Data Loading and Exploration

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

         data = pd.read_csv(r"C:\Users\hp\Downloads\Heart Disease data\Heart Disease data\He
         with open(r"C:\Users\hp\Downloads\Heart Disease data\Heart Disease data\Dataset Det
             data_info = f.read()
```

```
In [2]:  print("First few rows of the dataset:")
         print(data.head())
```

```
First few rows of the dataset:
   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
0   2     3       0
1   0     3       0
2   0     3       0
3   1     3       0
4   3     2       0
```

Basic information about the dataset

```
In [3]:  print("\nDataset Information:\n")
         print(data.info())
         print("\nGiven info:\n\n",data_info)
```

```
Dataset Information:

<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
None


Given info:

 Attribute Information:

age
sex
chest pain type (4 values)
resting blood pressure
serum cholestoral in mg/dl
fasting blood sugar > 120 mg/dl
resting electrocardiographic results (values 0,1,2)
maximum heart rate achieved
exercise induced angina
oldpeak = ST depression induced by exercise relative to rest
the slope of the peak exercise ST segment
number of major vessels (0-3) colored by flourosopy
thal: 0 = normal; 1 = fixed defect; 2 = reversable defect
```

Data Transformation and Handling Missing Values

```python
In [4]: print("\nMissing Values:")
        print(data.isnull().sum())

        data = data.dropna()
```
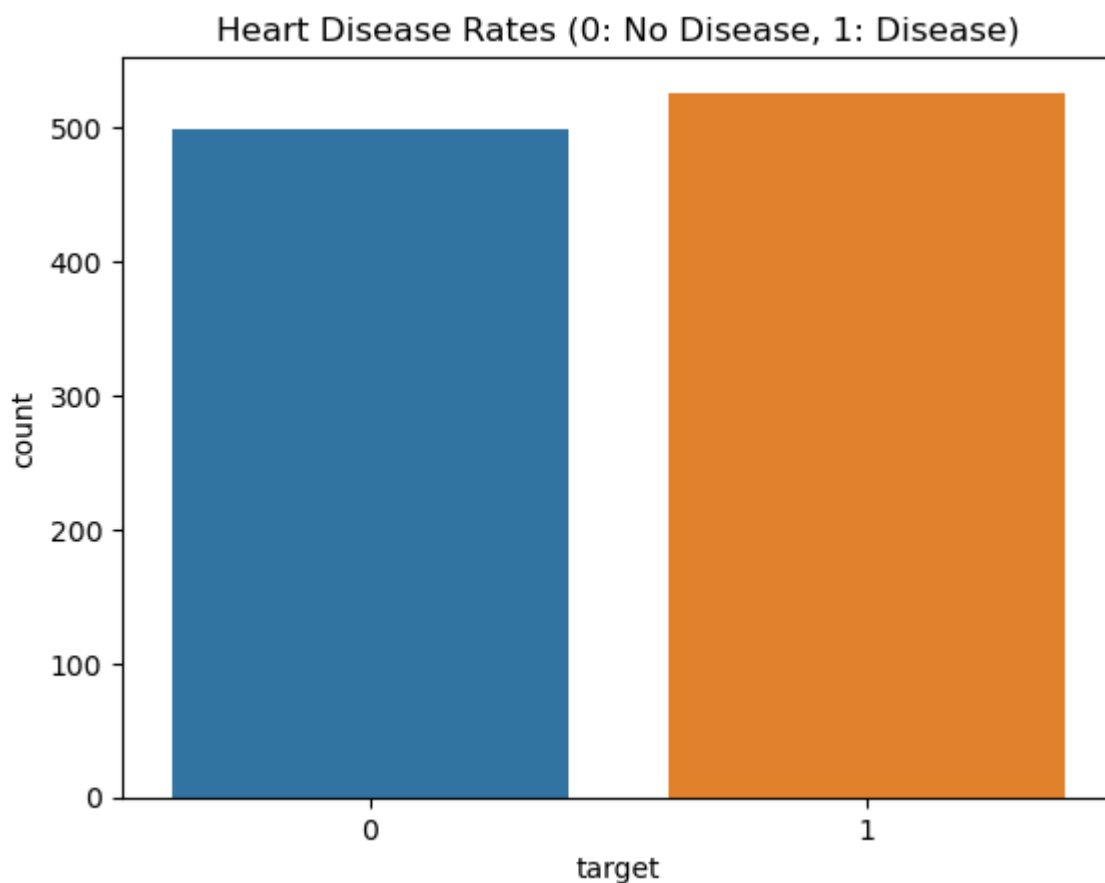
```
Missing Values:
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```
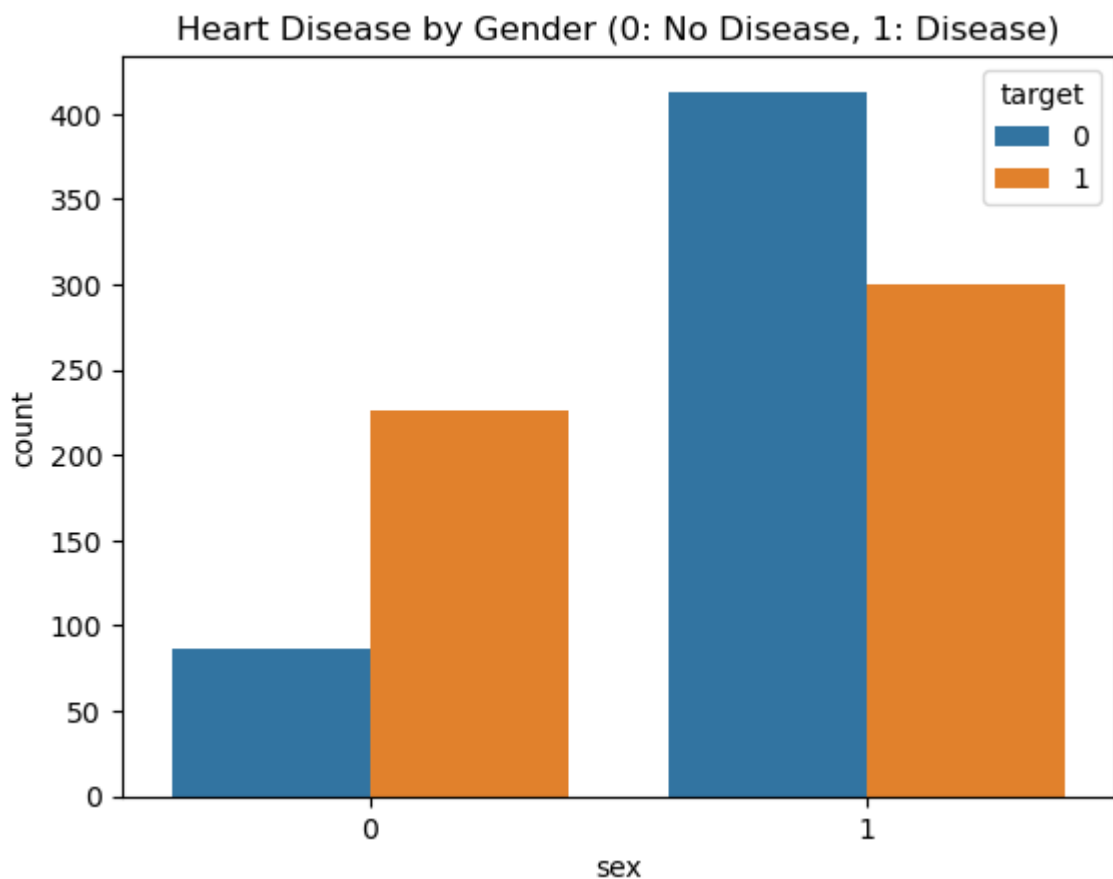
Visualization of Heart Disease Data

```
In [5]:  sns.countplot(x='target', data=data)
         plt.title('Heart Disease Rates (0: No Disease, 1: Disease)')
         plt.show()
```



```
In [6]:  # Visualize heart disease by gender
         sns.countplot(x='sex', hue='target', data=data)
         plt.title('Heart Disease by Gender (0: No Disease, 1: Disease)')
         plt.show()
```

Heart Disease by Gender (0: No Disease, 1: Disease)

In [7]:
```python
# Visualize heart disease by age
sns.histplot(x='age', hue='target', data=data, bins=20, kde=True)
plt.title('Heart Disease by Age')
plt.show()
```

Heart Disease by Age
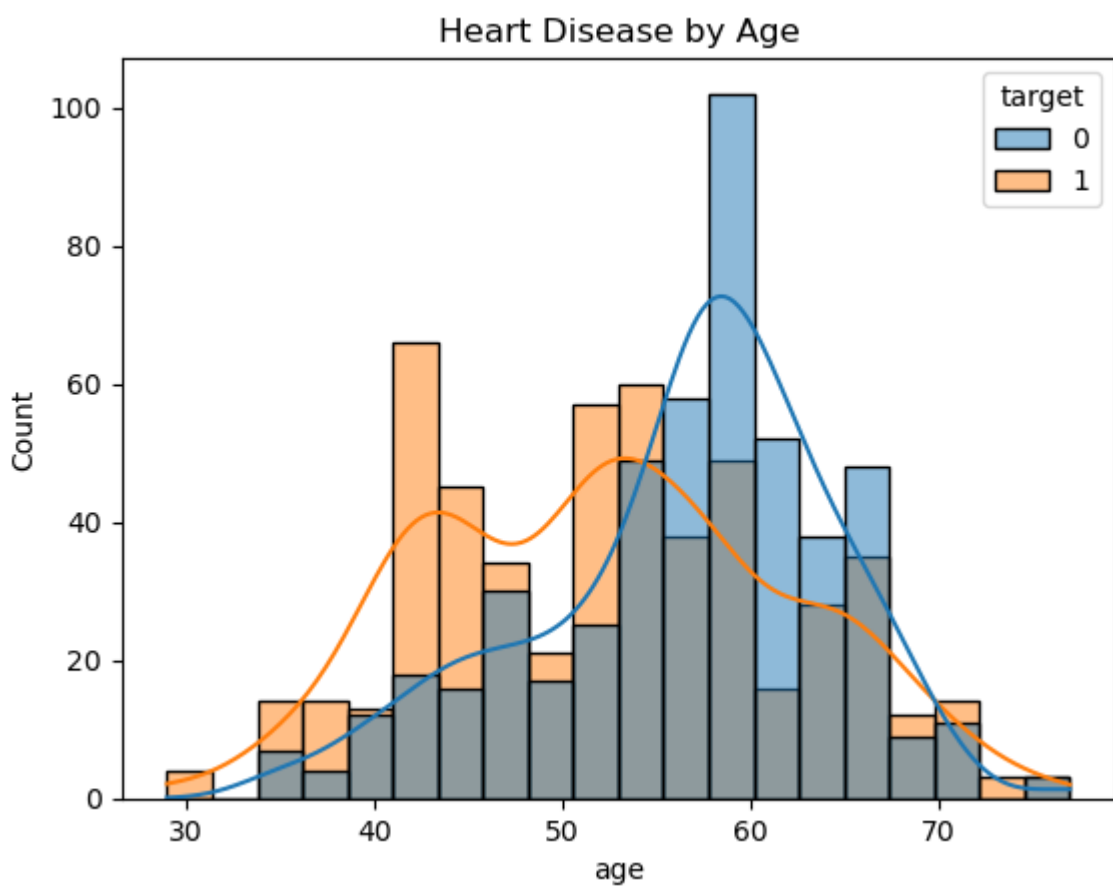
Correlation Heatmap and Distribution Plots

In [8]:
```python
# Correlation Heatmap
plt.figure(figsize=(8, 8))
sns.heatmap(data.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```


Correlation Heatmap
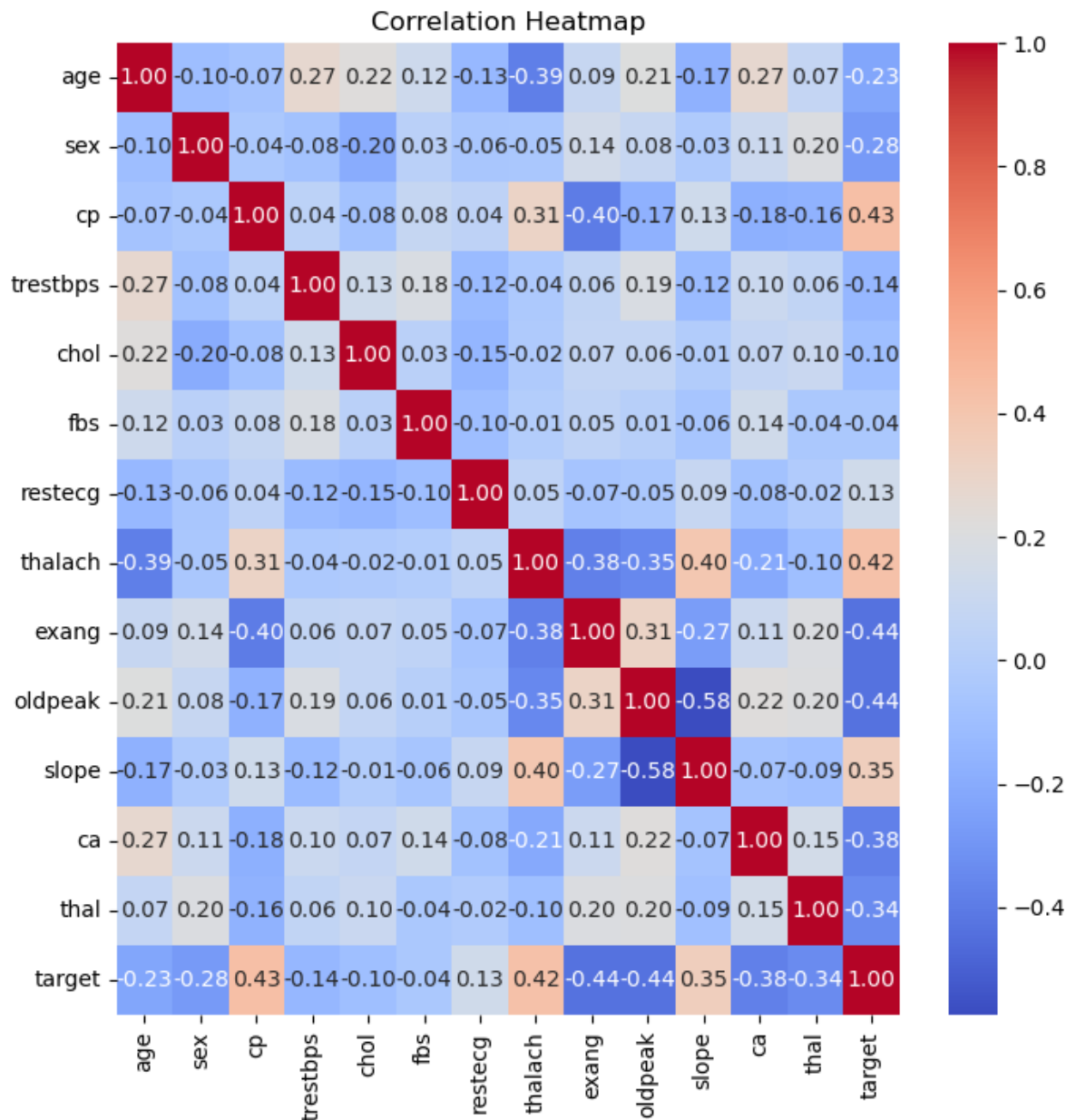
In [9]:
```python
# Distribution plots for numerical features
numerical_features = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak']
for feature in numerical_features:
    plt.figure(figsize=(6, 4))
    sns.histplot(x=feature, data=data, kde=True)
    plt.title(f'Distribution of {feature}')
    plt.show()
```

