Mini Project Using Heart Disease Dataset

This project is based on a real-world heart disease dataset. Students are expected to explore the dataset, perform statistical analysis, and visualize various attributes to understand factors that may influence heart disease risk.

# Dataset Overview

The dataset contains the following attributes:  
- age: Age of the patient  
- sex: Gender (1 = male, 0 = female)  
- cp: Chest pain type (0–3)  
- trestbps: Resting blood pressure  
- chol: Serum cholesterol (mg/dl)  
- fbs: Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)  
- restecg: Resting electrocardiographic results (0–2)  
- thalach: Maximum heart rate achieved  
- exang: Exercise induced angina (1 = yes; 0 = no)  
- oldpeak: ST depression induced by exercise relative to rest  
- slope: Slope of the peak exercise ST segment  
- ca: Number of major vessels (0–3) colored by fluoroscopy  
- thal: Thalassemia (1 = normal; 2 = fixed defect; 3 = reversible defect)  
- target: 1 = presence of heart disease; 0 = absence

# Project Tasks

Students should complete the following tasks:

* 1. Load and explore the dataset using pandas.
* 2. Perform basic statistical analysis (mean, median, mode, standard deviation) for numerical features.
* 3. Visualize the distribution of key variables using histograms and boxplots.
* 4. Create a correlation heatmap to find relationships between features.
* 5. Use scatter plots to explore the relationship between age and max heart rate (thalach), or age and cholesterol.
* 6. Visualize the proportion of patients with heart disease using a pie chart.
* 7. Group data by chest pain type and analyze the average cholesterol and heart rate values.
* 8. Use bar plots to compare the average resting blood pressure for males and females.
* 9. Use parallel coordinates to visualize multivariate relationships for a subset of the data.
* 10. Summarize the findings and prepare a brief report including visualizations and interpretations.

## Optional Challenge

Try to use a machine learning model (like Logistic Regression or Decision Tree) to predict the presence of heart disease using the available features.

Dataset link: https://www.kaggle.com/code/farzadnekouei/heart-disease-prediction/input