

# SLEEP HEALTH & LIFESTYLE

---

# CONTENTS

- About This Dataset
- Data Analysis
- Data Visualization
- Algorithms And Machine Learning
- Summary

# ABOUT THIS DATASET

---

# DATASET OVERVIEW

The Sleep Health and Lifestyle Dataset is a comprehensive collection of data on sleep patterns , lifestyle factors, and cardiovascular health  for 375 individuals. The dataset comprises 13 key variables, categorized as follows:

# 1. DEMOGRAPHICS:

- Person ID: Unique identifier for each individual.
- Gender:  (Male/Female)
- Age:  (Years)
- Occupation:  (e.g., Software Engineer, Doctor, Teacher)
- BMI Category:  (Underweight, Normal, Overweight, Obese)

BMI category of the person (e.g., Underweight, Normal, Overweight). The BMI is defined as the body mass divided by the square of the body height, and is expressed in units of kg/m<sup>2</sup>. Major adult BMI classifications are underweight (under 18.5 kg/m<sup>2</sup>), normal weight (18.5 to 24.9), overweight (25 to 29.9), and obese (30 or more).

# 1. DEMOGRAPHICS:

	Person ID	Gender	Age	Occupation	BMI Category
1	1	Male	34	Doctor	Underweight
2	2	Female	27	Software Engineer	Normal
3	3	Male	50	Engineer	Overweight
4	4	Female	65	Nurse	Obese

## 2. CARDIOVASCULAR HEALTH:

- Blood Pressure:  (Systolic/Diastolic)

Blood pressure measurement of the person, indicated as systolic pressure over diastolic pressure.

- **Systolic BP (Top Number):** Measures the pressure when the heart contracts (pumps blood). This is the higher of the two numbers. Normal range: 90–120 mmHg.
- **Diastolic BP (Bottom Number):** Measures the pressure when the heart relaxes (between beats). This is the lower number. Normal range: 60–80 mmHg

- Heart Rate:  (Beats per minute)

Heart Rate (bpm): The resting heart rate of the person in beats per minute.

## 2. CARDIOVASCULAR HEALTH: ❤️

	Blood Pressure	Heart Rate
1	126/83	77
2	125/80	75
3	140/90	68
4	120/80	84

## 3. LIFESTYLE:

- Physical Activity Level:  (Minutes per day)  
number of minutes the person engages in physical activity daily.
- Stress Level:  (Scale of 1 to 10)  
subjective rating of the stress level experienced by the person, ranging from 1 to 10.
- Daily Steps:  (Number of steps)  
number of steps the person takes per day.

### 3. LIFESTYLE:

	Physical Activity Level	Stress Level	Daily Steps
1	75	4	8000
2	30	6	7000
3	45	8	4000
4	60	7	5000

## 4. SLEEP-RELATED FACTORS:

- Sleep Duration:  (Hours per day)  
The number of hours the person sleeps per day.
- Quality of Sleep:  (Scale of 1 to 10)  
A subjective rating of the quality of sleep, ranging from 1 to 10.

## 4. SLEEP-RELATED FACTORS:

	Sleep Duration	Quality of Sleep
1	6.5	9
2	7.3	7
3	8.1	6
4	6.2	8

## 5. SLEEP DISORDER: 😴 (NONE, INSOMNIA, SLEEP APNEA)

THE PRESENCE OR ABSENCE OF A SLEEP DISORDER IN THE PERSON (NONE, INSOMNIA, SLEEP APNEA).

	Sleep Disorder
1	Insomnia
2	None
3	None
4	Sleep Apnea

# DATA ANALYSIS

---

- Import the necessary libraries.
- Read the data file.
- Understanding the data format.
- General information about the data.
- Check for missing values.

- Check for duplicate rows.
- Number of unique values in each column.
- Display unique values in each column.
- Data cleaning.
- Statistical data description.
- Encoding.

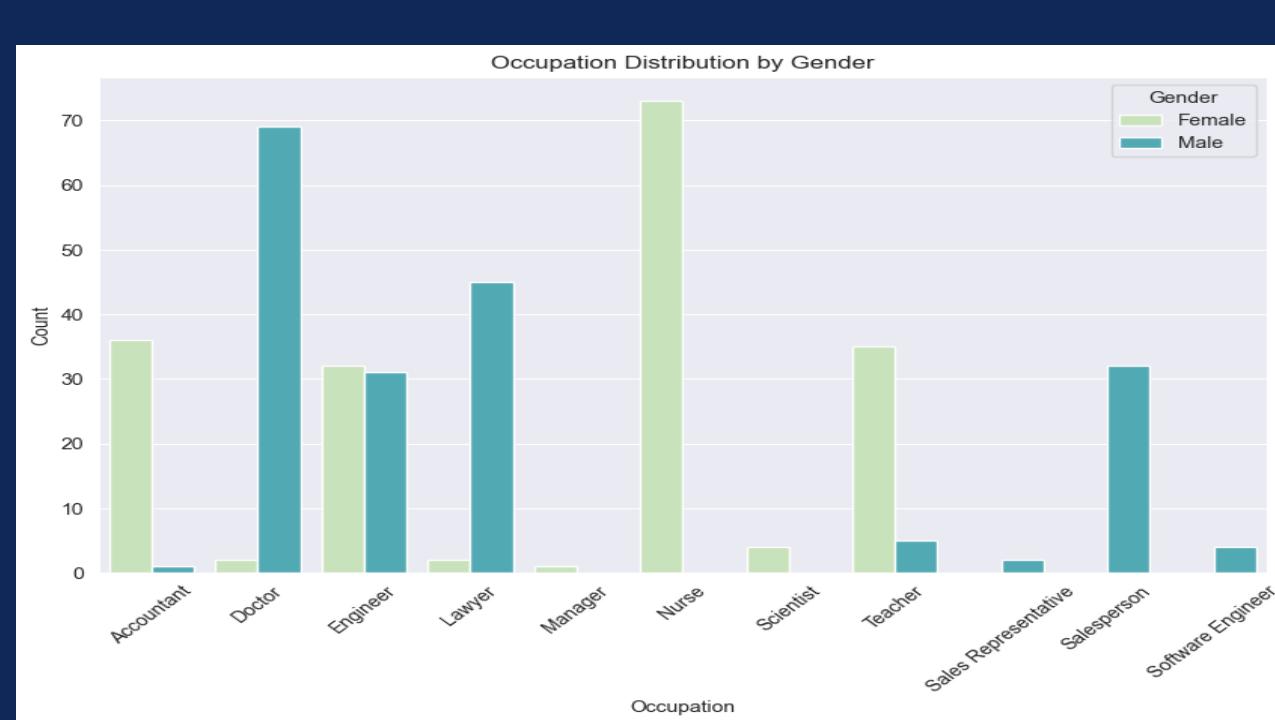
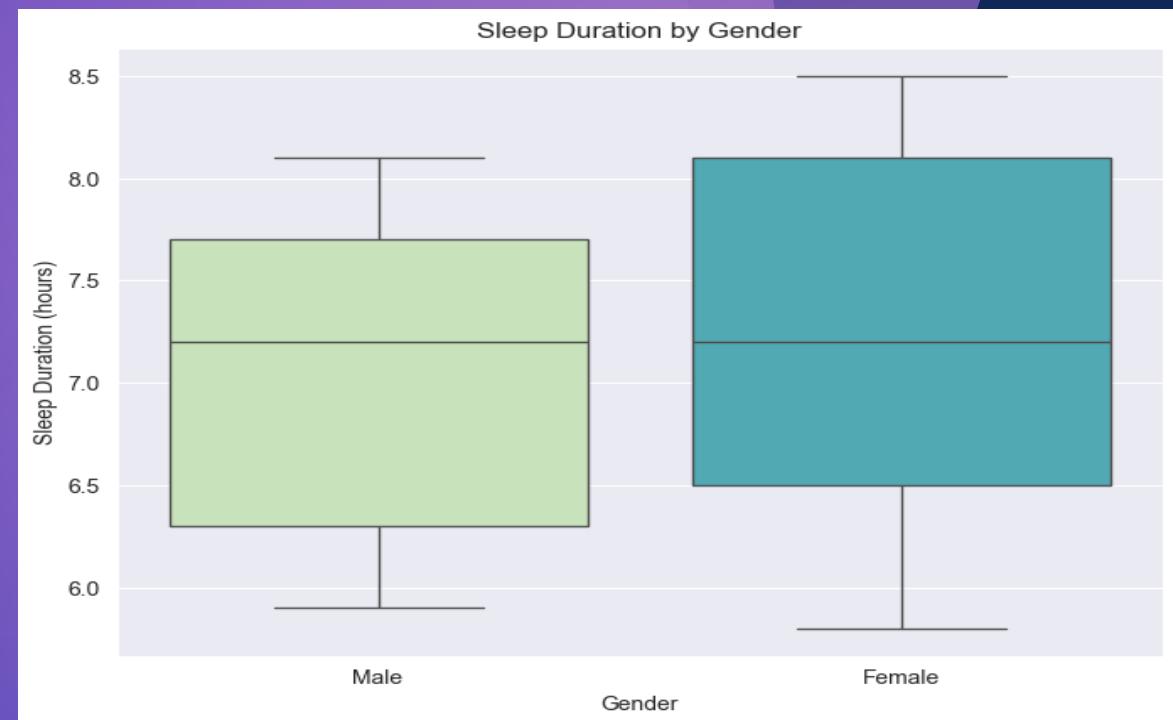
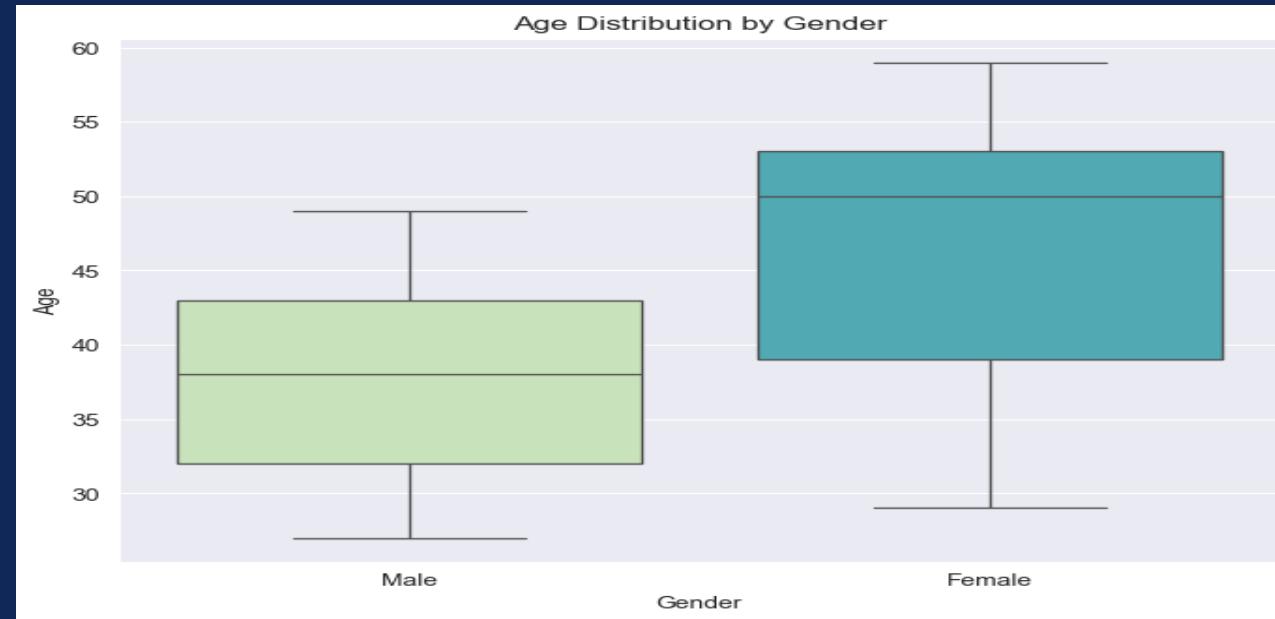
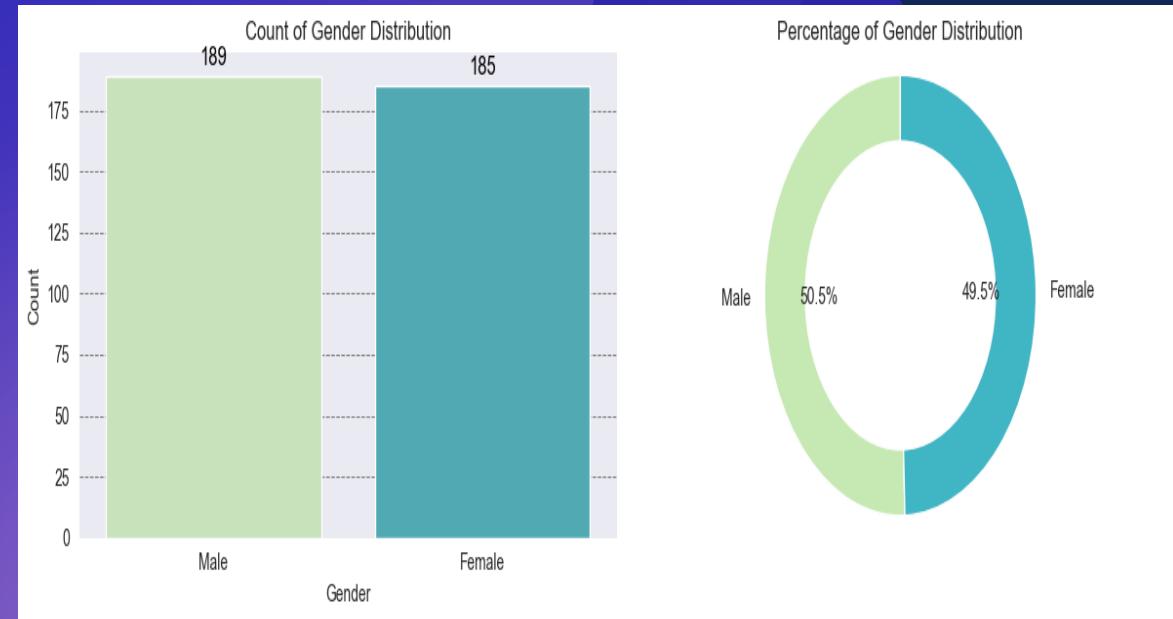


# DATA VISUALIZATION

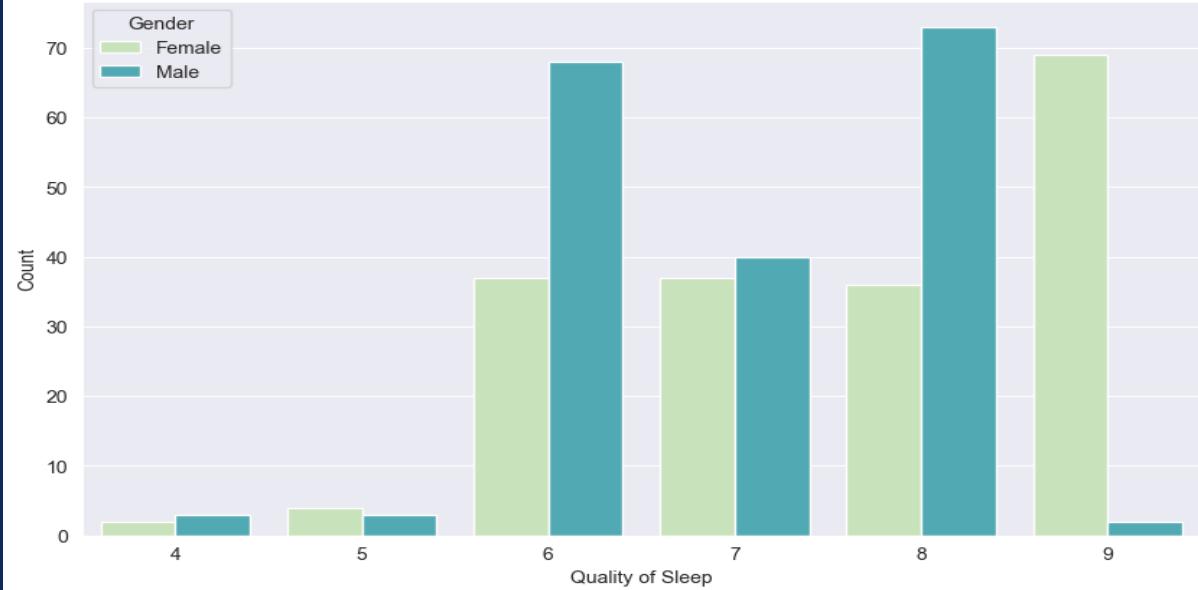
---

# GENDER

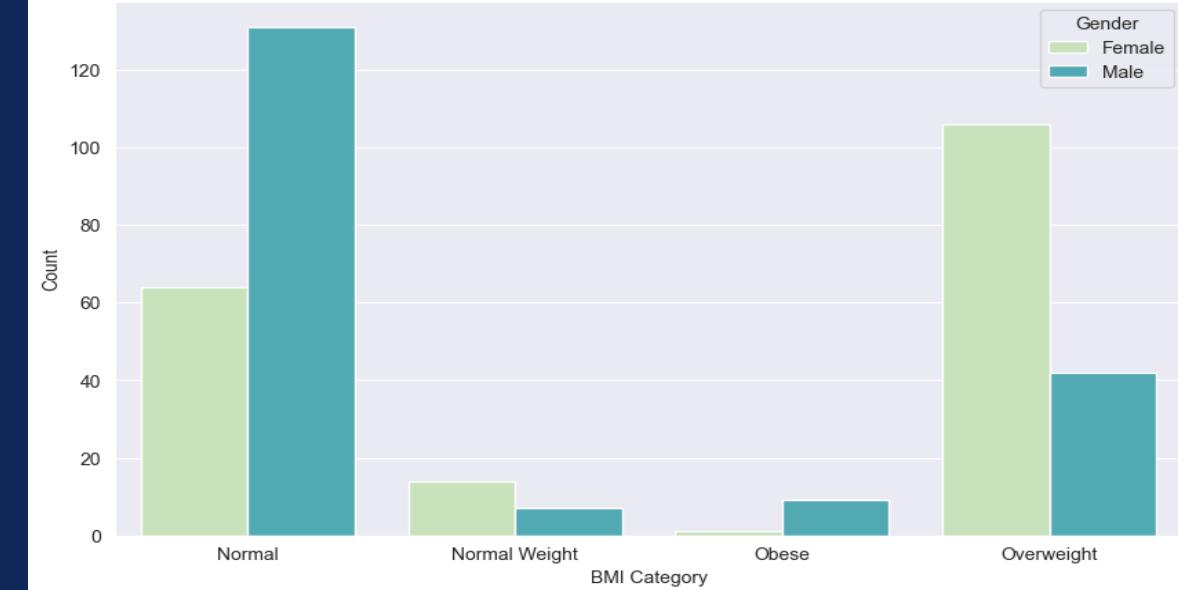
---



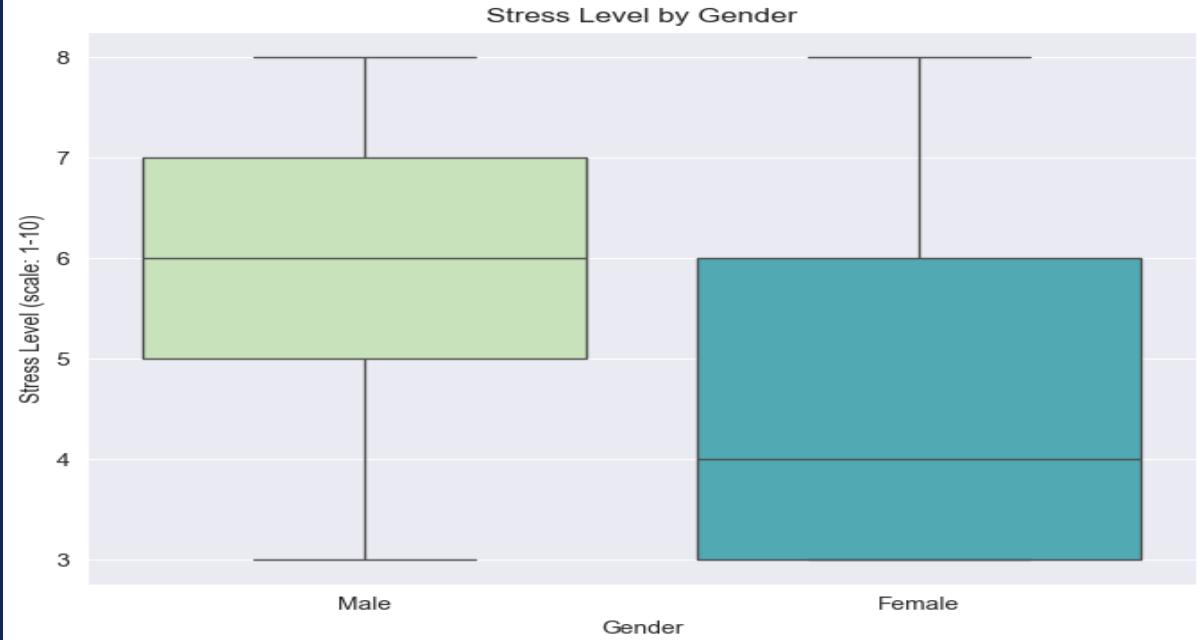
### Quality of Sleep Distribution by Gender



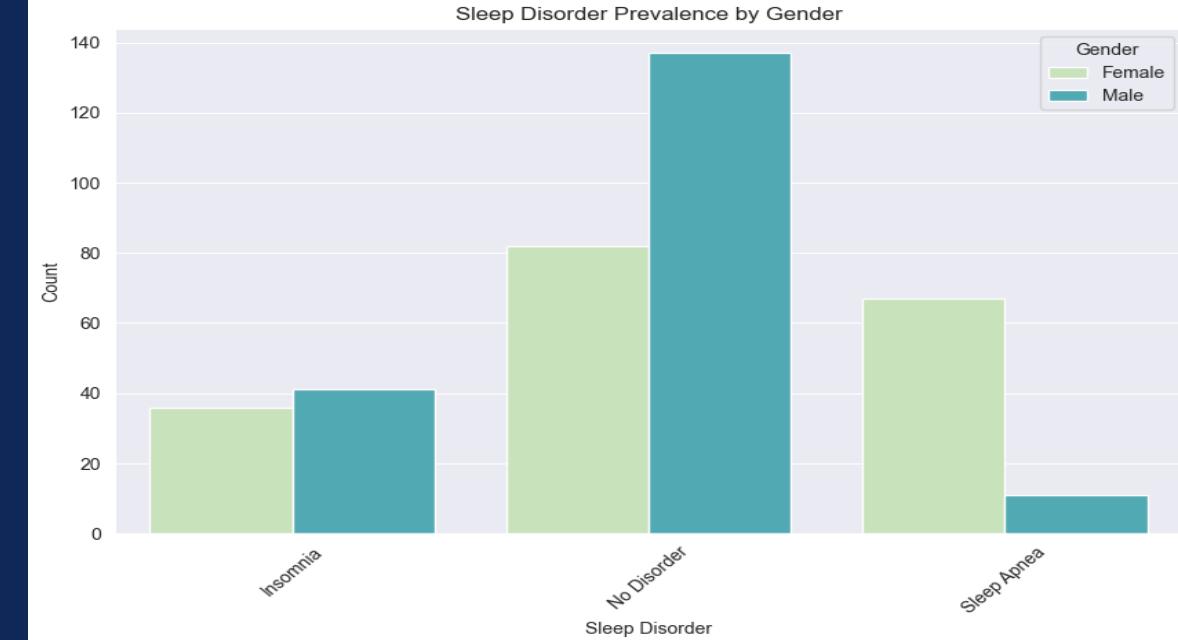
### BMI Category Distribution by Gender



### Stress Level by Gender



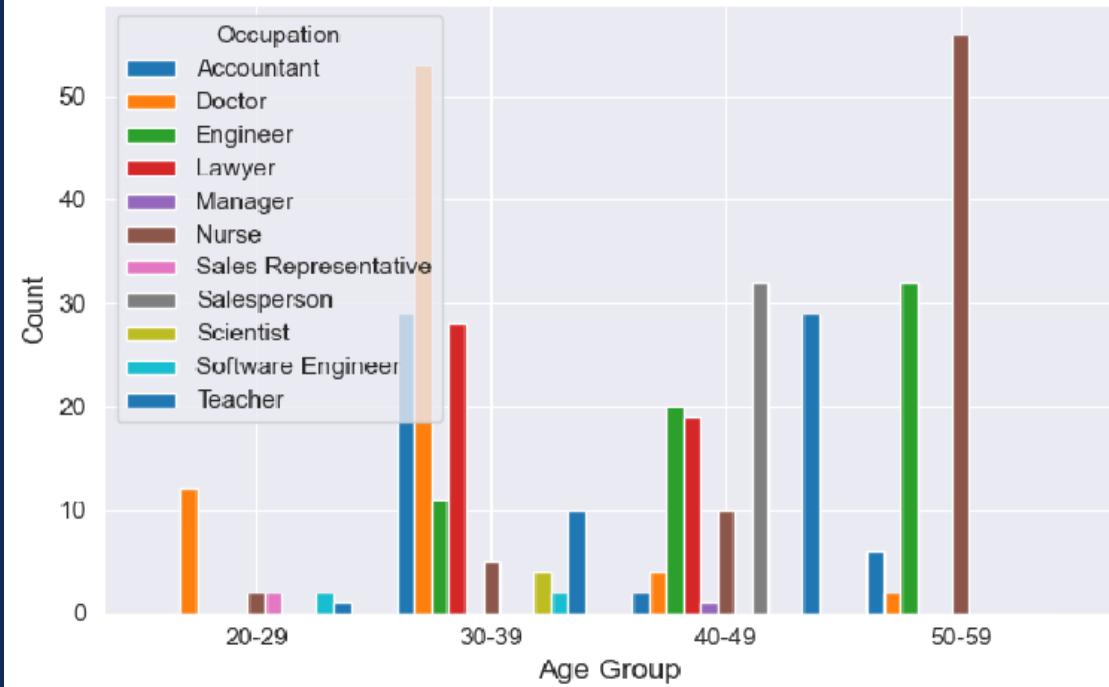
### Sleep Disorder Prevalence by Gender



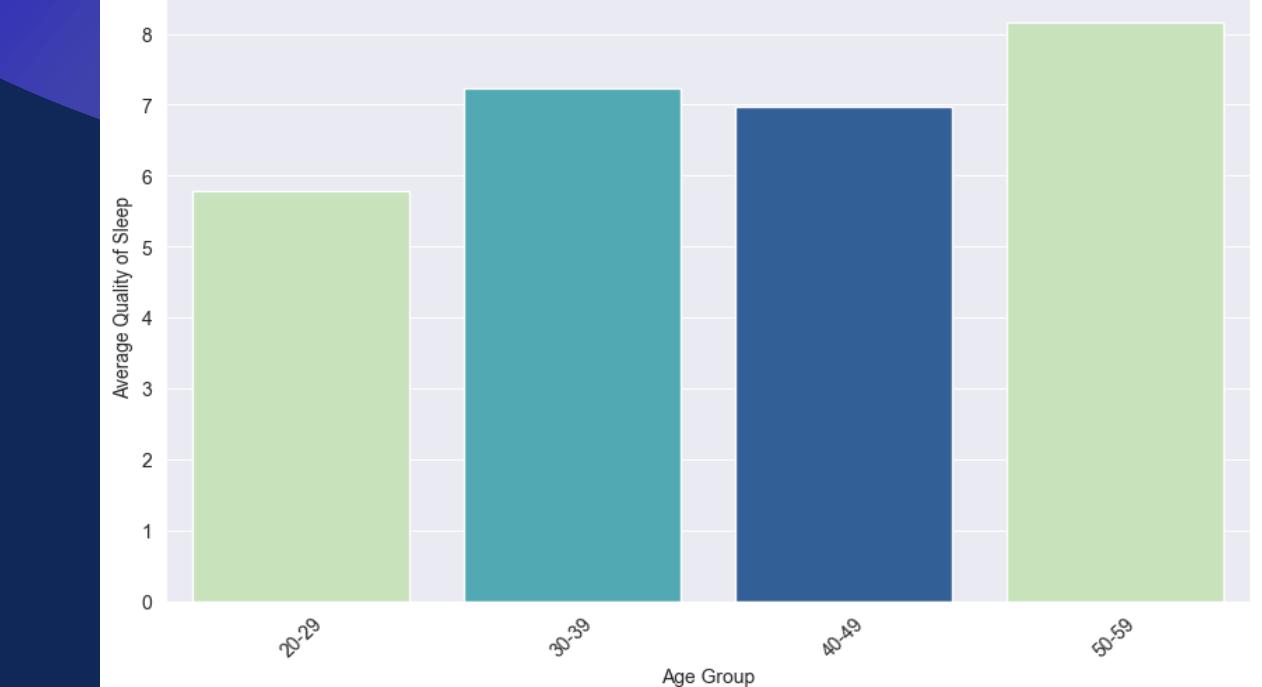
# AGE

---

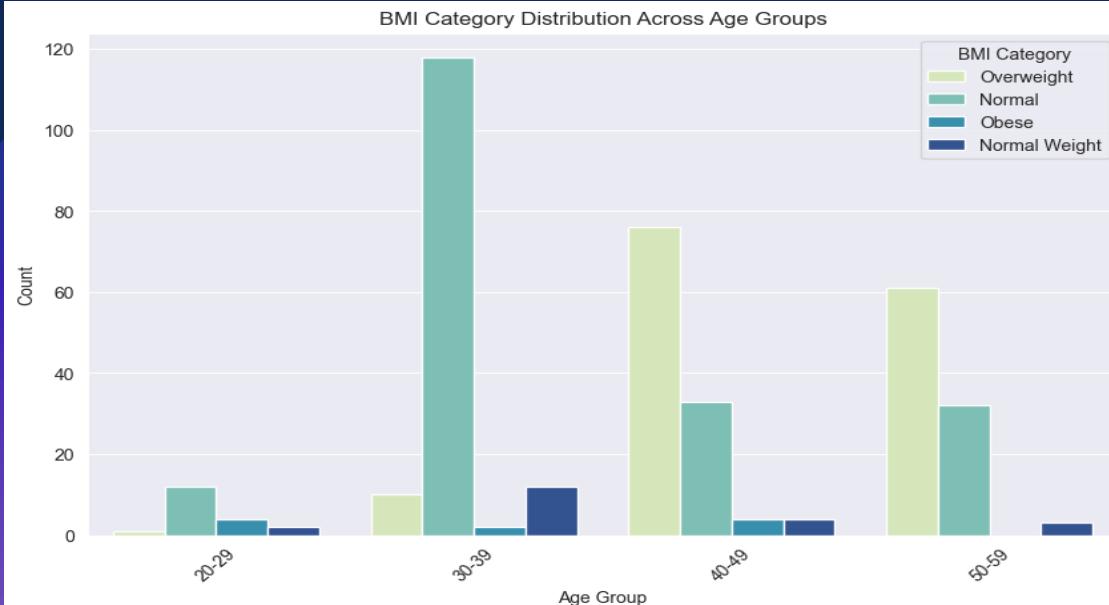
### Most Common Occupations by Age Group



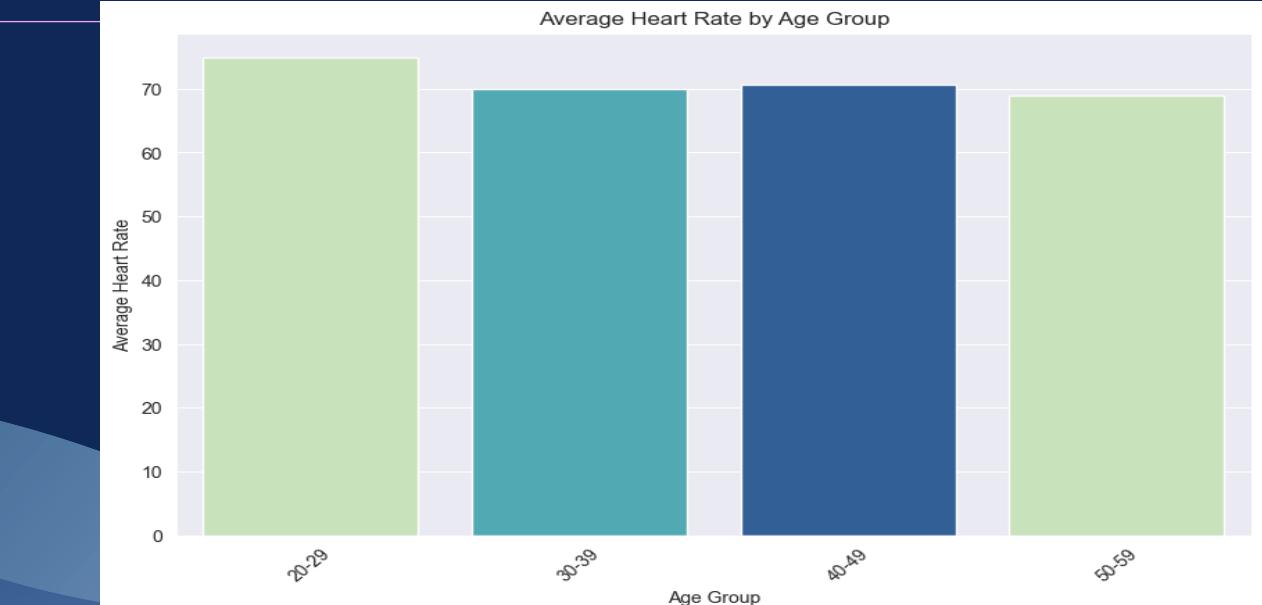
### Average Quality of Sleep by Age Group



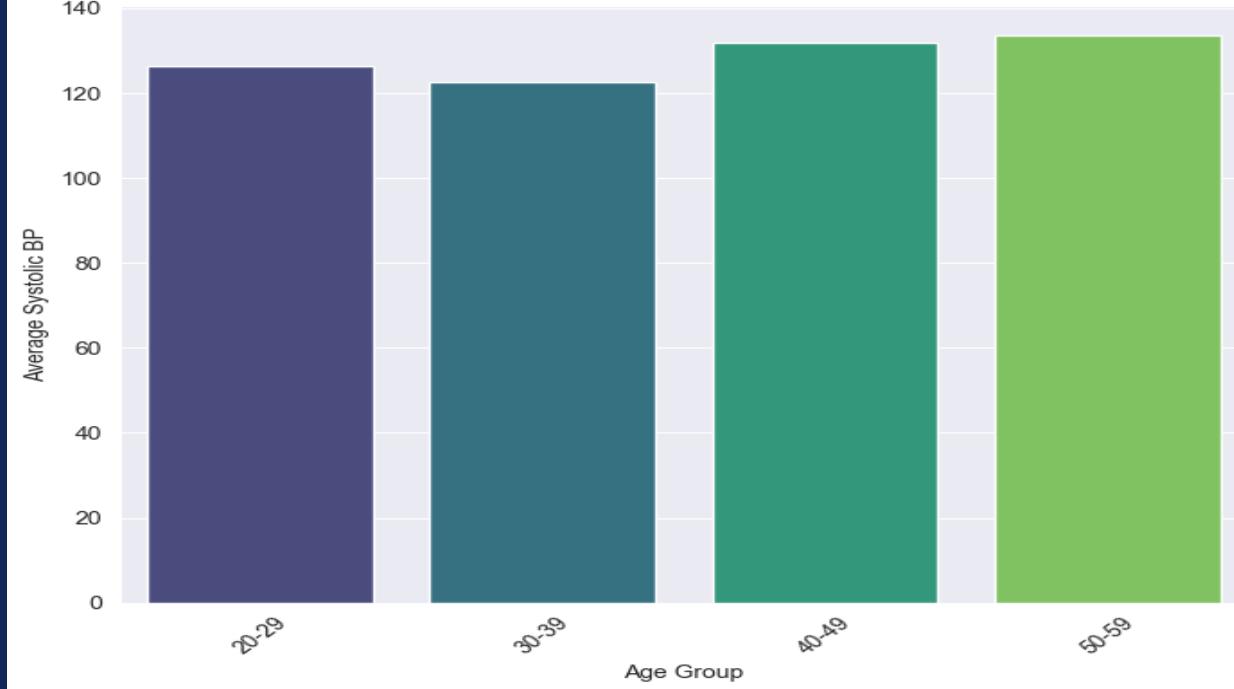
### BMI Category Distribution Across Age Groups



