

TOPIC : Healthcare Data Exploration

NAME : Harsh Kumar Mishra

UNIV. ROLL NO. : 202401100300114

BRANCH : CSE-AI

SECTION : B

SUBJECT : INTRODUCTION TO AI

SUBJECT CODE : AI101B

# Introduction

In the modern healthcare industry, data analysis plays a crucial role in monitoring patients' health and identifying potential risks. This project focuses on analyzing healthcare data, including attributes such as Patient ID, Age, Blood Pressure, Sugar Level, and Weight. The goal is to compute basic statistics, identify patients at risk based on abnormal readings, and visualize the data for better insights. This report outlines the methodology used, provides the full Python code, and includes screenshots of the output.

## Methodology

1. **Data Loading:** A dataset containing 20 sample patients was loaded using the Pandas library. The dataset was extracted from a CSV file and structured into a DataFrame.
2. **Statistical Analysis:** The `describe()` function was used to compute summary statistics, including mean, standard deviation, minimum, and maximum values for each attribute.
3. **Risk Identification:** Patients with blood pressure exceeding 140 or sugar levels above 180 were flagged as high-risk individuals.
4. **Data Visualization:** Two histograms were plotted using Matplotlib to represent the distributions of blood pressure and sugar levels across the sample population.

## Code Typed

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
def load_data():
```

```
    """Load healthcare data from a predefined dataset."""
```

```
    #So here below we have the data of 20 sample patients with attributes Patient ID, Age, Bloodpressure, sugar level, Weight, this data is extracted from CSV file
```

```
    data = {
```

```
        "PatientID": [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20],
```

```
        "Age": [44, 39, 49, 58, 35, 25, 46, 28, 60, 55, 41, 48, 58, 35, 67, 70, 43, 74, 19, 56],
```

```
        "BloodPressure": [118, 109, 149, 121, 109, 129, 132, 93, 145, 125, 143, 141, 93, 145, 176, 109, 148, 122, 147, 119],
```

```
        "SugarLevel": [87.89, 177.32, 144.15, 90.36, 126.42, 95.27, 146.61, 109.75, 103.19, 197.73, 180.58, 181.97, 181.78, 133.39, 87.01, 193.27, 135.94, 129.41, 125.48, 160.72],
```

```
        "Weight": [105.57, 105.70, 77.79, 115.24, 70.38, 119.05, 62.18, 81.79, 94.64, 118.59, 103.58, 61.45, 50.68, 113.19, 84.94, 77.72, 106.58, 83.30, 74.08, 111.87]
```

```
    }
```

```
    return pd.DataFrame(data)
```

```
def display_statistics(df):
```

```
    #Display basic statistics of the healthcare data
```

```
    print("\nBasic Statistics:")
```

```
    print(df.describe())
```

```
def identify_risk_patients(df):
```

```
    #Identify abnormal readings which indicates patients at risk
```

```
    risk_patients = df[(df['BloodPressure'] > 140) | (df['SugarLevel'] > 180)]
```

```
    print("\nPatients at Risk:")
```

```
print(risk_patients)
```

```
def plot_data(df):
```

```
    #Plot data for better visualization
```

```
    plt.figure(figsize=(10, 5))
```

```
    plt.subplot(1, 2, 1)
```

```
    plt.hist(df['BloodPressure'], bins=10, color='skyblue', edgecolor='black')
```

```
    plt.xlabel('Blood Pressure')
```

```
    plt.ylabel('Frequency')
```

```
    plt.title('Blood Pressure Distribution')
```

```
    plt.subplot(1, 2, 2)
```

```
    plt.hist(df['SugarLevel'], bins=10, color='salmon', edgecolor='black')
```

```
    plt.xlabel('Sugar Level')
```

```
    plt.ylabel('Frequency')
```

```
    plt.title('Sugar Level Distribution')
```

```
    plt.tight_layout()
```

```
    plt.show()
```

```
def main():
```

```
    df = load_data()
```

```
    display_statistics(df)
```

```
    identify_risk_patients(df)
```

```
    plot_data(df)
```

```
if __name__ == "__main__":
```

```
    main()
```

## Screenshots Output

### 1. Basic Statistics Output:

Basic Statistics:					
	PatientID	Age	BloodPressure	SugarLevel	Weight
count	20.00000	20.000000	20.000000	20.000000	20.00000
mean	10.50000	47.500000	128.650000	139.412000	90.91600
std	5.91608	14.968388	20.893905	37.010835	21.12455
min	1.00000	19.000000	93.000000	87.010000	50.68000
25%	5.75000	38.000000	115.750000	108.110000	76.81000
50%	10.50000	47.000000	127.000000	134.665000	89.79000
75%	15.25000	58.000000	145.000000	178.135000	107.90250
max	20.00000	74.000000	176.000000	197.730000	119.05000

### 2. Patients at Risk Output:

Patients at Risk:					
	PatientID	Age	BloodPressure	SugarLevel	Weight
2	3	49	149	144.15	77.79
8	9	60	145	103.19	94.64
9	10	55	125	197.73	118.59
10	11	41	143	180.58	103.58
11	12	48	141	181.97	61.45
12	13	58	93	181.78	50.68
13	14	35	145	133.39	113.19
14	15	67	176	87.01	84.94
15	16	70	109	193.27	77.72
16	17	43	148	135.94	106.58
18	19	19	147	125.48	74.08

### 3. Blood Pressure and Sugar Level Distribution:

