Your client, SleepInc, has shared anonymized sleep data from their hot new sleep tracking app SleepScope. As their data science consultant, your mission is to analyze the lifestyle survey data with Python to discover relationships between exercise, gender, occupation, and sleep quality. See if you can identify patterns leading to insights on sleep quality.

💾 The data: sleep_health_data.csv

SleepInc has provided you with an anonymized dataset of sleep and lifestyle metrics for 374 individuals. This dataset contains average values for each person calculated over the past six months. The data is saved as `sleep_health_data.csv`.

The dataset includes 13 columns covering sleep duration, quality, disorders, exercise, stress, diet, demographics, and other factors related to sleep health.

Column	Description
`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 (hours)`	The average number of hours the person sleeps per day.
`Quality of Sleep (scale: 1-10)`	A subjective rating of the quality of sleep, ranging from 1 to 10.
`Physical Activity Level (minutes/day)`	The average number of minutes the person engages in physical activity daily.
`Stress Level (scale: 1-10)`	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 (systolic/diastolic)`	The average blood pressure measurement of the person, indicated as systolic pressure over diastolic pressure.
`Heart Rate (bpm)`	The average resting heart rate of the person in beats per minute.
`Daily Steps`	The average 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).

```python
import pandas as pd
import numpy as np

# Load data
sleep_data = pd.read_csv("sleep_health_data.csv")

# Check the columns of the dataframe
print(sleep_data.columns)

# Find average sleep duration per occupation
occ_avg_sleep = sleep_data.groupby("Occupation")["Sleep Duration"].mean()

# Find the occupation with the lowest average sleep duration
lowest_sleep_occ = occ_avg_sleep.idxmin()
print("Lowest Average Sleep Duration by Occupation:", lowest_sleep_occ)

# Find average sleep quality per occupation
occ_avg_quality = sleep_data.groupby("Occupation")["Quality of Sleep"].mean()

# Find the occupation with the lowest average sleep quality
lowest_sleep_quality_occ = occ_avg_quality.idxmin()
print("Lowest Average Sleep Quality by Occupation:", lowest_sleep_quality_occ)

same_occ = (lowest_sleep_occ == lowest_sleep_quality_occ )
print("Is it the same occupation as the one with the lowest individual record?", same_occ)

# Explore BMI Category effect on insomnia
# Calculate ratio of people in each BMI Category diagnosed with insomnia
bmi_insomnia_ratios = (
    sleep_data.groupby("BMI Category")
    .apply(lambda x: round((x["Sleep Disorder"].eq("Insomnia").mean()), 2))
    .to_dict()
)

print("BMI Category Insomnia Ratios:", bmi_insomnia_ratios)
```

```
Index(['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'],
      dtype='object')
Lowest Average Sleep Duration by Occupation: Sales Representative
Lowest Average Sleep Quality by Occupation: Sales Representative
Is it the same occupation as the one with the lowest individual record? True
BMI Category Insomnia Ratios: {'Normal': 0.04, 'Obese': 0.4, 'Overweight': 0.43}
```