Project - Phase 1

Visual Insights into Heart Disease:Exploring Risk Factors and Outcomes

Team Members (Name with Roll.No.):
1. Lakshya Babel (22BDS033)
2. Rachit Gupta (22BDS047)
3. Abhijit Singh (22BDS054)
4. Suryansh Ayush (22BDS057)

Github Repo Link: https://github.com/invinciblecoder9/Data_Visualisation
StreamLit Link: https://heartdisease-dashboard.streamlit.app/

Overview of the Project

Project Description :

1. Dataset : Heart disease dataset
2. Source of the Dataset : https://www.kaggle.com/datasets/sid321axn/heart-statlog-cleveland-hungary-final/data
3. Brief Description of Dataset : Dataset has 15 features and 920 observations
4. Tools / Frameworks for interactive app/dashboard development : Voila/Streamlit ,Power BI

Overview of the Project

- The dataset is the Cleveland Heart Disease dataset taken from the UCI repository. The dataset consists of approx 2000 individuals' data. There are 14 columns in the dataset.

- This database contains 13 attributes and a target variable. It has 8 nominal values and 5 numeric values. The detailed description of all these features are as follows:
- Age, Sex, Chest pain, resting BP, cholestrol, fasting blood sugar, resting ECG, max heart rate, exercise angina, old peak, ST slope, target.

Overview of the Project

- This project aims to visualize various scenarios in which individuals may be at risk for heart disease, using a heart disease dataset. By analyzing key factors such as age, cholesterol, blood pressure, and lifestyle choices, the project will uncover patterns and relationships through visualizations like scatter plots, histograms, and heatmaps. The goal is to provide clear insights into heart disease risk factors, supporting better understanding and prevention efforts.

Overview of the Project

- Tools and Technologies used for the visualisations are: Jupyter notebook, matplotlib, seaborn, pandas.

Progress Achieved (Final Phase)

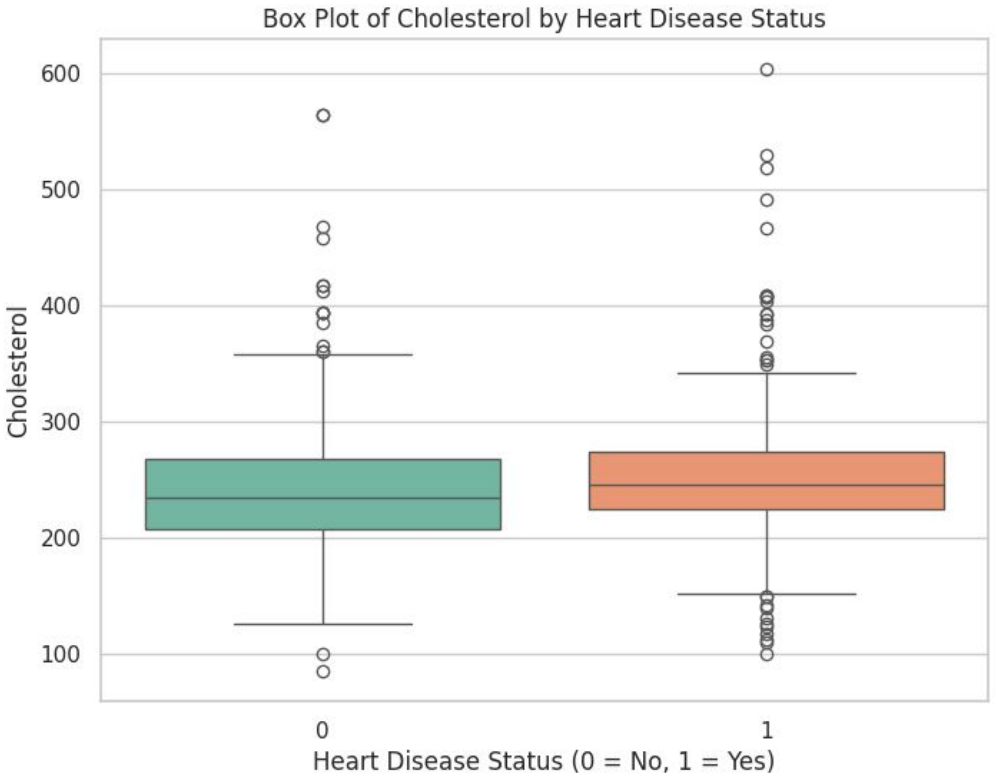
- The dataset contained many missing and duplicate values, which we addressed through data pre-processing. After cleaning the dataset, we discovered several important insights.

- Challenges faced and solutions implemented.

1. Handling values with missing data.
2. Did preprocessing and Eda to get some valuable insights.
3. It was hard and time-consuming to select proper visualisation.

Progress Achieved (Final Phase)

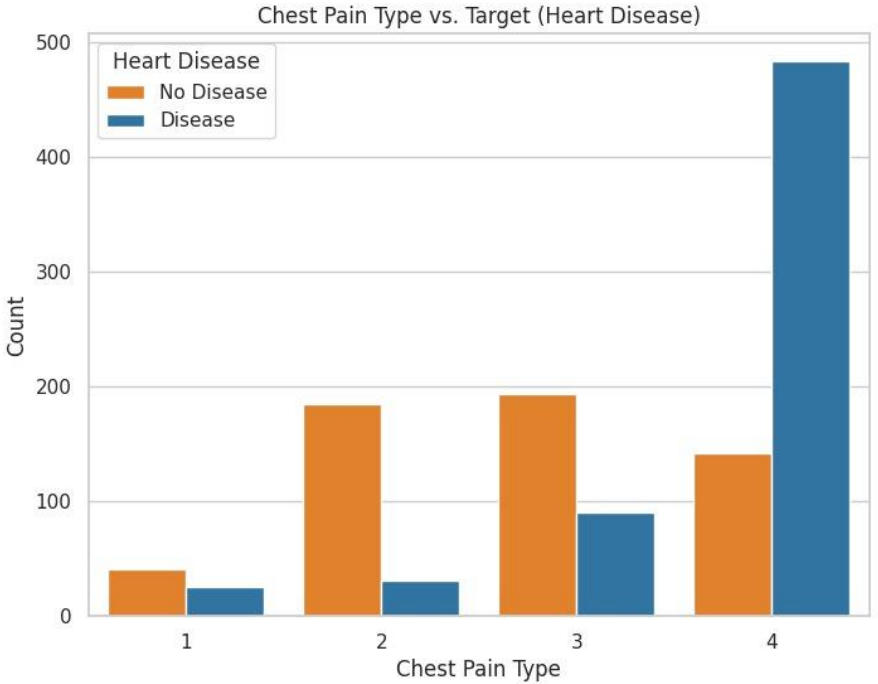
- The box plot represents the distribution of data among cholesterol. Also, it tells us about the outliers present in the data features. It helps us to prevent overfitting of the model by eliminating the outliers or extreme values.



Progress Achieved (Final phase)

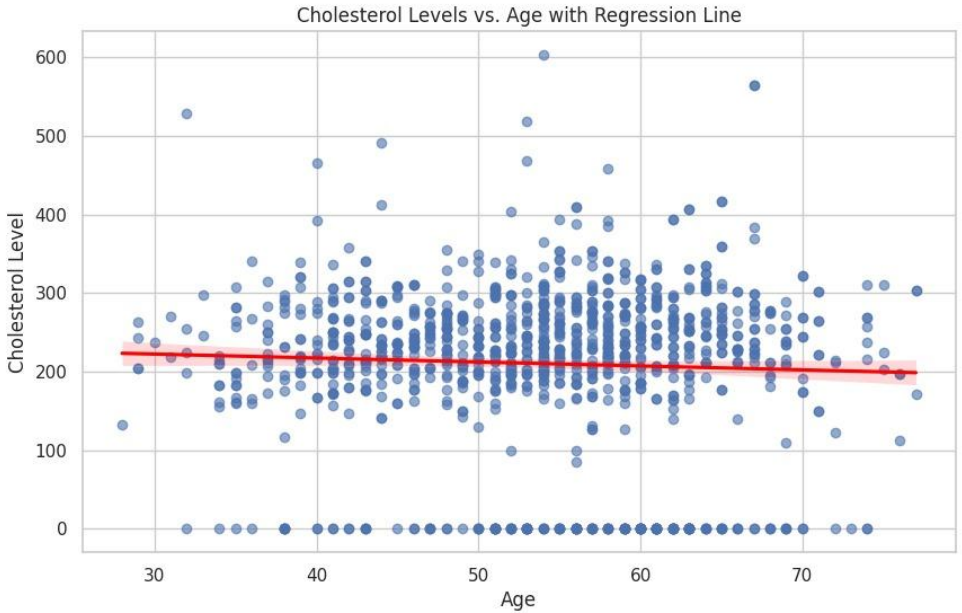
- **Association of Chest Pain Type with Heart Disease**: The chart shows a clear relationship between chest pain types and the presence of heart disease. Individuals with chest pain type 4 (asymptomatic) are significantly more likely to have heart disease, while other chest pain types are less associated with it.
- **Distinct Patterns in Chest Pain Distribution**:

  Most individuals without heart disease report chest pain types 2 or 3, while chest pain type 4 is predominantly observed in those with heart disease. This highlights chest pain type as a potentially important indicator in heart disease diagnosis.



Progress Achieved (Final phase)

- **Weak Negative Correlation between Age and Cholesterol**: The scatter plot with a regression line indicates a slight decline in cholesterol levels as age increases, suggesting that cholesterol levels are not strongly age-dependent in this dataset.
- **Wide Variability in Cholesterol Levels**:
  Cholesterol levels vary widely across all ages, ranging from below 100 to above 500. This variability indicates that factors other than age may have a more significant impact on cholesterol levels in the population.



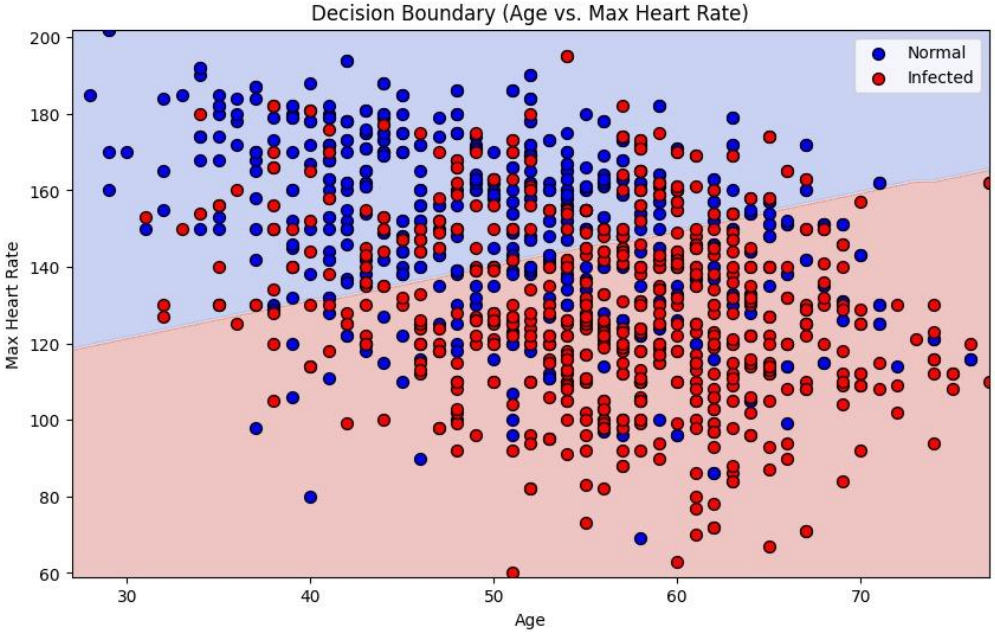
Progress Achieved (Final phase)

Point 1: Decision Boundary for Heart Disease Prediction

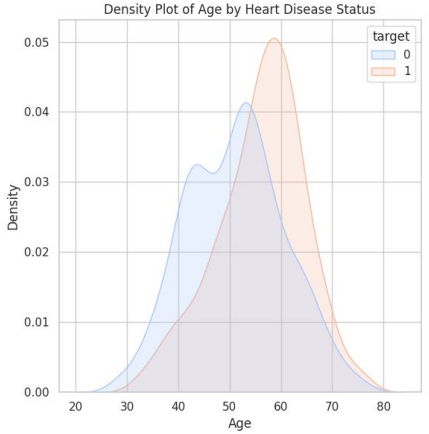
- The graph illustrates a decision boundary that separates individuals with normal heart health from those with heart disease based on their age and maximum heart rate.
- The blue region represents individuals classified as "Normal," while the red region indicates those classified as "Infected" (or having heart disease).

