[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
df=pd.read_csv("heart.csv")
```

[3]:
```
df.head()
```

[3]:

	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

[4]:
```
df.tail()
```

[4]:

	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

[5]:
```
df.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
```

[6]:
```
df.describe()
```

[6]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.00000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	102
mean	54.434146	0.695610	0.942439	131.611707	246.00000	0.149268	0.529756	149.114146	0.336585	1.071512	1.385366	0.754146	
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	
min	29.000000	0.000000	0.000000	94.000000	126.00000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	
25%	48.000000	0.000000	0.000000	120.000000	211.00000	0.000000	0.000000	132.000000	0.000000	0.000000	1.000000	0.000000	
50%	56.000000	1.000000	1.000000	130.000000	240.00000	0.000000	1.000000	152.000000	0.000000	0.800000	1.000000	0.000000	
75%	61.000000	1.000000	2.000000	140.000000	275.00000	0.000000	1.000000	166.000000	1.000000	1.800000	2.000000	1.000000	
max	77.000000	1.000000	3.000000	200.000000	564.00000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	

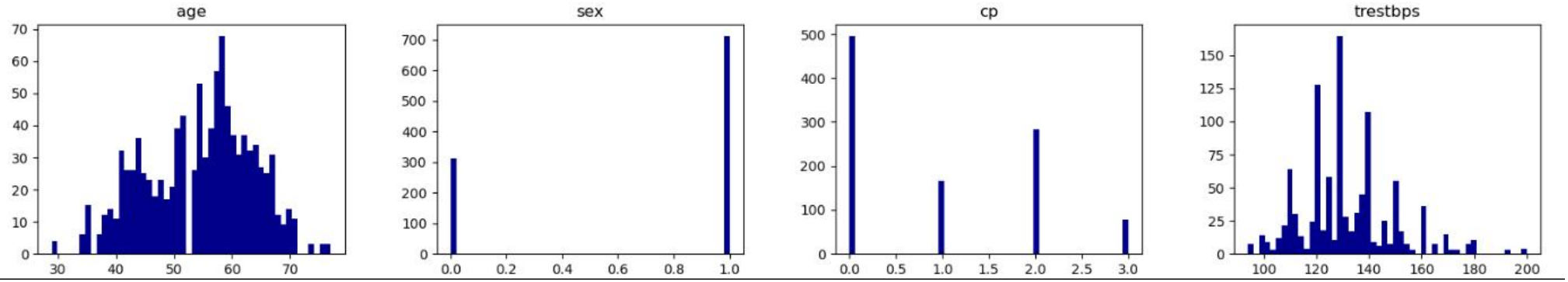
[7]:
```
df.columns.values
```

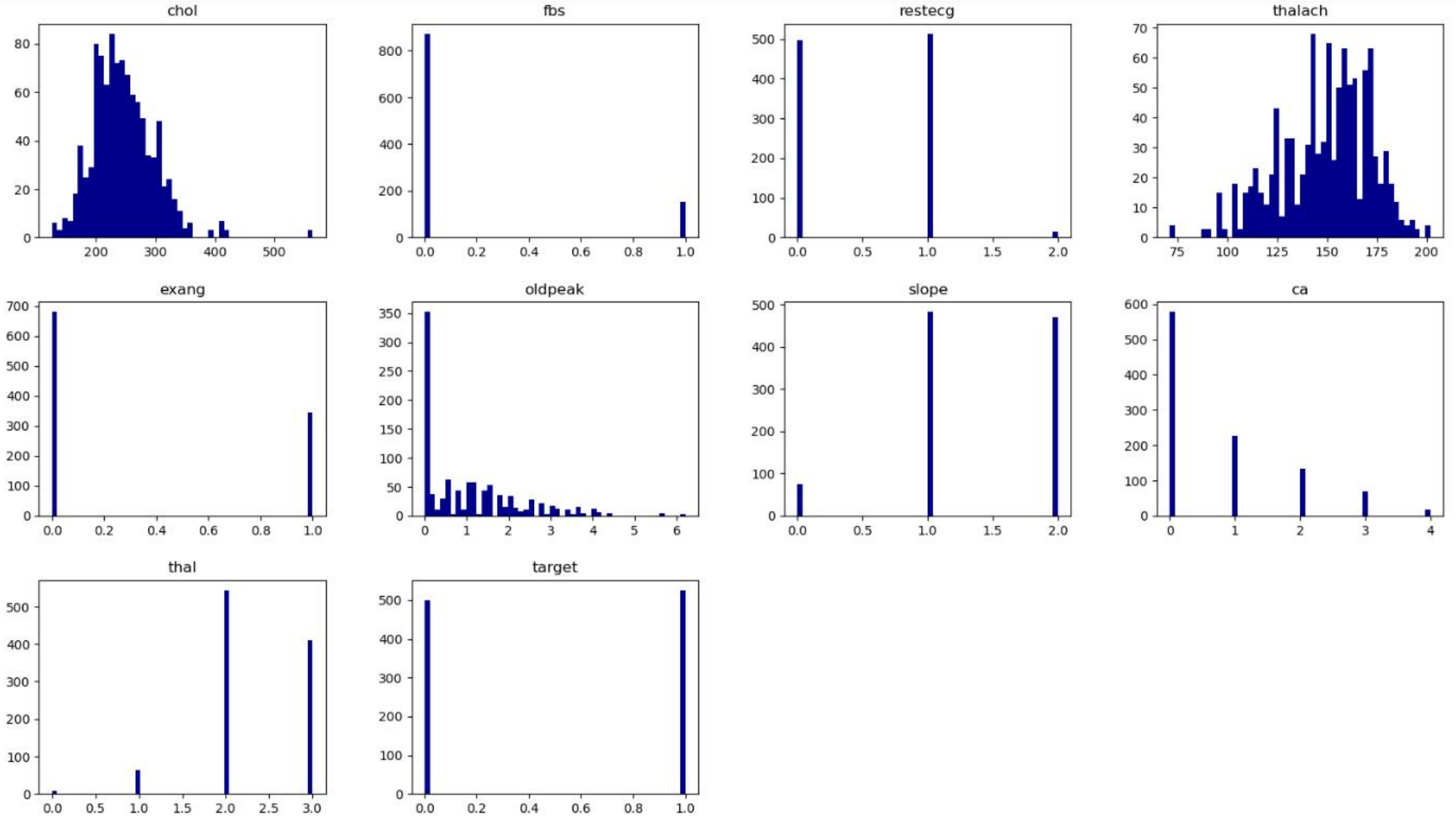
[7]:
```
array(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg',
       'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
      dtype=object)
```

[8]:
```
df.isnull().sum()
```

[8]:
```
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
```

[9]:
```
