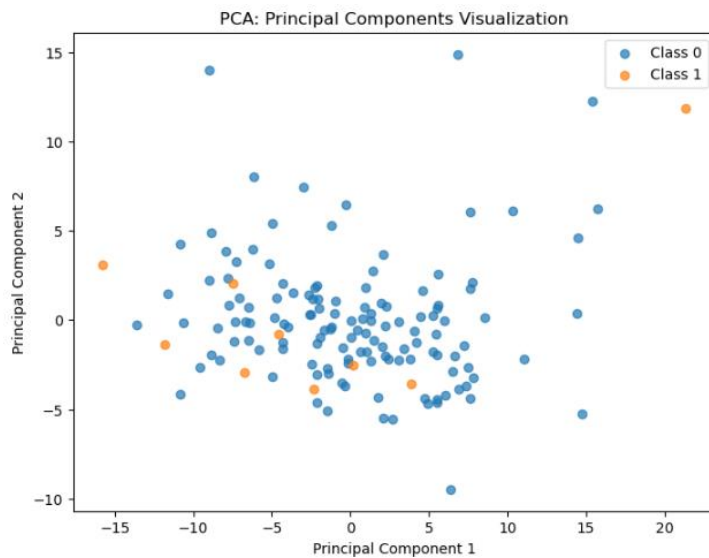


* TOTAL은 제외함

* agg Feature로 mean, max, min, var, median을 사용

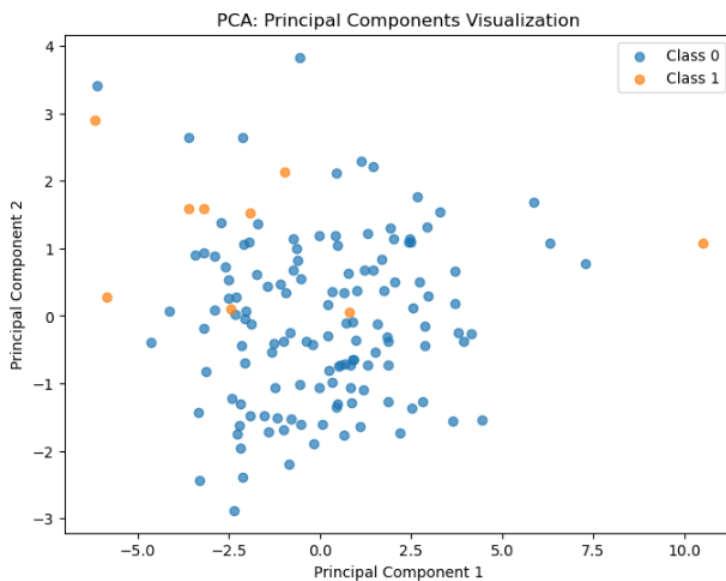
실험 1. Activity Groupby만 사용

- 원본 데이터 PCA



- 중요 피쳐 PCA

'activity_rest_median', 'activity_met_min_low_mean', 'activity_low_median', 'activity_non_wear_mean',
'activity_rest_mean', 'activity_total_median', 'activity_met_min_low_median',
'activity_daily_movement_median', 'activity_steps_median', 'activity_cal_active_median' 사용



1. 로지스틱 회귀

Logit Regression Results						
=====						
Dep. Variable:	DIAG_NM	No. Observations:	141			
Model:	Logit	Df Residuals:	130			
Method:	MLE	Df Model:	10			
Date:	Sat, 30 Nov 2024	Pseudo R-squ.:	0.4237			
Time:	15:55:10	Log-Likelihood:	-19.289			
converged:	True	LL-Null:	-33.470			
Covariance Type:	nonrobust	LLR p-value:	0.001579			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	-2.8173	4.937	-0.571	0.568	-12.493	6.859
activity_rest_median	0.0716	0.043	1.680	0.093	-0.012	0.155
activity_met_min_low_mean	-0.2526	0.111	-2.281	0.023	-0.470	-0.036
activity_low_median	0.0244	0.074	0.329	0.742	-0.121	0.170
activity_non_wear_mean	-0.0436	0.034	-1.283	0.200	-0.110	0.023
activity_rest_mean	-0.0607	0.040	-1.504	0.133	-0.140	0.018
activity_total_median	-0.0521	0.073	-0.711	0.477	-0.196	0.092
activity_met_min_low_median	0.2575	0.100	2.563	0.010	0.061	0.454
activity_daily_movement_median	-0.0007	0.002	-0.416	0.677	-0.004	0.003
activity_steps_median	0.0002	0.001	0.370	0.712	-0.001	0.001
activity_cal_active_median	0.0202	0.024	0.854	0.393	-0.026	0.067
=====						
Logistic Regression Accuracy: 0.8787878787878788						
Classification Report:						
	precision	recall	f1-score	support		
0	0.96	0.90	0.93	30		
1	0.40	0.67	0.50	3		
accuracy			0.88	33		
macro avg	0.68	0.78	0.72	33		
weighted avg	0.91	0.88	0.89	33		
Confusion Matrix:						
[[27 3]						
[1 2]]						

2. 나이브 베이즈

Naive Bayes Accuracy: 0.9696969696969697				
Classification Report:				
	precision	recall	f1-score	support
0	0.97	1.00	0.98	30
1	1.00	0.67	0.80	3
accuracy			0.97	33
macro avg	0.98	0.83	0.89	33
weighted avg	0.97	0.97	0.97	33
Confusion Matrix:				
[[30 0]				
[1 2]]				

3. XGBoost

XGBoost Accuracy: 0.9090909090909091

Classification Report:

	precision	recall	f1-score	support
0	0.94	0.97	0.95	30
1	0.50	0.33	0.40	3
accuracy			0.91	33
macro avg	0.72	0.65	0.68	33
weighted avg	0.90	0.91	0.90	33

Confusion Matrix:

```
[[29  1]
 [ 2  1]]
```

4. LightGBM

LightGBM Accuracy: 0.9393939393939394

Classification Report:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	30
1	1.00	0.33	0.50	3
accuracy			0.94	33
macro avg	0.97	0.67	0.73	33
weighted avg	0.94	0.94	0.93	33

Confusion Matrix:

```
[[30  0]
 [ 2  1]]
```

5. Decision Tree

Decision Tree Accuracy: 0.8787878787878788

Classification Report:

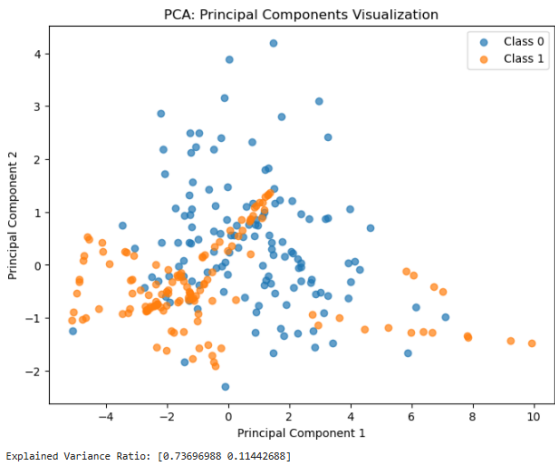
	precision	recall	f1-score	support
0	0.91	0.97	0.94	30
1	0.00	0.00	0.00	3
accuracy			0.88	33
macro avg	0.45	0.48	0.47	33
weighted avg	0.82	0.88	0.85	33

Confusion Matrix:

```
[[29  1]
 [ 3  0]]
```

실험 1-2. Activity Groupby + SMOTE

- PCA



1. 로지스틱 회귀

Logit Regression Results						
Dep. Variable:	DIAG_NM	No. Observations:	264			
Model:	Logit	Df Residuals:	253			
Method:	MLE	Df Model:	10			
Date:	Sat, 30 Nov 2024	Pseudo R-squ.:	0.5884			
Time:	15:55:12	Log-Likelihood:	-75.323			
converged:	True	LL-Null:	-182.99			
Covariance Type:	nonrobust	LLR p-value:	1.011e-40			
	coef	std err	z	P> z	[0.025	0.975]
const	-5.1910	3.768	-1.378	0.168	-12.577	2.195
activity_rest_median	0.1310	0.031	4.274	0.000	0.071	0.191
activity_met_min_low_mean	-0.2857	0.062	-4.613	0.000	-0.407	-0.164
activity_low_median	0.0317	0.044	0.724	0.469	-0.054	0.118
activity_non_wear_mean	-0.0183	0.013	-1.414	0.157	-0.044	0.007
activity_rest_mean	-0.1121	0.029	-3.814	0.000	-0.170	-0.055
activity_total_median	-0.0823	0.044	-1.867	0.062	-0.169	0.004
activity_met_min_low_median	0.2981	0.057	5.233	0.000	0.186	0.410
activity_daily_movement_median	-0.0021	0.001	-1.858	0.063	-0.004	0.000
activity_steps_median	1.951e-05	0.000	0.046	0.963	-0.001	0.001
activity_cal_active_median	0.0616	0.019	3.262	0.001	0.025	0.099

Logistic Regression Accuracy: 0.8484848484848485

Classification Report:				
	precision	recall	f1-score	support
0	0.96	0.87	0.91	30
1	0.33	0.67	0.44	3
accuracy			0.85	33
macro avg	0.65	0.77	0.68	33
weighted avg	0.91	0.85	0.87	33

Confusion Matrix:
[[26 4]
[1 2]]

2. 나이브 베이즈

Naive Bayes Accuracy: 0.81818181818182

Classification Report:

	precision	recall	f1-score	support
0	1.00	0.80	0.89	30
1	0.33	1.00	0.50	3
accuracy			0.82	33
macro avg	0.67	0.90	0.69	33
weighted avg	0.94	0.82	0.85	33

Confusion Matrix:

```
[[24  6]
 [ 0  3]]
```

3. XGBoost

XGBoost Accuracy: 0.8787878787878788

Classification Report:

	precision	recall	f1-score	support
0	0.91	0.97	0.94	30
1	0.00	0.00	0.00	3
accuracy			0.88	33
macro avg	0.45	0.48	0.47	33
weighted avg	0.82	0.88	0.85	33

Confusion Matrix:

```
[[29  1]
 [ 3  0]]
```

4. LightGBM

LightGBM Accuracy: 0.9090909090909091

Classification Report:

	precision	recall	f1-score	support
0	0.94	0.97	0.95	30
1	0.50	0.33	0.40	3
accuracy			0.91	33
macro avg	0.72	0.65	0.68	33
weighted avg	0.90	0.91	0.90	33

Confusion Matrix:

```
[[29  1]
 [ 2  1]]
```

5. Decision Tree

Decision Tree Accuracy: 0.8787878787878788

Classification Report:

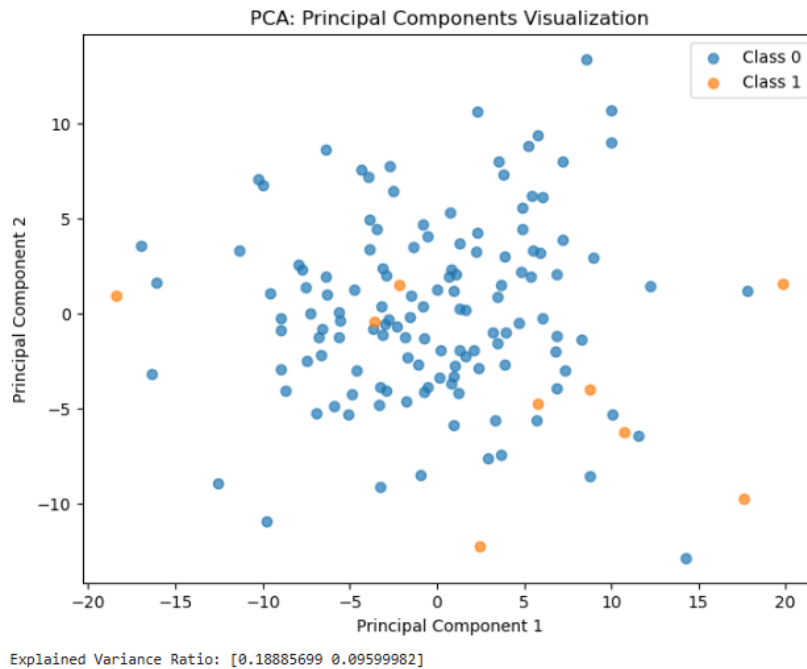
	precision	recall	f1-score	support
0	0.91	0.97	0.94	30
1	0.00	0.00	0.00	3
accuracy			0.88	33
macro avg	0.45	0.48	0.47	33
weighted avg	0.82	0.88	0.85	33

Confusion Matrix:

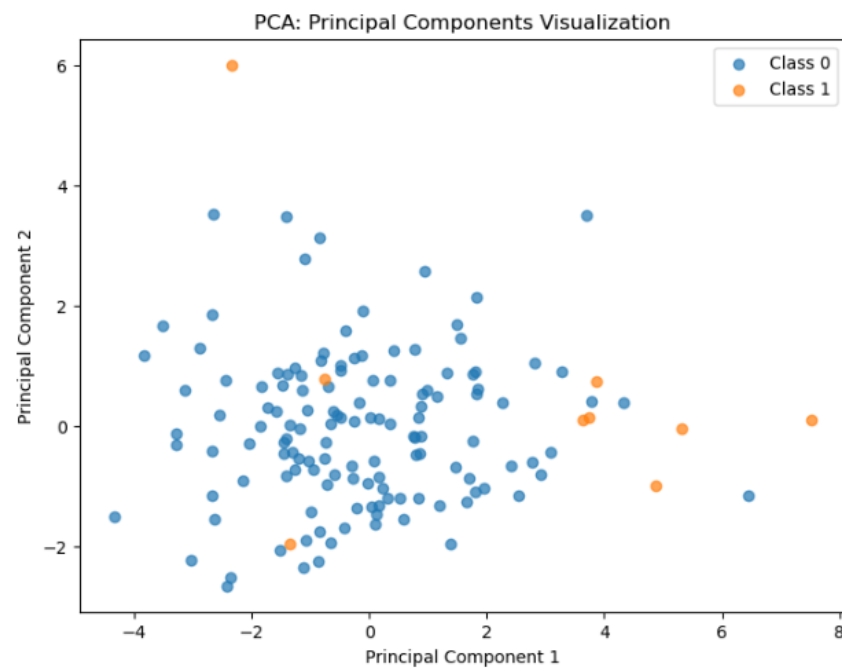
```
[[29  1]
 [ 3  0]]
```

실험 2. Activity Groupby + Sleep Groupby 사용

- 원본 데이터 PCA



- 중요 피쳐 PCA



1. 로지스틱 회귀

Logit Regression Results						
=====						
Dep. Variable:	DIAG_NM	No. Observations:	141			
Model:	Logit	Df Residuals:	130			
Method:	MLE	Df Model:	10			
Date:	Sat, 30 Nov 2024	Pseudo R-squ.:	0.4474			
Time:	15:55:16	Log-Likelihood:	-18.495			
converged:	True	LL-Null:	-33.470			
Covariance Type:	nonrobust	LLR p-value:	0.0008728			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	1.8613	9.961	0.187	0.852	-17.661	21.384
sleep_midpoint_time_median	0.0020	0.001	1.556	0.120	-0.001	0.005
sleep_duration_median	-0.0009	0.001	-1.210	0.226	-0.002	0.001
sleep_light_median	6.792e-06	0.001	0.005	0.996	-0.003	0.003
activity_cal_active_median	0.0024	0.003	0.885	0.376	-0.003	0.008
sleep_light_mean	0.0002	0.001	0.134	0.894	-0.002	0.003
sleep_score_latency_mean	-0.1567	0.115	-1.358	0.175	-0.383	0.069
sleep_rem_var	2.431e-07	4.17e-07	0.583	0.560	-5.74e-07	1.06e-06
activity_low_var	-0.0004	0.000	-1.439	0.150	-0.001	0.000
sleep_score_deep_var	0.0042	0.002	1.932	0.053	-6.1e-05	0.008
activity_met_min_low_median	0.0060	0.011	0.540	0.589	-0.016	0.028
=====						

Logistic Regression Accuracy: 0.8787878787878788

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.90	0.93	30
1	0.40	0.67	0.50	3
accuracy			0.88	33
macro avg	0.68	0.78	0.72	33
weighted avg	0.91	0.88	0.89	33

Confusion Matrix:

```
[[27  3]
 [ 1  2]]
```

2. 나이브 베이즈

Naive Bayes Accuracy: 0.9696969696969697

Classification Report:

	precision	recall	f1-score	support
0	0.97	1.00	0.98	30
1	1.00	0.67	0.80	3
accuracy			0.97	33
macro avg	0.98	0.83	0.89	33
weighted avg	0.97	0.97	0.97	33

Confusion Matrix:

```
[[30  0]
 [ 1  2]]
```

3. XGBoost

XGBoost Accuracy: 0.9393939393939394

Classification Report:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	30
1	1.00	0.33	0.50	3
accuracy			0.94	33
macro avg	0.97	0.67	0.73	33
weighted avg	0.94	0.94	0.93	33

Confusion Matrix:

```
[[30  0]
 [ 2  1]]
```

4. LightGBM

LightGBM Accuracy: 0.9393939393939394

Classification Report:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	30
1	1.00	0.33	0.50	3
accuracy			0.94	33
macro avg	0.97	0.67	0.73	33
weighted avg	0.94	0.94	0.93	33

Confusion Matrix:

```
[[30  0]
 [ 2  1]]
```

5. Decision Tree

Decision Tree Accuracy: 0.9090909090909091

Classification Report:

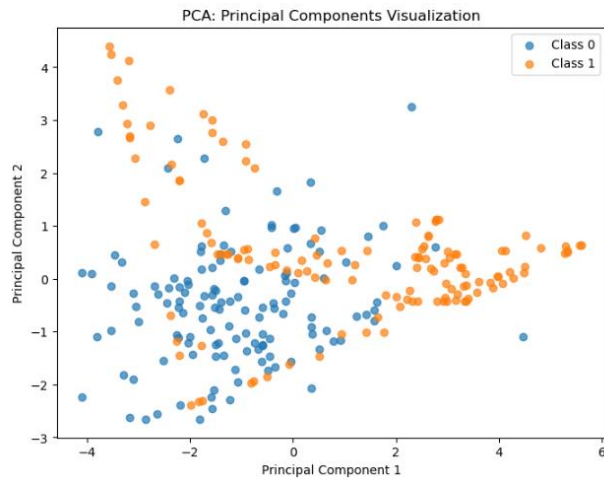
	precision	recall	f1-score	support
0	0.91	1.00	0.95	30
1	0.00	0.00	0.00	3
accuracy			0.91	33
macro avg	0.45	0.50	0.48	33
weighted avg	0.83	0.91	0.87	33

Confusion Matrix:

```
[[30  0]
 [ 3  0]]
```


실험 2-2. Activity Groupby + Sleep Groupby + SMOTE

- PCA



>> SMOTE는 이상치 데이터 또한 크게 증강시킨다.

1. 로지스틱 회귀

Logit Regression Results						
Dep. Variable:	DIAG_NM	No. Observations:	264			
Model:	Logit	Df Residuals:	253			
Method:	MLE	Df Model:	10			
Date:	Sat, 30 Nov 2024	Pseudo R-squ.:	0.5398			
Time:	15:55:17	Log-Likelihood:	-84.218			
converged:	True	LL-Null:	-182.99			
Covariance Type:	nonrobust	LLR p-value:	5.243e-37			
	coef	std err	z	P> z	[0.025	0.975]
const	-3.4192	6.188	-0.553	0.581	-15.548	8.710
sleep_midpoint_time_median	0.0029	0.001	4.057	0.000	0.002	0.004
sleep_duration_median	-0.0015	0.000	-3.711	0.000	-0.002	-0.001
sleep_light_median	-0.0005	0.001	-0.737	0.461	-0.002	0.001
activity_cal_active_median	0.0043	0.002	2.310	0.021	0.001	0.008
sleep_light_mean	0.0010	0.001	1.504	0.133	-0.000	0.002
sleep_score_latency_mean	-0.0616	0.069	-0.894	0.371	-0.197	0.073
sleep_rem_var	4.355e-08	2.7e-07	0.161	0.872	-4.86e-07	5.73e-07
activity_low_var	-0.0006	0.000	-3.676	0.000	-0.001	-0.000
sleep_score_deep_var	0.0057	0.001	5.072	0.000	0.004	0.008
activity_met_min_low_median	0.0026	0.006	0.415	0.678	-0.010	0.015

Logistic Regression Accuracy: 0.8787878787878788

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.90	0.93	30
1	0.40	0.67	0.50	3
accuracy			0.88	33
macro avg	0.68	0.78	0.72	33
weighted avg	0.91	0.88	0.89	33

Confusion Matrix:

```
[[27  3]
 [ 1  2]]
```

2. 나이브 베이즈

```
Naive Bayes Accuracy: 0.9090909090909091
Classification Report:
              precision    recall  f1-score   support

     0           1.00       0.90       0.95        30
     1           0.50       1.00       0.67         3

   accuracy          0.91        33
  macro avg          0.75        0.95        0.81        33
 weighted avg          0.95        0.91        0.92        33

Confusion Matrix:
[[27  3]
 [ 0  3]]
```

3. XGBoost

```
XGBoost Accuracy: 0.9696969696969697
Classification Report:
              precision    recall  f1-score   support

     0           1.00       0.97       0.98        30
     1           0.75       1.00       0.86         3

   accuracy          0.97        33
  macro avg          0.88        0.98        0.92        33
 weighted avg          0.98        0.97        0.97        33

Confusion Matrix:
[[29  1]
 [ 0  3]]
```

4. LightGBM

```
LightGBM Accuracy: 0.9393939393939394
Classification Report:
              precision    recall  f1-score   support

     0           0.97       0.97       0.97        30
     1           0.67       0.67       0.67         3

   accuracy          0.94        33
  macro avg          0.82        0.82        0.82        33
 weighted avg          0.94        0.94        0.94        33

Confusion Matrix:
[[29  1]
 [ 1  2]]
```

5. Decision Tree

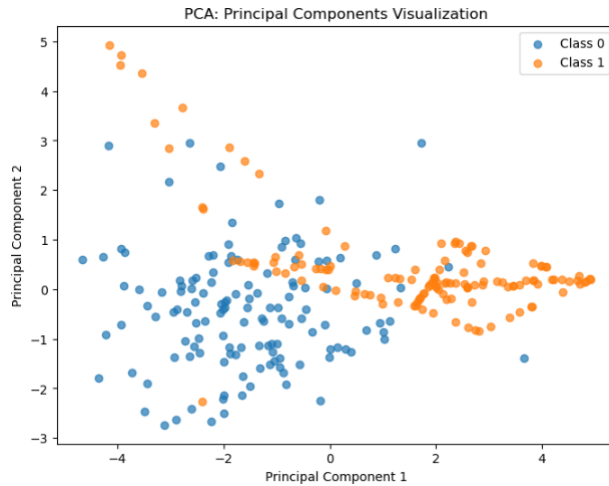
```
Decision Tree Accuracy: 0.8484848484848485
Classification Report:
              precision    recall  f1-score   support

     0           0.96       0.87       0.91        30
     1           0.33       0.67       0.44         3

   accuracy          0.85        33
  macro avg          0.65        0.77        0.68        33
 weighted avg          0.91        0.85        0.87        33

Confusion Matrix:
[[26  4]
 [ 1  2]]
```

실험 2-3. Activity Groupby + Sleep Groupby + BorderLine SMOTE



>> SMOTE 보다는 이상치를 증강하는 비율이 적다.

1. 로지스틱 회귀

Logit Regression Results						
Dep. Variable:	DIAG_NM	No. Observations:	264			
Model:	Logit	Df Residuals:	253			
Method:	MLE	Df Model:	10			
Date:	Sat, 30 Nov 2024	Pseudo R-squ.:	0.7514			
Time:	15:55:17	Log-Likelihood:	-45.488			
converged:	True	LL-Null:	-182.99			
Covariance Type:	nonrobust	LLR p-value:	2.945e-53			
	coef	std err	z	P> z	[0.025	0.975]
const	4.0979	8.916	0.460	0.646	-13.376	21.572
sleep_midpoint_time_median	0.0013	0.001	1.313	0.189	-0.001	0.003
sleep_duration_median	-8.7e-05	0.001	-0.146	0.884	-0.001	0.001
sleep_light_median	-0.0014	0.001	-1.341	0.180	-0.004	0.001
activity_cal_active_median	0.0029	0.002	1.208	0.227	-0.002	0.008
sleep_light_mean	0.0013	0.001	1.275	0.202	-0.001	0.003
sleep_score_latency_mean	-0.3548	0.110	-3.238	0.001	-0.570	-0.140
sleep_rem_var	4.155e-07	3.32e-07	1.251	0.211	-2.35e-07	1.07e-06
activity_low_var	-0.0010	0.000	-3.597	0.000	-0.002	-0.000
sleep_score_deep_var	0.0105	0.002	4.943	0.000	0.006	0.015
activity_met_min_low_median	0.0288	0.010	2.932	0.003	0.010	0.048

Logistic Regression Accuracy: 0.8787878787878788

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.90	0.93	30
1	0.40	0.67	0.50	3
accuracy			0.88	33
macro avg	0.68	0.78	0.72	33
weighted avg	0.91	0.88	0.89	33

Confusion Matrix:

```
[[27  3]
 [ 1 22]]
```

2. 나이브 베이즈

Naive Bayes Accuracy: 0.9090909090909091

Classification Report:

	precision	recall	f1-score	support
0	1.00	0.90	0.95	30
1	0.50	1.00	0.67	3
accuracy			0.91	33
macro avg	0.75	0.95	0.81	33
weighted avg	0.95	0.91	0.92	33

Confusion Matrix:

```
[[27  3]
 [ 0  3]]
```

3. XGBoost

XGBoost Accuracy: 0.9696969696969697

Classification Report:

	precision	recall	f1-score	support
0	1.00	0.97	0.98	30
1	0.75	1.00	0.86	3
accuracy			0.97	33
macro avg	0.88	0.98	0.92	33
weighted avg	0.98	0.97	0.97	33

Confusion Matrix:

```
[[29  1]
 [ 0  3]]
```

4. LightGBM

LightGBM Accuracy: 0.9696969696969697

Classification Report:

	precision	recall	f1-score	support
0	0.97	1.00	0.98	30
1	1.00	0.67	0.80	3
accuracy			0.97	33
macro avg	0.98	0.83	0.89	33
weighted avg	0.97	0.97	0.97	33

Confusion Matrix:

```
[[30  0]
 [ 1  2]]
```

5. Decision Tree

Decision Tree Accuracy: 1.0

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	30
1	1.00	1.00	1.00	3
accuracy			1.00	33
macro avg	1.00	1.00	1.00	33
weighted avg	1.00	1.00	1.00	33

Confusion Matrix:

```
[[30  0]
 [ 0  3]]
```