

## Project Proposal Group 23

**Dataset:** <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>

```
df = pd.read_csv("Sleep_health_and_lifestyle_dataset.csv", index_col=0)
print(df.head(5))
print(df.describe())
```

### **Description of Data:**

The Sleep Health and Lifestyle Dataset contains 400 rows and 13 columns with variables of sleep and lifestyle habits. These variables include gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders.

### **Questions:**

- **What lifestyle and health factors are the best indicators of a sleep disorder? Goal: Identify conditions under which a person should be screened for sleep disorders.**
- **What lifestyle and health factors translate to better quality of sleep?**

### **Variables:**

- Person ID (int): An identifier for each individual
- Gender (male/female): gender of the person
- Age (years): age of the person
- Occupation (categorical): occupation or profession of the person
- Sleep duration (hours): number of hours person sleeps per day
- Quality of sleep (1-10): subjective rating of quality of sleep
- Physical activity level (min/day): number of minutes the person engages in physical activity daily
- Stress level (1-10): subjective rating of stress level experienced by the person
- BMI category (categorical): BMI category of the person
- Blood pressure (systolic/diastolic): blood pressure measurement of the person
- Heart rate (bpm): resting heart rate of person in bpm
- Daily steps (# of steps): the number of steps the person takes per day
- Sleep disorder (type of sleep disorder): presence or absence of sleep disorder

### **Methods:**

#### **Question 1:**

- **We will create a decision tree classifier used with a maximum depth to understand which variables (sleep duration, BMI category, blood pressure, heart rate, etc.) are the most predictive of whether an individual has a sleep disorder.**
- **We intend to use logistic regression to understand a probability threshold for recommendation of screening for sleep disorders.**

#### **Question 2:**

- **We will use linear regression to model the influence of various lifestyle/health factors (physical activity level, occupation, BMI category, etc.) on sleep quality and sleep duration to understand the relationship between these variables.**