Data Science and Computer Programming
Final Project
Sleep and Health Data Analysis

I. Motivation

With the changes in modern lifestyles, staying up late and insomnia have become common issues. These sleep problems have a certain impact on people's physical health. Therefore, we aim to analyze data to explore what factors influence our sleep quality and how sleep affects various aspects of the human body.

II. Data Information

Using “Sleep Health and Lifestyle Dataset” from Kaggle

	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
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	NaN
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea

Read Data

```
Person ID                  int64
Gender                    object
Age                        int64
Occupation                object
Sleep Duration           float64
Quality of Sleep           int64
Physical Activity Level    int64
Stress Level               int64
BMI Category              object
Blood Pressure            object
Heart Rate                 int64
Daily Steps                int64
Sleep Disorder            object
dtype: object
```

Data Types

```
Person ID                  0
Gender                     0
Age                        0
Occupation                 0
Sleep Duration             0
Quality of Sleep           0
Physical Activity Level    0
Stress Level               0
BMI Category               0
Blood Pressure             0
Heart Rate                 0
Daily Steps                0
Sleep Disorder           219
dtype: int64
```

Missing Values

III. Data Cleaning

```
data['Sleep Disorder'] = data['Sleep Disorder'].fillna('No Disorder')
data.head()
```

	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
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	No Disorder
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	No Disorder
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	No Disorder
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea

2. Converting Data from Object to Numeric Type

(1) Gender (Male → 0, Female → 1)

	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	0	27	6.1	6	42	6	Overweight	126/83	77	4200	No Disorder
1	0	28	6.2	6	60	8	Normal	125/80	75	10000	No Disorder
2	0	28	6.2	6	60	8	Normal	125/80	75	10000	No Disorder
3	0	28	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	0	28	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea

(2) Sleep Disorder (No Disorder → 0, Sleep Apnea & Insomnia → 1)

	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	0	27	6.1	6	42	6	Overweight	126/83	77	4200	0
1	0	28	6.2	6	60	8	Normal	125/80	75	10000	0
2	0	28	6.2	6	60	8	Normal	125/80	75	10000	0
3	0	28	5.9	4	30	8	Obese	140/90	85	3000	1
4	0	28	5.9	4	30	8	Obese	140/90	85	3000	1

(3) BMI Category (Normal → 1, Overweight → 2, Obese → 3)

	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI_Level	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	0	27	6.1	6	42	6	2	126/83	77	4200	0
1	0	28	6.2	6	60	8	1	125/80	75	10000	0
2	0	28	6.2	6	60	8	1	125/80	75	10000	0
3	0	28	5.9	4	30	8	3	140/90	85	3000	1
4	0	28	5.9	4	30	8	3	140/90	85	3000	1

(4) Converting Blood Pressure into Systolic & Diastolic and then into Blood Pressure Level

	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI_Level	Heart Rate	Daily Steps	Sleep Disorder	Systolic	Diastolic
0	0	27	6.1	6	42	6	2	77	4200	0	126	83
1	0	28	6.2	6	60	8	1	75	10000	0	125	80
2	0	28	6.2	6	60	8	1	75	10000	0	125	80
3	0	28	5.9	4	30	8	3	85	3000	1	140	90
4	0	28	5.9	4	30	8	3	85	3000	1	140	90

Systolic & Diastolic

	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI_Level	Heart Rate	Daily Steps	Sleep Disorder	Blood_Pressure_Level
0	0	27	6.1	6	42	6	2	77	4200	0	3
1	0	28	6.2	6	60	8	1	75	10000	0	3
2	0	28	6.2	6	60	8	1	75	10000	0	3
3	0	28	5.9	4	30	8	3	85	3000	1	4
4	0	28	5.9	4	30	8	3	85	3000	1	4

Blood Pressure Level

BLOOD PRESSURE CATEGORY	SYSTOLIC mm Hg (upper number)	and/or	DIASTOLIC mm Hg (lower number)
NORMAL	LESS THAN 120	and	LESS THAN 80
ELEVATED	120 – 129	and	LESS THAN 80
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 1	130 – 139	or	80 – 89
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 2	140 OR HIGHER	or	90 OR HIGHER
HYPERTENSIVE CRISIS (consult your doctor immediately)	HIGHER THAN 180	and/or	HIGHER THAN 120

※ The classification of blood pressure levels

· Systolic < 120 & Diastolic < 80 → Blood Pressure: 1

· 120 < Systolic < 130 & Diastolic < 80 → Blood Pressure: 2

· 130 < Systolic < 140 & 80 ≦ Diastolic < 90 → Blood Pressure: 3

· Systolic ≧ 140 or Diastolic ≧ 90 → Blood Pressure: 4

· Systolic > 180 or Diastolic > 120 → Blood Pressure: 5

IV. Data Analysis

1. Correlation & Heatmap

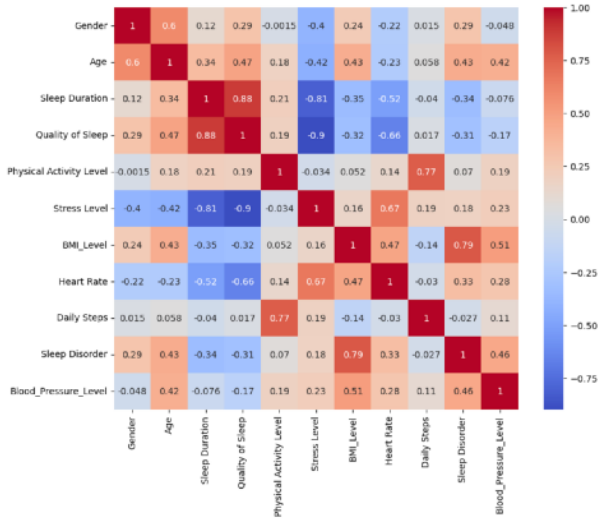
	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI_Level	Heart Rate	Daily Steps	Sleep Disorder	Blood_Pressure_Level
Gender	1.000000	0.596358	0.121579	0.291366	-0.001454	-0.396018	0.242743	-0.217105	0.014509	0.285824	-0.047597
Age	0.596358	1.000000	0.344709	0.473734	0.178993	-0.422344	0.427704	-0.225606	0.057973	0.432007	0.415846
Sleep Duration	0.121579	0.344709	1.000000	0.883213	0.212360	-0.811023	-0.349605	-0.516455	-0.039533	-0.338622	-0.075951
Quality of Sleep	0.291366	0.473734	0.883213	1.000000	0.192896	-0.898752	-0.324413	-0.659865	0.016791	-0.310984	-0.166965
Physical Activity Level	-0.001454	0.178993	0.212360	0.192896	1.000000	-0.034134	0.052076	0.136971	0.772723	0.069787	0.189633
Stress Level	-0.396018	-0.422344	-0.811023	-0.898752	-0.034134	1.000000	0.157649	0.670026	0.186829	0.181685	0.225487
BMI_Level	0.242743	0.427704	-0.349605	-0.324413	0.052076	0.157649	1.000000	0.469897	-0.142785	0.785389	0.505978
Heart Rate	-0.217105	-0.225606	-0.516455	-0.659865	0.136971	0.670026	0.469897	1.000000	-0.030309	0.330254	0.278124
Daily Steps	0.014509	0.057973	-0.039533	0.016791	0.772723	0.186829	-0.142785	-0.030309	1.000000	-0.026575	0.111964
Sleep Disorder	0.285824	0.432007	-0.338622	-0.310984	0.069787	0.181685	0.785389	0.330254	-0.026575	1.000000	0.457981
Blood_Pressure_Level	-0.047597	0.415846	-0.075951	-0.166965	0.189633	0.225487	0.505978	0.278124	0.111964	0.457981	1.000000

· Strong Positive Correlation:

1. Sleep Duration vs Quality of Sleep (0.88)
2. Sleep Disorder vs BMI (0.79)
3. Physical Activity Level vs Daily Steps (0.77)

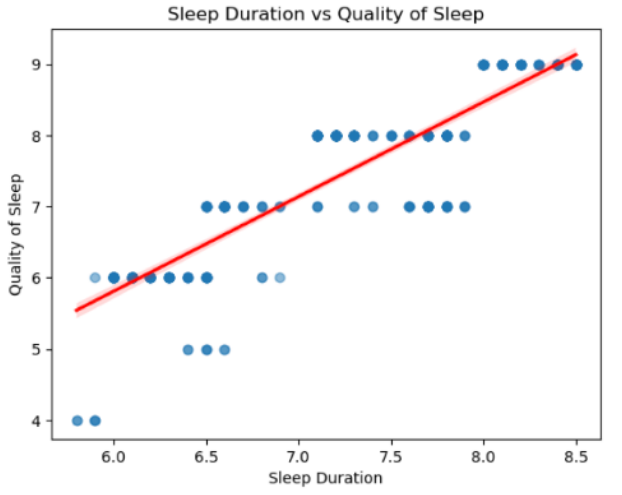
· Strong Negative Correlation:

1. Stress Level vs Quality of Sleep (-0.9)
2. Stress Level vs Sleep Duration (-0.81)



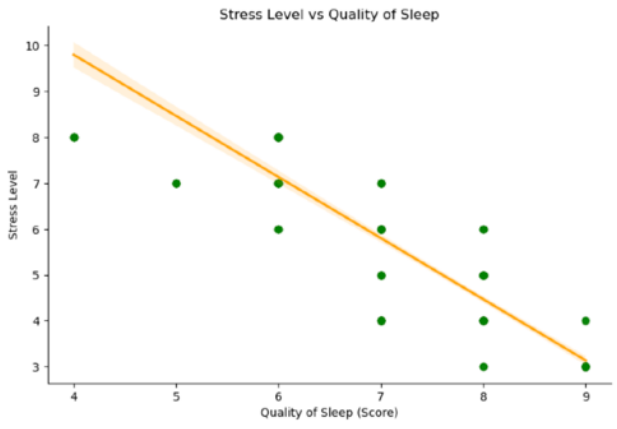
2. The Relationship Between Sleep Duration and Sleep Quality

The figure presented on the right illustrates a strong positive correlation of 0.88 between sleep duration and sleep quality. This substantial relationship indicates that individuals who experience longer sleep durations are likely to enjoy superior sleep quality. Such findings emphasize the critical role of adequate sleep duration in achieving restorative sleep. Therefore, prioritizing sufficient sleep is essential for enhancing overall sleep experiences and promoting well-being.



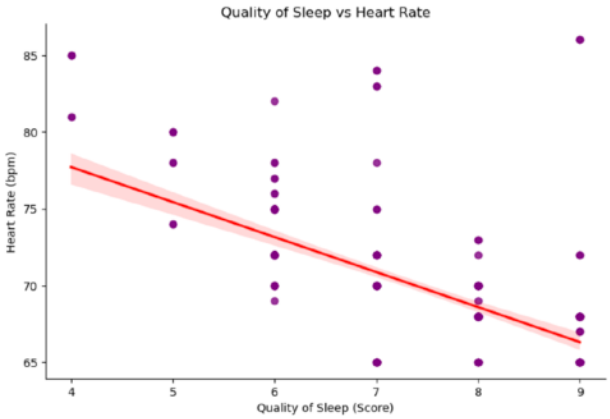
3. The Relationship Between Stress and Sleep Quality

The two variables demonstrate a strong negative correlation of -0.9, signifying that individuals with elevated stress levels typically experience diminished sleep quality. This significant relationship indicates that stress constitutes a primary factor influencing sleep quality. Consequently, addressing stress management strategies may be essential for enhancing overall sleep experiences. By prioritizing techniques aimed at stress reduction, individuals could improve their sleep quality and, subsequently, their overall well-being. Such findings underscore the critical importance of understanding the dynamic interplay between stress and sleep.



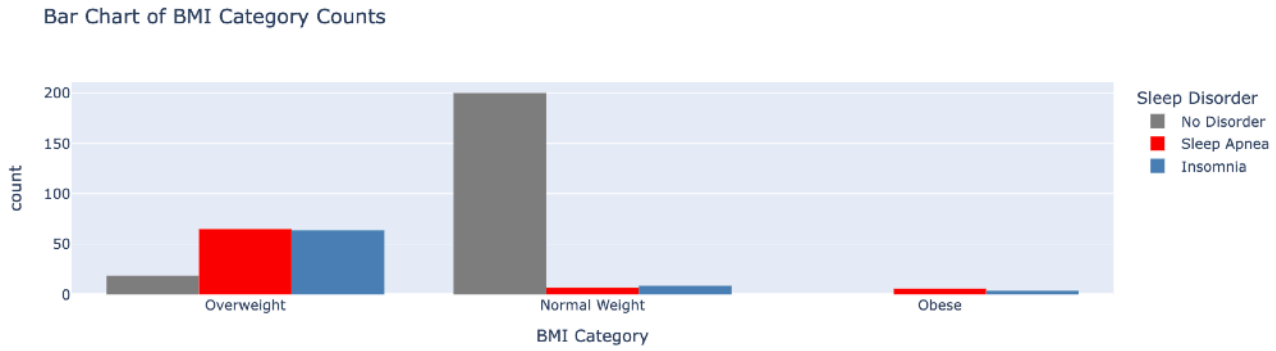
4. The Relationship Between Sleep Quality and Heart Rate

The chart demonstrates a negative correlation between sleep quality and heart rate; as sleep quality deteriorates, heart rate tends to increase. However, the scattered nature of the data complicates this relationship. Furthermore, factors such as stress, diet, exercise, and individual physiological differences significantly impact heart rate. Consequently, it is essential to consider these variables to improve the accuracy of predictions regarding fluctuations in heart rate and overall health outcomes.



5. The Impact of Sleep Disorders on Health

The chart illustrates that a majority of individuals suffering from sleep disorders are classified as overweight, whereas those who do not experience such disorders typically exhibit normal body weight. Consequently, one may infer that sleep disorders contribute to body weight complications, including obesity.

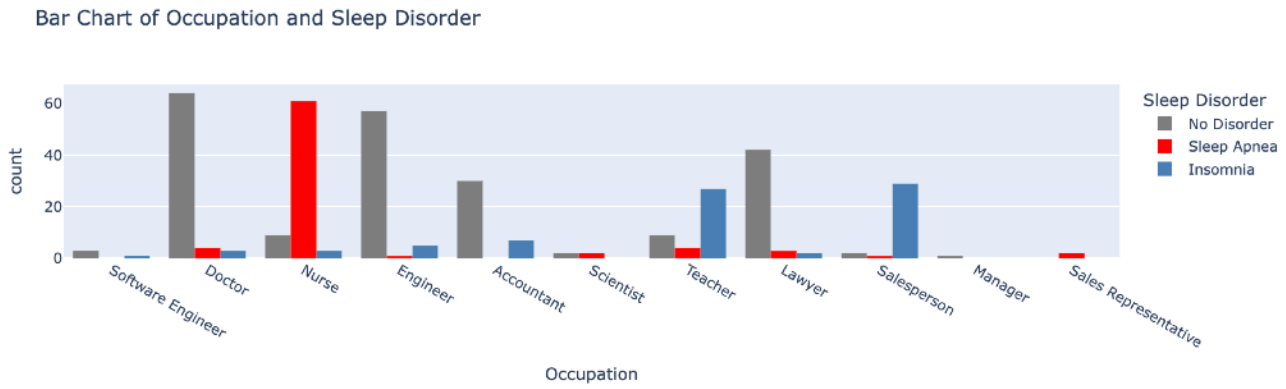


6. Sleep and Health Analysis by Profession

Among various professions, Sales Representatives experience the highest levels of stress and the shortest duration of sleep. In contrast, Engineers experience the least amount of stress and enjoy the longest duration of sleep. This disparity underscores the differing impacts of occupational roles on stress and sleep patterns.

7. Occupation and Sleep Disorders

The chart indicates that Nurses predominantly experience sleep disorders, with a significant number suffering from Sleep Apnea. In contrast, Teachers and Salespeople primarily encounter issues related to Insomnia. However, it is important to recognize that numerous factors contribute to sleep disorders. Therefore, further analysis utilizing additional data is necessary to identify the underlying causes of these conditions. A comprehensive understanding of these factors will facilitate targeted interventions to improve sleep health among affected professionals.



8. Linear Regression

· Feature Variables:

(1) Sleep Duration

(2) Stress Level

(3) Physical Activity Level

· Target Variable:

Heart Rate

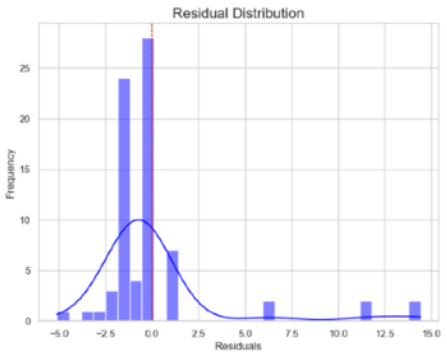
· Model Training Results:

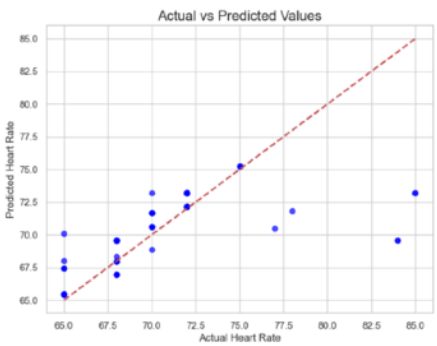
(1) MSE (Mean Squared Error): 12.03

(2) The model performs well, with predicted values closely matching the actual values.

(3) The features are strongly correlated with the target, and their influence on heart rate is significant.

(4) Model Application Value: The model can help predict heart rate and provide health-related recommendations for individuals.





9. Logistic Regression

· Feature Variables:

(1) Sleep Duration

(2) Stress Level

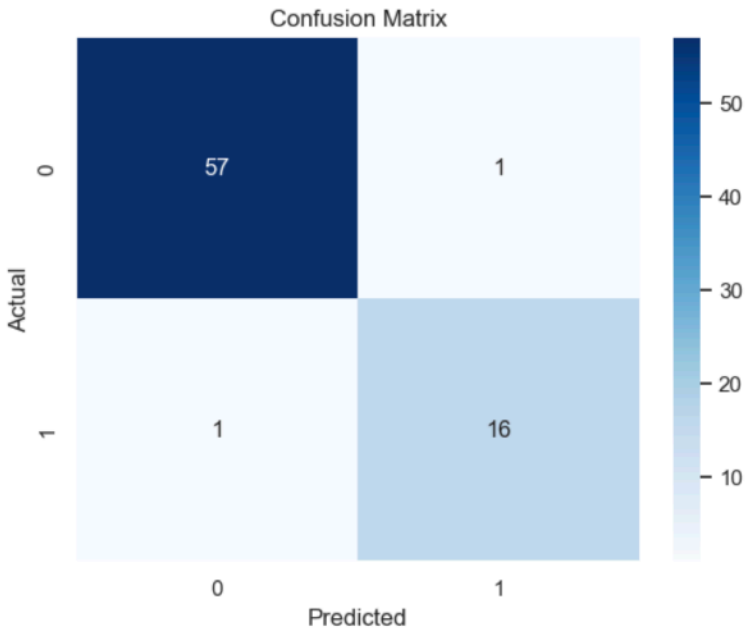
(3) Heart Rate

· Target Variable:

High Stress (Stress Level > 7 is considered 1, otherwise 0)

· Model Training Results:

```
Accuracy: 0.9733333333333334
Confusion Matrix:
 [[57  1]
 [ 1 16]]
```



(1) Accuracy: 97.3%, demonstratir levels.

(2) The classification is well-balanced, with only two misclassified samples.

(3) The features show a significant correlation with stress levels.

(4) Additional features that influence stress (e.g., diet, occupation) could be incorporated to improve the model.

V. Conclusion

1. The Importance of Adequate Sleep: Sleep duration is negatively correlated with heart rate, meaning that longer sleep can improve heart health.
2. The Impact of Stress on Sleep: Stress significantly affects sleep quality—higher stress levels lead to poorer sleep.
3. The Impact of Stress on Heart Rate: Stress has the most significant effect on heart rate, making it a key feature for predicting heart rate.
4. The Negative Correlation of Sleep Duration: The negative correlation of sleep duration highlights the importance of sufficient sleep in maintaining a stable heart rate.
5. Model Performance: While the overall model performance is satisfactory, there is still room for further optimization.

VI. Github Link

https://github.com/ethanlin1126/DSCP

VII. References

[Kaggle Dataset]: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

[Blood Pressure]: https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings