

# Project Proposal

- **Problem statement:**

Cardiovascular diseases constitute a large proportion of the mortality rate, and these diseases can be detected before the disease multiplies and leads to death. Therefore, through this data collected at the moment of the medical examination, The data will be analyzed to understand the proportion of the influence of factors in the incidence of cardiovascular disease.

- **Data Description:**

All data set values were collected at the moment of the medical examination, This data was obtained from Kaggle through this link:

<https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>

id	→ ID, int
age	→ Age, int (days)
gender	→ Gender, categorical code
height	→ Height, int (cm)
weight	→ Weight, float (kg)
ap_hi	→ Systolic blood pressure, int
ap_lo	→ Diastolic blood pressure, int
cholesterol	→ Cholesterol, int, (1: normal, 2: above normal, 3: well above normal)
gluc	→ Glucose, int, ( 1: normal, 2: above normal, 3: well above normal)
smoke	→ Smoking, int, binary
alco	→ Alcohol intake, int, binary
active	→ Physical activity, int binary
cardio	→ Presence or absence of cardiovascular disease, int, binary

(70000,13)

- **Tools:**

- Programs: Jupyter Notebook, SQLite.
- Libraries: pandas, numpy , SQLAlchemy, matplotlib.
- Functions: mean, info, shape, describe, sum.
- Plots: plot, bar, scatter , heatmap.

- **MVP Goal:**

The goal of this project is to analyze the data to explore the features that most affect the incidence of cardiovascular disease.