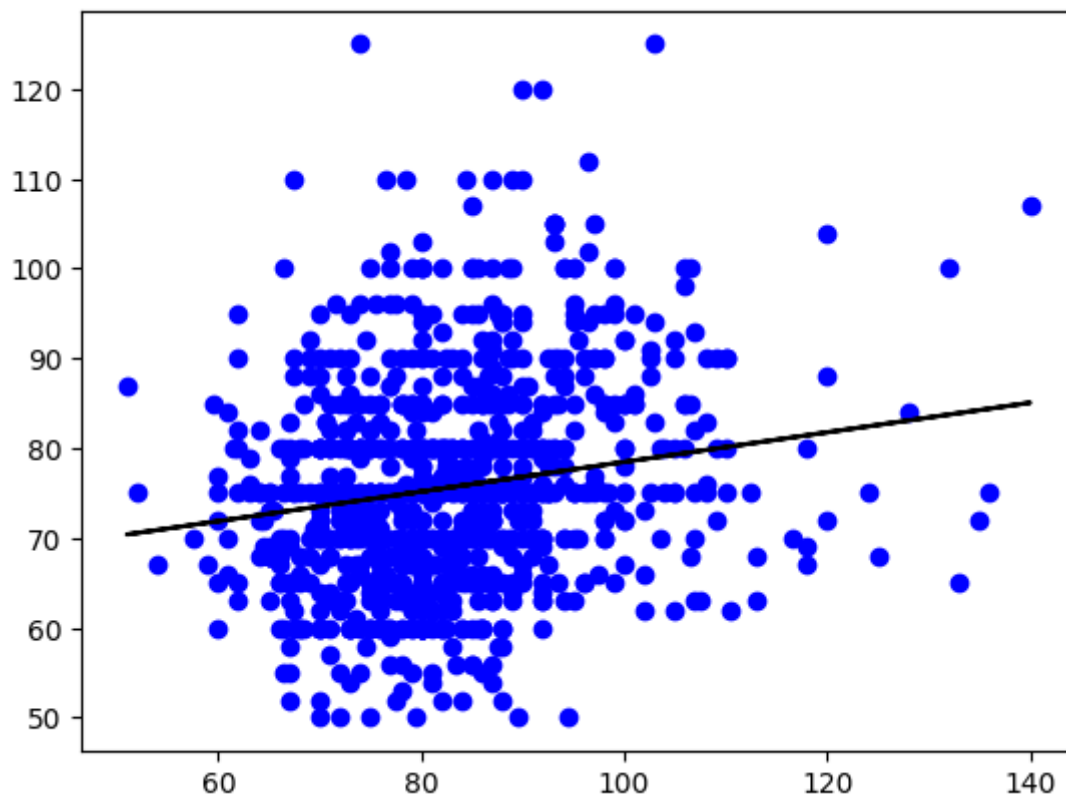
```
y_pred=lr.predict(x_test)  
lr.score(x_test,y_test)
```

Out[16]:

```
0.04663127392001265
```

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]:

```
data1=data[:2000]
data1
```

Out[19]:

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

2000 rows × 16 columns



In [20]:

```
data1.head()
```

Out[20]:

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



In [21]:

```
data.isna().any()
```

Out[21]:

```
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 [22]:

```
x=data1[['diaBP']]
y=data1['heartRate']
x.head()
```

Out[22]:

	diaBP
0	70.0
1	81.0
2	80.0
3	95.0
4	84.0

In [23]:

```
lr.fit(x_train,y_train)
```

Out[23]:

```
LinearRegression
LinearRegression()
```



In [24]:

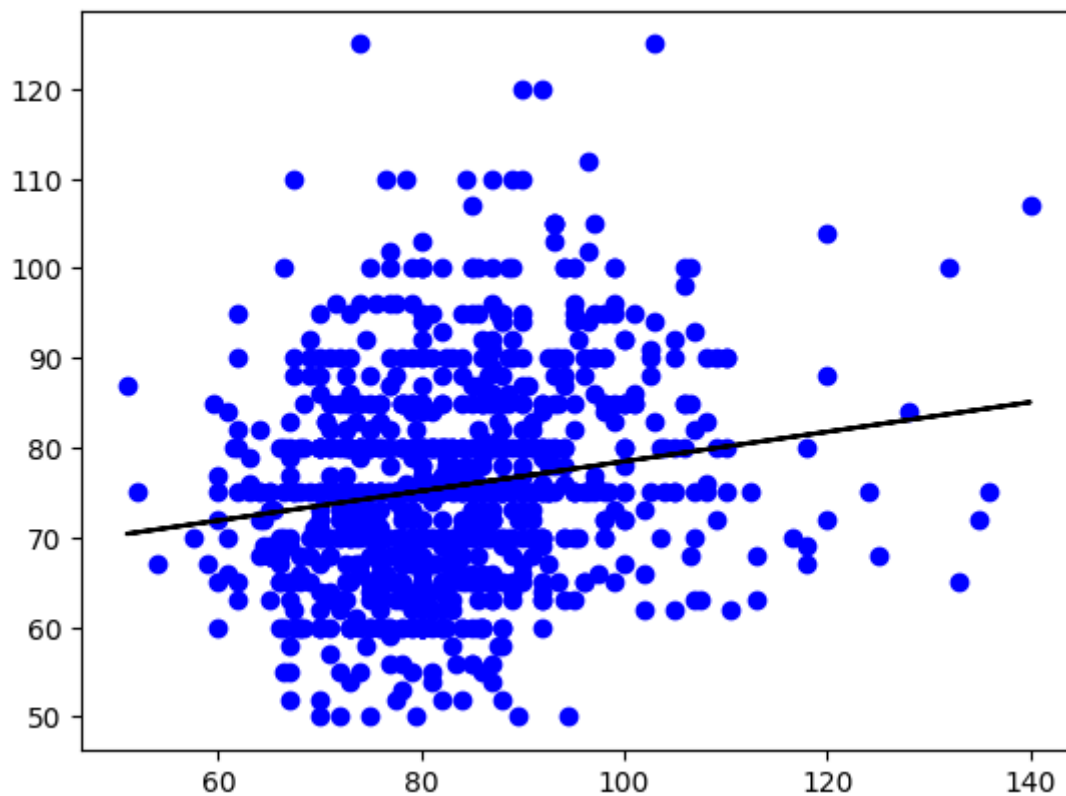
```
y_pred=lr.predict(x_test)  
lr.score(x_test,y_test)
```

Out[24]:

0.04663127392001265

In [25]:

```
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 [ ]: