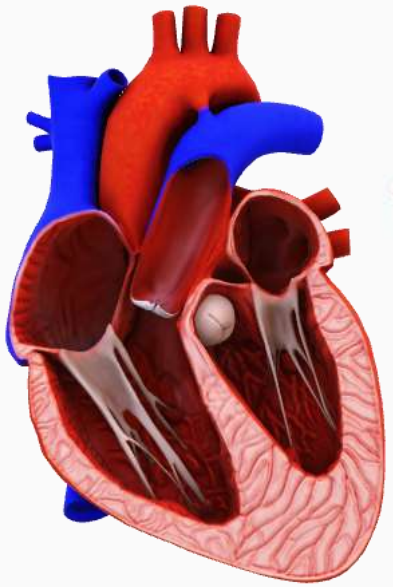
WELCOME TO

HEART DISEASE DIAGNOSTIC ANALYSIS PROJECT WITH PYTHON AND MACHINE LEARNING



PROJECT DETAILS

Title	Heart Disease Diagnostic Analysis
Technologies	Data Science
Domain	Healthcare & Cardiovascular Health
Project Difficulties Level	Intermediate
Programming Language Used	Python & Machine Learning
Other Tools	Jupyter Notebook, MS Power Point

Introduction

THE HEART IS A KIND OF MUSCULAR ORGAN WHICH PUMPS BLOOD INTO THE BODY AND IS THE CENTRAL PART OF THE BODY'S CARDIO VASCULAR SYSTEM WHICH ALSO CONTAINS LUNGS. CARDIO VASCULAR SYSTEM ALSO COMPRISES A NETWORK OF BLOOD VESSELS, FOR EXAMPLE, VEINS, ARTERIES, AND CAPILLARIES. THESE BLOOD VESSELS DELIVER BLOOD ALL OVER THE BODY. ABNORMALITIES IN NORMAL BLOOD FLOW FROM THE HEART CAUSE SEVERAL TYPES OF HEART DISEASES WHICH ARE COMMONLY KNOWN AS CARDIOVASCULAR DISEASES (CVD). HEART DISEASES ARE THE MAIN REASONS FOR DEATH WORLDWIDE.

ACCORDING TO THE SURVEY OF THE WORLD HEALTH ORGANIZATION (WHO), 17.5 MILLION TOTAL GLOBAL DEATHS OCCUR BECAUSE OF HEART ATTACKS AND STROKES. MORE THAN 75% OF DEATHS FROM CARDIO VASCULAR DISEASES OCCUR MOSTLY IN MIDDLE-INCOME AND LOW-INCOME COUNTRIES. ALSO, 80% OF THE DEATHS THAT OCCUR DUE TO CVDS ARE BECAUSE OF STROKE AND HEART ATTACK . THEREFORE, PREDICTION OF CARDIAC ABNORMALITIES AT THE EARLY STAGE AND TOOLS FOR THE PREDICTION OF HEART DISEASES CAN SAVE A LOT OF LIFE AND HELP DOCTORS TO DESIGN AN EFFECTIVE TREATMENT PLAN WHICH ULTIMATELY REDUCES THE MORTALITY RATE DUE TO CARDIOVASCULAR DISEASES.

Abstract

HEART DISEASE IS THE MAJOR CAUSE OF DEATHS WORLDWIDE. TO GIVE TREATMENT FOR HEART DISEASE, A LOT OF ADVANCED TECHNOLOGIES ARE USED. IN MEDICAL CENTER IT IS THE MOST COMMON PROBLEM THAT MANY OF MEDICAL PERSONS DO NOT HAVE EQUAL KNOWLEDGE AND EXPERTISE TO TREAT THEIR PATIENT SO THEY DEDUCE THEIR OWN DECISION AND AS A RESULT IT SHOW POOR OUTCOME AND SOMETIME LEADS TO DEATH.

TO OVERCOME THESE PROBLEMS PREDICTIONS OF HEART DISEASE USING MACHINE LEARNING ALGORITHMS AND DATA MINING TECHNIQUES, IT BECOME EASY TO AUTOMATIC DIAGNOSIS IN HOSPITALS AS THEY ARE PLAYING VITAL ROLE IN THIS REGARD. HEART DISEASE CAN BE PREDICTED BY PERFORMING ANALYSIS ON PATIENT'S DIFFERENT HEALTH PARAMETERS

Objective

The objective is to facilitate research and analysis aimed at developing predictive models for the detection and assessment of heart disease. Specially aims to;

- **Enable Prediction** : Provide a diverse set of medical attributes and corresponding diagnosis to enable development of machine learning models capable of predicting the likelihood of heart disease in patients.

- **Promote Healthcare Innovation** : Promote innovation in healthcare by empowering healthcare providers, businesses and policymakers with data-driven insights into heart disease risk assessment and management.

- **Improve Patient Outcomes** : Ultimately, the primary objective is to contribute to the improvement of patient outcomes by facilitating early detection, intervention and personalized treatment of heart disease.

