Humans spend around 1/3 of there lives a sleep: duration and quality of sleep is important to maintain optimal cognitive, psychological, and physical functioning. There has been research that has suggested increased mobile phone usage can result in sleep disturbances (Liu et al., 2019). One reason is excessive blue light can disrupt the brain’s ability to produce melatonin normally and this can interfere with sleep. Physical exercise has also been shown to improve sleep duration and quality (Wang & Boros, 2021). Caffeine worldwide is the most common psychoactive substance consumed and over 80% of people from the United States consume caffeine daily (Mitchell, Knight, Hockenberry, Teplanksy & Hartman, 2014). Although caffeine is an antagonist of adenosine receptors meaning it can dysregulate the sleep-wake cycle. There has been research that has shown caffeine can cause a reduction in sleep duration and quality (Clark & Landolt, 2017). The aim of the present study was to explore if a set of variables such as sleep direction, blue light exposure, exercise, screen time and caffeine could be used to predict the number of hours slept.

**Method**

**Dataset**

The sleep dataset was downloaded from Kaggle consisting of 46 samples. The mean age was 24.76 and there were 26 males, 18 females and 2 participants preferred not to reveal gender (Gajjar, 2021). There were no missing values while Python 3 was used to run the analysis and the main packages used were Pandas, NumPy and Sklearn. The features used from the dataset were meals per day, screen time per day, blue light filter, sleep direction, exercise, and beverage before sleep. The target was sleep duration.

**Features:**

**Meals per day**

The feature originally consisted of 6 unique values: the categories were one, two , three, four, five and more than five meals per day. Majority of values were either two or three meals. Considering four of the classes had little samples: the feature was transformed to a binary variable with one category consisting of 1-2 meals per day (19 samples) and the other category consisting of 3 or more meals per day (27 samples). This helped with class balancing.

**Screen time**

The feature originally consisted of 7 unique values measured in hours: the categories were 0-1, 1-2, 2, 2-3, 3-4, 4-5, and more than 5. The feature was transformed to a binary variable to help with class balancing considering some categories had little samples. The transformed binary feature consisted of 0-3 hours (17 samples) and the other category consisted of 3 or more hours (29 samples).

**Blue light filter, sleep direction and exercise**

Blue light filter was used as is and was not transformed. It was a binary variable with 25 samples using a blue light filter and 21 samples not using a blue light filter. Sleep direction was also used as is. 16 samples slept in the West direction, 15 slept in the east direction, 9 slept in the North direction while 6 slept in the south direction. Original exercise feature was also used as is: 23 participants exercised sometimes, 12 exercised regularly while 11 participants said they did not exercise.

**Beverage**

The feature originally consisted of 4 unique values: tea, none of the above, tea and coffee, and coffee. To simplify the feature and improve class balancing, tea and coffee, and coffee were merged into one category “coffee”.

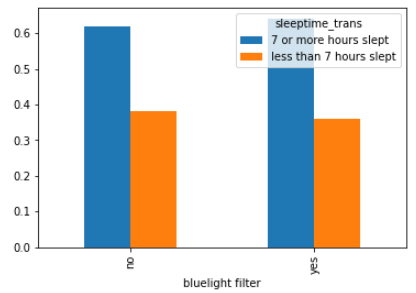
**Target transformation**

The original feature consisted of 10 unique values while majority of samples fell within 7-8 hours sleep. Sleep duration in nature is a continuous variable but the feature had been discretised. Due to this and many categories having little samples, the variable was transformed into a binary variable so a classification algorithm could be implemented. One category was 7 or more hours slept (29 samples) and the other category was less than 7 hours slept (17 samples).

**Logistic regression**

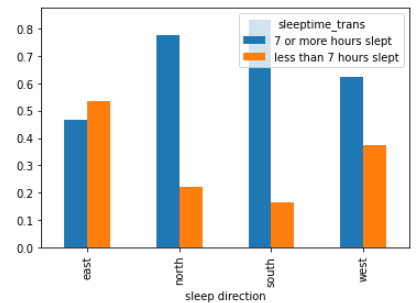
6 features were used in the logistic regression model. Sklearn Logistic Regression class and train\_test\_split function was used to run the analysis. There were 32 samples in the training set and 14 samples in the test set. The logistic regression fit function was used to fit the training data and the predict function was used to evaluate model performance on the test set. A second logistic regression model was used, and the exact same steps were used except blue light filter and meals per day were removed.

**Results**



*Figure 1*. Bar chart depicting blue filter on the x-axis and hours slept on the

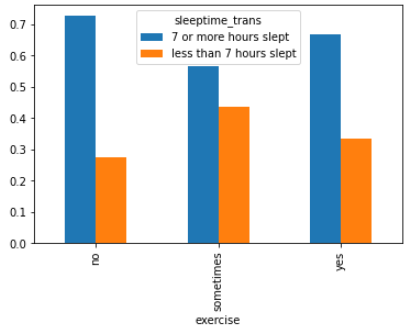
y-axis

 As can be seen from figure 1, using a blue light filter did not increase the number of hours slept. In other words, there is no indication to suggest blue light filters can increase sleep duration.

*Figure 2*. Bar chart depicting sleep direction on the x-axis and sleep

duration on the y-axis

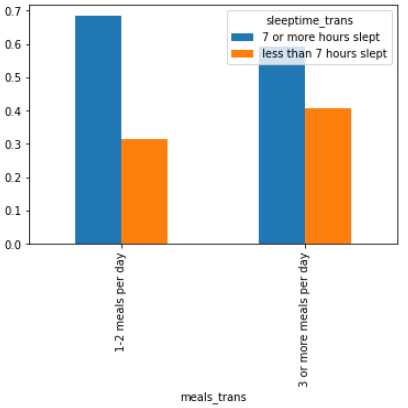
As can be seen from figure 2, people who slept in the north and south direction tended to sleep longer than people in the east and west direction.



*Figure 3*. bar chart depicting exercise on the x-axis and sleep duration on

the y-axis.

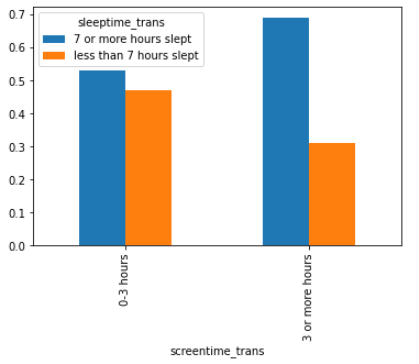
As can be seen from figure 3, the category with the most significant differences is people who didn’t exercise. People who didn’t exercise tended to sleep the longest.



*Figure 4*. bar chart depicting number of meals eaten per day on the x-axis

and sleep duration on the y-axis.

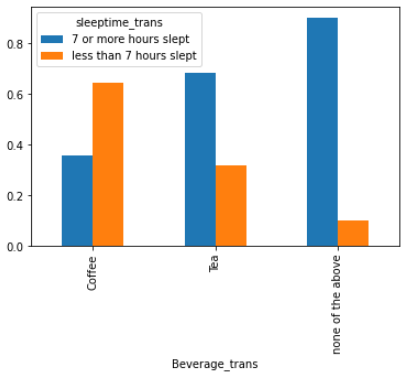
As can be seen from figure 4, There is no noticeable differences between eating 1-2 meals and eating 3 or more meals per day on sleep duration.



*Figure 5*. Bar chart depicting screen time in hours on the x-axis and sleep

duration on the y-axis.

As can be seen from figure 5, people who had 3 or more hours of screen time per day slept longer than people who had 0-3 hours of screen time.



*Figure 6*. Bar chart depicting beverage before bedtime on the x-axis and

sleep duration on the y-axis.

As can be seen from figure 6, people who didn’t have any caffeine (not drinking coffee or tea) slept the longest. People who had coffee slept the least.

*Table 1*. Comparison of two logistic regression models on test set accuracy

|  |  |
| --- | --- |
| Model | Classification accuracy |
| Logistic regression (6 features) | 71 % |
| Logistic regression (4 features) | 71 % |

Table 1 shows both models performed equally scoring 71 % on classification accuracy on the test set.

**Discussion**

The aim of the current research was to investigate whether a set of features such as blue light exposure, sleep direction, screen time, exercise and caffeine could collectively be used to classify sleep duration. Most participants were in there 20’s and guidelines suggest young adults need at least 7 hours sleep to function adequately. Anything less than 7 hours sleep is typically associated with negative outcomes such as psychological problems (attention, memory) and biological problems (fatigue). On this note, logistic regression algorithm was able to classify with reasonable accuracy whether an individual was sleeping 7 or more hours or sleeping less than 7 hours. Therefore, blue light exposure, sleep direction, screen time and caffeine were important in explaining sleep duration. There were various limitations in the research. Firstly, the sample size was small and many of the features and target had class imbalances. Therefore, categories were combined to reduce the imbalance issues. The limitation with this is information would have been loss. It is therefore important for future research to gather equal number of samples per class. The main implication in this research is sleep direction, screen time, exercise, caffeine, and blue light exposure can be useful at predicting whether someone is getting enough sleep.

References

Clark, I., & Landolt, H. P. (2017). Coffee, caffeine, and sleep: A systematic review of

epidemiological studies and randomized controlled trials. *Sleep medicine reviews*, *31*, 70-78.

Gajjar, K. (2021). Sleep Pattern. Retrieved from [https://www.kaggle.com/krupa1999/sleep- pattern](https://www.kaggle.com/krupa1999/sleep-%20%20%20%20%20%20pattern)

Liu, S., Wing, Y. K., Hao, Y., Li, W., Zhang, J., & Zhang, B. (2019). The associations of

long-time mobile phone use with sleep disturbances and mental distress in technical college students: a prospective cohort study. *Sleep*, *42*(2), zsy213.

Mitchell, D. C., Knight, C. A., Hockenberry, J., Teplansky, R., & Hartman, T. J. (2014).

Beverage caffeine intakes in the US. *Food and Chemical Toxicology*, *63*, 136-142.

Wang, F., & Boros, S. (2021). The effect of physical activity on sleep quality: a systematic

review. *European Journal of Physiotherapy*, *23*(1), 11-18.