

- Name : Wanave Chetan Dhanaji
- Branch : TEIT
- Roll no. : 3050
- Subject : Machine Learning
- Assignment no. 1

```
1 import pandas as pd
```

```
1 heart_data = pd.read_csv("heart.csv")
```

```
1 heart_data.shape
```

```
(303, 15)
```

```
1 heart_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 15 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0   303 non-null   int64
1   Age          303 non-null   int64
2   Sex          303 non-null   int64
3   ChestPain    303 non-null   object
4   RestBP       303 non-null   int64
5   Chol         303 non-null   int64
6   Fbs          303 non-null   int64
7   RestECG      303 non-null   int64
8   MaxHR        303 non-null   int64
9   ExAng        303 non-null   int64
10  Oldpeak      303 non-null   float64
11  Slope        303 non-null   int64
12  Ca           299 non-null   float64
13  Thal         301 non-null   object
14  AHD          303 non-null   object
dtypes: float64(2), int64(10), object(3)
memory usage: 35.6+ KB
```

```
1 heart_data = heart_data.drop(columns =['Unnamed: 0'])
```

```
2 heart_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Age          303 non-null   int64
1   Sex          303 non-null   int64
2   ChestPain    303 non-null   object
```

```

3   RestBP      303 non-null   int64
4   Chol        303 non-null   int64
5   Fbs         303 non-null   int64
6   RestECG     303 non-null   int64
7   MaxHR       303 non-null   int64
8   ExAng       303 non-null   int64
9   Oldpeak     303 non-null   float64
10  Slope       303 non-null   int64
11  Ca          299 non-null   float64
12  Thal        301 non-null   object
13  AHD         303 non-null   object
dtypes: float64(2), int64(9), object(3)
memory usage: 33.3+ KB

```

```
1 # descriptive statistics
```

```
2 heart_data.describe()
```

	Age	Sex	RestBP	Chol	Fbs	RestECG	Max
<b>count</b>	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.0000
<b>mean</b>	54.438944	0.679868	131.689769	246.693069	0.148515	0.990099	149.6072
<b>std</b>	9.038662	0.467299	17.599748	51.776918	0.356198	0.994971	22.8750
<b>min</b>	29.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.0000
<b>25%</b>	48.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.5000
<b>50%</b>	56.000000	1.000000	130.000000	241.000000	0.000000	1.000000	153.0000
<b>75%</b>	61.000000	1.000000	140.000000	275.000000	0.000000	2.000000	166.0000
<b>max</b>	77.000000	1.000000	200.000000	564.000000	1.000000	2.000000	202.0000

```
1 heart_data.isnull().sum()
```

```

Age          0
Sex          0
ChestPain    0
RestBP       0
Chol         0
Fbs          0
RestECG      0
MaxHR        0
ExAng        0
Oldpeak      0
Slope        0
Ca           4
Thal         2
AHD          0
dtype: int64

```

```
1 heart_data.head()
```

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope
0	63	1	typical	145	233	1	2	150	0	2.3	3
1	67	1	asymptomatic	160	286	0	2	108	1	1.5	2
2	67	1	asymptomatic	120	229	0	2	129	1	2.6	2
3	37	1	nonanginal	130	250	0	0	187	0	3.5	3

```
1 count = heart_data['Sex'].value_counts()[0]
2 print("No of 0s in sex col: ",count)
```

No of 0s in sex col: 97

```
1 countcp = heart_data['ChestPain'].value_counts()[0]
2 print("No of 0s in sex col: ",countcp)
```

No of 0s in sex col: 144

```
1 heartcolumns = ['Age','Sex','ChestPain','RestBP','Chol','Fbs','RestECG','Tha
2 for i in heartcolumns:
3     column = heart_data[i]
4     count = (column==0).sum()
5     print("Count of 0s in column ",i," is ",count)
6
```

```
Count of 0s in column Age is 0
Count of 0s in column Sex is 97
Count of 0s in column ChestPain is 0
Count of 0s in column RestBP is 0
Count of 0s in column Chol is 0
Count of 0s in column Fbs is 258
Count of 0s in column RestECG is 151
Count of 0s in column Thal is 0
Count of 0s in column ExAng is 204
Count of 0s in column Oldpeak is 99
Count of 0s in column Slope is 0
Count of 0s in column Ca is 176
Count of 0s in column AHD is 0
```

```
1 heart_data['Age'].mean()
```

54.43894389438944

```
1 part75 = heart_data.sample(frac=0.75)
2 print(part75)
```

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	\
92	62	1	nonanginal	130	231	0	0	146	0	
238	49	0	nontypical	134	271	0	0	162	0	
186	42	1	nonanginal	120	240	1	0	194	0	
237	46	1	asymptomatic	120	249	0	2	144	0	

130	54	1	nonanginal	120	258	0	2	147	0
..	...	...	...	...	...	...	...	...	...
83	68	1	nonanginal	180	274	1	2	150	1
164	48	1	nonanginal	124	255	1	0	175	0
73	65	1	asymptomatic	110	248	0	2	158	0
168	35	1	asymptomatic	126	282	0	2	156	1
66	60	1	nonanginal	140	185	0	2	155	0

	Oldpeak	Slope	Ca	Thal	AHD
92	1.8	2	3.0	reversable	No
238	0.0	2	0.0	normal	No
186	0.8	3	0.0	reversable	No
237	0.8	1	0.0	reversable	Yes
130	0.4	2	0.0	reversable	No
..	...	...	...	...	...
83	1.6	2	0.0	reversable	Yes
164	0.0	1	2.0	normal	No
73	0.6	1	2.0	fixed	Yes
168	0.0	1	0.0	reversable	Yes
66	3.0	2	0.0	normal	Yes

[227 rows x 14 columns]

```
1 part50 = heart_data.sample(frac = 0.5)
2 print(part50)
```

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	\
143	64	1	nonanginal	125	309	0	0	131	1	
161	77	1	asymptomatic	125	304	0	2	162	1	
263	44	1	nonanginal	120	226	0	0	169	0	
230	52	0	nonanginal	136	196	0	2	169	0	
160	46	1	nontypical	101	197	1	0	156	0	
..	...	...	...	...	...	...	...	...	...	
223	53	1	asymptomatic	123	282	0	0	95	1	
274	59	1	typical	134	204	0	0	162	0	
84	52	1	nontypical	120	325	0	0	172	0	
28	43	1	asymptomatic	150	247	0	0	171	0	
22	58	1	nontypical	120	284	0	2	160	0	

	Oldpeak	Slope	Ca	Thal	AHD
143	1.8	2	0.0	reversable	Yes
161	0.0	1	3.0	normal	Yes
263	0.0	1	0.0	normal	No
230	0.1	2	0.0	normal	No
160	0.0	1	0.0	reversable	No
..	...	...	...	...	...
223	2.0	2	2.0	reversable	Yes
274	0.8	1	2.0	normal	Yes
84	0.2	1	0.0	normal	No
28	1.5	1	0.0	normal	No
22	1.8	2	0.0	normal	Yes

[152 rows x 14 columns]

```
1 df=heart_data[['Age', 'Sex', 'ChestPain', 'RestBP', 'Chol']]
2 print(df)
3
```

	Age	Sex	ChestPain	RestBP	Chol
0	63	1	typical	145	233
1	67	1	asymptomatic	160	286
2	67	1	asymptomatic	120	229
3	37	1	nonanginal	130	250
4	41	0	nontypical	130	204
..	...	...	...	...	...
298	45	1	typical	110	264
299	68	1	asymptomatic	144	193
300	57	1	asymptomatic	130	131
301	57	0	nontypical	130	236
302	38	1	nonanginal	138	175

[303 rows x 5 columns]

```
1 df75 = df.sample(frac=0.75)
2 print(df75)
```

	Age	Sex	ChestPain	RestBP	Chol
87	53	0	nonanginal	128	216
178	43	1	nonanginal	130	315
186	42	1	nonanginal	120	240
88	53	0	asymptomatic	138	234
245	67	1	asymptomatic	120	237
..	...	...	...	...	...
133	51	1	asymptomatic	140	261
112	52	1	typical	118	186
97	60	0	asymptomatic	150	258
254	43	1	asymptomatic	115	303
204	43	1	asymptomatic	110	211

[227 rows x 5 columns]

```
1 df50 = df.sample(frac = 0.5)
2 print(df50)
```

	Age	Sex	ChestPain	RestBP	Chol
46	51	1	nonanginal	110	175
135	55	0	nontypical	135	250
183	59	1	typical	178	270
179	53	1	nonanginal	130	246
147	41	1	nonanginal	112	250
..	...	...	...	...	...
19	49	1	nontypical	130	266
124	65	1	typical	138	282
302	38	1	nonanginal	138	175
163	58	0	asymptomatic	100	248
169	45	0	nontypical	112	160

[152 rows x 5 columns]

```
1 from sklearn.model_selection import train_test_split
2 X = df
3 Y = heart_data['AHD']
```

```
4 Y = Y.map({'No':0, 'Yes':1})
5 Y
```

```
0      0
1      1
2      1
3      0
4      0
..
298    1
299    1
300    1
301    1
302    0
Name: AHD, Length: 303, dtype: int64
```

```
1 X_train, x_test, y_train, y_test = train_test_split(X,Y,test_size = 0.25 , r
```

```
1 X_train.head()
```

	Age	Sex	ChestPain	RestBP	Chol
118	63	1	asymptomatic	130	330
265	42	1	asymptomatic	136	315
301	57	0	nontypical	130	236
162	54	0	nonanginal	110	214
227	67	0	nonanginal	152	277

```
1 y_train.head()
```

```
118    1
265    1
301    1
162    0
227    0
Name: AHD, dtype: int64
```

```
1 x_test.head()
```

	Age	Sex	ChestPain	RestBP	Chol
281	47	1	nonanginal	130	253
81	53	0	asymptomatic	130	264
40	65	0	asymptomatic	150	225
175	57	1	asymptomatic	152	274
58	54	1	nonanginal	125	273

```
1 y_test.head()
```

```
281    0
81    0
40    1
175   1
58    0
Name: AHD, dtype: int64
```

```
1 data = heart_data.iloc[:,5:13]
2 data.head()
```

	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	
0	1	2	150	0	2.3	3	0.0	fixed	
1	0	2	108	1	1.5	2	3.0	normal	
2	0	2	129	1	2.6	2	2.0	reversable	
3	0	0	187	0	3.5	3	0.0	normal	
4	0	2	172	0	1.4	1	0.0	normal	

1