In [95]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [96]:
```
df = pd.read_csv('Sleep_health_and_lifestyle_dataset.csv')
df
```

Out[96]:

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	P
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	
1	2	Male	28	Doctor	6.2	6	60	8	Normal	
2	3	Male	28	Doctor	6.2	6	60	8	Normal	
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	
...	...	...	...	...	...	...	...	...	...	
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	

374 rows × 13 columns

In [97]:
```
df.drop_duplicates()
```

Out[97]:

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Pı
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	
1	2	Male	28	Doctor	6.2	6	60	8	Normal	
2	3	Male	28	Doctor	6.2	6	60	8	Normal	
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	
...	...	...	...	...	...	...	...	...	...	
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	

374 rows × 13 columns

In [98]:
```
df=df.dropna()
df
```

Out[98]:

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Pı
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	
5	6	Male	28	Software Engineer	5.9	4	30	8	Obese	
6	7	Male	29	Teacher	6.3	6	40	7	Obese	
16	17	Female	29	Nurse	6.5	5	40	7	Normal Weight	
...	...	...	...	...	...	...	...	...	...	
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	

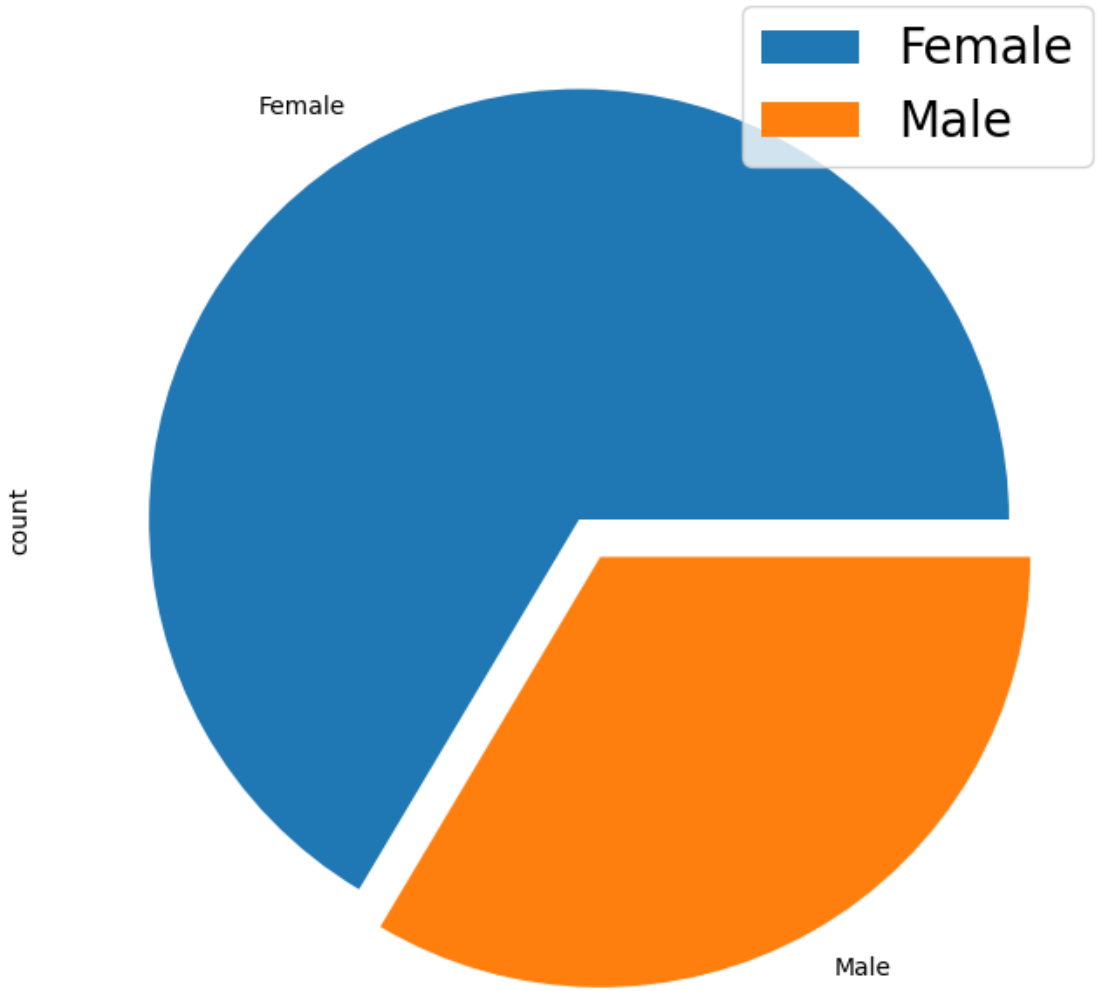
155 rows × 13 columns

In [99]:
```
x=df['Gender'].value_counts()
x
```

Out[99]:
```
Gender
Female    103
Male       52
Name: count, dtype: int64
```

In [100]:
```
x.plot(kind='pie',explode=(0,0.1) ,figsize=(8,8))
plt.legend( fontsize = 20)
```

Out[100]: <matplotlib.legend.Legend at 0x190528c4ed0>

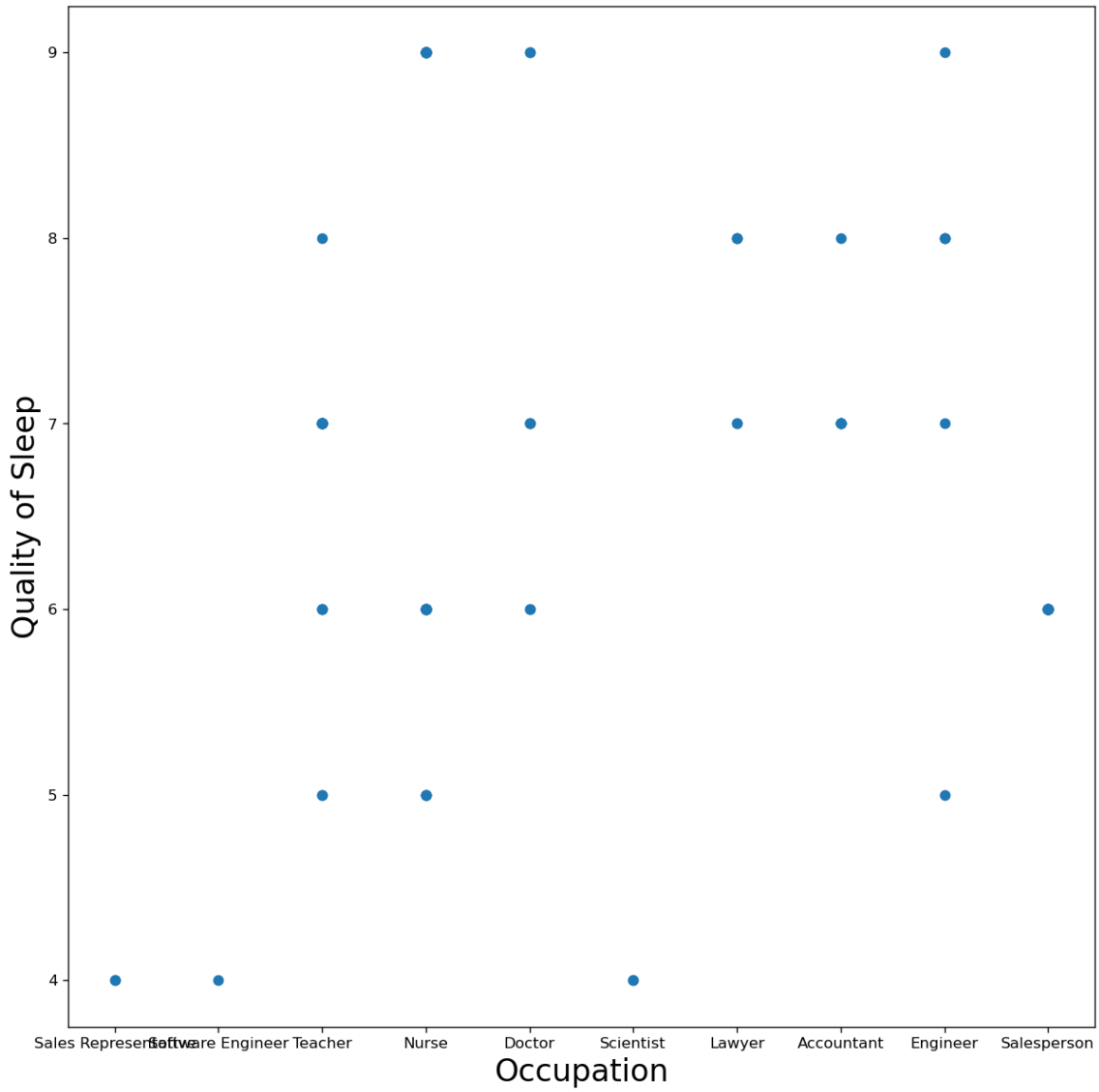


In [22]:
```
df.describe()
```

Out[22]:

	Person ID	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate
count	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000
mean	244.761290	46.632258	6.812258	6.870968	60.896774	5.767742	71.787097
std	91.282997	7.839311	0.773534	1.337325	20.634158	1.946757	5.187381
min	4.000000	28.000000	5.800000	4.000000	30.000000	3.000000	65.000000
25%	199.500000	43.000000	6.300000	6.000000	45.000000	4.000000	68.000000
50%	255.000000	45.000000	6.500000	7.000000	45.000000	7.000000	72.000000
75%	304.500000	51.000000	7.400000	8.000000	75.000000	7.000000	75.000000
max	374.000000	59.000000	8.300000	9.000000	90.000000	8.000000	86.000000

In [70]:
```
plt.figure(figsize=(12, 12), dpi=120)
plt.xlabel('Occupation',fontsize=20)
plt.ylabel('Quality of Sleep',fontsize=20)
x = df['Occupation']
y = df['Quality of Sleep']
plt.scatter(x,y)
plt.show()
```



In [58]:
```
data = df.select_dtypes(exclude='object')
```

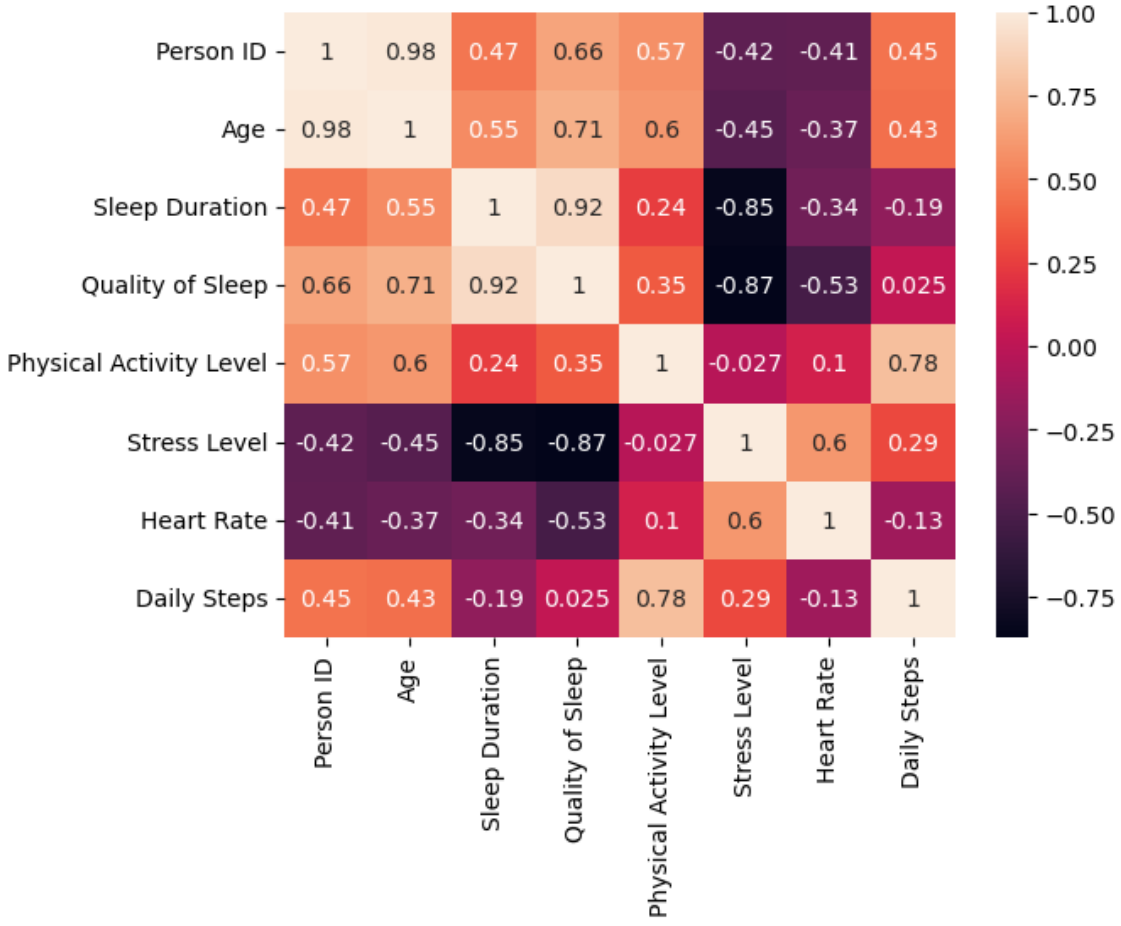
In [60]:
```
correlation=data.corr()
correlation
```

Out[60]:

	Person ID	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate	Daily Steps
Person ID	1.000000	0.981739	0.465877	0.661836	0.573090	-0.417244	-0.409011	0.44778
Age	0.981739	1.000000	0.550193	0.710221	0.604725	-0.452398	-0.373515	0.43427
Sleep Duration	0.465877	0.550193	1.000000	0.916748	0.240760	-0.848873	-0.337888	-0.19424
Quality of Sleep	0.661836	0.710221	0.916748	1.000000	0.352021	-0.872084	-0.534720	0.02516
Physical Activity Level	0.573090	0.604725	0.240760	0.352021	1.000000	-0.026627	0.101348	0.78068
Stress Level	-0.417244	-0.452398	-0.848873	-0.872084	-0.026627	1.000000	0.598860	0.28508
Heart Rate	-0.409011	-0.373515	-0.337888	-0.534720	0.101348	0.598860	1.000000	-0.13346
Daily Steps	0.447782	0.434277	-0.194249	0.025166	0.780685	0.285082	-0.133465	1.00000

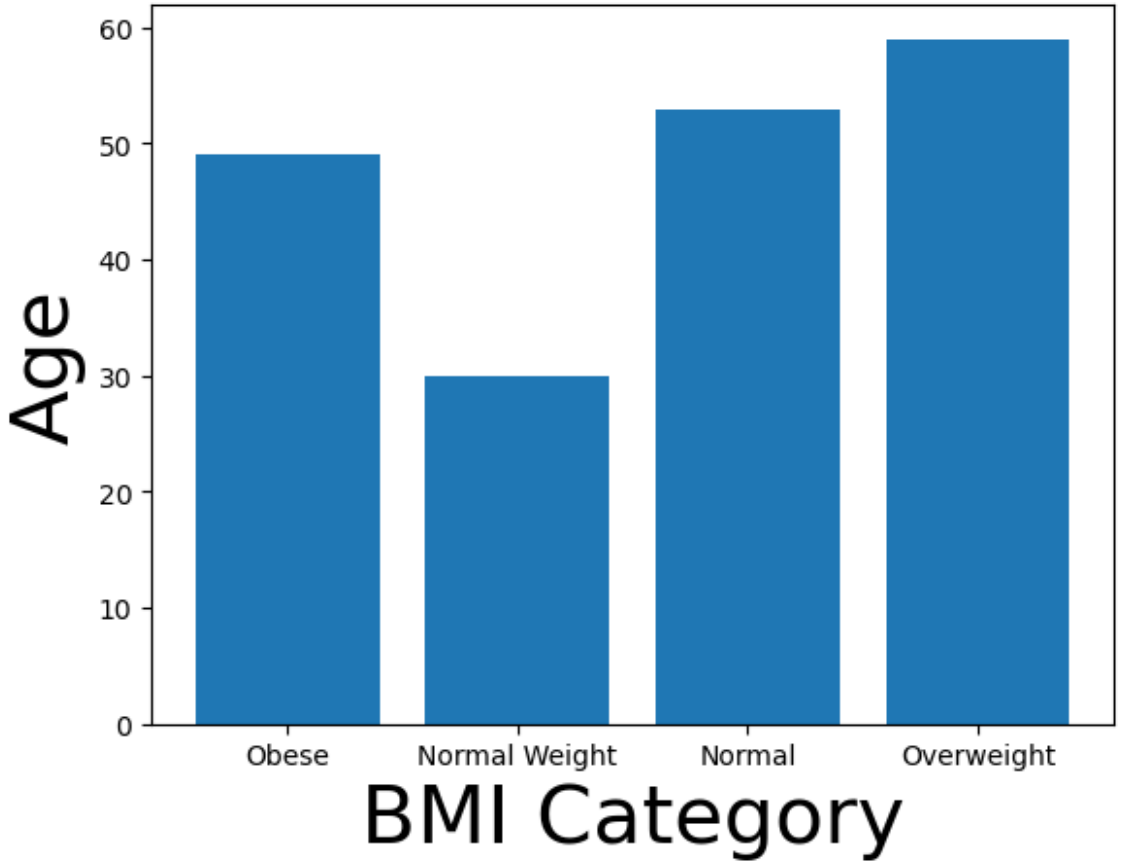
In [62]:
```
sns.heatmap(correlation,annot =True)
```

Out[62]: <Axes: >



In [90]:
```
c=df['BMI Category']
f =df['Age']
plt.xlabel('BMI Category',fontsize=30)
plt.ylabel('Age',fontsize=30)
plt.bar(c,f)
```

Out[90]: <BarContainer object of 155 artists>



In [ ]: