

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

```
In [2]: data = pd.read_csv('/content/HeartDisease.csv')
```

## Data Cleaning

```
In [3]: data=data.drop_duplicates()
```

```
In [4]: data
```

Out[4]:

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

303 rows × 14 columns



## Removing Duplicates

```
In [5]: data.isna().sum()
```

```
Out[5]:
```

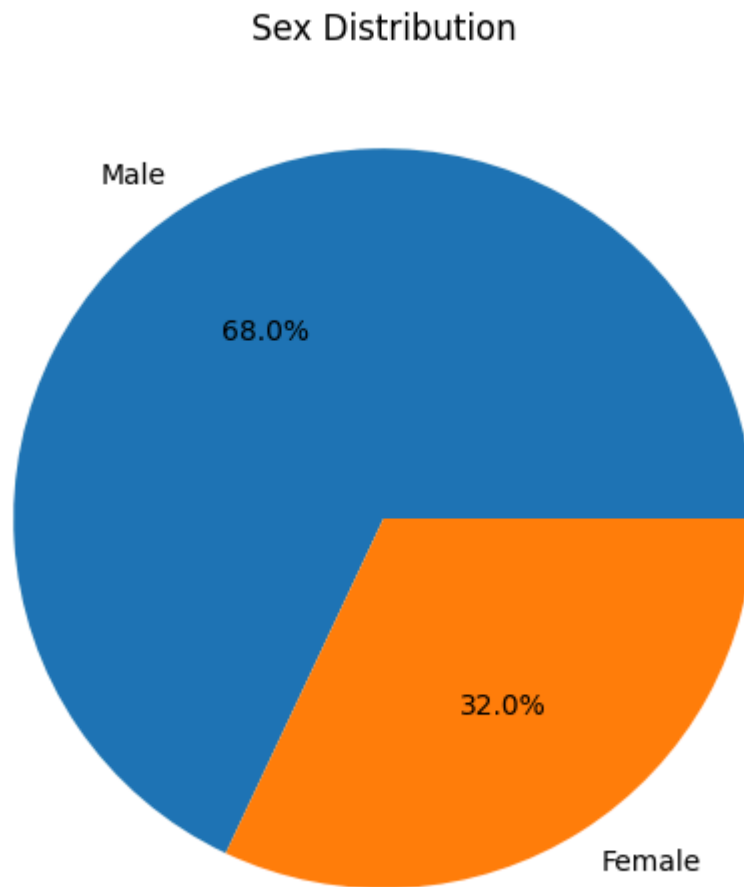
	0
age	0
sex	0
chest pain	0
trestbps	0
chol	0
fps	0
restecg	0
thalach	0
exang	0
oldpeak	0
slope	0
ca	0
thal	0
num	0

