



구내식당 식수 인원예측

4조 : 정현범 이상준 신종우



1.배경

안녕하세요 여러분! 🍌 구내식당 식수 인원 예측 AI 경진대회에 오신 것을 환영합니다.

지금까지는 단순한 시계열 추세와 담당자의 직관적 경험에 의존하여한국토지주택공사 구내식당 식수 인원을 예측하였으나, 빅데이터 분석으로 얻어지는 보다 정확도 높은 예측을 통해 잔반 발생량을 획기적으로 줄이고자 합니다.

2.목적

구내식당의 요일별 점심, 저녁식사를 먹는 인원을 예측

3.주최 / 주관

주최 : 한국토지주택공사

주관 : 데이콘

***평가기준 : MAE(Mean Absolute Error)
= 평균 절대 오차**

4.참가 대상

일반인, 학생 등 누구나

일자	요일	본사정원수	본사휴가자수	본사출장자수	본사시간외근무명령서...	현본사
2016-02-01	월	2601	50	150		238
2016-02-02	화	2601	50	173		319
2016-02-03	수	2601	56	180		111
2016-02-04	목	2601	104	220		355
2016-02-05	금	2601	278	181		34
		Sum 13005	Sum 538	Sum 904		Sum 1057

일자	사시간외근무명령서...	현본사소속재택근무자수	조식여유	중식여유	석식여유
2021-01-27	5	358.0	모닝콜/연유버터베이글 우...	발밥/죽미밥/찰현미밥 대...	죽미밥 열문순두부찌개 쇠...
2021-01-28	409	348.0	모닝콜/대만샌드위치 우유...	발밥/보리밥/찰현미밥 우...	중우김밥 우동국물 요징어...
2021-01-29	0	294.0	모닝콜/헛개떡 우유/두스...	발밥/죽미밥/찰현미밥 평...	죽미밥 물만두국 카레찜닭...
2021-02-01	538	322.0	모닝콜/축하한치즈떡 우...	발밥/죽미밥/찰현미밥 배...	죽미밥 동태탕 돈육짜리고...
2021-02-02	455	314.0	모닝콜/토마토샌드 우유/...	발밥/달밥/찰현미밥 부대...	죽미밥 바치락살국 쇠고기...
		Sum 1407	Sum 1636.0		

일자	중식계	석식계
2021-01-27	0	0
2021-01-28	0	0
2021-01-29	0	0
2021-02-01	0	0
2021-02-02	0	0
		Sum 0



휴일이 들어간 빨간 날 앞뒤로 휴가자 수가 많을 거 같다.

좋아하는 메뉴가

중식과 석식을 먹는 선택을 하는 데 영향을 미칠까?

날씨가 안 좋으면 국내식당을 이용하지 않을까?

휴일 전날에는 저녁을 안 먹지 않을까?

점심을 먹은 사람은 저녁도 먹을까?

“?”

겨울 저녁에는 밥을 먹고 나갈까?

2016~시간이 지나면서

재택근무 비율이 늘면서 석식 인원이 줄었다.

메뉴가 맛있으면 국내식당 이용 안한다.

요일 별 중식-석식 상관관계가 있을 것이다.



상관관계분석을 위한 데이터전처리

데이터 정보 확인 및 필요한 특성 컬럼 추가

- 메뉴 컬럼 삭제 (상관관계를 곧바로 수치화 할 수 없음)
- 일자, 요일 데이터 dtype변환 및 컬럼추가(연도, 월, 일)

데이터 타입 변경 및 확인

- Datatype object -> int로 변환
- 결측치 확인
- 새로운 피쳐 컬럼 추가(실제점심식사가능자수)

In [72]: df.head()

Out[72]:

	일자	요일	본사정원수	본사휴가자수	본사출장자수	본사시간외근무명령서승인건수	현본사소속재택근무자수	조식메뉴	중식메뉴	석식메뉴	중식계	석식계
0	2016-02-01	월	2601	50	150	238	0.0	모닝콜/진행 우유/두유/주스 계란후라이 호두죽/쌀밥 (쌀:국내산) 된장찌개 쥐...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 오징어찌개 쇠불고기 (쇠고기:호주산) 계란찜 ...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 육개장 자반고등어 구이 두부조림 건파래무침 ...	1039.0	331.0
1	2016-02-02	화	2601	50	173	319	0.0	모닝콜/단호박샌드 우유/두유/주스 계란후라이 팔죽/쌀밥 (쌀:국내산) 호박찐국 찌...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 김치찌개 가지미튀김 모듬소세지구이 마늘증무...	콩나물밥*양념장 (쌀,현미흑미:국내산) 어묵국 유산슬 (쇠고기:호주산) 아삭고추무...	867.0	560.0
2	2016-02-03	수	2601	56	180	111	0.0	모닝콜/베이글 우유/두유/주스 계란후라이 표고버섯죽/쌀밥 (쌀:국내산) 콩나물국...	카레덮밥 (쌀,현미흑미:국내산) 팽이장국 치킨핑거 (닭고기:국내산) 풀면야채무침 ...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 청국장찌개 황태양념구이 (황태:러시아산) 고기...	1017.0	573.0
3	2016-02-04	목	2601	104	220	355	0.0	모닝콜/토마토샌드 우유/두유/주스 계란후라이 닭죽/쌀밥 (쌀,닭:국내산) 근대국...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 쇠고기무국 주꾸미볶음 부추전 시금치나물 ...	미니김밥*겨자장 (쌀,현미흑미:국내산) 우동 멕시칸샐러드 군고구마 푸피클 포...	978.0	525.0
4	2016-02-05	금	2601	278	181	34	0.0	모닝콜/와플 우유/두유/주스 계란후라이 쇠고기죽/쌀밥 (쌀:국내산) 재첩국 방...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 떡국 돈육씨앗강정 (돼지고기:국내산) 우영잡채...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 차돌박이찌개 (쇠고기:호주산) 닭갈비 (닭고기:...	925.0	330.0



요일 데이터 레이블 인코딩

- 월요일 - 1, 화요일 - 2, ... 금요일 - 5

In [84]: #요일도 숫자로 잘 변환되었다. 인트만 .head()

Out[84]:

	요일	본사정원수	본사휴가자수	본사출장자수	본사시간외근무명령서승인건수	현본사소속재택근무자수	중식계	석식계	실제점심식사가능자수	연도	월	일
0	1	2601	50	150	238	0.0	1039.0	331.0	2401.0	2016	2	1
1	2	2601	50	173	319	0.0	867.0	560.0	2378.0	2016	2	2
2	3	2601	56	180	111	0.0	1017.0	573.0	2365.0	2016	2	3
3	4	2601	104	220	355	0.0	978.0	525.0	2277.0	2016	2	4
4	5	2601	278	181	34	0.0	925.0	330.0	2142.0	2016	2	5



상관관계분석을 위한 데이터전처리

	중식계	본사시간외근무 명령서승인건수	석식계	실제점심식사 가능자수	현본사소속재 택근무자수	연도	일	본사정원 수	월	본사휴가 자수	본사출장 자수	요일
중식계	1.000000	0.535611	0.508287	0.286810	0.076509	-0.078804	-0.097392	-0.115529	-0.154664	-0.391975	-0.512680	-0.731563
본사시간 외근무명 령서승인 건수	0.535611	1.000000	0.571168	0.309118	0.156792	0.181749	-0.093606	0.146544	0.018316	-0.230789	-0.308757	-0.554359
석식계	0.508287	0.571168	1.000000	0.172373	-0.057534	-0.194792	-0.185565	-0.173852	-0.127142	-0.316894	-0.188164	-0.312112
실제점심 식사가능 자수	0.286810	0.309118	0.172373	1.000000	-0.150393	0.306486	-0.028433	0.503771	0.006330	-0.552661	0.136389	-0.250092
현본사소 속재택근 무자수	0.076509	0.156792	-0.057534	-0.150393	1.000000	0.571133	-0.031583	0.452293	0.033586	0.065397	-0.352643	-0.026795
연도	-0.078804	0.181749	-0.194792	0.306486	0.571133	1.000000	-0.029843	0.764078	-0.092612	0.129875	-0.109583	-0.010725
일	-0.097392	-0.093606	-0.185565	-0.028433	-0.031583	-0.029843	1.000000	-0.020737	0.002120	-0.009210	0.144007	-0.002428
본사정원 수	-0.115529	0.146544	-0.173852	0.503771	0.452293	0.764078	-0.020737	1.000000	0.244705	0.244666	-0.061688	-0.008577
월	-0.154664	0.018316	-0.127142	0.006330	0.033586	-0.092612	0.002120	0.244705	1.000000	0.193627	0.210877	0.002540
본사휴가 자수	-0.391975	-0.230789	-0.316894	-0.552661	0.065397	0.129875	-0.009210	0.244666	0.193627	1.000000	-0.273318	0.155578
본사출장 자수	-0.512680	-0.308757	-0.188164	0.136389	-0.352643	-0.109583	0.144007	-0.061688	0.210877	-0.273318	1.000000	0.532962
요일	-0.731563	-0.554359	-0.312112	-0.250092	-0.026795	-0.010725	-0.002428	-0.008577	0.002540	0.155578	0.532962	1.000000

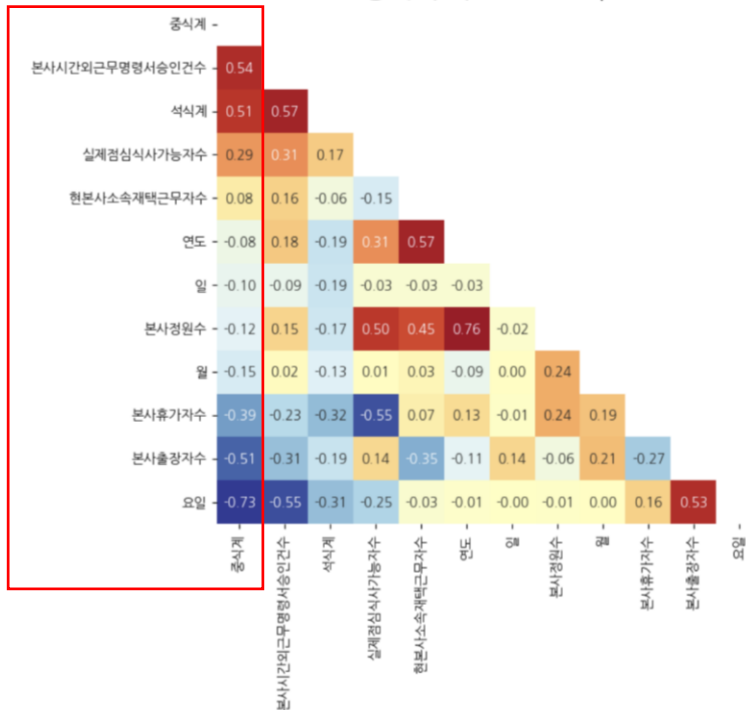
중식기준 피처별 상관관계 데이터 프레임

	석식계	본사시간외근무 명령서승인건수	중식계	실제점심식사 가능자수	현본사소속재 택근무자수	월	본사정원 수	일	본사출장 자수	연도	요일	본사휴가 자수
석식계	1.000000	0.571168	0.508287	0.172373	-0.057534	-0.127142	-0.173852	-0.185565	-0.188164	-0.194792	-0.312112	-0.316894
본사시간 외근무명 령서승인 건수	0.571168	1.000000	0.535611	0.309118	0.156792	0.018316	0.146544	-0.093606	-0.308757	0.181749	-0.554359	-0.230789
중식계	0.508287	0.535611	1.000000	0.286810	0.076509	-0.154664	-0.115529	-0.097392	-0.512680	-0.078804	-0.731563	-0.391975
실제점심 식사가능 자수	0.172373	0.309118	0.286810	1.000000	-0.150393	0.006330	0.503771	-0.028433	0.136389	0.306486	-0.250092	-0.552661
현본사소 속재택근 무자수	-0.057534	0.156792	0.076509	-0.150393	1.000000	0.033586	0.452293	-0.031583	-0.352643	0.571133	-0.026795	0.065397
월	-0.127142	0.018316	-0.154664	0.006330	0.033586	1.000000	0.244705	0.002120	0.210877	-0.092612	0.002540	0.193627
본사정원 수	-0.173852	0.146544	-0.115529	0.503771	0.452293	0.244705	1.000000	-0.020737	-0.061688	0.764078	-0.008577	0.244666
일	-0.185565	-0.093606	-0.097392	-0.028433	-0.031583	0.002120	-0.020737	1.000000	0.144007	-0.029843	-0.002428	-0.009210
본사출장 자수	-0.188164	-0.308757	-0.512680	0.136389	-0.352643	0.210877	-0.061688	0.144007	1.000000	-0.109583	0.532962	-0.273318
연도	-0.194792	0.181749	-0.078804	0.306486	0.571133	-0.092612	0.764078	-0.029843	-0.109583	1.000000	-0.010725	0.129875
요일	-0.312112	-0.554359	-0.731563	-0.250092	-0.026795	0.002540	-0.008577	-0.002428	0.532962	-0.010725	1.000000	0.155578
본사휴가 자수	-0.316894	-0.230789	-0.391975	-0.552661	0.065397	0.193627	0.244666	-0.009210	-0.273318	0.129875	0.155578	1.000000

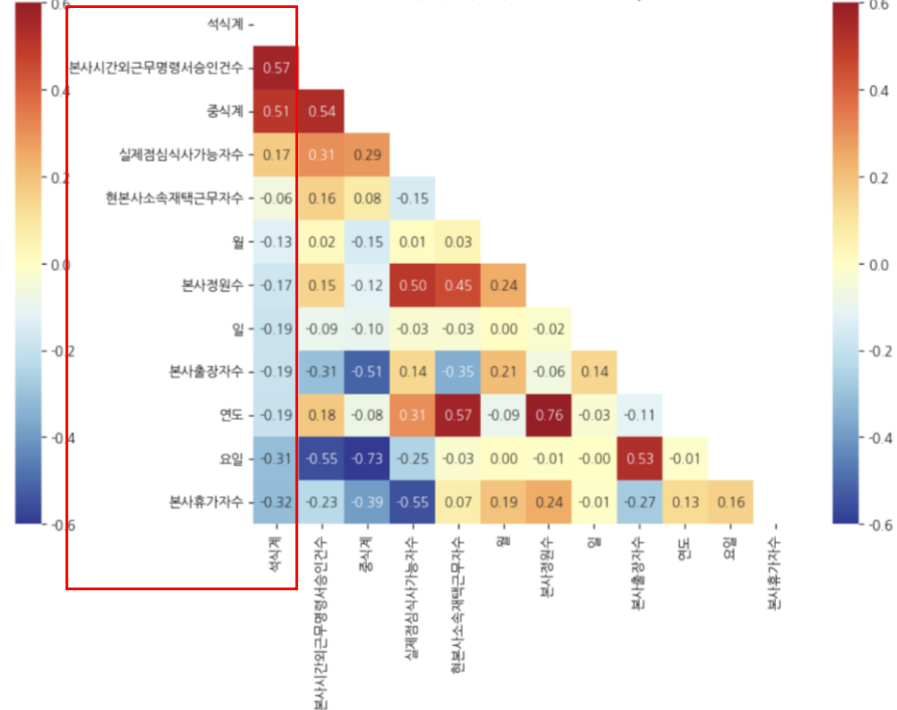
석식기준 피처별 상관관계 데이터 프레임



중식계 기준 heatmap



석식계 기준 heatmap



```
Out[37]: 종식계      1.000000
본사시간외근무명령서승인건수  0.535611
식식계      0.508287
실제점식식사가가능자수      0.286810
현본사소속채택근무자수      0.076509
연도      -0.078804
일      -0.097392
본사정원수      -0.115529
월      -0.154664
본사휴가자수      -0.391975
본사출장자수      -0.512680
요일      -0.731563
Name: 종식계, dtype: float64
```

```

Out[38]: 석식계      1.000000
본사시간외근무명령서승인건수  0.571168
중식계      0.508287
심재점삼식사가가능자수      0.172373
현본사소속재택근무자수      -0.057534
월      -0.127142
본사정원수      -0.173852
일      -0.185565
본사출장자수      -0.188164
연도      -0.194792
요일      -0.312112
본사휴가자수      -0.316894
Name: 석식계, dtype: float64

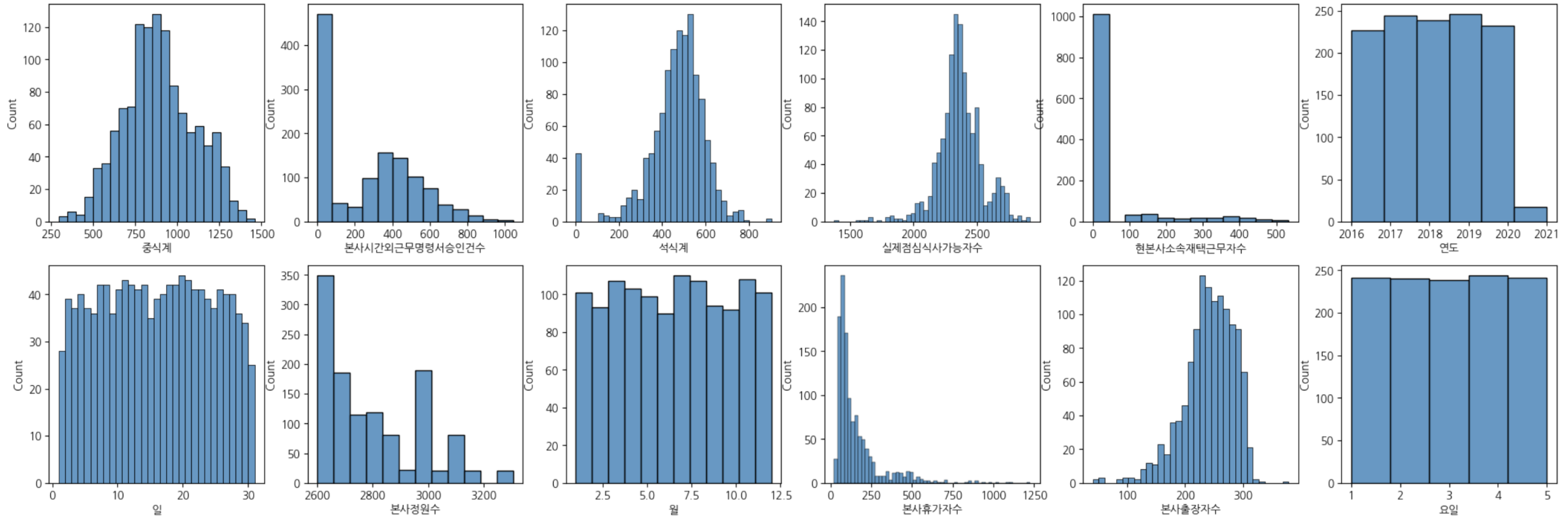
```

```
In [39]: # 인트만_sort 를 절대값으로 취하여 우선순위를 알아내고싶다.. ㅋㅋㅋ π
```

- 절대값 수치가 1에 가까울 수록 강한 상관관계를 보인다.
- 히트맵의 가장 왼쪽 컬럼 데이터의 순서대로 강한 상관관계를 보이도록 plot하였다.



Feature별 데이터 분포 시각화 & 분석



- 중식계, 석식계, 실제 점심식사 가능자수 그래프가 정규분포에 가까운 데이터 분포를 보였다.
- 본사시간외 근무명령서 승인건수, 본사정원수, 본사 휴가자수, 본사출장자수 그래프는 치우친 형태의 데이터 분포를 보였다.
- 연도, 일, 월, 요일 데이터의 분포는 유의미한 결과를 얻기 힘들었다.

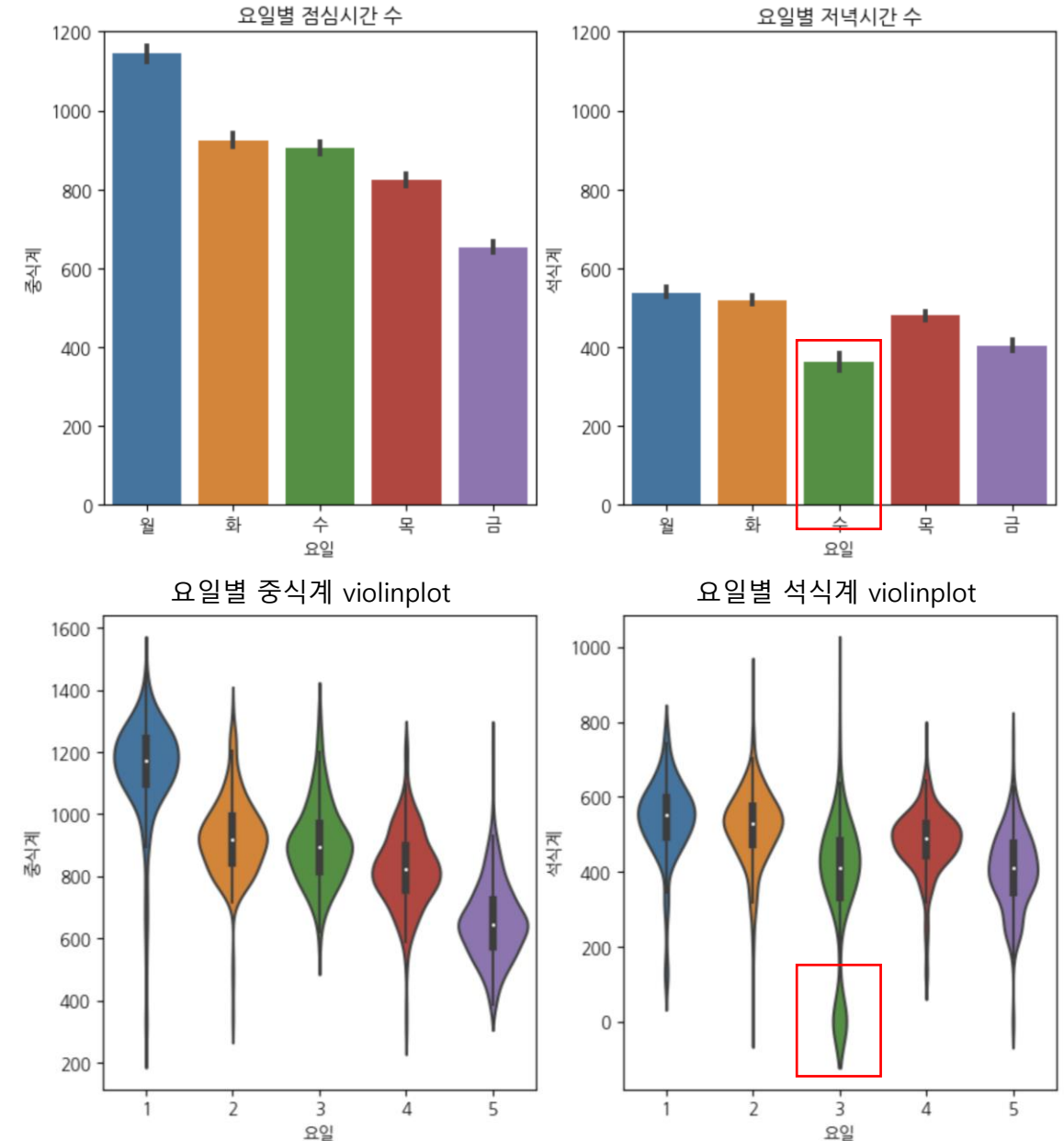


요일별 중식-석식(식사자수)분포 시각화 & 분석

요일 별 중식-석식 상관관계가 있을 것이다.

요일별 식사자 수 (Barplot, Violinplot)

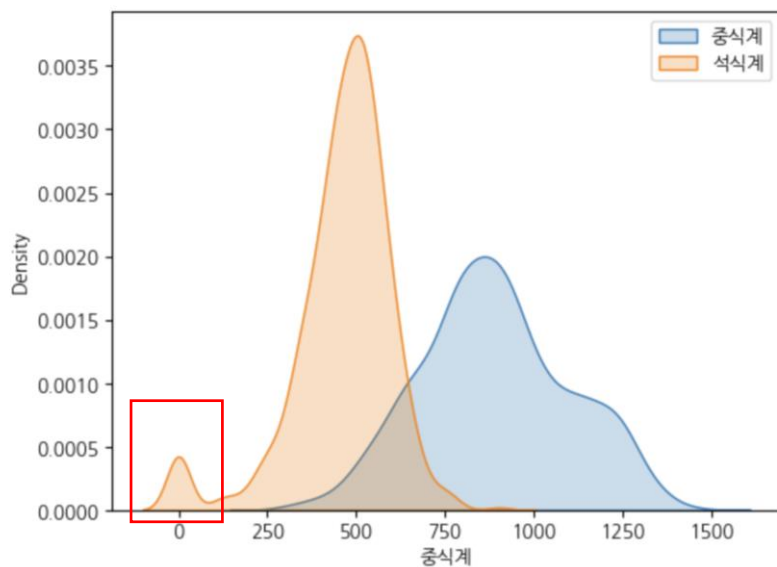
- Bar chart를 통해 중식식사자수 분포는 월요일부터 금요일 까지 순차적으로 값이 떨어지는 결과값을 확인할 수 있었고, 석식식사자도 비슷한 분포를 보였으나 수요일에 식사자가 줄어드는 중식과 다른 분포를 보였다.
- Violinplot chart를 통해 식사자수의 값 분포를 요일별로 확인할 수 있었고, 석식 수요일의 값 분포가 0에 근접한 것으로 보아 이상치가 있음을 감지할 수 있었다.



중식-석식(식사자수) 데이터 분포 시각화 & 분석

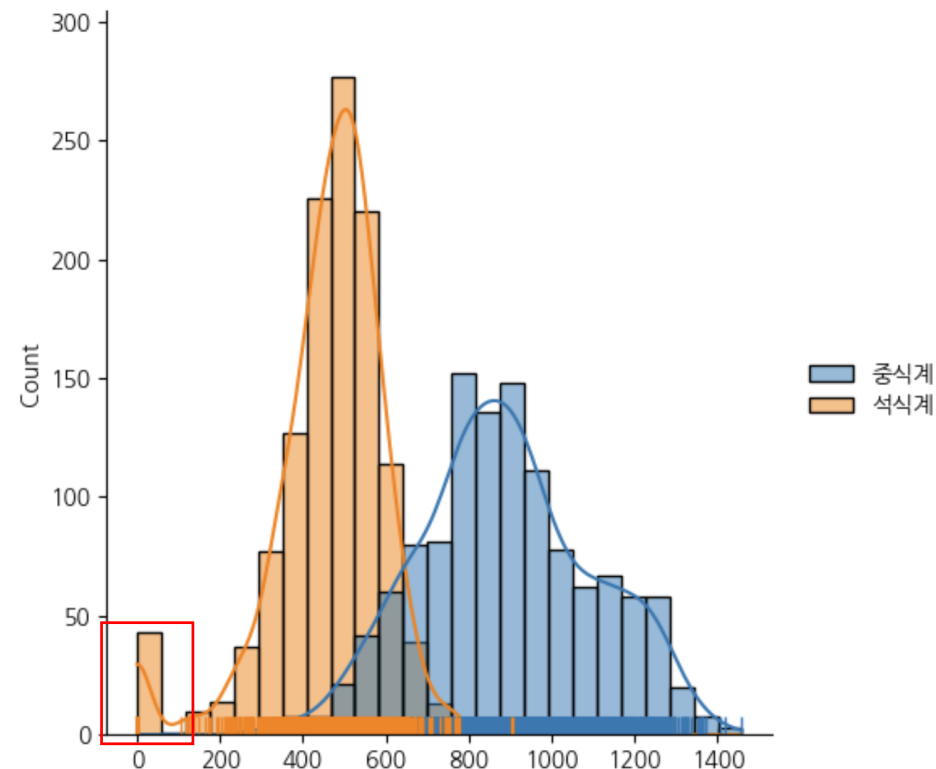
kdeplot, distplot을 통한 식사자수 정규분포도 확인

```
sns.kdeplot(df['중식계'], label='중식계', shade=True)
sns.kdeplot(df['석식계'], label='석식계', shade=True)
plt.legend()
plt.show()
```



```
In [56]: sns.displot(data=df[['중식계', '석식계']], kde=True, rug=True)
# sns.displot(data=df['석식계'], kde=True, color='blue')
```

```
Out[56]: <seaborn.axisgrid.FacetGrid at 0x1fde3fb7d00>
```



- Kdeplot과 distplot을 통해 중식 석식 데이터 분포 그래프를 한 차트에 표현하였으며, kde플롯의 y축은 데이터의 밀도, distplot의 y축은 데이터 값을 Count 해준다.
- 중식자수가 석식자수보다 많은 것을 볼 수 있고, 석식자 그래프에서 x축 0값에 그래프가 튀어나온 모양으로 보았을 때 이상치가 있음을 탐지할 수 있었다.



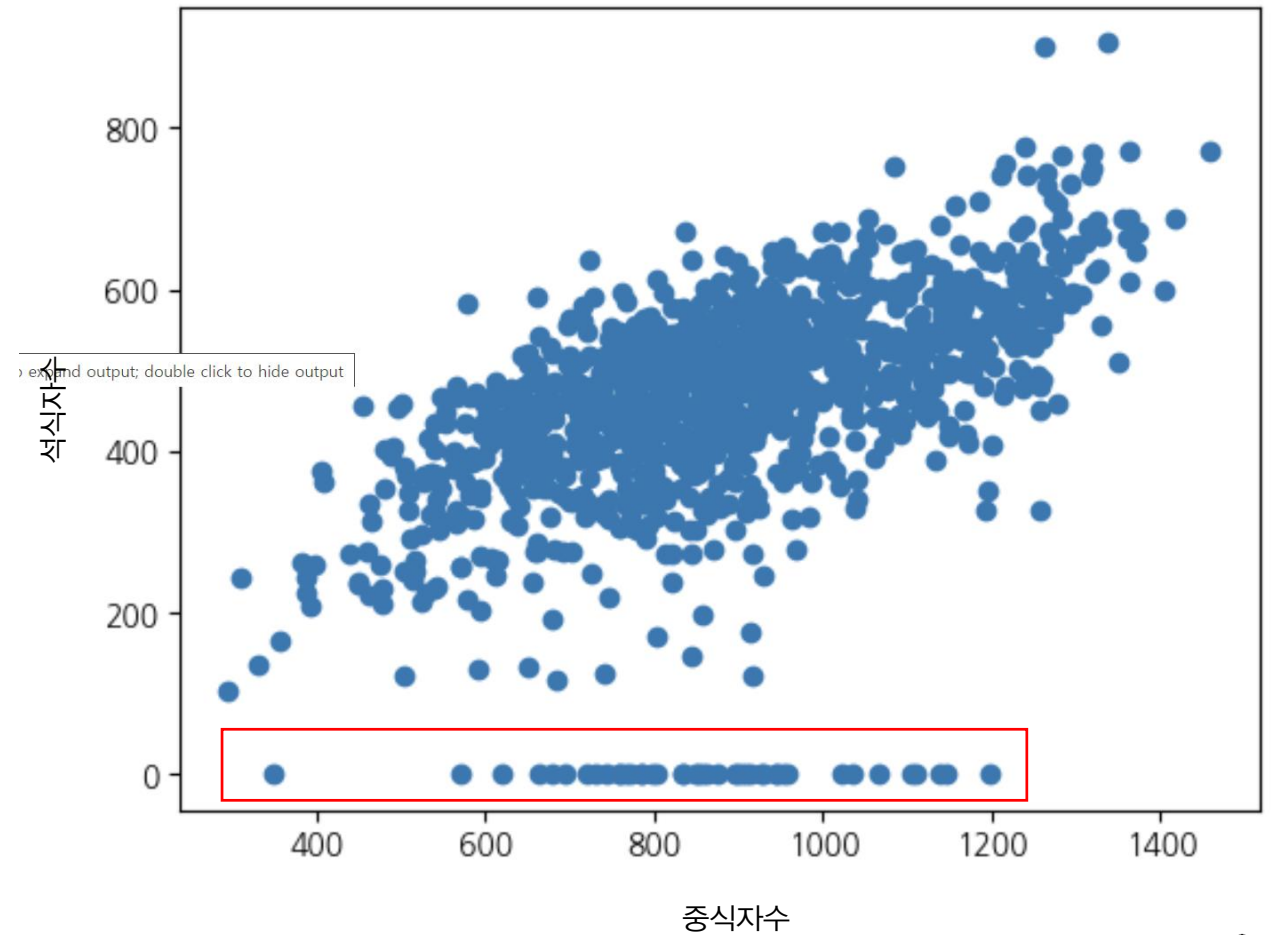
중식-석식간 상관관계 분석

점심을 먹은 사람은 저녁도 먹을까?

scatterplot을 통한 상관관계 확인

- Scatterplot을 통해 중식식사자와 석식 식사자 간 양의 상관관계를 발견할 수 있었다.
- 0 근처값의 분포를 통해 이상치를 탐지해볼 수 있었다.

```
plt.scatter(df['중식계'],df['석식계'])  
plt.show()
```



Ideation단계 vs 검증된 가설

요일 별 중식-석식 상관관계가 있을 것이다.

점심을 먹은 사람은 저녁도 먹을까?

중식 석식 과 피쳐별 상관관계를 통한 데이터 전처리 계획 수립

- 본사시간 외 근무명령서승인건수 수치로 보아 야근자수가 많을 수록 중식, 석식 식사를 할 확률이 높았다.
- 요일과 본사출장자수, 본사 휴가자수가 많을 수록 중식, 석식 식사를 안할 확률이 높았다.
- 현본사소속재택근무자수 값을 볼 때 초기 가설설정과는 다르게 재택근무자수는 중석식 식사자수에 영향을 크게 미치지 않는다는 것을 확인하였다.
- 크게 영향을 미치지 않는 데이터들은 예측결과값 향상을 위해 전처리 계획을 수립하였다.

```
In [37]: ▶ 인트만_sort = 인트만.corr()["중식계"].sort_values(ascending=False)
인트만_sort
```

```
Out[37]: 중식계 1.000000
본사시간외근무명령서승인건수 0.535611
석식계 0.508287
실제점심식사가능자수 0.286810
현본사소속재택근무자수 0.076509
연도 -0.078804
일 -0.097392
본사정원수 -0.115529
월 -0.154664
본사휴가자수 -0.391975
본사출장자수 -0.512680
요일 -0.731563
Name: 중식계, dtype: float64
```

```
In [38]: ▶ 인트만_sort = 인트만.corr()["석식계"].sort_values(ascending=False)
인트만_sort
```

```
Out[38]: 석식계 1.000000
본사시간외근무명령서승인건수 0.571168
중식계 0.508287
실제점심식사가능자수 0.172373
현본사소속재택근무자수 -0.057534
월 -0.127142
본사정원수 -0.173852
일 -0.185565
본사출장자수 -0.188164
연도 -0.194792
요일 -0.312112
본사휴가자수 -0.316894
Name: 석식계, dtype: float64
```

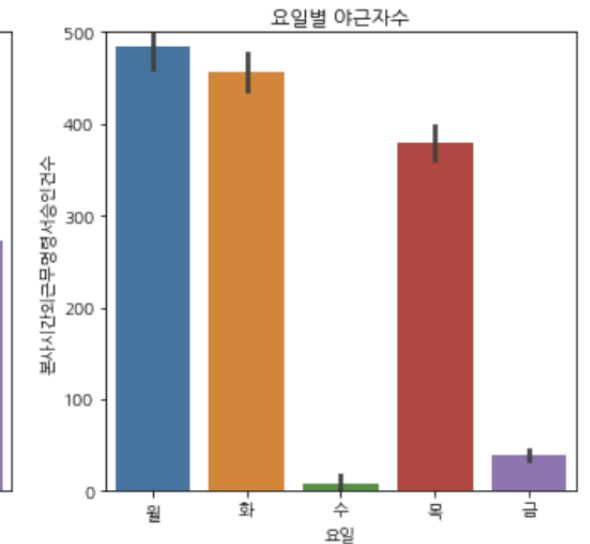
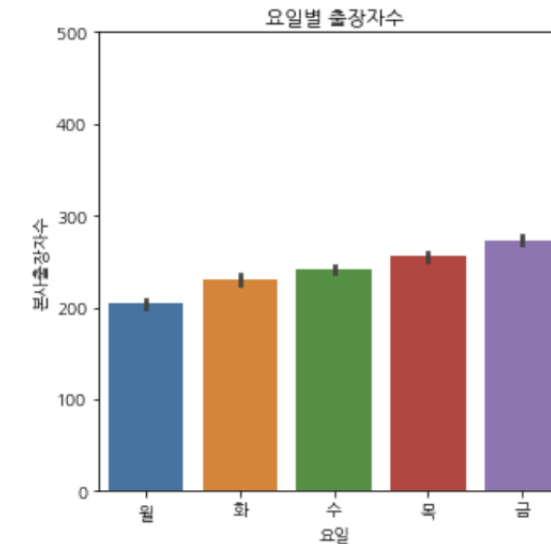
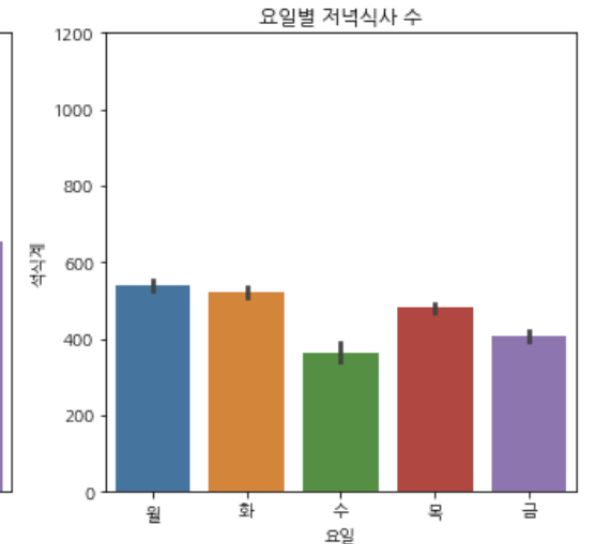
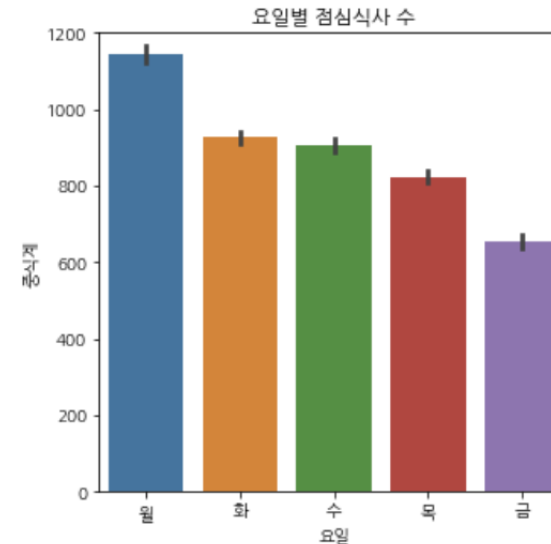
```
In [39]: ▶ # 인트만_sort 를 절대값으로 취하여 우선순위를 알아내고싶다.. ㅋㅋㅋ π
```



요일별 출장자수 및 야근자수 시각화 & 분석

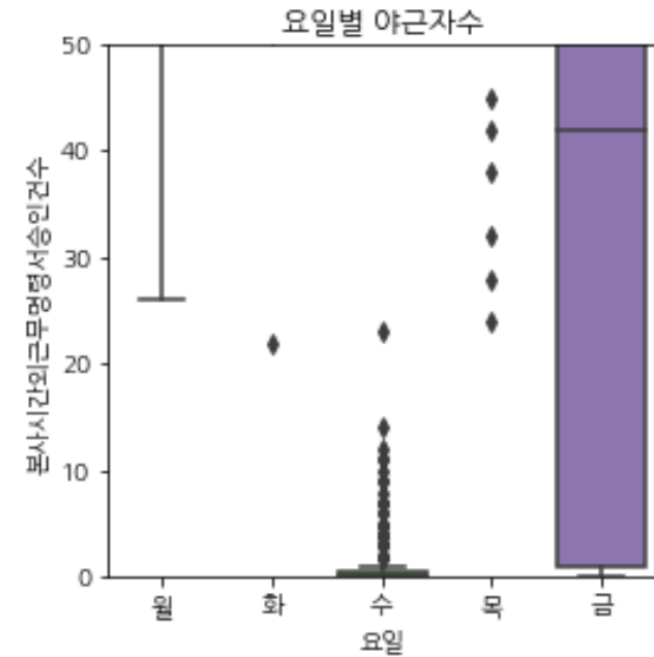
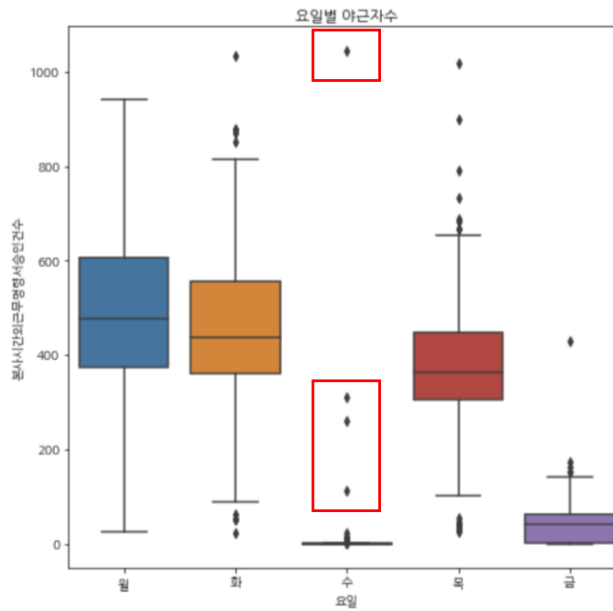
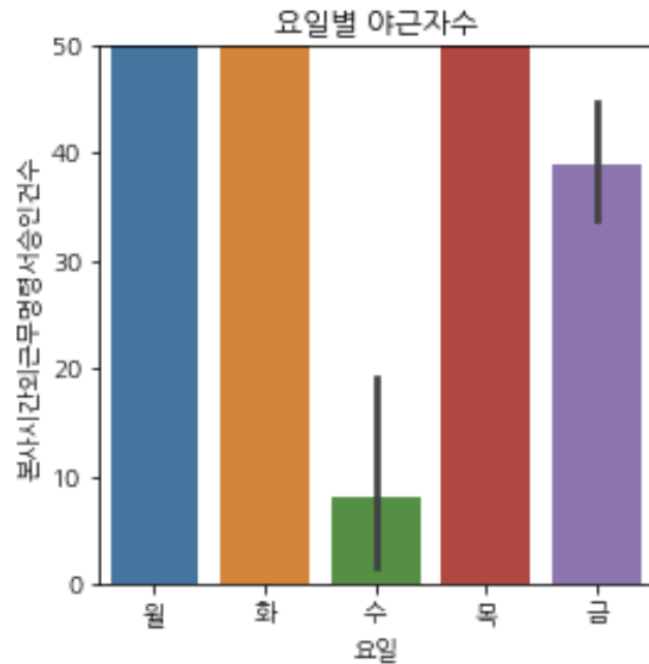
요일별 출장 및 야근자수 심층 분석(Barplot)

- 지난 슬라이드의 결과로 출장자수, 야근자수와 요일이 식수에 큰 영향을 미친다는 결과를 알 수 있었다.
- 심층적으로 요일별 출장자와 야근자수를 분석해보기 위해 Barplot을 통해 plot하였다.
- 출장자수는 일주일 내 비슷한 값들 사이에서도 월-금요일로 갈수록 많아지는 분포도를 보였다.
- 요일별 야근자수를 보았을 때 수요일과 금요일이 현저히 적은 것을 확인할 수 있었고, 수요일에는 어떤 관계가 있는지 심층적으로 분석해보았다.



요일별 출장자수 및 야근자수 시각화 & 심층 분석

요일별 야근자수 Boxplot



- 수요일에 야근하는 비율은 평균적으로 7~8명이다.
- 수요일에 야근하는 수는 중간값 0명에 가깝다.
- 1000명 넘게 야근한 하루의 이상치로 인해 평균이 7~8명으로 상승한 것으로 보인다.



요일별 야근자수 원본데이터

```

▶ # 원본 데이터를 살펴보자!!!
# pd.set_option('display.max_rows', None)
df.sort_values('석식계', ascending=True).head(50)

#석식메뉴가 비어있거나 메뉴가 써 있지 않으면, 저녁엔 국내식당이 문을 닫는다.

```

912	2019-10-30	수	3121	122	294	1	0.0	모닝롤/오렌지케익빵 우유/주스 계란후라이 누룽지탕/쌀밥 (쌀:국 내산) 황태국 (황...	마파두부덮밥 (쌀,돈육: 국내산) 맑은국 치킨핑 거 (닭:국내산) 시저샐 러드 무비트...	*	771.0	0.0	2705.0
828	2019-06-26	수	2758	69	282	0	0.0	모닝롤/고구마파이 우 유/주스 계란후라이 누 룽지탕/쌀밥 (쌀:국내 산) 콩가루배추국 ...	카레덮밥 (쌀,돈육:국내 산) 가쓰오장국 고구마 치즈구이 쫄면무침 무 말랭이 포기...	* 자기개발의날 *	946.0	0.0	2407.0
993	2020-02-26	수	2872	109	190	4	0.0	모닝롤/고구마파이 우 유/주스 계란후라이 누 룽지탕/쌀밥 (쌀:국내 산) 홍합탕 감자...	낙지비빔밥 (쌀:국내,낙 지:중국산) 팽이장국 치 킨텐더*콘소스D (닭:국 내산) 과일...	자기개발의날	1105.0	0.0	2573.0
549	2018-04-25	수	2714	66	285	0	0.0	모닝롤/롤케익 우유/주 스 계란후라이 누룽지 탕/쌀밥 (쌀:국내산) 아 욱된장국 감자...	비빔밥 (쌀:국내산) 유 부장국 오징어튀김 떡 밤초 요플레 포기김치 (김치:국내산)		851.0	0.0	2363.0
2040								모닝롤/야채샌드 우유/ 주스 스무디/라볶이 국	쌀밥/잡곡밥 (쌀:국내 산) 배춧단장국 다들...				

- 원본데이터를 확인하자 석식이 없는 수요일에는 '자기개발의 날' 이 포함되어 있는 것을 확인하였다.



Jupyter Notebook

Code Review



- 프로젝트의 목적은 구내식당의 식수를 예측하여 남은 잔반을 최소화 하는 것이었다.
- 변수별 상관관계 분석을 위하여 데이터 전처리를 하는 과정을 거쳤다.
- 다양한 시각화 방법을 이용하여 식수에 영향을 미치는 Feature들을 검증하였다.
- 시각화를 통해 이상치를 탐지하였고, 세부 시각화와 데이터 인덱싱을 통해 이상치를 검증해 내었다.
- 요일, 본사출장자수, 본사휴가자수 Feature들이 식수에 가장 큰 영향을 미치는 Feature로 검증되었다.
- 위 Feature들에 영향을 미치는 세부 상관관계를 시각화하여 출장자수, 야근자수와 요일의 세부 상관관계를 검증하였다.
- 출장자수, 휴가자수가 많을 수록 식수가 적을 확률이 높았다.
- 요일데이터의 세부 분석을 통해 수요일에 석식이 없는 경우를 확인하였으며 요일 Feature에서 ‘자기계발의 날’ 과같은 특수한 요일을 선별해 내었다.
- 위 과정과 크게 영향을 미치지 않는 Feature값들을 데이터 전처리 과정에 반영하였다.
- Dacon사이트에서 요구하였던 Random Forest모델을 이용하여 MSE값을 도출해 내었으며, 다른 모델로도 MSE값을 도출해보았다.
- 더 나은 결과값을 도출해내기위해 하이퍼파라미터 튜닝을 시도하였고, 시간이 많이 드는 GridsearchCV보다는 RandomsearchCV를 통해 최적의 하이퍼 파라미터값을 찾아내었다.
- Xgboost를 이용해서 결과값을 도출해보았다.
- 데이터 전처리 과정에서 보였던 그래프의 정규화 정도를 높이기 위해 로그화를 진행하여 모델을 학습시켜보았다.
- 결과적으로 Xgboost모델에서 최적의 MSE 값을 도출하였으며, 데이콘에서 Private : 112.7960591429 Public : 81.013793의 Score를 기록하였다.
- Randomforest 모델 시각화를 마지막으로 진행하였다.



데이터 자료

링크

<https://dacon.io/competitions/official/235743/overview/description>



4조

담당 퍼실 : 준아님

정현범 – 발표, 자료검색, 코드 검색

이상준 – 자료검색, 코드 구현

신종우 – 자료검색, 코드 검색, 발표 자료

