# About the dataset:

This dataset is related to the **cardiovascular disease.**

Such data is typically collected through **medical studies**, **healthcare systems**, or **public health surveys**. Specifically, for a dataset like this (the *Cardiovascular Disease Dataset*), the data usually comes from:

1. **Hospitals or Clinics** –  
   Doctors and nurses record patient details (age, height, weight, blood pressure, cholesterol, etc.) during routine checkups or medical exams.
2. **Government Health Programs** –  
   Public health agencies (like the Ministry of Health or WHO projects) collect anonymized data from clinics for research on chronic diseases such as heart disease or diabetes.
3. **Medical Research Studies** –  
   Universities or research institutions conduct clinical studies where participants volunteer to provide their medical measurements and lifestyle information (like smoking, alcohol, activity).
4. **Health Insurance or Wellness Programs** –  
   In some cases, health insurance or corporate wellness programs collect similar data to assess health risks (with user consent).

# About the columns in the data collection:

There are around 13 columns in the data collection. Below are the details about every column.

| **Column** | **Meaning** | **Type / Unit** | **Details** |
| --- | --- | --- | --- |
| id | Record ID | Integer | Unique identifier for each person/record. |
| age | Age in days | Integer | Person’s age (e.g., 18393 days ≈ 50.4 years). To convert to years: age / 365. |
| gender | Gender | Categorical (1 or 2) | 1 = Female, 2 = Male (in this dataset convention). |
| height | Height | Centimeters | Person’s height in cm. |
| weight | Weight | Kilograms | Person’s weight in kg. |
| ap\_hi | Systolic blood pressure | mmHg | Higher value of blood pressure (normal around 120). |
| ap\_lo | Diastolic blood pressure | mmHg | Lower value of blood pressure (normal around 80). |
| cholesterol | Cholesterol level | Ordinal (1–3) | 1 = normal, 2 = above normal, 3 = well above normal. |
| gluc | Glucose level | Ordinal (1–3) | 1 = normal, 2 = above normal, 3 = well above normal. |
| smoke | Smoking status | Binary (0 or 1) | 1 = currently smokes, 0 = does not smoke. |
| alco | Alcohol intake | Binary (0 or 1) | 1 = drinks alcohol, 0 = does not drink. |
| active | Physical activity | Binary (0 or 1) | 1 = physically active, 0 = inactive. |
| cardio | Cardiovascular disease | Target variable (0 or 1) | 1 = has cardiovascular disease, 0 = healthy. |