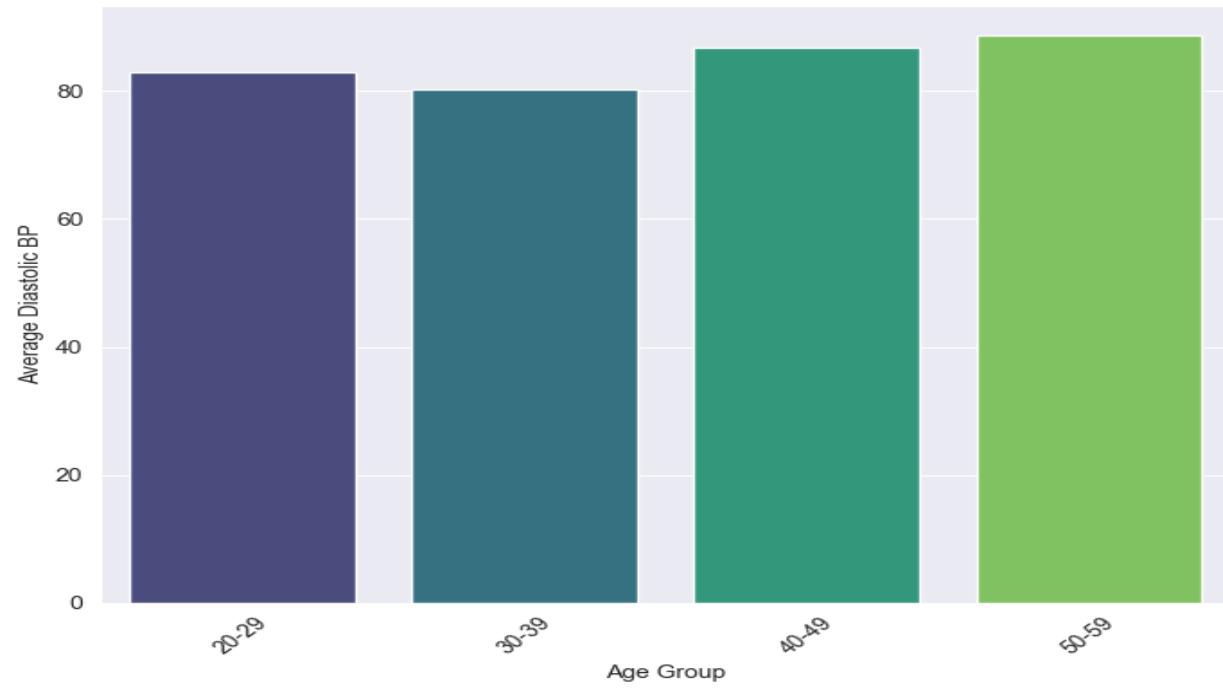
### Average Heart Rate by Age Group



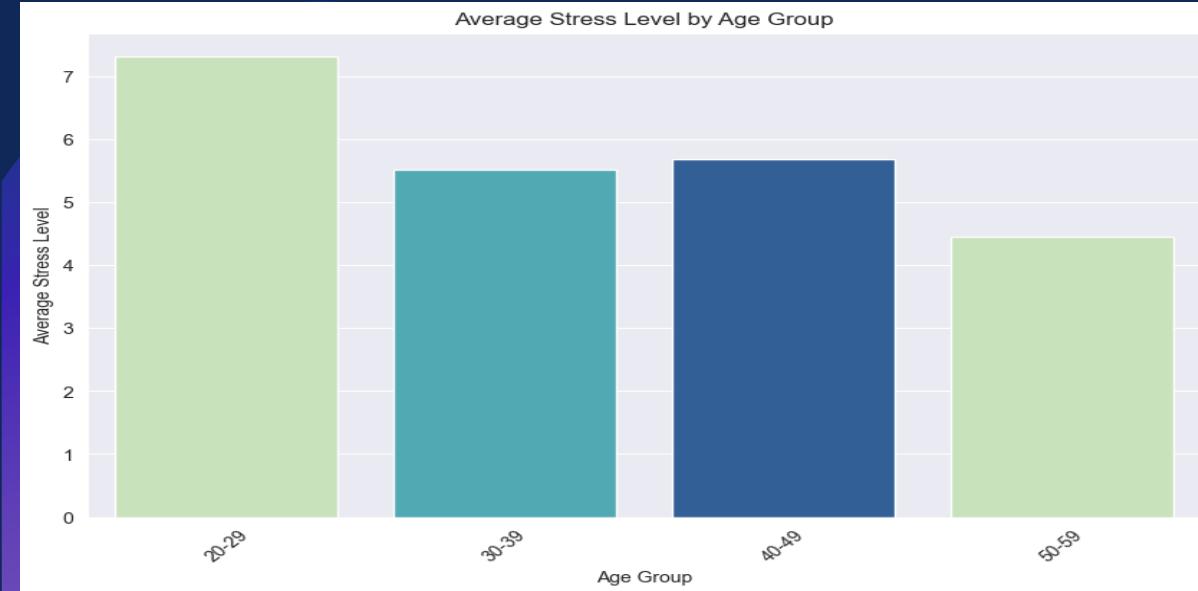
Average Systolic Blood Pressure by Age Group



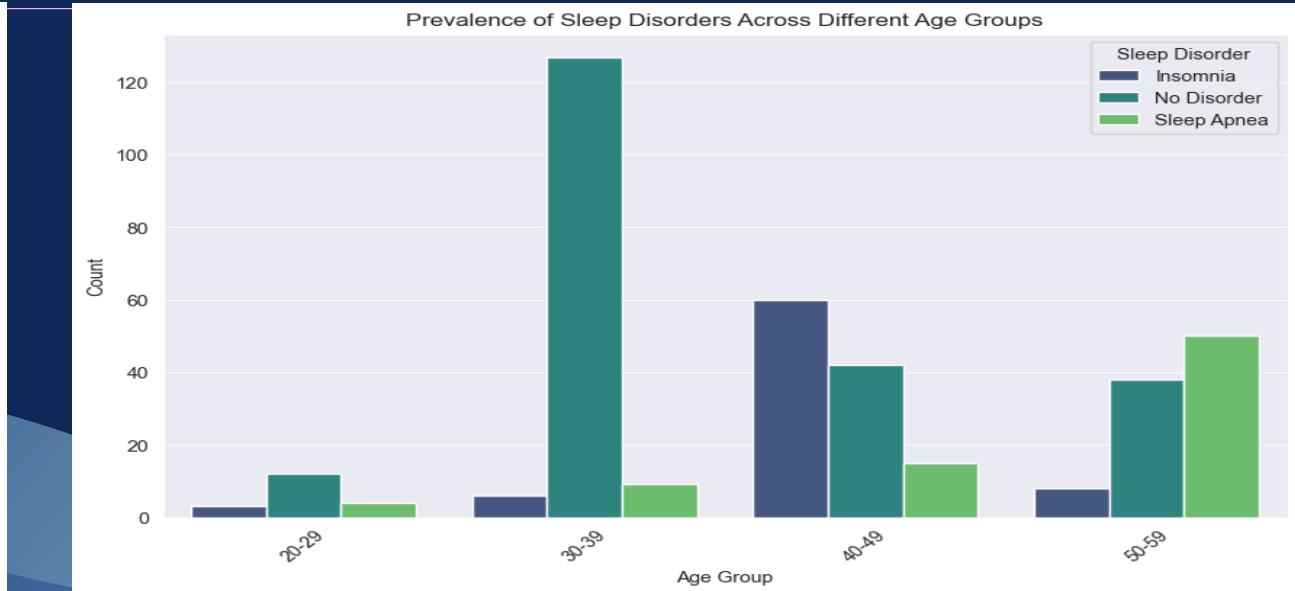
Average Diastolic Blood Pressure by Age Group



Average Stress Level by Age Group



Prevalence of Sleep Disorders Across Different Age Groups

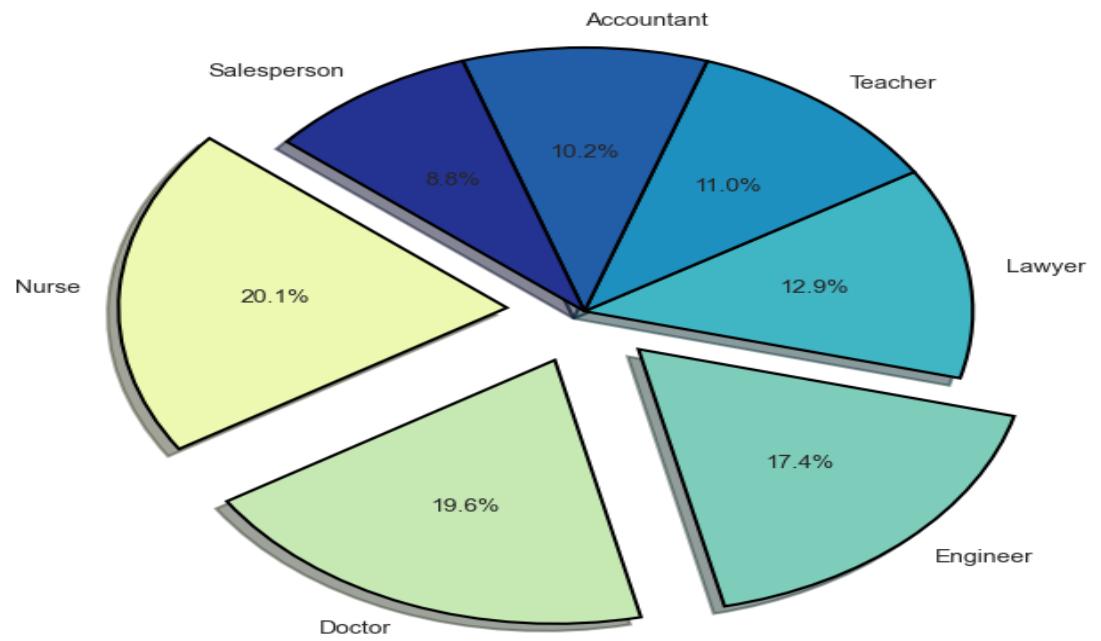


# OCCUPATION

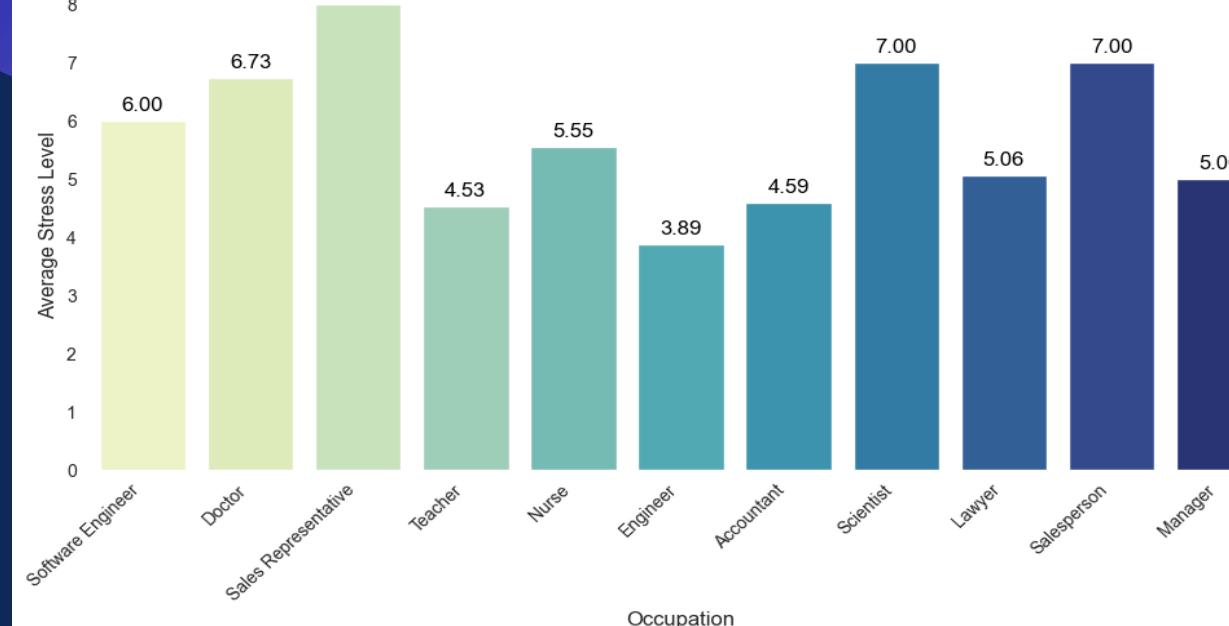
---

TEAM 3

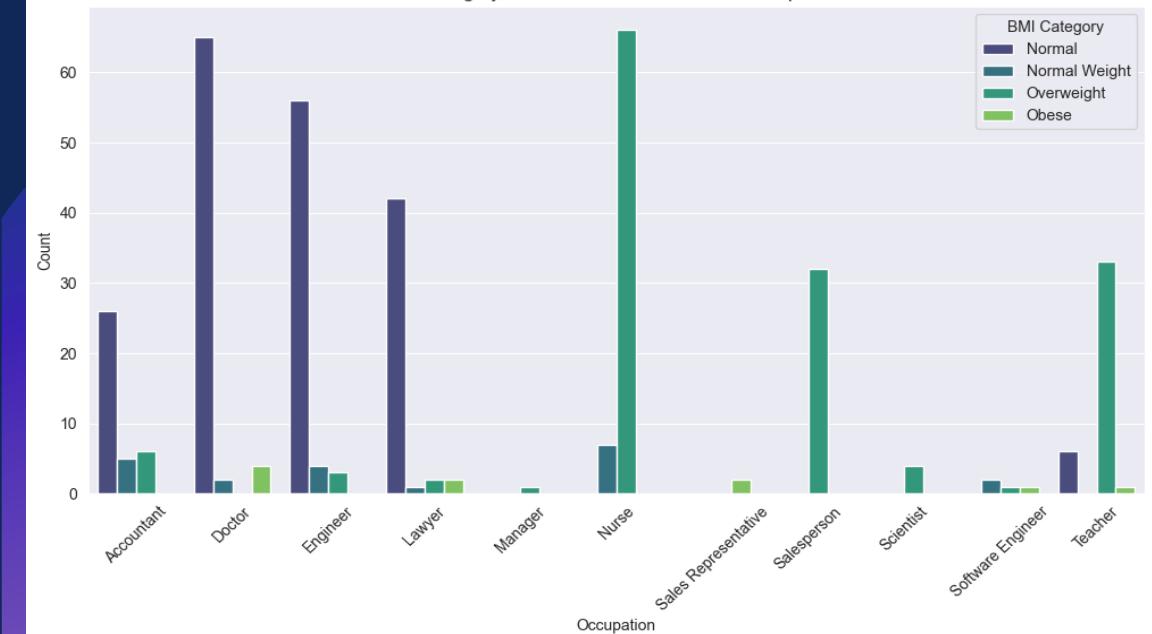
### Occupation Distribution (7 Largest Categories)



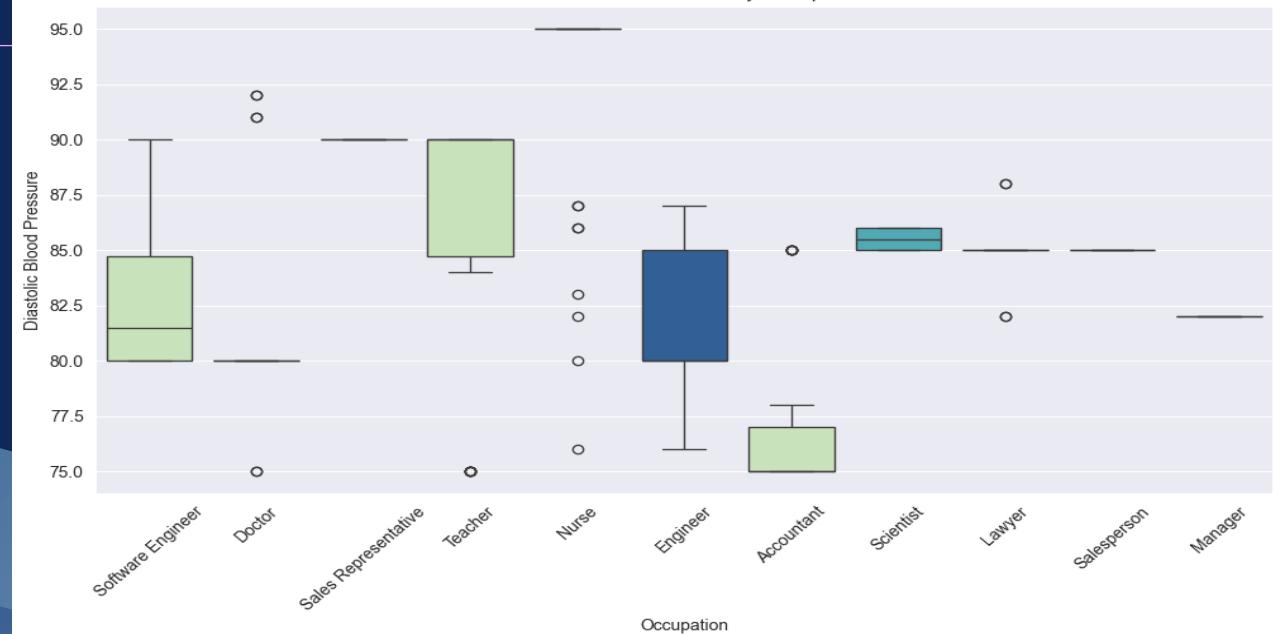
### Relationship Between Occupation and Stress Level

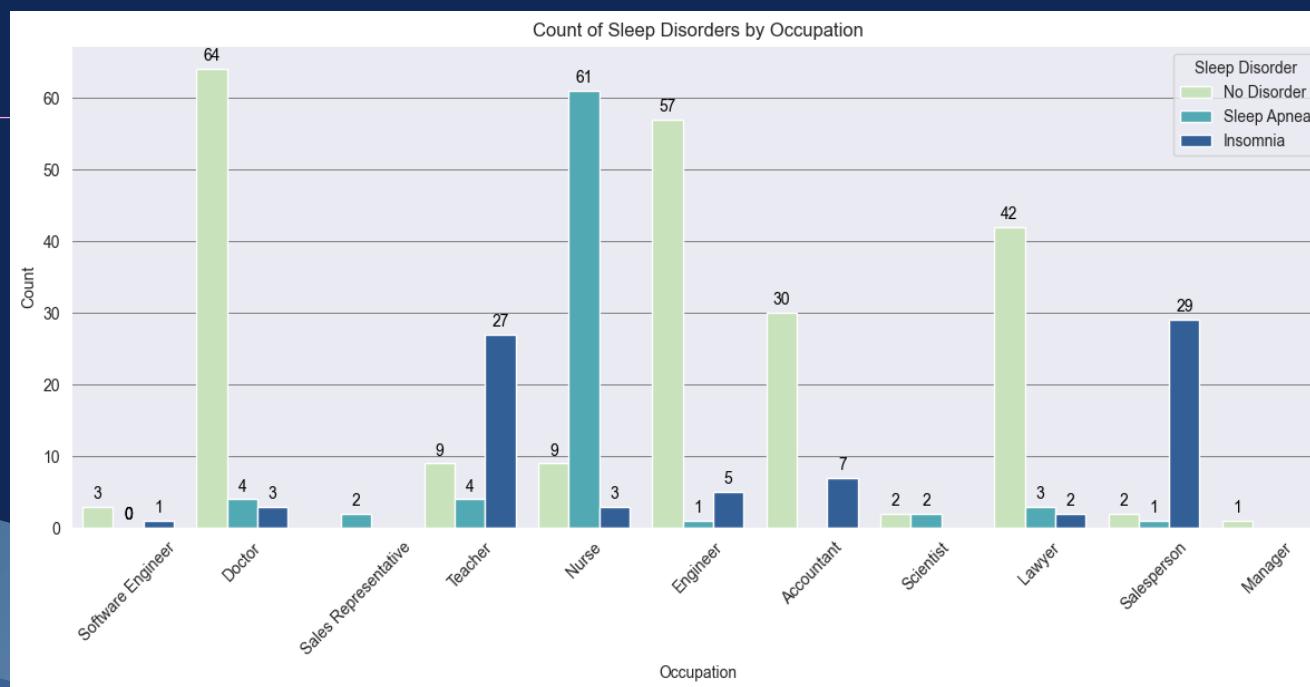
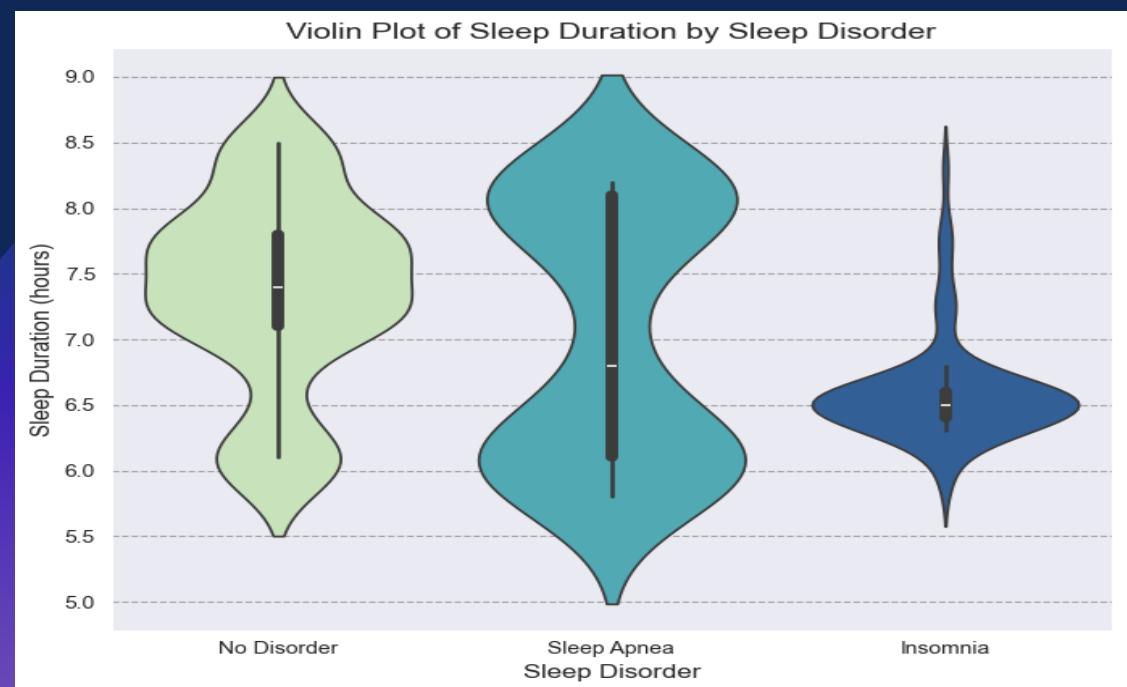
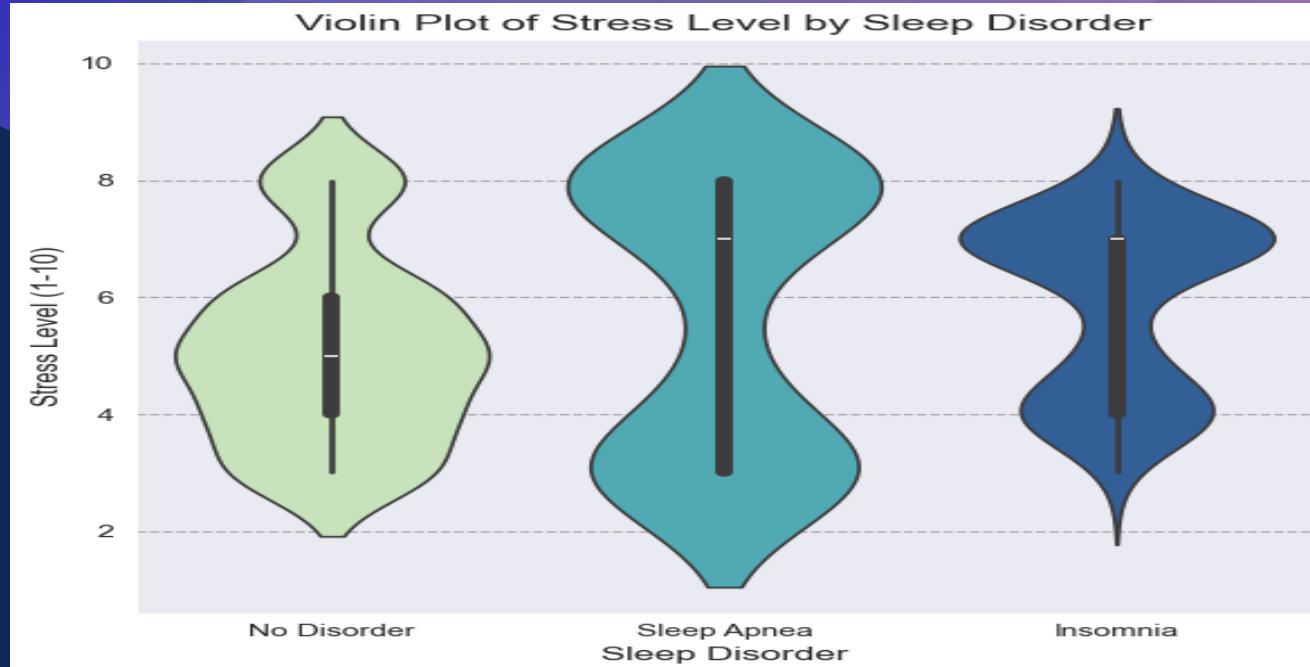
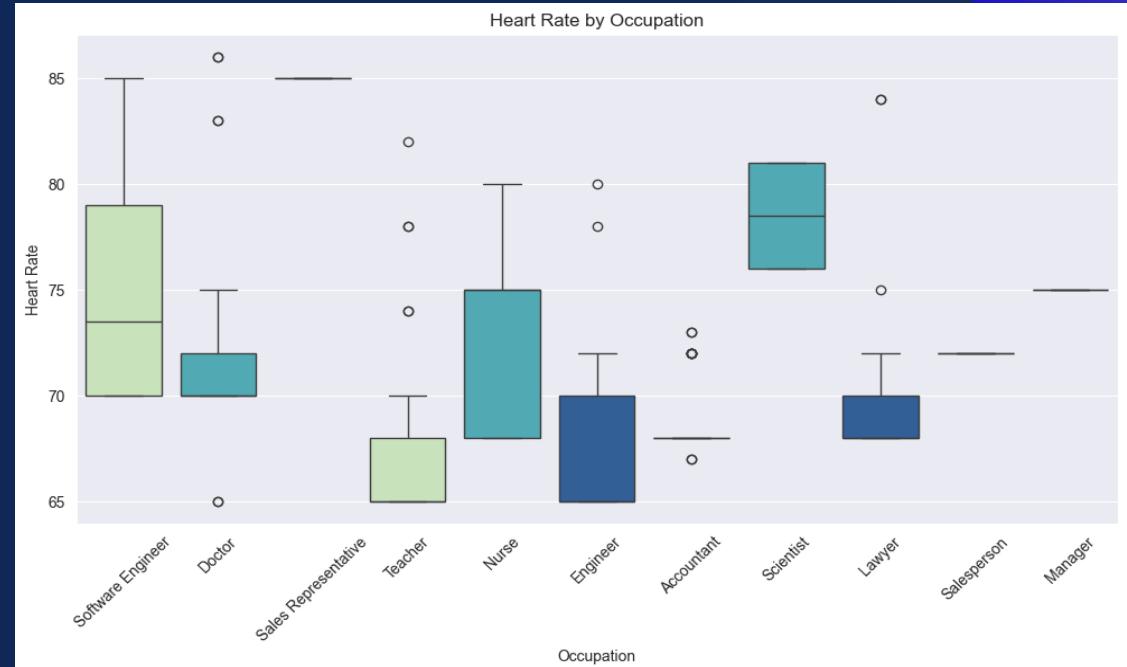


### BMI Category Distribution Across Different Occupations



### Diastolic Blood Pressure by Occupation







# ALGORITHMS AND MACHINE LEARNING

---

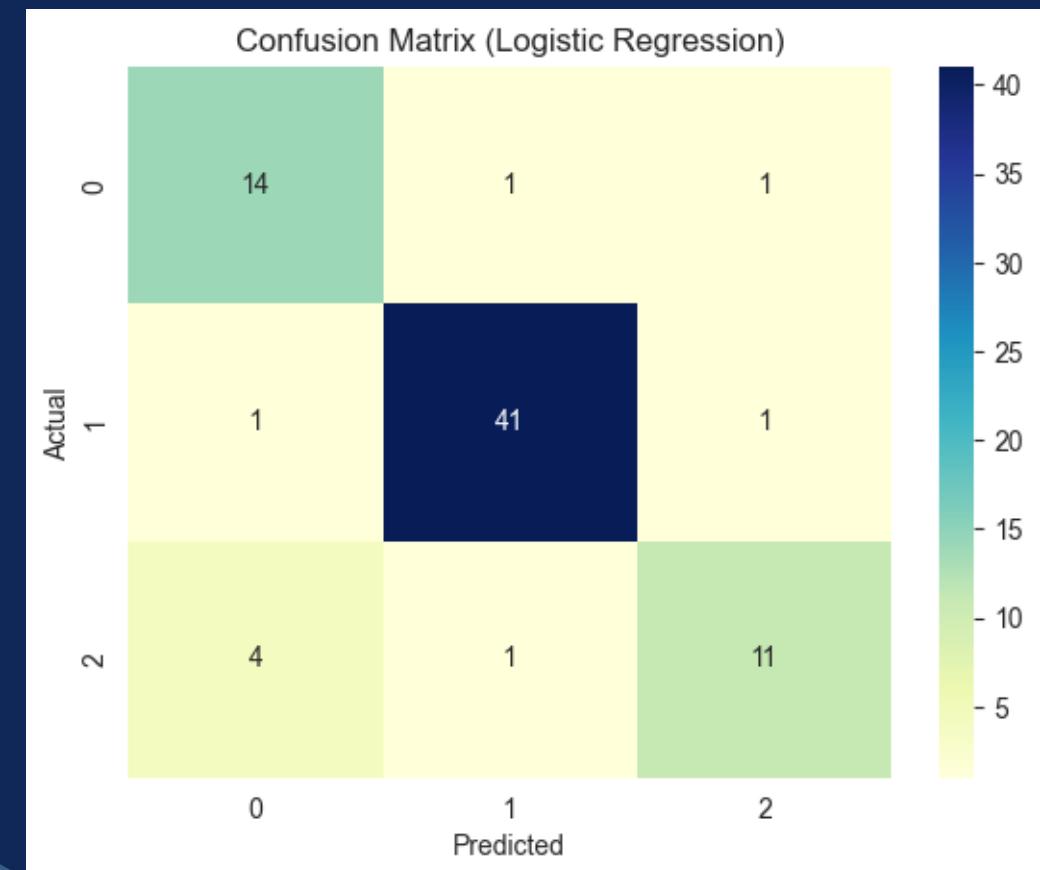
# 1. CORRELATION ANALYSIS:



# 2. PREDICTING SLEEP DISORDER USING LOGISTIC REGRESSION:

Classification Report:

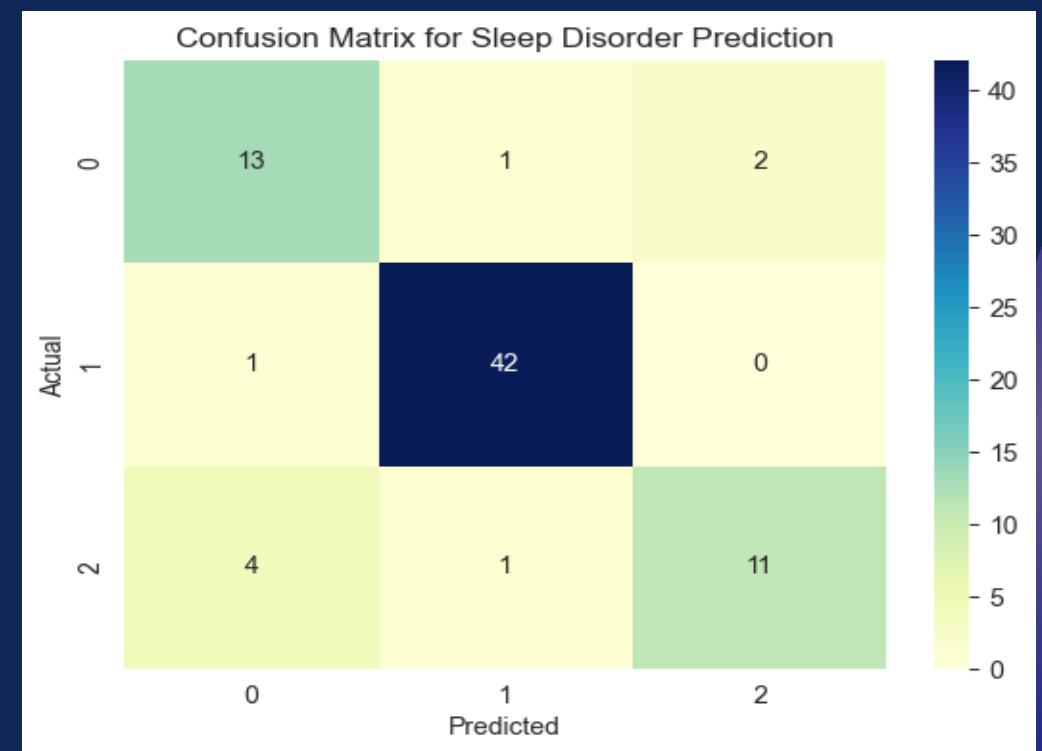
	precision	recall	f1-score	support
0	0.74	0.88	0.80	16
1	0.95	0.95	0.95	43
2	0.85	0.69	0.76	16
accuracy			0.88	75
macro avg	0.85	0.84	0.84	75
weighted avg	0.88	0.88	0.88	75



# 3. PREDICTING SLEEP DISORDER USING RANDOMFOREST:

Classification Report:

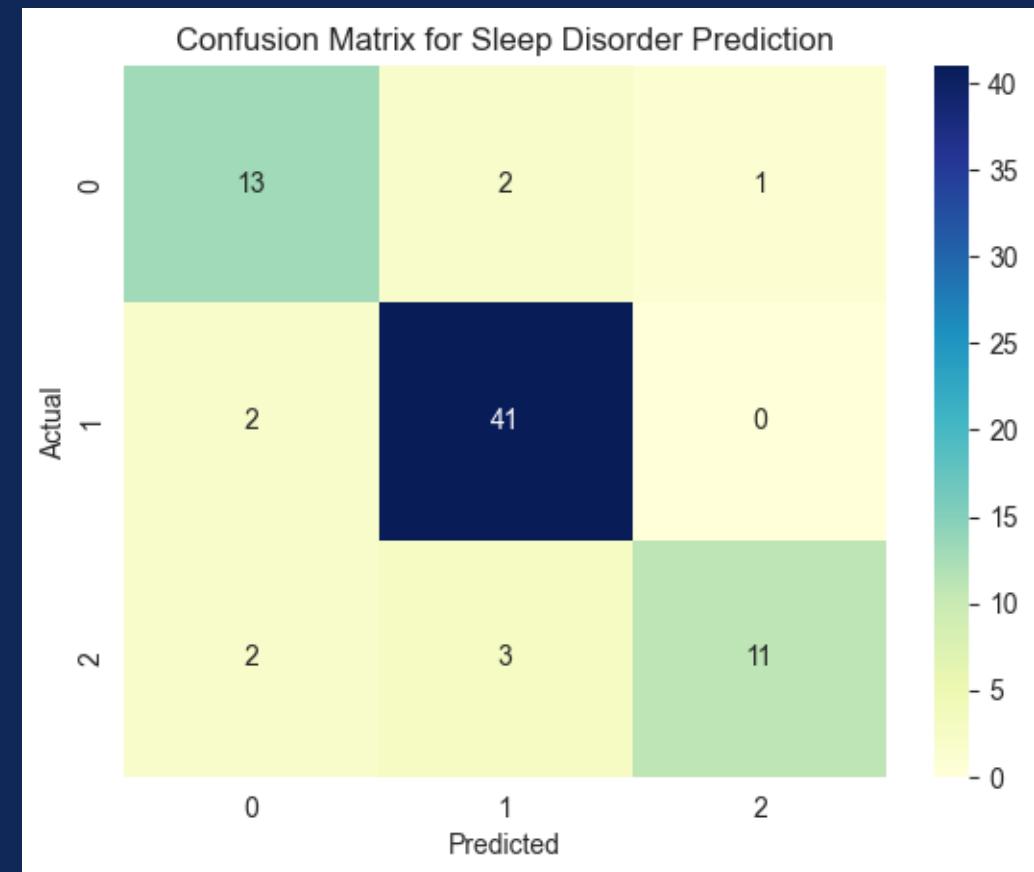
	precision	recall	f1-score	support
0	0.72	0.81	0.76	16
1	0.95	0.98	0.97	43
2	0.85	0.69	0.76	16
accuracy			0.88	75
macro avg	0.84	0.83	0.83	75
weighted avg	0.88	0.88	0.88	75



# 4. PREDICTING SLEEP DISORDER USING SVM:

Classification Report:

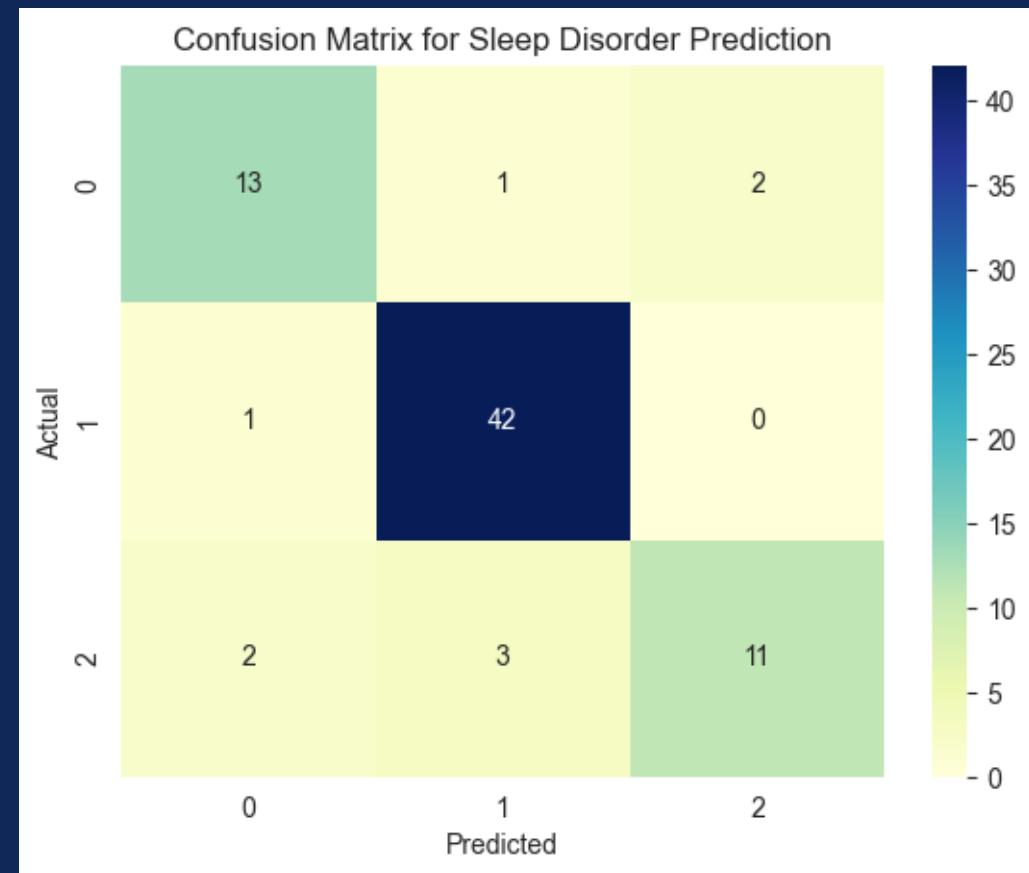
	precision	recall	f1-score	support
0	0.76	0.81	0.79	16
1	0.89	0.95	0.92	43
2	0.92	0.69	0.79	16
accuracy			0.87	75
macro avg	0.86	0.82	0.83	75
weighted avg	0.87	0.87	0.86	75



# 5. PREDICTING SLEEP DISORDER USING DECISION TREE:

Classification Report:

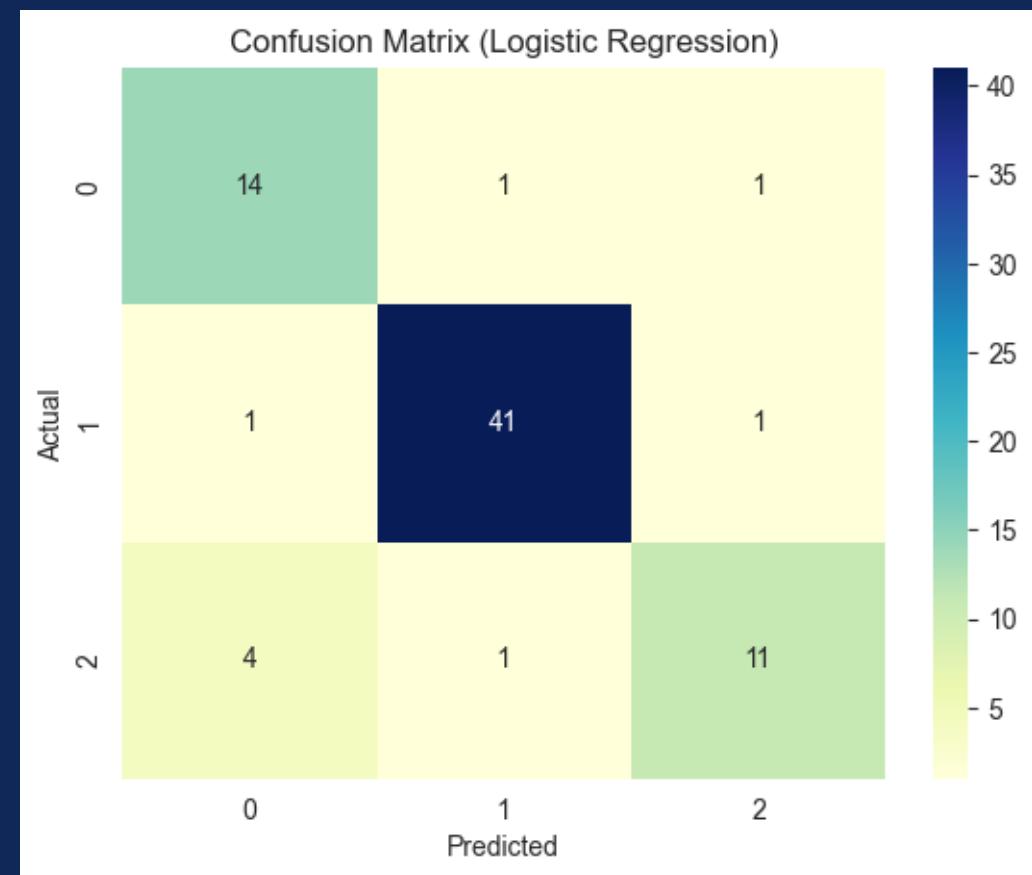
	precision	recall	f1-score	support
0	0.81	0.81	0.81	16
1	0.91	0.98	0.94	43
2	0.85	0.69	0.76	16
accuracy			0.88	75
macro avg	0.86	0.83	0.84	75
weighted avg	0.88	0.88	0.88	75



# 6. PREDICTING STRESS LEVEL USING LINEAR REGRESSION:

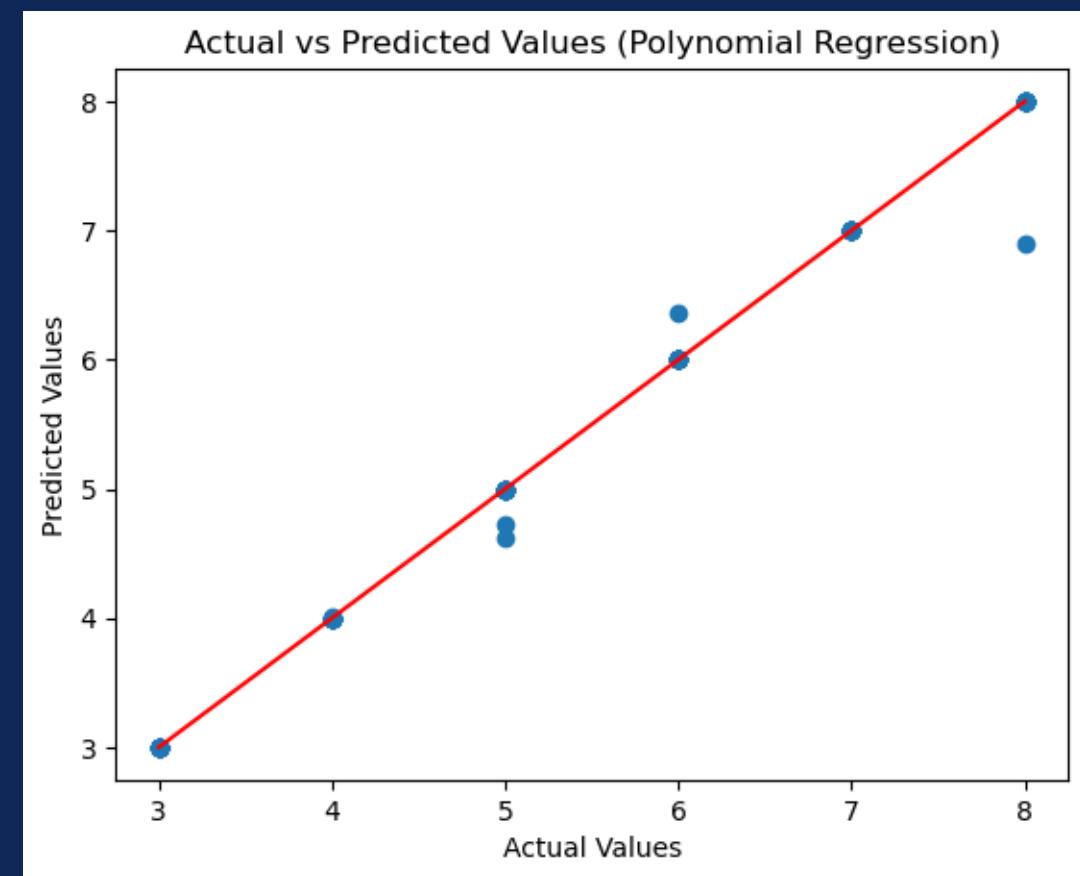
Classification Report:

	precision	recall	f1-score	support
0	0.74	0.88	0.80	16
1	0.95	0.95	0.95	43
2	0.85	0.69	0.76	16
accuracy			0.88	75
macro avg	0.85	0.84	0.84	75
weighted avg	0.88	0.88	0.88	75



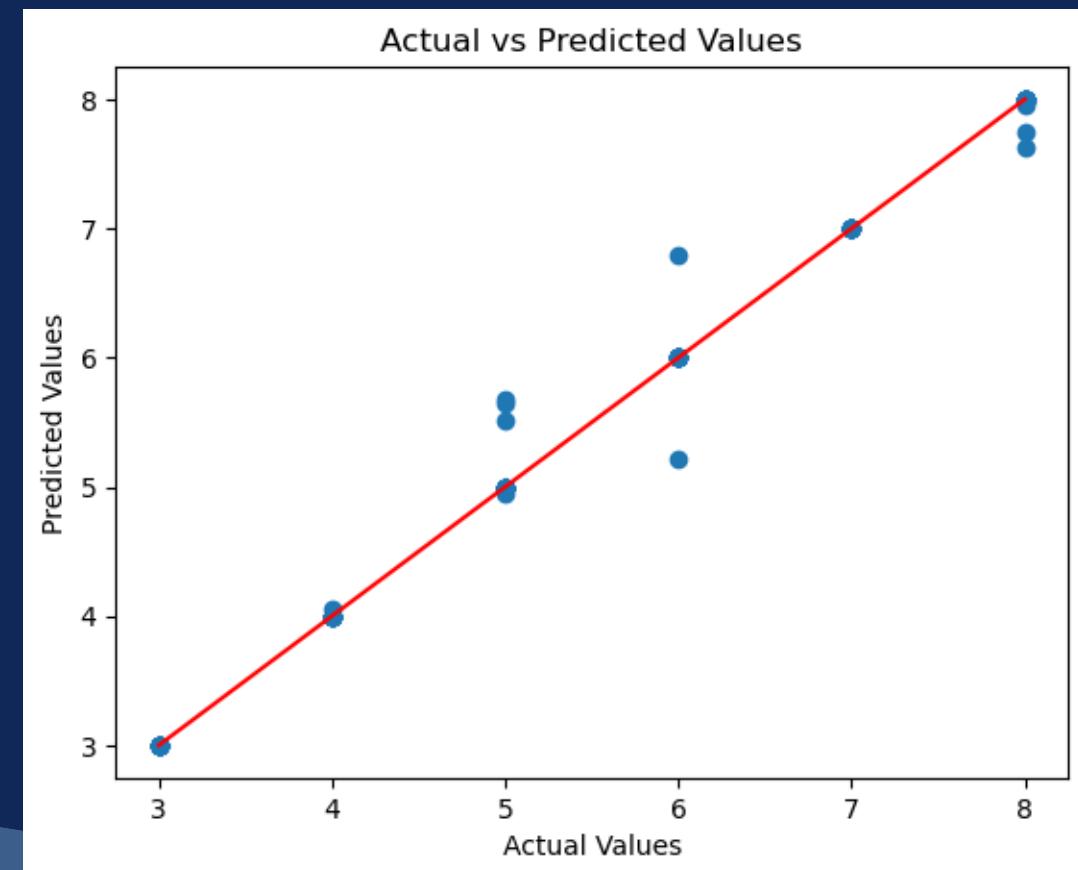
# 7. PREDICTING STRESS LEVEL USING POLYNOMIAL REGRESSION:

Polynomial Regression - Degree 6 -  
RMSE: 0.1435220582002766,  $R^2$   
Score: 0.9934069068399912



# 8. PREDICTING STRESS LEVEL USING RANDOMFORESTREGRESSOR:

Random Forest Regressor - RMSE:  
0.18657974166559455, MAE:  
0.0569333333333333, R<sup>2</sup> Score:  
0.9888575452372823

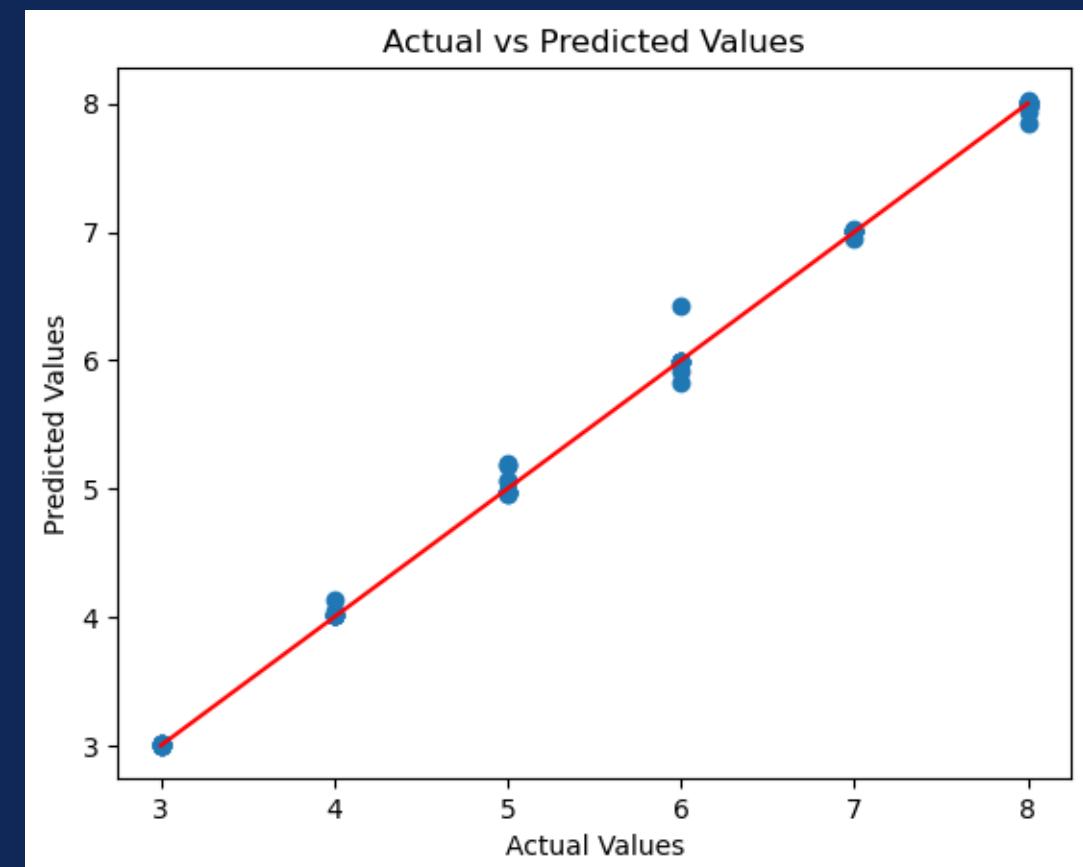


# 9. PREDICTING STRESS LEVEL USING GRADIENT BOOSTING REGRESSOR:

RMSE: 0.07242095377247026

Mean Absolute Error: 0.034520665442582976

R<sup>2</sup> Score: 0.9983212718039495

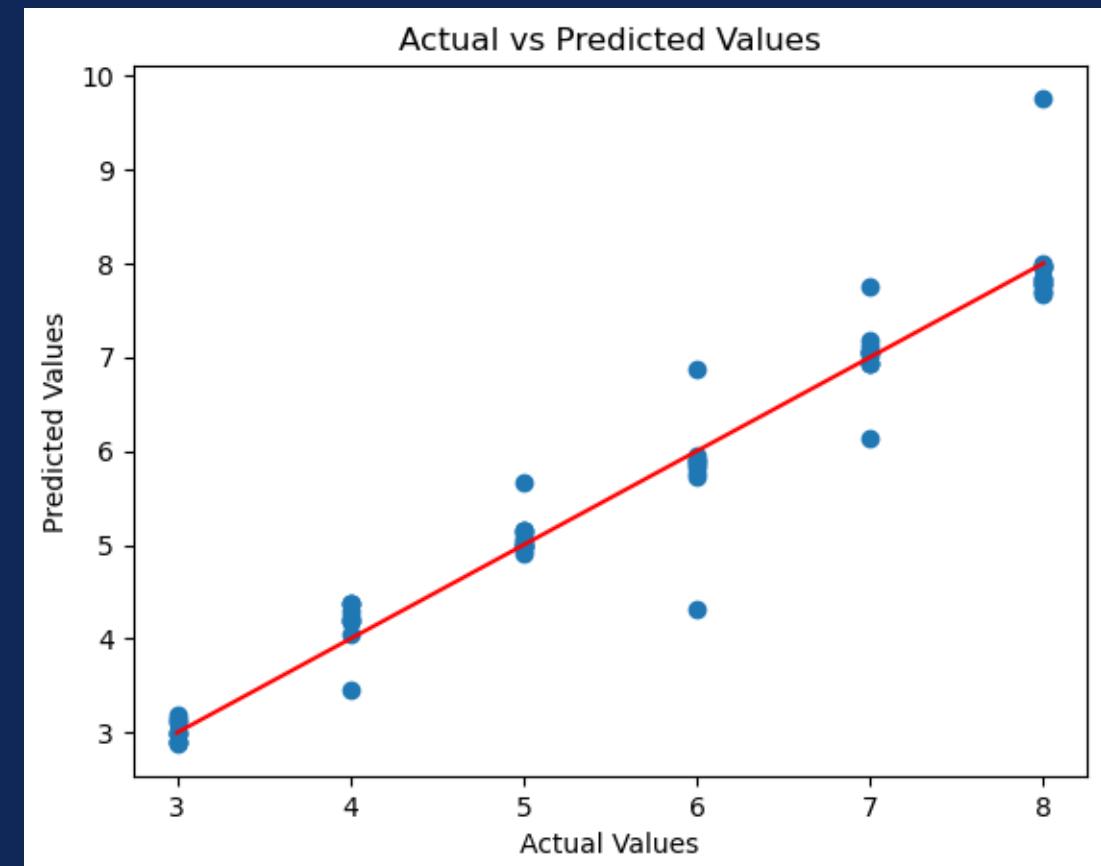


# 10. PREDICTING STRESS LEVEL USING LINEAR REGRESSION:

RMSE: 0.3736444666719237

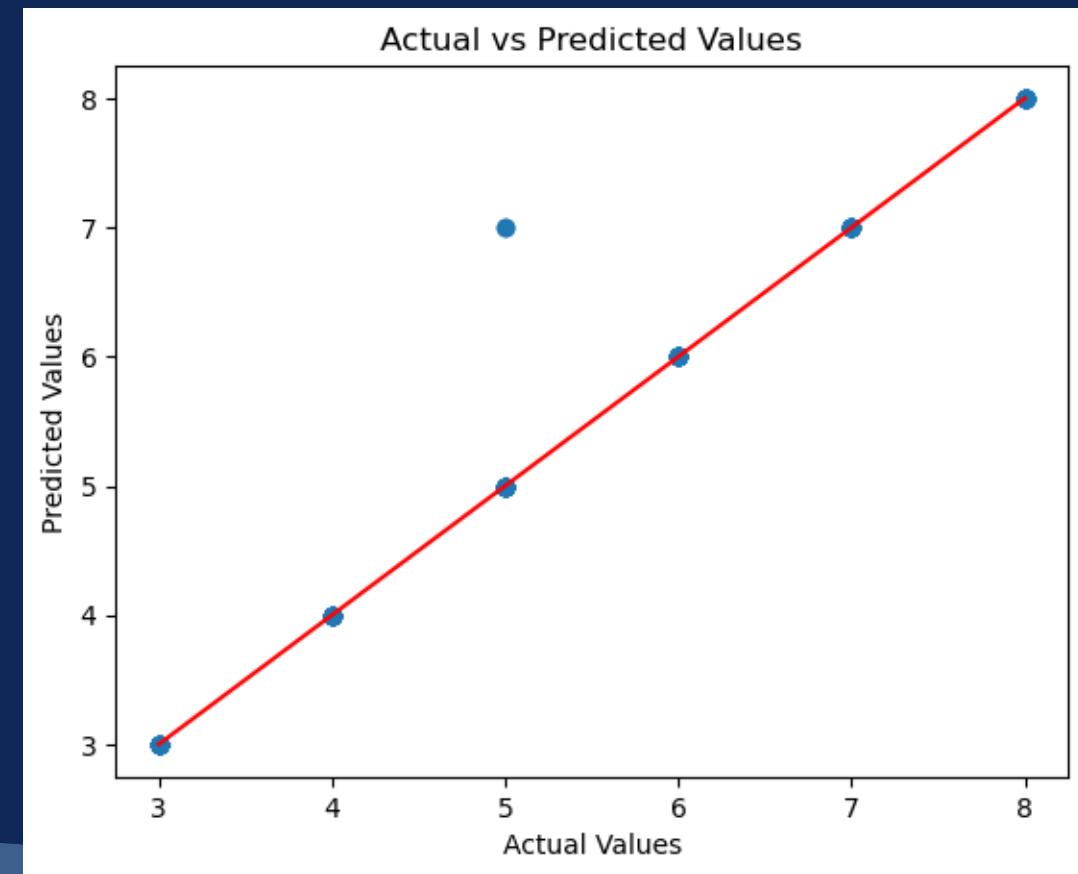
Mean Absolute Error: 0.20884365178983944

R<sup>2</sup> Score: 0.9553142537530259



# 11. PREDICTING STRESS LEVEL USING DECISION TREE:

Decision Tree Regressor -  
RMSE: 0.32659863237109044,  
MAE: 0.05333333333333334,  
R<sup>2</sup> Score: 0.9658586548310003



Model	Task	Precision	Recall	F1-Score	Accuracy	RMSE	MAE	R <sup>2</sup> Score
RandomForest	Sleep Disorder Prediction	0.72 (Class 0)	0.81 (Class 0)	0.76 (Class 0)	0.88	-	-	-
		0.95 (Class 1)	0.98 (Class 1)	0.97 (Class 1)				
		0.85 (Class 2)	0.69 (Class 2)	0.76 (Class 2)				
SVM	Sleep Disorder Prediction	0.76 (Class 0)	0.81 (Class 0)	0.79 (Class 0)	0.87	-	-	-
		0.89 (Class 1)	0.95 (Class 1)	0.92 (Class 1)				
		0.92 (Class 2)	0.69 (Class 2)	0.79 (Class 2)				
Decision Tree	Sleep Disorder Prediction	0.81 (Class 0)	0.81 (Class 0)	0.81 (Class 0)	0.88	-	-	-
		0.91 (Class 1)	0.98 (Class 1)	0.94 (Class 1)				
		0.85 (Class 2)	0.69 (Class 2)	0.76 (Class 2)				
Linear Regression	Stress Level Prediction	-	-	-	-	0.3736	0.2088	0.9553
Polynomial Regression (Degree 6)	Stress Level Prediction	-	-	-	-	0.1435	-	0.9934
RandomForestR	Stress Level egressor	Stress Level Prediction	-	-	-	0.1866	0.0569	0.9889
GradientBoostingRegressor	Stress Level Prediction	-	-	-	-	0.0724	0.0345	0.9983
Decision Tree Regressor	Stress Level Prediction	-	-	-	-	0.3266	0.0533	0.9659

## o Sleep Disorder Prediction:

RandomForest and Decision Tree performed similarly with an accuracy of 88%.

SVM had slightly lower accuracy (87%) but performed well in precision and recall for Class 1.

All models struggled with Class 2 (lower recall), indicating room for improvement.

## o Stress Level Prediction:

GradientBoostingRegressor outperformed all other models with the lowest RMSE (0.0724) and highest  $R^2$  Score (0.9983).

Polynomial Regression also performed well with an  $R^2$  Score of 0.9934.

Linear Regression had the highest RMSE (0.3736) and lowest  $R^2$  Score (0.9553), making it the least accurate for this task.

# SUMMARY

---

The Sleep Health and Lifestyle Dataset contains data from 375 individuals, focusing on sleep patterns, lifestyle, and cardiovascular health. It includes variables like demographics, sleep duration, quality, physical activity, stress levels, and sleep disorders (Insomnia, Sleep Apnea). The dataset aims to explore relationships between sleep disorders, cardiovascular health, and lifestyle factors. It also seeks to identify risk factors for sleep disorders and develop predictive models for health outcomes. This dataset is valuable for understanding how sleep and lifestyle impact overall health.



THANK YOU

---