REPORT AND ANALYSIS

Statistical Exploration of Health Indicators

Vital Signs Diagnosis

Prepared By

Brylle Matthew A. Lupac

Naomi Christienne Tiama

1. Introduction

The dataset titled "Vital Signs and Lifestyle Health Data" provides detailed health-related information for 26 individuals, each represented by a unique patient ID. The dataset includes key clinical measurements and lifestyle indicators that are commonly monitored in primary health care and preventive medicine. Core demographic attributes such as age and sex are provided alongside anthropometric data like weight (in kilograms), height (in centimeters), and BMI (Body Mass Index), which is a derived indicator of body fat. Cardiovascular health is assessed using systolic and diastolic blood pressure values, which are also categorized under a hypertension level legend. The heart rate column measures beats per minute, offering insight into cardiac function. Additionally, lifestyle and behavioral risk factors are included, such as smoking status, physical activity (in hours per week), and stress level, each with corresponding legend labels that help classify severity or frequency. Daily sleep hours are also tracked as a wellness indicator. Biochemical markers include glucose and cholesterol levels, both measured in mg/dL, which are critical for detecting metabolic risks like diabetes or cardiovascular disease. The dataset further includes a qualitative classification called "Elevated Risk", which summarizes the overall health risk category for each individual (e.g., Pre-Diabetes, Type 2 Diabetes, Normal/No Major Risk, etc.). Some entries also contain fields for medication status, though this variable is sparsely populated. Overall, this dataset is highly valuable for analyzing relationships between physiological measurements, lifestyle behaviors, and chronic disease risk—making it ideal for health data science projects focused on early detection, risk stratification, and patient profiling.

2. Methods

2.1. Data Cleaning and Preparation

- To ensure appropriate data types, numerical variables such as Age, Weight_kg, Height_cm, BMI, Systolic_BP, Diastolic_BP, Heart_rate, Glucose_mg.dL, Cholesterol_mg.dL, and Daily_Sleeping_hours were explicitly converted to numeric format using the mutate() function. The variable Sex was converted into a factor with labeled levels ("Female" for 0 and "Male" for 1) to enhance interpretability in plots.
- Additionally, a new variable Sleep_Hours_Rounded was created by rounding the Daily_Sleeping_hours to the nearest whole number using the round() function, for simplified categorical plotting.

2.2. Descriptive Statistical Analysis

- Descriptive statistics were generated using the psych::describe() function, summarizing central tendencies (mean, median), dispersion (standard deviation), and distribution properties (min, max, skew, kurtosis) of key continuous variables. These included age, anthropometric measures (weight, height, BMI), vital signs (systolic and diastolic blood pressure, heart rate), glucose, cholesterol, and daily sleeping hours.

2.3. Data Visualization

- **Boxplot:** A boxplot was created to compare Body Mass Index (BMI) between male and female patients. The plot utilized pastel colors for grouping and included clear axis labels and subtitles.

- **Histogram:** A histogram was generated to visualize the distribution of cholesterol levels across patients. It included 15 bins and was colored in dark orange with a black border for contrast.
- **Bar Plot:** A bar plot was created to display the frequency of rounded daily sleeping hours. The bars were colored sky blue and grouped by discrete sleep hour categories.
- **Scatter Plot:** A scatter plot was used to investigate the relationship between Age and Heart Rate. The plot included a fitted linear regression line with a 95% confidence interval using geom_smooth(method = "lm"), and point markers were styled with semi-transparency and larger size for clarity.

2.4. Advanced Statistical Test

- To assess the strength and direction of the linear relationship between systolic and diastolic blood pressure, a Pearson correlation test was conducted using the cor.test() function. The test provided a correlation coefficient (r), 95% confidence interval, and a p-value to evaluate statistical significance. The analysis helped determine whether higher systolic values were linearly associated with diastolic pressure among the sample population.

3. Results and Discussion

3.1. Summary Statistics

Table 1. Descriptive Statistics for Selected Variables

Variable	Mean	Median	Std. Dev	Min	Max
Age (years)	55.22	56.00	20.50	18	90.00
Weight (kg)	62.93	63.00	9.15	42	87.00
Height (cm)	152.95	152.50	11.77	130	175.00
BMI	27.25	26.58	5.39	0	46.75
Systolic BP (mm Hg)	134.02	134.00	10.28	97	161.00
Diastolic BP (mm Hg)	82.32	83.00	7.91	56	102.00

Heart Rate (bpm)	101.77	101.00	19.03	50	154.00
Glucose (mg/dL)	130.11	129.00	23.55	16	205.00
Cholesterol (mg/dL)	191.85	192.00	32.45	111	287.00
Daily Sleep (hours)	5.31	5.00	1.20	4	9.00

- The average age of patients was 55 years, ranging from 18 to 90 years.
- Weight averaged 63 kg, with a maximum of 87 kg, while height averaged 153 cm.
- The mean BMI was 27.25, but values ranged from 0 to 46.75, suggesting some extreme or possibly incorrect entries.
- The average systolic and diastolic blood pressures were 134 mm Hg and 82 mm Hg, respectively, indicating a generally elevated blood pressure trend.
- The mean heart rate was 102 bpm, with a wide range from 50 to 154 bpm, suggesting varied cardiovascular conditions.
- Glucose and cholesterol levels averaged 130 mg/dL and 192 mg/dL, respectively, with high maximum values, indicating possible hyperglycemia or hypercholesterolemia in some patients.
- On average, patients reported 5.3 hours of sleep per day, which is below the recommended amount for adults.

3.1. Data Visualizations

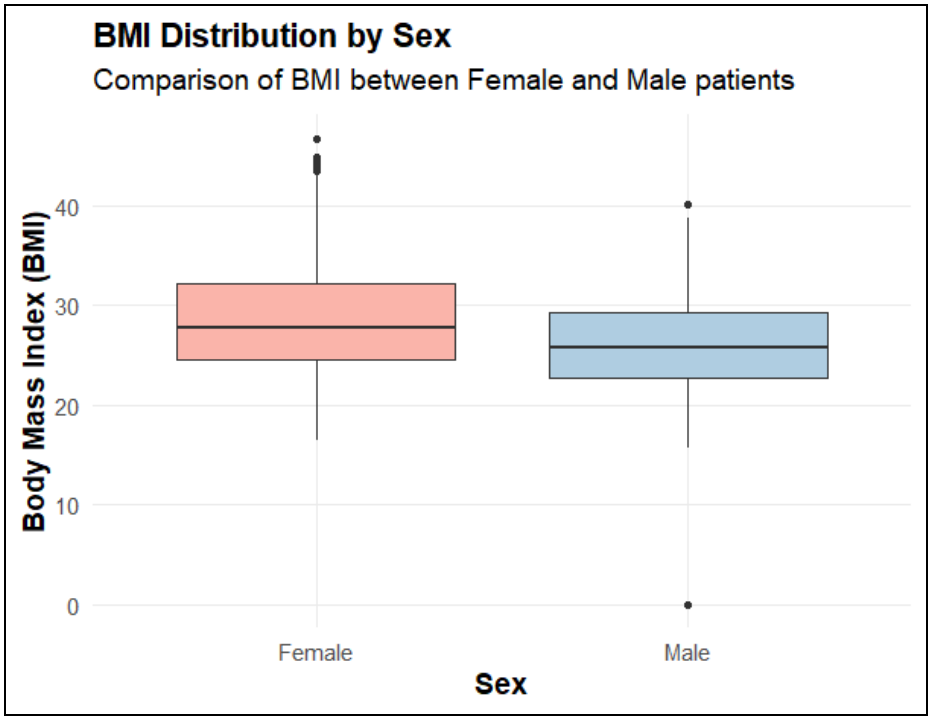

Figure 1. Boxplot: BMI Distribution by Sex

- The boxplot shows that both males and females have a close similar median BMI values.
- Females tend to have a slightly wider spread of BMI values compared to males.
- There are several outliers in both groups, especially among females, indicating individuals with unusually high or low BMI.
- Overall, the distribution suggests that BMI varies across sexes, but there is no extreme difference in the central tendency.

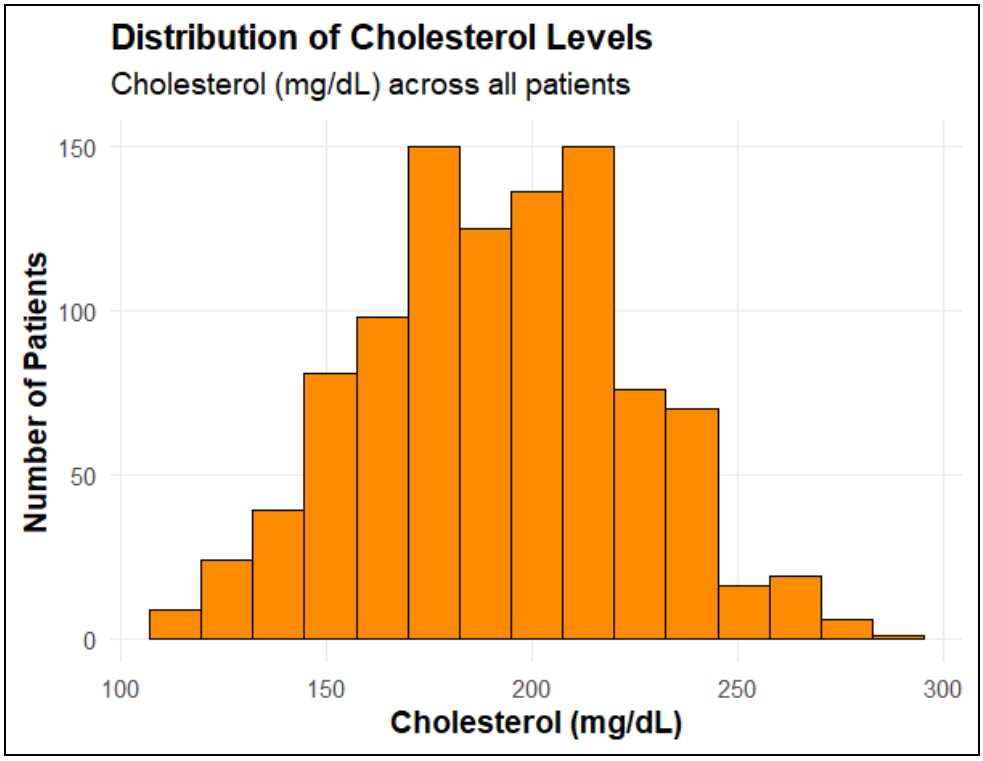

Figure 2. Histogram: Cholesterol Distribution

- The histogram shows that most individuals have cholesterol levels between 150 and 220 mg/dL, with the highest frequencies around 175 to 200 mg/dL.
- The distribution appears slightly right-skewed, indicating a few individuals with higher cholesterol levels above 250 mg/dL.
- Very low cholesterol values (below 130 mg/dL) are rare, while extremely high values above 270 mg/dL are uncommon but present.
- This pattern suggests that most patients fall within a normal to moderately elevated cholesterol range, with fewer extreme outliers.

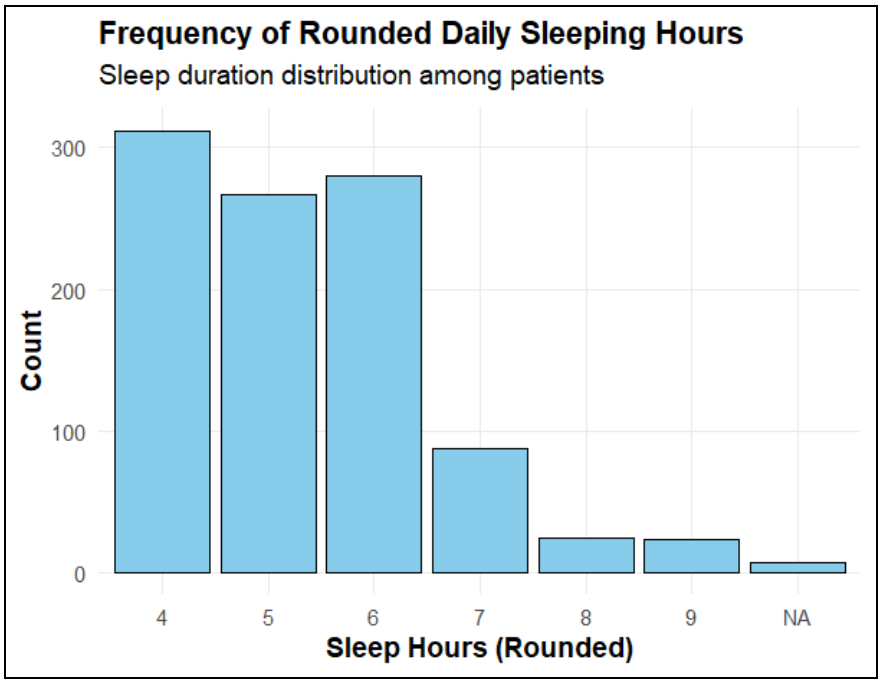

Figure 3. Bar Plot: Rounded Sleep Hours Distribution

- The bar plot shows that most individuals sleep between 4 to 6 hours per day, with 4 hours being the most common (312 individuals).
- Fewer individuals report getting 7 hours (87), 8 hours (24), or 9 hours (23) of sleep.
- This indicates a general trend of insufficient sleep, as the majority of patients sleep below the recommended 7–9 hours per night.
- The distribution suggests a potential concern for chronic sleep deprivation, which may negatively affect overall health.

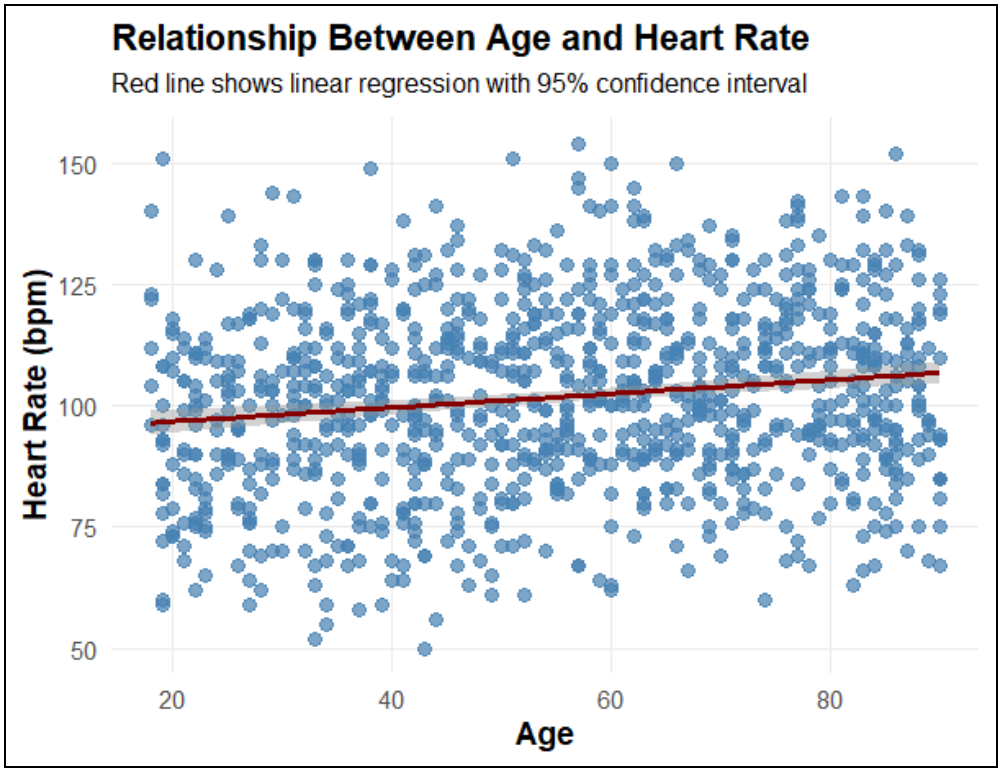

Figure 4. Scatter Plot: Age vs. Heart Rate

- The scatter plot shows a slightly upward trend, indicating a weak positive relationship between age and heart rate.
- This suggests that as age increases, heart rate tends to increase slightly, but the relationship is not strong.
- Other factors aside from age may be influencing heart rate more significantly in the dataset.

2.3 Advanced Statistical Insight

Table 2. Pearson Correlation Results Between Systolic and Diastolic Blood Pressure

Statistic	Value
Test Used	Pearson's correlation
Variables Tested	Systolic_BP vs Diastolic_BP
Correlation Coefficient (r)	0.7695

t-value	37.954
Degrees of Freedom (df)	992
p-value	< 2.2e-16
95% Confidence Interval	[0.7429, 0.7937]

- The correlation coefficient of 0.77 indicates a strong positive relationship between systolic and diastolic blood pressure.
- The very small p-value (< 2.2e-16) shows that this correlation is statistically significant.
- This means that as systolic BP increases, diastolic BP tends to increase as well.

4. Conclusion

This comprehensive analysis of the “Vital Signs and Lifestyle Health Data” reveals valuable insights into the health status and behavioral patterns of the sampled individuals. Descriptive statistics highlight an aging population with elevated average blood pressure and heart rate, moderately high glucose and cholesterol levels, and suboptimal sleep duration—all of which indicate potential cardiometabolic risk. Data visualizations supported these findings by showcasing distributions in BMI, cholesterol, and sleep hours, and illustrating relationships such as the weak positive trend between age and heart rate. Notably, a strong, statistically significant correlation ($r = 0.77$, $p < 0.001$) was found between systolic and diastolic blood pressure, confirming their interconnected behavior. Overall, the dataset underscores key risk areas, particularly sleep deprivation and blood pressure elevation, offering critical direction for targeted health interventions and future predictive modeling.