Distribution of age

Distribution of trestbps

Distribution of chol

Distribution of thalach

Distribution of oldpeak
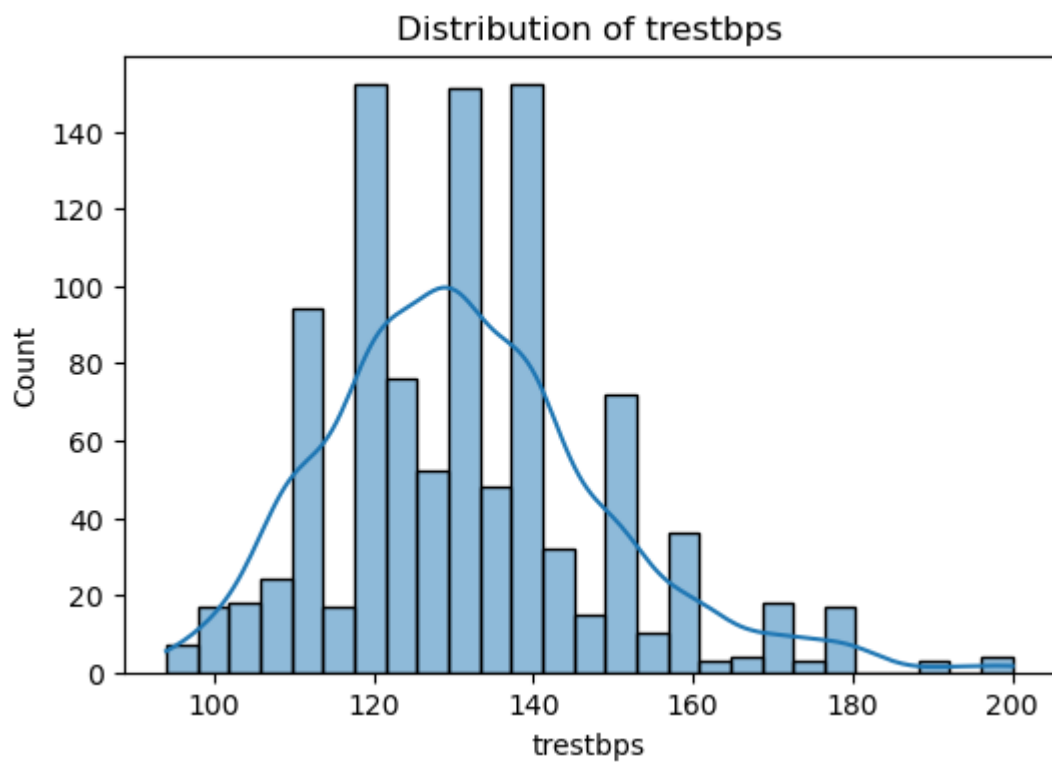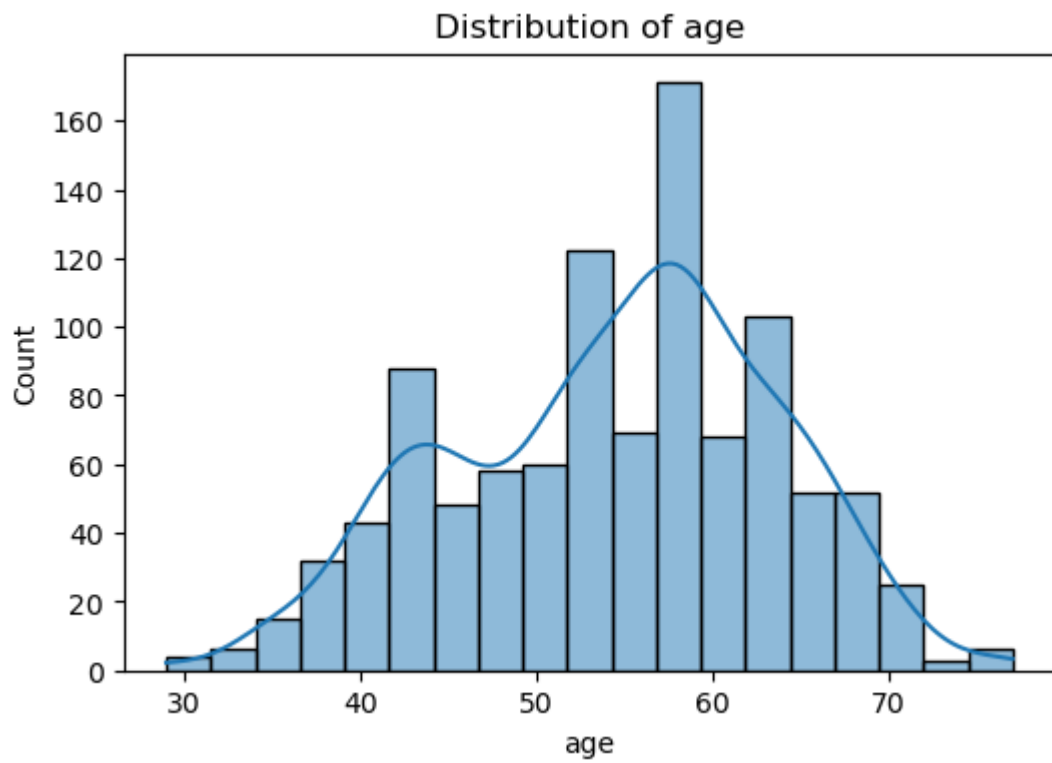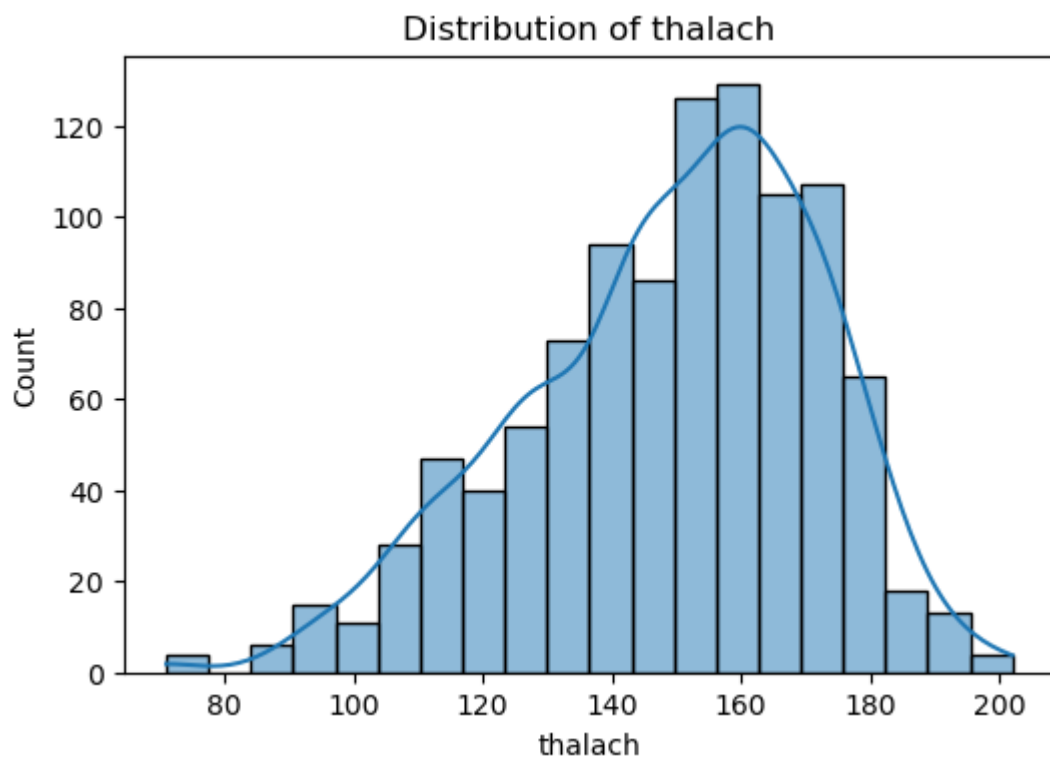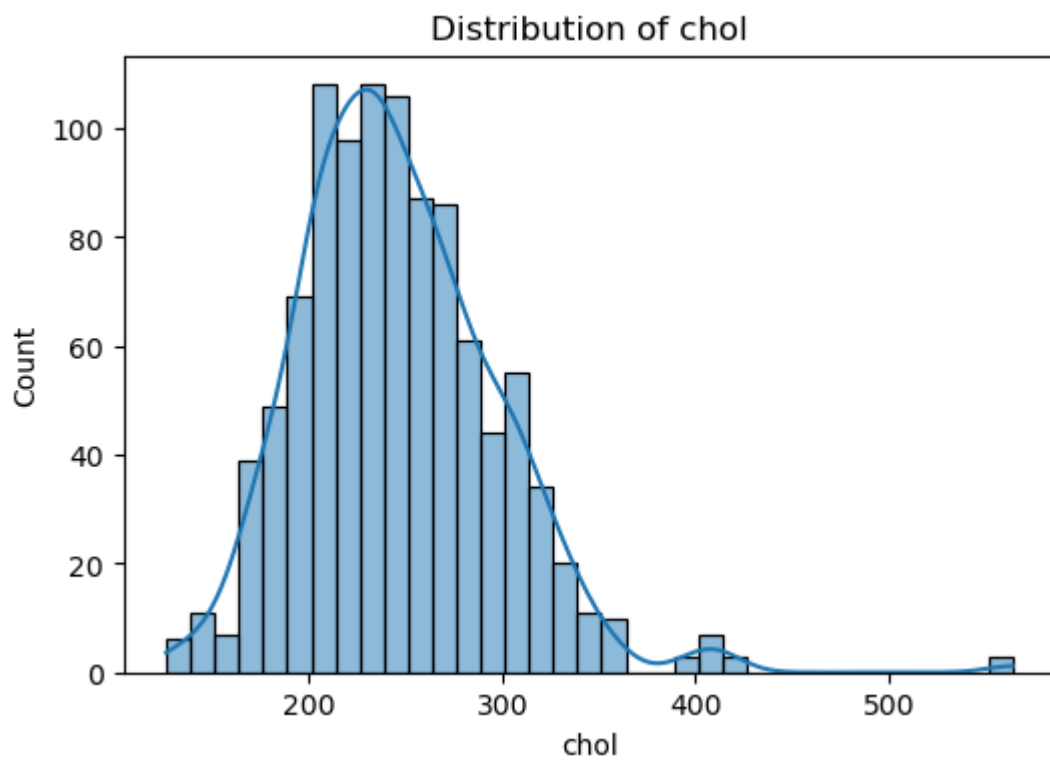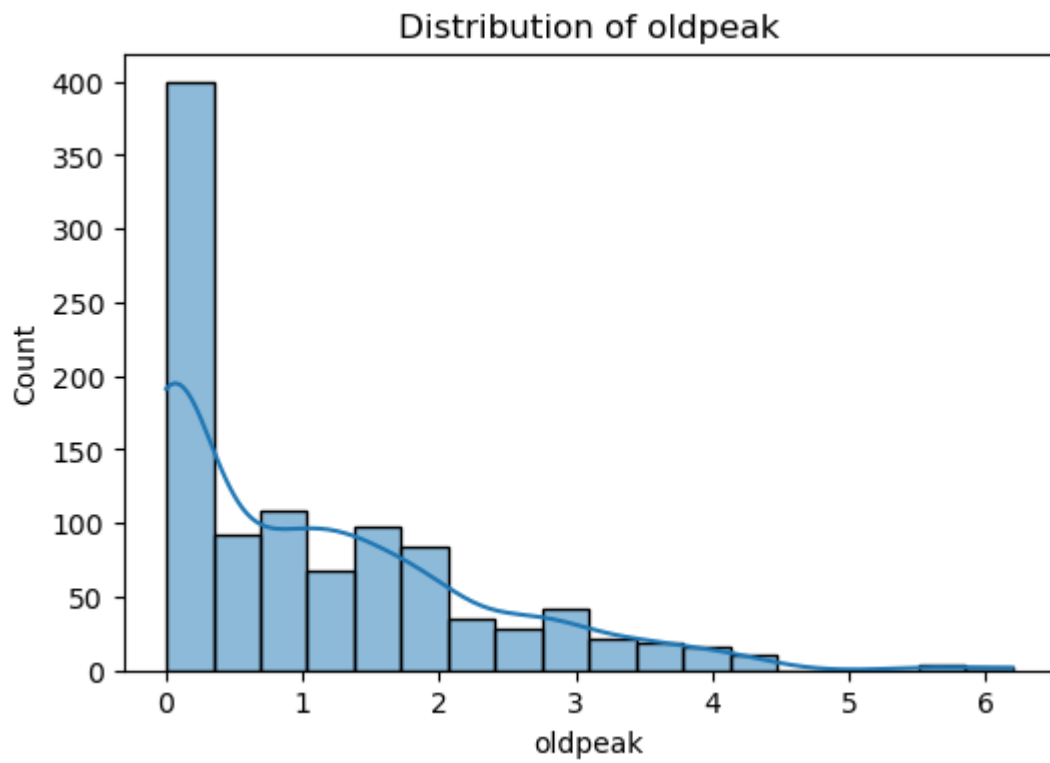
Pair Plot and Categorical Attribute Analysis

In [10]:
```python
# Pair plot for selected features
selected_features = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak', 'target']
sns.pairplot(data[selected_features], hue='target', diag_kind='kde')
plt.suptitle('Pair Plot of Selected Features')
plt.show()
```

Pair Plot of Selected Features
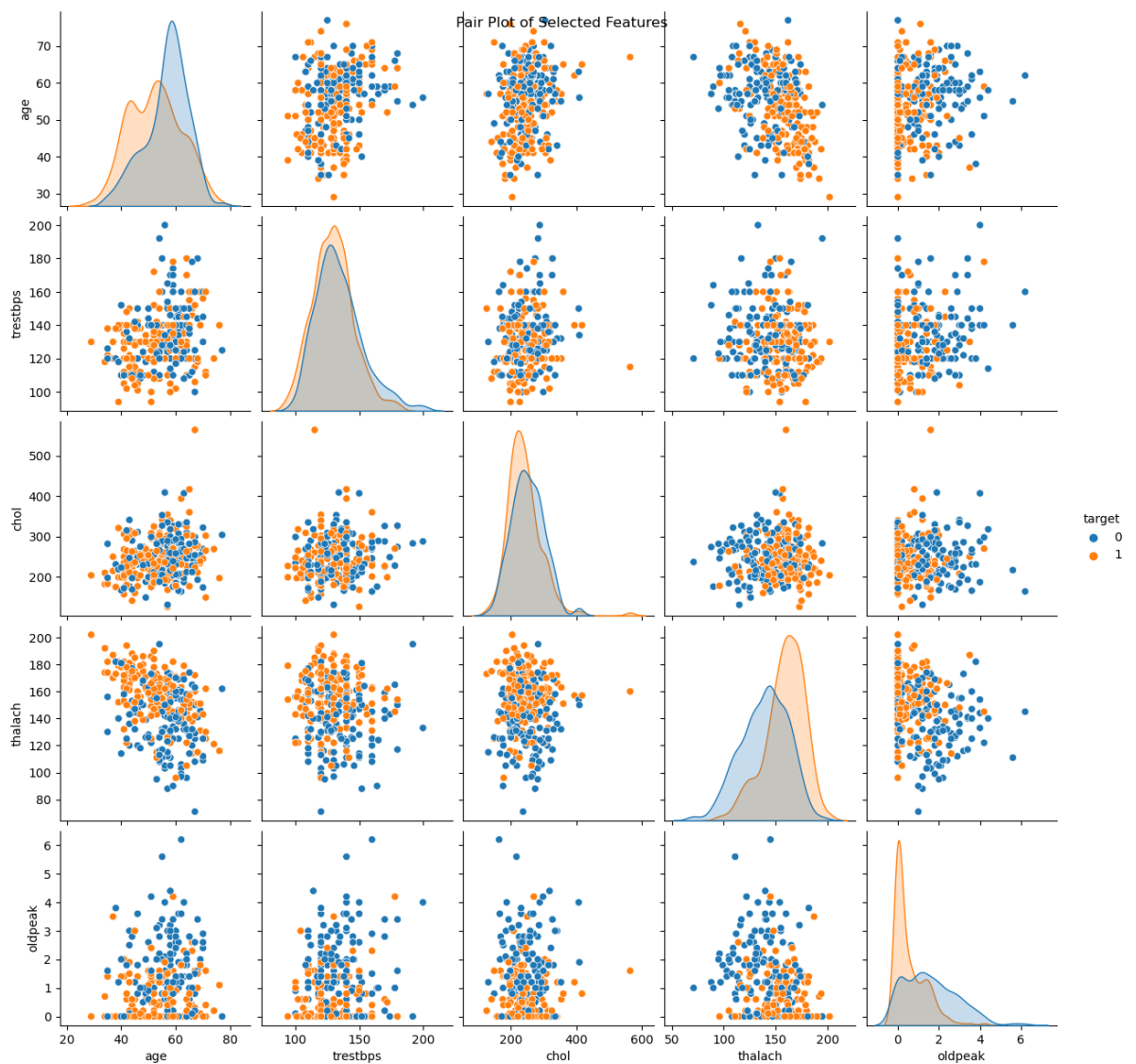
```
In [11]:  # Categorical attribute analysis
          categorical_attributes = ['cp', 'fbs', 'restecg', 'exang', 'slope', 'ca', 'thal']
          for feature in categorical_attributes:
              plt.figure(figsize=(8, 6))
              sns.countplot(x=feature, hue='target', data=data)
              plt.title(f'{feature} vs. Heart Disease')
              plt.show()
```
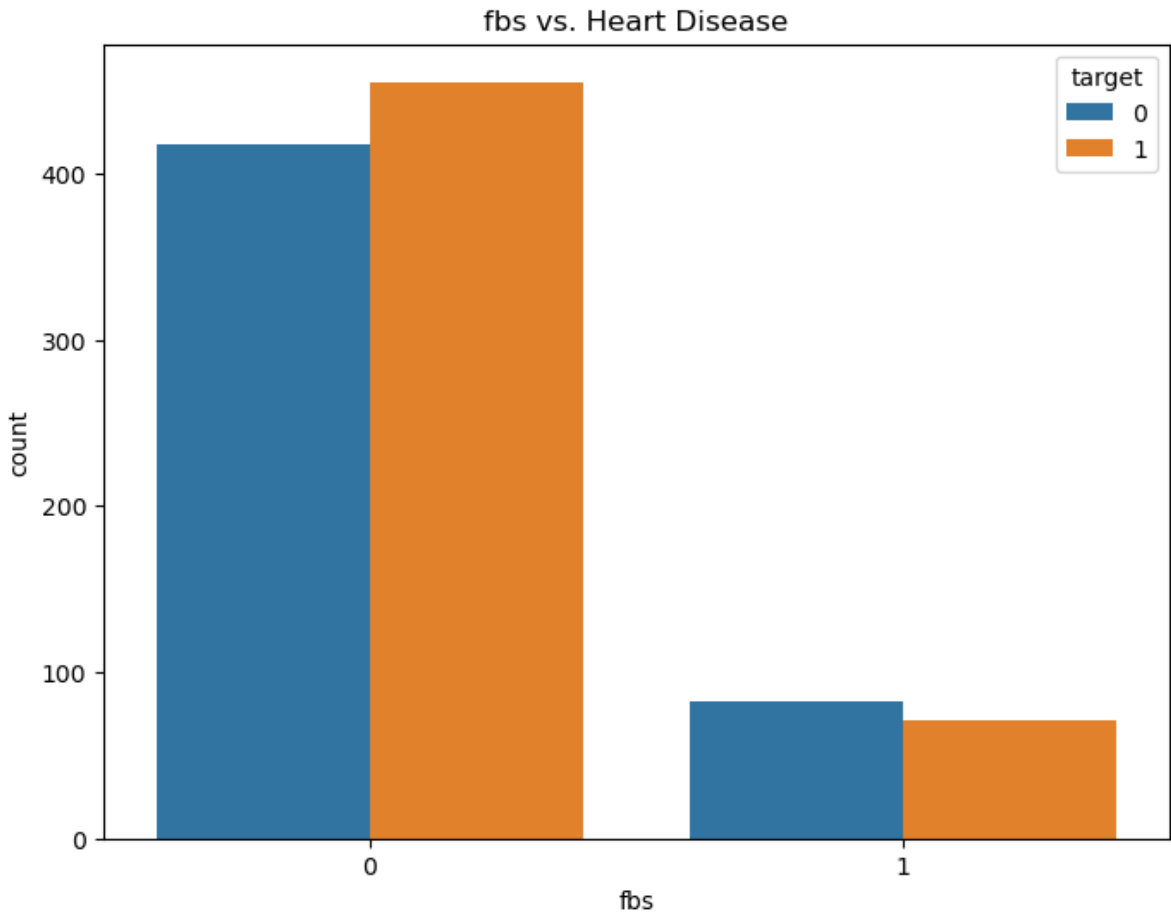
cp vs. Heart Disease

fbs vs. Heart Disease

restecg vs. Heart Disease

exang vs. Heart Disease

slope vs. Heart Disease

ca vs. Heart Disease

thal vs. Heart Disease
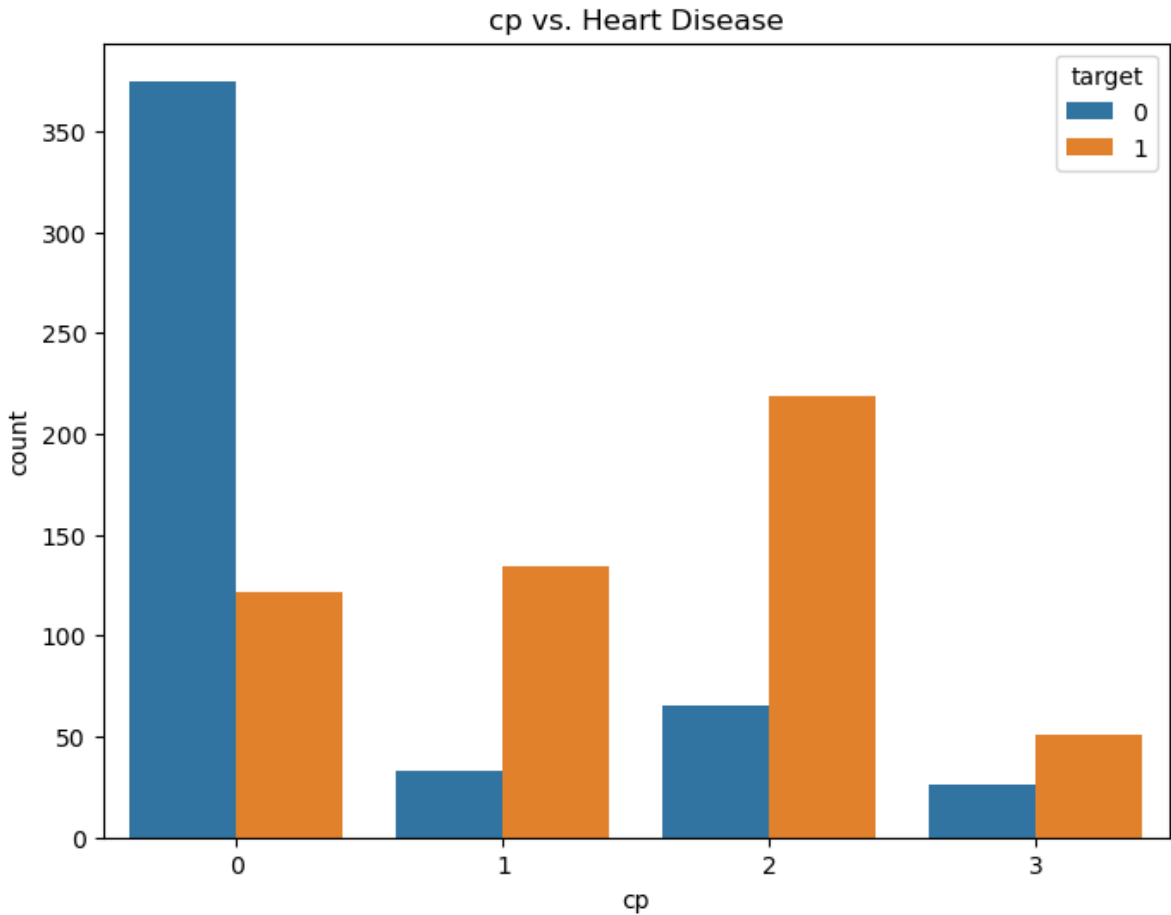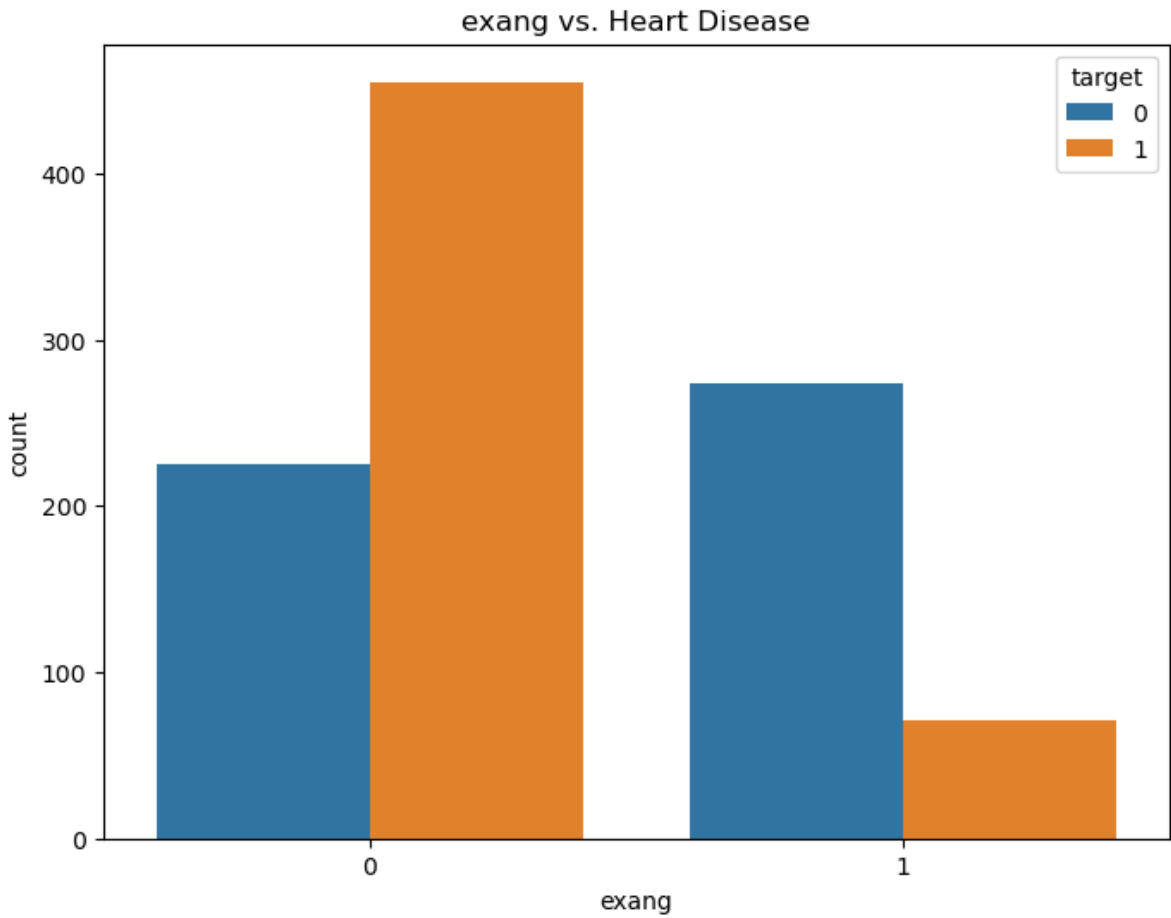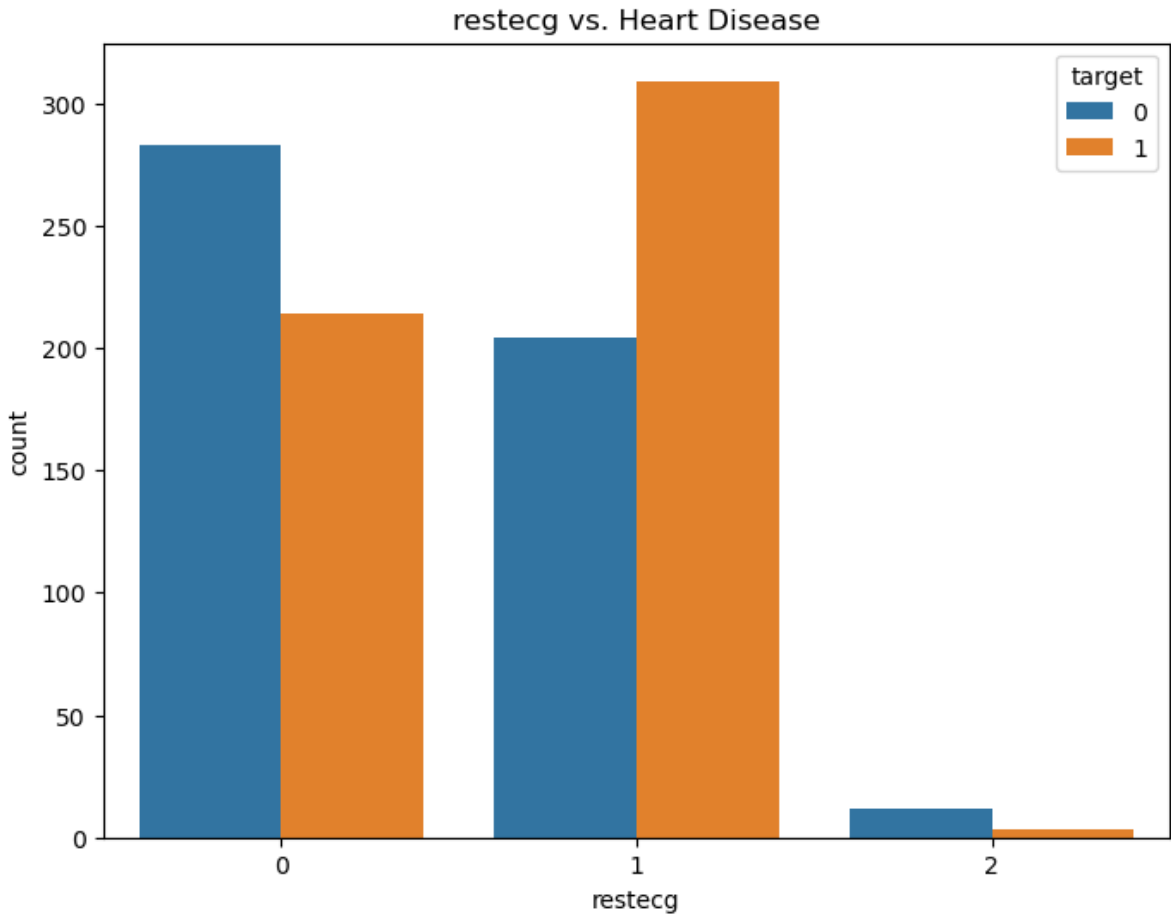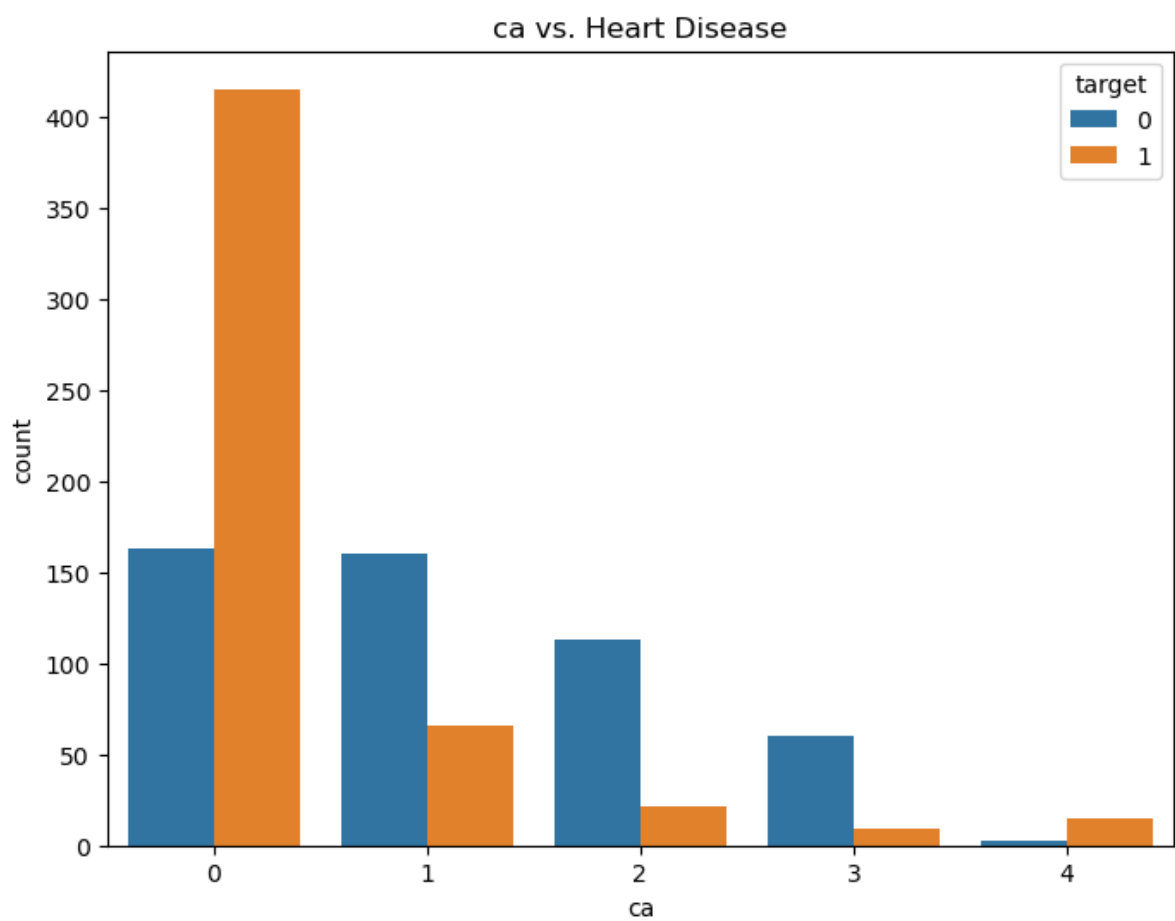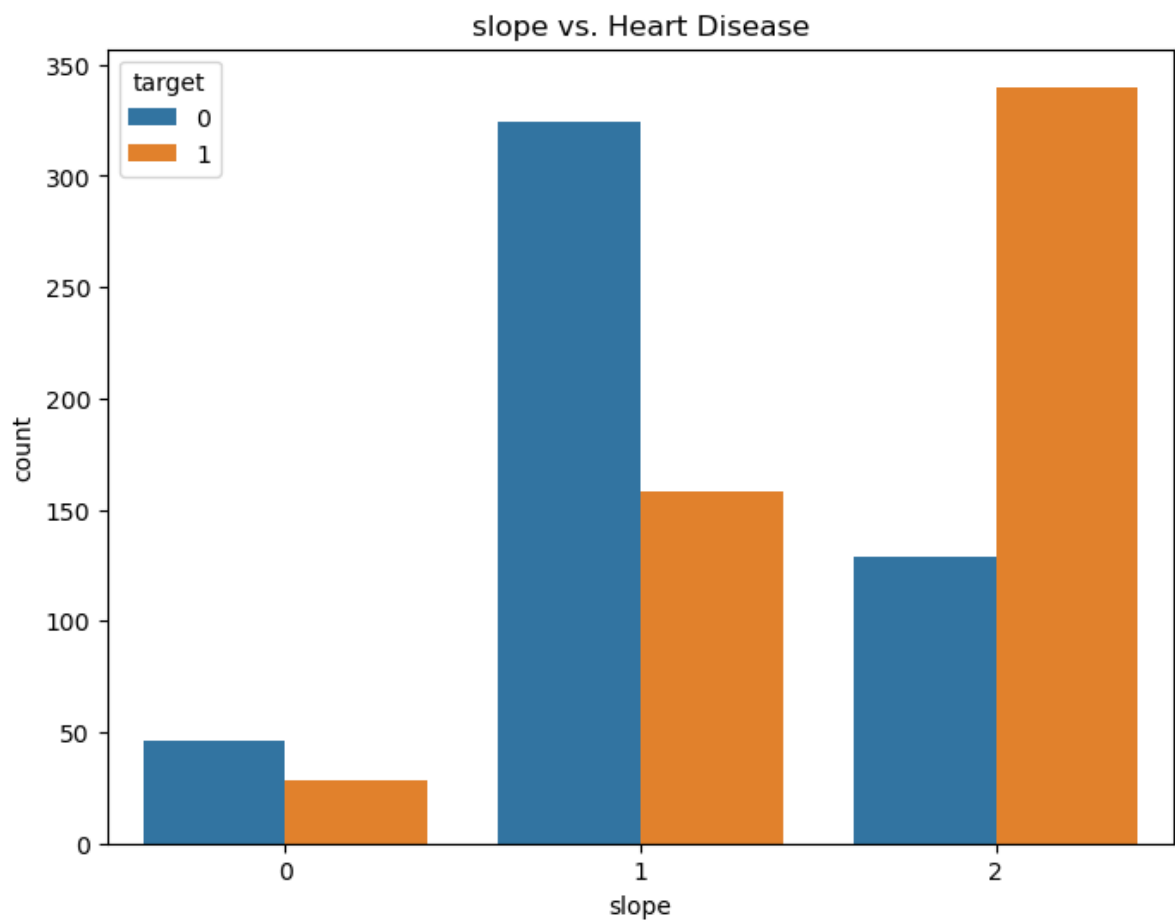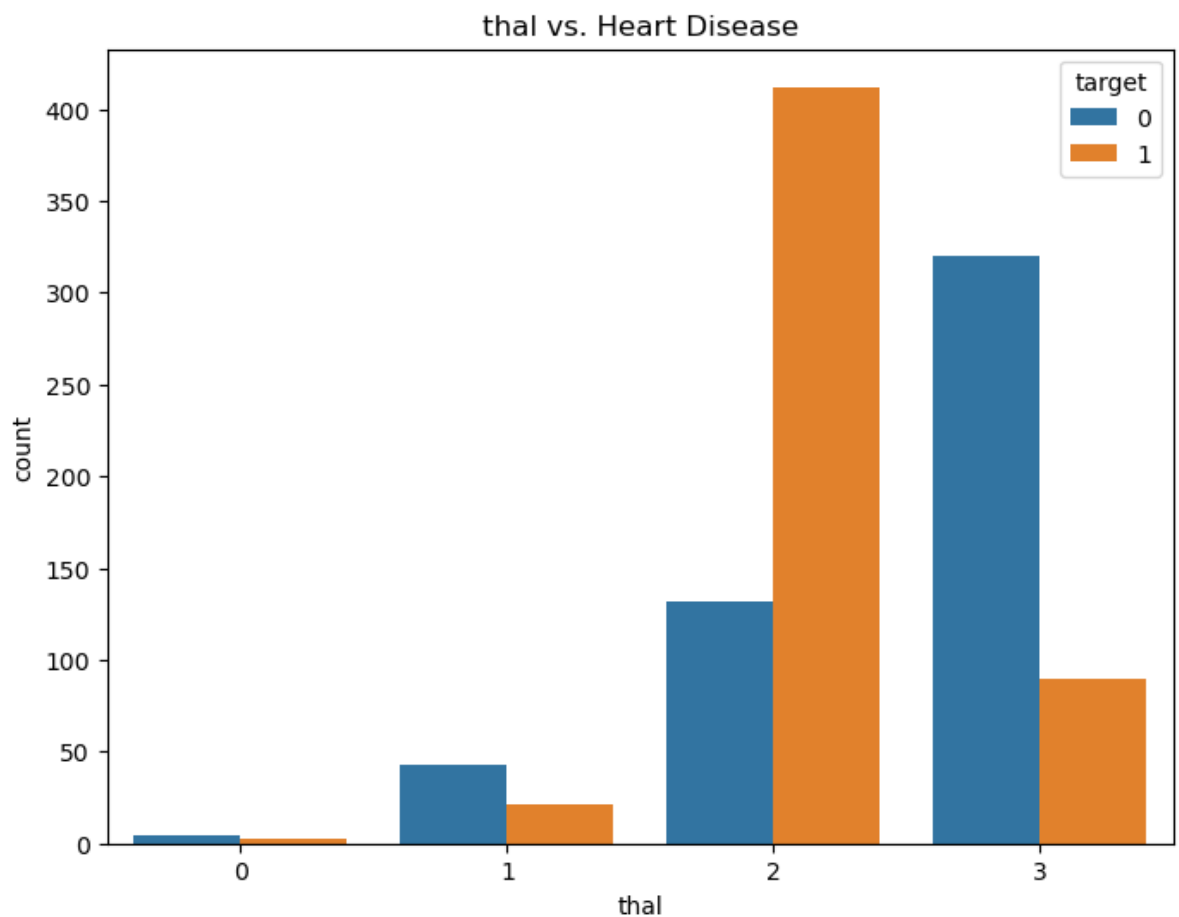
Function for Analysis Summary

```
In [12]:  def heart_disease_analysis_summary(data):
              """
              Summarizes key insights from the analysis of the heart disease dataset.

              Parameters:
              - data: DataFrame, the heart disease dataset.

              Returns:
              - summary: str, a textual summary of key findings.
              """
              heart_disease_rates = data['target'].value_counts(normalize=True)
              heart_disease_by_gender = data.groupby('sex')['target'].value_counts(normalize=
              heart_disease_by_age = data.groupby('target')['age'].describe()
              correlation_matrix = data.corr()

              summary = f"""
              Analysis Summary:
              ----------------

              1. Heart Disease Rates:
              {heart_disease_rates}

              2. Heart Disease by Gender:
              {heart_disease_by_gender}

              3. Heart Disease by Age:
              {heart_disease_by_age}

              4. Correlation Heatmap:
              {correlation_matrix}
              """
```

```python
    return summary
analysis_summary = heart_disease_analysis_summary(data)
print(analysis_summary)
```

```
        Analysis Summary:
        -----------------

        1. Heart Disease Rates:
        1     0.513171
0     0.486829
Name: target, dtype: float64

        2. Heart Disease by Gender:
        sex   target
0     1           0.724359
        0           0.275641
1     0           0.579243
        1           0.420757
Name: target, dtype: float64

        3. Heart Disease by Age:
                count        mean        std    min    25%    50%    75%    max
target
0         499.0  56.569138  7.908153  35.0  52.0  58.0  62.0  77.0
1         526.0  52.408745  9.631804  29.0  44.0  52.0  59.0  76.0

        4. Correlation Heatmap:
                        age        sex         cp   trestbps       chol        fbs  \
age           1.000000  -0.103240  -0.071966   0.271121   0.219823   0.121243
sex          -0.103240   1.000000  -0.041119  -0.078974  -0.198258   0.027200
cp           -0.071966  -0.041119   1.000000   0.038177  -0.081641   0.079294
trestbps  0.271121  -0.078974   0.038177   1.000000   0.127977   0.181767
chol          0.219823  -0.198258  -0.081641   0.127977   1.000000   0.026917
fbs           0.121243   0.027200   0.079294   0.181767   0.026917   1.000000
restecg  -0.132696  -0.055117   0.043581  -0.123794  -0.147410  -0.104051
thalach  -0.390227  -0.049365   0.306839  -0.039264  -0.021772  -0.008866
exang      0.088163   0.139157  -0.401513   0.061197   0.067382   0.049261
oldpeak   0.208137   0.084687  -0.174733   0.187434   0.064880   0.010859
slope     -0.169105  -0.026666   0.131633  -0.120445  -0.014248  -0.061902
ca            0.271551   0.111729  -0.176206   0.104554   0.074259   0.137156
thal          0.072297   0.198424  -0.163341   0.059276   0.100244  -0.042177
target   -0.229324  -0.279501   0.434854  -0.138772  -0.099966  -0.041164

                    restecg    thalach      exang     oldpeak      slope         ca  \
age          -0.132696  -0.390227   0.088163   0.208137  -0.169105   0.271551
sex          -0.055117  -0.049365   0.139157   0.084687  -0.026666   0.111729
cp            0.043581   0.306839  -0.401513  -0.174733   0.131633  -0.176206
trestbps  -0.123794  -0.039264   0.061197   0.187434  -0.120445   0.104554
chol         -0.147410  -0.021772   0.067382   0.064880  -0.014248   0.074259
fbs          -0.104051  -0.008866   0.049261   0.010859  -0.061902   0.137156
restecg   1.000000   0.048411  -0.065606  -0.050114   0.086086  -0.078072
thalach   0.048411   1.000000  -0.380281  -0.349796   0.395308  -0.207888
exang     -0.065606  -0.380281   1.000000   0.310844  -0.267335   0.107849
oldpeak   -0.050114  -0.349796   0.310844   1.000000  -0.575189   0.221816
slope      0.086086   0.395308  -0.267335  -0.575189   1.000000  -0.073440
ca           -0.078072  -0.207888   0.107849   0.221816  -0.073440   1.000000
thal         -0.020504  -0.098068   0.197201   0.202672  -0.094090   0.149014
target    0.134468   0.422895  -0.438029  -0.438441   0.345512  -0.382085

                    thal     target
age          0.072297  -0.229324
sex          0.198424  -0.279501
cp          -0.163341   0.434854
trestbps  0.059276  -0.138772
chol         0.100244  -0.099966
fbs         -0.042177  -0.041164
restecg  -0.020504   0.134468
thalach  -0.098068   0.422895
```

