IDS 400

Sleep Health and Lifestyle Analysis Project

Ngoc Ngan Ha Le

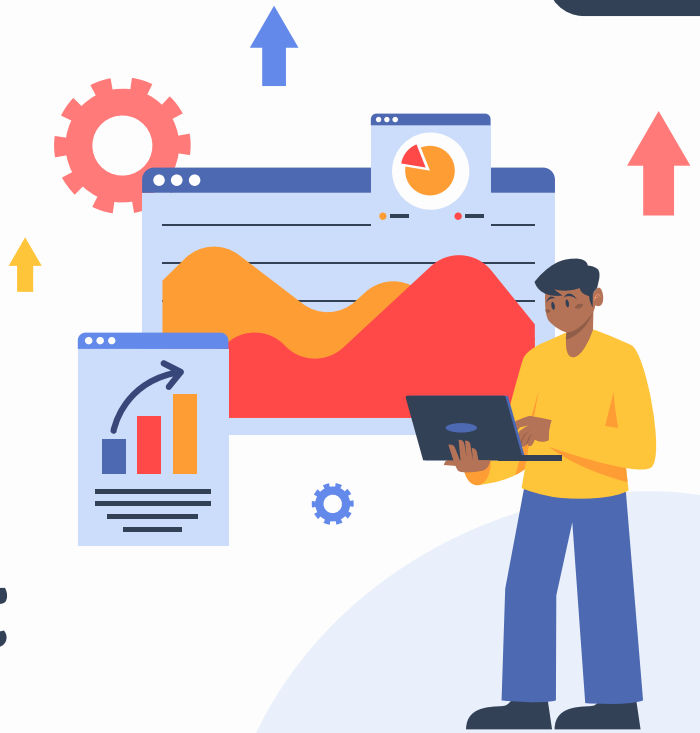
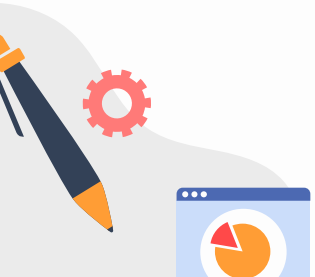


Table of contents



01

Introduction About Dataset and Project

Project Topic

Nowadays, in modern society, it's common to see people take part in busy,endless "gear". It's sadly to say that material life of human has been improved significantly, otherwise it reverse for health and mental life. Especially, sleep health has been reduced.

Why we come up with this topic:

- To explore relationship between lifestyle and sleep health
- To understand any hidden causes that lead to reduce sleep health in term of lifestyle
- Driven by our group's personal interest on health problem

Extract Extraction

- Explored Kaggle to identify a dataset related to lifestyle factors and sleep health.
- Selected the most relevant dataset based on scope and quality.
- Downloaded the dataset in CSV format for analysis.
- Prepared the data by cleaning, organizing, and preprocessing to ensure it was ready for accurate coding and analysis using Python.

Sleep_health_and_lifestyle_dataset-1

Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
6	Male	28	Software Engineer	5.9	4	30	8	Obese	140/90	85	3000	Insomnia
7	Male	29	Teacher	6.3	6	40	7	Obese	140/90	82	3500	Insomnia
8	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
9	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
10	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
11	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None
12	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
13	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None
14	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
15	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
16	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
17	Female	29	Nurse	6.5	5	40	7	Normal Weight	132/87	80	4000	Sleep Apnea
18	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	Sleep Apnea
19	Female	29	Nurse	6.5	5	40	7	Normal Weight	132/87	80	4000	Insomnia
20	Male	30	Doctor	7.6	7	75	6	Normal	120/80	70	8000	None
21	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
22	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
23	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
24	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
25	Male	30	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
26	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
27	Male	30	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
28	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
29	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
30	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
31	Female	30	Nurse	6.4	5	35	7	Normal Weight	130/86	78	4100	Sleep Apnea

Import Libraries and Load Dataset

- Mounting Google Drive to access files.
- Setting the working directory for project files.
- Importing libraries:
  - numpy for numerical operations.
  - pandas for data manipulation.
  - seaborn and matplotlib for visualization.
  - statsmodels for statistical modeling.

```
from google.colab import drive
import os
drive.mount("/content/drive")
```

```
Mounted at /content/drive
```

```
os.chdir("/content/drive/MyDrive/IDS400/")
```

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
```

```
health = pd.read_csv('Sleep_health_and_lifestyle_dataset.csv',delimiter=',')
```

Data Manipulation

- Renames "Normal Weight" to "Normal" in the BMI Category column for simplicity.
- Converting categorical data into numeric form for further coding.
  - Standardizes the Sleep Disorder column by replacing NaN with 0
  - Mapping disorders like "Sleep Apnea" and "Insomnia" to 1
- Displays the updated DataFrame to verify changes.

```python
health = pd.DataFrame(health)
health['BMI Category'] = health['BMI Category'].replace('Normal Weight', 'Normal')
```

Python

+ Code + Markdown

```python
health = pd.DataFrame(health)

health['Sleep Disorder'] = health['Sleep Disorder'].replace({np.NaN: 0, 'Sleep Apnea': 1, 'Insomnia' : 1})
print(health)
```

02

Multiple Linear Regression (With Category Variable)

Methodology

In this section, we want to investigate on relationship that to what extent physical activity level and sleep disorder impact on sleep quality.

Idea

In this section, we want to investigate on relationship that to what extent physical activity level and sleep disorder impact on sleep quality. To be specific, we want to know whether how present/absent of sleep disorder could impact to sleep quality in scale of (1-10)

Statistic

Field of knowledge from statistic that multiple regression with joint of category variable.

- Dependent variable:sleep quality.
- Independent variable:physical activity level
- Category variable: sleep disorder

Coding

- Selects Physical Activity Level and Sleep Disorder as independent variables, and Quality of Sleep as the dependent variable.
- Adds a constant term to the independent variables to include the intercept in the regression model.
- Fits an Ordinary Least Squares (OLS) regression model to quantify the relationship between the variables.
- Prints a summary of the model, showing key statistics and the strength of the relationships.

```
independent_variable = health[['Physical Activity Level', 'Sleep Disorder']]
dependent_variable = health['Quality of Sleep']

independent_variable = sm.add_constant(independent_variable)


model = sm.OLS(dependent_variable, independent_variable).fit()


print(model.summary())
```

Finding

- Run OLS Regression Result that show coefficients of Physical Activity Level and Sleep Disorder on Sleep Quality
- Regression equation = 6.9075 + PhysicalLevel x 0.0124 - SleepDisoder x 0.7911
    - For every 1 level change in Physical Activity Level result in increase 0.0124 in term of quality of sleep
    - If Sleep Disorder is appear (Insomnia, Sleep Apnea), the sleep quality reduce by 0.7911. Otherwise, if Sleep Disorder absent, the sleep quality are not going to be affected by Sleep Disorder coefficient

OLS Regression Results

Dep. Variable:	Quality of Sleep	R-squared:	0.143
Model:	OLS	Adj. R-squared:	0.138
Method:	Least Squares	F-statistic:	30.95
Date:	Mon, 16 Dec 2024	Prob (F-statistic):	3.71e-13
Time:	05:04:05	Log-Likelihood:	-568.57
No. Observations:	374	AIC:	1143.
Df Residuals:	371	BIC:	1155.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.9075	0.177	38.997	0.000	6.559	7.256
Physical Activity Level	0.0124	0.003	4.476	0.000	0.007	0.018
Sleep Disorder	-0.7911	0.117	-6.767	0.000	-1.021	-0.561

Omnibus:	23.302	Durbin-Watson:	0.426
Prob(Omnibus):	0.000	Jarque-Bera (JB):	12.189
Skew:	0.260	Prob(JB):	0.00226
Kurtosis:	2.284	Cond. No.	196.

Coding

- Converts the health dataset into a pandas DataFrame for compatibility.
- Creates a regression plot using Seaborn's lmplot, with:
  - Physical Activity Level on the x-axis.
  - Quality of Sleep on the y-axis.
  - Color differentiation (hue) based on the presence of sleep disorders, using the "Set1" color palette.
- Adds a title and labels to enhance readability.
- Displays the plot to analyze trends and correlations visually.

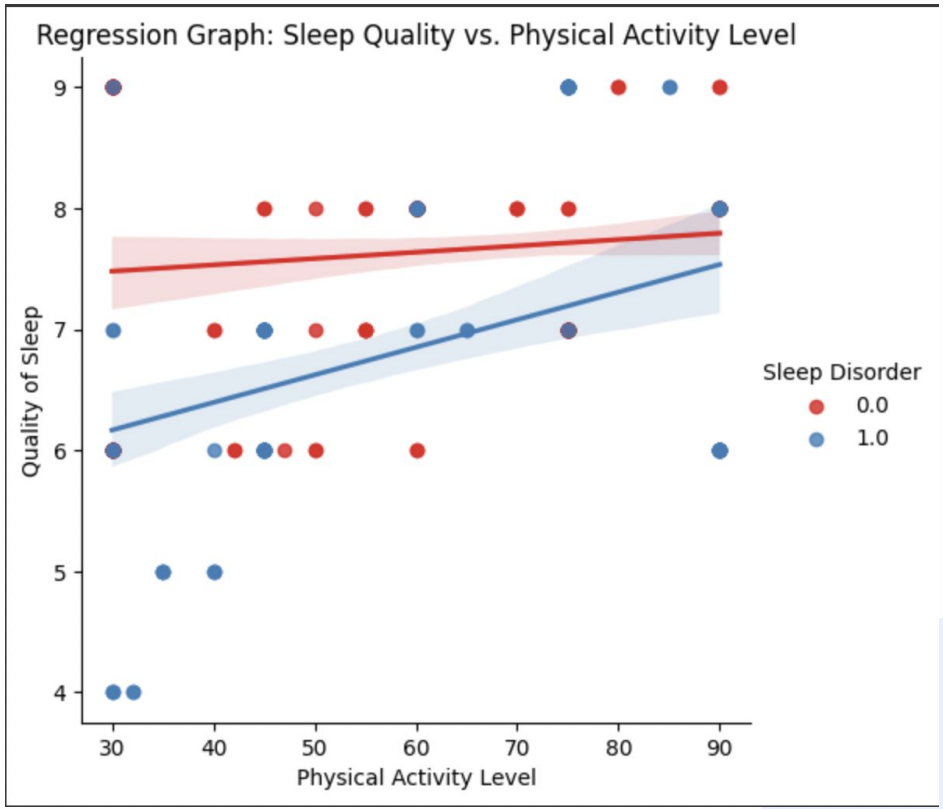
```
health = pd.DataFrame(health)
sns.lmplot(
    data=health,
    x='Physical Activity Level',
    y='Quality of Sleep',
    hue='Sleep Disorder',
    palette='Set1'
)

plt.title('Regression Graph: Sleep Quality vs. Physical Activity Level')
plt.xlabel('Physical Activity Level')
plt.ylabel('Quality of Sleep')

plt.show()
```

Finding

- Run sns.lmplot show the beside regression Graph that Physical Activity Level as independent variable, Quality of Sleep as dependent variable and Sleep Disorder as category variables.
- It support for the regression coefficients that:
    1. Present of sleep disorder result in lower Quality of Sleep Level.
    2. As Physical Activity Level increase, the quality of sleep increase doesn't matter sleep disorder present or not. However, the effect is much more significant within present of Sleep Disorder group



Further Research

I found out that both Physical Activity Level and Sleep Disorder has effect on Quality of Sleep. There are some majority causes:

1. Physical activity can help in improve quality of sleep because it increases the production of melatonin, a hormone that regulates sleep-wake cycles
2. Physical activity helps to regulate body temperature, which is necessary for falling asleep, as an increase in body temperature during physical activity aids the eventual drop 30-90 minutes post-exercises, facilitating easier sleepiness
3. The repeated awakenings associated with sleep apnea disorder make typical, restorative sleep impossible

03

Visualization (Clustered Columns Chart)

Methodology

In this section, we want to investigate on how Quality of Sleep and Stress Level across Age. Furthermore, looking for any pattern, trend.

Idea

In this section, we want to evaluate on how Quality of Sleep and Stress Level change across Age. Furthermore, we want to observe whether Quality of Sleep and Stress Level affect each other.

Visualization

- Conduct Clustered Column Chart in which Age on y-axis, level value of “Quality of Sleep” and “Stress Level” on the x-axis.
- Graph will show both Quality of Sleep and Stress Level respectively for each group of age.

Coding

- Reshapes the health DataFrame using pd.melt to consolidate Quality of Sleep and Stress Level into a single variable for comparison.
- Creates a bar plot where:
  - Age is on the x-axis.
  - The combined values for Quality of Sleep and Stress Level are on the y-axis.
  - The hue differentiates between the two variables.
- Enhances the plot with labels for the axes and a descriptive title.
- Displays the visualization to explore age-related trends in sleep and stress levels.

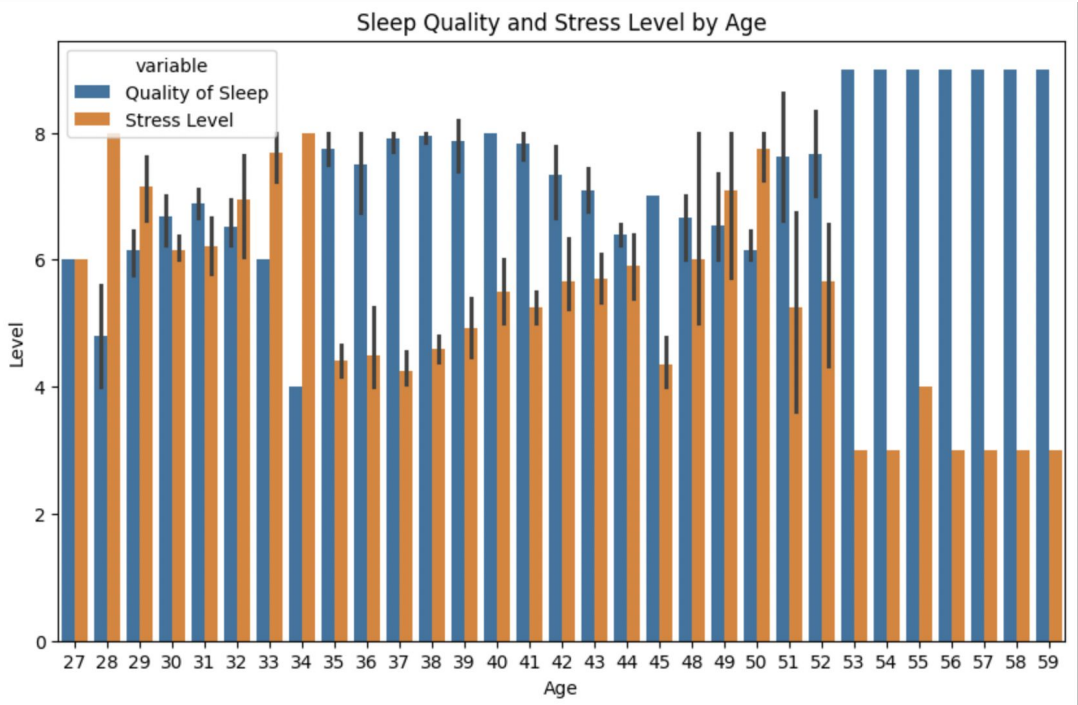
```
health = pd.DataFrame(health)
health_melted = pd.melt(health, id_vars=["Age"], value_vars=["Quality of Sleep", "Stress Level"],
                        var_name="variable", value_name="value")


plt.figure(figsize=(10, 6))
sns.barplot(x="Age", y="value", hue="variable", data=health_melted)

plt.xlabel('Age')
plt.ylabel('Level')
plt.title('Sleep Quality and Stress Level by Age')
```

Finding

- Run Visualization by sns.barplot show the beside graph.
- Observation from graph that:
  - Before age of 53, it seem like the Quality of Sleep has rise from age 27 to 41 then experience decline before rise again and reach stable after 52.
  - For Stress Level, it's high from 27-34 (peak at 28), decline during 35-48, then arise again around 48-50 then significant reduce and stable after 52.
  - It can be observed that there's not strong correlation relationship between "Quality of Sleep" and " Stress Level" as Quality of Sleep doesn't increase when Stress Level increase vice versa.



Further Research

I found out that has both “Quality of Sleep” and “Stress Level” do different across age and have pattern. There are several main causes:

1. For oldest people (after fifties), the body secretes less melatonin which important chemicals can coordinate circadian rhythms thus helps in promoting sleep

2. Younger people tending experience more concerning including jobs, families and homes, all of which create stress

3. 25-year-olds reported stressors on nearly 50% of days, while 70-year-olds reported stressors on only 30% of days

4. Older adults have more resilience than younger adults in the face of stressful events.

04

Visualization (Joint Plot)

Methodology

In this section, we want to investigate on how BMI Category and Daily Steps across Occupation. Furthermore, looking for any pattern, trend.

Idea

In this section, we want to evaluate on how BMI Category and Daily Steps change across Occupation. Furthermore, we want to observe whether there is correlation relationship between BMI Category and Daily Steps

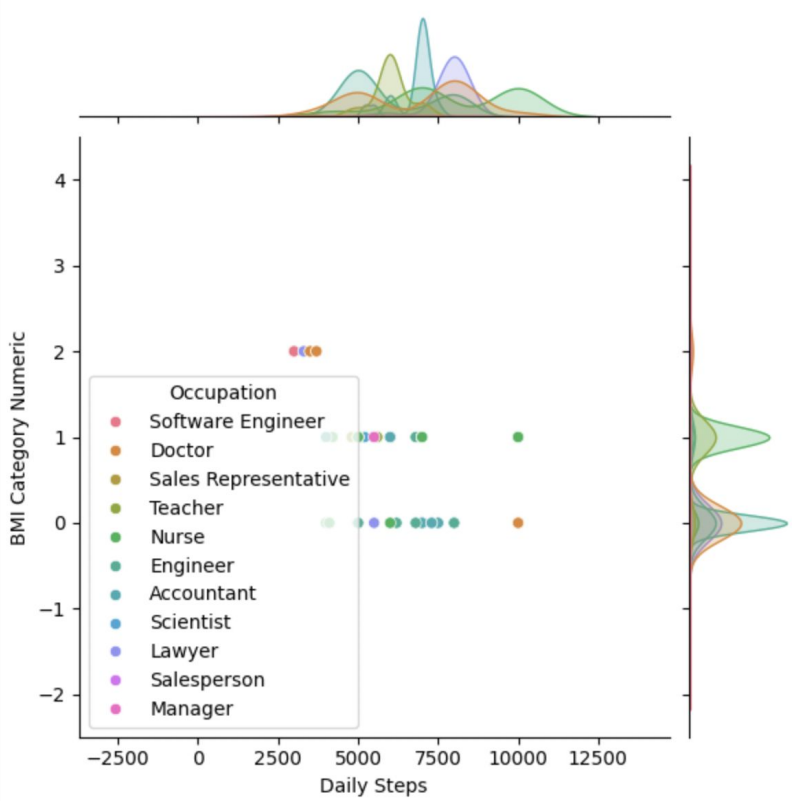
Visualization

Conduct Joint Plot in which Occupation on x-axis, BMI Category on the y-axis

- Convert BMI Category as categorical in to numeric (0 for Normal, 1 for Overweight , and 2 for Obese)
- Graph will show plot and line plot for each Occupations regarding to Daily Steps and BMI Category.

Finding

- Run Visualization by sns.jointplot show the beside graph.
- Observation from graph that:
  - Even though nurse tending have higher frequency that have large amount of daily steps (around 10,000), nurse have highest participation fall in Overweight of BMI Category.
  - Software Engineer have not a bad habitat within daily steps at around 5,000-10,000), however only this occupation show noticeable amount of obesity when look at the line plot
  - It seem like pattern in teacher is approachable in which they have low daily steps (about 2,500-6,000) and also considered as group have high frequency in overweight category



Further Finding

- Run correlation matrix to determine whether there is opposite relationship between Daily Steps and BMI Category within Nurse Occupation
- Observe from the matrix:
  - Correlation between Daily Steps and BMI Category is positive. Indicate that as Daily Steps increase, BMI Category also increase (after convert increase mean move toward obese).
  - That support for insight drawn joint plot from previous slides regarding Nurse Occupation

Correlation matrix for Nurse

	Daily Steps	BMI Category
Daily Steps	1.000000	0.479187
BMI Category	0.479187	1.000000

Further Research

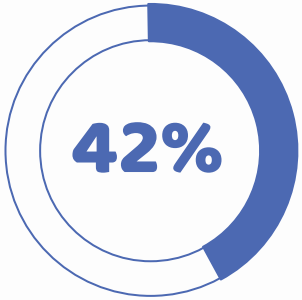
I found out that has both “Daily Steps” and “BMI Category” do different across Occupation and sometimes have pattern but sometimes doesn’t. There are several main causes:

1. Even people who work in nurse industry on their feet a lot at work, however because of job-specific that could contribute to poor dietary choices also reduced physical activity level thus lead to high probability of obesity

2. Software Engineer face with a lot of stressed, furthermore the job-specific doesn’t require much physical activities. Because of mental stress, it common leads to overeating and snacking.

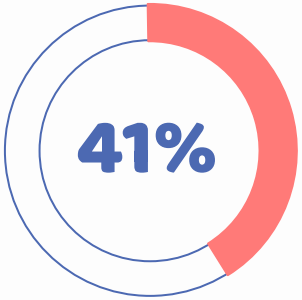
3. Similar situation happened with people whose jobs are teacher. Job’s specific that spend long time sitting, not too much physical activity and potential easy access to unhealthy snacks.

Let's use some percentages!



Abnormal BMI

41% of population of this dataset experience abnormal BMI



Sleep Disorder

41% of population of this dataset experience sleep disorder

Physical Activity Level

59% of population of this dataset experience over level 50 of physical activity level

Conclusions

Finding 1

There are relationship between sleep disorder and physical activity level affect to quality of sleep. Sleep Disorder does reduce quality of sleep level and Physical Activity Level do increase quality of sleep

Finding 2

There are not adequate evidence to conclude that there is relationship between stress level and quality of sleep across age from 27-59

Finding 3

Daily Steps amount do effect on BMI Category, however there are also exception with some specific occupation such as nurse due to job-specific

Resources

Aging and sleep. (2023). Retrieved from https://www.sleepfoundation.org/aging-and-sleep

Alnawwar, M. A., Alraddadi, M. I., Algethmi, R. A., Salem, G. A., Salem, M. A., & Alharbi, A. A. (2023). The effect of physical activity on sleep quality and sleep disorder: A systematic review. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC10503965/

Campbell, M. (n.d.). Study finds that experiences of daily stress decrease as people age: Penn State University. Retrieved from https://www.psu.edu/news/health-and-human-development/story/study-finds-experiences-daily-stress-decrease-people-age

Chin, D. L., Nam, S., & Lee, S.-J. (2016). Occupational factors associated with obesity and leisure-time physical activity among nurses: A Cross Sectional Study. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC4871118/

How to fix the health issues associated with the engineering lifestyle. (2020). Retrieved from https://www.ltvco.com/engineering/health-implications-engineer-lifestyle/#poor-food-decisions

The impact of stress on older adults and tips for managing it. (2024). Retrieved from https://www.hebrewseniorlife.org/blog/impact-stress-older-adults-and-tips-managing-it

Sleep apnea. (2023). Retrieved from https://www.mayoclinic.org/diseases-conditions/sleep-apnea/symptoms-causes/syc-20377631#:~:text=The%20repeated%20awakenings%20associated%20with,TV%20or%20even%20when%20driving.

Sleep health. (n.d.). Retrieved from https://www.nhlbi.nih.gov/health-topics/education-and-awareness/sleep-health

Thanks!

Do you have any questions?