



به نام خدا

درس:

داده‌کاوی

استاد درس:

دکتر حکیم داودی

اعضای گروه:

فاطمه جان‌نثاری

حوری دهش

خرداد ۱۴۰۳

پیش‌بینی اختلال خواب
افراد
بر اساس سبک زندگی آنها



کیفیت خواب یکی از محصولات اصلی سبک زندگی سالم است.

ما حدود یک سوم از عمرمان را در خواب می‌گذرانیم.

خواب به شدت هم روی سلامت جسم و هم سلامت روان تاثیرگذار است.

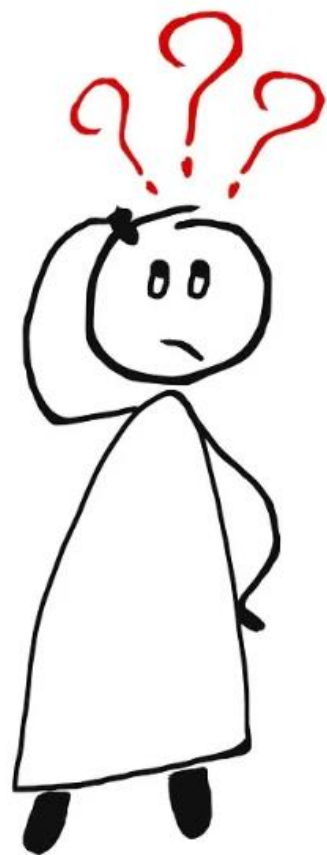
اختلالات خواب از جمله بی‌خوابی از مشکلاتی است که امروزه بسیار شایع شده و گریبان‌گیر بسیاری از مردم است.

مسئله این است که بفهمیم چه عواملی در سبک‌زندگی افراد روی کیفیت خواب و اختلال خواب آنها موثر است.

و می‌خواهیم بتوانیم با مشاهده سبک‌زندگی افراد میزان اختلال خواب آنها را بررسی کنیم.

سوال داده‌کاوی که برای حل مشکل مطرح شده است چیست؟

چگونه با استفاده از بررسی ویژگی‌های مختلف شخص که مربوط به سبک زندگی اوست بتوانیم اختلال خواب او را پیش‌بینی کنیم؟



داده‌های مسئله

RangeIndex: 374 entries, 0 to 373

Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	Person ID	374 non-null	int64
1	Gender	374 non-null	object
2	Age	374 non-null	int64
3	Occupation	374 non-null	object
4	Sleep Duration	374 non-null	float64
5	Quality of Sleep	374 non-null	int64
6	Physical Activity Level	374 non-null	int64
7	Stress Level	374 non-null	int64
8	BMI Category	374 non-null	object
9	Blood Pressure	374 non-null	object
10	Heart Rate	374 non-null	int64
11	Daily Steps	374 non-null	int64
12	Sleep Disorder	155 non-null	object

dtypes: float64(1), int64(7), object(5)

memory usage: 38.1+ KB

ستون‌ها:

سن

شغل

مدت زمان خواب

کیفیت خواب

سطح فعالیت فیزیکی

سطح استرس

دسته‌ی BMI

فشار خون

ضربان قلب

گام‌های روزانه

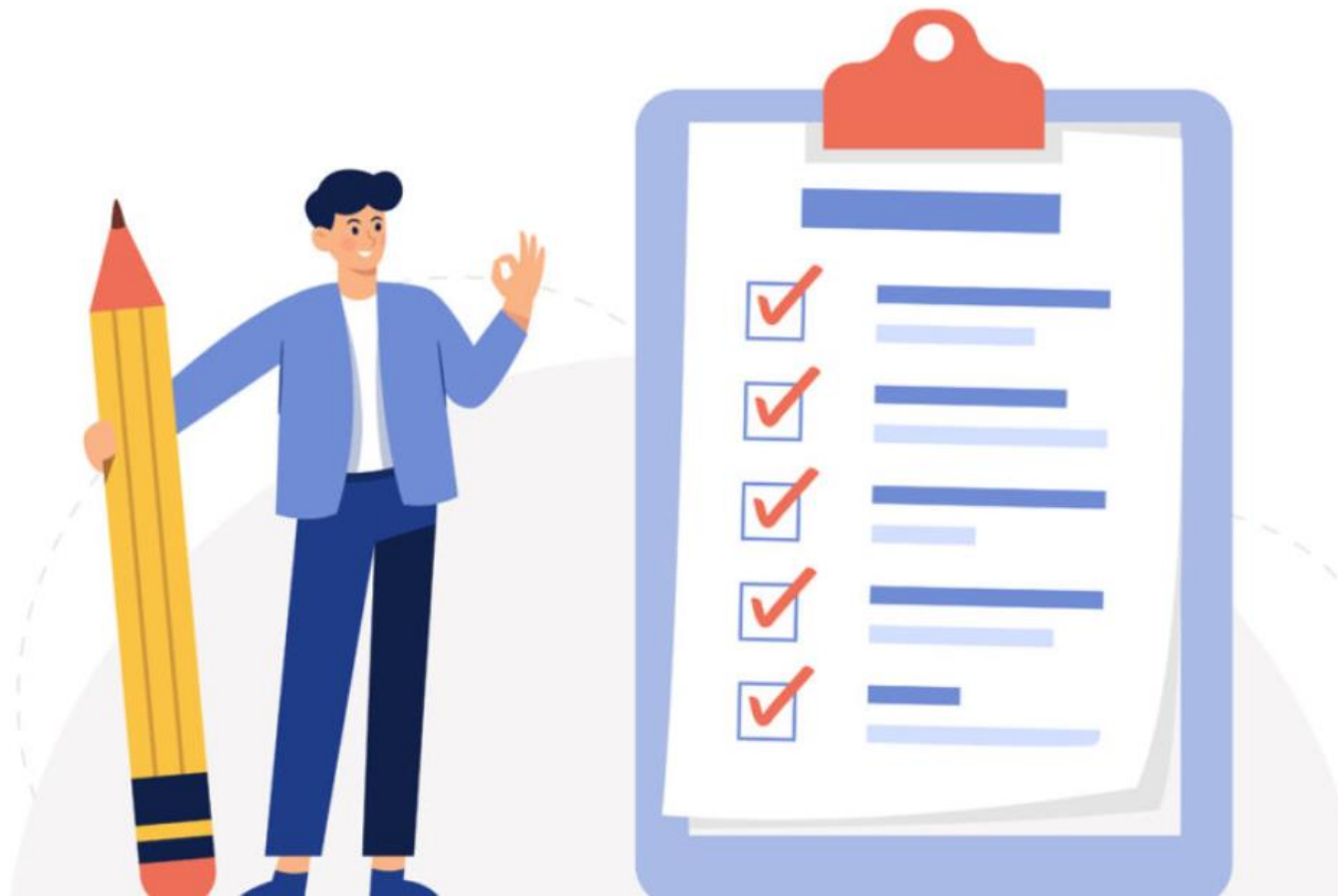
اختلال خواب

دیتاست:

شامل ۱۳ ستون و ۳۷۴ سطر

	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
0	1	Male	27	Software Engineer	6.100000	6	42	6	Overweight	126/83	77	4200	nan
1	2	Male	28	Doctor	6.200000	6	60	8	Normal	125/80	75	10000	nan
2	3	Male	28	Doctor	6.200000	6	60	8	Normal	125/80	75	10000	nan
3	4	Male	28	Sales Representative	5.900000	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.900000	4	30	8	Obese	140/90	85	3000	Sleep Apnea

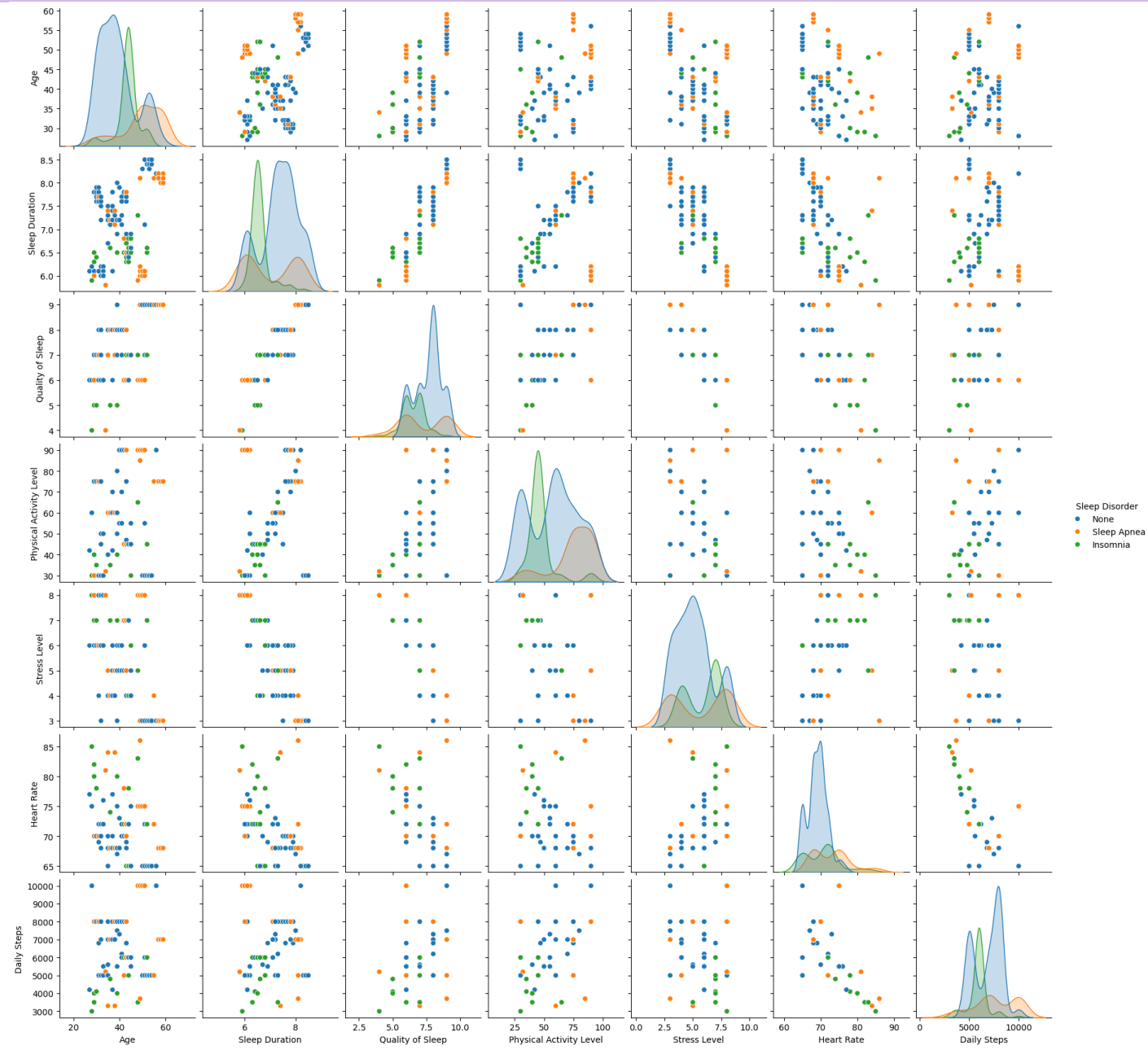
باید یک مدل می‌ساختیم که بتواند روی داده‌های ما آموزش ببیند و بعد از اتمام یادگیری بتواند اختلال خواب افراد را پیش‌بینی کند.



	Person ID	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate	Daily Steps
count	374.000000	374.000000	374.000000	374.000000	374.000000	374.000000	374.000000	374.000000
mean	187.500000	42.184492	7.132086	7.312834	59.171123	5.385027	70.165775	6816.844920
std	108.108742	8.673133	0.795657	1.196956	20.830804	1.774526	4.135676	1617.915679
min	1.000000	27.000000	5.800000	4.000000	30.000000	3.000000	65.000000	3000.000000
25%	94.250000	35.250000	6.400000	6.000000	45.000000	4.000000	68.000000	5600.000000
50%	187.500000	43.000000	7.200000	7.000000	60.000000	5.000000	70.000000	7000.000000
75%	280.750000	50.000000	7.800000	8.000000	75.000000	7.000000	72.000000	8000.000000
max	374.000000	59.000000	8.500000	9.000000	90.000000	8.000000	86.000000	10000.000000

	Gender	Occupation	BMI Category	Blood Pressure	Sleep Disorder
count	374	374	374	374	155
unique	2	11	4	25	2
top	Male	Nurse	Normal	130/85	Sleep Apnea
freq	189	73	195	99	78

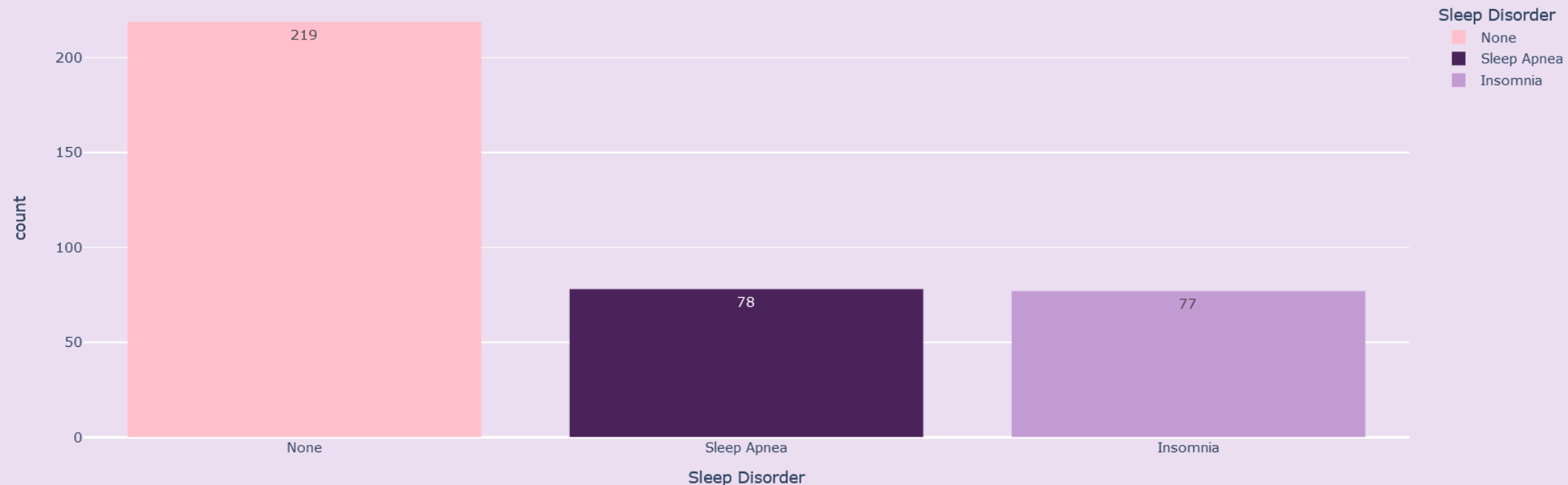
تجزیه و تحلیل داده‌های اکتشافی: نمودار Correlation



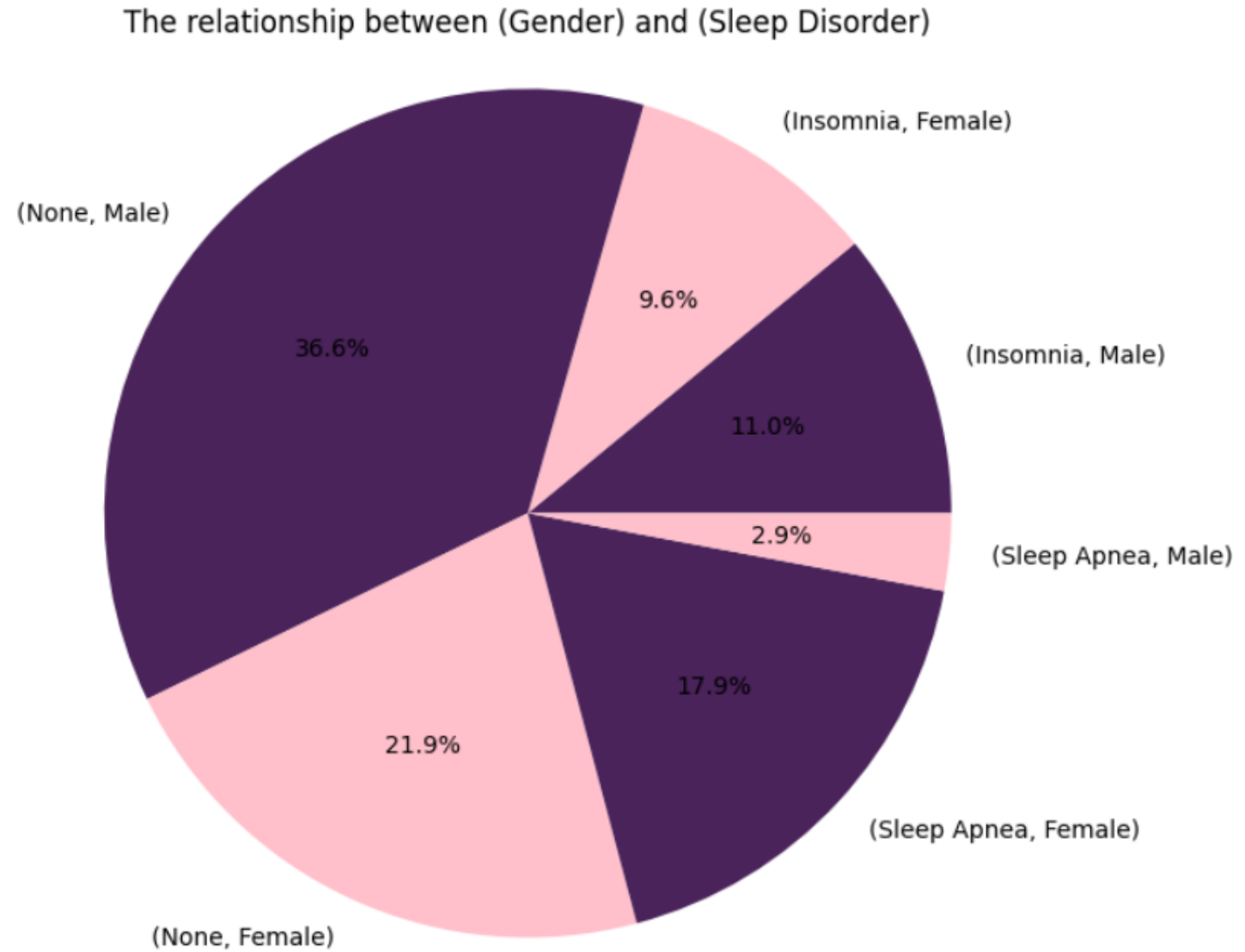
تجزیه و تحلیل داده‌های اکتشافی: انواع اختلال خواب

- هیچ: فرد هیچ اختلال خواب خاصی را نشان نمی دهد
- آپنه خواب: فرد در طول خواب از وقفه تنفسی رنج می برد که منجر به اختلال در الگوهای خواب و خطرات بالقوه سلامتی می شود.
- بی خوابی: فرد در به خواب رفتن یا به خواب ماندن دچار مشکل می شود که منجر به خواب ناکافی یا بی کیفیت می شود.

Distribution of persons have sleep disorder or not..



تجزیه و تحلیل داده‌های اکتشافی: رابطه جنسیت و اختلال خواب



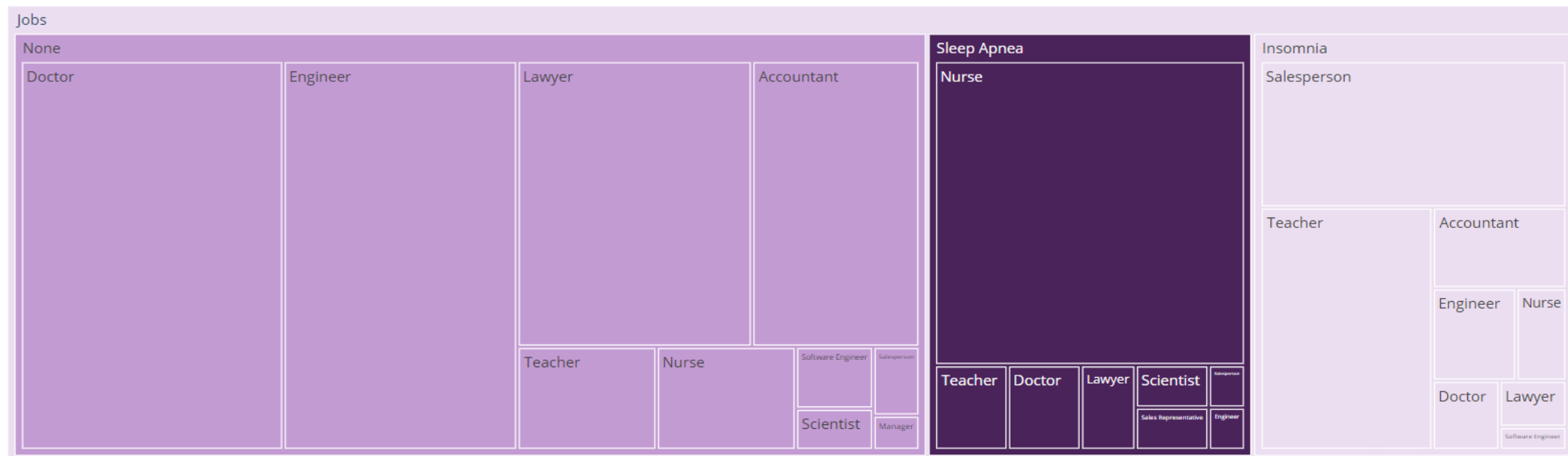
تجزیه و تحلیل داده‌های اکتشافی: رابطه شغل و اختلال خواب

Doctor	64
Engineer	57
Lawyer	42
Accountant	30
Nurse	9
Teacher	9
Software Engineer	3
Salesperson	2
Scientist	2
Manager	1

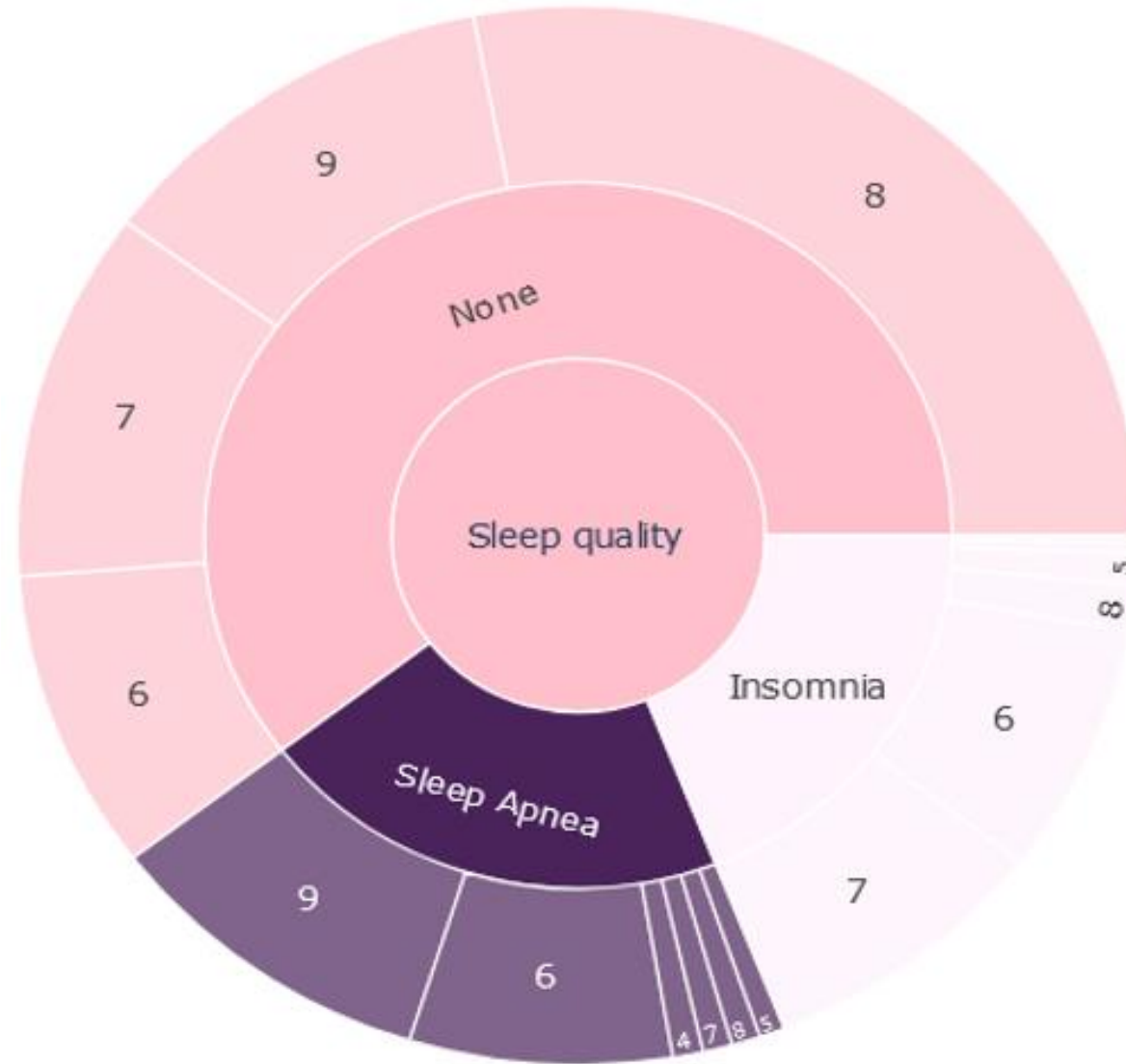
Nurse	61
Doctor	4
Teacher	4
Lawyer	3
Sales Representative	2
Scientist	2
Engineer	1
Salesperson	1

Occupation	
Salesperson	29
Teacher	27
Accountant	7
Engineer	5
Nurse	3
Doctor	3
Lawyer	2
Software Engineer	1

The effect of job on sleep..



تجزیه و تحلیل داده‌های اکتشافی: رابطه کیفیت خواب و اختلال خواب

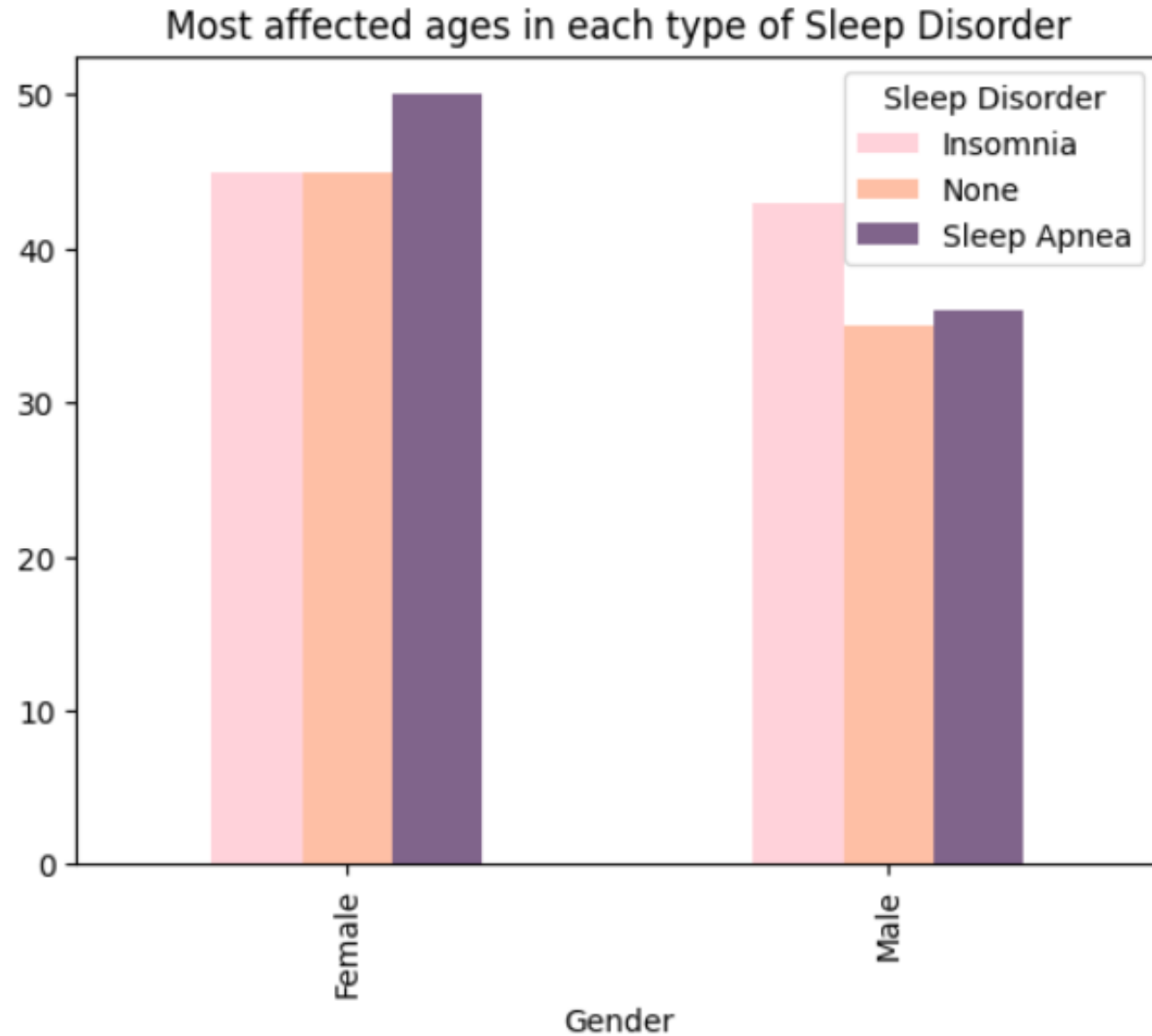


تجزیه و تحلیل داده‌های اکتشافی: رابطه فعالیت فیزیکی و اختلال خواب

The effect of activities on sleep ..

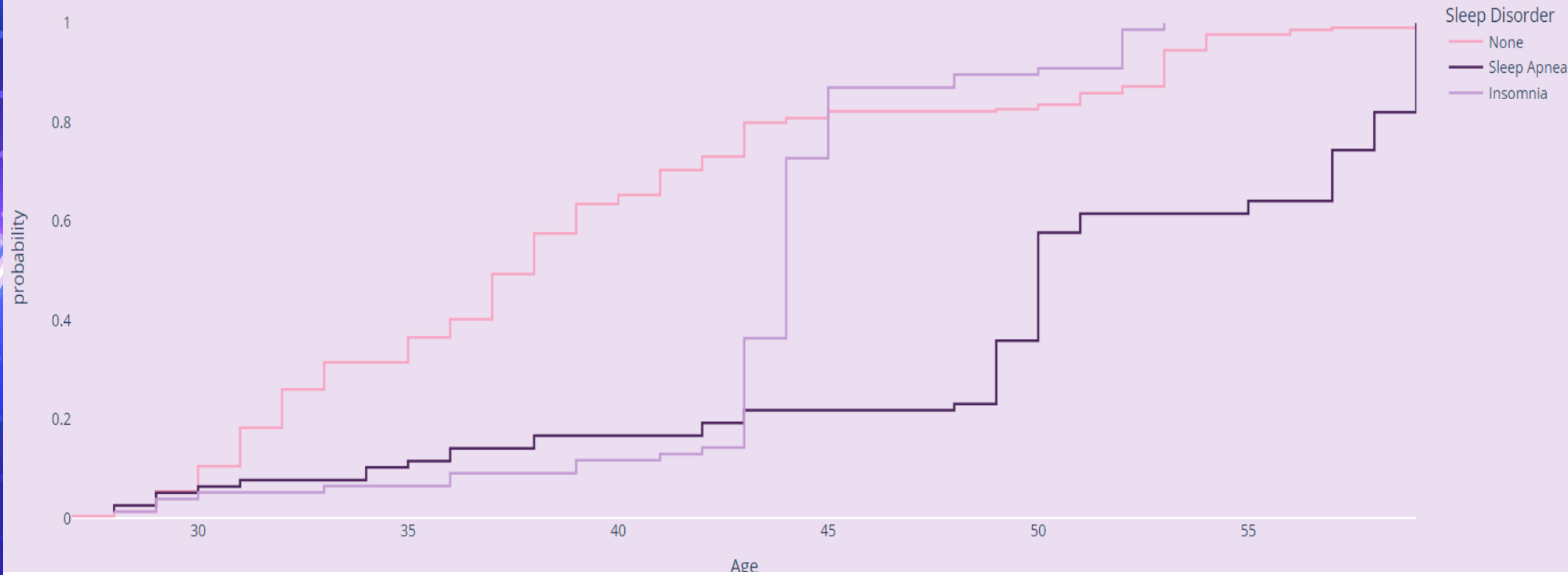


تجزیه و تحلیل داده‌های اکتشافی: رابطه سن و جنسیت و اختلال خواب



تجزیه و تحلیل داده‌های اکتشافی: رابطه سن و اختلال خواب

The effect of ages on sleep ..

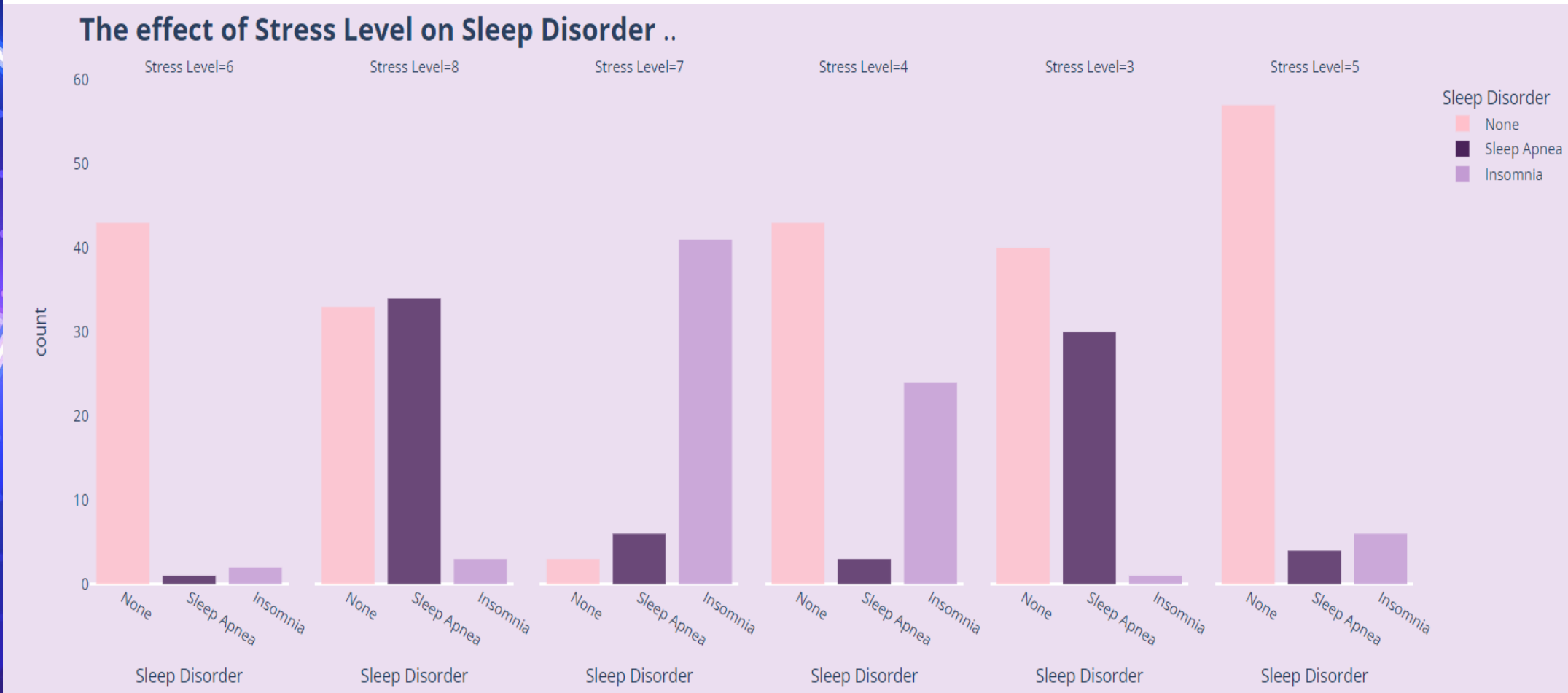


تجزیه و تحلیل داده‌های اکتشافی: رابطه طول مدت خواب و اختلال خواب

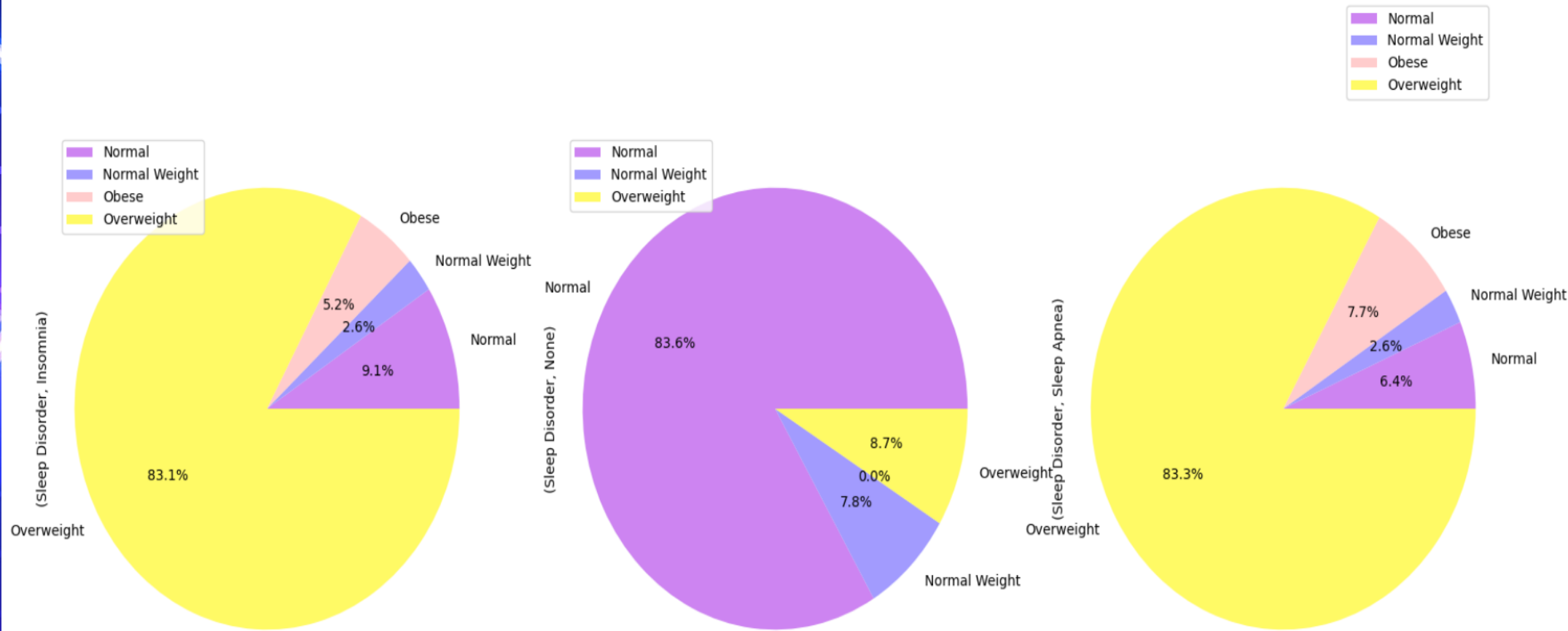
The Relationship Between Sleep Duration and Sleep Disorder



تجزیه و تحلیل داده‌های اکتشافی: رابطه سطح استرس و اختلال خواب



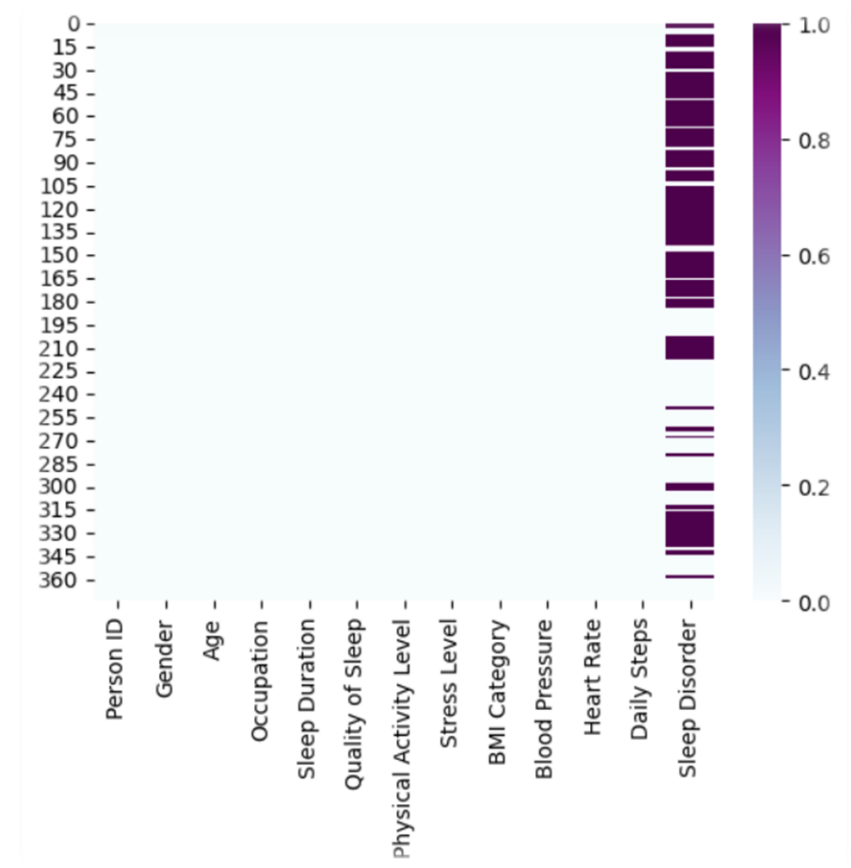
تجزیه و تحلیل داده‌های اکتشافی: رابطه BMI و اختلال خواب



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

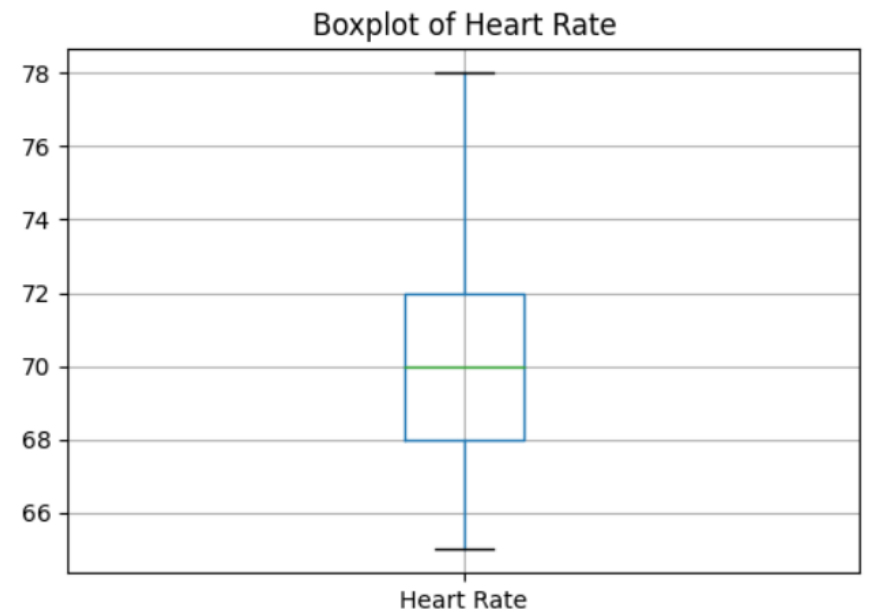
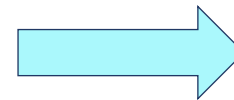
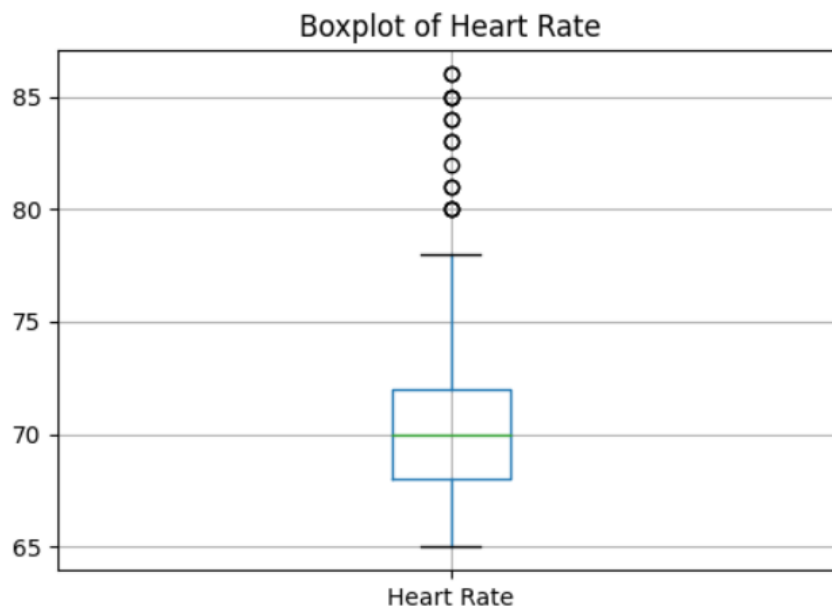
۱. پیدا کردن داده‌های nan

- نمایش دادن داده‌های nan



۲. تشخیص داده‌های outlier:

- استفاده از باکس‌پلات و پیدا کردن outlierها
- حذف داده‌های outlier



۳. تبدیل اطلاعات مربوط به فشار خون به دو دسته نرمال و غیرنرمال:

- فشار خون ایده آل سیستولیک (عدد بالا): کمتر از ۱۲۰، دیاستولیک (عدد پایین): کمتر از ۸۰
- سیستولیک طبیعی (عدد بالا): در محدوده (۱۲۰ - ۱۲۹)، دیاستولیک (عدد پایین): در محدوده (۸۰ - ۸۴)
- در غیر این صورت فشار خون بالاست

```
array(['126/83', '125/80', '120/80', '130/86', '117/76', '118/76',  
      '128/85', '128/84', '115/75', '129/84', '130/85', '115/78',  
      '119/77', '121/79', '125/82', '135/90', '122/80', '140/95',  
      '118/75'], dtype=object)
```

```
cleaned_data['Blood Pressure'] = cleaned_data['Blood Pressure'].apply(lambda x: 0 if x in ['120/80', '126/83', '125/80', '128/84',  
                                                                                          '129/84', '117/76', '118/76', '115/75', '125/82', '122/80'] else 1)  
  
# 0 = normal blood pressure  
# 1 = abnormal blood pressure
```

۴. تقسیم متغیرهای پیوسته به دسته‌های گسسته:

```
# Binning (dividing continuous variable into discrete intervals or categories)
cleaned_data["Age"] = pd.cut(cleaned_data["Age"],2)
cleaned_data["Heart Rate"] = pd.cut(cleaned_data["Heart Rate"],4)
cleaned_data["Daily Steps"] = pd.cut(cleaned_data["Daily Steps"],4)
cleaned_data["Sleep Duration"] = pd.cut(cleaned_data["Sleep Duration"],3)
cleaned_data["Physical Activity Level"] = pd.cut(cleaned_data["Physical Activity Level"],4)
```

۵. تبدیل داده‌های categorical به فرمت عددی:

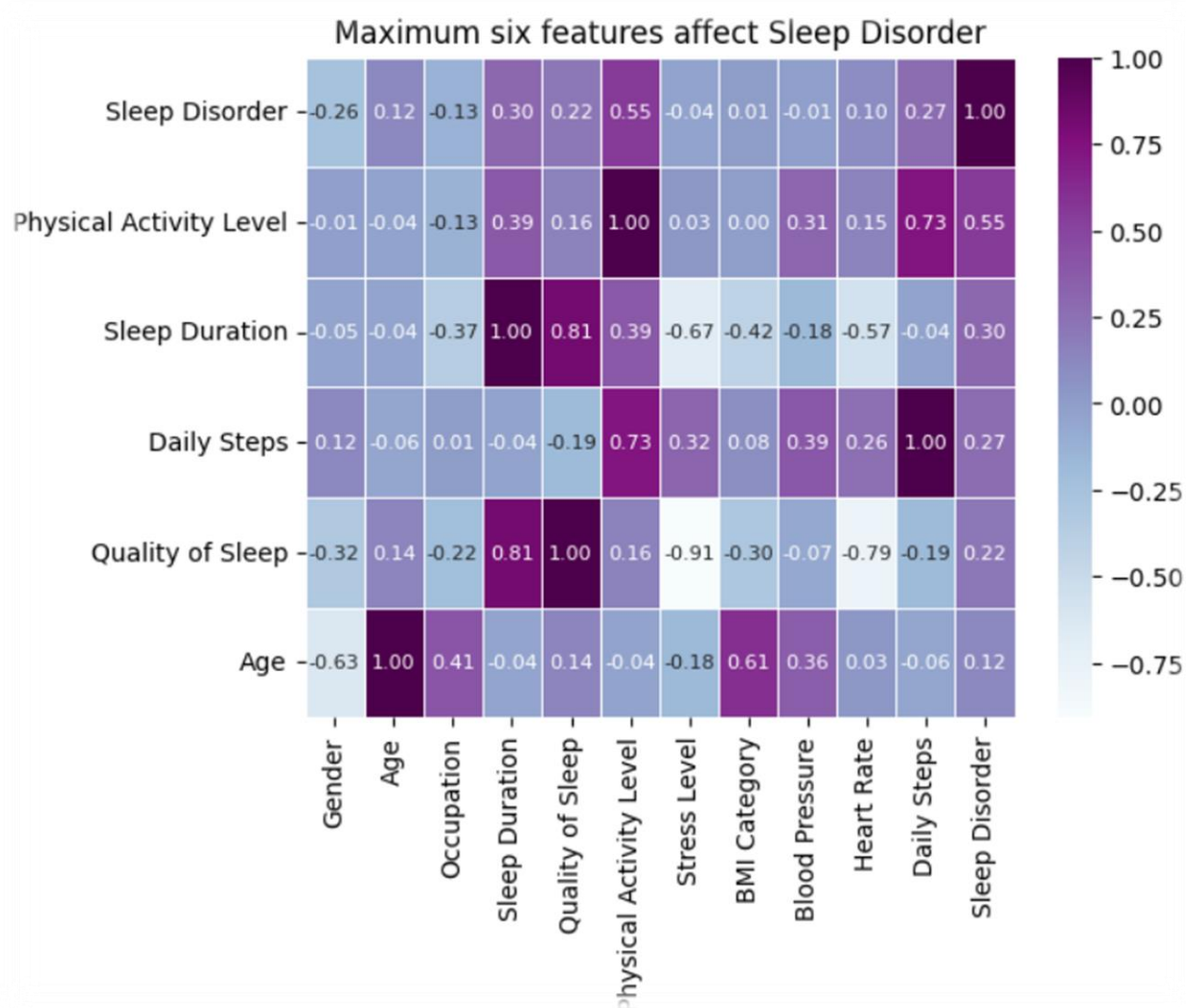
```
# convert categorical data into numerical format
LE = LabelEncoder()

categories=['Gender','Age','Occupation','Sleep Duration','Physical Activity Level','BMI Category','Heart Rate','Daily Steps','Sleep Disorder']
for label in categories:
    cleaned_data[label]=LE.fit_transform(cleaned_data[label])
```

```
# remove Person ID
cleaned_data.drop(['Person ID'], axis=1, inplace=True)
```

۶. حذف ویژگی Person ID:

۷. تحلیل و بررسی همبستگی بین متغیرها در داده‌های تمیز شده و شناسایی متغیرهایی که بیشترین تأثیر را بر اختلال خواب دارند:



- استفاده از روش Cross validation

```
Fold:1, Train set: 323, Test set:36  
Fold:2, Train set: 323, Test set:36  
Fold:3, Train set: 323, Test set:36  
Fold:4, Train set: 323, Test set:36  
Fold:5, Train set: 323, Test set:36  
Fold:6, Train set: 323, Test set:36  
Fold:7, Train set: 323, Test set:36  
Fold:8, Train set: 323, Test set:36  
Fold:9, Train set: 323, Test set:36  
Fold:10, Train set: 324, Test set:35
```


گام های طی شده در مدل سازی:

☐ انتخاب مدل

- Decision Tree
- Random Forest
- Gradient Boosting Classifier
- Extra Tree

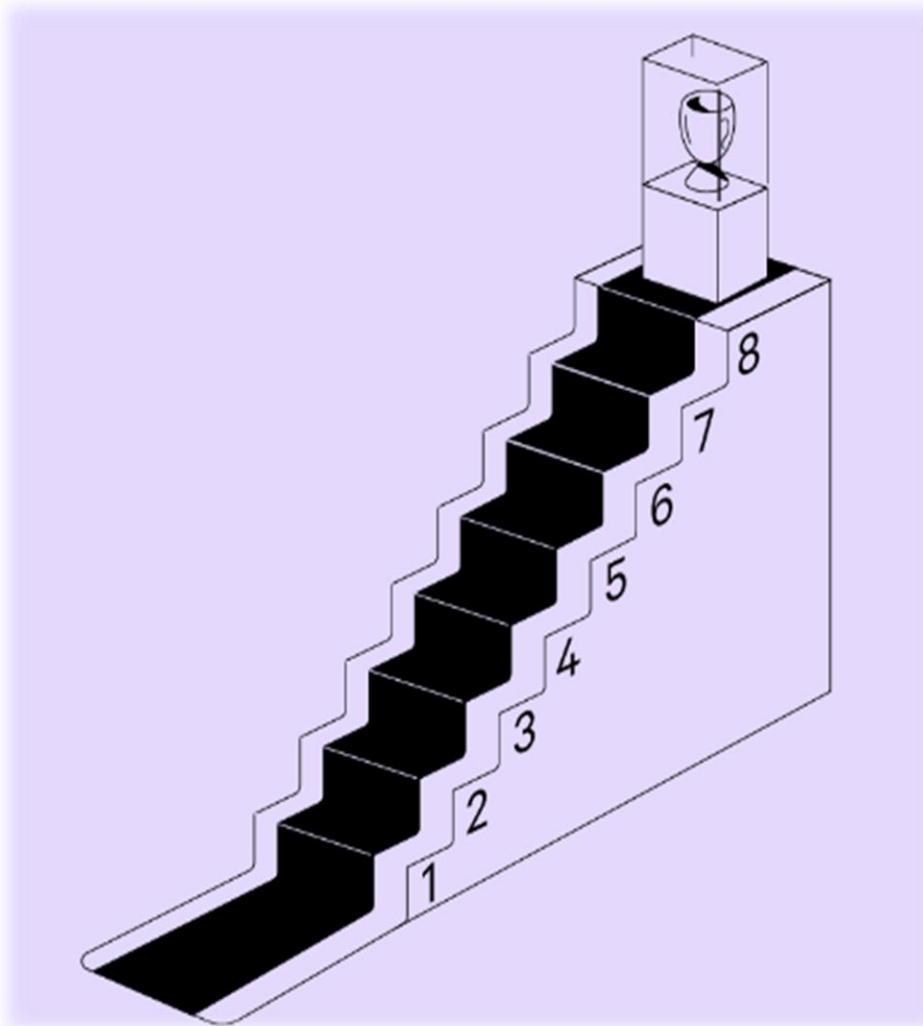
☐ انتخاب بهترین fold

☐ Hyperparameter Tunning

- Grid Search

☐ ساختن مدل نهایی با tune کردن

☐ ارزیابی مدل ها



```
Fold 1 - Train set Accuracy: 0.9255952380952381
Fold 1 - Test set Accuracy: 0.9473684210526315
```

```
=====
```

```
Fold 2 - Train set Accuracy: 0.9375
Fold 2 - Test set Accuracy: 0.7894736842105263
```

```
Overfitting detected in fold 2
```

```
=====
```

```
Fold 3 - Train set Accuracy: 0.9315476190476191
Fold 3 - Test set Accuracy: 0.9210526315789473
```

```
=====
```

```
Fold 4 - Train set Accuracy: 0.9345238095238095
Fold 4 - Test set Accuracy: 0.868421052631579
```

```
Overfitting detected in fold 4
```

```
=====
```

```
Fold 5 - Train set Accuracy: 0.9436201780415431
Fold 5 - Test set Accuracy: 0.8108108108108109
```

```
Overfitting detected in fold 5
```

```
Fold 6 - Train set Accuracy: 0.9228486646884273
Fold 6 - Test set Accuracy: 1.0
```

```
Overfitting detected in fold 6
```

```
=====
```

```
Fold 7 - Train set Accuracy: 0.9258160237388724
Fold 7 - Test set Accuracy: 0.918918918918919
```

```
=====
```

```
Fold 8 - Train set Accuracy: 0.9258160237388724
Fold 8 - Test set Accuracy: 0.972972972972973
```

```
=====
```

```
Fold 9 - Train set Accuracy: 0.9376854599406528
Fold 9 - Test set Accuracy: 0.8378378378378378
```

```
Overfitting detected in fold 9
```

```
=====
```

```
Fold 10 - Train set Accuracy: 0.9287833827893175
Fold 10 - Test set Accuracy: 0.918918918918919
```

The best fold is Fold 7 with a train accuracy of 0.9258160237388724 and a test accuracy of 0.918918918918919.

Report:

	precision	recall	f1-score	support
0	0.75	0.86	0.80	7
1	0.86	1.00	0.92	6
2	1.00	0.92	0.96	24
accuracy			0.92	37
macro avg	0.87	0.92	0.89	37
weighted avg	0.93	0.92	0.92	37

مدل DT با tune

```
Fold 1 - Train set Accuracy: 0.9196428571428571
Fold 1 - Test set Accuracy: 0.9736842105263158
```

```
Overfitting detected in fold 1
```

```
=====

Fold 2 - Train set Accuracy: 0.9345238095238095
Fold 2 - Test set Accuracy: 0.8157894736842105
```

```
Overfitting detected in fold 2
```

```
=====

Fold 3 - Train set Accuracy: 0.9255952380952381
Fold 3 - Test set Accuracy: 0.9210526315789473
```

```
=====

Fold 4 - Train set Accuracy: 0.9285714285714286
Fold 4 - Test set Accuracy: 0.868421052631579
```

```
Overfitting detected in fold 4
```

```
=====

Fold 5 - Train set Accuracy: 0.9376854599406528
Fold 5 - Test set Accuracy: 0.8108108108108109
```

```
Overfitting detected in fold 5
```

```
Fold 6 - Train set Accuracy: 0.9198813056379822
Fold 6 - Test set Accuracy: 0.972972972972973
```

```
Overfitting detected in fold 6
```

```
=====

Fold 7 - Train set Accuracy: 0.9258160237388724
Fold 7 - Test set Accuracy: 0.9459459459459459
```

```
=====

Fold 8 - Train set Accuracy: 0.9169139465875371
Fold 8 - Test set Accuracy: 0.9459459459459459
```

```
=====

Fold 9 - Train set Accuracy: 0.9287833827893175
Fold 9 - Test set Accuracy: 0.8378378378378378
```

```
Overfitting detected in fold 9
```

```
=====

Fold 10 - Train set Accuracy: 0.9258160237388724
Fold 10 - Test set Accuracy: 0.9459459459459459
```

The best fold is Fold 3 with a train accuracy of 0.9255952380952381 and a test accuracy of 0.9210526315789473.

Report:

	precision	recall	f1-score	support
0	0.86	0.86	0.86	7
1	1.00	0.82	0.90	11
2	0.91	1.00	0.95	20
accuracy			0.92	38
macro avg	0.92	0.89	0.90	38
weighted avg	0.93	0.92	0.92	38



	Model	Best fold	Train_accuracy	Test_accuracy	Cross Val Score	Difference Train & Test	precision_score	recall_score	f1_score	Description
0	DT	7	0.925816	0.918919	0.898578	0.006897	0.869048	0.924603	0.893200	
1	DT2	3	0.925595	0.921053	0.903841	0.004543	0.922078	0.891775	0.903175	

Fold 1 - Train set Accuracy: 0.9255952380952381
Fold 1 - Test set Accuracy: 0.9736842105263158

=====

Fold 2 - Train set Accuracy: 0.9375
Fold 2 - Test set Accuracy: 0.7894736842105263

Overfitting detected in fold 2

=====

Fold 3 - Train set Accuracy: 0.9315476190476191
Fold 3 - Test set Accuracy: 0.9210526315789473

=====

Fold 4 - Train set Accuracy: 0.9345238095238095
Fold 4 - Test set Accuracy: 0.8947368421052632

=====

Fold 5 - Train set Accuracy: 0.9436201780415431
Fold 5 - Test set Accuracy: 0.8108108108108109

Overfitting detected in fold 5

Fold 6 - Train set Accuracy: 0.9228486646884273
Fold 6 - Test set Accuracy: 1.0

Overfitting detected in fold 6

=====

Fold 7 - Train set Accuracy: 0.9258160237388724
Fold 7 - Test set Accuracy: 0.972972972972973

=====

Fold 8 - Train set Accuracy: 0.9258160237388724
Fold 8 - Test set Accuracy: 0.972972972972973

=====

Fold 9 - Train set Accuracy: 0.9376854599406528
Fold 9 - Test set Accuracy: 0.8648648648648649

Overfitting detected in fold 9

=====

Fold 10 - Train set Accuracy: 0.9287833827893175
Fold 10 - Test set Accuracy: 0.9459459459459459

The best fold is Fold 3 with a train accuracy of 0.9315476190476191 and a test accuracy of 0.9210526315789473.

Report:

	precision	recall	f1-score	support
0	0.86	0.86	0.86	7
1	1.00	0.82	0.90	11
2	0.91	1.00	0.95	20
accuracy			0.92	38
macro avg	0.92	0.89	0.90	38
weighted avg	0.93	0.92	0.92	38

مدل RF با tune

```
Fold 1 - Train set Accuracy: 0.9226190476190477
Fold 1 - Test set Accuracy: 0.9736842105263158
```

```
Overfitting detected in fold 1
```

```
=====

Fold 2 - Train set Accuracy: 0.9345238095238095
Fold 2 - Test set Accuracy: 0.7894736842105263
```

```
Overfitting detected in fold 2
```

```
=====

Fold 3 - Train set Accuracy: 0.9255952380952381
Fold 3 - Test set Accuracy: 0.9210526315789473
```

```
=====

Fold 4 - Train set Accuracy: 0.9255952380952381
Fold 4 - Test set Accuracy: 0.868421052631579
```

```
Overfitting detected in fold 4
```

```
=====

Fold 5 - Train set Accuracy: 0.9347181008902077
Fold 5 - Test set Accuracy: 0.7837837837837838
```

```
Overfitting detected in fold 5
```

```
Fold 6 - Train set Accuracy: 0.913946587537092
Fold 6 - Test set Accuracy: 0.9459459459459459
```

```
=====

Fold 7 - Train set Accuracy: 0.9228486646884273
Fold 7 - Test set Accuracy: 0.972972972972973
```

```
Overfitting detected in fold 7
```

```
=====

Fold 8 - Train set Accuracy: 0.9228486646884273
Fold 8 - Test set Accuracy: 0.972972972972973
```

```
Overfitting detected in fold 8
```

```
=====

Fold 9 - Train set Accuracy: 0.9287833827893175
Fold 9 - Test set Accuracy: 0.8378378378378378
```

```
Overfitting detected in fold 9
```

```
=====

Fold 10 - Train set Accuracy: 0.9228486646884273
Fold 10 - Test set Accuracy: 0.918918918918919
```

The best fold is Fold 10 with a train accuracy of 0.9228486646884273 and a test accuracy of 0.918918918918919.

Report:

	precision	recall	f1-score	support
0	0.89	1.00	0.94	8
1	0.75	0.86	0.80	7
2	1.00	0.91	0.95	22
accuracy			0.92	37
macro avg	0.88	0.92	0.90	37
weighted avg	0.93	0.92	0.92	37

	Model	Best fold	Train_accuracy	Test_accuracy	Cross Val Score	Difference Train & Test	precision_score	recall_score	f1_score	Description
0	RF	3	0.931548	0.921053	0.914651	0.010495	0.922078	0.891775	0.903175	
1	RF2	10	0.922849	0.918919	0.898506	0.003930	0.879630	0.922078	0.897852	



Fold 1 - Train set Accuracy: 0.9315476190476191
Fold 1 - Test set Accuracy: 0.9210526315789473

=====

Fold 2 - Train set Accuracy: 0.9226190476190477
Fold 2 - Test set Accuracy: 1.0

Overfitting detected in fold 2

=====

Fold 3 - Train set Accuracy: 0.9345238095238095
Fold 3 - Test set Accuracy: 0.868421052631579

Overfitting detected in fold 3

=====

Fold 4 - Train set Accuracy: 0.9315476190476191
Fold 4 - Test set Accuracy: 0.9210526315789473

=====

Fold 5 - Train set Accuracy: 0.9287833827893175
Fold 5 - Test set Accuracy: 0.9459459459459459

=====

Fold 6 - Train set Accuracy: 0.9287833827893175
Fold 6 - Test set Accuracy: 0.9459459459459459

=====

Fold 7 - Train set Accuracy: 0.9347181008902077
Fold 7 - Test set Accuracy: 0.8918918918918919

=====

Fold 8 - Train set Accuracy: 0.9317507418397626
Fold 8 - Test set Accuracy: 0.918918918918919

=====

Fold 9 - Train set Accuracy: 0.9376854599406528
Fold 9 - Test set Accuracy: 0.8648648648648649

Overfitting detected in fold 9

=====

Fold 10 - Train set Accuracy: 0.9287833827893175
Fold 10 - Test set Accuracy: 0.9459459459459459

The best fold is Fold 1 with a train accuracy of 0.9315476190476191 and a test accuracy of 0.9210526315789473.

Report:

	precision	recall	f1-score	support
0	1.00	0.73	0.84	11
1	0.90	1.00	0.95	9
2	0.90	1.00	0.95	18
accuracy			0.92	38
macro avg	0.93	0.91	0.91	38
weighted avg	0.93	0.92	0.92	38

Fold 1 - Train set Accuracy: 0.9315476190476191
Fold 1 - Test set Accuracy: 0.9210526315789473

=====

Fold 2 - Train set Accuracy: 0.9404761904761905
Fold 2 - Test set Accuracy: 0.8421052631578947

Overfitting detected in fold 2

=====

Fold 3 - Train set Accuracy: 0.9315476190476191
Fold 3 - Test set Accuracy: 0.9210526315789473

=====

Fold 4 - Train set Accuracy: 0.9315476190476191
Fold 4 - Test set Accuracy: 0.8157894736842105

Overfitting detected in fold 4

=====

Fold 5 - Train set Accuracy: 0.9317507418397626
Fold 5 - Test set Accuracy: 0.918918918918919

=====

The best fold is Fold 9 with a train accuracy of 0.9258160237388724 and a test accuracy of 0.918918918918919.

Fold 6 - Train set Accuracy: 0.9347181008902077
Fold 6 - Test set Accuracy: 0.8918918918918919

=====

Fold 7 - Train set Accuracy: 0.9317507418397626
Fold 7 - Test set Accuracy: 0.8648648648648649

Overfitting detected in fold 7

=====

Fold 8 - Train set Accuracy: 0.9287833827893175
Fold 8 - Test set Accuracy: 0.9459459459459459

=====

Fold 9 - Train set Accuracy: 0.9258160237388724
Fold 9 - Test set Accuracy: 0.918918918918919

=====

Fold 10 - Train set Accuracy: 0.9258160237388724
Fold 10 - Test set Accuracy: 0.972972972972973

Report:

	precision	recall	f1-score	support
0	0.86	0.86	0.86	7
1	0.89	0.89	0.89	9
2	0.95	0.95	0.95	21
accuracy			0.92	37
macro avg	0.90	0.90	0.90	37
weighted avg	0.92	0.92	0.92	37

مقایسه GB با GB2

	Model	Best fold	Train_accuracy	Test_accuracy	Cross Val Score	Difference Train & Test	precision_score	recall_score	f1_score	Description
0	GB	1	0.931548	0.921053	0.908962	0.010495	0.933333	0.909091	0.912281	
1	GB2	9	0.925816	0.918919	0.919915	0.006897	0.899471	0.899471	0.899471	




```

Fold 1 - Train set Accuracy: 0.9315476190476191
Fold 1 - Test set Accuracy: 0.9210526315789473

=====

Fold 2 - Train set Accuracy: 0.9285714285714286
Fold 2 - Test set Accuracy: 0.8947368421052632

=====

Fold 3 - Train set Accuracy: 0.9345238095238095
Fold 3 - Test set Accuracy: 0.8947368421052632

=====

Fold 4 - Train set Accuracy: 0.9285714285714286
Fold 4 - Test set Accuracy: 0.8947368421052632

=====

Fold 5 - Train set Accuracy: 0.9258160237388724
Fold 5 - Test set Accuracy: 0.972972972972973

=====

Fold 6 - Train set Accuracy: 0.9287833827893175
Fold 6 - Test set Accuracy: 0.918918918918919

=====

Fold 7 - Train set Accuracy: 0.9258160237388724
Fold 7 - Test set Accuracy: 0.972972972972973

=====

Fold 8 - Train set Accuracy: 0.9317507418397626
Fold 8 - Test set Accuracy: 0.918918918918919

=====

Fold 9 - Train set Accuracy: 0.9406528189910979
Fold 9 - Test set Accuracy: 0.8378378378378378

Overfitting detected in fold 9

=====

Fold 10 - Train set Accuracy: 0.9376854599406528
Fold 10 - Test set Accuracy: 0.8648648648648649

Overfitting detected in fold 10

=====

The best fold is Fold 6 with a train accuracy of 0.9287833827893175 and a test accuracy of 0.918918918918919.

```

Report:					
	precision	recall	f1-score	support	
0	0.75	0.86	0.80	7	
1	0.88	0.88	0.88	8	
2	1.00	0.95	0.98	22	
accuracy			0.92	37	
macro avg	0.88	0.90	0.88	37	
weighted avg	0.93	0.92	0.92	37	

مدل ET با tune

Fold 1 - Train set Accuracy: 0.9315476190476191
Fold 1 - Test set Accuracy: 0.868421052631579

Overfitting detected in fold 1

Fold 2 - Train set Accuracy: 0.9285714285714286
Fold 2 - Test set Accuracy: 0.9473684210526315

Fold 3 - Train set Accuracy: 0.9315476190476191
Fold 3 - Test set Accuracy: 0.9210526315789473

Fold 4 - Train set Accuracy: 0.9285714285714286
Fold 4 - Test set Accuracy: 0.9473684210526315

Fold 5 - Train set Accuracy: 0.9317507418397626
Fold 5 - Test set Accuracy: 0.8918918918918919

Fold 6 - Train set Accuracy: 0.9287833827893175
Fold 6 - Test set Accuracy: 0.9459459459459459

Fold 7 - Train set Accuracy: 0.9258160237388724
Fold 7 - Test set Accuracy: 0.918918918918919

Fold 8 - Train set Accuracy: 0.9376854599406528
Fold 8 - Test set Accuracy: 0.8648648648648649

Overfitting detected in fold 8

Fold 9 - Train set Accuracy: 0.9317507418397626
Fold 9 - Test set Accuracy: 0.918918918918919

Fold 10 - Train set Accuracy: 0.9317507418397626
Fold 10 - Test set Accuracy: 0.918918918918919

The best fold is Fold 7 with a train accuracy of 0.9258160237388724 and a test accuracy of 0.918918918918919.

Report:

	precision	recall	f1-score	support
0	0.67	1.00	0.80	6
1	1.00	0.75	0.86	12
2	1.00	1.00	1.00	19
accuracy			0.92	37
macro avg	0.89	0.92	0.89	37
weighted avg	0.95	0.92	0.92	37

	Model	Best fold	Train_accuracy	Test_accuracy	Cross Val Score	Difference Train & Test	precision_score	recall_score	f1_score	Description
0	ET	6	0.928783	0.918919	0.908962	0.009864	0.875000	0.895563	0.883915	
1	ET2	7	0.925816	0.918919	0.911664	0.006897	0.888889	0.916667	0.885714	



	Model	Train_accuracy	Test_accuracy	precision_score	recall_score	f1_score	Description
0	DT	0.925595	0.921053	0.922078	0.891775	0.903175	
1	RF	0.922849	0.918919	0.879630	0.922078	0.897852	
2	GB	0.925816	0.918919	0.899471	0.899471	0.899471	
3	ET	0.925816	0.918919	0.888889	0.916667	0.885714	



DT

۱. نتیجه خروجی نامناسب بعد از tune کردن

راه حل:

- تغییر هایپرپارامترهای Grid Search
- گذاشتن مقدار دیفالت هایپرپارامترها
- تغییر روش تقسیم‌بندی داده از holdout به Cross validation

۲. در روش Cross validation کدام fold برای tune کردن در نظر گرفته شود؟

راه حل:

- در اینجا به علت زمان‌بر بودن tune کردن ما بهترین fold را از بین تمامی foldها در نظر گرفتیم



✓ استفاده از Feature Engineering

✓ استفاده از روش Ensemble Learning

✓ Hyperparameter Tuning:

• Bayesian Optimization



https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset?select=Sleep_health_and_lifestyle_dataset.csv

<https://www.kaggle.com/code/amalyasser/shhh-i-want-to-sleep#Data-modeling-%E2%9A%99%EF%B8%8F>

<https://www.kaggle.com/code/satishgunjal/tutorial-k-fold-cross-validation>

<https://www.analyticsvidhya.com/blog/2021/06/tune-hyperparameters-with-gridsearchcv/>