

2019 빅 콘테스트 공모전 Innovation 분야 출품작

서울특별시 '청정 버스 노선 및 정류소' 사업 제안

- 빅데이터를 활용한 미세먼지로 인한 서울 시민의 사회, 경제적 행동 변화 및 비즈니스 아이디어 도출

CONTENT

1. 사업 주제 선정 배경
2. 연구 가설 및 데이터 분석을 통한 증명
3. 비즈니스 아이템 소개
4. 사업지 선정 과정
 - 4.1 종로구 사업지 선정
 - 4.2 노원구 사업지 선정
5. 한계점 및 향후 계획
6. 팀원 소개

1. 사업 주제 선정 배경

미세먼지와 서울 시민의 버스 이용



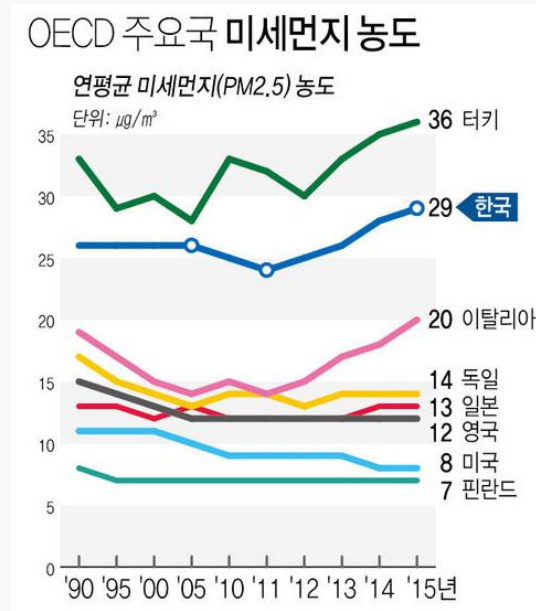
01 사업 주제 선정 배경

대한민국 미세먼지 현황 및 원인 통계

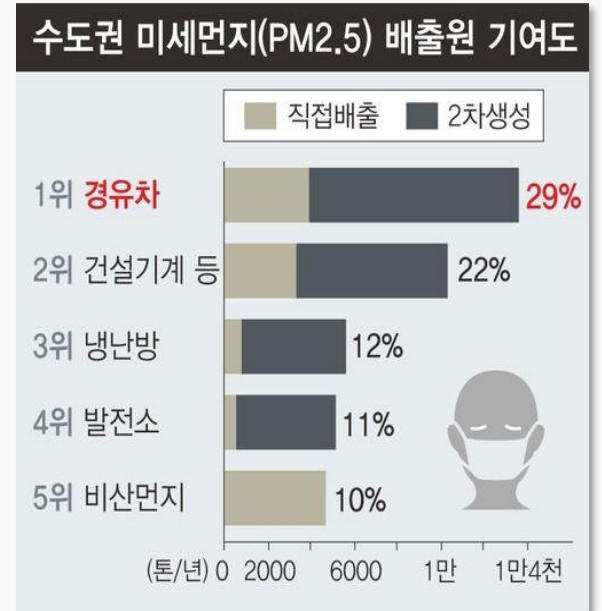
대한민국의 미세먼지 현황과 수도권 미세먼지 원인

지금 대한민국은?

대한민국은 현재 OECD 주요국 중 미세먼지 농도 2위로 공기 오염 국가의 오명을 안고 있습니다. 그 중 수도권 미세먼지의 배출원의 주요 원인은 경유차로, 도로의 승용차 통행을 줄이는 것이 미세먼지 대책의 주요 사안이라고 할 수 있습니다.



출처 연합뉴스



출처 arcgis

01 사업 주제 선정 배경

서울특별시 대중교통 이용 현황

그럼에도 불구하고 ...

서울시의 승용차 등록 대수는 해를 거듭하며 늘어나고 있는 추세지만, 그와 대비해 대중교통 이용객은 줄어들고 있습니다.

서울시 대중교통 이용현황 및 버스이용객 변화



출처 서울신문

01 사업 주제 선정 배경

서울 시민의 든든한 벗, 버스



서울시의 버스

버스는, 서울의 구석 구석을 훑으며 시민의 생활에 가장 밀접하게 맞닿아 있는 대중교통 수단입니다.

따라서, 저희는 버스 이용과 관련된 서울 시민의 사회, 경제적 행동과 미세 먼지 사이의 관계성에 주목하였습니다.

더 나아가, 그로부터 도출할 수 있는 비즈니스 인사이트를 고민해보기로 하였습니다.

2. 연구 가설 및 증명

빅데이터 분석을 통한 미세먼지로 인한
서울 시민의 행동 변화 도출



02

연구 가설

연구 가설 1 : 평일 미세먼지와 유동인구의 상관성



평일, 서울 시민들은 정말 미세먼지가 많은 날에 외출을 덜 할까?

연구 가설

서울 시민의 평일 유동인구량과 미세먼지와의 상관관계는 낮을 것이다.

분석 방법

활용 데이터	미세먼지 데이터, 유동인구 데이터
분석 언어	Python, R
활용 Tool	Jupyter Notebook, Pycharm, Rstudio

02 연구 가설

연구 가설 1 : 분석 데이터 소개

DATA 1

성,연령별 유동 인구 데이터 (노원구, 종로구)

- 출처 : 대회 제공 데이터
- 수집 및 활용 목적 : 평일 서울 시내의 유동인구 추이 파악
- 기간 범위 : 20180401~20190331
- 지역 범위 : 서울특별시 노원구, 종로구
- 변수 범위 : Date(날짜), Dong(법정동), 그 외 성과 나이별 유동인구

	A	B	C	D	E	F	G	H	I	J
1	STD_YM STD_YMD HDONG_CD HDONG_NM MAN_FLOW_POP_CNT_0004 MAN_FLOW_POP_CNT_0509 MA									
2	201804 20180401 1111051500 청운효자동 0.05 78.93 403.26 642.66 984.24 1792.3 1936.1 2060.54 1787.98 236									
3	201804 20180401 1111053000 사직동 1.25 262.54 1207.22 2177.88 4076.89 6887.72 6786.94 6873.7 5920.97 7									
4	201804 20180401 1111054000 삼청동 0.78 255.24 331.09 654.32 1264.86 1375.14 1445.4 1237.07 1658.35 1									
5	201804 20180401 1111055000 부암동 0.228.5 753.31 981.54 1615.77 2732 3390.58 4288.35 3857.72 4912.69 4									
6	201804 20180401 1111056000 평창동 0.354.38 1500.63 1963.08 2593.07 4927.24 5718.38 6997.92 6618.99 906									
7	201804 20180401 1111057000 무악동 0.18 39.69 204.43 377.5 509.92 766.21 786.83 868.99 832.5 1077.11 105									
8	201804 20180401 1111058000 교남동 0.1 71.96 296.03 424.87 644.91 1135.05 1332.61 1431.27 1346.62 1758.0									
9	201804 20180401 1111060000 가회동 0.37.87 196.3 535.94 472.73 893.54 937.81 1051.83 856.7 1095.21 952.3									
10	201804 20180401 1111061500 종로1.2.3.4가동 0.513.62 2338.1 5480.67 13268.22 22024.63 19405.86 17433.96									
11	201804 20180401 1111063000 종로5.6가동 0.110.6 521.4 1783.97 4146.69 5968.6 4939.16 4641.93 3904.82 47									
12	201804 20180401 1111064000 이화동 0.105.25 436.49 1302.85 3341.51 4938.03 3803.46 3105.04 2513.91 2978									
13	201804 20180401 1111065000 혜화동 0.165.65 718.89 1741.96 3934.77 5789.92 4392.22 4271.59 3801.47 4726									
14	201804 20180401 1111067000 창신1동 0.80.96 315.65 805.39 1885.79 2591.53 2233.44 2272 1882.6 2393.21 2									

DATA 2

미세먼지 데이터 (노원구, 종로구)

- 출처 : 대회 제공 데이터
- 수집 및 활용 목적 : 평일 서울 시내의 미세먼지 추이
- 기간 범위 : 20180401~20190331
- 지역 범위 : 서울특별시 노원구, 종로구
- 변수 범위 : tm(데이터측정날짜시간), pm10(미세먼지 pm-10)

	A	B	C	D	E	F	G	H	I	J
1	tm	serial	flag	pm10	co2	vocs	noise	temp	humid	pm25
2	2.02E+11	V10O1610	1	33	-9999	-9999	50	15.3	52	16
3	2.02E+11	V10O1610	1	33	-9999	-9999	50	15.3	52	16
4	2.02E+11	V10O1610	1	28	-9999	-9999	49	15.3	52	14
5	2.02E+11	V10O1610	1	33	-9999	-9999	49	15.3	52	16
6	2.02E+11	V10O1610	1	29	-9999	-9999	49	15.3	52	14
7	2.02E+11	V10O1610	1	36	-9999	-9999	50	15.3	52	18
8	2.02E+11	V10O1610	1	38	-9999	-9999	49	15.2	52	19
9	2.02E+11	V10O1610	1	33	-9999	-9999	49	15.3	53	16
10	2.02E+11	V10O1610	1	34	-9999	-9999	49	15.2	53	17
11	2.02E+11	V10O1610	1	37	-9999	-9999	49	15.2	54	18
12	2.02E+11	V10O1610	1	30	-9999	-9999	49	15.2	54	15
13	2.02E+11	V10O1610	1	33	-9999	-9999	50	15.2	54	16
14	2.02E+11	V10O1610	1	33	-9999	-9999	49	15.3	53	16

02 연구 가설

연구 가설 1 : 분석 과정 - Step 1

STEP 1

유동인구 데이터 전처리

전처리 포인트

1. 15분마다 측정된 데이터를 일별 평균 계산
2. 0~25세 범위가 5세 기준으로 나뉜 데이터를 하나의 카테고리로 합침

코드파일 위치

각 분석마다 해당 분석에서 사용한 코드가 위치한
코드 파일의 **번호**를 적어두었습니다.

In code file #B-1

노원구와 종로구의 성,연령별 유동인구의 일 평균 값이 저장된 **단일 데이터 프레임**이 생성됨

전처리 결과

	Date	Dong	M_0024	M_2529	M_3034	M_3539	M_4044	M_4549	M_5054	M_5559	...
1	20180401	청운호 자동	2109.14	1792.30	1936.10	2060.54	1787.98	2369.93	2348.95	2123.78	...
2	20180401	사직동	7725.78	6887.72	6786.94	6873.70	5920.97	7406.48	6956.00	7294.07	...
3	20180401	삼청동	1318.72	1264.86	1375.14	1445.40	1237.07	1658.35	1402.99	1360.18	...
4	20180401	부암동	3579.12	2732.00	3390.58	4288.35	3857.72	4912.69	4756.22	4692.38	...
5	20180401	평창동	6411.16	4927.24	5718.38	6997.92	6618.99	9064.13	9493.46	9873.77	...
6	20180401	무악동	1131.72	766.21	786.83	868.99	832.50	1077.11	1051.22	1024.64	...

⋮

Int64Index: 6205 entries, 1 to 1097
Data columns (total 22 columns):

02 연구 가설

연구 가설 1 : 분석 과정 - Step 1

STEP 1

미세먼지 데이터 전처리 1 ; 관측소 선정하기

전처리 포인트

동별로 어떤 관측소의 데이터를 사용할 것인가?

1. 종로구와 노원구의 미세먼지 관측소 위치 표시
2. 동 내 관측소 개수에 따라 전처리 방법 결정

- 관측소가 있는 동 → 그 관측소의 데이터 사용
- 관측소가 여러 개 있는 동 → 각 관측소 데이터의 평균 계산해 사용, 사직동의 경우 한 관측소가 모두 결측치므로 해당 관측소는 제외함
- 관측소가 없는 동 → 주변 관측소와의 거리를 계산해 가장 가까운 관측소의 데이터를 사용
- 다른 구의 관측소가 있는 경우 → 해당 데이터는 가까운 거리 계산에만 사용

두 위치의 위경도를 넣으면 거리(km)를 반환하는 함수 **haversine**

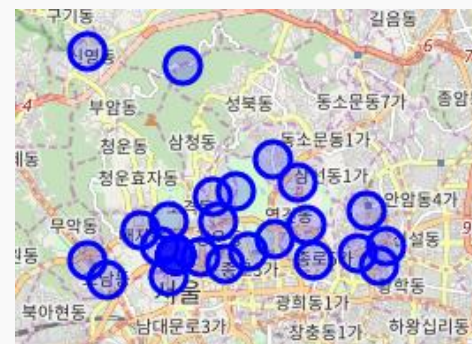
```
from haversine import haversine
n1 = (lat,lon)
n2 = (lat,lon)
haversine(n1,n2)
```

In code file #A-1

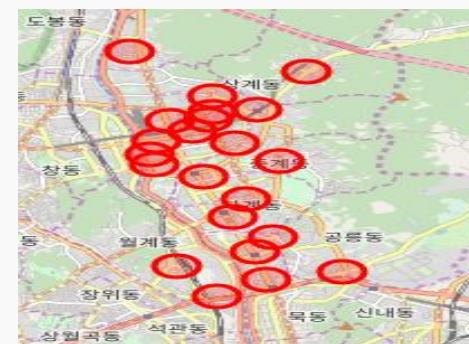
<관측소 기준에 따른 결과표>

기준	노원구	종로구
관측소가 없는 동	상계8동, 상계9동, 월계2동, 중계1동, 중계4동, 하계2동	무악동, 삼청동, 송인1동, 창신2동
타구 관측소 있는 동	없음	명동, 천연동

<종로구>



<노원구>



전처리결과 (시각화)

02 연구 가설

연구 가설 1 : 분석 과정 - Step 1

STEP 1

미세먼지 데이터 전처리 2 ; pm10/pm25 데이터 상관성 분석

pm10 데이터와 pm25 데이터의 상관 분석

상관성이 높을 경우 : 두 데이터가 비슷한 수치임을 가정하고 pm 10 데이터만 분석에 사용
상관성이 낮을 경우 : 두 데이터가 상이한 수치이므로 두 데이터 모두 분석에 사용

분석 포인트

1. 월별 데이터를 확인할 것이므로, 기존 분당 기록된 미세먼지 데이터를 일별로 통합하고자 함.
2. 주어진 미세먼지 데이터인 pm10과 pm25 중 하나의 데이터만 쓰기 위해 상관성을 분석함

In code file #A-3

노원구와 종로구 모두 pm10 데이터와 pm25데이터의 상관성이 매우 강했음. 두 데이터 모두 상관계수가 0.8~0.9를 상회하며 매우 강한 상관성을 입증함. 따라서, 수치적 비교를 명확히 하기 위해 pm10 데이터만 분석에 사용하기로 결정함

분석 결과

<노원구 미세먼지 측정기 데이터 분석 결과>

공릉2동172	0.913725
상계10동	0.909963
상계1동	0.893681
상계2동332	0.923806
상계2동397	0.915845
상계34동	0.938319
상계5동156	0.894596
상계5동456	0.917454
상계67동734	0.900994

상계67동740	0.903835
상계67동767	0.926091
월계1동	0.921774
월계3동	0.902237
중계23동	0.923896
중계본동	0.915543
하계1동251	0.917893
하계1동280	0.855487

<종로구 결과>

교남동	0.902209
명동	0.904946
송인2동	0.907810
창신1동170	0.873851
평창동	0.902455
혜화동1	0.906246
혜화동72	0.893715

* 본 데이터 분석은 데이터 상관성 입증이므로 관측소 제외 여부와 상관 없이 모든 데이터를 사용하였음

02 연구 가설

연구 가설 1 : 분석 과정 - Step 1

STEP 1

미세먼지 데이터 전처리 3 ; 데이터 합치기

결측치와 이상치 제거 기준

결측치 제거
-999와 -9999값을 제거함
이상치 제거
IQR*1.5 범위 밖 데이터를 이상치로 간주하고 제거함

전처리
포인트

1. 모든 열에 대해 결측치와 이상치가 들어있는 행의 데이터는 제거함
2. 위의 전처리가 포함된 함수를 생성하여 전체 동별 파일에 적용하여 데이터 프레임 생성
3. 동별로 pm10>80은 1, 그 외는 0의 값을 부여함
4. 월별로 미세먼지 나쁨인 날인 1의 값이 부여된 날의 수를 계산하여 하나로 통합

In code file #A-2

노원구와 종로구의 월별/동별 미세먼지 나쁨(기준 pm10>80)인 날의 빈도수가 저장된 단일 데이터 프레임이 생성됨

전처리
결과

	청운효자 동	사직 동	삼청 동	부암동	평창동	무악동	교남동	가회동	종로1234가	종로56가	이화동	혜화동	창신1동	창신2동	창신3동	송인1동	송인2동
pm10																	
201804	3	3	2	1	10	8	2	2	2	2	1	2	2	2	2	2	8
201805	7	7	1	3	10	4	7	1	1	1	4	1	1	1	5	5	1
201806	0	0	0	0	2	1	0	0	0	1	1	0	0	0	0	0	0

⋮

RangeIndex: 12 entries, 0 to 12
Data columns (total 17 columns)

02 연구 가설

연구 가설 1 : 분석 과정 - Step 2

STEP 2 유동인구, 미세먼지 데이터 합치기

데이터 처리 내용

1. 유동인구 데이터와 미세먼지 데이터 중 평일에 해당하는 행만 추출함
2. 평일에 해당하는 유동인구 데이터와 미세먼지 데이터를 동별로 합친 개별 데이터프레임 생성

날짜를 넣으면 요일을 반환하는
Datetime의 **weekday** 함수

날짜를 넣으면 요일에 해당되는 숫자를 반환
토요일, 일요일의 경우 5,6을 반환하므로 해당 값이
담긴 행은 모두 삭제하는 방식으로 처리

In code file #C

평일에 해당하는 유동인구 데이터와 미세먼지 데이터를 동별로 합친 개별 데이터프레임 생성

처리 결과

	Dong	유동인구	pm10
20180402	사직동	228497.58	67.281250
20180403	사직동	235732.21	70.256944
20180404	사직동	242992.17	21.386917
20180405	사직동	236220.14	12.054167
20180406	사직동	232095.14	81.513966
20180408	사직동	129433.90	33.235417
20180409	사직동	225989.72	48.705556

20180410	사직동	241528.70	51.075694
20180411	사직동	238391.24	76.484028
20180412	사직동	249065.74	51.911111
20180413	사직동	243617.07	37.961806
20180415	사직동	143549.43	70.556944
20180416	사직동	233095.47	33.765278
20180417	사직동	247186.77	56.141667
20180418	사직동	246145.70	66.204861

...

Index: 312 entries, 20180402
to 20190331
Data columns (total 3 columns):

02 연구 가설

연구 가설 1 : 분석 과정 - Step 3

STEP 3

심층 분석

상관계수 구하기

corr 메소드와 method속성을 pearson으로 주어
각각의 동의 피어슨 상관계수를 구함

분석 내용

각 동별 데이터프레임의 유동인구와 미세먼지의 상관계수를 구한 각 값을 합친 데이터프레임을 생성하여
유동인구와 미세먼지와의 상관계수를 통해 둘의 상관성을 살펴봄

In code file #C

처리 및
분석 결과

종로구

	가회동	교남동	무악동	부암동	사직동	삼청동	송인1동	송인2동	이화동	종로 1234가 동	종로56가 동	창신1동	창신2동	창신3동	청운효자 동	평창동	혜화동
0	-0.041066	-0.162183	0.097532	-0.537474	-0.270163	0.210014	-0.371013	-0.14543	-0.065525	-0.12863	-0.187432	-0.146648	0.207217	0.14699	-0.228789	-0.215641	-0.429102

노원구

	공릉1동	공릉2동	상계10동	상계1동	상계2동	상계34동	상계5동	상계67동	상계8동	상계9동	월계1동	월계2동	월계3동	중계1동
0	-0.219826	0.109592	-0.221312	-0.151253	-0.281383	-0.387641	-0.299013	-0.090964	0.018345	-0.347932	-0.150993	-0.289722	0.005481	-0.220597

	상계8동	상계9동	월계1동	월계2동	월계3동	중계1동	중계23동	중계4동	중계본동	하계1동	하계2동
0.018345	-0.347932	-0.150993	-0.289722	0.005481	-0.220597	-0.047657	-0.01865	0.284461	-0.114706	0.049462	

분석 결론

처리 결과, 종로구와 노원구 모두 피어슨 상관계수가 0.4 미만이며, 전반적으로 매우 낮은 값을 보이고
있으므로, 종로구와 노원구 모두 평일 유동인구와 미세먼지의 상관관계는 약하다고 판단됨

02

연구 가설

연구 가설 2 : 미세먼지와 공기 여과에 대한 관심도의 상관성



insight

미세먼지가 심해지면서 일반 시민들의 공기 여과 시설에 대한 관심이 증대되지 않았을까?

연구 가설

SNS상의 미세먼지의 언급 횟수와 공기 여과와 관련된 단어의 언급 횟수는 비슷한 추이를 보일 것이다.

분석 방법

활용 데이터	미세먼지 데이터, SNS 데이터
분석 언어	Python, R
활용 Tool	Jupyter Notebook, Pycharm, Rstudio

02 연구 가설

연구 가설 2 : 분석 데이터 소개

DATA 3

SNS 데이터

- 출처 : 대회 제공 데이터
- 수집 및 활용 목적 : 미세먼지와 공기 청정 설비에 대한 관심도 측정
- 지역 및 기간 범위 : 지역 범위 없음, 20180401~20190331, 미세먼지 키워드가 있는 모든 게시물
- 변수 범위 : DATE(날짜), SECTION(블로그/카페/뉴스), TITLE(게시글 제목), CONTENT(게시글 내용)

	A	B	C	D	E	F	G	H	I	J	K	L
1	DOCID	SECTION	DATE	TITLE	CONTENT							
2	204172453	블로그	20181102	신동탄 이	분 양 정 보 신동탄 이안파밀리에 눈에 보이는 프리미엄 진짜네요 유쾌한 분양김							
3	204172474	블로그	20181102	느린 점심	생활 속의 단상 느린 점심點心 묘광거사 2018 11 2 2357 URL 복사 이웃추가 본							
4	204172506	블로그	20181102	한화 태양	한화프렌즈 한화 태양의 숲 조성으로 미세먼지도 막고 환경도 지켜요 아르엘 20							
5	204172546	블로그	20181102	섬유유연제	소핑 섬유유연제추천 미세먼지까지 잡아 주는 착한 샤프란 섬유유연제 꽃담초							
6	204172587	블로그	20181102	핑거루트	리뷰 핑거루트 허브 다이어트 남자 몸매관리 후기 김중성유통 2018 11 2 2352 U							
7	204172624	블로그	20181102	어메의 28	280일 임신기록 어메의 280일 기록 임신26주 써니훈택 2018 11 2 2352 URL 복							
8	204172654	블로그	20181102	당진국화축	제 지난 주말 사진입니다 당진국화축제 10월 27일 11월 4일까지 서산국화축제 10월							
9	204172688	블로그	20181102	삼성전자	정보 삼성전자 에어드레서 의류청정기는 미세먼지 심한 겨울철에 필수템이네요							
10	204172729	블로그	20181102	코타키나발	코타키나발루 날씨 코타키나발루 1월 날씨 여행옷차림 강수량우기 매월 기후 정							
11	204172756	블로그	20181102	베이비템	사안녕하세요 지은지성맘입니다 워킹맘이다보니깐 빨래는 포기할 수 없고 집안							
12	204172797	블로그	20181102	유아마스크	제품리뷰 유아마스크 별과모래 면마스크 매일 써서 재구입이 시급하다 마이애플							
13	204172841	블로그	20181102	방학동용달	게시판 방학동용달이사수유동미아동용달이사원룸이사 용이 2018 11 2 2343 UR							

02 연구 가설

연구 가설 2 : 분석 과정 - Step 1

STEP 1 SNS 데이터 전처리

전처리 포인트

1. SNS데이터 파일을 필요한 열로 뽑아 정리하는 함수 생성하여 8개로 나뉜 파일을 데이터프레임으로 변환
2. 8개의 데이터프레임을 하나의 SNS 데이터프레임으로 합치고, 월별로 12개의 데이터프레임 생성

In code file #D

전체 SNS 데이터가 포함된 월별 단일 데이터프레임 생성

처리 결과

	date	section	year	month	day	title	content
0	20181102	블로그	2018	11	02	신동탄 이안파밀리에 눈에 보이는 프리미엄 진짜네요	분 양 정 보 신동탄 이안파밀리에 눈에 보이는 프리미엄 진짜네요 유쾌한 분양킴 2...
1	20181102	블로그	2018	11	02	느린 점심點心	생활 속의 단상 느린 점심點心 모광거사 2018 11 2 2357 URL 복사 이...
2	20181102	블로그	2018	11	02	한화 태양의 숲 조성으로 미세먼지도 막고 환경도 지켜요	한화프렌즈 한화 태양의 숲 조성으로 미세먼지도 막고 환경도 지켜요 아르엘 2018...
3	20181102	블로그	2018	11	02	섬유유연제추천 미세먼지까지 잡아 주는 착한 샤프란 섬유유연제 quote; 꽃담초 quote;	쇼링 섬유유연제추천 미세먼지까지 잡아 주는 착한 샤프란 섬유유연제 꽃담초 요미요...
4	20181102	블로그	2018	11	02	핑크루트 허브 다이어트 남자 몸매관리 후기	리뷰 핑거루트 허브 다이어트 남자 몸매관리 후기 김중성유통 2018 11 2 23...
5	20181102	블로그	2018	11	02	어메의 280일 기록 임신26주	280일 임신기록 어메의 280일 기록 임신26주 써니훈덕 2018 11 2 2...

Int64Index:

777832 entries, 0 to 21688

Data columns (total 7 columns):

02 연구 가설

연구 가설 2 : 분석 과정 - Step 1

STEP 1

미세먼지 데이터 전처리

전처리 포인트

1. 연구가설 1에서 도출한 일별/동별 데이터 프레임을 모두 불러와 단일 데이터로 통합함
2. 통합된 데이터를 일괄 평균 처리함
3. 년-월-일 로 지정된 일자 데이터를 년-월로 자른 후, group_by 함수를 사용해 월별로 평균 처리함

In code file #A-2

처리 결과

월별 평균 미세먼지 값이 담긴
단일 데이터프레임이 생성됨

	Date	pm10
0	201804	46.685984
1	201805	45.424308
2	201806	41.373570
3	201807	34.654231
4	201808	37.530600
5	201809	34.126963

6	201810	38.826354
7	201811	55.999444
8	201812	59.483938
9	201901	72.169308
10	201902	68.277820
11	201903	59.584523

02 연구 가설

연구 가설 2 : 분석 과정 - Step 2

STEP 2

심층 분석 ; 미세먼지 관련 단어 횟수 세기

전처리 포인트

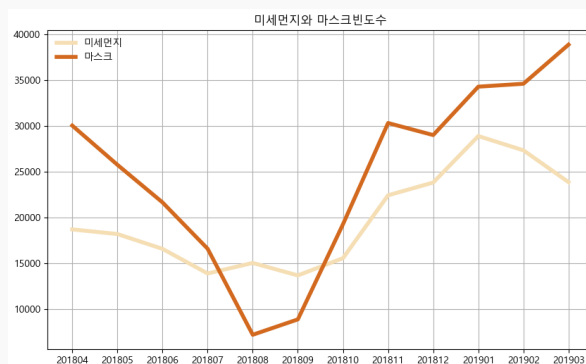
1. 키워드와 월을 input하면 횟수를 세주는 함수 생성
 - 키워드 선정 기준 : 공기 청정(여과)와 관련된 단어
2. 해당 함수를 월별로 반복문을 사용하여 결과값을 도출하고, 해당 결과값을 list화 하여 저장
3. 월별 단어 빈도 값과 월별 미세먼지 농도를 비교하는 선 그래프 작성
 - 미세먼지 농도와 빈도 값에 큰 차이가 있으므로, 미세먼지 농도값에 400 가중치를 두어 비교를 편하게 함

가중치 선정은 가장 추이 비교를 하기에
편리한 값인 400으로 선정함

```
plt.plot(date_list,asarray(dust_list)*400,  
        'wheat', label = '미세먼지')
```

In code file #D

처리 결과



월별 미세먼지 농도와
키워드 언급 추이가 비교 가능한
선 그래프 생성

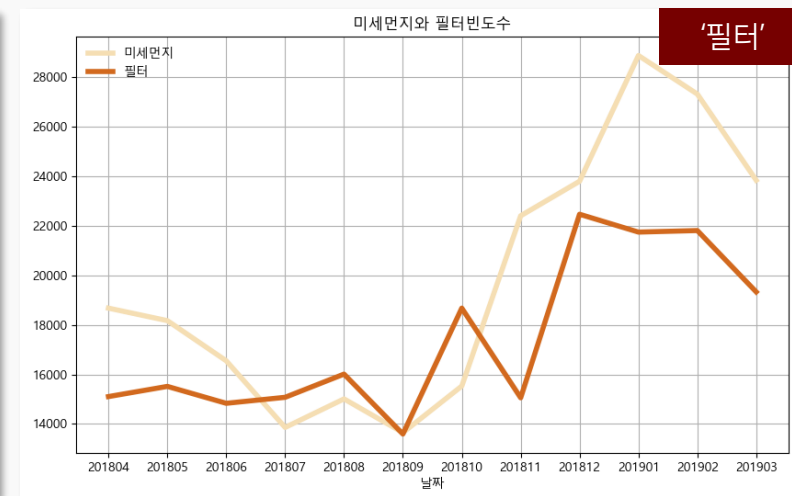
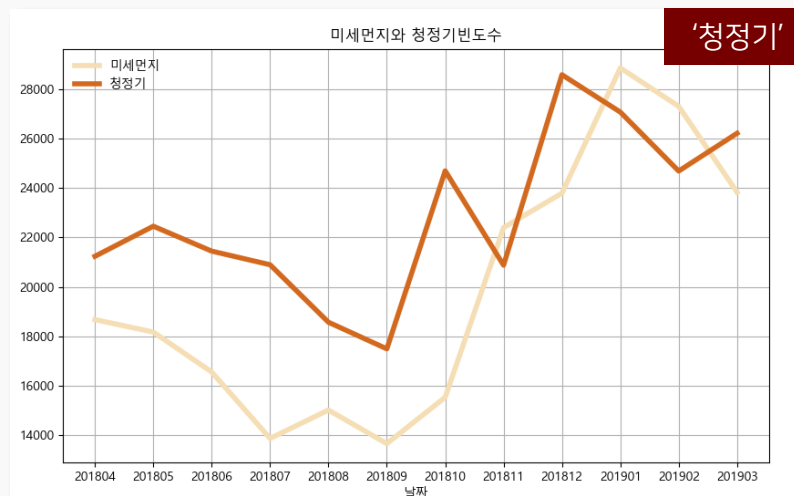
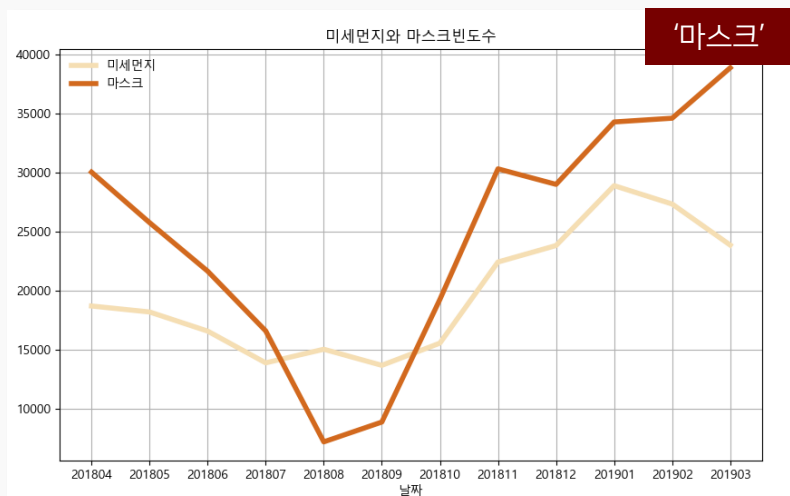
02 연구 가설

연구 가설 2 : 분석 과정 - Step 3

STEP 3

시각화 결과 및 분석 결론

공기청정(여과)와 관련된 키워드를 넣어 content 항목에서 빈도수 추출



In code file #D

* 세 그래프의 가중치는 모두 동일하게 처리하였습니다.

분석 결론

'청정기', '필터', 특히 '마스크' 키워드의 빈도 수는 미세먼지 추이와 비슷하게 나타남

3. 사업 아이템 소개

청정 버스 노선 및 정류장 설치 제안

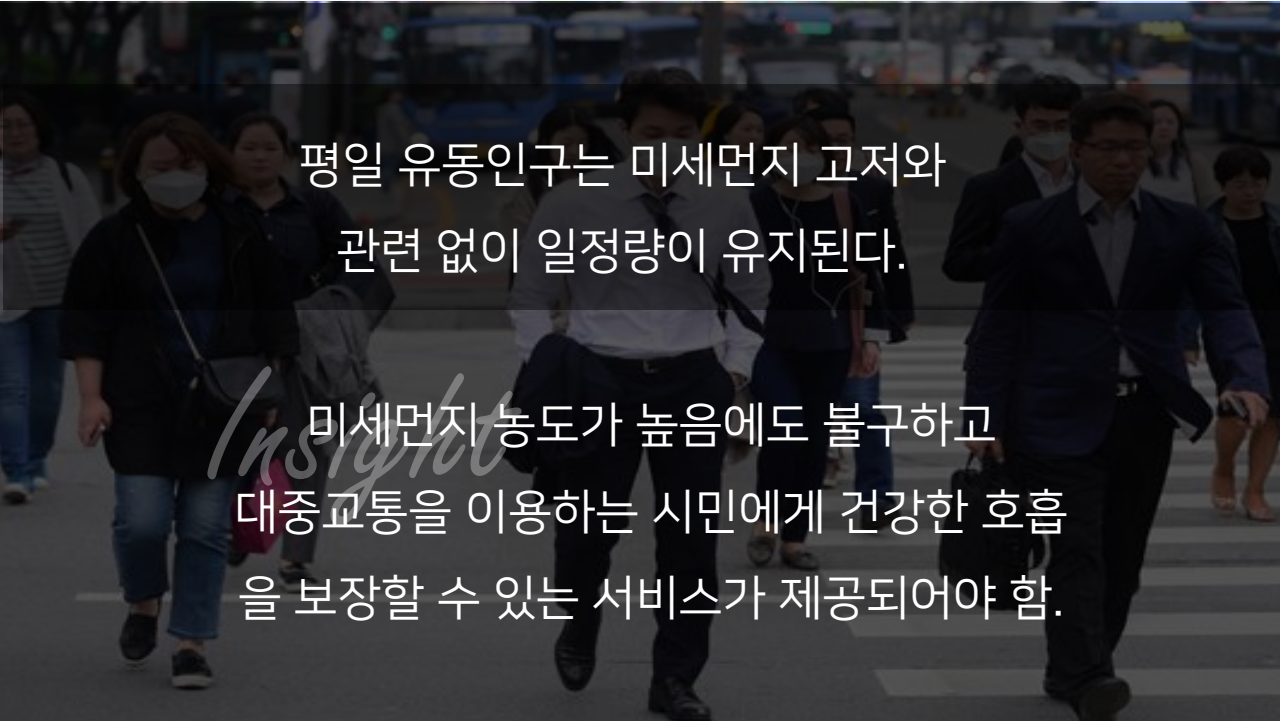


03

사업 아이템 소개

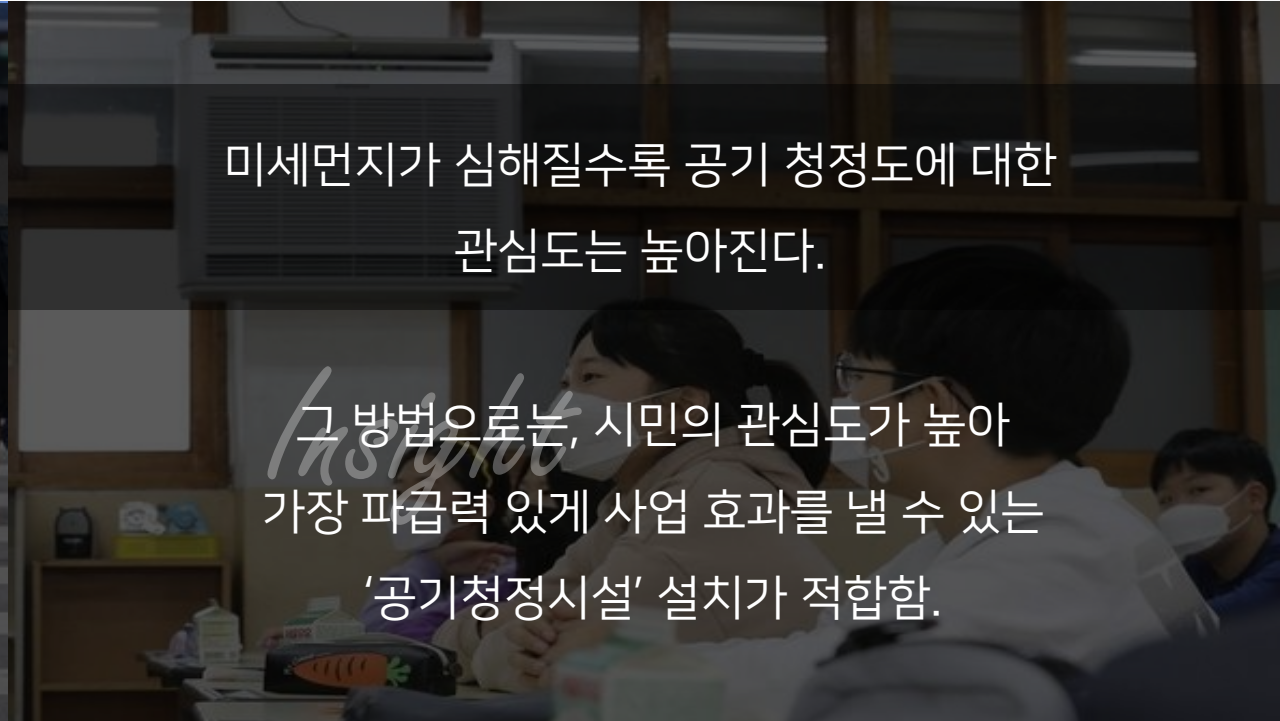
사업 아이템 도출 과정

연구 가설 분석 결과 및 인사이트 도출



평일 유동인구는 미세먼지 고저와
관련 없이 일정량이 유지된다.

Insight
미세먼지 농도가 높음에도 불구하고
대중교통을 이용하는 시민에게 건강한 호흡
을 보장할 수 있는 서비스가 제공되어야 함.



미세먼지가 심해질수록 공기 청정도에 대한
관심도는 높아진다.

Insight
그 방법으로는, 시민의 관심도가 높아
가장 파급력 있게 사업 효과를 낼 수 있는
'공기청정시설' 설치가 적합함.

03 사업 아이템 소개

사업 개요



ITEM 1 '미세먼지 프리 버스' 아이템

- 개요 : 버스 내 공기청정기를 설치하여 이용객들에게 미세먼지 없는 대중교통 이용을 할 수 있도록 함
- 선정 : 구별로 1개의 마을버스 노선 선정
(노선이 단일 구 내에서만 한정되는 마을버스로 선정)
- 대상 : 버스 탑승객 및 기사님



ITEM 2 '우리동네 공기쉼터' 아이템

- 개요 : 기존 정류장을 보강하여 폐쇄형 버스 정류장을 설치하고 공기 청정 서비스 및 데이터 기반 편의 서비스를 제공하여 이용객 및 시민의 복합 쉼터로 활용할 수 있도록 함
- 선정 : 구별로 5개의 정류장 선정
- 대상 : 버스 정류장 이용객 및 보행자

03

사업 아이템 소개

사업 목적 및 의의

사업 실행 목적

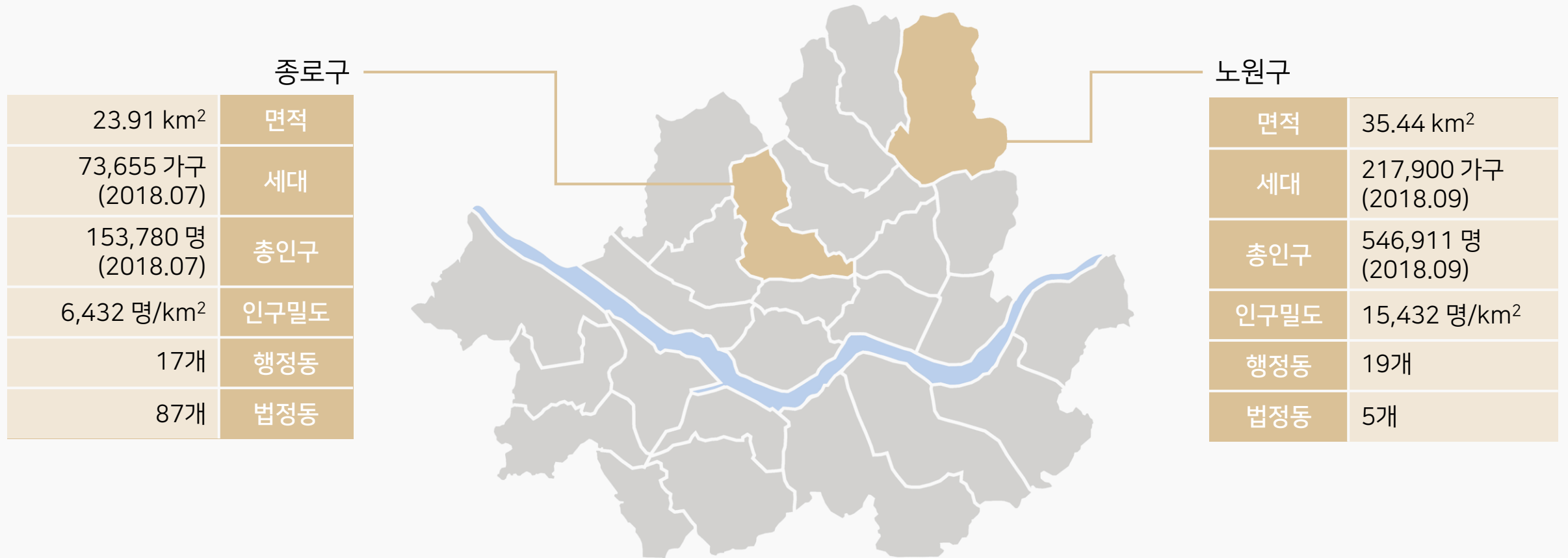
서울 시민의 삶의 질을 위협하는 미세먼지에 대처하기 위해 이용률이 높은 버스 노선과 버스 정류장을 대상으로 공기 질을 높일 수 있는 공기 청정 시설을 설치하고자 함. 특히, 노약자 비율이 높은 구의 경우 노약자 이용률이 높은 지점을 중점적으로 선정하여 **미세먼지에 취약한 노약자 계층의 건강과 삶의 질을 개선하는데 도움이 되고자 함.**

사업 의의

대중교통 이용률이 많은 곳을 중심으로 공기 청정 구역을 설치함으로써 **서울 시민의 생활의 질을** 높이고, 장기적으로 시민의 **대중 교통 이용을 장려함.** 또한, 노약자의 보행 시 건강한 호흡을 돕는 **쉽터의 역할도 겸할 것**으로 기대되므로 서울시 미세먼지 취약 계층의 건강 증진에 도움이 될 것으로 예상됨.

03 사업 아이템 소개

시범 사업 구 - 종로구, 노원구



* 시범 사업 구 선정 기준 : 대회 자료 기반

03 사업 아이템 소개

사업지 선정 및 사업 구분 기준

종로구

노원구 대비 거주 인구 수는 낮으나, 유동인구 수는 높음.
따라서, 상업 지역에 해당함

→ 유동인구 비율이 높기 때문에 두개의 아이템을 모두
활용해 사업지 선정

노원구

노원구 대비 거주 인구 수가 월등히 높음.
또한, 노약자 유동인구 비율이 종로구보다 높음.
따라서, 거주 지역에 해당함

→ 노약자 거주인구를 배려하여 보행자 쉼터 기능을 가
미한 '우리동네 공기쉼터' 아이템만 선정

사업지 선정 및 사업 구분 기준

‘미세먼지 프리 버스’ 사업지 선정 기준

- 1 미세먼지 많은 상위 5개 동을 선정하여 노선을 1차 선정
- 2 1차 선정된 노선들 중 승하차 인원 모두 많은 정류소가 포함된 노선을 최종 선정

‘우리동네 공기쉼터’ 사업지 선정 기준

- 1 미세먼지 많은 상위 5개 동을 선정하여 정류소를 1차 선정
- 2 1차 정류소 선정 시 노약자 유동인구도 함께 고려함
- 3 앞서 선정된 정류소 중 승차 인원이 많은 상위 5개 정류소 최종 선정

노원구만 해당

4. 사업지 선정

빅데이터 분석을 통한 노원/종로구 최적 시범사업지 선정



04 사업지 선정 분석 데이터 소개

DATA 4

서울시 버스노선정보 데이터

- 출처 : 열린데이터광장
- 수집 및 활용 목적 : 노원구, 종로구 마을/지선버스 노선명과 정류소명의 위치 좌표 파악
- 지역 및 기간 범위 : 서울특별시 (2018.04~2019.03)
- 변수 범위 : 노선명, 정류소명, X좌표, Y좌표

노선ID	노선명	순번	구간ID	정류소ID	ARS-ID	정류소명	X좌표	Y좌표
100100124	0017	1	0	102000271	03689	청암자이아파트	126.9465174884	37.5343626071
100100124	0017	2	102700549	102000204	03298	청암동강변삼성아파트	126.9493037945	37.5339606747
100100124	0017	3	102700550	102000227	03321	청심경로당	126.9504485667	37.5337439118
100100124	0017	4	102700551	102000210	03304	원효2동주민센터	126.9509043881	37.5342784061
100100124	0017	5	102700552	102000212	03306	산전동	126.9539838422	37.5354201502
100100124	0017	6	102700553	102000214	03308	신창동세방아파트	126.9550322215	37.5357487714
100100124	0017	7	102700554	102000216	03310	남이장군사당	126.9579568939	37.5367069612
100100124	0017	8	102700555	102000218	03312	새마을금고	126.9595027769	37.5371259582
100100124	0017	9	102700779	102000153	03247	용문시장	126.9607348708	37.5366304221
100100124	0017	10	102700557	102000156	03250	용산전자상가입구	126.9620433894	37.5344828174
100100124	0017	11	102700489	102000158	03252	신용산.지하차도	126.963900286	37.5326572802
100100124	0017	12	102701354	102000161	03255	용산역	126.966155	37.529349
100100124	0017	13	102700390	102000162	03256	용산푸르지오써밋	126.9645186276	37.527193494
100100124	0017	14	102700560	102000029	03122	한강대교북단	126.964005685	37.5252500866
100100124	0017	15	102700780	102000200	03294	서부이촌동입구	126.959676039	37.5236972645

DATA 5

서울시 버스노선별 정류장별 승하차 데이터

- 출처 : 열린데이터광장
- 수집 및 활용 목적 : 노원구, 종로구 내 전체 정류장별 승하차 인원 파악
- 지역 및 기간 범위 : 서울특별시 (2018.04~2019.03)
- 변수 범위 : 노선명, 정류소명, 승차, 하차

사용일자	노선ID	노선번호	노선명	표준버스정류장ID	정류소명	역명	승차	하차	승차+하차
20190301	11110336	421	421번(염곡동~옥수동)	105000019	6102	70032 신철동역	142	260	20190304
20190301	41110216	108	108번(양주 덕정~종로)	207000054	61031	74761 자동차검문소	14	16	20190304
20190301	11110672	성동02	성동02(마장동현대아파트~성동02)	103900178	4818	9014088 벽산아파트	26	25	20190304
20190301	41110205	507	507번(석수역~동대문)	117000081	18167	10758 박미마을	98	2	20190304
20190301	11110196	2413	2413번(성수동~개포)	122000123	23226	9002653 대치역	44	108	20190304
20190301	41110152	N37	N37번(송파공영차고지~신정)	122000123	23226	9002653 대치역	11	1	20190304
20190301	11110990	5522A	5522번(A,난곡방향,난곡역~신정)	120000068	21170	9673 상장아파트	19	64	20190304
20190301	41110151	N26	N26번(강서공영차고지~신정)	113000418	14064	8002642 이대역	7	7	20190304
20190301	11110002	101	101번(화계사~동대문)	100000367	1044	8000821 동묘앞	89	301	20190304
20190301	41110079	7728	7728번(대화동~신촌)	219000316	36025	9004977 탄현큰마루	127	11	20190304
20190301	11110013	142	142번(도봉동~고속터미널)	107000003	8003	71613 길음뉴타운	210	150	20190304
20190301	41110078	7727	7727번(설문동~신촌)	219000320	36302	8501446 상감천마루	0	28	20190304
20190301	11110016	145	145번(번동~강남역)	121000016	22016	9006333 신사역	75	95	20190304
20190301	41110150	N26	N26번(중랑공영차고지~신정)	115000022	16118	7384 공향시장	0	3	20190304
20190301	11110039	360	360번(송파차고지~신정)	122000180	23283	7923 역삼역	292	116	20190304

04 사업지 선정

분석 데이터 소개

DATA 6

서울시 동별 JSON 정보 데이터

- 출처 : GitHub의 southkorea-maps
- 수집 및 활용 목적 : 노원구, 종로구 버스 노선 및 정류소 위치 지도 시각화
- 지역 및 기간 범위 : 서울특별시 (2018 ver)
- 변수 범위 : 노원구와 종로구 행정동 내의 모든 데이터

```
← → ↺ raw.githubusercontent.com/southkorea/southkorea-maps/master/kostat/2013/json/skorea_submunicipalities_geo_simple.json
앱 Ewha CyberCamp...

{"type": "FeatureCollection", "features": [{"type": "Feature", "properties": {"code": "3902062", "name": "예래동", "name_eng": "Yeraedong", "base_year": "2013"}, "geometry": {"type": "Polygon", "coordinates": [[[126.4507554288519, 33.35524840263458], [126.43109137134583, 33.31469904417688], [126.41904652914948, 33.238204238225975], [126.37030085931454, 33.22738115481318], [126.38172437948549, 33.29900798215933], [126.40359940546149, 33.28975729195352], [126.42548871172248, 33.31077300041721], [126.4354320681953, 33.33648480736013], [126.4507554288519, 33.35524840263458]]]}}, {"type": "Feature", "properties": {"code": "3902061", "name": "중문동", "name_eng": "Jungmun-dong", "base_year": "2013"}, "geometry": {"type": "Polygon", "coordinates": [[[126.5073141568715, 33.354405740648], [126.47668302051052, 33.30988326907272], [126.48128253168437, 33.28413626673973], [126.45861334654401, 33.235743925105254], [126.41904652914948, 33.238204238225975]]]}}
```

04

사업지 선정

최적 사업지 선정 데이터 전처리 과정

STEP 1

버스노선 정보 데이터 전처리

전처리 포인트

1.

서울시 정류장 위치 좌표로 주소 추출하여 구만 분리하여 열 추가
2.

노원구, 종로구의 마을, 지선버스만 추출하여 4개의 데이터프레임 생성

In code file #E

전처리 결과

마을버스와 지선버스의 정보가 저장된 데이터프레임이 생성됨

마을버스 데이터(노원구)

노선명	순번	ARS-ID	정류소명	X좌표	Y좌표	구
노원01	1	11577	보람사거리	127.067094	37.664367	노원구
노원01	2	11531	계상초등학교	127.069051	37.662815	노원구
노원01	3	11926	24시편의점	127.070720	37.661440	노원구
노원01	4	11515	상계역	127.073501	37.660741	노원구
노원01	5	11521	상계3동파출소	127.075068	37.663827	노원구

RangeIndex: 317 entries, 0 to 316
Data columns (total 5 columns):

지선버스 데이터(노원구)

노선명	순번	ARS-ID	정류소명	X좌표	Y좌표	구
1017	1	11482	삼화상운차고지	127.052922	37.626070	노원구
1017	2	11331	인덕삼거리	127.057411	37.628515	노원구
1017	3	11332	신도브래뉴.한전노원변전소	127.059124	37.628191	노원구
1017	4	11336	월계삼거리	127.060135	37.622992	노원구
1017	5	11335	광운대학교	127.058205	37.619818	노원구

RangeIndex: 1204 entries, 0 to 1203
Data columns (total 5 columns):

04 사업지 선정

최적 사업지 선정 데이터 전처리 과정

STEP 2 버스 승하차 데이터 전처리

전처리 포인트

1. 노원구, 종로구의 2019년 1~3월 버스 승하차 일별 데이터만 병합하여 1개의 데이터프레임 생성
2. 정류소마다 중복되는 노선이 있으므로 정류소명으로 그룹화하여 합산

정류소명	노선명	승차	하차	승하차
14단지상가	노원02	16060	16763	32823
14단지상가	노원11	13405	12793	26198



정류소명	승차	하차	승하차
14단지상가	29465	29556	59021

In code file #F

노원구, 종로구에 위치한 3237개 버스 정류소의 1~3월 일별 승하차 합계 인원의 정보가 저장된 데이터프레임 생성

전처리 결과

정류소명	승차	하차	승하차
14단지상가	29465	29556	59021
2001아울렛중계점	4091	7979	12070
202번중점	8873	20459	29332
24시편의점	18928	26087	45015
⋮			

RangeIndex: 3237 entries, 0 to 3236
Data columns (total 4 columns)

04 사업지 선정

최적 사업지 선정 데이터 전처리 과정

STEP 3 서울시 동별 위치 JSON 파일에서 노원, 종로구 추출

전처리 포인트

1. 노원구, 종로구의 동에 해당하는 json 정보를 웹 페이지 내에서 로드하여 복사함
2. 해당 내용을 dic 형태로 변수 저장 후 dumps 메서드 사용해 str 형태로 저장
3. str 형태의 정보를 지도 시각화에 사용할 수 있도록 각각 json 파일로 저장

In code file #Z

노원구, 종로구의 법정동 위치 데이터가 json 파일로 각각 저장됨 (순서대로 노원구, 종로구 예시)

전처리 결과

```
노원구 json.py x nowon_dong.json x
1 {
2   "type": "FeatureCollection",
3   "features": [
4     {
5       "type": "Feature",
6       "properties": {
7         "code": "1111079",
8         "name": "공릉1동",
9         "name_eng": "Gongneung 1(il)-dong",
10        "base_year": "2013"
11      },
12      "geometry": {
13        "type": "Polygon",
14        "coordinates": [
```

```
노원구 json.py x nowon_dong.json x jongno_dong.json x
1 {
2   "type": "FeatureCollection",
3   "features": [
4     {
5       "type": "Feature",
6       "properties": {
7         "code": "1101073",
8         "name": "혜화동",
9         "name_eng": "Hyehwa-dong",
10        "base_year": "2013"
11      },
12      "geometry": {
13        "type": "Polygon",
14        "coordinates": [
```

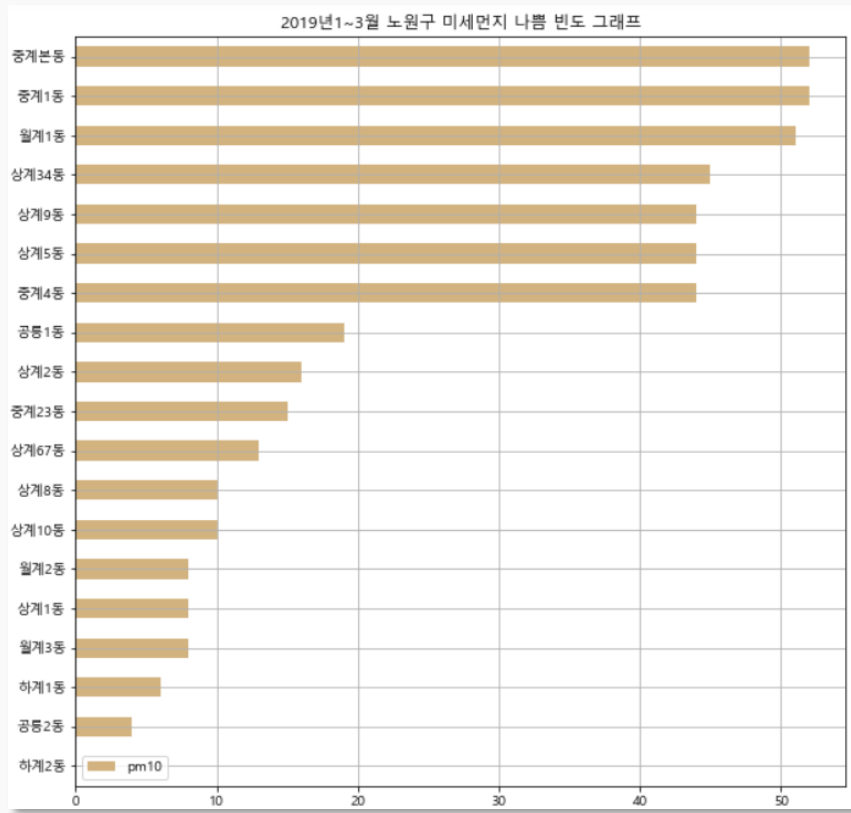

04 사업지 선정

최적 사업지 선정 - 노원구

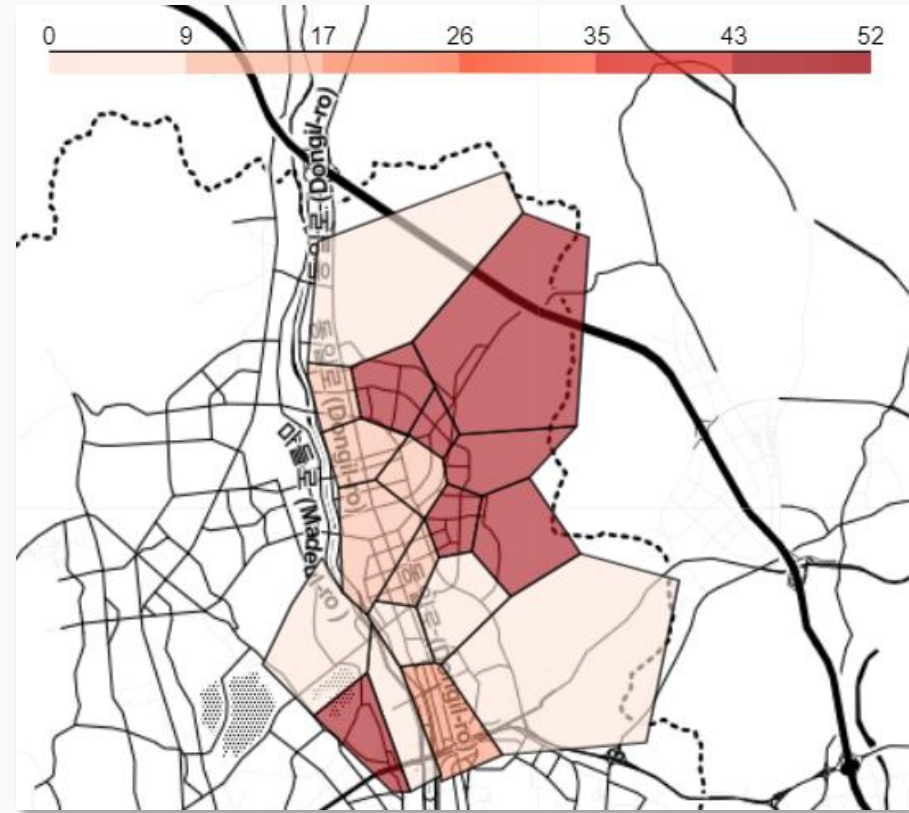
STEP 1

1~3월 미세먼지 농도에 따라 그래프 및 지도 시각화

막대 그래프 시각화



지도 시각화



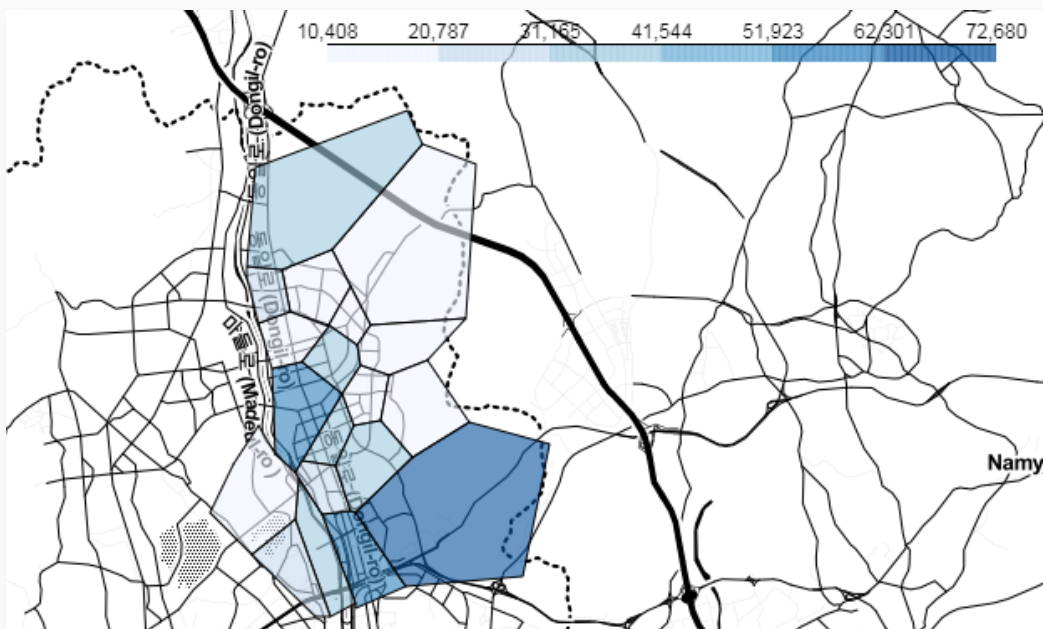
In code file #G

04 사업지 선정

최적 사업지 선정 - 노원구

STEP 2

1~3월 노약자 유동인구에 따라 지도로 시각화



In code file #G

STEP 3

최적 사업지 후보 상위 5개동 선정

	pm10	old_sum
중계1동	52	16967.144778
중계본동	52	16416.606444
월계1동	51	30078.651111
상계34동	45	20455.105556
중계4동	44	16254.684556
상계5동	44	12667.493000
상계9동	44	10407.832556
공릉1동	19	59007.702111

정렬 기준
1순위 : 미세먼지
2순위 : 노약자 유동인구

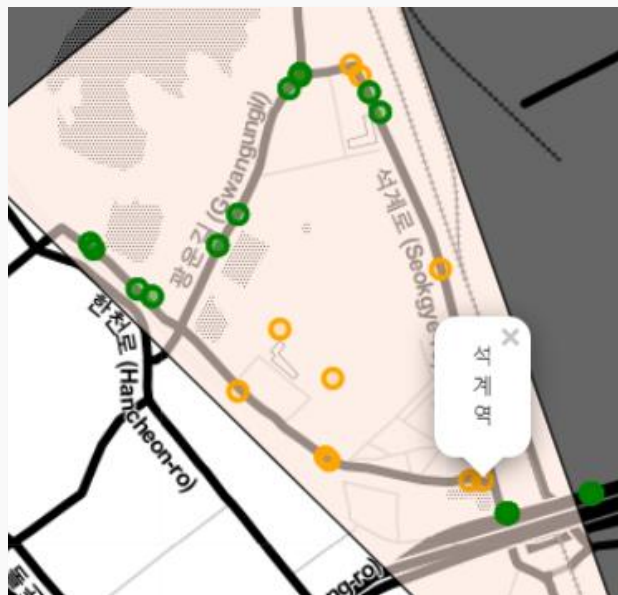
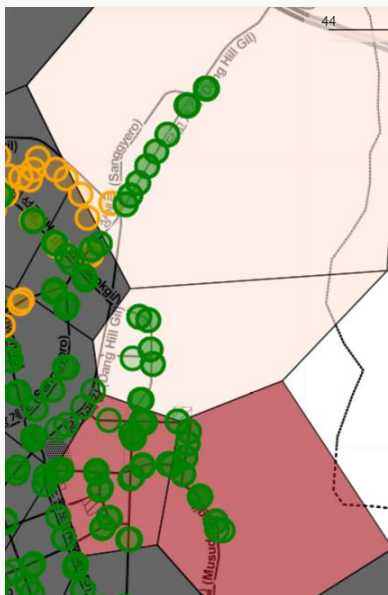
※ 노원구의 경우, 노약자 거주인구가 많기 때문에 노약자 유동인구를 함께 고려

04 사업지 선정

최적 사업지 선정 - 노원구

STEP 4

Top5동만 표시한 지도에 마을/지선버스 정류소 표시



In code file #F

최종 버스 정류소 사업지 선정

- 석계역1번출구 A
- 은행사거리
- 중계1동주민센터
- 광운대역
- 석계역1번출구 B

상위 5개동에 위치한 모든 마을/지선 버스 정류소를 노원구 전체 승차 인원 수로 정렬하여 최적 버스 정류소 사업지 5곳을 선정

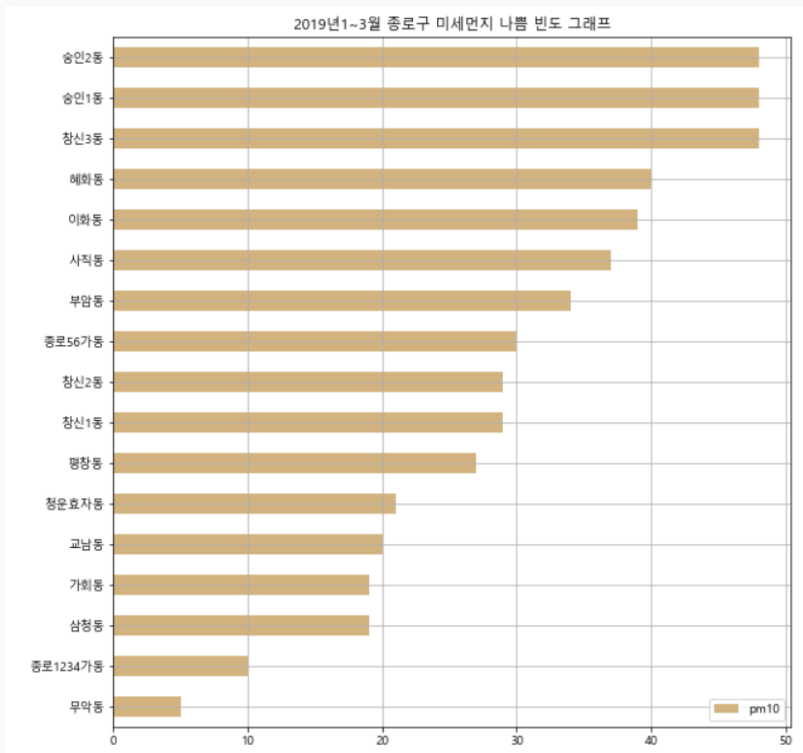
04 사업지 선정

최적 사업지 선정 - 종로구

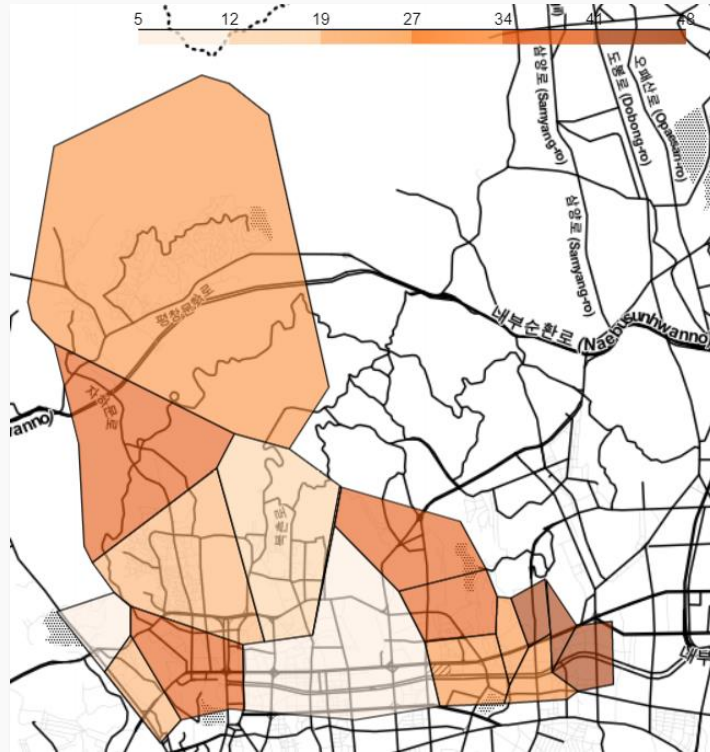
STEP 1

1~3월 미세먼지 농도에 따라 그래프 및 지도 시각화

막대 그래프 시각화



지도 시각화



In code file #G

STEP 2

최적 사업지 후보 상위 5개동 선정

pm10	
송인2동	48
송인1동	48
창신3동	48
혜화동	40
이화동	39
사직동	37
부암동	34
종로5.6가동	30

04 사업지 선정

최적 사업지 선정 - 종로구

STEP 3

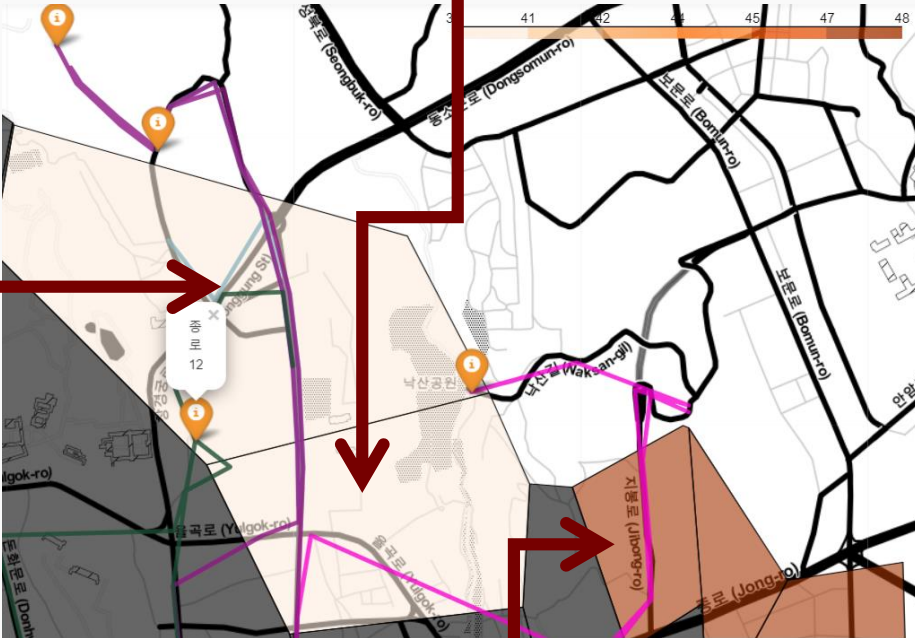
Top5동만 표시한 지도에 마을버스 노선 표시

혜화동

종로07, 종로08, 종로12

이화동

종로03, 종로07, 종로08, 종로12

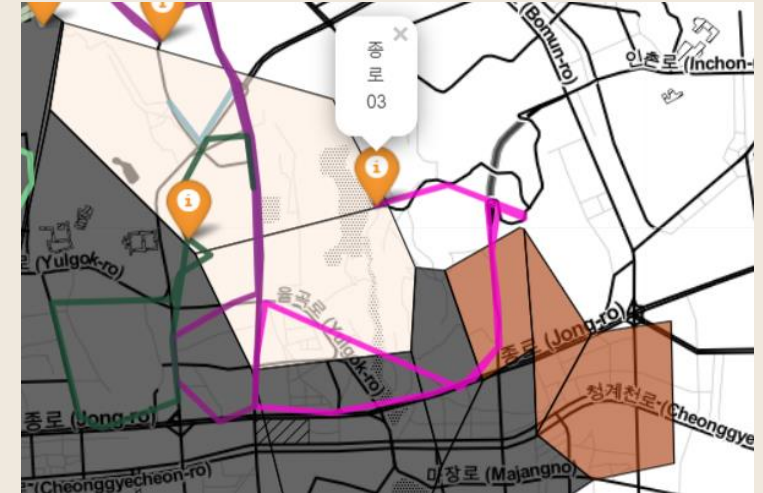


In code file #G

종로03

창신3동

최종 버스 노선 사업지 선정



승하차

노선명

종로08 1570471

종로03 1194445

종로12 399540

종로07 52698

1. 5개 동 지나는 버스노선 선정
2. 종로구 전체 승하차 인원 수로 정렬

분석 결과, 종로 03이 가장 적합함

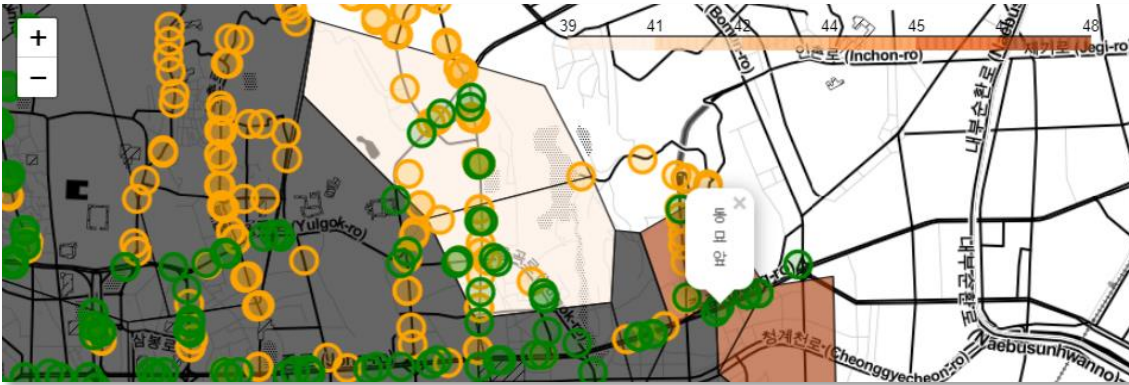
04 사업지 선정

최적 사업지 선정 - 종로구

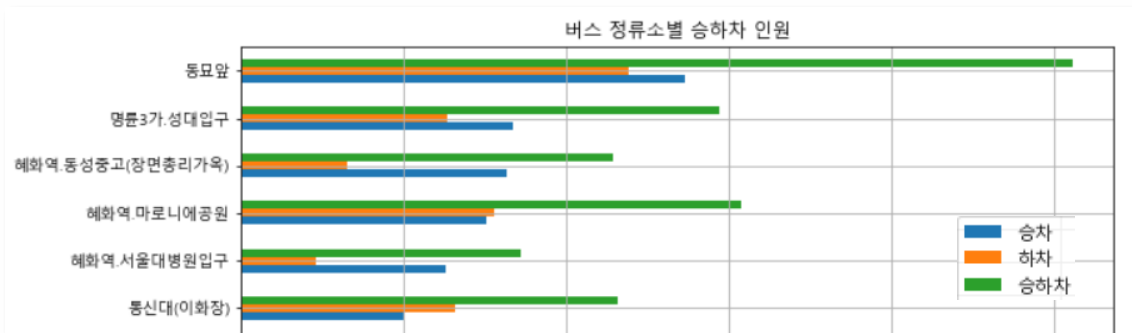
STEP 4

Top5동만 표시한 지도에 마을/지선버스 정류소 표시

지도 시각화



막대 그래프 시각화



In code file #F,G

최종 버스 정류소 사업지 선정

- 동묘앞
- 명륜3가.성대입구
- 혜화역.동성중고
- 혜화역.마로니에공원
- 혜화역.서울대병원입구

Top5동에 위치한 모든 마을/지선 버스 정류소를 종로구 전체 승차 인원 수로 정렬하여 최적 버스 정류소 사업지 5곳을 선정

5. 한계점 및 향후 계획

동부간선도로
Dongbu Expwy 북 (N)
성수대교
Seongsudaegyo (Br) 남 (S) 150m

한남대교
Hannamdaegyo
용비교
Yongbidaegyo
강변북로
Gangbyeonbuk-ro
(천호대교)

성수대교 사고희생자
위령비 (주차장) →

05

한계점 및 향후 계획

분석 및 사업의 한계점 및 보완 계획

가설 검증 데이터가 종로구, 노원구에 한정되어 있음

→ 향후 사업 유치할 경우, 서울시 전체 데이터로 확장하여 재분석할 예정임.

‘미세먼지 프리 버스’ 아이템의 경우 노선이 마을버스에 한정되어 있음

→ 전체 구로 확장하여 사업을 유치하게 될 경우, 서울시 전체 구의 데이터를 포괄하여 재분석하여 간선, 지선 버스까지 확대 적용할 예정임.

사업지 선정 시 실제 도로 사정이나 정류소 형태를 고려하지 않았음

→ 향후 사업 유치할 경우, 정류소 타입과 입지를 기준을 두고 나눠 분석에 활용할 예정임.

05

한계점 및 향후 계획

1 '미세먼지 프리 버스' 아이템과 '우리동네 공기쉼터' 아이템을 결합한 단일 사업으로 적극적인 공공 홍보 사업을 진행하여 **사업의 실효성**을 높이하고자 함. 홍보를 통해 **대중 교통 이용을 장려**하는 효과도 기대됨.

2 사업 진행 시 국내 공기 청정 및 여과 산업에 종사하는 **중소기업과 협력**하고자 함. 이를 통해 **국내 공기청정관련 기업의 발전**의 길을 마련하여줌으로써 관련 **내수 산업의 경제력 증진**을 도모함.

3 '우리동네 공기쉼터' 아이템의 경우, 여러 **스마트 기술**을 도입하여 시민에게 더욱 큰 편의를 제공하고자 함. 이를 통해 해당 사업이 단순한 공기 청정 공간이 아닌 **스마트 공기쉼터 제공 서비스**로 시민들에게 인식되게 하고자 함.

6. 팀원 소개



06 팀원 소개

향후 계획

TEAM NAME

역삼동 불주먹

역삼동에서 만난 3인방이 뭉쳤다!

경영, 경제, 통계 세 전공의 장점을 살려
데이터 탐색, 분석, 시각화 전반에 걸친
빅데이터 분석을 팀워크를 통해 수행해왔다.

향후 데이터 전문가로서의 길을
걷기 위해 노력하고있다.



Leader 강승연

이화여자대학교 경영학과
기획, 데이터 탐색, 데이터 분석, 데이터
시각화, PPT

이 프로젝트에서 프로젝트 아이디
어링 및 기획과 팀 리딩을 했으며,
전반적 분석 과정에 참여하였다.



서소미

숙명여자대학교 경제, 통계학과
기획, 데이터 탐색, 데이터 분석, 데이터
시각화, PPT

이 프로젝트에서 연구가설의 수
립과 증명을 주도하였으며, 전반
적 분석 과정에 참여하였다.



조유민

숙명여자대학교 경제, 통계학과
기획, 데이터 탐색, 데이터 분석, 데이터
시각화, PPT

이 프로젝트에서 사업지 선정과 시
각화를 주도하였으며, 전반적 분석
과정에 참여하였다.

Reference

데이터 출처

<https://www.bigcontest.or.kr/>

<https://data.seoul.go.kr/dataList/datasetView.do?infId=OA-12912&srvType=S&serviceKind=1¤tPageNo=1&searchValue=&searchKey=null>

<https://data.seoul.go.kr/dataList/datasetView.do?infId=OA-1095&srvType=F&serviceKind=1¤tPageNo=1&searchValue=&searchKey=null>

https://raw.githubusercontent.com/southkorea/southkorea-maps/master/kostat/2013/json/skorea_submunicipalities_geo_simple.json

이미지 출처

<https://www.yna.co.kr/view/AKR20170215159600009>

<://www.arcgis.com/apps/MapJournal/index.html?appid=a98ac517a0984f4cbb68c29a62a040c1>

<https://m.seoul.co.kr/news/newsView.php?id=20161026009011&cp=seoul>

https://upload.wikimedia.org/wikipedia/commons/thumb/f/f6/%ED%83%80%EC%9A%94%EB%B2%84%EC%8A%A4_9401_%EA%B0%80%EB%8B%88_A01.JPG/1600px-%ED%83%80%EC%9A%94%EB%B2%84%EC%8A%A4_9401_%EA%B0%80%EB%8B%88_A01.JPG

http://www.choicenews.co.kr/news/photo/201705/33236_17816_1731.jpg

https://mblogthumb-phinf.pstatic.net/20160311_3/batt7424_1457656931791SBiun_JPEG/image.JPEG?type=w800

https://upload.wikimedia.org/wikipedia/ko/2/22/Inside_the_bus.jpg
