Sai Navya. Tadisetti

University of North Texas

CORE CONCEPT OF THE PROJECT : An evaluation of sleep patterns and performance that takes into consideration for few factors like age, alcohol use, smoking status, and frequency of exercise.

DATASET : Sleep Efficiency Dataset , Extracted from Kaggle.

INTRODUCTION :

Sleep is essential for sustaining both physical as well as mental health, making it a vital aspect of human health and well-being. A key element in measuring the level of nap is sleep efficiency, a statistic that calculates the proportion of time spent sleeping while in bed. Being obese, cardiovascular illnesses, and melancholy are just a few of the health issues that have been related to ineffective sleep. Therefore, it's crucial to comprehend what influences the quality of sleep and how to enhance it.

In the discipline of sleep medicine, sleep efficiency has received substantial study, and researchers have found a number of variables that influence it. People are having trouble sleeping in today's fast-paced society for a number of reasons, including workplace stress, lifestyle choices, issues with sleep, and the ubiquitous use of electronics. Understanding the factors that impact the degree of sleep and stepping up with practical solutions to enhance it are therefore crucial.

RELATED WORK :

In medical terms, the proportion of time spent sleeping when in bed is known as sleep efficiency. We may figure it out by combining the total length of time in bed (in minutes) by the amount of time spent sleeping . A sleep efficiency of at least 85% is regarded as typical. For instance, An adult who rests for six of the eight hours they spends on bed, would have a sleep quality of 75%. However, a person is said to be sleep efficient or have high sleep efficiency if they consume the bulk of their time in bed truly sleeping. Whereas, an individual is not regarded to have a high or adequate amount of sleep if they spend a large portion of their time in bed awake. This frequently happens with sleeplessness (insomnia). Sleep efficiency tends to be less than 75% when there is insomnia.

According to few studies , having a sound sleep can enhance your memory, innovation, efficiency, and ability to make choices. It also aids in maintaining emotional stability and alertness. Adopting to few other rituals, such as catching up on reading, having a shower, meditation, might help in relaxing our mind and body and get better restful sleep. It's vital to use devices as little as possible in the hour before bed. As the brain may be stimulated by these electronic devices, making it complicated to get a sleep. Hence, this cycle of sleep and awakening may be disrupted by the blue radiation generated by gadgets devices, which is regarded as one of the main negative sleeping behaviours.

Ageing and genetics are one of the most important variables that impact how well people sleep. People often suffer alterations to their sleep patterns as they become older, such as reduction in the quantity of prolonged sleep and rise in the number of awakenings. Seniors people spend shorter hours in sleeping and more time awake on bed as a result, which might reduce their ability to function when awake.

The ambient temperature, noise level, and illumination in sleeping space are all aspects that might affect how well we sleep. The ideal atmosphere for encouraging quality sleep is often one that is cold, calm, and gloomy. On top of it, having excessive levels of anxiety and depression, Being obese, diabetes, cardiovascular diseases may have trouble falling asleep, which may eventually affect the individuals mental health.

ABOUT THE DATASET :

The "Sleep Efficiency" dataset, which was extracted from Kaggle https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency includes data on the effectiveness of sleep and various variables that may influence it, such as age, gender, sleep duration, caffeine and alcohol intake, cigarette consumption, frequency of exercise, and the proportion of time spent in REM, deep, and light sleep. 452 observations and 12 factors make up the dataset. (as mentioned above). The dataset's goal is to pinpoint all the aspects that influence sleep quality and investigate their interrelationships.

Column Name	Data Type
ID	Integer
Age	Integer
Gender	String
Sleep duration	Float
Sleep efficiency	Float
REM sleep percentage	Integer
Deep sleep percentage	Integer
Light sleep percentage	Integer
Awakenings	Integer or Float
Caffeine consumption	Integer or Float
Alcohol consumption	Integer or Float
Smoking status	String
Exercise frequency	Integer or Float

METHODS :

As mentioned above , the initial step is obtained by considering a dataset and then , the entire process is followed by involving a number of phases, most notably the tools that needs to be used to deliver an accurate analysis showing the potential of the learner. Data visualization tools like Tableau and Flourish are employed as support for the major tools in this scenario. Prior this visualization, the incorrect data is first cleared out through some techniques like data clean-up and data mining.

TOOLS EMPLOYED FOR THIS PROJECT :

TABLEAU

With Tableau, a well-liked tool for displaying data, users can quickly link up, examine, and represent data from various sources. It offers a range of visualization choices, namely multimedia dashboards, charts, and infographics that may be tailored to meet the individual needs of users. Tableau's intuitive layout makes it simple for users to generate fundamental visualisations without any programming expertise while simultaneously providing more complex functionality for more seasoned users. Regardless of its numerous benefits, Tableau does have a few drawbacks that customers should take into account before purchasing the program. Because of its high price and limited modification choices compared to other data tools, it may not be affordable for certain firms. In addition, Tableau has certain capabilities for data planning, so users might have to use other programs to carry out these tasks earlier than importing data into Tableau. In the end, robust equipment is needed for Tableau to function properly because it can be asset-intensive. In spite of these constraints, Tableau is still a very powerful tool for data analysis, and for many individuals, its advantages surpass its cons.

FLOURISH

Creating dynamic visuals that are both educational and aesthetically pleasing is made simple with the aid of Flourish, an engaging web-based data visualization tool. It is a simple to operate application that doesn't require technical expertise and offers a number of templates, editing possibilities, and interaction with other data sources, include databases and spreadsheets. Flourish is a fantastic choice for collaborative projects since users can quickly turn data into graphs, maps, and animations and work in actual time with other fellow users.

The key benefits of Flourish are its low cost, extensive library of templates, and flexible customisation possibilities. Flourish does, however, have significant drawbacks, such as its constrained data integration capability, dearth of sophisticated data analysis functions, and constrained privacy management choices. Compared to other data visualization tools, it also provides a restricted selection of chart kinds, and only a few of them are really complicated. Keeping aside these drawbacks, Flourish continues to be a popular choice for people and businesses bidding to easily effortlessly produce striking and successful visuals from data.

DATA CLEANING :

In order to guarantee that the data remains accurate and complete , this data cleaning is a vital stage in the procedure for analysing data. In this instance, taking into account my database for sleep efficiency, some typical data cleaning operations that are included are sorting off all the null values or the values that have been left as empty bars, all of which have been filtered on numbers application and are then further processed through the next stages, which include data Interpreter.

DATA INTERPRETER :

The act of analysing and making logical sense of data with the goal to derive learnings and inferences from it is known as data interpretation. This interpretation function aids users in exploring and comprehending the data being displayed in data analysis.

The dataset that has been filtered is further molded in this stage by removing a significant number of rows based on its time and place. Exploring this dataset indicates that it has about 3,000 and above data entries thus, in order to fit these elements in the best way feasible, we must delete a few rows and alter accordingly for exhibiting the ideal visualisations and comprehending it in an accessible manner. The file was cleaned up in Microsoft Excel after it had been exported from Numbers, and it was then applied to Tableau for the implementing of visualisations of data as a precise CSV structured file.

EXPLORATORY DATA ANALYSIS (EDA) :

Exploratory data analysis (EDA) is a technique for condensing, visualizing, and examining datasets in order to discover their fundamental properties. As a means to obtain knowledge and guide choices on data cleansing, feature design, and modelling purposes, it is often carried out at the onset of a data science project. On top of it, it also uses some statistical techniques like hypothesis, to evaluate a claim based on a sample of data. All these test statistic will be computed to ascertain the actual results.

DESIGN OF RESEARCH QUESTIONS / HYPOTHESIS :

To work on the main hypothesis incorporated for my chosen dataset , we need to consider various variables that point out towards the outcomes in this complete project. Therefore, if we lean toward the visuals below, so as to validate the hypothesis, we begin by gathering data on sleep efficiency from both males and females who are smokers and non-smokers, their physical activity frequency, and REM% and contrast, analyse, and assess the correlation between the findings to see if there is a discernible difference in sleep proficiency between each of these categories. Besides that, these forecast analytic approaches may be used by researchers to pinpoint particular causes of sleep disorders and provide remedial measures.

Few of my Research questions here , include :

- Is there any line of difference in sleep efficiency depending on gender and smoking status ?
- Scale of sleep duration correlated with caffeine intake for the first 100 datasets.
- For a subset of 30 individuals in the sleep efficiency data set, what is the relationship between the total of REM sleep % and the sum of exercise frequency ?
- How does Alcohol intake impacts the Individual Awakenings schedule and can we identify any prediction analysis that overcome these restless sleep ?

TECHNIQUES and ALGORITHMS USED :

- In addition to using the flourish tool to generate eye-catching pie charts and other word cloud artworks that demonstrated in a much better way, I used tableau as the primary visualization tool for this project, as indicated in the sections above.
- A dashboard designed to be responsive was constructed to assist with project interface.
- I developed four sheets chart graphics, including a vertical bar graph, thematic maps, circle charts or pie diagrams and other scatter plots with bubbles, and a word cloud representation using the flourish tool, employing all the measurements, values, and column attributes.

ANALYSIS AND VISUALISATIONS :

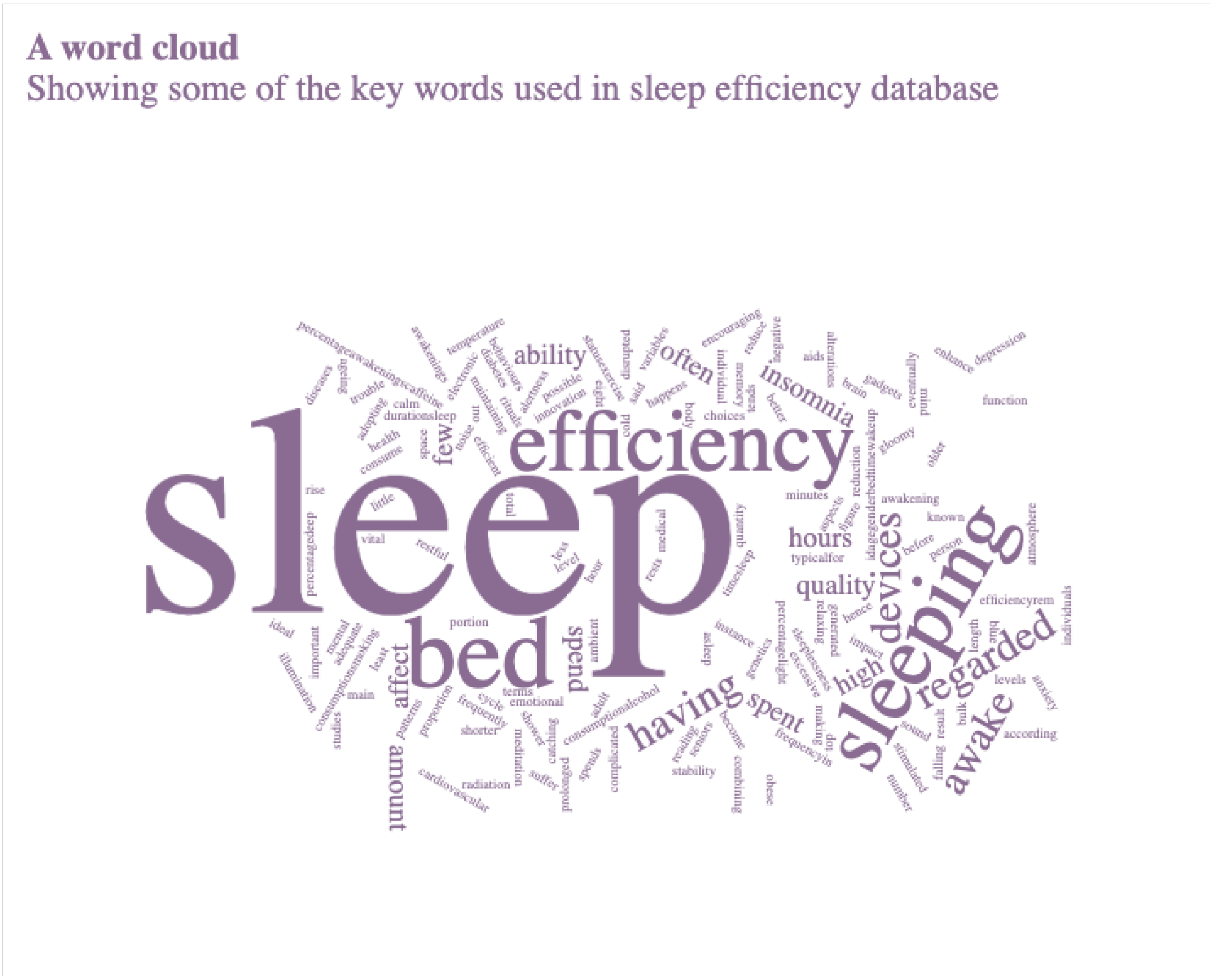

Figure 1: Word Cloud visualizing the frequency of keywords in the sleep efficiency dataset.

The occurrence of terms or phrases in a dataset can be displayed using a word cloud, a common style of illustration. A word cloud might be used to find the words or terms that are most frequently used to describe sleep efficiency in the context of this sleep performance database, such as "deep sleep," "REM sleep," "awakenings," and "light sleep." It can be particularly helpful in spotting patterns or outliers that would be hard to see through other kinds of visuals.

1. Is there any line of difference in sleep efficiency depending on gender and smoking status ?

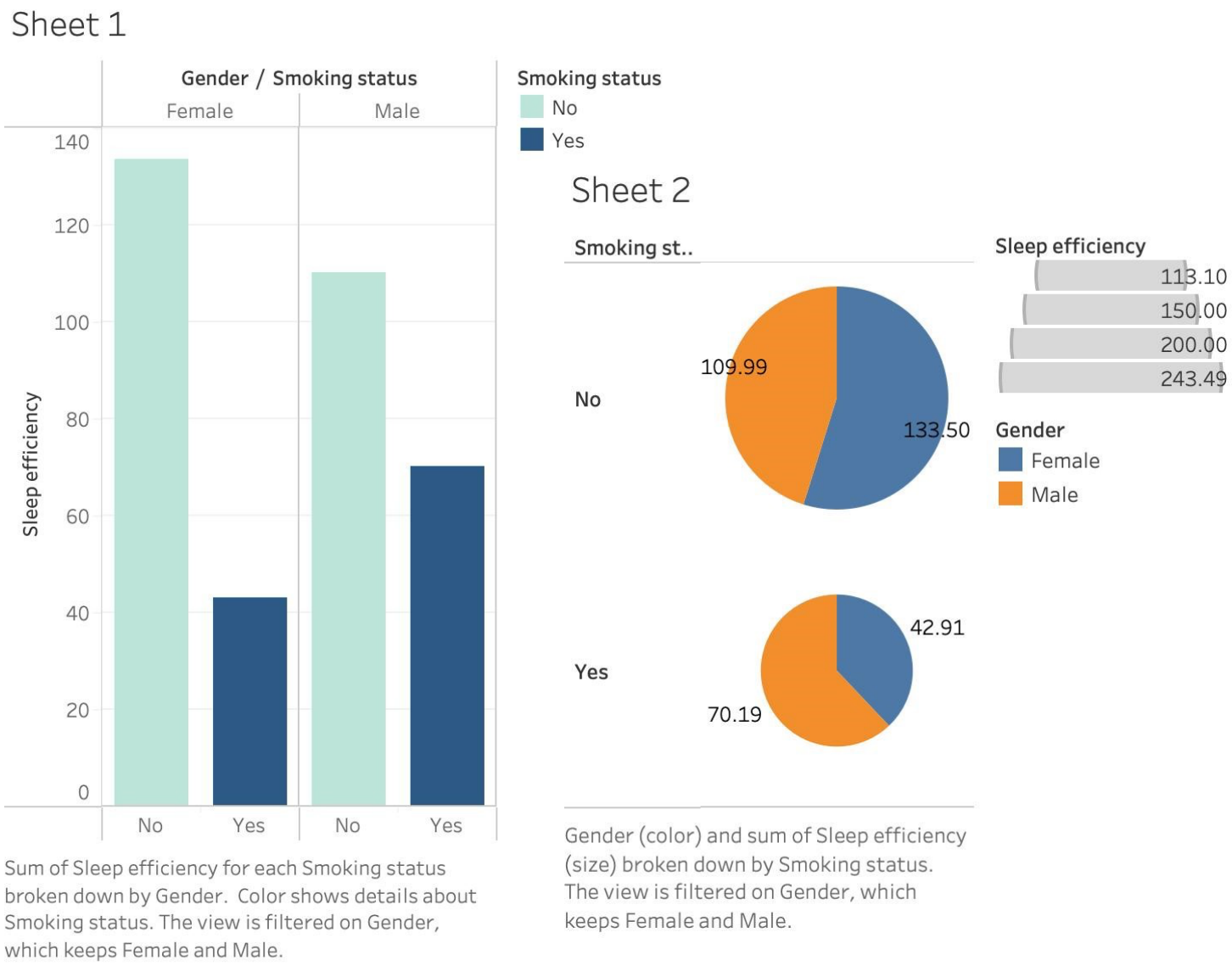

Figure 2 : The pie diagram and horizontal bar graph point to the impacts on sleep efficiency.

The pie chart and bar graph that are displayed above are commonly used to show several types of visualization that may be utilised to contrast the total of sleep efficiency for each smoking status segregated by gender. Pie charts are used to display the proportion of sleep efficiency in this situation, while bar graphs are able to show the total amount of sleep efficiency. The depiction is gender-filtered and marked by color, allowing for a deeper look of the distinctions between male and female category. In sum, both visualisations show that people who avoid cigarettes have better sleep quality than people who smoke. On top of that, women who don't smoke have higher sleep efficiency rates (133.50) than men who refrain from cigarettes, with an average rates of 109.99, thus these images enable rapid observations and analysis.

2. Scale of sleep duration correlated with caffeine intake for the first 100 datasets.

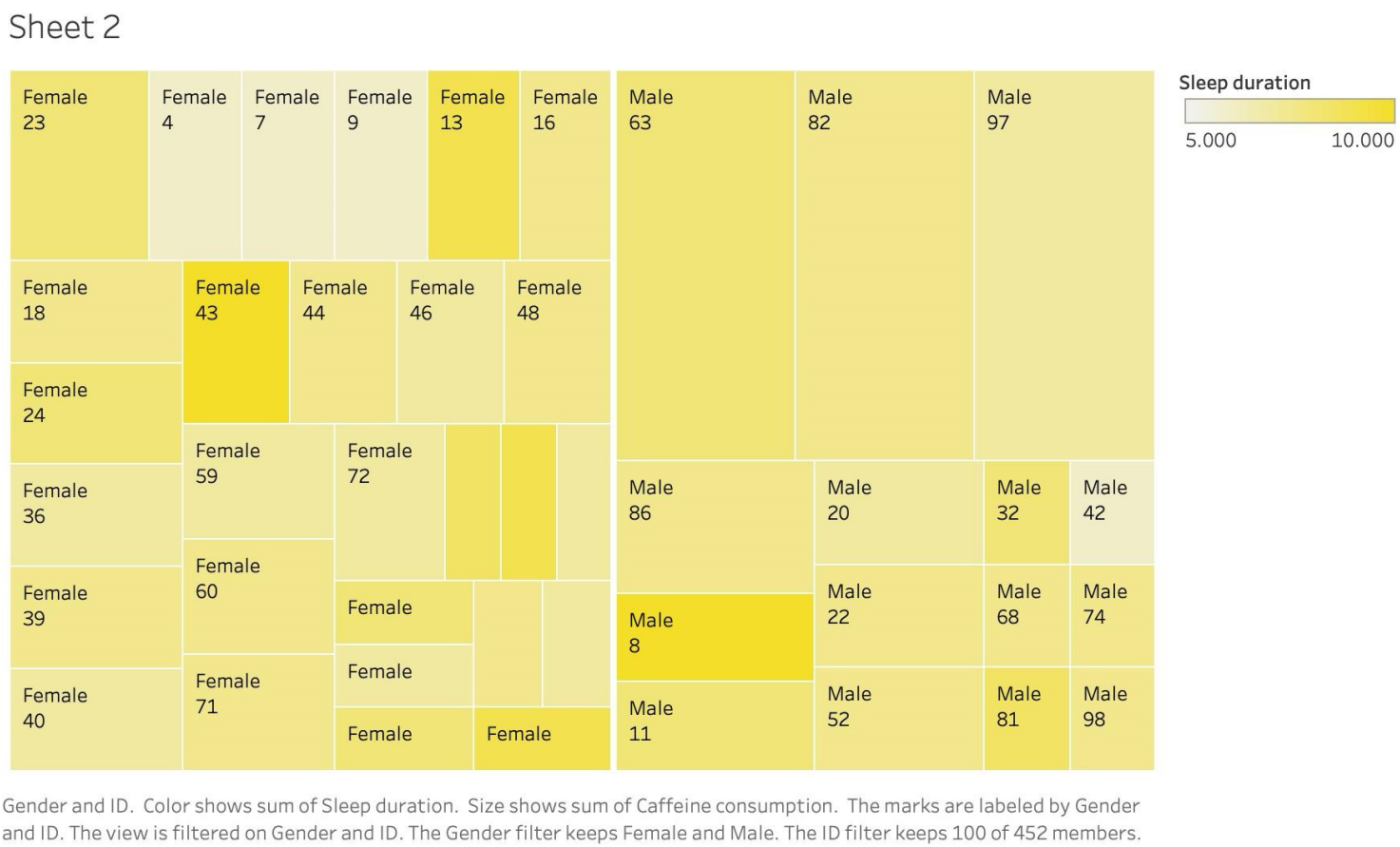

Figure 3 - In accordance with caffeine consumption, this heatmap depicts information on the first 100 datasets sleep duration.

One style of visualization that makes use of color tones to indicate variables in a data collection is the heat map depiction. By classifying gender (Male and Female) and ID tags, this heat map can be used to illustrate the total amount of sleep time in relation to the sleep efficiency data set. Firstly , there are 4 ratios (25 , 50 , 75 and 200) considered for caffeine intake, it is presented by the size of the blocks by symbolising it as : Small - 25 , Medium small - 50 , Medium - 75 and Large - 200. While the yellow color blocks are used to indicate the total amount of sleep time on a scale of 5.0 to 10.0 and this is represented by different tints of yellow color. For instance , all the darker yellow blocks indicate the highest 9 or 10 value, denoting the longest duration of sleep. On the other hand, the lighter to pale yellow blocks denote the average to lowest sleep duration (7.5, 6.0, and 5.7). The results of this chart reveals that, in comparison to those who use the stimulant caffeine in ratios of 75 and 200, with those who consume caffeine sparsely or not at all possess sound and decent sleep duration. These works can assist in detecting any patterns that may be helpful to locate areas where sleeping routines and general health could be improved.

3. For a subset of 30 individuals in the sleep efficiency data set, what is the relationship between the total of REM sleep % and the sum of exercise frequency ?

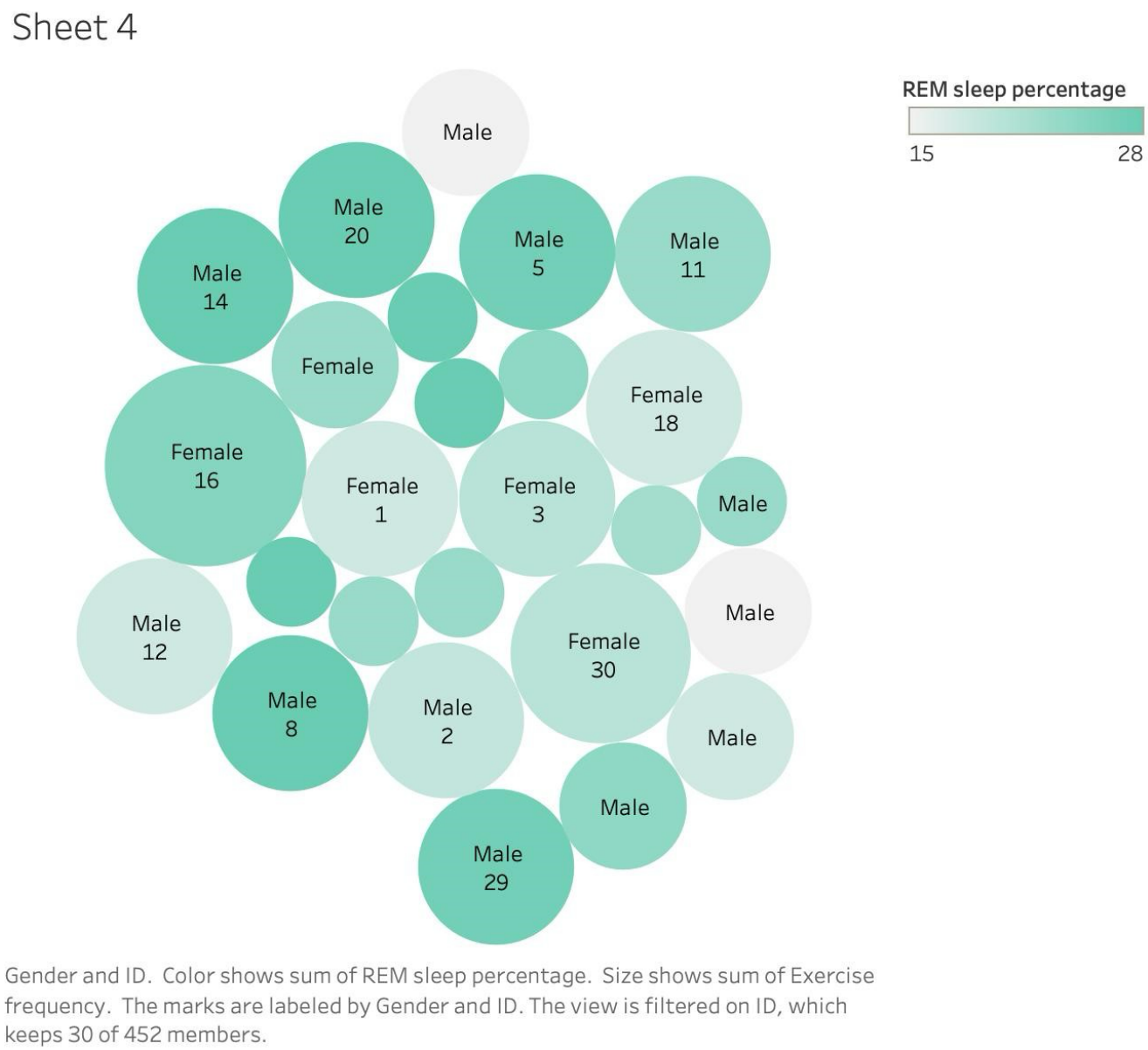

Figure 4: Packed bubble highlighting the link across the REM sleep percentage and exercise frequency.

The above packed - bubble chart graphic, which considers 30 datasets of the sleep efficiency, enables an interactive and captivating approach to investigate the association between the total of REM sleep % and the exercise frequency. It exposes ID and gender in the bubbles. Each bubble's size corresponds the total amount of exercise activity a person has done, with larger bubbles signifying people who have exercised more frequently overall. This parameter aligns with the bubble color's opacity while the REM percentage is stated. Low opacity shades connect to lower REM sleep % scales (15–20), higher than average opacity shades lean toward better REM sleep scales (21–25), and much greater opacity shades indent to the highest REM sleep scales (25–28).

According to the estimated findings, those who engage in a lot of aerobic activity are likely to experience good or excellent REM sleep %. Eventually, indolent people who exercise rarely or not fall under the scanty % REM sleep threshold.

4. How does Alcohol intake impacts the Individual Awakenings schedule and can we identify any prediction analysis that overcome these restless sleep ?

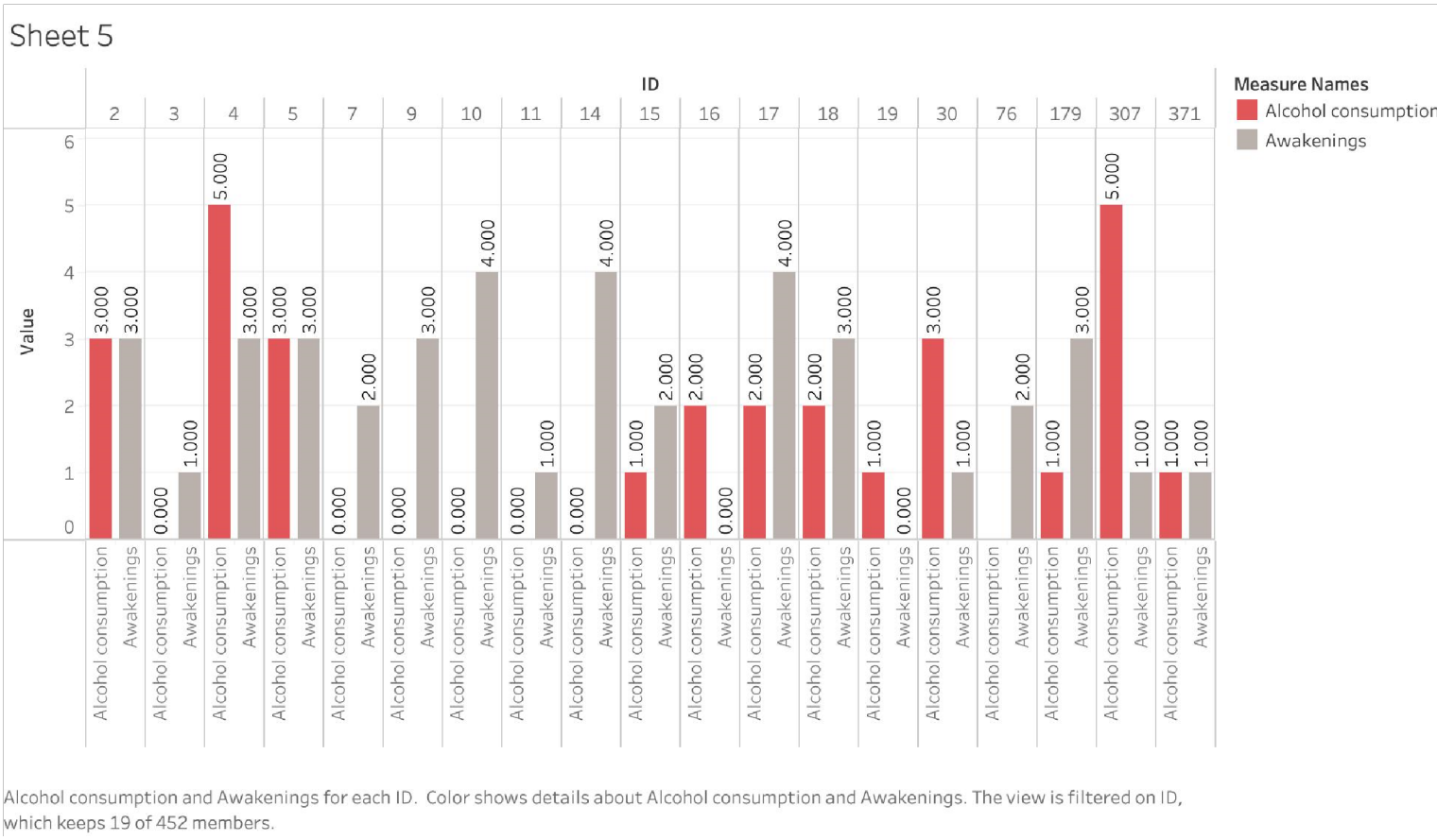

Figure 5 - Side - by - Side bar graph featuring the divergence of Alcohol consumption and Awakenings

One can examine the correlation between these two variables by comparing alcohol intake and awakenings for all the twenty IDs in a side-by-side bar chart. We may visually evaluate any potential connections among the two parameters by contrasting the red and grey bars. We can narrow on certain people and investigate their patterns of alcohol intake and awakenings in more depth via sorting on ID. Additionally integrating information, colorcoding tactics enables us to distinguish between the two variables with ease. This bar graph predicts with certainty that people who occasionally or rarely drink alcohol experience some altered awakening timings and can be rated on a scale of 1 to 3, but for those who consume alcoholic beverages desperately, they encounter a totally interrupted or negative sense of waking timings which in fact has an impact on their health. Last but not least, those who do not drink alcohol inclined to have healthy sleep habits that include consistent bedtime and awakening times and a scale level of 2 to 5.

Constructed on these graphs, we can predict unhealthy behaviours related to an active or sedentary lifestyle, and then we determine which norm has the largest and lowest percentages so that one can optimally alter or avoid them for a better standards of sleep productivity and an healthy living.

DISCUSSION :

Probing further into some general appraisal works, it indicates that the specific aims of the portrayal will influence the choice of chart style for the sleep efficiency dataset. picking the right type of chart, can assist to convey efficiently conclusions drawn from the data. For the very first visualization, I went on with bar graph and a pie diagram in order to visualise the sleep efficiency dataset by taking into account the gender and smoking status of the participants. The graph, with the bars representing the variables Yes or No on the X-axis and the two blue color hues designating gender male and female, shows the average sleep efficiency % for each group. For instance, the graph can reveal that smoking men have the lowest sleep efficiency % while non-smoking women have the greatest. By examining its proportions and the marked value that is raised on it, a pie diagram, on the other hand, explains the same concept in the most straightforward manner.

This sleep efficiency statistics may also be effectively shown via heat maps. The heat map can show the % of sleep efficiency dependent on the effects of caffeine ingestion, with coloration that is darker denoting longer sleep duration. Heat maps can help in finding trends by making it simple to compare productivity levels based on gender and quantity of sleep. But on a general note, these heat maps aren't always the ideal option, because it may prove difficult to identify between minute color-based alterations when reviewing huge datasets.

The last rendering that was affirmed was underpinned by bubble-packed charts, which were a great approach to show how a few consecutive objects merged with the bubble representation. while the color and size of the bubbles stand for the frequency of exercise for each person taken into consideration, and the measure of REM sleep efficiency is expressed by the volumes of the bubbles. The overarching patterns in the dataset may be successfully communicated by going into more detail regarding these Bubble packed graphics. Nevertheless, if there are numerous data categories scattered throughout them, they tend to get messy and disorganised making it difficult to compare the relationships between these bubbles.

FUTURE WORK :

Regardless of these limited number of illustrations that I added for my project, we can still use a few of the aspects to expand this dataset in a more sophisticated manner. We may concentrate on machine learning as an example, where algorithms may be employed to forecast sleep quality or detect probable sleep problems based on trends in the data input. The creation of therapies to enhance sleep quality, which could be assessed for efficacy using the dataset, is another crucial avenue. Last but not least, using wearable devices with sensor technology may present the chance to gather actual time sleep data and connect it with the already-existing information, enabling a more precise and in-depth investigation of sleep patterns over time.