

[Spring 2023] DATA MINING THEORY AND APPLICATION (IIE 4102)

Practice-HW2

Principal Component Analysis / Factor Analysis

논문참조

강세정, 김규현, 신재욱, 장현우



01

Introduction

Research Objective
Data



02

Body

PCA
FA



03

Conclusion

Result
Limitation & Development

Introduction

Contents

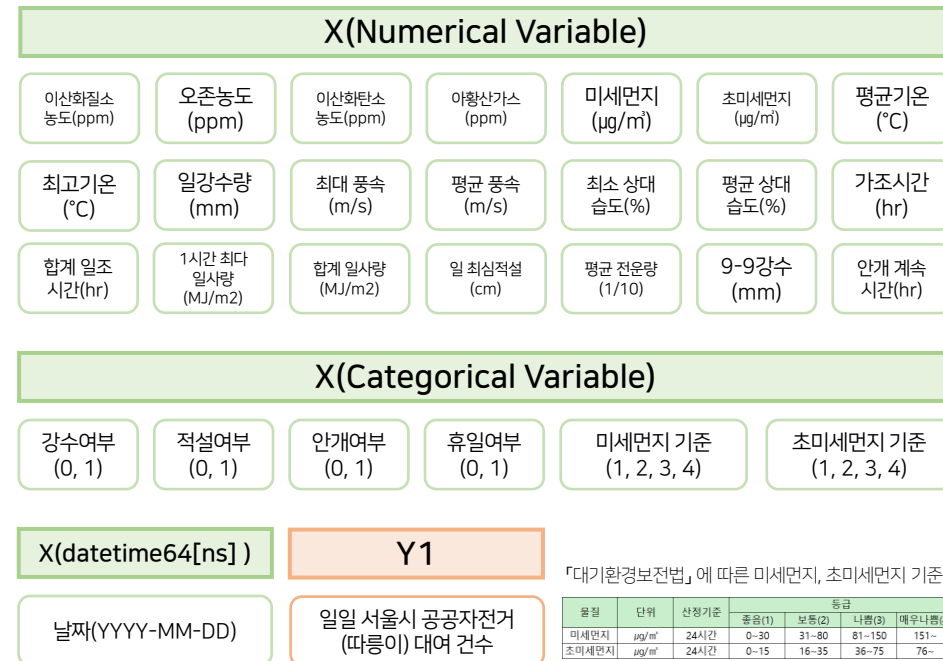
PCA / FA

01. Research Objective

- PCA와 FA를 모두 사용하여 날씨와 관련된 변수들의 영향을 조사하고, 이를 공유자전거 대여량 예측 모델에 반영하여 높은 예측 정확도를 달성하는 것이 목표
- 이를 통해, 효율적인 공유자전거 운영을 위한 정책적 제언을 제공할 수 있을 것으로 예상됨

02. Data

저번 주 Variable Selection에서 성능이 좋았던 따름이 예측 관련 데이터 사용



① data.head()

| 날짜 | 이산화질소 농도 (ppm) | 오존 농도 (ppm) | 이산화탄소 농도 (ppm) | 아황산가스 농도 (ppm) | 미세먼지 농도 (μg/m ³) | 초미세먼지 농도 (μg/m ³) | 평균기온 (℃) | 최고기온 (℃) | 일강수량 (mm) | 최대 풍속 (m/s) | 평균 풍속 (m/s) | 최소 상대 습도 (%) | 평균 상대 습도 (%) | 가조시간 (hr) | 합계 일조 시간(hr) | 1시간 최다 일사량 (MJ/m ²) | 합계 일사량 (MJ/m ²) | 일 최심적설 (cm) | 평균 전운량 (1/10) | 9-9강수 (mm) | 안개 계속 시간(hr) | | | | | | |
|------------|----------------------|-------------------|----------------------|----------------------|------------------------------------|-------------------------------------|-------------|-------------|--------------|-------------------|-------------------|--------------------|--------------------|--------------|-----------------|---------------------------------------|--------------------------------|----------------|------------------|---------------|-----------------|---|---|---|---|---|-------|
| 2019-01-01 | 0.033 | 0.016 | 0.5 | 0.003 | 42 | 27 | -0.5 | -0.6 | 0.0 | 4.3 | 2.1 | 34 | 48.0 | 9.6 | 7.5 | 1.42 | 7.56 | 0.0 | 5.4 | 0.0 | 0.0 | 2 | 2 | 0 | 0 | 1 | 3607 |
| 2019-01-02 | 0.043 | 0.010 | 0.6 | 0.004 | 36 | 24 | -4.9 | -5.2 | 0.0 | 3.8 | 1.7 | 20 | 40.0 | 9.6 | 9.7 | 1.81 | 10.48 | 0.0 | 9.0 | 0.0 | 0.0 | 2 | 2 | 0 | 0 | 0 | 14040 |
| 2019-01-03 | 0.049 | 0.008 | 0.6 | 0.004 | 33 | 22 | -3.3 | -3.2 | 0.0 | 2.9 | 1.4 | 19 | 38.0 | 9.7 | 9.7 | 1.79 | 10.28 | 0.0 | 9.1 | 0.0 | 0.0 | 2 | 2 | 0 | 0 | 0 | 10344 |
| 2019-01-04 | 0.065 | 0.005 | 1.0 | 0.005 | 35 | 42 | -1.1 | -1.1 | 0.0 | 3.0 | 1.4 | 18 | 35.0 | 9.7 | 9.9 | 1.27 | 6.26 | 0.0 | 9.0 | 0.0 | 0.0 | 2 | 2 | 0 | 0 | 0 | 17300 |
| 2019-01-05 | 0.034 | 0.017 | 0.6 | 0.004 | 40 | 41 | -2.8 | -3.1 | 0.0 | 4.3 | 2.2 | 20 | 40.0 | 9.7 | 9.6 | 1.76 | 10.09 | 0.0 | 9.0 | 0.0 | 0.0 | 2 | 2 | 0 | 0 | 1 | 12000 |

③ data.describe()

| | 이산화질소 NO2 | 오존 O3 | 이산화탄소 CO2 | 아황산가스 SO2 | 미세먼지 PM10 | 초미세먼지 PM2.5 | 평균기온 Tavg | 최고기온 Tmax | 일강수량 Precip | 최대 풍속 Vmax | 평균 풍속 Vavg | 최소 상대 습도 RHmin | 평균 상대 습도 RHavg | 가조시간 HDD | 합계 일조 시간 Hsum | 1시간 최다 일사량 I1max | 합계 일사량 Isum | 일 최심적설 Dmax | 평균 전운량 Tavg | 9-9강수 R99 | 안개 계속 시간 FogDur |
|-------|--------------|----------|--------------|--------------|--------------|----------------|--------------|--------------|----------------|---------------|---------------|-------------------|-------------------|-------------|------------------|---------------------|----------------|----------------|----------------|--------------|--------------------|
| count | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 |
| mean | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| std | 0.017 | 0.007 | 174.44 | 0.001 | 11.03 | 7.03 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -5.00 | -5.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| max | 0.100 | 0.020 | 1000.00 | 0.010 | 100.00 | 50.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| sum | 14.15 | 5.21 | 137074.68 | 1.46 | 12066.05 | 8066.56 | -630.00 | 6090.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 | 365.00 |
| avg | 0.039 | 0.014 | 374.74 | 0.004 | 33.03 | 22.07 | -1.70 | 16.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00</ | | | | |

④ Z 표준화 후 평균

| | |
|--------------------------------|--------------|
| 이산화질소 농도(ppm) | 2.75369e-16 |
| 오존농도(ppm) | -3.59046e-16 |
| 이산화탄소 농도(ppm) | 2.33603e-16 |
| 아황산가스(ppm) | 0.00000e+00 |
| 미세먼지(μg/m ³) | 5.60077e-17 |
| 초미세먼지(μg/m ³) | -7.73677e-17 |
| 평균기온(°C) | -1.18091e-16 |
| 최고기온(°C) | -7.73677e-17 |
| 일강수량(mm) | 4.68973e-17 |
| 최대 풍속(m/s) | 7.59210e-16 |
| 평균 풍속(m/s) | 2.53700e-16 |
| 최소 상대 습도(%) | 0.00000e+00 |
| 평균 상대 습도(%) | -7.76770e-17 |
| 가조시간(hr) | 9.344124e-16 |
| 합계 일조 시간(hr) | -1.28754e-16 |
| 1시간 최다 일사량(MJ/m ²) | 3.89335e-17 |
| 합계 일사량(MJ/m ²) | 0.00000e+00 |
| 일 최심적설(cm) | 1.94692e-17 |
| 평균 전운량(1/10) | 9.73403e-17 |
| 9-9강수(mm) | 3.89335e-17 |
| 안개 계속 시간(hr) | 1.460019e-17 |
| 오존농도(ppm) | float64 |

⑤ 공분산 행렬

| | | | | | | | | | | | | | | | | | | | |
|---------------|---------------|------------|--------------------------|---------------------------|----------|----------|----------|------------|------------|-------------|-------------|----------|--------------|--------------------------------|----------------------------|------------|--------------|-----------|--------------|
| 이산화질소 농도(ppm) | 이산화탄소 농도(ppm) | 아황산가스(ppm) | 미세먼지(μg/m ³) | 초미세먼지(μg/m ³) | 평균기온(°C) | 최고기온(°C) | 일강수량(mm) | 최대 풍속(m/s) | 평균 풍속(m/s) | 최소 상대 습도(%) | 평균 상대 습도(%) | 가조시간(hr) | 합계 일조 시간(hr) | 1시간 최다 일사량(MJ/m ²) | 합계 일사량(MJ/m ²) | 일 최심적설(cm) | 평균 전운량(1/10) | 9-9강수(mm) | 안개 계속 시간(hr) |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

⑥ data.info()

| | |
|---|-------------------------------|
| RangeIndex: 365 entries, 0 to 364 | |
| Data columns (total 23 columns): | |
| # | Column |
| 0 | 날짜 |
| 1 | 이산화질소농도(ppm) |
| 2 | 오존농도(ppm) |
| 3 | 이산화탄소농도(ppm) |
| 4 | 아황산가스(ppm) |
| 5 | 미세먼지(μg/m ³) |
| 6 | 초미세먼지(μg/m ³) |
| 7 | 평균기온(°C) |
| 8 | 최고기온(°C) |
| 9 | 일강수량(mm) |
| 10 | 최대 풍속(m/s) |
| 11 | 평균 풍속(m/s) |
| 12 | 최소 상대습도(%) |
| 13 | 평균 상대습도(%) |
| 14 | 가조시간(hr) |
| 15 | 합계 일조시간(hr) |
| 16 | 1시간 최다일사량(MJ/m ²) |
| 17 | 합계 일사량(MJ/m ²) |
| 18 | 일 최심적설(cm) |
| 19 | 평균 전운량(1/10) |
| 20 | 9-9강수(mm) |
| 21 | 안개 계속시간(hr) |
| 22 | 미세먼지 기준 |
| 23 | 초미세먼지 기준 |
| 24 | 강수여부 |
| 25 | 적설여부 |
| 26 | 안개여부 |
| 27 | 휴일여부 |
| 28 | 타종이 |
| dtypes: datetime64[ns](1), float64(18), int64(10) | |

데이터를 서울 열린 데이터 광장, 공공데이터포털, 기상청 등에서 직접 수집함
데이터 기간 : 2019.01.01 ~ 2019.12.31
처음에는 2016-2019 데이터를 사용해 프로젝트를 진행했으나, 성능이 좋지 않았음
→ 시계열 데이터의 특성을 고려해 2019년 만의 데이터를 이용하여 성능 개선

※ 추가 설명

- 진행하려는 연구와 관련된 이론적인 배경·선행 연구 사례를 찾기 위해 논문을 참고해 어떠한 변수들이 고려되어야 하는지, 어떠한 방법론을 사용을 했는지 검토하고, 이를 기반으로 데이터에 대한 전반적인 기초통계량을 확인하는 과정을 거침
(김동진 신희철 et al. 박준식 임형준 "날씨가 자전거 이용에 미치는 영향 분석 - 고양시 공공자전거를 대상으로 -" 교통연구 19.3 pp. 77-88 (2012): 77.)

Body

Contents

PCA / FA

PCA

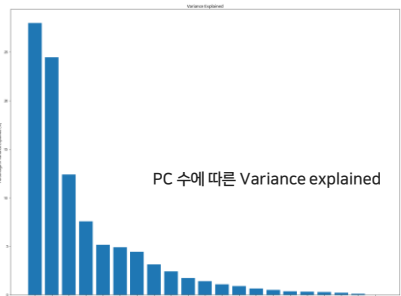
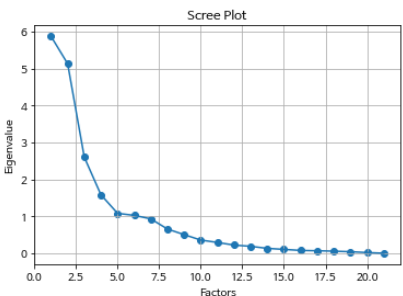
01. PCA의 factor 수 결정

① 방안 1. Eigen value > 1일 때까지 선택 : 6개

② 방안 2. Scree plot 그래프 기울기가 완만해지기 전의 값까지 선택 : 5개

- 5번째 factor에서 6번째 factor로 갈 때 기울기가 완만해짐

- 따라서 기울기가 완만해지기 전인 5번째 factor까지 사용



③ 방안 3. 약 80% 이상의 분산을 설명하는 PC 건수 채택 : 6개

▶ 그래프의 기울기가 완만해지는 지점은 5번째 factor이나, 6번째 factor까지 eigenvalue가 1 이상의 값을 가지며 또한 6번째 factor까지 포함했을 때 설명 가능한 분산의 비율이 0.8 이상이 되므로 factor 수 = 6개로 설정

누적 분산 설명 비율:

[0.27989592 0.52443275 0.64847754 0.72401527 0.77560572 0.82457398
0.86900646 0.90035299 0.92448992 0.94161838 0.95566771 0.96633211
0.97539198 0.98178089 0.98688991 0.99063253 0.99397208 0.9967524
0.99886972 0.99986106 1.]

02. PCA 진행

① PCA 진행

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| 아산화질소농도(ppm) | 0.346618 | 0.062099 | -0.197529 | -0.089846 | -0.129862 | -0.016841 |
| 오존농도(ppm) | -0.179436 | -0.251962 | -0.115588 | 0.365387 | 0.174302 | -0.024036 |
| 아산화탄소농도(ppm) | 0.348779 | 0.069645 | -0.223789 | 0.082868 | -0.034904 | 0.080598 |
| 아황산가스(ppm) | 0.276626 | -0.090698 | -0.182169 | 0.203521 | -0.103159 | -0.037892 |
| 미세먼지(μg/m³) | 0.283448 | -0.013585 | -0.243908 | 0.403272 | 0.051624 | 0.052131 |
| 초미세먼지(μg/m³) | 0.269638 | 0.021873 | -0.295821 | 0.385436 | 0.082993 | 0.055959 |
| 평균기온(°C) | -0.339944 | -0.050321 | -0.280908 | 0.053786 | 0.019117 | 0.037544 |
| 최고기온(°C) | -0.322360 | -0.096698 | -0.296479 | 0.040739 | -0.006377 | 0.049856 |
| 일강수량(mm) | -0.151558 | 0.236310 | -0.002519 | 0.148141 | -0.480592 | -0.050012 |
| 최대 풍속(m/s) | -0.029676 | -0.069268 | 0.432503 | 0.472617 | 0.028518 | 0.051502 |
| 평균 풍속(m/s) | -0.054174 | -0.032440 | 0.481466 | 0.390777 | 0.081496 | 0.090424 |
| 최소 상대습도(%) | -0.204510 | 0.325643 | -0.097950 | 0.063967 | 0.055686 | 0.100269 |
| 평균 상대습도(%) | -0.199349 | 0.306431 | -0.121444 | 0.062362 | 0.128271 | 0.116443 |
| 가조시간(hr) | -0.321821 | -0.124440 | -0.253489 | 0.133622 | 0.070886 | -0.058629 |
| 한계 일조시간(hr) | 0.024418 | -0.400196 | 0.024927 | -0.083608 | -0.179506 | 0.051071 |
| 1시간 최대강수량(mm) | -0.098266 | -0.397282 | -0.094004 | -0.007664 | -0.052863 | 0.031399 |
| 한계 일사량(MJ/m²) | -0.113976 | -0.401095 | -0.116281 | 0.013290 | -0.070923 | 0.007939 |
| 일 최대일사량(MJ/m²) | 0.036772 | 0.064923 | 0.012708 | -0.033029 | 0.484251 | -0.549180 |
| 평균 강수량(1/기) | -0.141227 | 0.312524 | -0.138954 | 0.117806 | 0.210962 | -0.058575 |
| 9~9강수량(mm) | -0.141775 | 0.216036 | -0.024563 | 0.189145 | 0.536990 | -0.081027 |
| 한계 강수량(1/기) | 0.014666 | 0.049552 | 0.003954 | -0.120862 | 0.239930 | 0.789991 |

④ PCA 적용 결과 분석

- PCA 적용 전

```
OLS Regression Results
Dep. Variable:  (선택)      R-squared:  0.001
Model:  OLS              Adj. R-squared:  0.000
Method:  Least Squares    F-statistic:  0.014
Date:    Wed, 22 Mar 2023    Prob (F-statistic):  0.914
Time:    12:08:01          Log Likelihood:  -101.4
No. Observations: 292      AIC:      6427
DF Residuals: 290         BIC:      6490
DF Model: 21
Covariance Type: nonrobust

Intercept    coef    std err   t    P>|t|    [0.025    0.975]
(Intercept)  1.047936    0.012086    86.703    0.000    1.023764    1.072108
아산화질소  0.346618    0.062099    5.582    0.000    0.218425    0.474811
오존농도     -0.179436    0.025196   -7.126    0.000   -0.229544   -0.129328
아산화탄소  0.348779    0.069645    5.008    0.000    0.209404    0.488154
아황산가스   0.276626    0.090698    3.047    0.003    0.095404    0.457848
미세먼지     0.283448    0.013585   20.861    0.000    0.256189    0.310707
초미세먼지   0.269638    0.021873   12.330    0.000    0.225892    0.313384
평균기온     -0.339944    0.050321   -6.756    0.000   -0.440281   -0.239607
최고기온     -0.322360    0.096698    3.335    0.001   -0.514581   -0.129139
일강수량     -0.151558    0.236310   -0.641    0.520   -0.624044    0.320928
최대 풍속     -0.029676    0.069268    0.428    0.672   -0.107434    0.048082
평균 풍속     -0.054174    0.032440   -1.670    0.098   -0.119444    0.011100
최소 상대습도 -0.204510    0.325643   -0.628    0.530   -0.954481    0.545461
평균 상대습도 -0.199349    0.306431   -0.651    0.517   -0.914481    0.515783
가조시간     -0.321821    0.124440   -2.586    0.011   -0.564481    -0.079161
한계 일조시간 0.024418    0.400196    0.061    0.948   -0.774481    0.823481
1시간 최대강수량 -0.098266    0.397282   -0.247    0.804   -0.884481    0.687949
한계 일사량   -0.113976    0.401095   -0.284    0.776   -0.904481    0.586429
일 최대일사량 0.036772    0.064923    0.566    0.577   -0.090481    0.163754
평균 강수량   -0.141227    0.312524   -0.452    0.651   -0.754481    0.271827
9~9강수량     -0.141775    0.216036   -0.656    0.513   -0.564481    0.281230
한계 강수량   0.014666    0.049552    0.296    0.771   -0.074481    0.103813
```

- PCA 적용 후

```
OLS Regression Results
Dep. Variable:  (선택)      R-squared:  0.677
Model:  OLS              Adj. R-squared:  0.660
Method:  Least Squares    F-statistic:  38.65
Date:    Wed, 22 Mar 2023    Prob (F-statistic):  6.37e-59
Time:    12:21:08          Log Likelihood:  -3228.7
No. Observations: 292      AIC:      6489
DF Residuals: 276         BIC:      6548
DF Model: 15
Covariance Type: nonrobust

Intercept    coef    std err   t    P>|t|    [0.025    0.975]
(Intercept)  6.2377e+04  3435.580  18.121  0.000  5.58e+04  6.93e+04
C(아산화질소)  0.346618  0.062099  5.582  0.000  0.218425  0.474811
C(아산화탄소)  0.348779  0.069645  5.008  0.000  0.209404  0.488154
C(아황산가스)  0.276626  0.090698  3.047  0.003  0.095404  0.457848
C(미세먼지)  0.283448  0.013585  20.861  0.000  0.256189  0.310707
C(초미세먼지)  0.269638  0.021873  12.330  0.000  0.225892  0.313384
C(평균기온)  -0.339944  0.050321  -6.756  0.000  -0.440281  -0.239607
C(최고기온)  -0.322360  0.096698  3.335  0.001  -0.514581  -0.129139
C(일강수량)  -0.151558  0.236310  -0.641  0.520  -0.624044  0.320928
C(최대 풍속)  -0.029676  0.069268  0.428  0.672  -0.107434  0.048082
C(평균 풍속)  -0.054174  0.032440  -1.670  0.098  -0.119444  0.011100
C(최소 상대습도) -0.204510  0.325643  -0.628  0.530  -0.954481  0.545461
C(평균 상대습도) -0.199349  0.306431  -0.651  0.517  -0.914481  0.515783
C(가조시간)  -0.321821  0.124440  -2.586  0.011  -0.564481  -0.079161
C(한계 일조시간) 0.024418  0.400196  0.061  0.948  -0.774481  0.823481
C(1시간 최대강수량) -0.098266  0.397282  -0.247  0.804  -0.884481  0.687949
C(한계 일사량)  -0.113976  0.401095  -0.284  0.776  -0.904481  0.586429
C(일 최대일사량) 0.036772  0.064923  0.566  0.577  -0.090481  0.163754
C(평균 강수량)  -0.141227  0.312524  -0.452  0.651  -0.754481  0.271827
C(9~9강수량)    -0.141775  0.216036  -0.656  0.513  -0.564481  0.281230
C(한계 강수량)  0.014666  0.049552  0.296  0.771  -0.074481  0.103813
```

② 누적 분산 설명 비율

| PC 수 | 누적 분산 설명 비율 |
|------|-------------|
| 1 | 0.20537245 |
| 2 | 0.3154477 |
| 3 | 0.40115552 |
| 4 | 0.47524658 |
| 5 | 0.53716937 |
| 6 | 0.57996796 |

③ 정규화된 observation에 미치는 각

PC의 영향력

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|----------------------|----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 2.271962 | 0.859127 | 1.955583 | -0.805607 | -0.053974 | -0.042886 |
| 1 | 3.419612 | -0.245713 | 1.215544 | -1.742047 | -0.936706 | -0.205647 |
| 2 | 3.511271 | -0.248265 | 0.587579 | -2.433296 | -1.125087 | -0.332426 |
| 3 | 4.965181 | 2.174880 | -1.489308 | -1.055540 | -0.526938 | -0.058662 |
| 4 | 3.437423 | -0.471820 | 1.374651 | -0.059552 | -0.493292 | -0.020124 |
| ... | ... | ... | ... | ... | ... | ... |
| 360 | 2.345956 | -0.300279 | 2.019338 | -1.157044 | -0.535905 | -0.032629 |
| 361 | 2.873856 | 1.034874 | 0.806869 | -1.155156 | -0.439519 | -0.056182 |
| 362 | 2.136826 | 3.229242 | 1.719310 | 0.217015 | 0.553901 | 0.002616 |
| 363 | 1.323629 | 3.261106 | 3.182388 | 1.498414 | 0.990926 | 0.301557 |
| 364 | 1.307547 | -0.810679 | 5.141601 | 0.842221 | -0.008309 | 0.057114 |
| 365 rows x 6 columns | | | | | | |

-----PCA 적용 후-----
훈련 세트의 정확도 : 0.68
테스트 세트의 정확도 : 0.70

-----PCA 적용 전-----
훈련 세트의 정확도 : 0.82
테스트 세트의 정확도 : 0.83

▶ feature의 수 감소에
따라 사진과 같은 차이만큼
의 원 데이터의 정보(분산)
의 손실 발생

-----PCA 적용 후-----
r2_score : 0.70
MAE : 12219.00
MSE : 235121530.66

-----PCA 적용 전-----
r2_score : 0.83
MAE : 8925.71
MSE : 134538299.21

Body

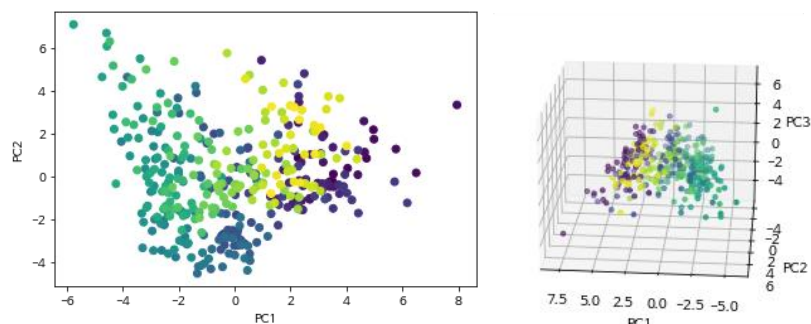
Contents

PCA(CONT.)

03. PCA PLOT(CONT.)

① ScorePlot

Principal Component Pattern Plot : 2D / 3D



▶ 주성분 (PC1, PC2 / PC1, PC2, PC3)과 원데이터 간 관계 파악 가능

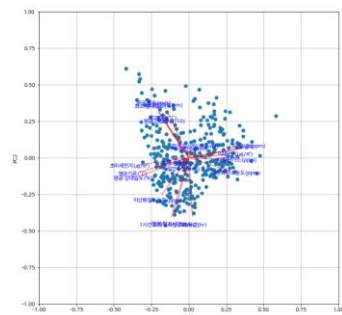
② Biplot

▶ 주성분 (PC1과 PC2)과 원데이터의 변수 간의 상관 관계 파악 가능

- 이산화탄소농도(ppm), 합계 일사량(MJ/m2)가
PC1, PC2에 영향을 가장 많이 주는 변수

- 각 빨간 선의 길이는 원변수의 분산을 표현. 길이가
길수록 분산이 큼

- 각 빨간 선이 가까울수록 서로 상관관계 높음



FA

01. 요인 분석 전 검토

① KMO Test

0.78 : 요인분석을 위한 변수 선정이 적당함

② Bartlett Test

chi_square_value = 9522.41, p_value = 0.0 < 0.05

→ 귀무가설 기각 → 요인분석 모델 사용가능

02. Factor 수 결정

PCA에서 6개의 factors를 선택하였으므로, FA에서도 동일하게 6개의 factors 선택

03. Factor Analysis 시행 : Rotation 방식 비교 *절댓값 기준

① varimax

Factor 1 : 기온

Factor 2 : 미세먼지 및 대기질

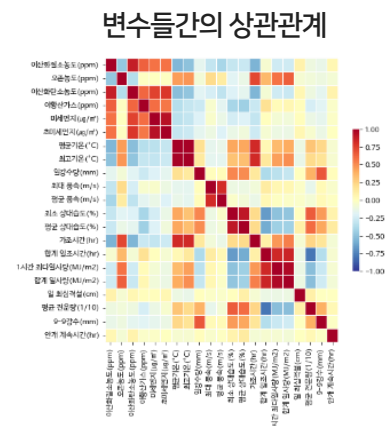
Factor 3 : 일조시간 및 일사량

Factor 4 : 풍속

Factor 5 : 강수

Factor 6 : 습도

- 모든 Factor에서 높은 적재량을 가진 특징적인 값이 존재하므로, 6개의 요인 모두 최종 선택



| | Factor1 | Factor2 | Factor3 | Factor4 | Factor5 | Factor6 |
|----------------|----------|----------|----------|----------|----------|----------|
| SS Loadings | 3.979419 | 3.509280 | 3.270060 | 1.909546 | 1.880331 | 1.381277 |
| Proportion Var | 0.189496 | 0.167109 | 0.155717 | 0.090931 | 0.089540 | 0.065775 |
| Cumulative Var | 0.189496 | 0.356605 | 0.512322 | 0.603253 | 0.692792 | 0.758567 |

Body

Contents

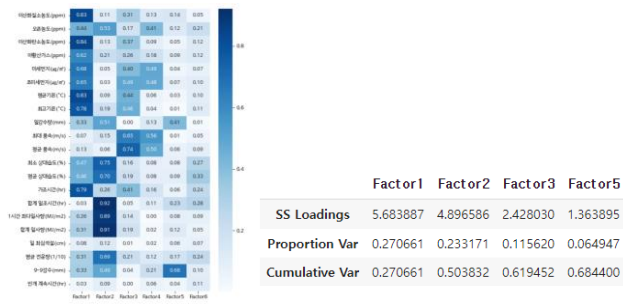
PCA / FA

FA(CONT.)

03. Factor Analysis 시행 : Rotation 방식 비교(CONT.) *절댓값 기준

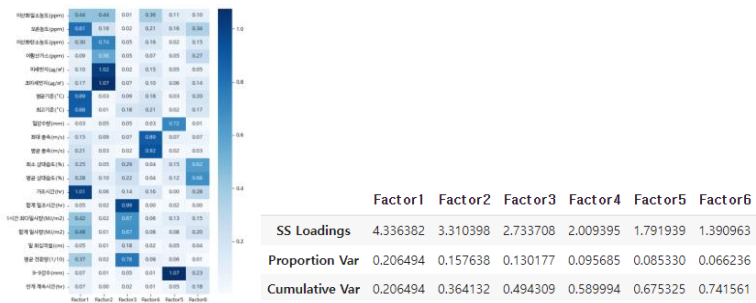
② oblimax

- Factor 1 : 기온, 미세먼지 및 대기질
Factor 2 : 습도, 일조시간 및 일사량
Factor 3 : 풍속
Factor 4 : -
Factor 5 : 강수
Factor 6 : -
- Varimax 방식에 비해 각 factor의 원인이 명확히 분류되지 않으며, factors 4 / 6에서는 특징적인 값이 보이지 않음



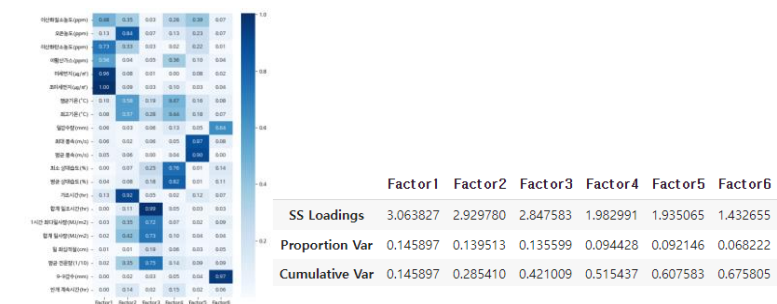
③ promax

- Factor 1 : 기온
Factor 2 : 미세먼지 및 대기질
Factor 3 : 일조시간 및 일사량
Factor 4 : 풍속
Factor 5 : 강수
Factor 6 : 습도
- varimax와 동일한 결과를 보임



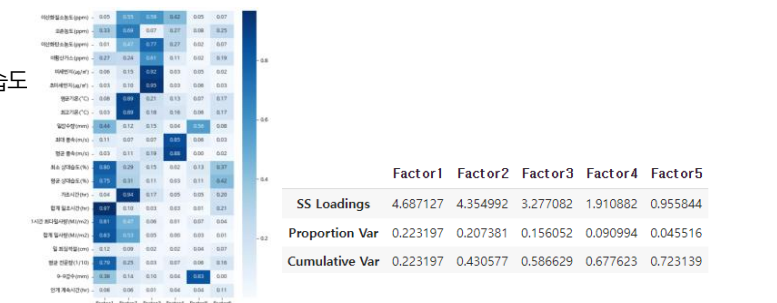
④ oblimin

- Factor 1 : 미세먼지 및 대기질
Factor 2 : 기온
Factor 3 : 일조시간 및 일사량
Factor 4 : 습도
Factor 5 : 풍속
Factor 6 : 강수
- varimax와 비교했을 때 factor가 의미하는 요인은 동일하지만, factor의 순서가 변함



⑤ quartimax

- Factor 1 : 일조시간 및 일사량, 습도
Factor 2 : 기온
Factor 3 : 미세먼지 및 대기질
Factor 4 : 풍속
Factor 5 : 강수량
Factor 6 : -
- varimax와 비교했을 때 일조시간과 습도가 모두 Factor 1을 설명하여 Factor 6을 설명하는 유의미한 변수가 존재하지 않음



Body / Conclusion

Contents

PCA / FA

FA(CONT.)

04. Result

가장 분산을 잘 설명하는(75.86%) Varimax rotation 사용

(1) Naming 결과

Factor 1 : Temperature (기온)

Factor 2 : FineDust (미세먼지 및 대기질)

Factor 3 : Sunshine (일조시간 및 일사량)

Factor 4 : Wind (풍속)

Factor 5 : Rain (강수)

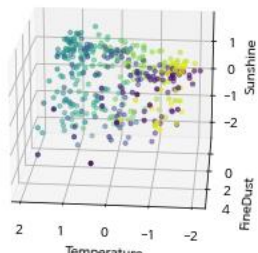
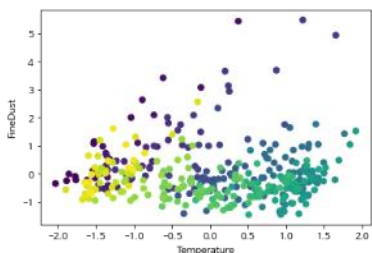
Factor 6 : Humidity (습도)

| | Factor1 | Factor2 | Factor3 | Factor4 | Factor5 | Factor6 |
|----------------|----------|----------|----------|----------|----------|----------|
| SS Loadings | 3.979419 | 3.509280 | 3.270060 | 1.909546 | 1.880331 | 1.381277 |
| Proportion Var | 0.189496 | 0.167109 | 0.155717 | 0.090931 | 0.089540 | 0.065775 |
| Cumulative Var | 0.189496 | 0.356605 | 0.512322 | 0.603253 | 0.692792 | 0.758567 |

| | Temperature | FineDust | Sunshine | Wind | Rain | Humidity |
|----------------------|-------------|-----------|-----------|-----------|-----------|-----------|
| 0 | -1.760737 | -0.218244 | 0.208757 | 0.352227 | -0.162400 | 0.280749 |
| 1 | -1.882598 | -0.251930 | 0.350146 | -0.598990 | -0.217219 | -0.492100 |
| 2 | -2.029454 | -0.346760 | 0.335377 | -1.240361 | -0.161482 | -0.709403 |
| 3 | -1.521864 | 1.147924 | -0.660044 | -1.588139 | -0.117820 | 0.028627 |
| 4 | -1.318566 | 0.518779 | 0.267412 | 0.456086 | -0.308061 | -0.347788 |
| ... | ... | ... | ... | ... | ... | ... |
| 360 | -1.642482 | -0.306126 | 0.508599 | 0.189706 | -0.224694 | 0.038111 |
| 361 | -1.551791 | -0.026201 | -0.226642 | -0.374405 | -0.243240 | 0.118795 |
| 362 | -1.069277 | 0.219879 | -1.666225 | 0.707326 | -0.331706 | -0.194285 |
| 363 | -1.011370 | 0.136085 | -1.685760 | 1.940966 | -0.393466 | 0.434639 |
| 364 | -1.556918 | -0.704649 | 0.413439 | 2.508610 | -0.379592 | -0.422040 |
| 365 rows x 6 columns | | | | | | |

Factor Analysis가 적용된 dataset (scaled)

(2) Score Plot : 2D / 3D



Temperature
0.908648960016989

FineDust
0.9177120574354731

Sunshine
0.381401843036471

Wind
0.8996742390902774

Rain
0.8036690314200539

Humidity
0.9456331525643407

신뢰도계수 (Cronbach's alpha 계산)

- Factor 1,2,4,5,6의 Cronbach's Alpha값 매우 우수

Conclusion

01. Result

(1) 날씨와 관련된 변수들 간의 상호작용 및 공유자전거 대여량에 영향을 미치는 요인들을 파악함

- 공유자전거 대여량 예측에 있어서 PCA와 FA가 유용한 분석 방법이며, 날씨와 관련된 변수들이 공유자전거 대여량 예측에 중요한 역할을 하는 것을 확인

(2) 차원 축소된 결과로 공유자전거 대여량을 적은 정보 손실로 예측함

- 변수 : 21개 → 6개, R_squared : 0.83 → 0.7

(3) FA를 통해 변수 간 상관 구조를 설명하기 위한 공통 요인을 추출

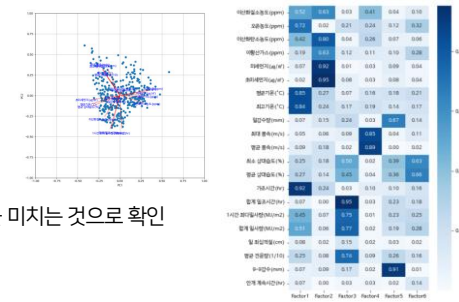
- 기온, 미세먼지, 일조시간, 풍속, 강수 및 습도와 같은 요소가 공유자전거 대여량에 영향을 미치는 것으로 확인

(4) 공유자전거 운영 기관 등에게 유용한 정보를 제공 가능

- 날씨 정보를 실시간으로 모니터링하고, 이를 바탕으로 공유자전거 대여량을 예측하는 시스템을 도입하면 효율적인 운영 가능

(5) 효율적인 공유자전거 운영을 위한 정책적 제언 제공 가능

- 날씨에 따른 공유자전거 대여요금 차등 적용 등



02. Limitation & Development

(1) 대여량 데이터의 시계열적 특성을 고려하지 않음

→ 시간에 따른 대여량의 변화를 고려할 수 있는 기법 적용 / 데이터 확보 필요

(2) 날씨 이외의 공유자전거 대여량에 영향을 미치는 다른 변수들이 분명히 존재(대여 요금, 지역별 특성, 이벤트 등)

→ 이러한 변수들을 고려한다면 개선된 예측 정확도 및 변수 관계 파악 가능

Data Source

Contents

PCA / FA

[데이터 출처]

(1) 서울 열린 데이터 광장

- 서울시 일별 평균 대기오염도 정보 <http://data.seoul.go.kr/dataList/OA-2218/S/1/datasetView.do#>
- 서울시 공공자전거 이용현황 <https://data.seoul.go.kr/dataList/OA-14994/F/1/datasetView.do>

(2) 기상청

- 일별 종관기상관측(ASOS) 자료 <https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36>

(3) 2016 ~ 19년 공휴일 데이터

- <https://superkts.com/day/holiday/2019>

[참고 문헌]

- (1) 김동준 신희철 et al. 박준식 임형준 “날씨가 자전거 이용에 미치는 영향 분석 - 고양시 공공자전거를 대상으로 -” 교통연구 19.3 pp. 77-88 (2012): 77.