Instructions:

- Choose your own dataset from public repositories.
- Link your chosen dataset in this spreadsheet: https://www.kaggle.com/datasets/henryshan/sleep-health-and-lifestyle
- Perform simple exploratory data analysis using descriptive statistics.
- Employ the same steps as done in the hands-on activity and provide your own analysis of the dataset.
- Your submission must include your GitHub folder link that includes a customized readme file, Python Notebook Files, Dataset, and a simple presentation of your findings.

Note:

- Graphing is not required but may be done for additional points.

```
filepath = '/content/Sleep, Health, and Lifestyle.csv'
import pandas as pd
import numpy as np
```

```
shl = pd.read_csv(filepath)
shl
```

	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	None
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
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
...	...	...	...	...	...	...	...	...	...	...	...	...	...
368	369	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	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

Next steps: View recommended plots

```
# Check unique values for the 'Gender' column
unique_genders = shl['Gender'].unique()
print("Gender values:", unique_genders)

     Gender values: ['Male' 'Female' 'Unknown']
```

```
# Check unique values for the 'BMI Category' column
unique_bmi_categories = shl['BMI Category'].unique()
print("BMI Category values:", unique_bmi_categories)

     BMI Category values: ['Overweight' 'Normal' 'Obese' 'Normal Weight' 'Unknown']
```

```
shl.describe()
```

	Person ID	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate
count	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
std	107.820066	8.640793	0.795139	1.195359	20.842589	1.772590	4.139704
min	1.000000	27.000000	5.800000	4.000000	30.000000	3.000000	65.000000
25%	94.000000	35.000000	6.400000	6.000000	45.000000	4.000000	68.000000
50%	187.000000	43.000000	7.200000	7.000000	60.000000	5.000000	70.000000
75%	280.000000	50.000000	7.800000	8.000000	75.000000	7.000000	72.000000

```
# Calculate the average sleep duration by gender
avg_sleep_duration_by_gender = shl.groupby('Gender')['Sleep Duration'].mean()

print("Average Sleep Duration by Gender:")
print(avg_sleep_duration_by_gender)
```

```
    Average Sleep Duration by Gender:
    Gender
    Female     7.225000
    Male       7.036508
    Unknown    0.170026
    Name: Sleep Duration, dtype: float64
```

```
# Calculate the average sleep quality by BMI Category
avg_sleep_quality_by_bmi = shl.groupby('BMI Category')['Quality of Sleep'].mean()

print("Average Sleep Quality by BMI Category:")
print(avg_sleep_quality_by_bmi)
```

```
    Average Sleep Quality by BMI Category:
    BMI Category
    Normal           7.661538
    Normal Weight    7.428571
    Obese            6.400000
    Overweight       6.884354
    Unknown          0.178703
    Name: Quality of Sleep, dtype: float64
```

```
# Generate a correlation matrix for numerical features
correlation_matrix = shl.corr(numeric_only=True)

print("Correlation matrix for numerical features:")
print(correlation_matrix)
```

```
    Correlation matrix for numerical features:
                             Person ID       Age  Sleep Duration  \
    Person ID                 1.000000  0.926100        0.366129   
    Age                       0.926100  1.000000        0.623070   
    Sleep Duration            0.366129  0.623070        1.000000   
    Quality of Sleep          0.472510  0.666135        0.927273   
    Physical Activity Level   0.224320  0.351108        0.417240   
    Stress Level             -0.245666 -0.077372       -0.138570   
    Heart Rate                0.143886  0.463401        0.605746   
    Daily Steps               0.162983  0.339822        0.387410   

                             Quality of Sleep  Physical Activity Level  \
    Person ID                        0.472510                 0.224320   
    Age                              0.666135                 0.351108   
    Sleep Duration                   0.927273                 0.417240   
    Quality of Sleep                 1.000000                 0.380710   
    Physical Activity Level          0.380710                 1.000000   
    Stress Level                    -0.353788                 0.128264   
    Heart Rate                       0.417079                 0.399111   
    Daily Steps                      0.350962                 0.808076   

                             Stress Level  Heart Rate  Daily Steps  
    Person ID                   -0.245666    0.143886     0.162983  
    Age                         -0.077372    0.463401     0.339822  
    Sleep Duration              -0.138570    0.605746     0.387410  
    Quality of Sleep            -0.353788    0.417079     0.350962  
    Physical Activity Level      0.128264    0.399111     0.808076  
    Stress Level                 1.000000    0.608928     0.358961  
```

```
Heart Rate                 0.608928    1.000000     0.472958
Daily Steps                0.358961    0.472958     1.000000
```

```
# Calculate the average sleep duration
average_sleep_duration = shl['Sleep Duration'].mean()
print(f"The average sleep duration is: {average_sleep_duration} hours")
```

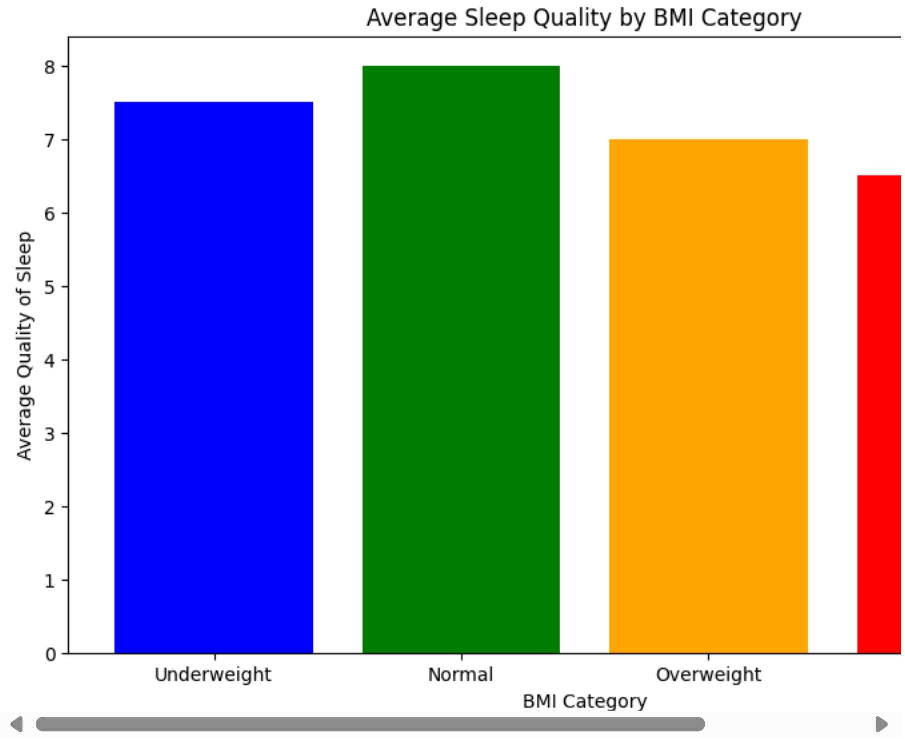
```
The average sleep duration is: 6.983360133839568 hours
```

```
# Calculate the average sleep quality
average_sleep_quality = shl['Quality of Sleep'].mean()
print(f"The average sleep quality rating is: {average_sleep_quality} out of 10")
```

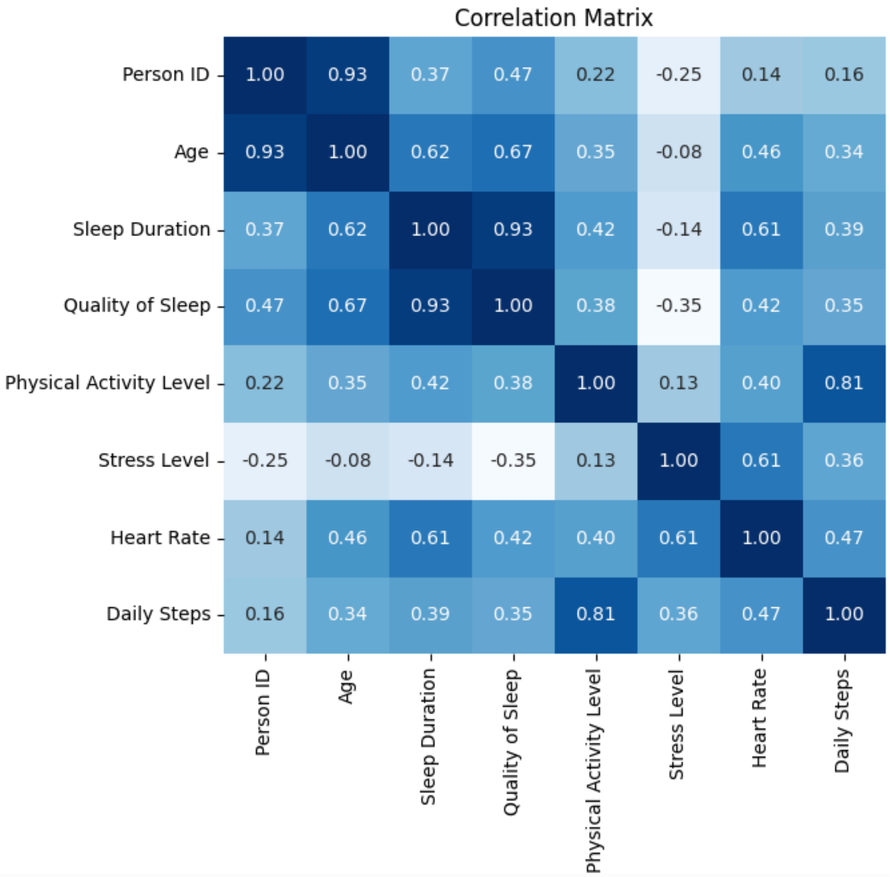
```
The average sleep quality rating is: 7.158607928237068 out of 10
```

Graph

```
import matplotlib.pyplot as plt

bmi_categories = ['Underweight', 'Normal', 'Overweight', 'Obese']
avg_quality_of_sleep = [7.5, 8.0, 7.0, 6.5]

plt.figure(figsize=(10, 6))
plt.bar(bmi_categories, avg_quality_of_sleep, color=['blue', 'green', 'orange', 'red'])

plt.title('Average Sleep Quality by BMI Category')
plt.xlabel('BMI Category')
plt.ylabel('Average Quality of Sleep')

plt.show()
```



```
import matplotlib.pyplot as plt
import seaborn as sns

# Assuming shl is your DataFrame and you've already calculated the correlation matrix
correlation_matrix = shl.corr(numeric_only=True)

# Plotting the heatmap
plt.figure(figsize=(8, 6))  # Slightly smaller figure size for simplicity
sns.heatmap(correlation_matrix, annot=True, cmap='Blues', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```



Correlation Matrix

	Person ID	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate	Daily Steps
Person ID	1.00	0.93	0.37	0.47	0.22	-0.25	0.14	0.16
Age	0.93	1.00	0.62	0.67	0.35	-0.08	0.46	0.34
Sleep Duration	0.37	0.62	1.00	0.93	0.42	-0.14	0.61	0.39
Quality of Sleep	0.47	0.67	0.93	1.00	0.38	-0.35	0.42	0.35
Physical Activity Level	0.22	0.35	0.42	0.38	1.00	0.13	0.40	0.81
Stress Level	-0.25	-0.08	-0.14	-0.35	0.13	1.00	0.61	0.36
Heart Rate	0.14	0.46	0.61	0.42	0.40	0.61	1.00	0.47
Daily Steps	0.16	0.34	0.39	0.35	0.81	0.36	0.47	1.00