

UCI Heart Disease Report – Technical Analysis (Excel-Based Project)

1. Outline

This report presents a detailed analysis of the UCI Heart Disease dataset using Microsoft Excel. It follows a structured flow covering:

- Introduction
- Story of Data
- Data Splitting and Preprocessing
- Pre-Analysis Insights
- In-Analysis Observations
- Post-Analysis Recommendations
- Data Visualizations & Charts
- Conclusion

The goal of this project is to uncover demographic and clinical patterns that explain the prevalence of heart disease and to demonstrate how Excel can be used as a practical analytical tool for healthcare datasets.

2. Introduction

The main objective of this analysis is to identify the major factors associated with heart disease among patients using the UCI dataset. The project focuses on understanding how variables such as age, gender, cholesterol level, and blood pressure correlate with the presence of heart disease.

Problem Being Addressed

Heart disease remains one of the leading causes of death worldwide. Many hospitals and health research teams rely on data analysis to recognize early warning signs. This report aims to translate raw medical data into clear insights that can support clinical awareness, patient screening, and early diagnosis.

Key Datasets and Methodologies

The dataset used is the **UCI Heart Disease dataset**, containing multiple patient attributes such as age, sex, chest pain type, cholesterol, resting blood pressure, and maximum heart rate achieved.

All analysis was done in Microsoft Excel, leveraging:

- **PivotTables** for summarization
- **IF and COUNTIFS** formulas for conditional analysis

- **Charts and conditional formatting** for visual storytelling
- **Descriptive statistics** to uncover patterns and relationships

3. Story of Data

This data tells the story of UCI heart disease. It shows us the prevalence of heart disease, the age bracket that are affected by heart disease, the type of heart diseases from different countries, differences in the kinds of heart diseases in male and female. It also includes the relationship of heart diseases with exercise, giving us a full idea of measure events with heart failure.

Stakeholders of the project

Public health officers for educating the general public on the dangers and prevalence of heart disease

Value to the Industry

This analysis demonstrates how even a non-coding tool like Excel can generate meaningful health insights that guide policy and clinical awareness.

Data Source

The data was downloaded from [Kaggle.com](https://www.kaggle.com)

Data Structure

Each row in the dataset represents a single patient, while columns describe demographic and clinical details. Some of the major variables include:

- **Age** – Patient's age in years
- **Sex** – Gender (Male/Female)
- **Cholesterol** – Serum cholesterol in mg/dl
- **Trestbps** – Resting blood pressure in mm Hg
- **Thalach** – Maximum heart rate achieved
- **Exang** – Exercise-induced angina (Yes/No)
- **Oldpeak** – ST depression induced by exercise relative to rest
- **Target** – Indicates the presence (1) or absence (0) of heart disease

Important Features and Their Significance

- **Age & Sex:** Basic demographic indicators that influence cardiovascular risk.
- **Cholesterol & Blood Pressure:** Primary medical indicators linked with heart conditions.
- **Exercise and ECG Data:** Reveal how the heart performs under stress and physical activity.

Data Limitations or Biases

The dataset may contain missing or incomplete records. It also represents a specific group of patients, so results might not generalize globally. Despite that, it remains a reliable dataset for pattern exploration and clinical insight extraction.

4. Data Splitting and Preprocessing

Data Cleaning

Initial cleaning involved removing duplicate records, fixing inconsistent entries, and ensuring all numeric fields were properly formatted. Excel's "Remove Duplicates" and ISBLANK checks were used to identify and clean empty cells.

Handling Missing Values

- Numeric missing values were replaced with the median of the column to prevent bias.
- Categorical missing data were labeled as "Unknown" for clarity.

Data Transformations

To make the data easier to interpret, new categorical variables were introduced:

- **Age Groups:** <40, 40–55, 56–70, >70
- **Cholesterol Categories:** Normal, Borderline, High
- **Risk Score:** A simple score combining high-risk attributes (e.g., high BP, high cholesterol, exercise-induced angina).

Data Splitting

The dependent variable was Heart Disease Presence (Target), while independent variables included age, sex, cholesterol, blood pressure, and exercise-related attributes.

Industry Context

The project fits within the healthcare and public health analytics sector, offering insight into how clinical and demographic data can be used for preventive care.

Stakeholders

- Hospitals and clinical researchers
- Public health policy teams
- Healthcare data analysts and data science learners

5. Pre-Analysis Insights

- a. Prevalence of heart disease in the respective gender
- b. Percentage of heart disease in male and female
- c. Age percentage analysis by gender
- d. Age percentage analysis by country
- e. The country with the highest cases of heart disease
- f. Analysis of gender percentage by country
- g. Age bracket with the highest heart failure
- h. Rate of blood pressure in male and female
- i. The type of chest pain in different country
- j. Cholesterol measure by type of chest pain
- k. Age group and the type of chest pain they experience
- l. Gender that performs exercise the most
- m. Blood pressure rate by type of chest pain
- n. Maximum heart rate by type of chest pain
- o. Age bracket by heart disease stages
- p. Maximum heart rate y respective genders
- q. Cholesterol measure by gender
- r. Cholesterol measure by exercise performances
- s. Gender by ECG

6. In-Analysis Observations

- Older women form a much smaller percentage compared to men in this dataset
- This confirms that the majority of men suffering from heart disease are within the older age group
- Older men are at higher risk, as they dominate the group most affected by heart problems that can potentially lead to death
- The female population is significantly smaller across all countries, representing only 21.09% in total
- Overall, male population accounts for 78.91%, making them more prone to heart disease cases than females
- The trend is consistent across all countries, where men are recorded to have the most heart disease cases
- Cleveland stands out with the highest percentage of elderly cases, accounting for 33.56% of the population
- Hungary ranks second with 28.50% of elderly individuals suffering from heart disease
- VA Long Beach is third, representing 24.11% of the population with heart disease
- Switzerland has the lowest share at 13.82% of the elderly population with heart disease
- Asymptomatic cases are the most prevalent type across all countries, totaling 496 cases
- Non-anginal cases rank second in prevalence, especially in Cleveland, Switzerland, and VA Long Beach
- Typical angina remains the least reported with only 47 cases overall

- Asymptomatic cases dominate with a cumulative total of 27,260 cases, highlighting their silent but dangerous spread
- Non-anginal cases follow with 10,880 cases across the dataset
- Atypical angina is the third most common, with 8,512 cases recorded
- Typical angina is the least prevalent with just 2,578 cases in total
- The dominance of asymptomatic cases highlights the urgent need for early screening programs, as patients often go undiagnosed until complications arise
- The consistently high male-to-female ratio across all countries suggests that gender-specific health interventions should be prioritized
- The sharp difference between Cleveland and Switzerland shows the importance of localized strategies since prevalence varies greatly by region
- The data also emphasizes the need for targeted elderly healthcare programs, especially since men are the most vulnerable groups between patients with and without heart disease.

Analysis Techniques Used in Excel

- PivotTables for age–gender–disease cross-tab analysis.
- COUNTIFS and AVERAGEIFS for group statistics.
- Conditional formatting to visually emphasize high-risk values.
- Bar and Line Charts for trend visualization.

7. Post-Analysis Recommendations

- Special care should be prioritized for older men since data shows they form the largest group affected by heart disease
- Preventive strategies should be developed and emphasized for men, targeting conditions that can be controlled or avoided
- Treatment centers should be equipped with gender-specific resources, ensuring that men's health needs are properly met with adequate medical tools and facilities
- Public awareness campaigns should be intensified to educate society on the higher risk of heart disease among men
- More focused attention should be given to Cleveland, as it has the highest proportion of elderly individuals with heart disease
- Asymptomatic cases are the most common, hence they should be prioritized with early detection programs and appropriate treatment options
- Special emphasis should be placed on asymptomatic patients to prevent sudden complications due to late diagnosis
- Scientific research should be expanded to discover better treatment methods and medications for the most prevalent types of heart disease
- Regular community-based screening programs should be introduced to identify early signs of heart disease, especially in high-risk groups
- Lifestyle modification programs should be promoted, encouraging healthier diets, regular exercise, reduced alcohol intake, and smoking cessation

- Collaborations between local governments, hospitals, and NGOs should be strengthened to improve funding and resource allocation for heart disease management
- Mental health support should also be included in heart disease care since stress and depression significantly contribute to cardiovascular risks
- Data collection systems should be improved to track trends in heart disease prevalence, treatment outcomes, and high-risk populations for better planning
- Special training should be provided for healthcare workers to improve diagnosis, gender-sensitive treatment, and patient education strategies

Comparison with Initial Findings

The hypotheses were confirmed most risk indicators identified early were validated through deeper Excel exploration.

8. Data Visualizations & Charts

Charts included:

- **Bar Chart:** Disease rate by gender and age group.
- **Scatter Plot:** Cholesterol vs. Max heart rate, colored by disease status.
- **Pie Chart:** Distribution of patients with and without heart disease.
- **Dashboard:** A simple Excel dashboard using slicers for interactive filtering.

Each chart was accompanied by short interpretations to make the findings easy to digest

9. Conclusion

This project highlights how simple Excel tools can extract meaningful clinical insights from structured data. It confirmed that age, gender, cholesterol, and exercise response play key roles in heart disease risk.

While Excel has limitations compared to advanced statistical software, it remains a valuable tool for early-stage healthcare analysis and internal reporting.