

```
import pandas as pd

df=pd.read_csv("Heart.csv")

print(df.to_string())
```

225	226	34	0	nontypical	118	210	0	0	192	0	0.7	1	0.0	normal	No
226	227	47	1	asymptomatic	112	204	0	0	143	0	0.1	1	0.0	normal	No
227	228	67	0	nonanginal	152	277	0	0	172	0	0.0	1	1.0	normal	No
228	229	54	1	asymptomatic	110	206	0	2	108	1	0.0	2	1.0	normal	Yes
229	230	66	1	asymptomatic	112	212	0	2	132	1	0.1	1	1.0	normal	Yes
230	231	52	0	nonanginal	136	196	0	2	169	0	0.1	2	0.0	normal	No
231	232	55	0	asymptomatic	180	327	0	1	117	1	3.4	2	0.0	normal	Yes
232	233	49	1	nonanginal	118	149	0	2	126	0	0.8	1	3.0	normal	Yes
233	234	74	0	nontypical	120	269	0	2	121	1	0.2	1	1.0	normal	No
234	235	54	0	nonanginal	160	201	0	0	163	0	0.0	1	1.0	normal	No
235	236	54	1	asymptomatic	122	286	0	2	116	1	3.2	2	2.0	normal	Yes
236	237	56	1	asymptomatic	130	283	1	2	103	1	1.6	3	0.0	reversable	Yes
237	238	46	1	asymptomatic	120	249	0	2	144	0	0.8	1	0.0	reversable	Yes
238	239	49	0	nontypical	134	271	0	0	162	0	0.0	2	0.0	normal	No
239	240	42	1	nontypical	120	295	0	0	162	0	0.0	1	0.0	normal	No
240	241	41	1	nontypical	110	235	0	0	153	0	0.0	1	0.0	normal	No
241	242	41	0	nontypical	126	306	0	0	163	0	0.0	1	0.0	normal	No
242	243	49	0	asymptomatic	130	269	0	0	163	0	0.0	1	0.0	normal	No
243	244	61	1	typical	134	234	0	0	145	0	2.6	2	2.0	normal	Yes
244	245	60	0	nonanginal	120	178	1	0	96	0	0.0	1	0.0	normal	No
245	246	67	1	asymptomatic	120	237	0	0	71	0	1.0	2	0.0	normal	Yes
246	247	58	1	asymptomatic	100	234	0	0	156	0	0.1	1	1.0	reversable	Yes
247	248	47	1	asymptomatic	110	275	0	2	118	1	1.0	2	1.0	normal	Yes
248	249	52	1	asymptomatic	125	212	0	0	168	0	1.0	1	2.0	reversable	Yes
249	250	62	1	nontypical	128	208	1	2	140	0	0.0	1	0.0	normal	No
250	251	57	1	asymptomatic	110	201	0	0	126	1	1.5	2	0.0	fixed	No
251	252	58	1	asymptomatic	146	218	0	0	105	0	2.0	2	1.0	reversable	Yes
252	253	64	1	asymptomatic	128	263	0	0	105	1	0.2	2	1.0	reversable	No
253	254	51	0	nonanginal	120	295	0	2	157	0	0.6	1	0.0	normal	No
254	255	43	1	asymptomatic	115	303	0	0	181	0	1.2	2	0.0	normal	No
255	256	42	0	nonanginal	120	209	0	0	173	0	0.0	2	0.0	normal	No
256	257	67	0	asymptomatic	106	223	0	0	142	0	0.3	1	2.0	normal	No
257	258	76	0	nonanginal	140	197	0	1	116	0	1.1	2	0.0	normal	No
258	259	70	1	nontypical	156	245	0	2	143	0	0.0	1	0.0	normal	No
259	260	57	1	nontypical	124	261	0	0	141	0	0.3	1	0.0	reversable	Yes
260	261	44	0	nonanginal	118	242	0	0	149	0	0.3	2	1.0	normal	No
261	262	58	0	nontypical	136	319	1	2	152	0	0.0	1	2.0	normal	Yes
262	263	60	0	typical	150	240	0	0	171	0	0.9	1	0.0	normal	No
263	264	44	1	nonanginal	120	226	0	0	169	0	0.0	1	0.0	normal	No
264	265	61	1	asymptomatic	138	166	0	2	125	1	3.6	2	1.0	normal	Yes
265	266	42	1	asymptomatic	136	315	0	0	125	1	1.8	2	0.0	fixed	Yes
266	267	52	1	asymptomatic	128	204	1	0	156	1	1.0	2	0.0	NaN	Yes
267	268	59	1	nonanginal	126	218	1	0	134	0	2.2	2	1.0	fixed	Yes
268	269	40	1	asymptomatic	152	223	0	0	181	0	0.0	1	0.0	reversable	Yes
269	270	42	1	nonanginal	130	180	0	0	150	0	0.0	1	0.0	normal	No
270	271	61	1	asymptomatic	140	207	0	2	138	1	1.9	1	1.0	reversable	Yes
271	272	66	1	asymptomatic	160	228	0	2	138	0	2.3	1	0.0	fixed	No
272	273	46	1	asymptomatic	140	311	0	0	120	1	1.8	2	2.0	reversable	Yes
273	274	71	0	asymptomatic	112	149	0	0	125	0	1.6	2	0.0	normal	No
274	275	59	1	typical	134	204	0	0	162	0	0.8	1	2.0	normal	Yes
275	276	64	1	typical	170	227	0	2	155	0	0.6	2	0.0	reversable	No
276	277	66	0	nonanginal	146	278	0	2	152	0	0.0	2	1.0	normal	No
277	278	39	0	nonanginal	138	220	0	0	152	0	0.0	2	0.0	normal	No
278	279	57	1	nontypical	154	232	0	2	164	0	0.0	1	1.0	normal	Yes
279	280	58	0	asymptomatic	130	197	0	0	131	0	0.6	2	0.0	normal	No
280	281	57	1	asymptomatic	110	335	0	0	143	1	3.0	2	1.0	reversable	Yes
281	282	47	1	nonanginal	130	253	0	0	179	0	0.0	1	0.0	normal	No
282	283	55	0	asymptomatic	128	205	0	1	130	1	2.0	2	1.0	reversable	Yes
283	284	55	1	asymptomatic	122	122	0	0	174	0	0.0	1	0.0	normal	No

```
#a) Find Missing Values and replace the missing values with suitable alternative.
df.isnull()
```



	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
299	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
300	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
301	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
302	False	False	False	False	False	False	False	False	False	False	False	False	True	False	False

303 rows × 15 columns

df.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

```

df.head(5)

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	1	63	1	typical	145	233	1	2	150	0	2.3	3	0.0	fixed	No
1	2	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3.0	normal	Yes
2	3	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2.0	reversible	Yes
3	4	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0.0	normal	No
4	5	41	0	nontypical	130	204	0	2	172	0	1.4	1	0.0	normal	No

df.tail(5)

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
298	299	45	1	typical	110	264	0	0	132	0	1.2	2	0.0	reversible	Yes
299	300	68	1	asymptomatic	144	193	1	0	141	0	3.4	2	2.0	reversible	Yes
300	301	57	1	asymptomatic	130	131	0	0	115	1	1.2	2	1.0	reversible	Yes
301	302	57	0	nontypical	130	236	0	2	174	0	0.0	2	1.0	normal	Yes
302	303	38	1	nonanginal	138	175	0	0	173	0	0.0	1	NaN	normal	No

```
df1=df.isnull().sum()
```

```
print(df1)
```

```

Unnamed: 0    0
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

```

```
df2=df.isnull().sum().sum()
```

```
print(df2)
```

```
6
```

```
df.tail(10)
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
293	294	63	1	asymptomatic	140	187	0	2	144	1	4.0	1	2.0	reversable	Yes
294	295	63	0	asymptomatic	124	197	0	0	136	1	0.0	2	0.0	normal	Yes
295	296	41	1	nontypical	120	157	0	0	182	0	0.0	1	0.0	normal	No
296	297	59	1	asymptomatic	164	176	1	2	90	0	1.0	2	2.0	fixed	Yes
297	298	57	0	asymptomatic	140	241	0	0	123	1	0.2	2	0.0	reversable	Yes
298	299	45	1	typical	110	264	0	0	132	0	1.2	2	0.0	reversable	Yes
299	300	68	1	asymptomatic	144	193	1	0	141	0	3.4	2	2.0	reversable	Yes
300	301	57	1	asymptomatic	130	131	0	0	115	1	1.2	2	1.0	reversable	Yes
301	302	57	0	nontypical	130	236	0	2	174	0	0.0	2	1.0	normal	Yes
302	303	38	1	nonanginal	138	175	0	0	173	0	0.0	1	NaN	normal	No

```
df3=df.fillna(value=0)
```

```
print(df3)
```

	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	reversable	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	reversable	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  0.0    normal    No

```

```
[303 rows x 15 columns]
```

```
df3=df.fillna(value=5)
```

```
print(df3)
```

```

      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  5.0    normal  No

```

```
[303 rows x 15 columns]
```

```
df4=df.fillna(method='pad')
```

```
print(df3)
```

```

      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  5.0    normal  No

```

```
[303 rows x 15 columns]
```

```
print(df4)
```

```

      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	1.0	normal	No

[303 rows x 15 columns]

df5=df.fillna(method='bfill')

print(df5)

	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]

df6=df.fillna(value=df['Ca'].mean())

print(df6)

	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	0.000000	fixed	No
1	1	1.5	2	3.000000	normal	Yes

2	1	2.6	2	2.000000	reversable	Yes
3	0	3.5	3	0.000000	normal	No
4	0	1.4	1	0.000000	normal	No
...	...	...	...	...	...	...
298	0	1.2	2	0.000000	reversable	Yes
299	0	3.4	2	2.000000	reversable	Yes
300	1	1.2	2	1.000000	reversable	Yes
301	0	0.0	2	1.000000	normal	Yes
302	0	0.0	1	0.672241	normal	No

[303 rows x 15 columns]

df7=df.fillna(value=df['Ca'].min())

print(df7)

	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	reversable	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	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	0.0	normal	No

[303 rows x 15 columns]

df8=df.fillna(value=df['Ca'].max())

print(df8)

	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	reversable	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	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	3.0	normal	No

[303 rows x 15 columns]

#e) Find data type of each column.

```
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
AHD            object
dtype: object

```

```
df9=df.fillna(value=df['Age'].min())
```

```
print(df9)
```

```

      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  reversable  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  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  29.0    normal  No

```

```
[303 rows x 15 columns]
```

#g) Find Mean age of patients considering above dataset.

```
print(df['Age'].mean())
```

```
54.43894389438944
```

#h) Find shape of data.

```
df10=df.shape
```

```
print(df10)
```

```
(303, 15)
```

#f) Finding out Zero's.

```
zeros_count = (df == 0).sum()
```

```
print(zeros_count)
```

```

Unnamed: 0      0
Age            0
Sex            97
ChestPain      0
RestBP         0

```

```

Chol      0
Fbs       258
RestECG   151
MaxHR     0
ExAng     204
Oldpeak   99
Slope     0
Ca        176
Thal      0
AHD       0
dtype: int64

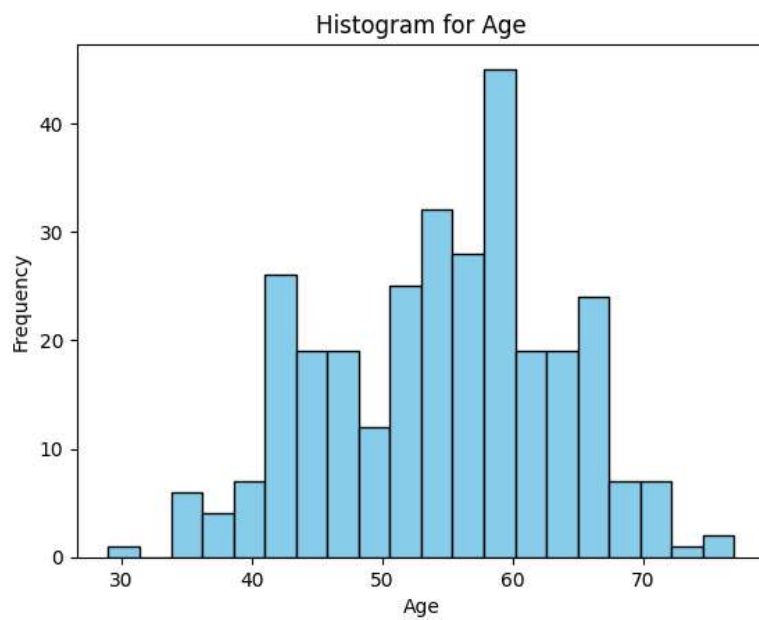
```

#d) Draw histogram for any two suitable attributes (E.g.age and Chol attributes for above dataset)  
import matplotlib.pyplot as plt

```

selected_column = 'Age'
plt.hist(df[selected_column], bins=20, color='skyblue', edgecolor='black')
plt.title(f'Histogram for {selected_column}')
plt.xlabel(selected_column)
plt.ylabel('Frequency')
plt.show()

```



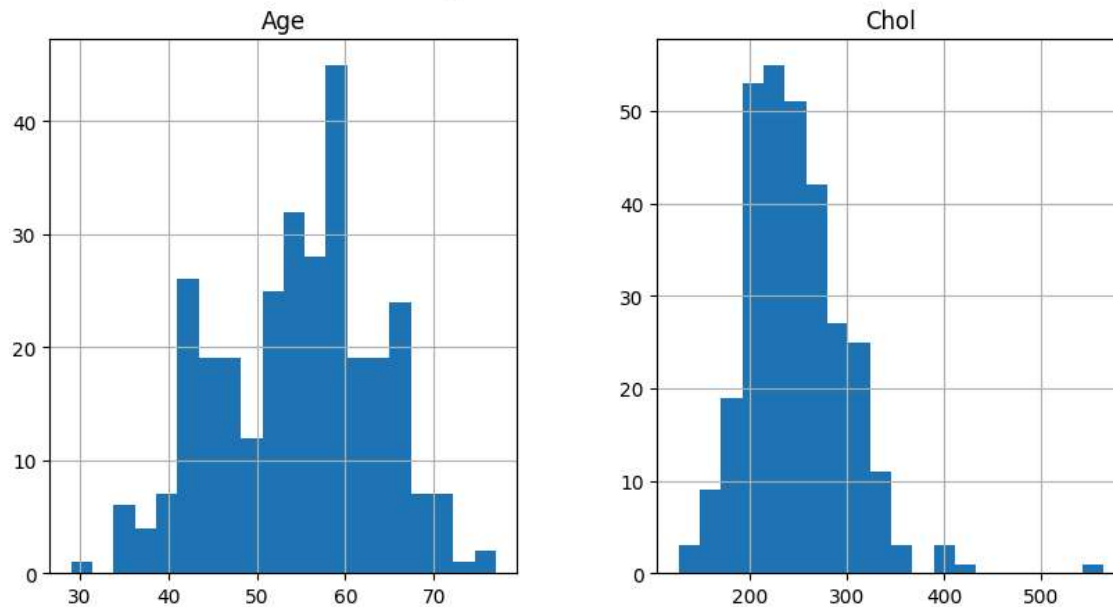
```

selected_columns = ['Age', 'Chol']
df[selected_columns].hist(bins=20, figsize=(10, 5))
plt.suptitle('Histogram for Selected Columns')
plt.show()

```



Histogram for Selected Columns



#c) Prepare boxplot analysis for each numerical attribute. Find outliers (if any) in each attribute in the dataset.

#seaborn is used here

# Assuming 'df' is your DataFrame

# Convert non-numeric values to NaN

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

df = df.apply(pd.to\_numeric, errors='coerce')

plt.figure(figsize=(15, 10))

sns.boxplot(data=df)

plt.title("Boxplot Analysis for Numerical Attributes")

plt.xticks(rotation=45)

plt.show()

# Outlier detection and handling

Q1 = df.quantile(0.25)

Q3 = df.quantile(0.75)

IQR = Q3 - Q1

outliers = ((df < (Q1 - 1.5 \* IQR)) | (df > (Q3 + 1.5 \* IQR))).sum()

print("Outliers:")

print(outliers)

