독립변수 '요일', '공휴일전후', '공휴일길이', '출근', '휴가비율', '출장비율', '야근비율', '재택비율', 'week', ‘year’, ‘month’, ‘day’, '강수량', '불쾌지수', '체감온도', '요일점심평균', '요일저녁평균', '월점심평균', '월저녁평균', '전일대비증감', 'breakfast\_0', 'breakfast\_1', 'breakfast\_2', 'lunch\_0', 'lunch\_1', 'lunch\_2', 'dinner\_0', 'dinner\_1', 'dinner\_2', 'oov\_cnt.b', 'oov\_cnt.l', 'oov\_cnt.d' 에 대해

1. 상관관계 분석
2. VIF 분석

진행했습니다.

1. 상관관계 분석

* 가장 높은 상관관계를 보인 순서

week와 day (+0.975)

체감 온도와 불쾌지수(+0.965)

요일 저녁 평균과 야근 비율(+0.831)

요일 점심 평균과 요일 (+0.769)

월 저녁 평균과 월 점심 평균 (+0.745)

월 점심 평균과 month (-0.687)

요일 저녁 평균과 요일 (+0.611)

휴가 비율과 출근 (-0.603)

year와 재택비율 (+0.571)

…

첨부파일 result.xlsx에 모든 결과가 나와있습니다. 위쪽 테이블은 correlation 원본값, 그 아래 테이블은 절대값이에요!

1. VIF 분석

아래는 VIF가 10 이상인 특성들입니다. 나머지 값 역시 result.xlsx에서 확인할 수 있습니다.

|  |  |  |
| --- | --- | --- |
|  | **VIF** | **features** |
| **14** | 118.5423 | 체감온도 |
| **10** | 92.91604 | day |
| **13** | 85.18913 | 불쾌지수 |
| **11** | 65.78397 | week |
| **3** | 59.23144 | 출근 |
| **5** | 28.22035 | 출장비율 |
| **16** | 18.70446 | 요일저녁평균 |
| **15** | 17.72211 | 요일점심평균 |
| **17** | 14.40999 | 월점심평균 |
| **18** | 13.31034 | 월저녁평균 |
| **9** | 11.75255 | month |
| **6** | 10.87064 | 야근비율 |
| **0** | 10.32045 | 요일 |