

```
In [1]: # importing python libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt # visualizing data
%matplotlib inline
import seaborn as sns
```

```
In [2]: # Import CSV File
df=pd.read_csv('Heart Disease data.csv')
```

```
In [3]: df.shape
```

```
Out[3]: (1025, 14)
```

```
In [4]: #Check null values
df.isnull().sum()
```

```
Out[4]: 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
```

```
In [5]: df=df.drop_duplicates()
```

```
In [6]: #shape of Data
df.shape
```

```
Out[6]: (302, 14)
```

In [7]: `df.head()`

```
Out[7]:
```

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

In [8]: `df.corr()`

```
Out[8]:
```

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| age | 1.000000 | -0.094962 | -0.063107 | 0.283121 | 0.207216 | 0.119492 | -0.111590 | -0.395235 | 0.093216 | 0.206040 | -0.164124 | 0.302261 | 0.065317 | -0.221476 |
| sex | -0.094962 | 1.000000 | -0.051740 | -0.057647 | -0.195571 | 0.046022 | -0.060351 | -0.046439 | 0.143460 | 0.098322 | -0.032990 | 0.113060 | 0.211452 | -0.283609 |
| cp | -0.063107 | -0.051740 | 1.000000 | 0.046486 | -0.072682 | 0.096018 | 0.041561 | 0.293367 | -0.392937 | -0.146692 | 0.116854 | -0.195356 | -0.160370 | 0.432080 |
| trestbps | 0.283121 | -0.057647 | 0.046486 | 1.000000 | 0.125256 | 0.178125 | -0.115367 | -0.048023 | 0.068526 | 0.194600 | -0.122873 | 0.099248 | 0.062870 | -0.146269 |
| chol | 0.207216 | -0.195571 | -0.072682 | 0.125256 | 1.000000 | 0.011428 | -0.147602 | -0.005308 | 0.064099 | 0.050086 | 0.000417 | 0.086878 | 0.096810 | -0.081437 |
| fbs | 0.119492 | 0.046022 | 0.096018 | 0.178125 | 0.011428 | 1.000000 | -0.083081 | -0.007169 | 0.024729 | 0.004514 | -0.058654 | 0.144935 | -0.032752 | -0.026826 |
| restecg | -0.111590 | -0.060351 | 0.041561 | -0.115367 | -0.147602 | -0.083081 | 1.000000 | 0.041210 | -0.068807 | -0.056251 | 0.090402 | -0.083112 | -0.010473 | 0.134874 |
| thalach | -0.395235 | -0.046439 | 0.293367 | -0.048023 | -0.005308 | -0.007169 | 0.041210 | 1.000000 | -0.377411 | -0.342201 | 0.384754 | -0.228311 | -0.094910 | 0.419955 |
| exang | 0.093216 | 0.143460 | -0.392937 | 0.068526 | 0.064099 | 0.024729 | -0.068807 | -0.377411 | 1.000000 | 0.286766 | -0.256106 | 0.125377 | 0.205826 | -0.435601 |
| oldpeak | 0.206040 | 0.098322 | -0.146692 | 0.194600 | 0.050086 | 0.004514 | -0.056251 | -0.342201 | 0.286766 | 1.000000 | -0.576314 | 0.236560 | 0.209090 | -0.429146 |
| slope | -0.164124 | -0.032990 | 0.116854 | -0.122873 | 0.000417 | -0.058654 | 0.090402 | 0.384754 | -0.256106 | -0.576314 | 1.000000 | -0.092236 | -0.103314 | 0.343940 |
| ca | 0.302261 | 0.113060 | -0.195356 | 0.099248 | 0.086878 | 0.144935 | -0.083112 | -0.228311 | 0.125377 | 0.236560 | -0.092236 | 1.000000 | 0.160085 | -0.408992 |
| thal | 0.065317 | 0.211452 | -0.160370 | 0.062870 | 0.096810 | -0.032752 | -0.010473 | -0.094910 | 0.205826 | 0.209090 | -0.103314 | 0.160085 | 1.000000 | -0.343101 |
| target | -0.221476 | -0.283609 | 0.432080 | -0.146269 | -0.081437 | -0.026826 | 0.134874 | 0.419955 | -0.435601 | -0.429146 | 0.343940 | -0.408992 | -0.343101 | 1.000000 |

In [9]: `df.describe()`

Out[9]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca |
|-------|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| count | 302.00000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 | 302.000000 |
| mean | 54.42053 | 0.682119 | 0.963576 | 131.602649 | 246.500000 | 0.149007 | 0.526490 | 149.569536 | 0.327815 | 1.043046 | 1.397351 | 0.718543 |
| std | 9.04797 | 0.466426 | 1.032044 | 17.563394 | 51.753489 | 0.356686 | 0.526027 | 22.903527 | 0.470196 | 1.161452 | 0.616274 | 1.006748 |
| min | 29.00000 | 0.000000 | 0.000000 | 94.000000 | 126.000000 | 0.000000 | 0.000000 | 71.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 48.00000 | 0.000000 | 0.000000 | 120.000000 | 211.000000 | 0.000000 | 0.000000 | 133.250000 | 0.000000 | 0.000000 | 1.000000 | 0.000000 |
| 50% | 55.50000 | 1.000000 | 1.000000 | 130.000000 | 240.500000 | 0.000000 | 1.000000 | 152.500000 | 0.000000 | 0.800000 | 1.000000 | 0.000000 |
| 75% | 61.00000 | 1.000000 | 2.000000 | 140.000000 | 274.750000 | 0.000000 | 1.000000 | 166.000000 | 1.000000 | 1.600000 | 2.000000 | 1.000000 |
| max | 77.00000 | 1.000000 | 3.000000 | 200.000000 | 564.000000 | 1.000000 | 2.000000 | 202.000000 | 1.000000 | 6.200000 | 2.000000 | 4.000000 |

In [10]: `df.info`

```
Out[10]: <bound method DataFrame.info of
0      52      1      0      125      212      0      1      168      0      1.0
1      53      1      0      140      203      1      0      155      1      3.1
2      70      1      0      145      174      0      1      125      1      2.6
3      61      1      0      148      203      0      1      161      0      0.0
4      62      0      0      138      294      1      1      106      0      1.9
..      ...      ...      ..      ...      ...      ...      ...      ...      ...      ...
723    68      0      2      120      211      0      0      115      0      1.5
733    44      0      2      108      141      0      1      175      0      0.6
739    52      1      0      128      255      0      1      161      1      0.0
843    59      1      3      160      273      0      0      125      0      0.0
878    54      1      0      120      188      0      1      113      0      1.4

      slope      ca      thal      target
0          2      2          3          0
1          0      0          3          0
2          0      0          3          0
3          2      1          3          0
4          1      3          2          0
..      ...      ..      ...      ...
723        1      0          2          1
733        1      0          2          1
739        2      1          3          0
843        2      0          2          0
878        1      1          3          0

[302 rows x 14 columns]>
```

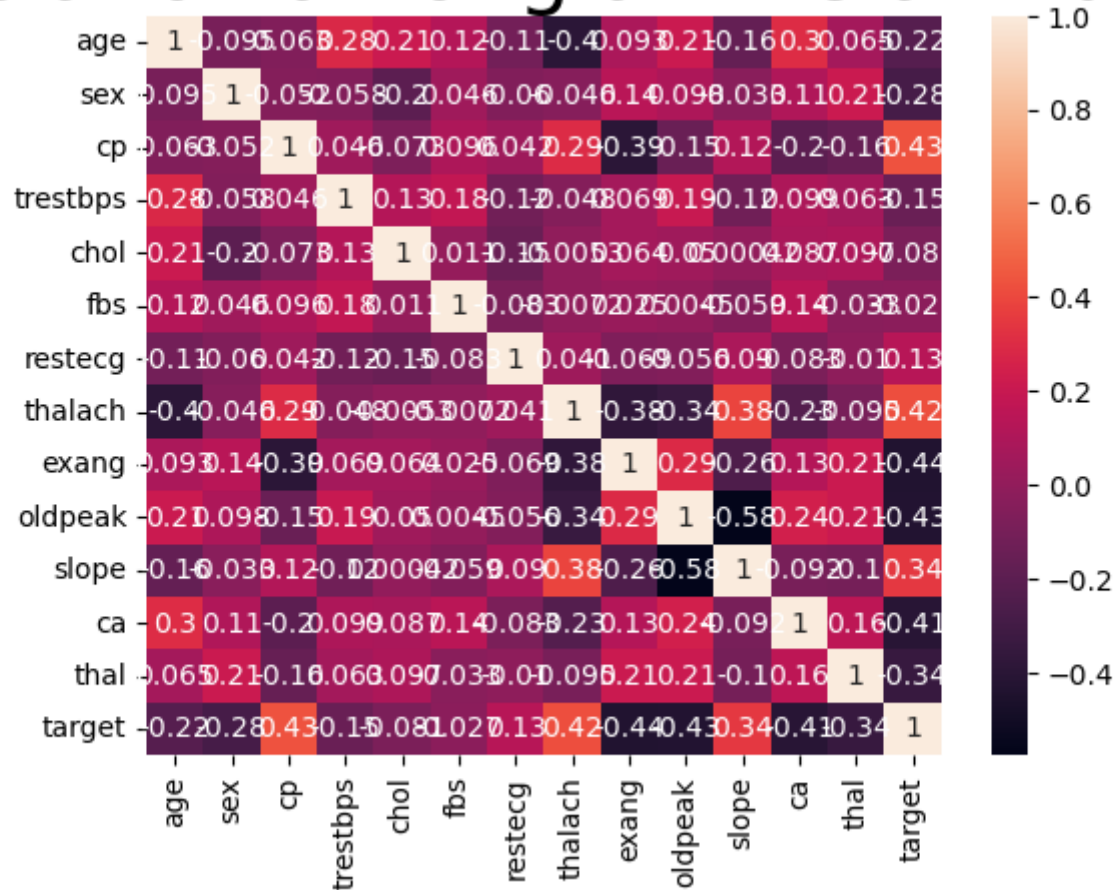
Min Age:29 Max Age:77

Exploratory Data Analysis

```
In [11]: sns.heatmap(df.corr(),annot=True)
plt.title('Correlation among all the attributes',fontsize=30)
```

```
Out[11]: Text(0.5, 1.0, 'Correlation among all the attributes')
```

