

# ML02

August 14, 2024

```
[2]: import pandas as pd
```

```
[3]: df = pd.read_csv("Heart.csv")
```

```
[4]: df.head()
```

```
[4]:
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	\
0	1	63	1	typical	145	233	1	2	150	
1	2	67	1	asymptomatic	160	286	0	2	108	
2	3	67	1	asymptomatic	120	229	0	2	129	
3	4	37	1	nonanginal	130	250	0	0	187	
4	5	41	0	nontypical	130	204	0	2	172	

  

	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	0	2.3	3	0.0	fixed	No
1	1	1.5	2	3.0	normal	Yes
2	1	2.6	2	2.0	reversable	Yes
3	0	3.5	3	0.0	normal	No
4	0	1.4	1	0.0	normal	No

```
[5]: df.tail()
```

```
[5]:
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	\
298	299	45	1	typical	110	264	0	0	132	
299	300	68	1	asymptomatic	144	193	1	0	141	
300	301	57	1	asymptomatic	130	131	0	0	115	
301	302	57	0	nontypical	130	236	0	2	174	
302	303	38	1	nonanginal	138	175	0	0	173	

  

	ExAng	Oldpeak	Slope	Ca	Thal	AHD
298	0	1.2	2	0.0	reversable	Yes
299	0	3.4	2	2.0	reversable	Yes
300	1	1.2	2	1.0	reversable	Yes
301	0	0.0	2	1.0	normal	Yes
302	0	0.0	1	NaN	normal	No

```
[6]: x = df.iloc[ :, 1:14]
```

```
[7]: x
```

```
[7]:      Age  Sex    ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  ExAng  \
0      63   1      typical    145   233   1      2     150    0
1      67   1  asymptomatic    160   286   0      2     108    1
2      67   1  asymptomatic    120   229   0      2     129    1
3      37   1   nonanginal    130   250   0      0     187    0
4      41   0   nontypical    130   204   0      2     172    0
..    ...  ...
298    45   1      typical    110   264   0      0     132    0
299    68   1  asymptomatic    144   193   1      0     141    0
300    57   1  asymptomatic    130   131   0      0     115    1
301    57   0   nontypical    130   236   0      2     174    0
302    38   1   nonanginal    138   175   0      0     173    0

      Oldpeak  Slope  Ca      Thal
0          2.3     3  0.0    fixed
1          1.5     2  3.0    normal
2          2.6     2  2.0  reversable
3          3.5     3  0.0    normal
4          1.4     1  0.0    normal
..         ...   ...  ...
298         1.2     2  0.0  reversable
299         3.4     2  2.0  reversable
300         1.2     2  1.0  reversable
301         0.0     2  1.0    normal
302         0.0     1  NaN    normal
```

[303 rows x 13 columns]

```
[8]: df.isnull
```

```
[8]: <bound method DataFrame.isnull of      Unnamed: 0  Age  Sex    ChestPain
RestBP  Chol  Fbs  RestECG  MaxHR  \
0          1  63   1      typical    145   233   1      2     150
1          2  67   1  asymptomatic    160   286   0      2     108
2          3  67   1  asymptomatic    120   229   0      2     129
3          4  37   1   nonanginal    130   250   0      0     187
4          5  41   0   nontypical    130   204   0      2     172
..         ...  ...
298        299  45   1      typical    110   264   0      0     132
299        300  68   1  asymptomatic    144   193   1      0     141
300        301  57   1  asymptomatic    130   131   0      0     115
301        302  57   0   nontypical    130   236   0      2     174
302        303  38   1   nonanginal    138   175   0      0     173

      ExAng  Oldpeak  Slope  Ca      Thal  AHD
```

0	0	2.3	3	0.0	fixed	No
1	1	1.5	2	3.0	normal	Yes
2	1	2.6	2	2.0	reversible	Yes
3	0	3.5	3	0.0	normal	No
4	0	1.4	1	0.0	normal	No
..	...	...	...	...	...	...
298	0	1.2	2	0.0	reversible	Yes
299	0	3.4	2	2.0	reversible	Yes
300	1	1.2	2	1.0	reversible	Yes
301	0	0.0	2	1.0	normal	Yes
302	0	0.0	1	NaN	normal	No

[303 rows x 15 columns]>

```
[9]: df.isna()
```

```
[9]:      Unnamed: 0    Age    Sex  ChestPain  RestBP    Chol    Fbs  RestECG  \
0          False  False  False         False  False  False  False  False
1          False  False  False         False  False  False  False  False
2          False  False  False         False  False  False  False  False
3          False  False  False         False  False  False  False  False
4          False  False  False         False  False  False  False  False
..          ...    ...    ...          ...    ...    ...    ...    ...
298        False  False  False         False  False  False  False  False
299        False  False  False         False  False  False  False  False
300        False  False  False         False  False  False  False  False
301        False  False  False         False  False  False  False  False
302        False  False  False         False  False  False  False  False
```

	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
..	...	...	...	...	...	...	...
298	False	False	False	False	False	False	False
299	False	False	False	False	False	False	False
300	False	False	False	False	False	False	False
301	False	False	False	False	False	False	False
302	False	False	False	False	True	False	False

[303 rows x 15 columns]

```
[10]: df.isnull()
```

```
[10]:      Unnamed: 0      Age      Sex  ChestPain  RestBP      Chol      Fbs  RestECG  \
0      False  False  False      False  False  False  False  False
1      False  False  False      False  False  False  False  False
2      False  False  False      False  False  False  False  False
3      False  False  False      False  False  False  False  False
4      False  False  False      False  False  False  False  False
..      ...      ...      ...      ...      ...      ...      ...
298     False  False  False      False  False  False  False  False
299     False  False  False      False  False  False  False  False
300     False  False  False      False  False  False  False  False
301     False  False  False      False  False  False  False  False
302     False  False  False      False  False  False  False  False

      MaxHR  ExAng  Oldpeak  Slope      Ca  Thal      AHD
0      False  False      False  False  False  False  False
1      False  False      False  False  False  False  False
2      False  False      False  False  False  False  False
3      False  False      False  False  False  False  False
4      False  False      False  False  False  False  False
..      ...      ...      ...      ...      ...      ...
298  False  False      False  False  False  False  False
299  False  False      False  False  False  False  False
300  False  False      False  False  False  False  False
301  False  False      False  False  False  False  False
302  False  False      False  False   True  False  False

[303 rows x 15 columns]
```

```
[11]: df.loc[ : , "Age":"RestBP"]
```

```
[11]:      Age  Sex      ChestPain  RestBP
0      63   1      typical      145
1      67   1  asymptomatic      160
2      67   1  asymptomatic      120
3      37   1   nonanginal      130
4      41   0   nontypical      130
..      ...  ...      ...      ...
298   45   1      typical      110
299   68   1  asymptomatic      144
300   57   1  asymptomatic      130
301   57   0   nontypical      130
302   38   1   nonanginal      138

[303 rows x 4 columns]
```

```
[12]: y = df['Age']
```

```
[13]: y
```

```
[13]: 0      63
      1      67
      2      67
      3      37
      4      41
      ..
     298     45
     299     68
     300     57
     301     57
     302     38
      Name: Age, Length: 303, dtype: int64
```

```
[14]: x = df.iloc[ 1:11, :]
```

```
[15]: x
```

```
[15]:   Unnamed: 0  Age  Sex  ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  \
1           2   67   1  asymptomatic    160   286   0         2    108
2           3   67   1  asymptomatic    120   229   0         2    129
3           4   37   1   nonanginal    130   250   0         0    187
4           5   41   0   nontypical    130   204   0         2    172
5           6   56   1   nontypical    120   236   0         0    178
6           7   62   0  asymptomatic    140   268   0         2    160
7           8   57   0  asymptomatic    120   354   0         0    163
8           9   63   1  asymptomatic    130   254   0         2    147
9          10   53   1  asymptomatic    140   203   1         2    155
10         11   57   1  asymptomatic    140   192   0         0    148

      ExAng  Oldpeak  Slope  Ca      Thal  AHD
1         1      1.5      2  3.0    normal  Yes
2         1      2.6      2  2.0  reversable  Yes
3         0      3.5      3  0.0    normal  No
4         0      1.4      1  0.0    normal  No
5         0      0.8      1  0.0    normal  No
6         0      3.6      3  2.0    normal  Yes
7         1      0.6      1  0.0    normal  No
8         0      1.4      2  1.0  reversable  Yes
9         1      3.1      3  0.0  reversable  Yes
10        0      0.4      2  0.0     fixed  No
```

```
[16]: z = df[ : :2]
```

```
[17]: z
```

```
[17]:
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	\
0	1	63	1	typical	145	233	1	2	150	
2	3	67	1	asymptomatic	120	229	0	2	129	
4	5	41	0	nontypical	130	204	0	2	172	
6	7	62	0	asymptomatic	140	268	0	2	160	
8	9	63	1	asymptomatic	130	254	0	2	147	
..	...	...	...	...	...	...	...	...	...	
294	295	63	0	asymptomatic	124	197	0	0	136	
296	297	59	1	asymptomatic	164	176	1	2	90	
298	299	45	1	typical	110	264	0	0	132	
300	301	57	1	asymptomatic	130	131	0	0	115	
302	303	38	1	nonanginal	138	175	0	0	173	

  

	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	0	2.3	3	0.0	fixed	No
2	1	2.6	2	2.0	reversible	Yes
4	0	1.4	1	0.0	normal	No
6	0	3.6	3	2.0	normal	Yes
8	0	1.4	2	1.0	reversible	Yes
..	...	...	...	...	...	...
294	1	0.0	2	0.0	normal	Yes
296	0	1.0	2	2.0	fixed	Yes
298	0	1.2	2	0.0	reversible	Yes
300	1	1.2	2	1.0	reversible	Yes
302	0	0.0	1	NaN	normal	No

[152 rows x 15 columns]

```
[18]: y = df[df['Age'] > 60]
#get
```

```
[19]: y
```

```
[19]:
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	\
0	1	63	1	typical	145	233	1	2	150	
1	2	67	1	asymptomatic	160	286	0	2	108	
2	3	67	1	asymptomatic	120	229	0	2	129	
6	7	62	0	asymptomatic	140	268	0	2	160	
8	9	63	1	asymptomatic	130	254	0	2	147	
..	...	...	...	...	...	...	...	...	...	
284	285	61	1	asymptomatic	148	203	0	0	161	
290	291	67	1	nonanginal	152	212	0	2	150	
293	294	63	1	asymptomatic	140	187	0	2	144	
294	295	63	0	asymptomatic	124	197	0	0	136	
299	300	68	1	asymptomatic	144	193	1	0	141	

  

	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	0	2.3	3	0.0	fixed	No
1	1	2.6	2	2.0	reversible	Yes
2	0	1.4	1	0.0	normal	No
6	0	3.6	3	2.0	normal	Yes
8	0	1.4	2	1.0	reversible	Yes
..	...	...	...	...	...	...
284	1	0.0	2	0.0	normal	Yes
290	0	1.0	2	2.0	fixed	Yes
293	0	1.2	2	0.0	reversible	Yes
294	1	1.2	2	1.0	reversible	Yes
299	0	0.0	1	NaN	normal	No

0	0	2.3	3	0.0	fixed	No
1	1	1.5	2	3.0	normal	Yes
2	1	2.6	2	2.0	reversible	Yes
6	0	3.6	3	2.0	normal	Yes
8	0	1.4	2	1.0	reversible	Yes
..	...	...	...	...	...	...
284	0	0.0	1	1.0	reversible	Yes
290	0	0.8	2	0.0	reversible	Yes
293	1	4.0	1	2.0	reversible	Yes
294	1	0.0	2	0.0	normal	Yes
299	0	3.4	2	2.0	reversible	Yes

[79 rows x 15 columns]

```
[20]: z = df.iloc[2:11, 5:11]
```

```
[21]: z
```

```
[21]:
```

	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak
2	229	0	2	129	1	2.6
3	250	0	0	187	0	3.5
4	204	0	2	172	0	1.4
5	236	0	0	178	0	0.8
6	268	0	2	160	0	3.6
7	354	0	0	163	1	0.6
8	254	0	2	147	0	1.4
9	203	1	2	155	1	3.1
10	192	0	0	148	0	0.4

```
[22]: y = df["AHD"]
```

```
[23]: y
```

```
[23]:
```

0	No
1	Yes
2	Yes
3	No
4	No
...	
298	Yes
299	Yes
300	Yes
301	Yes
302	No

Name: AHD, Length: 303, dtype: object

```
[24]: y= df.isnull().any(axis=1)
```

```
[25]: y
```

```
[25]: 0      False
      1      False
      2      False
      3      False
      4      False
      ...
      298    False
      299    False
      300    False
      301    False
      302     True
      Length: 303, dtype: bool
```

```
[26]: df[y]
```

```
[26]:      Unnamed: 0  Age  Sex  ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  \
      87          88   53    0   nonanginal    128   216    0         2    115
      166         167   52    1   nonanginal    138   223    0         0    169
      192         193   43    1  asymptomatic    132   247    1         2    143
      266         267   52    1  asymptomatic    128   204    1         0    156
      287         288   58    1   nontypical    125   220    0         0    144
      302         303   38    1   nonanginal    138   175    0         0    173

      ExAng  Oldpeak  Slope  Ca      Thal  AHD
      87      0      0.0     1  0.0      NaN  No
      166      0      0.0     1  NaN      normal  No
      192      1      0.1     2  NaN  reversable  Yes
      266      1      1.0     2  0.0      NaN  Yes
      287      0      0.4     2  NaN  reversable  No
      302      0      0.0     1  NaN      normal  No
```

```
[27]: df = df.drop('AHD', axis=1)
```

```
[28]: df
```

```
[28]:      Unnamed: 0  Age  Sex  ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  \
      0          1   63    1      typical    145   233    1         2    150
      1          2   67    1  asymptomatic    160   286    0         2    108
      2          3   67    1  asymptomatic    120   229    0         2    129
      3          4   37    1   nonanginal    130   250    0         0    187
      4          5   41    0   nontypical    130   204    0         2    172
      ..      ...  ...  ...      ...      ...  ...  ...
      298        299   45    1      typical    110   264    0         0    132
      299        300   68    1  asymptomatic    144   193    1         0    141
      300        301   57    1  asymptomatic    130   131    0         0    115
```



301	302	57	0	nontypical	130	236	0	2	174
302	303	38	1	nonanginal	138	175	0	0	173

	ExAng	Oldpeak	Slope	Ca	Thal
0	0	2.3	3	0.0	fixed
1	1	1.5	2	3.0	normal
2	1	2.6	2	2.0	reversible
3	0	3.5	3	0.0	normal
4	0	1.4	1	0.0	normal
..	...	...	...	...	...
298	0	1.2	2	0.0	reversible
299	0	3.4	2	2.0	reversible
300	1	1.2	2	1.0	reversible
301	0	0.0	2	1.0	normal
302	0	0.0	1	NaN	normal

[303 rows x 14 columns]

```
[29]: pip install scikit-learn
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: scikit-learn in
/home/admin1/.local/lib/python3.10/site-packages (1.5.1)
Requirement already satisfied: numpy>=1.19.5 in
/home/admin1/.local/lib/python3.10/site-packages (from scikit-learn) (2.0.1)
Requirement already satisfied: scipy>=1.6.0 in
/home/admin1/.local/lib/python3.10/site-packages (from scikit-learn) (1.14.0)
Requirement already satisfied: joblib>=1.2.0 in
/home/admin1/.local/lib/python3.10/site-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/home/admin1/.local/lib/python3.10/site-packages (from scikit-learn) (3.5.0)
```

[notice] A new release of pip is  
available: 24.0 -> 24.2

[notice] To update, run:

```
python3 -m pip install --upgrade pip
```

Note: you may need to restart the kernel to use updated packages.

```
[30]: from sklearn.model_selection import train_test_split
```

```
[31]: x = df.iloc[:, :-1]
      y = df.iloc[:, -1]
```

```
[32]: xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.5)
```

```
[33]: xtrain.shape
```

```
rows , columns = xtrain.shape

print(f"rows = {rows} and columns = {columns}")
```

rows = 151 and columns = 13

```
[34]: ytrain.shape
      rows = ytrain.shape

      print(f"rows = {rows} ")
```

rows = (151,)

```
[35]: xtest.shape
```

[35]: (152, 13)

```
[36]: ytest.shape
```

[36]: (152,)

```
[37]: rows
```

[37]: (151,)

```
[45]: #get the mean age of the patients.
      mean_age = df["Age"].mean()
      print(mean_age)
```

54.43894389438944

```
[44]: #get only Age, Sex, ChestPain, RestBP, Chol from the dataset.
      test = df.iloc[ :, 1:6]
      print(test)
```

	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]

```
[49]: #get the datatypes of each column
print(df.dtypes)
```

```
Unnamed: 0      int64
Age             int64
Sex             int64
ChestPain       object
RestBP          int64
Chol            int64
Fbs             int64
RestECG         int64
MaxHR           int64
ExAng           int64
Oldpeak         float64
Slope           int64
Ca              float64
Thal            object
dtype: object
```

```
[50]: zero_row = df[df.eq(0).any(axis = 1)]
print(zero_row)
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	\
0	1	63	1	typical	145	233	1	2	150	
1	2	67	1	asymptomatic	160	286	0	2	108	
2	3	67	1	asymptomatic	120	229	0	2	129	
3	4	37	1	nonanginal	130	250	0	0	187	
4	5	41	0	nontypical	130	204	0	2	172	
..	...	...	...	...	...	...	...	...	...	
298	299	45	1	typical	110	264	0	0	132	
299	300	68	1	asymptomatic	144	193	1	0	141	
300	301	57	1	asymptomatic	130	131	0	0	115	
301	302	57	0	nontypical	130	236	0	2	174	
302	303	38	1	nonanginal	138	175	0	0	173	

  

	ExAng	Oldpeak	Slope	Ca	Thal
0	0	2.3	3	0.0	fixed
1	1	1.5	2	3.0	normal
2	1	2.6	2	2.0	reversable
3	0	3.5	3	0.0	normal
4	0	1.4	1	0.0	normal
..	...	...	...	...	...
298	0	1.2	2	0.0	reversable
299	0	3.4	2	2.0	reversable
300	1	1.2	2	1.0	reversable

```

301      0      0.0      2  1.0      normal
302      0      0.0      1 NaN      normal

```

[299 rows x 14 columns]

```

[51]: zero_entries = df.eq(0)
      print(zero_entries)

```

```

      Unnamed: 0  Age  Sex  ChestPain  RestBP  Chol  Fbs  RestECG  \
0      False  False  False      False  False  False  False  False
1      False  False  False      False  False  False  True  False
2      False  False  False      False  False  False  True  False
3      False  False  False      False  False  False  True  True
4      False  False  True      False  False  False  True  False
..      ...      ...      ...      ...      ...      ...      ...
298     False  False  False      False  False  False  True  True
299     False  False  False      False  False  False  False  True
300     False  False  False      False  False  False  True  True
301     False  False  True      False  False  False  True  False
302     False  False  False      False  False  False  True  True

```

```

      MaxHR  ExAng  Oldpeak  Slope  Ca  Thal
0      False  True  False  False  True  False
1      False  False  False  False  False  False
2      False  False  False  False  False  False
3      False  True  False  False  True  False
4      False  True  False  False  True  False
..      ...      ...      ...      ...      ...
298  False  True  False  False  True  False
299  False  True  False  False  False  False
300  False  False  False  False  False  False
301  False  True  True  False  False  False
302  False  True  True  False  False  False

```

[303 rows x 14 columns]

```

[55]: y = df[df["ExAng"] == 0]
      print(y)

```

```

      Unnamed: 0  Age  Sex  ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  \
0              1  63   1      typical    145   233   1        2    150
3              4  37   1  nonanginal    130   250   0        0    187
4              5  41   0  nontypical    130   204   0        2    172
5              6  56   1  nontypical    120   236   0        0    178
6              7  62   0  asymptomatic  140   268   0        2    160
..      ...      ...      ...      ...      ...      ...
296          297  59   1  asymptomatic    164   176   1        2     90
298          299  45   1      typical    110   264   0        0    132

```

299	300	68	1	asymptomatic	144	193	1	0	141
301	302	57	0	nontypical	130	236	0	2	174
302	303	38	1	nonanginal	138	175	0	0	173

	ExAng	Oldpeak	Slope	Ca	Thal
0	0	2.3	3	0.0	fixed
3	0	3.5	3	0.0	normal
4	0	1.4	1	0.0	normal
5	0	0.8	1	0.0	normal
6	0	3.6	3	2.0	normal
..	...	...	...	...	...
296	0	1.0	2	2.0	fixed
298	0	1.2	2	0.0	reversable
299	0	3.4	2	2.0	reversable
301	0	0.0	2	1.0	normal
302	0	0.0	1	NaN	normal

[204 rows x 14 columns]

```
[57]: #get the shape of the data
df.shape
```

[57]: (303, 14)

```
[58]: #get the max value of age from the age column in the dataset.
y = df["Age"].max()
print(y)
```

77

```
[61]: #covid patients problem

TP = 45
TN = 395
FP = 55
FN = 5

Accuracy = (TP + TN) / (TP + TN + FN + FP)
print(f"Accuracy is: {Accuracy}")

Precision = TP / (TP + FP)
print(f"Precision is: {Precision}")

Recall = TP / (TP + FN)
print(f"Recall is: {Recall}")
```

```
F1_score = 2 / ( (1/Precision) + ( 1/Recall))  
print(f"f1-score is {F1_score}")
```

Accuracy is: 0.88  
Precision is: 0.45  
Recall is: 0.9  
f1-score is 0.6

[ ]:

[ ]: