Final Project Report

BOWLING GREEN STATE UNIVERSITY
BGSU
A PUBLIC UNIVERSITY FOR THE PUBLIC GOOD.

Sleep Health and Lifestyle Analyses.

A physical therapist would want to make use of available historic data from a physiotherapy facility she has just started to better respond to clients' needs.

Student: Isaac Kobby Anni
Major: Data Science

MATH 6490 Final Project
Fall 2024

Contents

1. Introduction

The Sleep Health and Lifestyle Dataset provides a comprehensive view of various factors that impact individual health, focusing on sleep quality, lifestyle choices, and cardiovascular health. With 374 entries, the dataset includes demographic information, sleep metrics, physical activity levels, stress levels, body mass index (BMI) categories, blood pressure readings, and the presence of sleep disorders. This data presents an invaluable resource for analyzing the interconnections between sleep, physical activity, stress, and overall health. Such insights are crucial for healthcare providers, especially in a physical therapy setting, where lifestyle adjustments and preventive measures play a significant role in improving patient outcomes.

In this report, we aim to answer several research questions relevant to physical therapists, using statistical graphics to uncover patterns and trends within the dataset. We will explore how factors like physical activity and stress impact sleep quality and cardiovascular health, assess the prevalence of sleep disorders across different BMI categories and demographics, and analyze how these insights might influence patient care.

By understanding these relationships, a physical therapy facility can develop more targeted wellness programs, advise patients on lifestyle modifications, and offer evidence-based recommendations that address specific health concerns. This report leverages the power of data visualization to make complex relationships more accessible and actionable for healthcare professionals, ultimately aiming to support improved patient health and well-being.

1.1 Objectives

Specifically, this project aims to

- **Analyze sleep patterns with recommendation:** Examine how factors such as physical activity, stress levels, and BMI categories influence sleep duration and quality.
- **Assess physical activity and health outcomes:** Explore the relationship between physical activity levels, cardiovascular indicators (such as heart rate and blood pressure), and overall health outcomes to understand the benefits of regular activity on cardiovascular health.
- **Investigate stress and blood pressure:** Evaluate the impact of stress levels on blood pressure categories to determine if higher stress correlates with hypertension, providing evidence for stress management as a tool for blood pressure control
- **Identify prevalence of sleep disorders:** Determine the prevalence of sleep disorders (e.g., Insomnia, Sleep Apnea) across different demographics and BMI categories, offering insights into which groups are most at risk.
- **Explore age-based trends:** Analyze sleep duration, quality, and other health metrics across different age groups to understand how sleep and health behaviors vary by age.
- **Examine occupation and lifestyle factors:** Investigate how different occupations correlate with sleep duration, physical activity levels, and stress, identifying high-risk occupations that may benefit from targeted wellness programs.

1.2 Research Questions

- How does physical activity level relate to the quality of sleep across different BMI categories?
- What is the association between stress levels and blood pressure categories?
- What is the distribution of sleep disorders across different BMI categories?
- How does physical activity level influence heart rate?
- How do sleep duration and quality of sleep vary across different age groups?
- What is the prevalence of sleep disorders by gender?
- How does occupation impact sleep duration and other health indicators?
- What is the relationship between stress levels and physical activity when visualized together with BMI categories?

2. Background

Sleep quality, physical activity, and stress levels are key indicators of overall health and well-being. Research has shown that inadequate sleep and high stress can increase the risk of chronic conditions such as hypertension, obesity, and cardiovascular diseases. Sleep disorders, particularly common among those with high body mass index (BMI) or sedentary lifestyles, contribute to further health complications, affecting mental and physical performance in daily life. In recent years, there has been a growing interest in understanding the interrelationships between lifestyle factors, such as physical activity, sleep quality, and health outcomes, to guide effective interventions.

This study leverages a comprehensive dataset containing information on sleep patterns, lifestyle behaviors, and health metrics to explore these relationships. By identifying patterns among variables like physical activity, sleep disorders, BMI, and stress, this analysis aims to provide insights that can support healthcare providers and physical therapists in recommending tailored interventions for improving clients' health outcomes.

3. Dataset

This project utilizes the "Sleep Health and Lifestyle Dataset," from kaggle which includes 400 records with 13 columns. It is under public license and available on kaggle for research. Each row represents an individual's profile, capturing a wide range of information about their sleep patterns, lifestyle choices, and health metrics. The dataset is designed to support analyses of how various lifestyle factors correlate with sleep quality, physical activity, stress levels, and health indicators like BMI and blood pressure.

Dataset Columns and Descriptions:

- **Person ID:** A unique identifier for each individual in the dataset.
- **Gender:** The gender of the individual (Male/Female).
- **Age:** The age of the individual in years.
- **Occupation:** The individual's occupation or profession, allowing analysis of lifestyle differences by profession.
- **Sleep Duration (hours):** The average number of hours the individual sleeps per night.
- **Quality of Sleep (scale: 1-10):** A subjective rating of sleep quality, with higher values indicating better sleep quality.
- **Physical Activity Level (minutes/day):** The daily duration of physical activity in minutes.
- **Stress Level (scale: 1-10):** A subjective rating of stress levels, with higher values representing higher stress.
- **BMI Category:** The BMI classification of the individual (e.g., Underweight, Normal Weight, Overweight, Obese).
- **Blood Pressure (systolic/diastolic):** The blood pressure reading, recorded as systolic over diastolic pressure.
- **Heart Rate (bpm):** The resting heart rate of the individual, measured in beats per minute.
- **Daily Steps:** The average number of steps taken by the individual per day.
- **Sleep Disorder:** The presence or absence of a sleep disorder, categorized as None, Insomnia, or Sleep Apnea.

This dataset provides a rich source of information, making it possible to explore the relationships between lifestyle factors, sleep habits, and health outcomes. The diverse variables offer a foundation for analyzing patterns and correlations that could inform health interventions and personalized care plans.

Description: df [6 × 13]

	Person.ID <int>	Gender <chr>	Age <int>	Occupation <chr>	Sleep.Duration <dbl>	Quality.of.Sleep <int>	Physical.Activity.Level <int>	Stress.Level <int>
1	1	Male	27	Software Engineer	6.1	6	42	6
2	2	Male	28	Doctor	6.2	6	60	8
3	3	Male	28	Doctor	6.2	6	60	8
4	4	Male	28	Sales Representative	5.9	4	30	8
5	5	Male	28	Sales Representative	5.9	4	30	8
6	6	Male	28	Software Engineer	5.9	4	30	8

6 rows | 1–9 of 13 columns

Description: df [6 × 13]

Physical.Activity.Level <int>	Stress.Level <int>	BMI.Category <chr>	Blood.Pressure <chr>	Heart.Rate <int>	Daily.Steps <int>	Sleep.Disorder <chr>
42	6	Overweight	126/83	77	4200	None
60	8	Normal	125/80	75	10000	None
60	8	Normal	125/80	75	10000	None
30	8	Obese	140/90	85	3000	Sleep Apnea
30	8	Obese	140/90	85	3000	Sleep Apnea
30	8	Obese	140/90	85	3000	Insomnia

6 rows | 8-14 of 13 columns

Figure 1: Data Overview

3.1 Data preparation

In preparing the "Sleep Health and Lifestyle Dataset" for analysis, several data cleaning and feature engineering steps were undertaken to ensure the dataset was suitable for visualizations and insights generation.

1. Checking for Missing Values: The dataset was initially examined for any missing or NaN (Not a Number) values across all columns. However, this dataset contaiṇd no nan values.

2. Feature Engineering:

    - **Age Group:** The "Age" variable was segmented into groups, creating an "Age Group" feature to allow for group comparisons. Age was divided into intervals such as 20-30, 30-40, 40-50, and 50-60 years. This grouping facilitates analysis by age cohort and provides a clearer understanding of how sleep, lifestyle habits, and health outcomes differ across age groups.

    - **BP Category:** A "BP Category" (Blood Pressure Category) feature was derived from the "Systolic" and "Diastolic" blood pressure readings. Using clinical guidelines, individuals were categorized into blood pressure levels such as "Normal," "Elevated," "Hypertension Stage 1," "Hypertension Stage 2," and "Hypertensive Crisis." This classification provides insights into the prevalence of different blood pressure conditions within the dataset and enables exploration of correlations between blood pressure and other health variables.

These data preparation steps laid a solid foundation for the analysis, allowing for the generation of insightful visualizations and meaningful patterns. The additional features, "Age Group" and "BP Category," enhance the dataset's utility by enabling more nuanced subgroup analyses, ultimately leading to more actionable insights for healthcare and lifestyle recommendations.

4. Exploratory Data Analysis

The exploratory data analysis (EDA) phase involved generating a series of visualizations to uncover patterns, distributions, and relationships within the "Sleep Health and Lifestyle Dataset." Various statistical graphics techniques were applied to understand how factors like sleep duration, physical activity level, stress levels, and heart rate vary across different age groups, BMI categories, and blood pressure levels.

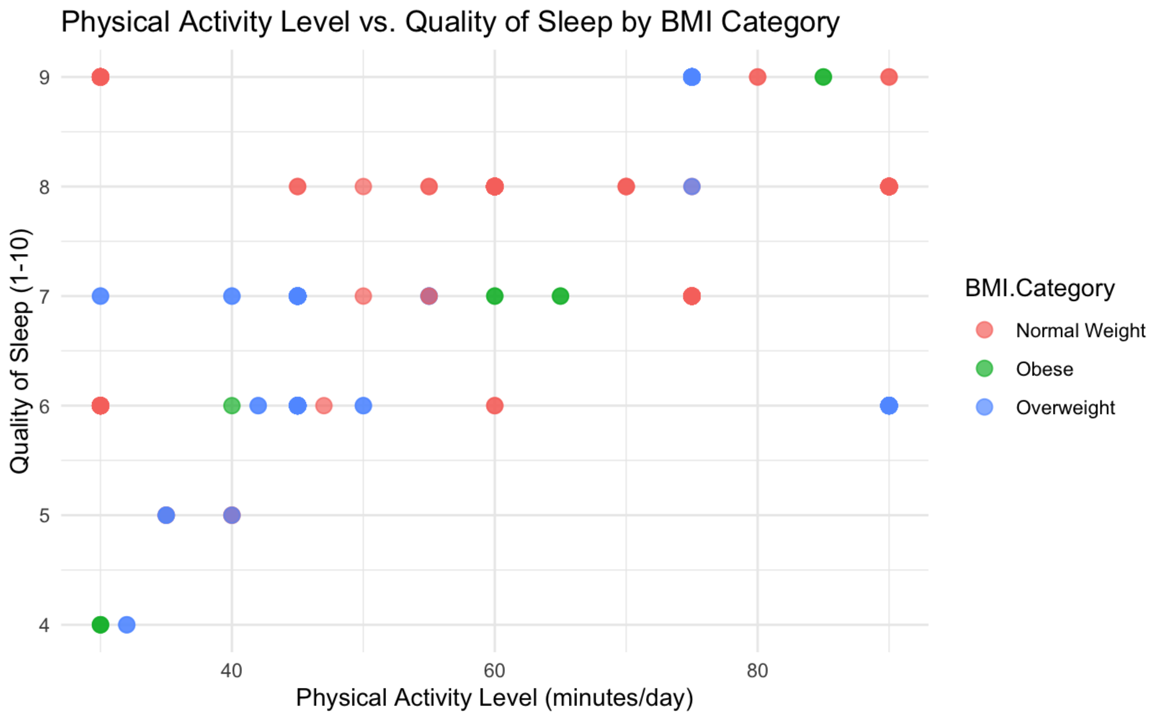

Figure 2: Impact of physical activity on sleep quality across different BMI Category

- Higher Physical Activity and Sleep Quality: Individuals with higher physical activity levels (70–90 minutes per day) tend to report higher sleep quality (around 8 to 9 on a 1–10 scale), regardless of BMI category. This suggests a positive relationship between physical activity and sleep quality.
- Variability in Lower Physical Activity: Among those with lower physical activity (30–60 minutes per day), there's a wider range of sleep quality scores. This range includes lower sleep quality (4 to 6), particularly among individuals categorized as overweight (blue dots). This indicates that lower physical activity may be associated with more variability in sleep quality, especially for individuals who are overweight.

- Normal Weight Group and High Sleep Quality: Individuals with a ”Normal Weight” BMI (represented by red dots) generally maintain higher sleep quality ratings (7 to 9), even with varying physical activity levels. This might suggest that normal weight individuals experience more stable sleep quality.

- Lower Sleep Quality in Obese and Overweight Categories: Some individuals in the obese (green) and overweight (blue) categories report lower sleep quality (4–6), especially at lower physical activity levels. This trend may highlight the impact of BMI on sleep quality, where higher BMI could correlate with poorer sleep outcomes, especially if physical activity is minimal.

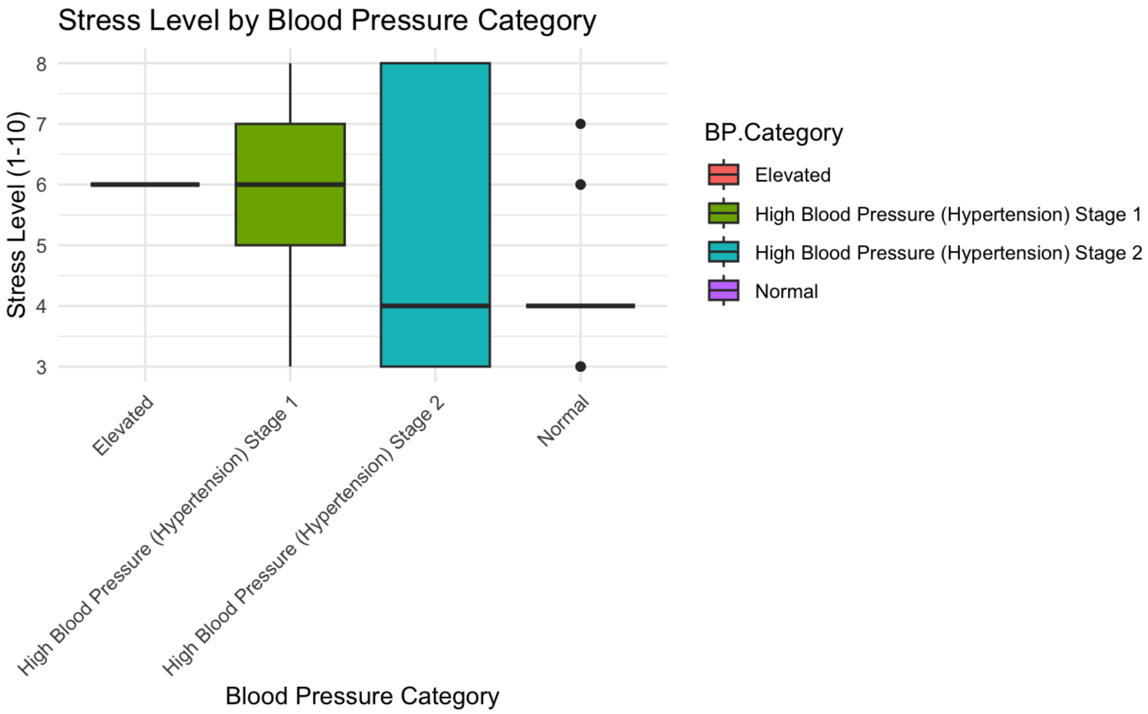

Figure 3: Stress levels by blood pressure category

- **Higher Stress in Hypertension Stage 1:** Individuals with ”High Blood Pressure (Hypertension) Stage 1” tend to report moderately high stress levels, with the middle 50% of their stress scores ranging around 5 to 7 on a 1–10 scale. This suggests a correlation between moderate hypertension and higher stress.

- **Broader Range of Stress in Hypertension Stage 2:** Those with ”Hypertension Stage 2” show a broader spread in stress levels, ranging from 3 to 8. This category includes individuals with both high and low stress, indicating that Stage 2 hypertension may encompass a wider range of stress responses, possibly due to variability in lifestyle or coping mechanisms

- **Lower Stress in Normal and Elevated Categories:** Individuals in the ”Normal” and ”Elevated" blood pressure categories report generally lower stress levels, with narrower distributions. The "Normal" group, in particular, shows low and stable stress levels, reinforcing the association between lower stress and normal blood pressure.

- **Stress and Blood Pressure Relationship:** Overall, the plot suggests a positive relationship between blood pressure levels and stress, with higher blood pressure categories (especially hypertension stages) linked to higher or more variable stress levels.

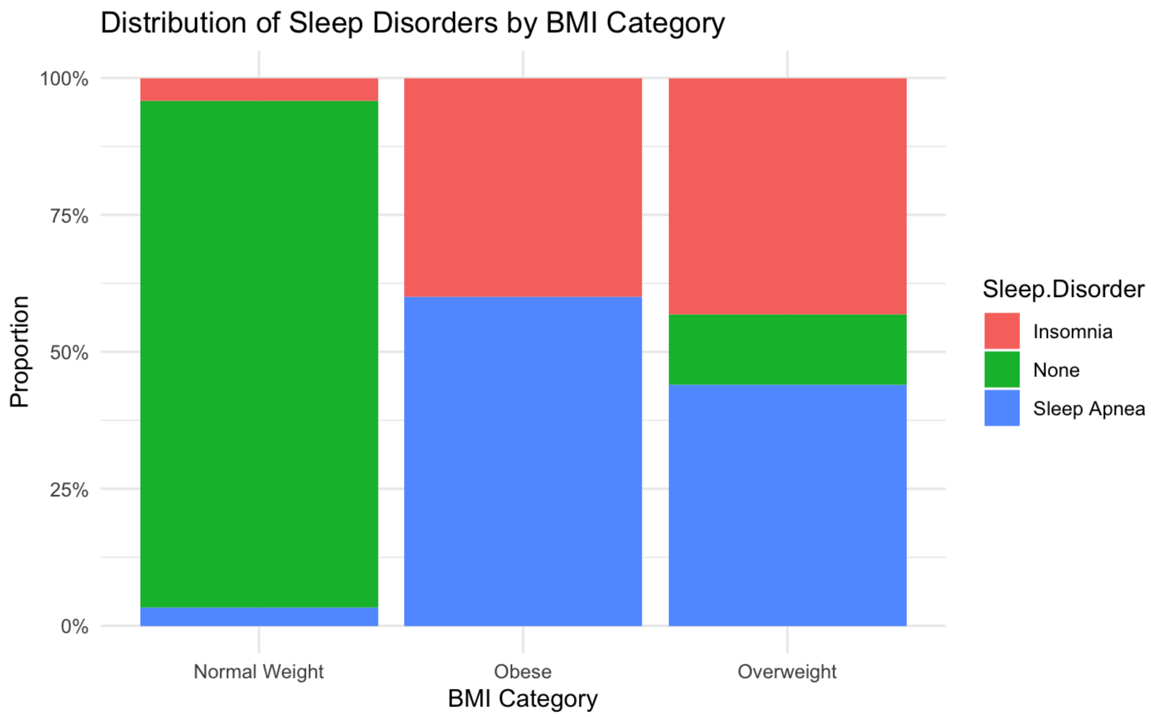

Figure 4: Distribution of Sleep Disorders by BMI Category

- **Normal Weight:** The vast majority of individuals with a "Normal Weight" BMI category have no sleep disorders (green), with only a small portion reporting sleep apnea (blue) and an even smaller portion reporting insomnia (red). This suggests that maintaining a normal BMI may be associated with fewer sleep-related health issues.

- **Obese:** Among individuals in the "Obese" category, there is a significant proportion experiencing sleep apnea (blue), followed by those with insomnia (red). Very few in this category report having no sleep disorder (green). This trend indicates a possible link between obesity and higher prevalence of sleep disorders, particularly sleep apnea.

- **Overweight:** In the "Overweight" group, there is a more balanced distribution between sleep apnea and insomnia, with a moderate portion reporting no sleep disorders. Although this group has fewer cases of sleep disorders than the obese group, it still shows a higher prevalence compared to the normal weight group.

- **BMI and Sleep Disorders Correlation:** Overall, the chart suggests a trend where individuals with higher BMI categories (Obese and Overweight) are more likely to report sleep disorders, particularly sleep apnea and insomnia. In contrast, those with a normal BMI are less likely to experience sleep-related issues.

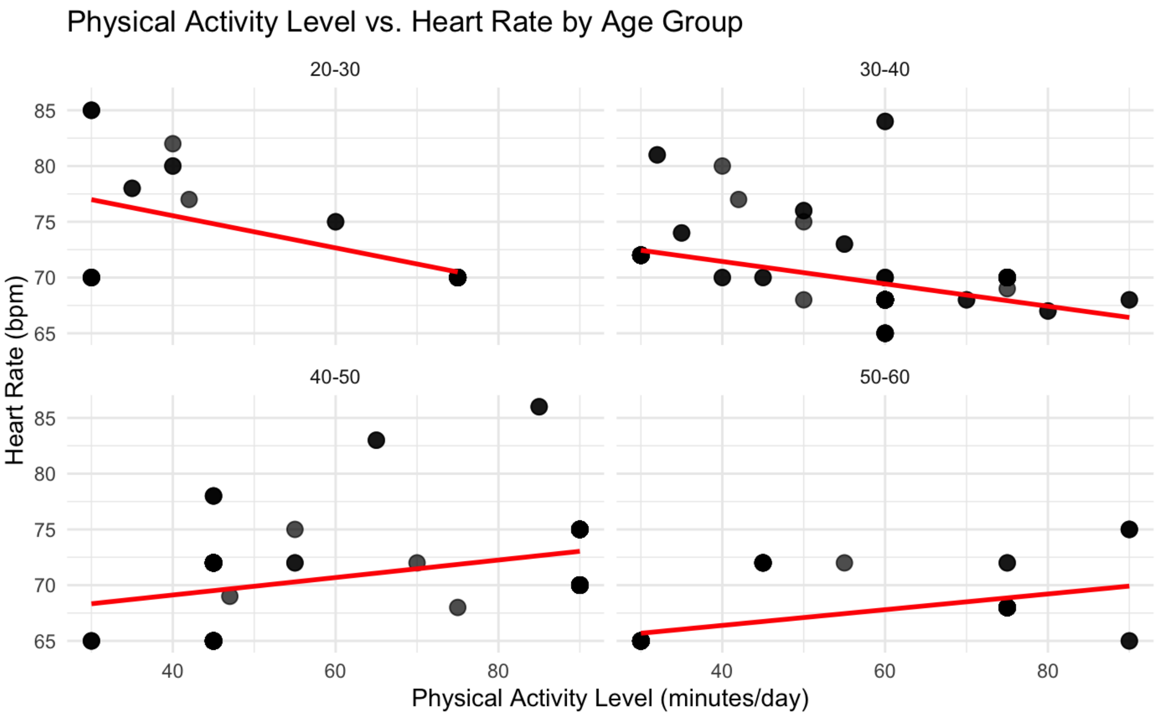

Figure 5: Physical Activity Level vs. Heart Rate

- **20-30 Age Group:** There is a noticeable negative trend between physical activity level and heart rate. This suggests that in this age group, higher physical activity might be associated with a lower resting heart rate, indicating potentially better cardiovascular fitness.

- **30-40 Age Group:** Similar to the 20-30 group, there is a negative correlation between physical activity and heart rate, though it appears less pronounced. This trend suggests that increased activity may also contribute to a lower heart rate in this group, albeit with more variability..

- **40-50 Age Group:** Interestingly, this age group shows a slight positive trend, where physical activity level and heart rate increase together. This may suggest that in this age group, higher physical activity might be associated with a slightly higher heart rate, potentially due to different fitness levels or response to physical activity.

- **50-60 Age Group:** This group also shows a slight positive trend. However, the trend is weak and indicates that physical activity and heart rate do not strongly correlate in this age range.

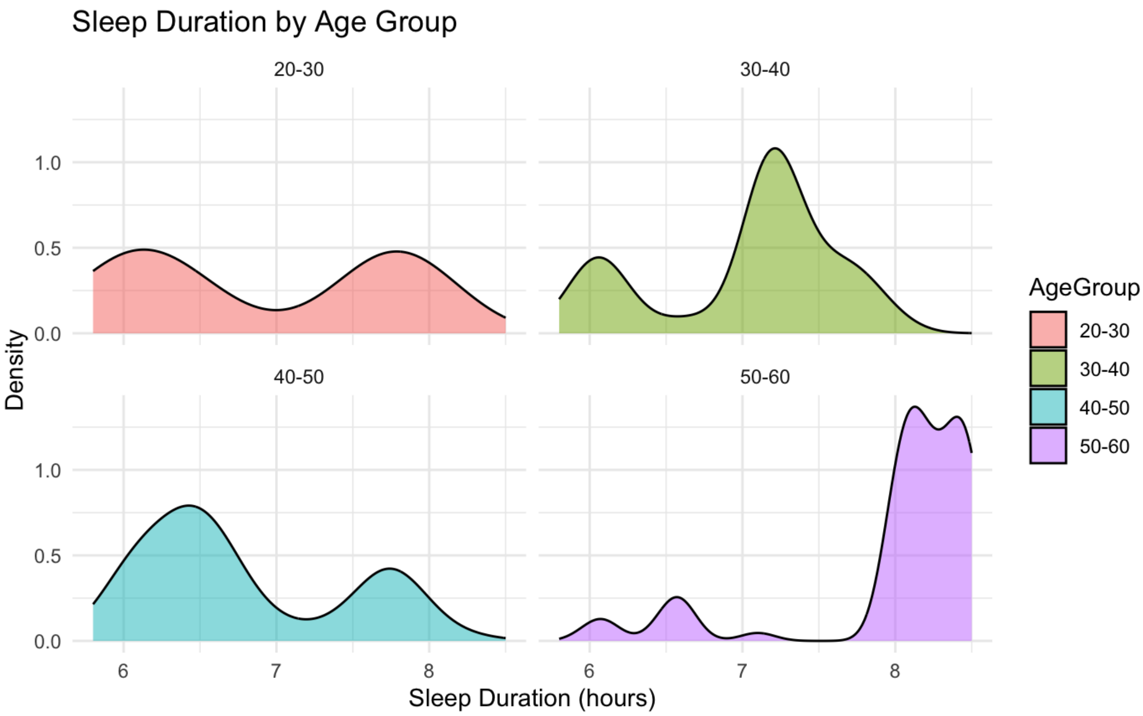

Figure 6: Sleep focus by Age group

- **20-30 Age Group:** Sleep duration in this age group is fairly evenly distributed, with a slight preference for durations between 6 and 8 hours. This indicates that younger adults may have more varied sleep patterns, potentially influenced by lifestyle factors.
- **30-40 Age Group:** This age group shows a pronounced peak around 7-8 hours, suggesting that most individuals within this age range achieve a relatively stable and adequate sleep duration. This could indicate an increased prioritization of sleep as individuals in this age group may focus on maintaining a work-life balance.
- **40-50 Age Group:** The distribution here is bimodal, with peaks around 6 and 8 hours. This could reflect different subgroups within this age range, with some individuals experiencing shorter sleep durations, possibly due to work or family responsibilities, while others achieve longer sleep durations.
- **50-60 Age Group:** This group displays a strong peak at 7-8 hours, suggesting a more consistent sleep pattern with most individuals achieving adequate sleep duration. This may reflect an increased focus on health and rest in this age group.

Figure 7: Sleep duration by occupation

- **Accountant:** Accountants display a consistent sleep duration around 7 hours, with a small range and minimal outliers, suggesting stable work hours.
- **Doctor:** Doctors show a wide range of sleep durations, with a higher median.
- **Engineer:** Engineers generally get higher sleep durations, clustering around the upper range with a more consistent pattern, suggesting stable work-life balance in terms of sleep.
- **Lawyer:** Lawyers show moderate sleep duration with a somewhat narrow range, suggesting that they generally maintain a regular sleep schedule.
- **Manager:** Managers have a similar pattern to Lawyers, with moderate sleep durations, although with slightly higher variability, possibly due to varying work demands.
- **Nurse:** Nurses also exhibit a wide range of sleep durations with a high median, though some outliers indicate variability, likely due to shifts or irregular working hours.
- **Salesperson and Sales Representative:** These roles show limited data points but indicate moderate sleep duration with less variability
- **Software Engineer:** Software Engineers show lower median sleep durations with a narrow range, possibly reflecting long or irregular working hours common in the tech industry.
- **Teacher:** Teachers exhibit higher sleep durations with minimal variability, suggesting a more

stable schedule conducive to regular sleep.

5. Modeling

The model was fit to the dataset using the following equation:

Model Equation

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon$$

where:

- Y: Sleep Duration,
- X_1: Age group (*AgeGroup*),
- X_2: Gender (*Gender*),
- X_3: Stress level (*Stress.Level*),
- X_4: Physical activity level (*Physical.Activity.Level*),
- X_5): BMI Category (*BMI.Category*),
- $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4\, \beta_5$: Model coefficients,
- ϵ: Residual error term.

To analyze the factors influencing sleep duration, a multiple linear regression model was developed. The dependent variable was Sleep Duration, measured in hours, while the predictors included:

- **AgeGroup (categorical):** representing different age brackets of individuals,
- **Gender (categorical):** denoting male or female participants,
- **Stress.Level (numeric):** measured on a scale of 1 to 10,
- **Daily.Steps (numeric):** the average number of steps taken per day, and
- **Physical.Activity.Level (numeric):** measured in minutes per day.
- **BMI.Category:** Body Mass Index (BMI) classifications.

5.1 Model Summary

```
Call:
lm(formula = Sleep.Duration ~ AgeGroup + Gender + Stress.Level +
    Daily.Steps + Physical.Activity.Level + BMI.Category, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.76815 -0.17307 -0.05607  0.17251  0.83556

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              8.321e+00  1.042e-01  79.839  < 2e-16 ***
AgeGroup30-40           -2.470e-01  5.348e-02  -4.619 5.37e-06 ***
AgeGroup40-50           -7.845e-02  5.886e-02  -1.333    0.183
AgeGroup50-60            6.816e-01  7.027e-02   9.700  < 2e-16 ***
GenderMale               3.652e-01  3.530e-02  10.345  < 2e-16 ***
Stress.Level            -2.753e-01  1.074e-02 -25.629  < 2e-16 ***
Daily.Steps             -7.113e-05  1.736e-05  -4.098 5.15e-05 ***
Physical.Activity.Level  1.346e-02  1.245e-03  10.805  < 2e-16 ***
BMI.CategoryObese       -5.413e-01  1.040e-01  -5.206 3.23e-07 ***
BMI.CategoryOverweight  -5.317e-01  3.919e-02 -13.567  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2555 on 364 degrees of freedom
Multiple R-squared:  0.8994,    Adjusted R-squared:  0.8969
F-statistic: 361.4 on 9 and 364 DF,  p-value: < 2.2e-16
```

Figure 8: Regression Model Output Summary

The regression analysis revealed the following:

- **Intercept (8.321)**: The baseline sleep duration for the reference group (AgeGroup: 20-30, Gender: Female, BMI.Category: Normal Weight, Stress Level: 0, Daily Steps: 0, Physical Activity Level: 0) is approximately 8.32 hours.
- **AgeGroup** significantly influenced sleep duration. Compared to the baseline age group (20-30 years):
  - i Participants aged 30-40 had decreases in sleep duration while 40-50 do not affect the sleep duration.
  - ii Participants aged 50-60 experienced a 0.068-hour increase in sleep duration.
- **Gender** also played a significant role, with male participants sleeping 0.365 hours longer than female participants, on average.
- **Stress.Level** exhibited a strong negative association, with each unit increase in stress level leading to a 0.275-hour reduction in sleep duration.
- **Physical.Activity.Level** was positively associated with sleep duration, with each additional minute of physical activity per day contributing to a 0.013-hour increase in sleep duration.
- **Daily.Steps:** An increase of 1,000 daily steps is associated with a negligible decrease in sleep duration by 0.071 hours (4.26 minutes).

- **BMI.Category:**
  - Individuals in the Obese category sleep approximately 0.541 hours (32.46 minutes) less compared to those in the Normal Weight category.
  - Individuals in the Overweight category sleep approximately 0.531 hours (31.86 minutes) less compared to those in the Normal Weight category.

The model achieved a **Multiple R-squared value of 0.8994**, indicating that 89.94% of the variance in sleep duration is explained by the predictors. The **Adjusted R-squared of 0.8969** suggests a high degree of explanatory power while accounting for the number of predictors in the model. The overall model was statistically significant **(F-statistic: 361, p-value $<$ 2.2e-16)**, confirming the relevance of the included predictors.

5.2 Residual Analysis

To validate the regression assumptions, residual diagnostics were conducted

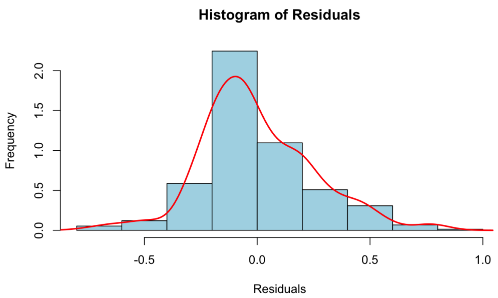

Figure 9: Normality of Residuals

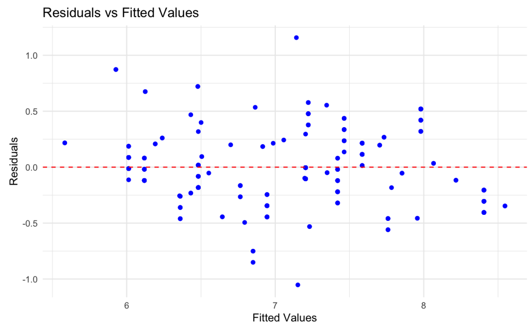

Figure 10: Randomness of residuals

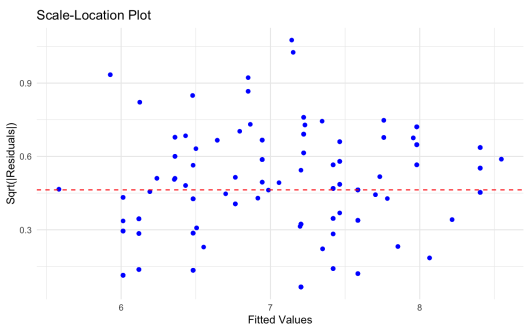

Figure 11: Constant Variance - Homoscedasticity

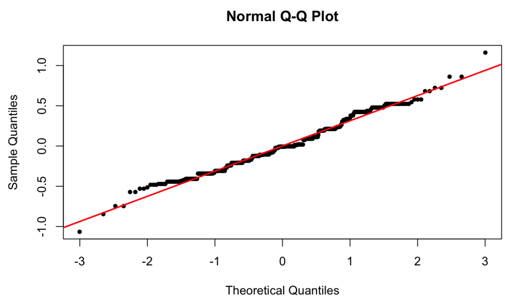

Figure 12: Residual QQ plot

Residual analysis is a critical step in evaluating the adequacy of the linear regression model and ensuring the underlying assumptions are met. In this section, we examine the residuals of the fitted model using four key diagnostic plots: Normality of residuals, randomness of residuals, constant variance, and the QQ plot.

i **Figure 10** presents the histogram of residuals overlaid with a density curve. The distribution of the residuals appears approximately symmetric and unimodal. However, slight deviations from normality can be observed in the tails.

ii **Figure 11** displays the residuals plotted against the fitted values. The residuals are randomly scattered around the horizontal reference line at zero, indicating no discernible patterns. This supports the assumption of independence of residuals and suggests that the model captures the systematic variation in the data effectively.

iii **Figure 12** depicts the Scale-Location plot, which assesses the homoscedasticity assumption. The plot shows the square root of the absolute residuals against the fitted values. The vertical spread of the points remains roughly constant across the range of fitted values, suggesting that the residuals exhibit constant variance. No clear funnel or fan shape is visible, which would indicate heteroscedasticity.

iv **Figure 13** illustrates the normal Q-Q plot of the residuals. The residual points closely follow the diagonal reference line, except for slight deviations in the extreme tails. This indicates that the residuals are approximately normally distributed, which aligns with the assumption of normality required for valid hypothesis testing and confidence interval estimation

Overall, the residual analysis supports the adequacy of the fitted model. The residuals exhibit randomness, independence, and constant variance, while showing only slight deviations from normality. These results suggest that the model provides a reliable representation of the underlying data, though further investigation or transformations might be considered to address any minor departures from assumptions if necessary.

6. Limitations and Challenges

While the project involved a comprehensive analysis and modeling of the data, several limitations were encountered that may impact the conclusions drawn:

- **Violation of Normality Assumption:** Despite applying various transformations to the response variable and predictors, the residuals consistently failed the Shapiro-Wilk test for normality. This suggests potential issues with the suitability of linear regression for this dataset, as normality is a key assumption for valid hypothesis testing and inference. Although the visual QQ plot indicated approximate normality, the statistical test's failure highlights a limitation in fully satisfying the model assumptions.

- **Data Suitability for Regression:** The nature of the data and relationships between variables might be better suited for classification tasks rather than regression. For example, categorical variables such as AgeGroup and Gender are often used in classification models, and the presence of discrete levels in Sleep.Duration might indicate a categorical outcome rather than a continuous one.
- The reliability of predictions is contingent on the quality and completeness of the dataset. Missing or inaccurate data may introduce bias and reduce the model's efficacy.
- **Residual Analysis and Model Adequacy:** While residual analysis suggested approximate randomness and independence, slight deviations in normality and variability raise concerns about whether the linear regression model is the best approach for capturing the data's underlying structure.
- **Limited Generalizability:** The data is limited to a specific population or context, which might not generalize to other groups. For instance, the variables measured might be highly specific to the dataset's origin, limiting applicability to broader populations.

7. Conclusion

This project explored the relationships between sleep health and lifestyle factors using a comprehensive dataset. Through exploratory data analysis and statistical modeling, we identified key factors such as age, gender, stress levels, and physical activity that significantly influence sleep duration. The analyses revealed that higher stress levels are associated with shorter sleep duration, while increased physical activity tends to promote better sleep. Demographic factors like age and gender also exhibited clear patterns in sleep behavior.

Although the multiple linear regression model performed well overall, with an adjusted R-squared of 0.89, residual diagnostics indicated that the assumption of normality was not fully met. This suggests that the dataset may benefit from alternative modeling approaches, such as classification or machine learning techniques, to capture complex relationships. Nevertheless, this study provides valuable insights into the interplay between lifestyle factors and sleep health, offering a foundation for future research and potential applications in personalized health interventions.

8. Recommendation

- Exploring Alternative Models: Given the potential suitability of the data for classification tasks, models such as logistic regression, decision trees, or ensemble methods could be explored. For instance, classifying individuals based on their sleep disorder or BMI category using the predictors could be more insightful.
- Dimensionality Reduction: Methods like Principal Component Analysis (PCA) could reduce the dataset's dimensionality, potentially improving the model's fit and interpretability.
- Testing for Multicollinearity: Using variance inflation factors (VIF) to assess multicollinearity could help determine whether certain predictors should be excluded or combined for better interpretability and performance.

9. Future Directions

- Employ classification models to better capture group-specific trends and outcomes, such as predicting sleep disorders based on demographic and lifestyle variables.
- Explore alternative modeling approaches, such as generalized additive models (GAMs) or machine learning algorithms, to address potential non-linear relationships and interactions between variables.
- Collect more granular data to enhance generalizability and provide deeper insights into health and sleep patterns across diverse populations.

References

[1] Di Milia, L., Vandelanotte, C., and Duncan, M. J. (2013). The association between short sleep and obesity after controlling for demographic, lifestyle, work and health related factors. *Sleep medicine*, 14(4):319–323.

[2] Dzierzewski, J. M., Sabet, S. M., Ghose, S. M., Perez, E., Soto, P., Ravyts, S. G., and Dautovich, N. D. (2021). Lifestyle factors and sleep health across the lifespan. *International Journal of Environmental Research and Public Health*, 18(12):6626.

[3] Mitchell, L. J., Davidson, Z. E., Bonham, M., O'Driscoll, D. M., Hamilton, G. S., and Truby, H. (2014). Weight loss from lifestyle interventions and severity of sleep apnoea: a systematic review and meta-analysis. *Sleep medicine*, 15(10):1173–1183.

[4] Ohida, T., Kamal, A., Uchiyama, M., Kim, K., Takemura, S., Sone, T., and Ishii, T. (2001). The influence of lifestyle and health status factors on sleep loss among the japanese general population. *Sleep*, 24(3):333–338.