❤️ Heart Disease Exploratory Data Analysis Report

1. Overview

This project focuses on performing **Exploratory Data Analysis (EDA)** on a heart disease dataset to understand patient demographics, clinical indicators, and their relationship with heart disease presence. The objective is to identify meaningful patterns and risk indicators through statistical summaries and visual analysis, without building predictive models.

---

2. Dataset Description

The dataset consists of patient-level medical and diagnostic data, including:

- Demographic features such as age and sex

- Clinical measurements like resting blood pressure, cholesterol levels, and maximum heart rate

- Diagnostic indicators including chest pain type, ECG results, and exercise-induced angina

- A target variable indicating the presence or absence of heart disease

The dataset was checked for missing values and duplicate records, ensuring data quality before analysis.

---

3. Data Preparation & Initial Exploration

Initial exploration included:

- Reviewing dataset structure, dimensions, and data types

- Identifying and removing duplicate records

- Generating descriptive statistics to understand feature distributions

The dataset shows significant variability in clinical measurements such as cholesterol, blood pressure, and heart rate, indicating diverse health profiles among patients.

---

4. Correlation Analysis

Correlation analysis was performed to examine relationships between clinical variables and heart disease presence.

Key observations:

- Maximum heart rate achieved shows a negative correlation with heart disease

- Exercise-induced angina and ST depression exhibit stronger associations with heart disease

- Several features demonstrate weak individual correlations, highlighting the importance of multivariate analysis

---

5. Target Variable Distribution

Analysis of the target variable shows a relatively balanced distribution of patients with and without heart disease. This balance enables meaningful comparison across demographic and clinical features.

---

6. Demographic Insights

- Male patients are more prevalent in the dataset compared to female patients

- Heart disease occurrence is higher among male patients

- Age distribution analysis indicates that heart disease prevalence increases with age, particularly in middle-aged and older groups

---

7. Clinical Feature Analysis

7.1 Chest Pain Type

Certain chest pain types show a stronger association with heart disease, emphasizing their diagnostic importance in clinical assessments.

7.2 Fasting Blood Sugar

Fasting blood sugar levels show limited separation between patients with and without heart disease, suggesting it may not be a strong standalone indicator.

7.3 Resting Blood Pressure

Resting blood pressure varies across patients, with slightly higher values observed in heart disease cases. Gender-wise comparison shows variation but no extreme separation.

7.4 Serum Cholesterol

Cholesterol levels exhibit high variance across the dataset, with overlapping distributions between heart disease and non-heart disease patients, indicating limited standalone predictive power.

7.5 Continuous Variables

Continuous variables such as age, cholesterol, and heart rate display diverse distributions, reinforcing the need for combined feature analysis rather than isolated interpretation.

---

8. Key Insights & Outcomes

- Age, chest pain type, and maximum heart rate emerge as important indicators associated with heart disease

- Exercise-induced angina and ST depression show meaningful relationships with heart disease presence

- Several clinical features demonstrate high variability, requiring multivariate consideration for effective analysis

- The analysis highlights the importance of data understanding before applying predictive models

---

9. Conclusion

This exploratory analysis provides a structured understanding of demographic and clinical factors related to heart disease. The insights derived from visual and statistical exploration form a strong foundation for future predictive modeling or advanced analytics. The project demonstrates the role of EDA in uncovering meaningful patterns and guiding data-driven decision-making in healthcare analytics.