

Heart Disease Analysis: Understanding Key Risk Indicators

Problem Statement:

Heart disease is one of the leading causes of death worldwide. Identifying risk factors and understanding their impact can help develop strategies for early detection and prevention. This project aims to analyze a dataset of heart disease patients to explore patterns and correlations among key health indicators, such as cholesterol levels, blood pressure, and age, to determine their relationship with heart disease.

Dataset:

<https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data>

This dataset contains 303 records with 14 attributes, including patient demographics, medical test results, and a target variable indicating the presence of heart disease.

EDA Questions:

1. General Information

- How many rows and columns are in the dataset?
- Are there any missing values in the dataset?

2. Target Variable Analysis

- What is the distribution of the **target** variable (presence of heart disease: 0 or 1)?
- How many patients have heart disease, and how many don't?

3. Demographic Insights

- What is the age range of patients in the dataset?

- What is the gender distribution of the patients?

4. Health Metrics

- What are the average and median values of:
 - Resting blood pressure (**trestbps**)?
 - Serum cholesterol (**chol**)?
 - Maximum heart rate (**thalach**)?

5. Categorical Features

- How many patients have exercise-induced angina (**exang**)?
- What are the counts of different chest pain types (**cp**)?

6. Visual Analysis

- Plot the age distribution of the patients.
- Compare the average cholesterol levels between patients with and without heart disease.

7. Outliers

- Are there any outliers in cholesterol (**chol**) or resting blood pressure (**trestbps**)?