Rianna McDuffus
Na'jae Batts
Alana Manuel
Sijan Shrestha

Short Report

Once learning about the group project that we were assigned, we searched on Kaggle for datasets and decided as a group on Sleep and Health Statistics. This was important to us because we all thought about the importance of sleep as college students and the type of sleep schedules and patterns we have. This dataset helps us to better understand the correlation of how other factors (age, dietary habits, exercise, medication, etc.) interact with sleep.

We were able to load the data after importing numpy and pandas using the read command. Once uploaded within the notebook we could familiarize ourselves with the data information discovering the dtypes within the data frame and displaying some rows.

Now familiar with the data frame, we selected 2 aspects, both having integer data types. Selecting "Age" helps us better understand the data pool, and "Calories Burned" helps us better understand the energy consumption. Utilizing the numpy operations we found:
- Median age=35
- Mean Age=36
- Standard Deviation of Age= 8
- Mean Calories Burned=2200
- Median Calories Burned=2421
- Standard Deviation of Calories Burned= 280

Though the aspects of the data frame included the "Bedtime" and "Wake-up Time" , there was no specific column for hours of sleep. To find this information we had to convert the object dtypes of "Bedtime" and "Wake-up Time" into datetime dtypes. Once converted we were able to subtract the "Wake-up Time" from the "Bedtime" to find the total hours slept. We then utilized numpy operations and discovered:
- Mean sleep = 7 hours

This information would have been helpful in its own column of the original dataset in addition to "Bedtime" and "Wake-up Time".

During data cleaning using Pandas, there weren't any missing null values or duplicate values in our dataset which was good and everything as far as collecting data was good quality. We do believe that adding more demographic factors such as their socioeconomic status would be important because that can affect their sleep and also their health. Expanding the health metrics can also be an improvement that the data can have.

The data filtering and selection, coupled with the sorting and ranking allowed for a deeper understanding of trends and patterns in the health and sleep behaviors of different age groups. One trend that stood out was the fact that younger individuals had a higher quality of sleep when compared to the older individuals in the study. It was also evidenced that having "healthy" dietary habits contributed to a higher quality of sleep, along with burning more qualities and taking more daily steps. The data would give a better representation, however, if it accounted for a wider range of ages, i.e children and the elderly.

In the process of grouping and aggregating we grouped the data on:
- Gender,
- Age Range,
- Dietary Habits,
- Physical Activity Level,
- Medication Usage, and
- Sleep Disorder

For instance, our analysis reveals significant connections between **sleep quality** and other factors such as **physical activity**, **calorie expenditure**, and **age**, highlighting the following:

1. Gender and Sleep Quality

- Females report better sleep quality, with an average score of 8.42 compared to 5.58 for males. This is linked to their higher physical activity levels (447,000 total steps vs. 236,000 for males) and calories burned (132,700 vs. 109,400 for males).
- Implication: Increased physical activity and energy expenditure may positively influence sleep quality in females. The variability in males' sleep and physical activity suggests a need for more targeted interventions to stabilize both.

2. Age and Sleep Quality

- Younger individuals (19-30) report the highest sleep quality (average 8.7), with those aged 41-50 reporting significantly lower scores (5.35). Younger individuals also take more steps daily (9,333 steps) and burn more calories (2,711 calories on average).
- Implication: Higher levels of physical activity and energy expenditure in younger age groups could contribute to better sleep quality. Encouraging older individuals to increase their activity may help improve their sleep patterns.

3. Physical Activity and Sleep Quality

- Individuals with high physical activity report better sleep quality (8.55) compared to those with low activity levels (4.61). High activity participants also burn more calories (2,686 daily vs. 2,057 for low activity individuals).

- **Implication**: There is a clear link between physical activity and sleep quality. Those with higher physical activity not only sleep better but also have more consistent physical health indicators like higher calorie burn.

In conclusion, the dataset that we chose overall had good quality and was interesting to look at. However, the dataset's small size may raise concern about the statistical significance when attempting to generalize findings. The absence of geographical data also leaves out the potential environmental influences on health and sleep. Furthermore incorporating mental health data, including stress levels and mood, could enhance the understanding of sleep disorder and quality. A large dataset would help improve the analysis in the real world. The categorical variables like Physical Activity Level and Dietary Habits can also be broken down into granular categories to facilitate more detailed quantitative analysis. The process of exporting and analyzing the health sleep statistics dataset has provided a structured approach to uncover key insights into sleep patterns across different demographic and behavioral segments.