

Health Data Analysis and Visualization Report

Tools Used: Python, Pandas, Matplotlib, Seaborn

Dataset: Health Data

1. Introduction

Health data analysis plays a crucial role in understanding key health indicators and identifying trends that can guide preventive healthcare measures. This project focuses on analyzing various health parameters, including **age, blood pressure, heart rate, and oxygen saturation**, to categorize individuals into relevant health risk groups and visualize trends among different demographics.

The objectives of this analysis include:

- Categorizing **age, blood pressure, heart rate, and oxygen saturation** to better interpret health risks.
 - Visualizing the distribution of different health metrics.
 - Comparing health indicators across different demographic groups, particularly by age and gender.
-

2. Data Processing and Feature Engineering

2.1 Data Cleaning and Transformation

The dataset included a combined blood pressure reading in the format "Systolic/Diastolic." This was split into two separate columns, **Systolic Blood Pressure (SBP)** and **Diastolic Blood Pressure (DBP)**, to enable better analysis.

2.2 Categorization of Health Metrics

To enhance interpretability, the numerical health metrics were classified into meaningful categories:

Age Groups

- **Young:** Individuals aged **35 years or below**.
- **Middle-aged:** Individuals aged **between 36 and 55 years**.
- **Senior:** Individuals aged **above 55 years**.

Blood Pressure Categories

- **Normal:** Systolic <120 mmHg and Diastolic <80 mmHg.
- **Elevated:** Systolic between **120–139 mmHg** or Diastolic between **80–89 mmHg**.
- **Hypertension Stage 1:** Systolic between **140–159 mmHg** or Diastolic between **90–99 mmHg**.
- **Hypertension Stage 2:** Systolic **≥160 mmHg** or Diastolic **≥100 mmHg**.

Heart Rate Categories

- **Low:** Below **60 bpm**.
- **Normal:** Between **60–100 bpm**.
- **High:** Above **100 bpm**.

Oxygen Saturation Categories

- **Low:** Below **94%**.
- **Normal:** **94% and above**.

These classifications provided a structured way to interpret different health conditions and compare trends across groups.

3. Data Visualization and Insights

3.1 Distribution of Health Metrics

Several visualizations were created to examine the distribution of health indicators.

- **Age Distribution:** The dataset showed that a significant portion of individuals were **middle-aged**, followed by **seniors** and **young adults**.
- **Blood Pressure Trends:** Many individuals exhibited **elevated blood pressure levels**, with a considerable number falling into **hypertension stages 1 and 2**.
- **Heart Rate Patterns:** Most individuals had **normal heart rates**, though some exhibited **high heart rates**, indicating potential risks.
- **Oxygen Saturation Levels:** A small percentage of individuals had **low oxygen saturation**, which could indicate underlying respiratory concerns.

3.2 Health Category Distributions

Count-based visualizations revealed the following patterns:

- **Age Group Distribution:** Middle-aged individuals formed the largest group, with fewer young and senior participants.
- **Blood Pressure Classification:** A significant portion of the dataset showed **hypertension**, suggesting a widespread health concern.
- **Heart Rate Trends by Gender:** While most individuals had normal heart rates, **males had slightly higher instances of high heart rates**.

- **Oxygen Saturation by Gender:** The majority of individuals had **normal oxygen saturation levels**, but a few cases of low oxygen were observed.
-

4. Key Findings and Observations

4.1 Age-Based Health Trends

- **Middle-aged and senior individuals** showed a higher prevalence of **hypertension**, emphasizing the importance of regular monitoring.
- **Young individuals** generally had normal blood pressure but some exhibited **elevated levels**, suggesting lifestyle factors at play.

4.2 Blood Pressure Patterns

- Many individuals exhibited **hypertension (stages 1 and 2)**, which is a major health concern linked to cardiovascular diseases.
- Very few individuals had **normal blood pressure**, indicating that lifestyle changes or medical intervention might be needed.

4.3 Heart Rate and Oxygen Saturation Insights

- **High heart rates** were observed in some individuals, with a slightly higher prevalence among males.
 - Most individuals maintained **healthy oxygen levels**, but a small subset had **low oxygen saturation (<94%)**, which could indicate potential respiratory issues.
-

5. Conclusion and Recommendations

5.1 Conclusion

This analysis provided a detailed examination of **age, blood pressure, heart rate, and oxygen saturation** trends. The categorization of these metrics helped in identifying risk factors across different demographic groups.

5.2 Recommendations

- **Hypertension Prevention:** A high number of individuals fell into the hypertensive category, emphasizing the need for **regular blood pressure monitoring, dietary changes, and lifestyle modifications**.
- **Cardiac Health Awareness:** Since some individuals had **high heart rates**, promoting **regular physical activity and stress management techniques** could help maintain heart health.

- **Respiratory Health Monitoring:** Individuals with **low oxygen saturation** should be encouraged to seek **medical consultation**, as this could be an early indicator of respiratory conditions.
-

6. Future Scope

This project can be extended by:

- Incorporating **machine learning models** to predict health risks based on patterns in the dataset.
- Expanding the dataset to include **additional health indicators** like BMI, cholesterol levels, and lifestyle factors.
- Conducting **comparative analysis across regions** to understand geographical trends in health conditions.

This project highlights the importance of health data analysis in identifying key health risks and promoting preventive healthcare measures. The insights gained can aid in better decision-making for individuals and healthcare professionals alike. 🚀