zerotopandas-course-project

November 30, 2023

1 Sleep Health and Lifestyle

This project explores the relationship between sleep, health and lifestyle in an enigmatic population. I will be analyzing the dataset collected from https://www.kaggle.com, shedding light on the fascinating interplay between individuals' lifestyle choices and their sleep patterns. As I embark on this analytical journey, I aim to unravel the mysteries surrounding sleep behaviors, uncovering correlations between daily routines, habits, and the quality of sleep among this diverse community. Join me as I navigate through this intriguing dataset, striving to illuminate the intricate dynamics shaping the sleep habits and lifestyles of this population.

This project is part of my Jovian course, Data Analysis with Python: Zero to Pandas

1.1 Downloading the Dataset

I download my required dataset using Python's opendatasets library. To do so, I first install the opendatasets library

[5]:
```
!pip install jovian opendatasets --upgrade --quiet
```

Let's begin by downloading the data, and listing the files within the dataset.

[6]:
```
dataset_url = 'https://www.kaggle.com/datasets/uom190346a/
↪sleep-health-and-lifestyle-dataset'
```

[7]:
```
import opendatasets as od
od.download(dataset_url)
```

```
Please provide your Kaggle credentials to download this dataset. Learn more:
http://bit.ly/kaggle-creds
Your Kaggle username: insiyahujjainwala
Your Kaggle Key: ·····
Downloading sleep-health-and-lifestyle-dataset.zip to ./sleep-health-and-
lifestyle-dataset

100%|      | 2.54k/2.54k [00:00<00:00, 2.21MB/s]
```

The dataset has been downloaded and extracted.

[8]:
```
data_dir = './sleep-health-and-lifestyle-dataset'
```

[9]:
```
import os
os.listdir(data_dir)
```

[9]:
```
['Sleep_health_and_lifestyle_dataset.csv']
```

Let us save and upload our work to Jovian before continuing.

[10]:
```
project_name = "sleep-health-and-lifestyle-project"
```

[11]:
```
!pip install jovian --upgrade -q
```

[12]:
```
import jovian
```

[13]:
```
jovian.commit(project=project_name)
```

```
<IPython.core.display.Javascript object>
```

```
[jovian] Updating notebook "insiyah-yu/sleep-health-and-lifestyle-project" on
https://jovian.com
[jovian] Committed successfully! https://jovian.com/insiyah-yu/sleep-health-and-
lifestyle-project
```

[13]:
```
'https://jovian.com/insiyah-yu/sleep-health-and-lifestyle-project'
```

1.2 Installing Libraries

We first install/ upgrade all necessary python libraries.

[14]:
```
!pip install pandas --upgrade
!pip install matplotlib
!pip install seaborn
```

```
Requirement already satisfied: pandas in /opt/conda/lib/python3.9/site-packages
(2.1.3)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.9/site-
packages (from pandas) (2021.3)
Requirement already satisfied: tzdata>=2022.1 in /opt/conda/lib/python3.9/site-
packages (from pandas) (2023.3)
Requirement already satisfied: numpy<2,>=1.22.4 in
/opt/conda/lib/python3.9/site-packages (from pandas) (1.22.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/opt/conda/lib/python3.9/site-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.9/site-
packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Requirement already satisfied: matplotlib in /opt/conda/lib/python3.9/site-
packages (3.4.3)
Requirement already satisfied: kiwisolver>=1.0.1 in
```

```
/opt/conda/lib/python3.9/site-packages (from matplotlib) (1.3.2)
Requirement already satisfied: pyparsing>=2.2.1 in
/opt/conda/lib/python3.9/site-packages (from matplotlib) (2.4.7)
Requirement already satisfied: pillow>=6.2.0 in /opt/conda/lib/python3.9/site-
packages (from matplotlib) (9.5.0)
Requirement already satisfied: numpy>=1.16 in /opt/conda/lib/python3.9/site-
packages (from matplotlib) (1.22.4)
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.9/site-
packages (from matplotlib) (0.11.0)
Requirement already satisfied: python-dateutil>=2.7 in
/opt/conda/lib/python3.9/site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.9/site-
packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Requirement already satisfied: seaborn in /opt/conda/lib/python3.9/site-packages
(0.11.2)
Requirement already satisfied: numpy>=1.15 in /opt/conda/lib/python3.9/site-
packages (from seaborn) (1.22.4)
Requirement already satisfied: scipy>=1.0 in /opt/conda/lib/python3.9/site-
packages (from seaborn) (1.7.2)
Requirement already satisfied: matplotlib>=2.2 in /opt/conda/lib/python3.9/site-
packages (from seaborn) (3.4.3)
Requirement already satisfied: pandas>=0.23 in /opt/conda/lib/python3.9/site-
packages (from seaborn) (2.1.3)
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.9/site-
packages (from matplotlib>=2.2->seaborn) (0.11.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/opt/conda/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn) (1.3.2)
Requirement already satisfied: python-dateutil>=2.7 in
/opt/conda/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn) (2.8.2)
Requirement already satisfied: pillow>=6.2.0 in /opt/conda/lib/python3.9/site-
packages (from matplotlib>=2.2->seaborn) (9.5.0)
Requirement already satisfied: pyparsing>=2.2.1 in
/opt/conda/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn) (2.4.7)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.9/site-
packages (from pandas>=0.23->seaborn) (2021.3)
Requirement already satisfied: tzdata>=2022.1 in /opt/conda/lib/python3.9/site-
packages (from pandas>=0.23->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.9/site-
packages (from python-dateutil>=2.7->matplotlib>=2.2->seaborn) (1.16.0)
```

[15]:
```
import pandas as pd
import numpy as np
```

1.3 Exploring the Dataset

We read the csv file and load it into a dataframe. Then, we explore the dataset, find out information such as the number of rows, columns, and range of values.

[16]:
```
sleep_raw_df = pd.read_csv(data_dir + "/Sleep_health_and_lifestyle_dataset.csv")
```

[17]:
```
sleep_raw_df
```

[17]:

	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
..	…	…	…	…	…	…	…	…	…	…	…	…	…
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea

[374 rows x 13 columns]

1.4 Understanding our Data

We use the info() and describe() methods to find out the structure of data in our dataset.

Info() tells us the datatype of each of our columns, along with the count of non-null values in each column.

Describe() gives us various statistics of our dataset (only considers those columns which are of type int/float).

[18]:
```
sleep_raw_df.shape
```

[18]:
```
(374, 13)
```

[19]:
```
sleep_raw_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 13 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Person ID                374 non-null    int64  
 1   Gender                   374 non-null    object 
 2   Age                      374 non-null    int64  
 3   Occupation               374 non-null    object 
 4   Sleep Duration           374 non-null    float64
 5   Quality of Sleep         374 non-null    int64  
 6   Physical Activity Level  374 non-null    int64  
 7   Stress Level             374 non-null    int64  
 8   BMI Category             374 non-null    object 
 9   Blood Pressure           374 non-null    object 
 10  Heart Rate               374 non-null    int64  
 11  Daily Steps              374 non-null    int64  
 12  Sleep Disorder           155 non-null    object 
dtypes: float64(1), int64(7), object(5)
memory usage: 38.1+ KB
```

[20]:
```
sleep_raw_df.describe()
```

[20]:

	Person ID	Age	Sleep Duration	Quality of Sleep	\
count	374.000000	374.000000	374.000000	374.000000	
mean	187.500000	42.184492	7.132086	7.312834	
std	108.108742	8.673133	0.795657	1.196956	
min	1.000000	27.000000	5.800000	4.000000	
25%	94.250000	35.250000	6.400000	6.000000	
50%	187.500000	43.000000	7.200000	7.000000	
75%	280.750000	50.000000	7.800000	8.000000	
max	374.000000	59.000000	8.500000	9.000000	

	Physical Activity Level	Stress Level	Heart Rate	Daily Steps
count	374.000000	374.000000	374.000000	374.000000
mean	59.171123	5.385027	70.165775	6816.844920
std	20.830804	1.774526	4.135676	1617.915679
min	30.000000	3.000000	65.000000	3000.000000
25%	45.000000	4.000000	68.000000	5600.000000
50%	60.000000	5.000000	70.000000	7000.000000
75%	75.000000	7.000000	72.000000	8000.000000
max	90.000000	8.000000	86.000000	10000.000000

We see from the results above that we do not have any NULL values in our columns (except for sleep disorder which only implies that the person is not suffering from any sleep disorders, and hence we allow those NULL values).

1.5 Data Cleaning

Checking for discrepancies in our data and getting rid of them.

Dropping columns that are potentially of no use for the purpose of our analysis.

[21]:
```
sleep_clean_df = sleep_raw_df
```

[22]:
```
print(sleep_clean_df['BMI Category'].unique())
```

```
['Overweight' 'Normal' 'Obese' 'Normal Weight']
```

We see from the result above that the entry to indicate normal BMI is some times stated as 'Normal' and other times stated as 'Normal Weight'. Let's change all entries to 'Normal'.

[23]:
```
sleep_clean_df['BMI Category'].replace("Normal Weight", "Normal", inplace =␣
↪True)

# Checking to make sure the entries have been replaced
sleep_clean_df['BMI Category'].unique()
```

[23]:
```
array(['Overweight', 'Normal', 'Obese'], dtype=object)
```

With this, we are done with the cleaning of our dataset.

[24]:
```
import jovian
```

[25]:
```
jovian.commit()
```

```
<IPython.core.display.Javascript object>
```

```
[jovian] Updating notebook "insiyah-yu/sleep-health-and-lifestyle-project" on
https://jovian.com
[jovian] Committed successfully! https://jovian.com/insiyah-yu/sleep-health-and-
lifestyle-project
```

[25]: 'https://jovian.com/insiyah-yu/sleep-health-and-lifestyle-project'

1.6 Exploratory Analysis and Visualization

The dataset includes a variety of variables that can be categorized into demographic, health-related, and lifestyle factors. These include age, gender, occupation, sleep duration, physical activity level, heart rate, daily steps, BMI category, blood pressure, and the presence of sleep disorders.

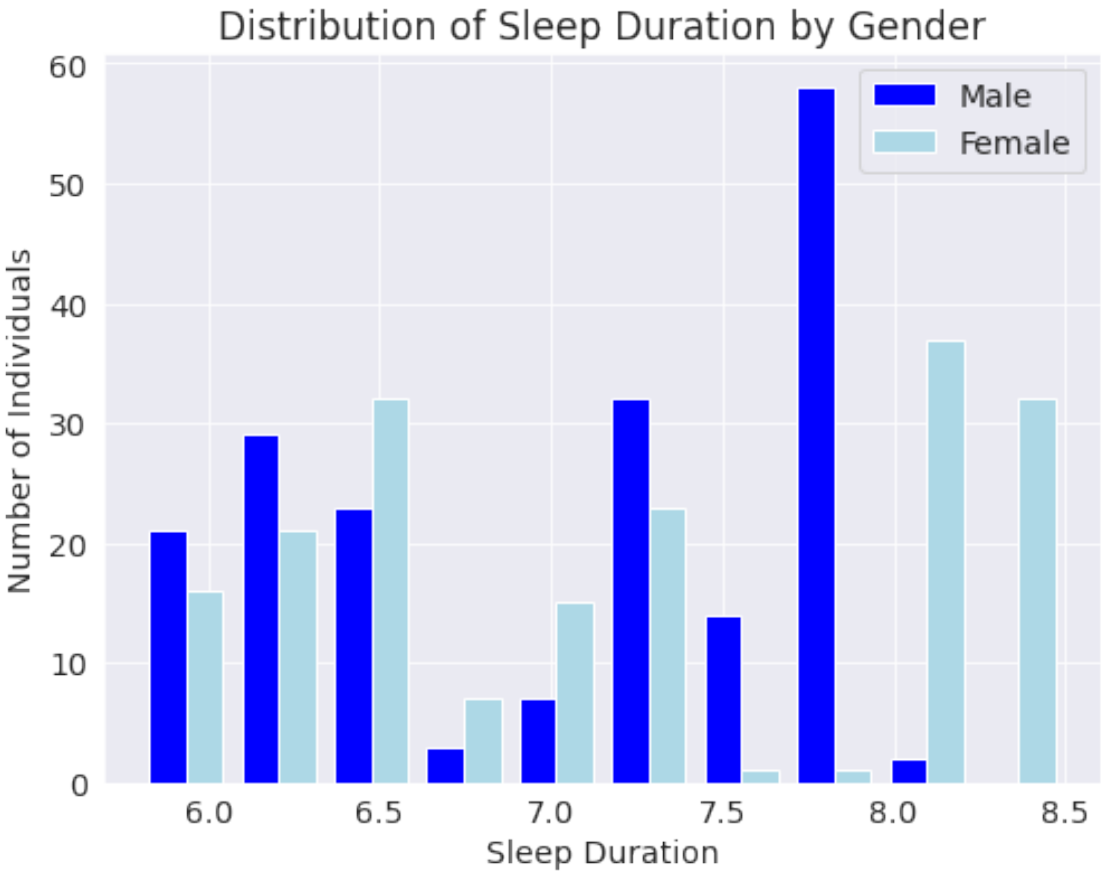
In this section, we aim to undertake a comprehensive exploration to discern how various factors potentially influence an individual's sleeping habits. This analysis will delve into the intricate interplay between diverse elements, ranging from physiological parameters to lifestyle choices, and their cumulative impact on sleep patterns. By systematically examining these aspects, we seek to shed light on the multifaceted nature of sleep behavior and its underlying determinants.

Let's begin by importing`matplotlib.pyplot` and `seaborn`.

[26]:
```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

1.7 Graph 1: Sleep Duration and Gender

[27]:
```
plt.figure(figsize=(8, 6))


plt.hist([sleep_clean_df[sleep_clean_df['Gender'] == 'Male']['Sleep Duration'],
          sleep_clean_df[sleep_clean_df['Gender'] == 'Female']['Sleep␣
 ↪Duration']],
         bins=10, alpha=1, label=['Male', 'Female'], color=['blue',␣
 ↪'lightblue'])
plt.xlabel('Sleep Duration')
plt.ylabel('Number of Individuals')
plt.title('Distribution of Sleep Duration by Gender')
plt.legend()
plt.show()
```

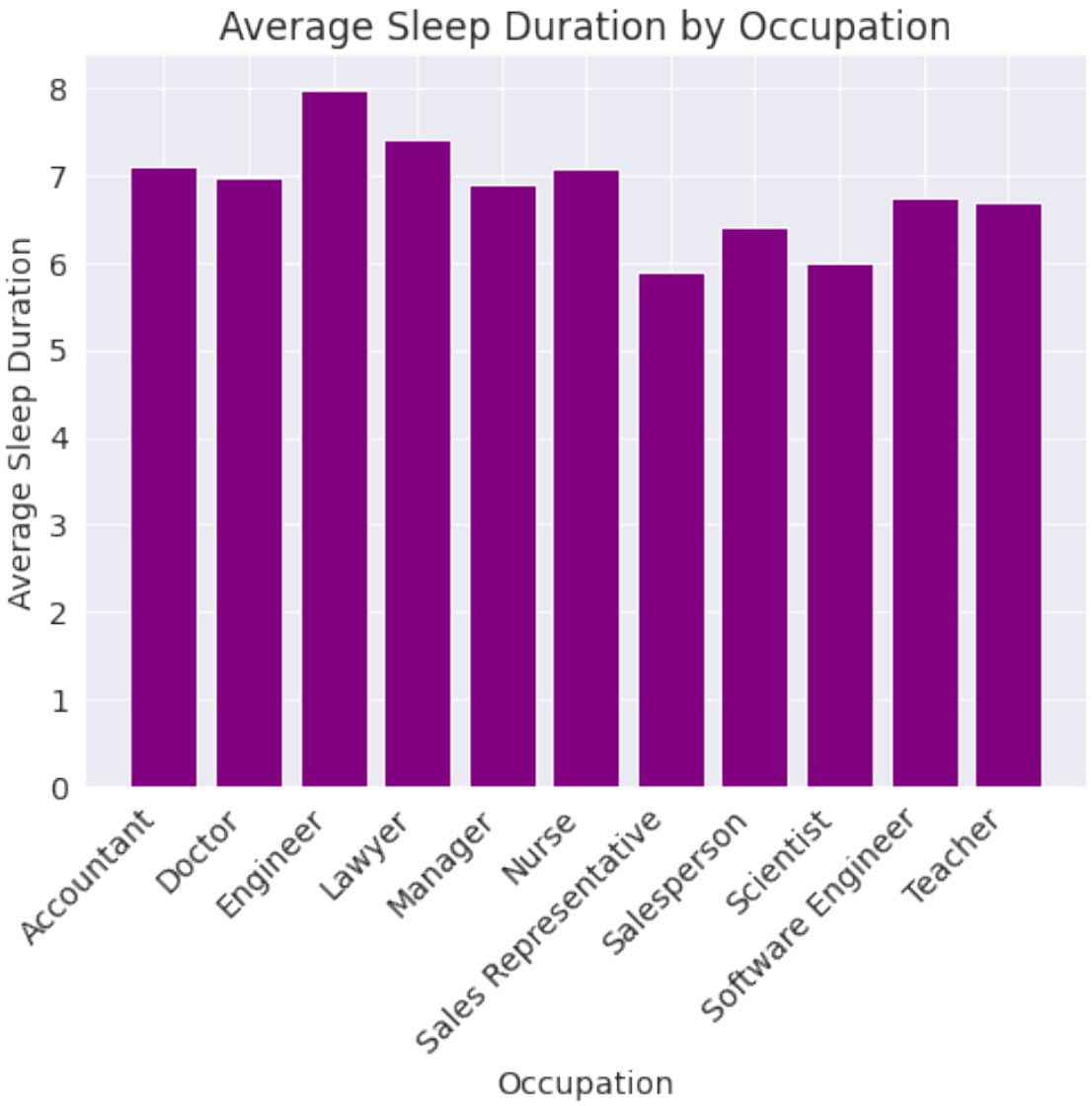


The bar graph delineates the distribution of sleep duration by gender, indicating that the most common sleep duration for both males and females in this dataset is around 7.5 hours. There is a discernible difference between genders: males show a secondary peak at 6.5 hours, whereas females tend to have a more evenly distributed frequency across different sleep durations, with a slight increase at 8.0 hours. Both genders have a sleep duration range from 6.0 to 8.5 hours, with no entries outside this interval. The majority of participants, regardless of gender, report sleeping between 7.0 to 8.0 hours, aligning with standard sleep recommendations for adults. This visualization highlights gender-specific sleep patterns, suggesting a potential need for tailored sleep duration recommendations.

1.8 Graph 2: Sleep Duration and Occupation

[28]:
```
avg_sleep = sleep_clean_df.groupby('Occupation')['Sleep Duration'].mean().
↪reset_index()
plt.figure(figsize=(8, 6))
plt.bar(avg_sleep['Occupation'], avg_sleep['Sleep Duration'], color='purple')
plt.xlabel('Occupation')
plt.ylabel('Average Sleep Duration')
plt.title('Average Sleep Duration by Occupation')
```

```
plt.xticks(rotation=45, ha='right')  # Rotating x-labels for better readability
plt.show()
```

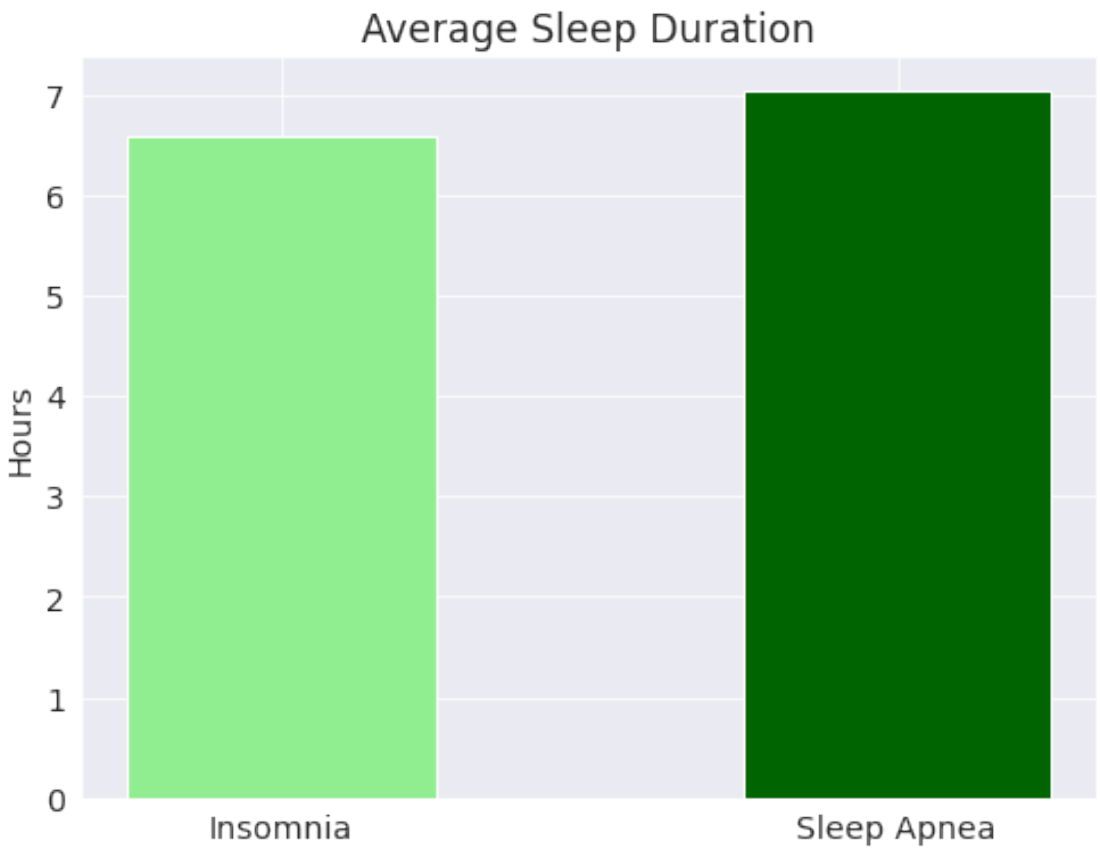


The bar chart illustrates the average sleep duration among various occupations, revealing some intriguing patterns that may challenge common stereotypes. Contrary to the prevailing belief about the grueling nature of engineering, individuals in this profession report the highest average sleep duration compared to other occupations. This observation defies the typical expectation of engineers experiencing shorter sleep periods due to demanding workloads. Other professions, such as lawyers and doctors, also show relatively high sleep averages, which may reflect successful management of work-life balance or prioritization of rest within these roles. On the other end of the spectrum, accountants and sales representatives are depicted with marginally lower sleep averages, which could be indicative of the pressures and unpredictable schedules often associated with these careers. The data presented here underscores the importance of not relying on assumptions about

occupational demands and their impact on sleep, as actual sleep durations may reflect a complex interplay of work culture, individual habits, and lifestyle choices.

1.9 Graph 3: Sleeping Disorders

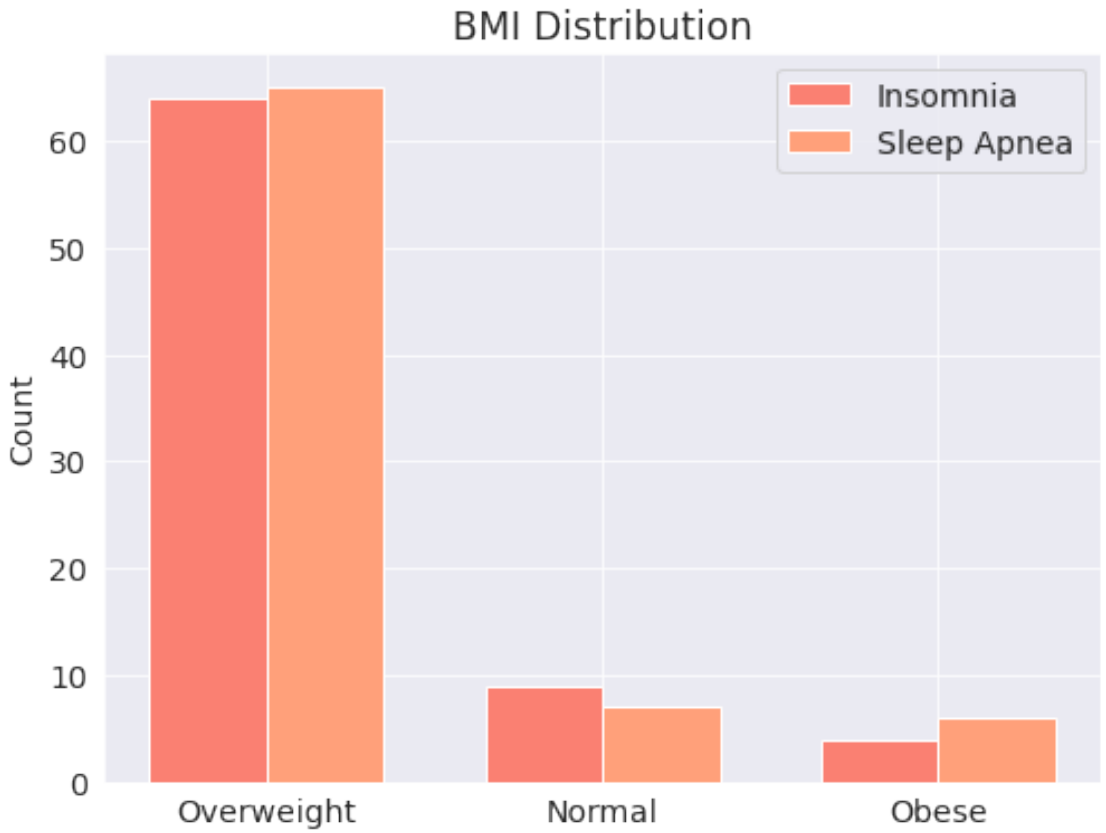
[29]:
```
# Filtering data for people with insomnia and sleep apnea
insomnia_data = sleep_clean_df[sleep_clean_df['Sleep Disorder'] == 'Insomnia']
sleep_apnea_data = sleep_clean_df[sleep_clean_df['Sleep Disorder'] == 'Sleep␣
↪Apnea']

# Plot 1: Comparison of average sleep duration
avg_sleep_insomnia = insomnia_data['Sleep Duration'].mean()
avg_sleep_apnea = sleep_apnea_data['Sleep Duration'].mean()

labels = ['Insomnia', 'Sleep Apnea']
average_sleep = [avg_sleep_insomnia, avg_sleep_apnea]

plt.figure(figsize=(8, 6))
plt.bar(labels, average_sleep, width = 0.5, color=['lightgreen', 'darkgreen'])
plt.title('Average Sleep Duration')
plt.ylabel('Hours')
plt.show()
```



The bar chart displays average sleep durations for individuals with insomnia compared to those with sleep apnea. Surprisingly, individuals with insomnia average over 6 hours of sleep, which challenges the common perception of this condition involving severely reduced sleep. Those with sleep apnea, a disorder characterized by disrupted breathing during sleep, show a higher average of nearly 7 hours. This suggests that while sleep apnea patients may achieve longer sleep, the restorative quality of their sleep could be compromised, consistent with the condition's symptomatic nighttime disturbances. Insomnia's impact, often associated with prolonged sleep onset and frequent waking, appears less severe in terms of sleep quantity in this sample. Both conditions contribute to significant health challenges, such as cardiovascular risks and daytime impairment, highlighting the complex nature of sleep disorders where both duration and quality are crucial for understanding their full impact on well-being.

1.10 Graph 4: BMI and Sleeping Disorders

[30]:
```
# Calculate BMI distribution
bmi_insomnia = insomnia_data['BMI Category'].value_counts()
bmi_sleep_apnea = sleep_apnea_data['BMI Category'].value_counts()

labels = bmi_insomnia.index.union(bmi_sleep_apnea.index).tolist()
x = np.arange(len(labels))  # Label locations
width = 0.35  # Width of the bars

# Create bar plot
plt.figure(figsize=(8, 6))
plt.bar(x - width/2, bmi_insomnia.reindex(labels).fillna(0), width, color =␣
↪'salmon', label='Insomnia')
plt.bar(x + width/2, bmi_sleep_apnea.reindex(labels).fillna(0), width, color =␣
↪'lightsalmon',label='Sleep Apnea')

plt.ylabel('Count')
plt.title('BMI Distribution')
plt.xticks(x, labels)
plt.legend()
plt.show()
```

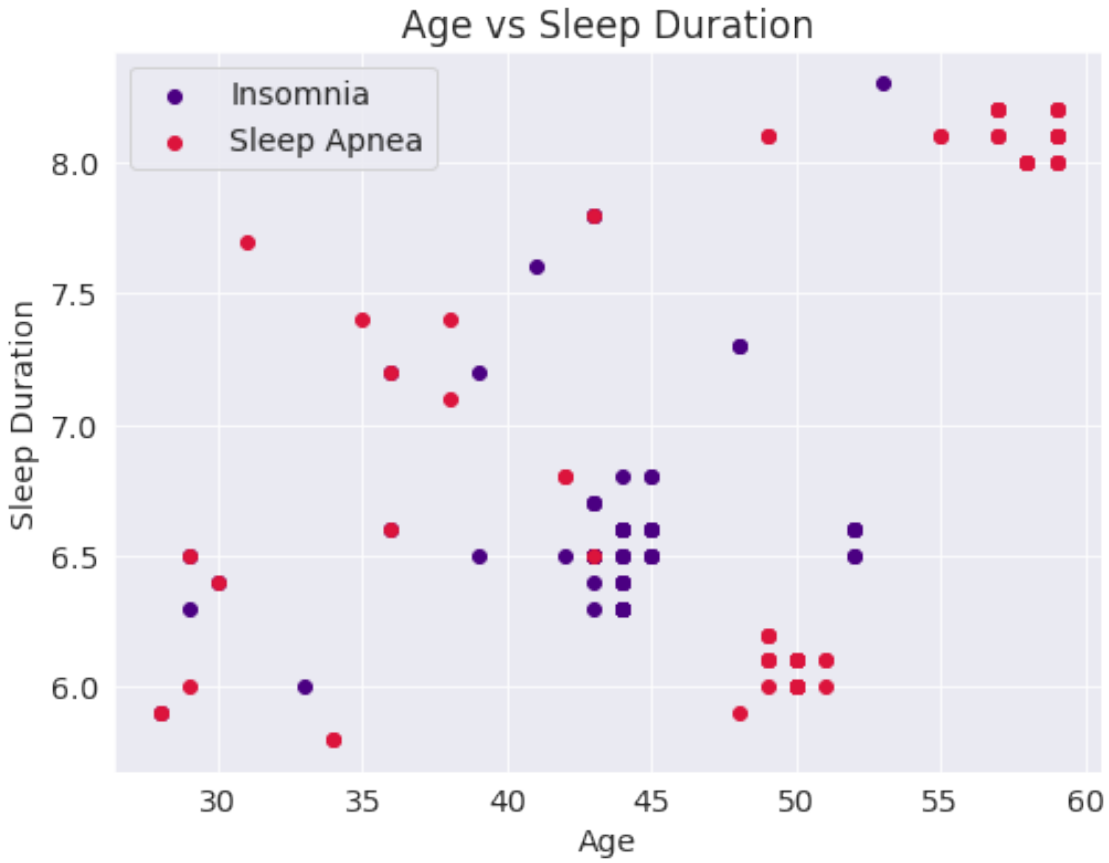


The bar chart illustrates the BMI distribution among individuals with insomnia and sleep apnea. Both conditions most frequently occur in individuals classified as overweight, followed by those in the normal and obese categories. Notably, the count of overweight individuals is roughly equal for both sleep disorders, suggesting that excess weight might be a common risk factor. However, the representation of obese individuals is noticeably higher in the sleep apnea group, which aligns with clinical observations that sleep apnea is often associated with higher obesity rates. This could be due to the fact that excess body weight can contribute to obstructed breathing during sleep. In contrast, the distribution of normal BMI among individuals with insomnia and sleep apnea is significantly lower, indicating that normal weight individuals are less represented in this dataset's sleep disorder population. The data underscores the potential link between body weight and the prevalence of sleep disorders, emphasizing the importance of considering BMI in the management and treatment of sleep-related conditions.

1.11 Graph 5: Age and Sleeping Disorders

[31]:
```
scatter_insomnia = insomnia_data[['Age', 'Sleep Duration']]
scatter_sleep_apnea = sleep_apnea_data[['Age', 'Sleep Duration']]

plt.figure(figsize=(8, 6))
plt.scatter(scatter_insomnia['Age'], scatter_insomnia['Sleep Duration'],␣
↪label='Insomnia', color = 'indigo')
```

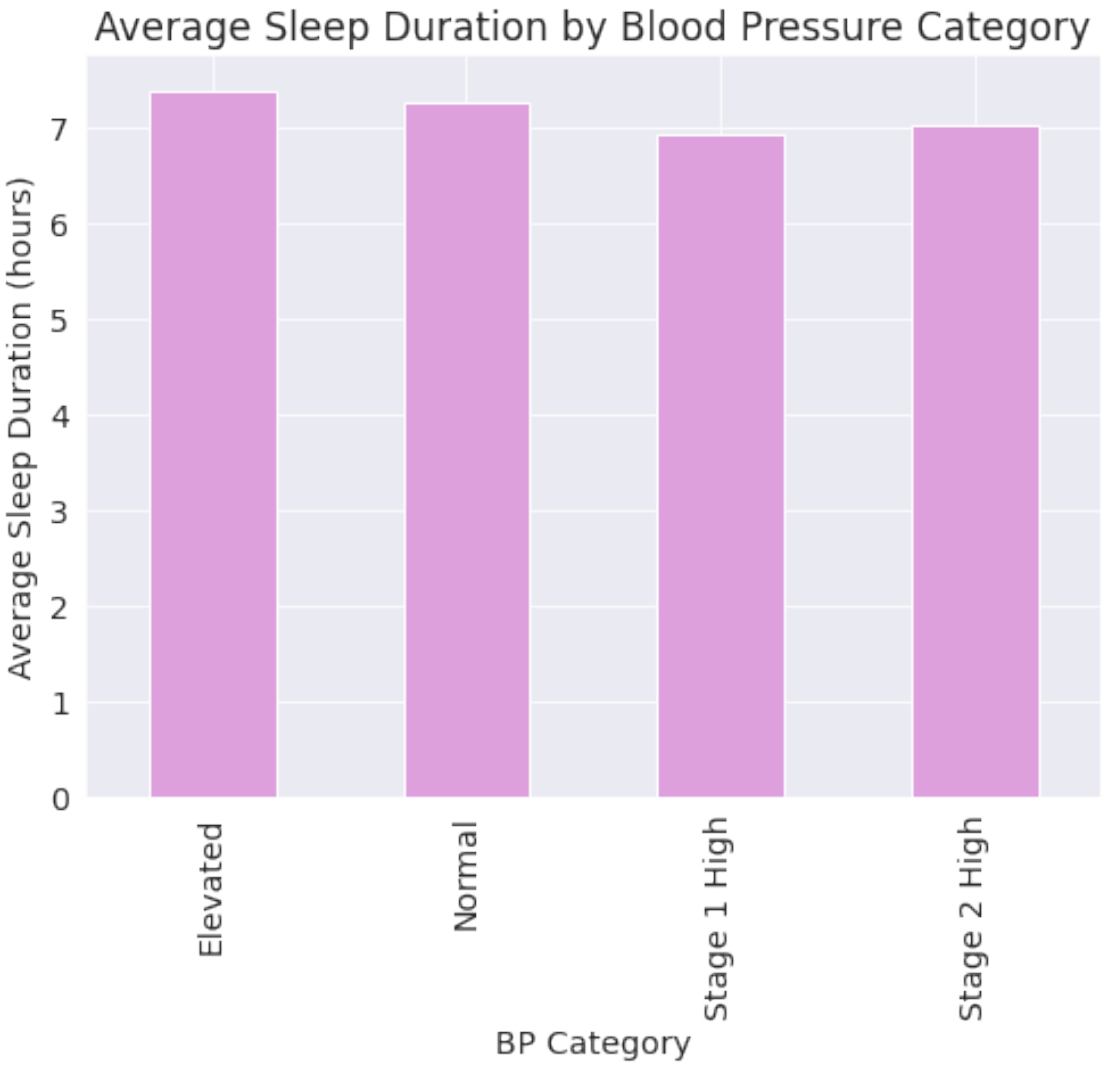
```
plt.scatter(scatter_sleep_apnea['Age'], scatter_sleep_apnea['Sleep Duration'],␣
 ↪label='Sleep Apnea', color = 'crimson')

plt.xlabel('Age')
plt.ylabel('Sleep Duration')
plt.title('Age vs Sleep Duration')
plt.legend()
plt.show()
```



The scatter plot depicts the relationship between age and sleep duration for individuals with insomnia and sleep apnea. Points are dispersed across the age spectrum from 30 to 60 years, with sleep duration ranging from 6 to 8 hours. Notably, there is no clear trend indicating that age has a direct influence on sleep duration for either condition. Individuals with sleep apnea tend to cluster more densely at higher sleep durations across all ages compared to those with insomnia, who display a broader spread of sleep durations. This could reflect the varied impact of insomnia on sleep quantity, whereas sleep apnea's effect might be more consistent in terms of sleep duration. However, both conditions show presence across the full age range examined, suggesting that neither condition is confined to a specific age group. The plot reinforces the complexity of sleep disorders and the necessity of personalized approaches in understanding and treating these conditions, as age alone does not appear to be a determinant factor for sleep duration among the affected individuals.

1.12 Graph 6: Sleep Duration and Blood Pressure

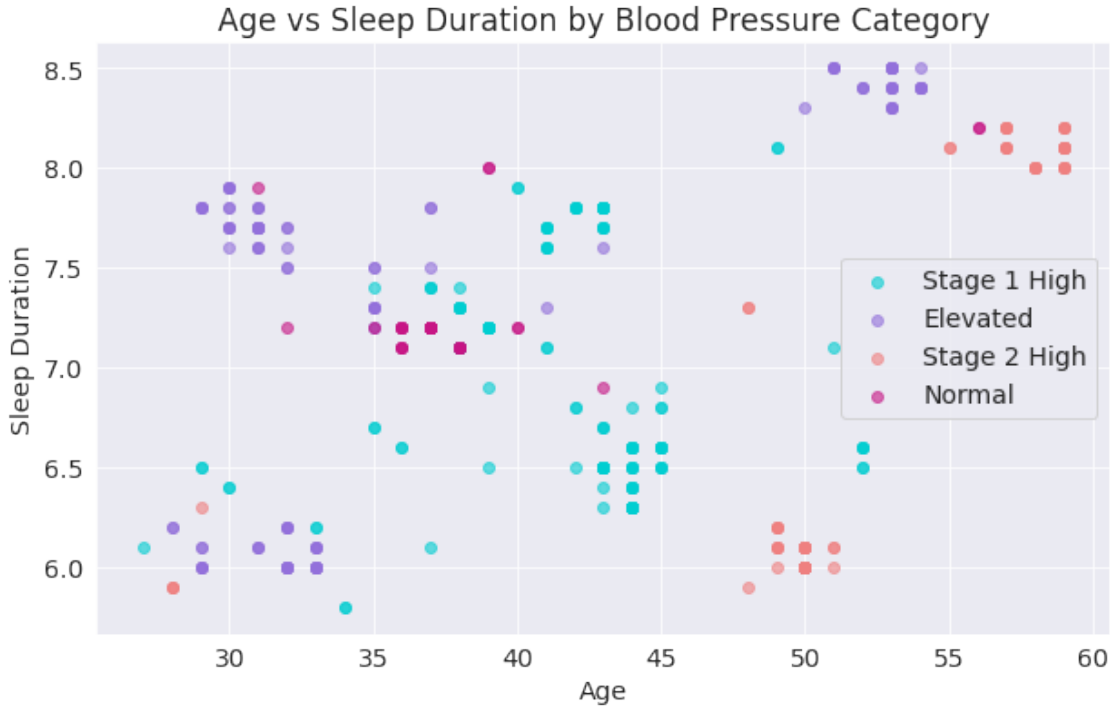
[32]:
```
def categorize_blood_pressure(bp):
    systolic, diastolic = map(int, bp.split('/'))
    if systolic < 120 and diastolic <= 80:
        return 'Normal'
    elif 120 <= systolic <= 129 and diastolic <= 80:
        return 'Elevated'
    elif 130 <= systolic <= 139 or 80 <= diastolic <= 89:
        return 'Stage 1 High'
    elif systolic >= 140 or 80 <= diastolic <= 89:
        return 'Stage 2 High'

# Apply the function to the Blood Pressure column
sleep_clean_df['BP Category'] = sleep_clean_df['Blood Pressure'].
 ↪apply(categorize_blood_pressure)

# Group by BP Category and calculate average sleep duration
avg_sleep_by_bp = sleep_clean_df.groupby('BP Category')['Sleep Duration'].mean()

# Plotting
plt.figure(figsize=(8, 6))
avg_sleep_by_bp.plot(kind='bar', color = 'plum')
plt.title('Average Sleep Duration by Blood Pressure Category')
plt.ylabel('Average Sleep Duration (hours)')
plt.show()
```



The bar chart presents average sleep duration across different blood pressure categories: Elevated, Normal, Stage 1 High, and Stage 2 High. Remarkably, the sleep duration does not significantly vary with blood pressure categories, as all groups report an average sleep duration of approximately 6 to just over 6.5 hours. This uniformity in sleep duration suggests that blood pressure levels, as classified here, may not be a major differentiator in how long people sleep on average. The consistency across categories could imply that individuals with high blood pressure manage to achieve a similar quantity of sleep as those with normal levels, contrary to common concerns that high blood pressure could substantially disrupt sleep. However, it's important to consider that these averages do not reflect sleep quality, which can be significantly affected by blood pressure issues.

1.13 Graph 7: Blood Pressure and Age

[33]:
```
category_colors = {
    'Normal': 'mediumvioletred',
    'Elevated': 'mediumpurple',
    'Stage 1 High': 'darkturquoise',
    'Stage 2 High': 'lightcoral'
}

# Scatter plot for age vs. sleep duration for each BP category
plt.figure(figsize=(10, 6))
for category in sleep_clean_df['BP Category'].unique():
    subset = sleep_clean_df[sleep_clean_df['BP Category'] == category]
    plt.scatter(subset['Age'], subset['Sleep Duration'], label=category,␣
 ↪alpha=0.6,color=category_colors[category])

plt.title('Age vs Sleep Duration by Blood Pressure Category')
plt.xlabel('Age')
plt.ylabel('Sleep Duration')
plt.legend()
plt.show()
```



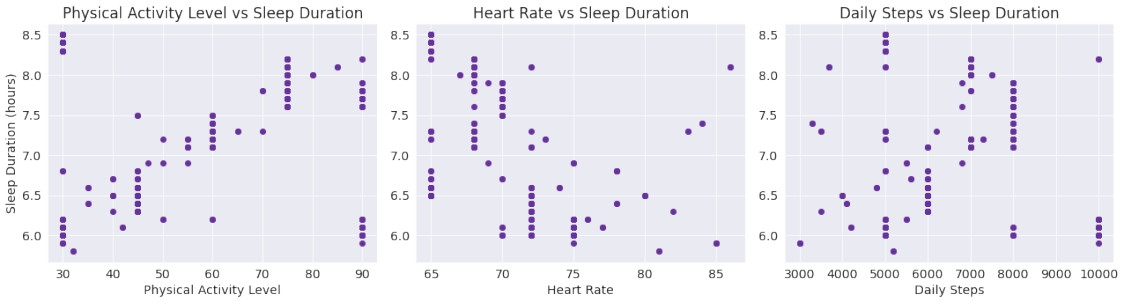
This scatter plot visualizes the relationship between age and sleep duration across different blood pressure categories. The data points are spread across ages 30 to 60 and show a range of sleep durations from just above 6 hours to 8.5 hours, regardless of blood pressure status.

The distribution does not indicate a clear pattern or trend of sleep duration varying with age within any blood pressure category. Individuals with normal blood pressure, elevated blood pressure, and stages 1 and 2 high blood pressure appear to have a wide variation in sleep duration, with no distinct differences between the categories.

Moreover, there's no evident clustering or correlation that suggests older or younger age groups experience shorter or longer sleep due to blood pressure. All categories scatter broadly across the sleep duration spectrum, implying that within this dataset, blood pressure status is not a strong predictor of sleep duration, and both younger and older adults experience a similar range of sleep lengths. This plot reinforces the complex nature of sleep and its multifactorial influences, suggesting that factors other than blood pressure and age may play more significant roles in determining sleep duration.

1.14 Graphs 8-10: Sleep and Lifestyle Choices

[34]:
```
import matplotlib.pyplot as plt

# Assuming 'df' is your DataFrame and it contains 'Sleep Duration' along with␣
 ↪the other variables
fig, axes = plt.subplots(1, 3, figsize=(18, 5))

# Scatter plot for Physical Activity Level vs Sleep Duration
axes[0].scatter(sleep_clean_df['Physical Activity Level'],␣
 ↪sleep_clean_df['Sleep Duration'], color ='rebeccapurple')
axes[0].set_title('Physical Activity Level vs Sleep Duration')
axes[0].set_xlabel('Physical Activity Level')
axes[0].set_ylabel('Sleep Duration (hours)')

# Scatter plot for Heart Rate vs Sleep Duration
axes[1].scatter(sleep_clean_df['Heart Rate'], sleep_clean_df['Sleep Duration'],␣
 ↪color ='rebeccapurple')
axes[1].set_title('Heart Rate vs Sleep Duration')
axes[1].set_xlabel('Heart Rate')

# Scatter plot for Daily Steps vs Sleep Duration
axes[2].scatter(sleep_clean_df['Daily Steps'], sleep_clean_df['Sleep␣
 ↪Duration'], color ='rebeccapurple')
axes[2].set_title('Daily Steps vs Sleep Duration')
axes[2].set_xlabel('Daily Steps')

plt.tight_layout()
plt.show()
```



The set of scatter plots represents sleep duration against three variables: physical activity level, heart rate, and daily steps. Across the board, sleep duration varies without a clear or consistent pattern in relation to these variables.

In the first plot, physical activity levels range widely, but there's no distinct trend suggesting higher activity levels lead to longer sleep durations. Similarly, the second plot shows no obvious correlation between heart rate and sleep duration. The third plot, concerning daily steps, also fails to display a clear relationship with sleep duration. While there are clusters of data points at higher step counts correlating with longer sleep durations, the overall spread does not indicate a strong or linear association.

These visualizations suggest that within the dataset, the simple measures of physical activity, heart rate, and daily steps may not be directly predictive of sleep duration. It's possible that the interrelations between sleep duration and these variables are more nuanced, potentially influenced by other factors not captured in these plots. It underscores the complexity of sleep as a physiological process influenced by a multitude of interdependent factors.

Let us save and upload our work to Jovian before continuing

[35]:
```
import jovian
```

[36]:
```
jovian.commit()
```

```
<IPython.core.display.Javascript object>
```

```
[jovian] Updating notebook "insiyah-yu/sleep-health-and-lifestyle-project" on
https://jovian.com
[jovian] Committed successfully! https://jovian.com/insiyah-yu/sleep-health-and-
lifestyle-project
```

[36]:
```
'https://jovian.com/insiyah-yu/sleep-health-and-lifestyle-project'
```

1.15 Asking and Answering Questions

Q1: Quality of Sleep and Sleeping Disorders In the third visualization, we analyzed the influence of sleep disorders on the duration of sleep among individuals, revealing that conditions such as insomnia and sleep apnea may not significantly reduce the total hours of sleep. However, these disorders could potentially compromise the quality of sleep, which is a crucial component of

overall health. To further investigate this, we will delve into the dataset to examine the nuances of how sleep disorders, notably insomnia and sleep apnea, impact the subjective quality of sleep as reported by the individuals. By scrutinizing this aspect, we aim to uncover the subtler, qualitative aspects of sleep that quantitative measures alone may overlook. This analysis could provide valuable insights into the broader implications of sleep disorders on well-being and daily functioning.

[50]:
```
average_quality_sleep = pd.pivot_table(
    sleep_clean_df[sleep_clean_df['Sleep Disorder'].isin(['Insomnia', 'Sleep␣
 ↪Apnea'])],
    values='Quality of Sleep',
    index='Sleep Disorder',
    aggfunc='mean'
).reset_index()
average_quality_sleep
```

[50]:

	Sleep Disorder	Quality of Sleep
0	Insomnia	6.532468
1	Sleep Apnea	7.205128

The findings from our dataset yield a somewhat surprising narrative: individuals who are affected by sleep apnea not only tend to achieve longer periods of sleep but also report, on average, a superior quality of sleep when compared to those grappling with insomnia. This revelation is counterintuitive, as sleep apnea is typically associated with fragmented sleep that could be presumed to diminish sleep quality. Conversely, despite insomnia being characterized by significant difficulties in initiating and maintaining sleep, it is traditionally not associated with the frequent nocturnal awakenings characteristic of sleep apnea. These unexpected results underscore the complexity of sleep disorders and their multifaceted impact on sleep quality, suggesting that the subjective experience of sleep quality may not always align with the clinical symptoms presented by these conditions.

Q2: Quality of Sleep, Age and Sleeping Disorders In the fifth graph, our analysis revealed a lack of a distinct trend correlating age with sleep duration for individuals suffering from either insomnia or sleep apnea. This intriguing observation prompts a deeper exploration into the dataset to understand how the quality of sleep is influenced by a combination of factors - namely, the type of sleeping disorder and the age of the individual.

[56]:
```
pd.pivot_table(
    sleep_clean_df[sleep_clean_df['Sleep Disorder'].isin(['Insomnia', 'Sleep␣
 ↪Apnea'])],
    values='Quality of Sleep',
    index=['Sleep Disorder', 'Age'],
    aggfunc='mean'
).reset_index()
```

[56]:

	Sleep Disorder	Age	Quality of Sleep
0	Insomnia	28	4.000000
1	Insomnia	29	5.500000
2	Insomnia	30	5.000000

3	Insomnia	33	6.000000
4	Insomnia	36	6.500000
5	Insomnia	39	6.500000
6	Insomnia	41	8.000000
7	Insomnia	42	6.000000
8	Insomnia	43	6.411765
9	Insomnia	44	6.428571
10	Insomnia	45	7.000000
11	Insomnia	48	7.000000
12	Insomnia	50	6.000000
13	Insomnia	52	7.000000
14	Insomnia	53	9.000000
15	Sleep Apnea	28	4.000000
16	Sleep Apnea	29	5.500000
17	Sleep Apnea	30	5.000000
18	Sleep Apnea	31	7.000000
19	Sleep Apnea	34	4.000000
20	Sleep Apnea	35	7.000000
21	Sleep Apnea	36	6.500000
22	Sleep Apnea	38	7.500000
23	Sleep Apnea	42	6.000000
24	Sleep Apnea	43	7.000000
25	Sleep Apnea	48	6.000000
26	Sleep Apnea	49	6.600000
27	Sleep Apnea	50	6.000000
28	Sleep Apnea	51	6.000000
29	Sleep Apnea	55	9.000000
30	Sleep Apnea	57	9.000000
31	Sleep Apnea	58	9.000000
32	Sleep Apnea	59	9.000000

The dataset reveals intriguing insights into the quality of sleep among individuals with insomnia and sleep apnea across different ages. For those with insomnia, a trend of gradually improving sleep quality is observed with increasing age. Young adults in their late 20s and early 30s report lower quality scores (around 4 to 5), but this improves to scores of 6 to 7 by the late 40s and peaks at age 53 with a score of 9.

In contrast, individuals with sleep apnea exhibit a less consistent pattern. While similarly lower quality scores are noted in the late 20s, there's a quicker improvement to scores of 7 in the early 30s. The quality of sleep fluctuates in the 40s, but significantly, older age groups, particularly in the mid to late 50s, consistently report high sleep quality scores of 9.

These findings suggest that while age might influence sleep quality in those suffering from sleep disorders, the trajectory of this influence differs between insomnia and sleep apnea. The marked improvement in sleep quality among older individuals with sleep apnea is especially notable, indicating potential age-related adaptations or effective management strategies for the disorder in later life.

Q3: Heart Rate and Sleep How does heart rate impact an individual's sleep patterns?

[55]:
```
sleep_clean_df.groupby('Heart Rate').agg({'Sleep Duration': 'mean', 'Quality of␣
↪Sleep': 'mean'})
```

[55]:

Heart Rate	Sleep Duration	Quality of Sleep
65	7.577612	8.104478
67	8.000000	9.000000
68	7.523404	8.329787
69	7.400000	7.000000
70	7.561842	7.394737
72	6.385507	6.246377
73	7.200000	8.000000
74	6.600000	5.000000
75	6.119444	6.055556
76	6.200000	6.000000
77	6.100000	6.000000
78	6.640000	5.800000
80	6.500000	5.000000
81	5.800000	4.000000
82	6.300000	6.000000
83	7.300000	7.000000
84	7.400000	7.000000
85	5.900000	4.000000
86	8.100000	9.000000

The results indicate a relationship between heart rate and sleep metrics such as sleep duration and quality of sleep. As heart rate values vary, so do the corresponding sleep patterns.

For instance, individuals with heart rates around 67 and 86 have notably higher sleep quality ratings, with quality scores of 9.0. In contrast, heart rates around 81 and 85 seem to correlate with lower sleep quality, with quality scores of 4.0.

Regarding sleep duration, heart rates around 67 and 68 show the highest mean sleep durations, around 8 hours, while heart rates around 72 and 80 correspond to the lowest sleep durations, averaging around 6 hours.

These findings suggest that there may be a connection between heart rate and sleep, with lower heart rates generally associated with longer sleep duration and higher sleep quality. However, it's important to note that these are observational results, and many factors can influence sleep, so further research and consideration of confounding variables would be necessary to draw definitive conclusions about causation.

Q3: Stress Levels and Quality of Sleep Do gender-related differences exist in susceptibility to higher stress levels, and if so, how does this gender disparity impact sleep quality?

[52]:
```
sleep_clean_df.groupby('Gender')[['Stress Level', 'Quality of Sleep']].mean()
```

[52]:

Gender	Stress Level	Quality of Sleep
Female	4.675676	7.664865
Male	6.079365	6.968254

The analysis of stress levels and sleep quality by gender reveals notable differences between females and males.

On average, females in the dataset report lower stress levels (mean of approximately 4.68) compared to males, who report higher stress levels (mean of approximately 6.08). This suggests that, within this sample, females tend to experience lower levels of perceived stress.

In terms of sleep quality, females also report higher mean scores (approximately 7.66) compared to males (approximately 6.97). This indicates that, on average, females in the dataset tend to report better sleep quality.

These findings suggest a potential correlation between gender, stress levels, and sleep quality, with females in this dataset reporting lower stress levels and better sleep quality compared to males.

Q4: Men and Women in the Workplace: Sleep and Stress This compelling heading explores the intricate interplay between gender, professional life, sleep patterns, and stress levels. Delving into various occupations, it sheds light on how men and women navigate their work environments, revealing intriguing insights into their unique sleep habits and the stress they encounter along the way. Discover the disparities and commonalities in the way gender impacts the world of work, rest, and well-being.

[54]:
```
sleep_clean_df.groupby(['Occupation', 'Gender'])[['Sleep Duration', 'Quality of␣
↪Sleep', 'Stress Level']].mean()
```

[54]:

Occupation	Gender	Sleep Duration	Quality of Sleep	Stress Level
Accountant	Female	7.111111	7.888889	4.555556
	Male	7.200000	8.000000	6.000000
Doctor	Female	8.200000	9.000000	3.000000
	Male	6.934783	6.579710	6.840580
Engineer	Female	8.425000	9.000000	3.000000
	Male	7.535484	7.806452	4.806452
Lawyer	Female	7.150000	7.000000	5.500000
	Male	7.422222	7.933333	5.044444
Manager	Female	6.900000	7.000000	5.000000
Nurse	Female	7.063014	7.369863	5.547945
Sales Representative	Male	5.900000	4.000000	8.000000
Salesperson	Male	6.403125	6.000000	7.000000
Scientist	Female	6.000000	5.000000	7.000000
Software Engineer	Male	6.750000	6.500000	6.000000
Teacher	Female	6.705714	7.114286	4.285714
	Male	6.580000	6.000000	6.200000

The analysis of sleep-related factors, including sleep duration, quality of sleep, and stress level,

across various occupations and genders reveals interesting patterns.

1. **Sleep Duration**: Across different occupations, females generally tend to have slightly longer sleep durations compared to males. However, the differences are relatively small and may not be statistically significant in some cases.
2. **Quality of Sleep**: Females in occupations such as doctors, engineers, and nurses report higher quality of sleep compared to their male counterparts. This suggests that females in certain professions may experience better sleep quality.
3. **Stress Level**: Notably, males in sales-related occupations, such as Sales Representatives and Salespersons, report higher stress levels compared to other professions, regardless of gender. On the other hand, females in professions like doctors, engineers, and nurses report lower stress levels.

These results highlight potential gender and occupation-related variations in sleep and stress.

Q5: TODO - ask a question here and answer it below

[57]:
```
sleep_clean_df.groupby(['Age', 'Gender'])[['Sleep Duration', 'Quality of␣
↪Sleep', 'Stress Level']].mean()
```

[57]:

Age	Gender	Sleep Duration	Quality of Sleep	Stress Level
27	Male	6.100000	6.000000	6.000000
28	Male	6.020000	4.800000	8.000000
29	Female	6.500000	5.000000	7.000000
	Male	6.700000	6.363636	7.181818
30	Female	6.400000	5.000000	7.000000
	Male	7.781818	7.000000	6.000000
31	Female	7.900000	8.000000	4.000000
	Male	7.417647	6.823529	6.352941
32	Male	6.588235	6.529412	6.941176
33	Female	6.200000	6.000000	6.000000
	Male	6.036364	6.000000	8.000000
34	Female	5.800000	4.000000	8.000000
35	Female	7.200000	8.000000	4.000000
	Male	7.218182	7.727273	4.454545
36	Female	7.160000	8.000000	4.000000
	Male	6.600000	5.000000	7.000000
37	Female	7.146667	7.866667	4.133333
	Male	7.560000	8.000000	4.600000
38	Female	7.133333	7.888889	4.111111
	Male	7.227273	8.000000	5.000000
39	Female	7.633333	8.333333	4.000000
	Male	7.141667	7.750000	5.166667
40	Female	7.200000	8.000000	6.000000
	Male	7.900000	8.000000	5.000000
41	Male	7.533333	7.833333	5.250000
42	Female	6.800000	6.000000	7.000000

	Male	7.614286	7.714286	5.285714
43	Female	6.700000	7.000000	4.000000
	Male	7.193548	7.096774	5.870968
44	Female	6.563636	7.000000	4.000000
	Male	6.368421	6.052632	7.000000
45	Female	6.621429	7.000000	4.357143
48	Female	5.900000	6.000000	8.000000
	Male	7.300000	7.000000	5.000000
49	Female	6.133333	6.000000	8.000000
	Male	8.100000	9.000000	3.000000
50	Female	6.155000	6.150000	7.750000
51	Female	7.412500	7.625000	5.250000
52	Female	7.177778	7.666667	5.666667
53	Female	8.423529	9.000000	3.000000
54	Female	8.414286	9.000000	3.000000
55	Female	8.100000	9.000000	4.000000
56	Female	8.200000	9.000000	3.000000
57	Female	8.155556	9.000000	3.000000
58	Female	8.000000	9.000000	3.000000
59	Female	8.093750	9.000000	3.000000

The data presents a comprehensive overview of sleep duration, quality of sleep, and stress levels across various age groups and genders. Several key observations can be made:

1. **Age Trends**: Sleep duration and quality of sleep seem to fluctuate with age. Generally, sleep duration remains relatively stable or increases slightly with age, particularly among females. However, stress levels show less consistency, with some age groups reporting higher stress levels than others.
2. **Gender Differences**: In many age groups, females tend to report better quality of sleep than males. Stress levels, on the other hand, exhibit variations, with no clear gender-related pattern in stress levels.
3. **Outliers**: Notably, at age 49, males report a remarkably high stress level of 9.0, suggesting a significant stressor in that particular age group.
4. **Age and Sleep**: As individuals approach their late 40s and early 50s, both genders tend to report higher stress levels. However, females in this age group also report higher sleep quality.

In conclusion, this data underscores the complex interplay between age, gender, sleep patterns, and stress levels. While some trends emerge, such as females generally reporting better sleep quality, the relationship between age and stress is less straightforward.

Let us save and upload our work to Jovian before continuing.

[58]:
```
import jovian
```

[59]:
```
jovian.commit()
```

```
<IPython.core.display.Javascript object>
```

```
[jovian] Updating notebook "insiyah-yu/sleep-health-and-lifestyle-project" on
https://jovian.com
[jovian] Committed successfully! https://jovian.com/insiyah-yu/sleep-health-and-
lifestyle-project
```

[59]: 'https://jovian.com/insiyah-yu/sleep-health-and-lifestyle-project'

1.16 Inferences and Conclusion

The relationship between sleep health and lifestyle is multifaceted, with various parameters such as age, occupation, stress level, BMI, heart rate, sleep disorders, and blood pressure all playing pivotal roles.

General lifestyle choices, including daily routines, exercise habits, and stress management, significantly impact sleep quality and duration. Different age groups and occupational demands can influence sleep patterns, and stress levels often correlate with sleep disturbances.

Additionally, factors like BMI, heart rate, and the presence of sleep disorders can further complicate sleep health. Engaging in regular physical activity and monitoring daily steps can positively affect sleep, while blood pressure and other health indicators may provide insights into overall sleep and wellness. Recognizing the intricate interplay of these parameters is crucial for fostering healthier sleep habits and ultimately enhancing overall well-being.

It's important to note that all the analyses conducted are based on a single dataset and should be interpreted as specific to the data within that set. These findings may not necessarily be indicative of trends or characteristics within the broader general population, as the dataset's composition, size, and demographics can influence the observed results. For a more comprehensive and reliable understanding of sleep health and lifestyle patterns in the general population, broader and more diverse datasets, along with additional research, would be needed.

[60]:
```
import jovian
```

[61]:
```
jovian.commit()
```

```
<IPython.core.display.Javascript object>
```

```
[jovian] Updating notebook "insiyah-yu/sleep-health-and-lifestyle-project" on
https://jovian.com
[jovian] Committed successfully! https://jovian.com/insiyah-yu/sleep-health-and-
lifestyle-project
```

[61]: 'https://jovian.com/insiyah-yu/sleep-health-and-lifestyle-project'