df.hist(bins=50, grid=False,color='darkblue',figsize=(20,15));
```




chol
fbs
restecg
thalach
exang
oldpeak
slope
ca
thal
target

[10]:
```
#nuber of people having heart disease and number of people having not heart disease
df.target.value_counts()
```

[10]:
```
target
1    526
0    499
Name: count, dtype: int64
```

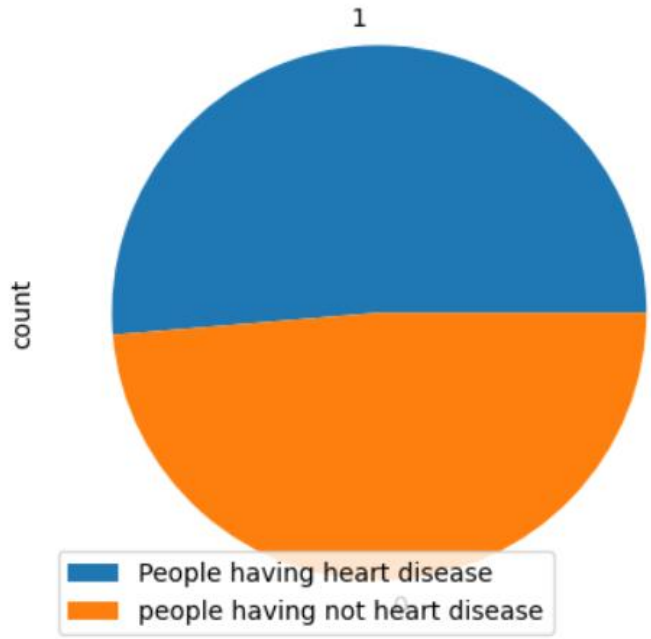
[11]:
```
df.target.value_counts().plot(kind='bar',color=['darkred','darkgreen'])
plt.title("Number of people having heart disease and not having heart disease")
plt.xlabel("1 = Having heart disease 0 = Not having heart disease")
plt.ylabel("people")
```

[11]:
```
Text(0, 0.5, 'people')
```


Number of people having heart disease and not having heart disease
people
500
400
300
200
100
0
1
0
1 = Having heart disease 0 = Not having heart disease

[12]:
```
df.target.value_counts().plot(kind='pie',color=['darkred','darkgreen'])
plt.legend(["People having heart disease" ,"people having not heart disease"])
```

[12]:
```
<matplotlib.legend.Legend at 0x282fa2d6c60>
```

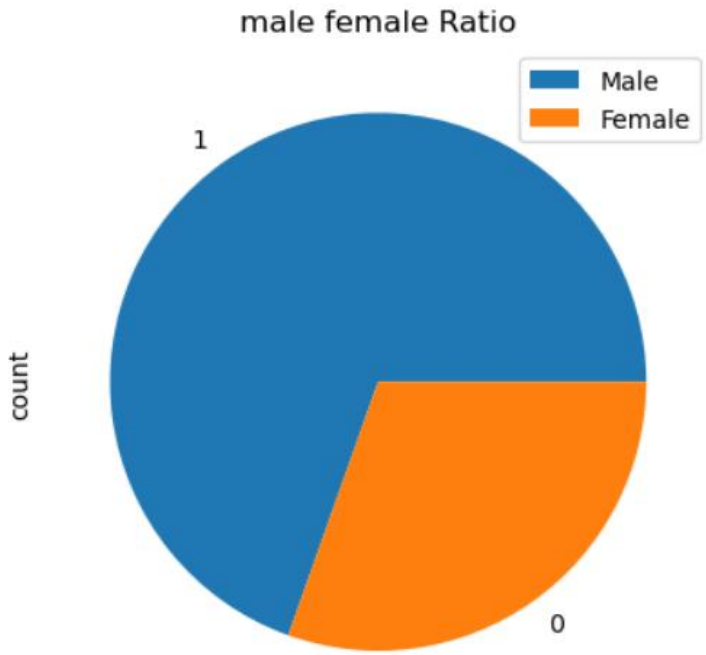


[13]:
```
#checking number of male and number of females
df.sex.value_counts()
```

[13]:
```
sex
1    713
0    312
Name: count, dtype: int64
```

[14]:
```
df.sex.value_counts().plot(kind='pie',color=['darkred','darkgreen'])
plt.title("male female Ratio")
plt.legend(["Male","Female"])
```

[14]:
```
<matplotlib.legend.Legend at 0x282f9437bc0>
```



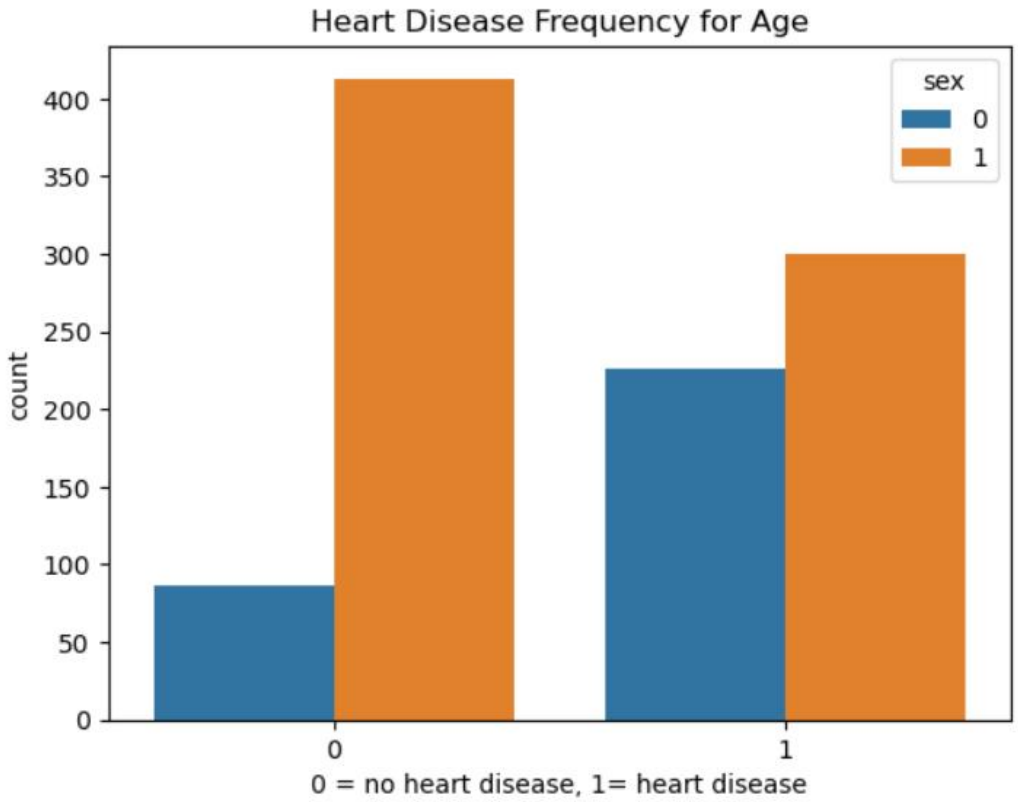
[15]:
```
#gender having most heart disease
pd.crosstab(df.target,df.sex)
```

[15]:

sex	0	1
target		
0	86	413
1	226	300

[16]:
```
sns.countplot(x='target' , data=df,hue='sex')
