Heart Disease Analysis Report

📌 Project Objective

To explore a heart disease dataset and identify key medical factors that correlate with the likelihood of having heart disease. This report summarizes the data analysis, key insights, and possible implications.

📄 Dataset Overview

- Source: Kaggle (Heart Disease Dataset)

- Observations: 918 records

- Features: 12 health-related attributes including age, cholesterol, chest pain type, blood pressure, fasting blood sugar, and target (disease presence)

🛠️ Tools & Technologies Used

- Python 3

- Jupyter Notebook

- Pandas, Numpy

- Seaborn, Matplotlib

- MySQL

🔍 Data Exploration Highlights

- The data types for all columns were appropriate and required minimal correction.

- Outliers were observed, especially a cholesterol value of 0, which is clinically invalid.

- Most individuals had cholesterol levels between 150–360 mg/dL.

- Distribution plots showed balanced spread in age and chest pain types, with notable skew in fasting blood sugar.

📊 Correlation Analysis

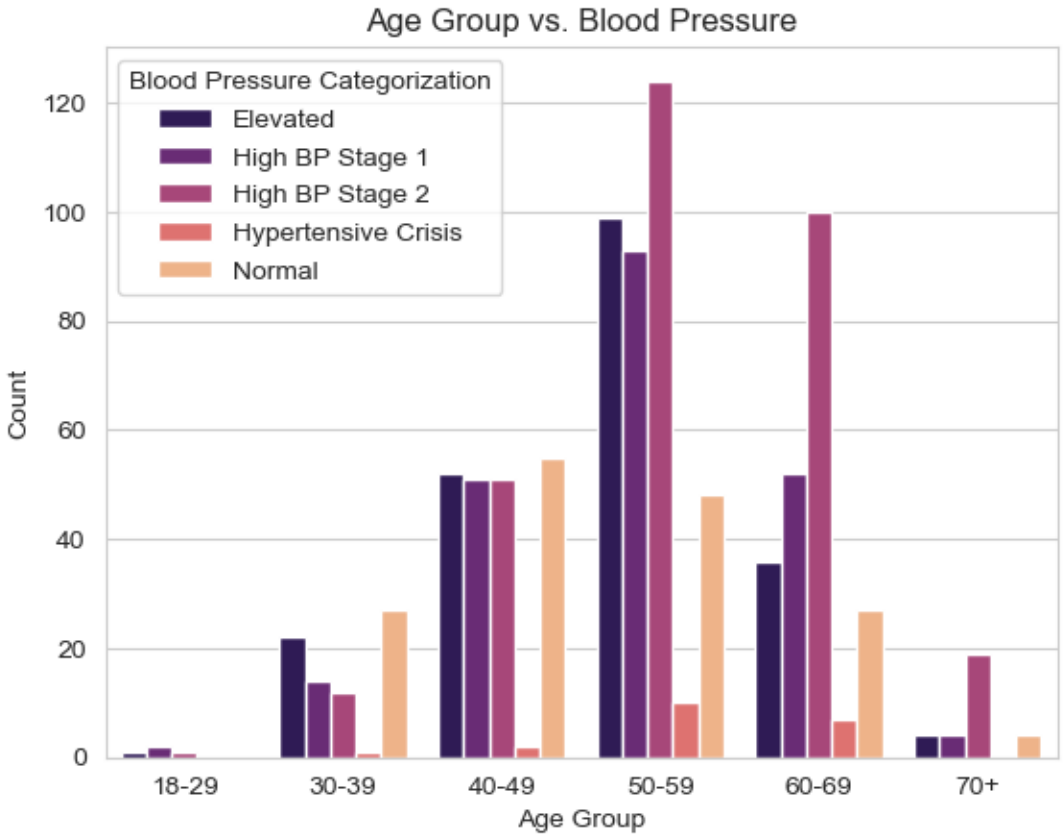
- Oldpeak showed the strongest positive correlation with heart disease.

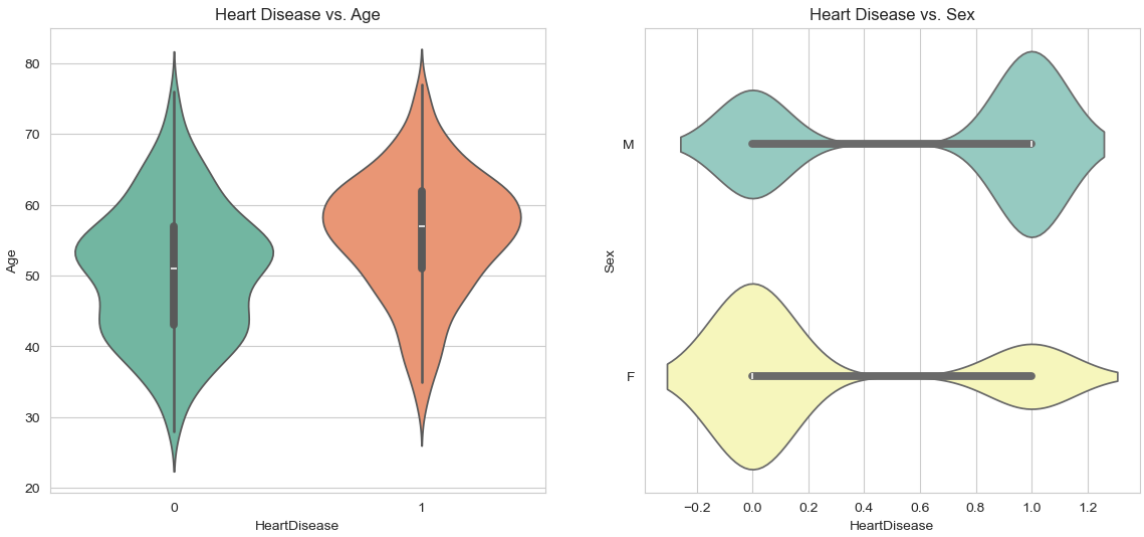
- Fasting blood sugar & Resting blood pressure showed almost zero correlation with the heart disease.

- Max heart rate showed a strong negative correlation with heart disease.

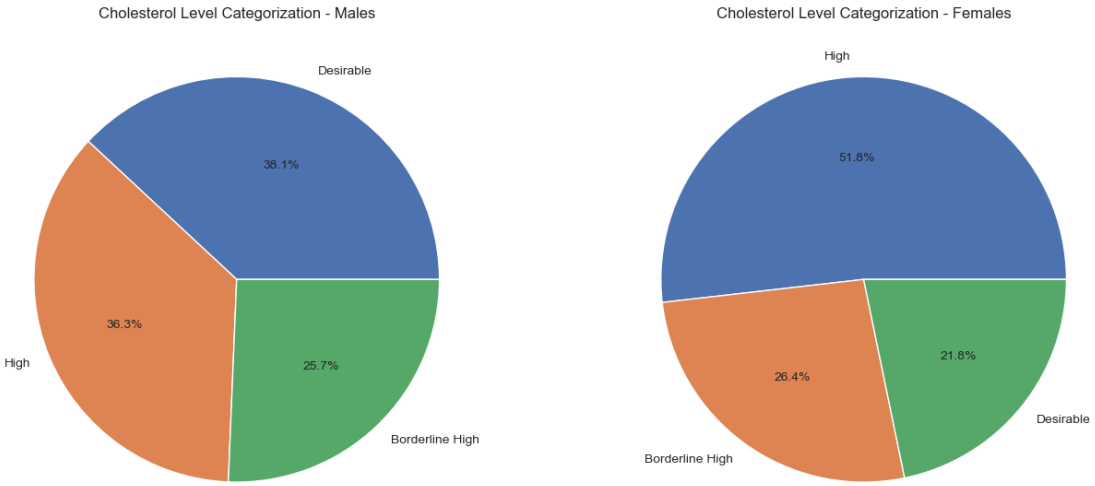
- Surprisingly, cholesterol had a weak negative correlation, possibly due to external variables (e.g., medication).

Sample Visual Insights

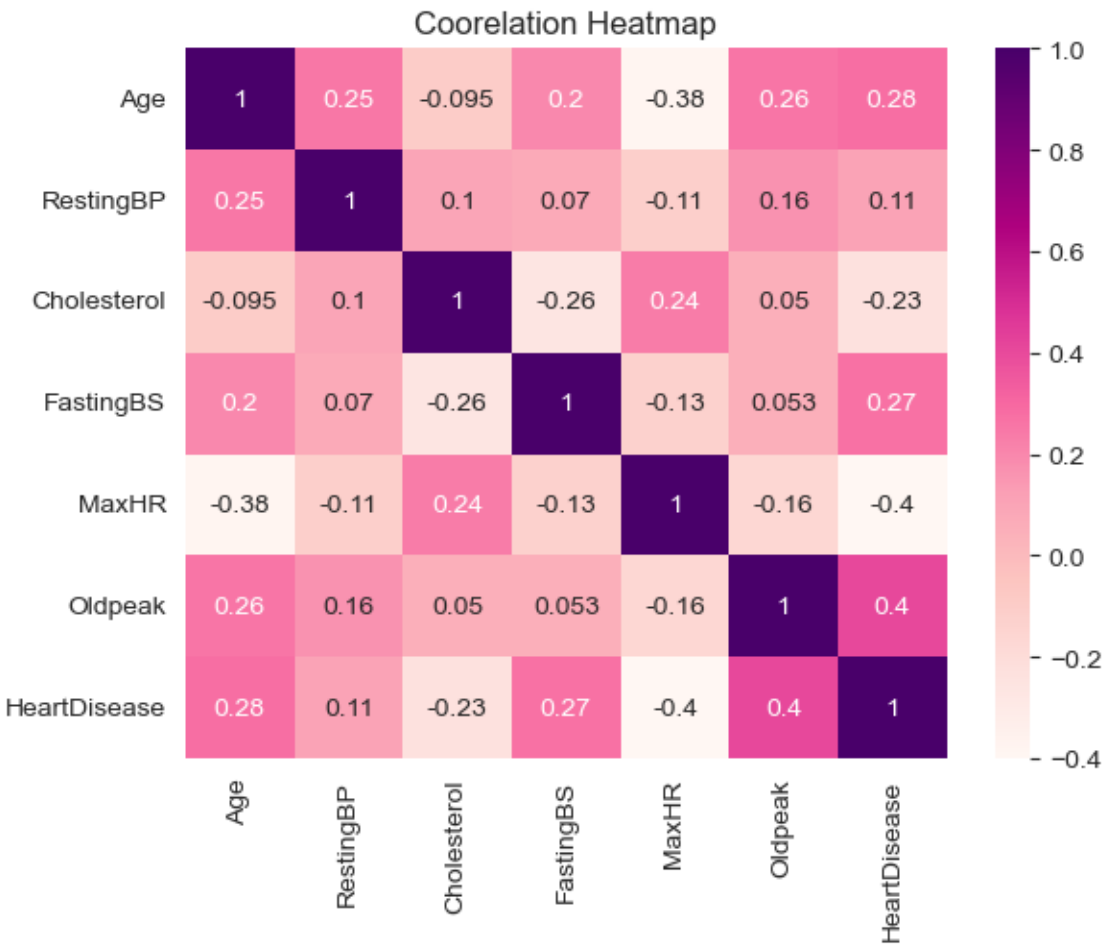




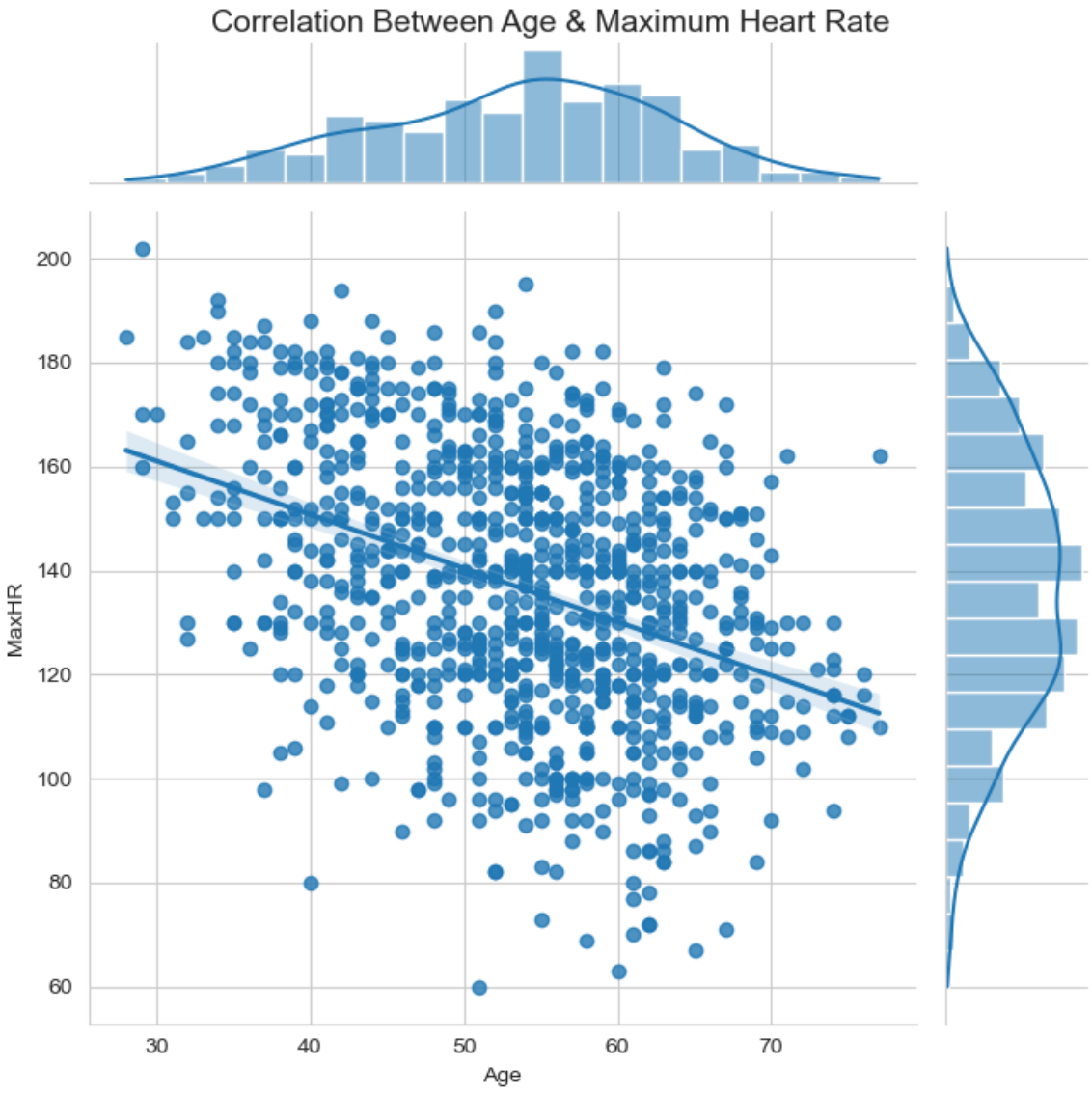
Heart Disease – Age & Sex



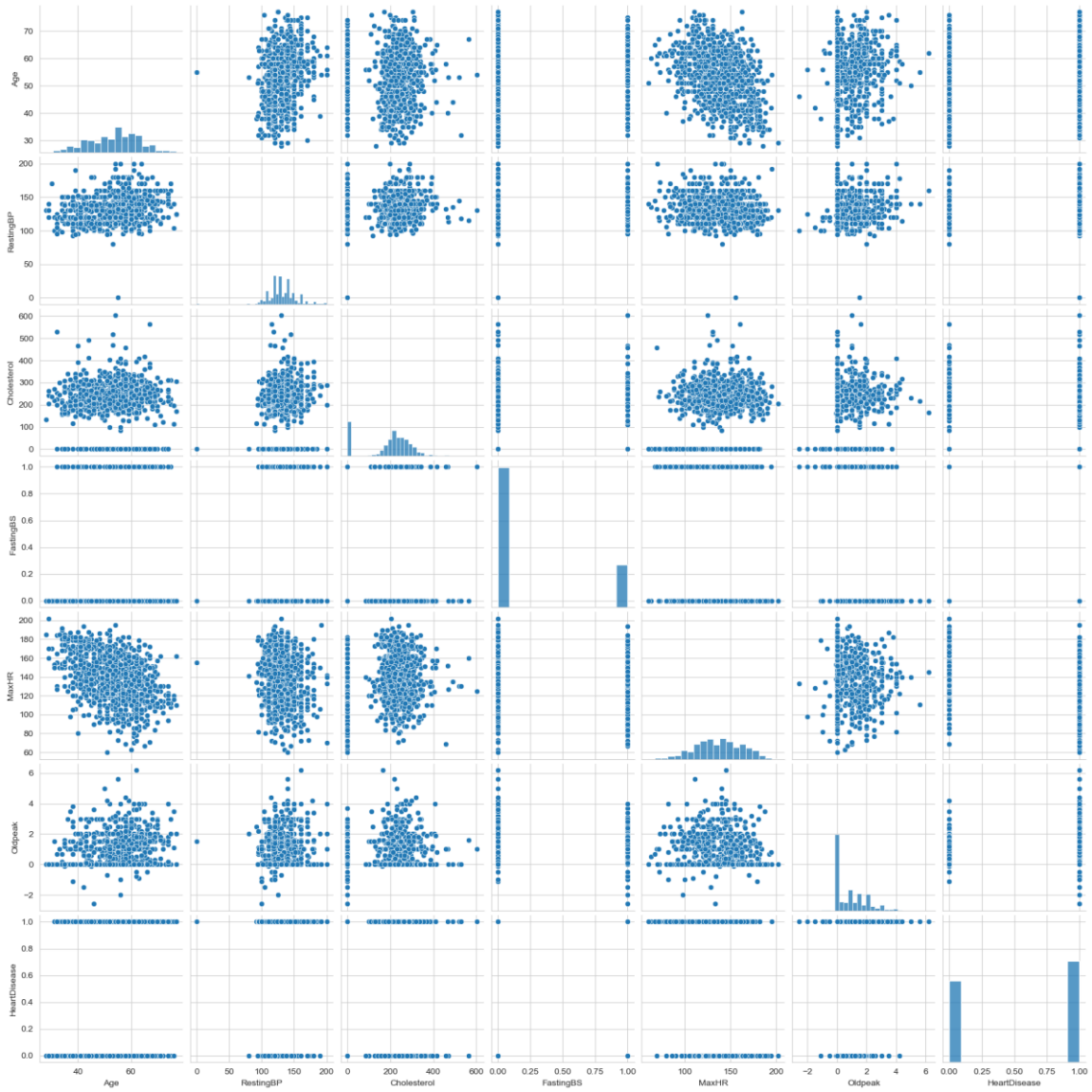
Cholesterol Level Categorization – Males & Females



Correlation Between Each Factor



Correlation Between Age & Maximum Heart Rate



Correlation Between Each Factor

☑ Final Conclusion

This exploratory data analysis on the heart disease dataset provided several important observations:

1. Age, Fasting Blood Sugar, and Oldpeak showed a positive correlation with the presence of heart disease, suggesting they are potentially important risk indicators.

2. Maximum Heart Rate (MaxHR) had a strong negative correlation with heart disease, indicating that individuals with lower MaxHR were more likely to be diagnosed.

3. Oldpeak emerged as the strongest positive predictor of heart disease in the dataset.

4. Some surprising results were found, such as a negative correlation between cholesterol and heart disease, which may be due to data imbalance, medication usage, or lifestyle variations not captured in the dataset.

5. The relationship between Age and MaxHR confirms known medical insights, as age increases, the maximum heart rate typically decreases.

While this analysis offers meaningful trends, it's important to note that:

- Correlation does not imply causation

- The dataset might not represent all age groups or clinical variations equally

- Further analysis using statistical models or machine learning could provide more predictive insights

- Turning numeric insights into real-world implications

Author

Shijin Ramesh

LinkedIn: https://www.linkedin.com/in/shijin-ramesh-268b051a

Date: 26th June, 2025