

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

```
In [5]: file_path = 'C:\\Users\\Sejal\\Desktop\\data science projects\\heart disease\\H  
heart_disease_data = pd.read_csv(file_path)
```

```
In [6]: missing_values = heart_disease_data.isnull().sum()
print(missing_values)

# Get basic statistics of the dataset
basic_stats = heart_disease_data.describe()
print(basic_stats)
```

```
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
```

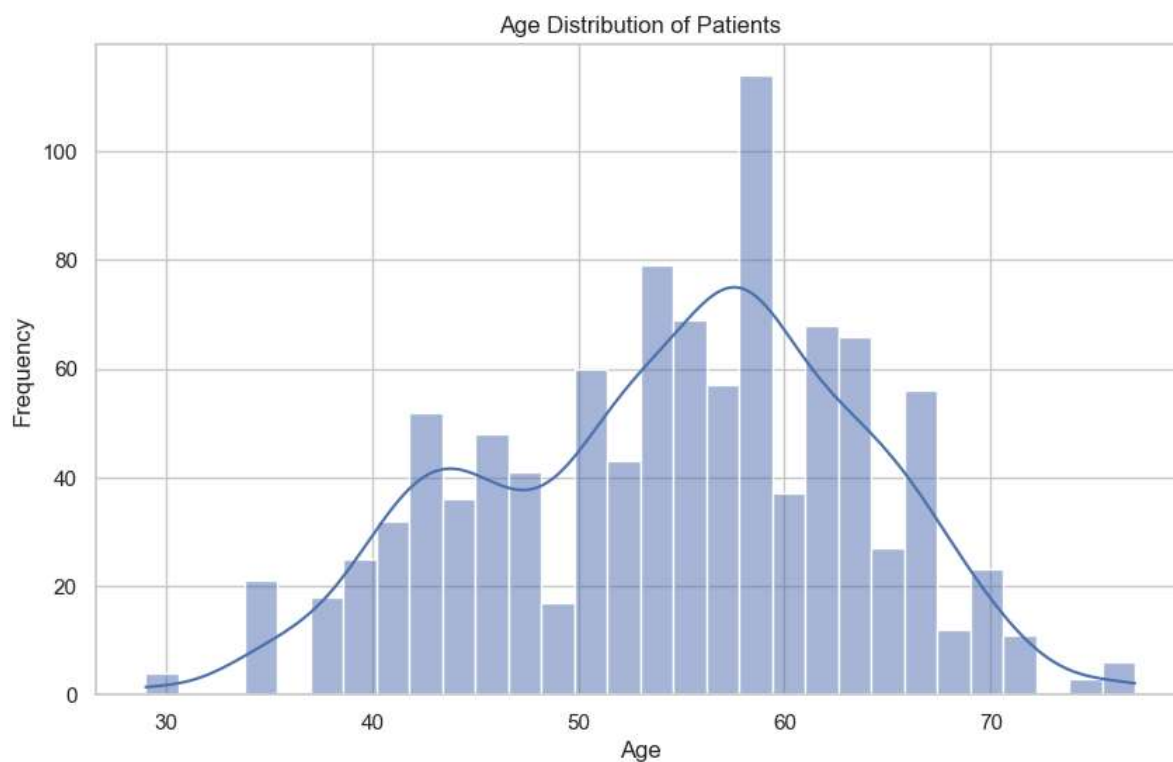
	age	sex	cp	trestbps	chol \
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	54.434146	0.695610	0.942439	131.611707	246.000000
std	9.072290	0.460373	1.029641	17.516718	51.59251
min	29.000000	0.000000	0.000000	94.000000	126.000000
25%	48.000000	0.000000	0.000000	120.000000	211.000000
50%	56.000000	1.000000	1.000000	130.000000	240.000000
75%	61.000000	1.000000	2.000000	140.000000	275.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000

	fbs	restecg	thalach	exang	oldpeak \
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	0.149268	0.529756	149.114146	0.336585	1.071512
std	0.356527	0.527878	23.005724	0.472772	1.175053
min	0.000000	0.000000	71.000000	0.000000	0.000000
25%	0.000000	0.000000	132.000000	0.000000	0.000000
50%	0.000000	1.000000	152.000000	0.000000	0.800000
75%	0.000000	1.000000	166.000000	1.000000	1.800000
max	1.000000	2.000000	202.000000	1.000000	6.200000

