



KIET
GROUP OF INSTITUTIONS

Connecting Life with Learning

Assesment Report

on

“Exploratory Data Analysis of Heart Disease Risk Factors”

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY
DEGREE**

SESSION 2024-25

in

Name of discipline

By

Divyanshi Gautam (20240110040087)

Under the supervision of

“Mr. Abhishek Shukla”

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow
(Formerly UPTU)
May, 2025



A

Introduction

Heart disease remains one of the leading causes of mortality worldwide. Early detection and understanding of the contributing risk factors are crucial for prevention and treatment. This report presents an exploratory data analysis (EDA) of a heart disease dataset to uncover significant patterns, distributions, and correlations among the clinical features. The goal is to visualize how certain health indicators relate to the presence or absence of heart disease and lay the groundwork for future predictive modeling.

Methodology

1. Dataset Upload: The CSV file is uploaded and read into a Pandas DataFrame in Google Colab.

2. Data Exploration: Basic dataset properties such as shape, column names, and data types are displayed.

3. Data Cleaning: We check for missing or duplicate values.

4. Visualization: We employ Seaborn and Matplotlib to create meaningful plots including:

- Heart disease distribution
- Age distribution
- Chest pain type vs. heart disease
- Correlation heatmap

5. Tools Used:

- Python programming language.
- Libraries: Pandas, Matplotlib.

CODE

```
# =====  
# UPLOAD CSV FILE (Google Colab)  
# =====  
from google.colab import files  
import pandas as pd  
import io  
  
# Upload CSV file  
uploaded = files.upload()  
filename = next(iter(uploaded))  
df = pd.read_csv(io.BytesIO(uploaded[filename]))  
  
# =====  
# IMPORT LIBRARIES  
# =====  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
sns.set(style='whitegrid', palette='Set2')
```

```
plt.rcParams['figure.figsize'] = (8, 5)

# =====

# 3 BASIC INFO

# =====

print("Dataset Shape:", df.shape)

print("Column Names:", df.columns.tolist())

df.head()

# Identify the target column

target_col = 'target' if 'target' in df.columns else 'output'

# =====

# 4 GRAPH 1: Heart Disease Distribution

# =====

sns.countplot(data=df, x=target_col)

plt.title(" ❤️ Heart Disease Distribution")

plt.xlabel("Heart Disease (1 = Yes, 0 = No)")

plt.ylabel("Count")

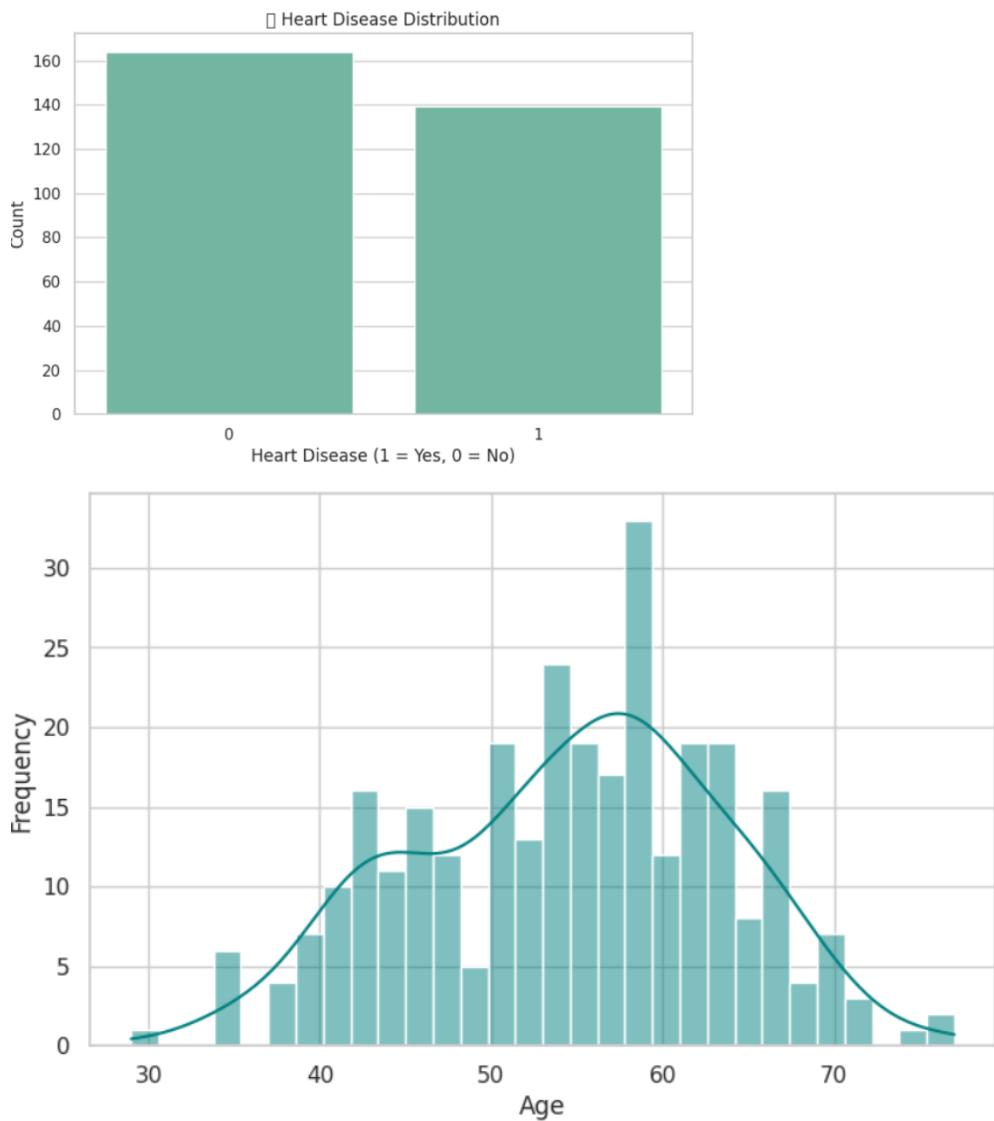
plt.show()

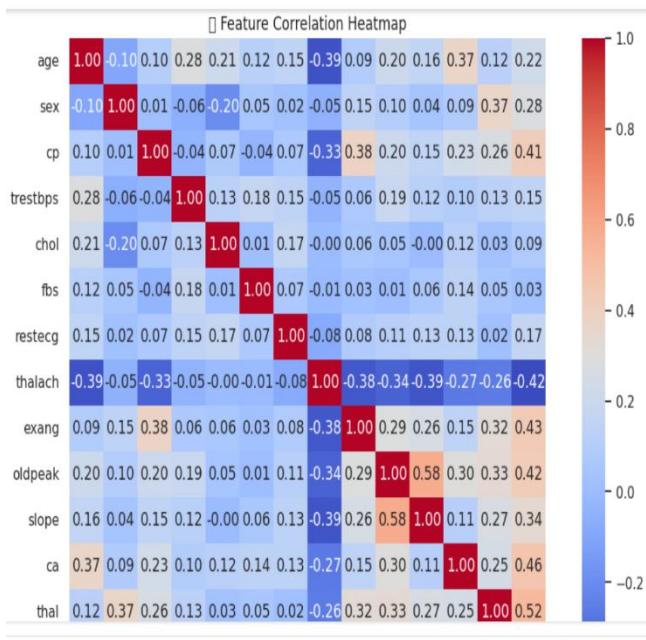
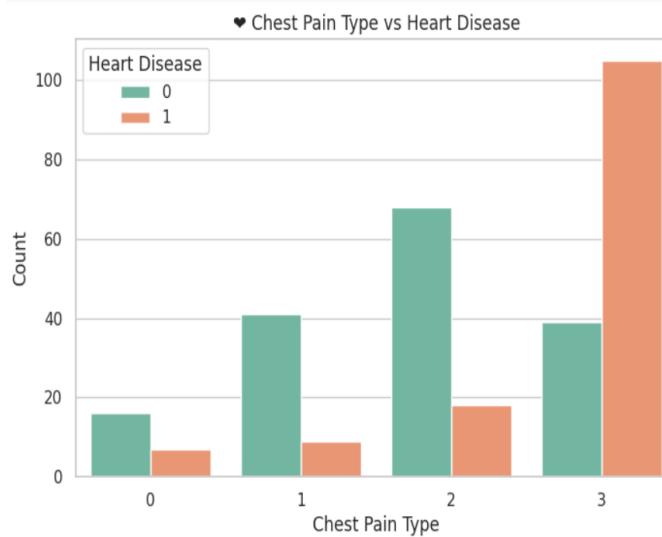
# =====
```

```
# 5 GRAPH 2: Age Distribution  
# =====  
  
if 'age' in df.columns:  
  
    sns.histplot(df['age'], kde=True, bins=30, color='teal')  
  
    plt.title("🎂 Age Distribution")  
  
    plt.xlabel("Age")  
  
    plt.ylabel("Frequency")  
  
    plt.show()  
  
# =====  
  
# 6 GRAPH 3: Chest Pain Type vs Heart Disease (if available)  
# =====  
  
# Chest pain column could be named 'cp'  
  
if 'cp' in df.columns:  
  
    sns.countplot(data=df, x='cp', hue=target_col)  
  
    plt.title("❤️ Chest Pain Type vs Heart Disease")  
  
    plt.xlabel("Chest Pain Type")  
  
    plt.ylabel("Count")  
  
    plt.legend(title="Heart Disease")  
  
    plt.show()  
  
# =====
```

```
# █GRAPH 4: Correlation Heatmap  
# =====  
plt.figure(figsize=(12, 8))  
sns.heatmap(df.corr(), annot=True, fmt=".2f", cmap='coolwarm', square=True)  
plt.title("🔗 Feature Correlation Heatmap")  
plt.show()
```

Output





References/Credits

- . Dataset Source: chatgpt**
 - . Python Libraries Used: Pandas,
Matplotlib**
 - . Images/Graphs: Generated using
Matplotlib**
-

