snoozeless💤

Colbe Chang, Jocelyn Ju, Jethro Ronald Lee, Michelle Wang, Ceara Zhang

Northeastern University, Boston, MA, USA

Link to Github Repository: https://github.com/jethrorlee731/sleep_efficiency

Abstract

Snoozeless, a sleep efficiency dashboard, is a data visualization tool built using the Plotly Dash library. The dashboard is designed to help individuals track and analyze their sleep patterns to improve their sleep quality. The dashboard displays relationships between key sleep metrics such as sleep duration and sleep efficiency and also includes several charts to display how different variables impact sleep. Many charts are included, from ones as intuitive as a scatter plot and histogram to ones more profound as a density contour plot and radar graph. The user can interact with the dashboard to change the variables and values shown on the plots through dropdown menus and sliders to help them discover sleep data that interests them. From there, users can identify trends and patterns in people's sleep behavior. For this project, a user-friendly dashboard was created for users of any age to interact with and gain more information concerning factors that impact sleep and see what they can do to improve their sleep. A machine-learning model was also implemented to allow users to input factors like their bedtime, sleep duration, caffeine consumption, and exercise frequency and afterward see stats about their sleep quality. It was hypothesized that sleeping longer, drinking less caffeine and alcohol, not smoking, and exercising more would help one maximize their sleep quality. Even though these findings were supported, the dashboard presented other interesting findings such as how gender seems to play a minimal direct role in one's sleep quality and that age, awakening frequency, and alcohol consumption majorly affect sleep quality overall.

Introduction

With sleep being an essential part of human life, *Snoozeless* is a useful tool for anyone looking to improve their sleep and overall health. The purpose of this project was to create a sleep efficiency dashboard that would, by extension, help users understand the reasons behind their sleep quality and how to maximize certain parts of their sleep cycle. College students in particular, as they are typically away from home for the first extended period of time of their life, face difficulties in prioritizing sleep. This project is intended to help college students, as well as anyone looking to improve their sleep efficiency, do so.

The dashboard will focus on helping users maximize their sleep efficiency, REM sleep percentage, and deep sleep percentage, as these factors are associated with better sleep. Sleep efficiency refers to the ratio of time that one rests in bed while actually asleep [1]. REM sleep is responsible for helping people process new knowledge and execute motor skills to their fullest potential [2]. Deep sleep enables the body to release vital growth hormones that work to build muscles, tissues, and bones [3]. The working hypothesis was that focusing on different factors such as awakenings and age will help to maximize these metrics. Specifically, it was thought that sleeping longer, drinking less caffeine and alcohol, not smoking, and exercising more would help one maximize their sleep efficiency, causing one's sleep duration, alcohol and caffeine consumption habits, smoking status, and exercising habits serve the most significant roles in determining their sleep quality.

Unlike other sleep analyses, this work provides visualizations for user exploration as well as a predictor model. Our dashboard contains two parts. The first tab allows users to *explore* a variety of plots regarding factors that impact sleep efficiency, deep sleep percentage, and REM sleep percentage. The second tab allows users to *input current information about themselves* to see their predicted sleep efficiency, deep sleep percentage, and REM sleep percentage. Thus, this work urges users to engage with its content by connecting with it personally rather than only having them passively observe the sleep metrics of other people.

Methods

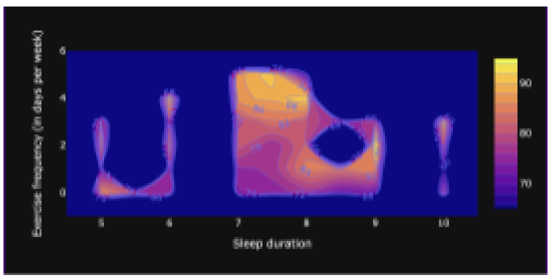
Kaggle and other websites were explored when looking for datasets that would align with the working hypothesis. This dashboard relies on a dataset found on Kaggle that contains data on a group of 452 test subjects and their sleep patterns as well as 15 distinct attributes. Attributes include a subject's ID (unique identifier), age, gender, bedtime, wakeup time, sleep duration, sleep efficiency, REM sleep percentage, deep sleep percentage, light sleep percentage, awakenings, caffeine consumption in the 24 hours before bedtime in milligrams, alcohol consumption in the 24 hours before bedtime in ounces, smoking status, and exercise frequency in days per week [1].

For pre-processing, times were parsed properly and converted into military time to make the graphs simpler to build, columns were renamed for clarification, and rows with NA values were removed. We believe removing the rows with NA values was the most simplistic as well as the best way to address them because each subject is too different in terms of age, gender, caffeine consumption habits, and other factors for imputation through a statistical measure such as the mean or median to be reasonable. It also doesn't make sense to impute the missing data with zeros as doing so could cause the user to make inaccurate conclusions regarding the trends across certain sleep metrics. There was still a substantial amount of data remaining afterward (388 rows instead of the original 452). Sleep efficiency values, which were decimals out of 1, were also multiplied by 100 so they could be represented as percentages (out of 100%).

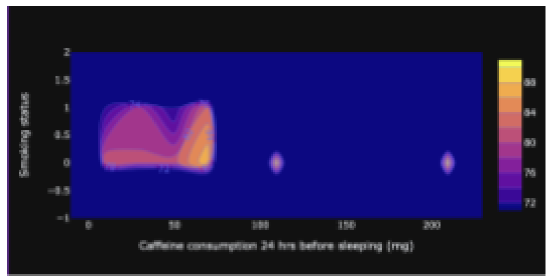
Analysis

Based on the data used, there seem to be multiple factors that impact sleep quality and different trends between sleep variables. There are many conclusions our dashboard can provide regarding the sleep quality of an individual. For conciseness, this section will only mention some, but users have a lot of opportunities with the dashboard to explore other conclusions by observing the relationships among different combinations of variables in each plot.

The density contour plot presents how two variables influenced the average sleep efficiency of the test subjects. The color map on the right marks which colors correlate with certain sleep efficiency averages. Spots on the plots that are brighter and more yellow indicate combinations of sleep variables that contribute to higher average sleep efficiencies. Hence, the left plot shows that to maximize sleep efficiency, one should get seven to eight hours of sleep and exercise for at least 4 days a week. The right one stresses that limiting caffeine consumption to 60 milligrams in the 24 hours before sleeping or less and not smoking (smoking status = 0) also helps to maximize sleep efficiency. Interestingly, it doesn't seem like having the lowest caffeine consumption amount possible maximizes one's sleep efficiency. Perhaps caffeine affects one's ability to sleep more based on when it's consumed, not how much of it's consumed.

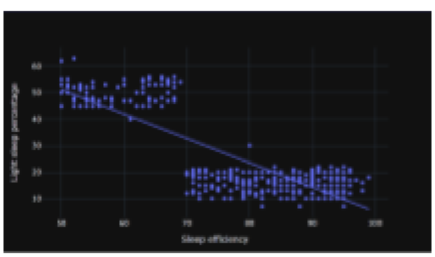


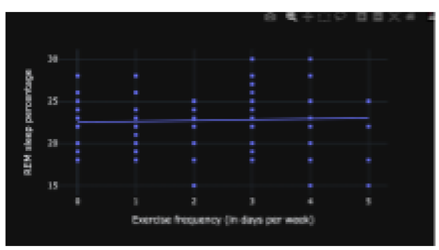
Sleeping for 7-8 hours and exercising more optimizes sleep efficiency



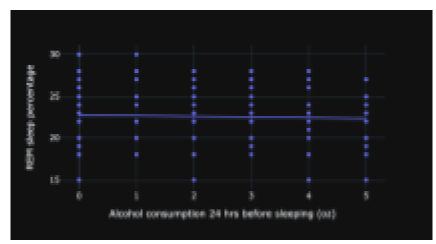
Not smoking and less caffeine improves sleep efficiency

There is an inverse relationship between light sleep percentage and sleep efficiency. There appears to be a moderate correlation between these variables as well as apparent clustering. The scatterplots also suggest that increasing the amount of exercise one has in a week and decreasing alcohol consumption in the 24 hours before sleep can raise REM sleep percentage as well.


Sleep efficiency and light sleep percentage are inversely correlated

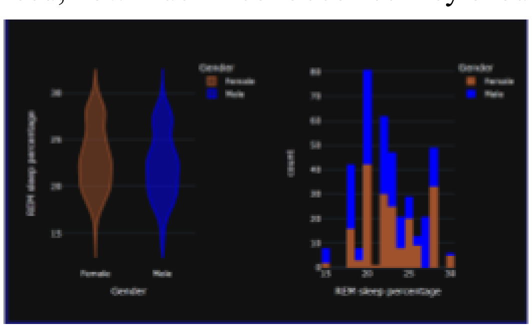


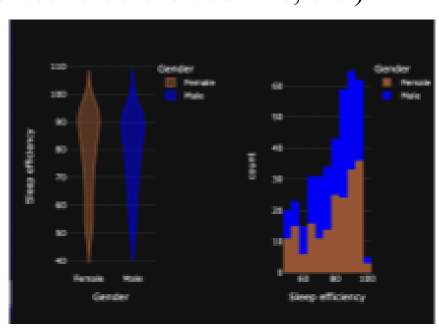
More exercise leads to more REM sleep



More alcohol leads to less REM sleep

With gender and sleep-related variables, there doesn't appear to be much variability between the sexes. For example, the images below display that females and males have a similar distribution of data proportion-wise regarding REM sleep percentages and sleep efficiencies. These graphs suggest that gender alone is not critical in determining one's sleep quality. However, gender can indirectly influence other factors that influence one's sleep quality (e.g., how much sleep they need, how much alcohol/coffee they should limit themselves to before bedtime, etc.)

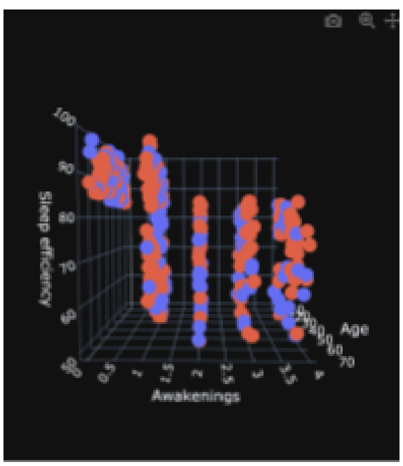

REM sleep percentages between males vs. females are similar proportion-wise


Sleep efficiency between males vs. females are similar proportion-wise

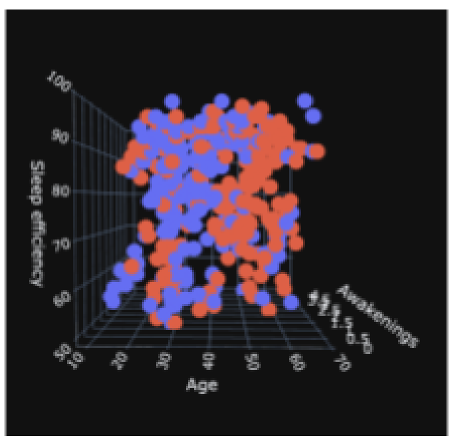
Overall, the strip chart shows that non-smokers seem to sleep better than smokers. The points for the non-smokers are skewed toward the right, showing they tend to have high sleep efficiencies.

Not smoking leads to better sleep

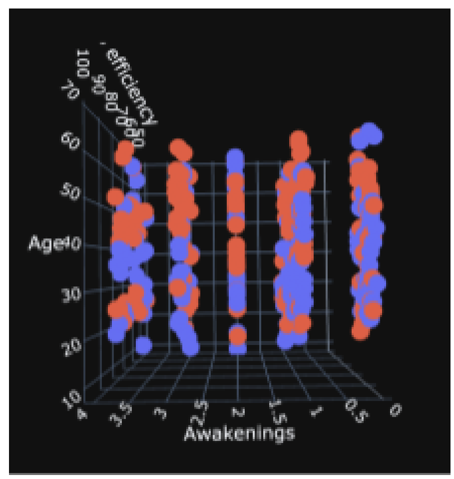
The 3D scatterplot shows the relationship between three sleep factors rather than two. Blue points represent biological females and red points represent biological males. These three plots show that an increase in awakenings correlates with a lower sleep efficiency for both genders. Furthermore, there is a weak correlation between awakenings and age as well as between age and sleep efficiency for both genders.



Awakenings and sleep efficiency seem to be inversely correlated



No clear relationship between age and sleep efficiency

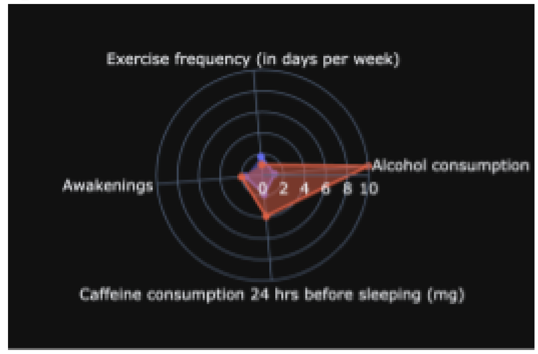


No clear relationship between awakenings and age

The radar graph allows a user to visually compare their sleep statistics with that of the average person in the UK study. The left graph indicates someone who typically wakes up 2 times during their sleep, drinks 50 milligrams of caffeine in the 24 hours before sleeping, consumes 1 ounce of alcohol in the 24 hours before sleeping, and exercises just once per week. The red plane, which represents the user, closely aligns with the blue plane, which represents the average test subject, indicating that the user's habits generally align with the average participant in the study. Conversely, the right graph represents someone who drinks an abnormal amount of alcohol in the

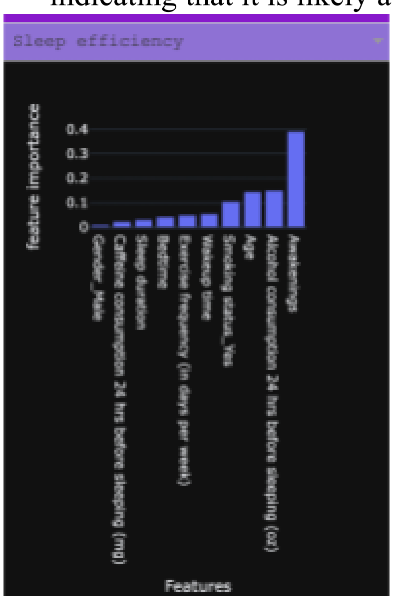
24 hours before bedtime (10 ounces) compared to the average subject, as suggested by how the user's red diamond skews toward alcohol consumption a lot more than the blue one.

A user with habits comparable to the average person in the UK sleep study

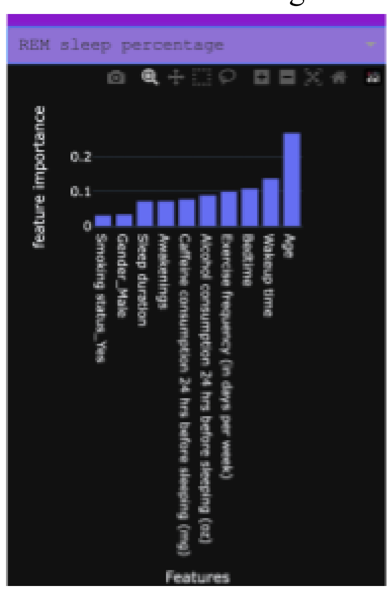


A user who drinks an abnormal amount of alcohol before bedtime compared to a typical person in the UK sleep study

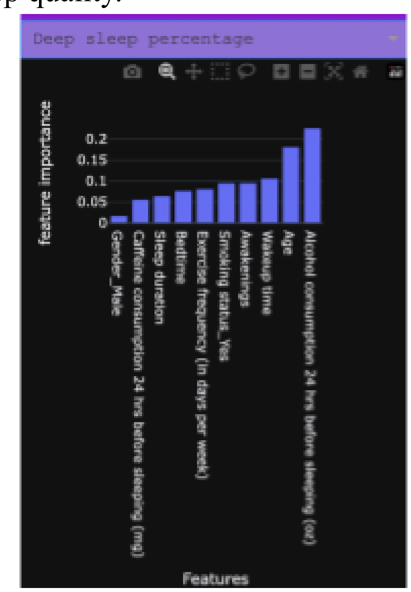
The feature importance graphs present the degrees to which a random forest regressor finds a factor critical for predicting sleep efficiency, REM sleep percentage, or deep sleep percentage. The number of awakenings, age, and alcohol consumption in the 24 hours before bedtime seem to appear in the top 3 features for predicting sleep efficiency, REM sleep percentage, and/or deep sleep percentage. These results suggest that users should focus more on these features when trying to maximize their sleep metrics. In particular, age shows up in the top 3 for all three, indicating that it is likely a critical factor in determining one's sleep quality.



Feature importance graph for predicting sleep efficiency

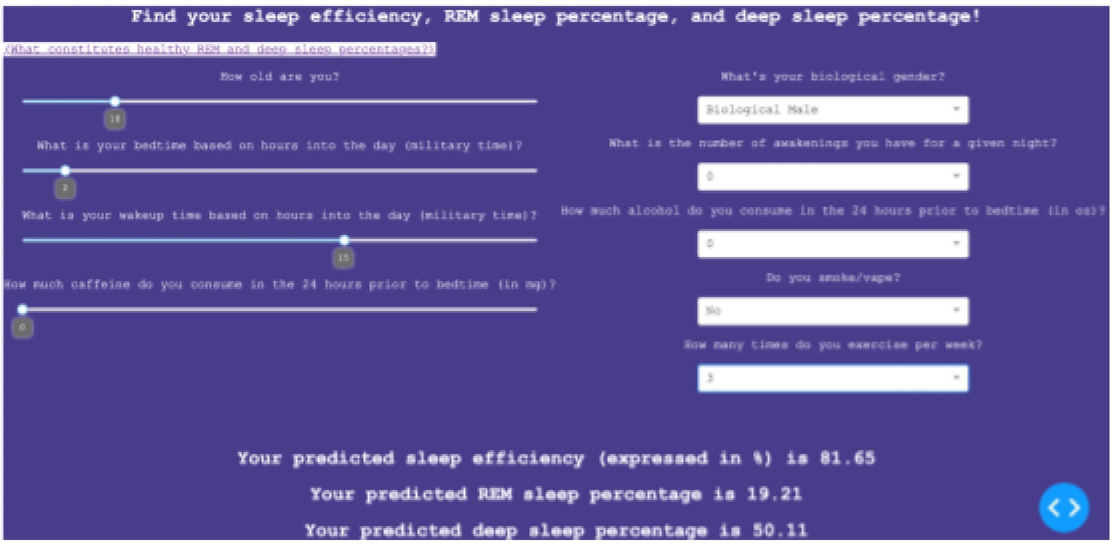


Feature importance graph for predicting REM sleep percentage

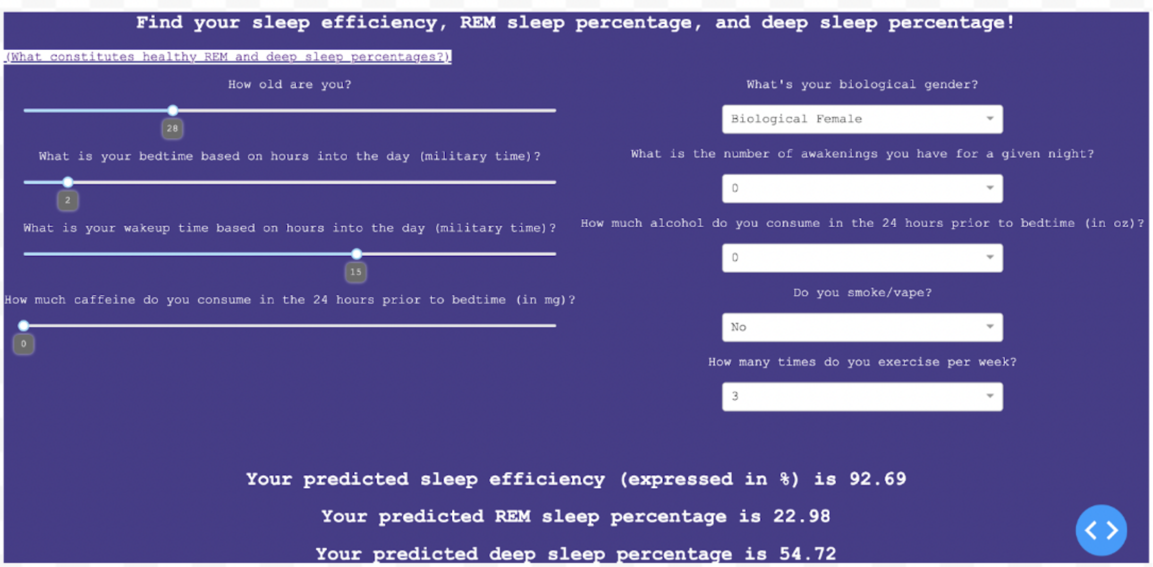


Feature importance graph for predicting deep sleep percentage

The random forest regressor allows one to see their predicted sleep efficiency, REM sleep percentage, and deep sleep percentage based on key factors about them. Factors typically associated with higher predicted sleep efficiencies, REM sleep percentages, and deep sleep percentages include getting around 8-10 hours of sleep, having fewer awakenings, drinking less alcohol in the 24 hours before sleeping, exercising more in a week, and not smoking. Caffeine's effect on sleep quality is a little unclear. There are many reasons why caffeine may have had a small effect on people's sleep quality, such as maybe caffeine affects sleep quality more based on when it's consumed and not how much of it's consumed The values of the metrics that correlate with certain sleep efficiencies, REM sleep percentages, and deep sleep percentages somewhat depend on one's biological gender and age. The top image presents an 18-year-old biological male who has a slightly lower sleep efficiency, REM sleep percentage, and deep sleep percentage compared to a 28-year-old biological female who exhibits many of the same behaviors as the biological male. Also, sleep duration (the difference between wake-up time and bedtime) interestingly does not seem to have a clear effect on sleep quality, emphasizing the need to get good deep and REM sleep rather than just time in bed with closed eyes.



A theoretical user who inputs values for the sleep quality predictor



Another theoretical user who inputs the same values as the previous user, except those for age and biological gender

**Note: We built two machine learning models, a random forest regressor and a multiple linear regression model. We ended up only using the random forest regressor as it yielded higher R^2 scores for predicting sleep efficiency, REM sleep percentage, and deep sleep percentage. You can find two individual .py files (sleep_forest.py and sleep_mult_reg.py) that present how the models resulted in those R^2 scores. The final random forest regressor we used for the sleep quality predictor model is directly implemented in random_forest_assets.py, sleep.py, and utils.py.

** Note: We also include a help tab on our dashboard with helpful informational dropdowns and videos to aid users if they are confused about how to use our dashboard.

Conclusions

In conclusion, *Snoozeless* provides valuable insights into an individual's sleep patterns and overall sleep health. By tracking important metrics such as sleep efficiency, REM sleep percentage, and deep sleep percentage, users can gain a deeper understanding of their sleep habits to make adjustments to improve their sleep quality and overall health.

The visualizations on the dashboard make it easy to see patterns and trends in sleep data over time. Age appears to be within the top three most important variables impacting sleep efficiency, REM sleep percentage, and deep sleep percentage. Additionally, other variables that appear in the top five for the sleep metrics include awakenings, age, and alcohol consumption. While age and awakenings are factors out of a user's direct control, the dashboard supports that they can still attain the best sleep routine by waking up early, exercising often, and refraining from excessive alcohol, caffeine, and smoking.

As with all statistical analyses, a potential limitation of this sleep dashboard could be user bias. For example, users who are particularly interested in improving their sleep efficiency may be more likely to have had their data collected for this purpose, which could skew the data and insights in certain directions. Furthermore, this data comes from one study that may not be representative of everybody's sleeping patterns in the real world. No information about the race of the participants is known either. The data could potentially be biased to represent only certain races, limiting the extent to which the dashboard's results are extensible to broad audiences. To address other limitations, this dashboard could potentially explore other large life factors that impact sleep, such as stress levels and screen time. It's also important to note that the predictions of the random forest regressor should be taken with caution. While the random forest regressor predicts sleep efficiency with a fairly high cross-validated R^2 score of ~0.67, it only predicts REM sleep percentage with a cross-validated R^2 score of ~0.16 and deep sleep percentage with a cross-validated R^2 score of ~0.35. We believe our models are valuable in predicting these complex metrics but we still want users to use the models with caution in terms of the conclusions that are drawn.

Overall, the sleep efficiency dashboard is a good starting step for those looking to improve their sleep health and gain insights into their sleep patterns. By using it to track and monitor sleep data, users can take proactive steps to improve their sleep quality and overall well-being.

Author Contributions

Below are all of each author's main contributions to this project.

Colbe Chang: radar graph, sliders, proofreading, debugging, commenting, report

Jocelyn Ju: Helper functions, dropdowns, scatterplots, help tab and demo videos, report, commenting, slides for in-class science fair

Jethro Ronald Lee: Gender plots, strip chart, density contour plot, sliders + dropdowns, random forest regressor, dashboard layout (e.g., colors and tabs), report, commenting

Michelle Wang: random forest regressor, multiple linear regression, feature importance plot, 3d scatter, 2d scatter, helped with making the framework more reusable and more functions, report

Ceara Zhang: Modified/debugged visualizations (scatterplots, strip plot, contour), optimized utility functions, report, and conclusions

References

1. Equilibriumm, (2023, Feb. 21). Sleep Efficiency Dataset [Dataset]. Kaggle. Retrieved from https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency

2. Summer, J. (2023, March 2). *What is REM Sleep and How Much Do You Need?* Sleep Foundation. Retrieved April 11, 2023, from https://www.sleepfoundation.org/stages-of-sleep/rem-sleep

3. Pacheco, D. (2023, March 22). *Deep Sleep: How Much Do You Need?* Sleep Foundation. Retrieved April 11, 2023, from https://www.sleepfoundation.org/stages-of-sleep/deep-sleep