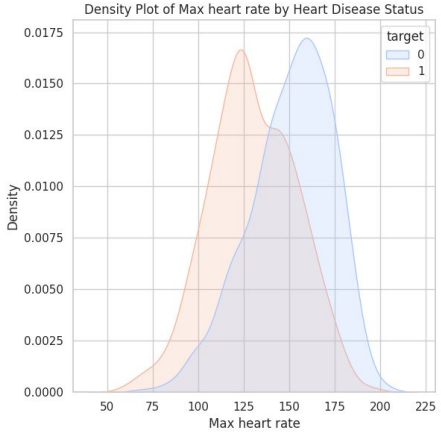
Point 2: Age and Max Heart Rate as Key Predictors

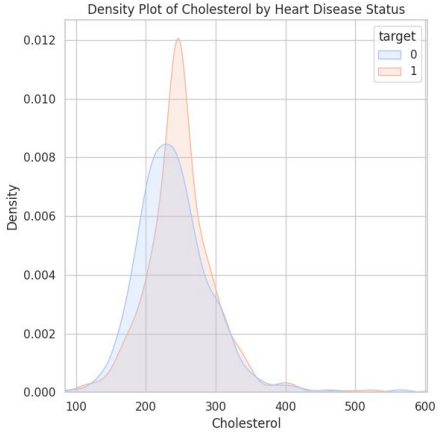
- The graph highlights that age and maximum heart rate are significant predictors of heart disease risk.
- The decision boundary suggests that individuals with lower maximum heart rates, particularly at older ages, are more likely to be classified as having heart disease
-

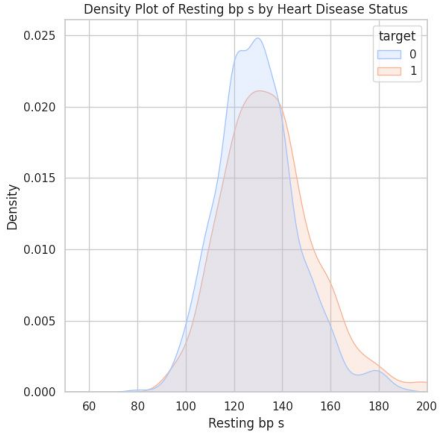


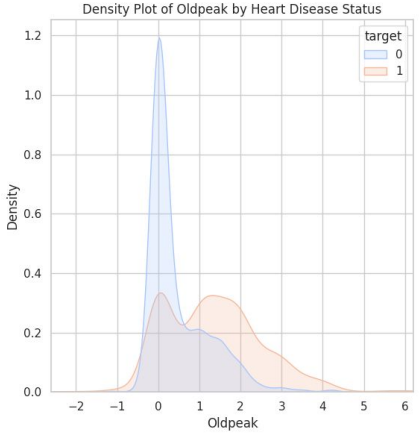
Density Plot











Progress Achieved (Final phase)

Point 1: Key Predictors for Heart Disease

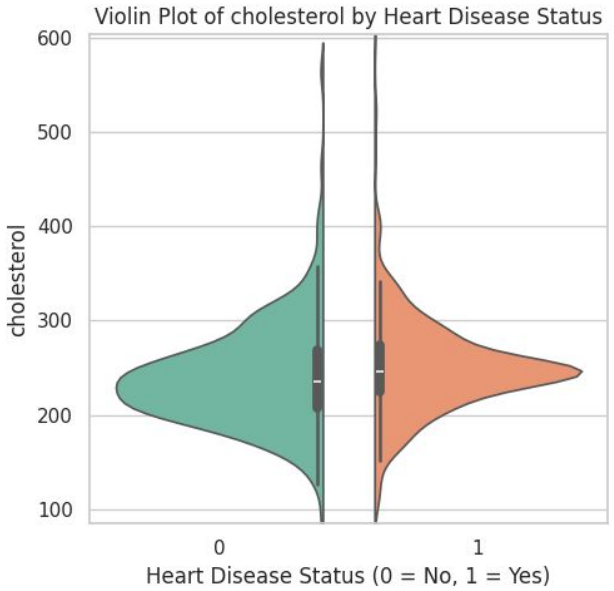
- The density plots reveal several key predictors of heart disease risk:
  - **Age:** Older individuals have a higher likelihood of heart disease.
  - **Maximum Heart Rate:** Lower maximum heart rates are associated with increased risk.
  - **Cholesterol:** Higher cholesterol levels are linked to higher risk.
  - **Resting BP:** Elevated resting blood pressure increases the risk.
  - **Oldpeak:** A higher value for ST depression induced by exercise relative to rest also indicates higher risk.

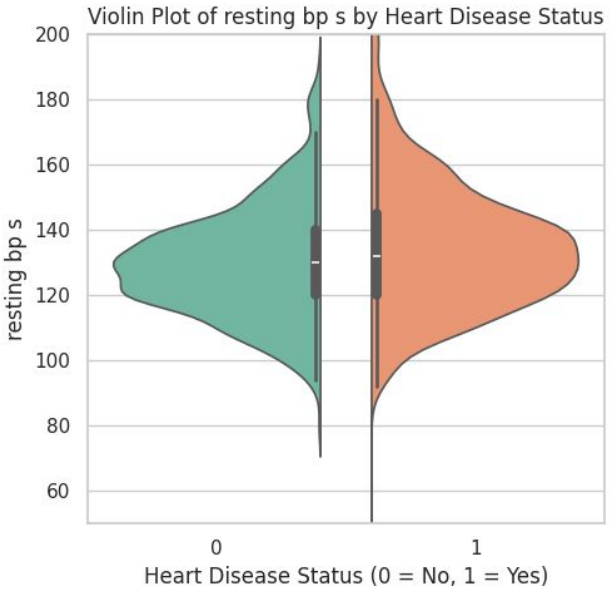
Point 2: Distribution of Predictors Across Heart Disease Status

- The plots show how the distribution of these predictors differs between individuals with and without heart disease.
- For example, the distribution of age is shifted to the right for individuals with heart disease, indicating a higher average age in this group.
- Similarly, the distributions of maximum heart rate, cholesterol, resting BP, and Oldpeak are shifted to the right for individuals with heart disease, suggesting higher values of these predictors are associated with increased risk.

Progress Achieved (Final phase)

Violin Plot





Progress Achieved (Final phase)

Point 1: Key Predictors for Heart Disease

- The violin plots reveal several key predictors of heart disease risk:
  - **Age:** Older individuals have a higher likelihood of heart disease.
  - **Cholesterol:** Higher cholesterol levels are linked to higher risk.
  - **Resting BP:** Elevated resting blood pressure increases the risk.

Point 2: Distribution of Predictors Across Heart Disease Status

- The plots show how the distribution of these predictors differs between individuals with and without heart disease.
- For example, the distribution of age is shifted to the right for individuals with heart disease, indicating a higher average age in this group.
- Similarly, the distributions of cholesterol and resting BP are shifted to the right for individuals with heart disease, suggesting higher values of these predictors are associated with increased risk.

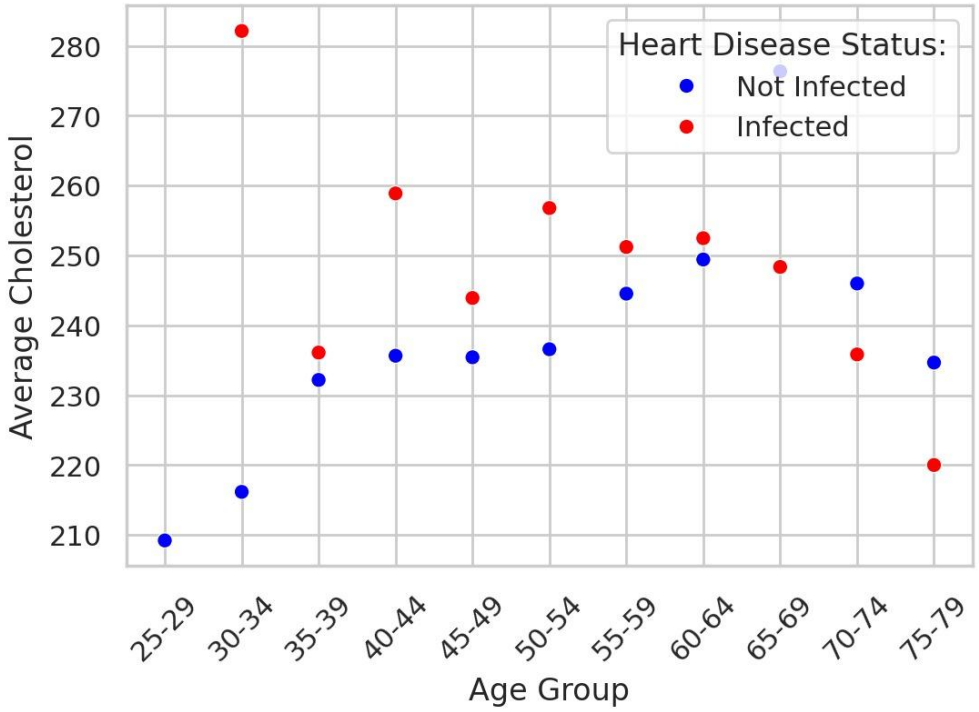
Progress Achieved (Final phase)

Point 1: Max Heart Rate Trends Across Age Groups

- The graph shows the average maximum heart rate for individuals with and without heart disease across different age groups.
- Both groups exhibit a general downward trend in maximum heart rate with increasing age, as expected due to physiological aging.

Point 2: Heart Disease Impact on Max Heart Rate

- Individuals with heart disease (red line) tend to have lower maximum heart rates compared to those without heart disease (blue line) across all age groups.
- This difference becomes more pronounced in older age groups, suggesting that heart disease may further impair cardiovascular function and limit exercise capacity in older individuals.



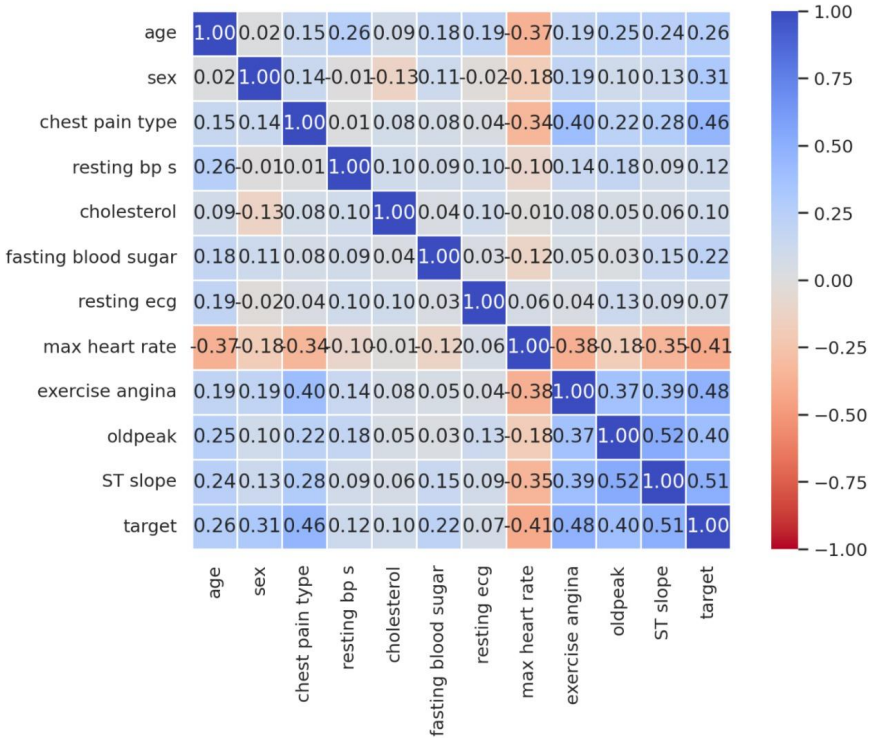
Progress Achieved (Final phase)

Point 1: Key Predictors and Their Interactions

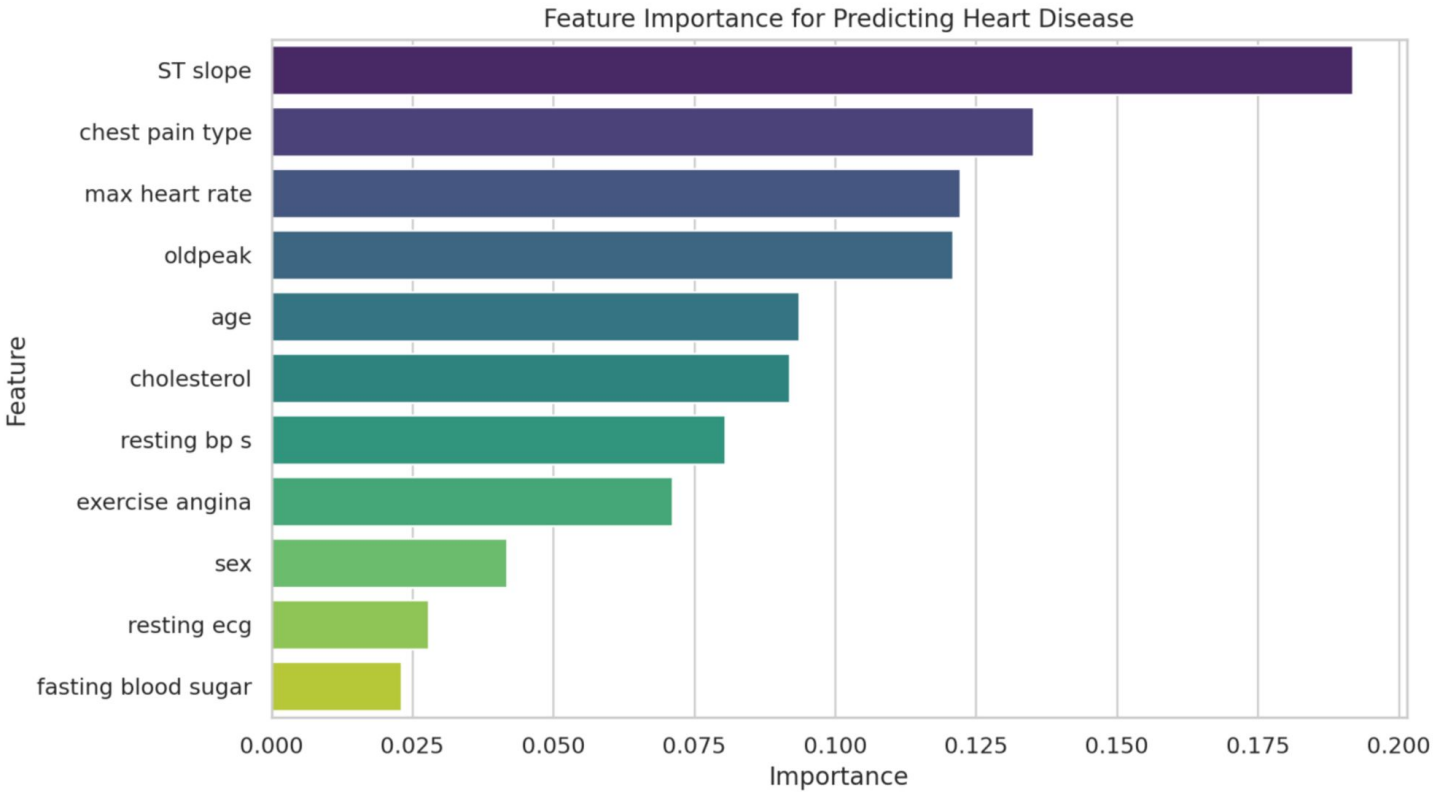
- The heatmap reveals several key predictors of heart disease risk, including:
- **Age:** Older individuals have a higher likelihood of heart disease.
- **Cholesterol:** Higher cholesterol levels are linked to higher risk.
- **Resting BP:** Elevated resting blood pressure increases the risk.
- **Max Heart Rate:** Lower maximum heart rates are associated with increased risk.
- **Exercise Angina:** Presence of exercise-induced angina is a strong predictor.
- **ST Slope:** The slope of the ST segment during exercise is also a significant factor.

Point 2: Feature Interactions and Complexity

- The heatmap also highlights the complex interplay between different predictors.
- For example, the interaction between `max heart rate` and `exercise angina` is particularly strong, suggesting that the combination of these two factors is a powerful predictor of heart disease.
- Understanding these feature interactions can help build more accurate and robust predictive models.



Progress Achieved (Final phase)



Thank You !!!