HEART DISEASE DIAGNOSTIC ANALYSIS

In [1]:
```
#IMPORTING LIBRARIES
```

In [2]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
sns.set_style('whitegrid')
```

In [3]:
```
# EXTRACT DATASET
data=pd.read_csv("C:\\Users\\DELL\\Downloads\\heart.csv")
data
```

Out[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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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

1025 rows × 14 columns

In [4]:
```
# ALL COULUMNS IN THE DATASET
```

In [5]:
```
data.columns
```

Out[5]:
```
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 PERSONS AGE IN YEARS
- SEX : TEH PERSONS SEX (1 = MALE, 0 = FEMALE)
- CP : THE CHEST PAIN EXPERIENCED (VALUE 1 : TYPICAL ANGINA, VALUE 2: ATYPICAL ANGINA, VALUE3: NON-ANGINAL PAIN, VALUE4: ASYMPTOMMATIC)
- TRESTBPS : THE PERSONS RESTING BLOOD PRESSURE (MM HG ON ADMISSION TO THE HOSPITAL)
- CHOL: THE PERSONS CHOLESTEROL MEASUREMENT IN MG/DI
- FBS : THE PERSONS FASTING BLOOD SUGAR(>120 MG/DI, 1=TRUE, 0 = FALSE)
- RESTECG : RESTING ELECTROCARDIOGRAPHIC MEASUREMENT (0 = NORMAL, 1 = HAVING ST-T WAVE ABNORMALLY, 2 = SHOWING PROBABLE OR DEFINITE LEFT VENTRICULAR HYPERNTROPHY BY ESTES`CRITERIA)
- THALACH : THE PERSONS MAXIMUM HEART RATE ACHIEVED
- EXANG : EXERCISE INDUCED ANGINA(1= YES, 0= NO)
- OLDPEAK : ST DEPRESSION INDUCED BY EXCERCISE RELATIVE TO REST
- SLOPE : THE SLOPE OF THE PEAK EXERCISE ST SEGMENT (VALUE 1: UPSTOPING, VALUE 2: FLAT, VALUE3: DOWNSLOPING)
- CA : THE NUMBER OF MAJOR VESSELS (0.3)
- THAL : A BLOOD DISORDER CALLED THALSSEMIA (3 =NORMAL, 6 = FIXED DEFECT, 7 = REVERSABLE DEFECT)
- TARGET : HEART DISEASE(0=NO, 1=YES)

In [6]:
```
#CHECK THE NULL VALUES
```

In [7]:
```
data.isnull().sum()
```

Out[7]:
```
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
```

NO MISSING VALUES IN DATASET

In [8]:
```
num=data.groupby('target').size()
num
```

Out[8]:
```
target
0    499
1    526
dtype: int64
```

In [9]:
```
# CONVERT NUMERICAL DATA INTO CATEGORICAL DATA
```

In [10]:
```
def heart_disease(row):
    if row==0:
        return 'Absence'
    elif row==1:
        return 'presence'
```

In [11]:
```
data['Heart_Disease']=data['target'].apply(heart_disease)
data.head()
```

Out[11]:

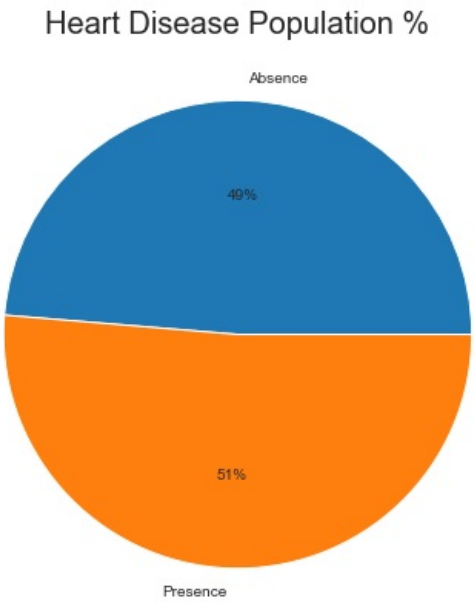
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	Heart_Disease
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0	Absence
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0	Absence
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0	Absence
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0	Absence
4	62	0	0	138	294	1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

In [12]:
```
hd=data.groupby('Heart_Disease')['target'].count()
hd
```

Out[12]:
```
Heart_Disease
Absence     499
presence    526
Name: target, dtype: int64
```

