Heart Disease Dataset – Problem Statements (Python)

Basic Level (Data Understanding & Cleaning)

1 Load and Explore the Dataset

Load the dataset using **pandas**
Display first 10 rows
Check data types
Find missing values

Display summary statistics

2 Gender Distribution Analysis

Count number of males and females
Calculate percentage distribution using **NumPy**
Plot a bar chart using **Matplotlib**

3 Age Analysis

Find:
- Minimum age
- Maximum age
- Mean age
- Median age

Plot histogram of age distribution

4 Target Variable Analysis

Count number of patients with and without heart disease
Plot pie chart
Calculate disease percentage

5 Correlation Between Age and Cholesterol

Calculate correlation using df.corr()
Plot scatter plot (Age vs Cholesterol)
Interpret relationship

Intermediate Level (Grouped Analysis)

6 Chest Pain Type vs Disease

Group by cp and calculate disease rate
Plot grouped bar chart
Identify which chest pain type is most risky

7 Average Cholesterol by Gender

Group by sex
Calculate mean cholesterol
Visualize using bar plot

8 Resting Blood Pressure Analysis

Find:
- Average BP
- Patients with BP > 140

Compare disease presence in high BP group

9Maximum Heart Rate vs Disease

Compare average thalach for:
- Disease patients
- Non-disease patients

Plot boxplot

10. Exercise Induced Angina Impact

Calculate disease percentage in:
- exang = 1
- exang = 0

Visualize using bar chart

Advanced Level (Statistical + NumPy Focus)

11. ST Depression (oldpeak) Analysis

Calculate mean oldpeak by target
Plot histogram for both classes
Identify trend

12. Number of Major Vessels (ca) Impact

Group by ca
Calculate disease probability
Plot line chart

13. Thalassemia vs Disease

Cross-tabulate thal and target
Convert to percentage
Plot stacked bar chart

14. Multi-Factor Risk Analysis

Find patients with:
- Age > 50
- Cholesterol > 240
- BP > 140

Calculate percentage having disease

Use NumPy filtering

15. Create Risk Score (Custom Analysis)

Create new column:
risk_score = (chol/200) + (trestbps/120) + (oldpeak)

Classify patients as:
- Low Risk
- Medium Risk
- High Risk

Visualize distribution

Skills Covered