
Heart Disease Dataset Analysis

Project Overview

This project offers an analytical study of a heart disease dataset using Power BI and SQL, aiming to identify patterns and key predictors of heart disease through visualizations and data-driven queries.

Dataset

The dataset (heart_disease_prediction_data.csv) is a cleaned and structured version of a publicly available heart disease dataset. It contains various clinical features related to cardiovascular health.

Features

- **age**: Patient's age
- **sex**: Gender (1 = male, 0 = female)
- **cp**: Type of chest pain (0–3)
- **trestbps**: Resting blood pressure (mm Hg)
- **chol**: Serum cholesterol level (mg/dl)
- **fbs**: Fasting blood sugar > 120 mg/dl (1 = true, 0 = false)
- **restecg**: Resting electrocardiogram results
- **thalachh**: 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 colored by fluoroscopy (0–3)
- **thal**: Thalassemia type (1 = normal, 2 = fixed defect, 3 = reversible defect)
- **target**: Presence of heart disease (1 = yes, 0 = no)

Data Source

The original dataset was sourced from:

- Kaggle – Heart Disease Prediction Dataset

Data Cleaning Process

1. **Handled Missing Values**: Removed rows with null or undefined entries.
2. **Standardized Data Types**: Correctly encoded categorical variables such as sex, target, and cp.
3. **Column Renaming**: Enhanced readability and consistency by renaming columns.

Tools Used

- **Power BI** – For creating interactive dashboards and visualizations
- **SQL** – For executing exploratory queries and discovering patterns

---

Key Insights

- **Age** and **chest pain type** emerge as significant predictors of heart disease.
- **Males** exhibit a slightly higher prevalence of heart disease compared to females.
- **Exercise-induced angina** and elevated **oldpeak** values are strongly associated with heart disease.
- Higher **cholesterol levels** and reduced **maximum heart rate** are common among at-risk patients.

Recommendations

- Promote routine cardiovascular check-ups, particularly for individuals over 40 years old and those experiencing chest pain.
- Encourage regular physical activity and monitoring of heart rate to mitigate risks.
- Give special attention to patients presenting with exercise-induced angina or elevated ST depression.
- Consider implementing gender-specific strategies for early detection and intervention.

Limitations

- The dataset is relatively small (approximately 300 entries), limiting its broader applicability.
- As historical data, it may not fully capture recent advancements in heart disease diagnosis or treatment.
- Dropping records with missing values may have led to the loss of critical information.
- The dataset lacks lifestyle variables (such as diet, smoking habits, and stress levels), which are important for comprehensive heart health analysis.