

# Sleep Health and Lifestyle Analysis with Python

Shareen Banu

Problem Statement Delving into a comprehensive analysis of sleep patterns, factors affecting our daily lives, and their impacts on cardiovascular health. From sleep duration to stress levels, we'll uncover valuable insights to enhance well-being.

## [1]: 1 Import Library

```
[2]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

C:\Users\Syed Arif\anaconda3\lib\site-packages\scipy\\_\_init\_\_.py:146:  
UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this version  
of SciPy (detected version 1.25.1  
warnings.warn(f"A NumPy version >={np\_minversion} and <{np\_maxversion}")

## 2 Uploading Csv file

```
[3]: df = pd.read_csv(r"C:\Users\Syed_
Arif\Desktop\Sleep_health_and_lifestyle_dataset.csv")
```

## 3 Data Preprocessing

### 4 .head()

head is used show to the By default = 5 rows in the dataset

```
[4]: df.head()
```

```
[4]:   Person ID Gender  Age      Occupation  Sleep Duration \
0         1   Male   27  Software Engineer           6.1
```

|   |   |      |    |       |                |     |
|---|---|------|----|-------|----------------|-----|
| 1 | 2 | Male | 28 |       | Doctor         | 6.2 |
| 2 | 3 | Male | 28 |       | Doctor         | 6.2 |
| 3 | 4 | Male | 28 | Sales | Representative | 5.9 |
| 4 | 5 | Male | 28 | Sales | Representative | 5.9 |

|   | Quality of Sleep | Physical Activity Level | Stress Level | BMI Category \ |
|---|------------------|-------------------------|--------------|----------------|
| 0 | 6                | 42                      | 6            | Overweight     |
| 1 | 6                | 60                      | 8            | Normal         |
| 2 | 6                | 60                      | 8            | Normal         |
| 3 | 4                | 30                      | 8            | Obese          |
| 4 | 4                | 30                      | 8            | Obese          |

|   | Blood Pressure | Heart Rate | Daily Steps | Sleep Disorder |
|---|----------------|------------|-------------|----------------|
| 0 | 126/83         | 77         | 4200        | None           |
| 1 | 125/80         | 75         | 10000       | None           |
| 2 | 125/80         | 75         | 10000       | None           |
| 3 | 140/90         | 85         | 3000        | Sleep Apnea    |
| 4 | 140/90         | 85         | 3000        | Sleep Apnea    |

## 6 .tail()

tail is used to show rows by Descending order

[5]: `df.tail()`

[5]:

| Person ID | Gender | Age    | Occupation | Sleep Duration | Quality of Sleep \ |
|-----------|--------|--------|------------|----------------|--------------------|
| 369       | 370    | Female | 59         | Nurse          | 8.1                |
| 370       | 371    | Female | 59         | Nurse          | 8.0                |
| 371       | 372    | Female | 59         | Nurse          | 8.1                |
| 372       | 373    | Female | 59         | Nurse          | 8.1                |
| 373       | 374    | Female | 59         | Nurse          | 8.1                |

|     | Physical Activity Level | Stress Level | BMI Category | Blood Pressure \ |
|-----|-------------------------|--------------|--------------|------------------|
| 369 | 75                      | 3            | Overweight   | 140/95           |
| 370 | 75                      | 3            | Overweight   | 140/95           |
| 371 | 75                      | 3            | Overweight   | 140/95           |
| 372 | 75                      | 3            | Overweight   | 140/95           |
| 373 | 75                      | 3            | Overweight   | 140/95           |

|     | Heart Rate | Daily Steps | Sleep Disorder |
|-----|------------|-------------|----------------|
| 369 | 68         | 7000        | Sleep Apnea    |
| 370 | 68         | 7000        | Sleep Apnea    |
| 371 | 68         | 7000        | Sleep Apnea    |
| 372 | 68         | 7000        | Sleep Apnea    |
| 373 | 68         | 7000        | Sleep Apnea    |

## 7 .shape

It show the total no of rows & Column in the dataset

```
[6]: df.shape
```

```
[6]: (374, 13)
```

## 8 .Columns

It show the no of each Column

```
[7]: df.columns
```

```
[7]: 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')
```

## 9 .dtypes

This Attribute show the data type of each column

```
[8]: df.dtypes
```

```
[8] : Person ID          int64  
    Gender              object  
    Age                int64  
    Occupation          object  
    Sleep Duration      float64  
    Quality of Sleep    int64  
    Physical Activity Level int64  
    Stress Level        int64  
    BMI Category        object  
    Blood Pressure      object  
    Heart Rate          int64  
    Daily Steps         int64  
    Sleep Disorder      object  
    dtype: object
```

## 10 .unique()

In a column, It show the unique value of specific column.

```
[9]: df["BMI Category"].unique()
```

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

## 11 .nunique()

It will show the total no of unique value from whole data frame

```
[10]: df.nunique()
```

```
[10] : Person ID          374
      Gender             2
      Age               31
      Occupation        11
      Sleep Duration     27
      Quality of Sleep    6
      Physical Activity Level 16
      Stress Level        6
      BMI Category       4
      Blood Pressure     25
      Heart Rate         19
      Daily Steps        20
      Sleep Disorder      3
      dtype: int64
```

## 12 .describe()

It show the Count, mean , median etc

```
[11]: df.describe()
```

```
[11]:
```

|       | 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 |

### 13 .value\_counts

It Shows all the unique values with their count

```
[12]: df["Occupation"].value_counts()
```

```
[12]: Nurse          73
      Doctor         71
      Engineer        63
      Lawyer          47
      Teacher         40
      Accountant       37
      Salesperson      32
      Software Engineer  4
      Scientist         4
      Sales Representative 2
      Manager           1
      Name: Occupation, dtype: int64
```

### 14 .isnull()

It shows the how many null values

```
[13]: df.isnull()
```

```
[13]:
```

|     | Person ID | Gender | Age   | Occupation | Sleep Duration | Quality of Sleep | \ |
|-----|-----------|--------|-------|------------|----------------|------------------|---|
| 0   | False     | False  | False | False      | False          | False            |   |
| 1   | False     | False  | False | False      | False          | False            |   |
| 2   | False     | False  | False | False      | False          | False            |   |
| 3   | False     | False  | False | False      | False          | False            |   |
| 4   | False     | False  | False | False      | False          | False            |   |
| ..  | ...       | ...    | ...   | ...        | ...            | ...              |   |
| 369 | False     | False  | False | False      | False          | False            |   |
| 370 | False     | False  | False | False      | False          | False            |   |
| 371 | False     | False  | False | False      | False          | False            |   |
| 372 | False     | False  | False | False      | False          | False            |   |
| 373 | False     | False  | False | False      | False          | False            |   |

|   | Physical Activity | Level | Stress | Level | BMI Category | Blood Pressure | \ |
|---|-------------------|-------|--------|-------|--------------|----------------|---|
| 0 |                   | False |        | False | False        | False          |   |
| 1 |                   | False |        | False | False        | False          |   |
| 2 |                   | False |        | False | False        | False          |   |
| 3 |                   | False |        | False | False        | False          |   |

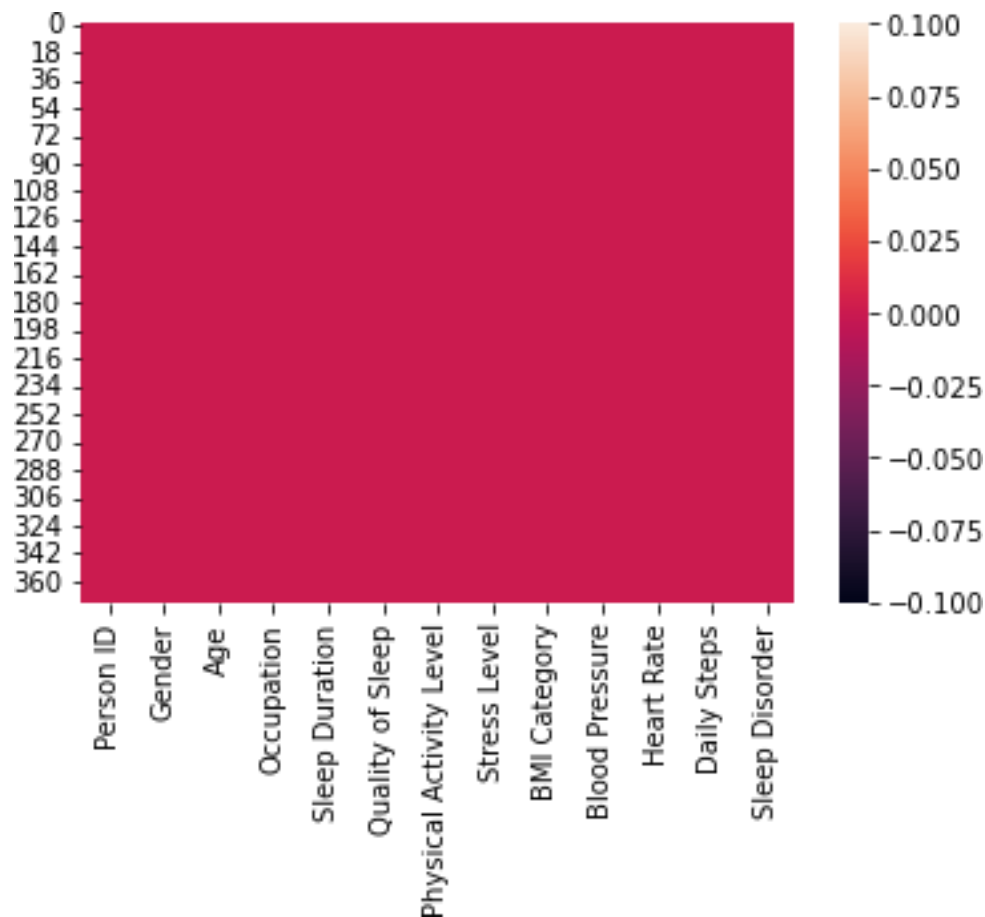
|     |       |       |       |       |
|-----|-------|-------|-------|-------|
| 4   | False | False | False | False |
| --  | ...   | ...   | ...   | ...   |
| 369 | False | False | False | False |
| 370 | False | False | False | False |
| 371 | False | False | False | False |
| 372 | False | False | False | False |
| 373 | False | False | False | False |

|     | Heart Rate | Daily Steps | Sleep Disorder |
|-----|------------|-------------|----------------|
| 0   | False      | False       | False          |
| 1   | False      | False       | False          |
| 2   | False      | False       | False          |
| 3   | False      | False       | False          |
| 4   | False      | False       | False          |
| --  | ...        | ...         | ...            |
| 369 | False      | False       | False          |
| 370 | False      | False       | False          |
| 371 | False      | False       | False          |
| 372 | False      | False       | False          |
| 373 | False      | False       | False          |

[374 rows x 13 columns]

```
[14]: sns.heatmap(df.isnull())
```

```
[14]: <AxesSubplot:>
```



```
[15]: df["Sleep Disorder"] = df["Sleep Disorder"].replace("None", "Nothing")
```

```
[16]: df
```

```
[16]:
```

|     | Person ID | Gender | Age | Occupation           | Sleep Duration \ |
|-----|-----------|--------|-----|----------------------|------------------|
| 0   | 1         | Male   | 27  | Software Engineer    | 6.1              |
| 1   | 2         | Male   | 28  | Doctor               | 6.2              |
| 2   | 3         | Male   | 28  | Doctor               | 6.2              |
| 3   | 4         | Male   | 28  | Sales Representative | 5.9              |
| 4   | 5         | Male   | 28  | Sales Representative | 5.9              |
| ..  | ...       | ...    | ... | ...                  | ...              |
| 369 | 370       | Female | 59  | Nurse                | 8.1              |
| 370 | 371       | Female | 59  | Nurse                | 8.0              |
| 371 | 372       | Female | 59  | Nurse                | 8.1              |
| 372 | 373       | Female | 59  | Nurse                | 8.1              |
| 373 | 374       | Female | 59  | Nurse                | 8.1              |

```

Quality of Sleep Physical Activity Level Stress Level BMI Category \

```

|     |     |     |     |            |
|-----|-----|-----|-----|------------|
| 0   | 6   | 42  | 6   | Overweight |
| 1   | 6   | 60  | 8   | Normal     |
| 2   | 6   | 60  | 8   | Normal     |
| 3   | 4   | 30  | 8   | Obese      |
| 4   | 4   | 30  | 8   | Obese      |
| --  | ... | ... | ... | ...        |
| 369 | 9   | 75  | 3   | Overweight |
| 370 | 9   | 75  | 3   | Overweight |
| 371 | 9   | 75  | 3   | Overweight |
| 372 | 9   | 75  | 3   | Overweight |
| 373 | 9   | 75  | 3   | Overweight |

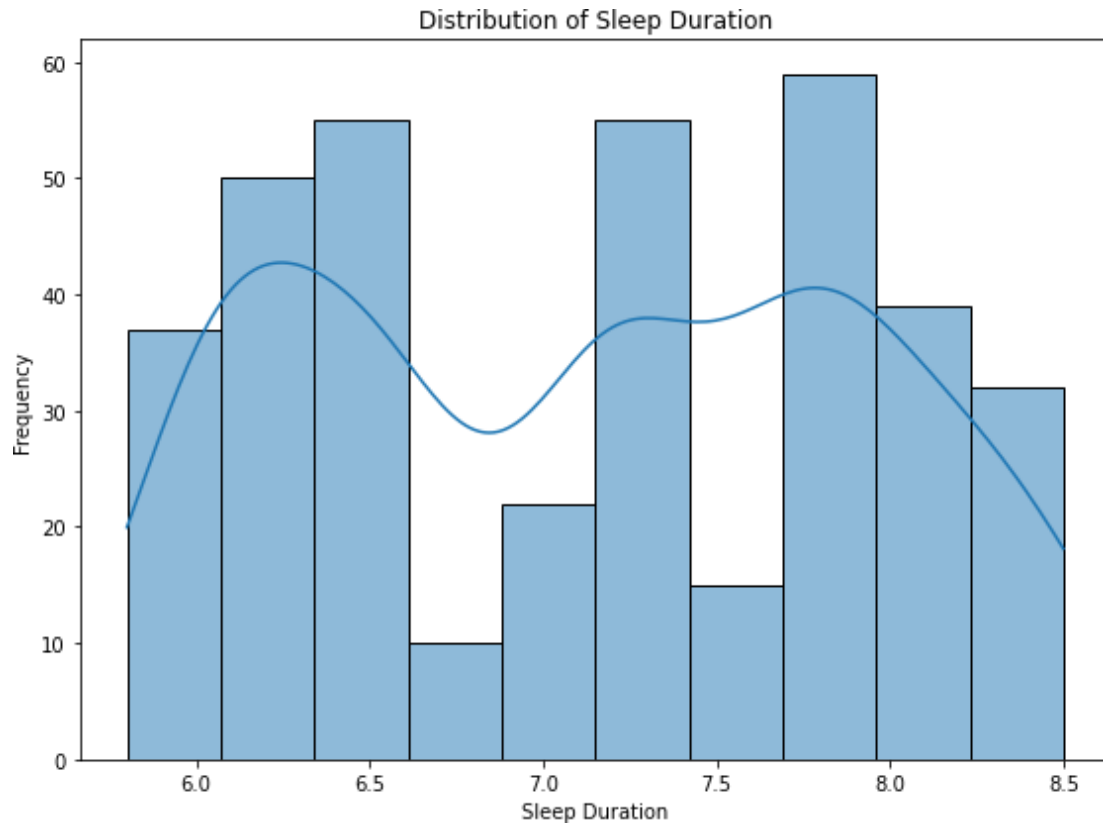
|     | Blood Pressure | Heart Rate | Daily Steps | Sleep Disorder |
|-----|----------------|------------|-------------|----------------|
| 0   | 126/83         | 77         | 4200        | Nothing        |
| 1   | 125/80         | 75         | 10000       | Nothing        |
| 2   | 125/80         | 75         | 10000       | Nothing        |
| 3   | 140/90         | 85         | 3000        | Sleep Apnea    |
| 4   | 140/90         | 85         | 3000        | Sleep Apnea    |
| --  | ...            | ...        | ...         | ...            |
| 369 | 140/95         | 68         | 7000        | Sleep Apnea    |
| 370 | 140/95         | 68         | 7000        | Sleep Apnea    |
| 371 | 140/95         | 68         | 7000        | Sleep Apnea    |
| 372 | 140/95         | 68         | 7000        | Sleep Apnea    |
| 373 | 140/95         | 68         | 7000        | Sleep Apnea    |

[374 rows x 13 columns]

## 15 1. What is the distribution of sleep duration among the individuals?

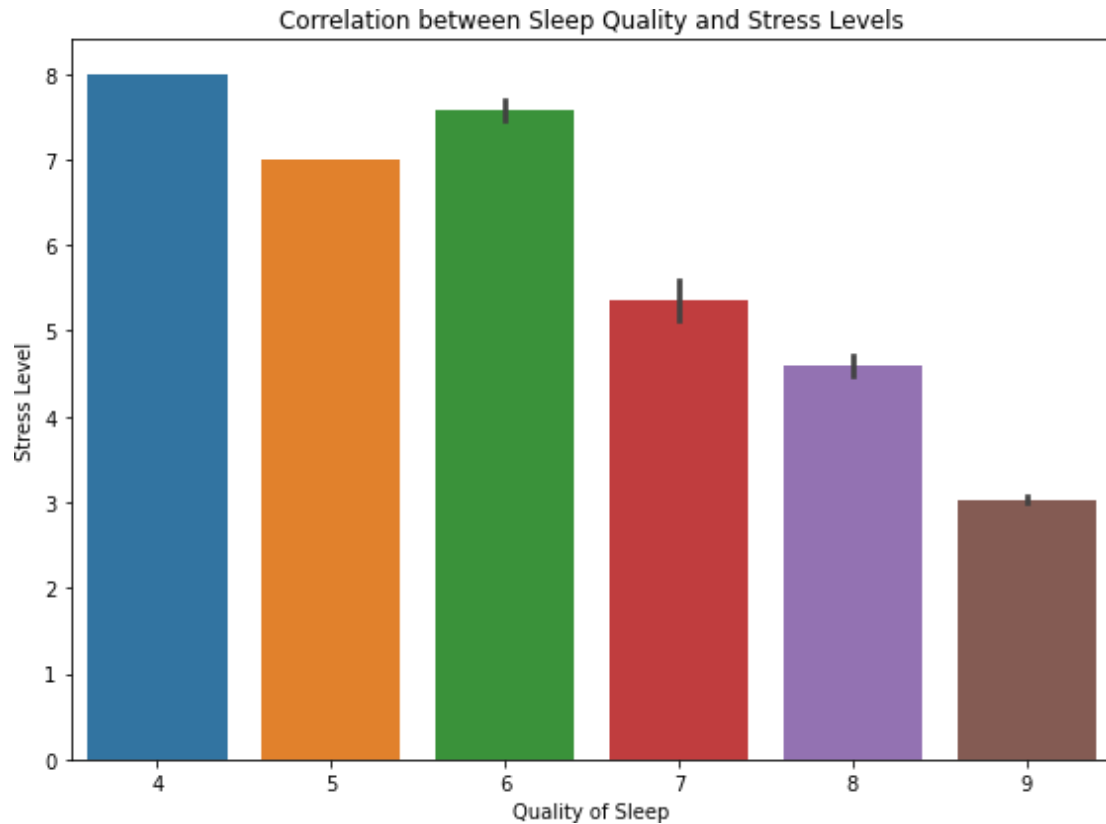
```
[17]: # Create a histogram of sleep duration
plt.figure(figsize=(8, 6))
sns.histplot(data=df, x="Sleep Duration", kde=True)
plt.title("Distribution of Sleep Duration")
plt.xlabel("Sleep Duration")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```





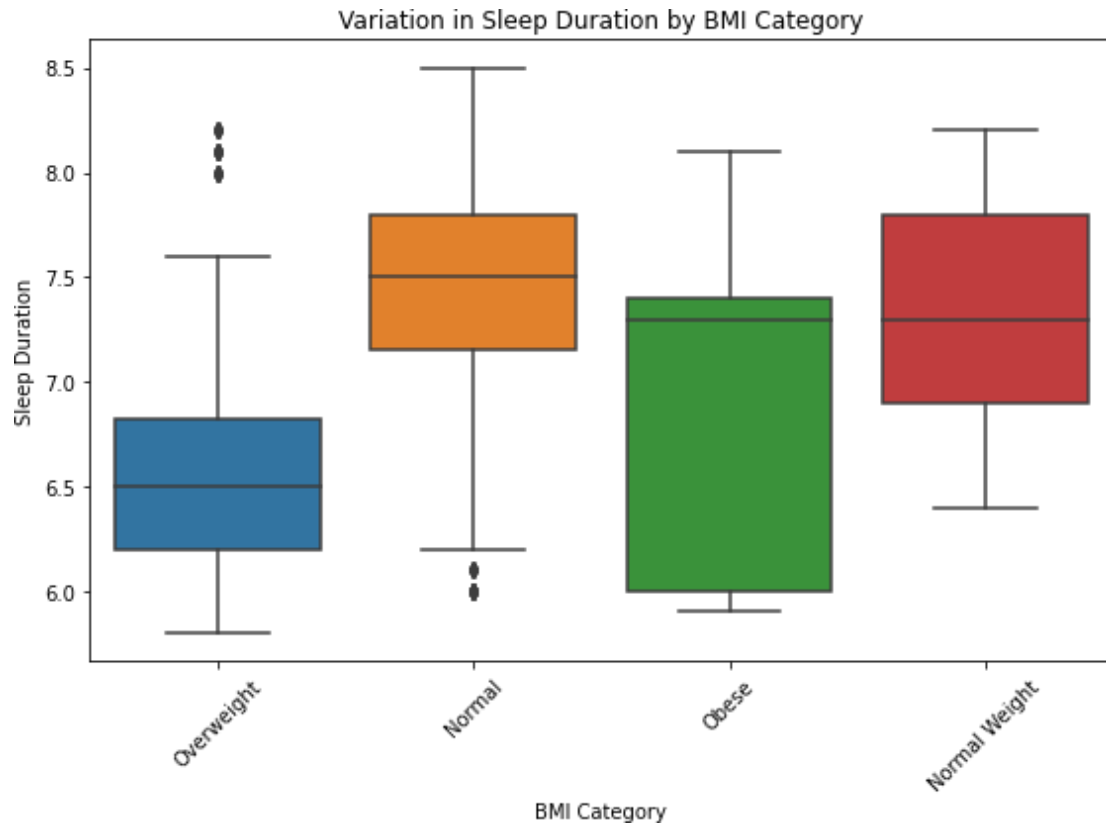
## 16 2. Is there a correlation between sleep quality and stress levels?

```
[18]: # Create a scatter plot to visualize the correlation
plt.figure(figsize=(8, 6))
sns.barplot(data=df, x="Quality of Sleep", y="Stress Level")
plt.title("Correlation between Sleep Quality and Stress Levels")
plt.xlabel("Quality of Sleep")
plt.ylabel("Stress Level")
plt.tight_layout()
plt.show()
```



### 17 3. How does sleep duration vary by BMI category?

```
[19]: # Create a box plot to show the distribution of sleep duration by BMI category
plt.figure(figsize=(8, 6))
sns.boxplot(data=df, x="BMI Category", y="Sleep Duration")
plt.title("Variation in Sleep Duration by BMI Category")
plt.xlabel("BMI Category")
plt.ylabel("Sleep Duration")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



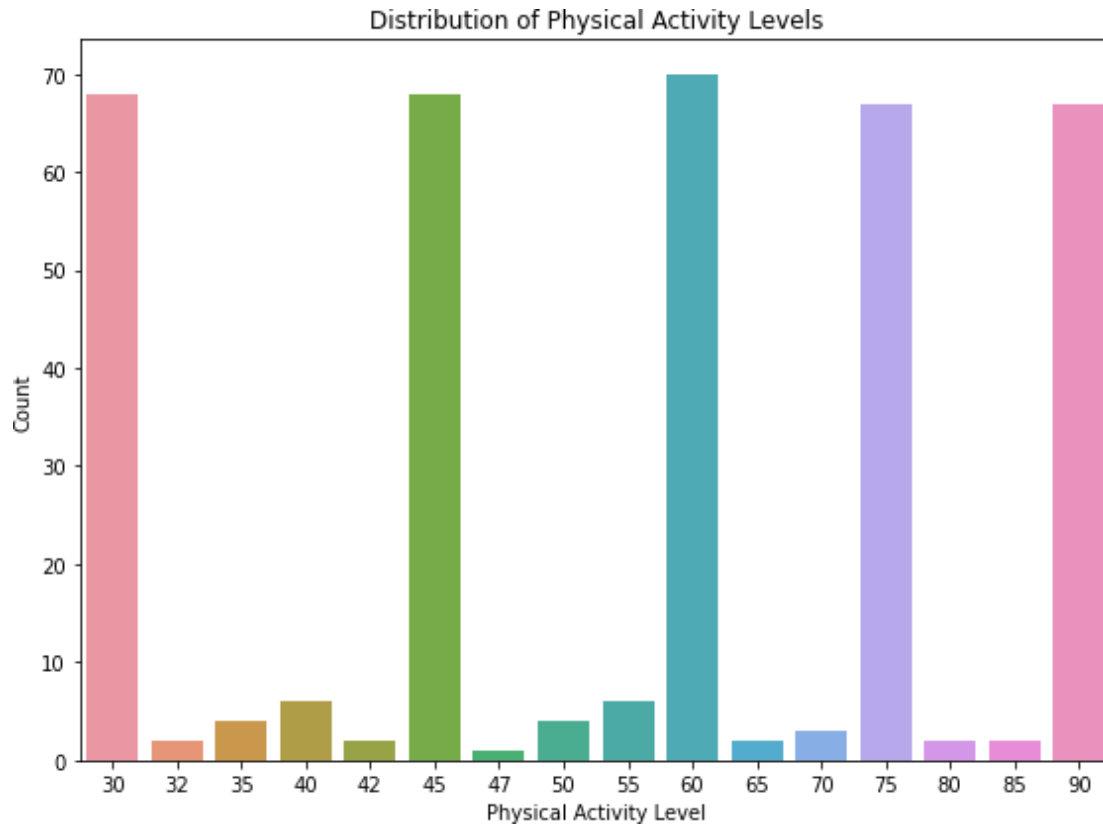
## 18 4. Are there any patterns in heart rate based on age?

```
[20]: # Create a line plot to visualize the relationship between age and heart rate
plt.figure(figsize=(8, 6))
sns.lineplot(data=df, x="Age", y="Heart Rate")
plt.title("Heart Rate Patterns by Age")
plt.xlabel("Age")
plt.ylabel("Heart Rate")
plt.tight_layout()
plt.show()
```



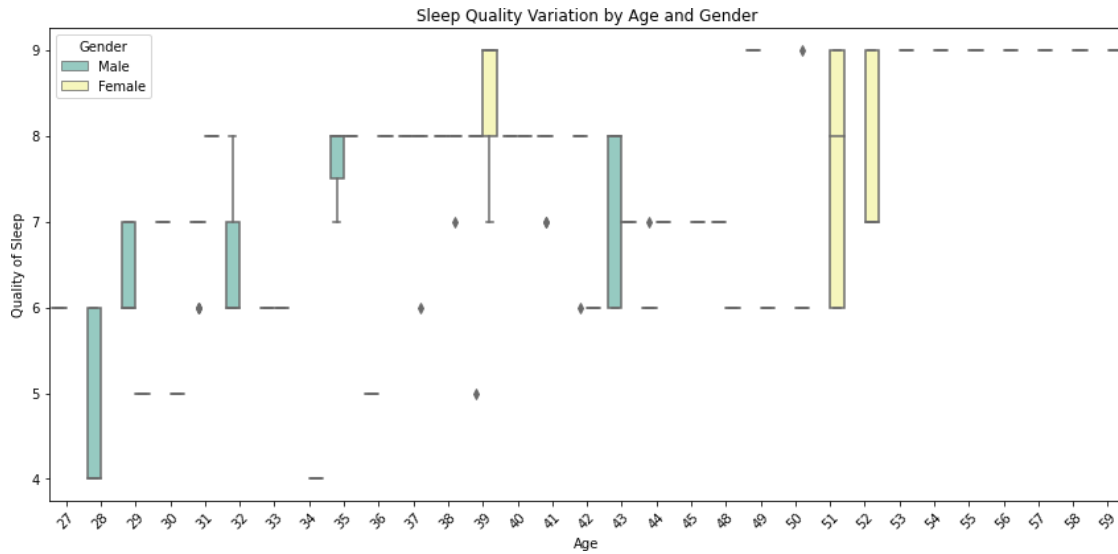
## 19 5. What is the distribution of physical activity levels?

```
[21]: # Create a bar plot to show the distribution of physical activity levels
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x="Physical Activity Level")
plt.title("Distribution of Physical Activity Levels")
plt.xlabel("Physical Activity Level")
plt.ylabel("Count")
plt.tight_layout()
plt.show()
```



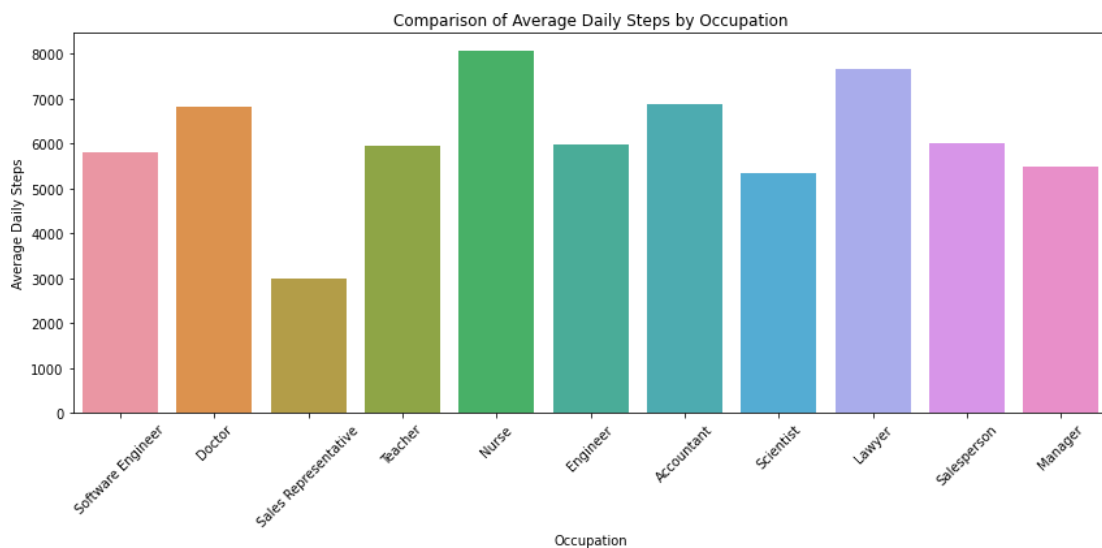
## 20 How does sleep quality vary with age and gender?

```
[22]: plt.figure(figsize=(12, 6))
sns.boxplot(x="Age", y="Quality of Sleep", hue="Gender", data=df,
            palette="Set3")
plt.title("Sleep Quality Variation by Age and Gender")
plt.xlabel("Age")
plt.ylabel("Quality of Sleep")
plt.xticks(rotation=45)
plt.legend(title="Gender")
plt.tight_layout()
plt.show()
```



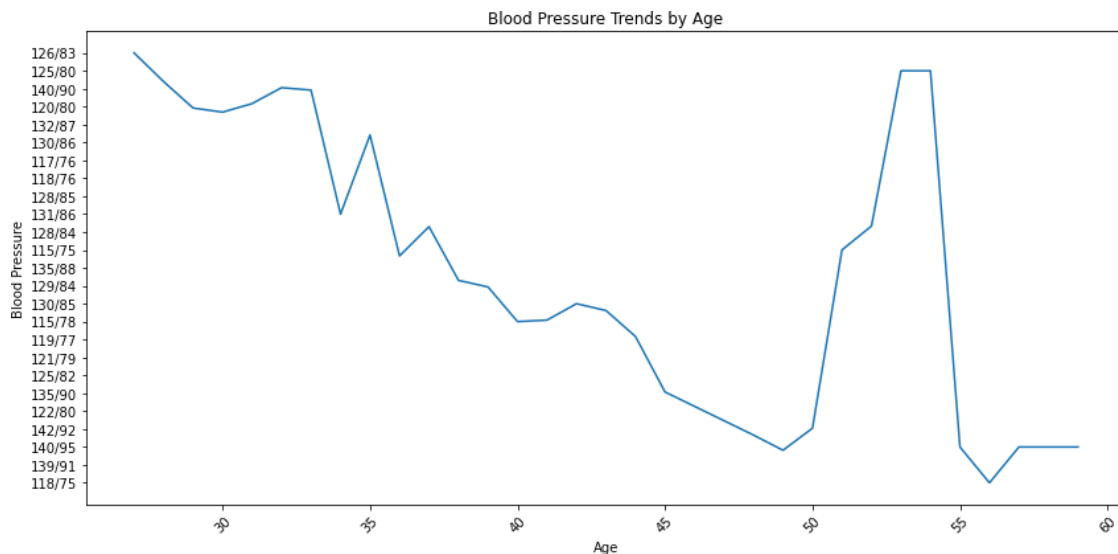
## 21 How does occupation affect daily steps taken?

```
[23]: plt.figure(figsize=(12, 6))
sns.barplot(x="Occupation", y="Daily Steps", data=df, ci=None)
plt.title("Comparison of Average Daily Steps by Occupation")
plt.xlabel("Occupation")
plt.ylabel("Average Daily Steps")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



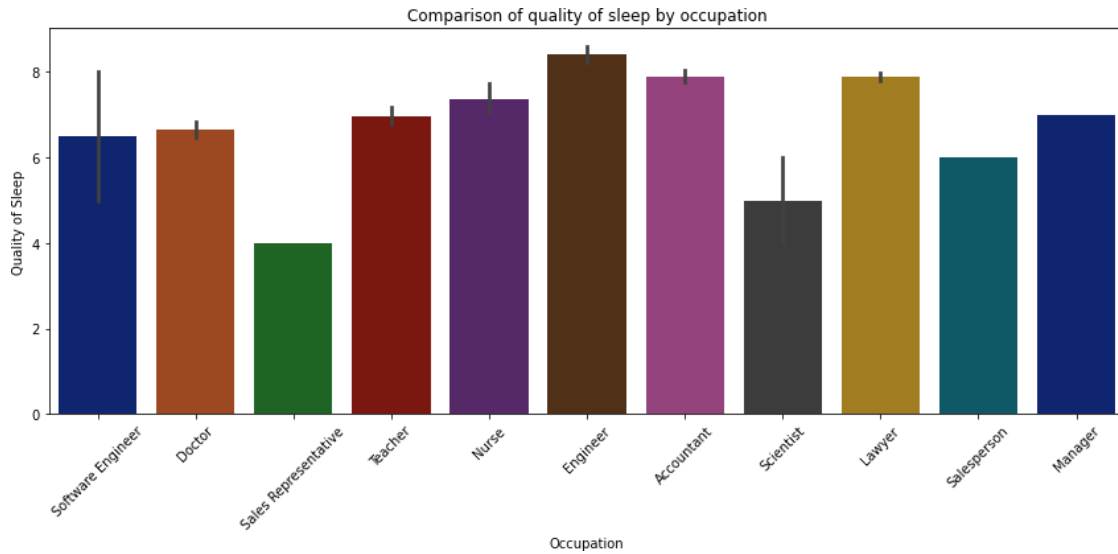
## 22 Are there any patterns in blood pressure across different age groups?

```
[24]: plt.figure(figsize=(12, 6))
sns.lineplot(x="Age", y="Blood Pressure", data=df, ci=None)
plt.title("Blood Pressure Trends by Age")
plt.xlabel("Age")
plt.ylabel("Blood Pressure")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



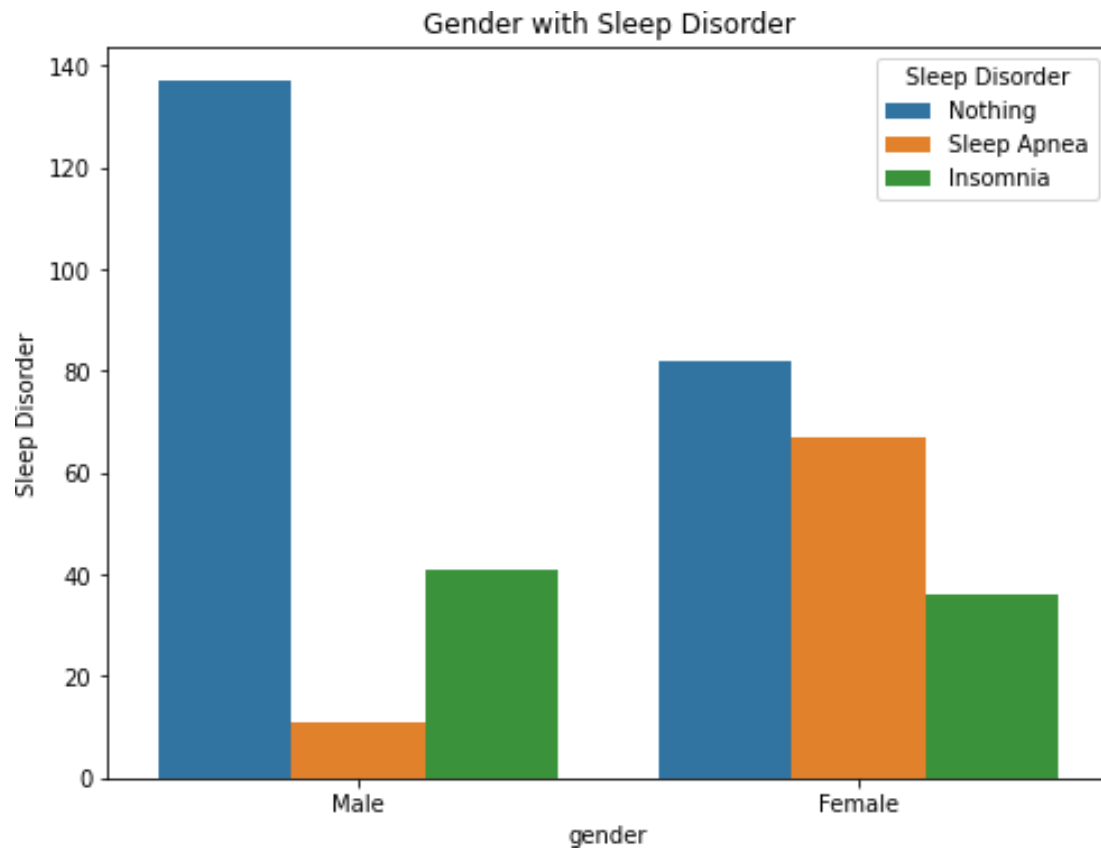
## 23 Do people in specific occupations tend to have better or worse quality of sleep?

```
[25]: plt.figure(figsize=(12, 6))
sns.barplot(data=df, x="Occupation", y="Quality of Sleep", palette="dark")
plt.xticks(rotation=45)
plt.xlabel("Occupation")
plt.ylabel("Quality of Sleep")
plt.title("Comparison of quality of sleep by occupation")
plt.tight_layout()
plt.show()
```

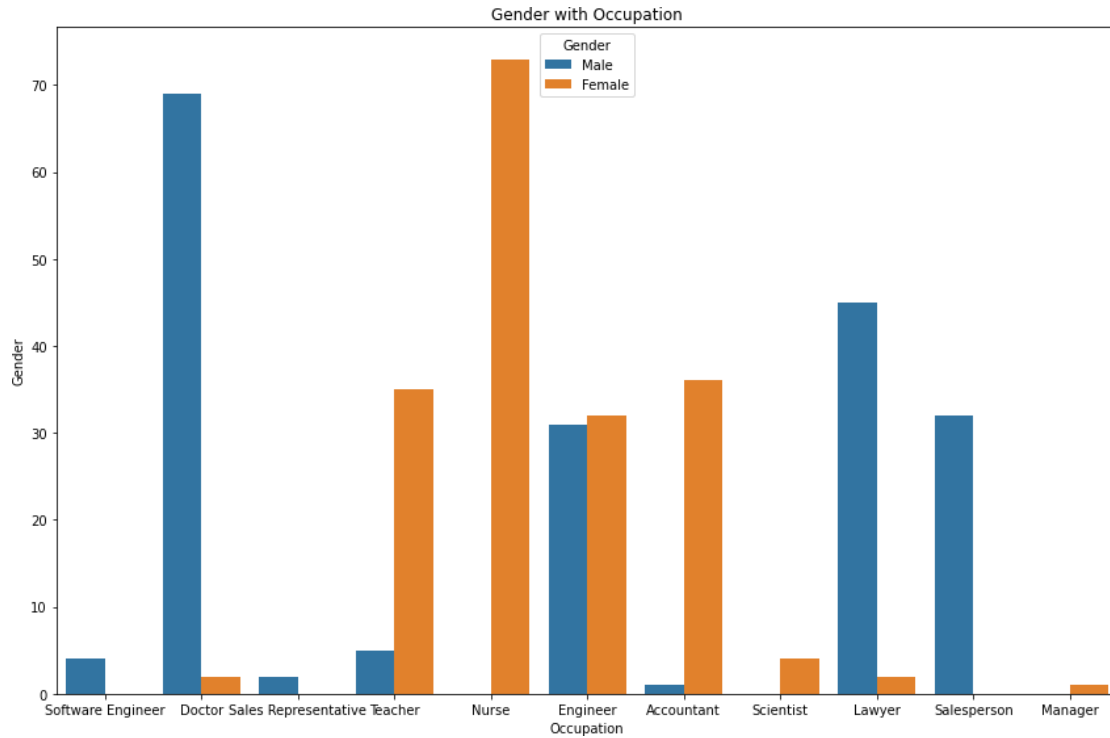


```
[26]: plt.figure(figsize=(8, 6))
sns.countplot(data=df, x="Gender", hue="Sleep Disorder")
plt.xlabel("gender")
plt.ylabel("Sleep Disorder")
plt.title("Gender with Sleep Disorder")
plt.show()
```

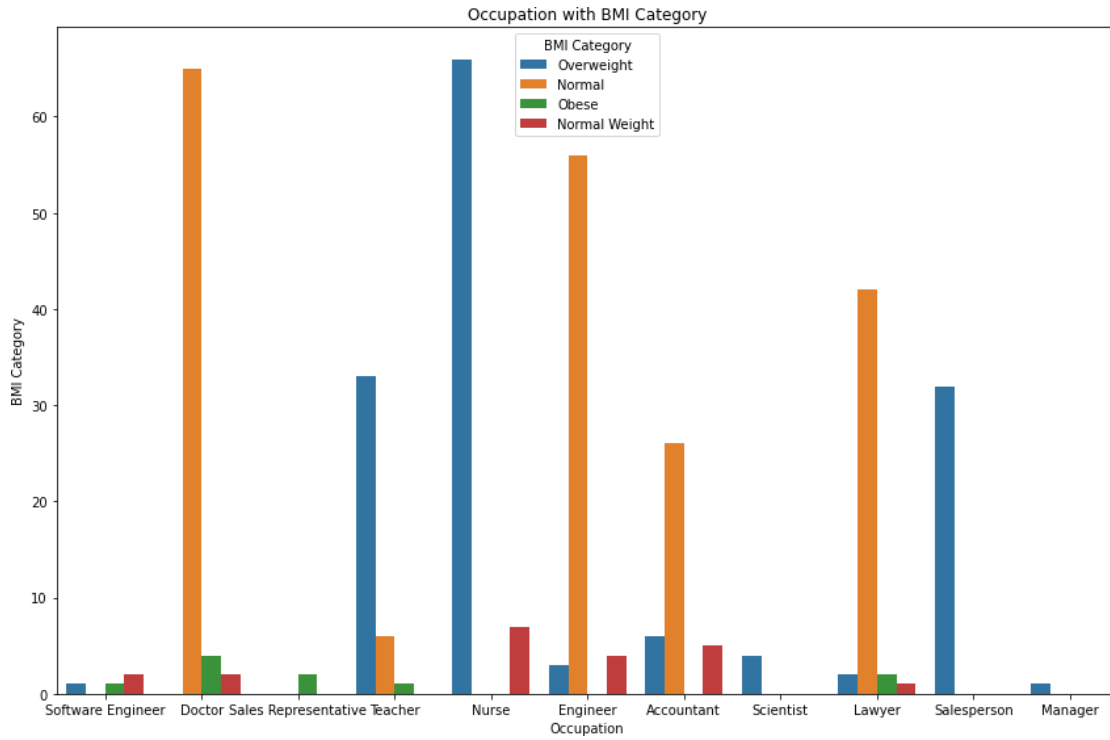




```
[35]: plt.figure(figsize=(12, 8))
sns.countplot(data=df, x='Occupation', hue='Gender')
plt.xlabel('Occupation')
plt.ylabel('Gender')
plt.title('Gender with Occupation')
plt.tight_layout()
plt.show()
```



```
[36]: plt.figure(figsize=(12, 8))
sns.countplot(data=df, x='Occupation', hue='BMI Category')
plt.xlabel('Occupation')
plt.ylabel('BMI Category')
plt.title('Occupation with BMI Category')
plt.tight_layout()
plt.show()
```



[ ]: