



DACON

# 구내식당 식수 인원 예측 AI 경진대회



으쌔으쌔 4조 (1석 4조)



# 1석 4조 팀원 소개



김 성 우



김 조 은



김 준 호



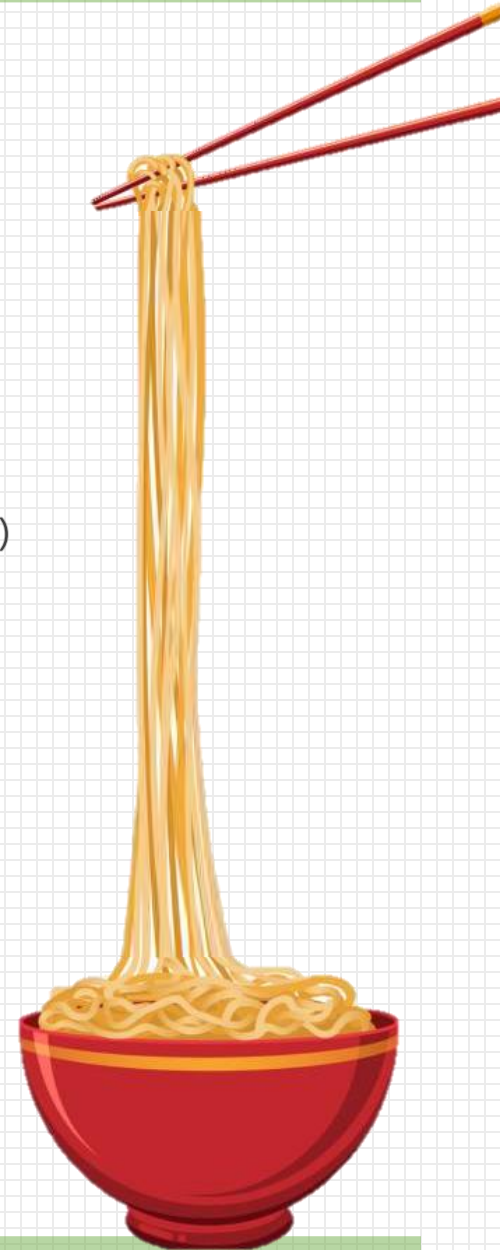
김 현 아 (PPT 제작 및 발표)



김 희 목



정 지 원



# 목 차

1 ; 프로젝트 주제 소개  
; DACON 구내식당 식수 인원 예측 AI 경진대회

2 ; 가설 설정  
; 참고 논문 / 피처 예측

3 ; 선정된 최적의 피처  
; 선택된 소수의 피처

4 ; 선정되지 못한 피처들  
; 국가별 메뉴 범주화 / 반찬 개수 / 날씨 관련 피처

5 ; 결론 및 회고  
; 머신러닝 시행착오 / 팀원별 자기평가 / 개선점





# 1

## 프로젝트 주제 소개

DACON 구내식당 식수 인원 예측 AI 경진대회



으쌔으쌔



4팀



1석 4조

# 데이터 소개

## train 데이터

- 일자
- 요일
- 본사정원수
- 본사휴가자수
- 본사출장자수
- 시간외근무명령서승인건수
- 현본사소속재택근무자수
- 조식메뉴
- 중식메뉴
- 석식메뉴
- 중식계
- 석식계

## test 데이터셋

- 일자
- 요일
- 본사정원수
- 본사휴가자수
- 본사출장자수
- 시간외근무명령서승인건수
- 현본사소속재택근무자수
- 조식메뉴
- 중식메뉴
- 석식메뉴

## sample\_submission 데이터

- 일자
- 중식계
- 석식계

	일자	요일	본사 정원 수	본사 휴가 자수	본사 출장 자수	본사시간 외근무명 령서승인 건수	현본사 소속재 택근무 자수	조식메뉴	중식메뉴	석식메뉴	중식 계	석식 계
0	2016-02-01	월	2601	50	150	238	0	모닝롤/편빵 우유/두유/주스 계란후라이 호두죽/알밥 (쌀:국내산) 된장찌개 쥐...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 오징어찌개 쇠불고기 (쇠고기:호주산) 계란찜 ...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 육개장 자반고등어구이 두부조림 건파래무침 ...	1039	331



## 아이디어 회의

인기 있는 메뉴?  
 메뉴를 다양하게 범주화해보기  
 새로 내보이는 메뉴명 앞에는 (New)가 붙는다  
 메뉴별 중요소/선호도에 따라 다른 가중치?  
 일별 식사인원수 대비 중식/석식계 비율 계산하기  
 코로나 전/후 식수 비교하기  
 코로나 거리두기가 식수에 영향을 미칠까?  
 매주 수요일 야근 없는 자기계발의 날이 있다  
 월별/요일별 식수 비교해보기  
 1,2,3월과 9,10월 중식계가 높다  
 석식은 2,3월과 10월이 높다  
 2,3월에 식수가 많은 이유가 날이 추워서인지?  
 회사가 바쁜 기간이 식수와 관련있는지?  
 요일에 따른 피처의 특징 탐색해보기  
 요일은 숫자로 변환하여 사용하기  
 메뉴도 원핫인코딩하여 사용하기  
 ...

## ‘1석 4조’의 프로젝트 주제

# 정확한 식수 예측



→ 이에 대한 기준은 DAICON 점수로 결정함.





# 2

## 가설설정

참고 논문 / 피쳐 예측



으쌔으쌔



4팀



1석 4조

# 피처 선택

## 참고 논문)

Original Article

대한영양사협회 학술지  
J Korean Diet Assoc 25(1):44-58, 2019  
<https://doi.org/10.14373/JKDA.2019.25.1.44>

### 기계학습방법을 활용한 대형 집단급식소의 식수 예측: S시청 구내직원식당의 실데이터를 기반으로

전종식 · 박은주<sup>1</sup> · 권오병<sup>†</sup>  
경희대학교 경영학과 · <sup>1</sup>경남대학교 식품영양학과

Predicting the Number of People for Meals of an Institutional  
Foodservice by Applying Machine Learning Methods: S City Hall Case

Jongshik Jeon · Eunju Park<sup>1</sup> · Ohbyung Kwon<sup>†</sup>  
School of Management, Kyung Hee University, Seoul 02447, Korea  
<sup>1</sup>Dept. of Food & Nutrition, Kyungnam University, Changwon 51767, Korea

#### 1. 전체적인 구조

식수 예측을 위해서 먼저 식사 가능 인원 변수, 시계열 인원 변수, 메뉴 특성 변수, 요일 변수, 전후 휴일 여부 변수, 날씨 변수, 연휴 연말 변수, 이벤트 변수 등 8개 카테고리 63개의 변수를 선정하였다.

#### 2. 데이터

본 연구는 식수 예측 모델링을 위해 S시청 구내 식당에서 제공한 메뉴, 식사 가능 인원 및 날씨 등  
에 대한 데이터를 대상으로 연구를 진행하였다.

식수 예측 시 고려하는 요인으로 요일, 제공 메뉴 선호도, 날씨, 전주 동일 요일 식수, 운영 끼니, 고객사 행사, 복수 메뉴의 선호도, 전월 식수 자료, 계절, 동일 메뉴 과거 식수, 이벤트, 전년 해당 월(주) 식수, 기온, 주변 식당 이용 가능성, 동일 메뉴 과거 잔반량 등이 식수 예측 향상을 위한 요인으로 파악되었다.

#### 3. 구성 변수

식수 예측에 미치는 변수 파악을 위해 크게 내부적 요인과 외부적 요인의 2가지 요인으로 구분

하였으며 내부적 요인으로 고려할 변수로는 식사 가능 인원(출근 인원, 출장자, 휴가자 등),

시계열 인원 변수(전주 식사 인원, 요일별 평균 식사 인원, 월 평균 식사 인원, 일일 식사 가능 인원), 메뉴 특성(일일 메뉴) 등 3가지 범주를 고려하여 분류를 하였다. 외부적 요인으로서는 요일 및 계절, 전일 · 익일의 휴일 여부, 날씨, 이벤트 여부 및 연휴 · 연말 등 5가지

범주로 분류를 하였다.

다음으로 식수 예측에 영향을 미치는 변수를 파악하기 위해 현장 조사를 수행하였다. S시청사 내 집단급식소의 경우 영양사 및 조리사는 경험에 의해 식수를 예상하고 있었는데, 요일별 식수를 주요결정요인으로 인식하고 있었다. 식수는 요일 별로 일정한 패턴이 있는 것으로 알려져 있다.



## 피처별 가설설정



### 날짜 관련

- 공휴일은 식수 인원에 영향을 줄 것이다.
- 요일은 식수 인원에 영향을 줄 것이다.



### 인원수 관련

- 출근 인원은 식수 인원에 영향을 줄 것이다.
- 야근 인원은 석식 인원에 영향을 줄 것이다.



### 메뉴 관련

- 선호도가 높은 음식이 나오는 날 식수가 높을 것이다.
- 반찬 개수가 식수 인원에 영향을 줄 것이다.



### 날씨 관련

- 기온의 변화는 식수 인원에 영향을 줄 것이다.
- 강수량은 식수 인원에 영향을 줄 것이다.

# 팀원별 선택 피처

공통

요일, 식사가능인원, 식사비율

개인



김 성 우

- 식단 메뉴
- 재료별 범주화

닭, 돼지, 소, 떡,  
오리, 해물 등



김 조 은

- 식단 메뉴
- 종류별 범주화

인기 있는 메뉴,  
메뉴별 선호도



김 준 호

- 반찬 개수
- 인원 수 스케일링

Min-Max, log1p



김 현 아

- 계절
- 날씨별 범주화

불쾌지수, 폭염,  
한파, 강수량



김 회 목

- 식단 메뉴
- 조리방법별 범주화

볶음, 찜, 구이 등



정 지 원

- 공휴일 자료
- 연차쓰기 좋은 날





# 3

## 선정된 최적의 피처

선택된 소수의 피처



으쌔으쌔



4팀



1석 4조

## 최종 선택 피처

[ train\_lunch 데이터셋 ]

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날
6	7	8	9	10	11
식사가능인원	야근수	휴가자수	재택근무자수	출장자수	반찬1
12	13	14			
국	메인메뉴범주 _볶음	중식계			

[ train\_dinner 데이터셋 ]

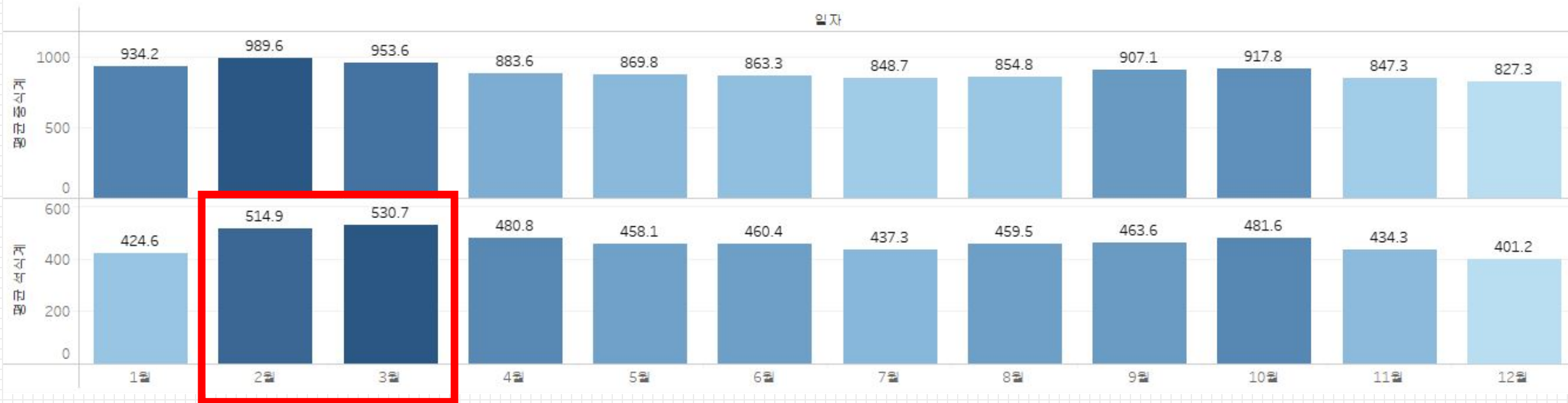
0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날
6	7	8	9	10	11
식사가능인원	야근수	휴가자수	재택근무자수	출장자수	반찬1
12	13	14			
국	메인메뉴범주 _볶음	석식계			



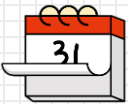
## 날짜 관련 피쳐

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날

월 별

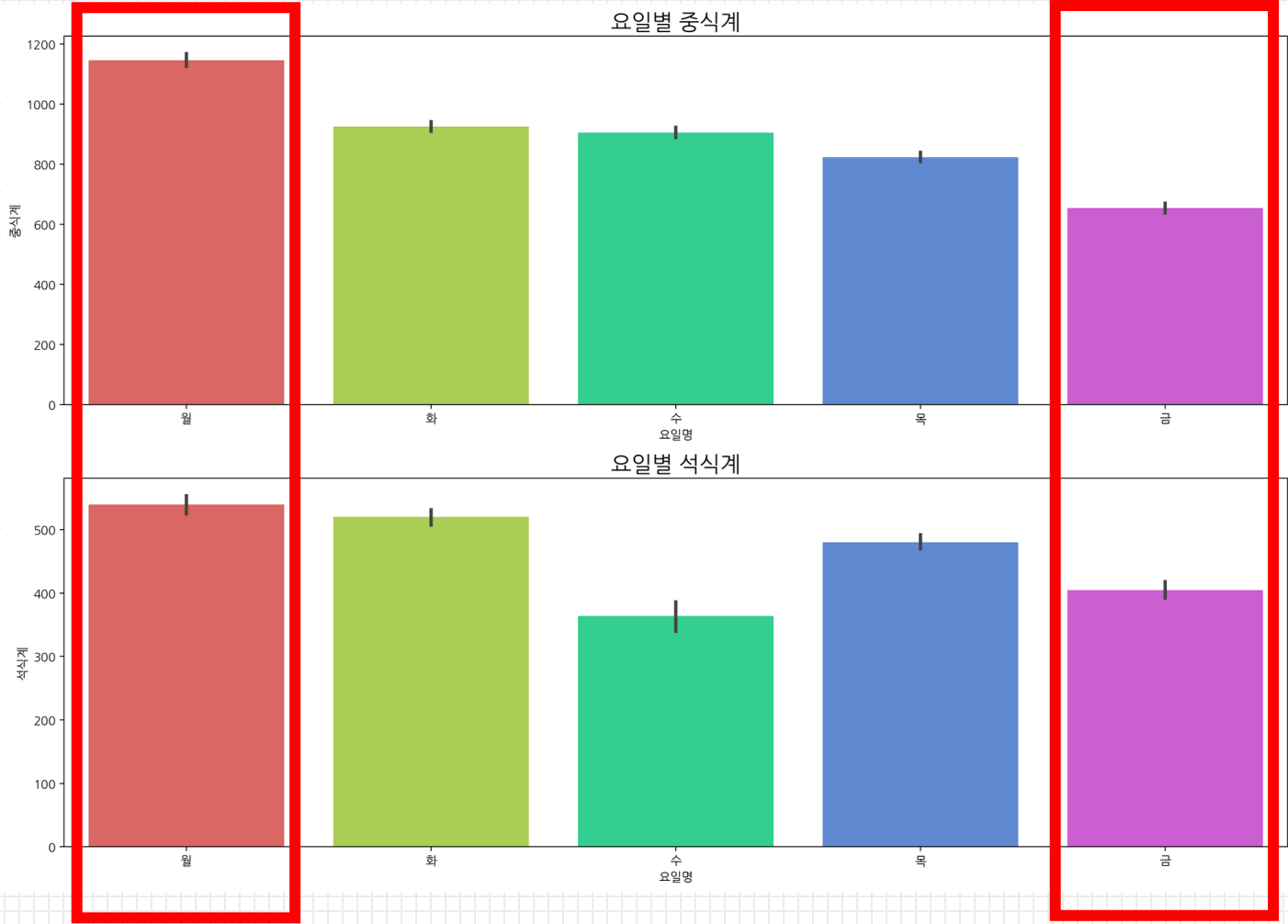






날짜 관련 피쳐

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날

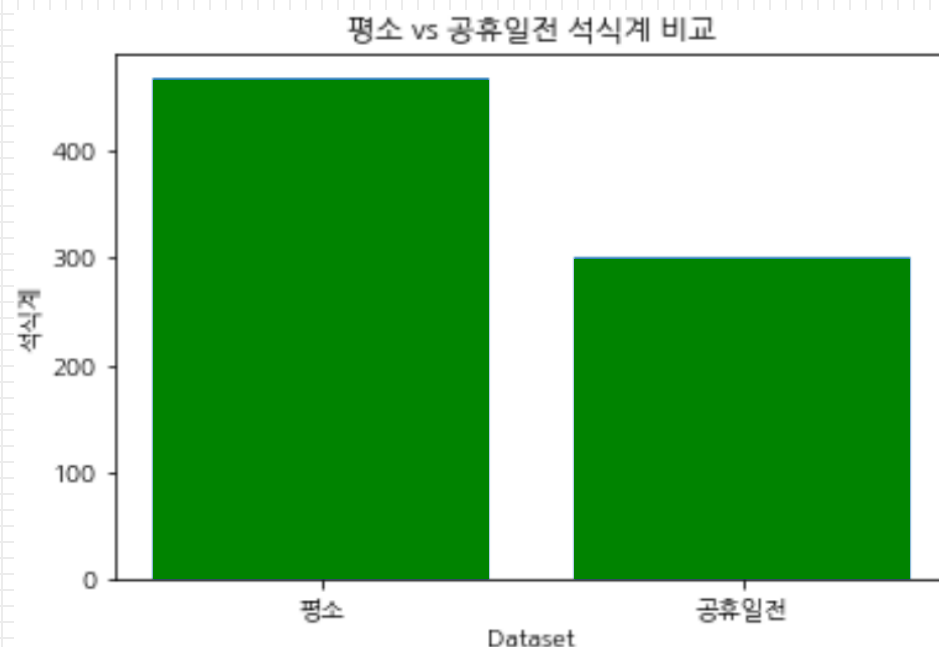
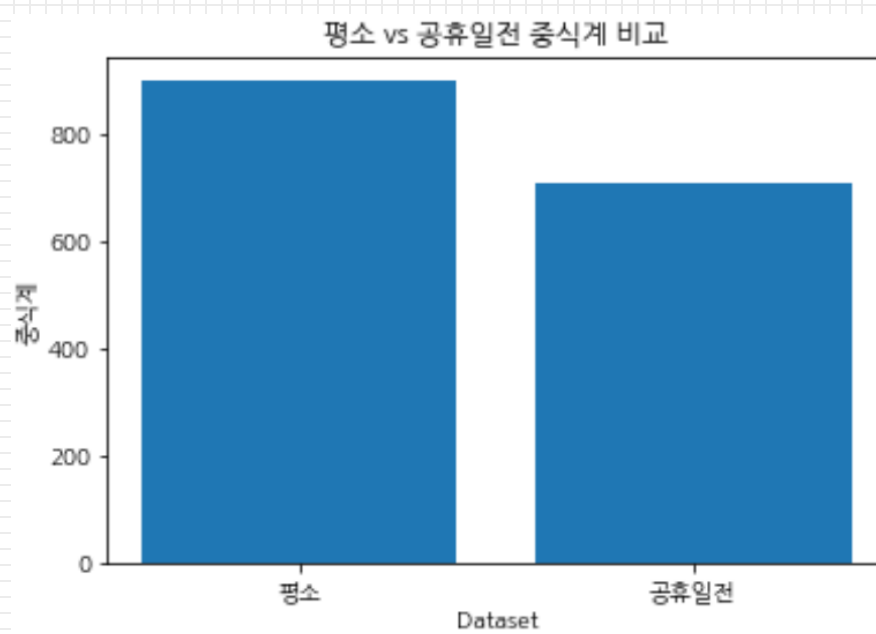




## 날짜 관련 피쳐

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날

## 평소 vs 공휴일 전 '중식계', '석식계' 비교



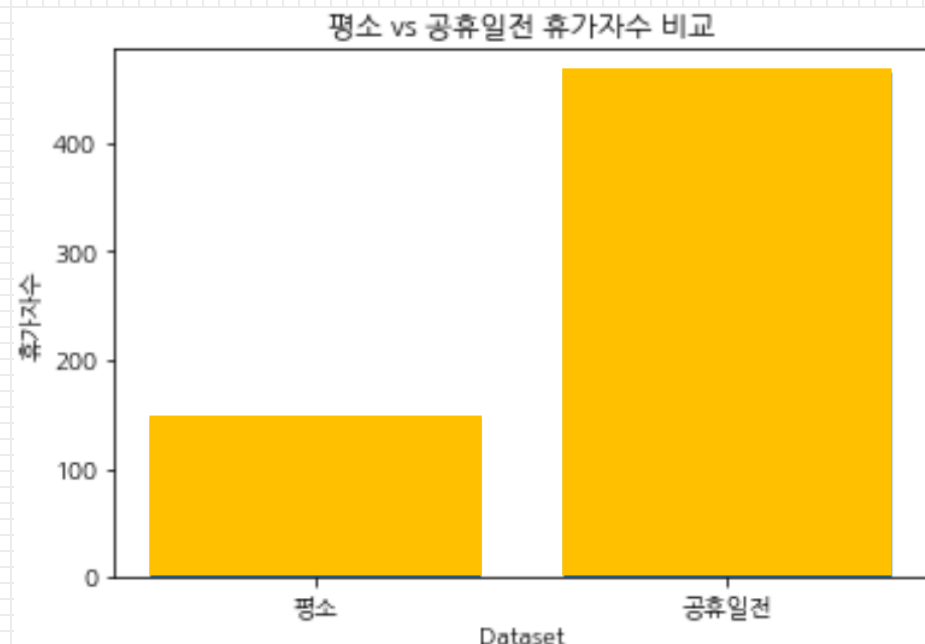
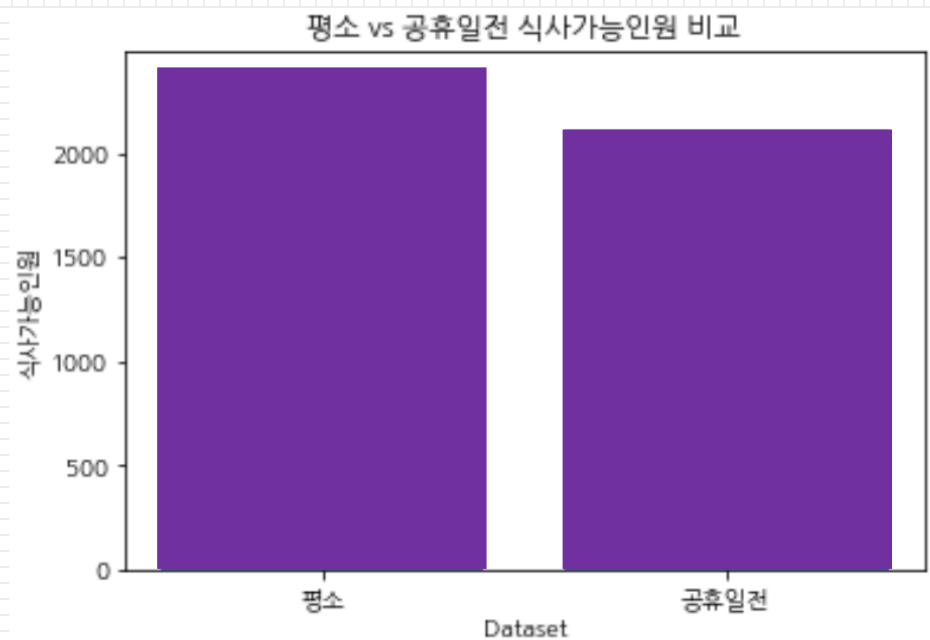
- 공휴일 전에 평소보다 중식계 수요가 200명 가량 적은 것을 확인할 수 있다. 석식계는 약 150명이 줄었다.
- 공휴일 전에 회사 내에 인원이 더 적어서 그런 것은 아닐지 알아보았다.



## 날짜 관련 피쳐

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날

### 평소 vs 공휴일 전 '식사가능인원', '휴가자수' 비교



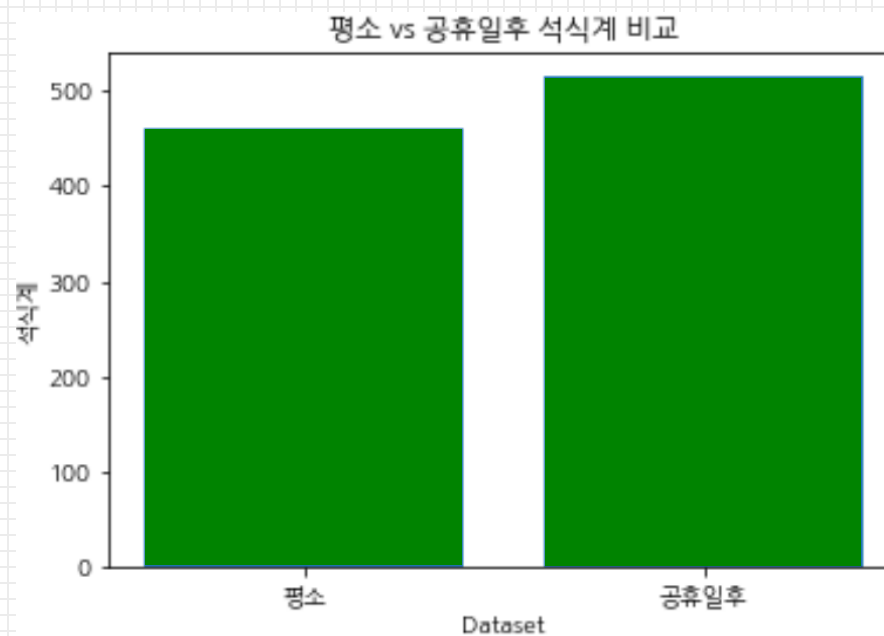
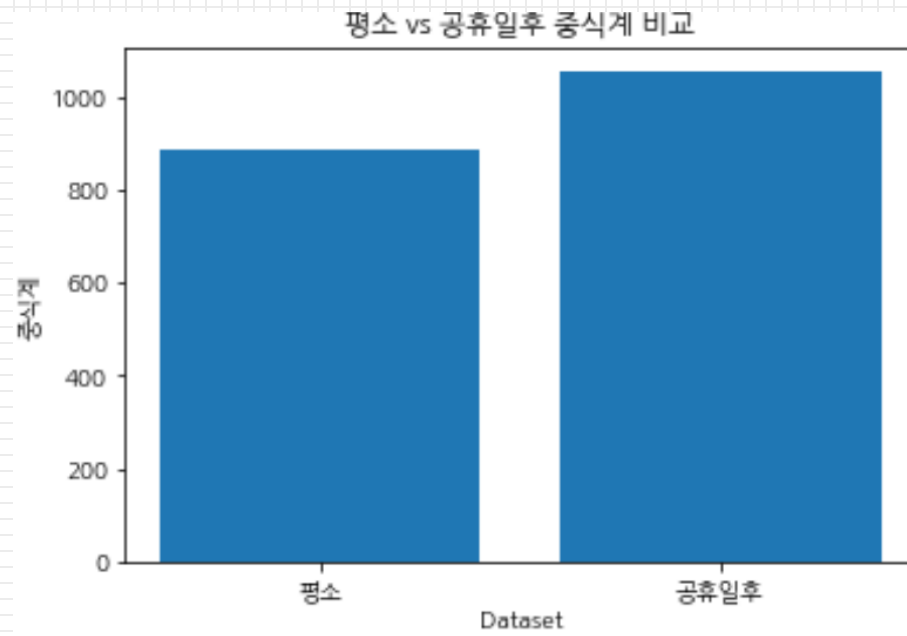
- 회사 내에 남아있는 인원이 평소보다 공휴일 전에 260명 가량 더 적었다.
- 공휴일 전에 식사가능인원이 적은 이유로 휴가자수를 예측해보았다.



## 날짜 관련 피쳐

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날

## 평소 vs 공휴일 후 '중식계', '석식계' 비교



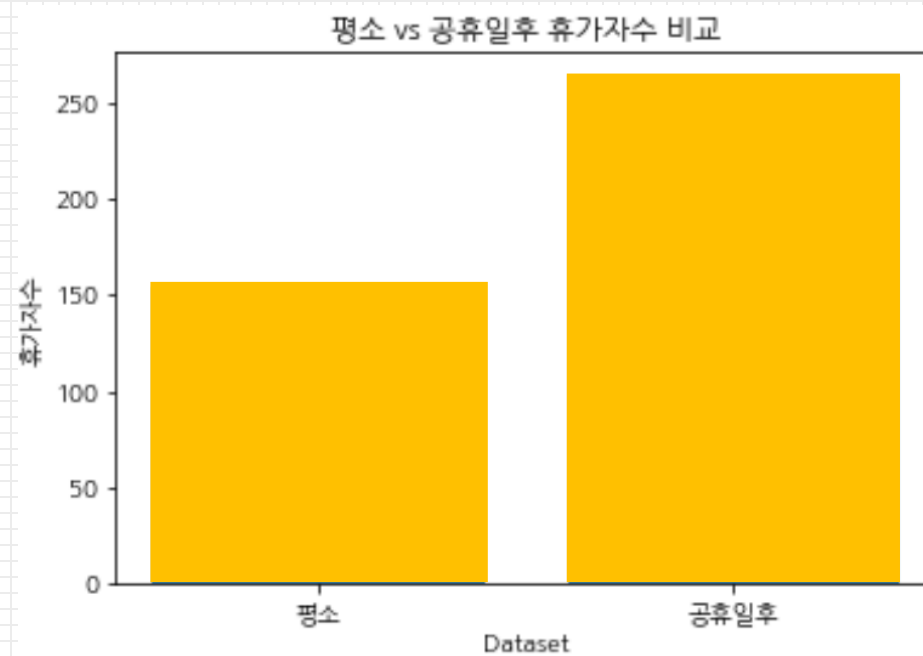
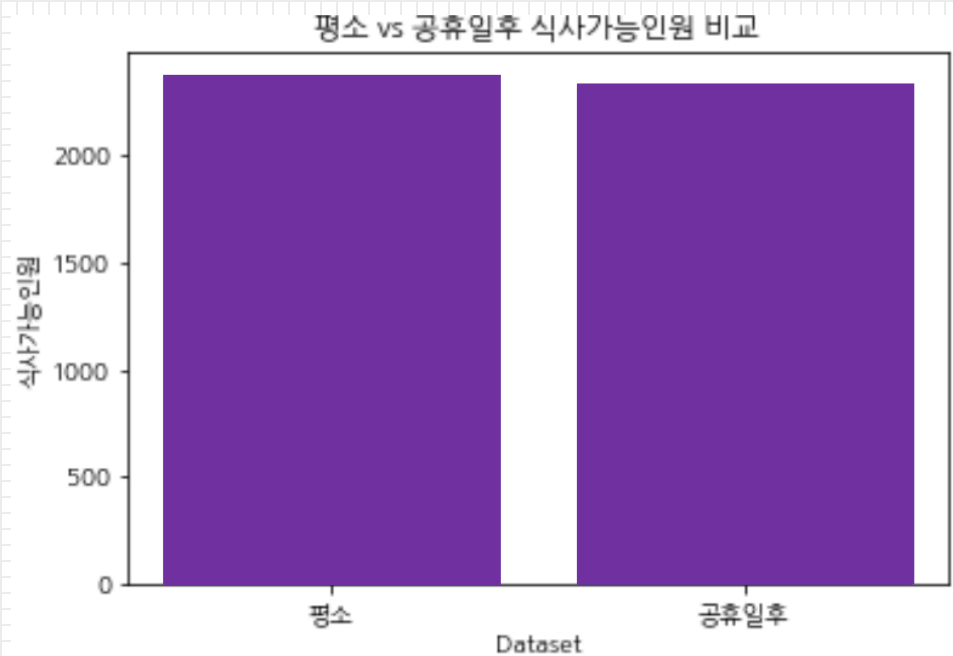
- 공휴일 후에는 평소보다 중식계, 석식계 모두 수요가 많았다. 이는 공휴일 전에 식사 수요가 평소보다 적었던 것과 대비되는 결과이다.



## 날짜 관련 피쳐

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날

## 평소 vs 공휴일 후 '식사가능인원', '휴가자수' 비교



- 공휴일 후의 식사가능인원은 평소보다 약간 적은 것을 확인할 수 있었다.
- 그럼에도 중식계, 석식계 수가 높았던 것을 통해 공휴일 후에는 직원들이 사내식당에서 식사를 해결하려는 경향이 있다는 것을 알 수 있다.

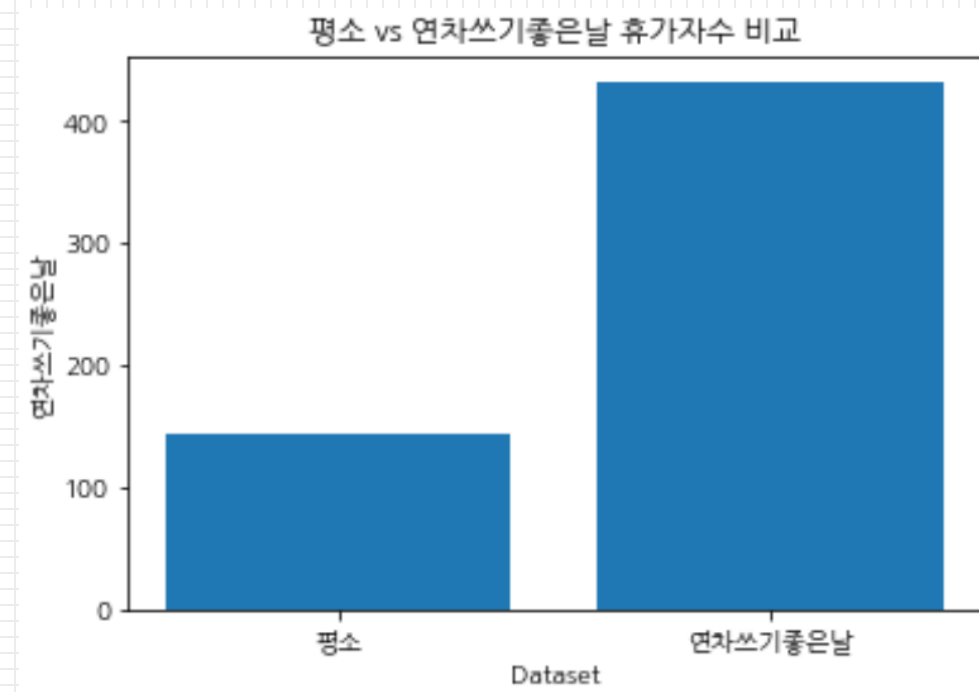
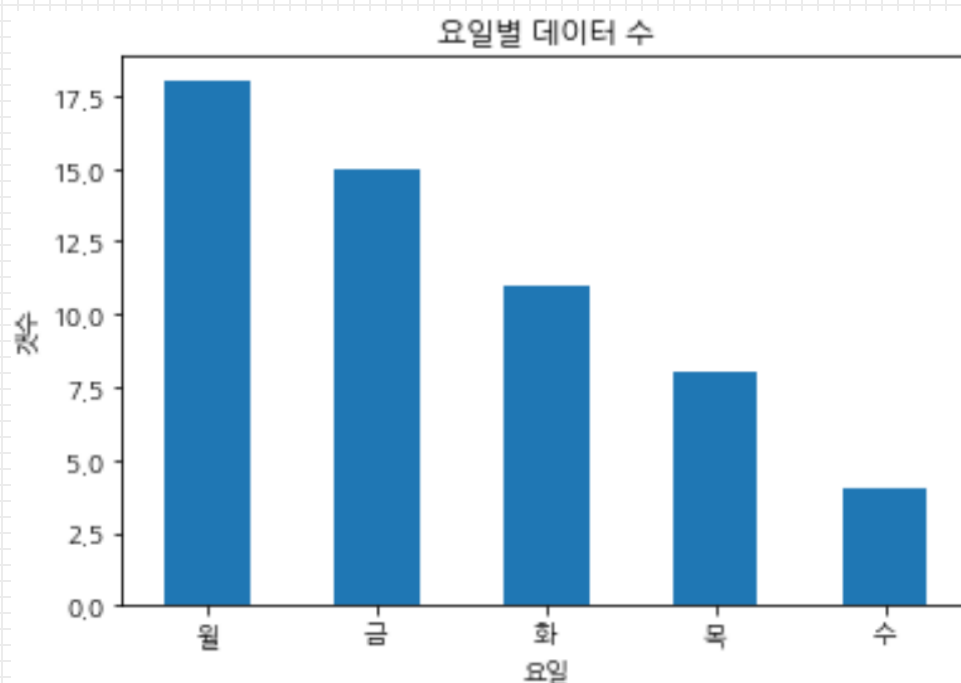




## 날짜 관련 피쳐

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날

## 연차 쓰기 좋은 날 EDA



- 선정 결과 연차쓰기 좋은 날은 주말 전/후인 월요일, 금요일이 가장 많이 선정되었고 화요일, 목요일, 수요일이 뒤를 이었다.
- 연차 쓰기 좋은 날의 휴가자 수는 실제로 평소보다 약 3배 정도 더 많았다.



## 인원수 관련 피처

6	7	8	9	10
식사가능인원	야근수	휴가자수	재택근무자수	출장자수

### 피처 설명

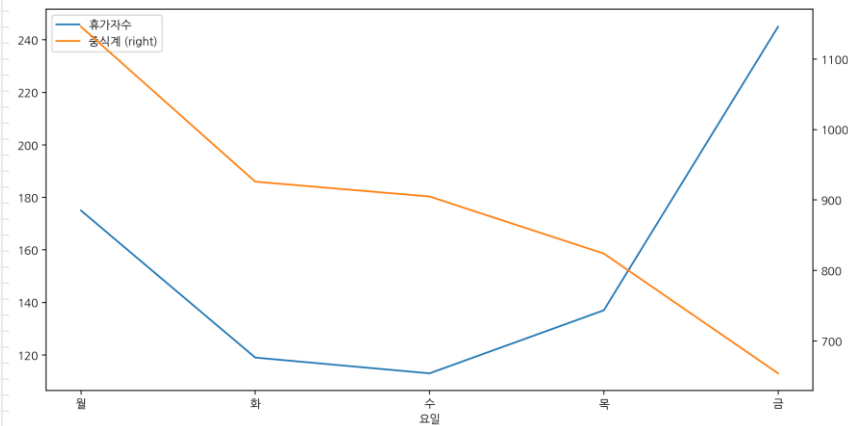
- **식사가능인원**       $\text{식사가능인원} = \text{본사정원 수} - (\text{본사휴가자수} + \text{본사출장자수} + \text{현본사소속재택근무자수})$
- **야근수**       $= \text{본사시간외근무명령서승인건수}$
- **휴가자수**       $= \text{본사휴가자수}$
- **재택근무자수**       $= \text{현본사소속재택근무자수}$
- **출장자수**       $= \text{본사출장자수}$



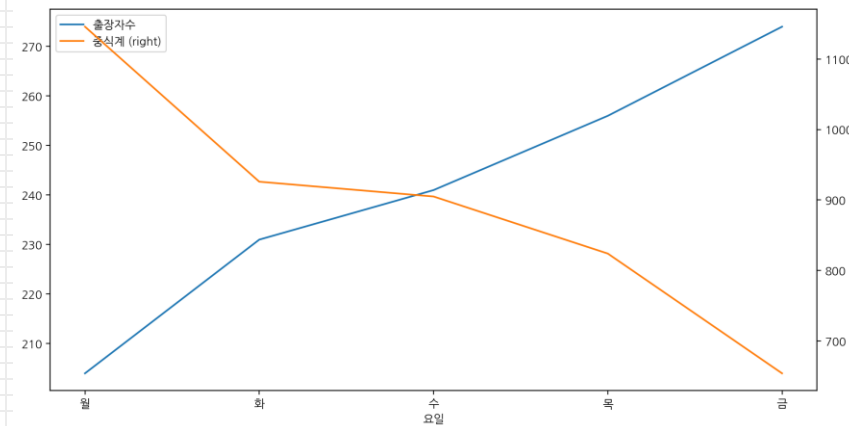
## 인원수 관련 피쳐

6	7	8	9	10
식사가능인원	야근수	휴가자수	재택근무자수	출장자수

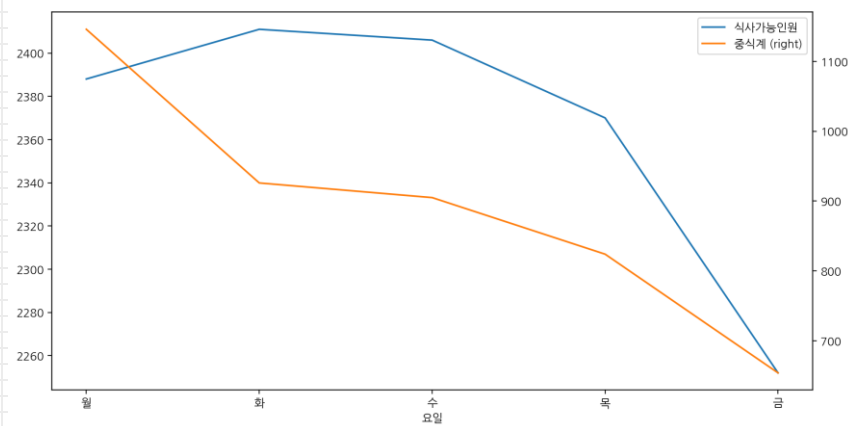
### 휴가자수와 종식계



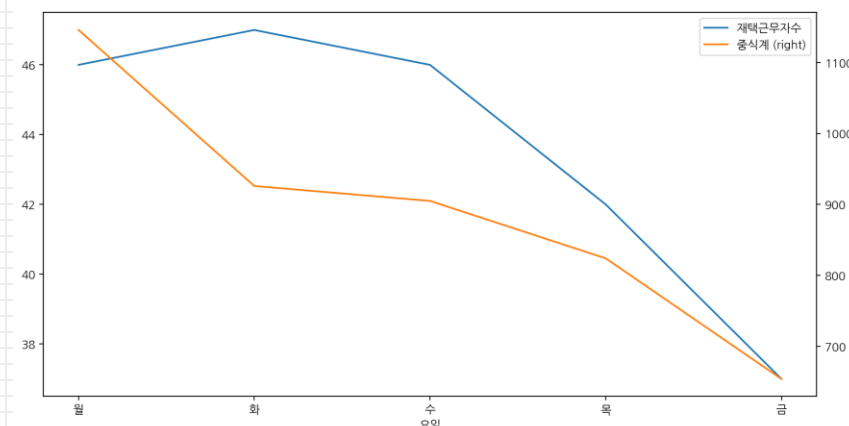
### 출장자수와 종식계



### 식사가능인원과 종식계



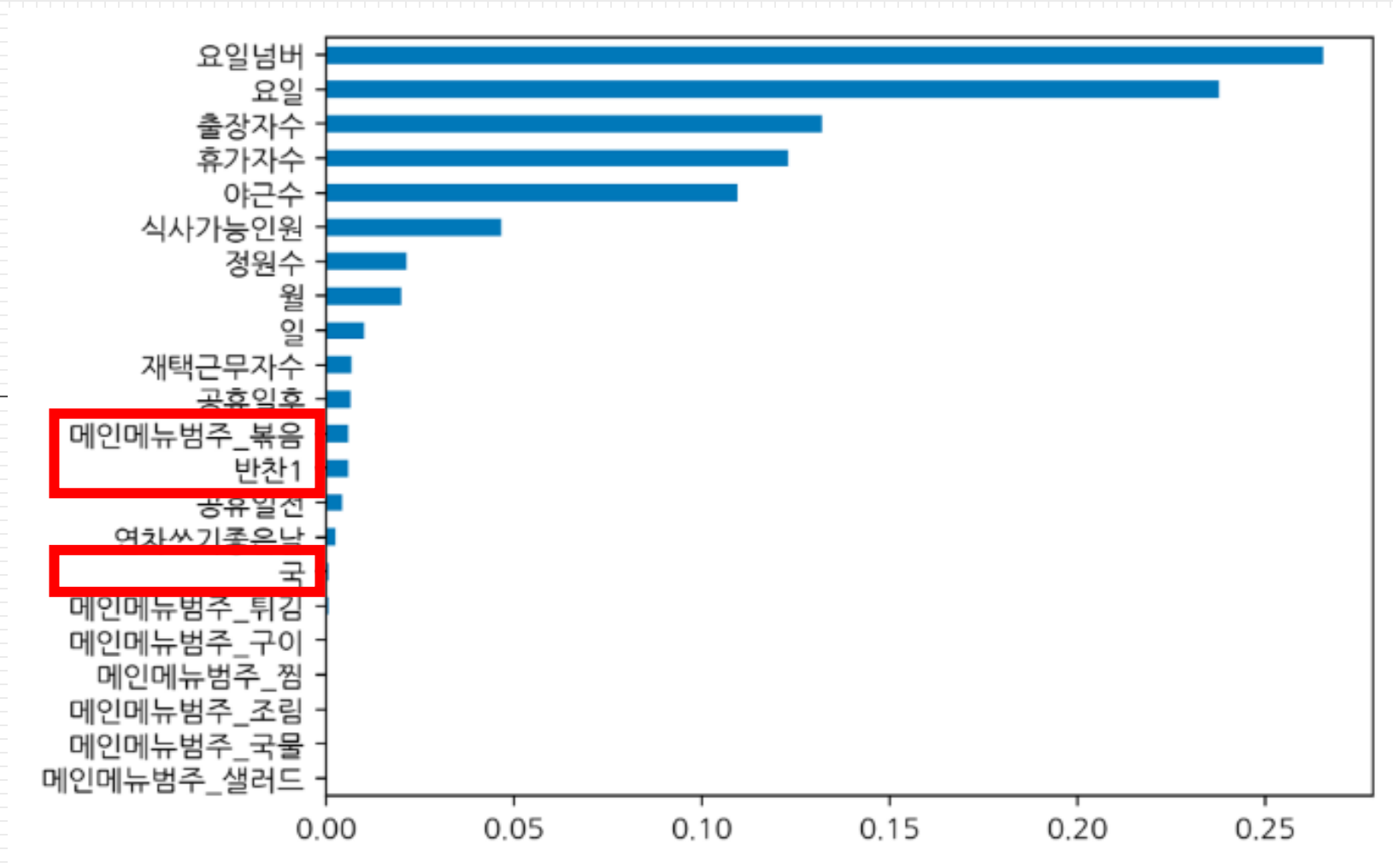
### 재택근무자수와 종식계





## 메뉴 관련 피쳐

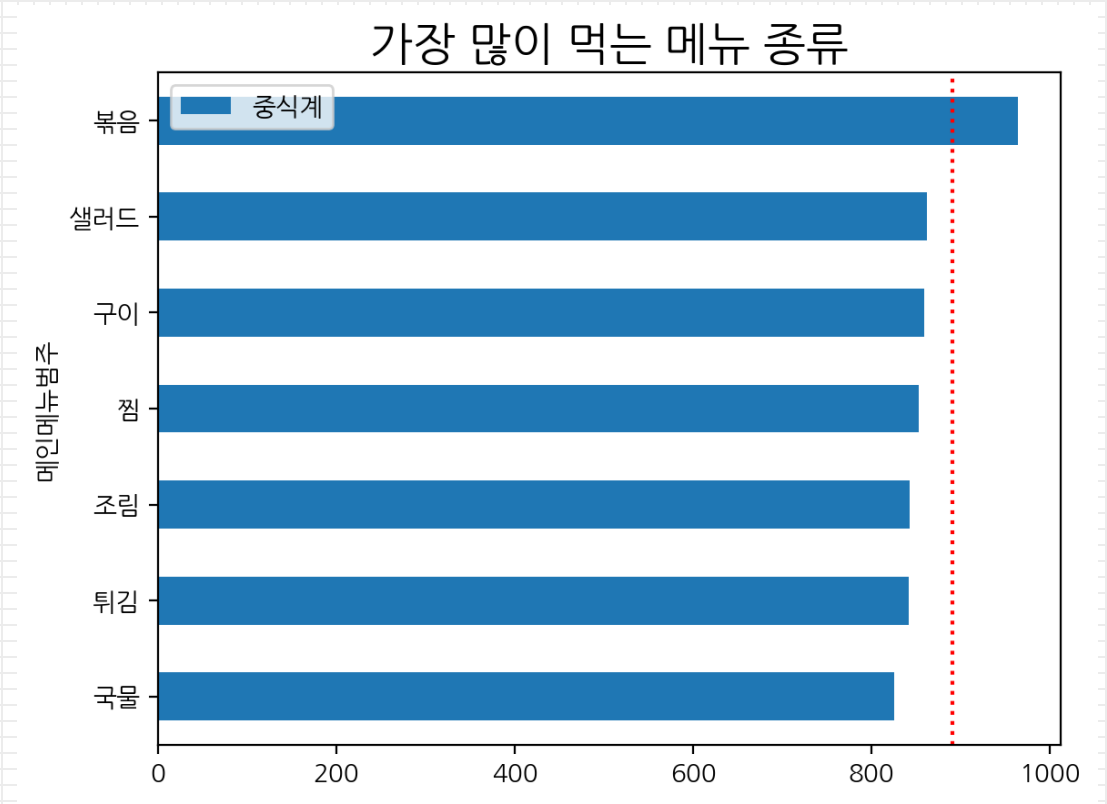
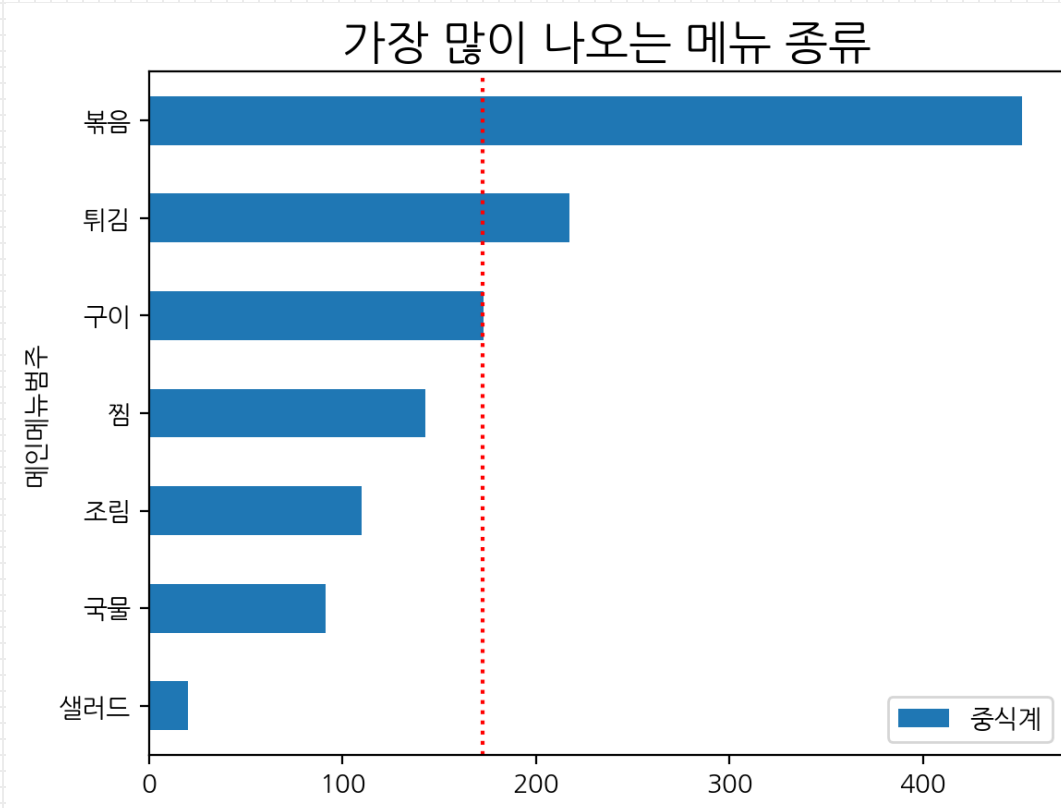
11	12	13
반찬1	국	메인메뉴범주_볶음





## 메뉴 관련 피쳐

11	12	13
반찬1	국	메인메뉴범주 _볶음



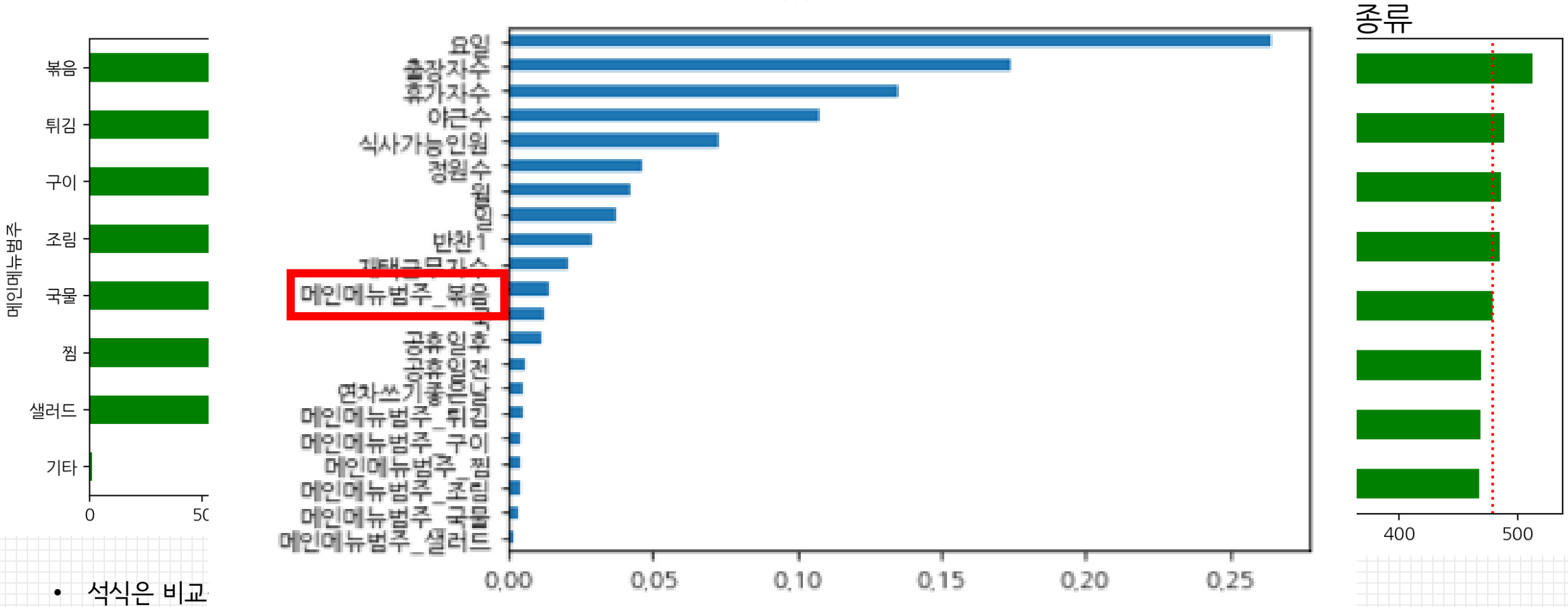
- 볶음류가 압도적으로 많이 나오지만, 중식계 식수에 조리법은 큰 영향을 미치지 못하는 것으로 보인다.





메뉴 관련 피쳐

11	12	13
반찬1	국	메인메뉴범주 _볶음





# 4

## 선정되지 못한 피처들

국가별 메뉴 범주화 / 반찬 개수 / 날씨 관련 피처



으쌔으쌔



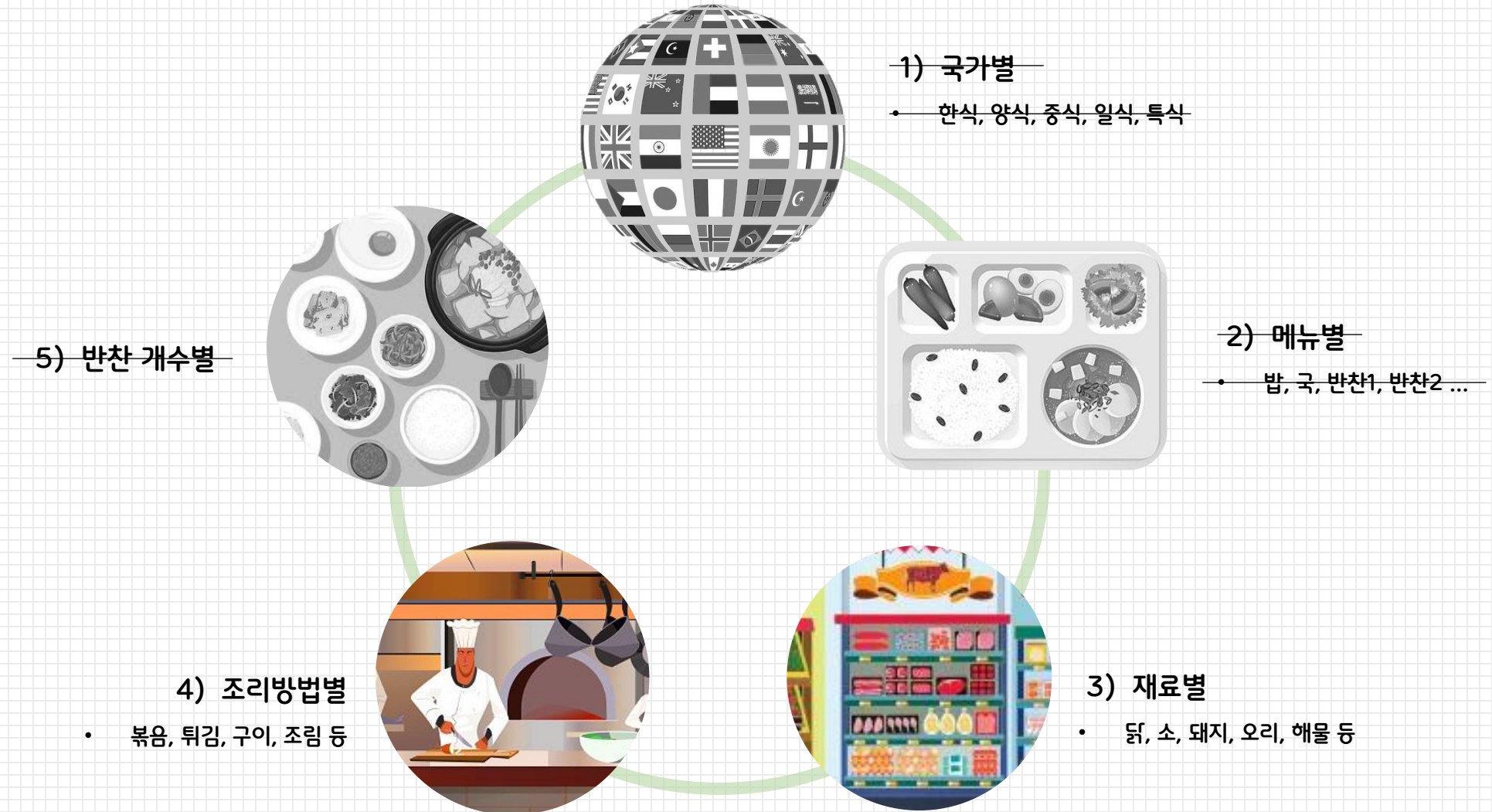
4팀



1석 4조

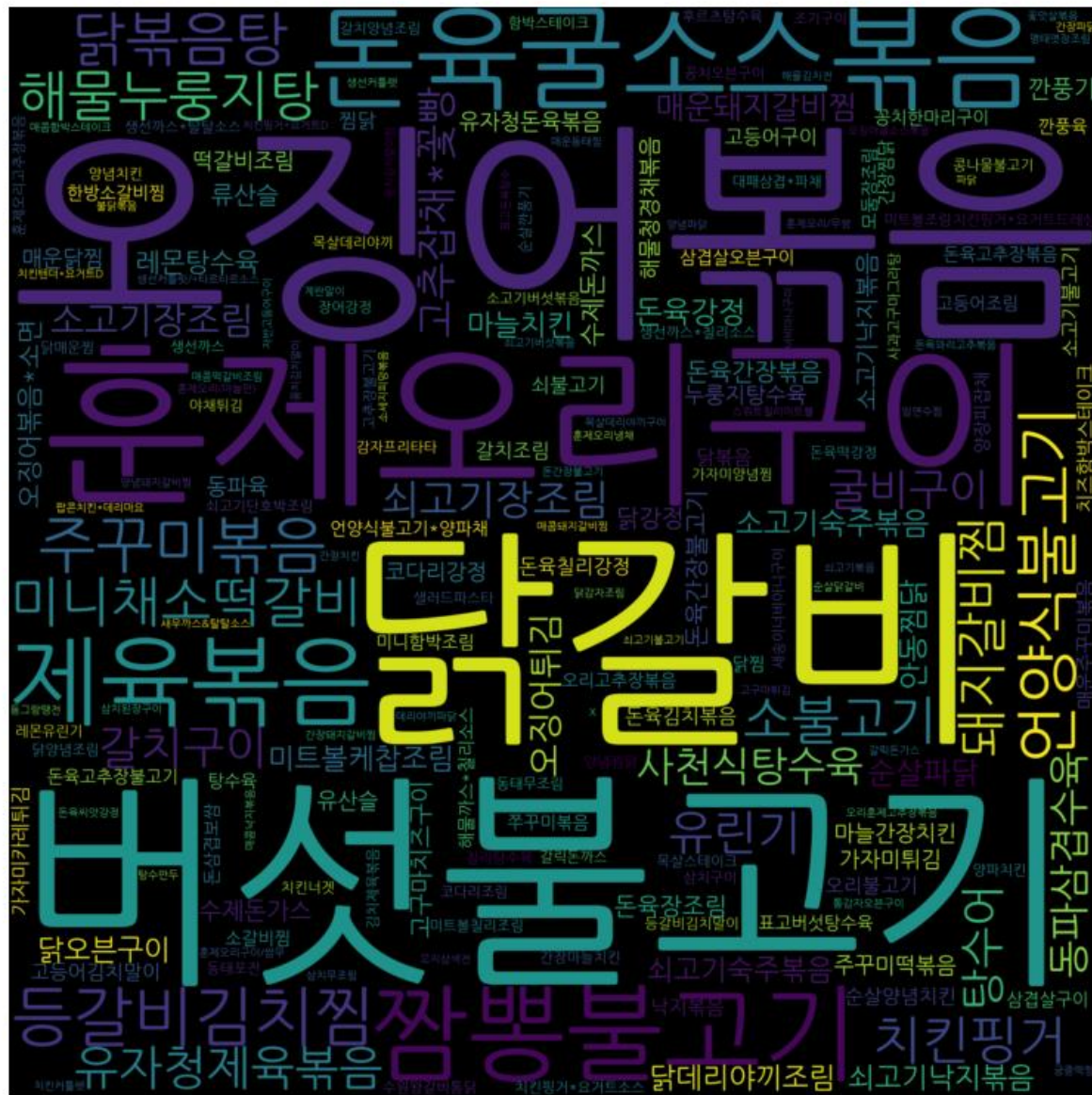


## 메뉴 관련 피처





## 메뉴 관련

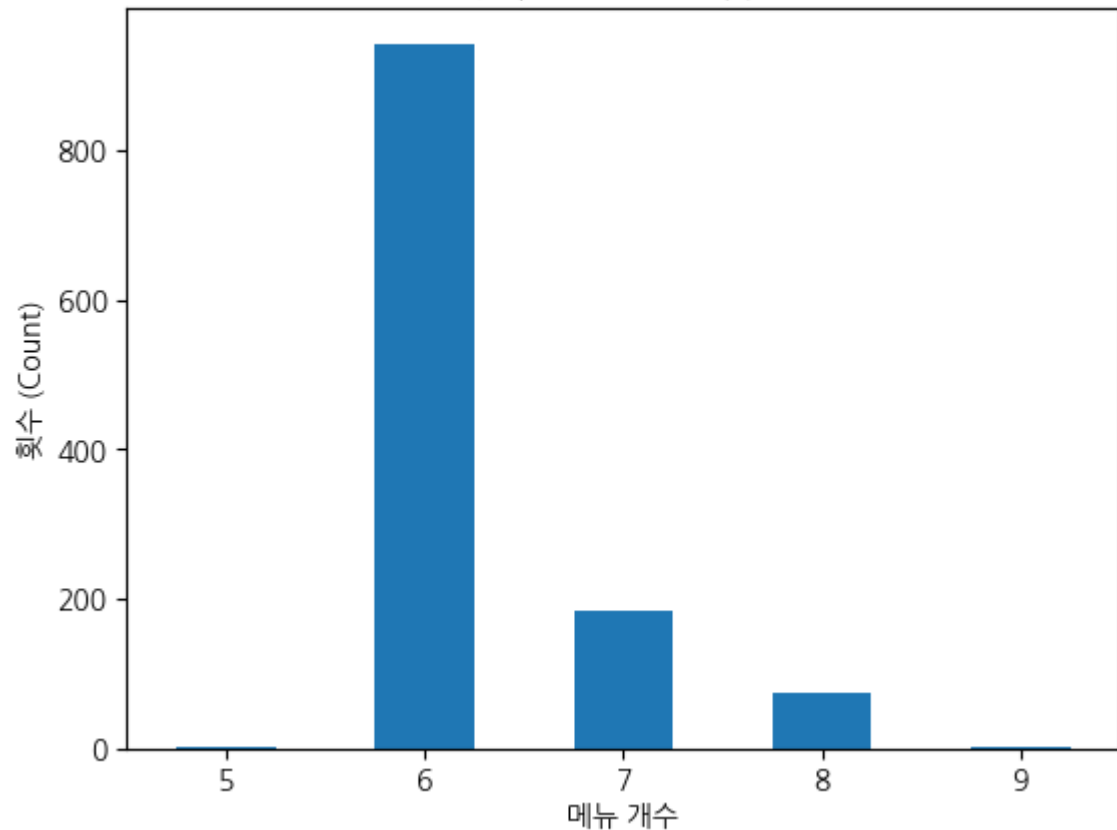




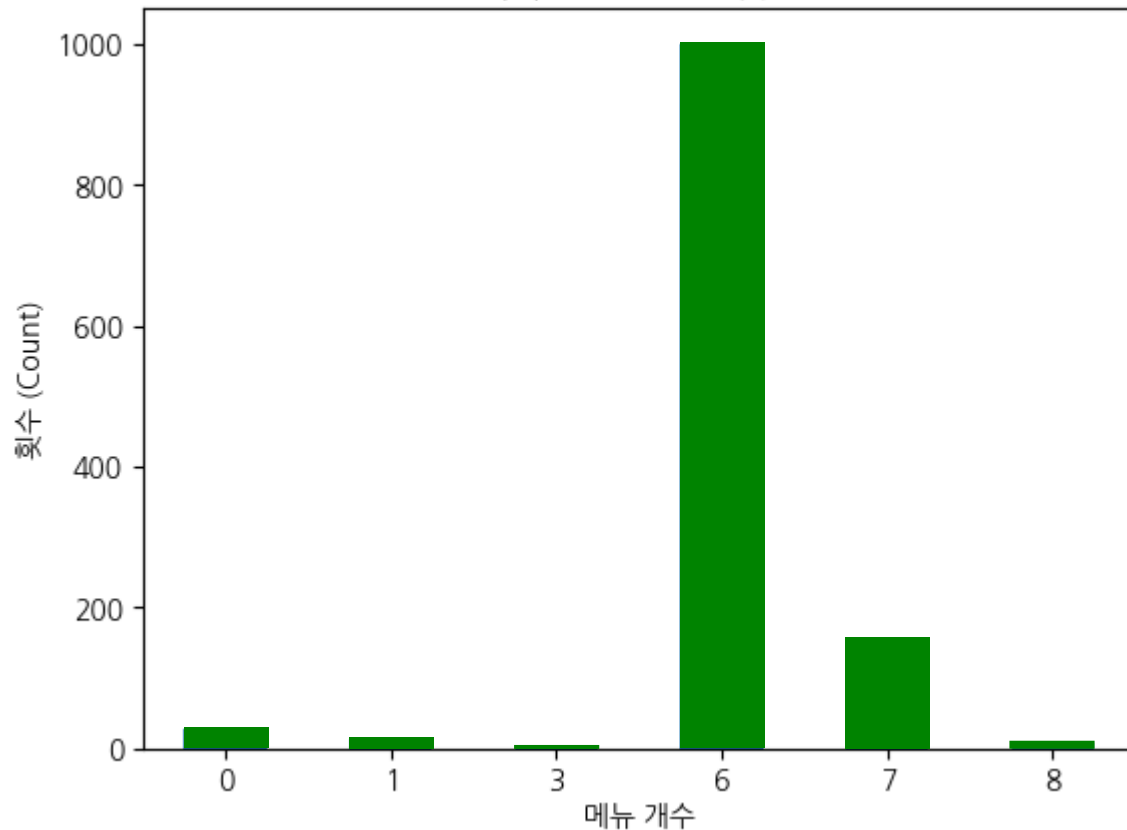
메뉴 관련

## 반찬 개수와 식사인원

중식 메뉴 개수 별 횟수



석식 메뉴 개수 별 횟수



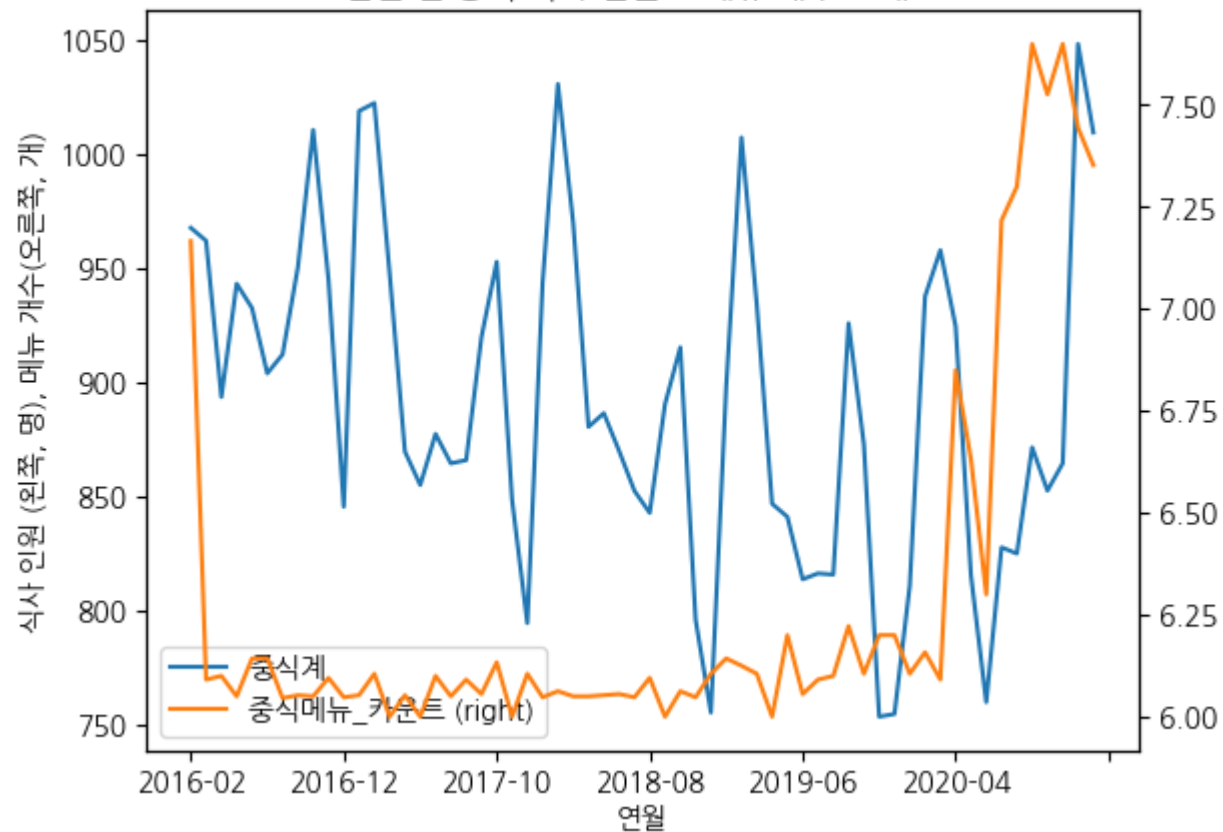




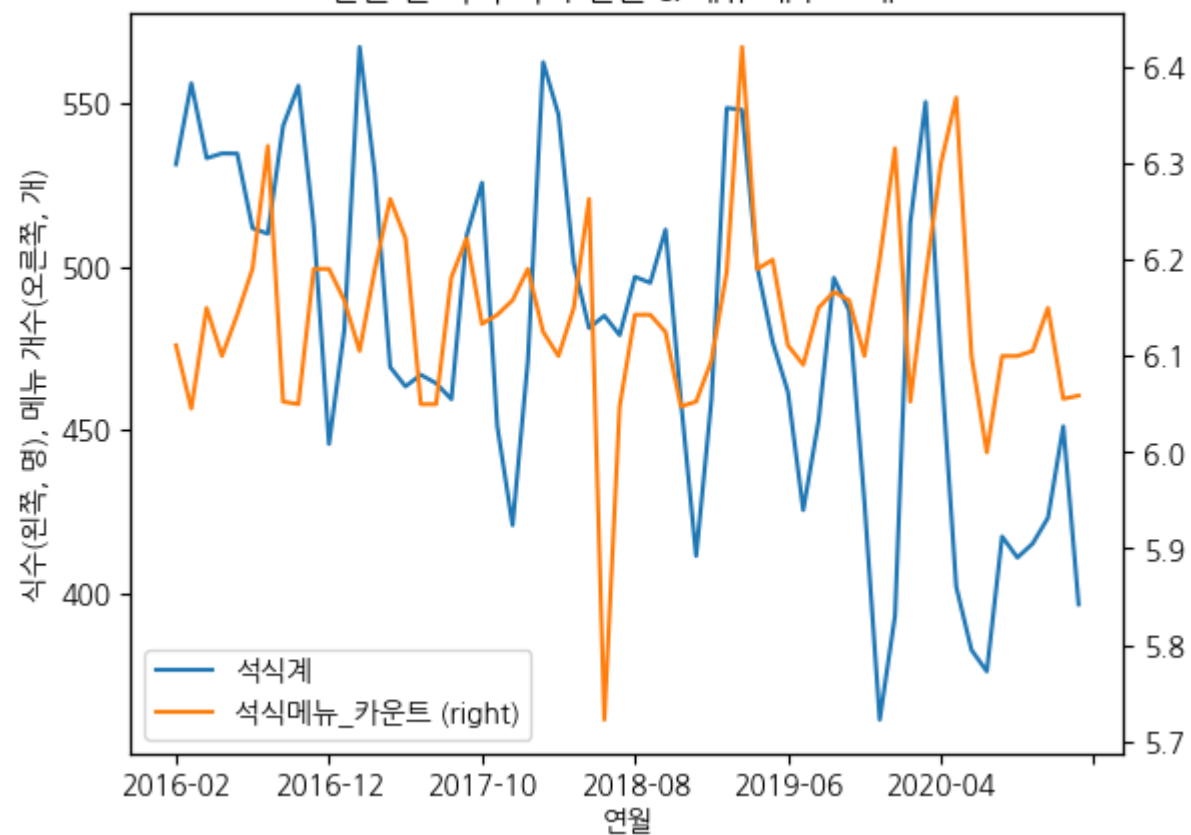
메뉴 관련

## 반찬 개수와 식사인원

연월 별 중식 식사 인원 & 메뉴 개수 그래프



연월 별 석식 식사 인원 & 메뉴 개수 그래프

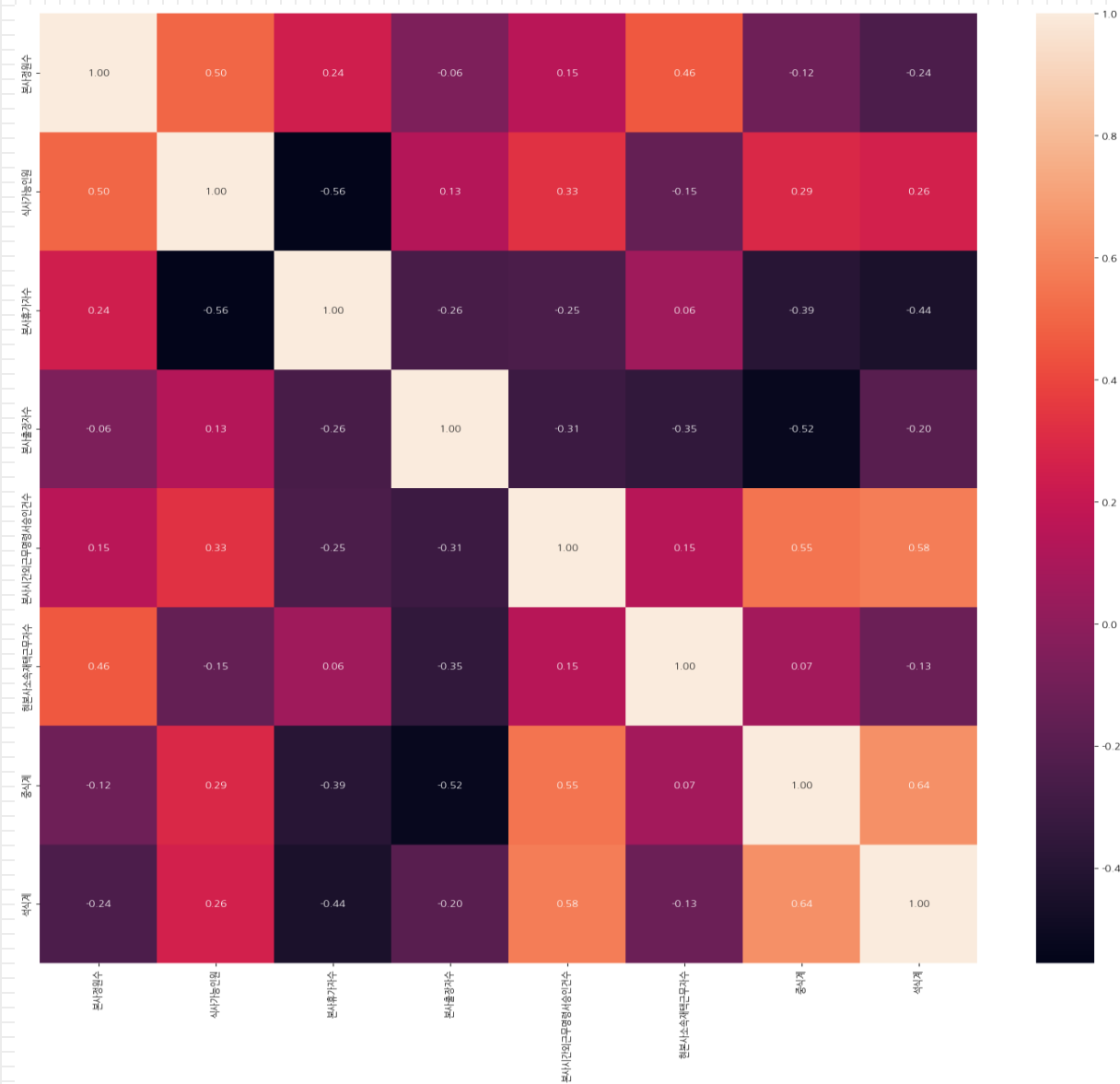




메뉴 관련

## 반찬 개수와 식사인원

	중식계	석식계	중식메뉴_카운트	석식메뉴_카운트
중식계	1.000000	0.644331	0.050389	-0.047345
석식계	0.644331	1.000000	-0.179723	0.047036
중식메뉴_카운트	0.050389	-0.179723	1.000000	-0.035670
석식메뉴_카운트	-0.047345	0.047036	-0.035670	1.000000

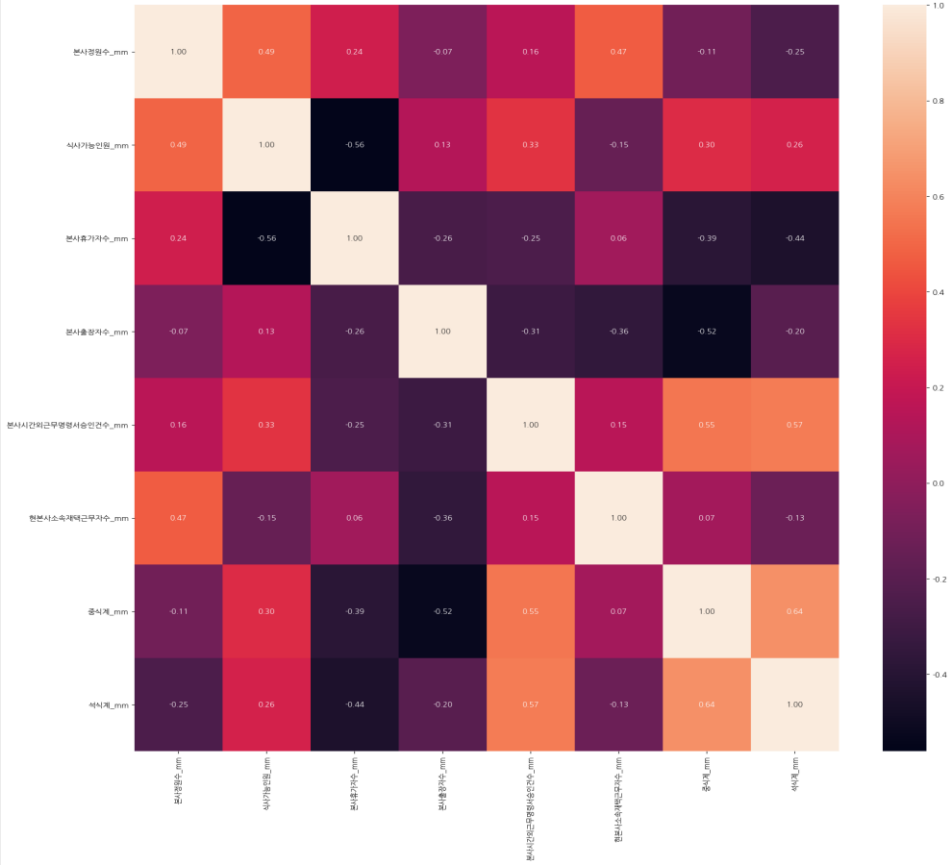




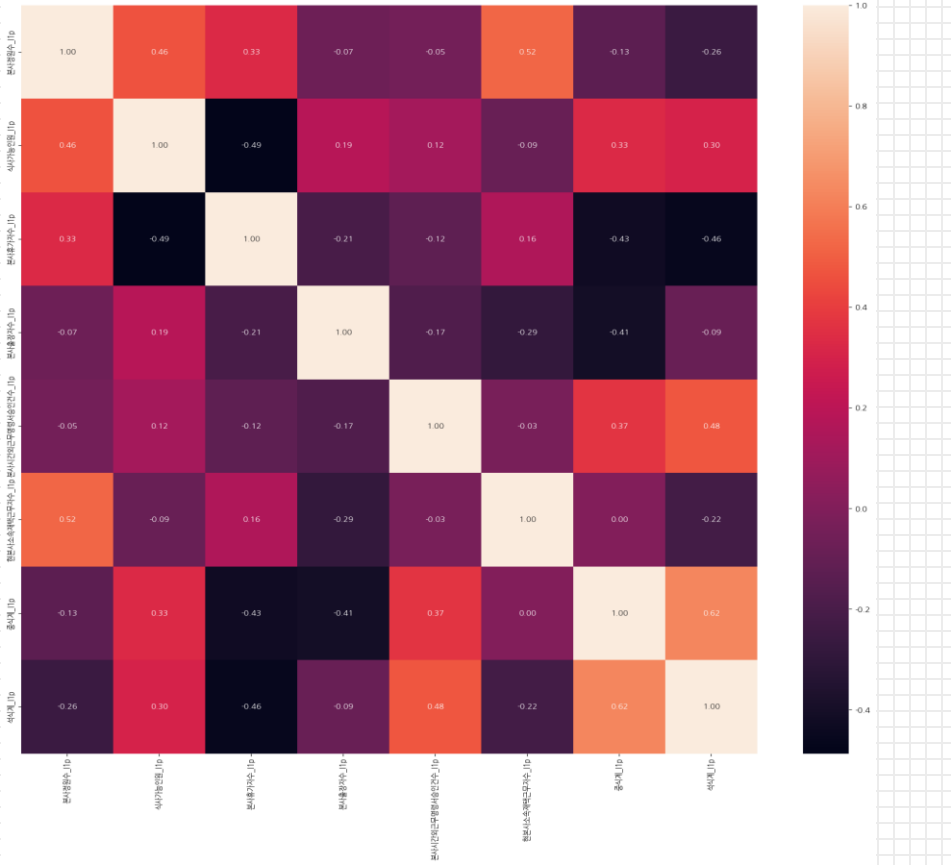
메뉴 관련

## 반찬 개수와 식사인원

인원수 Min-Max Scaling



인원수 log1p Scaling





날씨 관련

## 날씨 관련 피쳐

### [ 폭염 ]



#### 데이터) 여름 데이터

- 안전 : 33°C 미만
- 폭염 : 33°C 이상
- 폭염경보 : 35°C 이상

### [ 한파 ]



#### 데이터) 겨울 데이터

- 안전 : -12°C 초과
- 한파 : -12°C 이하
- 한파경보 : -15°C 이하

### [ 불쾌지수 ]



#### 데이터) 여름 데이터

- 데이터 범주화 기준  
: 상위 100일, 하위 100일의 불쾌지수의 평균

### [ 강수량 ]



#### 데이터) 전체 데이터

- 약한 비 (0) : 시간당 3mm 미만
- 보통 비 (1) : 시간당 15mm 미만
- 강한 비 (2) : 시간당 30mm미만
- 매우 강한 비 (3) : 시간당 30mm이상



날씨 관련

## 날씨 관련 데이터와 상관계수 확인하기

	정원수	휴가자수	출장자수	야근수	재택근무자수	중식계	석식계	식사가능인원	불쾌지수	최고기온(℃)	최저기온(℃)	rain
정원수	1.000000	0.244666	-0.061688	0.146544	0.452293	-0.115529	-0.173852	0.503771	0.070706	0.062000	0.079941	0.034379
휴가자수	0.244666	1.000000	-0.273318	-0.230789	0.065397	-0.391975	-0.316894	-0.552661	0.219685	0.205482	0.212695	0.022678
출장자수	-0.061688	-0.273318	1.000000	-0.308757	-0.352643	-0.512680	-0.188164	0.136389	0.108485	0.134143	0.100638	-0.024423
야근수	0.146544	-0.230789	-0.308757	1.000000	0.156792	0.535611	0.571168	0.309118	-0.058311	-0.055188	-0.053149	-0.007575
재택근무자수	0.452293	0.065397	-0.352643	0.156792	1.000000	0.076509	-0.057534	-0.150393	-0.025736	-0.025459	-0.031830	0.015028
중식계	-0.115529	-0.391975	-0.512680	0.535611	0.076509	1.000000	0.508287	0.286810	-0.096058	-0.122591	-0.072575	0.088758
석식계	-0.173852	-0.316894	-0.188164	0.571168	-0.057534	0.508287	1.000000	0.172373	0.007206	0.008034	0.003260	0.027613
식사가능인원	0.503771	-0.552661	0.136389	0.309118	-0.150393	0.286810	0.172373	1.000000	-0.121523	-0.124890	-0.101150	0.011491
불쾌지수	0.070706	0.219685	0.108485	-0.058311	-0.025736	-0.096058	0.007206	-0.121523	1.000000	0.956074	0.976952	0.167374
최고기온(℃)	0.062000	0.205482	0.134143	-0.055188	-0.025459	-0.122591	0.008034	-0.124890	0.956074	1.000000	0.901720	0.065174
최저기온(℃)	0.079941	0.212695	0.100638	-0.053149	-0.031830	-0.072575	0.003260	-0.101150	0.976952	0.901720	1.000000	0.232453
rain	0.034379	0.022678	-0.024423	-0.007575	0.015028	0.088758	0.027613	0.011491	0.167374	0.065174	0.232453	1.000000

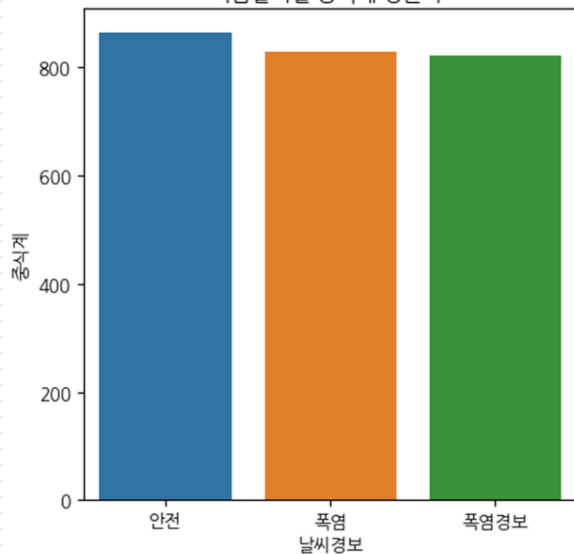


날씨 관련

## 날씨 관련 데이터와 **중식계** 시각화

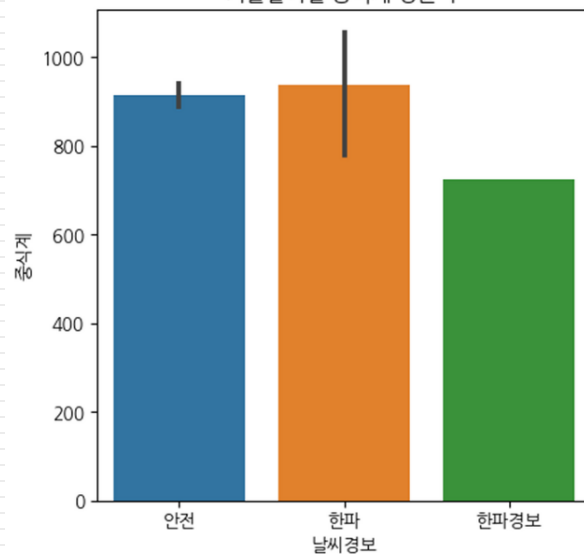
[ 폭염 ]

여름날씨별 중식계 평균 수



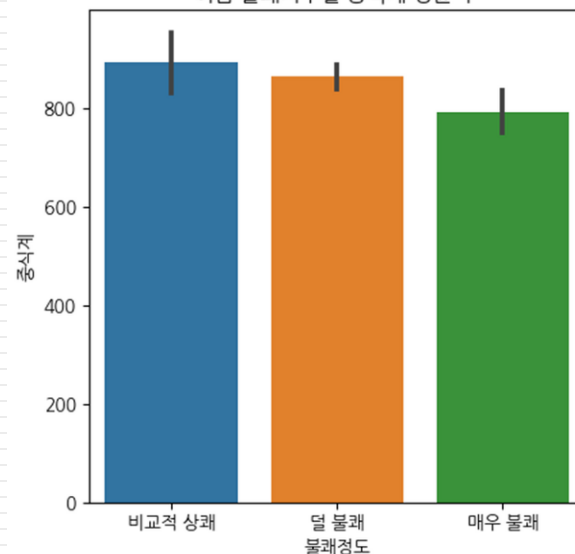
[ 한파 ]

겨울날씨별 중식계 평균 수



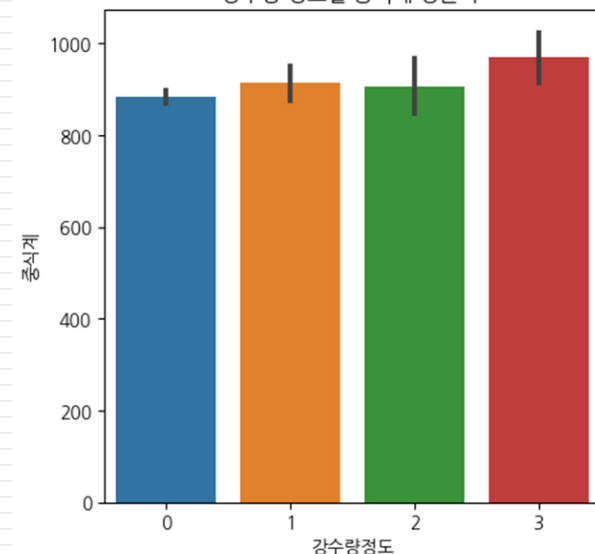
[ 불쾌지수 ]

여름 불쾌지수별 중식계 평균 수



[ 강수량 ]

강수량 정도별 중식계 평균 수



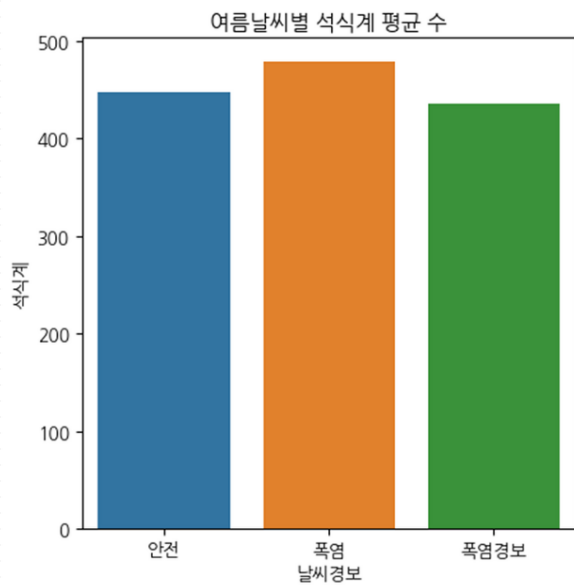
- 폭염은 미세한 차이이지만 더울수록 중식계 수가 줄어듦을 확인할 수 있다.
- 한파의 경우, 기상청 기준으로 범주화하여 '한파경보'에 해당하는 경우가 3개였다. 따라서 정확한 분석이 어려울 것으로 판단된다.
- 불쾌지수는 높을수록 중식계 수가 줄어들었고, 강수량은 많을 수록 중식계 수가 증가하였다.



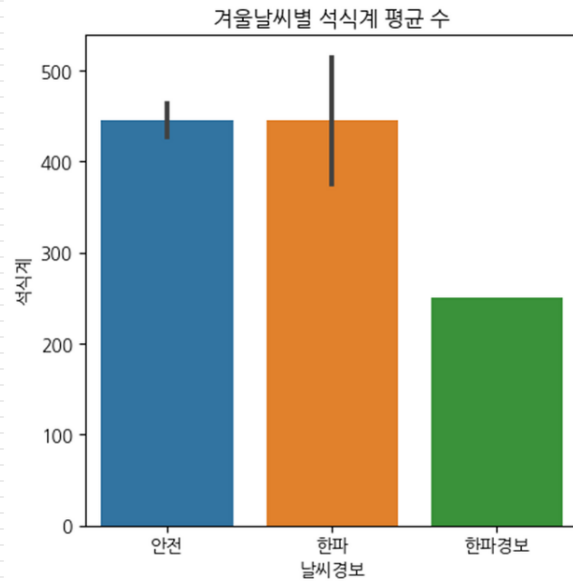
날씨 관련

## 날씨 관련 데이터와 석식계 시각화

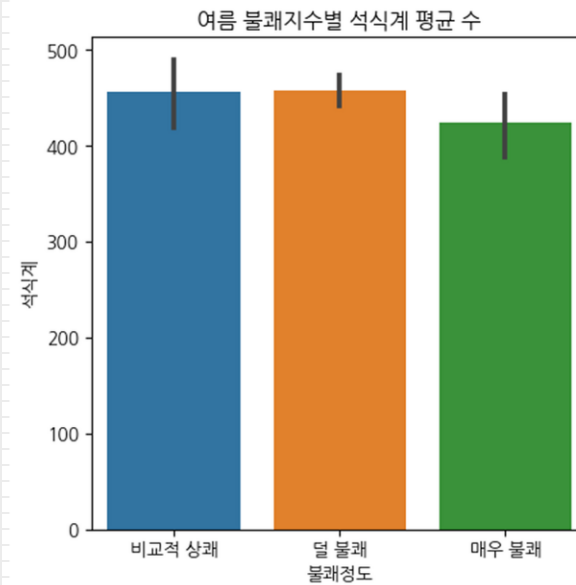
[ 폭염 ]



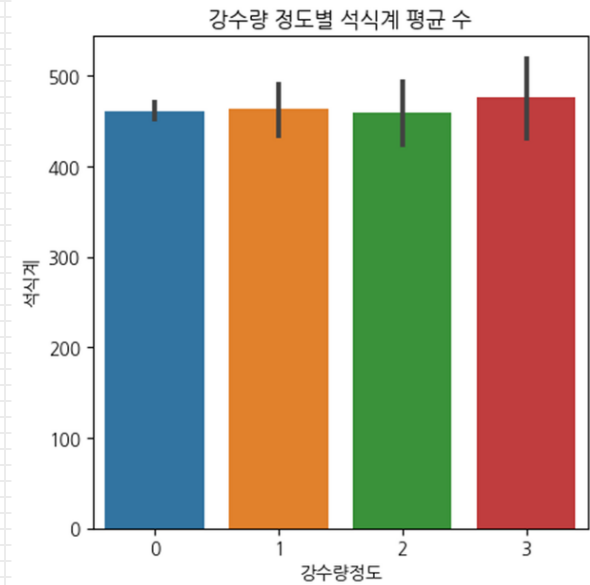
[ 한파 ]



[ 불쾌지수 ]



[ 강수량 ]



- 그러나, 전체적으로 중식계/석식계에 큰 영향을 끼치는 피처가 아니라고 판단되었으며, test셋에 적용하기 위해서는 날씨를 예측하기 위해 시계열 예측을 사용해야 하므로 최종적으로는 제외하기로 결정하였다.





# 5

## 결론 및 회고

머신러닝 시행착오 / 팀원별 자기평가 / 개선점



으쌔으쌔



4팀



1석 4조



## 머신러닝 돌리기

No.	모델명	피쳐 특징	■피쳐범주	피쳐
1	랜덤포레스트	메뉴 범주화(종류별)	메뉴,인원비율,날짜	'월', '요일', '정원수', '종류', '중식계', '석식계'
2	랜덤포레스트	메뉴다백고 공휴일전/후,연차칼럼추가	인원,날짜,공휴일	'월','일','출장자수','차쓰기틀은날'
3	randomforest	메인메뉴범주화(조리방법별)	메뉴,인원,날짜	연도,월,일,요일,가능인원,야근수,조임)
4	auto gluon	인원 피쳐만 사용	인원	'정원수', '휴가자수', '석식계'
5	auto gluon	월,요일 피쳐 추가	인원,날짜	'월', '요일', '정원수', '중식계', '석식계'
6	auto gluon	국(6가지), 반찬(19가지) 범주화	인원,날짜,메뉴	'월', '요일', '정원수', '국', '반찬1', '반찬2', '반찬3', '반찬4', '반찬5', '반찬6', '반찬7', '반찬8', '반찬9', '반찬10', '반찬11', '반찬12', '반찬13', '반찬14', '반찬15', '반찬16', '반찬17', '반찬18', '반찬19'
7	Catboost	메뉴 범주화(종류별) + 점심에는 종류 빼주고 들림	인원비율,날짜,저녁만메뉴	점심: '연', '월', '요일', '택비율', '중식계', '저녁: '연', '월', '요일', '택비율', '중류', '석식계'
8	catboost	메인메뉴범주화(조리방법별)	날짜,인원,메뉴	연도,월,일,요일,가능인원,야근수,조임)
9	Catboost	메뉴를 모두 빼고 인원수와 시간 주기성을 예측할 수 있는 데이터	인원,날짜	점심: '월', '요일', '저녁: '연', '월', '요일', '택비율', '중식계', '저녁: '연', '월', '요일', '택비율', '중류', '석식계'
10	randomforest	메뉴빼고	인원,날짜	연도,월,일,요일,가능인원,야근수

19	랜덤포레스트	최종 모은 피쳐 용량			-----점심----- 최적 하이퍼 파라미터: {'max_depth': 10, 'max_features': 0.4, 'min_samples_leaf': 8, 'n_estimators': 500} Lunch_MAE: 76.273 -----저녁----- 최적 하이퍼 파라미터: {'max_depth': 15, 'max_features': 1.0, 'min_samples_leaf': 8, 'n_estimators': 500} Dinner_MAE: 47.571	80.50874	274	김조은
20	auto gluon			'반찬1', '요일', '메인메뉴범주_볶음', '재택근무자수', '공휴일전', '정원수', '공휴일후', '휴가자수', '메인메뉴범주_조림', '메인메뉴범주_샐러드', '메인메뉴범주_찜', '메인메뉴범주_국물', '식사가능인원', '메인메뉴범주_튀김', '출장자수', '월', '연', '연차쓰기좋은날', '메인메뉴범주_구이', '야근수'	problem_type='regression', eval_metric='mean_absolute_error' presets='best_quality', num_stack_levels=3	90.235001	388	김성우
21	랜덤포레스트	최종 모은 피쳐 용량		'월','일','요일','휴가자수','국','출장자수','야근수','식사가능인원', '정원수','공휴일전','공휴일후','연차쓰기좋은날','반찬1','재택근무자수', '메인메뉴범주_샐러드','메인메뉴범주_튀김','메인메뉴범주_찜','메인메뉴범주_볶음', '메인메뉴범주_구이','메인메뉴범주_국물','메인메뉴범주_조림'	{'max_depth': 21, 'max_features': 6, 'n_estimators': 850}	90.740294 1176	395	정지원
22	auto gluon	이전 머신러닝 에서 중요 피쳐 만 포함하여 돌려봄		'야근수', '연', '휴가자수', '출장자수', '월', '요일', '정원수', '식사가능인원', '재택근무자수', '반찬1'	problem_type='regression', eval_metric='mean_absolute_error' presets='best_quality', num_stack_levels=3	99.402249 6667		김성우
23	랜덤포레스트	피쳐: 조리법별 + 공휴일 파라미터: 런치/ 디너 파라미터 따로		'월','일','요일','휴가자수','출장자수','야근수','식사가능인원','공휴일전','공휴일후', '연차쓰기좋은날','메인메뉴범주_구이','메인메뉴범주_국물', '메인메뉴범주_볶음','메인메뉴범주_샐러드', '메인메뉴범주_조림','메인메뉴범주_찜','메인메뉴범주_튀김'	lunch: criterion='absolute_error', max_depth=9, max_features=0.88000000000000000001, max_samples=0.80000000000000000003, n_estimators=600, random_state=42  dinner: criterion='absolute_error', max_depth=9, max_features=0.96000000000000000002, max_samples=0.75000000000000000002, n_estimators=500, random_state=42	81.508033 3333	292	김희복
24	켓부스트 + 랜덤포레스트			'월','일','출장자수','야근수','식사가능인원','공휴일전','공휴일후','연차쓰기좋은날'	켓부스트 : 파라미터 없음 랜덤포레스트 : n_estimators=1000, max_features=0.95	68.87633	125	정지원

그리드서치로 찾은거 'max_depth': 10, 'max_features': 0.4, 'min_samples_leaf': 8, 'n_estimators': 1000	80.508743 7015		김조은
random_state=42, criterion='absolute_error', n_estimators=300, n_jobs=-1	84.398166 6667		김희록
random_state=42, criterion='absolute_error', n_estimators=300, n_jobs=-1	84.355166 6667	329	김희록
random_state=42, criterion='absolute_error', n_estimators=300, n_jobs=-1	84.941666 6667		김희록
점심 - n_estimators=250 - learning_rate=0.36	71.43333	Public 기준 161	김준호
저녁 - n_estimators=250 - learning_rate=0.27			
problem_type='regression', eval_metric='mean_absolute_error'	84.672856 6667	340	김성우
problem_type='regression', eval_metric='mean_absolute_error', presets='best_quality', num_stack_levels=3	78.344033 6667	227	김성우
problem_type='regression', eval_metric='mean_absolute_error', presets='best_quality', num_stack_levels=3	84.586613 6667	339	김성우

## 최종 선택 피처

[ train\_lunch 데이터셋 ]

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날
6	7	8	9	10	11
식사가능인원	야근수	휴가자수	재택근무자수	출장자수	반찬1
12	13	14			
국	메인메뉴범주 _볶음	중식계			

[ train\_dinner 데이터셋 ]

0	1	2	3	4	5
월	일	요일	공휴일전	공휴일후	연차 쓰기 좋은 날
6	7	8	9	10	11
식사가능인원	야근수	휴가자수	재택근무자수	출장자수	반찬1
12	13	14			
국	메인메뉴범주 _볶음	석식계			


## 머신러닝 결과 최종 피쳐

모델명	피쳐	train_lunch 파라미터	train_dinner 파라미터	public	private
catboost	최종피쳐	'learning_rate': 0.035743, 'loss_function': 'RMSE', 'subsample': 0.719163, 'max_depth': 4, 'n_estimators': 539, 'colsample_bylevel': 0.65762, 'reg_lambda': 74.378856	'learning_rate': 0.1335167628101091, 'loss_function': 'RMSE', 'subsample': 0.5587791418211517, 'max_depth': 3, 'n_estimators': 345, 'colsample_bylevel': 0.9510656252991059, 'reg_lambda': 8.885771302364363	84.5916149757	120.4436
랜덤포레스트	최종 피쳐	최적 하이퍼 파라미터: {'max_depth': 21, 'max_features': 5, 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 300} Lunch MAE: 72.193	최적 하이퍼 파라미터: {'max_depth': 17, 'max_features': 5, 'min_samples_leaf': 2, 'min_samples_split': 4, 'n_estimators': 100} Dinner MAE: 45.685	82.7887349102	124.4969
catboost	최종 피쳐	n_estimators: 590 learning_rate: 0.02	n_estimators: 530 learning_rate: 0.04	95.9666666667	128.0286


## 머신러닝 결과 그 외

모델명	피쳐	train_lunch 파라미터	train_dinner 파라미터	public	private
catboost(중식) 랜덤포레스트(식 식)	'월', '일', '출장자수', '야근수', '식사가능인 원', '공휴일전', '공휴일후', '연차쓰기좋은 날'	CatBoost 사용, 파라미터 x	랜덤포레스트 사용, n_estimators=1000, max_features=0.95	68.87634	129.1456
catboost	식사가능인원 추가 후, 피쳐 중요도 확인 해서 5 미만의 피쳐는 제거 후 다시 학습 -----점심----- '월', '요일넘버', '식사가능인원', '휴가자 수', '출장자수', '야근수' -----저녁----- '연', '월', '일', '식사가능인원', '휴가자 수', '출장자수', '야근수'	n_estimators: 250 learning_rate: 0.36	n_estimators: 250 learning_rate: 0.27	71.43333	126.7429
auto gluon	월, 요일, 공휴일전, 공휴일후, 정원수, 식 사인원수, 야근수	problem_type='regression', eval_metric='mean_absolute_error'	problem_type='regression', eval_metric='mean_absolute_error'	84.29615	119.4641

## 팀원별 자기평가

 김 성 우


만족도 ★★☆☆☆  
완성도 ★★☆☆☆  
성장률 ★★★★★

 김 조 은


만족도 ★★★★★  
완성도 ★★☆☆☆  
성장률 ★★★★★

 김 준 호

만족도 ★★★★★  
완성도 ★★☆☆☆  
성장률 ★★★★★

 김 현 아

만족도 ★★★★★  
완성도 ★★☆☆☆  
성장률 ★★★★★

 김 희 목

만족도 ★★★☆☆  
완성도 ★★☆☆☆  
성장률 ★★★★★

 정 지 원

만족도 ★★★☆☆  
완성도 ★★☆☆☆  
성장률 ★★★★★

## 개선점

### 개선점(팀 내에서 논의 및 합의된 개선 방향)

- 메뉴 범주화 기준을 좀더 다듬고, 분류에 좀더 신경을 써서 이상치가 되지 않게 하면 좋을 것 같다
- 피처와 모델, 사용할 파라미터를 지정한 후 팀원마다 그리드서치 범위를 나누어 진행하면 다수의 컴퓨터 사용 이점을 살려 시간을 절약할 수 있었을 것 같습니다.
- 이 두가지를 더 개선한다면 더 정확한 예측을 가능하게 할 것이다.

### 추후 개선 계획

추가적으로 파라미터 튜닝을 더 해보고, 다른 모델도 더 돌려보겠다!

THE END

---

감사합니다

으쌔으쌔 4조 (1석 4조)