universidade de aveiro
theoria poiesis praxis

# Exploring Sleep Health and Lifestyle Patterns Through Python Data Analysis

Afonso Duarte Do Fundo Ruela Branco Carreira[1] (107988)

Beatriz Nunes Lavado[1] (102495)

Tomás Vasconcelos Branco Serras Geraldes[1] (107508)

1 – Department of Medical Sciences, Master's in Clinical Bioinformatics

**Abstract**

Sleep is essential for maintaining physical, mental, and emotional well-being, yet sleep disorders such as insomnia and sleep apnoea are increasingly common, affecting daily performance and relationships. This study explores the relationship between sleep health and lifestyle patterns using the "Sleep Health and Lifestyle" dataset, which includes 374 observations and 12 variables. Data analysis was performed using Python, focusing on Sleep Duration, Quality of Sleep, and Sleep Disorder. The dataset was cleaned and processed to correct inconsistencies, such as categorizing blood pressure levels and resolving inconsistencies in the BMI and Sleep Disorders variables. Descriptive and exploratory data analysis revealed key relationships, including a strong negative correlation between stress levels and both sleep duration and quality. A user-friendly Python program was developed, allowing interactive exploration of variables, detailed analyses, and visualizations. This study shows how Python can be used to process and analyse data effectively, making it a useful tool for finding insights in complex datasets Future work could incorporate inferential statistics and machine learning to uncover deeper patterns and improve our understanding of the factors affecting sleep quality.

## 1. Introduction

Sleep is a fundamental biological need, playing a crucial role in physical health, cognitive function, and emotional well-being. However, sleep disorders such as sleep apnoea, and insomnia can significantly disrupt this essential process. Sleep apnoea, characterized by repeated interruptions in breathing during sleep, can lead to fragmented rest and an increased risk of cardiovascular issues [1]. Insomnia, on the other hand, often manifests as difficulty falling or staying asleep, resulting in fatigue, reduced concentration, and a lower quality of life [2]. Both conditions emphasize the importance of understanding and addressing the factors that influence sleep quality. Additionally, sleep disorders are becoming more common, impacting people's ability to perform effectively at work or school and maintain healthy interpersonal relationships [3,4].

In this work, we apply data analysis in Python to evaluate factors that influence sleep quality and the presence of sleep disorders. By examining relevant indicators—such as physiological metrics, sleep patterns, and environmental conditions—we aim to uncover significant correlations and insights. This analytical approach provides a valuable foundation for understanding the complexities of sleep health and supporting evidence-based strategies to enhance well-being.

## 2. Materials and Methods

This study utilizes a “Sleep Health and Lifestyle” dataset, which provides detailed information on various factors related to sleep and general health. The dataset was processed and analysed using Python, with its robust packages for data manipulation, visualization, and statistical analysis. The primary focus was on three key variables related to sleep: Sleep Duration, Quality of Sleep, and Sleep Disorders. These variables were selected based on their direct relevance to understanding sleep health.

The methodology adopted in this work involved four main steps. First, the dataset was cleaned to address any missing or inconsistent data, ensuring the accuracy of subsequent analysis. Secondly, descriptive statistic methods were employed to get a full perspective of the data and its variables, with emphasis on the percentage of people with sleep disorders. Next, exploratory data analysis (EDA) was conducted. As a preliminary step, correlations between all variables in the dataset were

calculated, aiming to identify significant relationships. Particular attention was given to understanding how the three primary sleep variables interact with other features, such as lifestyle habits and health indicators. Additionally, and as an example of a first approach to inferential statistics, chi-squared tests were applied to check the relation of BMI and sleep quality with sleep disorders. Lastly, all components were consolidated into a single main function with an intuitive menu, giving users the flexibility to explore the dataset in their preferred way. The packages used were pandas, seaborn, scipy, tabulate and matplotlib.

## 3. Results and discussion

### 3.1. Data Processing and Explanation

The dataset contains 374 observations and 12 variables. Some variables required processing to simplify their use. For example, the BMI category variable needed adjustments to address inconsistencies (**Fig. 1A.**), and the "None" level in the Sleep Disorder variable was incorrectly recognized as NA, requiring appropriate corrections (**Fig. 1B.**). Blood Pressure was transformed into a categorical variable and grouped into levels, as shown in **Figure 1C**.

After cleaning the data, **Figure 2** in appendix shows descriptive statistics for some of the variables, including the mean, standard deviation, minimum and maximum values, quantiles, and frequencies for qualitative measures. **Figure 2A.** shows that, on average, the individuals in the dataset are 42 years old and sleep about 7 hours per night. This amount may or may not be enough for full recovery, as it varies greatly between people. The average rest level, on a scale from 0 to 10, was 7, which is considered quite good. It is also clear that most individuals are male, and more than 50% do not suffer from any sleep disorders (**Fig. 2B.**).

**Figure 3** illustrates the prevalence of sleep disorders within the sample. It can be observed that approximately 41% of individuals present with either apnoea or insomnia, with each disorder exhibiting a prevalence of approximately 21%. In contrast, approximately 59% of the sample does not exhibit any sleep disorder.

### 3.2.Exploratory Data Analysis (EDA)

The variables with the highest correlation were stress level with quality of sleep and stress level with sleep duration, indicating a negative relationship between these variables (-0.9 and -0.81, respectively), and a positive correlation

between sleep duration and sleep quality (0.88) (**Fig. 4**). The analysis of the correlation table revealed insights into the dependence of sleep on other variables, providing a foundation for further EDA.

In order to understand the potential influence of all variables on sleep disorders, a series of graphics were developed and incorporated into the main function. As an example, **Figure 5** shows sample distribution by BMI category and proportion of sleep disorders by BMI category, allowing to conclude that overweight and obesity have great impact on sleep (all obese individuals in the sample have a sleep disorder, and overweight have a small proportion of individuals with no disorder when compared to normal weight). In terms of sleep duration by sleep disorder, **Figure 6** shows that individuals with insomnia exhibit reduced sleep duration compared to other groups, while individuals with apnoea display greater variation of sleep duration than those without any disorder.

Furthermore, chi-squared independence tests were applied to check the association between sleep disorders and other variables, namely BMI categories and sleep quality. This kind of testing was used to show the suitability of inferential statistics for the analysis of this data, which would represent the next step of this research.

### 3.3. Main Function Structure

The main function of the Python code is built around a user menu, as shown in **Figure 7**, which provides 9 options for interaction. A While True loop is used to keep the menu running indefinitely until the user chooses to exit, ensuring continuous navigation through the program.

When options 2 or 3 are selected from the main menu, new sub-menus are opened with additional options for the user. If option 2 is chosen, the new menu displayed can be seen in **Figure 8A**. This menu is designed for a more detailed analysis of individual variables in relation to the main variable, Sleep Disorder, including the presentation of graphics and relevant statistics. On the other hand, selecting option 3 opens the menu shown in **Figure 8B**, which corresponds to the exploratory data analysis and allows users to visualize graphical relationships between multiple variables under study. These sub-menus provide a structured way to explore the dataset, making the analysis interactive and user-friendly.

## 4. Conclusions

The data processing and EDA in Python were successful, giving useful insights into the relationships between key variables like Sleep Duration, Quality of Sleep, and Sleep Disorder. These analyses created a solid foundation for understanding the dataset and the factors that influence sleep health.

The user-friendly menu provides a quick and easy way to explore the dataset. This allowed for both detailed analysis of individual variables in relation to Sleep Disorder and the visualization of relationships between multiple variables, helping users gain initial insights without needing advanced technical skills.

In the future, more advanced methods such as inferential statistics and machine learning techniques could be used to deepen the analysis. These approaches would enable the discovery of more complex patterns and stronger conclusions, offering a better understanding of the factors affecting sleep quality.

## 5. References


**1**. Jean-Louis G, Zizi F, Clark LT, Brown CD, McFarlane SI. Obstructive sleep apnea and cardiovascular disease: role of the metabolic syndrome and its components. J Clin Sleep Med 2008; 4: 261–72.

**2**. Fernandez-Mendoza J, Vgontzas AN. Insomnia and its Impact on Physical and Mental Health. Curr Psychiatry Rep 2013; 15: 418.

**3**. Ahn E, Baek Y, Park J-E, Lee S, Jin H-J. Elevated prevalence and treatment of sleep disorders from 2011 to 2020: a nationwide population-based retrospective cohort study in Korea. BMJ Open 2024; 14: e075809.

**4**. Peng J, Zhang J, Wang B *et al.* The relationship between sleep quality and occupational well-being in employees: The mediating role of occupational self-efficacy. Front Psychol 2023; 14.


## 7. Figures

A.
```python
df['BMI Category'] = df['BMI Category'].replace('Normal', 'Normal Weight')
```

B.
```python
df['Sleep Disorder'] = df['Sleep Disorder'].fillna('None')
df['Sleep Disorder'] = pd.Categorical(df['Sleep Disorder'],
                                      categories=['None', 'Sleep Apnea', 'Insomnia'],
                                      ordered=True)
```

C.
```python
# Função para categorizar a pressão sistólica
def categorizar_pressao(leitura):
    sistolica = int(leitura.split('/')[0])

    if 110 <= sistolica < 120:
        return "110-119"
    elif 120 <= sistolica < 130:
        return "120-129"
    elif 130 <= sistolica < 140:
        return "130-139"
    else:
        return "140-149"

# Aplicando a função e criando a nova coluna 'BloodPressureGroups'
df['BloodPressureGroups'] = df['Blood Pressure'].apply(categorizar_pressao)
```

**Figure 1** – A) Treatment of BMI Category variable; B) Treatment of Sleep Disorder variable; C) Categorizing and grouping Blood Pressure variable.

A.

| | Age | Sleep Duration | Quality of Sleep |
|---|---|---|---|
| mean | 42.18 | 7.13 | 7.31 |
| std | 8.67 | 0.8 | 1.2 |
| min | 27 | 5.8 | 4 |
| 25% | 35.25 | 6.4 | 6 |
| 50% | 43 | 7.2 | 7 |
| 75% | 50 | 7.8 | 8 |
| max | 59 | 8.5 | 9 |

B.

| | Gender | Sleep Disorder |
|---|---|---|
| unique | 2 | 3 |
| top | Male | None |
| freq | 189 | 219 |

**Figure 2** – A) and B) Statistics for some of the variables.

A.
```python
#percentagem de doentes
percent = df['Sleep Disorder'].value_counts(normalize=True) * 100
percent=percent.round(2)
print(percent)

#Distribuição da amostra por sleep disorder
plt.figure(figsize = (10, 5))
sleepD_order = df['Sleep Disorder'].value_counts().index
sns.set_style("whitegrid")
plt.gca().set_title('Sample Distribution per Sleep Disorder')
sns.countplot(y = 'Sleep Disorder', orient='h',
              palette = 'deep', data = df, order=sleepD_order)
plt.xlabel('Frequency')
plt.show()
```

B.

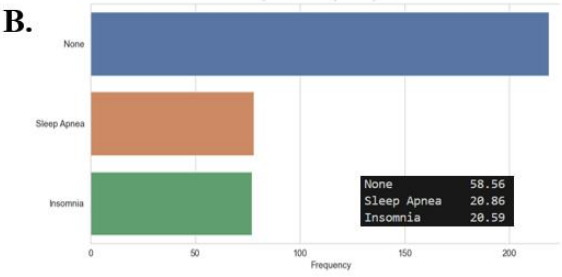

**Figure 3** – A) Code for sleep disorder percentage and code for the distribution of sleep disorder graphic; B) Distribution of Sleep Disorder.

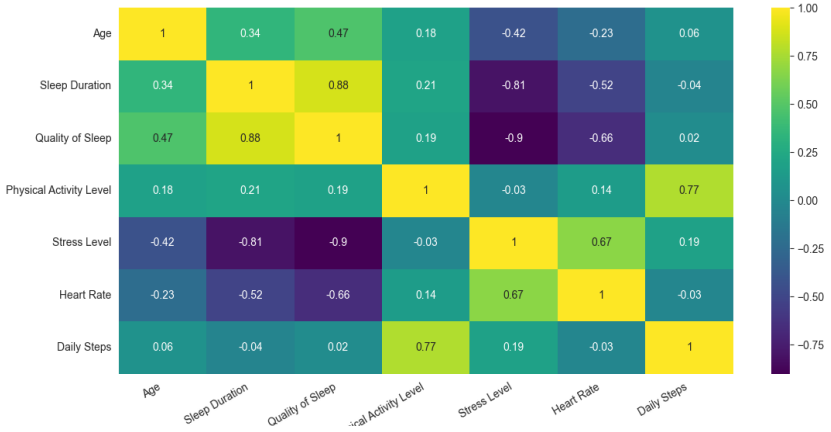


**Figure 4** – A) Correlations between all variables (except Sleep Disorders) .

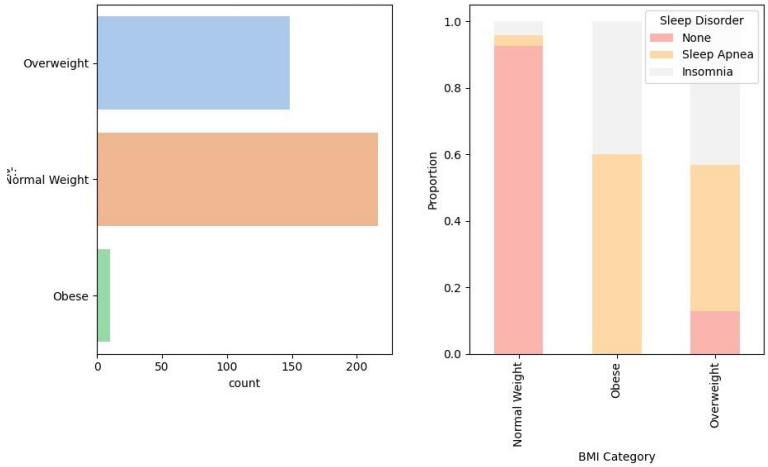


**Figure 5** – Sample distribution by BMI category (left); Sample distribution (proportion) by sleep disorder and BMI category (right).

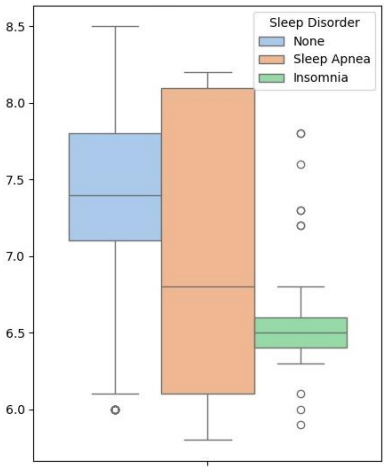


**Figure 6** – Boxplot of sleep duration by sleep disorder.

```
Sleep is a physiological need present in animals, but it is increasingly being neglected, particularly due to work pressures, he
avy workloads, and productivity demands. In summary, this study aims to show how lifestyles can be associated with sleep quality
 and vice versa, using Python code to visualize the results and conclusions of the work.
To learn more about the work, the following options are available:
1 - Contextualization about sleep.
2 - Study variables. (see graphics and dispersion measures)
3 - Most relevant variables' associations
4 - Qui-Square Test between BMI Categories and Sleep Disorders
5 - Qui-Square Test between Sleep Quality and Sleep Disorders
6 - See dataset structure
7 - Save data on a file
8 - See credits
99 - Terminate
Your answer:
```

**Figure 7** – Main menu of the program.

**A.**

```
This menu allows the individual analysis of each variable and their association with the main variable: Sleep Disorders
The possible options for analysis are:
1  -  Gender
2  -  Age
3  -  Occupation
4  -  Sleep Duration
5  -  Quality of Sleep
6  -  Physical Activity Level
7  -  Stress Level
8  -  BMI Category
9  -  Blood Pressure
10  -  Heart Rate
11  -  Daily Steps
12  -  Sleep Disorder
13  -  Blood Pressure Groups
99 - Go back to the main menu
Write the number of the pretended option:
```

**B.**

```
Welcome to the exploratory data analysis (EDA) between variables page.
The available options are:
1 - All variables
2 - Occupation vs Sleep disorder
3 - Occupation vs Sleep Quality
4 - Blood Pressure vs Sleep disorder
5 - Stress level vs Sleep Disorder
6 - Stress level vs Sleep Quality
7 - Stress Level vs Sleep Duration
99 - Go back to the main menu.
Your choice:
```

**Figure 8** – A) User menu opened if option 2 is selected; B) User menu if option 3 is selected.