CS - C4100 Digital Health and Human Behavior
Project report
10.12.2025

Daily Activity Analysis with Fitbit Tracker Data

1. Introduction

This project aims to study the data collected by the Fitbit activity trackers and analyze users daily activity, steps, energy consumption, and sleep. The goal is to determine which daily and weekly rhythms can be observed among different users, both at the group and individual levels. This topic is important, as physical activity is critical for maintaining and promoting health. Physical activity is essential for good health and well-being, and its lack is one of the most significant risk factors for the development of non-communicable diseases (World Health Organization, 2024). Regular physical activity and high enough step count are directly associated with lower mortality risk and developing diseases such as cardiovascular diseases and type 2 diabetes (Kraus et al., 2019). It is worrying that globally 31% of adults and up to 80% of young people do not meet the recommended levels of physical activity (World Health Organization, 2024).

In recent years various smart devices, such as activity trackers have gained popularity, which has increased the amount of continuous monitoring of physical activity, sleep, and energy consumption in everyday life. Studies show that modern devices such as Fitbit provide reliable information on users activity trends and sleep patterns, making them useful for both individual health monitoring and large-scale population-level data collection (Haghayegh et al., 2019).

Kraus et al. (2019) mention that previous studies have shown that increasing physical activity improves health in multiple ways, and that the number of steps taken daily is linearly associated with reduced mortality and disease risk. In addition to that, previous studies also show that FitBit trackers can reliably monitor sleep and activity in large populations, although they do not replace clinical methods (Haghayegh et al., 2019).

It is still unclear how well Fitbit users actually follow the recommendations for physical activity and sleep. For adults, it is recommended that they get at least 7 hours of sleep a night in order to promote optimal health (Watson et al., 2015). It is not known what kind of patterns can be identified from the data, and what type of behavioral patterns can reveal about users health habits. In addition, the relationship between activity and sleep is relatively unexplored among the users of this dataset, which provides motivation for this project and leaves a clear research gap.

Although the monitoring of physical activity, sleep, and energy consumption has improved, it remains unclear how different behavioral patterns appear in everyday life among different users. Wearables enable this research with new accuracy, but the data they produce is still underused. In addition, previous research has shown that users commitment to daily activity can vary significantly (Rayward et al., 2021). This highlights the need to identify behavioral differences at both the individual and group levels. These differences underscore the motivation for this project to better understand how activity and sleep are actually distributed among different users over days and weeks.

The main research question is what kind of physical activity and sleep patterns can be identified from the data collected by the FitBit trackers at the group and individual levels, and how do these patterns align with the WHO's health recommendations? The aim is to find answers to how many steps users take on average per day, what kind of daily patterns the users have, and how different levels of activity affect energy consumption. The project also aims to superficially examine users sleep habits and whether their daily activity is somehow related to their sleep durations.

This report is divided into the following sections: Introduction, Problem formulation, Dataset description, Methods, Results, and Conclusion & Discussion.

2. Problem Formulation

Tracking daily activity provides users with information about their own health behavior. With digitalization, FitBit trackers and other health tracking devices have grown in popularity and an increasing number of people are tracking their daily habits. Unfortunately, traditional monitoring methods that are based only on averages do not reveal all activity changes or sleep stages, which is why FitBit tracker monitoring can be considered useful. Haghayegh et al. (2019) mention that although Fitbit is not as accurate as laboratory-based polysomnography, it provides a reliable and cost-effective tool for monitoring sleep quantity and structure in large data sets.

The main idea of the project is to find out how well different FitBit users actually follow the WHO recommendations regarding sleep and activity. WHO emphasizes that even a small amount of increased physical activity brings significant health benefits, while being sedentary is an independent risk factor for chronic illnesses (World Health Organization, 2020).

This project aims to create a multi-level analysis of data produced by FitBit trackers, which can be used to identify different individual patterns. These results will be compared with the official health recommendations by the WHO. The aim is to

identify possible hidden patterns by distinguishing between users with different sleep and activity profiles and investigating the differences between them.
The visualizations try to highlight the activity trends and step counts clearly.

The project offers an achievable and centralized solution to main research questions, observing behavioral patterns from data at the group and individual levels. The project's analyses can be easily compared to the recommendations, providing information on how this group of 30 people has performed.

3. Dataset description

The dataset used and analyzed in this project is called "FitBit Fitness tracker data," which was retrieved from Kaggle. The data was collected through the Amazon Mechanical Turk service between 3.12.-5.12. 2016. The data contains information on the physical activity, heart rate, and sleep monitoring of 30 volunteer FitBit users over a period of 31 days from 12.4-12.5.2016. The data was collected automatically and passively using FitBit device sensors. The dailyActivity.merged data set contains 940 observations, while the other data sets contain varying numbers of observations.

This material has been divided into several csv folders based on different data collected. The main categories are: daily data, hourly data, minute data, and background data (weight/physical). Different users can be identified in each file using their user ID.

Daily files contain information about users daily activities. The file includes variables such as TotalSteps, TotalDistance, Calories, and ActivityDate. Hourly files contain similar data to daily files, but the information is divided into hourly intervals. They can be used to examine, for example, at what time of day users are most active and how physical activity is distributed throughout the day. This data can be used to analyze users daily rhythms and activity profiles. The minute by minute data is the most accurate level of detail in the dataset, as it allows us to examine, for example, the users sleep status and how awake/asleep they are with minute-by-minute accuracy. The dataset also includes heart rate data, which allows us to examine users heart rate data with second by second accuracy.

The dataset contains also composite files, such as dailyactivity_merged.csv, which combines data from several files. The database contains this information in separate files, but the project mainly used data from combined files. For this project, the following features were analyzed from the collected data:

- TotalSteps: daily step count, which serves as a measure of physical activity
- Calories: calories burned during the day, which serves as a measure of energy consumption

- TotalMinutesAsleep: total duration of sleep, which serves as a measure of recovery
- SedentaryMinutes: minutes spent passively by the user, which serves as a measure of inactivity
- TotalActiveMinutes: a combination of the columns VeryActiveMinutes, FairlyActiveMinutes, and LightlyActive minutes. With this combination, it was possible to observe overall activity as a whole.

3.1 Data Preprocessing

Data preprocessing is an important step of the project, as it makes sure that the material is consistent and error-free for the intended visualizations. Since the project data consisted of several different files, different steps had to be done before the visualizations could be created.

First, the dailyActivity.merged and sleepDay.merged files were printed to see what kind of data they contained. It was checked that all the data that was used in this project was from the period 12.4-12.5.2016. The DailyActivity-merged file contained 940 data points and the sleepDay.merged file contained 413 observations. The main analysis for the project was done by combining these two files.

After that, duplicate values were removed and date changes were made by converting the files to datetime format so that sns-lineplot could be created. After that, the main analysis was created from the actual combined data by combining the activity data and unit data into df format. This was done so that all 940 daily activity rows were retained. Extra columns such as LoggedActivitiesDistance and TrackerDistance were removed because they were not relevant.

There were some missing values in the unit data, so a separate data set df_sleep was created, by removing rows where TotalMinutesAsleep was empty. This was done using the .dropna command.

In order to obtain data on overall activity, the VeryActiveMinutes, FairlyActiveMinutes, and LightlyActiveMinutes columns were combined. This new variable enabled the observation of physical activity throughout the day. Finally, the times in the TotalTimeInBed column were converted into hours, which would be easier to interpret in regression analysis.

4. Methods

The methods section of the project is divided into two parts: basic methods and advanced methods. Basic methods cover basic description, visualization, and correlation-based analysis, while advanced methods cover differences between user groups using clustering. The purpose of the methods is to produce interesting and useful insights from the data used about user activity and sleep behavior, as well as to identify possible recurring behavior patterns.

4.1 Basic methods

The data used in this project is visualized in many different ways to gain a better understanding of different behavior patterns. First, a group-level visualization was created as a bar chart showing the average number of steps taken by users compared to the 10,000 steps recommended by the WHO. The bar chart was created using the sns.barplot command and was designed to provide an overview of how active users are in relation to the recommendations.

Next, a timeline was used to visualize users daily activity compared to international health recommendations. The timeline (sns.lineplot) was used to visualize the groups daily rhythm by calculating the average number of steps per hour using data collected with minute by minute accuracy. This showed at what time of day the users were the most active. The timeline was also used to visualize the daily step count of an individual test subject throughout the entire observation period in order to obtain information about the variation in individual activity.

Pearson's correlation was also used in the project to evaluate the possible linear relationship between sleep, activity and energy consumption. The results were visualized using a correlation heat map, which made it easy to see which variables were related to each other and how strongly. In the sleep analysis, the results were compared to the health recommendations, which mention that adults should sleep at least 7 hours per night (Watson et al., 2015). The relationship between users total activity minutes and calories was described using sns.regplot, which is a linear regression model. This revealed that activity explained part of the energy consumption, even though the variation was relatively large.

4.2 Advanced methods

In order to gain depth in the project's findings, a more advanced analysis method was also used: K-means clustering. K-means clustering is an unsupervised machine learning method that classifies observations into groups based on similarity (CS-C4100 Digital Health and Human Behavior assignment material, 2025). For this visualization, the TotalSteps and TotalMinutesAsleep variables were used as input to obtain a deeper analysis of whether there was any grouping, e.g., between users who exercised little and those who slept longer. The variables were standardized using StandardScaler so that their different units of measurement would not affect the clustering. K-means was chosen as a more advanced analysis method because it can identify structures that basic analysis cannot produce

5. Results

The aim of the project was to find out what kinds of physical activity and sleep patterns can be identified from the data collected at the group and individual levels, and how do these patterns align with the WHO's health recommendations? Multiple visualizations were made to find answers to these questions.

The first visualization is a bar chart showing the average number of steps taken by users compared to WHO recommendations. This can be seen in Figure 1. The average number of steps taken by users per day over 31 days was 7637, which is approximately 76% of the WHO's recommendation of 10,000 steps. This shows that, on average, users did not reach the recommended amount.

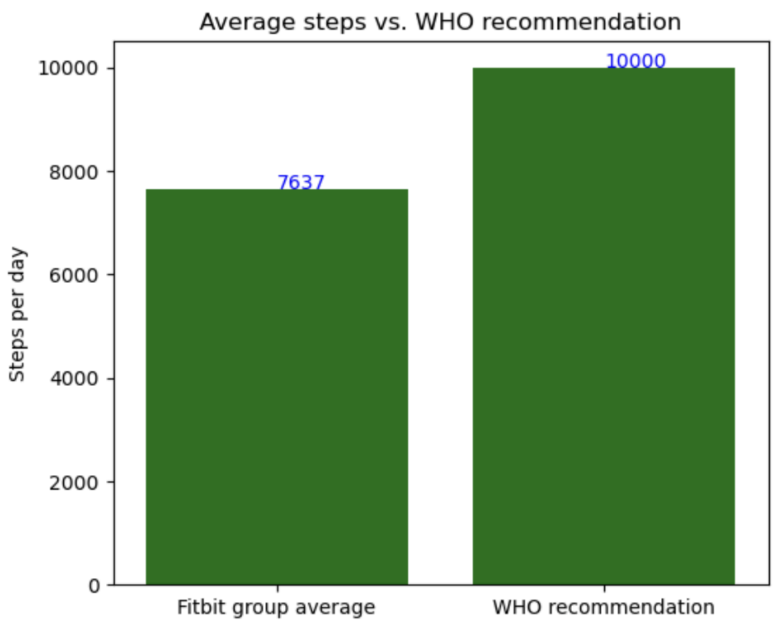

Figure 1: Average steps of the users compared to WHOs recommendation.

Figure 2 shows the variations in the average daily step count of users throughout the observation period. The graph shows that the number of steps taken by users varies significantly between days and remains on average between 7000 and 8200 steps. The drop at the end of the graph is probably due to the very low activity levels of individual users or not using the Fitbit tracker. Overall, there was a significant amount of daily variation, with individual days deviating from the average in both directions. The average number of steps taken by users did not reach the recommended 10,000 steps on any single day.

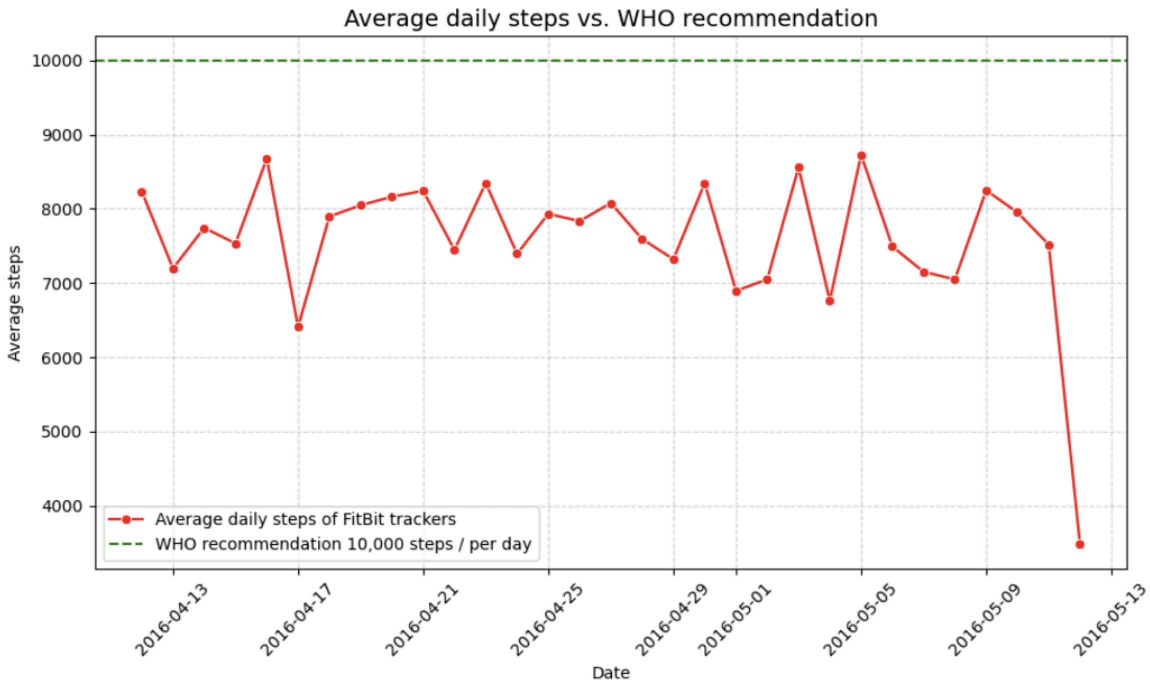

Figure 2: Average daily steps compared to WHO recommendation over the observation period.

Figure 3 shows the average number of steps compared to each hour of the day, providing an overview of the users activity patterns. The graph shows that activity is almost non-existent between 0 and 5 a.m., suggesting that users are asleep. Between 6 and 7 a.m., activity increases as users start their daily activities. Between 10 a.m. and 1 p.m., activity peaks. Day Activity remains fairly steady between 10 a.m. and 6 p.m., which indicates a typical workday rhythm and the activity it brings. As the day moves into the evening, activity decreases significantly between 8 p.m. and 11 p.m., indicating that users are calming down.

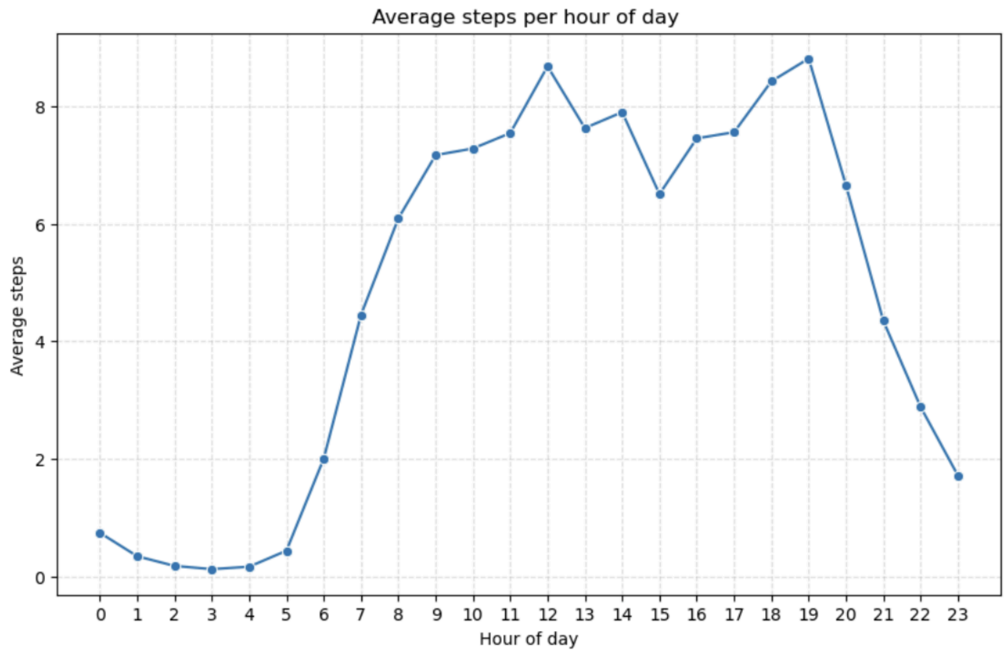

Figure 3: Average steps per hour of day.

Figure 4 shows a comparison at the individual level when looking at the average number of steps taken by a single user over a period of 31 days. It can be seen that the user is clearly more active than average, as their steps vary between 10,000 and 15,000 steps per day, which exceeds the WHO recommendation. Such high step counts indicate a high level of physical activity. Although there is variation between days, activity remains at a consistently high level. When comparing this individual with the average in Figure 2, this user is clearly more active than the average user.

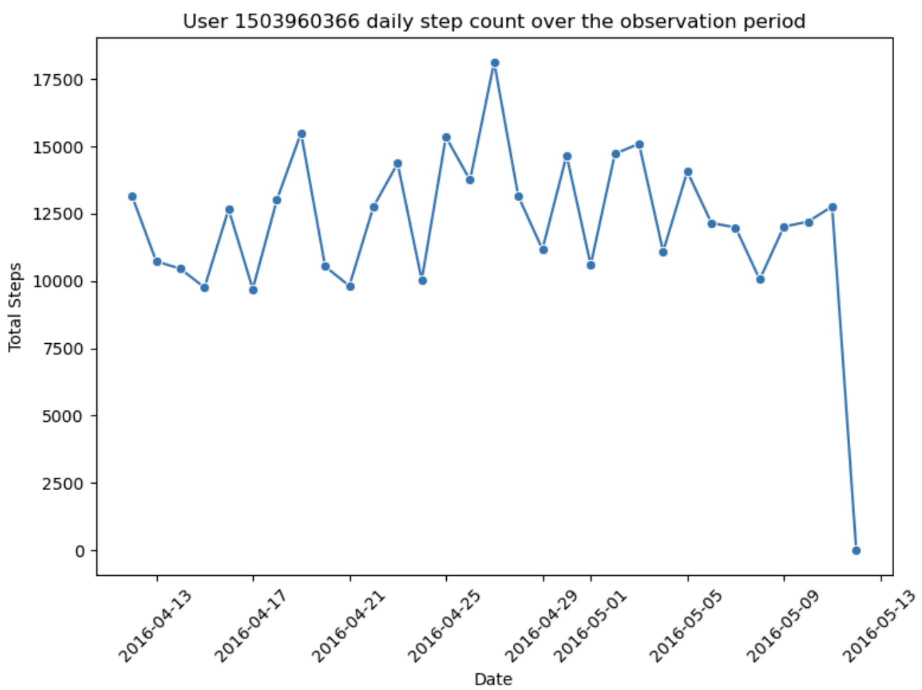

Figure 4: Daily step count of one user over the observation period.

To gain more insight into the project results, the correlation matrix shown in Figure 5 was created. The matrix can be used to examine the linear relationships between activity, sleep, and energy consumption. The heat map shows that TotalMinutesAsleep and TotaltimeInBed are strongly correlated, with a correlation coefficient of r=0.93. This observation was expected, as spending more time in bed often affects the amount of sleep.

The graph also shows that there is a clear positive correlation between Calories and VeryActiveMinutes (r= 0.61), which suggests that physical activity explains part of the energy consumption. These variables also correlate positively with the TotalSteps variable, as the correlation between TotalSteps and Calories is r=0.41 and between VeryActiveMinutes r=0.54. These findings support the conclusion that a higher number of steps is associated with both higher energy consumption and a more active lifestyle.

The heat map also shows a negative correlation, which means that no linear relationship can be observed between the variables. For example, TotalMinutesAsleep correlates only lightly with steps (r = −0.19) and calories (r = −0.03). This suggests that the users sleep time is not directly dependent on their daily activity level.

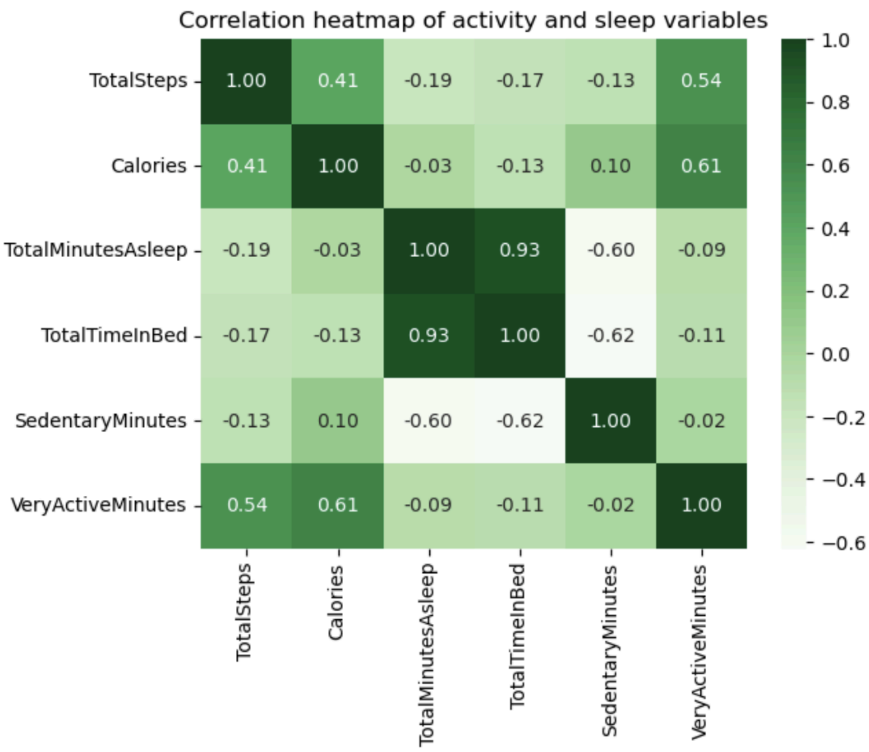

Figure 5: Correlation heatmap.

Figure 6 shows the relationship between total activity minutes and daily calories burned using a linear regression model. The graph shows that there is a clear positive trend between the two variables. The more active the users were during the day, the more calories they burned. The graph also shows the correlation coefficient between the two (r=0.47), which indicates a positive linear relationship. On average, users burned around 1800-2500 calories, although activity levels varied greatly between users. This indicates that basic metabolic rate and lifestyle factors are highly individual.

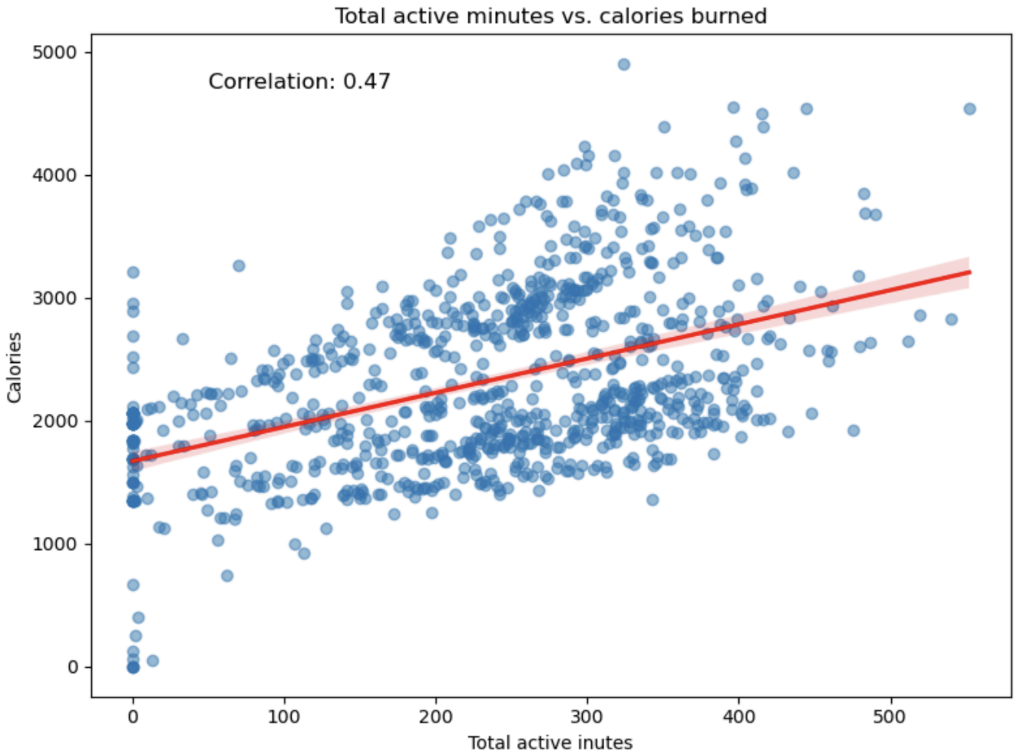

Figure 6: Total active minutes compared to calories burned.

Figure 7 shows the relationship between time spent in bed and actual sleep duration. It can be seen that the relationship is linear and most of the observations fall close to the regression line, suggesting that the length of sleep corresponds fairly logically to the time spent in bed. The average time spent in bed varies from 6-10 hours for most of the users, which seems appropriate for the recommendations. The observations are also in line with the correlation coefficient of the heat map (r=0.93). A few observations are seen far from the line, which suggest that sleep time is significantly shorter than time spent in bed, which may indicate sleep problems or, for example, the use of a phone in bed.

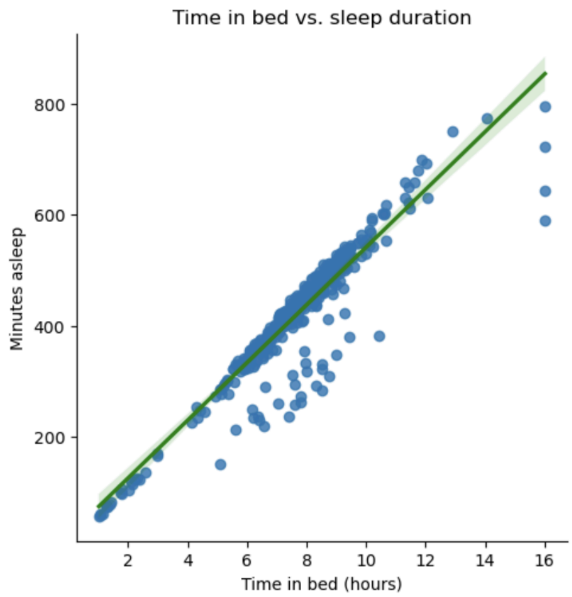

Figure 7: Time in bed compared to sleep duration.

In order to observe sleep more closely, the distribution of sleep across different days of the week is examined in Figure 8. The figure shows that the amount of sleep remains fairly consistent on weekdays, but there is more variation in sleep duration on weekends. People sleep significantly longer on Saturdays and Sundays than on weekdays. This suggests that users try to "catch up" on sleep on weekends, which is quite understandable. On average , it shows that the user's sleep is around 7 hours, which is in line with the recommendations.

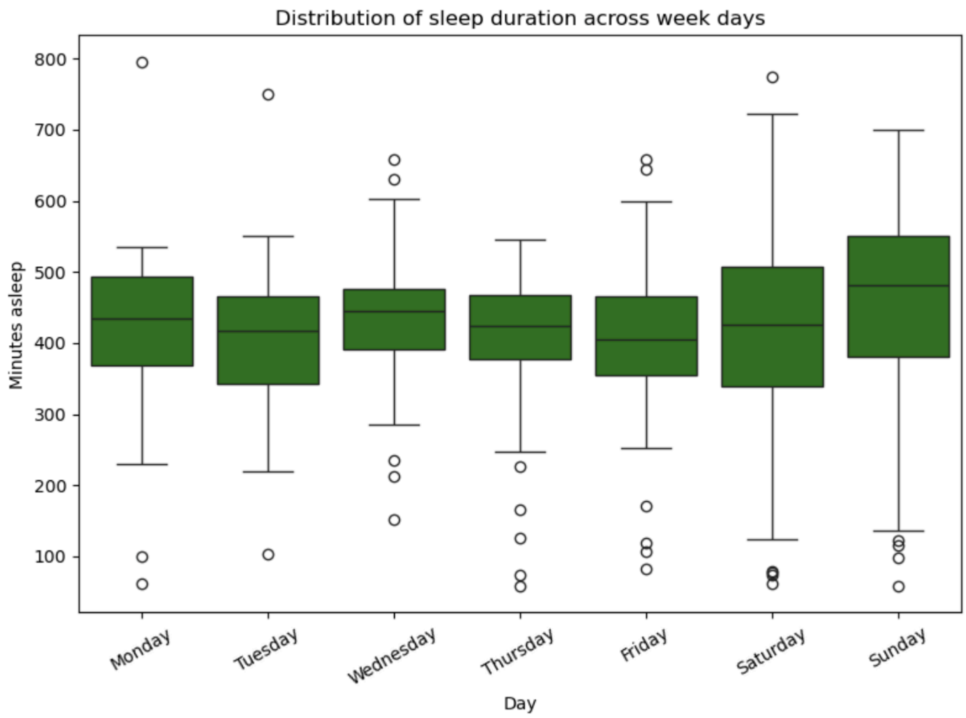

Figure 8: Distribution of sleep duration across weekdays.

For a more advanced analysis, K-means clustering was used to group users based on their daily steps and minutes slept. Figure 9 shows that users can be divided into three different groups:

Yellow: the first cluster, which indicates low step count but long average sleep time. This group includes users who sleep a long time on average but are not physically that active.

Green: the second cluster, which refers to moderate activity and medium-length duration of sleep. This group includes users who represent a balanced daily rhythm in terms of activity and sleep. The majority of users fall into this group.

Purple: the third cluster, which refers to short sleep duration but high activity. This group includes users who sleep less than other users but are clearly more active. The results for this group should be compared, for example, with recovery data in terms of possible challenges.

This graph reveals clear behavioral patterns among users that cannot be detected using normal averages. The graph supports the assumption that users have very different behavioral patterns that combine activity and sleep.

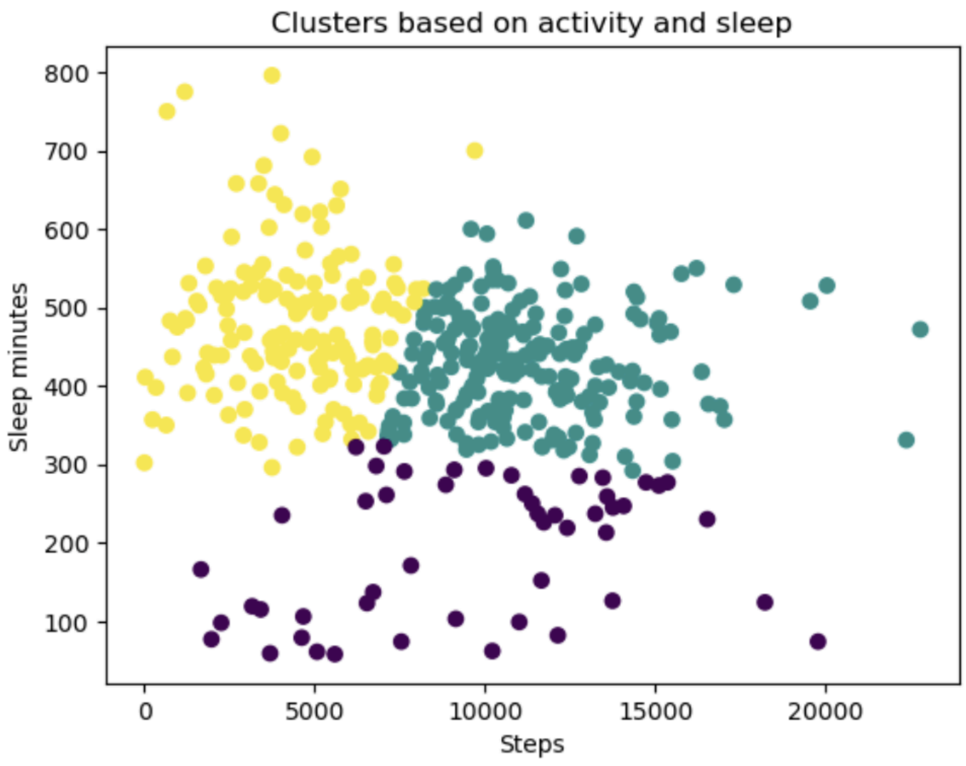

Figure 9: K-means clusters based on activity and sleep.

6. Conclusion & Discussion

To conclude the findings, the average number of steps taken by the users fell short of the WHO recommendations, but there were differences between the users. The results showed that physical activity was highest during the day and early evening and total activity minutes were related to energy consumption. The amount of sleep users got remained fairly consistent between weekdays, but on weekends the amount of sleep was higher. On average, the sleep hours were in line with the 7 hour recommendation. In general, the relationship between sleep and activity was varied, and clustering revealed that the most active users did not sleep significantly more or less than average. Using clustering, we identified three different user groups: 1. those who are not very active and sleep for a long time, 2. those who are moderately active and sleep on average, and 3. those who are very active and sleep little.

Although the results provide a broad picture of users daily activity and sleep, the analysis has several limitations. The observation period of 31 days was quite short and the sample size of 30 users was relatively small, which limits the generalizability of the findings. In addition, the lack of demographic information prevents the results from being correlated with, for example, age or health status, and the large amount of missing data in the sleep data reduces the reliability of the analysis.

A future aspect for this type of project could be to study the topic with a larger number of users, with more variables and a longer observation period to get a deeper analysis of different user patterns. Also, for example, studying user recovery could be the next step in producing higher quality analyses, because as can be interpreted from figure 9, one user group was very active but had short sleep duration, which may cause health concerns in the long term.

7. References

Aalto University. (2025). *CS-C4100: Digital Health and Human Behavior* [Unpublished course material]. School of Science.

Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R., & Castriotta, R. J. (2019).Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review andMeta-Analysis. *Journal of Medical Internet Research.*, *21*(11). https://doi.org/10.2196/16273

Kraus, W. E., Janz, K. F., Powell, K. E., Campbell, W. W., Jakicic, J. M., Troiano, R. P.,Sprow, K., Torres, A., & Piercy, K. L. (2019). Daily Step Counts for Measuring Physical Activity Exposure and Its Relation to Health. *Medicine & Science in Sports & Exercise.*, *51*(6), 1206–1212. https://doi.org/10.1249/MSS.0000000000001932

Rayward, A. T., Vandelanotte, C., Van Itallie, A., & Duncan, M. J. (2021). The Association Between Logging Steps Using a Website, App, or Fitbit and Engaging With the 10,000 Steps Physical Activity Program: Observational Study. *Journal of Medical Internet Research.*, *23*(6). https://doi.org/10.2196/22151

Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., Gangwisch, J., Grandner, M. A., Kushida, C., Malhotra, R. K., Martin, J. L., Patel, S. R., Quan, S. F., & Tasali, E. (2015). Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. *Sleep*, *38*(6), 843–844. https://doi.org/10.5665/sleep.4716

World Health Organization. (2024, 26 June). *Physical activity* [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/physical-activity Maailman terveysjärjestö

World Health Organization. (2020). *WHO guidelines on physical activity and sedentary behaviour*. Geneva: World Health Organization.

Arashnic. (n.d.). Fitbit Fitness Tracker Data. [Data set]. Kaggle. Retrieved November 19, 2025, from https://www.kaggle.com/datasets/arashnic/fitbit