Required Packages And Libraries

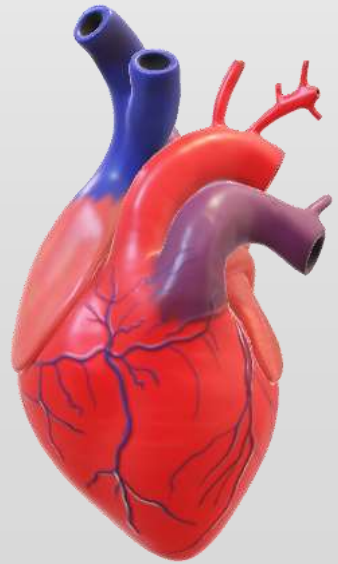
Data Preparation Process

Data Collection

- Gathering the data from the required source
- Loading the dataset

Data Preprocessing

- Checking for Missing Values
- Checking for Duplicates and dropping them



Data Understanding

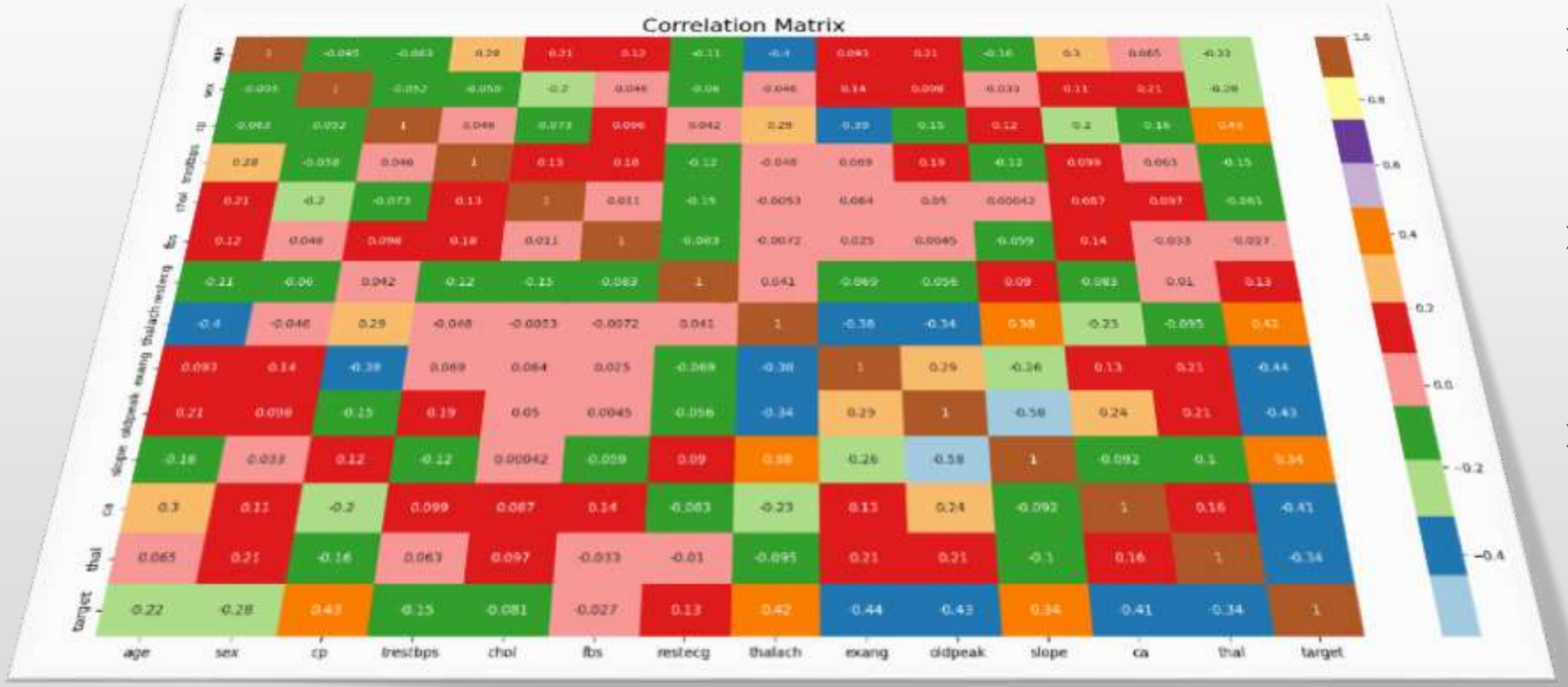
Dataset Features Description

- There are 14 features in Dataset
- **Age** : The person's age in years
- **Sex** : The person's gender (1 = Male, 0 = Female)
- **cp** : Chest Pain Type (It has 4 values) Value 0 - typical angina Value 1 - atypical angina Value 2 - non-anginal pain Value 3 - asymptomatic
- **trestbps** : The person's resting blood pressure (mm Hg on admission to the hospital)
- **chol** : The person's serum cholesterol measurement in mg/dl
- **fbs** : The person's fasting blood sugar (> 120 mg/dl, 1 = True, 0 = False)
- **restecg** : Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)
- **thalach** : The person's maximum heart rate achieved
- **exang** : Exercise induced angina (1 = yes; 0 = no)
- **oldpeak** : ST depression induced by exercise relative to rest
- **slope** : the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)
- **ca**: The number of major vessels (0- colored by fluoroscopy
- **thal** : A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
- **target** : Heart Disease (0 = Less chance of Heart Attack) (1 = More chance of Heart Attack)

