

A photograph of a person's hands typing on a white laptop keyboard. The laptop screen displays a solid pink color. The background shows a wooden desk with a white coffee cup, a power strip, and some papers. A smartphone with a pink screen is also on the desk.

Portfolio

국동민

프로젝트 소개

- **프로젝트 명 :** Fine Dust
- **수행기간 :** 2018. 12. 06 ~ 2019. 01. 23
- **프로젝트 목표 :** 미세먼지와 질병과의 상관관계를 분석하여 소비자에게 관련제품의 필요성을 인식시키고 그에 따른 판매전략 수립
- **담당 역할 :** 데이터분석 및 시각화
- **향후 프로젝트 계획 :** Node.JS를 통한 크롤링 및 머신러닝 / SNS API를 이용하여 사람들의 관심도 및 홍보방법 조사 / 조사한 자료에 대한 데이터베이스 구축

개발환경

- **OS** : VMware Workstation 14 Player/ Ubuntu 17.10 64bit/ Windows 7 64bit
- **Server** : AWS EC2 / Xshell 6
- **Language** : Python 3.6.4 / R 3.5.1 / HTML5 / CSS3 / JavaScript / PHP
- **DataBase** : MariaDB 10.3 / MySQL 5.7
- **Tools** : Jupyter 5.0.0 / VSCode / 공공데이터포털 OpenAPI / R Studio 1. 1. 4 / Sublime Text 3
- **Github** : Git Bash / Github Desktop

<https://github.com/Bgroupbb/b-store>



Contents

국외

OECD 국외기여도 중국

국내

대한민국 주요도시

대구

구별 오염원별

질병

미세먼지와 질병 상관관계

취약계층 통계와 분포

판매

관련제품의 필요성

판매타겟층 판매전략

개선사항

현재 진행중인 작업

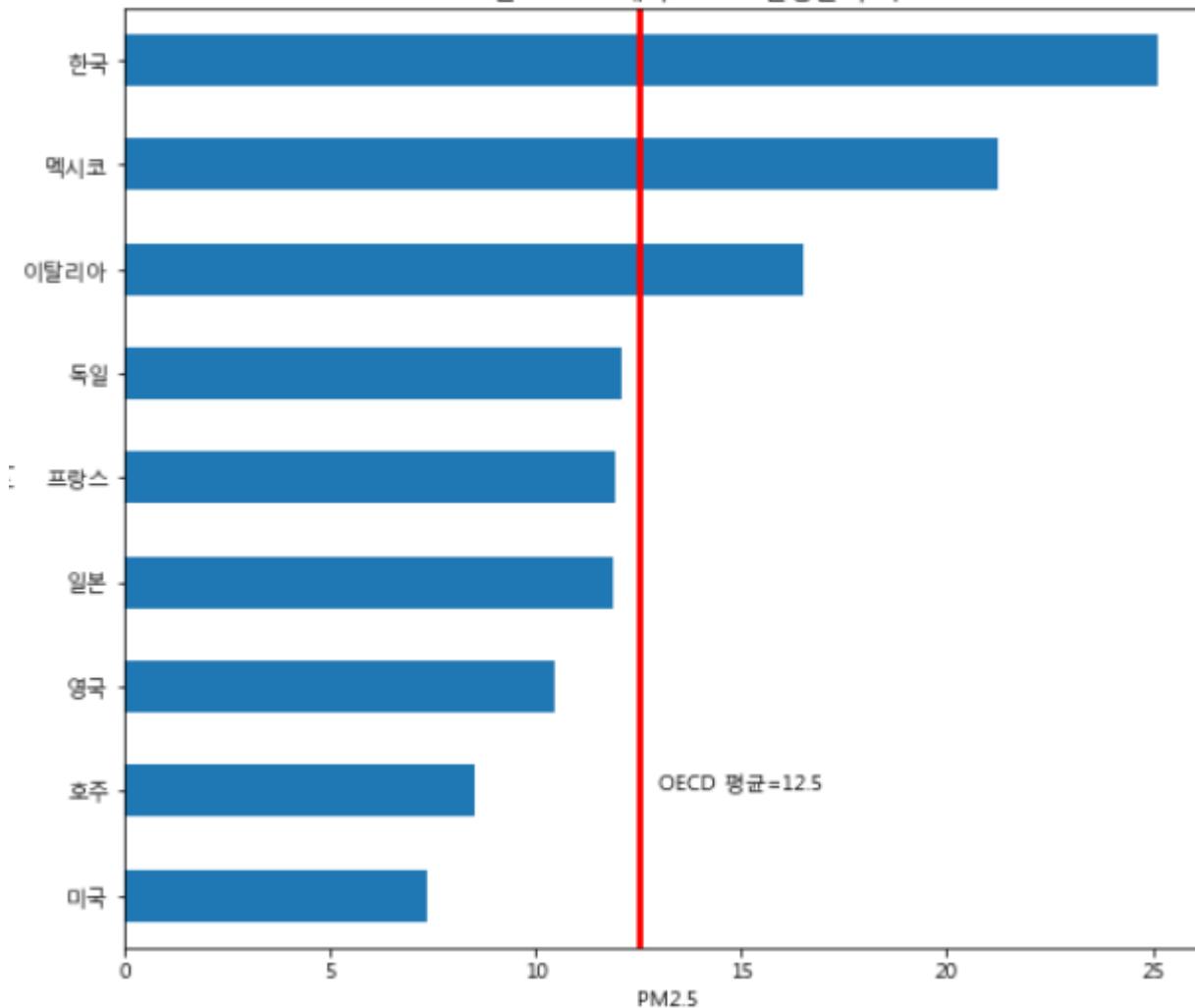
그 외 사항



국외

OECD
국외기여도
중국

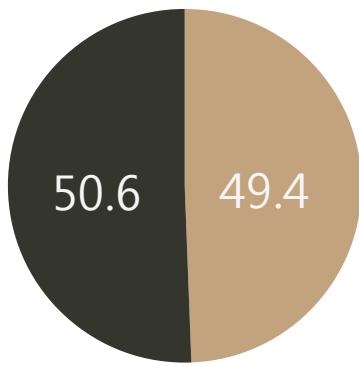
2017년 OECD 9개국 PM2.5 연평균 수치



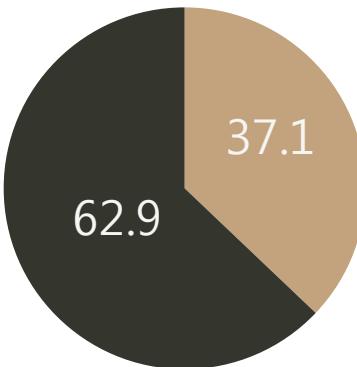
대한민국은 OECD 국가중
초미세먼지 수치가 가장 높다.

■ 국외 기여도

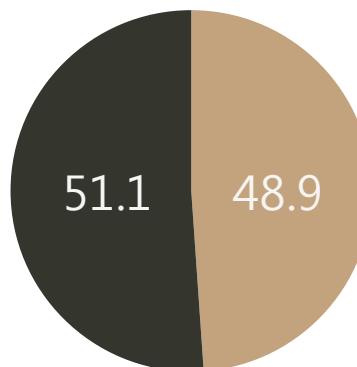
2018-11-24



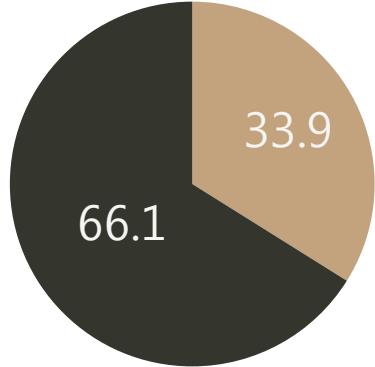
2018-11-25



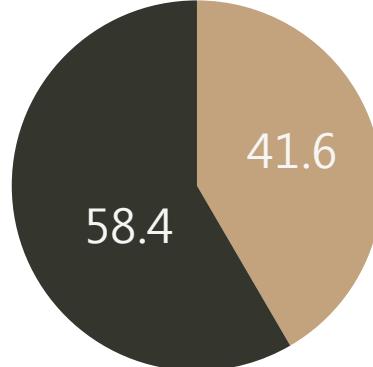
2018-11-26



2018-11-27



2018-11-28



국외(주로 중국)의 영향이 크다.



국내 국외

전국 초미세먼지
단위 (%)



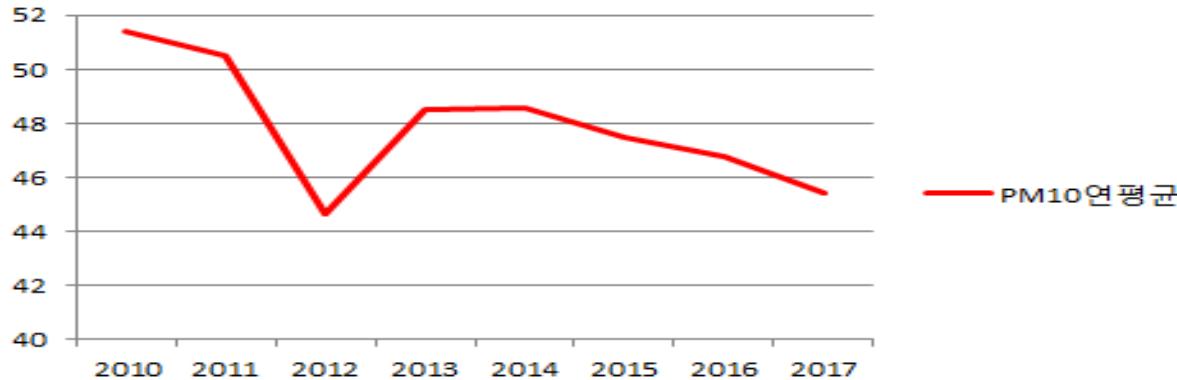
중국이 미세먼지가 심한 이유



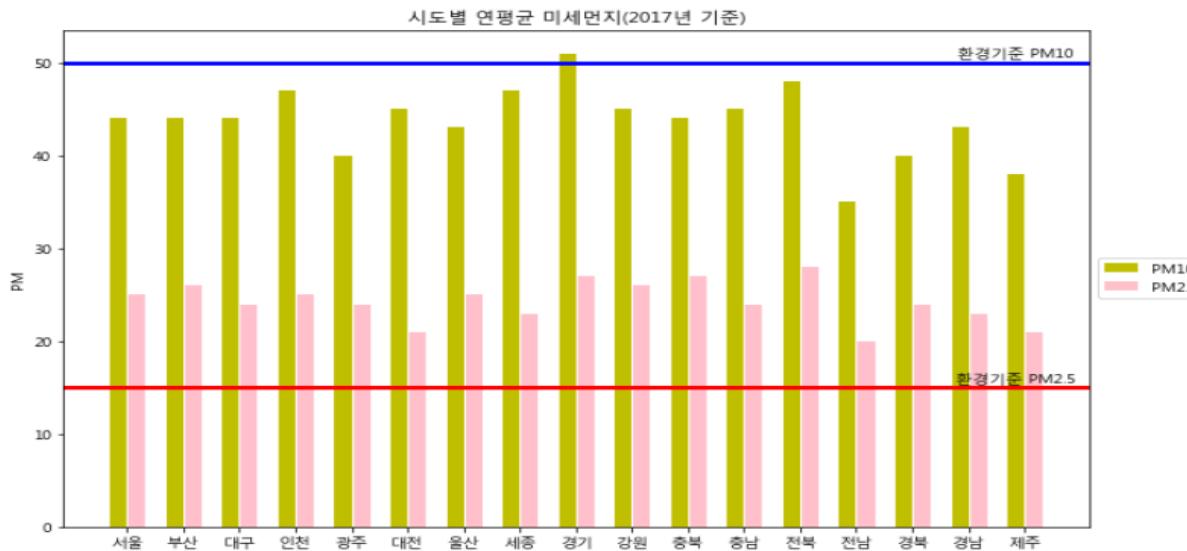
국내

대한민국
주요도시
주요도시 오염원별

2010-2017 대한민국 PM10 연평균



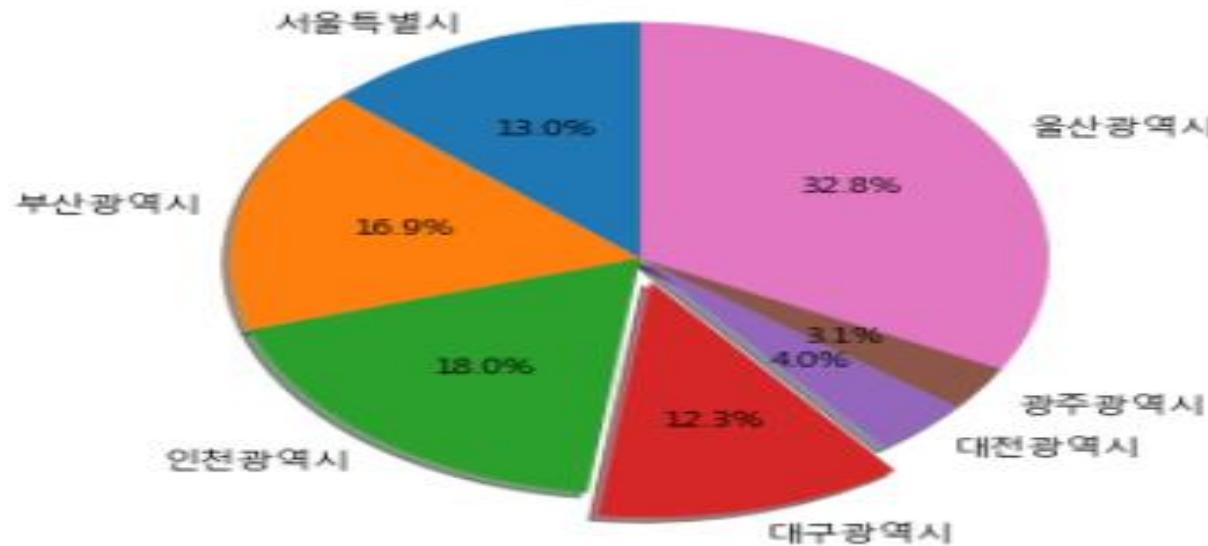
대한민국 미세먼지 연평균 농도가 2012~2013년 사이 풍속이 약해져 나타난 악영향으로 잠시 올랐으나 전반적으로 하락 추세를 보임을 알 수 있다.



PM 2.5 환경기준이 3월 개정전 25에서 세계기준인 15로 강화되고나서 모든 도시가 다 초과중인 상황
(경기지역은 PM10 PM2.5 모두 초과하였다.)

