

ANALYSIS OF DISORDER DATASET SLEEP

An Exploratory and Predictive Study

Prepared by: Sahar Alshahrani

Data Science department

December 2024

Instructor: Dr. Ayman Qahmash



INTRODUCTION

- This statistical project aims to study the factors affecting sleep quality, which is a crucial element for overall health and daily performance. The project analyses diverse data, including age, gender, physical activity, stress level, and blood pressure, to achieve accurate predictions and clear insights into how these factors influence sleep quality. The findings will help identify the most impactful factors, enabling the development of effective recommendations to improve sleep patterns.



OBJECTIVES

- Explore and clean the Sleep Quality dataset.
- Conduct descriptive analysis of attributes.
- Investigate relationships between attributes and their effect on sleep quality.
- Build and evaluate predictive models.
- Identify key factors influencing sleep quality.



EXPLORE AND CLEAN THE DISORDER SLEEP DATASET

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	NaN
2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
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	NaN
9	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	NaN
10	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	NaN



EXPLORE AND CLEAN THE DISORDER SLEEP DATASET

	Person ID	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate	Daily Steps
count	373.000000	373.000000	373.000000	373.000000	373.000000	373.000000	373.000000	373.000000
mean	187.000000	42.139410	7.129491	7.308311	59.128686	5.391421	70.171582	6816.353887
std	107.820066	8.640793	0.795139	1.195359	20.842589	1.772590	4.139704	1620.060932
min	1.000000	27.000000	5.800000	4.000000	30.000000	3.000000	65.000000	3000.000000
25%	94.000000	35.000000	6.400000	6.000000	45.000000	4.000000	68.000000	5600.000000
50%	187.000000	43.000000	7.200000	7.000000	60.000000	5.000000	70.000000	7000.000000
75%	280.000000	50.000000	7.800000	8.000000	75.000000	7.000000	72.000000	8000.000000
max	373.000000	59.000000	8.500000	9.000000	90.000000	8.000000	86.000000	10000.000000



EXPLORE AND CLEAN THE DISORDER SLEEP DATASET

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 373 entries, 0 to 372
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Person ID                            373 non-null    int64
1   Gender                               373 non-null    object
2   Age                                   373 non-null    int64
3   Occupation                           373 non-null    object
4   Sleep Duration                       373 non-null    float64
5   Quality of Sleep                     373 non-null    int64
6   Physical Activity Level              373 non-null    int64
7   Stress Level                         373 non-null    int64
8   BMI Category                         373 non-null    object
9   Blood Pressure                       373 non-null    object
10  Heart Rate                           373 non-null    int64
11  Daily Steps                          373 non-null    int64
12  Sleep Disorder                       154 non-null    object
dtypes: float64(1), int64(7), object(5)
memory usage: 38.0+ KB
None
```



EXPLORE AND CLEAN THE DISORDER SLEEP DATASET

```
#Missing Value: Replase nan values to (Unknown)
df['Sleep Disorder'] = df['Sleep Disorder'].fillna('No Sleep Disorde')
print(df['Sleep Disorder'])
```

✓ 0.0s

```
0      No Sleep Disorde
1      No Sleep Disorde
2      No Sleep Disorde
3          Sleep Apnea
4          Sleep Apnea
...
368      Sleep Apnea
369      Sleep Apnea
370      Sleep Apnea
371      Sleep Apnea
372      Sleep Apnea
```

```
Name: Sleep Disorder, Length: 373, dtype: object
```



DATASET OVERVIEW

The dataset contains the following columns:

- Person ID (Unique Identifier)
- Gender, Age, Occupation
- Sleep Duration, Quality of Sleep, Sleep Disorder
- Physical Activity Level, Stress Level
- BMI Category, Blood Pressure, Heart Rate, Daily Steps

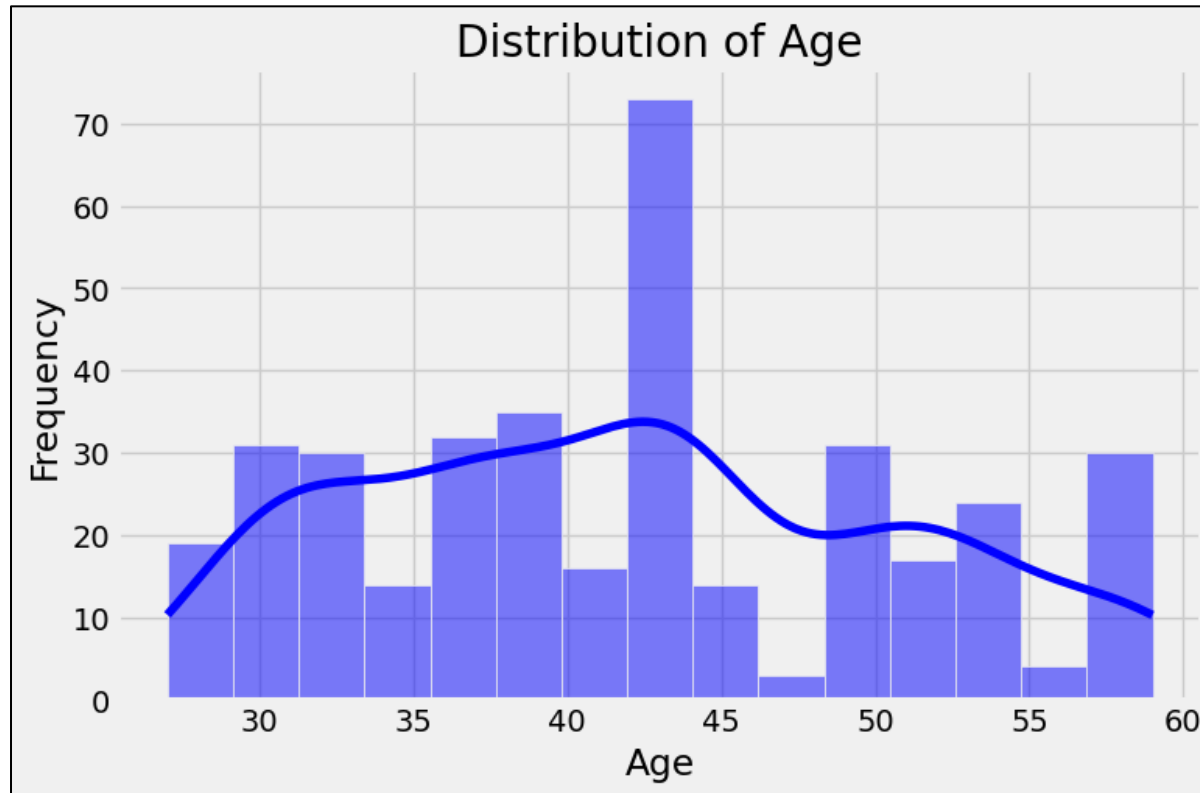


DESCRIPTIVE ANALYSIS

- Distribution of Age, Gender, and BMI Category analyzed.
- Boxplots and histograms plotted for numerical variables.
- Sleep Duration and Quality of Sleep distributions explored.



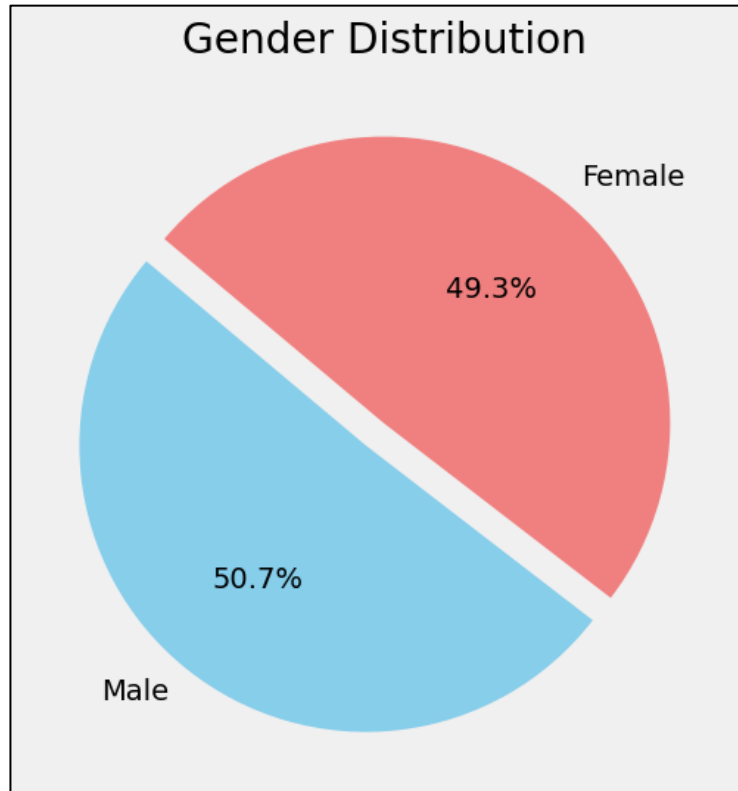
DISTRIBUTION OF AGE



- The age distribution shows a peak around the mid-40s, suggesting that the majority of the individuals in the dataset belong to this age group. The variability across other age ranges ensures a diverse representation, though the data is slightly skewed toward middle-aged participants. This distribution allows for meaningful insights into sleep quality trends across different age categories



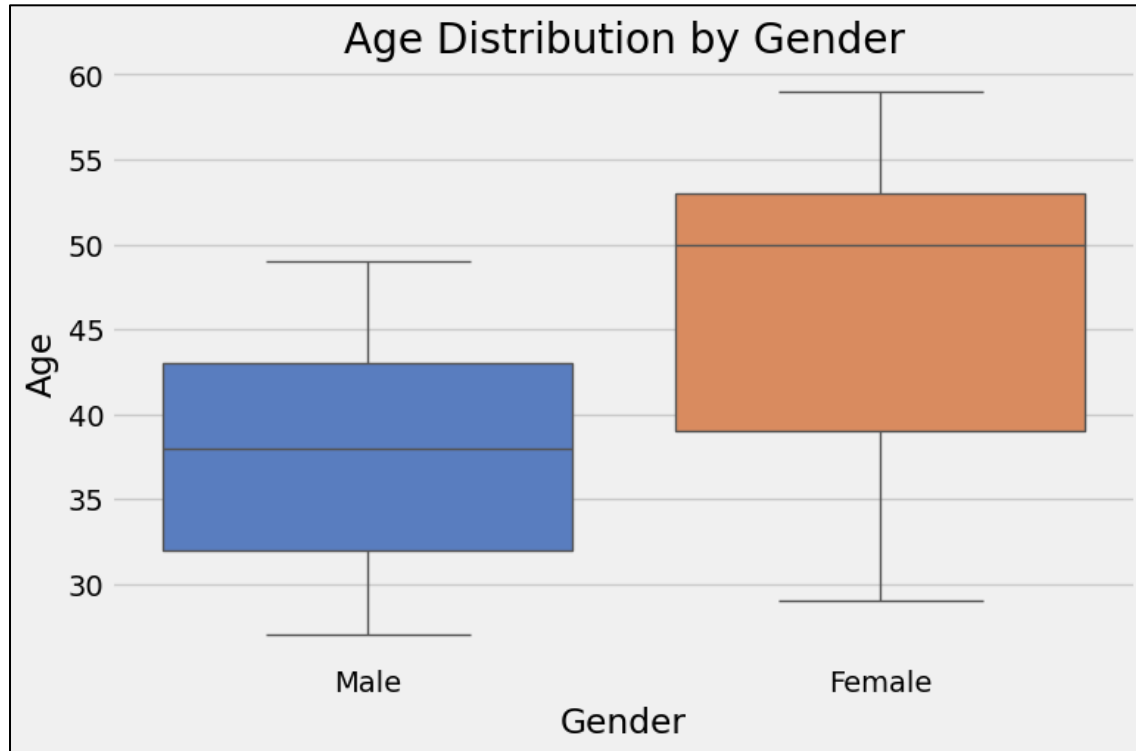
DISTRIBUTION OF GENDER



- **Male:** 49.3%
- **Female:** 50.7%



DISTRIBUTION OF AGE BY GENDER



- Male:

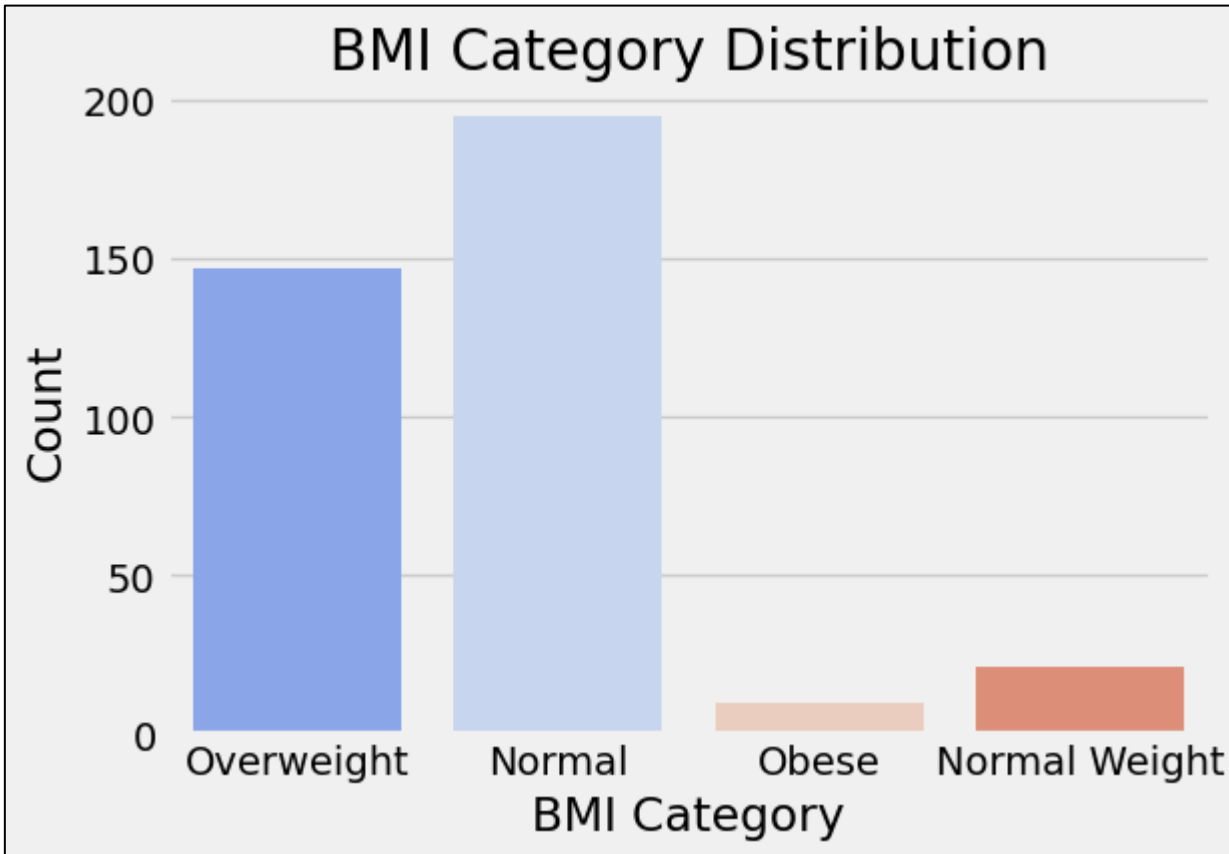
The median age is slightly below 40.

- Female:

The median age is around 45.



DISTRIBUTION OF BMI CATEGORY

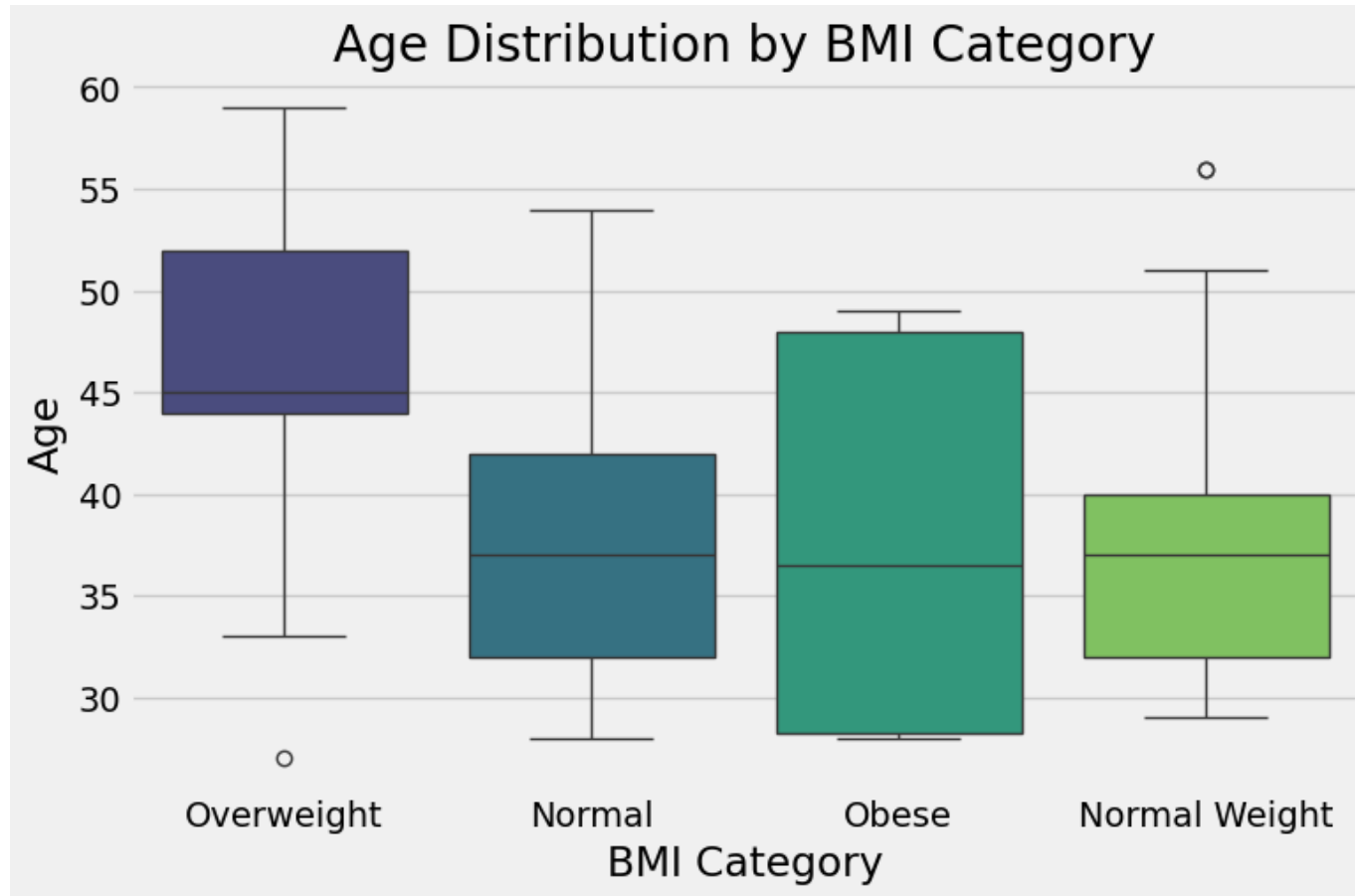


CATEGORY	BMI
Normal	
Normal Weight	18.5-24.9
Overweight	25-29.9
Obese	30 or higher

- Use the Bar Plot to find out how many people are in each BMI category



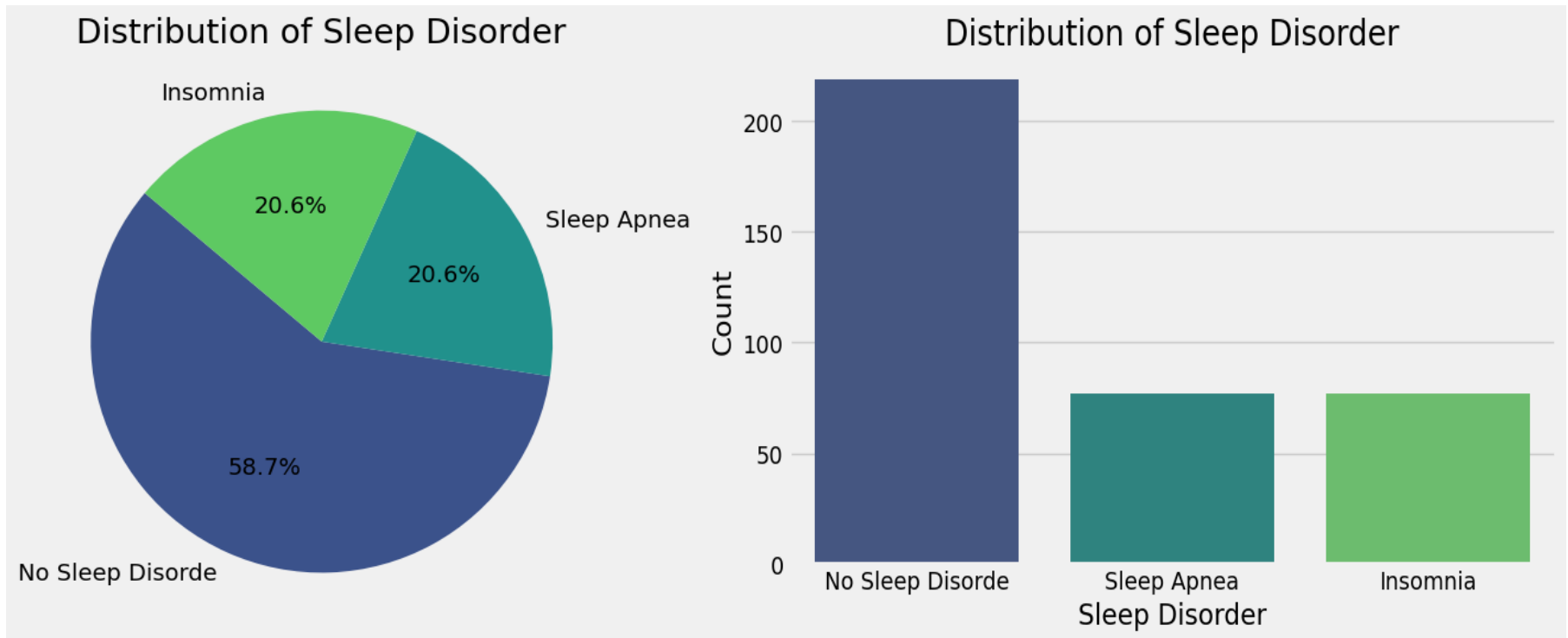
DISTRIBUTION OF BMI CATEGORY



- Use the Bar Plot to find out how many people are in each BMI category



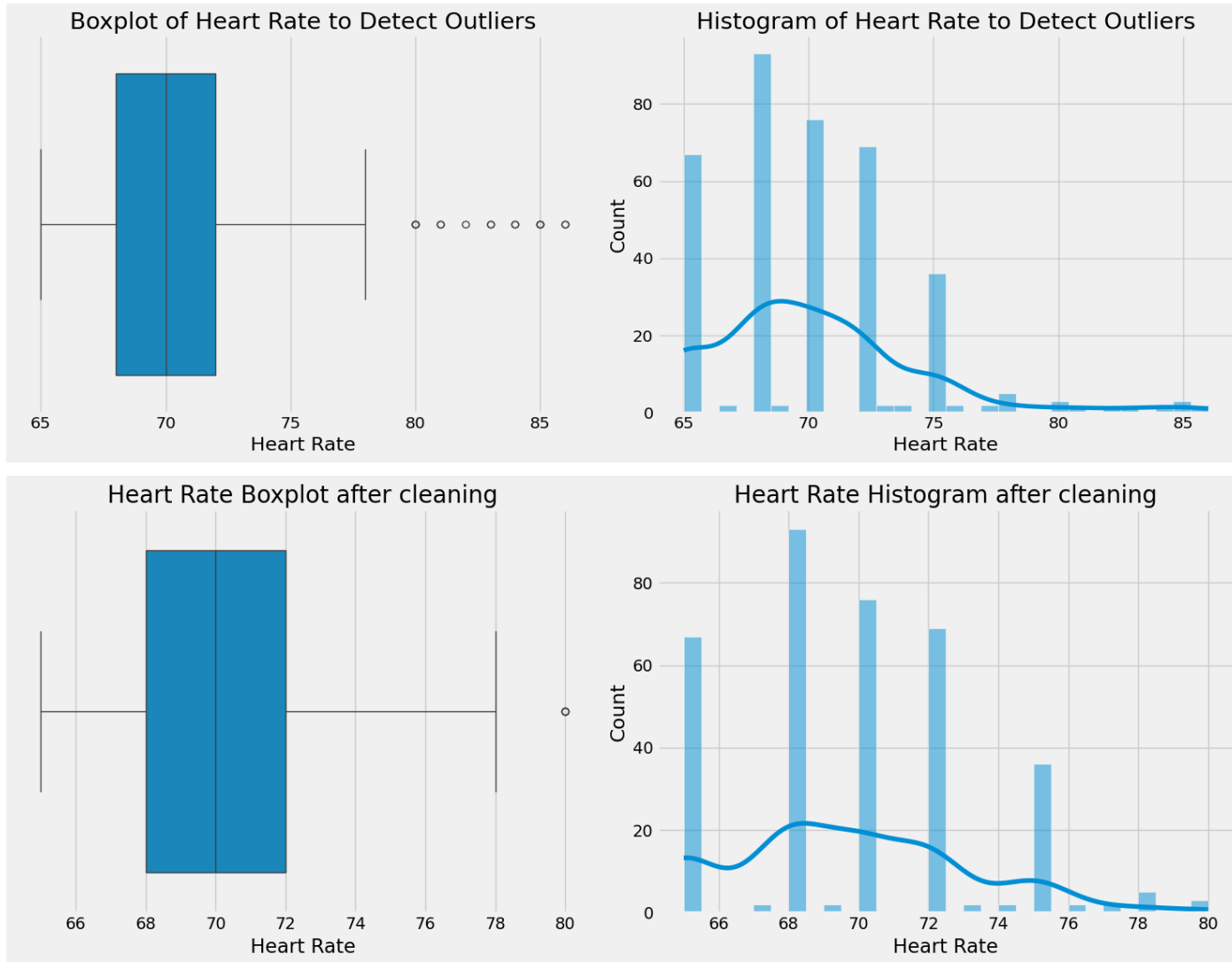
DISTRIBUTION OF SLEEP DISORDER



- To understand the percentage of people who suffer from sleep apnea, whether insomnia or sleep disorder, compared to those who do not suffer



OUTLIERS :



- Quartile was applied to identify and remove outliers.

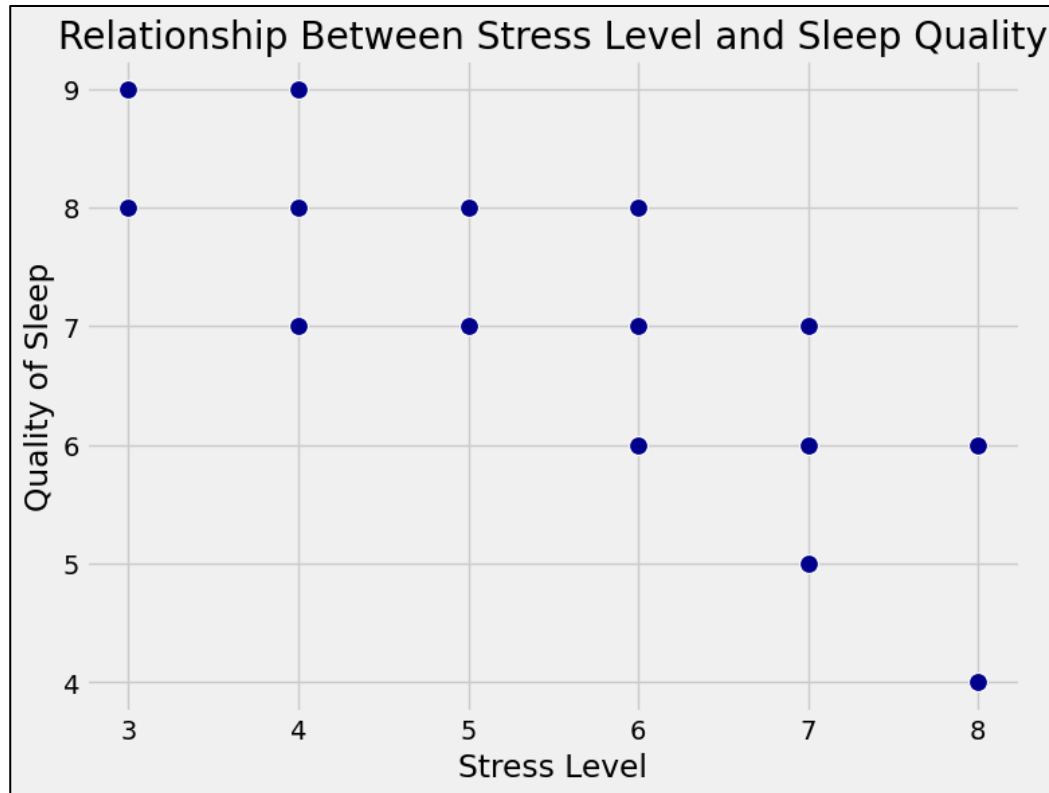


CORRELATION ANALYSIS

- Investigated the relationship between:
- Stress Level and Sleep Quality.
- BMI Category and Sleep Disorders.
- Used heatmaps visualize correlations.



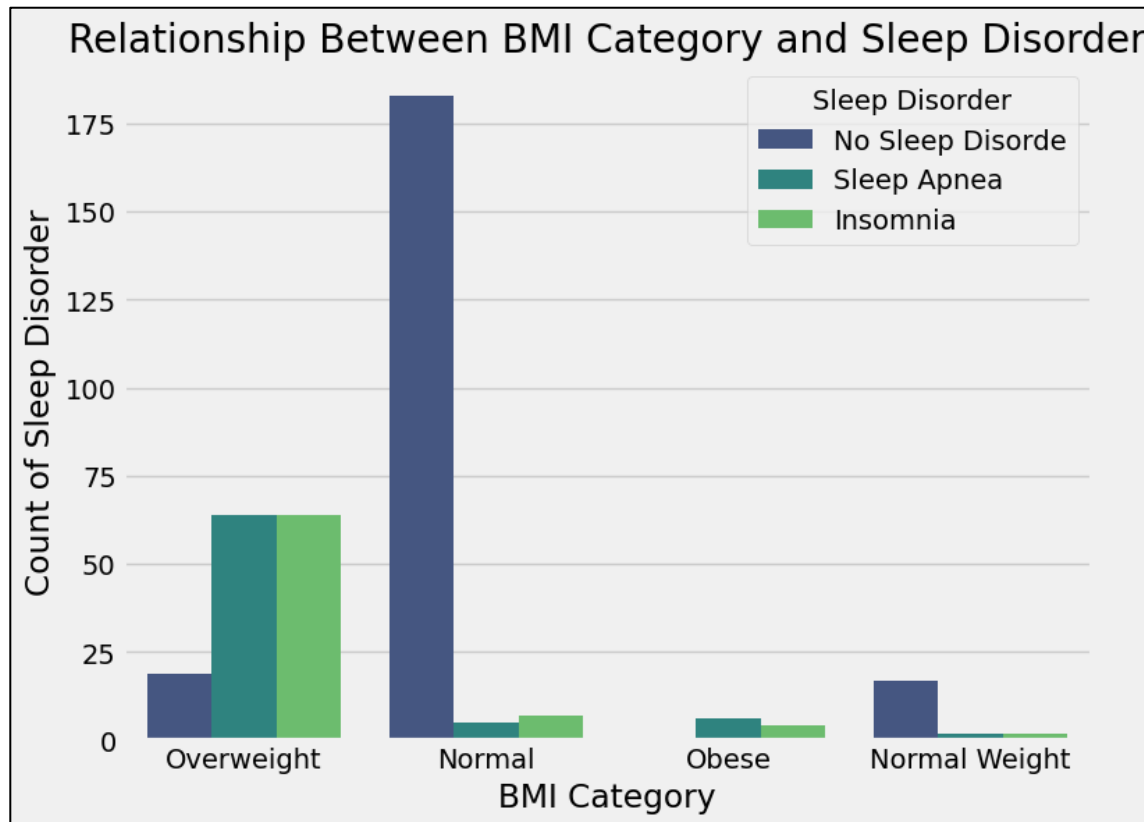
RELATIONSHIP BETWEEN STRESS LEVEL AND SLEEP QUALITY



- Scatter plot to investigate the relationship between Stress Level and Sleep Quality



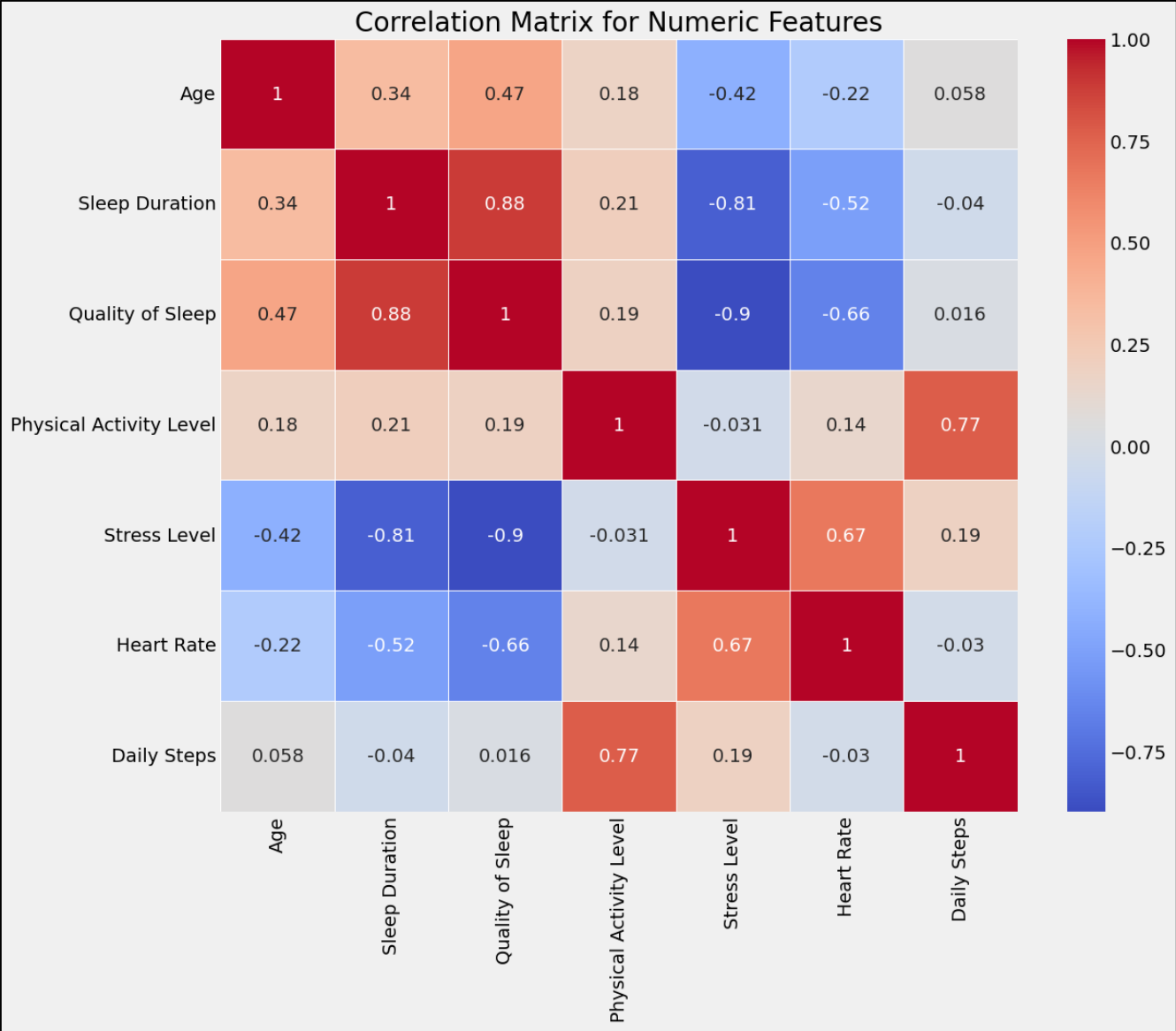
RELATIONSHIP BETWEEN BMI CATEGORY AND SLEEP DISORDER



- Bar plot to investigate the relationship between BMI Category and Sleep Disorder



CORRELATION HEAT MAP



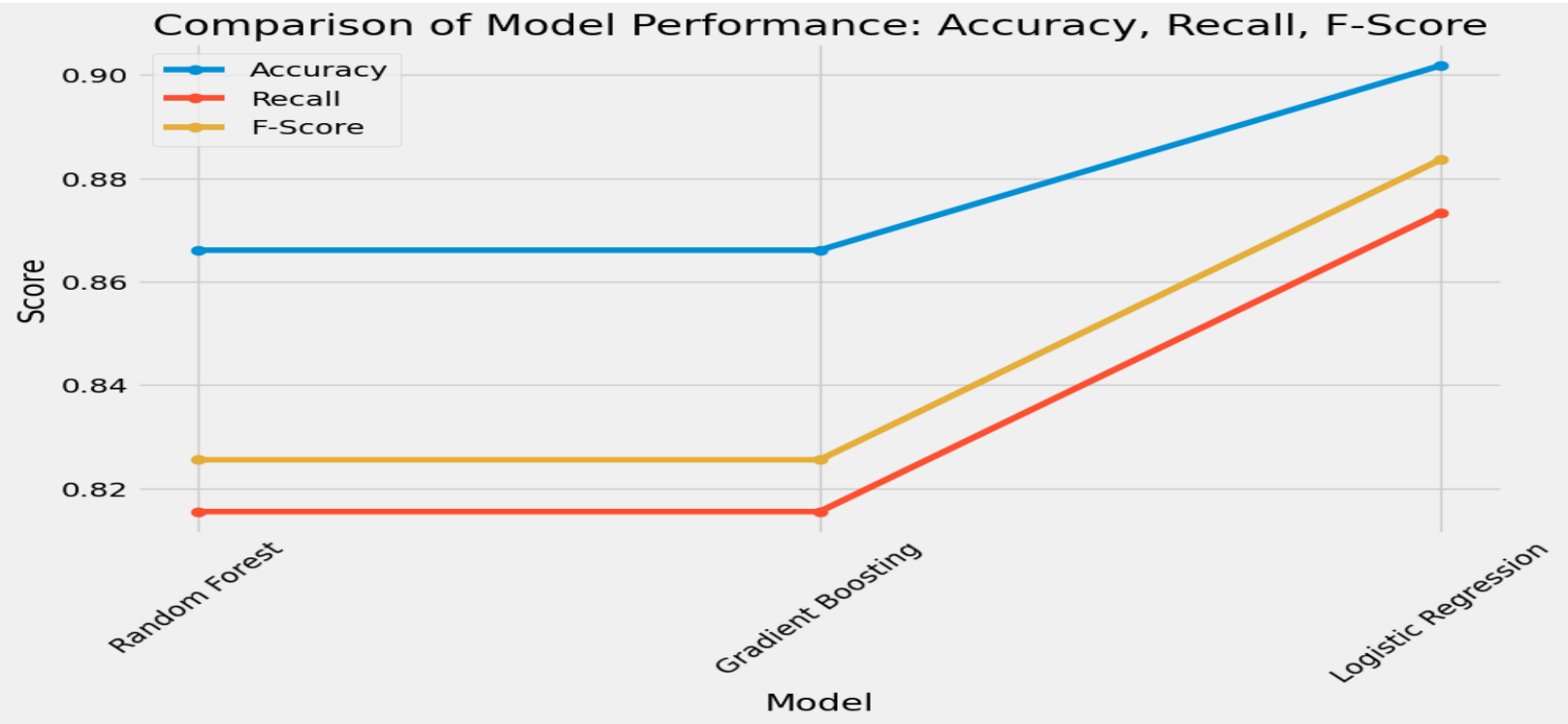
PREDICTIVE MODELS

- Models compared: Logistic Regression, Random Forest and Gradient Boosting.
- Logistic Regression showed the highest performance.
- Evaluations based on Accuracy, Recall, and F1-Score.



MODELS COMPARED

Accuracy	Recall	F-Score	Model
0.866071	0.815523	0.825543	Random Forest
0.866071	0.815523	0.825543	Gradient Boosting
0.901786	0.873330	0.883598	Logistic Regression



Logistic Regression is considered the best model because it has the highest values of precision, recall, and F-Score compared to Random Forest and Gradient Boosting.



CONCLUSIONS

1. **Model Performance:** Logistic Regression performed the best in terms of **accuracy**, **recall**, and **F-score**, indicating its superior ability to predict sleep disorders compared to Random Forest and Gradient Boosting, which showed slightly lower results.
 2. **Key Features:** Factors like **stress level**, **physical activity**, and **BMI category** significantly impact **sleep quality** and **sleep disorders**. Additionally, **age** and **gender** play a role in predicting sleep outcomes.
 3. **Practical Implications:** Understanding these relationships can help design better interventions for improving sleep quality by addressing factors like stress and physical activity levels.
 4. **Future Recommendations:** Future models could benefit from incorporating additional features and exploring advanced algorithms or ensemble methods for enhanced accuracy.
- In summary, Logistic Regression is the most reliable model for predicting sleep disorders in this dataset, but further model refinement could improve predictions.



ANY QUESTION ?

Prerped by: Sahar Alshahrani