```
exang      0.197201 -0.438029
oldpeak    0.202672 -0.438441
slope     -0.094090  0.345512
ca         0.149014 -0.382085
thal       1.000000 -0.337838
target    -0.337838  1.000000
```

In [13]: `pip install jupyter-dash`

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: jupyter-dash in c:\users\hp\appdata\roaming\python
\python39\site-packages (0.4.2)
Requirement already satisfied: flask in c:\programdata\anaconda3\lib\site-packages
(from jupyter-dash) (1.1.2)
Requirement already satisfied: retrying in c:\users\hp\appdata\roaming\python\pyth
on39\site-packages (from jupyter-dash) (1.3.4)
Requirement already satisfied: ansi2html in c:\users\hp\appdata\roaming\python\pyt
hon39\site-packages (from jupyter-dash) (1.9.1)
Requirement already satisfied: nest-asyncio in c:\programdata\anaconda3\lib\site-p
ackages (from jupyter-dash) (1.5.5)
Requirement already satisfied: ipykernel in c:\programdata\anaconda3\lib\site-pack
ages (from jupyter-dash) (6.15.2)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packa
ges (from jupyter-dash) (2.28.1)
Requirement already satisfied: dash in c:\users\hp\appdata\roaming\python\python39
\site-packages (from jupyter-dash) (2.14.2)
Requirement already satisfied: ipython in c:\programdata\anaconda3\lib\site-packag
es (from jupyter-dash) (7.31.1)
Requirement already satisfied: setuptools in c:\programdata\anaconda3\lib\site-pac
kages (from dash->jupyter-dash) (63.4.1)
Requirement already satisfied: typing-extensions>=4.1.1 in c:\users\hp\appdata\roa
ming\python\python39\site-packages (from dash->jupyter-dash) (4.8.0)
Requirement already satisfied: dash-table==5.0.0 in c:\users\hp\appdata\roaming\py
thon\python39\site-packages (from dash->jupyter-dash) (5.0.0)
Requirement already satisfied: Werkzeug<3.1 in c:\programdata\anaconda3\lib\site-p
ackages (from dash->jupyter-dash) (2.0.3)
Requirement already satisfied: plotly>=5.0.0 in c:\programdata\anaconda3\lib\site-
packages (from dash->jupyter-dash) (5.9.0)
Requirement already satisfied: dash-core-components==2.0.0 in c:\users\hp\appdata
\roaming\python\python39\site-packages (from dash->jupyter-dash) (2.0.0)
Requirement already satisfied: importlib-metadata in c:\programdata\anaconda3\lib
\site-packages (from dash->jupyter-dash) (4.11.3)
Requirement already satisfied: dash-html-components==2.0.0 in c:\users\hp\appdata
\roaming\python\python39\site-packages (from dash->jupyter-dash) (2.0.0)
Requirement already satisfied: click>=5.1 in c:\programdata\anaconda3\lib\site-pac
kages (from flask->jupyter-dash) (8.0.4)
Requirement already satisfied: itsdangerous>=0.24 in c:\programdata\anaconda3\lib
\site-packages (from flask->jupyter-dash) (2.0.1)
Requirement already satisfied: Jinja2>=2.10.1 in c:\programdata\anaconda3\lib\site
-packages (from flask->jupyter-dash) (2.11.3)
Requirement already satisfied: packaging in c:\programdata\anaconda3\lib\site-pack
ages (from ipykernel->jupyter-dash) (21.3)
Requirement already satisfied: debugpy>=1.0 in c:\programdata\anaconda3\lib\site-p
ackages (from ipykernel->jupyter-dash) (1.5.1)
Requirement already satisfied: jupyter-client>=6.1.12 in c:\programdata\anaconda3
\lib\site-packages (from ipykernel->jupyter-dash) (7.3.4)
Requirement already satisfied: traitlets>=5.1.0 in c:\programdata\anaconda3\lib\si
te-packages (from ipykernel->jupyter-dash) (5.1.1)
Requirement already satisfied: pyzmq>=17 in c:\programdata\anaconda3\lib\site-pack
ages (from ipykernel->jupyter-dash) (23.2.0)
Requirement already satisfied: matplotlib-inline>=0.1 in c:\programdata\anaconda3
\lib\site-packages (from ipykernel->jupyter-dash) (0.1.6)
Requirement already satisfied: tornado>=6.1 in c:\programdata\anaconda3\lib\site-p
ackages (from ipykernel->jupyter-dash) (6.1)
Requirement already satisfied: psutil in c:\programdata\anaconda3\lib\site-package
s (from ipykernel->jupyter-dash) (5.9.0)
Requirement already satisfied: backcall in c:\programdata\anaconda3\lib\site-packa
ges (from ipython->jupyter-dash) (0.2.0)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in
c:\programdata\anaconda3\lib\site-packages (from ipython->jupyter-dash) (3.0.20)
Requirement already satisfied: pygments in c:\users\hp\appdata\roaming\python\pyth
on39\site-packages (from ipython->jupyter-dash) (2.17.2)
Requirement already satisfied: colorama in c:\users\hp\appdata\roaming\python\pyth

```
on39\site-packages (from ipython->jupyter-dash) (0.4.6)
Requirement already satisfied: decorator in c:\programdata\anaconda3\lib\site-pack
ages (from ipython->jupyter-dash) (5.1.1)
Requirement already satisfied: jedi>=0.16 in c:\programdata\anaconda3\lib\site-pac
kages (from ipython->jupyter-dash) (0.18.1)
Requirement already satisfied: pickleshare in c:\programdata\anaconda3\lib\site-pa
ckages (from ipython->jupyter-dash) (0.7.5)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\programdata\anaconda3\l
ib\site-packages (from requests->jupyter-dash) (1.26.11)
Requirement already satisfied: idna<4,>=2.5 in c:\programdata\anaconda3\lib\site-p
ackages (from requests->jupyter-dash) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib
\site-packages (from requests->jupyter-dash) (2022.9.14)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\programdata\anaconda
3\lib\site-packages (from requests->jupyter-dash) (2.0.4)
Requirement already satisfied: six>=1.7.0 in c:\programdata\anaconda3\lib\site-pac
kages (from retrying->jupyter-dash) (1.16.0)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\programdata\anaconda3\lib
\site-packages (from jedi>=0.16->ipython->jupyter-dash) (0.8.3)
Requirement already satisfied: MarkupSafe>=0.23 in c:\programdata\anaconda3\lib\si
te-packages (from Jinja2>=2.10.1->flask->jupyter-dash) (2.0.1)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\programdata\anaconda3
\lib\site-packages (from jupyter-client>=6.1.12->ipykernel->jupyter-dash) (2.8.2)
Requirement already satisfied: entrypoints in c:\programdata\anaconda3\lib\site-pa
ckages (from jupyter-client>=6.1.12->ipykernel->jupyter-dash) (0.4)
Requirement already satisfied: jupyter-core>=4.9.2 in c:\programdata\anaconda3\lib
\site-packages (from jupyter-client>=6.1.12->ipykernel->jupyter-dash) (4.11.1)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\hp\appdata\roaming\pyth
on\python39\site-packages (from plotly>=5.0.0->dash->jupyter-dash) (8.2.3)
Requirement already satisfied: wcwidth in c:\programdata\anaconda3\lib\site-packag
es (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython->jupyter-dash) (0.
2.5)
Requirement already satisfied: zipp>=0.5 in c:\programdata\anaconda3\lib\site-pack
ages (from importlib-metadata->dash->jupyter-dash) (3.8.0)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\programdata\anaconda
3\lib\site-packages (from packaging->ipykernel->jupyter-dash) (3.0.9)
Requirement already satisfied: pywin32>=1.0 in c:\programdata\anaconda3\lib\site-p
ackages (from jupyter-core>=4.9.2->jupyter-client>=6.1.12->ipykernel->jupyter-das
h) (302)
Note: you may need to restart the kernel to use updated packages.
```

Dash App for Interactive Visualization

```python
In [14]:   import dash
           from dash import dcc, html
           from dash.dependencies import Input, Output
           import pandas as pd
           import plotly.express as px
           import seaborn as sns
           import plotly.graph_objects as go

           # Initialize the Dash app
           app = dash.Dash(__name__)

           app.layout = html.Div([
               html.H1("Heart Disease Analysis Dashboard"),

               dcc.Markdown(children=heart_disease_analysis_summary(data)),

               dcc.Graph(
                   id='heart-disease-rates',
                   figure=px.bar(data, x=data['target'].value_counts().index, y=data['target']
                           labels={'x': 'Heart Disease', 'y': 'Count'},
```

```python
                            title='Heart Disease Rates (0: No Disease, 1: Disease)')
        ),

        dcc.Graph(
            id='heart-disease-gender',
            figure=px.bar(data, x='sex', color='target',
                          labels={'sex': 'Gender', 'target': 'Heart Disease'},
                          title='Heart Disease by Gender (0: No Disease, 1: Disease)')
        ),

        dcc.Graph(
            id='heart-disease-age',
            figure=px.histogram(data, x='age', color='target', nbins=20,
                                labels={'age': 'Age', 'target': 'Heart Disease'},
                                title='Heart Disease by Age')
        ),

        dcc.Graph(
            id='correlation-heatmap',
            figure=px.imshow(data.corr(),
                             labels=dict(x='Features', y='Features', color='Correlation
                             title='Correlation Heatmap')
        ),

        dcc.Graph(
            id='distribution-plots',
            figure={}
        ),

        dcc.Graph(
            id='pair-plot',
            figure={}
        ),

        dcc.Graph(
            id='categorical-attributes',
            figure={}
        )
    ])

@app.callback(
    [Output('distribution-plots', 'figure'),
     Output('pair-plot', 'figure'),
     Output('categorical-attributes', 'figure')],
    [Input('heart-disease-age', 'hoverData')]
)
def update_plots(hover_data):
    distribution_fig = go.Figure()
    numerical_features = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak']
    for feature in numerical_features:
        distribution_fig.add_trace(go.Histogram(x=data[feature], name=feature, nbin

    pair_plot_fig = px.scatter_matrix(data, dimensions=numerical_features, color='t

    categorical_fig = go.Figure()
    categorical_attributes = ['cp', 'fbs', 'restecg', 'exang', 'slope', 'ca', 'thal
    for feature in categorical_attributes:
        grouped_data = data.groupby([feature, 'target']).size().unstack()
        categorical_fig.add_trace(go.Bar(x=grouped_data.index, y=grouped_data[1], n
        categorical_fig.add_trace(go.Bar(x=grouped_data.index, y=grouped_data[0], n

    return distribution_fig, pair_plot_fig, categorical_fig
app.run_server(mode="inline")
```