Final Report

Understanding Heart Disease Risk Factors through Health and Demographic Analysis

---

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, responsible for an estimated 17.9 million deaths each year—equivalent to 31% of all global deaths. Four out of five CVD-related deaths result from heart attacks and strokes, and one-third occur prematurely in individuals under the age of 70.

Heart failure is a common outcome of CVDs, making early identification of risk factors critical for prevention, timely diagnosis, and improved patient outcomes. Individuals with cardiovascular risk factors such as hypertension, diabetes, or hyperlipidaemia require early detection and management. In this context, predictive modeling and machine learning can enhance clinical screening and inform preventive healthcare strategies.

This report analyzes medical and demographic data to:

- Identify the strongest predictors of heart disease.
- Compare risk across different patient groups.
- Provide evidence-based recommendations for clinical and preventive practices.

---

2. Objectives

The main objectives of this project are:

1. To determine the most significant demographic and medical predictors of heart disease.
2. To analyze relationships between patient characteristics (e.g., age, gender, cholesterol, blood pressure, chest pain type, and exercise tolerance) and heart disease prevalence.
3. To generate actionable insights supporting early clinical screening and preventive healthcare strategies.

---

3. Data Source and Citation

- **Dataset Name:** Heart Failure Prediction Dataset
- **Source:** Kaggle – *Heart Failure Prediction*
- **Citation:**
  fedesoriano. (September 2021). *Heart Failure Prediction Dataset*. Retrieved [Date Retrieved] from Kaggle.

---

4. Acknowledgements

This dataset is based on research contributions from multiple institutions and experts:

- **Hungarian Institute of Cardiology, Budapest** — Andras Janosi, M.D.
- **University Hospital, Zurich, Switzerland** — William Steinbrunn, M.D.
- **University Hospital, Basel, Switzerland** — Matthias Pfisterer, M.D.
- **V.A. Medical Center, Long Beach and Cleveland Clinic Foundation** — Robert Detrano, M.D., Ph.D.
- **Donor:** David W. Aha (aha '@' ics.uci.edu), (714) 856-8779

---

5. Data Overview

The dataset contains **11 attributes** relevant to cardiovascular health:

- **Age** — Patient's age (years)
- **Sex** — Gender
- **ChestPainType** — Type of chest pain
- **RestingBP** — Resting blood pressure
- **Cholesterol** — Serum cholesterol level
- **FastingBS** — Fasting blood sugar
- **RestingECG** — Resting electrocardiogram results
- **MaxHR** — Maximum heart rate achieved
- **ExerciseAngina** — Presence of exercise-induced angina
- **Oldpeak** — ST depression induced by exercise relative to rest
- **ST_Slope** — Slope of the peak exercise ST segment
- **HeartDisease** — Target variable (presence of heart disease)

---

6. Methodology

The analysis followed a structured workflow using Python:

- **Tools & Libraries:**
    1. *Pandas* — Data cleaning, manipulation, exploration
    2. *Matplotlib & Seaborn* — Visualization of distributions, correlations, and trends
    3. *Statistical Analysis* — Correlation and hypothesis testing
- **Workflow:**
    1. Data cleaning (duplicates, missing values)
    2. Exploratory Data Analysis (EDA)
    3. Correlation and predictor evaluation
    4. Insights generation and recommendations

---

7. Key Insights

- **Age:** Heart disease prevalence rises sharply with age, from ~34% in patients aged 30–39 to over 70% in patients aged 60+.
- **Gender:** Males consistently exhibit higher prevalence. In the 30–39 group, 40% of males had heart disease compared to only 16% of females.
- **Exercise-Induced Angina:** Strong predictor — ~85% of patients with angina also had heart disease.
- **Chest Pain Type:** Asymptomatic (ASY) chest pain is the most common among patients with heart disease (~80%).
- **Medical Indicators:**
    - *MaxHR:* Strong negative correlation (-0.40)
    - *Oldpeak:* Strong positive correlation (+0.40)
    - *RestingBP & Cholesterol:* Weak correlations

---

8. Recommendations

8.1 Target High-Risk Groups

- Prioritize screening for males and individuals over 50 years old.

8.2 Prioritize Strong Predictors

- Emphasize **MaxHR** and **Oldpeak** over traditional measures like cholesterol or resting BP.

- Treat **exercise-induced angina** as a clinical red flag.

8.3 Maintain Core Diagnostics

- Continue routine use of blood pressure, cholesterol, and ECG tests.
- Proactively screen patients presenting with **asymptomatic chest pain**.

8.4 Preventive Strategies

- Encourage lifestyle modifications: diet, exercise, smoking cessation.
- Provide **tailored exercise programs** for borderline patients.
- Launch **awareness campaigns** for healthcare providers and the public to highlight that heart disease often presents without classic chest pain symptoms.

Summary:
By focusing on high-risk groups, emphasizing stronger predictors, and reinforcing prevention, healthcare systems can reduce both the prevalence of heart disease and delays in diagnosis.

---

9. Conclusion

This study underscores the need to move beyond traditional predictors (cholesterol, resting BP) and place greater emphasis on stronger indicators such as **exercise-induced angina, MaxHR, and Oldpeak**.

Effective prevention requires a dual approach:

1. **Clinical Precision** — Improved risk stratification and early detection.
2. **Population-Wide Prevention** — Lifestyle interventions and education.

Together, these strategies can significantly reduce premature mortality and the global burden of cardiovascular disease.