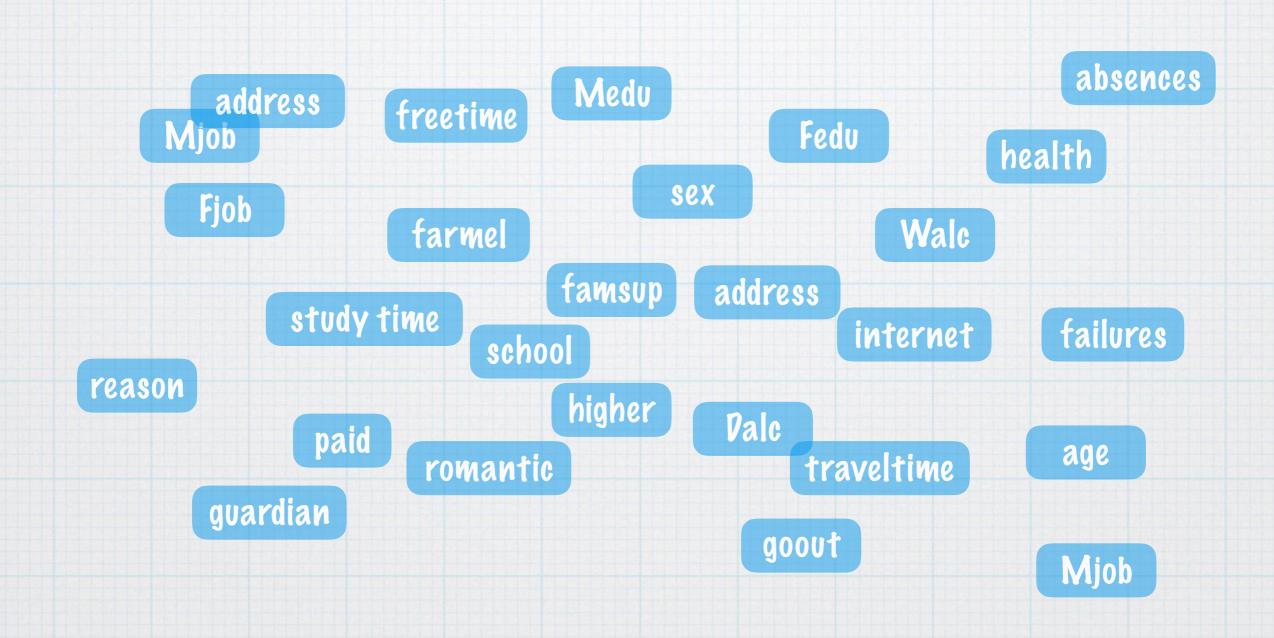
## 다금바리를 꿈꾸는 송사리

조수익이경환정순주

서른

### 데이터 개요



### 데이터개요

가정적 요인 개인적 요인 사회적 요인 Palc absences Medu address sex Fedu freetime study time Fjob health failures Mjob paid reason Walc famsup guardian traveltime address age goout internet higher school romantic