Correlation among all the attributes

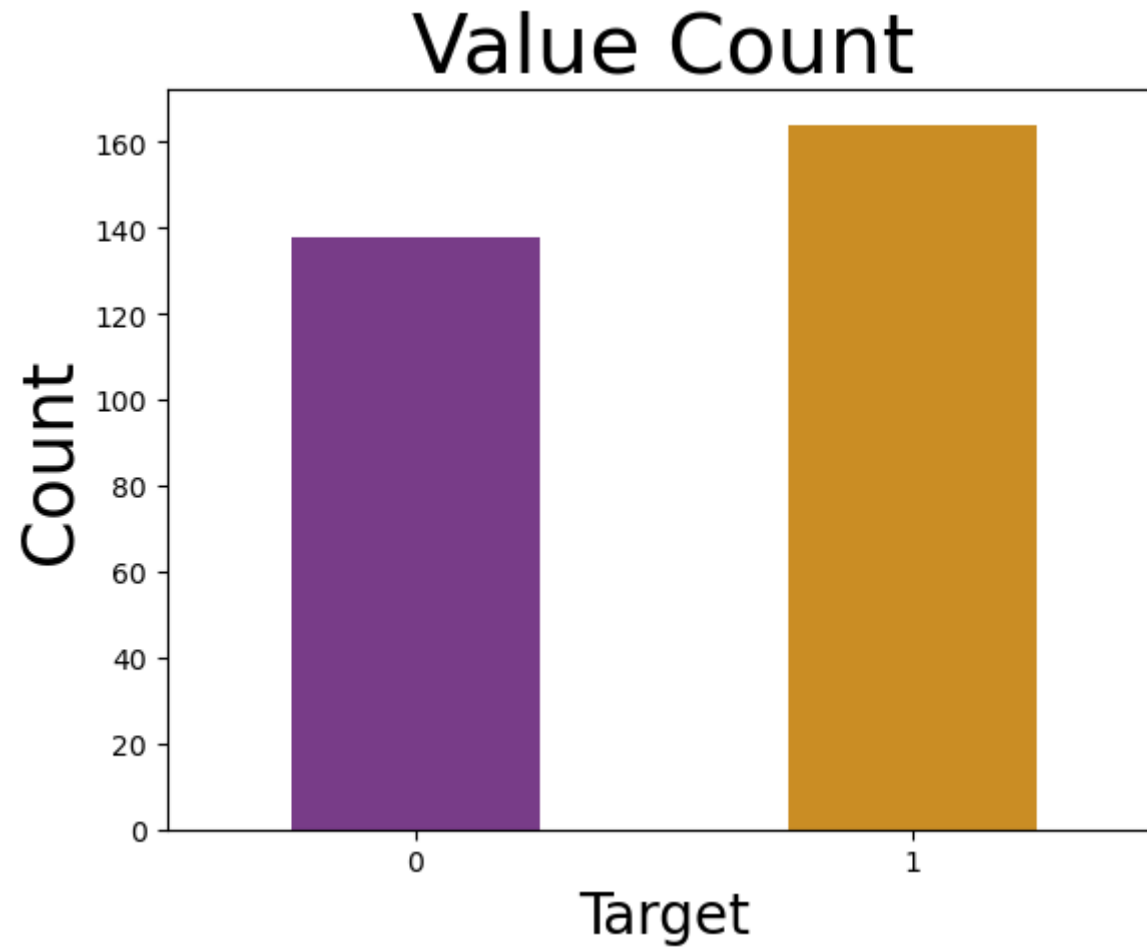


```
In [12]: df['target'].value_counts()
```

```
Out[12]: target
1      164
0      138
Name: count, dtype: int64
```

```
In [15]: sns.countplot(x='target',data=df,palette='CMRmap',width=0.5)
plt.xlabel('Target',fontsize=20)
plt.ylabel('Count',fontsize=25)
plt.title('Value Count',fontsize=30)
```

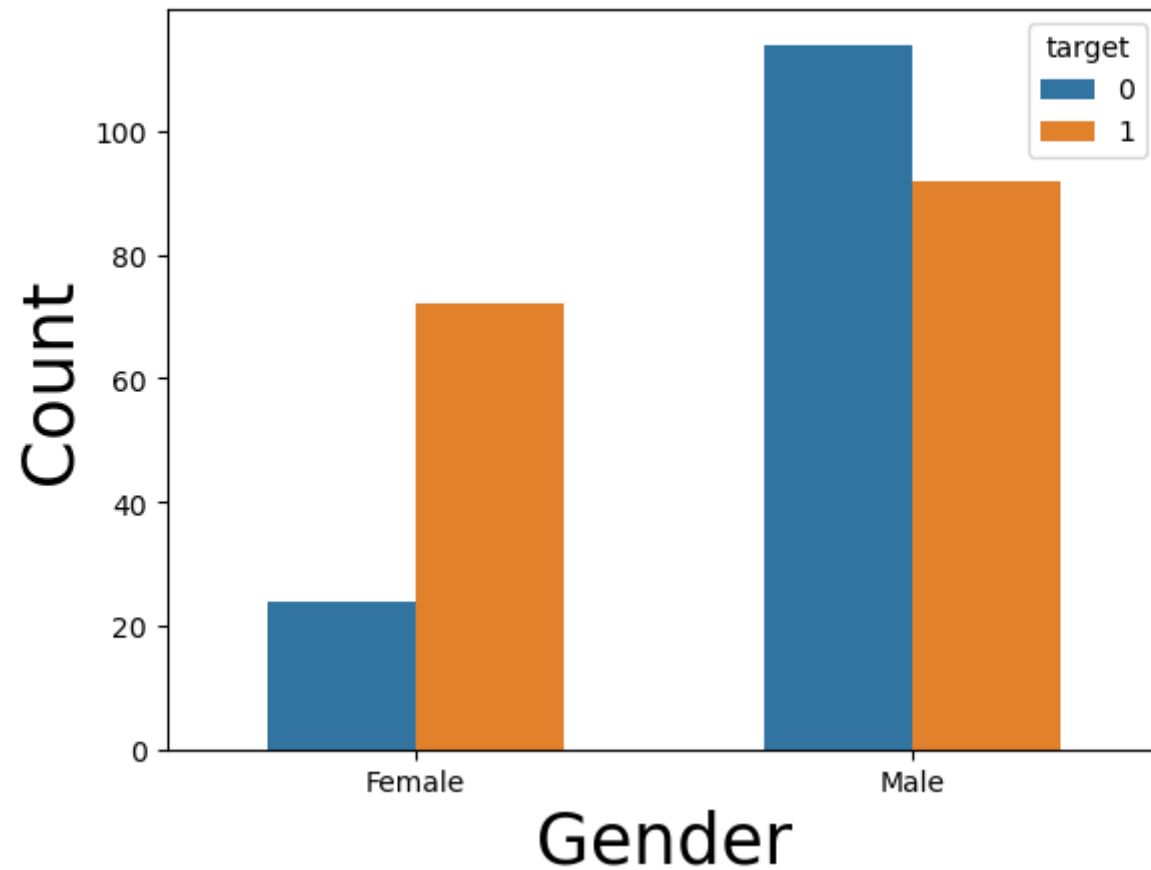
Out[15]: Text(0.5, 1.0, 'Value Count')



```
In [16]: sns.countplot(x='sex',hue='target',data=df,width=0.6)
plt.xlabel('Gender',fontsize=25)
plt.ylabel('Count',fontsize=25)
plt.title('Gender Count based on heart disease',fontsize=30)
plt.xticks([1,0],['Male','Female'])
plt.show
```

Out[16]: <function matplotlib.pyplot.show(close=None, block=None)>

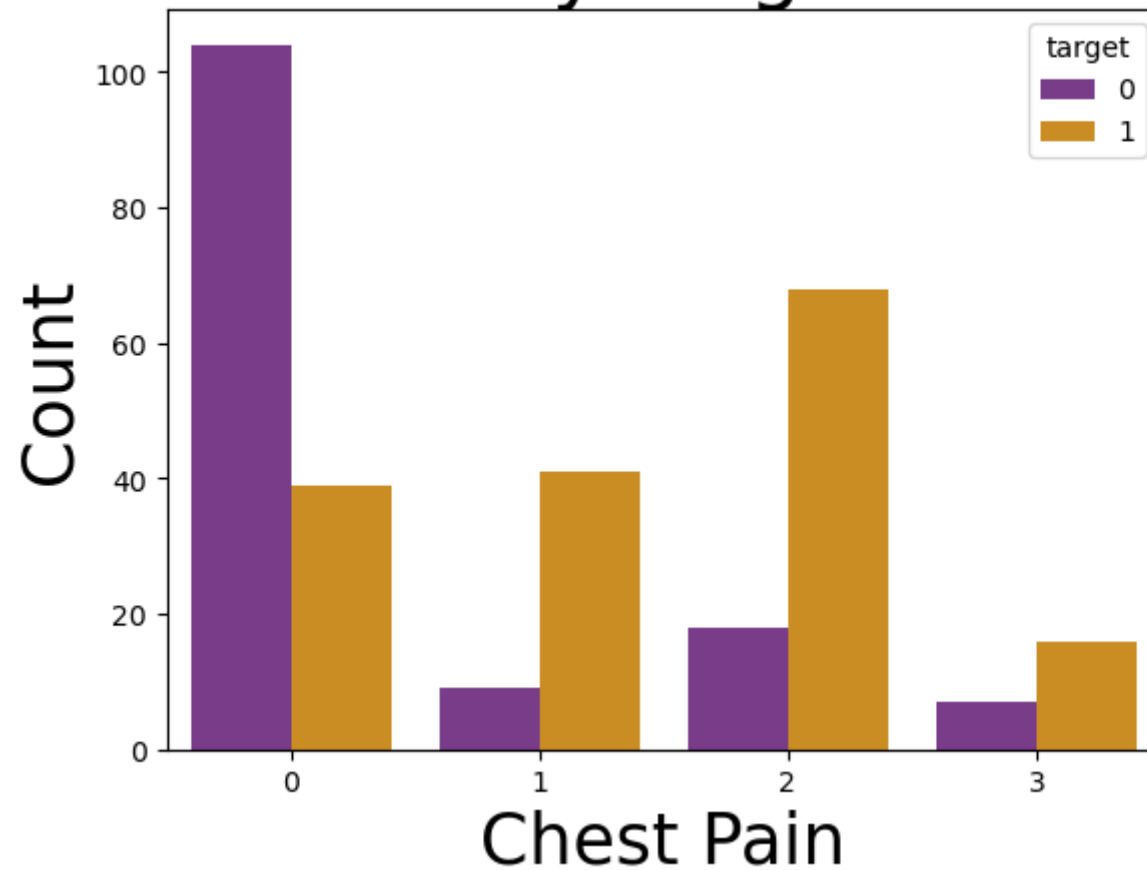
Gender Count based on heart disease



```
In [17]: sns.countplot(data=df,x='cp',hue='target',palette='CMRmap')
plt.xlabel('Chest Pain',fontsize=25)
plt.ylabel('Count',fontsize=25)
plt.title('Chest Pain by Targeted People',fontsize=30)
```

```
Out[17]: Text(0.5, 1.0, 'Chest Pain by Targeted People')
```

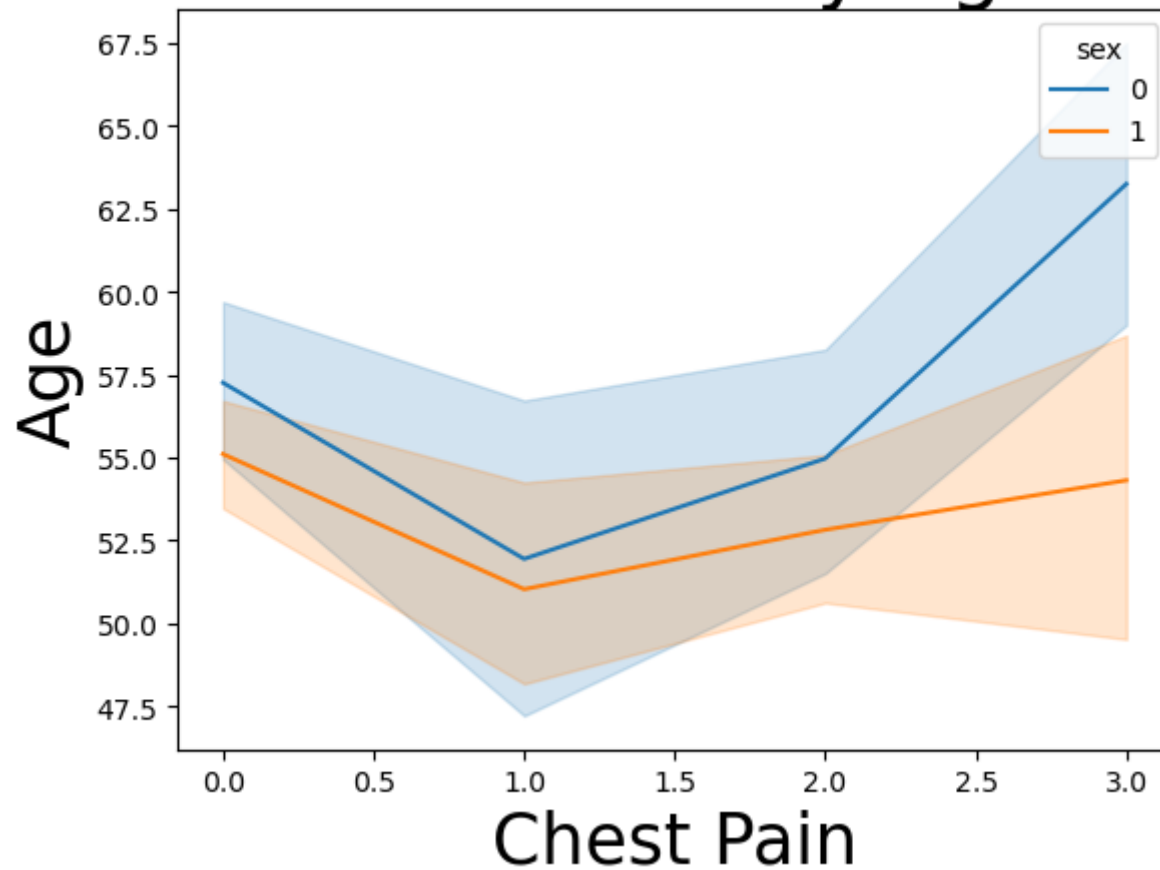
Chest Pain by Targeted People



```
In [18]: sns.lineplot(x='cp',data=df,y='age',hue='sex')
#sns.set(rc={'figure.figsize':(15,5)})
plt.xlabel('Chest Pain',fontsize=25)
plt.ylabel('Age',fontsize=25)
plt.title('Chest Pain by Age',fontsize=30)
```

```
Out[18]: Text(0.5, 1.0, 'Chest Pain by Age')
```

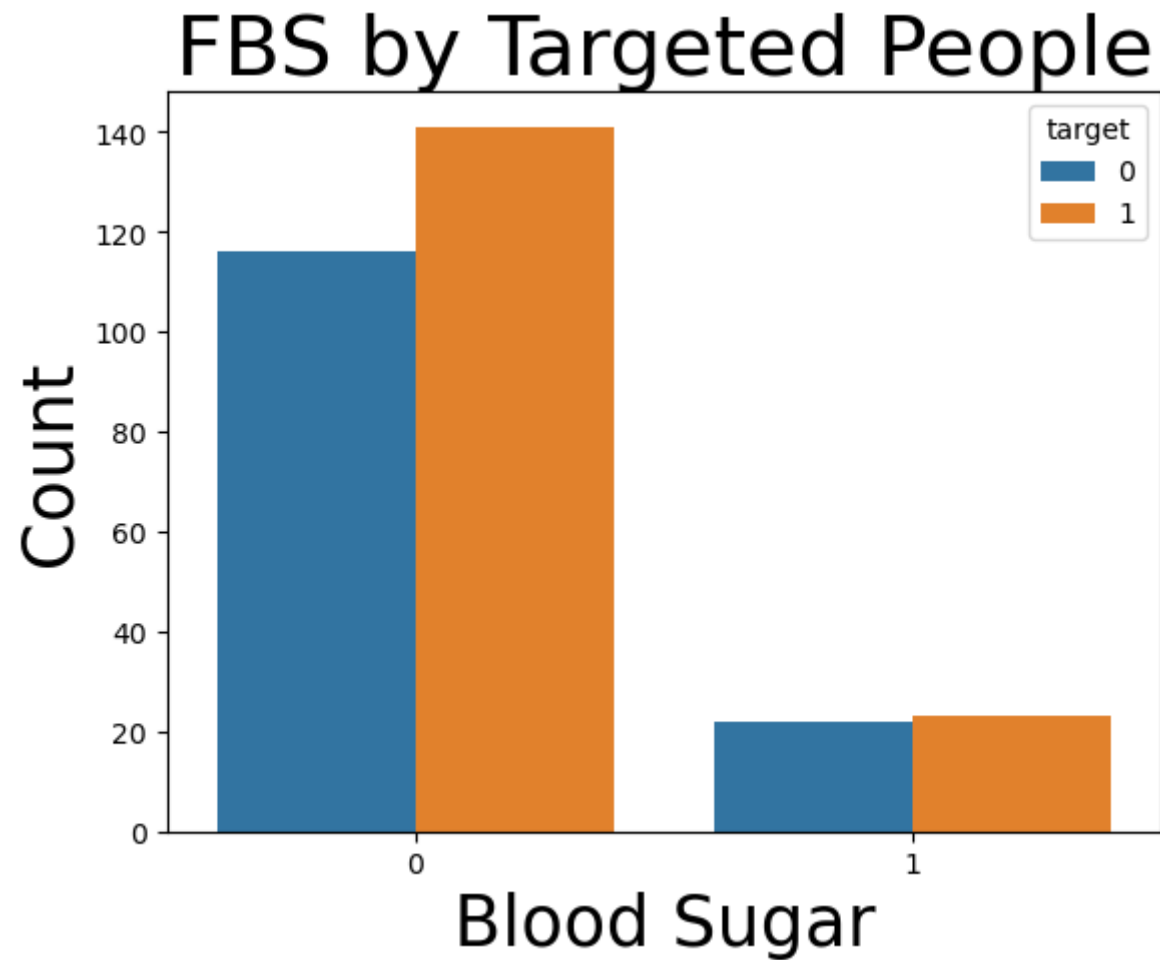

Chest Pain by Age



```
In [20]: # fasting blood sugar distribution according

sns.countplot(data=df,x='fbs',hue='target')
plt.xlabel('Blood Sugar',fontsize=25)
plt.ylabel('Count',fontsize=25)
plt.title('FBS by Targeted People',fontsize=30)
```

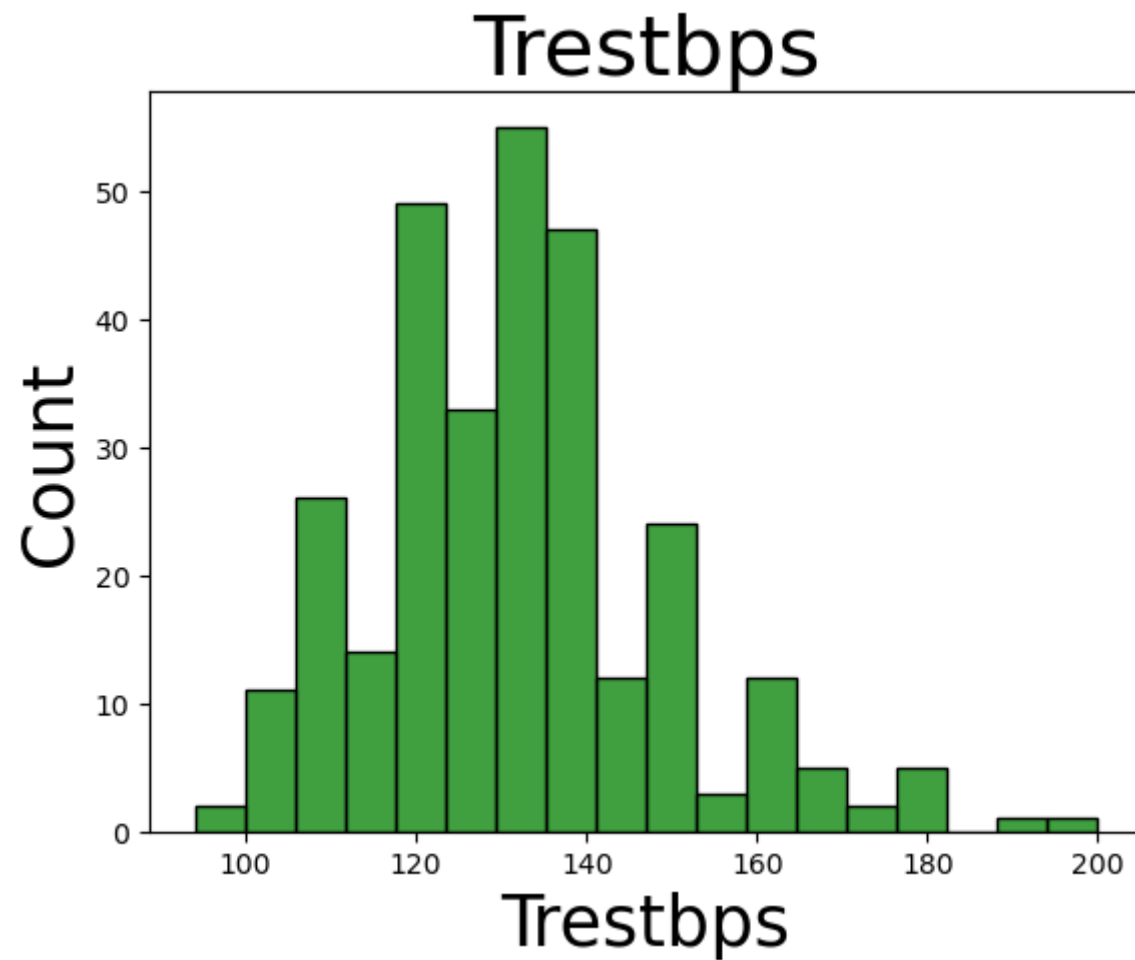
```
Out[20]: Text(0.5, 1.0, 'FBS by Targeted People')
```



```
In [23]: #Trestbps Count

sns.histplot(x='trestbps',data=df,color='Green')
plt.show
plt.xlabel('Trestbps',fontsize=25)
plt.ylabel('Count',fontsize=25)
plt.title('Trestbps',fontsize=30)
```

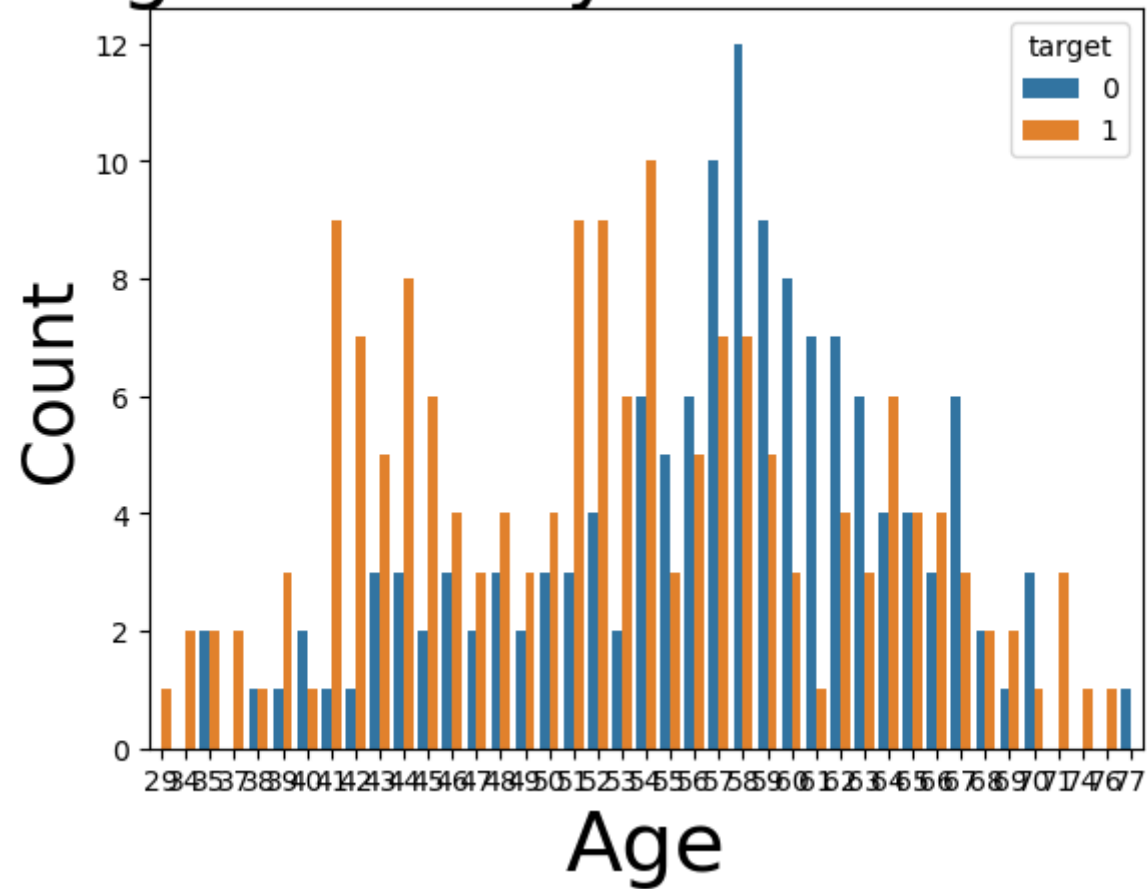
```
Out[23]: Text(0.5, 1.0, 'Trestbps')
```



```
In [25]: sns.countplot(x='age', data=df, hue='target')
plt.xlabel('Age', fontsize=30)
plt.ylabel('Count', fontsize=25)
plt.title('Age count by Heart Disease', fontsize=30)
plt.show
```

```
Out[25]: <function matplotlib.pyplot.show(close=None, block=None)>
```

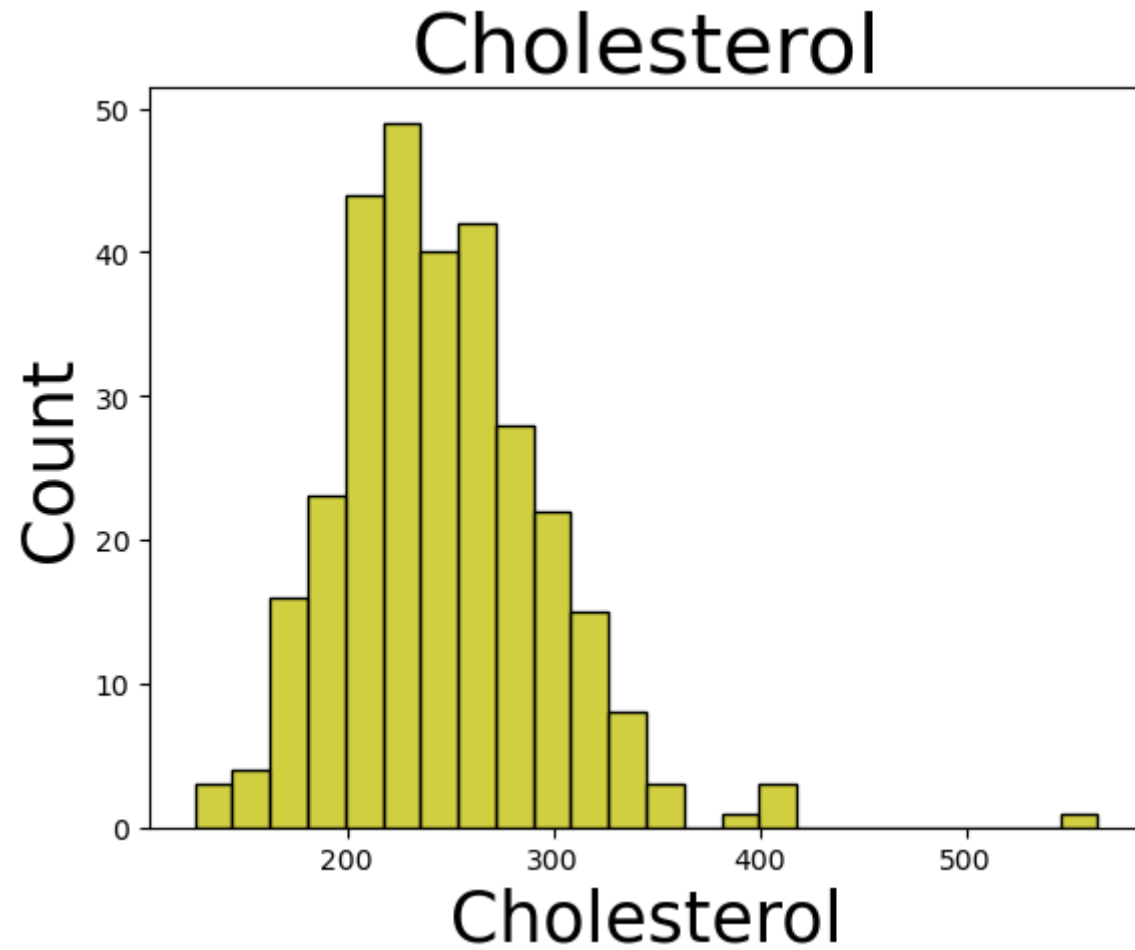
Age count by Heart Disease



In [27]: *# show distribution of serum cholesterol*

```
sns.histplot(x='chol',data=df,color='y')
plt.xlabel('Cholesterol',fontsize=25)
plt.ylabel('Count',fontsize=25)
plt.title('Cholesterol',fontsize=30)
plt.show
```

Out[27]: <function matplotlib.pyplot.show(close=None, block=None)>



Deepak Mengal

End Of The Project

Thank You...