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. This dataset is structured as a CSV file and contains 400 rows and 13 columns, featuring a wide array of sleep and lifestyle variables. These variables encompass information such as gender, age, occupation, sleep duration, sleep quality, physical activity, stress levels, BMI category, blood pressure, heart rate, daily steps, and sleep disorder status. We selected this dataset because, as students, we can readily relate to the common sleep-related challenges that many of us encounter. Our project places a particular emphasis on exploring the relationship between sleep and stress, rendering a dataset that focuses on sleep and health the most apt 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, our primary objective is to examine the correlation between key attributes such as sleep duration, sleep quality, BMI, blood pressure, and occupation, and their influence on stress levels. This analysis will help us better understand and predict the dynamics between these factors and identify the attributes that have the most significant impact on stress levels. As an integral part of the project, we will develop an accurate regression model for stress prediction based on the highly correlated attributes in our dataset. Our last goal, which is somewhat a reach goal due to its complexity, is to determine the extent to which additional sleep reduces an individual's stress level and the optimal duration for stress reduction. These defined project goals will serve as a structured guide throughout our research, data collection, analysis, and model development, enabling us to gain valuable insights into the complex relationship between sleep health and stress levels while delivering practical 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: Researching our problem and finding related work
2. Exploratory Data Analysis: Understanding our dataset in relation to findings 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 selecting the best predictor
5. Developing front-end to display and run models: Create a user interface and invoke our models
6. Finish Report and Minor Adjustments: Clean up our code and final report paper
7. Presentation Practice: Practice our presentation and prepare for questions