

환경특별시 서울

구별 특성을 고려한 서울시 종합환경대책



- 초미세먼지와 열섬, 녹지 및 태양광에너지를 중심으로 -

팀명 : MSG (Make Seoul Green)

신민용 이유찬 박예리 송수빈

Index

01 분석 배경 및 목적

- 환경문제에 대한 서울시민들의 인식
- 분석목적

02 데이터 분석

- 데이터 전처리 및 전체 분석 순서도
- 트렌드 및 소셜 분석
- EDA
- 모델링(xgboost)

03 분석 결과 활용 및 해결책 제시

- 클러스터링
- 구별 해결책 제시
- 태양광 미니 발전소
- 환경 문제 예측 모델 제안(LSTM)

04 부록

- 한계점
- 데이터 목록
- 분석 툴 및 참고문헌

1. 분석배경 및 목적 - (1) 분석배경

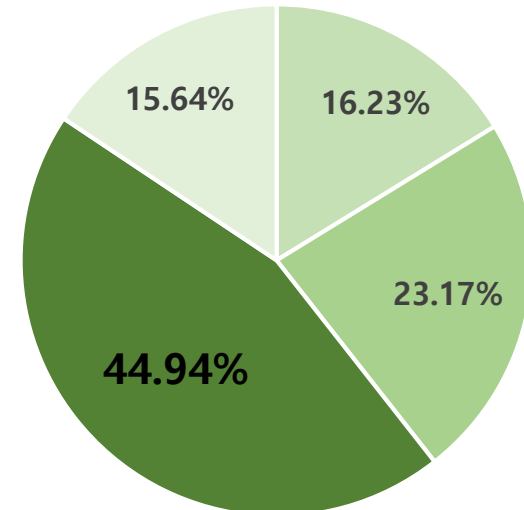
1) 환경문제에 대한 서울시민들의 인식

초열대야+폭염+미세먼지+오존... '헬서울' 4중고

서울 초미세먼지 주의보...서울광장·스케이트장 중단

최대전력수요 작년 여름 최고치 경신...서울 첫 폭염
경보 등 무더위에 전력소비 급증

가장 불안하게 느껴지는
서울시 환경 문제는?



- 지구온난화로 인한 기후변화
- 유해화학물질 방사능 등 유출
- 황사, 미세먼지 유입
- 농약, 화학비료 사용

통계청에서 조사한 서울시 환경문제에 대한 인식 통계에 따르면,

2016년 서울시민이 가장 불안해하는 환경문제 1위는 미세먼지, 2위는 원전에 의한 방사능 유출임을 알 수 있다.

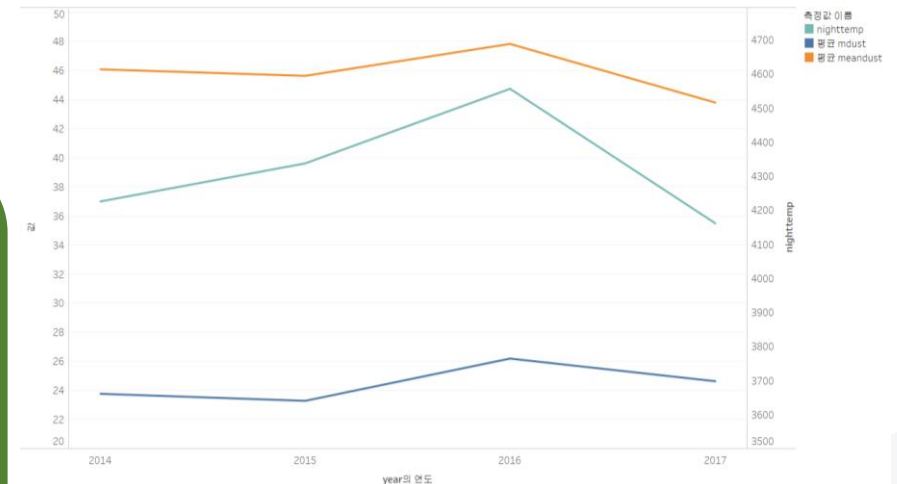
1. 분석배경 및 목적 – (2) 분석 목적

서울시가 2010년 이후 환경 문제를 해결하기 위해 세운 정책들 총 27개 중 대기환경과 관련된 환경 정책은 9개이며, 에너지관련 환경 정책은 7개이다.

ex) '초미세먼지 20% 감축', '원전 하나 줄이기'

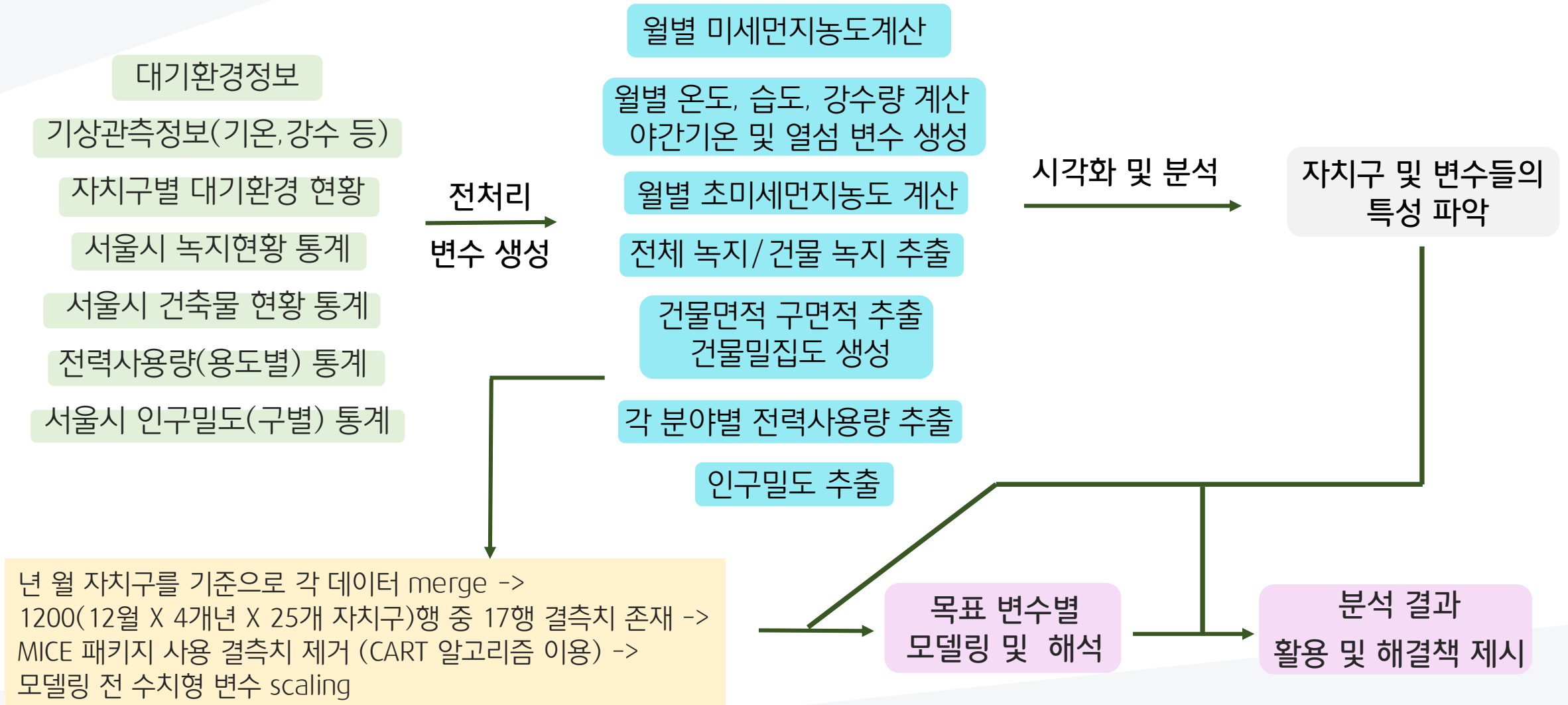
- 미세먼지는 그 자체로도 문제지만(호흡기 질환 유발 등) 열섬현상과 같은 다른 환경문제를 야기하는 원인이 되기도 함.
- 이처럼 환경문제는 여러 원인들이 복합적으로 엮여져 있기 때문에 비단 한 현상만 분석할 것이 아닌 각 현상 간의 연관 관계를 분석하여 종합적인 해결책을 제시할 수 있어야 한다.

초미세먼지, 미세먼지와 열섬현상과의 관계



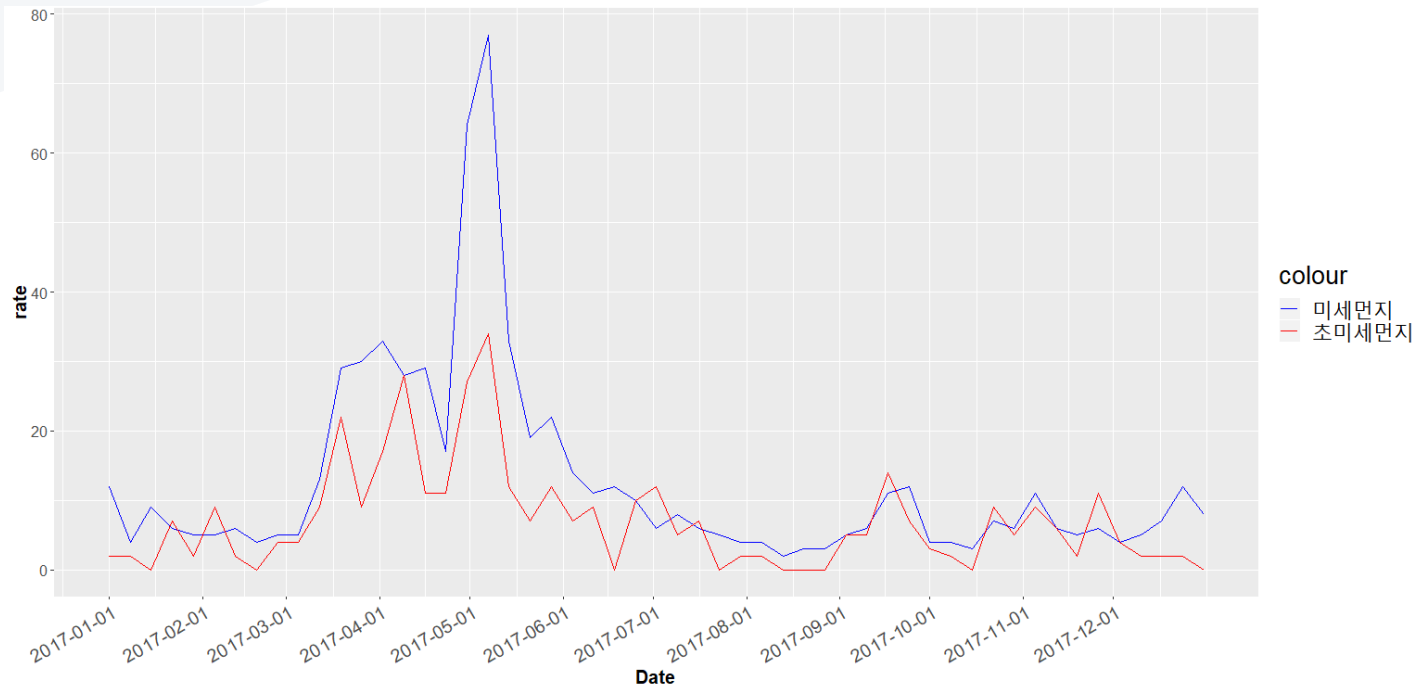
최종적으로는 미세먼지 감축과 신재생 에너지 사용량을 늘려 원전을 줄이는 것을 목표로 자치구별로 분석하고, 각 구마다 정책이 효율적으로 수행될 수 있는 방안을 제시하고자 한다.

2. 데이터 분석 - 데이터 전처리 및 전체 분석 순서도



2. 데이터 분석 - 트렌드 및 소셜 분석

미세먼지, 초미세먼지 구글트렌드



미세먼지에 대한 연관검색어 분석

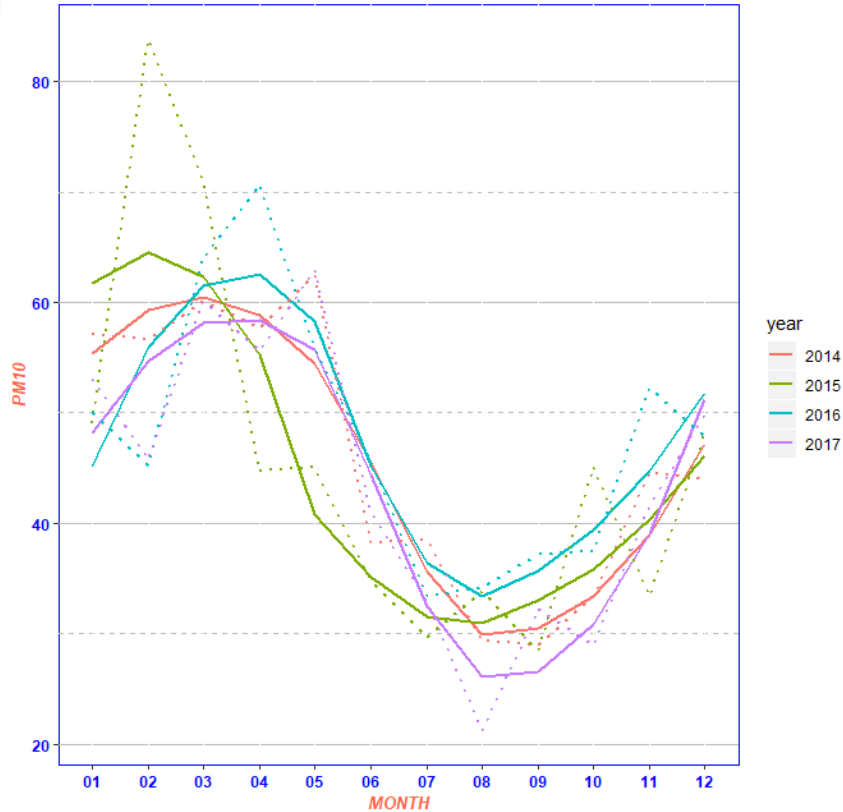


미세먼지/초미세먼지에 대한 관심도를 조사하기 위하여 구글 트렌드를 활용해 데이터를 수집한 후 시각화 ➤ 3월 이후 관심도가 급증 ➤ 확인 결과 3월에 미세먼지 수치가 높게 나타남. ➤ 트렌드의 전체적인 추이를 살피면 봄철마다 관심도가 높은 것을 확인. ➤ 추가로 연관어 분석을 진행 봄과의 연관도 높음 ➤➤ 이를 바탕으로 **초미세먼지는 3월** **미세먼지는 3,4,5월**을 기준으로 데이터를 시각화하기로 함

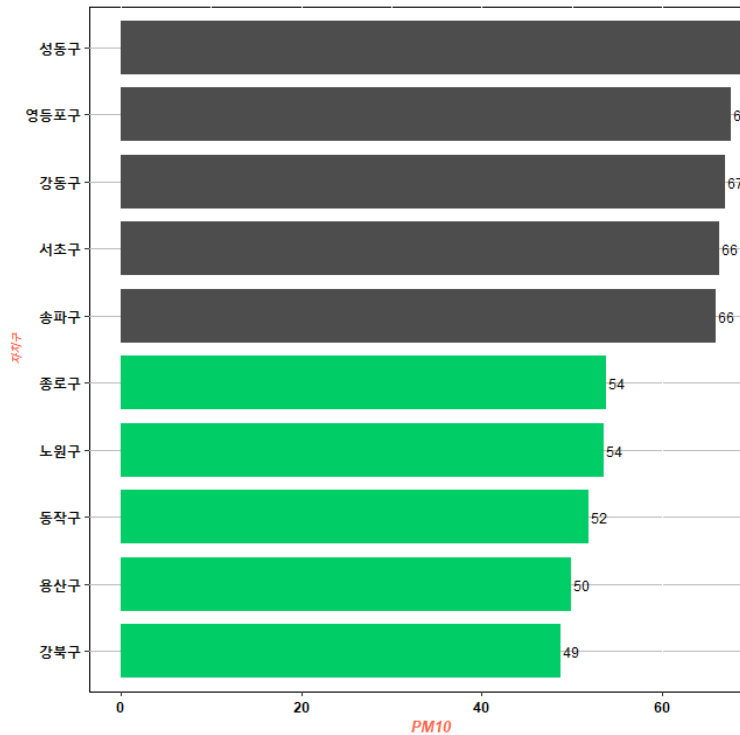
➡ 이처럼 변수의 측정 기준이나 구간을 선정하는데 트렌드 및 소셜 분석 결과를 보충 지표로서 활용하였음

2. 데이터 분석 - 트렌드 및 소셜 분석

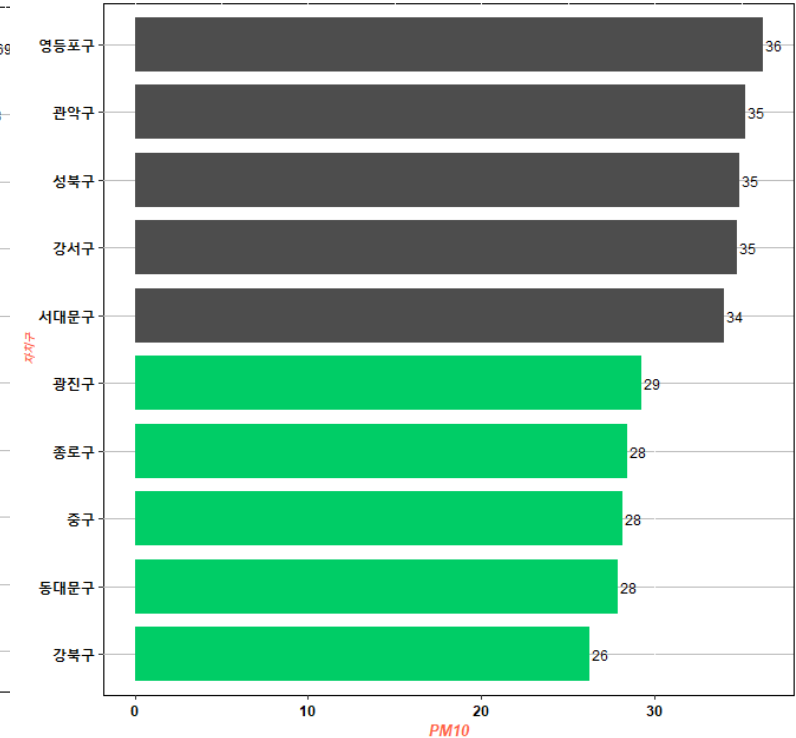
< 연도별 미세먼지 추이 >



2017 봄철 구별 미세먼지 상-하위 5개

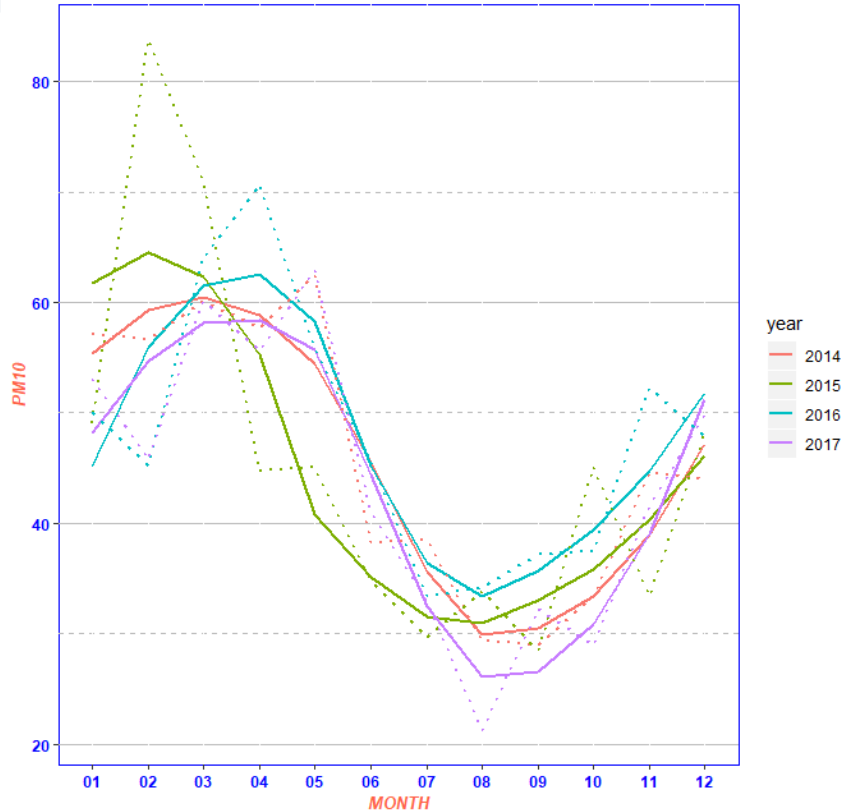


2017 여름철 구별 미세먼지 상-하위 5개

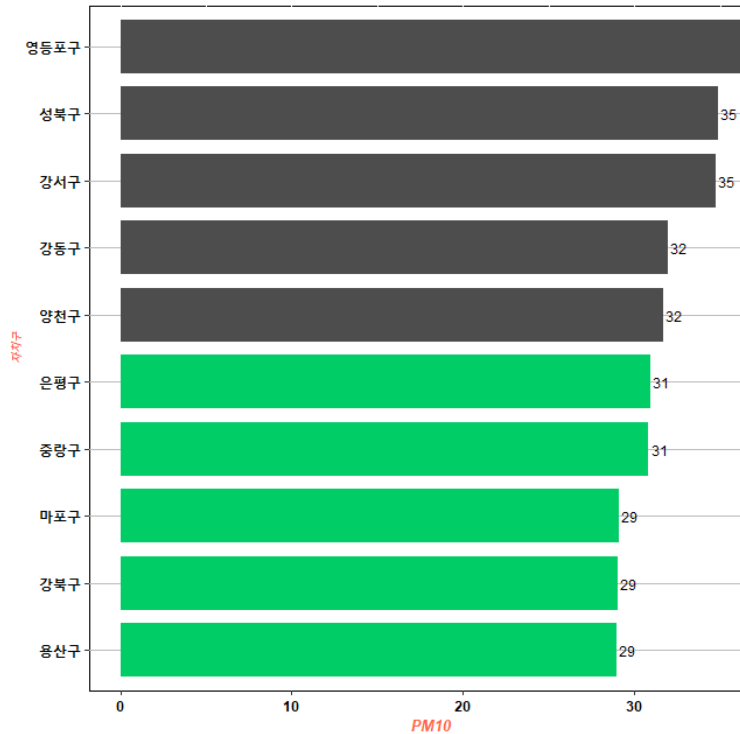


2. 데이터 분석 - 트렌드 및 소셜 분석

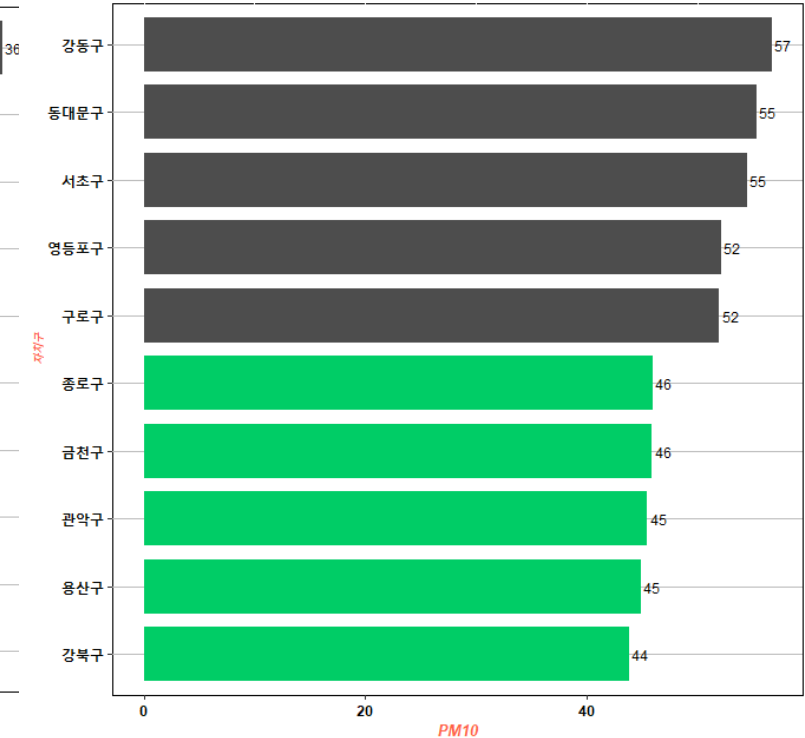
< 연도별 미세먼지 추이 >



2017 가을철 구별 미세먼지 상-하위 5개

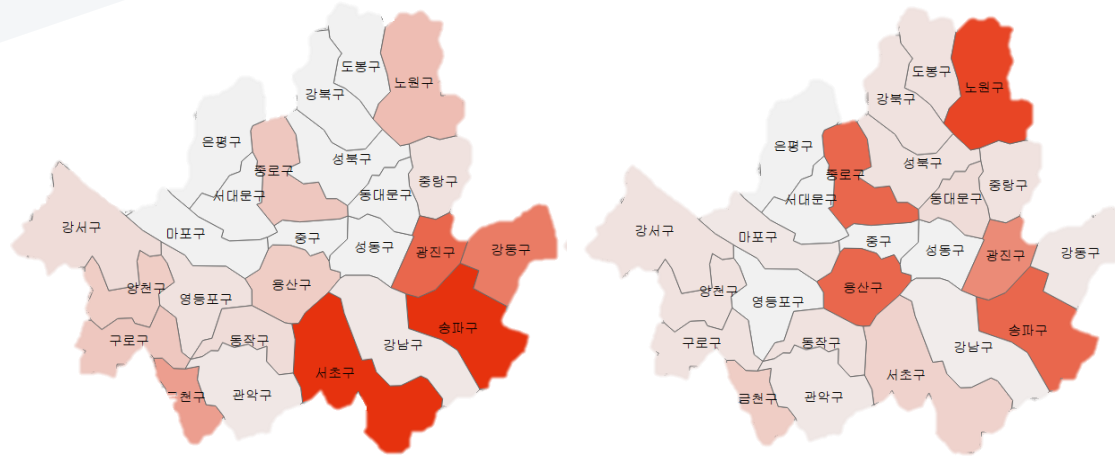


2017 겨울철 구별 미세먼지 상-하위 5개

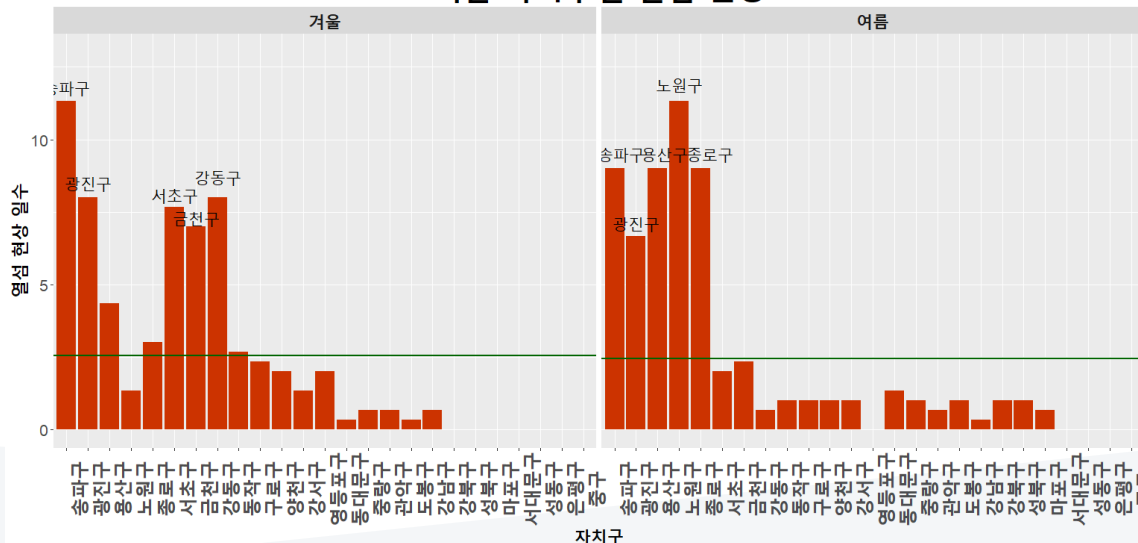


2. 데이터 분석

구별 계절별 열섬 현상 평균횟수



서울 자치구별 열섬 현상



“보편적인 열섬현상의 정의”

: 도심지역이 외곽지역과 3~4도 차이나는 현상



그러나 구별로 그 현상의 심각성을 알아보기 위해서는 새로운 기준이 필요.

“새롭게 정의한 열섬현상”

: 서울시 야간온도의 평균보다 0.75가 높은 현상

서울 내 열섬강도(도시화·비도시화 온도 비교)

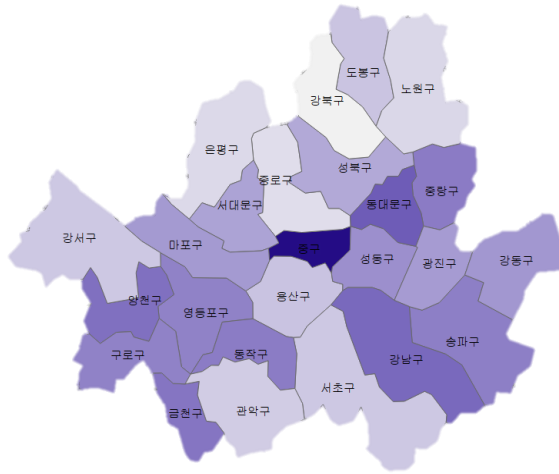
- 구별로 비교하기 위하여 열섬현상이 더 심해지는 밤 시간대의 온도를 사용.
- 이 기준을 바탕으로 열섬현상이 발생한 횟수를 구별, 월별로 합산하고 여름철, 겨울철만 따로 평균으로 비교함.

< 결과 >

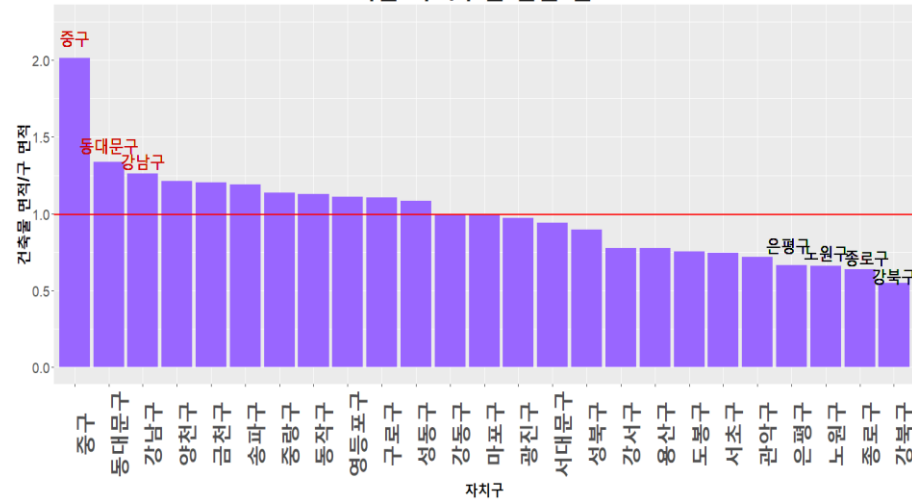
- 구별차이가 뚜렷한 편이며, 그 중에서도 송파구, 광진구는 겨울과 여름 모두 열섬현상 발생횟수가 높음.
- 서초구, 금천구, 강동구의 경우 겨울에는 열섬현상이 높게 나타나지만 여름에는 굉장히 드물게 나타남.
- 광진구, 용산구, 노원구, 종로구는 여름에는 열섬현상이 빈번히 일어나지만 겨울에는 적게 나타남.

2. 데이터 분석

구별 건물밀도



서울 자치구별 건물 밀도



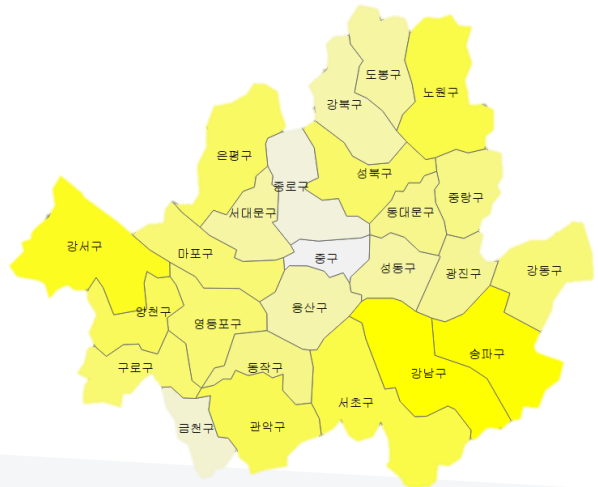
- 건물밀도가 높을수록 복사열 증가 및 바람길을 막는 원인이 되며 인공열을 증가 시킴.
- > 이것이 열섬현상과 밀접한 관계가 있다고 생각함.

2. 데이터 분석

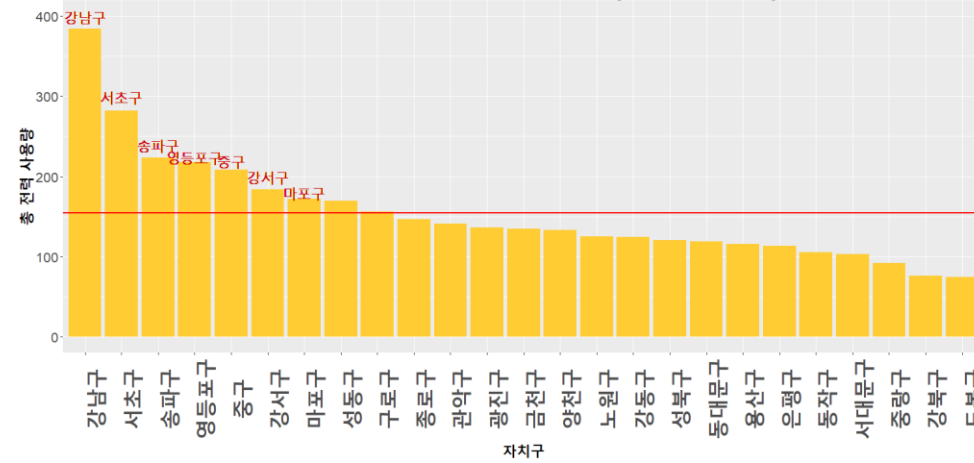
구별 전력 사용량



구별 가정용 전력 사용량

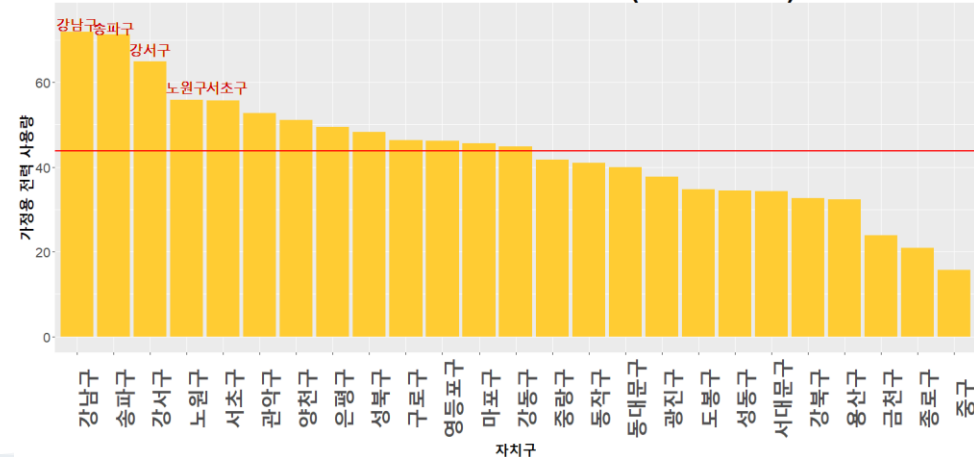


서울 자치구별 총 전력 사용량(단위 : GWh)



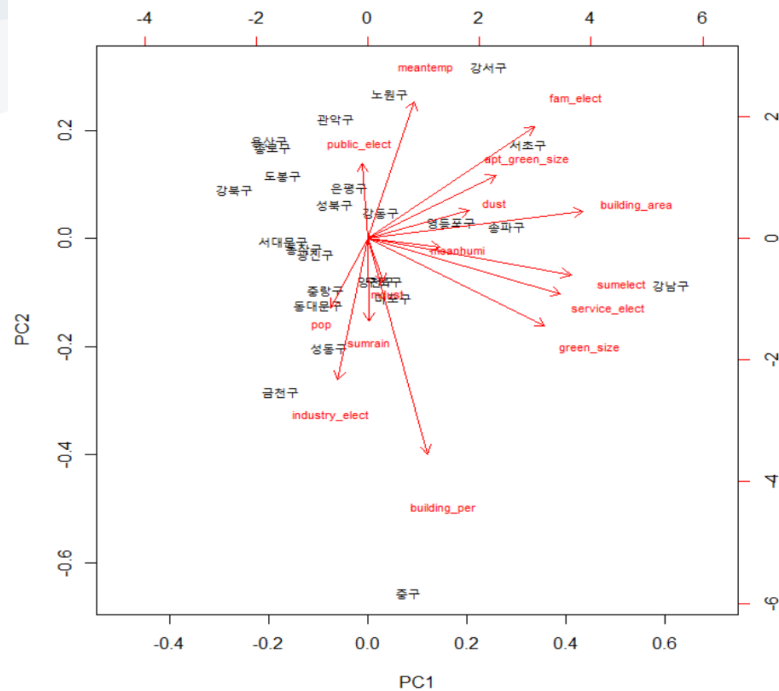
- 자치구별 총 전력 사용량은 강남구, 서초구, 송파구, 영등포구, 중구, 강서구, 마포구가 높음.
- 이들의 특징은 해당 구들의 녹지면적이 다른 구들에 비해 이미 높은 자치구들임.

서울 자치구별 가정용 전력 사용량(단위 : GWh)



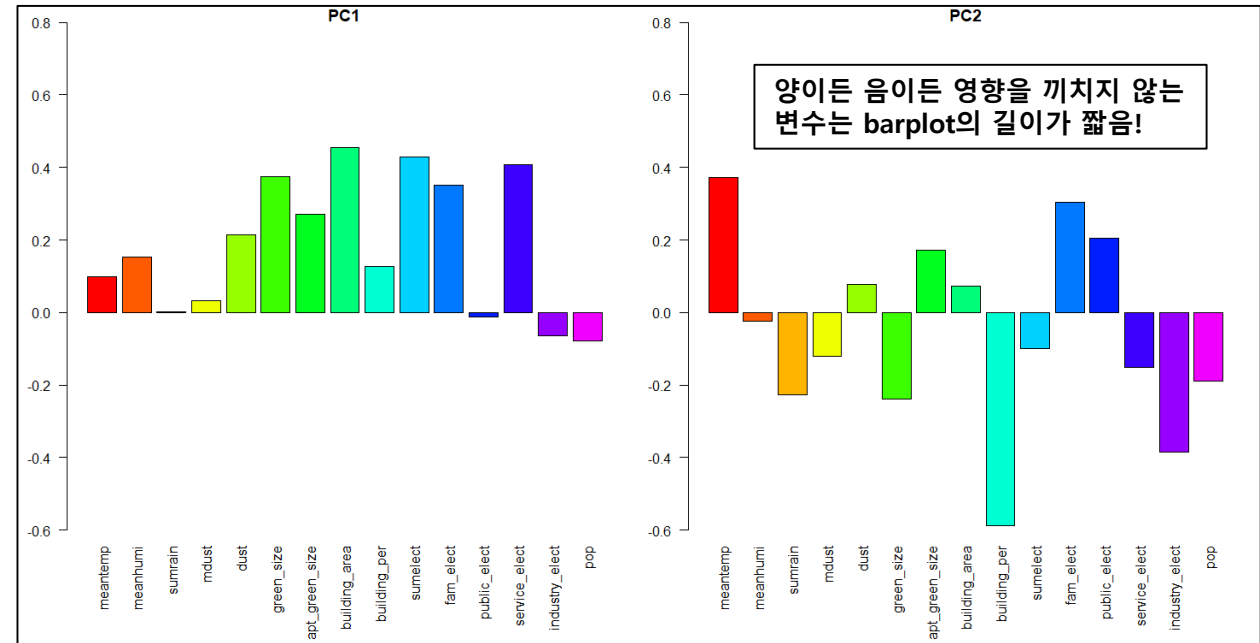
- 가정용 전력 사용량은 강남구, 송파구, 강서구, 노원구, 서초구가 평균적으로 높은 사용량을 보임.
- 총 전력 사용량과 비교해보면 강남구, 송파구, 강서구, 서초구는 총 전력 사용량에서 가정용 전력 사용량이 높은 것을 알 수 있음.

2. 데이터 분석 – 기타분석(PCA)



• PC1에 영향을 끼치는 변수는 Barplot으로 확인해보면 온도강수량, 초 미세먼지 건물 밀도, 공공, 산업 전력 사용량, 인구밀도의 변수를 제외한 나머지 변수들이 PC1에 영향을 끼치는 변수임. Biplot로 확인하면 PC1의 축과 수직이 되는 축은 PC1에 영향을 주지 않는 변수이며 축과 평행인 변수들은 PC1에 영향을 주는 변수임을 알 수 있으며 Barplot과 Biplot의 영향을 주는 변수에 대한 결과는 동일함

• PC2에 영향을 끼치는 변수로는 온도, 강수량, 초 미세먼지, 건물 밀도, 공공, 산업 전력 사용량, 인구밀도의 변수임. 마찬가지로 Biplot로 확인해보면 PC2에 영향을 주는 변수는 해당 축과 평행이 되는 벡터 변수들임. 따라서 Barplot으로 분석한 결과와 동일함을 알 수 있음



```
> summary(pca)
```

Importance of components:									
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Standard deviation	2.1208	1.5086	1.4698	1.11928	1.05558	0.98637	0.92815	0.80042	0.71554
Proportion of Variance	0.2999	0.1517	0.1440	0.08352	0.07428	0.06486	0.05743	0.04271	0.03413
Cumulative Proportion	0.2999	0.4516	0.5956	0.67913	0.75342	0.81828	0.87571	0.91842	0.95255

	PC10	PC11	PC12	PC13	PC14	PC15
Standard deviation	0.58804	0.49715	0.29889	0.16010	0.06175	6.946e-07
Proportion of Variance	0.02305	0.01648	0.00596	0.00171	0.00025	0.000e+00
Cumulative Proportion	0.97560	0.99208	0.99804	0.99975	1.00000	1.000e+00

- 주성분 분석을 통해 변동성(standard deviation)이 1보다 작고, 설명력(cumulative proportion)이 0.7~0.8인 차원까지 축소하여 PC6까지 선택
- 분석시에는 PC1과 PC2를 이용하여 PC별 구별 어떤 변수가 영향을 주는지 Biplot과 barplot을 통해 알아봄, Biplot 특성상 PC1과 PC2만을 이용하여 분석

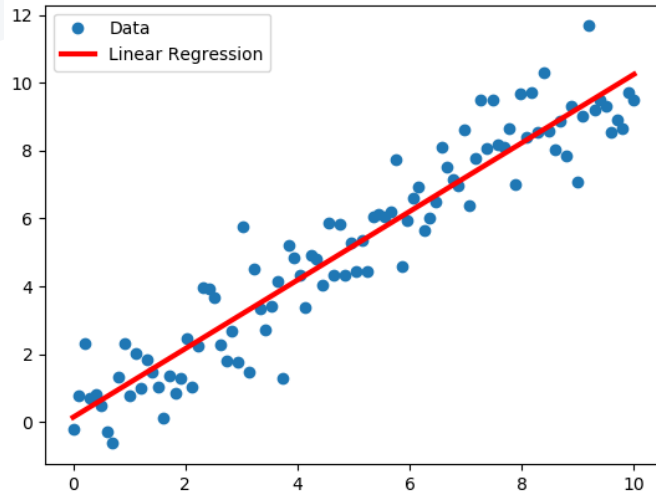
2. 데이터 분석 - 모델링



- 모델링을 할 때 변수 간 상관관계가 높으면 다중공선성의 문제가 발생하므로 이를 시각화로 한 번에 확인하고, 0.7이상인 변수를 뺌(평균온도, 건물면적 등)
- 파란색일수록 상관계수가 높고, 빨간색일수록 낮은 상관계수임.

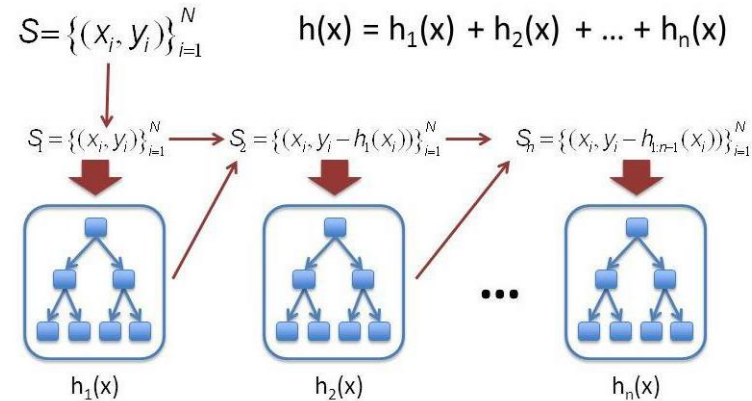
2. 데이터 분석 - 모델링

Linear regression



- 주어진 데이터를 대표하는 하나의 직선 즉, 회귀식을 찾는 시스템.

Xgboost



- 유연하고 end-to-end tree boosting 시스템.
- Tree boosting은 매우 효과적이고 머신러닝 방법에서 많이 사용됨.

Linear regression 을 사용했으나 유의성과 다중공선성을 띄는 변수들을 제거하여도 충분한 설명력을 갖지 못함.
설명력을 확보하기 위해 보다 효과적인 머신러닝 기법인 xgboost를 사용하였고 변수 중요도를 확인함.

트리 기반 모델의 경우 과적합이 될 가능성이 높아 학습 횟수를 높이고 학습률을 낮추었으며 5-fold 교차 검증을 실시함.
또한 객관적인 검증을 위해 train데이터와 test데이터를 7:3으로 나누었고 학습과 평가를 나누어 진행함.

2. 데이터 분석 - xgboost

목표 변수별 모델 정확도(설명력) R-squared(adjusted)

	Linear Regression	Xgboost ✓
초미세먼지	58%	78%
열섬(야간온도)	48%	75%
전력(가정)	51%	87%

변수별 모델 기여도 - 초미세먼지

순위	변수	해석
1	기온	<ul style="list-style-type: none"> 기온이 1위로 나타난 것은 초미세먼지와 마찬가지로 계절성이 있기 때문이라 추측, 이를 xgboost가 시간적 개념으로 사용한 것으로 보임. 3,4,5월은 미세먼지로 인한 피해가 심각한 달. 강수량 및 습도의 경우 비가 내릴 경우 대기 중의 먼지가 씻겨 내려가면서 미세먼지 농도가 감소했기 때문이라 추측 서비스전력 및 산업전력의 경우 해당 지수가 높을수록 생산활동이 활발하기 때문에 이에 영향을 받은 것으로 보임.
2	3,4,5월	
3	강수량 및 습도	
4	서비스전력 및 산업전력	
5	건물면적 및 아파트녹지면적	

2. 데이터 분석 – xgboost

변수별 모델 기여도 – 열섬(야간온도)

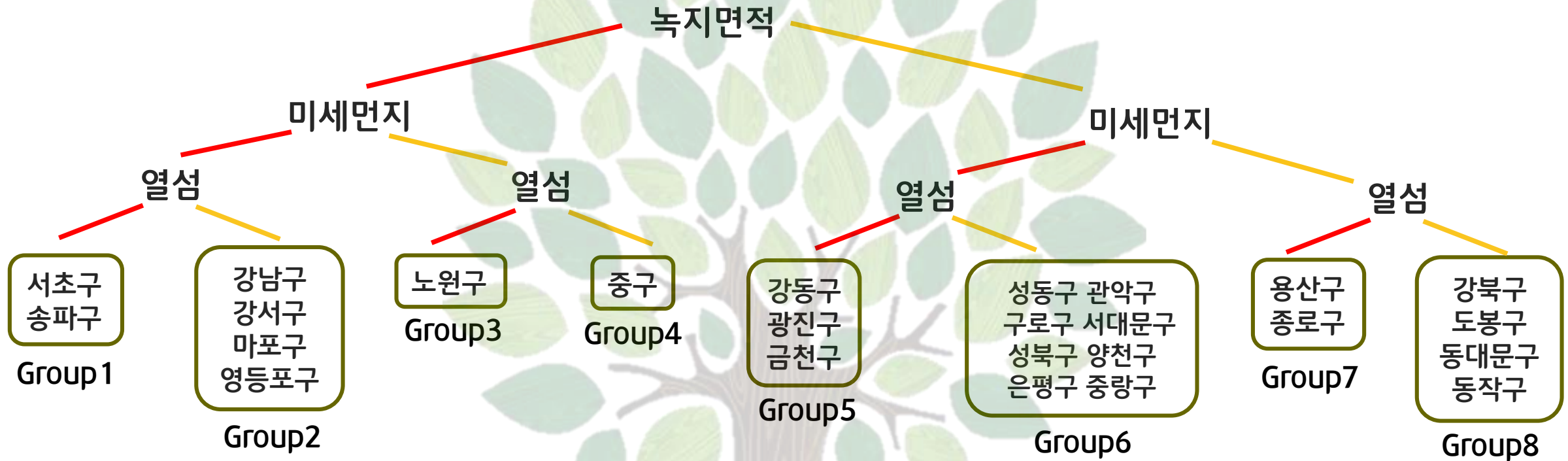
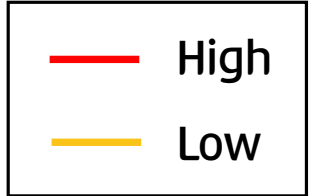
순위	변수	해석
1	미세먼지	<ul style="list-style-type: none"> 미세먼지가 열섬현상에 영향을 끼치는 것으로 보임. 실제로 연구 사례를 보면 미세먼지에 의해 온도가 보존되어 열섬현상이 일어난다고 설명. 전력의 경우 기온이 높을수록 가정에서 냉방사용량이 증가. 건물밀도의 경우 이는 도시화로 인한 열섬 현상 발생을 보여줌. 녹지비율의 경우 밀집한 도심 속 조성된 녹지와 일반 녹지가 열섬 현상 해소에 효과를 보임.
2	강수량 및 습도	
3	가정,서비스,공공 전력	
4	건물밀도	
5	건물 녹지비율 및 녹지 비율	

변수별 모델 기여도 – 가정전력

순위	변수	해석
1	건물밀도	<ul style="list-style-type: none"> 건물밀도가 높으면 전력사용량 합이 많을 수 밖에 없기 때문에 의미가 없다고 생각함. 최대한 구의 면적과 관련된 지표는 빼려고 노력하였으나 여전히 부족함 여름과 겨울의 경우를 보면 냉난방으로 인해 전력사용량이 높을 수 밖에 없다고 판단. 즉 평균온도와 관련이 있다고 판단.
2	평균온도 및 8월	


3. 분석 결과 활용 및 해결책 제시 - (1) 클러스터링


< 구별 특성에 따른 클러스터링 >



-> 각 그룹마다 해당 특성에 따라 보다 맞춤형 해결책 및 보완책 제시 가능!

3. 분석 결과 활용 및 해결책 제시 - (2) 구별 해결책 제시

1	2	3	4	5	6	7	8
서초구 송파구	강남구 강서구 마포구 영등포구	노원구	중구 	강동구 광진구 금천구	성동구 관악구 구로구 서대문구 성북구 양천구 은평구 종로구	용산구 종로구	강북구 도봉구 동대문구 동작구

1. 녹지면적 자체를 확대할 필요가 있는 구: Group 5,6,7 
-> 그 중에서도 금천구, 성동구, 구로구, 양천구는 건물녹지나 가로수 위주로 확대해야 함.
2. **녹지면적은 높으나 다른 대안이 필요한 구**: Group 1,2
3. 녹지면적은 낮으나 환경문제가 뚜렷하지 않은 구: Group 8
4. **중구**: 녹지가 다른 구보다 월등히 높음. 그만큼 환경문제가 없어서 녹지의 순기능을 제일 잘 보여주는 구임

3. 분석 결과 활용 및 해결책 제시 - (2) 구별 해결책 제시

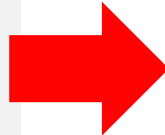
✓ Group 1: 서초구, 송파구

✓ Group 2: 강남구, 강서구, 마포구, 영등포구

- 분석 결과 다른 구보다 녹지면적이 넓음에도 불구하고, 미세먼지 및 열섬 현상이 두드러짐.
-> 녹지면적을 확대하는 것보다 다른 요인을 고려하여 환경문제를 해결하려는 접근이 필요



- 앞에서 구별 전력 총량을 비교해 봤을 때 위 두 그룹에 해당하는 모든 구가 전력사용량이 구별 평균 사용량보다 높았음
-> 전력 사용량이 높은 특징을 이용하면서도 미세먼지 및 열섬현상을 해결할 수 있는 대안이 필요



- 녹지면적을 늘리지 않고 높은 전력사용량, 미세먼지 발생, 열섬현상을 완화할 수 있는 대안이 필요함
-> “태양광 미니 발전소”를 확대 보급하는 것으로 접근!

• 태양광 미니발전소란?



서울시의 원전하나 줄이기 사업 중 하나로, 태양광 발전기를 가정의 베란다, 주택이나 건물의 옥상에 설치할 수 있는 크기로 제작한 것으로, 자투리 공간을 활용하여 에너지를 생산할 수 있음

3. 분석 결과 활용 및 해결책 제시 - 태양광 미니 발전소(참고)

• 환경문제 개선 측면에서의 기대효과

1) 250W 1가구 설치시 소나무 2그루 식재효과



2) 250W 16가구 설치시 경유차 1대당 연간
초미세먼지 저감효과

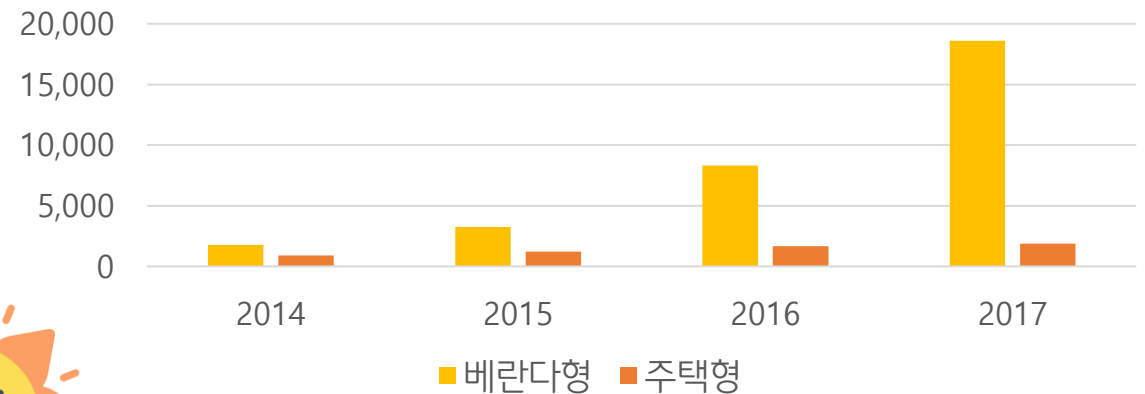


➡ 태양광에너지 보급 증진뿐만 아니라, 녹지부족이나 초미세먼지에 대한 또 다른 방안이 될 수 있음.

• 개선되어야 할 점

- 2022년까지 서울시의 태양광 미니 발전소 보급 목표는 100만 가구(베란다형 63만, 주택형 15만, 건물형 22만)
- 17년까지의 누적 보급 현황 중 건물형이 설치된 개수는 23개에 불과 -> 집중적으로 확대보급 필요

서울시 태양광 미니발전소 보급 현황



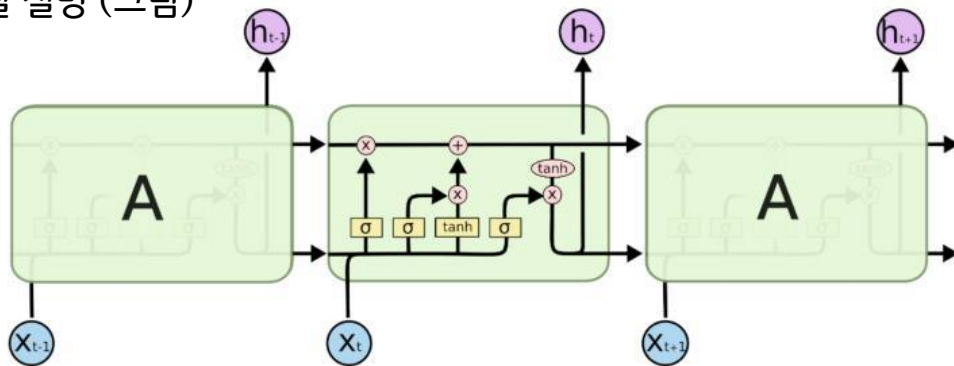
- 기존 서울시 태양광 미니 발전소 설치 지원은 선착순이었음
- 기존 선착순 보급 방식을 바꿔 Group 1,2에 해당하는 구에 우선순위를 두고 '건물형' 설치를 유도하기 위해 지원금을 늘리는 방향으로 정책을 수정하여 시행하면 더 효과적일 것임. + 습도와 강수량 등을 사용하여 **입지 선정**에 활용.
- '초미세먼지 20%감축'과 '원전 하나 줄이기 정책'을 동시에 좀 더 효율적으로 시행할 수 있음.

3. 분석 결과 활용 및 해결책 제시 - (3) 환경 문제 예측 모델 제안

해결 방안 중 하나로 각 문제 현상 별 중장기 예측 모델 개발 제안 (일 단위가 아닌 월 단위)

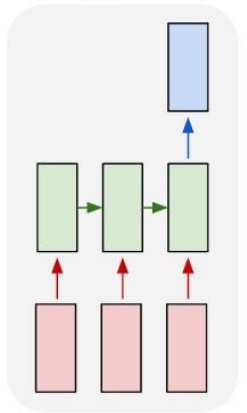
데이터의 특성 파악: 변수 중 대부분이 시계열 자료 (계절성)

모델 설명 (그림)



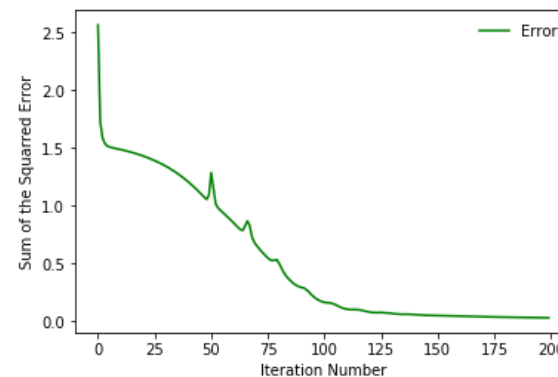
메모리 셀이 있어서 시계열 자료에 강하다! + 교차상관을 반영할 수 있음

many to one

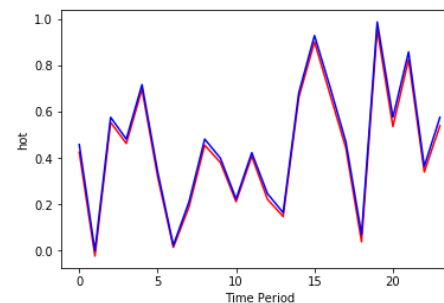


- 여러 가지 변수로 한 변수를 예측하는 과정이므로 Many to One 형태의 LSTM 모델 선택!
- 예시에서는 12 to 1 형태로 학습.
총 학습 기간은 2014.1 ~ 2016.12
검증 기간은 2017년도 로 사용

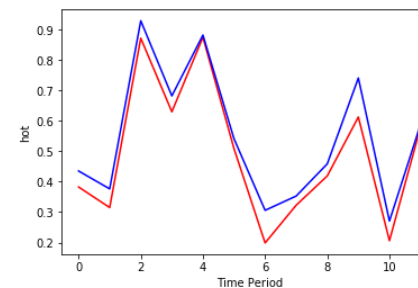
• 열섬 현상 예측 모델 예시 (LSTM)



Epoch당 loss 변화



트레인 데이터 학습 결과










테스트 데이터 검증 결과

-> 예측 결과를 통한 사전 대비 및 계획 수립 가능

4. 부록 – (1) 한계점

- 산업용, 공공용, 서비스용 전력 사용량은 공간상의 제약으로 실지 못함
- 옥상 녹화 사업의 데이터는 2011년도 까지의 자료밖에 존재하지 않음 (최근의 데이터가 없음)
- 태양광을 설치하는데 있어서 일조량, 일사량 데이터가 중요한 입지 기준으로 작용하는데 "기상관측정보" 데이터의 일조량, 일사량 결측치가 160만개가 있어 사용할 수 없었음 그래서 다른 요인들인 전력사용량, 습도, 강수량 등으로 확인 했음
- 태양광 미니 발전소가 설치된 구체적인(적어도 구단위까지) 장소가 명시된 데이터가 존재하지 않음
- 열섬 완화 방안으로 쿨루프 정책(COOL ROOF)도 살펴보았지만 어느 위치에 쿨루프 정책이 시행 되었는지 확인할 수가 없어서 대안으로 활용할 수 없었음
- 공원 및 사유지 위치 정보를 통해 구별로 나무의 밀집도를 분석하여 더 나은 정책 제안을 하고 싶었으나 데이터가 2012년까지 밖에 업데이트되지 않아 현재에 맞춰 분석하기 어려웠음

4. 부록 - (2) 데이터 목록

데이터	출처	사용변수(파생변수)
 기상관측정보(기온, 강수 등)	서울시 빅데이터캠퍼스 반출	야간기온, (열섬)
 대기환경정보	서울시 빅데이터캠퍼스 반출	구별 월별 미세먼지
 서울시 인구밀도(구별) 통계	서울열린데이터광장	인구밀도
 서울시 녹지현황 통계	서울열린데이터광장	구별 녹지면적, 건물 녹지 면적
 서울시 실시간 자치구별 대기환경 현황	서울열린데이터광장	월별 초미세먼지
 서울시 건축물 현황 통계	서울열린데이터광장	건물면적, (건물면적/구면적)
 서울시 전력사용량(용도별) 통계	서울열린데이터광장	전력사용량 총합, 월별 가정·공공·서비스·산업 전력사용량

4. 부록 - (3) 참고문헌 및 분석 툴

참고문헌

- 서울에너지공사 - 태양광 미니발전소 QnA
- 환경부 - 미세먼지, 도대체 뭘까(환경부 소책자)
- 도시열섬 지역에 대한 정의 및 구분 방법론에 관한 비교연구 - 김기중 (Kim Kijung), 안영수 (An Youngsoo),
발행기관 : 한국지역학회
- 단지 유형에 따른 열섬현상 저감 방안에 관한 연구 - 정주리(丁周利)

분석 툴



Excel - 전처리



R STUDIO - 전처리 및 시각화 + 모델링 (lm, xgboost)



Tableau - 데이터 시각화



Python - 신경망 모델 제작 (LSTM)



구글트렌드 - 검색빈도 파악



소셜 매트릭스 - 검색어 연관도 파악

감사합니다