THEJASWINI J N
M.Sc [MATHEMATICS]
(COMPLETED)
JSS COLLEGE OF ARTS COMMERCE & SCIENCE,
OOTY ROAD, MYSORE.

#EXPLORATORY DATA ANALYICS

#CREATE A DATAFRAME

```
import pandas as pd
df = pd.read_csv('https://raw.githubusercontent.com/ameenmanna8824/DATASETS/main/heart_dis
df
```

	Unnamed: 0	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	s
0	0	63	1	3	145	233	1	0	150	0	2.3	
1	1	37	1	2	130	250	0	1	187	0	3.5	
2	2	41	0	1	130	204	0	0	172	0	1.4	
3	3	56	1	1	120	236	0	1	178	0	0.8	
4	4	57	0	0	120	354	0	1	163	1	0.6	
...	...	...	...	...	...	...	...	...	...	...	...	
298	298	57	0	0	140	241	0	1	123	1	0.2	
299	299	45	1	3	110	264	0	1	132	0	1.2	
300	300	68	1	0	144	193	1	1	141	0	3.4	
301	301	57	1	0	130	131	0	1	115	1	1.2	
302	302	57	0	1	130	236	0	0	174	0	0.0	

303 rows × 15 columns

```
type(df)
```

```
pandas.core.frame.DataFrame
```

```
df.shape
```

```
(303, 15)
```

```
df.size
```

```
4545
```

```
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   cp          303 non-null    int64  
 4   trestbps    303 non-null    int64  
 5   chol        303 non-null    int64  
 6   fbs         303 non-null    int64  
 7   restecg     303 non-null    int64  
 8   thalach     303 non-null    int64  
 9   exang       303 non-null    int64  
 10  oldpeak     303 non-null    float64
 11  slope       303 non-null    int64  
 12  ca          303 non-null    int64  
 13  thal        303 non-null    int64  
 14  target      303 non-null    int64  
dtypes: float64(1), int64(14)
memory usage: 35.6 KB
```

```
#SLICING
#SLICE ROW INDEXES FROM 30 TO 50
df[30:51]
```

	Unnamed: 0	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	sl
30	30	41	0	1	105	198	0	1	168	0	0.0	
31	31	65	1	0	120	177	0	1	140	0	0.4	
32	32	44	1	1	130	219	0	0	188	0	0.0	

```
#SLICE ROW INDEXES FROM 30 TO 50 AND COLUMN INDEXES FROM 0 TO 2
df.iloc[30:51,0:3]
```

	Unnamed: 0	age	sex
30	30	41	0
31	31	65	1
32	32	44	1
33	33	54	1
34	34	51	1
35	35	46	0
36	36	54	0
37	37	54	1
38	38	65	0
39	39	65	0
40	40	51	0
41	41	48	1
42	42	45	1
43	43	53	0
44	44	39	1
45	45	52	1
46	46	44	1
47	47	47	1
48	48	53	0
49	49	53	0
50	50	51	0

```
df.nunique()
```

```
Unnamed: 0    303
age            41
```

```
sex            2
cp             4
trestbps      49
chol         152
fbs            2
restecg        3
thalach       91
exang          2
oldpeak       40
slope          3
ca             5
thal           4
target         2
dtype: int64
```

```
df = df.drop(columns = 'Unnamed: 0')
df
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca
0	63	1	3	145	233	1	0	150	0	2.3	0	0
1	37	1	2	130	250	0	1	187	0	3.5	0	0
2	41	0	1	130	204	0	0	172	0	1.4	2	0
3	56	1	1	120	236	0	1	178	0	0.8	2	0
4	57	0	0	120	354	0	1	163	1	0.6	2	0
...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2
301	57	1	0	130	131	0	1	115	1	1.2	1	1
302	57	0	1	130	236	0	0	174	0	0.0	1	1

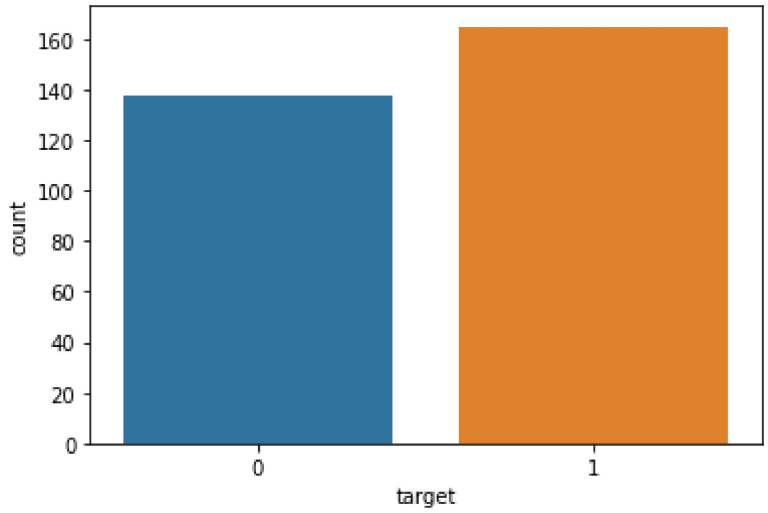
303 rows × 14 columns

```
df = df.drop(columns = 'restecg')
df
```

	age	sex	cp	trestbps	chol	fbs	thalach	exang	oldpeak	slope	ca	thal	tar
0	63	1	3	145	233	1	150	0	2.3	0	0	1	
1	37	1	2	130	250	0	187	0	3.5	0	0	2	
2	41	0	1	130	204	0	172	0	1.4	2	0	2	
3	56	1	1	120	236	0	178	0	0.8	2	0	2	
4	57	0	0	120	354	0	163	1	0.6	2	0	2	

```
#VISUALISATION
import seaborn as sns
sns.countplot(x = 'target',data = df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6ebe950fd0>
```




```
#NOW I WANT EXACT COUNT OF HOW MANY PEOPLES ARE TRAGET AND HOW MANY PEOPLE ARE NOT TARGET

df.groupby('target').size()
```

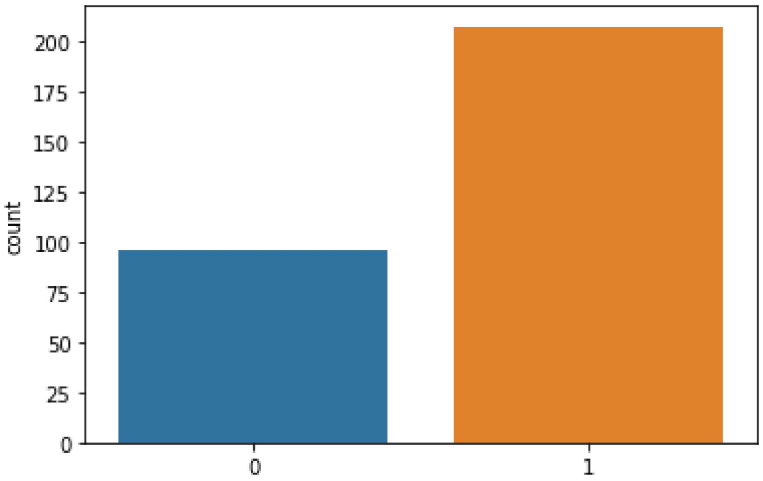
```
target
0    138
1    165
dtype: int64
```

```
df['target'].value_counts()
```

```
1    165
0    138
Name: target, dtype: int64
```

```
#THIS COUNT PLOT WILL TELL US HOW MANY MALES HOW MANY FEMALE WERE THERE ON HEART DISEASE
sns.countplot(x = 'sex',data = df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6ebe8357d0>
```




```
#FIND OUT HOW MANY MALES HAVE TARGET AND NOT TARGET AND SIMULTANEOUSLY
#FIND OUT HOW MANY FEMALES HAVE TARGET ANFD NOT TARGET
import numpy as np
target_1 = np.sum((df['sex']=='1')&['target']==0)
target_1
```

```
303
```

```
df.groupby(['sex','target']).size()
```

```
sex  target
0    0          24
     1          72
1    0         114
     1          93
dtype: int64
```

```
import numpy as np
young = np.sum((df['age']>=0)&(df['age']<20))
adult = np.sum((df['age']>=20)&(df['age']<40))
midage = np.sum((df['age']>=40)&(df['age']<60))
old = np.sum((df['age']>=60))
print(young)
print(adult)
print(midage)
print(old)
```

