

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import random as rd
```

```
df=pd.read_csv("/content/heart (1).csv")
```

```
df
```

```
ds=pd.read_csv("/content/IndiaAirQualityData.csv",encoding="ISO-8859-1")
ds
```

```
ds.head()
```

	stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	spm
0	150.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	4.8	17.4	NaN	NaN
1	151.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	3.1	7.0	NaN	NaN
2	152.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.2	28.5	NaN	NaN
3	150.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.3	14.7	NaN	NaN

```
df.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2

Next steps: [View recommended plots](#)

```
ds.info()
```

```
df.info()
```

```
ds.isnull().sum()
```

```
stn_code      144077
sampling_date      3
state          0
location        3
agency      149481
type          5393
so2          34646
no2          16233
rspm         40222
spm          237387
location_monitoring_station  27491
pm2_5         426428
date           7
dtype: int64
```

```
df.isnull().sum()
```

```
age          0
sex          0
```

```
cp 0
trestbps 0
chol 0
fbs 0
restecg 0
thalach 0
exang 0
oldpeak 0
slope 0
ca 0
thal 0
target 0
dtype: int64

ds.dropna()
```

stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	spm	location_
----------	---------------	-------	----------	--------	------	-----	-----	------	-----	-----------

```
df.dropna()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	th
0	52	1	0	125	212	0	1	168	0	1.0	2	2	
1	53	1	0	140	203	1	0	155	1	3.1	0	0	
2	70	1	0	145	174	0	1	125	1	2.6	0	0	
3	61	1	0	148	203	0	1	161	0	0.0	2	1	
4	62	0	0	138	294	1	1	106	0	1.9	1	3	
...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	

1025 rows × 14 columns

```
ds1=ds.loc[111:999,['state','location','so2','rspm']]
```

```
ds2=ds.iloc[[1,3,5,2,4,22,43,54,67,7,8,9,50,10,11]]
```

```
ds1
```

	state	location	so2	rspm
111	Andhra Pradesh	Hyderabad	4.9	NaN
112	Andhra Pradesh	Vishakhapatnam	NaN	NaN
113	Andhra Pradesh	Vishakhapatnam	11.2	NaN
114	Andhra Pradesh	Vishakhapatnam	4.5	NaN
115	Andhra Pradesh	Hyderabad	6.2	NaN
...
995	Andhra Pradesh	Hyderabad	2.8	NaN
996	Andhra Pradesh	Hyderabad	5.0	NaN
997	Andhra Pradesh	Hyderabad	5.5	NaN
998	Andhra Pradesh	Hyderabad	5.8	NaN
999	Andhra Pradesh	Hyderabad	5.9	NaN

889 rows × 4 columns

Next steps: [View recommended plots](#)

ds2

	stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	
1	151.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	3.1	7.0	NaN	↑
3	150.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.3	14.7	NaN	↑
5	152.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.4	25.7	NaN	↑
2	152.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.2	28.5	NaN	↑
4	151.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.7	7.5	NaN	↑
22	152.0	September - M091990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	8.1	17.8	NaN	16
43	152.0	May - M051991	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	12.3	38.6	NaN	21
54	151.0	September - M091991	Andhra Pradesh	Hyderabad	NaN	Industrial Area	13.3	11.9	NaN	5
67	203.0	January - M011992	Andhra Pradesh	Hyderabad	Andhra Pradesh Pollution Control Board	NaN	35.8	12.5	NaN	26
7	151.0	April - M041990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.7	8.7	NaN	↑
8	152.0	April - M041990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	4.2	23.0	NaN	↑
9	151.0	May - M051990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.0	8.9	NaN	↑
50	150.0	August - M081991	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	8.5	12.5	NaN	11
10	152.0	May - M051990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	3.6	18.6	NaN	↑
11	150.0	June - M061990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and	3.9	14.1	NaN	13

```
ds_integration=pd.concat([ds1,ds2])
```

```
ds_integration
```

	state	location	so2	rspm	stn_code	sampling_date	agency	type	nc
111	Andhra Pradesh	Hyderabad	4.9	NaN	NaN	NaN	NaN	NaN	Na
112	Andhra Pradesh	Vishakhapatnam	NaN	NaN	NaN	NaN	NaN	NaN	Na
113	Andhra Pradesh	Vishakhapatnam	11.2	NaN	NaN	NaN	NaN	NaN	Na
114	Andhra Pradesh	Vishakhapatnam	4.5	NaN	NaN	NaN	NaN	NaN	Na
115	Andhra Pradesh	Hyderabad	6.2	NaN	NaN	NaN	NaN	NaN	Na
...
8	Andhra Pradesh	Hyderabad	4.2	NaN	152.0	April - M041990	NaN	Residential, Rural and other Areas	23
9	Andhra Pradesh	Hyderabad	4.0	NaN	151.0	May - M051990	NaN	Industrial Area	8
50	Andhra Pradesh	Hyderabad	8.5	NaN	150.0	August - M081991	NaN	Residential, Rural and other Areas	12
10	Andhra Pradesh	Hyderabad	3.6	NaN	152.0	May - M051990	NaN	Residential, Rural and other Areas	18
Residential.									

```
ds_integration.transpose()
```

	111	112	113	114	
state	Andhra Pradesh	Andhra Pradesh	Andhra Pradesh	Andhra Pradesh	F
location	Hyderabad	Vishakhapatnam	Vishakhapatnam	Vishakhapatnam	Hyd
so2	4.9	NaN	11.2	4.5	
rspm	NaN	NaN	NaN	NaN	
stn_code	NaN	NaN	NaN	NaN	
sampling_date	NaN	NaN	NaN	NaN	
agency	NaN	NaN	NaN	NaN	
type	NaN	NaN	NaN	NaN	
no2	NaN	NaN	NaN	NaN	
spm	NaN	NaN	NaN	NaN	
location_monitoring_station	NaN	NaN	NaN	NaN	
pm2_5	NaN	NaN	NaN	NaN	
date	NaN	NaN	NaN	NaN	

```
13 rows x 904 columns



ds.drop(columns="so2")
```

	stn_code	sampling_date	state	location	agency	type	no2	rspr
0	150.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	17.4	NaN
1	151.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	7.0	NaN
2	152.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	28.5	NaN
3	150.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	14.7	NaN
4	151.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	7.5	NaN
...
435737	SAMP	24-12-15	West Bengal	ULUBERIA	West Bengal State Pollution Control Board	RIRUO	50.0	143.0
435738	SAMP	29-12-15	West Bengal	ULUBERIA	West Bengal State Pollution Control Board	RIRUO	46.0	171.0
435739	NaN	NaN	andaman-and-nicobar-islands	NaN	NaN	NaN	NaN	NaN

ds2.drop(1)

	stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	
3	150.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.3	14.7	NaN	1
5	152.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.4	25.7	NaN	1
2	152.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.2	28.5	NaN	1
4	151.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.7	7.5	NaN	1
22	152.0	September - M091990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	8.1	17.8	NaN	16
43	152.0	May - M051991	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	12.3	38.6	NaN	21
54	151.0	September - M091991	Andhra Pradesh	Hyderabad	NaN	Industrial Area	13.3	11.9	NaN	5
67	203.0	January - M011992	Andhra Pradesh	Hyderabad	Andhra Pradesh Pollution Control Board	NaN	35.8	12.5	NaN	26
7	151.0	April - M041990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.7	8.7	NaN	1
8	152.0	April - M041990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	4.2	23.0	NaN	1
9	151.0	May - M051990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.0	8.9	NaN	1
50	150.0	August - M081991	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	8.5	12.5	NaN	11
10	152.0	May - M051990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	3.6	18.6	NaN	1
11	150.0	June - M061990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	3.9	14.1	NaN	13

```
ds.melt()
```

	variable	value	
0	stn_code	150.0	
1	stn_code	151.0	
2	stn_code	152.0	
3	stn_code	150.0	
4	stn_code	151.0	
...	
5664641	date	2015-12-24	
5664642	date	2015-12-29	
5664643	date	NaN	
5664644	date	NaN	
5664645	date	NaN	

5664646 rows × 2 columns

```
ds_merged=pd.concat([ds,df])
```

ds_merged

	stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	
0	150.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	4.8	17.4	NaN	I
1	151.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	3.1	7.0	NaN	I
2	152.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.2	28.5	NaN	I
3	150.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.3	14.7	NaN	I
4	151.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.7	7.5	NaN	I
...
1020	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	I
1021	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	I
1022	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	I
1023	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	I
1024	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	I

436767 rows × 27 columns

df['ca'].unique()

array([2, 0, 1, 3, 4])

df.ca.value_counts()

```
0    578
1    226
2    134
3     69
4     18
Name: ca, dtype: int64
```

from sklearn import linear_model, metrics

```
x=df[['age']]
y=df[['thal']]
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
```

len(x_train)

820

len(x_test)

205

len(y_train)

820

len(y_test)

205

df.shape

```

(1025, 14)

reg=linear_model.LinearRegression()

print(x_train)

model=reg.fit(x_train,y_train)

r_sq=reg.score(x_train,y_train)

print("Determination Coefficient:",r_sq)
↳ Determination Coefficient: 0.008792008347529245

print("Intercept:",model.intercept_)

Intercept: [1.97461867]

print("slope:",model.coef_)

slope: [[0.00633286]]

y_pred=model.predict(x_test)
print("Predicted Response:",y_pred,sep="\n")

plt.scatter(x_train,y_train,color="black")
plt.plot(x_train,reg.predict(x_train),color="blue",linewidth=3)
plt.title("Age vs thal")
plt.xlabel('age')
plt.ylabel('thal')
plt.show()

```

