



미세먼지 발생 영향인자 분석 및 발생량 예측

A반 이소희

분석 배경

미세먼지란?

아황산가스, 질소 산화물, 납, 오존, 일산화 탄소 등을 포함하는 대기오염 물질
자동차, 공장, 조리 과정 등에서 발생하여
대기 중 장기간 떠다니는 입경 $10\mu\text{m}$ 이하의 미세한 먼지

미세먼지가 우리 삶에 미치는 영향

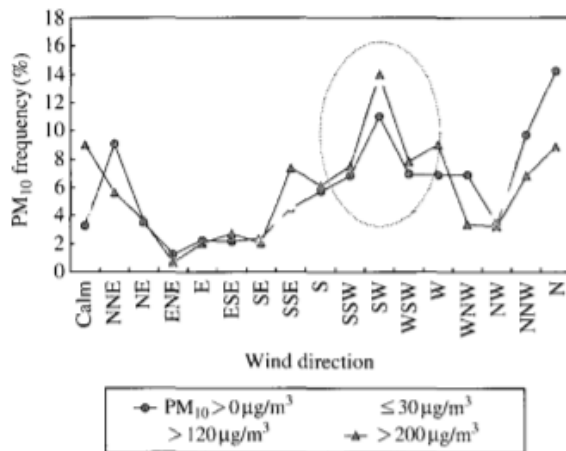
- 건강
미세먼지에 장기간 노출되게 되면 사람의 인체 건강에 상당히 안 좋은 영향을 끼친다.
특히 유아나 노인들의 경우 더 위험할 수 있다.
- 환경
- 경제

문헌 조사 – 기상정보

신문기 등이 진행한 기상인자가 미세먼지 농도에 미치는 영향 연구에서는 풍향, 풍속, 상대습도, 일기유형(기온) 등을 살펴봄으로써 미세먼지 현상의 원인이 무엇인지 분석하였다.

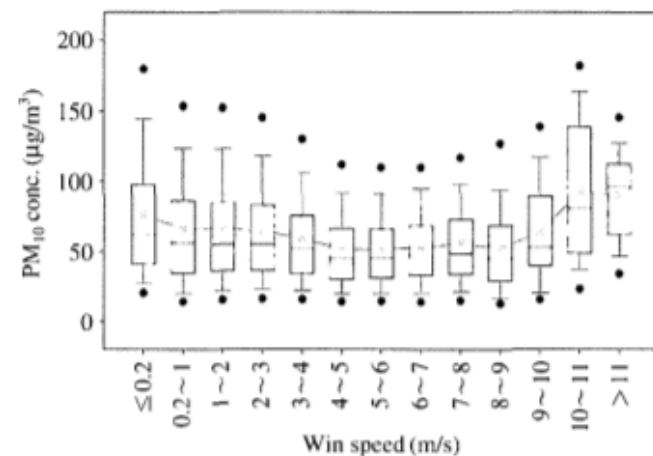
분석결과를 간단히 정리하면 다음과 같았다.

1. 풍향



서풍계열 바람이 불 때 미세먼지의 농도가 높다.

2. 풍속

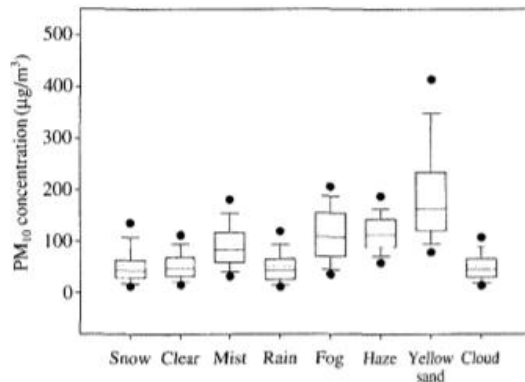


0~6 m/s 구간은 무풍시 가장 높은 농도, 풍속이 증가함에 따라 농도 감소, 6m/s 초과 구간에서는 증가함에 따라 농도도 상승

문헌 조사 – 기상정보

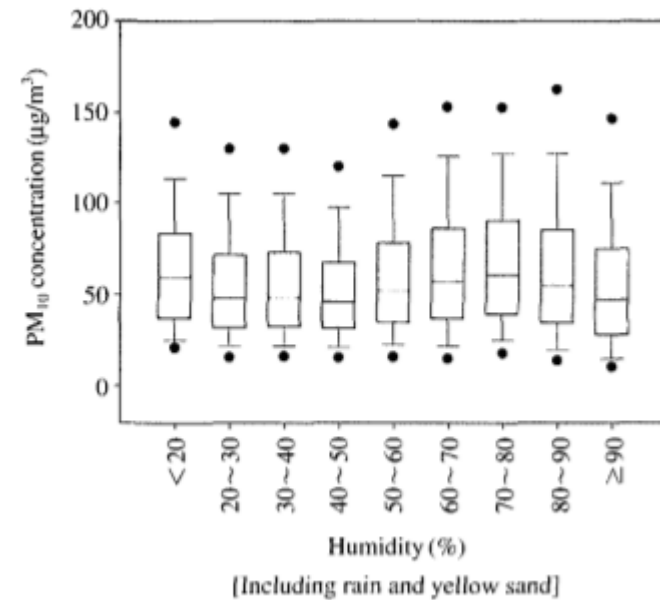
3. 일기 유형

Weather condition	N	Mean ($\mu\text{g}/\text{m}^3$)	S.D. ($\mu\text{g}/\text{m}^3$)
Clear	18,666	54.0	31.0
Cloud	19,591	51.2	30.4
Rain	6,557	50.6	34.7
Snow	791	53.1	38.2
Mist	13,583	92.3	47.2
Fog	955	114.1	54.5
Haze	140	117.4	40.2
Yellow sand	875	196.0	113.4



황사와 안개가 낀 날씨일 때 미세먼지 농도가 높다.

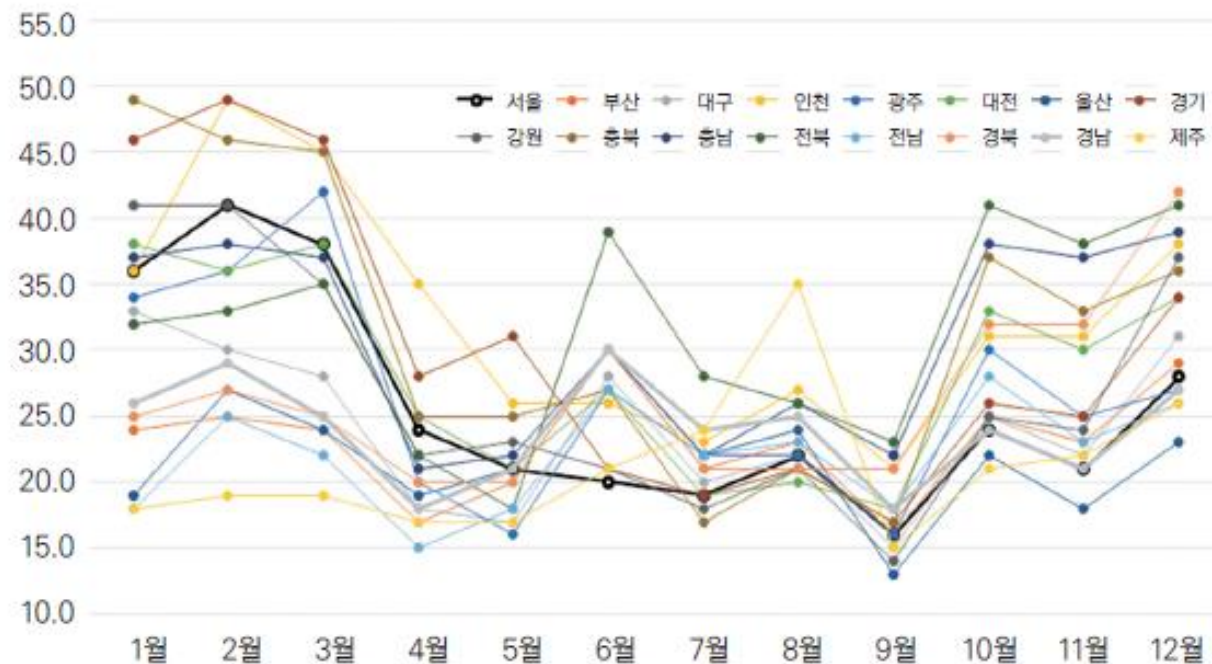
4. 습도



60~90%일 때 미세먼지 농도가 높다.
높은 습도가 대기중의 황산염과 질산염 등의
2차 먼지 생성을 촉진하기 때문이다.

문헌 조사 – 기상정보

2015년 월별 초 미세먼지 현황은 다음과 같았다. 보통 가을부터 초봄까지의 농도가 높다는 것을 확인할 수 있었다. 즉, 따뜻할 때보다 추울 때 미세먼지의 농도가 높다.



자료 : 에어코리아 홈페이지(<http://www.airkorea.or.kr/index>)

문헌 조사 – 대기오염

환경부의 국내 환경동향 보고에서 미세먼지의 발생원인을 다음과 같이 분석하고 있다.

1) 중국 발 황사와 대기 오염물질 문제

국내 미세먼지의 43% 정도가 중국에서 날아오는 것.

공장의 매연이나 자동차의 배기가스, 사막의 황사 등

2) 국내 발 미세먼지 문제

국내에서 생성된 미세먼지의 비중 50~70%, 서울의 미세먼지 가운데 51%는 국내에서 생성된 것

국내 미세먼지 배출 원인 중 68.2%는 화석연료를 연소시키면서 배출하는 입자들

분석 목표

문헌조사 결과 여러 요인들이 미세먼지의 농도에 영향을 미치고 있음을 알 수 있었다.
하지만 이 문헌들은 특정 시점, 특정 공간의 조사결과를 기반으로 분석한 것이기 때문에
우리 데이터도 이러한 문헌의 **분석결과와 동일한지 확인할 필요**가 있다.
따라서 이번 분석의 목표를 다음과 같이 정하였다.

| 목표 |

기상정보와 오염농도의 데이터를 활용해 문헌조사를 통해 세운 가설이 맞는지 확인
분석 후 미세먼지 문제 해결책 수립

가설 설정

앞의 문헌 조사 결과들을 바탕으로 다음과 같이 가설을 세웠다.

가설 1) 대기오염

- 화석연료를 연소시키면서 나오는 입자들이 미세먼지의 주요 원인 중 하나이다.
- 이산화질소, 일산화탄소, 아황산가스, 오존 농도가 모두 높으면 미세먼지가 많이 발생할 것이다.

→ 대기오염 정도가 높을수록 미세먼지가 많이 발생할 것이다.

가설 2) 기상정보

- 기온, 기압, 습도가 낮을수록 미세먼지가 많이 발생할 것이다.

→ 여름보다 겨울에 미세먼지가 많이 발생할 것이다.

계획 수립

분석을 시작하기 앞서 간단히 계획을 수립하였다.

- 1) 데이터 가져오기
- 2) 데이터 품질 확인 – 결측치 / 이상치 / 분포 확인
- 3) 데이터 정제
- 4) 가설 검정
- 5) 상관관계 분석
- 6) 회귀분석 모델링
- 7) 모델 평가
- 8) 모델 개선안

데이터 확인

데이터는 14개의 변수로 구성된 366개 데이터로 이루어져 있다.

```
df_raw = pd.read_csv("../data/AIR_POLLUTION.csv")
df_raw
```

	MeasDate	PM10	O3	NO2	CO	SO2	TEMP	RAIN	WIND	WIND_DIR	HUMIDITY	ATM_PRESS	SNOW	CLOUD
0	2019-07-01	29.0	0.054	0.021	0.5	0.003	24.03	0.00	2.30	249	63.2	995.1	0.0	5.70
1	2019-07-02	26.0	0.053	0.020	0.5	0.003	24.29	0.00	2.26	265	63.2	998.6	0.0	3.83
2	2019-07-03	30.0	0.042	0.023	0.4	0.003	24.18	0.00	1.79	280	65.3	998.3	0.0	6.29
3	2019-07-04	28.0	0.034	0.026	0.4	0.003	25.35	0.00	2.04	263	58.6	996.6	0.0	2.54
4	2019-07-05	29.0	0.045	0.035	0.5	0.003	27.30	0.00	1.45	175	45.5	993.5	0.0	3.92
...
361	2020-06-26	19.0	0.039	0.016	0.4	0.003	21.66	0.41	3.12	228	84.0	996.0	0.0	8.73
362	2020-06-27	22.0	0.044	0.017	0.4	0.004	23.94	0.00	1.93	217	69.8	995.8	0.0	6.21
363	2020-06-28	27.0	0.044	0.009	0.4	0.003	25.03	0.00	2.35	283	71.3	994.7	0.0	2.63
364	2020-06-29	36.0	0.026	0.028	0.6	0.003	24.06	1.26	2.48	103	75.5	992.9	0.0	7.58
365	2020-06-30	6.0	0.039	0.009	0.3	0.002	20.60	5.60	4.00	50	92.0	983.8	0.0	10.00

366 rows × 14 columns

데이터 확인 – 결측치 / 이상치 / 데이터분포 확인

데이터 타입을 확인해보니 MeasDate 변수를 제외한 모든 데이터가 연속형 데이터이다.
MeasDate는 날짜를 나타내는 식별자이기 때문에 학습에 혼동을 줄 수 있으므로 삭제한다.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   MeasDate    366 non-null   object
1   PM10        365 non-null   float64
2   O3          365 non-null   float64
3   NO2         365 non-null   float64
4   CO          311 non-null   float64
5   SO2         365 non-null   float64
6   TEMP        366 non-null   float64
7   RAIN        366 non-null   float64
8   WIND        366 non-null   float64
9   WIND_DIR    366 non-null   int64
10  HUMIDITY    366 non-null   float64
11  ATM_PRESS   366 non-null   float64
12  SNOW        366 non-null   float64
13  CLOUD       366 non-null   float64
dtypes: float64(12), int64(1), object(1)
memory usage: 40.2+ KB
```

데이터 확인 – 결측치 / 이상치 / 데이터분포 확인

결측치가 있어 확인해본 결과 각 변수 당 결측치 개수는 오른쪽 그림과 같았다.
PM10은 미세먼지 농도로 목표변수이기 때문에 제거하여 처리하였고,
나머지는 데이터에 영향을 주지 않도록 평균값으로 채워주기로 결정하였다.

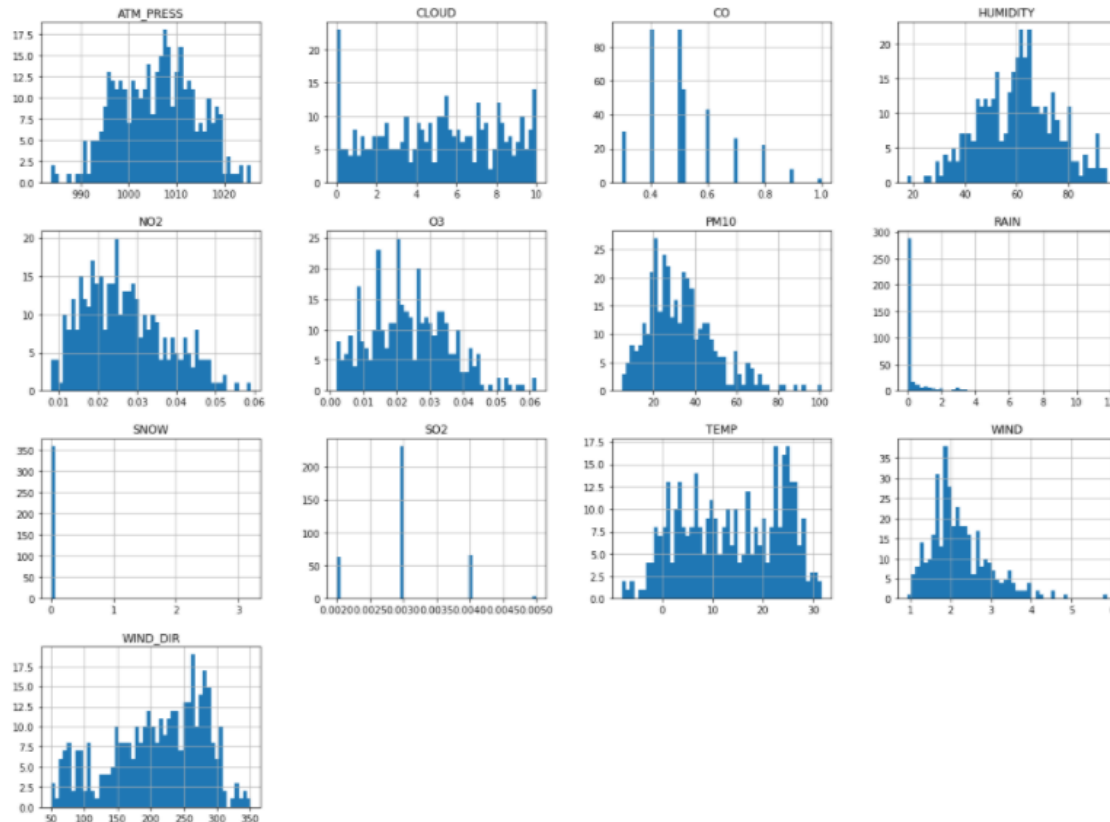
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   MeasDate    366 non-null   object
1   PM10        365 non-null   float64
2   O3          365 non-null   float64
3   NO2         365 non-null   float64
4   CO          311 non-null   float64
5   SO2         365 non-null   float64
6   TEMP        366 non-null   float64
7   RAIN        366 non-null   float64
8   WIND        366 non-null   float64
9   WIND_DIR    366 non-null   int64
10  HUMIDITY    366 non-null   float64
11  ATM_PRESS   366 non-null   float64
12  SNOW        366 non-null   float64
13  CLOUD       366 non-null   float64
dtypes: float64(12), int64(1), object(1)
memory usage: 40.2+ KB
```

MeasDate	0	
PM10	1	→ 제거
O3	1	
NO2	1	
CO	55	→ 평균값으로 채우기
SO2	1	
TEMP	0	
RAIN	0	
WIND	0	
WIND_DIR	0	
HUMIDITY	0	
ATM_PRESS	0	
SNOW	0	
CLOUD	0	
dtype:	int64	

데이터 확인 – 결측치 / 이상치 / 데이터분포 확인

변수 별로 데이터 분포를 확인해보았다.

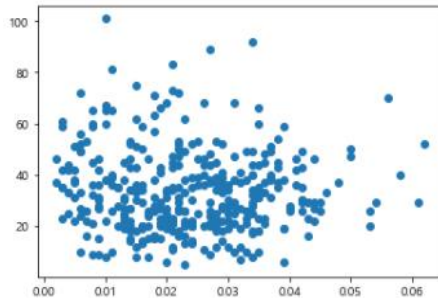
대체적으로 정규분포를 따르고 있으나 눈이나 비의 데이터는 눈이 오는 날이 적어 0의 분포가 많았다.



그래프 분석 및 가설 검정

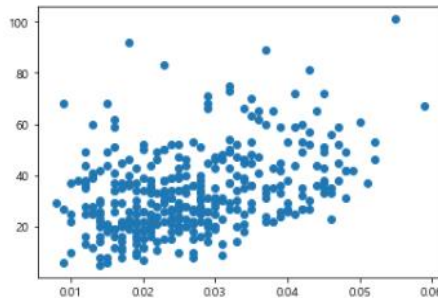
가설 1) 대기정보 : 대기오염 정도가 높을수록 미세먼지가 많이 발생할 것이다.
산점도로 대기오염 정도에 따른 미세먼지 발생량을 확인해 선형관계가 있는지 확인하였다.

오존의 농도



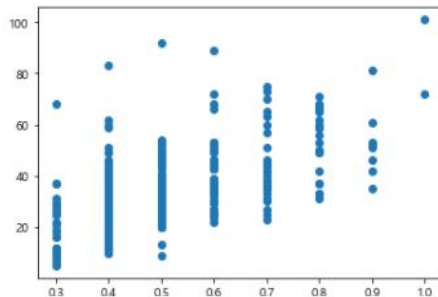
오존의 농도는 미세먼지와 선형관계가 있어보이지 않는다.

NO2



미세하게 선형관계가 관찰된다.
NO2농도가 높을수록 미세먼지 발생량이 늘어난다.

CO

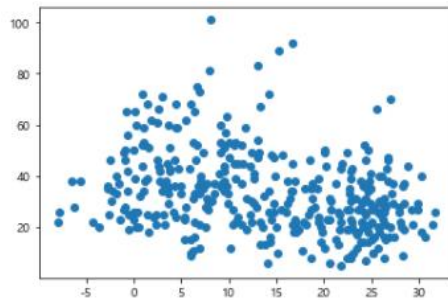


마찬가지로 미세하게 선형관계가 관찰된다.
CO농도가 높을수록 미세먼지 발생량이 증가한다.

그래프 분석 및 가설 검정

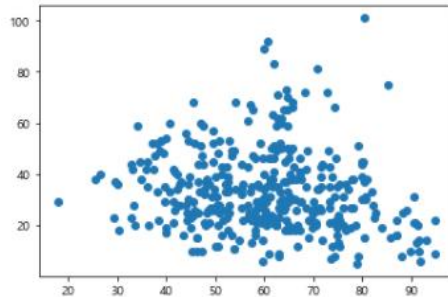
가설 2) 기상정보 : 기온, 기압, 습도가 낮을수록 미세먼지가 많이 발생할 것이다.
마찬가지로 산점도로 선형관계가 있는지 분석해 보았다.

기온



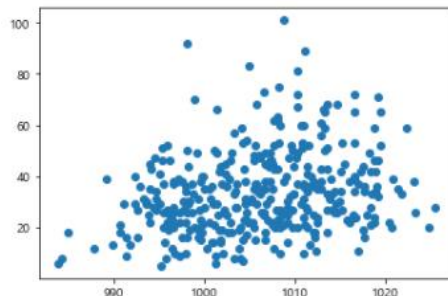
선형관계가 잘 보이지 않는다.

습도



선형관계가 잘 보이지 않는다.

기압



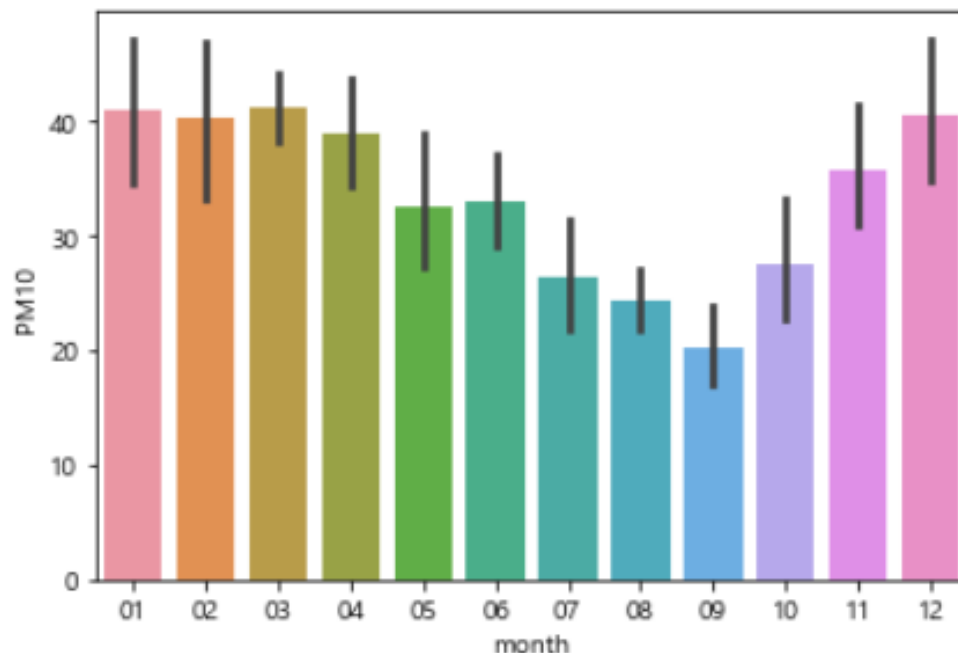
미세하게 선형관계가 관찰된다.
기압이 높을수록 미세먼지의 발생량이 증가할까?

그래프 분석 및 가설 검정

가설 2) 기상정보 : 기온, 기압, 습도가 낮을수록 미세먼지가 많이 발생할 것이다.

다음은 월별 미세먼지 농도의 변화를 나타내는 그래프이다.

추울 때보다 따뜻할 때 미세먼지의 농도가 낮다는 가설을 확인하고자 하였다.



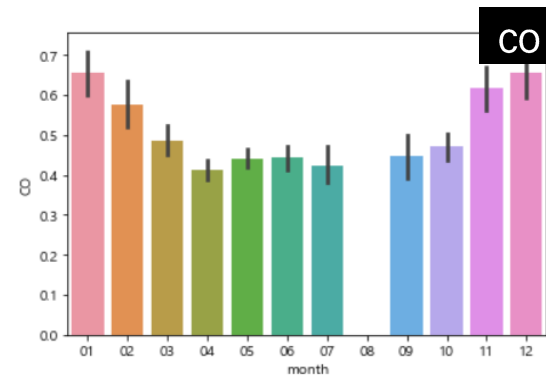
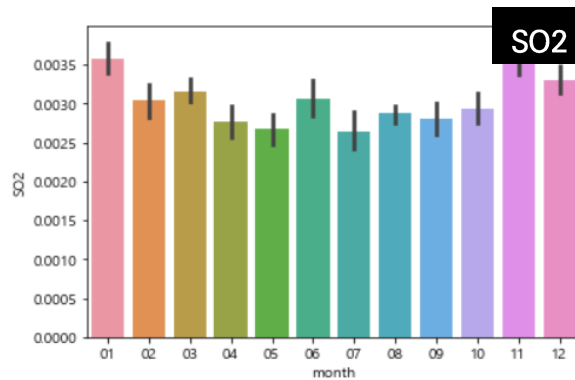
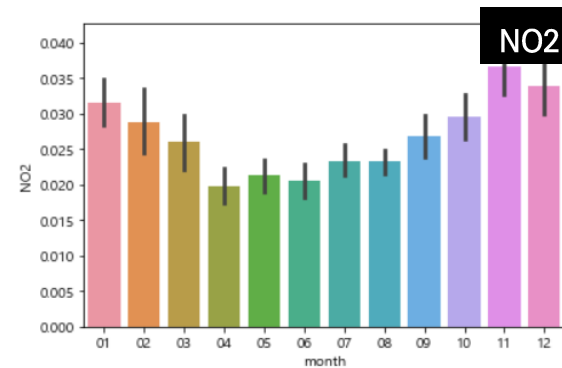
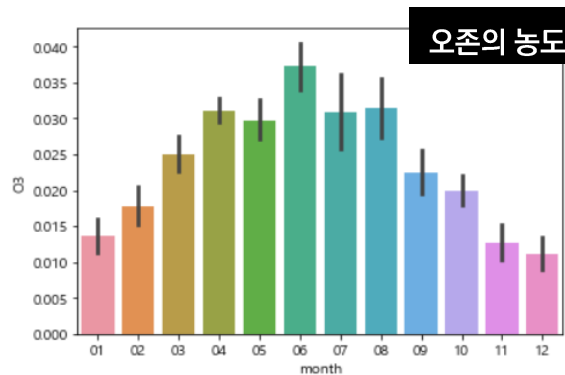
늦가을부터 초여름까지 미세먼지 농도가 높다. 가설이 참일 확률이 높다.

그래프 분석 및 가설 검정

추가적으로 날씨에 따라서 오존의 농도나 대기오염 물질의 변화가 어떠한지 알아보았다.

오존은 따뜻할 때 더 농도가 높고, NO₂, CO는 따뜻할 때 농도가 낮다.

SO₂는 날씨에 영향을 받지 않는 것으로 보인다.



(CO는 8월의 데이터가 없어 비어 있다.)

상관관계 확인 - 목표변수

목표변수인 PM10과 다른 설명변수 간의 관계를 살펴보니 다음과 같았다.

CO와 SO2, NO2, 대기압이 높은 양의 상관관계를 보이고 있음을 확인할 수 있었다.

