```
#importing libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#loading the dataset
df = pd.read_csv('Sleep_health_and_lifestyle_dataset.csv')
df.head()

   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            NaN
1         125/80          75        10000            NaN
2         125/80          75        10000            NaN
3         140/90          85         3000    Sleep Apnea
4         140/90          85         3000    Sleep Apnea

#checking for missing values
df.isnull().sum()

Person ID                  0
Gender                     0
Age                        0
Occupation                 0
Sleep Duration             0
Quality of Sleep           0
Physical Activity Level    0
Stress Level               0
BMI Category               0
Blood Pressure             0
```

```
Heart Rate                   0
Daily Steps                  0
Sleep Disorder             219
dtype: int64
```

```python
#replacing the null values with 'None' in the column 'Sleep Disorder'
df['Sleep Disorder'].fillna('None', inplace=True)
```

```python
#drop column Person ID
df.drop('Person ID', axis=1, inplace=True)
```

```python
# print the number of unique values for each column to get a quick sense of data variety
print("Unique values in each column:")
for col in df.columns:
    unique_count = df[col].nunique()  # count unique entries in the column
    print(f"{col}: {unique_count}")   # show column name and its unique count
```

```
Unique values in each column:
Gender: 2
Age: 31
Occupation: 11
Sleep Duration: 27
Quality of Sleep: 6
Physical Activity Level: 16
Stress Level: 6
BMI Category: 4
Heart Rate: 19
Daily Steps: 20
Sleep Disorder: 3
systolic_bp: 18
diastolic_bp: 17
```

```python
# clean any sneaky spaces in column names
df.columns = df.columns.str.strip()
```

```python
# split bp if it’s there, else warn me
if 'Blood Pressure' in df.columns:
    df[['systolic_bp', 'diastolic_bp']] = df['Blood Pressure'].str.split('/', expand=True)
    df.drop('Blood Pressure', axis=1, inplace=True)
else:
    print("no 'Blood Pressure' column found")
```

```
no 'Blood Pressure' column found
```

```python
df['BMI Category'] = df['BMI Category'].replace('Normal Weight', 'Normal')
```

```
df.head()
```

```
   Gender  Age  Occupation  Sleep Duration  Quality of Sleep  \
0       1   27           9             6.1                 6   
1       1   28           1             6.2                 6   
2       1   28           1             6.2                 6   
3       1   28           6             5.9                 4   
4       1   28           6             5.9                 4   

   Physical Activity Level  Stress Level  BMI Category Blood Pressure 
\
0                       42             6             3         126/83   

1                       60             8             0         125/80   

2                       60             8             0         125/80   

3                       30             8             2         140/90   

4                       30             8             2         140/90   


   Heart Rate  Daily Steps  Sleep Disorder
0          77         4200               1
1          75        10000               1
2          75        10000               1
3          85         3000               2
4          85         3000               2
```

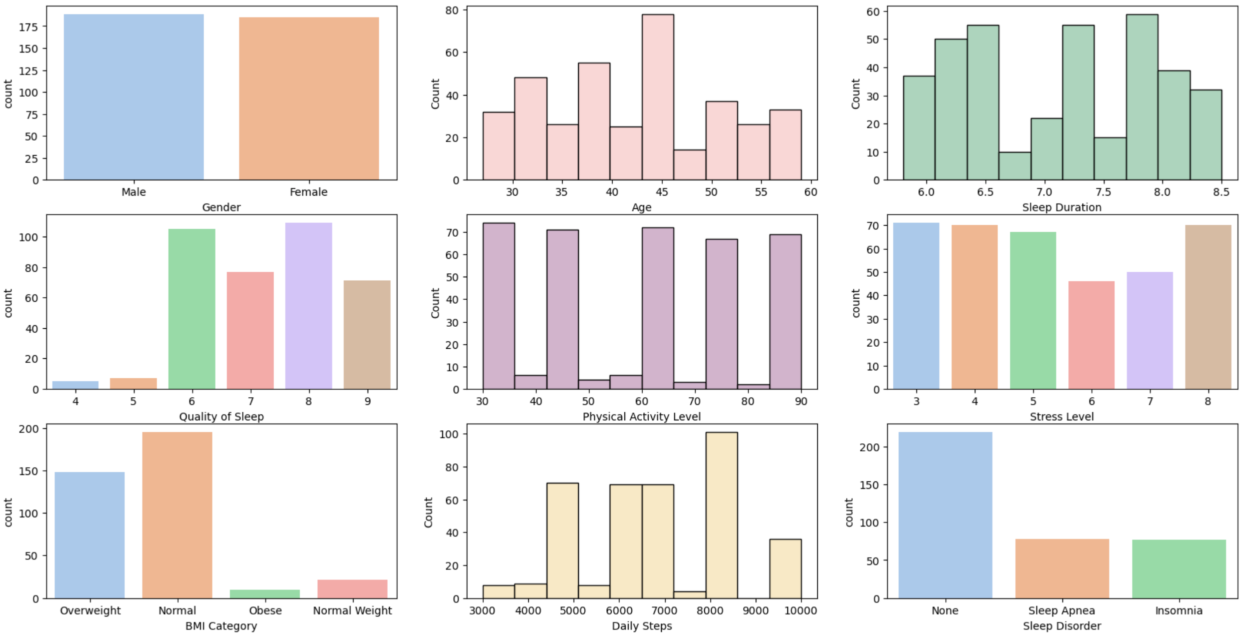
```
#unique values from categorical columns
print(df.Occupation.unique())
print('\n')
print(df['BMI Category'].unique())
print('\n')
print(df['Sleep Disorder'].unique())
```

```
['Software Engineer' 'Doctor' 'Sales Representative' 'Teacher' 'Nurse'
 'Engineer' 'Accountant' 'Scientist' 'Lawyer' 'Salesperson' 'Manager']


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


['None' 'Sleep Apnea' 'Insomnia']
```

```
fig, ax = plt.subplots(3,3, figsize=(20,10))

sns.countplot(x='Gender', data=df, ax=ax[0,0], hue='Gender', 
palette='pastel', legend=False)
sns.histplot(x='Age', data=df, ax=ax[0,1], bins=10, color='#f7cac9') 
# soft pink coral-ish
```

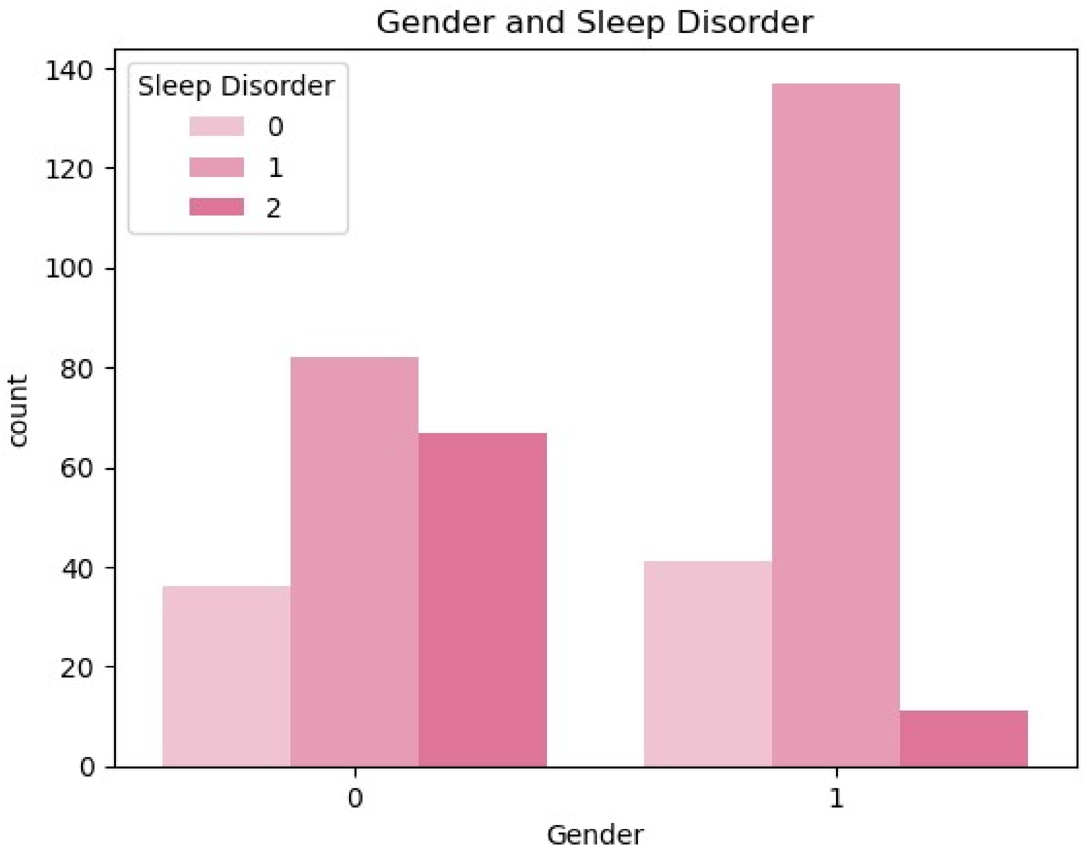
```
sns.histplot(x='Sleep Duration', data=df, ax=ax[0,2], bins=10,
color='#92c6a7')  # soft green

sns.countplot(x='Quality of Sleep', data=df, ax=ax[1,0], hue='Quality
of Sleep', palette='pastel', legend=False)
sns.histplot(x='Physical Activity Level', data=df, ax=ax[1,1],
bins=10, color='#c49bbb')  # muted orchid-ish
sns.countplot(x='Stress Level', data=df, ax=ax[1,2], hue='Stress
Level', palette='pastel', legend=False)

sns.countplot(x='BMI Category', data=df, ax=ax[2,0], hue='BMI
Category', palette='pastel', legend=False)
sns.histplot(x='Daily Steps', data=df, ax=ax[2,1], bins=10,
color='#f6e2b3')  # light gold
sns.countplot(x='Sleep Disorder', data=df, ax=ax[2,2], hue='Sleep
Disorder', palette='pastel', legend=False)
```

```
<Axes: xlabel='Sleep Disorder', ylabel='count'>
```



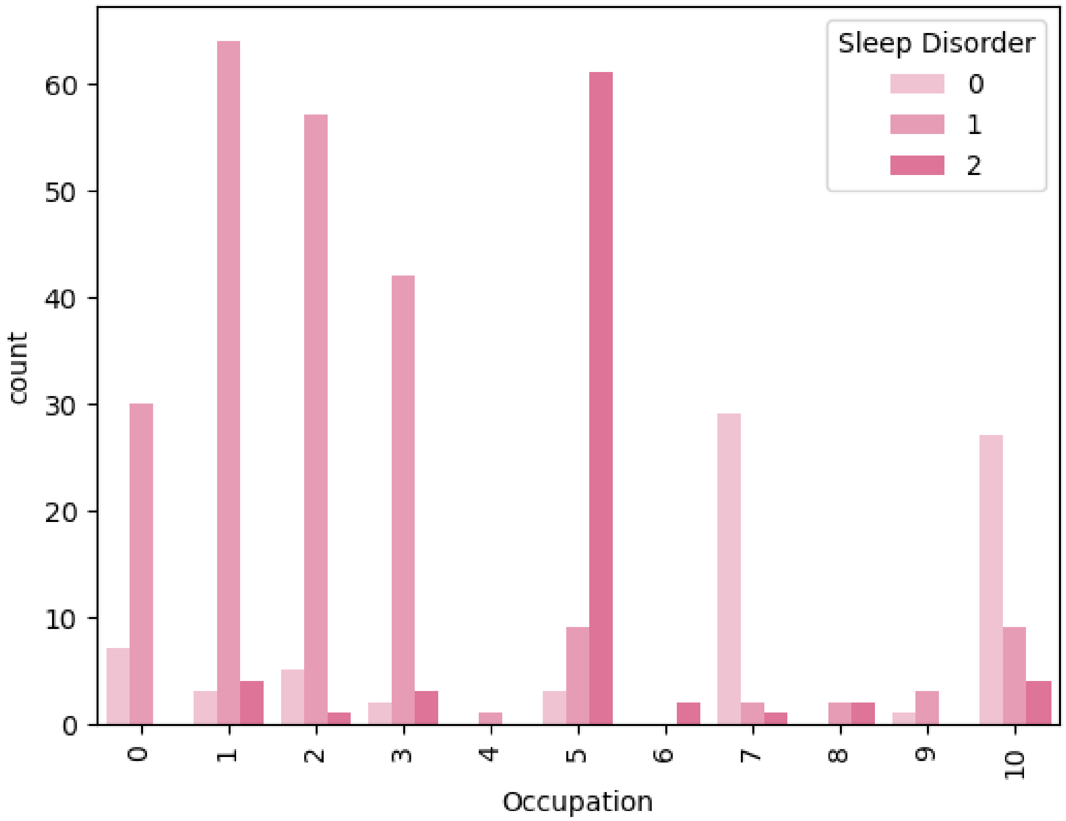
```
pink_pastel = ['#f8bbd0', '#f48fb1', '#f06292']  # 3 pastel pink
shades

sns.countplot(x='Gender', data=df, hue='Sleep Disorder',
palette=pink_pastel).set_title('Gender and Sleep Disorder')
```

```
Text(0.5, 1.0, 'Gender and Sleep Disorder')
```



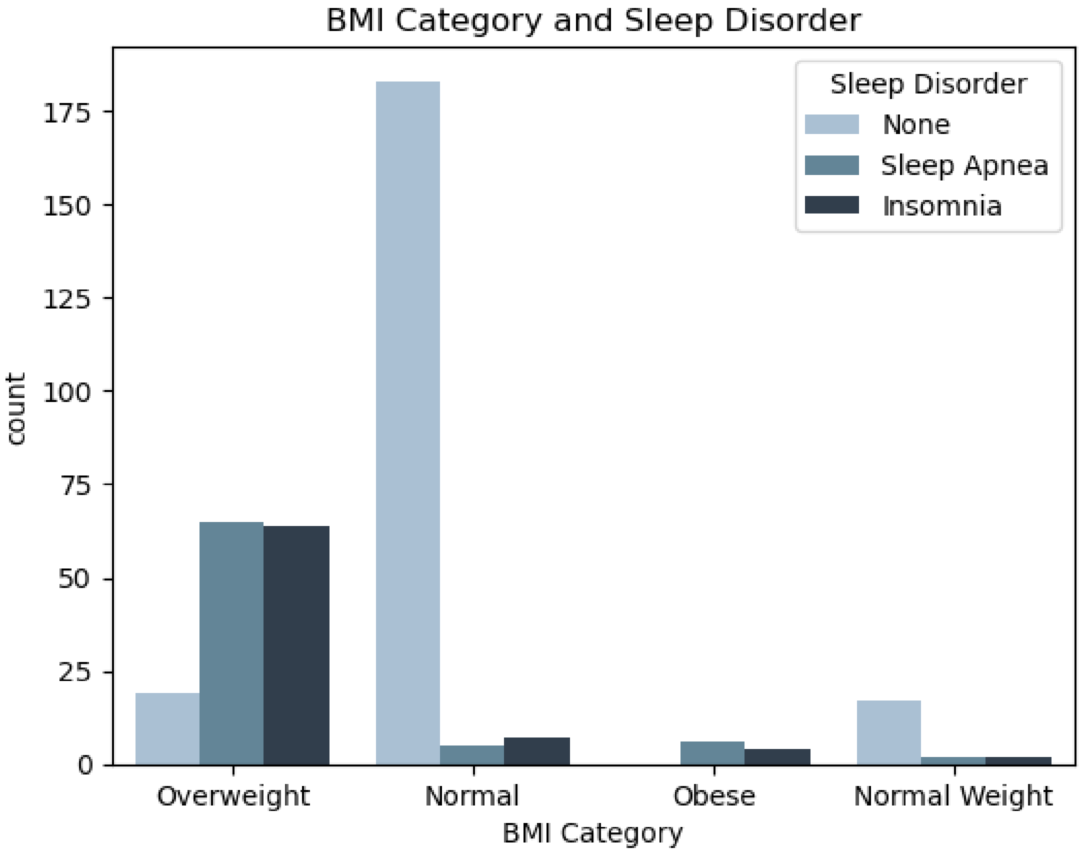
```
pink_pastel = ['#f8bbd0', '#f48fb1', '#f06292']  # 3 soft pink shades
ax = sns.countplot(x='Occupation', data=df, hue='Sleep Disorder',
palette=pink_pastel)
plt.xticks(rotation=90)
plt.legend(title='Sleep Disorder')
```

```
<matplotlib.legend.Legend at 0x16a268200>
```



```
blue_shades = ['#a3c1da', '#5a87a0', '#2c3e50']  # light to dark blue
shades

sns.countplot(x='BMI Category', hue='Sleep Disorder', data=df,
palette=blue_shades).set_title('BMI Category and Sleep Disorder')
```

```
Text(0.5, 1.0, 'BMI Category and Sleep Disorder')
```

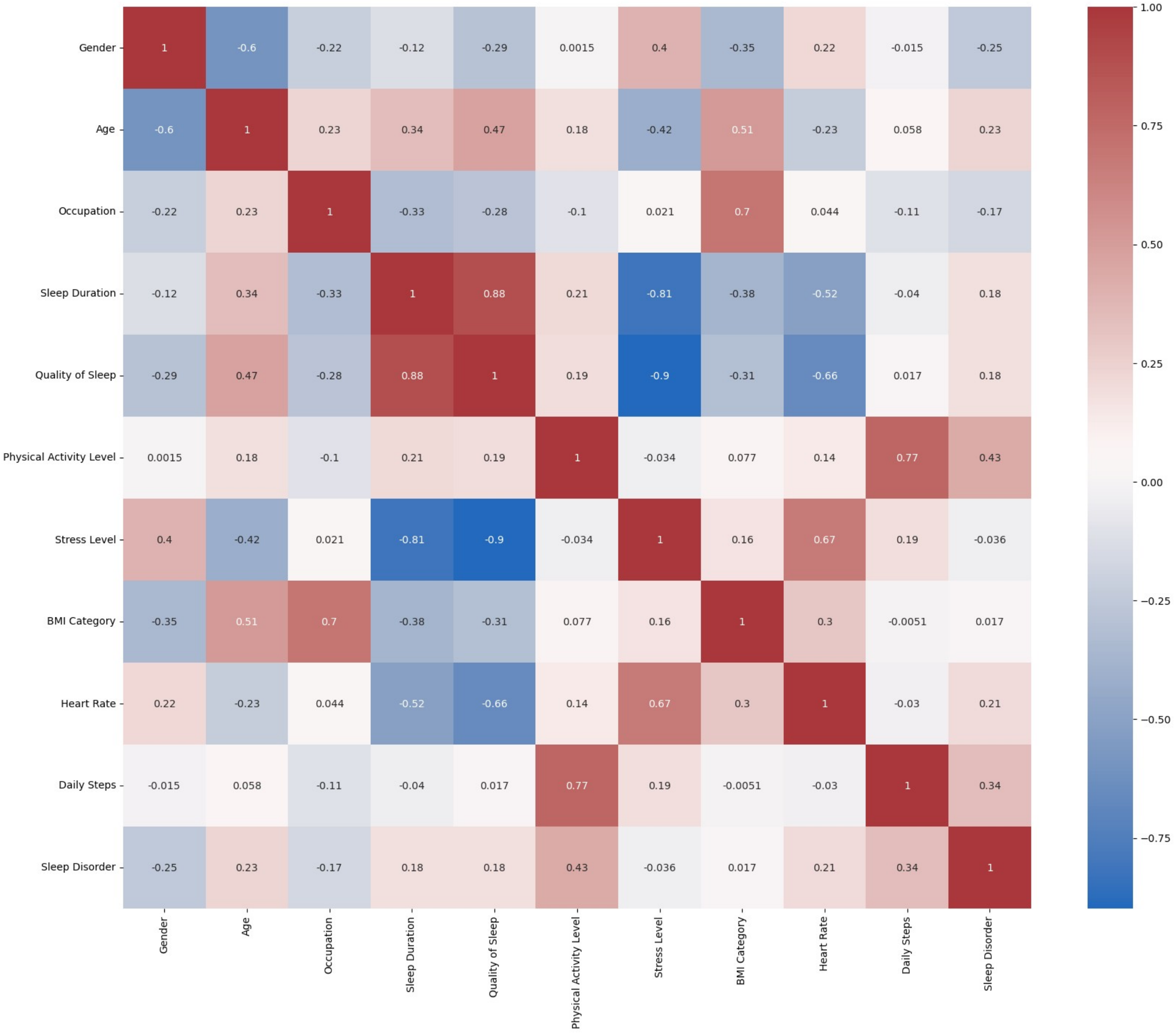


```python
from sklearn.preprocessing import LabelEncoder

vars = ['Gender', 'Occupation', 'BMI Category', 'Sleep Disorder']
for i in vars:
    label_encoder = LabelEncoder()  # make a fresh encoder each loop
    df[i] = label_encoder.fit_transform(df[i])
    print(i, ':', df[i].unique())
```

```
Gender : [1 0]
Occupation : [ 9  1  6 10  5  2  0  8  3  7  4]
BMI Category : [3 0 2 1]
Sleep Disorder : [1 2 0]
```

```python
plt.figure(figsize=(20,16))
sns.heatmap(df.select_dtypes(include='number').corr(), annot=True,
cmap='vlag')
```

```
<Axes: >
```



	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Heart Rate	Daily Steps	Sleep Disorder
Gender	1	-0.6	-0.22	-0.12	-0.29	0.0015	0.4	-0.35	0.22	-0.015	-0.25
Age	-0.6	1	0.23	0.34	0.47	0.18	-0.42	0.51	-0.23	0.058	0.23
Occupation	-0.22	0.23	1	-0.33	-0.28	-0.1	0.021	0.7	0.044	-0.11	-0.17
Sleep Duration	-0.12	0.34	-0.33	1	0.88	0.21	-0.81	-0.38	-0.52	-0.04	0.18
Quality of Sleep	-0.29	0.47	-0.28	0.88	1	0.19	-0.9	-0.31	-0.66	0.017	0.18
Physical Activity Level	0.0015	0.18	-0.1	0.21	0.19	1	-0.034	0.077	0.14	0.77	0.43
Stress Level	0.4	-0.42	0.021	-0.81	-0.9	-0.034	1	0.16	0.67	0.19	-0.036
BMI Category	-0.35	0.51	0.7	-0.38	-0.31	0.077	0.16	1	0.3	-0.0051	0.017
Heart Rate	0.22	-0.23	0.044	-0.52	-0.66	0.14	0.67	0.3	1	-0.03	0.21
Daily Steps	-0.015	0.058	-0.11	-0.04	0.017	0.77	0.19	-0.0051	-0.03	1	0.34
Sleep Disorder	-0.25	0.23	-0.17	0.18	0.18	0.43	-0.036	0.017	0.21	0.34	1