This line chart is created by using the updated dataset, which is cleaned above. This line chart shows the Activity Score over time, from May 2021 through March 2024. The green and orange dashed lines represent the thresholds for low and medium activity levels, respectively. According to the example thresholds provided, below the green line describes the scores up to 33 which is considered low activity, and above the orange line describes the scores above 66 which is categorized as high activity. Also, between the green and orange lines means the scores between 34 and 66 indicate the medium activity. These thresholds are illustrative and depend on the specific definitions or thresholds for low, medium, and high activity times in our cleaned dataset.

This part has multiple graphics as it iterates through a list of explanatory variables such as Activity Score, Steps, and Inactive Time. When creating the scatter plot for each variable, it always compares it to Sleep Score. The scatter plots visually represent the relationship between daily physical activity and sleep quality, allowing the user to observe trends. For example, when viewing the correlation between Activity Score and Sleep Score, it is positive, which associates that higher levels of activity is better for sleep quality. On the other hand, Activity Burn demonstrates a negative correlation due to the fact that when burn increases, the sleep score decreases. The scatterplots provide a good understanding of how different aspects of daily activity impact sleep, which can be explored thoroughly through analysis.

For our project, we decided to look at sleep data over time, and how certain factors might have influenced a person’s sleep quality. Our data was taken from an Oura Ring that a person wore all day, every day for roughly 3 years. In our EDA, we first handled any missing or incorrect information. For example, if the Oura Ring had run out of battery overnight, there would be gaps in the sleep data. We took those days out of our data. The same went if the person hadn't worn the ring at all—we left out those days, even if we had other data from their daytime activities. This ensured our research stayed accurate, especially since we were looking at how what you do during the day affected your sleep.

Next, we tossed out any data that looked weird, like sleep numbers that just didn't make sense or were way off what we expected. These odd bits of data could have messed up our findings and they weren't a good reflection of the person’s normal sleep.

Also, we made sure that the records of daytime activities lined up exactly with the sleep scores from the same day. Each record of what the person did during the day should have matched with the sleep data for that night. It was important that we had the same number of records across all the data we were looking at—sleep, daytime activities, and readiness—to keep things accurate.

Finally, we reformatted some data so that it would be easier to work with. For example, the Bedtime start and Bedtime end columns contained the current year, which was already provided in another column, so it was removed. Also, the type of those columns were converted to datetime.

First, we have the histogram of Sleep Scores. This histogram displays a strong left skew, which means that this person was more likely to have a higher sleep score than a lower sleep score. Also, from the histogram, we can see that the person most frequently got a sleep score from 70-80. Understanding the frequency of each sleep score is important because it provides insights into the overall quality of sleep the individual typically experiences. If most of the scores are clustered within the 70-80 range, it suggests that the person generally has good sleep quality, with fewer instances of poor sleep. This can help us to identify patterns or trends in the data, such as whether the sleep quality is consistently high or varies significantly from night to night. Moreover, by analyzing the distribution of sleep scores, we can also assess the impact of different daytime activities or stressors on sleep quality. For example, if we notice that lower sleep scores occur more frequently after certain events or activities, it may indicate a correlation that warrants further investigation. Understanding these nuances allows us to make more informed conclusions and can help us predict future trends

Finally, The analysis of the sleep dataset has revealed that behaviors such as bedtime start, bedtime end, and sleep timing are not just arbitrary elements of an individual's sleep routine, but critical indicators of their alignment with circadian rhythms. When sleep patterns are in sync with the circadian rhythm, individuals tend to experience better sleep quality, as reflected in higher Sleep Scores.

Our findings support the hypothesis that earlier bedtimes and wake times, which are more in tune with the natural light-dark cycle, are associated with improved sleep quality. As you can see, as both bedtime start and bedtime end increases, sleep score decreases. In other words, having a sleep schedule that is aligned with sunlight tends to improve sleep score. This is in line with research suggesting that exposure to light in the morning and darkness at night helps to reinforce our natural circadian signals. Also, the absolute value for the correlation coefficient for bedtime start (0.55) is much larger than the correlation coefficient for bedtime end (.25). This suggests that these trends have more statistical support for bedtime start than bedtime end.

Scroll down

The disruption of these patterns, such as with irregular sleep schedules, can cause misalignment with the circadian rhythm. This misalignment can lead to difficulty falling asleep, staying asleep, or waking up, which in turn results in lower Sleep Scores. Such disruptions are common in modern society due to factors like the use of artificial lighting and the demands of work or social schedules that conflict with natural sleep patterns. To explore such disruptions, we used the standard deviation of sleep score as a proxy for the person's deviation from average sleep and wake times. As you can see, there is a moderate negative correlation between sleep score and the standard deviation of both bedtime start and bedtime end. This suggests that deviating from a normal bedtime decreases sleep quality, which matches our prediction.

Scroll down

Our analysis also identified a moderate linear relationship between sleep timing and sleep score, suggesting that individuals who maintain a longer duration of sleep tend to have higher Sleep Scores. Note that sleep timing values less than 3000 were ignored, since we determined that these values represented naps, not a full night's rest, which is what we are interested in measuring. This relationship underscores the importance of not just sleep timing, but also sleep duration, in achieving high-quality rest. Adequate sleep duration ensures that individuals progress through the necessary sleep cycles, including deep sleep and REM sleep, which are crucial for cognitive functions and overall health.

In conclusion, our analysis confirms the importance of aligning sleep with the body's circadian rhythm and the benefits of consistent, sufficient sleep. These insights could be instrumental in developing interventions for individuals seeking to improve their sleep quality. By encouraging regular sleep schedules that align with natural circadian rhythms and ensuring adequate sleep duration, we can help individuals optimize their sleep health and, consequently, their overall well-being.