KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning

PROBLEM STATEMENT

Healthcare Data Exploration : Analyzing Blood Pressure, Sugar Levels, and Weight Trends

This report investigates the relationships between **blood pressure**, **blood sugar levels**, and **weight** in patients over time. The goal is to identify the health trends that may help in early detection of problems like hypertension, diabetes, and obesity. Through the analysis of patient data, the objective is to provide insights into how these parameters affects.

NAME – GAURI DHINGRA

BRANCH- CSEAIML

UNIVERSITY ROLL NO – 202401100400090

CLASS ROLL NO. 17

Introduction

Nowadays, there has been an increasing focus on preventive healthcare, where continuous monitoring of key health parameters can help identify potential risks at an early stage. Blood pressure, blood sugar levels, and weight are three crucial metrics that directly influence the overall health and well-being of individuals. For example:

Hypertension (high blood pressure) is a major risk factor for heart disease and stroke.

Diabetes, which is elevated blood sugar levels, can lead to serious long-term health complications.

Obesity is known to contribute to various diseases, including heart disease and diabetes.

Considering these facts, it is essential to identify patterns between these health indicators to improve the understanding between them. Monitoring these variables can help in early detection, personalized healthcare plans, and improving patient outcomes.

This report focuses on a dataset containing patient information, specifically their blood pressure (both systolic and diastolic), sugar levels (measured in mg/dL), and weight (measured in kg). By analyzing this dataset over a period of time, the objective is to identify trends that can provide valuable insights into how these health metrics evolve together.

Through data visualization, this report aims to highlight:

The trends in blood pressure and sugar levels over time.

The relationship between blood pressure and sugar levels.

The impact of weight fluctuations on blood pressure and sugar levels

Methodology

The approach to analyze the patient health data involves several key steps that include data preparation, data analysis, and visualization. Below are the specific techniques and tools used for this analysis:

Data Collection and Preparation:

Dataset Creation: A sample dataset was taken which represents patient data. The dataset consists of the following features:

Patient ID

Date

Blood Pressure (Systolic and Diastolic)

Sugar Level

Weight

This dataset was created for four patients over a period of three days. This simplified dataset was used for demonstrating the methodology and visualizations.

Data Preprocessing:

The **Date** column was converted into a datetime format to facilitate easier time-based analysis and plotting.

The dataset was stored as a Pandas DataFrame.

Analysis and Visualization:

Blood Pressure Trends:

Line plots were used to track **Systolic** and **Diastolic** blood pressure levels for each patient over time. This helped identify any fluctuations or patterns in blood pressure that might signal potential health issues.

Blood Pressure vs. Sugar Level:

A **scatter plot** was generated to examine the relationship between **Systolic Blood Pressure** and **Sugar Levels**. The goal was to explore whether higher blood pressure is associated with higher sugar levels, which is a common trend in conditions like hypertension and diabetes.

Blood Pressure Distribution:

A **box plot** was used to analyze the distribution of **Systolic Blood Pressure** for each patient. The box plot highlighted the median blood pressure values, the spread of the data, and any potential outliers for each patient.

Correlation Analysis:

A **correlation matrix** was generated to explore the relationships between **Systolic Blood Pressure**, **Diastolic Blood Pressure**, **Sugar Levels**, and **Weight**. This heatmap allowed us to visually examine the strength of correlations between these parameters, helping identify any significant associations.

Weight Change over Time:

A line plot was used to track the change in **Weight** for each patient over time. This allowed for the identification of trends such as weight gain or loss, which could have a direct impact on the patients' blood pressure and sugar levels.

Tools Used:

Python programming language

Pandas for data manipulation and analysis

Matplotlib for data visualization (line plots, scatter plots, box plots, and heatmaps)

NumPy for numerical calculations and handling arrays

CODE

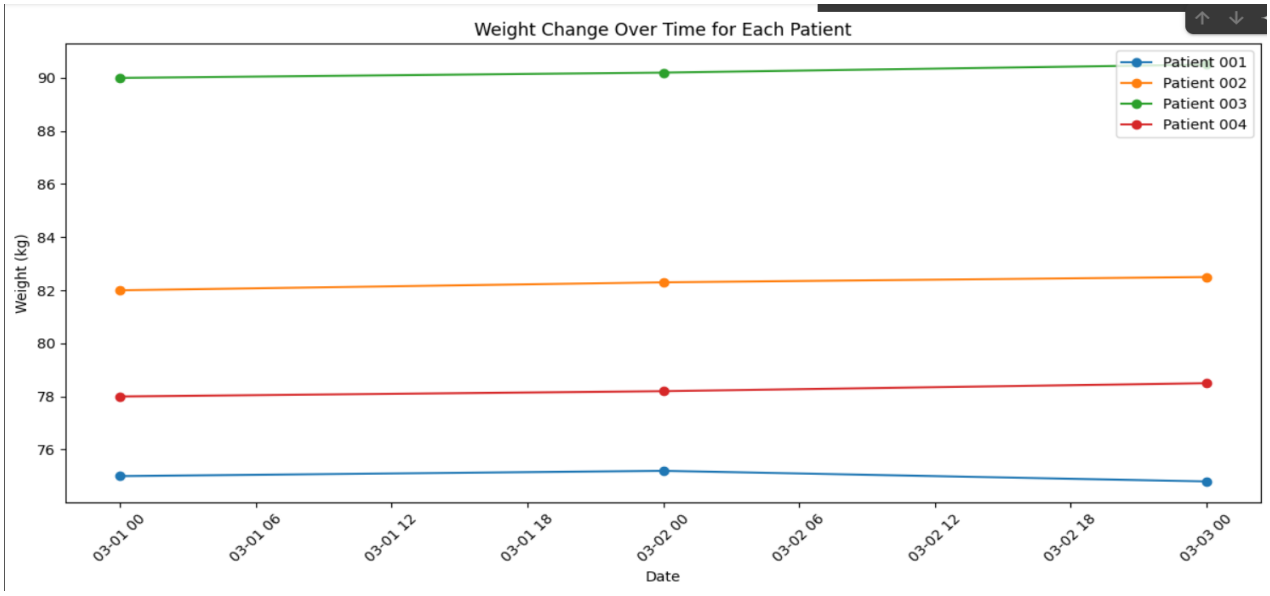
```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt


# Step 1: Create a sample dataset for patient data

# This dataset includes Patient ID, Date, Blood Pressure (Systolic and Diastolic), Sugar Level, and Weight

data = {

    'Patient ID': ['001', '001', '001', '002', '002', '002', '003', '003', '003', '004', '004', '004'],

    'Date': ['2025-03-01', '2025-03-02', '2025-03-03', '2025-03-01', '2025-03-02', '2025-03-03', '2025-03-01', '2025-03-02', '2025-03-03', '2025-03-01', '2025-03-02', '2025-03-03'],

    'Blood Pressure (Systolic)': [120, 122, 118, 135, 137, 130, 140, 142, 138, 125, 128, 123],

    'Blood Pressure (Diastolic)': [80, 82, 78, 85, 87, 85, 90, 92, 88, 80, 82, 80],

    'Sugar Level (mg/dL)': [95, 98, 92, 105, 110, 108, 115, 120, 118, 100, 105, 102],

    'Weight (kg)': [75, 75.2, 74.8, 82, 82.3, 82.5, 90, 90.2, 90.5, 78, 78.2, 78.5]

}


# Step 2: Convert the data into a Pandas DataFrame

df = pd.DataFrame(data)


# Step 3: Convert the 'Date' column to a datetime object for easier manipulation

df['Date'] = pd.to_datetime(df['Date'])


# Display the first few rows of the DataFrame

print(df)


# ---- Visualizations ----


# Step 4: Plot Blood Pressure (Systolic and Diastolic) over Time for Each Patient
```

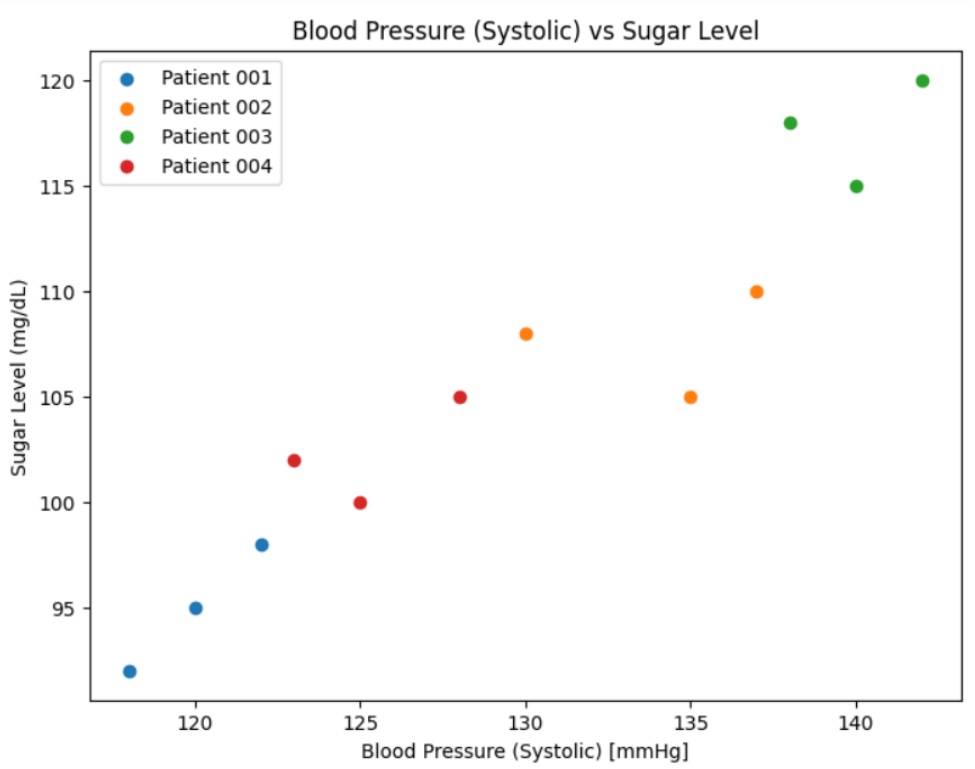
```
plt.figure(figsize=(12, 6))  # Create a figure with specific size
for patient in df['Patient ID'].unique():  # Loop through each unique patient ID
    patient_data = df[df['Patient ID'] == patient]  # Filter the dataset for this patient
    # Plot Systolic Blood Pressure over time
    plt.plot(patient_data['Date'], patient_data['Blood Pressure (Systolic)'], label=f'Patient {patient} - Systolic', marker='o')
    # Plot Diastolic Blood Pressure over time
    plt.plot(patient_data['Date'], patient_data['Blood Pressure (Diastolic)'], label=f'Patient {patient} - Diastolic', marker='x')

# Adding labels and title to the plot
plt.xlabel('Date')  # X-axis label
plt.ylabel('Blood Pressure (mmHg)')  # Y-axis label
plt.title('Blood Pressure (Systolic and Diastolic) over Time')  # Title of the plot
plt.legend()  # Show legend for different lines
plt.xticks(rotation=45)  # Rotate the x-axis labels for better readability
plt.tight_layout()  # Adjust layout to avoid label overlap
plt.show()  # Show the plot
# Step 5: Scatter plot for Blood Pressure vs. Sugar Level
plt.figure(figsize=(8, 6))  # Create a new figure
# Loop through each patient and plot their Blood Pressure vs Sugar Level
for patient in df['Patient ID'].unique():
    patient_data = df[df['Patient ID'] == patient]  # Filter data for the patient
    # Scatter plot: Systolic Blood Pressure vs Sugar Level
    plt.scatter(patient_data['Blood Pressure (Systolic)'], patient_data['Sugar Level (mg/dL)'], label=f'Patient {patient}')

# Adding labels and title to the plot
plt.xlabel('Blood Pressure (Systolic) [mmHg]')  # X-axis label
plt.ylabel('Sugar Level (mg/dL)')  # Y-axis label
plt.title('Blood Pressure (Systolic) vs Sugar Level')  # Title of the plot
plt.legend()  # Show legend to identify patients
```

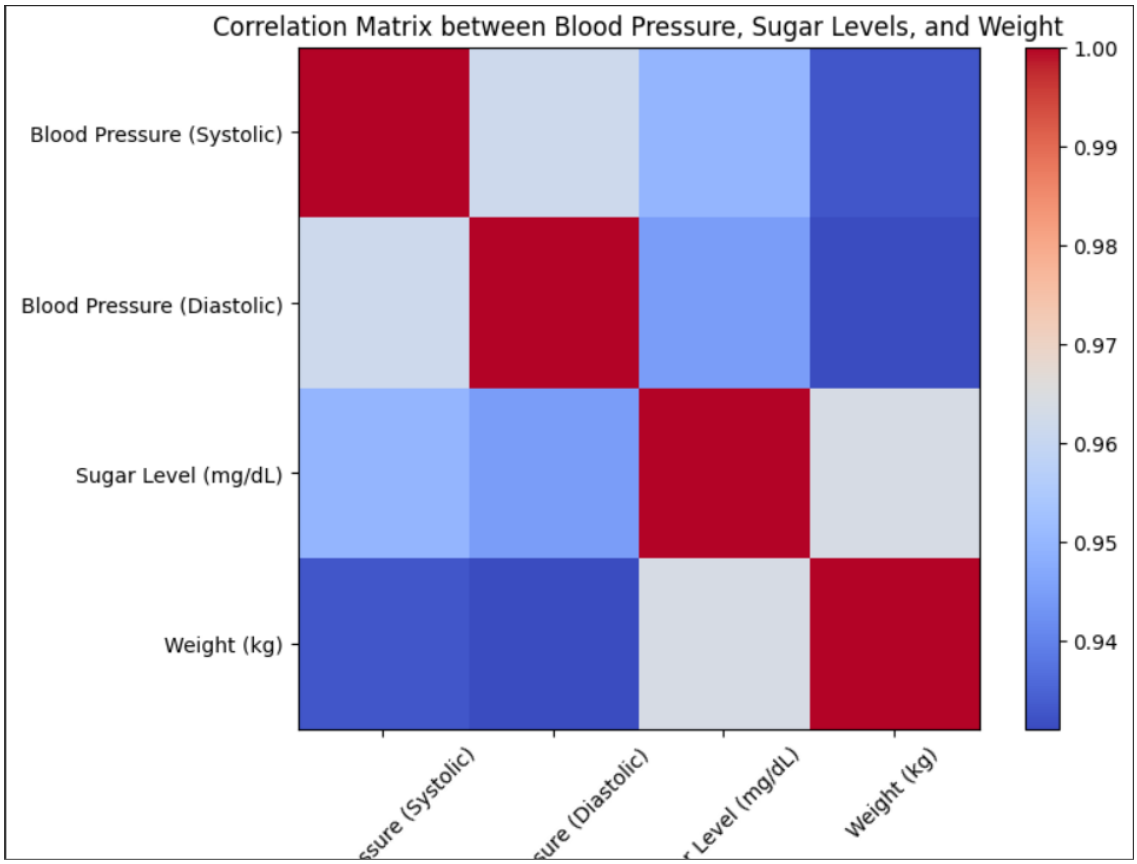
```
plt.show()  # Show the plot


# Step 6: Box plot for Blood Pressure Distribution (Systolic) for Each Patient
plt.figure(figsize=(8, 6))  # Create a new figure for the box plot
# Box plot for Blood Pressure (Systolic) by Patient ID
plt.boxplot([df[df['Patient ID'] == patient]['Blood Pressure (Systolic)'] for patient in df['Patient ID'].unique()],
        labels=df['Patient ID'].unique())  # Each patient gets a box plot
# Adding labels and title to the plot
plt.xlabel('Patient ID')  # X-axis label
plt.ylabel('Blood Pressure (Systolic) [mmHg]')  # Y-axis label
plt.title('Blood Pressure (Systolic) Distribution by Patient')  # Title of the plot
plt.show()  # Show the plot


# Step 7: Correlation Heatmap between Blood Pressure, Sugar Levels, and Weight
correlation_matrix = df[['Blood Pressure (Systolic)', 'Blood Pressure (Diastolic)', 'Sugar Level (mg/dL)', 'Weight (kg)']].corr()


# Create a heatmap to visualize the correlation between these features
plt.figure(figsize=(8, 6))  # Create a new figure for the heatmap
plt.imshow(correlation_matrix, cmap='coolwarm', interpolation='none')  # Heatmap display
plt.colorbar()  # Add a color bar to show correlation scale
# Set labels for the x and y axes based on the correlation matrix columns
plt.xticks(np.arange(4), correlation_matrix.columns, rotation=45)
plt.yticks(np.arange(4), correlation_matrix.columns)
plt.title('Correlation Matrix between Blood Pressure, Sugar Levels, and Weight')  # Title of the plot
plt.show()  # Show the heatmap


# Step 8: Weight Change Over Time for Each Patient
plt.figure(figsize=(12, 6))  # Create a figure with specific size
for patient in df['Patient ID'].unique():  # Loop through each unique patient ID
    patient_data = df[df['Patient ID'] == patient]  # Filter data for this patient
```

```
    # Plot Weight over time for each patient
    plt.plot(patient_data['Date'], patient_data['Weight (kg)'], label=f'Patient {patient}', marker='o')

# Adding labels and title to the plot
plt.xlabel('Date')  # X-axis label
plt.ylabel('Weight (kg)')  # Y-axis label
plt.title('Weight Change Over Time for Each Patient')  # Title of the plot
plt.legend()  # Show legend for different patients
plt.xticks(rotation=45)  # Rotate x-axis labels
plt.tight_layout()  # Adjust layout
plt.show()  # Show the plot
```

	Patient ID	Date	Blood Pressure (Systolic)	Blood Pressure (Diastolic)	Sugar Level (mg/dL)	Weight (kg)
0	001	2025-03-01	120	80	95	75.0
1	001	2025-03-02	122	82	98	75.2
2	001	2025-03-03	118	78	92	74.8
3	002	2025-03-01	135	85	105	82.0
4	002	2025-03-02	137	87	110	82.3
5	002	2025-03-03	130	85	108	82.5
6	003	2025-03-01	140	90	115	90.0
7	003	2025-03-02	142	92	120	90.2
8	003	2025-03-03	138	88	118	90.5
9	004	2025-03-01	125	80	100	78.0
10	004	2025-03-02	128	82	105	78.2
11	004	2025-03-03	123	80	102	78.5




Blood Pressure (Systolic) vs Sugar Level
Patient 001
Patient 002
Patient 003
Patient 004
Sugar Level (mg/dL)
120
115
110
105
100
95
120
125
130
135
140
Blood Pressure (Systolic) [mmHg]


Correlation Matrix between Blood Pressure, Sugar Levels, and Weight
Blood Pressure (Systolic)
Blood Pressure (Diastolic)
Sugar Level (mg/dL)
Weight (kg)
1.00
0.99
0.98
0.97
0.96
0.95
0.94
Weight (kg)

REFERENCES-

1. Dataset from google
2. NumPy Documentation
3. Pandas Documentation
4. Matplotlib Documentation