Which Lifestyle Habits Most Strongly Predict Sleep Quality?

FINC 405 Research Project

By Dalia Poblano

Professor Vijay Kumar

December 03, 2025

Introduction

Beep, beep, beep — the alarm sounds, signifying the start of a new day. It's a sound most people dread, adults and children alike, especially after a night of little or poor quality sleep. Sleep is a fundamental part of human health, yet it is often undervalued in everyday life. According to an article by Cultivating Health, "The three pillars of health are nutrition, physical exercise, and sleep. All three of these are connected."Adequate sleep not only prevents illness but also supports growth, heart health, weight management, injury prevention, attention, memory, and learning (Cultivating Health). Without getting quality sleep, people are at a higher risk for Coronary Heart Disease, High Blood Pressure, Obesity, and Strokes (National Heart, Lung, and Blood Institute).

Despite these well-documented benefits, sleep deprivation remains widespread. One study found that only about 15% of adults regularly achieved the recommended 7–9 hours of sleep for at least five nights per week (Sleep Health). If there are numerous benefits that come with sleep, why do most people neglect their sleeping schedules? Are there ways to improve the prioritization of sleep in our life? All these questions reflect the problem that people face in our society and that is getting quality sleep. Whether it's homework assignments, children, or work responsibilities, people are not getting proper sleep quality and we are putting ourselves at risk of facing complications. Many scientists recommend reducing caffeine intake and limiting screen time before bed, but these are only part of the picture. Daily routines and lifestyle choices also influence how well we sleep. This research project aims to explore which lifestyle habits have the strongest association with sleep quality. Ultimately, it seeks to answer the guiding question of this study: **Which lifestyle habits most strongly predict sleep quality?**

Materials and Methods

Data:

Data Link: Click Here!

The dataset I used for this project came from Kaggle, a reliable source that provides people with various datasets for data analytics and projects. This dataset is composed of data from 374 participants and 13 columns. The column names are Person ID, Gender, Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Blood Pressure, Heart Rate, Daily Steps, and Sleep Disorder.

Numerical Values: Person ID, Age, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, Heart Rate, Daily Steps

Categorical Values: Gender, Occupation, BMI Category, Blood Pressure, Sleep Disorder

Focus Variables:

To answer the research question, I will be focusing on most numerical values such as **Stress Level, Physical Activity Level, Daily Steps, Heart Rate, and Blood Pressure**. In addition to these factors, the analysis will also examine whether demographic variables such as Gender, BMI Category, Occupation, and Age show any significant relationship with predicting sleep quality, even though these variables are not considered lifestyle habits.

Techniques Used:

- Data Importing and Cleaning
- Data Wrangling (Filtering, Grouping, Data Transformations)
- Exploratory Data Analysis (Detecting Outliers and Trends)
- Statistical Models
- DataVisualization (Scatter Plots, Box Plots, Correlation Heatmaps)

Equations:

Pearson's Correlation Coefficient

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Used Pearson's correlation to measure the strength and direction of the relationship between lifestyle habits (daily steps, stress level, physical activity level, blood pressure, heart rate) and sleep quality.

Linear Regression Model

$$\hat{y} = \beta_0 + \beta_1 x$$

While I didn't conduct a full regression analysis, I did run a simple linear regression model on the lifestyle habits I focused on in order to view their coefficients

Random Forest Model

In addition to exploratory and statistical techniques, I used a Random Forest classification model to determine which lifestyle habits most strongly predict sleep quality. Random Forests are ensembles of decision trees, where each tree selects optimal splits by minimizing Gini Impurity, defined as:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2$$

The overall prediction of the forest is computed through majority voting across all trees:

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \ldots, h_n(x))$$

Results with Interpretations

Data Prep:

Before starting to work with the data, I checked the data types that I would be working with and if there were any missing values in the data. I discovered that the variable “Sleep Disorder” had 219 missing values.

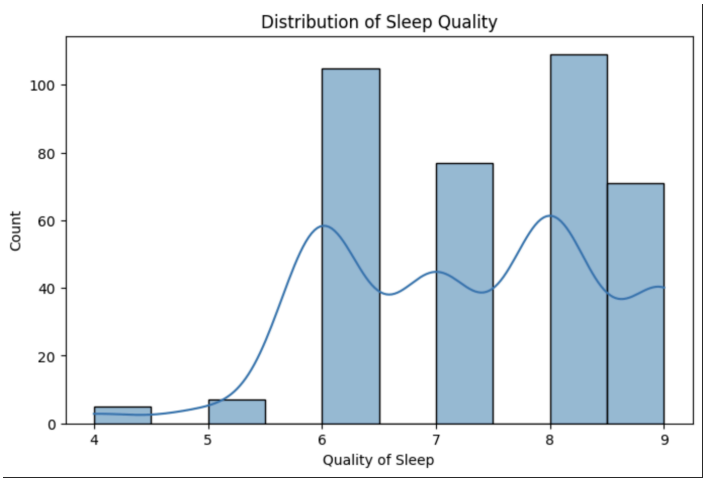
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 13 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Person ID                374 non-null    int64  
 1   Gender                   374 non-null    object 
 2   Age                      374 non-null    int64  
 3   Occupation               374 non-null    object 
 4   Sleep Duration           374 non-null    float64
 5   Quality of Sleep         374 non-null    int64  
 6   Physical Activity Level  374 non-null    int64  
 7   Stress Level             374 non-null    int64  
 8   BMI Category             374 non-null    object 
 9   Blood Pressure           374 non-null    object 
 10  Heart Rate               374 non-null    int64  
 11  Daily Steps              374 non-null    int64  
 12  Sleep Disorder           155 non-null    object 
dtypes: float64(1), int64(7), object(5)
memory usage: 38.1+ KB
None

Missing values:
 Person ID                  0
Gender                      0
Age                         0
Occupation                  0
Sleep Duration              0
Quality of Sleep            0
Physical Activity Level     0
Stress Level                0
BMI Category                0
Blood Pressure              0
Heart Rate                  0
Daily Steps                 0
Sleep Disorder            219
dtype: int64
```

For this reason I opted to completely drop this column as I would not be using it to answer my research question since sleep disorders like sleep apnea or insomnia are medical conditions not lifestyle habits.

Exploratory Data Analysis:

After cleaning the data from any missing values, I took a look at the distribution of sleep in a histogram plot to identify any potential trends and understand the data a bit more.



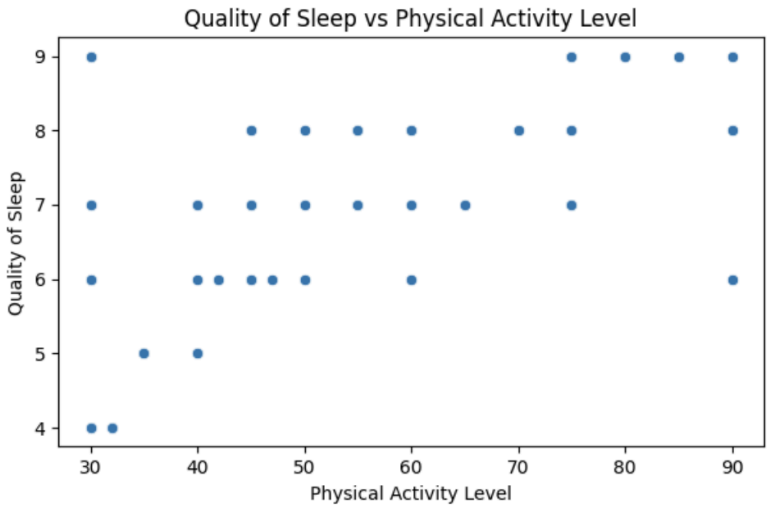
This histogram demonstrates a peak in the 6th and 8th level of quality sleep. This means that most people reported this level of sleep while another high percentage of the participants also reported a 7th and 9th level. Due to the high frequency amongst the 6-9 level range, we can also see that the distribution of sleep quality is slightly right skewed. At this point I did not see any outliers and overall I was able to conclude that most participants reported a good level of sleep quality with relatively few low quality sleepers.

Exploring Lifestyle Habits:

Since the dataset included many variables, I categorized them into two groups: Lifestyle Habits and Health Indicators. While the primary focus of this research is on lifestyle habits, health indicators were also examined as secondary predictors of sleep quality.

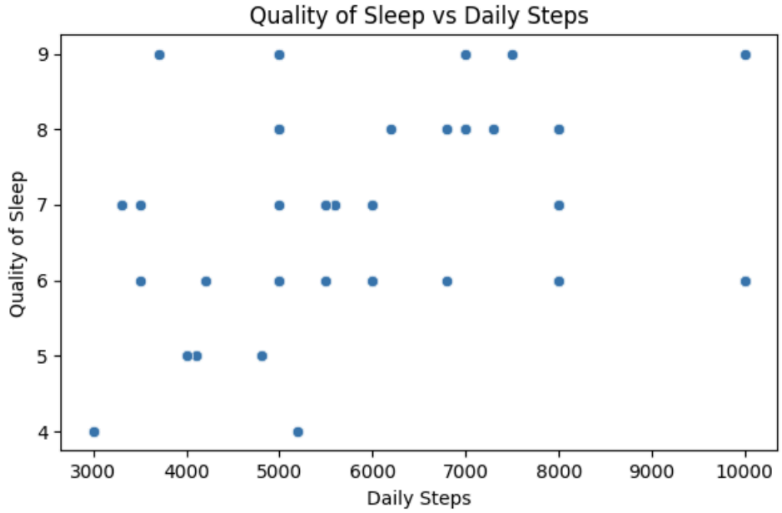
After creating scatter plots of lifestyle habits (Physical Activity Level, Daily Steps, and Stress Level) vs Quality of Sleep, we can make a few observations based on the following plots.

Quality of Sleep vs Physical Activity Level



For the Quality of Sleep vs Physical Activity Level plot , we can see that there's no strong linear relationship but there is a trend or pattern that appears. It seems that people that reported a moderate to high level of physical activity (50-90) tend to have sleep quality between 6 and 9. Similarly, the slowest sleep quality of 4 and 5 appear mostly at lower activity levels of 30-40. In conclusion, higher physical activity may be associated with slightly better sleep quality.

Quality of Sleep vs Daily Steps



For the Quality of Sleep vs Daily Steps plot, we can see that the data points are widely scattered, meaning daily steps alone do not strongly predict sleep quality. Additionally, we can see that sleep quality between 6 and 8 occurs across a wide range of step counts (3000–10,000) and very low sleep quality (4–5) tends to appear only when steps are below 5000. In conclusion, daily steps have a weak positive relationship with sleep quality, although more steps can help but it's not a strong predictor.

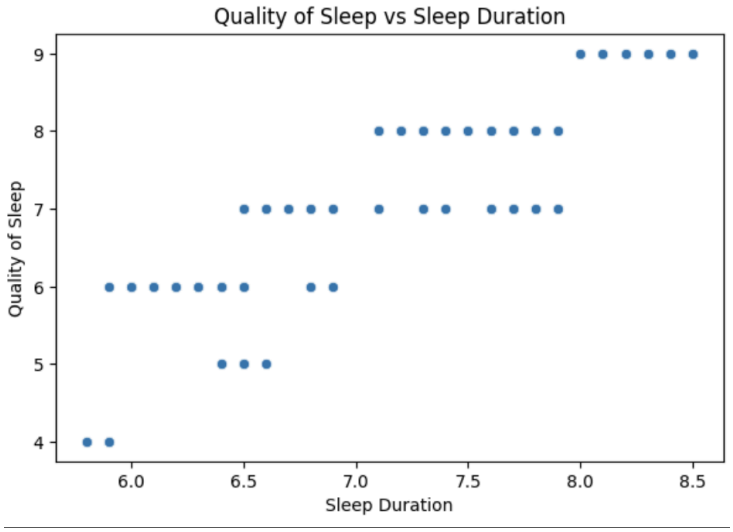
Quality of Sleep vs Stress Level

For the Quality of Sleep vs Stress Level plot, we can see that this plot shows the clearest pattern out of the three. Higher sleep quality levels (8-9) occur mostly when stress levels are low (3-4) and lower sleep quality levels (4-6) appear more often when stress levels are high (7-8). In conclusion, this is the strongest relationship among the three graphs, indicating a strong negative relationship between quality of sleep and stress level.

Exploring Health Indicators:

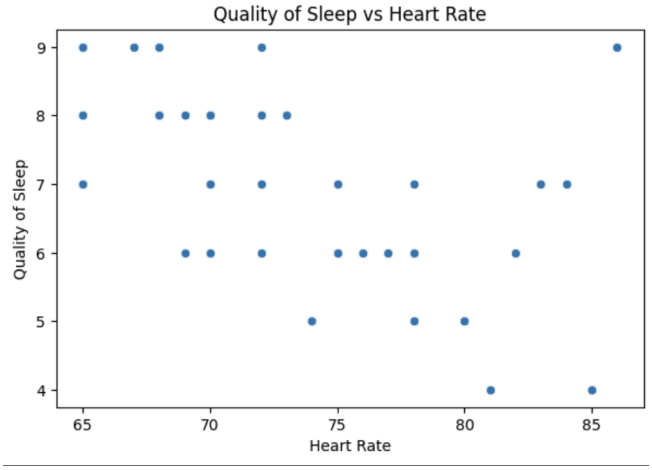
For this section, I examined Sleep Duration, Heart Rate, and Blood Pressure. Although these are not explicit lifestyle habits, they can be influenced by daily behaviors and may contribute to overall sleep quality.

Quality of Sleep vs Sleep Duration



For the Quality of Sleep vs Sleep Duration plot, we can see a positive relationship between sleep duration and sleep quality. Individuals who sleep 7–8.5 hours generally have higher sleep quality (7–9) and lower sleep quality (4–5) tends to occur when sleep duration is below ~6.5 hours. In conclusion, the plot shows that shorter sleep duration correlates with lower sleep quality and longer sleep duration correlates with higher sleep quality.

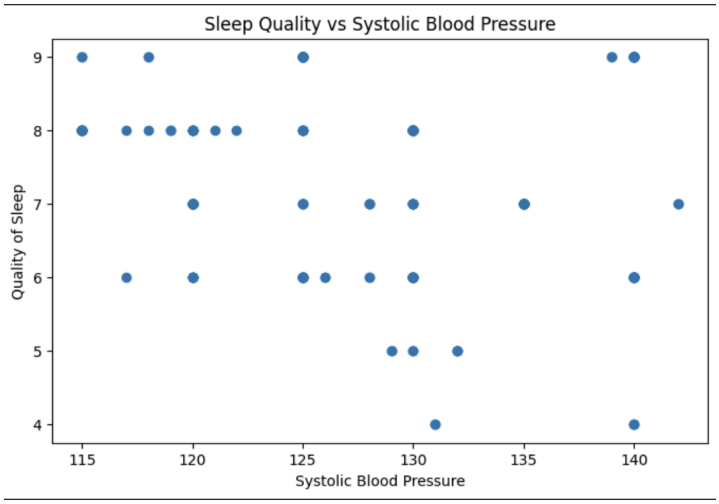
Quality of Sleep vs Heart Rate



For the Quality of Sleep vs Heart Rate plot, we can see that there is no strong linear relationship. However, there is a trend where lower sleep quality is more common at higher resting rates of 78-85 bpm and higher sleep quality is seen more in those with a lower to mid range heart rate of 65-75 bpm. In conclusion, the relationship is weak but higher resting heart rate might be associated with poorer sleep, possibly due to stress or physical fatigue.

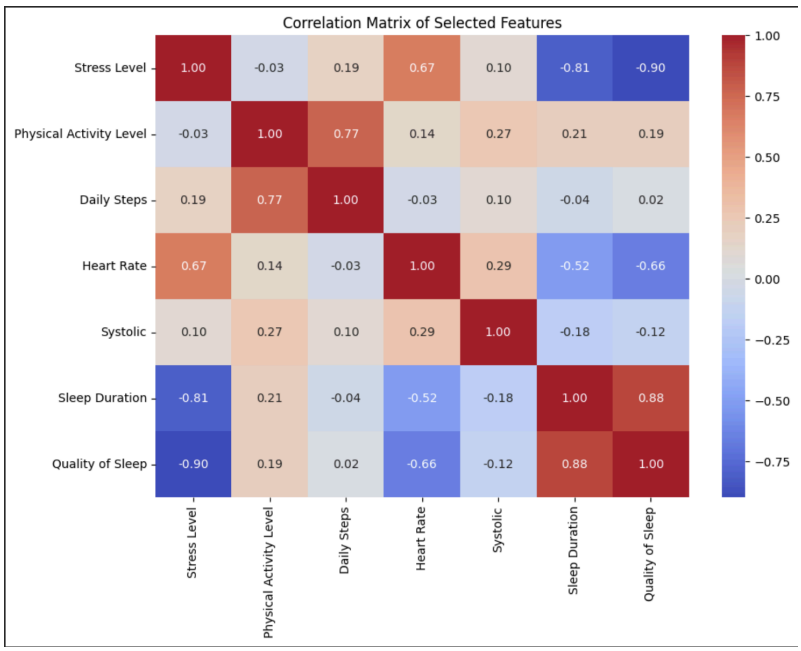
Quality of Sleep vs Blood Pressure

In order to plot the relationship between Quality of Sleep and Blood Pressure, I first had to split blood pressure into systolic and diastolic columns. Systolic pressure is the force of the blood flow when blood is pumped out of the heart. Diastolic pressure is the measurement between heartbeats when the heart is filling with blood. However, experts say that the systolic pressure is the best way to predict future cardiovascular events and death so we should be more concerned about this pressure.



For the Quality of Sleep vs Systolic Blood Pressure plot, we can see no clear linear correlation between both. However, lower sleep quality is slightly more common in individuals with higher systolic blood pressure but this trend is very inconsistent. In conclusion, systolic blood pressure does not appear to be a strong predictor of sleep quality in this dataset.

Correlation Heatmap:



The correlation matrix heatmap is a great way to showcase how our selected features relate to our target variable which is quality of sleep. From our heatmap we can make the following observations:

Strongest Relationships

- Sleep Duration (0.88)
- Stress Level (-0.90)
- Heart Rate (-0.66)

Sleep Duration has a strong positive correlation with Quality of Sleep. This makes sense because longer sleep duration is strongly associated with better sleep quality.

Stress Levels have a very strong negative correlation with Quality of Sleep. This also makes sense because experts have always supported the idea that higher stress levels are strongly linked to worse sleep quality.

Heart Rate also has a strong negative correlation with Quality of Sleep. Things like not enough relaxation or intense stress can heighten your heart rate and result in lower sleep quality.

Weak Relationships

- Daily Steps (0.02)
- Physical Activity Level (0.19)
- Systolic Blood Pressure (-0.12)

These lifestyle habits demonstrated some of the weakest relationships suggesting that none of them directly influence sleep quality.

Simple Linear Regression Model:

Although we already plotted a heatmap and saw how the pairwise relationships correlate to each other, linear regression offers a clearer way to evaluate how strongly each lifestyle habit predicts sleep quality while controlling for the influence of the others.

To apply a multi linear regression model on my data, I first selected the features for modeling. I made sure to include Physical Activity Level, Daily Steps, Stress Level, Systolic, Sleep Duration, and Heart Rate. As always, I made sure to include our target variable, Quality of Sleep and set it up as my Y.

```
# Selecting Features For Modeling

features = ['Physical Activity Level', 'Daily Steps', 'Stress Level','Systolic', 'Sleep Duration','Heart Rate']

X = df[features]
y = df['Quality of Sleep']
```

```
# Splitting the Data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

After applying the linear regression model, these were the results that we observed:

```
Linear Regression Results:
R² Score: 0.906316260994345
MSE: 0.14133337052479783
```

	Feature	Coefficient
0	Physical Activity Level	-0.000672
1	Daily Steps	0.000101
2	Stress Level	-0.355713
3	Systolic	0.003417
4	Sleep Duration	0.669910
5	Heart Rate	-0.017772

Based on the coefficients of each feature, I can make the following conclusions:

- Sleep Duration is the biggest and most important coefficient with a coefficient of **0.669910** indicating a positive predictor. Sleep duration is the strongest driver of better sleep quality. This aligns with what we saw on the heatmap.
- Stress Level is the strongest negative predictor with a coefficient of **-0.355713** meaning that higher stress levels are strongly associated with worse sleep quality. This also matches with what we saw on the heatmap.
- Systolic Blood Pressure is a very weak positive predictor with a coefficient of **0.003417** indicating an extremely small and not meaningful predictor.
- Heart Rate, Daily Steps, and Physical Activity Level have no meaningful effect on sleep quality due to their low coefficients of **-0.017772**, **0.000101**, and **-0.000672**.

Conclusion: The strongest positive predictor is Sleep Duration, meaning the more sleep you get, the higher your sleep quality will be. The strongest negative predictor is Stress Level, meaning higher stress consistently lowers sleep quality. Physical Activity Level, Daily Steps, Systolic Blood Pressure, and Heart Rate are all minor or insignificant predictors and contribute very little to sleep quality according to the linear regression model.

Model Performance: The model has an r squared score of 0.9063 or 90.6%. This measurement indicates that the model explains 90.6% of the variation in sleep quality. This is a very good score and further demonstrates that the predictors captured most of the patterns. The model also received an MSE score of 0.14 which shows low error meaning the model predicts sleep quality pretty well.

Regression Equation

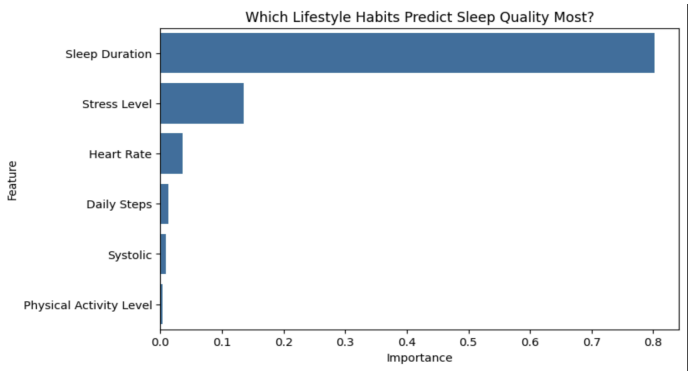
Sleep Quality = b − 0.000672(Physical Activity Level)+0.000101(Daily Steps)−0.355713(Stress Level)+0.003417(Systolic) +0.669910(Sleep Duration)−0.017772(Heart Rate)

<u>Random Forest Model for Feature Importance:</u>

In addition to the linear regression model, I also explored a random forest model to strengthen the evidence for my findings. I chose a random forest specifically because it provides feature importance rankings, allowing us to identify which variables contribute most to predicting the target outcome which in this case is sleep quality. Additionally, random forest models are great at handling skewed data, correlated features, and non-linear trends which is amazing for this dataset

The model showed the following:

	Feature	Importance
4	Sleep Duration	0.801758
2	Stress Level	0.135553
5	Heart Rate	0.036232
1	Daily Steps	0.013683
3	Systolic	0.009679
0	Physical Activity Level	0.003096



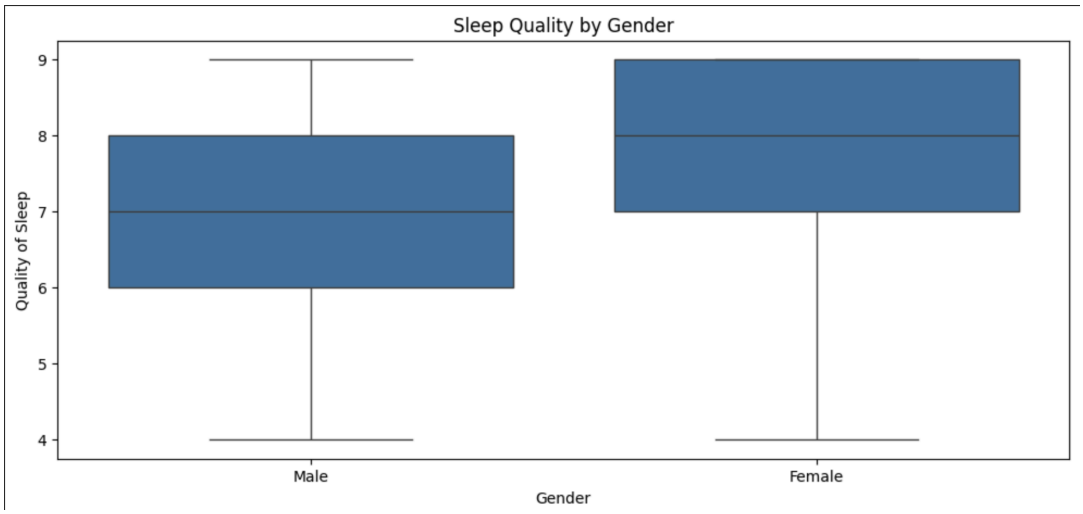
Conclusion

Based on the random forest findings, once again we can conclude that Sleep Duration and Stress Levels are the two most dominant predictors when it comes to sleep quality in all 3 models that we've explored.

Additional Findings:

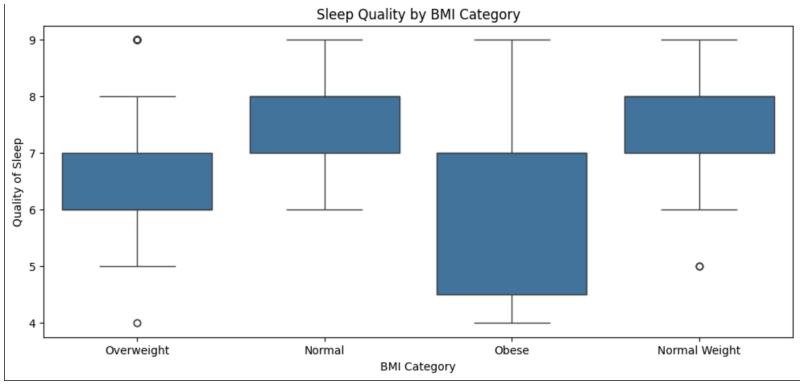
While the concentrations of this research were lifestyle habits and how they influence sleep quality, the dataset also included other variables like BMI Category, Gender, Age, and Occupation. I also explored these variables and consider them secondary predictors since they are not meant to be the main focus for my project.

Quality of Sleep vs Gender



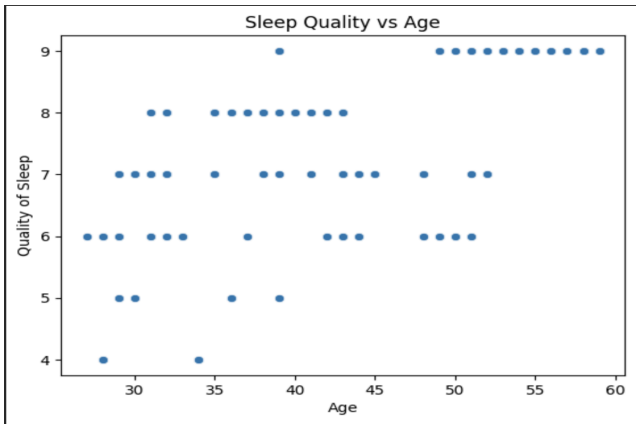
This boxplot compares the distribution of sleep quality scores between male and female participants. From the plot we can see that the median sleep quality for females is around 8, compared to about 7 for males. This suggests that females in the dataset tend to report slightly better sleep quality. However, the quality of sleep range is similar for both genders and indicates that both groups experience a wide range of sleep quality levels.

Quality of Sleep vs BMI Category



The boxplot shows how sleep quality varies across different BMI categories: Overweight, Normal, Obese, and Normal Weight. Overall, individuals in the Normal BMI categories tend to report the highest sleep quality with their median sleep quality being around 8. Additionally, people classified as Obese show noticeably lower sleep quality with their median being much lower (around 5). These observations can indicate that there's a trend where normal BMI can generally mean better sleep quality and higher BMI can mean lower sleep quality. This supports the idea that BMI may be linked to sleep quality, with healthier weight ranges corresponding to better sleep.

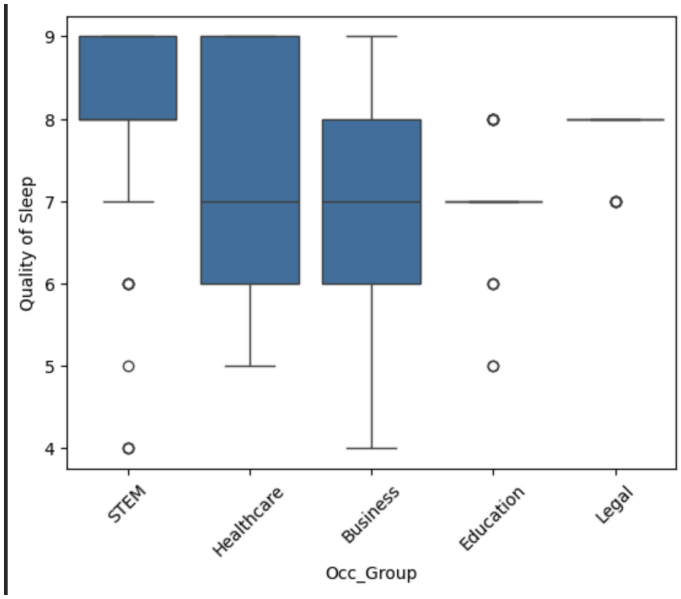
Quality of Sleep vs Age



While there are no strong linear trends between the two variables, there are a few observations worth noting. Younger adults, ages 27-35, have a wide range of sleep quality, meaning they have more variability in sleep habits. Middle-aged adults, ages 35 to 45, reported more scores of 6-8 for their sleep quality but there's still some variation. Lastly, older adults, ages 50-59, show more consistent sleep quality, suggesting better or more stable sleep patterns.

Quality of Sleep vs Occupation

Before starting our exploration of the correlation between these 2 variables, I had to categorize or group the occupations of the participants in the dataset. At the end, we had grouped the occupations into 5 groups: STEM, Healthcare, Business, Education, and Legal work.



From the boxplot we can make the following conclusions:

- STEM and Legal professionals report the highest and most consistent sleep quality.
- Healthcare workers show the most variability, possibly due to demanding or inconsistent work schedules.

- Business and Education fall in the middle, with moderate but stable sleep patterns.

Conclusions:

In this project, we wanted to answer the question, “Which Lifestyle Habits Most Strongly Predict Sleep Quality?” Based on the correlation heatmap, the linear regression model, and the random forest model, I can conclude that the two habits that are most influential towards sleep quality are **Sleep Duration** and **Stress Levels.** Both variables consistently showed strong predictive power across all analyses, providing robust, evidence-based support for this conclusion. In contrast, the other habits that we rigorously explored, Physical Activity Level, Daily Steps, Heart Rate, and Blood Pressure, showed minimal to insignificant predictive power. Therefore, I can conclude that the only habits worth mentioning are Sleep Duration and Stress Levels. These findings align with what many doctors and experts are always advocating for and that is for people to get their proper hours of sleep and to reduce any stress inducing activities in order to achieve better sleep quality. By identifying these as the strongest predictors, this study highlights practical lifestyle adjustments that can meaningfully improve sleep and overall health.

References

Health, Cultivating. “Better Sleep: Why It’s Important for Your Health and Tips to Sleep Soundly: Cultivating Health.” *Health*, UC Davis Health, 25 Aug. 2025, health.ucdavis.edu/blog/cultivating-health/better-sleep-why-its-important-for-your-healt-and-tips-to-sleep-soundly/2023/03.

Sleep Health | Journal | ScienceDirect.Com by Elsevier, www.sciencedirect.com/journal/sleep-health.

“Why Is Sleep Important?” *National Heart Lung and Blood Institute*, U.S. Department of Health and Human Services, www.nhlbi.nih.gov/health/sleep/why-sleep-important.