



Shinchon Ultra ML

# 서울시 종로구 익일 미세먼지 농도 예측 모델

KoreaIT Final Project

# INDEX

01

조원 및 프로젝트 소개

02

데이터 설명 및 전처리

03

데이터 분석

04

베이스라인 모델

05

피처 선정 및 모델 최적화

5-1

Machine Learning

5-2

Deep Learning

06

최종 모델 소개

07

결론



장서운

데이터 분석

## Key Role

프로젝트 매니저

데이터 분석



정이슬

프론트엔드, 백엔드

## Key Role

데이터베이스 구축



김종엽

머신러닝 및 딥러닝 모델링

## Key Role

데이터 전처리

머신러닝 모델링



김승주

머신러닝 및 딥러닝 모델링

## Key Role

데이터 전처리

딥러닝 모델링



출처 : 뉴시스, 2022 ([https://mobile.newsis.com/view.html?ar\\_id=NISX20221213\\_0002121544#\\_PA](https://mobile.newsis.com/view.html?ar_id=NISX20221213_0002121544#_PA))

직전 주 대비

71%

공기청정기 판매량

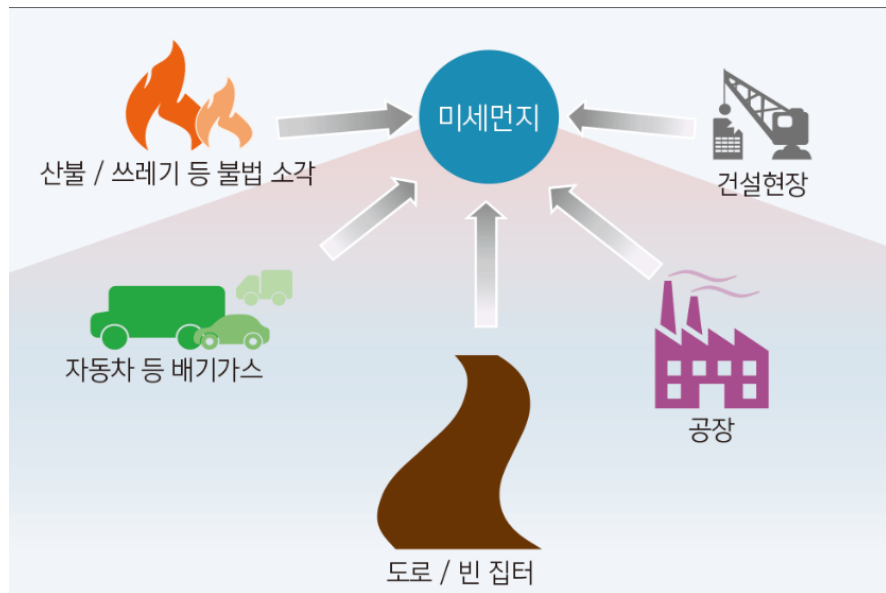
미세먼지 주의보

1월 1일부터 7일까지 일주일 간

공기청정기 판매량

직전 주 대비 71% 증가

(Jang, 2023.02, inews24 )



출처 : 미세먼지특별대책위원회, 2022 (<https://www.cleanair.go.kr/dust/dust/dust-define.do>)

미세먼지:  $10\mu\text{g}$  이하의 대기질에 떠 다니는 물질

발원지

- 국내 발생
  - 산업시설, 자동차, 난방 및 에너지 사용
- 국외 발생
  - 중국의 사막과 공업지역에서 편서풍을 타고 넘어오는 먼지
  - 일본에서 넘어오는 먼지

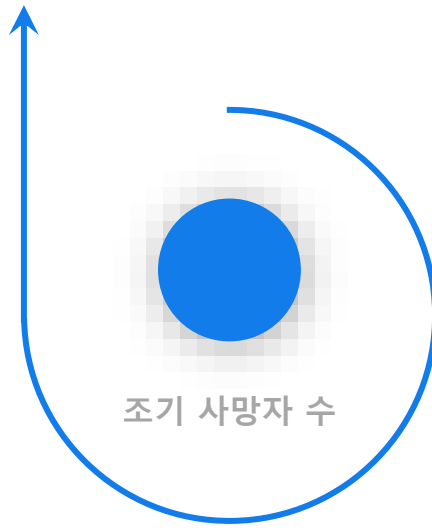
“미세먼지로 해마다 서울  
시민 1만명 넘게  
**조기 사망**”

(Yim et al., 2023, 환경과학기술저널)



전세계 인류 평균수명  
**2.2년**

2010 대비  
**1만명**



“2024년 조기사망자  
**2만 5천명** 예상”

(Lim, 2015)

“ 내일 미세먼지 농도를 미리 알 수 있으면 마스크를 잊지 않을 텐데...”

Objective

1

프로젝트 목표 1

**중요 변수 추출**

서울시 익일 미세먼지  
예측에 가장 중요한 변수 선정

Objective

2

프로젝트 목표 2

**예측모델 생성**

알고리즘 별 비교 평가  
정확도 (Accuracy) 기반 평가

Objective

3

프로젝트 목표 3

**예측 서비스 배포**



Objective

1

프로젝트 목표 1

중요 변수 추출

서울시 익일 미세먼지  
예측에 가장 중요한 변수 선정

- 현재의 한정된 자원으로 모든 서울 지역구의 미세먼지 농도를 예측하는 것은 한계가 존재.
- 또한, 서울시 평균 미세먼지 농도를 예측하는 것은 '평균의 함정'에 빠져 예측값의 효용성이 떨어짐

=> 예측을 종로구 미세먼지 농도로 한정

- 서울시 기상관측소가 위치해 있음
- 종로구가 평균과의 오차가 적음 (Song, 2016)

가설1 : 기온과 일산화질소(NO2) 수치가 미세먼지 농도 예측에 가장 큰 영향력을 미칠 것이다. (Park, 2021)

가설2 : 편서풍의 영향으로 인천과 경기 서북부 지역들이 미세먼지 농도 예측에 가장 큰 영향력을 미칠 것이다. (Kang, 2008, Jeong et al., 2011)

2<sup>nd</sup> Objective

50%

프로젝트 목표 2

예측모델 생성

알고리즘 별 비교 평가

정확도 (Accuracy) 기반 평가

- Sung et al.(2020)의 연구에 따르면 국내 미세먼지 예측에 XGBoost, Random Forest, SVM, ANN 중 ANN 이 가장 정확도가 높았다.
- Kristiani et al. (2022)의 연구에 따르면 단기 초미세먼지 예측에 LSTM < Bi-LSTM < Bi-GRU < CNN < RNN 순으로 정확도가 낮았다.
- Cha and Kim(2018) ANN과 KNN 그리고 둘을 합친 모델을 비교했을 때, 제안된 모델이 서울지역 미세먼지 예측 정확도가 크게 높았다.
- Hwang and Sin(2022) 베이징의 초미세먼지 예측에 CNN과 LSTM을 결합한 모델을 생성해 향상된 결과를 도출
- Cho et al. (2021) 천안시의 12시간 뒤 미세먼지 농도 예측에서 LSTM < RNN < DNN 순으로 정확도가 낮았다

CNN 모델 선정 이유 : Convolutional Neural Network(CNN)은 데이터의 공간정보를 유지하며 공간적 특성을 추출하는데 적합하다. (Park, 2022)

가설 : 지리적 특성을 유지하여 분석 할 수 있는 CNN 기반 딥러닝 모델이 정확도가 가장 높을 것으로 예상된다.

3<sup>rd</sup> Objective

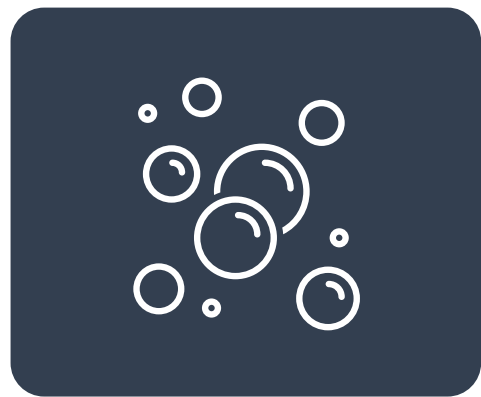
25%

프로젝트 목표 3

예측 서비스 배포

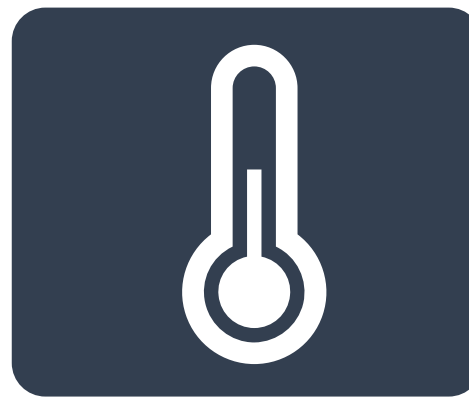
- OPEN API로 실시간 관측 데이터 가져와 예측
- 24시간 뒤 종로구 미세먼지 농도를 기상청 예보 범위로 제공
- FAST API : Pythonanywhere로 배포

## 01 수도권 대기환경 측정자료



- 수도권 측정소별 측정자료
- 시간별 (정각)
- 출처 : 국립환경과학원
- 기간 : 6년 6개월
- 2015.01.01 ~ 2022.08.31

## 02 수도권 기상관측 자료



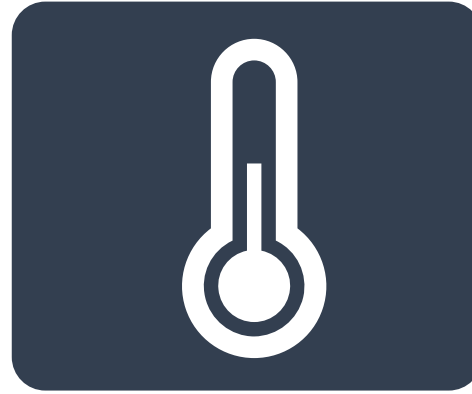
- 수도권 관측소 지상관측
- 종관기상관측(ASOS)
- 시간별 (시 평균)
- 출처 : 기상청
- 기간 : 6년 6개월
- 2015.01.01 ~ 2022.08.31

## 01 수도권 대기환경 측정자료

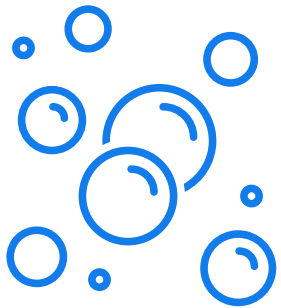


PM10(미세먼지, 종속변수),  
PM25 (초미세먼지)  
O3(오존),  
NO2(이산화질소),  
CO(일산화탄소),  
SO2(아황산가스)

## 02 수도권 기상관측 자료



기온, 강수량, 풍속, 풍향,  
습도, 증기압, 이슬점온도,  
현지기압, 해면기압, 일조,  
일사, 적설, 3시간신적설,  
전운량, 중하층운량, 운형,  
최저운고, 시정, 지면상태,  
현상번호, 지면온도, 지중온도



미세먼지 측정자료 전처리

## &lt;미세먼지 농도별 예보 등급&gt;

**좋음** PM<sub>10</sub> 0~30(μg/m<sup>3</sup>) PM<sub>2.5</sub> 0~15(μg/m<sup>3</sup>)

대기오염 관련 질환자군에서도 영향이 유발되지 않을 수준

**보통** PM<sub>10</sub> 31~80(μg/m<sup>3</sup>) PM<sub>2.5</sub> 16~35(μg/m<sup>3</sup>)

환자군에게 만성 노출시 경미한 영향이 유발될 수 있는 수준

**나쁨** PM<sub>10</sub> 81~150(μg/m<sup>3</sup>) PM<sub>2.5</sub> 36~75(μg/m<sup>3</sup>)

환자군 및 민감군(어린이, 노약자 등)에게 유해한 영향 유발, 일반인도 건강상 불쾌감을 경험할 수 있는 수준

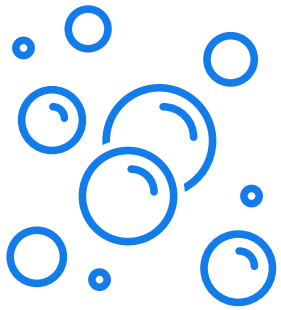
**매우나쁨** PM<sub>10</sub> 151(μg/m<sup>3</sup>)이상 PM<sub>2.5</sub> 76(μg/m<sup>3</sup>)이상

환자군 및 민감군에게 급성 노출시 심각한 영향 유발, 일반인도 약한 영향이 유발될 수 있는 수준



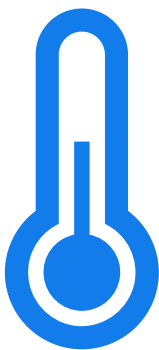
## • 미세먼지 (종속변수) 전처리

- 종로구 익일 PM10을 기상청 예보 기준에 따라 범주화
- ( 0 : 좋음, 1 : 보통, 2 : 나쁨, 3 : 매우 나쁨)



대기환경 측정자료 전처리

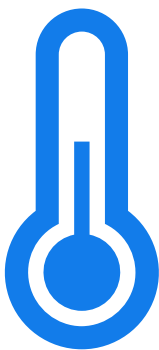
- 결측치 존재 이유
  - 행정구역의 변동이나 기존 측정소의 폐쇄와 측정소 신설
  - 관측기 오류로 인한 기록 누락
- 결측치 처리
  - 1시간 단위 관측 자료이므로 이전 값과 이후 값의 평균으로 대체
  - 초미세먼지(PM25)는 결측치 비율이 너무 커서 제거
  - 결측치가 너무 많은 관측소는 임의로 채우는 게 의미 없다고 판단하여 제거
- 총 118개 측정소의 측정 데이터 사용
  - 서울 : 38개, 인천 : 18개, 경기 : 68개
- 각 측정소의 위치적 특성을 보존하기 위해, 각 측정소의 측정 값들을 측정일시를 기준으로 열 방향으로 병합
- 최종 데이터 크기: (67176, 126)



기상관측 자료 전처리

- 결측치 존재 이유
  - 행정구역의 변동이나 기존 측정소의 폐쇄와 측정소 신설
  - 관측기 오류로 인한 기록 누락
- 결측치 0으로 대체
  - 강수량, 적설은 비와 눈이 안 올 경우 값을 nan으로 기록
  - 일사, 일조는 태양의 영향이 없는 경우, nan으로 기록
  - 최저운고는 관측장소에 구름이 없는 경우, nan으로 기록
- 결측치 전과 후 값의 평균으로 대체
  - 나머지 피쳐는 연속적인 피쳐로 판단
- 결측치 다음 값으로 대체
  - 각 관측소의 제일 첫 값이 결측치인 경우, 이전과 이후 값의 평균으로 대체할 수 없음
  - 측정일시가 누락된 경우도 이후 값으로 대체

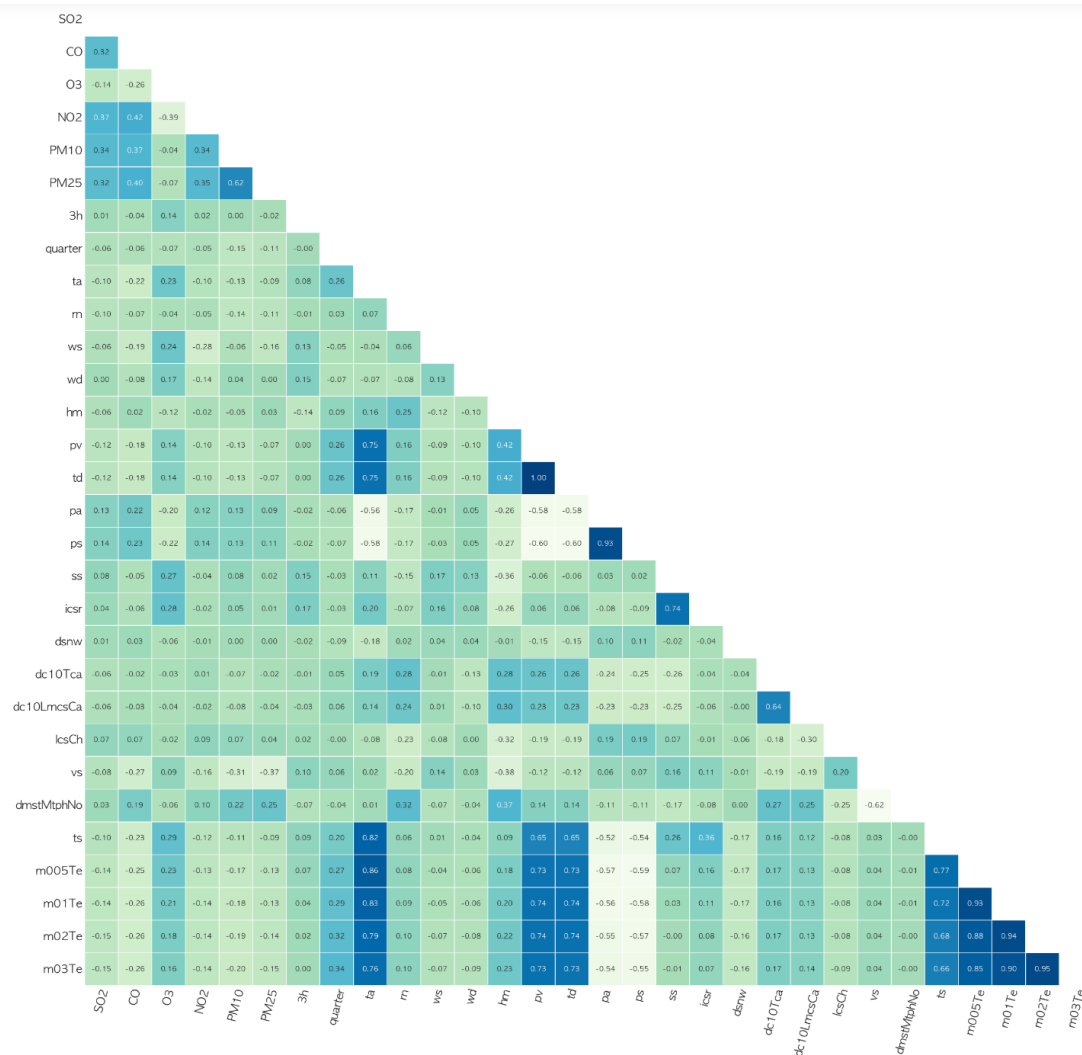




기상관측 자료 전처리

- 변수 제거
  - 결측치가 많거나 사용이 불가능하다고 판단한 변수
  - 지면상태, 3시간신적설, 운형, 지점번호
- 눈, 비, 안개 등의 기상 현상을 번호로 나타낸 현상번호 변수는 원핫인코딩 사용하여 해당 시간의 날씨를 변수화
- 총 9개 관측소의 관측 데이터 사용
  - 서울 : 1개, 인천 : 3개, 경기 : 5개
- 각 측정소의 위치적 특성을 보존하기 위해, 각 측정소의 측정 값들을 측정일시를 기준으로 열 방향으로 병합
- 최종 데이터 크기: (67200, 1927)

# 모든 독립변수 상관관계도



- Kendall's Tau Correlation Coefficient으로 독립변수들 간의 상관관계 정도 확인
  - 비모수적 데이터
- 상관계수 (Coefficient)란?
  - -1 ~ 1사이의 값 반환
  - 1에 가까울 수록 서로 상관 관계가 큼
  - 음수 값은 음의 상관관계
  - 양수 값은 양의 상관관계

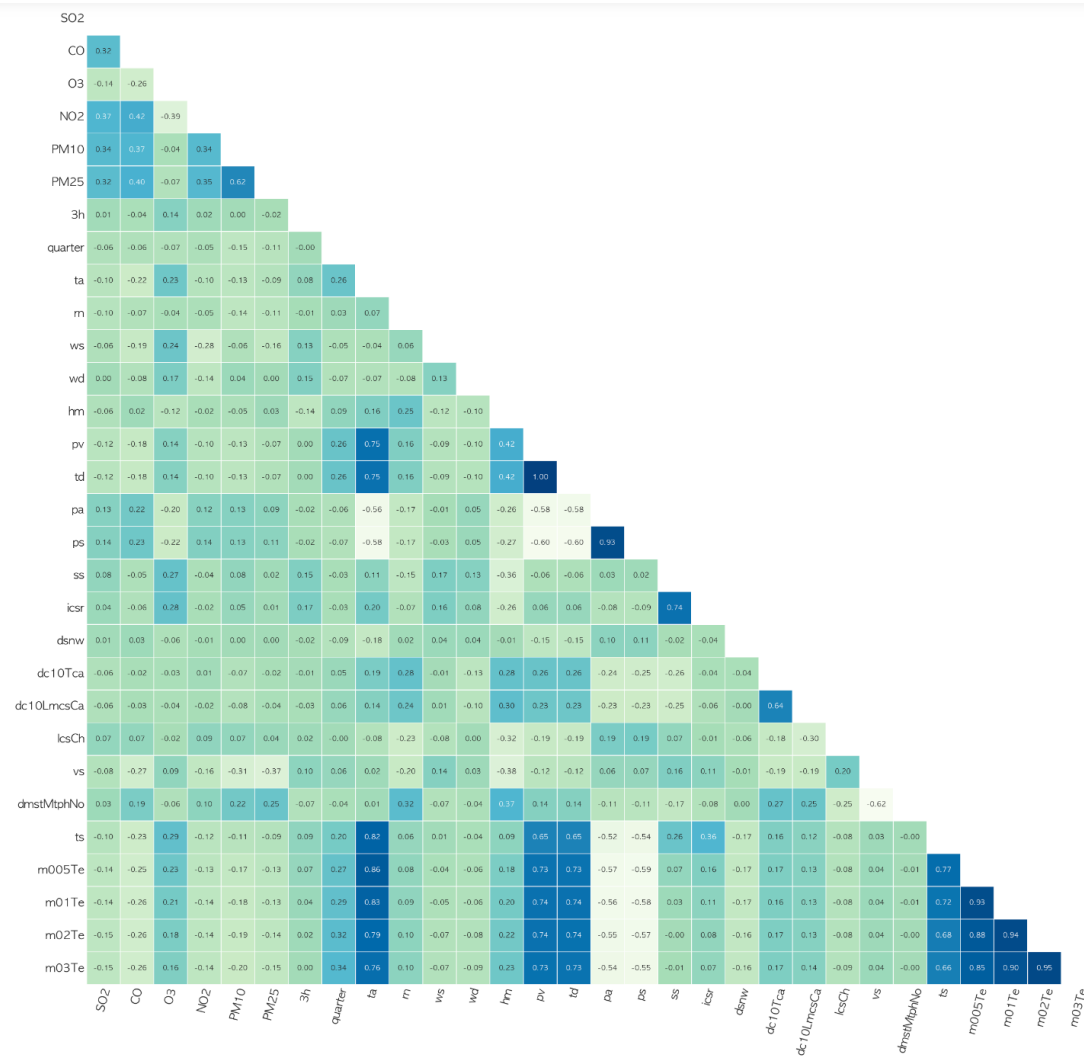
변수	$\tau$
이슬점 온도, 증기압	0.99
지중온도(30cm) 지중온도 (20cm)	0.95
지중온도(20cm) 지중온도 (10cm)	0.93
해면기압 현지기압	0.93
지중온도(10cm) 지중온도 (5cm)	0.93
지중온도(30cm) 지중온도 (10cm)	0.90

변수	$\tau$
지중온도(20cm) 지중온도 (5cm)	0.87
지중온도 (5cm), 온도	0.86
지중온도(30cm) 지중온도 (5cm)	0.84
지중온도(30cm) 지중온도 (5cm)	0.82
지면온도 온도	0.81
지중온도(20cm) 온도	0.78



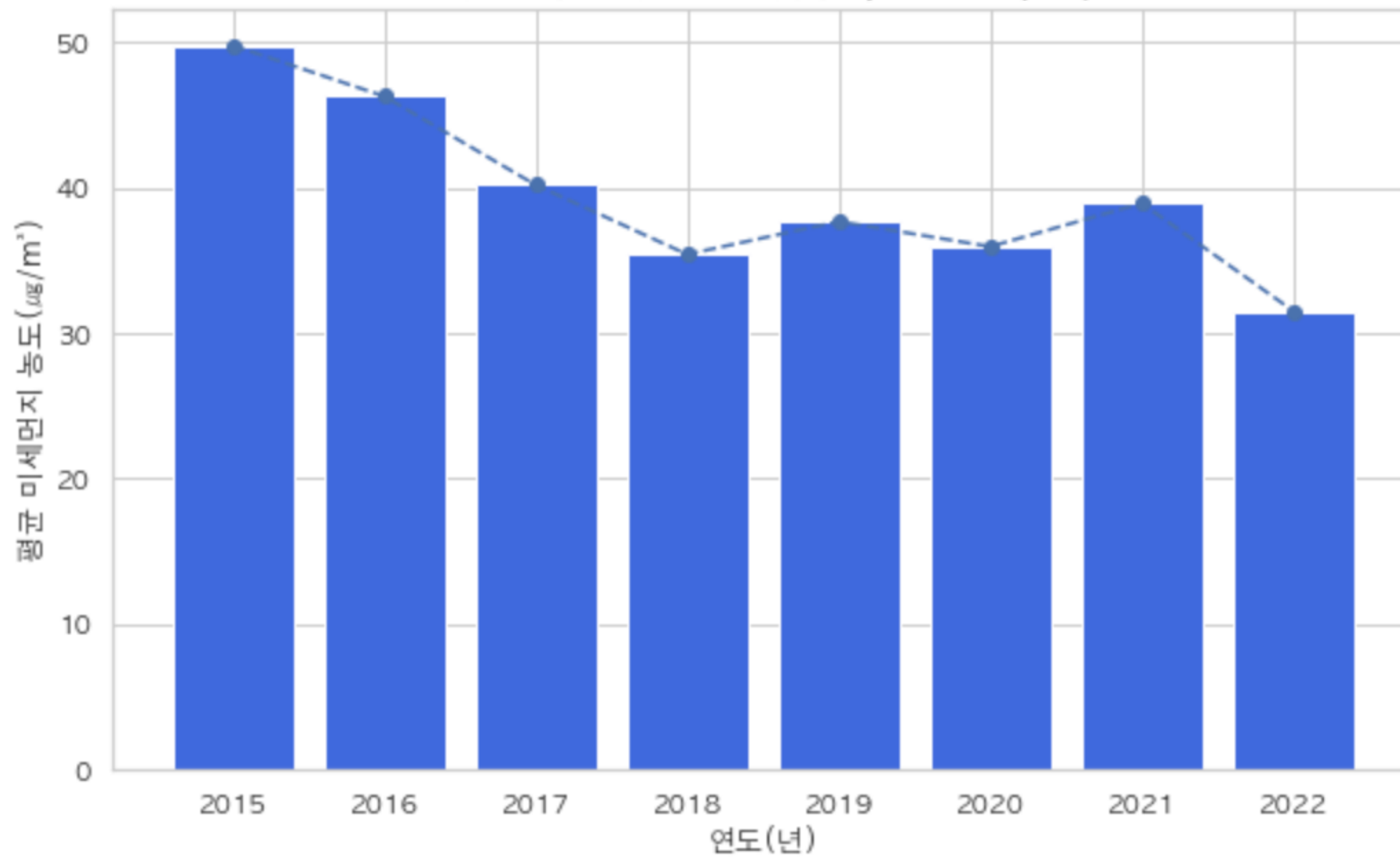
온도에 관련된 변수들간의 상관계수가 높다

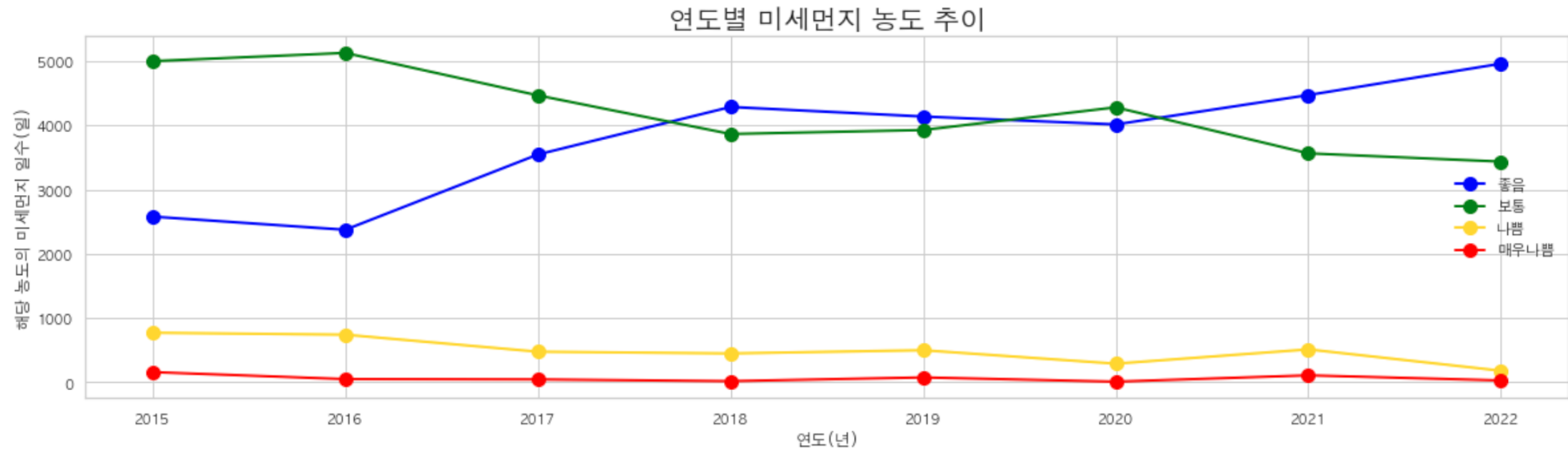
## 모든 독립변수 상관관계도

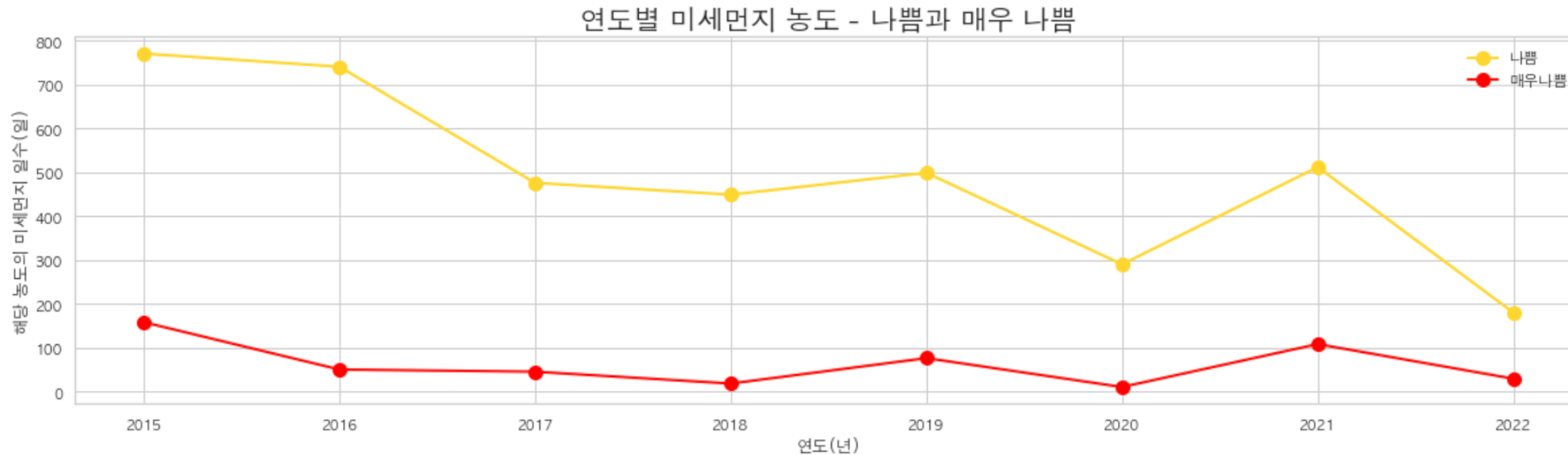


- $\tau > 0.7$  이상인 변수들
  - 지중 온도 변수들
  - 해면기압과 현지기압
  - 기온과 증기압
  - 일조량과 일사량
- 지중온도 변수들이 다중공상성이 의심되어 모델링에서 제외

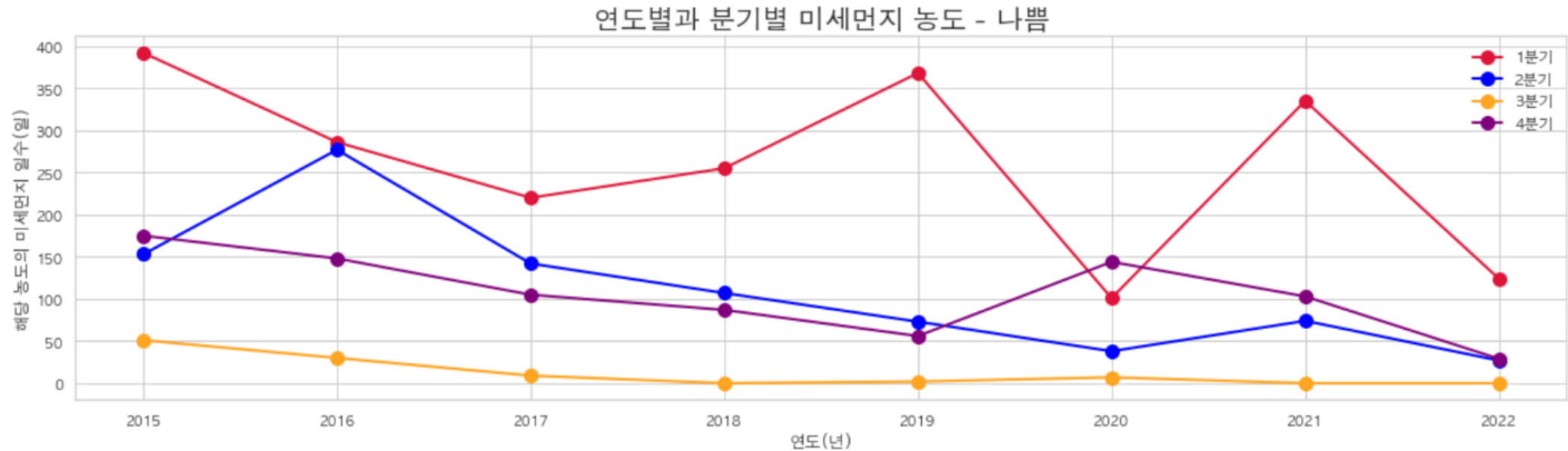
연도별 평균 미세먼지 농도 추이





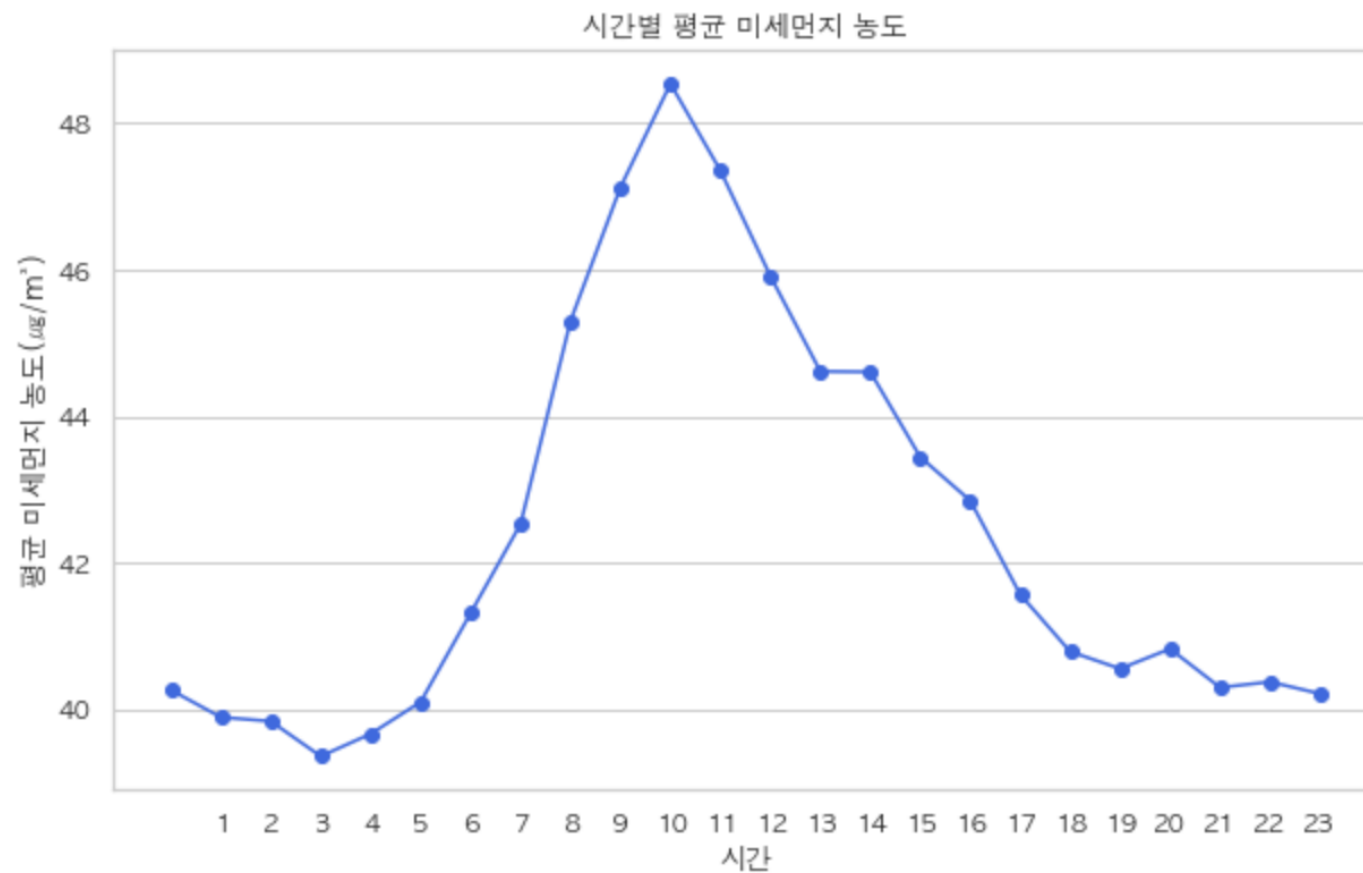


정부의 대기질 개선 정책으로 미세먼지와 대기질이 개선 되었음 (서울경제, 2022.1)



2014년의 월별 미세먼지 농도변화에서도 1월이 가장 농도가 높고 8월에 농도가 가장 낮았다. (Cho, 2014)

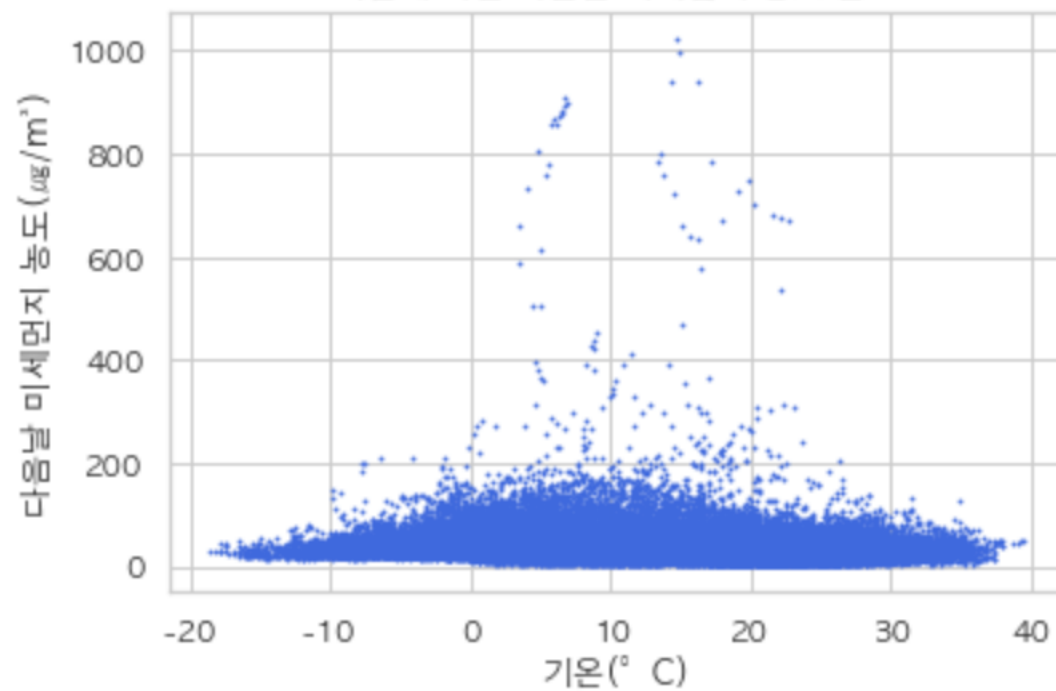




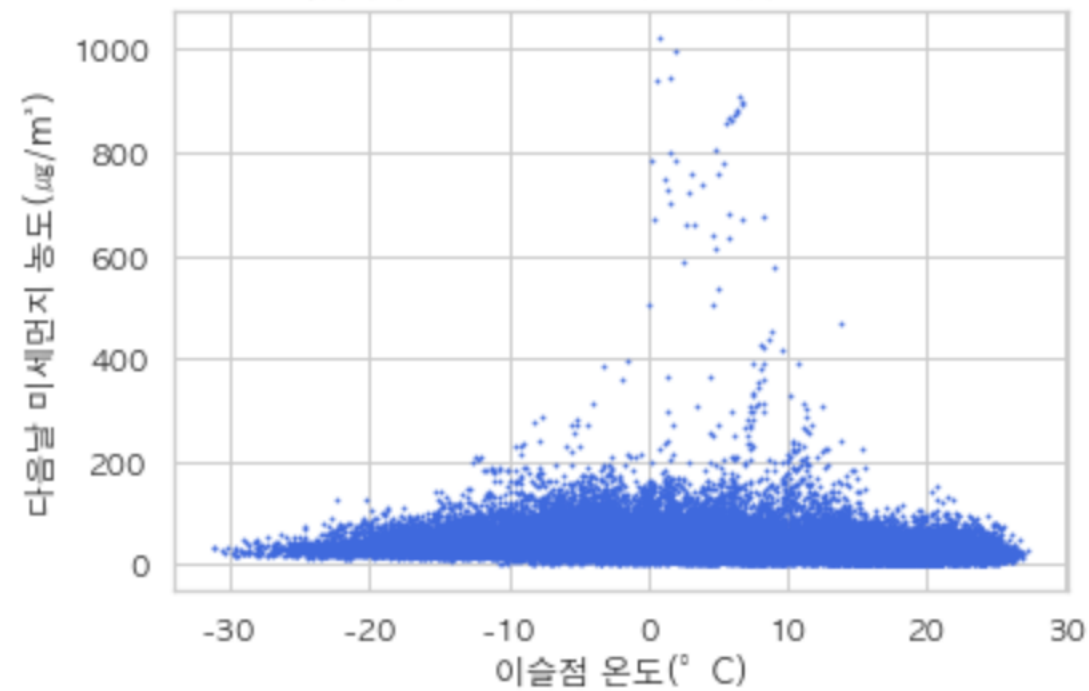
5시부터 급격한 상승을 보이는데 이것은 교통량과 연관이 있다. (Kang, 2008)

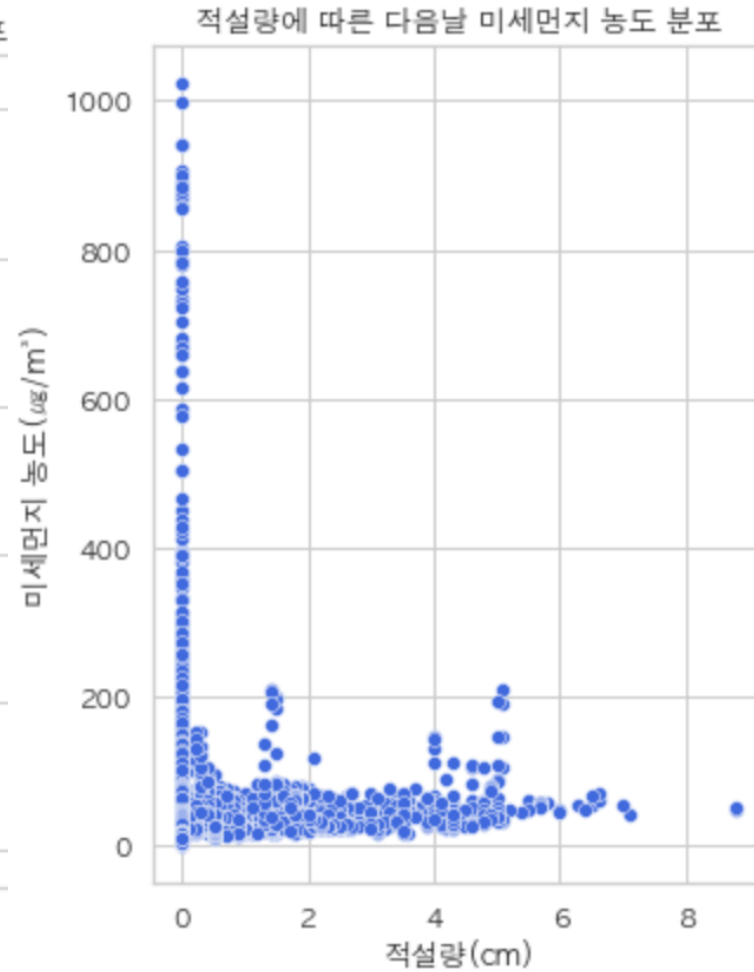
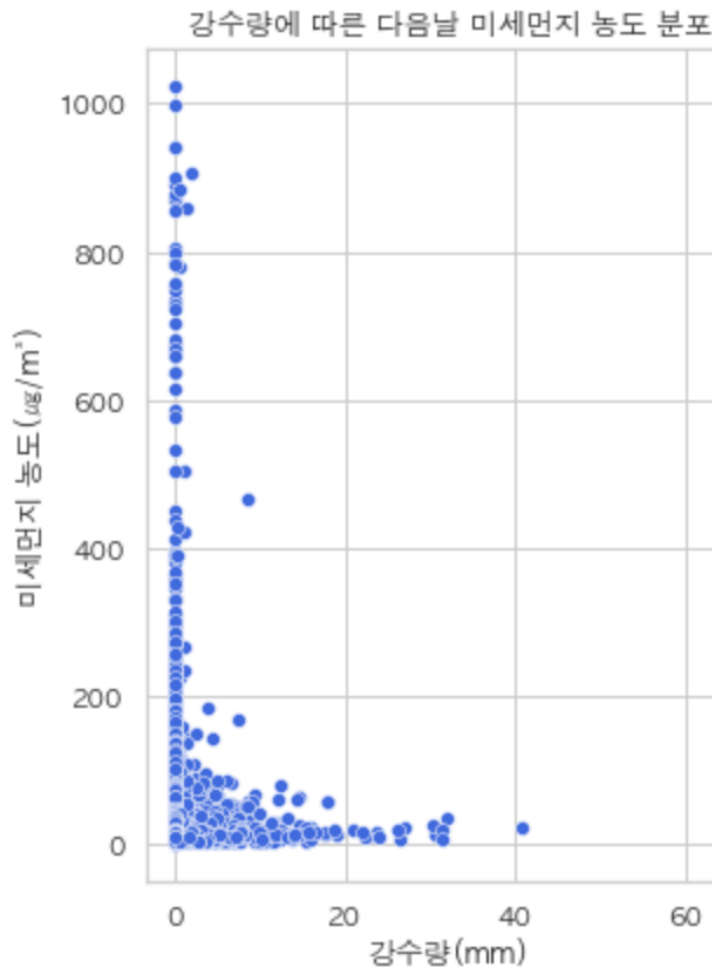
## 기상 관측 데이터와 미세먼지 - 온도

기온에 따른 다음날 미세먼지 농도 분포

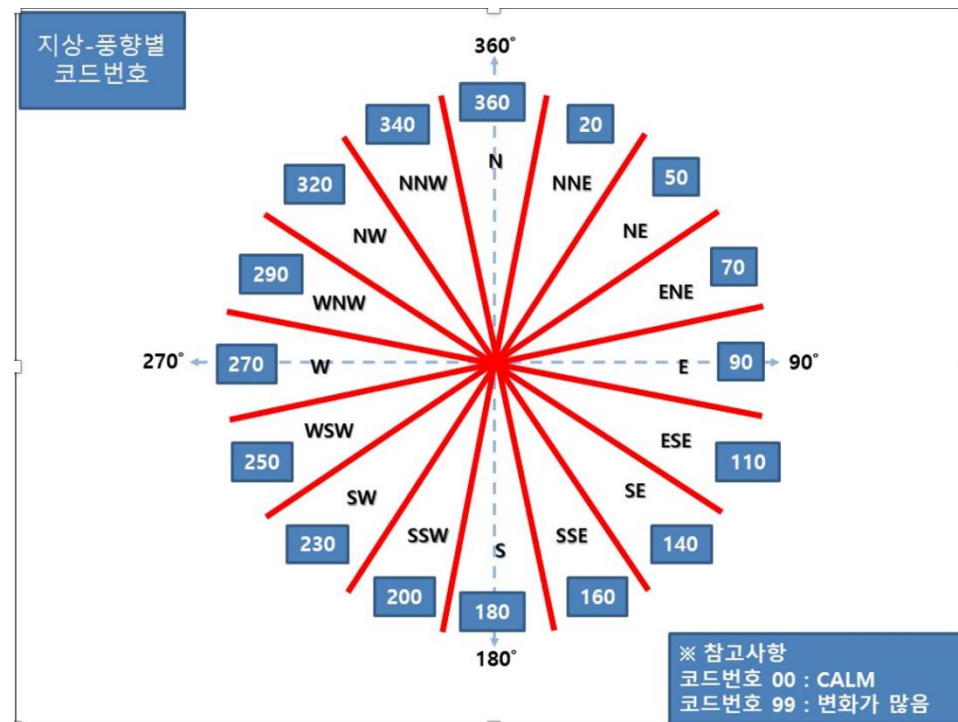
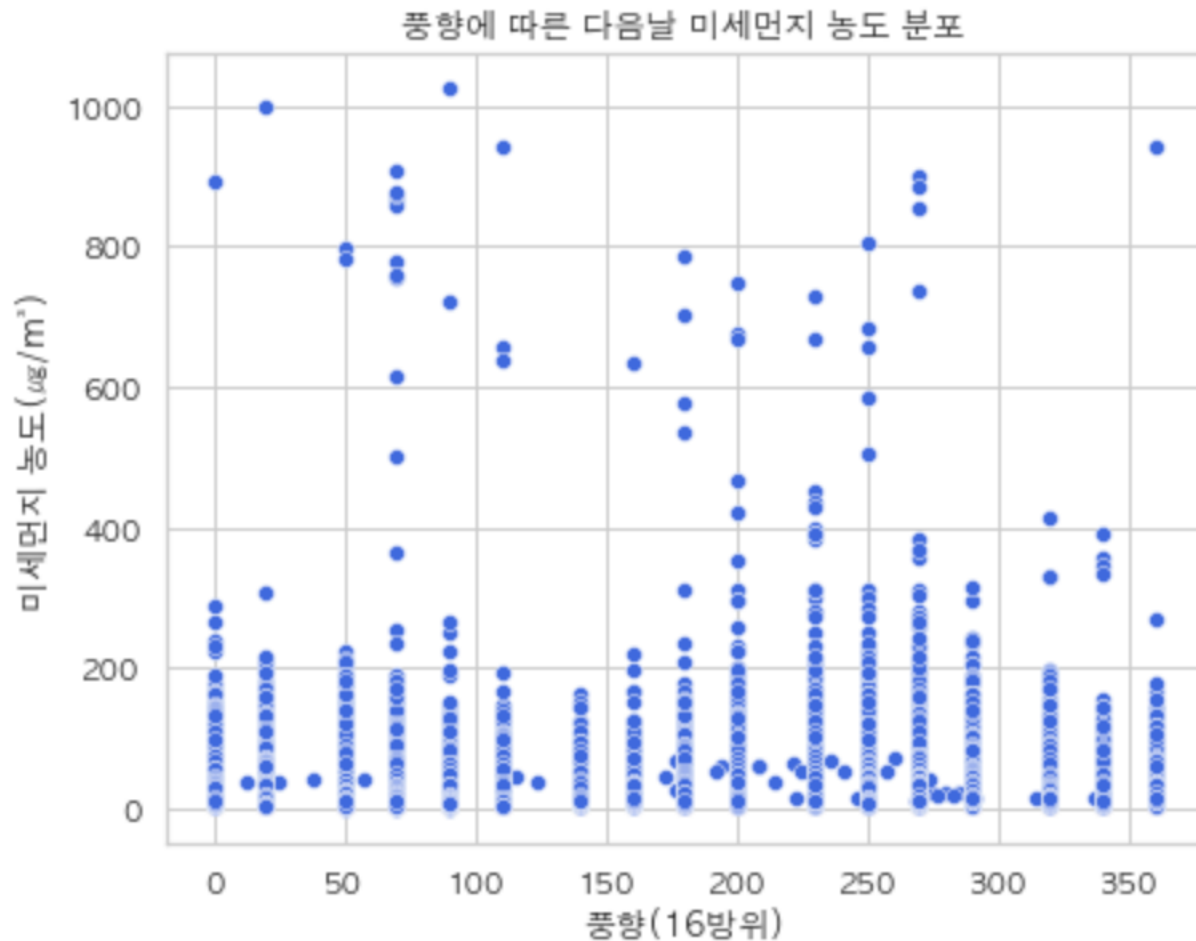


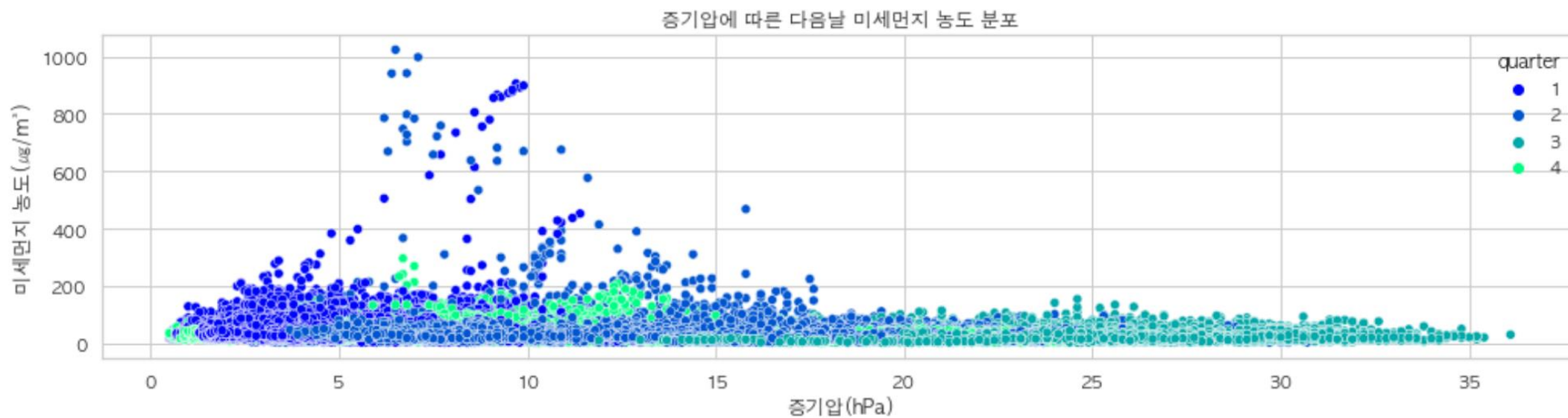
이슬점 온도에 따른 다음날 미세먼지 농도 분포

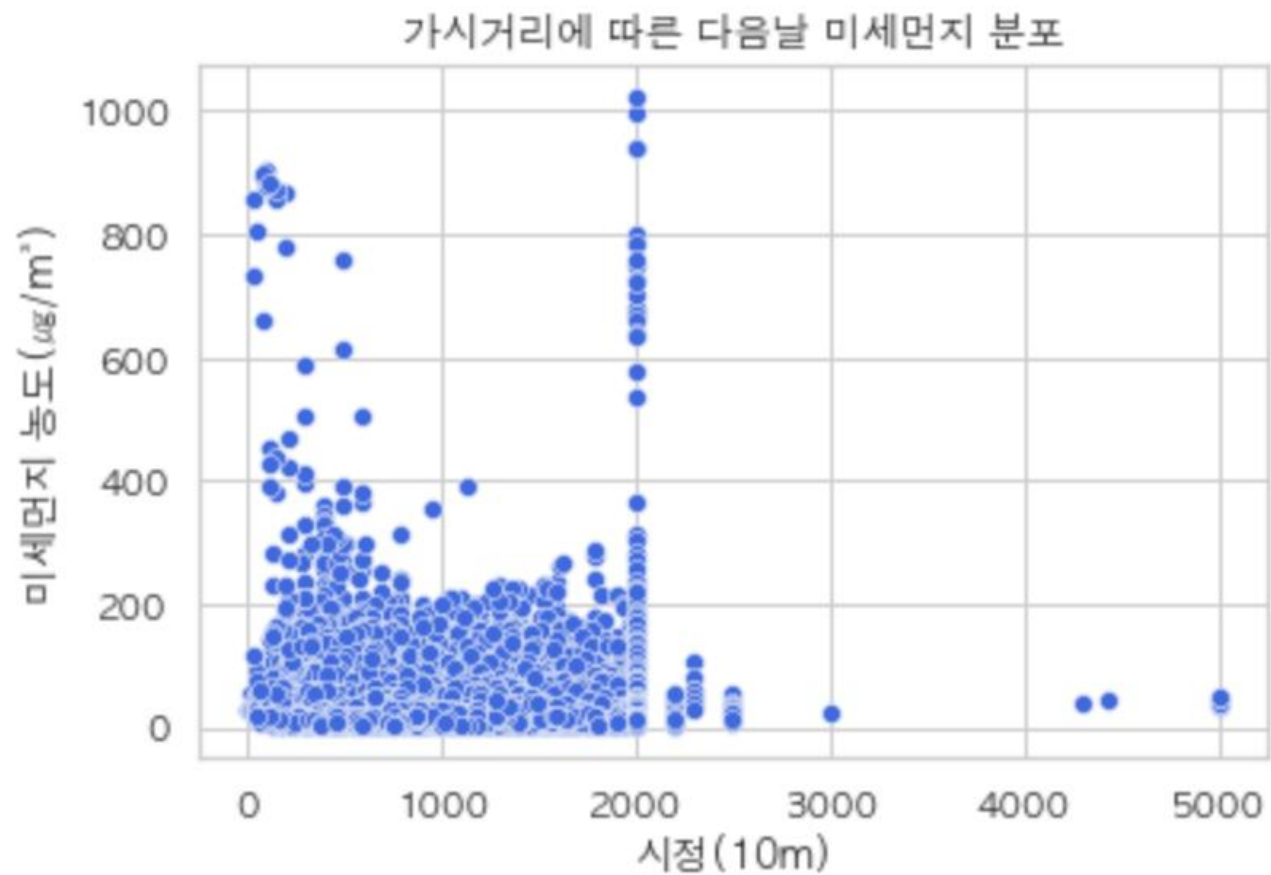




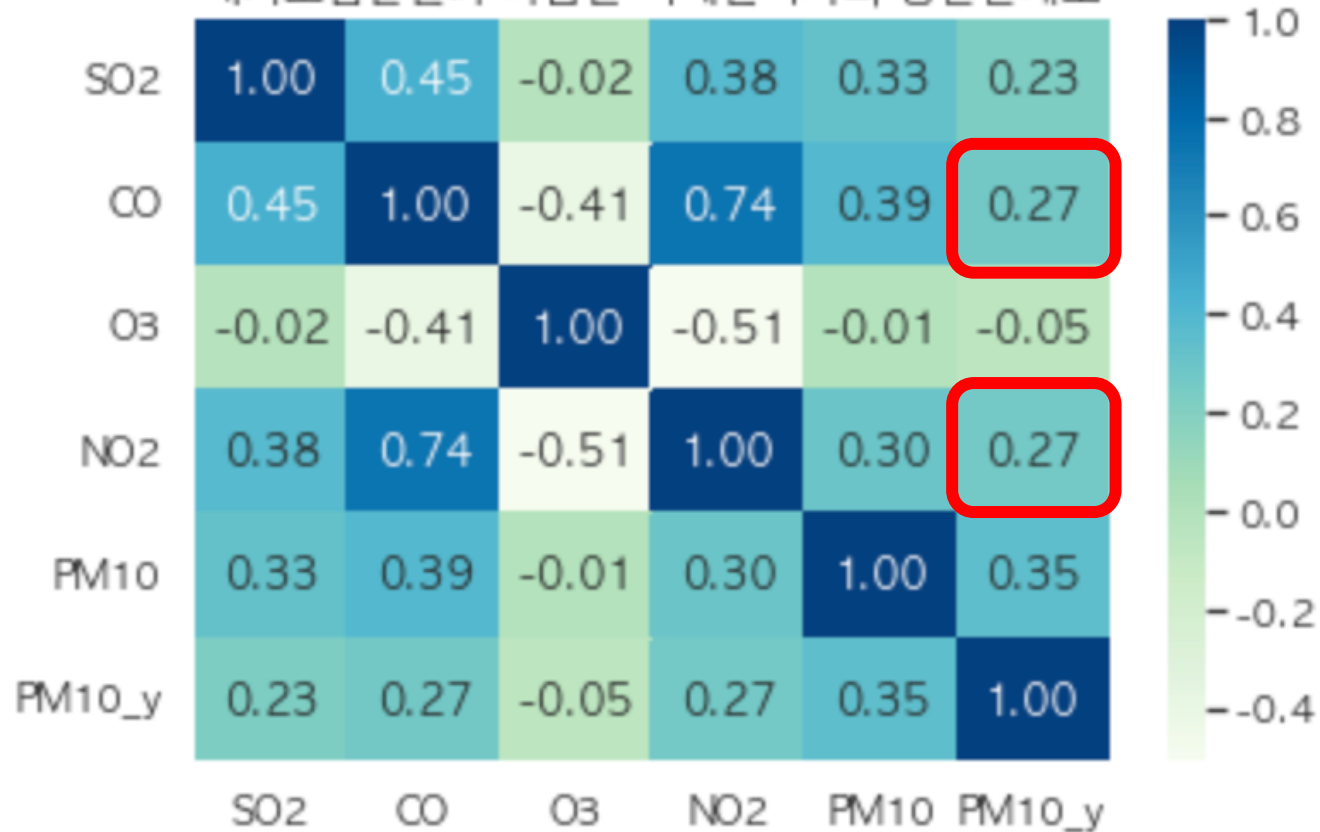
강수 세정효과  
: 강수와 미세먼지의 음의 상관 관계 (Oh, 2013)







대기오염물질과 다음날 미세먼지와의 상관관계도



SO2 = 이산화 황  
 CO = 일산화탄소  
 O3 = 오존  
 NO2 = 일산화질소  
 PM10 = 당일 미세먼지 농도  
 PM10\_y = 종속변수

기존 연구와 같이, 24시간 뒤 미세먼지와 NO2과의 양의 상관 관계를 발견.

그러나, CO도 같은 크기의 양의 상관 관계가 발견 되었다.

## ▼ ML Classification

KNeighborsClassifier 0.652

RandomForestClassifier 0.811

ExtraTreesClassifier 0.836

XGBClassifier 0.814

LGBMClassifier 0.788



83.6%

ACCURACY

## ▼ ML Regression

LinearRegressor 0.706

RandomForestRegressor 0.808

ExtraTreesRegressor 0.835

XGBRegressor 0.785

CatBoostRegressor 0.776



83.5%

ACCURACY

	실제값	Linear	RandomForest	ExtraTrees	XGBoost	LightGBM	CatBoost
count	20153.000000	20153.000000	20153.000000	20153.000000	20153.000000	20153.000000	20153.000000
mean	0.642386	-0.001153	-0.004653	-0.005268	-0.002418	-0.003475	-0.004744
std	0.624277	0.494079	0.389824	0.365905	0.409887	0.442784	0.414111
min	0.000000	-1.960904	-1.280000	-1.350000	-1.646268	-1.807272	-1.902160
25%	0.000000	-0.336507	-0.240000	-0.220000	-0.247000	-0.289529	-0.270490
50%	1.000000	-0.014090	-0.010000	-0.010000	-0.018083	-0.027323	-0.020394
75%	1.000000	0.309088	0.220000	0.200000	0.240946	0.272486	0.254808
max	3.000000	2.668727	2.420000	2.180000	2.301509	2.422898	2.123658

(실제 값 - 예측 값)으로 차이를 알아본 결과,  
 최솟값과 최댓값은 절댓값이 1 이상으로 나타났지만  
 25%, 75% 구간에서는 대체로 0.3 미만의 오차가 나타났음  
 이에 따라 예측 값을 반올림해서 종속변수에 맞춰 정확도를 구함

대체로 분류 모델에서 정확도가 높았지만, 큰 차이는 보이지 않았음  
 따라서 성능이 가장 좋았던 ExtraTreesClassifier를 베이스라인 모델로 선정



## 베이스라인모델 - Cross Validate

81.8%

## Cross Validate

0.818 -&gt; 0.821

교차검증 스코어 개선

## ExtraTreesClassifier

feature\_importances &gt; 0.0001

총 776개 피처 선별

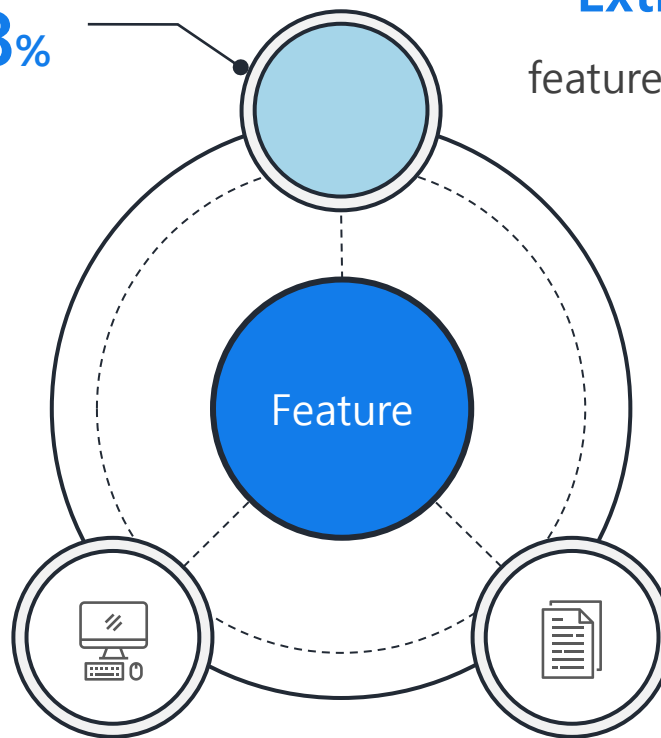
## POINT

총 1773개 피처로 이루어진 현상번호 피처 중 36개만

0.0001 이상의 중요도를 가짐

(서울, 수원, 인천, 백령도에 속한 일부 현상번호)

비 눈 황사 안개, 박무



### Hyperparameter Tuning

훈련용 데이터가 검증용 데이터보다  
스코어가 현저히 높게 나타나  
과대적합이 일어났다고 판단

결정트리의 개수를 늘리고 노드를 분할  
하기 위한 최소 향상도를 낮춤으로써  
과대적합이 방지되는 걸 확인했으나  
성능을 베이스라인 모델 이상으로  
끌어올리는 것은 컴퓨터 성능과  
긴 학습 시간 문제로 실패함

### Voting

분류기로 랜덤포레스트와  
엑스트라트리, XGBoost 분류모델 사용

하드 보팅과 소프트 보팅 모두  
성능 개선에 기여하지 못함

81.7%

ACCURACY

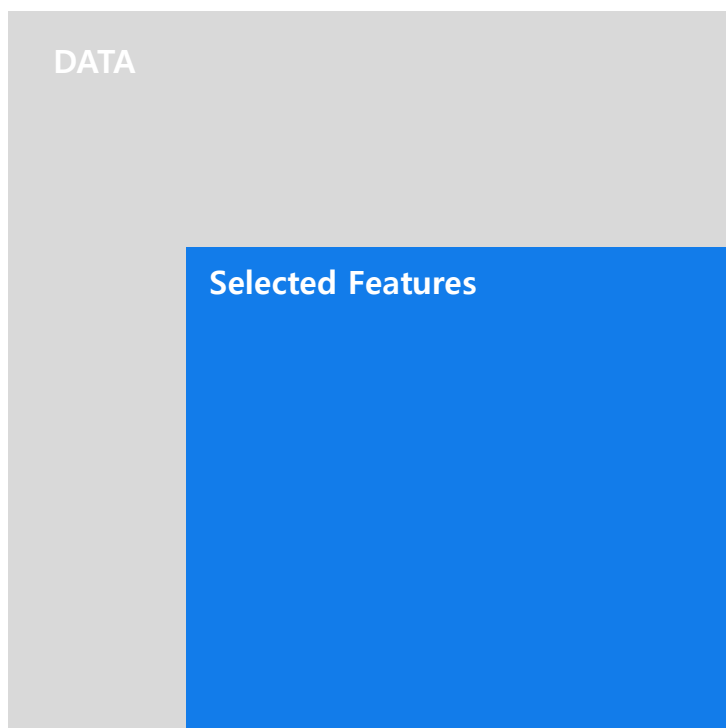
### Stacking

랜덤포레스트와 XGBoost, 엑스트라트리  
분류, 회귀 모델을 이용해  
도출한 예측 값을 이용해서  
GradientBoost의 분류 모델로  
예측 값을 구함

그 결과, 성능 개선이 크게 이루어짐.  
이에 따라 최종 모델에서는 더 많은  
모델을 스택킹으로 앙상블하기로 결정

85.3%

ACCURACY



전체 피처 **2528**개

### 피처 선정

랜덤포레스트의 feature importances를 이용  
중요도가 0인 피처 확인 후 제거  
선정된 피처 **951**개

## \* 성능지표

- 분류 : **Accuracy** (전체 데이터 수 중 예측 결과와 실제 값이 동일한 건수가 차지하는 비율)
- 회귀 : **RMSE** (실제값에서 예측값을 뺀 제곱의 합을 표본수로 나눈 뒤 그 값에 루트를 씌운 것)

## ▼ DL Classification

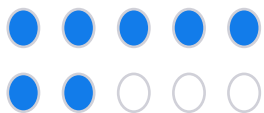
Model 1  
(Conv1D, Conv1D, Dense, Dense)  
(MaxPooling(3), Dropout(0.3))  
(epoch : 1000, optimizer : Adam) **0.765**

Model 2  
(epoch : 25) **0.701**

Model 3  
(Dense layer 1 제거) **0.771**

Model 4  
(epoch : 500) **0.756**

Model 5  
(dropout : 0.3 → 0.5) **0.752**



**77%**  
ACCURACY

## ▼ DL Regression

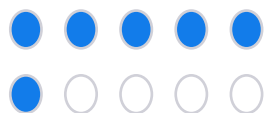
Model 1  
(Conv1D, Conv1D, Dense, Dense)  
(MaxPooling(3), Dropout(0.3))  
(epoch : 1000, optimizer : Adam) **0.785**  
RMSE : 0.417

Model 2  
(epoch : 25) **0.698**

Model 3  
(Dense layer 1 제거) **0.784**  
RMSE : 0.414

Model 4  
(epoch : 500) **0.781**

Model 5  
(dropout : 0.3 → 0.5) **0.758**



**78%**  
ACCURACY

## 구분 (회귀모델)

## Accuracy

## RMSE

Model1

0.785

0.417

Model3

0.784

0.414

차이

0.001

0.003

DL 모델의 분류와 회귀의 성능을 비교하고자 Accuracy 지표로 확인.

(머신러닝 베이스라인 모델 선정시와 동일하게

예측 값을 반올림해서 종속변수에 맞춰 정확도를 구함)

그 결과, 회귀모델의 Accuracy가 더 높아서 회귀모델로 결정

회귀모델 내에서는 Accuracy와 RMSE 지표를 확인.

Accuracy는 Model1이 0.001 높지만 RMSE는 Model3이 0.003 낮다.

(accuracy는 높을수록, RMSE는 낮을수록 성능이 우수)

따라서 회귀의 성능지표인 RMSE 기준으로 더 낮은 Model3을

베이스모델로 선정함.

\* 정확도는 전체 데이터 중에서 몇 개를 맞추었는가, 손실은 실제 정답과 모델이 예측한 값 사이의 차이이다. 따라서 두 지표는 서로 관련이 없다.

**1. Strides = 3**

MaxPooling 대신 Strides = 3을 사용  
 -> 최근에는 Maxpooling보다 Strides를 적용한 모델의 성능이 우수한 경우가 있어 많이 사용한다고 함. (Lee, 2020)

**2. Kernel size = 5**

특징을 한번에 학습하는 크기를 넓히고자 kernel\_size = 5 사용  
 -> 커널의 개수가 많아질수록 정확도가 늘어날 수 있으며 반대로 커널 개수가 일정이상 많아지면 실질적으로 학습에 참여하지 않는 잉여 커널이 생기게 되어 학습에 방해를 주므로 오히려 정확도가 낮아질 수 있다. (Kong & Jang, 2019)

**3. Batch size = 128**

Learning\_rate 0.0001에 최적화된 batch size 찾기  
 Batch\_size를 64에서 128로 변경  
 (Kong & Jang, 2019)  
 RMSE : 0.395

**4. RobustScaler**

기존에 MinMaxScaler를 사용,  
 RobustScaler로 변경  
 -> 평균과 분산 대신에 중간값과 사분위값을 사용,  
 아웃라이어 영향을 최소화  
 (Kang, 2020)

**최종 Model**

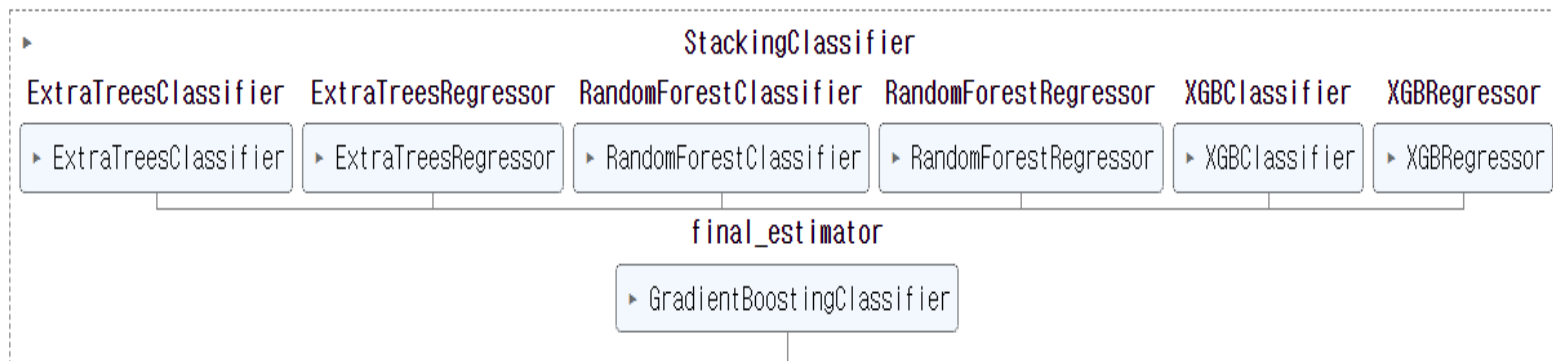
Layer (type)	Output Shape	Param #
conv1d_16 (Conv1D)	(None, 317, 64)	384
dropout_24 (Dropout)	(None, 317, 64)	0
conv1d_17 (Conv1D)	(None, 106, 64)	20544
dropout_25 (Dropout)	(None, 106, 64)	0
flatten_8 (Flatten)	(None, 6784)	0
dense_16 (Dense)	(None, 64)	434240
dropout_26 (Dropout)	(None, 64)	0
dense_17 (Dense)	(None, 1)	65
Total params: 455,233		
Trainable params: 455,233		
Non-trainable params: 0		

**82.9%**

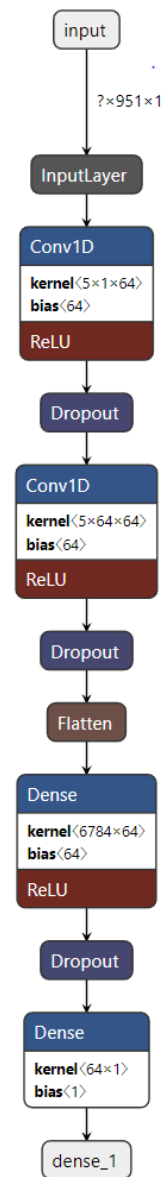
ACCURACY

## 모델 구조

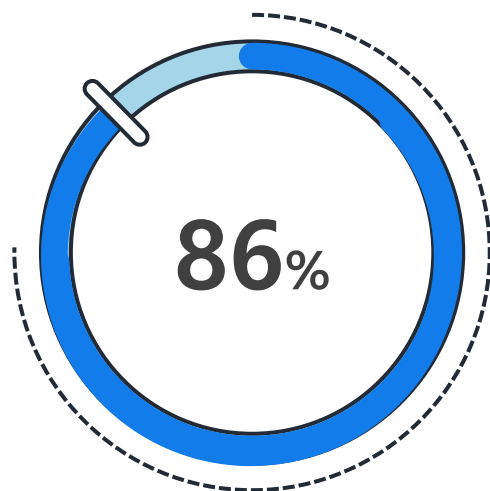
## 머신러닝



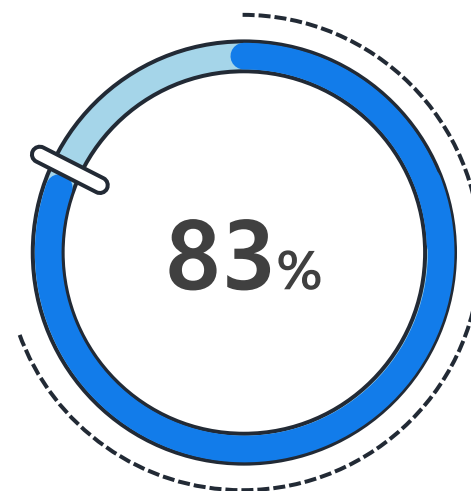
## 딥러닝



## 모델 비교

**MachineLearning**

ACCURACY : 0.858

**DeepLearning**

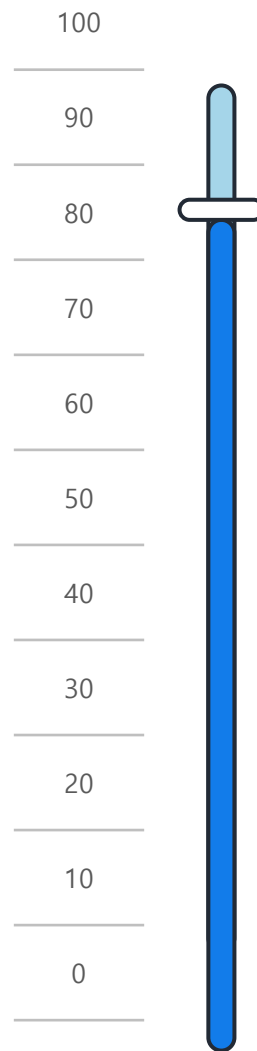
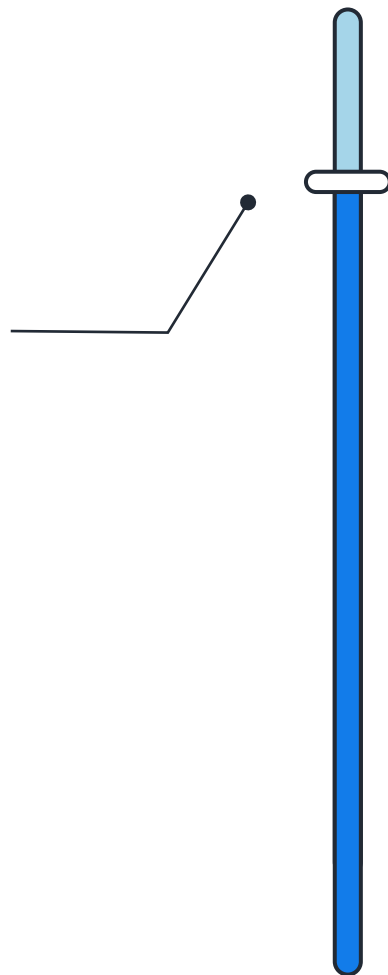
ACCURACY : 0.829

교차검증 결과

**ACCURACY 82%**

베이스라인 모델

ExtraTreesClassifier



교차검증 결과

**ACCURACY 86%**

최종 모델

StackingClassifier





목표 1

Tree 알고리즘의 Feature Importance를 통해  
가장 예측력이 높은 요소 선정



목표 2

스태킹을 이용해서 회귀과 분류 모델을  
앙상블한 모델이 가장 정확도가 높았다



익일 미세먼지 예측에 주요하게 영향을 주는 피처 (영향력 순)

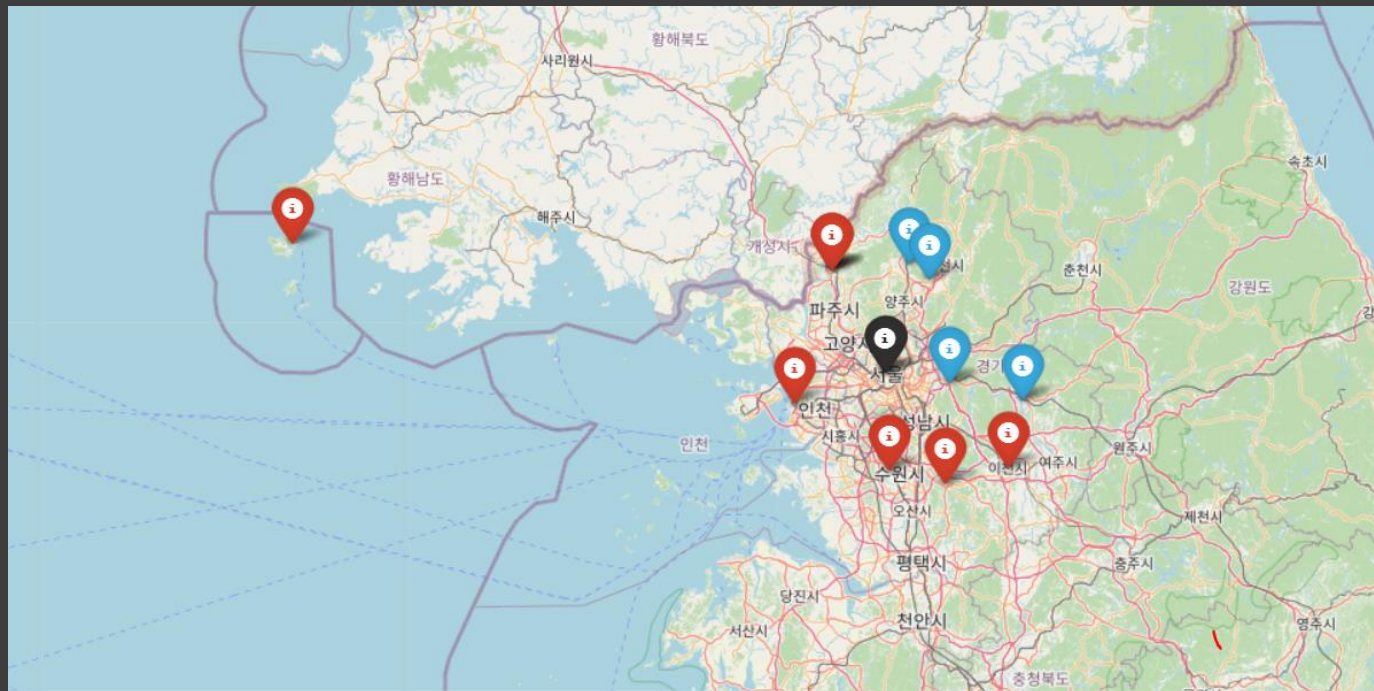
1. 파주의 이슬점 온도
2. 경기도 선단동의 PM10, 경기도 신장동의 CO, 동두천의 이슬점 온도, 양평의 이슬점 온도
3. 경기도 천천동의 CO, 경기도 김량장동의 PM10, 백령도의 이슬점 온도, 인천의 기온, 수원의 기온, 수원의 이슬점 온도, 파주의 증기압, 양평의 증기압

기상환경 요소는 백령도, 인천 중구, 파주, 강변북로의 비, 눈, 안개, 황사, 백무



목표 1 결과

## 예측력 높은 관측소 위치





### 목표 1 - 가설 1 결과 해석

- 가설 1: 기온과 일산화질소(NO<sub>2</sub>) 수치가 미세먼지 농도 예측에 가장 큰 영향력을 미칠 것이다.
- 가설 1과 같이 기온에 관련된 변수들이 예측력이 높았다.
- 가설 1과 다르게 일산화탄소(CO) 수치가 예측력이 높았다.
  - 고온 연소 공정과 화학물질 제조 공정에서 생성되는 NO<sub>2</sub>보다 연료 연소, 자동차 운행에서 생성되는 CO가 다음날 종로구 미세먼지 예측에 더 큰 영향력을 가지고 있는 것으로 볼 수 있다.
  - 이는 종로구의 특성상, 큰 공장이 없고 유동인구와 자동차 운행량이 많아, 공장에서 생성되는 NO<sub>2</sub> 보다 CO가 예측에 직접적인 영향력을 미치는 것으로 판단 된다. (Park et al. 2011)



### 목표 1 – 가설2 결과 해석

- 가설2 : 편서풍의 영향으로 인천과 경기 서북부 지역들이 미세먼지 농도 예측에 가장 큰 영향력을 미칠 것이다.
- 가설2와 다르게 인천과 경기 서북부, 그리고 경기 동남부 지역들이 예측력이 높았다.
  - 이는 경기 남부의 택지 개발과 소규모 사업장에서 배출되는 오염물질, 경기 남부와 인접한 충남지역에 위치한 발전소 등에서 배출된 공해가 영향을 미친 것으로 판단된다. (Lee, 2019)



## 목표 2 결과

목표 2 결과

- 스택킹을 이용해서 다양한 모델을 앙상블한 모델이 가장 좋은 성능
- 랜덤포레스트와 엑스트라트리, xgb의 분류와 회귀 모델을 estimator로 사용
- Final estimator로 그래디언트 부스팅을 사용한 스택킹 모델 생성

목표 2 해설

- 가설 : 지리적 특성을 유지하여 분석 할 수 있는 CNN 기반 딥러닝 모델이 정확도가 가장 높을 것으로 예상된다.
- 가설과 다르게 CNN 보다 머신러닝 알고리즘 중 Tree 알고리즘이 더 높은 성능을 보임.
- 이는 이 실험의 데이터 양이 적고 피처가 적은 것으로 판단 된다. (Kim, 2021)



## 웹 결과

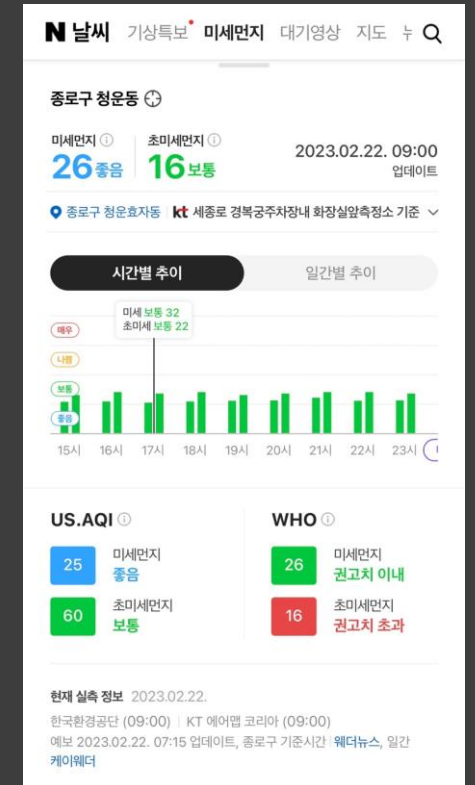
결과예측

보통

## 네이버(케이웨더) 예보



## 네이버(케이웨더) 실제 농도





목표 3 결과

결과 도출 : 7초 소요

<https://seoyoonjoannechang.github.io/>