Exploratory Data Analysis on Heart Disease Dataset

By Afiya Abbasi

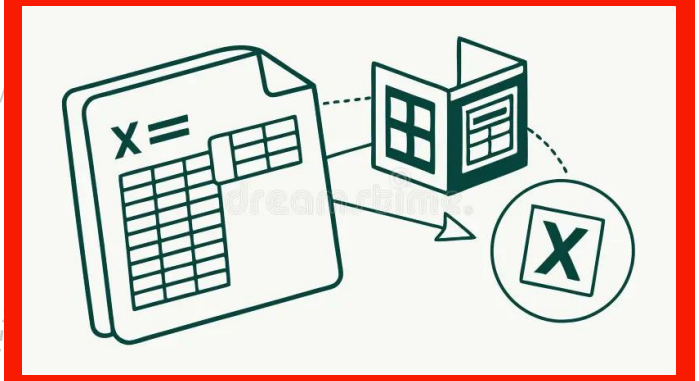
Introduction

- This project focuses on analyzing health and demographic data to identify patterns and traits associated with cardiovascular disease. The dataset contains various health indicators, and through exploratory data analysis (EDA), we aim to discover insights that could help in early diagnosis and prevention.

Dataset Overview

- - Source: Kaggle – Cardiovascular Disease Dataset
- - Rows: 69,000+ observations
- - Columns: 14 attributes
- - Key features: Age, Gender, BMI, Blood Pressure, Cholesterol, Glucose, Lifestyle habits
- - Target: 'cardio' (1 = heart disease, 0 = no disease)

Data Preparation



- - Converted age from days to years
- - Calculated BMI using height and weight
- - Dropped unnecessary columns
- - Checked for missing values
- - Ensured categorical variables were properly formatted

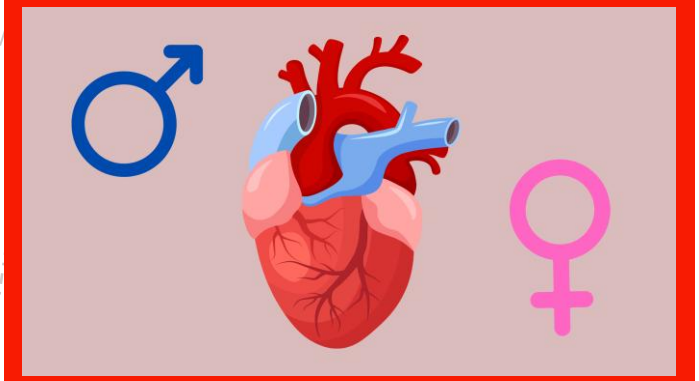
Univariate Analysis:

Age Distribution

- Histogram of age distribution among people with and without heart disease.

Univariate Analysis:

Gender vs Heart Disease



- Distribution of heart disease across genders.

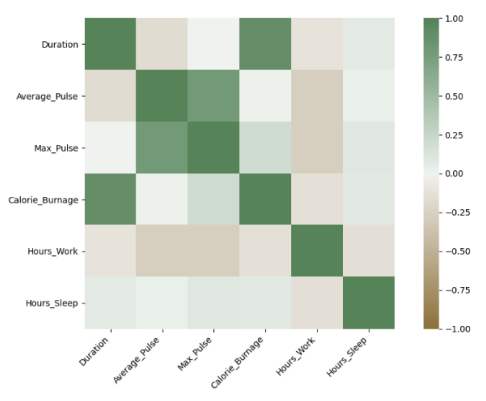
Univariate Analysis:

BMI

- Box plot showing BMI values for individuals with and without heart disease.

Bivariate Analysis:

Correlation Heatmap



- Heatmap showing correlation between numerical variables.

Key Insights

- - Age and BMI are positively associated with heart disease.
- - Males showed a slightly higher risk than females.
- - High cholesterol and glucose levels were strongly correlated with heart disease.
- - Lifestyle habits like smoking and physical inactivity were relevant indicators.

Conclusion

- The EDA process revealed several important associations between demographic and health factors and the likelihood of heart disease. These findings are foundational for building predictive models and informing public health strategies.

- By Afiya Abbasi