■ 주요도시

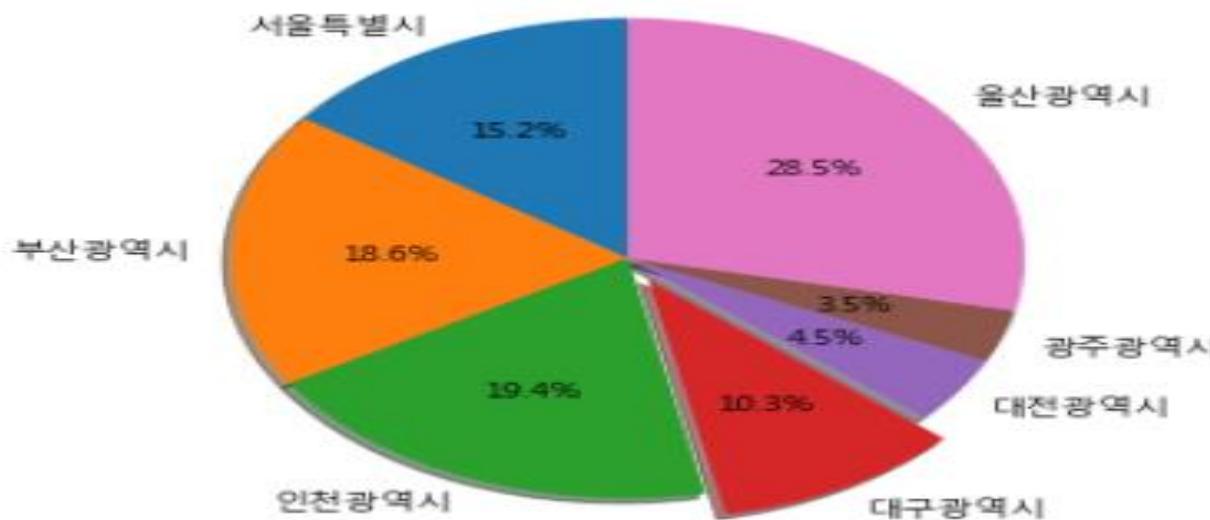
특별&광역시 별 PM10 비율



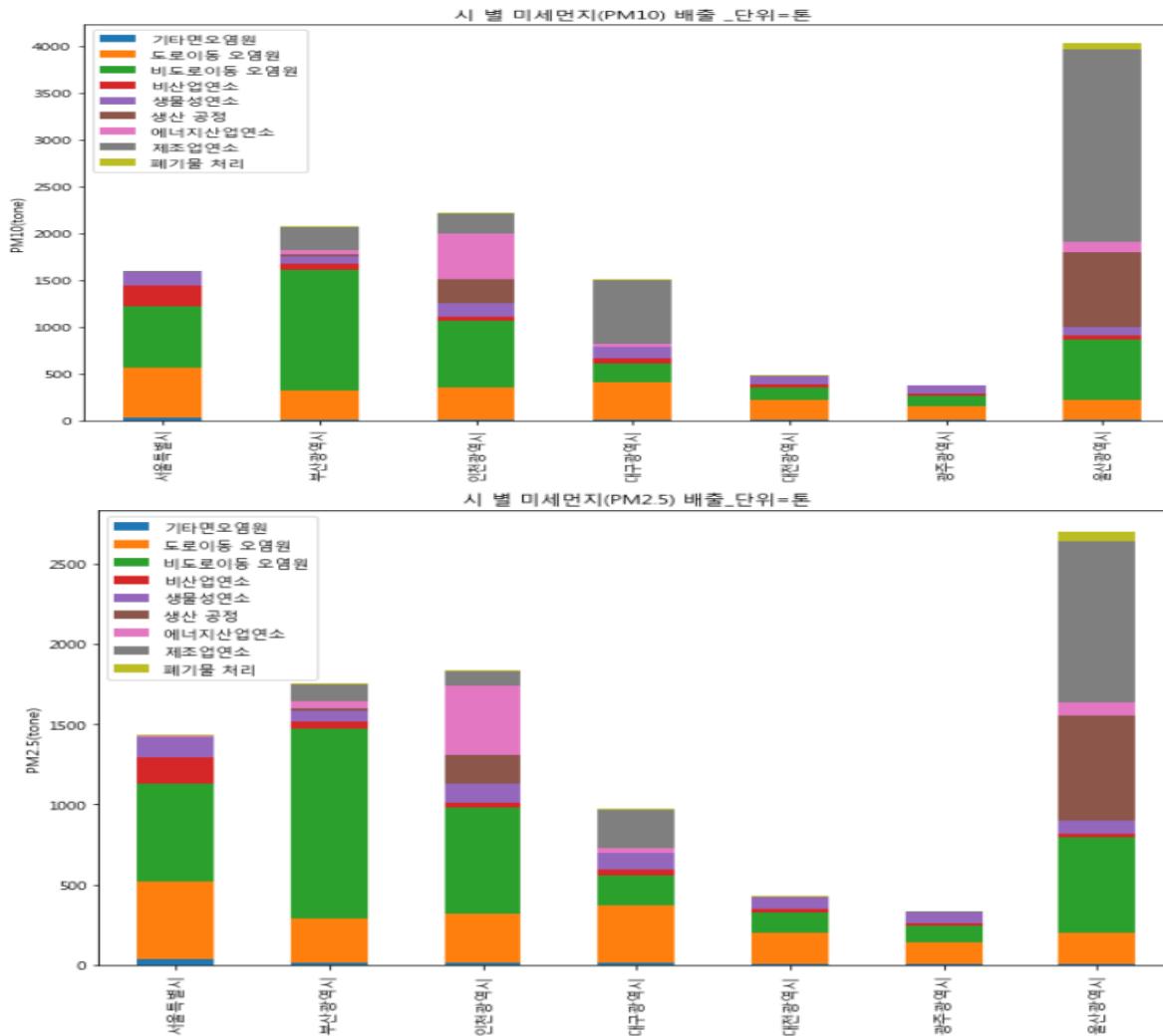
배출비율이 **울산**이 가장 높고
광주 대전이 낮음을 알 수 있
다.

2015년 기준

특별&광역시 별 PM2.5 비율



주요도시 오염원별



울산 대구는
제조업연소로 인한
미세먼지 배출이 눈에 띄고

서울 부산 인천은
이동오염원으로 인한
미세먼지 배출이 비중이 높다.

2015년 기준

구현설명

시도	시군구	배출원대분류	PM10	PM2.5	count
서울특별시 종로구	비산업 연소	126	120	1	
서울특별시 종로구	비산업 연소	5	2	1	
서울특별시 종로구	비산업 연소	2,795	1,436	1	
서울특별시 종로구	비산업 연소	126	81	1	
서울특별시 종로구	비산업 연소	5	4	1	
서울특별시 종로구	비산업 연소	16	16	1	
서울특별시 종로구	비산업 연소	147	147	1	
서울특별시 종로구	비산업 연소			1	
서울특별시 종로구	비산업 연소	537	280	1	
서울특별시 종로구	비산업 연소	14	13	1	
서울특별시 종로구	비산업 연소	3	3	1	
서울특별시 종로구	비산업 연소	1,585	1,585	1	
서울특별시 종로구	비산업 연소	9	6	1	
서울특별시 종로구	제조업 연소			1	
서울특별시 종로구	제조업 연소			1	



```
df = pd.read_excel (r'C:\Users\USER\Desktop\프로젝트\csv\test\특별시&광역시오염배출량.xls', encoding = 'utf-8')
pv = pd.pivot_table(df, index=['시도'], columns=["배출원대분류"], values=['PM10'], aggfunc=sum, fill_value=0)
pv2 = pd.pivot_table(df, index=['시도'], columns=["배출원대분류"], values=['PM2.5'], aggfunc=sum, fill_value=0)
```

pv

PM10

배출원대분류	기타 연소	농업	도로이동오염	비도로이동오염	비산업 연소	생물성 연소	생산공정	에너지산업 연장	에너지수송 및 저용	유기용제 사용	제조업 연소	폐기물처리
시도												
광주광역시	6178	0	143590	115615	26511	78296	1280	2333	0	0	2422	2807
대구광역시	11145	0	394542	200922	52275	125813	1255	33114	0	0	680370	7629
대전광역시	8151	0	209392	137579	32190	87031	873	3780	0	0	603	5849
부산광역시	12182	0	303659	1293586	63678	81777	14525	49420	0	0	246688	6461
서울특별시	36303	0	531068	657393	221928	136675	0	7342	0	0	1586	9127
울산광역시	5497	0	213007	648966	37037	97487	791367	117385	0	0	2046827	67141
인천광역시	12079	0	336132	715536	41406	145796	258932	492070	0	0	203069	9711

일반적인 엑셀파일에서 '시도'를 index, '배출원대분류'를 columns, 'PM10'과 'PM2.5'를 각각 values로 잡아 pivot_table을 이용해 합산함.

시도별 어떤 배출원을 통해 얼마만큼의 미세먼지가 나오는지 파악 후 matplotlib를 이용해 파이그래프와 누적그래프를 구현



대구

대구 특수성
오염원 별 배출량
오염원 별 비중
구별 비율
미세먼지 분포

날로 심화되는 대구 초미세먼지...대구시 대응 '미흡'

정창오 2018.06.18. 11:29



=뉴스= 우종록 기자 = 자료
미세먼지 농도가 '나쁨' 수준
30일 오후 대구 중구 대구백
크를 찾은 시민들이 등성로
모습이다. 2018.06.18...

【대구=뉴시스】정창오 기자 = 최근 초미세먼지가 숨 쉬는 공기를 공포의 대상으로 바꾸어 놓고 있는 재난이 되어가고 있는데 대구에서도 주의보와 경보가 잇따라 내려지는 등 초미세먼지가 심각한 이슈로 부상하고 있다.

18일 대구시 등에 따르면 1급 발암물질인 초미세먼지는 호흡기 질환과 우울증, 치매 등 각종 질병을 야기하는 무서운 매개체임에도 불구하고 그 피해가 단시간 내에 나타

**도심의 서북쪽에 위치한 산업단지에서 발생된
오염물질이 도시전체로 이동하는 열악한 구조**

특히 분지지형인 대구는 다양한 미세먼지 배출원과 도심의 서북쪽에 위치한 산업단지에서 발생된 오염물질이 도시전체로 이동하는 열악한 구조를 가지고 있어 특단의 대책 마련이 요구되고 있다.

최근 3년간 대구의 미세먼지 평균농도는 m^3 당 42~46 μg 이고 초미세먼지는 m^3 당 23~26 μg 정도다.

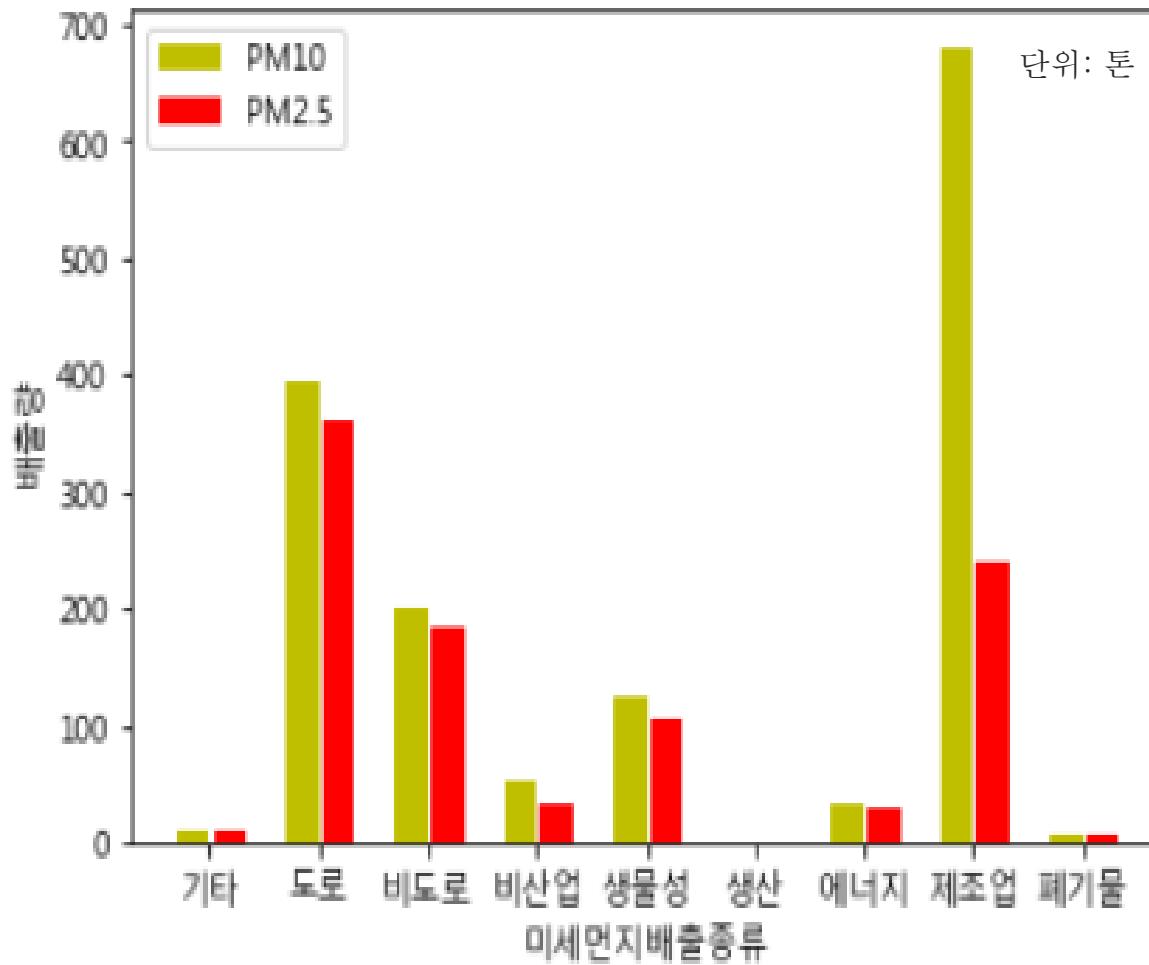
FIND

ie SCHOOL
ARCHITECTURE & DESIGN

추천 항목
후원 링크 제공



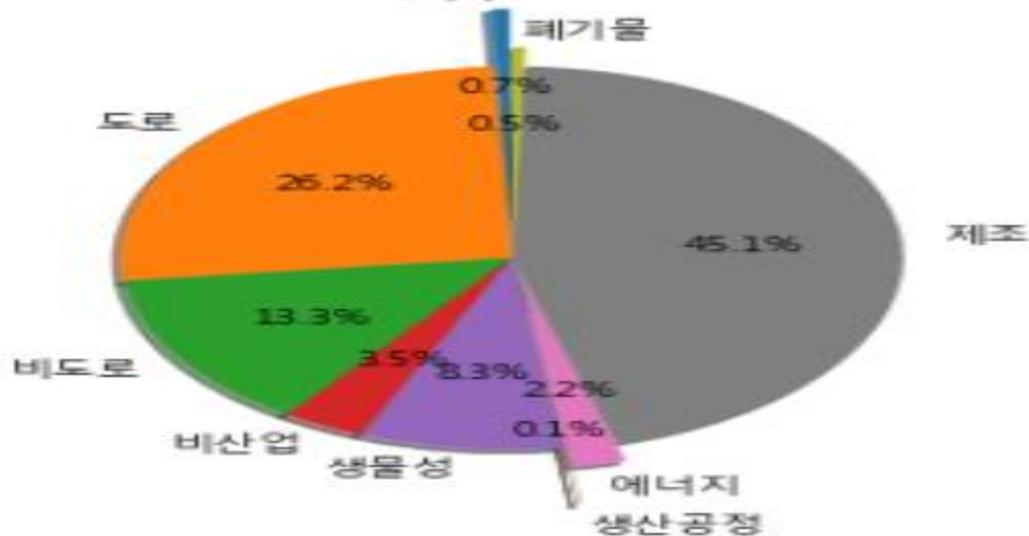
대구시 미세먼지 배출(2015년 기준)



대구 미세먼지는
제조업에서 많이 발생되고
초미세먼지는 도로와
제조업 모두 많이 발생된다.