plt.title("Heart Disease Frequency for Age")
plt.xlabel("0 = no heart disease, 1= heart disease")
```

[16]:
```
Text(0.5, 0, '0 = no heart disease, 1= heart disease')
```

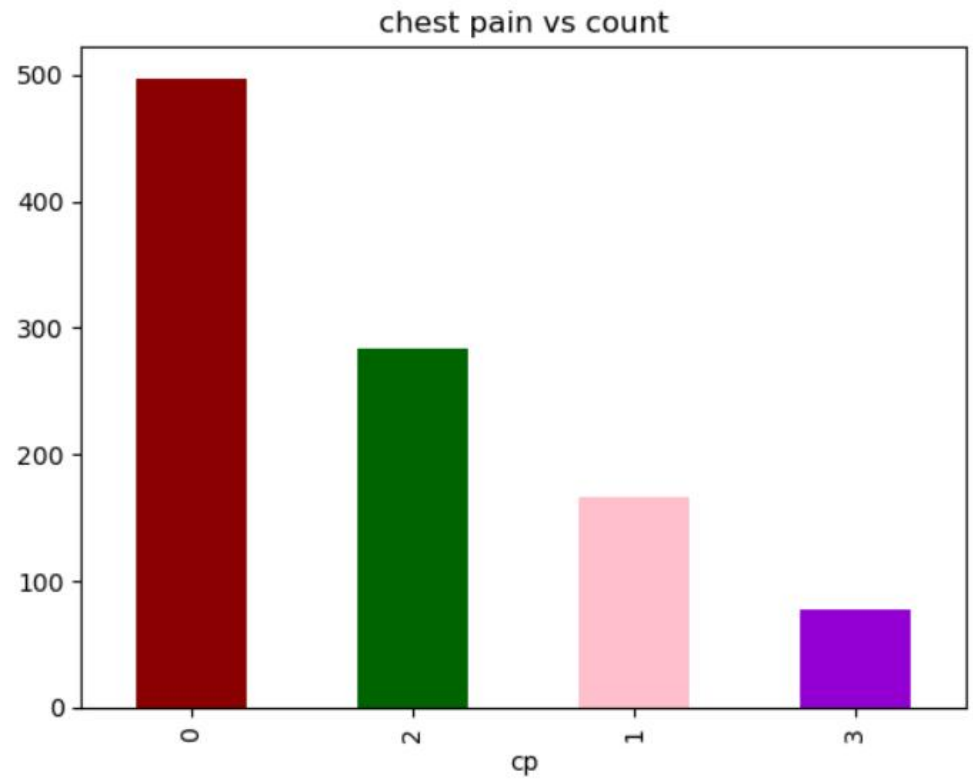


[17]:
```
#count values for different chest pains
df.cp.value_counts()
```

[17]:
```
cp
0    497
2    284
1    167
3     77
Name: count, dtype: int64
```

[18]:
```
df.cp.value_counts().plot(kind='bar',color=['darkred','darkgreen','pink','darkviolet'])
plt.title('chest pain vs count')
```

[18]: Text(0.5, 1.0, 'chest pain vs count')



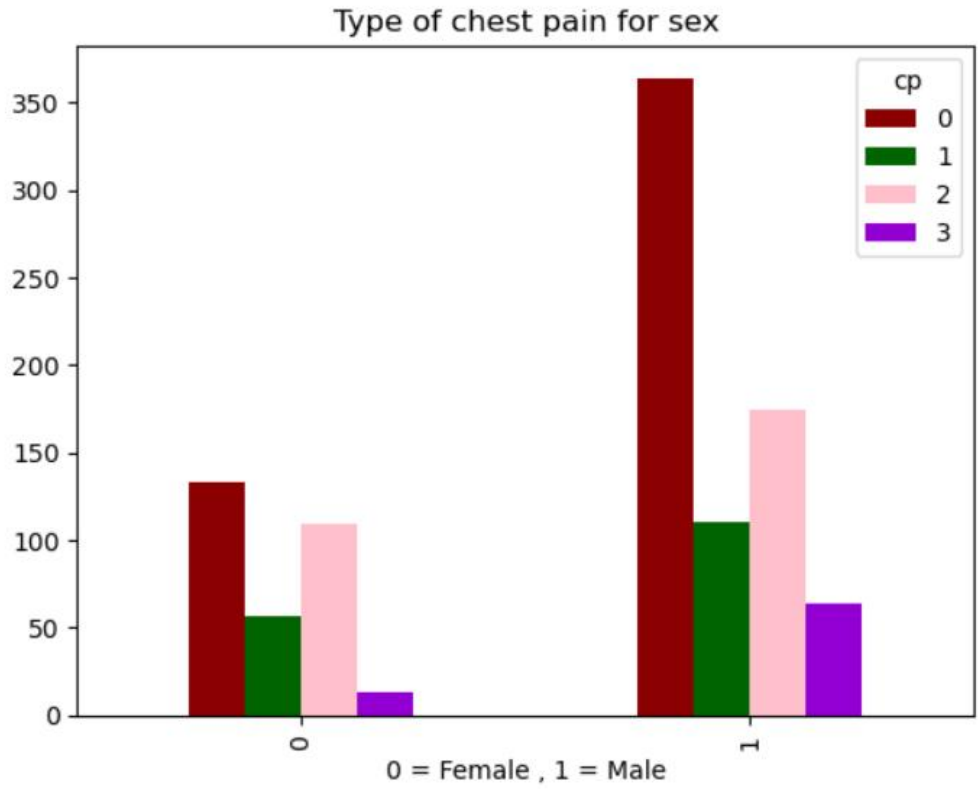
[19]:
```
#gender having most chest pains
pd.crosstab(df.sex,df.cp)
```

[19]:

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

[20]:
```
pd.crosstab(df.sex,df.cp).plot(kind='bar',color=['darkred','darkgreen','pink','darkviolet'])
plt.title('Type of chest pain for sex')
plt.xlabel('0 = Female , 1 = Male')
```

[20]:
```
Text(0.5, 0, '0 = Female , 1 = Male')
```



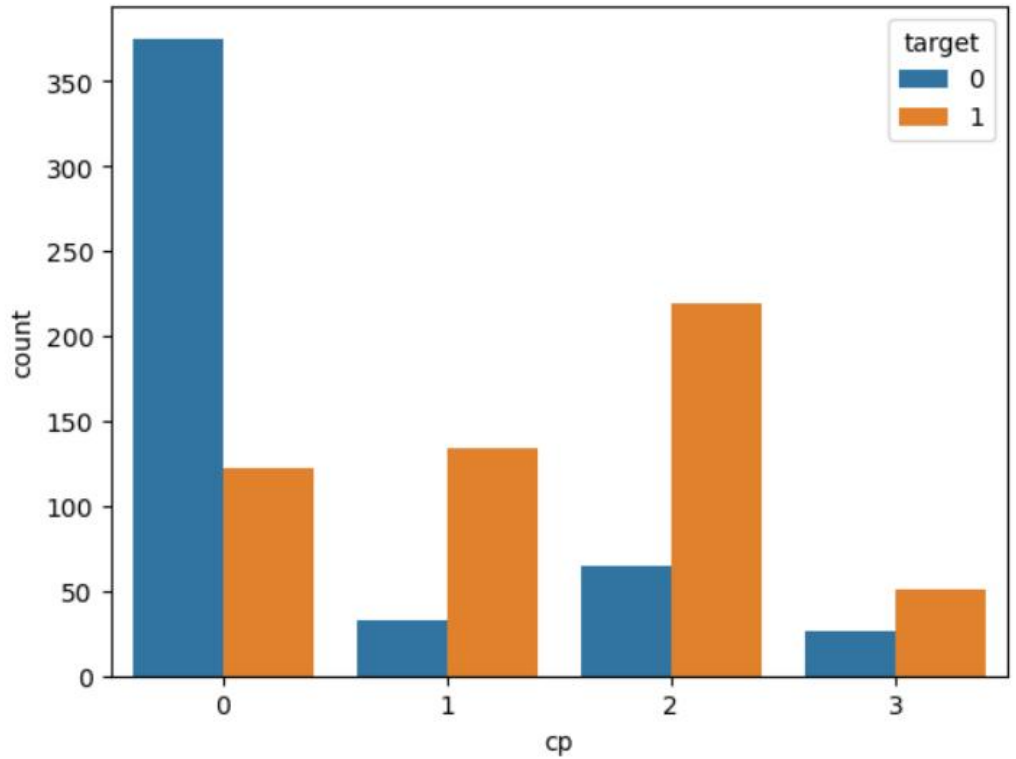
[21]:
```
#people with which chest pain have pron to have heart disease
pd.crosstab(df.cp,df.target)
```

[21]:

target	0	1
cp		
0	375	122
1	33	134
2	65	219
3	26	51

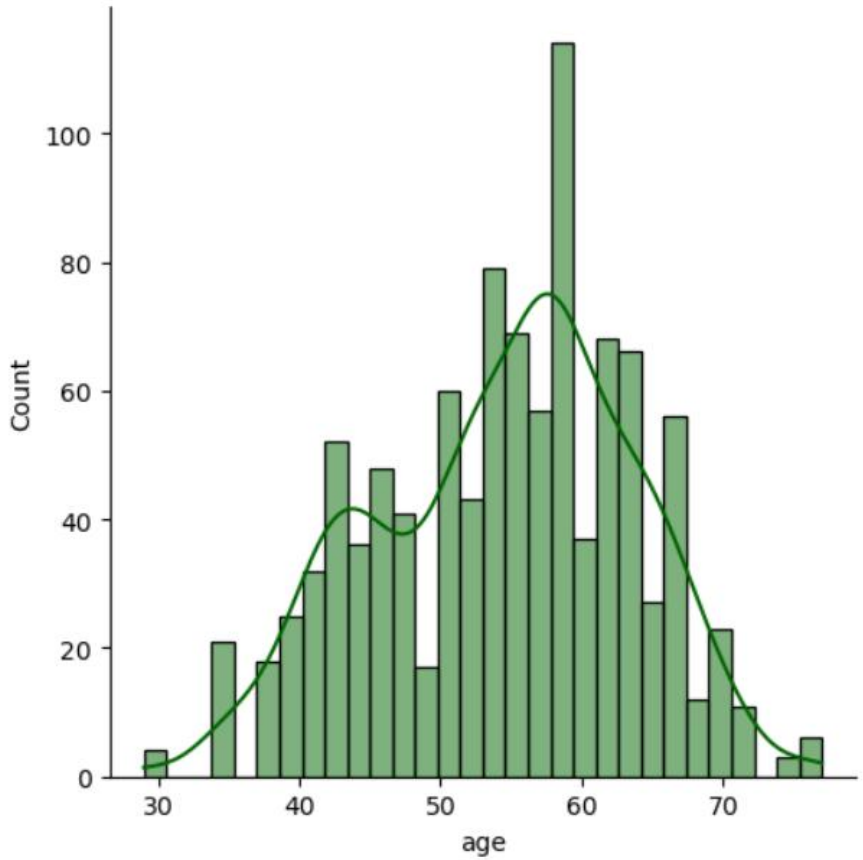
[22]:
```
sns.countplot(x='cp' , data=df ,hue='target')
```

[22]: <Axes: xlabel='cp', ylabel='count'>



[23]:
```
sns.displot(x='age',data=df,bins=30,kde=True,color='darkgreen')
```

[23]:
```
<seaborn.axisgrid.FacetGrid at 0x282f9b4e180>
```



[24]:
```
sns.displot(x='thalach',data=df,bins=30,kde=True,color='darkred')
```

[24]:
```
<seaborn.axisgrid.FacetGrid at 0x282f9b375c0>
```