



# ESC FINAL PROJECT

김해 시 화재 예측 - EDA까지

ESC SPRING 2021 6조



# 목차

01 조원 소개

02 데이터 및 프로젝트 소개

03 EDA

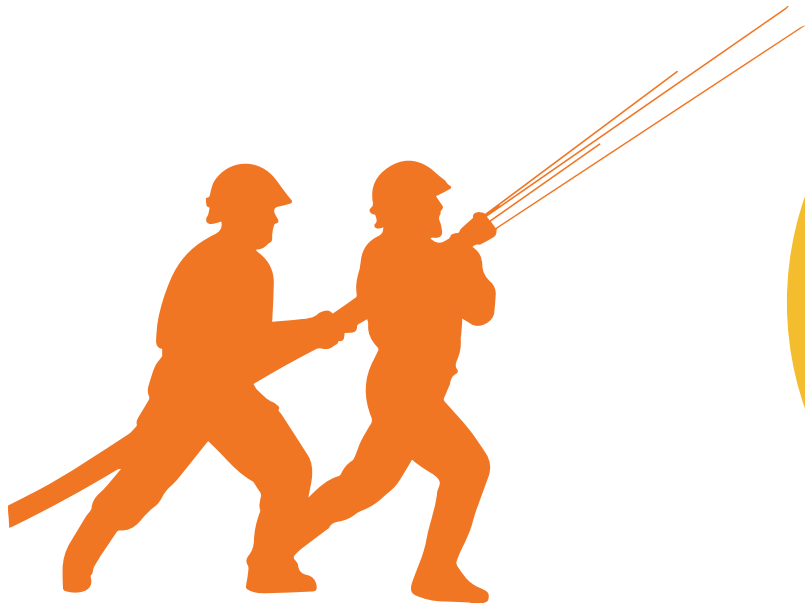
04 추후 계획



A dramatic scene of two firefighters in full protective gear, silhouetted against a massive, intense fire. One firefighter is holding a hose, directing a stream of water towards the flames. The fire is a bright, turbulent mass of orange and yellow, filling the background. The scene is set at night or in a dark environment, emphasizing the brightness of the fire.

# 조원 소개

ESC 소방대



박중창

서상수



김정규

신입 맞아?



서경덕



김내히



A dramatic scene of two firefighters in full protective gear, silhouetted against a massive, intense fire. They are positioned on the left, holding a hose that extends across the frame, spraying a powerful stream of water towards the right. The fire is a towering wall of orange and yellow flames, filling the background and creating a high-contrast, fiery atmosphere. The ground in the foreground is dark and appears to be covered in ash or soot.

# 프로젝트 및 데이터 소개

### 프로젝트 개요

#### 배경

김해지역의 높은 화재 발생량

#### 목적

김해시 내 화재 위험도에 대해 분석 및 예측

#### 해결 과제

김해시가 수집한 경상남도의 소방 및 건물 관련 데이터를 활용하여 건축물의 화재 위험도 분석 및 예측 모델을 제시

## 데이터 개요

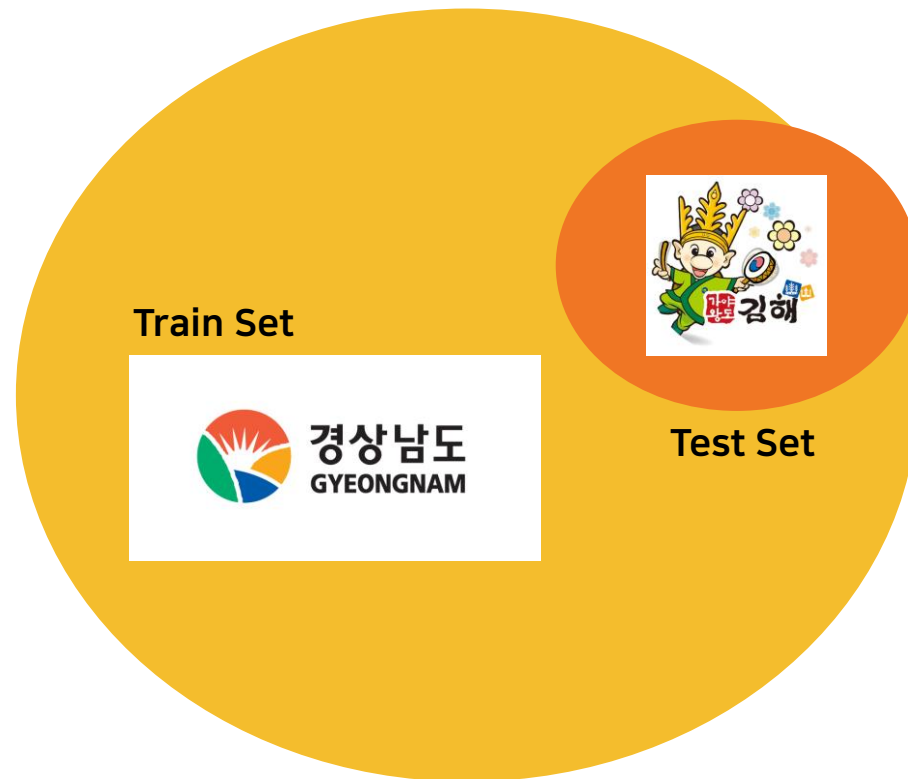
### Train Set

경상남도 지역의 **날짜, 시간, 건물화재여부, 건물정보** 등에 대한 정보

불 안 났으면 데이터를 어떻게 뽑을까? 무작위 샘플링!

### Test Set

Train Set과 동일하나 김해시 지역으로 한정 및 예측해야 하는 화재발생여부 변수를 제거







EDA



## NA imputation – Data Overview

### Dataset statistics

Number of variables	180
Number of observations	59199
Missing cells	4750511
Missing cells (%)	44.6%
Total size in memory	81.3 MiB
Average record size in memory	1.4 KiB

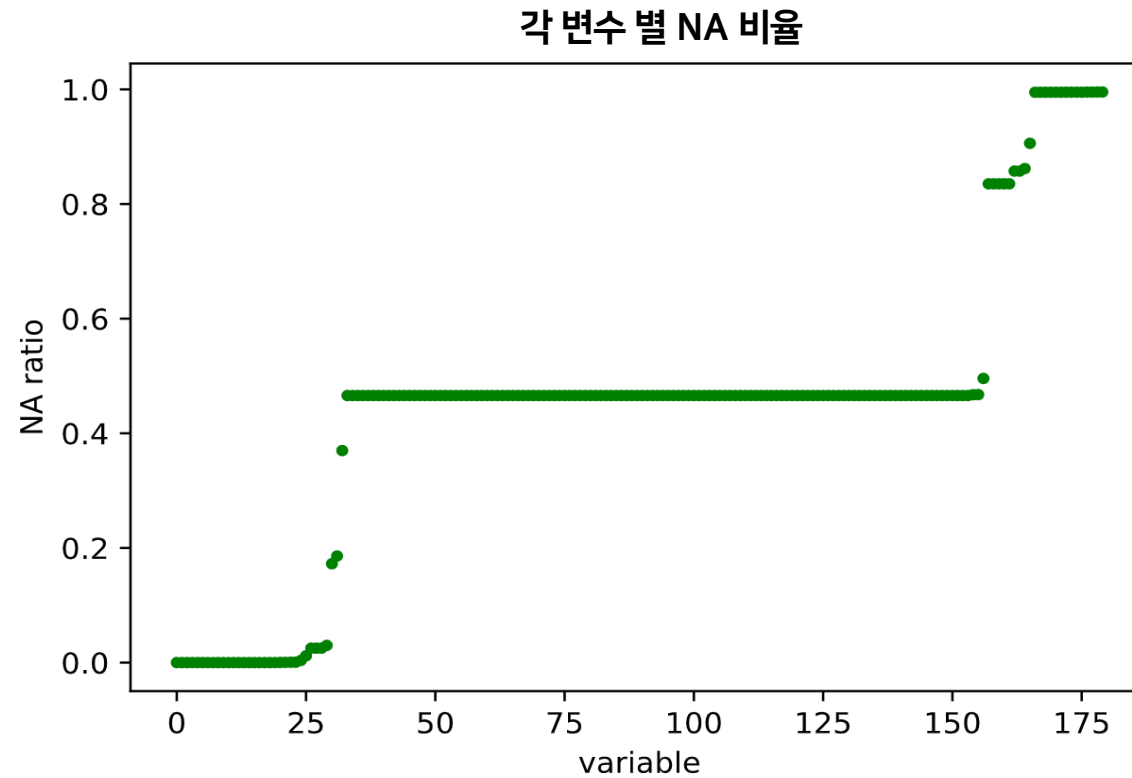
### Variable types

Categorical	21
Numeric	159

강수량

```
print(df.isnull().sum().head(30))
shape: (59199, 180)
dt_of_fr      0
fr_yn         0
bldng_us     27677
bldng_archtctr 27665
bldng_cnt     0
bldng_ar      0
ttl_ar        0
lnd_ar        0
dt_of_athrzttn 27581
ttl_grnd_flr  10210
ttl_dwn_flr   11005
bldng_us_clsfcctn 29343
tmprtr        13
prcpttn      53629
wnd_spd       33
wnd_drctn     221
hmdt          22
gas_engry_us_201401 27581
ele_engry_us_201401 27581
gas_engry_us_201402 27581
ele_engry_us_201402 27581
gas_engry_us_201403 27581
ele_engry_us_201403 27581
```

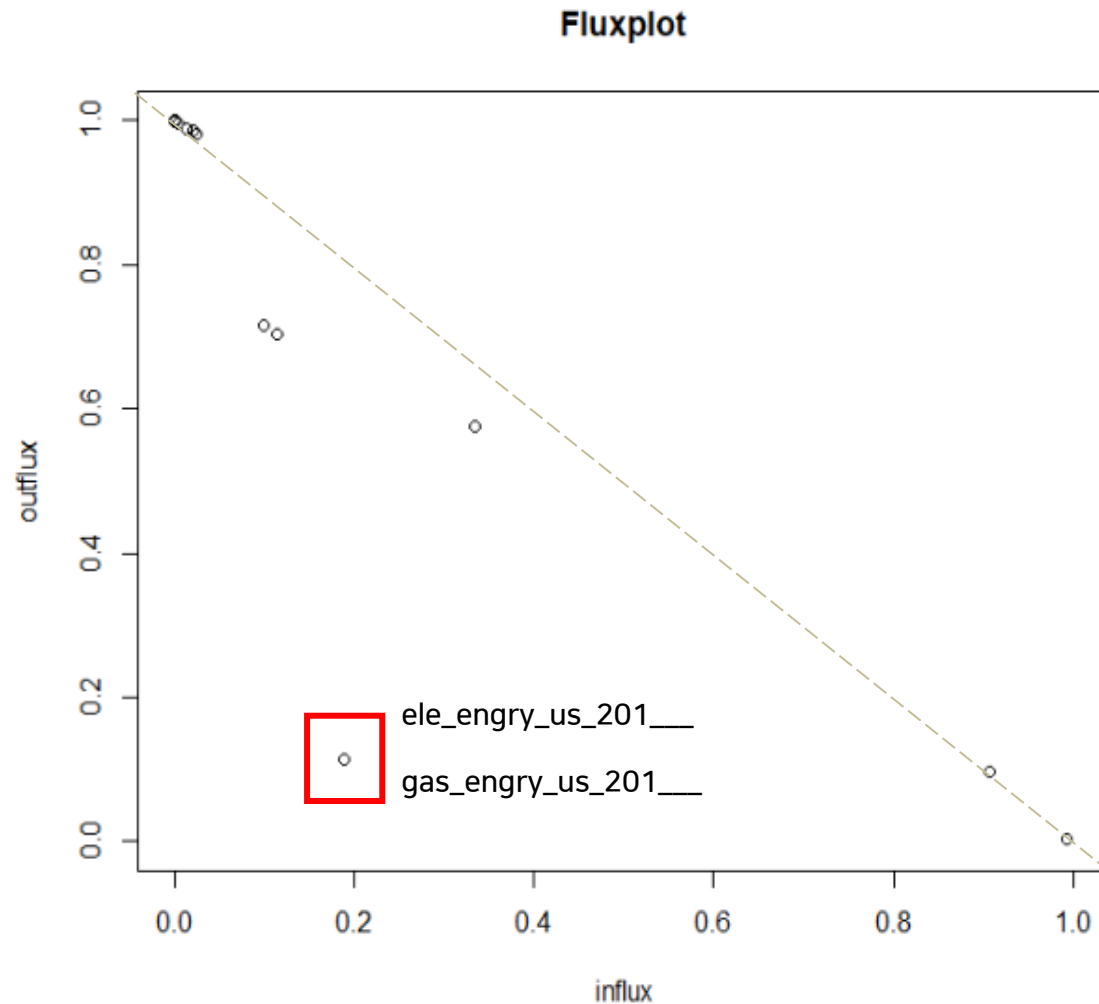
## NA imputation – NA ratio



180개 변수 중 161개 변수에 결측치 존재



## NA imputation - Fluxplot



월별 가스, 전기 사용량 관련 변수의 influx와 outflux가 0.2보다 낮음



NA Imputation에 영향이 적은 변수 확인

## NA imputation - Lw\_131\_\_\_\_

Lw\_131\_\_\_\_ - (결측률 99.49%~99.55%)

의미 있는 NA인지 판단 필요

해당 변수들을 두가지 대체방법을 통해 로지스틱 회귀분석 실시

**Imputation 1**  
NA는 0, 관측치일 경우 1



**Imputation 2**  
MICE imputation

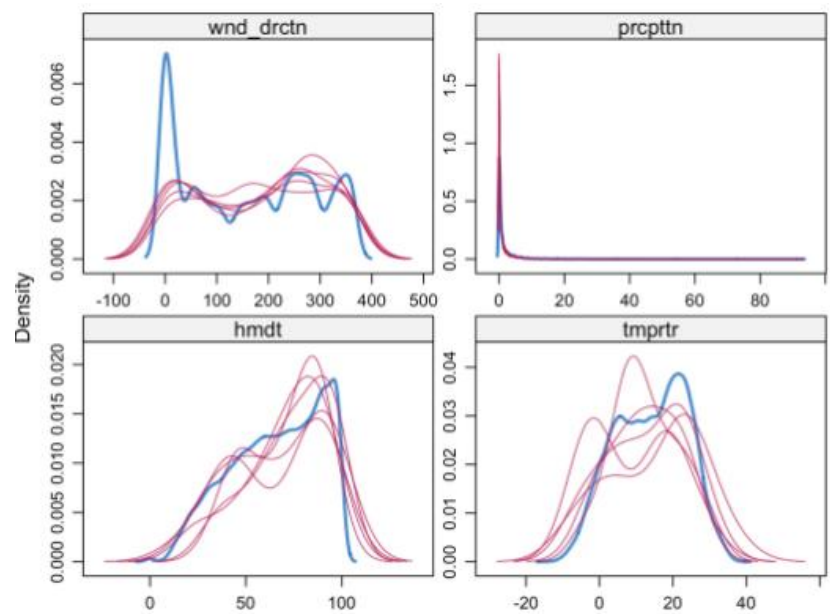
Imputation 1이 F1 Score가 높게 나타나, NA 자체가 의미가 있다고 해석



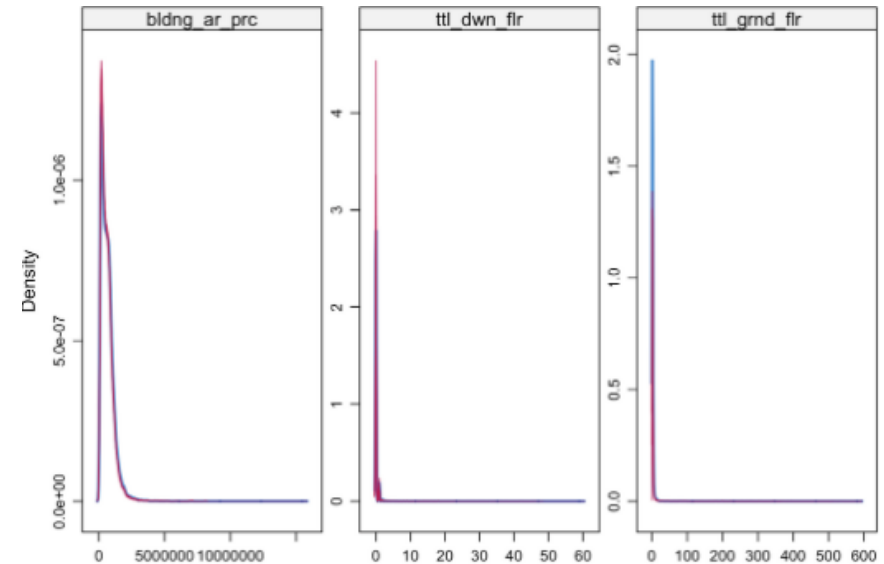
## NA imputation – multivariate imputation

Prcpptn, wnd\_drctn, wnd\_spd, hdmt, tmprtr - (결측률 0.02%~90.59%)

Bldng\_ar\_prc, ttl\_dwn\_fir, ttl\_gmd\_flr - (결측률 17.25%~36.99%)



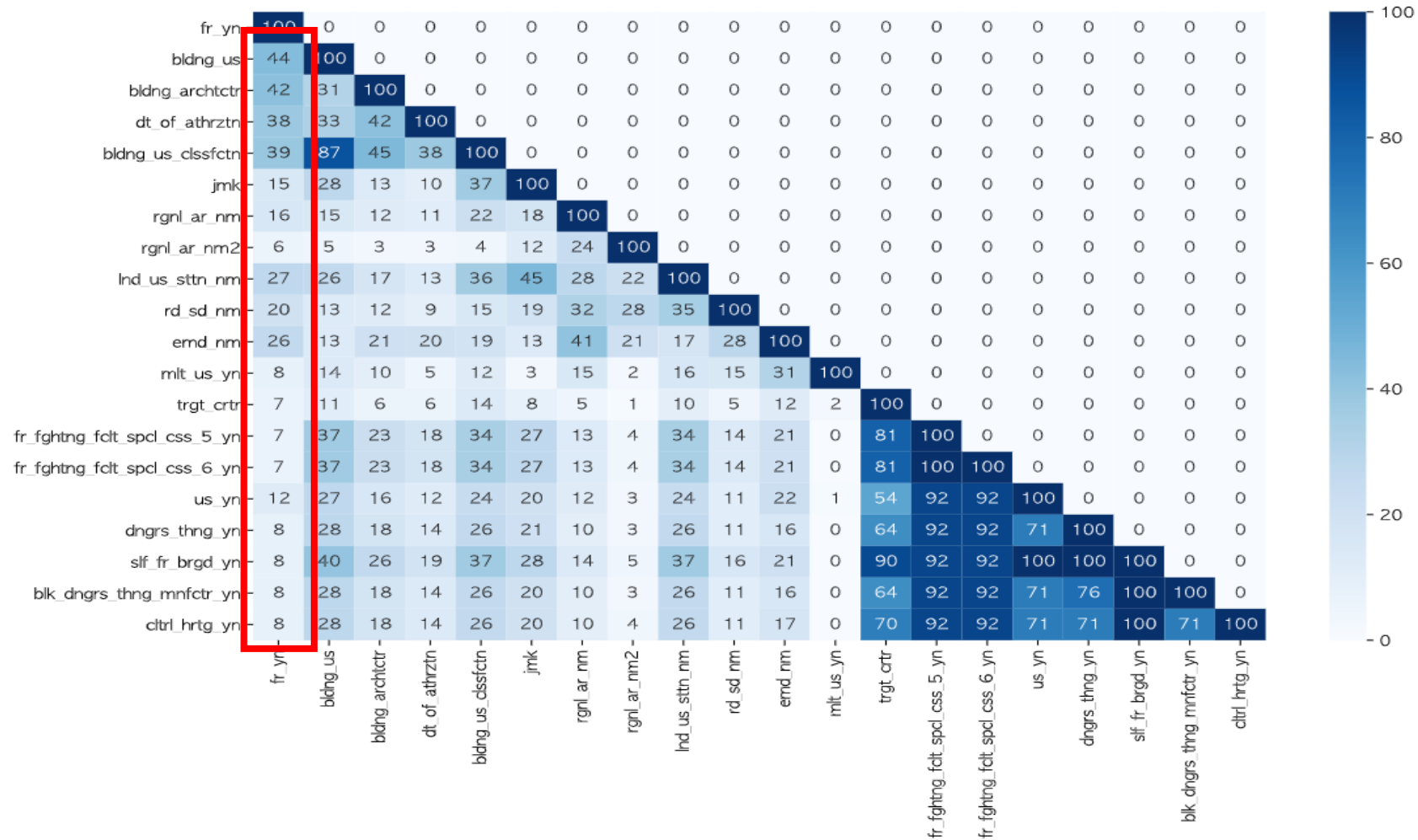
날씨 관련 변수



건물 가격 관련 변수

## NA imputation – NA ratio

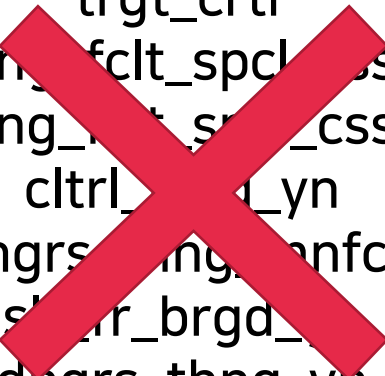
Crammer's V : 범주형 변수간의 상관관계를 나타냄





## NA imputation - 변수 삭제

화재발생여부 변수와의 상관관계 이용

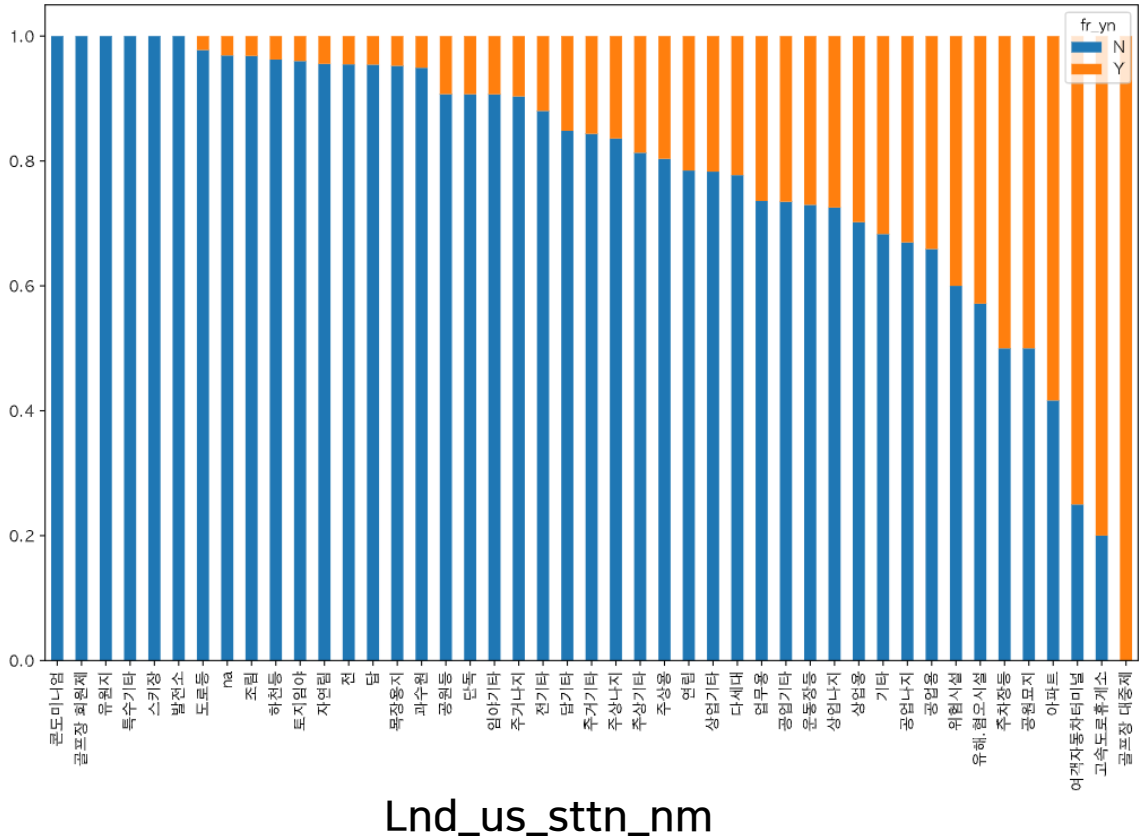
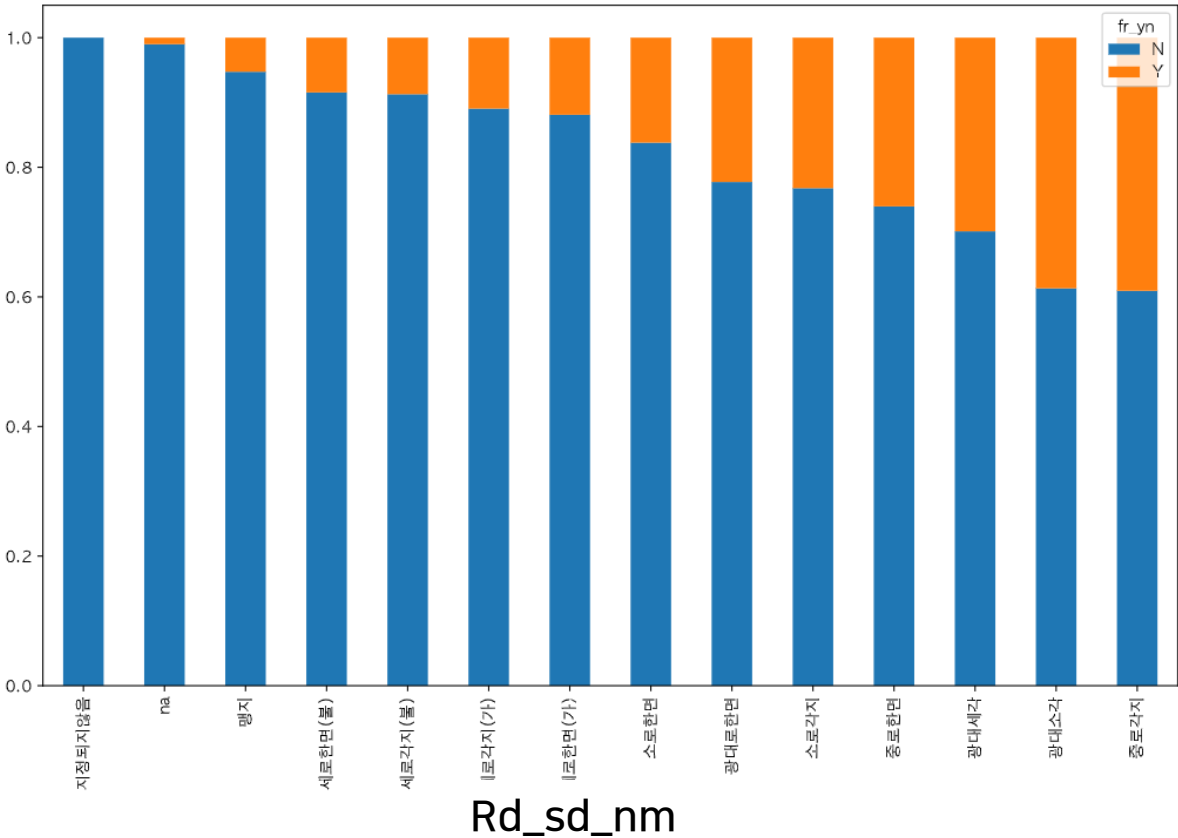


rgnl\_ar\_nm2  
trgt\_crtr  
fr\_fghtng\_fclt\_spcl\_css\_6\_yn  
fr\_fghtng\_fclt\_spcl\_css\_5\_yn  
ctrl\_yn  
blk\_dngrs\_thng\_infctr\_yn  
shpr\_brgd\_yn  
dngrs\_thng\_yn  
us\_yn

8개 변수 삭제!

# NA imputation

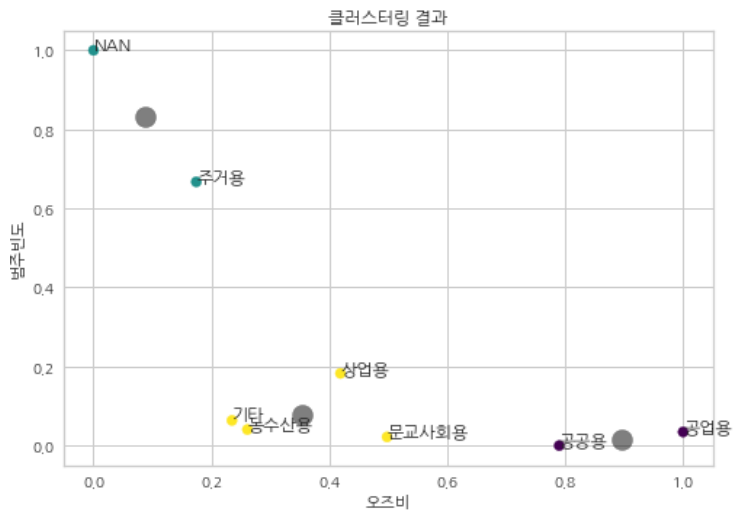
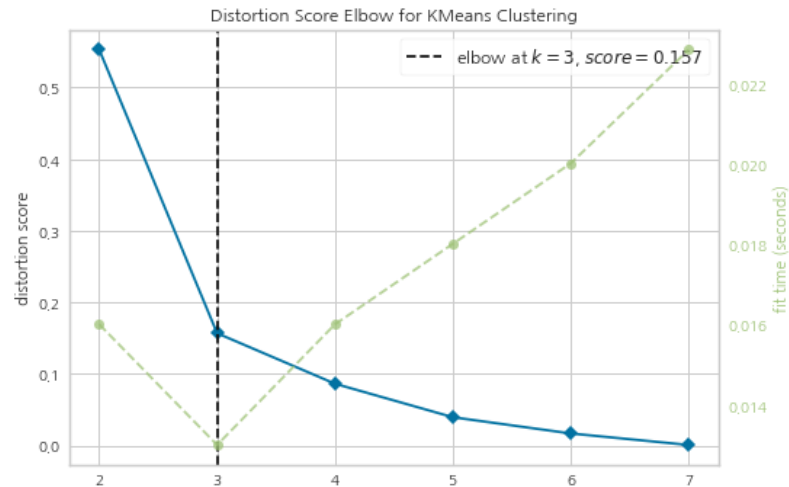
각 범주 별 화재발생비율 확인



각각 4개 카테고리로 묶음



## NA imputation - clustering



Collapse Category level

다음의 변수에 대하여 범주를 묶는 시도를 진행함

bldng\_us\_clssfctn : 건물 사용 용도

bldng\_us : 건물 유형

bldng\_archtctr : 건물 구조

임베딩 벡터 = (반응변수에 대한 오즈비, 수준이 관측된 빈도수)

{0: ['공공용', '공업용'], 1: ['NAN', '주거용'], 2: ['기타', '농수산용', '문교사회용', '상업용']}

## NA imputation - 에너지, 가스 시계열 데이터

0과 na를 다른 범주로 가정(범주: 0: 0, na: 1), 0과 na 이외는 범주 2로 묶음

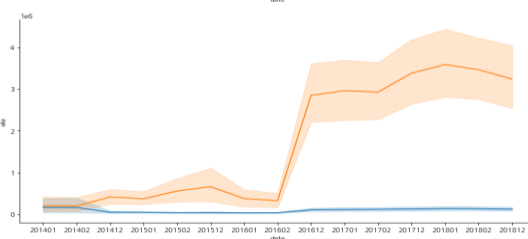
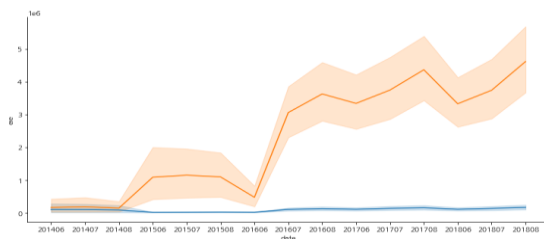
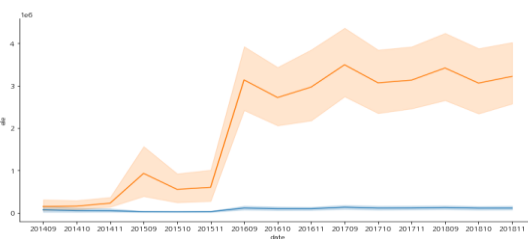
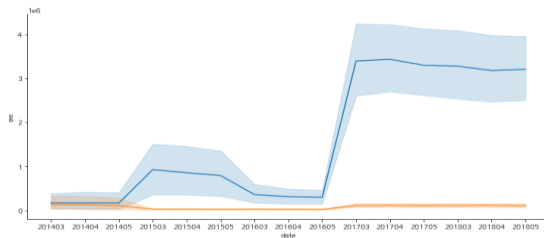
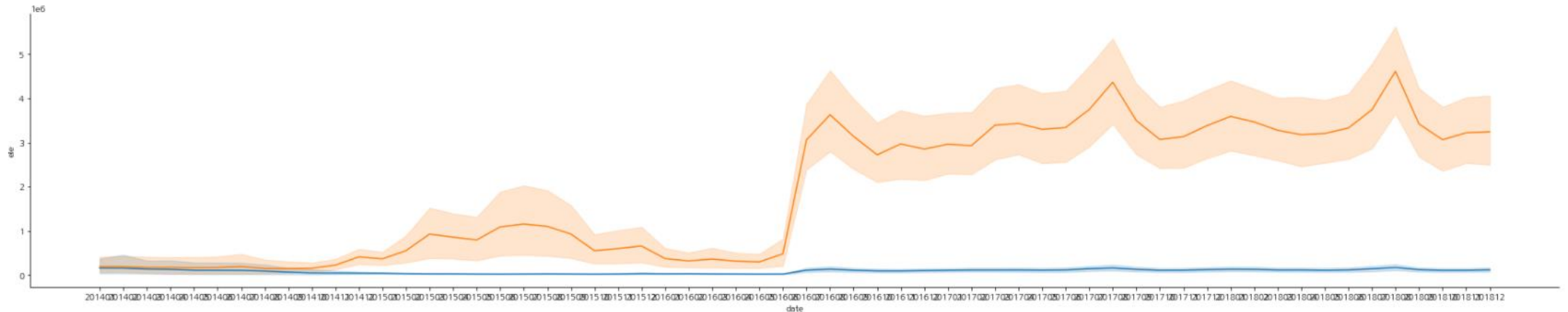
0과 na를 같은 범주로 가정(범주: 0, na: 0), 0과 na 이외는 범주 2로 묶음

Value	Count	Frequency (%)
0	30753	51.9%
70455168	6	< 0.1%
1028221320	6	< 0.1%
1078168968	6	< 0.1%
21826944	6	< 0.1%
1056662200	5	< 0.1%
154440600	4	< 0.1%
96390360	4	< 0.1%
364526032	4	< 0.1%
7996920	4	< 0.1%
Other values (704)	820	1.4%
(Missing)	27581	46.6%

engry	0.0	1.0	2.0
yn			
N	23520	27581	441
Y	7273	0	384

## NA imputation - 에너지, 가스 시계열 데이터

### 화재발생여부별 가스사용량 time series plot



gas, ele data 120개의 변수  
계절 단위로 파생 변수 생성 -> 4개!



## 진행상황 요약





A dramatic scene of two firefighters in full protective gear, silhouetted against a massive, intense fire. The firefighters are positioned on the left, with one holding a hose that extends towards the right. The fire is a large, bright orange and yellow wall of flames that fills the background. The text '추후 계획' is overlaid in white on the right side of the image.

추후 계획

### 추후 계획

Outlier deletion

Scaling, Skewness 조정

추가 파생 변수 생성

