Kaunas University of Technology
Faculty of Mathematics and Natural Sciences

Personalized Assessment of Human Physical Activity Impact on Sleep Quality Using Data from Wearable Device

Master's Final Degree Project

Urtė Liutkevičiūtė
Project Author

Assoc. Prof. Dr. Vytautas Janilionis
Supervisor

Assoc. Prof. Dr. Asta Daunorienė
Supervisor

Kaunas, 2024

Kaunas University of Technology
Faculty of Mathematics and Natural Sciences

Personalized Assessment of Human Physical Activity Impact on Sleep Quality Using Data from Wearable Device

Master's Final Degree Project
Business Big Data Analytics (6213AX001)

Urtė Liutkevičiūtė
Project Author

Assoc. prof. dr. Vytautas Janilionis
Supervisor

Assoc. prof. dr. Asta Daunorienė
Supervisor

Prof. dr. Tomas Ruzgas
Reviewer

Prof. dr. Eglė Staniškienė
Reviewer

Kaunas, 2024

Kaunas University of Technology
Faculty of Mathematics and Natural Sciences
Urtė Liutkevičiūtė

Personalized Assessment of Human Physical Activity Impact on Sleep Quality Using Data from Wearable Device

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;
2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;
3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;
4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Urtė Liutkevičiūtė
Confirmed electronically

Urtė Liutkevičiūtė. Personalized Assessment of Human Physical Activity Impact on Sleep Quality Using Data from Wearable Device. Master's Final Degree Project/ supervisors Vytautas Janilionisand Asta Daunorienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Mathematics, Applied mathematics.

Keywords: sleep quality, physical activity, wearable devices, individualized recommendations, time series clustering, multiple regression analysis, Python.

Kaunas, 2024. 62 p.

Summary

The object of this project is to provide a personalized evaluation of the relationship between physical activity and sleep quality. Poor sleep quality is associated not only with health and social problems, such as an increased risk of conditions such as diabetes, obesity, and depression, reduced productivity and engagement in workplaces, but also with the impact on the economy. In the United States, poor sleep costs up to $411 billion annually. In this project, a novel methodology is created and implemented that includes various metrics of physical activity and sleep quality and finds the strongest relationships on an individual level. The results of the personalized multiple regression analysis showed distinct associations from different persons. In addition, the results of the time series cluster analysis with physical activity data showed to be a significant regressor in multiple regression models, showing the effectiveness of the newly recommended approach. This project demonstrated the importance of the need for a personalized analysis of physical activity and sleep quality. The developed methodology can be extended by incorporating additional biometrics or new methods to expand the research area.

Urtė Liutkevičiūtė. Individualizuotas žmogaus fizinio aktyvumo įtakos miego kokybei vertinimas naudojant dėvimų prietaisų duomenis. Magistro studijų baigiamasis projektas/vadovai Vytautas Janilionisir Asta Daunorienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Matematikos mokslai, taikomoji matematika.

Reikšminiai žodžiai: miego kokybė, fizinis aktyvumas, dėvimi prietaisai, individualizuotos rekomendacijos, laiko eilučių klasterizavimas, daugialypė regresinė analizė, Python.

Kaunas, 2024, 62 p.

Santrauka

Šio projekto objektas – individualizuotas žmogaus fizinio aktyvumo ir miego kokybės ryšio įvertinimas. Prasta miego kokybė siejama ne tik su sveikatos ir socialinėmis problemomis, tokiomis kaip padidėjusi rizika susirgti tokiomis ligomis kaip diabetas, depresija ar širdies ligos, sumažėjęs produktyvumas ir įsitraukimas į darbą, tačiau prastas miegas siejamais ir su neigiamu poveikiu ekonomikai. Jungtinėse Valstijose prastas miegas kasmet kainuoja iki 411 milijardų dolerių. Šiame projekte sukurta ir įdiegta nauja metodika, apimanti įvairius fizinio aktyvumo ir miego kokybės rodiklius bei padedanti surasti stipriausius personaliztus ryšius tarp žmogaus fizinio aktyvumo ir miego kokybės. Personalizutos daugialypės regresijos analizės rezultatai parodė, kad ryšiai taip rodiklių priklauso nuo žmogaus ir gali skirtis. Be to, laiko eilučių klasterinės analizės su fizinio aktyvumo duomenimis rezultatai parodė, kad tai yra reikšmingas regresorius daugialypės regresijos modeliuose, rodantis, kad naujai rekomenduotas metodas, kuris klasterizuoja dienas, yra veiksmingumas ryšio tyrimuose. Šis projektas parodė individualizuoto fizinio aktyvumo ir miego kokybės analizės poreikio svarbą. Sukurta metodika gali būti išplėsta įtraukiant papildomus biometrinius duomenis arba naujus metodus, leidžiant praplėsti tyrimo sritį.

Table of contents

List of Figures

List of Tables

Introduction

The problem of this project is human physical activity impact on sleep quality. Recently, the significance of sleep has been increasingly highlighted by the World Health Organization (WHO) and other national health organizations as a key recommendation to improve human health. Poor quality sleep not only increases the risks of depression and obesity, reduces productivity and engagement on the job, but also has a huge impact on the economy. For example, in the United States, poor sleep costs up to $411 billion annually. In recent years, various scientific disciplines have begun to closely examine ways to improve sleep quality, and physical activity is one of the areas. The studies mainly analyzed larger populations without approaching each person individually. Creating a methodology to analyze and identify the strongest associations represents a new approach. This project was carried out in collaboration with the company *dhealthIQ* which specialized in remote patient monitoring and personalized therapy solutions.

The object of this project is to provide a personalized evaluation of the relationship between physical activity and sleep quality. This interdisciplinary project integrates multiple fields: it addresses social and economic problems by improving sleep quality, utilizes applied mathematics to develop methodologies and analyze data through modeling, and leverages informatics to create a realization of data processing and a semi-automated tool.

The goal of this project is to develop a methodology to evaluate the personalized relationship between human physical activity and sleep quality using data from wearable devices.

Objectives:

1. Conduct a literature review on the topic of human physical activity and sleep quality.
2. Select and calculate key metrics for physical activity and sleep quality using available data from company-provided data.
3. Develop and implement methodology to assess the impact of human physical activity on sleep quality based on selected metrics.
4. Apply the developed methodology and software tools to company-provided data.
5. Offer suggestions for further improvement of the methodology and software tools.

1. Literature review

1.1. Importance of quality sleep

Sleep quality impact on human health. Human health is one of the complex fields that includes various physiological, psychological, and social aspects. Although health has been associated with physical activity and a balanced diet, nowadays more and more attention is also paid to quality of sleep. For a better understanding of why sleep is an important component of human health, it is important to know what happens in the body during sleep.

Sleep varies throughout the night and progresses through 3-5 sleep cycles. Each cycle lasts approximately 90 to 120 minutes, which consists of several stages of sleep. It is important to note that the length of sleep cycles changes with age and sleep becomes shorter and with less quality. During each cycle, sleep goes through different stages that are called NREM 1, NREM 2, NREM 3, and REM.

The first stage of NREM 1 is the shortest, during which the body relaxes and prepares for sleep. During this stage, a person is still sensitive, so he can wake up from the slightest external stimulus. A person spends about half of their total sleep time in the NREM 2 sleep stage. At this stage, the person already becomes unconscious, but the sleep is still light and the person can easily wake up.

NREM 3 - the third stage sleep is also called deep sleep, because it is the stage of sleep during which the body completely relaxes, blood pressure drops, heart rate drops to a minimum, breathing slows down, and a person falls into a deep sleep. At this stage, it is quite difficult to awaken a person, and if a person is awakened, the person does not immediately orient himself to the environment. In this stage, the body, muscle, and immune systems recover after a full day's work, and growth hormones are also produced.

Finally, in the fourth stage, also known as Rapid Eye Movement (REM), brain activity is activated. From the physiological sensations, a person begins to breathe heavily, the body temperature and blood pressure rise, and it is also distinguished by the fact that the person's eyes begin to involuntarily move suddenly. Currently, unnecessary memories are being removed and important information is being transferred from short-term memory to long-term memory. It is also observed that the most of the dreaming happens during this stage of sleep, and it takes about 25% of the total sleep time [3].

A review study showed how the level of hormones dynamically changes during the day and at night based on the circadian cycles [4]. For example, there is a growth hormone that is the most secreted during sleep, which explains the restoration role of sleep. Melatonin, which can be called the sleep regulator, dynamically changes through the circadian rhythm, helping the person regulate sleep during the day. It is important to mention that melatonin can be affected by caffeine intake, which pushes sleepiness later in the day. Cortisol, a stress hormone, is reduced during sleep, which gradually increases before the person wakes up to support the person during the day. Leptin and ghrelin that promote or suppress food intake are also impacted by circadian cycles. It is important to understand that any sleep disturbances, such as shift work or not regulating bed and wake up times, can negatively impact the natural dynamics of hormones that are crucial for human health.

Studies show that it is important to take care of sleep quality. In recent years, there has been more interest in how good sleep affects overall physical and mental health. Sleep disruption can cause short-term problems, such as reduced alertness and memory, degraded attention, poorer decision making, or an increased risk of accidents [5]. It is important to mention that studying long-term effects on sleep is challenging, because human life is uncertain and may be impacted by multiple factors, however, there are some potential associations between poor sleep and conditions such as cardiovascular disease, high blood pressure, diabetes, depression, and memory disorders [6, 7]. Some other examples were seen in a study in South Korea in which they saw associations of poor sleep with anxiety and depression, which can also lead to fatigue [8]. Looking at the broader analysis, one study performed meta-analysis on poor sleep outcomes and found significant medium associations with sleep problems and loneliness [9]. In addition, some sleep disorders were linked to a higher risk of cardiovascular diseases, including hypertension, congestive heart failure, coronary artery disease, and metabolic disorders [10]. High-quality sleep has been shown to benefit both physical and mental health, as well as overall quality of life, especially for those with type 2 diabetes [11].

Poor sleep quality impact on economy. Although sleep quality might impact only the individual himself, the research shows that the increasing lack of sleep brings great losses to the country itself. In Australia, lost working days due to illnesses such as depression or other heart conditions, as well as road traffic accidents due to sleepiness, could cost the state from 41.38 billion AUD to 49.21 billion AUD [12]. Having healthy and rested citizens in the country could save a huge amount of money for every government.

This is also true in the workspace. A study calculated that an employee with less than 6 hours of sleep gets on average 2.36% higher productivity loss, and employees who sleep 6 to 7 hours get on average 1.47% higher productivity loss than the recommended 7 to 9 hours per night [13]. That means that the worker who sleeps less than 6 hours loses on average around 6 working days per year, and the worker who sleeps from 6 to 7 hours, loses on average 3.7 days. The study created a predictive model that calculated that such sleep deprivation can cause between $280 billion US dollars and $411 billion US dollars of economic loss to the USA or $88 billion and $138 billion to Japan.

The same study created some scenarios on how increased sleep duration of working individuals may change expenses due to poor sleep. The first scenario was if all individuals were able to extend their sleep to the recommended seven to nine hours. In this case, all losses due to poor sleep would be eliminated. The second scenario, which would be more possible, is to have moderate improvement, such as people who sleep less than six hours, would be able to increase the amount of sleep to at least six to seven hours. Although this scenario is less ambitious than the previous one, it can still offer significant benefits, changing the productivity loss in those individuals from 2.36% to only 1.47%. And the third scenario focused on people who sleep six to seven hours and improve their sleep to reach at least seven hours, which could reduce productivity loss for those individuals[13].

Unfortunately, there is no comprehensive survey on the sleep habits of Lithuanian citizens and whether they sleep the recommended number of hours. Therefore, it is necessary to refer to the situation in neighboring countries. In the same study, they also calcualted that in Germany 9% of people sleep less than 6 hours and 20% sleep between 6-7 hours, which is below the recommended amount [13]. The study estimates that people who sleep less than 6 hours may lose up to 2.36% of their working days per year, while those who sleep 6-7 hours may lose 1.47%. If assuming that the situation in Lithuania is similar to Germany, it can be estimated that in 2021, out of 1,079,000 full-time workers [14], those who slept an average of 6-7 hours lost 793,065 days and those who slept less than 6 hours lost 572,949 days. This totals to 1,336,014 lost working days. That year, the average salary was 1,579.4€ per month, which translates to approximately 75.2 € per day. By calculating lost days and their expenses, this could have cost Lithuania 102,737,262.5€.

Similarly, it would be possible to adapt a similar but less ambitious scenario to what would happen if there was a possibility to take one percent of people who sleep six to seven hours and reduce the overall 20% of people to only 19% of people who sleep six to seven hours. If taking the same assumed percentage of Lithuanians and having only 19% of people who sleep six to seven hours, then Lithuania could save almost 3 million euros per year. Although this calculation is only an approximate idea, it can show the need for such studies in the future. Understanding the loss that is caused every year due to poor sleep, the Lithuanian government could invest in sleep health programs that could lead to better sleep quality and potential savings in the health care system.

The same study combined different recommendations for each level of affected parties, such as individuals, employers, and public authorities [13]. For individuals, the recommendation is to have consistent waking-up time even during the weekends, also, limiting electronic devices before bed and any substances, such as caffeine and alcohol. Also, exercise. Employers should recognize the importance of sleep quality, since it benefits not only person's health, but also increases productivity and engagment in the workplave. In addition, it would be good to help employees support their daily routines based on sleep hygiene practices, such as discouragement after waking up communication. Public authorities can also show a good example in the promotion of sleep education, such as creating awareness campaigns to encourage health professionals to assist patients with sleep disorders. Also, since children also require a good amount of sleep, it would be possible to adjust school times to delay waking up time.

Well-rested employees are not only beneficial to the country's economy, but also have other positive outcomes in the workplace [15]. Good quality sleep affects the employee's mood, positively affecting job satisfaction and customer service. The tasks that require creativity, innovation, and strategic thinking, also are performed better by workers with sufficient sleep. In addition, insufficient sleep leads workers to be less engaged at work and may even cause unethical behavior in the workspace. Having motivated employees who are more creative and innovative and produce fewer errors at work is a success for any business.

Due to a poor night of sleep, people often experience fatigue and lack of productivity, making it difficult to perform their regular tasks. For example, one research conducted in a hospital in South Korea discovered a correlation between poor sleep quality and decreased productivity among nurses[16]. This is concerning, since nurses must have decision-making skills due to their profession. Another study in the USA using similar tools also found that poor sleep and sleep diseases reduce productivity at work and can cost 1967 USD per employee per year [17]. Besides direct productivity losses, poor sleep is associated with an increase in workplace accidents and errors, which also can add financial losses. Also, raise health concerns. For instance, fatigue-related incidents are common in high-risk industries such as healthcare, transportation, and manufacturing, where errors can have serious consequences.

Seeing the results that good quality sleep has a positive impact on productivity and health and can save money for companies and governments, shows the importance that not only the individual himself should take care of his sleeping habits but also companies and governments must be interested in the benefits of good night sleep. Of course, it is difficult to manage the time of employees after work hours; however, a review study concluded that the attempt by employers to educate workers about sleep hygiene can produce positive results in better sleep for workers [18, 19]. Similar results were observed with police officers, where sleep health promotion improved sleep quality, also reducing the frequency of accidents at work and near misses [20]. A method to encourage better sleep and possibly increase productivity could be to gift wearable devices to employees, so they could track sleep and other metrics [21].

Studies have shown that people who learn and implement recommendations on how to take care of their sleep often experience better sleep quality [22, 18]. However, as one study stated, sleep recommendations must be renewed because current sleep hygiene recommendations have not changed in 40 years [23]. The authors highlighted the importance of an individualized approach because recommendations should be relevant to person's current life events and circumstances. Certain guidelines, with few exceptions, apply broadly to most people, such as spending time in natural light in the morning and avoiding intense light in the evening. Yet, exercising overall has a positive effect on sleep, but it depends on person, his activity times, intensity, and exercise type.

Many factors can influence how well humans sleep, from daily routines to stress levels. This project will focus specifically on the connection between physical activity and sleep quality. Physical activity shows various results on sleep quality [24] that can vary due to the individualization mentioned above.

1.2. Sleep and physical activity monitoring and assessment methods review

Polysomnography (PSG) is a golden standard process to measure the quality of sleep, identify sleep patterns, and also explain what can cause discrepancies. Usually, it is done in laboratories in medical centers where a patient comes and spends the night there. The sensors are stuck or taped all over the body and connected to a computer. This computer can keep an eye on various body measurements such as brain waves, heart rate, breathing, body and eye movements, blood oxygen levels, and even snoring (Figure 1). Although using these data it is possible to differentiate sleep stages and patterns, it is very expensive and because all laboratory setups can cause a patient's sleep to be disruptive itself, they also do not have adequate readings [25]. In addition, this method has a lot of data; however, a doctor or specialist is needed who could interpret the results to detect sleep disturbances and assess sleep quality.

Without accurate physiological data, it is hard to objectively measure sleep quality. In the literature, it is very common to find that sleep is subjectively measured with the Pittsburgh Sleep Quality Index (PSQI) [26], which is a self-reported standardized questionnaire that includes different types of questions related to sleep in the last month. The person has to rate different parts of his or her

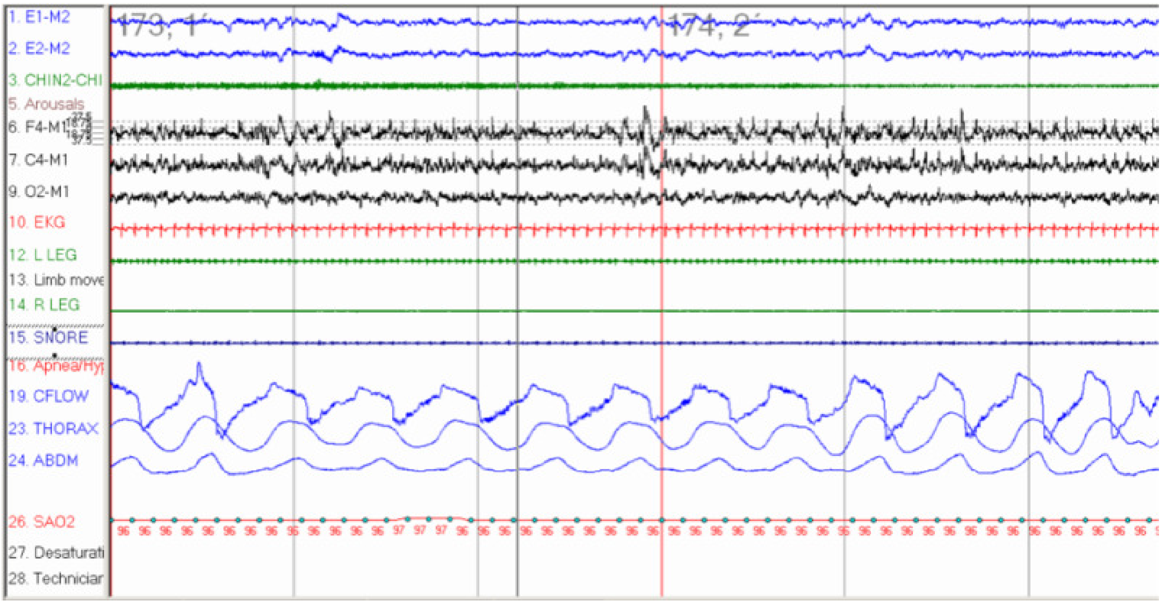


Fig. 1. A polygraph from in-laboratory polysomnography. Source: Maggard MD, Sankari A, Cascella M. Upper Airway Resistance Syndrome [1].

sleep, such as subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medications and daytime dysfunction [27]. After answering all questions, the score can range from 0 to 21, where the higher score indicates poor quality sleep and the lower score indicates better quality sleep. Using this standardized measurement, a global PSQI greater than 5 can distinguish poor-quality sleepers from good ones. Since this approach is easy and inexpensive, it has become popular and well-accepted by the scientific community [28], [26], [29]. Other similar approaches such as sleep questionnaires or sleep diary include sleep-related questions [30]. Nevertheless, relying on subjective scores for sleep quality or other parameters provides only a singular numerical representation of the past sleep period, which is not enough to detect any sleep disorders [31]. Given that sleep quality is influenced by various factors, including the duration of different phases and the proportion of REM or deep sleep, it becomes crucial to monitor sleep continuously throughout the entire night.

Concerning sleep quality, it is also important to talk about physical activity. One systematic study review on this topic has shown very different results that, although there is a link between physical activities and more studies have shown the positive impact on sleep quality, however, it is also interesting that some negative impacts also appear in those studies depending on various factors, such as intensive exercises late in the evening [26]. Exercise science has different ways to measure physical activity by applying different discipline fields, such as physiology or biomechanics. For example, some of the most common methods are used to estimate the intensity of physical activity: VO2R, HRR, percent of the maximum HR, percent of VO_2max and percent of the metabolic equivalent of the task (MET) [32]. These measurements are more individualized because they use the metrics of a person. On the other hand, there are some measurements like caloric expenditure, absolute oxygen uptake, and METs, yet they can miss classify the results because those methods do not include individual measurements like sex, body weight, or current fitness level. Most of the metrics discussed require specialized equipment and tools for accurate measurement. Additionally, recent advancements in wearable technology allow for the monitoring of Oxygen Saturation (SpO2), skin temperature, breathing rate, and heart rate variability, thereby improving the comprehension of fitness levels and the precision of exercise assessments.

Physical activity can be analyzed based on four dimensions: frequency (how often a person exercises), duration (the length of the exercise), intensity (the physiological effort), and type of activity [33]. Focusing more on intensity, one method suggested for measuring it is to categorize intensity into moderate and vigorous levels, which, in other words, can be found as moderate-to-vigorous-intensity physical activity (MVPA). Usually, it is measured by how many minutes were spent in each intensity per specific period. The Canadian Society for Exercise Physiology guidelines for adults (aged 18-64 years) to achieve health benefits are to have at least 150 min of moderate-to-vigorous-intensity aerobic physical activity per week. Similarly, the Ministry of Health of The Republic of Lithuania recommends at least 30 minutes of moderate-intensity activity 5 times a week for adults.

There is no strict way to measure when the activity reaches moderate or vigorous levels [34], but there are some flexible guidelines. For example, moderate-intensity exercises can be called the ones that make a person breathe harder but still allow holding a conversation, such as brisk walking or dancing, and vigorous-intensity can be defined by activities that significantly increase the breathing and heart rate. Focusing on more measurable metrics, heart rate is one of the measurements that can be used to classify the intensity, such as that for moderate intensity classification can be used 40-59% of maximum heart rate, and for vigorous intensity could be 60-84% of maximum heart rate.

1.3. Wearable technologies in sleep and activity monitoring and assessment

Due to limitations in monitoring sleep and physical activity measurements, recently in the consumer market a new approach to wearable technology gained more popularity in the field of health research which includes technologies with actigraphy and photoplethysmography [35]. Actigraphy can be described as a technology that can be worn on the wrist and can track a person's activity over long periods, and based on the body movements can detect sleep and awake cycles [36]. Although actigraphy cannot beat the accuracy of laboratory measurements, studies have shown a high correlation with the results of polysomnography [37]. Furthermore, heart rate monitors (Photoplethysmography - PPG) play a big part in sleep pattern detection because heart rate during the different stages of sleep has different patterns, for example, during the deep sleep stage, heart rate gets the slowest throughout the night.

The data from wearable devices is not yet treated as a sufficient way to detect sleep abnormalities or based on that give medical advice, however, studies show that the accuracy of wearable devices, such as Fitbit, has advanced rapidly in recent years [38]. For example, studies compared non-sleep-staging Fitbit models with polysomnography (PSG), and they did not have high accuracy, however, after 2017 Fitbit introduced a sleep-staging feature. With the help of machine learning (i.e., linear discriminant classifier) based on motion, heart rate variability (HRV), and respiratory rate, it still showed just moderate accuracy results in detecting sleep stages, however, it improved total sleep time (TST), sleep efficiency (SE) and wake after sleep onset (WASO) better than previous devices. Another study reviewed low-cost devices that showed similar patterns. The step count and sleep minutes tend to be more accurate overall, while detecting sleep phases is more challenging. [39]

As of today, even with improved technology and records accuracy, they still cannot be considered to be clinically approved, however, it is an easy and cheap way to find long-term patterns and tendencies of a person's sleep, as well as it does not require active user engagement because the only requirement is just to wear a small device through the day and night. It is harder to fully trust and interpret the results when the algorithms that companies use are not disclosed; however, the affordability and extended duration might bring opportunities to discover new patterns. Also, science community also tries to invent new ways to detect sleep phases, and other metrics, while it is confirmed to be a very challenging task. Moreover, this approach has a promising potential, as accuracy is increasing every year with new technology.

Fitbit and Garmin also provide their Sleep Scores that can identify their sleep quality index where a sleep score of 90-100 is assigned for excellent sleep, 80-89 for good sleep, 60-79 for fair sleep, and below that for poor sleep [40]. Garmin mentions on its website that the score is calculated based on Sleep Duration, Average stress score during sleep, Total deep sleep, Total light sleep, Total REM sleep, Awake time, and Restlessness. Garmin includes less - total duration, quality deep and REM sleep duration, restlessness, and heart rate, which they group as restoration[41].

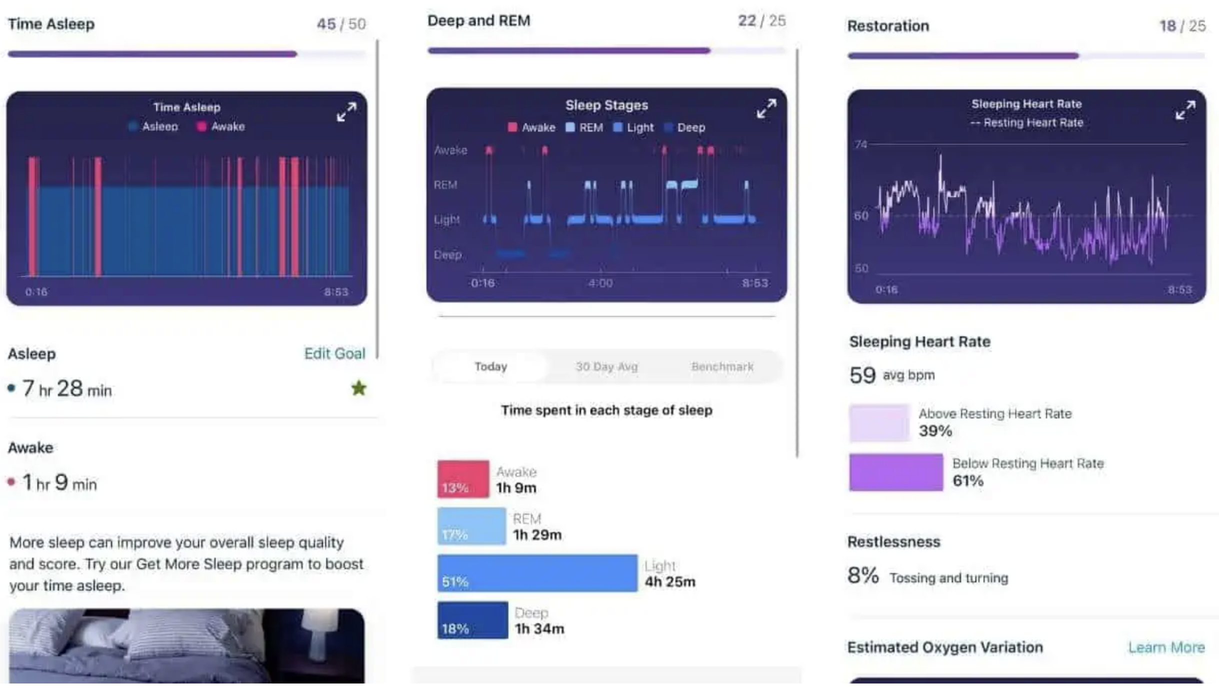


Fig. 2. Fitbit sleep tracker aggregated data display. Image source: Digital Health Central [2].

Wearable devices provide users with a lot of data that are usually accessible by their apps, or it is possible to get access to the data through their API or data can be downloaded on request. The data that is possible to retrieve from the companies usually include minute-by-minute heart rate averages and step counts, fall asleep, and wake-up times, also, aggregated data such as Total Bed in Time (TIB) - the duration of spent time in bed, Total Sleep Time (TST) - the duration of time in bed which was spent sleeping, Wake After Sleep Onset (WASO) - time of being awake during the period of sleep, Sleep Onset Latency (SOL) - the duration how long it took to move from being awake to asleep.

$$TST = TIB - (SOL + WASO)$$

Those measurements are handy for studying sleep quality. There is no single objective definition to describe sleep quality, and without using subjective PSQI index or expensive PSG readings, in studies that rely on wearable devices, it has become popular to use these aggregated readings mentioned.

It is possible to find more detailed metrics, for example, some studies on sleep quality [42] [43] also considered the number of awakenings and their duration, Sleep Onset (Son) and Sleep Offset (Soff) - actual times of falling asleep and waking up, also, various rations to Total Sleep Time, such as REM or deep sleep stages [30]. In addition, it is common to find the Sleep Efficiency (SE) ratio which is the proportion of Total Sleep Time and Total Time in Bed.

$$SE = \frac{TST}{TIB}$$

Notably, these metrics are typically derived from sleep devices, which can be considered secondary data. This is because these devices frequently predict various stages of sleep, as well as the times of falling asleep and waking up. Therefore, the reliability of these metrics in studies that were not based on polysomnography (PSG) relies on the accuracy of the device and the algorithms used by the company.

Wearable devices designed for the consumer market, such as Garmin and Fitbit, calculate metrics such as "Intensity Minutes" and "Active Minutes" which can be linked to the intensity measurements mentioned before. However, their user documentation does not explicitly outline the methodology behind these calculations. The instructions hint at the importance of keeping the heart rate tracker active throughout the day for the best accuracy. In particular, Garmin devices offer users the option to customize their heart rate thresholds, allowing more individual ways to score Intensity Minutes.

The common practice in research to measure the activity of a person on similar topics is to calculate the total steps per day or month [44]. It showed that increasing the daily amount of steps might increase sleep quality and increase the duration of sleep. The same research incorporated Fitbit-provided active minutes.

Evaluating sleep is not a very well-defined task. Even laboratory results are not able to have the final metric that could define a perfect sleep. From the review in similar studies with wearable device data it was seen that total sleep time was one of the main metric to evaluate sleep, the more people sleep, the better quality it has. In addition, sleep was classified as "good" or "poor" based on duration thresholds. There might be a couple of reasons why total sleep time is a common metric. First, the majority of the wearable devices measure this metric and can provide as aggregated results. Also, it is pretty accurate. It looks that the detection of sleep duration appears to be one of the most accurate in wearable technology compared to others. The amount spent in deep and REM sleep phases is also very important, and not having enough of it can affect human health [3], therefore, to see sleep quality as a whole, it is important to take that into consideration. Although these data are provided in the most common wearable devices, they have not been used often. It might be because there is not enough evidence of accuracy, or it can depend a lot for each person, and it did not show an realtionship with other factors.

The evaluation of physical activity is even narrower. Without special devices to measure muscles, oxygen saturation, or overall physical condition of the person, it is harder to measure physical activity of the day. Steps, distance, and heart rate are the most common ways to evaluate it. Even though studies have not explicitly stated this, but there could be situations that are difficult to interpret. For instance, a high heart rate with a low step count during an interval could indicate that the person was lifting weights or cycling, or it could simply be the result of a stressful situation, such as giving a presentation, which typically raises heart rate.

Wearable technology opened the way to view a more detailed real-time view of human health that enables medical doctors to view a more holistic view of their patients. This study reviewed the history of wearable devices in healthcare and what benefits and risks they could bring [45]. For example, it gives personalized healthcare insight, where doctors can analyze whether assigned treatment plans may work or not. Also, new prospects of big data allow the creation of better prediction models that could enable people to foresee any abnormalities in their health metrics in real-time. It also brings risks and challenges when dealing with health data from the data security and privacy questions to continuous transmission and storage of large amounts of data. In addition, current AI development could shape predictive health analytics and digital remote therapies to the new level.

1.4. Mathematical methods used in similar projects

Overall, in the medicine literature, it is more common to see basic statistical methods, such as linear correlation, linear regression, logistic regression, and some time series analysis methods [26]. They are well established and easy to interpret, giving direct insights into the relationship between multiple variables. The advantages they have are that the results of such methods is that it gives clear hypothesis testing, which allows researchers to understand if the results are significant or not. In addition, those methods are easier to implement, do not require large computational power, and do not require a large data set. On the other hand, basic statistical methods assume linearity, and this is harder in complex data where the relationship is not necessarily going to be linear.

Machine learning methods are able to handle more complex relationships, and not require linearity, which gives more opportunity for forecasting exercises. It is also more flexible, it offers many different algorithms where any different domain or challenge might find a suitable method. However, it is limited to interpretability, since not all algorithms are able to explain the relationships between variables. Also, it requires larger datasets where it is necessary to have test and validation subsets in order to reduce overfitting possibility. However, it is important to mention that in some areas, such as the detection of cancer by scanning, machine learning is very effective. The interpretability in this case is not necessary to understand why algorithm decided on the result, because it is visually seen. For example, machine learning is very effective in detecting different cancers such as lung, breast, and skin cancers [46]. However, in the tasks where it is important to understand the relationship, statistical methods provide more advantages. One study did a systematic review and found that machine learning models did not perform better than basic statistic models [47].

And this pattern was seen in the majority of the studies where basic statistic models were chosen for the analysis. Many reviews were done in a simplified way in which data input takes aggregated data (such as total count step, total sleep minutes, or average heart rate throughout the day from different wearable devices). Alternatively, more elaborate studies can be explored that utilize raw time-series data, such as body movements, heart rate, and step counts. This data can be provided at minute or second intervals, allowing for the application of time series analysis. It became easier to conduct research like this because some of the companies are more willing to supply digital health data structurally. For example, Fitbit, Inc. offers their Web API where users of the device can conveniently download the data of their collected measurements of the day. Having almost raw full data at a minute-by-minute level allows mathematicians and data scientists to uncover more insights from Big Data.

One of the study that focused on physical activity and sleep assessment using machine learning tools [48]. The purpose of the study was to analyze sedentary behaviors with sleep patterns, where they used commercially available data from Harward University. The data consisted of personal information such as height, weight, height, and also amounts of distance, steps, burned calories, and heart rate during the days. The goal was to analyze the accuracy of different classification methods, such as Random Forest, XGBoost, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). For implementation Python was used with sklearn.preprocessing, sklearn.modelselection, sklearn.preprocessing, sklearn.ensemble, matplotlib, pyplot, sklearn.metric libraries and functions. The study found that different methods had different accuracy in different data sets of the device. For example, Random Forests performed the best with Apple Watch data, while XGBoost showed the best results with both Apple Watch and Fitbit.

One of those studies used labeled raw acceleration time series data and created new models to detect sleep patterns or activity intensity [49]. In this study, the authors wanted to train the model to detect the levels of physical activity of toddlers. The toddlers wore accelerometers, while their activities during that period were annotated with eight different activities, such as walking, crawling, standing, sitting, etc. The accelerometer data was then segmented into 2-second windows, which together with annotations were used to train the models. The study used static window classifiers (K-nearest neighbors, SVM, Logistic regression, Decision tree, Random forest), and then to improve those models, they applied a hidden Markov model (HMM) to add more context for each observation. The results of HMM showed a slight improvement in accuracy for each classifier where the random forest performed the best.

However, having labeled data or comparing wearable device measurements with actigraphy or golden standard polysomnography (PSG) increases the cost and difficulty of doing this research. Another approach to handling this data is to use unsupervised classification through clustering [50]. Studies such as this eliminate costly (in Lithuania starting at 200€ per night) and labor-intensive expert labels and provides an automated approach to detect wake and sleep cycles. Several studies have used unsupervised machine learning techniques on accelerometer data to identify clusters of physical activity [51]. These studies used algorithms such as K-means, DBSCAN, and Hidden Markov Models (HMM). It is noteworthy that dimensionality reduction studies lean towards the utilization of basic techniques like Principal Component Analysis (PCA) and correlation analysis.

Unfortunately, it is costly to have annotated data and create classifiers, so it is important to find a way to analyze data in an unsupervised way. One study that worked with unsupervised machine learning models tried to distinguish the activity patterns of people during weekdays and weekends. Specifically, looking at wearable device data, it is possible to distinguish activity patterns using the Timer-Series Anytime Density Peak (TADPole) clustering method [52]. In this study, people wore smart watches for some time, from the extracted variables, such as step count, distance, energy expenditure, and duration with starting and ending points, then performed a clustering analysis with the TADPole clustering method from *dtwclust* package in R. In the results, it is seen that the energy expenditure in this study was the main variable for the cluster analysis. The data was divided into weekdays and weekends to mimic better human behavior, and the authors found two groups of people with more stable activities throughout weekdays and weekends and more shifting activities where different patterns occur throughout the week. It was also seen that age was a major factor in patterns of physical activity.

Another study was to investigate complex chronic patients (people with functional decline that can be seen in changes in mobility patterns) [53]. Those patients were measured with the Barthel Index which measures the extent of a person's independent mobility. The authors of this study collected patient data such as step count and heart rate, where from these hour-by-hour measurements they created the Mean Activity Profiles of taking median measurements of each entry within every hour of the day. Those profiles, as separate time series entries, were analyzed using the K-means algorithm with the distance function Dynamic Time Warping (DTW) from *tslearn* library. Four distinct clusters of Mean Activity Profiles were discovered where one of the clusters had a high correlation with a spectrum of patients with the Barthel index. The discovery showed that by analyzing the person's activity and his or her behavior, it is possible to detect complex chronic patients of one of the spectrum.

With all the opportunities that a large amount of personal data offers, new artificial intelligence methods, and much higher computing power, now it is possible to find more personalized studies that can provide recommendations or detect abnormalities on the personal level. The new study showed how new technologies help to create better personalized prediction models to detect blood pressure spikes [54]. This incorporated data from Fitbit API such as heart rate, breathing rate, blood oxygen levels (SpO2), and skin temperature, also Omron's Healthcare API which provides time-stamped measurements of blood pressure. Humans are very different, and their biosignals differ from person to person, and to avoid generalized predictions, the model was trained with individual data separately. For building models, long short-term memory (LSTM) and Transformer Neural Networks were applied by using library *Tensorflow*. This study was just a pilot project to see if it is possible to predict blood pressure spikes by training models in not generalized, but personalized approach, however, studies like this bring good prospects for human health by applying machine learning techniques on biosignals for detecting illnesses or abnormalities on the personal level.

In another study [55] the data from Fitbit Web API was employed to detect sleep/wake cycles. The authors stated that they were one of the first unsupervised and personalized approaches in the literature. The Fitbit data used heart rate (minute-by-minute) and steps (total count in 15-minute periods) to create a model utilizing the hidden Markov model (HMM). As a result, this unsupervised model was able to detect sleep/wake cycles, and later the results were checked against Fitbit-generated cycles, and the model had 87. 31% agreement with the results. In addition, they experimented and created personalized algorithms for each person. This time they tried to do with the day-to-day approach, and they found a significant difference in sleep/wake cycles between weekdays and weekends.

Another interesting study developed an algorithm U-BEHAVED which can detect changes in activity and behavior over time and uses the moving mean step number and the difference in step number per hour, allowing the algorithm to detect even new habits [56].

An interesting application of data from wearable devices was mentioned in one study in which, based on some biometric data, it was possible to detect symptoms of COVID-19 [57]. The study took wearable device data from participants who have reported obtaining symptoms of the COVID-19 virus and then divided into groups by people who actually received a diagnosis of COVID-19 and those who did not receive a diagnosis. Retrospectively, data from several months were taken to analyze what changes or disturbances were detected before seeing visual symptoms of COVDI-19. Two methods were developed to detect differences: resting heart rate difference and heart rate in steps (HROS). Using Gaussian density estimation, anomalies were calculated. The sickness detection window was created from 14 days before to 7 days after the first symptoms. With this method 63% of cases would be detected in real-time data tracking before COVID-19 symptoms appeared.

Regarding the impact of physical activity on sleep quality, there are not many similar studies with applications of machine learning or mathematics. One of the studies tried to predict sleep quality that was defined as good or bad based on slept minutes (402 minutes and more was considered good sleep) [58]. For physical activity, the study did not approach time series data and took aggregated data of calories, steps, distance, and minutes of different types of activities that were generated by Fitbit. They used Long-short Term Memory (LSTM), Gated Recurrent Unit (GRU), and Convolution Neural Network (CNN) for model building and implemented them with Python and the most common libraries like Tensorflow, Anaconda, Keras, and Scikit-learn. Even though the GRU model performed the best, it still had low accuracy which was explained by the authors by not a sufficient amount of data. Also, they marked that the day window was used from 00:00 to 23:59, which does not consider people still being active after midnight.

Another study in similar context as this project, used 10-minute intervals of heart rate and step count to predict sleep quality [59]. This study proposed four different LSTM models: a multivariate stacked LSTM network, a dual-input stacked LSTM network, an LSTM-based autoencoder model, and a stacked LSTM with a one-dimensional convolution block. The first model was the best, while more complex models performed worse because of overfitting the data. The quality of sleep was defined by the number of deep sleeps because the evaluation of sleep quality based on deep sleep minutes has evidence of a correlation between deep sleep and overall sleep quality [60]. Another drawback that can be visible is that it was done only on one person's data, which as well might lead to overfitting.

To this day, there are still not many personalized approaches to similar topic studies. It may include weight, gender, and some other parameters, but it mainly takes a generic approach to measure the impact, and which subgroup of participants sees the bigger impact on good quality sleep. One of the studies tried to take a more personalized approach and collected more personal data to see the impact of physical activity on sleep [61]. This study collected the weight, height, BMI, fat mass, and lean mass of the participants. Furthermore, they monitored activity data using an accelerometer and with the R package GGIR determined periods of sedentary behavior and physical activity. For sleep quality, they had subjective (PSQI) and objective (Wake After Sleep Onset (WASO), Total Sleep Time (TST), and Sleep Efficiency (SE)) measurements of sleep quality. They also used individually measured data such as VO_2max and the muscular strength of the body. Using all this data, the authors were able to apply statistical analysis - descriptive analysis, simple linear regression, and multiple linear regression. Those models were adjusted by the person's measurements. For statistical analysis, they used SPSS, v.23.0. The results showed that overall physical activity, better muscular strength, and VO_2max have a positive impact on quality and quantity of sleep.

Another study that used steps and sleep quality from Fitbit devices conducted research on the day-to-day association between both measurements in overweight and obese participants [62]. This study decided that the evaluation of physical activity uses only heart rate, heart rate variability, and step data, and did not include physical intensity measurement, which could also be provided by Fitbit. The reason for this decision was the lack of evidence on the accuracy of other activity measurements, while the step count is calculated accurately based on the evidence. For sleep measurements, they used Total Sleep Time (TST) and Sleep Efficiency (SE), which was derived from the formula mentioned above. The other measurements such as weight and height were taken, and because of potentially different behaviors during different days, the days were marked as public holidays, weekdays versus weekends, and seasons.

The study created different linear regression models, however, the most important for this study were two: where the independent variable was selected as the total count of steps per previous day, and the dependent variables for two models were Sleep Efficiency and Total Sleep Time. It is worth mentioning that each model was trained separately for each individual, which created a more personalized approach. With the same variables, other models were created by using Generalized Additive Models (GAMs), which are recommended for nonlinear person-specific time series analysis. For the analysis, the authors used the R packages *mgcv* and *visreg*. The results that were received were very complex, and even though they found just some association of daily total steps and sleep time and sleep efficiency, it varied a lot from person to person, and as discussed, they recommended looking at each person on an individual level.

Human health and its improvement recommendations are not one-size-fits-all, but to understand the individual itself, it requires a lot of data to provide good medical conclusions. To solve this problem, wearable technology allowed scientists to collect large amounts of sequential life-logging data and provide healthcare analytics at the individual level [63]. This study focused on predicting physical activity from time series data from wearable devices. They took the activity data from mobile phones that consisted of activity type, distance traveled and duration during the activity, as well as each activity was labeled by intensity from 1 to 5. To create a model which could predict activity level, the authors of the study used a hybrid approach by combining two models into one MOGP-HMM: Multi-objective Genetic Programming (MOGP) helped to predict the class based on time series data, and the Hidden Markov Model (HMM) predicted the activity status over time-based observations. The study aimed to compare this new model's performance with already known SVM-HMMs, and the results showed that the new model was able to achieve comparable performance.

In general, a mixture of different methods was observed by analyzing how scientists in recent studies were able to approach this problem of detecting the relationship between human physical activity and sleep quality. It is seen the pattern than usually the studies are designed to review the relationship in larger groups of people, sometimes specifically to one or another group, like diabetic patients, students, specific occupations, such as policeman or nursers, etc. In addition to this, the same metrics were used most of the time, either a Pittsburgh Sleep Quality Index (PSQI) or the time spent sleeping to improve sleep quality, or steps and average heart rate to measure physical activity. Based on that, different methods were applied to analyze the relationship.

The majority of studies, especially from medical or social sciences, used simple statistical models due to ease of use and good interpretability, which is very important when the goal is to explain the relationship and provide the necessary recommendations. On the other hand, new technology with larger amounts of data gave an opportunity to mathematicians explore machine learning techniques to create different models for detection of abnormalities, also detection and classification of physical activity types, also very often trying to detect sleep times and sleep phases.

Although the accuracy of sleep time with detection of wake-up and sleep times is getting better, there is still a lot of room for improvement to work on sleep phases detections, which is expensive because of the necessity of polysomnography results to be compared with. However, emerging consumer technologies now allow for the collection of additional biosignals, such as brain waves through integrated electroencephalogram (EEG) trackers. These advances could potentially improve visibility of sleep patterns in the future, complementing heart rate and accelerometer-based movement sleep phases detections.

What is interesting to see how this topic became interdisciplinary combining knowledge from medicine, neuroscience about the body, then psychology, and economy to help understand the value of those studies with findings, and how the recommendations must be translated to people because it may impact their quality of sleep to bring a positive impact to the economy. Also, mathematical experts are trained in understanding data analysis with statistics, to show actual relationships, help to easily discover patterns, and apply complex calculations that might unlock interesting insights. And on top of that, informatics that help to create a technical solution which would not only allow processing and analyzing large data amounts, but also create data integrations with medical institutions, provide user-friendly feedback to the individual, and even bring prompt real-time notifications about reminders of working out or go to bed on time.

1.5. Foundation for project's objectives and relevance

The relationship between human physical activity and sleep quality is a popular topic due to its importance for both human health and economic impact. Poor sleep has been associated with substantial losses to healthcare systems at the governmental level, as well as increased expenses for employers. Workers who do not get adequate sleep may be less productive and creative, participate less in work activities, and have a higher risk of accidents at work. Educating people about sleep hygiene and providing personalized recommendations was also shown to be important and demonstrated the relevance of the focus of this project.

Multiple studies have shown that physical activity affects sleep quality. However, most of these studies are based on subjective sleep and activity questionnaires, or used only few metrics. The reason could be that the relationship of some metrics was not strong between multiple people and they were not widely used, but with this project it could be possible to validate metrics on individual level that are less often or never used before. It shows the importance of having a semi-automated framework that would enable quick detection of personal relationships between physical activity and sleep quality metrics.

Recent studies also attempted to obtain a more complete view of daily activities by applying various time series analysis and machine learning methods. This was made possible by improvements in technology that allows to gather extensive data sets, not only on a larger population but also at an individual level. These methods have shown the ability to detect daily activities or identify deeper patterns, for example, applying time series clustering analysis.

2. Research Methods

After completing the literature review, it was found that the topic of the impact of physical activity on sleep quality and overall recommendations to improve sleep quality are very relevant topics. Unfortunately, most studies only investigated the relationship in large groups of people, and they often overlooked individual differences. This project aims to address this gap and open the possibility of finding a more meaningful relationship between physical activity and sleep quality based on an individualized level. In addition, in recent years it became possible due to the new wave of wearable technology which allows collecting a lot of biometrics that can unlock the patterns of each person separately. The key feature of this research is its focus on individual needs by trying multiple different biomarkers that have not been explored together in other studies. This section of the project will propose a methodology that uses various metrics of physical activity and sleep quality to find more individualized relationships. This section includes data pre-processing steps, creation of the Activity-Sleep Metric table, the analysis methods to use, and its technical implementation.

2.1. Human physical activity and sleep quality metrics

This methodology will focus on the available Fitbit data that was provided by the company. Many studies specifically reviewed the accuracy of the wearable devices, and in some of the metrics provided, there might still be lack of accurate records [39]. However, it is important to note that this project focus is not on accuracy or phase detection, hence the idea is to use the already aggregated and processed Fitbit data. Fitbit provides multiple data sets on daily activity and sleep, and for this project, specific data sets will be used.

The werable device data provided by the company contains multiple files with physical activity and sleep data. Each sleep and activity data source is in different formats, hence, there is a need for a better structure that would help to automate the analysis. First, for physical activity analysis, the time series data file is provided which contains two minute-by-minute time series, one for steps count and another average heart rate. For sleep data, one file provides aggregated data, such as total sleep minutes, total deep sleep minutes, sleep start and end times, etc. Also, in this project, another sleep file will be used which provides exact timings when and what sleep phase occurred during the night. For this matter, the goal is to create an Activity-Sleep metrics table which would include all the daily physical activity and sleep quality metrics. Before going into the details of the data preparation, it is important to define the metrics of physical activity and sleep quality that will be used in this project based on literature review and can be derived from Fitbit data.

2.1.1. Sleep quality metrics

- **Sleep Minutes or Total Sleep Time (TST):** Widely used in medical studies, TST serves as an indicator of sleep duration. Longer sleep duration is often associated with better sleep quality. Fitbit calculates sleep minutes by excluding awake minutes from the total recorded time in bed.

For instance, if a person slept for 8 hours but 1 hour was recorded as awake, the calculated sleep hours would be 7 hours, identifying only the restful sleep period. For classification exercises, it would be possible to classify sleep based on the threshold, where the most popular threshold is 420 minutes[58].

- **Awake minutes or Wake After Sleep Onset (WASO):** WASO measures the duration of wakefulness after the initial onset of sleep. In the Fitbit data, this measurement is referred to as "Awake Minutes." Higher amounts of awake minutes recorded during the sleep period may suggest poorer sleep quality.

- **Awakening Frequency:** Similar to awake minutes, this measure indicates the frequency of awakenings during sleep. A higher number of awakening episodes suggests less restful sleep and may indicate poorer sleep quality.

- **Sleep Efficiency:** It is a ratio of Total Sleep Time (TST) and Total Time in Bed (TTB). This measurement helps to identify if a person slept through all the time in bed or if he or she had trouble falling asleep. Fitbit API provides its Sleep Efficiency measurement, however, it does not follow the same logic, and it could be their internal calculation, which is not clearly explained in their documentation. However, it would be good to analyze both measurements (Fitbit calculated and literature-based) as well as a part of this project.

- **Light, Deep, REM Sleep:** As it was already mentioned in some studies [59] [60], deep sleep minutes had promising results to be used as the measure of quality. Evaluating sleep not only by duration but also based on sleep structure gives additional insights into the sleep quality differences

- **Deep, Light, and REM Sleep ratios:** Ideally, sleep stages should be balanced, and it must have around 25% of sleeping time for REM sleep, and the same part of the sleep for deep sleep [3]. This means that if a person has a 7-hour sleep, ideally, there should be around 100 minutes of each sleep stage. The rest of the sleep should be around 50%. To my knowledge, there are no such examples in current studies, but knowing the importance of sleep phase composition, it could bring additional value in classification models where more than 420 minutes of sleep with around 25% of deep and 25% of REM sleep could be classified as good, and if the ratio is lower, then just moderate quality. Based on Garmin's documentation, they consider a good balance for deep sleep around 16%-33% of their sleeping time, and for REM around 21% to 31% of total sleep. Taking into consideration the company's recommendation, it would be possible based on the structure of sleep classify sleep quality as good or poor quality sleep if it meets the required ratios.

2.1.2. Physical activity metrics

- **Total Steps Count:** Steps count is a common metric for measuring personal physical activity. Typically in studies, it was aggregated as a daily step count. Addressing the gap in previous studies where aggregation was done based on calendar date, in this project the Awake Period is introduced to cut the day into the windows from waking up to sleep times.

Table 1. Sleep quality metrics

Feature	Unit of Metric	Name in Models	Metric Scale
Sleep minutes	min	y1	Quantitative
Deep, Light, REM minutes	min	y2, y3, y4	Quantitative
Awake minutes	min	y5	Quantitative
Efficiency (provided by device)	%	y6	Quantitative
Efficiency (calculated TST/TTB)	%	y7	Quantitative
Deep, REM, Light ratio	%	y8, y9, y10	Quantitative
Frequency of awakenings	count	y11	Quantitative

- **Heart Rate Measurement:** Heart rate can also be aggregated as an average of the Awake Period. Also, the maximum heart rate during the Awake Period can identify if there was any more intense activity during the day.

- **Intensity Score Calculation (moderate-to-vigorous):** To get a summarized view of the intensity of the day, it is possible to calculate the Active Minutes based on heart rate during each minute. There is no single rule, and it is based on each individual, but as a general rule of thumb, it can be classified into Moderate activity (50-70% of maximum heart rate) and Vigorous activity (70-90% of maximum heart rate). Then it is possible to score each minute: 1 point for moderate activity and 2 points for vigorous activity. To calculate maximum heart rate, an equation can be used 208 - 0.7age [64].

- **Activity Score Calculation:** Each class of minutes (vigorous, moderate, light) can be treated quantitatively, in this case not losing the importance of light activity. The activity score combines minutes into points: Vigorous activity (+3 points per minute), Moderate activity (+2 points per minute), and Light activity (+1 points per minute). Sedentary minutes are excluded. Offers a holistic view of daily physical activity.

- **Detailed Daily Time Series Data:** Time series data can be used effectively for cluster analysis to identify distinct groups of days based on their patterns. Methods such as k-shape clustering and k-means with Dynamic Time Warping (DTW) distance metrics are commonly employed to measure similarities within the data. These techniques help to discover clusters that exhibit similar dynamics over time.

- **Intensive Minutes Before Bed:** Based on the medical literature, it is not recommended to exercise intensively 2-3 hours before bedtime [23], hence, this metric was added to this methodology to test the real data. This metric can be calculated by taking 2-3 hours before recorded sleep and applying the Intensity Score Calculation.

It is also important to mention that humans behave differently during the weekends [52], so it is important to include categorical variables for the weekends.

Table 2. Physical activity metrics

Feature	Unit of Metric	Name in Models	Metric Scale
Period Steps Count	count/period	x1	Quantitative
Average Awake Period Heart Rate	beats/min	x2	Quantitative
Maximum Awake Period Heart Rate	beats/min	x3	Quantitative
Intensity Score	points	x4	Quantitative
Activity Score	points	x5	Quantitative
Intensity Score Before Bed	points	x6	Quantitative
Intensive Activity Before Bed	boolean	c1	Categorical
Activity Cluster Number (DTW, k-shape)	number	c2, c3	Categorical
Is Day off	boolean	c4	Categorical

2.2. Data pre-processing

The first step before combining the mentioned fields is to remove invalid, inconsistent, or insufficient data. Since the focus is to understand how physical activity might affect sleep quality, both active and sleep period data must be valid and present together to be included in the analysis.

Fitbit provides some identifications for sleep data where the data might not be sufficient. For example, the field *isMainSleep* identifies if the record is the main sleep or not, which is helpful for detecting naps or other shorter sleep episodes. In addition, the Fitbit API provides another value *infoCode* which indicates the records that were calculated based on insufficient data, or the sleep record was less than 3 hours of sleep. Those records (nights) should be excluded for further analysis.

Regarding physical activity data, Fitbit does not provide aggregated or similar information that would help clean the data. For this reason, planned physical activity metrics must be calculated. But before that, it is important to define what is Awake Period. In previous studies on similar subjects, it was seen that there was a gap in how step count or any activity was calculated about the day. Usually, the day was defined as the period from midnight to midnight, without taking into account that a person might go to sleep after midnight, and steps will be counted for the following day. For this reason, in this methodology, it is suggested to introduce the Awake Period definition, which would explicitly refer to the time from waking up to the time a person falls asleep. This Awake Period can be easily calculated by approaching sleep's data of Sleep Start and Sleep End timings.

For Awake Period data there are no specific identifications provided by Fitbit, and similar studies did not come up with one rule on how much of non-wear time could be a threshold to be removed from the analysis. There are multiple examples in wearable device data where different studies set different exclusion criteria for non-wear periods: less than 1000 minutes (around 16 hours) of the recorded full day, [65], less than 1000 steps for the day [66], less than 8 hours monitored of the awake period [67], less than 10 hours monitored of the awake period [68], or more strict less than 1400 minutes (23 hours 20minutes) [62]. These variations suggest that the choice of threshold largely depends on the specific objectives and context of the project, but typically requires at least 60% of the data to be available throughout the day. To ensure robust data, it is feasible to increase this threshold up to a point where the proportion of removed records does not exceed 5-10% of the total data.

Outliers should also be removed in this data cleaning section to improve the robustness of the models. There are two options to find outliers: Z-score or Interquartile Range (IQR) methods. The Z-score represents the number of standard deviations that is different from the mean. As a common practice, the outliers are beyond the thresholds of -3 and 3. This method can be used if the data is normally distributed. However, the IQR method can be used on skewed data. The IQR is the difference between the 75th percentile (Q3) and the 25th percentile (Q1). Observations that fall below Q1 - 1.5*IQR or above Q3 + 1.5*IQR can be defined as outliers. In this methodology, it is important to check the data before applying any of the methods and then choose accordingly.

2.3. Awake Periods clustering based on physical activity data time series

The received minute-by-minute activity data received for steps and heart rate can present challenges in trying to clean and aggregate them for analysis. However, it can also offer some opportunities by deepening the understanding of the daily patterns of human activity. Although aggregated data such as total steps and average with maximum heart rates per day can provide valuable information and overall health trends, it might overlook some patterns that could happen within the day. To unlock these hidden patterns of the data, clustering analysis can add value to this methodology.

In clustering analysis, objects are grouped into clusters that are more similar to each other than to those in other clusters. This method is particularly useful with data that was not labeled before, and might hold interesting patterns within different data points. Calculating the distance in time series clustering analysis might be a challenge, since similar days and similar activities will not necessarily be the same length or at the same time. For example, there could be two very similar days, but a person starts exercising at slightly different times, the exercise might be of different lengths, and the calculated Euclidean distance might be very high 3. To solve this problem, it is possible to use k-

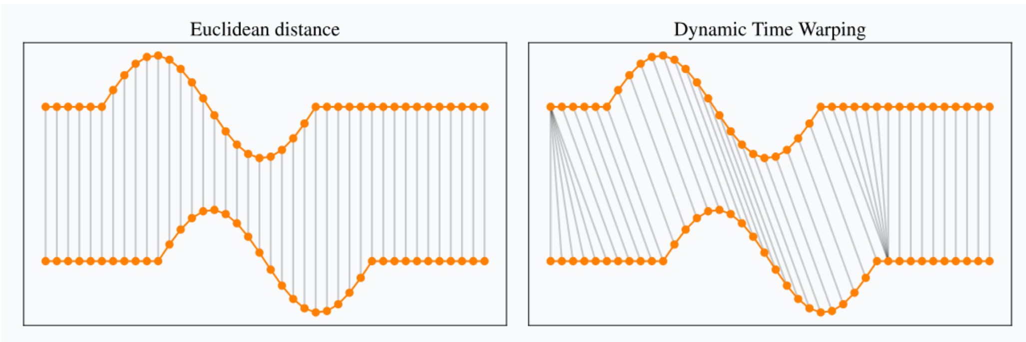


Fig. 3. Comparison of distance metrics: Euclidean distance and Dynamic Time Warping. Source: `https://geoenergymath.com/wp-content/uploads/2024/03/image.png`

means clustering method, but for distance metrics it is necessary to use the one that is dedicated to time series. Dynamic Time Warping (DTW) algorithm can take into account sequences that may vary in time or speed, and algorithm allows "elastic" transformations in order to minimize the distance of similar time series [69]. Historically, DTW was used for speech recognition to detect similar words or sentences to manage different speech speeds, but it found a very good application in the healthcare field on biometrics time series data. To calculate the distance of two time series, the goal is to align them by stretching and compressing until the minimum distance is found.

Given two sequences $X = [x_1, x_2, \ldots, x_n]$ and $Y = [y_1, y_2, \ldots, y_m]$, the cost matrix C is defined as the pairwise distance of all time series points:

$$C \in \mathbb{R}^{N \times M} : c_{i,j} = \|x_i - y_j\|, \quad i \in [1:N],\ j \in [1:M]$$

Time series points need to create an alignment path where x_i will have its corresponding y_j. Basically, the alignment of two time series is a sequence of points $p = (p_1, p_2, \ldots, p_K)$ with $p_l = (i_l, j_l) \in [1 : N] \times [1 : M]$ for $l \in [1 : K]$. The boundary condition must be met where the first and last points are linked to each other $p_1 = (1,1)$ and $p_K = (N,M)$. Other conditions, such as monotonicity and step-size conditions, must be met. For distance calculation, it is necessary to have the cost function calculated:

$$c_p(X,Y) = \sum_{l=1}^{L} c(x_{n_l}, y_{m_l})$$

where the final DTW distance calculation is done using this function:

$$\mathrm{DTW}(X,Y) = c_{p^*}(X,Y) = \min\{c_p(X,Y) \mid p \in P^{N \times M}\}$$

K-means method uses this distance metric for its calculations and is an effective way to find similar time series.

Another clustering method that is used mainly on time series data is the k-shape method. This algorithm focuses more on the shape of the data, and by calculating the distance between two time series it continues considering the shape. It is less computationally demanding, but still requires recalculation of the distances, hence, it can be heavy on large datasets. It uses shape-based distance (SBD). SBD between two z-normalized time series X and Y is calculated by:

$$\mathrm{SBD}(X,Y) = 1 - \max_{\tau} \left(\frac{1}{n} \sum_{i=1}^{n} (Z(X)_i \cdot Z(Y)_{i+\tau}) \right)$$

where $Z(X)$ and $Z(Y)$ are the z-normalized time series, and τ represents the lag. This technique has been shown to exceed other clustering methods, is domain independent, and is a scalable solution for time series clustering tasks [70].

In this methodology, it is important to unlock deeper patterns within days, and both distance metrics might help to find different clusters that were grouped by shape or differ in time or speed. DTW algorithm requires the same size time series data while the k-shape method can deal with different lengths. But for the current simplicity of the software implementation creation, for both metrics time series of the Awake Period will be cut into the same size windows. Since individuals are not regular and do not necessarily go to bed and wake up at the same time, it is better to calculate the usual wake-up and sleep times, as a usual start and end of the Awake Period.

Instead of taking an average or mean time not to lose some crucial information, the decision is for the usual wake-up time to take the 25th quantile by including a little bit more than the average of the morning, and for the usual sleep time, the 75th quantile will be assigned. Calculating quantiles in times is not straightforward. Knowing that a person can go to sleep before or after midnight, where linearity can be lost. For this reason, to calculate the usual sleep time, the actual time should be converted to *minutes to midnight*. In other words, if the sleep time was 23:45, then minutes to midnight would be 15 minutes, and if midnight is passed, it would get a value below zero. Then a calculation of the 75th quantile is performed, and the value of the usual sleep time can be assigned.

Having usual wake-up and sleep times, it is possible to create a matrix where rows will identify different Awake periods (different days), and each column will represent the value of the same time during that day. Each day's time series should be aggregated into 10-minute, 15-minute, or 1-hour intervals to reduce computational power. There is no single approach to this topic, and different studies applied different aggregations.

2.4. Physical activity's and sleep quality relationship model

The goal of this project is to find a relationship between physical activity and sleep quality. Although many studies approached similar topics with multiple linear regression, quantile regression, and logistic regression models, they used models on multiple people and derived results that fit a more generic group of people. In this project, an individualized approach will be taken in which each person might have different models that can identify the areas of physical activity that can be worked on to improve sleep quality.

2.4.1. Multiple linear regression analysis

Multiple linear regression is a statistical technique that helps to understand the relationship between one dependent variable (Y) and two or more independent variables ($X_i,\ i = 1, 2, \ldots, p$). The goal of this method is to find a linear relationship between variables by applying ordinary least-squares (OLS) regression. The model assumes that by modifying independent variables (also called regressors), it should linearly impact the result (Y). The model can be written as the formula:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p,$$

where:

- $E(Y)$ is the expected value (mean) of the dependent variable Y,
- $\beta_0, \beta_1, \ldots, \beta_p$ are the coefficients,
- $X_1, X_2, \ldots, X_p$ are the independent variables.

Since there are multiple metrics of physical activity, it is possible to generate a large number of different multiple regression models for each sleep metric. To avoid the time-consuming task of manually fitting each model, the All Subsets Regression method is used. This method creates all possible combinations of predictor variables (X) and automatically fits regression models for each outcome variable (Y) separately. An additional condition is added to require a minimum of three predictors to accelerate the process. The final results table for multiple linear regression filters out models where the adjusted R-squared is less than 0.1 and where multicollinearity is present (i.e., at least one variable has a Variance Inflation Factor (VIF) greater than 10). The same process is applied to multiple quantile regression, but filtering is based on the Pseudo R-squared value. The threshold for Pseudo R-squared is selected manually, as it does not have the same interpretation as adjusted R-squared.

Once All Subset Regression is completed, only the best-fitting models will be taken for further review. For creating the model, a standard procedure is going to be taken:

1. Model fiting: Choose `Y` and `X` which had the highest adjusted R-squared and has potential for stronger relationship models. Fit the variables into the model function (categorical variables must be converted to dummy variables).
2. Iterative Refinement:
   - Outlier and Influence Points Detection: Identify and remove outliers; during the first iteration, the standardized residual values can be chosen 3, and adjusted if necessary with other iterations. Use leverage and Cook's distance to identify Influence points.
   - Re-fit the Model: After cleaning the data, refit the model with the revised data set and compare the adjusted R-squared, AIC, and BIC values to assess improvement.
3. Validate Model and Performance: If the model reached the highest adjusted R-squared and lowest AIC and BIC values, the model should be validated for the linear regression assumptions, such as no multicollinearity (variance inflation factor (VIF) < 10), homoscedasticity (White's test), no auto-correlations (the Durbin-Watson statistics) and no strong outliers (standartized residual values within the range of -3 and 3)

2.4.2. Multiple quantile regression analysis

If the created linear regression model does not satisfy all necessary assumptions, it is possible to apply multiple quantile regression, which is recommended when data have medium outliers or do not satisfy the normality of residuals assumption. This model estimates the conditional median or other τ quantiles of the response variable

$$Q_\tau(Y) = \beta_0(\tau) + \beta_1(\tau)X_1 + \cdots + \beta_p(\tau)X_p,$$

where:

- $Q_\tau(Y)$ is the τ-th quantile of the dependent variable Y,
- $\beta_0(\tau), \beta_1(\tau), \ldots, \beta_p(\tau)$ are the estimated coefficients for the τ-th quantile,

- $X_1, X_2, \ldots, X_p$ are the independent variables,
- $0 < \tau < 1$.

The application process is very similar.

1. Model fitting: Choose Y and the independent variables X which had the highest Pseudo R-squared and has potential to create stronger relationship models. Fit the variables into the model function (categorical variables must be converted to dummy variables). While fitting the model, specify the quantile(s) at which to estimate the model, starting with 50th.

2. Iterative Refinement:

   - Outlier and Influence Points Detection: Identify and remove outliers; during the first iteration the standartized residual values can be chosen 3, and adjusted if necessary with other iterations. Use leverage and Cook's distance to identify Influence points.
   - Re-fit the Model: After cleaning the data, refit the quantile model with the revised data set and compare the adjusted R-squared, AIC, BIC values to assess the improvement.

3. Validate Model and Performance: If the model reached the highest pseudo R-squared and lowest AIC and BIC values, the model should be validated for for the quantile regression assumptions, such as no multicollinearity (variance inflation factor (VIF) < 10), no auto-correlations (the Durbin-Watson statistics) and no strong outliers (standartized residual values within the range of -3 and 3).

By applying this model, it is possible to obtain a more detailed understanding of how predictors affect different segments of the outcome distribution, providing a comprehensive view that may not be possible with linear regression.

2.5. Solution implementation with Python

Python was selected to implement the methodology of this project due to its comprehensive libraries, ease of use, and extensive support and tutorials. This implementation aims to create a proof of concept (POC) of the methodology as a semi-automated process to load data from any individual, prepare the data for analysis, apply the data to multiple models, and provide the results. If this concept, created as a base using Jupyter Notebook or Google Colab Notebook, proves successful, the code can be extended in the future to develop an application with a user interface, enhanced flexibility, and improved visualization of key findings.

Data pre-processing. One of the primary challenges of this project is the nature of the data itself. Wearable device companies, such as Fitbit, provide data in their structures and formats. Although the data is somewhat aggregated and prepared for initial use, it still requires significant cleaning and preprocessing to fit the specific needs of this methodology. This project will utilize four specific datasets supplied by Fitbit:

- **User Info:** This dataset contains the user identification code, his or her birthday, and gender.

- **Minute Data:** This dataset contains full-day information on step count and heart rate per each recorded minute.

- **Sleep Data Entries:** This list contains multiple aggregated records of sleep quality for each night. It includes the total duration of sleep, the duration of the deep, REM and light phases of sleep, the calculated efficiency (using a proprietary algorithm not disclosed by the company), and the total time spent in bed.

- **Sleep Data Series:** This dataset contains a more detailed view of sleep over time of the night, providing the actual time and duration of each phase.

The goal of the pre-processing step is to gather all necessary metrics from different sources and save them to the Activity-Sleep metrics table. The process is shown in Figure 4. In the first step of data loading, all files are loaded into *pandas* dataframes for easier data manipulation and built-in functions, such as data cleaning, arithmetic calculations, and data merging. Sleep data entries contain indicators of insufficient data that can be easily filtered out. For outlier detection, additional functions must be added for the IQR and z-score (*SciPy* library) methods to detect outliers. Based on the methodology, the selection of the method should be based on domain knowledge; both results are printed and then the best option is chosen for the final removal.

Fitbit initially provides only a few metrics such as sleep minutes, deep/REM/light sleep minutes, efficiency, and awake minutes with start and end times of sleep. Additional features are generated and calculated within the feature engineering step. Awake times are counted and taken from another sleep data series file.

Fig. 4. Process flow of Activity-Sleep metrics table creation and data pre-processing

The second part of the data pre-processing involves cutting the time series data of steps and heart rate into Awake period windows, which are calculated based on sleep end and the next day's sleep start times. Days with less than 40% non-wear time (entries without heart rate recorded) are removed. For each day, the total steps, average, and maximum heart rate are calculated using the *pandas* functions. Functions are created for Activity and Intensity scores, involving maximum heart rate calculated from the date of birth found in the user information file. Outlier detection and removal follow the same process as in the sleep part.

Both tables are merged together, creating the Activity-Sleep metrics table based on the date, with the understanding that sleep metrics should be merged with the previous day's activity. An additional feature, Day Off, is calculated by combining manually inserted public holidays and weekends using the *datetime* library.

Visualizations for outlier detection and descriptive analysis are implemented with *seaborn* and *matplotlib.pyplot* libraries.

Time series clustering analysis. For clustering analysis, the *tslearn* library is used along with the *TimeSeriesKMeans* and *KShape* classes. Two functions are created for time series data matrix preparation: one for k-means, which requires the same-size time series (the window is cut based on the quantiles explained in the previous section), and one for k-shape, where it is possible to keep the actual wake-up and sleep times, requiring only padding to create a matrix with the *numpy* library. Once time series matrices are prepared, they are used for clustering analysis, where the results of each assigned cluster each day are included in the final Activity-Sleep metrics table (Fig. 5).

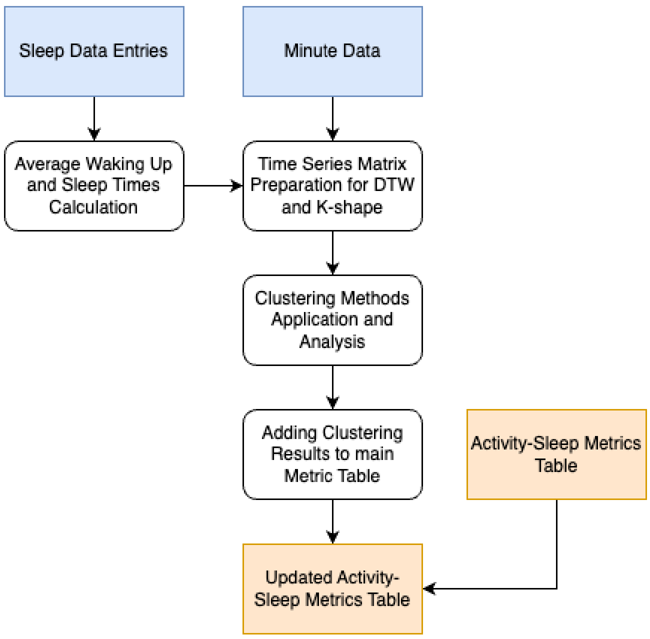


Fig. 5. Clustering analysis process flow and results merging to the final Activity-Sleep metrics table

Relationship analysis with regression models. In this step, the *statsmodels* library is used primarily for model fitting and the required assumption testing. To keep the code cleaner, additional functions are created for data preparation (including dummy variable creation), experiment execution, assumption testing, and outlier detection. All Subset Regression is performed using *itertools* to create all possible combinations.

This part still requires some manual input and cannot be fully automated. Based on the best models selected from All Subset Regression, each model is manually fit separately. The first model is initially fitted according to the automated solution. Outlier diagnostics are run, and if any outliers are detected, they are removed, and the model is re-fitted. Once outliers are no longer visible or there are no more strong leverage points, model diagnostics are run to test if the model fulfills all the assumptions. Since all the functions are already written, the model fitting process works more as a configuration of steps taken in the final model creation.

Final remarks. This implemented methodology solution allows an in-depth analysis of each person separately using Fitbit data. Currently, as a pilot version, it includes a semi-automated process to load data files, clean insufficient data, remove outliers, and use already aggregated Fitbit data to engineer additional features and create an Activity-Sleep metrics table. Then it performs a daily activity time series clustering analysis and performs a relationship analysis from all gathered data. Steps such as outlier removal and feature selection for modeling require domain knowledge and must be reviewed manually to ensure that they are performed correctly. In the next section, this methodology will be tested on real data.

3. Experiments

The new technology of wearable devices now enables scientists to look deeper into each individual and, based on that, provide more tailored recommendations on sleep quality. One of the objectives of this project is to apply methodology on real user data, which would be able to provide insights of the most influential individualized metrics based on different models that could explain the individual's relationship between physical activity and sleep quality. First, the data will be reviewed to better understand the domain, the quality of the data, and detect initial differences between each user. The data will then be cleaned and processed based on the recommendations of the methodology in the second section. Then for each person, an Activity-Sleep Metrics table will be created which will consist of cleaned and structured data that include processed and engineered features of physical activity and sleep quality; this can be seen as a table that consists of all possible Xs and Ys for modeling and relationship analysis. In addition, the daily time series of the Awake Period will go through time series clustering analysis, which would help to unlock new patterns. With all the prepared metrics, different models will be created to find which metrics of physical activity may impact or relate the most to sleep quality metrics. The goal is to find the strongest relationship for each individual separately, which might differ from person to person. The results and recommendations will be presented and limitations with future work recommendations will be discussed.

3.1. Descriptive data analysis

Person A data pre-processing. Starting with the person A data analysis, all CSV files were loaded into the Jupyter Notebook and filtered by UserId. Person A is a female born in the 1980s. A total of 331 lines were loaded from the sleep data file, and a total of 466560 lines were loaded from the activity file, which contained minute-by-minute information of the 324 unique days.

For sleep data, it was necessary to clean it based on Fitbit indicators of insufficiently measured sleep entries, also removing the entries that are not marked as main sleep. This step removed 11 entries from the data set. The key metrics gathered from the device will be checked for outliers. After checking the potential outliers with both methods, IQR picked more relevant outliers based on different metrics, which, in total, removed 10 rows (3. 13% of the filtered data set). In addition, in this step new metrics were engineered as calculated sleep efficiency, ratios of sleep phases, and frequency of awakening.

The next step was to clean up and aggregate the data set of physical activity metrics. To aggregate data of the Awake Period (from the time the person woke up to the time it fell asleep), the starting and ending times were taken from the sleep entries data set. If the previous day sleep entry with the waking time was not available, the median value of the waking time was used. Having time series windows of Awake Periods it was possible to check how much non-wear time was detected (non-wear time was indicated by records with missing heart rate). Since no direct recommendation was found on the acceptable amount of non-wear data for modeling, a quick analysis was conducted to determine how many records would be removed at different thresholds. Only four records were found to not meet the 60% threshold, indicating that increasing the threshold could provide more robust data. Therefore, a 75% threshold for available data was set, resulting in the removal of only 2.27% records from the dataset.

The total steps, the average, and the maximum heart rate were calculated. For intensity score, the person's age was used to calculate the moderate heart rate range, from 89 to 125 beats per minute. Using these parameters, intensity and activity minutes were calculated. The intensity score before bed was filtered based on two hours before the sleep time. This score was also used to create a categorical variable to indicate late night activity.

For outlier detection, the key focus was on aggregated data (such as steps and heart rate). After reviewing the z-score and IQR methods for outlier detection, both methods provided similar results. It detected 11 outliers, which were 3. 33% of the physical activity data set. After removal of outliers, the tables of physical activity and sleep quality metrics were merged on the date of entry, taking into account the difference of one day on the side of sleep quality metrics (activity happened on Monday and should be merged with sleep data that occurred on Tuesday). The final Activity-Sleep metrics table was created with a total of 290 entries.

Having first view of the relationship between physical activity and sleep quality metrics, the linear correlations are generally weak to moderate (Fig. 6). The average heart rate (x2) shows more noticeable correlations with certain sleep quality metrics. For example, it shows a moderate positive correlation of r = 0.19 with sleep minutes (y1) and r = 0.20 with awake minutes (y5), along with a moderate negative correlation of r = -0.19 with sleep efficiency (y6). The awake minutes (y5) also have a weak negative correlation with the maximum heart rate (x3) the day before and a weak positive correlation with the intensity score (x4). The correlation matrix revealed only weak to moderate linear correlations. Furthermore, examining the scatter plots for each pair of metric of physical activity and sleep quality (as shown in Figure 14 in Appendix A), no non-linear relationships were identified.

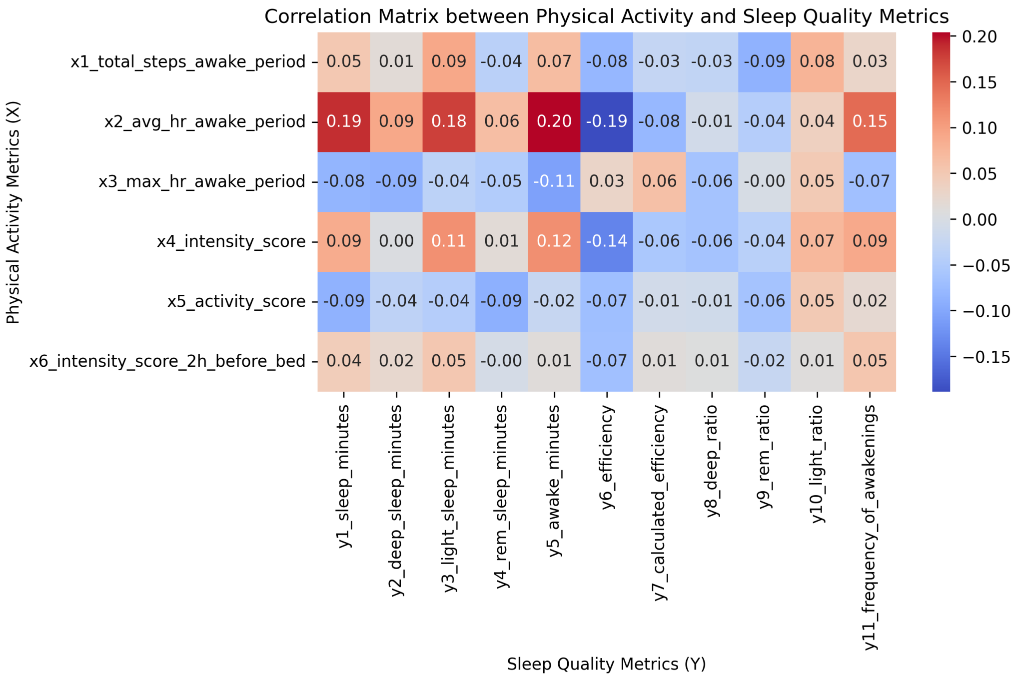


Fig. 6. Correlation matrix of physical activity and sleep quality metrics of person A

After analyzing the correlation among the physical activity metrics presented in Figure 7, it was found that the majority of these metrics have moderate correlations. Specifically, the intensity score (x4) and the activity score (x5) demonstrate a very strong linear correlation (r = 0.87), suggesting that these metrics are highly connected and likely reflect similar aspects of physical activity, therefore, one of them will be removed from further analysis.

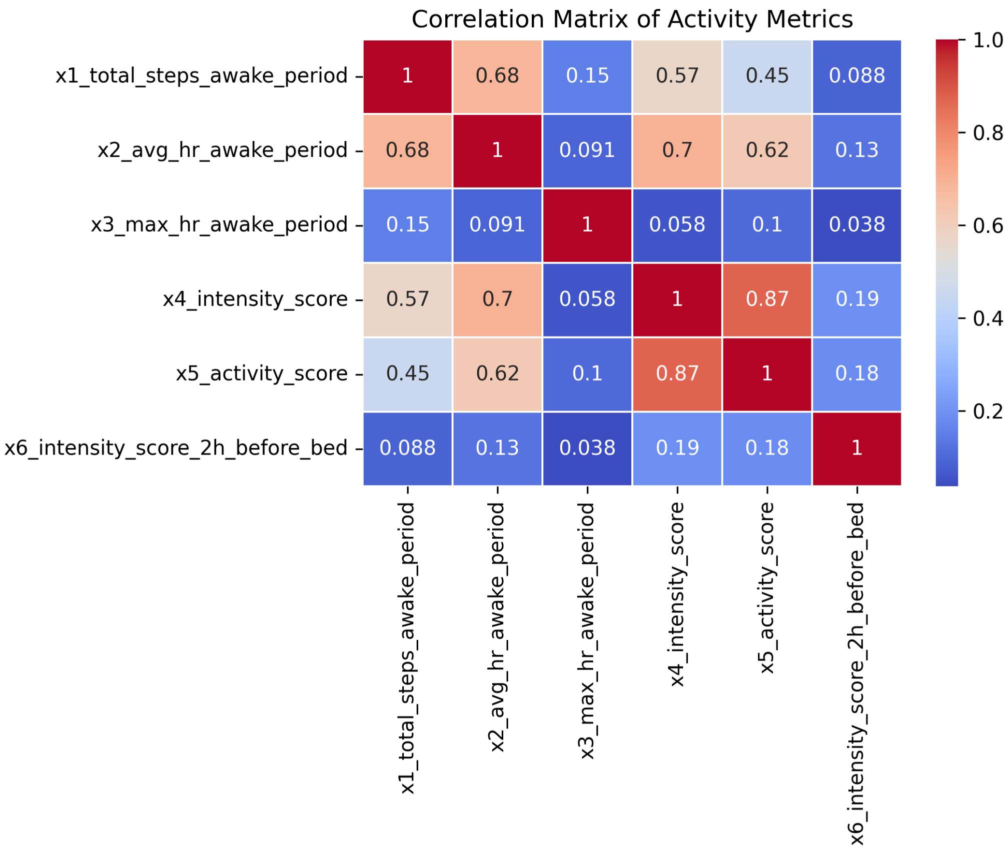


Fig. 7. Correlation matrix of physical activity metrics of person A

Given this high correlation, it will be essential to check for multicollinearity when including these variables in the regression analysis, as multicollinearity can inflate the variance of the coefficient estimates and make the model unstable. Although only one intensity or activity score will be used, the other metrics should still be carefully checked to ensure that they do not collectively introduce multicollinearity into the models.

Reviewing the categorical variables in Table 3, such as the day off variable (c4) (0 - working days, 1 - weekends and public holidays), it is evident that there is a difference between the averages. For example, during working days, the person averages 376 minutes of sleep (y1), which is 46.7 minutes less than during weekends. This pattern is consistent in other sleep phases, with more minutes on average spent in each phase on weekends. Interestingly, awake minutes (y5) also increase on weekends, indicating potential interruptions in sleep that should be investigated further. Regarding the standard deviation, it suggests that sleep patterns during weekends and public holidays are less consistent. This variability could be due to differences in daily routines, social activities, and other factors.

Table 3. Comparison of sleep metrics between days off and working days for person A

	Sleep Minutes (y1)		**Deep Sleep Minutes** (y2)		**Light Sleep Minutes** (y3)		**REM Sleep Minutes** (y4)		**Awake Minutes** (y5)	
day off (c4)	mean	std	mean	std	mean	std	mean	std	mean	std
0	376.0	39.0	69.9	19.5	215.7	30.6	90.4	19.9	52.0	10.4
1	422.7	54.9	85.3	19.9	235.2	35.0	102.2	24.6	62.4	13.2

Sleep phase ratios do not variate similarly between weekends and weekdays (Fig. 4), however, it is possible to see that for this person during the weekend on average the deep sleep ratio (y8) is higher, meaning more restful sleep. The average light ratio (y10) is lower during the weekend, and the REM sleep ratio (y9) remains the same.

Table 4. Comparison of sleep phase ratios between days off and working days for person A

	Deep Ratio (y8)		**REM Ratio** (y9)		**Light Ratio** (y10)	
day off (c4)	**mean**	**std**	**mean**	**std**	**mean**	**std**
0	18.6	4.7	24.0	4.2	57.5	6.6
1	20.2	4.1	24.0	4.2	55.8	5.9

Person B data pre-processing. Person B, a male born in the 1990s, had approximately 473 sleep data records loaded for analysis. The outlier detection methods identified similar outliers, but ultimately the z-score method was chosen for its ability to pinpoint specific outliers. This method removed 21 outliers, accounting for approximately 4% of the original data. Activity data was also prepared for aggregation, and the days with more than 40% of non-wear time were removed. Then data was aggregated, and activity and intensity scores were calculated based on a personal calculated moderate intensity heart rate (94.9 to 132.8 beats per minute). Then outliers were detected by using z-score and IQR methods. Both provided similar results, but IQR method was picked just because it identified more outliers that were seen from the boxplots. Since there were less valid data in activity data, after merging physical activity and sleep quality metrics in total there were 401 entries in Activity-Sleep metrics table.

Reviewing the correlation matrix of pairs of physical activity and sleep quality metrics (Fig. 8), only weak correlations were found, similarly, as observed for person A. There is a positive correlation (0.16) between average heart rate (x2) and deep sleep minutes (y2), indicating that higher average heart rates during awake periods might be associated with longer durations of deep sleep and more restful sleep. In contrast, light sleep minutes (y3) show a slight negative correlation (-0.22) with the average heart rate (x2), suggesting that higher average heart rates might reduce light sleep duration. This pattern is also reflected in the deep and light sleep ratios, indicating that the average heart rate could influence the composition of the sleep phases, potentially leading to more restful sleep. No other linear relationships were detected. Furthermore, the scatter plots in Figure 15 in the appendix did not reveal any linear or non-linear relationships.

After analyzing the correlation among the physical activity metrics presented in Figure 9, it was found

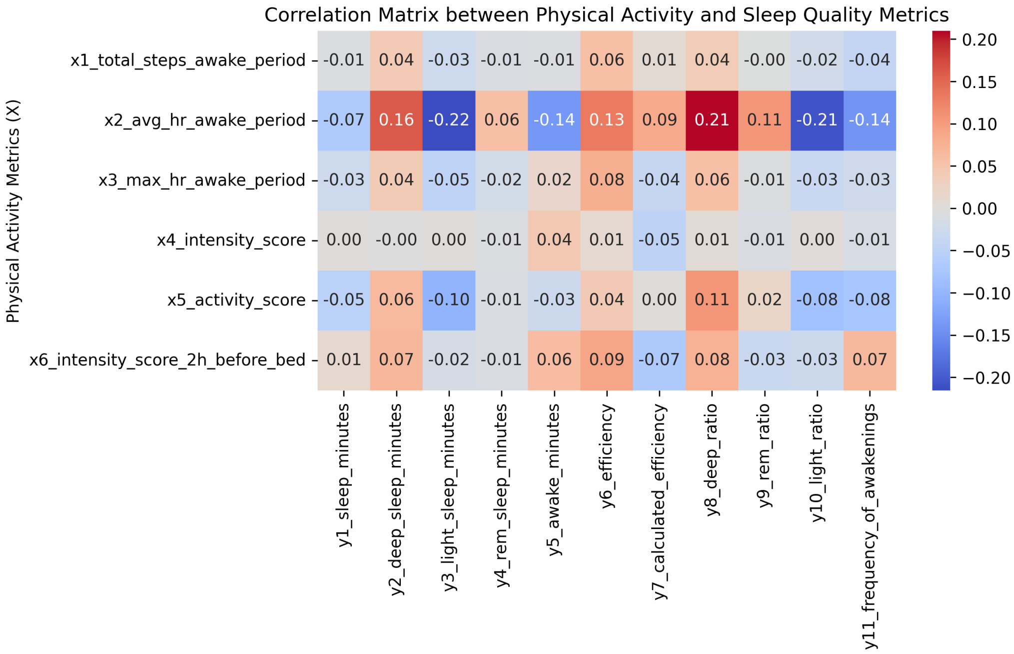


Fig. 8. Correlation matrix of physical activity and sleep quality metrics of person B

that the majority of these metrics have moderate correlations, similarly to the previous person. The same approach will be taken that activity score (x5) is not going to be included in analysis, and each model will need to be checked for potential multicollinearity.

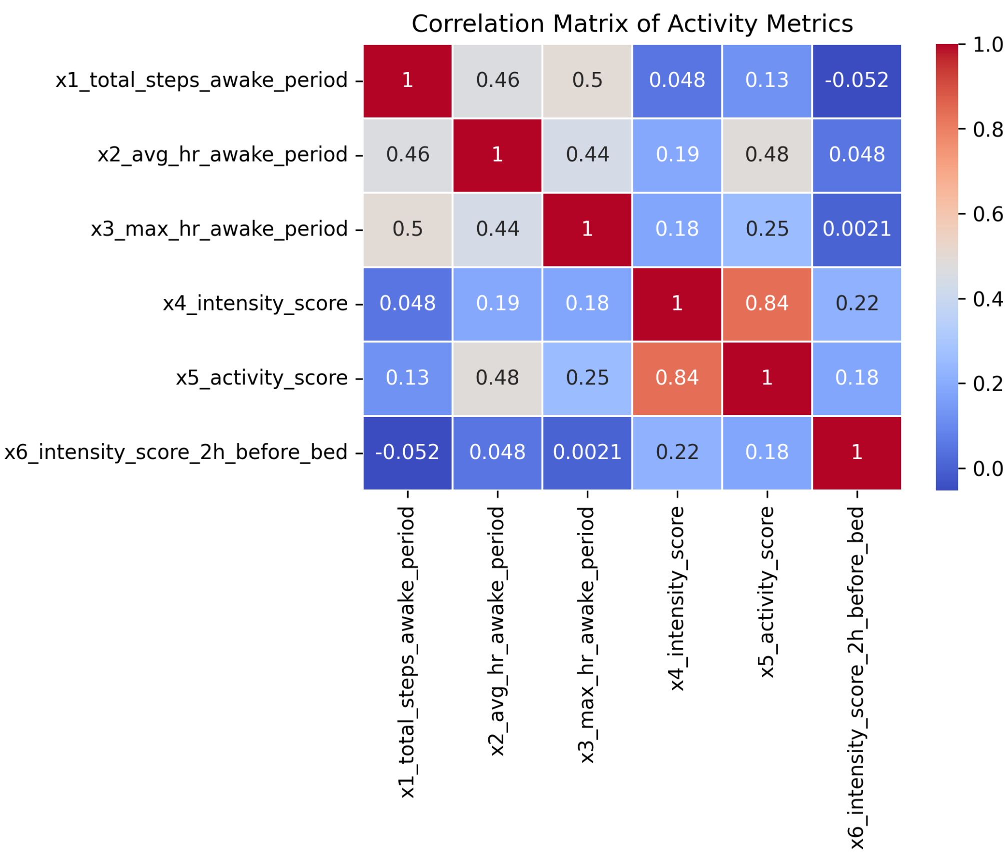


Fig. 9. Correlation matrix of physical activity metrics of person B

Tables 5 and 6 show a similar pattern in which during the weekend sleep in all minutes on average is longer than during the weekends. Although, the ratios of sleep phases did not change much between weekends and weekdays.

Table 5. Comparison of sleep metrics between weekends or public holidays and working days for person A

	Sleep Minutes (y1)		**Deep Sleep Minutes** (y2)		**Light Sleep Minutes** (y3)		**REM Sleep Minutes** (y4)		**Awaken Minutes** (y5)	
day off (c4)	mean	std	mean	std	mean	std	mean	std	mean	std
0	407.8	40.2	76.2	21.9	243.0	35.5	88.5	22.4	56.3	14.6
1	458.1	57.1	83.2	22.3	274.0	44.3	100.9	26.1	67.9	16.2

Table 6. Comparison of sleep phase ratios between weekends or public holidays and working days for person B

	Deep Ratio (y8)		**REM Ratio** (y9)		**Light Ratio** (y10)	
day off (c4)	**mean**	**std**	**mean**	**std**	**mean**	**std**
0	18.7	5.0	21.6	4.8	59.7	7.4
1	18.1	4.3	22.0	4.9	60.0	7.1

Comparison on data. Looking separately at both persons it was seen that there are similar behaviour patterns, and similar weak correlations between physical activity and sleep quality metircs. Checking both of them side by side (in Tables 7 and 8) main difference was seen that person A is overall more active, as on average daily steps is around twice more than person B. This can give stronger relationships in person A analysis.

Table 7. Descriptive statistics comparison of physical activity metrics for person A and person B data

	x1 Total Steps	**x2 Avg. HR**	**x3 Max HR**	**x4 Intensity Score**	**x5 Activity Score**	**x6 Int. Sc. Before Bed**
A Mean	16543.2	79.2	133.4	273.2	551.8	21.1
B Mean	7358.6	73.3	124.7	177.6	424.0	10.9
A Median	15358	78.8	129.2	229	496	10
B Median	6783	73.6	124.6	107	362	4
A Min	5057	64.5	99.8	61	121	0
B Min	185	59.4	95.6	13	59	0
A Max	37169	95.6	181.1	910	1453	242
B Max	23830	90.6	190.0	910	1638	242
A Std	5706.5	4.8	14.9	142.3	238.1	27.4
B Std	3287.6	4.4	15.0	175.0	268.3	29.2

Table 8. Descriptive statistics comparison of sleep quality metrics for person A and person B data

	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11
A Mean	390.6	74.0	222.7	94.0	55.8	90.9	87.4	18.9	57.2	23.9	5.0
B Mean	420.2	78.3	249.7	92.3	59.3	94.4	87.7	18.6	59.6	21.8	4.4
A Median	384	75	223	90	55	91	88	19	58	22	4
B Median	415	78	249	90	57	95	88	19	59	22	4
A Min	158	20	111	24	28	82	73	7	39	7	2
B Min	171	4	109	30	12	85	78	1	41	5	0
A Max	598	180	392	182	141	99	98	35	88	36	12
B Max	548	138	306	175	102	97	93	32	78	34	10
A Std	50.9	21.3	34.2	22.9	12.9	2.8	2.5	4.6	6.7	4.4	1.5
B Std	58.4	23.3	44.1	26.1	17.5	2.6	2.8	5.0	7.5	5.1	1.6

y1 - Sleep Minutes, y2 - Deep Sleep Minutes, y3 - Light Sleep Minutes, y4 - REM Sleep Minutes, y5 - Awake Minutes, y6 - Sleep Efficiency (provided by Fitbit), y7 - Sleep Efficiency (calculated based on literature), y8 - Deep Sleep Ration, y9 - REM Sleep Ration, y10 - Light Sleep Ratio, y11 - Frequency of Awakenings.

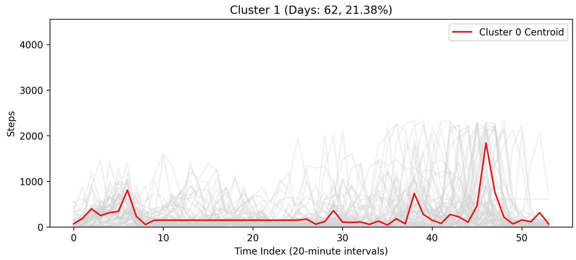
3.2. Person A data analysis

3.2.1. Awake Period clustering based on steps time series

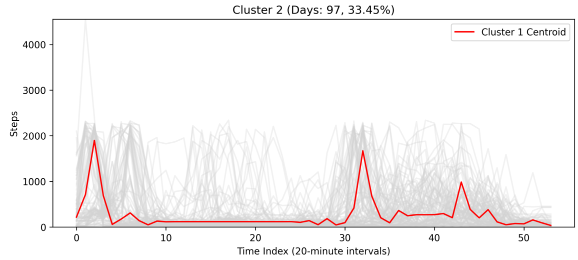
Next step is to get a better view of the person's daily activity by applying a time series clustering analysis. First, the k-means method with the Dynamic Time Warping (DTW) distance metric requires equal size time series to be analyzed and added to clusters; hence, it is not possible to use actual wake-up and sleep times. For this reason, the average timings of waking up and going to sleep were calculated. To avoid missing some of the crucial data, the 25th quantile was taken for wake-up time, and for bedtime the 75th quantile. This ensures to include a little extra time to the day not to seriously cut activity where a person might wake up earlier or go to bed later as an average day. For person A, the usual awake time was 08:49 and the sleep time was 02:05.

Then, daily minute-by-minute data was cut into a set of daily time series, forming a matrix. Each daily time series was aggregated into 20-minute intervals. With the data prepared, clustering analysis was performed using the k-means method with the DTW distance metric. The goal was to find clusters that could be easily explained or interpreted, especially if they would be significant in regression modeling. After different trials, 4 clusters were selected, and the results are shown in Figure 10, with calculated and plotted centroids for better visualization. Cluster 2 was the most distinct, showing activity spread throughout the day without any specific pattern. The other three clusters showed similarities, such as having more activity in the first part of the day, then lower activity, and slightly higher activity in the evening. Although they appeared similar, there were some distinctions. Cluster 0 had only around 1000 steps in the first part of the day, with more steps in the late evening. Cluster

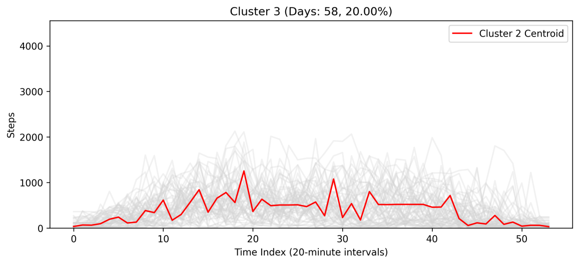
3 was similar, but activity in the evening tended to be a little earlier. Cluster 1 had twice the activity of the morning steps compared to the others.



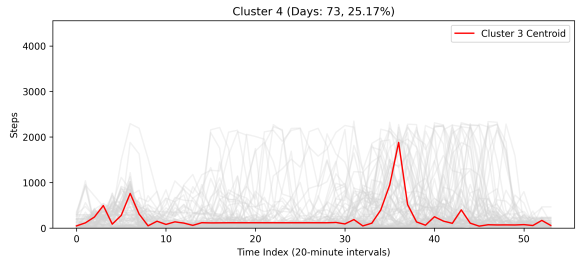
(a) Cluster 0



(b) Cluster 1



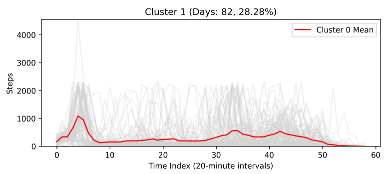
(c) Cluster 2



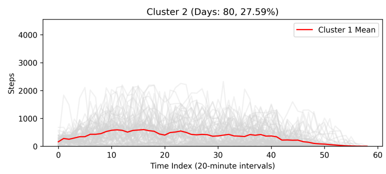
(d) Cluster 3

Fig. 10. Time series clustering results using k-means with DTW distance metric for person A

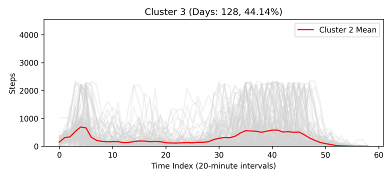
Applying the k-shape method, the identification of distinct clusters was more challenging. However, three interpretable clusters were identified (Figure 11). The second cluster exhibited a greater variety of days, while clusters 0 and 2 shared similarities to the previously discussed clusters identified using DTW. Despite multiple attempts, the creation of visually distinct clusters remained difficult.



(a) Cluster 0



(b) Cluster 1



(c) Cluster 2

Fig. 11. Time series clustering results using k-shape method for person A

More detailed comparisons of clusters and sleep metrics are provided in the Appendix A Tables 17 and 18, and only sleep minutes (y1) are shown to have a small difference in their averages. For example, DTW created cluster 2 with the highest average sleep minutes, while cluster 3 had the lowest average. The K-shaped clusters did not show very different averages between the clusters.

Following the cluster analysis, new columns representing the identified clusters were created and added to the comprehensive Activity-Sleep metrics table. These metrics will be treated as categorical variables and in the model itself converted to dummy variables according to the regression modeling requirements.

3.2.2. Investigation of the relationship between physical activity and sleep quality

After compiling all the metrics into a single table, the next step involved analyzing the relationships between them. Various metrics will be used to identify the strongest relationships. To expedite this process and pinpoint the most significant relationships, an automated script, as an all subsets regression with multicollinearity control, will be run to explore different combinations of physical activity metrics, including categorical variables, and determine which sleep quality metrics are most affected.

The performance of multiple linear regression models will be measured using adjusted R-squared, AIC, and BIC, while pseudo R-squared, AIC, and BIC will be used for multiple quantile regression models. This initial all subsets regression round aims to identify the strongest relationships from the first model fit, acknowledging that these models might still contain outliers or fail to meet all assumptions. Once potentially good models are identified, they will be examined more closely, and further tuning may be performed to enhance their accuracy and reliability.

For person A, the strongest relationships were identified and are presented in Table 9. The most robust models of initial fit relate to sleep minutes (y1), which included categorical variables such as the results of the clustering of the two clustering methods used (c2 and c3), weekends (c4) and whether there was intensive activity before bed (c1). Continuous metrics such as total steps (x1), maximum heart rate (x3), and intensity score before bed (x6) were also included. Similar metrics influenced the models for deep sleep minutes (y2). For awake minutes (y5), similar metrics were chosen, but with few differences. Only the DTW clustering results (c2) were included in the top models, with the average heart rate (x2) being used instead of the maximum heart rate.

Table 9. Top multiple linear regression models for different sleep quality metrics for person A

Features	Target	Adj. R-squared
[x6, c4, c1, c2, c3]	y1 sleep minutes	0.20
[x1, x6, c4, c1, c2]	y1 sleep minutes	0.19
[x3, x6, c4, c1, c3]	y1 sleep minutes	0.19
[x6, c4, c1, c3]	y2 deep sleep minutes	0.11
[x3, x4, c4, c3]	y2 deep sleep minutes	0.11
[x3, c4, c2, c3]	y2 deep sleep minutes	0.11
[x1, c4, c1, c2]	y5 awake minutes	0.17
[x1, x6, c4, c1, c2]	y5 awake minutes	0.17
[x2, c4, c1, c2]	y5 awake minutes	0.17

As the all subsets regression identified the highest adjusted R-Squared was detected by analyzing the strongest relationship with sleep minutes (y1). One of the models that was able to fulfill all the assumptions is presented in Table 10. The model heavily depended on the categorical variable indicating weekdays versus weekends (c4). The results indicated that sleep minutes (y1) increased by almost an hour on weekends compared to weekdays. Other physical activity metrics did not show a significant impact on sleep duration in this model.

One of the models that showed interpretable results is presented in Table 11. The chosen sleep quality metric was awake minutes (y5). Although total steps were not statistically significant due to their

Table 10. Multiple linear regression model of sleep minutes (y1) relationship with physical activity metrics for person A (adjusted R-squared = 0.325, AIC = 2548, BIC = 2566)

Variable	Coefficient	Std. Error	t-value	P-value
const	377.5	9.094	41.511	0.000
x1_total_steps_awake_period	-0.0004	0.001	-0.732	0.465
c4_day_off	54.9	5.033	10.914	0.000
c3_clusters_kshape_1	0.65	5.884	0.110	0.913
c3_clusters_kshape_2	8.12	5.051	1.609	0.109

p-value, removing them resulted in slightly worse model performance with an adjusted R-squared = 0.230, AIC = 1921, BIC = 1942.

The categorical variable indicating intense activity two hours before bed (c1) was statistically significant, suggesting that such activity increases the average number of awake minutes by 3.5 minutes. In addition, weekends (c4) added approximately 10 minutes to the awake time. This model is very important because the results of the cluster analysis also showed significance. Cluster 0 was removed from the model to avoid multicollinearity and acts as a base. Clusters 2 and 3 showed statistical significance and can therefore add 5.9 or 3.3 minutes compared to Cluster 0. Cluster 0 represented the days with a smaller step count in the morning but more active in the evening. Cluster 2 represented the day with full-day steps, which can tell that those days with more activity also increase the average awaken minutes. Cluster 3 is very similar to Cluster 0, but shows a little bit longer activity in the evenings.

Table 11. Multiple linear regression results for awake minutes (Y5) with physical activity metrics for person A (adjusted R-squared = 0.244, AIC = 1895, BIC = 1920)

Variable	Coefficient	Std. Error	t-Value	p-Value
const	48.6	2.086	23.291	0.000
x1_total_steps_awake_period	-0.0001	0.000	-0.843	0.400
c1_intensive_activity_before_bed	3.5	1.113	3.156	0.002
c4_day_off	10.2914	1.269	8.112	0.000
c2_clusters_dtw_1	1.6	1.538	1.025	0.306
c2_clusters_dtw_2	5.9	2.041	2.892	0.004
c2_clusters_dtw_3	3.3	1.573	2.116	0.035

Since both models met all assumptions, there will be no further steps with quantile regression. However, it is possible to review 12 and see that very similar models were recommended.

3.3. Person B data analysis

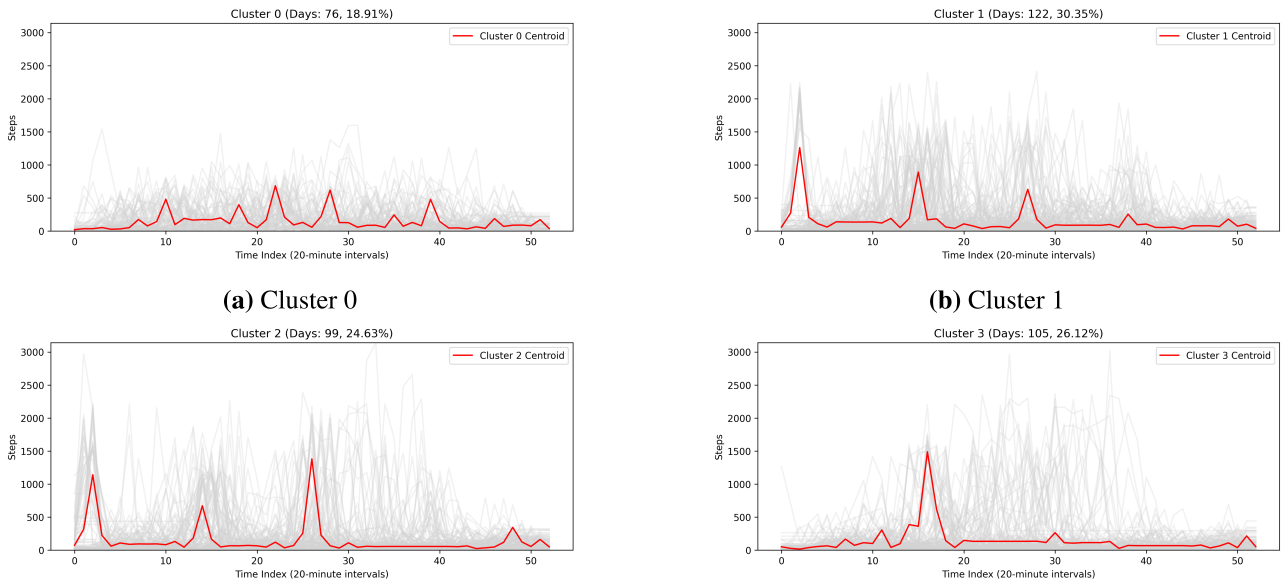
3.3.1. Awake Period clustering based on steps time series

To prepare data for time series clustering analysis, the same approach is used as for person A. First, for the k-means method, the Awake Period windows must be cut into equal parts. For the start Awake period, 0.25 quantiles of wake up time were used, and for the end Awake period 0.75 quantiles of sleep time. It cut the windows from 08:49 to 02:00 a.m.

Table 12. Top multiple quantile regression models for different sleep quality metrics for person A

Features	Pseudo R-squared	Target	AIC	BIC
[x1, x2, x4, x6, c4, c1]	0.19	y1 sleep minutes	2204	2229
[x1, x3, x4, x6, c4, c1]	0.19	y1 sleep minutes	2205	2231
[x3, x6, c4, c1, c2, c3]	0.21	y1 sleep minutes	2205	2242
[x2, x4, x6, c4, c1, c3]	0.12	y2 deep sleep minutes	1737	1766
[x1, x2, x4, c4, c1, c3]	0.12	y2 deep sleep minutes	1737	1766
[x1, x4, x6, c4, c2, c3]	0.13	y2 deep sleep minutes	1737	1773
[x1, x2, x6, c4, c1, c2]	0.18	y5 awake minutes	1407	1440
[x1, x2, x4, c4, c1, c2]	0.18	y5 awake minutes	1408	1441
[x1, x2, x3, c4, c1, c2]	0.18	y5 awake minutes	1408	1441

The results of the k-means method with DTW distance metrics are presented in Figure 12). Multiple trials were performed to find the clusters that could be interpreted, and these results showed the best distinguishing ones. Although clusters 1 and 2 looked very similar, with small differences in the evening. They showed some steps activity regularly in the morning, afternoon, and early evening. Cluster 1 showed some more activity later in the evening. Cluster 0 can indicate days with some random activities throughout the day, and cluster 3 showed days that mainly had only one spike throughout the day. K-shape method the best results presented with three clusters shown in figure 13.



(a) Cluster 0

(b) Cluster 1

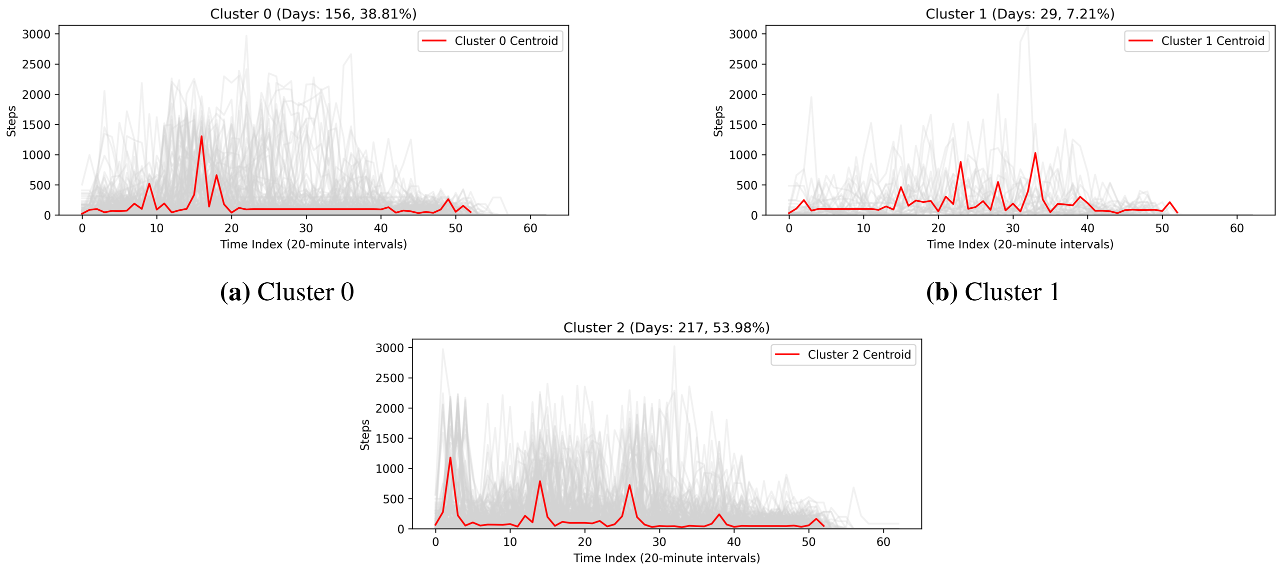
(c) Cluster 2

(d) Cluster 3

Fig. 12. Time series clustering results using k-means with DTW distance metric for person B

Cluster 2 showed a similar pattern that was seen in k-means as well, where key activities occurred three times per day. It also had a similar amount of days included in the cluster as with the k-mean method per two similar clusters. Cluster 0 combines similar days that had more intensive activity around the middle of the day. And Cluster 1 connected fewer days that did not look having an interesting pattern.

The tables of sleep metrics statistics based on clustering results are provided in Appendix Tables 19 and 20. Those clustering results were added to the final Activity-Sleep metrics table and will be used



(a) Cluster 0

(b) Cluster 1

(c) Cluster 2

Fig. 13. Time series clustering results using k-shape method for person B

for relationship analysis.

3.3.2. Investigation of the relationship between physical activity and sleep quality

The same methodology was applied to person B. The results for person B, as summarized in Table 13, indicate similar patterns. The top models for Person B also included both categorical variables and continuous metrics. These models were evaluated based on adjusted R-squared values to determine the strength of the relationships between physical activity and sleep quality metrics. Here it is possible to see some differences as well, for example, awake minutes (y5) initial fit models are not that strong as it was for person A, and instead of deep minutes (y2), here the light minutes (y3) models were selected instead.

Table 13. Top multiple linear regression models for different sleep quality metrics for person B

Features	**Target**	**Adj. R-squared**
Top 3 Models for y1_sleep_minutes		
[x4, c4, c2,]	y1_sleep_minutes	0.22
[x1, x4, x6, c4, c2]	y1_sleep_minutes	0.21
[x1, x6, c4, c1, c2]	y1_sleep_minutes	0.21
Top 3 Models for y3_light_sleep_minutes		
[x2, x4, x6, c4, c2,]	y3_light_sleep_minutes	0.17
[x2, x4, c4, c2,]	y3_light_sleep_minutes	0.17
[x2, x4, c4, c1, c2]	y3_light_sleep_minutes	0.17
Top 3 Models for y5_awake_minutes		
[x2, x4, x6, c4, c2]	y5_awake_minutes	0.13
[x2, x4, c4, c1, c2]	y5_awake_minutes	0.13
[x2, x4, c4, c2]	y5_awake_minutes	0.13

The first regression model was developed to identify the strongest physical activity metrics associated with sleep minutes (Table 14). Additionally, this model met all the necessary assumptions for multiple linear regression.

The model indicates a significant relationship between sleep minutes (y1) and weekends (c4), with individual spending, on average, 55 more minutes sleeping during weekends compared to weekdays. Cluster 3 showed a significant negative relationship with sleep minutes (y1), where days characterized by a single spike in step activity rather than multiple smaller spikes that were associated with 24.7 fewer sleep minutes on average during the awake period.

Although the exact nature of the difference between these days and others is not immediately clear from the clustering analysis, it suggests that days with a singular spike in activity could negatively impact sleep duration. More research with the individual may be necessary to understand specific activities or events on these days that could have negatively influenced sleep.

The intensity score (x4) was not statistically significant, but its inclusion helped to improve the overall fit of the model and met all assumptions.

Table 14. Multiple linear regression model of sleep minutes (y1) relationship with physical activity metrics for person B (R-squared = 0.297, AIC = 3933, and the BIC = 3957)

Variable	Coefficient	Std. Error	t-value	P-value
const	410.8	5.344	76.881	0.000
x4_intensity_score	0.02	0.015	1.608	0.109
c4_day_off	55.4414	4.58	12.094	0.000
c2_clusters_dtw_1	-5.53	5.754	-0.961	0.337
c2_clusters_dtw_2	2.25	6.010	0.375	0.708
c2_clusters_dtw_3	-24.66	5.928	-4.159	0.000

Another model identified from the all subsets regression was for light sleep minutes (y3). It was possible to create a model (Table 15) that meets all linear regression assumptions. This time, the clustering results were removed because their exclusion slightly improved the model. In this model, all physical activity metrics are significant.

Starting with the average heart rate (x2), each additional beat per minute added to the average heart rate can reduce the number of minutes of light sleep (y3) by 0.63. Each unit increase in the intensity score (x4), which measures each minute in which the person's heart rate is higher than 50% of their maximum heart rate, is associated with an increase of 0.025 minutes in light sleep. Furthermore, weekends (c4) add an average of 12.383 more minutes of light sleep.

Table 15. Multiple linear regression model of sleep minutes (y1) relationship with physical activity metric for person B (adjusted R-squared = 0.2, AIC = 2667, BIC = 2682)

Variable	Coefficient	Std. Error	t-value	P-value
const	98.1800	12.566	7.813	0.000
x2_avg_hr_awake_period	-0.6348	0.174	-3.652	0.000
x4_intensity_score	0.0254	0.007	3.617	0.000
c4_day_off	12.3830	1.467	8.440	0.000

Two models were created for the relationship between sleep minutes (y1) and light sleep minutes (y3) with physical activity. The awake minutes (y5) model did not perform well and did not meet the majority of the assumptions, and if additional review would be needed, multiple quantile regression could be performed. Top models of quantile regression all subest regressions results are in Table 16.

Table 16. Top multiple quantile regression models for different sleep quality metrics for person B

Features	Pseudo R-squared	Target
[x4, c4, c2]	0.13	y1_sleep_minutes
[x2, c4, c2]	0.13	y1_sleep_minutes
[x3, c4, c2]	0.13	y1_sleep_minutes
[x2, x4, c4, c2]	0.08	y3_light_sleep_minutes
[x1, x2, x4, c4, c2]	0.09	y3_light_sleep_minutes
[x2, c4, c2]	0.08	y3_light_sleep_minutes
[x1, x2, x3, x6, c4, c2, c3]	0.05	y4_rem_sleep_minutes
[x1, x2, x4, x6, c4, c2, c3]	0.05	y4_rem_sleep_minutes
[x1, x2, x3, x4, c4, c2, c3]	0.05	y4_rem_sleep_minutes
[x2, x3, x4, c4, c2]	0.10	y5_awake_minutes
[x1, x2, x4, c4]	0.09	y5_awake_minutes
[x1, x2, x4, c4, c3]	0.10	y5_awake_minutes
[x1, x2, x4, c1, c2]	0.06	y10_light_ratio
[x1, x2, c4, c1, c2]	0.05	y10_light_ratio
[x1, x2, x4, c2]	0.05	y10_light_ratio
[x2, x3, x6, c4, c2, c3]	0.05	y11_frequency_of_awakenings
[x2, x3, x6, c4, c1, c2, c3]	0.05	y11_frequency_of_awakenings
[x1, x2, x3, x6, c4, c2, c3]	0.05	y11_frequency_of_awakenings

3.4. Discussion

The methodology was successfully tested on real data from two different persons and this experiment gave individual results, even though both experiments had similarities. From the descriptive analysis, both persons shared similar amounts of sleep minutes in all phases, only the average minutes of sleep was 30 minutes lower for person A, which also was under the recommended 7 hours of sleep. The structure of the sleep phases also did not differ much. However, in the physical activity data it was seen that person A has about two times more steps, which could be the reason why the created models had higher exploratory power.

With regard to sleep quality metrics, both people were able to have stronger regression models with sleep minutes metric. According to other studies in the literature review, this metric was frequently used in various disciplines and appears to be usually influenced by other factors, such as physical activity. Not an exception in this project - both individuals were able to have a sleep minutes model. However, for person A, the greatest impact of sleep minutes was whether the day was a weekend (or public holiday) or a weekday. On the other hand, for person B, the sleep minutes model was influenced by time series clustering analysis results, which is a unique piece of the experiment that was brought about by this project.

The application of the results of the time series cluster analysis in regression models demonstrated promise for future research, as they proved to be significant in certain customized models within this project. As a pilot version, this method was used only on steps data within the Awake period, yet it could be extended and tested on heart rate, or even on both combined. In addition, various modifications can be introduced, such as different aggregation intervals. The nuance with clustering analysis is that it cannot be fully automated and requires additional check to see if the results of the clusters can be interpret to understand what type of days may impact different sleep metrics. It gives a deeper look at the activity, behavior, and routines of each person that can impact sleep quality.

It was interesting to see that some of the sleep quality metrics differed for both people. For example, for person A the model for awake minutes was created, which was impacted by dtw clustering results, whether she has more intensive activity two hours before bed, and whether it was a weekend. However, for person B, the light sleep minutes model showed more explanatory power, which was affected by the average heart rate, the intensity score and whether it was a weekend. This showed the importance that each person is different and different factors can influence their sleep. This created tool enables people to find the metrics that are impacted the most within their own help without trying to follow general recommendations that were made on multiple people or even specific groups.

Unfortunately, the new metrics included in this work that were related to sleep structure, such as REM, deep, and light sleep ratios, did not show a strong relationship with physical activity with any person. However, it is important to acknowledge that wearable devices currently lack precision in detecting sleep phases, with sleep minutes being the only metric measured with relatively higher accuracy. Knowing how fast technology involves, it is valuable to keep these metrics in the methodology.

From an economic perspective, this experiment demonstrated that it is possible to detect relationships between physical activity and sleep quality using wearable devices, without the need for expensive laboratory experiments. It enables people to track their biometrics by pinpointing physical activity metrics that can impact sleep quality, make actionable changes, and observe if these changes improve the selected metrics. As some reviewed studies showed, just by being informed about how to take care of your sleep and being track that, it actually can increase sleep quality. This project also highlighted the importance of personalized analysis. For example, some studies found a correlation between a higher number of steps and improved sleep quality in larger populations. However, for individuals A and B, the step count did not significantly affect their sleep duration. This discrepancy could be due to the fact that general studies often capture a brief period of a person's life, suggesting that more active individuals with higher step counts generally sleep better. However, this does not necessarily mean that increasing your step count in a single day will improve your quality of sleep.

Implementing a personalized approach in workplaces can improve employee satisfaction by showing that the company cares about their health and individuality. Improved sleep quality among employees can lead to increased productivity, engagement, and creativity. In addition, good sleep is associated with fewer health problems, resulting in fewer sick leave and fewer workplace accidents. Furthermore, if individuals can extend their sleep duration to the suggested 7 hours, it could potentially reduce healthcare expenses related to poor sleep quality.

Since this method uses data from the Fitbit wearable device, it can be applied to any commonly used wearable device, making the solution scalable and reused on current wearable devices without the need to buy a new one, to make it a more sustainable option. It can be developed into a user-friendly application in which individuals can select different metrics to see how physical activity impacts their sleep quality and identify areas of focus.

Finally, although this finding was not the primary focus of the project, it should be mentioned the significant impact of weekends on sleep quality. According to the models, weekends add almost an hour to the total sleep time on average for both individuals. This metric was included in most models and has consistently been proven to be significant. Unfortunately, it is a metric that individuals have limited control over. However, future studies in Lithuania or Europe could explore the broader implications of poor sleep on the economy. Such studies might calculate how the additional sleep gained from non-working days can save costs, and enhance productivity and engagement at the workplace.

3.5. Limitations and future work recommendations

One significant limitation in understanding the relationship between various factors and sleep quality is the lack of comprehensive biometrics. Sleep is influenced by multiple factors, including physical activity, stress, diet, and alcohol consumption. To improve the explanation of what most significantly affects sleep, it would be beneficial to collect additional data on diet, alcohol consumption, and stress levels daily.

Also, it is important to mention that the accuracy of wearable devices is still questionable, and while sleep minutes and step count already show acceptable accuracy, other metrics are still not reliable, and the results can still vary for technical reasons.

For future work, this methodology can be extended not only to include more biometrics but also to incorporate additional methods. For example, logistic regression could be used to classify sleep quality into categories such as poor, moderate, and good, or just poor-good. The previous studies usually used standard thresholds, such as 420 minutes (7 hours), but knowing that each individual is different, and sometimes progress, and not perfection, is more important, those thresholds can be adjusted based on person's goals - longer sleep, longer deep sleep minutes.

In addition, machine learning approaches could be used for forecasting. The provided Activity-Sleep metrics table can also be treated as time series data, enabling time series analysis. However, missing data and outlier detection should be reviewed with time-series practices. Code implementation can also be created to load not only Fitbit data, but also any other wearable device data that contains simple metrics such as minute-by-minute (or other interval) steps and heart rate data, overall sleep, deep sleep, REM sleep, light sleep and awake minutes, and total time in bed with waking up and sleep times.

Conclusion

1. The literature review confirmed the economic and health importance of addressing sleep quality, emphasizing the high costs for government associated due to the poor sleep outcomes. In addition, for companies, sleep quality affects worker productivity, engagement, creativity, and the frequency of accidents at work. The literature revealed various findings on human's physical activity impact on sleep quality, which in some cases had opposite relationships, showing the importance of more personalized approach in physical activity and sleep quality relationship analysis.

2. A set of indicators was proposed for the analysis of the data provided by the company. This is a new approach to use various metrics to find a personalized evaluation of the relationship between physical activity and sleep quality metrics.

3. After applying the developed and implemented Python methodology to the real data of two individuals, the results demonstrated the impact of human physical activity on sleep quality metrics and highlighted the importance of personalized models, which revealed varying results between individuals.

4. The inclusion of time-series clustering results using the k-means method combined with the Dynamic Time Warping (DTW) distance metric in multiple regression models has proven to be a significant regressor, highlighting the effectiveness of this novel approach.

5. The findings advocate for a shift towards personalized sleep medicine, where customized recommendations can significantly improve sleep quality and improve overall productivity and health outcomes. This personalized strategy has economic implications by potentially reducing healthcare costs and improving workplace efficiency and productivity.

This project was carried out in collaboration with the company *dhealthIQ*, which specializes in remote patient monitoring and personalized therapy solutions. Furthermore, this project was presented at the student conference *Mathematics and Natural Sciences: Theory and Application*, hosted by the Faculty of Mathematics and Natural Sciences at Kaunas University of Technology on May 17, 2024.

List of References

[1] Margaret D. Maggard, Abdulghani Sankari, and Marco Cascella. *Upper Airway Resistance Syndrome*. StatPearls Publishing, Treasure Island, FL, 2024. URL https://www.ncbi.nlm.nih.gov/books/NBK564402/figure/article-30790.image.f1/.

[2] Digital Health Central. Fitbit sleep tracker accuracy, 2022. URL https://digitalhealthcentral.com/2022/04/05/fitbit-sleep-tracker-accuracy/.

[3] A. K. Patel, V. Reddy, K. R. Shumway, and et al. *Physiology, Sleep Stages*. StatPearls Publishing, Treasure Island, FL, 2024. Updated 2024 Jan 26.

[4] Tae Won Kim, Jong-Hyun Jeong, and Seung-Chul Hong. The impact of sleep and circadian disturbance on hormones and metabolism. *International Journal of Endocrinology*, 2015:591729, 2015. doi: 10.1155/2015/591729. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4377487/. Accessed: Your Access Date Here.

[5] John A. Caldwell, J. Lynn Caldwell, Lauren A. Thompson, and Harris R. Lieberman. Fatigue and its management in the workplace. *Neuroscience Biobehavioral Reviews*, 96:272–289, 2019. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2018.10.024. URL https://www.sciencedirect.com/science/article/pii/S0149763418305220.

[6] Micheline Wille Goran Medic and Michiel EH Hemels. Short- and long-term health consequences of sleep disruption. *Nature and Science of Sleep*, 9:151–161, 2017. doi: 10.2147/NSS.S134864. URL https://www.tandfonline.com/doi/abs/10.2147/NSS.S134864. PMID: 28579842.

[7] S. Chokroverty. Overview of sleep & sleep disorders. *Interactive*, 131(2), 2010. Available online at https://journals.lww.com/ijmr/Fulltext/2010/31020/Overview_of_sleep___sleep_disorders.4.aspx.

[8] Sujin Lee, Ji Hyun Kim, and Jae Ho Chung. The association between sleep quality and quality of life: a population-based study. *Sleep Medicine*, 84:121–126, 2021. ISSN 1389-9457. doi: https://doi.org/10.1016/j.sleep.2021.05.022. URL https://www.sciencedirect.com/science/article/pii/S1389945721003014.

[9] Melanie A. Hom, Carol Chu, Megan L. Rogers, and Thomas E. Joiner. A meta-analysis of the relationship between sleep problems and loneliness. *Clinical Psychological Science*, 8 (5):799–824, 2020. doi: 10.1177/2167702620922969. URL https://doi.org/10.1177/2167702620922969.

[10] Eleonora Tobaldini, Giorgio Costantino, Monica Solbiati, Chiara Cogliati, Tomas Kara, Lino Nobili, and Nicola Montano. Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. *Neuroscience Biobehavioral Reviews*, 74:321–329, 2017. ISSN 0149-7634.

doi: https://doi.org/10.1016/j.neubiorev.2016.07.004. URL https://www.sciencedirect.com/science/article/pii/S0149763416302184. Stress, Behavior and the Heart.

[11] Marie-Rachelle Narcisse, Pearl A. McElfish, Mario Schootman, James P. Selig, Tracie Kirkland, Samy I. McFarlane, Holly C. Felix, Azizi Seixas, and Girardin Jean-Louis. Type 2 diabetes and health-related quality of life among older medicare beneficiaries: The mediating role of sleep. *Sleep Medicine*, 117:209–215, 2024. ISSN 1389-9457. doi: 10.1016/j.sleep.2024.03.015. URL https://www.sciencedirect.com/science/article/pii/S1389945724001138.

[12] D. Hillman et al. The economic cost of inadequate sleep. *Interactive*, 41(8):zsy083, 2018. URL https://doi.org/10.1093/sleep/zsy083. Accessed 2023-.

[13] Marco Hafner, Martin Stepanek, James Taylor, Wendy M. Troxel, and Christian van Stolk. Why sleep matters-the economic costs of insufficient sleep: A cross-country comparative analysis. *Rand health quarterly*, 6(4):11, 2017.

[14] Statistics Lithuania. Darbuotojų skaičius, 2021. URL https://osp.stat.gov.lt/darbo-rinka-lietuvoje-2021/darbo-uzmokestis-darbo-sanaudos-ir-streikai/darbuotoju-skaicius. Accessed: 2024-05-23.

[15] Christopher M. Barnes and Nathaniel F. Watson. Why healthy sleep is good for business. *Sleep Medicine Reviews*, 47:112–118, 2019. ISSN 1087-0792. doi: https://doi.org/10.1016/j.smrv.2019.07.005. URL https://www.sciencedirect.com/science/article/pii/S1087079219300449.

[16] Eunyoung Park, Hye Young Lee, and Chang Sook Park. Association between sleep quality and nurse productivity among korean clinical nurses. *Journal of Nursing Management*, 26(8):1051–1058, 2018. doi: 10.1111/jonm.12634. URL https://pubmed.ncbi.nlm.nih.gov/29855101/. Epub 2018 Jun 1.

[17] Mark R. Rosekind, Kevin B. Gregory, Melissa M. Mallis, Stephanie L. Brandt, Brenda Seal, and Debra Lerner. The cost of poor sleep: workplace productivity loss and associated costs. *Journal of Occupational and Environmental Medicine*, 52(1):91–98, 2010. doi: 10.1097/JOM.0b013e3181c78c30. URL https://pubmed.ncbi.nlm.nih.gov/20042880/.

[18] Nancy S. Redeker, Claire C. Caruso, Shireen D. Hashmi, Janet M. Mullington, Michael Grandner, and Timothy I. Morgenthaler. Workplace interventions to promote sleep health and an alert, healthy workforce. *Journal of Clinical Sleep Medicine*, 15(4):649–657, 2019. doi: 10.5664/jcsm.7734. URL https://pubmed.ncbi.nlm.nih.gov/30952228/.

[19] Steven E Lerman, Eleyna Eskin, David J Flower, Elizabeth C George, Bill Gerson, Natalie Hartenbaum, Steven R Hursh, and Martin Moore-Ede. Fatigue risk management in the workplace. *Journal of Occupational and Environmental Medicine*, 54(2):231–258, 2012. doi: 10.1097/JOM.0b013e318247a3b0.

[20] Sergio Garbarino, Giovanni Tripepi, and Nicola Magnavita. Sleep health promotion in the workplace. *International Journal of Environmental Research and Public Health*, 17(21):7952, 2020. ISSN 1660-4601. doi: 10.3390/ijerph17217952. URL https://www.mdpi.com/1660-4601/17/21/7952.

[21] Kateryna Maltseva. Wearables in the workplace: The brave new world of employee engagement. *Business Horizons*, 63(4):493–505, 2020. ISSN 0007-6813. doi: https://doi.org/10.1016/j.bushor.2020.03.007. URL https://www.sciencedirect.com/science/article/pii/S0007681320300367.

[22] S. Johnson and Sami Al Abdulkareem. 362 efficacy of sleep hygiene advices for individuals with poor sleep quality. *Sleep*, 44, 2021. doi: 10.1093/SLEEP/ZSAB072.361.

[23] Lucie Urbanová, Martina Sebalo Vňuková, Martin Anders, Radek Ptáček, and Jitka Bušková. The updating and individualizing of sleep hygiene rules for non-clinical adult populations. *Prague Medical Report*, 124(4):329–343, 2023. doi: 10.14712/23362936.2023.26. URL https://doi.org/10.14712/23362936.2023.26.

[24] Feifei Wang and Szilvia Boros. The effect of physical activity on sleep quality: a systematic review. *European Journal of Physiotherapy*, 23(1):11–18, 2021. doi: 10.1080/21679169.2019.1623314. URL https://doi.org/10.1080/21679169.2019.1623314.

[25] Evan D Chinoy, Joseph A Cuellar, Kirbie E Huwa, Jason T Jameson, Catherine H Watson, Sara C Bessman, Dale A Hirsch, Adam D Cooper, Sean P A Drummond, and Rachel R Markwald. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*, 44(5):zsaa291, 12 2020. ISSN 0161-8105. doi: 10.1093/sleep/zsaa291. URL https://doi.org/10.1093/sleep/zsaa291.

[26] Aamir R. Memon, Charlotte C. Gupta, Meagan E. Crowther, Sally A. Ferguson, Georgia A. Tuckwell, and Grace E. Vincent. Sleep and physical activity in university students: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 58:101482, 2021. ISSN 1087-0792. doi: https://doi.org/10.1016/j.smrv.2021.101482. URL https://www.sciencedirect.com/science/article/pii/S1087079221000678.

[27] Daniel J. Buysse, Charles F. Reynolds, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, 28(2):193–213, 1989. ISSN 0165-1781. doi: https://doi.org/10.1016/0165-1781(89)90047-4. URL https://www.sciencedirect.com/science/article/pii/0165178189900474.

[28] Marco Fabbri, Alessia Beracci, Monica Martoni, Debora Meneo, Lorenzo Tonetti, and Vincenzo Natale. Measuring subjective sleep quality: A review. *International Journal of Environmental Research and Public Health*, 18(3), 2021. ISSN 1660-4601. doi: 10.3390/ijerph18031082. URL https://www.mdpi.com/1660-4601/18/3/1082.

[29] Justyna Godos, Giuseppe Grosso, Sabrina Castellano, Fabio Galvano, Filippo Caraci, and Raffaele Ferri. Association between diet and sleep quality: A systematic review. *Sleep Medicine Reviews*, 57:101430, 2021. ISSN 1087-0792. doi: https://doi.org/10.1016/j.smrv.2021.101430. URL https://www.sciencedirect.com/science/article/pii/S1087079221000150.

[30] Antonino Crivello, Paolo Barsocchi, Michele Girolami, and Filippo Palumbo. The meaning of sleep quality: A survey of available technologies. *IEEE Access*, 7:167374–167390, 2019. doi: 10.1109/ACCESS.2019.2953835.

[31] Takeshi Nishiyama, Tomoki Mizuno, Masayo Kojima, Sadao Suzuki, Tsuyoshi Kitajima, Kayoko Bhardwaj Ando, Shinichi Kuriyama, and Meiho Nakayama. Criterion validity of the pittsburgh sleep quality index and epworth sleepiness scale for the diagnosis of sleep disorders. *Sleep Medicine*, 15(4):422–429, 2014. ISSN 1389-9457. doi: https://doi.org/10.1016/j.sleep.2013.12.015. URL `https://www.sciencedirect.com/science/article/pii/S1389945714000768`.

[32] Garber, Carol Ewing, Ph.D., FACSM, Blissmer, Bryan, Ph.D., Deschenes, Michael R., PhD, FACSM, Franklin, Barry A., Ph.D., FACSM, Lamonte, Michael J., Ph.D., FACSM, Lee, I-Min, M.D., Sc.D., FACSM, Nieman, David C., Ph.D., FACSM, and Swain, David P., Ph.D., FACSM. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine & Science in Sports & Exercise*, 43(7):1334–1359, July 2011. doi: 10.1249/MSS.0b013e318213fefb.

[33] Gina Sprint, Diane J. Cook, and Maureen Schmitter-Edgecombe. Unsupervised detection and analysis of changes in everyday physical activity data. *Journal of Biomedical Informatics*, 63: 54–65, 2016. ISSN 1532-0464. doi: https://doi.org/10.1016/j.jbi.2016.07.020. URL `https://www.sciencedirect.com/science/article/pii/S1532046416300740`.

[34] Brian R. MacIntosh, Juan M. Murias, Daniel A. Keir, and Jamie M. Weir. What is moderate to vigorous exercise intensity? *Frontiers in Physiology*, 12:682233, September 2021. doi: 10.3389/fphys.2021.682233. This article is part of the Research Topic "Exercise Prescription in Metabolic Diseases: An Efficient Medicine Towards Prevention and Cure".

[35] Janina Janurek, Sascha Abdel Hadi, Andreas Mojzisch, and Jan Alexander Häusser. The association of the 24 hour distribution of time spent in physical activity, work, and sleep with emotional exhaustion. *International Journal of Environmental Research and Public Health*, 15 (9), 2018. ISSN 1660-4601. doi: 10.3390/ijerph15091927. URL `https://www.mdpi.com/1660-4601/15/9/1927`.

[36] Mario A. Leocadio-Miguel and John Fontenele-Araújo. *Actigraphy*, pages 411–424. Springer International Publishing, Cham, 2022. ISBN 978-3-030-85074-6. doi: 10.1007/978-3-030-85074-6_37. URL `https://doi.org/10.1007/978-3-030-85074-6_37`.

[37] A. Castro, W.M. Anderson, and R. Nakase-Richardson. Actigraphy. In Clete A. Kushida, editor, *Encyclopedia of Sleep*, pages 88–91. Academic Press, Waltham, 2013. ISBN 978-0-12-378611-1. doi: https://doi.org/10.1016/B978-0-12-378610-4.00145-5. URL `https://www.sciencedirect.com/science/article/pii/B9780123786104001455`.

[38] Shahab Haghayegh, Sepideh Khoshnevis, Michael H Smolensky, Kenneth R Diller, and Richard J Castriotta. Accuracy of wristband fitbit models in assessing sleep: Systematic review and meta-analysis. *J Med Internet Res*, 21(11):e16273, Nov 2019. ISSN 1438-8871. doi: 10.2196/16273. URL `http://www.jmir.org/2019/11/e16273/`.

[39] Laurent Degroote, Gilles Hamerlinck, Karolien Poels, Carol Maher, Geert Crombez, Ilse De Bourdeaudhuij, Ann Vandendriessche, Rachel G Curtis, and Ann DeSmet. Low-cost

consumer-based trackers to measure physical activity and sleep duration among adults in free-living conditions: Validation study. *JMIR Mhealth Uhealth*, 8(5):e16674, May 2020. ISSN 2291-5222. doi: 10.2196/16674. URL `http://www.ncbi.nlm.nih.gov/pubmed/32282332`.

[40] Fitbit. How is my sleep score calculated in the fitbit app?, 2024. URL `https://support.google.com/fitbit/answer/14236513?hl=en#zippy=%2Chow-is-my-sleep-score-calculated-in-the-fitbit-app`. Accessed: 2024-05-19.

[41] Garmin. How does garmin assign sleep quality scores?, 2024. URL `https://support.garmin.com/en-US/?faq=DWcdBazhr097VgqFufsTk8`. Accessed: 2024-05-19.

[42] Jerome T. Galea, Karen Ramos, Julia Coit, Lauren E. Friedman, Carmen Contreras, Milagros Dueñas, Noris Hernandez, Caroline Muster, Leonid Lecca, and Bizu Gelaye. The use of wearable technology to objectively measure sleep quality and physical activity among pregnant women in urban lima, peru: A pilot feasibility study. *Maternal and Child Health Journal*, 24(7):823–828, July 2020. doi: 10.1007/s10995-020-02931-5. URL `https://doi.org/10.1007/s10995-020-02931-5`.

[43] Robert P. Smith, Cole Easson, Sarah M. Lyle, Ritishka Kapoor, Chase P. Donnelly, Eileen J. Davidson, Esha Parikh, Jose V. Lopez, and Jaime L. Tartar. Gut microbiome diversity is associated with sleep physiology in humans. *PLOS ONE*, October 7 2019. doi: 10.1371/journal.pone.0222394. URL `https://doi.org/10.1371/journal.pone.0222394`.

[44] Alycia N. Sullivan Bisson, Stephanie A. Robinson, and Margie E. Lachman. Walk to a better night of sleep: testing the relationship between physical activity and sleep. *Sleep Health*, 5 (5):487–494, 2019. ISSN 2352-7218. doi: https://doi.org/10.1016/j.sleh.2019.06.003. URL `https://www.sciencedirect.com/science/article/pii/S2352721819301056`.

[45] Ehizogie Paul Adeghe, Chioma Anthonia Okolo, and Olumuyiwa Tolulope Ojeyinka. A review of wearable technology in healthcare: Monitoring patient health and enhancing outcomes. *Open Access Research Journal of Multidisciplinary Studies*, 7(01):142–148, 2024. doi: 10.53022/oarjms.2024.7.1.0019.

[46] Muhammad Irfan Sharif, Jian Ping Li, Javeria Naz, and Iqra Rashid. A comprehensive review on multi-organs tumor detection based on machine learning. *Pattern Recognition Letters*, 131: 30–37, 2020. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2019.12.006. URL `https://www.sciencedirect.com/science/article/pii/S0167865519303691`.

[47] Evangelos Christodoulou, Jing Ma, Gary S Collins, Ewout W Steyerberg, Joris Y Verbakel, and Ben Van Calster. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, 110:12–22, 2019. doi: 10.1016/j.jclinepi.2019.02.004. URL `https://www.jclinepi.com/article/S0895-4356(18)31081-3/abstract#articleInformation`.

[48] Payel Bhattacharjee, Sambit Prasad Kar, and Nirmal Kumar Rout. Sleep and sedentary behavior analysis from physiological signals using machine learning. *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pages 240–243, 2020. doi: 10.1109/ICIMIA48430.2020.9074883.

[49] Mark V Albert, Albert Sugianto, Katherine Nickele, Patricia Zavos, Pinky Sindu, Munazza Ali, and Soyang Kwon. Hidden markov model-based activity recognition for toddlers. *Physiological Measurement*, 41(2):025003, March 5 2020. doi: 10.1088/1361-6579/ab6ebb. Published 5 March 2020 • © 2020 Institute of Physics and Engineering in Medicine.

[50] Yasser El-Manzalawy, Orfeu Buxton, and Vasant Honavar. Sleep/wake state prediction and sleep parameter estimation using unsupervised classification via clustering. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 718–723, 2017. doi: 10.1109/BIBM.2017.8217742.

[51] Petra J. Jones, Mike Catt, Melanie J. Davies, Charlotte L. Edwardson, Evgeny M. Mirkes, Kamlesh Khunti, Tom Yates, and Alex V. Rowlands. Feature selection for unsupervised machine learning of accelerometer data physical activity clusters – a systematic review. *Gait & Posture*, 90:120–128, 2021. ISSN 0966-6362. doi: https://doi.org/10.1016/j.gaitpost.2021.08.007. URL `https://www.sciencedirect.com/science/article/pii/S0966636221002824`.

[52] Junhyoung Kim, Jin-Young Choi, Hana Kim, Taeksang Lee, Jaeyoung Ha, Sangyi Lee, Jungmi Park, Gyeong-Suk Jeon, and Sung-il Cho. Physical activity pattern of adults with metabolic syndrome risk factors: Time-series cluster analysis. *JMIR Mhealth Uhealth*, 11:e50663, Dec 2023. ISSN 2291-5222. doi: 10.2196/50663. URL `https://mhealth.jmir.org/2023/1/e50663`.

[53] Alejandro Polo-Molina, Eugenio F. Sánchez-Úbeda, José Portela, Rafael Palacios, Carlos Rodríguez-Morcillo, Antonio Muñoz, Celia Alvarez-Romero, and Carlos Hernández-Quiles. Analyzing mobility patterns of complex chronic patients using wearable activity trackers: A machine learning approach. *Engineering Proceedings*, 39(1), 2023. ISSN 2673-4591. doi: 10.3390/engproc2023039092. URL `https://www.mdpi.com/2673-4591/39/1/92`.

[54] Ali Kargarandehkordi and Peter Washington. Personalized prediction of stress-induced blood pressure spikes in real time from fitbit data using artificial intelligence: A research protocol. *medRxiv*, 2023. doi: 10.1101/2023.12.18.23300060. URL `https://www.medrxiv.org/content/early/2023/12/20/2023.12.18.23300060`.

[55] Jiaxing Liu, Yang Zhao, Boya Lai, Hailiang Wang, and Kwok Leung Tsui. Wearable device heart rate and activity data in an unsupervised approach to personalized sleep monitoring: Algorithm validation. *JMIR Mhealth Uhealth*, 8(8):e18370, Aug 2020. ISSN 2291-5222. doi: 10.2196/18370. URL `https://mhealth.jmir.org/2020/8/e18370`.

[56] Claudio Diaz, Corinne Caillaud, and Kalina Yacef. Unsupervised early detection of physical activity behaviour changes from wearable accelerometer data. *Sensors*, 22(21), 2022. ISSN 1424-8220. doi: 10.3390/s22218255. URL `https://www.mdpi.com/1424-8220/22/21/8255`.

[57] T. Mishra, M. Wang, A.A. Metwally, et al. Pre-symptomatic detection of covid-19 from smartwatch data. *Nature Biomedical Engineering*, 4:1208–1220, 2020. doi: 10.1038/s41551-020-00640-6. URL `https://doi.org/10.1038/s41551-020-00640-6`.

[58] Dinh-Van Phan, Chien-Lung Chan, and Duc-Khanh Nguyen. Applying deep learning for prediction sleep quality from wearable data. In *Proceedings of the 4th International Conference on*

Medical and Health Informatics, ICMHI '20, page 51–55, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450377768. doi: 10.1145/3418094.3418114. URL https://doi.org/10.1145/3418094.3418114.

[59] Alam Ahmad Hidayat, Arif Budiarto, and Bens Pardamean. Long short-term memory-based models for sleep quality prediction from wearable device time series data. *Procedia Computer Science*, 227:1062–1069, 2023. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2023.10.616. URL https://www.sciencedirect.com/science/article/pii/S1877050923017830. 8th International Conference on Computer Science and Computational Intelligence (ICCSCI 2023).

[60] Andrew D. Krystal and Jack D. Edinger. Measuring sleep quality. *Sleep Medicine*, 9:S10–S17, 2008. ISSN 1389-9457. doi: https://doi.org/10.1016/S1389-9457(08)70011-X. URL https://www.sciencedirect.com/science/article/pii/S138994570870011X. The Art of Good Sleep Proceedings from the 5th International Sleep Disorders Forum: Novel Outcome Measures of Sleep, Sleep Loss and Insomnia.

[61] Sandra Mochón-Benguigui, Araceli Carneiro-Barrera, María J. Castillo, et al. Role of physical activity and fitness on sleep in sedentary middle-aged adults: the fit-ageing study. *Scientific Reports*, 11(539), 2021. doi: 10.1038/s41598-020-79355-2. URL https://www.nature.com/articles/s41598-020-79355-2.

[62] Guillaume Chevance, Dimitri Baretta, Ahmed Jérôme Romain, et al. Day-to-day associations between sleep and physical activity: a set of person-specific analyses in adults with overweight and obesity. *Journal of Behavioral Medicine*, 45:14–27, 2022. doi: 10.1007/s10865-021-00254-6. URL https://link.springer.com/article/10.1007/s10865-021-00254-6#citeas.

[63] Ji Ni, Bowei Chen, Nigel M. Allinson, and Xujiong Ye. A hybrid model for predicting human physical activity status from lifelogging data. *European Journal of Operational Research*, 281(3):532–542, 2020. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2019.05.035. URL https://www.sciencedirect.com/science/article/pii/S0377221719304655. Featured Cluster: Business Analytics: Defining the field and identifying a research agenda.

[64] Hirofumi Tanaka, K. Monahan, and D. Seals. Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37 1:153–6, 2001. doi: 10.1016/S0735-1097(00)01054-8.

[65] J. M. Radin, N. Wineinger, E. Topol, and S. Steinhubl. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the usa: a population-based study. *The Lancet. Digital health*, 2 2:e85–e93, 2020. doi: 10.1016/s2589-7500(19)30222-5.

[66] Stephanie L. Orstad, Lauren Gerchow, Nikhil R. Patel, Meghana Reddy, Christina Hernandez, D. Wilson, and Melanie R. Jay. Defining valid activity monitor data: A multimethod analysis of weight-loss intervention participants' barriers to wear and first 100 days of physical activity. *Informatics (MDPI)*, 8, 2021. doi: 10.3390/informatics8020039.

[67] S. Prescott, J. Traynor, I. Shilliday, T. Zanotto, R. Rush, and T. Mercer. Minimum accelerometer wear-time for reliable estimates of physical activity and sedentary behaviour of people receiving haemodialysis. *BMC Nephrology*, 21, 2020. doi: 10.1186/s12882-020-01877-8.

[68] Samuel R LaMunion, R. Brychta, Pedro Saint-Maurice, Charles E. Matthews, and Kong Y Chen. Does wrist-worn accelerometer wear compliance wane over a free-living assessment period? an nhanes analysis. *Medicine and science in sports and exercise*, 2023. doi: 10.1249/MSS.0000000000003301.

[69] Pavel Senin. Dynamic time warping algorithm review. *ResearchGate*, 2008. URL `https://www.researchgate.net/profile/Pavel-Senin/publication/228785661_Dynamic_Time_Warping_Algorithm_Review/links/02bfe5100f11a7929f000000/Dynamic-Time-Warping-Algorithm-Review.pdf`. Accessed: 2024-03-28.

[70] John Paparrizos and Luis Gravano. k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1855–1870, 2015. URL `https://www.cs.columbia.edu/~gravano/Papers/2015/sigmod2015.pdf`. Accessed: 2024-03-28.

Appendices

1. Plots representing data analysis of person A

Fig. 14. Correlation matrix of physical activity and sleep quality metrics of person A

Table 17. Comparison of sleep metrics across different DTW clusters of person A

	y1		**y2**		**y3**		**y4**		**y5**	
c2_clusters_dtw	mean	std	mean	std	mean	std	mean	std	mean	std
0	390.5	49.6	74.3	19.4	221.7	32.7	94.5	22.7	52.0	10.2
1	386.2	46.1	71.6	19.9	222.1	30.5	92.4	22.0	53.9	12.0
2	404.1	52.8	79.3	22.1	228.9	35.5	95.9	22.3	59.1	12.4
3	379.6	45.8	73.5	21.5	213.3	33.4	92.9	21.1	55.2	13.0

Table 18. Comparison of sleep metrics between different K-Shape clusters for person A

	y1		**y2**		**y3**		**y4**		**y5**	
c3_clusters_kshape	mean	std	mean	std	mean	std	mean	std	mean	std
0	384.1	48.4	71.0	22.1	219.9	31.1	93.2	21.4	55.8	12.7
1	396.7	53.3	76.7	22.6	225.7	33.0	94.3	22.9	57.8	13.0
2	387.4	45.5	74.7	18.5	219.2	34.1	93.6	21.8	52.5	10.9

2. Plots representing data analysis of person B

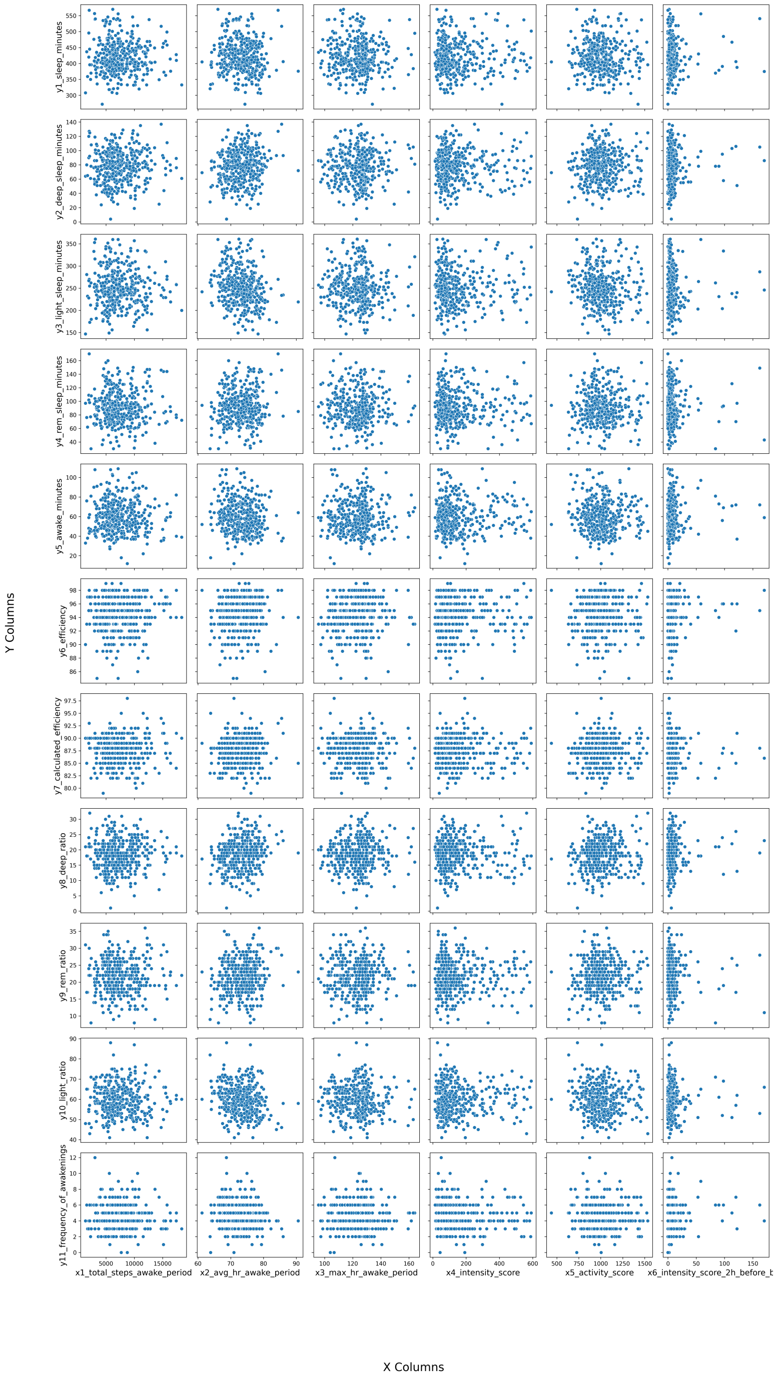


Fig. 15. Correlation matrix of physical activity and sleep quality metrics of person B

Table 19. Comparison of sleep metrics across different DTW clusters of person B

	y1		**y2**		**y3**		**y4**		**y5**	
c2_clusters_dtw	mean	std	mean	std	mean	std	mean	std	mean	std
0	428.1	52.8	80.5	22.5	250.7	41.1	96.9	25.9	61.8	17.2
1	421.6	46.8	75.2	20.9	255.6	37.9	90.1	25.1	58.7	14.5
2	426.7	48.9	80.8	21.6	254.3	41.9	92.6	22.9	60.8	16.5
3	410.8	53.2	77.2	23.8	244.2	40.9	89.4	22.4	57.1	15.9

Table 20. Comparison of sleep metrics between different K-Shape clusters for person B

	y1		**y2**		**y3**		**y4**		**y5**	
c3_clusters_kshape	mean	std	mean	std	mean	std	mean	std	mean	std
0	384.1	48.4	71.0	22.1	219.9	31.1	93.2	21.4	55.8	12.7
1	396.7	53.3	76.7	22.6	225.7	33.0	94.3	22.9	57.8	13.0
2	387.4	45.5	74.7	18.5	219.2	34.1	93.6	21.8	52.5	10.9