

Sleep Efficiency Analysis

16th November 2025

1 Dataset Overview

The dataset being used is a Sleep Efficiency dataset from Kaggle. It is a study about sleep efficiency and sleep patterns. The dataset contains information about a group of test subjects and their sleep patterns. Each test subject is identified by a unique "Subject ID" and their age and gender are also recorded. The "Sleep efficiency" feature is a measure of the proportion of time spent in bed that is actually spent asleep. The "REM sleep percentage", "Deep sleep percentage", and "Light sleep percentage" features indicate the amount of time each subject spent in each stage of sleep. The "Awakenings" feature records the number of times each subject wakes up during the night. Additionally, the dataset includes information about each subject's caffeine and alcohol consumption in the 24 hours prior to bedtime, their smoking status, and their exercise frequency. The following is the link to the dataset: <https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency/data>.

2 Data Exploration

I used a variety of methods to explore the data. I used the `.describe()` and `.head()` methods of Pandas to get an idea of the data. I also used the `.columns` method of a DataFrame to get an idea of the columns in the dataset. The main issue I found with the dataset was the existence of empty/NaN values, which is why I spent time cleaning the data. Using the `.isna()` method and a for loop, I was able to figure out the number of NaNs in each column, and use the `.dropna()` method to get rid of rows with NaNs in them. I also used the `.cut()` method to group the ages into different ranges.

3 Question

The question I will try and answer regarding the data is as follows: What are the factors that affect sleep efficiency? Elaborate on the role of gender in any/all of the factors.

4 Visualisation and Interpretation

Figure 1 presents the relationships between several behavioural and physiological factors—caffeine consumption, exercise frequency, alcohol intake, REM sleep percentage, and age—and sleep efficiency, separated by gender. From these graphs, certain patterns emerge: both genders show improved sleep efficiency with increased exercise frequency, which aligns with research showing that regular physical activity enhances sleep quality [1]. Furthermore, higher alcohol consumption is associated with lower efficiency for both groups. REM sleep percentage appears to fluctuate without a strong linear pattern, and age shows substantial individual variability. The inclusion of gender as a hue across all plots makes it possible to compare how strongly each factor affects men and women.

The graphs illustrate that males generally show slightly higher sleep efficiency at most levels of caffeine and exercise, while females show more variability across the same factors. Alcohol consumption

affects both genders similarly, though females display a steeper decline at moderate intake levels. These trends may be influenced by external factors not captured in the dataset-such as stress, work schedules, sleep timing, or measurement noise-which may explain some irregular patterns, particularly in the REM percentage plot. Gender differences in sleep patterns have been documented in prior work, with women often showing more sensitivity to lifestyle factors [2]. It can also be inferred that males have a higher median sleep efficiency for the age groups 0-25 and 51-75, while the median sleep efficiency for females is higher in the 26-50 age group. It is also interesting to note the median female sleep efficiency for the age group 51-75, showing greater variability and suggesting increased heterogeneity in sleep outcomes among older females compared to males. Overall, the visualisations indicate that lifestyle factors such as exercise and alcohol consumption meaningfully affect sleep efficiency, and gender can moderate these relationships, though the dataset does not provide enough evidence to establish causal effects.

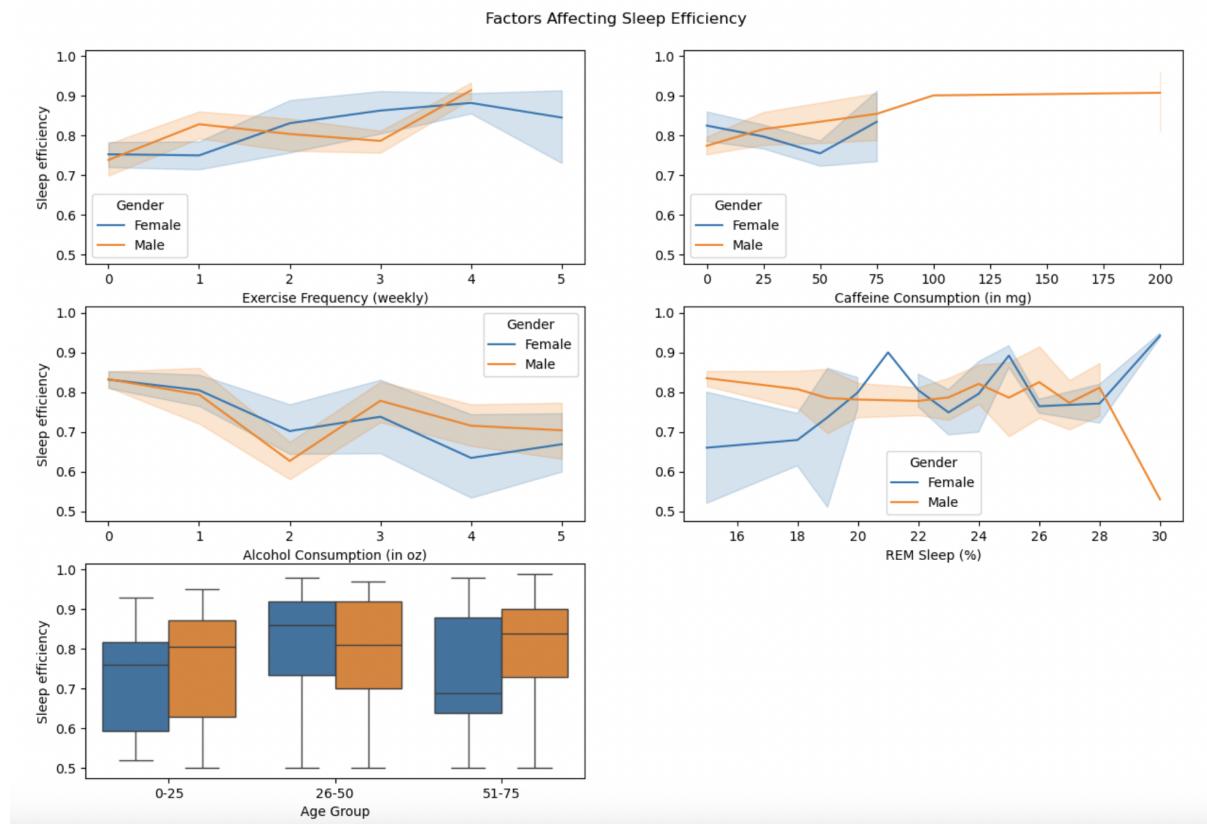


Figure 1: This figure highlights the various factors affecting sleep quality. Also includes an additional gender characteristic.

References

- [1] Shawn D Youngstedt. ‘The effects of exercise on sleep’. In: *Sleep Medicine Reviews* 3.3 (1999), pp. 213–229.
- [2] Zhen Zhang et al. ‘Gender differences in the association between sleep quality and lifestyle factors’. In: *Nature and Science of Sleep* 12 (2020), pp. 389–401.