PIE CHART CREATING OF HEART DISEASE POPULATION % USING MATPLOTLIB

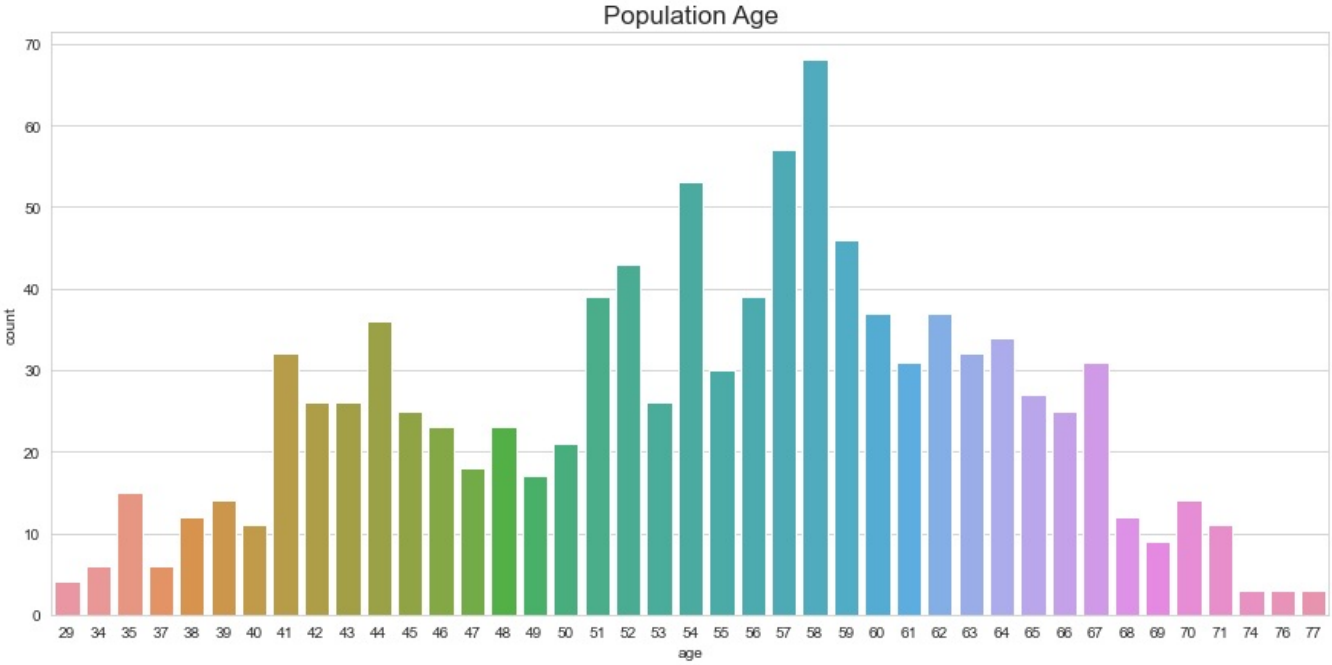
In [13]:
```
plt.figure(figsize=(10,7))
plt.pie(hd, labels=['Absence','Presence'], autopct='%0.0f%%')
plt.title('Heart Disease Population %', fontsize=20)
plt.show()
```



FROM THE OVERALL POPULATION PEOPLE HAVING HEART DISEASE (51%) ARE GREATER THAN THOSE WHO HAVE HEART DISEASE(49%)

In [14]:
```
# count plot creating 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.xlable('Age', fontsize=15)
plt.ylable('count', fontsize=15)
plt.show()
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Input In [14], in <cell line: 6>()
      4 sns.countplot(x='age', data=data)
      5 plt.title('Population Age', fontsize=17)
----> 6 plt.xlable('Age', fontsize=15)
      7 plt.ylable('count', fontsize=15)
      8 plt.show()

AttributeError: module 'matplotlib.pyplot' has no attribute 'xlable'
```



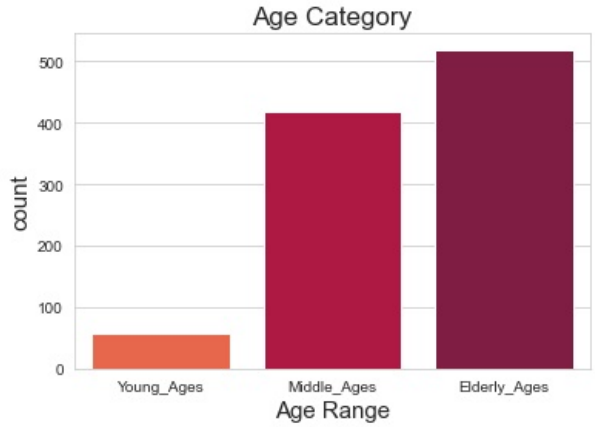
By examining the age distribution and segmenting it into different age groups such as the elderly, middle-aged, and young individuals, we can perform a more comprehensive analysis.

