

ECS 171 Group 10 Project

Leader: Ryan Yu

Amira Basyouni, Calvin Chen, Alexis Lydon, Tianming Tan

October 16, 2023

Problem Statement

As a student, we are in a battle between the many stressors we have in our lives and our strive towards academic success. We are faced with deadlines, exams, and the pressure to perform well. These factors can have a negative impact on our quality of learning and thus hinder our progress towards acquiring a degree and entering the workforce. Our team would like to alleviate this problem using sleep. We would like to predict the effects poor sleep quality can have on our stress levels. Would longer sleep duration result in a decrease in stress? After we analyze our data, we will be able to confirm the correlation between those two attributes. We will also look into other factors affecting stress such as physical activity, occupation, and blood pressure. By finding various correlations, we can narrow down the factors affecting stress and ultimately propose a healthy set of habits to help manage our mental health. The data will involve ages that vary from 27 to 59 years of age. Although it does not cover the typical age of a college student, our analysis can still help us build awareness that would last us for years to come.

Dataset

The dataset being used in this project is the "Sleep Health and Lifestyle Dataset" by Laksika Tharmalingam on Kaggle. The dataset comes in the form of a CSV file and contains 400 rows and 13 columns, encompassing various sleep and lifestyle variables. These include gender, age, occupation, sleep duration, sleep quality, physical activity, stress levels, BMI category, blood pressure, heart rate, daily steps, and sleep disorder status. We chose this dataset because, as students, we can relate to the common sleep issues many of us face. Our project focuses on investigating the relationship between sleep and health, making a dataset on sleep and health the most suitable choice. A limitation of the dataset is its synthetic nature, generated artificially rather than from real-world observations. While it may lack some real-world nuances, high-quality synthetic data can effectively train and test machine learning models in this context. Here is the link of our dataset for further reference: [Sleep Health and Lifestyle Dataset](#)

Goals

In this machine learning project centered on the intersection of sleep health and stress, we aim to achieve several key objectives to better understand and predict the dynamics between these two factors. First, we will dive into the relationship between sleep duration and stress. Our investigation will involve analyzing how varying durations of sleep impacts an individual's stress levels. Next, we will explore the influence of sleep quality on stress levels. As an integral part of the project, we will develop an accurate regression model for stress prediction based on factors of sleep quality. We will also develop models that predict stress levels based on other highly correlated attributes in our dataset other than just sleep. Our last goal is somewhat of a reach goal, and it involves investigating the amount of extra sleep required for stress reduction. These defined project goals will be a structured guide throughout our research, data collection, analysis, and model development. They will enable us to gain valuable insights into the complex relationship between sleep health and stress levels while delivering solutions and predictions for this area of health.

Project Timeline

For our project, we plan to roughly follow the timeline below:

Project Timeline			
Task	Start Date	End Date	Duration (days)
Background/Literature Review	Oct. 16	Oct. 22	7
Exploratory Data Analysis	Oct. 23	Oct. 29	7
Developing prediction models	Oct. 30	Nov. 5-8	7-10
Evaluation of the model and testing performance	Nov. 6-9	Nov. 12-14	7-9
Developing front-end to display and run models	Nov. 13-15	Nov. 19-21	7
Finish Report and Minor Adjustments	Nov. 20-22	Nov. 23-25	4
Presentation Practice	Nov. 24-26	Nov. 28-30	5

Table 1: Tasks, Start and End Dates, and Planned Duration

Task Breakdown

1. Background/Literature Review: Research our problem and describe it scientifically
2. Exploratory Data Analysis: Understanding our dataset and filtering out noise
3. Developing prediction models: Creating 3+ models to quantify and predict desired outcomes
4. Evaluation of the model and testing performance: Testing our models and deciding on the best performing one
5. Developing front-end to display and run models: Invoke our models and display results
6. Finish Report and Minor Adjustments: Clean up our code and final report paper
7. Presentation Practice: Practice our presentation and prepare for questions