

기계학습

서울대학교 빅데이터 핀테크 과정

# 개인의 심리 및 인적 정보를 활용한 유권자의 국가 선거 투표 여부 예측

1조 | 김서현 김연성 서문홍 조건우 최지은

---

# INDEX

---

01 Data

02 EDA(Exploratory Data Analysis)

03 Preprocessing

04 Modeling

05 Result

# 01. 데이터 소개

## DATA 설명

---

### ● 마키아벨리즘, TIPI 심리테스트를 활용하여 테스트 참가자의 국가 선거 투표 여부 예측

Q\_A : 마키아벨리즘 심리 테스트 문항

Q\_E : 마키아벨리즘 심리 테스트 질문을 답할때 까지의 시간

tp\_ : tipi 심리테스트-자가 진단 문항

wr\_ : 특정 단어 정의 인지 여부

wf\_ : 허구의 단어 정의 인지 여부

연령 / 모국어 / 가족 구성원의 수 / 성별 / 필기하는 손 / 혼인상태 / 인종 / 거주지역

Voted : 지난 해 국가 선거 투표 여부

 <https://www.dacon.io/competitions/official/235647/data/>

## 01 | 데이터 소개

# DATA 설명

심리 테스트(주관적 정보)		객관적 정보
마키아벨리즘 테스트	tipi 심리테스트	인적 정보
<p><b>“마키아벨리즘 성격”</b> : 목적을 위해서 수단과 방법을 가리지 않는 교활한 성격</p> <p>Score</p> <ul style="list-style-type: none"><li>● 높은 수치: 계산적, 신중함</li><li>● 낮은 수치: 개인적, 감정이입적</li></ul>	<p><b>“TIPI”</b> (Ten - Item Personality Inventory) : 10개 질문으로 하는 성격 검사</p> <p>Score (성격 유형별 점수)</p> <ul style="list-style-type: none"><li>● 성실성</li><li>● 친화성</li><li>● 정서적 안정성</li><li>● 경험 개방성</li><li>● 외향성</li></ul>	<ul style="list-style-type: none"><li>● 연령 / 모국어 / 가족 구성원의 수 / 성별 / 필기하는 손 / 혼인상태 / 인종 / 거주지역</li><li>● 지난 해 국가 선거 투표 여부</li></ul>

# 데이터 변수 설명(1/5)

변수명	내용	데이터 타입
QaA	Secret	범주형
QbA	The biggest difference between most criminals and other people is that the criminals are stupid enough to get caught.	범주형
QcA	Anyone who completely trusts anyone else is asking for trouble.	범주형
QdA	Secret	범주형
QeA	P.T. Barnum was wrong when he said that there's a sucker born every minute.	범주형
QfA	There is no excuse for lying to someone else.	범주형
QgA	Secret	범주형
QhA	Most people forget more easily the death of their parents than the loss of their property.	범주형
QiA	Secret	범주형
QjA	It is safest to assume that all people have a vicious streak and it will come out when they are given a chance.	범주형

## 데이터 변수 설명(2/5)

변수명	내용	데이터 타입
QkA	All in all, it is better to be humble and honest than to be important and dishonest.	범주형
QIA	Secret	범주형
QmA	It is hard to get ahead without cutting corners here and there.	범주형
QnA	Secret	범주형
QoA	The best way to handle people is to tell them what they want to hear.	범주형
QpA	Secret	범주형
QqA	Most people are basically good and kind.	범주형
QrA	One should take action only when sure it is morally right.	범주형
QsA	It is wise to flatter important people.	범주형
QtA	Secret	범주형

## 데이터 변수 설명(3/5)

변수명	내용	데이터 타입
Q_E	문항 “__” 의 질문을 답할 때 까지의 시간	연속형
age_group	연령	범주형
education	교육 수준	범주형
engnat	모국어가 영어	범주형
familysize	형제자매 수	범주형
gender	성별	범주형
hand	필기하는 손	범주형
married	혼인 상태	범주형
race	인종	범주형
religion	종교	범주형

## 데이터 변수 설명(4/5)

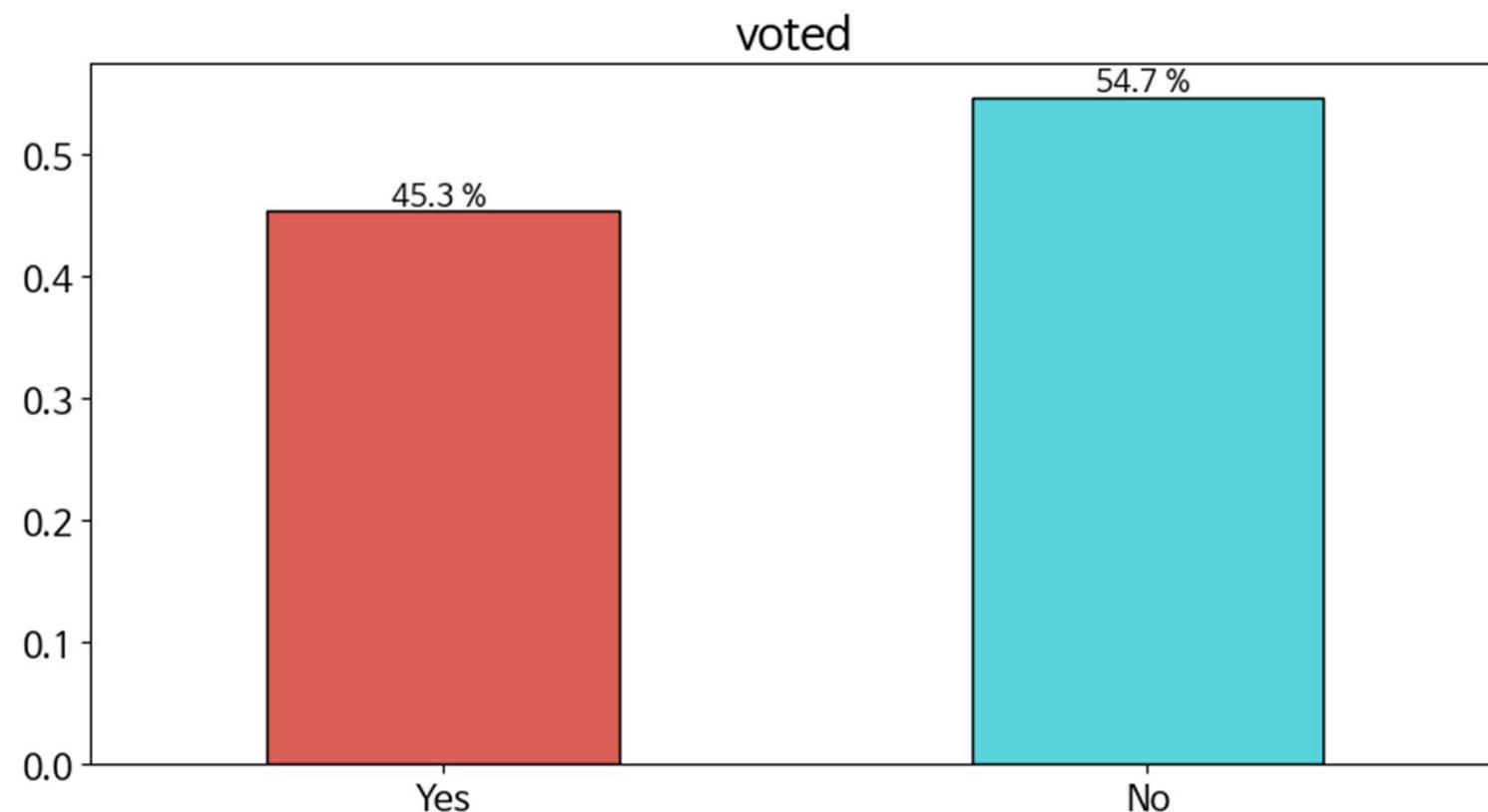
변수명	내용	데이터 타입
tp01	items were rated "I see myself as:" [Extraverted, enthusiastic]	범주형
tp02	items were rated "I see myself as:" [Critical, quarrelsome]	범주형
tp03	items were rated "I see myself as:" [Dependable, self-disciplined]	범주형
tp04	items were rated "I see myself as:" [Anxious, easily upset]	범주형
tp05	items were rated "I see myself as:" [Open to new experiences, complex]	범주형
tp06	items were rated "I see myself as:" [Reserved, quiet]	범주형
tp07	items were rated "I see myself as:" [Sympathetic, warm]	범주형
tp08	items were rated "I see myself as:" [Disorganized, careless]	범주형
tp09	items were rated "I see myself as:" [Calm, emotionally stable]	범주형
tp10	items were rated "I see myself as:" [Conventional, uncreative]	범주형

## 데이터 변수 설명(5/5)

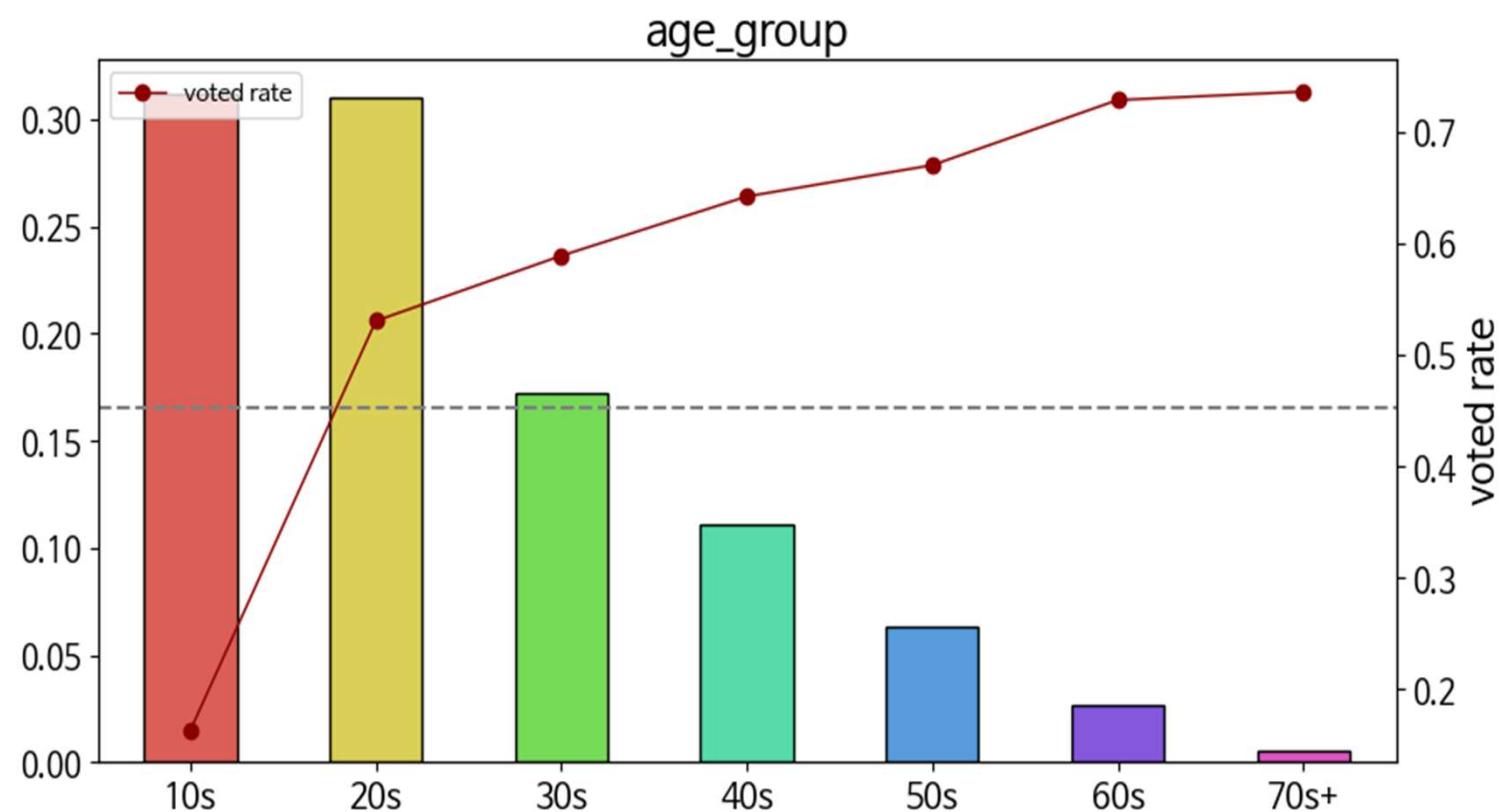
변수명	내용	데이터 타입
urban	유년기의 거주 구역	범주형
wr_	실존하는 해당 단어의 정의를 암	범주형
wf_	허구인 단어의 정의를 암	범주형
voted	지난 해 국가 선거 투표 여부	범주형

## 02. EDA (탐색적 데이터 분석)

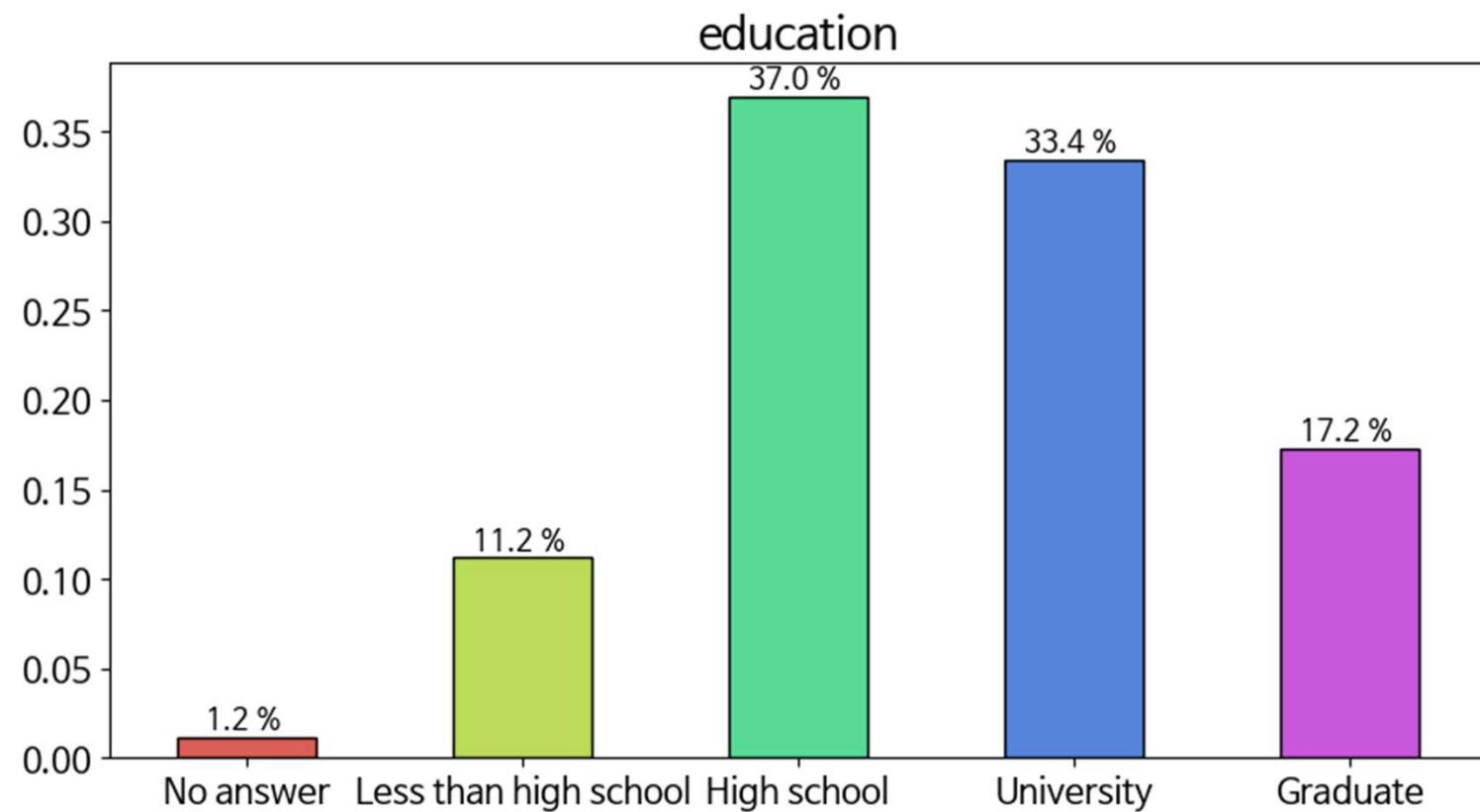
## 시각화-[voted]



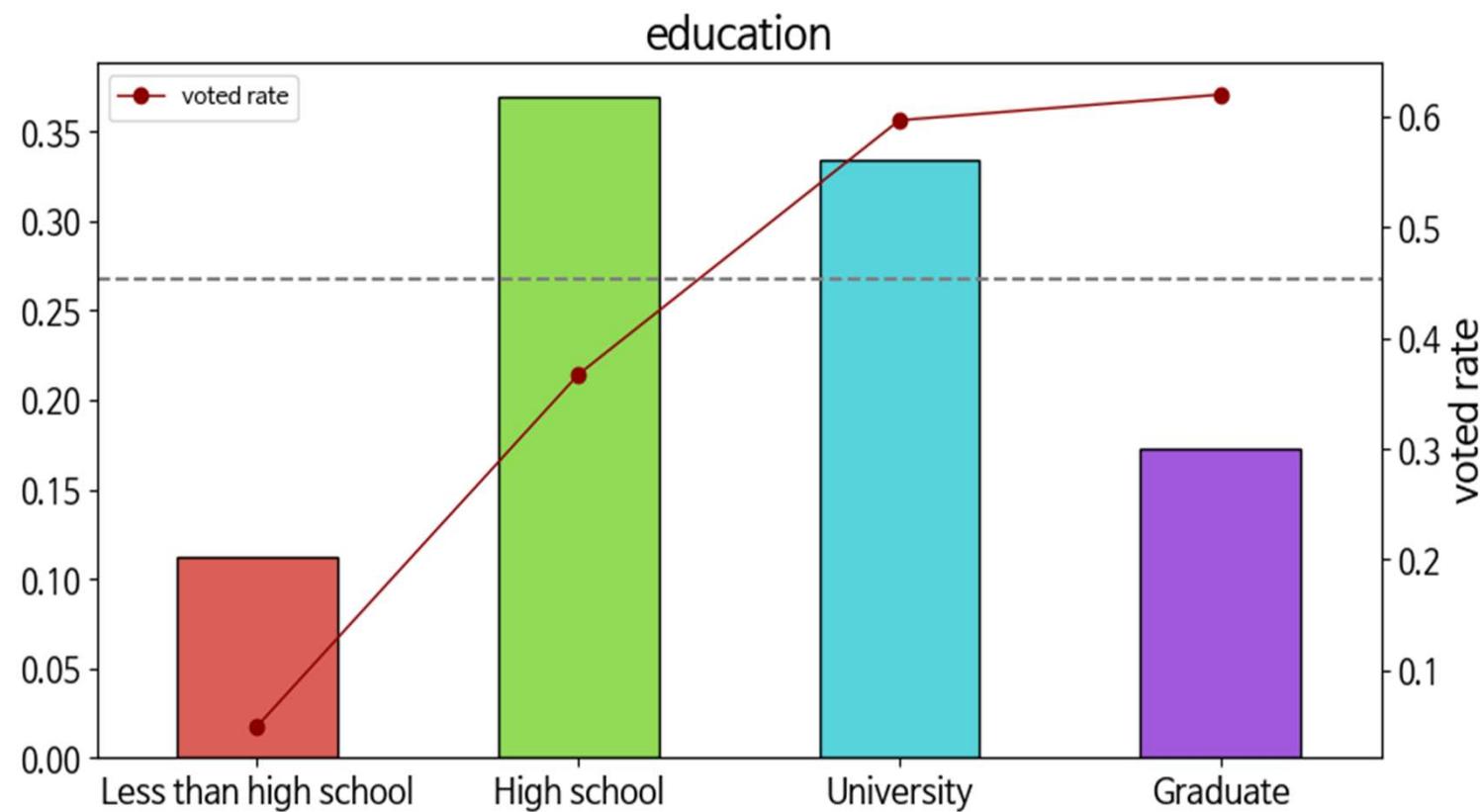
## 시각화-[age\_group]



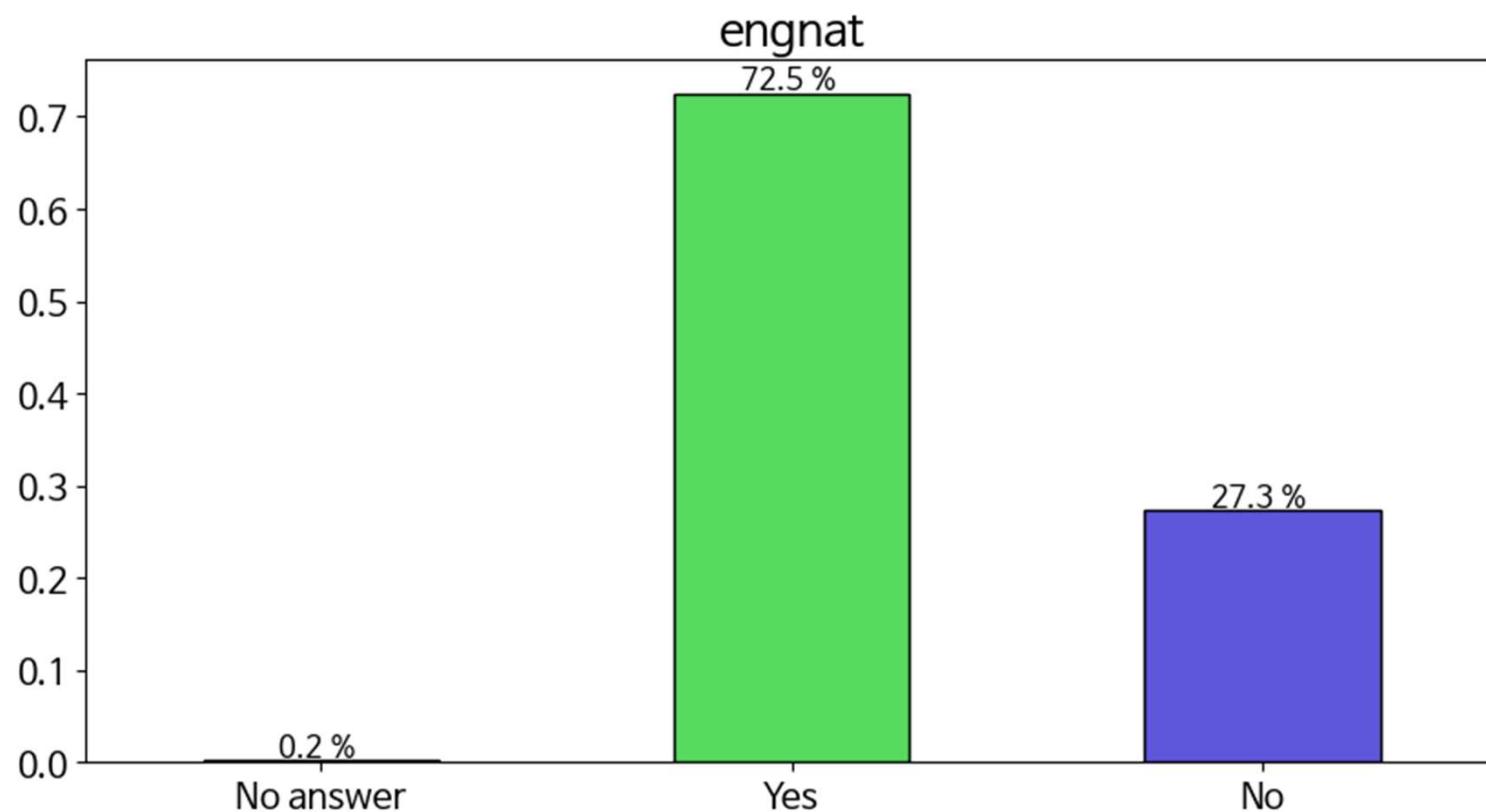
## 시각화-[education]



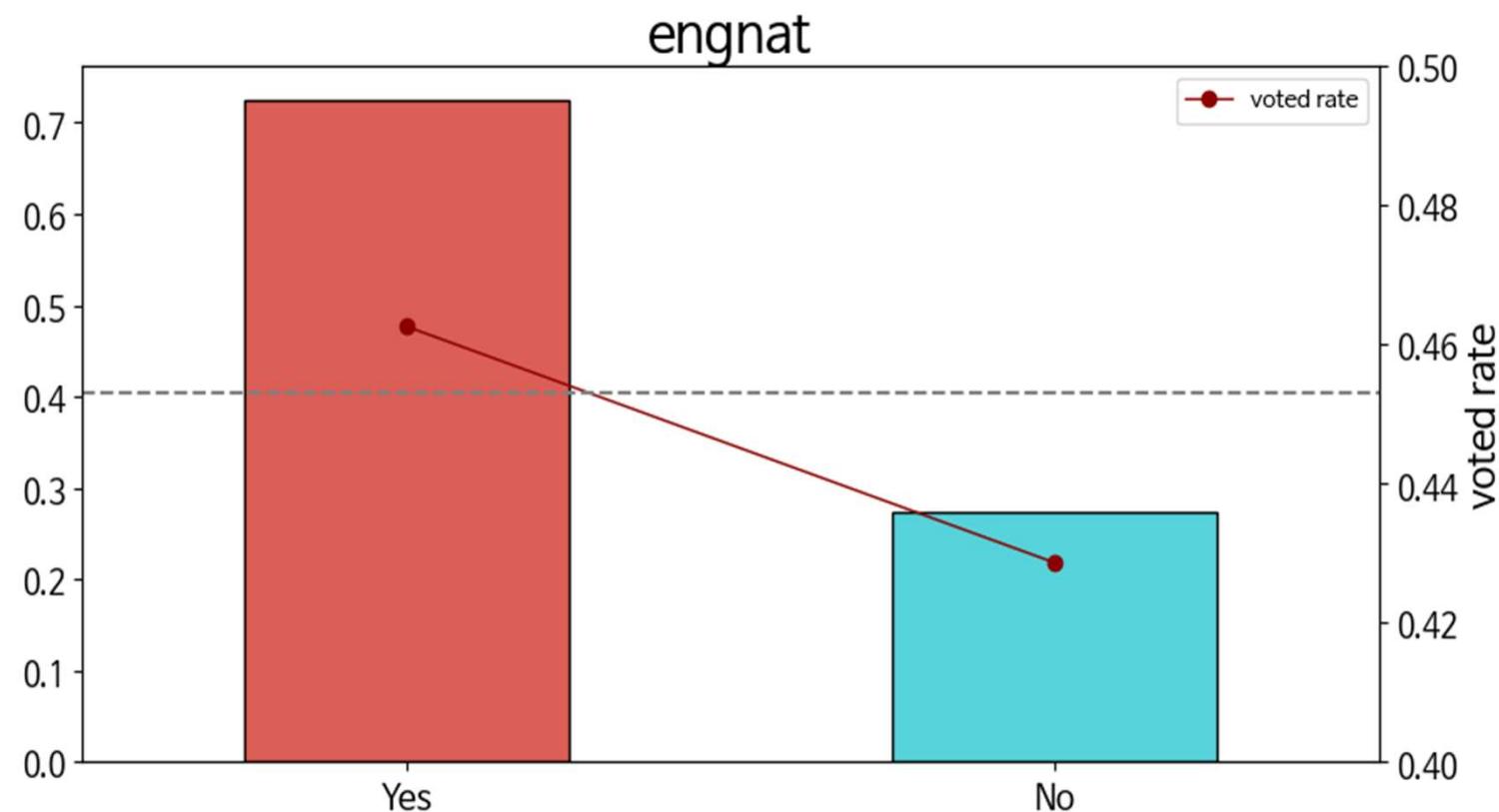
## 시각화-[education]



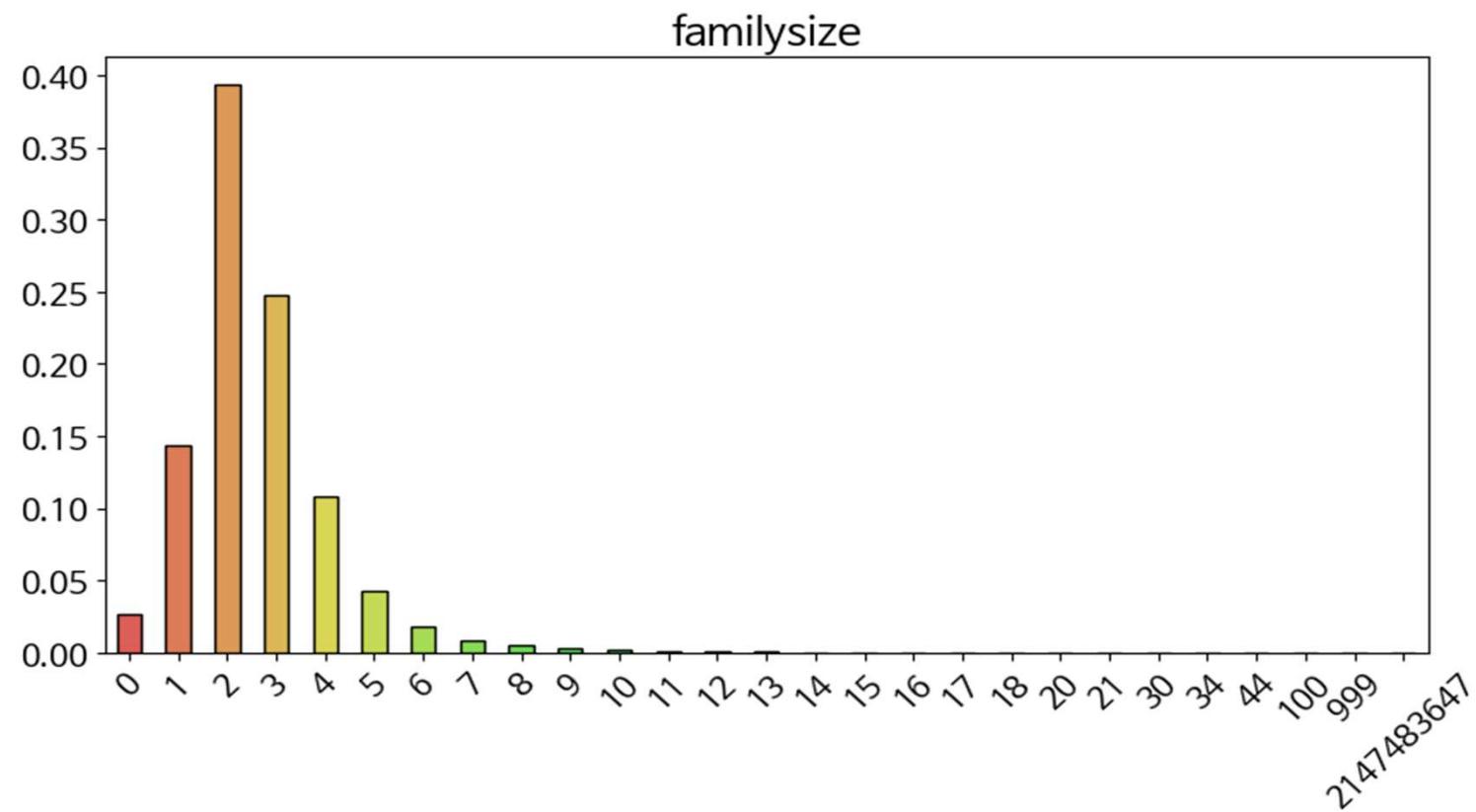
## 시각화-[engnat]



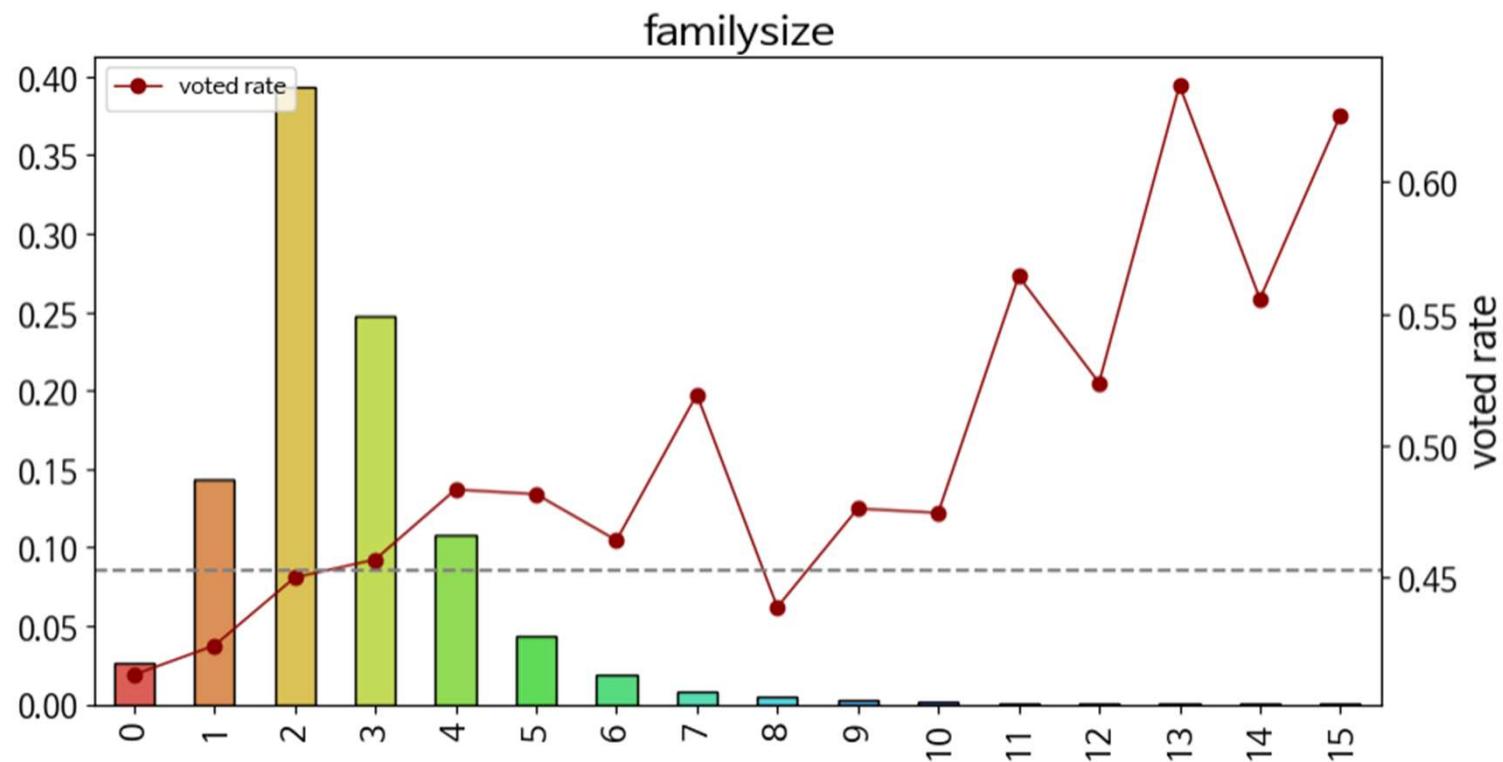
## 시각화-[engnat]



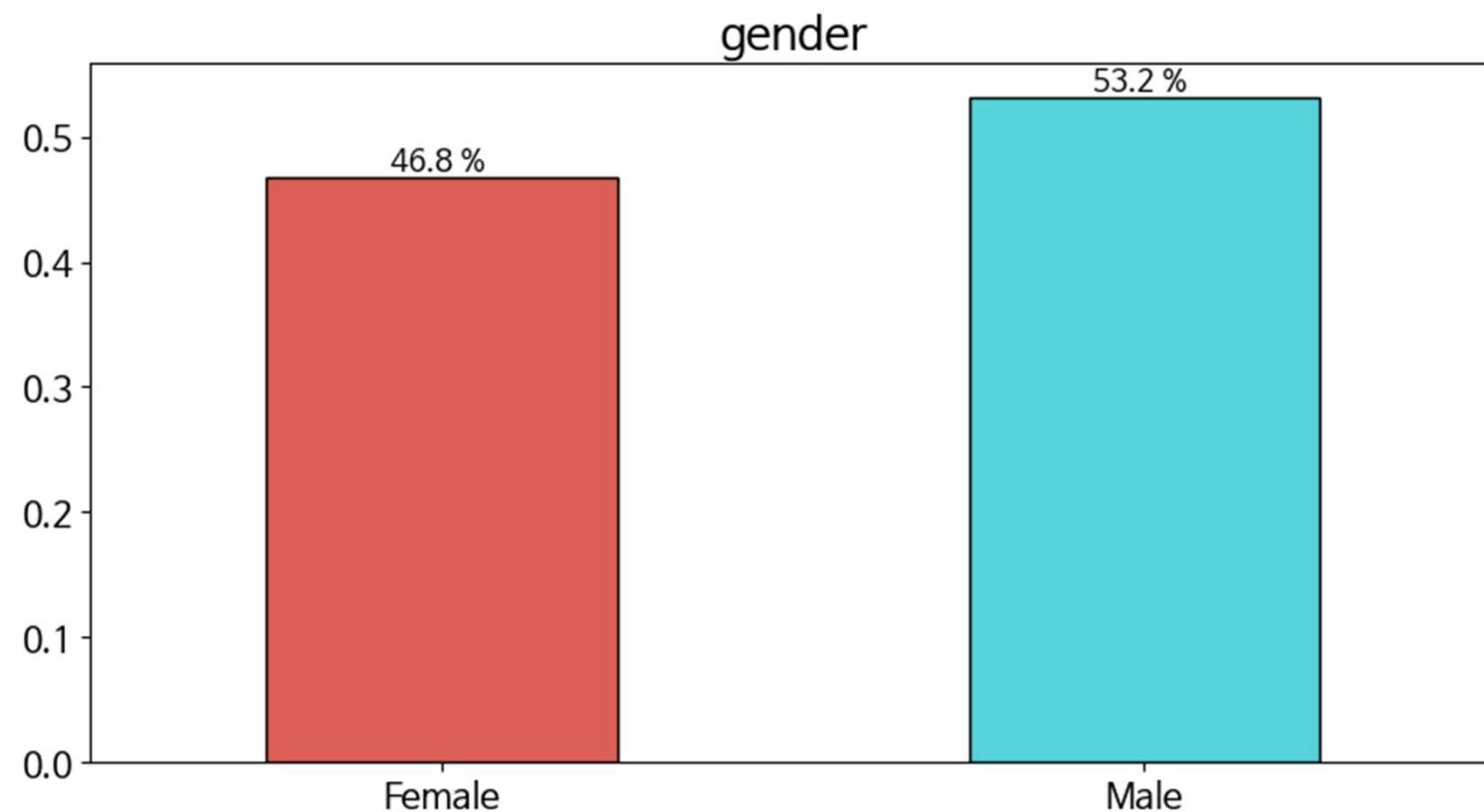
## 시각화-[familysize]



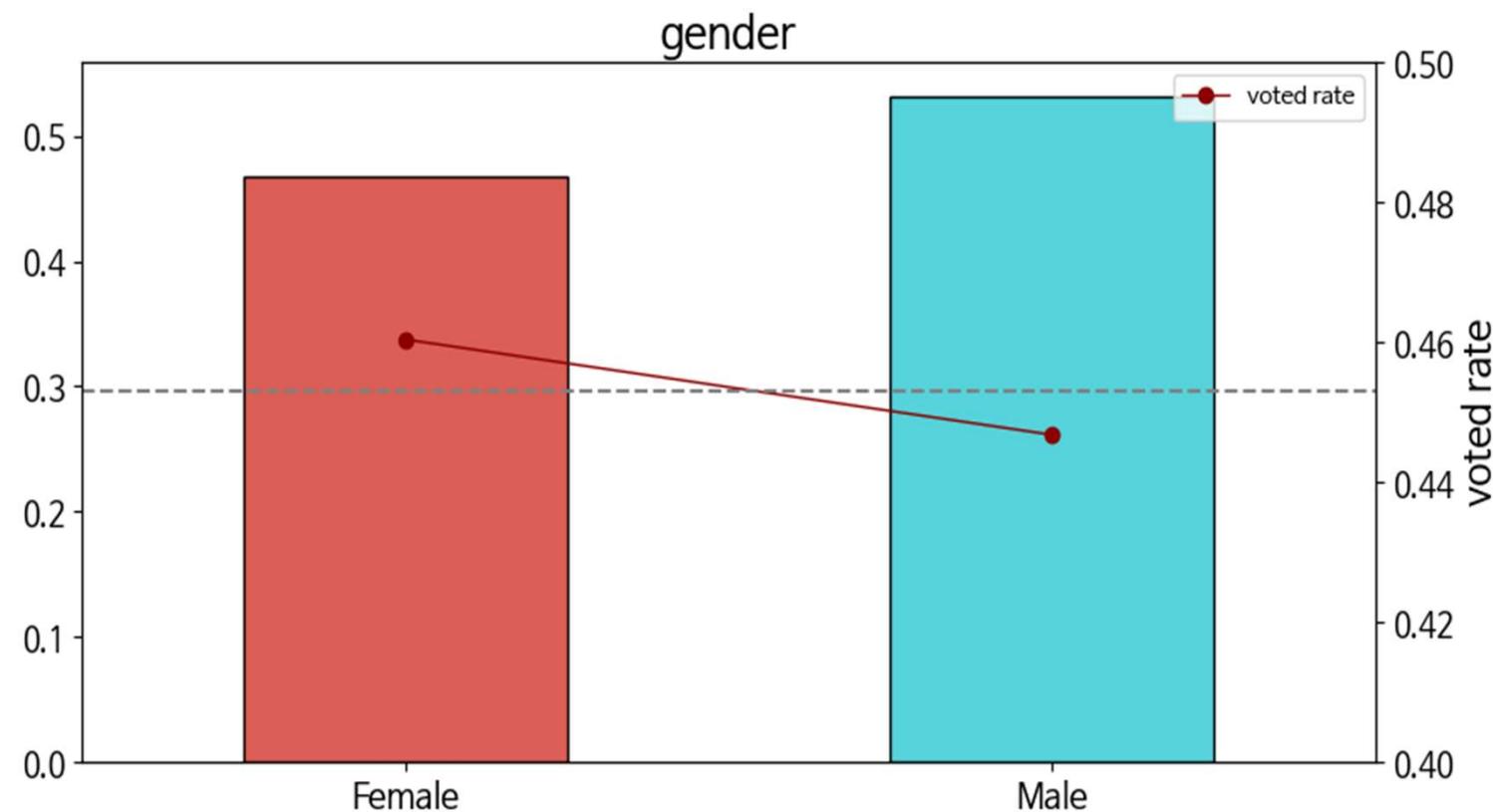
## 시각화-[familysize]



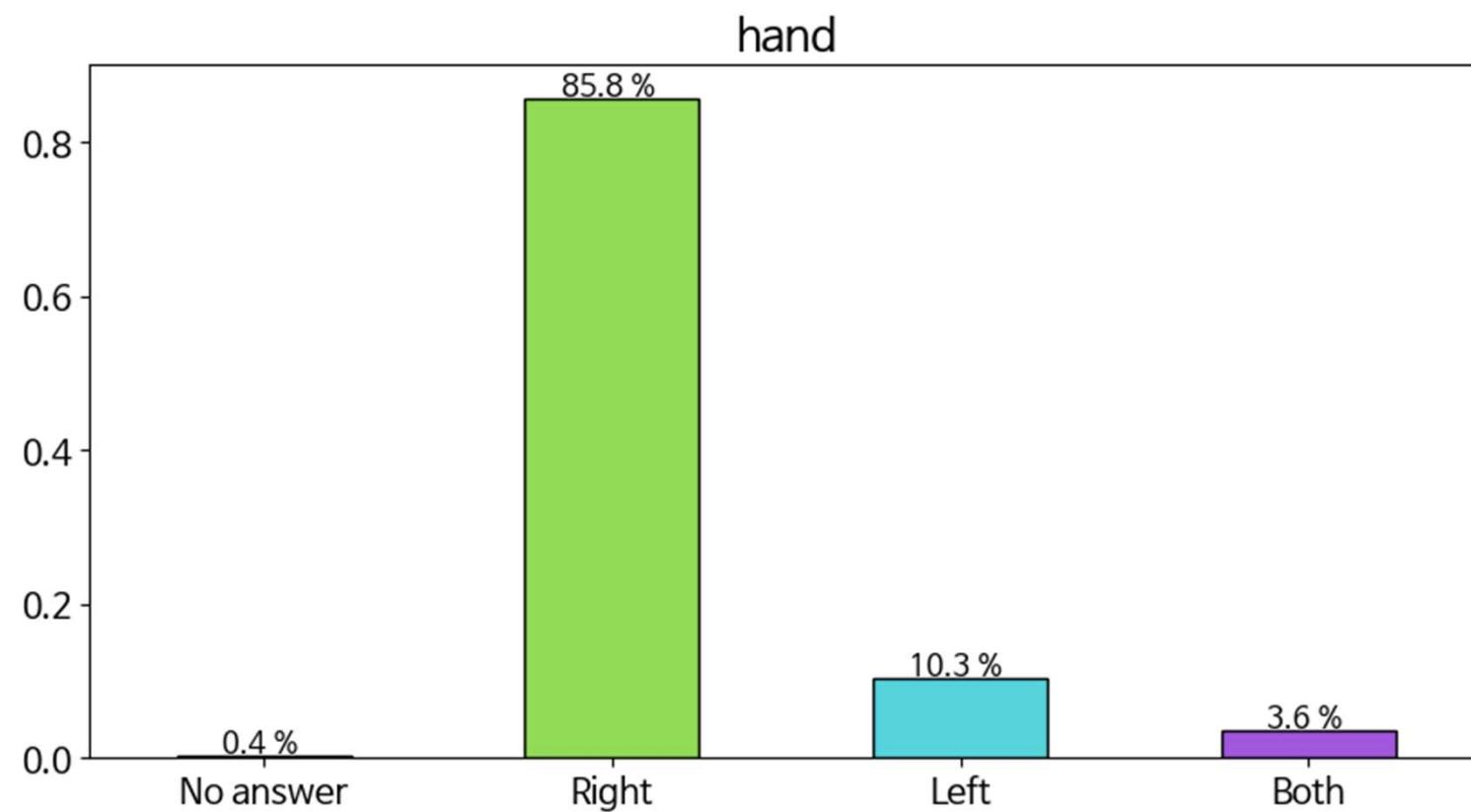
## 시각화-[gender]



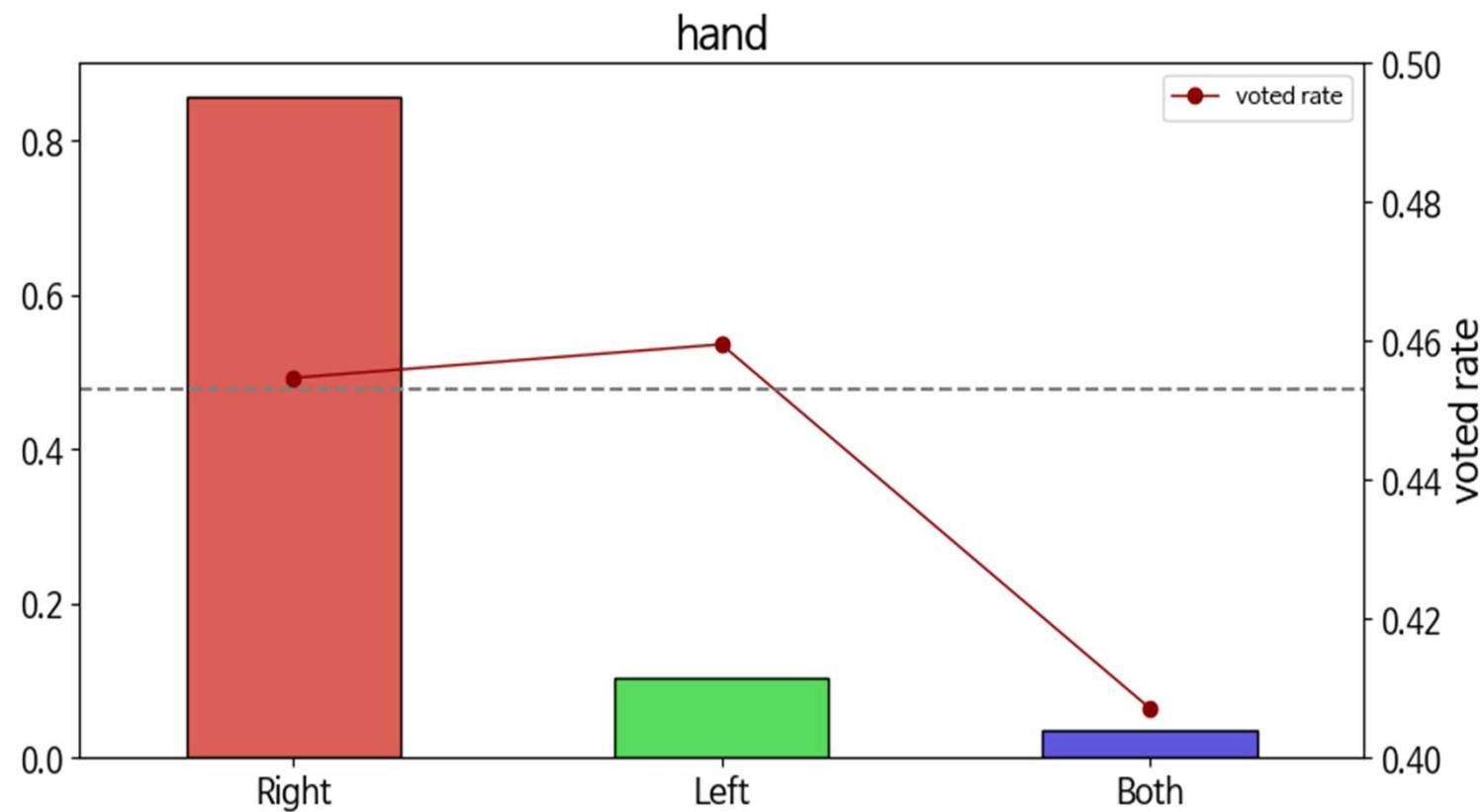
## 시각화-[gender]



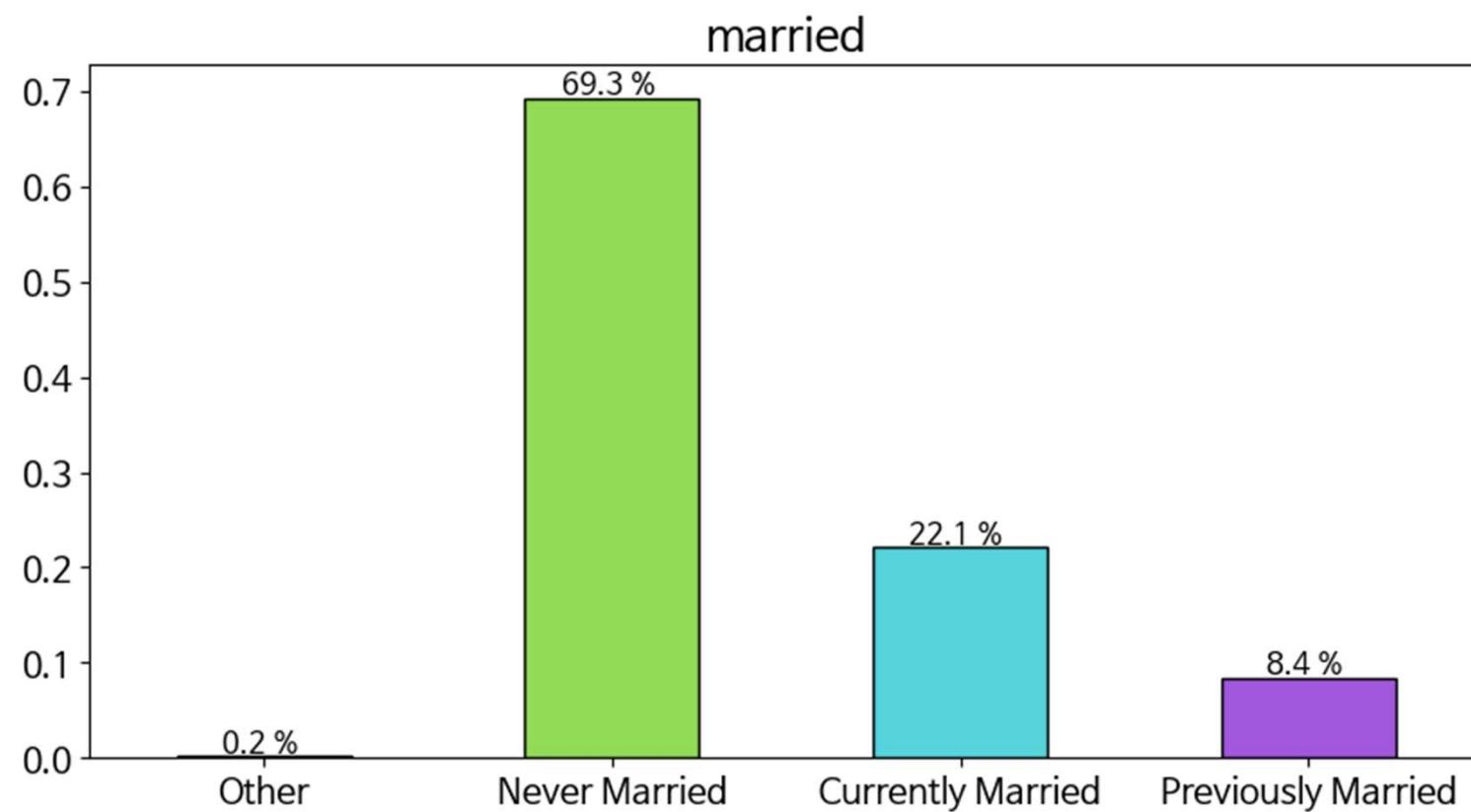
## 시각화-[hand]



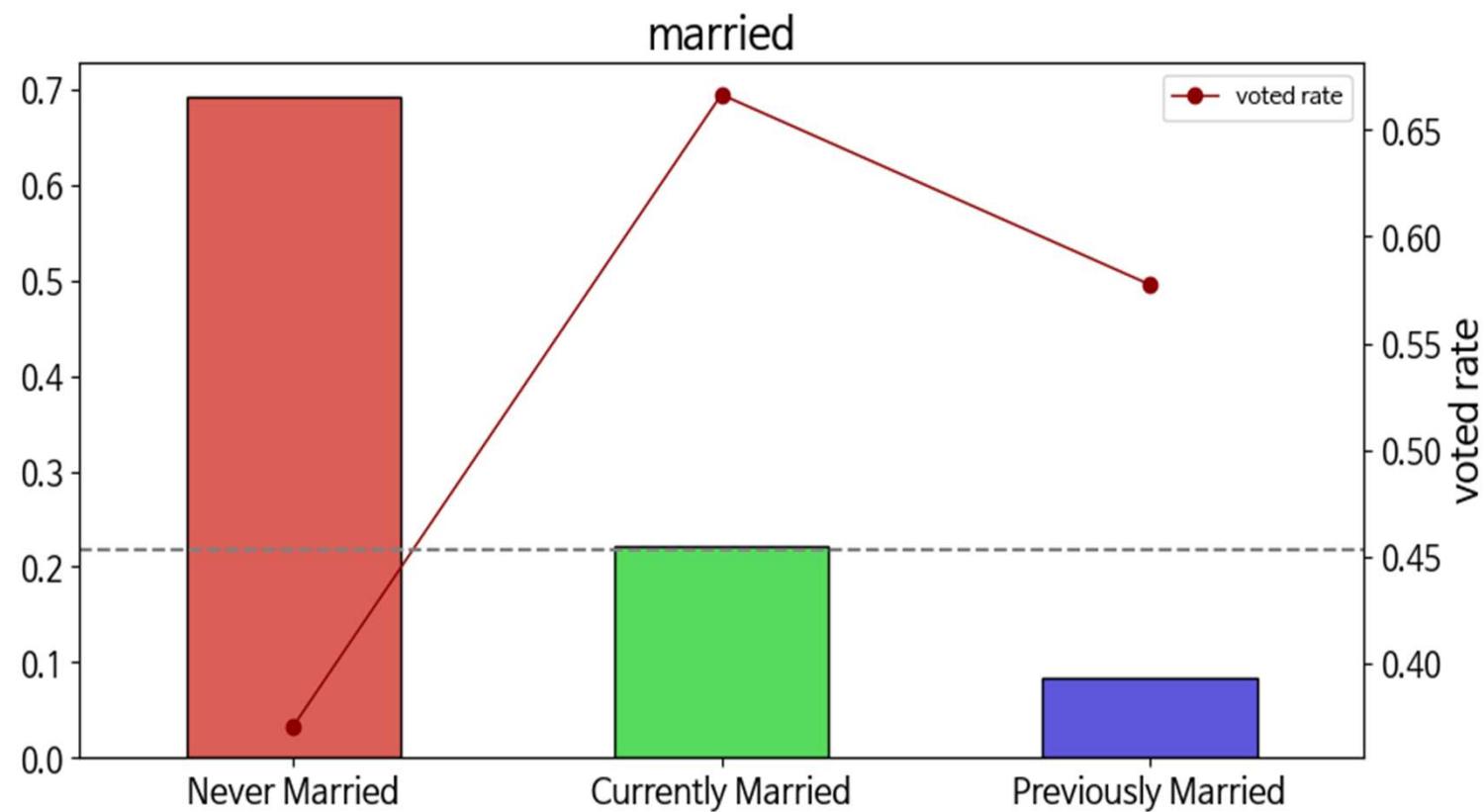
## 시각화-[hand]



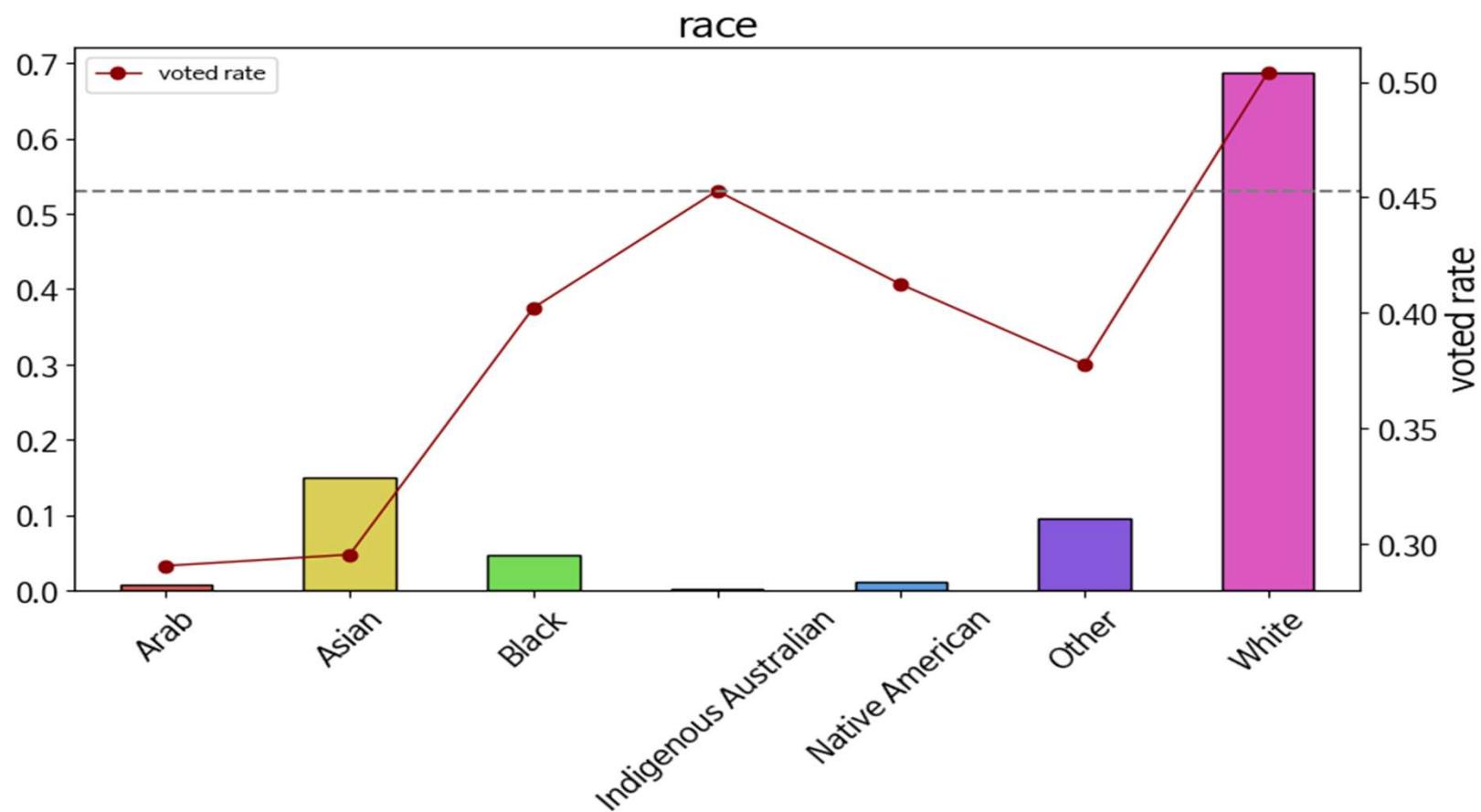
## 시각화-[married]



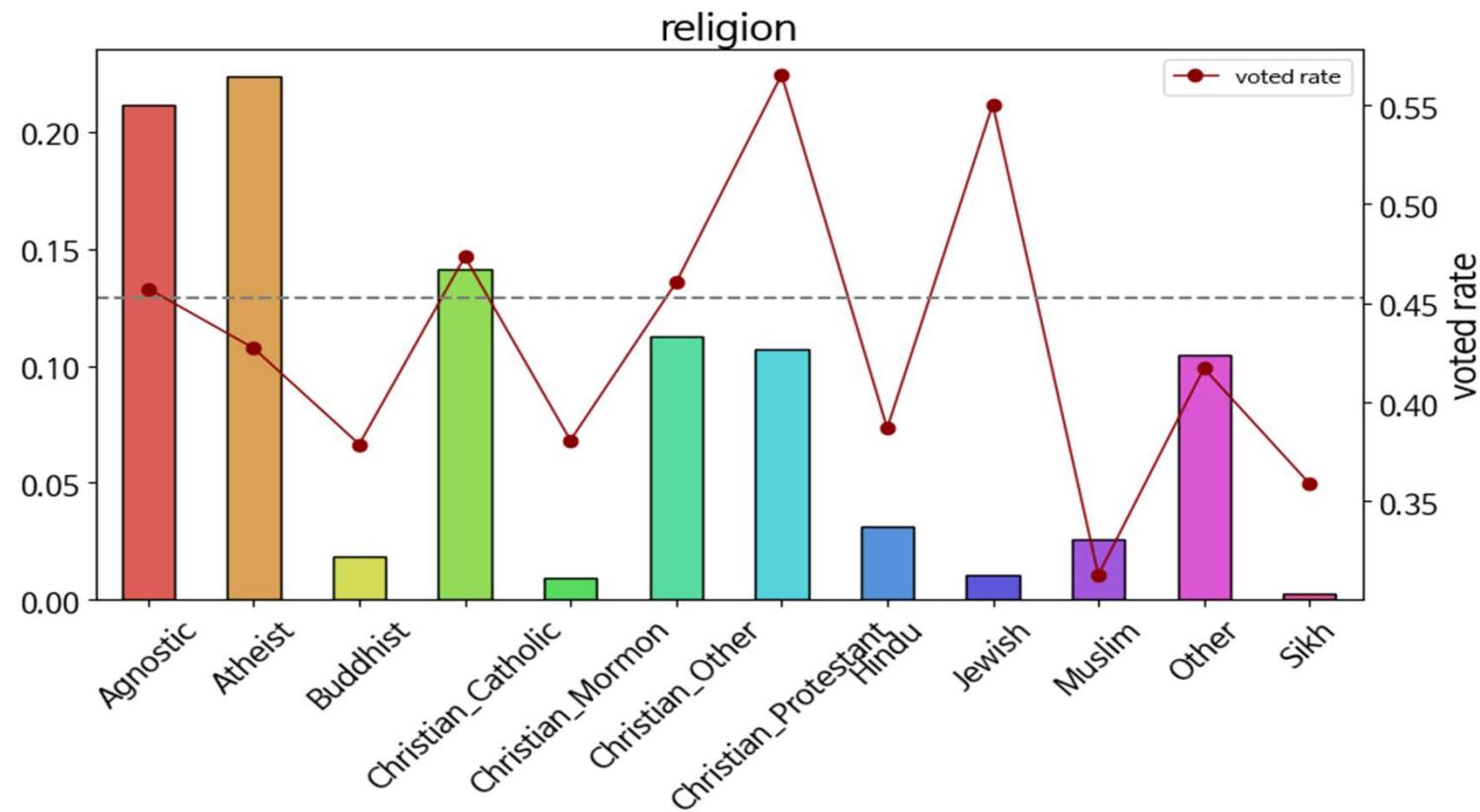
## 시각화-[married]



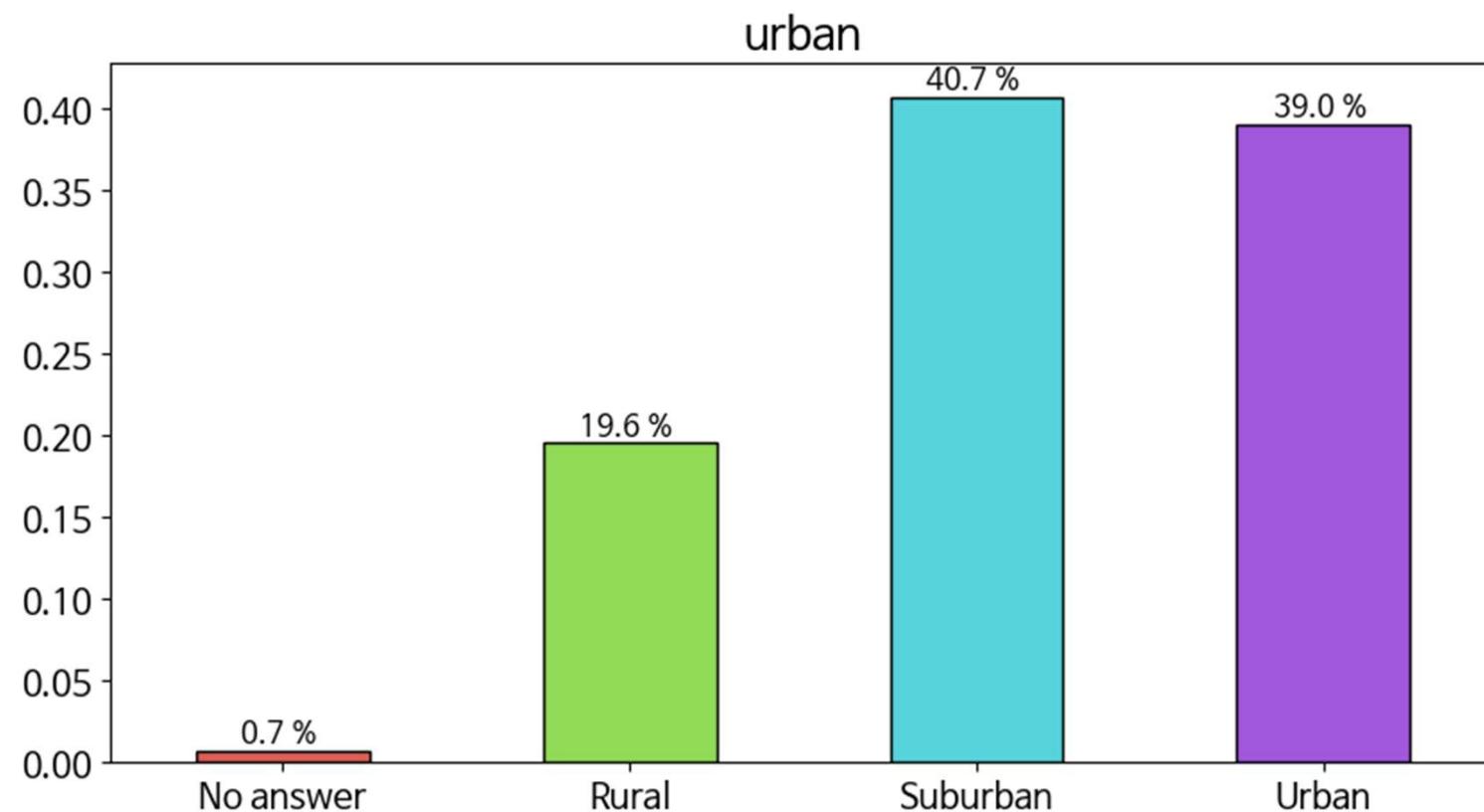
## 시각화-[race]



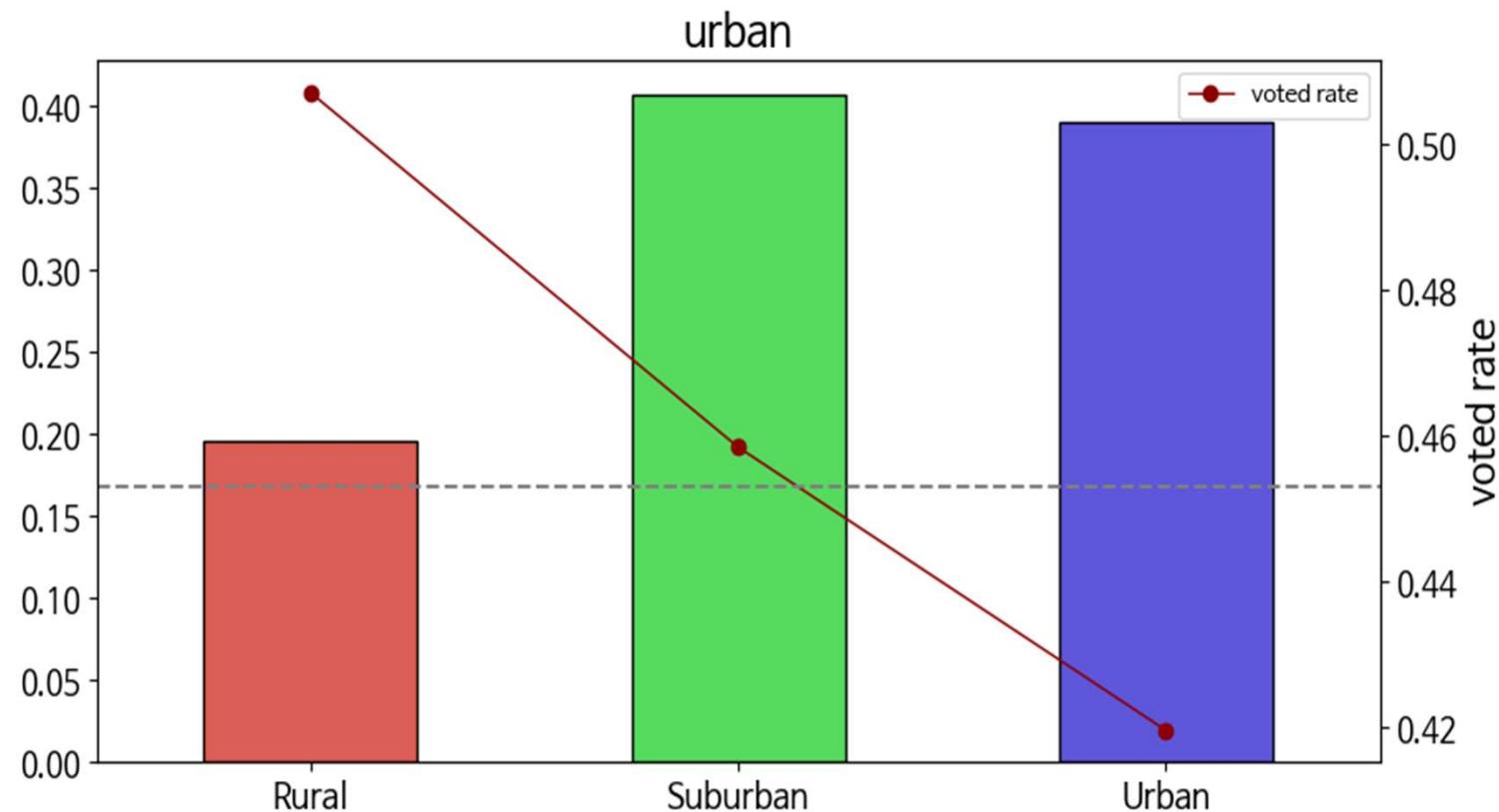
## 시각화-[religion]



## 시각화-[urban]

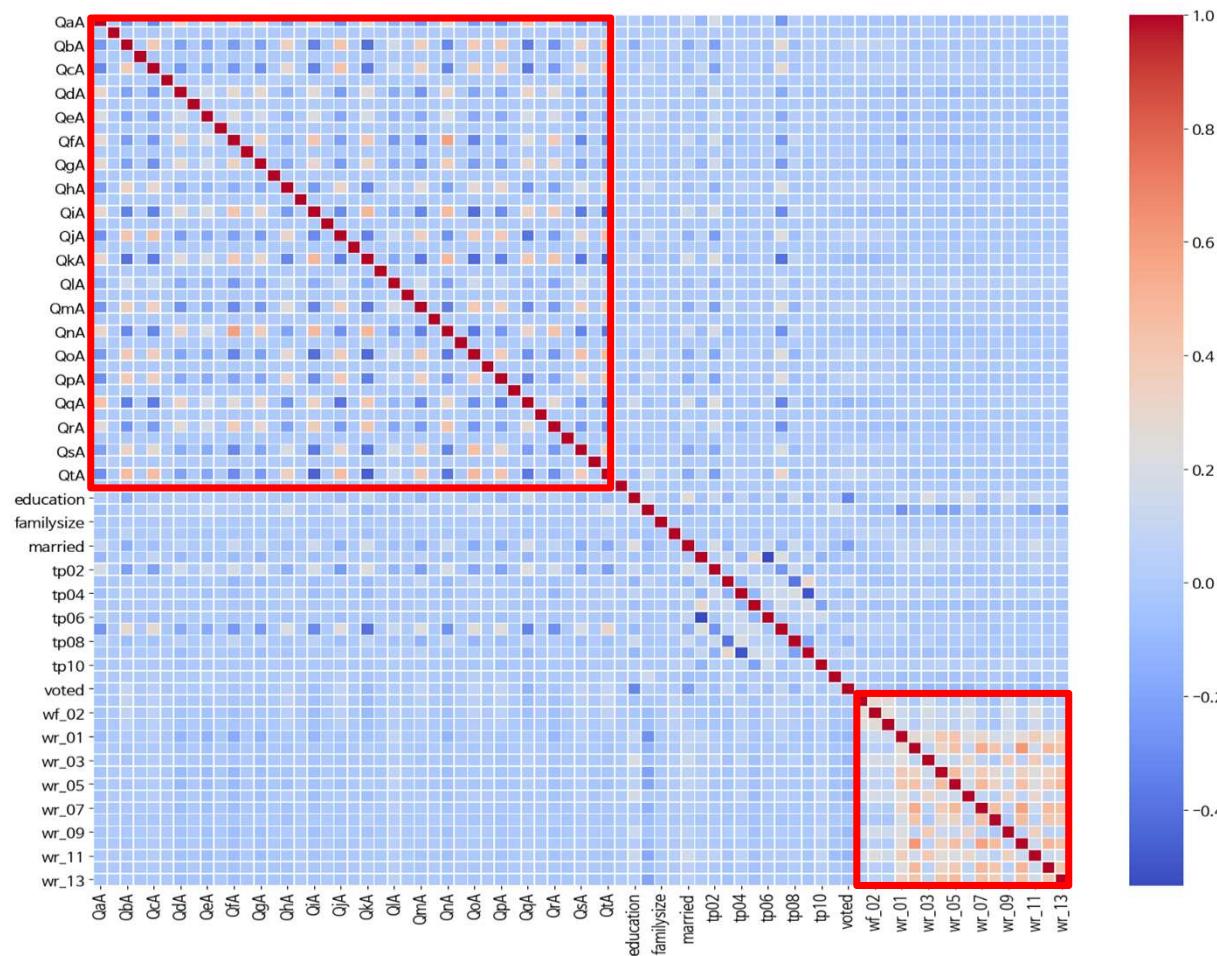


## 시각화-[urban]



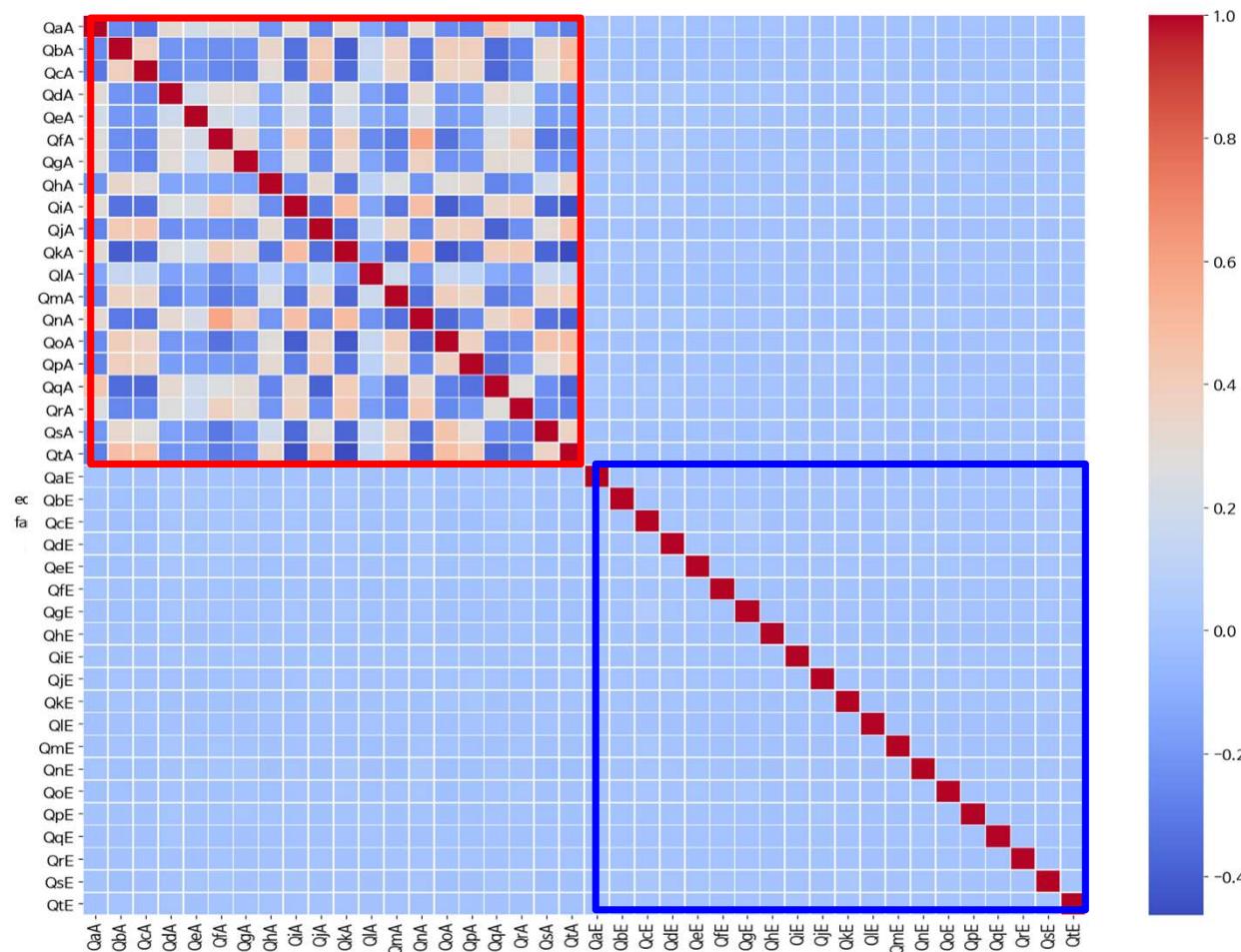
## 02 | EDA(탐색적 데이터 분석)

# 시각화



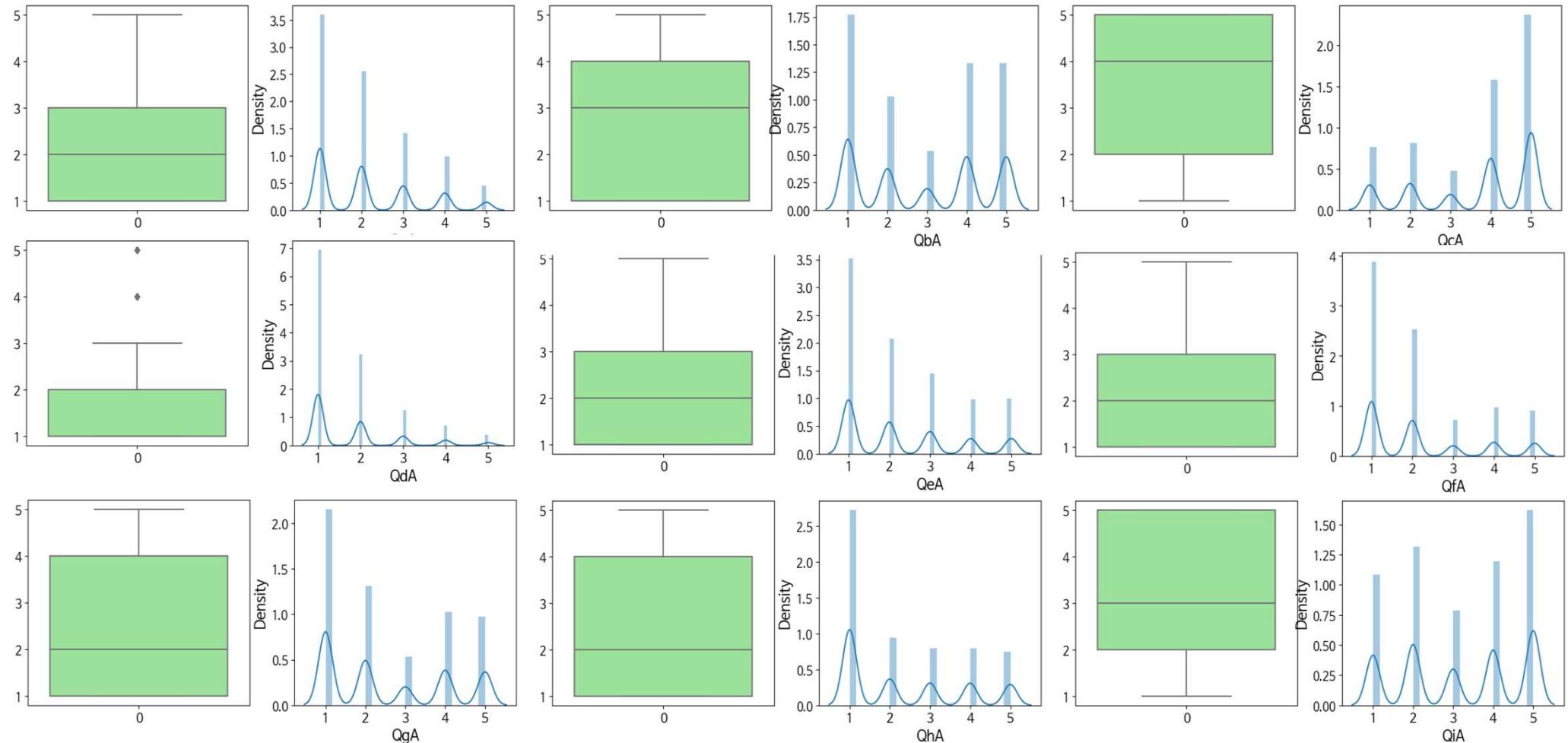
## 02 | EDA(탐색적 데이터 분석)

### 시각화



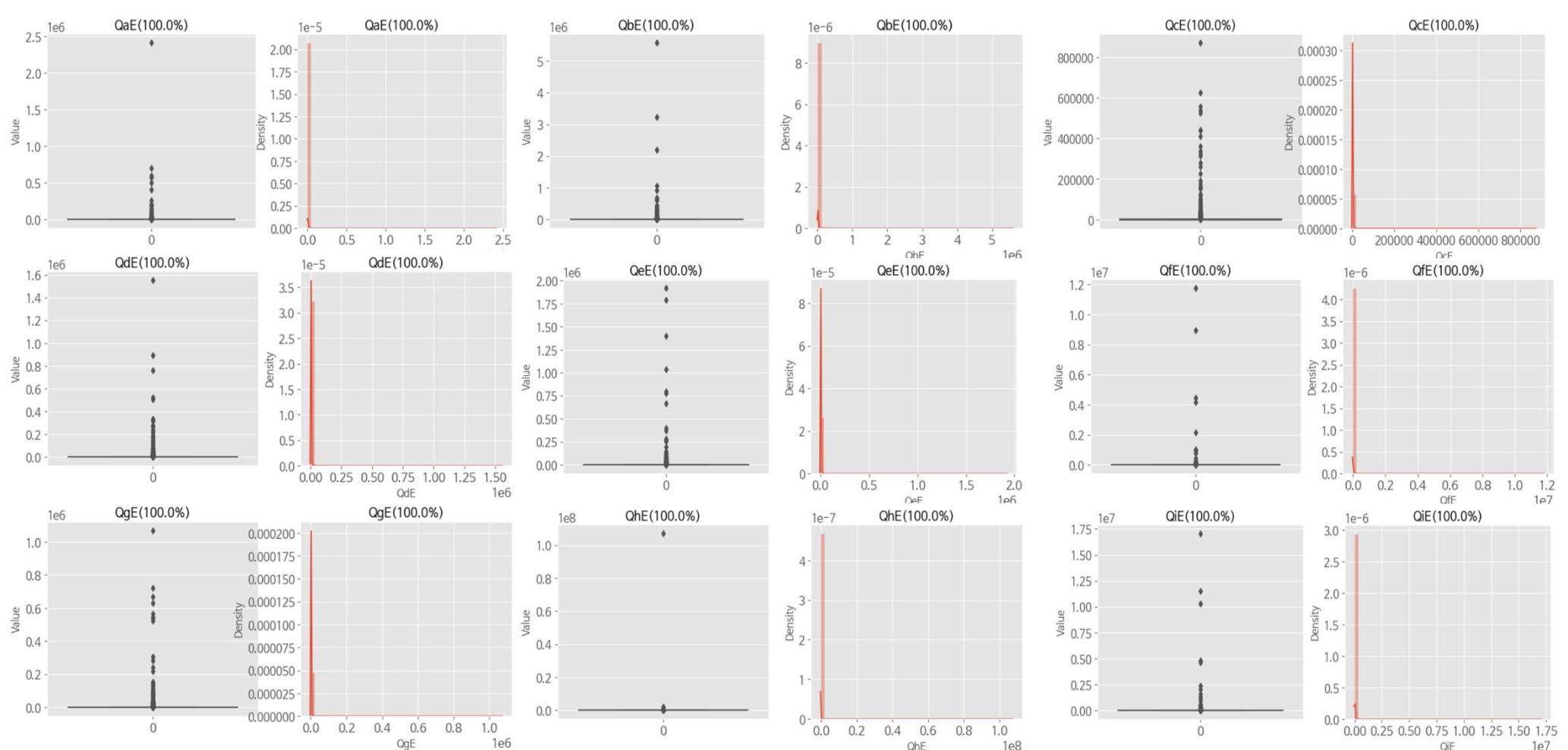
## 02 | EDA(탐색적 데이터 분석)

### 시각화-[QA]



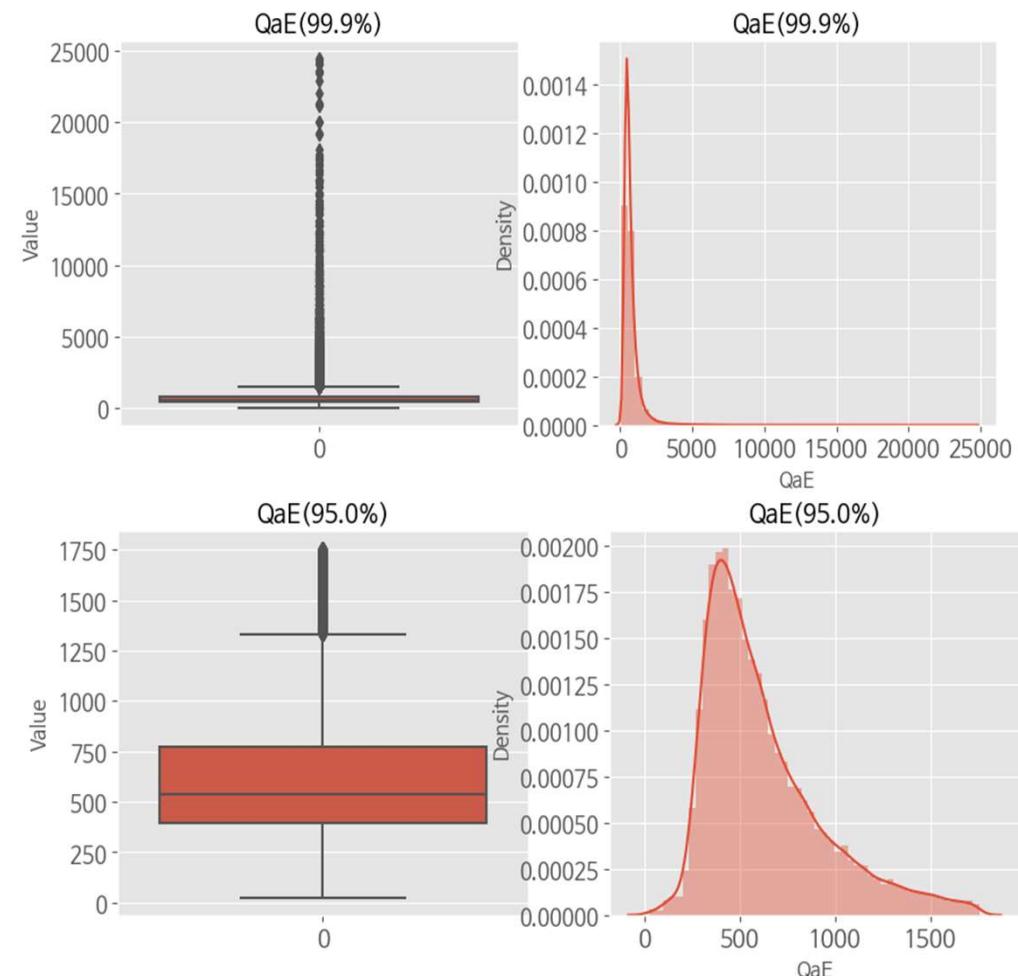
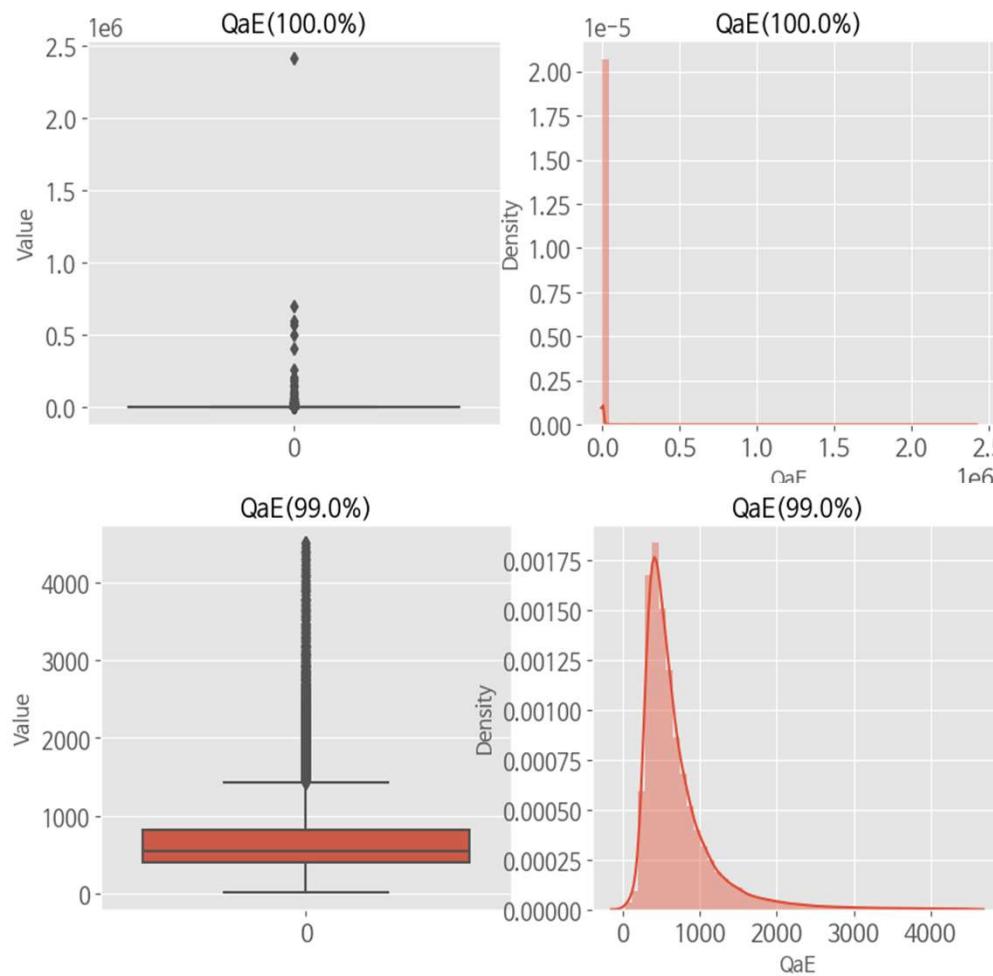
## 02 | EDA(탐색적 데이터 분석)

### 시각화-[QE]



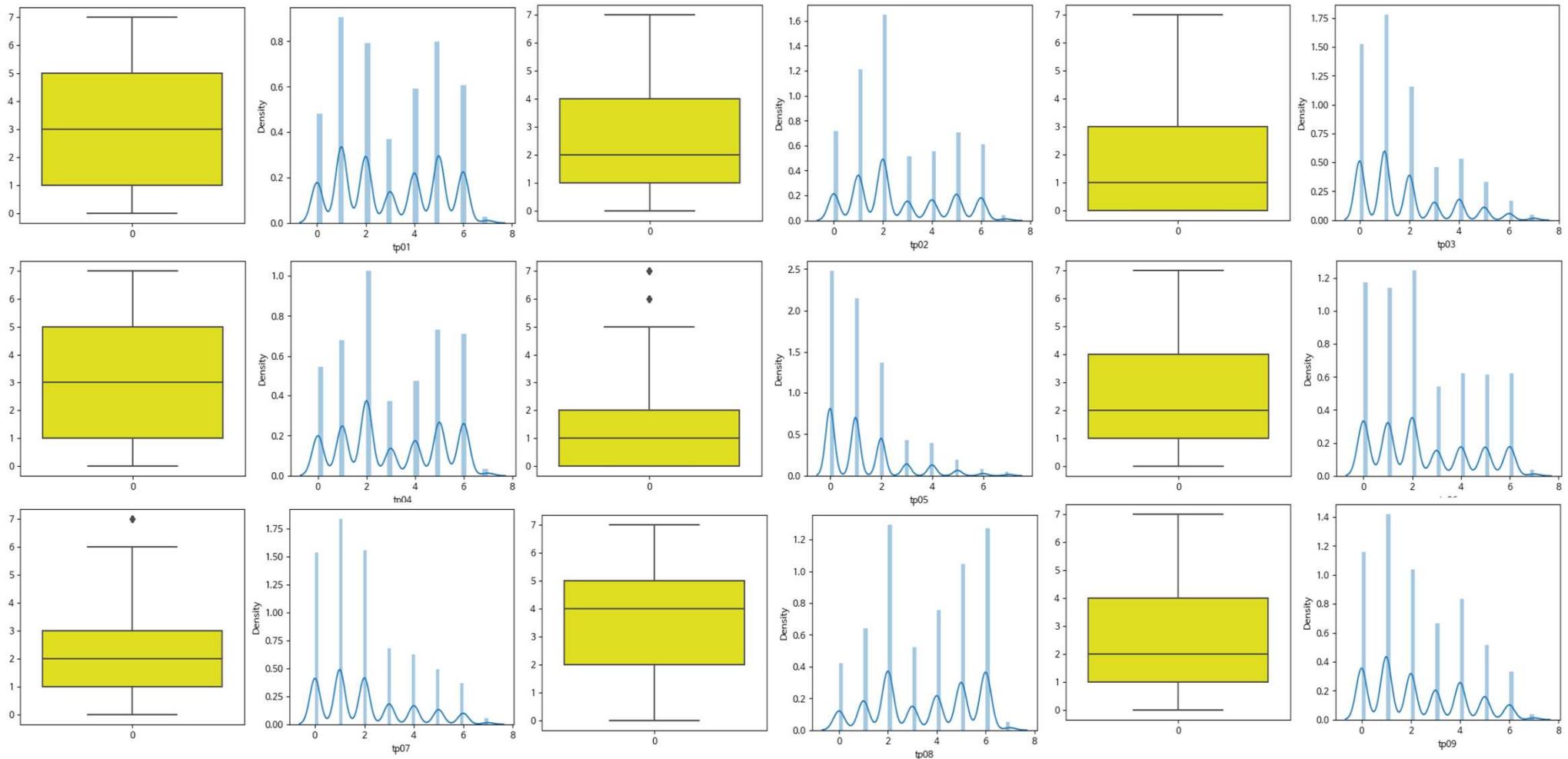
## 02 | EDA(탐색적 데이터 분석)

# 시각화-[QE]



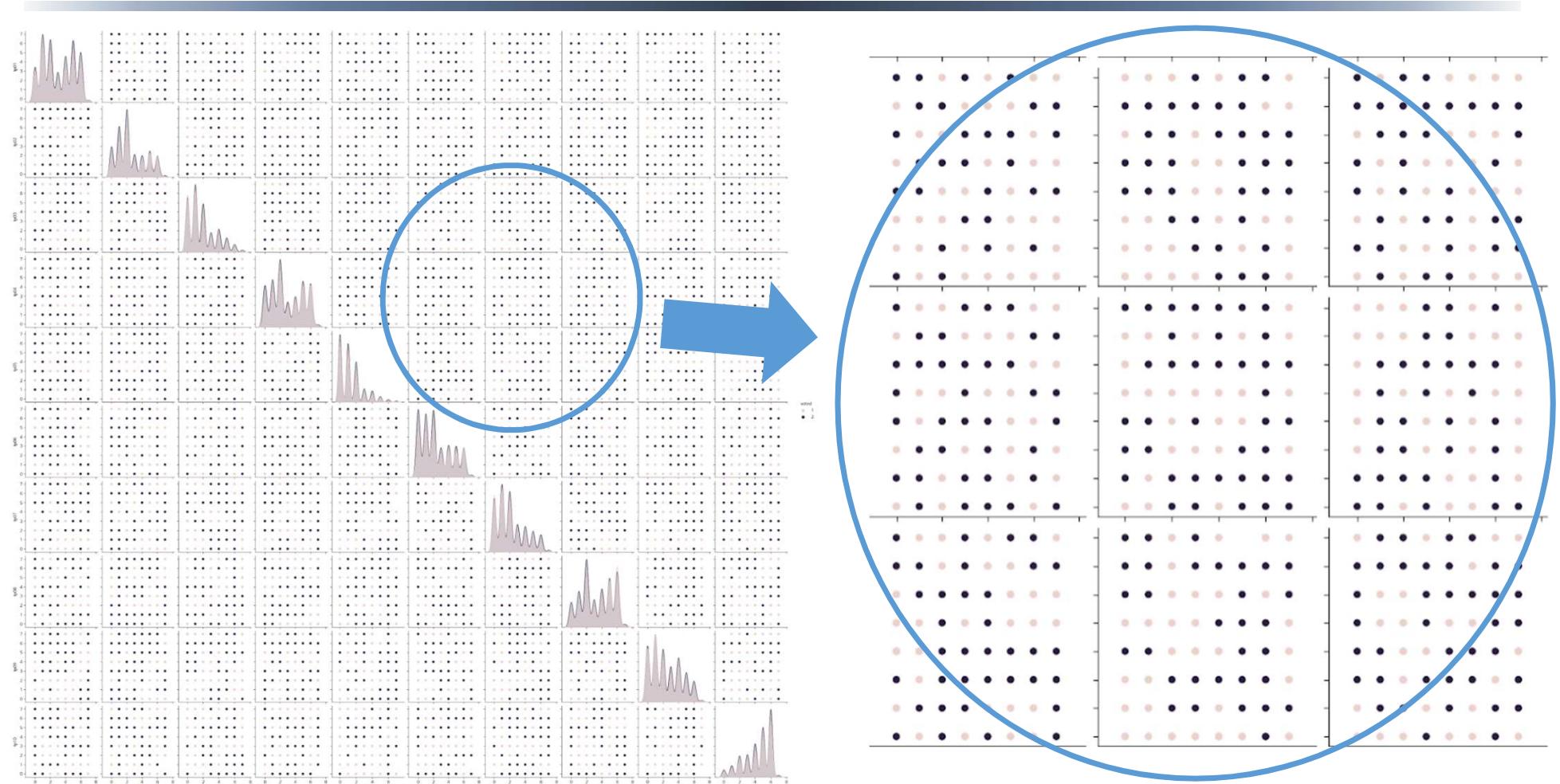
## 02 | EDA(탐색적 데이터 분석)

### 시각화-[tp]



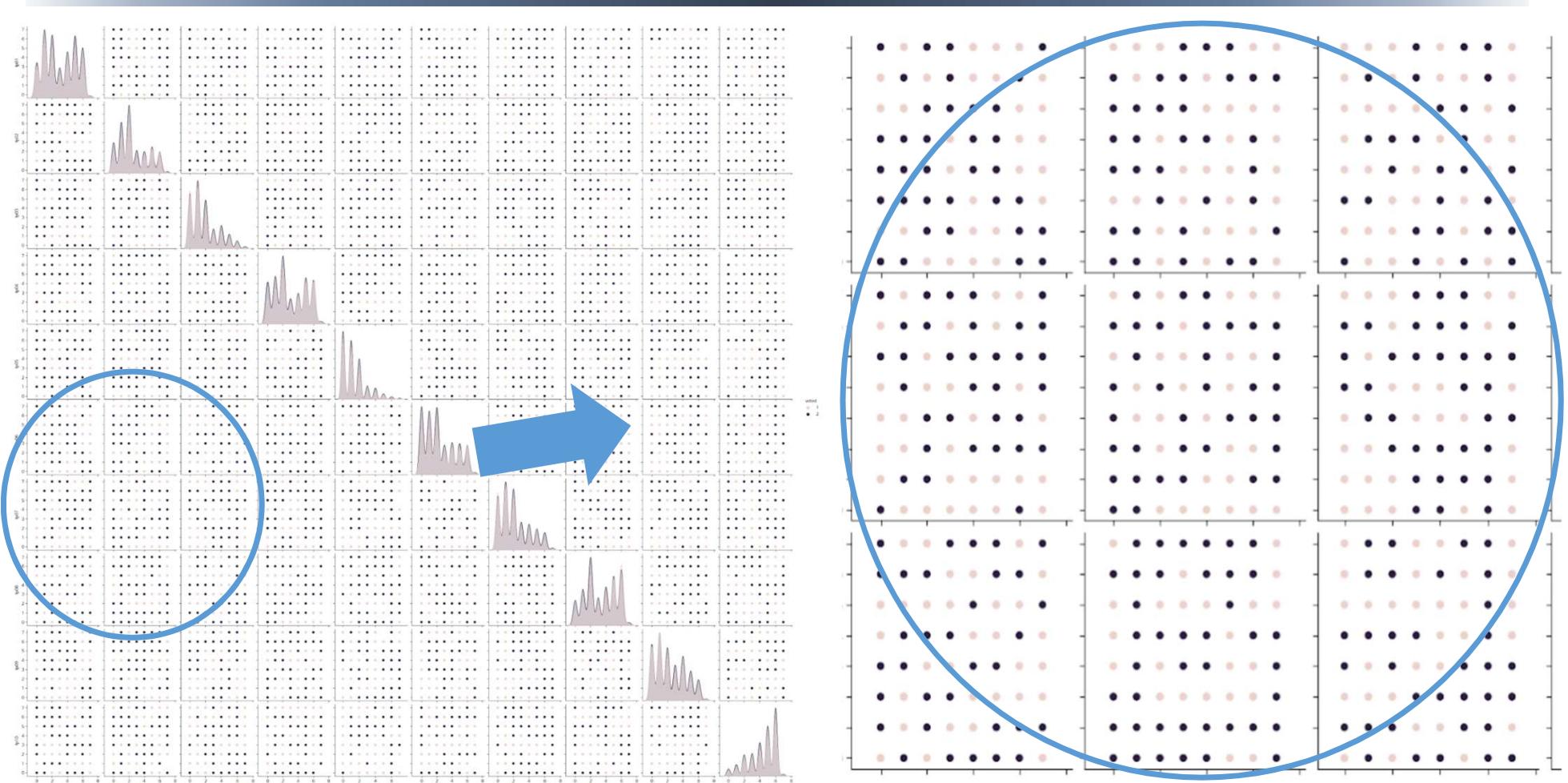
## 02 | EDA(탐색적 데이터 분석)

### 시각화-[tp]

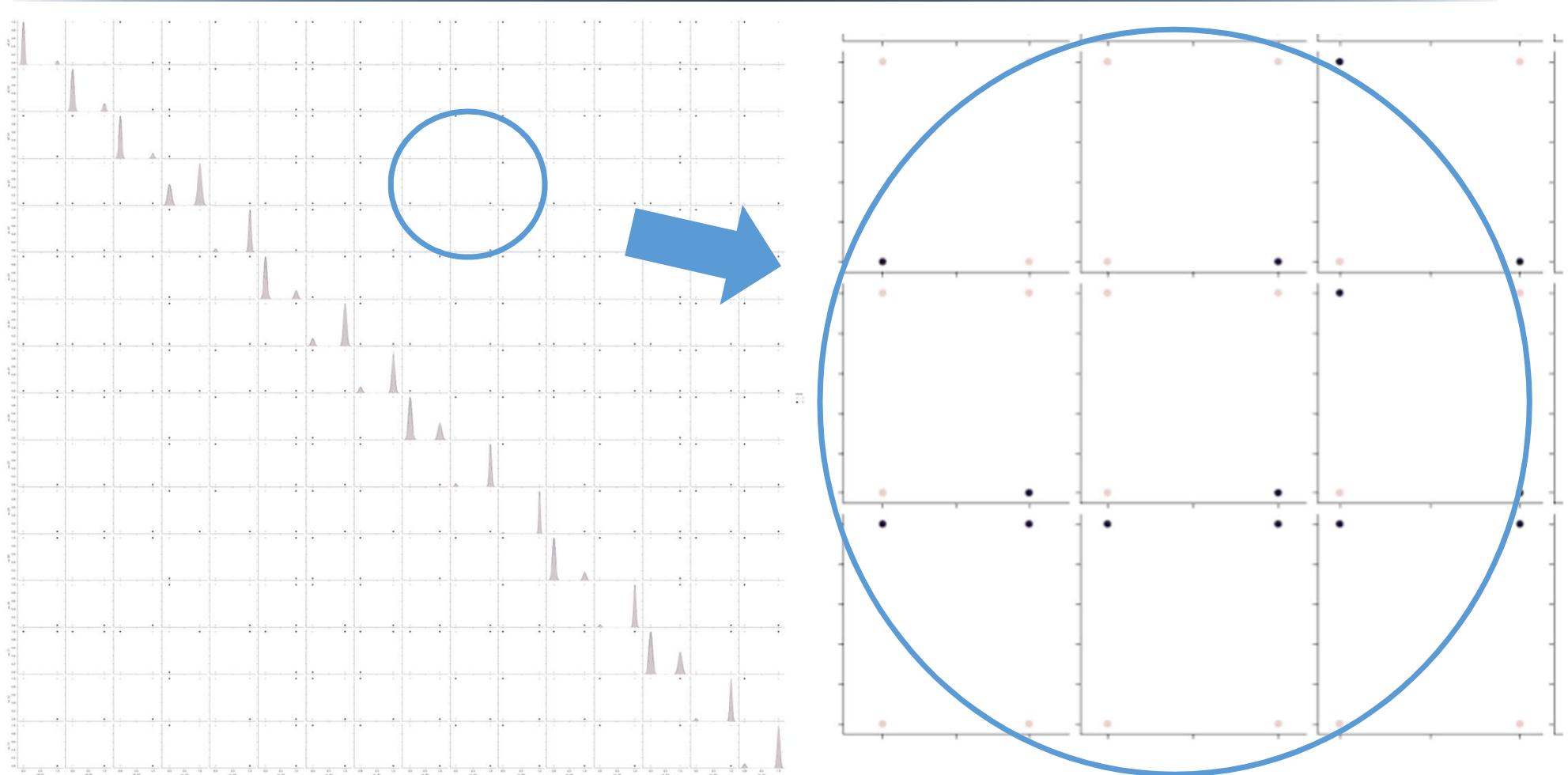


## 02 | EDA(탐색적 데이터 분석)

### 시각화-[tp]

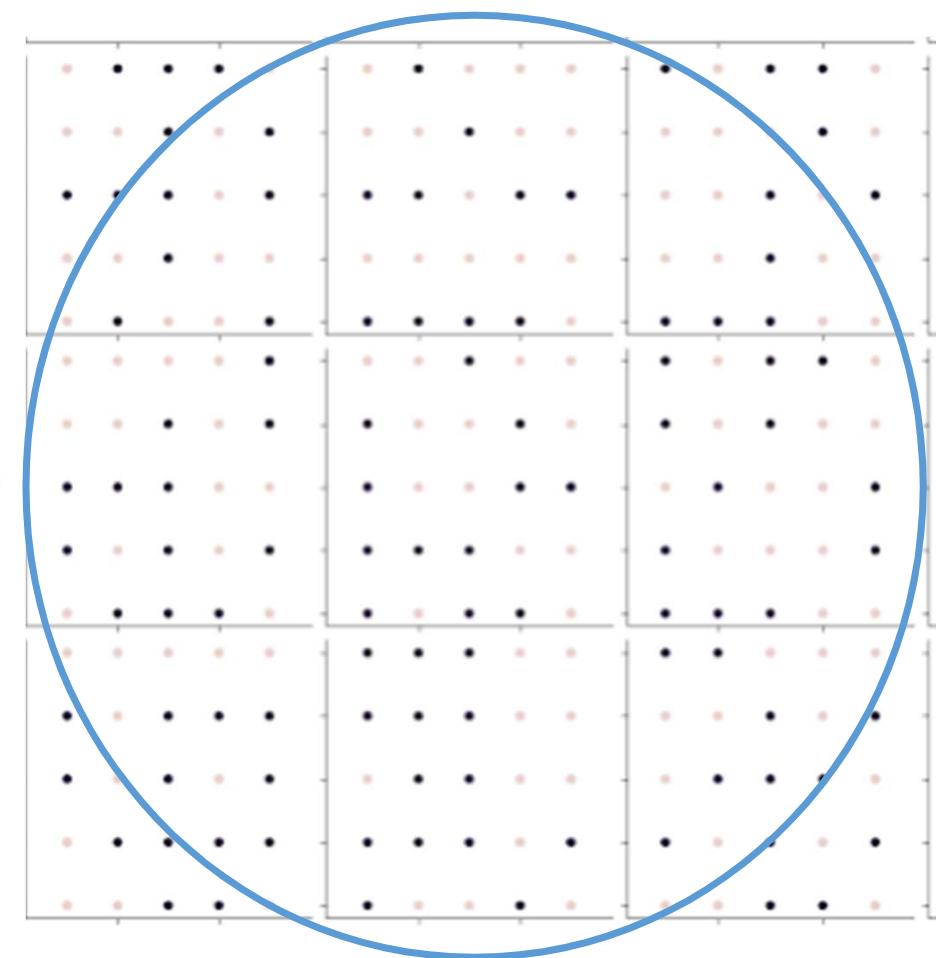
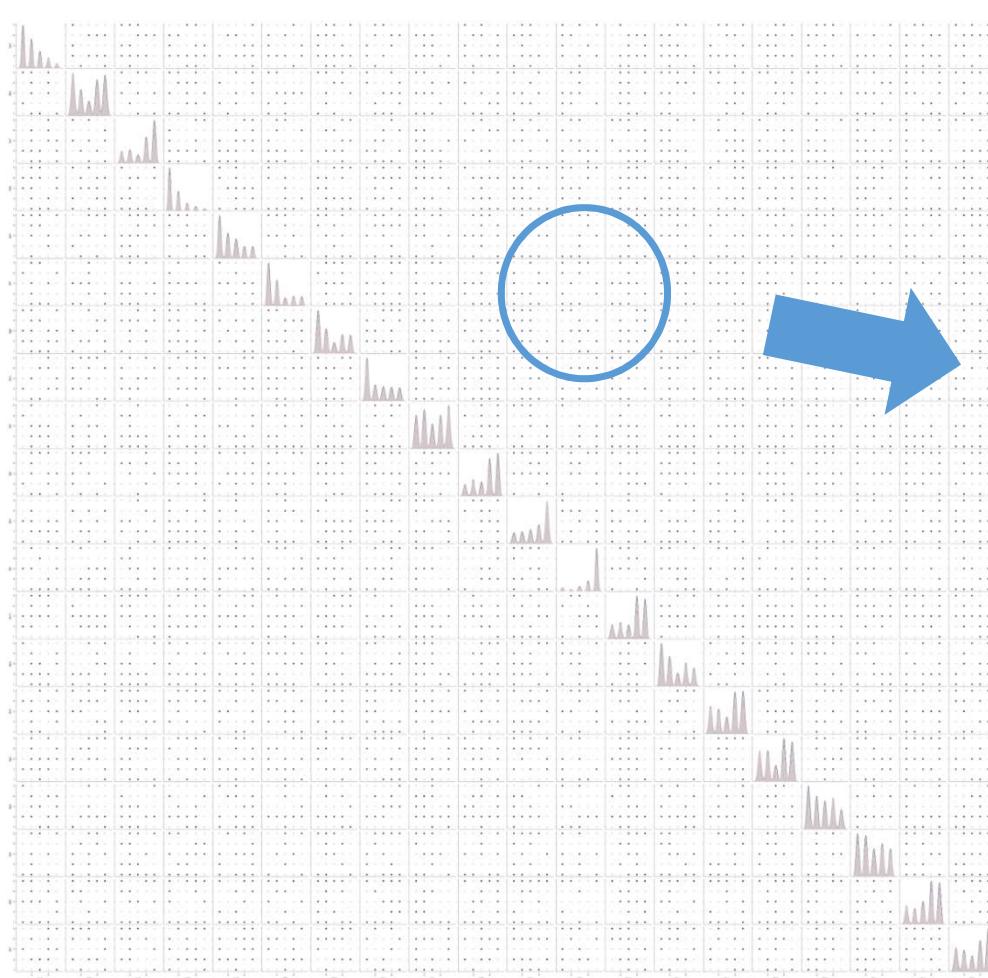


## 시각화-[wf]&[wr]



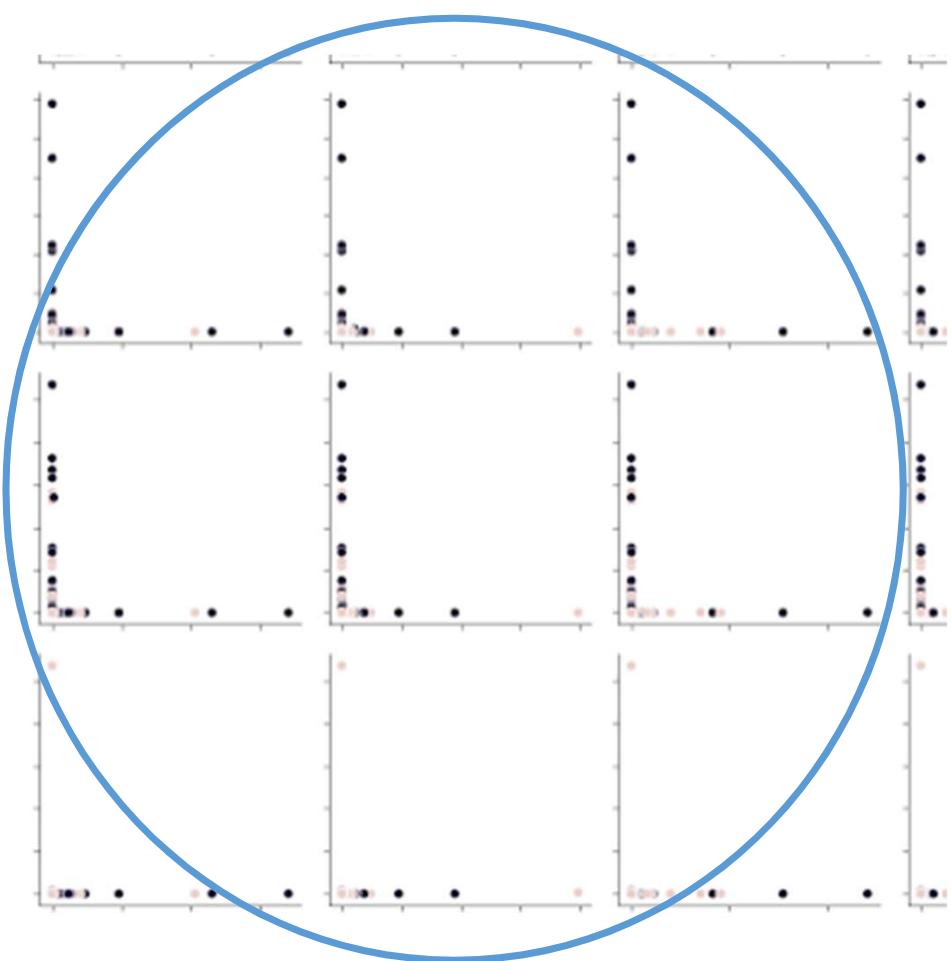
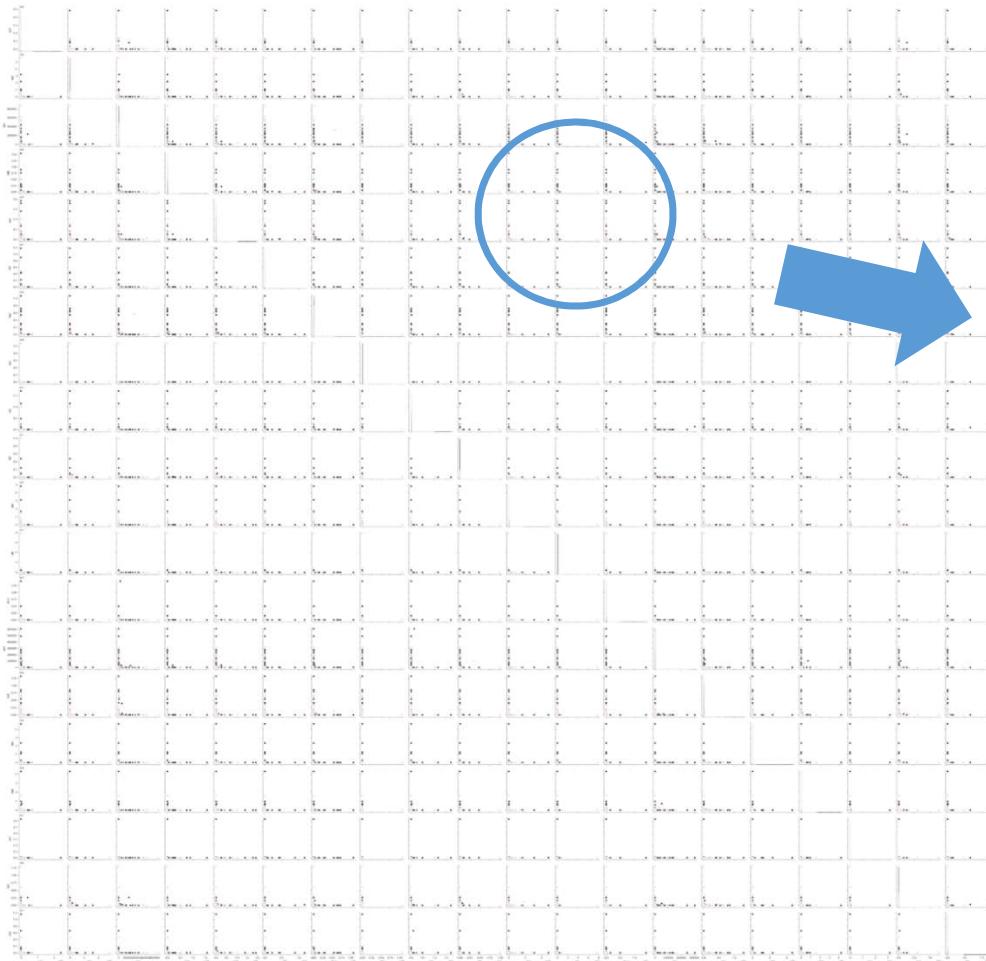
## 02 | EDA(탐색적 데이터 분석)

### 시각화-[QA]



## 02 | EDA(탐색적 데이터 분석)

### 시각화-[QE]



# 03. 데이터 전처리

# Raw data - 데이터 전처리

변수명	변경 전	변경 후	비고
age_group	10s, 20s, 30s, 40s, 50s, 60s, +70s	10, 20, 30, 40, 50, 60, 70	object형을 int로 변환
engnat	0(무응답), 1(모국어가 영어), 2(모국어 영어 X)	No(모국어 영어X), Yes(모국어가 영어)	0(비중小)을 1로 변환 one-hot 인코딩 준비
familysize	0, 1, 2, 3, 5, 6, 7, 8, 9, 10, …, 100, 999, …	0, 1, 2, 3, 4, 5, 6,	6이상(비중小)을 6로 변환
hand	1(Right), 2(Left), 3(Both)	Right, Left, Both	one-hot 인코딩 준비
married	0(other), 1(Never married) 2(Currently married) 3(Previously married)	other, Never married, Currently married, Previously married	one-hot 인코딩 준비
urban	0(무응답), 1(Rural), 2(Suburban), 3(Urban)	Rural, Suburban, Urban	0(비중小)을 2로 변환 one-hot 인코딩 준비

one-hot 인코딩 대상 범주형 변수: engnat(2), gender(2), hand(3), married(4), race(7), religion(12), urban(3)

feature의 총 개수: 102개

# Refined data - 데이터 전처리

## 삭제한 변수

삭제 변수	이유
engnat	EDA 분석 결과, 투표 여부에 유의미한 차이 없음.
hand	EDA 분석 결과, 투표 여부에 유의미한 차이 없음.
gender	EDA 분석 결과, 투표 여부에 유의미한 차이 없음.
QA(a~t)	마키아벨리즘 테스트 스코어 합산 방식에 따라 계산한 마키아벨리즘 스코어(Mach_score)를 변수에 추가
QE(a~t)	EDA 분석 결과, 투표 여부에 유의미한 차이 없음.
tp(01~10)	tipi 테스트 스코어 합산 방식에 따라 5가지 성격 유형에 대한 스코어를 변수에 추가
wf(01~03)	3개 항목의 평균값(wf_mean)을 변수에 추가
wr(01~13)	13개 항목의 평균값(wr_mean)을 변수에 추가

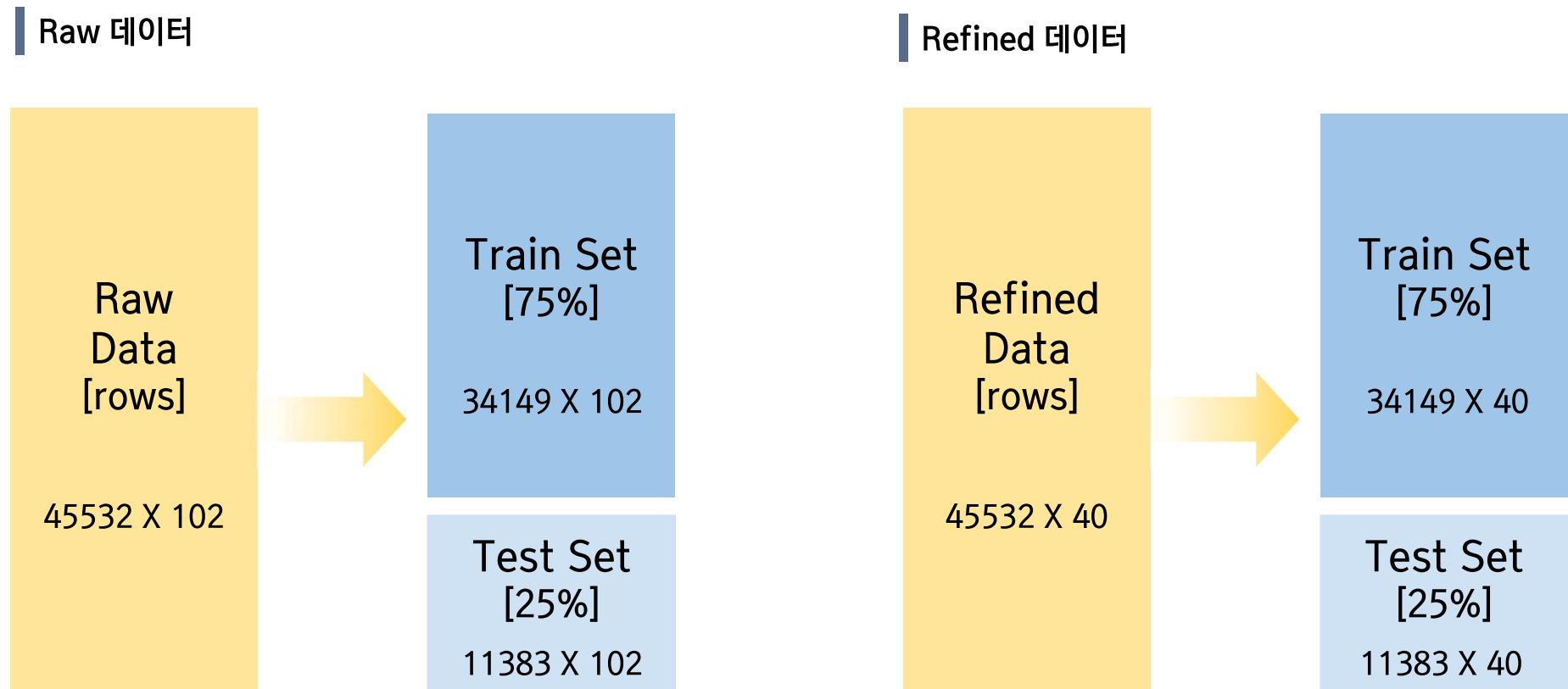
# Refined data - 데이터 전처리

## 추가한 변수

추가 변수	설명	추가 이유
Mach_score	마키아벨리즘 테스트 종합 스코어	개별 질문에 대한 응답을 '종합 점수'로 표현
Extraversion	TIPI 테스트 - 외향성	개별 질문에 대한 응답을 '외향성 점수'로 표현
Agreeableness	TIPI 테스트 - 친화성	개별 질문에 대한 응답을 '친화성 점수'로 표현
Conscientiousness	TIPI 테스트 - 성실성	개별 질문에 대한 응답을 '성실성 점수'로 표현
EmotionalStability	TIPI 테스트 - 정서적 안정성	개별 질문에 대한 응답을 '정서적 안정성 점수'로 표현
Openness	TIPI 테스트 - 경험 개방성	개별 질문에 대한 응답을 '경험 개방성 점수'로 표현
wf_mean	wf 3개 항목의 평균값	wf 개별 항목이 유의미하지 않을 것이라 생각하여 평균값 도출
wr_mean	wr 13개 항목의 평균값	wr 개별 항목이 유의미하지 않을 것이라 생각하여 평균값 도출

Refined data - feature의 총 개수: 38개

## Train-Test-Split



# Scaler

## Standard Scaler

```
from sklearn.preprocessing import StandardScaler  
  
def _standard(train, test):  
  
    scaler = StandardScaler()  
    scaler.fit(train)  
    X_train_scaled = scaler.transform(train)  
    X_test_scaled = scaler.transform(test)  
  
    return X_train_scaled, X_test_scaled
```

## Minmax Scaler

```
from sklearn.preprocessing import MinMaxScaler  
  
def _minmax(train, test):  
  
    scaler = MinMaxScaler()  
    scaler.fit(train)  
    X_train_scaled = scaler.transform(train)  
    X_test_scaled = scaler.transform(test)  
  
    return X_train_scaled, X_test_scaled
```

## Robust Scaler

```
from sklearn.preprocessing import RobustScaler  
  
def _robust(train, test):  
  
    scaler = RobustScaler()  
    scaler.fit(train)  
    X_train_scaled = scaler.transform(train)  
    X_test_scaled = scaler.transform(test)  
  
    return X_train_scaled, X_test_scaled
```

## 차원 축소

고차원 데이터 특성을 저차원 데이터의 특징으로 표현하는 작업

투영

PCA

PCA (0.9)

PCA(0.95)

PCA(0.99)

매니폴드

LLE

가까운 k개의 이웃들 ( $n_{neighbors}$ )로부터  
선형적으로 연관된 정도를 측정하여, 관계가  
가장 잘 보존되는 저차원 표현  
( $n_{components}$ )으로 축소



과도한 소요 시간  
낮은 성능

기타

t-SNE

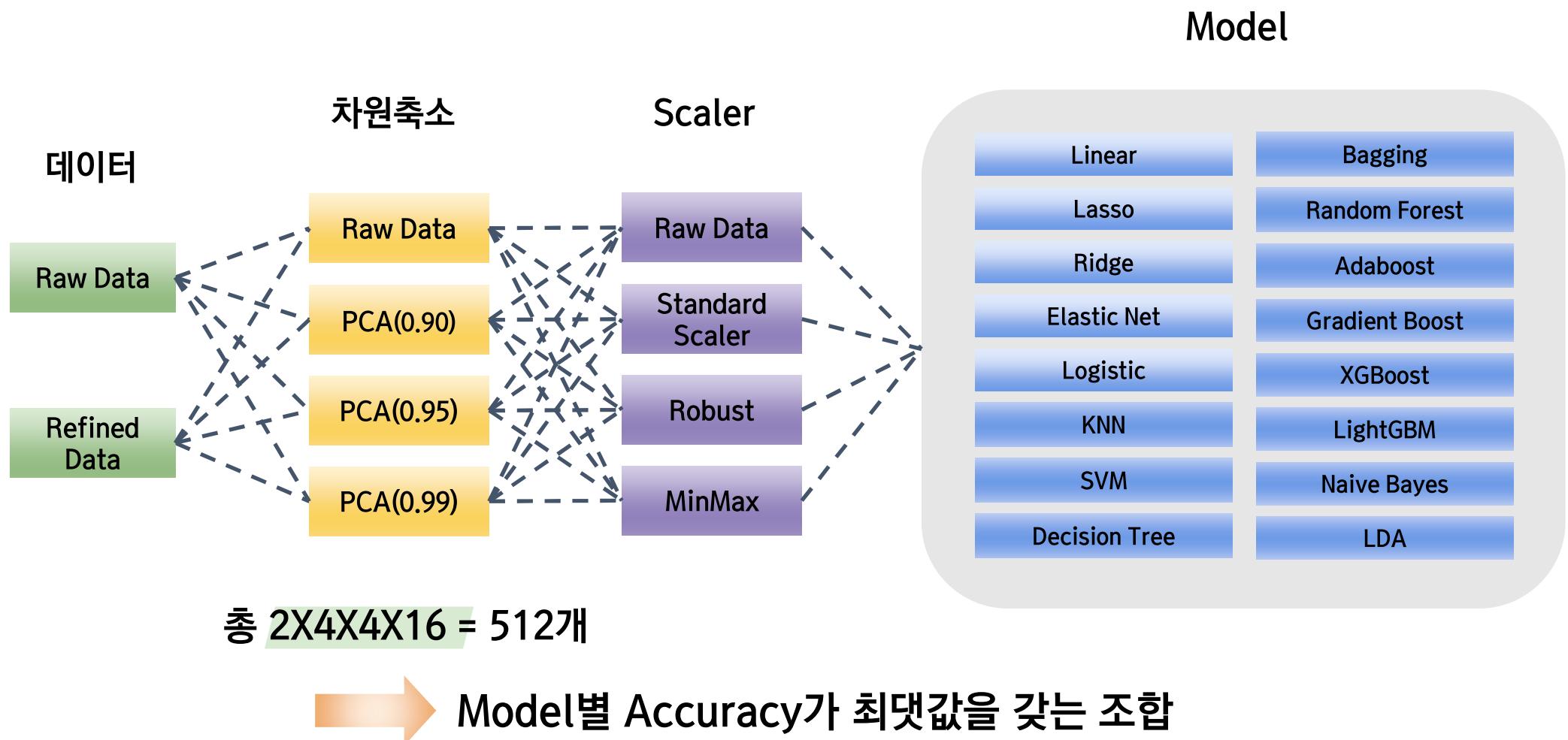
비슷한 샘플은 가까이, 비슷하지 않은 샘플은 멀리 떨어지도록 하면서  
차원 축소



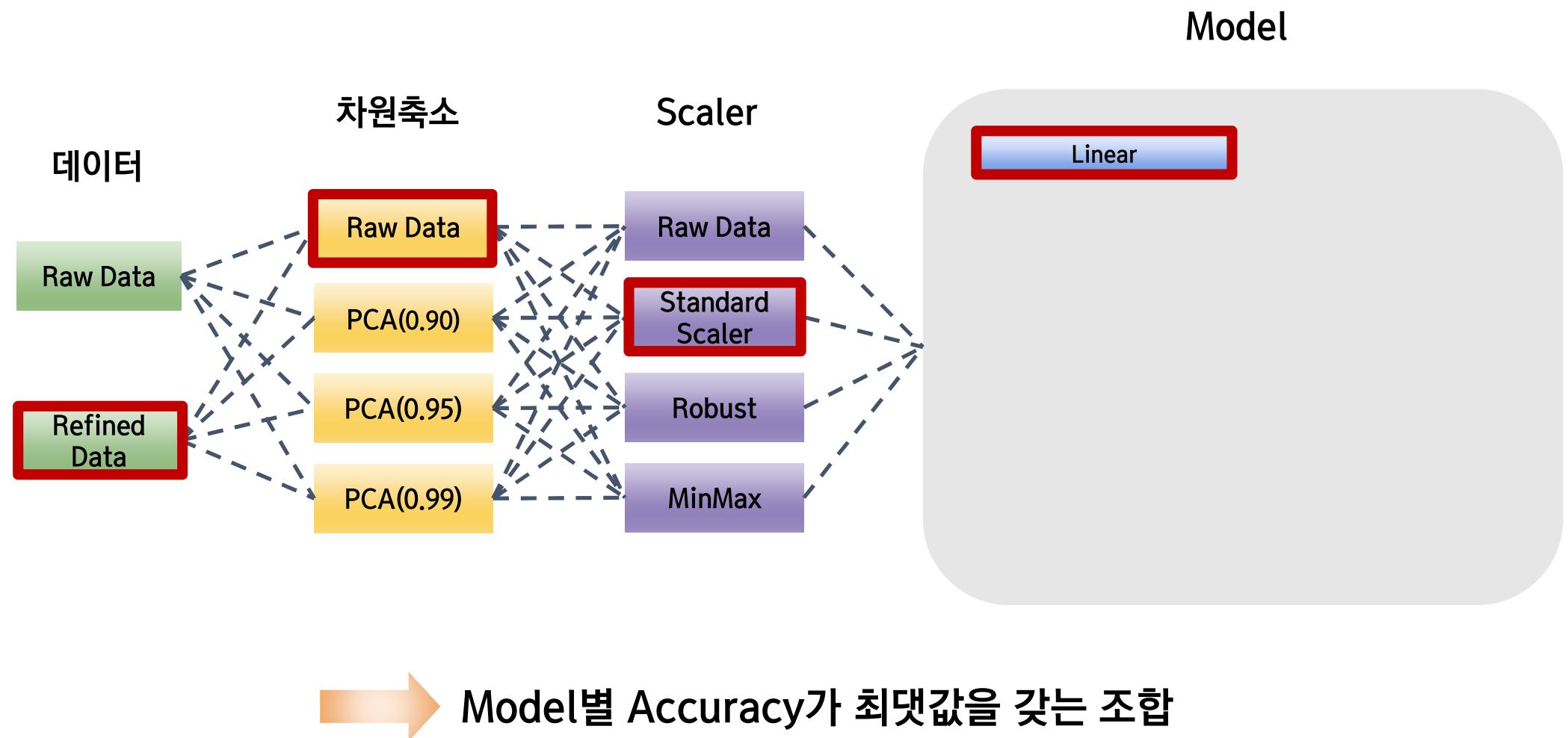
과도한 소요 시간  
낮은 성능

# 04. Modeling

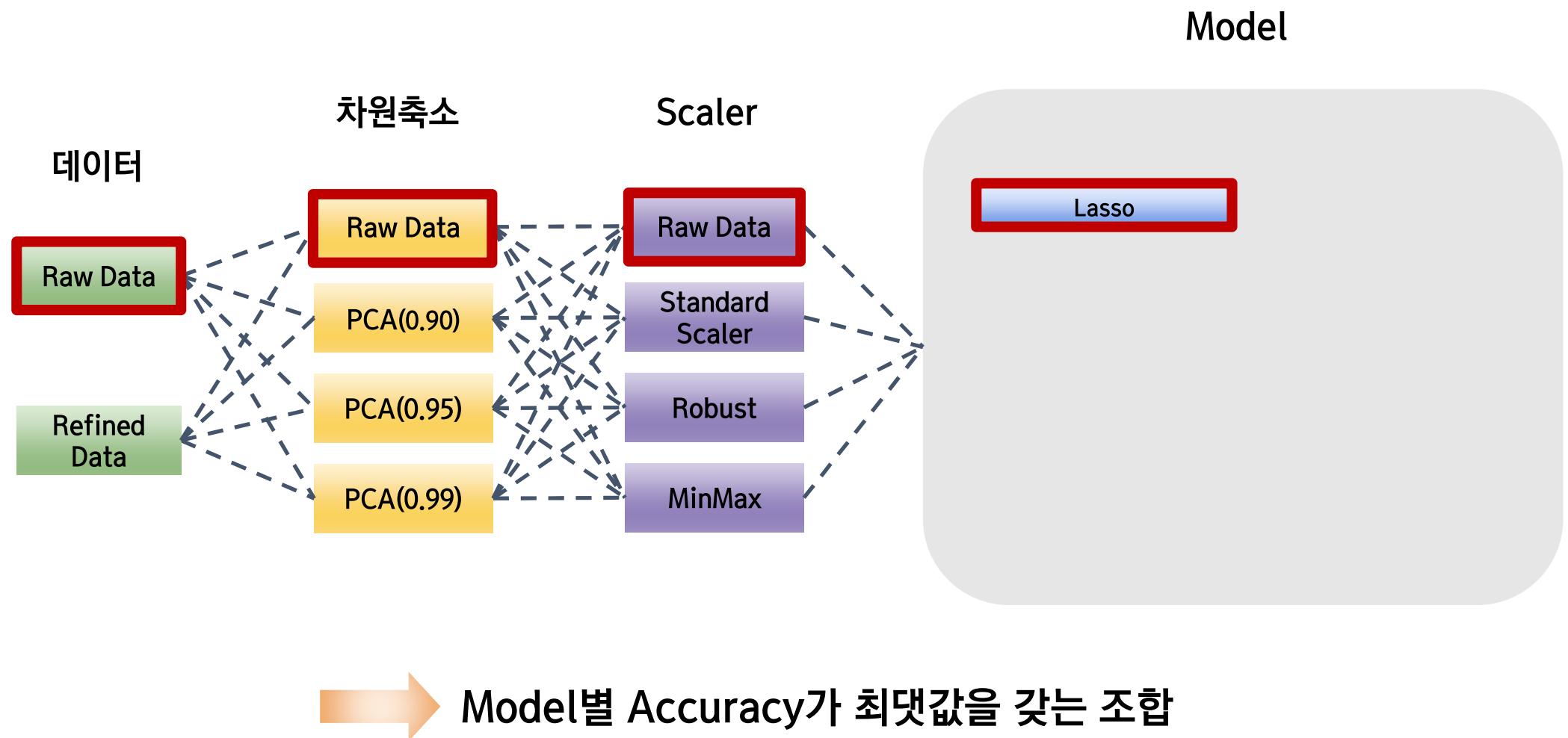
## 04 | Modeling



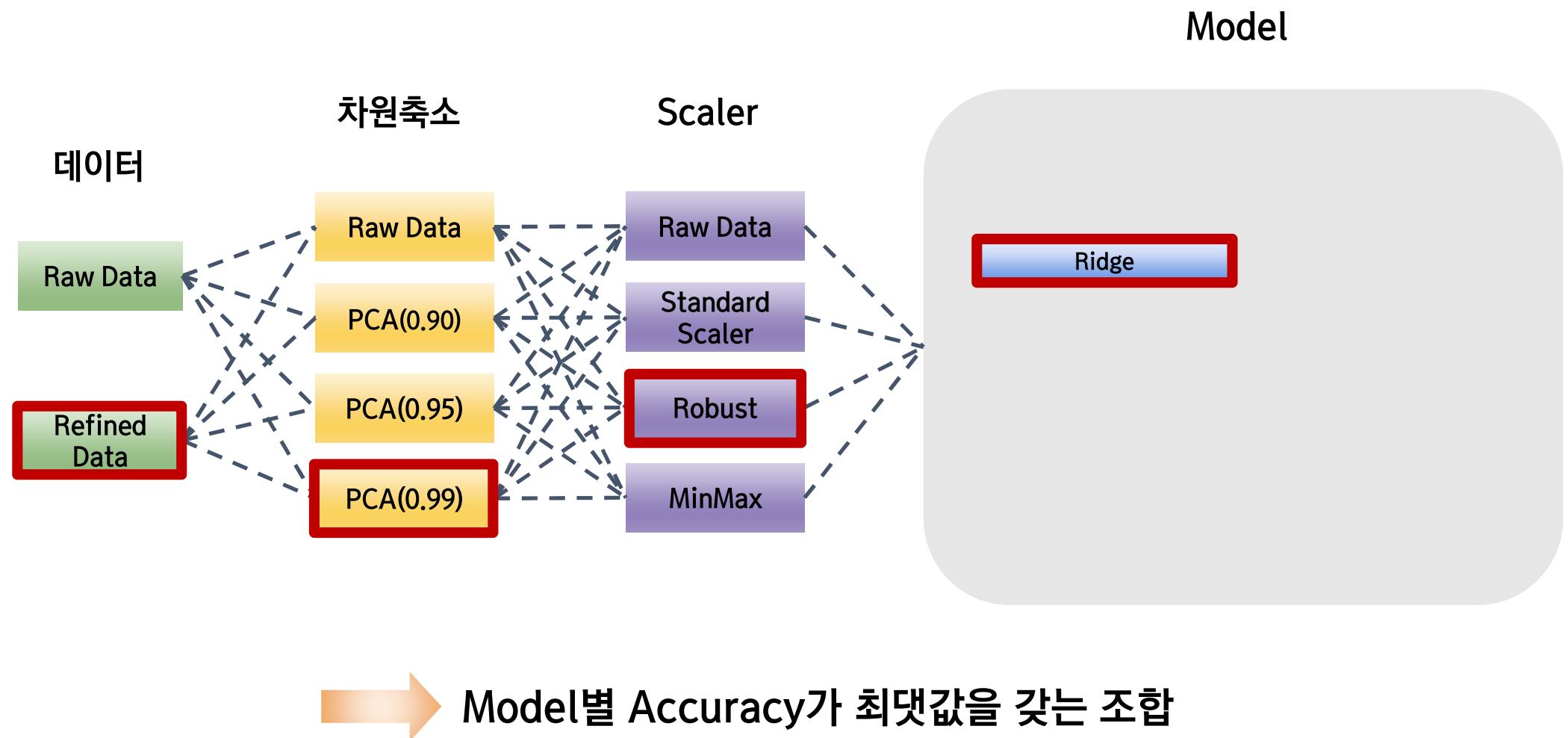
## 04 | Modeling



## 04 | Modeling



## 04 | Modeling



## 04 | Modeling

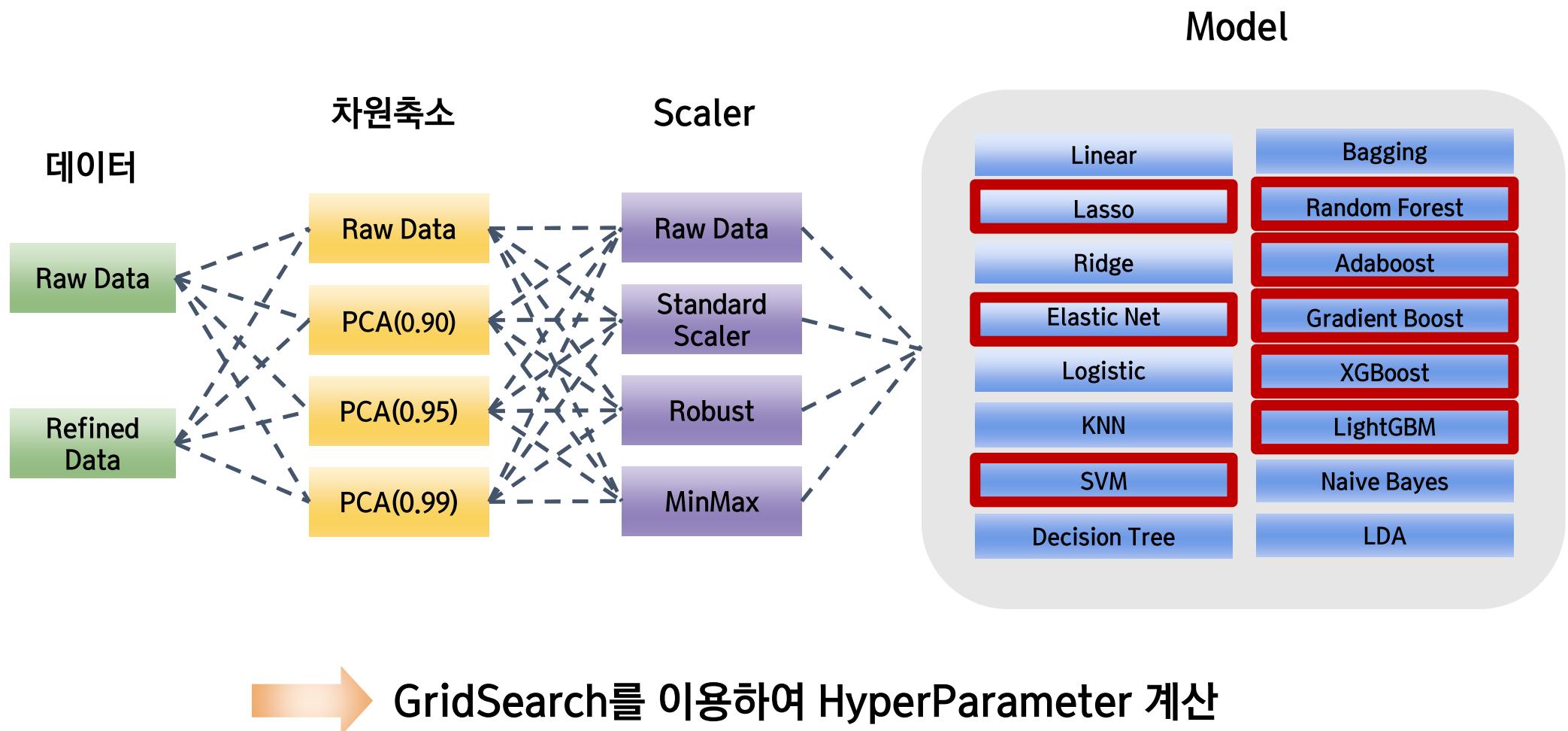
```
1 print_max_list(max_list)

*****
*****          SUMMARY          *****
*****

[ LinearRegression      / StandardScaler / Processed Data / PCA( X ) ] f1 score : 0.6279, acc : 0.6760, auc_score : 0.6699
[ Ridge                 / RobustScaler  / Processed Data / PCA( X ) ] f1 score : 0.6259, acc : 0.6745, auc_score : 0.6683
[ Lasso                 / NoScaler     / Raw Data       / PCA( X ) ] f1 score : 0.4197, acc : 0.6147, auc_score : 0.5872
[ ElasticNet            / NoScaler     / Processed Data / PCA(99%) ] f1 score : 0.5800, acc : 0.6482, auc_score : 0.6391
[ LogisticRegression    / NoScaler     / Processed Data / PCA( X ) ] f1 score : 0.6317, acc : 0.6779, auc_score : 0.6721
[ GaussianNB            / RobustScaler / Processed Data / PCA(99%) ] f1 score : 0.6228, acc : 0.6654, auc_score : 0.6606
[ LinearDiscriminantAnalysis / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6260, acc : 0.6744, auc_score : 0.6683
[ LinearSVC             / NoScaler     / Processed Data / PCA( X ) ] f1 score : 0.7173, acc : 0.6826, auc_score : 0.6978
[ DecisionTreeClassifier / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.5740, acc : 0.6159, auc_score : 0.6120
[ KNeighborsClassifier   / MinMaxScaler / Processed Data / PCA(99%) ] f1 score : 0.6239, acc : 0.6552, auc_score : 0.6528
[ RandomForestClassifier / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6792, acc : 0.6924, auc_score : 0.6938
[ AdaBoostClassifier    / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6841, acc : 0.6929, auc_score : 0.6954
[ GradientBoostingClassifier / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6908, acc : 0.6963, auc_score : 0.6997
[ BaggingClassifier     / NoScaler     / Raw Data       / PCA( X ) ] f1 score : 0.6021, acc : 0.6547, auc_score : 0.6479
[ LGBMClassifier        / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6928, acc : 0.6951, auc_score : 0.6993
[ XGBClassifier          / RobustScaler / Processed Data / PCA( X ) ] f1 score : 0.6717, acc : 0.6829, auc_score : 0.6849
```

→ Model별 Accuracy가 최댓값을 갖는 조합

## 04 | Modeling



## 04 | Modeling

```
1 param_lix = ['lix', {'alpha': [0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 1000]}]
2 param_ada = ['ada', {'n_estimators':list(range(100, 600, 100)),
3                   'algorithm':['SAMME.R'],
4                   'learning_rate':[0.3, 0.5, 0.7]
5                   }]
6 param_ran = ['ran', {'n_estimators':list(range(100, 600, 100)),
7                     'max_depth': [2, 3, 4],
8                     'min_samples_split': [2, 5, 10, 100],
9                     'min_samples_leaf': [1, 4, 8],
10                    }]
11 param_gra = ['gra', {'n_estimators': [100, 500],
12                      'max_depth': [6, 9],
13                      'learning_rate': [0.01, 0.1]
14                     }]
15 param_bag = ['bag', {'n_estimators':list(range(100, 600, 100)),
16                   'max_samples':[100, 200],
17                   'bootstrap': [True, False],
18                  }]
19 param_lgb = ['lgb', {'n_estimators':list(range(100, 600, 100)),
20                 'max_depth':[1, 5, 10],
21                 'learning_rate': [0.1, 0.4, 0.7],
22                }]
23 param_xgb = ['xgb', {'n_estimators':list(range(100, 600, 100)),
24                   'max_depth':[1, 5, 10],
25                   'learning_rate': [0.1, 0.4, 0.7],
26                  }]
27 param_svc = ['svc', {'C':[0.1, 0.5, 1, 5, 10]}]
```

```
1 print_max_cv_list(max_cv_list)
```

```
*****
*****          After Grid Search      *****
*****
[ Ridge           / RobustScaler   / Processed Data / PCA( X ) ] f1 score : 0.6254, acc : 0.6741, auc_score : 0.6679
[ Lasso            / NoScaler     / Raw Data       / PCA( X ) ] f1 score : 0.5963, acc : 0.6564, auc_score : 0.6480
[ ElasticNet        / NoScaler     / Processed Data / PCA(99%) ] f1 score : 0.6014, acc : 0.6561, auc_score : 0.6491
[ RandomForestClassifier / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6657, acc : 0.6870, auc_score : 0.6885
[ AdaBoostClassifier / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6867, acc : 0.6953, auc_score : 0.6979
[ GradientBoostingClassifier / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6954, acc : 0.6974, auc_score : 0.7018
[ BaggingClassifier / NoScaler     / Raw Data       / PCA( X ) ] f1 score : 0.6883, acc : 0.6778, auc_score : 0.6879
[ LGBMClassifier    / MinMaxScaler / Processed Data / PCA( X ) ] f1 score : 0.6928, acc : 0.6951, auc_score : 0.6993
[ XGBClassifier     / RobustScaler / Processed Data / PCA( X ) ] f1 score : 0.6934, acc : 0.6969, auc_score : 0.7009
[ LinearSVC         / NoScaler     / Processed Data / PCA( X ) ] f1 score : 0.6839, acc : 0.6868, auc_score : 0.6909
```

# 05. Result

## 05 | Results

### 모델 성능 비교

	Linear	Lasso	Ridge	EN	Logic	KNN	NB	LDA	SVM	DT	Bag	RF	Ada	GB	XGB	LGBM
Data	Refined	Raw	Refined	Refined	Refined	Refined	Refined	Refined	Refined	Refined	Raw	Refined	Refined	Refined	Refined	Refined
Reduction	X	X	X	PCA 0.99	X	PCA 0.99	PCA 0.99	X	X	X	X	X	X	X	X	X
Scaler	Standard	X	Robust	X	X	MinMax	Robust	MinMax	X	MinMax	X	MinMax	MinMax	MinMax	Robust	MinMax
Accuracy	0.6760	0.6564	0.6741	0.6561	0.6779	0.6552	0.6654	0.6744	0.6868	0.6159	0.6778	0.6870	0.6953	0.6974	0.6969	0.6951
F1 score	0.6279	0.5963	0.6254	0.6014	0.6317	0.6239	0.6228	0.6260	0.6839	0.5740	0.6883	0.6657	0.6867	0.6954	0.6934	0.6928
AUC	0.6699	0.6480	0.6679	0.6491	0.6721	0.6528	0.6606	0.6683	0.6909	0.6120	0.6879	0.6865	0.6979	0.7018	0.7009	0.6993

## 05 | Results

### 모델 성능 비교

	Linear	Lasso	Ridge	EN	Logic	KNN	NB	LDA	SVM	DT	Bag	RF	Ada	GB	XGB	LGBM
Data	Refined	Raw	Refined	Refined	Refined	Refined	Refined	Refined	Refined	Refined	Raw	Refined	Refined	Refined	Refined	Refined
Reduction	X	X	X	PCA 0.99	X	PCA 0.99	PCA 0.99	X	X	X	X	X	X	X	X	X
Scaler	Standard	X	Robust	X	X	MinMax	Robust	MinMax	X	MinMax	X	MinMax	MinMax	MinMax	Robust	MinMax
Accuracy	0.6760	0.6564	0.6741	0.6561	0.6779	0.6552	0.6654	0.6744	0.6868	0.6159	0.6778	0.6870	0.6953	0.6974	0.6969	0.6951
F1 score	0.6279	0.5963	0.6254	0.6014	0.6317	0.6239	0.6228	0.6260	0.6839	0.5740	0.6883	0.6657	0.6867	0.6954	0.6934	0.6928
AUC	0.6699	0.6480	0.6679	0.6491	0.6721	0.6528	0.6606	0.6683	0.6909	0.6120	0.6879	0.6865	0.6979	0.7018	0.7009	0.6993

# (참조) 각 모델별 예측 성능 도표

# Linear Regression

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6727	0.5495	0.5494	0.5494	0.6724	0.6488	0.6506	0.6498	0.6734	0.6381	0.6410	0.6396	0.6727	0.5495	0.5494	0.5494	
f-1 score	0.6248	0.0000	0.0004	0.0004	0.6245	0.5712	0.5728	0.5676	0.6257	0.5562	0.5419	0.5424	0.6248	0.0000	0.0004	0.0004	
AUC	0.666	0.5000	0.4999	0.4999	0.6663	0.6371	0.6388	0.6372	0.6673	0.6260	0.6257	0.6247	0.6666	0.5000	0.4999	0.4999	

# Linear Regression

## 2. Refined Data

Scaler	기본					Standard			MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99												
accuracy	0.6745	0.6441	0.6436	0.6598	0.6760	0.6521	0.6535	0.6555	0.6748	0.6349	0.6346	0.6356	0.6744	0.6602	0.6597	0.6634
f-1 score	0.6259	0.5557	0.5539	0.6098	0.6279	0.5651	0.5680	0.5731	0.6252	0.5182	0.5173	0.5202	0.6241	0.6075	0.6070	0.6078
AUC	0.6683	0.6319	0.6312	0.6536	0.6699	0.6399	0.6414	0.6439	0.6684	0.6190	0.6186	0.6198	0.6679	0.6535	0.6530	0.6562

# Lasso

## 1. Raw Data

Scaler	기본				Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99
accuracy	0.6147	0.5495	0.5494	0.5494	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495
f-1 score	0.4197	0.0000	0.0004	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AUC	0.5872	0.5000	0.4999	0.4999	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

# Lasso

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6144	0.6144	0.6144	0.6144	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	
f-1 score	0.4204	0.4204	0.4204	0.4204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
AUC	0.5905	0.5905	0.5905	0.5905	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	

# Ridge

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6726	0.5495	0.5494	0.5494	0.6727	0.6489	0.6503	0.6494	0.6726	0.6381	0.6410	0.6393	0.6727	0.5495	0.5494	0.5494	
f-1 score	0.6248	0.0000	0.0004	0.0004	0.6248	0.5713	0.5723	0.5670	0.6248	0.5562	0.5419	0.5420	0.6249	0.0000	0.0004	0.0004	
AUC	0.6665	0.5000	0.4999	0.4999	0.6666	0.6372	0.6385	0.6368	0.6665	0.6260	0.6258	0.6244	0.6666	0.5000	0.4999	0.4999	

# Ridge

## 2. Refined Data

Scaler	기본				Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99
accuracy	0.6745	0.6441	0.6436	0.6598	0.6745	0.6490	0.6537	0.6562	0.6744	0.6349	0.6346	0.6356	0.6745	0.6601	0.6598	0.6634
f-1 score	0.6259	0.5557	0.5539	0.6098	0.6259	0.5598	0.5698	0.5730	0.6258	0.5182	0.5173	0.5202	0.6259	0.6073	0.6071	0.6078
AUC	0.6683	0.6319	0.6312	0.6536	0.6683	0.6365	0.6419	0.6444	0.6682	0.6190	0.6186	0.6198	0.6683	0.6534	0.6531	0.6562

# ElasticNet

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6410	0.5495	0.5494	0.5494	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	
f-1 score	0.5664	0.0000	0.0004	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
AUC	0.6301	0.5000	0.5494	0.4999	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	

# ElasticNet

## 2. Refined Data

Scaler	기본				Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99
accuracy	0.6482	0.6482	0.6482	0.6482	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419	0.5419
f-1 score	0.5800	0.5800	0.5800	0.5800	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AUC	0.6391	0.6391	0.6391	0.6391	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

# Logistic Regression

## 2. Refined Data

Scaler	기본				Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99
accuracy	0.6779	0.6454	0.6447	0.6601	0.6771	0.6504	0.6527	0.6562	0.6302	0.6347	0.6345	0.5193	0.6771	0.6606	0.6596	0.6642
f-1 score	0.6317	0.5609	0.5588	0.6106	0.6303	0.5634	0.5682	0.5763	0.6771	0.5176	0.5166	0.6354	0.6303	0.6082	0.6070	0.6106
AUC	0.6721	0.6338	0.6329	0.6540	0.6712	0.6382	0.6408	0.6449	0.6712	0.6187	0.6185	0.6195	0.6712	0.6540	0.6529	0.6573

# KNN

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.5162	0.5120	0.5144	0.5073	0.6166	0.6050	0.6092	0.6221	0.6157	0.6053	0.6045	0.6135	0.6011	0.5065	0.5100	0.5085	
f-1 score	0.4664	0.4364	0.4448	0.4507	0.5596	0.5596	0.5644	0.5815	0.5587	0.5523	0.5505	0.5644	0.5584	0.4335	0.4363	0.4490	
AUC	0.5120	0.5037	0.5069	0.5021	0.6098	0.6007	0.6049	0.6185	0.6089	0.5994	0.5985	0.6083	0.5974	0.4987	0.5020	0.5027	

# KNN

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6372	0.6229	0.6294	0.6353	0.6462	0.6396	0.6445	0.6469	0.6451	0.6468	0.6461	0.6552	0.6466	0.6271	0.6352	0.6424	
f-1 score	0.6106	0.5886	0.5940	0.6054	0.6034	0.6031	0.6081	0.6102	0.6038	0.6096	0.6110	0.6239	0.6104	0.5897	0.5969	0.6038	
AUC	0.6359	0.6203	0.6265	0.6334	0.6417	0.6364	0.6412	0.6436	0.6409	0.6433	0.6431	0.6528	0.6433	0.6238	0.6317	0.6387	

# Decision Tree

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6028	0.5058	0.6210	0.5111	0.6063	0.5638	0.5651	0.5716	0.6047	0.5572	0.5648	0.5687	0.6061	0.5035	0.5195	0.5101	
f-1 score	0.5576	0.4535	0.4718	0.4559	0.5624	0.5180	0.5264	0.5354	0.5629	0.5106	0.5216	0.5304	0.5613	0.4550	0.4737	0.4617	
AUC	0.5986	0.5012	0.5168	0.5060	0.6022	0.5599	0.5625	0.5695	0.6011	0.5532	0.5614	0.5662	0.6019	0.4996	0.5160	0.5061	

# Decision Tree

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6139	0.5969	0.5930	0.6042	0.6132	0.5918	0.5935	0.6029	0.6159	0.6019	0.6071	0.6032	0.6149	0.5900	0.6000	0.6031	
f-1 score	0.5718	0.5559	0.5571	0.5655	0.5706	0.5500	0.5481	0.5598	0.5740	0.5644	0.5667	0.5600	0.5735	0.5479	0.5622	0.5614	
AUC	0.6100	0.5934	0.5903	0.6010	0.6092	0.5882	0.5892	0.5989	0.6120	0.5989	0.6035	0.5992	0.6110	0.5863	0.5970	0.5993	

# SVM

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.5572	0.4540	0.4600	0.5402	0.6715	0.6482	0.6500	0.6492	0.6730	0.6381	0.6410	0.6395	0.6713	0.5466	0.5471	0.5452	
f-1 score	0.6325	0.6127	0.5914	0.1783	0.6247	0.5702	0.5722	0.5670	0.6254	0.5561	0.5419	0.5422	0.6228	0.0220	0.0319	0.0365	
AUC	0.5832	0.4995	0.4967	0.5015	0.6657	0.6364	0.6382	0.6366	0.6670	0.6259	0.6258	0.6245	0.6651	0.4984	0.4993	0.4978	

# SVM

## 2. Refined Data

Scaler	기본				Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99
accuracy	0.6826	0.6480	0.6460	0.6605	0.6746	0.6512	0.6508	0.6550	0.6743	0.6348	0.6346	0.6356	0.6743	0.6599	0.6598	0.6633
f-1 score	0.7173	0.5710	0.5644	0.6151	0.6262	0.5634	0.5622	0.5705	0.6259	0.5180	0.5173	0.5202	0.6259	0.6072	0.6072	0.6076
AUC	0.6978	0.6376	0.6347	0.6553	0.6685	0.6389	0.6383	0.6430	0.6682	0.6189	0.6186	0.6198	0.6682	0.6533	0.6532	0.6560

# Bagging

## 1. Raw Data

Scaler	기본				Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99	기본	PCA 0.90	PCA 0.95	PCA 0.99
accuracy	0.6547	0.5225	0.5215	0.5293	0.6425	0.6082	0.6109	0.6303	0.6446	0.6048	0.6100	0.6202	0.6434	0.5230	0.5293	0.5249
f-1 score	0.6021	0.3799	0.3725	0.3764	0.5863	0.5285	0.5284	0.5507	0.5871	0.5136	0.5217	0.5332	0.5873	0.3738	0.3833	0.3602
AUC	0.6479	0.5047	0.5029	0.5100	0.6353	0.5973	0.5995	0.6164	0.6307	0.5920	0.5976	0.6077	0.6362	0.5043	0.5109	0.5044

# Bagging

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6539	0.6220	0.6216	0.6512	0.6520	0.6332	0.6303	0.6418	0.6531	0.6349	0.6396	0.6432	0.6428	0.6286	0.6320	0.6391	
f-1 score	0.6030	0.5588	0.5602	0.5932	0.5966	0.5635	0.5678	0.5763	0.5993	0.5717	0.5774	0.5846	0.5867	0.5610	0.5701	0.5738	
AUC	0.6477	0.6143	0.6142	0.6438	0.6450	0.6242	0.6226	0.6333	0.6464	0.6269	0.6317	0.6359	0.6359	0.6200	0.6243	0.6307	

# Random Forest

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6799	0.5253	0.5280	0.5402	0.6818	0.6289	0.6389	0.6605	0.6793	0.6215	0.6292	0.6417	0.6822	0.5257	0.5332	0.5337	
f-1 score	0.6643	0.4084	0.3960	0.3791	0.6675	0.5807	0.5896	0.6195	0.6624	0.5636	0.5707	0.5855	0.6684	0.4045	0.4006	0.3658	
AUC	0.6820	0.5107	0.5114	0.5196	0.6843	0.6236	0.6332	0.6562	0.6811	0.6144	0.6218	0.6345	0.6848	0.5106	0.5163	0.5125	

# Random Forest

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6894	0.6423	0.6451	0.6742	0.6861	0.6558	0.6670	0.6766	0.6924	0.6539	0.6592	0.6666	0.6916	0.6541	0.6587	0.6693	
f-1 score	0.6753	0.6192	0.6250	0.6583	0.6708	0.6233	0.6384	0.6466	0.6792	0.6716	0.6231	0.6367	0.6777	0.6238	0.6344	0.6492	
AUC	0.6906	0.6417	0.6451	0.6750	0.6870	0.6532	0.6650	0.6742	0.6938	0.6505	0.6558	0.6644	0.6928	0.6520	0.6578	0.6692	

# Adaboost

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6902	0.5485	0.5444	0.5510	0.6902	0.6520	0.6522	0.6551	0.6918	0.6351	0.6431	0.6452	0.6902	0.5458	0.5457	0.5504	
f-1 score	0.6803	0.1852	0.1945	0.2474	0.6803	0.6003	0.6007	0.5960	0.6822	0.5657	0.5637	0.5683	0.6803	0.1899	0.1886	0.2796	
AUC	0.6939	0.5094	0.5064	0.5161	0.6939	0.6455	0.6458	0.6470	0.6956	0.6254	0.6312	0.6338	0.699	0.5073	0.5071	0.5182	

# Adaboost

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6929	0.6715	0.6710	0.6735	0.6929	0.6700	0.6694	0.6676	0.6929	0.6481	0.6555	0.6675	0.6929	0.6699	0.6692	0.6706	
f-1 score	0.6841	0.6968	0.6946	0.6644	0.6841	0.6420	0.6378	0.6345	0.6841	0.5862	0.5997	0.6248	0.6841	0.6491	0.6463	0.6422	
AUC	0.6954	0.6833	0.6823	0.6759	0.6954	0.6682	0.6668	0.6647	0.6954	0.6400	0.6483	0.6626	0.6954	0.6696	0.6684	0.6687	

# Gradient Boost

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6924	0.5451	0.5443	0.5513	0.6923	0.6608	0.6592	0.6667	0.6909	0.6383	0.6421	0.6468	0.6924	0.5463	0.5442	0.5513	
f-1 score	0.6868	0.1786	0.1804	0.2244	0.6866	0.6123	0.6104	0.6166	0.6841	0.5655	0.5608	0.5691	0.6868	0.1844	0.1783	0.2270	
AUC	0.6975	0.5059	0.5053	0.5147	0.6974	0.6548	0.6532	0.6602	0.6956	0.6279	0.6299	0.6352	0.6975	0.5073	0.5051	0.5148	

# Gradient Boost

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6963	0.6736	0.6735	0.6860	0.6963	0.6782	0.6804	0.6795	0.6963	0.6667	0.6714	0.6802	0.6963	0.6773	0.6766	0.6826	
f-1 score	0.6908	0.6910	0.6876	0.6903	0.6908	0.6576	0.6617	0.6582	0.6908	0.6367	0.6406	0.6554	0.6908	0.6694	0.6676	0.6691	
AUC	0.6997	0.6831	0.6820	0.6920	0.6997	0.6779	0.6805	0.6791	0.6997	0.6644	0.6689	0.6790	0.6997	0.6801	0.6791	0.6840	

# XGBoost

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6754	0.5271	0.5310	0.5323	0.6754	0.6498	0.6469	0.6562	0.6755	0.6331	0.6358	0.6407	0.6754	0.5325	0.5313	0.5395	
f-1 score	0.6754	0.5271	0.5310	0.5323	0.6611	0.6063	0.6054	0.6206	0.6612	0.5695	0.5713	0.5843	0.6611	0.3579	0.3743	0.4109	
AUC	0.6779	0.5063	0.5109	0.5165	0.6779	0.6452	0.6428	0.6533	0.6779	0.6246	0.6270	0.6335	0.6779	0.5106	0.5115	0.5230	

# XGBoost

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6829	0.6652	0.6600	0.6757	0.6829	0.6751	0.6721	0.6733	0.6829	0.6722	0.6709	0.6800	0.6829	0.6710	0.6620	0.6674	
f-1 score	0.6716	0.6701	0.6580	0.6652	0.6716	0.6575	0.6519	0.6525	0.6716	0.6486	0.6508	0.6638	0.6717	0.6566	0.6470	0.6482	
AUC	0.6848	0.6712	0.6642	0.6779	0.6848	0.6755	0.6719	0.6730	0.6849	0.6722	0.6631	0.6675	0.6848	0.6713	0.6708	0.6807	

# LightGBM

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6916	0.5424	0.5466	0.5461	0.6919	0.6578	0.6590	0.6665	0.6936	0.6400	0.6403	0.6488	0.6899	0.5424	0.5473	0.5454	
f-1 score	0.6876	0.2852	0.3162	0.3344	0.6885	0.6164	0.6163	0.6327	0.6896	0.5783	0.5758	0.5918	0.6864	0.2906	0.3138	0.3342	
AUC	0.6971	0.5118	0.5183	0.5197	0.6977	0.6535	0.6544	0.6639	0.6992	0.6317	0.6315	0.6412	0.6956	0.5123	0.5187	0.5191	

# LightGBM

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6951	0.6698	0.6714	0.6861	0.6951	0.6810	0.6838	0.6804	0.6951	0.6797	0.6792	0.6890	0.6951	0.6787	0.6770	0.6837	
f-1 score	0.6928	0.6824	0.6791	0.6870	0.6928	0.6679	0.6697	0.6670	0.6928	0.6632	0.6635	0.6838	0.6928	0.6711	0.6702	0.6736	
AUC	0.6993	0.6779	0.6782	0.6912	0.6993	0.6825	0.6851	0.6818	0.6993	0.6804	0.6800	0.6925	0.6993	0.6816	0.6800	0.6859	

# Naive Bayes

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.4524	0.4516	0.5492	0.4519	0.5195	0.6511	0.6486	0.6538	0.5195	0.6349	0.6396	0.6393	0.5200	0.4516	0.5488	0.4519	
f-1 score	0.6172	0.6200	0.0089	0.6183	0.6205	0.5912	0.5905	0.5844	0.6205	0.5518	0.5547	0.5549	0.6208	0.6201	0.0062	0.6182	
AUC	0.4999	0.5004	0.5001	0.5000	0.5513	0.6429	0.6408	0.6436	0.5513	0.6226	0.6268	0.6267	0.5517	0.5005	0.4996	0.4999	

# Naive Bayes

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6610	0.6422	0.6410	0.6524	0.6610	0.6485	0.6518	0.6592	0.6610	0.6344	0.6341	0.6308	0.6610	0.6634	0.6641	0.6654	
f-1 score	0.6568	0.5404	0.5394	0.5802	0.6568	0.5486	0.5557	0.5801	0.6568	0.5130	0.5119	0.5094	0.6568	0.6237	0.6258	0.6228	
AUC	0.6646	0.6281	0.6270	0.6425	0.6646	0.6344	0.6381	0.6480	0.6646	0.6178	0.6175	0.6144	0.6646	0.6592	0.6601	0.6606	

# LDA

## 1. Raw Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6731	0.5495	0.5494	0.5494	0.6731	0.6489	0.6504	0.6497	0.6731	0.6382	0.6410	0.6394	0.6731	0.5495	0.5494	0.5494	
f-1 score	0.6257	0.0000	0.0004	0.0004	0.6257	0.5715	0.5727	0.5678	0.6257	0.5565	0.5420	0.5423	0.6257	0.0000	0.0004	0.0004	
AUC	0.6671	0.5000	0.4999	0.4999	0.6671	0.6372	0.6387	0.6372	0.6671	0.6261	0.6257	0.6245	0.6671	0.5000	0.5999	0.4999	

# LDA

## 2. Refined Data

Scaler	기본					Standard				MinMax				Robust			
차원축소	기본	PCA 0.90	PCA 0.95	PCA 0.99													
accuracy	0.6744	0.6445	0.6436	0.6596	0.6744	0.6532	0.6526	0.6541	0.6744	0.6350	0.6347	0.6355	0.6744	0.6603	0.6596	0.6636	
f-1 score	0.6260	0.5564	0.5540	0.6098	0.6260	0.5674	0.5671	0.5678	0.6260	0.5184	0.5175	0.5202	0.6260	0.6078	0.6070	0.6082	
AUC	0.6682	0.6323	0.6312	0.6535	0.6683	0.6411	0.6406	0.6419	0.6683	0.6191	0.6187	0.6197	0.6683	0.6537	0.6530	0.6564	

감사합니다

