UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

BSD2223
DATA SCIENCE PROGRAMMING II
2022/2023 SEMESTER II

GROUP PROJECT

SECTION : 01G

TITLE: SLEEP HEALTH AND LIFESTYLE ANALYSIS

PREPARED FOR: PROFESOR MADYA DR. ROSLINAZAIRIMAH BINTI ZAKARIA

GROUP MEMBERS:

NAME	STUDENT ID
NUR NABILA BINTI ABD RAHMAN	SD22037
SITI MAISARAH BINTI SUHARDI	SD22006
AMIRAH AISYAH BINTI ABDULLAH	SD22055
YU JIA ANN	SD22054

TABLE OF CONTENT

1.0 INTRODUCTION

The dataset in our project was obtained by collecting data from Kaggle. It comprises a collection of 374 participants to investigate the relationship between sleep health and various lifestyle factors. It includes comprehensive information on sleep quality, daily steps, physical activities level, stress levels, BMI categories, heart rates and prevalence of sleep disorders. These variables provide a comprehensive overview of each participant's lifestyle and health status.

The primary objective of this analysis is to understand the factors that contribute to healthy lifestyle and sleep patterns and to identify potential areas for improving overall health. By examining the data, we aim to uncover trends and correlations that can inform better sleep hygiene practices and lifestyle modifications. This analysis could be beneficial for healthcare providers, researchers, and individuals seeking to enhance their sleep quality and overall well-being.

Furthermore, this dataset allows us to explore how different demographic factors such as gender, age and occupation influence sleep health. By understanding these relationships, we can develop targeted strategies to address sleep-related issues across various population segments. This analysis provides a valuable resource for making informed decisions regarding sleep health interventions and lifestyle improvements, ultimately contributing to enhanced quality of life.

Last but not least, we aim our project will be provided to healthcare providers to develop targeted treatment plans and also preventive measures based on an individual's lifestyle factors. Researchers can further our study to explore the association between lifestyle factors and sleep health and be able to find out new discoveries about healthy lifestyle. On the other hand, public health officials also can design some useful campaigns to promote better sleep hygiene practices while employers can develop workplace wellness programs to address sleep health issues and reduce work-related stress. Furthermore, we wish that entrepreneurs are able to develop innovative products and services like personalized sleep coaching applications or stress management programs to meet specific needs by users while all of our users can apply insights that we do to improve their daily habits and routines to achieve better sleep health and maintain a healthy lifestyle.

2.0 PROJECT DESCRIPTION

This project proposes to develop a comprehensive analytical and visual dashboard that reveals the relationship between sleep health and various lifestyle factors by using Rshiny applications. In our dashboard, it will represent all of the insights about how different variables such as gender, age, occupation, sleep duration, sleep quality, physical activity level, stress level, BMI category, heart rate, daily steps and sleep disorders interact with each other. By leveraging data visualization and statistical techniques, our project aims to discover patterns and correlations that can inform better health and lifestyle decisions.

Our project was motivated by the growing awareness of the importance of sleep health in overall health. It can be explained like sleep and healthy lifestyle are very vital aspects of health that will bring effects to various physiological and physiological functions. However, a lot of people are getting unhealthy lifestyles, high stress levels, poor quality of sleep and suffer from sleep disorders. All of these situations are affected by lifestyle factors. Hence, by analyzing sleep health and lifestyle dataset and combining it with all lifestyle variables, our project seeks to provide some valuable insights to help individuals to improve their sleep quality and their overall health.

The significance of our project lies in its potential to improve public health by providing actionable insights into sleep and lifestyle. By simplifying the process of discovering how various factors interact with sleep health, our project will aim to enhance understanding of the impact of lifestyle on BMI category, stress level, quality of sleep and sleep disorders. Not only that, we aim to help individuals make informed decisions to improve their sleep and lifestyle.

The primary objective of this project is to develop an analytical and visual dashboard using RShiny applications to reveal the intricate relationship between sleep health and various lifestyle factors. Hence, we aim to achieve correlation analysis**,** which is to examine the relationship between multiple variables, including sleep duration, sleep quality, physical activity level, stress level, BMI category, heart rate, and daily steps.

3.0 DATA DESCRIPTION

The Sleep Health and Lifestyle Dataset comprises 400 rows and 13 columns, offering a comprehensive insight into various aspects of sleep and daily habits. It encompasses information on sleep patterns, lifestyle factors, cardiovascular health, and the prevalence of sleep disorders. The dataset includes demographic details such as gender and age, alongside occupation, sleep metrics, physical activity levels, stress ratings, BMI categories, and cardiovascular measurements. It contains 7 categorical data and 6 numerical data.

- **Categorical (Nominal) :** Person ID, Gender, Occupation, Sleep Disorder
- **Categorical (Ordinal) :** Quality of Sleep, Stress Level, BMI Category
- **Numerical (Discrete) :** Age, Daily Steps
- **Numerical (Continuous) :** Sleep Duration, Physical Activity Level, Blood Pressure

The dataset includes the following information as stated below:

Variable Name	Description
Person ID	Unique identifier for each individual
Gender	Male or Female
Age	Age of the person in years
Occupation	Profession or occupation of the person
Sleep Duration	Number of hours slept per day
Quality of Sleep (scale: 1-10)	Subjective rating of sleep quality from 1 (very poor) to 10 (excellent)
Physical Activity Level	Daily minutes of physical activity
Stress Level (scale: 1-10)	Subjective rating of stress level from 1 (very low) to 10 (very high)
BMI Category	Body Mass Index category (Normal, Overweight, Obese)
Blood Pressure (systolic/diastolic)	Systolic pressure over diastolic pressure
Daily Steps	Number of steps taken per day
Sleep Disorder	Presence or absence of sleep disorders categorized as None, Insomnia, or Sleep Apnea

4.0 DATA PREPARATION

Data gathering, combination, structure, and organizing are all steps in the process of preparing data for use in business intelligence (BI), analytics, and data visualization applications. Data scope , data cleansing, pre-processing and profiling are some of the preparatory steps to data. Often it involves integration of information from multiple Internal systems and external sources. At the same time, for making analysis in the mental health and lifestyle system, the data preparation category includes cleaning and preprocessing of data to make it with high quality.

- Define the Scope and Collect Data

This dataset focuses on defining exactly which aspect of mental health and the lifestyle we want to analyze. It involves the selection of the target variables which is health status for instance, mental health status represented by an aspect like sleep disorder, lifestyle parameters including sleep duration and physical activity levels, and participant characteristics including age and gender. Once defined, we collect the data from relevant sources like surveys, public datasets, or health records.

- Cleanse Data

All the data cleaning processes were cleaned in rnotebook with the respective codes.In order to maintain data quality one of the procedures performed included deduplication where obviously repeated rows were deleted. In addition to this one that was called for in the first sort in the computer programme there was a need to drop an ID variable that was a replica of the indexing column. This column was taken out so as not to complicate matters and confuse the reader by having the same information stated in other ways. The 'BMI_Category' column was actually inclusive of redundant categories which included 'Normal' and 'Normal Weight'. To make these categories more standard the field was converted from a column type 'objects' into character data and removing any white space and all the 'Normal 'were changed to 'Normal'.It was then recorded back to factor in order to clean the specified column.

Filtering Specific for all occupations and 'Scientist' and 'Manager' were excluded from further consideration and were deleted from the dataset. To maintain the accuracy during filtering, the feature in the 'Occupation' column was changed to be in character type, then it was converted back to a factor type. Descriptive statistics like the maximum stress values, minimum stress values, and the average values were computed to describe the stress distribution pattern in the study data set.

- Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) involves creating visualizations to understand data patterns and relationships. A histogram of age shows the distribution of the dataset. A box plot of BMI Category between daily steps. The scatter plot shows a negative correlation since as the quality of sleep increases, the stress level tends to decrease.

5.0 DATA ANALYSIS, RESULTS AND DISCUSSION

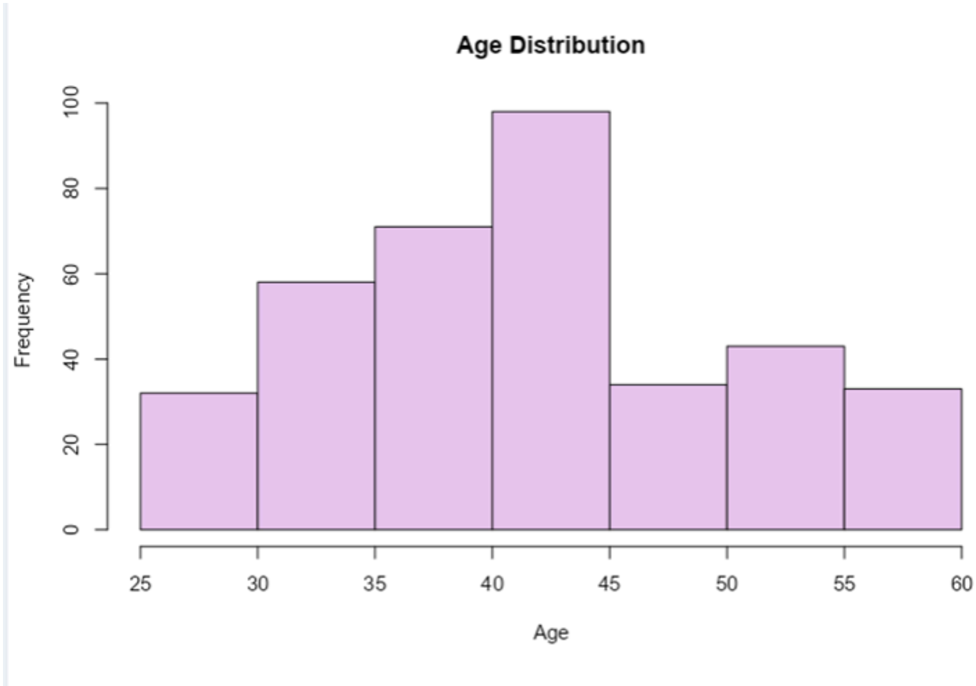


Figure 1 Age Distribution

Result:.

The plot shows at age 40-50 have a higher proportion of people in this age group compared to other ages. This also a minor right skew, that older people are more frequently questioned regarding mental health issues.

Discussion:

This result shows that it could be due to several factors. First of all, people in their 40s have midlife crises concerning career, family, or personal development, which may negatively influence mental health. Further, this group of people may develop a higher sensitivity to their health status making them to participate more in research on mental health, and lifestyle. Also, this age group may be of interest to researchers because it is relevant to mental health issues, which resulted in data collection bias. The appearance of the right skewing, although minor, suggests that older people are more often investigated or questioned regarding mental health issues, presumably due to life changes, medical check-ups, or inherent interest in learning and addressing issues most likely to prevail among this population. It is crucial, though, to admit that it provides findings based on current information and may warrant additional consideration for a definitive assessment.

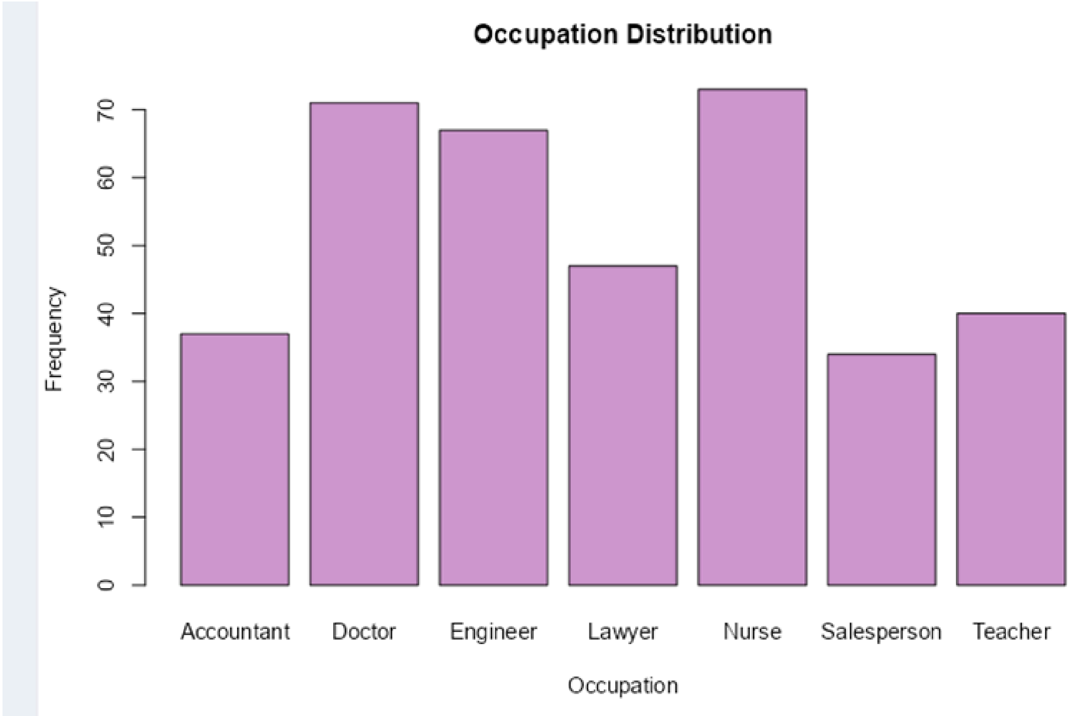


Figure 2 Occupation Distribution

Result:

One can identify that there are more positions, which refer to a specific occupation, such as “Doctor” and “Nurse” show the highest value that join this mental health system.

Discussion:

In the given dataset highest value is due to several reasons related to the peculiarities of health care employees. Doctors are important health care facilities and due to the demand of their services in the provision of medical services they are well represented in the dataset. Furthermore, being involved in medical research and data collection, they also help to affect the composition of the dataset. Nevertheless, doctors are also exposed to high levels of stress because of long working hours, mandatory caring for the patients, and the degree of decision-making, which influences their well-being and life in general. In the same vein, nurses featured prominently owing to their central status in the provision of care across various contexts. Through their working processes they have a lot of contact with patients and have to do shift work, which influences their lifestyle and mental state. Nurses feel physical demands, pressure and burnout more so because their working schedules infringe on their sleep-wake cycle. Altogether, the fact that doctors and nurses occupy the highest positions within the given dataset reveals the crucial and stressful nature of healthcare occupations, therefore calling for enhanced consideration of employees' mental well-being and lifestyles.

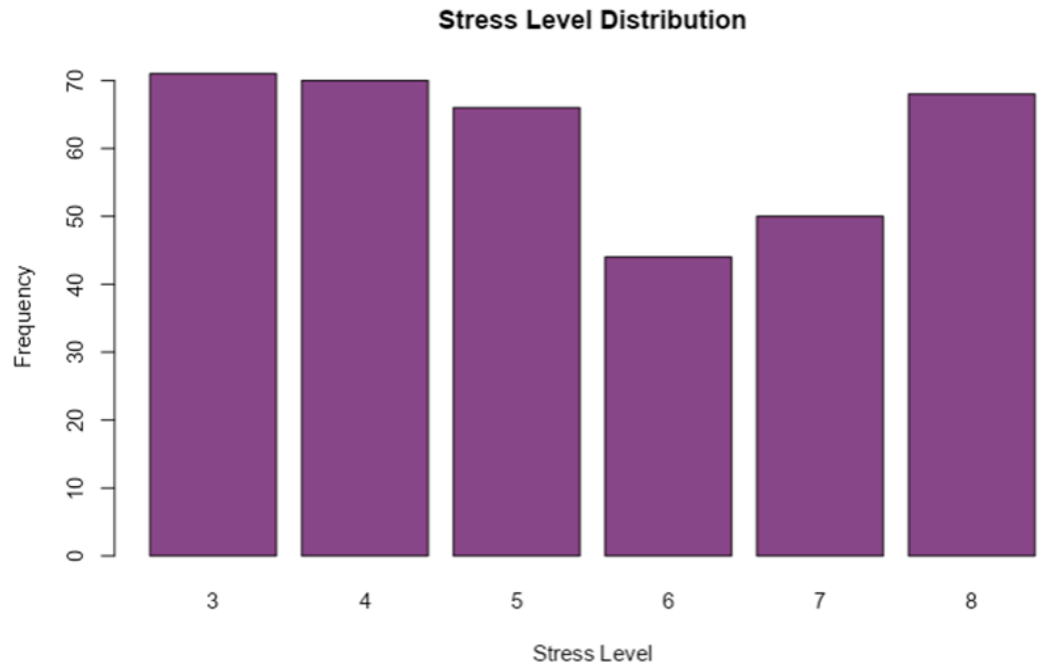


Figure 3 Stress Level Distribution

Result:

The skewness is a slight left skew suggested due to stress level by individuals. The bar chart show the high frequency of respondent indicating high stress level particularly at stress level 8.

Discussion:

From stress level, there is a high frequency of the persons responding high stress levels indicated by the spike in the bar chart at stress level 8 due to numerous factors. Firstly, work stress including work pressure due to challenging occupations, personal commitments, or other factors prevalent in life can bring about high stress levels. Nevertheless, change may also increase stress, where conditions such as loss or financial pressure extend the experience. Also, prevalence of chronic diseases is also a factor able to have an influence on mental health dispensing and contributing to stress levels. As for the stress level distribution, based on the graph below, one can say that creating an average graph without double-peak doesn't distort the picture much as the double top at stress levels 3 and 8 is evident Nevertheless, looking at the aspect of skewness one by one it is possible to say that there is a slight left-skew suggesting that the people in the graphic experienced different degrees of stress. Daily working hours and lack of leisure time, and improper relaxation and coping all add to the high-stress life while the low-stress life involves pursuing self-care, exercising and socializing. But now we know that chronic stress affects mental health negatively and results in anxiety, depression and burnout, thus making it very important to have measures such as meditation and therapy in managing stress in our day to day lives.

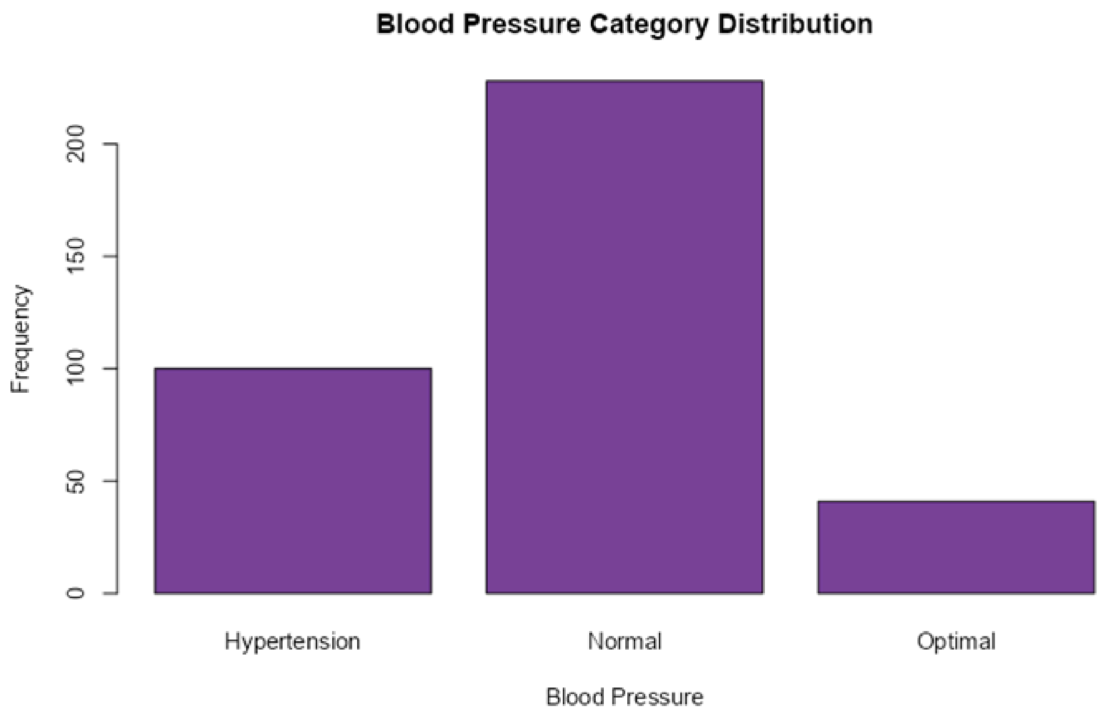


Figure 4 Blood Pressure Distribution

Result:

The bar that highlights the “Normal” blood pressure indicates that a large number of patients in the sample fall under this category, and the relative peak in the blood pressure of approximately 175 implies that many patients have normal blood pressure readings. The distribution is left skewed, suggesting the high and low blood presure are less frequent compare to high frequency of normal value.

Discussion:

This abundance may be blamed on the fact that patient awareness has improved over time, thereby making those with high blood pressure more vigilant. Other nonpharmacologic interventions on the prognosis of normal blood pressure include adoption of healthy diets, engaging in regular physical activities, and stress. However, other aspects such as medications and other health care services which involve doctors, nurses and other personnel play part in treatment of hypertension. Additionally, given that the distribution is somewhat left-skewed, meaning that extreme values of significantly high or significantly low blood pressures are less frequent than the high-frequency normal values, it can be inferred that there is still something which could be done in order to achieve better levels of blood pressure control. Stress-related practices, exercise regime, and food choice commitments, as well as the mental health activities, including calming techniques like mindfulness and muscle relaxation, have an impact on blood pressure. In addition, the availability of a healthcare system to attend occasional check-ups or consulting with a physician while the availability of awareness programs for the importance of a healthy diet and exercises helps to keep normal blood pressure within the population.

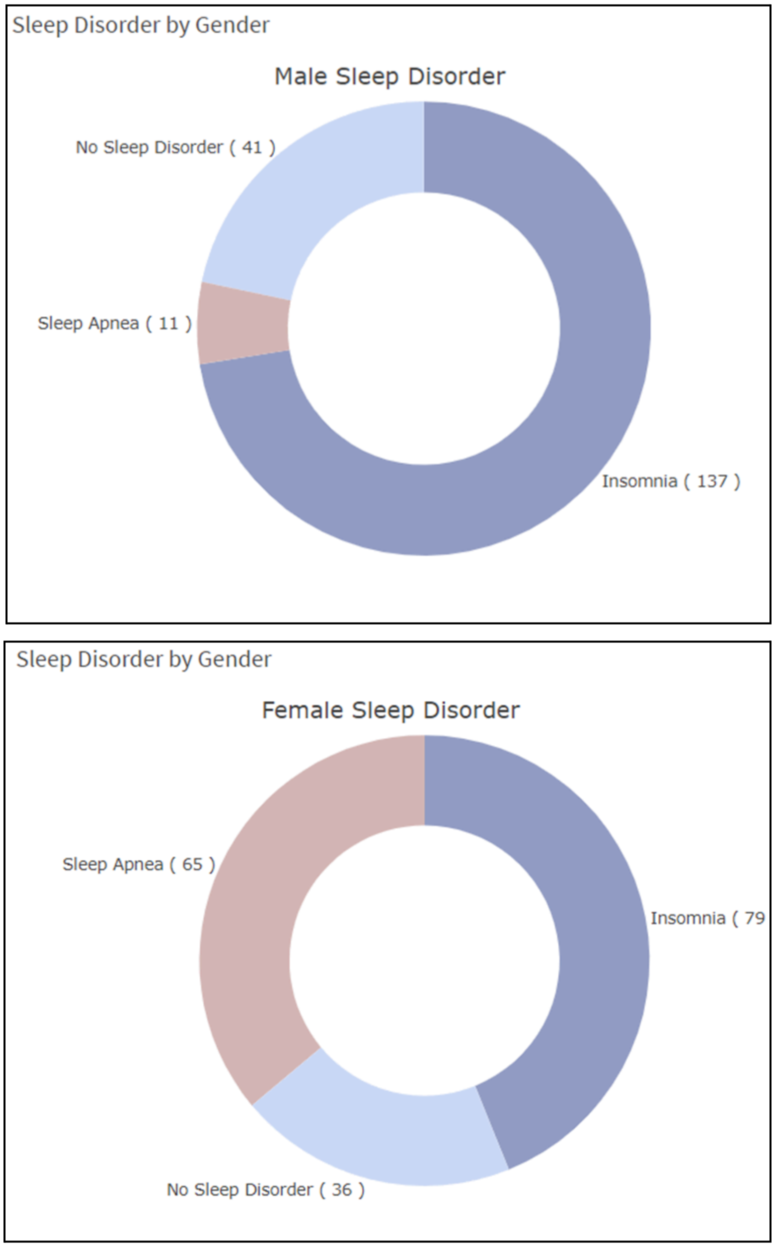


Figure 5 Sleep Disorder by Gender

Result: This donut chart represents the percentage distribution of sleep disorders by gender, categorized to males and females. Each slice of the donut chart corresponds to one of the three sleep disorder categories which is insomnia, sleep apnea and no sleep disorder. The size of each slice represents the proportion of individuals in that gender group who experience each type of sleep disorder. This visualization allows for a clear comparison of the prevalence of different sleep disorders in males and females, highlighting gender-specific patterns in sleep health.

Result and Discussion: The donut chart reveals significant differences in the prevalence of sleep disorders between males and females. Among females, 44.3% of them suffered from insomnia, 36.2% had sleep apnea, and 19.5% reported no sleep disorder. On the other hand, among males, 72.5% of males suffered from insomnia, 5.8% had sleep apnea, and 21.7% had no sleep disorder. These results suggest that insomnia is the most common sleep disorder in both sexes, but is particularly prevalent in males. Sleep apnea, on the other hand, is more common in females than in males. The distribution of sleep disorders between males and females shows that insomnia symptoms are the most common sleep disorder for both sexes, with a significantly higher prevalence in men (72.5%) than in women (44.3%). This significant difference suggests that men may be more susceptible to factors that contribute to insomnia, such as higher stress levels or lifestyle habits.

Furthermore, sleep apnea is more prevalent in females, with 36.2% of women having sleep apnea compared to 5.8% of males. The higher prevalence in females can be attributed to physiological differences, such as smaller larynx and airway size, which may increase the risk of airway obstruction. In addition, hormonal changes associated with the menstrual cycle, pregnancy, and menopause can affect the sleep breathing control center, increasing the risk of sleep apnea in females. Addressing these physiological and hormonal factors through a dedicated healthcare approach can help reduce the chances of sleep apnea in females.

A relatively small percentage of males and females reported no sleep disorders, with 19.5% of females and 21.7% of males saying they did not have any sleep disorders. This situation shows that the vast majority of males and females have sleep disorders, so this issue must be concerned and taken seriously by everyone because good quality of sleep and without sleep disorders can bring physical and mental health.

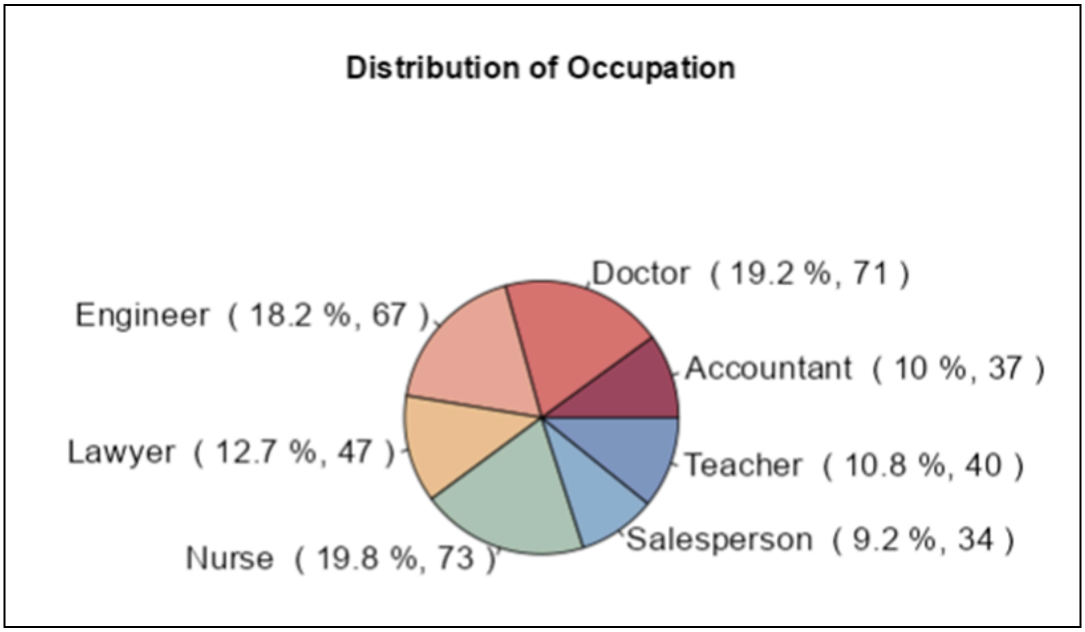


Figure 6 Distribution of Occupation

Result: This bar chart represents the distribution of different occupations. Our dataset includes 7 types of occupation which are "Accountant", "Doctor", "Engineer", "Nurse", "Salesperson" and "Teacher". Each slice of the pie chart corresponds to one of the occupation categories while the size of each slice represents the proportion of individuals in a certain occupation.

Result and Discussion: The pie chart reveals "Nurse" is the most with 73 people (19.8%) while "salesperson" is the least with 34 people (9.2%). The participation of individuals from various occupations in this survey is able to enhance the comprehensiveness of our results. The diverse occupational representation allows users to gain a clearer understanding of the health and well-being of professionals across different fields. The data highlights distinctions in healthcare needs and other aspects among various professions, providing a robust foundation for targeted health interventions and policy decisions. This diverse dataset not only enhances the credibility of our findings but also ensures that the conclusions drawn are relevant to a wide range of occupational groups, thereby offering valuable insights into the health status and needs of different professional sectors.

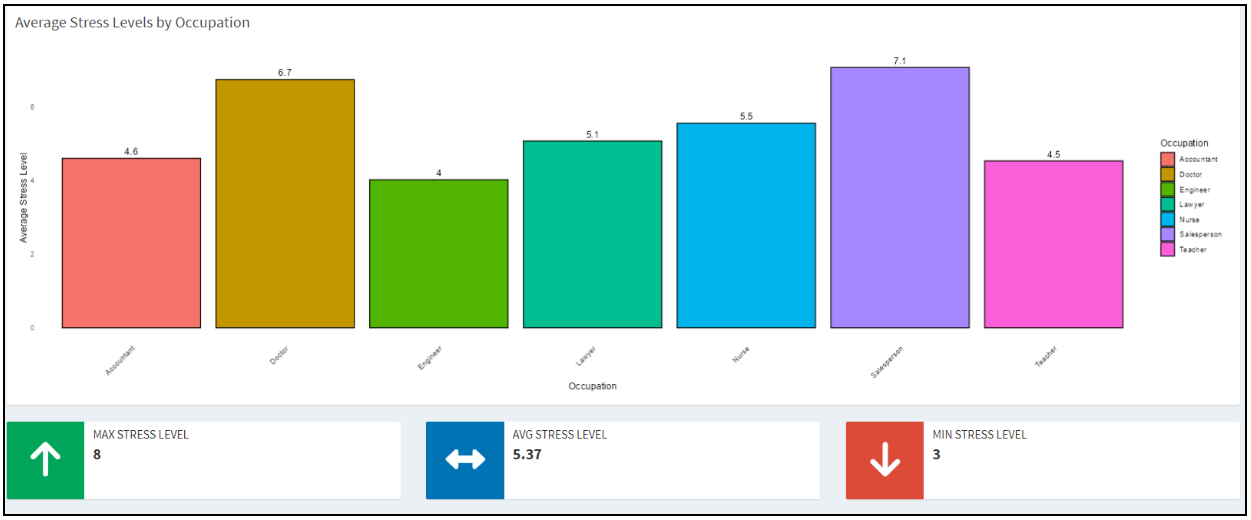
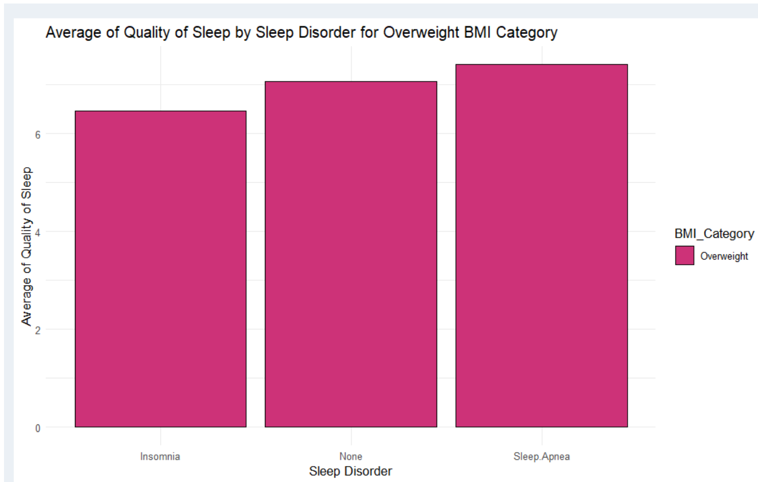
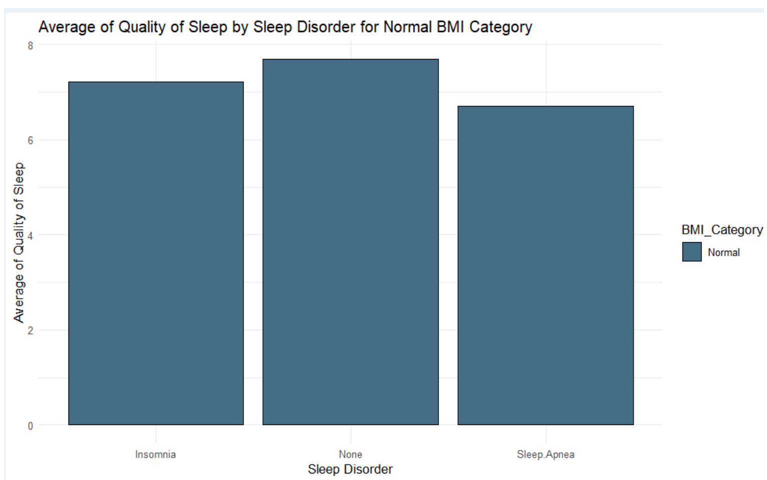


Figure 7 Average Stress Levels by Occupation

Result: This bar chart represents the average stress levels by different occupations. In our dataset, the stress level is divided into 10 levels which are from 1 until 10 where level 1 represents the least stress level and level 10 shows the highest stress level. Not only that, there exist 7 types of occupations which are “Accountant”, “Doctor”, “Engineer”, “Nurse”, “Salesperson” and “Teacher”. This visualization allows for a clear comparison of the prevalence of different average stress levels by different occupations, highlighting occupation-specific patterns in mental health.

Discussion: The bar chart reveals highest average stress levels is represented by “Salesperson” which is 7.1 followed by “Doctor” with average stress levels of 6.7. “Engineer” in our dataset shows the lowest average stress level which is 4.0. Below the bar chart, three info boxes also represent maximum stress levels of investigators in our dataset is about 8.0, minimum stress level is about 3.0 while the overall stress level of investigators is about 5.37. The “Salesperson” has the highest stress level due to their working environments. For example, salesperson always need to work in environments with high-performance targets and commissions. Hence, the pressure to reach or exceed the Key Performance Indicator (KPI) may become a significant source of stress. On the other hand, doctors also represent the second high stress level because they need to carry immense responsibility for patient health and lives, they need to work long hours for their surgeries and these kinds of working hours lead to fatigue and high levels of stress. Furthermore, engineers have the lowest stress level due to their structured and predictable work environment, their jobs also involve less direct interaction with clients or customers. Hence, these factors will reduce potential stress from dealing with external pressures.





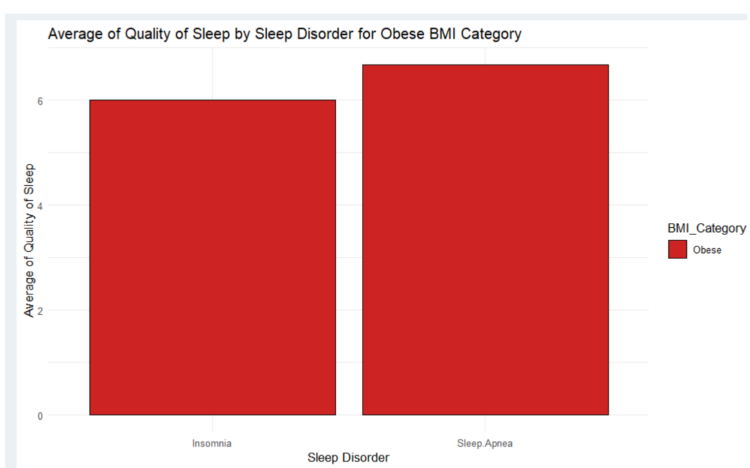


Figure 8 Average of Quality of Sleep by Sleep Disorder for BMI category

Analysis of data: These bar charts represent the average quality of sleep by sleep disorder for different BMI categories (Normal, Overweight, and Obese). The quality of sleep is measured on a scale from 1 to 10, where 1 is the worst and 10 is the best. The charts provide a visual comparison of the impact of different sleep disorders on sleep quality across various BMI groups.

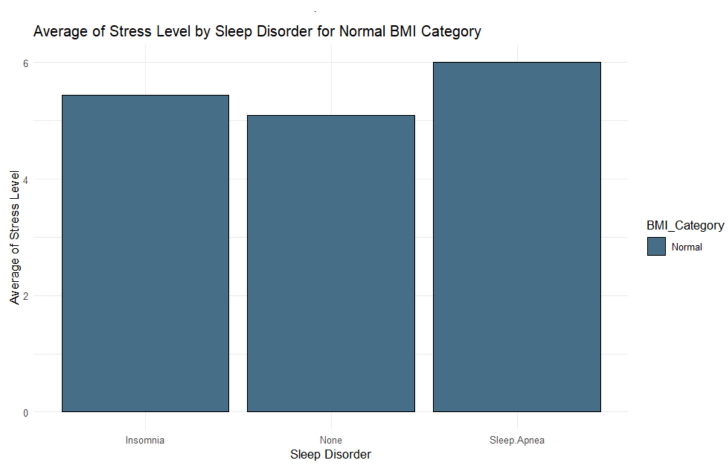
Result and Discussion:

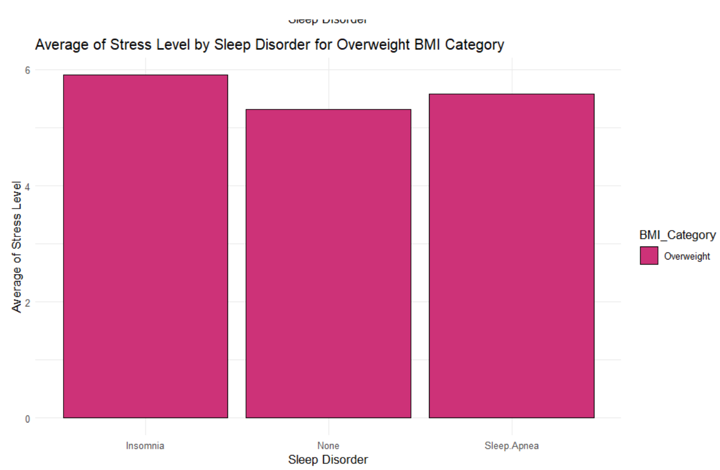
The first graph shows the average quality of sleep by sleep disorder for the normal BMI category. Individuals with normal BMI who have insomnia show an average sleep quality of around 7. Those without any sleep disorder have a slightly higher average sleep quality, approaching 8. In contrast, those with sleep apnea have an average sleep quality slightly above 6, which is lower than those with insomnia and no sleep disorder. This indicates that in the normal BMI category, individuals without any sleep disorder have the highest sleep quality, followed by those with insomnia, and those with sleep apnea having the lowest quality of sleep among the three groups.

The second graph shows the average quality of sleep by sleep disorder for the overweight BMI category. Overweight individuals with insomnia have an average sleep quality of around 6. For those without any sleep disorder, the average sleep quality is slightly higher than 6, close to 7. Overweight individuals with sleep apnea have an average sleep quality of around 7, similar to those with no sleep disorder. This suggests that in the overweight category, individuals with insomnia have slightly lower sleep quality compared to those with no sleep disorder or sleep apnea. Interestingly, sleep apnea does not seem to significantly reduce sleep quality for this group compared to those with no sleep disorder.

The last graph shows the average quality of sleep by sleep disorder for the obese BMI category. Obese individuals with insomnia have an average sleep quality of around 6. Similarly, those with sleep apnea also have an average sleep quality of around 6. Noticeably, none of the obese individuals are without a sleep disorder, indicating that sleep disorders are prevalent in this group. For the obese category, both insomnia and sleep apnea result in a similar average sleep quality, which is lower than the sleep quality observed in the normal and overweight categories. This indicates that obesity, regardless of the type of sleep disorder, might be associated with poorer sleep quality.

The analysis of these bar charts reveals distinct patterns in the sum of quality of sleep scores across different BMI categories and sleep disorders. Individuals with normal BMI generally report the highest sleep quality, especially those without any sleep disorder. Sleep apnea tends to reduce sleep quality more significantly in individuals with normal BMI compared to those who are overweight or obese. Both insomnia and sleep apnea have a noticeable impact on sleep quality across all BMI categories, but the extent of this impact varies. This analysis suggests a complex relationship between BMI, sleep disorders, and sleep quality, highlighting the need for tailored interventions to improve sleep health across different BMI groups.





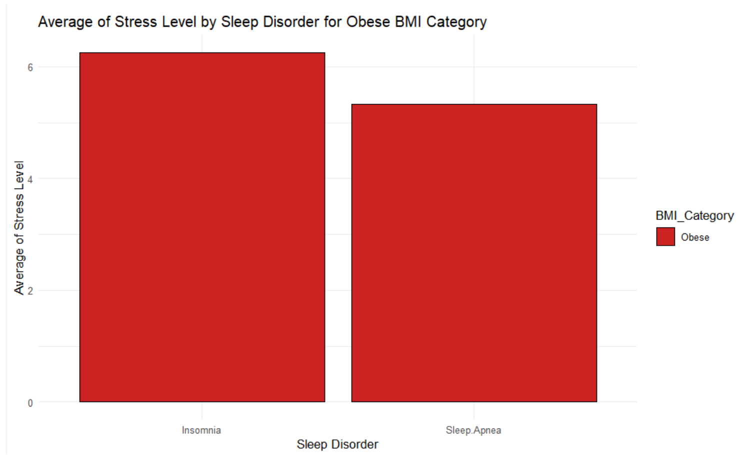


Figure 9 Sum of Stress Level by Sleep Disorder for BMI category

Result: This set of bar charts represents the average stress levels associated with different sleep disorders across various BMI categories, including normal, overweight, and obese. Stress levels are measured on a scale from 1 to 10, where 1 represents the lowest stress and 10 the highest stress. These visualizations allow for a clear comparison of stress levels within different BMI categories, highlighting the impact of various sleep disorders on stress across these groups.

Result and Discussion:

The first chart, depicting individuals with a normal BMI, reveals that those suffering from sleep apnea experience the highest average stress levels, with an average around 6. Insomnia sufferers follow closely, with an average stress level of approximately 5.5. Individuals without any sleep disorders report the lowest stress levels, averaging around 5. This indicates that among individuals with a normal BMI, sleep disorders such as sleep apnea and insomnia are associated with higher stress levels compared to those without any sleep disorder.

In the second chart, which focuses on the overweight BMI category, it is evident that individuals with insomnia report the highest average stress level, around 6. Those with sleep apnea and those without sleep disorders have similar stress levels, with the former slightly lower. Specifically, the average stress level for individuals with sleep apnea is around 5.5, while for those without any sleep disorder it is around 5. This suggests that in the overweight BMI category, insomnia is a significant stress factor, whereas sleep apnea and the absence of sleep disorders have a relatively similar impact on stress levels.

The third chart presents data for the obese BMI category, showing that individuals with insomnia have the highest average stress levels, around 6. In contrast, those with sleep apnea report a lower average stress level, around 5. Notably, data for individuals without sleep disorders in this BMI category is not provided, indicating a focus on those with diagnosed sleep disorders. This pattern suggests that, among obese individuals, insomnia is a more significant contributor to higher stress levels than sleep apnea.

In conclusion, these charts highlight the significant impact of sleep disorders on stress levels across different BMI categories. Insomnia consistently results in the highest stress levels in both the overweight and obese categories. For individuals with a normal BMI, sleep apnea is associated with the highest stress. These findings underscore the importance of addressing sleep disorders, particularly in overweight and obese populations, to manage and potentially reduce stress levels. The data emphasizes the need for targeted interventions aimed at improving sleep quality to enhance overall well-being across different BMI groups.

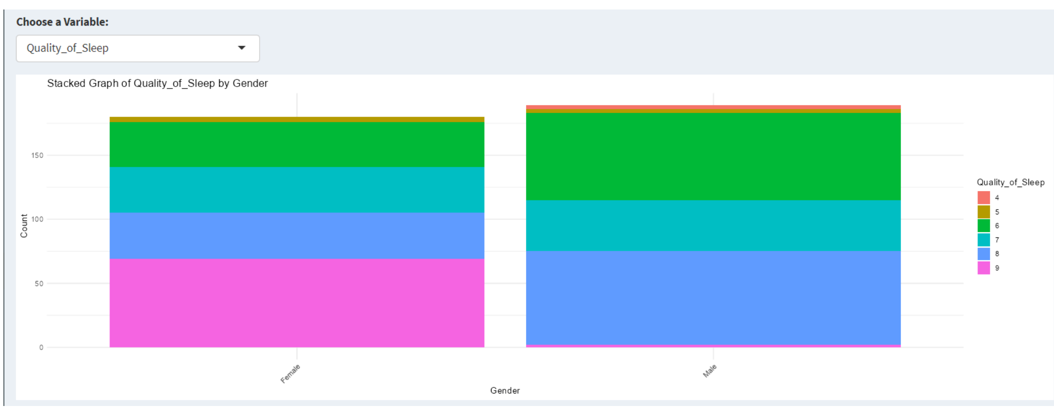


Figure 10 Quality of Sleep by Gender

Result:

The graph shows the quality of sleep of each gender with pink segment is the maximum quality of sleep measured as 9. Overall indicating that good quality sleep is less common. Most people fall into the mid range of sleep quality which is 6 and 7 , where 7 is the most common in blue color. Smaller segment is the lower sleep quality in level 5.

Discussion:

While there are few cases of both genders having excellent sleep, this distribution shows that these are rare, with the majority of people having mediocre sleep. The lifestyle factor of potential causes for mediocre sleep quality is stress from work, personal life and commitments that contribute to the sleep quality of most people.This is also because use of electric devices before bedtime and lack of physical activity might be the factor affecting sleep quality. Additionally, the details of gender difference have to consider how societal roles and responsibilities might be different for men and women. For example women might experience sleep disturb due to caregiving and multitasking demands for family and work.

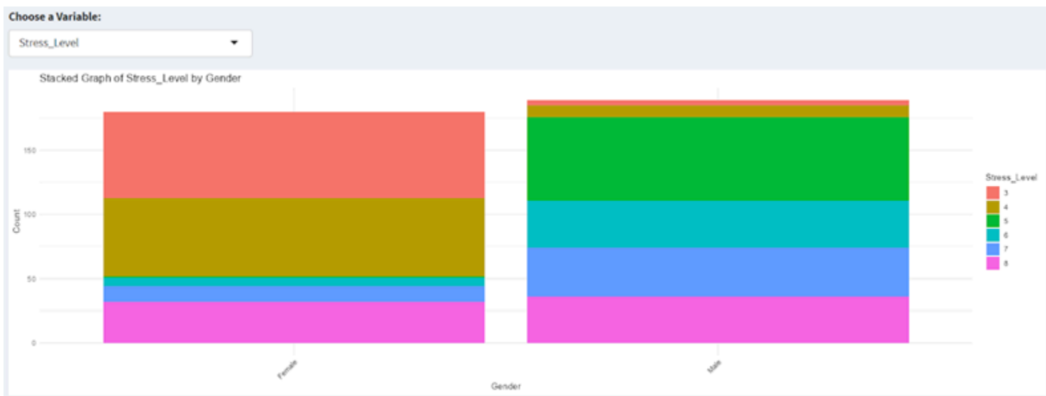


Figure 11 Stress Level by Gender

Result:

The provided graph distribution of stress levels across different genders, with the x-axis representing gender and the y-axis representing the count of individuals. The colors within the bars difference various stress levels, as indicated by the legend levels 3, 4, 5, 6, 7, and 8 .In figure 5.8, the highest level is level 4, the biggest section of the bar in brown in the female bar, showing that many of them have a stress level of four.

Discussion:

One possible reason is the stress between family responsibilities and the demands of their jobs as women are usually expected to perform highly in their professional pursuit and at the same time provide care in their homes. Social expectations and gender roles may also involve women in caregiving roles, which also lead to stress, especially when one has to handle both professional and household roles. For males, the highest level is level 5, the biggest section of the bar in green, showing that many of them tend to report a stress level of five. Most men tend to be stressed due to pressure at work, expectations related to career advancement, and the need to provide for their families. Being the major breadwinner may attract more stress, particularly in an environment when the job market requires workers to be highly productive and competitive.

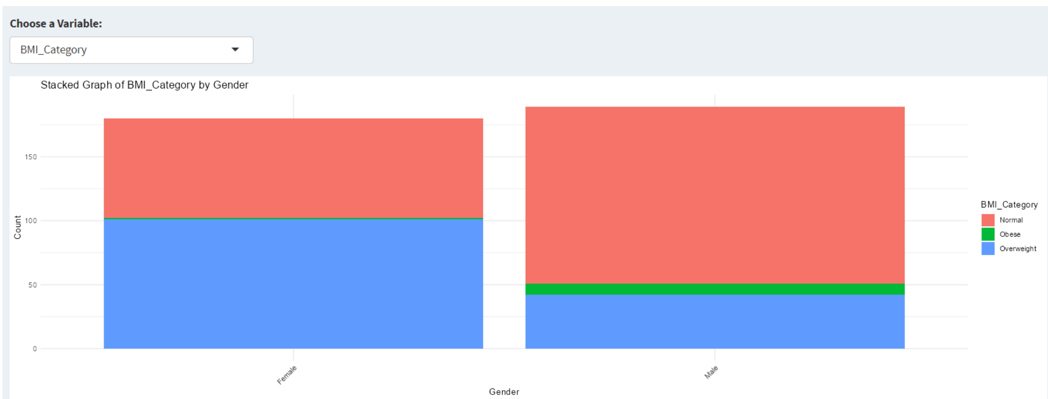


Figure 12 BMI Category by Gender

Result:

The BMI range for female is too short with the highest class being “Overweight,” this implies that many females in the data set are overweight. In the case of BMI distribution with the current population data, the highest prevalence of the BMI category for males is also “Normal” , meaning that a significant portion of males have a normal BMI.

Discussion:

There are several variables that help explain the indicated BMI categories. As for food and nutrition, females may be more likely to gain weight because of a high-calorie diet or a choice between accessible unhealthy foods/café meals and less healthy home-cooked meals; culturally specific food choices. On the other hand, males might reach a normal weight because they control their food intake, eating the right foods and being conscious of their nutrition. Physical activity the part of females may have lower inclination or less time or inadequate facilities or caring for families while on the other hand males engage in regular workouts deemed productive for normal BMI.

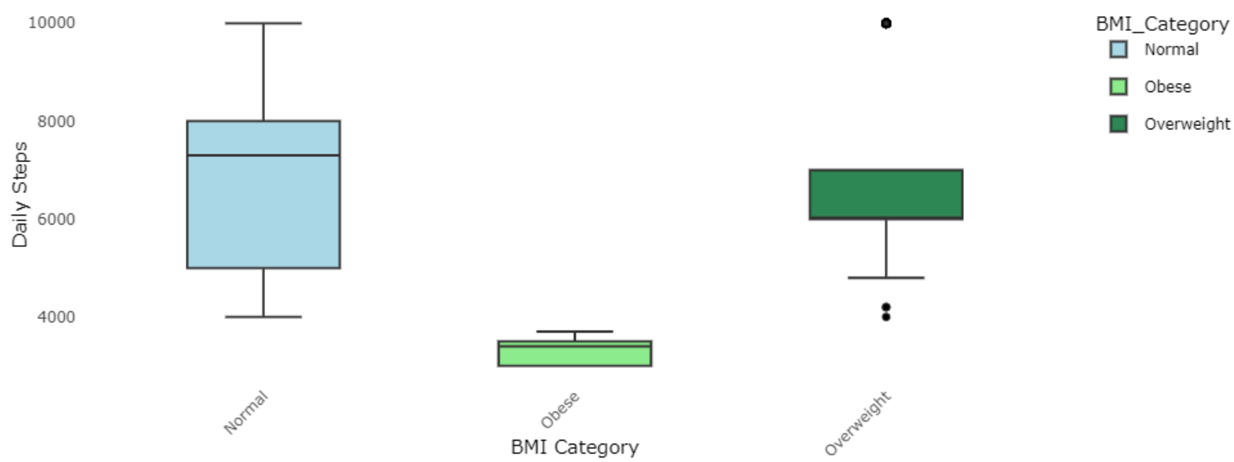


Figure 13 Daily Step by BMI Category

Result:

The Normal BMI category shows that the boxplot is left-skewed, and 50% of individuals with normal BMI have daily step counts less than or equal to 7300 steps. The IQR (Q3 - Q1) = 8000 - 5000 = 3000 steps. The absence of outliers indicates no extreme values beyond the upper or lower whiskers.

The Obese BMI category shows that the boxplot is left-skewed, and 50% of individuals with obese BMI have daily step counts less than or equal to 3400 steps. The IQR (Q3 - Q1) = 3500 - 3400 = 100 steps. The absence of outliers indicates no extreme values beyond the upper or lower whiskers.

The Overweight BMI category shows that the boxplot is right-skewed, and 50% of individuals with normal BMI have daily step counts less than or equal to 6000 steps. The lower fence is 4800 steps, indicating the lower limit beyond which data points are considered outliers. The outliers values, Max = 10000 steps, and Min = 4000 steps indicate that these values lie beyond the lower and upper fences and are considered outliers.

Discussion:

The left-skewed Normal BMI boxplot indicates that the distribution of daily steps within the normal BMI category is skewed toward higher step counts. Individuals with a normal BMI tend to have higher daily step counts. The boxplot shows that they engage in more physical activity, which helps them maintain their BMI. The left-skewed Obese boxplot suggests that a higher proportion of individuals within the obese BMI category have lower daily step counts compared to those with higher step counts. Since the boxplot is right-skewed, overweight people have the lowest step counts. The lower fence (4800 steps) gives an indication of the lower boundary of the data, and the outliers (4000 and 10000 steps) suggest the spread of extreme values. Outliers indicate that some individuals engage in significantly higher physical activity levels while others stay moderate. In summary, maintaining higher daily steps improves your health.

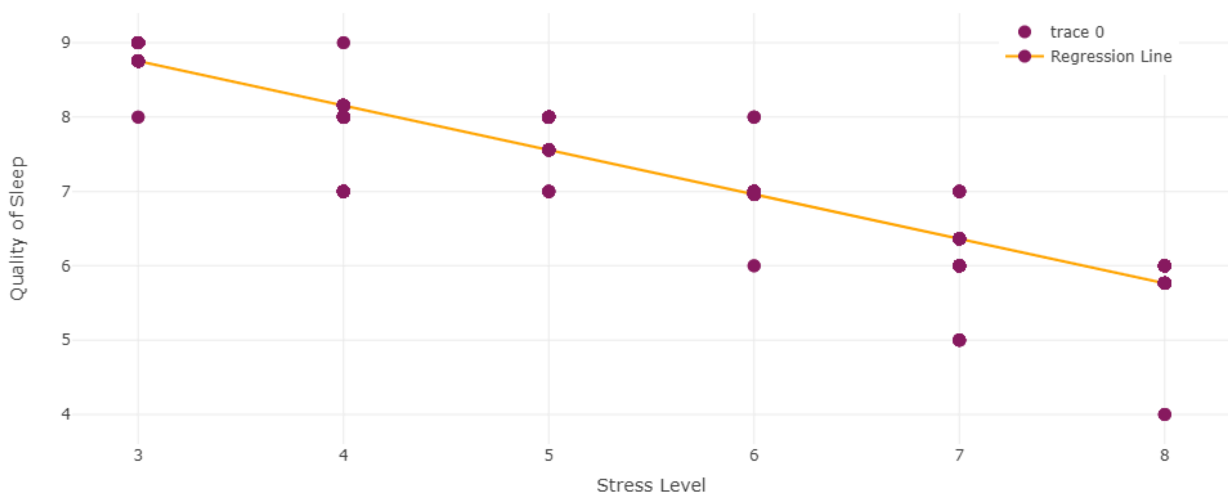


Figure 14 Stress Level vs Quality of Sleep

Result:

The scatter plot shows a negative correlation since as the stress level increases, sleep quality tends to decrease.

Discussion:

This negative correlation shows that lower stress levels are associated with better sleep quality. Individuals who experience lower levels of stress tend to report higher-quality sleep, which highlights the potential impact of stress levels on sleep quality. Although there is variability among individual points of view, the correlation pattern supports the idea that improvements in reductions in stress levels are linked with sleep quality.

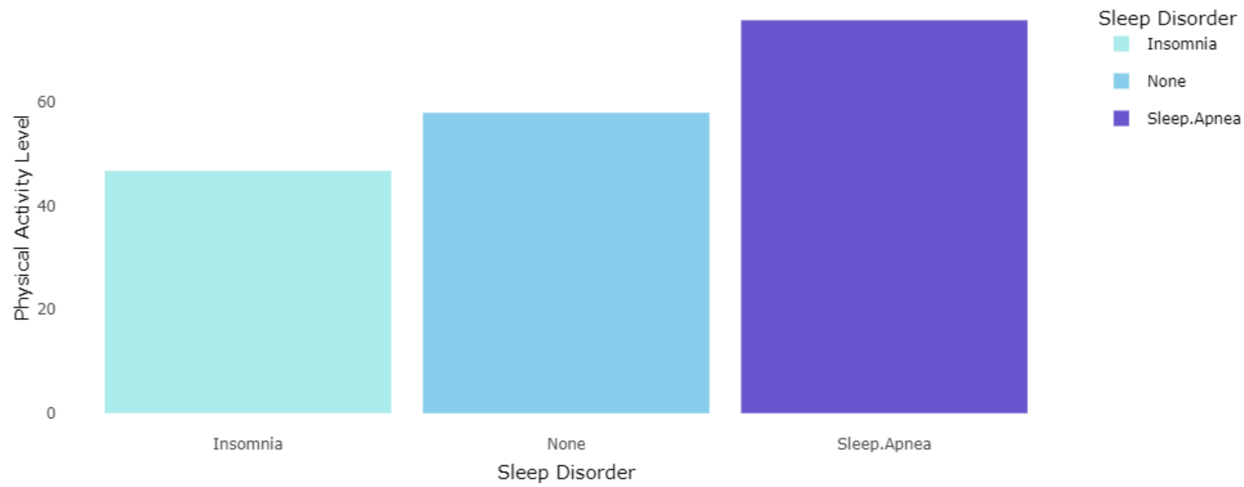


Figure 15 Average of Physical Activities Level by Sleep Disorder

Result:

The bar plot shows that insomniacs have the lowest average physical activity level, while sleep apnea sufferers have the highest.

Discussion:

On average, individuals with insomnia have a lower physical activity level compared to those without sleep disorders. Insomnia occurs when individuals have difficulty falling or staying asleep. It can lead to fatigue and reduced energy levels, which may result in lower physical activity levels.

Individuals without sleep disorders may experience better sleep quality and, consequently, higher energy levels, which could lead to increased physical activity participation.

Sleep apnea often leads to fragmented sleep and daytime sleepiness, which may affect physical activity participation. However, some individuals with sleep apnea may engage in higher physical activity levels to improve overall health and manage symptoms.

Individual lifestyle factors, health conditions, and personal preferences can influence sleep quality and physical activity levels. Promoting physical activity and addressing sleep disorders through lifestyle modifications, behavioral interventions, and medical treatments can contribute to overall well-being and quality of life.

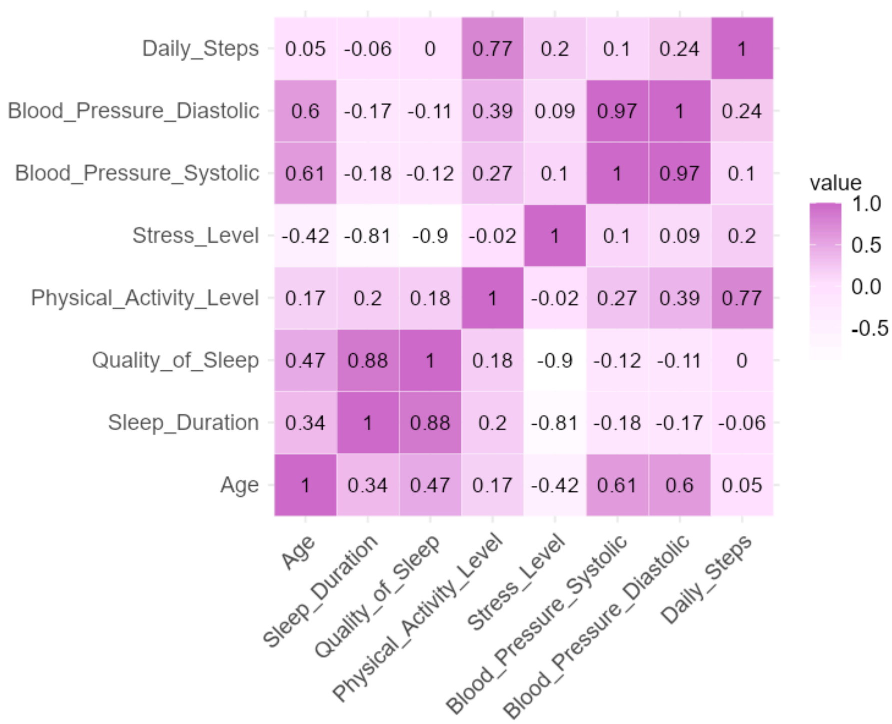


Figure 16 Heatmap

Result:

There was a strong negative correlation (-0.90) between stress level and sleep quality.

There was a strong positive correlation (0.88) between the sleep duration and quality of sleep.

Discussion:

Higher stress levels are strongly associated with poorer sleep quality. Stress can significantly disrupt sleep patterns thus, managing stress through relaxation techniques, mindfulness, or physical activity could improve sleep quality. Longer sleep duration is strongly associated with better sleep quality, which is expected as getting adequate sleep often leads to feeling more rested and having better sleep quality.

6.0 CONCLUSION

In conclusion, the analysis of sleep health and lifestyle dataset provide meaningful insights of the various lifestyle factors and sleep health. By using the RShiny, for data visualization and statistical analysis, we achieved a deeper understanding of these relationships. Our analysis showed that different BMI, gender, occupation affect their sleep quality, stress level, sleep disorder, daily steps and physical activity level. Every variable differently affects sleep health and lifestyle.

The insights from this project can be used by various stakeholders. Healthcare providers can therefore develop treatment plans and preventive measures on very specific lifestyle factors and individual demographic characteristics. The association between lifestyle factors and sleep health is therefore another area of study that researchers can continuously do leading to new and more specific discovery. Public health campaigns may therefore be specifically designed in ways that will encourage better sleep health practices. Employers and occupational health experts can develop workplace wellness programs to address sleep health and reduce work-related stress. Entrepreneurs can create innovative products and services, such as personalized sleep coaching apps and stress management programs, that could directly impact the needs highlighted by the research and could be implemented in everyday life by individuals to improve day to day habits or routine for their sleep health.

Overall, this project provides a resource for informed decision-making regarding sleep health interventions, contributing to improved public health and opening new avenues for entrepreneurial ventures aimed at enhancing sleep and lifestyle quality.

7.0 LIMITATIONS OF THE STUDY

1. **Data Bias**

   The dataset includes many subjective and self-reported variables, such as sleep quality, stress level, and physical activity. The questionnaire dataset may introduce the risk of bias and inaccuracies due to individual perception and may have the potential of misreporting. Consequently, the reliability of the variables in this dataset might be compromised, impacting the accuracy of the insights and correlation from the analysis.

2. **Data Limitation**

   The dataset is cross-sectional and captures data at a single point in time for each person. This limits the ability to causality or observe changes and trends over time. Longitudinal data would provide a more robust foundation for understanding the dynamic relationships between sleep health and lifestyle factors, enabling the study of temporal patterns and causative effects.

3. **Limited Generalizability**

   The dataset has 400 samples, which might differ from the broader population. The sample size and demographic characteristics, like age distribution and occupation types, could limit the generalizability of the findings. Besides, this dataset does not observe specific lifestyle or cultural factors that might influence sleep and health, so the results may only apply to certain populations or geographical regions.

8.0 APPENDIX

Dataset:

https://drive.google.com/file/d/1plEiu0ddDa5HbMBRSrY4zfT_iBIG5-2S/view?usp=drive_link

Codes:

https://drive.google.com/file/d/1ZCCqCQyMFwh15Op4ZICVaIICOuq71-20/view?usp=drive_link

GUI:

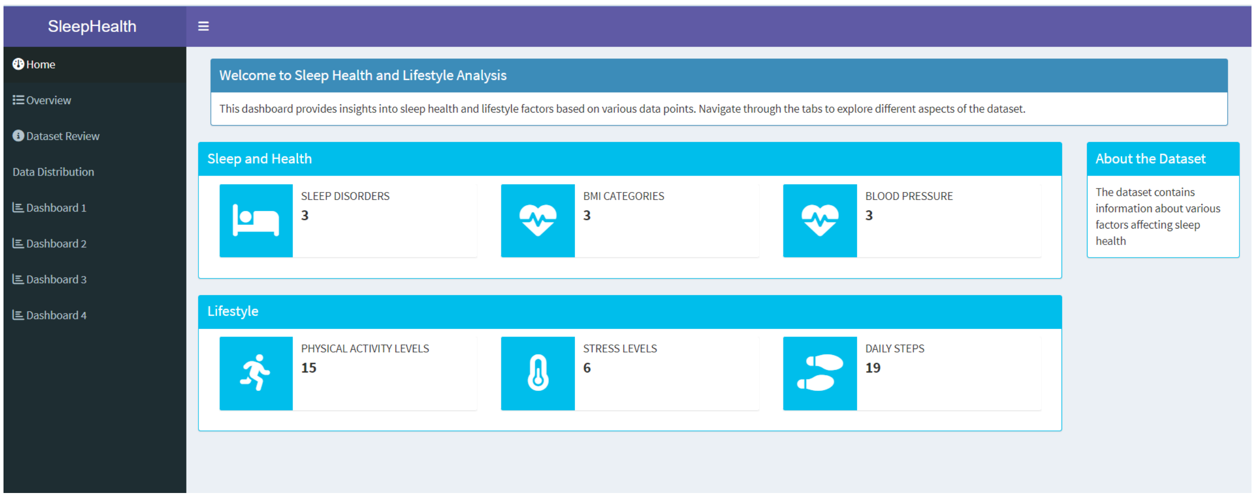


Figure 17 Home page of Dashboard

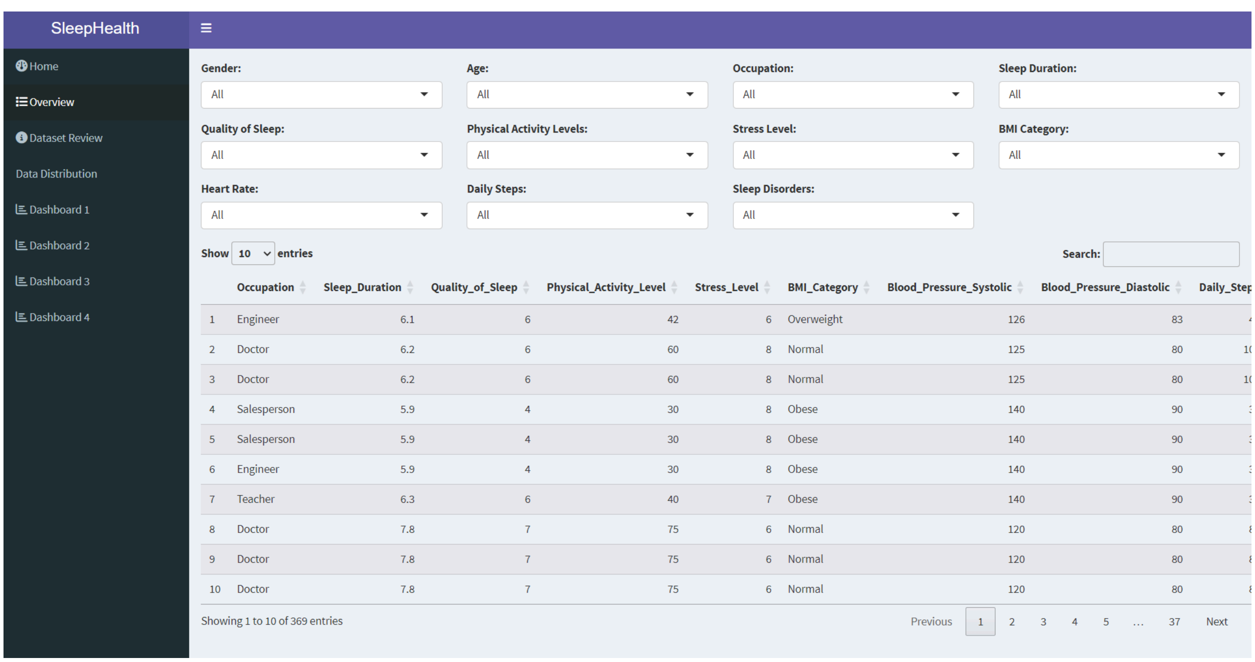


Figure 18 Overview of Dashboard

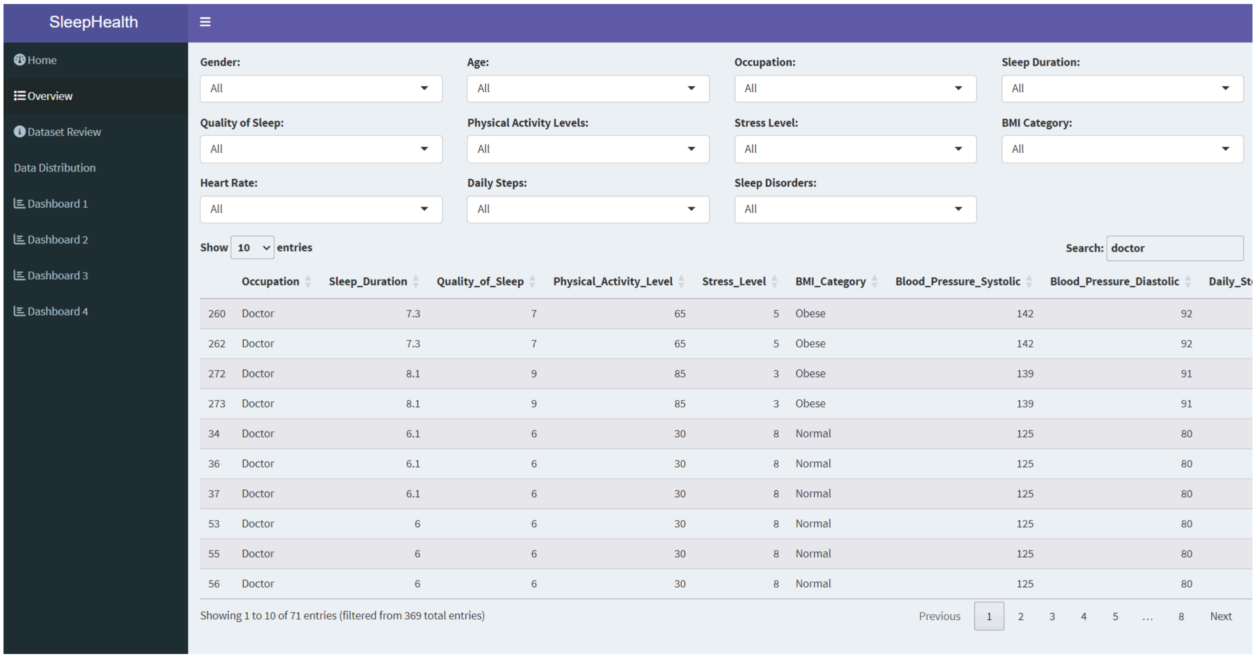


Figure 19 Search button to find data

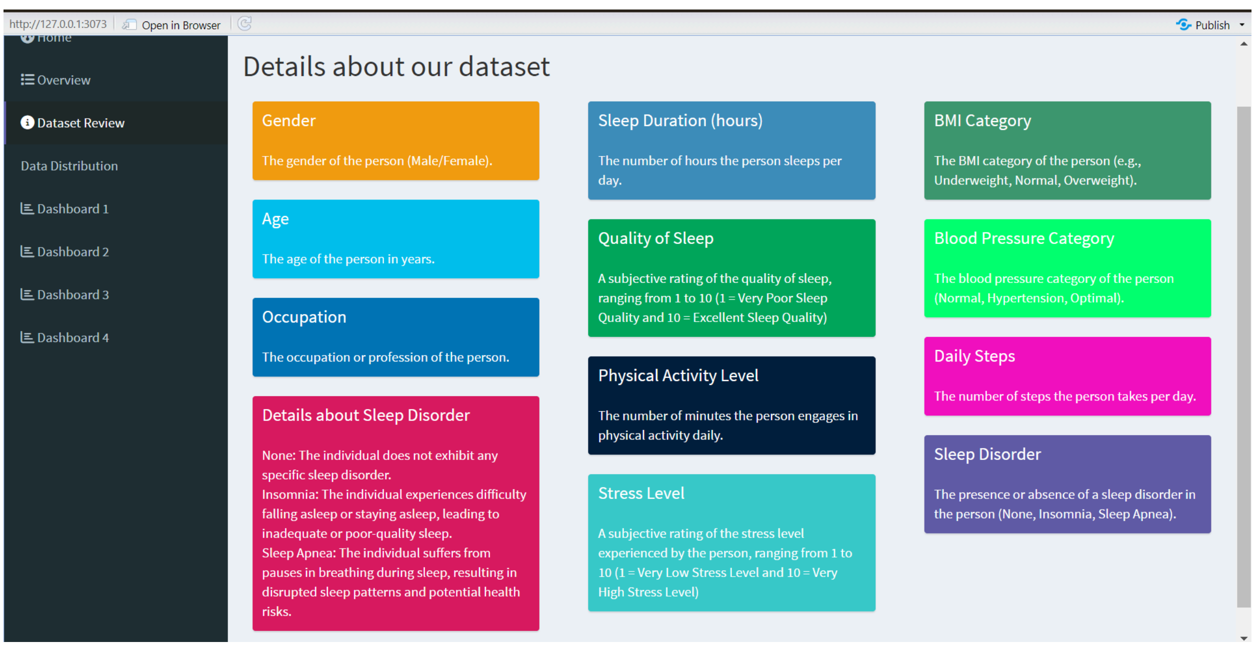


Figure 20 Dashboard of the detail about our dataset variable

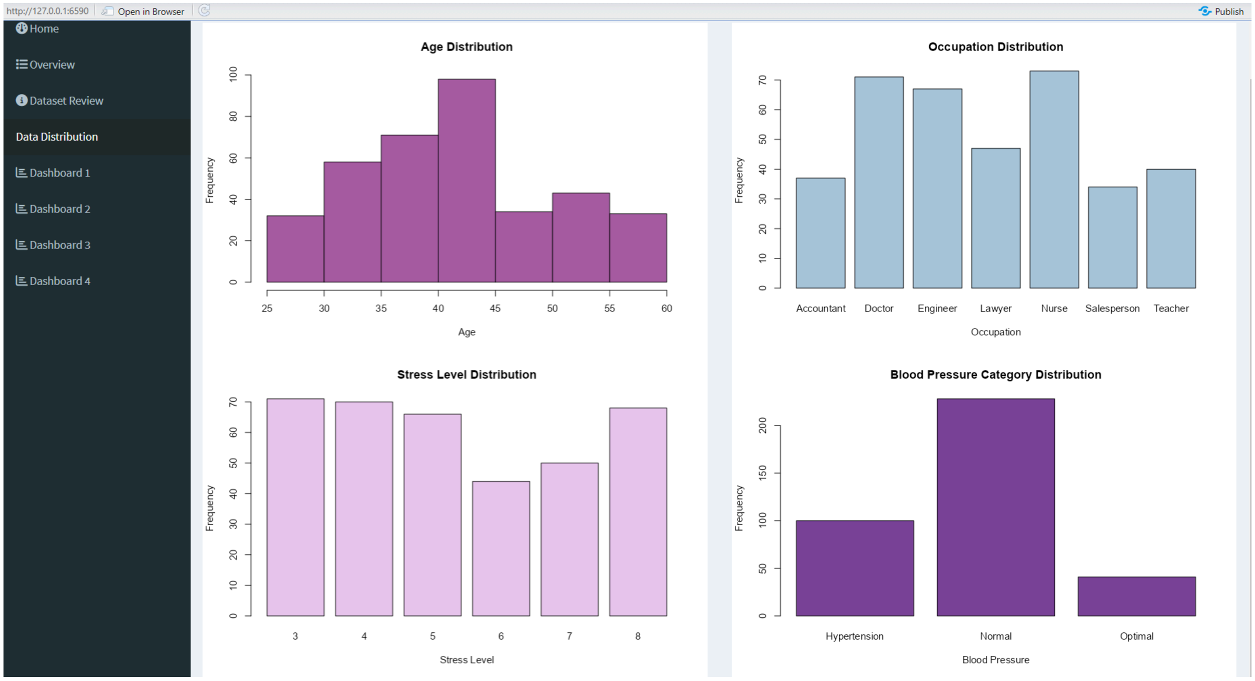


Figure 21 Dashboard Data Distribution show some of variable Distribution

Figure 22 Dashboard 1 Bar Chart of Quality Sleep By sleep Disorder by BMI Category

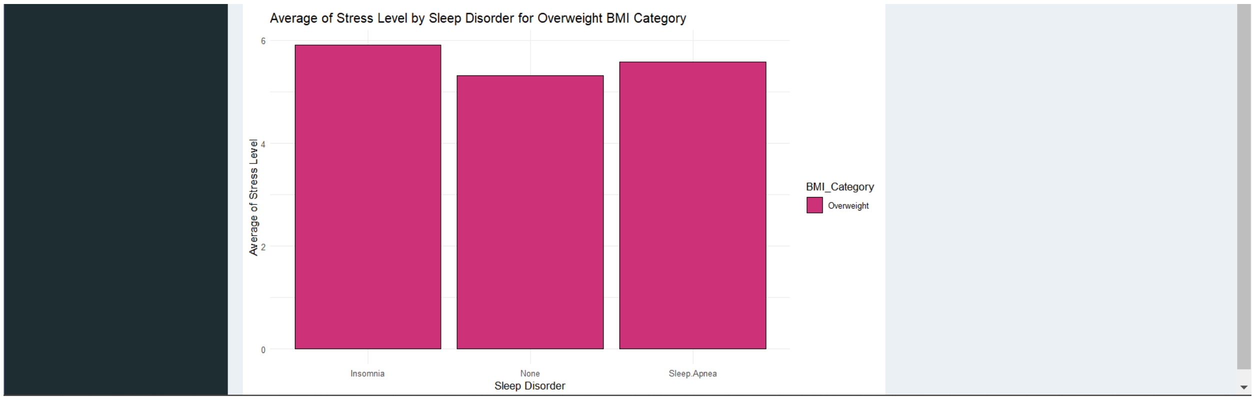


Figure 23 Dashboard 1 Bar Chart of Stress Lvel by Sleep Disorder by BMI Category

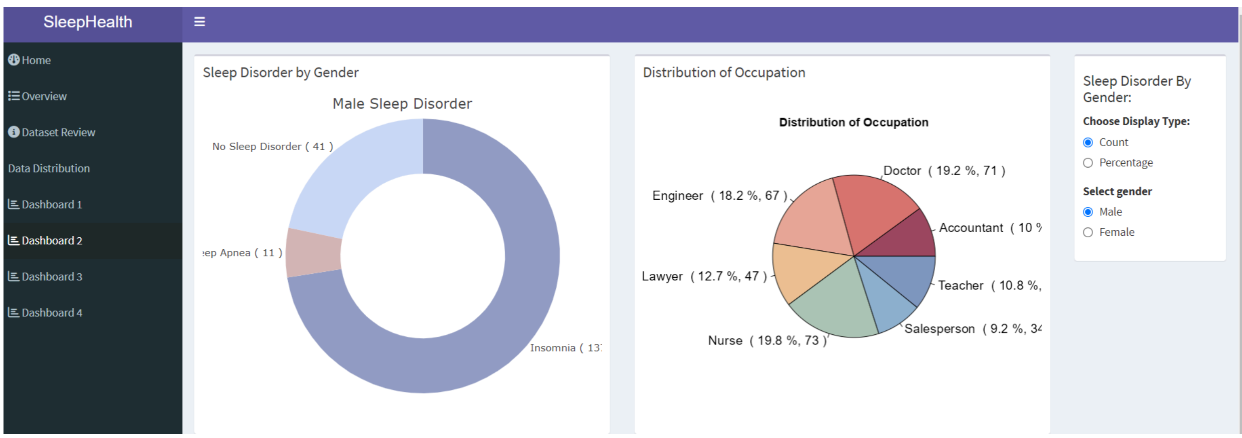


Figure 24 Dashboard 2 Sleep Disorder by Gender and Pie Chart of Occupation Distribution

Figure 31 Dashboard 2 Average Stress Level by Occupation

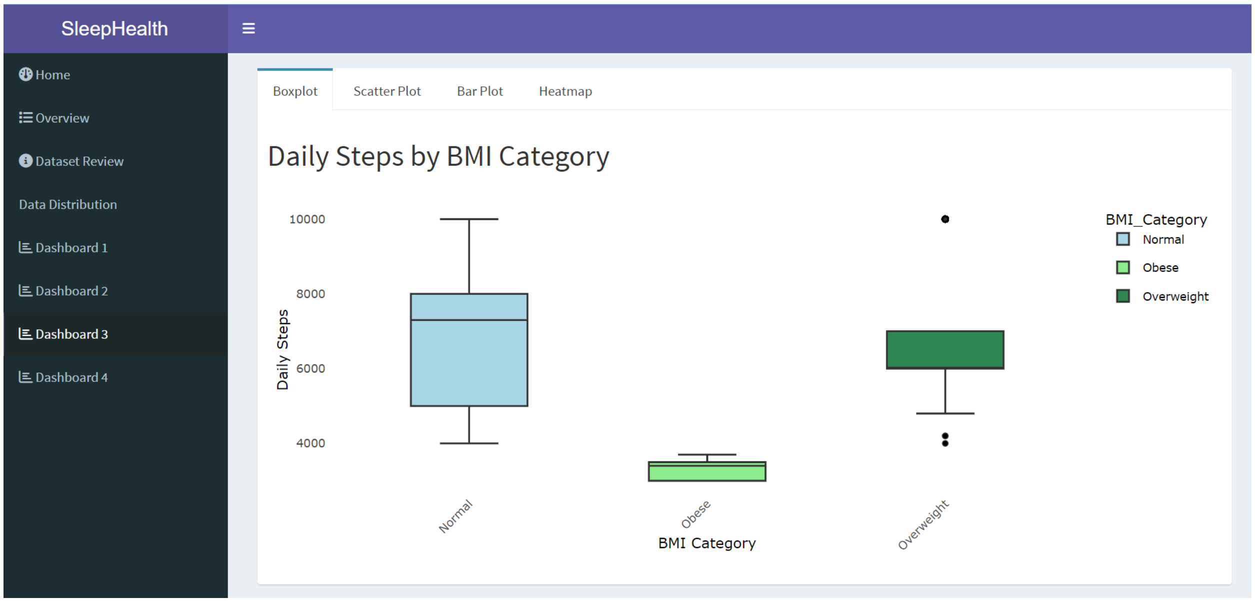


Figure 32 Dashboard 3 Boxplot Graph

Figure 33 Dashboard 3 Scatter Plot Graph

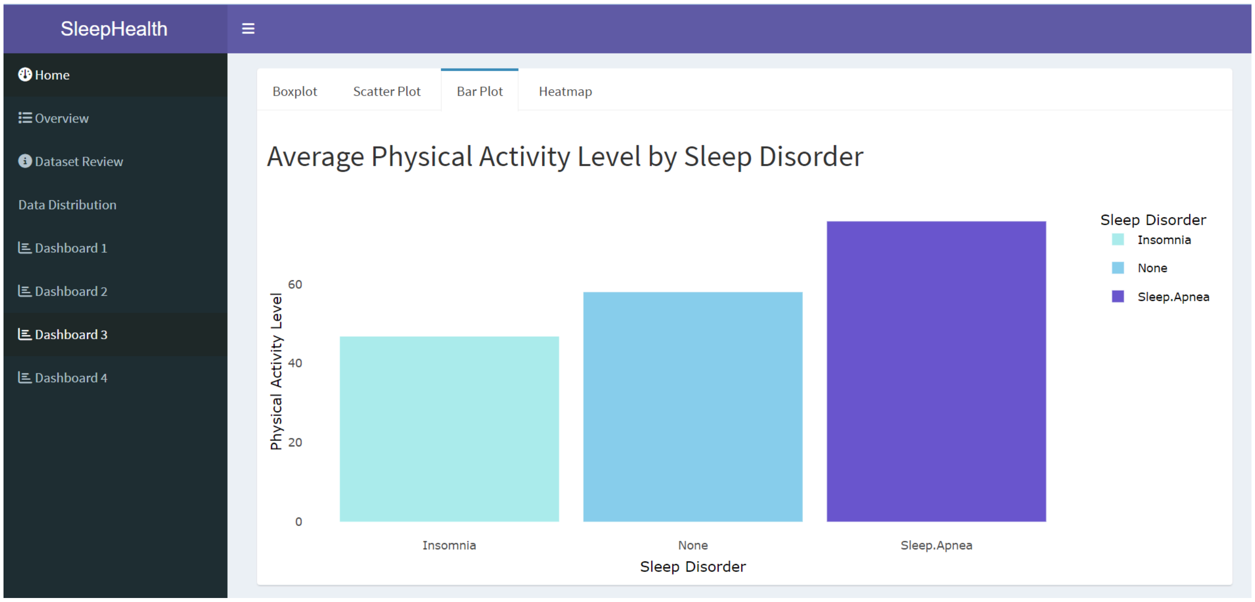


Figure 34 Dashboard 3 Bar Plot Graph

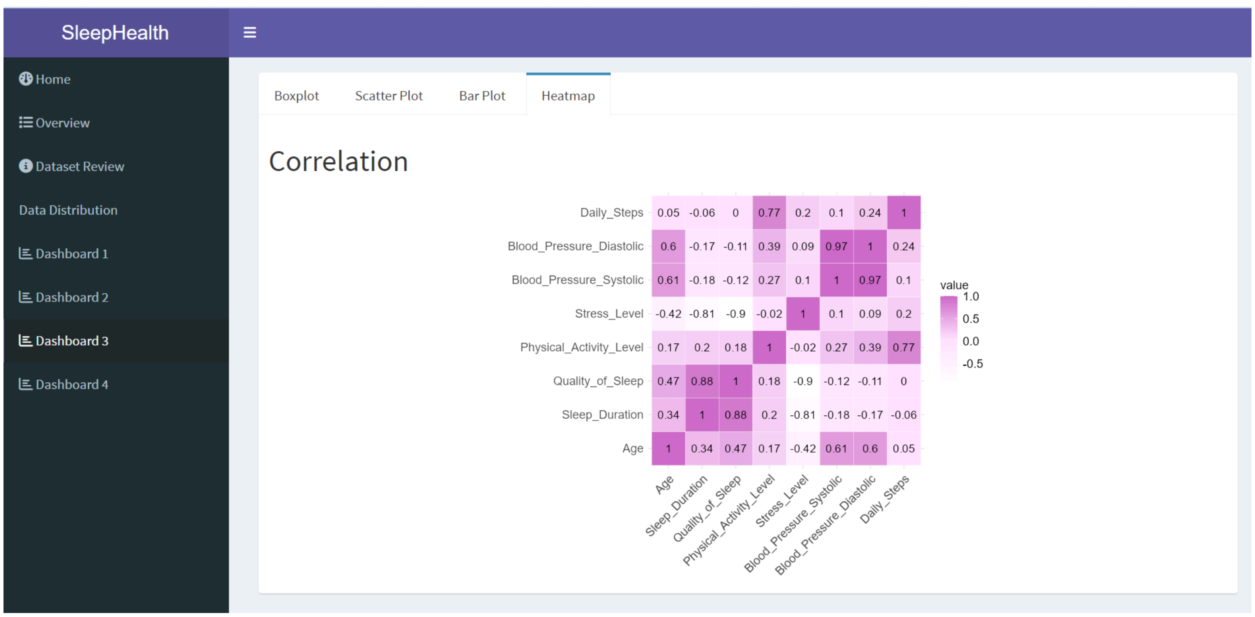


Figure 35 Dashboard 3 Heatmap

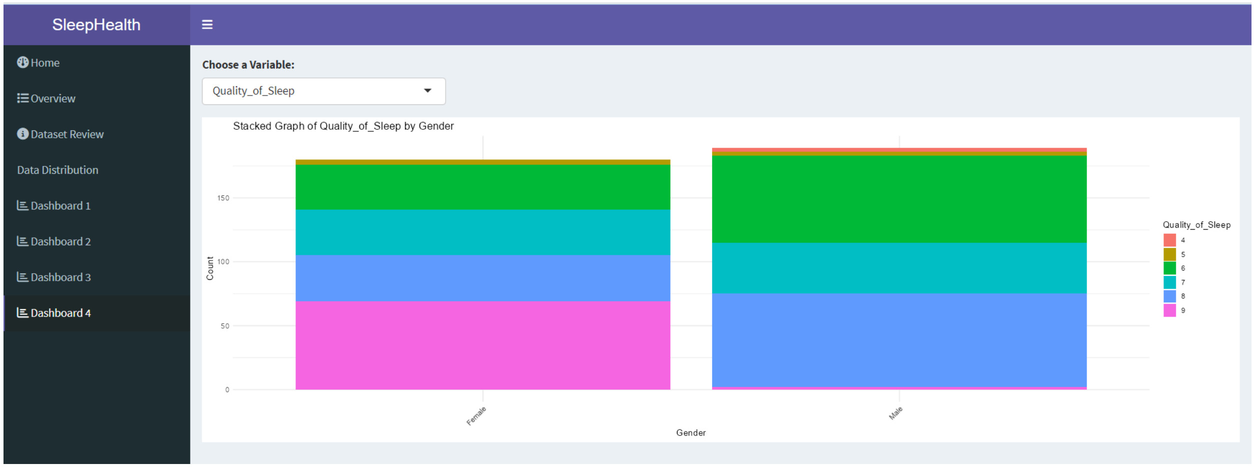


Figure 36 Dashboard 4 show Quality of Sleep by Gender

Figure 37 Dashboard 4 show Stress Level by Gender

Figure 38 Dashboard 4 show BMI Category by Gender