■ 오염원별 비중

대구 오염원 비중(PM10)
기타



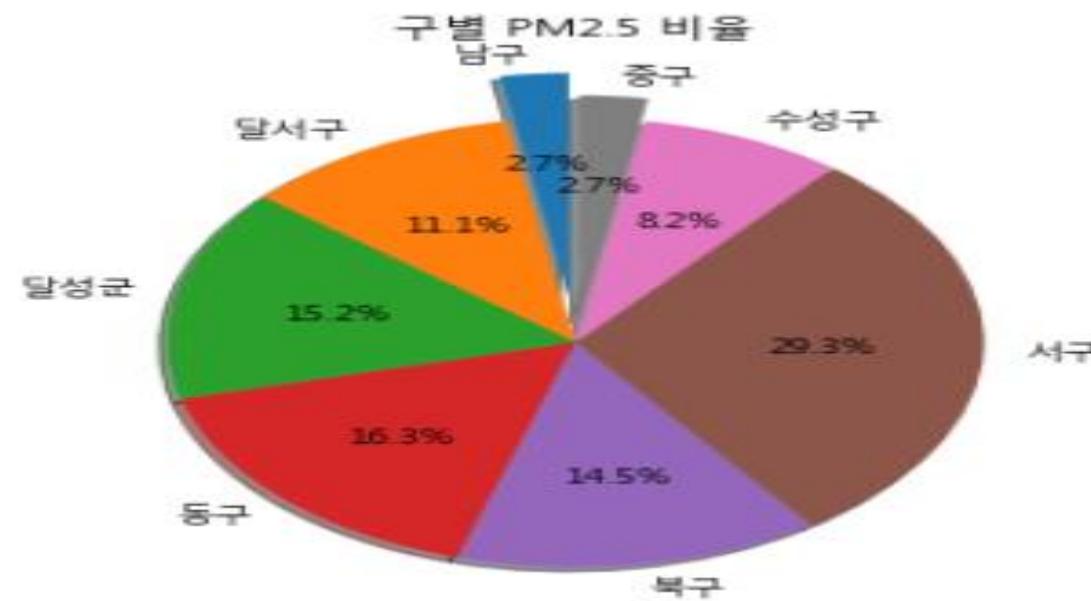
대구 오염원 비중(PM2.5)
기타



대구 미세먼지는
제조업에서 45.1%를 차지하고
초미세먼지는 도로에서
~7.4% 제조업 24.6%를 보인
다.

2015년 기준

구별 비율



미세먼지를 가장 많이 배출하는 곳은 서구이고 가장 적게 배출하는 곳은 남구 중구임을 알 수 있다.

2015년 기준

구별 비율

시도	시군구	배출원대분류	PM10	PM2.5
대구광역시	중구	비산업 연소	12	8
대구광역시	중구	비산업 연소	5	5
대구광역시	중구	비산업 연소	17	17
대구광역시	중구	비산업 연소	25	25
대구광역시	중구	비산업 연소		
대구광역시	중구	비산업 연소	15	15
대구광역시	중구	비산업 연소	2,406	1,253
대구광역시	중구	비산업 연소	3	2
대구광역시	중구	비산업 연소	40	37
대구광역시	중구	비산업 연소	38	38
대구광역시	중구	비산업 연소	401	401



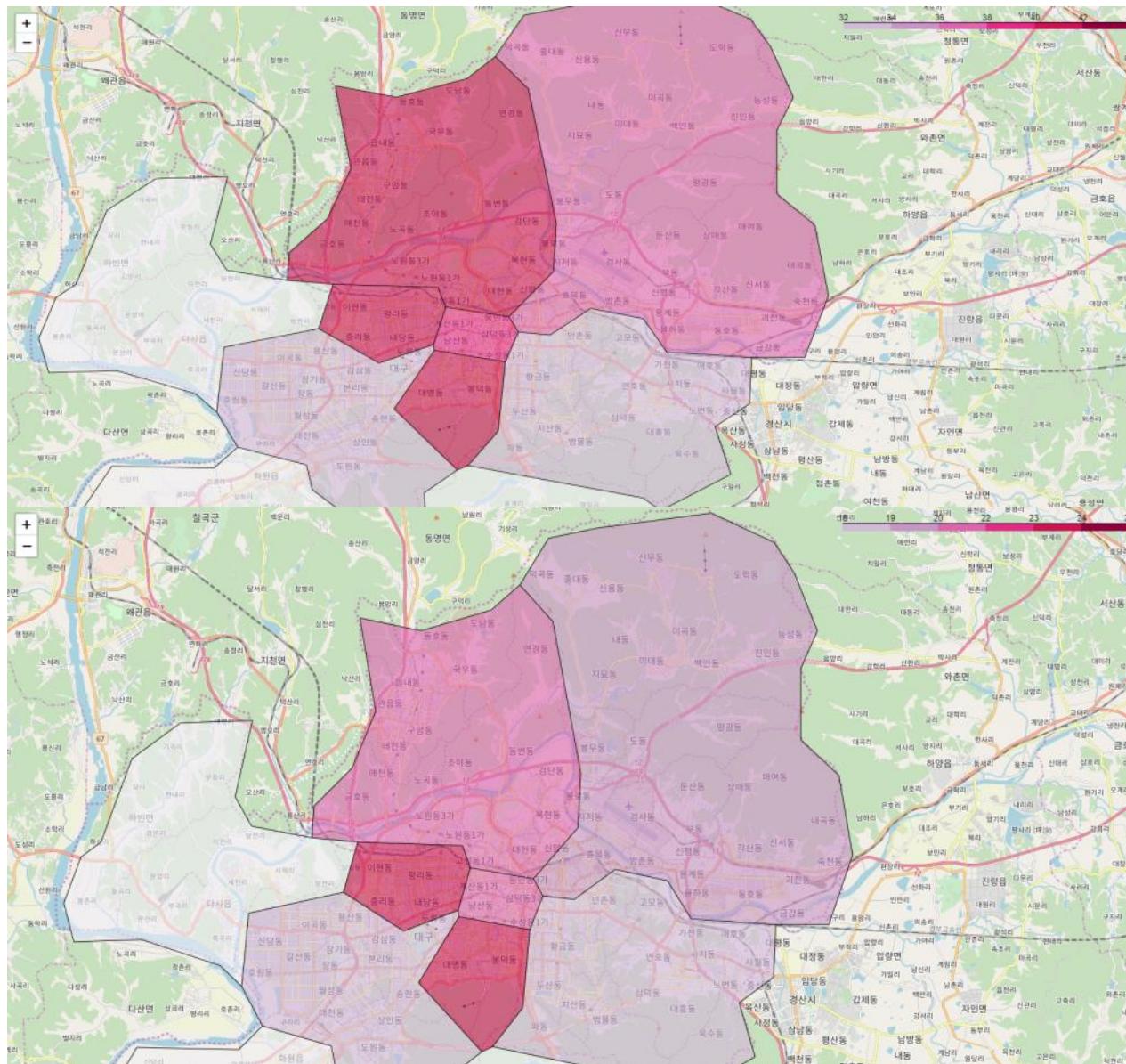
주요도시의 미세먼지 배출원을 파악할 때 사용한 엑셀 파일에서 시도가 대구광역시인 부분만 다른 엑셀 파일로 만들어 똑같이 pivot_table을 적용해 그래프 구현

```
In [3]: pv = pd.pivot_table(df, index=['시군구'], columns=["배출원대분류"], values=['PM10'], aggfunc=sum, fill_value=0)
pv
```

```
Out[3]: PM10
```

배출원대분류	기타 면오염 원	농업 원	도로이동오염 원	비도로이동오염 원	비산업 연소	생물성 연소	생산공정	에너지산업 연장	에너지수송 및 저용	유기용제 사	제조업 연소	폐기물처리
시군구												
남구	536	0	13882	7634	4705	3413	4	0	0	0	0	0
달서구	2452	0	65716	27299	10024	13564	451	1924	0	0	772	2298
달성군	1344	0	60471	60033	4638	33876	406	186	0	0	1447	4222
동구	1468	0	70685	44762	9495	34295	123	15222	0	0	30	0
북구	2218	0	101365	29304	8108	14145	191	0	0	0	1065	0
서구	1147	0	22292	7142	5014	4283	43	15782	0	0	677017	1109
수성구	1376	0	47767	15315	7329	17950	9	0	0	0	30	0
중구	604	0	12364	9433	2962	4287	28	0	0	0	9	0

미세먼지 분포



PM 10

서구와 북구는
공단이 밀집하여
자체배출이 높으나

남구 중구는
자체배출은 낮은데도
미세먼지농도는 높음

PM 2.5

2018 연평균 기준

지도구현설명

```
import pandas as pd  
import numpy as np
```

```
air = pd.read_excel(r'C:\Users\USER\Desktop\프로젝트\csv\map용\미세먼지(PM_2.5).xlsx', encoding='utf-8')
```

```
import folium  
import googlemaps  
from tqdm import tqdm_notebook
```

```
gmaps_key = "REDACTED"  
gmaps = googlemaps.Client(key=gmaps_key)
```

```
place_name = []  
for name in air['측정동']:  
    place_name.append('대구광역시' + str(name))  
place_name
```

```
['대구광역시 대구광역시 중구 달성로 22길 30',  
'대구광역시 대구광역시 수성구 무학로 209',  
'대구광역시 대구광역시 동구 안심로 49길 70',  
'대구광역시 대구광역시 서구 국채보상로 136',  
'대구광역시 대구광역시 남구 성당로 30길 55',  
'대구광역시 대구광역시 북구 3공단로 14길 31',  
'대구광역시 대구광역시 동구 마양로 37길 92',  
'대구광역시 대구광역시 북구 칠곡중앙대로 52길 56',  
'대구광역시 대구광역시 수성구 국채보상로 1000',  
'대구광역시 대구광역시 달서구 성서공단로 11길 32',  
'대구광역시 대구광역시 달성군 현풍면 현풍중앙로 144-34',  
'대구광역시 대구광역시 수성구 노변공원로 52',  
'대구광역시 대구광역시 달서구 월배로 131']
```

대구시의 각 구별 측정소에서 측정한 미세먼지 농도를 측정한 데이터가 있는 엑셀파일을 가져와 각 측정소 도로명주소를 주피터에 입력하여 리스트화

지도구현설명

```
place_address = []
place_lat = []
place_lng = []

for name in place_name:
    tmp = gmaps.geocode(name, language='ko')
    place_address.append(tmp[0].get("formatted_address"))

    tmp_loc = tmp[0].get("geometry")

    place_lat.append(tmp_loc['location']['lat'])
    place_lng.append(tmp_loc['location']['lng'])

    print(name + '-->' + tmp[0].get("formatted_address"))
```

대구광역시대구광역시 중구 달성로 22길 30→대한민국 대구광역시 중구 성내3동 달성로22길 30
대구광역시대구광역시 수성구 무학로 209→대한민국 대구광역시 수성구 지산1동 무학로 209
대구광역시대구광역시 동구 안심로 49길 70→대한민국 대구광역시 동구 안심3,4동 안심로49길 70
대구광역시대구광역시 서구 국채보상로 135→대한민국 대구광역시 서구 상중이동 국채보상로 135
대구광역시대구광역시 남구 성당로 30길 55→대한민국 대구광역시 남구 대명동 성당로30길 55
대구광역시대구광역시 북구 3공단로 14길 31→대한민국 대구광역시 북구 노원동 3공단로14길 31
대구광역시대구광역시 동구 미양로 37길 92→대한민국 대구광역시 동구 신일동 미양로37길 92
대구광역시대구광역시 북구 철곡중앙대로 52길 56→대한민국 대구광역시 북구 태전동 철곡중앙대로52길 56
대구광역시대구광역시 수성구 국채보상로 1000→대한민국 대구광역시 수성구 만촌2동 국채보상로 1000
대구광역시대구광역시 달서구 성서공단로 11길 32→대한민국 대구광역시 달서구 신당동 성서공단로11길 32
대구광역시대구광역시 달성군 현풍면 현풍중앙로 144-34→대한민국 대구광역시 달성군 현풍면 현풍중앙로 144-34
대구광역시대구광역시 수성구 노변공원로 52→대한민국 대구광역시 수성구 시지동 노변공원로 52
대구광역시대구광역시 달서구 월배로 131→대한민국 대구광역시 달서구 진천동 월배로 131

```
air['lat'] = place_lat
air['lng'] = place_lng
```

```
air.head()
```

리스트화 한 미세먼지 측정소들의 주소를 이용하여 googlemapAPI를 통해 위도와 경도를 가져옴

	측정동	측정일자	연_평균값	측정항목	단위	구	측정소	lat	lng
0	대구광역시 중구 달성로 22길 30	201803	22.636364	미세먼지	$\mu\text{g}/\text{m}^3$	중구	중구 수창동	35.874296	128.584137
1	대구광역시 수성구 무학로 209	201803	18.000000	미세먼지	$\mu\text{g}/\text{m}^3$	수성구	수성구 지산동	35.830470	128.631555
2	대구광역시 동구 안심로 49길 70	201803	22.272727	미세먼지	$\mu\text{g}/\text{m}^3$	동구	동구 서호동	35.869103	128.711015
3	대구광역시 서구 국채보상로 135	201803	25.454545	미세먼지	$\mu\text{g}/\text{m}^3$	서구	서구 이현동	35.869552	128.545324
4	대구광역시 남구 성당로 30길 55	201803	24.500000	미세먼지	$\mu\text{g}/\text{m}^3$	남구	남구 대명동	35.845499	128.571068

지도구현설명

```
import json
```

```
geo_path = r'C:\Users\USB\Desktop\프로젝트\jupyter\json\02. skorea_municipalities_geo_simple_2.json'  
geo_data = json.load(open(geo_path, encoding='utf-8'))
```

```
gu_data = pd.pivot_table(air, index=['구'], values=['연_평균값'], aggfunc=np.mean)  
gu_data
```

연_평균값

구

남구 24.50000

달서구 20.17500

달성군 17.818182

동구 21.590909

북구 22.465909

서구 25.454545

수성구 19.840909

중구 22.636364

대구시의 경계 데이터를
저장한 json파일을 불러옴

pivot table을 이용하여
'구'를 인덱스, '연 평균값'
을 values 삼아 구별 미세
먼지 연평균 농도를 정리

Folium을 이용하여 구별
미세먼지 농도를 지도위에
도출함

```
mapping = folium.Map(location=[air['lat'].mean(), air['lng'].mean()], zoom_start=11)  
mapping.choropleth(geo_data=geo_data,  
                  data=gu_data['연_평균값'],  
                  columns=[gu_data.index, '연_평균값'],  
                  fill_color='PuRd',  
                  key_on='feature.id')
```



질병

미세먼지와 질병
미세먼지와 질병 상관관계
상관계수 그래프
질환별 조기사망률
취약계층
환자 연령분포
어린이집 분포
경로당 분포

[김봉순]

"그냥 미세먼지가 많으면 숨쉬기 곤란하고 목이 칼칼하고 그렇잖아요."

그런데 이렇게 미세먼지가 많은 날 건강에 특히 영향을 받는 질병이 따로 있는 것으로 확인됐습니다.

미세먼지가 많은 날 건강에 특히 영향을
받은 질병이 따로 있는 것으로 확인

성균관대 연구진이 질병관리본부 등록을 같이 하는 미세먼지 관찰 대상 1000명을 조사했습니다.

미세먼지가 기승을 부리면 폐에 염증이 생기는 폐렴과 호흡기능이 떨어지는

는

국가응급진료망에 접수된 환자 4000명을 분석해 봤더니 4개의 질병군에서 미세먼지가 몰려오기 전후로 폐에 염증이 생기는 만성 폐쇄성 폐질환, 심장 관상동맥에 문제가 생기는 허혈성 심장질환과 심장 수축이완 능력이

폐에 염증이 생기는 폐렴과 호흡기능이 떨어지는 만성 폐쇄성 폐질환, 심장 관상동맥에 문제가 생기는 허혈성 심장질환과 심장 수축이완 능력이 저하되는 심부전 4가지였습니다.

초미세먼지가 '나쁨'일 때 폐렴은 11%, 만성 폐질환 환자는 9% 정도가 늘었습니다.

심장관련 질환도 초미세먼지가 농도가 짙어질수록 평균 3%에서 7%까지 환자가 증가했습니다.

미세먼지 농도와 질병환자수와의 상관관계

	PM10	PM2.5	C34	I20	I50	I64	J18
PM10	1.0000000	0.8315156	-0.387220458	0.20192382	0.3975649	0.29913714	0.82596620
PM2.5	0.8315156	1.0000000	-0.302565221	0.27783479	0.2957580	0.31671769	0.71479554
C34	-0.3872205	-0.3025652	1.000000000	0.53847864	0.5239977	0.34968087	-0.08172730
I20	0.2019238	0.2778348	0.538478641	1.00000000	0.9060248	0.84382789	0.08399057
I50	0.3975649	0.2957580	0.523997696	0.90602484	1.0000000	0.83164942	0.32191412
I64	0.2991371	0.3167177	0.349680875	0.84382789	0.8316494	1.00000000	0.05281766
J18	0.8259662	0.7147955	-0.081727302	0.08399057	0.3219141	0.05281766	1.00000000
J44	0.7916848	0.7347171	-0.007118199	0.61951078	0.7522035	0.65015171	0.65359746
J45	0.8240417	0.7834161	-0.128731840	0.12967552	0.2927814	0.07010298	0.96087105
	J44	J45					
PM10	0.791684843	0.82404169					
PM2.5	0.734717149	0.78341608					
C34	-0.007118199	-0.12873184					
I20	0.619510783	0.12967552					
I50	0.752203480	0.29278144					
I64	0.650151714	0.07010298					
J18	0.653597458	0.96087105					
J44	1.000000000	0.66214793					
J45	0.662147935	1.00000000					

J18 폐렴

J44 만성 폐색성 폐질환

J45 천식

0.1이상 무시 선형관계

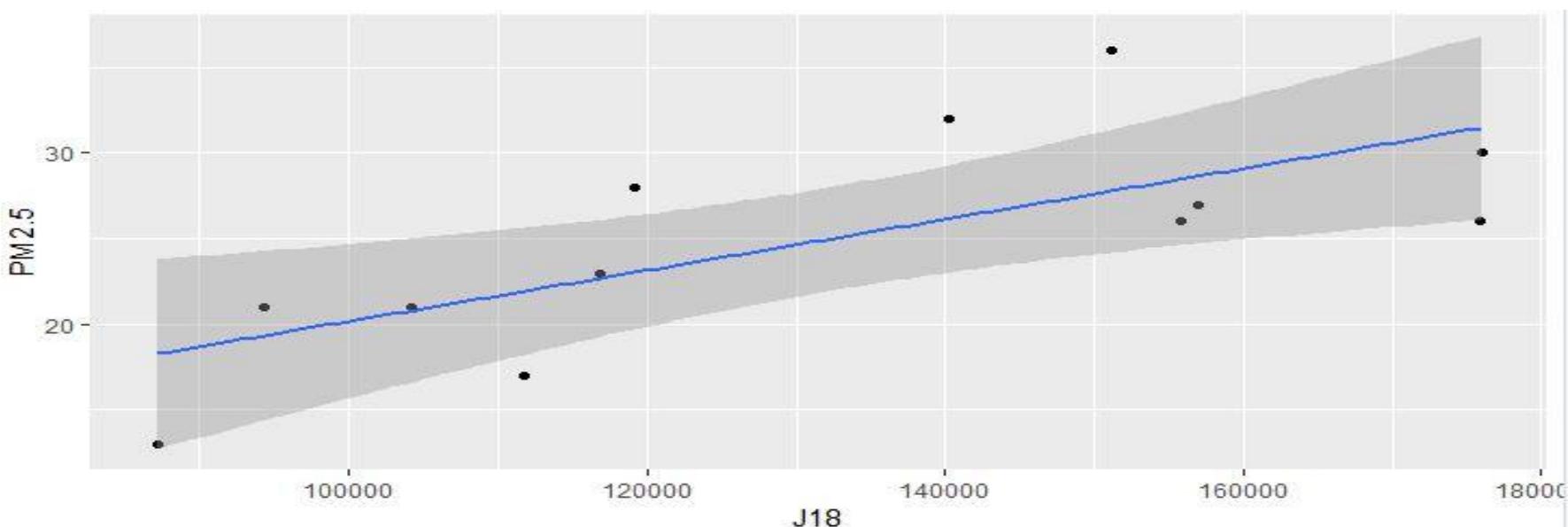
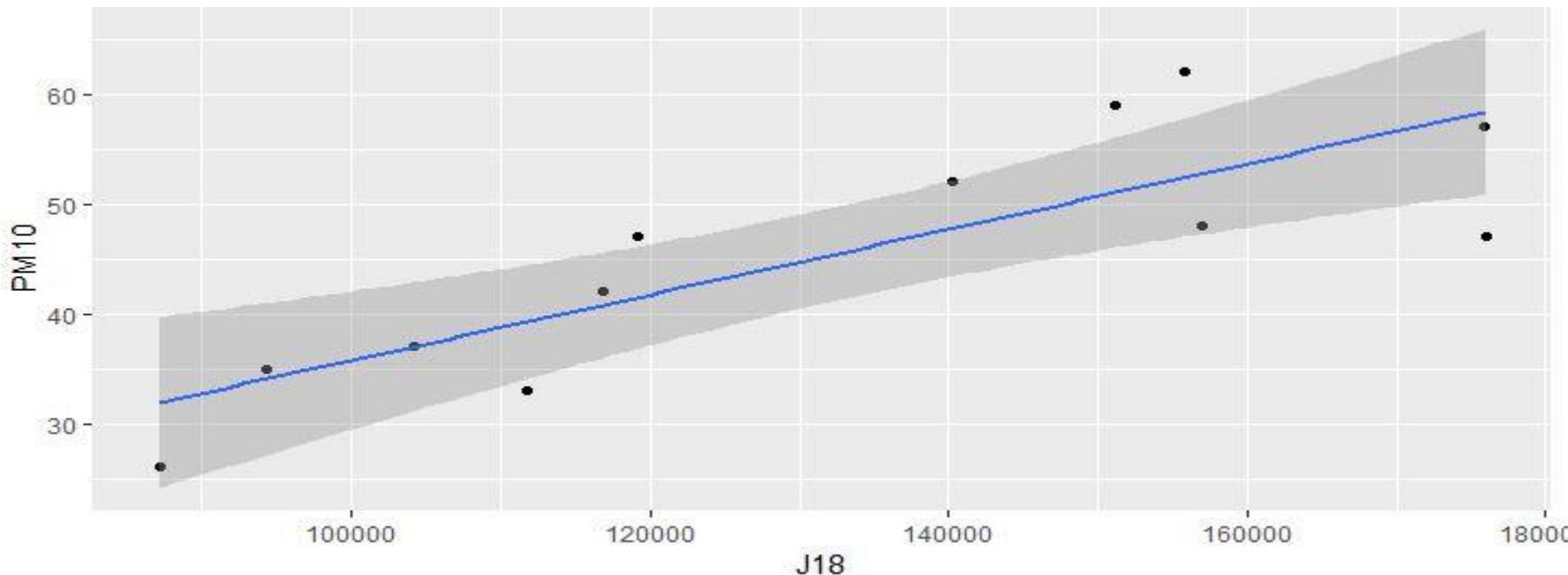
0.3이상 약한 선형관계

0.7이상 뚜렷한 선형관계

1.0이상 강한 선형관계

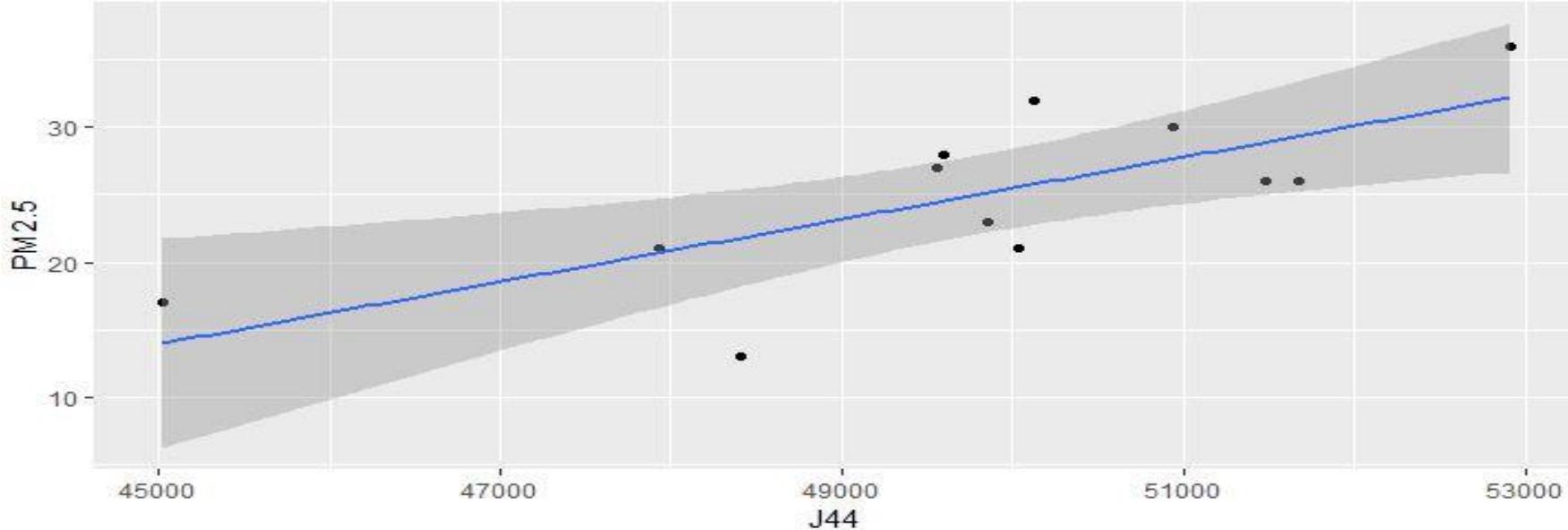
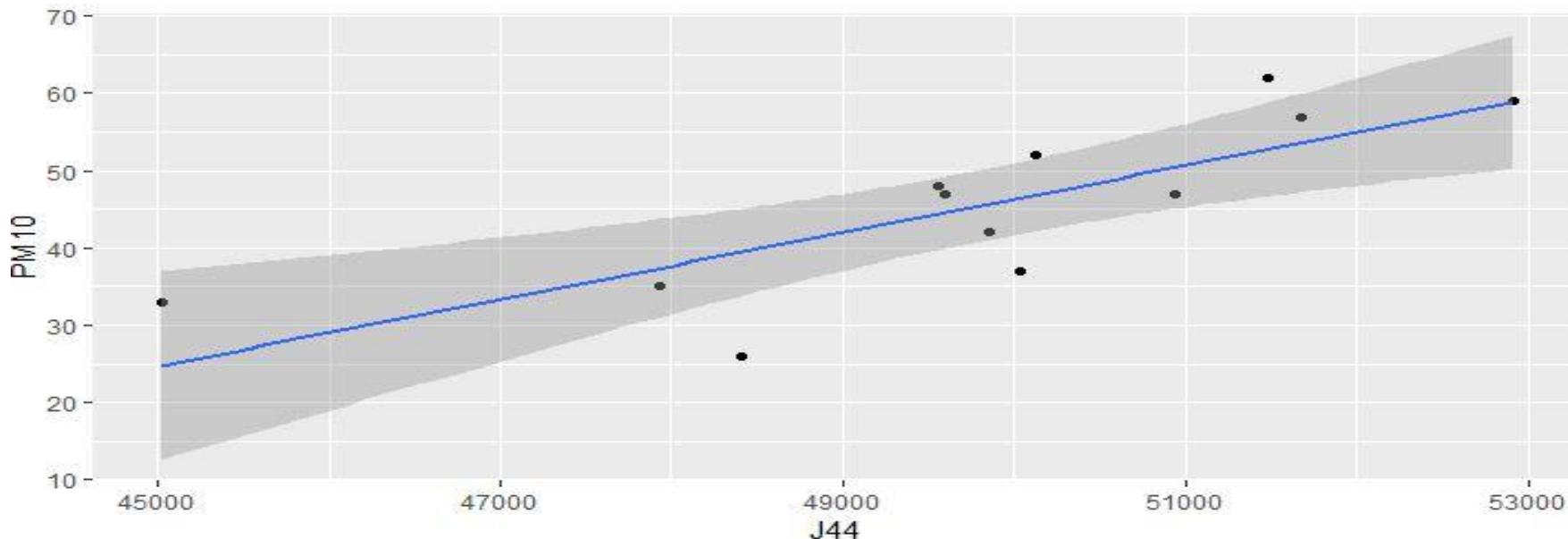
2017년 기준

상관계수 그래프

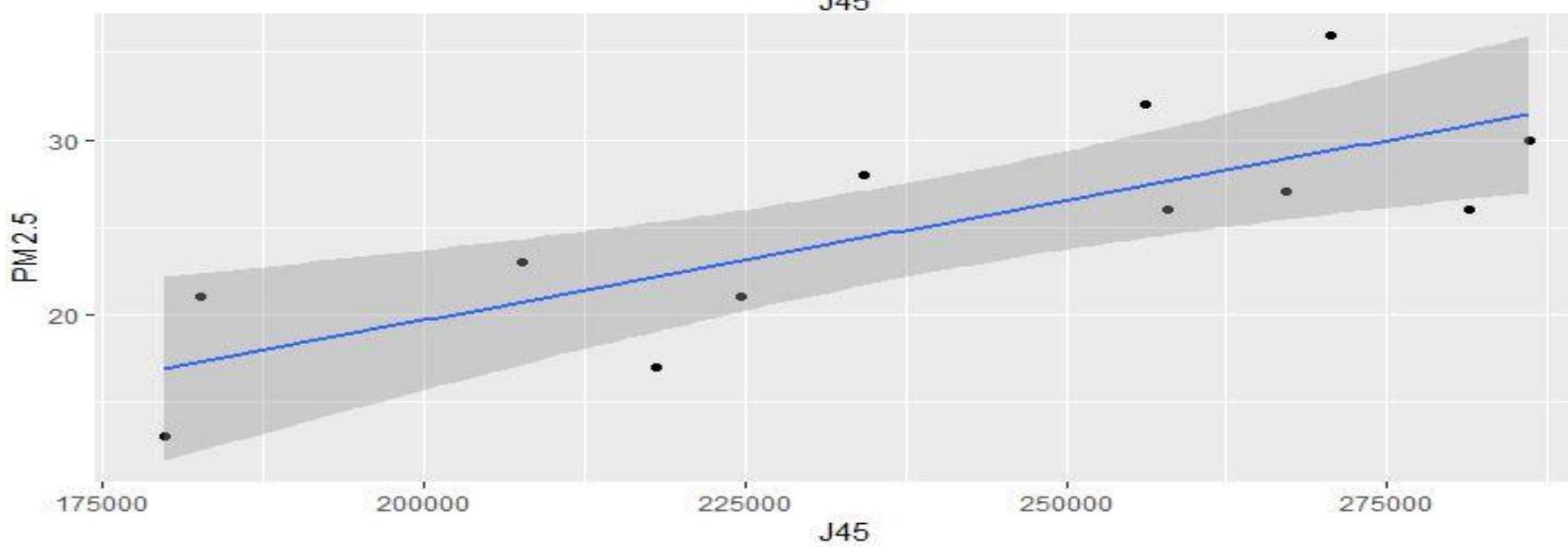
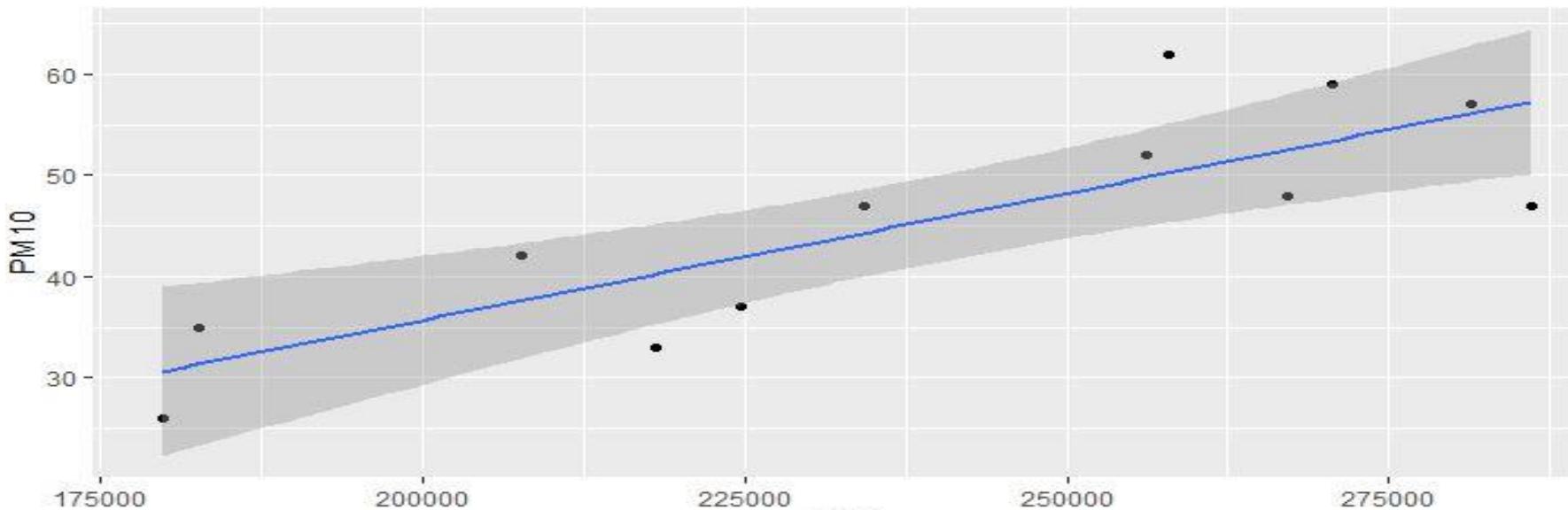


출처 : 통계청 중앙응급의료센터

상관계수 그래프



상관계수 그래프



상관관계 구현설명

PM10	PM2.5	C34	I20	I50	I64	J18	J44	J45
52	32	30167	169198	30112	4386	140274	50127	256159
47	28	30405	172736	30834	4256	119192	49597	234246
59	36	31901	183898	33255	4687	151137	52910	270564
57	26	30916	171690	32130	4382	175924	51680	281354
62	26	31433	176095	32719	4497	155820	51479	257833
42	23	32074	176045	32088	4480	116799	49860	207655
35	21	31801	175432	31447	4482	94281	47936	182643
26	13	32989	175983	31720	4418	87107	48405	179798
37	21	32607	185494	32748	4575	104207	50037	224746
33	17	31287	156873	28936	4152	111731	45019	218080
48	27	33031	183374	32705	4437	156992	49553	267142
47	30	32994	174497	31932	4353	175969	50939	285994

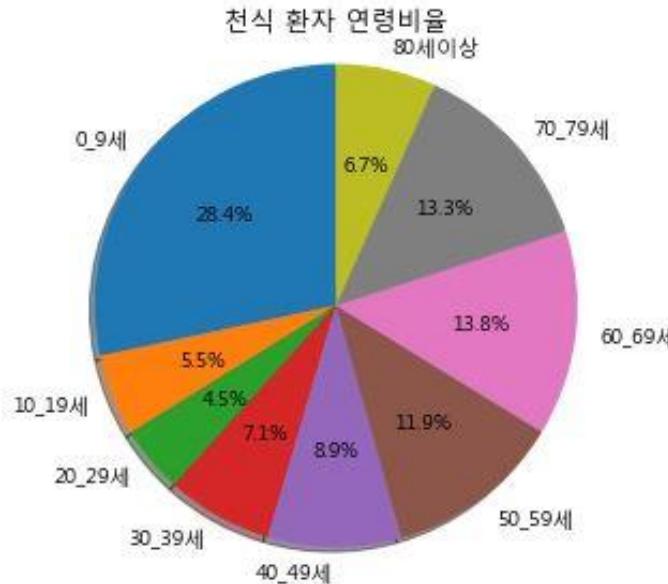
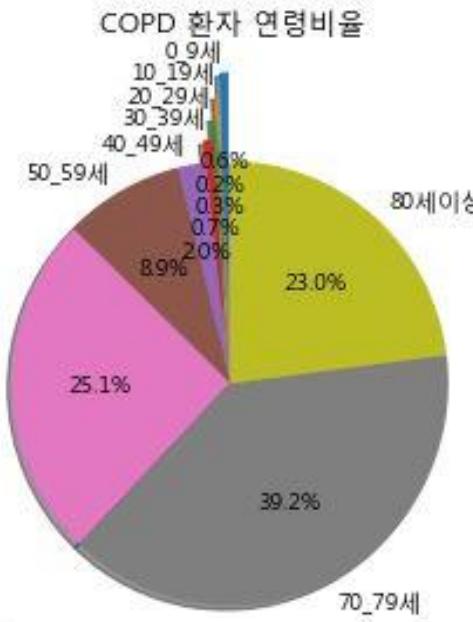


```
dat <- read.csv('C:/Users/USER/Desktop/프로젝트/csv/R그래프용/i11.csv'
cor(dat)
library(ggplot2)
ggplot(dat, aes(J18, PM10))+ geom_point() + geom_smooth(method = "lm")
ggplot(dat, aes(J18, PM2.5))+ geom_point() + geom_smooth(method = "lm")
ggplot(dat, aes(J44, PM10))+ geom_point() + geom_smooth(method = "lm")
ggplot(dat, aes(J44, PM2.5))+ geom_point() + geom_smooth(method = "lm")
ggplot(dat, aes(J45, PM10))+ geom_point() + geom_smooth(method = "lm")
ggplot(dat, aes(J45, PM2.5))+ geom_point() + geom_smooth(method = "lm")
corrplot(dat)
```

2017년 1월부터 12월까지
의 미세먼지농도와 각종
질병 환자수 데이터를 가
진 엑셀파일을 불러와 R을
이용해 상관관계를 도출

상관관계가 높은 질병은
그래프와 함께 도출

■ 환자 연령분포



미세먼지 관련 질병별
취약계층 분석 결과
폐렴은 어린이
COPD는 노인
천식은 노약자 모두
발병률이 높음을 확인

2017년 기준

질병 별 환자 연령분포 그래프 구현설명

```
In [1]: import pandas as pd  
import numpy as np
```

```
In [2]: sick = pd.read_excel(r'C:\Users\USB\Desktop\프로젝트\csv\그래프용\환자연령비율\J18.xls', encoding = 'UTF-8')
```

```
In [3]: sick
```

```
Out [3]:
```

코드	성별구분	연령구분10세	환자수	환자수.1	환자수.2	환자수.3	환자수.4	환자수.5	환자수.6	환자수.7	환자수.8	환자수.9	환자수.10	환자수.11	총환자수	
0	J18	남	소계	66661	57803	73542	65025	75710	57545	47715	44510	52300	55820	77780	84628	779039
1	J18	남	0_9세	30559	28701	39833	47572	41184	30124	24440	21232	27025	31204	48660	46788	417322
2	J18	남	10_19세	3232	2248	3829	5266	3634	2584	2023	1964	3015	2320	3371	5508	38994
3	J18	남	20_29세	1687	1317	1595	1822	1496	1106	912	1053	1282	1212	1362	1672	15604
4	J18	남	30_39세	2941	2369	2575	2773	2515	1754	1445	1620	1881	2025	2555	3020	27473
5	J18	남	40_49세	3606	2663	3174	3464	3054	2224	1812	1897	2244	2376	3047	3960	33521
6	J18	남	50_59세	4894	3884	4298	4788	4484	3411	2901	2769	2858	2848	3500	5005	45640
7	J18	남	60_69세	6607	5366	5999	6425	6281	5106	4156	4087	4195	4215	4869	6424	63730
8	J18	남	70_79세	7658	6556	7171	7679	7602	6510	5591	5546	5570	5529	5974	7226	78612
9	J18	남	80세이상	5507	4722	5102	5260	5481	4741	4456	4367	4250	4098	4463	5052	57499
10	J18	여	소계	73613	61389	77595	90899	80110	59254	46566	42597	51907	55911	79212	91341	810394

```
In [4]: si = pd.pivot_table(sick, index=['성별구분'], columns=['연령구분10세'], values=['총환자수'], aggfunc=sum, fill_value=0)
```

```
In [5]: si
```

```
Out [5]:
```

성별구분	총환자수									
	연령구분10세	0_9세	10_19세	20_29세	30_39세	40_49세	50_59세	60_69세	70_79세	80세이상
남	417322	38994	15604	27473	33521	45640	63730	78612	57499	779039
여	371915	36253	23462	52240	50601	68996	70999	66497	69750	810394
총	789237	75247	39978	79713	84122	114636	134729	145109	127249	1589433

```
In [6]: import matplotlib.pyplot as plt
```

```
%matplotlib inline
```

```
import platform
```

```
path = 'c://Windows/Fonts/malgun.ttf'  
from matplotlib import font_manager, rc  
if platform.system() == 'Darwin':  
    rc('font', family='AppleGothic')  
elif platform.system() == 'Windows':  
    font_name = font_manager.FontProperties(fname=path).get_name()  
    rc('font', family=font_name)  
else:  
    print('error')
```

```
In [11]: labels = ['0_9세', '10_19세', '20_29세', '30_39세', '40_49세', '50_59세', '60_69세', '70_79세', '80세이상']
```

```
sizes3 = [789237, 75247, 39978, 79713, 84122, 114636, 134729, 145109, 127249]
```

```
explode = (0, 0, 0, 0, 0, 0, 0, 0, 0)
```

```
fig3, ax3 = plt.subplots()
```

```
ax3.pie(sizes3, explode=explode, labels=labels, autopct='%1.1f%%',  
        shadow=True, startangle=90)
```

```
ax3.axis('equal')
```

```
plt.title ('폐렴 환자 연령비율')
```

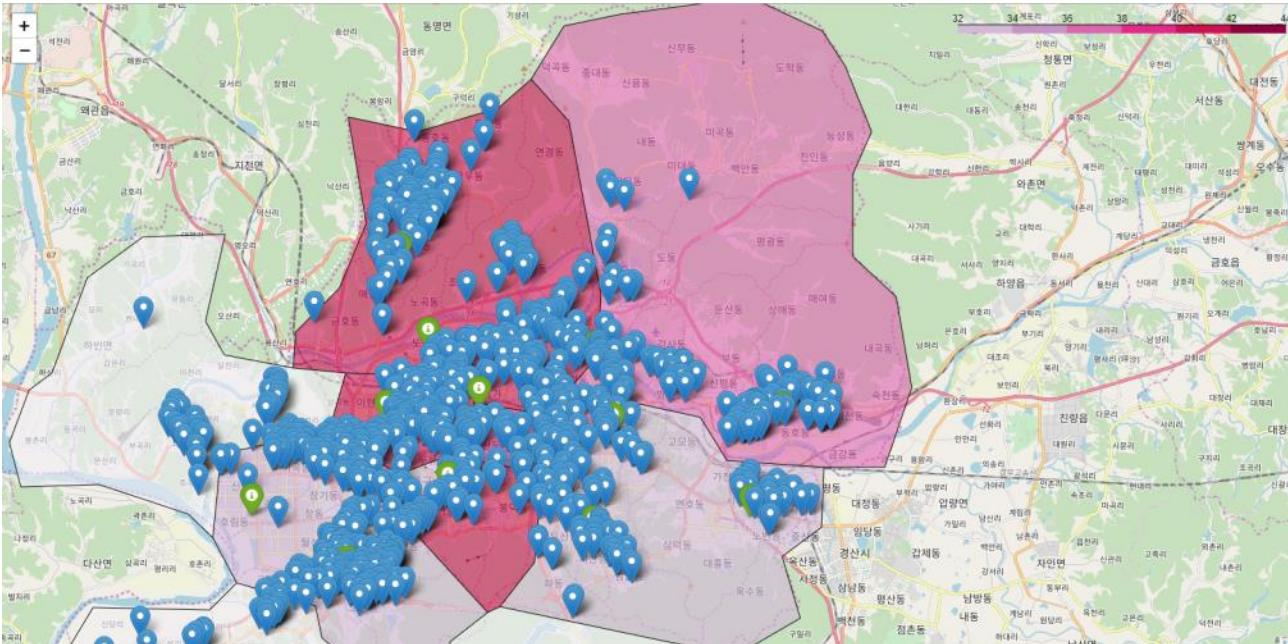
```
plt.tight_layout()
```

```
plt.show()
```

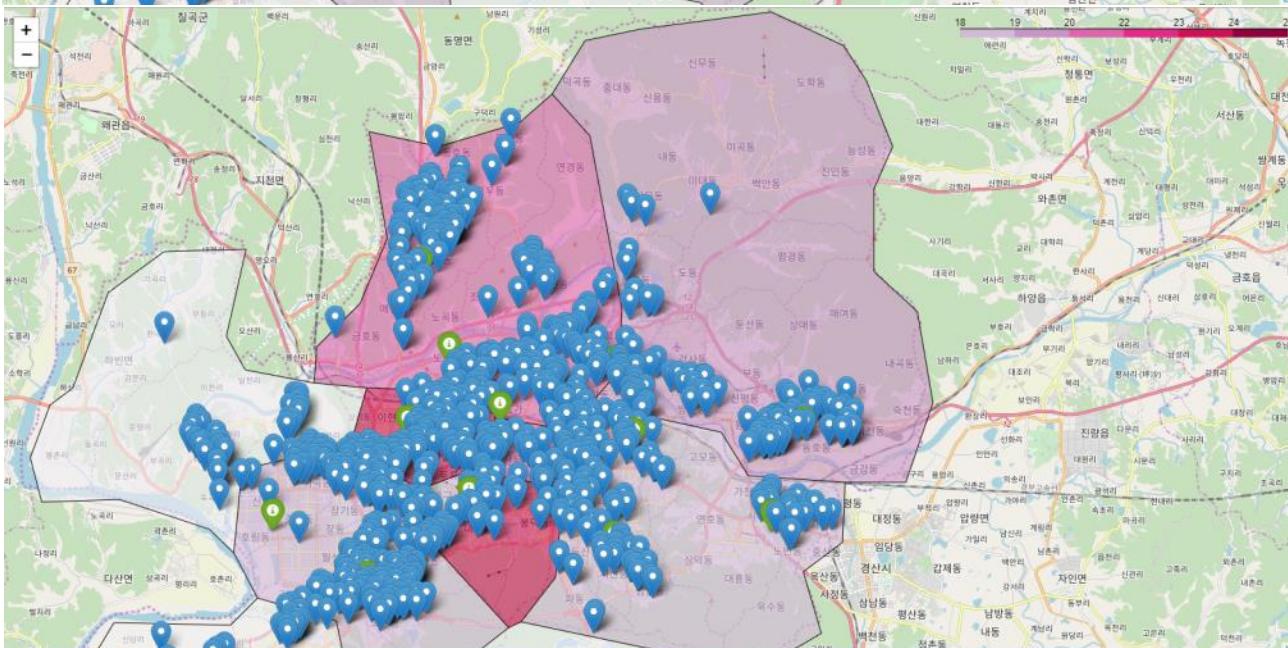
2017년 1월부터 12월까지의 각 질병 별 월간 환자수를 연령별로 나눈 데이터를 가진 엑셀파일을 불러 와pivot table로 정리

2017년 총 환자수를 파악하고 연령에 따른 비율을 보여주는 파이그래프 구현

어린이집 분포



[PM10](#)



대구 어린이집 위치와
미세먼지 구별 분포도

2018 연평균 기준

[PM2.5](#)

출처 : 통계청

어린이집 분포도 설명

1	상가업소번호	상호명	지점명	상권업종중분류코드	시도명	도로명주소	위도
2	19917773	대구통닭	통통치킨상	Q05	닭/오리요리	27 대구광역시 대구광역시	128.546 35.81755
3	19919475	빵마루		Q08	제과제빵점	27 대구광역시 대구광역시	128.6288 35.83087
4	19925178	오구피자		Q07	패스트푸드	28 인천광역시 인천광역시	126.6786 37.44568
5	19935229	무진장횟집		Q03	일식/수산	31 울산광역시 울산광역시	129.3097 35.54567
6	19927968	우리집양념		Q09	유흥주점	28 인천광역시 인천광역시	126.7009 37.47282
7	19932479	콩나물술뚜껑오겹살		Q01	한식	28 인천광역시 인천광역시	126.6753 37.41928
8	19920606	보리.보리		Q01	한식	28 인천광역시 인천광역시	126.6931 37.61841

457162 16111898 일신양꼬치 Q09 유흥주점 28 인천광역시 인천광역시 126.7455 37.48448



```
MariaDB [project]> load data infile '/var/lib/mysql/유아교육.csv' into table 유아교육 fields terminated by ',' lines terminated by '\n' (상가업소번호, 상호명, 상권업종중분류코드, 상권업종중분류명, 상권업종분류명, 시도코드, 도로명주소);
```

약 45만개의 상가정보 데이터를 가진 엑셀파일을 ubuntu 환경에서 mysql의 데이터베이스 내부의 테이블에 입력



```
MariaDB [project]> create view 아동교육 as select 상가업소번호, 상호명, 상권업종분류명, 도로명주소 from 유아교육 where 상권업종중분류명 = '유아교육' and 시도코드 = 27;
```

Sql구문을 이용해 대구의 유아교육 사업장 데이터를 view로 만듬



상가업소번호	상호명	상권업종중분류명	도로명주소
4606150	브라운스톤어린이집	유아교육	대구광역시 북구 내곡
8076000	강림어린이집	유아교육	대구광역시 달성군 옥
10361800	꼬마시티어린이집	유아교육	대구광역시 동구 동북
10382300	꼬마대통령어린이집	유아교육	대구광역시 달성군 옥
10398300	퀸즈베이비어린이집	유아교육	대구광역시 서구 평리
10407500	하모니어린이집	유아교육	대구광역시 달서구 송
10408800	대현어린이집	유아교육	대구광역시 북구 대현
10418400	해달별어린이집	유아교육	대구광역시 남구 대명

어린이집 분포도 설명

```
</head>
<body>

<?php
$conn = mysqli_connect(
'localhost',
'',
'',
'project');

$sql = "SELECT *FROM 아동교육";
$result = mysqli_query($conn, $sql);
echo "<table border='1'><tr><th>상가업소번호</th><th>상호명</th><th>상권업종분류명</th><th>도로명주소</th>";
$n=1;

while($row= mysqli_fetch_array($result))
{
echo "<tr>";
echo "<td class='상가업소번호 '>".$row[0]."</td>";
echo "<td class='상호명 '>".$row[1]."</td>";
echo "<td class='상권업종분류명 '>".$row[2]."</td>";
echo "<td class='도로명주소 '>".$row[3]."</td>";
echo "</tr>";
$n++;
}
mysqli_close($con);
```



```
In [3] : url_base = "http://192.168.0.82"
url_sub = "/test4.php"
url = url_base + url_sub

html = urlopen(url)
soup = BeautifulSoup(html, "html.parser")

soup
```

```
Out [3] : <!DOCTYPE html>
<html>
<head>
<meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>
<style>
table{border:1px solid gray; border-collapse:collapse;}
td{border:1px solid gray;padding:5px;}
</style>
</head>
<body>
<table <tr="" border="1"><th>상가업소번호</th><th>상호명</th><th>상권업종분류명</th><th>도로명주소</th></tr><td class="상가업소번호 ">4806148</td><td class="상호명 ">보라운스톤어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">대구광역시 북구 내곡로 89</td></tr><tr><td class="상가업소번호 ">8076001</td><td class="상호명 ">강림어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">대구광역시 달성군 록포면 울미로 51</td></tr><tr><td class="상가업소번호 ">10361846</td><td class="상호명 ">꼬마시티어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">대구광역시 동구 풍족로 500</td></tr><tr><td class="상가업소번호 ">10382256</td><td class="상호명 ">교마대통령어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">대구광역시 달성군 록포면 비슬로 447길 11</td></tr><tr><td class="상가업소번호 ">10398806</td><td class="상호명 ">린즈베이비어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">10407515</td></tr><tr><td class="상호명 ">하모니어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">대구광역시 서구 펄리로 236</td></tr><tr><td class="상호명 ">10418419</td><td class="상호명 ">해달별어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">대구광역시 달구벌대로 47-2</td></tr><tr><td class="상가업소번호 ">10425404</td><td class="상호명 ">워너스타디어린이집</td><td class="상권업종분류명 ">유아교육</td><td class="도로명주소 ">대구광역
```

Mysql에 저장된 뷰를 php
와 연동하여 데이터베이스
의 내용을 웹으로 만듬

Html의 body부분에 php
구문을 넣고 각 칼럼별로
클래스를 주어 크롤링 하
기 편리하도록 작성

어린이집 분포도 구현설명

```
from bs4 import BeautifulSoup
import pandas as pd
import numpy as np
from urllib.request import urlopen
```

```
In [4]: 상가업소번호 = []
상호명 = []
상권업종분류명 = []
도로명주소 = []

list_soup = soup.find_all('tr')

for item in list_soup:
    상가업소번호.append(item.find(class_='상가업소번호').get_text())
    상호명.append(item.find(class_='상호명').get_text())
    상권업종분류명.append(item.find(class_='상권업종분류명').get_text())
    도로명주소.append(item.find(class_='도로명주소').get_text())
```

위의 웹에서 텍스트를 크롤링하여 DataFrame을 만듬

```
In [5]: data = {'상가업소번호':상가업소번호, '상호명':상호명, '상권업종분류명':상권업종분류명, '도로명주소':도로명주소}
df = pd.DataFrame(data)
df.head()
```

Out [5]:

	도로명주소	상가업소번호	상권업종분류명	상호명
0	대구광역시 북구 내곡로 89	4606148	유아교육	브라운스톤어린이집
1	대구광역시 달성군 옥포면 돌미로 51	8076001	유아교육	강림어린이집
2	대구광역시 동구 동북로 500	10361846	유아교육	꼬마시티어린이집
3	대구광역시 달성군 옥포면 비슬로447길 11	10382256	유아교육	꼬마대통령어린이집
4	대구광역시 서구 광리로 236	10398306	유아교육	원즈베이비어린이집

어린이집 분포도 구현설명

```
In [94]: import folium  
import googlemaps  
from tqdm import tqdm_notebook
```

```
In [95]: gmaps_key = ████████████████████████████████████████  
gmaps = googlemaps.Client(key=gmaps_key)
```

```
In [96]: df.set_index('상가업소번호', inplace=True)  
df.head()
```

상호명	상권업종분류명	도로명주소
상가업소번호		
4606148	브라운스톤어린이집	유아교육 대구광역시 북구 내곡로 89
8076001	강림어린이집	유아교육 대구광역시 달성군 옥포면 돌미로 51
10361846	꼬마시티어린이집	유아교육 대구광역시 동구 동북로 500
10382256	꼬마대통령어린이집	유아교육 대구광역시 달성군 옥포면 비슬로447길 11
10398306	퀸즈베이비어린이집	유아교육 대구광역시 서구 평리로 236

Googlemaps API를 통하여 위도와 경도 추출

```
In [97]: lat = []  
lng = []  
  
for n in tqdm_notebook(df.index):  
    if df['도로명주소'][n] != 'Multiple':  
        target_name = df['도로명주소'][n] + ', 대구광역시'  
        gmaps_output = gmaps.geocode(target_name)  
        location_output = gmaps_output[0].get('geometry')  
        lat.append(location_output['location']['lat'])  
        lng.append(location_output['location']['lng'])  
  
    else:  
        lat.append(np.nan)  
        lng.append(np.nan)
```

어린이집 분포도 구현설명

```
In [98]: df['lat'] = lat  
df['lng'] = lng  
df.head()
```

Out [98]:

상호명	상권업종분류명	도로명주소	lat	lng
상가업소번호				
4606148	브라운스톤어린이집	유아교육 대구광역시 북구 내곡로 89	35.902610	128.514867
8076001	강림어린이집	유아교육 대구광역시 달성군 옥포면 돌미로 51	35.789632	128.463792
10361846	꼬마시티어린이집	유아교육 대구광역시 동구 둘복로 500	35.880064	128.639470
10382256	꼬마대통령어린이집	유아교육 대구광역시 달성군 옥포면 비슬로447길 11	35.790807	128.465422
10398306	퀸즈베이비어린이집	유아교육 대구광역시 서구 광리로 236	35.860811	128.548963

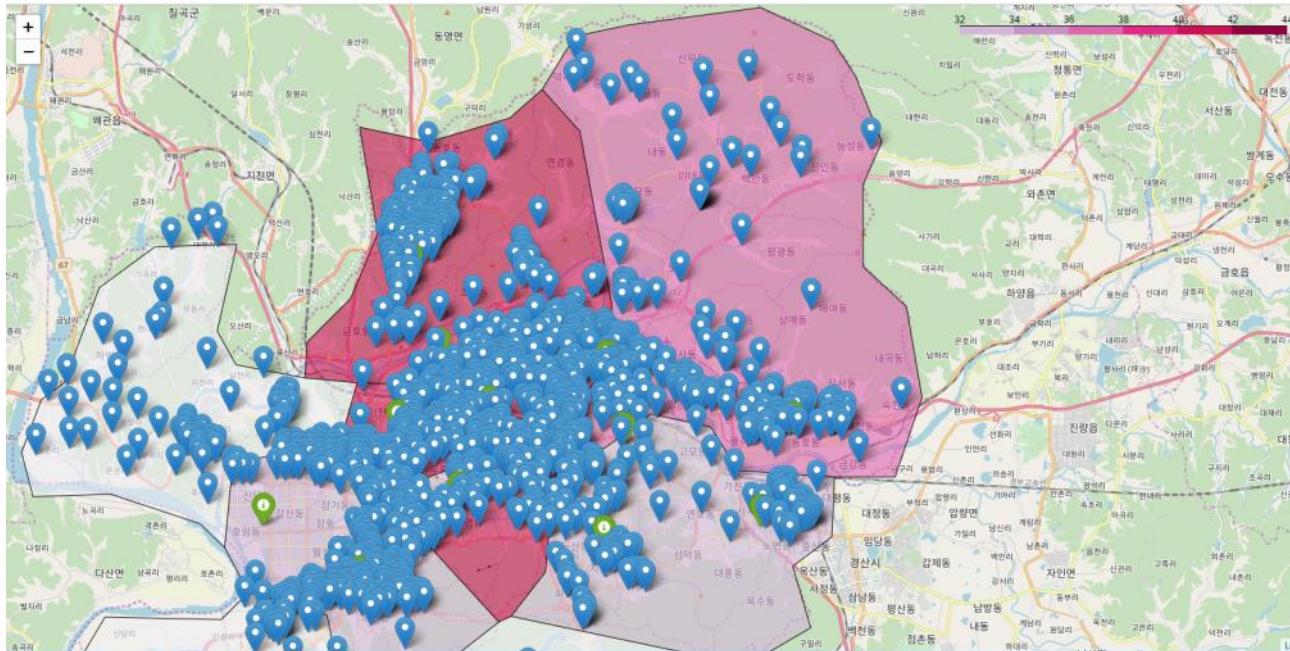


```
In [109]: mapping = folium.Map(location=[df['lat'].mean(), df['lng'].mean()], zoom_start=11)  
  
mapping.choropleth(geo_data=geo_data,  
                   data=gu_data['연_평균값'],  
                   columns=[gu_data.index, '연_평균값'],  
                   fill_color='PuRd',  
                   key_on='feature.id')  
  
for n in df.index:  
    if df['도로명주소'][n] != 'Multiple':  
        folium.Marker([df['lat'][n], df['lng'][n]], popup=df['상호명'][n]).add_to(mapping)  
  
for x in air.index:  
    if air['측정소'][x] != 'Multiple':  
        folium.Marker([air['lat'][x], air['lng'][x]], popup=air['측정소'][x], icon=folium.Icon(color='green', icon='info-sign')).add_to(mapping)  
  
mapping
```

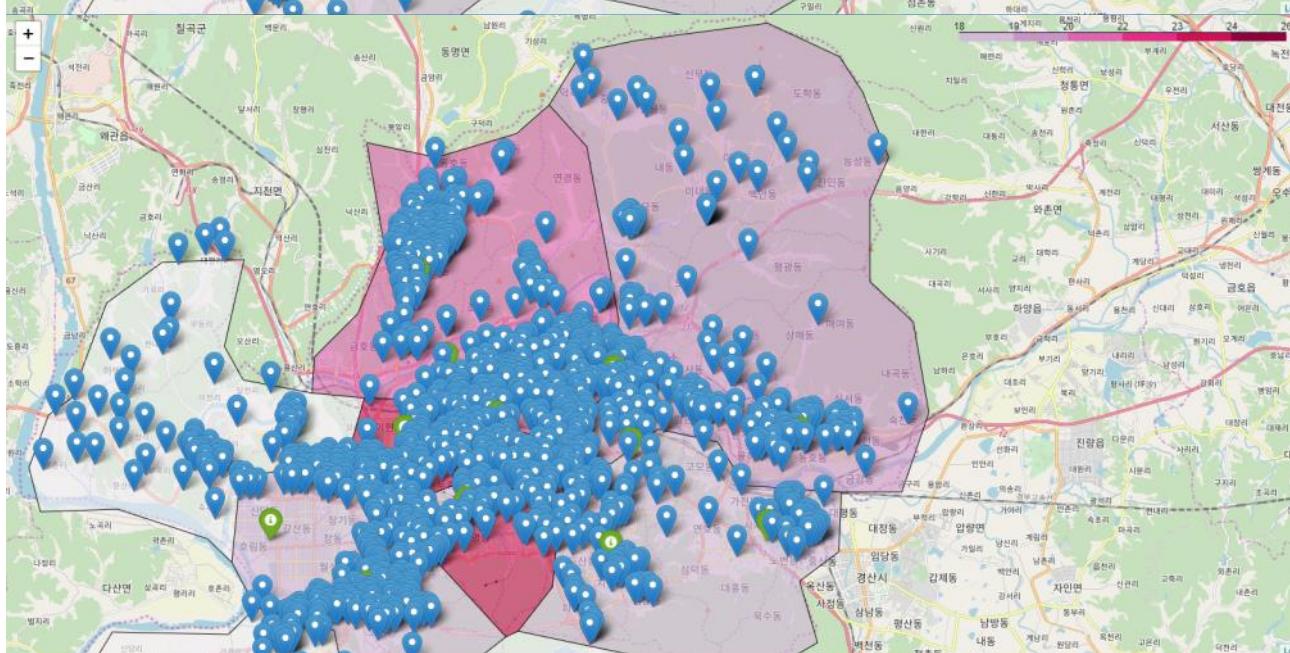
추출한 위도, 경도를
DataFrame에 삽입

20페이지의 folium지도와
합쳐 어린이집은 기본마커,
측정소는 녹색마커로 지도
에 표시

경로당 분포



[PM10](#)



대구 경로당 위치와
미세먼지 구별 분포도

2018 연평균 기준

[PM2.5](#)

출처 : 통계청

```
In [4]: from tqdm import tqdm_notebook
```

```
In [5]: for name in tqdm_notebook(raw_data['주소']):  
    tmp = gmaps.geocode(name, language='ko')  
    try:  
        raw_data.loc[raw_data['주소'] == name, 'formatted_address'] = tmp[0].get("formatted_address")  
        tmp_loc = tmp[0].get("geometry")  
        raw_data.loc[raw_data['주소'] == name, 'lat'] = tmp_loc['location']['lat']  
        raw_data.loc[raw_data['주소'] == name, 'lng'] = tmp_loc['location']['lng']  
    except IndexError:  
        print(tmp)  
        pass  
raw_data.head()
```

어린이집 분포도와 방법은
같으나 위도, 경도를 추출
하는 코드만 바뀜



판매

관련제품의 필요성
미세먼지 관련 뉴스
판매전략
판매전략(연령)
판매전략(지역)
판매전략(마스크)
판매전략(공기청정기)
결론
CF 콘티

공기청정기를 사용하는 것은 실내 미세먼지를 줄이는데 도움이 됩니다. 특히 최근 공기청정기를 사용하는 것은 실내 미세먼지 농도를 줄이는데 도움이 됩니다. 물론 주기적으로 공기청정기 필터를 교체하고 적정하게 관리하는 것은 반드시 필요합니다.

물론 주기적으로 공기청정기 필터를 교체하고 적정하게 관리하는 것은 반드시 필요합니다.



식약처 인증 마스크
염화나트륨 투과 테스트 결과
(99.14%의 차단률을 보임)

미세먼지관련상품판매 기획안

1. 담당부서

판매업체 기획, 경영, 마케팅팀

2. 판매기간

미세먼지 인식이 높은 지금, 봄 3~4월 황사철 집중판매

3. 판매대상

취약계층(어린이, 노인)의 가구주인 4050, 미세먼지 심한 곳

4. 판매목록

공기청정기 마스크등 관련제품들

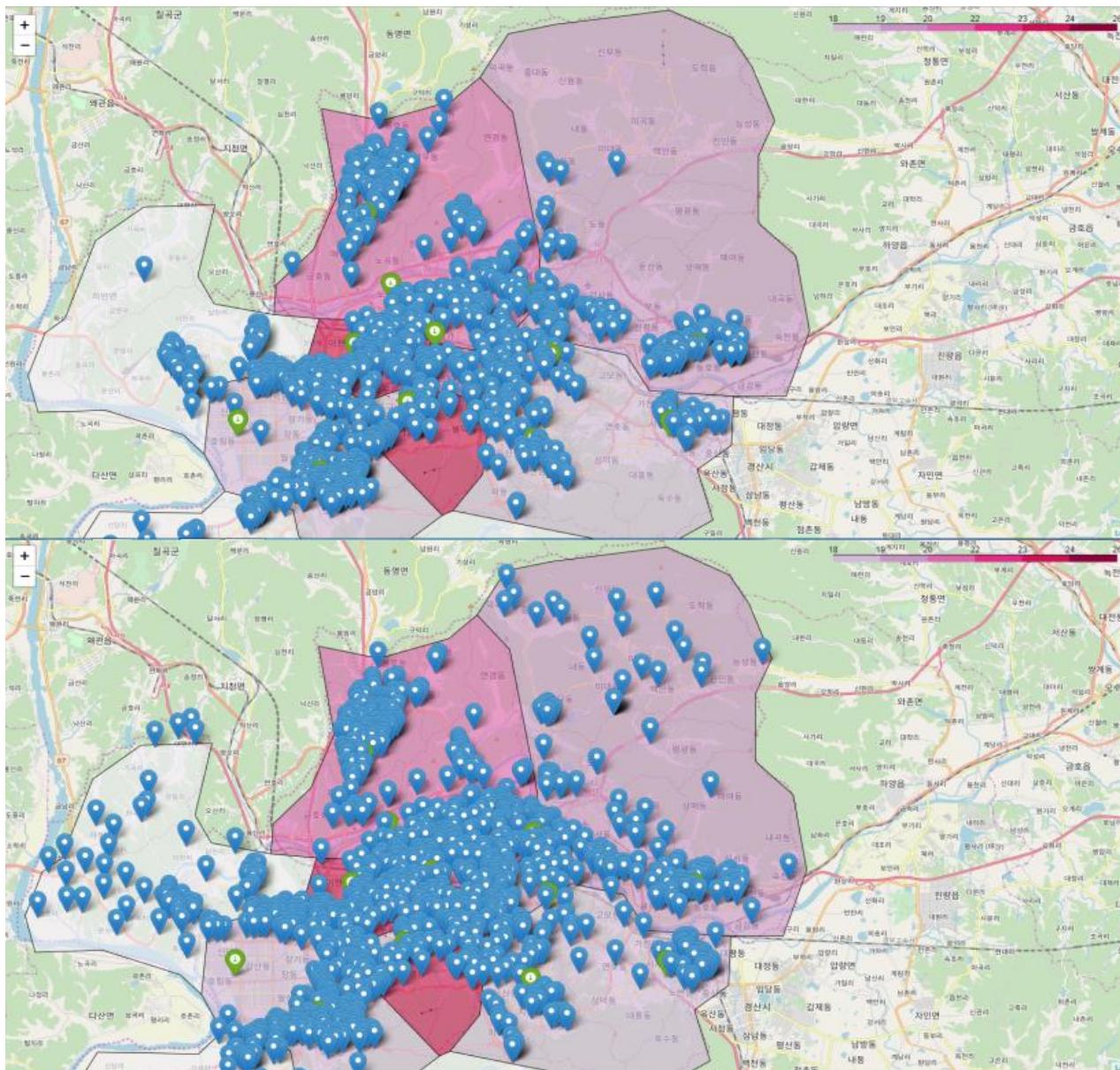
5. 판매전략

홈페이지(<http://Bgroup.ml>), PDF파일, SNS, YOUTUBE, TV 광고등을 이용하여 홍보 후 판매

6. 판매목적

회사매출증가, 사회적기업, 국민건강증진

판매전략(지역)



PM2.5

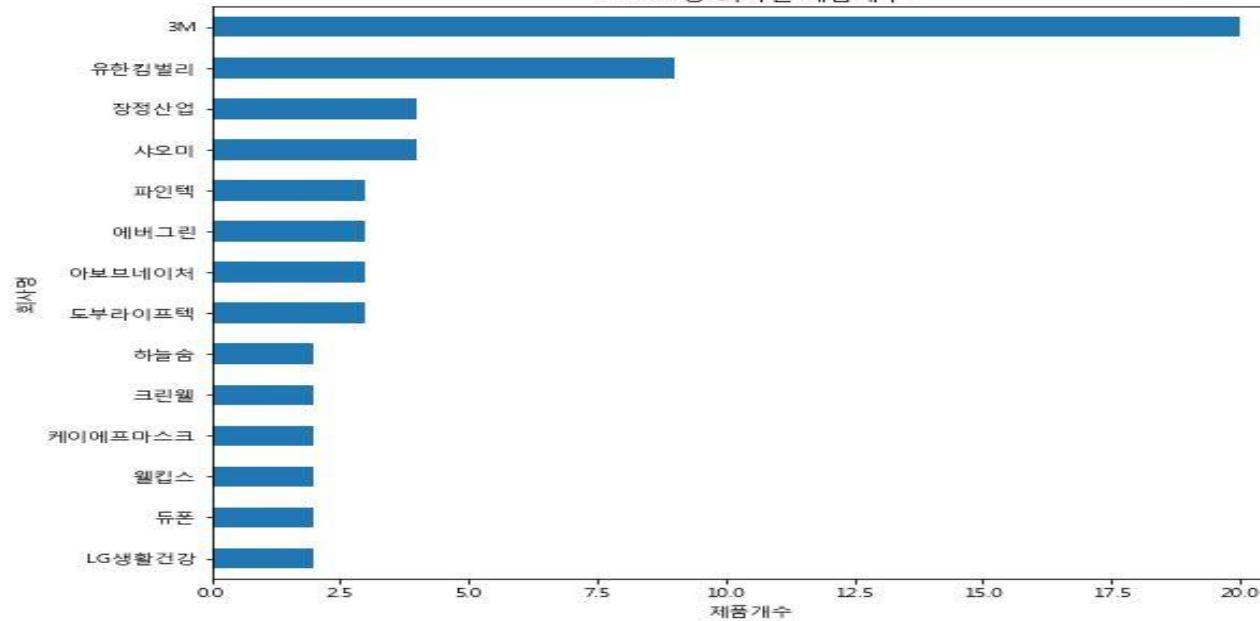
영유아를 양육하는 3040
고령자 부양하는 4050 타겟팅
지역은 서구 북구 남구 중구
위주

2018 연평균 기준

PM2.5

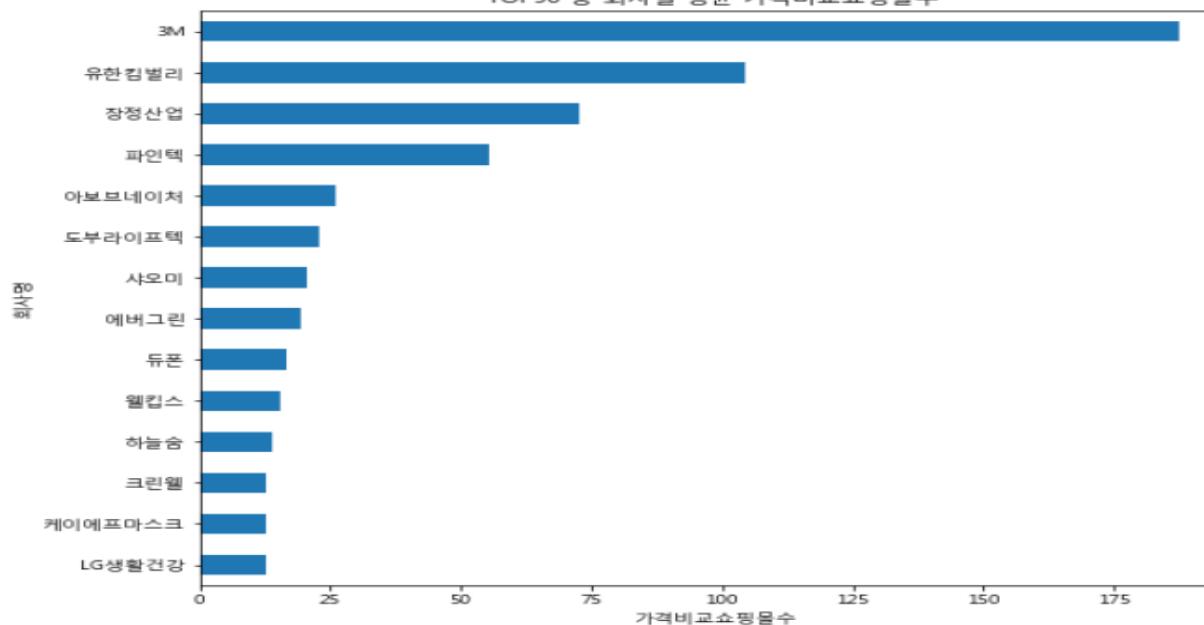
판매전략(마스크)

TOP90 중 회사별 제품개수

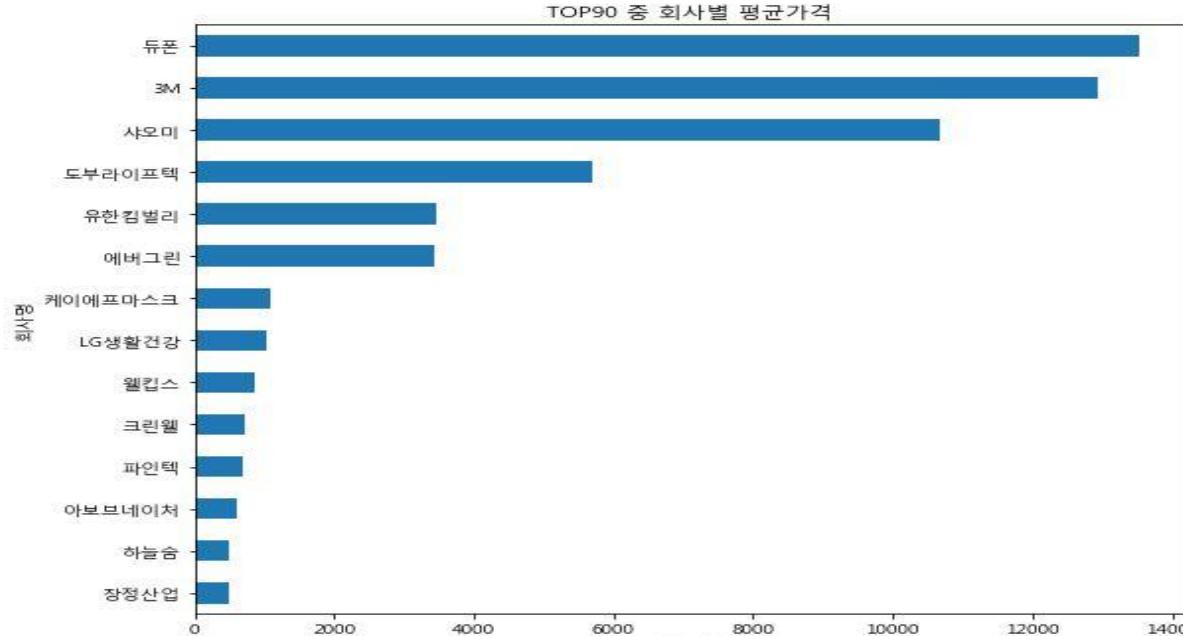


3M 유한킴벌리 장정산업이
인기상품 점유율이 높다.

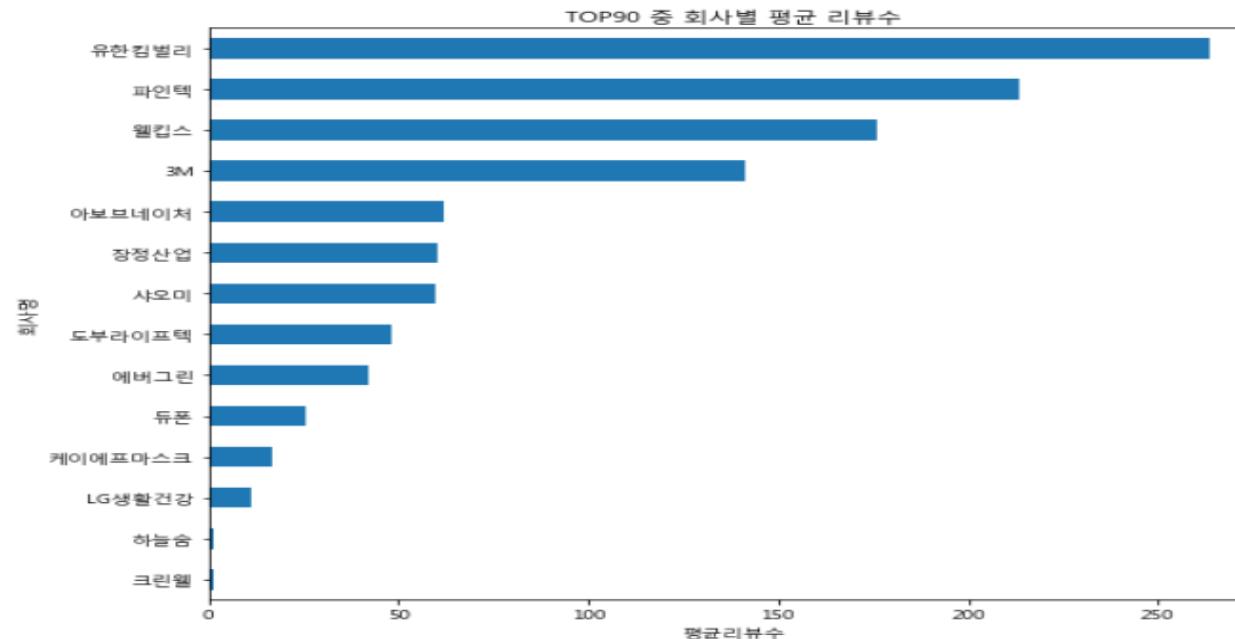
TOP90 중 회사별 평균 가격비교쇼핑몰수



판매전략(마스크)



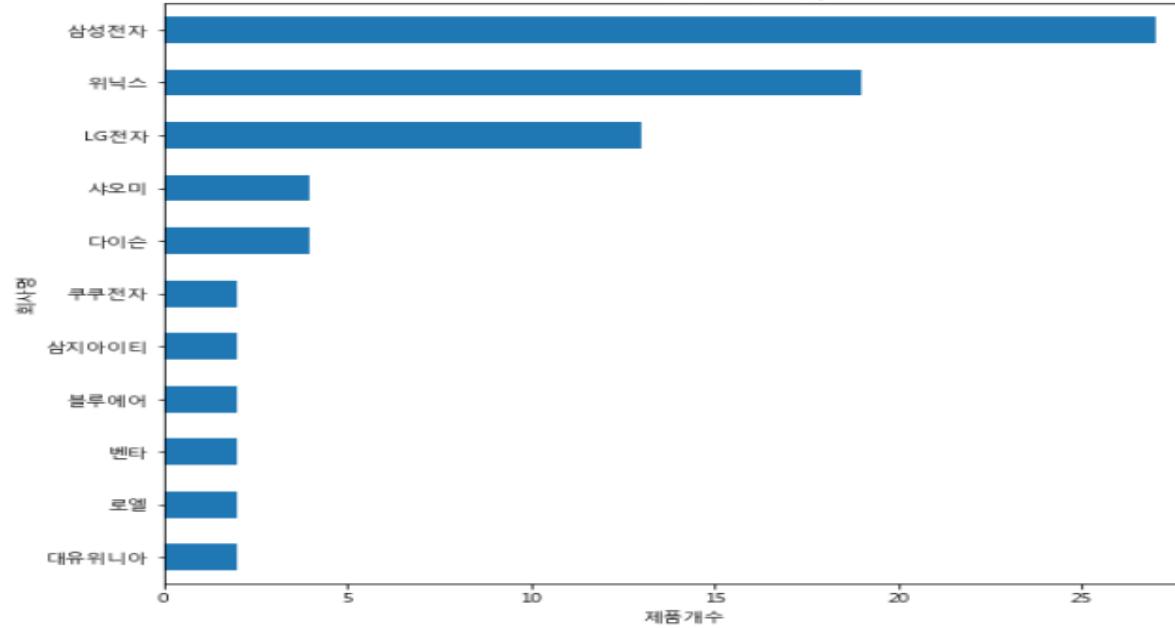
파인텍 웰킵스 제품이
가격은 싸지만



유한킴벌리 제품을 제외한
다른 제품들에 비해
리뷰수(구매수)가 압도적임

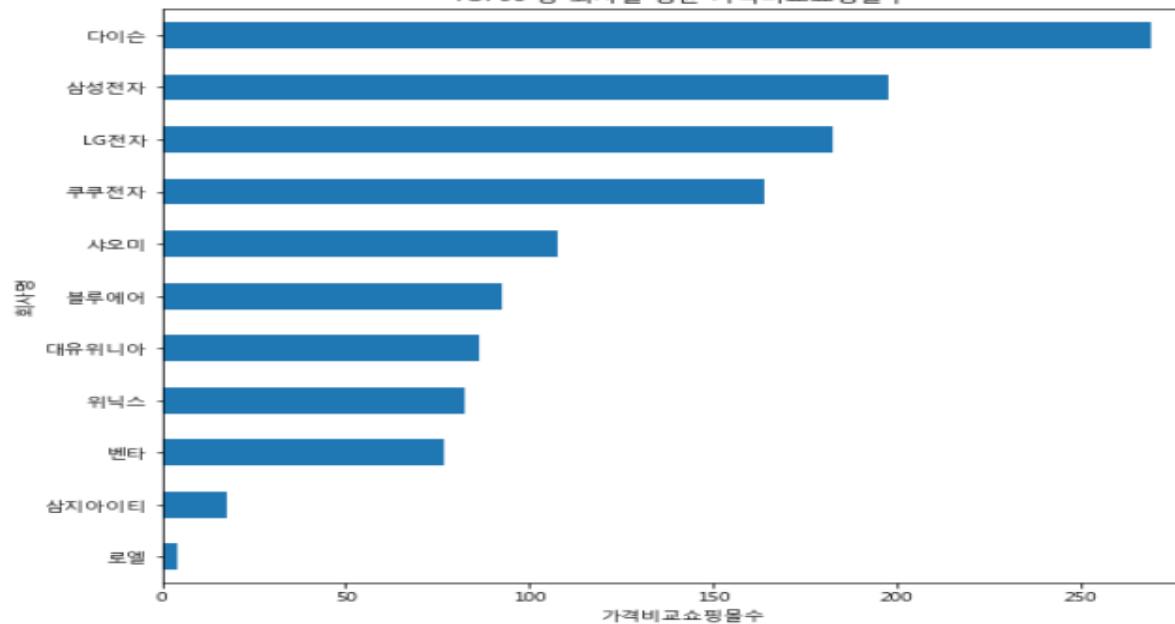
판매전략(공기청정기)

TOP90 중 회사별 제품개수



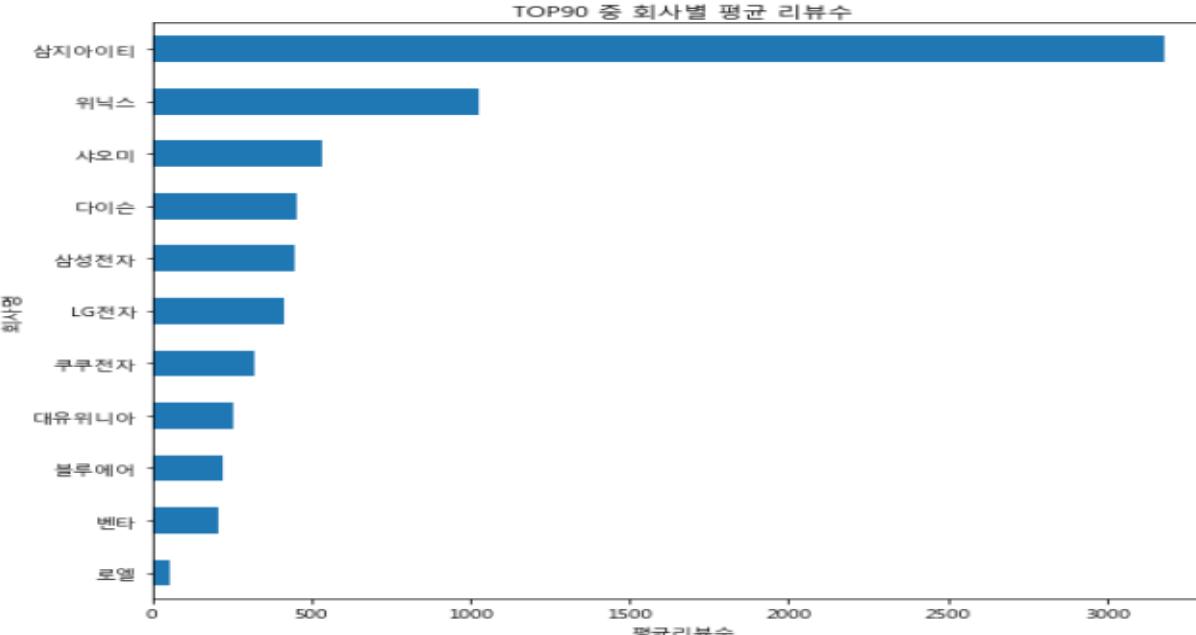
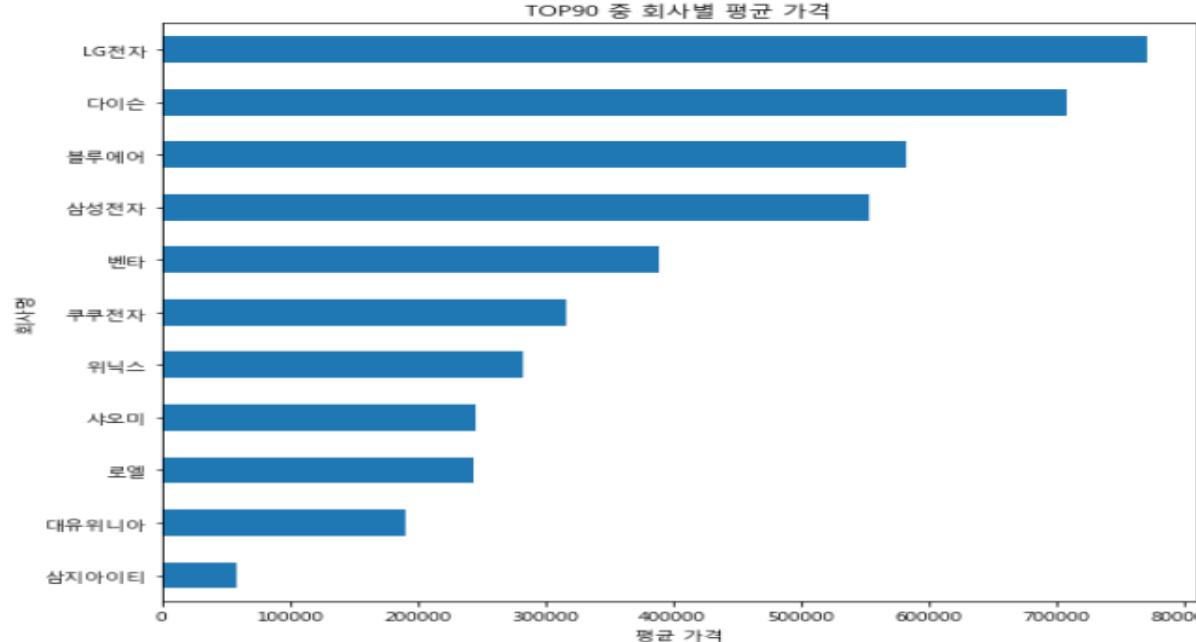
삼성 LG가 인기상품
점유율이 높다

TOP90 중 회사별 평균 가격비교쇼핑몰수



출처 : 다나와

판매전략(공기청정기)



삼지아이티가 가격도 싸고
리뷰수(구매수)도
압도적으로 많다.

미세먼지관련상품판매 기획안 결론

1. 판매지역 : 대구광역시 서구, 북구, 남구, 중구

2. 타겟 연령층 : 30~50대

3. 홍보 및 판매전략 집중이 필요한 주요 회사상품 :

마스크 : 3M, 유한킴벌리, 파인텍, 웰킵스, 장정산업

가성비 – 장정산업 애니가드 방진마스크 VC201G

최고 인기상품 – 파인텍 네퓨어 황사방역용 마스크 KF94 대형

공기청정기 : 삼성, LG, 위닉스, 삼지아이티

가성비 – 삼지아이티 PISnet 피스넷 에어, 퓨어 360

최고 인기상품 – 삼성전자 블루스카이 AX60N5580WDD

홈페이지를 통해 ‘다나와’ 사이트를 연결해서

자세한 인기상품을 볼 수 있음

구매 가이드라인 세일정보등 도움되는 정보들을 제공

크롤링 구현설명

```
{ launch.json      다나와서치.py x    릴리스 정보: 1.31.0
1 import time
2 from selenium import webdriver as wd
3 from bs4 import BeautifulSoup as bs
4 from selenium.webdriver.common.by import By
5 # 명시적 대기를 위해
6 from selenium.webdriver.support.ui import WebDriverWait
7 from selenium.webdriver.support import expected_conditions as EC
8 from Item import ItemInfo
9
```



```
for page in range(1, 4):
    try:
        # 자바스크립트 구동하기
        driver.execute_script("getPage(%s); return false;" % page)
        time.sleep(1)
        print("%s 페이지 이동" % page)
        #####
        # 상품명, 가격, 별점, 리뷰수, 가격비교불
        prod_item = driver.find_elements_by_css_selector('.product_')

        for li in prod_item:
            print ('상품명', li.find_element_by_css_selector('.prod_')
            print ('가격', li.find_element_by_css_selector('.prod_ma
            print ('별점', li.find_element_by_css_selector('.prod_ma
            print ('리뷰수', li.find_element_by_css_selector('.prod_
            print ('가격비교불', li.find_element_by_css_selector('.p
except Exception as e1:
    pass
```

Vscode사용

Selenium과
chromedriver를 이용해
상품, 가격, 별점, 리뷰수,
가격비교불을 크롤링하여
엑셀파일로 추출

For문 구현에 문제가 있어
추출이 안된 부분을 손으
로 작업하는 상황이 있었
음. 보완이 필요함



판매

현재 진행중인 작업
그 외 사항

현재 진행중인 작업

```
1 import requests
2 from urllib.parse import quote
3
4 def get1000Result(keyword):
5     list = []
6     for num in range(1, 11):
7         list = list + call("공기청정기", num*100+1)[‘items’]
8     return list
9
10 def call(keyword, start):
11     encText = quote(keyword)
12     url = "https://openapi.naver.com/v1/search/blog?query=" + encText
13     url += "&display=100" + "&start=" + str(start)
14     result = requests.get(url=url,
15                           headers = {"X-Naver-Client-Id" : "REDACTED",
16                                       "X-Naver-Client-Secret" : "REDACTED"})
17     print(result) # Response [401]
18     return result.json()
```

Vscode 사용

홍보를 협력할 수 있는 블로거를 찾기 위해 네이버 API를 통해 1000개의 블로그를 검색 후 그 정보를 json파일로 만듭니다.

현재 진행중인 작업

```
1 import pandas as pd
2
3 df = pd.read_json("./air.json")
4 print(df.count())
5
6 # 상위 5개 블로거 이름
7 dfTop5 = df.head(5)
8 print [(dfTop5['bloggername'])]
```

문제 6 출력 디버그 콘솔 터미널

1: cmd

```
C:\Users\USER\Desktop\크롤링연습\crawler>python air.py
bloggerlink    1000
bloggername    1000
description    1000
link           1000
postdate       1000
title          1000
dtype: int64
0
.
1    나란 사람의 일상
2    be yourself
3    가구인테리어친구
4    루르의 인테리어
Name: bloggername, dtype: object
```

Json파일을 pandas로 불러와 상위 5개의 블로거
파악

- 크롤러의 성능 개선을 위하여 Node.JS를 통해 웹을 공부 중입니다.
- 판매 전략에 대한 부분이 부실하기 때문에 방법을 모색할 필요가 있어 보입니다.
- R을 이용하여 p-value 등을 이용해 더욱 설득력이 있는 데이터를 제작해야 한다고 생각합니다.

A wide-angle photograph of a rural landscape. In the foreground, there are rolling hills covered in sparse vegetation and small trees. The middle ground shows more hills, some with vegetation and others appearing as dark silhouettes against a hazy sky. The background features a range of mountains visible through a thick layer of fog or mist. The overall atmosphere is serene and somewhat mysterious due to the low light and haze.

THANK YOU
FOR WATCHING