

# 시민이 공감하는 치안 체감안전도 예측 및 분석

SSAC 3조

# 목차 Table of Contents

- 1 분석 배경 및 목표
- 2 데이터 소개 및 전처리
- 3 모델링
- 4 분석 결과
- 5 결론



---

Part 1  
분석배경 및 목표

# 1 분석 배경

## 체감 안전도란?

- 각 경찰서관할지역당 주민일부에게 설문조사를 실시하여 치안체감안전도를 측정
- 1년에 상/하반기로 나누어 총 2번 측정하며 점수는 6월과 12월에 발표, 2020년도부터는 연1회 실시
- 현장 치안인력 보강, 치안서비스 개선 등 다양한 경찰활동 및 치안정책을 개발하는데 활용

### [체감안전도 관련 기사]

[동네방네]구로구, 안전귀가 돕는 '스마트 보안등' 설치

"선제·예방적 경찰활동 펼쳐 주민 체감 안전도 높ی겠다"

인천시, 공중화장실 '안심비상벨' 설치..."범죄 예방강화나서"

"안전한 도시, 행복한 시민, 함께하는 인천경찰"...체감안전도 높이기 전력

상주경찰서, 주민 체감안전도 향상을 위한 '원스톱 안심채움센터 실시'

> >

국민의 체감안전도를 높이기 위한  
경찰의 적극적인 **노력**

"현재 진행중"

# 1 분석 배경

## 경찰 체감안전도 설문지 내용

(0~10점, 높을수록 매우 안전)

- Q1. 생활하시는 지역의 **절도, 폭력**과 같은 범죄로부터 얼마나 안전하다고 생각하시나요?
- Q2. 생활하시는 지역의 **살인, 강도**와 같은 범죄로부터 얼마나 안전하다고 생각하시나요?
- Q3. 생활하시는 지역의 **교통사고**로부터 얼마나 안전하다고 생각하시나요?
- Q4. 생활하시는 지역의 **기초질서, 집회시위질서 등 법질서**가 현재 얼마나 잘 지켜진다고 생각하시나요?
- Q5. 위의 문항을 고려해 볼 때, **생활하시는 지역이 얼마나 안전**하다고 생각하시나요?



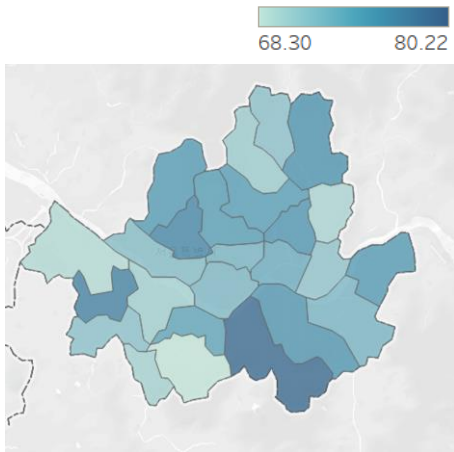
여러분들은 해당 질문에 몇 점을 주실 것 같은가요?

그렇게 생각하는 이유는 무엇이었나요?

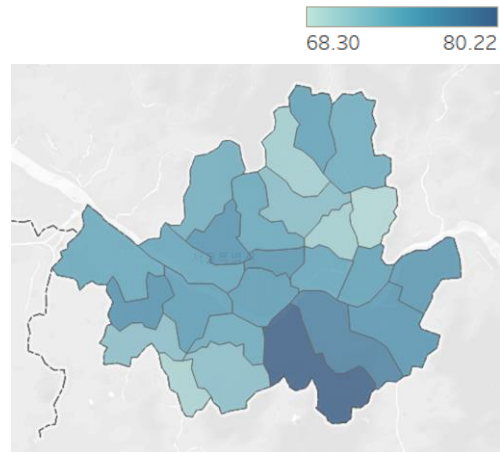
여러분의 동네 체감안전도는 몇점이 나왔을까요?

# 1 분석 배경

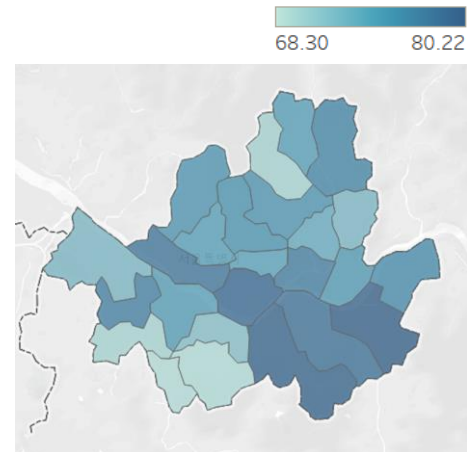
여러분의 동네의 체감 안전도는 이렇다고 합니다



[ 2017 종합체감안전도 ]



[ 2018 종합체감안전도 ]



[ 2019 종합체감안전도 ]



[ 2020 종합체감안전도 ]

2020년 안전도는 어떻게 될 것 같나요?  
우리동네 체감안전도에 공감이 되시나요?

# 1 분석 목표

뉴스홈 | 최신기사

## 주민 불편듣고 순찰 위치 변경했더니...체감안전도 상승

송고시간 | 2019-02-18 10:35

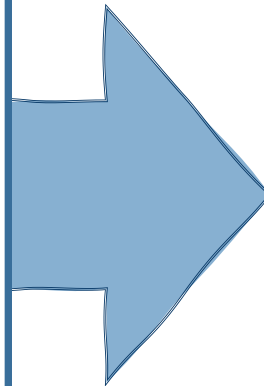
이 중 체감안전도가 많이 오른 상위 10개 경찰서를 분석한 결과 다른 경찰서에 비교해 탄력순찰과 문안 순찰 등 경찰관이 지역주민들의 민원을 듣고 반영하는 활동이 많았던 것으로 확인됐다.

### 경찰, 자치경찰 시대 앞두고 체감안전도 제고 방안 모색

글 이성진 | ⓒ 승인 2020.07.13 16:22 | 댓글 0

민회경 연구관은 「서울청 체감안전도 영향요인 분석 및 향상방안 제언」이라는 주제로 경찰 치안서비스 평가의 영향요인을 발표했다.

시민들의 체감안전도에 영향을 미치는 다양한 요인들을 분석한 결과, 관서별로 영향요인이 다른 것으로 나타났다. 예를 들어 주민들의 체감안전도에 가장 큰 영향을 미치는 요소가 A 관서는 '높은 여성 대상 범죄 발생량'인 반면, B 관서는 '높은 경제 취약인구 비율'로 영향요인이 서로 다른 것을 확인할 수 있었다.



- ✓ 머신러닝을 사용한 체감안전도 분석에 대한 연구가 조명을 받고 있음
- ✓ 현실적으로 체감안전도에 영향을 미치는 실질적인 요인을 찾는 것이 필요함

기사출처 : <http://www.lec.co.kr/news/articleView.html?idxno=720342>

# 1 분석 목표

“

체감안전도에 영향을 주는 것 요인을 파악하고 나아가 예측 모델을 확보하여  
실질적인 체감안전도를 높이기 위한 경찰활동 및 정책을 개발하는데 활용하자 ”

목표 1. 경찰데이터, 외부데이터를 활용하여 체감안전도 예측 모델을 만든다

목표 2. 체감안전도 설문 각 항목에 대해 실질적인 영향을 주는 요소를 파악한다



---

Part 2

# 데이터 소개 및 전처리

## 2 데이터 소개

데이터명	형식	추출정보	비고
5개시도_경찰서_관할경계	geojson	-	내부데이터
체감안전도_통계(2017~2019)	csv	체감안전도	
112신고	csv	신고건수	
범죄발생원표	csv	범죄발생 비율	
범죄검거원표	csv	검거발생 비율	
5개시도_최종학력통계	csv	고학력자 수	
5개시도_교통사고현황	csv	연령별, 성별 교통사고 건수	
5개시도_보안등현황	csv	보안등 개수	
5개시도_CCTV설치현황	csv	CCTV 개수	
5개시도_유흥업소_단란주점 현황	csv	유흥업소 개수	
5개시도_비상벨 현황	geojson	비상벨 개수	
5개시도_성연령별 총 인구분포	csv	연령별, 성별 인구수, 외국인 인구수	
5개시도_1인 가구수	csv	1인 가구수, 여성 1인 가구수	
체감안전도_통계(2016)	csv	체감안전도	외부데이터 (스마트치안빅데이터플랫폼)
경찰서별_지구대별 인원현황	csv	근무인원수	
어린이보호구역	csv	어린이보호구역 개수	외부데이터 (공공데이터포털)
지역별 인구현황 데이터(2020)	csv	연령별 성비, 연령별, 성별 인구수	
지역별 1인가구 통계(2020)	csv	1인 가구수, 여성 1인 가구수	

[ 공간적 수집 단위 ]

[ 시간적 수집 단위 ]

좌표  
(위도,경도)

경찰서 관할 경계와 Spatial Join

시군구

관할 경찰서로 매핑

jur_stn	year
서울중부경찰서	2017(상반기)
서울종로경찰서	2017(상반기)
서울남대문경찰서	2017(상반기)
서울서대문경찰서	2017(상반기)
서울혜화경찰서	2017(상반기)
서울용산경찰서	2017(상반기)
서울성북경찰서	2017(상반기)
서울동대문경찰서	2017(상반기)

외부데이터 활용

시간 데이터 없음

상/하반기 55로 값 분할

연도  
(yy)

상/하반기로 나눠 계산

날짜  
(yymm | yymmdd)

# 2 전처리

## 2. 인구 데이터

sido	sgg_nm	age	date	popu_num	popu_female_num	local_num	local_female_num	jur_stn	year	sx_rate_합계	sx_rate_u20	sx_rate_2030	sx_rate_4050	sx_rate_o60	for_sx_rate_합계	for_sx_rate_u20	for_sx_rate_2030		
0	서울특별시	종로구	합계	2017	157277	80607	146298	74541	0	서울중부경찰서	2017(상반기)	94.7	95.1	99.1	99.8	81.5	88.0	92.5	89.7
1	서울특별시	종로구	0~4세	2017	4137	2038	4011	1972	1	서울남대문경찰서	2017(상반기)	94.7	95.1	99.1	99.8	81.5	88.0	92.5	89.7
2	서울특별시	종로구	0세	2017	694	358	675	348	2	서울중부경찰서	2017(하반기)	94.7	95.1	99.1	99.8	81.5	88.0	92.5	89.7
3	서울특별시	종로구	1세	2017	836	407	805	393	3	서울남대문경찰서	2017(하반기)	94.7	95.1	99.1	99.8	81.5	88.0	92.5	89.7
4	서울특별시	종로구	2세	2017	878	419	850	403	4	서울종로경찰서	2017(상반기)	95.1	105.7	98.0	97.0	82.0	81.0	67.1	75.1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13819	서울특별시	중구	15세미만	2017	11035	5431	10565	5231	241	마산동부경찰서	2019(하반기)	100.3	109.2	114.4	100.6	82.7	127.9	90.8	142.3
13820	서울특별시	중구	15~64세	2017	97649	49317	88374	44349	242	진주경찰서	2019(상반기)	99.7	110.8	115.7	99.3	77.1	149.4	114.1	161.8
13821	서울특별시	중구	65세이상	2017	19212	10953	18692	10660	243	진주경찰서	2019(하반기)	99.7	110.8	115.7	99.3	77.1	149.4	114.1	161.8
13822	서울특별시	중구	평균연령	2017	42.8	43.5	43.2	43.9	244	진해경찰서	2019(상반기)	105.7	109.1	120.0	109.0	82.6	178.1	116.3	214.2
13823	서울특별시	중구	중위연령	2017	42.6	43.3	43.2	43.9	245	진해경찰서	2019(하반기)	105.7	109.1	120.0	109.0	82.6	178.1	116.3	214.2

- 유사한 특성을 가지는 연령대로 구간을 나눔
- 성별 인구조로 성비를 계산해 파생변수로 사용

서울중부경찰서	2017(상반기)	중구	127896
서울남대문경찰서	2017(상반기)	중구	127896
서울중부경찰서	2017(상반기)	중구	63948
서울남대문경찰서	2017(상반기)	중구	63948

### [ 전처리 특이사항 ]

- 하나의 행정구역을 다수의 경찰서가 담당하는 경우, 담당 값을 관찰서 수만큼 1/N 처리  
=> 중구는 중부경찰서와 남대문경찰서 두 곳이 담당하므로 인구를 1/2로 나누어 대입

	jur_stn	year	report_sx_total	vio_cnt	mur_rob_cnt	car_cnt
0	서울중부경찰서	2017(상반기)	22574	1324	0	1412
1	서울종로경찰서	2017(상반기)	19674	1102	0	867
2	서울남대문경찰서	2017(상반기)	14499	949	0	736
3	서울서대문경찰서	2017(상반기)	38769	1924	0	1788
4	서울혜화경찰서	2017(상반기)	18857	1127	0	899

- 내용확인불가 data drop (전체의 15% 분량)

### [ 전처리 특이사항 ]

- 2018년(상반기)의 경우 6월 데이터만 존재해서 편차가 크게 발생
- 2019년, 2020년 상반기 데이터와 비교해본 결과, 6월의 비중은 매년 20%로 일정
- 2018년(상반기)의 데이터는 비율을 고려하여 Count \* 5 작업으로 결측값을 대체

jur_stn	year	총범죄발생 대비검거수	절도폭력범 죄발생대비 검거수	강도살 인범죄 발생대 비검거 수	교통사고범 죄발생대비 검거수	기초질서범 죄발생대비 검거수	인구수대 비검거수	외국인인구 수대비검거 수	인구수대 비경찰수	
0	서울중부경찰서	2017(상반기)	79.206107	41.258741	75.0	96.829971	0.000000	4.056421	50.540672	0.803778
1	서울종로경찰서	2017(상반기)	73.816060	39.619048	200.0	110.891089	400.000000	2.735301	39.183897	0.815122
2	서울남대문경찰서	2017(상반기)	69.114126	54.237288	0.0	97.580645	128.571429	2.708451	33.745738	0.772503
3	서울서대문경찰서	2017(상반기)	81.954535	56.379822	0.0	106.567164	0.000000	1.088238	26.589112	0.205076
4	서울혜화경찰서	2017(상반기)	80.970670	39.812646	125.0	104.612546	0.000000	2.948937	42.244285	0.604030

- 내용확인불가 data drop (전체의 3% 분량)

- ✓ 범죄 데이터는 신고 데이터와 동일하게 작업했다
- ✓ "각 범죄 발생 대비 검거 수"를 파생변수로 생성  
=> 변수의 논리적 의미를 고려

	sido	sgg_nm	emd_nm	sx	age	inv_edu	inv_detail	num_in_2015	jur_stn
0	서울특별시	종로구	사직동	남자	합계	계	6세 이상 내국인	3669.0	서울종로경찰서
1	서울특별시	종로구	사직동	남자	합계	초등학교	6세 이상 내국인	339.0	서울종로경찰서
2	서울특별시	종로구	사직동	남자	합계	중학교	6세 이상 내국인	237.0	서울종로경찰서
3	서울특별시	종로구	사직동	남자	합계	고등학교	6세 이상 내국인	629.0	서울종로경찰서
4	서울특별시	종로구	사직동	남자	합계	대학교(2,3년제)	6세 이상 내국인	196.0	서울종로경찰서
...	...	...	...	...	...	...	...	...	...
80339	경상남도	창원시진해구	웅동2동	여자	70세 이상	고등학교	6세 이상 내국인	57.0	진해경찰서
80340	경상남도	창원시진해구	웅동2동	여자	70세 이상	대학교(2,3년제)	6세 이상 내국인	8.0	진해경찰서
80341	경상남도	창원시진해구	웅동2동	여자	70세 이상	대학교(4년제 이상)	6세 이상 내국인	NaN	진해경찰서
80342	경상남도	창원시진해구	웅동2동	여자	70세 이상	대학원(석박사 과정)	6세 이상 내국인	NaN	진해경찰서
80343	경상남도	창원시진해구	웅동2동	여자	70세 이상	받지 않았음(미취학 포함)	6세 이상 내국인	234.0	진해경찰서

	jur_stn	edu
0	마산동부경찰서	71730.0
1	마산중부경찰서	39117.0
2	서울강남경찰서	327316.0
3	서울강동경찰서	199874.0
4	서울강북경찰서	111917.0
5	서울강서경찰서	256653.0
6	서울관악경찰서	281728.0
7	서울광진경찰서	169627.0
8	서울구로경찰서	177404.0

- ✓ 모든 연령을 대상으로 지역별 고학력자(대학교 이상)을 취득한 총합계
- ✓ 최종학력 데이터는 5년마다 측정하는 값으로, 연도 구분 없이 같은 값으로 처리

	date	acc_case	ofn_sx	ofn_age	vic_sx	vic_age	jur_stn
0	2018년 1월	차대차 - 측면충돌	남	20대	남	50대	서울중부
1	2018년 1월	차대차 - 측면충돌	여	50대	남	70대	서울중부
2	2018년 1월	차대차 - 측면충돌	남	40대	남	50대	서울중부
3	2018년 1월	차대차 - 기타	기타불명	미분류	여	40대	서울중부
4	2018년 1월	차대차 - 기타	남	60대	남	50대	서울중부
...	...	...	...	...	...	...	...
140949	2019년 8월	차대사람 - 횡단중	남	40대	여	90대 이상	마산중부
140950	2020년 12월	차대사람 - 차도통행중	여	40대	여	90대 이상	창원중부
140951	2019년 4월	차대사람 - 횡단중	남	60대	남	90대 이상	진해경찰서
140952	2019년 5월	차대사람 - 차도통행중	남	20대	남	90대 이상	진해경찰서
140953	2020년 9월	차대차 - 기타	여	40대	남	90대 이상	진해경찰서

	jur_stn	year	ofn_10	ofn_20	ofn_340	ofn_5670	vic_10	vic_20	vic_340	vic_5670
0	서울중부경찰서	2017(상반기)	0.015209	0.152091	0.296578	0.528517	0.030418	0.129278	0.349810	0.471483
1	서울종로경찰서	2017(상반기)	0.012000	0.096000	0.328000	0.560000	0.012146	0.129555	0.404858	0.429150
2	서울남대문경찰서	2017(상반기)	0.004000	0.100000	0.312000	0.576000	0.012346	0.111111	0.399177	0.465021
3	서울서대문경찰서	2017(상반기)	0.038462	0.128205	0.286325	0.538462	0.037363	0.167033	0.342857	0.430769
4	서울혜화경찰서	2017(상반기)	0.007968	0.143426	0.314741	0.525896	0.012146	0.141700	0.340081	0.485830
5	서울용산경찰서	2017(상반기)	0.014184	0.122340	0.335106	0.523050	0.019573	0.156584	0.405694	0.405694
6	서울성북경찰서	2017(상반기)	0.024259	0.121294	0.336927	0.509434	0.038567	0.159780	0.363636	0.413223
7	서울동대문경찰서	2017(상반기)	0.032401	0.122239	0.318115	0.515464	0.037425	0.130240	0.344311	0.458084
8	서울마포경찰서	2017(상반기)	0.026352	0.152566	0.334258	0.481276	0.035862	0.195862	0.379310	0.377931
9	서울영등포경찰서	2017(상반기)	0.033748	0.134991	0.332149	0.495560	0.026478	0.144748	0.396293	0.421889
10	서울성동경찰서	2017(상반기)	0.030769	0.132692	0.340385	0.488462	0.034749	0.160232	0.376448	0.411197

- ✓ 변수간의 상관관계를 확인하여 작업 => 피해자\_연령별, 가해자\_연령별, 사고케이스로 column 생성
- ✓ 지역별 교통사고건수 편차를 고려하기 위해 “사고 발생 비율”을 파생변수로 생성

✓ 보안등 / CCTV / 비상벨 / 유흥업소 같은 형태의 데이터로 모두 동일하게 처리

	address	purpose	cctv_cnt	set_yr	lat	lon
0	세종특별자치시 조치원을 교리 129-1	다목적	3	NaN	36.605060	127.296286
1	세종특별자치시 조치원을 남리 144	다목적	2	NaN	36.597678	127.302452
2	세종특별자치시 조치원을 남리 425	다목적	3	NaN	36.594863	127.300492
3	세종특별자치시 조치원을 명리 82	다목적	3	NaN	36.600620	127.301301
4	세종특별자치시 조치원을 상리 18-10	다목적	4	NaN	36.602434	127.303145



geometry
POINT (127.28769 36.44347)
POINT (127.28881 36.44418)
POINT (127.28957 36.44471)
POINT (127.29007 36.44445)
POINT (127.29000 36.44419)



jur_stn	year
세종경찰서	1900
세종경찰서	1900
세종경찰서	1900
세종경찰서	1900
세종경찰서	1900

### 추출결과

	jur_stn	year	light_cnt	cctv	adult	bell
0	서울중부경찰서	2017(상반기)	210.0	1563	198.0	241.0
1	서울종로경찰서	2017(상반기)	0.0	676	148.0	38.0
2	서울남대문경찰서	2017(상반기)	10.0	725	87.0	127.0
3	서울서대문경찰서	2017(상반기)	0.0	2938	74.0	311.0
4	서울혜화경찰서	2017(상반기)	0.0	625	136.0	20.0
5	서울용산경찰서	2017(상반기)	10.0	2051	121.0	890.0
6	서울성북경찰서	2017(상반기)	0.0	1158	36.0	402.0
7	서울동대문경찰서	2017(상반기)	0.0	13	143.0	4.0
8	서울마포경찰서	2017(상반기)	94.0	1245	118.0	570.0

### [ 전처리 특이사항 ]

- 과거에 설치된 CCTV 수가 대다수를 차지하고 있는 점을 감안
- 신규 설치된 시설물 => 해당 시점에 설치되어 있는 시설물(설치현황)으로 변수 변경



# 2 전처리

## 7. 외부데이터(어린이 보호구역, Lag변수 생성)

	jur_stn	year	child
0	서울중부경찰서	2017(상반기)	39
1	서울종로경찰서	2017(상반기)	46
2	서울남대문경찰서	2017(상반기)	39
3	서울서대문경찰서	2017(상반기)	39
4	서울혜화경찰서	2017(상반기)	46

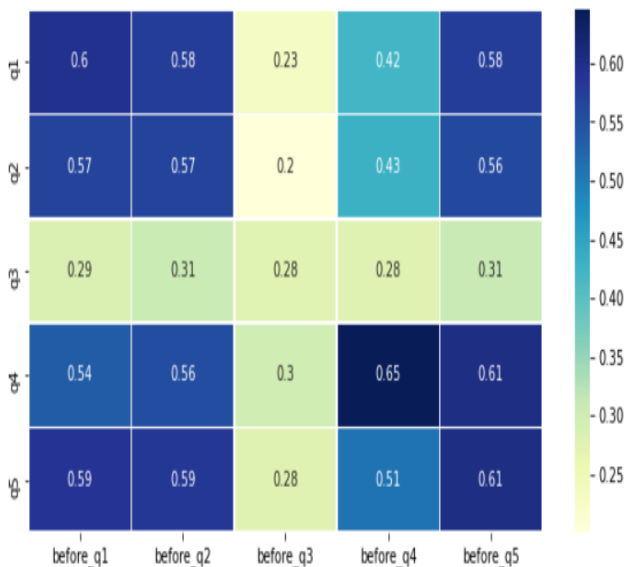
### [ 관련 기사 ]

[서행하면 사람이 보입니다] (1)한달 평균 7.6명 사망...안전하지 못한 인천 교통환경  
어린이보호구역(스쿨존)에서조차 교통사고 사망자가 잇따르고 있다.

### [ 어린이 보호구역 ]

- 분석과정중,어린이보호구역과안전도와관련있다는기사를확인
- 설치물 처리와동일하게현황으로데이터처리

	jur_stn	year	q1		jur_stn	year	before_q1
0	서울중부경찰서	2017(상반기)	75.2	82	서울중부경찰서	2018(상반기)	75.2
1	서울종로경찰서	2017(상반기)	81.6	83	서울종로경찰서	2018(상반기)	81.6
2	서울남대문경찰서	2017(상반기)	77.3	84	서울남대문경찰서	2018(상반기)	77.3
3	서울서대문경찰서	2017(상반기)	78.0	85	서울서대문경찰서	2018(상반기)	78.0
4	서울혜화경찰서	2017(상반기)	78.6	86	서울혜화경찰서	2018(상반기)	78.6
5	서울용산경찰서	2017(상반기)	77.8	87	서울용산경찰서	2018(상반기)	77.8
6	서울성북경찰서	2017(상반기)	77.9	88	서울성북경찰서	2018(상반기)	77.9
7	서울동대문경찰서	2017(상반기)	76.3	89	서울동대문경찰서	2018(상반기)	76.3
8	서울마포경찰서	2017(상반기)	76.8	90	서울마포경찰서	2018(상반기)	76.8
9	서울영등포경찰서	2017(상반기)	71.3	91	서울영등포경찰서	2018(상반기)	71.3
10	서울성동경찰서	2017(상반기)	76.0	92	서울성동경찰서	2018(상반기)	76.0



### [ Lag 변수 ]

- 체감안전도가전년도 점수에비해큰변화가없음을확인
- 각문항의전년도 점수가현재년도 모든 변수에 영향을미침
- 모델의성능을개선하기위한파생변수생성

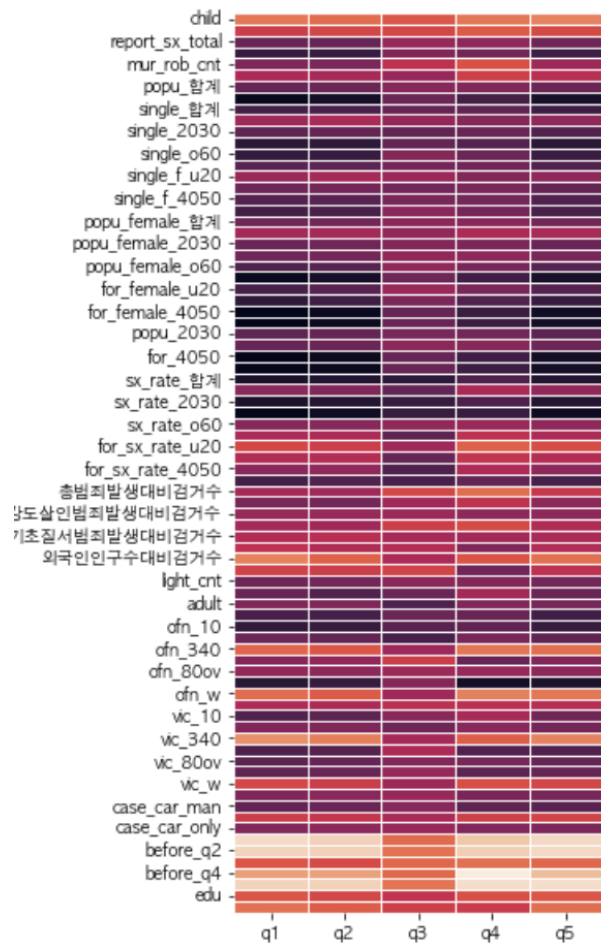
=> 전년도 안전도가 현재년도 안전도의 상관관계가 대부분 0.5 이상으로 밀접한 관련이 있다

데이터명	형식	추출정보	형태
체감안전도 관할경찰서 지역 데이터	jur_stn	관할경찰서(41개)	Category
	year	2017~2020 상/하반기 구분	Category
어린이보호구역 데이터	child	어린이보호구역	Count
최종학력 데이터	edu	지역별 고학력자 수	Count
112신고 데이터	report_sx_total, vio_cnt, mur_rob_cnt, car_cnt	신고데이터(전체, 폭력, 살인, 교통사고)	Count
범죄발생원표 데이터 범죄검거원표 데이터	총범죄발생대비검거수, 절도폭력범죄발생대비검거수, 강도살인범죄발생대비검거수, 교통사고범죄발생대비검거수	변수명과 동일	Rate
	기초질서범죄발생대비검거수, 인구수대비검거수, 외국인인구수대비검거수, 인구수대비경찰수		
1인 가구 데이터	single_합계, single_u20, single_2030, single_4050, single_o60, single_f_합계, single_f_u20, single_f_2030, single_f_4050	1인 가구수(전체, 연령별, 여성연령별)	Count
인구 데이터	popu_2030, popu_4050, popu_2030, popu_4050	인구수(전체, 연령별, 여성연령별)	Count
	popu_female_합계, popu_female_u20, popu_female_2030, popu_female_u20, popu_female_2030, popu_female_4050, popu_female_o60		Count
	for_female_합계, for_female_u20, for_female_2030, for_female_4050, for_female_o60	외국인 인구수(연령별, 여성연령별)	Count
	for_4050, for_o60		Count
	sx_rate_합계, sx_rate_u20, sx_rate_2030, sx_rate_4050, sx_rate_o60	인구 성비(전체, 연령별)	Rate
	for_sx_rate_합계, for_sx_rate_u20, for_sx_rate_2030, for_sx_rate_4050, for_sx_rate_o60	외국인 성비(전체, 연령별)	Rate
교통사고 데이터	case_car_man, case_car_car, case_car_only	차대사람, 차대차, 차량단독 사고수	Count
	ofn_10, ofn_20, ofn_340, ofn_5670, ofn_80ov, ofn_m, ofn_w, ofn_unknown	교통사고 가해자 비율(연령별, 성별)	Rate
	vic_10, vic_20, vic_340, vic_5670, vic_80ov, vic_m, vic_w, vic_unknown	교통사고 피해자 비율(연령별, 성별)	Rate
보안등, CCTV 유형업소, 비상벨 데이터	light_cnt, cctv, adult, bell	보안등, CCTV, 유형업소, 비상벨 수	Cumsum
체감안전도 데이터(2016-2018)	before_q1, before_q2, before_q3, before_q4, before_q5	전년도 체감안전도 점수(현재년도에 추가)	-
체감안전도 데이터(2017~2019)	q1, q2, q3, q4, q5	체감안전도(2017~2019)	18 -

# Part 3

## 모델링

## Step1. Question별 Corr 확인



AIC 최소일때  
Feature 조합 체크

Feature importance 체크

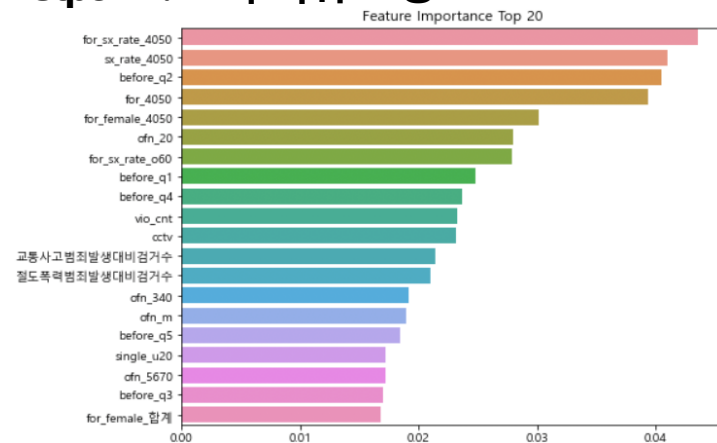
## Step2-1. 선형 회귀 모형

```
Step: AIC=361.12
q1 ~ sx_rate_4050 + ofn_w + sx_rate_2030 + 인구수대비경찰수 +
for_sx_rate_합계 + 인구수대비검거수 + single_u20 + mur_rob_cnt +
bell + ofn_340 + case_car_only + for_female_u20 + single_f_u20 +
for_female_o60 + 외국인인구수대비검거수 + sx_rate_u20 + vio_cnt +
vic_80ov + before_q2

Step: AIC=360.9
q1 ~ sx_rate_4050 + ofn_w + sx_rate_2030 + 인구수대비경찰수 +
for_sx_rate_합계 + 인구수대비검거수 + single_u20 + mur_rob_cnt +
bell + ofn_340 + case_car_only + for_female_u20 + single_f_u20 +
for_female_o60 + 외국인인구수대비검거수 + sx_rate_u20 + vio_cnt +
vic_80ov + before_q2 + vic_unknown

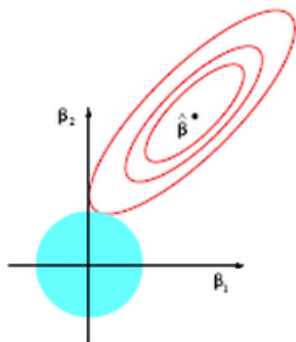
Step: AIC=360.59
q1 ~ sx_rate_4050 + ofn_w + sx_rate_2030 + 인구수대비경찰수 +
for_sx_rate_합계 + 인구수대비검거수 + single_u20 + mur_rob_cnt +
bell + ofn_340 + case_car_only + for_female_u20 + single_f_u20 +
for_female_o60 + 외국인인구수대비검거수 + sx_rate_u20 + vio_cnt +
vic_80ov + before_q2 + vic_unknown + ofn_80ov
```

## Step2-2. 트리 회귀 모형

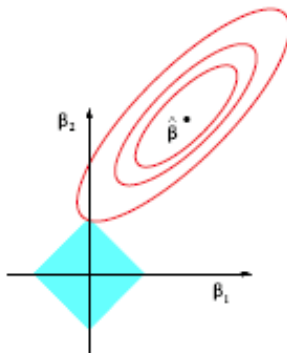


Step1 <-> Step2 과정을 기준으로 이후, 변수 수기로 추가/제거 과정 반복

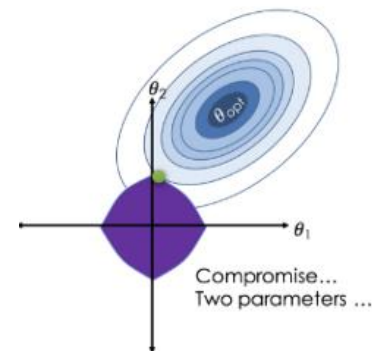
Ridge



Lasso



Elastic



Random Forest

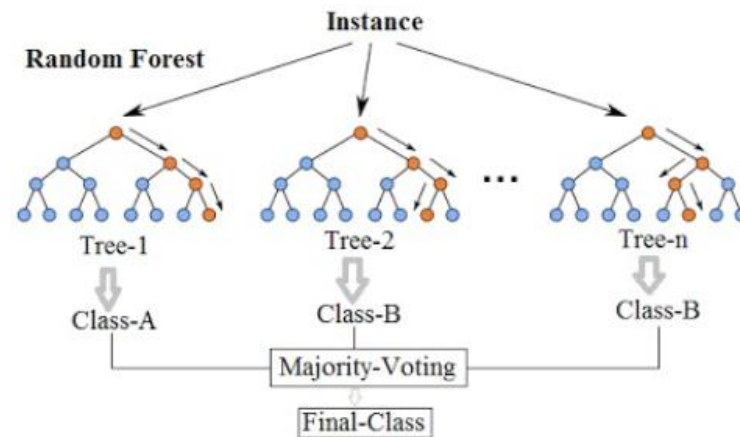
LightGBM



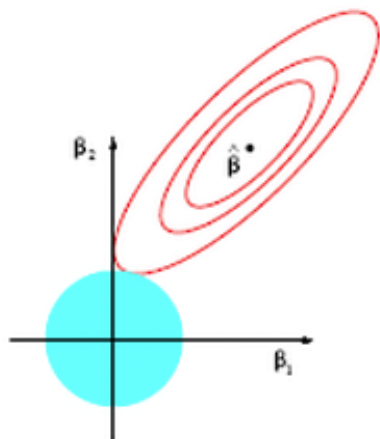
XGBoost



CatBoost



## Ridge

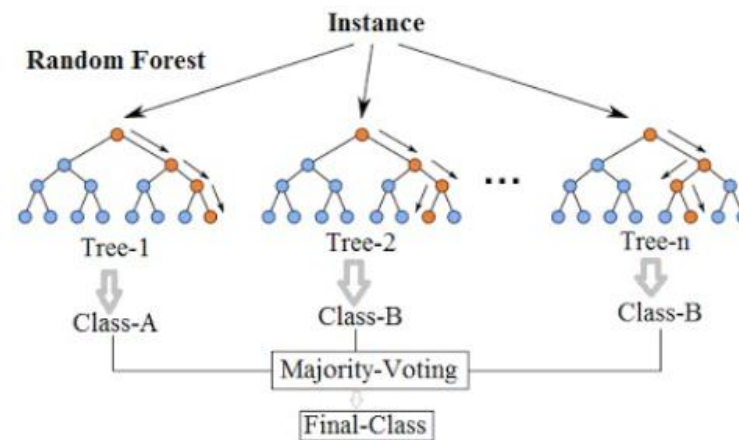


$$\min_{\beta_0, \beta} \left\{ \frac{1}{N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 \right\} + \sum_{j=1}^p (\beta_j)^2 \leq t.$$

cost =  $\sum e_i^2 + \lambda \sum w_i^2$  를 최소화 하는 방식

✓ 다중공선성 완화

## Random Forest



✓ 여러 트리의 의사결정을 반영하여 과적합 방지

**Q 1**  
**Ridge**  
**model**

**Q 2**  
**Random**  
**Forest**

**Q 3**  
**Ridge**  
**model**

**Q 4**  
**Ridge**  
**model**

**Q 5**  
**Ridge**  
**model**

구분	영향 변수
<b>Q 1</b> (절도, 폭력)	'vio_cnt', 'mur_rob_cnt', '인구수대비경찰수', 'bell', 'cctv', 'child', 'popu_o60', 'for_2030', 'single_2030', 'sx_rate_합계', 'for_o60', 'for_sx_rate_4050', 'single_4050', 'single_o60', 'single_f_u20', 'popu_2050', 'sx_rate_4050', 'single_u20', '인구대비고학력자'
<b>Q 2</b> (강도, 살인)	'sx_rate_2030', 'cctv', 'for_female_o60', 'for_female_4050', 'for_o60', 'for_4050', 'before_q4', 'bell', 'adult', '인구수대비경찰수', 'before_q2', '외국인인구수대비검거수'
<b>Q 3</b> (교통사고)	'인구수대비경찰수', '총범죄발생대비검거수', 'ofn_unknown', 'mur_rob_cnt', '인구수대비검거수', 'ofn_m', 'vic_10', 'case_car_car', 'case_car_man', 'sx_rate_4050', 'ofn_10', '교통사고범죄발생대비검거수', 'ofn_20', 'for_sx_rate_u20'
<b>Q 4</b> (법질서)	'before_q4', 'ofn_m', 'sx_rate_합계', 'before_q2', 'car_cnt', '인구수대비검거수', 'sx_rate_u20', 'popu_female_2030', 'sx_rate_o60', 'bell', 'for_female_합계', 'single_f_합계', 'before_q5', 'ofn_20', 'vio_cnt', 'case_car_car'
<b>Q 5</b> (전반적)	'for_o60', 'ofn_m', 'sx_rate_4050', 'sx_rate_2030', 'single_2030', 'sx_rate_합계', 'car_cnt', 'popu_female_2030', 'single_4050', 'before_q5', 'before_q2', 'single_u20', 'for_female_o60', 'sx_rate_o60', 'mur_rob_cnt', 'child', '인구대비고학력자', 'single_o60', 'before_q4', 'case_car_car', 'single_f_합계', 'bell', '인구수대비경찰수', 'single_f_u20', 'for_female_4050', '외국인인구수대비검거수', 'for_sx_rate_4050', '인구수대비검거수', 'adult', 'ofn_20', 'sx_rate_u20', 'for_female_합계', 'cctv', 'vio_cnt'

2019년 예측값과 실제값의 **평균절대오차점수(MAE)**를 비교한 결과 모든 모델이 **1.8점대**를 기록

구분	Q1	Q2	Q3	Q4	Q5
채택모델	Ridge	RF	Ridge	Ridge	Ridge
MAE	1.81	1.97	1.85	1.88	1.73

해당 체감안전도 예측대회 TOP 5에 들어가는 점수대로 좋은 모델을 찾았다고 판단



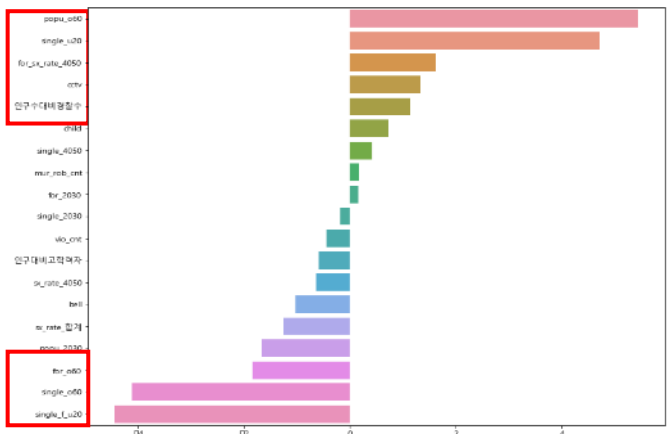
---

## Part 4

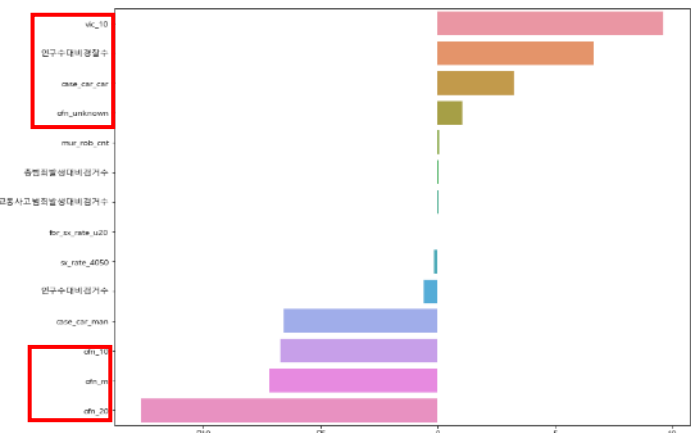
# 분석 결과

## 분석 결과

### 1. 모델 결과 시각화



[도로폭력 문항의 Coefficient]

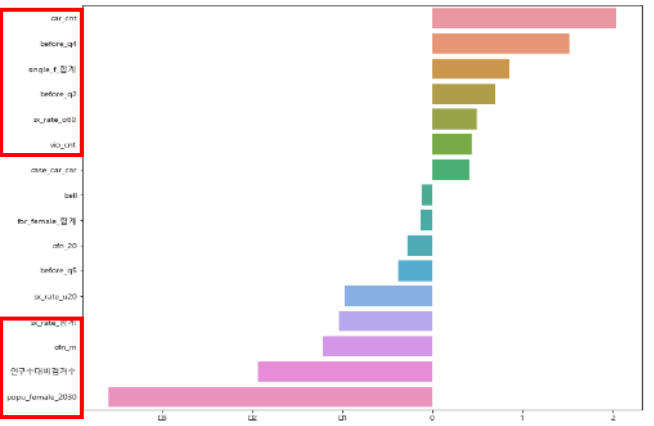


[교통사고의 Coefficient]

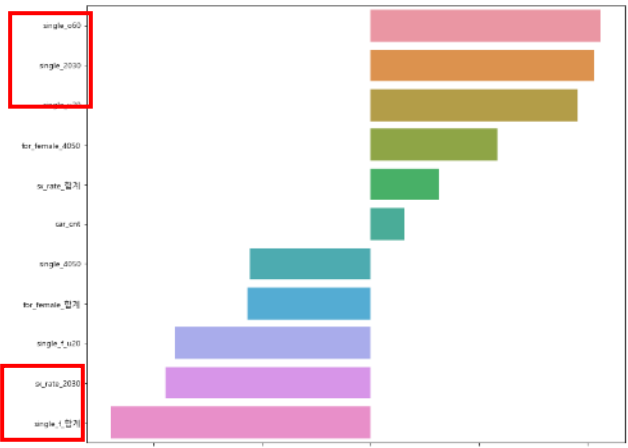
문항별 변수의 중요도 시각화

선형 회귀 모델 : Coef

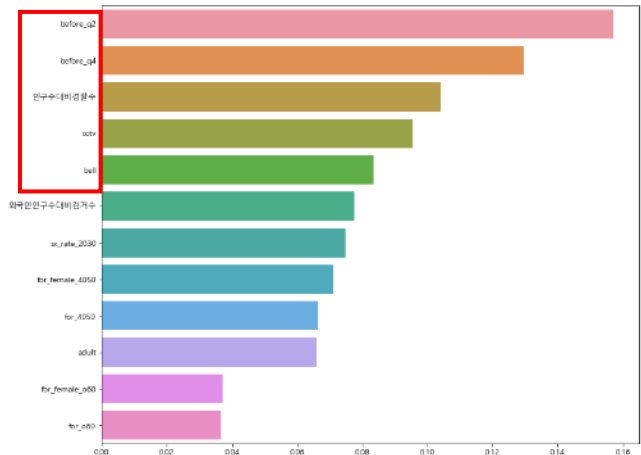
트리 회귀 모델 : Feature importances



[법질서의 Coefficient]



[전반적의 Coefficient]



[도로폭력 문항의 Feature Importances]

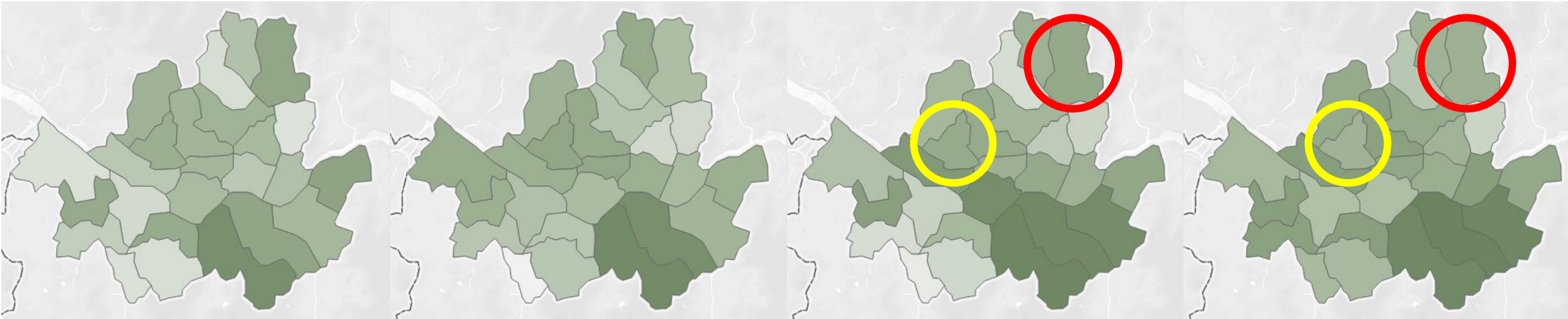
- 전반적으로 1인가구, 여성인구 비율이 많은 지역이 안전도에 영향이 많음

[ 2017년 ]

[ 2018년 ]

[ 2019년 ]

[ 2020년 ]



- 2019년 -> 2020년에 서대문경찰서, 노원경찰서 절도, 폭력 안전도가 감소되어 예측됨
- 데이터로 확인해 본 결과, 이는 20년에 **여성 1인가구**수가 19년보다 급격히 늘어남에 따른 감소될 것으로 예측됨  
=> 서대문경찰서 12점 감소 / 노원경찰서 1.1점 감소
- 전반적으로 **여성 1인가구, 20세 이하 인구**가 해당 체감안전도에 부정적인 영향을 끼치는 것을 알 수 있음
- 사회적 취약 계층이 상대적으로 절도, 폭력에 대한 불안함을 많이 나타냄

## 4

## 분석 결과

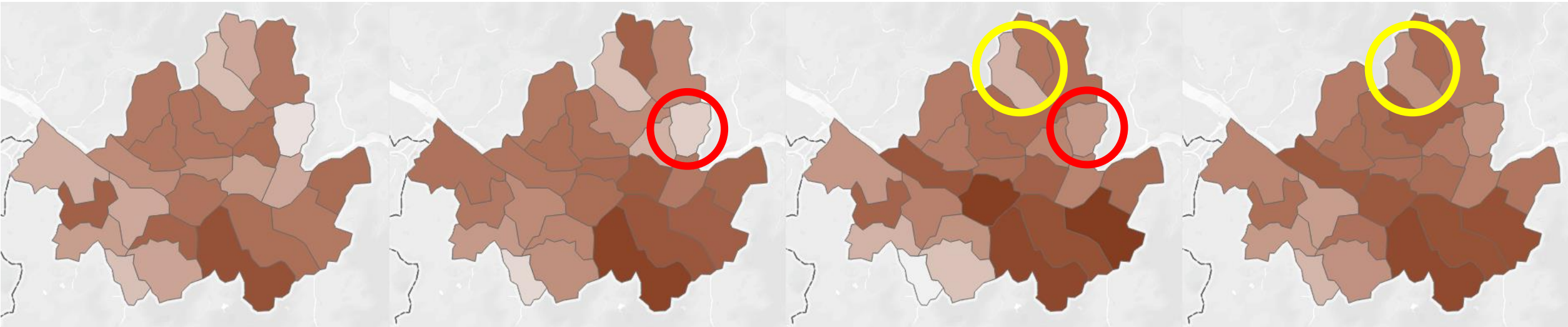
2. 예측 결과 분석 (Q2. 살인, 강도 안전도)

[ 2017년 ]

[ 2018년 ]

[ 2019년 ]

[ 2020년 ]



- 2018년 -> 2019년 **중랑경찰서** 살인, 강도 안전도는 증가
- 2019년 -> 2020년 **강북 경찰** 살인, 강도 안전도는 증가되어 예측됨
- 데이터로 확인해본 결과, **중랑경찰서는 인구수대비경찰수가 2% 증가함에 따라 해당 안전도가 증가함을 알 수 있음**
- 또한, **강북경찰서는 노인 인구가 8.5% 증가함에 따라 해당 안전도가 증가 될 것이라 예측됨**  
=> 중랑경찰서 4점 증가 / 강북경찰서 3점 증가
- 노인 인구수의 증가와 인구수대비경찰수의 증가는 해당 체감안전도에 긍정적인 영향이 있음

## 4

## 분석 결과

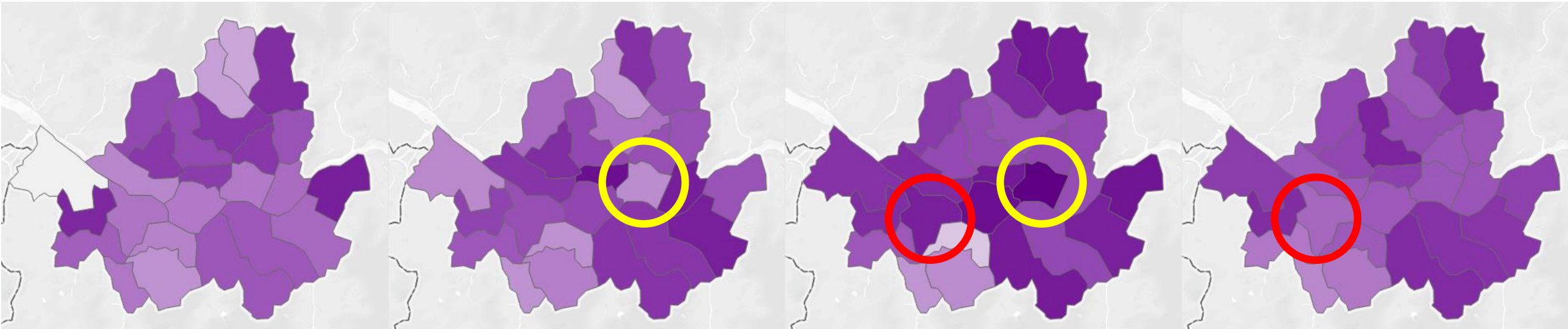
2. 예측 결과 분석 (Q3. 교통사고 안전도)

[ 2017년 ]

[ 2018년 ]

[ 2019년 ]

[ 2020년 ]



- 2018년 -> 2019년 성동경찰서는 교통사고 안전도 증가
- 2019년 -> 2020년 영등포경찰서는 교통사고 안전도가 감소되는 것으로 예측됨
- 데이터로 확인해본 결과, 교통사고 젊은연령층 가해자(10대~20대)가 늘어날수록 급격히 감소하는 것을 알 수 있음
- 또한, 인구수 대비 경찰수 비율이 증가함에 따라 증가되는 것으로 예측 됨  
=> 성동경찰서 5점 증가 / 영등포경찰서 3.8점 감소
- 전반적으로 젊은 연령층의 교통사고 가해자의 수가 많을수록  
교통사고 안전도 점수에 대한 부정적인 영향이 크고, 인구수 대비 경찰 수 비율은 긍정적인 영향이 큼

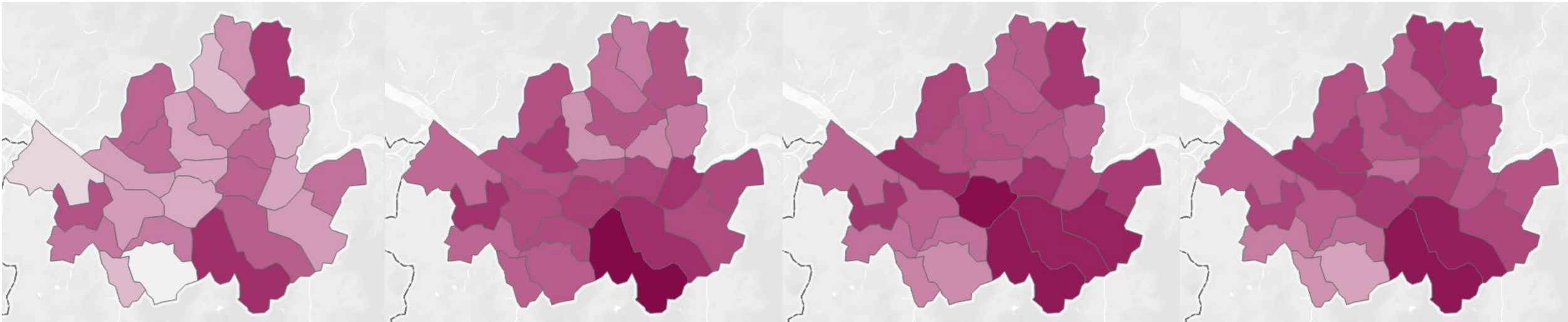


[ 2017년 ]

[ 2018년 ]

[ 2019년 ]

[ 2020년 ]



- 법질서 안전도의 경우, 기초질서와 시위질서와 관련이 있는 변수가 뚜렷하게 나타나지 않음
- 단, 특별한 변수의 영향보다 **시간 흐름에 따라 상승하는 추세**가 분명함을 확인할 수 있음
- 추가적으로 법질서 체감안전도를 평가할 수 있는 실질적인 데이터와 해당 안전도에 대한 보다 직관적인 명시가 필요해 보임

## 4

## 분석 결과

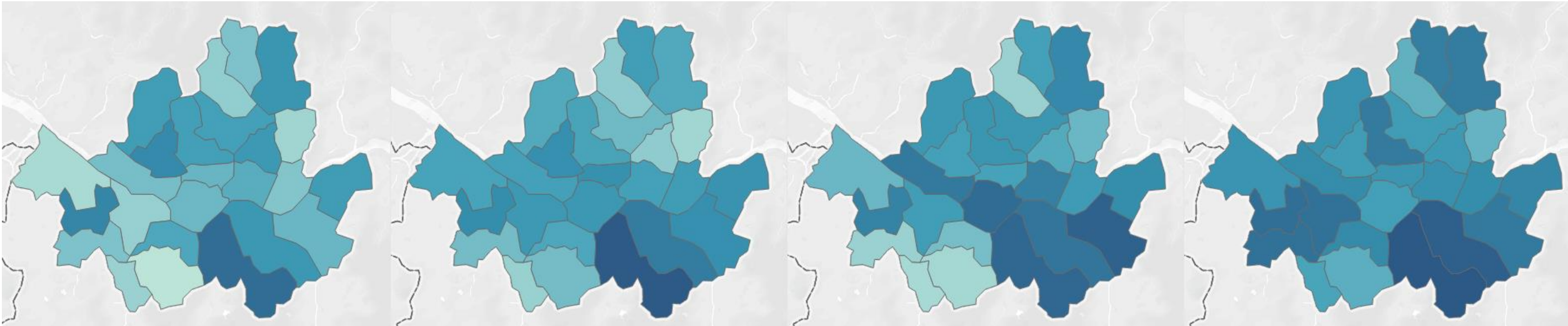
## 2. 예측 결과 분석 (Q5. 전반적 안전도)

[ 2017년 ]

[ 2018년 ]

[ 2019년 ]

[ 2020년 ]



- 전반적 안전도(Q5)는 Q1, Q2, Q4와 큰 상관 관계가 있음을 확인  
따라서, 앞에 나온 Q1, Q2, Q4의 영향력과 유사한 경향을 보임  
(\* Q1. 절도, 폭력 안전도 / Q2. 살인, 강도 안전도 / Q4. 법질서 안전도)

---

Part 5

결론

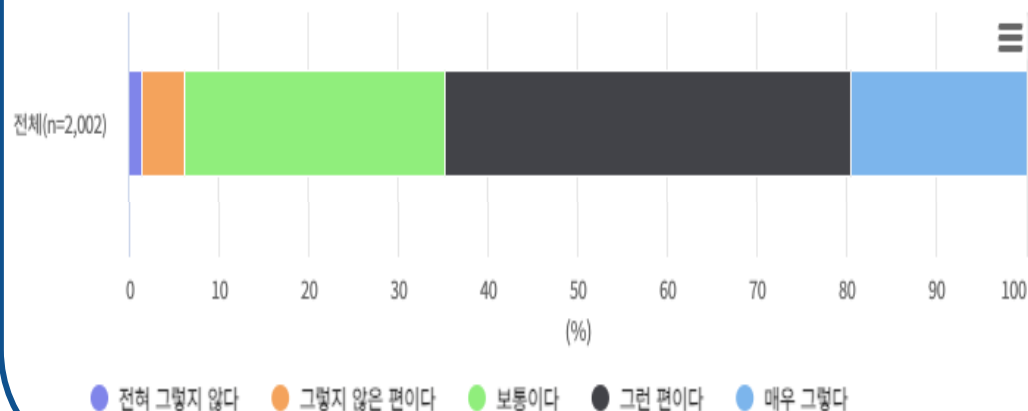


## 기대효과

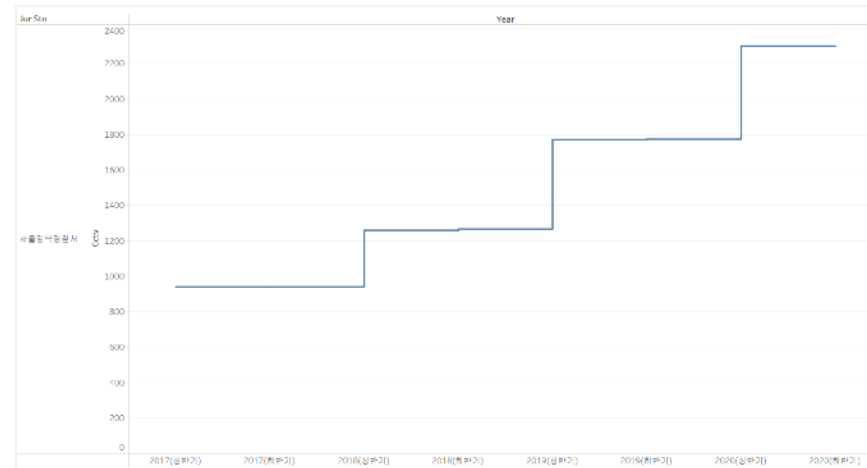
- ✓ 별도의 설문조사를 통하지 않고 데이터를 이용하여 치안 체감 안전도를 구할 수 있으며 이에 따른 시간, 비용의 절감될 것으로 예상
- ✓ 치안 체감 안전도에 영향 끼치는 요인을 파악 할 수 있으므로 치안 서비스를 보다 실질적으로 개선 할 수 있을 것으로 예상

## CCTV 설치구역, 5대 강력범죄 26.6% 감소효과

-국민안전처·경찰청, CCTV 범죄 감소효과 분석결과 공개-



- CCTV가 5대 범죄 관련해서 감소효과를 크게 나타낸다는 분석결과가 많이 나타남  
=> CCTV설치는 체감 안전도를 느끼는데 중요한 요인 중 하나라 생각
- 모든 문항에 CCTV가 어느정도 영향이 있다는 결론을 얻기는 했음
- 하지만, 우리가 가지고 있는 데이터 특성상 CCTV데이터의 경향을 살리는 것이 어려웠음



=> CCTV의 설치가 많아지면, 체감안전도도 상승한다고 생각해 볼 수 있음

=> CCTV데이터의 활용도가 높아진다면 체감안전도 모델을 개선 할 수 있음

## ■ 논문

형사정책연구 Korean Criminological Review  
제21권 제3호(통권 제83호, 2010 · 가을호)

### 범죄두려움에 관한 다수준적 접근 : 주요 모델들의 검증

박 정 선\* · 이 성 식\*\*

#### 국 | 문 | 요 | 약

이 연구는 2009년도 한국범죄피해조사 자료를 통해 우리나라 사회구성원들의 범죄두려움 및 그 원인을 설명하기 위해 주요 모델을 검증하고 있다. 이 연구는 개인수준의 변인을 고려하는 취약성/피해모델뿐 아니라 지역수준의 무질서모델, 사회통합모델, 그리고 지역사회경찰활동모델을 검증하는 다수준적 접근에 관심을 갖는다. 본 연구결과 개인수준변인들의 영향력이 크게 유의미하여 취약성/피해모델이 어느 정도 가장 크게 지지되었고, 아울러 무질서모델과 지역사회경찰활동모델도 지지된 것으로 나타났다. 하지만 사회통합모델에서는 지역구조적 특성으로 지역경제수준이외에는 유의미한 영향력을 갖는 변인은 없었으며, 그나마 지역경제수준도 사회통합모델과 기존 외국연구와는 다르게 지역경제수준이 높은 지역의 구성원들이 범죄두려움을 더 겪는 것으로 조사되어 그 지지도는 가장 낮았다. 이 연구를 통해 범죄두려움에 관한 기존의 주요 모델들이 여전히 유효하다는 것을 확인하였고, 개인수준의 변인들 이외에 지역수준의 변인들도 범죄두려움에 중요한 영향을 준다는 것을 확인함으로써 다수준적 접근의 필요성을 재인식할 수 있었다. 하지만 기존 모델들에서의 주장과는 상이한 결과도 도출되는 바 앞으로 이에 대한 세밀한 분석과 최신 모델의 구성도 필요하다고 할 것이다.

❖ 주제어 : 범죄두려움, 다수준적 접근, 취약성/피해모델

- 분석 과정 중, 성비 구분을 하기전후를 기준으로 모델의 성능이 굉장히 개선 됨을 확인함
- 해당 요인을 찾아보기 위해 논문을 찾아보니  
여성이 느끼는 체감안전도가 남성에 비해 낮다는 점을 발견
- 추가적으로, 여성이 가장 위험하다고 생각하는 '성폭력'은 체감안전도 설문에 포함이 되지 않음

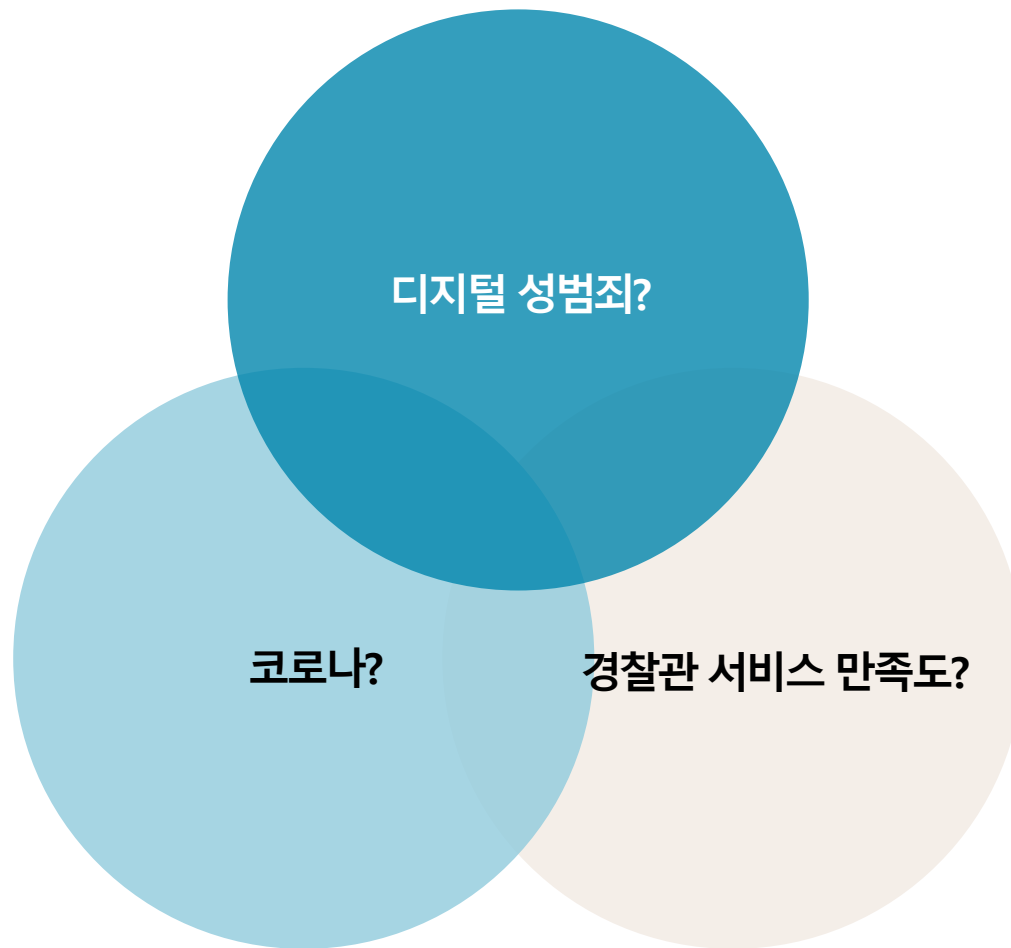
이런 **성폭력** 관련된 문제가 집중도 있게 해결된다면,

남성에 비해 안전도 평가가 낮은 여성들의 체감안전도가 보다 실질적으로 개선될 수 있을 것임

논문 출처: 다수준적 접근의 범죄학적 활용에 대한 연구 - 박정선

# 5 마치며

- 시시각각 변화하는 현 상황을 반영한다면 더욱 정확한 예측모델과 요인을 파악할 수 있을 것



새로운 요인은...?



THANK YOU