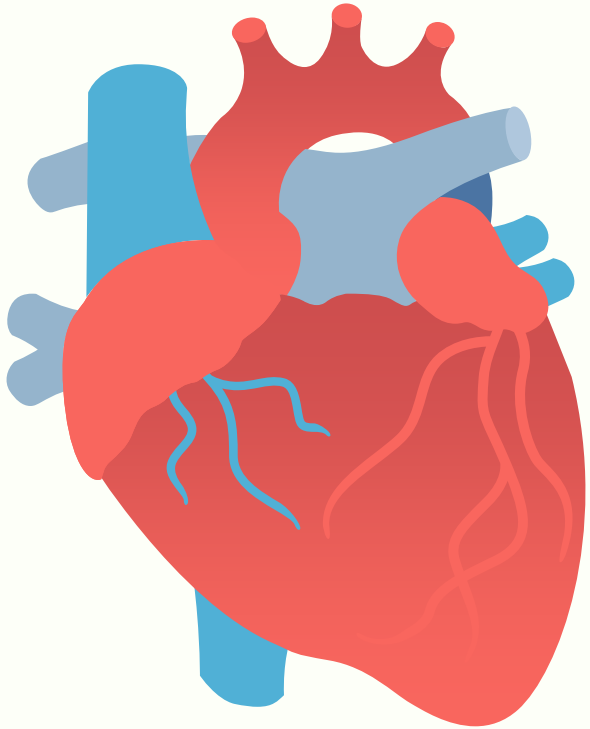
Heart Disease Data Analysis : Report

Unified Mentor

By: Israr mohammed
Email : israr4075@gmail.com
Mobile: +91 7737136423

Looking for Work Opportunities



LinkedIn : https://www.linkedin.com/in/israrmohammed/

01

Introduction

Project Description , Data and Libraries used

Problem Statement

Overview

Health is real wealth, and during the pandemic, we all realized the brute effects of COVID-19 on everyone, irrespective of status; therefore, you are required to analyze health and medical data for better future preparation to understand and mitigate the risks associated with heart disease.

Tools and Libraries

ETL Process:

- Use Python libraries such as pandas for data manipulation.
- Utilize Jupyter Notebook for database interactions.

Exploratory Data Analysis (EDA):

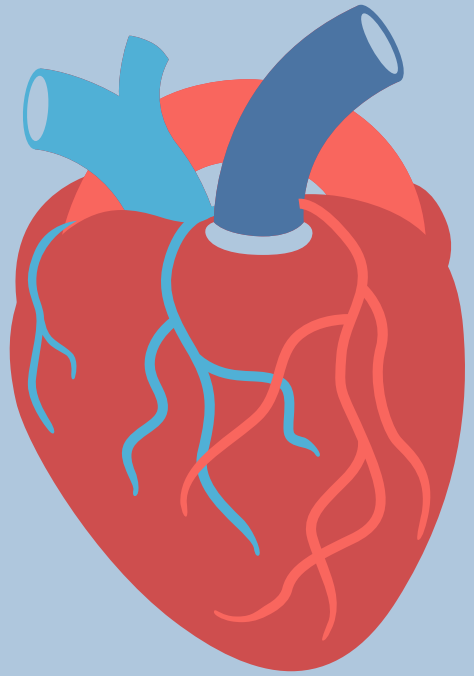
- Perform EDA with **pandas** using Python.
- Visualize data with **matplotlib** and **seaborn**.
- Focus on insights related to heart disease rates, gender, and age.

Data: Attributes

- **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

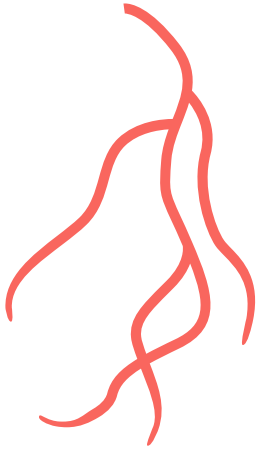
Codes Link :
https://drive.google.com/file/d/1ilzeTrCtM3claA6VlDw2dpZgEnJHll9x/view?usp=sharing

Null Value Check!



```
[5]: missing_values = df.isnull().sum()
     missing_values
```

```
[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
     target      0
     dtype: int64
```



Attribute Name Change

old names

```
df.columns
```

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

new names

age	sex
chest_pain_type	resting_blood_pressure
cholesterol	blood_sugar
electrocardiographic	maximum_heat_rate
exercise_induced_agina	oldpeak
slope	number_of_major_vessels
thalassemia	target

```
new_headers=["age","sex","chest_pain_type","resting_blood_pressure","cholestoral","blood_sugar","electrocardiographic","maximum_heart_rate","exercise_ind
df.columns=new_headers
df.columns
```

```
Index(['age', 'sex', 'chest_pain_type', 'resting_blood_pressure',
       'cholestoral', 'blood_sugar', 'electrocardiographic',
       'maximum_heart_rate', 'exercise_induced_angina', 'oldpeak', 'slope',
       'number_of_major_vessels', 'Thalassemia', 'target'],
      dtype='object')
```

02

Analysis and Insights

Statistics , Graphs and charts d

Basic Statistics

summary statistics (mean, median, standard deviation, min and max) for numerical attributes like age, blood pressure, cholesterol, and heart rate.

```
[13]: summary_stats=df.describe()
      attributes=['age','resting_blood_pressure','cholestoral','maximum_heart_rate']
      selected_stats=summary_stats.loc[['mean','50%','std',"max","min"],attributes]
      print("\nsummary statics:")
      print(selected_stats)
```

```
summary statics:
            age  resting_blood_pressure  cholestoral  maximum_heart_rate
mean  54.434146              131.611707    246.00000          149.114146
50%   56.000000              130.000000    240.00000          152.000000
std    9.072290               17.516718     51.59251           23.005724
max   77.000000              200.000000    564.00000          202.000000
min   29.000000               94.000000    126.00000           71.000000
```

It shows a middle-aged population with slightly elevated average blood pressure and cholesterol levels. Maximum heart rate varies widely but generally falls within expected physiological ranges.

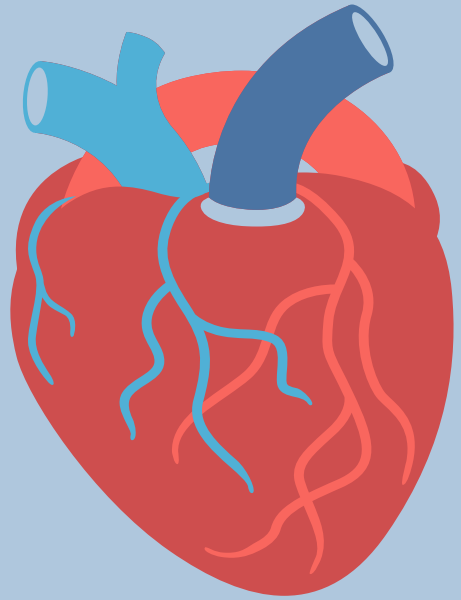
Basic Statistics

Prevalence : The prevalence of heart disease in the dataset is calculated as 51.32%, indicating that slightly over half of the individuals in the sample have been diagnosed with heart disease.

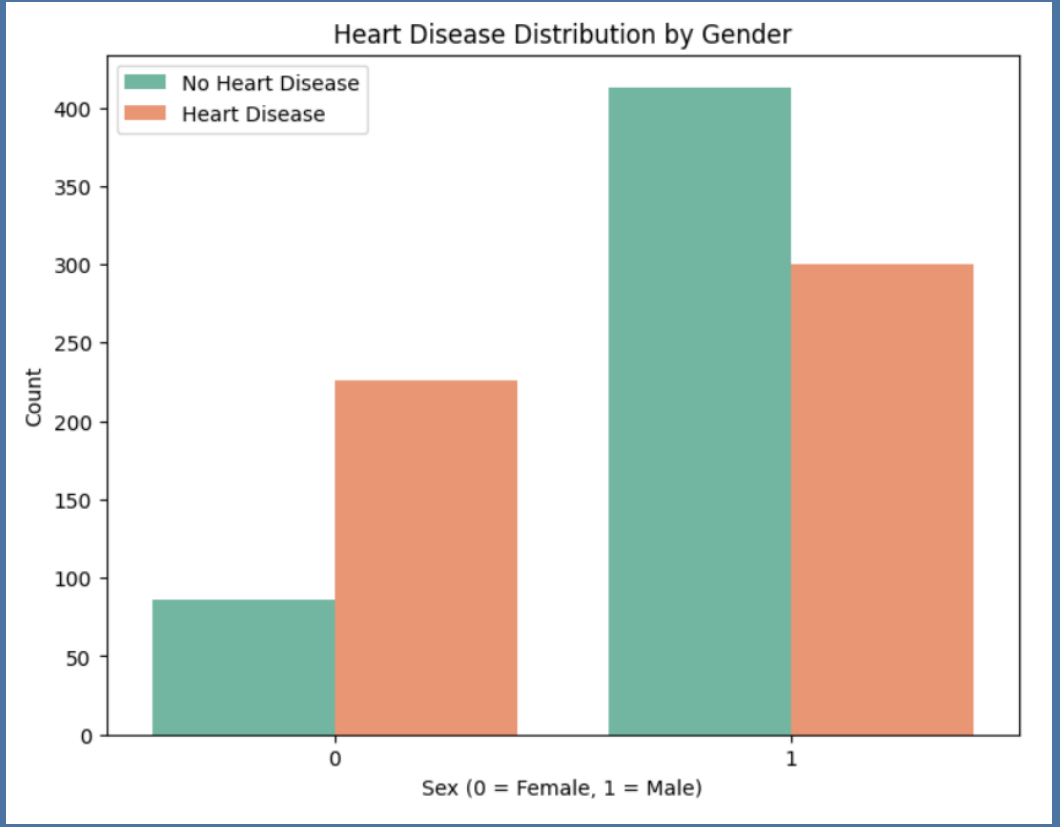
```
[15]: prevalence=df["target"].mean()*100
      print(f'Prevalence of Heart Disease in the Dataset:{prevalence:.2f}%')

      Prevalence of Heart Disease in the Dataset:51.32%
```

Distribution : Out of 1025 individuals in the dataset, approximately 69.56% are male (713 individuals), and 30.44% are female (312 individuals). This gender distribution is crucial for understanding the prevalence of heart disease across genders in the sample.

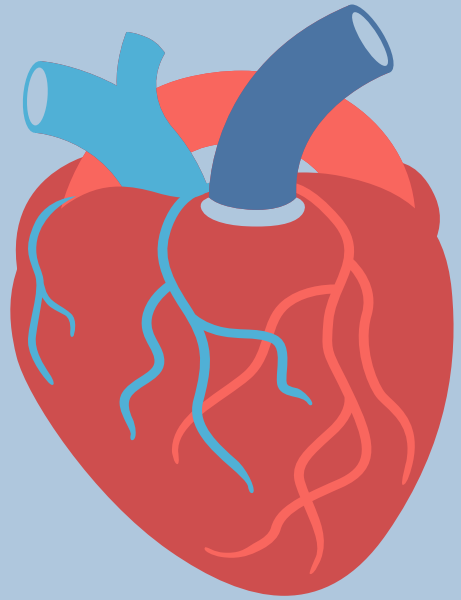
```
[53]: males = df['sex'].sum()
      males

[53]: 713
```



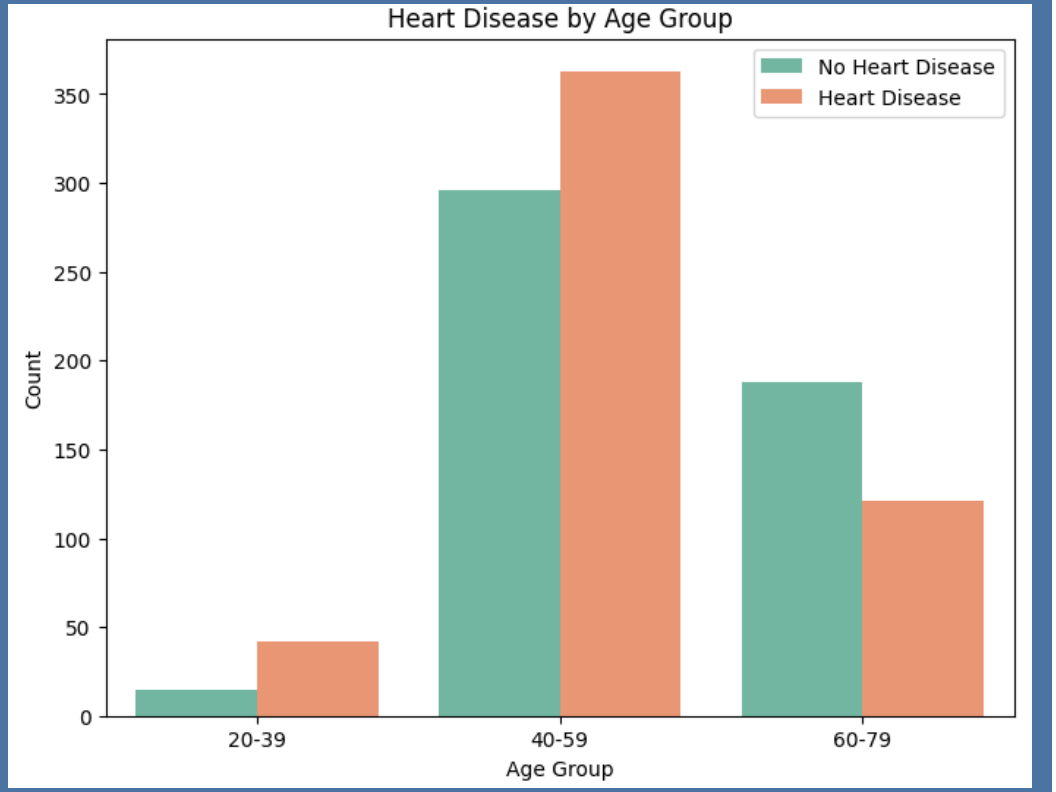
Distribution by Gender



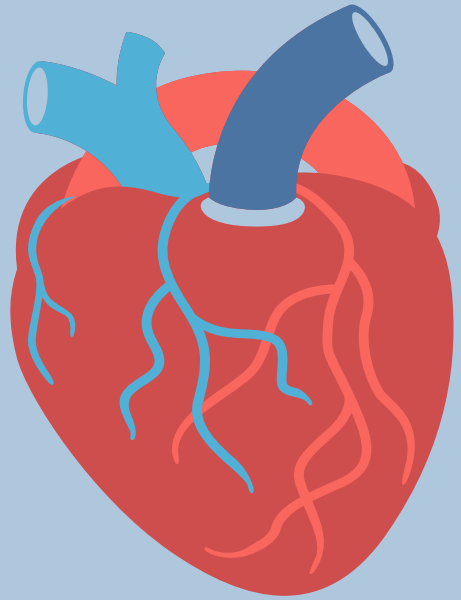
Heart disease is more prevalent among men, consistent with a higher proportion of male samples in the dataset. This suggests that while males outnumber females in the dataset, both genders exhibit similar percentages of heart disease prevalence.



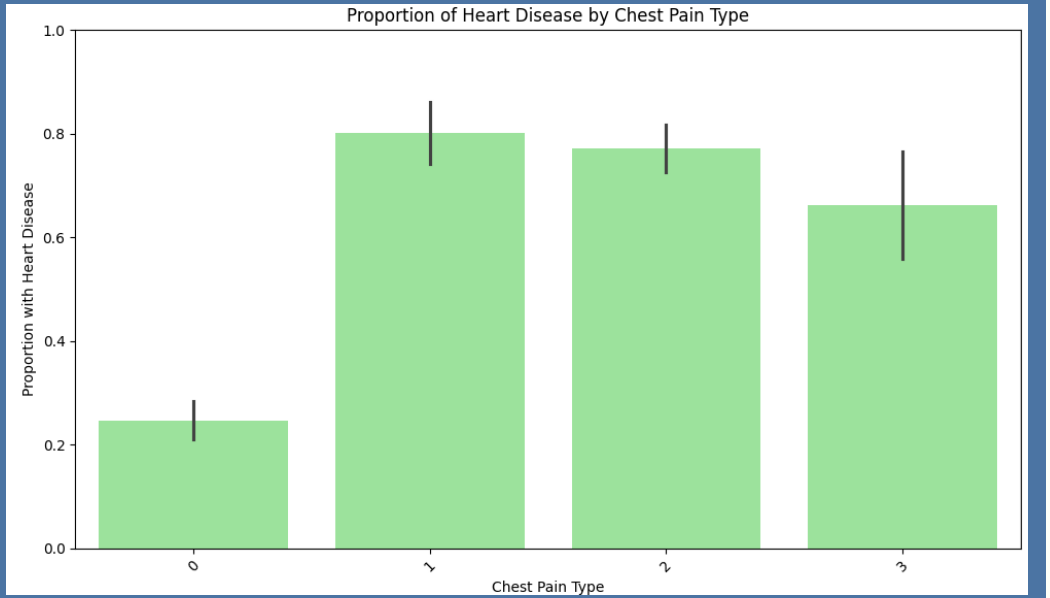
Distribution by Age Group



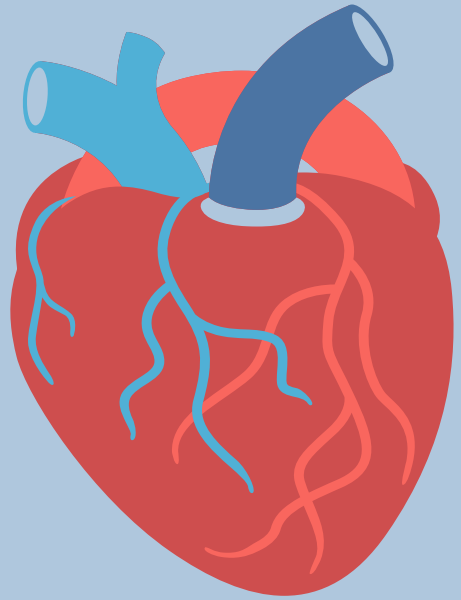
The prevalence of heart disease increases with age, peaking in midlife.



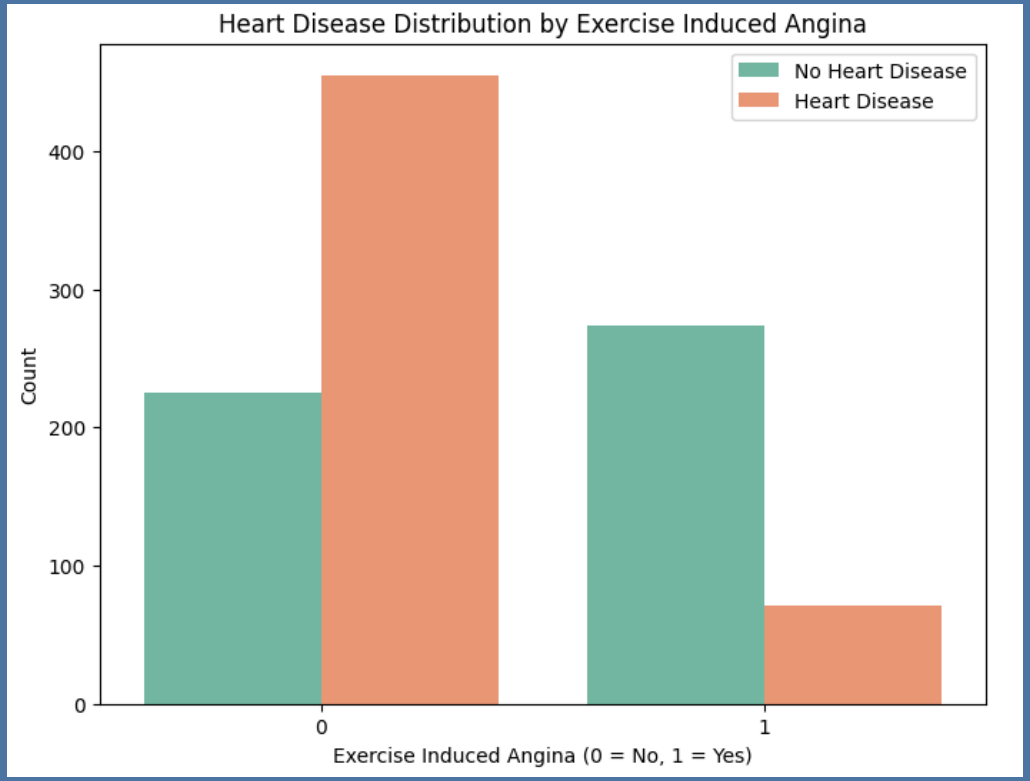
Distribution with Type of Chest Pain



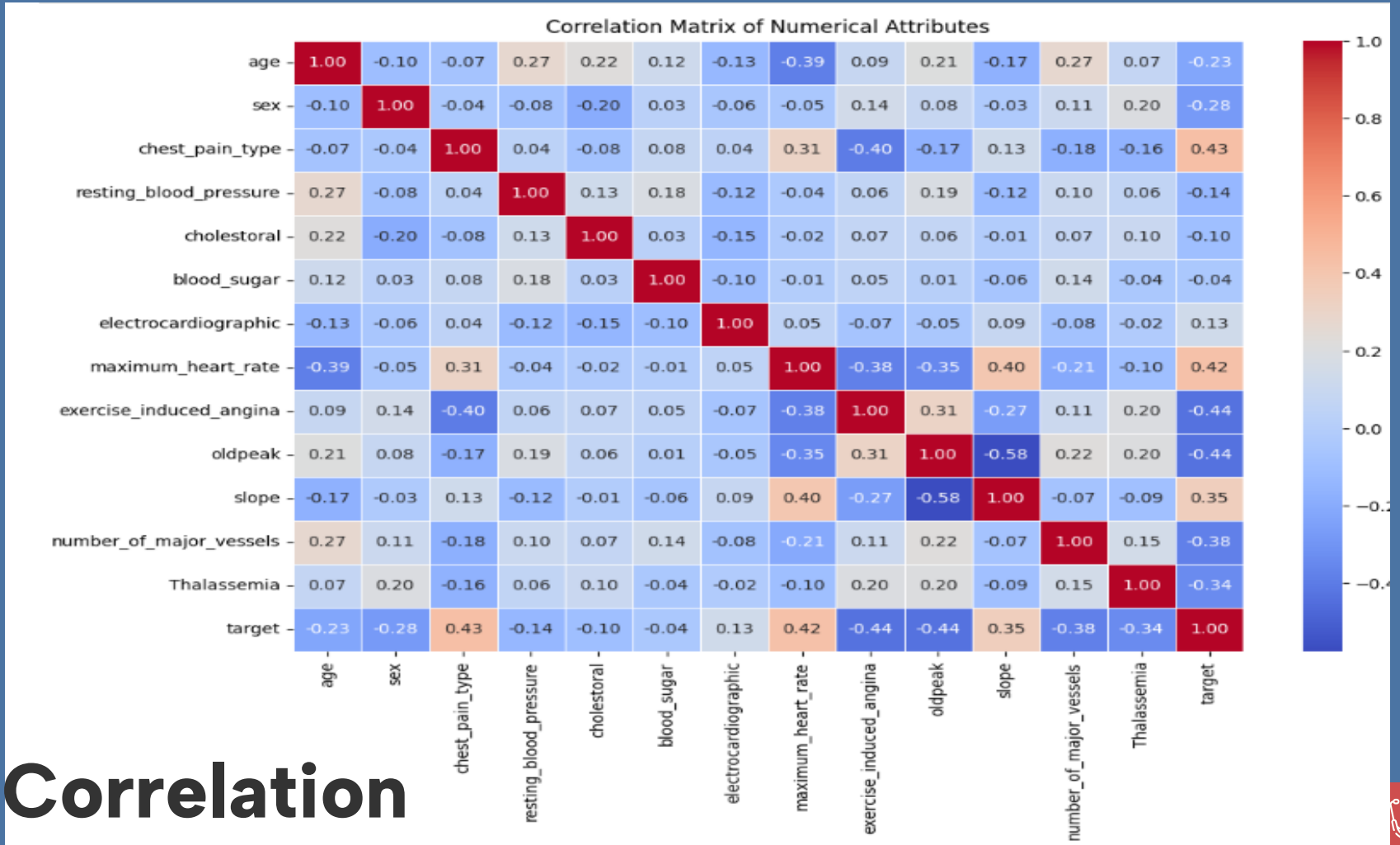
Type 1 and 2 chest pain types indicate higher chances of having heart disease, with type 1 showing approximately 80% and type 2 around 70% likelihood based on data analysis.



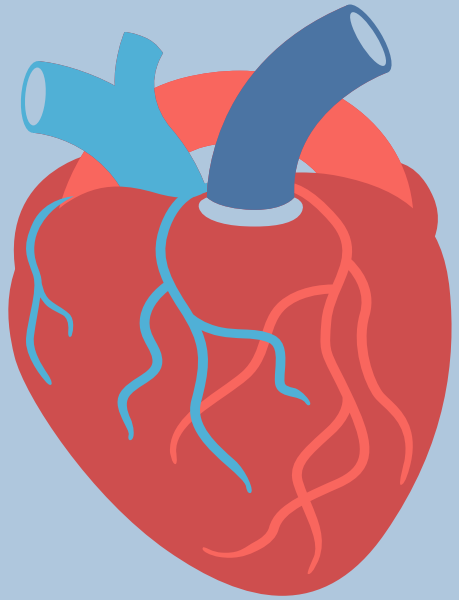
Exercise induced Agina by heart Disease



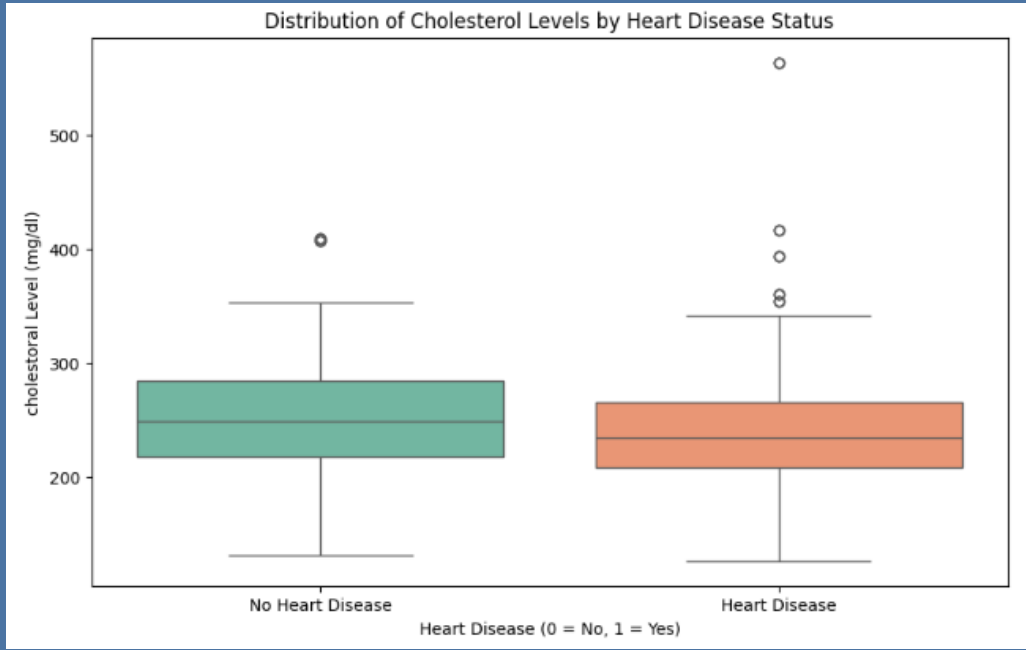
Most of the sample does not have exercise-induced angina, suggesting that heart disease factors related to this condition are minimal in this particular sample.



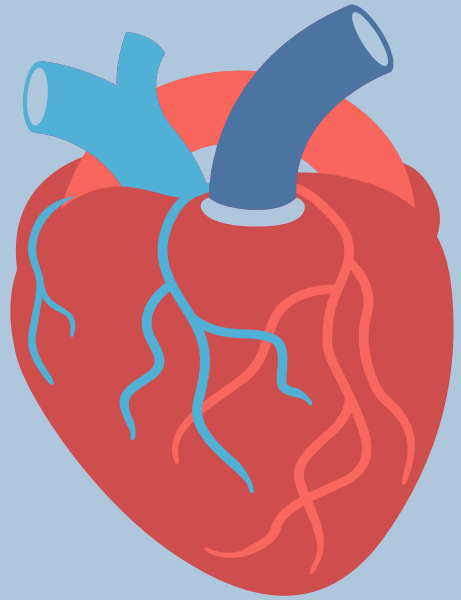
Correlation



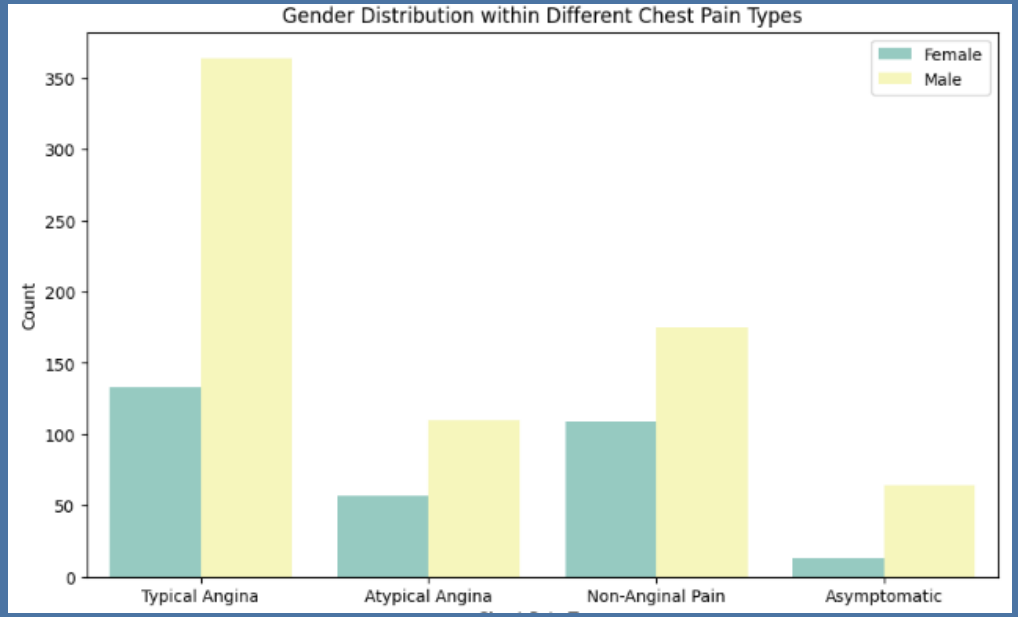
Heart Disease by Cholesterol



Cholesterol level among those are way over range for Heart Disease Patient.
There are higher outlier with heart disease and cholesterol.



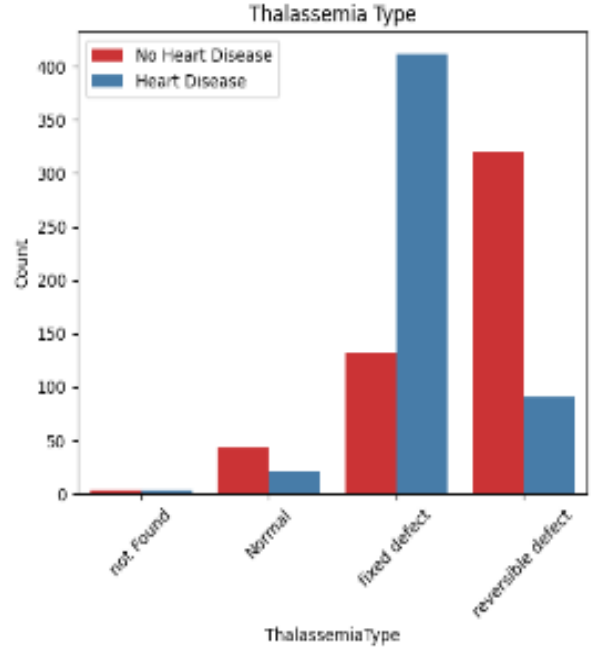
Chest Pain with Gender

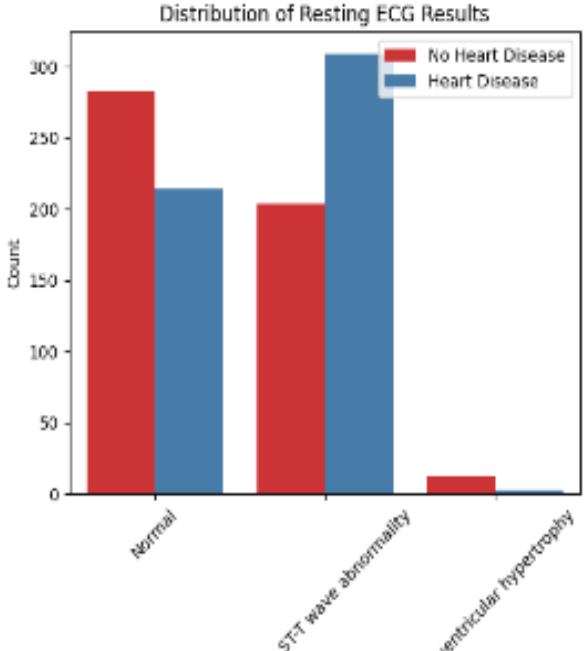


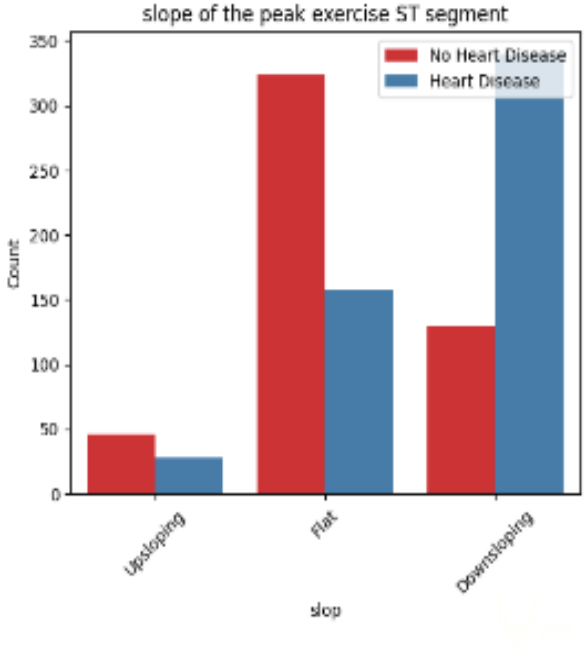
Most Cases are of Typical Agina and in which too are more of male patient.

More Insights

Following shows more heart Disese with different attribute distribution.

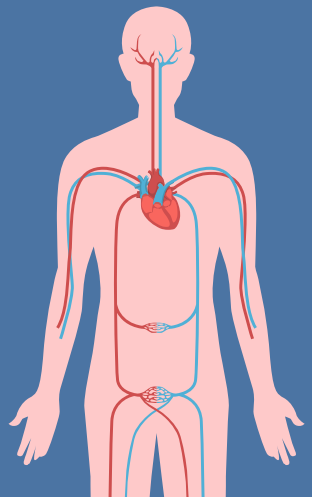


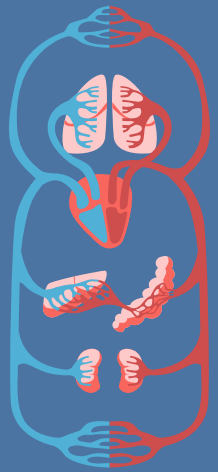


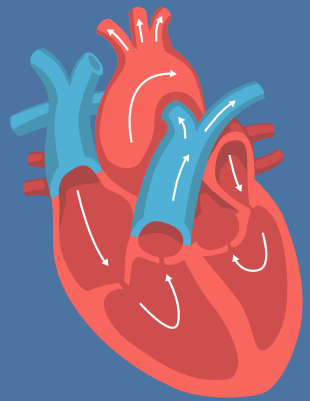


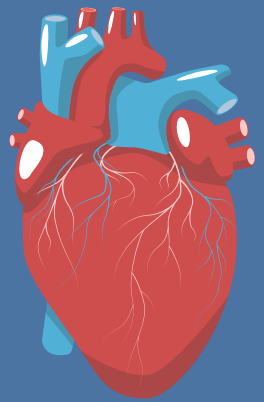
Conclusion

- **Age**: Average 54.4 years (Range: 29–77)
- **Resting Blood Pressure**: Average 131.6 mmHg (Range: 94–200 mmHg)
  - Suggests slight elevation compared to normal (<120/80 mmHg)
- **Cholesterol Level**: Average 246 mg/dL (Range: 126–564 mg/dL)
  - Elevated above the healthy range (<200 mg/dL)
- **Maximum Heart Rate**: Average 149.1 bpm (Range: 71–202 bpm)
  - Slightly lower than expected for similar age groups

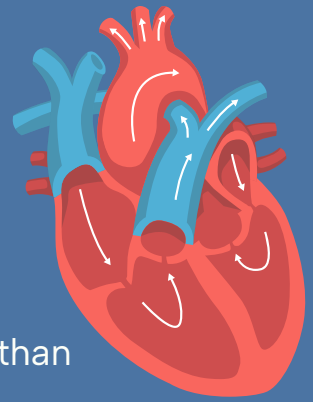




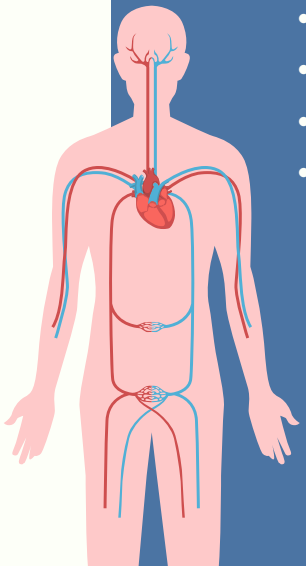


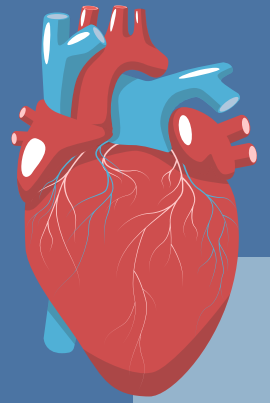


Conclusion

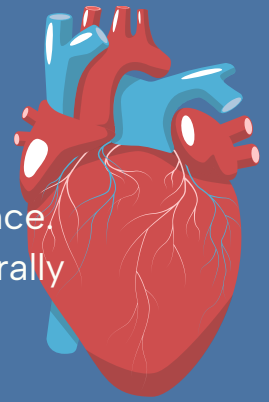


- **Prevalence of Heart Disease**: 51.32% of the sample has heart disease.
- **Gender Analysis**:
  - Among those with heart disease, females outnumber males.
  - Conversely, among those without heart disease, there are more males than females.
- **Age Groups and Heart Disease**:
  - Individuals aged 40–59 years show the highest prevalence of heart disease.
  - The 20–39 age group has lower prevalence compared to 40–59 years.
  - The 60–79 age group exhibits the lowest prevalence.
  - The 40–59 age group is not only most affected by heart disease but also constitutes the largest segment of the sample.





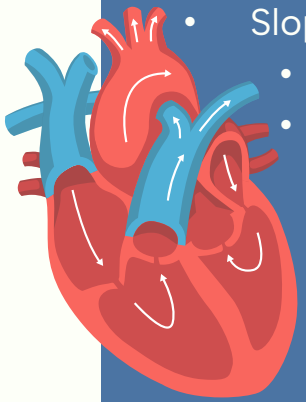
Conclusion

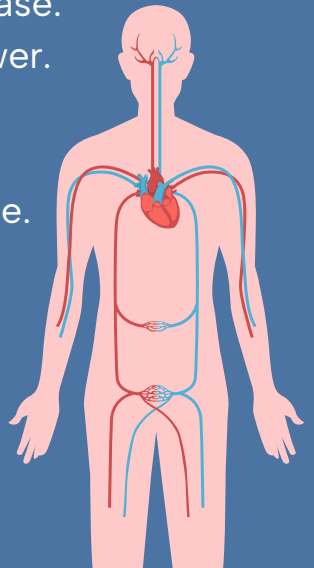


- Thalassemia Type :
  - Reversible' or 'fixed defect' types link to higher heart disease prevalence.
  - 'Normal' type shows lower incidence, 'not found' type mixed but generally lower prevalence.

- Resting ECG Results :
  - 'Left ventricular hypertrophy' correlates with higher heart disease.
  - 'Normal' and 'ST-T wave abnormality' vary, 'Normal' usually lower.

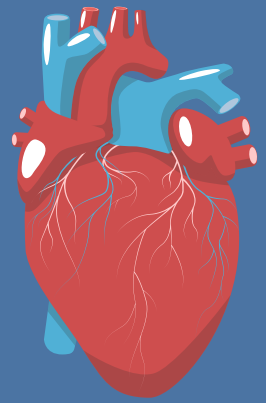
- Slope of Peak Exercise ST Segment :
  - 'Downsloping' ST segment associates with higher heart disease.
  - 'Flat' and 'upsloping' vary, 'upsloping' generally lower.

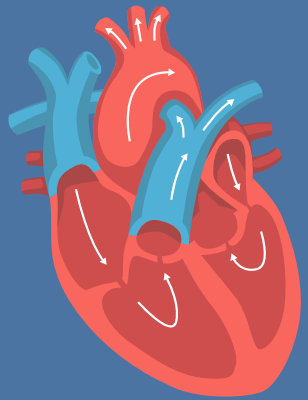


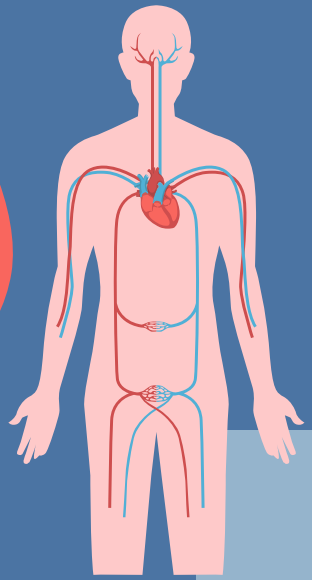


Conclusion

- Correlation Matrix :
  - Age, max heart rate, and possibly cholesterol correlate with heart disease.
- Scatter Plot: Age vs. Max Heart Rate :
  - Heart disease linked to lower max heart rates with age.
- Gender and Chest Pain Types :
  - Gender impacts chest pain reporting.
- Cholesterol Levels (Box Plot) :
  - Higher in heart disease cases.
- Major Vessels Colored (Bar Plot) :
  - More vessels colored indicates more heart disease cases.







Thanks!

Do you have any questions?

israr4075@gmail.com
+34 654 321 432
yourwebsite.com