

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sb
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import warnings
warnings.simplefilter(action='ignore')
```

```
In [2]: df=pd.read_csv(r"C:\Users\pappu\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

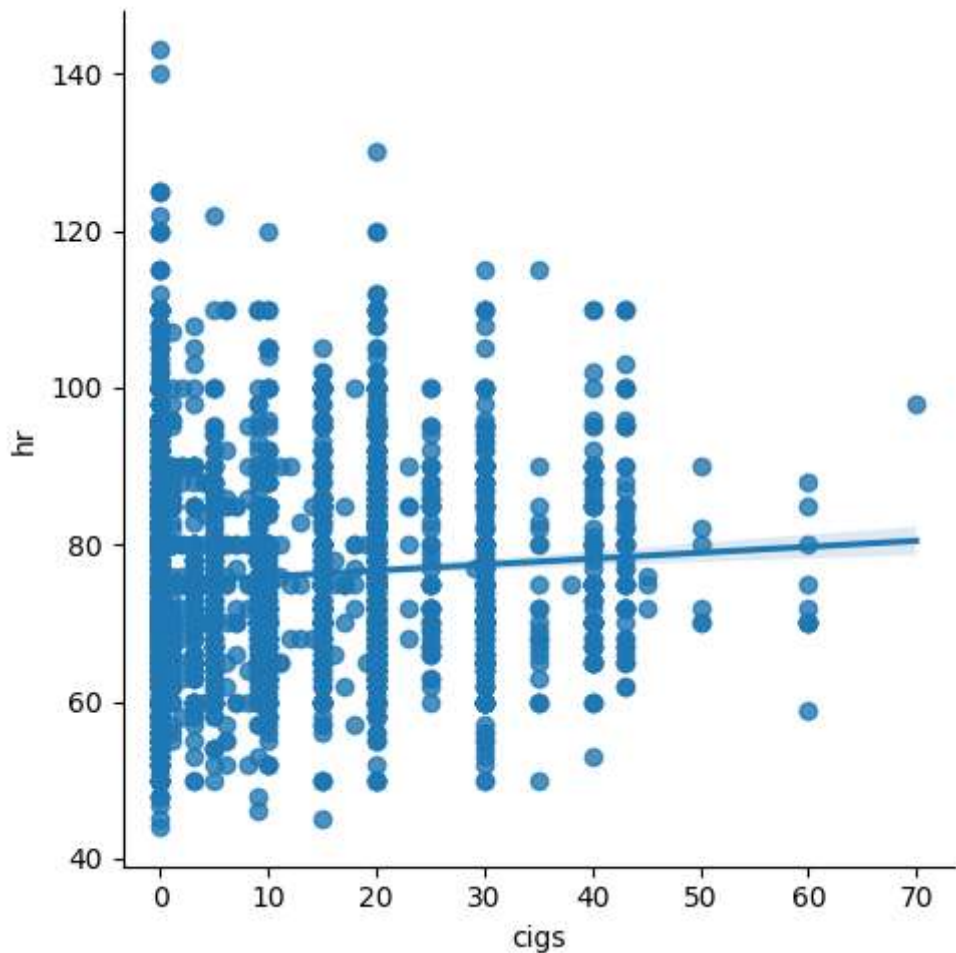


```
In [3]: df=df[['cigsPerDay', 'heartRate']]
```

```
In [4]: df.columns=['cigs', 'hr']
```

```
In [5]: sb.lmplot(x='cigs',y='hr',data=df)
```

```
Out[5]: <seaborn.axisgrid.FacetGrid at 0x1f82f1e4370>
```



```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 4238 entries, 0 to 4237  
Data columns (total 2 columns):  
#   Column  Non-Null Count  Dtype  
---  -  
0   cigs     4209 non-null     float64  
1   hr       4237 non-null     float64  
dtypes: float64(2)  
memory usage: 66.3 KB
```

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

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

```
Out[8]: cigs    False  
        hr      False  
        dtype: bool
```

```
In [9]: x=df[['cigs']]
        y=df['hr']
        x.head()
```

Out[9]:

	cigs
0	0.0
1	0.0
2	20.0
3	30.0
4	23.0

```
In [10]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
         x_train
```

Out[10]:

	cigs
2743	20.0
2933	30.0
3602	0.0
1449	30.0
1992	0.0
...	...
2627	0.0
4190	0.0
2377	0.0
1990	0.0
335	30.0

2945 rows × 1 columns

```
In [11]: from sklearn.linear_model import LinearRegression
         lr=LinearRegression()
```

```
In [12]: lr.fit(x_train,y_train)
```

Out[12]:

LinearRegression

LinearRegression()

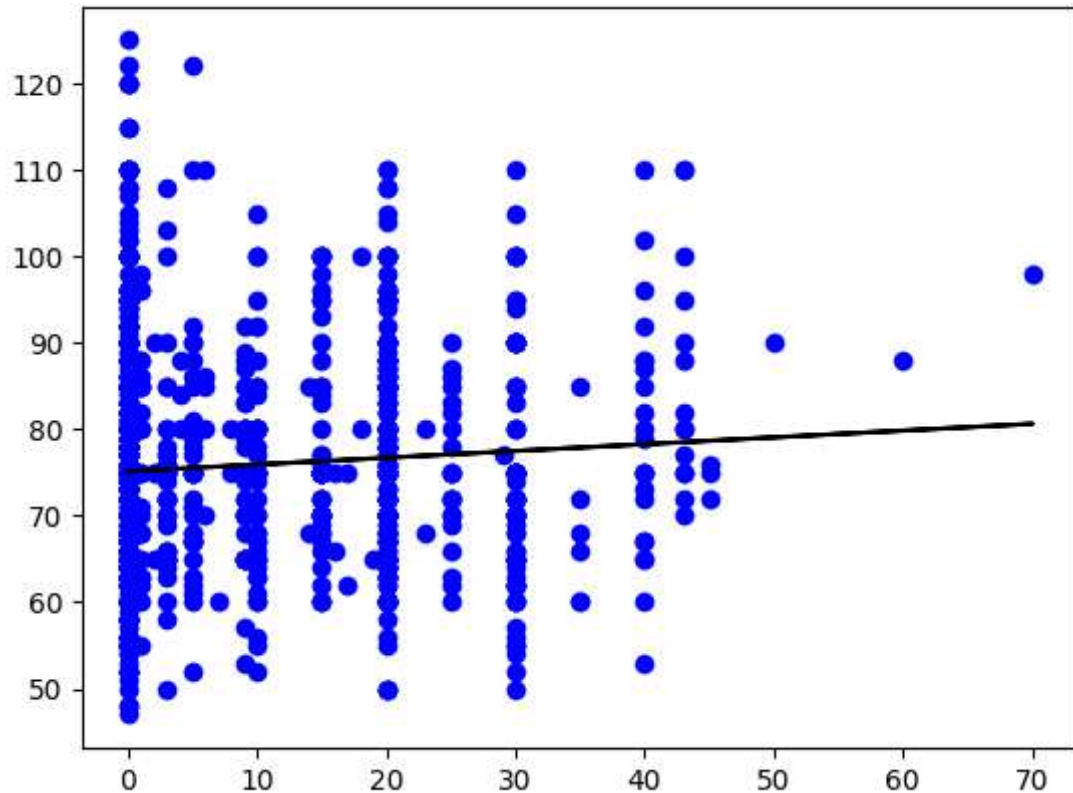
```
In [13]: lr.predict(x_test)
```

Out[13]: array([75.13560177, 75.13560177, 75.21406464, ..., 75.13560177,  
76.3125448 , 75.13560177])

```
In [14]: print(lr.score(x_test,y_test))
```

0.004274344513437667

```
In [15]: 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 [16]: df100=df[:100]
```

```
In [17]: df100.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 100 entries, 0 to 99
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  -
 0   cigs    100 non-null    float64
 1   hr      100 non-null    float64
dtypes: float64(2)
memory usage: 2.3 KB
```

```
In [18]: df100.dropna(inplace=True)
```

```
In [19]: df100.isna().any()
```

```
Out[19]: cigs    False  
         hr      False  
         dtype: bool
```

```
In [20]: x=df100[['cigs']]  
         y=df100['hr']  
         x.head()
```

```
Out[20]:
```

	cigs
0	0.0
1	0.0
2	20.0
3	30.0
4	23.0

```
In [21]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4)
```

```
In [22]: from sklearn.linear_model import LinearRegression  
         lr=LinearRegression()
```

```
In [23]: lr.fit(x_train,y_train)
```

```
Out[23]:
```

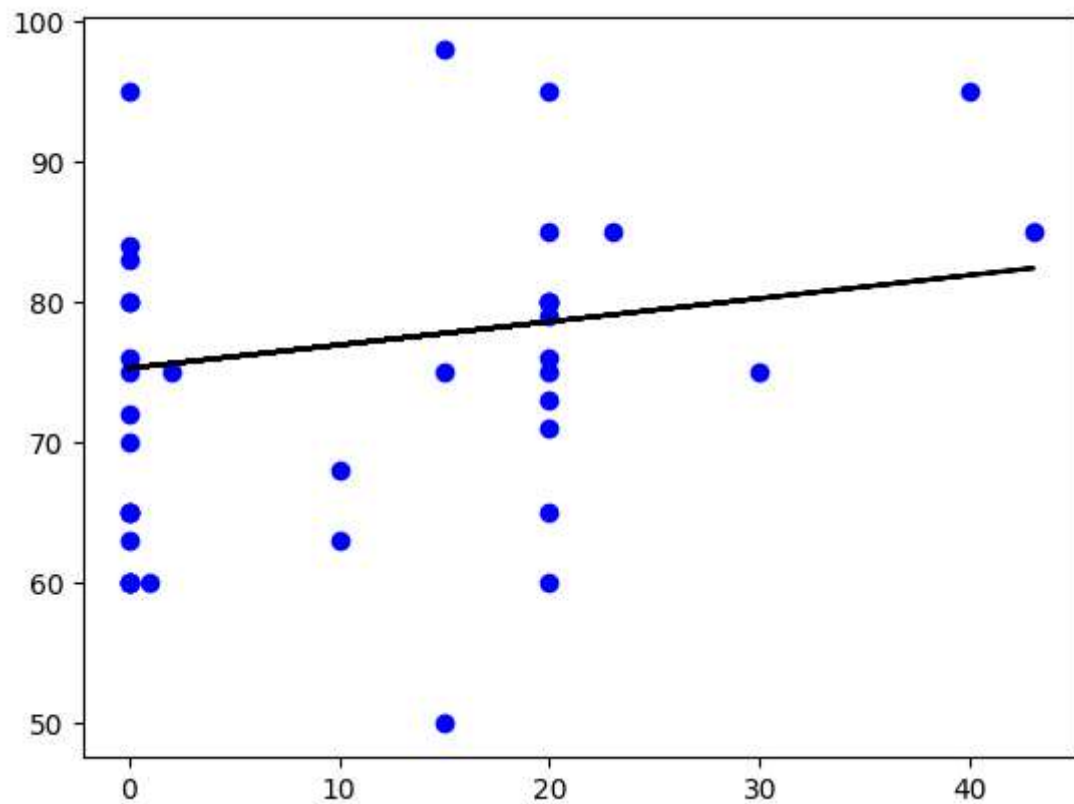
▼ LinearRegression

LinearRegression()

```
In [24]: print(lr.score(x_test,y_test))
```

```
0.024701017388460644
```

```
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 [ ]:

In [ ]: