## Status Update

**Data Source Name**: Data source file names are SleepData\_2021-05-30\_2024-03-01, ActivityData\_2021-05-30\_2024-03-01, ReadinessScores\_2021-05-30\_2024-03-01 [ALL files can be found in this folder.](https://drive.google.com/drive/folders/147MTlz61D0aKu6BOXW-iufuKw2vCscAL?usp=sharing)

**Related Work**: A public [GitHub repository](https://github.com/crystoll/oura-ring) shows someone analyzing their sleep data using the oura ring, specifically investigating the # meetings in a day effects on various sleep metrics. [LINK to specific code in repo](https://github.com/crystoll/oura-ring/blob/main/external_correlations.ipynb)

Other than that project we found no projects similar to ours in investigating primarily activity levels during the day effects on sleep score calculated by Oura, and other sleep metrics.

**Libraries**: NumPy, Pandas, MatPlotLib,

### Initial EDA

**Data Cleaning**

In the data cleaning section of our exploratory data analysis, we will first address any instances where data is incomplete or inaccurate. On days when the Oura Ring's battery died during the night, leading to incomplete sleep data, we will remove these entries from our analysis. Similarly, if the user forgot to wear the ring, leaving us without sleep data for that night, we will exclude those days even if daytime activity data is available. This step ensures the integrity of our study, which focuses on the relationship between daytime activities and their effects on sleep.

We will also identify and discard any data that appears to be an outlier due to glitches in the sleep tracker, such as sleep metrics that are implausible or significantly deviate from the expected range. These anomalies can skew our results and do not represent the user's typical sleep patterns.

Furthermore, we will synchronize the datasets to ensure there is a one-to-one correspondence between the days of recorded activity and the subsequent sleep scores. Each row in the activity dataset will need to align with the corresponding sleep data row to accurately assess the impact of that day's activities on the night's sleep quality. Ensuring an equal number of rows across the sleep, activity, and readiness datasets is crucial for maintaining this alignment and for the validity of our subsequent analyses.

Let's consider an example of data matching. The activities recorded on a given day, say February 28th, are expected to influence the sleep quality of the following night. Therefore, we must align the activity data from February 28th with the sleep data recorded for that night February 28th, which the data shows as the sleep data for February 29th. This precise matching is essential to ensure that our analysis accurately reflects the relationship between daytime activities and sleep outcomes. Any mismatch or discrepancy here could lead to incorrect conclusions about the effects of exercise and activity levels on sleep quality.

**Exploring Relationships**

In the subsequent phase of our analysis, we will delve into pattern and trend identification. By employing Pearson's correlation coefficient and linear regression, our goal is to discern the relationships between potential explanatory variables and response variables.

Firstly, we will focus on the explanatory variables within the activity dataset, which include Activity Score, Activity Burn, Total Burn, Steps, Inactive Time, Rest Time, Low Activity Time, Medium Activity Time, High Activity Time, Average MET, and Long Periods of Inactivity. These variables are crucial in understanding the extent to which daily physical activity impacts sleep quality.

Additionally, within the sleep dataset, we recognize certain behaviors that could significantly affect sleep, such as bedtime start, bedtime end, and sleep timing (found in Sleep Data set). These variables are indicative of a person's sleep alignment with their circadian rhythm, which is hypothesized to have a substantial impact on sleep quality.

From our analysis, it appears that earlier bedtimes and earlier wake times tend to result in higher Sleep Scores. This confirms the hypothesis that aligning sleep patterns with the body’s natural circadian rhythm can improve sleep quality. Furthermore, irregular sleep schedules can disrupt a person’s circadian rhythm and lead to poorer sleep results.

We also took a look at how sleep timing affects sleep score. From our analysis, there appears to be a moderate linear relationship between sleep timing and sleep score. In essence, higher sleep timing tends to result in higher sleep scores. This result further confirms our expectations by suggesting that longer periods of sleep results in higher sleep qualities.

The primary dependent variable we will be measuring is the [Sleep Score](https://support.ouraring.com/hc/en-us/articles/360025445574-Sleep-Score), which ranges from 1 to 100 and serves as a direct measure of sleep quality for the corresponding night. In parallel, we will also examine secondary response variables that act as proxies for sleep quality. These include sleep latency, which measures the time it takes for one to fall asleep after lying in bed, and sleep efficiency, which calculates the percentage of time in bed that was spent sleeping. How Total sleep durations effect on sleep quality is also of interest.

Moreover, we will analyze other columns in the sleep dataset that could yield insights into the recovery process and the impact of daily activities as potential response variables. These include total sleep duration, total bedtime, durations of REM, light, and deep sleep stages, average resting heart rate, and average Heart Rate Variability (HRV). We hypothesize that increased physical activity might lead to a lower HRV, as a higher HRV typically indicates a well-recovered body, and conversely, that it may result in a higher average resting heart rate as the body works to recover from the day's activities. For a comprehensive analysis, we will establish baseline measures for mean HRV and mean heart rate.

Visualizations will play a key role in our exploration, allowing us to present and interpret the relationships between the activities of the day and the sleep metrics. Through these visualizations, we will be able to better understand and communicate the relationship between exercise, activity and sleep quality.