**Alexandra Perez**

**Deepak Neelam**

**Nathan Tatko**

**Multivariate Analysis: Project Proposal**

**Project Goal**

The project aims to use the selected dataset below to explore the relationship between health data and the presence of a sleep disorder.

Hypothesis

1. Increased activity levels would lead to a decrease in the probability of having a sleep disorder.
2. Increase in stress levels would lead to an increase in the probability of having a sleep disorder.
3. As the participants get older, they have a higher probability of having a sleep disorder.

**Data Source**

The Sleep Health and Lifestyle Dataset, sourced from Kaggle and curated by Lakisha Tharmalingam, provides a comprehensive collection of data on 400 individuals. This dataset includes a range of variables, such as demographic information, sleep metrics, lifestyle factors, health metrics, and sleep disorder diagnoses. By analyzing this dataset, we can explore the intricate relationship between sleep patterns, lifestyle habits, and overall health.  
[Dataset Link](https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset)

**Proposed Analysis Method(s)**  
a. PCA (How do age, physical activity level, stress level, heart rate, and daily step count influence sleep quality?)

b. Data Reduction-Based on the correlation matrix, we can select variables.

c. Classification

d. Compare the means to see if there are significant differences between groups

**Potential Conclusions**

Through the analysis, we will conclude if a person’s stress levels, age, physical activity level, heart rate, and daily step count influence the presence of a sleep disorder.