Data Visualization

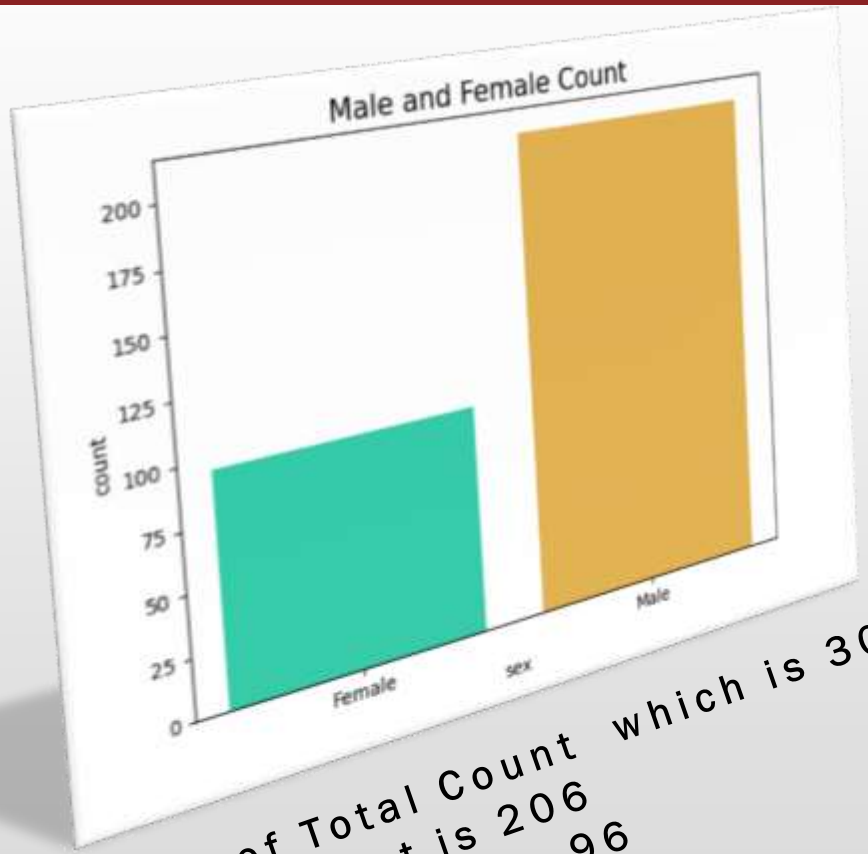
Analyzing The Dataset

Correlation of Features with Target column



- cp, thalach and slope show good positive correlation with target column
- oldpeak, exang, ca, thal, age, sex show negative correlation with target column
- fbs, chol, trestbps show very low correlation with target column

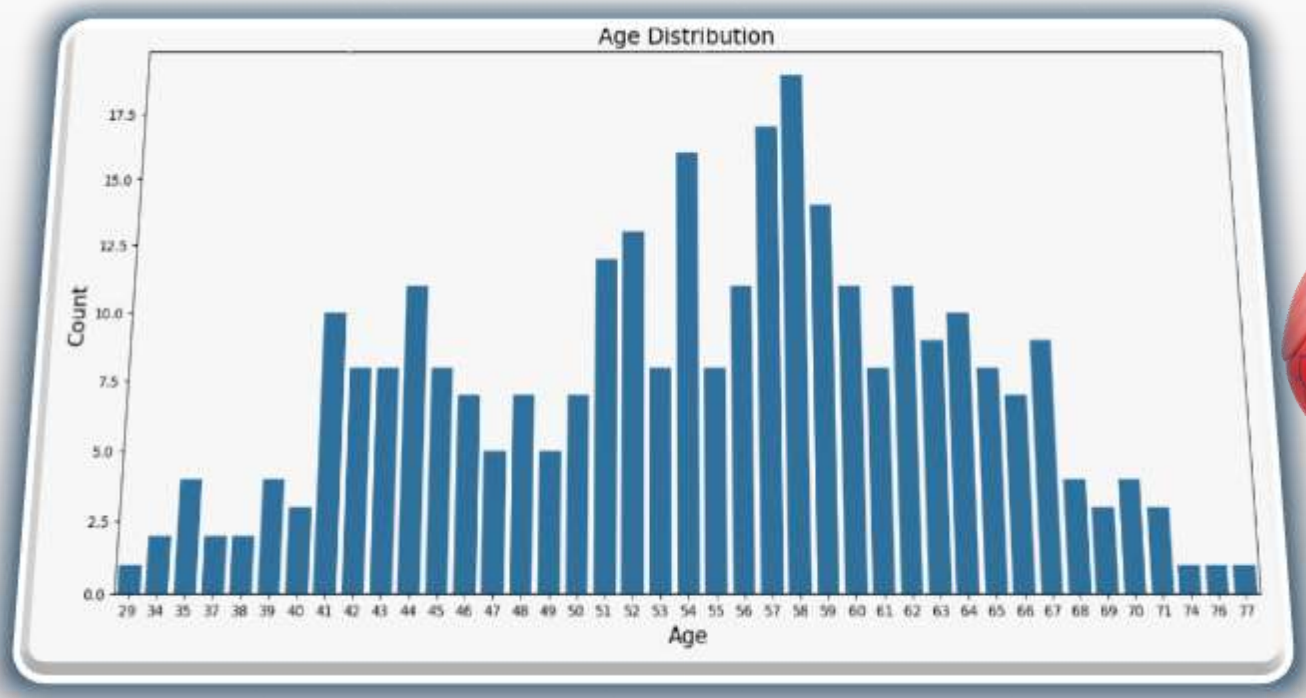
Gender Distribution

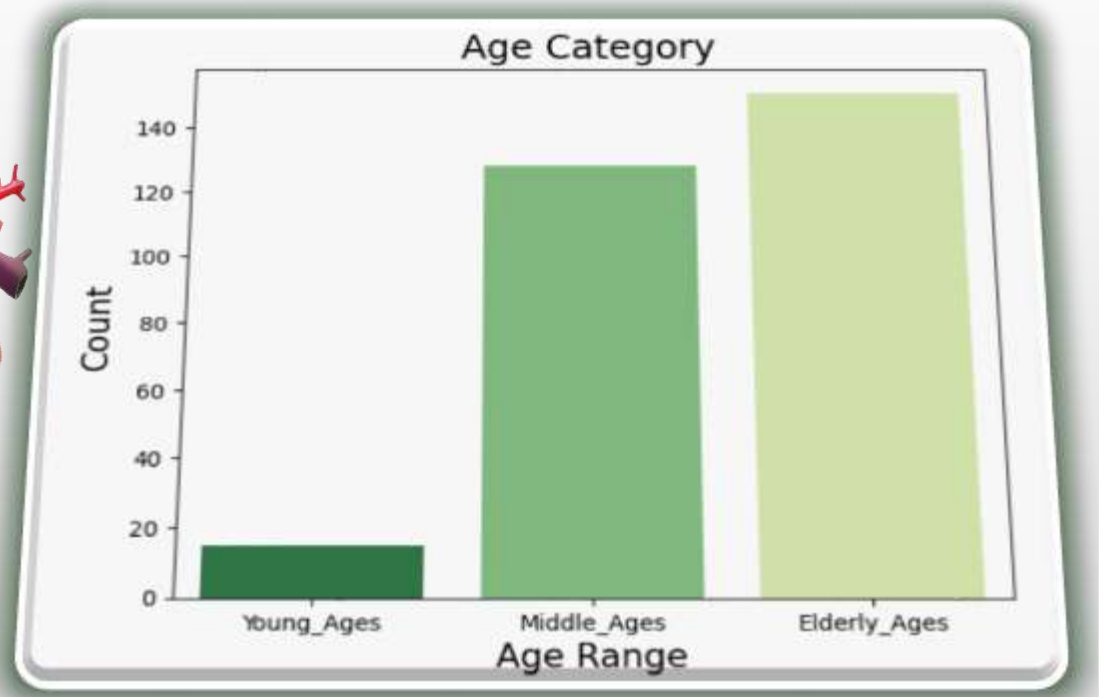


- Out of Total Count which is 303
- Male count is 206
- Female count is 96

- From the above chart we can see that Male's are more prone to Heart disease than Female's

Age Distribution

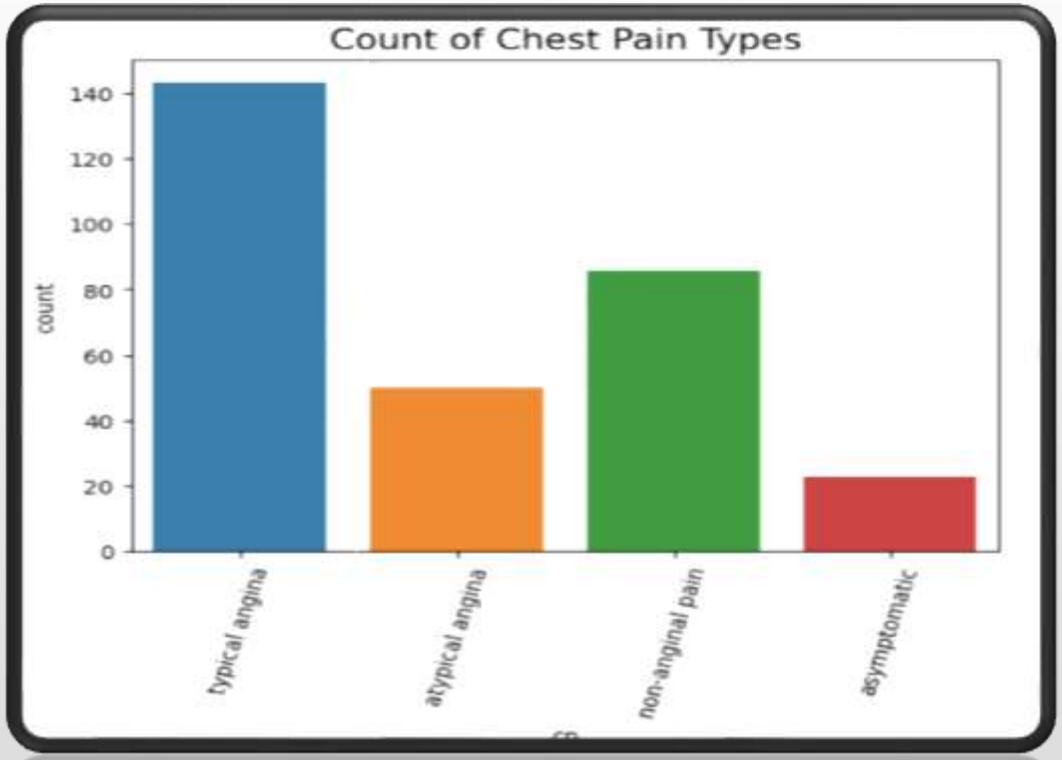


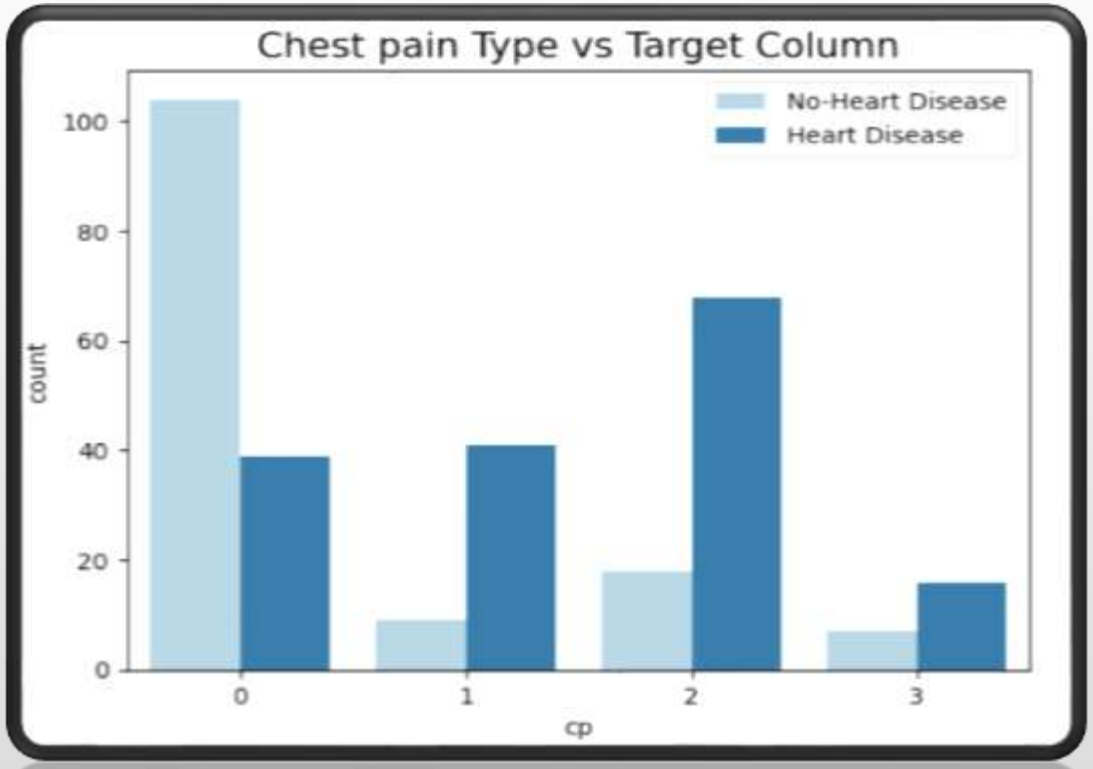


- Statistical Analysis of Age Distribution :
- Minimum Age is 29
- Middle Age is 55
- Maximum Age is 77

- Young Ages (29-40)yr , count is 15
- Middle Age (41-55)yr , count is 128
- Elderly Age (55-77)yr , count is 151

Chest Pain Type Distribution

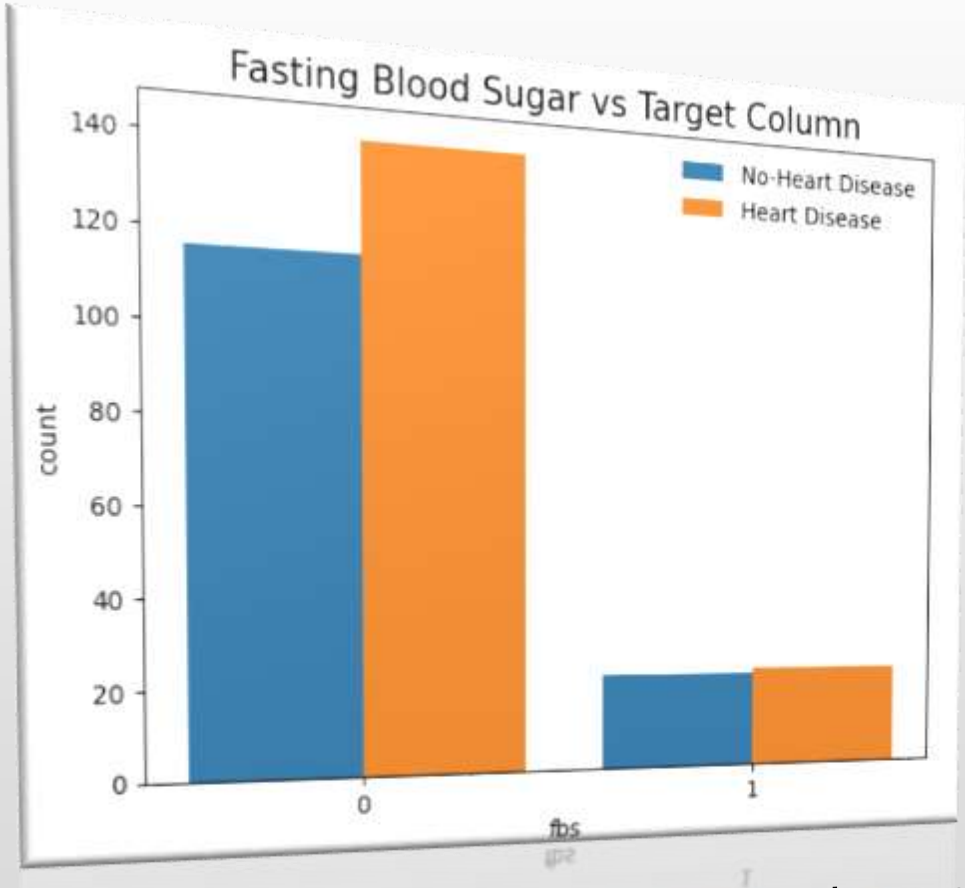


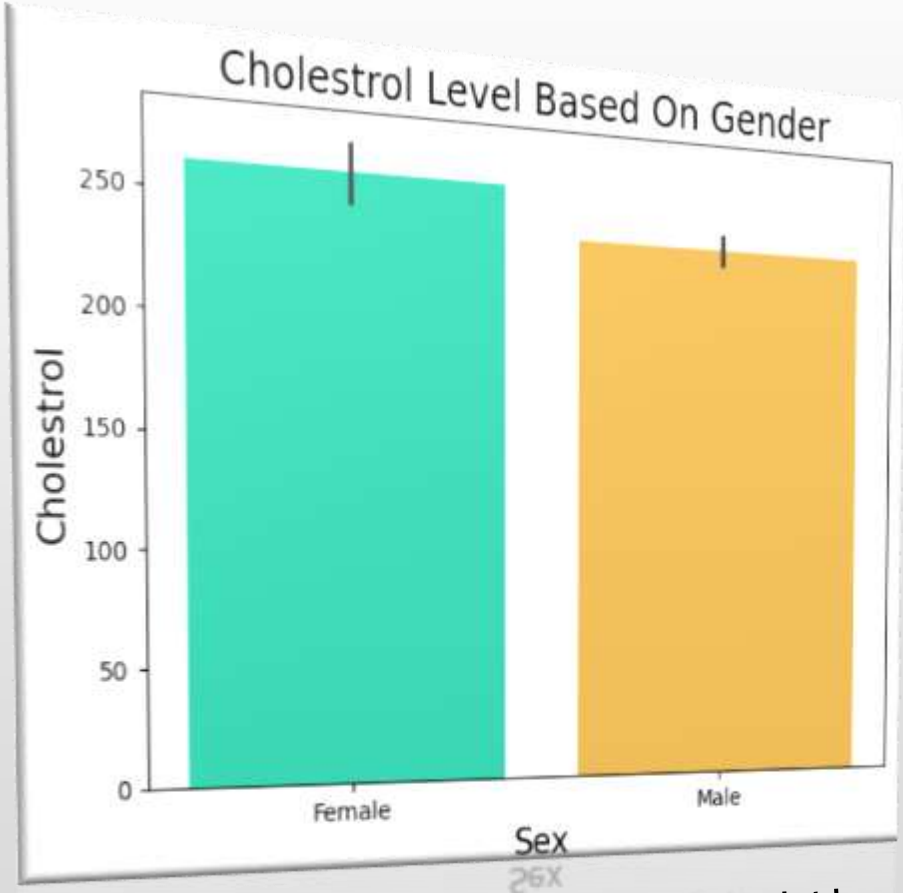


- Typical Angina Chest Pain Type is most common among people in this dataset

- From this chart we can see healthy people are more affected by Chest Pain. Chest Pain can be subjective due to stress, physical activities & many more & also it varies between Gender

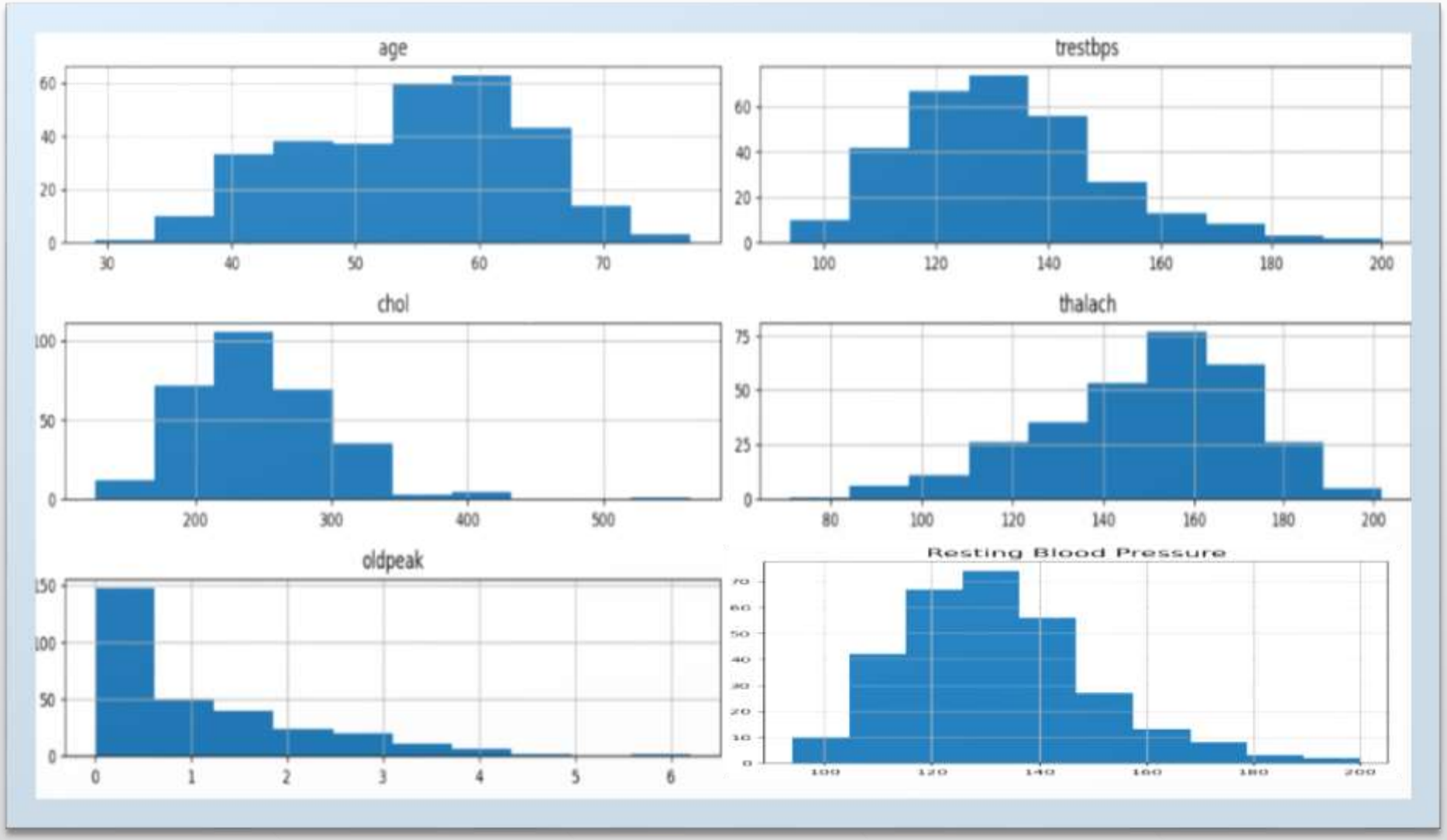
Analyzing Fasting Blood Sugar & Cholesterol Level

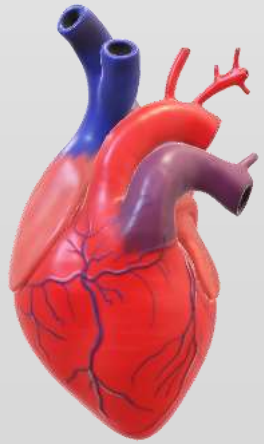




- It can seen that Non – diabetic are more affected with Heart diseases
- Cholesterol Levels are slightly higher in Female's than Male

Visualizing the Continuous Variables



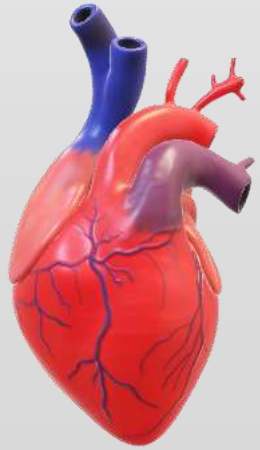


Process of Model Training and Model Evaluation

Step 1	
Defining input and output i.e. x and y	Splitting the dataset into Training and Testing

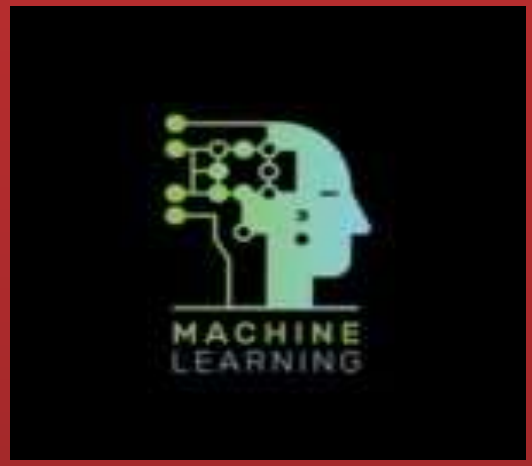
Step 2	
Model Training	Training model with Logistic Regression

Step 3	
Model Evaluation	Evaluating the model performance by checking accuracy on training data and testing data



```
LogisticRegression
LogisticRegression()
```

- Logistic regression is a statistical algorithm which analyze the relationship between two data factors.
- Logistic regression is used for binary classification where we use sigmoid function, that takes input as independent variables and produces a probability value between 0 and 1.
- With Model Evaluation we can see that we can build a Logistic Regression Model for Predictions of Heart Attack based on training given to the model as it gives accuracy of 86% on training dataset i.e. model classifies 86 datapoints correctly and 14 are misclassified datapoints. For testing dataset it gives accuracy of 82% i.e. model can classify 82 datapoints correctly and misclassifies 18 datapoints. As compared to other algorithms of classification Logistic Regression works well for the given dataset and there is no overfitting.
- Heart Disease dataset provides all useful information required to build a model that will give better outcomes and helps health professionals in predicting heart disease efficiently and effectively.



Machine Learning Algorithm

Logistic Regression

Real - life Applications

- **Clinical Decision Support** – Healthcare professionals can use these models as decision support tools during patient consultations. By inputting patient data into the model, clinicians can obtain risk scores and recommendations for further evaluation or treatment.

- **Public Health Initiatives** – Public health authorities can utilize predictive models to identify populations at high risk of heart disease and implement targeted prevention strategies, such as educational campaigns, screening programs or policy interventions.

- **Remote Monitoring** – Remote monitoring devices equipped with heart disease prediction algorithms can continuously monitor individuals at risk and alert them or their caregivers of any significant changes or warning signs, enabling timely medical intervention.

- **Personalized Medicine** – Predictive models can facilitate the shift towards personalized medicine by enabling healthcare providers to tailor treatment plans based on an individual's risk profile and genetic predisposition to heart disease.

Conclusion

This project provides the deep insight into machine learning techniques for classification of heart diseases. The role of classifier is crucial in healthcare industry so that the results can be used for predicting the treatment which can be provided to patients. The existing techniques are studied and compared for building a model in detecting efficiently and in effective accurate systems. Machine learning techniques significantly improves accuracy of cardiovascular risk prediction through which patients can be identified during an early stage of disease and can be benefitted by preventive treatment. It can be concluded that there is a huge scope for machine learning algorithms. With this Logistic regression model many applications can be build that will help in predicting cardiovascular diseases and can prevent life threats.

Future Scope of this Project

- **Healthcare Providers** : Hospitals and clinics can use these models to assess the risk of heart disease in patients during routine check –ups. This can lead to early detection and intervention, ultimately improving patients outcomes and reducing healthcare costs.

- **Insurance Companies** : Insurance Companies can utilize these models to assess the risk of heart disease in their policyholders. By identifying high –risk individuals , they can offer targeted interventions or wellness programs to mitigate the risk and reduce claims.

- **Pharmaceutical Companies** : Pharmaceutical companies can use predictive models to identify potential candidates for clinical trials of new drugs aimed at preventing or treating heart disease. This can streamline the drug development process and bring new treatments to market more efficiently.

- **Health tech Startups** : Health Tech Startups focus on digital health and wellness can develop applications or wearable devices that utilize heart disease prediction models to provide personalized health recommendations to users. This can empower individuals to take proactive steps toward preventing heart disease.

thank you

Presentation by
Saniya Shaikh