PROJECT_NAME

- *Cardio_Predict*

TEAM_NAME

- *HeartWise*

TEAM_MEMBERS

- *MUHAHHAM_OWAIS.*
- *ABDUL_HANNAN.*
- *ZAHID_HUSSAIN.*

Project Title:

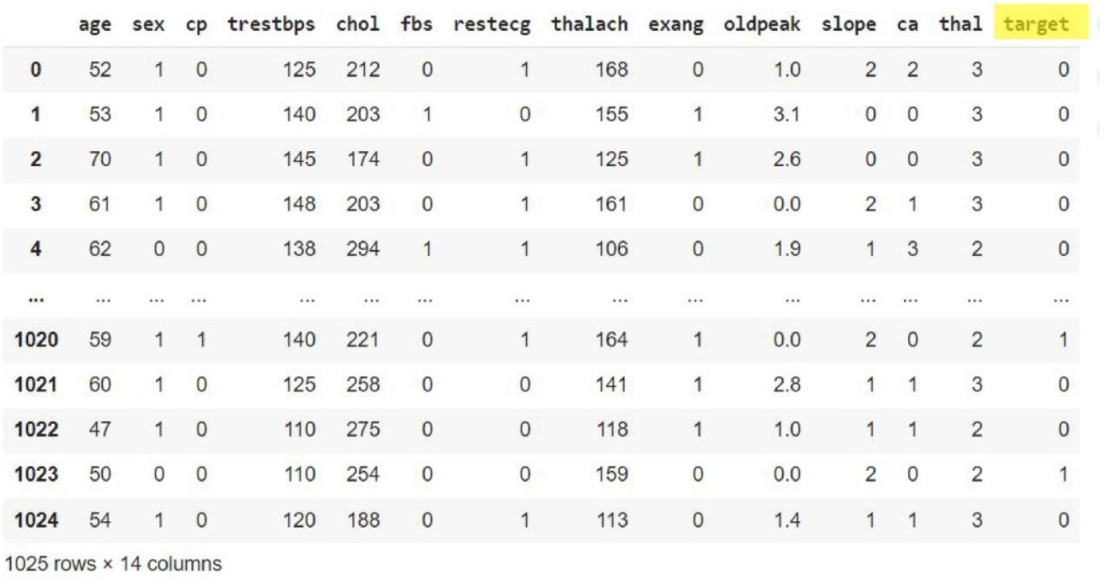
- **Predicting Heart Disease Risk Using Machine Learning Models: A Feature-Based Analysis**

Project Objective:

The goal of this project is to develop a predictive model that accurately classifies individuals as at-risk or not at-risk for heart disease based on clinical and demographic features. The project will involve analyzing the dataset, selecting important features, and training machine learning models to predict the presence of heart disease.

Understanding the Heart Disease Dataset

- **Shape**: `(1025, 14)` means we have **1025 rows**and **14 columns** (features).



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

1025 rows × 14 columns

To determine how we can extract **useful information** from a dataset and how each feature (or column) will help us in making **predictions** or deriving **insights**, we need to consider several factors. Here’s a step-by-step guide on how to analyze a dataset to understand what kind of **information** you can extract and how to assess the **usefulness** of each column:

Understand the Dataset and Its Purpose

•**Features**: The dataset usually consists of multiple features that describe patient demographics, clinical measurements, and health history.

- **Target Variable**: The main outcome is often a binary variable indicating the presence (1) or absence (0) of heart disease.

Types of Columns and Their Usefulness

1. Age (Numerical)

- **Usefulness**: Age is an essential factor in assessing heart disease risk, as the likelihood of heart problems increases with age.

2. Sex (Categorical: 0 = Female, 1 = Male)

Description: The patient's gender.

- Gender can impact the likelihood and type of heart disease. Males tend to have a higher risk of heart disease at a younger age compared to females.

3. cp (chest pain type):

- Different chest pain types can indicate various heart conditions, and this could be a critical predictive feature.

4. trestbps (resting blood pressure)

- High blood pressure is a risk factor for heart disease.

5. chol (serum cholesterol):

- Elevated cholesterol levels are associated with heart disease

6. restecg (resting electrocardiographic results):

- Electrocardiographic abnormalities are indicators of heart issues

7. thalach (maximum heart rate achieved):

- Highly useful. The maximum heart rate can be indicative of cardiovascular fitness and the heart's response to stress.

8. exang (exercise-induced angina):

- **Angina induced by exercise may indicate heart-related issues under stress.**

9. oldpeak (ST depression induced by exercise relative to rest):

- Helps assess abnormal heart stress responses, linked to heart disease risk.

10. slope (slope of the peak exercise ST segment):

- Slope observed in the ST segment can indicate different types of heart disease.

11. ca (number of major vessels colored by fluoroscopy):

- Highly useful. This directly measures the extent of arterial blockage, which is a major factor in heart disease.

12. thal (thalassemia):

- This variable indicates certain blood disorder states, some of which may impact heart health.

13. Target (Outcome)

- **Full Form**: Target Variable (Heart Disease)

- **Description**: The presence (1) or absence (0) of heart disease, which is the outcome

the model aims to predict.

Correlation Analysis:

- Helps determine which features are most closely associated with heart disease (e.g., high correlation between cholesterol and disease).

Visualize Relationships

- Visualizing the relationship between two variables is essential because it provides insights into how one variable may influence or be associated with the other.

Check for Missing Values

- Ensure proper handling of missing data, potentially using median or mode filling strategies depending on the column.

.

Evaluate Columns for Usefulness

Important Features:

1. **Age (Numerical)** Age is a well-known risk factor for heart disease. As people age, their risk of developing heart-related conditions increases.
2. **Sex** Gender likely plays a big role Males are generally at higher risk of heart disease compared to females.
3. **Chest Pain Type** Different types of chest pain are strongly associated with varying levels of heart disease risk. For example, typical angina (chest pain due to reduced blood flow) is a significant indicator of heart disease.

DROP_COLUMN

- **Fasting Blood Sugar (FBS):**

  The **`fbs`** feature may be considered for exclusion or lower priority in heart disease prediction:
  Limited Information
  Weak Correlation with Heart Disease
  Model Performance Impact:

Less Useful Features:

1. **Fasting Blood Sugar (FBS)** While elevated blood sugar can be related to heart disease, it's not as directly predictive as other features like `age`, `chol`, or `ca`. Moreover, it's binary, which limits its ability to capture nuances in blood sugar levels.