IBM Machine Learning
Analysis on Sleep Efficiency

Yu-Hsuan HSIEH
Mar 2023

Summary

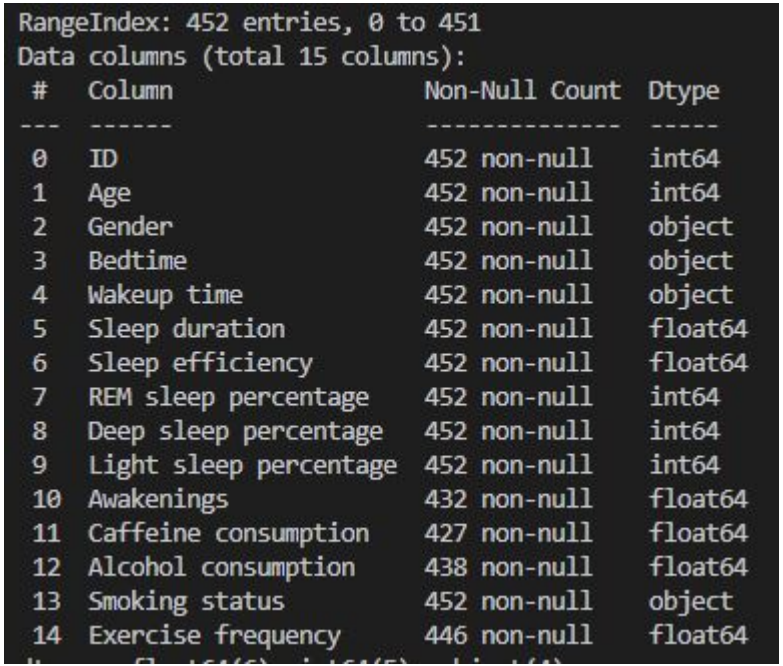
This is a presentation for the course project for the IBM Machine Learning Course

This presentation is a study of sleep efficiency, and sleep patterns using the Sleep Efficiency Dataset from Kaggle.

In this study, we'll first have an overview of the dataset, then evaluate the data integrity, data cleaning will be performed.

Dataset Overview

The dataset contains information about a group of test subjects and their sleep patterns, with 15 columns and 452 rows of data.



```
RangeIndex: 452 entries, 0 to 451
Data columns (total 15 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   ID                      452 non-null    int64
 1   Age                     452 non-null    int64
 2   Gender                  452 non-null    object
 3   Bedtime                 452 non-null    object
 4   Wakeup time             452 non-null    object
 5   Sleep duration          452 non-null    float64
 6   Sleep efficiency        452 non-null    float64
 7   REM sleep percentage    452 non-null    int64
 8   Deep sleep percentage   452 non-null    int64
 9   Light sleep percentage  452 non-null    int64
 10  Awakenings              432 non-null    float64
 11  Caffeine consumption    427 non-null    float64
 12  Alcohol consumption     438 non-null    float64
 13  Smoking status          452 non-null    object
 14  Exercise frequency      446 non-null    float64
```

Data Cleaning - Null Values & Duplicated Value

1. Check for columns with null values
2. Columns are: Awakenings, Caffeine consumption, Alcohol consumption, and Exercise frequency
3. Replace null values with column median.
4. Check for duplicated rows.
    a. Not found.

ID	0
Age	0
Gender	0
Bedtime	0
Wakeup time	0
Sleep duration	0
Sleep efficiency	0
REM sleep percentage	0
Deep sleep percentage	0
Light sleep percentage	0
Awakenings	20
Caffeine consumption	25
Alcohol consumption	14
Smoking status	0
Exercise frequency	6

ID	0
Age	0
Gender	0
Bedtime	0
Wakeup time	0
Sleep duration	0
Sleep efficiency	0
REM sleep percentage	0
Deep sleep percentage	0
Light sleep percentage	0
Awakenings	0
Caffeine consumption	0
Alcohol consumption	0
Smoking status	0
Exercise frequency	0

Data Cleaning - Data Type Convertion

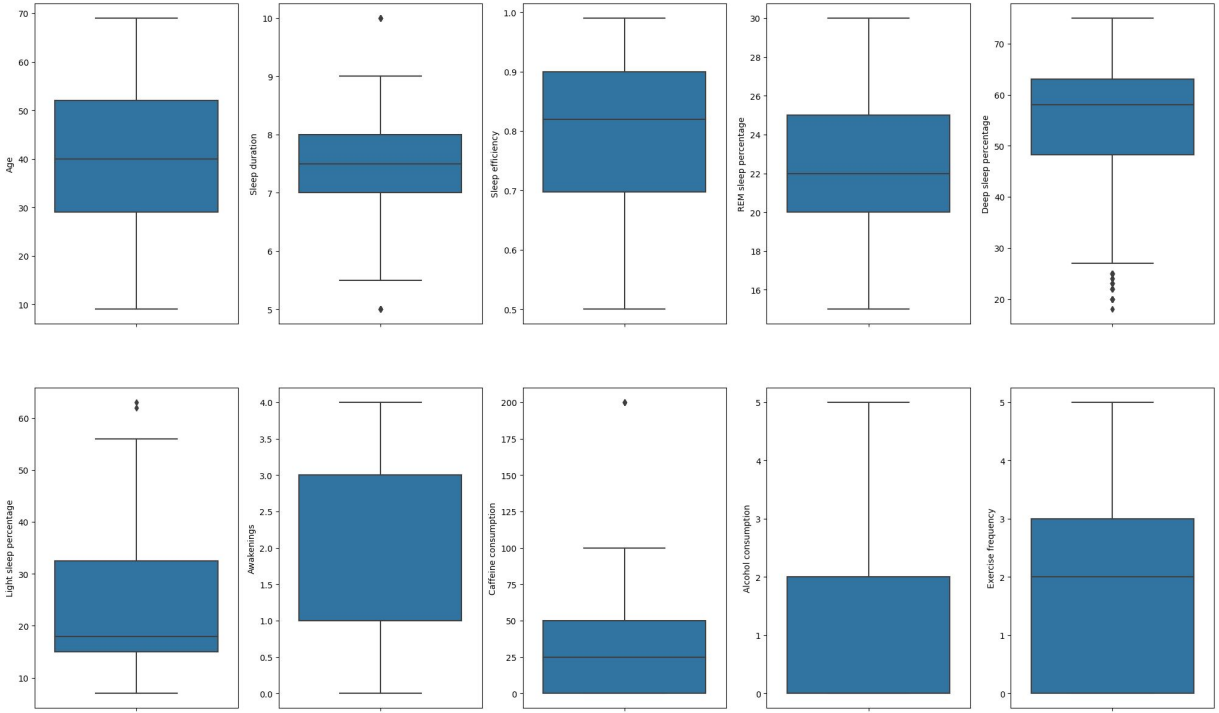
1. Column Bedtime, Wakeup time should be datetime.
    a. Format as datatime
2. Column Awakenings, Exercise Frequency should be int.
    a. Meaning # awakenings during night & # exercise per week.

```
RangeIndex: 452 entries, 0 to 451
Data columns (total 15 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   ID                      452 non-null    int64
 1   Age                     452 non-null    int64
 2   Gender                  452 non-null    object
 3   Bedtime                 452 non-null    datetime64[ns]
 4   Wakeup time             452 non-null    datetime64[ns]
 5   Sleep duration          452 non-null    float64
 6   Sleep efficiency        452 non-null    float64
 7   REM sleep percentage    452 non-null    int64
 8   Deep sleep percentage   452 non-null    int64
 9   Light sleep percentage  452 non-null    int64
 10  Awakenings              452 non-null    int32
 11  Caffeine consumption    452 non-null    float64
 12  Alcohol consumption     452 non-null    float64
 13  Smoking status          452 non-null    object
 14  Exercise frequency      452 non-null    int32
dtypes: datetime64[ns](2), float64(4), int32(2), int64(5), object(2)
```

Data cleaning - Outlier Treatment

Columns with outliers:

- Sleep Duration
- Deep Sleep Percentage
- Light sleep Percentage
- Caffeine Consumption



Data cleaning - Outlier Treatment

q1 = quantile(0.25)

q3 =quantile(0.75)

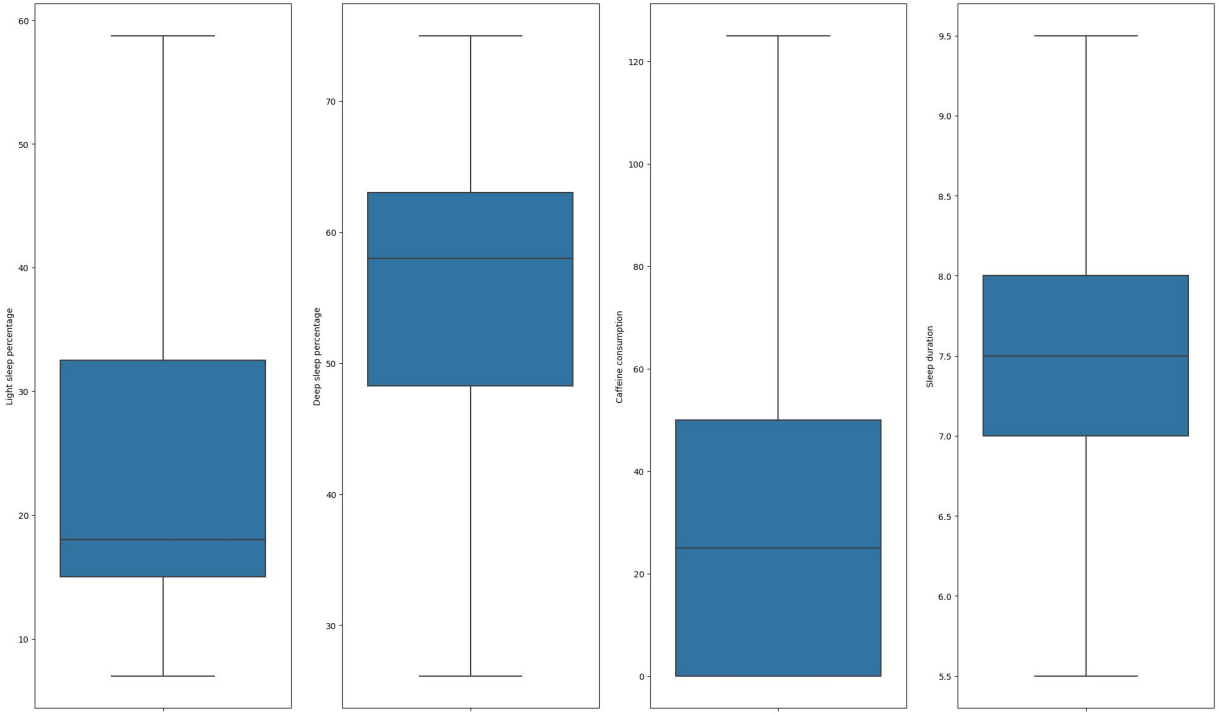
iqr = q3-q1

min = q1 - 1.5 * iqr

max = q3 + 1.5 * iqr

if value > max, then value = max

if value < min, then value = min



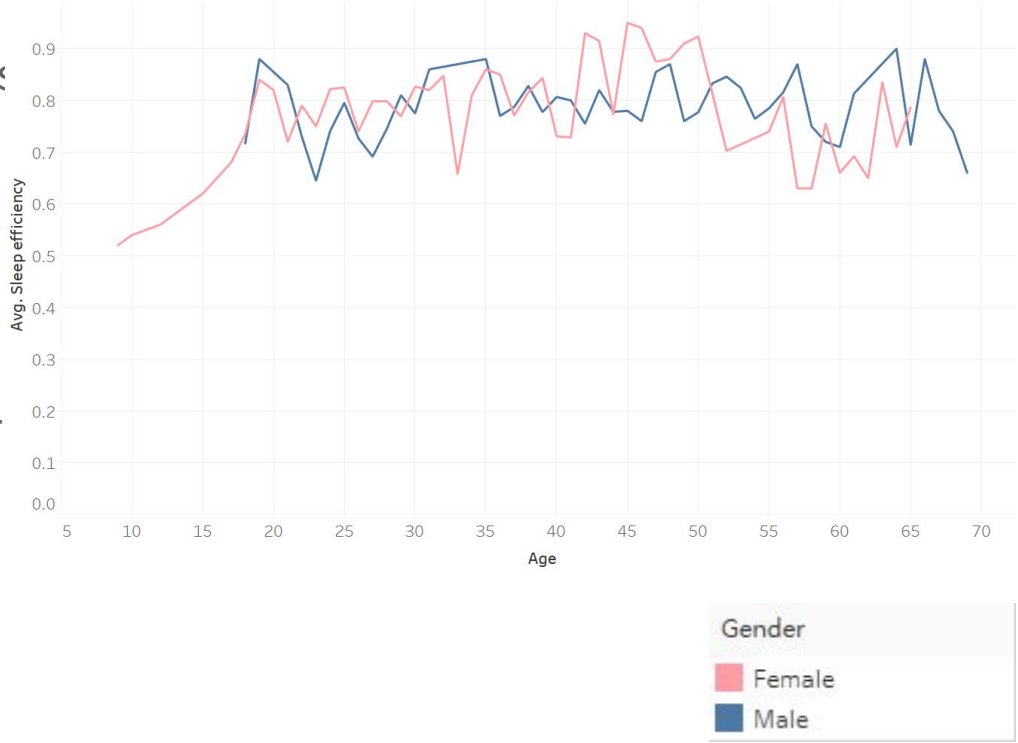
Age vs Sleep Efficiency

Female:

- sleep efficiency increases in theirs 40s.
- Drop in 30s and 50s

Male:

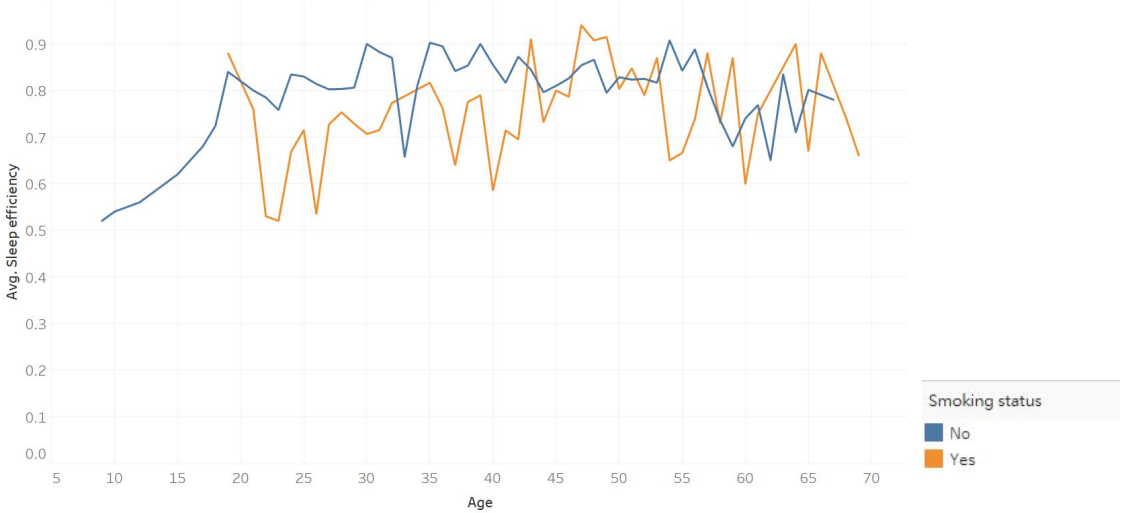
- Sleep efficiency increases in their 60s.
- Drop in 20s

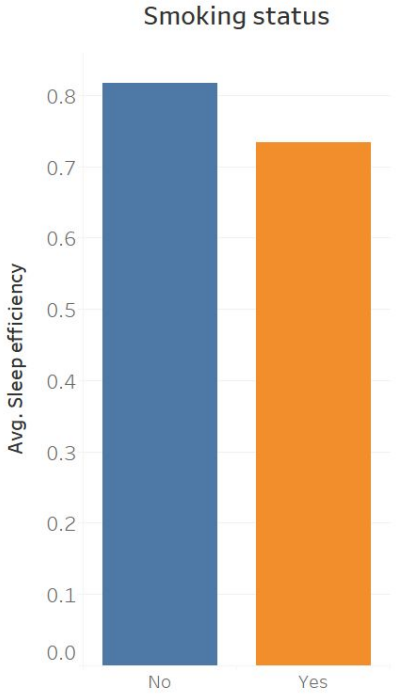


Smoking Status vs Sleep Efficiency

On average and most of the time, non-smokers have better sleep efficiency than smoker.

Opposite trend appears in age 45 - 50 .



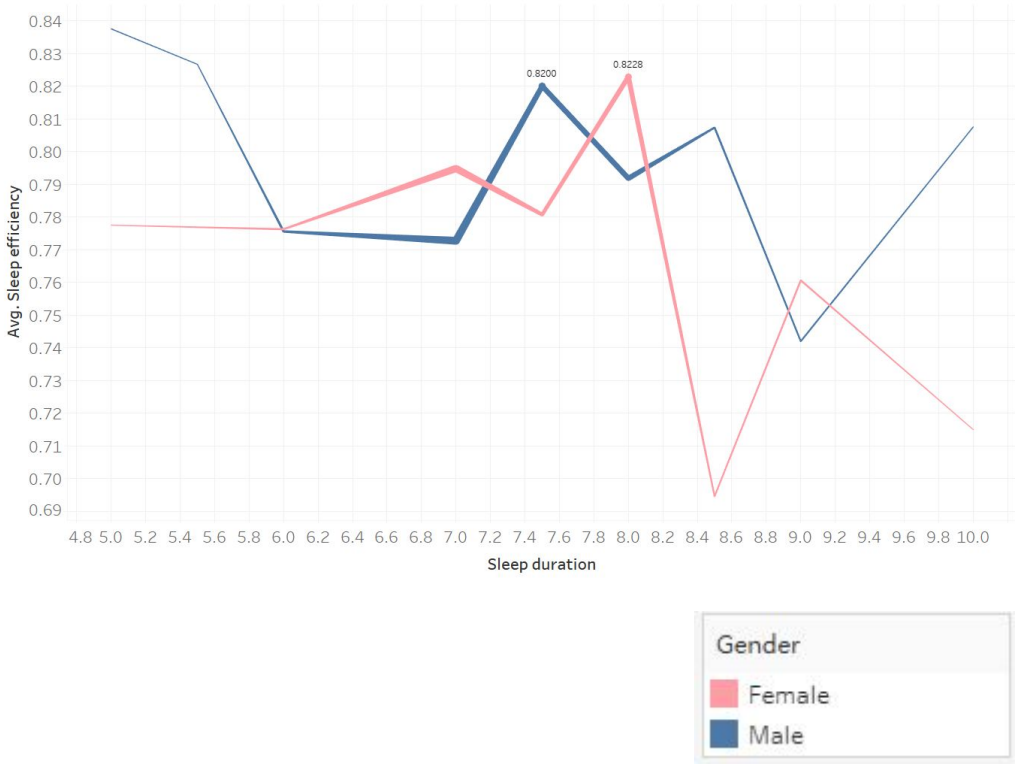


Sleep Duration vs Sleep Efficiency

Sleep duration of 7.4 - 8.2 hr can get best sleep efficiency.

For male, 7.5 hour of sleep gets highest sleep efficiency of 0.8200.

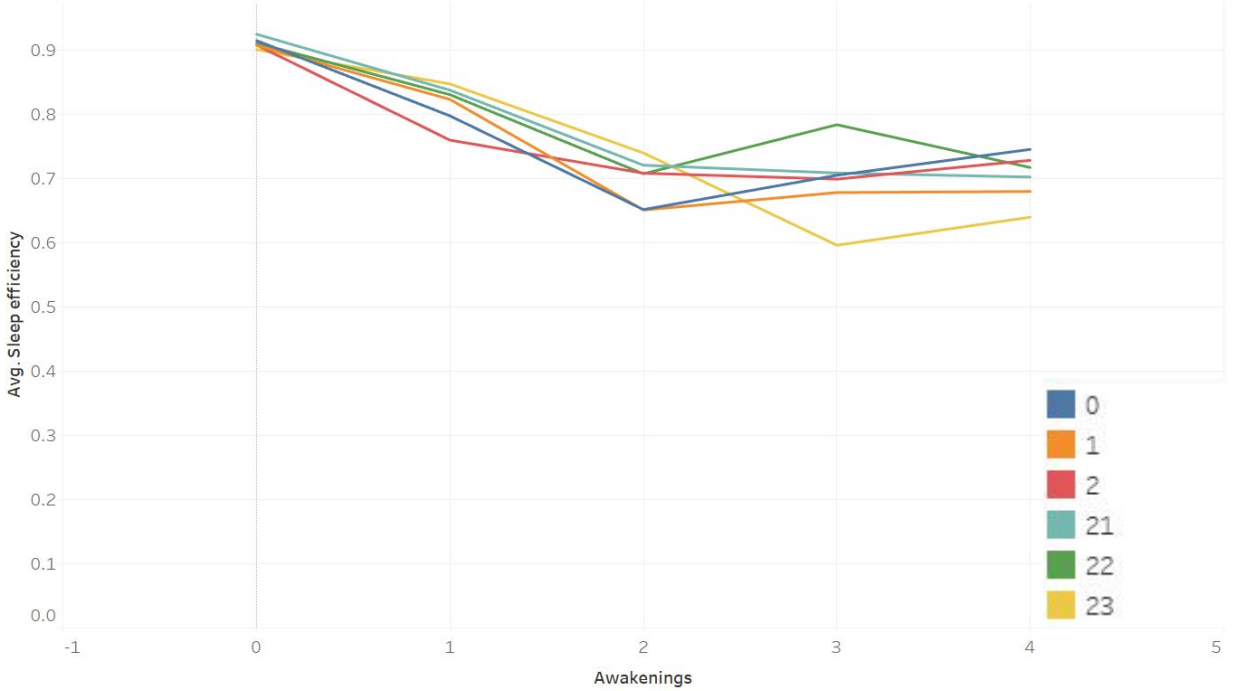
For female, 8 hour of sleep gets highest sleep efficiency of 0.8228.



Awakenings and Bedtime Hour vs Sleep Efficiency

The lesser awakenings, the better sleep efficiency.

Sleeping at 22, awaking 3 times has higher sleep efficiency than 2 times.



Further Research

- Research on Caffeine consumption vs Sleep efficiency
- Research on Light sleep & REM vs Sleep efficiency
- Research on possible reason for awaking 3 times having higher sleep efficiency than 2 times when sleeping at 22.