

Healthcare Data Exploration: Visualizing Patient Data for Health Trend Analysis

Date: March 10, 2025

Author: Sakshi Kumari

University Roll No. : 202401100400163

Institution: KIET Group of Institutions

Course: Introduction to AI

Introduction

Healthcare data is a rich source of insights, providing valuable information on patient health, treatment effectiveness, and disease patterns. In this report, we aim to explore key patient health metrics such as **blood pressure**, **blood sugar levels**, and **weight** to identify any significant health trends. By visualizing these data points, we can gain a better understanding of how health parameters fluctuate over time, as well as detect correlations or patterns that could indicate areas of concern for healthcare providers.

With the increasing availability of electronic health records (HER) and wearable health devices, analysing this data has become easier and more crucial for early detection and prevention of various medical conditions. The goal of this project is to demonstrate how visual analytics tools can help identify trends, patterns, and anomalies in patient health data.

Methodology

To effectively analyze and visualize the health data, we followed a systematic approach. Below are the key steps in our methodology:

1. **Data Collection:**

- A sample dataset of patient health records was created, including key metrics such as **blood pressure**, **blood sugar levels**, and **weight**.
- Data was organized into a table with patient IDs, dates of measurements, and the corresponding health metrics.

2. **Data Preprocessing:**

- The dataset was cleaned by removing any missing or erroneous data points.
- Time-series data was sorted by date to ensure that measurements were in chronological order.
- Any outliers or extreme values that could distort the analysis were identified and marked for further investigation.

3. **Visualization:**

- A variety of visualization techniques were used to analyze the data:
 - **Line charts** to track changes over time.
 - **Scatter plots** to observe correlations between different health metrics.
 - **Histograms** to analyze the distribution of specific health measurements.
 - **Heatmaps** to visualize correlations between all health metrics.
- These visualizations were created using **Python** with the libraries **Matplotlib**, **Seaborn**, and **Plotly** for interactive charts.

4. **Analysis:**

- Patterns in the data were identified through visual inspections of trends, distributions, and correlations.
- The insights gained were used to formulate conclusions about the patients' health trends, including identifying possible correlations (e.g., between weight and blood pressure).

5. **Interpretation:**

- After visualizing the data, conclusions were drawn regarding the relationships between health parameters, helping to understand potential causes of changes in patient health over time.

4. Code Typed

```
# Importing necessary libraries
import matplotlib.pyplot as plt
```

```
import seaborn as sns
import pandas as pd
import plotly.express as px

# Sample dataset
data = {
    'Date': pd.date_range('2022-01-01', periods=6, freq='M'),
    'Blood Pressure': [120, 122, 118, 125, 130, 128],
    'Blood Sugar': [90, 92, 95, 91, 89, 85],
    'Weight': [70, 72, 73, 75, 76, 77]
}

df = pd.DataFrame(data)

# Line Chart: Health Trends Over Time
plt.figure(figsize=(10, 6))

# Plot Blood Pressure, Blood Sugar, and Weight
plt.plot(df['Date'], df['Blood Pressure'], label='Blood Pressure', marker='o')
plt.plot(df['Date'], df['Blood Sugar'], label='Blood Sugar', marker='o')
plt.plot(df['Date'], df['Weight'], label='Weight', marker='o')

plt.title('Health Trends Over Time')
plt.xlabel('Date')
plt.ylabel('Measurements')
plt.legend()
plt.xticks(rotation=45)
plt.show()

# Scatter Plot: Weight vs Blood Pressure
sns.scatterplot(data=df, x='Weight', y='Blood Pressure')
plt.title('Weight vs Blood Pressure')
```

```
plt.xlabel('Weight (kg)')
plt.ylabel('Blood Pressure (mmHg)')
plt.show()

# Histogram: Distribution of Blood Pressure
plt.hist(df['Blood Pressure'], bins=5, edgecolor='black')
plt.title('Distribution of Blood Pressure')
plt.xlabel('Blood Pressure (mmHg)')
plt.ylabel('Frequency')
plt.show()

# Correlation Heatmap: Health Metrics Correlation
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix of Health Metrics')
plt.show()

# Interactive Line Chart with Plotly
fig = px.line(df, x='Date', y=['Blood Pressure', 'Blood Sugar', 'Weight'],
              title='Health Trends Over Time',
              labels={'value': 'Measurements', 'variable': 'Metrics'})
fig.show()
```

Explanation of Code:

- **Matplotlib and Seaborn** are used to create static visualizations like line charts, scatter plots, and histograms.
- **Plotly** is used to generate interactive charts, providing a more engaging user experience.

- The data includes three health metrics: **Blood Pressure**, **Blood Sugar**, and **Weight**, which are tracked over six months.
- The **line chart** shows trends for each health metric over time.
- The **scatter plot** explores the correlation between **weight** and **blood pressure**.
- The **histogram** visualizes the distribution of blood pressure values.
- The **heatmap** provides insights into the correlations between different health metrics.

Screenshots:



