Heart Disease Data Analysis (Project Report)

Identifying Risk Factors and Predicting Heart Disease

Presented by : Swetha Srinivas

Project Overview

- Analyzing heart disease data to identify the factors that influence heart disease presence.
- Goal: To predict heart disease based on key health indicators for early detection and intervention.

Business Problem

- Key Challenges in Healthcare:
- • Identifying high-risk individuals based on health data.
- • Understanding key factors like cholesterol, heart rate, and blood sugar.
- • Improving medical interventions and preventive care.

Dataset

- Health records of individuals with the following key features:
- • Age, Sex, Cholesterol (chol), Max Heart Rate (thalach), Fasting Blood Sugar (fbs), and Target.

Tools Used

- Technologies and Tools:
- • Python: For data cleaning and analysis.
- • Pandas: For data manipulation.
- • Matplotlib & Seaborn: For data visualization.
- • Power BI: For interactive dashboards.

Data Preparation

- Steps involved in preparing the data:
- • Converted numerical and categorical columns to correct formats.
- • Handled missing and zero values.
- • Detected and removed outliers in cholesterol and heart rate.

Key Metrics

- Important metrics for analysis:
- • Target: Presence or absence of heart disease.
- • Key Indicators: Age, cholesterol, max heart rate, and fasting blood sugar.
- • Sex Distribution: Analyzed heart disease by gender.

Analysis Insights

- Key Insights from the analysis:
- • Age: Heart disease is prevalent in ages 40-60, peaking at 50-60.
- • Cholesterol: Levels above 250 mg/dl are associated with higher risk.
- • Max Heart Rate: Lower rates correlate with heart disease, especially in ages 50-60.
- • Gender: Males have a slightly higher prevalence.

Recommendations

- Recommended strategies:
- • Targeted screenings for ages 40-60.
- • Cholesterol management for levels above 250 mg/dl.
- • Encourage cardiovascular exercises for better heart health.
- • Gender-specific health campaigns for men aged 40-60.

Limitations

- Limitations of this analysis:
- • Data Size: Limited dataset, affecting generalizability.
- • Static Data: Analysis does not consider changes over time.
- • External Factors: Lifestyle and medication use not included.