HEART DISEASE DIAGNOSTIC ANALYSIS

RAGI RAJESH

July - 2024

Project Title

Heart Disease Diagnostic Analysis

Technologies

- Data Science
- Python
- Matplotlib
- Seaborn
- Plotly

Domain

Healthcare

Project Overview

Health is a real wealth, and during the pandemic, we all realized the severe effects of COVID-19 on everyone, regardless of status. This project aims to analyze health and medical data to better prepare for the future.

ETL (Extract, Transform, Load)

Extract: The dataset is extracted from a heart disease diagnostic database, which includes various attributes such as age, gender, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, and the presence of heart disease.

Transform: The data is cleaned and transformed for analysis, including handling missing values and converting numerical data into categorical data where necessary.

Load: The transformed data is loaded into a dataframe for analysis.

Exploratory Data Analysis (EDA)

EDA is performed to understand the distribution and relationships of the data attributes.

Statistical Analysis:

- Minimum Age: `Min_Age`
- Maximum Age: `Max_Age`
- Mean Age: `Mean_Age` **Age Distribution:**

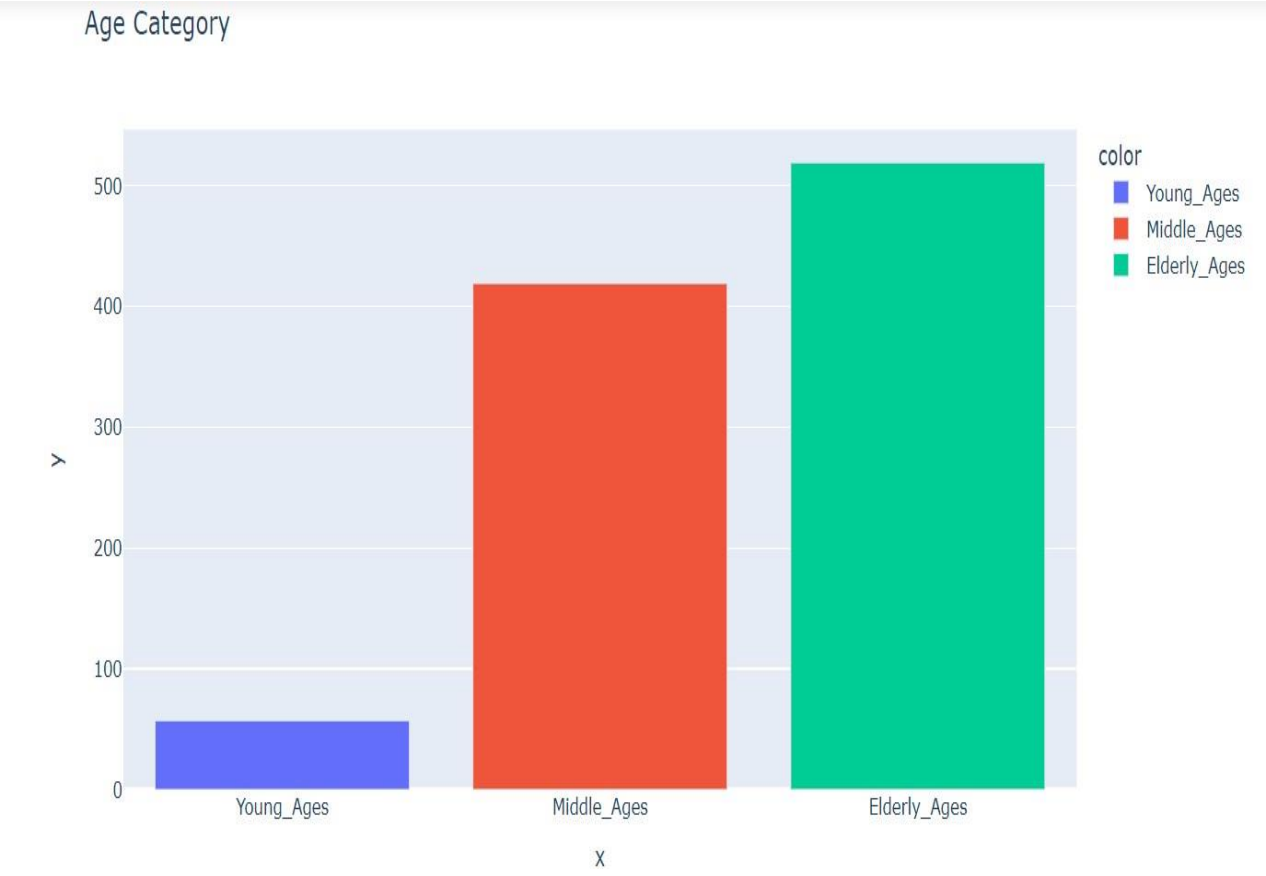
- Young Ages (29-40): `len(Young_Ages)`
- Middle Ages (40-55): `len(Middle_Ages)`
- Elderly Ages (>55): `len(Elderly_Ages)` **Gender Distribution:**

- Male: `count`
- Female: `count`

Data Visualization

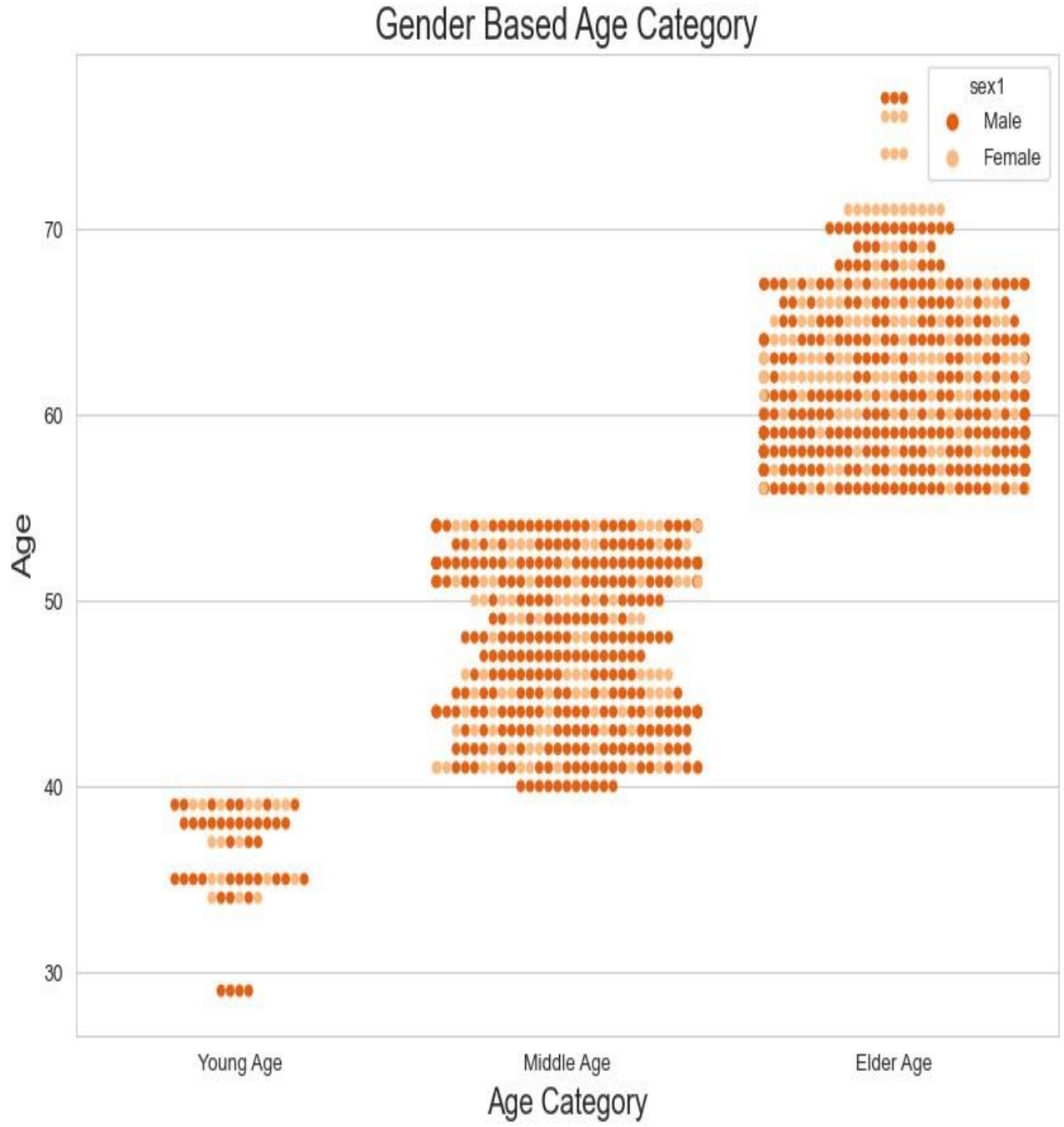
Various visualizations are created to understand the data better.

Bar Plot of Age Categories:



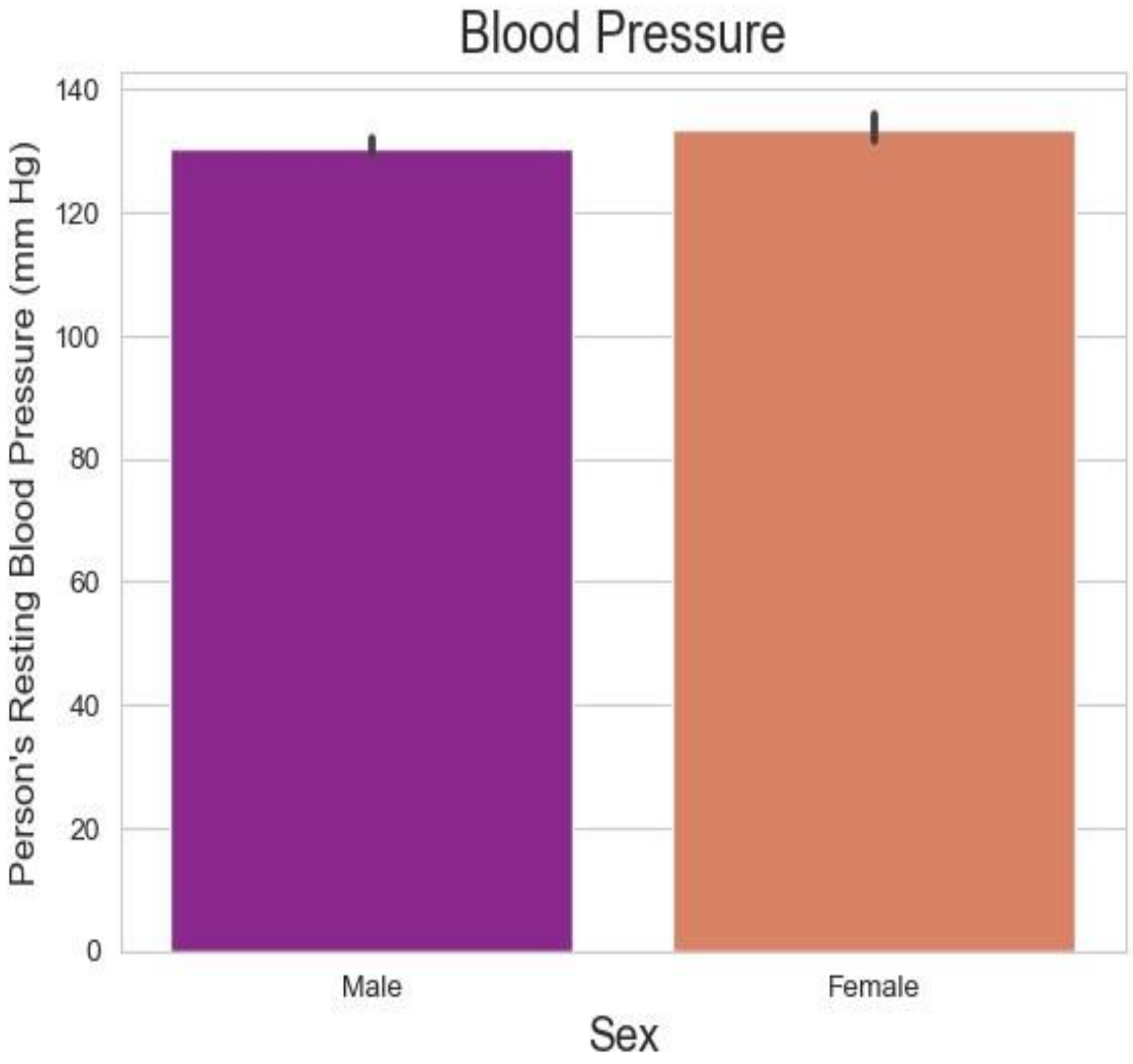
The bar plot illustrates the distribution of patients across different age categories: Young Ages (29-40), Middle Ages (40-55), and Elderly Ages (>55). It helps in understanding which age group has the highest number of patients.

Swarm Plot of Gender-Based Age Categories:



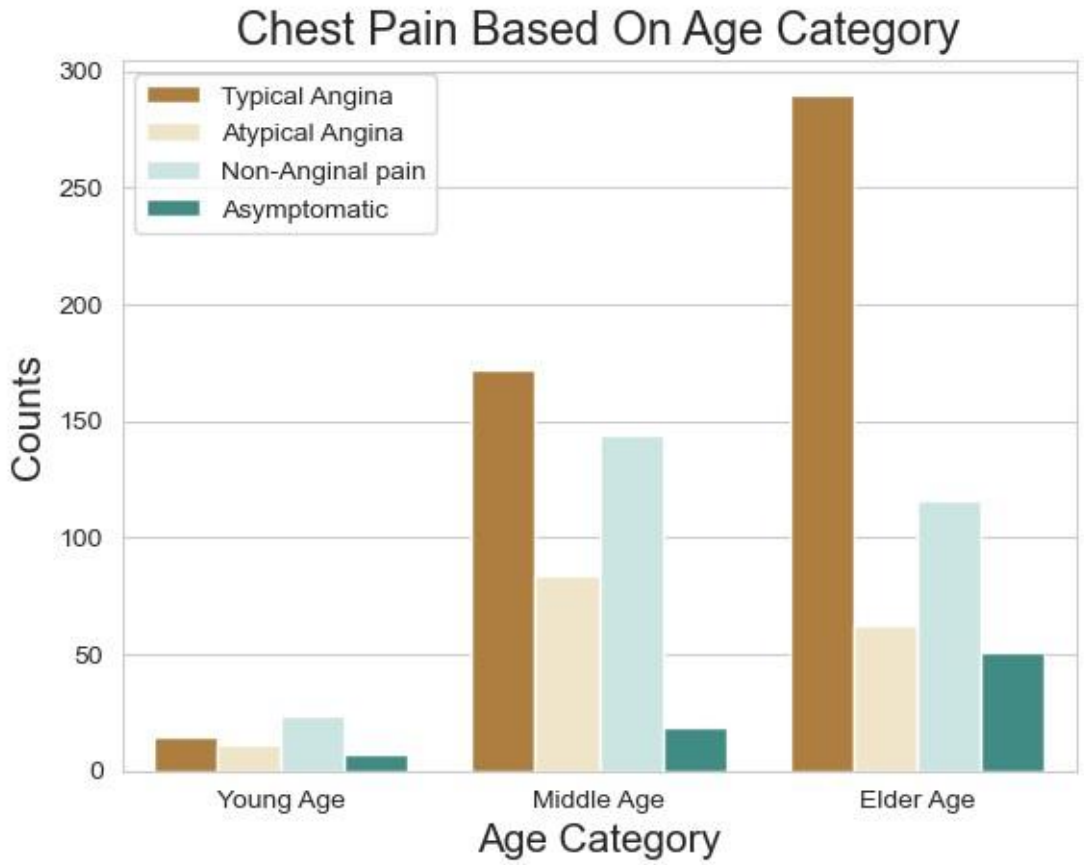
The swarm plot shows the distribution of patients across age categories based on gender, highlighting differences in age distribution between males and females.

Bar Plot of Blood Pressure by Gender:



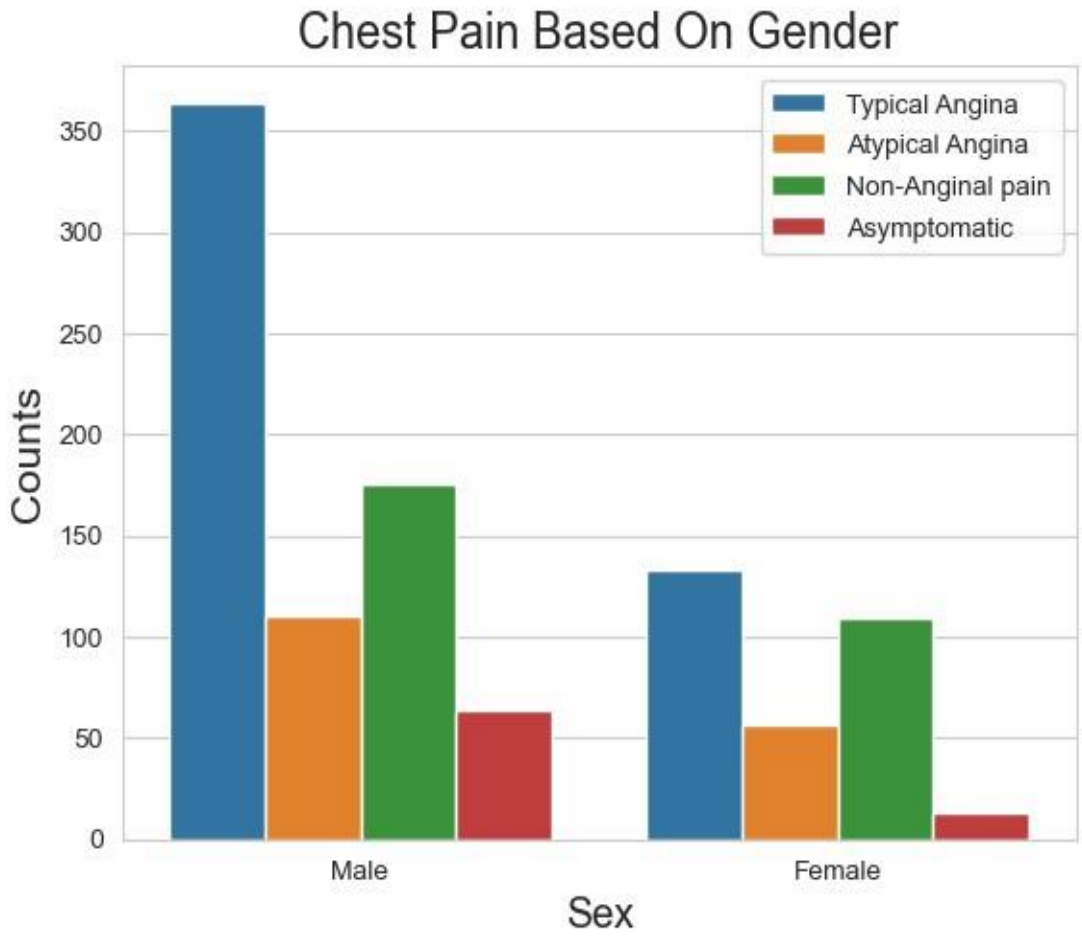
The bar plot displays the average resting blood pressure of patients categorized by gender, allowing comparison of resting blood pressure between males and females.

Count Plot of Chest Pain by Age Category:



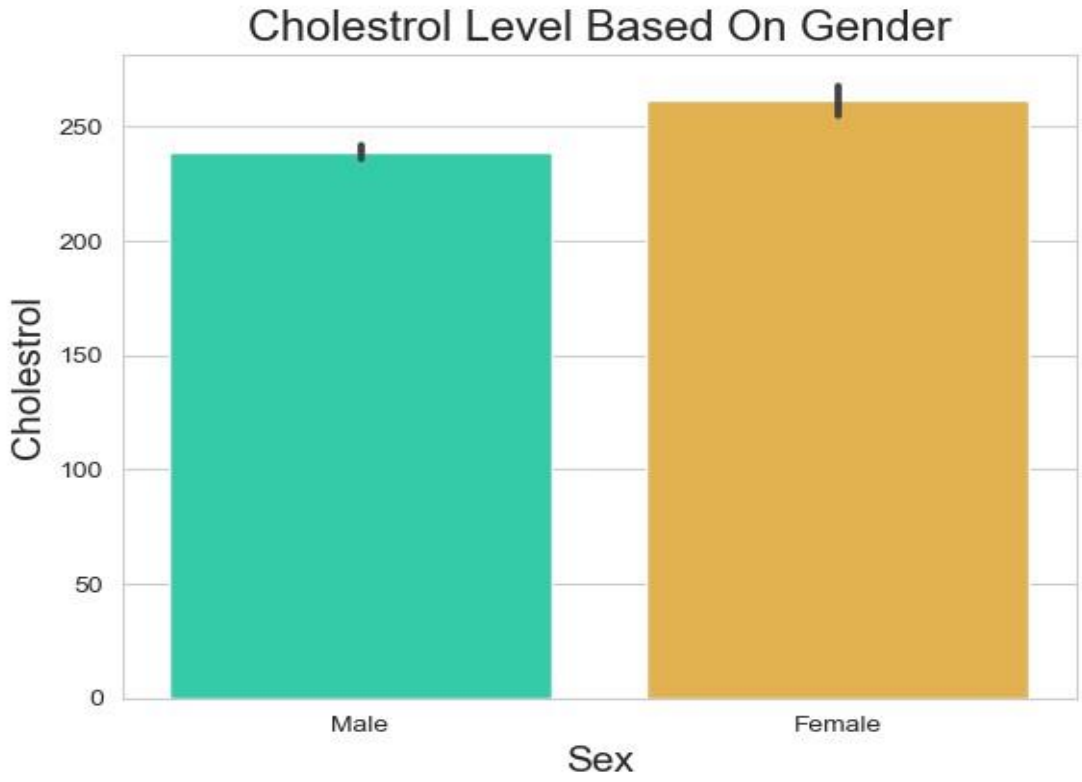
The count plot shows the distribution of chest pain types (typical angina, atypical angina, non-anginal pain, asymptomatic) across different age categories. It helps identify which age group experiences more chest pain.

Count Plot of Chest Pain by Gender:



A count plot showing chest pain types across genders compares how often each type of chest pain occurs in males and females. It helps us see if certain chest pain types are more common in one gender than the other. For example, the plot might show that males more frequently have typical angina, while females might experience atypical angina more often. This visualization highlights any gender-specific patterns in chest pain.

Bar Plot of Cholesterol Levels by Gender:

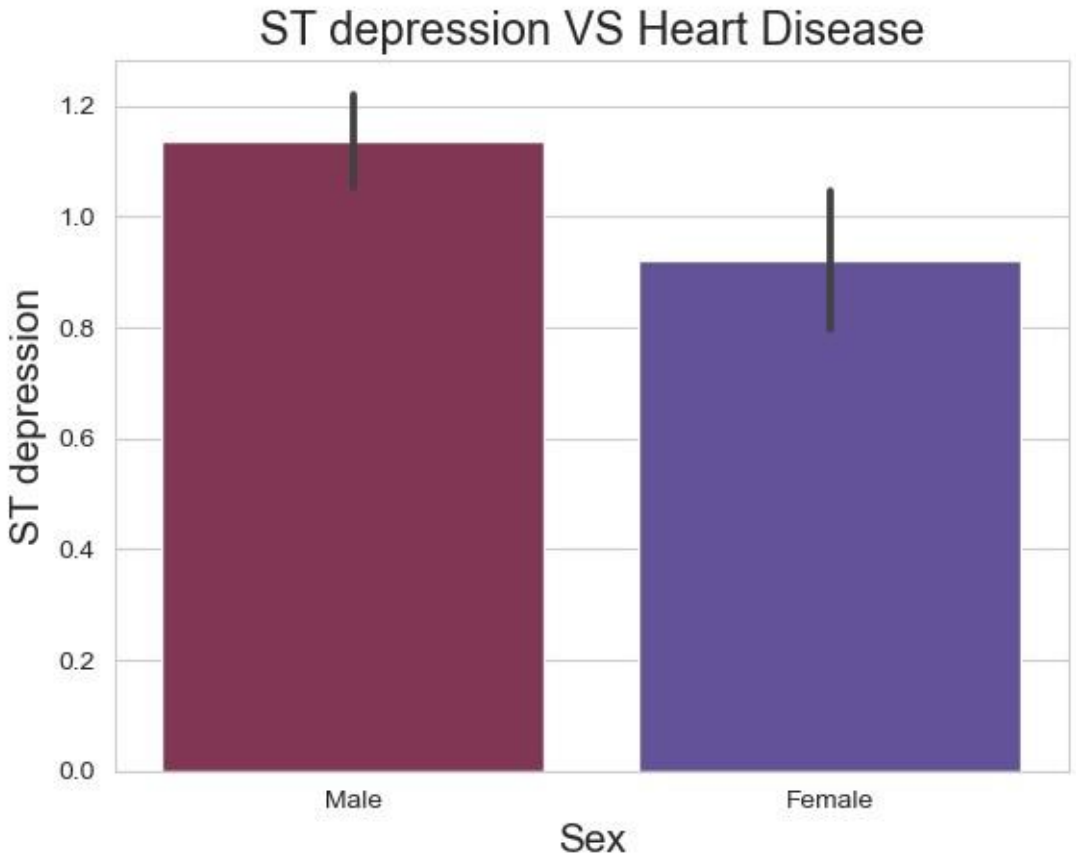


The bar plot compares the average cholesterol levels between male and female patients, providing insights into gender-based differences in cholesterol levels.

Line Plot of ST Depression vs Age:

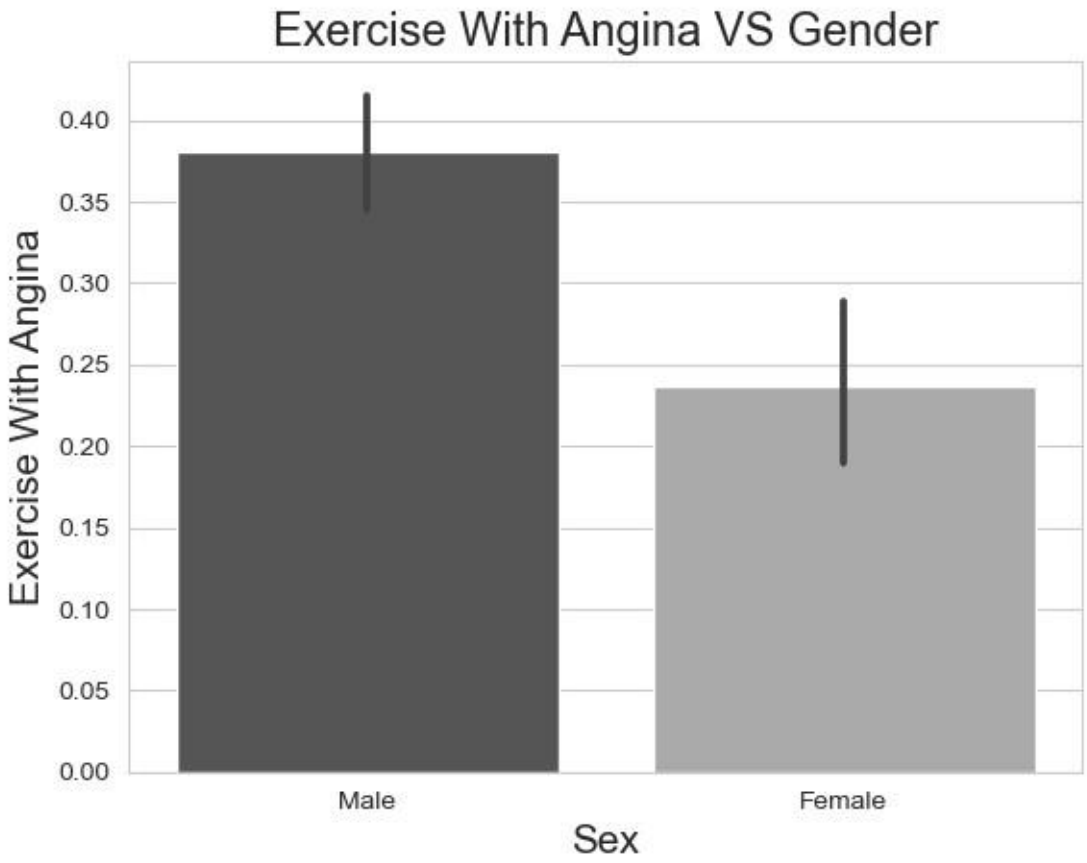
The line plot depicts the relationship between age and ST depression values, showing how ST depression varies across different age groups.

Bar Plot of ST Depression vs Heart Disease:



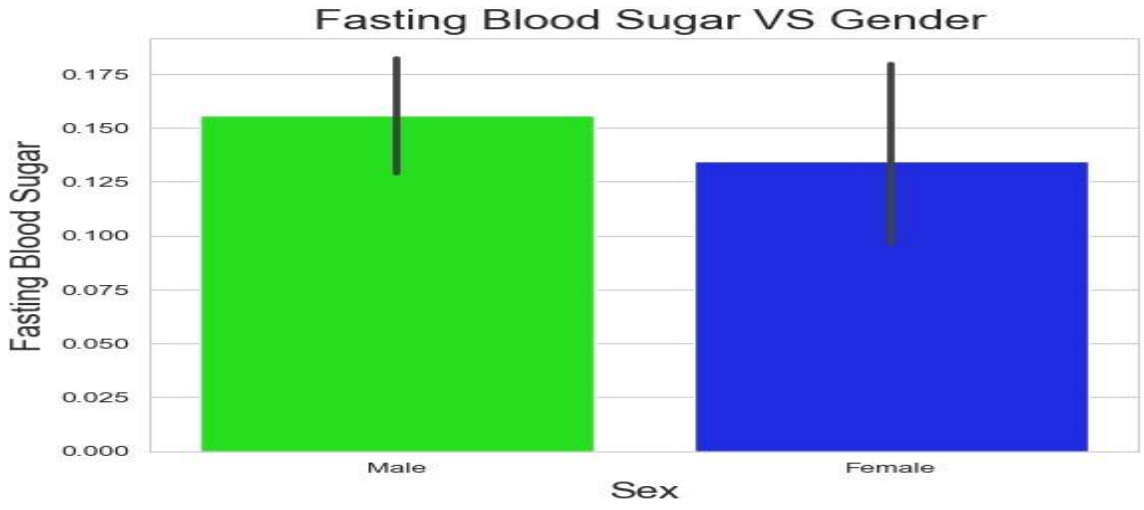
This bar plot shows the average ST depression levels in patients with and without heart disease, highlighting the relationship between ST depression and heart disease.

Bar Plot of Exercise-Induced Angina by Gender:



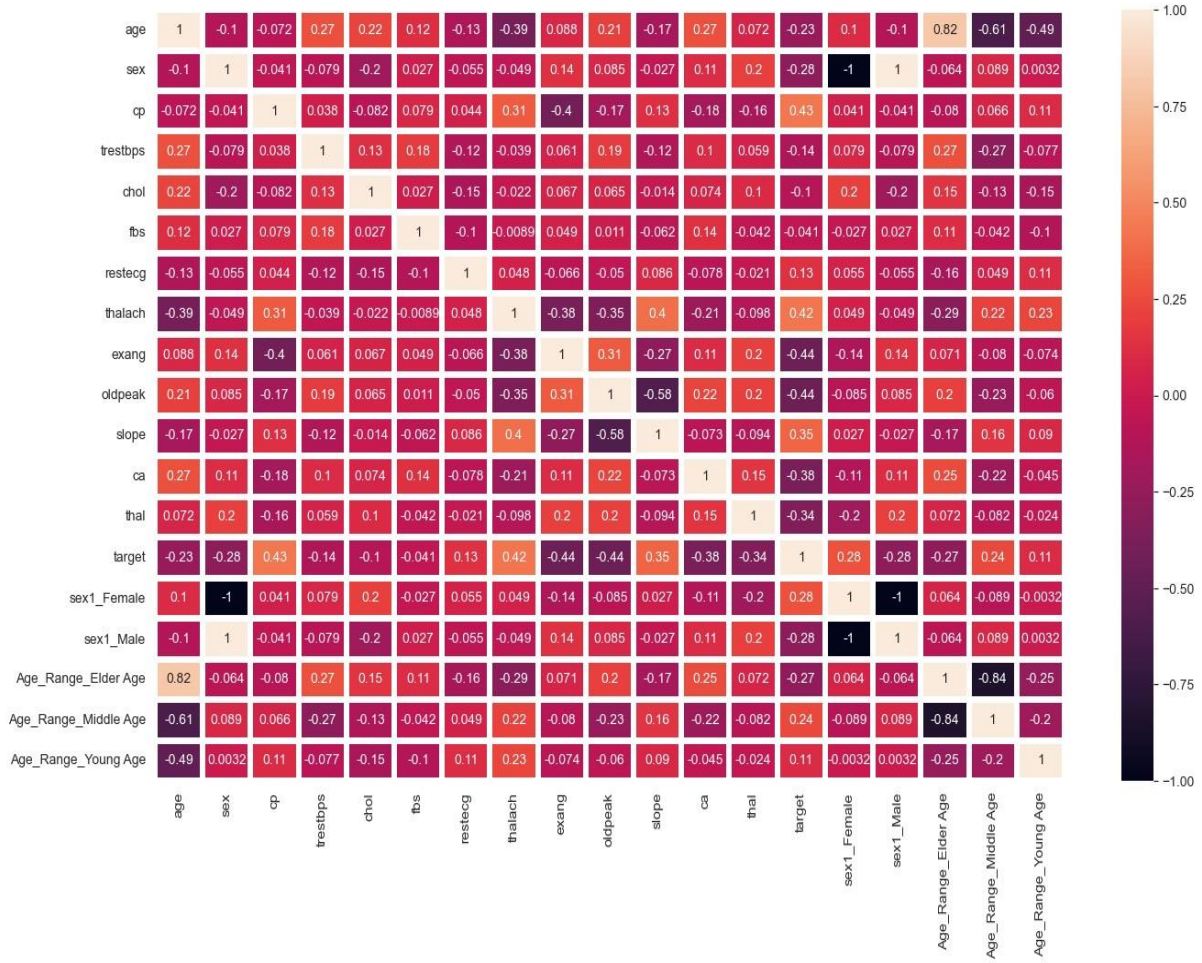
The bar plot compares the prevalence of exercise-induced angina between male and female patients, helping to understand gender-based differences in angina occurrence.

Bar Plot of Fasting Blood Sugar by Gender:



The bar plot shows the average fasting blood sugar levels for male and female patients, illustrating the differences in fasting blood sugar levels between genders.

Heatmap of Correlations:



The heatmap visualizes the correlation between different variables in the dataset, helping to identify which variables are strongly or weakly correlated with each other.

Machine Learning Model

A Logistic Regression model is trained to predict heart disease based on the attributes.

Model Training and Evaluation:

- The dataset is split into training and testing sets.
- The model is trained and evaluated using Mean Squared Error (MSE), R-squared, and accuracy.

Model Performance:

- Mean Squared Error: `mse`
- R-squared: `r2`
- Accuracy: `accuracy`

Prediction Results: Error! Filename not specified.

Predictive Analysis

The model is used to predict heart disease status for new data.

```
new_data = pd.DataFrame({
    'age': [63],
    'sex': [0],
    'cp': [2],
    'trestbps': [135],
    'chol': [252],
    'fbs': [0],
    'restecg': [0],
    'thalach': [172],
    'exang': [0],
    'oldpeak': [0],
    'slope': [2],
    'ca': [0],
    'thal': [2]
})  heart_prediction =
model.predict(new_data) if
heart_prediction == 1:
    print("The person has a heart problem") else:
print("The person does not have a heart problem")
```

Conclusion

This project provides a comprehensive analysis of heart disease diagnostics, highlighting key factors such as age, gender, cholesterol levels, and blood pressure. The insights derived from the data can aid in better understanding and managing heart disease risks.