### **4.1 Dataset Description**

The dataset that we used in this project is the Sleep Efficiency dataset, obtained from [Kaggle](https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency) **( apa citaat van maken?)** . The dataset description states that it contains real-life sleep tracking data from individuals. The goal of the dataset is to explore the relationship between some of the variables and sleep efficiency, which is defined as the proportion of time in bed actually spent sleeping as stated in the description of the dataset.

#### **Source**

The dataset was originally uploaded by the user “equilibriumm” on Kaggle. It is publicly accessible and intended for educational and research purposes. The actual source of the data is unknown, so we do not really know if the data is collected naturally or if this dataset was fabricated.

#### **Dataset Structure**

The dataset consists of 452 observations (rows) and 15 variables (columns) **(INCLUDE CODE IN DOCUMENT FOR THESE CLAIMS?)** Each row represents an individual and their corresponding sleep and lifestyle data for a single day. The variables are structured as follows:

**(Add in discrete continuous etc?)**

| Variable | Type | Description |
| --- | --- | --- |
| ID | Categorical | Unique identifier for each subject |
| Age | Numerical | Age of the individual in years |
| Gender | Categorical | Gender of the individual: "Male" or "Female" |
| Bedtime | Time | Time the person went to bed |
| Wakeup time | Time | Time the person woke up |
| Sleep duration | Numerical | Total number of hours spent sleeping |
| Sleep efficiency | Numerical | Percentage of time in bed actually spent sleeping (target) |
| REM sleep percentage | Numerical | Proportion of sleep spent in REM stage |
| Deep sleep percentage | Numerical | Proportion of sleep spent in deep sleep stage |
| Light sleep percentage | Numerical | Proportion of sleep spent in light sleep stage |
| Awakenings | Numerical | Number of times the individual woke up during the night |
| Caffeine consumption | Numerical | Number of caffeine units consumed per day |
| Alcohol consumption | Numerical | Number of alcohol units consumed per day |
| Smoking status | Categorical | Indicates if the person smokes: "Yes" or "No" |
| Exercise frequency | Numerical | Number of days per week with physical activity |