Healthcare Data Exploration

Author:Kartikey Teotia

Date:11 March 2025

Submitted to:Mr.Shivansh Prasad

# 1. Introduction

This report explores healthcare data, analyzing key trends, missing values, and correlations. It includes visualizations to help understand the dataset.

# 2. Methodology

The dataset is loaded using pandas, and various exploratory data analysis (EDA) techniques are applied. We check for missing values, visualize distributions, and analyze feature correlations.

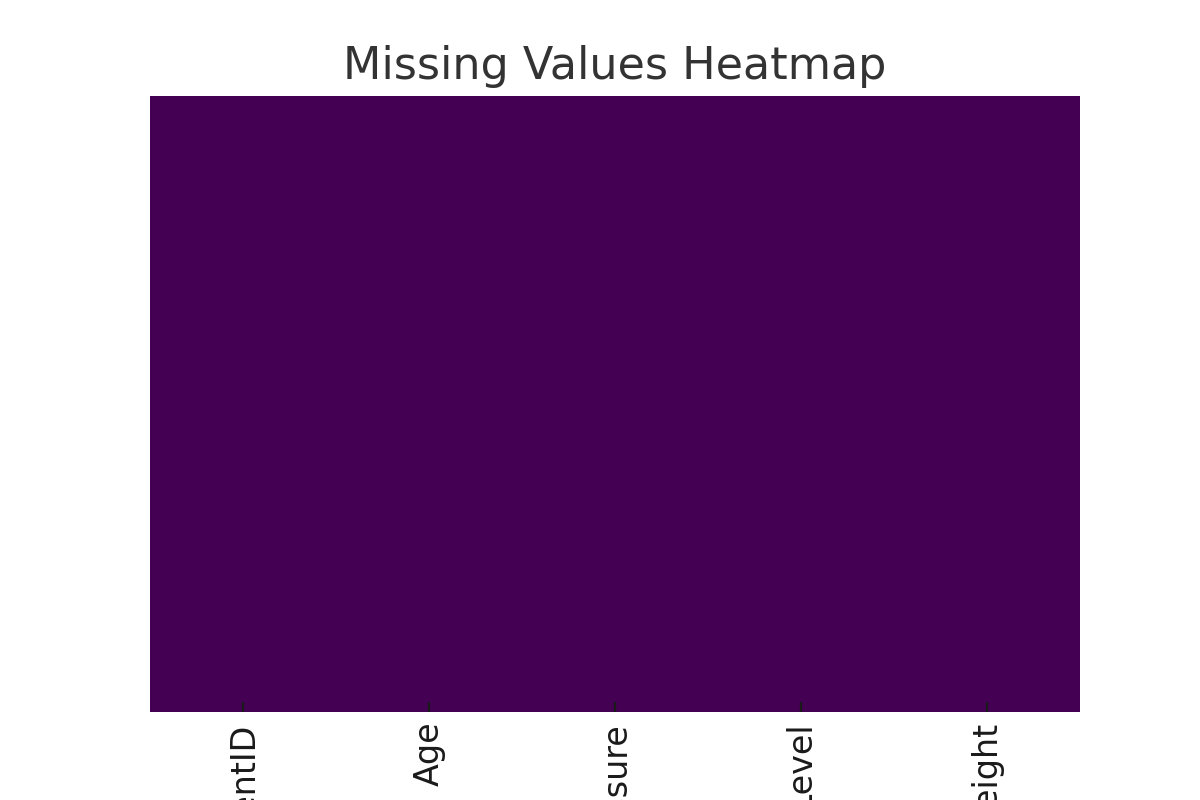
# 3. Python Code

import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
# Load the dataset  
file\_path = '/content/healthcare\_data.csv'  
df = pd.read\_csv(file\_path)  
  
# Display basic information  
print("Dataset Information:")  
print(df.info())  
  
# Display basic statistics  
print("  
Summary Statistics:")  
print(df.describe())  
  
# Check for missing values  
print("  
Missing Values:")  
print(df.isnull().sum())  
  
# Visualizing missing values  
plt.figure(figsize=(10, 5))  
sns.heatmap(df.isnull(), cmap='viridis', cbar=False, yticklabels=False)  
plt.title("Missing Values Heatmap")  
plt.show()  
  
# Visualizing data distribution  
plt.figure(figsize=(10, 5))  
df.hist(bins=30, figsize=(10, 8))  
plt.suptitle("Feature Distributions", fontsize=14)  
plt.show()  
  
# Pairplot to check relationships (if dataset is not too large)  
if df.shape[1] <= 5: # Avoid excessive plotting for large datasets  
 sns.pairplot(df)  
 plt.show()  
  
# Correlation heatmap  
plt.figure(figsize=(8, 6))  
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt='.2f')  
plt.title("Feature Correlation Heatmap")  
plt.show()  
  
print("Exploration completed!")

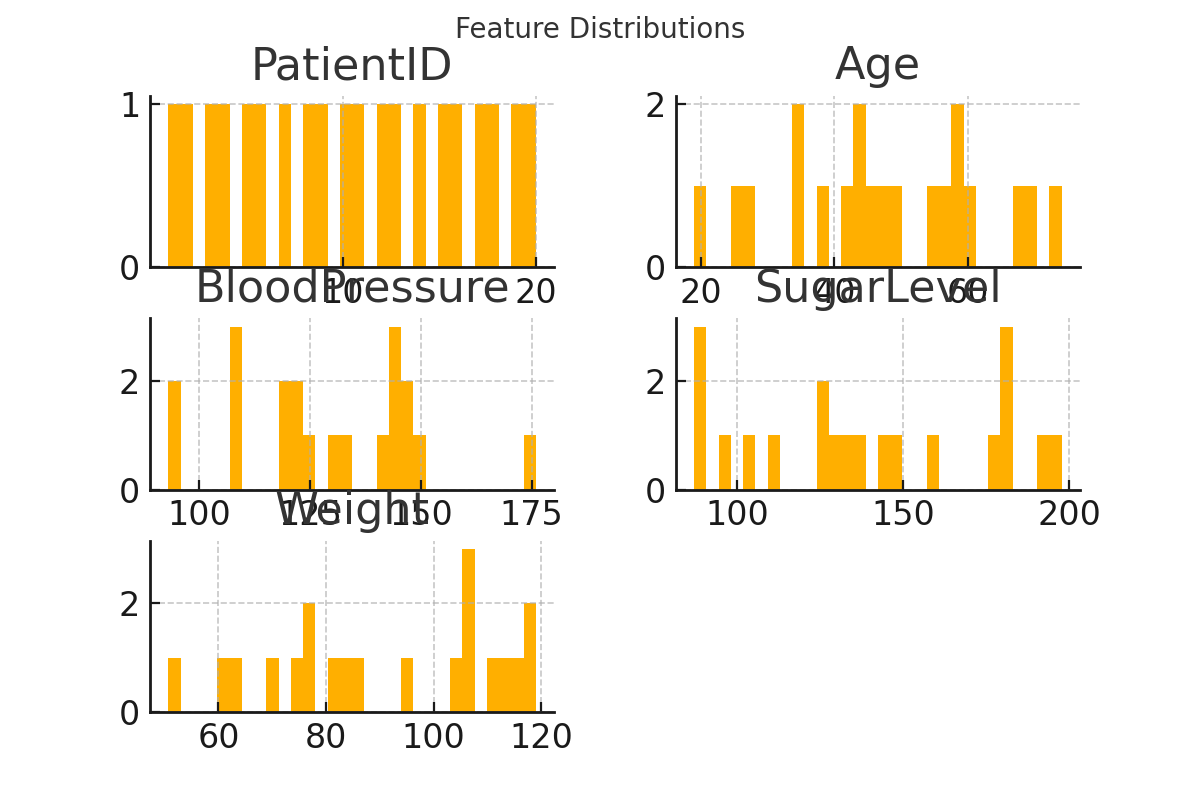
# 4. Screenshots & Output

Below are visualizations representing the healthcare data exploration results.

1. Missing Values Heatmap:



2. Feature Distribution:



3. Correlation Heatmap:

