

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

```
[3]: data = pd.read_csv('C:\\Users\\SRINIVASA SESHADRI K\\OneDrive\\Documents\\MainFlow Services and Technologies Internship\\Task 5 (Heart Disease Analysis
```

```
[4]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   age         1025 non-null   int64
1   sex         1025 non-null   int64
2   cp          1025 non-null   int64
3   trestbps    1025 non-null   int64
4   chol        1025 non-null   int64
5   fbs         1025 non-null   int64
6   restecg     1025 non-null   int64
7   thalach     1025 non-null   int64
8   exang       1025 non-null   int64
9   oldpeak     1025 non-null   float64
10  slope       1025 non-null   int64
11  ca          1025 non-null   int64
12  thal        1025 non-null   int64
13  target      1025 non-null   int64
dtypes: float64(1), int64(13)
memory usage: 112.2 KB
```

```
[5]: data.shape
```

```
[5]: (1025, 14)
```

```
[6]: data.head(6)
```

	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
5	58	0	0	100	248	0	0	122	0	1.0	1	0	2	1

```
[7]: data.describe()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1
mean	54.434146	0.695610	0.942439	131.611707	246.000000	0.149268	0.529756	149.114146	0.336585	1.071512	1.385366	0.754146	2.323902	
std	9.072290	0.460373	1.029641	17.516718	51.59251	0.356527	0.527878	23.005724	0.472772	1.175053	0.617755	1.030798	0.620660	
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	48.000000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	132.000000	0.000000	0.000000	1.000000	0.000000	2.000000	
50%	56.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	152.000000	0.000000	0.800000	1.000000	0.000000	2.000000	
75%	61.000000	1.000000	2.000000	140.000000	275.000000	0.000000	1.000000	166.000000	1.000000	1.800000	2.000000	1.000000	3.000000	
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	

```
data.tail()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

```
data.columns.values
```

```
array(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg',  
      'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],  
      dtype=object)
```

```
data.isnull()
```

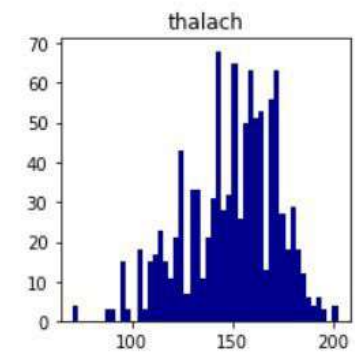
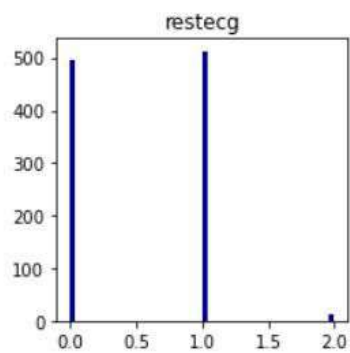
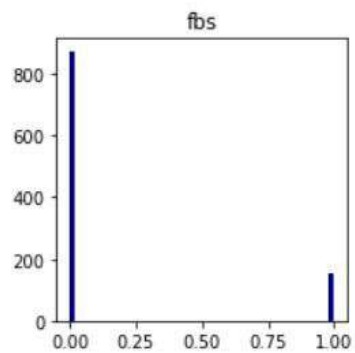
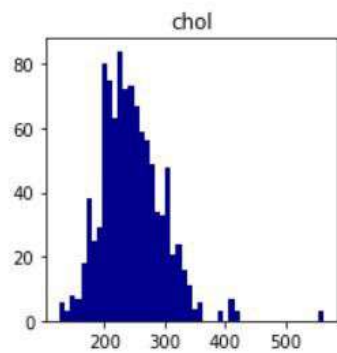
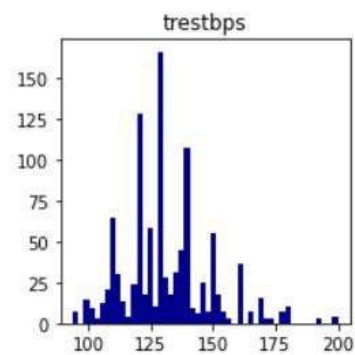
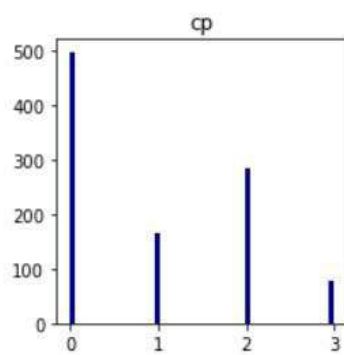
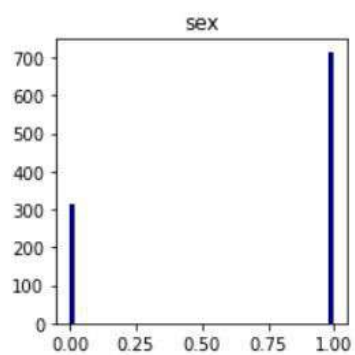
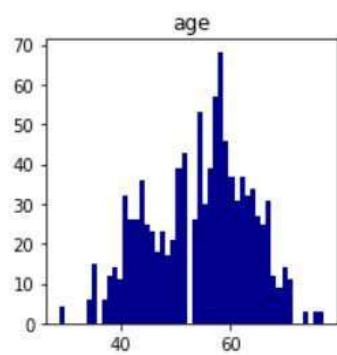
[illegible]

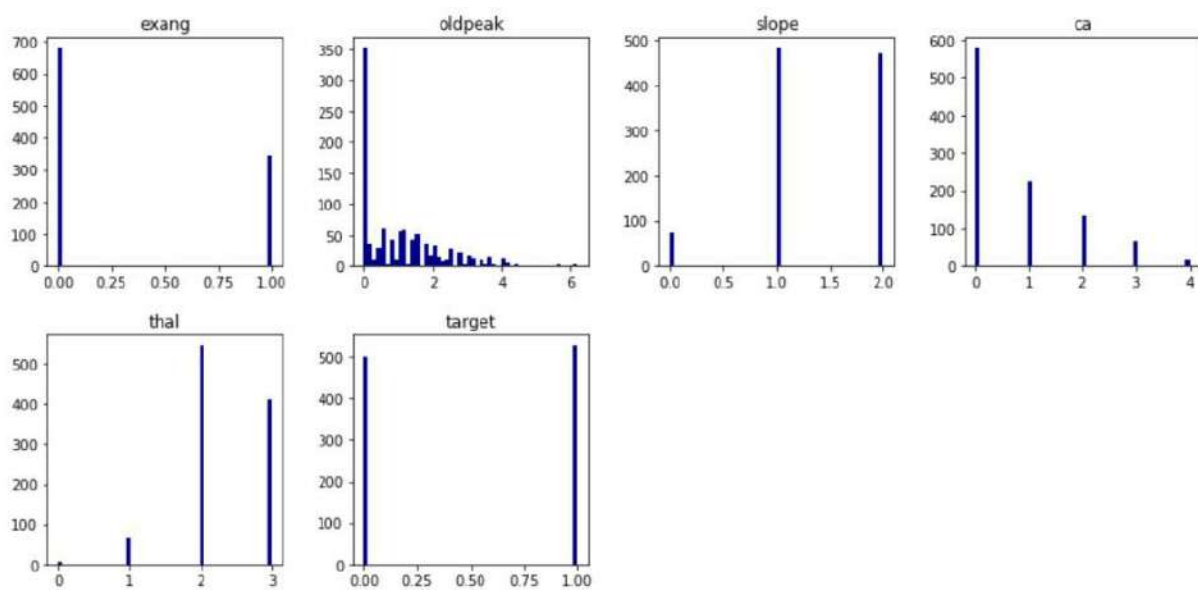
```
[11]: data.isnull().sum()
```

```
[11]: 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
```

```
[12]: import matplotlib.pyplot as plt  
import seaborn as sns  
%matplotlib inline
```

```
[13]: data.hist(bins=50, grid=False, color='darkblue', figsize=(20,15));
```



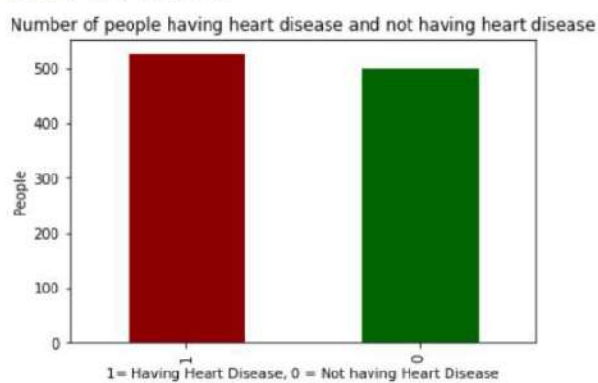


```
[14]: #Number of people having heart disease and not having heart disease
data.target.value_counts()
```

```
[14]: 1    526
      0    499
      Name: target, dtype: int64
```

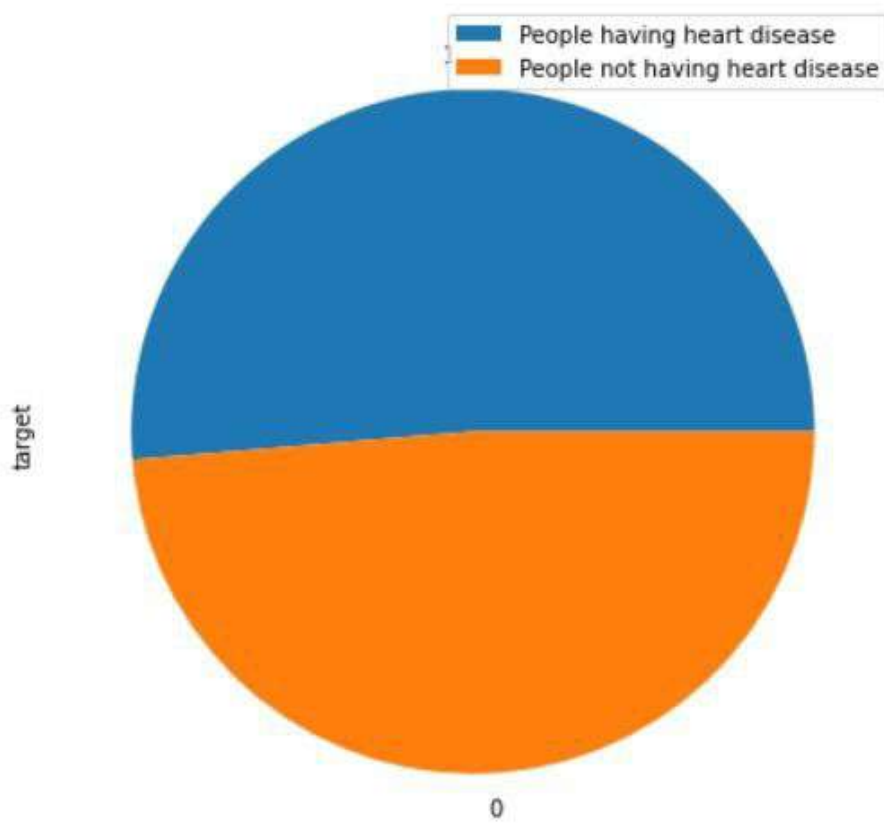
```
[15]: #plotting graph for people having heart disease and not having heart disease
data.target.value_counts().plot(kind='bar',color = ["darkred","darkgreen"])
plt.title("Number of people having heart disease and not having heart disease")
plt.xlabel("1= Having Heart Disease, 0 = Not having Heart Disease")
plt.ylabel("People")
```

```
[15]: Text(0, 0.5, 'People')
```



```
[22]: data.target.value_counts().plot(kind='pie', figsize=(10,8))
plt.legend(["People having heart disease", "People not having heart disease"])
```

[47]: <matplotlib.legend.Legend at 0x1eb6cc1b7f0>



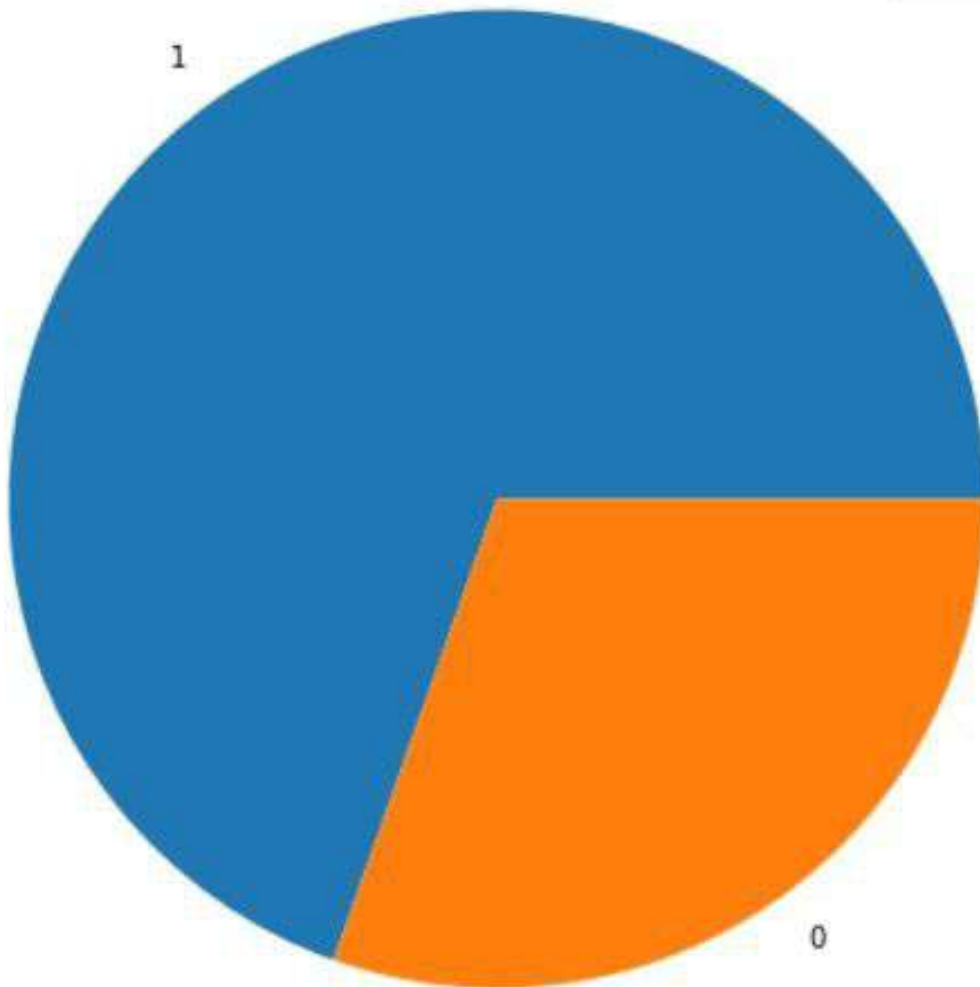

```
[23]: #Checking for number of male and female in the dataset
data.sex.value_counts()
```

```
[23]: 1    713
      0    312
      Name: sex, dtype: int64
```

```
[24]: #Visualization of Male Female Ratio
data.sex.value_counts().plot(kind='pie', figsize=(10,8))
plt.title("Male Female Ratio")
plt.legend(["Male", "Female"])
```

```
[24]: <matplotlib.legend.Legend at 0x1eb69324130>
```

Male Female Ratio



sex

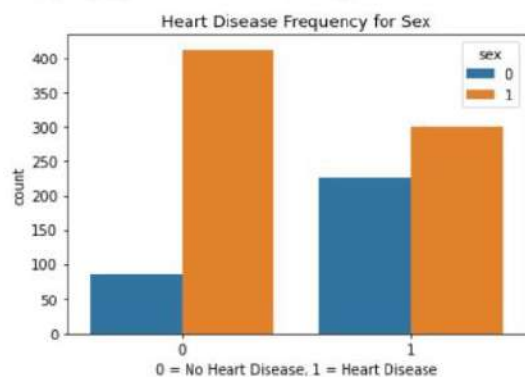
```
[25]: #Gender having most heart disease  
pd.crosstab(data.target, data.sex)
```

```
[25]:
```

sex	0	1
target		
0	86	413
1	226	300

```
[26]: sns.countplot(x='target', data = data, hue = 'sex')  
plt.title("Heart Disease Frequency for Sex")  
plt.xlabel("0 = No Heart Disease, 1 = Heart Disease")
```

```
[26]: Text(0.5, 0, '0 = No Heart Disease, 1 = Heart Disease')
```

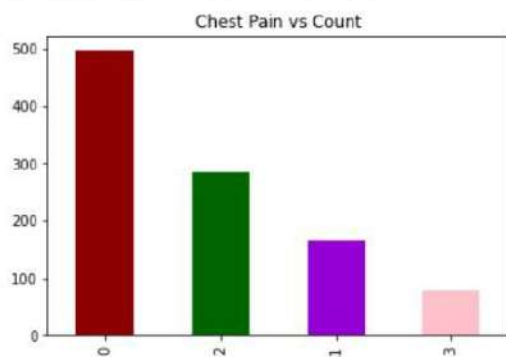


```
[27]: #Counting values for different chest pain  
data.cp.value_counts()
```

```
[27]: 0    497  
     2    284  
     1    167  
     3     77  
     Name: cp, dtype: int64
```

```
[29]: #plotting bar chart  
data.cp.value_counts().plot(kind='bar',color = ["darkred","darkgreen","darkviolet", "pink"])  
plt.title("Chest Pain vs Count")
```

```
[29]: Text(0.5, 1.0, 'Chest Pain vs Count')
```



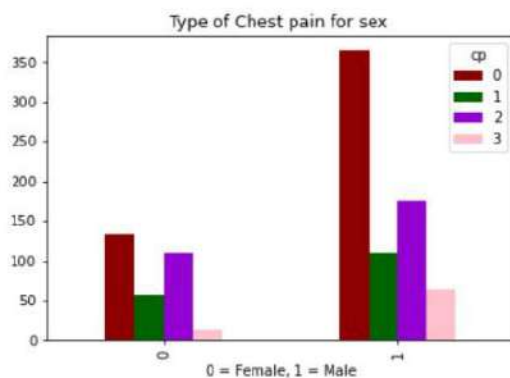
```
[30]: pd.crosstab(data.sex, data.cp)
```

```
[30]:
```

cp	0	1	2	3
sex				
0	133	57	109	13
1	364	110	175	64

```
[32]: pd.crosstab(data.sex, data.cp).plot(kind='bar', colors=['darkred', 'darkgreen', 'darkviolet', 'pink'])  
plt.title("Type of Chest pain for sex ")  
plt.xlabel("0 = Female, 1 = Male")
```

```
[32]: Text(0.5, 0, '0 = Female, 1 = Male')
```

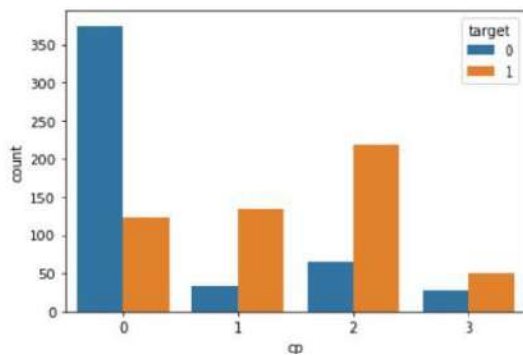


```
[33]: #People with which chest pain are most pron to have heart disease  
pd.crosstab(data.cp, data.target)
```

```
[33]: target    0    1  
cp  
0    375  122  
1     33  134  
2     65  219  
3     26   51
```

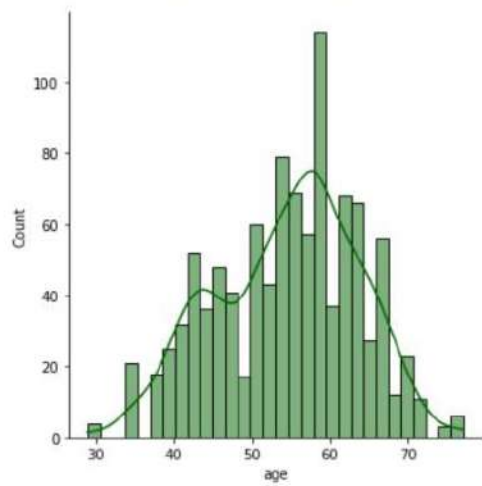
```
[37]: sns.countplot(x='cp', data=data, hue='target')
```

```
[37]: <matplotlib.axes._subplots.AxesSubplot at 0x1eb68f0e040>
```



```
[40]: sns.displot(x='age', data=data, bins=30, kde=True, color='darkgreen')
```

```
[40]: <seaborn.axisgrid.FacetGrid at 0x1eb66a25be0>
```



```
sns.displot(x='thalach', data=data, bins=30, kde=True, color='darkred')
```

```
[41]: sns.displot(x='thalach', data=data, bins=30, kde=True, color='darkred')
```

```
[41]: <seaborn.axisgrid.FacetGrid at 0x1eb66795a30>
```

