CRACKING THE CODE OF QUALITY SLEEP: A JOURNEY INTO SLEEP HEALTH PATTERNS

- Sleep health is a critical aspect of overall well-being, and disruptions in sleep patterns can significantly impact an individual's physical and mental health.

- Understanding and predicting sleep disorders, such as insomnia and sleep apnea, are crucial for early intervention and personalized health management.

MOTIVATION BEHIND THE PROJECT

- Sleep health is a cornerstone of overall well-being, influencing physical health, cognitive function, and emotional balance.
- Recognizing the importance of addressing sleep health, our motivation is to develop a predictive model that empowers individuals to understand and manage their sleep patterns effectively.
- By leveraging machine learning techniques, we aim to provide a tool that can predict the likelihood of common sleep disorders, namely "Insomnia" and "Sleep Apnea."

PROBLEM STATEMENT

- The problem is to design a machine learning model that accurately classifies individuals into three categories: those without any sleep disorder ("None"), those with insomnia, and those with sleep apnea.

AIM

- The model will utilize relevant features extracted from sleep-related data, encompassing aspects such as sleep duration, efficiency, bedtime patterns, and other metrics.

PROJECT MODULES

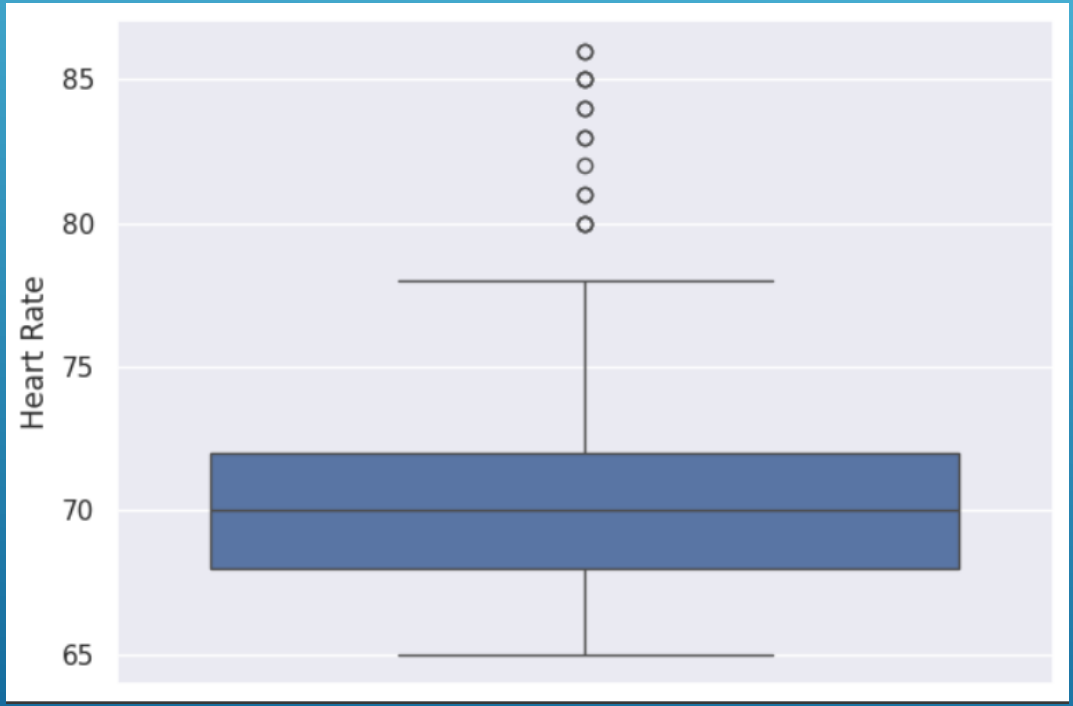
Importing Data: Data understanding
Exploratory Data Analysis: Data Exploration, Univariate/Multivariate analysis

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

- **Person ID:** An identifier for each individual.
- **Gender:** The gender of the person (**Male/Female**).
- **Age:** The age of the person in years.
- **Occupation:** The occupation or profession of the person.
- **Sleep Duration:** The number of hours the person sleeps per day.
- **Quality of Sleep:** A subjective rating of the quality of sleep, ranging from **1 to 10**.
- **Physical Activity:** The number of minutes the person engages in physical activity daily.
- **Stress Level:** A subjective rating of the stress level experienced by the person, ranging from **1 to 10**.
- **BMI Category:** The BMI category of the person (e.g., **Underweight, Normal, Overweight**).
- **Blood Pressure:** The blood pressure measurement of the person, indicated as systolic pressure over diastolic pressure.
- **Heart Rate:** The resting heart rate of the person in beats per minute.
- **Daily Steps:** The number of steps the person takes per day.
- **Sleep Disorder:** The presence or absence of a sleep disorder in the person (**None, Insomnia, Sleep Apnea**).

TREATING OUTLIERS

While exploring, found that Heart rate has some outliers.
So, treating the outliers using IQR (Inter Quartile Range)
After treating, most of the outliers were resolved.



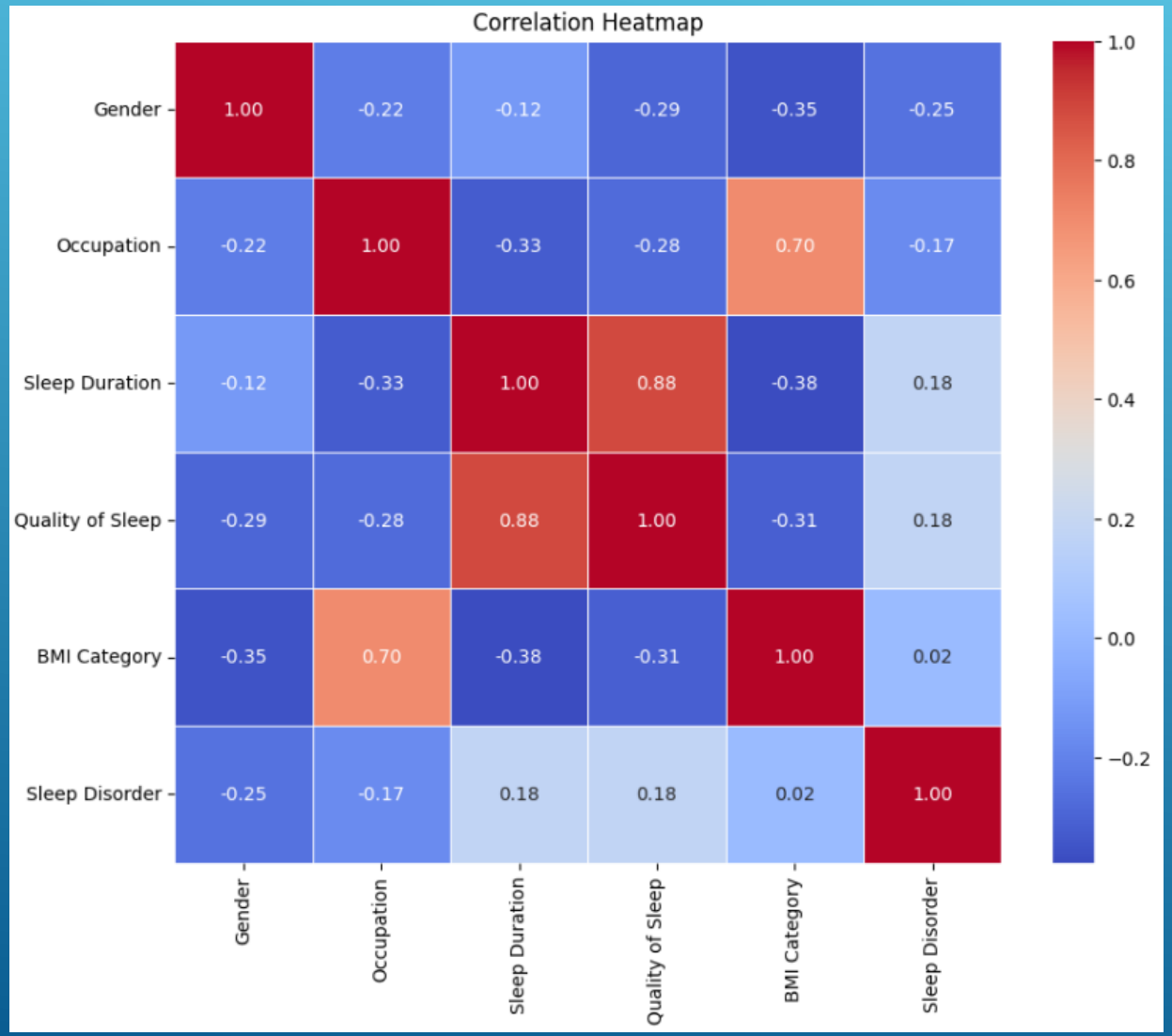

```
def treat_outliers(df, colname):
  q1, q3 = df[colname].quantile([0.25, 0.75])
  iqr = q3-q1
  min = q1-(1.5*iqr)
  max = q3+(1.5*iqr)
  return min, max

treat_outliers(df, 'Heart Rate')
```

```
(62.0, 78.0)
```

LABEL ENCODING/HEATMAP

Applied **Label Encoding** as we have three unique values for our dependent categorical variable 'Sleep Disorder' : 'None', 'Insomnia', 'Sleep Apnea'



Heatmap Evaluation:

- There is a correlation between Occupation and BMI Category.
- Quality of sleep is highly correlated with Sleep duration.

FEATURE SELECTION

Variable	Chi-Square	P-Value
Person ID	748.000000	4.724994e-01
Gender	54.306020	1.612863e-12
Age	447.074575	1.457533e-60
Occupation	421.362755	7.481314e-77
Sleep Duration	434.548392	8.440547e-62
Quality of Sleep	156.959670	1.373652e-28
Physical Activity Level	343.957876	5.051001e-55
Stress Level	240.199368	6.221717e-46
BMI Category	246.968163	1.822932e-50
Heart Rate	245.009940	1.429509e-39
Daily Steps	432.401706	2.311053e-68
BP High	426.076724	2.799307e-69
BP Low	435.550225	2.546427e-72

- A small p-value (typically below 0.05) indicates that the relationship is statistically significant.
- Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Heart Rate, Daily Steps, BP High, BP Low:
- All these variables have extremely low p-values (close to zero), indicating a highly significant association with the target variable.

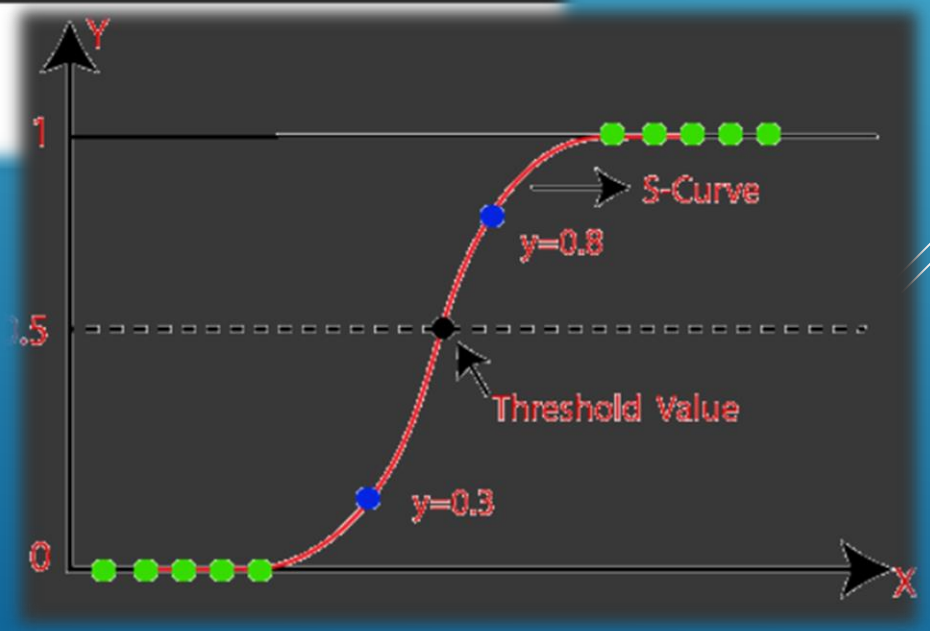
LOGISTIC REGRESSION

•Logistic regression classifies samples by using the idea of predictive modelling as a regression technique; as a result, it is categorized as a classification algorithm.

```
[ ]  ### Creating a Logistic regression object
     logistic_reg=LogisticRegression(random_state = 0)

[ ]  ### Passing independant and dependant training data to the model
     logistic_reg.fit(X_train,y_train)
```

```
LogisticRegression
LogisticRegression(random_state=0)
```



MODEL BUILDING

- Test size = 0.2 implies we have taken 20% of the total data as test data and random state = 42 implies different train and test sets will be generated across different executions.
- if random state = True which means fixed sets will be generated.

```
# Assuming X is your feature matrix and y is the encoded label
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and fit the model
model = LogisticRegression(multi_class='multinomial', solver='lbfgs')
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)
```

Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Heart Rate	Daily Steps	BP High	BP Low
-1.486605	0.989361	-1.406737	-0.908497	0.840575	-0.261708	0.760896	0.347021	-0.907284	0.009757	0.732263	-1.105421	-0.755640
0.319551	0.989361	0.209606	1.057431	-0.921335	-1.098280	-0.681208	0.911306	1.191629	0.571163	-0.505551	0.186944	0.056923

PERFORMANCE METRICS

Accuracy: 0.91

	precision	recall	f1-score	support
0	0.78	0.88	0.82	16
1	0.95	0.95	0.95	43
2	0.93	0.81	0.87	16
accuracy			0.91	75
macro avg	0.89	0.88	0.88	75
weighted avg	0.91	0.91	0.91	75

ADVANTAGES

- This project aligns with the growing interest in digital health solutions, leveraging technology to address prevalent health challenges.
- Through the development and deployment of an accurate sleep health prediction model, we aim to empower individuals to take proactive steps toward achieving better sleep and, consequently, better overall health.

ADVANTAGES

- This project aligns with the growing interest in digital health solutions, leveraging technology to address prevalent health challenges.
- Through the development and deployment of an accurate sleep health prediction model, we aim to empower individuals to take proactive steps toward achieving better sleep and, consequently, better overall health.

- Improve Sleep Health Awareness
- Enable Early Intervention
- Personalized Health Management
- Contribute to Public Health
- Promote Proactive Self-Care

CHALLENGES FACED

Data Quality and Availability:
•Obtaining a comprehensive and high-quality dataset with diverse sleep-related features can be challenging. Incomplete or biased data may impact the model's performance.

Labeling and Class Imbalance:
•Accurate labeling of sleep health patterns, especially for disorders like insomnia and sleep apnea, can be subjective. Class imbalance also affected model training.

Feature Selection and Engineering:
•Identifying and selecting relevant features from sleep-related data is complex. Choosing the right set of features that truly contribute to predicting sleep health patterns is crucial.

Get enough sleep every night. An exhausted mind is rarely productive.

Occupation	Max of BP High	Max of BP Low
Accountant	130	85
Doctor	142	92
Engineer	132	87
Lawyer	135	88
Manager	125	82
Nurse	140	95
Sales Representative	140	90
Salesperson	130	85

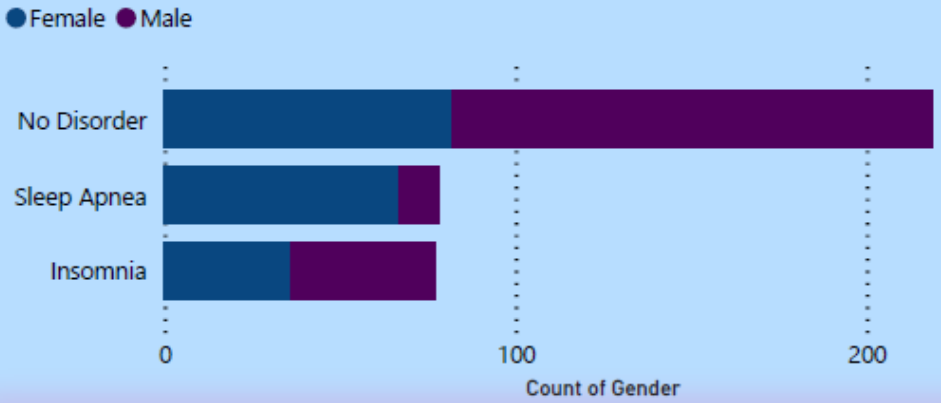
7.13
Average of Sleep Duration

70.17
Average of Heart Rate

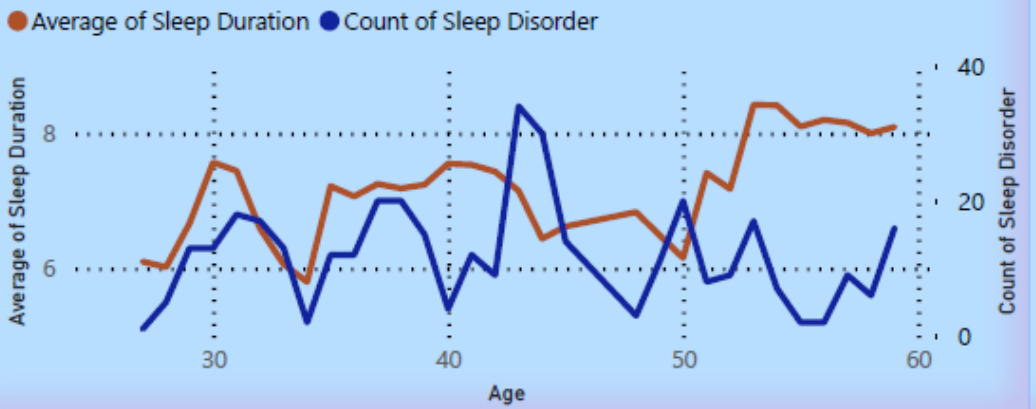
5.39
Average of Stress Level

42.18
Average of Age

Gender by Sleep Disorder

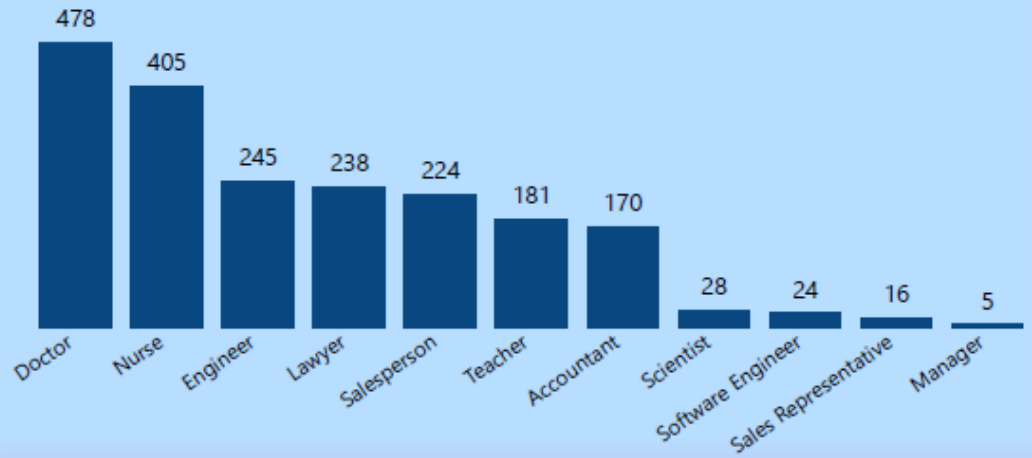


Average of Sleep Duration and Count of Sleep Disorder by Age

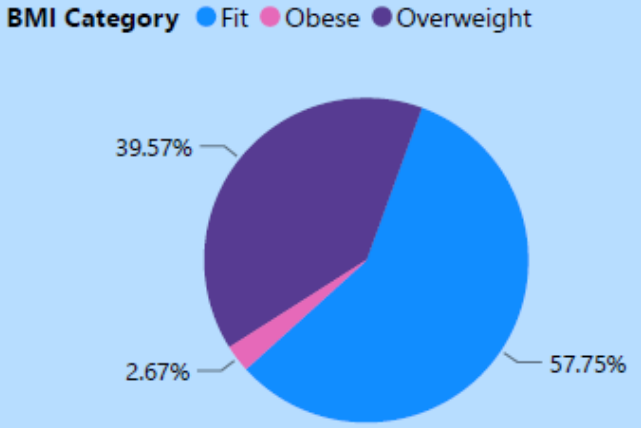


6.82K
Average of Daily Steps

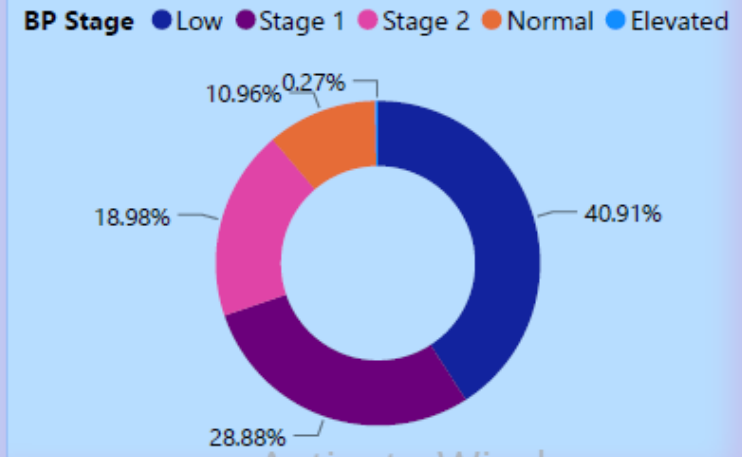
Total Stress Level by Occupation



Distribution by BMI Category



Distribution by BP STages



INSIGHT

- **Population Distribution:**
- The population in this dataset has a fairly balanced gender distribution, with 50.5% males and 49.5% females.
- The most prevalent occupation is nursing (19.5%), followed by doctors (19.0%), with the majority falling into the Normal BMI category (57.8%).
- **Relationship between Sleep Disorders and Other Factors:**
- Individuals without sleep disorders (None) have the highest distribution, while Insomnia has the lowest.
- Nurses are the occupational group with the highest rate of sleep disorders, while Managers, Engineers, and Salespersons have the lowest rates.

- **Influence of BMI and Occupation on Sleep Duration and Quality:**
- Nurses have the highest sleep duration and quality, while Managers have the lowest.
- Engineers and Nurses with a Normal BMI category have the highest sleep duration and quality, while Software Engineers and Scientists with an Obese BMI category have the lowest.
- **Characteristics of Sleep Disorders:**
- The average age of individuals without sleep disorders falls in the range of 35 to 40 years. Meanwhile, those with Sleep Apnea tend to have an average age of 50 years, and individuals with Insomnia typically fall within the average age range of 40 to 45 years.
- The average sleep duration and quality of sleep scores are lower for individuals with sleep disorders, with Sleep Apnea having a lower average than Insomnia.