In [ ]:
```
# STATISTICAL ANALYSIS

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

In [15]:
```
# 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)), ('Middle Ages =',len(Middle_Ages)), ('Elderly Ages =',len(Elderly_Ages)
```

```
('Young Ages =', 57) ('Middle Ages =', 419) ('Elderly Ages =', 519)
```

In [16]:
```
# BAR PLOT CREATION OF AGE CATEGORY USING MATPLOTLIB & SEABORN

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



In [17]:
```
# CONVERTING NUMERICAL DATA INTO CATEGORICAL DATA

def gender(row):
    if row==1:
        return 'Male'
    elif row==0:
        return 'Female'
```

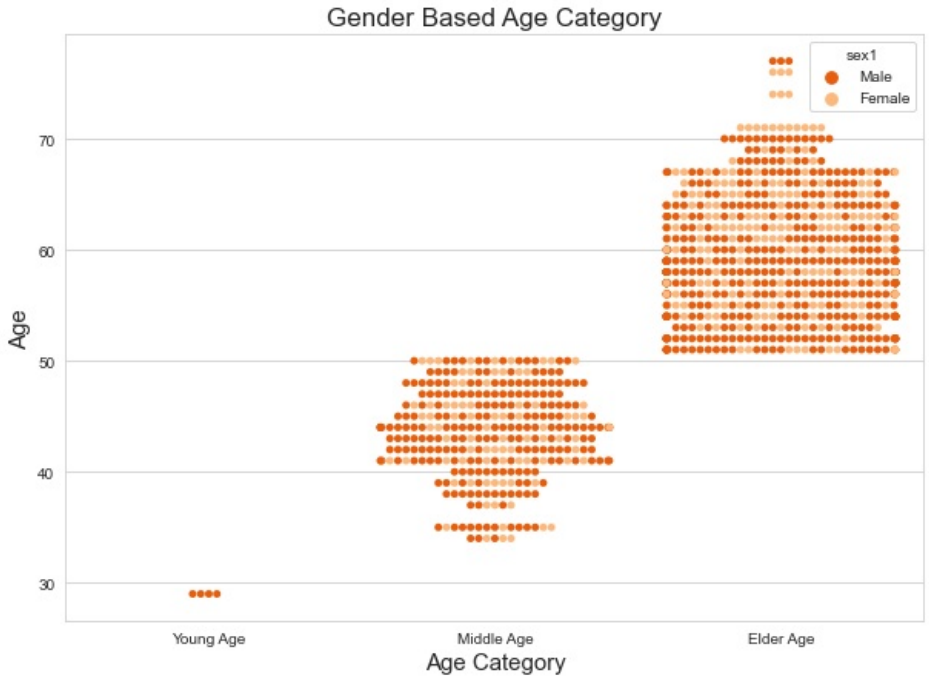
In [18]:
```
# APPLYING CONVERTED DATA INTO OUR DATASET WITH NEW COLUMN - SEX1

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

Out[18]:

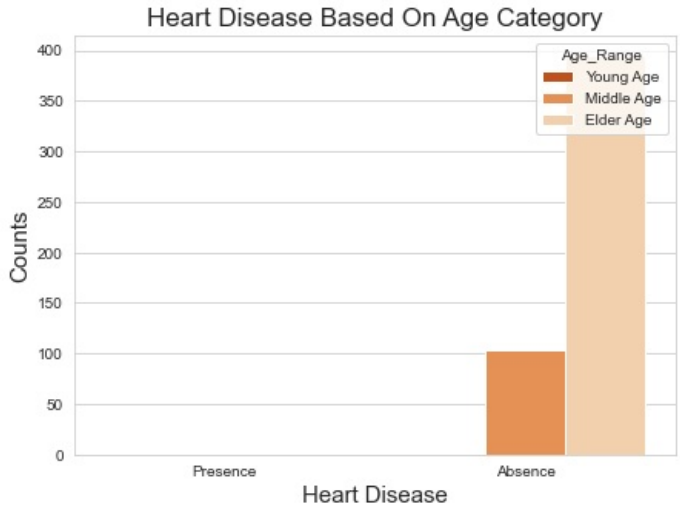
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	Heart_Disease	sex1
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0	Absence	Male
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0	Absence	Male
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0	Absence	Male
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0	Absence	Male
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0	Absence	Female

In [19]:
```
# SWARM PLOT CREATION OF GENDER BASED AGE CATEGORY USING MATPLOTLIB AND SEABORN

data['Age_Range'] = pd.cut(data['age'], bins=[0, 30, 50, 100], labels=['Young Age', 'Middle Age', 'Elder Age'])

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

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 24.1% of the points cannot
be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
```

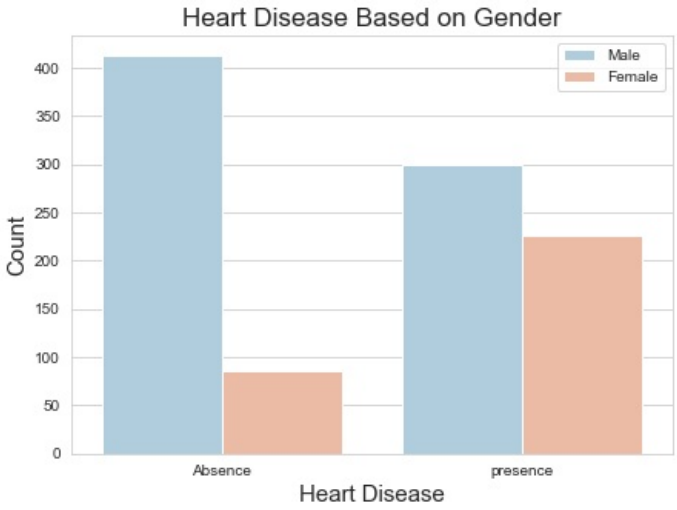


In [20]:
```
 # COUNT PLOT CREATION OF HEART DISEASE BASED ON AGE CATEGORY USING MATPLOTLIB AND SEABORN
plt.figure(figsize=(7,5))
hue_order=['Young Age', 'Middle Age', 'Elder Age']
sns.countplot(x='Heart_Disease', hue='Age_Range', data=data, order=['Presence', 'Absence'], hue_order=hue_order
plt.title('Heart Disease Based On Age Category', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.show()
```



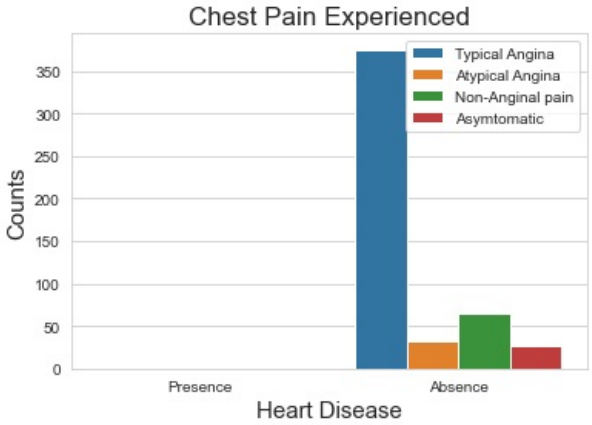
ELDER AGE PEOPLE ARE MOST AFFECTED BY HEART DISEASE AND MIDDLE AGE OF PEOPLE ARE MOSTLY FREE FROM ANY KIND OF HEART DISEASE

In [21]:
```
# 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='RdBu_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()
```



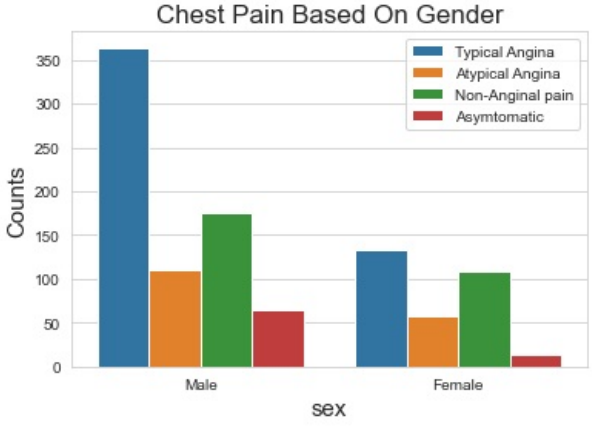
It can be observed that the incidence of heart disease is higher among males compared to females.

In [22]:
```
# COUNT PLOT CREATION OF CHEST PAIN EXPRIENCED 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','Asymtomatic'])
plt.show()
```



- IT SEEMS PEOPLE HAVING TYPICAL ANGINA CHEAST PAIN HAVE A HIGHER CHANCE OF HEART DISEASE
- TYPICAL ANGINA CHEAST PAIN MEANS OFTEN TRIGGERED BY PHYSICAL OR EMOTIONAL STRESS.

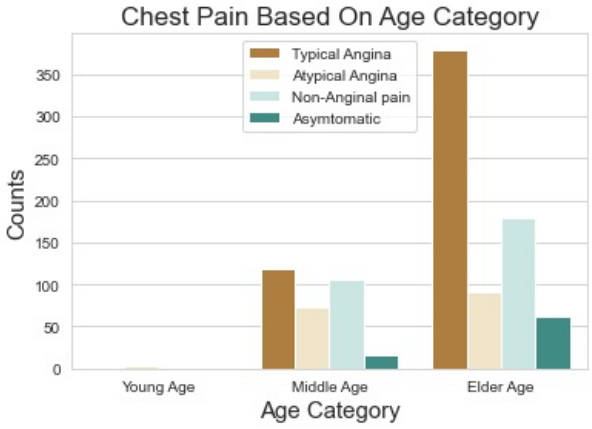
In [23]:
```
# COUNT PLOT CREATION OF CHEST PAIN BASED ON GENDER USING MATPLOTLIB AND SEABORN

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','Asymtomatic'])
plt.show()
```



WE CAN SEE THAT A HIGHER NUMBER OF MEN ARE SUFFERING FROM TYPICAL ANGINA TYPE OF CHEST PAIN

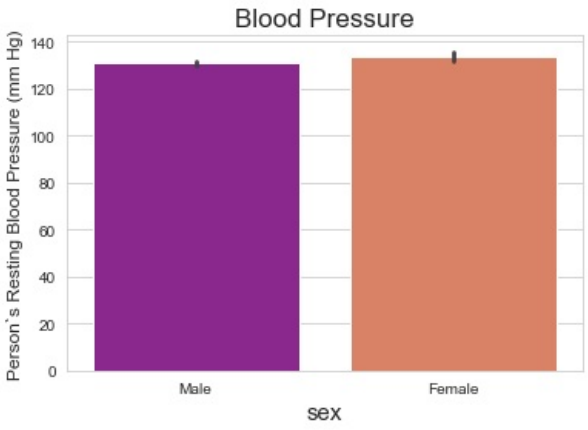
In [24]:
```
# COUNT PLOT CREATION OF CHEST PAIN BASED ON AGE CATEGORY USING MATPLOTLIB AND SEABORN
```

```
sns.countplot(x=data['Age_Range'], hue='cp', data=data, order=['Young Age','Middle Age','Elder Age'], palette='
plt.title('Chest Pain Based On Age Category', fontsize=17)
plt.xlabel('Age Category', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.legend(labels=['Typical Angina','Atypical Angina','Non-Anginal pain','Asymtomatic'])
plt.show()
```



- The incidence of typical angina pain is considerably higher among individuals in the elderly age group.

In [25]:
```
# BAR PLOT CREATION OF PERSONS RESTING BLOOD PRESURE USING MATPLOTLIB AND SEABORN

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()
```

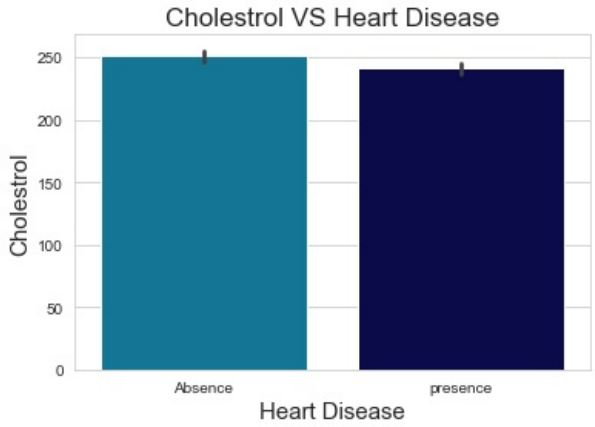


- The incidence of blood pressure is nearly identical between males and females

In [26]:
```
# 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("Cholesterol Level Based On Gender", fontsize=17)
plt.xlabel('sex', fontsize=15)
plt.ylabel("Cholesterol", fontsize=15)
plt.show()
```

- FEMALES HAVE LITTLE BIT OF HIGHER CHOLESTEROL THAN MALES

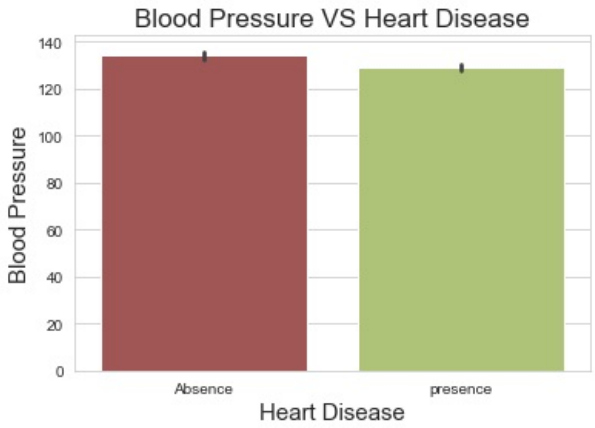
In [27]:
```
# BAR PLOT CREATION OF CHOLESTROL VS HEART DISEASE USING MATPLOTLIB AND SEABORN

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

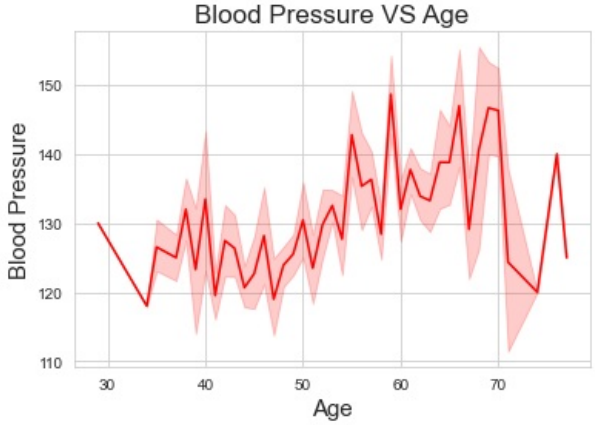
In [28]:
```
# BAR PLOT CREATION OF BLOOD PRESSURE VS HEART DISEASE USING MATPLOTLIB AND SEABORN

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
```

Out[28]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



- HIGHER BLOOD PRESSURE LEVEL RESULTS CHANCES OF HEART DISEASE

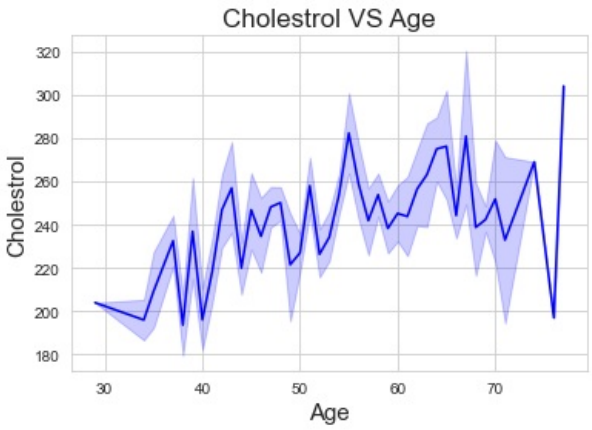
In [29]:
```
# LINE PLOT CREATION OF BLOOD PRESSURE VS AGE USING MATPLOTLIB AND SEABORN

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 SEE THAT BLOOD PRESSURE INCREASES BETWEEN AGE OF 50 TO 60 AND SOME CONTINUE THE

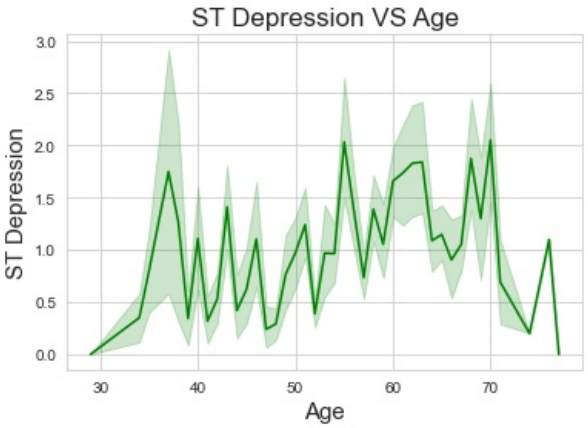
- HERE WE CAN SEE THAT BLOOD PRESSURE INCREASES BETWEEN AGE OF 50 TO 60 AND SOME CONTINUE THE PATTERN TILL 70

In [30]:
```
# 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()
```



- CHOLESTROL INCREASING IN THE AGE GROUP OF 50 TO 60.

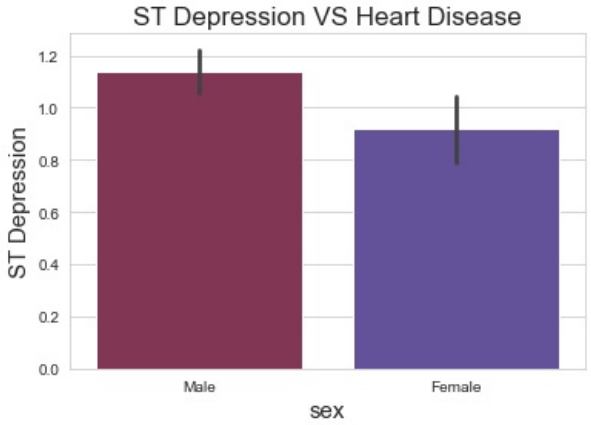
In [31]:
```
#LINE PLOT CREATION OF STRESS DEPRESSION VS AGE USING MATPLOTLIB AND SEABORN

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()
```



- STRESS DEPRESSION MOSTLY INCREASES THE AGE GROUP OF 30 TO 40.
- STRESS DEPRESSION REFERS TO A FINDING ON AN ELECTROCARDIOGRAM, WHERIN THE TRACE IN THE ST SEGMENT IS ABNOMALLY LOW BELOW THE BASELINE

In [32]:
```
# 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
```

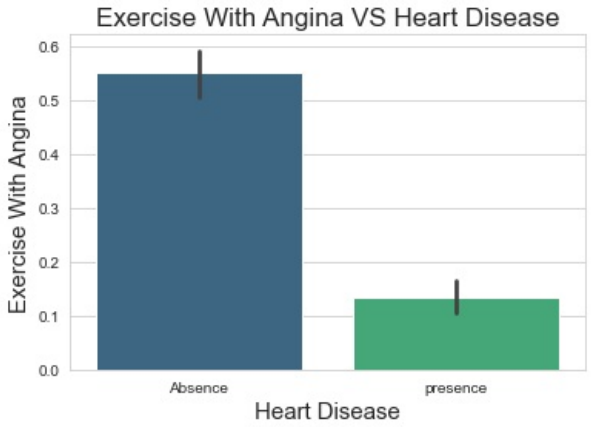
Out[32]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



- Males have a higher susceptibility to stress-related depression compared to females

In [33]:
```
# BAR PLOT CREATION OF EXERCISE WITH ANGINA VS HEART DISEASE USING MATPLOTLIB AND SEABORN

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
```

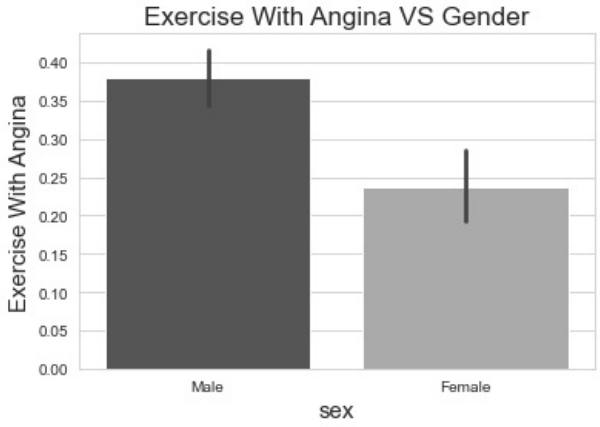
Out[33]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



- If you have angina, you might be worried that engaging in physical activity could exacerbate your symptoms.

In [34]:
```
# BAR PLOT CREATION OF EXERCISE WITH ANGINA VS GENDER USING MATPLOTLIB AND SEABORN

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

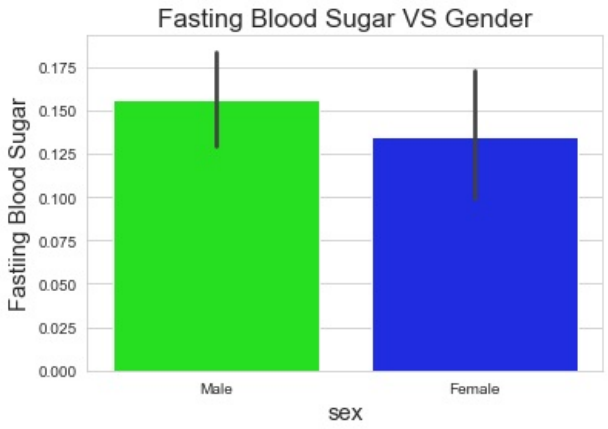
Out[34]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



- Men are more likely to experience angina symptoms during exercise.

In [35]:
```
# 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('Fastiing Blood Sugar', fontsize=15)
plt.show
```

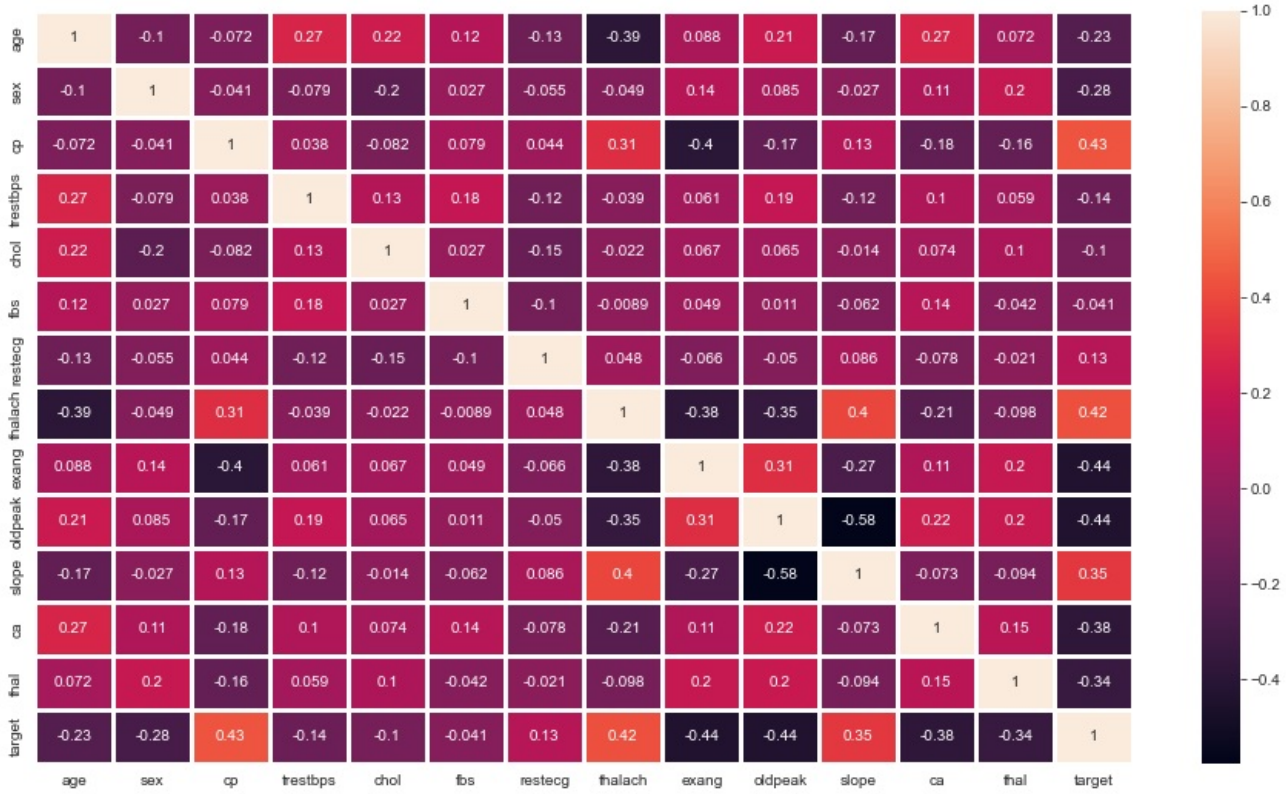
Out[35]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



- males exhibit a higher prevalence of elevated fasting blood sugar levels exceeding 120.

In [36]:
```
# HEATMAP CREATION USING SEABORN

plt.figure(figsize=(16,9))
sns.heatmap(data.corr(), annot=True, linewidth=3)
```

Out[36]:
```
<AxesSubplot:>
```



In [ ]: