다금바리를 꿈꾸는 송사리

조수익 이경환 정순주

서론

데이터 개요

address
Mjob
freetime
Medu
absences
Fedu
health
Fjob
sex
farmel
Walc
famsup
address
study time
internet
failures
school
reason
higher
Dalc
paid
age
romantic
traveltime
guardian
goout
Mjob

데이터 개요

개인적 요인

Dalc

sex

freetime

health

failures

Walc

age

goout

romantic

가정적 요인

Medu

Fedu

Fjob

Mjob

famsup

guardian

internet

사회적 요인

address

absences

study time

reason

paid

traveltime

address

higher

school

목표

카테고리 분류를 통해
개인 / 가정 / 학교
요소중 성적에 미치는 영향 분석

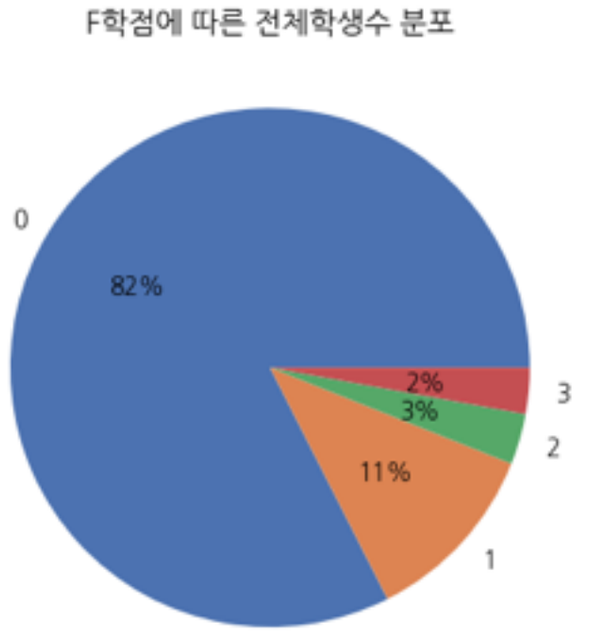
보로

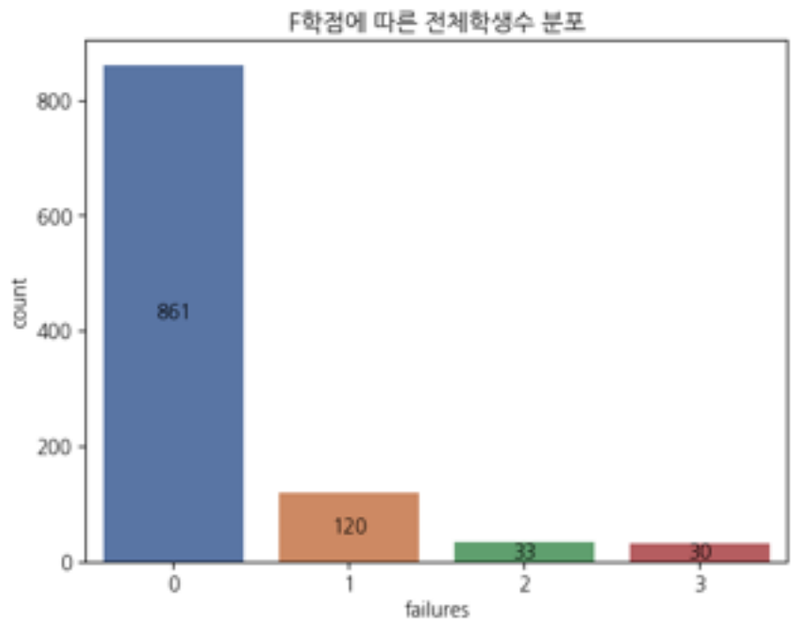
카테고리별 EDA

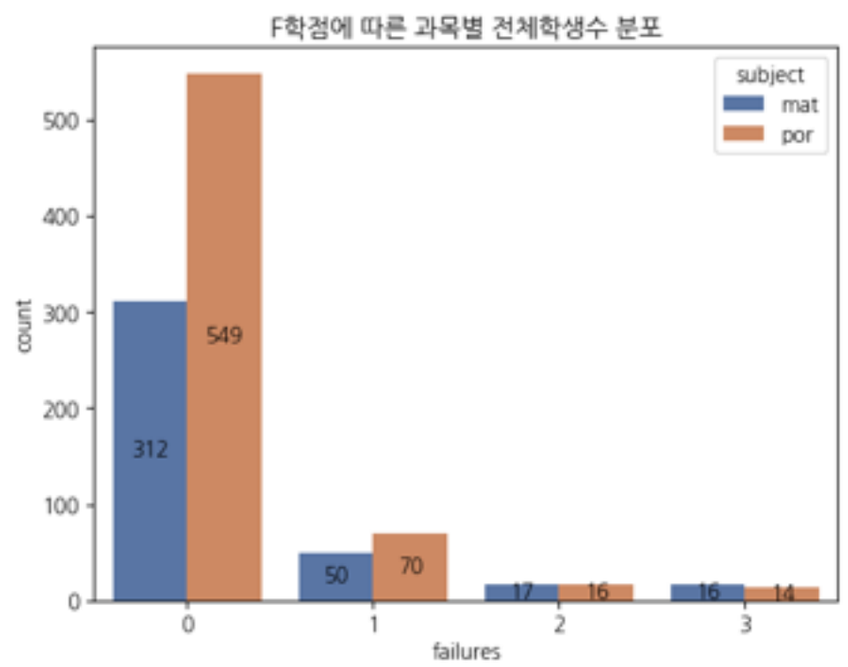
개별 컬럼들을 EDA

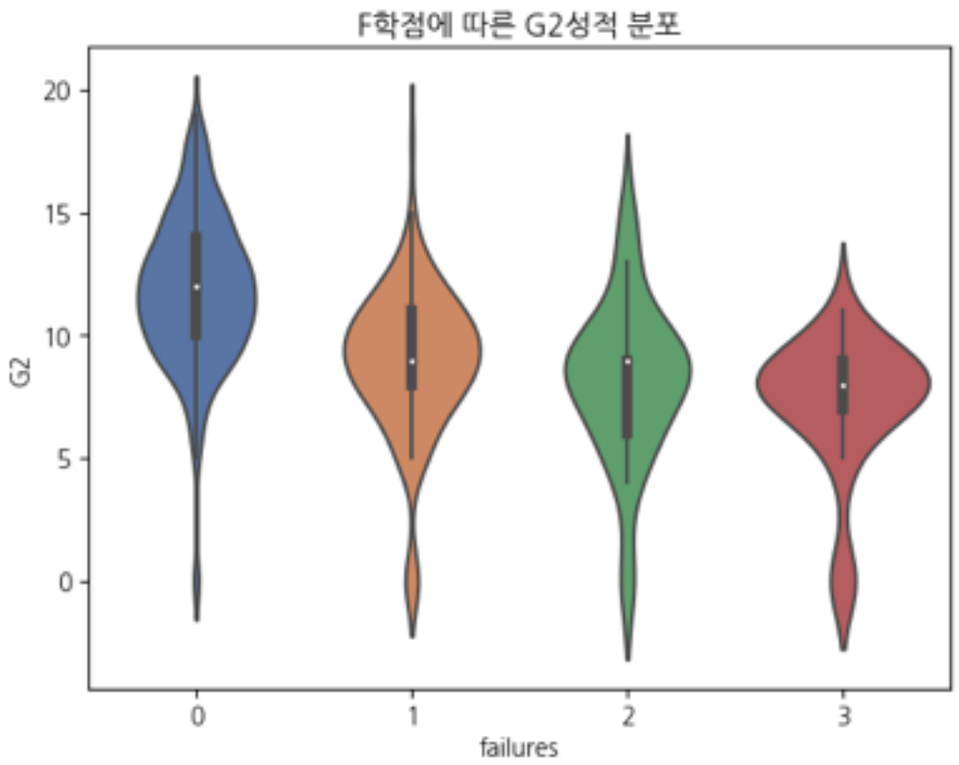
를 통해 성적과의 연관성을 분석

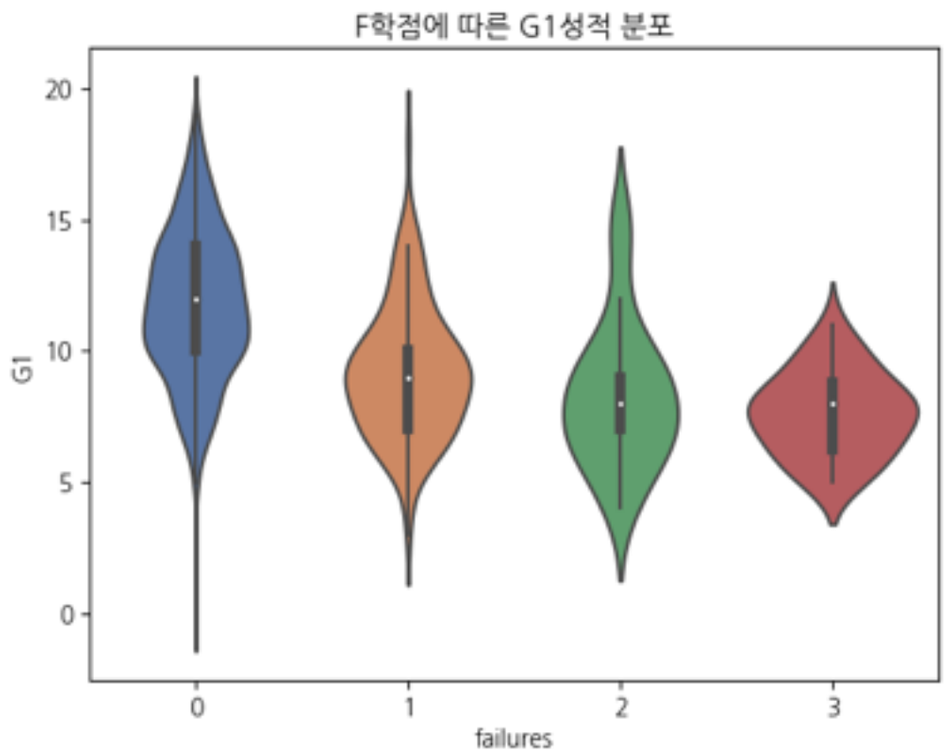
카테고리별 EDA분석(개인적 환경)
failures





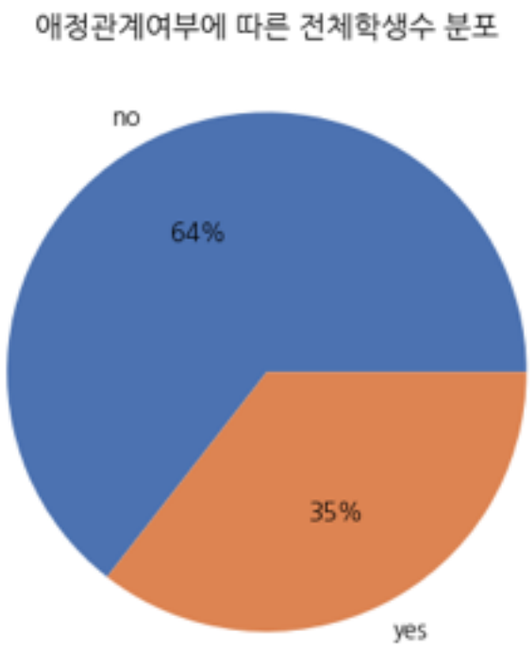


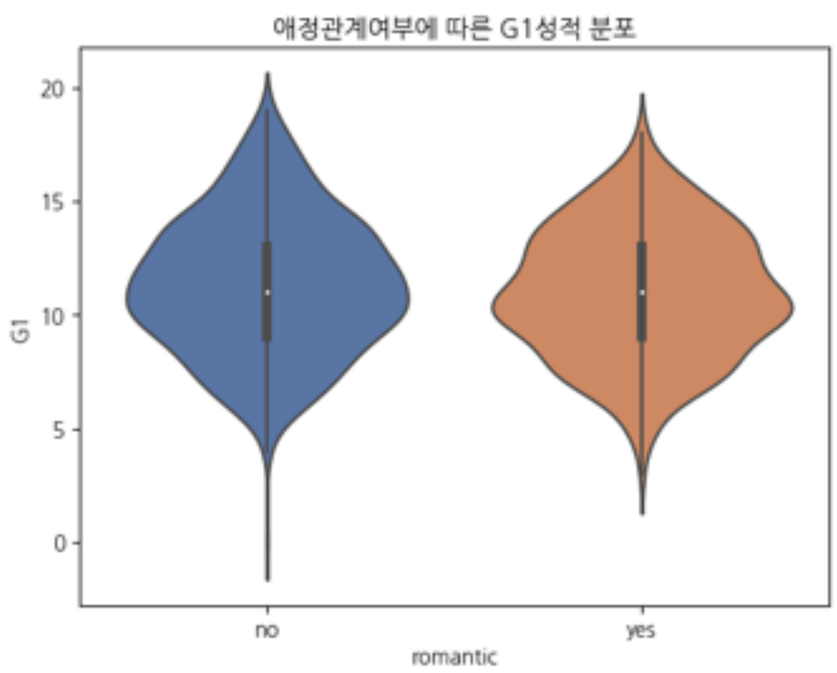


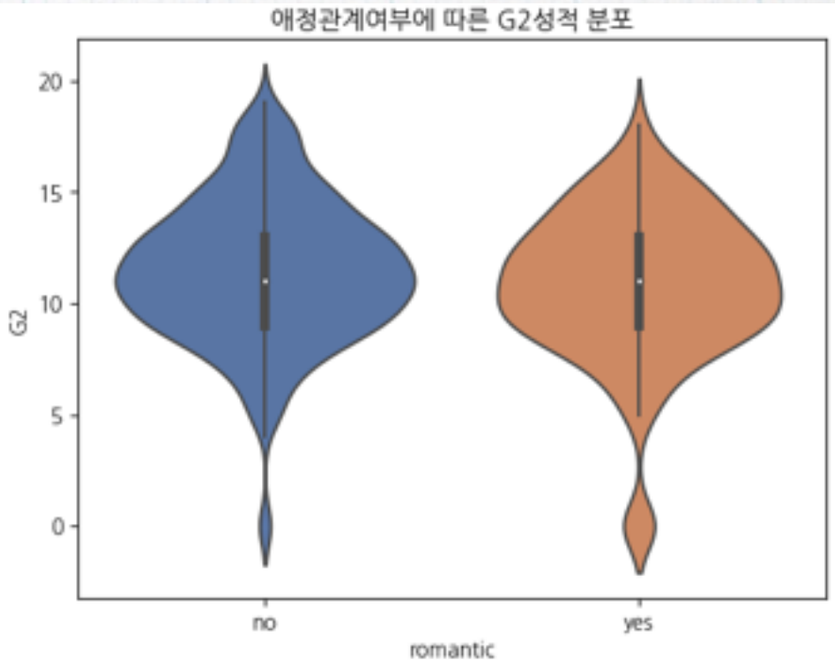


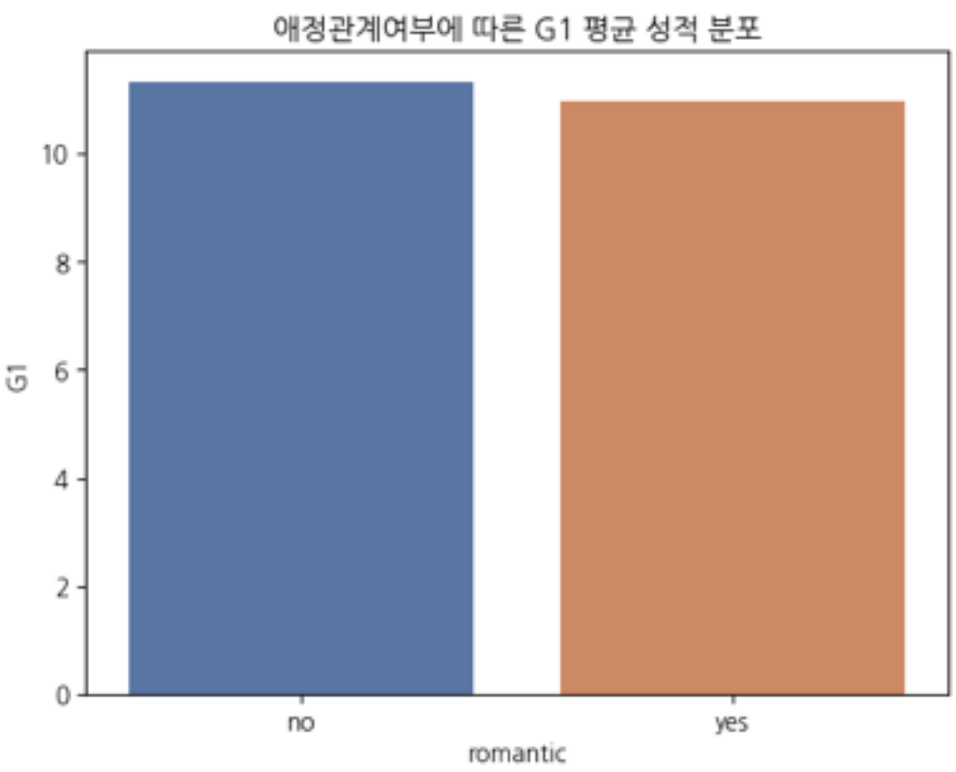
카테고리별 EDA분석(개인적 환경)

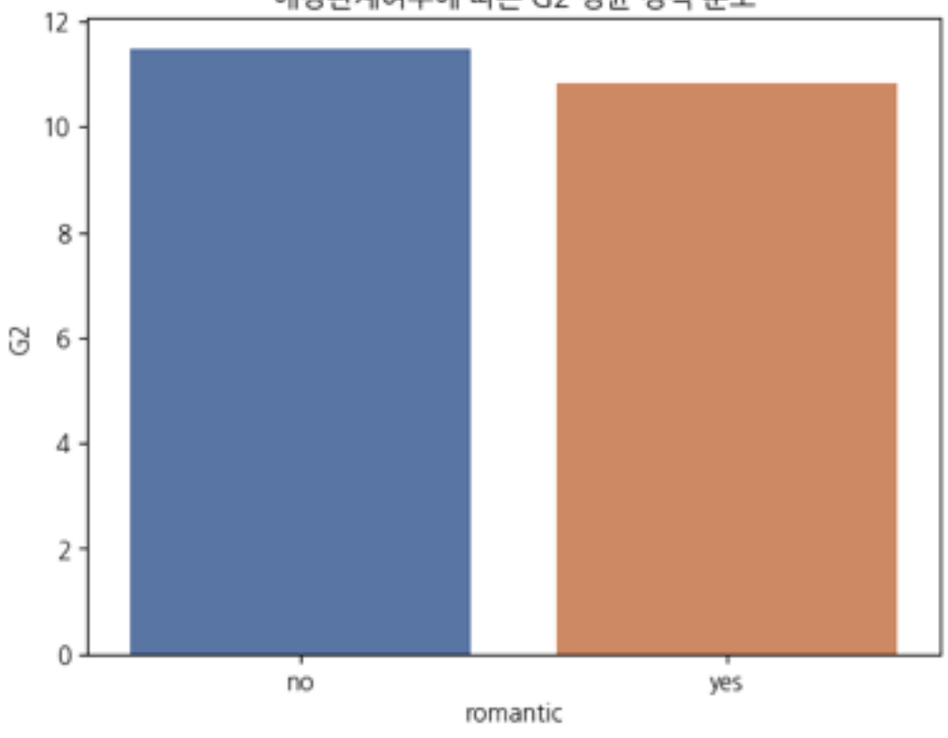
romantic



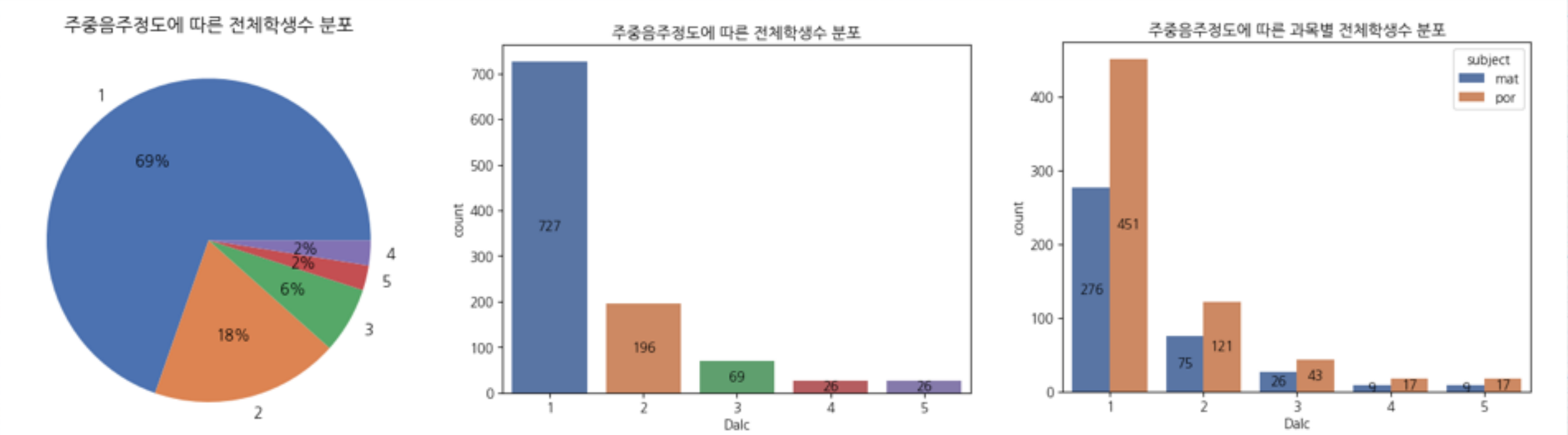


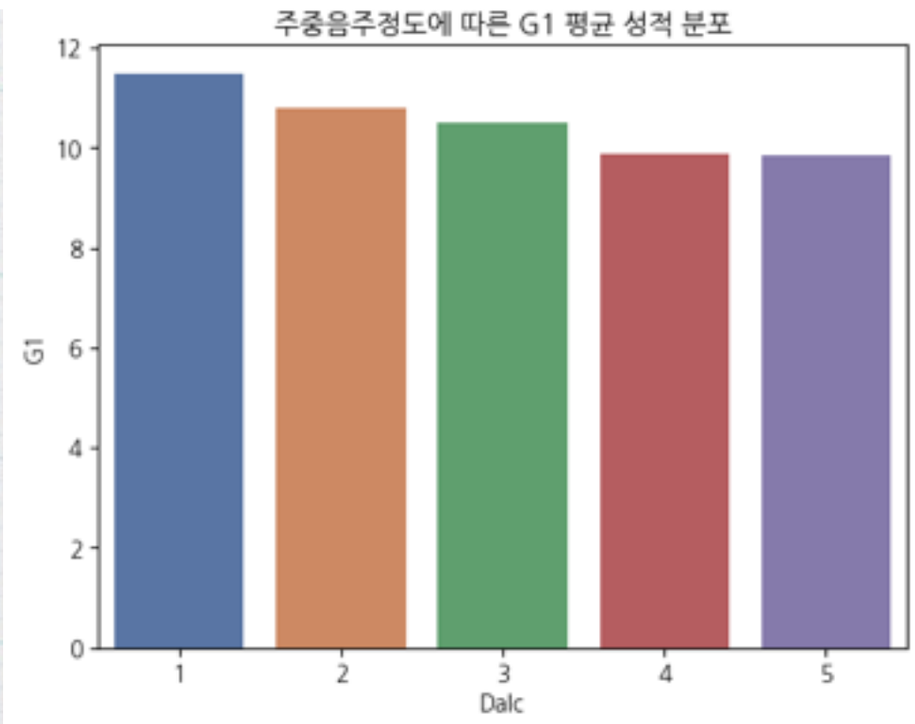


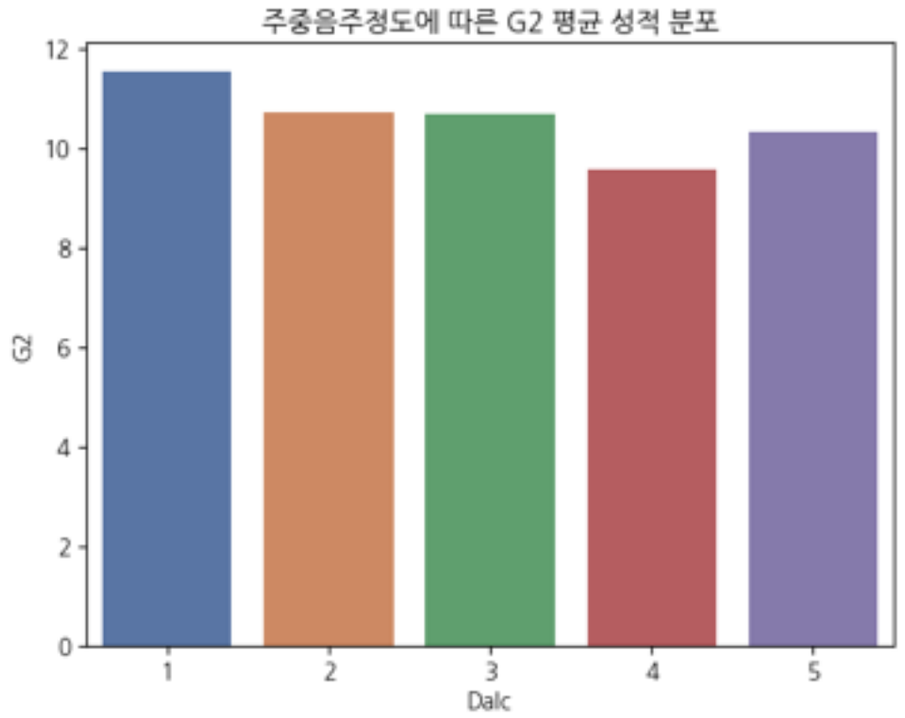




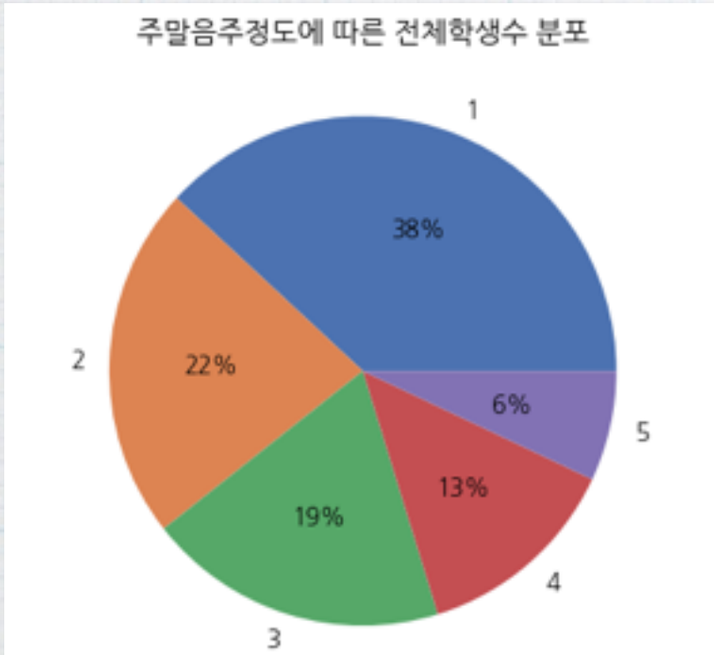
카테고리별 EDA분석(개인적 환경)
Dalc

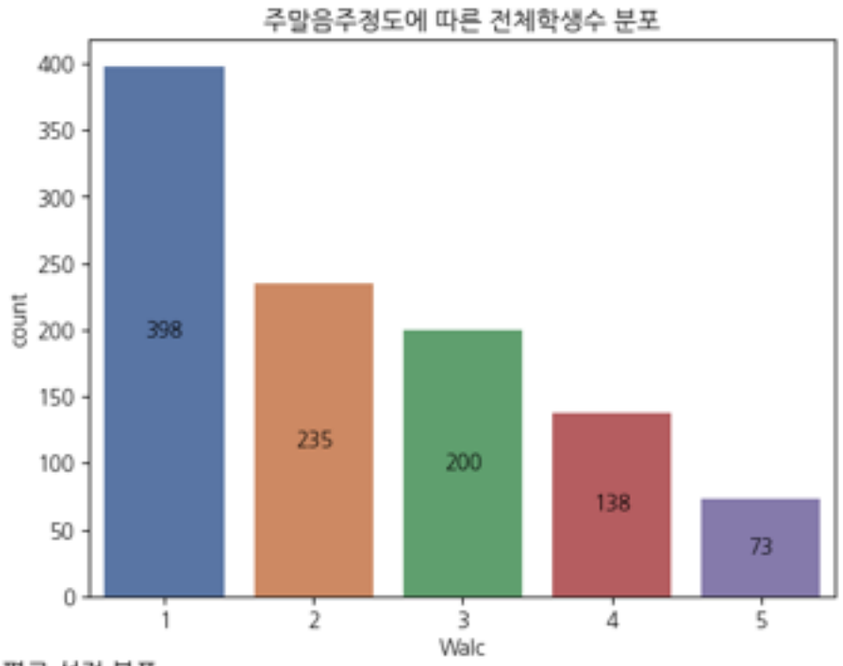


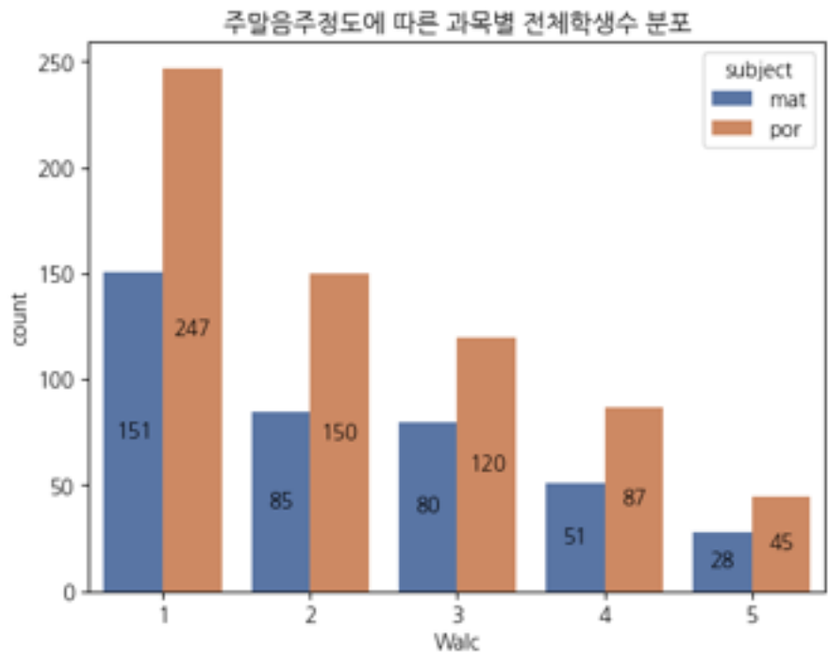


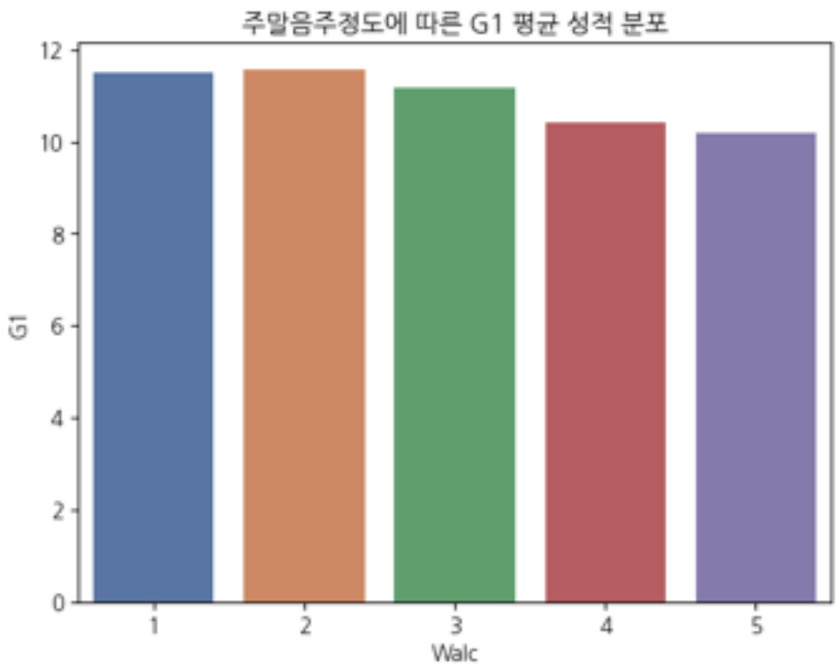


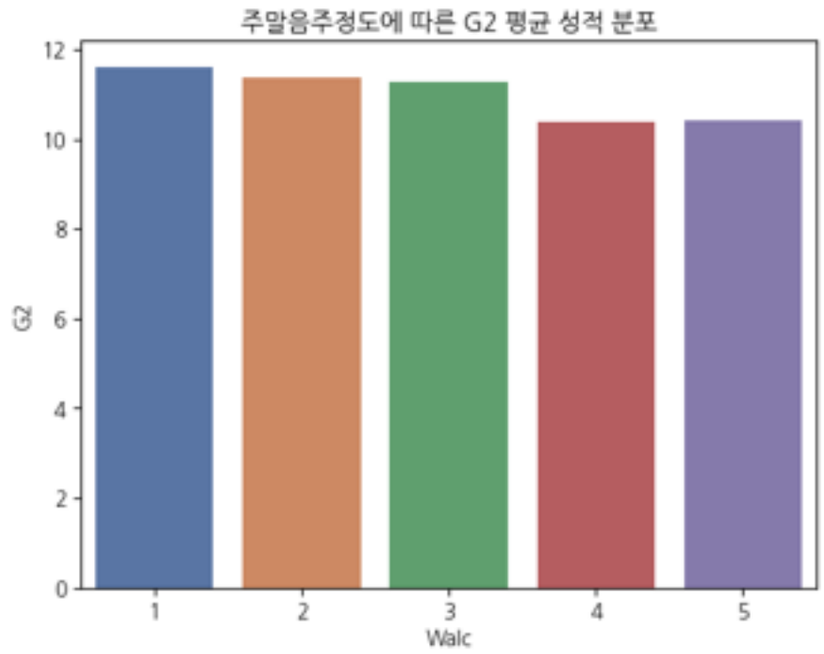
카테고리별 EDA분석(개인적 환경)
Walc





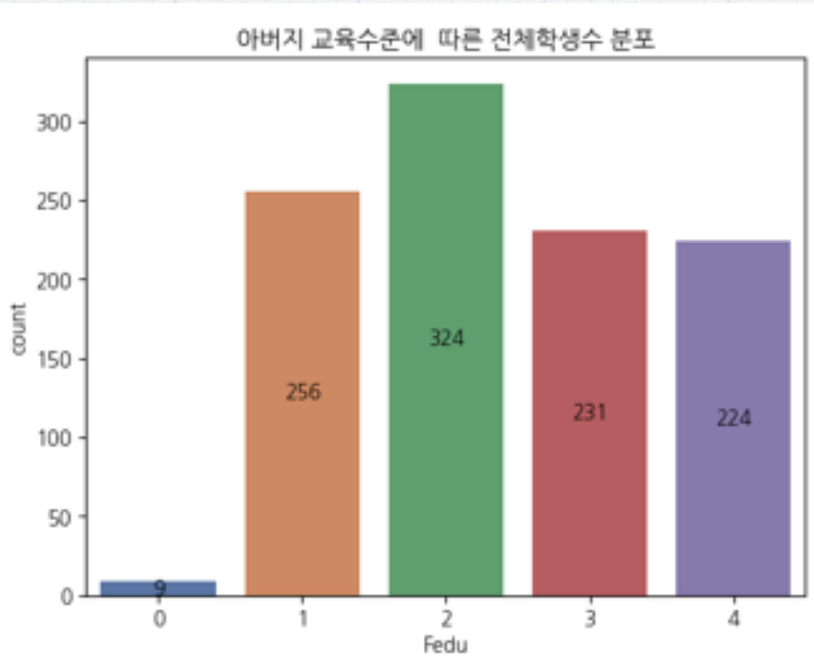


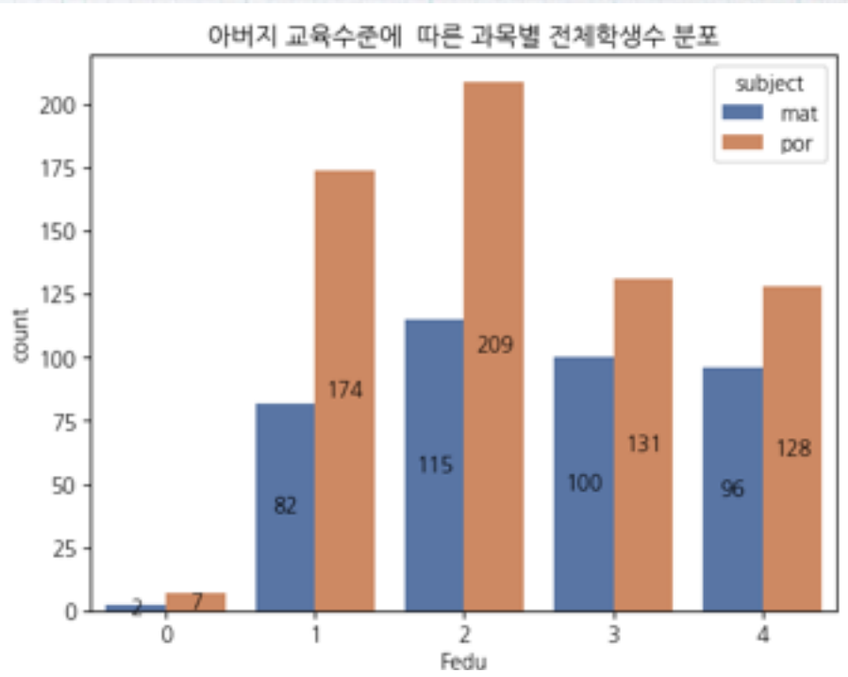


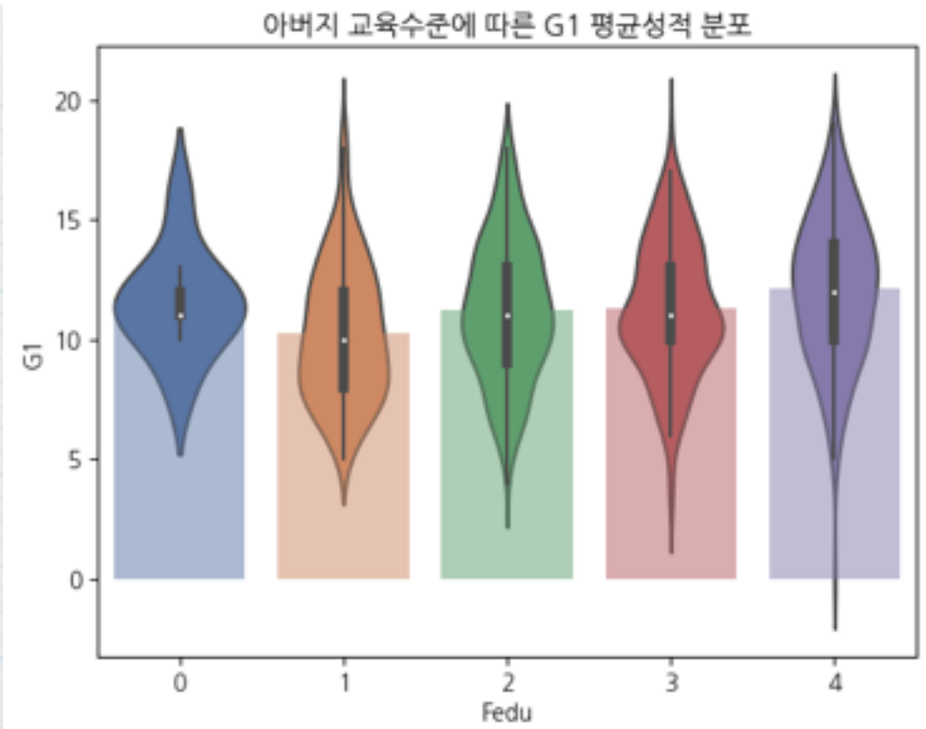


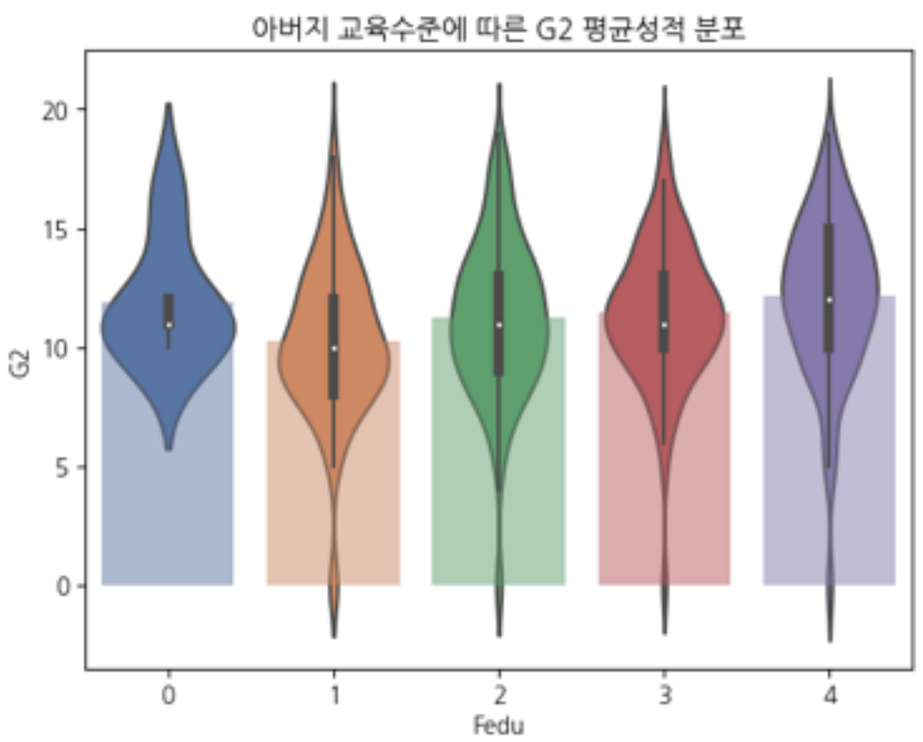
카테고리별 EDA분석(가정적 환경)

Fedu

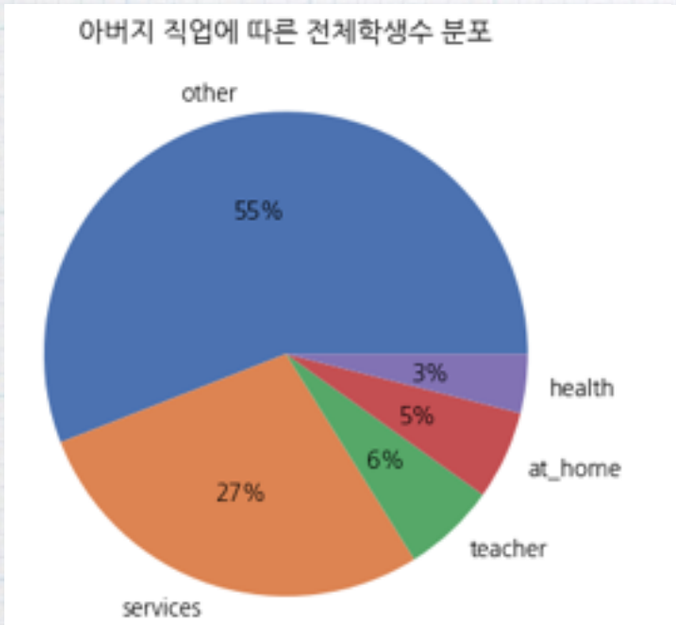


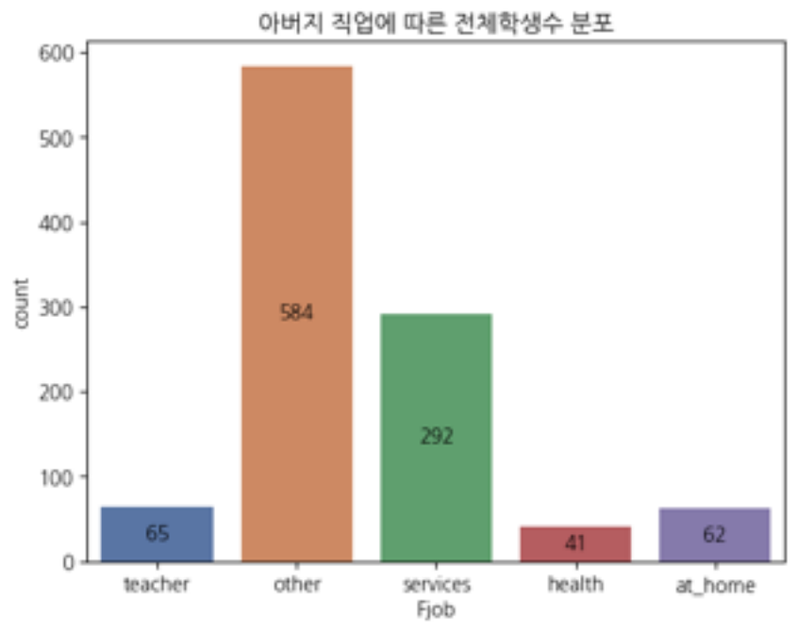


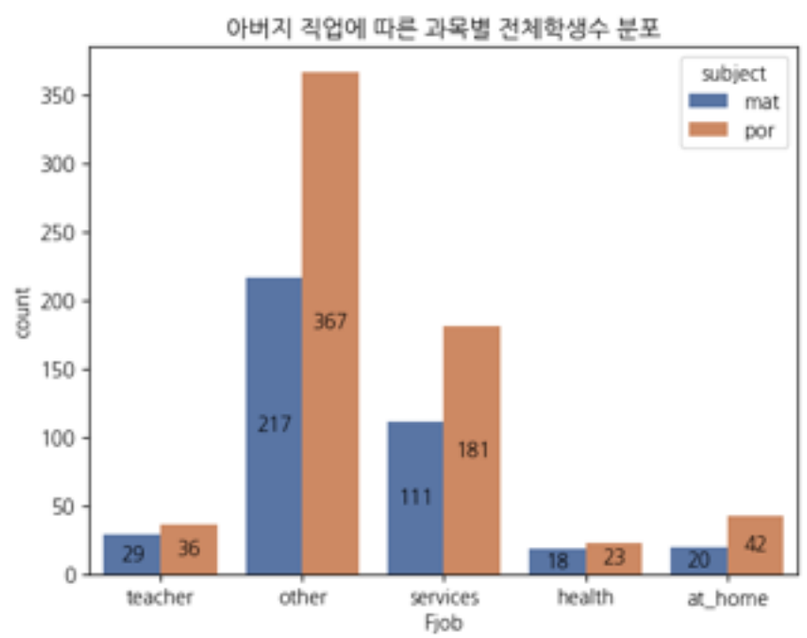


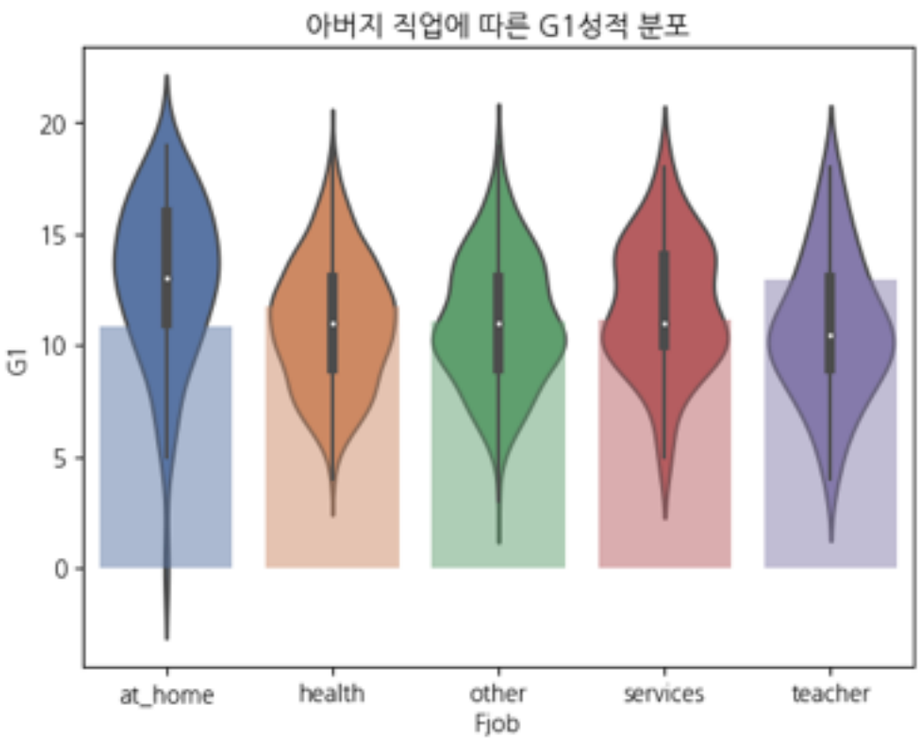


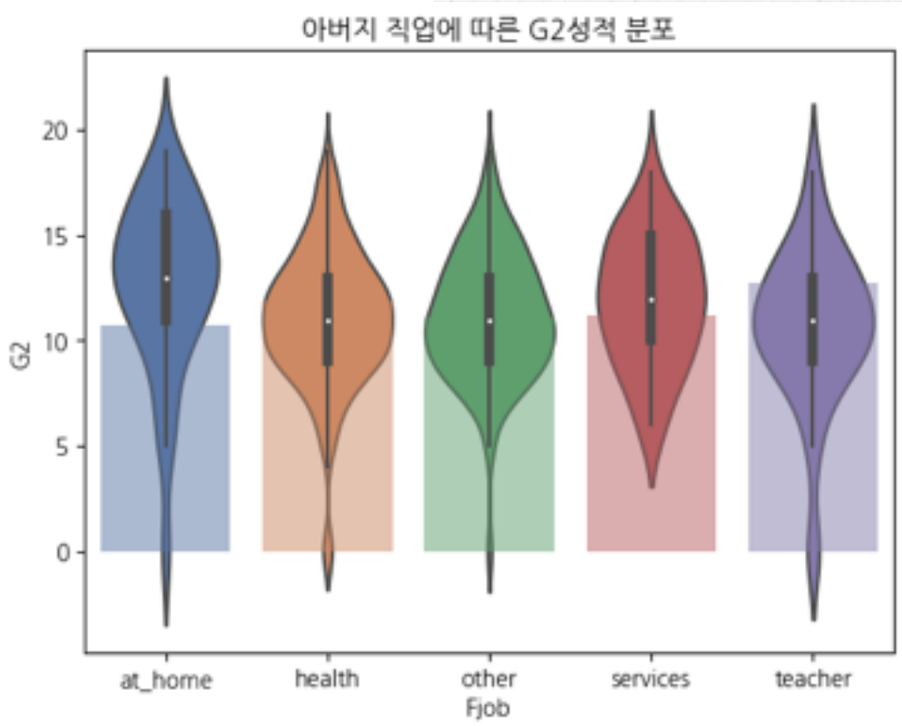
카테고리별 EDA분석(가정적 환경)
Fjob





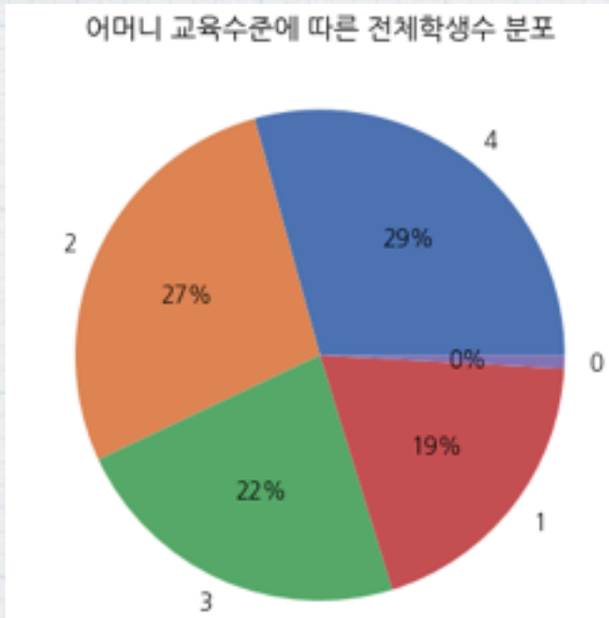


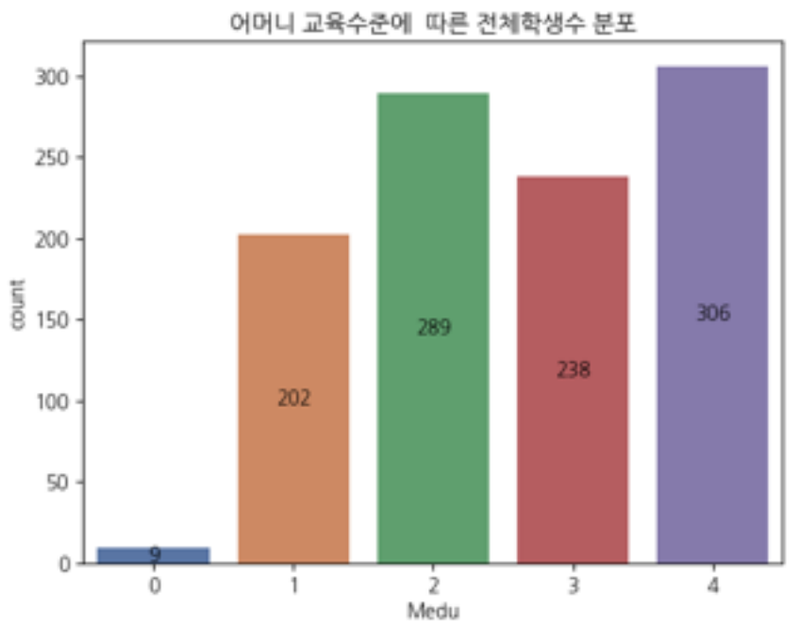


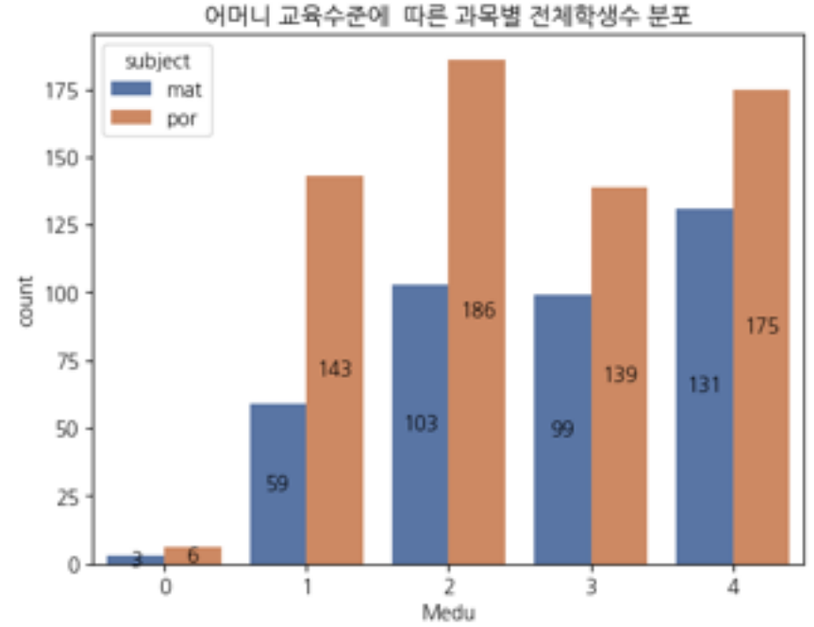


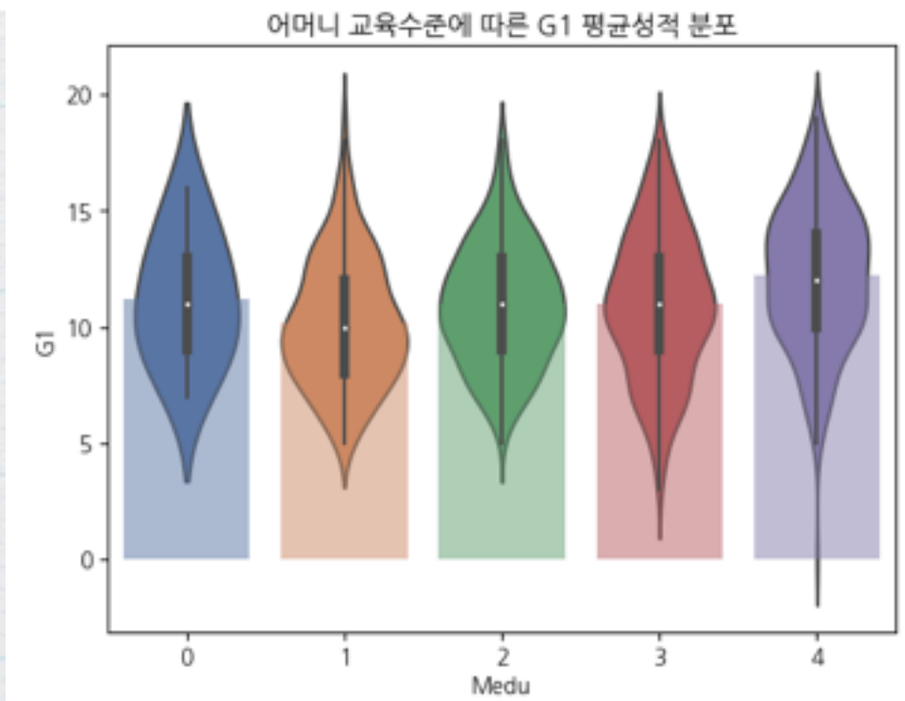
카테고리별 EDA분석(가정적 환경)

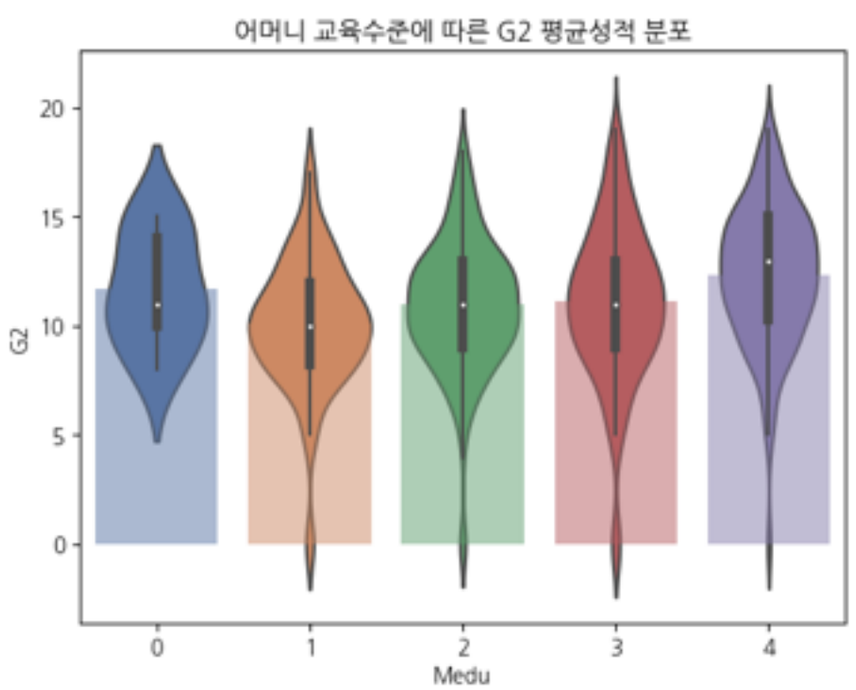
Medu





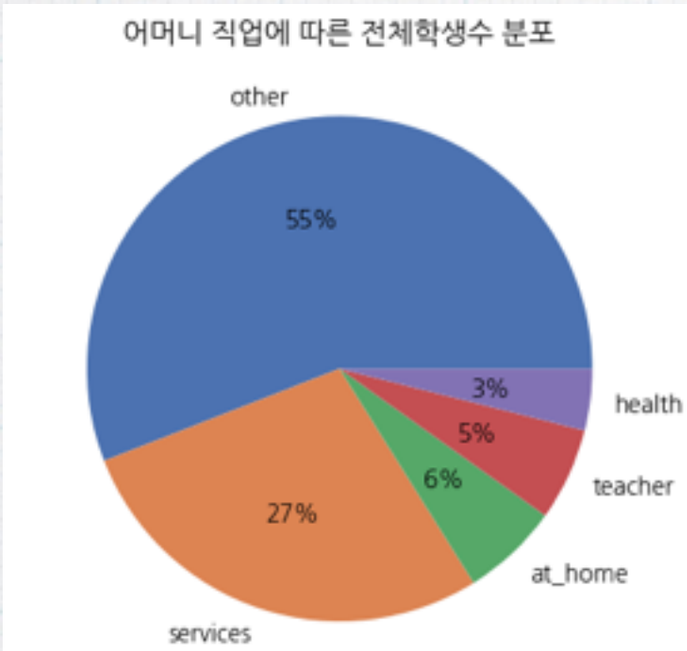


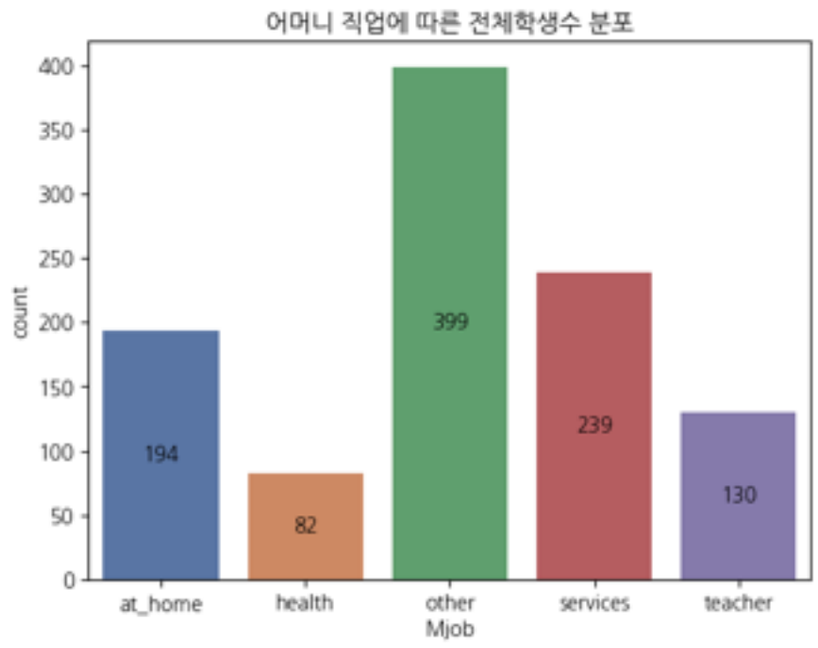


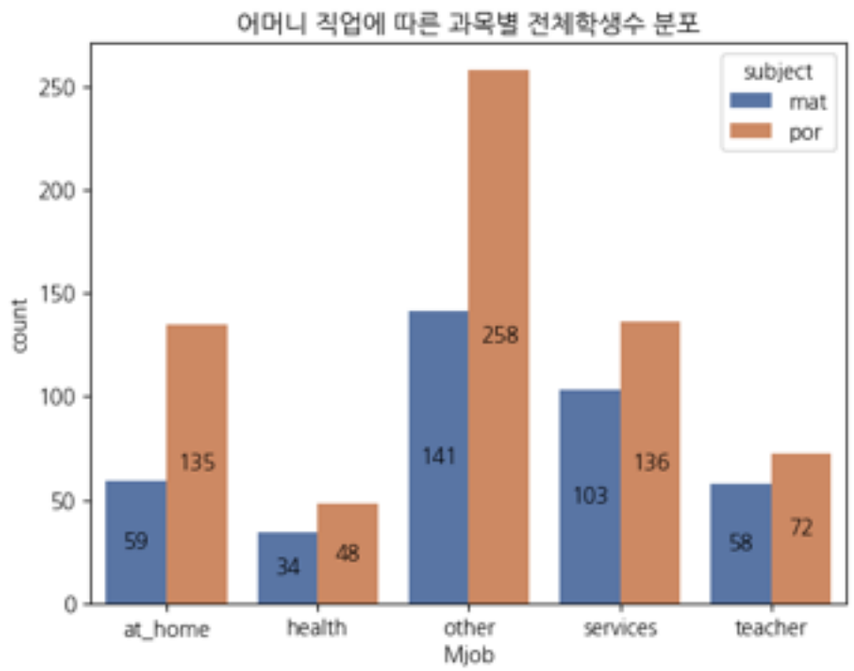


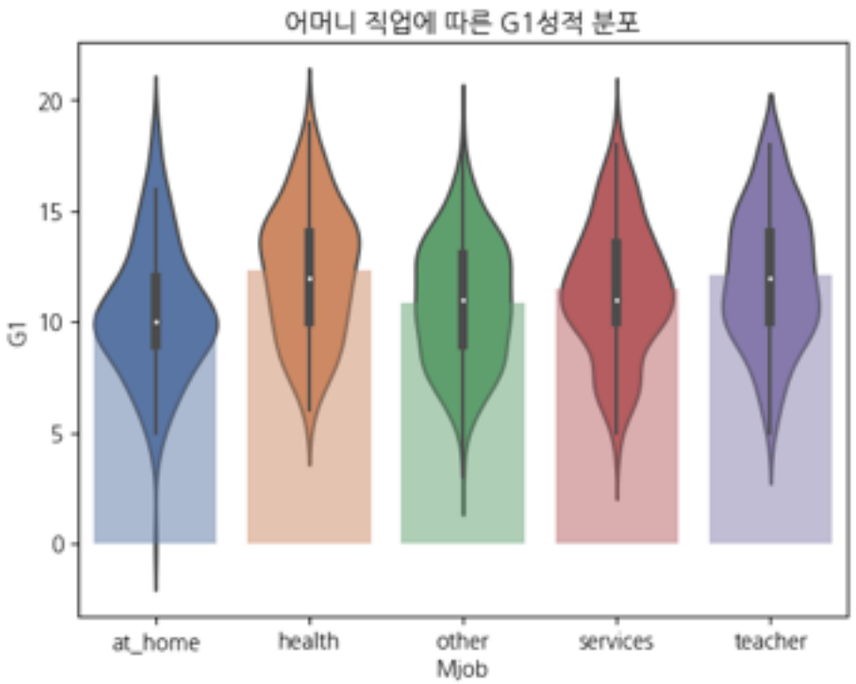
카테고리별 EDA분석(가정적 환경)

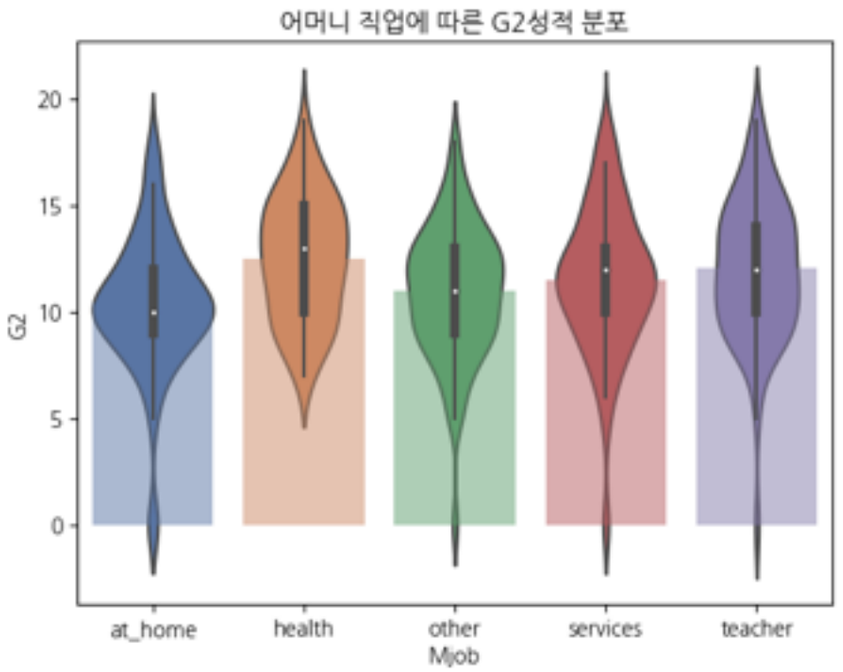
Mjob





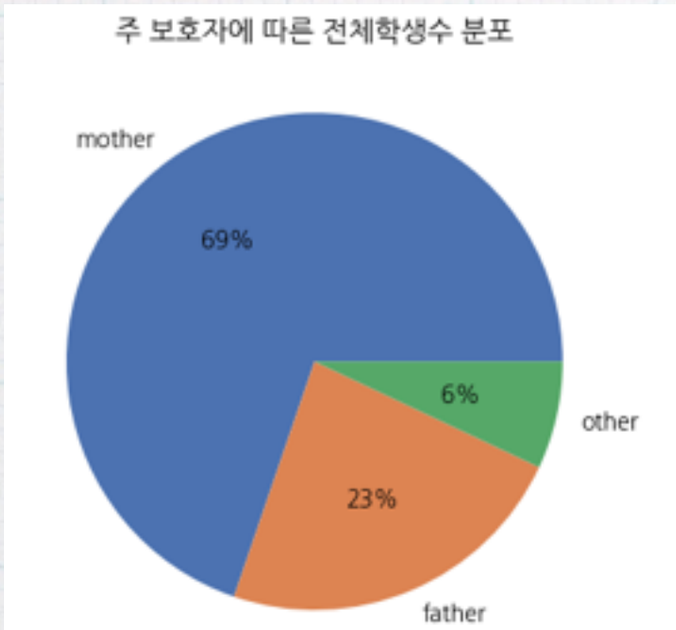


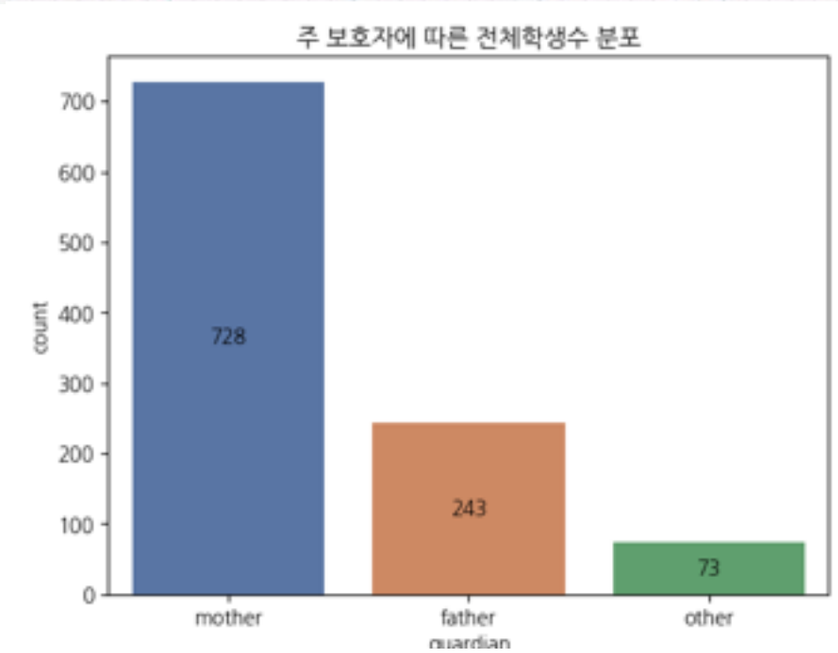


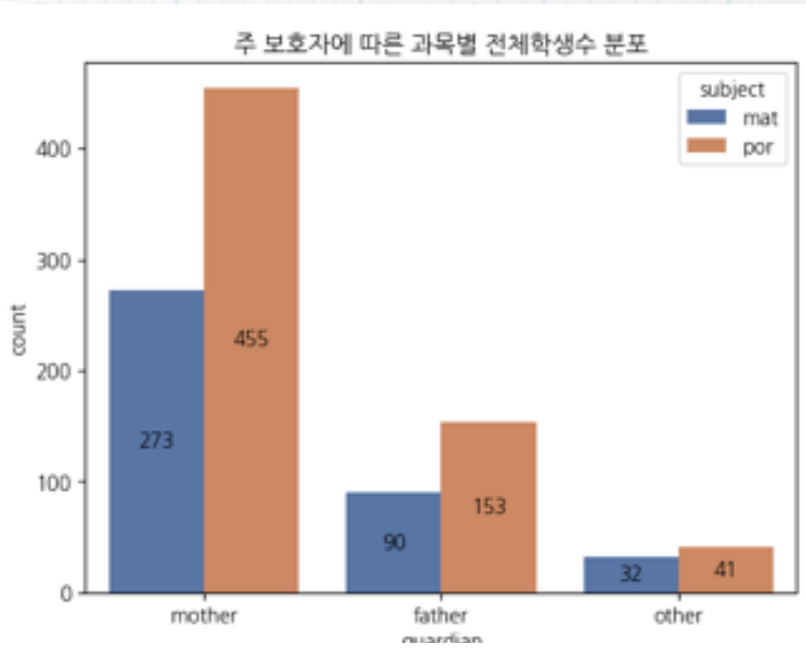


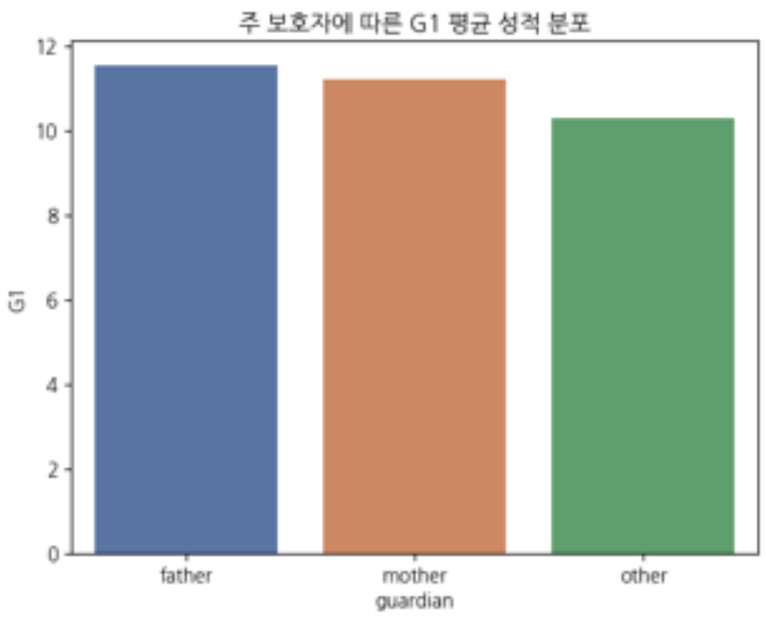
카테고리별 EDA분석(가정적 환경)

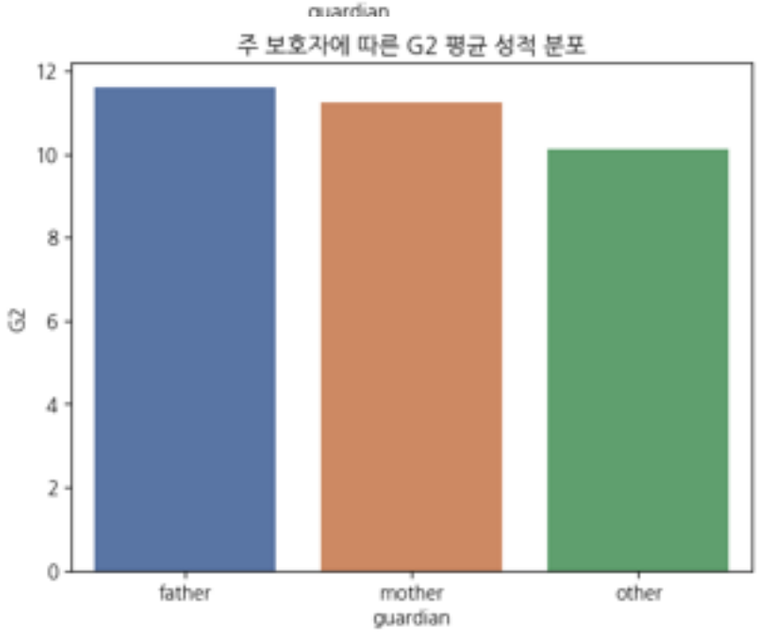
guardian

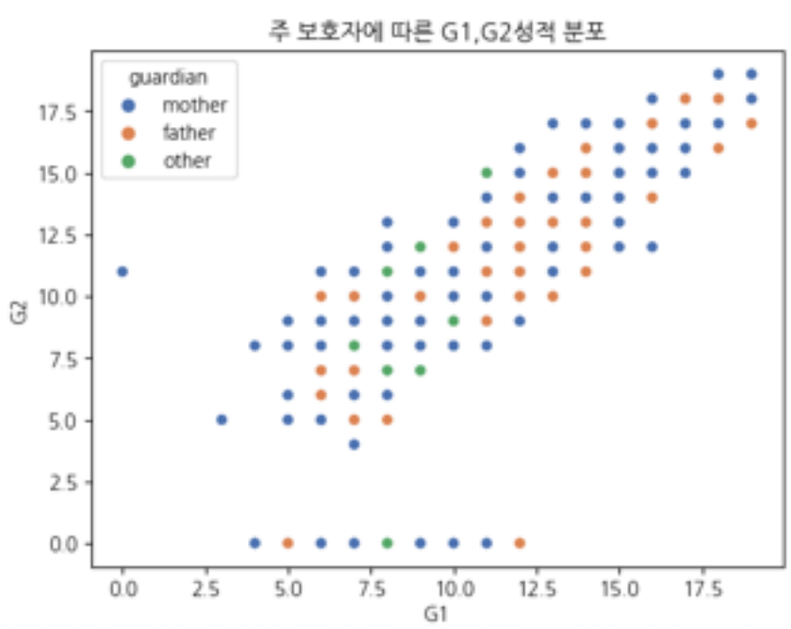




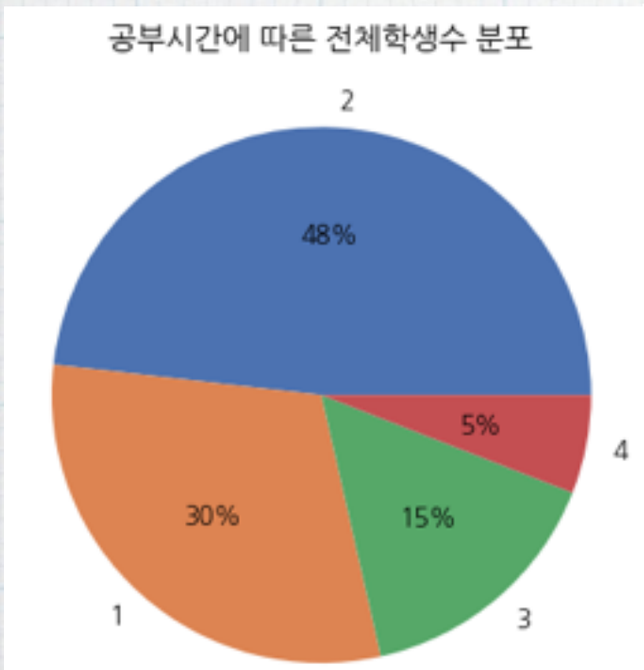


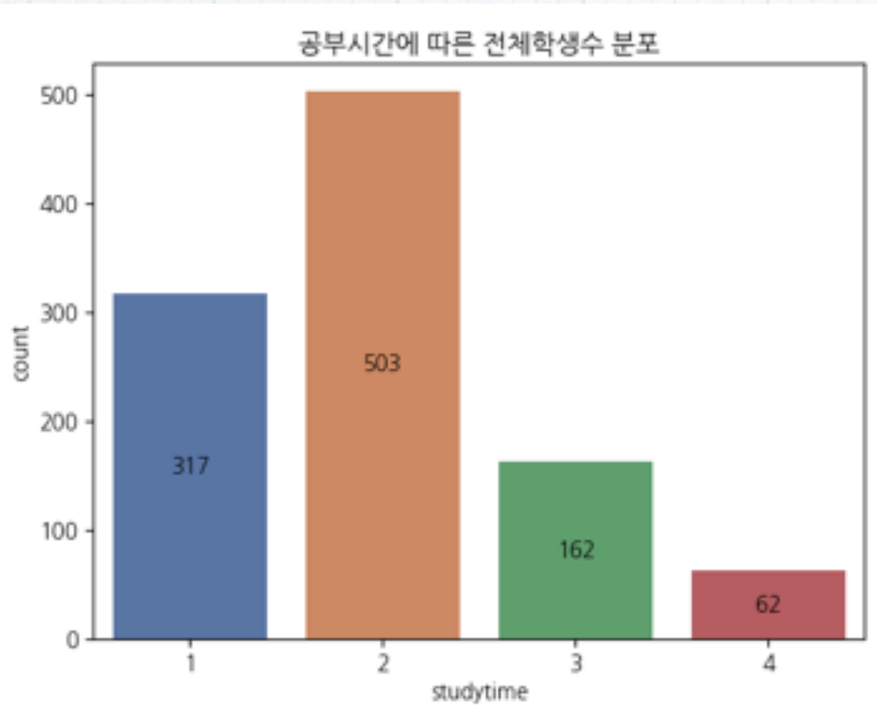


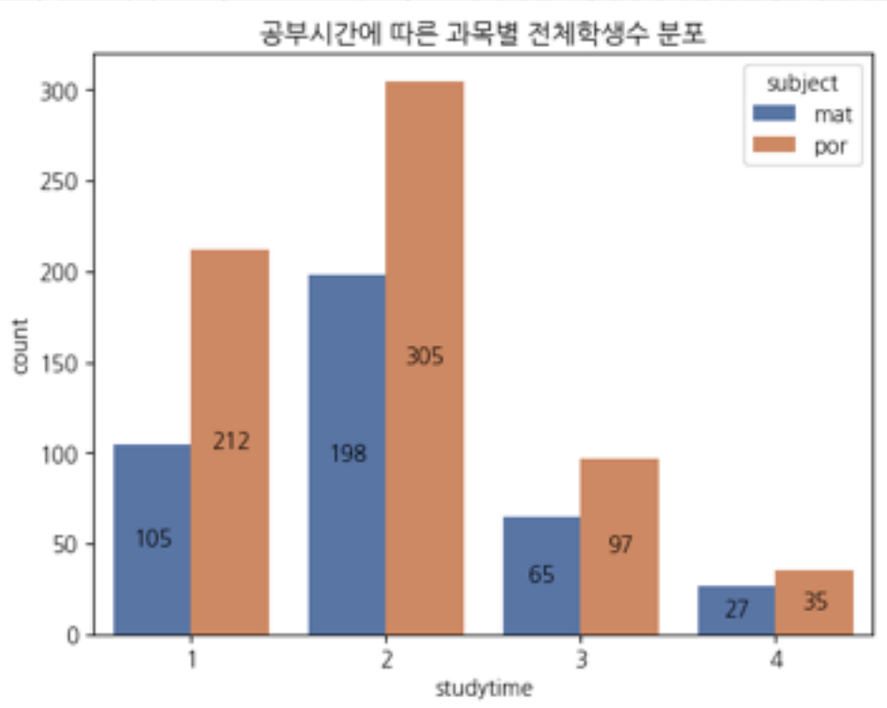


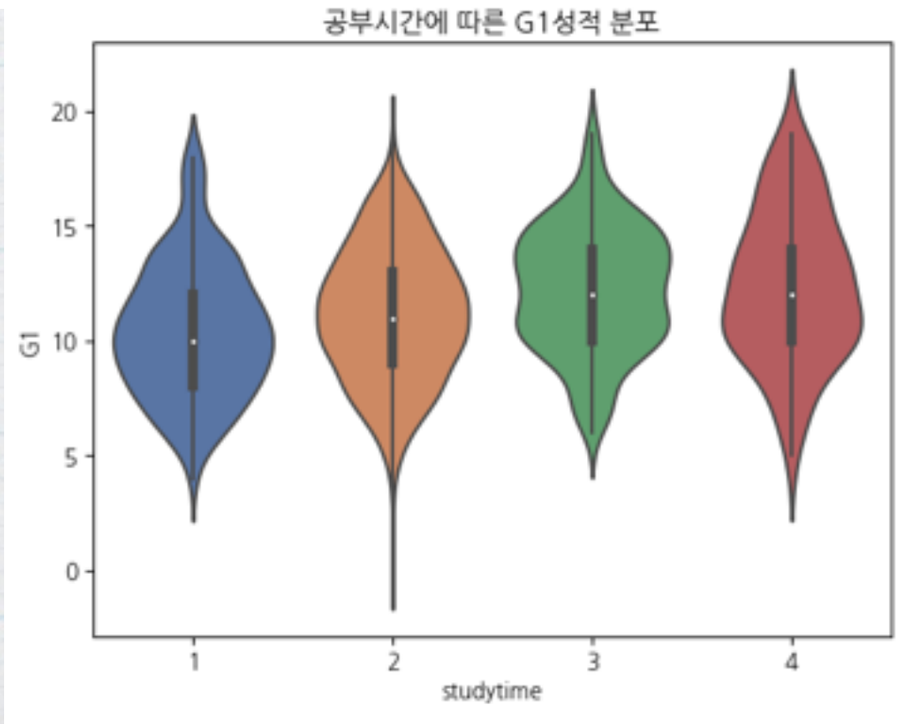


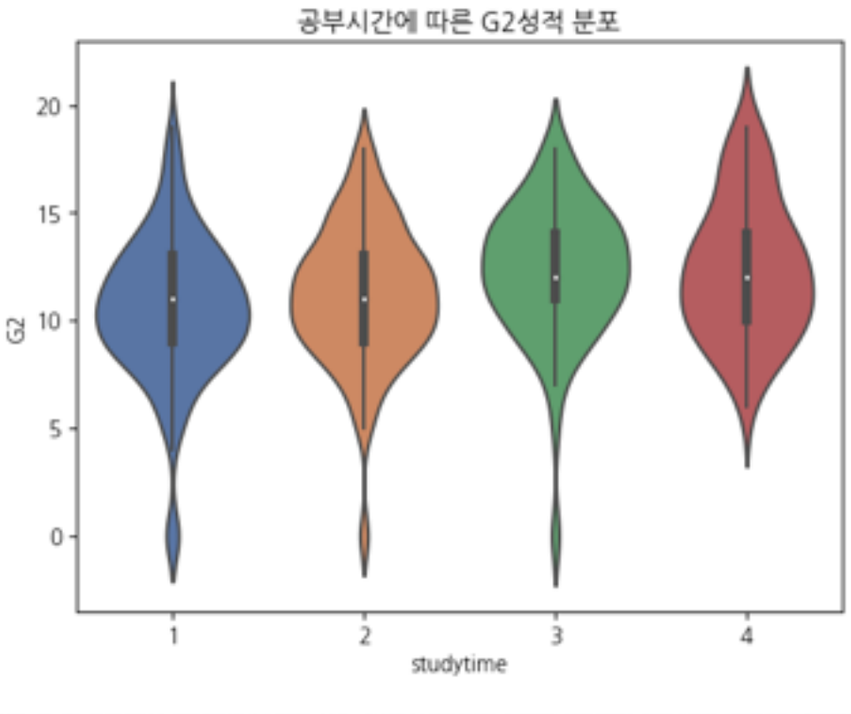
카테고리별 EDA분석(사회적 환경)
study time



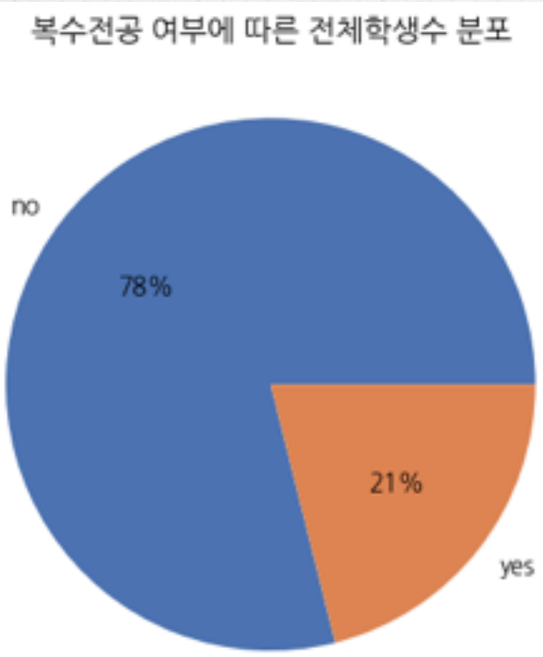


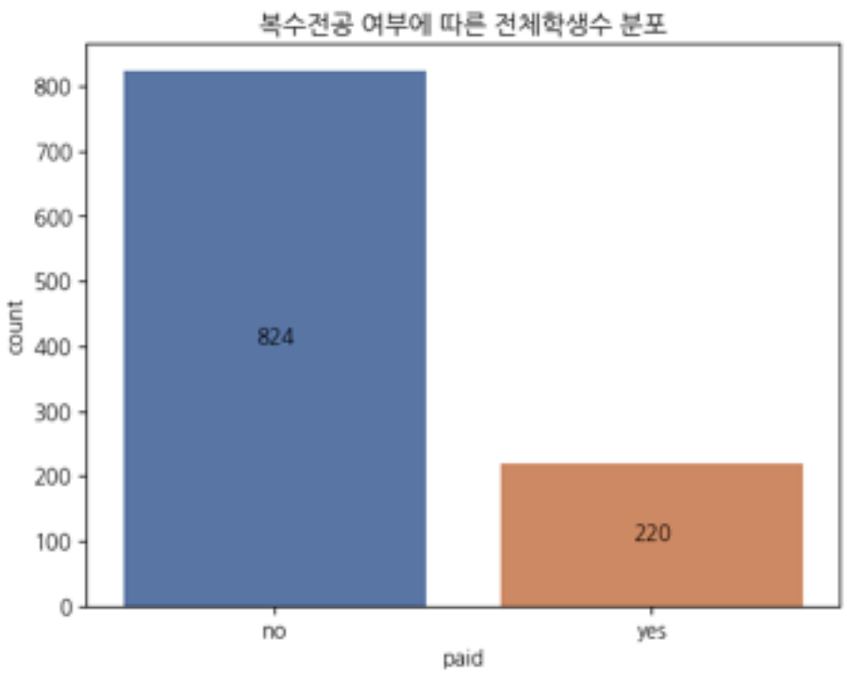


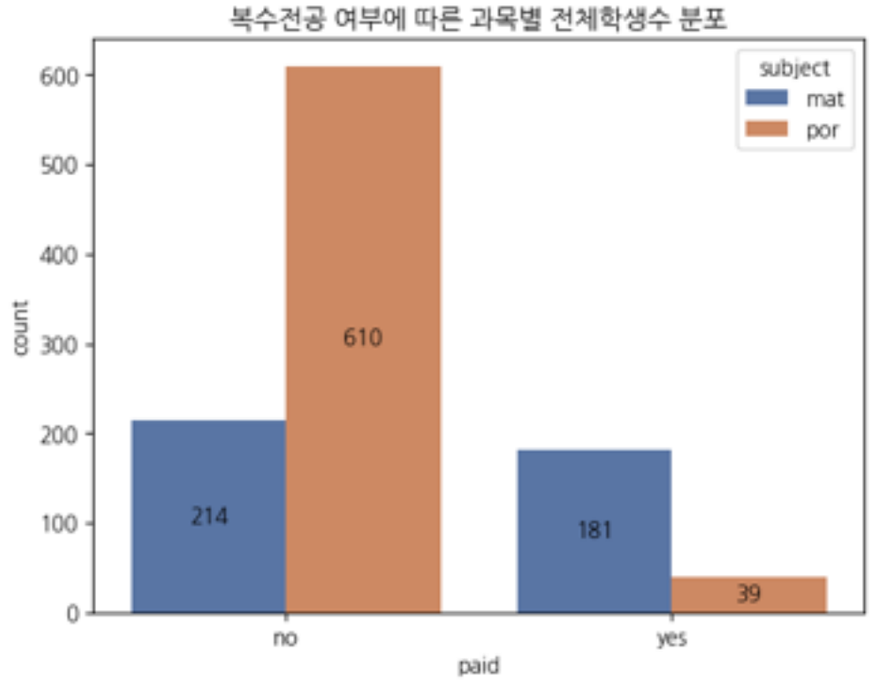


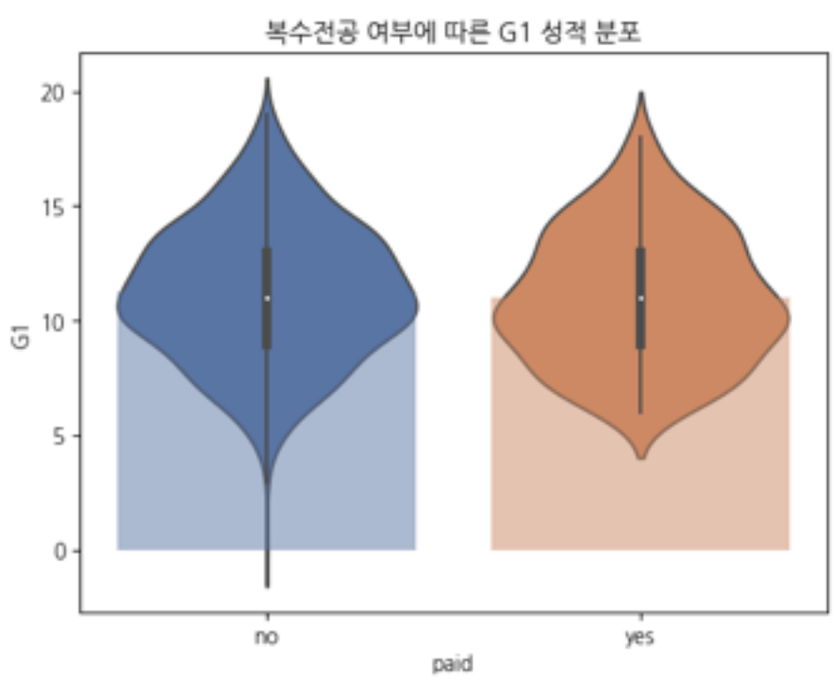


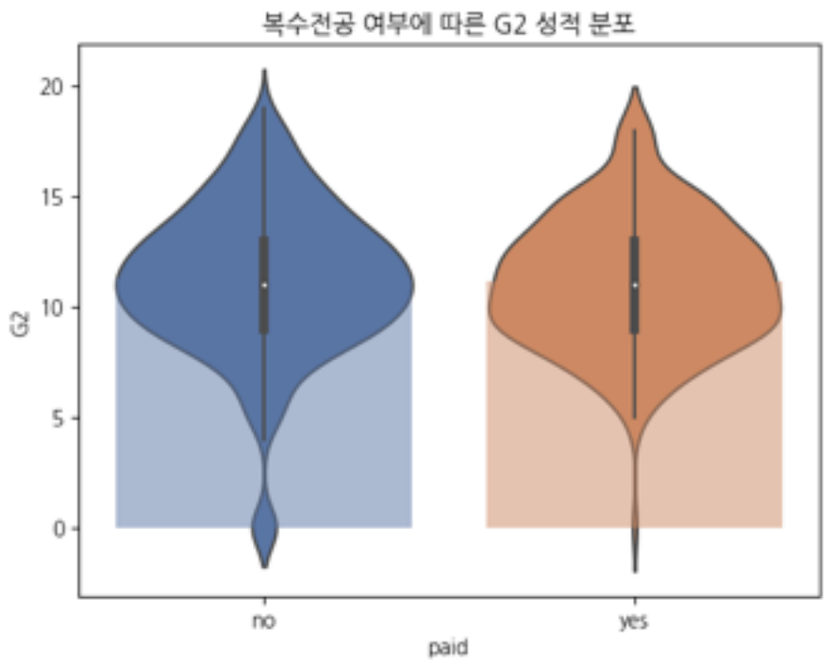
카테고리별 EDA분석(사회적 환경)
paid





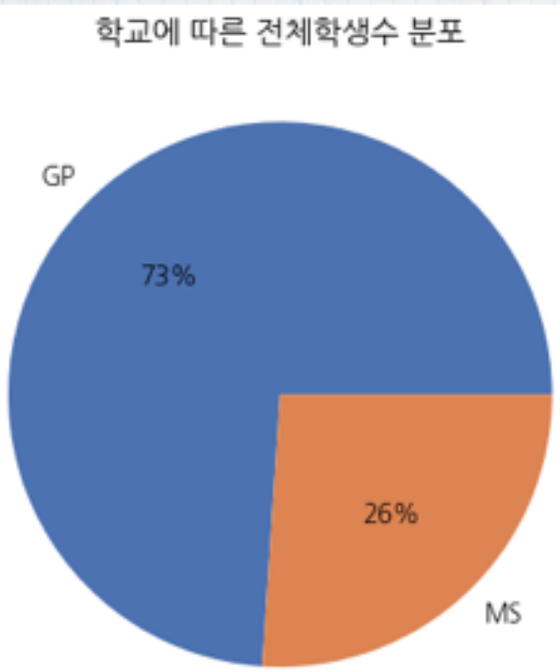


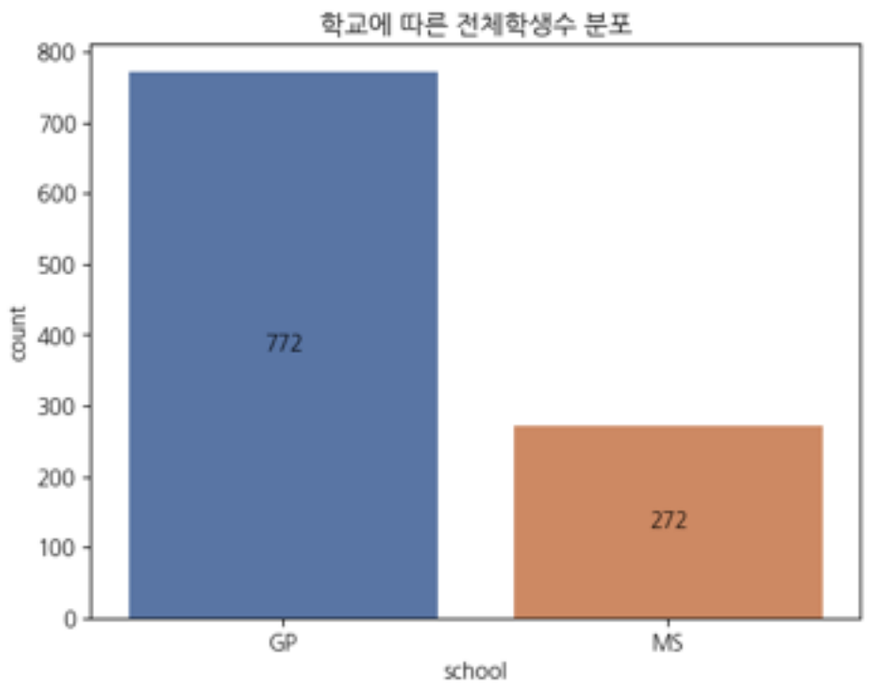


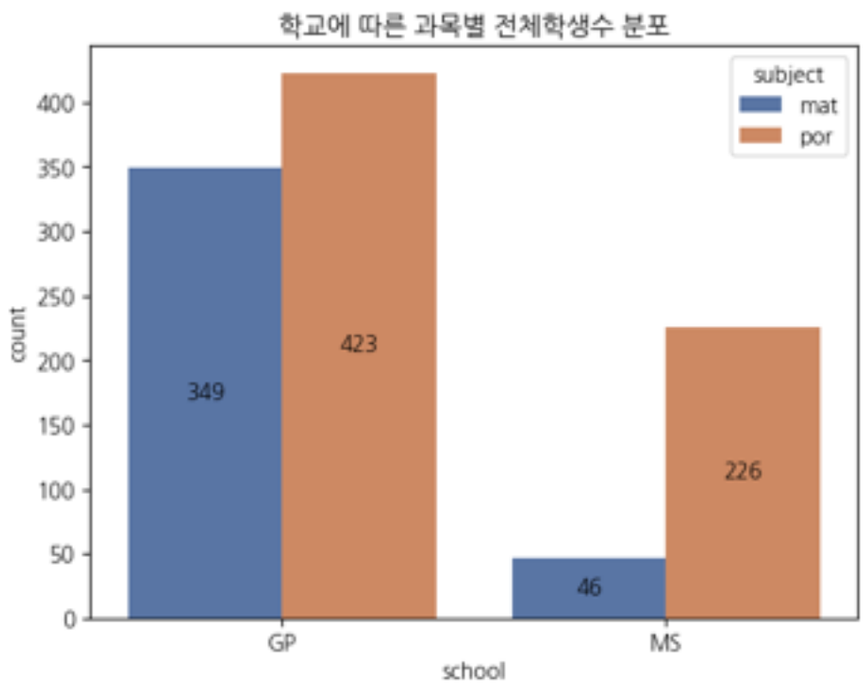


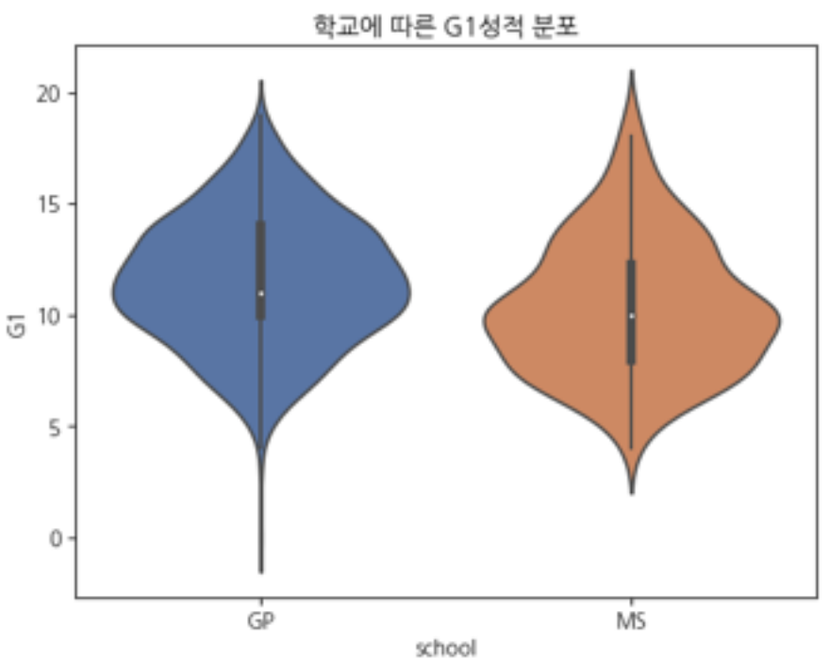
카테고리별 EDA분석(사회적 환경)

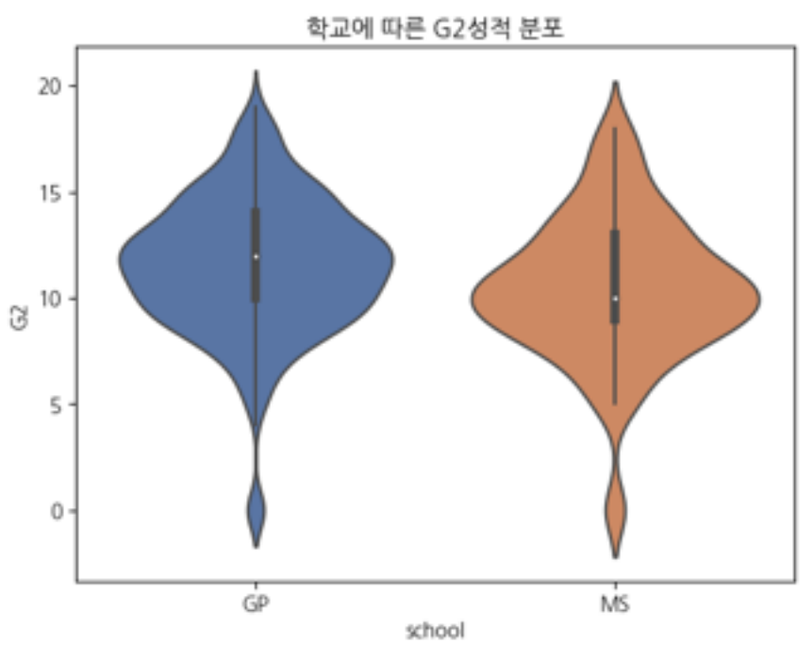
school





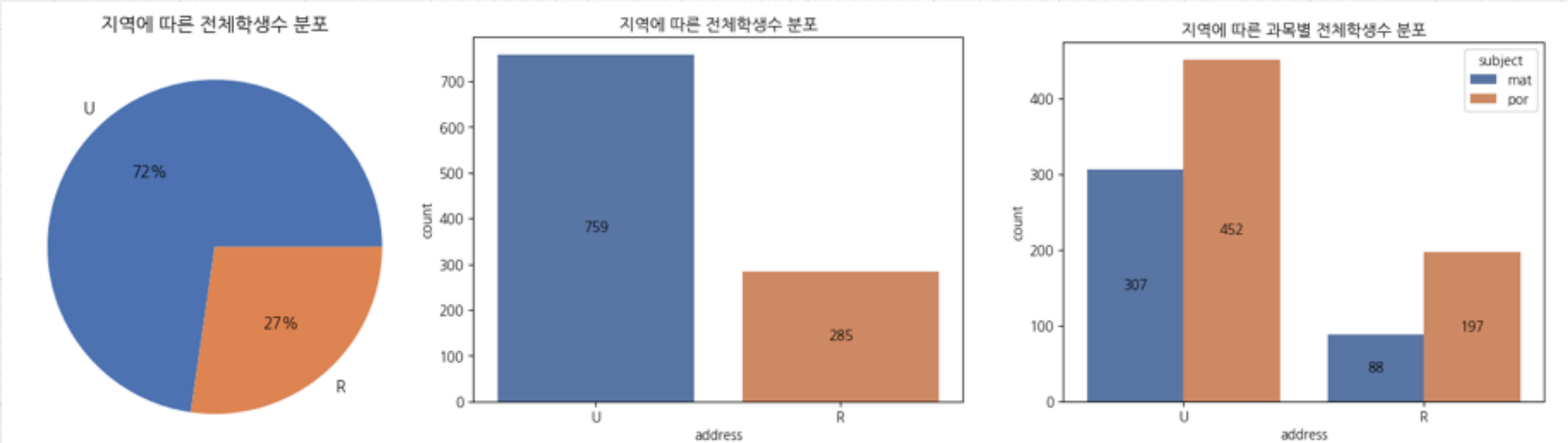


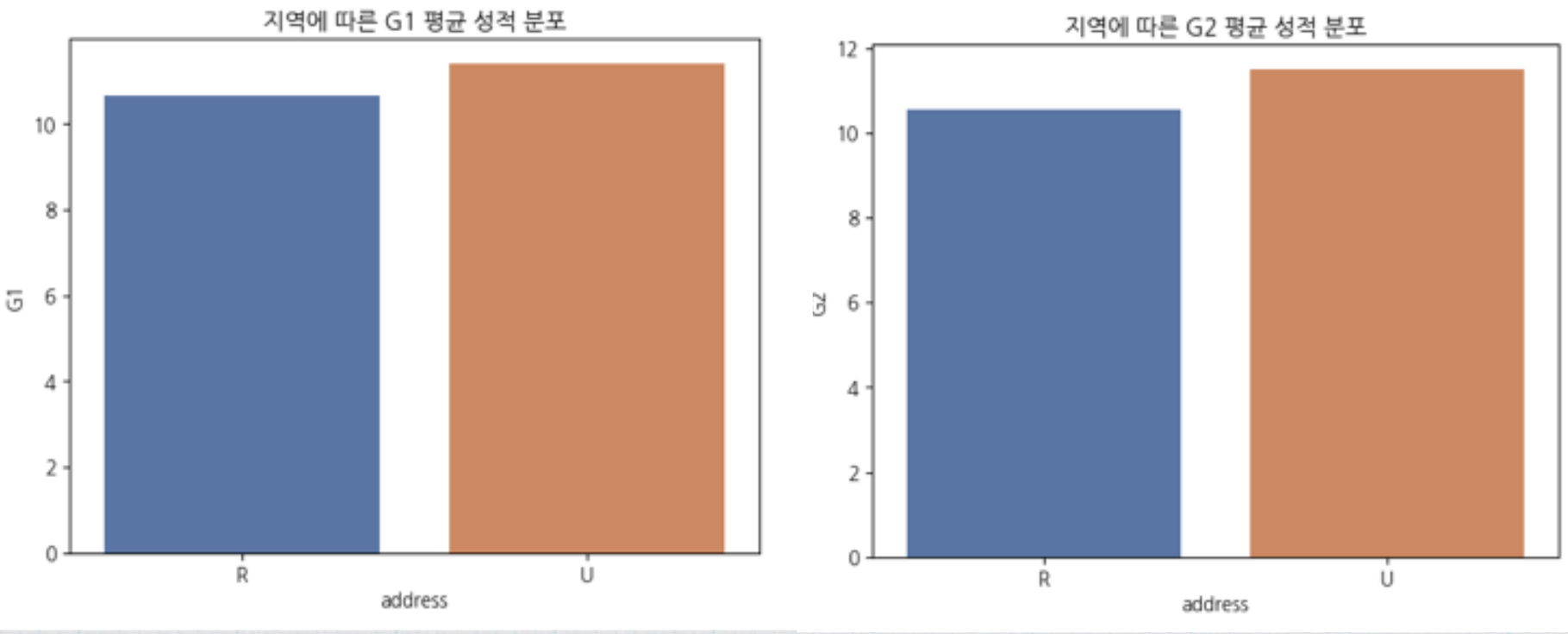




카테고리별 EDA분석(사회적 환경)

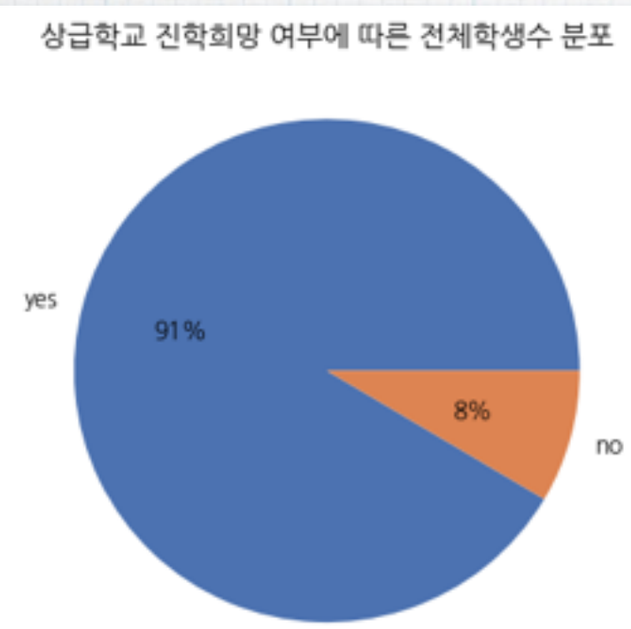
address

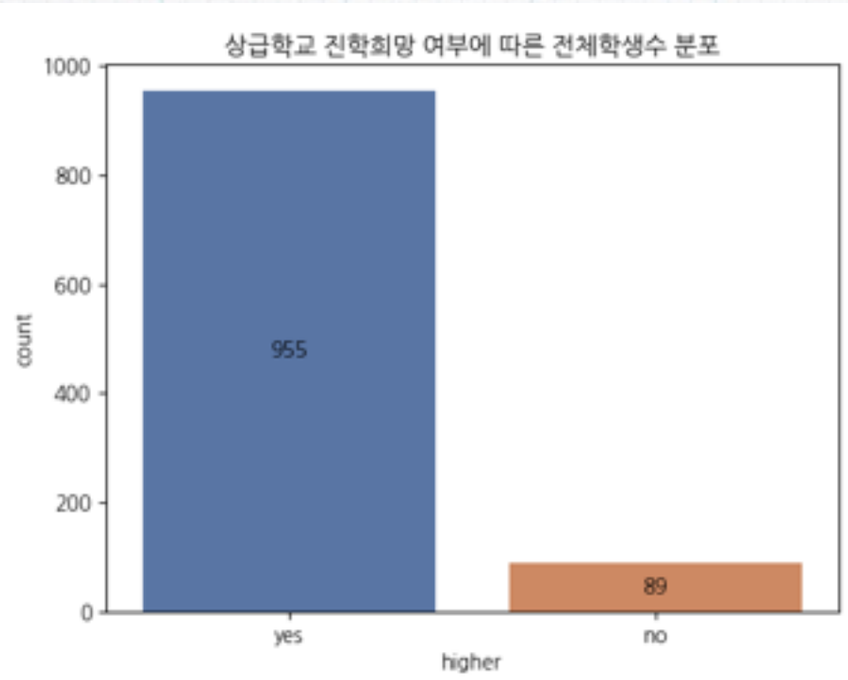


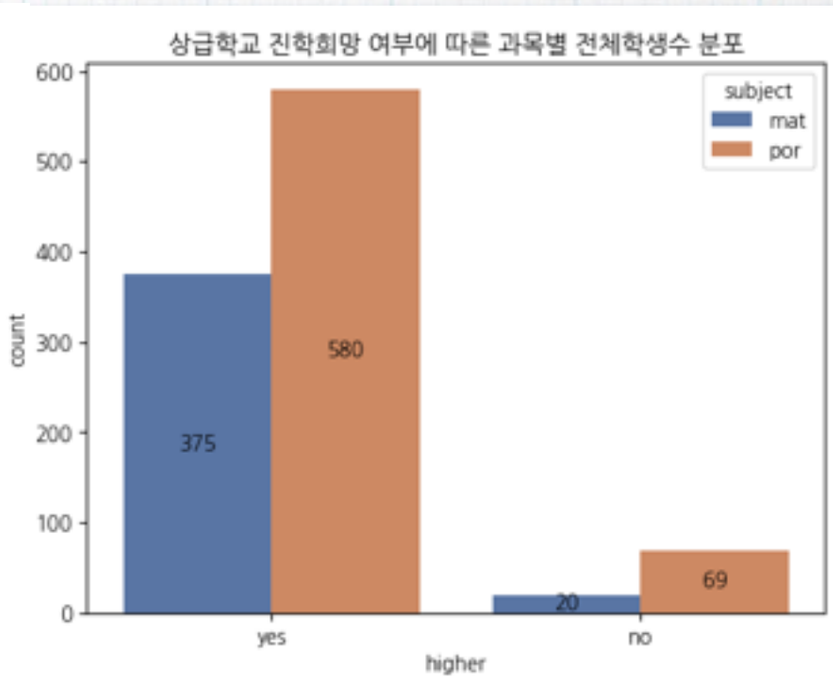


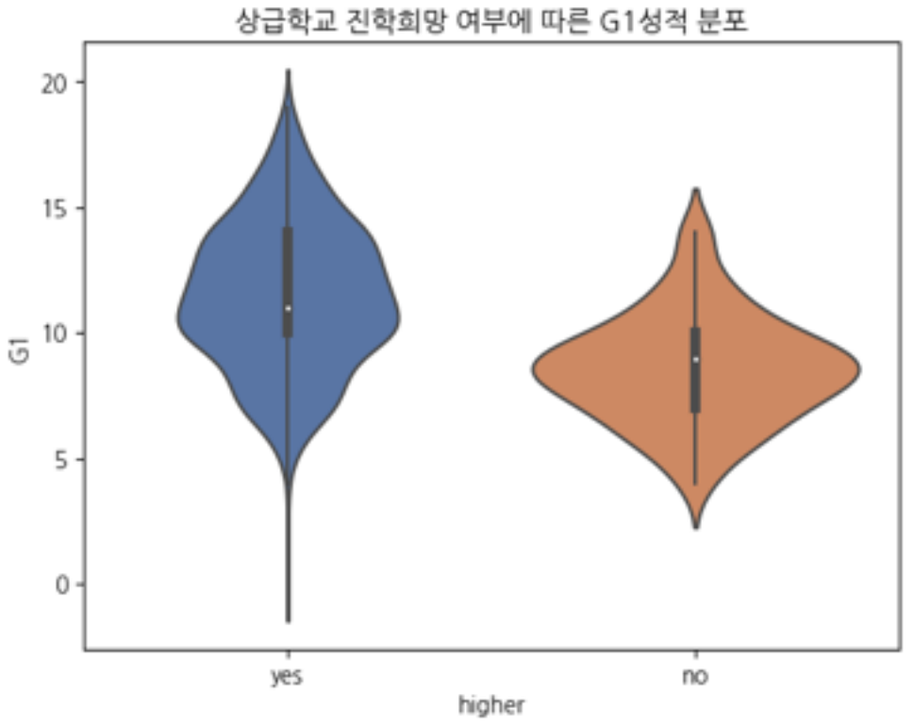
카테고리별 EDA분석(사회적 환경)

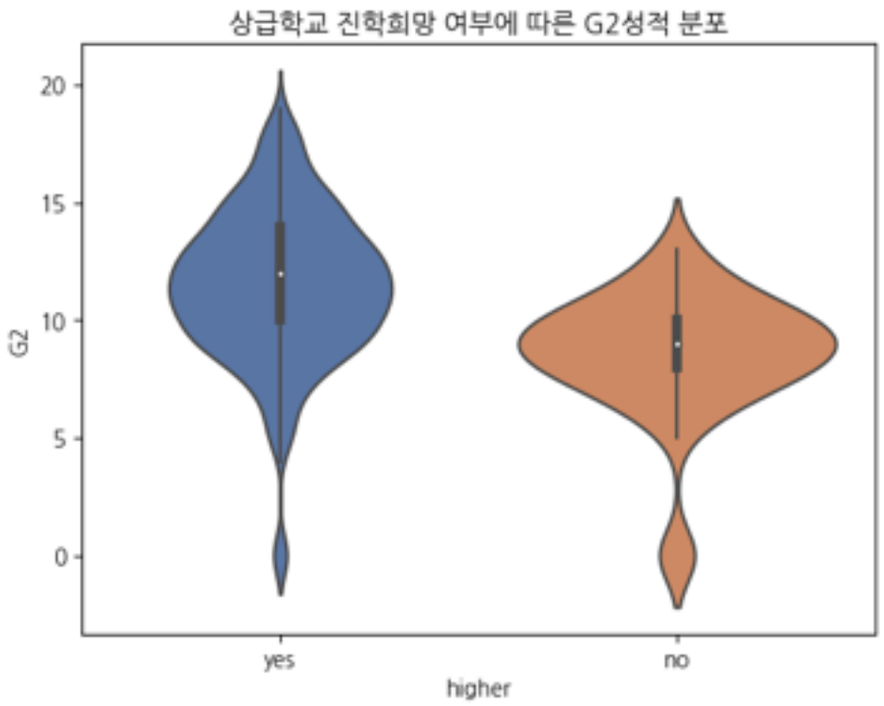
higher











카테고리별 컬럼 추출

각 카테고리별 성적과

가장 연관성 높다고 판단된 컬럼 추출

데이터셋

개인적 요인

Dalc
activities
sex
freetime
health
failures
Walc
age
goout
romantic

가정적 요인

famsize
Pstatus
farmel
internet
Fedu
Medu
Mjob
Fjob
guardian
famsup
internet

사회적 요인

schoolsup
nursery
address
absences
study time
paid
reason
traveltime
address
school
higher

데이터셋

개인적 요인

Dalc

failures

Walc

romantic

가정적 요인

Medu

Fedu

Fjob

Mjob

guardian

사회적 요인

address

study time

paid

address

school

higher

activities

goout

age

freetime

famsup

famsize

schoolsup

nursery

reason

health

sex

internet

Pstatus

internet

traveltime

absences

상관관계 분석

과 추출된 컬럼 비교

failures

higher

Medu

study time

Fedu

school

Dalc

Walc

Fjob

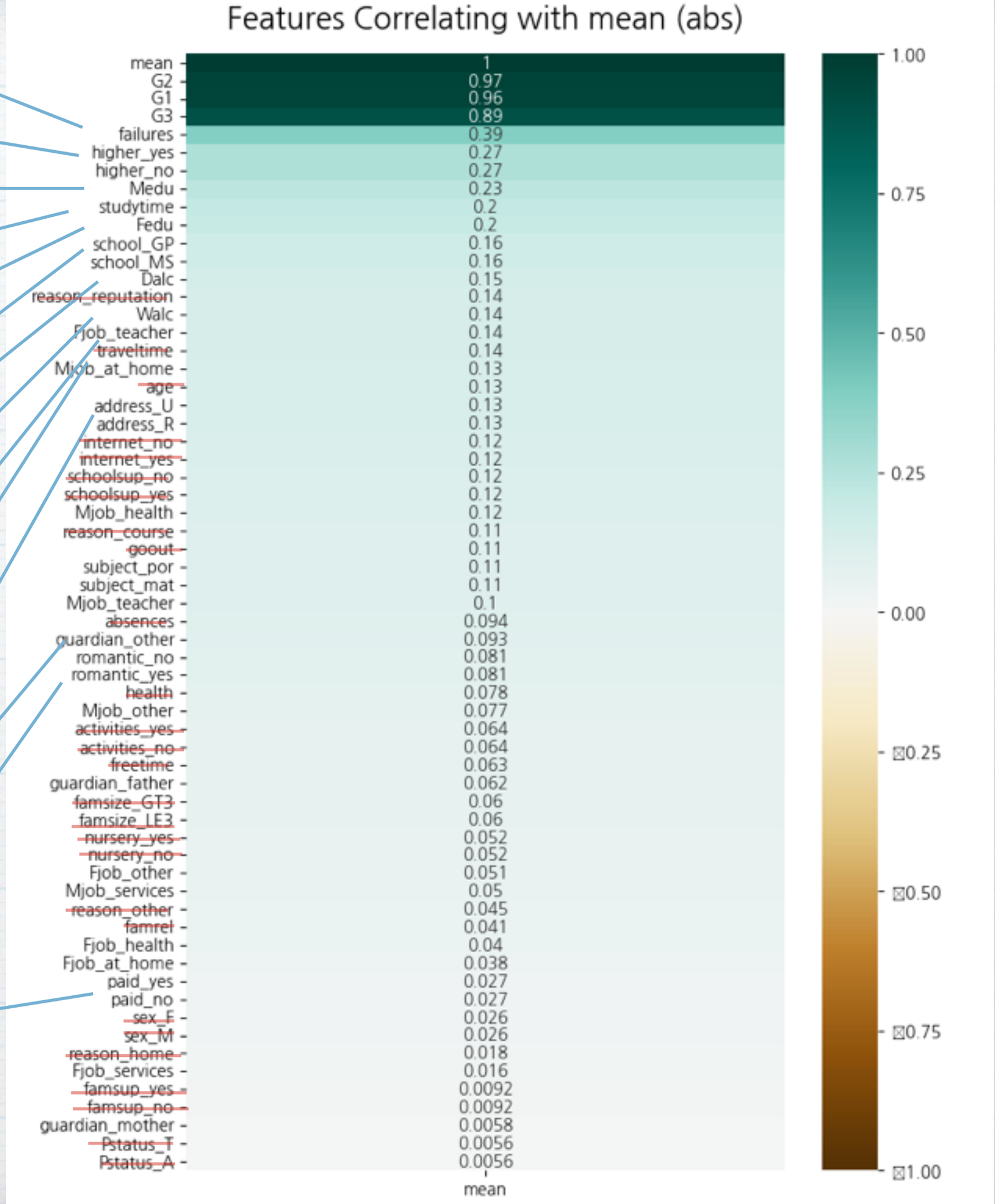
Mjob

address

guardian

romantic

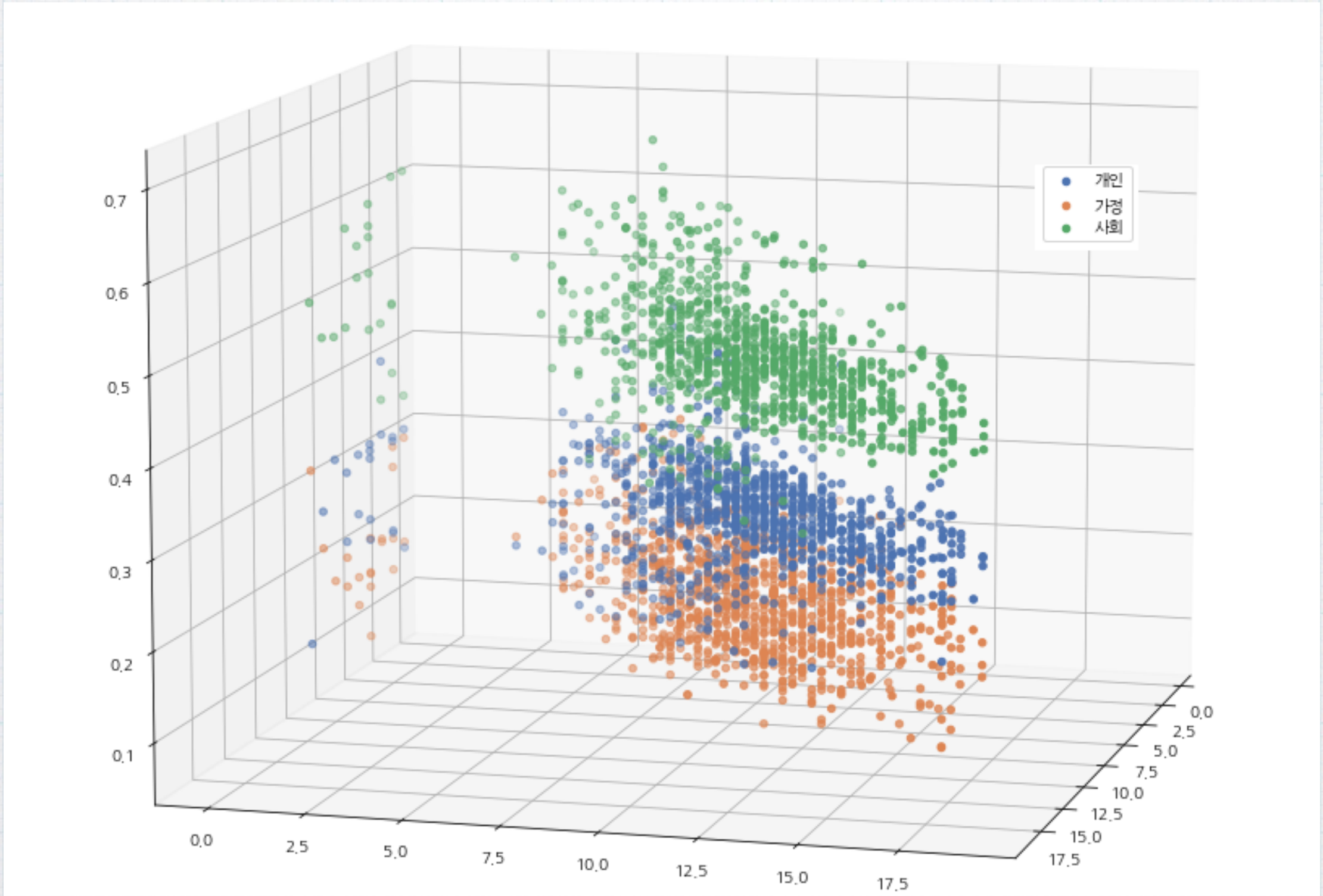
paid



추출된 컬럼

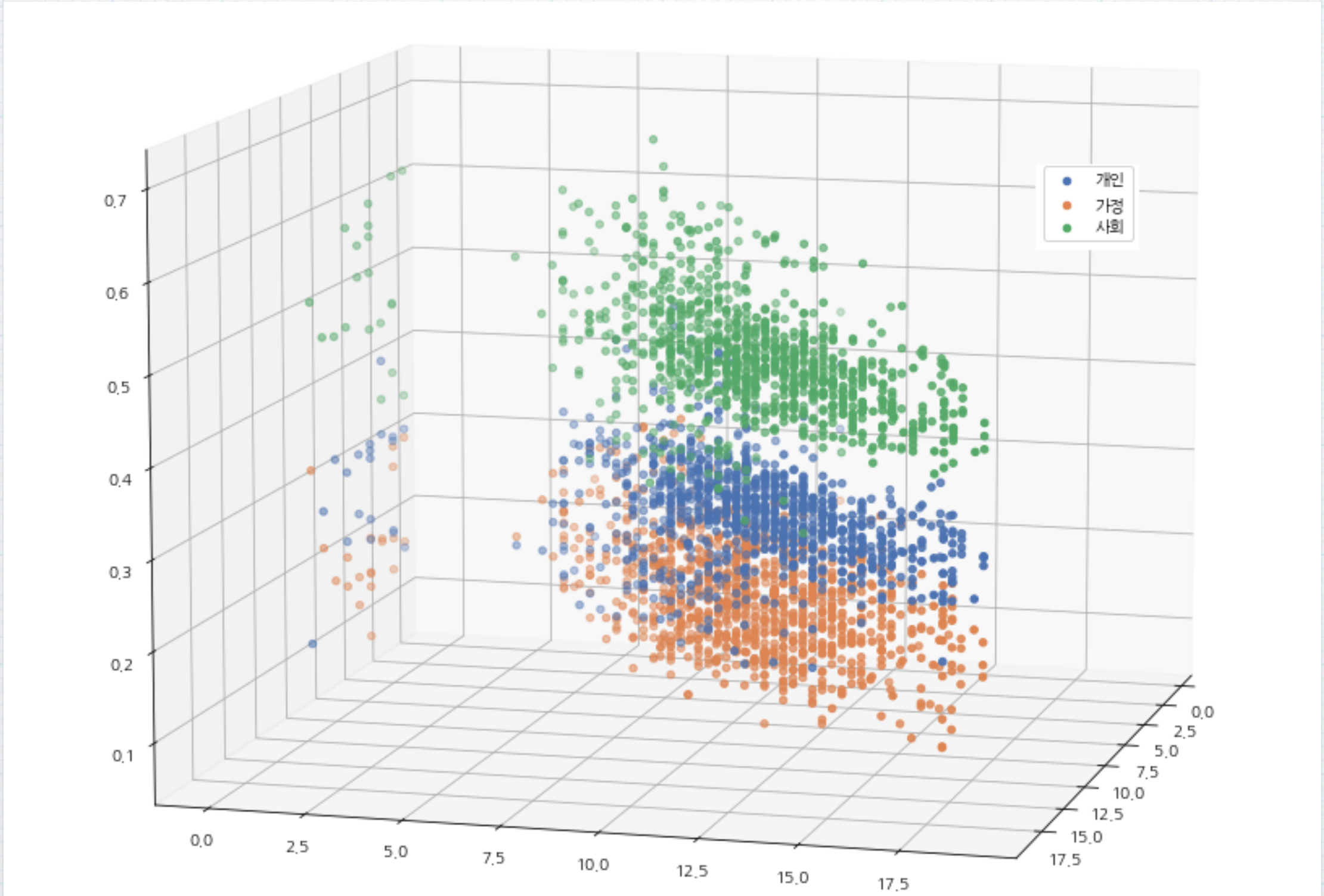
으로

클러스터링 분석


개인
가정
사회
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

결론

클러스터링 결과 고찰


개인
가정
사회
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

향후 과제

K-mean클러스터링 분석
을 통해 정교한 군집분석 제시

EDA를 통해 데이터
에 근거한 카테고리 분류

Roll

조수익

설계

학교 EDA분석

발표

이경환

기획

가정 EDA분석

데이터 해석

정순주

자료준비

개인 EDA분석

코드리뷰

Refer.

https://github.com/springwater00/Datathon

https://seaborn.pydata.org/

https://wikidocs.net/137778

https://datascienceschool.net/01%20python/05.04%20시본을%20사용한%20데이터%20분포%20시각화.html