Longitudinal Analysis of Sleep Quality and Related Factors

12.12.2024

Introduction

Setting and Context

This project focuses on dynamics of different sleep patterns and their relationships with various influencing factors, using interesting dataset: Four years of sleep data for one individual, collected with Sleep Cycle application. The study uses information on sleep quality, sleep duration, sleep consistency and associated lifestyle factors to get insights of individual sleep dynamics over a long period of time. This longitudinal approach allows analysis of trends and patterns that would be not so apparent in short-term dataset. Growing awareness of the importance of sleep to overall health and well-being is the rationale for this study. According to Worley (2018), poor sleep quality substantially influences many aspects of health and safety, which shows the need for more profound understanding of factors resulting in optimal sleep.

Relevant Literature

The literature relevant to this project underlines the growing interest in the complex interplay between sleep, physical activity and cardiometabolic health. Alnawwar et al. (2023) conducted a systematic review demonstrating the positive association of moderate physical activity with improved sleep quality. Nauha et al. (2024) investigated also this relationship, and found that irregular sleep is associated with more poor health outcomes, especially when combined with sedentary work.

This project aims to find if we can support these studies findings with single subject data.

Research Question

How do sleep quality, consistency, and duration vary over time, and what is the relationship between these sleep parameters and lifestyle factors, such as physical activity on single individual?

This research question is motivated by need to understand individual sleep dynamics beyond population-level averages and find long-term patterns affecting sleep quality.

Contribution Summary

This project contributes to existing literature by providing detailed, longitudinal analysis of sleep patterns in a single individual. The findings offers insights into the individual variability in sleep dynamics and the potential influence of lifestyle factors on sleep quality. Time series analysis was used to identify a more detailed understanding of various interactions. The outcomes are represented graphically for ease of interpretation, after which they will be related to the existing literature on the subject, to further the knowledge in this area of individual sleep patterns and their dynamics.

Problem Formulation

While existing research shows general relationships between sleep, activity and health, it relies on aggregate data and thus masks individual sleep dynamics. This project shows methods for understanding individual sleep patterns more deeply and in long-term.

The problem will focus on very limited understanding of how different aspects evolve over long periods and how these aspects relate to lifestyle factors. While studies at the population level may be very informative, they cannot show unique patterns within individual. This project fills this gap for this individual by answering these questions:

1. How does sleep quality, consistency and duration change over four years?
2. What are correlations on these parameters and physical activity?
3. Can patterns/clusters of sleep characteristics be identified?

This approach can inform further research and lead to more personalised ways to improve sleep health.

Dataset

The dataset (Diotte 2022) used in this project represents longitudinal sleep data from single individual, collected from 2014 to 2018. The data contains lot of details about sleep patterns and different factors that may influence sleep quality. This chapter gives thorough description of the dataset, its features and preprocessing that was done to it for further analysis.

Data source

Data was obtained using Sleep Cycle iOS application. Sleep Cycle predicts different sleep parameters based on motion and sound analysis. The dataset contain 887 rows, each representing one sleep session, and contains following columns:

1. **Start**: Timestamp of when the sleep session started.
2. **End**: Timestamp of when the session ended.
3. **Sleep Quality**: Subjective measure in percentage (0–100%) calculated by the app. It reflects how well the individual slept during this session. Calculated based on (Sleep Cycle 2024):
   - Amount of time spent in bed, Amount of time spent in deep sleep, The frequency of motion and intensity for each movement and Amount of times when the app registered you as fully awake
4. **Time in Bed:** Total time spent in bed during a session, including interruptions or wake periods.
5. **Wake Up**: User-input mood on waking up, either ':)' ':|' or ':('
6. **Sleep Notes:** User-provided notes about lifestyle factors, five options:
   - Stressful day, Drank coffee, Drank tea, Worked out, Ate late
7. **Heart Rate**: Optional heart rate data recorded after waking up
8. **Activity (Steps)**: Number of steps recorded during the corresponding day

There are missing values out of 887 in Wake Up (641), Sleep notes (235) and Heart rate (725). In addition to this, there are 418 values with value 0 in Activity (steps) which is not humanely possible so we can conclude that those are also missing values.

Data preprocessing

Preprocessing was done to prepare dataset for analysis. Final shape of dataset and rationale for each column selection is shown in figure 1. Following operations were done so that datapoints were in this format:

- **Parsing timestamps:** start and end columns were converted into datetime objects
- **Sleep quality:** from percentages to decimal form
- **Added Weekday**: Added weekday (0-6) so that it can be compared to e.g. sleep quality
- **Time in bed:** reformatting from 'HH:mm' to decimal form
- **Mood Encoding:** encoded. :)=1, :| = 0, :(=-1
- **Removing naps**: when time in bed was less than an hour (12 values) they were removed to prevent them skewing the data
- **Handling Missing Data:** The missing values were filled with zeros

	type	unique values	non-zero values	rationale for selection
Weekday	int	7	875	context
Time in bed	float	217	875	objective measure
Sleep quality	float	71	875	objective measure
Wake up	float	3	216	subjective measure
Sleep Notes	object	19	650	subjective measure
Activity (steps)	int	446	463	context
Heart rate	int	30	161	context

Figure 1: Preprocessed dataset info

Methods

This study analyses four years of sleep data on one individual. The analysis is done in two steps: firstly, basic methods providing overview and after, advanced methods for more detailed analysis.

1. Basic Methods

First step in analysis is descriptive statistics and exploratory data visualisation, which are important in understanding the basic features of dataset. Several visualisation techniques were used to show underlying patterns and relationships.

Sleep quality was assessed by several complementary approaches. A distribution histogram gave information about the general dispersion of the sleep quality measures and thus allowed us to understand the normal range and frequency of sleep quality experiences. A scatter plot of time in bed vs. sleep quality helped in the identification of possible relations between sleep duration and perceived sleep effectiveness. Weekday-based variations in sleep quality were visualised using box plots to possibly detect the weekly rhythm or inconsistencies in sleep. Another scatter plot was used to investigate the possible impact of physical activity on sleep, where days with zero recorded activity were excluded on purpose to make the analysis meaningful.

These visualisations served multiple purposes. Main reason was to provide a intuitive overview of data distributions, but also to identify correlations between variables and revealing some trends.

2. Advanced Methods

The advanced methodological approach makes use of more sophisticated techniques to find insights from the sleep dataset.

Correlation matrix

Correlation matrix was created to quantitatively investigate the relationship between key sleep parameters. This approach allows systematic investigation on how variables time in bed, sleep quality, activity (steps) and heart rate are interrelated. By understanding these correlations, underlying factors influencing sleep patterns can be found.

Linear regression

Regression analysis was used to assess linear predictive relationship of physical activity on sleep quality. It helps to determine the extent to which variations in physical activity can explain variations in sleep quality.

Time-series

Since the longitudinal nature of this dataset, a robust time series decomposition was done. This technique allows to separate systematic variation from random fluctuations to understand long-term sleep dynamics

Results

This chapter talks about different results this project found with different methods that were introduced in the last chapter.

Figure 2 shows that sleep quality differed across different weekdays. The highest average sleep quality was always on Sunday nights, with a mean of 0.81, while Wednesday had the lowest, with a mean of 0.74. This could indicate a kind of weekly rhythm where this individual may recover over the weekend, resulting in improved sleep quality by Sunday.

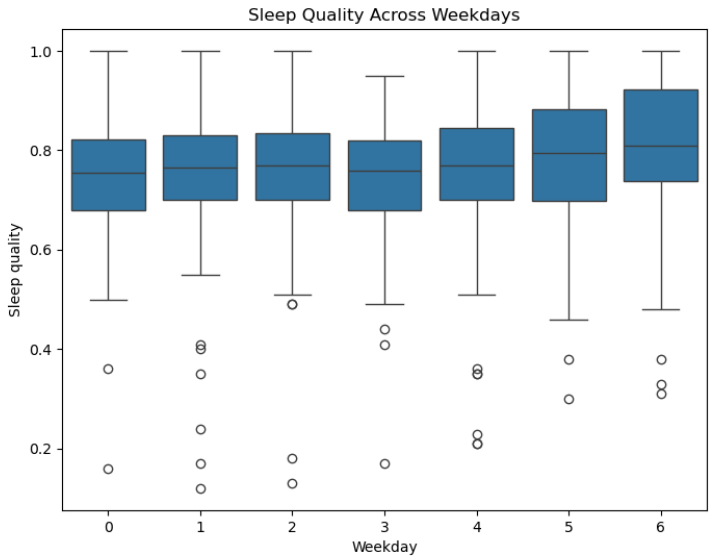

Figure 2: Sleep quality and different weekdays

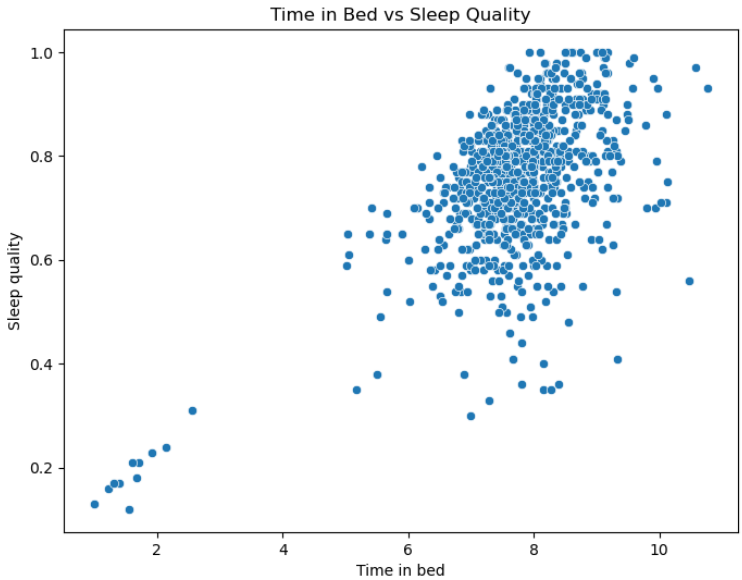

Figure 3: Time in bed to sleep quality

A scatter plot and correlation matrix analysis in Figure 3 and Figure 5 shows that there was a positive correlation, r=0.58 between time in bed and perceived sleep quality. When spending between 7.5 and 8.5 hours in bed generally resulted in higher sleep quality. This shows the importance of maintaining sufficient sleep duration.

Physical activity was weakly negatively related to sleep quality, as shown in Figure 4 and figure 5, with a correlation of -0.25. This was further supported through regression analysis, which had an R-squared value of 0.06, hence physical activity marginally predicts changes in sleep quality. This weak relationship contradicts the hypothesis of a strong positive relationship that was suggested by previous literature.

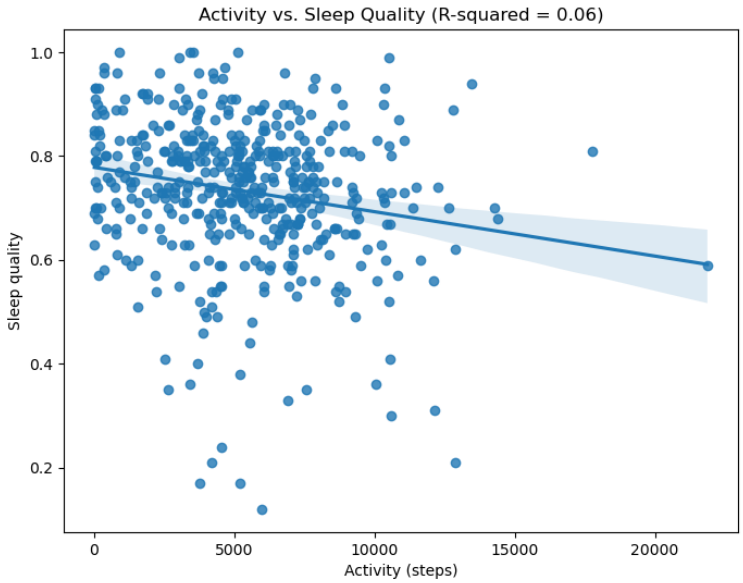

Figure 4: Activity to sleep quality

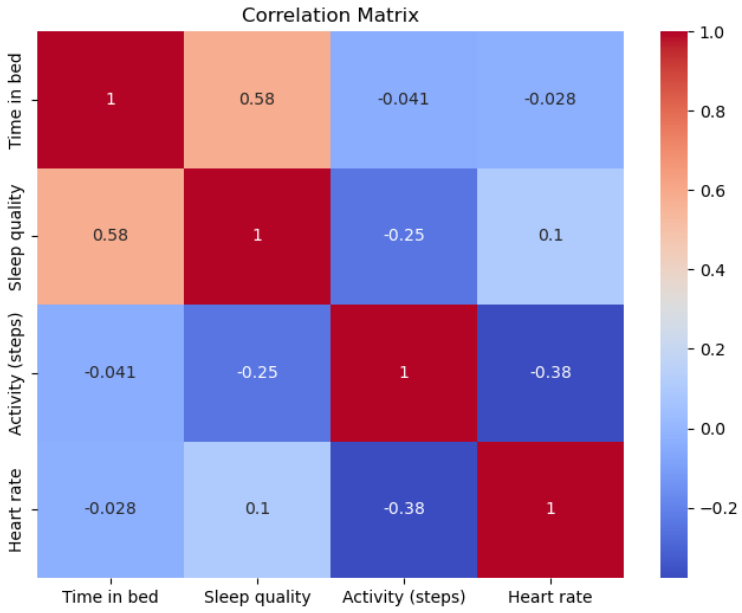

Figure 5: Correlation matrix

Time-series decomposition in figure 6 shows a slight downward trend in sleep quality and upward trend in time in bed during the four year period. The seasonal components showed some periodic like fluctuations, which most might be caused by external factors such as holidays or workload variations.

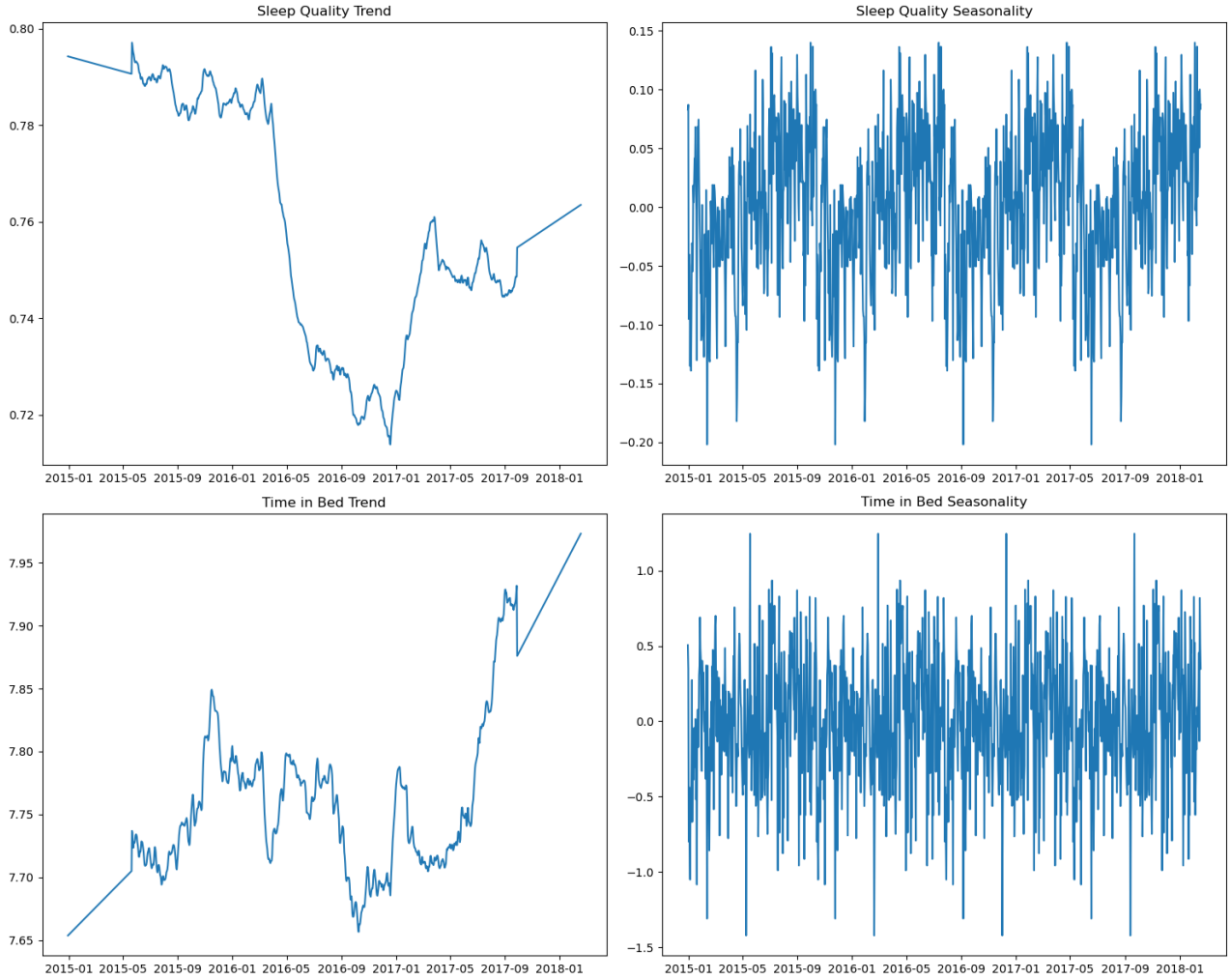

Figure 6: Time series

Conclusion & Discussion

Let's go back to original research question:

How do sleep quality, consistency, and duration vary over time, and what is the relationship between these sleep parameters and lifestyle factors, such as physical activity on single individual?

Analysis provided useful information into one individuals sleep dynamics over longitudinal period. Firstly, there is clear positive correlation between sleep duration and perceived sleep quality. More precisely, when individual spent between 7.5 and 8.5 hours in bed, he consistently reported higher sleep quality. This shows that maintenance of sufficient sleep duration is of critical importance for better rest.

However, the weak link between physical activity and sleep quality was more unexpected. Literature suggested a strong positive link between activity and sleep, but this analysis revealed just a weak negative correlation. This marginal relationship suggest that for this individual, physical activity does not directly predict their sleep quality as some population-level studies suggests.

These findings emphasise the significance of individualised strategies for sleep health. Identifying weekly rhythms allows for personal optimisation of schedules to improve sleep. As a specific example, one could compensate for reduced midweek sleep by allowing more time in bed during weekdays.

Limitations

There is some methodological limitations to this study. First is of course that this can't be generalised and used at large, since the data was collected from one individual only - but on the other hand, that was not the purpose of this study. Also the reliance on one app only can introduce some potential biases, since its algorithms might not fully capture the complexity of sleep physiology.

Other important issue is limitation in data collections. Some of the inputs in the data are subjective like mood and sleep notes which may bring some variability and hence potential reporting bias. Large number of zero values also create inconsistency in data collection. Furthermore, measures in sleep quality by this application using motions and sound might not be consistent with clinically valid measures.

Besides the clinical one, there is also some technological limitations concerning such sleep tracking application. In as much as they are great, the data derived from them cant compare to actual professional studies in sleep, since factors like environmental noise, partner movements and the way the device is positioned may affect accuracy

Future steps

Based on these findings and limitations, there are some clear directions for possible future research. First, using combination of data collection methods may lead to more comprehensive studies. For example, continuous monitoring device of heart rate, sensors for room temperature tracking, and even some personal markers may be used in

understanding individual variations. Data would already been more reliable if it had been captured with some sensory device like smartwatch

Advances in technology could also open possibilities for improving individual sleep research. Machine learning algorithms could build sophisticated models that would predict sleep quality by taking into consideration these various different factors. So, the future of research might involve more individually tailored strategies to optimise sleep. Adaptive algorithms could offer real-time advice based on one's sleep pattern, lifestyle and physiological responses.

Refining methodologies also remains important. In the future, research in this domain could be done by identifying generally accepted methods of objectively monitoring sleep, with a few scientifically valid frameworks.

References

Alnawwar MA, Alraddadi MI, Algethmi RA, Salem GA, Salem MA, Alharbi AA. 2023. The Effect of Physical Activity on Sleep Quality and Sleep Disorder: A Systematic Review. Cureus. doi: 10.7759/cureus.43595. PMID:

Diotte, D. 2022. Sleep Data. Kaggle. Retrieved from https://www.kaggle.com/datasets/danagerous/sleep-data/data?select=sleepdata.csv

Nauha, L., Farrahi, V., Jurvelin, H. et al. 2024. Regularity of bedtime, wake-up time, and time in bed in mid-life: associations with cardiometabolic health markers with adjustment for physical activity and sedentary time. JASSB. https://doi.org/10.1186/s44167-023-00040-6

Sleep Cycle. 2024. How does the app calculate Sleep Quality?. https://support.sleepcycle.com/hc/en-us/articles/206704659-How-does-the-app-calculate-Sleep-Quality

Worley SL. 2018. The Extraordinary Importance of Sleep: The Detrimental Effects of Inadequate Sleep. Health and Public Safety Drive an Explosion of Sleep Research. PMID: 30559589