

# Advanced heart disease prediction model

Presented by ISHU

# INTRODUCTION

This project delved deep into the heart disease diagnostic dataset, leveraging advanced data science techniques to unravel the intricate relationships between various health metrics and heart disease. Through a comprehensive analysis process, we have not only identified key risk factors but also laid the groundwork for predictive modeling that could transform how we approach heart disease prevention and management.

Health is real wealth, and during the pandemic, we all realized the profound impact of COVID-19 on everyone, regardless of status. This project focuses on analyzing heart disease diagnostic data to better prepare for future health challenges. By leveraging data-driven insights, we aim to understand the patterns and risk factors associated with heart disease, ultimately contributing to better health outcomes.

# Overview of Heart Disease Diagnostic Data

## Key Insights and Findings:

01.

**Age and Cholesterol as Major Predictors:** The analysis, particularly through the correlation matrix and interactive visualizations, confirmed that age and cholesterol levels are critical predictors of heart disease. Older individuals with high cholesterol levels are significantly more likely to be at risk.

02.

**Chest Pain Type (cp) as a Strong Indicator:** The chest pain type, classified in the dataset, emerged as a significant feature. The RandomForestClassifier model, which was trained using the preprocessed and feature-engineered data, highlighted chest pain type as one of the top predictors, as seen in the feature importance plot.

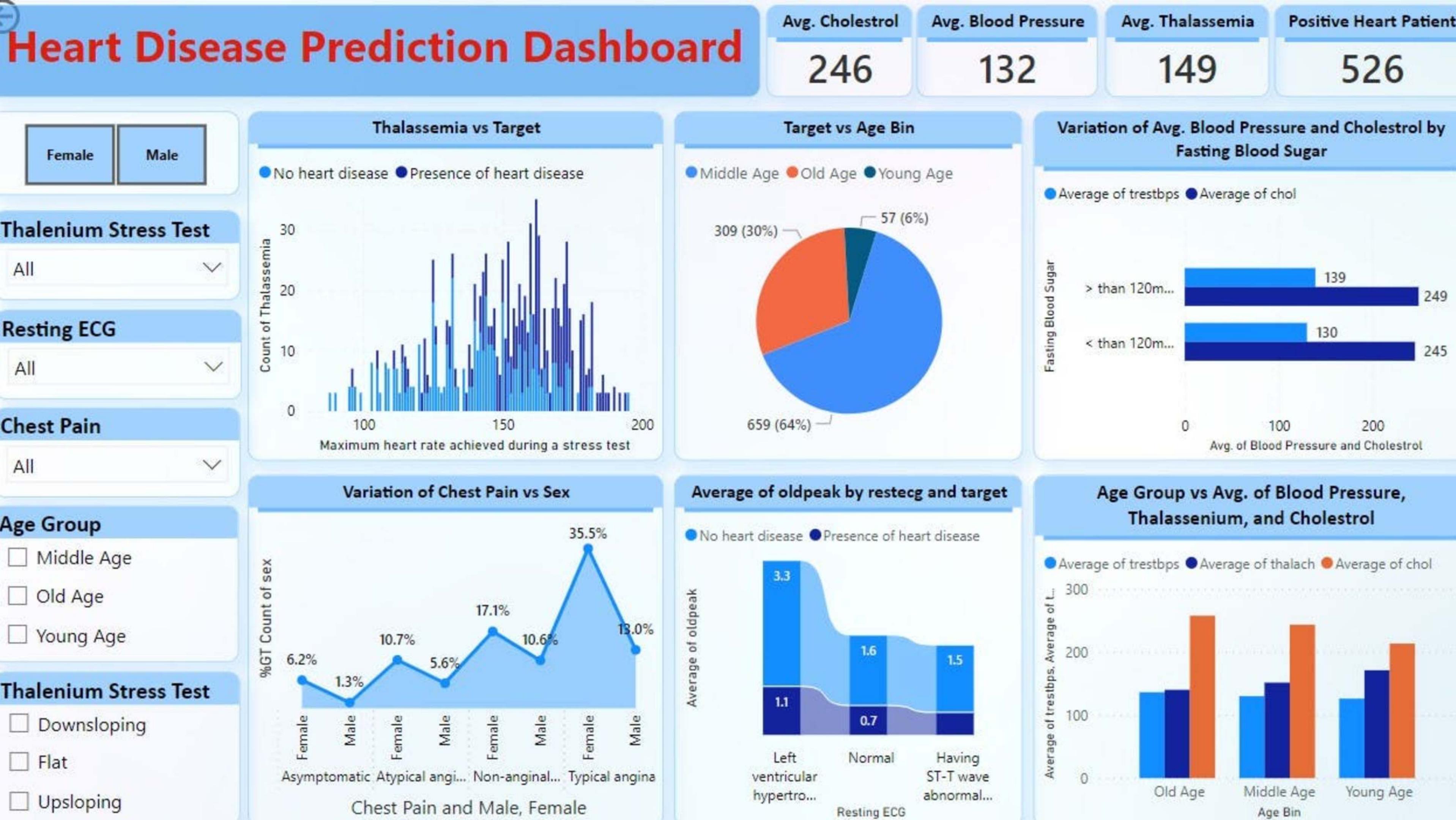
03.

**Max Heart Rate (thalach) as a Crucial Factor:** The violin plot and other visualizations indicated that maximum heart rate (thalach) is closely associated with the presence of heart disease. The KDE plot further reinforced that patients with heart disease tend to have a lower maximum heart rate.

# Leveraging the Power of Data:

- ETL Process: The ETL (Extract, Transform, Load) process was meticulously executed, where raw data was cleansed, transformed with new features such as age groups and cholesterol levels, and then loaded into a final, scalable DataFrame. This clean and enriched dataset was pivotal in driving meaningful insights and accurate model predictions.
- Exploratory Data Analysis (EDA): The EDA, powered by visual tools like seaborn and plotly, provided a rich understanding of the data distribution, relationships, and outliers. The correlation matrix, pairwise plots, and interactive dashboards were instrumental in identifying patterns that could otherwise be overlooked.
- Advanced Modeling Techniques: The use of polynomial features and feature selection through a RandomForestClassifier allowed us to capture non-linear relationships and distill the most critical predictors of heart disease. The model's ability to weigh these features accurately contributed to a robust predictive framework.





# Future Research Directions:

## Integrating Lifestyle and Behavioral Data:

Future studies could enrich the model by integrating lifestyle factors such as diet, physical activity, and stress levels, which were not included in the current dataset but are known to influence heart disease.



## Refining Predictive Models:

Further refinement of predictive models using ensemble methods or deep learning techniques could enhance accuracy and generalizability. For instance, integrating a GradientBoostingClassifier or a neural network could be explored to capture more complex patterns.

## Longitudinal Analysis:

Conducting longitudinal studies could provide insights into the progression of heart disease over time, allowing for early intervention strategies. Tracking patients' health metrics over several years could help in understanding how the risk evolves.

# Final Thought:

This project is a testament to the power of data in transforming healthcare. By systematically analyzing and modeling health data, we can identify at-risk populations, personalize treatment plans, and ultimately, save lives. The journey from raw data to actionable insights was guided by robust coding practices and advanced analytical techniques, as reflected in the provided Python code.

## Deliverables:

- 1).Python Code: Clean, modular, and well-documented code, ensuring reproducibility and scalability.
- 2).Power BI Dashboard: An interactive, user-friendly dashboard that highlights key findings and allows stakeholders to explore the data intuitively.

**Thank you  
very much!**