→ 대기 오염이 심해지면 미세먼지의 농도도 높아진다.

온도는 미세먼지와 음의 상관관계를 가지고 있다.

→ 온도가 낮을 수록 미세먼지 농도가 높다.

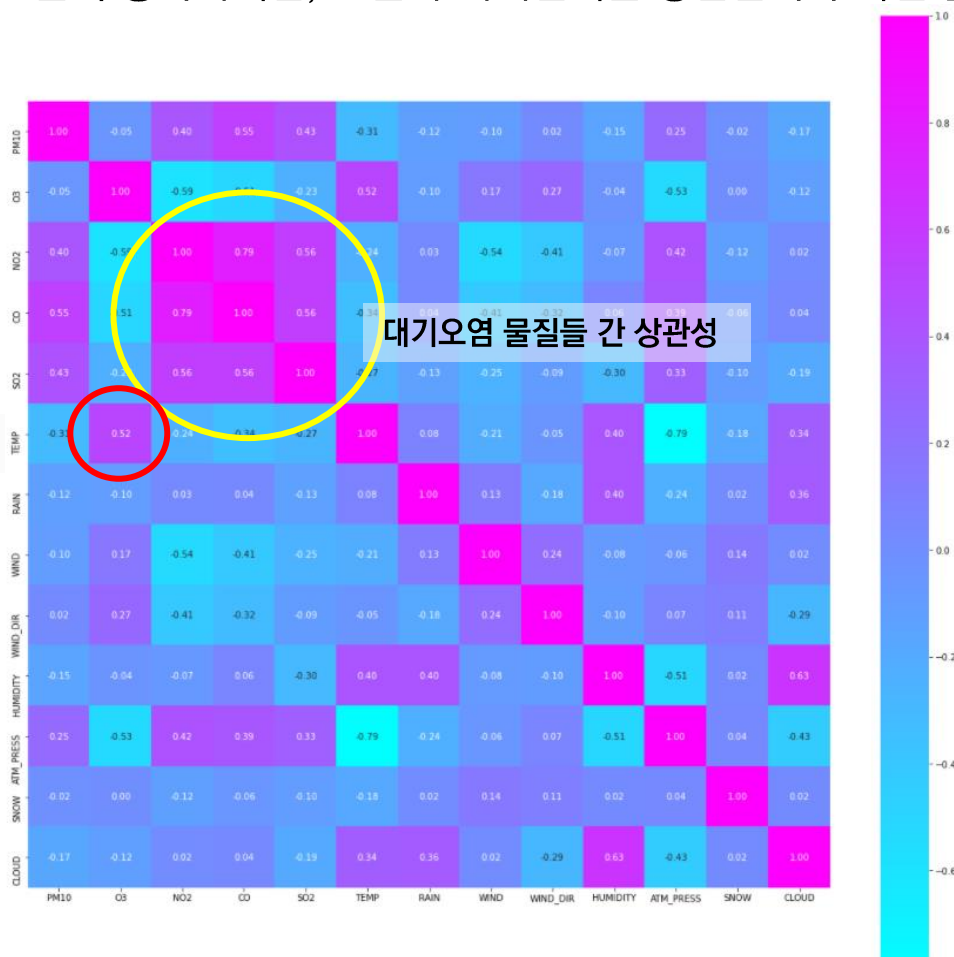
PM10	1.000
CO	0.548
SO2	0.429
NO2	0.396
ATM_PRESS	0.253
WIND_DIR	0.020
SNOW	-0.020
O3	-0.052
WIND	-0.100
RAIN	-0.121
HUMIDITY	-0.150
CLOUD	-0.172
TEMP	-0.310

Name: PM10, dtype: float64

상관관계 확인 - 설명변수

설명변수 간 상관관계를 확인해보았을 때, 대기 오염 물질들 간의 양의 상관관계가 강하다.
기온이 높을수록 오존이 증가하지만, 오존과 미세먼지는 상관관계가 약한 음의 상관관계를 가진다.

기온과 오존의 상관성



- 온도와 비는 습도와 양의 상관관계.
- 온도와 대기압은 강한 음의 상관관계.

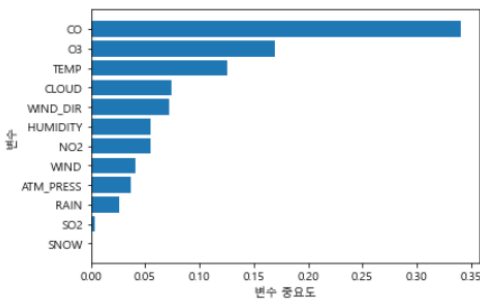
모델링 - 다중선형회귀

다중 선형 회귀의 결과를 요약하면 다음과 같다.

1. 모델의 설명력은 약 44%
2. 모델은 통계적으로 유의미
3. 변수별로 목표변수에 미치는 영향은 CO, O3, 기온, NO2, 대기압, 풍향 순으로 높으며 통계적으로 유의하다.

OLS Regression Results						
Dep. Variable:	PM10	R-squared:	0.456			
Model:	OLS	Adj. R-squared:	0.438			
Method:	Least Squares	F-statistic:	24.57			
Date:	Fri, 05 Mar 2021	Prob (F-statistic):	1.05e-39			
Time:	21:45:51	Log-Likelihood:	-1411.9			
No. Observations:	364	AIC:	2850.			
Df Residuals:	351	BIC:	2900.			
Df Model:	12					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	33.5481	0.625	53.707	0.000	32.320	34.777
O3	6.6617	1.035	6.438	0.000	4.626	8.697
NO2	5.2717	1.398	3.770	0.000	2.522	8.022
CO	7.9854	1.180	6.765	0.000	5.664	10.307
SO2	0.7485	0.870	0.861	0.390	-0.962	2.459
TEMP	-6.8236	1.326	-5.144	0.000	-9.432	-4.215
RAIN	-1.3593	0.724	-1.876	0.061	-2.784	0.066
WIND	1.9923	0.870	2.290	0.023	0.281	3.704
WIND_DIR	2.4251	0.767	3.162	0.002	0.917	3.933
HUMIDITY	0.0988	1.019	0.097	0.923	-1.904	2.102
ATM_PRESS	-4.0091	1.342	-2.986	0.003	-6.649	-1.369
SNOW	-0.8159	0.658	-1.240	0.216	-2.110	0.478
CLOUD	-0.1779	0.903	-0.197	0.844	-1.954	1.598
Omnibus:	84.273	Durbin-Watson:	1.281			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	207.910			
Skew:	1.124	Prob(JB):	7.13e-46			
Kurtosis:	5.942	Cond. No.	5.91			

모델링 – DT, RF, GB

모델	Decision Tree	Random Forest	Gradient Boosting																																																																														
정확도 (기본 모델 기준)	Training : 100% Test : 29%	Training : 61.5% Test : 42.9%	Training : 92.4% Test : 52.9%																																																																														
설명 변수 중요도	 <table><thead><tr><th>변수</th><th>중요도</th></tr></thead><tbody><tr><td>CO</td><td>0.45</td></tr><tr><td>O3</td><td>0.18</td></tr><tr><td>CLOUD</td><td>0.15</td></tr><tr><td>TEMP</td><td>0.14</td></tr><tr><td>ATM_PRESS</td><td>0.08</td></tr><tr><td>WIND_DIR</td><td>0.06</td></tr><tr><td>HUMIDITY</td><td>0.02</td></tr><tr><td>SNOW</td><td>0.01</td></tr><tr><td>WIND</td><td>0.01</td></tr><tr><td>RAIN</td><td>0.01</td></tr><tr><td>SO2</td><td>0.01</td></tr><tr><td>NO2</td><td>0.01</td></tr></tbody></table>	변수	중요도	CO	0.45	O3	0.18	CLOUD	0.15	TEMP	0.14	ATM_PRESS	0.08	WIND_DIR	0.06	HUMIDITY	0.02	SNOW	0.01	WIND	0.01	RAIN	0.01	SO2	0.01	NO2	0.01	 <table><thead><tr><th>변수</th><th>중요도</th></tr></thead><tbody><tr><td>CO</td><td>0.34</td></tr><tr><td>O3</td><td>0.18</td></tr><tr><td>TEMP</td><td>0.14</td></tr><tr><td>CLOUD</td><td>0.08</td></tr><tr><td>WIND_DIR</td><td>0.07</td></tr><tr><td>HUMIDITY</td><td>0.06</td></tr><tr><td>NO2</td><td>0.05</td></tr><tr><td>WIND</td><td>0.04</td></tr><tr><td>ATM_PRESS</td><td>0.03</td></tr><tr><td>RAIN</td><td>0.02</td></tr><tr><td>SO2</td><td>0.01</td></tr><tr><td>SNOW</td><td>0.01</td></tr></tbody></table>	변수	중요도	CO	0.34	O3	0.18	TEMP	0.14	CLOUD	0.08	WIND_DIR	0.07	HUMIDITY	0.06	NO2	0.05	WIND	0.04	ATM_PRESS	0.03	RAIN	0.02	SO2	0.01	SNOW	0.01	 <table><thead><tr><th>변수</th><th>중요도</th></tr></thead><tbody><tr><td>CO</td><td>0.31</td></tr><tr><td>O3</td><td>0.14</td></tr><tr><td>CLOUD</td><td>0.12</td></tr><tr><td>TEMP</td><td>0.09</td></tr><tr><td>WIND_DIR</td><td>0.08</td></tr><tr><td>WIND</td><td>0.07</td></tr><tr><td>NO2</td><td>0.06</td></tr><tr><td>HUMIDITY</td><td>0.05</td></tr><tr><td>ATM_PRESS</td><td>0.04</td></tr><tr><td>RAIN</td><td>0.03</td></tr><tr><td>SO2</td><td>0.02</td></tr><tr><td>SNOW</td><td>0.01</td></tr></tbody></table>	변수	중요도	CO	0.31	O3	0.14	CLOUD	0.12	TEMP	0.09	WIND_DIR	0.08	WIND	0.07	NO2	0.06	HUMIDITY	0.05	ATM_PRESS	0.04	RAIN	0.03	SO2	0.02	SNOW	0.01
변수	중요도																																																																																
CO	0.45																																																																																
O3	0.18																																																																																
CLOUD	0.15																																																																																
TEMP	0.14																																																																																
ATM_PRESS	0.08																																																																																
WIND_DIR	0.06																																																																																
HUMIDITY	0.02																																																																																
SNOW	0.01																																																																																
WIND	0.01																																																																																
RAIN	0.01																																																																																
SO2	0.01																																																																																
NO2	0.01																																																																																
변수	중요도																																																																																
CO	0.34																																																																																
O3	0.18																																																																																
TEMP	0.14																																																																																
CLOUD	0.08																																																																																
WIND_DIR	0.07																																																																																
HUMIDITY	0.06																																																																																
NO2	0.05																																																																																
WIND	0.04																																																																																
ATM_PRESS	0.03																																																																																
RAIN	0.02																																																																																
SO2	0.01																																																																																
SNOW	0.01																																																																																
변수	중요도																																																																																
CO	0.31																																																																																
O3	0.14																																																																																
CLOUD	0.12																																																																																
TEMP	0.09																																																																																
WIND_DIR	0.08																																																																																
WIND	0.07																																																																																
NO2	0.06																																																																																
HUMIDITY	0.05																																																																																
ATM_PRESS	0.04																																																																																
RAIN	0.03																																																																																
SO2	0.02																																																																																
SNOW	0.01																																																																																

요약

1) 모델 별 핵심 인자

다중선형회귀

CO
O3
기온
NO2
대기압
풍향

의사결정나무

CO
O3
구름 양
기온
대기압
풍향

랜덤 포레스트

CO
O3
구름 양
기온
풍향
풍속

그라디언트 부스팅

CO
O3
기온
구름 양
풍향
습도

여러 모델을 종합적으로 보고 영향 인자 5개를 선정한 결과는 다음과 같다.

[CO, 오존농도, 온도, 구름 양, 풍향]

2) 모델 성능

그라디언트 부스팅이 가장 정확도가 높다.

모델 파라미터 조절

파라미터 조절 이전 정확도

->

파라미터 조절 이후 정확도

Score on training set:0.924
Score on test set : 0.529

Score on training set:0.846
Score on test set : 0.509

- 선정한 모델인 그라디언트 부스팅을 이용하였다.
- 파라미터 조절 후 오히려 점수가 낮아져 파라미터 조절 이전을 이용하는 것이 더 좋은 예측 정확도를 낼 수 있을 것이라고 판단된다.
- Training 점수에 대해서는 과대적합이 조금 더 방지될 수 있다.

예측 및 평가

```
print(mse, mae)
```

```
[128.2488948086676] [8.454827808122369]
```

- 예측 후 MSE와 MAE를 계산해본 결과 위와 같았다.
- MSE가 상당히 높은데 이는 개선이 필요하다고 판단된다.

결론

- 미세먼지 발생에 영향을 미치는 요인 중 CO, 오존농도, 기온, 구름양, 풍향이 가장 큰 영향을 미친다.
- CO와 오존농도가 높으면 미세먼지의 발생량이 높다.
- 풍향은 상관성 분석을 했을 때는 영향을 미치는 것 같진 않았는데, 모델을 통한 분석을 했을 때는 공통적으로 영향 인자라는 결론을 냈다. 따라서 특정 방향 계열의 바람일수록 미세먼지의 농도가 높아진다고 추측된다.
- 그래프 분석 결과에 따르면 온도가 높을수록 오존 농도가 높아지고, 결국 미세먼지의 발생도 많아진다.

|가설 검정 결과|

- 1) 대기오염이 심해지면 미세먼지의 발생도 많아진다.
- 2) 여름보다 겨울에 미세먼지의 발생이 많아진다.

개선안 제시

- 모델 정확도

모델의 정확도가 50%정도로 예측 성능이 좋다고 할 수 없다. 분석을 하면서 크게 강한 상관성을 보인 변수들이 많지 않았는데 이는 데이터의 양을 늘려 해결할 수 있을 것이라고 생각된다.

- 분석 개선사항

시간이 부족해 미처 하지 못했지만 다음과 같이 분석하면 의미 있는 인사이트를 얻을 수 있을 것으로 생각된다.

- 비나 눈이 오는 날과 아닌 날을 나누어 미세먼지 농도를 비교해 분석

실무 적용 방향

- 날씨를 관측해 오늘의 미세먼지 발생 예측
- 대기 오염 정도에 따른 미세먼지 발생 예측

- 배운점 및 느낀점

데이터 분석에 대한 경험을 해볼 수 있어서 좋았고 분석 배경을 조사하면서 미세먼지에 대한 도메인 지식 또한 쌓을 수 있어서 좋았다. 개념을 조각조각 배우면 활용할 때 어려움을 겪곤 하는데 이번 기회에 한번 더 배웠던 내용을 정리하여 응용까지 할 수 있었던 좋은 기회였다.

- 에로사항

변수들의 설명력이 부족해 정확도가 낮게 나왔는데, 데이터의 양을 늘리거나 다른 변수를 추가하는 등의 작업을 할 필요가 있다고 생각한다.

- 아이디어

- 미세먼지 농도를 예측해 마스크의 생산량 조절.

미세먼지가 많은 날이 예측되면 마스크의 생산량을 늘리고, 반대의 경우에는 줄여 효율적인 생산을 가능하게 한다.

- 미세먼지 농도에 따른 환자 관리

미세먼지가 많은 날이 예측되면 미리 호흡기에 질환을 앓고 있는 환자들에게 알리고, 외출을 삼가하게 한다.

변수 번호	변수	변수 설명	변수 역할	변수 형태	분석 제외 사유	탐색적 기법		모델링 기법					총점	선정 (사유)
						그래프	상관분석	회귀분석	DT	RF	GB	사례연구		
1	MeasDate	측정일자	제외	연속형	학습에 필요하지 않다고 판단되어 제거									
2	PM10	미세먼지 $10\mu\text{g}/\text{m}^3$	목표변수	연속형										
3	O3	오존 농도	설명변수	연속형			○	○	○	○	○	○	6	선정
4	NO2	이산화질소 농도	설명변수	연속형		○	○	○				○	4	선정
5	CO	일산화탄소 농도	설명변수	연속형		○	○	○	○	○	○	○	7	선정 (모든 기법에서 중요하다고 판단)
6	SO2	아황산가스 농도	설명변수	연속형			○					○	2	
7	TEMP	기온($^{\circ}\text{C}$)	설명변수	연속형			○	○	○	○	○	○	6	선정
8	RAIN	강수량(mm)	설명변수	연속형									0	
9	WIND	풍속(m/s)	설명변수	연속형						○		○	2	
10	WIND_DIR	풍향(16방위)	설명변수	연속형				○	○	○	○	○	5	선정
11	HUMIDITY	습도(%)	설명변수	연속형							○	○	2	
12	ATM_PRESS	현지기압(hPa)	설명변수	연속형		○	○	○	○				4	선정
13	SNOW	적설(cm)	설명변수	연속형									0	
14	CLOUD	전운량(10분위)	설명변수	연속형					○	○	○		3	