**dtype:** int64

## Visualization Plot

### 1. Pie Graph

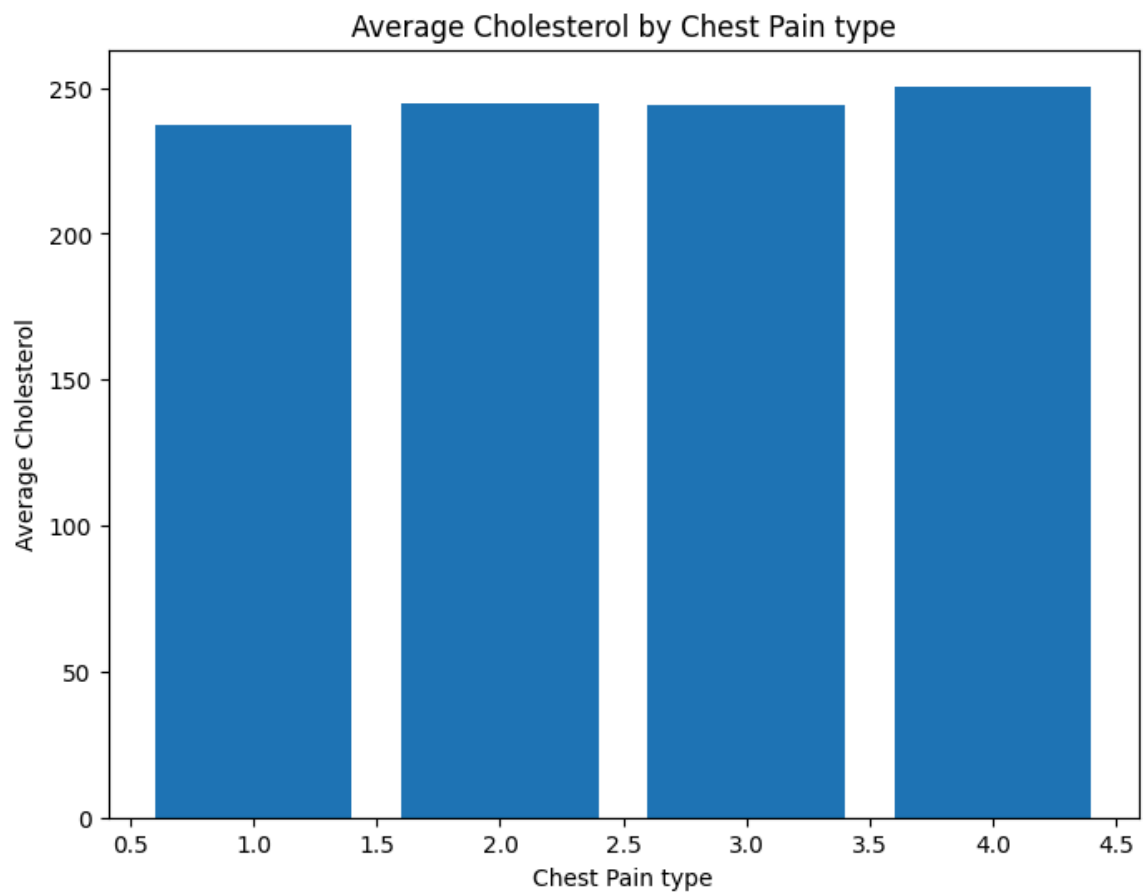
```
In [6]: sex_counts = data['sex'].value_counts()
plt.figure(figsize=(6,6))
plt.pie(sex_counts, labels=['Male', 'Female'], autopct='%1.1f%%')
plt.title("Sex Distribution")
plt.show()
```



## 1. Bar Plot

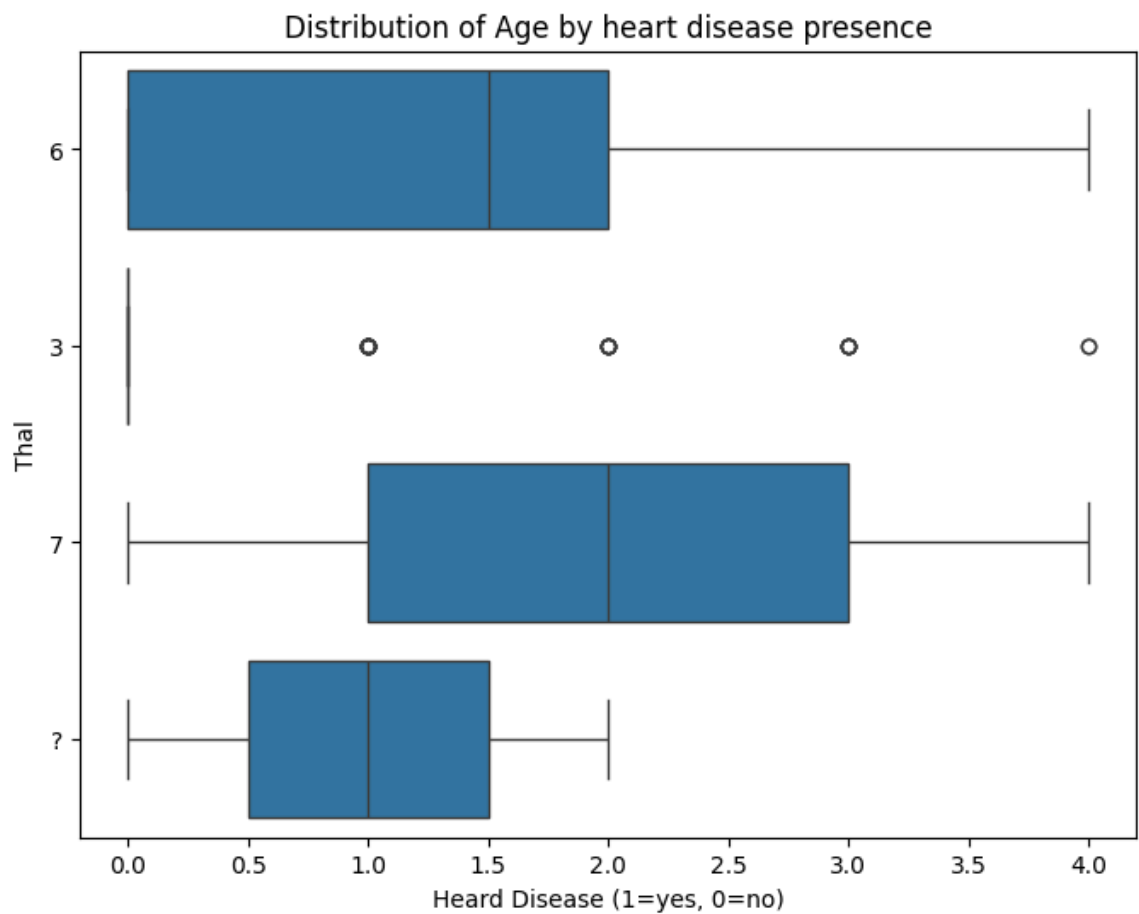
```
In [7]: avg_chol_by_cp = data.groupby('chest pain')['chol'].mean()
```

```
In [8]: plt.figure(figsize=(8,6))
plt.bar(avg_chol_by_cp.index, avg_chol_by_cp.values)
plt.xlabel('Chest Pain type')
plt.ylabel('Average Cholesterol')
plt.title('Average Cholesterol by Chest Pain type')
plt.show()
```



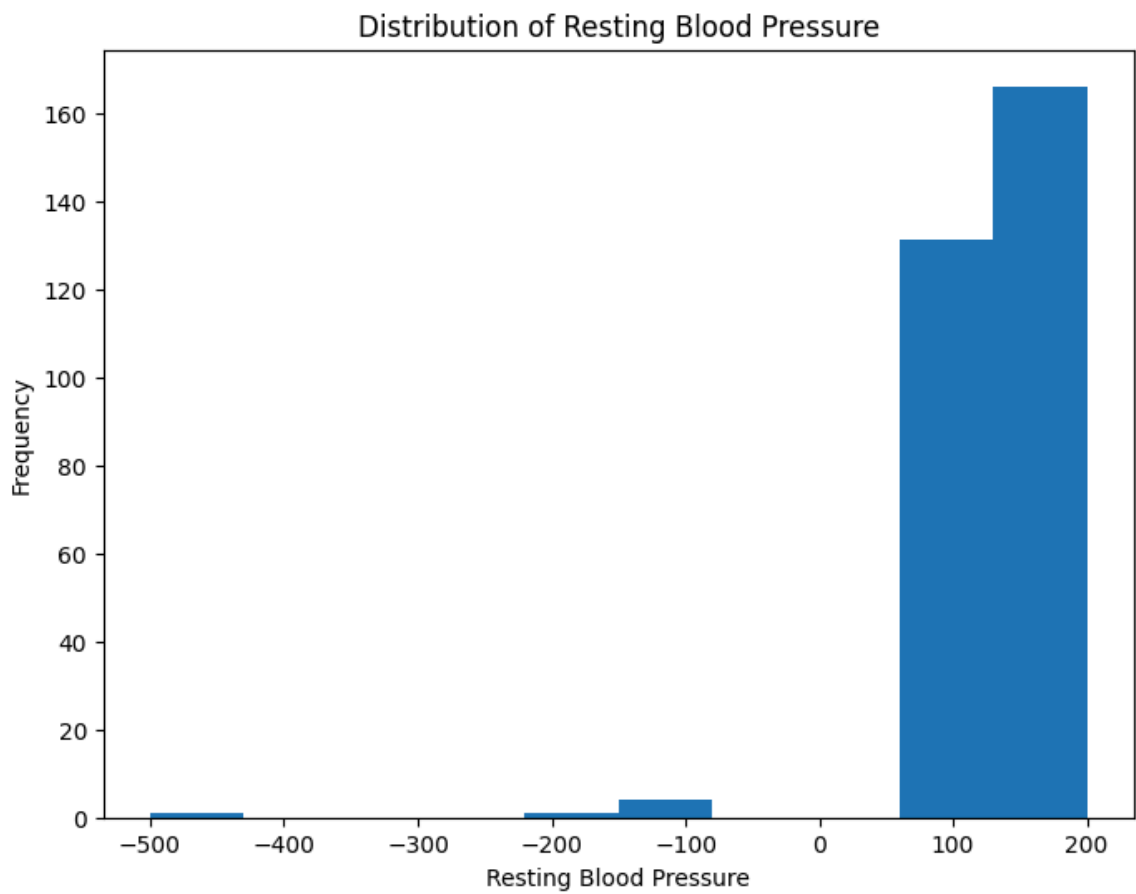
## 1. Box plot

```
In [9]: plt.figure(figsize=(8,6))
sns.boxplot(x='num', y='thal', data=data)
plt.xlabel('Heart Disease (1=yes, 0=no)')
plt.ylabel('Thal')
plt.title('Distribution of Age by heart disease presence')
plt.show()
```



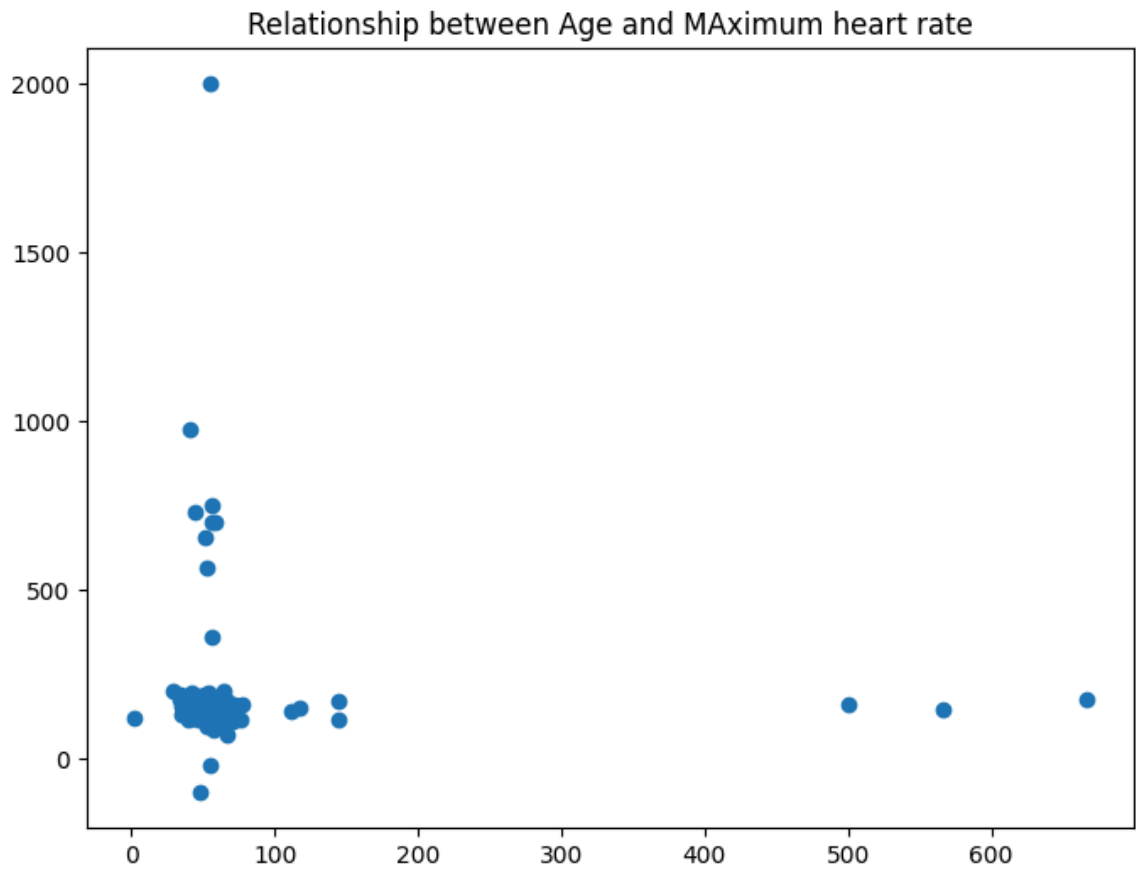
## 1. Histogram

```
In [10]: plt.figure(figsize=(8,6))
plt.hist(data['trestbps'], bins=10)
plt.xlabel('Resting Blood Pressure')
plt.ylabel('Frequency')
plt.title('Distribution of Resting Blood Pressure')
plt.show()
```



## 1. Scatter plot

```
In [11]: plt.figure(figsize=(8,6))  
plt.scatter(data['age'], data['thalach'])  
plt.title('Relationship between Age and MAXimum heart rate')  
plt.show()
```

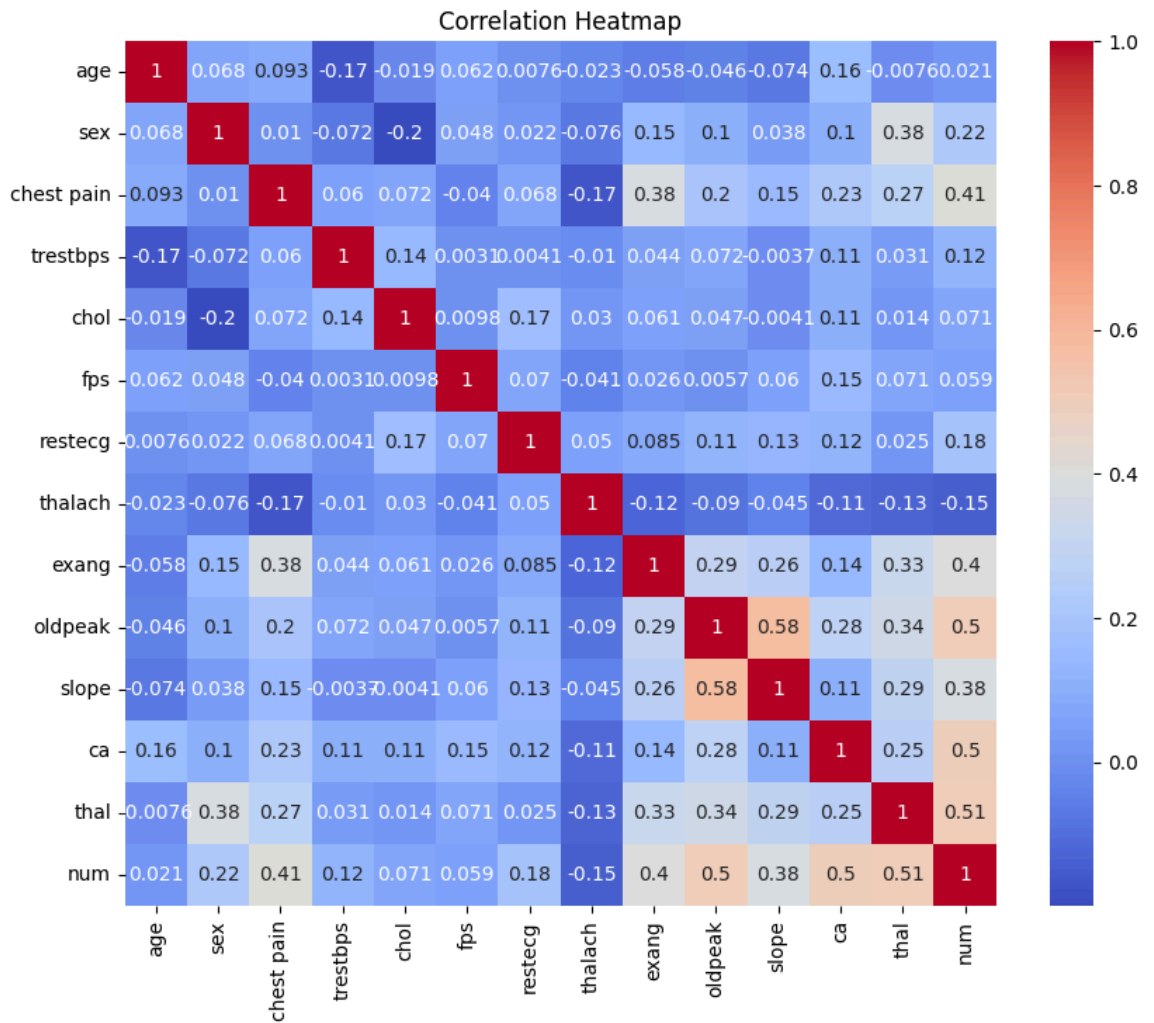


```
In [12]: data.replace('?', np.nan, inplace=True)
```

```
In [13]: data = data.apply(pd.to_numeric, errors='coerce')
```

## 1. Heat Plot

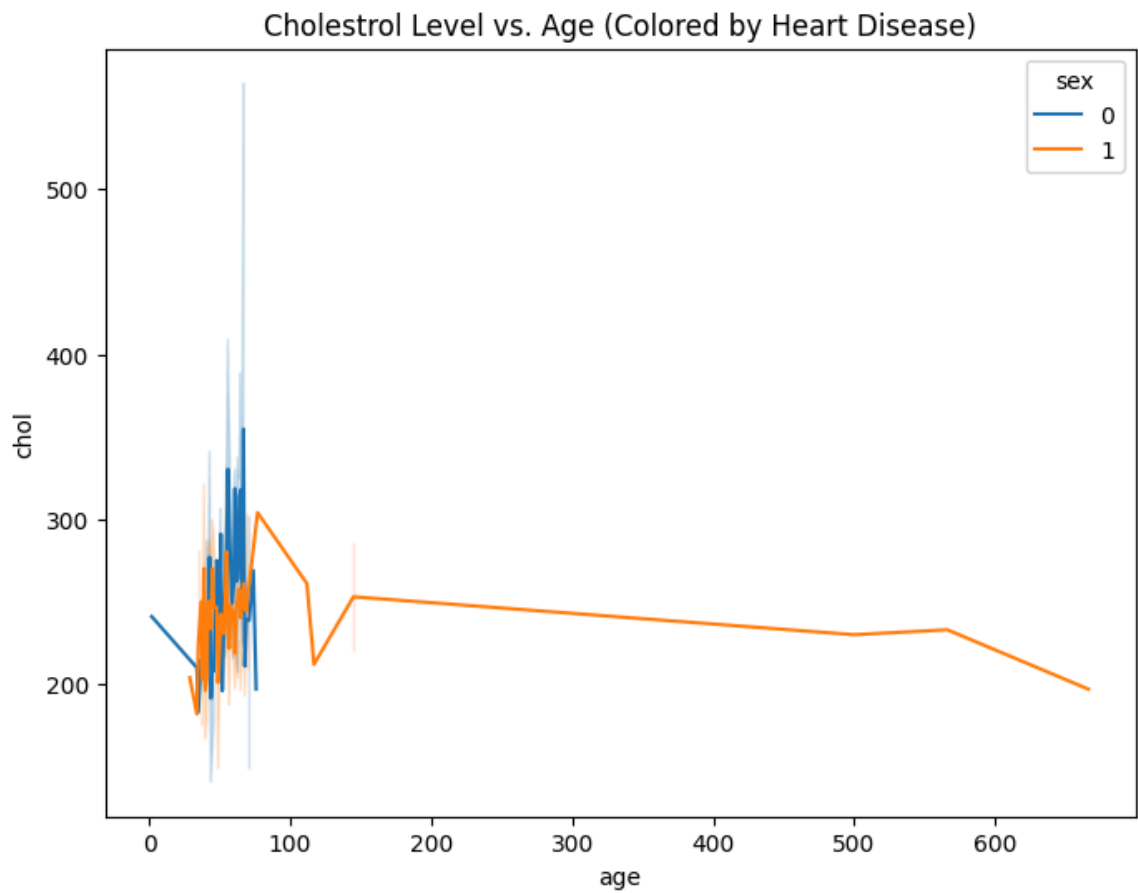
```
In [14]: plt.figure(figsize=(10,8))
sns.heatmap(data.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



## 1. Line plot

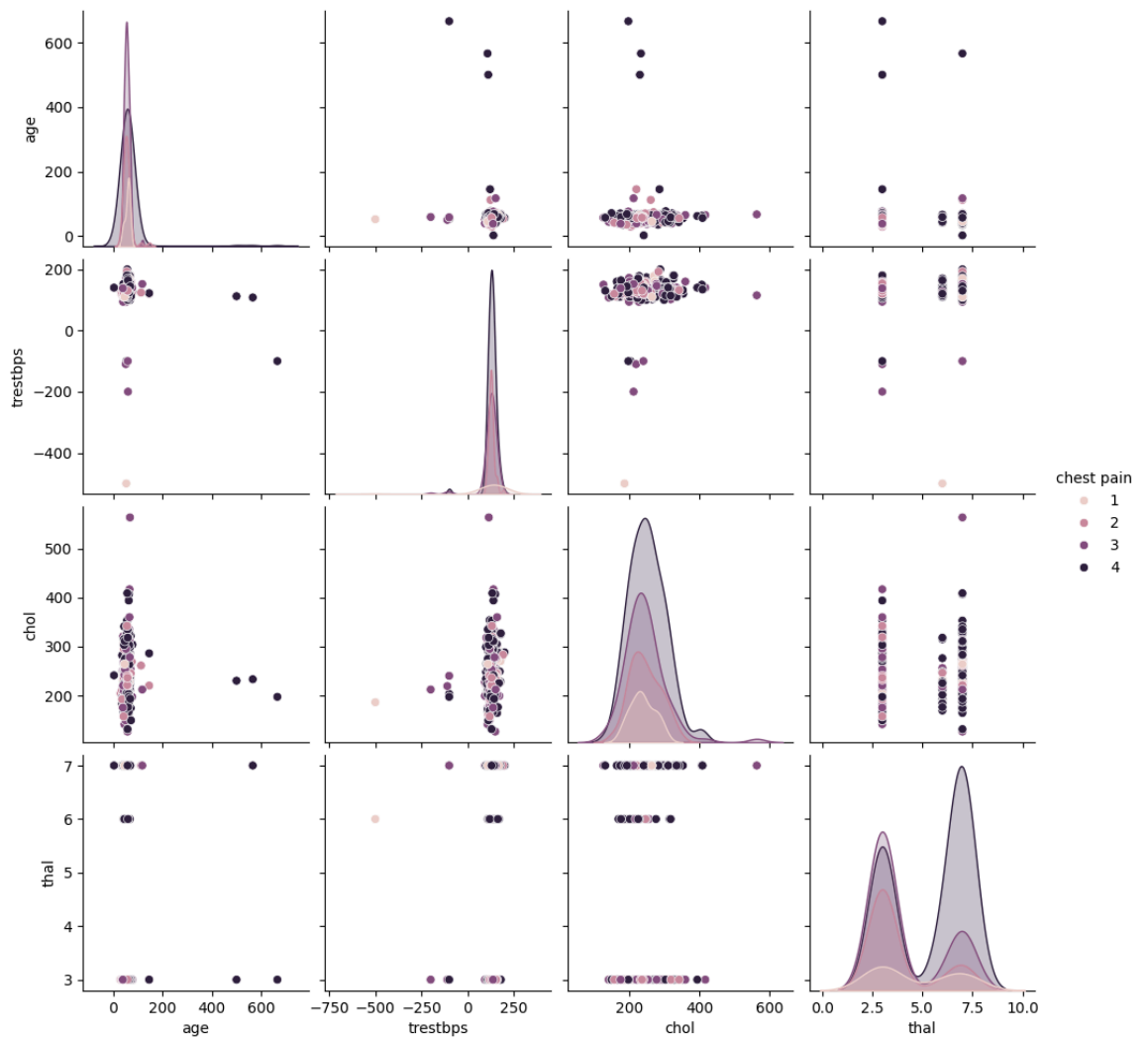


```
In [15]: plt.figure(figsize=(8,6))  
sns.lineplot(data=data, x='age', y='chol', hue='sex')  
plt.title('Cholestrol Level vs. Age (Colored by Heart Disease)')  
plt.show()
```



## 1. Pair Plot

```
In [16]: temp_df = data[['age', 'chest pain', 'trestbps', 'chol', 'thal']]
sns.pairplot(temp_df, hue='chest pain')
plt.show()
```



In [17]: `!pip install wordcloud`

```
Requirement already satisfied: wordcloud in /usr/local/lib/python3.11/dist-packages (1.9.4)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.11/dist-packages (from wordcloud) (2.0.2)
Requirement already satisfied: pillow in /usr/local/lib/python3.11/dist-packages (from wordcloud) (11.1.0)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (from wordcloud) (3.10.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (1.3.2)
Requirement already satisfied: cyclor>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (1.4.8)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (24.2)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil->matplotlib->wordcloud) (1.17.0)
```

In [18]: `from wordcloud import WordCloud`

In [19]: `with open('sample.txt', 'r') as f:  
 text = f.read()`

In [20]: `wordcloud = WordCloud().generate(text)`

In [21]: `import matplotlib.pyplot as plt  
  
plt.imshow(wordcloud, interpolation='bilinear')  
plt.axis("off")  
plt.show()`

