Heart Disease Diagnostic Analysis

```
#Importing Libraries

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# sns.set_style('whitegrid')

#Extracting CSV Dataset From System Using Pandas Library

data=pd.read_csv(r"C:\Users\pahuj\Downloads\Heart Disease data\Heart
Disease data\Heart Disease data.csv",encoding= 'unicode_escape')
data
```

```
      age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang
oldpeak  \
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
...   ...  ...  ..       ...   ...  ...      ...      ...    ...
...
1020   59    1   1       140   221    0        1      164      1
0.0
1021   60    1   0       125   258    0        0      141      1
2.8
1022   47    1   0       110   275    0        0      118      1
1.0
1023   50    0   0       110   254    0        0      159      0
0.0
1024   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
...     ...  ..   ...     ...
1020      2   0     2       1
1021      1   1     3       0
```

```
1022      1   1     2       0
1023      2   0     2       1
1024      1   1     3       0

[1025 rows x 14 columns]
```

```
#Features/Columns of dataset 'data'

data.columns
```

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

There are thirteen features in Dataset

age: The person's age in years

sex: The person's sex (0 = male, 1 = female)

cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)

trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)

chol: The person's cholesterol measurement in mg/dl

fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)

restecg: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)

thalach: The person's maximum heart rate achieved

exang: Exercise induced angina (1 = yes; 0 = no)

oldpeak: ST depression induced by exercise relative to rest

slope: the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)

ca: The number of major vessels (0-3)

thal: A blood disorder called thalassemia (0 = normal; 1 = fixed defect; 2 = reversable defect)

target: Heart disease (0 = no, 1 = yes)

```
#checking null values

data.isnull().sum()
```

```
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
```

There are No Null/Missing Values in the dataset.

Percentage of People Having Heart Disease

```
num=data.groupby('target').size()
num
```

```
target
0    499
1    526
dtype: int64
```

```
#changing numerical data into categorical form and applying it to new
column
def heart_disease(row):
    if row==0:
        return 'Absence'
    elif row==1:
        return 'Presence'

data['heart_disease']=data['target'].apply(heart_disease)
data.head()
```

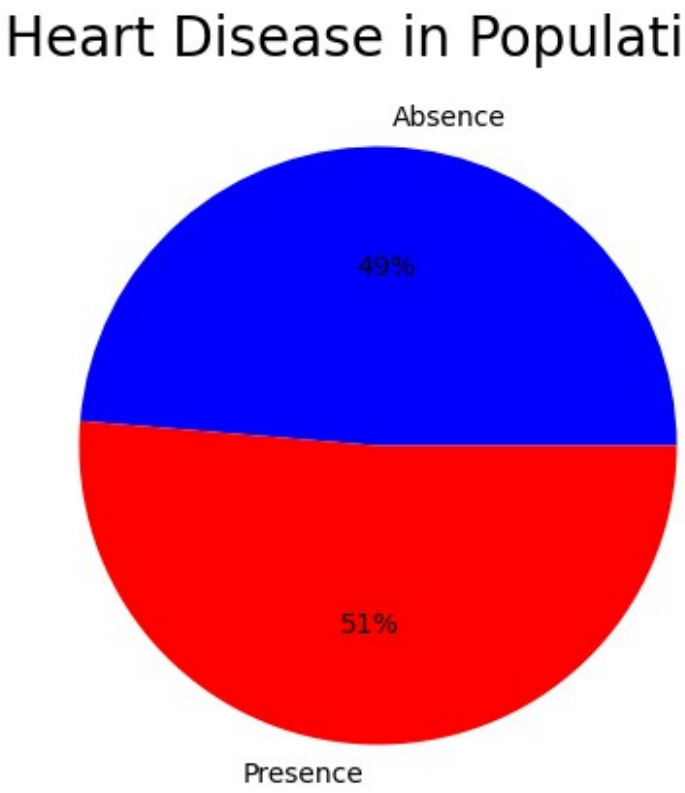
```
   age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang  oldpeak
slope  \
0   52    1   0       125   212    0        1      168      0      1.0
2
1   53    1   0       140   203    1        0      155      1      3.1
0
2   70    1   0       145   174    0        1      125      1      2.6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

```
   ca  thal  target heart_disease
0   2     3       0       Absence
1   0     3       0       Absence
2   0     3       0       Absence
3   1     3       0       Absence
4   3     2       0       Absence
```

```
hd=data.groupby('heart_disease')['target'].count()
hd
```

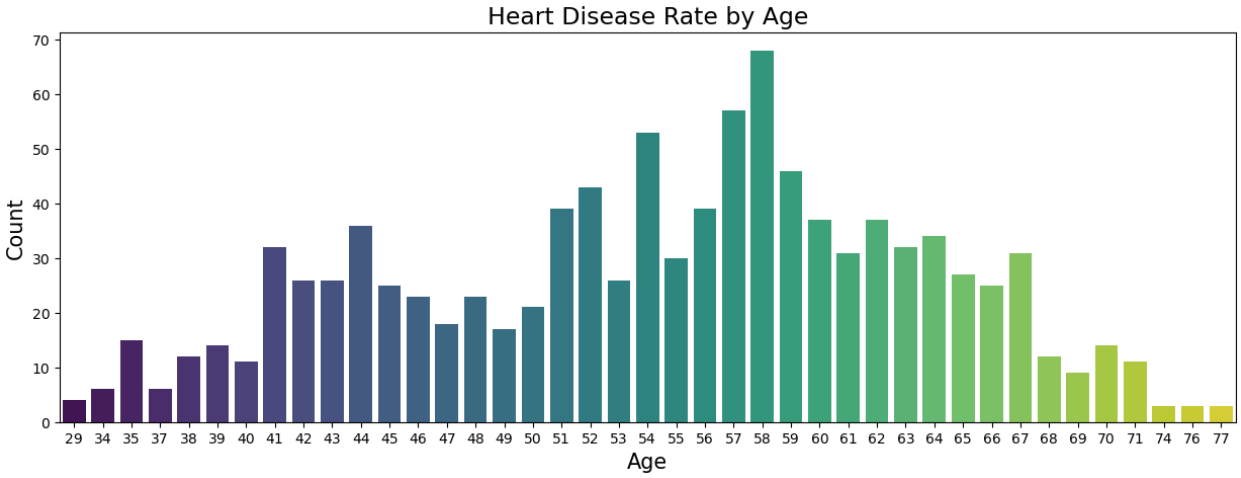
```
heart_disease
Absence     499
Presence    526
Name: target, dtype: int64
```

```
#showing distribution of heart disease by matplotlib pie chart

plt.figure(figsize=(7,5))
plt.pie(hd,labels=['Absence','Presence'],autopct='%0.0f%
%',colors=['blue','red'])
plt.title('Heart Disease in Population',fontsize=20)
plt.show()
```



From the Overall Population, 51% of population is suffering with Heart Disease.

```
#showing heart disease rate by age using countplot

plt.figure(figsize=(15,5))
sns.countplot(data=data,x='age',palette='viridis')
plt.title("Heart Disease Rate by Age",fontsize=17)
plt.xlabel('Age',fontsize=15)
plt.ylabel('Count',fontsize=15)
plt.show()
```



Best analysis can be done by dividing the age group into 3 age ranges-elderly,middle-aged, young people.

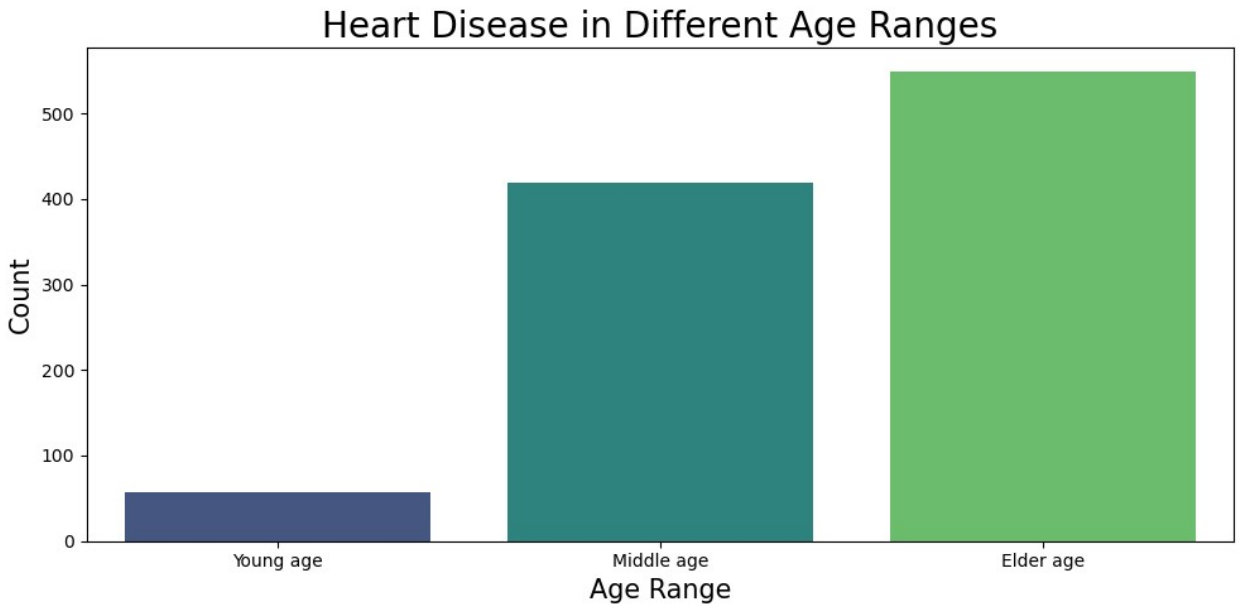
```
#Statistical Analysis

Min_Age=data['age'].min()
Max_Age=data['age'].max()
Mean_Age=data['age'].mean()
print("Minimum Age =",Min_Age)
print("Maximum Age =",Max_Age)
print("Mean Age =",Mean_Age)

Minimum Age = 29
Maximum Age = 77
Mean Age = 54.43414634146342

data.sort_values(by='age',ascending=True,inplace=True)

#grouping ages into 3 ranges and then applying it to the new column in
dataframe 'data'.

def age_range(num):
    if num>=29 and num<40:
        return 'Young age'
    elif num>=40 and  num<55:
```

```
        return 'Middle age'
    else:
        return 'Elder age'

data['age_groups']=data['age'].apply(age_range)
data
```

```
     age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang
oldpeak  \
60    29    1   1       130   204    0        0      202      0
0.0
118   29    1   1       130   204    0        0      202      0
0.0
668   29    1   1       130   204    0        0      202      0
0.0
64    29    1   1       130   204    0        0      202      0
0.0
572   34    1   3       118   182    0        0      174      0
0.0
..   ...  ...  ..       ...   ...  ...      ...      ...    ...      .
..
99    76    0   2       140   197    0        2      116      0
1.1
965   76    0   2       140   197    0        2      116      0
1.1
160   77    1   0       125   304    0        0      162      1
0.0
162   77    1   0       125   304    0        0      162      1
0.0
387   77    1   0       125   304    0        0      162      1
0.0

     slope  ca  thal  target heart_disease age_groups
60       2   0     2       1      Presence  Young age
118      2   0     2       1      Presence  Young age
668      2   0     2       1      Presence  Young age
64       2   0     2       1      Presence  Young age
572      2   0     2       1      Presence  Young age
..     ...  ..   ...     ...           ...        ...
99       1   0     2       1      Presence  Elder age
965      1   0     2       1      Presence  Elder age
160      2   3     2       0       Absence  Elder age
162      2   3     2       0       Absence  Elder age
387      2   3     2       0       Absence  Elder age

[1025 rows x 16 columns]
```

```
plt.figure(figsize=(10,5))

sns.countplot(data=data,x='age_groups',palette='viridis')
```

```
plt.title("Heart Disease in Different Age Ranges",fontsize=20)
plt.xlabel('Age Range', fontsize=15)
plt.ylabel('Count', fontsize=15)
plt.tight_layout()
plt.show()
```



The above plot shows that elder age group(>50) is more prone to heart diseases as compared to other age groups(29-55).

```
#Heart Disease spread on the basis of gender,adding new column

def sex1(row):
    if row==0:
        return 'Male'
    elif row==1:
        return 'Female'
data['gender']=data['sex'].apply(sex1)
data
```

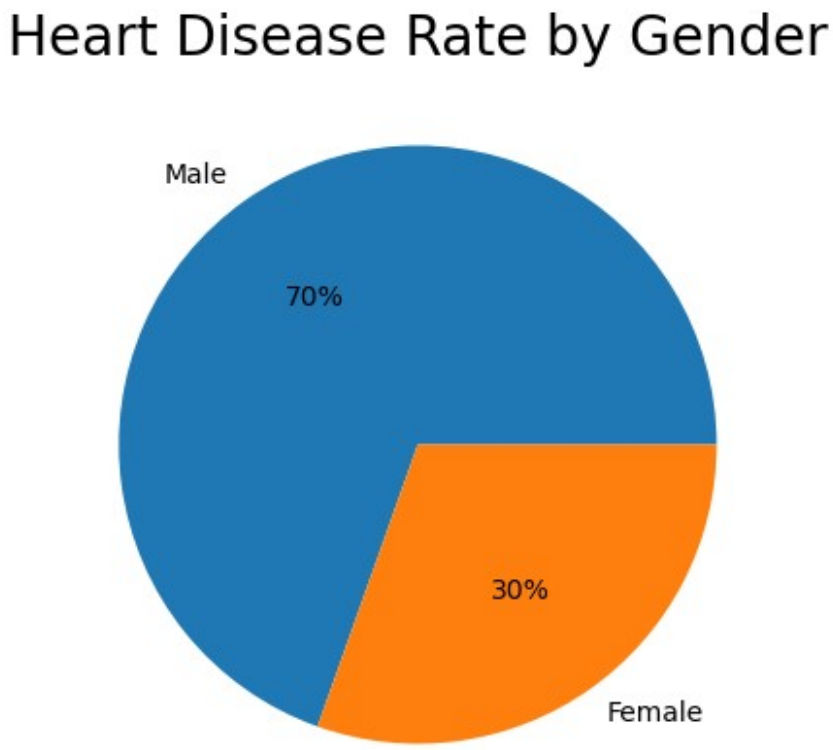
```
     age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang
oldpeak  \
60    29    1   1       130   204    0        0      202      0
0.0
118   29    1   1       130   204    0        0      202      0
0.0
668   29    1   1       130   204    0        0      202      0
0.0
64    29    1   1       130   204    0        0      202      0
0.0
572   34    1   3       118   182    0        0      174      0
```

```
0.0
..   ...  ...  ..       ...   ...  ...      ...      ...    ...      .
..
99    76    0   2       140   197    0        2      116      0
1.1
965   76    0   2       140   197    0        2      116      0
1.1
160   77    1   0       125   304    0        0      162      1
0.0
162   77    1   0       125   304    0        0      162      1
0.0
387   77    1   0       125   304    0        0      162      1
0.0

     slope  ca  thal  target heart_disease age_groups  gender
60       2   0     2       1      Presence  Young age  Female
118      2   0     2       1      Presence  Young age  Female
668      2   0     2       1      Presence  Young age  Female
64       2   0     2       1      Presence  Young age  Female
572      2   0     2       1      Presence  Young age  Female
..     ...  ..   ...     ...           ...        ...     ...
99       1   0     2       1      Presence  Elder age    Male
965      1   0     2       1      Presence  Elder age    Male
160      2   3     2       0       Absence  Elder age  Female
162      2   3     2       0       Absence  Elder age  Female
387      2   3     2       0       Absence  Elder age  Female

[1025 rows x 17 columns]
```

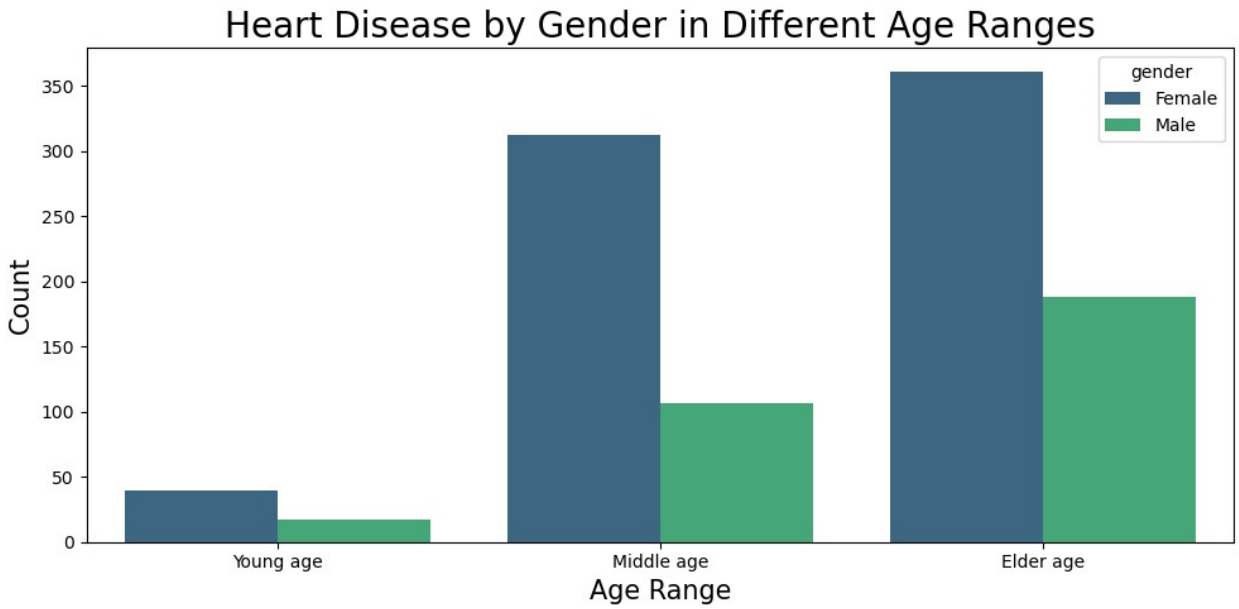
```
gender1=data.groupby('gender')['sex'].count()
gender1
```

```
gender
Female    713
Male      312
Name: sex, dtype: int64
```

```
#showing distribution of heart disease by gender using matplotlib pie
chart

plt.figure(figsize=(7,5))
plt.pie(gender1,labels=['Male','Female'],autopct='%0.0f%%')
plt.title('Heart Disease Rate by Gender',fontsize=20)
plt.show()
```

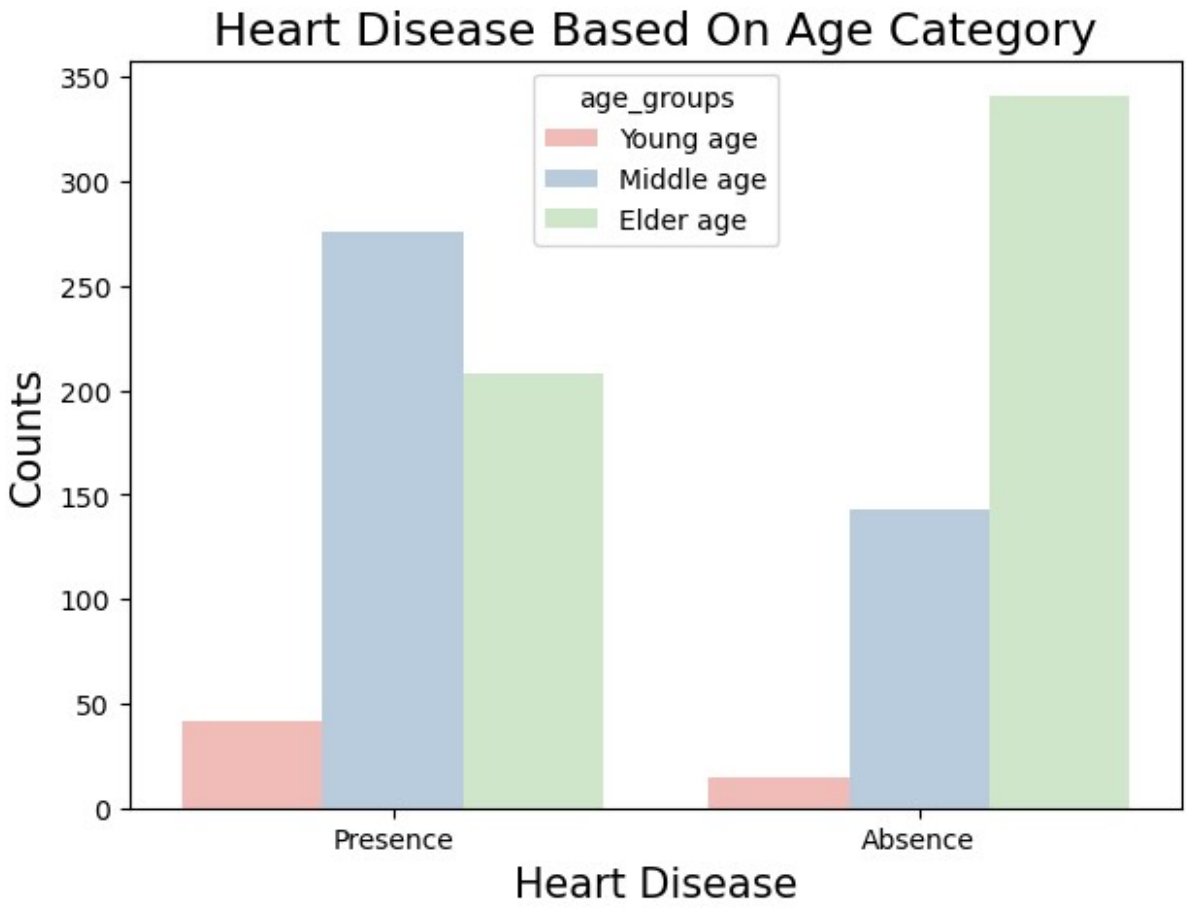



```
#creating countplot to show disease spread by gender in different age
groups.
plt.figure(figsize=(10,5))

sns.countplot(x='age_groups',data=data,hue='gender',palette='viridis')

plt.title("Heart Disease by Gender in Different Age
Ranges",fontsize=20)
plt.xlabel('Age Range', fontsize=15)
plt.ylabel('Count', fontsize=15)
plt.tight_layout()
plt.show()
```



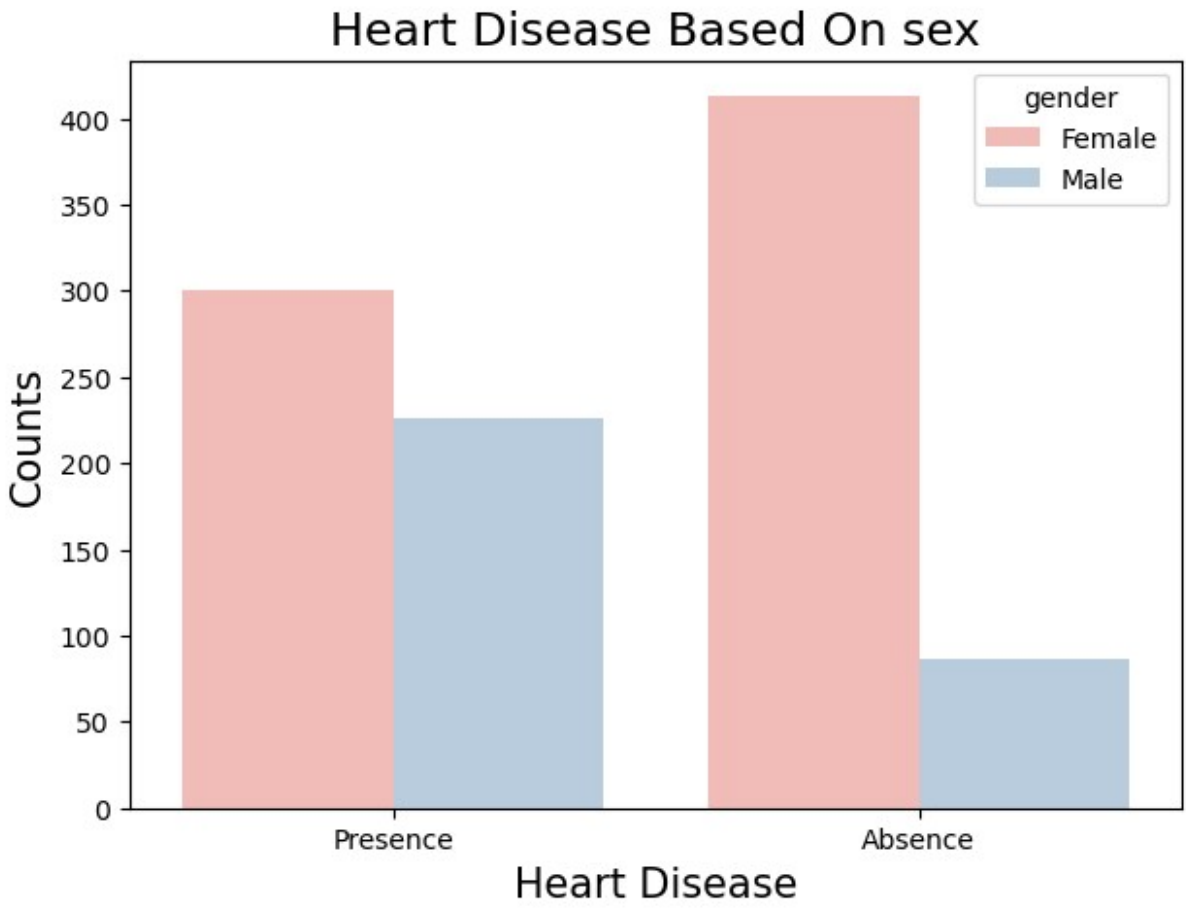
The above plots show that the heart diseases are spread more in females(70%) as compared to males(30%) in all age ranges and the greatest is in elder age group.

```
#percentage of people having Heart Disease based on age categry

plt.figure(figsize=(7,5))
sns.countplot(x='heart_disease', hue='age_groups', data=data,
order=['Presence','Absence'], palette='Pastel1')
plt.title('Heart Disease Based On Age Category', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.show()
```



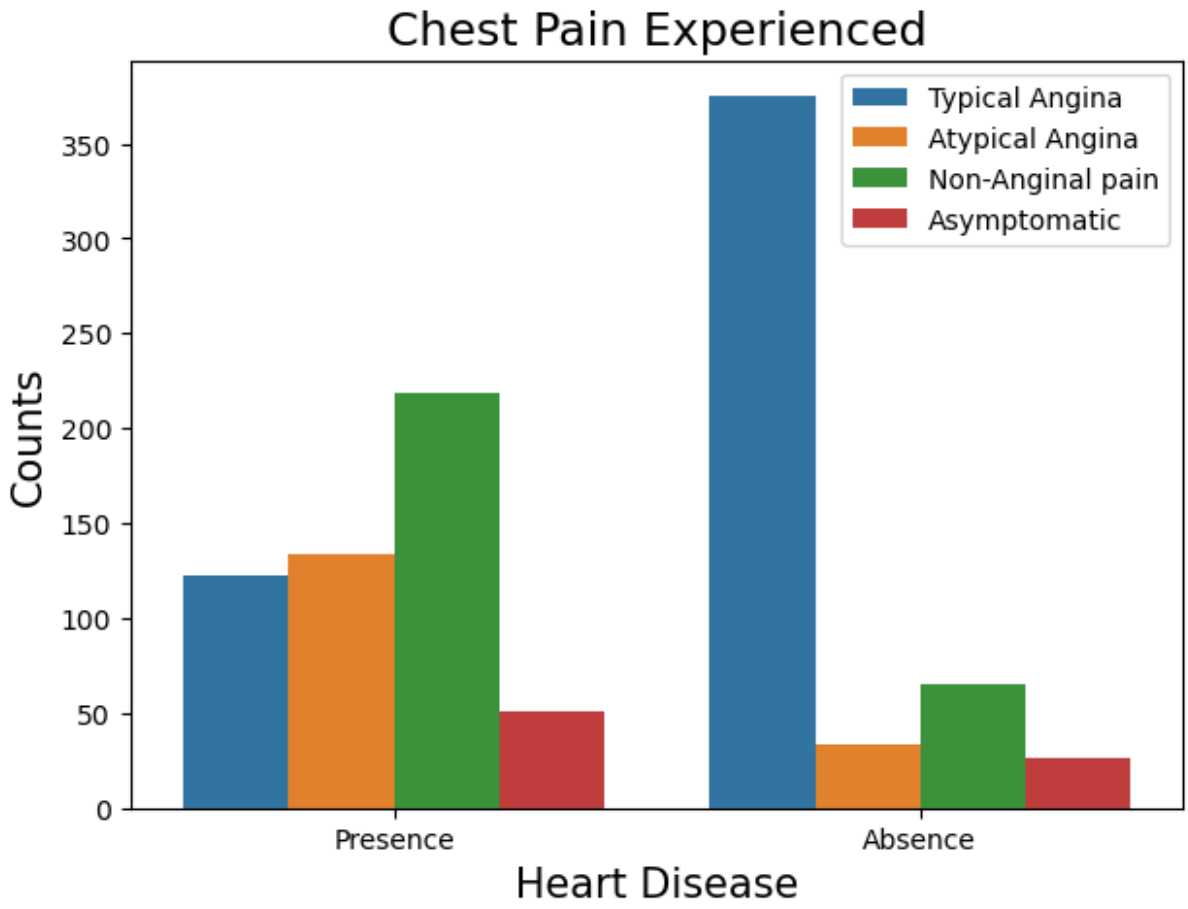
Middle Aged People are most affected by Heart Disease.

```
#percentage of people having Heart Disease based on sex

plt.figure(figsize=(7,5))
sns.countplot(x='heart_disease', hue='gender', data=data,
order=['Presence','Absence'], palette='Pastel1')
plt.title('Heart Disease Based On sex', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.show()
```



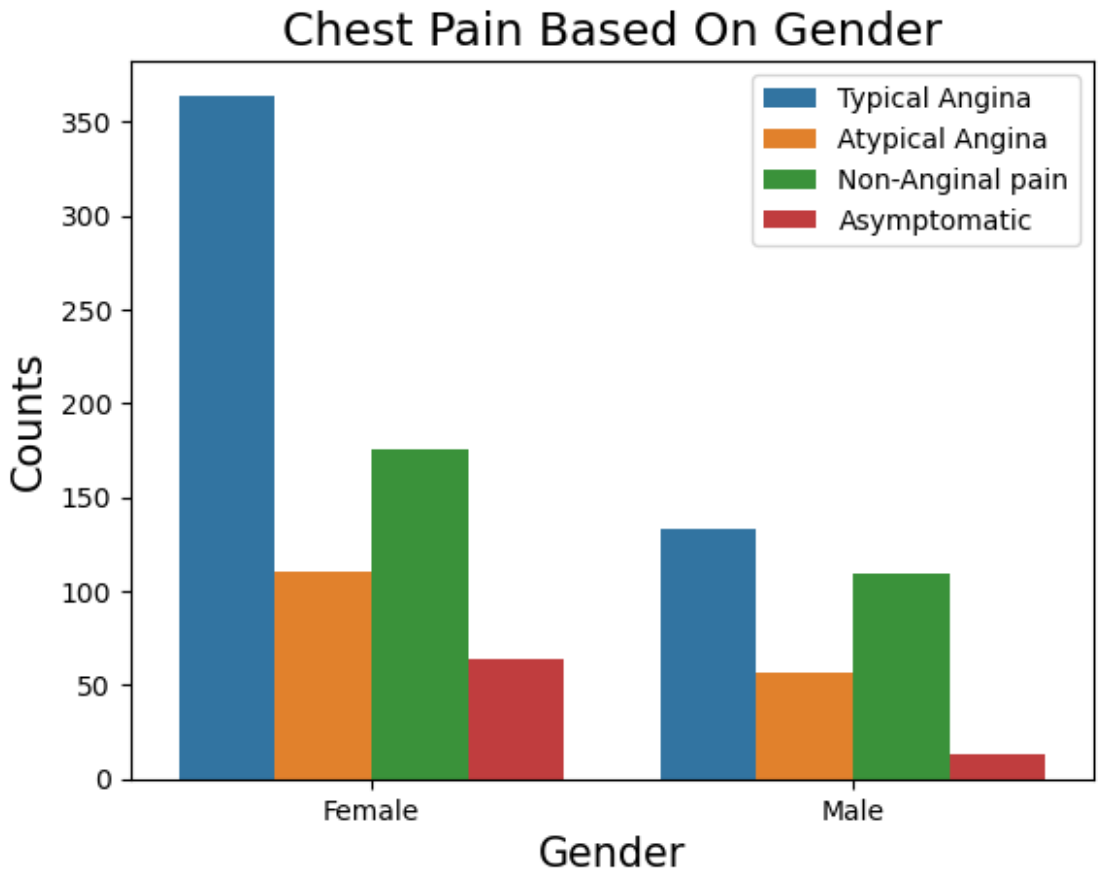
We can see that females are more prone to Heart Disease.

```
#Count Plot Creation of Chest Pain Experienced using MatplotLib and
Seaborn
plt.figure(figsize=(7,5))
sns.countplot(x='heart_disease', hue='cp', data=data,
order=['Presence','Absence'])
plt.title('Chest Pain Experienced', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.legend(labels=['Typical Angina','Atypical Angina','Non-Anginal
pain','Asymptomatic'])
plt.show()
```

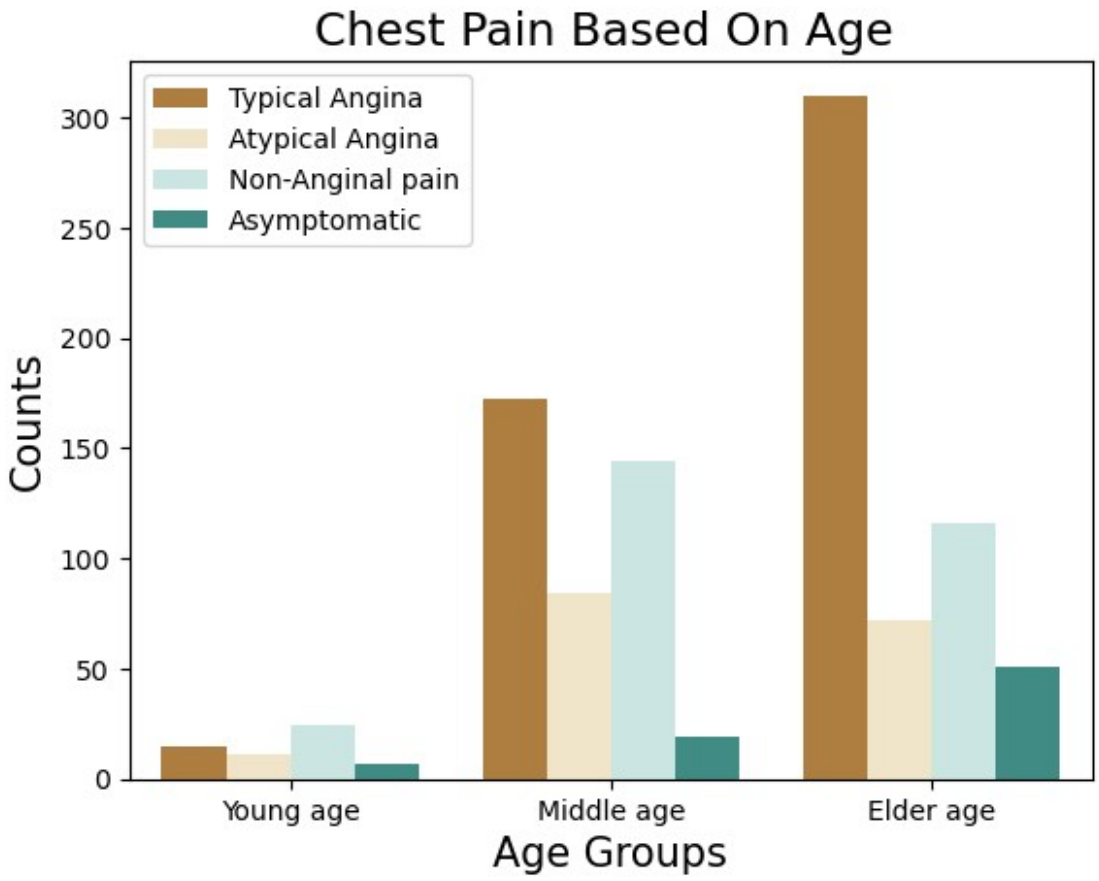


It seems people having Non-Anginal chest pain have a higher chance of heart disease.

```
#Count Plot Creation of Chest Pain Based On Gender

sns.countplot(x=data['gender'], hue='cp', data=data)
plt.title('Chest Pain Based On Gender', fontsize=17)
plt.xlabel('Gender', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.legend(labels=['Typical Angina','Atypical Angina','Non-Anginal
pain','Asymptomatic'])
plt.show()
```

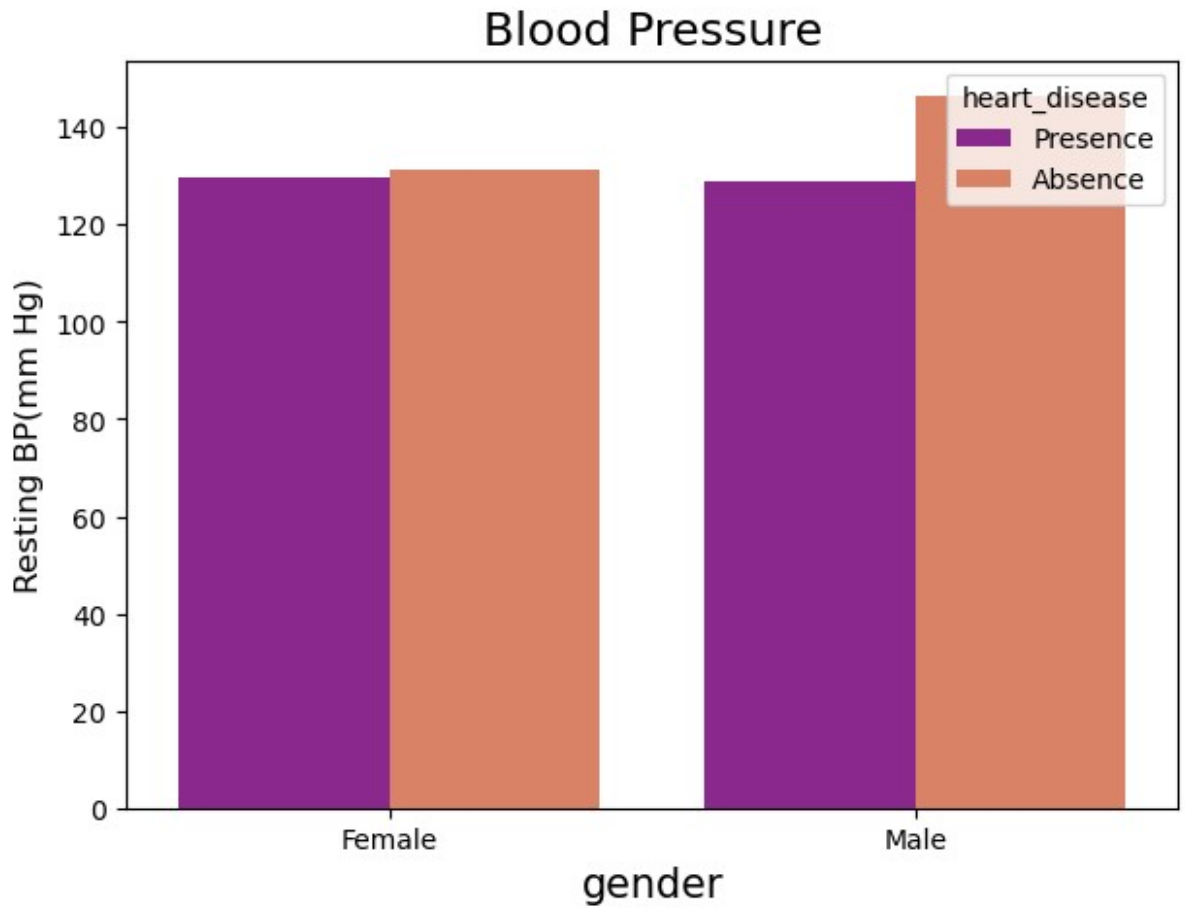



```
#Count Plot Creation of Chest Pain Based On Age

sns.countplot(x=data['age_groups'], hue='cp',
data=data,palette='BrBG')
plt.title('Chest Pain Based On Age ', fontsize=17)
plt.xlabel('Age Groups', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.legend(labels=['Typical Angina','Atypical Angina','Non-Anginal
pain','Asymptomatic'])
plt.show()
```



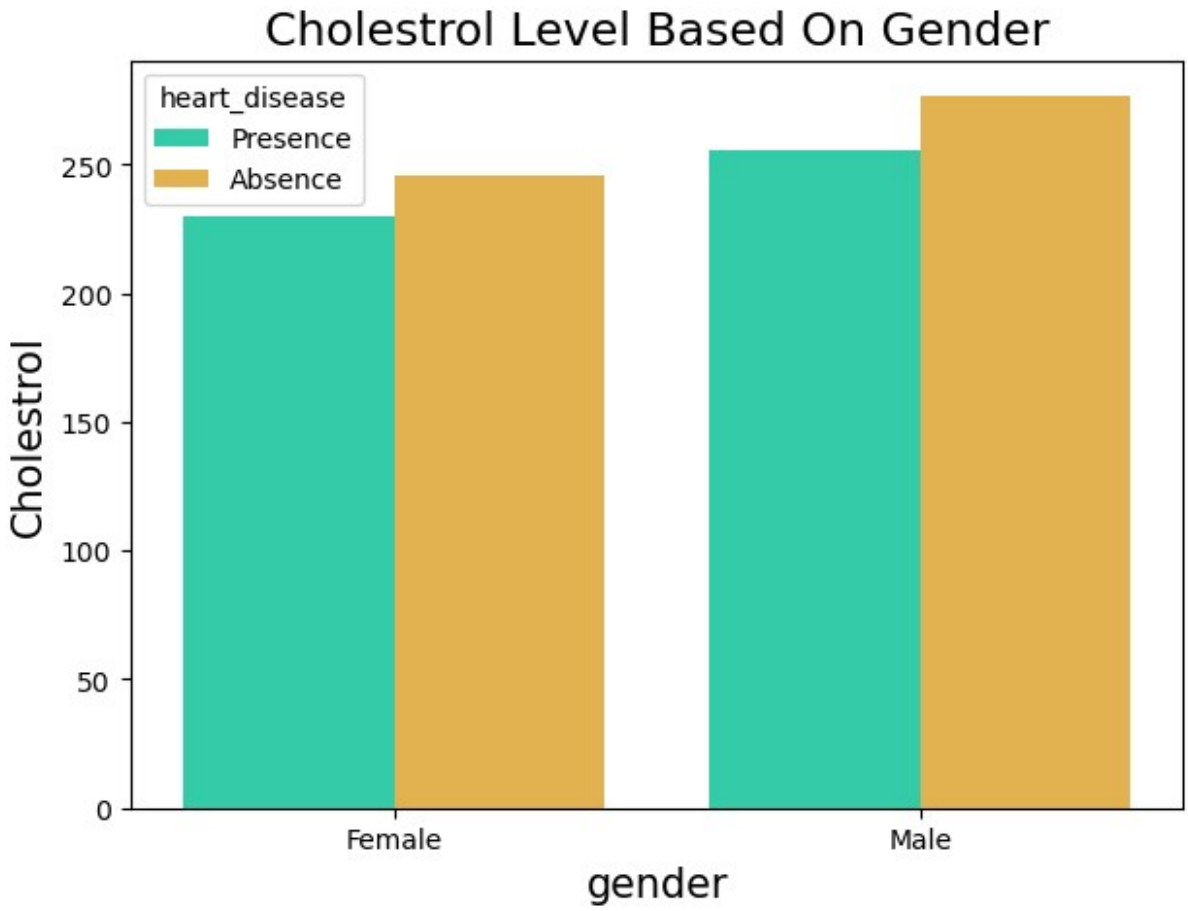
There is very high number of Typical Anginal Pain in Elderly age Category

```
#Bar Plot Creation of Person's Resting Blood Pressure (mm Hg)
plt.figure(figsize=(7,5))
sns.barplot(x='gender', y='trestbps',hue='heart_disease', data=data,
palette='plasma',ci=None)
plt.title("Blood Pressure", fontsize=17)
plt.xlabel('gender',fontsize=15)
plt.ylabel("Resting BP(mm Hg)", fontsize=12)
plt.show()
```



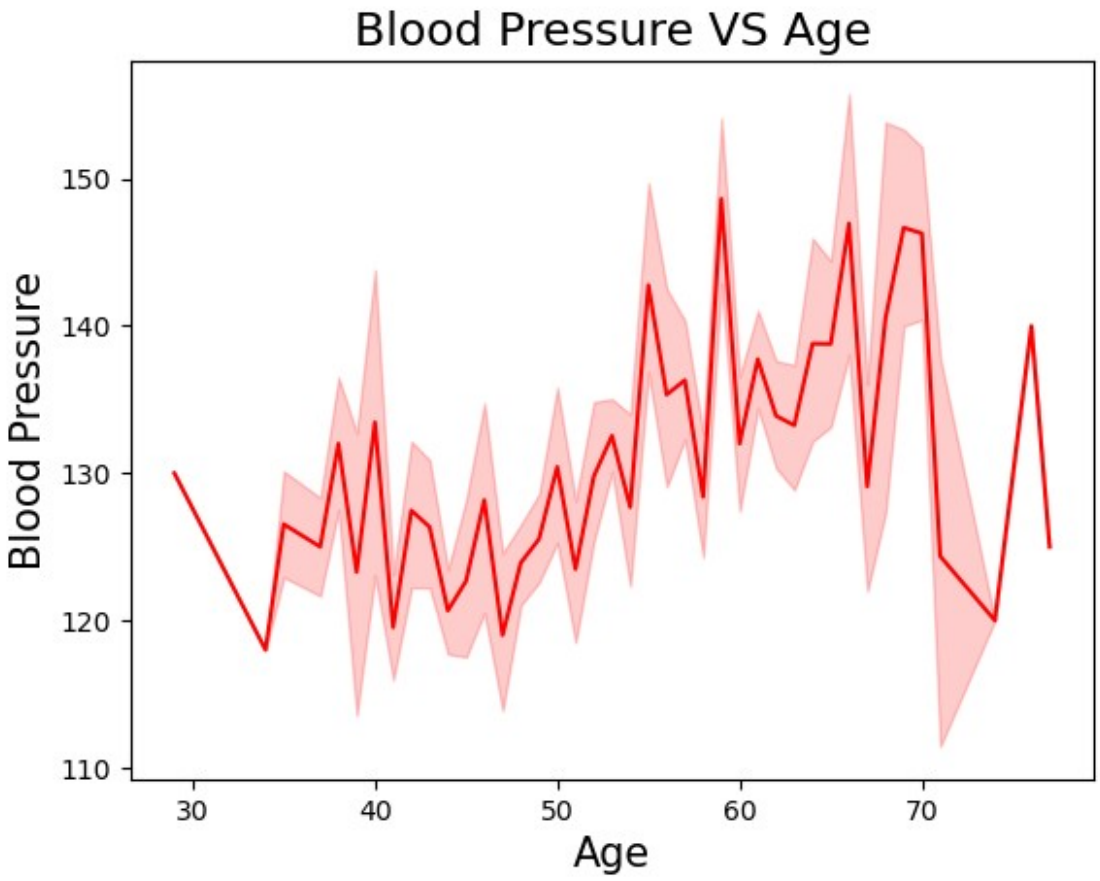
Blood Pressure Rate is almost equal in Males and Females

```
#Bar Plot Creation of Cholesterol Level Based On Gender
plt.figure(figsize=(7,5))
sns.barplot(x='gender', y='chol', data=data,hue='heart_disease',
palette='turbo',ci=None)
plt.title("Cholestrol Level Based On Gender", fontsize=17)
plt.xlabel('gender',fontsize=15)
plt.ylabel("Cholestrol", fontsize=15)
plt.show()
```



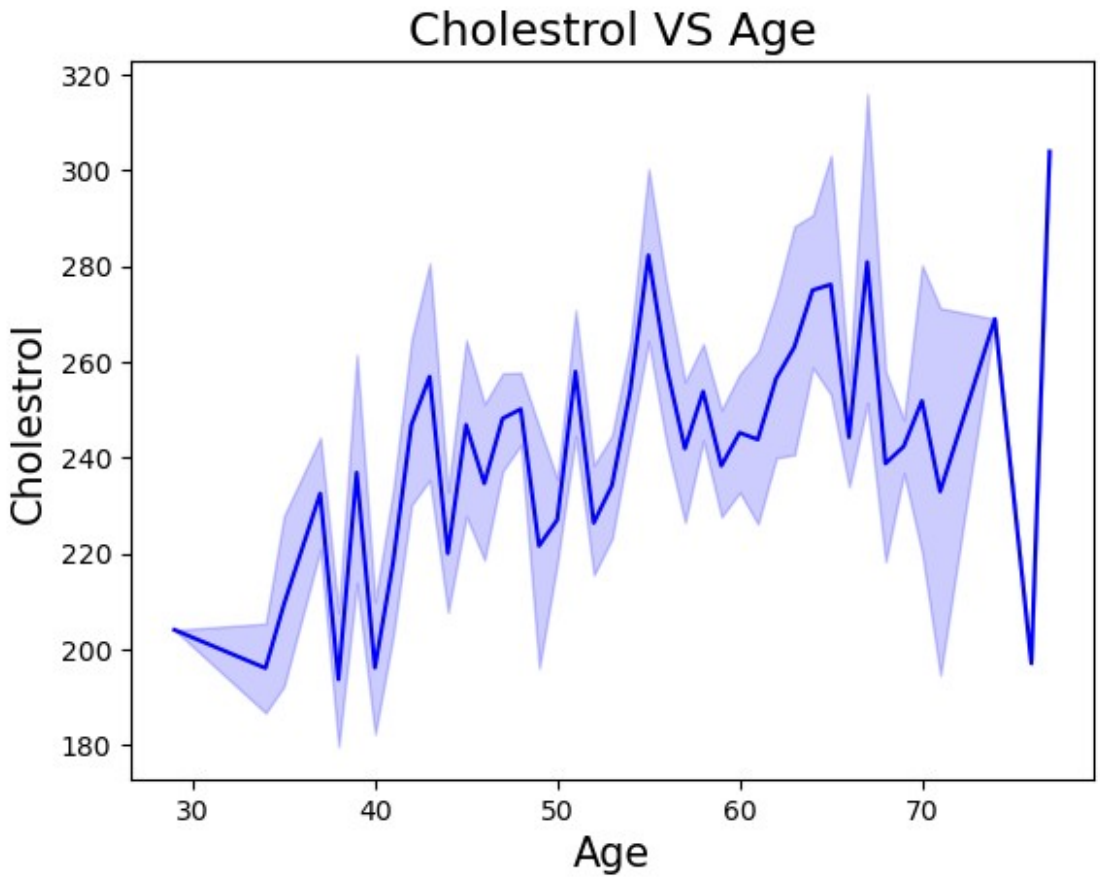
Males have little bit of higher cholesterol than females.

```
#Line Plot Creation of Blood Pressure by Age using MatplotLib and
Seaborn

sns.lineplot(x='age', y='trestbps' ,data=data, color='r')
plt.title('Blood Pressure VS Age', fontsize=17)
plt.xlabel('Age', fontsize=15)
plt.ylabel('Blood Pressure', fontsize=15)
plt.show()
```



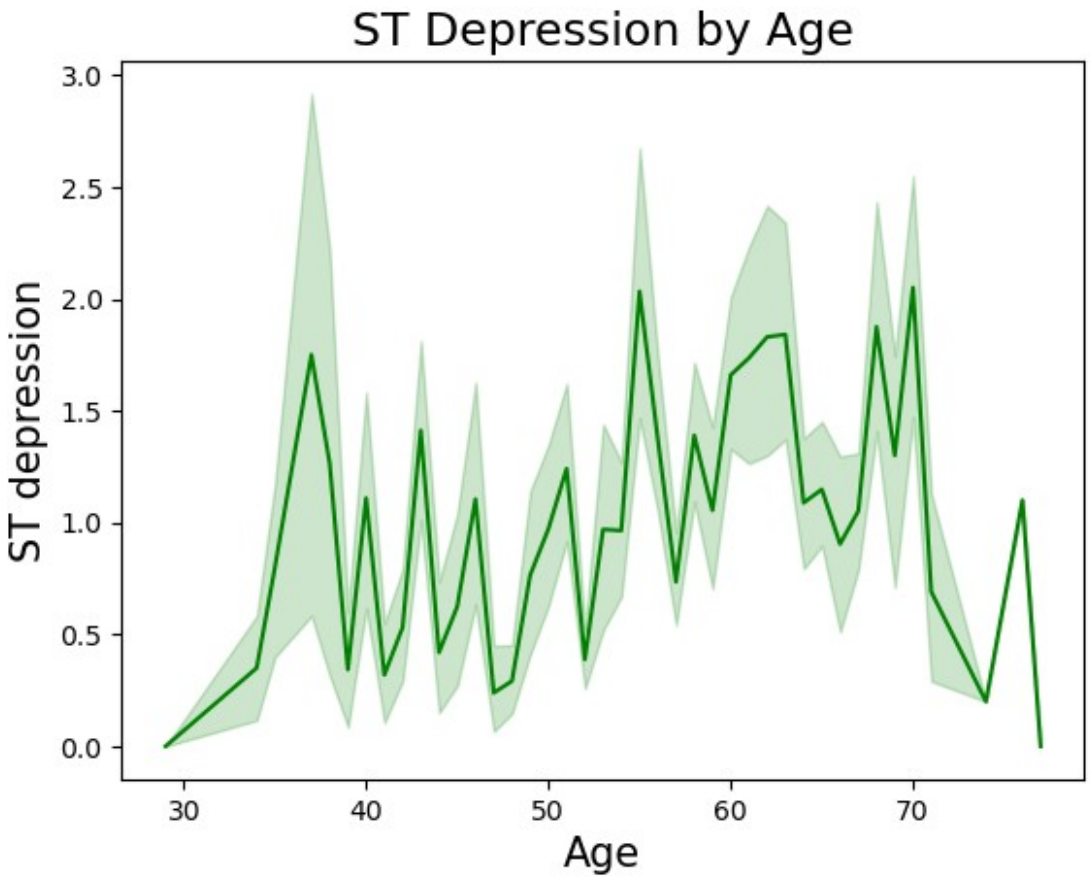
Here we can observe that Blood Pressure increases between age of 50 to 60 and somehow continue the pattern till 70

```
#Line Plot Creation of Cholestrol by Age

sns.lineplot(x='age', y='chol', data=data, color='b')
plt.title('Cholestrol VS Age', fontsize=17)
plt.xlabel('Age', fontsize=15)
plt.ylabel('Cholestrol', fontsize=15)
plt.show()
```



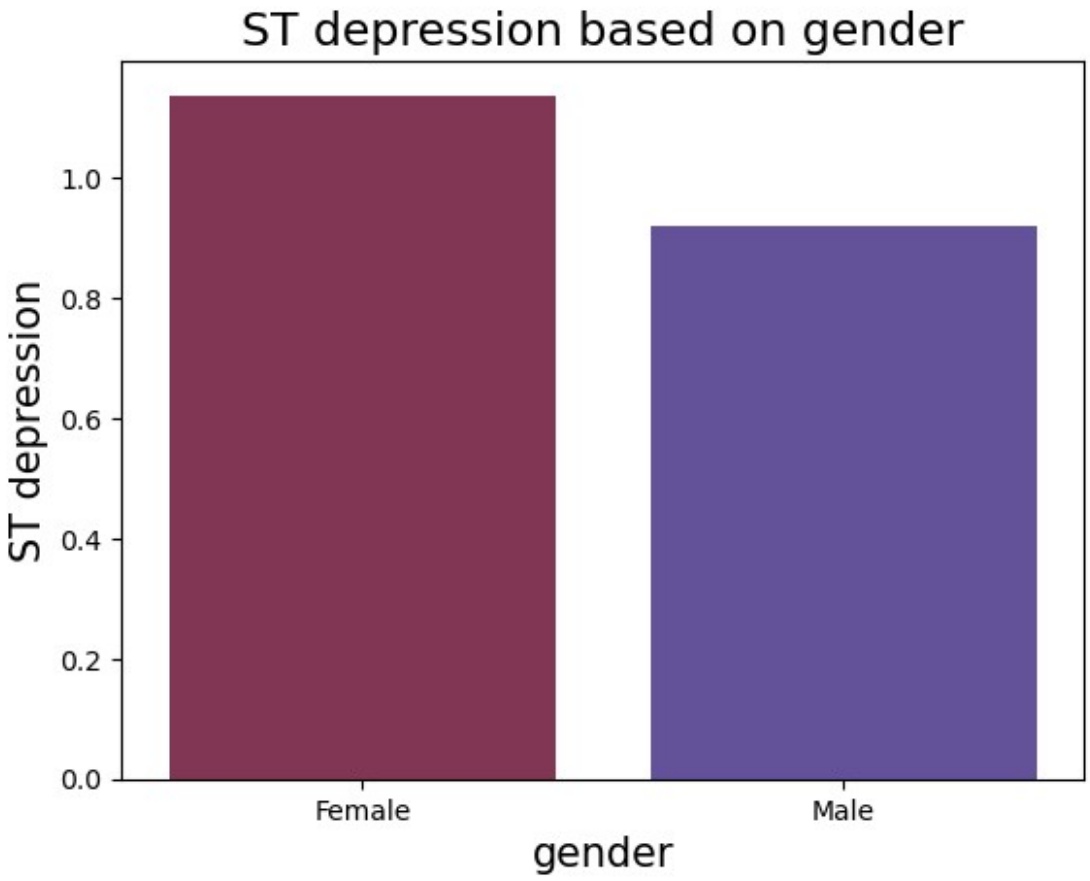
Similarly Cholesterol is Increasing in the age group of 50-60

```
#Line Plot Creation of ST Depression by Age

sns.lineplot(x='age', y='oldpeak',data=data, color='g')
plt.title('ST Depression by Age', fontsize=17)
plt.xlabel('Age', fontsize=15)
plt.ylabel('ST depression', fontsize=15)
plt.show()
```



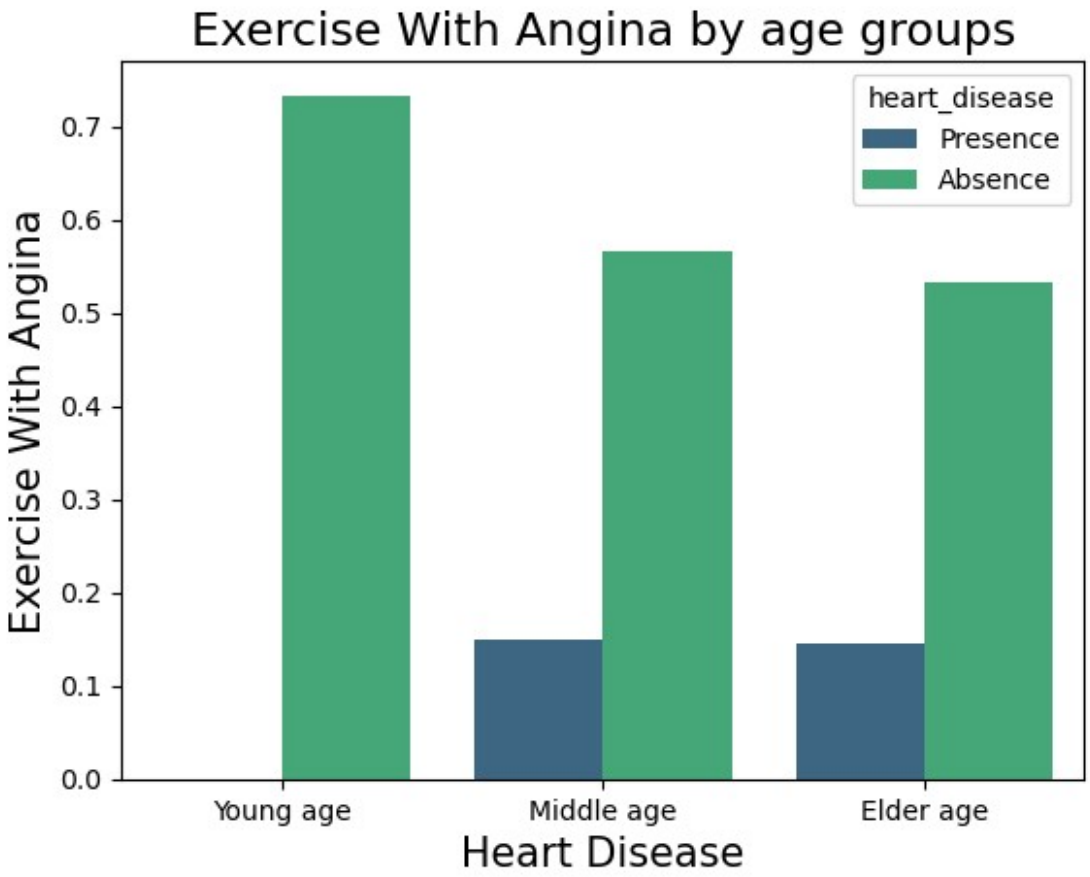
we can observe from here that ST depression mostly increases between the age group of 30-40.
-> ST depression refers to a finding on an electrocardiogram, wherein the trace in the ST segment is abnormally low below the baseline.

```
#Bar Plot Creation of ST depression VS Heart Disease using MatplotLib
and Seaborn

sns.barplot(x='gender', y='oldpeak', data=data,
palette='twilight_r',ci=None)
plt.title('ST depression based on gender', fontsize=17)
plt.xlabel('gender', fontsize=15)
plt.ylabel('ST depression', fontsize=15)
plt.show()
```



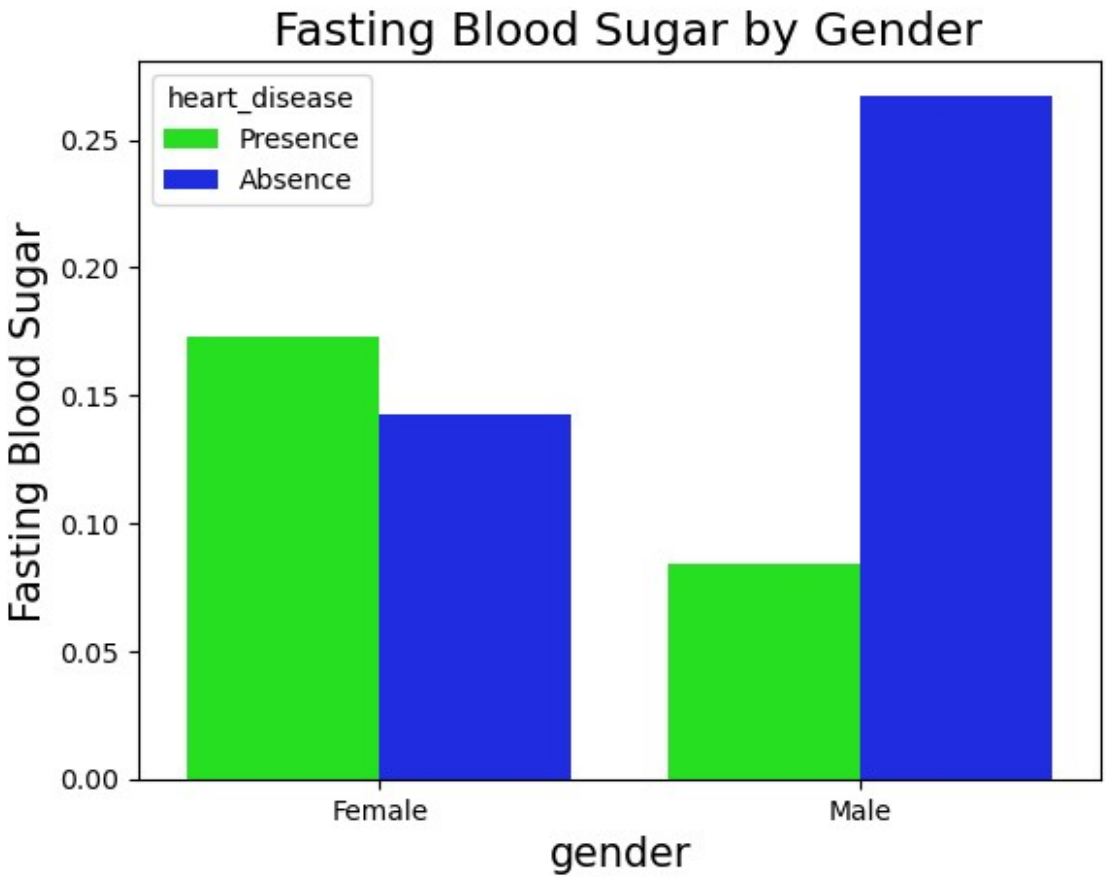
More Females are prone to ST depression as compare to males

```
#Bar Plot Creation of Exercise With Angina VS Heart Disease using
MatplotLib and Seaborn

sns.barplot(x='age_groups', y='exang',hue="heart_disease", data=data,
palette='viridis',ci=None)
plt.title('Exercise With Angina by age groups', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Exercise With Angina', fontsize=15)
plt.show()
```



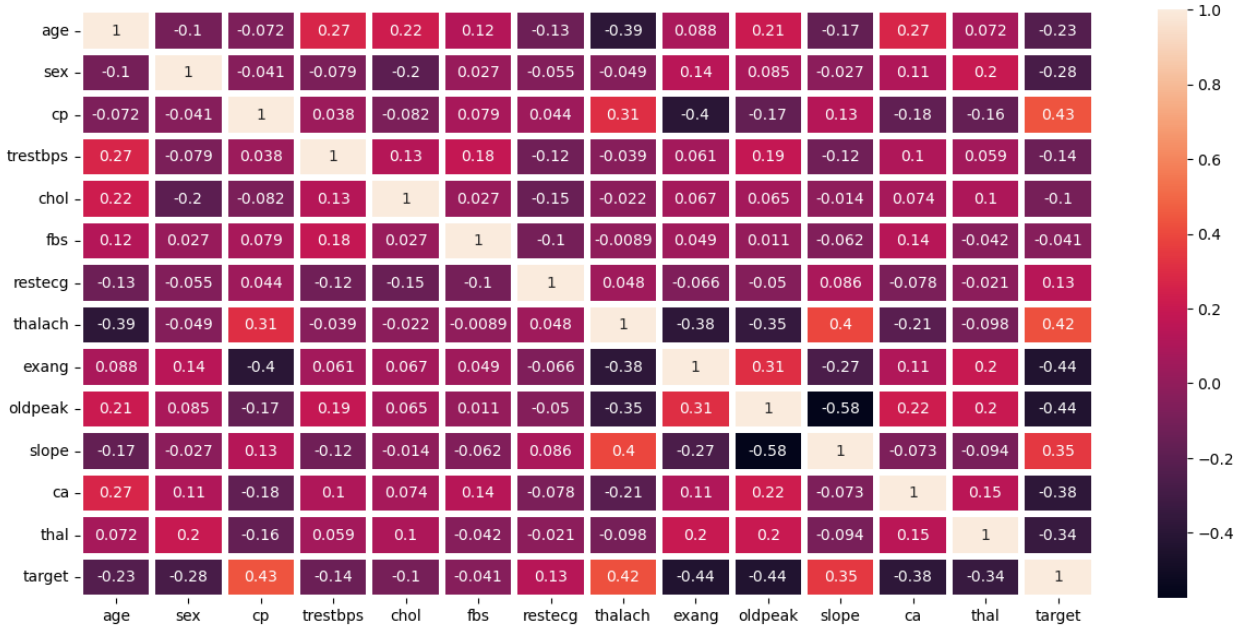
Middle and Elder age people are likely to suffer with exercise induced angina whereas this is negligent in young age people . ->exang is a type of chest pain caused by reduced blood flow to the heart.

```
#Bar Plot Creation of Fasting Blood Sugar by gender

sns.barplot(y='fbs', x='gender',hue='heart_disease' ,data=data,
palette='hsv',ci=None)
plt.title(' Fasting Blood Sugar by Gender', fontsize=17)
plt.xlabel('gender', fontsize=15)
plt.ylabel('Fasting Blood Sugar', fontsize=15)
plt.show()
```



Female have more fasting blood sugar over 120 mg/dl as compared to male.

```
#Creating a heatmap to show the dependency of features responsible for
Heart Diseases

plt.figure(figsize=(15,7))
sns.heatmap(data.corr(),linewidths=3,linecolor='white',annot=True)
plt.show()
```


age 1 -0.1 -0.072 0.27 0.22 0.12 -0.13 -0.39 0.088 0.21 -0.17 0.27 0.072 -0.23
sex -0.1 1 -0.041 -0.079 -0.2 0.027 -0.055 -0.049 0.14 0.085 -0.027 0.11 0.2 -0.28
cp -0.072 -0.041 1 0.038 -0.082 0.079 0.044 0.31 -0.4 -0.17 0.13 -0.18 -0.16 0.43
trestbps 0.27 -0.079 0.038 1 0.13 0.18 -0.12 -0.039 0.061 0.19 -0.12 0.1 0.059 -0.14
chol 0.22 -0.2 -0.082 0.13 1 0.027 -0.15 -0.022 0.067 0.065 -0.014 0.074 0.1 -0.1
fbs 0.12 0.027 0.079 0.18 0.027 1 -0.1 -0.0089 0.049 0.011 -0.062 0.14 -0.042 -0.041
restecg -0.13 -0.055 0.044 -0.12 -0.15 -0.1 1 0.048 -0.066 -0.05 0.086 -0.078 -0.021 0.13
thalach -0.39 -0.049 0.31 -0.039 -0.022 -0.0089 0.048 1 -0.38 -0.35 0.4 -0.21 -0.098 0.42
exang 0.088 0.14 -0.4 0.061 0.067 0.049 -0.066 -0.38 1 0.31 -0.27 0.11 0.2 -0.44
oldpeak 0.21 0.085 -0.17 0.19 0.065 0.011 -0.05 -0.35 0.31 1 -0.58 0.22 0.2 -0.44
slope -0.17 -0.027 0.13 -0.12 -0.014 -0.062 0.086 0.4 -0.27 -0.58 1 -0.073 -0.094 0.35
ca 0.27 0.11 -0.18 0.1 0.074 0.14 -0.078 -0.21 0.11 0.22 -0.073 1 0.15 -0.38
thal 0.072 0.2 -0.16 0.059 0.1 -0.042 -0.021 -0.098 0.2 0.2 -0.094 0.15 1 -0.34
target -0.23 -0.28 0.43 -0.14 -0.1 -0.041 0.13 0.42 -0.44 -0.44 0.35 -0.38 -0.34 1
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4