
GS! 미세먼지가 많은 날 고객을 부탁해!

TEAM 먼지가먼지 : 김미현, 박지아, 우나영

목차

1

공모배경

2

전처리/EDA

3

분석

4

비즈니스
아이디어

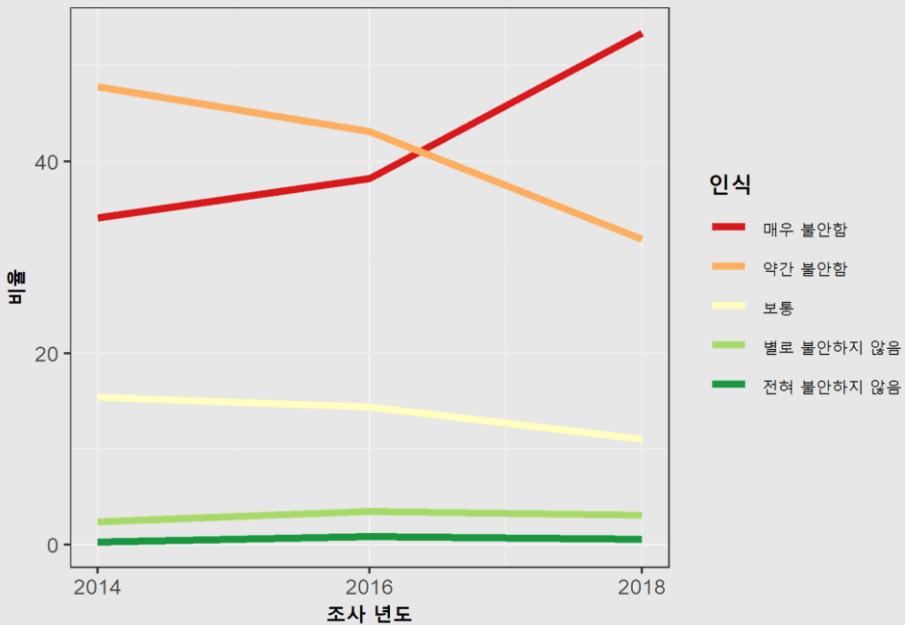
1

공모배경

1ST 공모배경

1. 서울시민의 미세먼지에 대한 인식

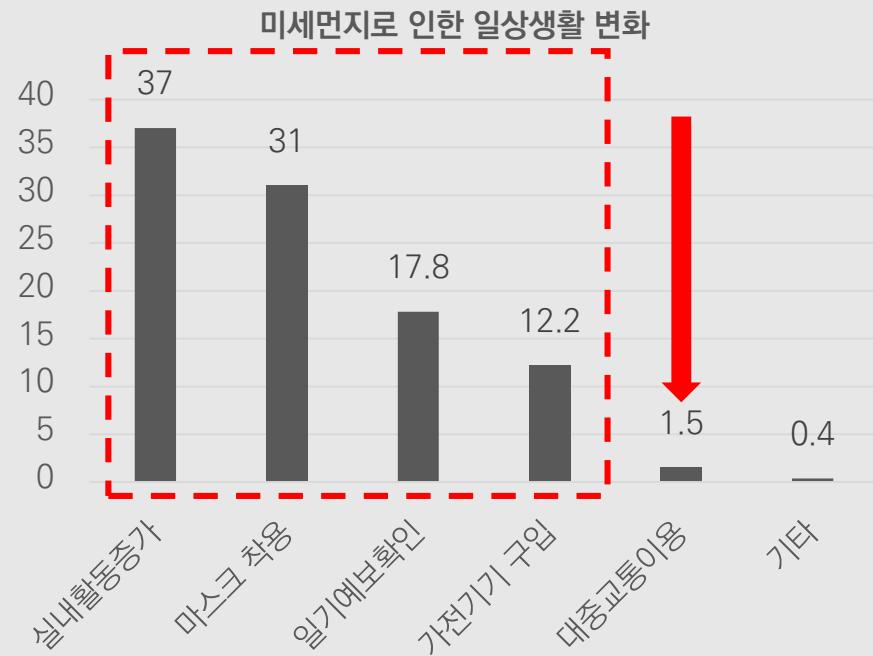
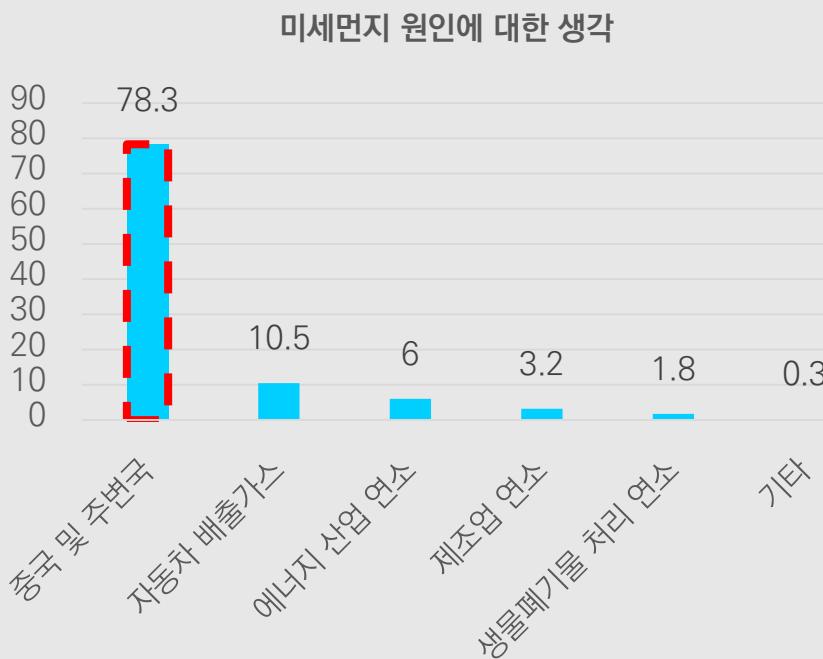
서울시민의 미세먼지에 대한 인식변화



서울시민의 미세먼지 인식: 최근 2-3년 들어 불안감 가중

1ST 공모배경

2. 미세먼지에 대한 대응



그러나 미세먼지 원인이 **외부**에 있다고 생각
따라서, 사람들은 **미세먼지 불안감을 느낌에도 불구하고**
미세먼지 경감을 위한 근본적인 노력보다 방어적인 해결책 선택

1ST 공모배경

3. 공모전 주제 : 미세먼지의 사회적 영향 분석 및 비즈니스 아이디어

출제문제

“미세먼지의 사회적 영향 분석 및 비즈니스 아이디어 제시”



빅데이터를 활용한 “미세먼지의 사회적 영향 분석 및 비즈니스 아이디어 제시”

- 유동인구데이터(SK텔레콤), 카드매출데이터(신한카드), SNS데이터(와이즈넛),
환경기상데이터(케이웨더), 유통데이터(GS리테일), 공공데이터 등 다양한
데이터를 활용하여 미세먼지로 인한 소비/경제/행동변화에 따른
사회적 영향 분석 및 예측 모델링을 통한 비즈니스 아이디어 제시



1ST 공모배경

3. 공모전 주제 : 미세먼지의 사회적 영향 분석 및 비즈니스 아이디어

???



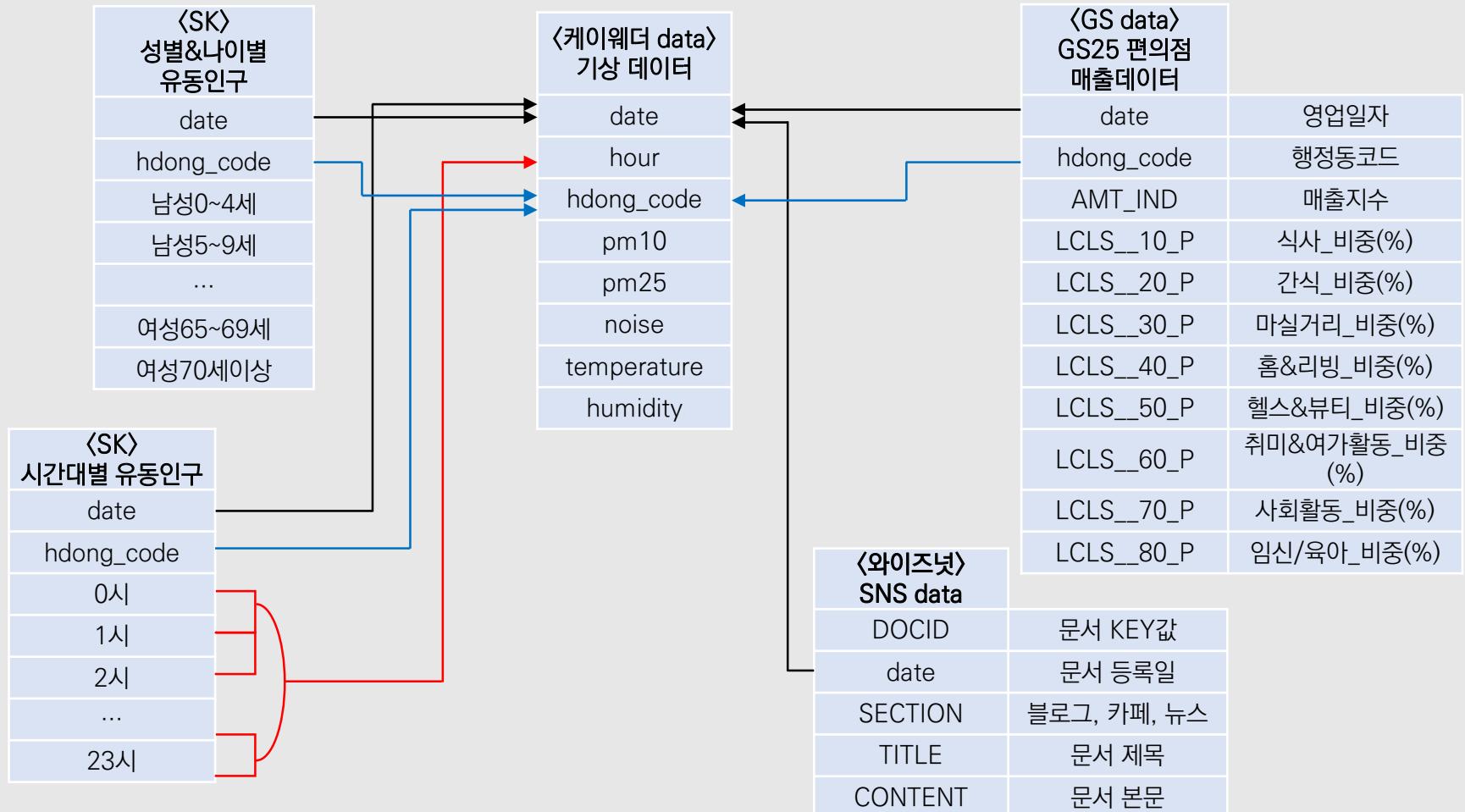
그렇다면, 미세먼지로 인해 사람들의 행동양식은 어떻게 **변화했는가?**

2

전처리 / EDA

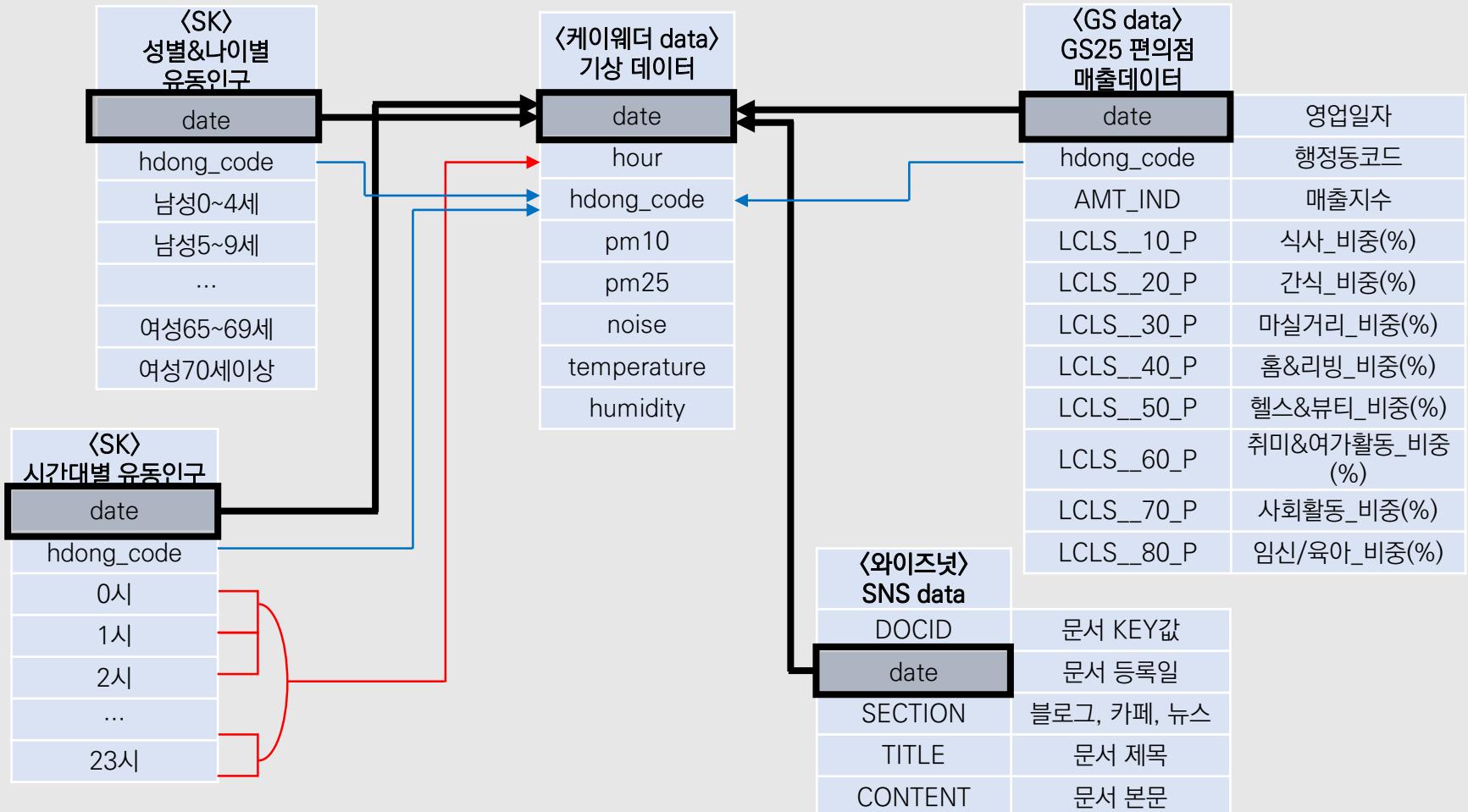
2ND 데이터 전처리/EDA

1.1 데이터 소개



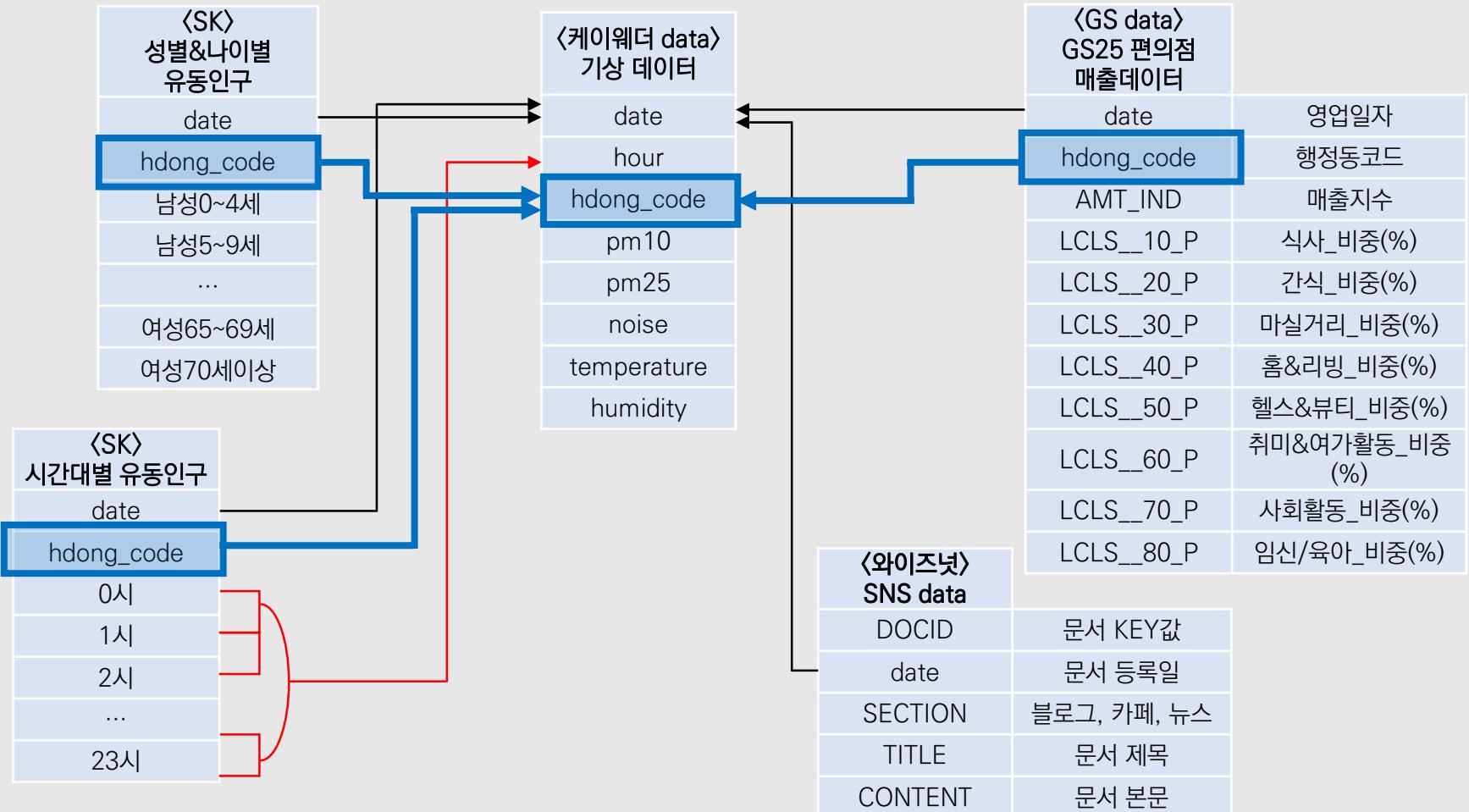
2ND 데이터 전처리/EDA

1.1 데이터 소개



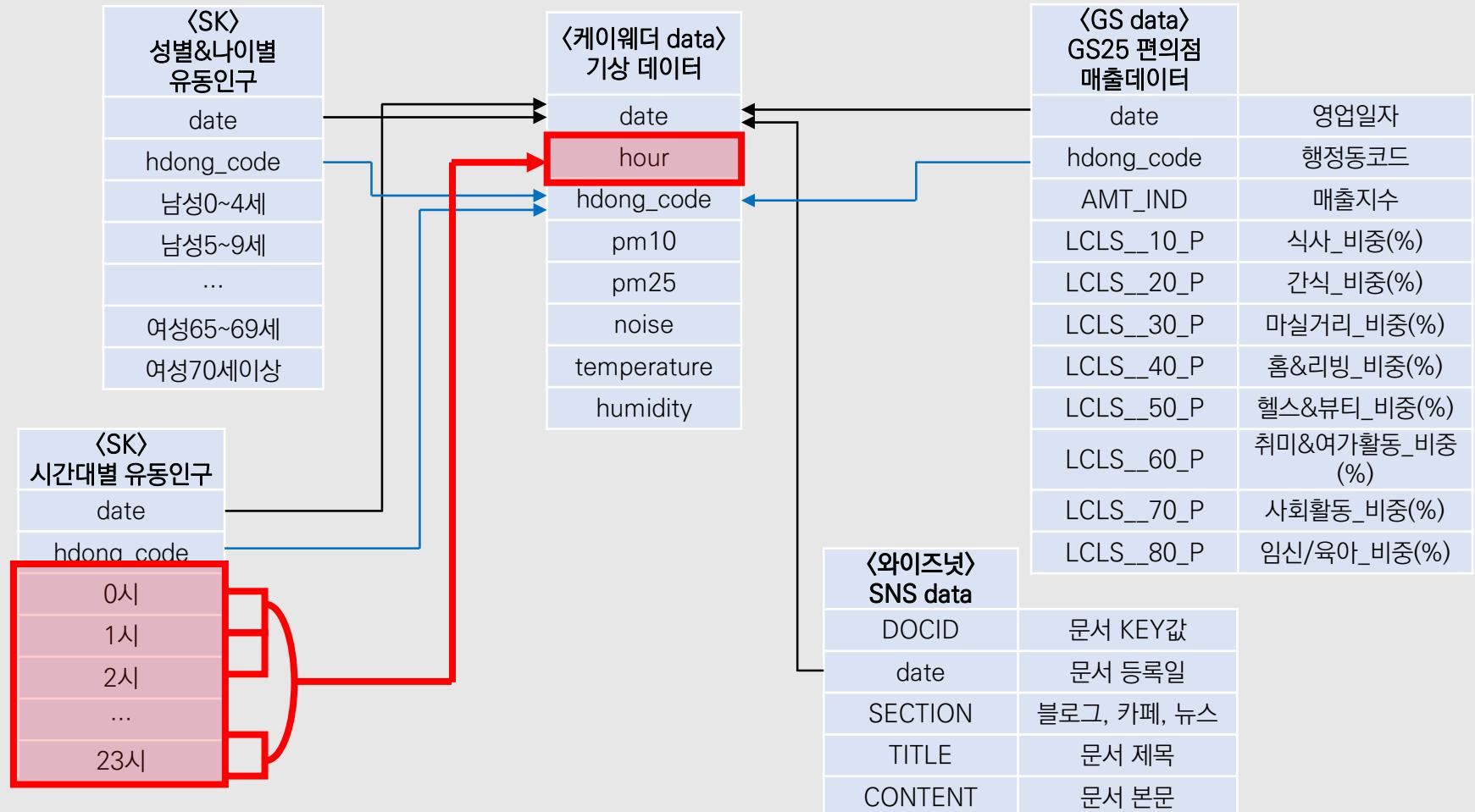
2ND 데이터 전처리/EDA

1.1 데이터 소개



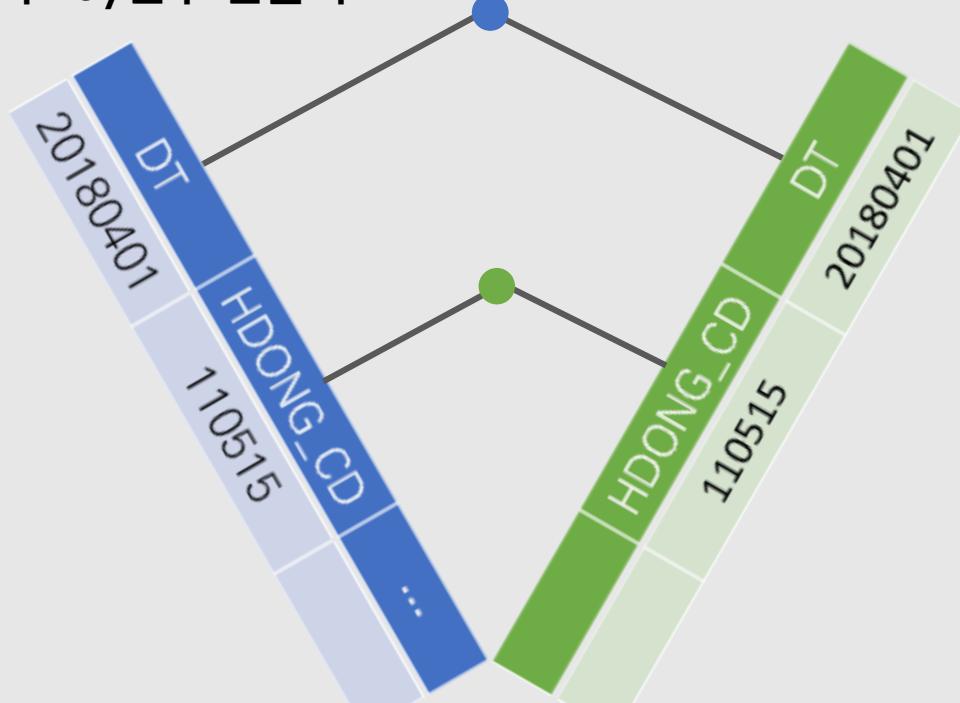
2ND 데이터 전처리/EDA

1.1 데이터 소개



2ND 데이터 전처리/EDA

1.2 전체 데이터 key 변수 만들기



각각의 데이터에서 동일한 특성을 가진 칼럼은 key값으로 merge하기 위해 변수명 통일

날짜는 DT

구 코드와 행정동 코드의 경우, 결합해 6자리의 HDONG_CD 생성



케이웨더 환경기상데이터

2ND 데이터 전처리/EDA

2.1 케이웨더 환경기상데이터 - 원본

제공받은 케이웨더 환경기상데이터 원본 소개:
종로구, 노원구쪽 53개 스테이션의 초별 관측 데이터

종로구	스테이션	측정기 등록일	위치	노원구	스테이션	측정기 등록일	위치
31지점	V10O1610252	2017-12-26	교남동	22지점	V10O1610610	#	하계동
	V10O1610546	2018-03-09	세종로 100		V10O1610376	2017-12-27	상계동
	V10O1610540	2018-03-14	세종로 100		V10O1610293	2017-12-27	상계동
	V10O1610542	2018-03-14	세종로 100		V10O1610356	2017-12-27	상계동
	V10O1610543	2018-03-09	세종로 100		V10O1610616	2017-12-23	상계2동

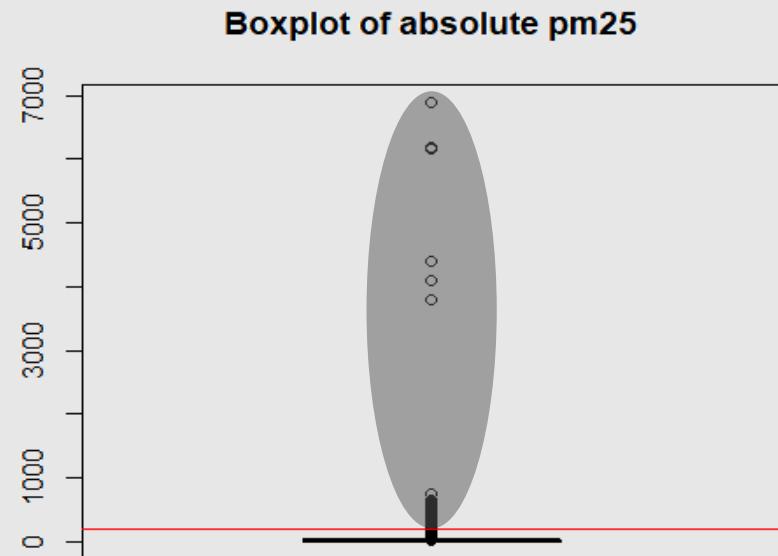
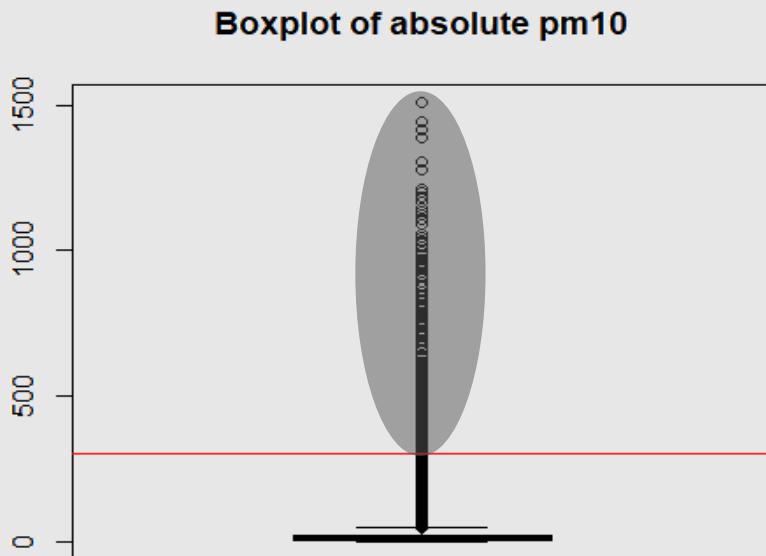
	V10O1611698	#	원서동		V10O1611652	2018-03-28	공릉동
	V10O1611722	2018-03-29	가회동		V10O1611150	2018-03-25	상계동
	V10O1611645	2018-03-28	명륜1가 72-3		V10O1611229	2018-03-25	월간

V10O1611229

serial	time	pm10	pm25	noise	temp	humid
V10O1611229	201804010000	36	18	50	14.4	51
V10O1611229	201804010001	40	20	50	14.4	51
V10O1611229	201804010002	53	27	50	14.4	51
V10O1611229	201804010003	49	24	50	14.4	51
V10O1611229	201804010004	33	16	50	14.4	51
V10O1611229	201804010005	52	26	50	14.4	52
V10O1611229	201804010006	46	23	50	14.4	52
V10O1611229	201804010007	38	19	50	14.4	52
V10O1611229	201804010008	28	14	50	14.4	52

2ND 데이터 전처리/EDA

2.2 케이웨더 환경기상데이터 - 이상치 처리



미세먼지 데이터는 시계열 데이터
시간의 흐름에 따른 차이(difference)를 계산
Difference 절대치의 quantile 값이 상위 0.005% 이상인 값을
이상치로 진단하고 NA처리

2ND 데이터 전처리/EDA

2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기

종로구	스테이션	측정기 등록일	위치	행정동
31지점	V1001610252	2017-12-26	교남동 46	교남동
	V1001610546	2018-03-09	세종로 100	종로 1,2,3,4가동
	V1001610540	2018-03-14	세종로 100	종로 1,2,3,4가동
	V1001610542	2018-03-14	세종로 100	종로 1,2,3,4가동
	V1001610543	2018-03-09	세종로 100	종로 1,2,3,4가동
	V1001610544	2018-03-09	세종로 100	종로 1,2,3,4가동
	V1001610545	2018-03-14	세종로 100	종로 1,2,3,4가동
	V1001610567	#	세종로 100	종로 1,2,3,4가동
	V01o1610468	2018-01-04	도렴동 (광화문역 7번 출구)	사직동
	V1001611289	2018-03-25	영천동 277	천연동
	V1001611172	2018-03-25	내자동 111-1	사직동
	V1001611634	2018-03-28	재동 95-2	가회동
	V1001611887	2018-03-30	태평로1가 68-2	명동
	V1001611639	2018-03-28	종로5가 20-2	종로5,6가동
	V1001611658	2018-03-28	연건동 187-1	이화동

한 행정동 내에 여러 스테이션이 있을 경우,
시간별 평균을 낸 데이터를
행정동의 기상데이터로 삼는다.
만약 스테이션이 없는 행정동의 경우,
다음 슬라이드와 같이 처리한다.

2ND 데이터 전처리/EDA

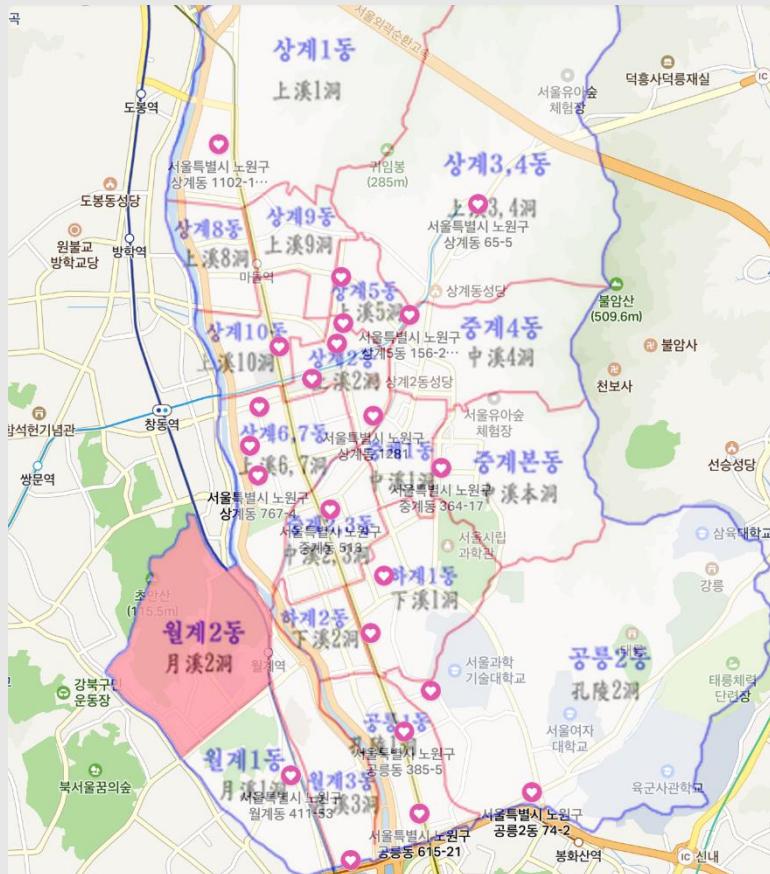
2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



1) 지도에 스테이션을 ❤️로 표시한다.

2ND 데이터 전처리/EDA

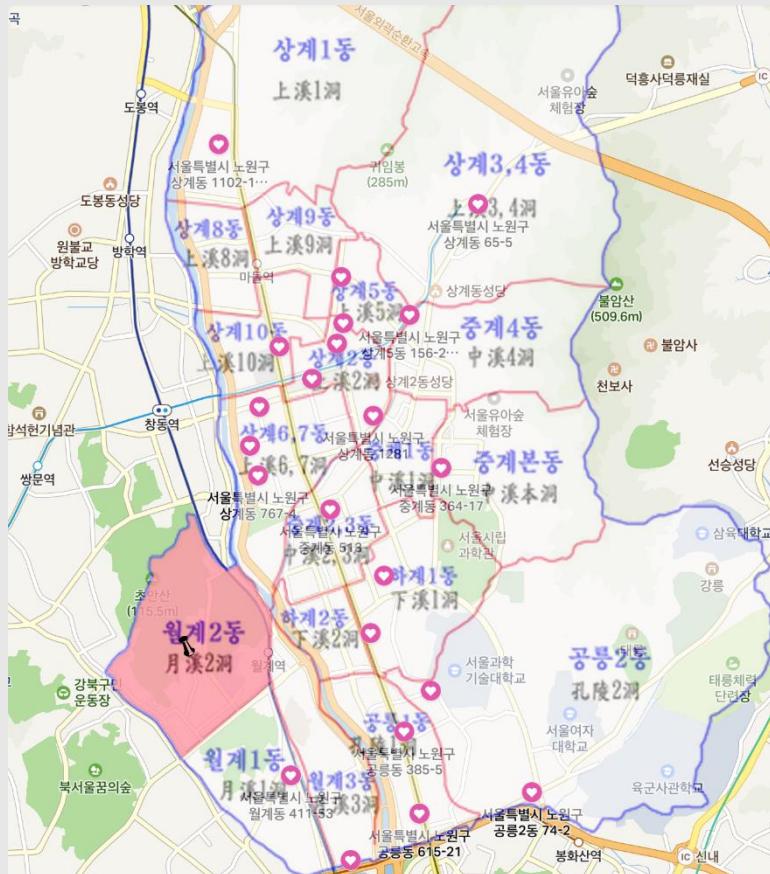
2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



2) 스테이션이 행정동 안에 없는 곳을 확인한다. Ex: 월계2동

2ND 데이터 전처리/EDA

2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



3) 해당 동 주민센터 위치를 확인한다

2ND 데이터 전처리/EDA

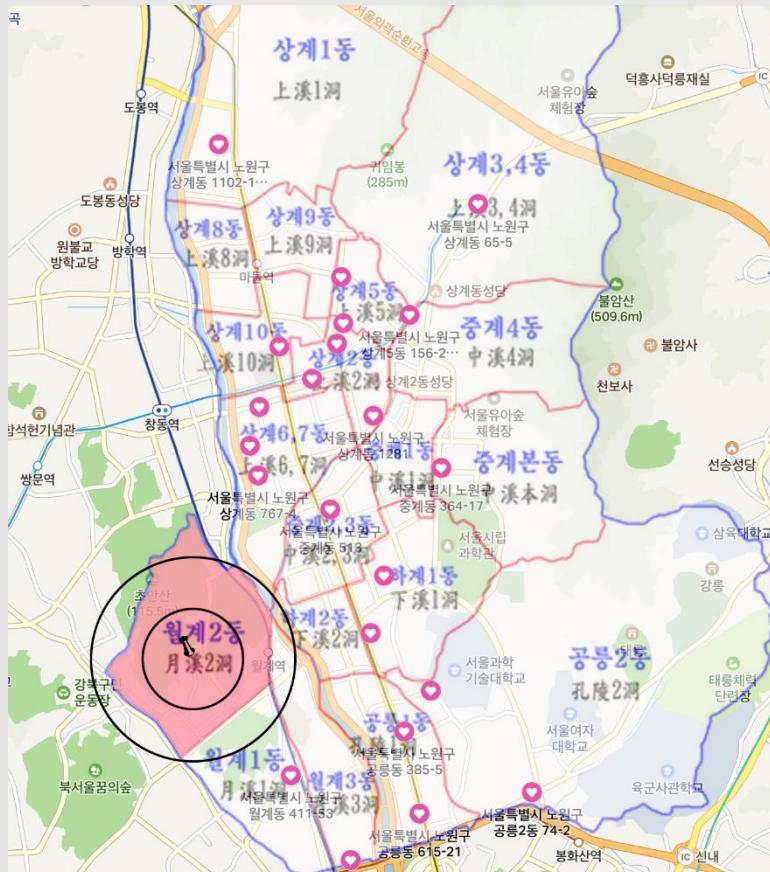
2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



4) 동 주민센터로부터 반경을 0.5km씩 늘려가며 인근 스테이션이 반경에 들어올 때까지, 주민센터와 스테이션 간 거리를 잰다.

2ND 데이터 전처리/EDA

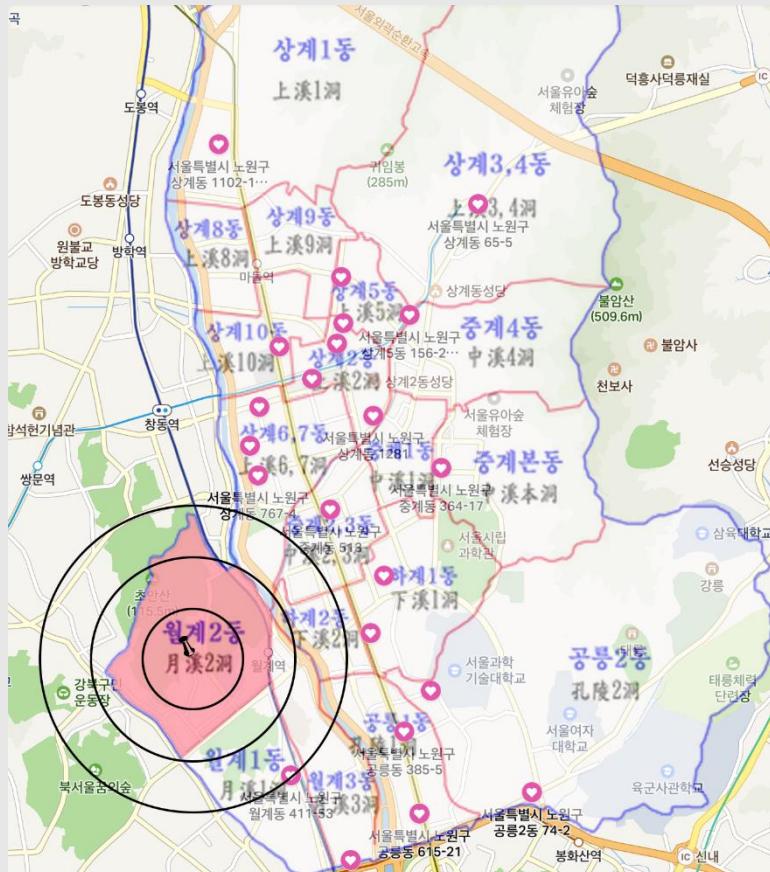
2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



4) 동 주민센터로부터 반경을 0.5km씩 늘려가며 인근 스테이션이 반경에 들어올 때까지, 주민센터와 스테이션 간 거리를 잰다.

2ND 데이터 전처리/EDA

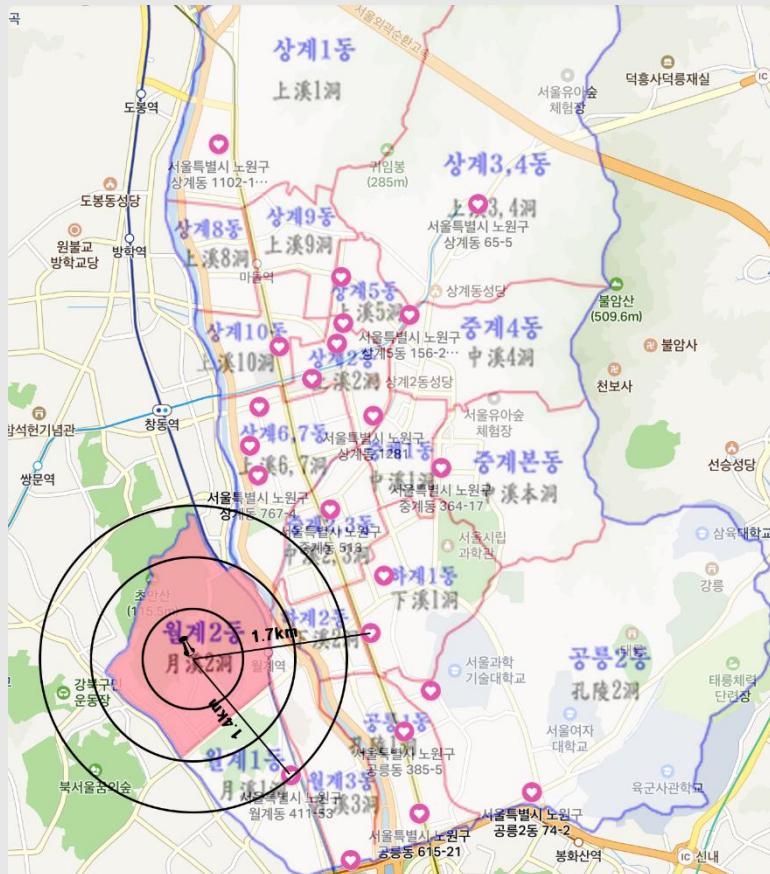
2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



4) 동 주민센터로부터 반경을 0.5km씩 늘려가며 인근 스테이션이 반경에 들어올 때까지, 주민센터와 스테이션 간 거리를 잰다.

2ND 데이터 전처리/EDA

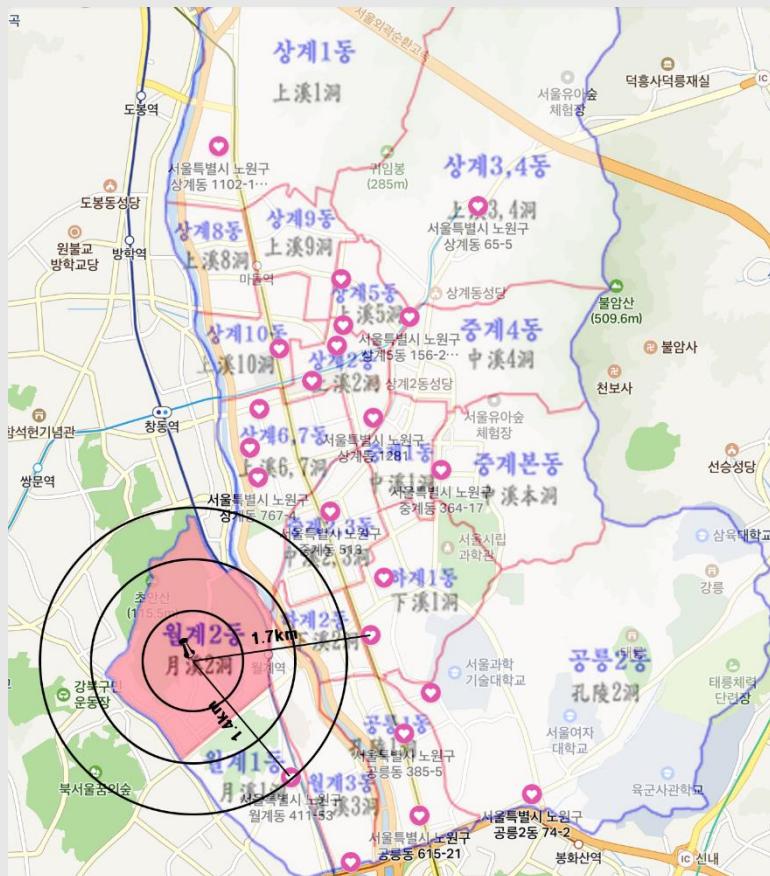
2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



4) 동 주민센터로부터 반경을 0.5km씩 늘려가며 인근 스테이션이 반경에 들어올 때까지, 주민센터와 스테이션 간 거리를 잰다.

2ND 데이터 전처리/EDA

2.3 케이웨더 환경기상데이터 - 행정동별 기상 데이터 만들기



5) 인근 스테이션이 0.5km씩 늘린 반경에 들어오면 해당 스테이션의 데이터를 사용한다.

기상 데이터가 missing 이 많은 경우의 행정동 역시 같은 방법을 사용하여, Missing hour 가 365일*24시간 = 8760시간 중 10시간 미만이 될 때까지 인근 스테이션 데이터를 끌어와 사용한다.

2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분



예보데이터	관측데이터
 The logo for Airkorea, featuring the word "Airkorea" in a stylized blue and green font. Above the text, there is Korean text: "한국실시간 대기오염도".	Total Weather Service Provider  The logo for Kweather, featuring the word "Kweather" in a bold, black, sans-serif font with a red sun-like graphic integrated into the letter "W".

2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분

일별 예보데이터의 경우

pm10, pm2.5 모두 2미만인 경우를 미세먼지 좋은날로,
Pm10, pm2.5 모두 2초과인 경우를 미세먼지 나쁜날로 지정

미세먼지
(pm10)

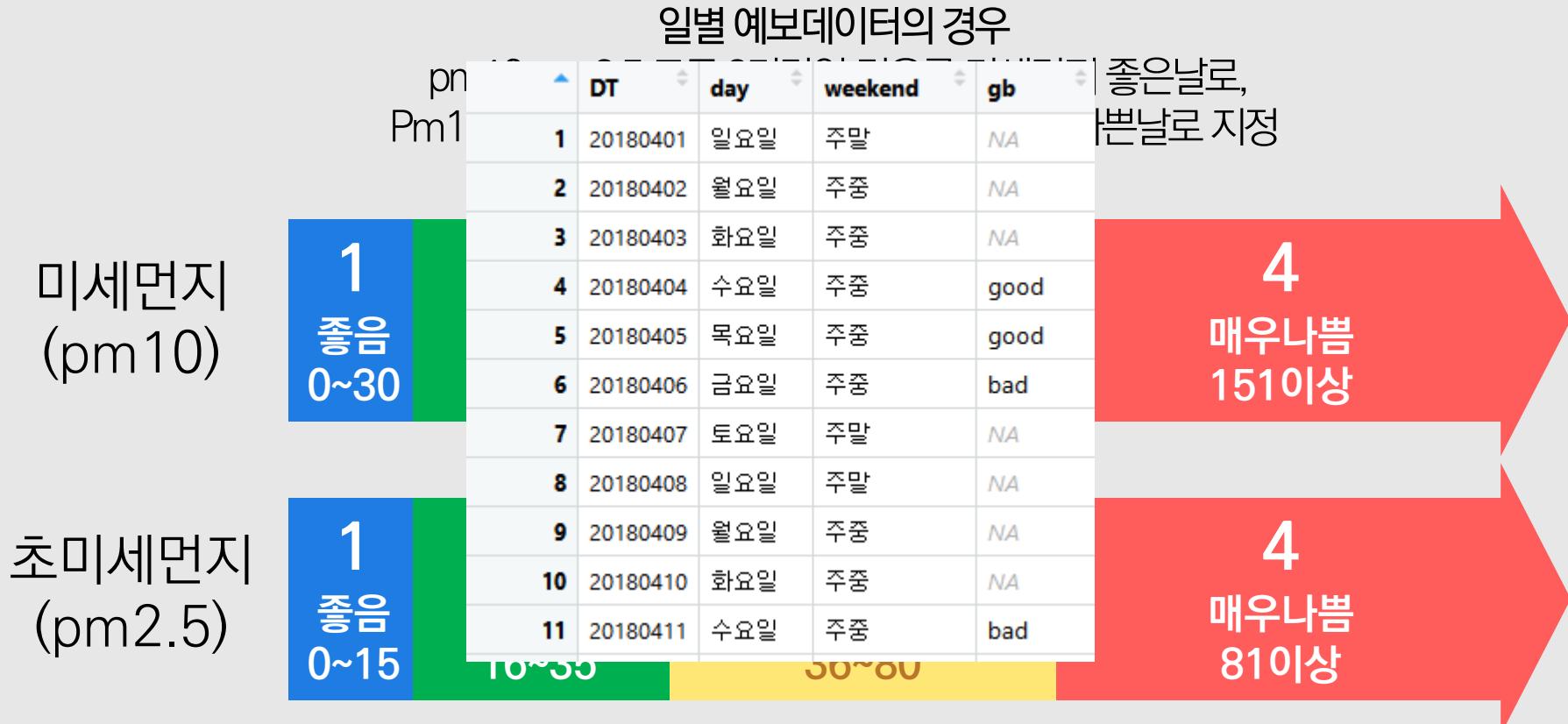


초미세먼지
(pm2.5)



2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분



2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분

예보데이터를 통한 좋은날/나쁜날 구분은
미세먼지 (pm10) 추후 SNS데이터 분석에 사용

초미세먼지
(pm2.5)

1
좋음
0~15



2	20180401	일요일	주말	NA
3	20180402	월요일	주중	NA
4	20180403	화요일	주중	NA
5	20180404	수요일	주중	NA
6	20180405	목요일	주중	good
7	20180406	금요일	주중	bad
8	20180407	토요일	주말	NA
9	20180408	일요일	주말	NA
10	20180409	월요일	주중	NA
11	20180410	화요일	주중	NA
12	20180411	수요일	주중	bad

4
매우나쁨
81이상

일별 예보데이터의 경우

좋은날로,
나쁜날로 지정

151이상

2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분

관측데이터의 경우

행정동별 시간별 미세먼지 수치에 따라 예보기준으로 1,2,3,4의 라벨링 후
일평균 pm10, pm2.5 모두 2미만인 경우를 미세먼지 좋은날로,
일평균 pm10, pm2.5 모두 2초과인 경우를 미세먼지 나쁜날로 지정

미세먼지
(pm10)



초미세먼지
(pm2.5)



2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분

기준데이터인 경우						
행정동	▲ DT	▼ day	weekend	HDONG_CD	gb	후
미세먼지 (pm10)	1	20180401	일요일	주말	110515	kbard
	2	20180402	월요일	주중	110515	NA
	3	20180403	화요일	주중	110515	kbard
	4	20180404	수요일	주중	110515	kgood
	5	20180405	목요일	주중	110515	kgood
	6	20180406	금요일	주중	110515	NA
	7	20180407	토요일	주말	110515	kgood
	8	20180408	일요일	주말	110515	kgood
	9	20180409	월요일	주중	110515	NA
	10	20180410	화요일	주중	110515	kgood
	11	20180411	수요일	주중	110515	kbard
초미세먼지 (pm2.5)	1	20180401	일요일	주말	110515	나쁨 이상
	2	20180402	월요일	주중	110515	나쁨 이상
	3	20180403	화요일	주중	110515	나쁨 이상
	4	20180404	수요일	주중	110515	나쁨 이상
	5	20180405	목요일	주중	110515	나쁨 이상
	6	20180406	금요일	주중	110515	나쁨 이상
	7	20180407	토요일	주말	110515	나쁨 이상
	8	20180408	일요일	주말	110515	나쁨 이상
	9	20180409	월요일	주중	110515	나쁨 이상
	10	20180410	화요일	주중	110515	나쁨 이상
	11	20180411	수요일	주중	110515	나쁨 이상

2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분

관측데이터를 통한 좋은날/나쁜날 구분은
주후 GS유통과 유동인구데이터 분석에 사용

초미세먼지
(pm2.5)

행정동	DT	day	weekend	HDONG_CD	gb	후
일	1 20180401	일요일	주말	110515	kbad	
0~	2 20180402	월요일	주중	110515	NA	
1	3 20180403	화요일	주중	110515	kgood	나쁨
2	4 20180404	수요일	주중	110515	NA	나쁨
3	5 20180405	목요일	주중	110515	kbad	나쁨
4	6 20180406	금요일	주중	110515	NA	나쁨
5	7 20180407	토요일	주말	110515	kgood	나쁨
6	8 20180408	일요일	주말	110515	kgood	나쁨
7	9 20180409	월요일	주중	110515	NA	나쁨
8	10 20180410	화요일	주중	110515	kgood	나쁨
9	11 20180411	수요일	주중	110515	kbad	나쁨

4
나쁨
이상

2ND 데이터 전처리/EDA

2.4 케이웨더 환경기상데이터 - 미세먼지 좋은 날과 나쁜 날 구분

예보 데이터로 본 미세먼지 좋은날 & 나쁜날

→ 서울 전역 365일 중 138일 & 45일

관측데이터로 본 미세먼지 좋은날 & 나쁜날

→ 종로구 노원구 평균 365일 중 125일 & 125일

각 행정동별 좋은날 & 나쁜날 일수

	J가회동	J교남동	J무악동	J부암동	J사직동	J삼청동	J승인1동	J승인2동	J이화동	J종로 1,2,3,4가동	J종로 5,6가동	J창신1동	J창신2동	J창신3동	J청운효자동	J평창동	J혜화동
좋은날	115	105	129	127	82	122	154	136	145	103	122	120	138	154	122	131	149
나쁜날	129	139	131	120	153	127	98	107	96	129	151	144	96	98	151	107	94

	N공릉1동	N공릉2동	N상계10동	N상계1동	N상계2동	N상계3,4동	N상계5동	N상계6,7동	N상계8동	N상계9동	N월계1동	N월계2동	N월계3동	N중계1동	N중계2,3동	N중계4동	N중계본동	N하계1동	N하계2동
좋은날	92	87	101	98	112	157	151	128	104	152	170	170	98	148	156	119	148	139	106
나쁜날	163	163	135	136	124	131	117	117	138	114	110	110	158	135	118	122	135	124	145



GS25 유통데이터

2ND 데이터 전처리/EDA

3.1 GS25 유통데이터 - 원본

GS25 유통데이터 원본 소개:

DT(date)	HDONG_CD	AMT_IND	LCLS_10_P	LCLS_20_P	LCLS_30_P	LCLS_40_P	LCLS_50_P	LCLS_60_P	LCLS_70_P	LCLS_80_P
영업일자	행정동코드	매출지수	식사_비중(%)	간식_비중(%)	마실거리_비중(%)	홈&리빙_비중(%)	헬스&뷰티_비중(%)	취미&여가활동_비중(%)	사회활동_비중(%)	임신/육아_비중(%)

▲	DT	HDONG_CD	AMT_IND	LCLS_10_P	LCLS_20_P	LCLS_30_P	LCLS_40_P	LCLS_50_P	LCLS_60_P	LCLS_70_P	LCLS_80_P
1	20180401	110515	0.7369329	0.2790179	0.2995689	0.3348155	0.015635754	0.053782440	0.0002648501	0.0130918695	0.0038227796
2	20180401	110530	1.2584726	0.2374031	0.2579840	0.3498257	0.045501936	0.092580946	0.0049875094	0.0024672924	0.0092495841
3	20180401	110540	0.6741641	0.2540806	0.2468443	0.3244343	0.065025245	0.089691441	NA	0.0132117587	0.0067122740
4	20180401	110550	1.0108743	0.4163683	0.2177150	0.2826133	0.026737568	0.036621671	0.0058931979	0.0122951542	0.0017557955
5	20180401	110560	1.0129422	0.3113209	0.2673872	0.2623582	0.014736400	0.135350211	NA	0.0072699316	0.0015771108
6	20180401	110570	1.9437500	0.3422831	0.2065746	0.3596700	0.022241870	0.048595631	0.0122819543	0.0038337490	0.0045190643
7	20180401	110580	0.8804933	0.3360336	0.2607941	0.2473888	0.077210136	0.046259803	0.0099750402	0.0191432105	0.0031953379
8	20180401	110600	0.7410180	0.2381865	0.2104442	0.4127694	0.016911631	0.090471264	0.0096374499	0.0205918511	0.0009877135
9	20180401	110615	1.2545984	0.2747816	0.2097499	0.3867400	0.024330760	0.076227504	0.0055390375	0.0174884222	0.0051427866
10	20180401	110630	0.7571405	0.2531996	0.2205302	0.3976594	0.021804458	0.046169977	0.0097986017	0.0192298645	0.0316078935

2ND 데이터 전처리/EDA

3.2 GS25 유통데이터 - 이상치

이상치 진단

18.11.10

18.11.11

19.02.14

19.03.14

4개의 날은 이상치로 판단하여 데이터에서 제외
(분석 파트에서 이상치 진단 기준 설명 예정)

2ND 데이터 전처리/EDA

3.3 GS25 유통데이터 - 상품별 매출 지수

우선, 상품별 상대지수(비중)을 절대지수로 바꾸기 위해 매출지수와 상품별 비중을 곱하였다.

$$(\text{매출 지수}) * (\text{상품별 비중}) = (\text{상품별 매출 지수})$$

- 변경 전(원본)

▲	DT	HDONG_CD	AMT_IND	LCLS_10_P	LCLS_20_P	LCLS_30_P	LCLS_40_P	LCLS_50_P	LCLS_60_P	LCLS_70_P	LCLS_80_P
1	20180401	110515	0.737	0.279	0.300	0.335	0.016	0.054	0.000	0.013	0.004
2	20180401	110530	1.258	0.237	0.258	0.350	0.046	0.093	0.005	0.002	0.009
3	20180401	110540	0.674	0.254	0.247	0.324	0.065	0.090	N/A	0.013	0.007

- 변경 후

▲	DT	HDONG_CD	AMT_IND	LCLS_10_P	LCLS_20_P	LCLS_30_P	LCLS_40_P	LCLS_50_P	LCLS_60_P	LCLS_70_P	LCLS_80_P
1	20180401	110515	0.737	0.206	0.221	0.247	0.012	0.040	0.000	0.010	0.003
2	20180401	110530	1.258	0.299	0.325	0.440	0.057	0.117	0.006	0.003	0.012
3	20180401	110540	0.674	0.171	0.166	0.219	0.044	0.060	N/A	0.009	0.005

2ND 데이터 전처리/EDA

3.4 GS25 유통데이터 - ratio 파생변수 생성

앞에서 분류한 케이웨더 미세먼지 좋은날&나쁜날을 이용해 GS ratio 생성
(행정동별, 주말주중별, 상품별 ratio값 생성)

$$\text{ratio} = \frac{\text{미세먼지 나쁜 날 GS 상품별매출지수 평균값}}{\text{미세먼지가 좋은 날 GS 상품별매출지수 평균값}} = \text{좋은날 대비 나쁜날 매출정도}$$

HDONG	weekend	gb	mAMT	m10	m20	m30	m40	m50	m60	m70	m80
J청운효자동	주중	ratio	0.91	0.91	1.01★	0.84	0.96	1.1★	0.82	0.96	0.68
J청운효자동	주말	ratio	0.94	0.95	0.99	0.89	1.05★	0.97	1.25★	0.94	0.66
J사직동	주중	ratio	0.99	0.98	1.03★	0.96	1.02★	1.03★	0.85	0.92	0.79
J사직동	주말	ratio	0.99	0.97	1.08★	0.93	1.03★	0.99	1.13★	0.96	0.86
J삼청동	주중	ratio	0.97	1.15★	1.02★	0.83	1.06★	0.94	0.67	0.81	0.46
J삼청동	주말	ratio	0.88	1.03★	0.91	0.81	0.83	0.84	0.72	0.74	0.71

2ND 데이터 전처리/EDA

3.4 GS25 유통데이터 - ratio 파생변수 생성

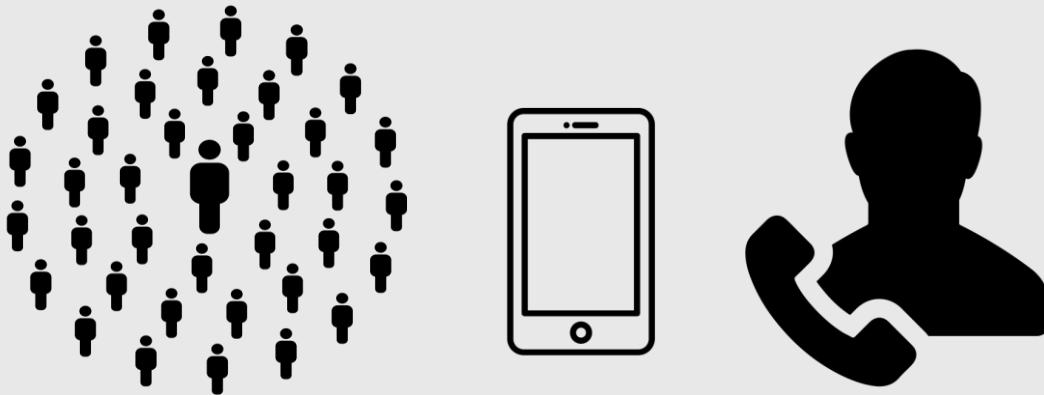
〈ratio 해석 예시〉

ratio값이 작을 수록, 해당 동의 주중 or 주말에, 미세먼지가 좋은 날에 비해 미세먼지가 나쁜 날 매출이 적다. ratio값이 1을 넘을 수록 미세먼지가 좋은 날에 비해 미세먼지가 나쁜 날 오히려 해당 품목 매출이 많다.

예를 들어, 종로구 청운효자동의 주말 주중 모두 10번 식사 품목의 ratio가 1보다 작으므로,
미세먼지가 좋은 날에 비해 나쁜 날 식사 품목의 매출 작다.

HDONG	weekend	gb	m10
J청운효자동	주중	ratio	0.91
J청운효자동	주말	ratio	0.95

HDONG	weekend	gb	mAMT	m10	m20	m30	m40	m50	m60	m70	m80
J청운효자동	주중	ratio	0.91	0.91	1.01★	0.84	0.96	1.1★	0.82	0.96	0.68
J청운효자동	주말	ratio	0.94	0.95	0.99	0.89	1.05★	0.97	1.25★	0.94	0.66
J사직동	주중	ratio	0.99	0.98	1.03★	0.96	1.02★	1.03★	0.85	0.92	0.79
J사직동	주말	ratio	0.99	0.97	1.08★	0.93	1.03★	0.99	1.13★	0.96	0.86
J삼청동	주중	ratio	0.97	1.15★	1.02★	0.83	1.06★	0.94	0.67	0.81	0.46
J삼청동	주말	ratio	0.88	1.03★	0.91	0.81	0.83	0.84	0.72	0.74	0.71



SK 유동인구 데이터

2ND 데이터 전처리/EDA

4.1 SK 유동인구 데이터 - 원본

SK 유동인구 데이터:

행정동 시간별 유동인구

◆	DT	▲ HDONG_CD	◆ HDONG_NM	◆ TIME	◆ VALUE
1	20180401	110515	청운효자동	0	609.11
2	20180401	110515	청운효자동	1	447.82
3	20180401	110515	청운효자동	2	301.75
4	20180401	110515	청운효자동	3	217.03
5	20180401	110515	청운효자동	4	237.07
6	20180401	110515	청운효자동	5	318.75
7	20180401	110515	청운효자동	6	519.28
8	20180401	110515	청운효자동	7	1131.14
9	20180401	110515	청운효자동	8	1588.58
10	20180401	110515	청운효자동	9	1881.46
11	20180401	110515	청운효자동	10	2520.06

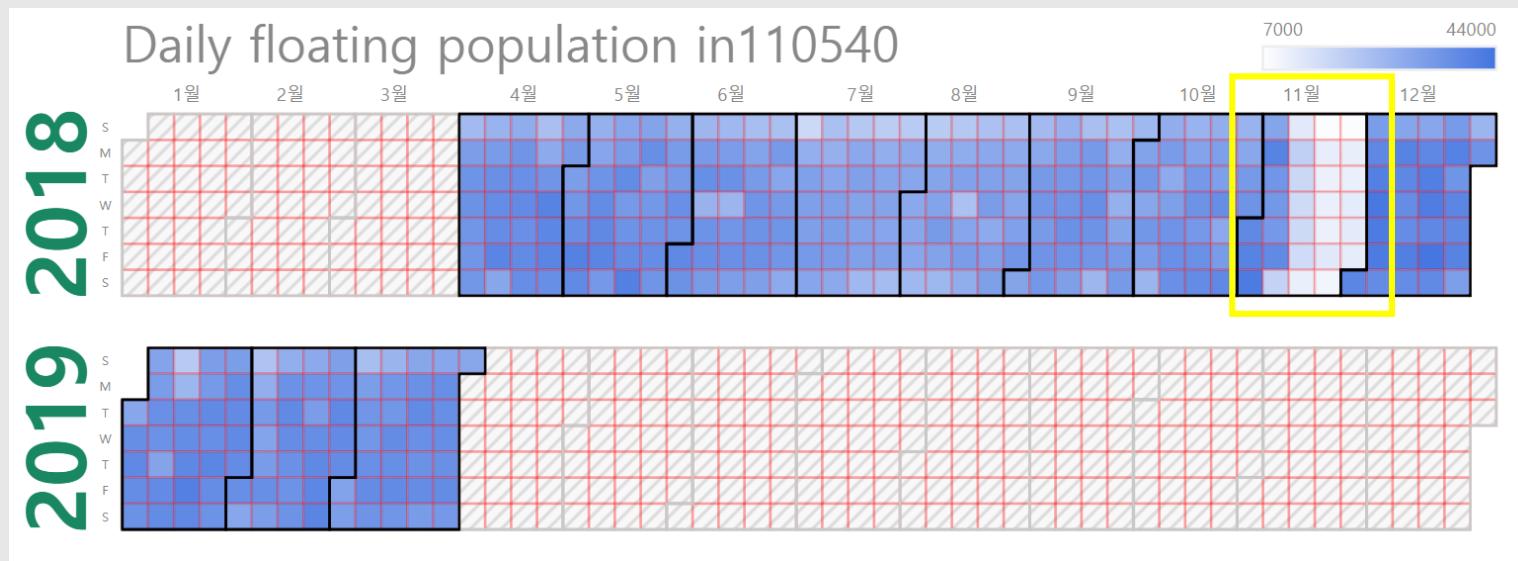
행정동 성별 연령별 유동인구

▲	DT	◆ HDONG_CD	◆ SEX	◆ AGE	◆ VALUE
1	20180401	110515	F	15	1301.69
2	20180401	110515	F	20	1038.43
3	20180401	110515	F	25	1437.80
4	20180401	110515	F	30	1356.08
5	20180401	110515	F	35	1598.14
6	20180401	110515	F	40	1495.53
7	20180401	110515	F	45	1768.21
8	20180401	110515	F	50	1611.47
9	20180401	110515	F	55	1433.06
10	20180401	110515	F	60	1212.56

2ND 데이터 전처리/EDA

4.2 SK 유동인구 데이터 - 이상치 처리

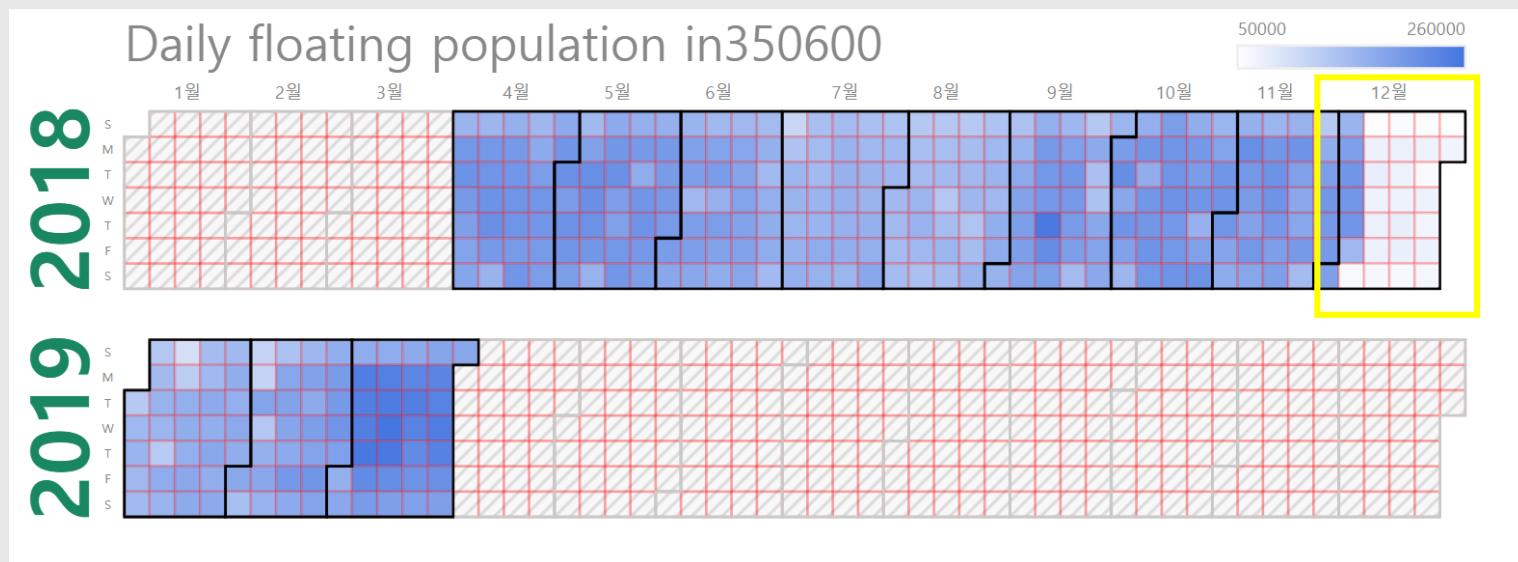
행정동 일별 유동인구



2ND 데이터 전처리/EDA

4.2 SK 유동인구 데이터 - 이상치 처리

행정동 일별 유동인구



2ND 데이터 전처리/EDA

4.2 SK 유동인구 데이터 - 이상치 처리

행정동 일별 유동인구

행정동별 NA처리 날짜

2018 2019

Daily floating

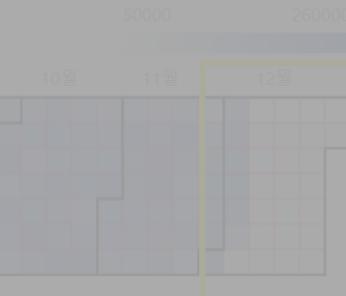
1월

2월

3월



- 110540: 11월10일~11월30일
- 110600: 11월10일~11월30일
- 350560: 12월19일~12월31일
- 350580: 12월28일~12월31일
- 350595: 12월22일~12월31일
- 350600: 12월08일~12월31일
- 350611: 12월14일~12월31일
- 350512: 12월14일~12월31일



2ND 데이터 전처리/EDA

4.3 SK 유동인구 데이터 - Ratio 파생변수 생성

앞에서 분류한 케이웨더 미세먼지 좋은날&나쁜날을 이용해, 유동인구 ratio 생성
(행정동별, 주말주중별, 시간별, 성별, 나이별 ratio값 생성)

$$\text{ratio} = \frac{\text{미세먼지 나쁜 날 유동인구 중앙값}}{\text{미세먼지가 좋은 날 유동인구 중앙값}} = \text{좋은날 대비 나쁜날 유동인구 정도}$$

(처리하지 못한 유동인구 이상치의 영향을 적게 받도록 하기 위해 중앙값을 사용)

HDONG	weekend	gb	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
J청운효자동	주중	ratio	0.87	0.83	0.83	0.83	0.95	0.97	0.95	0.93	0.94	0.91	0.9	0.87	0.9	0.93	0.91	0.92	0.88	0.89	0.89	0.87	0.87	0.92	0.91	0.92
J청운효자동	주말	ratio	0.94	0.9	0.91	0.9	0.92	0.87	0.87	0.91	0.94	0.98	1	0.99	1.03★	1.02★	1	0.98	1.03★	0.95	0.92	0.88	0.9	0.92	0.89	0.86
J사진동	주중	ratio	0.85	0.84	0.8	0.81	0.97	0.98	0.98	0.96	0.96	0.97	0.96	0.93	1	1.01★	0.95	0.92	0.93	0.94	0.9	0.88	0.88	0.91	0.88	0.88
J사진동	주말	ratio	0.93	0.93	0.92	0.89	0.92	0.94	0.95	0.93	0.91	0.95	0.97	1	1.01★	1.01★	1.01★	1.01★	0.97	0.93	0.93	0.89	0.9	0.9	0.86	0.84
J삼청동	주중	ratio	1.77★	1.86★	2.1★	2.21★	1.57★	1.34★	1.13★	1.02★	1.03★	1	1.02★	1.01★	1.04★	1.08★	1.07★	1.05★	1.01★	1.04★	1.03★	1.01★	1.17★	1.2★	1.29★	143★
J삼청동	주말	ratio	1.76★	1.94★	2.2★	2.17★	2.05★	1.39★	1.19★	1.15★	1.03★	1.01★	1.06★	1.07★	1.04★	1.06★	1.06★	1.02★	1.08★	1	1.02★	1.1★	1.13★	1.22★	1.38★	1.62★

2ND 데이터 전처리/EDA

4.3 SK 유동인구 데이터 - Ratio 파생변수 생성

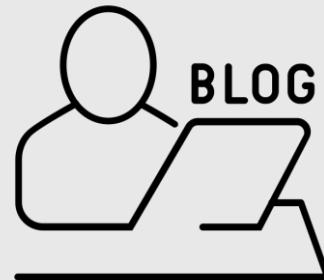
〈ratio 데이터 예시 - 시간별 유동인구 ratio〉

Ratio값이 작을 수록, 해당 동의 주중/주말 해당 시간대에, 미세먼지가 좋은 날에 비해 미세먼지가 나쁜 날 유동인구가 적다. ratio값이 1을 넘을 수록 미세먼지가 좋은 날에 비해 미세먼지가 나쁜 날 오히려 유동인구가 많다.

예를 들면, 종로구 청운효자동의 경우, 주중 주말 모두 0시에 ratio 값이 1보다 작으므로 미세먼지가 좋은 날에 비해 나쁜 날 유동인구가 적다.

HDONG	weekend	gb	0
J청운효자동	주중	ratio	0.87
J청운효자동	주말	ratio	0.94

HDONG	weekend	gb	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
J청운효자동	주중	ratio	0.87	0.83	0.83	0.83	0.95	0.97	0.95	0.93	0.94	0.91	0.9	0.87	0.9	0.93	0.91	0.92	0.88	0.89	0.89	0.87	0.87	0.92	0.91	0.92	
J청운효자동	주말	ratio	0.94	0.9	0.91	0.9	0.92	0.87	0.87	0.91	0.94	0.98	1	0.99	1.03★	1.02★	1	0.98	1.03★	0.95	0.92	0.88	0.9	0.92	0.89	0.86	
J사진동	주중	ratio	0.85	0.84	0.8	0.81	0.97	0.98	0.98	0.96	0.96	0.97	0.96	0.93	1	1.01★	1.01★	1.01★	1.01★	0.97	0.93	0.94	0.9	0.88	0.91	0.88	0.88
J사진동	주말	ratio	0.93	0.93	0.92	0.89	0.92	0.94	0.95	0.93	0.91	0.95	0.97	1	1.01★	1.01★	1.01★	1.01★	0.97	0.93	0.93	0.89	0.9	0.9	0.86	0.84	
J삼청동	주중	ratio	1.77★	1.86★	2.1★	2.21★	1.57★	1.34★	1.13★	1.02★	1.03★	1	1.02★	1.01★	1.04★	1.08★	1.07★	1.05★	1.01★	1.04★	1.03★	1.01★	1.17★	1.2★	1.29★	143★	
J삼청동	주말	ratio	1.76★	1.94★	2.2★	2.17★	2.05★	1.39★	1.19★	1.15★	1.03★	1.01★	1.06★	1.07★	1.04★	1.06★	1.06★	1.02★	1.08★	1	1.02★	1.1★	1.13★	1.22★	1.38★	1.62★	



와이즈넛 SNS 데이터

2ND 데이터 전처리/EDA

5.1 와이즈넛 SNS 데이터 - 원본 및 전처리 과정

REG_DT	SECTION	TITLE	CONTENT
20190401203000	뉴스	미세먼지 · 일교차에 건조한 날씨...소아 편도염 주의보	미세먼지 · 일교차에 건조한 날씨...소아 편도염 주의보 [앵커] 요즘같은 환절기는 몸 여가저기 탈이 나기 쉽죠. 특히나 목 관리 잘 하셔야겠습니다. 목 안에 편도가 부으면 목뿐 아니라 온몸이 아픈 편도염이 생기기 때문인데요. 면역력이 약한 어린이는 특히 조심해야 합니다. 이준호 기자입니다. [기자] 하루가 마다 하고 날아오는 미세먼지의 공습, 낮 다르고 저녁 다른 오락가락 기운, 여기에 건조한 대기까지. 관리한다고 해도 면역력이 약해지기 쉬운 것이 바로 요즘 환절기입니다. 특히 보통 감기몸살로 생각하기 쉬운 편도염 환자가 크게 늘어납니다. <신황애 / 국민건강보험공단 일산병원 이비인후과 교수> "급성 편도염은 갑작스럽게 시작돼 오한과 열이 동반되고 인후통, 두통이 있으면서 온몸이 쑤시는 통증이 있고 전신 쇠약감 등 신체 전반에 걸친 증상이 나타납니다." 편도염은 세균과 바이러스가 목 안에 있는 편도를 공격해 생깁니다. 한 해 환자만 600만명이 넘어 매년 한국인이 가장 많이 걸리는 병 3, 4위에 오를 정도인데 환자 5명 중 1명은 9살 이하 어린이입니다. 면역력이 약하고 학교, 어린이집처럼 단체생활을 하는 경우가 많기 때문입니다. 일주일 안에 증상이 사라지는 것이 대부분이지만 심하면 패혈증 같은 중한 합병증으로 이어질 수 있습니다. 이런 편도염을 막으려면 양치질을 자주하고 틈틈히 손을 씻어 입 안으로 세균과 바이러스가 못 들어가게 해야 막을 수 있습니다. 또 물을 자주 마시고 충분히 잠을 자는 것 역시 편도염 예방의 방법입니다. 연합뉴스TV 이준호입니다. humi@yna.co.kr 연합뉴스TV 기사문의 및 제보 : 카톡/라인 jebo23 ► 연합뉴스TV 네이버 채널 구독 ► 생방송 시청 ► 대한민국 뉴스의 시작, 연합뉴스TV 앱 다운받기

SECTION 컬럼
뉴스/카페/블로그
분류

CONTENT 컬럼만 선택
가정: TITLE은 CONTENT
의 요약문이므로 제거

미세먼지
좋은 날/ 나쁜 날
코퍼스 구축

TermDocumentMatrix
생성

두 글자 이상 명사추출
(SimplePos09)

맞춤법 교정 및
영어, 숫자,
특수문자 제거

2ND 데이터 전처리/EDA

5.2 SNS데이터 – 외부데이터 네이버 쇼핑 검색지수



분야

생활/건강 > 건강관리용품 > 먼지차단마스크

기간

일간 < 1개월 3개월 1년 직접입력 2018 04 01 – 2019 03 31 · 2017년 8월부터 조회할 수 있습니다.

기기별

전체 PC 모바일 | 성별 전체 여성 남성 | 연령 전체 10대 20대 30대 40대 50대 60대 이상

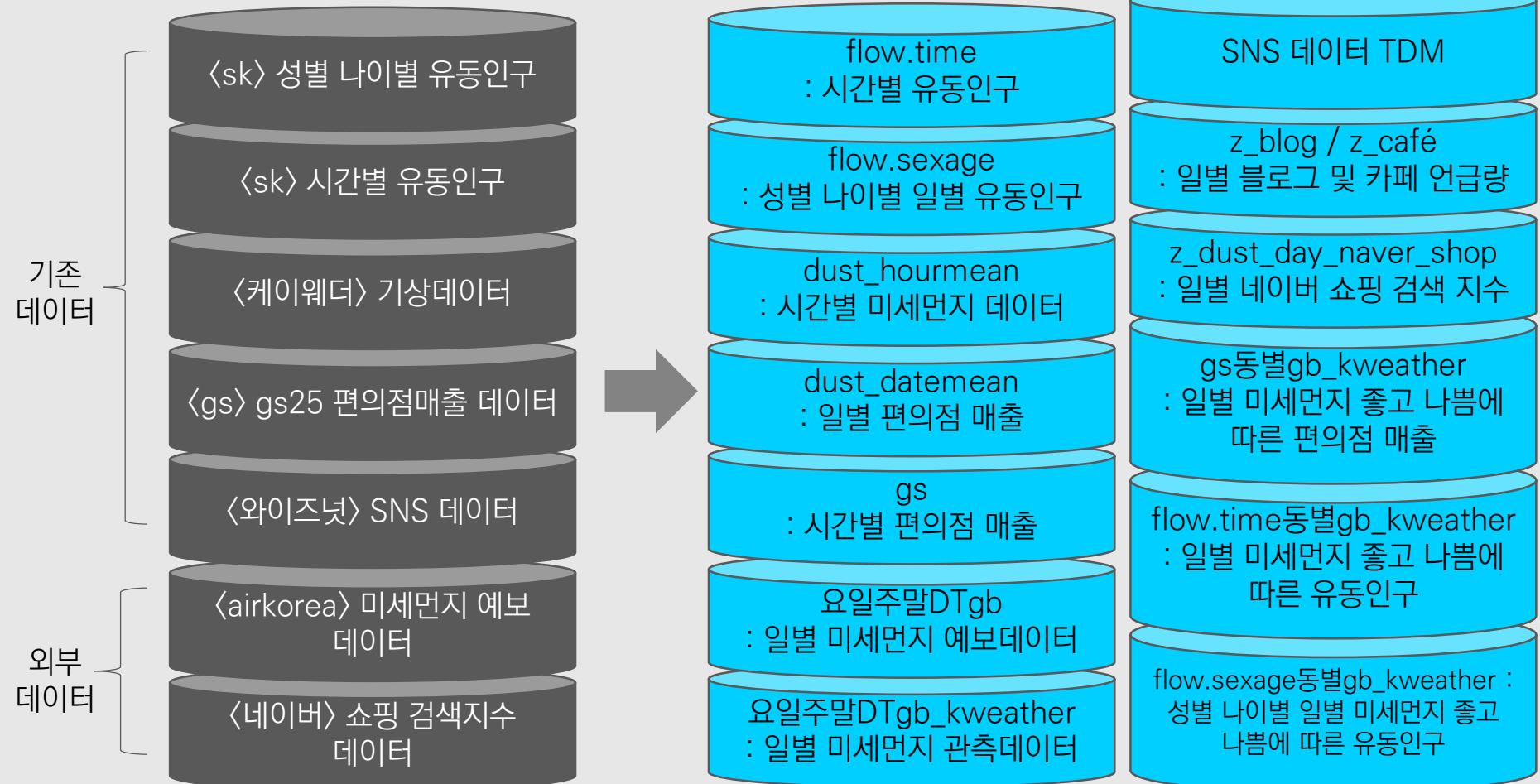
[조회하기](#)

미세먼지 농도에 따른 소비 심리를 알아보기 위해

외부데이터로 네이버의 쇼핑 검색 지수

2ND 데이터 전처리/EDA

6. 일별 및 시간별 데이터 정리 후 최종 데이터 베이스



3

분석



GS25 편의점 매출 데이터

&



케이웨더 환경기상데이터

3RD 분석

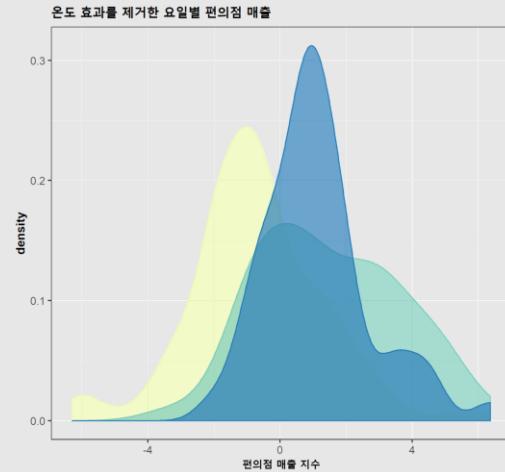
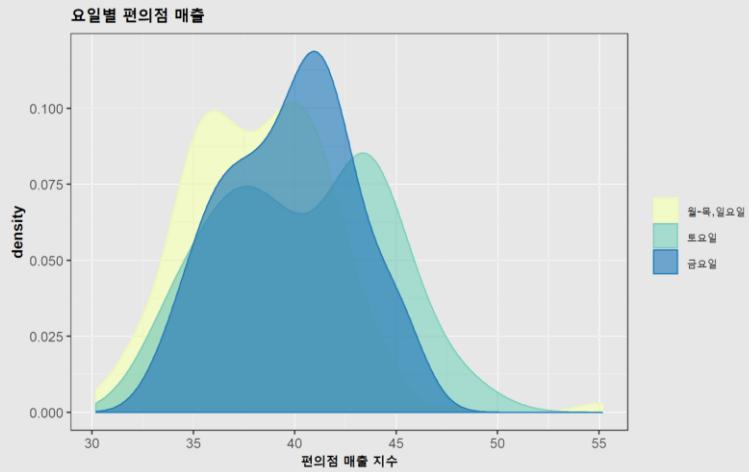
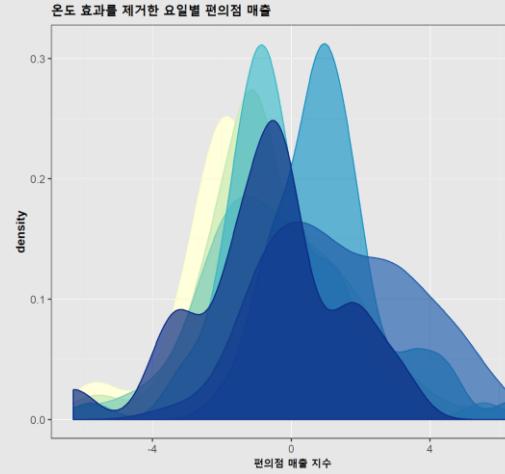
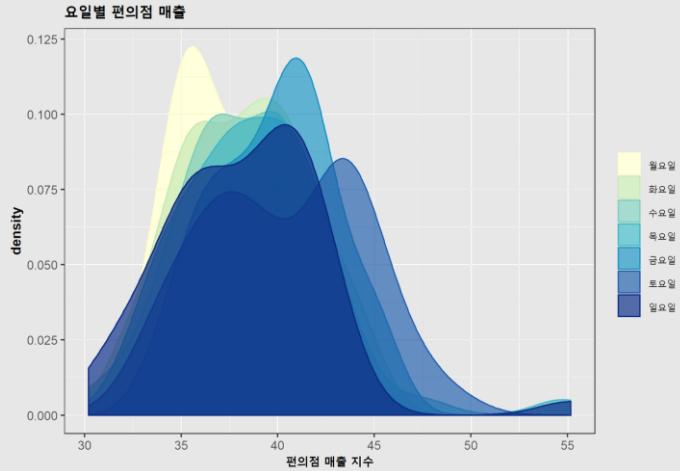
1. GS25 유통 데이터 -편의점 매출에 영향을 미치는 변수 찾기



GS25 편의점 매출에 영향을 끼치는 변수파악을 위해
요일, 온도, 그리고 미세먼지/초미세먼지 농도와의 관계를 살펴봄

3RD 분석

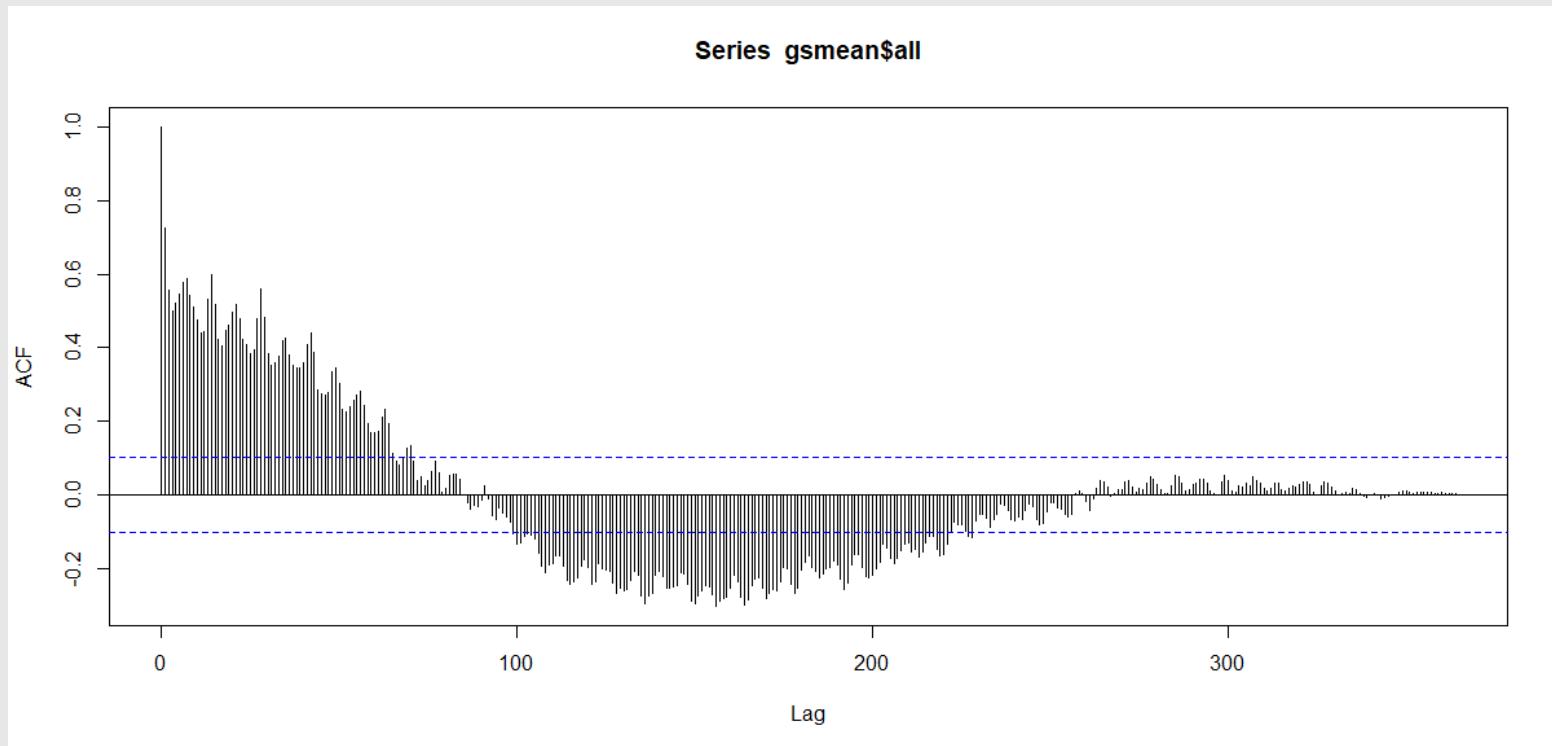
1.1 GS25 유통 데이터 - 요일



3RD 분석

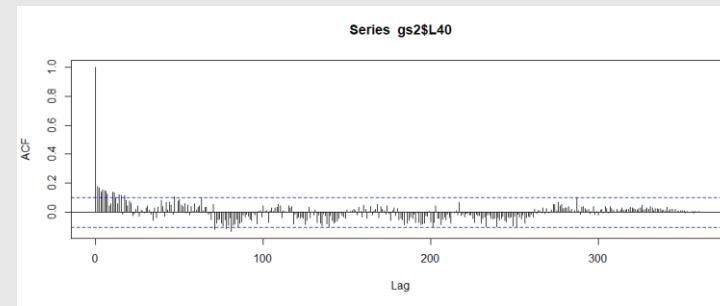
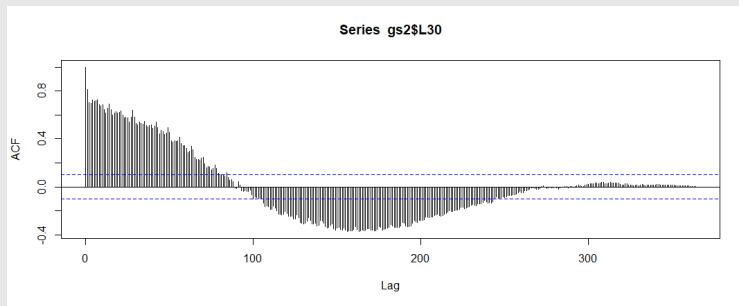
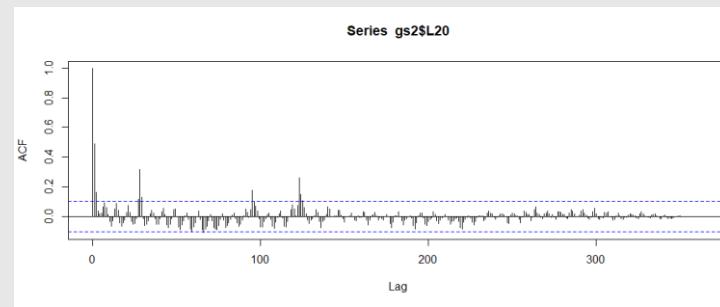
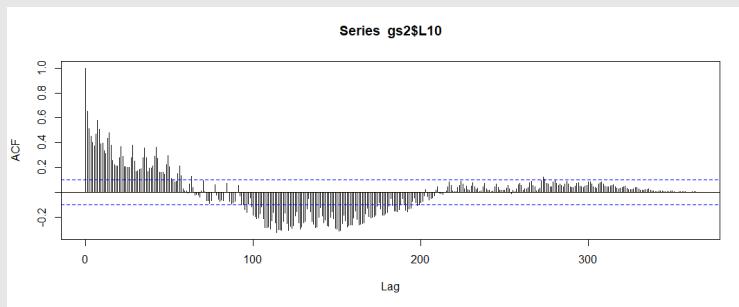
1.1 GS25 유통 데이터 - 요일

Time series data ACF



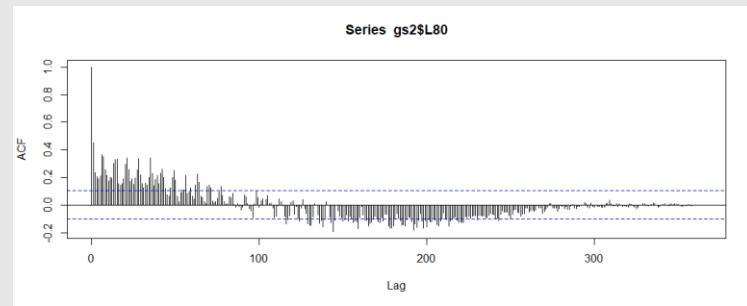
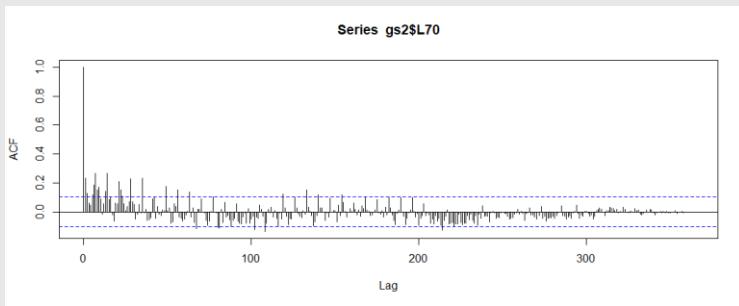
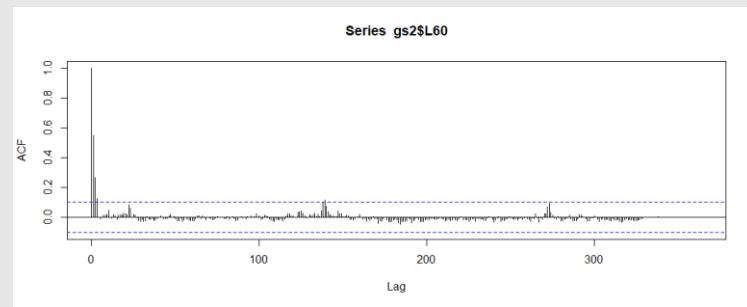
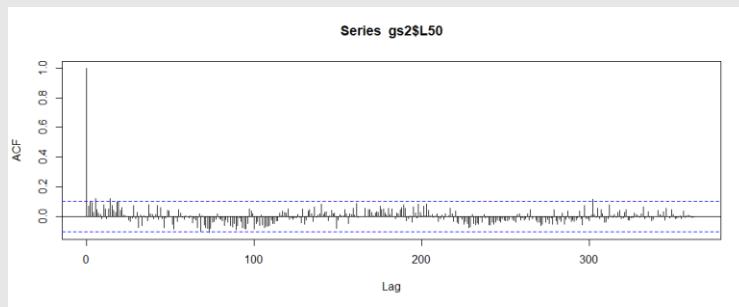
3RD 분석

1.1 GS25 유통 데이터 - 요일



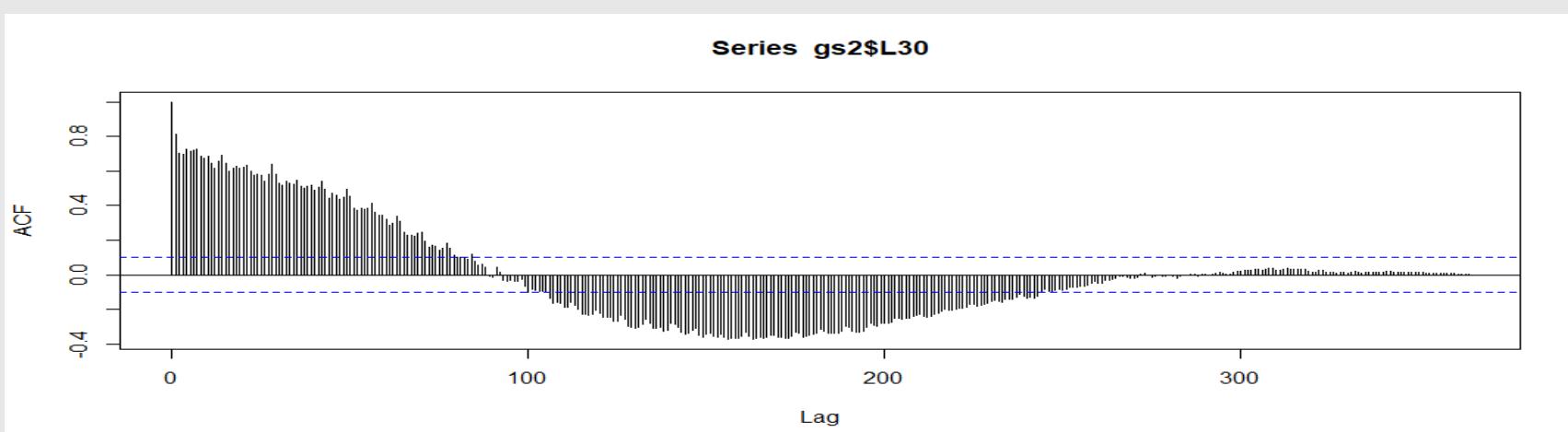
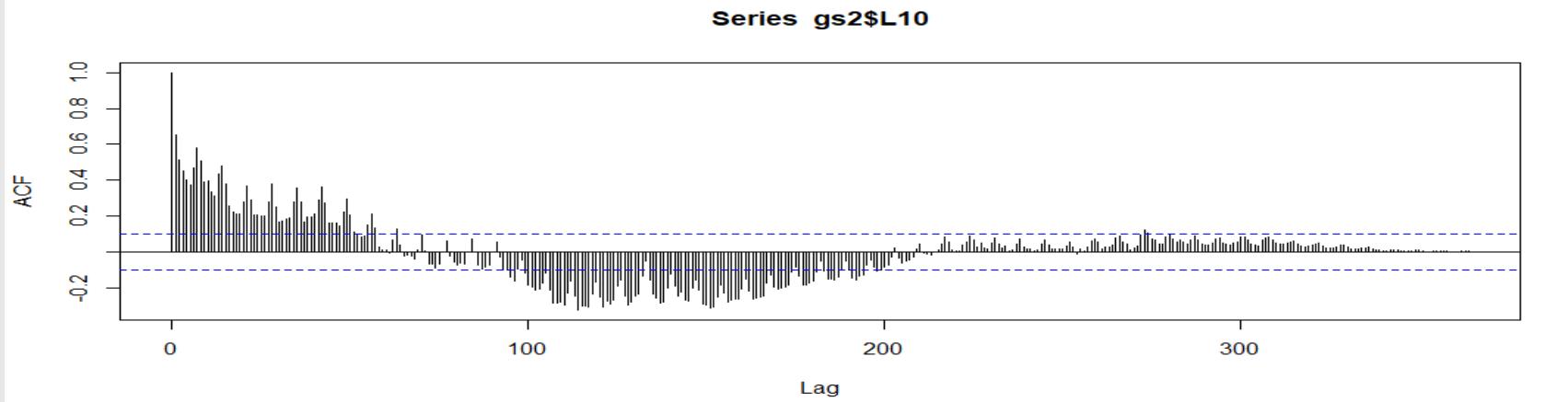
3RD 분석

1.1 GS25 유통 데이터 - 요일



3RD 분석

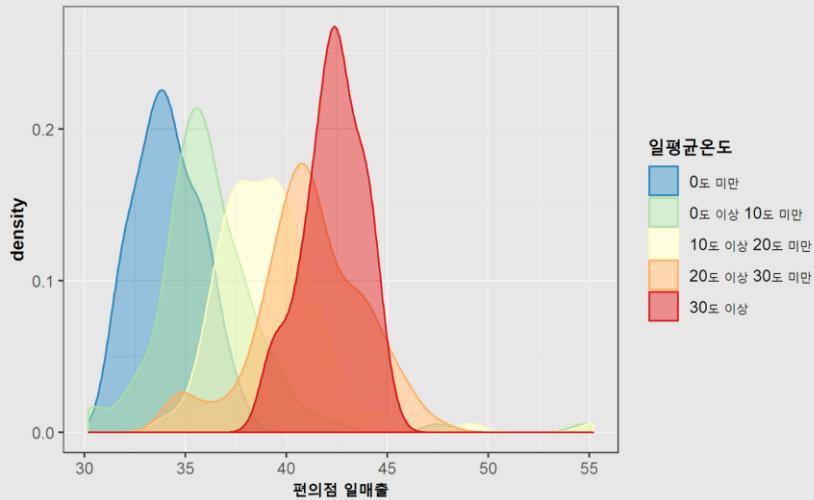
1.1 GS25 유통 데이터 - 요일



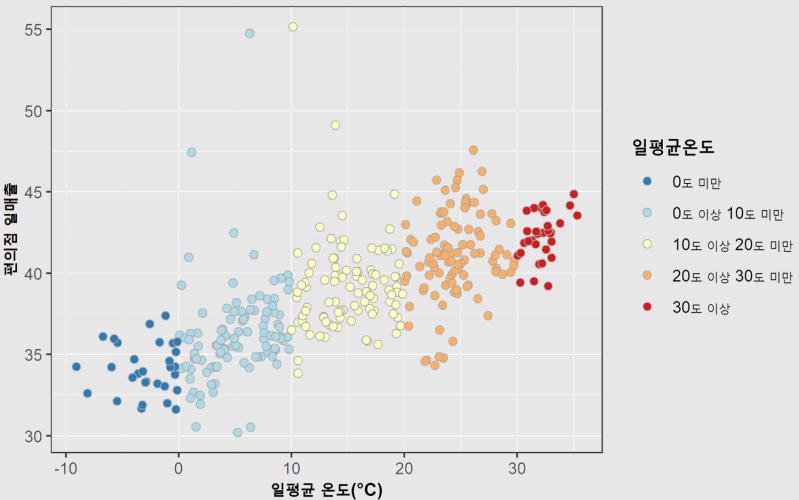
3RD 분석

1.2 GS25 유통 데이터 - 온도

일평균온도에 따른 편의점 매출의 분포



온도에 따른 편의점 매출



편의점 매출은 온도에 가장 큰 영향을 받는다

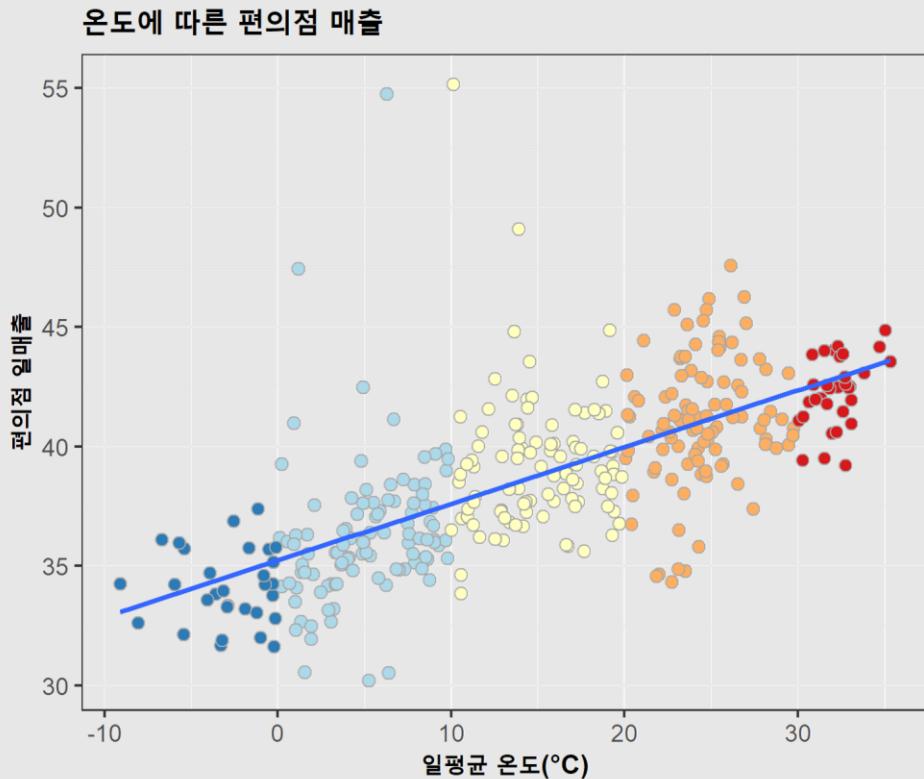
온도가 높아질 수록 매출이 증가하는 경향을 보인다

산점도와 상관계수를 통해 온도와 매출의 선형관계 확인 가능

상관계수: 0.686 (동떨어진 4개의 자료점 제외 시: 0.753)

3RD 분석

1.1 GS 유통 데이터 - 온도



선형 회귀분석을 통해 통계적으로 유의미한 선형관계가 있는지 파악

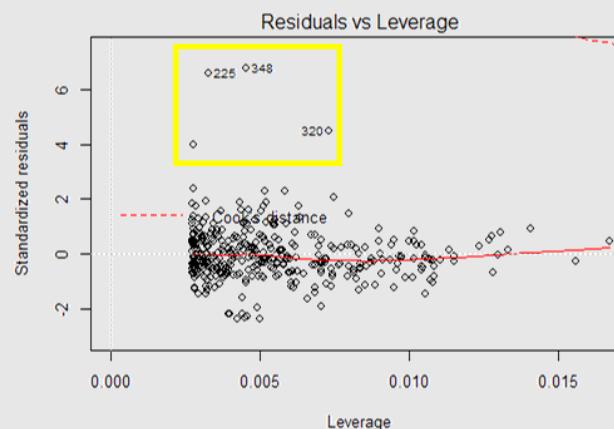
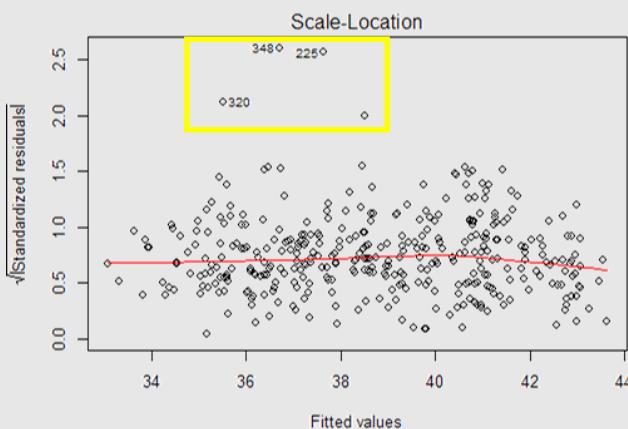
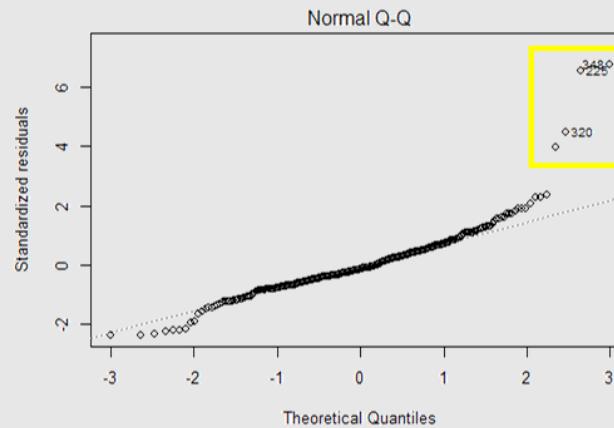
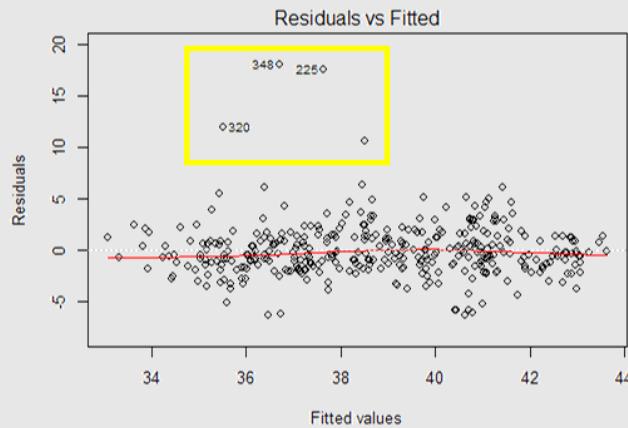
추정 회귀식:

$$\text{편의점 매출} = 35.234 + 0.237 * \text{온도}$$
$$(R^2 = 0.4703)$$

F검정결과 선형관계가 유의함
(F = 322.3, p-value = 0)

3RD 분석

1.2 GS25 유통 데이터 - 온도 : 이상치 진단



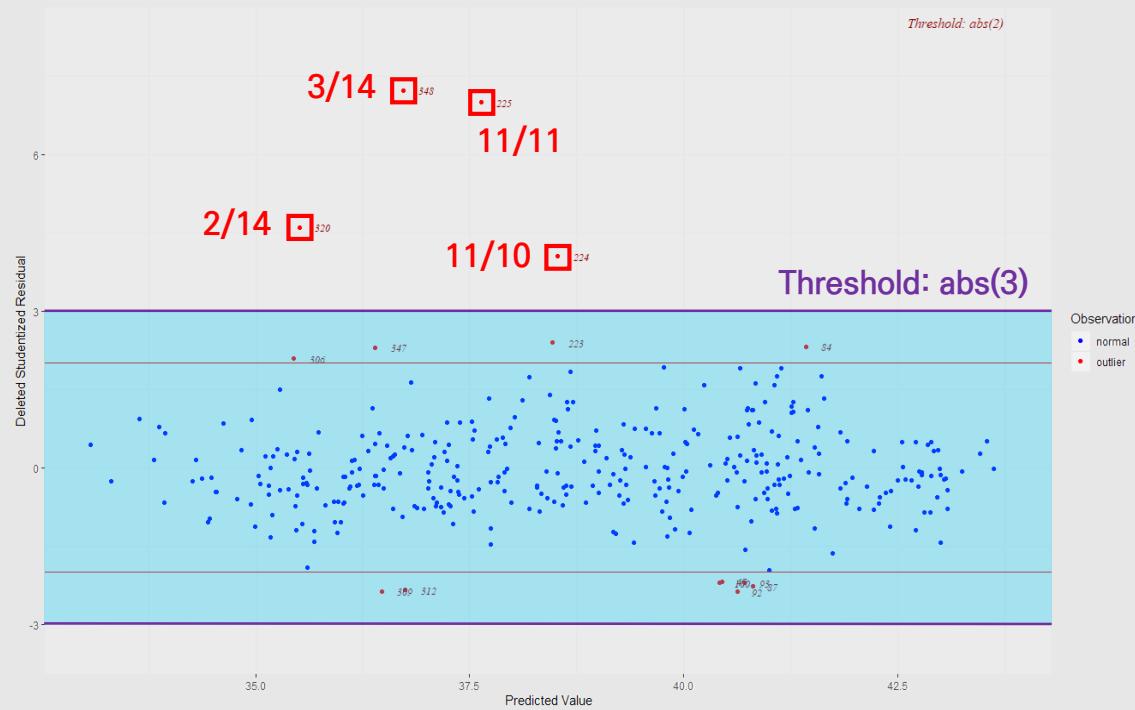
잔차그림으로
선형모형의 적합성
과 이상행동을 보이는
자료점을 확인

산점도와 마찬가지로
네 개의 자료점이
다른 자료점과
매우 동떨어져 보인다

3RD 분석

1.2 GS25 유통 데이터 - 온도 : 이상치 진단

스튜던트 삭제 잔차 대 적합치 그림



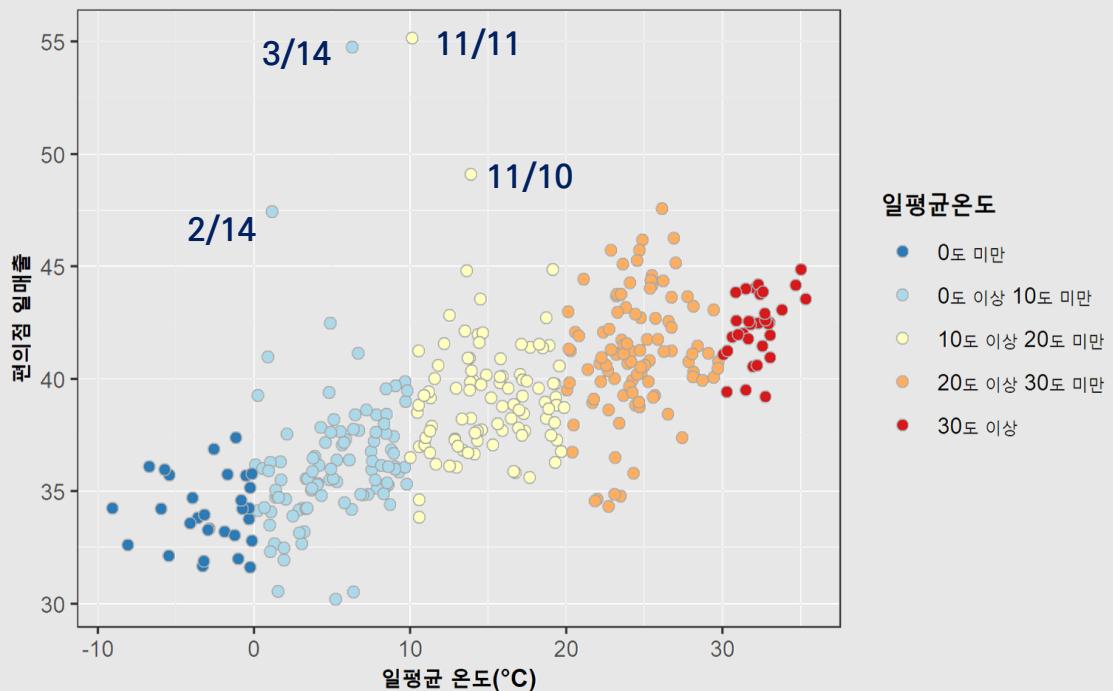
0을 중심으로 ± 3 안의 수평구간 안에 대부분의 잔차가 존재하고, 네 개의 자료점만 따로 떨어져 있으므로

이상점을 4개로 확정

3RD 분석

1.2 GS25 유통 데이터 - 온도 : 이상치 진단

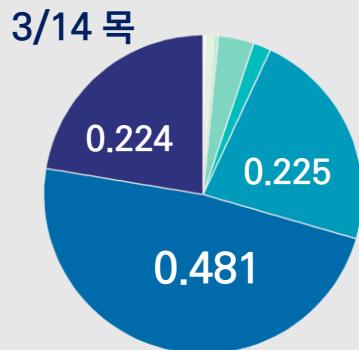
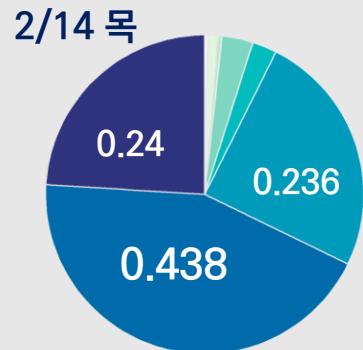
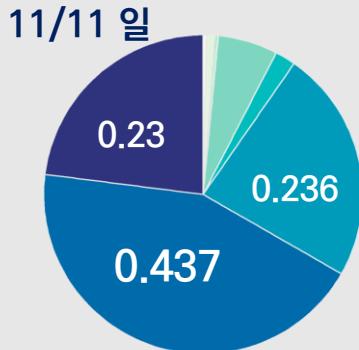
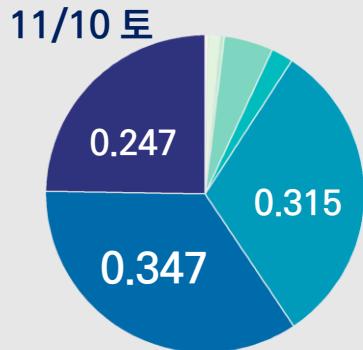
온도에 따른 편의점 매출



일평균온도와 매출의 선형
관계에서 벗어난 네 개의
자료점을 확인하면 각각
발렌타인 데이(2/14),
화이트 데이(3/14),
빼빼로 데이(11/11),
빼빼로 데이 전날(11.10)
이다.

3RD 분석

1.2 GS25 유통 데이터 - 온도 : 이상치 진단



각각의 날짜에 대한
항목 별 매출 비중을
살펴본 결과,
전체 매출의 간식항목
비중(0.212)에 비해
간식항목의 매출 비중이
월등히 높으므로 빼빼로
데이, 발렌타인 데이, 화
이트 데이 특수로 판단

온도 외의 요인이 작용하여 매출 이상치가 발생한 것으로 판단

3RD 분석

1.3 GS 유통 데이터 - 미세먼지

〈총 매출과 미세먼지와의 관계〉

각 행정동별 365일 GS 총매출의 중앙값보다 높은 날, 낮은 날을 나눈 후
미세먼지가 나쁜 날, 좋은 날을 전처리 데이터와 매칭시켜 2x2 table 을 만들었다.
그리고 미세먼지의 좋고나쁨과 총매출의 높고낮음이 연관이 있는지 2x2 빈도 교차분석을 실시했다.

귀무가설: 미세먼지의 좋고 나쁨과, 총매출의 높고 낮음은 서로 독립이다.

대립가설: 미세먼지의 좋고 나쁨과, 총매출의 높고 낮음은 서로 연관이 있다.

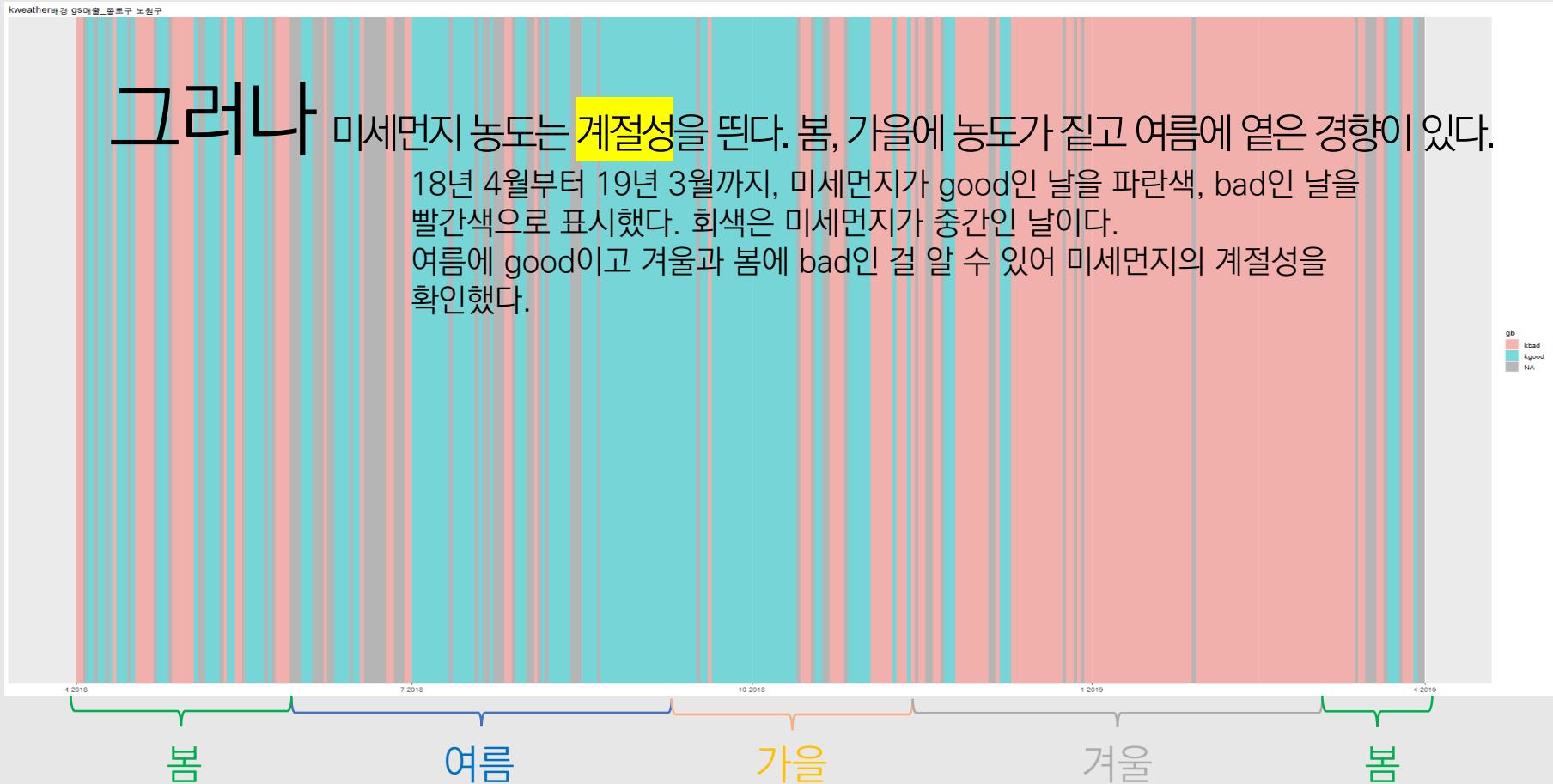
유의수준 0.05하에서, 32개의 행정동이 귀무가설을 기각했다. 즉 32개 동에서 미세먼지의 좋고나쁨과 총매출의 높고낮음은 연관이 있다.

행정동	J청운효자동	J사직동	J삼청동	J부암동	J평창동	J무악동	J교남동	J가회동	J종로1,2,3,4가동	J종로5,6가동	J이화동	J혜화동	J창신1동	J창신2동	J창신3동	J승인1동	J승인2동
p-value	0.000	0.853	0.046	0.756	0.001	0.000	0.000	0.000	0.000	0.001	0.610	0.001	0.000	0.000	0.969	0.000	0.000

행정동	N월계1동	N월계2동	N월계3동	N공릉1동	N공릉2동	N하계1동	N하계2동	N중계본동	N중계1동	N중계4동	N중계2,3동	N상계1동	N상계2동	N상계3,4동	N상계5동	N상계6,7동	N상계8동	N상계9동	N상계10동
p-value	0.112	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.000	0.000	

3RD 분석

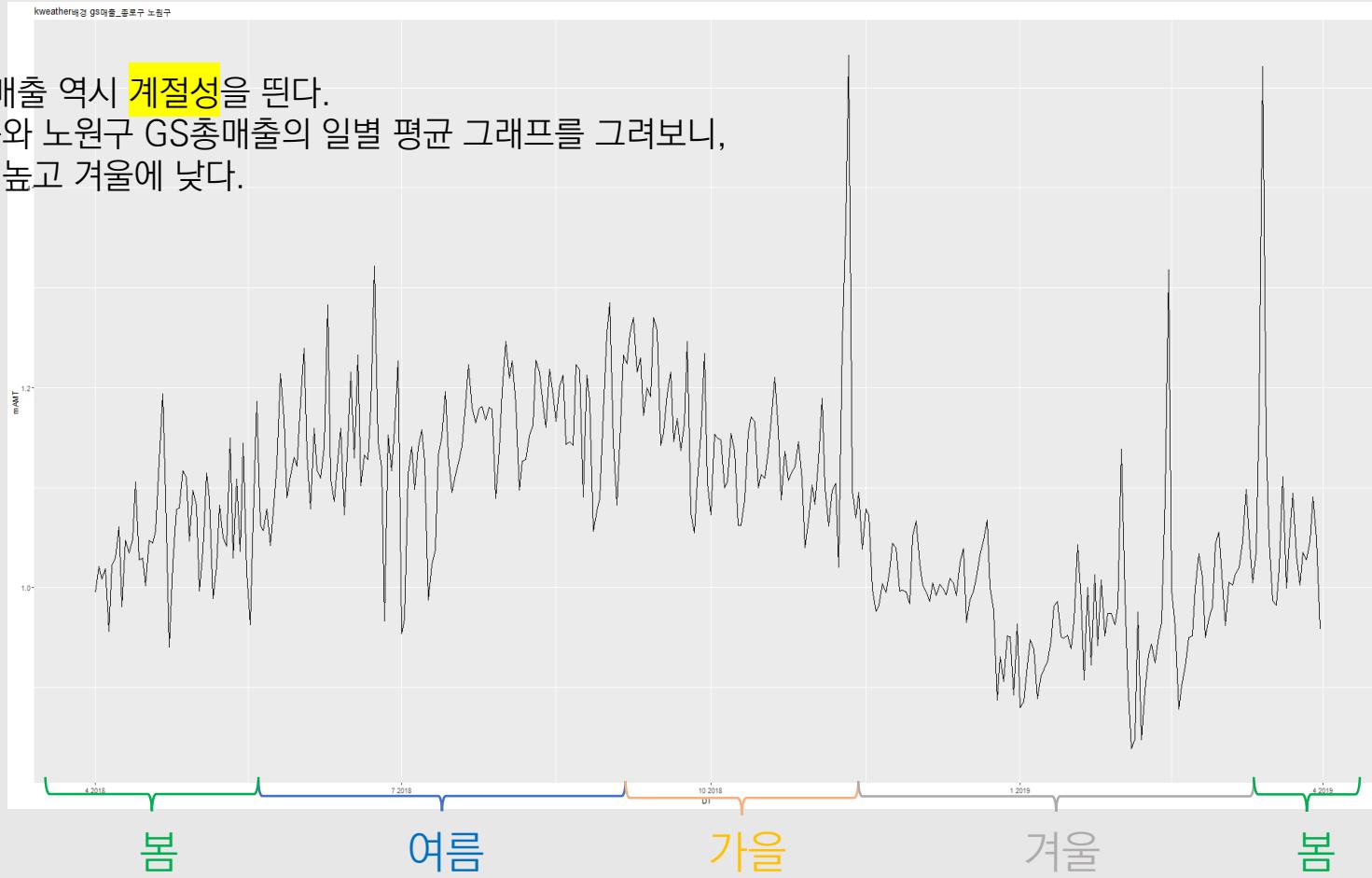
1.3 GS 유통 데이터 - 미세먼지 : 계절성



3RD 분석

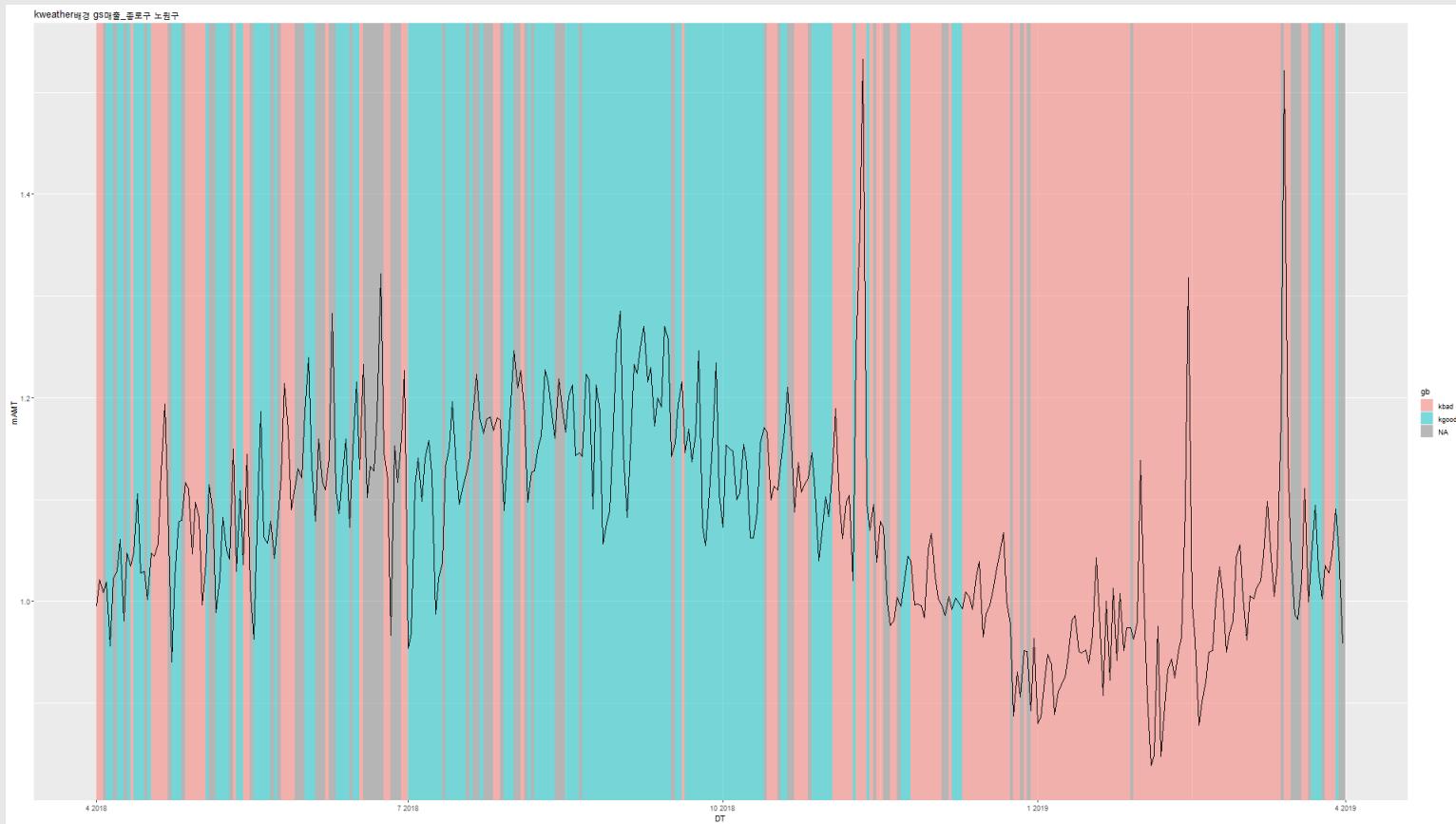
1.3 GS 유통 데이터 - 미세먼지 : 계절성

GS 총매출 역시 계절성을 띈다.
종로구와 노원구 GS총매출의 일별 평균 그래프를 그려보니,
여름에 높고 겨울에 낮다.



3RD 분석

1.3 GS 유통 데이터 - 미세먼지 : 계절성



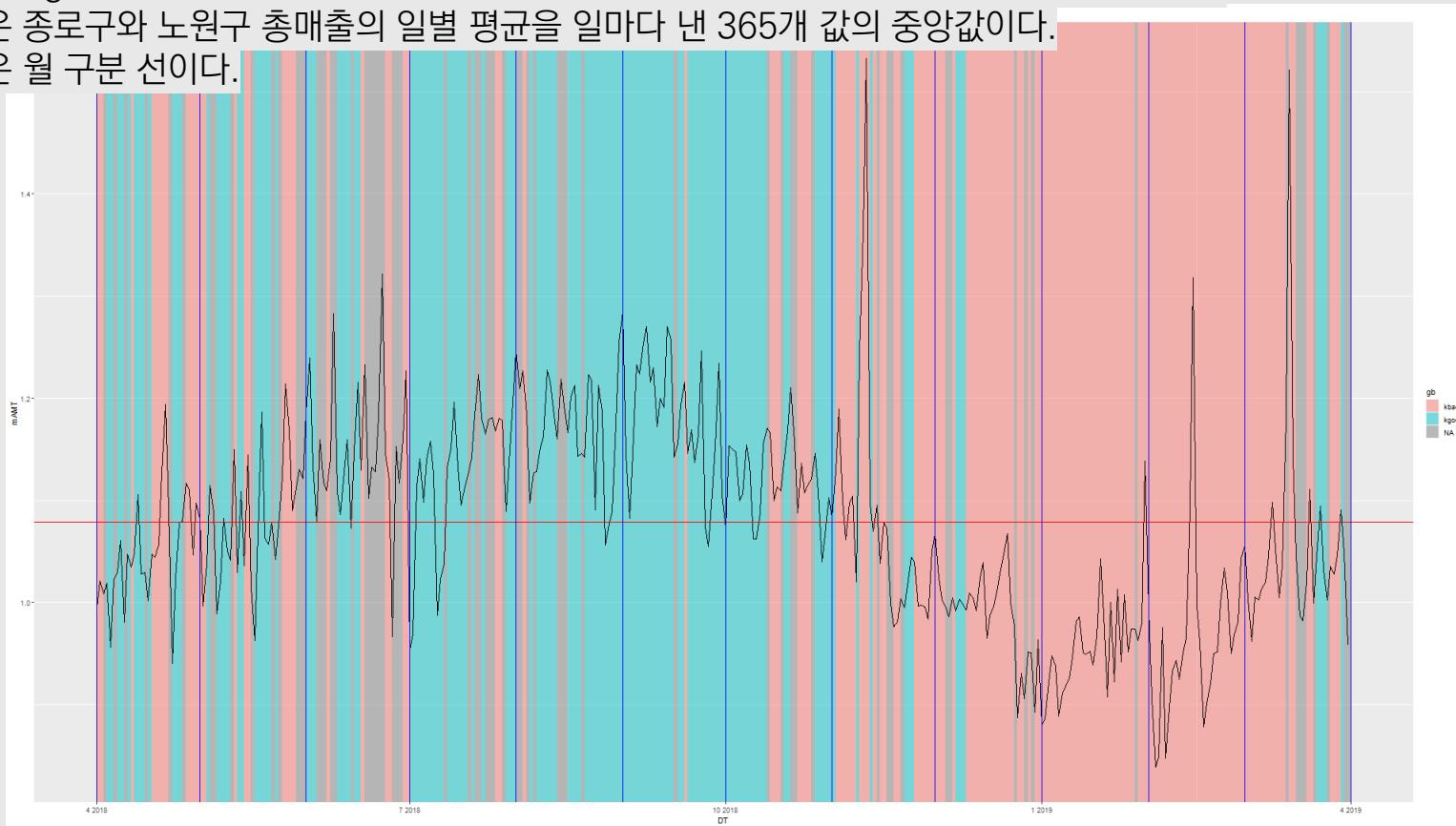
3RD 분석

1.3 GS 유통 데이터 - 미세먼지 : 계절성

미세먼지가 good, bad인 날을 배경으로 하고, 총매출을 그래프로 하여 겹쳐 그려보면 이와 같다.

빨간 선은 종로구와 노원구 총매출의 일별 평균을 일마다 낸 365개 값의 중앙값이다.

파란 선은 월 구분 선이다.



3RD 분석

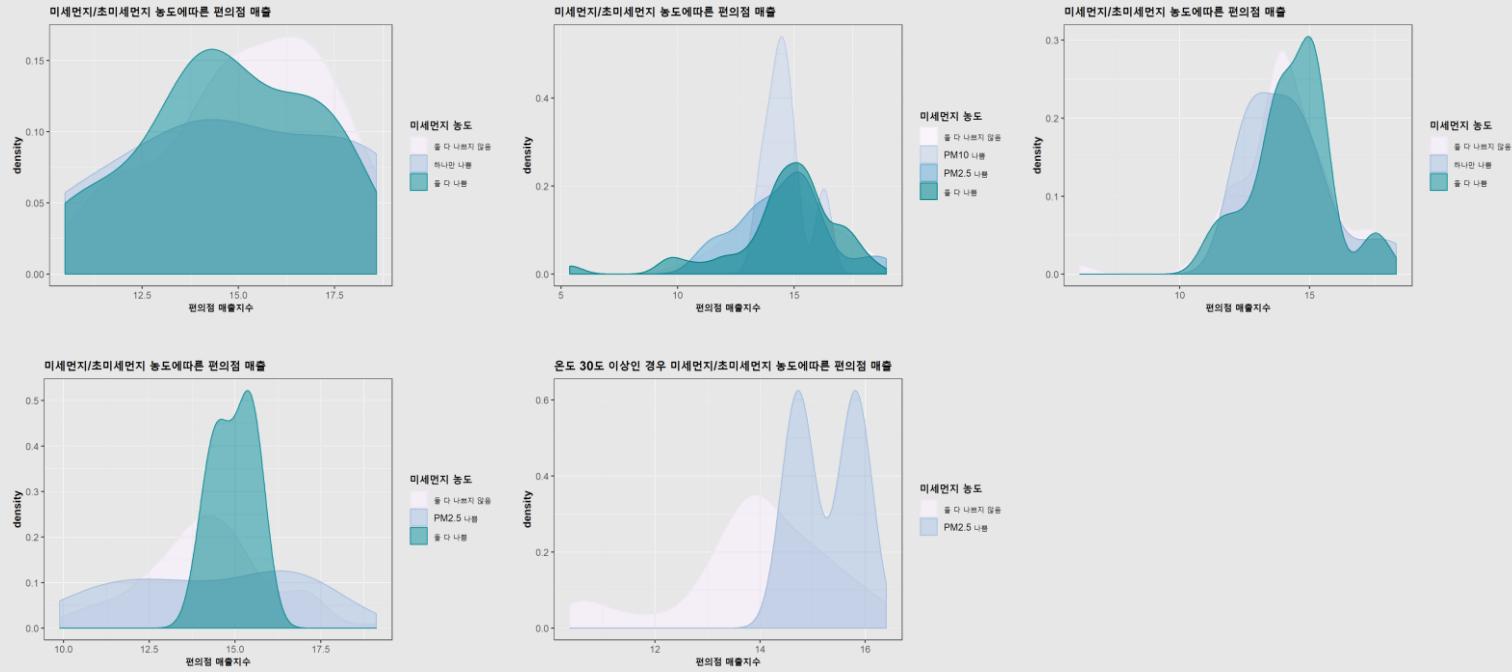
1.3 GS 유통 데이터 - 미세먼지 : 계절성 통제

미세먼지도 계절성을 따르고, 매출도 계절성을 따른다. 그러므로 미세먼지와 매출의 연관성은 계절을 통제했을 때 없을 수도 있다.

따라서, 온도 혹은 월(month) 변수를 통제했을 때도 미세먼지의 좋고나쁨과 GS총매출의 높고낮음이 서로 연관이 있는지 알아보았다.

3RD 분석

1.3 GS 유통 데이터 - 미세먼지 : (1) 온도 통제 시 미세먼지와 매출의 관계



온도 10도 간격으로 그룹을 나누어 온도의 영향을 제외하고
미세먼지와 편의점 매출의 관계가 있는지 살펴보았으나
유의미한 관계가 있다고 보기 어렵다

3RD 분석

1.3 GS 유통 데이터 - 미세먼지 : (2) 월(month) 통제 시 미세먼지와 매출의 관계

월별이라는 범주형 변수를 통제한 상태에서도 미세먼지의 좋고나쁨과 GS총매출의 높고낮음이 서로 연관이 있는지 Mantel Haenszel test를 해보자.

- * 귀무가설: 월 변수를 통제했을때, 미세먼지의 좋고나쁨과 GS총매출의 높고 낮음은 독립이다. 서로 연관이 없다.
- * 대립가설: 월 변수 요인을 제거했을 때, 미세먼지의 좋고나쁨과 GS총매출의 높고 낮음은 서로 독립이 아니고 연관이 있다.

그 결과, 36개 동 중 4개의 동을 제외한 32개의 동이 독립이라는 결과가 나왔다.

3RD 분석

1.3 GS 유통 데이터 - 미세먼지 : (2) 월(month) 통제 시 미세먼지와 매출의 관계

- 월 수준에 따라 미세먼지와 GS 총 매출이 차이가 나는지,
Mantel Haenszel test 검정 p-value값

행정동	J청운효자동	J사직동	J삼청동	J부암동	J평창동	J무악동	J교남동	J가회동	J종로 1,2,3,4가동	J종로 5,6가동	J이화동	J혜화동	J창신1동	J창신2동	J창신3동	J승인1동	J승인2동
p-value	0.036	0.088	0.531	0.244	0.338	0.621	0.827	0.264	0.246	0.164	0.113	0.534	0.981	0.182	0.257	0.132	0.116

행정동	N월계1동	N월계2동	N월계3동	N공릉1동	N공릉2동	N하계1동	N하계2동	N중계본동	N중계1동	N중계4동	N중계2,3동	N상계1동	N상계2동	N상계3,4동	N상계5동	N상계6,7동	N상계8동	N상계9동	N상계10동
p-value	0.747	0.011	0.284	0.128	0.869	0.533	0.911	0.783	0.428	0.942	0.020	0.264	0.894	0.902	0.439	0.138	0.469	0.425	0.033

월 변수를 통제 했을 때, 유의수준 0.05하에서 32개 동이 GS매출과 미세먼지 농도는 독립이다. 4개의 동은 GS매출과 미세먼지 농도가 연관이 있다. 이를 미세먼지 민감지역으로 분류할 수 있다.

4개의 동(미세먼지 민감 지역)

* 종로구-청운효자동

* 노원구-월계2동/중계2,3동/상계10동

3RD 분석

1.3 GS 유통 데이터 - 미세먼지와 상품군 별 매출의 관계

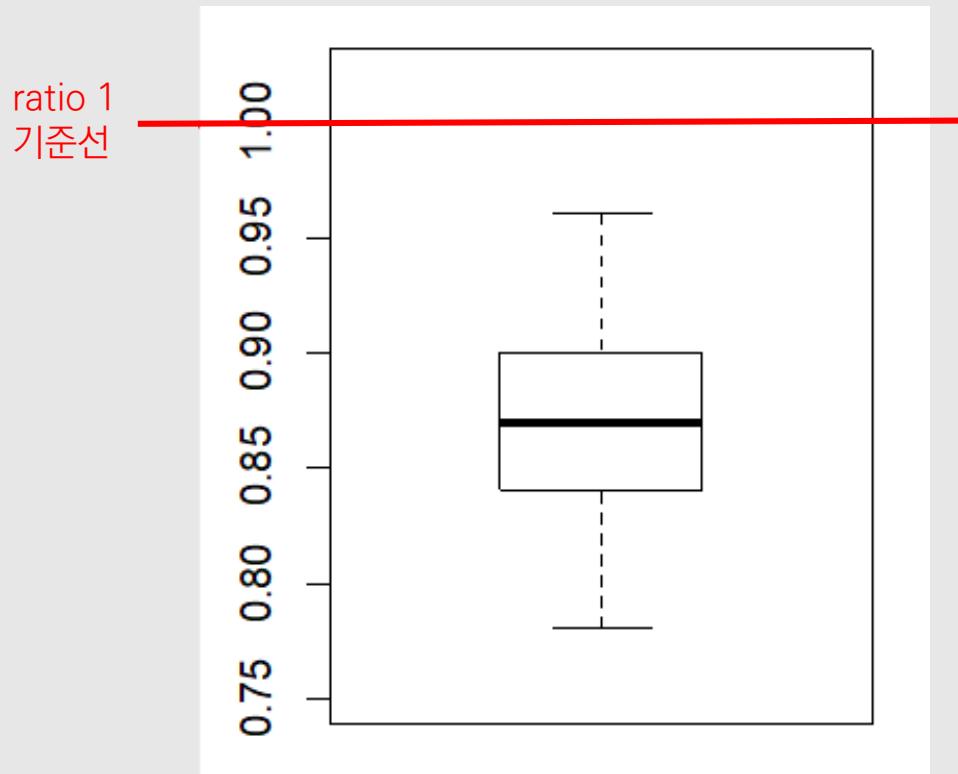
〈상품군 별 매출과 미세먼지 살펴보기〉

36개동의 미세먼지 좋은날, 나쁜날 GS 카테고리별 매출 평균을 비교해보았다. 미세먼지 나쁜 날의 매출 평균이 더 높은 동의 개수와, 36개 동 중의 percent로 표시했다. 주중, 주말로 나누어 살펴봤다. 음료는 항상 미세먼지 나쁜날 매출액 평균이 항상 더 낮고, 간식은 63.9~66.7%의 동이 미세먼지가 나쁜 날 매출량 평균이 더 높다.

	m10	m20	m30	m40	m50	m60	m70	m80	mAMT
주중 count	6	23	0	15	7	4	16	6	1
주중 percent	0.167	0.639	0	0.417	0.194	0.111	0.444	0.167	0.028
주말 count	2	24	0	15	2	16	11	10	1
주말 percent	0.056	0.667	0	0.417	0.056	0.444	0.306	0.278	0.028

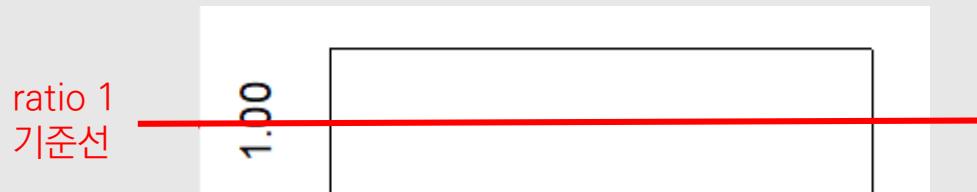
3RD 분석

1.3 GS 유통 데이터 - 미세먼지와 상품군 별 매출의 관계



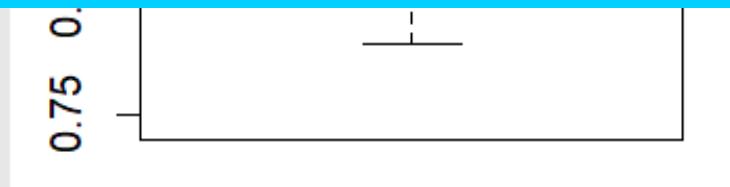
3RD 분석

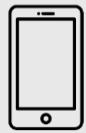
1.3 GS 유통 데이터 - 미세먼지와 상품군 별 매출의 관계



→ 공공보건의료사업단 김경남교수¹에 따르면 미세먼지의 영향을 줄여주는 식품은 없지만 코와 호흡기 점막의 수분량을 증가시켜 미세먼지를 흡착시켜 배출 할 수 있도록 미세먼지가 나쁜 날엔 물을 많이 마셔야 한다. 따라서, **미세먼지가 나쁜 날 생수&차의 판매를 촉진시킬 필요가 있다.**

¹ 〈미세먼지에 건강관리 '비상'… "물·마스크는 필수품"〉, 연합뉴스, 기사작성 2018/11/27, 접속일자 2019/09/09,
<https://www.yonhapnews.co.kr/view/AKR20181127113100017?input=1195m>





SK 유동인구 데이터

&



케이웨더 환경기상데이터

3RD 분석

2.1 SK유동인구 데이터 - 시간별 유동인구와 미세먼지

종로구의 17개 동과, 노원구의 19개 동을 살펴보았다. 미세먼지가 나쁜날 유동인구의 중앙값을 좋은 날의 유동인구 중앙값으로 나눈, 유동인구 ratio 데이터를 사용했다. (전처리에 기재한 데이터)

종로구, 노원구별로 ratio가 10이 넘는 동 개수(count)와, 해당 구의 동개수 대비 비율(percent)를 추출해봤다.

노원구는 미세먼지가 나쁜 날 유동인구가 적은 반면, 종로구는 유동인구가 더 많아지는 특정 시간대가 존재했다. 특히 주말의 종로구는 전반적으로 미세먼지가 나쁨에도 유동인구가 더 많았다.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
종로구 주중 count	3	10	8	5	11	7	4	6	3	5	8	6	8	8	8	7	7	7	1	2	4	7	3	2
종로구 주중 percent	0.176	0.588	0.471	0.294	0.647	0.412	0.235	0.353	0.176	0.294	0.471	0.353	0.471	0.471	0.471	0.412	0.412	0.412	0.059	0.118	0.235	0.412	0.176	0.118
종로구 주말 count	10	8	13	10	11	13	8	8	12	11	12	15	15	15	13	14	15	12	11	8	8	12	5	7
종로구 주말 percent	0.588	0.471	0.765	0.588	0.647	0.765	0.471	0.471	0.706	0.647	0.706	0.882	0.882	0.882	0.765	0.824	0.882	0.706	0.647	0.471	0.471	0.706	0.294	0.412

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
노원구 주중 count	4	5	4	6	8	7	3	1	3	7	7	9	9	11	11	9	6	10	7	1	3	5	2	3
노원구 주중 percent	0.211	0.263	0.211	0.316	0.421	0.368	0.158	0.053	0.158	0.368	0.368	0.474	0.474	0.579	0.579	0.474	0.316	0.526	0.368	0.053	0.158	0.263	0.105	0.158
노원구 주말 count	8	8	9	12	3	6	4	8	10	10	6	11	9	10	10	12	12	12	9	7	3	6	4	6
노원구 주말 percent	0.421	0.421	0.474	0.632	0.158	0.316	0.211	0.421	0.526	0.526	0.316	0.579	0.474	0.526	0.526	0.632	0.632	0.632	0.474	0.368	0.158	0.316	0.211	0.316

3RD 분석

2.1 SK유동인구 데이터 - 시간별 유동인구와 미세먼지

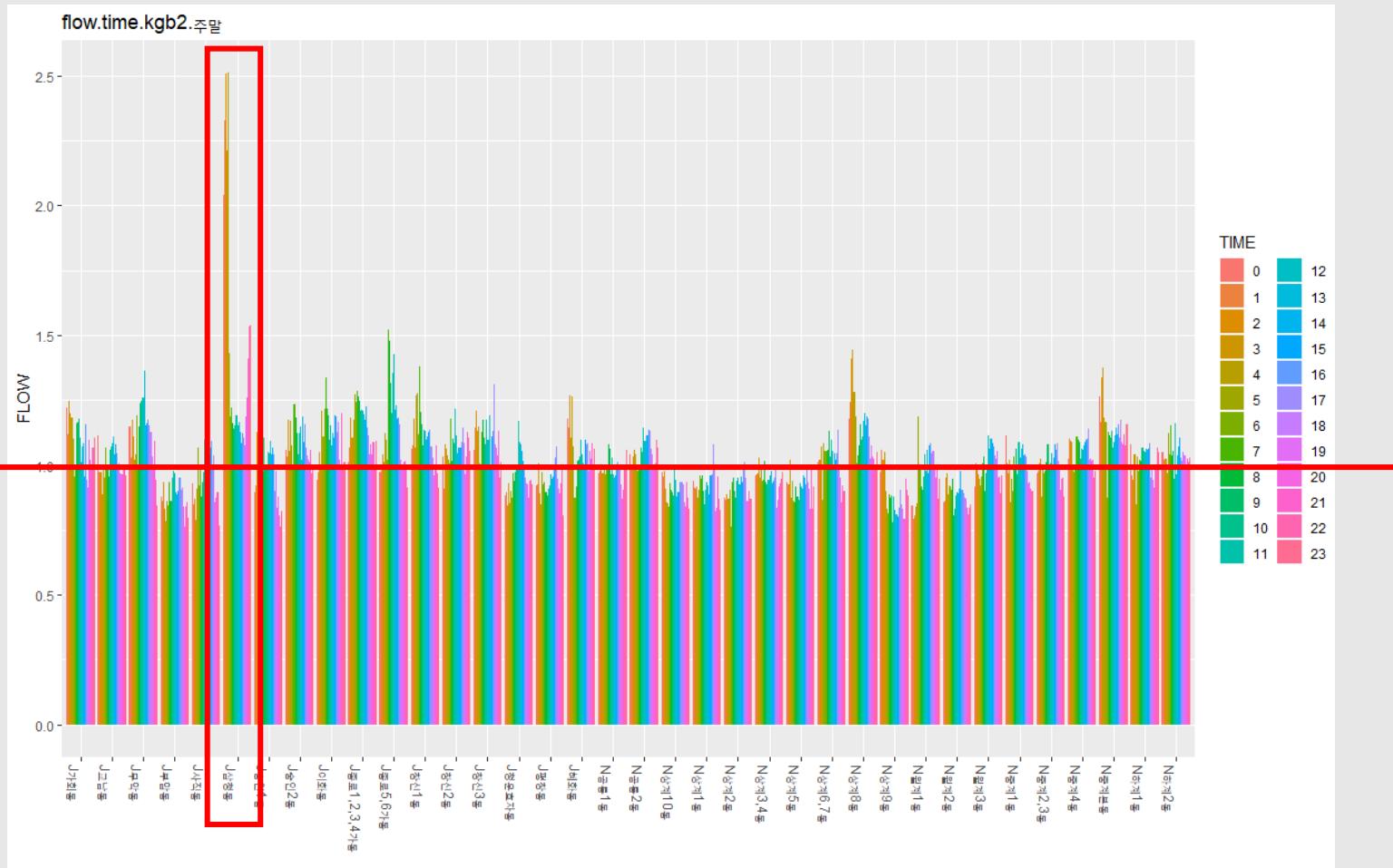
주말 미세먼지가 나쁘면 사람들이 잘 돌아다니지 않을 것이라 생각했으나, 유동인구가 늘어난 이유는 무엇일까?

종로구는 직장지&관광지가 많다는 특징이 있다.

특히 **삼청동**이 다른 구에 비해 **미세먼지가 나쁜 날 유동인구가 늘어난다.**

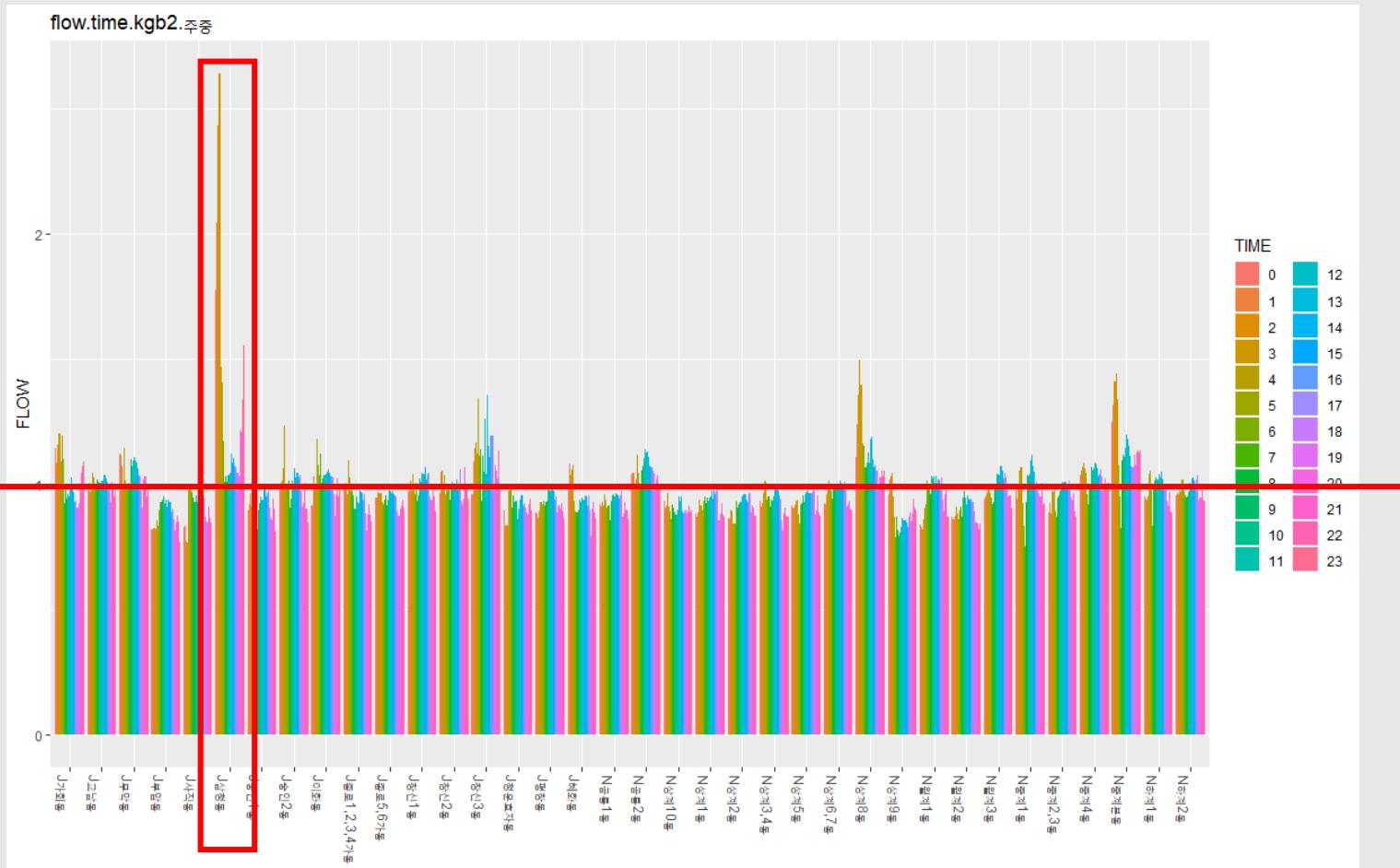
3RD 분석

ratio 1
기준선



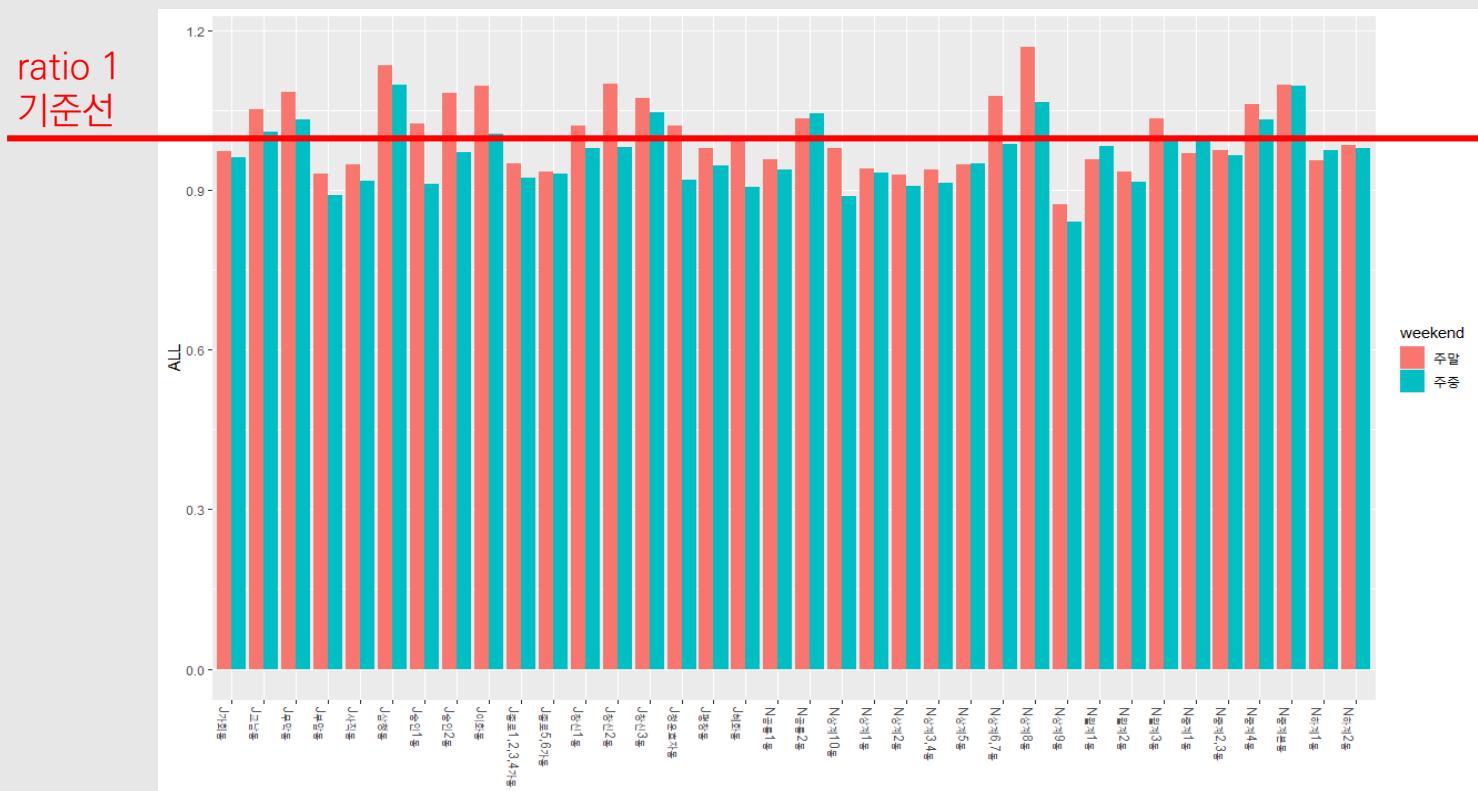
3RD 분석

ratio 1
기준선



3RD 분석

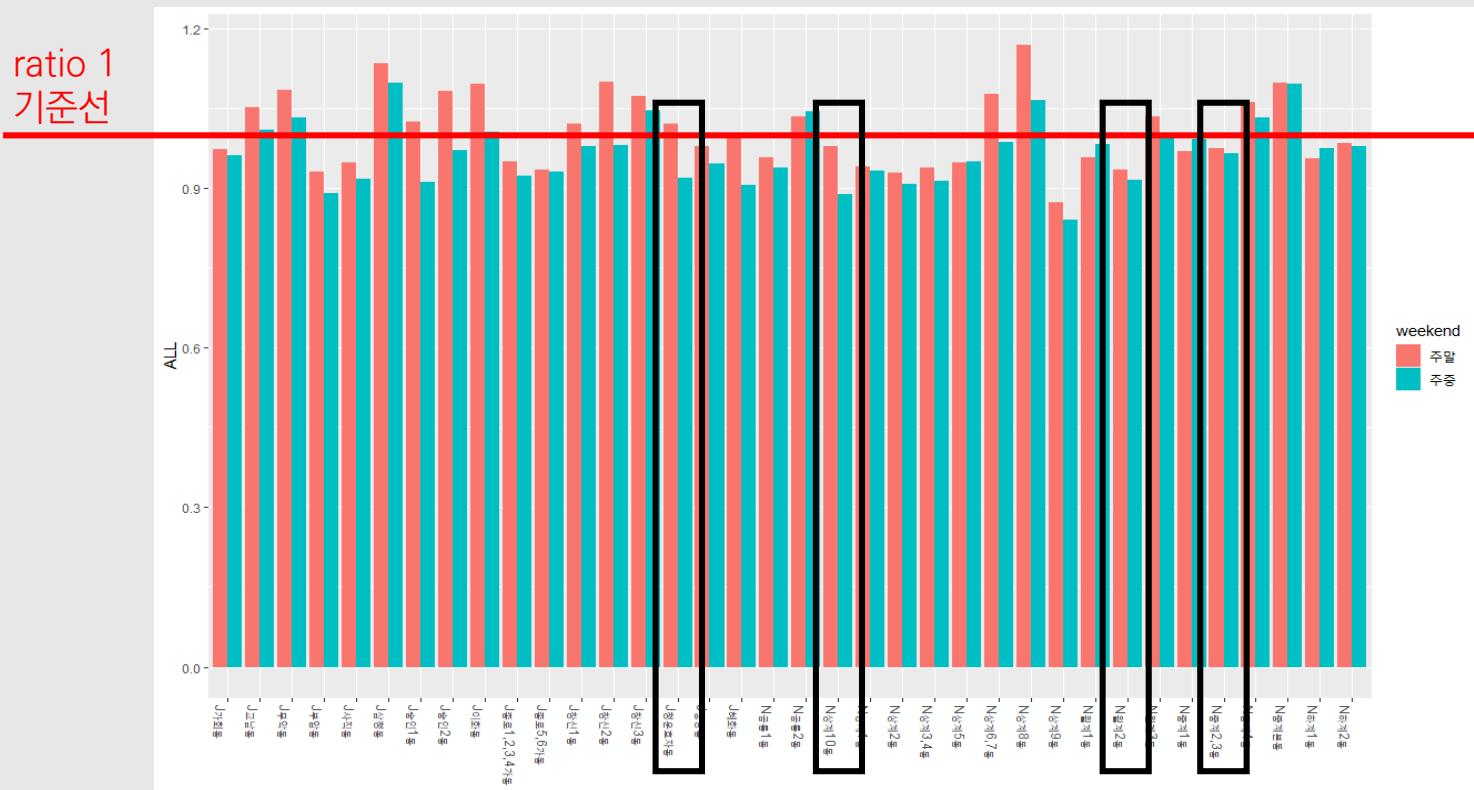
2.2 SK유동인구 데이터 - 일별 유동인구와 미세먼지



3RD 분석

2.2 SK유동인구 데이터 - 일별 유동인구와 미세먼지

GS매출 미세먼지 민감 지역 4개의 행정동(종로구-청운효자동 & 노원구-월계2동/중계2,3동/상계10동)은 유동인구에서도 ratio 지수가 대개 1미만이다. 즉 총 유동인구 중앙값이 미세먼지 나쁠때 더 낮으므로 유동인구에서도 미세먼지 민감 행동을 보인다고 할 수 있다. .



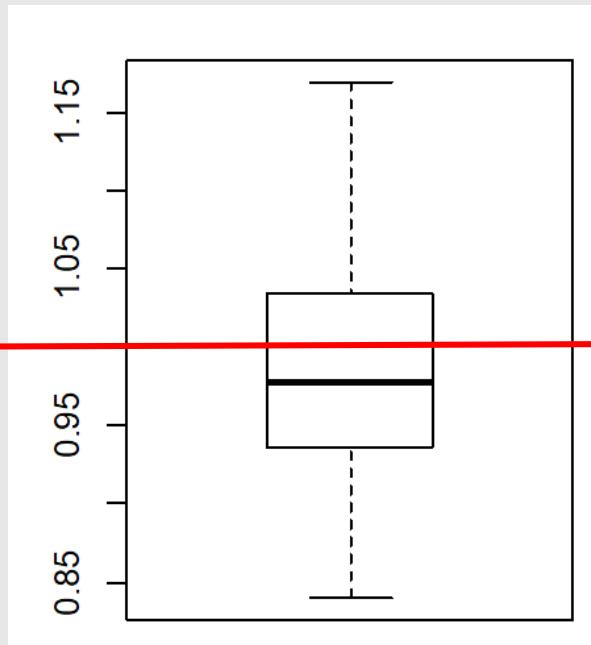
3RD 분석

2.2 SK유동인구 데이터 - 일별 유동인구와 미세먼지

	주말 Ratio	주중 Ratio
종로구+노원구	1.01	0.968
종로구만	1.02	0.967
노원구만	0.991	0.969

ratio 1
기준선

〈36개동 주말+주중 ratio 분포〉



3RD 분석

2.3 SK유동인구 데이터 - 분석 결과

일부 시간대와 일부 행정동을 제외하곤, 유동인구수가 크게 줄어들지 않는다.

그 이유는?

3RD 분석

2.3 SK유동인구 데이터 - 분석 결과

미세먼지 걱정과 우려와 달리,
일부 시간대와 일부 행정동을 제외하면, 유동인구수가 크게 줄어들지 않는다.
미세먼지 유동인구에 변화에 큰 영향X

3RD 분석

2.3 SK유동인구 데이터 - 분석 결과

미세먼지 현황을 일상 곳곳에 알려
경각심을 불러 일으켜야 한다



와이즈넛 SNS 데이터

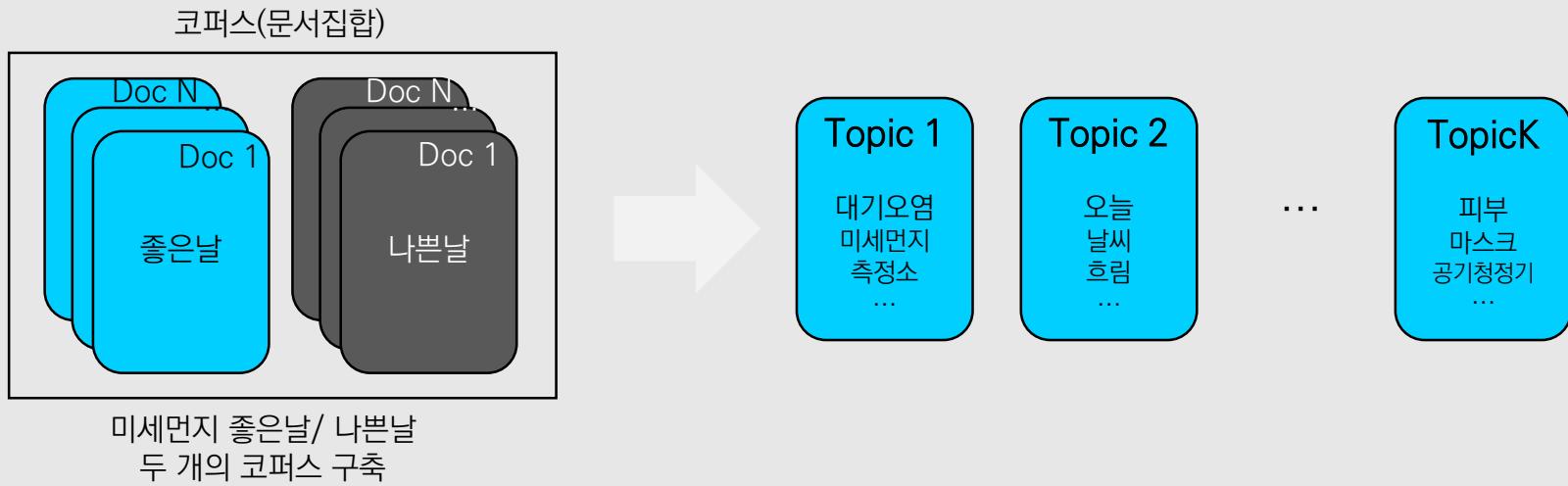
&



케이웨더 환경기상데이터

3RD 분석

3.1 와이즈넛 SNS 데이터 - 토픽모델 : LDA



토픽 모델: 문서 집합인 코퍼스로부터 추상적인 토픽을 찾아내는 모델

디리클레 분포와 깁스 샘플링을 이용한 단어 군집 방법인 **LDA**를 사용

파라미터와 토픽 수 지정 후 연구자가 단어 군집에 대한 토픽 명명

3RD 분석

3.1 와이즈넛 SNS 데이터 - (1) 뉴스 토픽

미세먼지 좋은날

topic	keyword
정부 정책	억원, 정부, 사업, 예산, 정책, 지원, 추진, 원전, 국회, 일자리
미세먼지 대응책	미세먼지, 차량, 경유차, 저감, 사업장, 보급, 자동차, 지원, 설치, 배출가스
선거	후보, 시민, 공약, 추진, 조성, 민선, 사업, 일자리, 시장, 정책
일기예보	기온, 아침, 최고기온, 날씨, 전국, 농도, 오늘, 기상청, 서울, 구름
기술	기술, 서비스, 개발, 에너지, 스마트시티, 시스템, 아파트, 가구, 플랫폼, 데이터
일기예보	기온, 최고기온, 아침, 기상청, 구름, 물결, 제주도, 앞바다, 태풍, 전국
헬스 케어	사람, 마스크, 치료, 라돈, 증상, 미세먼지, 질환, 생각, 문제, 환자
공기청정기 신제품 출시	제품, 피부, 공기청정기, 삼성, 브랜드, 출시, 이벤트, 성분, 가전, 기능

미세먼지 나쁜날 등장 토픽

미세먼지 고농도에 대한 예보
미세먼지 경보에 따른 조치

공통 토픽

일기예보
정부 정책,
헬스 케어 정보,
공기청정기 광고
미세먼지 대응책

미세먼지 나쁜날

topic	keyword
미세먼지 경보 대응	비상저감조치, 발령, 시행, 공공기관, 차량이부제, 수도권, 주의보, 초미세먼지, 차량, 조치
정부 정책	사업, 조성, 억원, 추진, 지원, 계획, 구축, 환경, 나무, 다양한
공기청정기 광고	제품, 피부, 공기청정기, 출시, 상품, 브랜드, 다양한, 판매, 기능, 고객
미세먼지로 인한 경기 취소	미세먼지, 중국, 농도, 경기, 마스크, 초미세먼지, 취소, 황사, 사람, 리그
미세먼지 농도 예보	초미세먼지, 서울, 농도, 등급, 주의보, 오늘, 날씨, 매우나쁨, 기온, 전국
정치	대통령, 정부, 의원, 원전, 문제, 청와대, 문재인, 후보, 정책, 국회
일기예보	기온, 아침, 최고기온, 평년, 물결, 강릉, 제주도, 춘천, 기상청, 전국
미세먼지 대응책	지원, 설치, 저감, 차량, 사업장, 자동차, 보급, 사업, 학교, 계획

3RD 분석

3.1 와이즈넛 SNS 데이터 - (2) 카페 토픽

미세먼지 좋은날

topic	keyword
공기청정기 광고	공기청정기, 시공, 미세먼지, 거실, 설치, 공기, 제품, 청소, 공간, 주방
안부	생각, 아이들, 미세먼지, 사람, 오늘, 날씨, 아이, 사진, 이야기, 작성
판매	택배비, 직거래, 판매, 가능, 이민, 세트, 토플, 구성, 시험, 전기
판매	연석, 장당, 블록, 판매자, 연락처, 등록, 물품, 네이비, 판매, 연락
판매	택배비, 직거래, 판매, 가능, 세트, 구성, 색상, 전기, 제품, 요리
헬스 케어 광고	피부, 매연, 연료, 엔진, 성분, 증상, 제품, 효과, 치료, 미세먼지
오늘 날씨	오늘, 미세먼지, 날씨, 구름, 아침, 하늘, 수치, 전국, 오후, 기온
정부 사업	억원, 정부, 시장, 사업, 중국, 개발, 상승, 논란, 국내, 확대

공통 토픽

소통
광고 및 판매글
가전제품 판매글

미세먼지 나쁜날

topic	keyword
안부	오늘, 아이들, 감사, 생각, 엄마, 친구, 사람, 사진, 친구들, 아이
안부	미세먼지, 오늘, 마스크, 다들, 날씨, 내일, 감기, 주말, 외출, 기침
판매	이벤트, 만원, 상품, 상담, 할인, 가능, 판매, 구매, 삭제, 제품
판매	판매, 가능, 판매자, 물품, 연락처, 등록, 연락, 내용, 등록한, 배송방법
투자	상승, 시장, 정부, 사업, 소식, 규모, 전망, 중국, 투자, 개발
미세먼지 수치	미세먼지, 오늘, 환기, 수치, 내일, 다들, 최악, 창문, 날씨, 주말
미세먼지 대책	미세먼지, 중국, 농도, 문제, 정부, 영향, 초미세먼지, 사람, 예보, 질환
안티더스트 제품 광고	피부, 제품, 시공, 사용, 공기청정기, 거실, 설치, 공간, 성분, 주방

미세먼지 나쁜날 등장 토픽

미세먼지에 대한 대책 촉구
수치에 대한 부정적 반응

3RD 분석

3.1 와이즈넛 SNS 데이터 - (3) 블로그 토픽

미세먼지 좋은날

topic	keyword
미용 광고	가역, 주근깨, 대학생, 축소, 여대생, 눈밑, 심술보, 추출물, 뾰띠, 턱살
방충망 광고	시공, 방충망, 차량, 청소, 아파트, 신축빌라, 공간, 작업, 설치, 거실
정부 정책	정부, 차량, 사업, 미세먼지, 지역, 정책, 지원, 추진, 확대, 중국
일기	오늘 하늘, 장소, 생각, 날씨, 사진, 엄마, 카페, 여행, 친구
공기청정기 광고	공기청정기, 공기, 사용, 제품, 미세먼지, 청소, 청소기, 삼성, 건조기, 마스크
지하철 역명	가역, 시청역, 을지로, 종로, 신설동역, 성수역, 동대문역, 충무로역, 신도림역, 교대역
헬스 케어	증상, 치료, 비염, 질환, 비타민, 미세먼지, 수술, 건강, 면역력, 호흡기
피부제품 광고	피부, 클렌징, 제품, 사용, 성분, 거품, 크림, 세안, 샴푸, 자극

공통 토픽

미세먼지에 대한 개인의 견해
피부 미용 관련 광고 유도글

미세먼지 나쁜날

미세먼지 나쁜날 등장 토픽

좋은 날과 나쁜 날의
토픽 차이 두드러지지 않음

topic	keyword
공기청정기 광고	시공, 공기청정기, 청소, 작업, 설치, 차량, 공기, 사용, 아파트, 고객님
공기청정기 광고	마스크, 제품, 공기청정기, 디자인, 사용, 삼성, 미세먼지, 차량, 구매, 고객님
일기	오늘, 엄마, 생각, 수업, 아빠, 친구, 아침, 하루, 오빠, 학교
피부제품 광고	피부, 클렌징, 제품, 성분, 사용, 세안, 샴푸, 크림, 거품, 수분
여행	여행, 카페, 사진, 숙소, 호텔, 도착, 아이들, 코스, 하늘, 날씨
맛집	맛집, 메뉴, 고기, 주문, 소스, 카페, 삼겹살, 국물, 테이블, 분위기
미용 광고	가역, 시청역, 을지로, 축소, 개선, 시술, 종로, 여대생, 주사, 대학생
차량 보험 광고	차량, 보험, 미세먼지, 가입, 운전, 정부, 지역, 중국, 지원

3RD 분석

3.1 와이즈넛 SNS 데이터 - (3) 블로그 토픽

미세먼지 좋은날

topic	keyword
미용 광고	가역, 주근깨, 대학생, 축소, 여대생, 눈밑, 심술보, 추출물, 뽀띠, 턱살
방충망 광고	시공, 방충망, 차량, 창문, 아파트, 세탁, 물건, 착용, 설치, 기밀
정부 정책	정부, 차량, 사업, 미네랄, 경영, 시장, 구조, 확대, 증권
일기	오늘, 하늘, 장소, 생각, 글씨, 사진, 암미, 카페, 여행, 친구
공기청정기 광고	공기청정기, 공기, 미세먼지, 환기, 청소기, 바이러스, 마스크
지하철 역명	가역, 시청역, 을지로, 동작, 지설동역, 경수역, 동대문역, 충무로역

공통 토픽

미세먼지 나쁜 날, 블로그 : 광고성 토픽

뉴스와 카페 : 미세먼지 대책 및 부정적 반응

미세먼지 나쁜날 등장 토픽

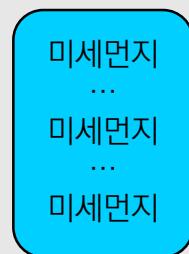
좋은 날과 나쁜 날의 토픽 차이 두드러지지 않음

topic	keyword
공기청정기 광고	시공, 공기청정기, 청소, 작업, 설치, 차량, 공기, 사용, 아파트, 고객님
공기청정기 광고	마스크, 제품, 공기청정기, 디자인, 사용, 삼성, 미세먼지, 차량, 구매, 고객님
일기	오늘, 엄마, 생각, 수업, 아빠, 친구, 아침, 하루, 오빠, 학교
피부제품 광고	피부, 클렌징, 제품, 성분, 사용, 세안, 샴푸, 크림, 거품, 수분
여행	여행, 카페, 사진, 숙소, 호텔, 도착, 아이들, 코스, 하늘, 날씨
맛집	맛집, 메뉴, 고기, 주문, 소스, 카페, 삼겹살, 국물, 테이블, 분위기
미용 광고	가역, 시청역, 을지로, 축소, 개선, 시술, 종로, 여대생, 주사, 대학생
차량 보험 광고	차량, 보험, 미세먼지, 가입, 운전, 정부, 지역, 중국, 지원

3RD 분석

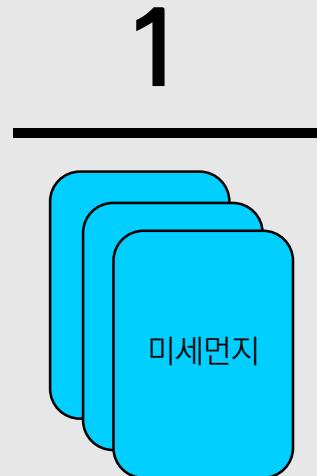
3.2 와이즈넛 SNS 데이터 - 단어 빈도표 TF-IDF 가중치

TF-IDF =



TF : Term Frequency
특정 단어가 문서내에 나타나는 빈도수

X



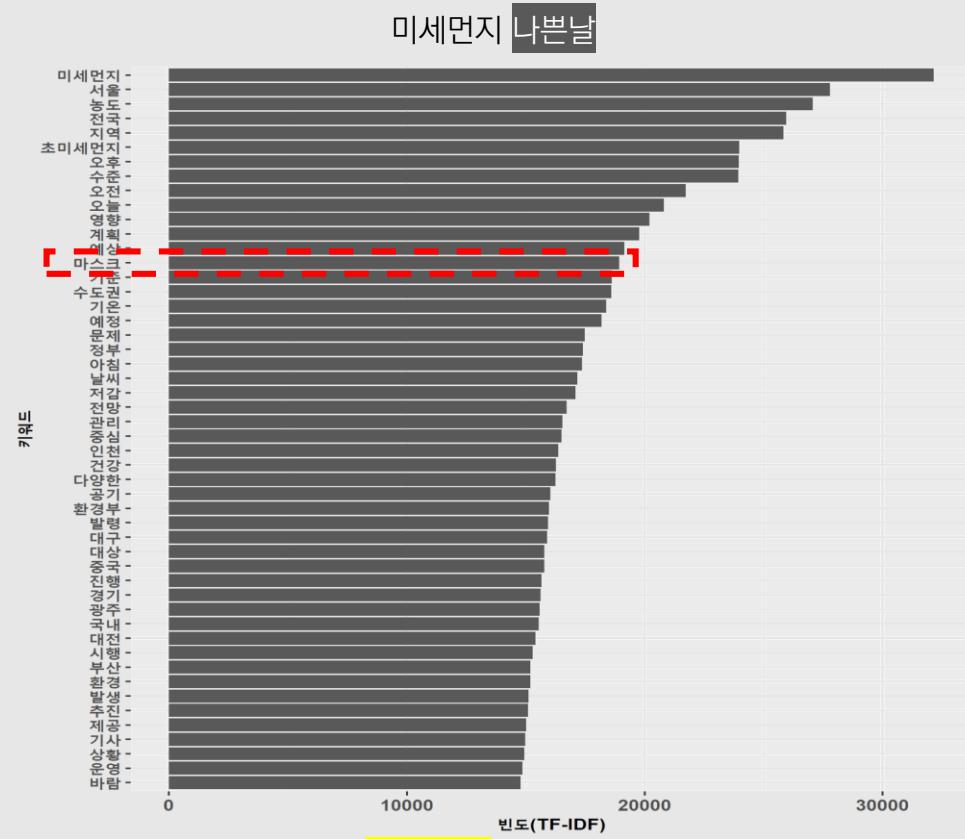
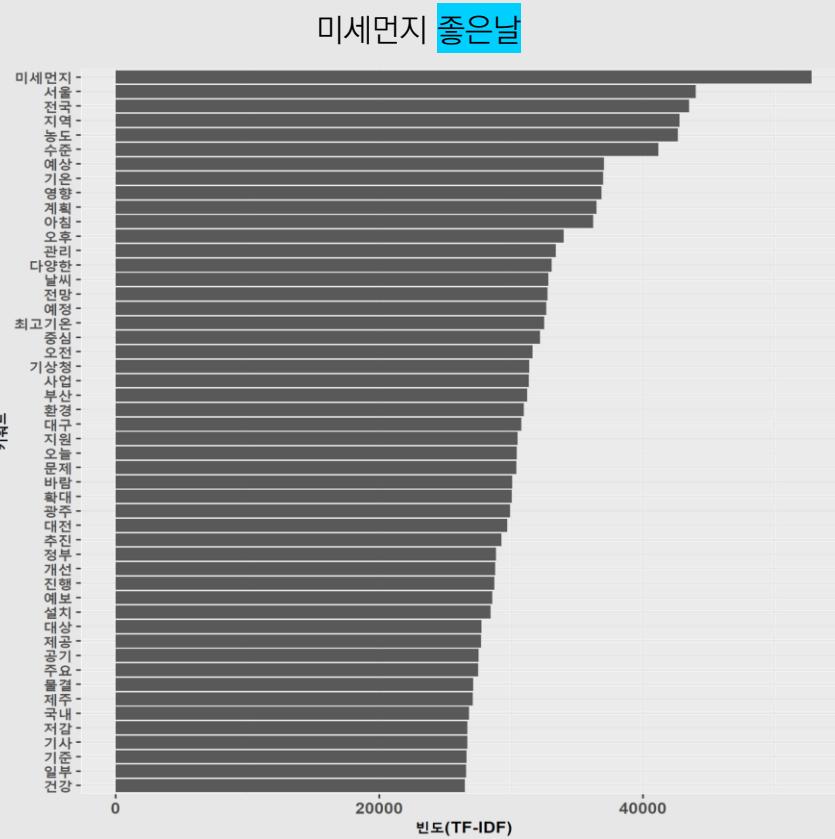
IDF : Inverse Document Frequency
단어가 나타난 문서의 빈도수(DF)의 역수

TF-IDF는 여러 문서에 빈출하는 단어보다
특정 문서에서 많이 사용된 단어에 가중치를 부여하는 방법

예) 미세먼지가 포함된 문서에서 “미세먼지”라는 단어에 페널티를 부여

3RD 분석

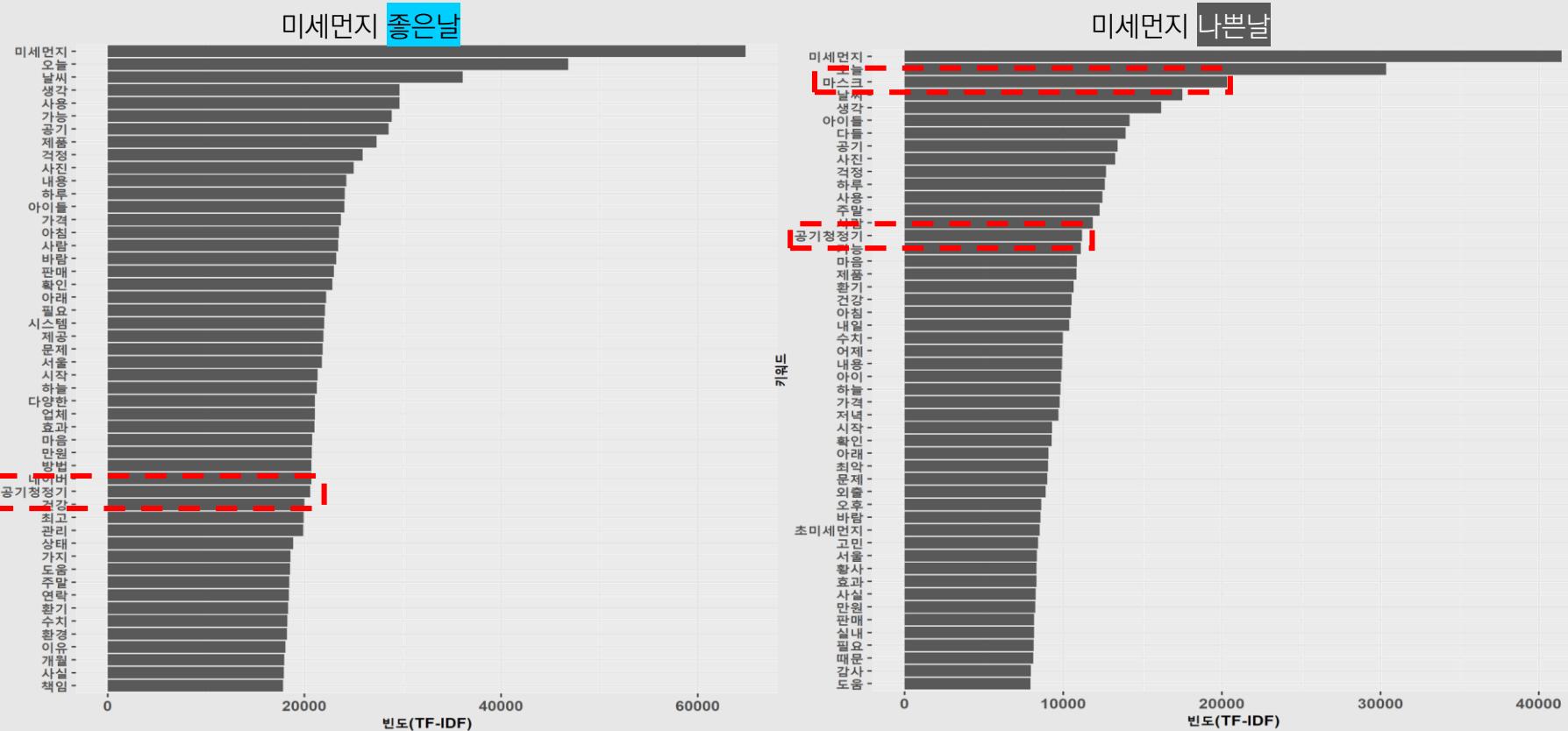
3.2 와이즈넛 SNS 데이터 - (1) 뉴스 단어 빈도표



미세먼지 나쁜 날 사람들의 대처 방안인 **마스크** 등장

3RD 분석

3.2 와이즈넛 SNS 데이터 - (2) 카페 단어 빈도표



미세먼지 좋은날 나쁜날 미세먼지 수동적인 대처 방안이 되는 공기청정기와 마스크

3RD 분석

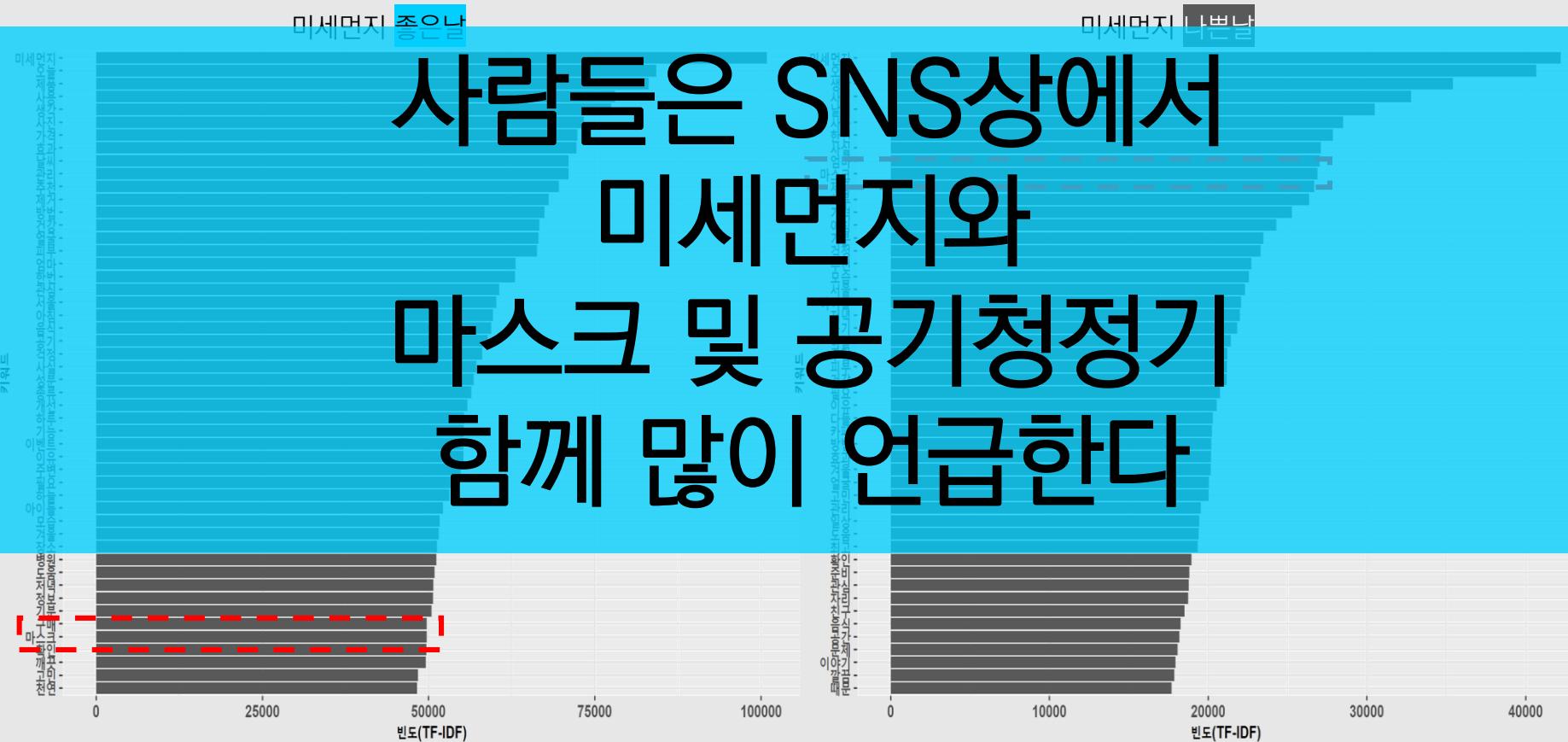
3.2 와이즈넛 SNS 데이터 - (3) 블로그 단어 빈도표



미세먼지 좋은 날 나쁜 날에 **마스크** 상위 키워드로 등장

3RD 분석

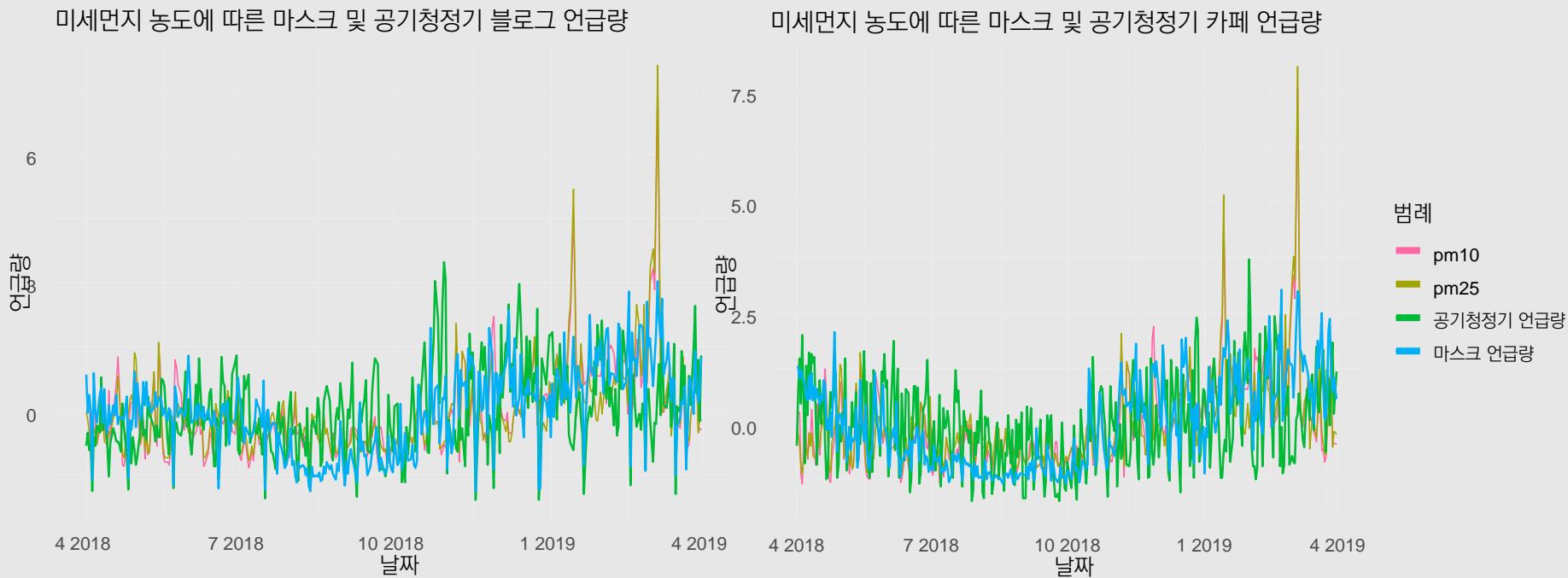
3.2 와이즈넛 SNS 데이터 - (3) 블로그 단어 빈도표



미세먼지 좋은 날 나쁜 날에 **마스크**만 빈출단어로 등장

3RD 분석

3.2 와이즈넛 SNS 데이터 - 미세먼지 농도와 마스크 및 공기청정기 SNS 언급량



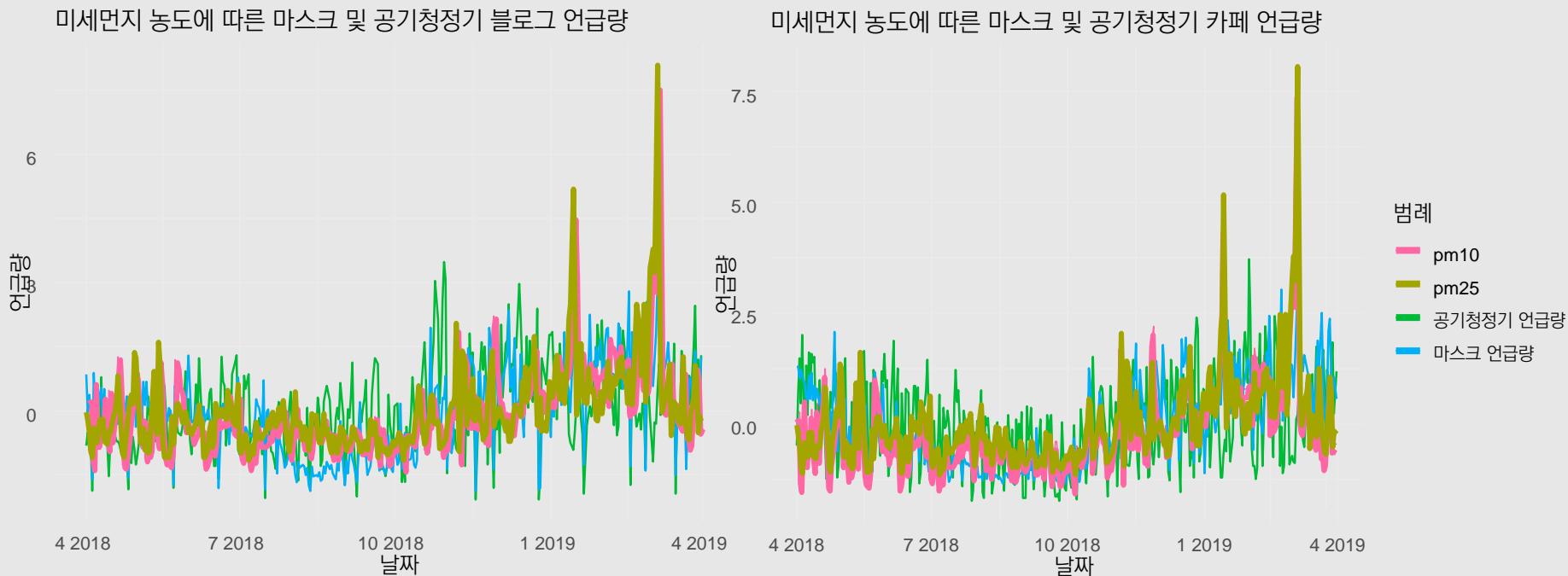
미세먼지 농도와 공기청정기 및 마스크의 sns 데이터 언급량 (표준화, 좌: 블로그 우: 카페)

미세먼지 농도 증가와 함께 공기청정기와 마스크 모두 언급량이 증가

미세먼지가 심한 날 사람들은 1차적으로 미세먼지 차단을 위해 공기청정기와 마스크를 찾는다.

3RD 분석

3.2 와이즈넛 SNS 데이터 - 미세먼지 농도와 마스크 및 공기청정기 SNS 언급량



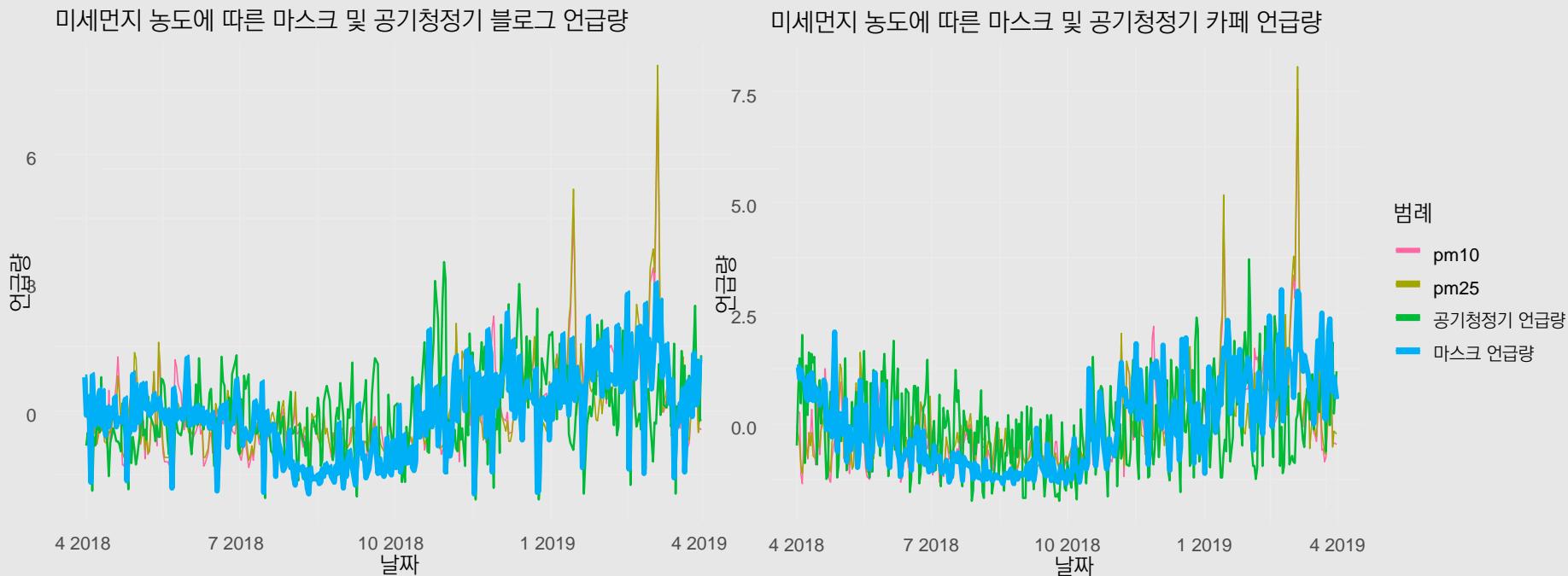
미세먼지 농도와 공기청정기 및 마스크의 sns 데이터 언급량 (표준화, 좌: 블로그 우: 카페)

미세먼지 농도 증가와 함께 공기청정기와 마스크 모두 언급량이 증가

미세먼지가 심한 날 사람들은 1차적으로 미세먼지 차단을 위해 공기청정기와 마스크를 찾는다.

3RD 분석

3.2 와이즈넛 SNS 데이터 - 미세먼지 농도와 마스크 및 공기청정기 SNS 언급량



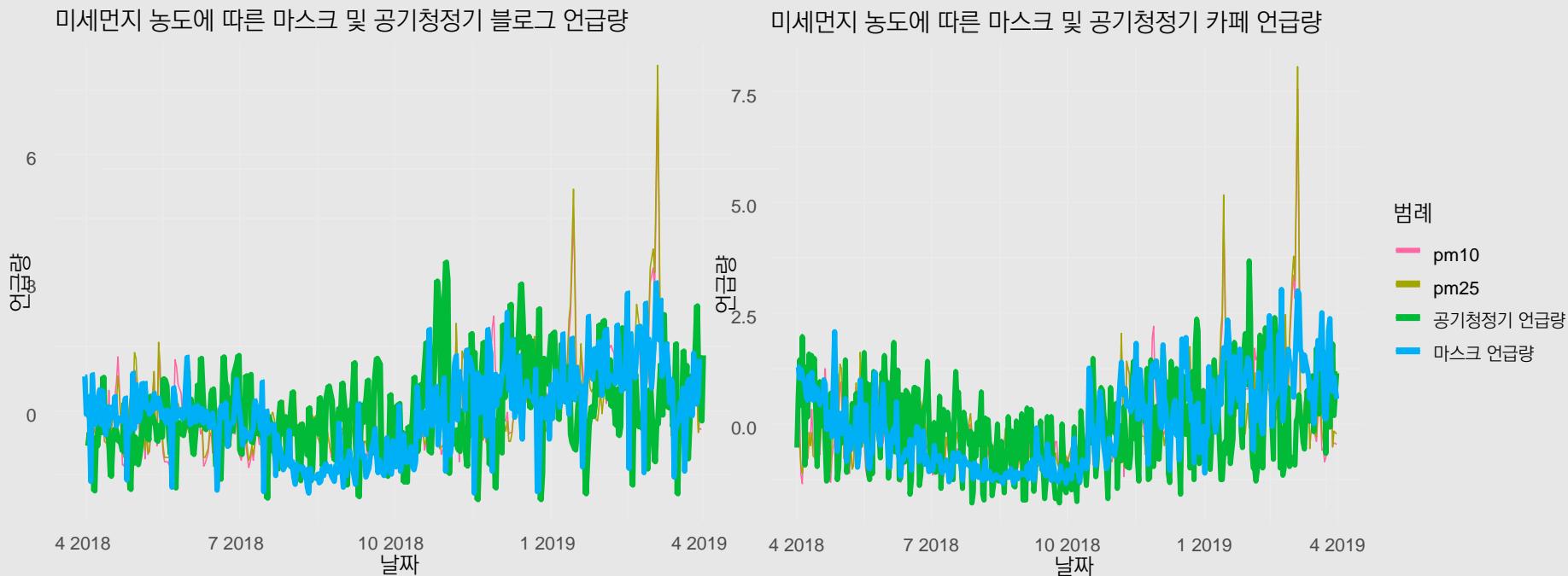
미세먼지 농도와 공기청정기 및 마스크의 sns 데이터 언급량 (표준화, 좌: 블로그 우: 카페)

미세먼지 농도 증가와 함께 공기청정기와 마스크 모두 언급량이 증가

미세먼지가 심한 날 사람들은 1차적으로 미세먼지 차단을 위해 공기청정기와 마스크를 찾는다.

3RD 분석

3.2 와이즈넛 SNS 데이터 - 미세먼지 농도와 마스크 및 공기청정기 SNS 언급량



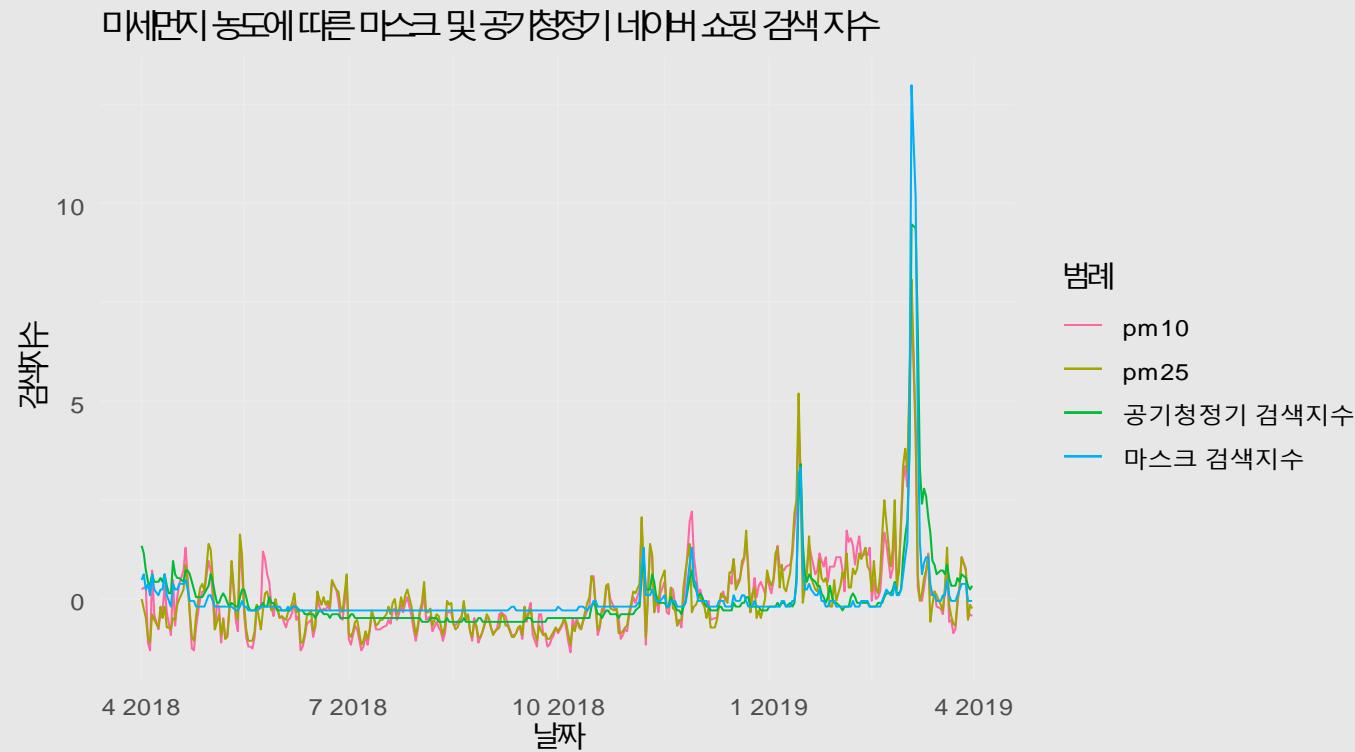
미세먼지 농도와 공기청정기 및 마스크의 sns 데이터 언급량 (표준화, 좌: 블로그 우: 카페)

미세먼지 농도 증가와 함께 공기청정기와 마스크 모두 언급량이 증가

미세먼지가 심한 날 사람들은 1차적으로 미세먼지 차단을 위해 공기청정기와 마스크를 찾는다.

3RD 분석

3.3 네이버 쇼핑 검색 지수 - 미세먼지 농도와 마스크 및 공기청정기 쇼핑 검색 지수

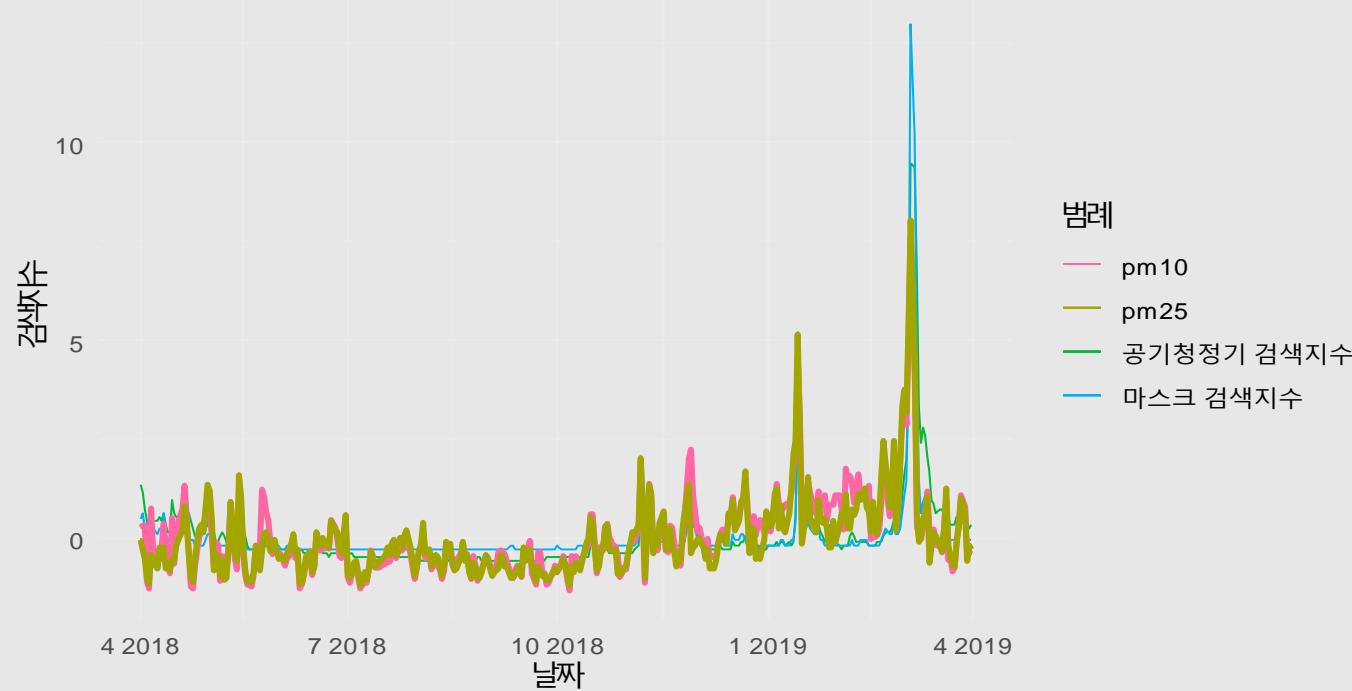


2018.04 ~ 2019.03 미세먼지 농도와 공기청정기 및 마스크 네이버 검색지수(표준화)
사람들이 미세먼지에 대응하기 위한 대표적인 방안 공기청정기와 마스크
미세먼지가 높아짐에 따라 공기청정기와 마스크의 쇼핑 검색어 지수가 비례하게 높아지고 있다.

3RD 분석

5. 미세먼지 농도에 따른 공기청정기 및 마스크의 쇼핑 검색어 트렌드

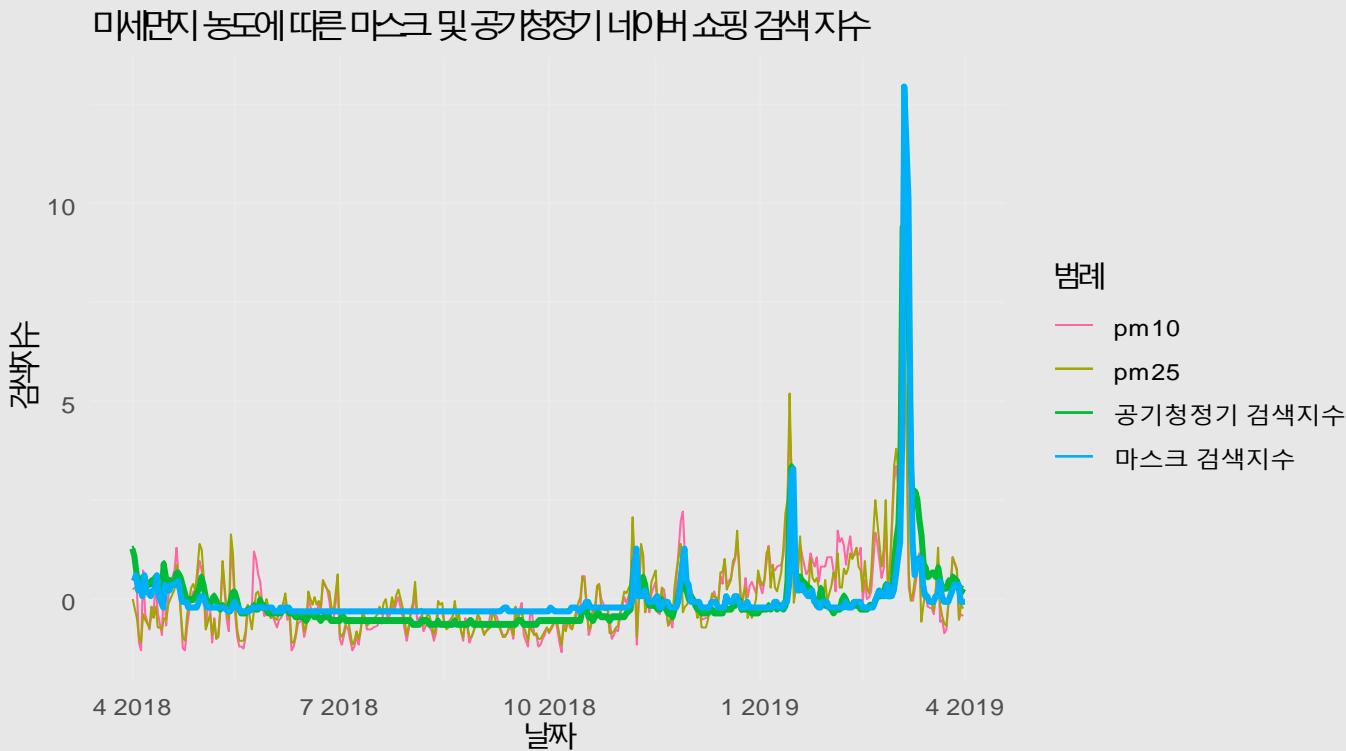
미세먼지 농도에 따른 마스크 및 공기청정기 네이버 쇼핑 검색지수



2018.04 ~ 2019.03 미세먼지 농도와 공기청정기 및 마스크 네이버 검색지수(표준화)
사람들이 미세먼지에 대응하기 위한 대표적인 방안 공기청정기와 마스크
미세먼지가 높아짐에 따라 공기청정기와 마스크의 쇼핑 검색어 지수가 비례하게 높아지고 있다.

3RD 분석

3.3 네이버 쇼핑 검색 지수 - 미세먼지 농도와 마스크 및 공기청정기 쇼핑 검색 지수



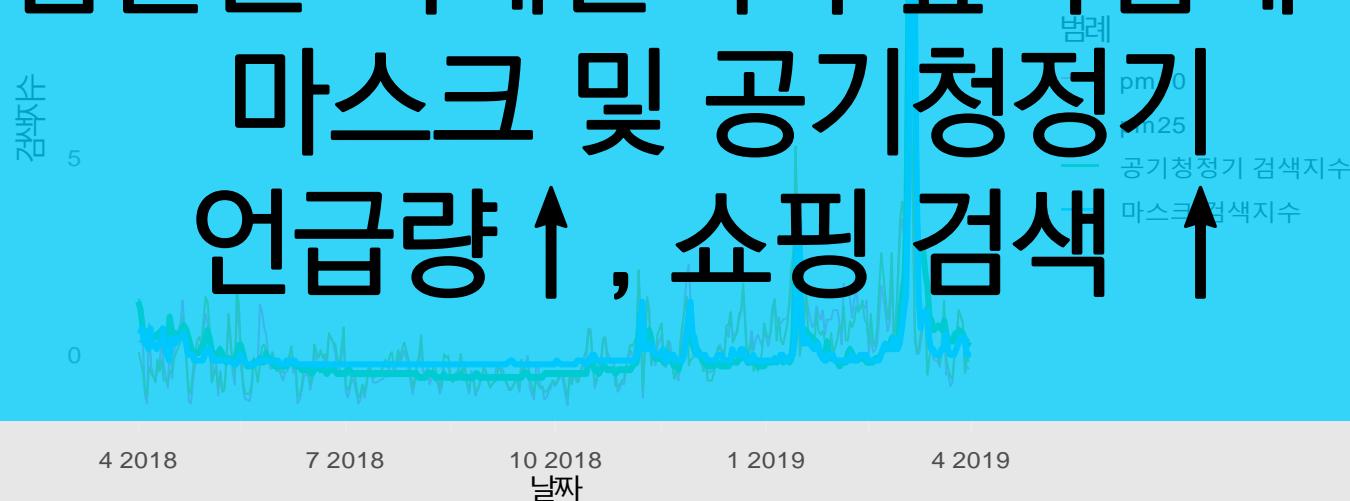
2018.04 ~ 2019.03 미세먼지 농도와 공기청정기 및 마스크 네이버 검색지수(표준화)
사람들이 미세먼지에 대응하기 위한 대표적인 방안 공기청정기와 마스크
미세먼지가 높아짐에 따라 공기청정기와 마스크의 쇼핑 검색어 지수가 비례하게 높아지고 있다.

3RD 분석

3.3 네이버 쇼핑 검색 지수 - 미세먼지 농도와 마스크 및 공기청정기 쇼핑 검색 지수

미세먼지 농도에 따른 마스크 및 공기청정기 네이버 쇼핑 검색 지수

사람들은 미세먼지가 높아짐에 따라
마스크 및 공기청정기
언급량 ↑, 쇼핑 검색 ↑



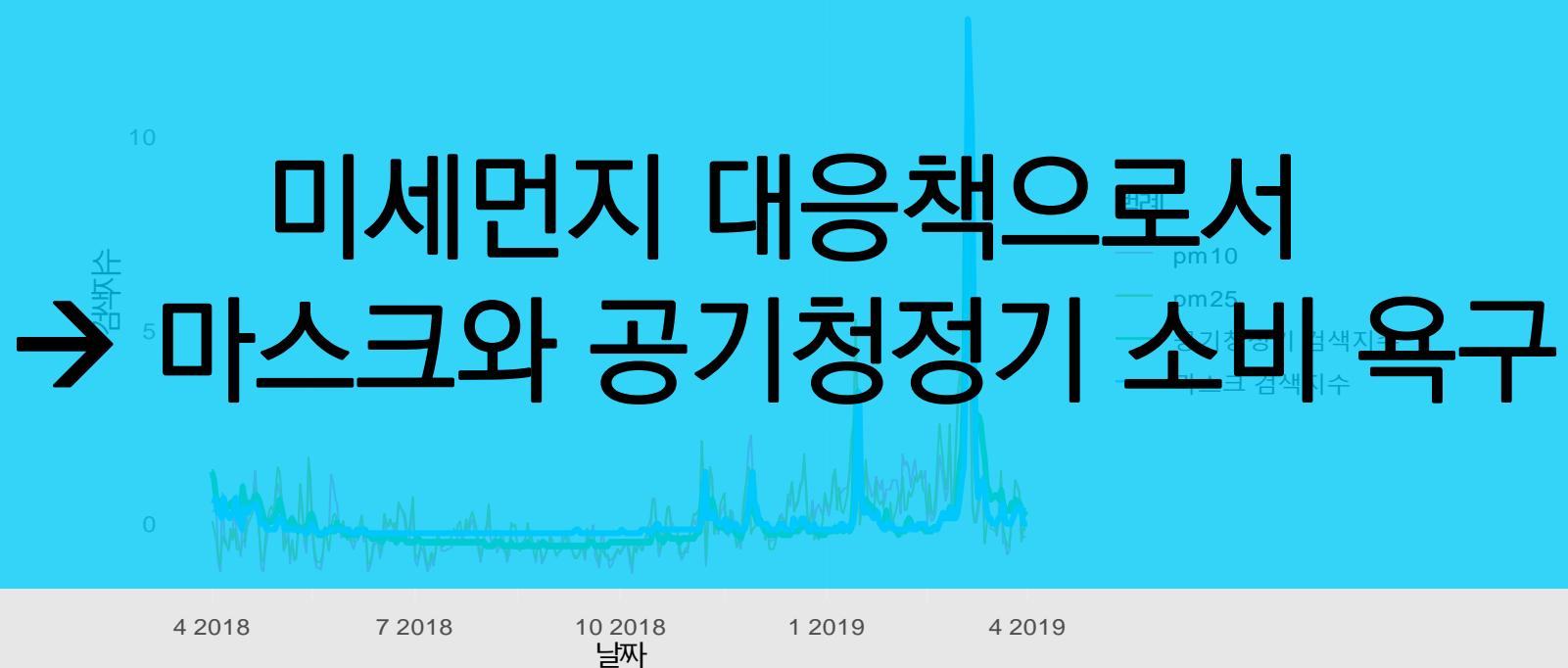
2018.04 ~ 2019.03 미세먼지 농도와 공기청정기 및 마스크 네이버 검색지수(표준화)

사람들이 미세먼지에 대응하기 위한 대표적인 방안 공기청정기와 마스크
미세먼지가 높아짐에 따라 공기청정기와 마스크의 쇼핑 검색어 지수가 비례하게 높아지고 있다.

3RD 분석

3.3 네이버 쇼핑 검색 지수 - 미세먼지 농도와 마스크 및 공기청정기 쇼핑 검색 지수

미세먼지 농도에 따른 마스크 및 공기청정기 네이버 쇼핑 검색 지수



2018.04 ~ 2019.03 미세먼지 농도와 공기청정기 및 마스크 네이버 검색지수(표준화)

사람들이 미세먼지에 대응하기 위한 대표적인 방안 공기청정기와 마스크
미세먼지가 높아짐에 따라 공기청정기와 마스크의 쇼핑 검색어 지수가 비례하게 높아지고 있다.

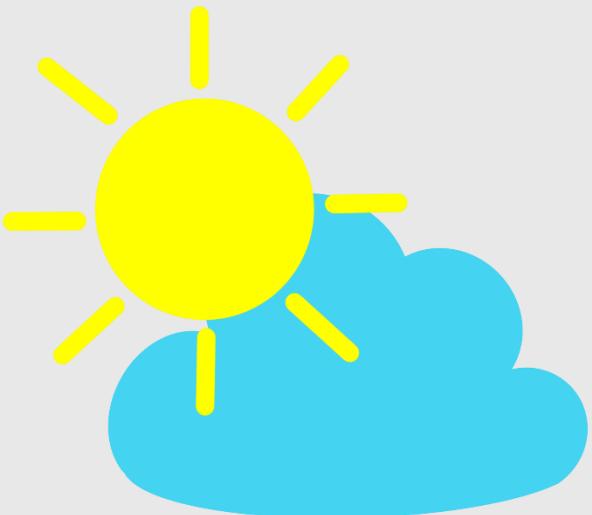
4

비즈니스
아이디어

4TH 비즈니스 아이디어

GS25가 추구하는 Fresh 이미지를 날씨로 확장하자!

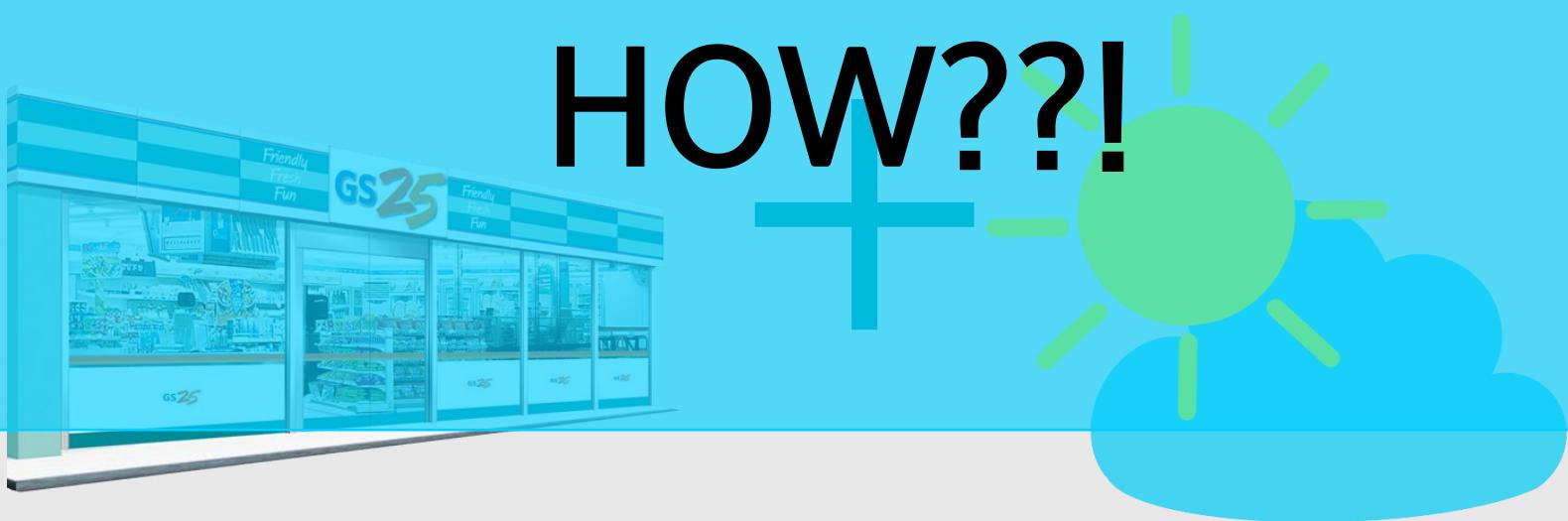
GS25의 Fresh한 이미지를 날씨와 연결! 미세먼지 현황을 알리고 Fresh 추구의 Care를 위한 소비를 유도!



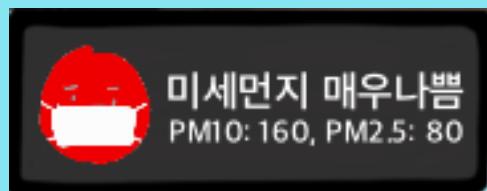
4TH 비즈니스 아이디어

GS25가 추구하는 Fresh 이미지를 날씨로 확장하자!

GS25의 Fresh한 이미지를 날씨와 연결! 미세먼지 현황을 알리고 Fresh 추구의 Care를 위한 소비를 유도!



4TH 비즈니스 아이디어



전광판

4TH 비즈니스 아이디어



미세먼지 매우나쁨
PM10: 160, PM2.5: 80

전광판



생수 야외
판매대

4TH 비즈니스 아이디어



미세먼지 매우나쁨
PM10: 160, PM2.5: 80

전광판



생수 야외
판매대



마스크 야외
판매대

전광판

좋음, 보통, 나쁨, 매우나쁨에
따라 하늘색, 초록색, 주황색,
빨강색으로 표시

- 색상을 이용해 직관적으로
미세먼지에 대한 정보를
빠르게 전달
- 사람들의 즉각적 행동
변화 유도



음료 야외 매대

분석결과, GS는 미세먼지가 나쁜 날 음료
매출이 줄어들었다

이는 미세먼지가 심할 땐 물과 차를
마셔야 한다는 상식에 대치되는 현상!

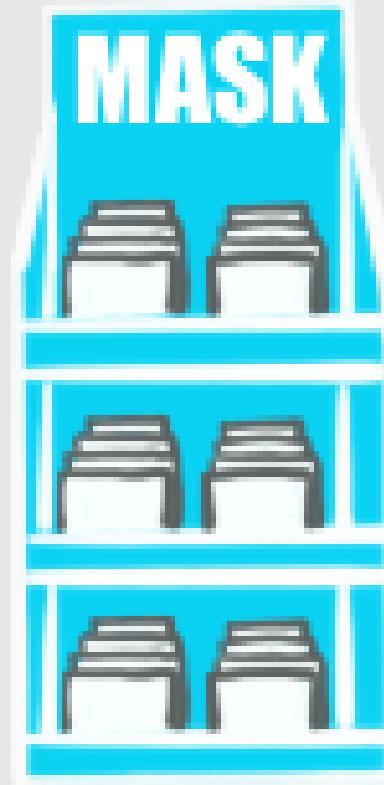
→ 미세먼지가 나쁜 날, 전광판 아래에
생수와 차 야외 매대를 비치해서 음료
판매를 촉진!



마스크 야외 매대

카페와 블로그의 미세먼지 농도에 따른
마스크 언급량 추이가 급격히 높아지는
것으로 보아 미세먼지가 나쁜 날 사람들은
마스크를 1차적인 해결책으로 찾는다는
분석 결과

→ 미세먼지가 나쁜 날, 전광판 아래에서
음료와 함께 마스크를 비치해서 판매 촉진!



즉...

〈미세먼지 나쁜 날〉



〈미세먼지 나쁜 날〉

미세먼지에 별다른 대응없이 걷는 사람들…



〈미세먼지 나쁜 날〉

미세먼지에 대응이 필요한 상황!



〈미세먼지 나쁜 날〉

전광판을 설치해 지나가는 사람들에게
미세먼지가 나쁨을 알리고



〈미세먼지 나쁜 날〉

음료와 마스크 야외 매대를 놓아
미세먼지에 대한 경각심을
바로 구매행동으로 바로 유도하자!



〈미세먼지 나쁜 날〉

내 몸에 필요한 정보를 실시간으로 전달!
지금 당장 필요한 소비 유도!



〈미세먼지 나쁜 날〉

GOOD!



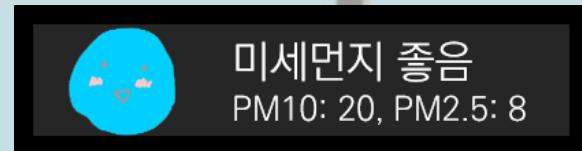
FRESH GS25!

GS25

전광판 & 음료매대 & 마스크 야외매대는

4개의 동(미세먼지 민감 지역): 종로구-청운효자동, 노원구-월계2동/중계2,3동/상계10동

시범 설치 권장!



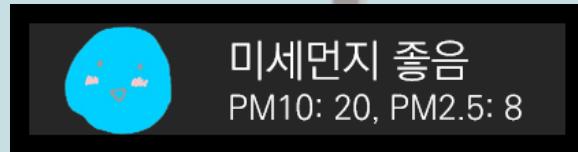
FRESH GS25!

GS25

전광판 & 음료매대 & 마스크 야외매대는

4개의 동(미세먼지 민감 지역): 종로구-청운효자동, 노원구-월계2동/중계2,3동/상계10동

시범 설치 권장!



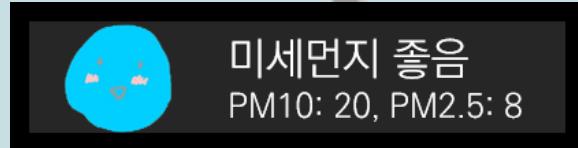
FRESH GS25!

GS25

전광판 & 음료매대 & 마스크 야외매대는

4개의 동(미세먼지 민감 지역): 종로구-청운효자동, 노원구-월계2동/중계2,3동/상계10동

시범 설치 권장!



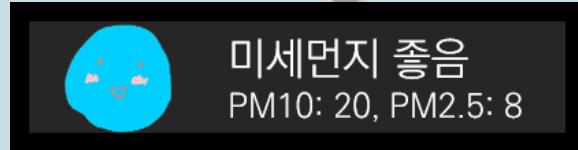
FRESH GS25!

GS25

전광판 & 음료매대 & 마스크 야외매대는

4개의 동(미세먼지 민감 지역): 종로구-청운효자동, 노원구-월계2동/중계2,3동/상계10동

시범 설치 권장!





Thank you