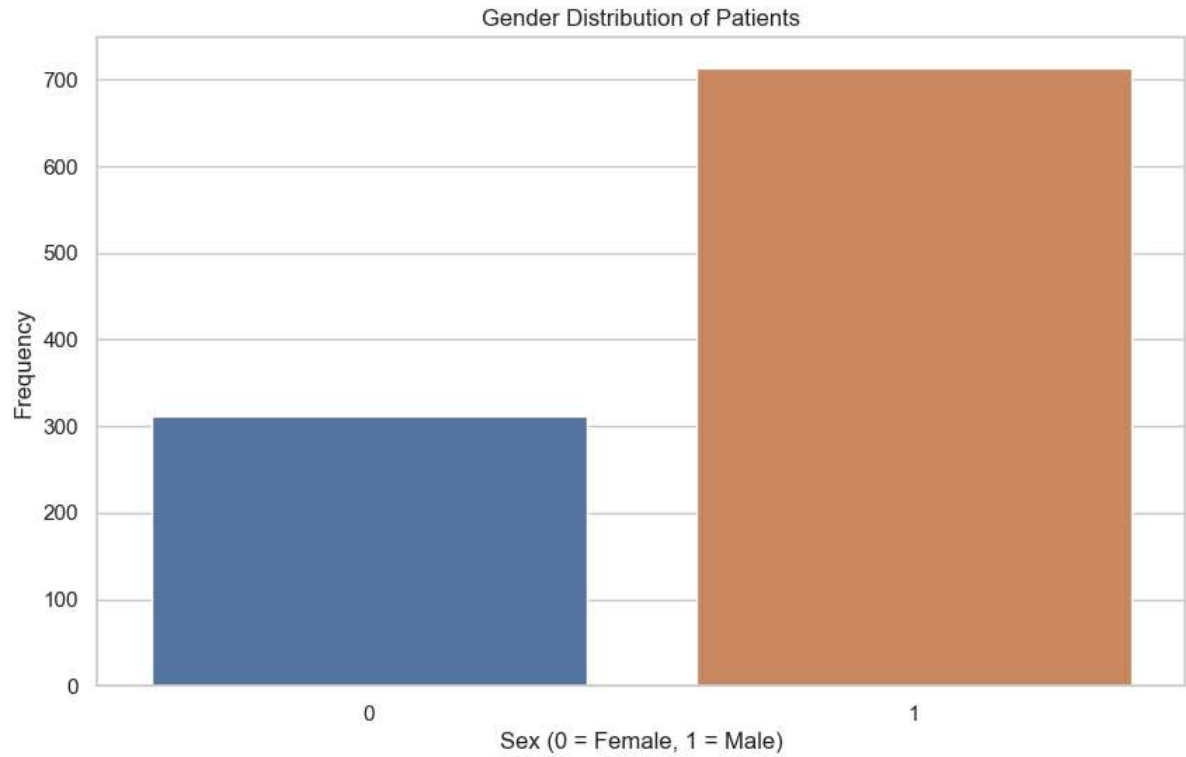
	slope	ca	thal	target
count	1025.000000	1025.000000	1025.000000	1025.000000
mean	1.385366	0.754146	2.323902	0.513171
std	0.617755	1.030798	0.620660	0.500070
min	0.000000	0.000000	0.000000	0.000000
25%	1.000000	0.000000	2.000000	0.000000
50%	1.000000	0.000000	2.000000	1.000000
75%	2.000000	1.000000	3.000000	1.000000
max	2.000000	4.000000	3.000000	1.000000

```
In [7]: # Set the style for the plots  
sns.set(style="whitegrid")
```

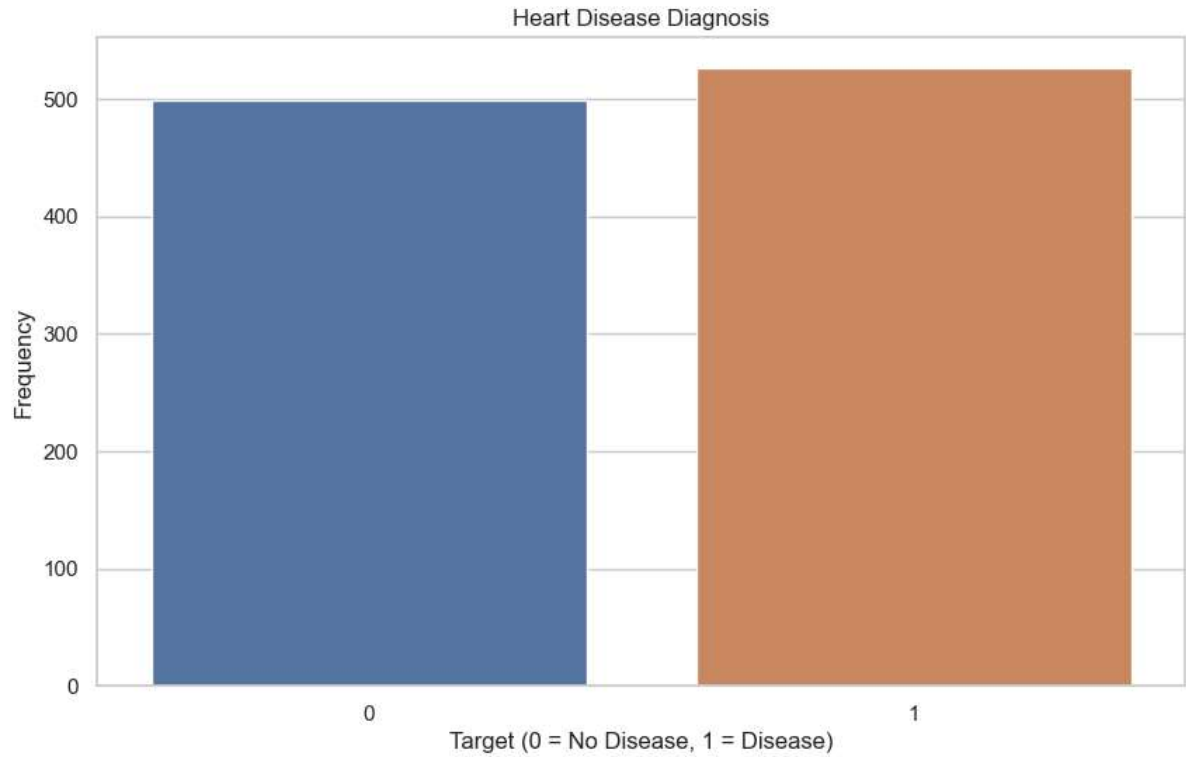
```
In [8]: # Plot the distribution of age  
plt.figure(figsize=(10, 6))  
sns.histplot(heart_disease_data['age'], bins=30, kde=True)  
plt.title('Age Distribution of Patients')  
plt.xlabel('Age')  
plt.ylabel('Frequency')  
plt.show()
```



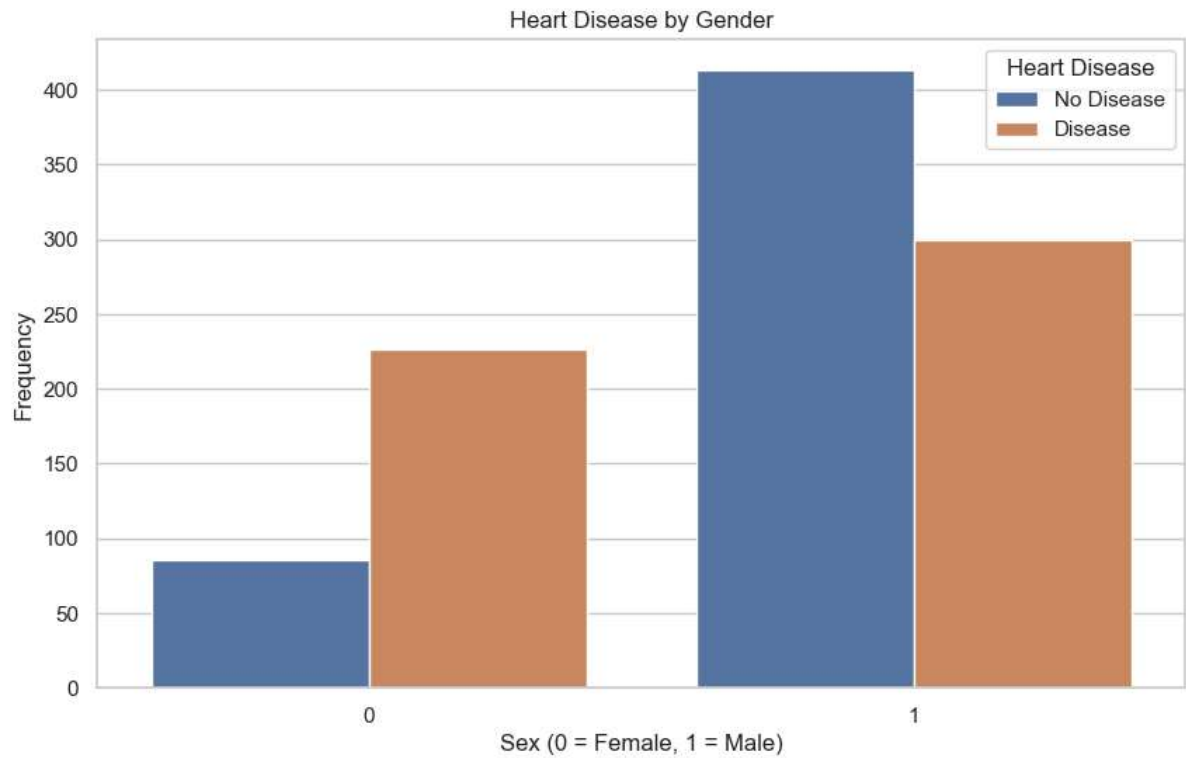
```
In [9]: # Plot the distribution of sex
plt.figure(figsize=(10, 6))
sns.countplot(x='sex', data=heart_disease_data)
plt.title('Gender Distribution of Patients')
plt.xlabel('Sex (0 = Female, 1 = Male)')
plt.ylabel('Frequency')
plt.show()
```



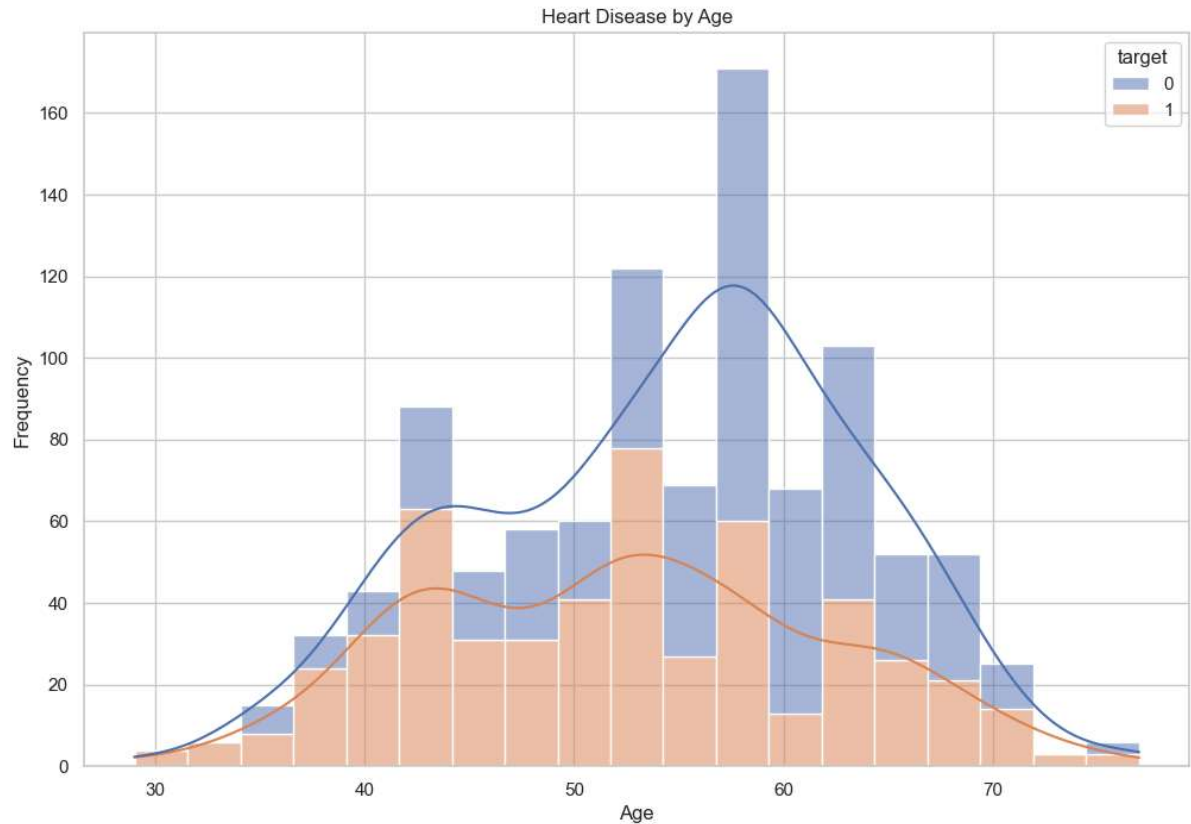
```
In [10]: # Plot the distribution of target variable
plt.figure(figsize=(10, 6))
sns.countplot(x='target', data=heart_disease_data)
plt.title('Heart Disease Diagnosis')
plt.xlabel('Target (0 = No Disease, 1 = Disease)')
plt.ylabel('Frequency')
plt.show()
```



```
In [11]: # Plot the relationship between gender and heart disease
plt.figure(figsize=(10, 6))
sns.countplot(x='sex', hue='target', data=heart_disease_data)
plt.title('Heart Disease by Gender')
plt.xlabel('Sex (0 = Female, 1 = Male)')
plt.ylabel('Frequency')
plt.legend(title='Heart Disease', loc='upper right', labels=['No Disease', 'Disease'])
plt.show()
```



```
In [12]: # Plot the relationship between age and heart disease
plt.figure(figsize=(12, 8))
sns.histplot(data=heart_disease_data, x='age', hue='target', multiple='stack',
plt.title('Heart Disease by Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



In []: