

In [2]:

```
#PROJECT 2 [Heart Disease Diagnostic Analysis]
```

In [11]:

```
import pandas as pd
import numpy as np
heart_data=pd.read_csv(r"C:\Users\user\Documents\Unified Mentor Projects\Heart D
heart_data
```

Out[11]:

|      | age | sex | cp  | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca  |
|------|-----|-----|-----|----------|------|-----|---------|---------|-------|---------|-------|-----|
| 0    | 52  | 1   | 0   | 125      | 212  | 0   | 1       | 168     | 0     | 1.0     | 2     | 2   |
| 1    | 53  | 1   | 0   | 140      | 203  | 1   | 0       | 155     | 1     | 3.1     | 0     | 0   |
| 2    | 70  | 1   | 0   | 145      | 174  | 0   | 1       | 125     | 1     | 2.6     | 0     | 0   |
| 3    | 61  | 1   | 0   | 148      | 203  | 0   | 1       | 161     | 0     | 0.0     | 2     | 1   |
| 4    | 62  | 0   | 0   | 138      | 294  | 1   | 1       | 106     | 0     | 1.9     | 1     | 3   |
| ...  | ... | ... | ... | ...      | ...  | ... | ...     | ...     | ...   | ...     | ...   | ... |
| 1020 | 59  | 1   | 1   | 140      | 221  | 0   | 1       | 164     | 1     | 0.0     | 2     | 0   |
| 1021 | 60  | 1   | 0   | 125      | 258  | 0   | 0       | 141     | 1     | 2.8     | 1     | 1   |
| 1022 | 47  | 1   | 0   | 110      | 275  | 0   | 0       | 118     | 1     | 1.0     | 1     | 1   |
| 1023 | 50  | 0   | 0   | 110      | 254  | 0   | 0       | 159     | 0     | 0.0     | 2     | 0   |
| 1024 | 54  | 1   | 0   | 120      | 188  | 0   | 1       | 113     | 0     | 1.4     | 1     | 1   |

1025 rows × 14 columns

In [12]:

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

Out[12]:

```
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 [13]:

```
heart_data.notnull().sum()
```

```
Out[13]: age      1025
sex      1025
cp       1025
trestbps 1025
chol     1025
fbs      1025
restecg  1025
thalach  1025
exang    1025
oldpeak  1025
slope    1025
ca       1025
thal     1025
target   1025
dtype: int64
```

```
In [142]: heart_data.head()
```

Out[142]:

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

```
In [140]: heart_data.describe()
```

Out[140]:

|       | age         | sex         | cp          | trestbps    | chol        | fbs         |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
| count | 1025.000000 | 1025.000000 | 1025.000000 | 1025.000000 | 1025.000000 | 1025.000000 |
| mean  | 54.434146   | 0.695610    | 0.942439    | 131.611707  | 246.000000  | 0.149268    |
| std   | 9.072290    | 0.460373    | 1.029641    | 17.516718   | 51.59251    | 0.356527    |
| min   | 29.000000   | 0.000000    | 0.000000    | 94.000000   | 126.000000  | 0.000000    |
| 25%   | 48.000000   | 0.000000    | 0.000000    | 120.000000  | 211.000000  | 0.000000    |
| 50%   | 56.000000   | 1.000000    | 1.000000    | 130.000000  | 240.000000  | 0.000000    |
| 75%   | 61.000000   | 1.000000    | 2.000000    | 140.000000  | 275.000000  | 0.000000    |
| max   | 77.000000   | 1.000000    | 3.000000    | 200.000000  | 564.000000  | 1.000000    |

