Faculty of Business Administration (in Foreign Languages)

International Master in Business Administration

Research Methods for Business Administration

**Sleep health influenced by lifestyle factors**

Professors: Students:

### Petre Caraiani Torjescu Ana-Maria

### Daniel Traian Pele Dumitru Laura Alexandra

Wolfgang Hardle Costea Cristina Bianca

Bucharest

2024

**Abstract**

*Sleep is a fundamental determinant of health, yet its quality is often compromised by lifestyle factors. This study aimed to investigate the relationship between sleep health and lifestyle factors, specifically sleep duration, physical activity level, and stress. Using a dataset of a sample of 374 , we employed descriptive analytics, visualized through scatterplots, to explore initial associations between the variables. Moreover, an OLS regression model was developed to quantify these relationships and predict the quality of sleep from the selected lifestyle factors. The main findings suggest that the model demonstrated that approximately 80.9% of the variance in sleep quality could be explained by sleep duration and stress level. Stress level was found to be a significant predictor of sleep quality. The results underscore the complex interplay between lifestyle habits and sleep health, highlighting the significance of stress management in improving sleep quality.*

**Keywords:** Sleep Quality, Lifestyle Factors, Physical Activity, Stress, Regression,Sleep Duration

**Introduction**

The topic of sleep is one of the most important subjects in terms of health, which can positively or negatively influence us in many ways. Sleep is essential for health, and sometimes it does not get the necessary, deserved attention. With the increasing prevalence of sleep disorders and the growing recognition of sleep's impact on health, productivity, and quality of life, it is important to understand the relationships between sleep and lifestyle factors. In this context, we have chosen as our research to explore the connections between sleep quality, sleep duration, physical activity levels, and stress. The motivation behind selecting this topic stems from a rising concern over the sleep health of individuals who often report sleep-related issues. These issues not only affect personal health but also have broader social and economic impact.

The subject has been explored and analyzed by many researchers. As such, it is considered that poor sleep has numerous negative effects on the health of each and every individual. Researchers consider that poor sleep increases the risk of appearance of different kinds of diseases, such as: “premature mortality, cardiovascular disease, hypertension, obesity, metabolic syndrome, diabetes and impaired glucose tolerance, immunosuppression, inflammation, cancer, cognitive impairment, and psychiatric disorders such as anxiety and depression” (Dedhia & Maurer, 2022).

Sleep is vital for our daily well-functioning of the organism and our mental, emotional, and physical health (Dedhia & Maurer, 2022).

Based on their study, the National Sleep Foundation Scientific Advisory Council created a recommended framework with sleep ranges for all the types of age groups as follow: newborns(0-3m) between 14h-17h, infants(4-12m) between 12h-16h, toddlers(1y-2y) between 11h-14h, preschoolers(3y-5y) between 10h-13h, school-age children(6y-12y) between 9h-12h, teen (13y,18y) between 8h-10h, adults(18y-60y) more than 7h, adults(61y,64y) between 7h-9h and adults (65+) between 7h-8h (National Sleep Foundation, 2020).

Sleep is a very complex topic which defines a crucial biological process that is considered to restore energy, accelerate the process of healing, strengthen the immune system, and influence both the brain and behavior (Perez-Pozuelo et al., 2020).

In terms of the lifestyle factors that influence sleep quality and duration, it can be enumerated the following: cardiovascular health, such as blood pressure and heart rate, stress level, physical activity, sleep disorders, and other relevant factors that tackles this issue (Perez-Pozuelo et al., 2020).

(Luyster et al., 2012) consider that our sleep tendencies and patterns are heritable from our family in terms of sleep duration, excessive sleepiness, and diurnal preferences.

**Methods**

The primary goal of this study is to assess sleep health and the lifestyle factors that influence it. In order to do that, a quantitative method is going to be used with the help of a Python programme. This type of analysis has been chosen due to the fact that many researchers consider Python to be one of the best choices for analyzing data by managing data structures and complex data manipulation.

In order to have a better understanding of the topic, we have selected and extracted data from the Kaggle database (Tharmalingam, 2023) that focuses on the creation of a Sleep Health and Lifestyle Dataset. The dataset has been extracted by downloading a csv version of it, which was further included and used in our creation of the Python code. The Sleep Health and Lifestyle Dataset comprises 374 rows and 13 columns, covering a wide range of variables related to sleep and daily habits. It includes details such as gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders applied to a sample of 374 individuals (male and female). The main sleep metrics are sleep duration, quality, and factors influencing sleep patterns. Factors are split into lifestyle factors( physical activity levels, stress levels, and BMI) cardiovascular factors (blood pressure and heart rate) and sleep disorder factors (insomnia and sleep apnea). For each individual we have a person ID (an identifier for each individual, gender, age, occupation, sleep duration(hours per day), quality of sleep(scale 1-10), physical activity level(minutes/day), stress level(scale 1-10), BMI category ( Underweight, Normal, Overweight), blood pressure(systolic/diastolic), heart rate (bpm), daily steps(per day), sleep disorder(None, Insomnia, Sleep Apnea).

Based on these data, four research questions have been developed:

1. What is the correlation between the main factors of the dataset?
2. What is the relationship/correlation between Sleep - Physical Activity Level? What about Sleep Duration and Stress Level?
3. What is the linear regression between quality of sleep and sleep duration and to what extent the predicted values differ from actual values?
4. To what extent does the data analyzed represent “a good fit” for the model?

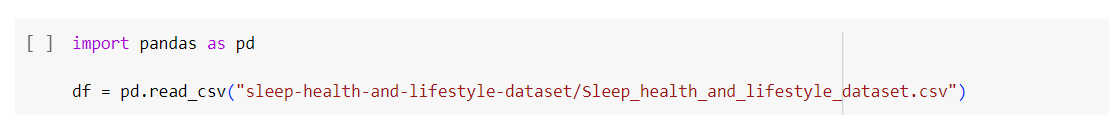
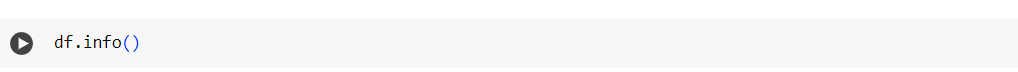
As stated before, for conducting the analysis we used different types of tools in Python such

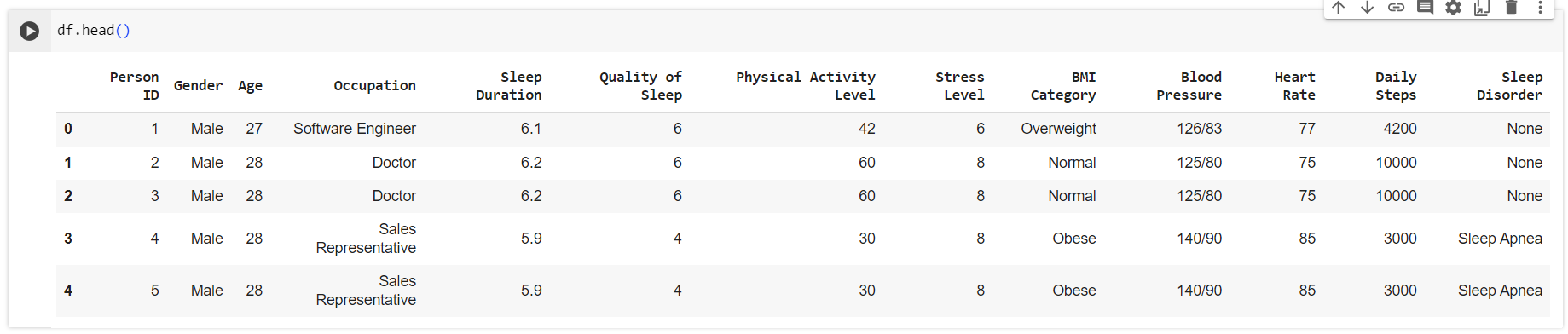
as: correlation matrix, correlation heatmap, pairplots, scatterplots and simple linear regression model through usage of baseline predictor mean average, scatterplot of residuals, prediction plot and Ordinary Least squares table. The first phase of our research involved a descriptive analysis using scatterplots to visualize the relationships between the quality and duration of sleep with physical activity and stress levels. This graphical analysis aimed to identify patterns and trends that could indicate potential areas of concern or interest for further investigation. Following the visual exploration, we conducted a more rigorous analysis employing an Ordinary Least Squares (OLS) regression model. This model helped us quantify the relationships between sleep duration, stress levels, and the quality of sleep. The OLS model's predictive accuracy was evaluated, and the distribution of residuals was examined to assess the model's fit.

**Results**

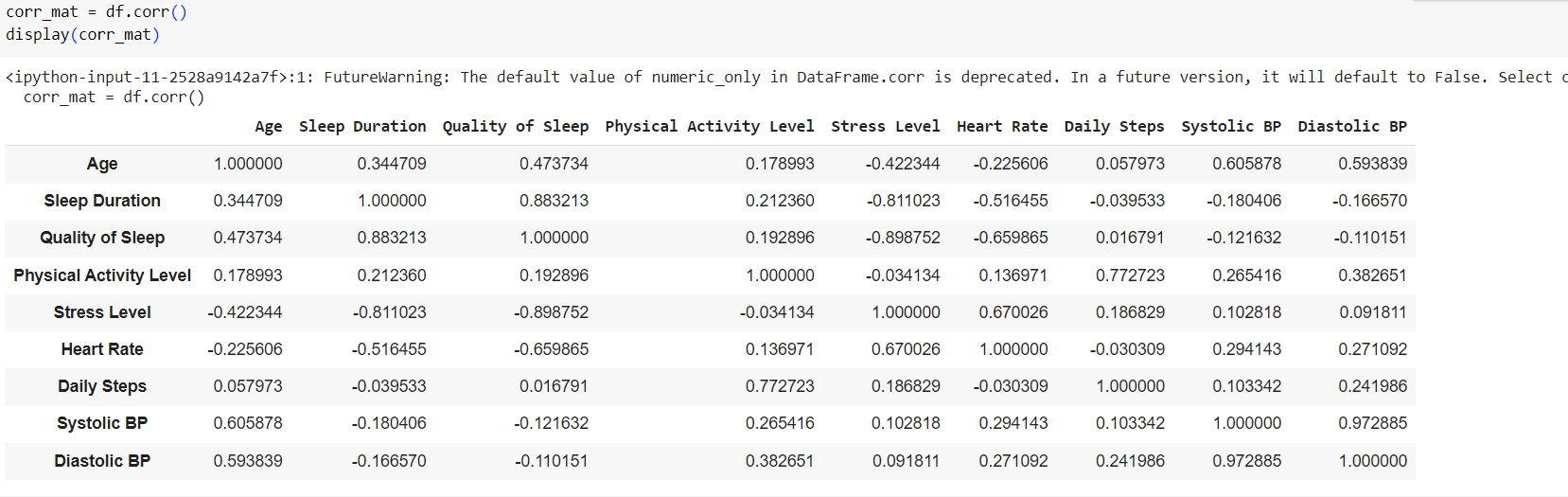
As discussed above, we downloaded the dataset from Kaggle website and inserted them into our Goggle Colab document.

Displays the first 5 rows of the dataframe with all of the factors:



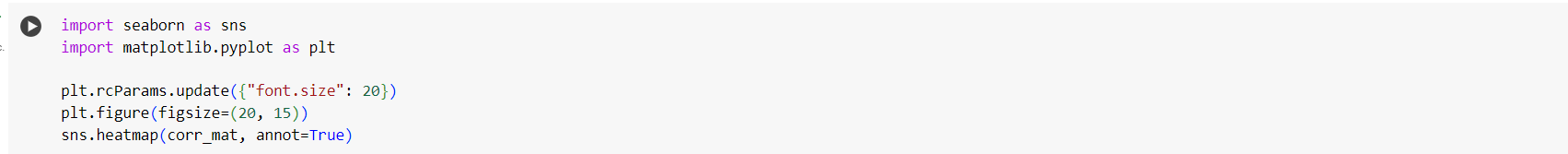


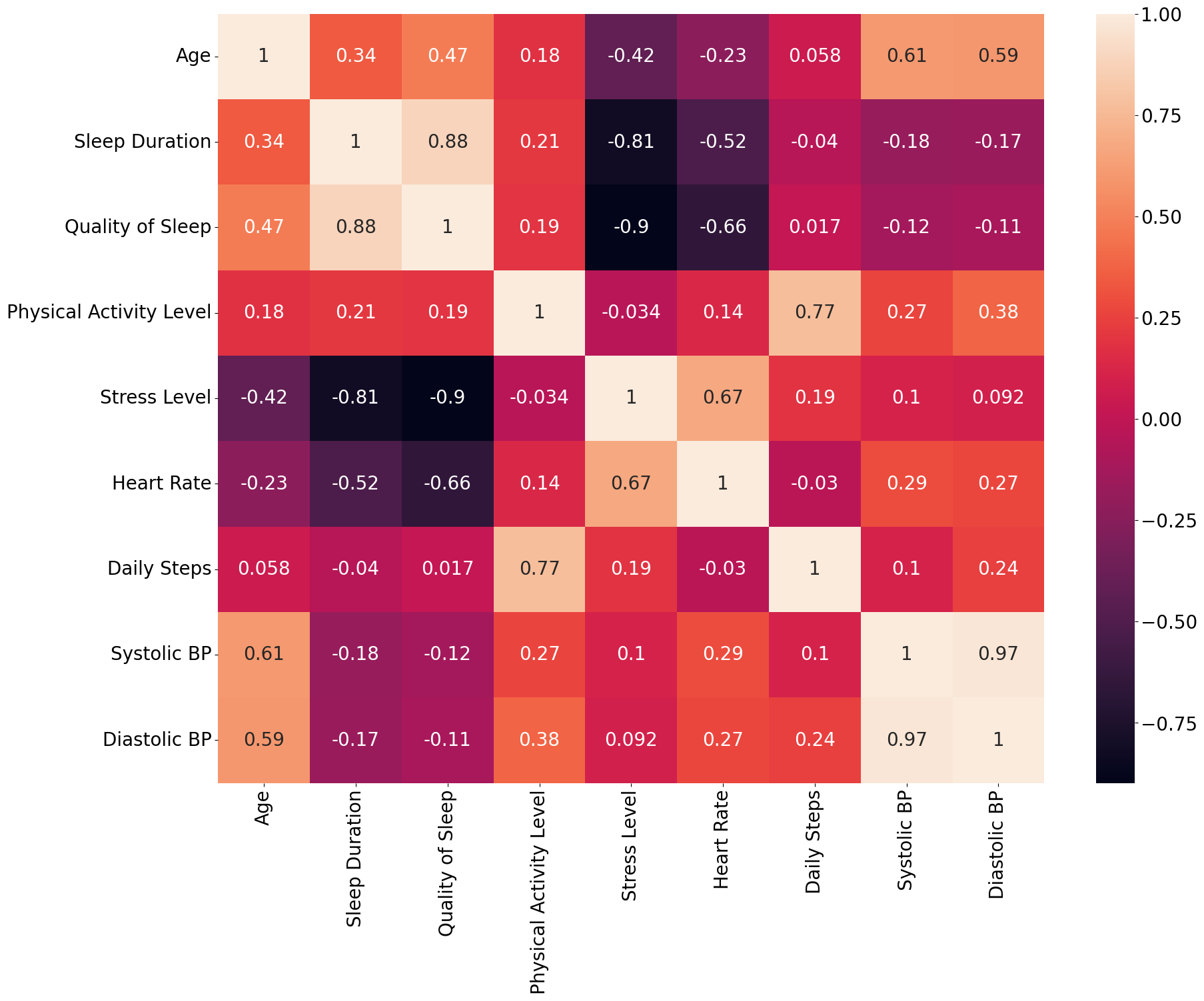
Then we generated the correlation matrix in order to see the correlation between the main factors that influence sleep and to try to respond to the first research question:



The correlation matrix presented above is shown as a table in which every cell contains a correlation coefficient, where 1 is considered a strong relationship between variables, 0 a neutral relationship and -1 a not strong relationship. In order to have a clear view and to be able to easily read the matrix we generated a heatmap.

A very strong correlation relation is between sleep duration and quality of sleep having the value 0.88 which represents a positive correlation. Moreover, quality of sleep is influenced by age having a 0.47 value of the correlation. A very interesting fact is that the stress level seems to not influence the sleep duration and quality of sleep due to the fact that the values are -0.81 and -0.9 which are very close to -1 meaning a quite strong negative correlation. But, the stress level has an influence on the heart rate having the value 0.67, one option could be that the heart rate is increasing as the stress level is increasing too. The sleep duration as well as the quality of sleep is a bit influenced ( 0.21 respectively 0.19) by the physical activity level.



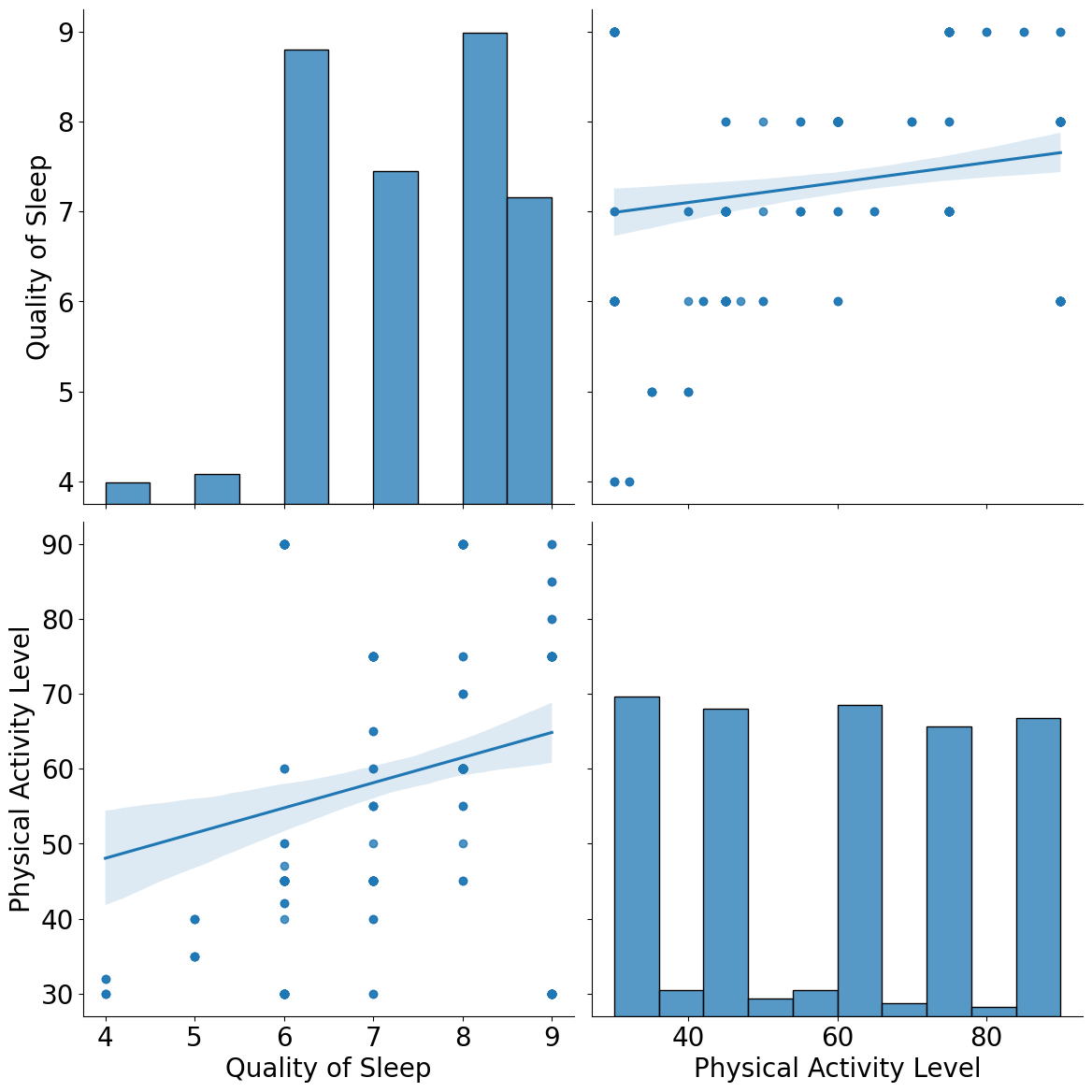


Moreover, to tackle the second research question we created two Pairplots, one between Quality of Sleep - Physical Activity Level and another one between Sleep Duration - Stress Level. Also, we created a Scatterplot between Quality of Sleep - Physical Activity Level and one between Sleep Duration - Stress Level.

From the histograms, we learn that the quality of sleep is heavily left-skewed, negative skew.

The frequency distribution of the various quality scores is displayed in the "Quality of Sleep" histogram, and the distribution of activity levels among the dataset's participants is displayed in the "Physical Activity Level" histogram. Individual data points illustrating the relationship between "Physical Activity Level" (Y-axis) and "Quality of Sleep" (X-axis) are displayed in the scatterplot in the lower-left corner. The regression line shown with a shaded area indicates a positive correlation: physical activity levels tend to rise in tandem with improved sleep quality. The same data points are displayed in the scatterplot in the upper-right corner, but the axes are reversed. Basically, the same information is being presented in a different way. Overall, this pair plot indicates that sleep quality and physical activity level are generally positively correlated, with higher sleep quality being linked to higher levels of physical activity.



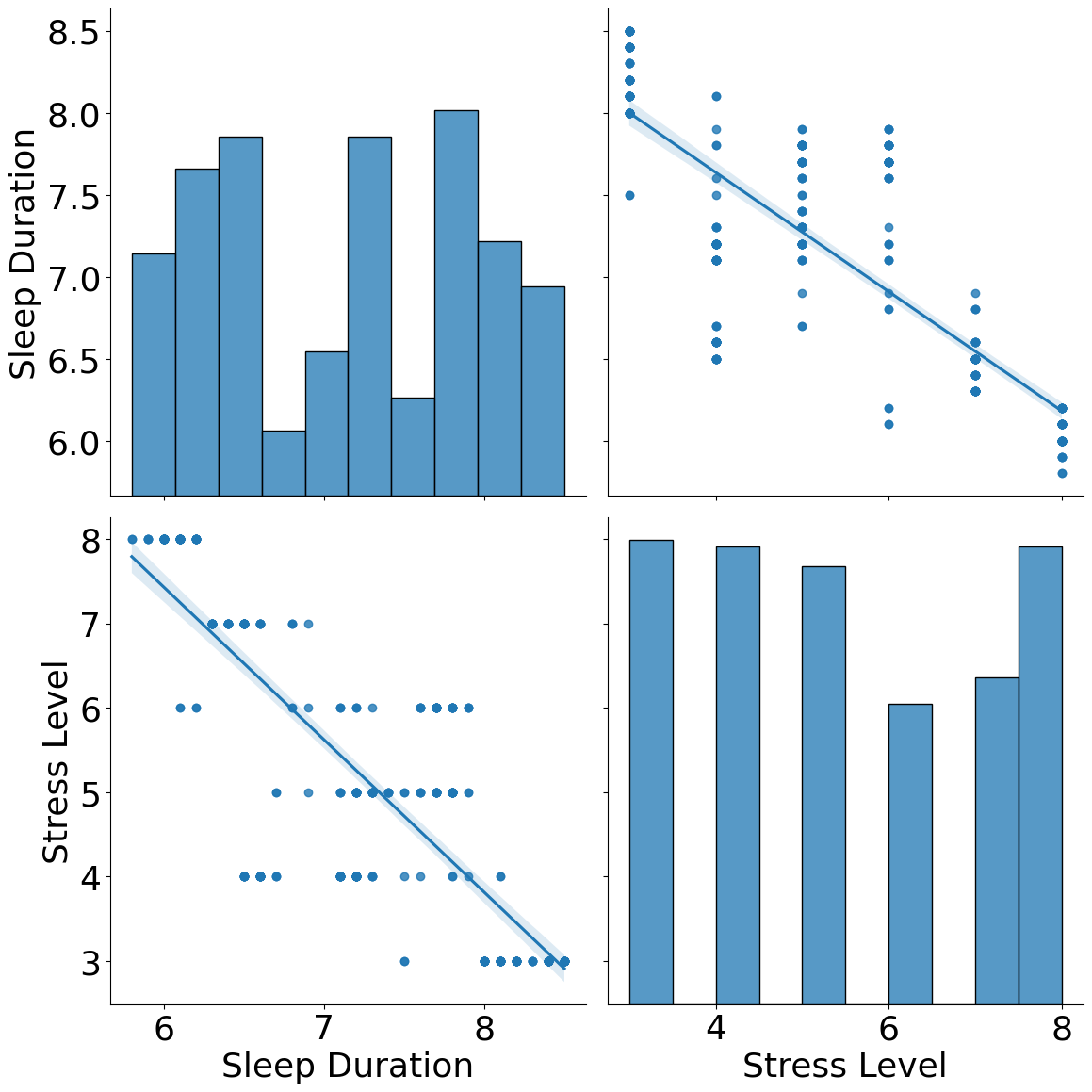


The sleep duration histogram for each member of the dataset indicates a reasonably normal distribution, with the majority of durations falling between 7 and 7.5 hours.

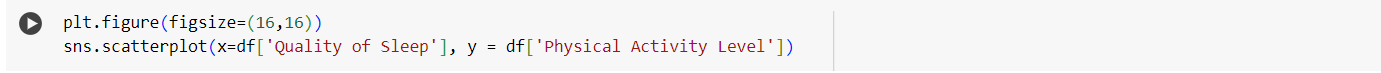
The distribution of the "Stress Level" histogram is more diversified and appears to be slightly skewed to the right. "Sleep Duration" is plotted on the X-axis and "Stress Level" is plotted on the Y-axis in the scatterplot in the lower left corner. According to the regression line, there appears to be a negative correlation between stress levels and sleep duration.

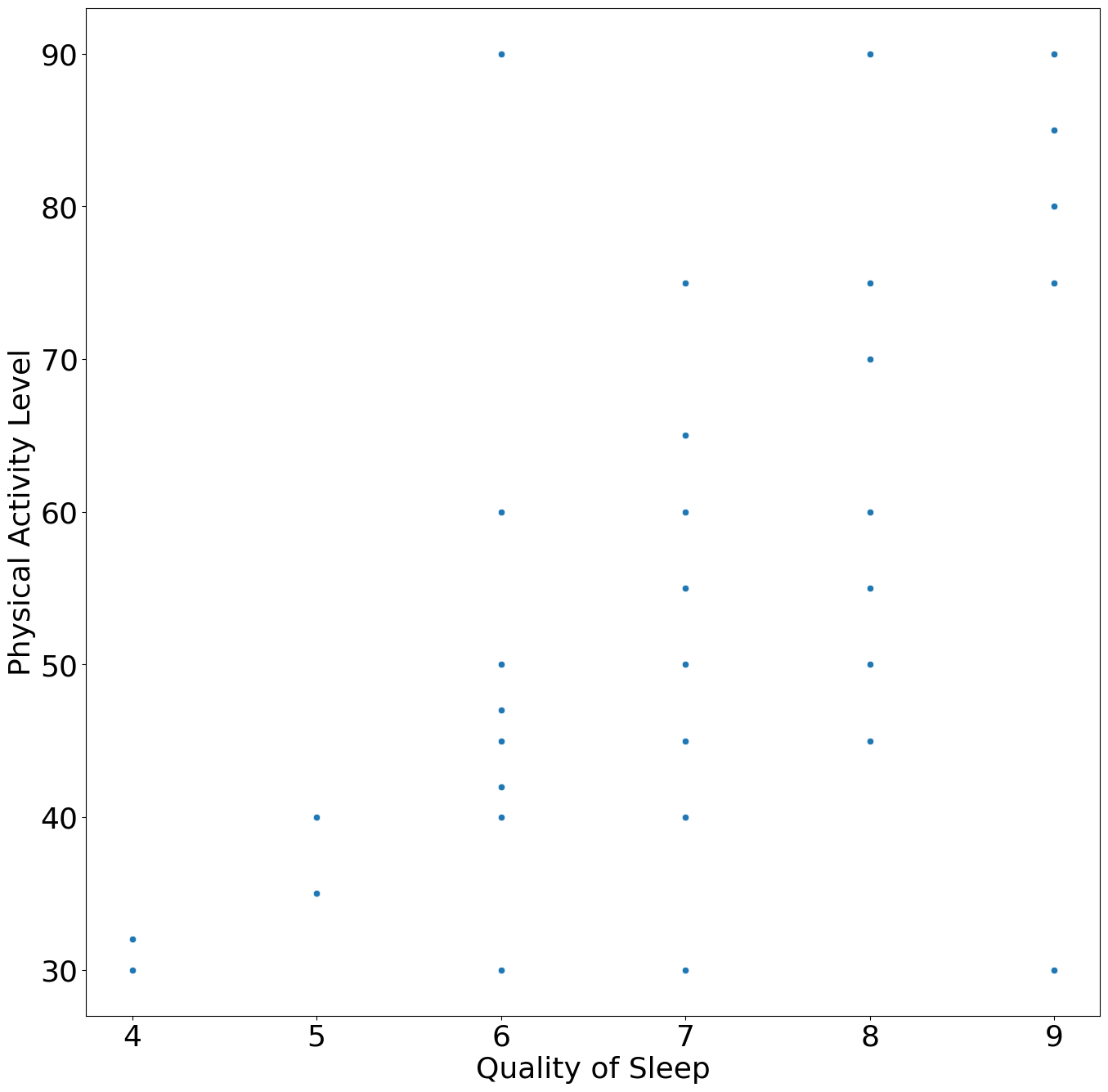
The same data are displayed in the top-right corner of the scatterplot, which displays the variables "Stress Level" on the X-axis and "Sleep Duration" on the Y-axis along with a negative regression line that shows the same relationship. In summary, the pair plot suggests that there is a trend where individuals with longer sleep durations tend to have lower stress levels.





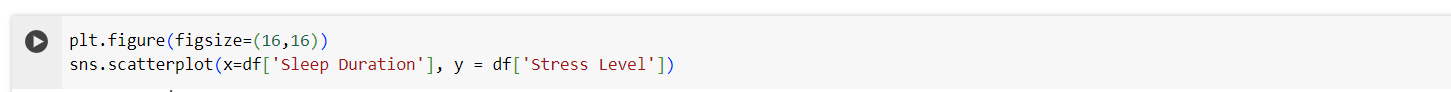
The scatterplot below represents the relationship between “Quality of Sleep” (on the X-axis) and “Physical Activity Level” (on the Y-axis) for a dataset. The data points do not form a distinct pattern that would indicate either a positive or negative correlation, it looks like it is more a neutral correlation that may be influenced by other factors.

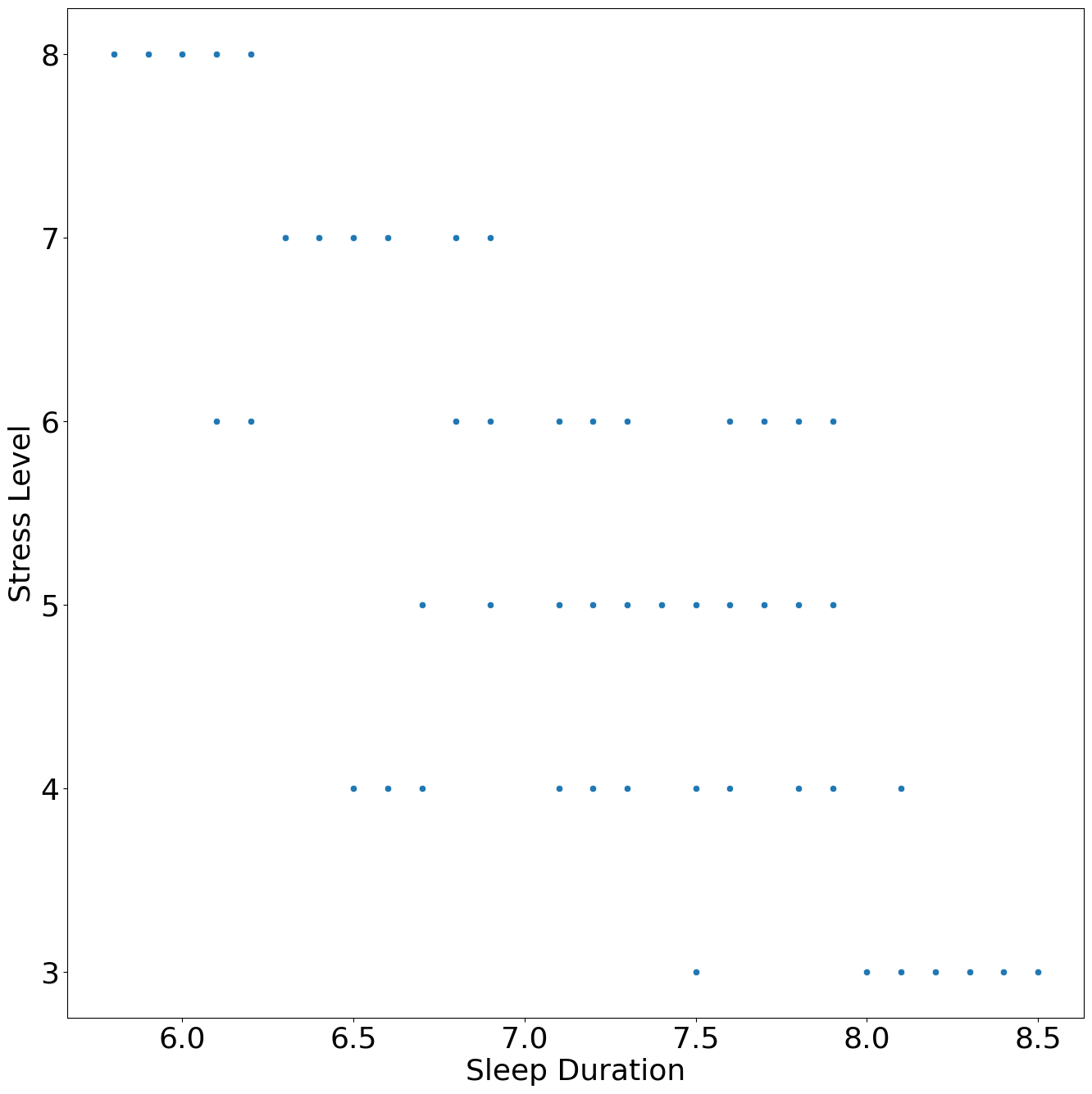




The “Sleep Duration” is shown to range from 6.0 to around 8.5 hours.

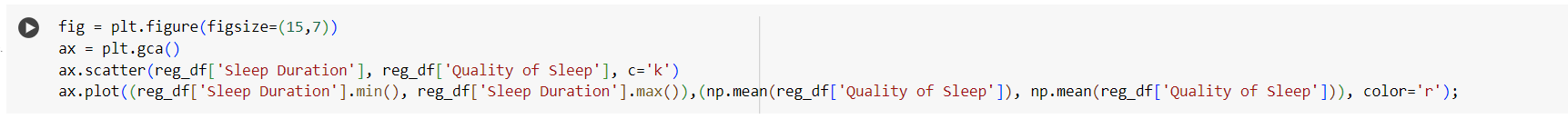
The “Stress Level” ranges from 3 to 8, assuming these are levels of stress. The data points do not form a pattern that would suggest a strong positive or negative relationship between “Sleep Duration” and “Stress Level”.

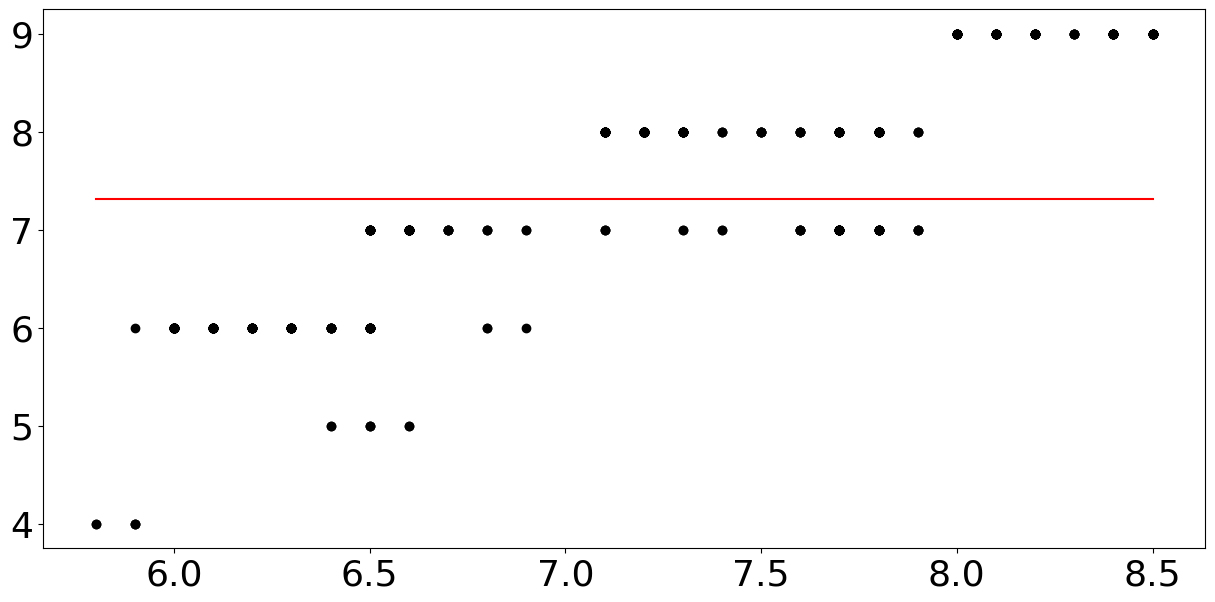




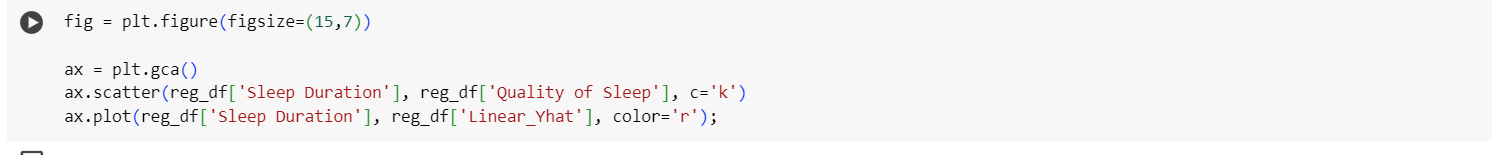
Furthermore, we generated a simple linear regression model for the most relevant two variables: sleep duration and quality of sleep.

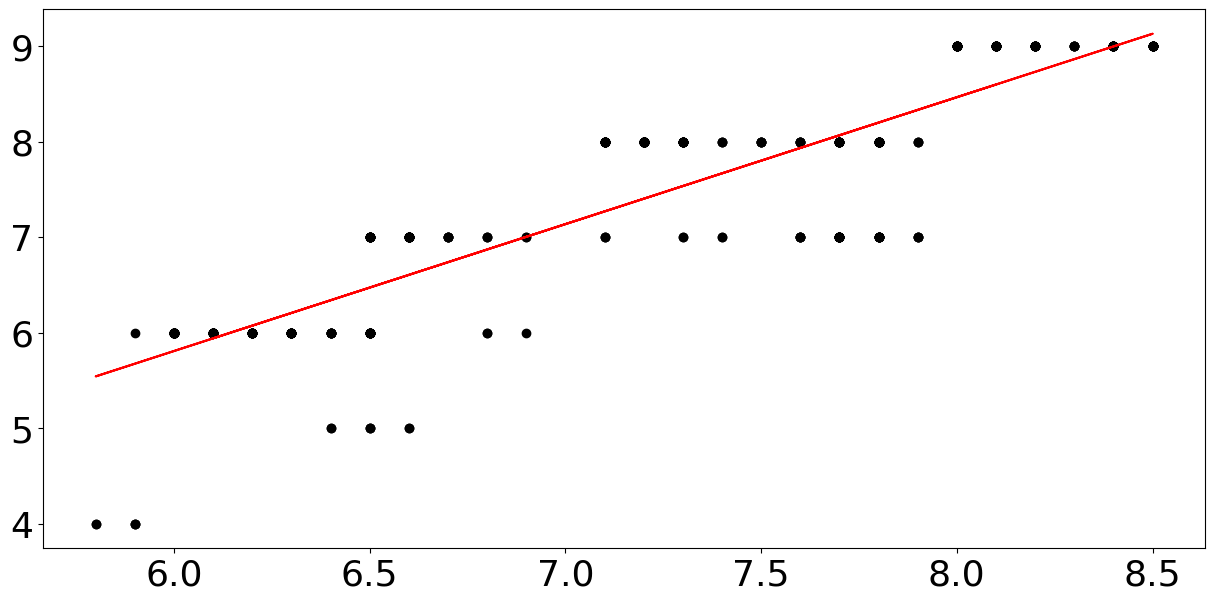
The below scatterplot shows baseline prediction , the relationship between “Sleep Duration” on the X-axis and “Quality of Sleep” on the Y-axis, along with a horizontal red line that represents the mean “Quality of Sleep” across all sleep durations which divides the data in two. There seems to be a concentration of data points at certain “Quality of Sleep” levels, particularly around levels 6 and 7. We can state that the red line suggests that no matter the variation in “Sleep Duration”, the “Quality of Sleep” remains, on average, around the same level (the level at which the red line is drawn). We may say that the “Quality of Sleep” is not strongly affected by the “Sleep Duration” within the ranges of sleep duration observed here.



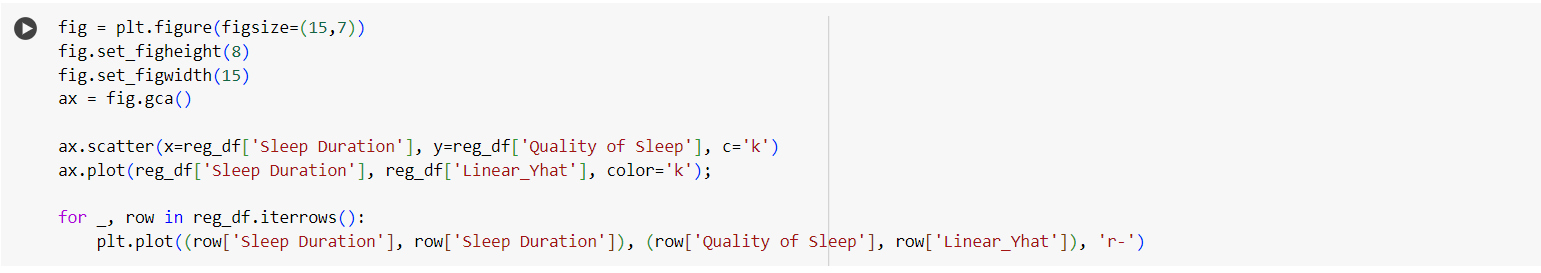


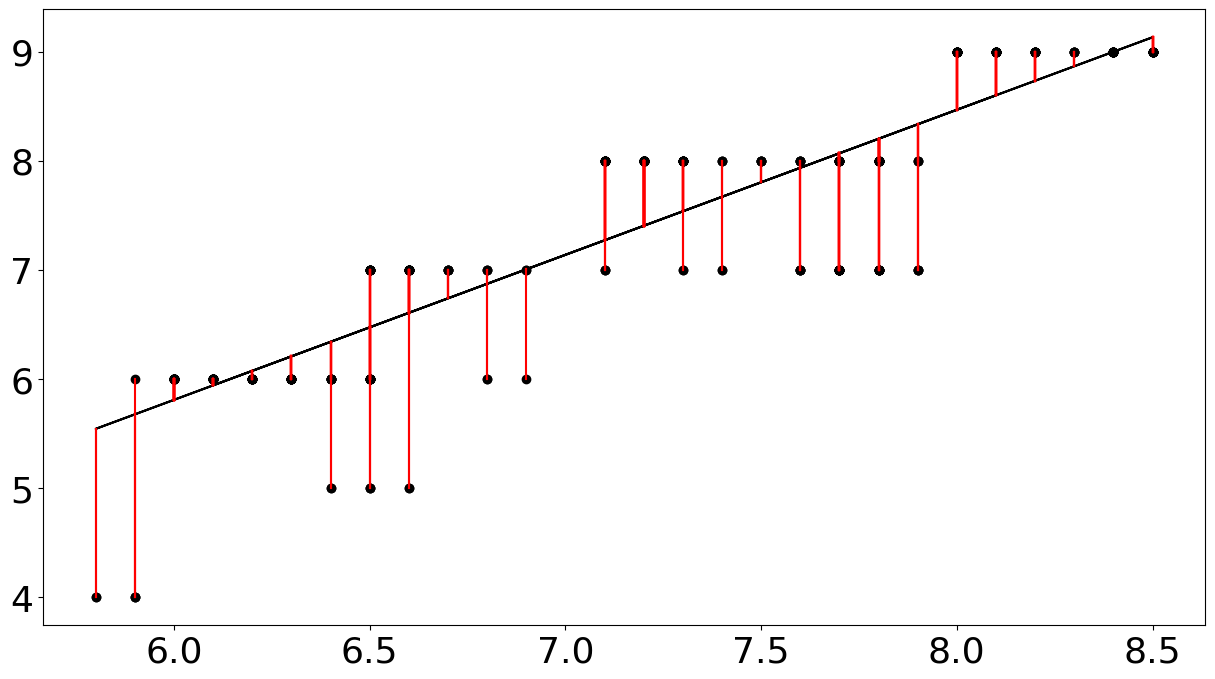
The below scatterplot with a red line that appears to represent a fitted linear regression model. The red line represents the predicted relationship between “Sleep Duration” and “Quality of Sleep” according to the linear regression model. The line suggests that there is a positive relationship between “Sleep Duration” and “Quality of Sleep”: as sleep duration increases, the model predicts an increase in the quality of sleep. If points are closely clustered around the line, it suggests that the model explains a significant portion of the relationship in “Quality of Sleep” based on “Sleep Duration”. We can conclude that the linear regression model predicts that longer sleep duration is associated with higher quality of sleep.



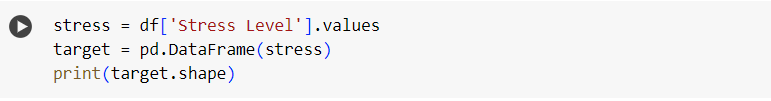


The below scatterplot represents the residuals/the errors from a linear regression analysis. The black line through the data points is the regression line, showing the best fit through the points according to the linear regression model. The length of the red line indicates the magnitude of the error: a longer line means a larger error. As sleep duration increases, the quality of sleep is predicted to improve.

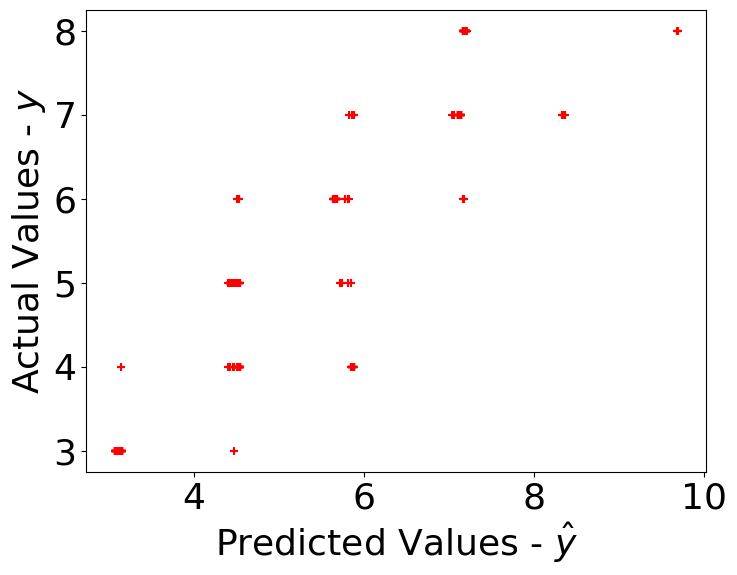




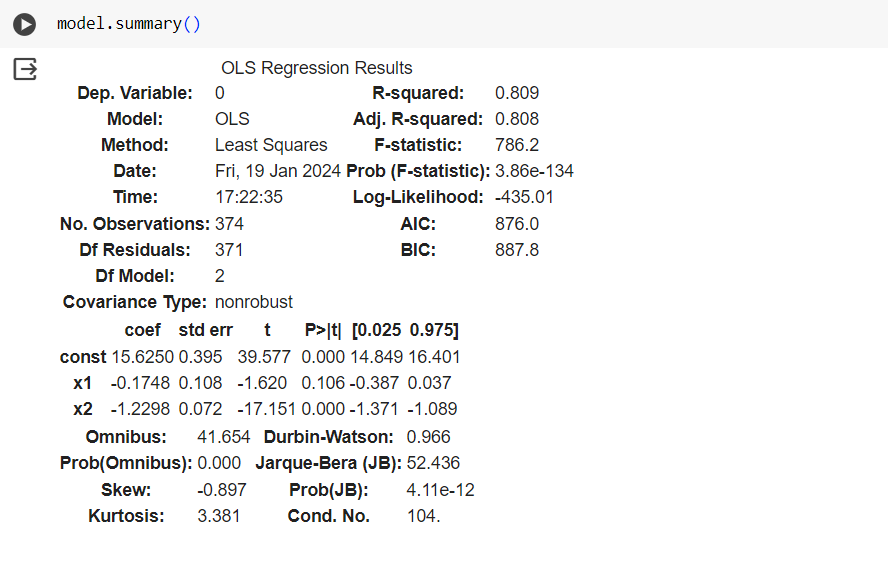
We created an Ordinary Least Squares (OLS) regression model. This type of plot is used to assess the performance of a regression model. The position on the Y-axis indicates the actual value and the position on the X-axis represents the value predicted by the regression model. The plot suggests that the predictions do not always align closely with the actual values and that the model is capable of variance in its predictions.







The number of observations is the size of our sample N=374. In our case, the independent/explanatory variables chosen are x1 = sleep duration, x2 = sleep quality and the dependent y is the stress level. From the table we can state the following: The R-squared value is 0.809, which means that approximately 80.9% of the variability in the dependent variable can be explained by the model. This is quite high, indicating a good fit. The coefficient for x1 is -0.1748 with a p-value of 0.106, which is bigger than 0.05 for statistical significance. This suggests that the relationship between x1 and the dependent variable is not statistically significant at the 5% level. The coefficient for x2 is -1.2298 and is highly significant with p-value= 0.000, indicating a strong relationship between x2 and the dependent variable for a statistically significant at the 5% level. The OLS regression model shows a good fit to the data with a high R-squared value. The variable x2 is significantly related to the dependent variable, while x1 is not significant at the 5% level.



**Discussion/Conclusion**

This study's findings underscore the significant impact of stress on sleep quality, highlighting the need for stress management interventions as part of a healthy lifestyle.

Despite the strong model fit, the lack of normality in the residuals suggests that future research should incorporate a broader range of variables to fully capture the determinants of sleep quality. The sleep duration as a predictor in the model raises critical questions about the prevailing assumptions regarding sleep quantity and quality. It suggests that interventions aimed at improving sleep health should focus more on the enhancement of sleep quality rather than merely increasing sleep duration. Overall, the pairplot indicates that sleep quality and physical activity level are generally positively correlated, with higher sleep quality being linked to higher levels of physical activity. Physical activity levels tend to rise in tandem with improved sleep quality. Based on the regression, as sleep duration increases, the model predicts an increase in the quality of sleep.

As sleep is a pillar of overall health, understanding and intervening in these relationships have profound implications for public health policy and individual well-being. Further research in this domain is essential to develop targeted strategies for the prevention and management of sleep disturbances in the adult working population. The limitations of the research are mainly on the limited access to data and the lack of previous numerous researches on this topic. The topic is a very complex one, which needs to be further investigated due to the fact that human nature, behavior and lifestyle is changing due to the rapid changes in the environment.

**References**

Dedhia, P., & Maurer, R. (2022). Sleep and Health— A Lifestyle Medicine Approach. *Supplement to The Journal of Family Practice*, *71*(1), 30-34. 10.12788/jfp.0295

Luyster, F. S., Strollo, P. J., Zee, P. C., & Walsh, J. K. (2012). Sleep: A Health Imperative. *Sleep*, *35*(6), 727-734. https://doi.org/10.5665/sleep.1846

National Sleep Foundation. (2020, October 1). *How Much Sleep Do You Really Need?* National Sleep Foundation. Retrieved January 19, 2024, from https://www.thensf.org/how-many-hours-of-sleep-do-you-really-need/

Perez-Pozuelo, I., Zhai, B., Palotti, J., Mall, R., Aupetit, M., Garcia-Gomez, J. M., Taheri, S., Guan, Y., & Fernandez-Luque, L. (2020). The future of sleep health: a data-driven revolution in sleep science and medicine. *npj Digit. Med*, *3*(42), 1-15. https://doi.org/10.1038/s41746-020-0244-4

Tharmalingam, L. (2023). *Sleep Health and Lifestyle Dataset*. Kaggle. Retrieved January 20, 2024, from https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset/data