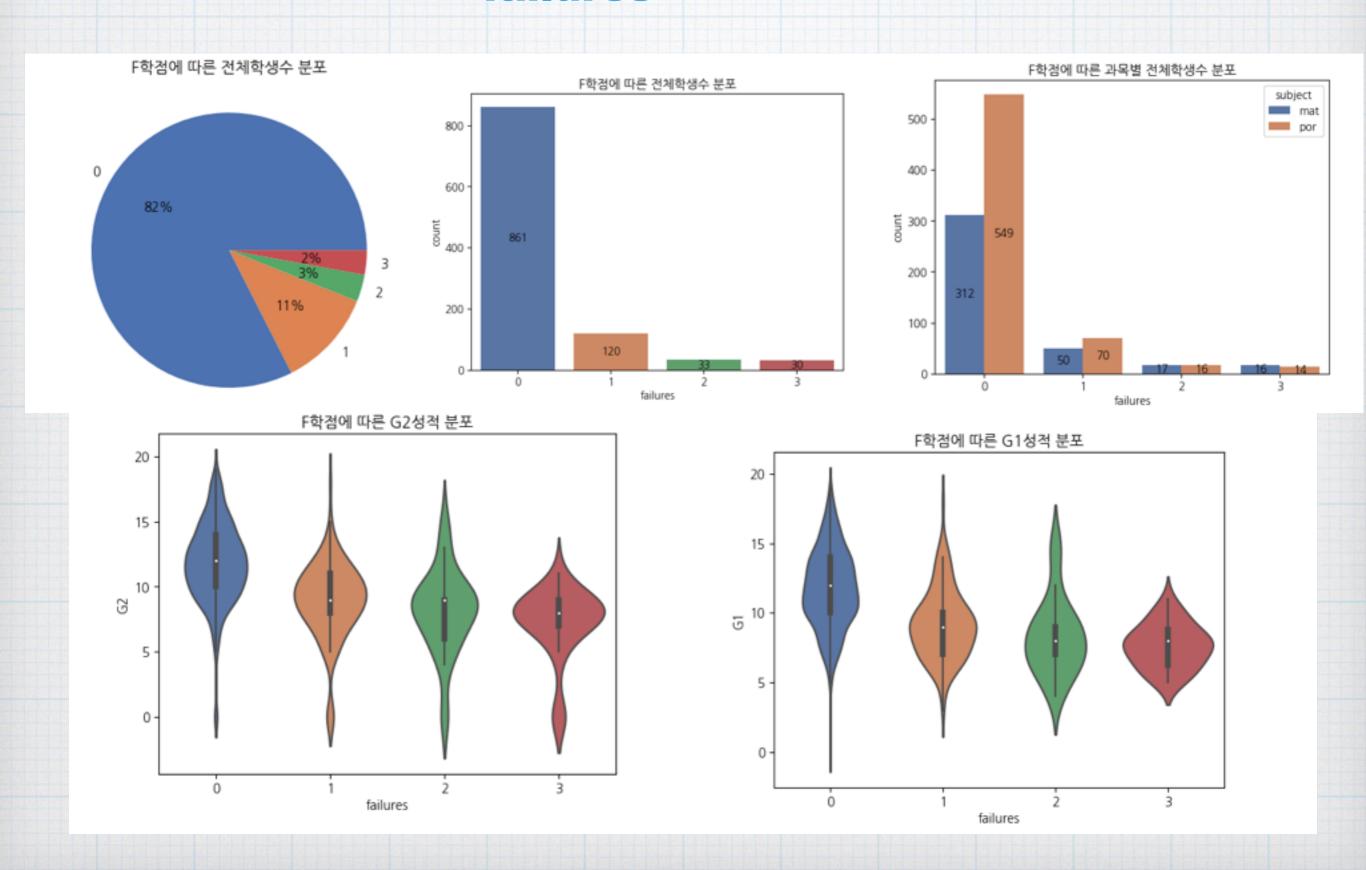
### 목표

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

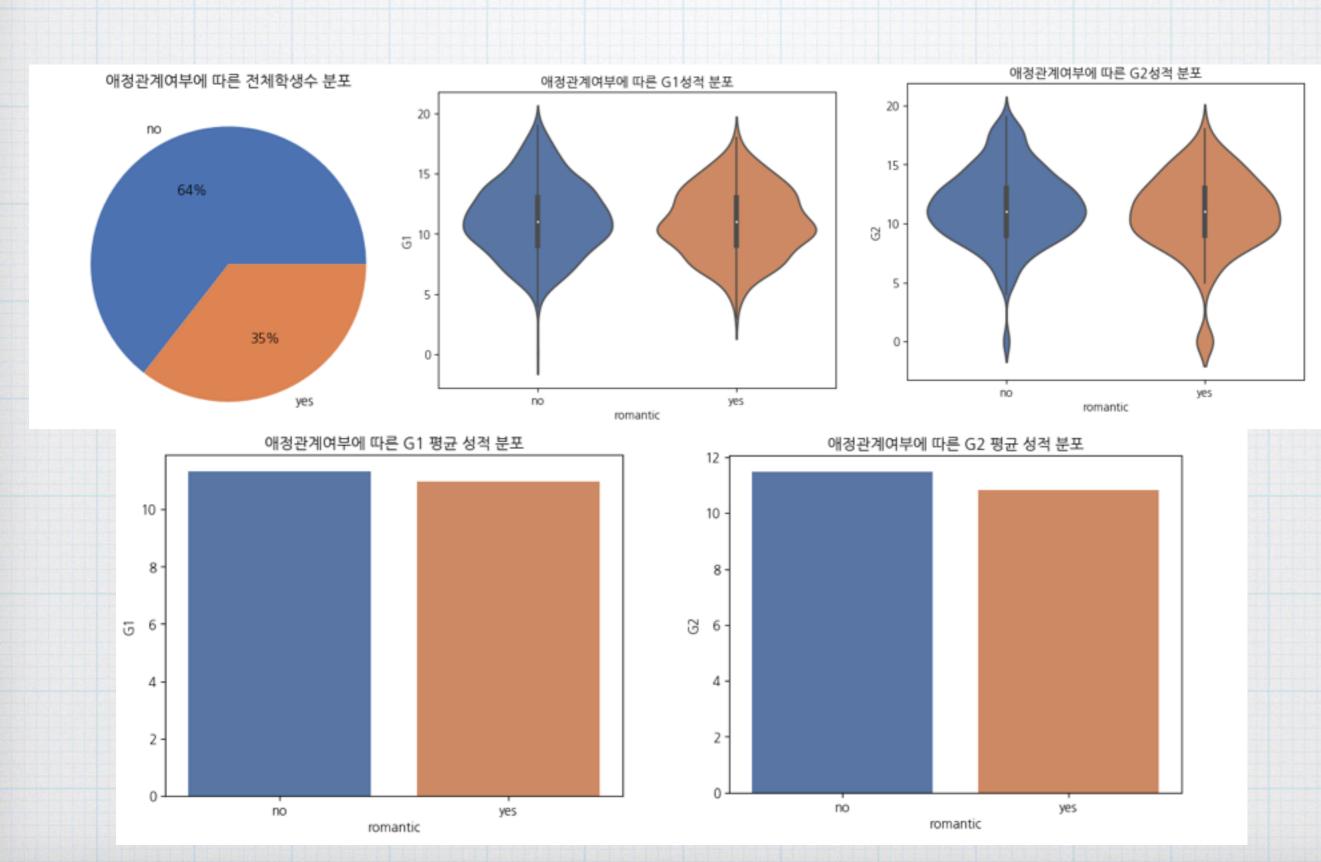
#### 카테고리별 EDA

### 개별 컬럼들을 EDA 를 통해 성적과의 연관성을 분석

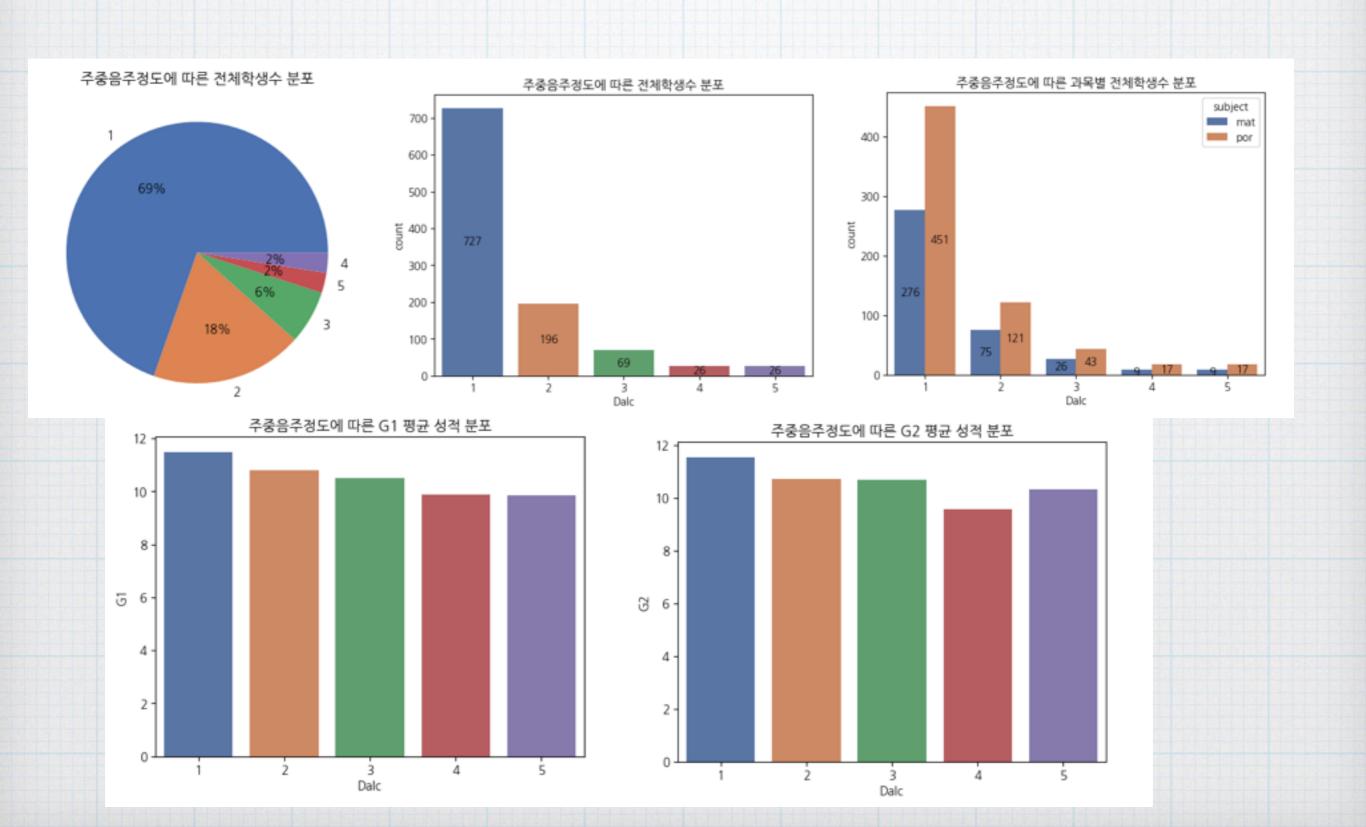
### 카테고리별 EDA분석(개인적 환경) failures



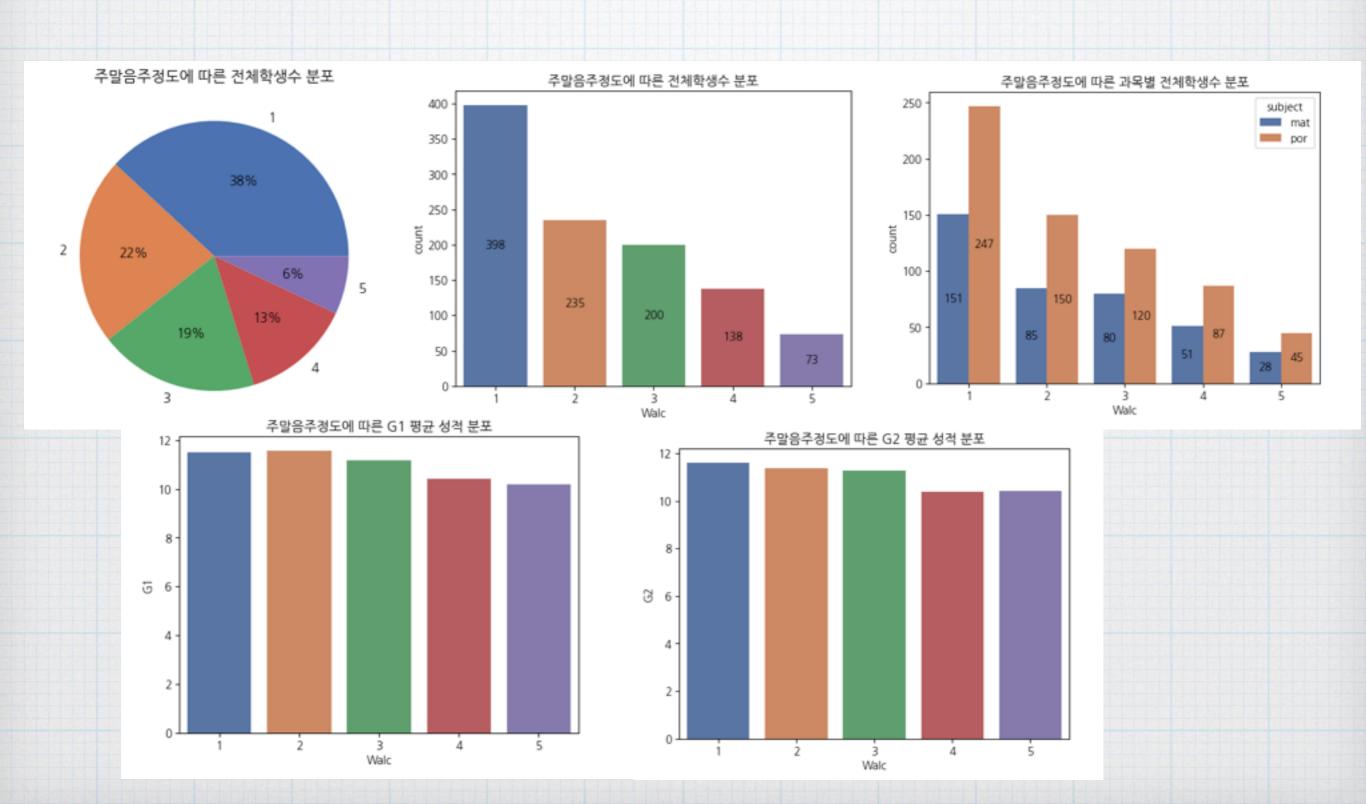
### 카테고리별 EDA분석(개인적 환경) romantic



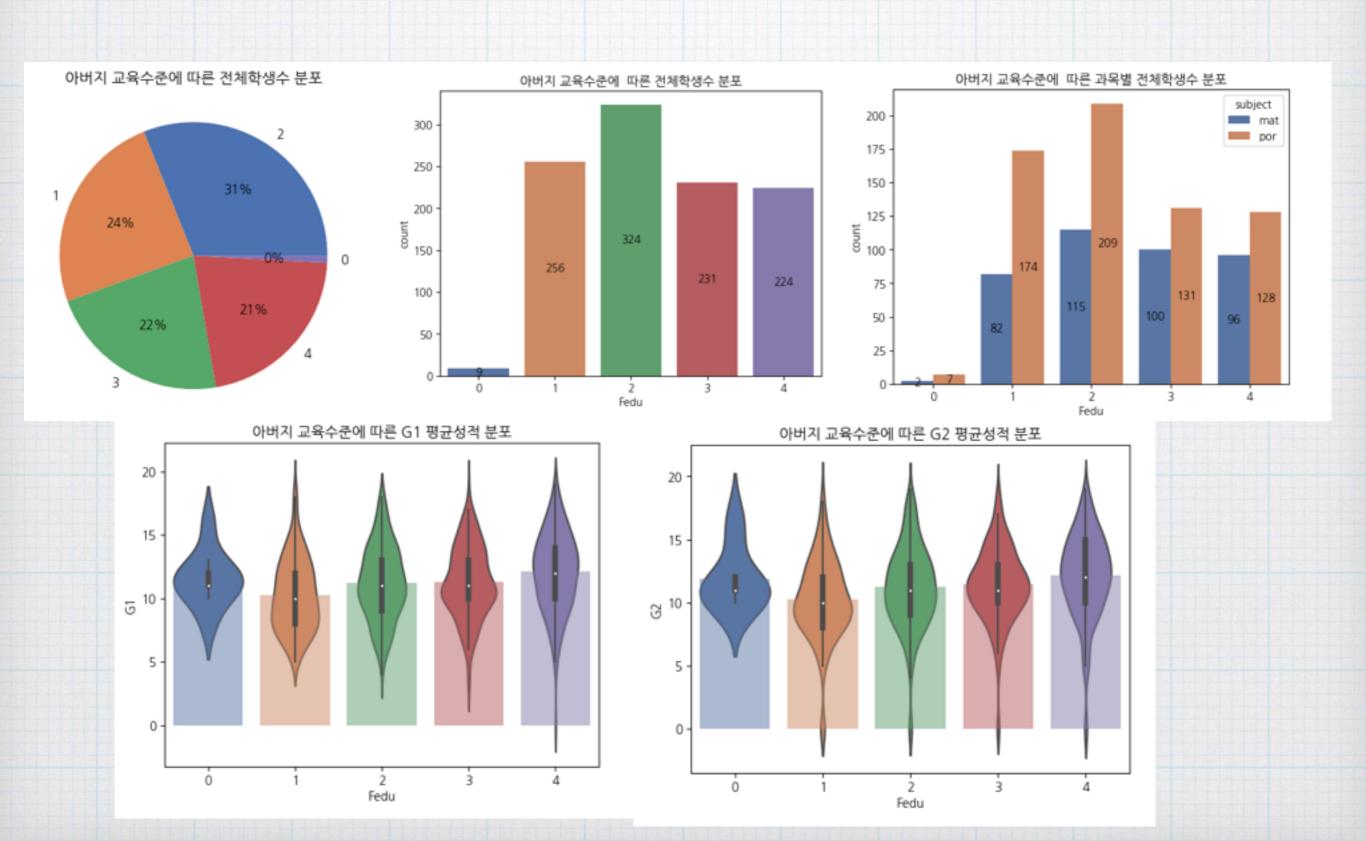
#### 카테고리별 EDA분석(개인적 환경) Dalc



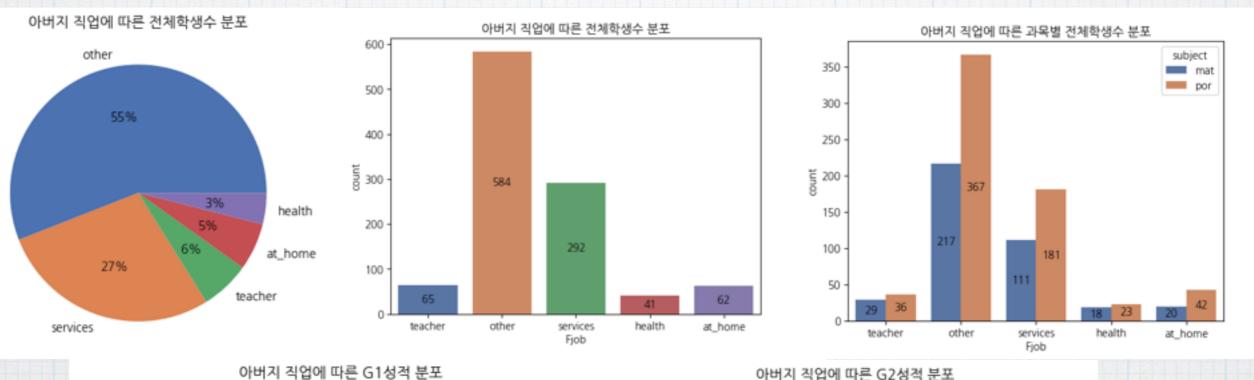
#### 카테고리별 EDA분석(개인적 환경) Walc



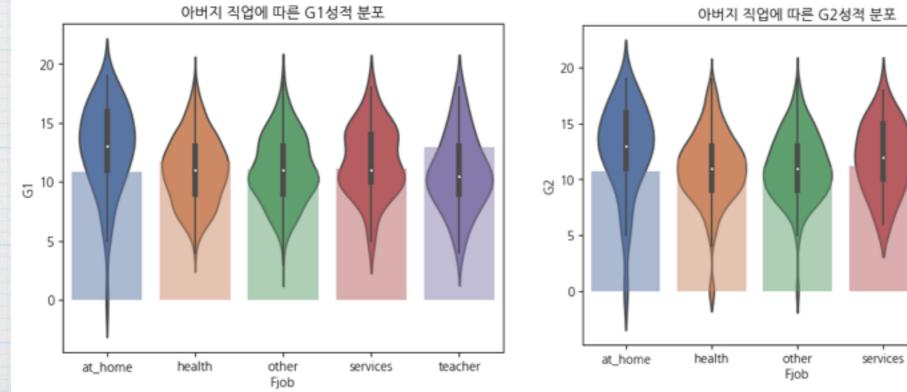
#### 카테고리별 EDA분석(가정적 환경) Fedu



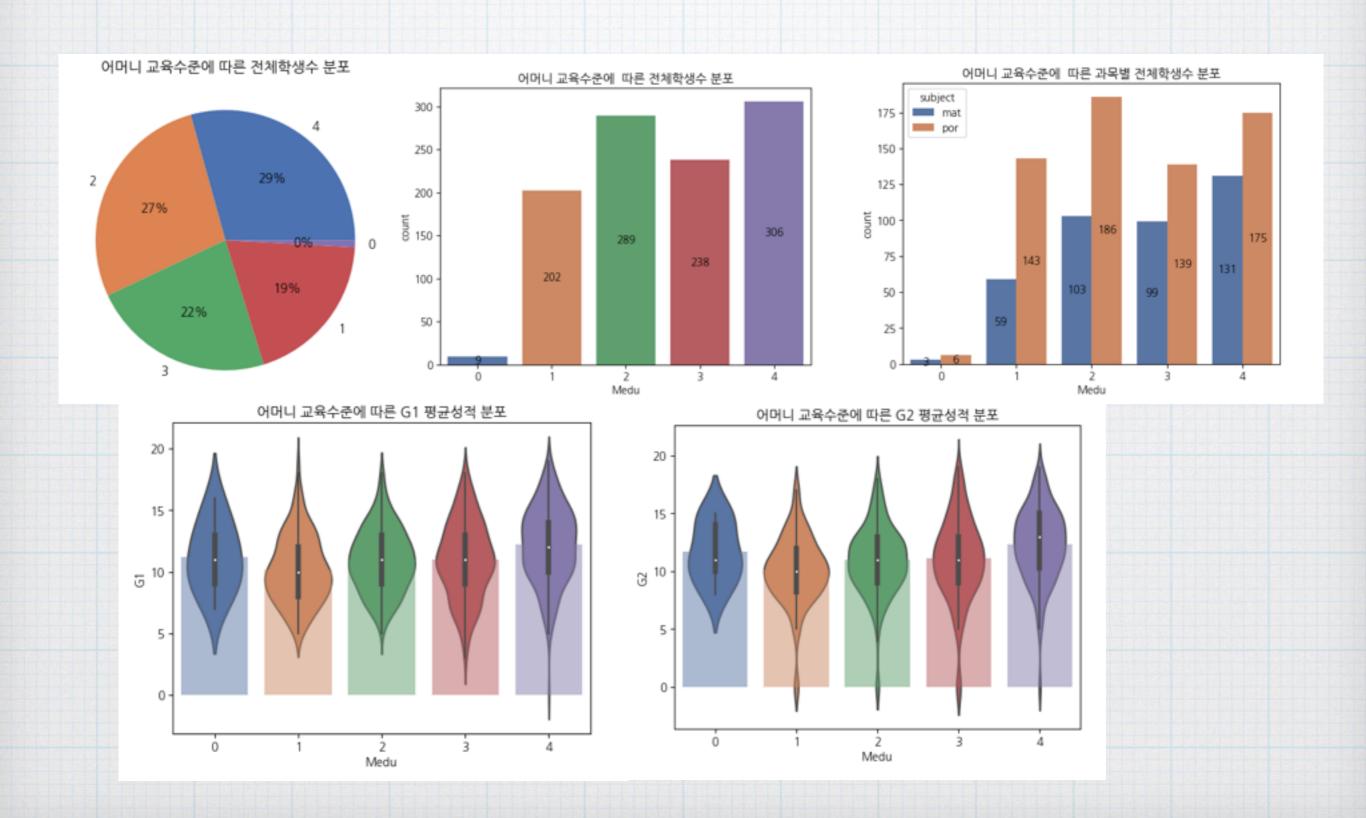
#### 카테고리별 EDA분석(가정적 환경) Fjob



teacher

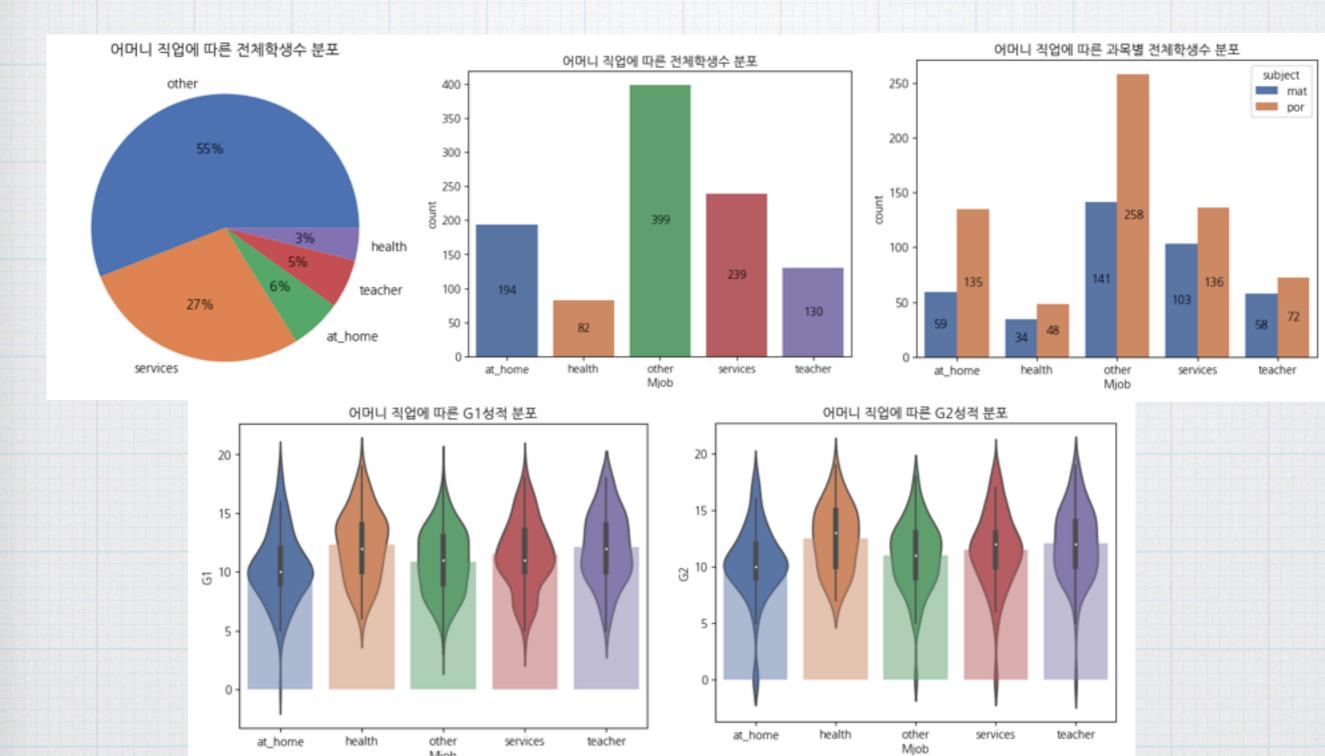


#### 카테고리별 EDA분석(가정적 환경) Medu

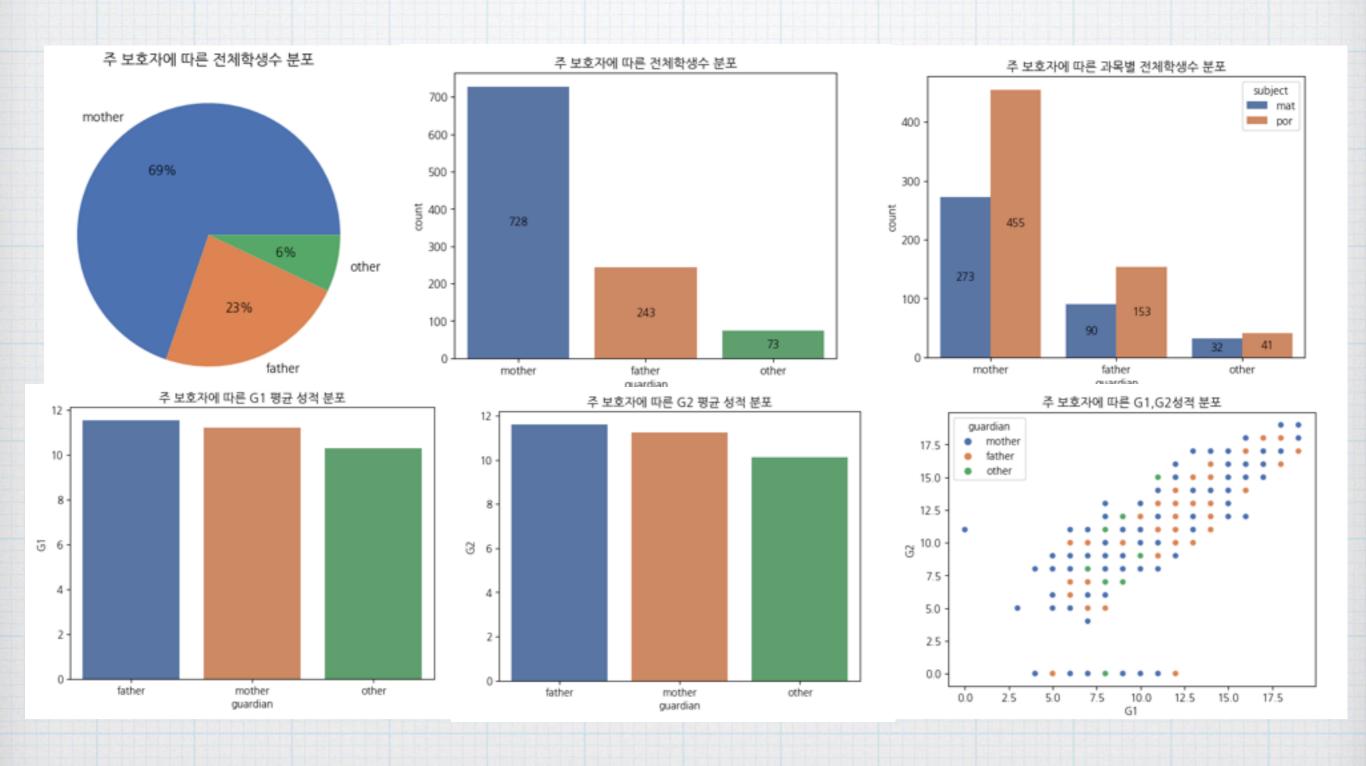


#### 카테고리별 EDA분석(가정적 환경) Mjob

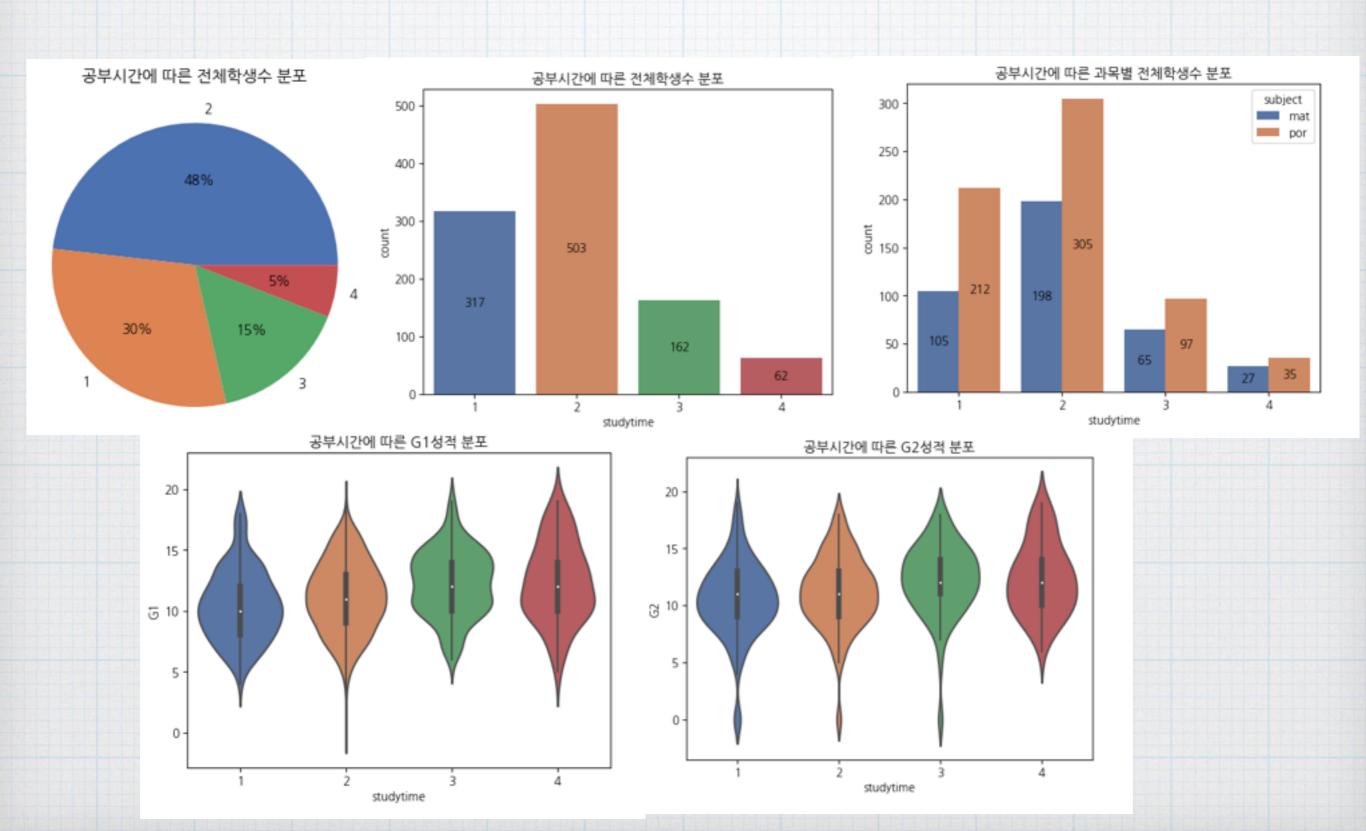
Mjob



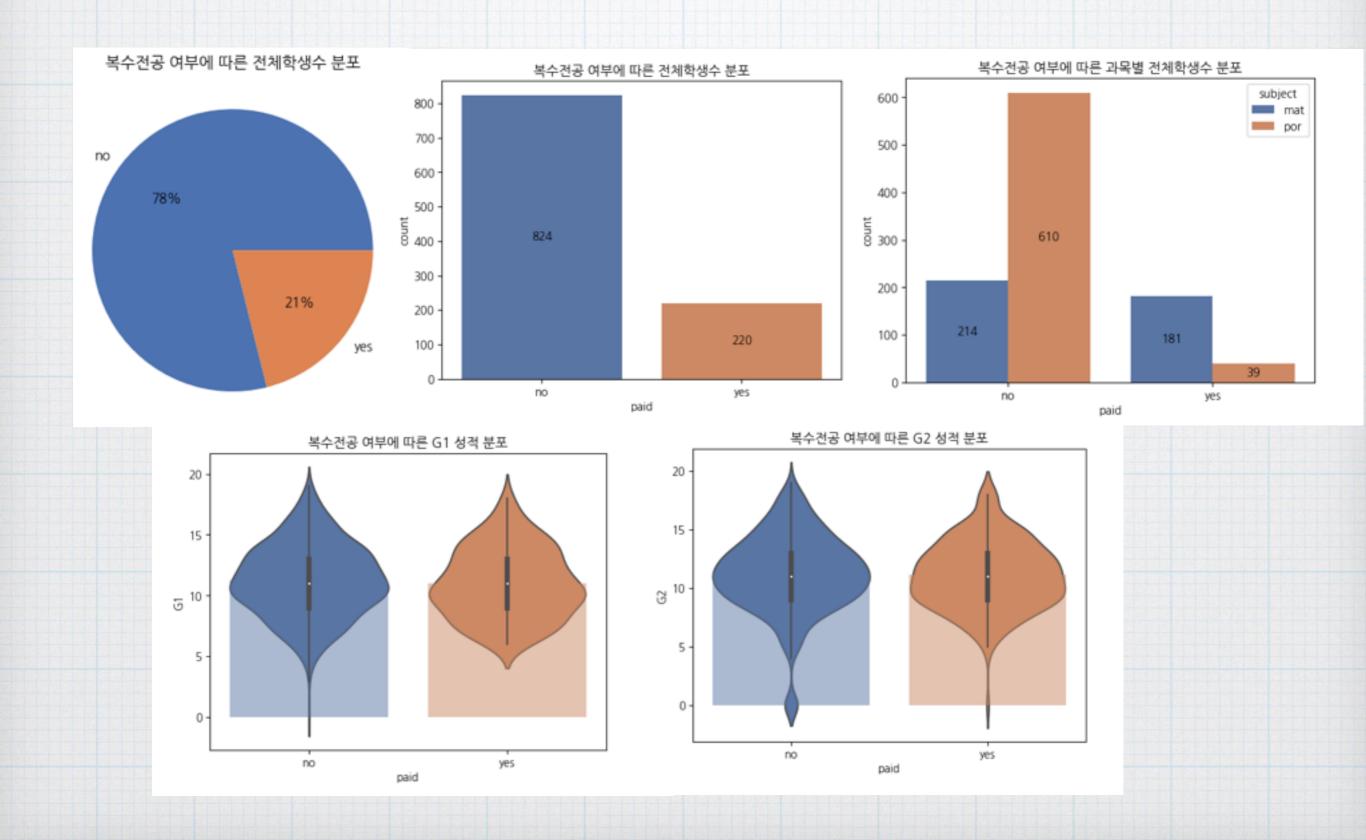
### 카테고리별 EDA분석(가정적 환경) guardian



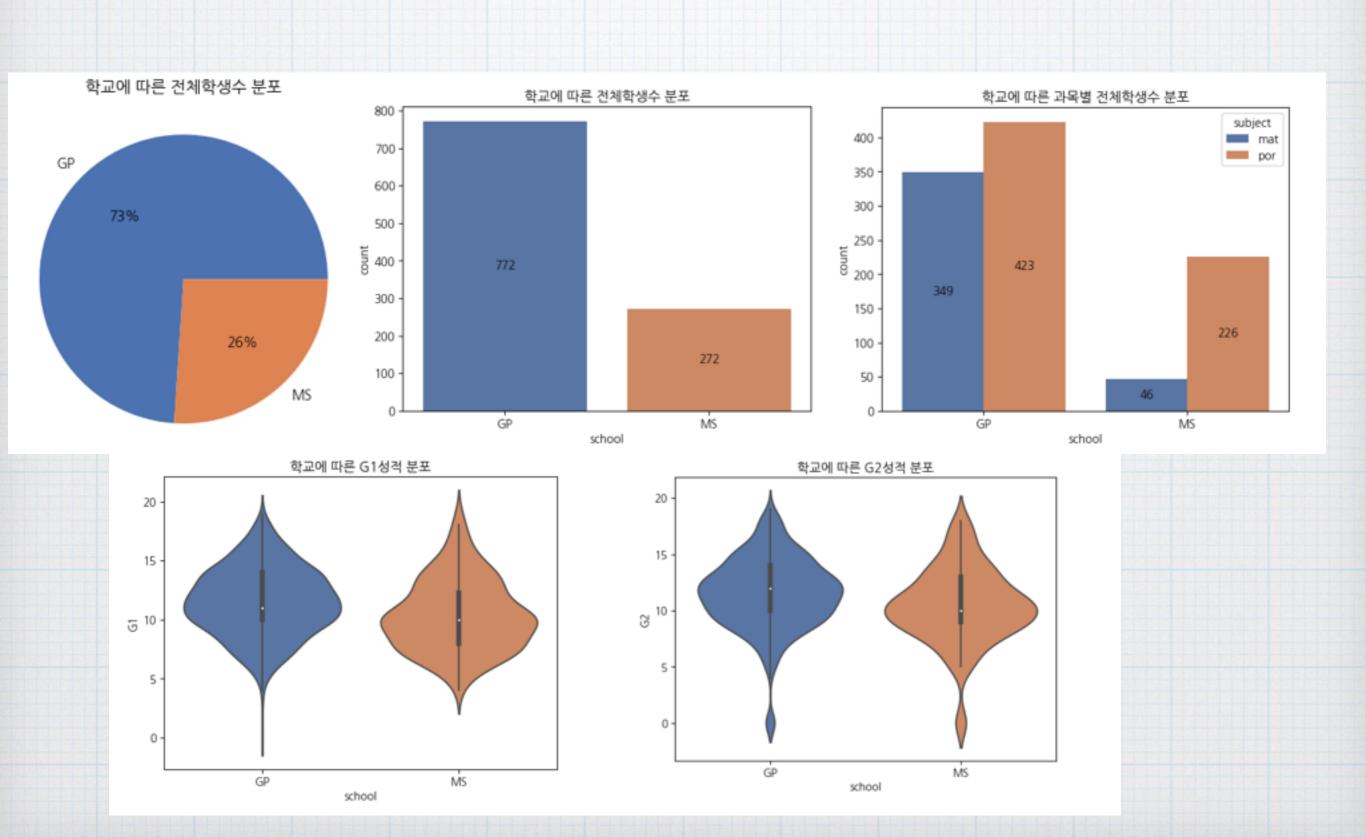
#### 카테고리별 EDA분석(사회적 환경) study time



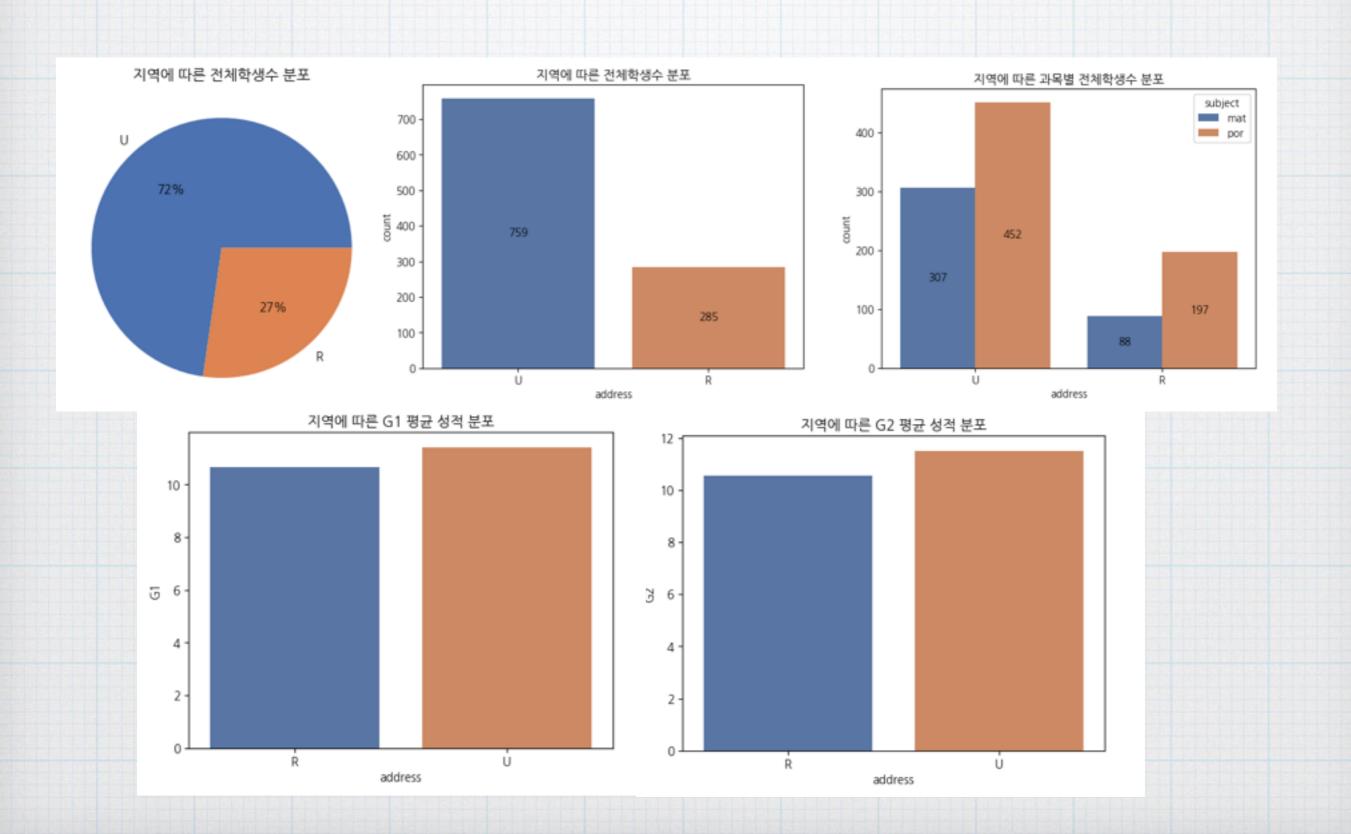
### 카테고리별 EDA분석(사회적 환경) paid



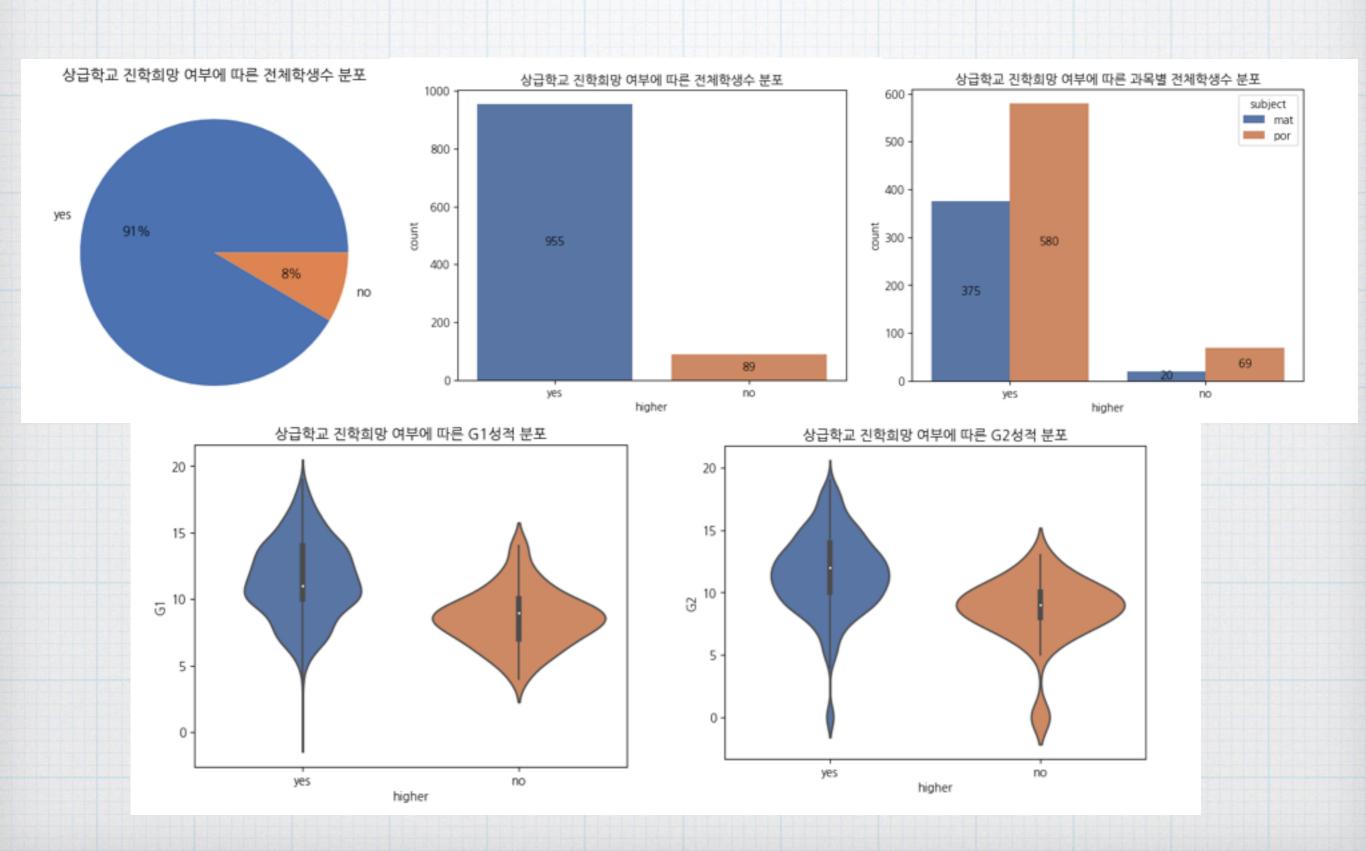
### 카테고리별 EDA분석(사회적 환경) school



#### 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

activities

Palc

goout

sex

freetime

health

Walc

age

romantic

failures

가정적 요인

famsize

farmel

Fedu

Mjob

guardian

internet

Pstatus

internet

Medu

Fjob

famsup

사회적 요인

schoolsup

nursery

address

absences

study time

paid

reason

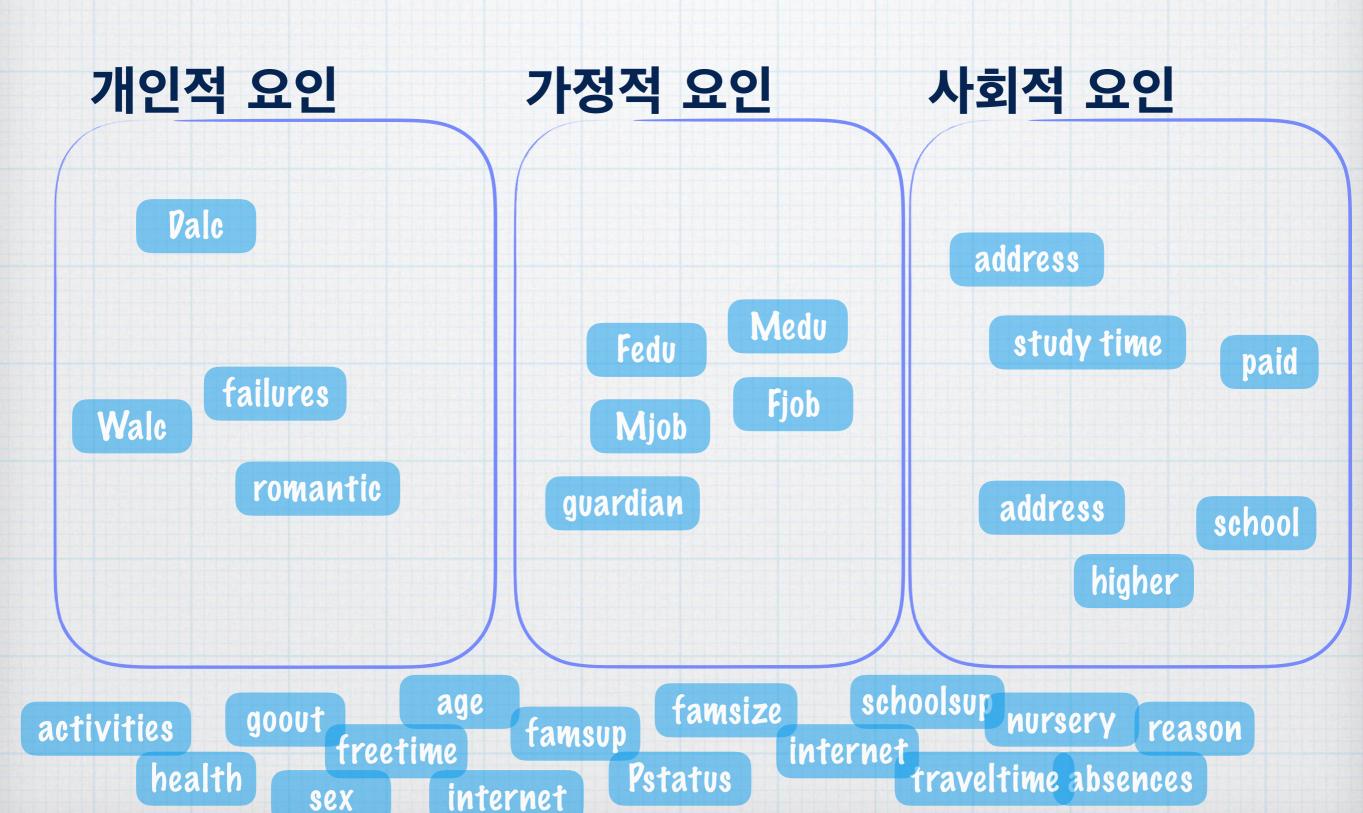
traveltime

address

school

higher

### 데이터셋

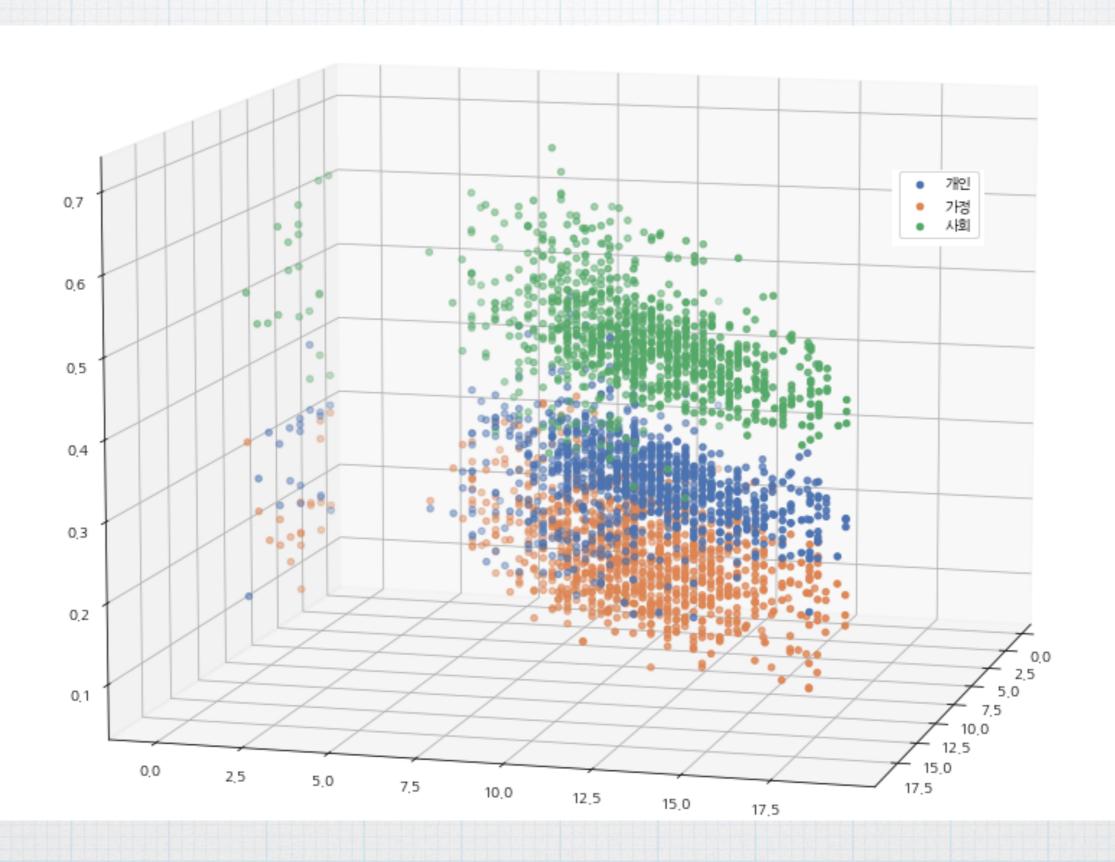


# 상관관계 분석

과 추용된 컬럼 비교

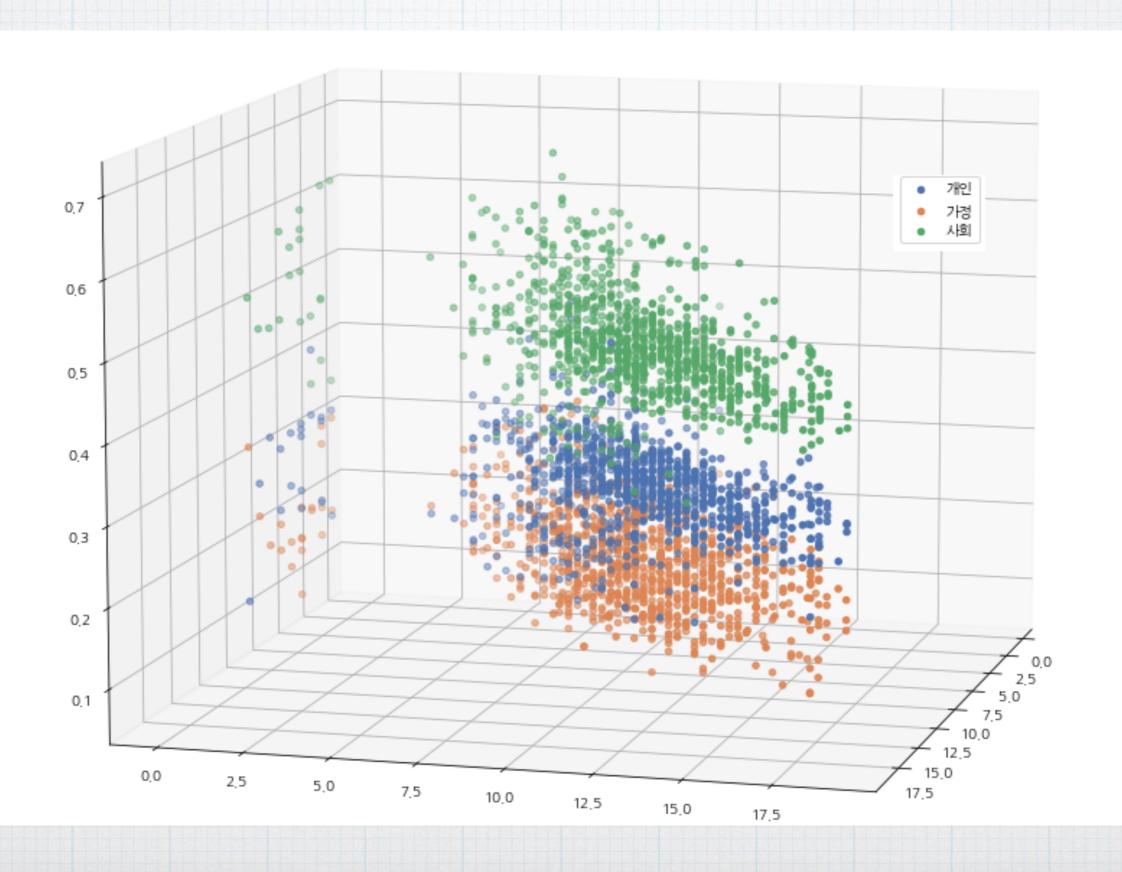
#### Features Correlating with mean (abs) failures 1.00 mean 0.97 0.96 G2 G1 higher G3 0.89 0.39 failures 0.27 higher\_yes higher\_no 0.27 Medu 0.23 Medu - 0.75 studytime 0.2 0.2 Fedu school\_GP 0.16 study time school\_MS -0.16 Dalc -0.15 reason\_reputation -Walc -0.14 Fedu 0.14 Job\_teacher 0.14 - 0.50 0.14 traveltime Miob\_at\_home 0.13 school 0.13 age -0.13 address\_U address\_R -0.13 0.12 internet\_no -Valc internet\_yes -0.12 - 0.25 0.12 schoolsup\_no schoolsup\_yes 0.12 Mjob\_health -0.12 Walc 0.11 reason\_course -0.11 goout -0.11 subject\_por -0.11 subject\_mat -Fjob Mjob\_teacher -0.1 - 0.00 0.094 absences -0.093 guardian\_other -0.081 Mjob romantic\_no -0.081 romantic\_yes -0.078 health -0.077 Mjob\_other address 0.064 activities\_yes -0.064 activities\_no - ⊠0.25 freetime -0.063 0.062 guardian\_father guardian 0.06 famsize\_GT3 -0.06 famsize LE3 -0.052 nursery\_yes -0.052 nursery\_no -Fjob\_other -0.051 romantic 0.05 Mjob\_services -- ⊠0.50 0.045 reason\_other -0.041 famrel -Fjob\_health -0.04 0.038 Fjob\_at\_home -0.027 paid\_yes -0.027 paid\_no paid sex\_F -0.026 - ⊠0.75 sex\_M -0.026 0.018 reason home -0.016 Fjob\_services famsup\_yes -0.0092 0.0092 - famsup\_no guardian\_mother -0.0058 Pstatus\_T -0.0056 0.0056 Pstatus A -⊠1.00 mean

## 추출된 컬럼 으로 클러스터링 분석



결론

#### 클러스터링 결과 고찰



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