

Loading data

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

plt.style.use('dark_background')
```

Loading the data

```
In [30]: file_path = './Sleep_Efficiency.csv'
data_frame = pd.read_csv(file_path)
data_frame.head()
```

Out[30]:

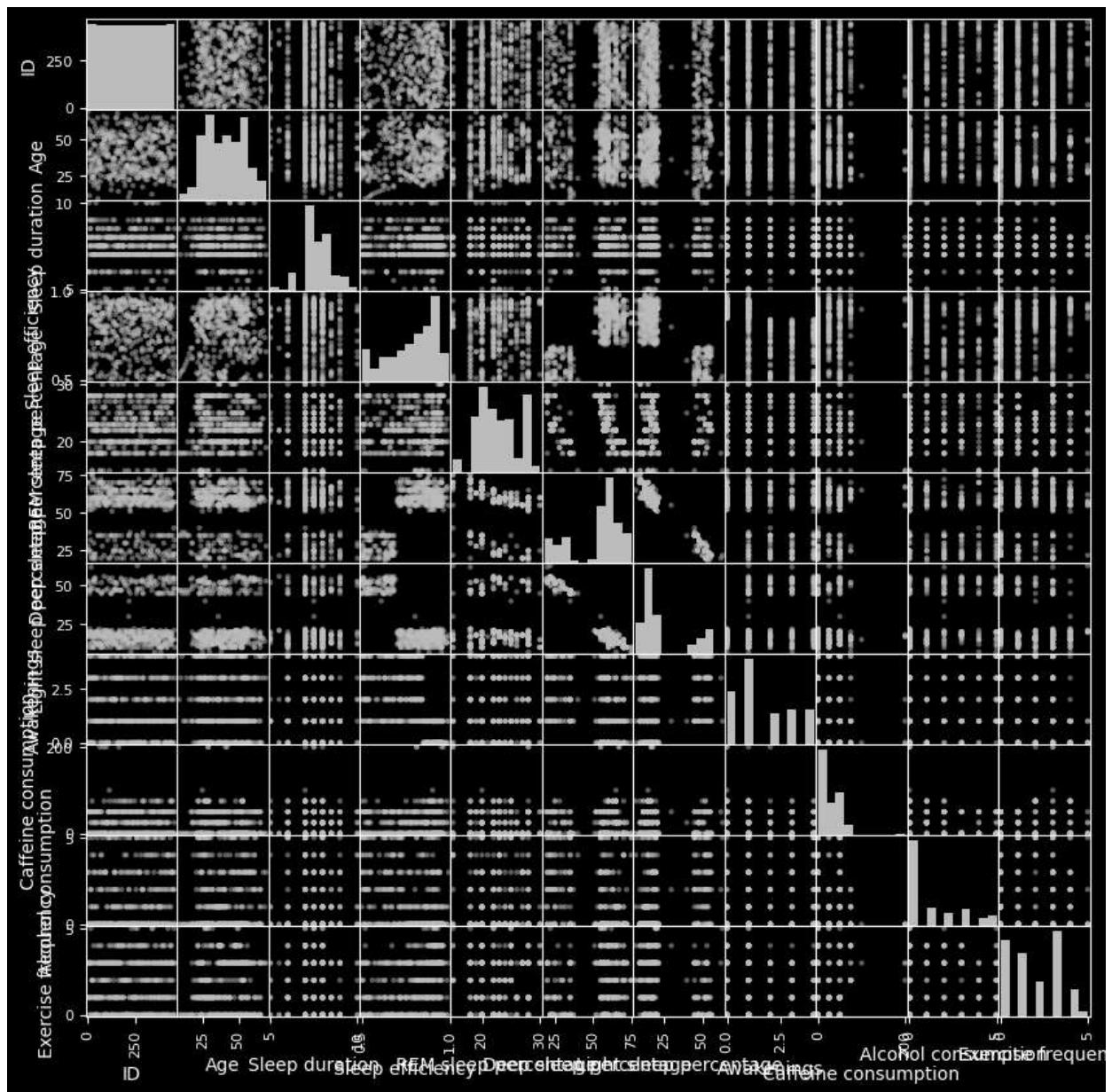
	ID	Age	Gender	Bedtime	Wakeup time	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	All
0	1	65	Female	2021-03-06 01:00:00	2021-03-06 07:00:00	6.0	0.88	18	70	12	
1	2	69	Male	2021-12-05 02:00:00	2021-12-05 09:00:00	7.0	0.66	19	28	53	
2	3	40	Female	2021-05-25 21:30:00	2021-05-25 05:30:00	8.0	0.89	20	70	10	
3	4	40	Female	2021-11-03 02:30:00	2021-11-03 08:30:00	6.0	0.51	23	25	52	
4	5	57	Male	2021-03-13 01:00:00	2021-03-13 09:00:00	8.0	0.76	27	55	18	



Preliminary Plots

```
In [3]: # Generate scatter plots for all pairs of variables
pd.plotting.scatter_matrix(data_frame, figsize=(10, 10))

plt.show()
```



```
In [4]: # Description  
data_frame.describe()
```

Out[4]:

	ID	Age	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	Awakenings
count	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000	432.000000
mean	226.500000	40.285398	7.465708	0.788916	22.615044	52.823009	24.561947	1.641200
std	130.625419	13.172250	0.866625	0.135237	3.525963	15.654235	15.313665	1.356100
min	1.000000	9.000000	5.000000	0.500000	15.000000	18.000000	7.000000	0.000000
25%	113.750000	29.000000	7.000000	0.697500	20.000000	48.250000	15.000000	1.000000
50%	226.500000	40.000000	7.500000	0.820000	22.000000	58.000000	18.000000	1.000000
75%	339.250000	52.000000	8.000000	0.900000	25.000000	63.000000	32.500000	3.000000
max	452.000000	69.000000	10.000000	0.990000	30.000000	75.000000	63.000000	4.000000



Plotting the correlation matrix

Plotting Failed hence the data is not clean.

Data Exploration

In [5]: `data_frame.head()`

	ID	Age	Gender	Bedtime	Wakeup time	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	Awakenings
0	1	65	Female	2021-03-06 01:00:00	2021-03-06 07:00:00	6.0	0.88	18	70	12	
1	2	69	Male	2021-12-05 02:00:00	2021-12-05 09:00:00	7.0	0.66	19	28	53	
2	3	40	Female	2021-05-25 21:30:00	2021-05-25 05:30:00	8.0	0.89	20	70	10	
3	4	40	Female	2021-11-03 02:30:00	2021-11-03 08:30:00	6.0	0.51	23	25	52	
4	5	57	Male	2021-03-13 01:00:00	2021-03-13 09:00:00	8.0	0.76	27	55	18	

Finding null rows using the `isna()` or `isnull()` function

```
In [6]: na_rows = data_frame[data_frame.isna().any(axis=1)]  
print(na_rows)
```

sleep

	ID	Age	Gender	Bedtime	Wakeup time	\
5	6	36	Female	2021-07-01 21:00:00	2021-07-01 04:30:00	
19	20	52	Male	2021-12-03 00:30:00	2021-12-03 07:30:00	
20	21	24	Male	2021-05-02 00:00:00	2021-05-02 08:00:00	
24	25	24	Male	2021-09-16 00:00:00	2021-09-16 07:00:00	
26	27	36	Female	2021-06-18 00:00:00	2021-06-18 07:00:00	
..	\
434	435	28	Female	2021-08-11 23:00:00	2021-08-11 07:00:00	
440	441	28	Female	2021-02-06 21:00:00	2021-02-06 05:00:00	
442	443	27	Female	2021-02-06 01:30:00	2021-02-06 08:30:00	
446	447	23	Male	2021-04-21 00:00:00	2021-04-21 07:00:00	
449	450	40	Female	2021-09-07 23:00:00	2021-09-07 07:30:00	

	Sleep duration	Sleep efficiency	REM sleep percentage	\
5	7.5	0.90	23	
19	7.0	0.89	28	
20	8.0	0.83	15	
24	7.0	0.88	15	
26	7.0	0.95	28	
..	\
434	8.0	0.90	22	
440	8.0	0.88	22	
442	7.0	0.68	23	
446	7.0	0.50	15	
449	8.5	0.55	20	

	Deep sleep percentage	Light sleep percentage	Awakenings	\
5	60	17	0.0	
19	52	20	NaN	
20	75	10	3.0	
24	75	10	1.0	
26	55	17	0.0	
..	\
434	63	15	NaN	
440	63	15	0.0	
442	22	55	1.0	
446	40	45	2.0	
449	32	48	1.0	

	Caffeine consumption	Alcohol consumption	Smoking status	\
5	NaN	0.0	No	
19	50.0	0.0	Yes	
20	0.0	NaN	No	
24	NaN	5.0	No	
26	NaN	0.0	No	
..	\
434	75.0	2.0	No	
440	75.0	NaN	Yes	
442	NaN	4.0	No	
446	0.0	4.0	Yes	
449	NaN	3.0	Yes	

	Exercise frequency
5	1.0
19	3.0
20	2.0
24	2.0
26	1.0
..	...
434	4.0

```
440          2.0
442          1.0
446         NaN
449          0.0
```

[64 rows x 15 columns]

Removing the NaN rows

```
In [7]: data_frame.dropna(inplace = True)
print(data_frame)
```

sleep

	ID	Age	Gender	Bedtime	Wakeup time	\
0	1	65	Female	2021-03-06 01:00:00	2021-03-06 07:00:00	
1	2	69	Male	2021-12-05 02:00:00	2021-12-05 09:00:00	
2	3	40	Female	2021-05-25 21:30:00	2021-05-25 05:30:00	
3	4	40	Female	2021-11-03 02:30:00	2021-11-03 08:30:00	
4	5	57	Male	2021-03-13 01:00:00	2021-03-13 09:00:00	
..	\
445	446	30	Female	2021-11-16 23:00:00	2021-11-16 06:30:00	
447	448	27	Female	2021-11-13 22:00:00	2021-11-13 05:30:00	
448	449	52	Male	2021-03-31 21:00:00	2021-03-31 03:00:00	
450	451	45	Male	2021-07-29 21:00:00	2021-07-29 04:00:00	
451	452	18	Male	2021-03-17 02:30:00	2021-03-17 10:00:00	

	Sleep duration	Sleep efficiency	REM sleep percentage	\
0	6.0	0.88	18	
1	7.0	0.66	19	
2	8.0	0.89	20	
3	6.0	0.51	23	
4	8.0	0.76	27	
..	\
445	7.5	0.53	28	
447	7.5	0.91	22	
448	6.0	0.74	28	
450	7.0	0.76	18	
451	7.5	0.63	22	

	Deep sleep percentage	Light sleep percentage	Awakenings	\
0	70	12	0.0	
1	28	53	3.0	
2	70	10	1.0	
3	25	52	3.0	
4	55	18	3.0	
..	\
445	20	52	4.0	
447	57	21	0.0	
448	57	15	4.0	
450	72	10	3.0	
451	23	55	1.0	

	Caffeine consumption	Alcohol consumption	Smoking status	\
0	0.0	0.0	Yes	
1	0.0	3.0	Yes	
2	0.0	0.0	No	
3	50.0	5.0	Yes	
4	0.0	3.0	No	
..	\
445	50.0	2.0	Yes	
447	0.0	0.0	No	
448	25.0	0.0	No	
450	0.0	0.0	No	
451	50.0	0.0	No	

	Exercise frequency
0	3.0
1	3.0
2	3.0
3	1.0
4	3.0
..	...
445	1.0

447	5.0
448	3.0
450	3.0
451	1.0

[388 rows x 15 columns]

In [8]: `data_frame.corr(numeric_only= True)`

Out[8]:

	ID	Age	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	Awakenings
ID	1.000000	0.012011	0.030504	0.025931	0.086319	-0.015469	-0.003608	-0.08
Age	0.012011	1.000000	-0.065951	0.124093	0.015449	0.058156	-0.062613	-0.00
Sleep duration	0.030504	-0.065951	1.000000	-0.019164	-0.015408	-0.035477	0.039536	-0.00
Sleep efficiency	0.025931	0.124093	-0.019164	1.000000	0.064038	0.789087	-0.816934	-0.56
REM sleep percentage	0.086319	0.015449	-0.015408	0.064038	1.000000	-0.185850	-0.035046	-0.02
Deep sleep percentage	-0.015469	0.058156	-0.035477	0.789087	-0.185850	1.000000	-0.975461	-0.32
Light sleep percentage	-0.003608	-0.062613	0.039536	-0.816934	-0.035046	-0.975461	1.000000	0.33
Awakenings	-0.084746	-0.004006	-0.009162	-0.567885	-0.024608	-0.327284	0.338397	1.00
Caffeine consumption	-0.056247	-0.169057	-0.030922	0.070804	0.113965	-0.024522	-0.000596	-0.11
Alcohol consumption	0.089620	0.069295	-0.048436	-0.397012	-0.039597	-0.374614	0.389895	0.21
Exercise frequency	0.009799	0.071241	-0.046671	0.266050	0.044200	0.171841	-0.184684	-0.23

Key Insights

- Hypothesis: Age and Sleep efficiency are positively correlated. (Counterintuitive)
Justification: The correlation coefficient between Age and Sleep efficiency is 0.124093, indicating a weak positive correlation.
- Hypothesis: Sleep duration and Sleep efficiency are negatively correlated. Justification: The correlation coefficient between Sleep duration and Sleep efficiency is -0.019164, suggesting a weak negative correlation.
- Hypothesis: p percentage and Sleep efficiency are positively correlated. Justification: The correlation coefficient between Deep sleep percentage and Sleep efficiency is 0.789087, indicating a strong positive correlation.

4. Hypothesis: Light sleep percentage and Sleep efficiency are negatively correlated.

Justification: The correlation coefficient between Light sleep percentage and Sleep efficiency is -0.816934, suggesting a strong negative correlation.

5. Hypothesis: Awakenings and Sleep efficiency are negatively correlated. Justification: The correlation coefficient between Awakenings and Sleep efficiency is -0.567885, indicating a moderate negative correlation.

Testing Hypothesis

Hypothesis 1:

Age and Sleep efficiency are positively correlated. (Counterintuitive) Justification: The correlation coefficient between Age and Sleep efficiency is 0.124093, indicating a weak positive correlation.

```
In [14]: age = data_frame.sort_values('Age')
age.head(5)
```

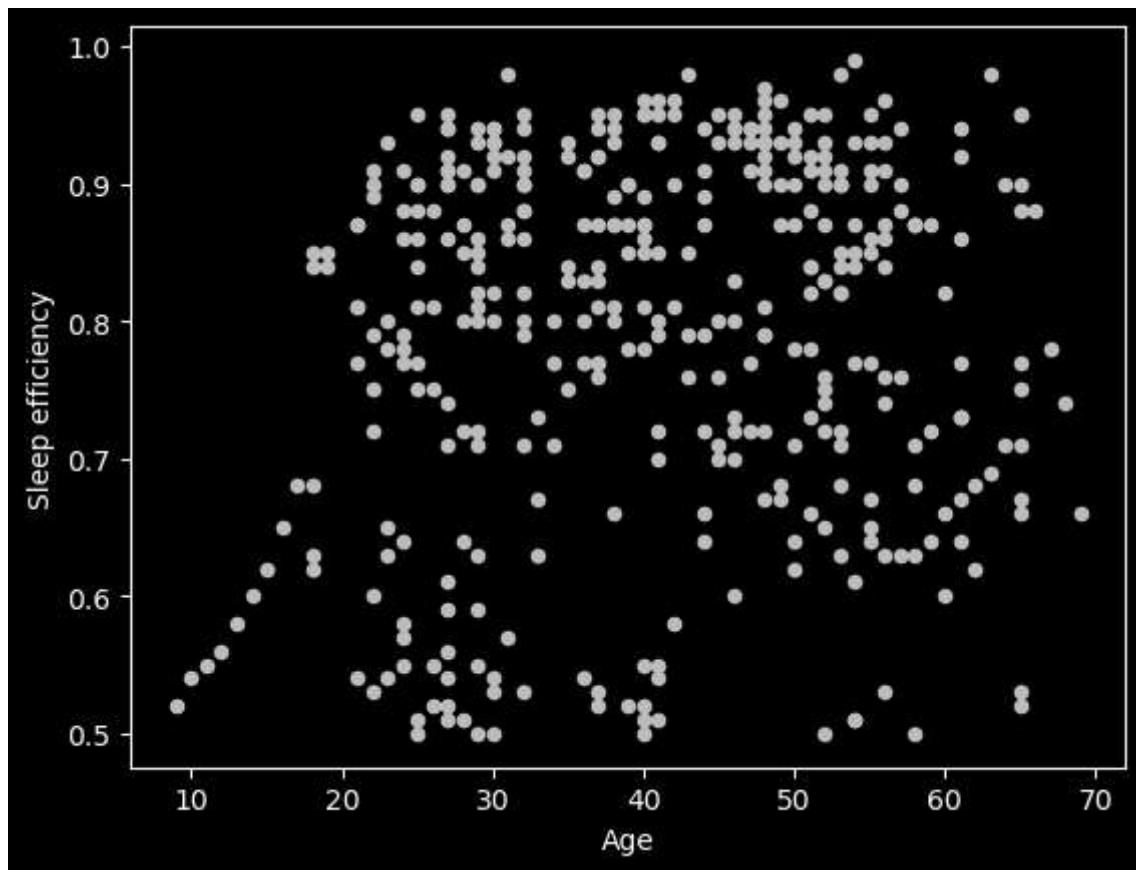
Out[14]:

	ID	Age	Gender	Bedtime	Wakeup time	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage
52	53	9	Female	2021-11-27 01:00:00	2021-11-27 09:30:00	8.5	0.52	18	35	47
211	212	10	Female	2021-07-15 01:00:00	2021-07-15 08:00:00	7.0	0.54	18	35	47
9	10	11	Female	2021-09-16 01:00:00	2021-09-16 10:00:00	9.0	0.55	18	37	45
332	333	12	Female	2021-11-01 01:00:00	2021-11-01 10:00:00	9.0	0.56	18	35	47
219	220	13	Female	2021-03-11 01:00:00	2021-03-11 09:00:00	8.0	0.58	18	37	45

◀ ▶

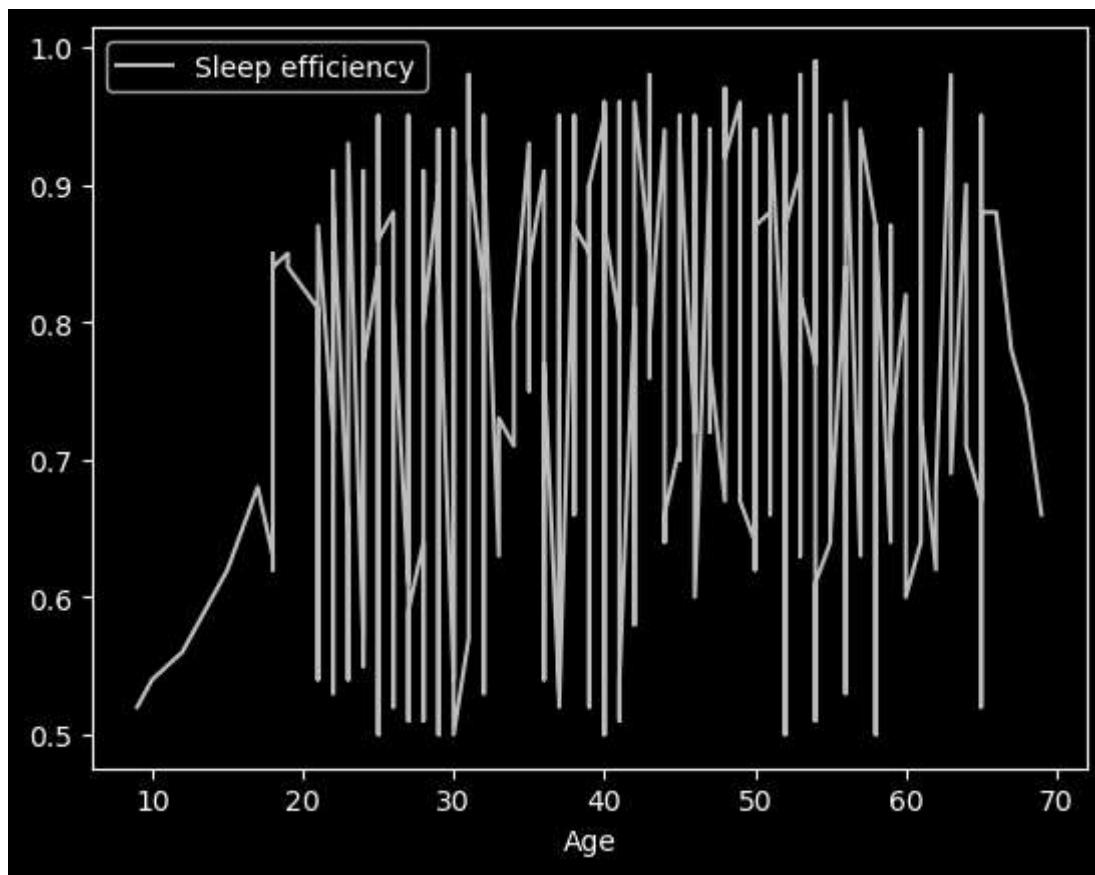
```
In [12]: age.plot(x='Age', y= 'Sleep efficiency', kind= 'scatter')
```

```
Out[12]: <Axes: xlabel='Age', ylabel='Sleep efficiency'>
```



```
In [13]: age.plot(x='Age', y= 'Sleep efficiency', kind= 'line')
```

```
Out[13]: <Axes: xlabel='Age'>
```



Conclusion 1:

This hypothesis seems controversial and the data does not suggest strong correlation

Hypothesis 2:

Sleep duration and Sleep efficiency are negatively correlated. Justification: The correlation coefficient between Sleep duration and Sleep efficiency is -0.019164, suggesting a weak negative correlation.

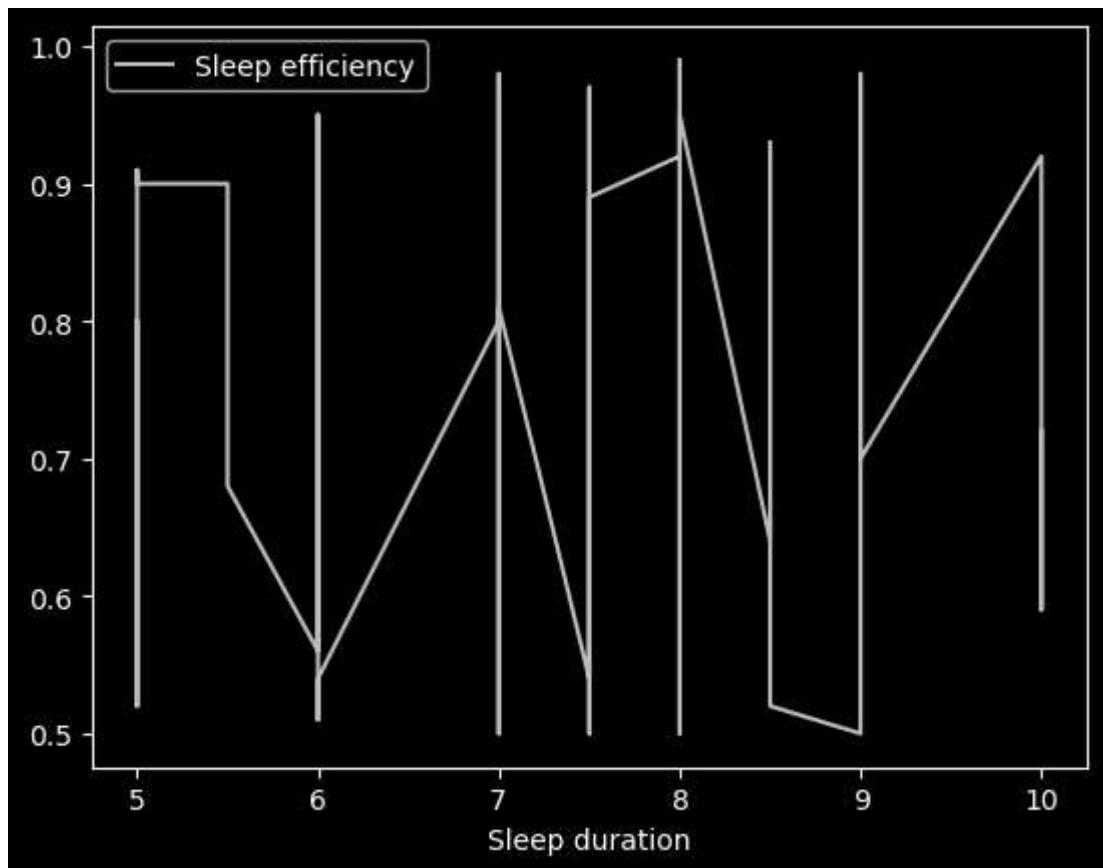
```
In [15]: sleep = data_frame.sort_values('Sleep duration')
sleep.head()
```

Out[15]:

	ID	Age	Gender	Bedtime	Wakeup time	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage
90	91	45	Male	2021-03-29 02:30:00	2021-03-29 07:30:00	5.0	0.80	24	60	16
367	368	56	Male	2021-08-03 02:30:00	2021-08-03 07:30:00	5.0	0.74	22	59	19
232	233	36	Female	2021-03-06 23:00:00	2021-03-06 04:00:00	5.0	0.87	22	65	13
198	199	32	Female	2021-12-31 01:30:00	2021-12-31 06:30:00	5.0	0.86	20	65	15
110	111	51	Male	2021-06-18 00:30:00	2021-06-18 05:30:00	5.0	0.91	23	57	20

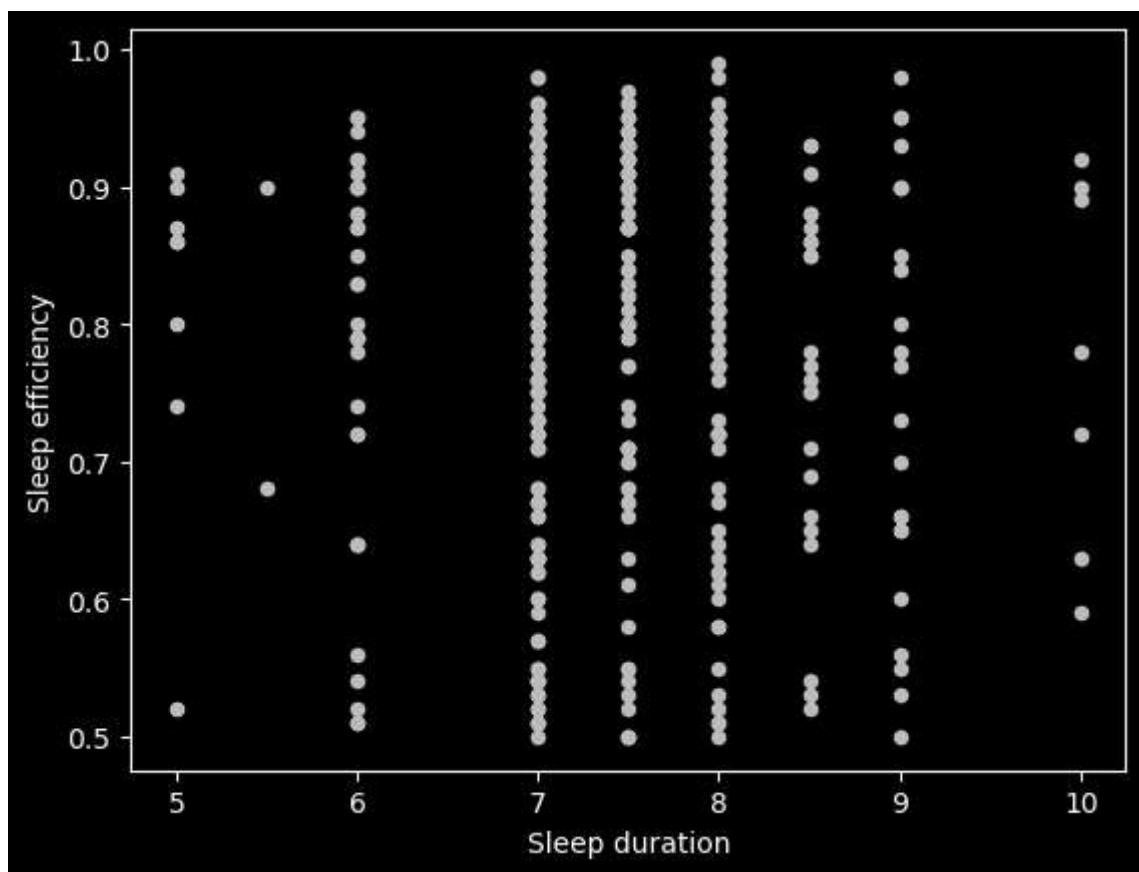
```
In [17]: sleep.plot(x= 'Sleep duration', y='Sleep efficiency')
```

Out[17]: <Axes: xlabel='Sleep duration'>



```
In [18]: sleep.plot(x= 'Sleep duration', y='Sleep efficiency', kind='scatter')
```

```
Out[18]: <Axes: xlabel='Sleep duration', ylabel='Sleep efficiency'>
```



Conclusion 2:

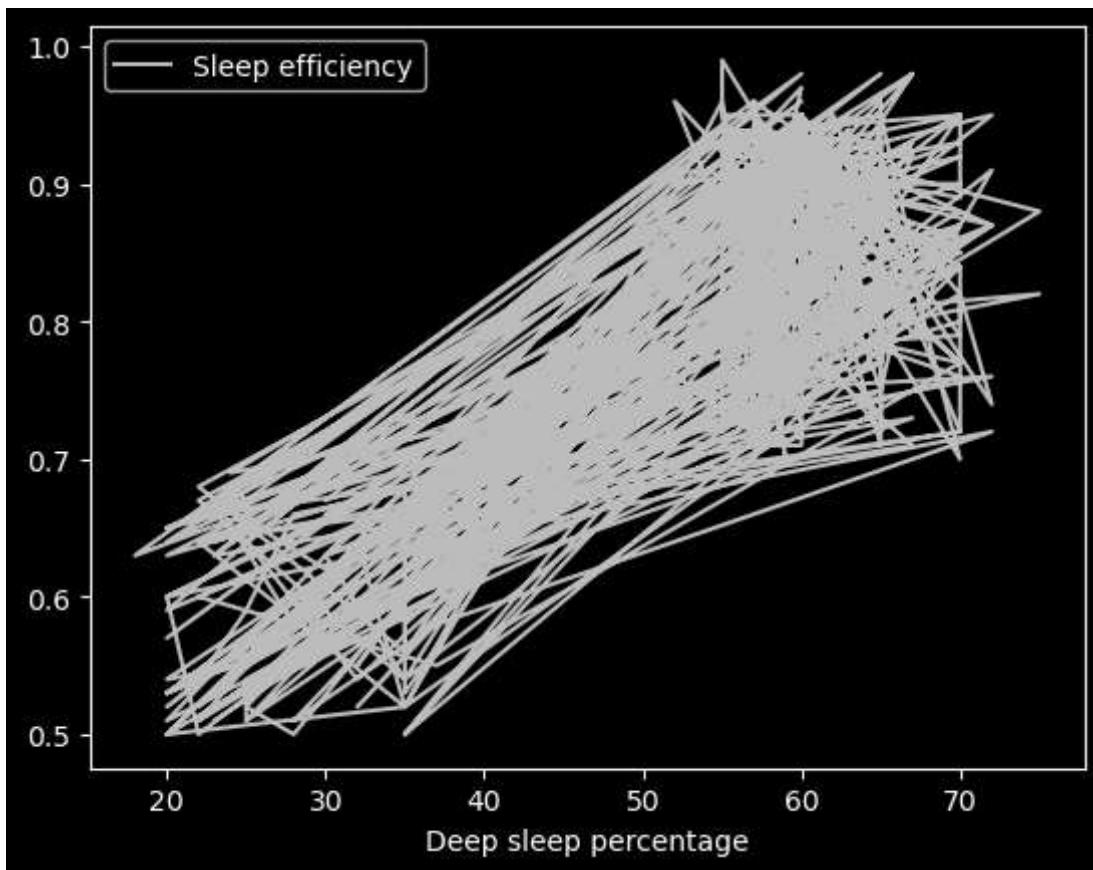
This hypothesis also seems counterintuitive, we must all factor in various other aspects...

Hypothesis 3:

deep sleep percentage and Sleep efficiency are positively correlated. Justification: The correlation coefficient between Deep sleep percentage and Sleep efficiency is 0.789087, indicating a strong positive correlation.

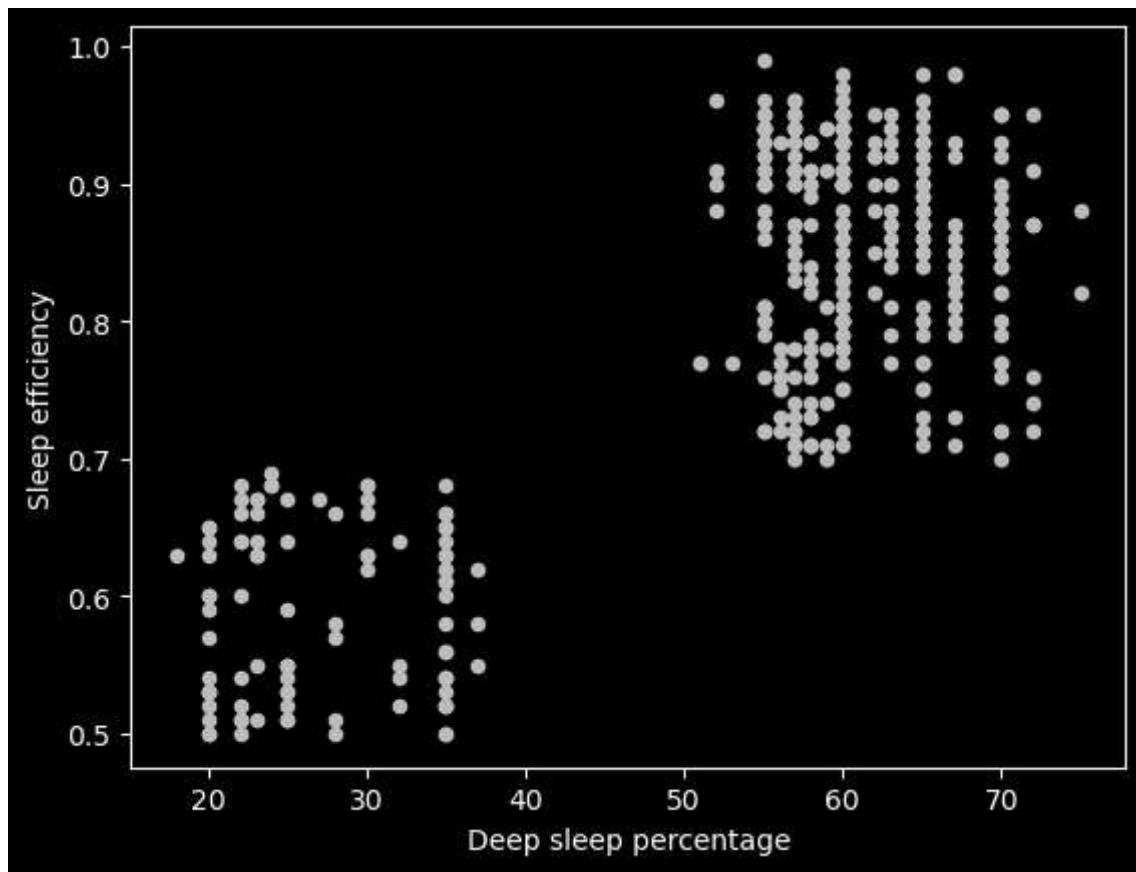
```
In [20]: sleep.plot(x='Deep sleep percentage', y='Sleep efficiency')
```

```
Out[20]: <Axes: xlabel='Deep sleep percentage'>
```



```
In [21]: sleep.plot(x='Deep sleep percentage', y='Sleep efficiency', kind='scatter')
```

```
Out[21]: <Axes: xlabel='Deep sleep percentage', ylabel='Sleep efficiency'>
```



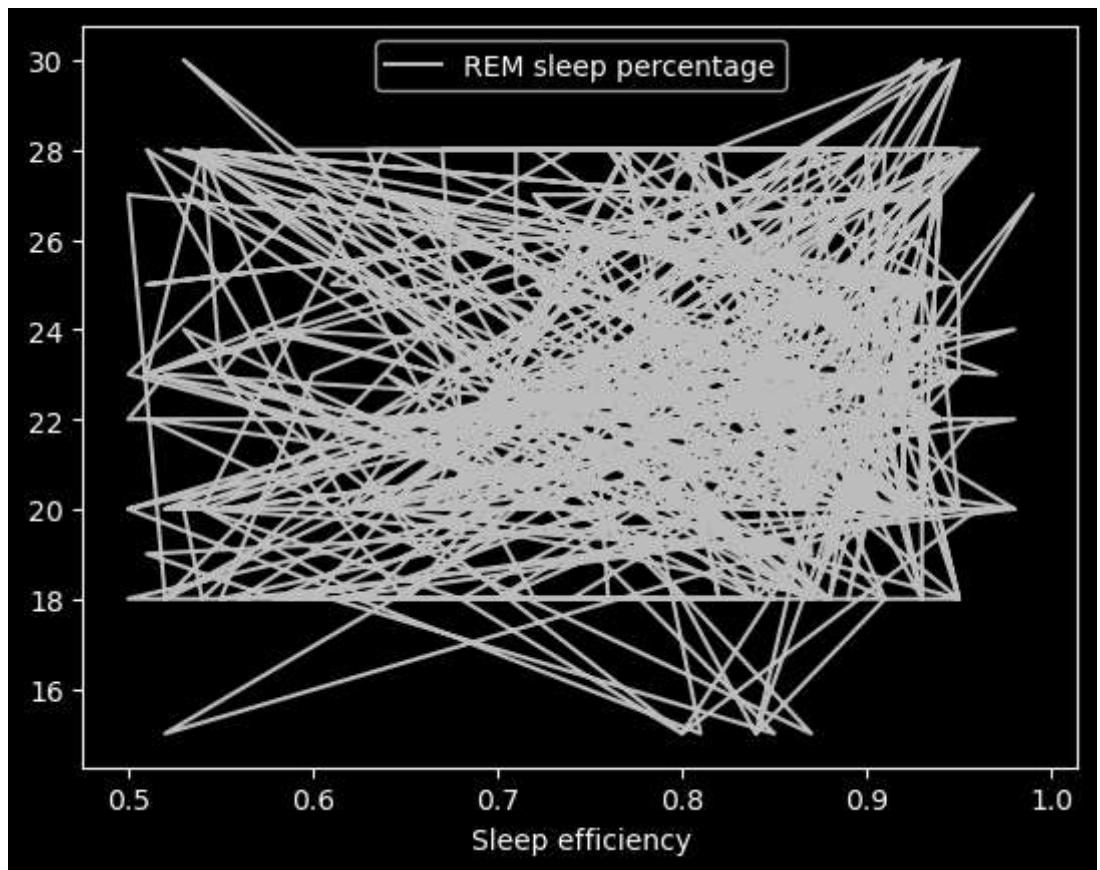
Conclusion 3:

seems very logical and strong data inference we can confidently stipulate that higher deep sleep percentage causes better sleep efficiency

On an off note we can also check against rem sleep

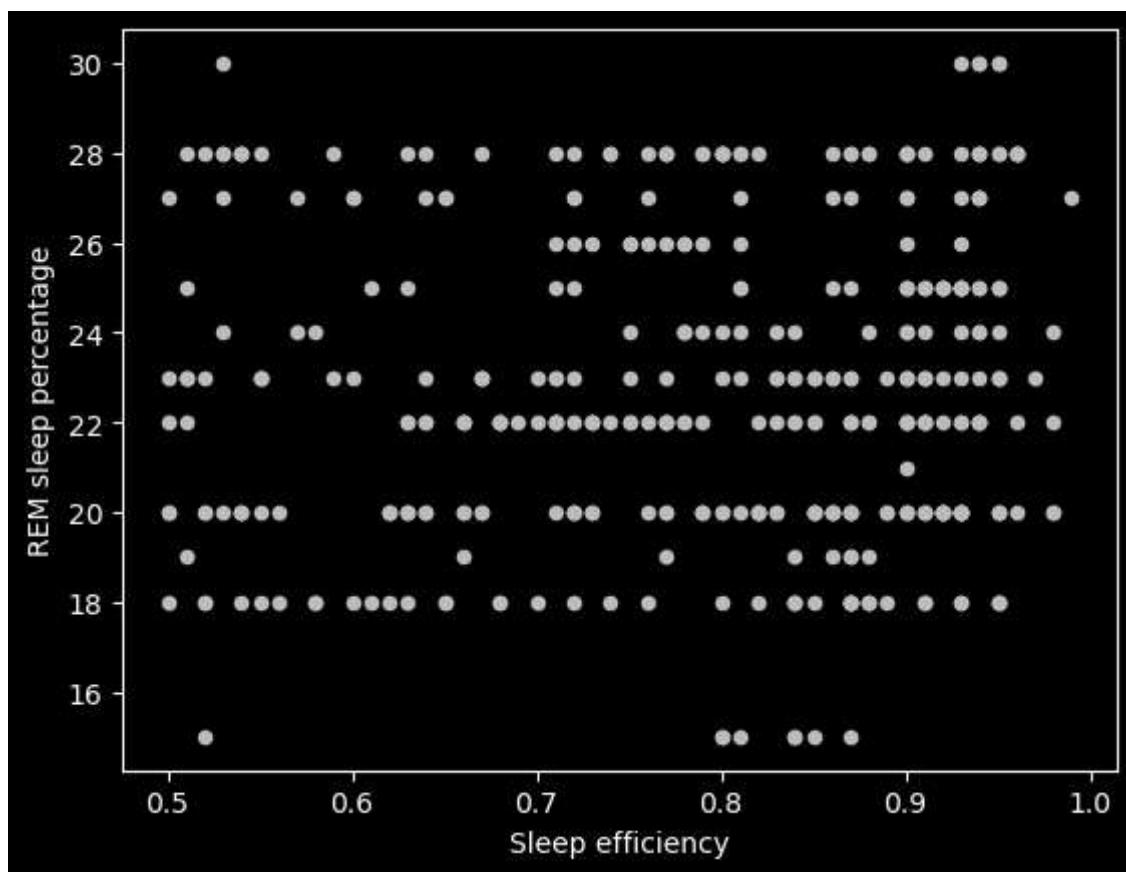
```
In [23]: sleep.plot(x='Sleep efficiency', y='REM sleep percentage')
```

```
Out[23]: <Axes: xlabel='Sleep efficiency'>
```



```
In [22]: sleep.plot(x='Sleep efficiency', y='REM sleep percentage', kind='scatter')
```

```
Out[22]: <Axes: xlabel='Sleep efficiency', ylabel='REM sleep percentage'>
```

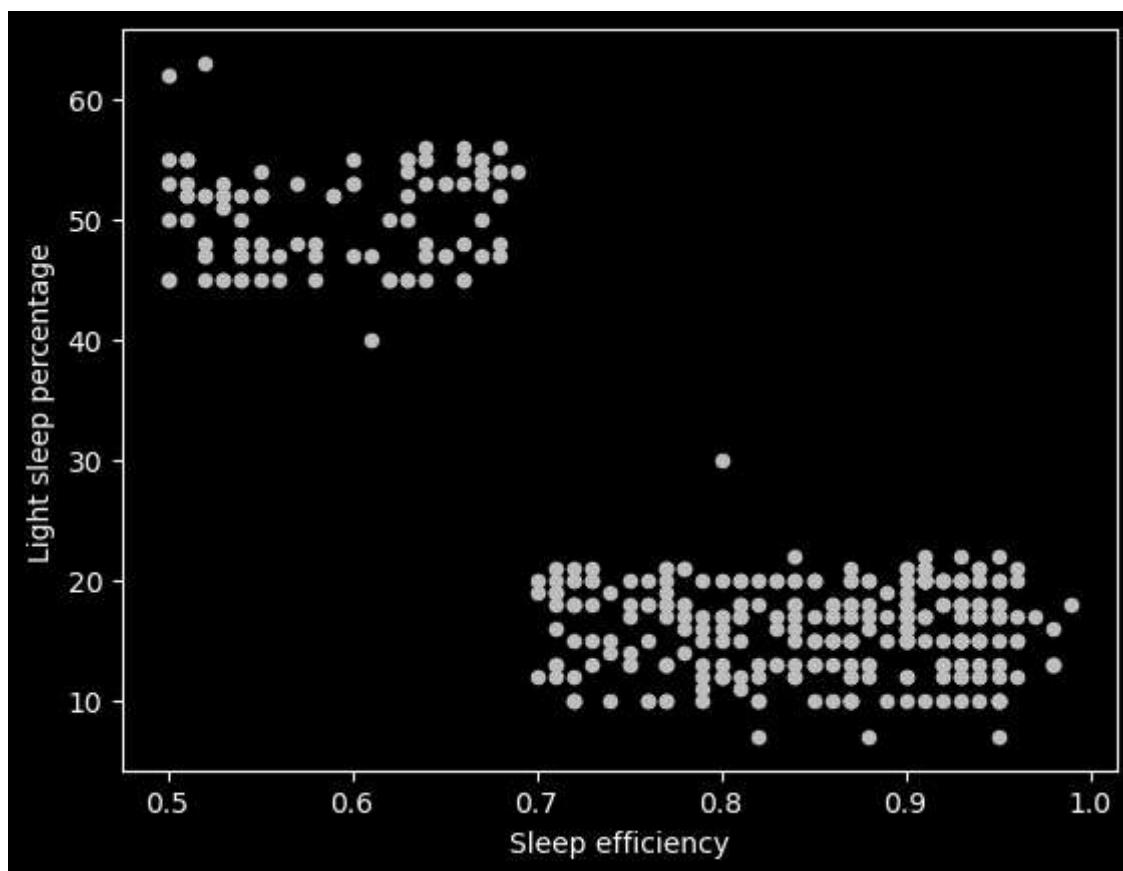


Hypothesis 4:

Light sleep percentage and Sleep efficiency are negatively correlated. Justification: The correlation coefficient between Light sleep percentage and Sleep efficiency is -0.816934, suggesting a strong negative correlation.

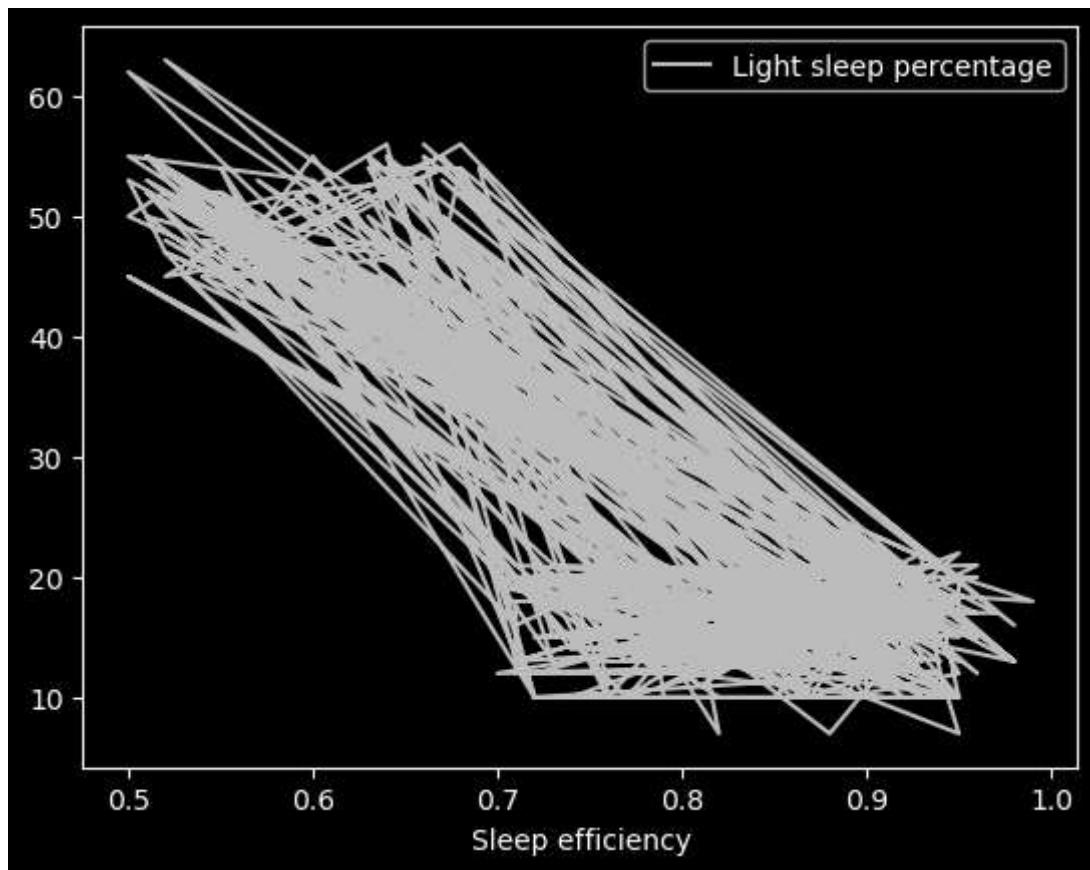
```
In [25]: sleep.plot(x='Sleep efficiency', y='Light sleep percentage', kind='scatter')
```

```
Out[25]: <Axes: xlabel='Sleep efficiency', ylabel='Light sleep percentage'>
```



```
In [26]: sleep.plot(x='Sleep efficiency', y='Light sleep percentage', kind='line')
```

```
Out[26]: <Axes: xlabel='Sleep efficiency'>
```

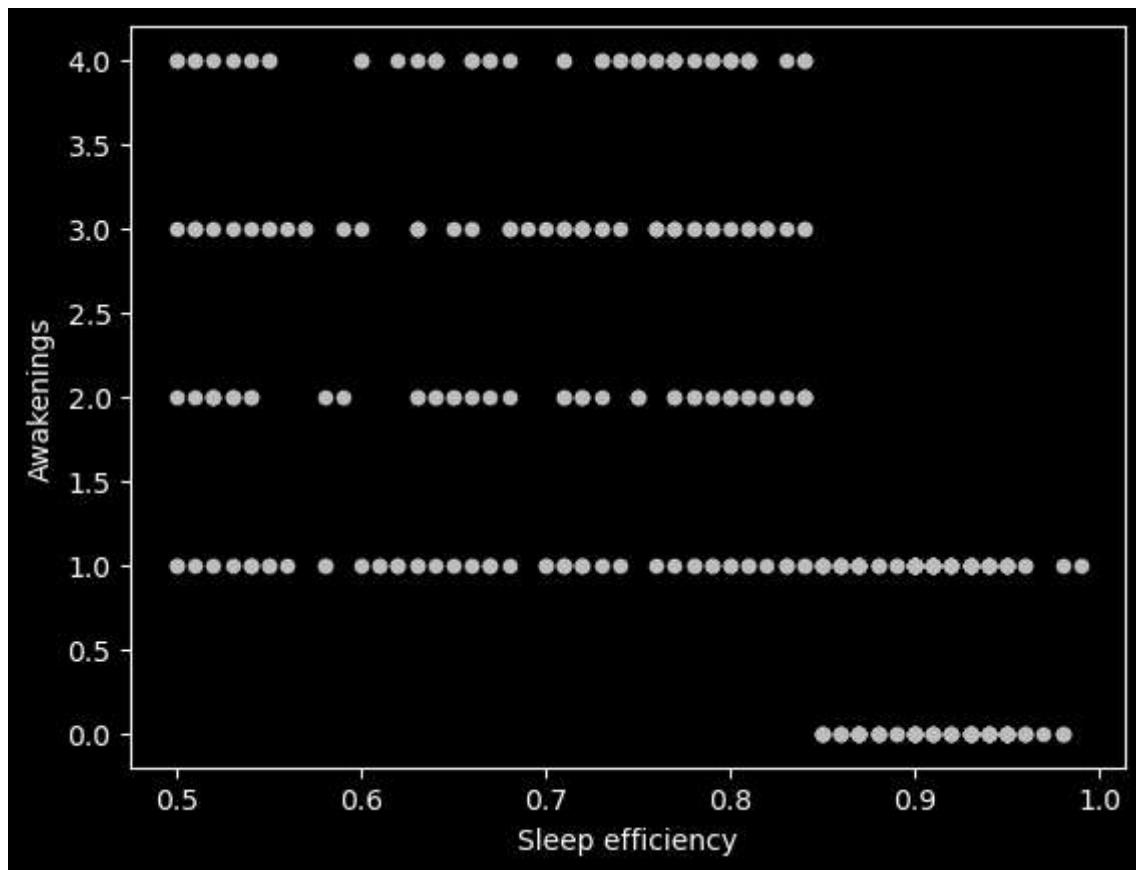


Hypothesis 5:

Awakenings and Sleep efficiency are negatively correlated. Justification: The correlation coefficient between Awakenings and Sleep efficiency is -0.567885, indicating a moderate negative correlation.

```
In [27]: sleep.plot(x='Sleep efficiency', y='Awakenings', kind='scatter')
```

```
Out[27]: <Axes: xlabel='Sleep efficiency', ylabel='Awakenings'>
```



Conclusion

No strong actionable points found hence further ahead we will segregate the data in age bands to find trends.

Filtering Data

Age group 0 - 10

```
In [33]: group1 = data_frame[data_frame.Age <= 10]  
group1.describe()
```

Out[33]:

	ID	Age	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	Awakenings
count	2.000000	2.000000	2.000000	2.000000	2.0	2.0	2.0	2.000000
mean	132.500000	9.500000	7.750000	0.530000	18.0	35.0	47.0	2.500000
std	112.429978	0.707107	1.06066	0.014142	0.0	0.0	0.0	0.707107
min	53.000000	9.000000	7.000000	0.520000	18.0	35.0	47.0	2.000000
25%	92.750000	9.250000	7.37500	0.525000	18.0	35.0	47.0	2.250000
50%	132.500000	9.500000	7.750000	0.530000	18.0	35.0	47.0	2.500000
75%	172.250000	9.750000	8.12500	0.535000	18.0	35.0	47.0	2.750000
max	212.000000	10.000000	8.500000	0.540000	18.0	35.0	47.0	3.000000

--	--

Too small data set

Age group 10 - 20

In [34]: `group2 = data_frame[data_frame.Age <= 20]
group2.describe()`

Out[34]:

	ID	Age	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	Awakenings
count	18.000000	18.000000	18.000000	18.000000	18.000000	18.000000	18.000000	17.000000
mean	186.388889	15.777778	8.000000	0.685556	19.611111	43.000000	37.388889	2.117647
std	139.046390	3.456546	0.747545	0.129352	2.569937	16.003676	15.882432	1.268974
min	10.000000	9.000000	7.000000	0.520000	15.000000	23.000000	12.000000	0.000000
25%	54.250000	13.250000	7.500000	0.585000	18.000000	35.000000	18.500000	1.000000
50%	176.500000	17.500000	8.000000	0.640000	18.500000	35.000000	46.000000	2.000000
75%	272.500000	18.000000	8.500000	0.835000	21.500000	56.750000	47.000000	3.000000
max	452.000000	20.000000	9.000000	0.910000	26.000000	70.000000	55.000000	4.000000

--	--

In [33]: `group1 = data_frame[data_frame.Age <= 10]
group1.describe()`

Out[33]:

	ID	Age	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	Awakenings
count	2.000000	2.000000	2.00000	2.000000	2.0	2.0	2.0	2.000000
mean	132.500000	9.500000	7.75000	0.530000	18.0	35.0	47.0	2.500000
std	112.429978	0.707107	1.06066	0.014142	0.0	0.0	0.0	0.707107
min	53.000000	9.000000	7.00000	0.520000	18.0	35.0	47.0	2.000000
25%	92.750000	9.250000	7.37500	0.525000	18.0	35.0	47.0	2.250000
50%	132.500000	9.500000	7.75000	0.530000	18.0	35.0	47.0	2.500000
75%	172.250000	9.750000	8.12500	0.535000	18.0	35.0	47.0	2.750000
max	212.000000	10.000000	8.50000	0.540000	18.0	35.0	47.0	3.000000

