

WHAT ARE THE BEST
PREDICTORS OF SLEEP
EFFICIENCY?

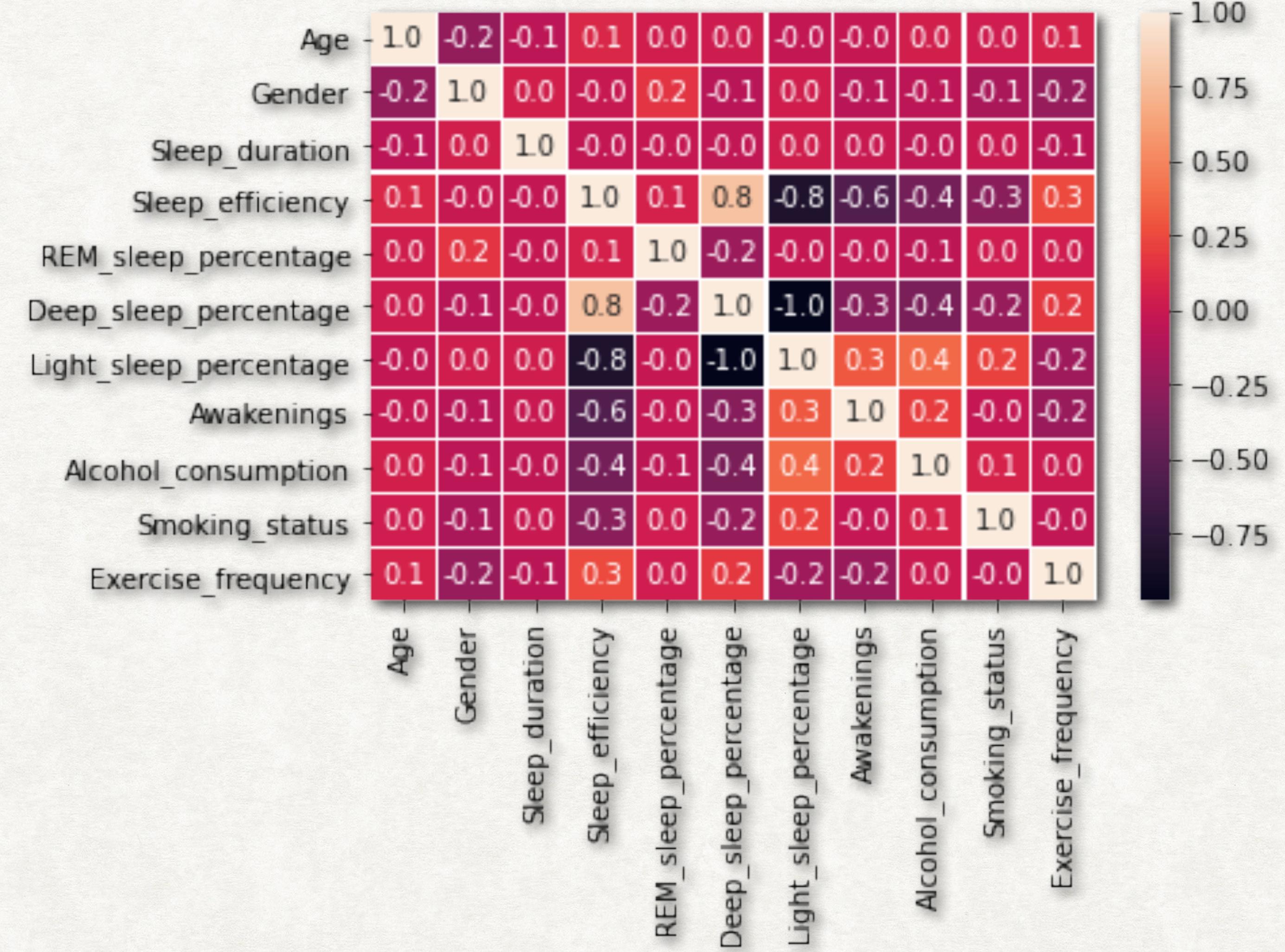
[HTTPS://WWW.KAGGLE.COM/
DATASETS/EQUILIBRIUMM/SLEEP-
EFFICIENCY](https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency)

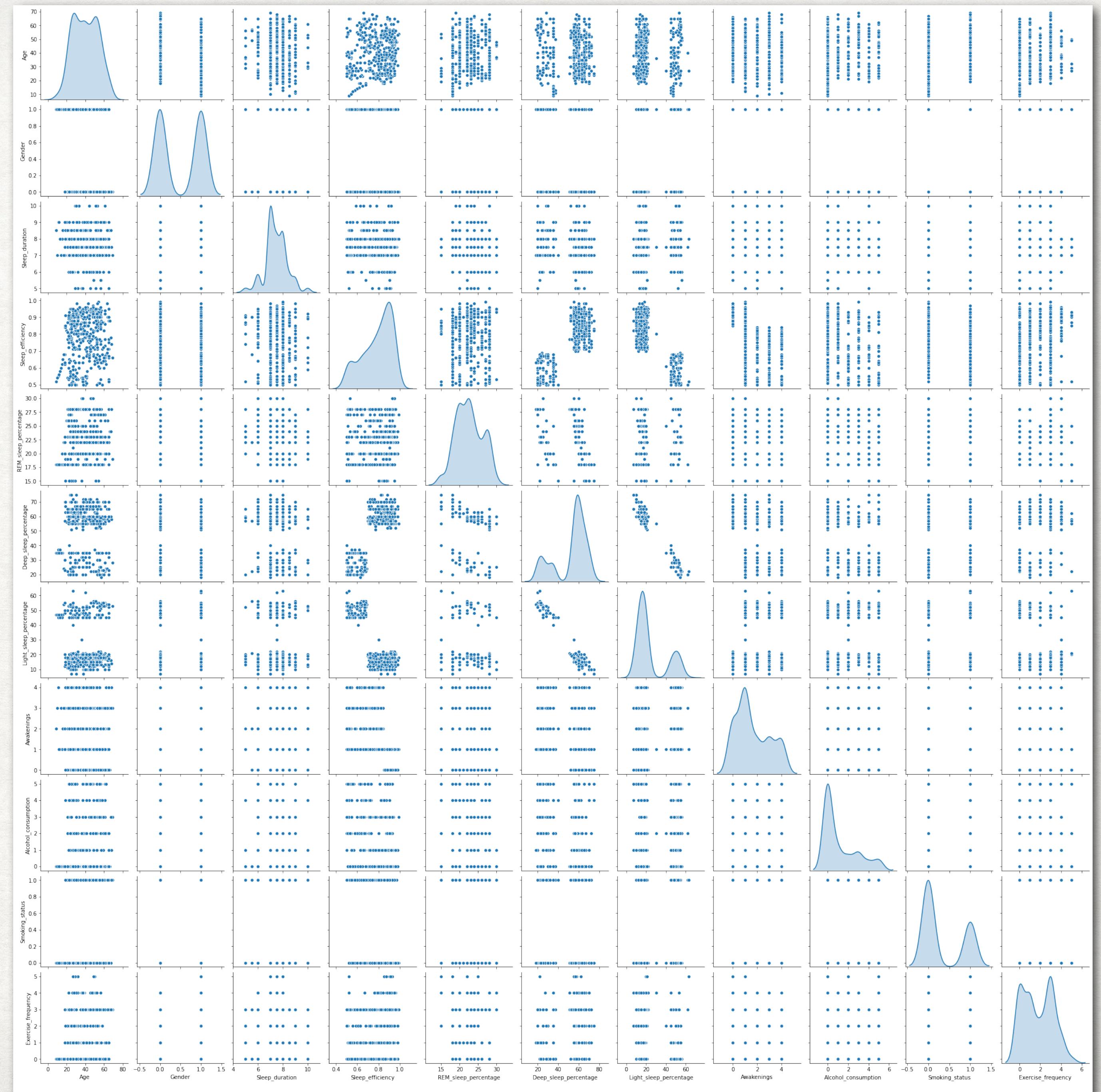


Variable	Description
Subject ID	Unique ID number assigned to each participant.
Age	Ages from 9 - 69
Gender	Male = 0, Female =1
Sleep_duration	Total number of hours slept; float
Sleep_efficiency	Response Variable
REM_sleep_percentage	Percentage of time spent in REM sleep
Deep_sleep_percentage	Percentage of time spent in Deep sleep
Light_sleep_percentage	Percentage of time spent in Light sleep
Akakenings	Number of times each subject wakes up during the night; int
Caffeine_consumption	How much caffeine subject consumed in previous 24 hours
Alcohol_consumption	How much alcohol subject consumed in previous 24 hours
Smoking_status	1 = Smoker, 0 = Nonsmoker
Exercise_frequency	Number times subject exercises per week; int

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 452 entries, 0 to 451
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   ID               452 non-null    int64  
 1   Age              452 non-null    int64  
 2   Gender            452 non-null    int64  
 3   Bedtime           452 non-null    object  
 4   Wakeup_time       452 non-null    object  
 5   Sleep_duration    452 non-null    float64 
 6   Sleep_efficiency 452 non-null    float64 
 7   REM_sleep_percentage 452 non-null  int64  
 8   Deep_sleep_percentage 452 non-null  int64  
 9   Light_sleep_percentage 452 non-null  int64  
 10  Awakenings        432 non-null    float64 
 11  Caffeine_consumption 427 non-null  float64 
 12  Alcohol_consumption 438 non-null  float64 
 13  Smoking_status    452 non-null    int64  
 14  Exercise_frequency 446 non-null    float64 
dtypes: float64(6), int64(7), object(2)
memory usage: 53.1+ KB
```





OLS Regression Results

Dep. Variable:	Sleep_efficiency	R-squared:	0.804
Model:	OLS	Adj. R-squared:	0.800
Method:	Least Squares	F-statistic:	180.9
Date:	Thu, 02 Mar 2023	Prob (F-statistic):	3.88e-149
Time:	22:40:14	Log-Likelihood:	631.74
No. Observations:	452	AIC:	-1241.
Df Residuals:	441	BIC:	-1196.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0002	1.08e-05	21.152	0.000	0.000	0.000
Age	0.0007	0.000	3.334	0.001	0.000	0.001
Gender	-0.0028	0.006	-0.437	0.663	-0.015	0.010
Sleep_duration	0.0011	0.003	0.319	0.750	-0.005	0.008
REM_sleep_percentage	0.0111	0.001	15.078	0.000	0.010	0.013
Deep_sleep_percentage	0.0089	0.000	21.416	0.000	0.008	0.010
Light_sleep_percentage	0.0028	0.001	4.540	0.000	0.002	0.004
Awakenings	-0.0334	0.002	-14.181	0.000	-0.038	-0.029
Alcohol_consumption	-0.0057	0.002	-2.915	0.004	-0.010	-0.002
Smoking_status	-0.0419	0.006	-6.599	0.000	-0.054	-0.029
Exercise_frequency	0.0048	0.002	2.237	0.026	0.001	0.009
Deep_sleep_percentage:Light_sleep_percentage	2.82e-05	1.98e-05	1.426	0.154	-1.07e-05	6.71e-05

Omnibus:	11.605	Durbin-Watson:	1.960
Prob(Omnibus):	0.003	Jarque-Bera (JB):	11.850
Skew:	-0.374	Prob(JB):	0.00267
Kurtosis:	2.736	Cond. No.	6.08e+18

OLS Regression Results

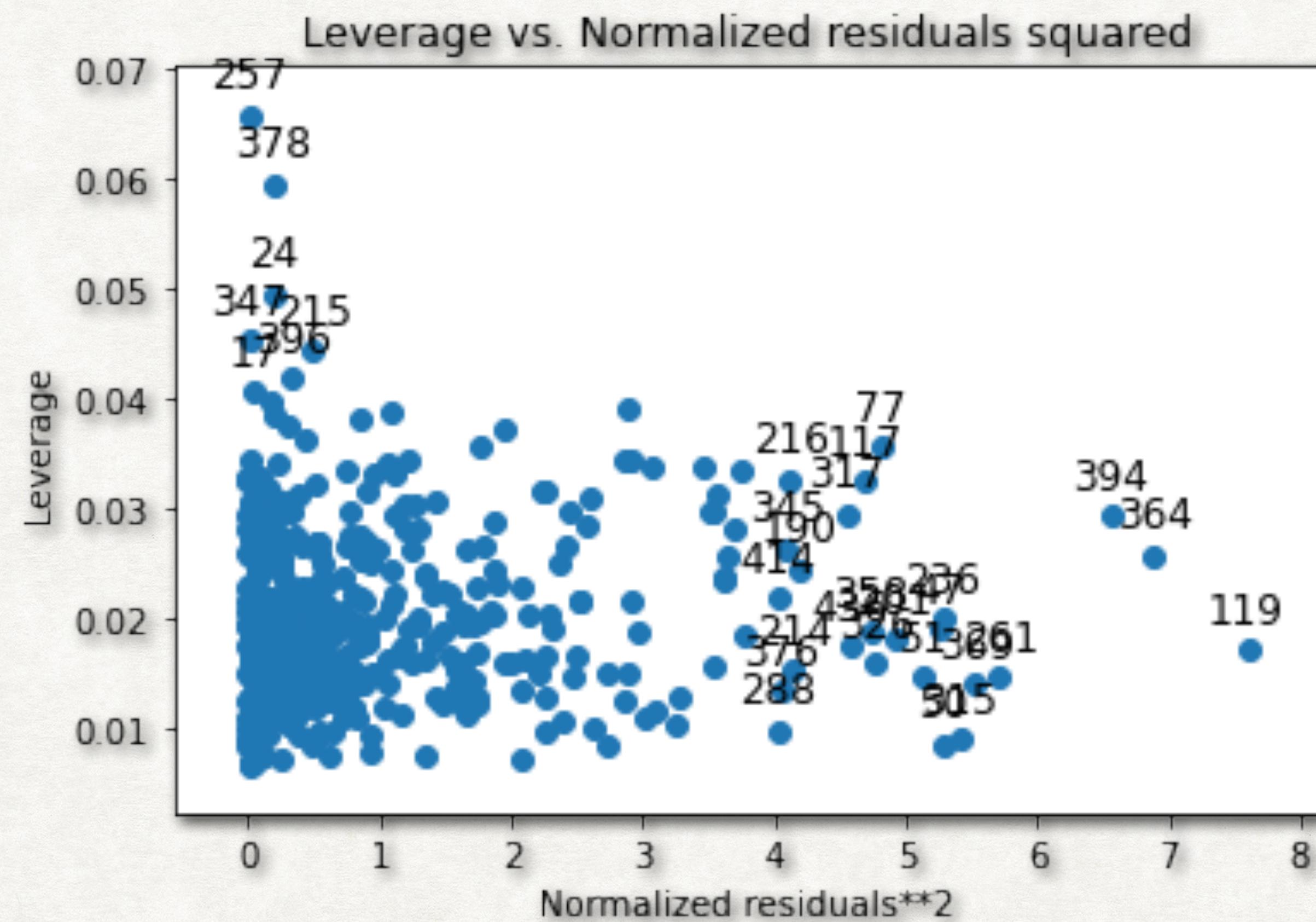
Dep. Variable:	Sleep_efficiency	R-squared:	0.804
Model:	OLS	Adj. R-squared:	0.800
Method:	Least Squares	F-statistic:	226.9
Date:	Thu, 02 Mar 2023	Prob (F-statistic):	1.97e-151
Time:	22:40:19	Log-Likelihood:	631.60
No. Observations:	452	AIC:	-1245.
Df Residuals:	443	BIC:	-1208.
Df Model:	8		
Covariance Type:	nonrobust		

		coef	std err	t	P> t	[0.025	0.975]
Intercept		0.0002	7.2e-06	31.816	0.000	0.000	0.000
Age		0.0008	0.000	3.541	0.000	0.000	0.001
REM_sleep_percentage		0.0111	0.001	16.656	0.000	0.010	0.012
Deep_sleep_percentage		0.0089	0.000	27.105	0.000	0.008	0.010
Light_sleep_percentage		0.0029	0.001	5.084	0.000	0.002	0.004
Awakenings		-0.0333	0.002	-14.317	0.000	-0.038	-0.029
Alcohol_consumption		-0.0057	0.002	-2.934	0.004	-0.010	-0.002
Smoking_status		-0.0414	0.006	-6.645	0.000	-0.054	-0.029
Exercise_frequency		0.0050	0.002	2.410	0.016	0.001	0.009
Deep_sleep_percentage:Light_sleep_percentage		2.875e-05	1.97e-05	1.459	0.145	-9.98e-06	6.75e-05

Omnibus:	11.826	Durbin-Watson:	1.957
Prob(Omnibus):	0.003	Jarque-Bera (JB):	11.981
Skew:	-0.373	Prob(JB):	0.00250
Kurtosis:	2.716	Cond. No.	7.20e+18

OLS Regression Results

Dep. Variable:	Sleep_efficiency	R-squared:	0.803			
Model:	OLS	Adj. R-squared:	0.800			
Method:	Least Squares	F-statistic:	258.4			
Date:	Thu, 02 Mar 2023	Prob (F-statistic):	3.40e-152			
Time:	22:40:24	Log-Likelihood:	630.51			
No. Observations:	452	AIC:	-1245.			
Df Residuals:	444	BIC:	-1212.			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0002	4.34e-06	54.739	0.000	0.000	0.000
Age	0.0008	0.000	3.504	0.001	0.000	0.001
REM_sleep_percentage	0.0108	0.001	16.844	0.000	0.010	0.012
Deep_sleep_percentage	0.0093	0.000	43.634	0.000	0.009	0.010
Light_sleep_percentage	0.0036	0.000	13.197	0.000	0.003	0.004
Awakenings	-0.0335	0.002	-14.474	0.000	-0.038	-0.029
Alcohol_consumption	-0.0057	0.002	-2.914	0.004	-0.010	-0.002
Smoking_status	-0.0424	0.006	-6.834	0.000	-0.055	-0.030
Exercise_frequency	0.0051	0.002	2.457	0.014	0.001	0.009
Omnibus:	11.123	Durbin-Watson:	1.975			
Prob(Omnibus):	0.004	Jarque-Bera (JB):	11.012			
Skew:	-0.349	Prob(JB):	0.00406			
Kurtosis:	2.687	Cond. No.	4.91e+17			



OLS Regression Results

Dep. Variable:	Sleep_efficiency	R-squared:	0.819			
Model:	OLS	Adj. R-squared:	0.816			
Method:	Least Squares	F-statistic:	275.3			
Date:	Thu, 02 Mar 2023	Prob (F-statistic):	1.02e-153			
Time:	22:40:32	Log-Likelihood:	628.67			
No. Observations:	434	AIC:	-1241.			
Df Residuals:	426	BIC:	-1209.			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0002	4.24e-06	56.915	0.000	0.000	0.000
Age	0.0006	0.000	3.018	0.003	0.000	0.001
REM_sleep_percentage	0.0113	0.001	17.879	0.000	0.010	0.013
Deep_sleep_percentage	0.0093	0.000	44.420	0.000	0.009	0.010
Light_sleep_percentage	0.0035	0.000	13.102	0.000	0.003	0.004
Awakenings	-0.0355	0.002	-15.822	0.000	-0.040	-0.031
Alcohol_consumption	-0.0030	0.002	-1.484	0.139	-0.007	0.001
Smoking_status	-0.0421	0.006	-7.045	0.000	-0.054	-0.030
Exercise_frequency	0.0049	0.002	2.401	0.017	0.001	0.009
Omnibus:	11.342	Durbin-Watson:	1.951			
Prob(Omnibus):	0.003	Jarque-Bera (JB):	10.501			
Skew:	-0.327	Prob(JB):	0.00524			
Kurtosis:	2.611	Cond. No.	3.26e+17			

OLS Regression Results

=====

Dep. Variable: Sleep_efficiency R-squared: 0.818

Model: OLS Adj. R-squared: 0.815

Method: Least Squares F-statistic: 319.9

Date: Thu, 02 Mar 2023 Prob (F-statistic): 1.64e-154

Time: 22:40:36 Log-Likelihood: 627.56

No. Observations: 434 AIC: -1241.

Df Residuals: 427 BIC: -1213.

Df Model: 6

Covariance Type: nonrobust
 =====

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0002	4.23e-06	56.881	0.000	0.000	0.000
Age	0.0006	0.000	2.874	0.004	0.000	0.001
REM_sleep_percentage	0.0114	0.001	17.928	0.000	0.010	0.013
Deep_sleep_percentage	0.0093	0.000	44.536	0.000	0.009	0.010
Light_sleep_percentage	0.0034	0.000	13.154	0.000	0.003	0.004
Awakenings	-0.0358	0.002	-16.071	0.000	-0.040	-0.031
Smoking_status	-0.0422	0.006	-7.040	0.000	-0.054	-0.030
Exercise_frequency	0.0045	0.002	2.233	0.026	0.001	0.008

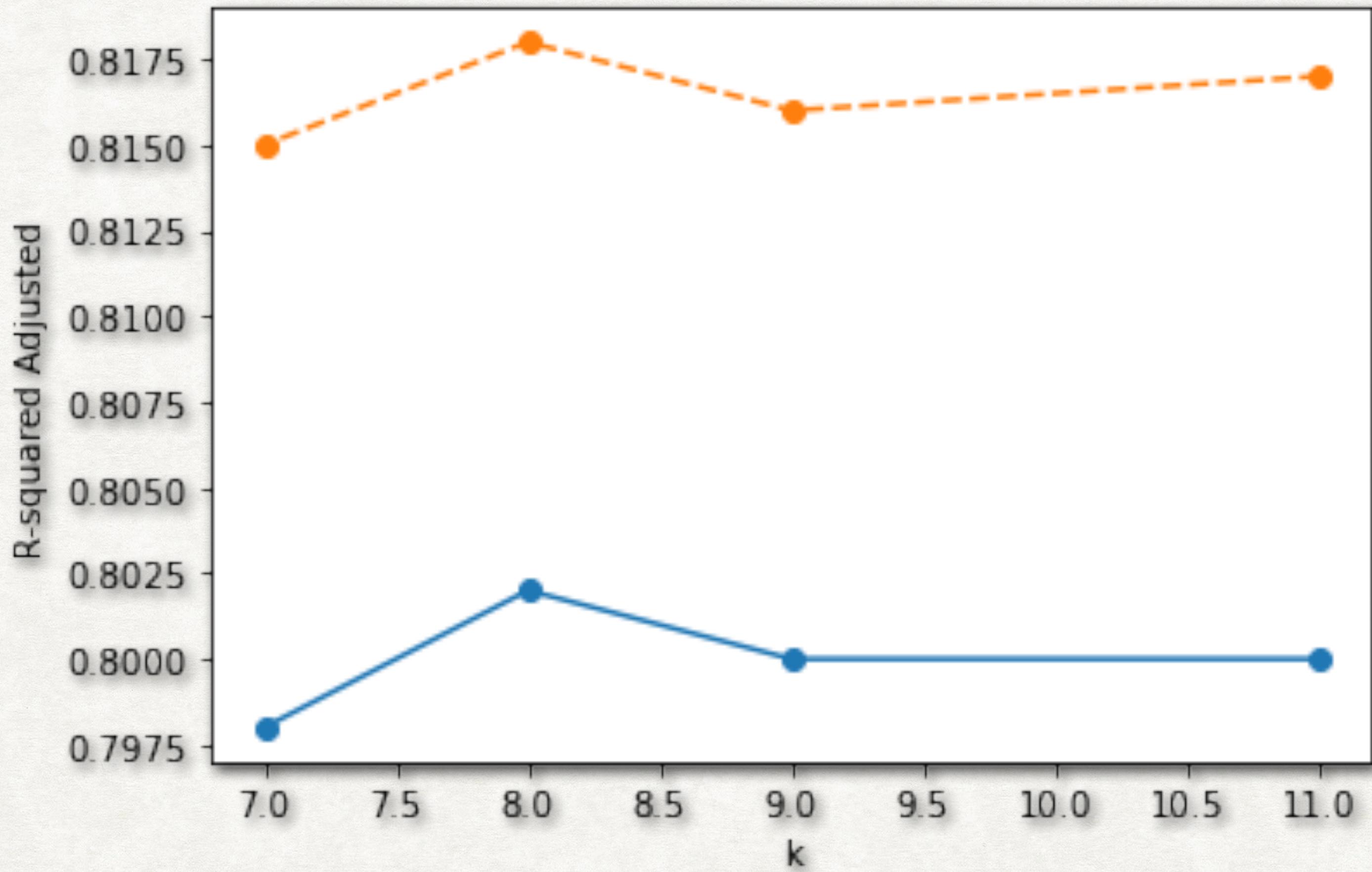
=====

Omnibus: 11.215 Durbin-Watson: 1.938

Prob(Omnibus): 0.004 Jarque-Bera (JB): 10.505

Skew: -0.331 Prob(JB): 0.00524

Kurtosis: 2.621 Cond. No. 3.27e+17
 =====



CONCLUSION

```
Age 0.00987329434169859
REM_sleep_percentage 0.005028312096599263
Deep_sleep_percentage 0.6228650215846004
Light_sleep_percentage 0.6774382093949173
Awakenings 0.33583160763021047
Alcohol_consumption 0.14699552384446923
Smoking_status 0.07489209563444288
Exercise_frequency 0.07939652809341702
best: ['Light_sleep_percentage', 0.6774382093949173]
```

<https://github.com/minesalum/msds-5509-final-project.git>