

미추홀구 흡연 민원지역 분석 및 금연구역 추천

2022.02.14

12196

정다인

인천광역시청

1. 분석 배경	3
1.1 분석 배경.....	3
1.2 분석 목적 및 방향.....	4
1.3 분석 결과 활용 방안	4
2. 분석 데이터.....	5
2.1 분석 데이터 목록	5
2.2. 데이터 상세 설명-	6
2.3. 데이터 정제 방안	9
3. 분석 내용	12
3.1 분석 프로세스	12
3.2 분석 내용 및 방법.....	13
4. 분석 결과	16
4.1 민원 분석.....	16
4.2 기초 분석 (EDA)	28
4.3 추천지 분석 결과	35
5. 참고자료.....	38

1. 분석 배경

1.1 분석 배경

■ 흡연 단속 인력 한계로 민원 지속적으로 발생

- 일선인 군구보건소가 코로나19 사태로 인력 문제 발생
- 금연구역 확대됨에 따라 인천시 관리, 단속 한계 증가
- 추후 발생할 민원 예방을 위해 민원 분석을 통한 시민의 요구사항 파악 필요

■ 비흡연자와 흡연자의 갈등 심화

- 실외 금연구역을 확대 <-> 공공 흡연시설 확충
- 민원 중심지를 고려한 과학적 근거 기반 금연구역 지정 필요

■ 인천광역시 중 미추홀구 흡연율과 음주율 심각성

○ 흡연율 / 음주율 현황

- 2020년 기준, 전국 행정구역 상위 10위 [음주율 & 흡연율]

순위	행정구역	음주율(%)	흡연율(%)
1	충청북도 종평군	64.3	29.3
2	서울특별시 마포구	63	17.6
3	인천광역시 미추홀구	63	20.7
4	충청북도 충주시	62.2	22.4
5	경상남도 진주시	61.6	18.4
6	강원도 동해시	61.2	20.5
7	서울특별시 양천구	61.1	14.4
8	서울특별시 영등포구	60.8	15
9	강원도 정선군	60.8	24.7
10	서울특별시 성동구	60.3	18.7

- 인천광역시 중에서 미추홀구가 가장 높은 음주율(%)과 흡연율(%)을 보인다.

1.2 분석 목적 및 방향

■ 민원 데이터 분석을 통해 시민의 요구사항 도출

○ 민원 빈도 분석을 통해 세부 내용 파악

- 핵심 주제어(흡연, 담배꽂초)를 기반 네트워크 분석을 통해 시민의 불만의 원인이 되는 키워드들의 변화와 연관성을 판별

○ 미추홀구 흡연 관련 Hotspot 지도 제작

- 흡연 관련 민원 및 기초 현황을 파악하여 시각화

■ 신규 금연구역 후보지 도출

- 흡연자와 비흡연자의 의견 상충을 과학적 근거로 도출한 금연구역으로 제안

1.3 분석 결과 활용 방안

- 시민의 직접적인 요구사항이인 민원은 추후 정책 개발을 위한 데이터로 활용 가능
- 집중 단속인력 배치 제안 지역으로 효율적인 인력 활용
- 민원 해결의 신속화 • 효율화 추구 및 향후 건의사항을 사전에 해결

2. 분석 데이터

2.1 분석 데이터 목록

구분	활용 데이터	구분	중요도	데이터 소스
개방형 공공데이터	교육/보육시설물 데이터 (초등학교, 중학교, 고등학교,어린이집)	정형/내부	필수	<ul style="list-style-type: none"> - 출처 : 인천광역시 미추홀구 - 유형, 학교명, 설립유형, 도로명주소, 전화번호, 각 학급 수, 위도, 경도 등 - csv파일 - 기준일자: 2021.09.01
	금연구역 데이터	정형/외부	필수	<ul style="list-style-type: none"> - 출처 : 공공데이터 포털 - 금연구역명, 범위, 시군구명, 구분, 위도, 경도, 소재지번호주소, 관리기관명 등 - csv파일 - 기준일자 : 2021.05.30
	흡연구역 데이터	정형/외부	필수	<ul style="list-style-type: none"> - 출처: 인천광역시 미추홀구 - 시설명, 위치, 관리기관 - csv파일 - 기준일자 : 2021.05.25
	교통시설물 지하철 위치정보	정형/외부	필수	<ul style="list-style-type: none"> - 출처: 국가철도공단 - 호선, 역명, 좌표 - csv 파일 - 기준일자: 2021.11.23
	교통시설물 지하철 수송정보	정형/외부	선택	<ul style="list-style-type: none"> - 출처: 국가철도공단 - 역명, 승차여객수, 하차여객수, 여객수 - csv 파일 - 기준일자 : 2019.12.31
	행정동경계	정형/외부	필수	<ul style="list-style-type: none"> - 출처 : 국토교통부 - 국가표준 데이터 - shp파일

구분	활용 데이터	구분	중요도	데이터 소스
	일반시설물 데이터 음식점, 담배소매업소, PC방)	정형/외부	필수	- 출처 : 인천광역시 미추홀구 - 상호명, 도로명주소, 전화번호, 위도, 경도 - csv파일 - 기준일자 : 2021.09.27 (pc) 2021.05.27(담배), 2021.05.31(음식점)
지자체 데이터	SKT 유동인구 데이터	정형/외부	필수	- 출처 : 인천광역시 미추홀구 - 날짜, 시간, 연령, 인덱스ID, H_POP, W_POP, V_POP, 시/군/구, 읍/면/동, 법정동 코드, 행정구역 코드, 면적, length - csv파일 - 2020.02.01 ~ 2021.08.31
	민원 데이터	비정형/ 정형 /외부	필수	- 출처 : 인천광역시 미추홀구 - 미추홀구 새올민원 데이터 - csv파일 - 2019.10.30 ~ 2020.11.23

2.2. 데이터 상세 설명-

- 민원 데이터, 총 547 건
- 교육/보육시설물 데이터, 총 259 건
- 유동인구 데이터(20.2.1 ~ 21.8.31), 총 19,922,925 건
- 금연구역 데이터, 총 50,000 개
- 인천광역시 미추홀구 음식점 데이터(2021), 총 4548 건
- 인천광역시 미추홀구 담배소매업소 데이터(2021), 총 864 건
- 인천광역시 미추홀구 PC 방 데이터(2021), 총 145 건

1) 민원 데이터[.CSV]

- 민원 일자, 민원 경로, 민원 주제, 내용, 처리 한계기간, 부서

ldadng_ym	recv_path	minwon_sj	minwon_c	process_limit	dept_g	dept_nm	ans_mod_dtime
20191030	모바일접수민원	[생활불편신고]쓰레기 방기	공원내 흡	2019-11-01 18:00	주관	건설교통국	2019-10-30 17:34
20191031	모바일접수민원	[생활불편신고]기타불편	기타불편	2019-10-31 18:00	주관	복지환경국	2019-10-31 17:23
20191031	이송이관민원	[이송이관] 간석역 1번출구	간석역 1번	2019-10-31 18:00	주관	보건소 건강	2019-10-31 17:59
20191031	자체접수민원	주안동3동 쓰레기무단투기	쓰레기무단	2019-11-06 18:00	주관	주안3동	2019-10-31 18:01
20191101	모바일접수민원	[생활불편신고]기타생활불	기타생활불	2019-11-05 18:00	주관	건설교통국	2019-11-01 15:58
20191101	이송이관민원	[이송이관] 동네 포장마차	첨부파일0	2019-11-01 18:00	주관	보건소 위생	2019-11-01 10:42
20191101	이송이관민원	[이송이관] 동네 포장마차	첨부파일0	2019-11-01 18:00	협조	도시재생국	2019-11-01 9:03
20191101	자체접수민원	용현동 메가커피	이 구역에	2019-11-01 18:00	주관	용현5동	2019-11-01 16:42
20191105	이송이관민원	[이송이관] 주안역 앞 흡연	주안역 앞	2019-11-05 18:00	주관	보건소 건강	2019-11-05 12:01
20191106	모바일접수민원	[생활불편스마트폰신고]가	※ 해당 민	2017-01-17 18:00	주관	보건소 건강	2017-01-12 18:07

2) 교육/보육시설물 데이터[.CSV]

- 유형, 등록군구, 학교명, 설립유형, 우편번호, 도로명주소, 전화번호, 학급 수, 위도, 경도, 구분

유형	등록군구	학교명	설립유형	우편번호	도로명주소	전화번호	일학년학급	이학년학급	삼학년학급	사학년학급	오학년학급	육학년학급	특수학급수	위도	경도
남부	미추홀구	인천광역시공립	22222	인천광역시032-875-25	4	6	5	5	6	2	37.44239	126.6737			
남부	미추홀구	인천광역시공립	22163	인천광역시032-887-98	8	9	9	8	8	2	37.46372	126.6557			
남부	미추홀구	인천광역시공립	22187	인천광역시032-629-08	7	8	9	7	7	3	37.45318	126.6408			
남부	미추홀구	인천광역시공립	22165	인천광역시032-874-27	7	6	6	6	7	2	37.46314	126.6678			
남부	미추홀구	인천광역시공립	22134	인천광역시032-424-07	6	7	6	6	5	3	37.46235	126.6867			
남부	미추홀구	인천광역시공립	22233	인천광역시032-425-25	4	5	5	5	5	2	37.43989	126.6828			
남부	미추홀구	인천광역시공립	22208	인천광역시032-874-03	4	5	4	5	4	2	37.45245	126.6634			
남부	미추홀구	인천광역시공립	22116	인천광역시032-867-03	4	4	4	4	5	2	37.46902	126.6685			
남부	미추홀구	인천광역시공립	22228	인천광역시032-424-74	4	4	5	5	4	3	37.44955	126.685			
남부	미추홀구	인천광역시공립	22148	인천광역시032-629-04	4	5	5	5	4	1	37.45465	126.6739			
남부	미추홀구	인천광역시공립	22114	인천광역시032-868-911	8	8	7	7	5	2	37.47399	126.6608			
남부	미추홀구	인천광역시공립	22226	인천광역시032-867-22	2	2	3	3	4	2	37.43892	126.6719			
남부	미추홀구	인천광역시공립	22168	인천광역시032-876-62	2	3	2	2	2	2	37.45944	126.6572			

3) 유동인구 데이터 [.CSV]

- TD_YMD, TIME, AGE, HCODE, H_POP(주거인구/야간체류지), W_POP(직장인구/주간체류지), V_POP(방문인구), DO_NAME, GU_CODE, DO_CODE, area, length

연번	구분	의료기관명	소재지	X	Y	표준신주소	표준구주소
1	종합병원	인하대학교의료원	인천광역시	126.6342	37.45886	인천광역시	인천광역시 중구
2	종합병원	인천기독병원	인천광역시	126.632	37.47156	인천광역시	인천광역시 중구
3	종합병원	의료법인성수	인천광역시	126.6458	37.4817	인천광역시	인천광역시 동구
4	종합병원	인천광역시의료원	인천광역시	126.6692	37.47841	인천광역시	인천광역시 동구
5	종합병원	현대유비스병원	인천광역시	126.6432	37.46176	인천광역시	인천광역시 미추홀
6	종합병원	인천사랑병원	인천광역시	126.6807	37.4622	인천광역시	인천광역시 미추홀

4) 금연구역 데이터[.CSV]

- 금연구역명, 금연구역범위상세_x, 시도명_x, 시군구명_x, 금연구역구분_x, 금연구역지정근거명_x, 금연구역면적_x, 위반과태료_x, 위반신고전화번호_x, 소재지도로명주소_x, 소재지지번주소_x, 관리기관명_x, Y, X

금연구역명	금연구역구분	시도명_x	시군구명_x	금연구역구분	금연구역구분	금연구역면적	위반과태료	위반신고전화번호	소재지도로명주소	소재지지번주소	관리기관명	Y	X
관교어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	1698.7	30000	032-880-5413	인천광역시	인천광역시	37.44593	126.6909			
관선어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	1610.6	30000	032-880-5413	인천광역시	인천광역시	37.43786	126.6875			
관청어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	1485.6	30000	032-880-5413	인천광역시	인천광역시	37.43979	126.6806			
나리어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	1504	30000	032-880-5413	인천광역시	인천광역시	37.44019	126.6746			
낙섬어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	655.7	30000	032-880-5413	인천광역시	인천광역시	37.45277	126.6363			
노적어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	2426.5	30000	032-880-5413	인천광역시	인천광역시	37.43759	126.6527			
다솔어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	2370.2	30000	032-880-5413	인천광역시	인천광역시	37.46086	126.695			
도천어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	1800	30000	032-880-5413	인천광역시	인천광역시	37.43887	126.6783			
동화어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	1555.7	30000	032-880-5413	인천광역시	인천광역시	37.45766	126.6726			
두리어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	1052.2	30000	032-880-5413	인천광역시	인천광역시	37.46	126.6926			
매소홀어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	3006.9	30000	032-880-5413	인천광역시	인천광역시	37.44192	126.6546			
모리포어린이공원 전체	인천광역시 미추홀구	도시공원	인천광역시	5589.7	30000	032-880-5413	인천광역시	인천광역시	37.43998	126.6594			

5) 인천광역시 미추홀구 음식점 데이터[.CSV]

- 유형, 상호명, 도로명주소, 지번주소, 위도, 경도

연번	유형	상호명	도로명주소	지번주소	위도	경도
1	일반음식점	논골고기촌	인천광역시	인천광역시	37.46391	126.6466
2	일반음식점	봉이빈대떡	인천광역시	인천광역시	37.45755	126.691
3	일반음식점	참설렁탕	인천광역시	인천광역시	37.44738	126.6784
4	일반음식점	영화반점		인천광역시	37.46319	126.641
5	일반음식점	동락반점	인천광역시	인천광역시	37.46053	126.6463
6	일반음식점	성진물텀벙	인천광역시	인천광역시	37.45601	126.6492
7	일반음식점	밥푸면&박스밥		인천광역시	37.46803	126.6614
8	일반음식점	동백식당	인천광역시	인천광역시	37.46547	126.6631
9	일반음식점	경안각	인천광역시	인천광역시	37.4656	126.6627

6) 인천광역시 미추홀구 담배소매업소 데이터[.CSV]

- 상호명, 도로명주소, 위도, 경도

연번	상호명	도로명주소	위도	경도
1	세븐일레븐	인천광역시	37.46654	126.6611
2	BOROO	인천광역시	37.45176	126.6552
3	늘푸른	인천광역시	37.45291	126.651
4	GS25 인천	인천광역시	37.45225	126.69
5	(주)이마트	인천광역시	37.46955	126.6645
6	씨유 문학점	인천광역시	37.43825	126.6873
7	슈퍼킹	인천광역시	37.45215	126.681
8	씨유 도화점	인천광역시	37.46323	126.6707

7) 인천광역시 미추홀구 PC 방 데이터[.CSV]

- 상호명, 도로명주소, 전화번호, 위도, 경도

연번	상호명	도로명주소	전화번호	위도	경도
1	철구PC주	인천광역시 미추홀구		37.46357	126.681
2	프리미엄본	인천광역시	070-4077-	37.4515	126.6472
3	라네이즈P	인천광역시 미추홀구		37.45705	126.6391
4	아프리카T	인천광역시	070-4119-	37.46285	126.6806
5	맥스피드P	인천광역시 미추홀구		37.46993	126.6602
6	킹덤PC카페	인천광역시	032-548-9	37.45262	126.6356
7	라이프PC방	인천광역시 미추홀구		37.47031	126.6627
8	더엑스PC방	인천광역시	070-5180-	37.46574	126.6639
9	애타PC카페	인천광역시 미추홀구		37.45238	126.6578

2.3. 데이터 정제 방안

2.3.1 민원 데이터 전처리

민원 텍스트 분석에 앞서 R을 이용하여 데이터 전처리를 진행

1) 형태소 분석 및 추출

-Excel, R을 이용하여 불필요한 문자, 연속된 공백, 기호 등을 제거하는 클렌징 작업 실시

-민원 내용을 SimplePos09() 함수를 적용하여 9개의 형태소 단위로 품사 구분하여 토큰화 실시

-민원 내용이 길기 때문에 출력 결과를 알아보기 쉽게 하기 위해 데이터를 tibble 구조로 변환

2) 불용어 제거

-str_match() 함수를 활용하여 '명사'만 추출하고, filter() 함수를 활용하여 '것', '수'와 같이 의미가 없는 한 글자로 된 의존명사 분석 대상에서 제외

3) 문서별 최소 단어 수 설정

-단어 빈도수가 5 이하, 일부 단어(민원방법 관련) 제외 후, 빈도 순서 변수 생성하여 빈도 확인

```

NEWS <- read.csv("D:/개 인 인 천 시 청 /데 이 터 /민 원 데 이 터 /saeol_minwon_B.csv", encoding = 'euc-kr', header = T)
summary(NEWS)

TEXTFILE <- NEWS$minwon_cn

tbl_TEXTFILE <- TEXTFILE %>%
  simplePos09 %>% # 품 사 구 분 함수. simplePos09()는 9개 품 사로, simplePos22()는 22개 품 사로 구 분
  melt %>% # 전 체 자 료를 tibble 형 태로 저 장
  as_tibble %>% # 전 체 자 료를 tibble 형 태로 저 장
  select(3, 1) # 실 제 분 석 에 필 요한 3열과 1열만 따 로 저 장

tbl_TEXTFILECOUNT0 <- tbl_TEXTFILE %>% # tbl_TEXTFILE 데 이 터 사 용
  mutate(noun=str_match(value, '([가-힣]+)/N')[,2]) %>% # 명 사 형 (/N) 자 료만 골 라 낸 후 /N 빼 고 저 장
  na.omit %>% # 비 어 있 는 값 (NA) 제 외
  filter(str_length(noun)>=2) %>% # '것', '수' 와 같 이 별 의 미 가 없 는 의 존 명 사 를 제 외 하 기 위 하 여 한 글 자 단 어 는 제 외
  count(noun, sort=TRUE)

## 명 사 형 자 료만 골 라 내 어 카운트
head(tbl_TEXTFILECOUNT0, 10)

tbl_TEXTFILECOUNT1 <- tbl_TEXTFILECOUNT0 %>% filter(n>=5) %>% filter(!noun %in% c(
  "관련", "첨부한", "저장", "갤러리", "정보시스템", "동영상", "일대", "주변",
  "첨부파일", "복사", "촬영", "붙여넣으면", "해당", "브라우저", "주소창", "개월", "별도",
  "보관해주시길", "담당자분", "인천광역시", "삭제", "부터"))

## 50대 키워드만 선정
tbl_TEXTFILECOUNT2 <- tbl_TEXTFILECOUNT1[1:50,]

# 워드클라우드 그리기
wordcloud2(tbl_TEXTFILECOUNT2, fontFamily="Malgun Gothic", size = 0.5, minRotation=0, maxRotation=0)

```

2.3.2 민원 데이터 공간 정보 처리

- 민원 데이터 공간 분석을 위해 상세 위치가 따로 분류되어 있지 않는 경우, 민원내 용에서 근처 위치 값 도출
- 위치정보, 민원 내용 속 장소를 유추할 수 있는 내용이 부재하거나 타 구역 민원으 로 이송이관 된 데이터를 제거(47개)
- GeocodingTool을 이용해 주소 값을 기준으로 위경도 데이터(X, Y) 추출
- QGIS [위치에 따라 속성 결합], [필드 값으로 속성 결합]을 사용

2.3.3 유동인구 데이터

- 대용량 파일이므로 python을 이용하여 민원 데이터와 같은 시기만 추출하여, 동 별 로 분류
- 행정동별로 월별 평균을 계산하여 변수 생성
- 미추홀구 행정동으로 한정

	A	B	C	D	E	F	G	H	I	J	K	L
1	행정동	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
2	가정1동	102.2241	89.83871	92.36806	108.0961	104.241	113.994	105.4651	154.5208	168.8965	163.9181	181.4698
3	가정2동	18.64009	19.79637	19.31389	24.41196	22.56597	23.89583	17.96505	17.16944	18.71237	18.78472	18.2379
4	가정3동	25.20546	21.08199	23.05208	23.67944	22.20417	25.02218	26.13374	25.77639	26.22043	23.21389	26.22782
5	가좌1동	174.9792	175.1761	170.2056	176.3629	182.5382	182.3253	162.1996	169.4576	156.2621	172.8146	170.9677
6	가좌2동	61.75934	56.29503	58.025	67.53091	66.74653	71.10417	69.70027	70.41111	70.71707	66.49236	64.00739
7	가좌3동	111.6573	107.5867	111.8701	129.334	129.9868	129.168	112.9388	122.2153	118.8676	118.0944	121.7547
8	가좌4동	78.37213	76.9664	81.84444	96.36223	95.97847	86.6875	75.99597	84.28889	77.75067	74.31528	79.69556
9	간석1동	125.8175	113.7702	105.4951	135.0551	122.1243	120.0417	119.8958	121.5625	123.1573	127.3882	116.1337
10	간석2동	118.8678	112.4556	138.8833	140.9651	143.8174	152.5027	142.5437	124.934	128.168	131.9833	121.7614

2.3.4 교육시설물 데이터 및 보육시설물 데이터

- 시군구명을 인천광역시 미추홀구로 한정하여 추출
- 데이터가 각 따로 존재하므로 초등학교, 중학교, 고등학교, 보육시설을 구분변수 생성하여 분류 후, '등록군, 학교명, 위도, 경도' 이외 컬럼 제거 후 데이터 결합
- 행정동별/격자별 교육시설물 건수를 구하기 위해 [폴리곤에 포함하는 포인트 개수 계산]을 이용하여 50m 격자 기준으로 교육시설물 위치데이터 count

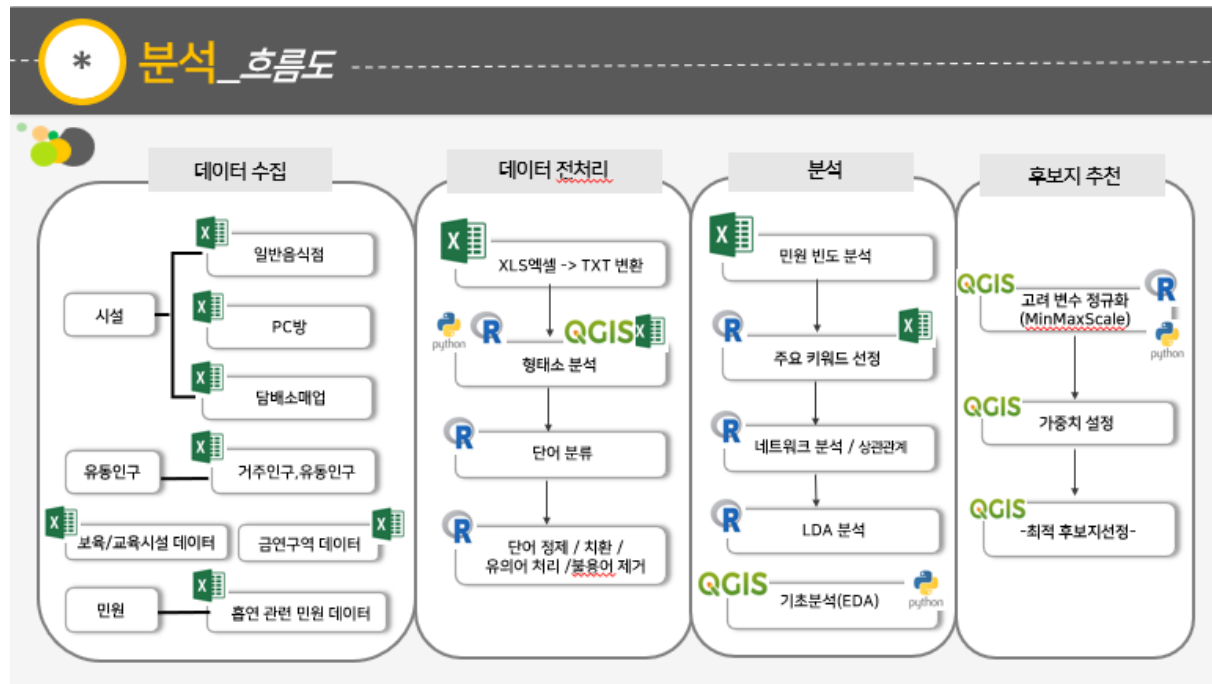
2.3.5 금연구역 데이터

- 시군구명을 인천광역시 미추홀구로 한정하여 추출(917개)하여 동별로 분류
- 행정동별/격자별 금연구역 데이터 건수를 구하기 위해 [폴리곤에 포함하는 포인트 개수 계산]을 이용하여 50m 격자 기준으로 금연구역 위치데이터 count

2.3.6 미추홀구 음식점, 담배소매업소, PC 방 데이터

- qgis 를 이용하여 행정동 경계 기준 공간 위치 데이터[.shp] 와 일반 시설물 데이터를 결합해 일반시설물 데이터를 시각화

3. 분석 내용



3.1 분석 프로세스

흡연 민원 지역 분석 및 금연구역 추천을 위해 데이터 수집, 전처리, 분석 그리고 시각화 순으로 진행

1) 데이터 준비 및 EDA 실시

2) 전처리 및 분석 :

- [민원데이터]를 정형화 데이터 형태로 재표현하기 위해 형태소 분석 및 추출, 불용어 제거, 문서별 최소 단어 수 설정
- QGIS를 활용한 버퍼분석, 중첩분석, 공간분할 실시

* 지오코딩(Geocoding): 주소지를 토대로 호출지와 목적지 X, Y 좌표 생성

* 기초분석(EDA) 및 시각화: 수집된 데이터를 기반으로 시각화를 통해 밀집지역 파악

3) 텍스트마이닝

- 민원 빈도 분석 : 형태소 분석기를 통해 자주 사용된 중요한 단어 파악
- 의미망 분석 : 단어의 관계를 이용해 의미망을 만들어 [동시 출현 단어 분석], [동시 출현 네트워크], [단어 간 상관 분석] 을 실시. 단어 간 관계를 파악하고 텍스트 맥락을 이해
- 토픽 모델링(LDA) : 텍스트를 토픽별로 분류해 최적의 토픽 수를 정하고, LDA 모델을 만들어 토픽별 주요 단어를 살펴 봄. 흡연 관련 민원의 세부 내용을 분류함

4) 추천지역 도출 및 시각화

- 민원 데이터를 고려한 금연구역 추천지 도출

3.2 분석 내용 및 방법

■ 민원 분석

○ 민원 빈도 분석

- 민원 중 명사만 추출 후 그래프 형성'
- 주요 단어 살펴보기

○ 의미망 분석

- 유의어 처리 후 동시 출현 빈도를 구하여 특정 단어와 자주 함께 사용된 단어 분석
- 빈도 수가 25이상이고 공백을 제외한 네트워크 그래프 데이터 생성 후 그래프 시각화
- 연결 중심성, 커뮤니티 변수 생성하여 표현
- 연결 중심성이 높은 네트워크의 주요 단어 살펴본 후, 관심단어 목록 설정

- 파이계수 생성하여 파이 상관분석 실시

○ 토픽 모델링(LDA)

- Corpus(말뭉치)를 통한 데이터 유형화
- LDA분석을 통한 최적 토픽 수 도출 및 토픽 구조화

-

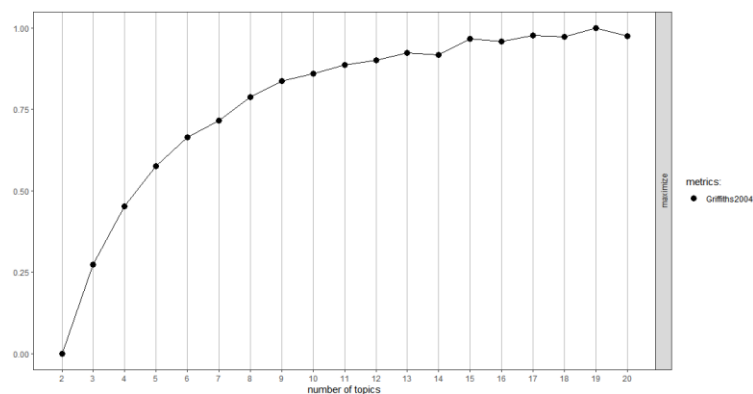


그림 - 최적 토픽 그래프

■ 미추홀구 흡연 hotspot 지도

- 행정동별 / 격자별 / 시간대별 민원데이터를 시각화하여 주요 흡연 관련 민원발생지 추출

■ 금연구역 후보지 도출

- 민원 데이터와 함께 현재 금연구역이 지정 된 곳 중 상위 3분류인 일반음식점, PC방, 담배소매업 수와 보호구역인 교육시설, 보육시설 수를 고려함
- 금연구역 지정 후 행정의 효율성을 위해 유동인구를 고려함

○ 가중치 부여 방법

→ 유동인구 수를 행정동 별로 월평균 수를 구한 후, 비율을 가중치로 사용

$$H_POP(\text{주거인구}) : W_POP(\text{직장인구}) : V_PO(\text{방문인구}) = 1 : 1 : 1$$

→ 격자(50m) 별 일반시설과 교육/보육시설 수를 count 후, minmax scaling

→ 격자에 지하철 위치 기준으로 거리별 차등 점수 부여한 y_4 생성

주위 버퍼 100m 생성 후, 해당 격자 [위치로 선택]하여 점수(1) 부여

주위 버퍼 500m 생성 후, 해당 격자 [위치로 선택]하여 점수(0.33) 부여

[금연 구역 필요 지수]

$$= y_1(\text{유동인구}) + y_2(\text{일반시설}) + y_3(\text{가중시설}) + y_4(\text{거리가점}) + y_5(\text{민원발생})$$

y	y_1 유동인구			y_2 일반시설			y_3 가중시설		y_4 거리가점	y_5 민원발생
	주거 인구	직장 인구	방문 인구	일반음식점	Pc방	담배소매업	교육시설	보육시설	지하철 위치	민원발생 건수
비율	1/3	1/3	1/3	1	1	1	1	1	1	2
총비율	1			3			2		1	2

그림 1 - 비율 표

■ 금연 구역 단속 강화 후보지 도출

- 금연 구역 필요 지수 상위 10위 구역 중 이미 금연구역으로 설치된 곳은 단속 강화 필요

4. 분석 결과

4.1 민원 분석

■ 핵심 키워드 도출 및 민원 빈도 분석

○ 2019년 ~ 2020년 다중출현 키워드와 빈도

-



Figure 1 -top 50 키워드 시각화

- 2019년 ~ 2020년 해에 따른 흡연 관련 자주 언급된 민원 상위 50개를 추출하여 [그림 1]과 같이 도출
- '담배', '쓰레기', '담배꽂초' 등 환경위험요소 단어가 높은 출현빈도를 보이고, '공원', '건물', '골목', '벤치' 등 주요 민원 장소에 관한 단어가 보임
- '학교', '아이들', '학생들', '청소년' 등 상대적으로 위협에 취약한 미성년자들이 자주 언급됨

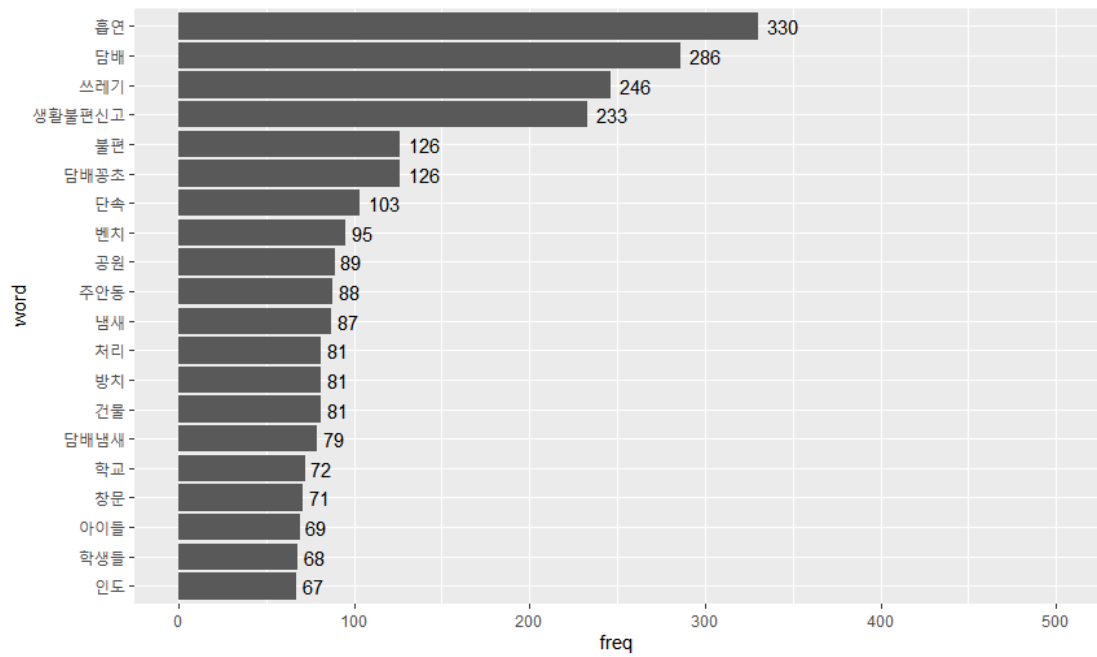


Figure 2 - 민원 빈도 그래프 [상위 20개]

순위	word	freq
1	흡연	330
2	담배	286
3	쓰레기	246
4	생활불편신고	233
5	담배공초	126
6	불편	126
7	단속	103
8	벤치	95
9	공원	89
10	주안동	88
11	냄새	87
12	건물	81
13	방치	81
14	처리	81
15	담배냄새	79
16	학교	72
17	창문	71
18	아이들	69
19	학생들	68
20	인도	67

■ 흡연 관련 핵심 단어 네트워크 분석

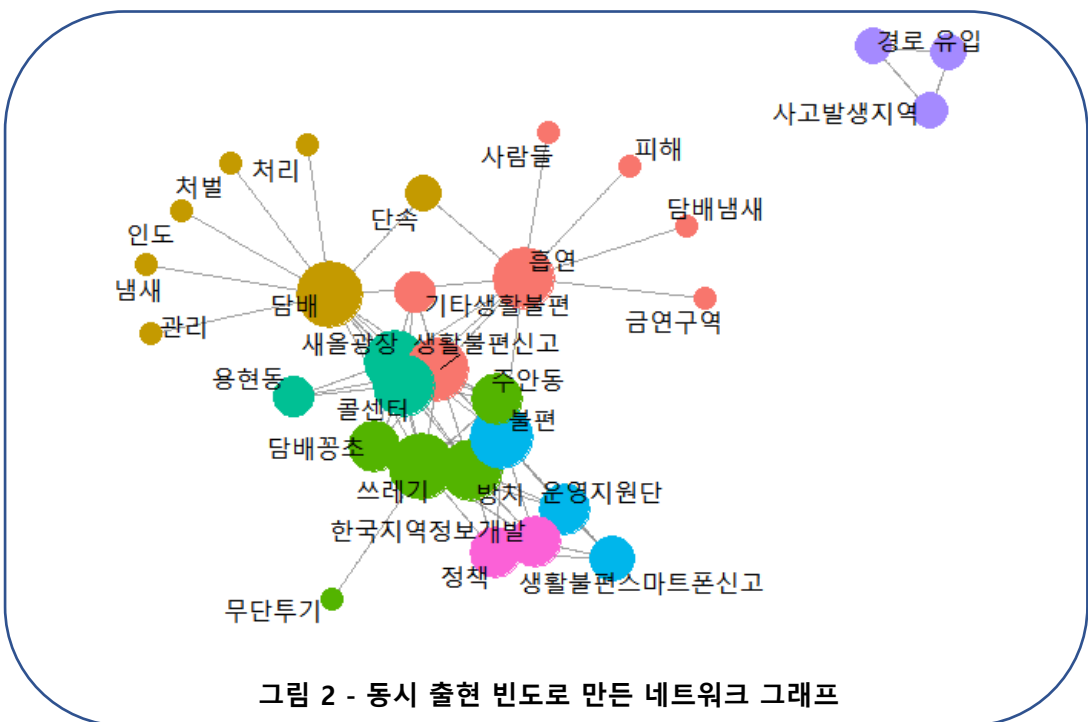
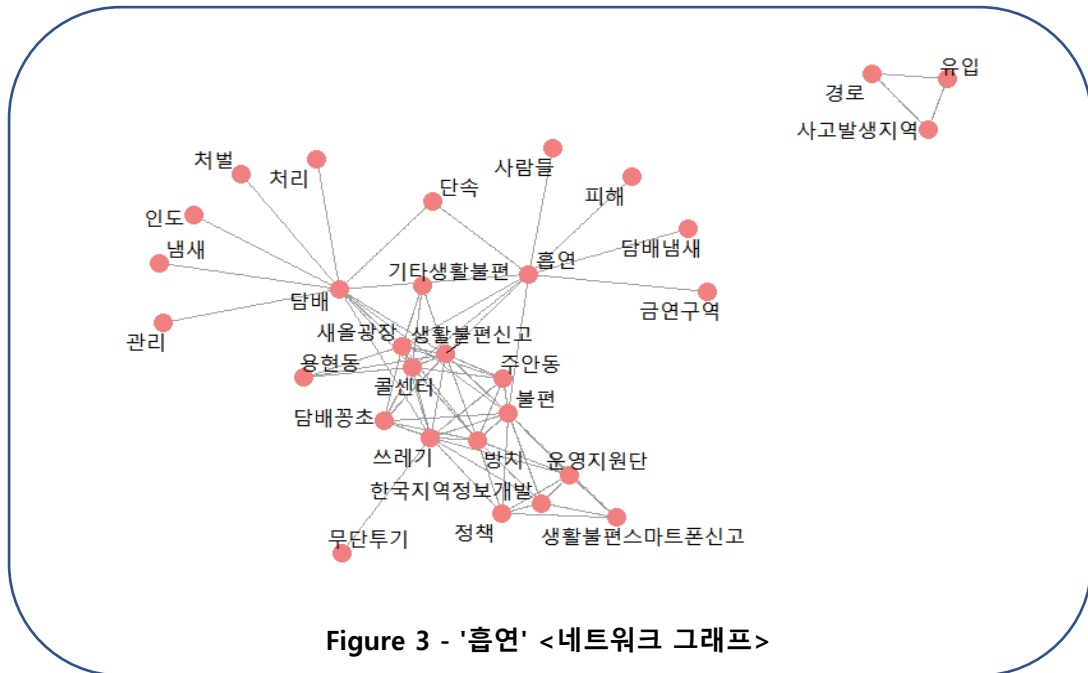
○ '흡연' 동시 출현 단어 및 빈도

순위	단어 1	단어 2	N(빈도)
1	흡연	불편	67
2	흡연	담배	57
3	흡연	단속	38
4	흡연	금연구역	27
5	흡연	피해	26
6	흡연	담배냄새	25
7	흡연	사람들	25
8	흡연	주민	24
9	흡연	건물	22
10	흡연	냄새	22

* '생활불편신고'와 '불편'은 동의어 처리

- ['흡연' - '불편'] 동시 출현 빈도가 67로, 감정적 단어가 가장 자주 언급됨'
- '흡연' 동시 출현 단어 중 '단속', '피해', '담배냄새'가 상위에 위치하는 것을 보아, 흡연 관련 민원은 대부분 간접흡연 피해 호소 및 단속요청 민원임
- '주민', '건물' 키워드가 동시 출현한 것으로 보아, 주택 내에서 간접흡연 고충이 있음
- ['흡연' - '금연구역']의 동시 출현 빈도가 27로, 금연구역에서 흡연자에 대한 신고가 다량 발생. 금연구역에 대한 흡연자의 인식이 부족함

○ 흡연 민원 키워드 간 네트워크 관계



- 동시 출현 빈도 : 자주 사용된 단어 (흡연) 중심으로 관계를 표현

- 연결 중심성 및 커뮤니티 구조 분석을 실시 함
- 도출된 키워드를 군집화한 결과, 전체 민원은 총 7개의 유형으로 군집화
- 빈번하게 함께 사용되고 있음을 알 수 있으며, 비슷한 유형의 조합이라고 해석할 수 있음

그룹1	그룹2	그룹3	그룹4	그룹5	그룹6	그룹7
생활불편신고	쓰레기	담배	새울광장	불편	경로	정책
흡연	방치	단속	콜센터	운영지원단	사고발생지역	한국지역정보개발
기타생활불편	담배꽂초	냄새	용현동	스마트폰신고	유입	
금연구역	주안동	처리				
피해	무단투기	관리				
담배냄새		처벌				

표 3 - 커뮤니티 구조

- <그룹1>에서 나타나는 키워드는 '생활불편신고', '흡연', '금연구역', '피해' 등이며, 간접흡연으로 인한 직접적인 피해 호소 관련 민원으로 나타남.
- <그룹2>에서 '쓰레기', '담배꽂초', '무단투기', '주안동'이 커뮤니티 구조를 이루는 것을 보아 주안동에서 흡연민원 중 담배꽂초 무단투기가 주요 민원임
- <그룹3>에서 금연구역에서 '담배'에 관한 처벌 및 관리 강화에 대한 민원을 나타남
- '단속'이 '관리'보다 높은 centrarity(중심성)를 갖는 것을 보아 [단속(4) > 관리(2)] 사전이행보다 사후이행인 단속에 대한 요구가 더 큰 민원 구조임
- <그룹4> '새울광장' - '콜센터' - '용현동' 키워드가 선으로 연결되어 삼각형 구조를 이뤄 단어가 빈번하게 함께 사용되고 있음
- <그룹5> 처리부서 '운영지원단'에서 생활불편에 관한 '스마트폰신고' 유형이 빈번함

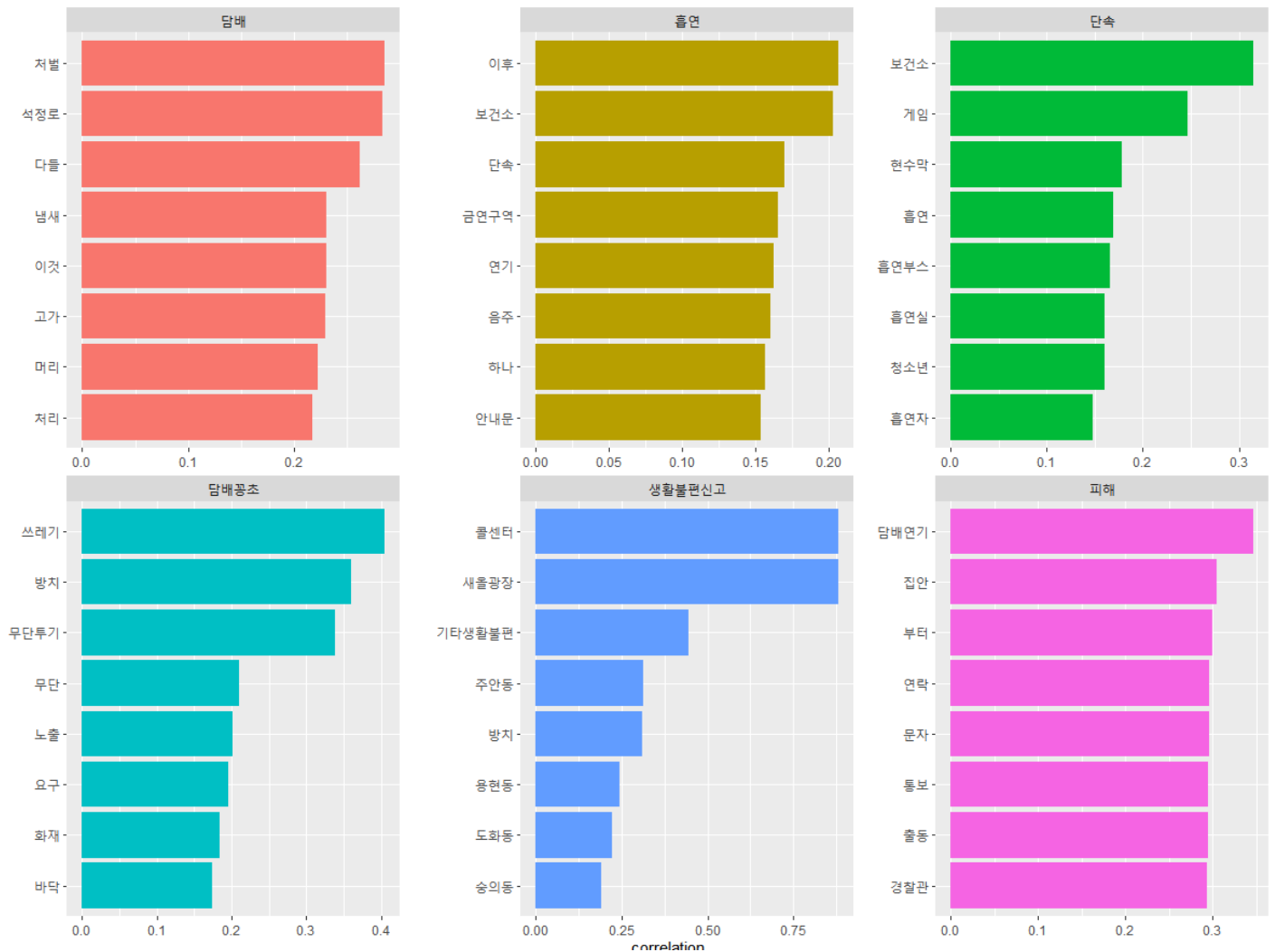
-

순위	word	eigen	centrality
1	쓰레기	1	24
2	담배	0.703	24
3	생활불편신고	0.956	22
4	불편	0.894	22
5	새울광장	0.859	20
6	콜센터	0.859	20
7	방치	0.91	20
8	흡연	0.578	20
9	담배꽂초	0.672	12
10	운영지원단	0.503	12
11	정책	0.503	12
12	한국지역정보개발	0.503	12
13	주안동	0.672	12
14	생활불편스마트폰신고	0.295	8
15	기타생활불편	0.328	6
16	용현동	0.328	6
17	단속	0.157	4
18	경로	0	4
19	사고발생지역	0	4
20	유입	0	4
21	냄새	0.086	2
22	처리	0.086	2
23	관리	0.086	2
24	처벌	0.086	2
25	금연구역	0.071	2
26	피해	0.071	2
27	무단투기	0.123	2
28	인도	0.086	2
29	담배냄새	0.071	2
30	사람들	0.071	2

- 고유벡터 중심성, 연결 중심성

○ 관심단어 6개 상관관계 분석

관심단어 6개 : '담배', '흡연', '단속', '담배꽂초', '생활불편신고', '피해'



- 높은 민원 빈도를 나타난 '담배', '흡연', '단속', '담배꽂초', '생활불편신고', '피해'를 흡연 관련 관심단어 키워드로 선정
- 각 단어의 상관계수가 0.2 이상인 것을 관계성이 있다고 간주함
- ['단속' - '게임'] 이 단어 간 파이 상관계수가 0.246으로, '단속'과 2번째로 관계성이 높음
- ['담배꽂초' - '방치'] 가 0.359으로 담배꽂초로 인한 환경에 관한 문제가 주로 제기 됨

표.

	item1	item2	correlation
1	단속	보건소	0.315301
2	단속	게임	0.246482
3	단속	현수막	0.177664
4	단속	흡연	0.169608
5	단속	흡연부스	0.166211
6	단속	흡연실	0.15982
7	단속	청소년	0.15982
8	단속	흡연자	0.147733

[단속]

minwon_cn

인천남구 주안역부터 신기시장 인근 방들중 일부가 나이에 안맞는 게임물 허용 만 세미만 이용불가 게임을 만 세 미만 인천광역시 남구 석정로 삼덕 상가 지하 층 인도에 풍선간판이 나와있습니다 처리 부탁드립니다 안에 기서는 청소년도 주안 남구 주안동에 부모님이 계신동내라 자주가는데 근처 피시방들 전부 담배 피게하고있습니다 특히 저녁부터 야간에 이스 게임랜드 신고합니다 단속시 문자가 간다하여 인터넷으로 신고합니다 인천광역시 남구 주안동 메인프라자 위 모든 방은 흡연부스 흡연구역 를 설치하여 부스에서만 흡연해야 하는걸로 알고있는데 왜 밤 시간 되면 부스를 이용하

Figure 4 - ['단속' -'게임'] 이 같이 쓰인 민원

- pc방이 금연구역으로 지정 및 흡연부스 설치 지역이지만, 제대로 이행되지 않고 있음
- pc방 중심으로 단속을 강화해야 함

표.

	item1	item2	correlation
1	담배꽁초	쓰레기	0.405234
2	담배꽁초	방치	0.359362
3	담배꽁초	무단투기	0.338888
4	담배꽁초	앨범	0.262994
5	담배꽁초	무단	0.210009
6	담배꽁초	노출	0.201601
7	담배꽁초	요구	0.195765
8	담배꽁초	화재	0.183847

[담배꽁초]

표

	item1	item2	correlation
1	담배	처벌	0.285676
2	담배	석정로	0.283115
3	담배	다들	0.26155
4	담배	냄새	0.230506
5	담배	이것	0.230476
6	담배	고가	0.22891
7	담배	머리	0.222209
8	담배	처리	0.21731

[담배]

- ['담배 - '처벌] 의 파이 상관계수가 0.285로 담배에 관해 부정적인 비흡연자와 흡연자의 갈등이 심각함

표.

	item1	item2	correlation
1	생활불편신고	콜센터	0.881589
2	생활불편신고	새올광장	0.881589
3	생활불편신고	기타생활불편	0.443082
4	생활불편신고	주안동	0.311628
5	생활불편신고	방치	0.310345
6	생활불편신고	용현동	0.242958
7	생활불편신고	도화동	0.222776
8	생활불편신고	앨범	0.222161

[생활불편신고]

표

	item1	item2	correlation
1	피해	담배연기	0.34689
2	피해	집안	0.304456
3	피해	문자	0.296501
4	피해	연락	0.296501
5	피해	출동	0.295081
6	피해	통보	0.282085
7	피해	경찰관	0.280367
8	피해	개인정보	0.27768

[피해]

- '생활불편신고'는 '주안동', '용현동', '도화동'에서 0.2 이상의 유의미한 상관계수를 갖으므로, '생활불편신고' 유형의 민원 주된 발생지임을 알 수 있음
- ['피해' - '집안] 파이 상관계수가 0.304으로 주택가에서 간접흡연이 심각함

표.

	item1	item2	correlation
1	흡연	보건소	0.20301
2	흡연	단속	0.169608
3	흡연	금연구역	0.165717
4	흡연	연기	0.162718
5	흡연	음주	0.160476
6	흡연	안내문	0.153586
7	흡연	벤치	0.153062
8	흡연	피해	0.151434

[흡연]

- '흡연'은 눈에 띄는 상관성을 갖는 단어가 없음

- 네트워크가 너무 복잡하지 않고 관련성이 큰 단어 중심으로 만들어지도록 word_correlation이 0.15이상인 단어쌍만 추출
- 네트워크 그래프를 만들고 중심성과 커뮤니티 추가
- 파이계수로 만든 네트워크 그래프 시각화

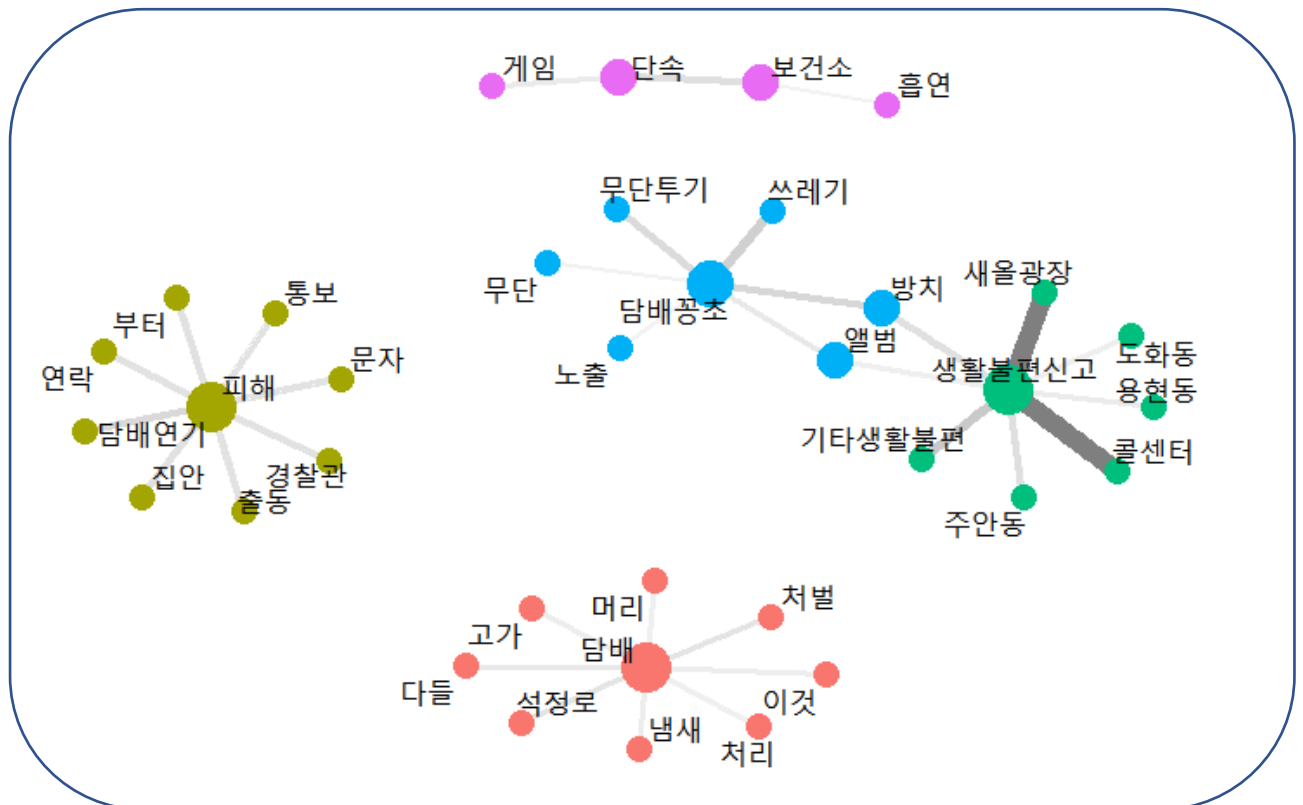
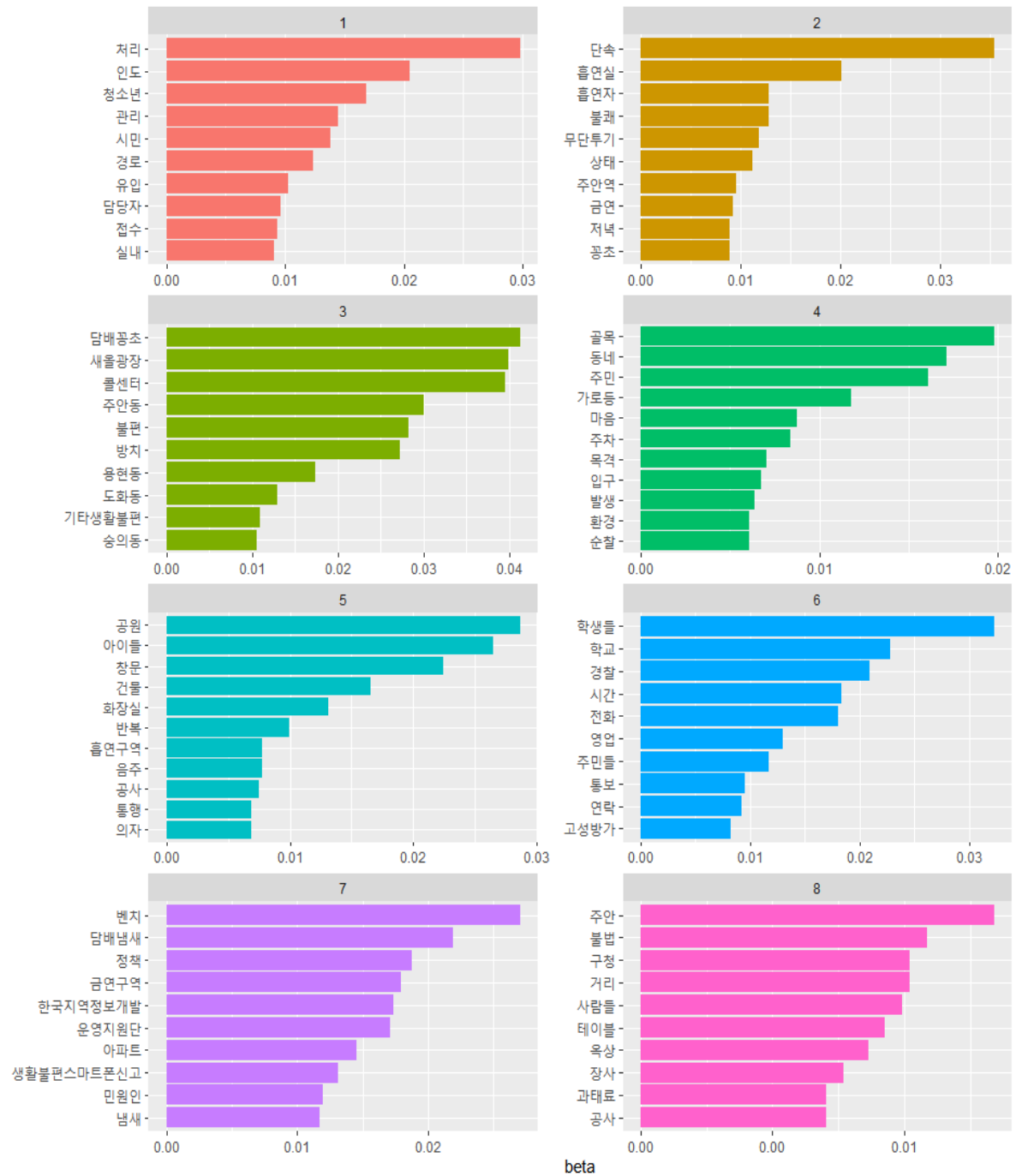


그림 3 - 파이 계수로 만든 네트워크 그래프

- 파이계수 : 관련쌍이 큰 단어쌍이 무엇인지, 단어 군집을 알 수 있음

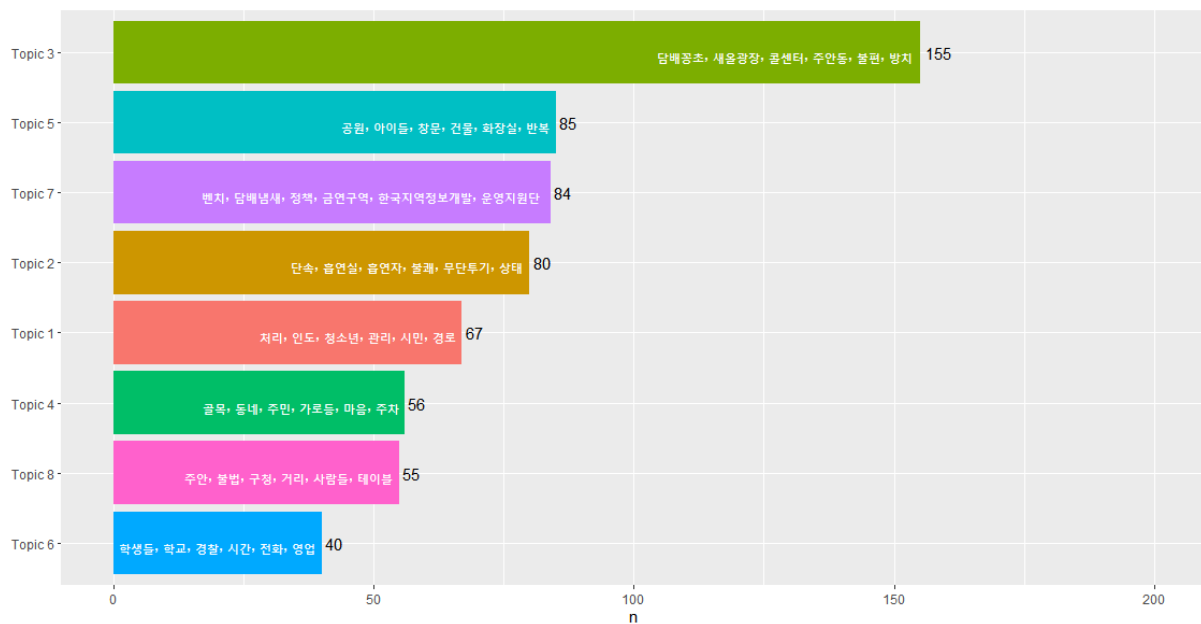
○ LDA



- 토픽2의 '단속', '흡연실', '저녁'을 보면 이 토픽은 단속에 대한 문제점과 관련됨을 알 수 있음
- 등장확률이 높은 토픽3의 "담배공조", "불편", "방치"를 보면 이 토픽은 담배공조로 인한 불쾌감과 관련됨을 알 수 있음

- Topic6의 '학생들', '학교', '경찰'을 보면 이 토픽은 취약계층 우려 문제와 관련됨을 알 수 있음
- 토픽 7의 '주안', '거리', '테이블', '장사'을 보면 이 토픽은 일반음식점으로 인한 불편 문제와 관련됨을 알 수 있음

○ 토픽별 민원 수

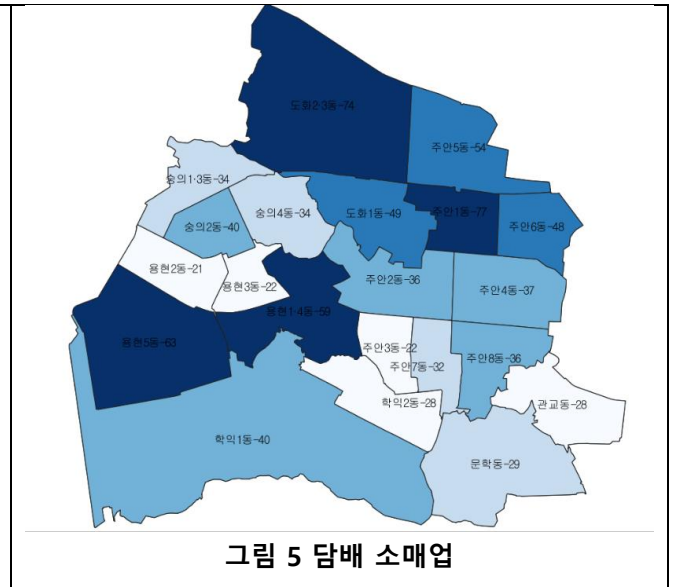
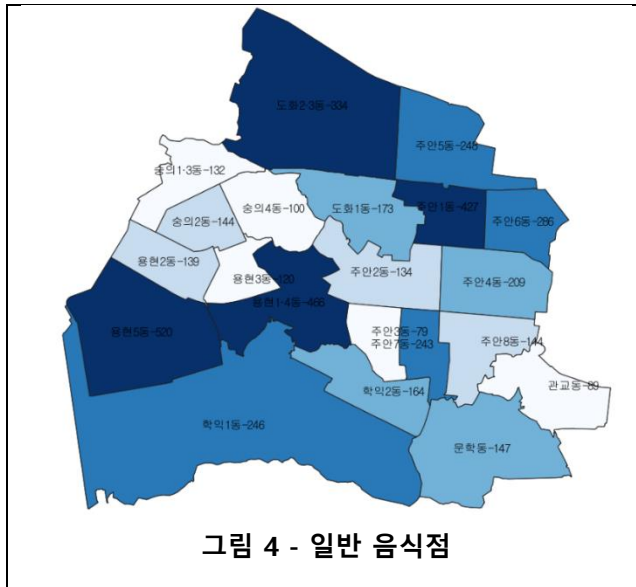


- Topic3이 156개로 민원 담배공초 관련 민원이 155건으로 가장 많음
- Topic5로 공원을 중심으로 건물 주변 담배연기로 인한 주택가 민원이 2번째로 많음
- Topic7도 벤치를 중심으로 간접흡연 불쾌 민원으로 84건이며 Topic5와 유사한 유형임
-

4.2 기초 분석 (EDA)

■ 행정동별 일반 시설물 현황

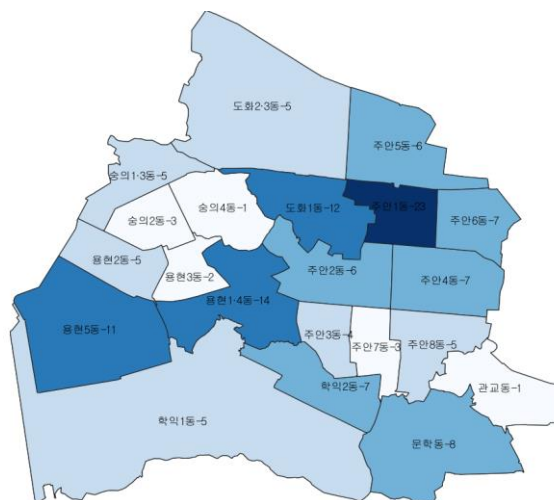
○ 일반 음식점 현황 & 담배 소매업 현황



- 일반 음식점은 용현5동(520), 용현1.4동(466), 주안1동(427), 도화2.3동 (334), 주안6동 (286) 순으로 상위에 위치함

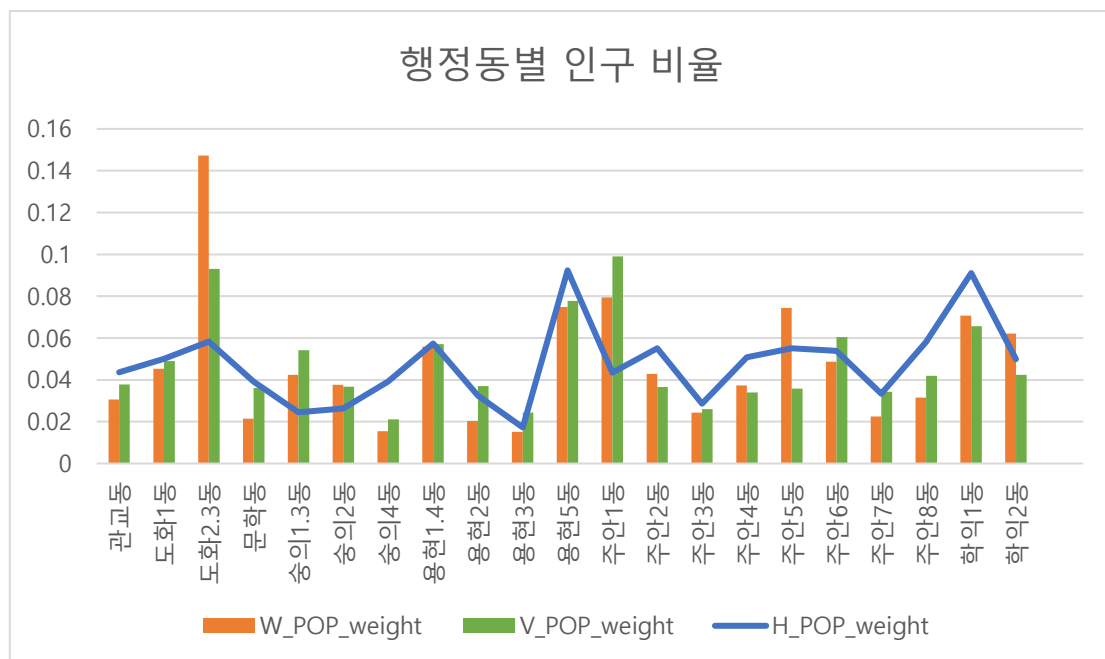
- 담배 소매업은 주안1동(77), 도화2.3동(74), 용현5동(63), 용현1.4동(59), 주안5동(54) 순으로 상위에 위치함

○ pc방 현황



- 주안1동(23)으로 압도적으로 높고, 용현1.4동(14), 도화1동(12), 용현5동(11), 문학동(8) 순으로 상위에 위치함

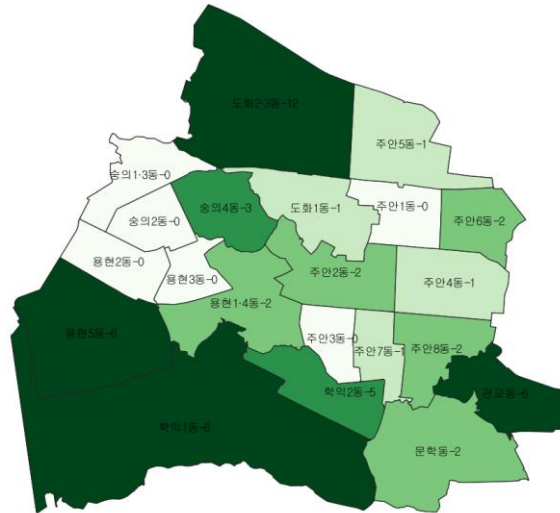
■ 행정동별 유동인구



- 유동인구를 H_POP(주거인구), W_POP(직장인구), V_PO(방문인구)로 분류
- H_POP(주거인구)는 용현 5동, 학익1동 순으로 상위에 위치하며, 도화2.3동, 주안8동, 용현1.4동 순으로 높다.
- W_POP(직장인구)는 도화2.3동이 압도적으로 높고, 주안1동, 용현5동, 주안5동 순으로 높다.
- V_POP(방문인구)는 주안1동, 주안2.3동 순으로 상위에 위치하며, 용현5동, 학익1동 순으로 높다.

■ 행정동별 교육시설물 현황

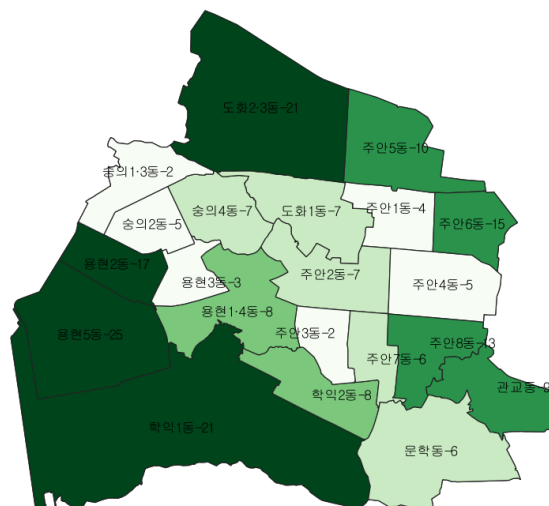
○ 교육시설물 (초등학교, 중학교, 고등학교)



- 도화2.3동(12), 관교동(6), 용현5동(6), 학익1동(6), 학익2동(5) 순으로 순으로 상위에 위치함

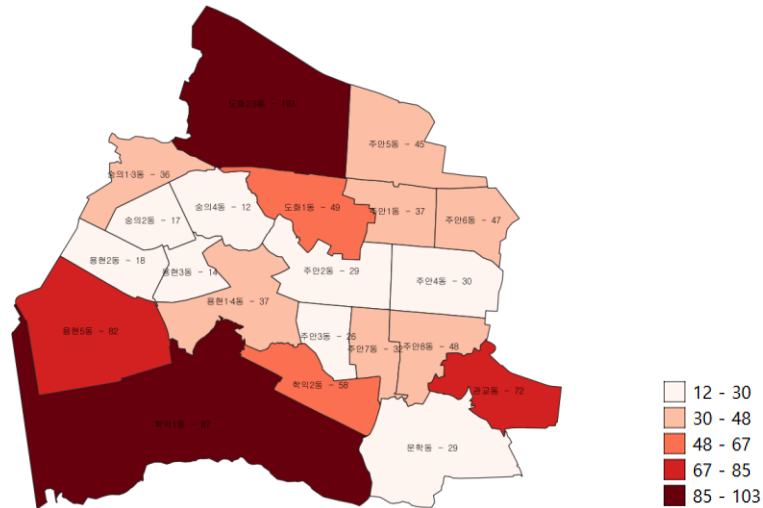
- 일반시설물 위치 중심지인 주안1동, 송의2동은 교육시설물이 존재하지 않음

○ 보육시설 (어린이)



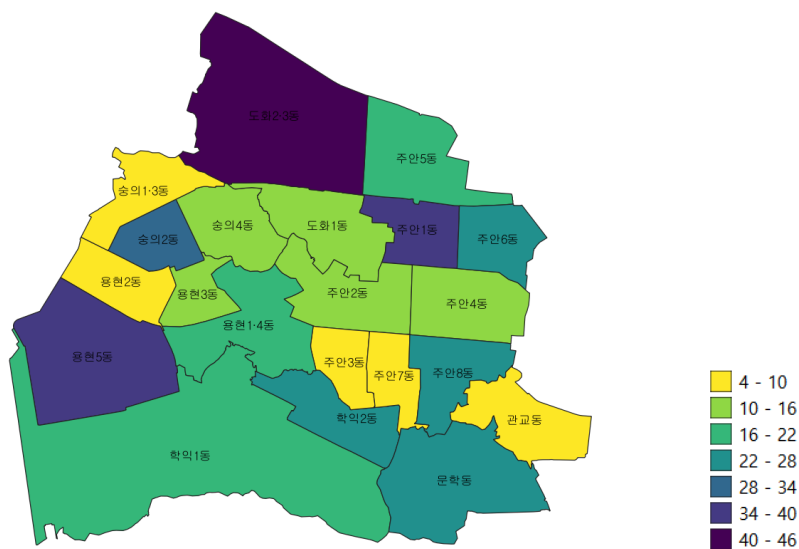
- 용현5동(25), 도화2.3동(21), 학익1동(21), 용현2동(17), 주안6동(15) 순으로 순으로 상위에 위치함

■ 행정동별 금연 구역 현황

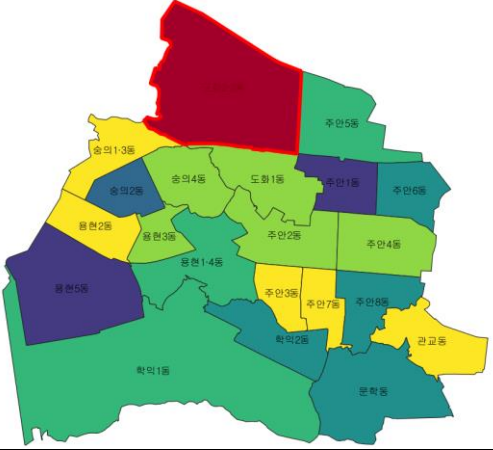
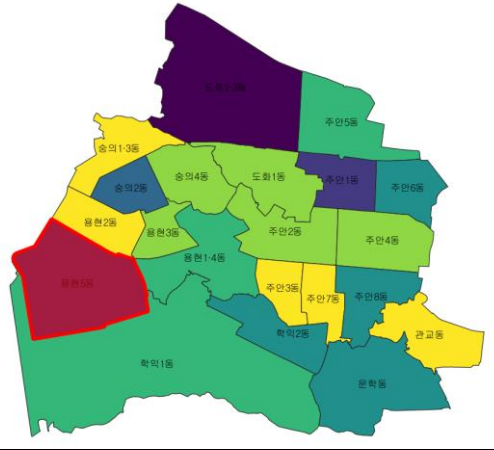
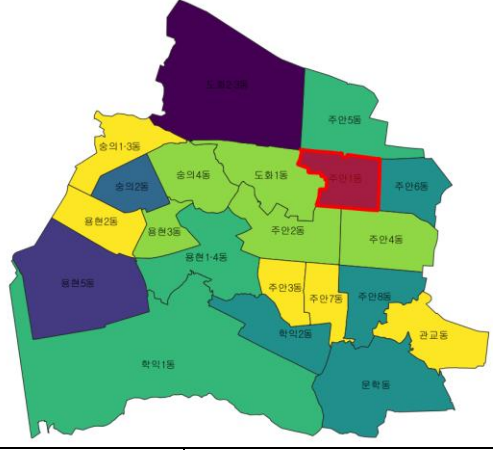
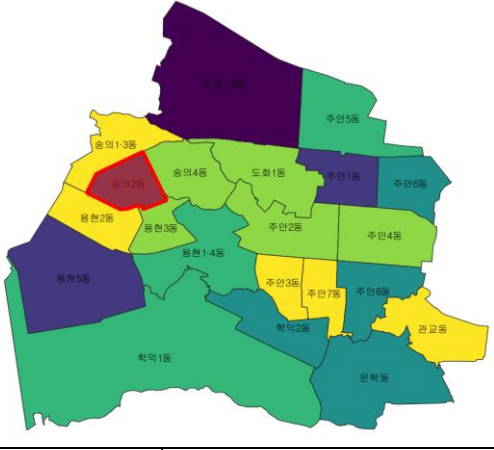


- 흡연 인원 구역인 주안동에는 비슷한 면적을 갖고 있는 도화1동보다 금연구역 지정 수가 낮음
- 도화2·3동, 학익1동이 주요 금연구역 행정동임

■ 행정동별 민원 발생 빈도



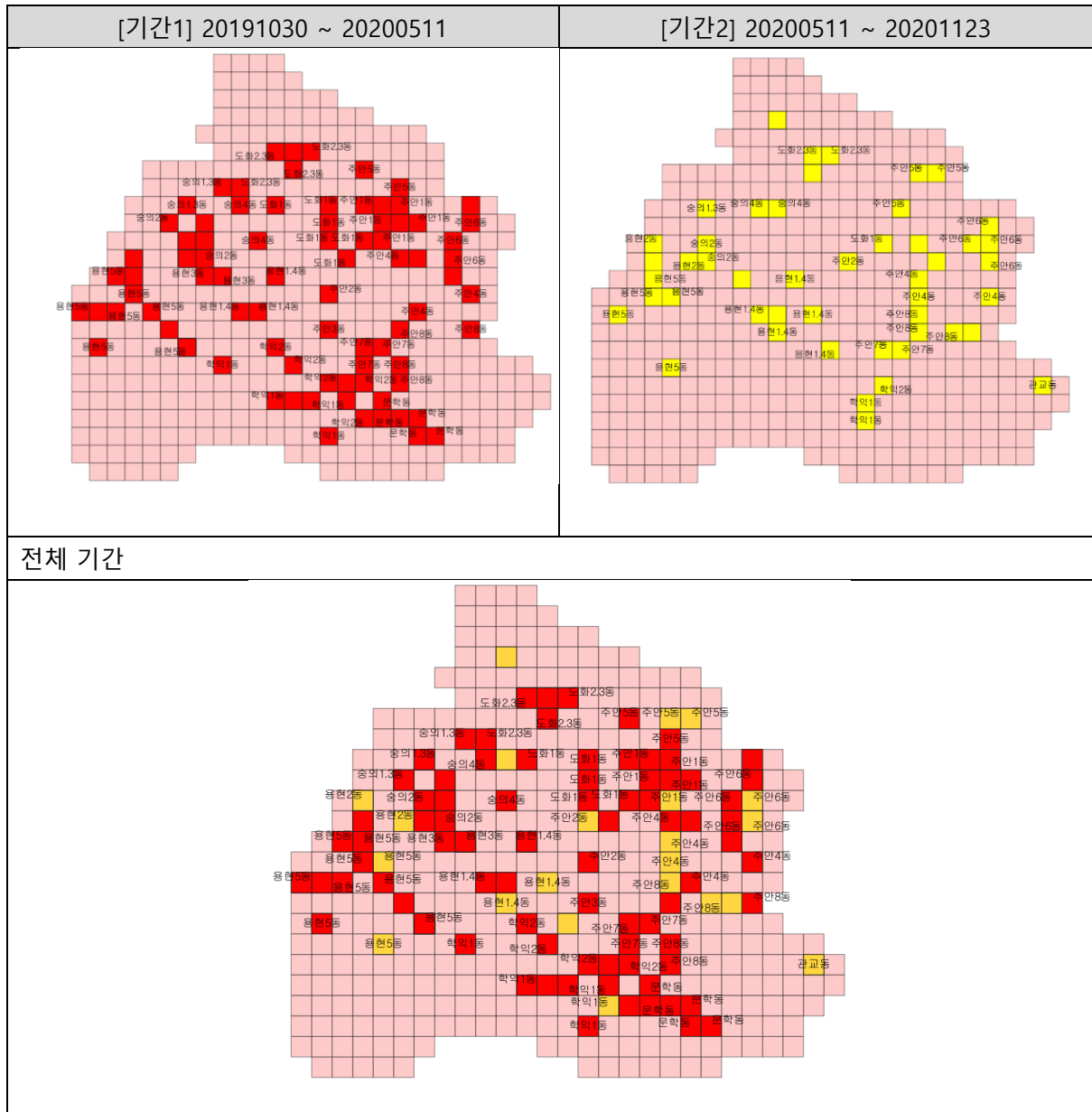
○ 상위 4개 행정동

1.		2.	
			
행정동	도화2.3동	행정동	용현5동
count	46	count	40
3		4	
			
행정동	주안1동	행정동	승의2동
count	39	count	32

- 도화2.3동, 용현5동, 주안1동, 승의2동 순으로 민원 발생 수가 높음
- 도화2.3동이 가장 많은 민원 발생지. 금연구역 수가 높음에도 불구하고 지속적으로 민원 발생함
- 주안 1동, 승의 2동은 높은 민원에 반해, 낮은 금연구역 수를 가짐

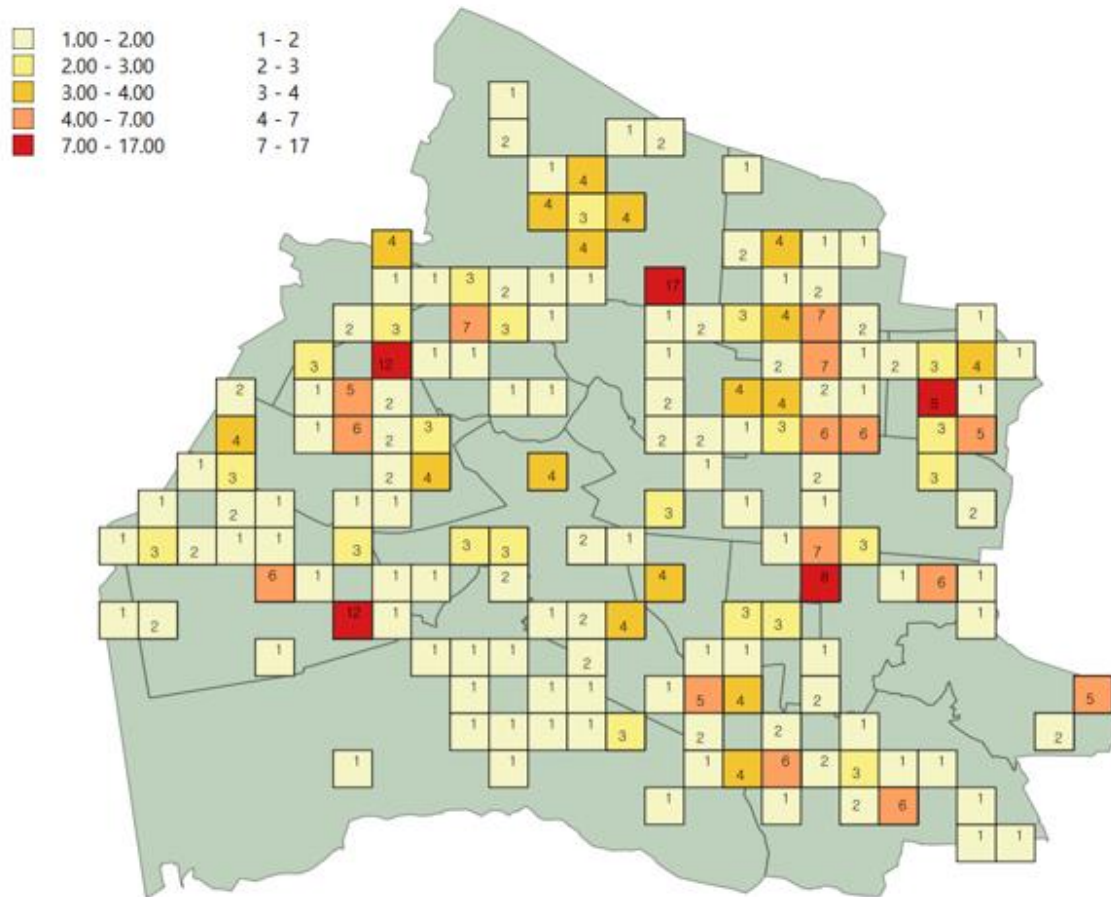
■ 민원 세부 발생지

○ 기간별 민원 세부 발생지



- 그리드 격자(50m) 기준 민원 발생지 분포 파악
- 기간1(20191030 ~ 20200511) 과 기간2 (20200511 ~ 20201123)의 민원발생지가 비슷한 위치 형태를 보임

○ 격자별 민원 세부 발생지



순위	위치	특징	건수
1	도화역 4번 출구, 인천대화초등학교 사이	유동인구 많음, 주변 버스정류장 다수 존재. 신태양 아파트 주변 빌라 단지 존재	17
2	미추홀구청, 미추홀구 청소년수련관 주변 (송의2동)	유동인구 많음. 주변 어린이집 존재,	12
2	e편한세상시티 인하대역 주변	인하대역 1번출구 주변. 유동인구 밀집 지역. 인천용학초등학교 주변	12
4	인천가정법원, 르네상스 주변 주변 (주안6동)	유동인구 많음. 주변 음식점 다수 존재	8
4	옥일아파트 ~ 청운아파트 주변	아파트 밀집 구역, 놀이터나 공터가 주변에 많음. 주변에 아림도서관 존재	8

4.3 추천지 분석 결과

■ 금연 필요 지수 - 상위 20개 지역



- 역 주변(제물포역, 도화역)과 학교나 주요 시설로 인한 일반음식점 밀집 지역이 많이 나타남

■ 금연구역 후보지 top3 도출

금연 필요 지수를 산출한 결과, 이미 선정된 금연구역을 제외하고 TOP3를 선정함

1) e편한세상시티 인하대역 메가커피 주변



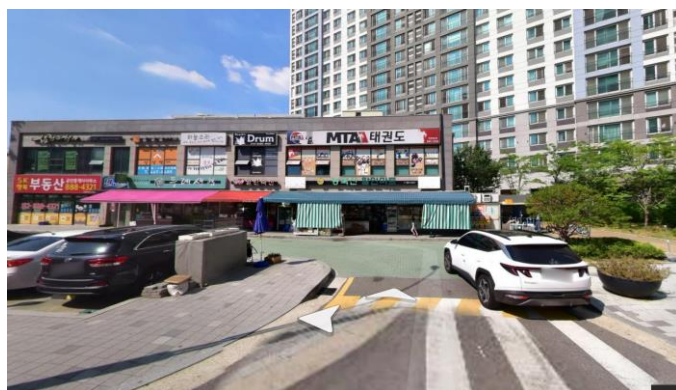
- 역 100M 이내, 민원신고발생지(4위), 주변 음식점 다량 존재(7위), PC방, 담배 소매업, 유동인구 높은 지역, 횡단보도 주변

2) 대화초등학교 주변



- 역 500m 이내에 위치하며 유동인구가 가장 많은 행정동(1위)/
지난 2년간 민원 발생(1위), 주변 초등학교 존재
아파트 단지 밀집 지역임

3) 인천SK스카이뷰, 동아아파트 사이 상가



- 주변 어린이집, 주변 일반음식점 다량 분포(2위),
담배소매업(1위) 다수 분포, 세부 민원 발생(0건)은 하지 않음

■ 금연 구역 단속 강화 후보지 도출

금연 필요 지수를 산출한 결과, 이미 선정된 금연구역을 기준으로 상위구역

1) 인천터미널 역 주변



- 담배소매업이 유독 많은 지역이면서 지하철입구 500m 이내
공원과 벤치가 다수 위치한 지역으로 담배꽂초 관련 민원 잠재적 요소가 존재

2) 제물포역 주변



- 일반음식점 지수, 유동인구가 많고, 지하철입구 100m 이내 위치한 지역.
2021년 4월 「도시철도법」에 따른 도시철도 출입구 10m이내 금연구역 확대를
했음에도 인식률이 낮아 지속적인 민원 발생이 고려됨

5. 참고자료

○ 인천광역시 건강증진과 보도자료 2021년 3월 17일

인천시, 금연구역 확대하고 과태료 5만원 부과, 4.7일 시행, 횡단보도·지하철 출입구·산책로·택시승차대 등, '흡연자는 코로나19 고위험군' 메시지 담은 대중교통 광고 추진

○ 유승의(2018) - 텍스트 네트워크 분석을 통한 부산시 버스민원 패턴분석

○ 박치성, 정지원(2013) - 사회적 인식 네트워크(socio - cognitive network)분석을 통한 정책이해 관계자 간 공유된 의미 파악 사례. 정부학연구, 19(2), 73 - 108

○ 조태민, 이지형(2015) - LDA 모델을 이용한 잠재 키워드 추출

○ LDA 토픽모델링 기법을 활용한 부산시 민원 빅데이터 분석