```
In [141]: heart_data.corr() #Correlation
```

Out[141]...

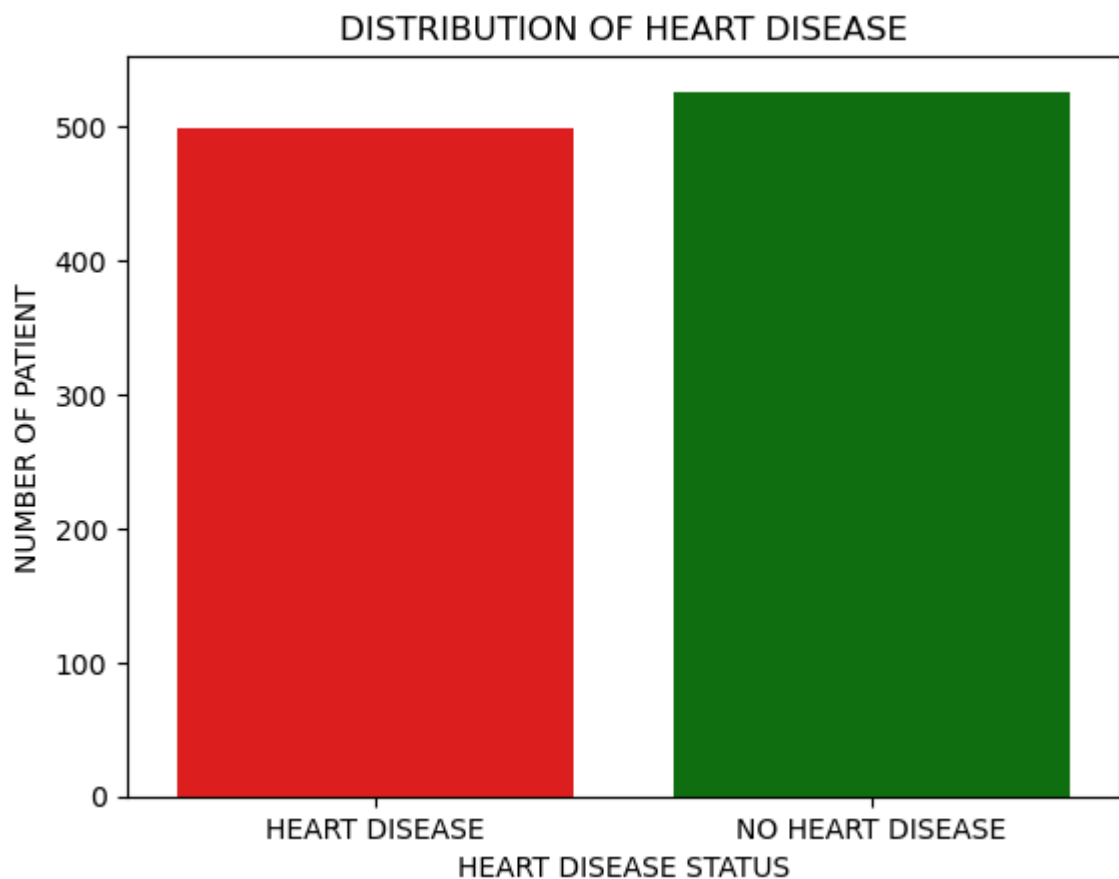
|                 | age       | sex       | cp        | trestbps  | chol      | fbs       | restecg   | t    |
|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------|
| <b>age</b>      | 1.000000  | -0.103240 | -0.071966 | 0.271121  | 0.219823  | 0.121243  | -0.132696 | -0.0 |
| <b>sex</b>      | -0.103240 | 1.000000  | -0.041119 | -0.078974 | -0.198258 | 0.027200  | -0.055117 | -0.0 |
| <b>cp</b>       | -0.071966 | -0.041119 | 1.000000  | 0.038177  | -0.081641 | 0.079294  | 0.043581  | 0.0  |
| <b>trestbps</b> | 0.271121  | -0.078974 | 0.038177  | 1.000000  | 0.127977  | 0.181767  | -0.123794 | -0.0 |
| <b>chol</b>     | 0.219823  | -0.198258 | -0.081641 | 0.127977  | 1.000000  | 0.026917  | -0.147410 | -0.0 |
| <b>fbs</b>      | 0.121243  | 0.027200  | 0.079294  | 0.181767  | 0.026917  | 1.000000  | -0.104051 | -0.0 |
| <b>restecg</b>  | -0.132696 | -0.055117 | 0.043581  | -0.123794 | -0.147410 | -0.104051 | 1.000000  | 0.0  |
| <b>thalach</b>  | -0.390227 | -0.049365 | 0.306839  | -0.039264 | -0.021772 | -0.008866 | 0.048411  | 1.0  |
| <b>exang</b>    | 0.088163  | 0.139157  | -0.401513 | 0.061197  | 0.067382  | 0.049261  | -0.065606 | -0.0 |
| <b>oldpeak</b>  | 0.208137  | 0.084687  | -0.174733 | 0.187434  | 0.064880  | 0.010859  | -0.050114 | -0.0 |
| <b>slope</b>    | -0.169105 | -0.026666 | 0.131633  | -0.120445 | -0.014248 | -0.061902 | 0.086086  | 0.0  |
| <b>ca</b>       | 0.271551  | 0.111729  | -0.176206 | 0.104554  | 0.074259  | 0.137156  | -0.078072 | -0.0 |
| <b>thal</b>     | 0.072297  | 0.198424  | -0.163341 | 0.059276  | 0.100244  | -0.042177 | -0.020504 | -0.0 |
| <b>target</b>   | -0.229324 | -0.279501 | 0.434854  | -0.138772 | -0.099966 | -0.041164 | 0.134468  | 0.0  |



```
In [23]: import matplotlib.pyplot as plt
import seaborn as sns
```

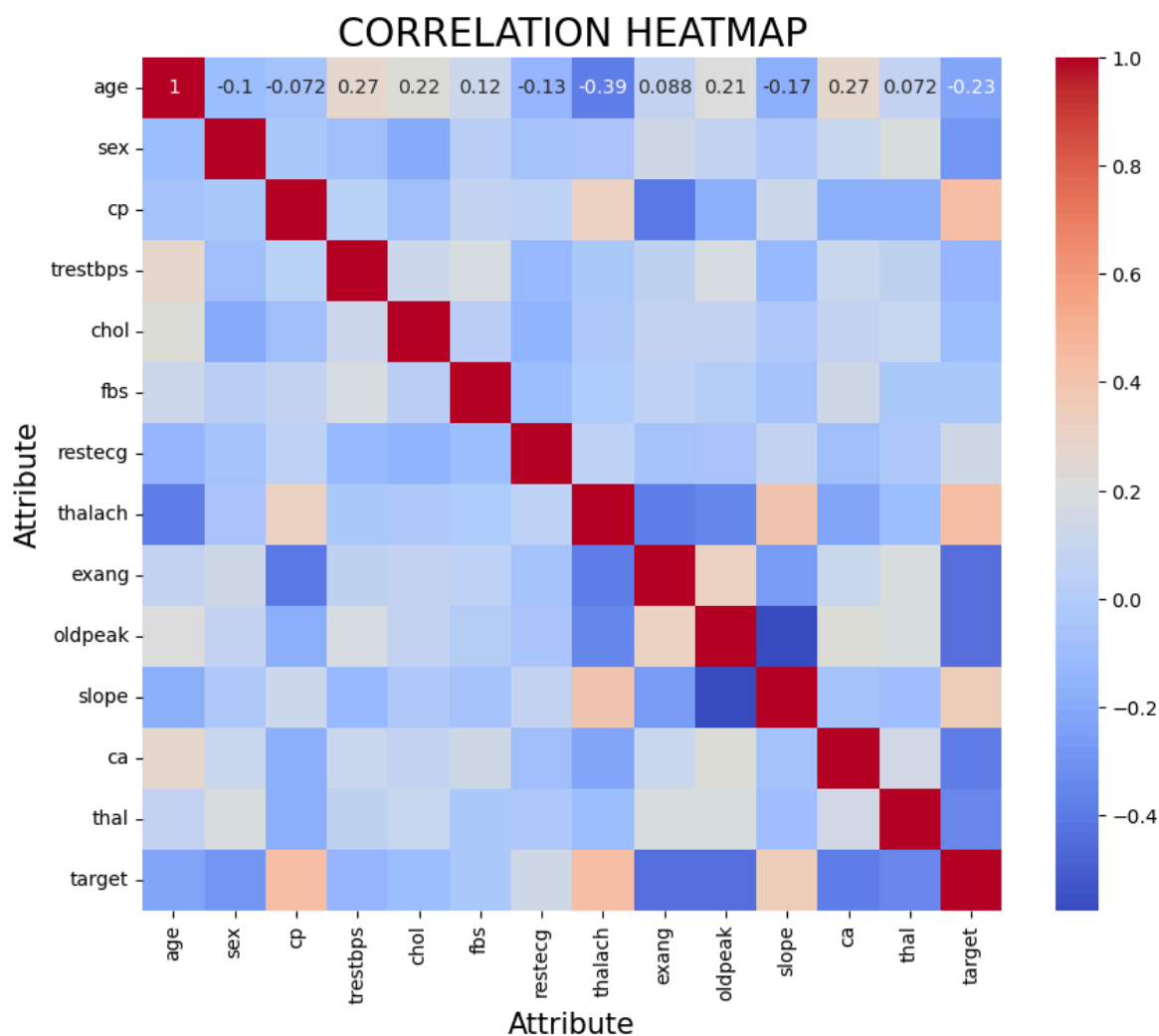
```
In [ ]: #ANALYSING THE KEY MATRICS AND RELATIONSHIPS
#Target Variable distribution [Heart disease present]
```

```
In [30]: sns.countplot(x='target',data=heart_data, hue_order=[1,0],palette=['red','green']
labels=["HEART DISEASE", "NO HEART DISEASE"]
plt.xlabel('HEART DISEASE STATUS')
plt.ylabel('NUMBER OF PATIENT')
plt.xticks(ticks=[0,1], labels=labels)
plt.title('DISTRIBUTION OF HEART DISEASE')
plt.show()
```

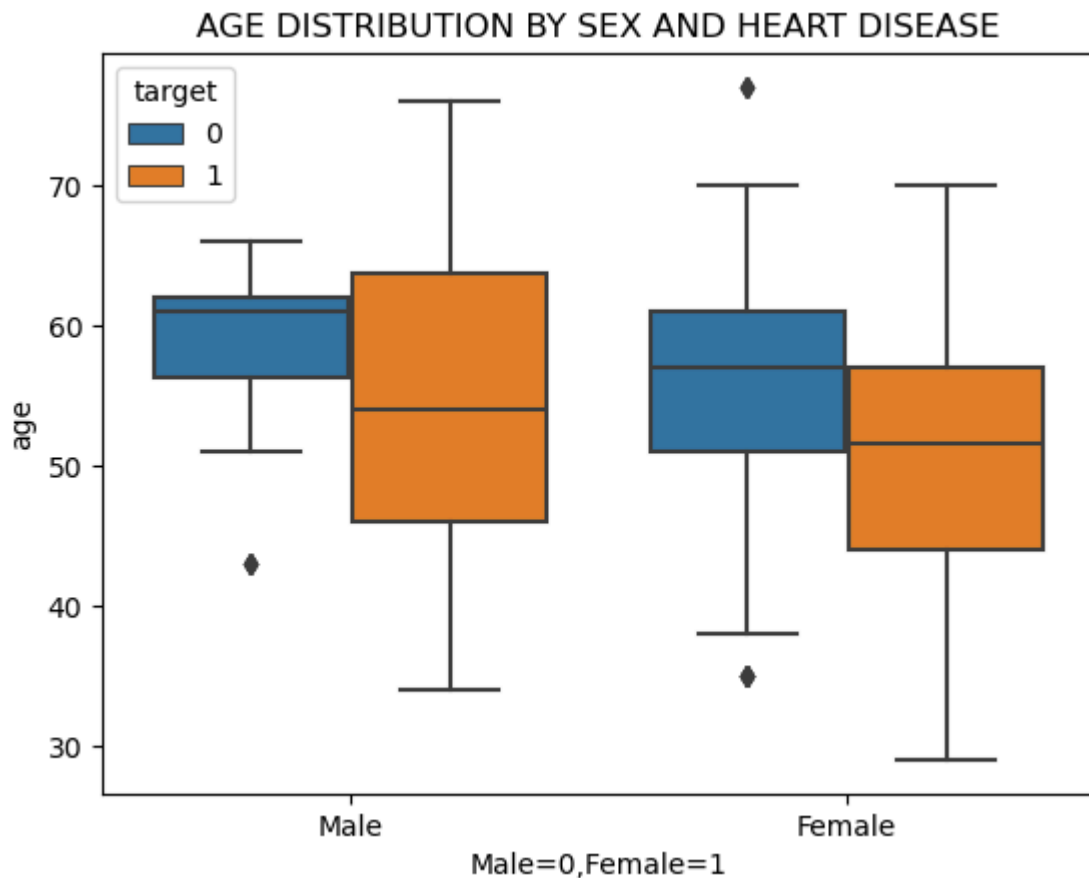


In [168...

```
#Created the digram for better understanding
import matplotlib.pyplot as plt
import seaborn as sns
#Calculate correlation matrix
plt.figure(figsize=(10,8))
sns.heatmap(heart_data.corr(), annot=True, cmap='coolwarm')
plt.xlabel('Attribute',fontsize=15)
plt.ylabel('Attribute',fontsize=15)
plt.title('CORRELATION HEATMAP',fontsize=20)
plt.show()
```



```
In [59]: #AGE DISTRIBUTION BY SEX AND HEART DISEASE
sns.boxplot(x='sex',y='age',data=heart_data, hue='target')
plt.xlabel('Male=0,Female=1')
plt.xticks([0,1],['Male','Female'])
plt.title('AGE DISTRIBUTION BY SEX AND HEART DISEASE')
plt.show()
```



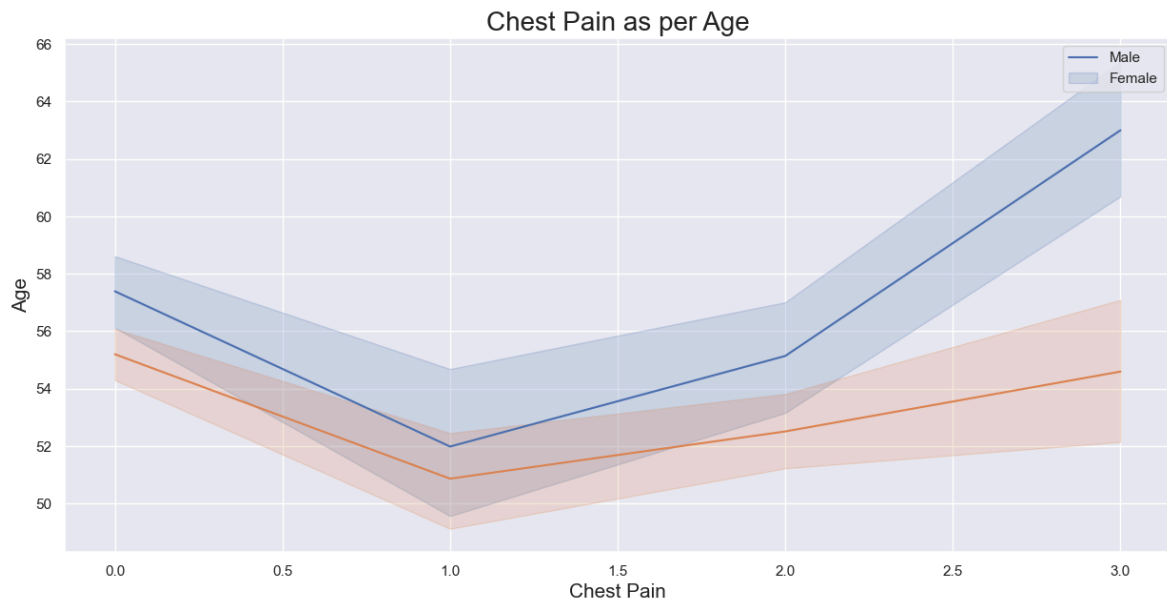
```
In [222... sns.lineplot(x='cp', data=heart_data, y='age', hue='sex')
plt.legend(labels=['Male', 'Female'])
sns.set(rc={'figure.figsize':(15,5)})
plt.xlabel('Chest Pain',fontsize=15)
plt.ylabel('Age',fontsize=15)
plt.title('Chest Pain as per Age',fontsize=20)
plt.show()
```

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\\_oldcore.py:1119: FutureWarning: use\_inf\_as\_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

with pd.option\_context('mode.use\_inf\_as\_na', True):

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\\_oldcore.py:1119: FutureWarning: use\_inf\_as\_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

with pd.option\_context('mode.use\_inf\_as\_na', True):



```
In [88]: #Age Distribution
sns.distplot(heart_data['age'])
plt.xlabel('Age')
plt.ylabel('Destiny')
plt.title('AGE DISTRIBUTION')
plt.show()
```

C:\Users\user\AppData\Local\Temp\ipykernel\_10344\2845624387.py:2: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

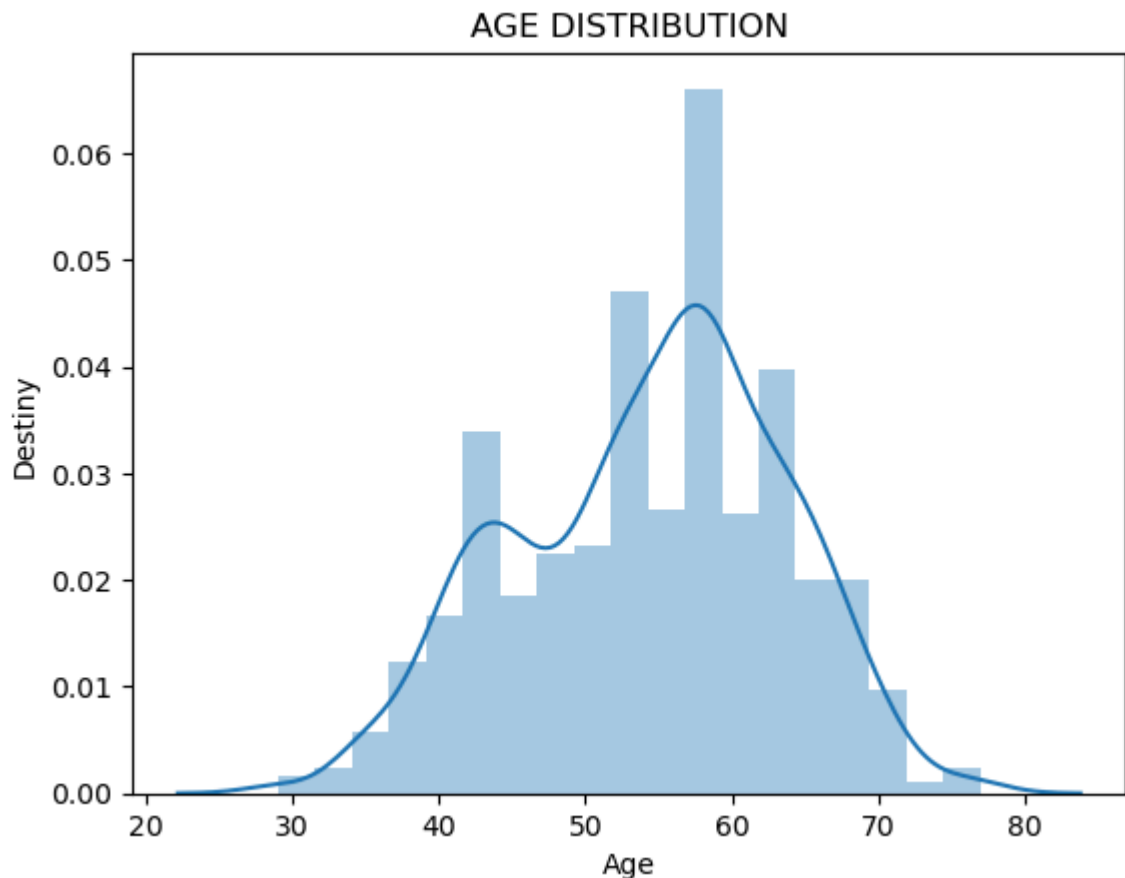
Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

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

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\\_oldcore.py:1119: FutureWarning: use\_inf\_as\_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```



```
In [89]: sns.distplot(heart_data['ca'])
plt.xlabel('Number of coloured vessel(0-3)')
plt.ylabel('Density')
plt.title('Number of coloured vessels Distribution')
plt.show()
```

C:\Users\user\AppData\Local\Temp\ipykernel\_10344\2679143531.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

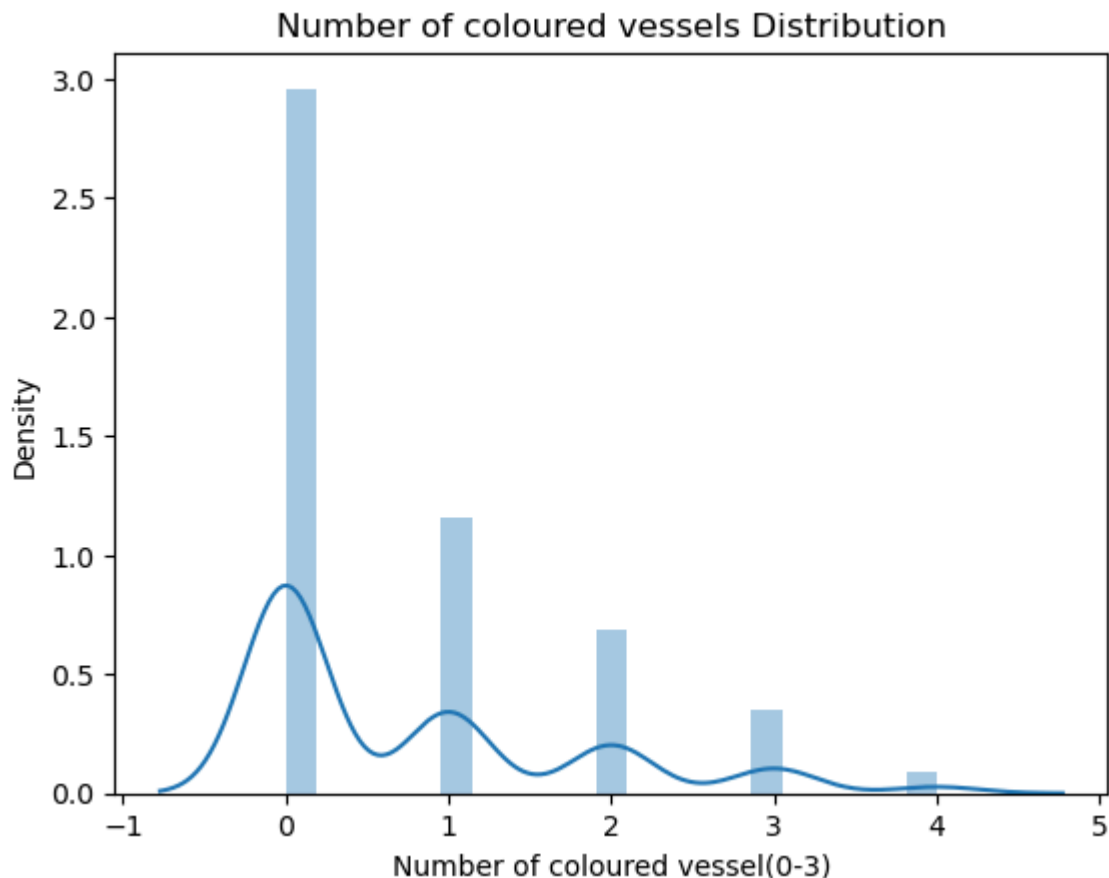
For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(heart_data['ca'])
```

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\\_oldcore.py:1119: FutureWarning: use\_inf\_as\_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

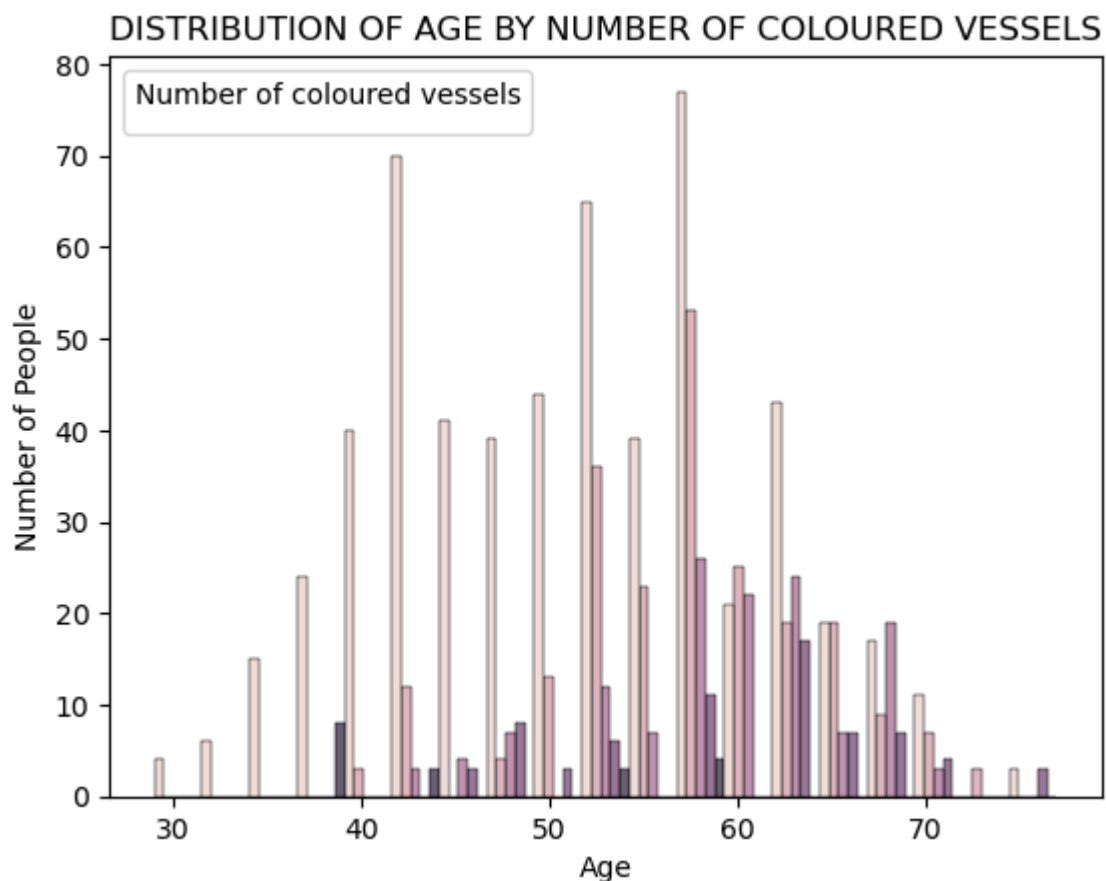




```
In [93]: #Distribution of age by the number of coloured vessels
sns.histplot(x='age', hue='ca', data=heart_data, multiple='dodge', alpha=.7)
plt.xlabel('Age')
plt.ylabel('Number of People')
plt.title('DISTRIBUTION OF AGE BY NUMBER OF COLOURED VESSELS')
plt.legend(title='Number of coloured vessels')
plt.show()
```

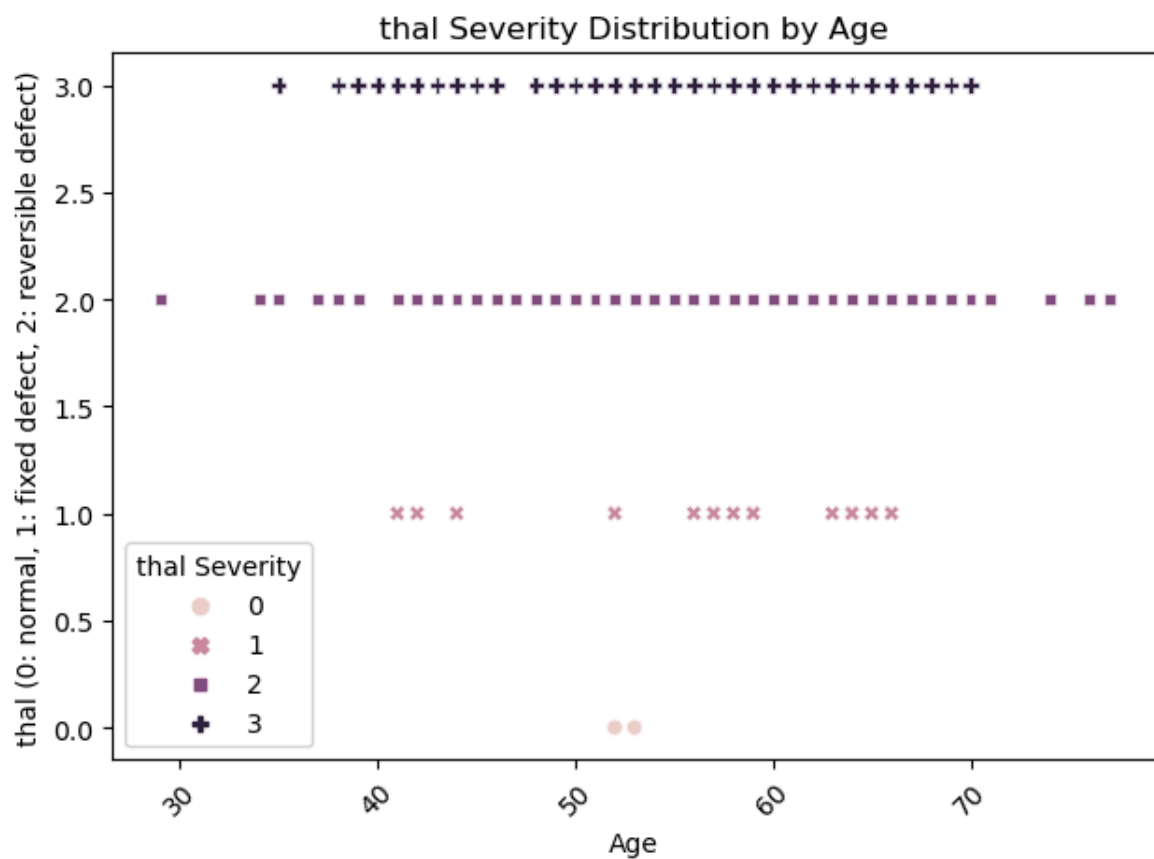
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\\_oldcore.py:1119: FutureWarning: use\_inf\_as\_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
No artists with labels found to put in legend. Note that artists whose label start with an underscore are ignored when legend() is called with no argument.
```



In [104...

```
sns.scatterplot(x='age', y='thal', hue='thal', data=heart_data, style='thal')
plt.xlabel('Age')
plt.ylabel('thal (0: normal, 1: fixed defect, 2: reversible defect)')
plt.title('thal Severity Distribution by Age')
plt.xticks(rotation=45)
plt.legend(title='thal Severity')
plt.tight_layout()
plt.show()
```

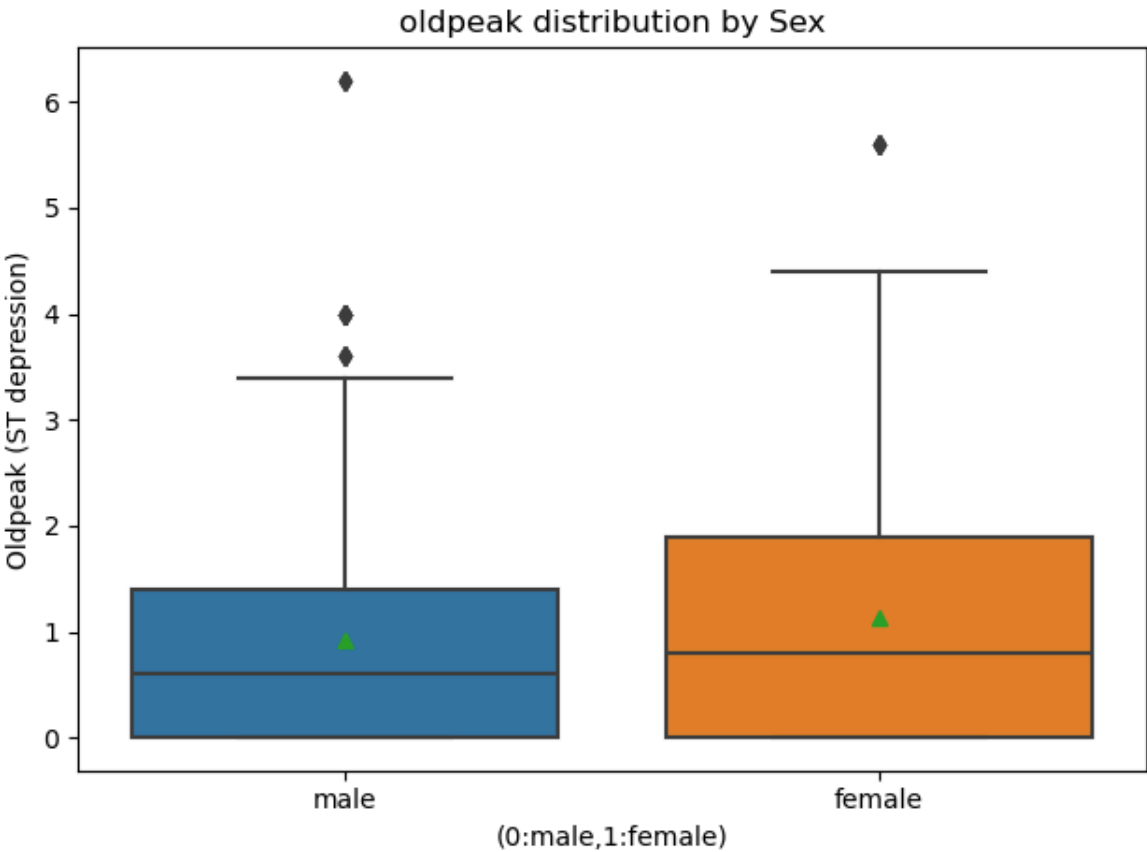


In [111...

```

sns.boxplot(x='sex',y='oldpeak',showmeans=True,data=heart_data)
plt.xlabel('(0:male,1:female)')
plt.ylabel('Oldpeak (ST depression)')
plt.title('oldpeak distribution by Sex')
plt.xticks([0,1],['male','female'])
plt.tight_layout()
plt.show()

```



```
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
```

In [ ]:

In [ ]:

In [ ]: