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

In [2]:
```
sns.set_style('whitegrid')
```

In [3]:
```
data = pd.read_csv('E:\Internship\Heart Disease Diagnostic Analysis\Data\heart_disease_
data
```

Out[3]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	45	1	1	110	264	0	0	132	0	1.2	2	0	7	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
300	57	1	4	130	131	0	0	115	1	1.2	2	1	7	1
301	57	0	2	130	236	0	2	174	0	0.0	2	1	3	1
302	38	1	3	138	175	0	0	173	0	0.0	1	-100000	3	0

303 rows × 14 columns

In [4]:
```
#All columns in the dataset
data.columns
```

Out[4]:
```
Index(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
       'exang', 'oldpeak', 'slope', 'ca', 'thal', 'num'],
      dtype='object')
```

Here are thirteen features in Dataset age: The person's age in years:-

sex: The person's sex (1 = male, 0 = 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 (3 = normal; 6 = fixed defect; 7 = reversable defect)

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

In [5]:
```
#Checking NULL Values

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

Out[5]:
```
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
num         0
dtype: int64
```

there is no Missing values in our dataset

Percentage of people having Heart Disease

In [7]:
```
num=data.groupby('num').size()
num
```

Out[7]:
```
num
```

```
0    164
1    139
dtype: int64
```

In [8]:
```
#Converting Numerical Data into Categorical Data

def heart_disease(row):
    if row==0:
        return 'Absence'
    elif row==1:
        return 'Presence'
```

In [9]:
```
data['Heart_Disease']=data['num'].apply(heart_disease)
data.head()
```

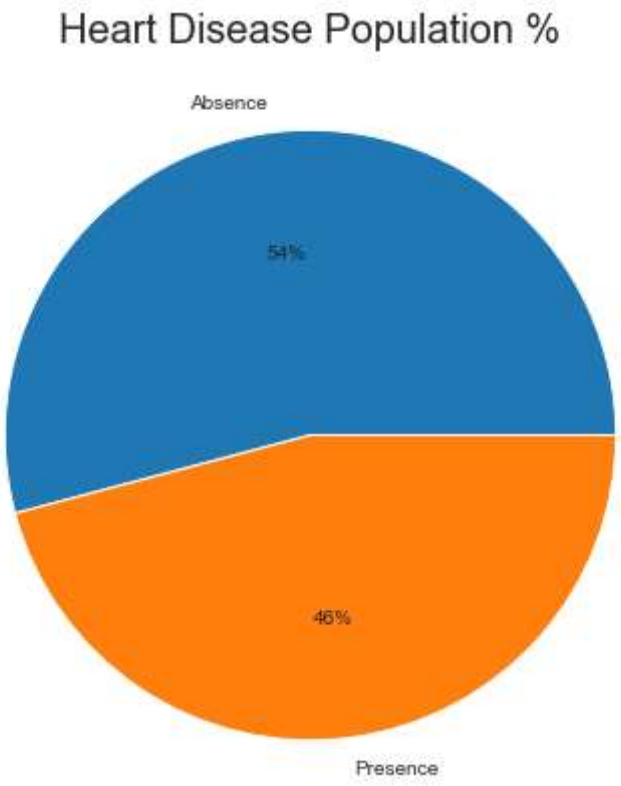
Out[9]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_D
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	A
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Pr
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Pr
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	A
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	A

In [10]:
```
hd=data.groupby('Heart_Disease')['num'].count()
hd
```

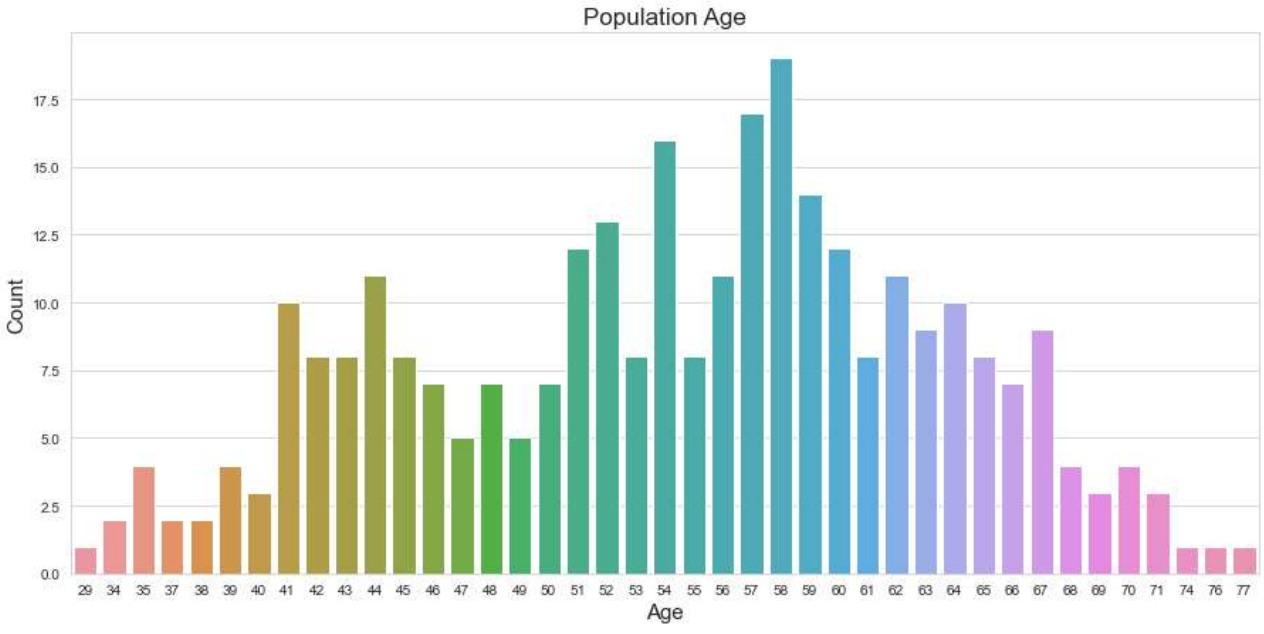
Out[10]:
```
Heart_Disease
Absence     164
Presence    139
Name: num, dtype: int64
```

In [11]:
```
#Pie Chart Creation of Heart Disease Population % using MatplotLib

plt.figure(figsize=(10,7))
plt.pie(hd, labels=['Absence','Presence'], autopct='%0.0f%%')
plt.title('Heart Disease Population %', fontsize=20)
plt.show()
```



From the overall population, people having heart disease are (46%) where it absent is (56%)

In [12]:
```
#Countplot Creation of Population Age using MatplotLib and Seaborn

plt.figure(figsize=(15,7))
sns.countplot(x='age', data=data)
plt.title('Population Age', fontsize=17)
plt.xlabel('Age', fontsize=15)
plt.ylabel('Count', fontsize=15)
plt.show()
```



In [13]:

```
#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 = 54.43894389438944
```

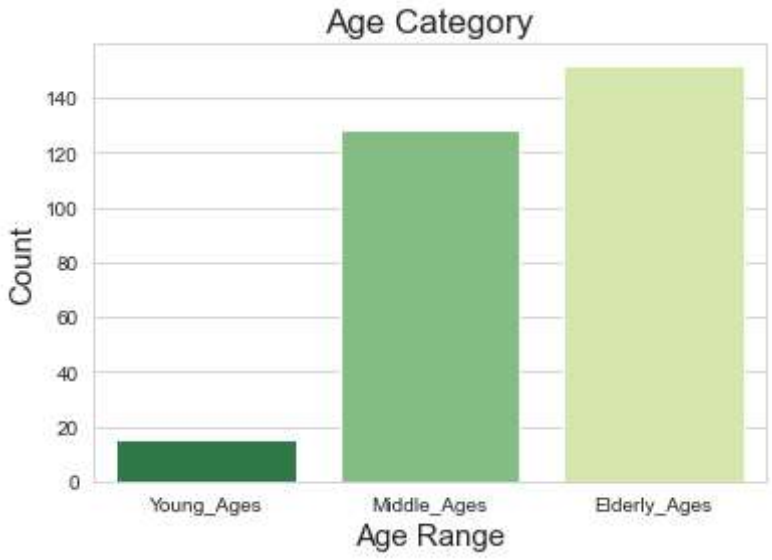
In [14]:

```
#Categorical Analysis

Young_Ages=data[(data['age']>=29) & (data['age']<40)]
Middle_Ages=data[(data['age']>=40) & (data['age']<55)]
Elderly_Ages=data[(data['age']>55)]
print('Young Ages =',len(Young_Ages))
print('Middle Ages =',len(Middle_Ages))
print('Elderly Ages =',len(Elderly_Ages))
```

```
Young Ages = 15
Middle Ages = 128
Elderly Ages = 152
```

In [15]:

```
#Bar Plot Creation of Age Category using MatplotLib and Seaborn

sns.barplot(x=['Young_Ages','Middle_Ages','Elderly_Ages'], y=[len(Young_Ages), len(Midd
plt.title('Age Category', fontsize=17)
plt.xlabel('Age Range', fontsize=15)
plt.ylabel('Count', fontsize=15)
plt.show()
```



In [16]:

```
#Converting Numerical Data into Categorical Data

def gender(row):
    if row==1:
        return 'Male'
```

```
    elif row==0:
        return 'Female'

#Applying converted data into our dataset with new column - sex1

data['sex1']=data['sex'].apply(gender)
data.head()
```

Out[16]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_D
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	A
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Pr
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Pr
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	A
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	A

In [17]:

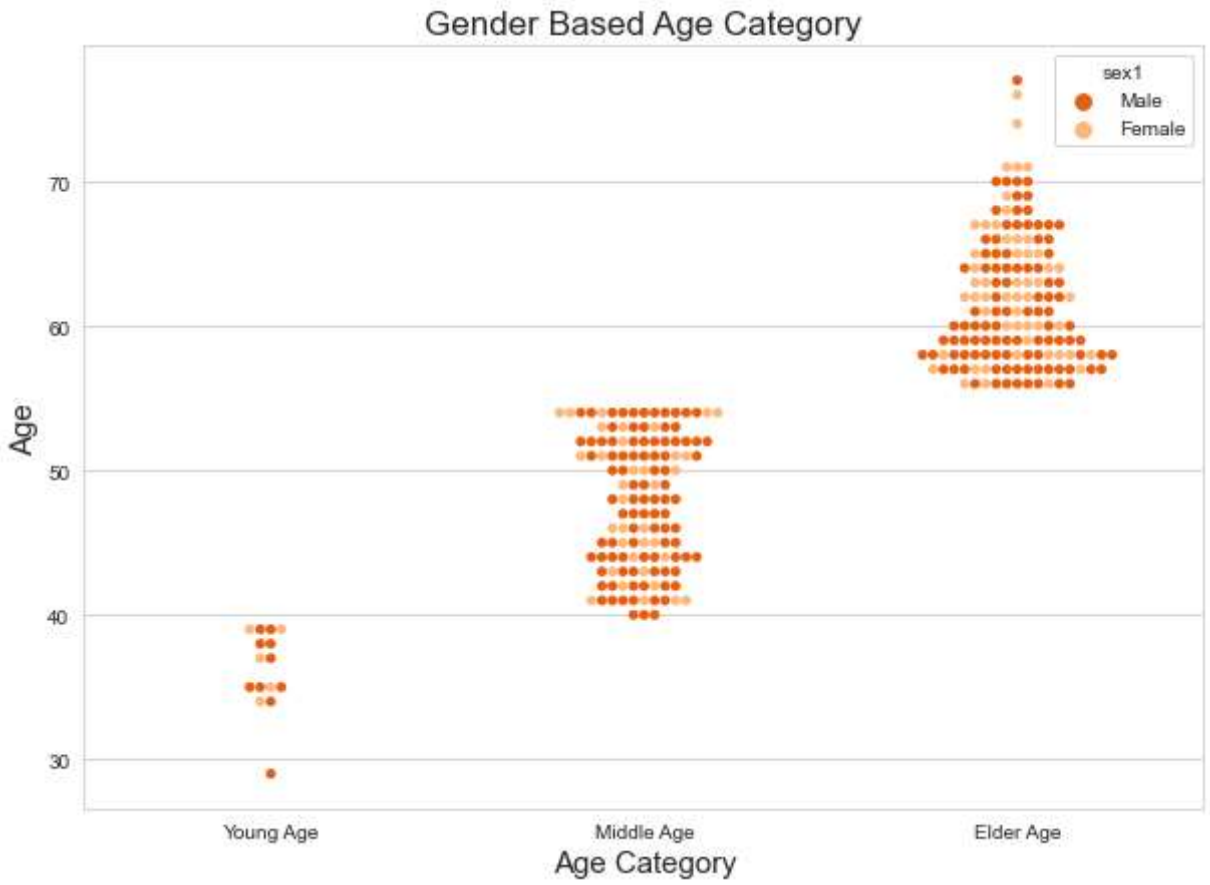
```
#Converting Numerical Data into Categorical Data

def age_range(row):
    if row>=29 and row<40:
        return 'Young Age'
    elif row>=40 and row<55:
        return 'Middle Age'
    elif row>55:
        return 'Elder Age'
#Applying converted data into our dataset with new column - Age_Range

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

Out[17]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_D
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	A
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Pr
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Pr
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	A
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	A

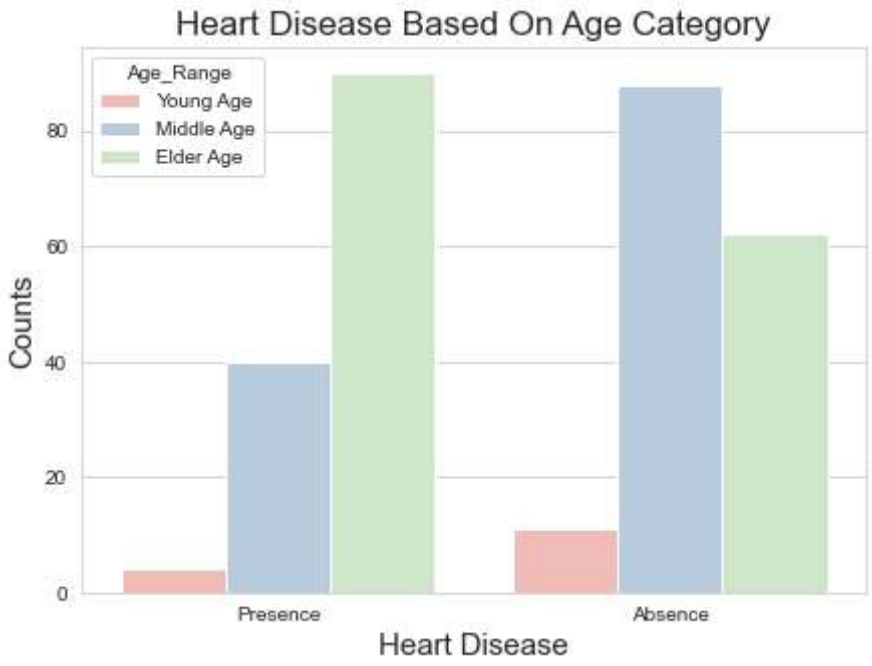
In [18]:

```
#Swarm Plot Creation of Gender Based Age Category using MatplotLib and Seaborn

plt.figure(figsize=(10,7))
sns.swarmplot(x='Age_Range', y='age', hue='sex1', data=data, order=['Young Age','Middle
plt.title('Gender Based Age Category', fontsize=17)
plt.xlabel('Age Category', fontsize=15)
plt.ylabel('Age', fontsize=15)
plt.show()
```



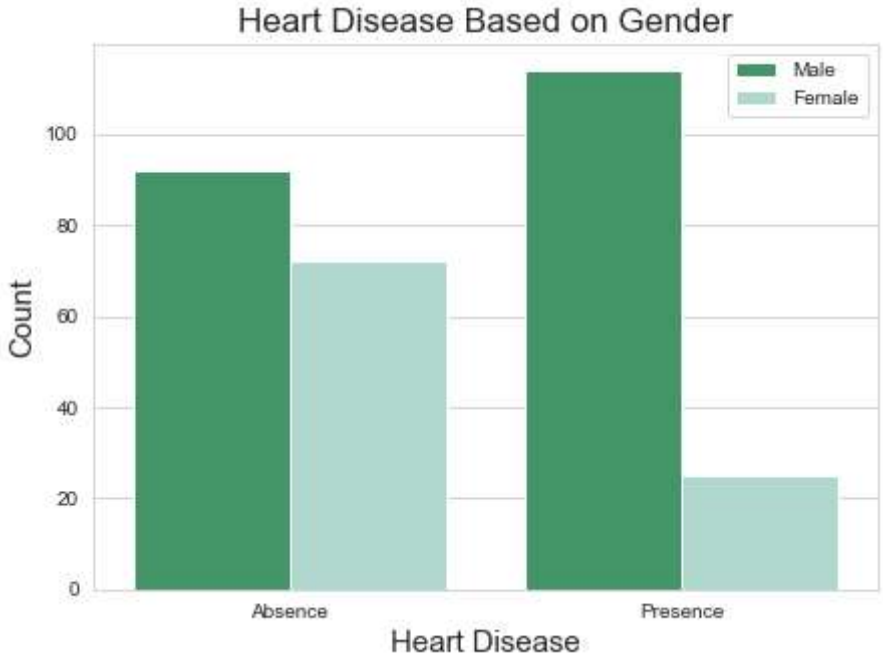
In our population number of males are more in middle age category and females are more in elder age category

In [19]:
```
#Count Plot Creation of Heart Disease Based On Age Category using MatplotLib and Seabor

plt.figure(figsize=(7,5))
hue_order=['Young Age', 'Middle Age', 'Elder Age']
sns.countplot(x='Heart_Disease', hue='Age_Range', data=data, order=['Presence','Absence
plt.title('Heart Disease Based On Age Category', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.show()
```



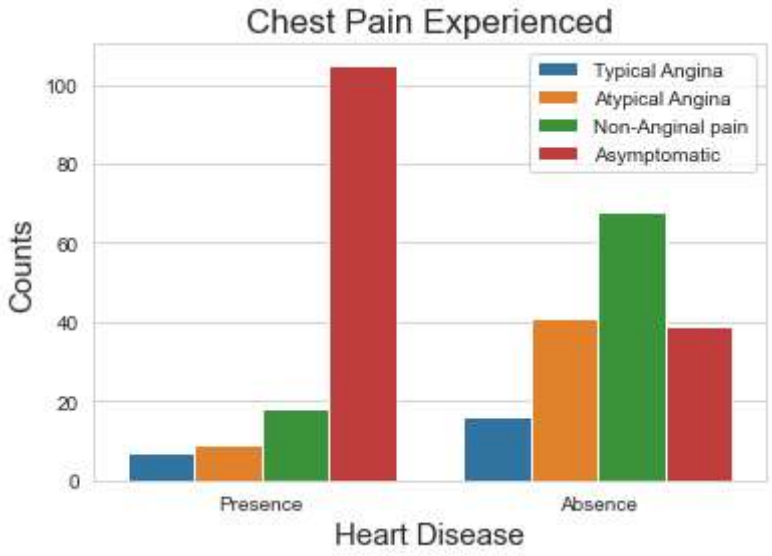
Elder age people are most affected by Heart Disease and Middle age people are mostly free from any kind of Heart Disease

In [20]:
```
#Count Plot Creation of Heart Disease Based on Gender using MatplotLib and Seaborn

plt.figure(figsize=(7,5))
sns.countplot(x=data['Heart_Disease'], hue='sex1', data=data, palette='BuGn_r')
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Count',fontsize=15)
plt.legend(labels=['Male','Female'])
plt.title('Heart Disease Based on Gender',fontsize=17)
plt.show()
```



In [21]:
```
#Count Plot Creation of Chest Pain Experienced using MatplotLib and Seaborn
```

```
sns.countplot(x=data['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()
```

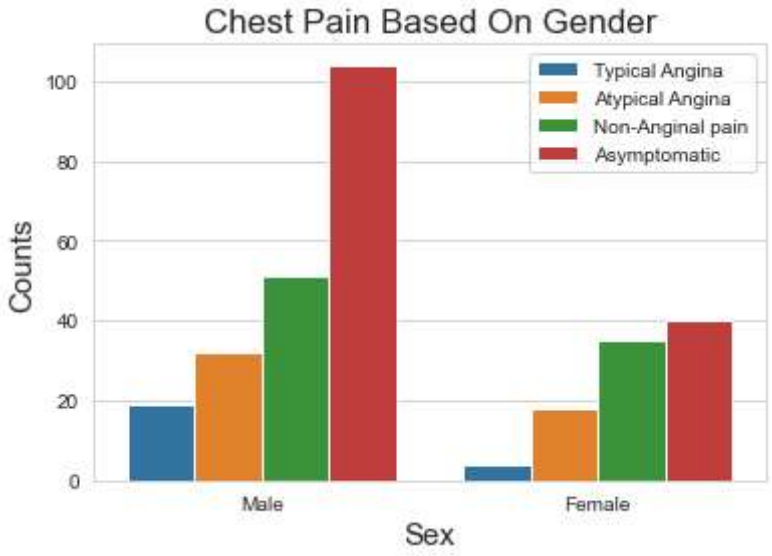


It seems people having asymptomatic chest pain have a higher chance of heart disease

Asymptomatic Chest pain means neither causing nor exhibiting symptoms of Heart disease.

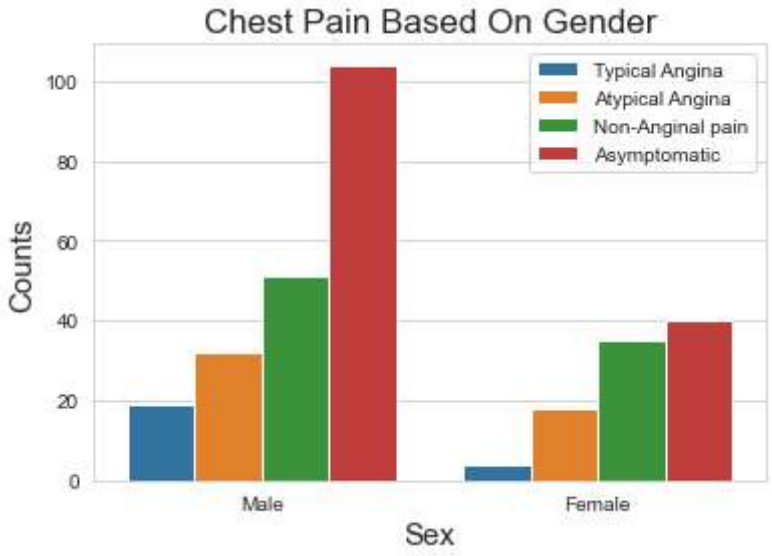
In [23]:

```
#Count Plot Creation of Chest Pain Based On Gender using MatplotLib and Seaborn

sns.countplot(x=data['sex1'], hue='cp', data=data)
plt.title('Chest Pain Based On Gender', fontsize=17)
plt.xlabel('Sex', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.legend(labels=['Typical Angina','Atypical Angina','Non-Anginal pain','Asymptomatic'
plt.show()
```



In [24]:

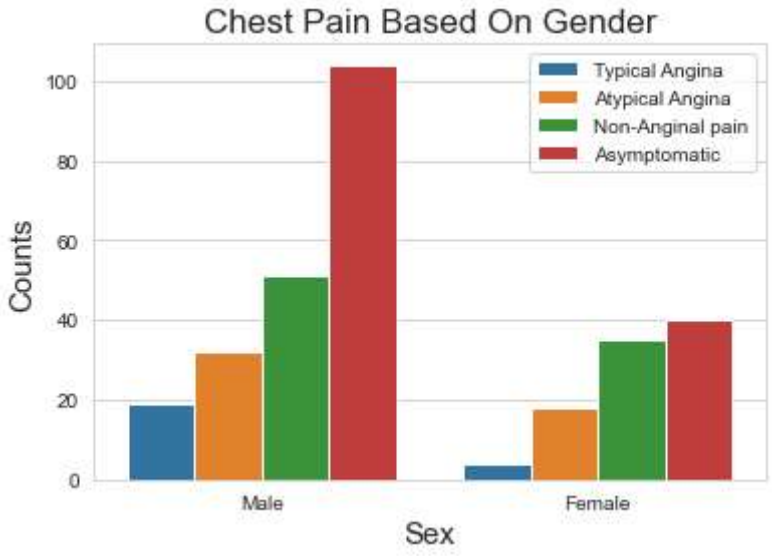
```
#Count Plot Creation of Chest Pain Based On Gender using MatplotLib and Seaborn

sns.countplot(x=data['sex1'], hue='cp', data=data)
plt.title('Chest Pain Based On Gender', fontsize=17)
plt.xlabel('Sex', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.legend(labels=['Typical Angina','Atypical Angina','Non-Anginal pain','Asymptomatic'
plt.show()
```



We can see that a higher number of men are suffering from Asymptomatic type of Chest Pain

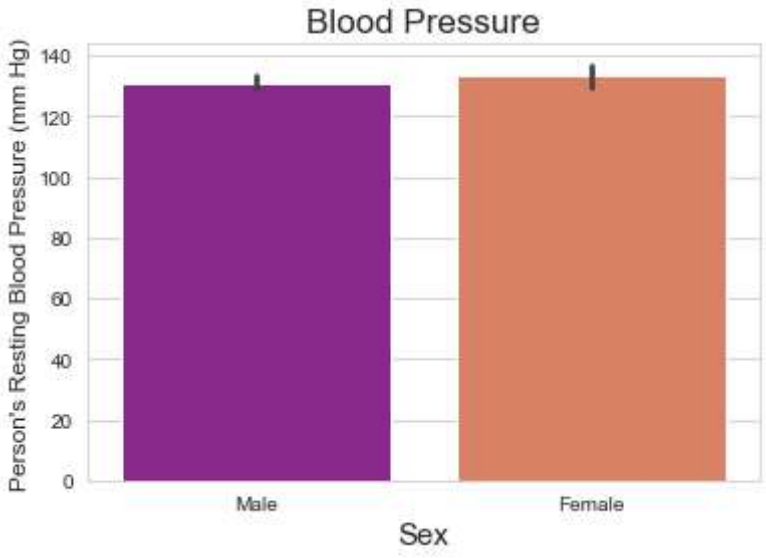
In [25]:

```
#Count Plot Creation of Chest Pain Based On Gender using MatplotLib and Seaborn

sns.countplot(x=data['sex1'], hue='cp', data=data)
plt.title('Chest Pain Based On Gender', fontsize=17)
plt.xlabel('Sex', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.legend(labels=['Typical Angina','Atypical Angina','Non-Anginal pain','Asymptomatic'
plt.show()
```



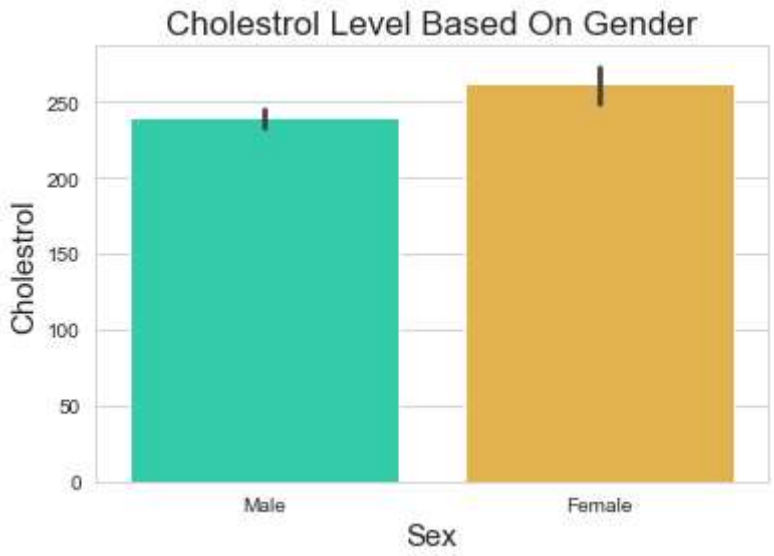
There is very high number of Asymptomatic Pain in Elderly age Category

In [26]:
```
#Bar Plot Creation of Person's Resting Blood Pressure (mm Hg) using MatplotLib and Seab

sns.barplot(x='sex1', y='trestbps', data=data, palette='plasma')
plt.title("Blood Pressure", fontsize=17)
plt.xlabel('Sex',fontsize=15)
plt.ylabel("Person's Resting Blood Pressure (mm Hg)", fontsize=12)
plt.show()
```



Blood Pressure Rate is almost equal in Males and Females

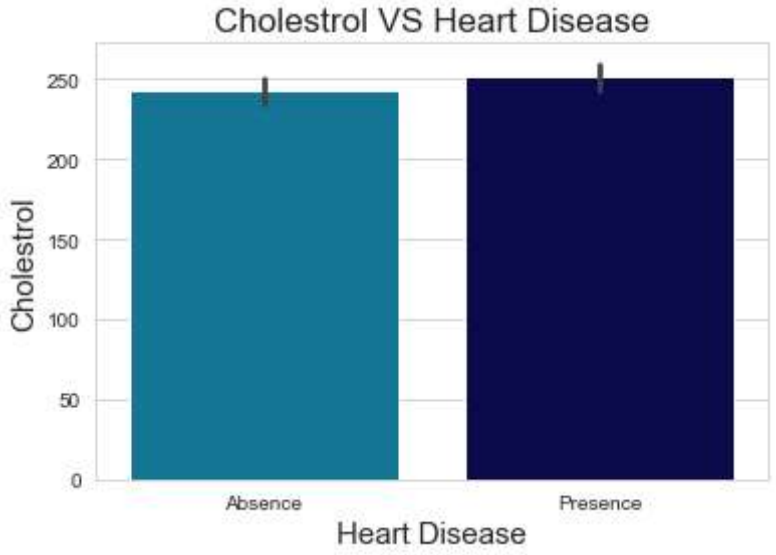
In [27]:
```
#Bar Plot Creation of Cholestrol Level Based On Gender using MatplotLib and Seaborn

sns.barplot(x='sex1', y='chol', data=data, palette='turbo')
plt.title("Cholestrol Level Based On Gender", fontsize=17)
plt.xlabel('Sex',fontsize=15)
plt.ylabel("Cholestrol", fontsize=15)
plt.show()
```



Females have little bit of higher cholesterol than Males

In [28]:
```
#Bar Plot Creation of Cholestrol VS Heart Disease using MatplotLib and Seaborn
```

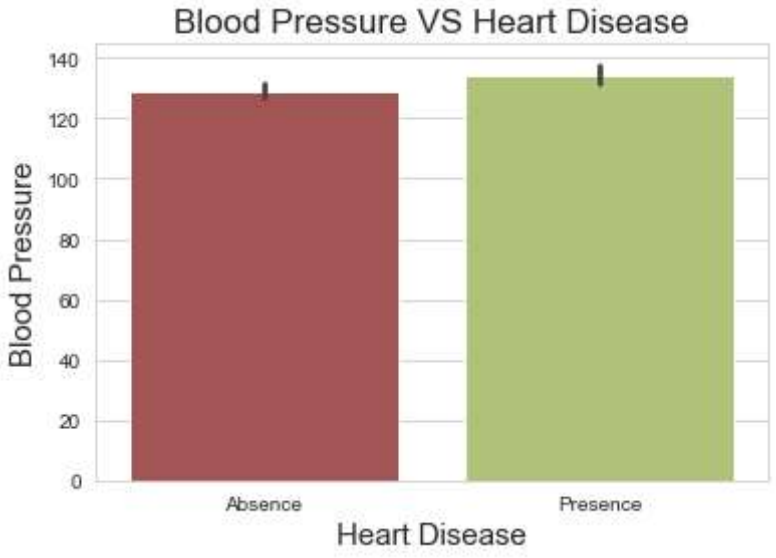
```
sns.barplot(x='Heart_Disease', y='chol', data=data, palette='ocean_r')
plt.title('Cholestrol VS Heart Disease', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Cholestrol', fontsize=15)
plt.show()
```



Higher Cholestrol Level results Chances Of Heart Disease

In [29]:

```
#Bar Plot Creation of Blood Pressure VS Heart Disease using MatplotLib and Seaborn

sns.barplot(x='Heart_Disease', y='trestbps', data=data, palette='tab20b_r')
plt.title('Blood Pressure VS Heart Disease', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Blood Pressure', fontsize=15)
plt.show()
```

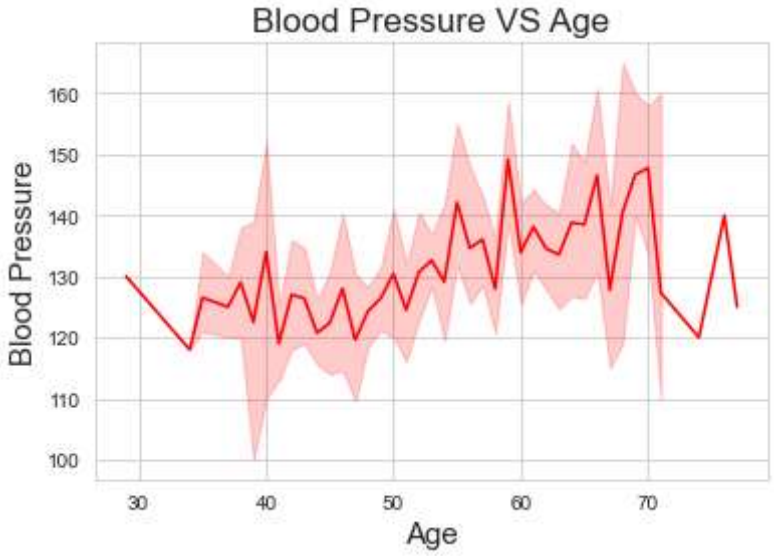


Higher Blood Pressure Level results Chances Of Heart Disease

In [30]:

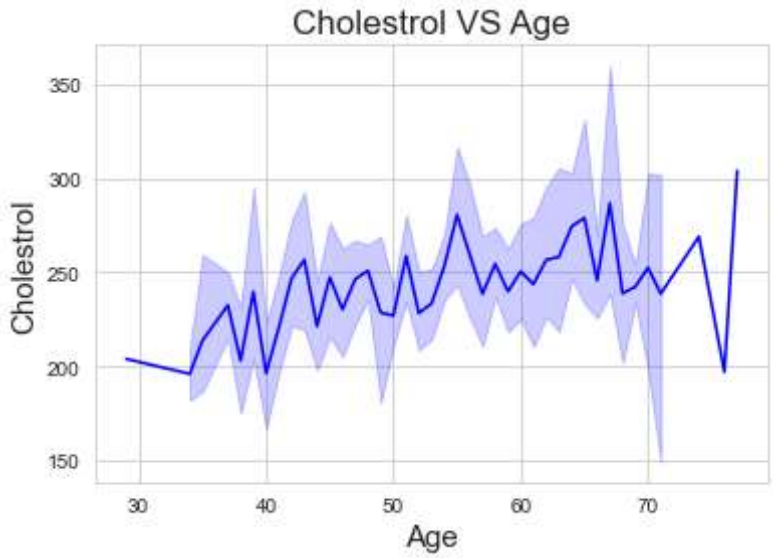
```
#Line Plot Creation of Blood Pressure VS Age using MatplotLib and Seaborn

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()
```



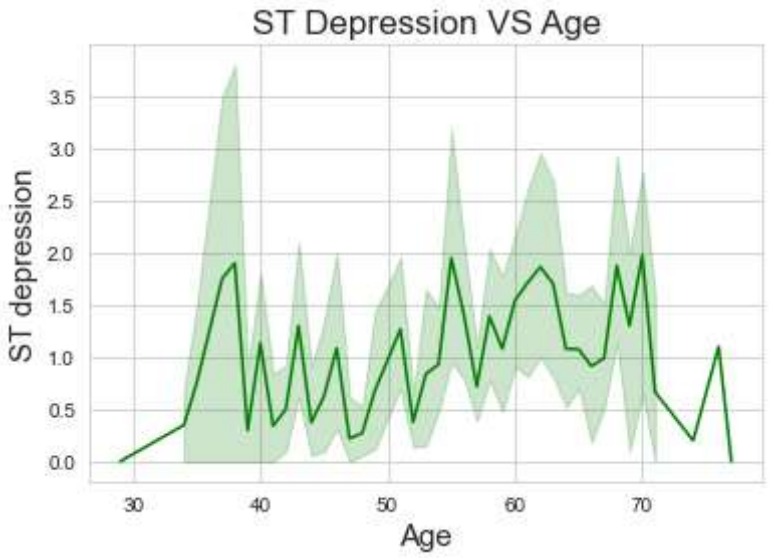
Here we can observe that Blood Pressure increases between age of 50 to 60 and somehow continue the pattern till 70

In [31]:
```
#Line Plot Creation of Cholestrol VS Age using MatplotLib and Seaborn

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()
```



In [32]:
```
#Line Plot Creation of ST Depression VS Age using MatplotLib and Seaborn

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

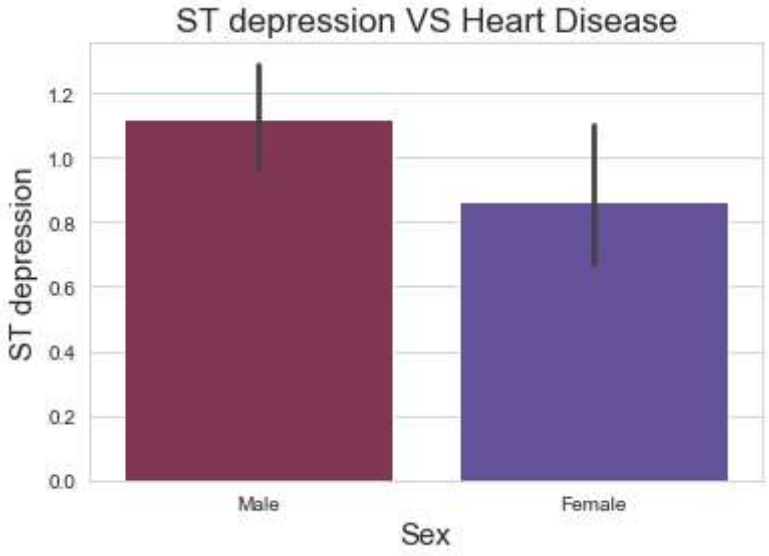
```
plt.ylabel('ST depression', fontsize=15)
plt.show()
```



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.

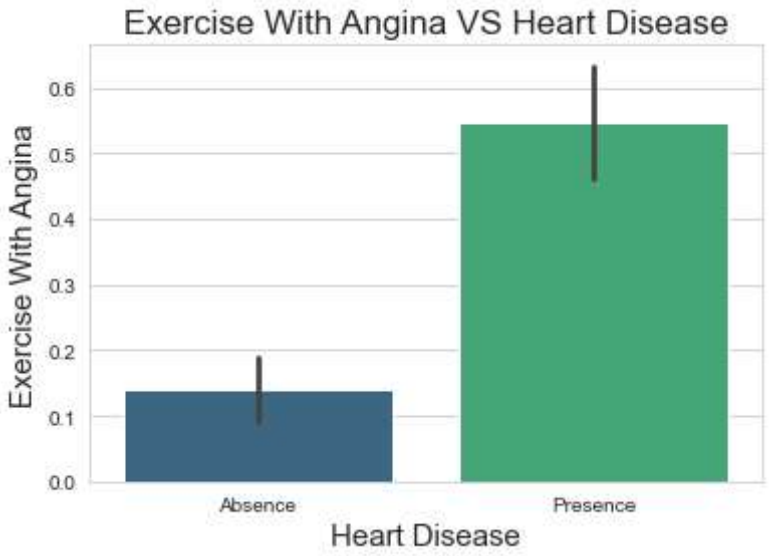
In [33]:
```
#Bar Plot Creation of ST depression VS Heart Disease using MatplotLib and Seaborn

sns.barplot(x='sex1', y='oldpeak', data=data, palette='twilight_r')
plt.title('ST depression VS Heart Disease', fontsize=17)
plt.xlabel('Sex', fontsize=15)
plt.ylabel('ST depression', fontsize=15)
plt.show()
```



Males are more prone to ST depression as compared to Females

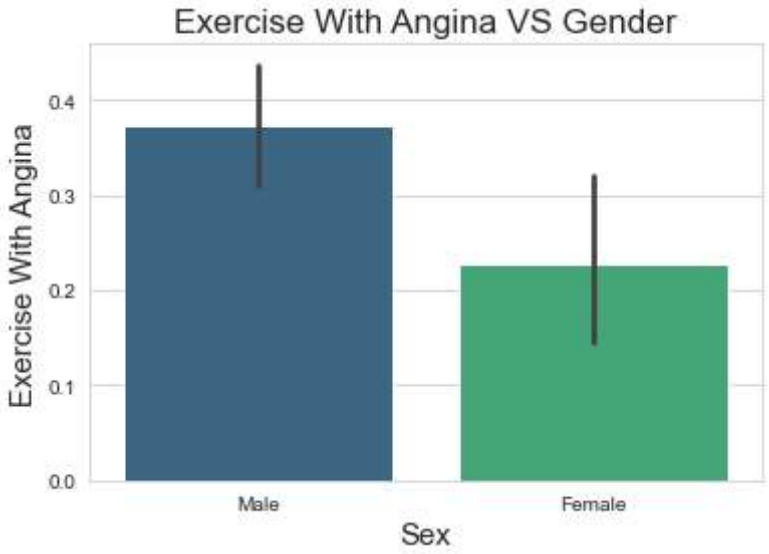
In [34]:
```
#Bar Plot Creation of Exercise With Angina VS Heart Disease using MatplotLib and Seabor

sns.barplot(x='Heart_Disease', y='exang', data=data, palette='viridis')
```

```
plt.title('Exercise With Angina VS Heart Disease', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Exercise With Angina', fontsize=15)
plt.show()
```



If you suffer from Angina, you may be concerned that exercise will make your symptoms worse

In [35]:
```
#Bar Plot Creation of Exercise With Angina VS Gender using MatplotLib and Seaborn

sns.barplot(x='sex1', y='exang', data=data, palette='viridis')
plt.title('Exercise With Angina VS Gender', fontsize=17)
plt.xlabel('Sex', fontsize=15)
plt.ylabel('Exercise With Angina', fontsize=15)
plt.show()
```

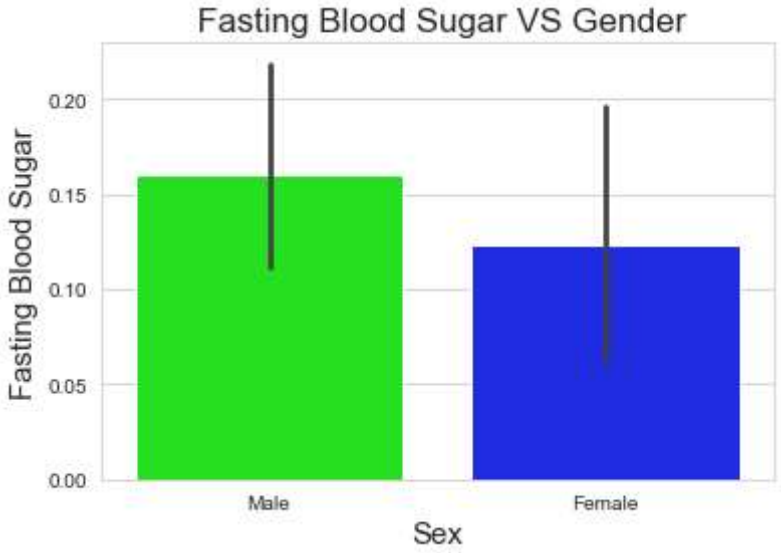


Males have have high Exercise Angina

A type of chest pain caused by reduced blood flow to the heart

In [36]:
```
#Bar Plot Creation of Fasting Blood Sugar VS Gender using MatplotLib and Seaborn
```

```
sns.barplot(y='fbs', x='sex1', data=data, palette='hsv')
plt.title(' Fasting Blood Sugar VS Gender', fontsize=17)
plt.xlabel('Sex', fontsize=15)
plt.ylabel('Fasting Blood Sugar', fontsize=15)
plt.show()
```



Males have high no of Fasting Blood Sugar over 120

In [ ]: