**Title Page**

**Problem Statement: Healthcare Data Exploration – Visualize patient data like blood pressure, sugar levels, and weight to identify health trends.**

* **Student Name**: Aditya Raj Patel
* **Branch:** CSE AIML
* **SECTION:** A
* **University Roll No:** 202401100400017

1. **Introduction**

Healthcare data analysis is an essential aspect of modern medical practices and research. By analyzing patient data, such as blood pressure, sugar levels, and weight, healthcare professionals can identify trends that can lead to early diagnosis, better treatment plans, and improved patient outcomes. In this project, we explore patient data to identify trends and relationships between health metrics using data

visualization and basic machine learning models.

The goal of this project is to:

1. Visualize trends in **blood pressure**, **sugar levels**, and **weight**.
2. Explore the relationship between these metrics to understand their impact on each other.
3. Develop a simple predictive model to forecast blood pressure based on sugar levels and weight.
4. **Methodology**

In this project, we used the following methodology:

1. **Data Collection**:
   * A synthetic dataset containing health information for 15 patients was used. The dataset includes the following attributes:
     + Patient ID: Unique identifier for each patient.
     + Age: The age of the patient.
     + Blood Pressure: Blood pressure readings for the patient.
     + Sugar Level: Blood sugar levels of the patient.
     + Weight: Weight of the patient.
     + Date: Date when the data was recorded.
2. **Data Preprocessing**:
   * The dataset was uploaded into Google Colab, and basic cleaning steps were performed, such as checking for missing values and ensuring proper data types.
3. **Exploratory Data Analysis (EDA)**:
   * Visualizations were created to explore the data:
     + **Blood pressure trends over time** were visualized using line plots.
     + **Sugar levels vs. weight** were visualized using scatter plots.
     + **Distributions of blood pressure and weight** were analyzed with histograms.
     + **Correlation heatmap** was generated to explore the relationships between health metrics
4. **Code**

# Importing necessary libraries for data analysis and visualization

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

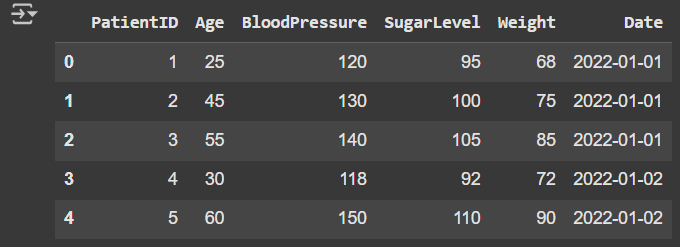
from sklearn.metrics import mean\_squared\_error

# Load the dataset

df = pd.read\_csv('/healthcare\_data.csv')  # Assuming the dataset is named 'healthcare\_data.csv'

# Display the first few rows of the dataset

df.head()



# Check for missing values

print("Missing values in each column:")

print(df.isnull().sum())

# Check the data types of each column

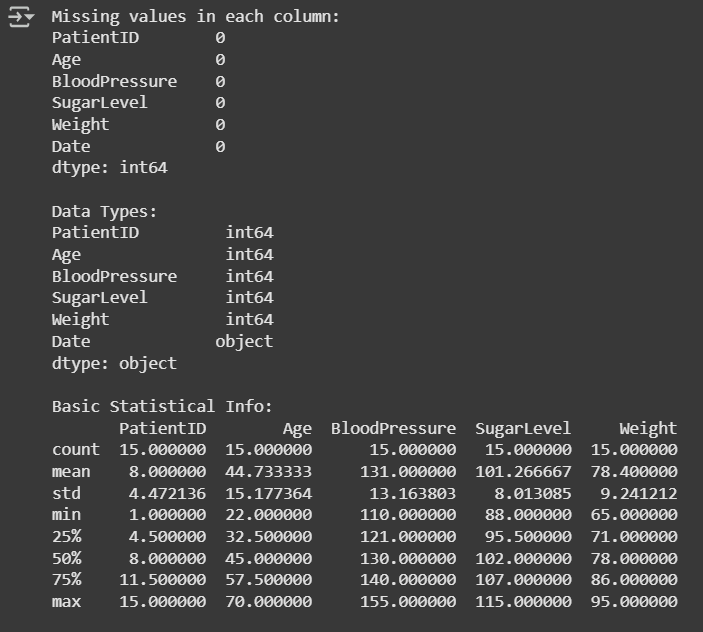
print("\nData Types:")

print(df.dtypes)

# Get basic statistical information about the dataset

print("\nBasic Statistical Info:")

print(df.describe())



# Scatter plot of Sugar Levels vs Weight

plt.figure(figsize=(10, 6))

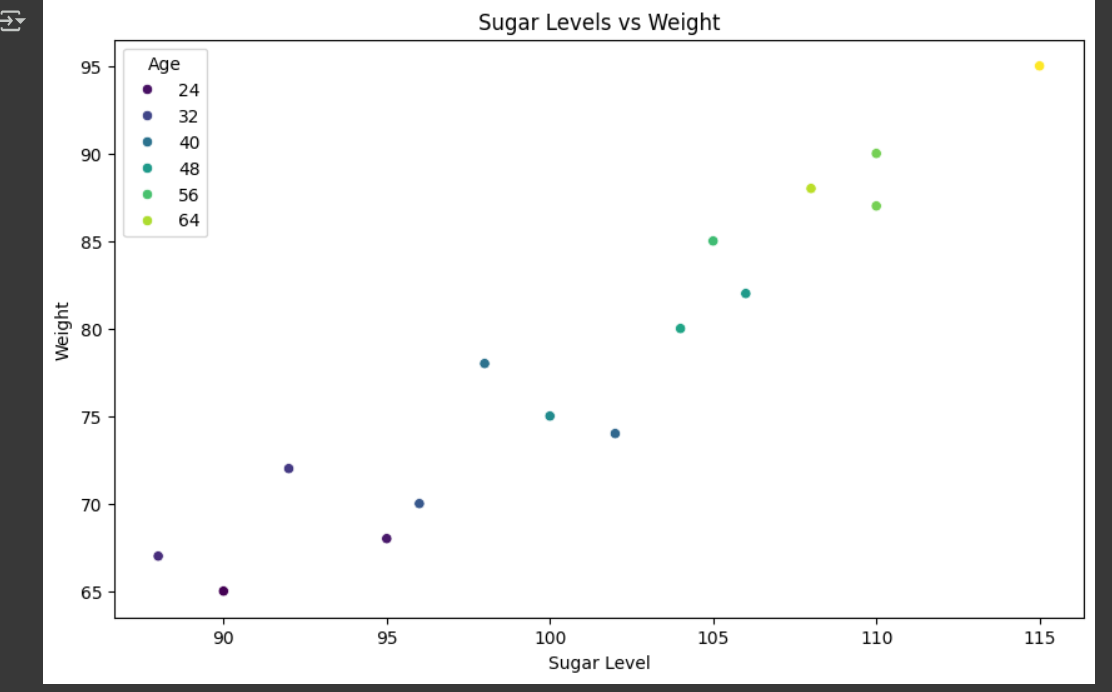
sns.scatterplot(data=df, x='SugarLevel', y='Weight', hue='Age', palette='viridis')

plt.title('Sugar Levels vs Weight')

plt.xlabel('Sugar Level')

plt.ylabel('Weight')

plt.show()



# Create an Age Group column (you can adjust the age ranges)

bins = [0, 20, 40, 60, 80, 100]

labels = ['0-20', '21-40', '41-60', '61-80', '81-100']

df['AgeGroup'] = pd.cut(df['Age'], bins=bins, labels=labels)

# Calculate the average blood pressure per age group

avg\_bp\_by\_age = df.groupby('AgeGroup')['BloodPressure'].mean()

# Plot the average blood pressure by age group

plt.figure(figsize=(8,5))

avg\_bp\_by\_age.plot(kind='bar', color='skyblue')

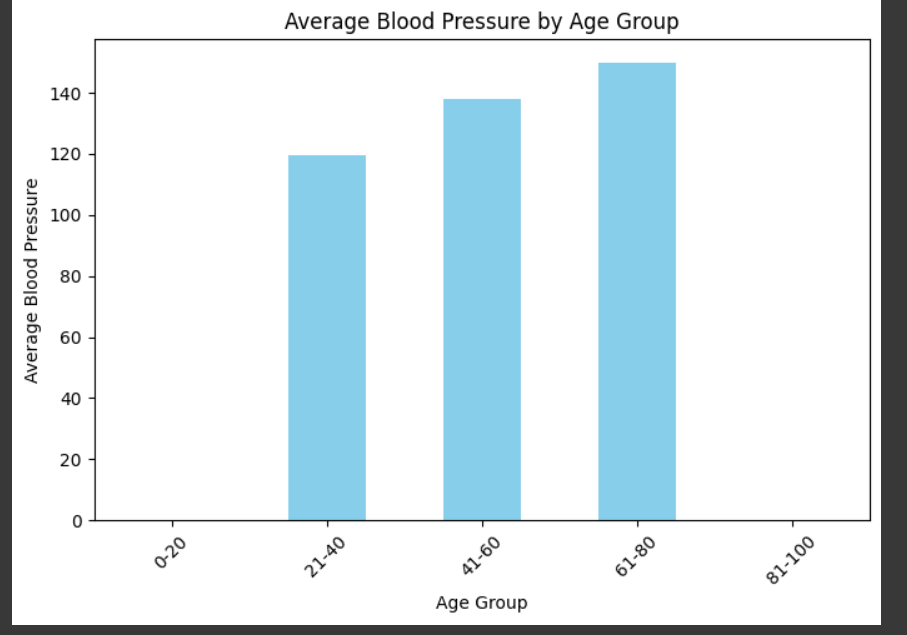
plt.title('Average Blood Pressure by Age Group')

plt.xlabel('Age Group')

plt.ylabel('Average Blood Pressure')

plt.xticks(rotation=45)

plt.show()



# Distribution of Blood Pressure

plt.figure(figsize=(10, 6))

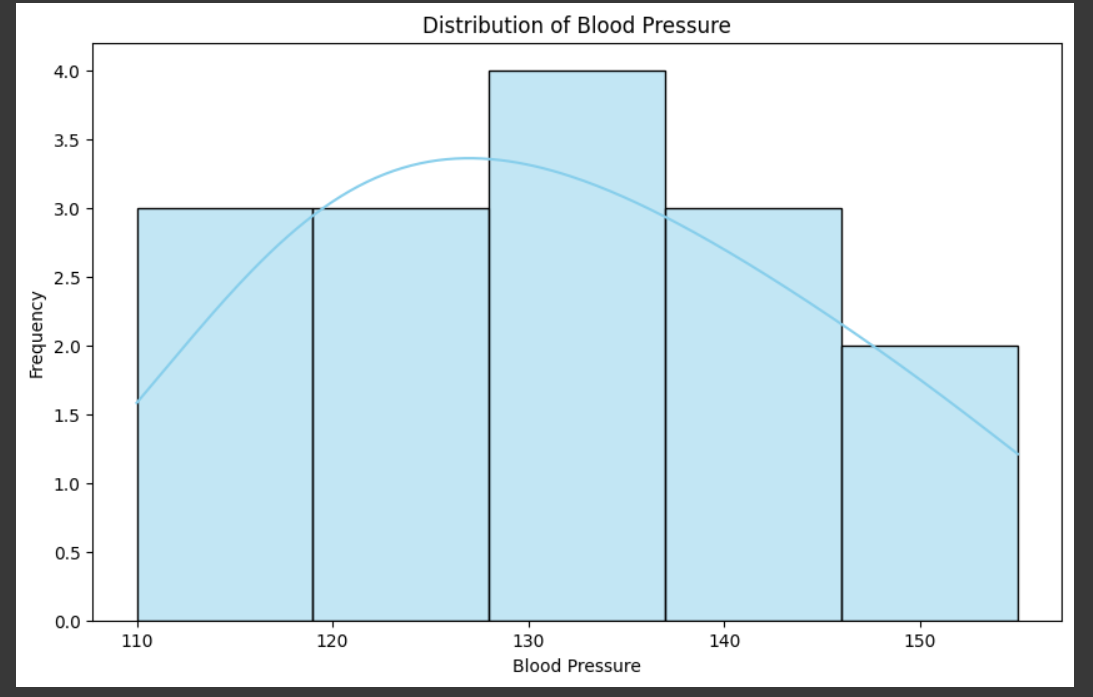
sns.histplot(df['BloodPressure'], kde=True, color='skyblue')

plt.title('Distribution of Blood Pressure')

plt.xlabel('Blood Pressure')

plt.ylabel('Frequency')

plt.show()



# Distribution of Weight

plt.figure(figsize=(10, 6))

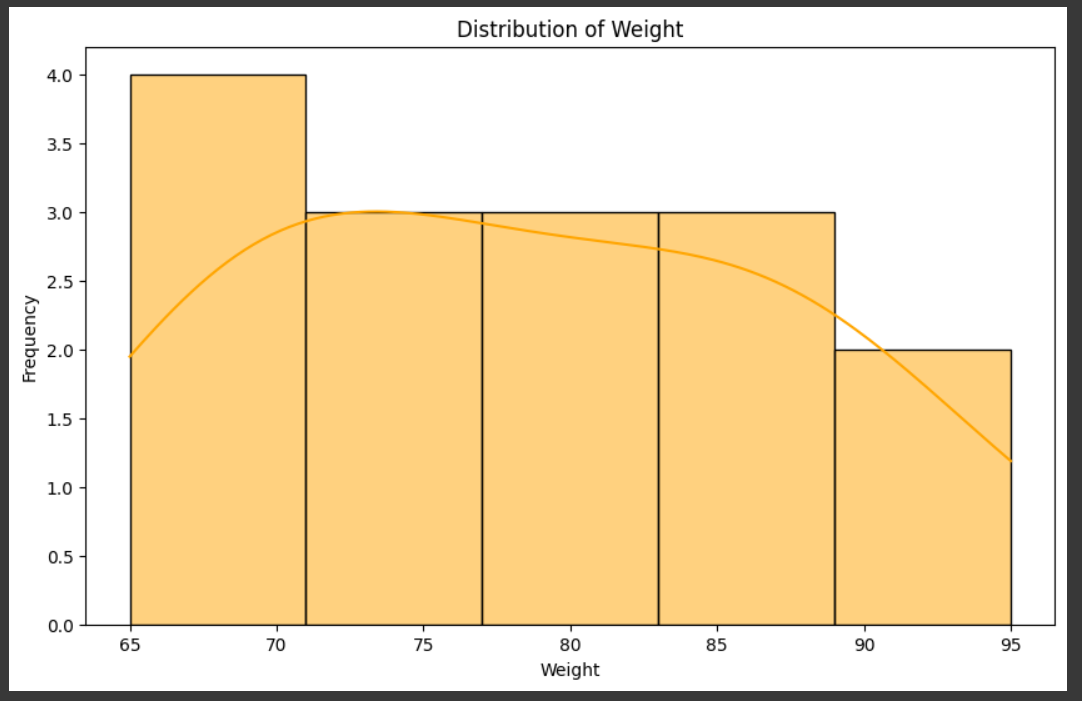
sns.histplot(df['Weight'], kde=True, color='orange')

plt.title('Distribution of Weight')

plt.xlabel('Weight')

plt.ylabel('Frequency')

plt.show()



# Make predictions using the test data

y\_pred = model.predict(X\_test)

# Calculate Mean Squared Error (MSE)

mse = mean\_squared\_error(y\_test, y\_pred)

print(f"Mean Squared Error of the model: {mse}")

# Plotting actual vs predicted blood pressure

plt.figure(figsize=(10, 6))

plt.scatter(y\_test, y\_pred, color='blue')

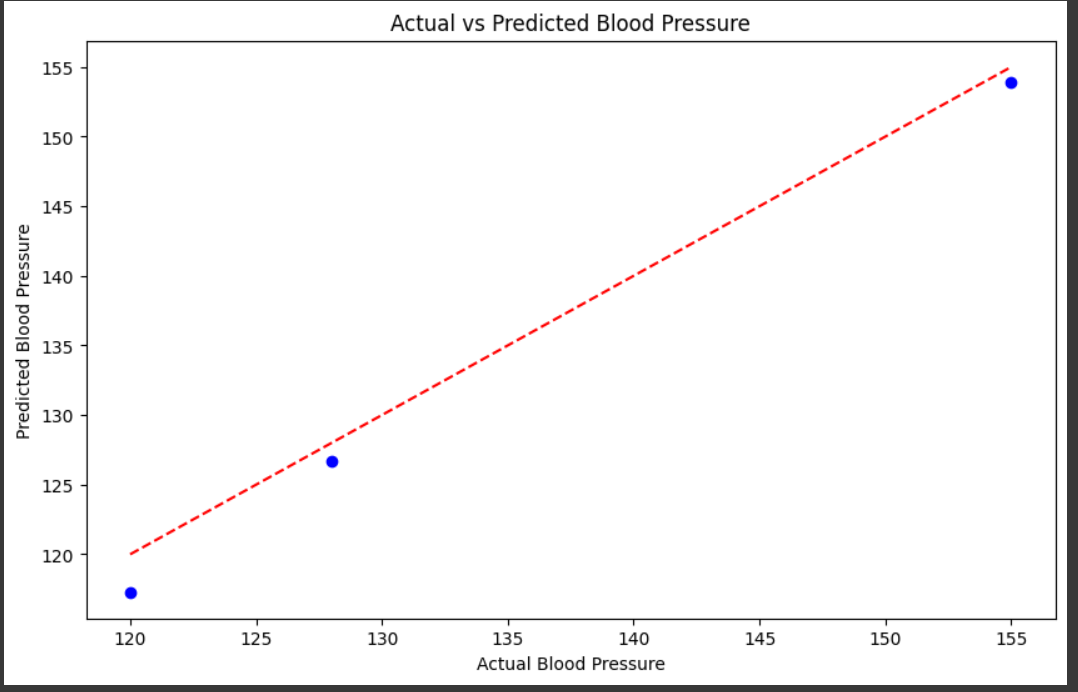
plt.plot([min(y\_test), max(y\_test)], [min(y\_test), max(y\_test)], color='red', linestyle='--')

plt.title('Actual vs Predicted Blood Pressure')

plt.xlabel('Actual Blood Pressure')

plt.ylabel('Predicted Blood Pressure')

plt.show()



1. References

**Pandas Documentation**: https://pandas.pydata.org/pandas-docs/stable/

**Matplotlib Documentation**: https://matplotlib.org/stable/contents.html