```
0
16
197
90
```

```
# FIND OUT THE YOUNGEST PERSON ABOARD HEART DISEASE
np.min(df['age'])
```

```
29
```

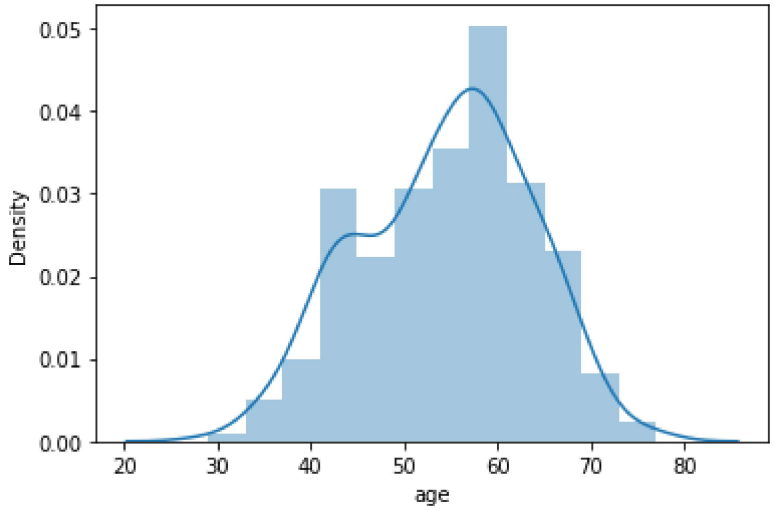
```
np.max(df['age'])
```

```
77
```

```
#DISTRIBUTION PLOT
```

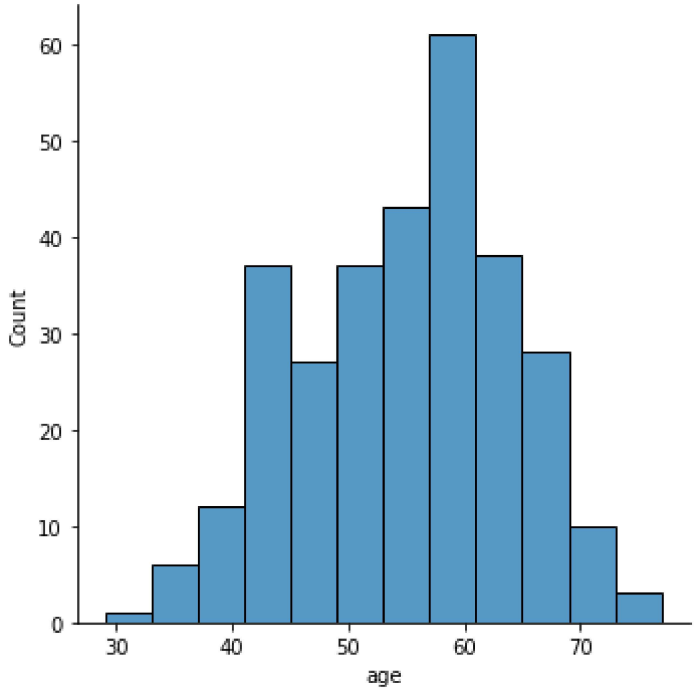
```
sns.distplot(df['age'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning:
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f6eb96c0710>
```




```
sns.displot(df['age'])
```

```
<seaborn.axisgrid.FacetGrid at 0x7f6eb96a3210>
```



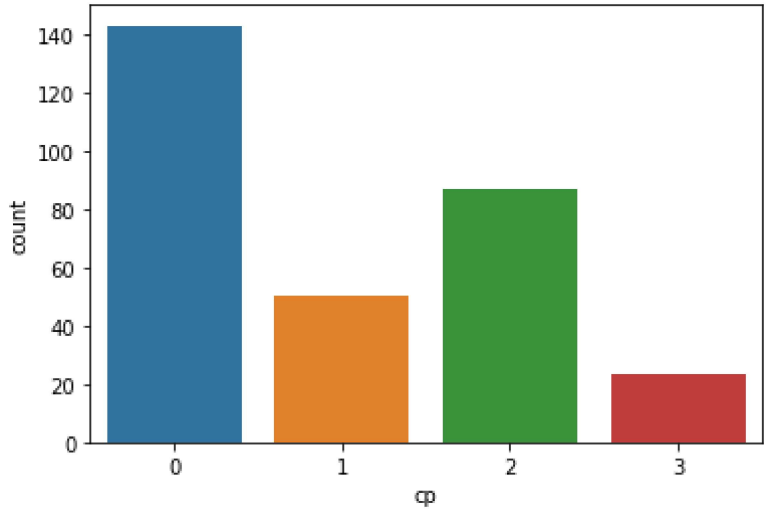

```
df.groupby('cp').size()
```

```
cp
0    143
```

```
1     50
2     87
3     23
dtype: int64
```

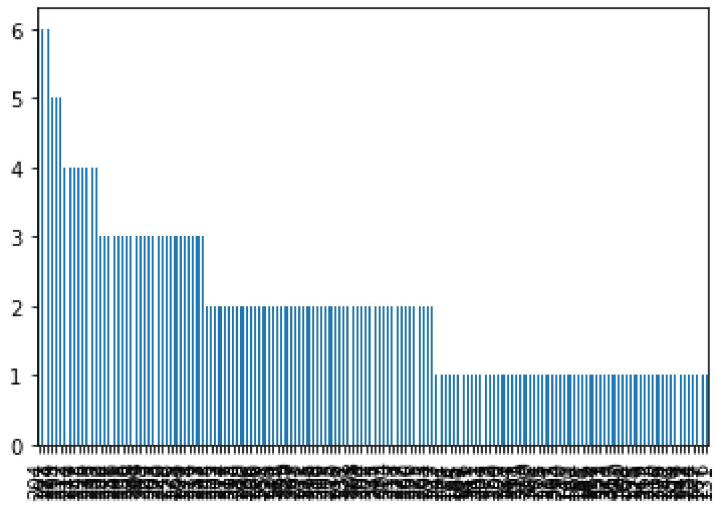
```
sns.countplot(x = 'cp',data = df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6eb958b4d0>
```



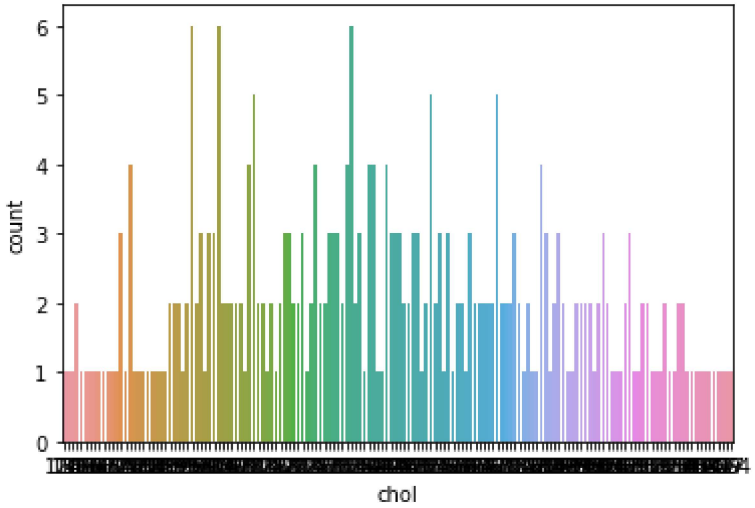

```
import matplotlib.pyplot as plt
df['chol'].value_counts().plot(kind = 'bar')
plt.xticks(rotation = '90')
```

```
(array([  0,   1,   2,   3,   4,   5,   6,   7,   8,   9,  10,  11,  12,
         13,  14,  15,  16,  17,  18,  19,  20,  21,  22,  23,  24,  25,
         26,  27,  28,  29,  30,  31,  32,  33,  34,  35,  36,  37,  38,
         39,  40,  41,  42,  43,  44,  45,  46,  47,  48,  49,  50,  51,
         52,  53,  54,  55,  56,  57,  58,  59,  60,  61,  62,  63,  64,
         65,  66,  67,  68,  69,  70,  71,  72,  73,  74,  75,  76,  77,
         78,  79,  80,  81,  82,  83,  84,  85,  86,  87,  88,  89,  90,
         91,  92,  93,  94,  95,  96,  97,  98,  99, 100, 101, 102, 103,
        104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116,
        117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129,
        130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142,
        143, 144, 145, 146, 147, 148, 149, 150, 151]),
 <a list of 152 Text major ticklabel objects>)
```



```
sns.countplot(x = 'chol',data = df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6eb8cce790>
```




```
#CONCLUSION:
#WE HAD COMPARE THE DIFFERENT AGES OF HEART DISEASE
#THE DIFFERENCE GIVEN US TO PREDICT THE AGES OF PATIENT
#THESE ANALYSIS HELP US TO KNOW ABOUT MORE DATA
# A MODEL DEPENDS ON NATURE OF THE DATA as well as the size of the data


THANK YOU
```