

- 데이터 분석 최종결과보고서 -

신용평가모형을 활용한 교통사고 위험 5등급 시스템 구축
: 2022년 대전 · 세종 · 충남 112 신고 데이터를 중심으로

팀 명 : 스트리트 교통파이터

작성자 : 박희진, 양문일, 최성원

목차

1 수행 개요

- ① 수행 개요
- ② 아이디어 소개
- ③ 사업 필요성
- ④ 성과요약

2 세부 수행 내용

- ① 수행 프로세스
- ② 교통사고 위험 평점화 모형 개발
 - 1) 종속변수 정의
 - 2) 독립변수 수집
 - 3) 모형 개발
 - 4) 5등급 시스템 구축
- ③ 대시보드
 - 1) 구성요소
 - 2) 시연영상

3 수행 성과

- ① 활용 방안
- ② 기대효과
- ③ 건의 사항



1

수행 개요

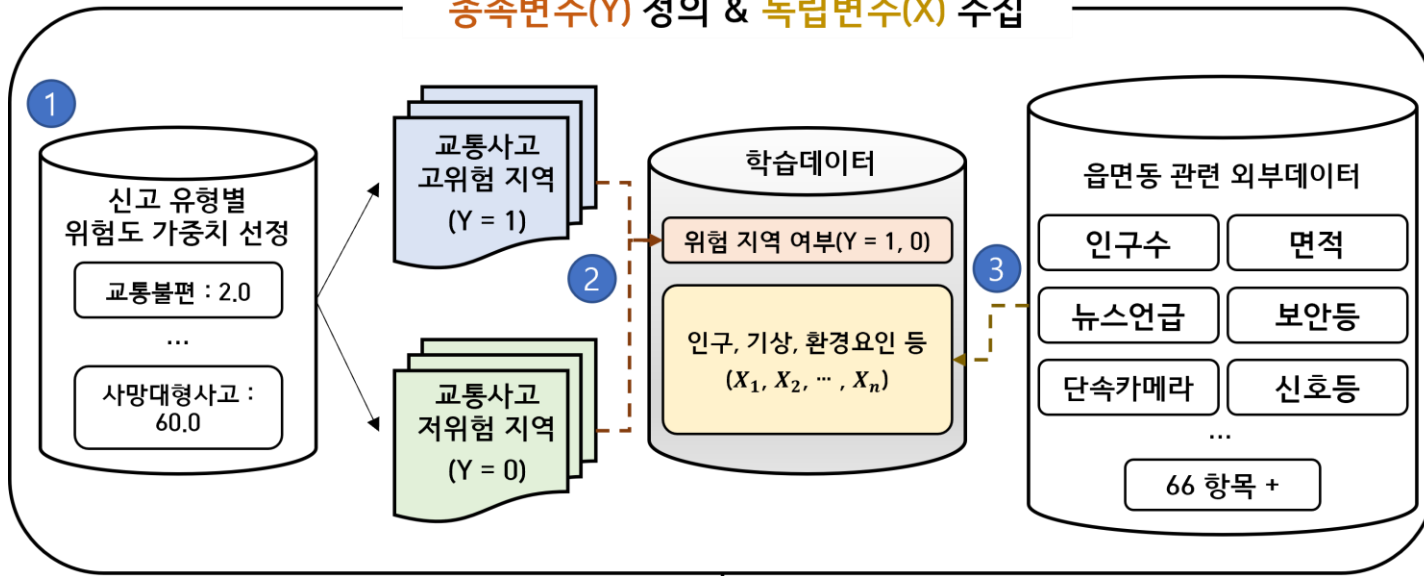
- ① 수행 개요
- ② 아이디어 소개
- ③ 사업 필요성
- ④ 성과요약

1 수행 개요

① 수행 개요

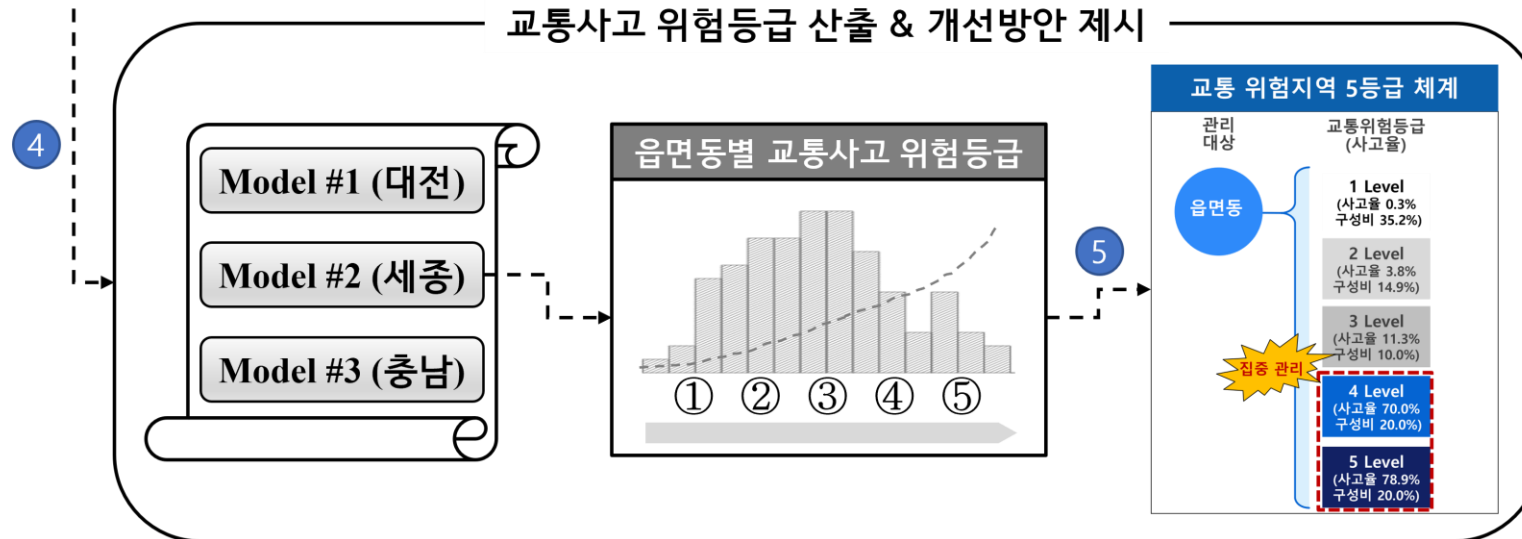
“ 대전·세종·충남 교통사고 위험 평점화 모형 제안을 위한 5단계 분석과정 요약 ”

종속변수(Y) 정의 & 독립변수(X) 수집



- 1 교통 관련 신고 유형마다 가중치 부여 (심각도가 높을수록 가중치 ↑)
- 2 읍면동별 가중치 합계에 따른 교통사고 고위험, 저위험 지역 구분
- 3 외부데이터 수집 & 매핑
- 4 교통사고 위험 평점모형 & 위험등급 산출
- 5 교통 위험지역 개선 정책 제안

교통사고 위험등급 산출 & 개선방안 제시



1 수행 개요

② 아이디어 소개

“ 교통사고 위험지역 우선 선별를 통해
교통사고 예방 및 사회적 비용 감소 ”

□ 아이디어 소개

구분	수행과제
아이디어명	신용평가모형을 활용한 교통사고 위험 5등급 시스템 구축: 2022년 대전·세종·충남 112 신고 데이터를 중심으로
목표 산출물	<ol style="list-style-type: none">대전·세종·충남 교통사고 위험평점모형 & 위험등급 → ex) 1등급 - 사고위험률 0.2%, ..., 5등급 - 사고위험률 88.7%현업이 활용가능한 대시보드 구축교통 위험지역 개선을 위한 정책제안 → ex) 3분기 교통량 증가 대비 교통경찰 배치 확대 등

□ 분석목적

- ① 축적된 112 신고 데이터를 활용하여 읍면동별 교통사고 위험지수·등급 산출 & 활용방안 제시
- ② 파편화된 공공데이터 수집, 112 신고 데이터와 융합하여 유의미한 분석결과 도출
- ③ 분석결과 대시보드 구축, 수집 필요 데이터 추가제언을 통해 고도화 방안 제시

1 수행 개요

③ 사업 필요성

“ 사회적 거리두기 전면 해제에 따른
교통사고 증가 → 사회적 비용 증가 ”

□ 가상 시나리오 설정

- ① 가정상황 : 지방자치단체에서 공모한 다음의 사업에 선정됨
- ② 사업명 : 『교통사고 예방 및 관련 정책 제안을 위한 데이터 분석 사업』

□ 분석 배경 (=사업 배경)

- ① 사회적 거리두기 전면 해제에 따른 교통사고 증가¹⁾
- ② 교통사고로 인한 사회적 비용 증가²⁾
- ③ 도로 교통사고 비용, GDP의 2.3% 비중_(2019년 기준)으로 문제 심각³⁾

□ 분석 목표

- ① 대전 · 세종 · 충남 읍면동, 분기별 교통사고 위험지역 선별 & 위험등급 산정 (5등급 체계 구축)
- ② 현업이 해석가능한 분석결과 제시
- ③ 위험등급별 교통사고 예방을 위한 정책 제안

1) 사회적 거리두기 해제 뒤 보행자 교통사고 27% 증가, <http://www.tdaily.co.kr/news/view.php?idx=39972>

2) 코로나19 시국에도 교통사고로 인한 사회적 비용은 오히려 늘어, https://www.ytn.co.kr/_ln/0103_202205121648435457

3) 'GDP 3%' 40조 길바닥서 날린다...한국인 삶 옥죄는 '이것' [뉴스원샷], <https://www.joongang.co.kr/article/24052724#home>

1 수행 개요

④ 성과요약(1/2)

“

본 사업의 산출물 종류는 5가지

”

□ 산출물 소개

1

교통사고
위험지수

- ① 교통사고 위험지수 산출 공식
- ② 대전,세종,충남 분기별, 읍면동
교통사고 위험지수 통합 리스트

2

교통사고
위험 평점화
모형

- ③ 읍면동 교통사고 위험 평점화 모형
(대전, 충남, 세종)
- ④ 5등급 체계

3

대시보드

- ⑤ 교통사고 위험지수 대시보드
(대전, 충남, 세종)

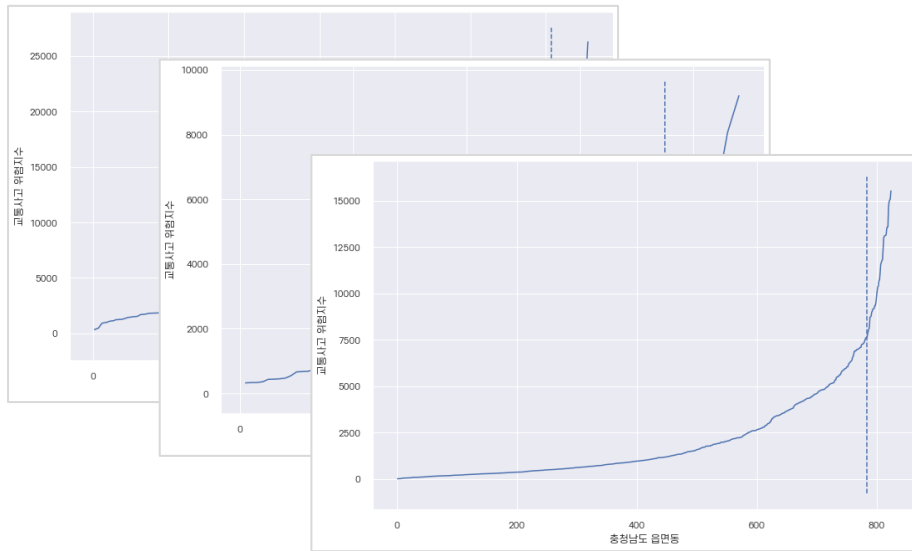
1 수행 개요

④ 성과요약(2/2)

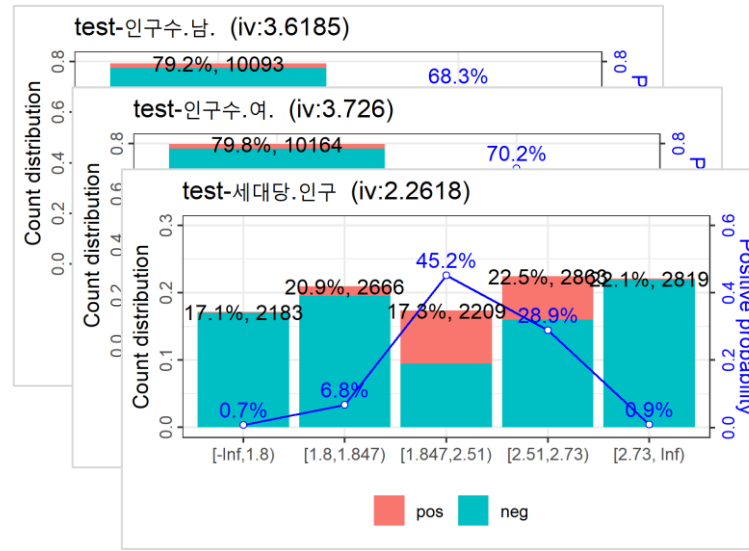
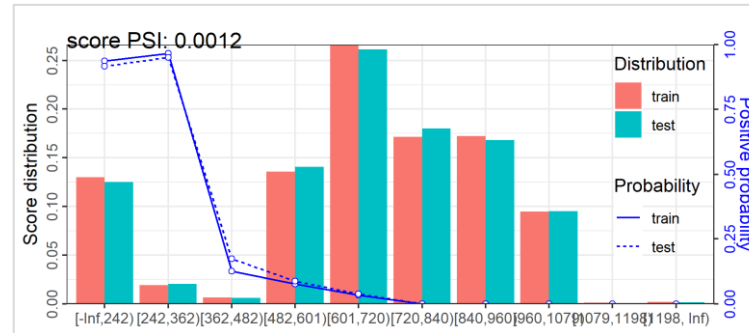
□ 산출물 요약

1) 교통사고 위험지수 & 고위험 읍면동 구분

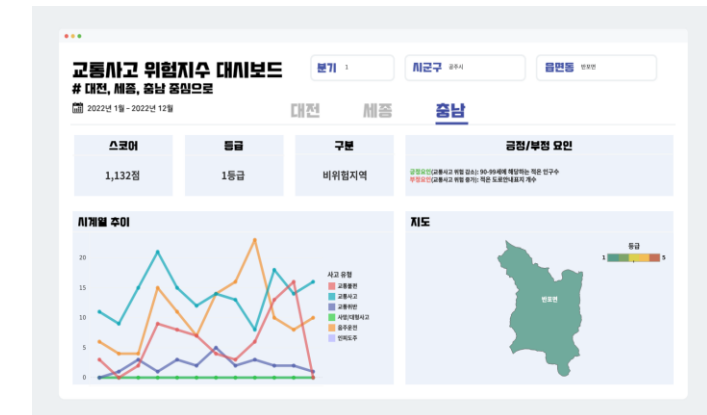
구분 (코드값)	주관적 심각도 (0 ~ 10, 고득점일수록 심각)			심각도 평균 (㉠)	가중치 (㉡)	합계 (㉠ + ㉡)
	박희진(㉠)	양문일(㉡)	최성원(㉢)			
교통불편(402)	1	1	1	1.0	1	2.0
교통위반(403)	2	3	3	2.7	2	4.7
음주운전(406)	5	4	5	4.7	5	9.7
인피도주(405)	8	7	8	7.7	10	17.7
교통사고(401)	6	10	7	7.7	30	37.7
사망대형사고(404)	10	10	10	10.0	50	60.0



2) 교통사고 위험 평점화 모형



3) 대시보드



2

세부 수행 내용

- ① 수행 프로세스
- ② 교통사고 위험 평점화 모형 개발
 - 1) 종속변수 정의
 - 2) 독립변수 수집
 - 3) 모형 개발
 - 4) 5등급 시스템 구축
- ③ 대시보드
 - 1) 구성요소
 - 2) 시연영상

2 세부 수행 내용

① 수행 프로세스

“과제 수행 프로세스를 5단계로 요약”

□ 목표 산출물 : 시도별 교통사고 위험 평점화 모형 & 5등급 체계

1

112 신고 데이터

✓ 교통사고 신고유형 : 2022년 전체

✓ 교통불편, 교통사고, 음주운전 등

※ 교통사고 위험지수 산출에 활용

2

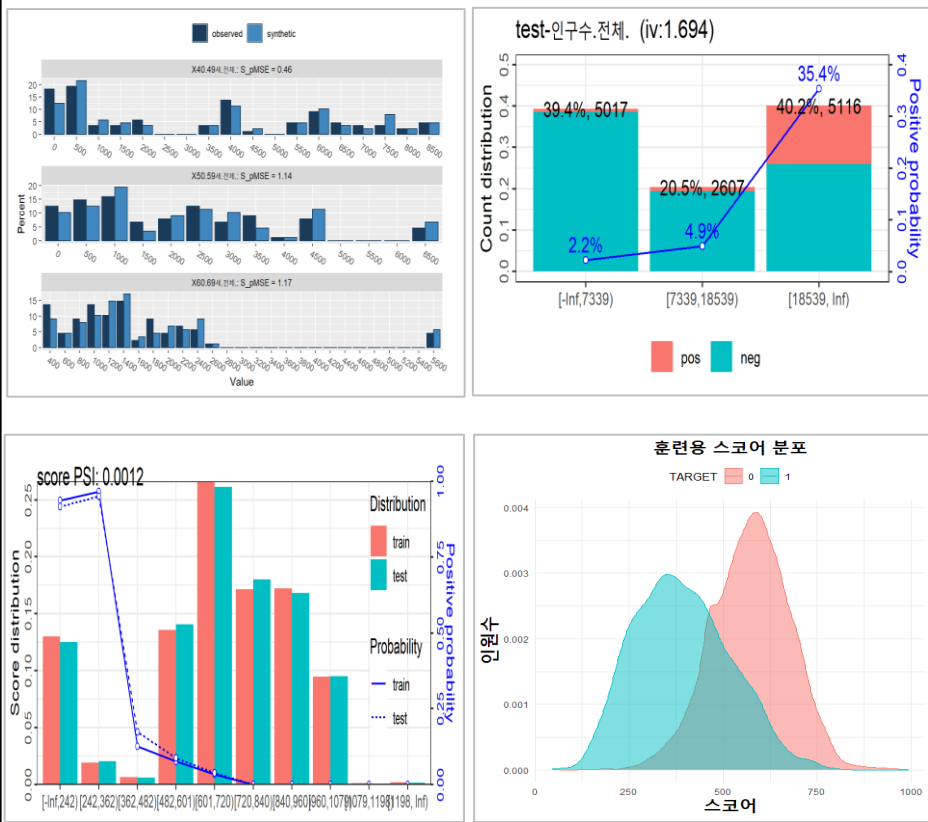
추가변수 탐색

✓ 공공데이터 수집

✓ 파생변수 생성 등

3

교통사고 위험평가 스코어카드 개발
(대전,충남,세종)



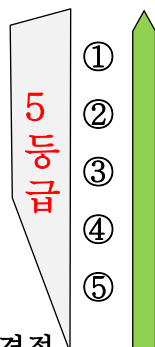
4

교통사고 위험지역 예측모형

✓ 시도별 특화모형
(대전, 충남, 세종)

✓ 예측 확률에 따른
5등급 체계
(1~5)

※ 사고위험률 파악 & 지역별 정책 결정



5

교통사고 위험 5등급 체계

등급	건수	비위험지역	위험지역	사고위험률
1등급	55	55	0	0.0%
2등급	11	10	1	9.1%
3등급	8	7	1	12.5%
4등급	3	2	1	33.3%
5등급	11	0	11	100.0%

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발

“ 모형 개발 프로세스를 5단계로 요약 ”



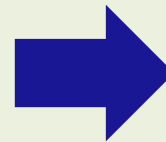
2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 종속변수 정의

“모형 개발을 위해 종속변수(Y) 정의”

□ 신용평가모형이란?

- ① 목적 : 신용대출 신청자중 『연체발생가능자』를 사전차단
- ② 기본사상 : 신용 우량 · 불량자를 구분하는 이진분류모형
- ③ 차별점
 - 예측 확률을 스코어 형태로 변환
 - 스코어를 활용하여 등급체계를 구성 → 정책 등 의사 결정에 활용



1. 시도별 교통사고 위험지역 선정
 2. 위험등급별 정책 제언
- 신용평가모형 방법론 적용

□ 종속변수(=교통사고 위험지역 여부) 정의 - 교통 위험 최종 가중치 도출

구분 (코드값)	주관적 심각도 (0 ~ 10, 고득점일수록 심각)			심각도 평균 (㉔)	가중치 (㉕)	합계 (㉔ + ㉕)
	박희진(㉖)	양문일(㉗)	최성원(㉘)			
교통불편(402)	1	1	1	1.0	1	2.0
교통위반(403)	2	3	3	2.7	2	4.7
음주운전(406)	5	4	5	4.7	5	9.7
인피도주(405)	8	7	8	7.7	10	17.7
교통사고(401)	6	10	7	7.7	30	37.7
사망대형사고(404)	10	10	10	10.0	50	60.0

- ✓ 사고 유형별 주관적 심각도 설문조사 수행
- ✓ 설문조사 기반 심각도 평균 산정
- ✓ 심각도 평균, 가중치 합산

“교통 위험 최종 가중치 도출”

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 종속변수 정의

“모형 개발을 위해 종속변수(Y) 정의”

□ 종속변수(=교통사고 위험지역 여부) 정의 - (분기, 읍면동)조합의 가중치 산출

① 대전, 세종, 충남의 분기별 읍면동 교통사고 위험 가중치 합산 산출

② 계절성 존재(유동인구, 날씨, 명절 등) → 읍면동 + 분기 추가로 관측



예시1) 세종시, 고운동

✓ 1월 3일 **교통사고** 신고 1건 : $1 \times 37.7 = 37.7$

✓ 1월 4일 **교통위반** 신고 1건 : $1 \times 4.7 = 4.7$

예시2) 세종시, 고운동 1분기(1월~3월)
교통사고 위험 가중치 = 2351.5

분기	시도명	읍면동	교통가중치 (1분기 총계)	위험지역여부
1	세종특별자치시	고운동	2351.5	0

시도명	읍면동	일자	신고건수 (총건수)	신고건수 (교통사고)	신고건수 (교통불편)	신고건수 (교통위반)	신고건수 (사망대형사고)	신고건수 (인피도주)	신고건수 (음주운전)	교통가중치 (총계)
세종특별자치시	고운동	2022-01-03	1	1	0	0	0	0	0	37.7
세종특별자치시	고운동	2022-01-04	1	0	0	1	0	0	0	4.7
세종특별자치시	고운동	2022-01-05	2	0	2	0	0	0	0	4

...

세종특별자치시	고운동	2022-03-29	4	3	0	0	0	0	1	122.8
세종특별자치시	고운동	2022-03-30	1	0	1	0	0	0	0	2
세종특별자치시	고운동	2022-03-31	3	3	0	0	0	0	0	113.1

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 종속변수 정의

“모형 개발을 위해 종속변수(Y) 정의”

□ 종속변수(=교통사고 위험지역 여부) 정의 - 교통사고 위험지역 구분 (위험지역 = 1, 비위험지역 = 0)

순번	시도명	분기	읍면동	교통가중치 (분기별 총계)	TARGET (위험지역 여부)
1	대전광역시	1	기성동	335.4	0
2	대전광역시	4	기성동	417.5	0
3	대전광역시	2	기성동	435.9	0
4	대전광역시	3	기성동	548.4	0
5	대전광역시	1	판암2동	759.2	0
...					
301	대전광역시	3	괴정동	7331.1	0
302	대전광역시	2	갈마2동	7369.4	0
303	대전광역시	3	오정동	7438.8	0
304	대전광역시	2	신성동	7517.1	1
305	대전광역시	2	오정동	7787.3	1
306	대전광역시	4	진잠동	7859	1
...					
324	대전광역시	4	온천1동	14970.6	1
325	대전광역시	1	중촌동	18292.5	1
326	대전광역시	2	중촌동	21178.6	1
327	대전광역시	4	중촌동	23746.6	1
328	대전광역시	3	중촌동	26255.3	1

1. 대전광역시:
(분기, 읍면동) 조합 328건
2. 교통가중치 7517.1 이상부터
교통사고 위험지역으로 정의
→ 종속변수 정의 완료

위험지역 구분 지점(=Elbow Point)

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 종속변수 정의

“모형 개발을 위해 종속변수(Y) 정의”

□ 종속변수(=교통사고 위험지역 여부) 정의 - 교통사고 위험지역 구분 (위험지역 = 1, 비위험지역 = 0)

순번	시도명	분기	읍면동	교통가중치 (분기별 총계)	TARGET (위험지역 여부)
1	세종특별자치시	1	전의면	322.5	0
2	세종특별자치시	4	전동면	334.7	0
3	세종특별자치시	3	연동면	337.3	0
4	세종특별자치시	4	연동면	356.8	0
5	세종특별자치시	1	연기면	432.3	0
...					
72	세종특별자치시	2	금남면	2719.2	0
73	세종특별자치시	2	장군면	2742.2	0
74	세종특별자치시	3	금남면	2770.4	0
75	세종특별자치시	3	고운동	2817.4	1
76	세종특별자치시	4	금남면	3398.9	1
77	세종특별자치시	1	새롬동	3988	1
...					
84	세종특별자치시	1	조치원읍	6942.9	1
85	세종특별자치시	4	새롬동	7106.1	1
86	세종특별자치시	2	조치원읍	8068.3	1
87	세종특별자치시	4	조치원읍	8635.3	1
88	세종특별자치시	3	조치원읍	9201.3	1

1. 세종특별자치시:

(분기, 읍면동) 조합 88건

2. 교통가중치 2817.4 이상부터

교통사고 위험지역으로 정의

→ 종속변수 정의 완료

위험지역 구분 지점(=Elbow Point)

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 종속변수 정의

“모형 개발을 위해 종속변수(Y) 정의”

□ 종속변수(=교통사고 위험지역 여부) 정의 - 교통사고 위험지역 구분 (위험지역 = 1, 비위험지역 = 0)

순번	시도명	분기	읍면동	교통가중치 (분기별 총계)	TARGET (위험지역 여부)
1	충청남도	4	남이면	2	0
2	충청남도	1	마산면	2	0
3	충청남도	1	이원면	2	0
4	충청남도	2	남이면	4.7	0
5	충청남도	3	미산면	9.7	0
...					
785	충청남도	2	신방동	7499.4	0
786	충청남도	1	취암동	7608.8	0
787	충청남도	3	예산읍	7619.1	0
788	충청남도	1	청룡동	7622.2	1
789	충청남도	4	신방동	7794	1
790	충청남도	1	홍성읍	7951.2	1
...					
824	충청남도	2	성정2동	14669	1
825	충청남도	2	부성2동	14962	1
826	충청남도	4	신안동	15053	1
827	충청남도	3	배방읍	15110	1
828	충청남도	4	배방읍	15545.3	1

1. 충청남도:

(분기, 읍면동) 조합 828건

2. 교통가중치 7622.2 이상부터

교통사고 위험지역으로 정의

→ 종속변수 정의 완료

위험지역 구분 지점(=Elbow Point)

2 세부 수행 내용

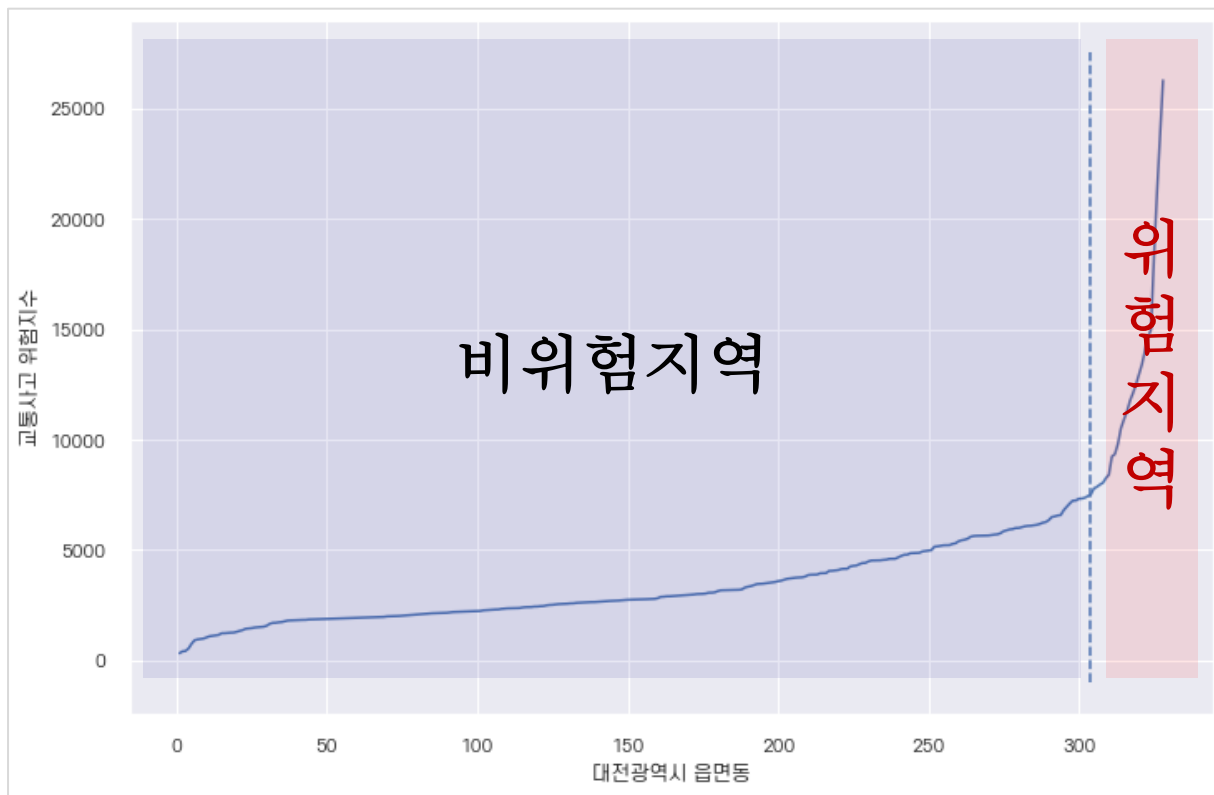
② 교통사고 위험 평점화 모형 개발 - 종속변수 정의

“모형 개발을 위해 종속변수(Y) 정의”

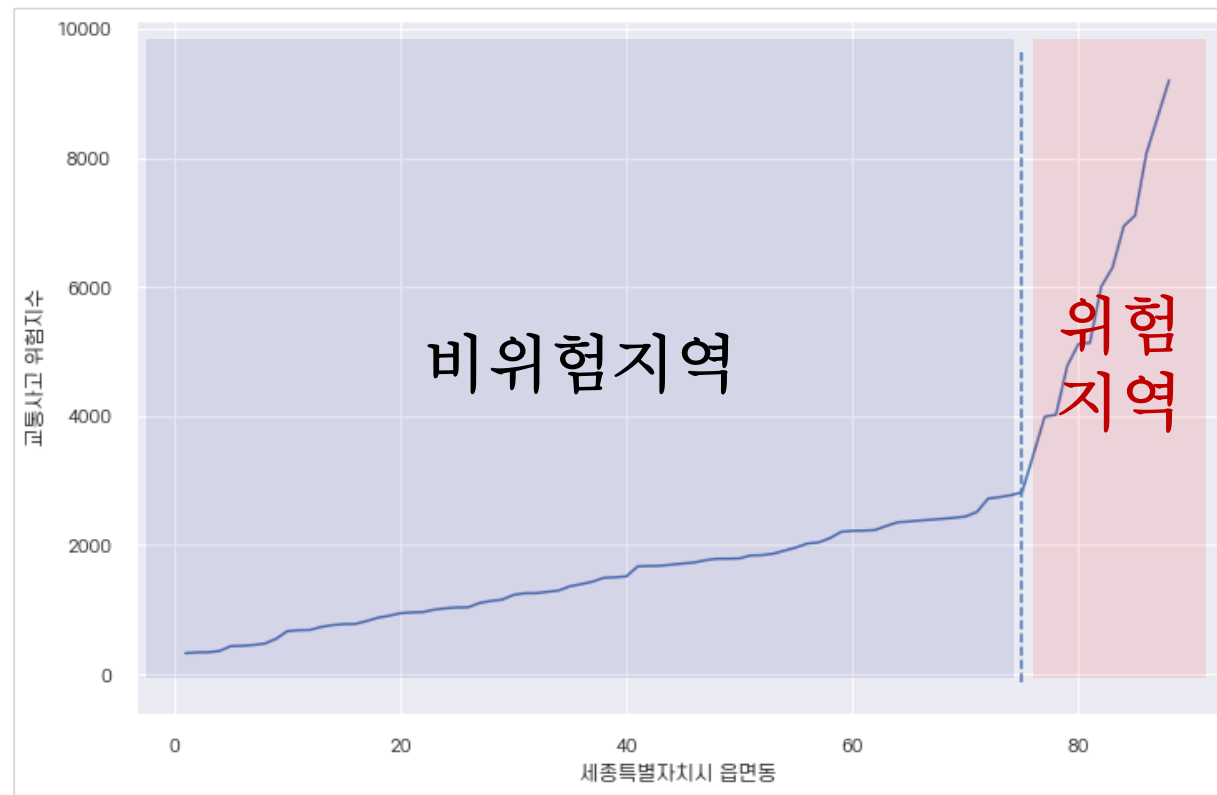
□ 종속변수(=교통사고 위험지역 여부) 정의 - 위험지역 구분 지점(=Elbow Point) 산출원리

➤ Elbow Point Detection 라이브러리 ‘Kneed⁴⁾’ 활용

<대전광역시>



<세종특별자치시>



4) Elbow Point Detection Library, <https://kneed.readthedocs.io/en/stable/>

2 세부 수행 내용

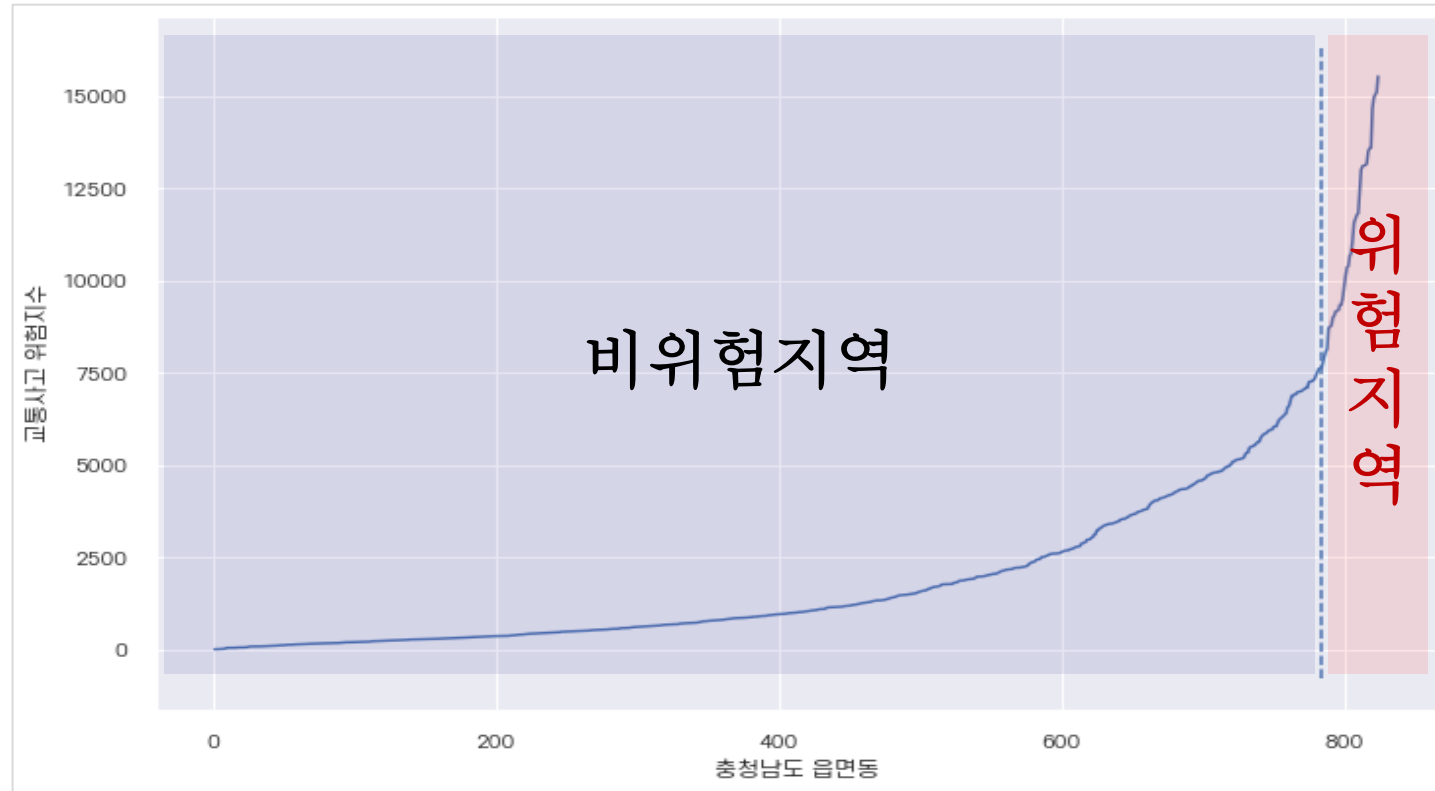
② 교통사고 위험 평점화 모형 개발 - 종속변수 정의

“모형 개발을 위해 종속변수(Y) 정의”

□ 종속변수(=교통사고 위험지역 여부) 정의 - 위험지역 구분 지점(=Elbow Point) 산출원리

➤ Elbow Point Detection 라이브러리 ‘Kneed⁴⁾’ 활용

<충청남도>



2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 독립변수 수집

“모형 개발을 위해 독립변수(X) 수집”

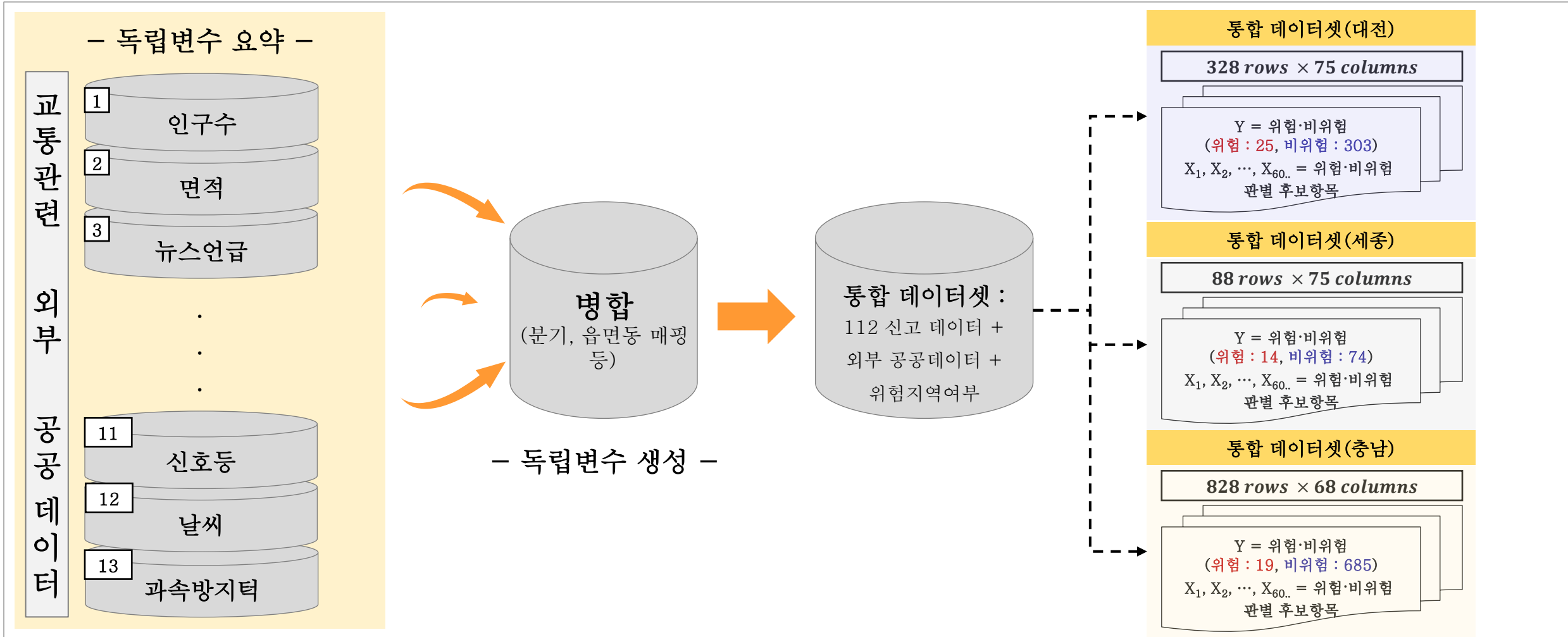
□ 수집 공공데이터 리스트

구분	No	대분류	원천	데이터명	항목 예시
외부데이터	1	인구수	행정안전부	주민등록 인구통계	세대수, 세대당 인구수, 인구수(남/여/전체), 남녀비율, 인구수(전체 나이대/남자 나이대/ 여자 나이대) (총 39 column)
	2	면적	KOSIS	지역통계 - 행정구역	면적
	3	뉴스언급	스마트치안 빅데이터 플랫폼	국내 교통사고	뉴스언급건수(시도,시군구,읍면동) 뉴스언급건수(시군구,읍면동)
	4	보안등	공공데이터	전국보안등 표준 데이터	보안등(설치개수)
	5	무인단속카메라	공공데이터	전국무인교통단속카메라표준데이터 경찰청 세종특별자치시경찰청_무인교통단속카메라	무인단속카메라(개수) 무인단속카메라(평균제한속도)
	6	도로안내표지	공공데이터	전국도로안내표지표준데이터	도로안내표지_개수
	7	음식점	지방행정 인허가 데이터 개방	식품 - 음식점 - 일반음식점	음식점(개수)
	8	버스정류장	공공데이터	국토교통부_전국 버스정류장 위치정보	버스정류장(개수)
	9	어린이보호구역	공공데이터	전국어린이보호구역표준데이터	어린이보호구역(개수)
	10	주차장	공공데이터	전국주차장정보표준데이터	주차장(개수)
	11	신호등	공공데이터	전국신호등표준데이터	신호등(개수)
	12	날씨	기상청 기상자료 개방포털	기후 통계분석	총 16Column 강수일수,결빙일수,강수량.mm.,평균기온 등
	13	과속방지턱	지방행정 인허가 데이터 개방	생활밀착데이터 - 과속방지턱정보	과속방지턱(개수)

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 독립변수 수집

✓ 통합 데이터셋 산출과정 요약



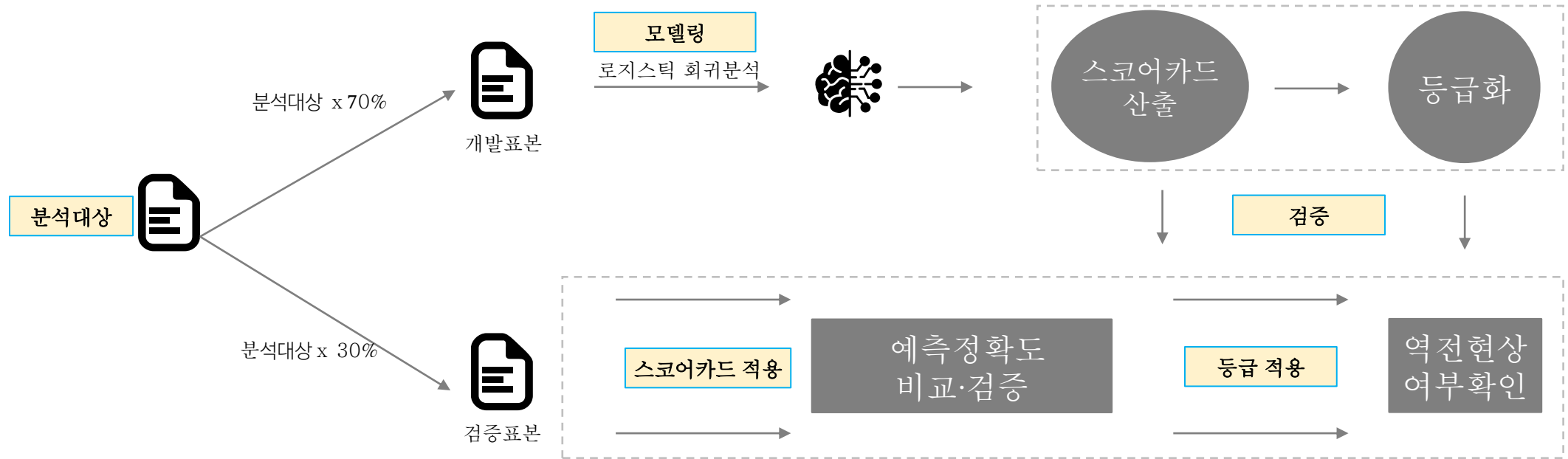
“ 모형 개발을 위해 독립변수(X) 수집 ”

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 모형 개발

“ 시도별 교통사고 위험 평점 모형 개발 ”

✓ 개발표본을 스코어카드 모델링에 활용 & 검증표본에 스코어카드를 적용하여 성능검증 진행



구분	분석대상(대전,세종,충남)			
	소계	위험지역	비위험지역	사고위험률(%)
(A) 개발표본 (70%)	총화추출법으로 종속변수 비율 균등하게 데이터 분할			
(B) 검증표본 (30%)				
총계				

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 대전광역시 위험 모형

“

평가항목 10개로 구성된
대전광역시 스코어카드

”

✓ 최종 스코어카드 (대전광역시)

No	항목명	구간	건수	구성비	비위험지역	위험지역	사고위험률	woe	bin_iv	total_iv	breaks	is_special_values	points
1	X20.29세.전체.	[-Inf,4552)	206	90.0%	201	5	2.4%	-1.1705	0.7655	2.4727	4552	FALSE	143
	X20.29세.전체.	[4552, Inf)	23	10.0%	11	12	52.2%	2.6104	1.7072	2.4727	Inf	FALSE	-318
2	무인단속카메라.개수.	missing	2	0.9%	2	0	0.0%	1.7636	0.0804	1.4017	missing	TRUE	-97
	무인단속카메라.개수.	[-Inf,12)	199	86.9%	192	7	3.5%	-0.8448	0.4364	1.4017	12	FALSE	47
	무인단속카메라.개수.	[12, Inf)	28	12.2%	18	10	35.7%	1.8790	0.8849	1.4017	Inf	FALSE	-103
3	도로안내표지_개수	[-Inf,15)	66	28.8%	65	1	1.5%	-1.6510	0.4091	2.1986	15	FALSE	135
	도로안내표지_개수	[15,27)	82	35.8%	80	2	2.4%	-1.1655	0.3027	2.1986	27	FALSE	95
	도로안내표지_개수	[27,56)	43	18.8%	40	3	7.0%	-0.0669	0.0008	2.1986	56	FALSE	5
	도로안내표지_개수	[56,72)	26	11.4%	23	3	11.5%	0.4865	0.0331	2.1986	72	FALSE	-40
	도로안내표지_개수	[72, Inf)	12	5.2%	4	8	66.7%	3.2165	1.4530	2.1986	Inf	FALSE	-263

...

9	무인단속카메라.평균제한속도.	missing	2	0.9%	2	0	0.0%	1.7636	0.0804	0.1038	missing	TRUE	-60
	무인단속카메라.평균제한속도.	[-Inf,43.33333333)	198	86.5%	184	14	7.1%	-0.1091	0.0098	0.1038	43.33333333	FALSE	4
	무인단속카메라.평균제한속도.	[43.33333333, Inf)	29	12.7%	26	3	10.3%	0.3073	0.0136	0.1038	Inf	FALSE	-10
10	세대당.인구	[-Inf,2.113333333)	125	54.6%	112	13	10.4%	0.3698	0.0874	0.2518	2.113333333	FALSE	-57
	세대당.인구	[2.113333333, Inf)	104	45.4%	100	4	3.8%	-0.6955	0.1644	0.2518	Inf	FALSE	108

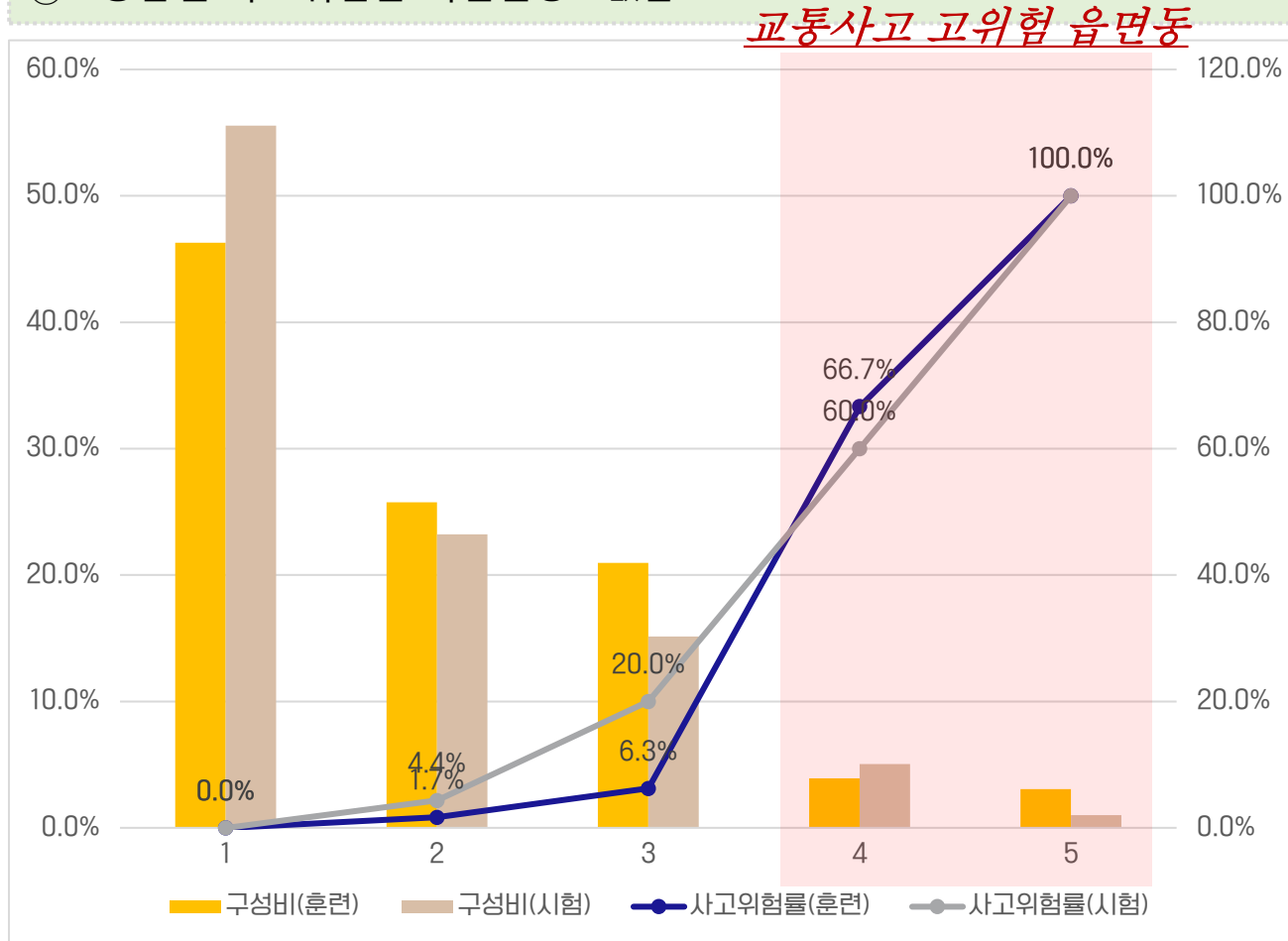
※ 전체 항목 '대전_위험평점_모형정의서_fn.xlsx'내 'scorecard(1)' sheet 참조요망

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 대전광역시 위험 모형

위험등급 적용결과 요약

① 등급별 사고위험률 역전현상⁵⁾ 없음



“

교통 관련 정책 결정을 위해 5등급 시스템 구축

”

▶ 개발표본 5등급 적용결과

등급	구간	전체건수	비위험지역	위험지역	구성비	사고위험률
1	[858, Inf)	106	106	-	46.3%	0.0%
2	[666, 858)	59	58	1	25.8%	1.7%
3	[473, 666)	48	45	3	21.0%	6.3%
4	[281, 473)	9	3	6	3.9%	66.7%
5	[-Inf, 281)	7	-	7	3.1%	100.0%

▶ 검증표본 5등급 적용결과

등급	구간	전체건수	비위험지역	위험지역	구성비	사고위험률
1	[858, Inf)	55	55	-	55.6%	0.0%
2	[666, 858)	23	22	1	23.2%	4.4%
3	[473, 666)	15	12	3	15.2%	20.0%
4	[281, 473)	5	2	3	5.1%	60.0%
5	[-Inf, 281)	1	-	1	1.0%	100.0%

5) 등급이 높음에도 불구하고 사고위험률(%)이 높은 현상

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 세종특별자치시 위험 모형

“

평가항목 21개로 구성된
세종특별자치시 스코어카드

”

✓ 최종 스코어카드 (세종특별자치시)

No	항목명	구간	건수	구성비	비위험지역	위험지역	사고위험률	woe	bin_iv	total_iv	breaks	is_special_values	points
1	세대수	[-Inf,3104)	5,189	17.5%	5,158	31	0.6%	-3.4608	0.6930	3.9619	3104	FALSE	165
	세대수	[3104,10820)	18,359	61.8%	17,897	462	2.5%	-2.0033	1.2431	3.9619	10820	FALSE	95
	세대수	[10820, Inf)	6,178	20.8%	1,896	4,282	69.3%	2.4682	2.0258	3.9619	Inf	FALSE	-117
2	세대당.인구	[-Inf,1.8)	5,129	17.3%	5,104	25	0.5%	-3.6654	0.7306	2.4786	1.8	FALSE	70
	세대당.인구	[1.8,1.847)	6,233	21.0%	5,865	368	5.9%	-1.1152	0.1762	2.4786	1.847	FALSE	21
	세대당.인구	[1.847,2.51)	5,237	17.6%	2,786	2,451	46.8%	1.5254	0.6127	2.4786	2.51	FALSE	-29
	세대당.인구	[2.51,2.73)	6,562	22.1%	4,681	1,881	28.7%	0.7418	0.1530	2.4786	2.73	FALSE	-14
	세대당.인구	[2.73, Inf)	6,565	22.1%	6,515	50	0.8%	-3.2163	0.8061	2.4786	Inf	FALSE	61
3	인구수.남.	[-Inf,13944)	23,557	79.2%	23,064	493	2.1%	-2.1920	1.7999	3.8305	13944	FALSE	-60
	인구수.남.	[13944, Inf)	6,169	20.8%	1,887	4,282	69.4%	2.4730	2.0306	3.8305	Inf	FALSE	67

...

21	보안등.설치개수.	missing	12,143	40.8%	8,740	3,403	28.0%	0.7101	0.2572	2.2636	missing	TRUE	-24
	보안등.설치개수.	[-Inf,710)	6,885	23.2%	6,885	-	0.0%	-7.1938	1.9836	2.2636	710	FALSE	241
	보안등.설치개수.	[710, Inf)	10,698	36.0%	9,326	1,372	12.8%	-0.2632	0.0228	2.2636	Inf	FALSE	9

※ 전체 항목 '세종_위험평점_모형정의서_fn.xlsx'내 'scorecard(1)' sheet 참조요망

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 세종특별자치시 위험 모형

“

평가항목 21개로 구성된
세종특별자치시 스코어카드

”

- 세종시 학습 데이터 **100건 미만**, 데이터 확보를 위해 데이터 증폭 알고리즘 ‘synthpop⁶⁾’ 기법을 활용
- 세종시 원본 교통데이터를 이용하여 유사한 분포를 지닌 세종시 ‘재현데이터’ 생성

재현자료의 각 컬럼은 원본자료, 조건부 분포의 매개 변수, 그리고 이미 생성된 변수의 조건부 분포에 따라 생성됨

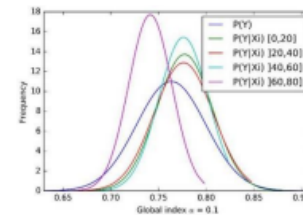
민감정보

원본자료

obs.No	y	duration	credit.amount	age	account balance	Credit history	purpose
1	good	6	1,169	67	<0 DM	critical account	radio/television
2	bad	48	5,951	22	0<=...<200 DM	credits paid back till now	radio/television
3	good	12	2,096	49	no account	critical account	education
4	good	42	7,882	45	<0 DM	credits paid back till now	furniture/equipment

재현자료

obs.No	y	duration	credit.amount	age	account balance	Credit history	purpose
1	good	33	2,672	43	<0 DM	critical account	radio/television
2	bad	32	4,751	39	0<=...<200 DM	credits paid back till now	radio/television
3	good	10	839	46	no account	critical account	education
4	good	41	4,753	23	<0 DM	credits paid back till now	furniture/equipment



변수 간 조건부 확률분포를
기반으로 하는 재현데이터 생성

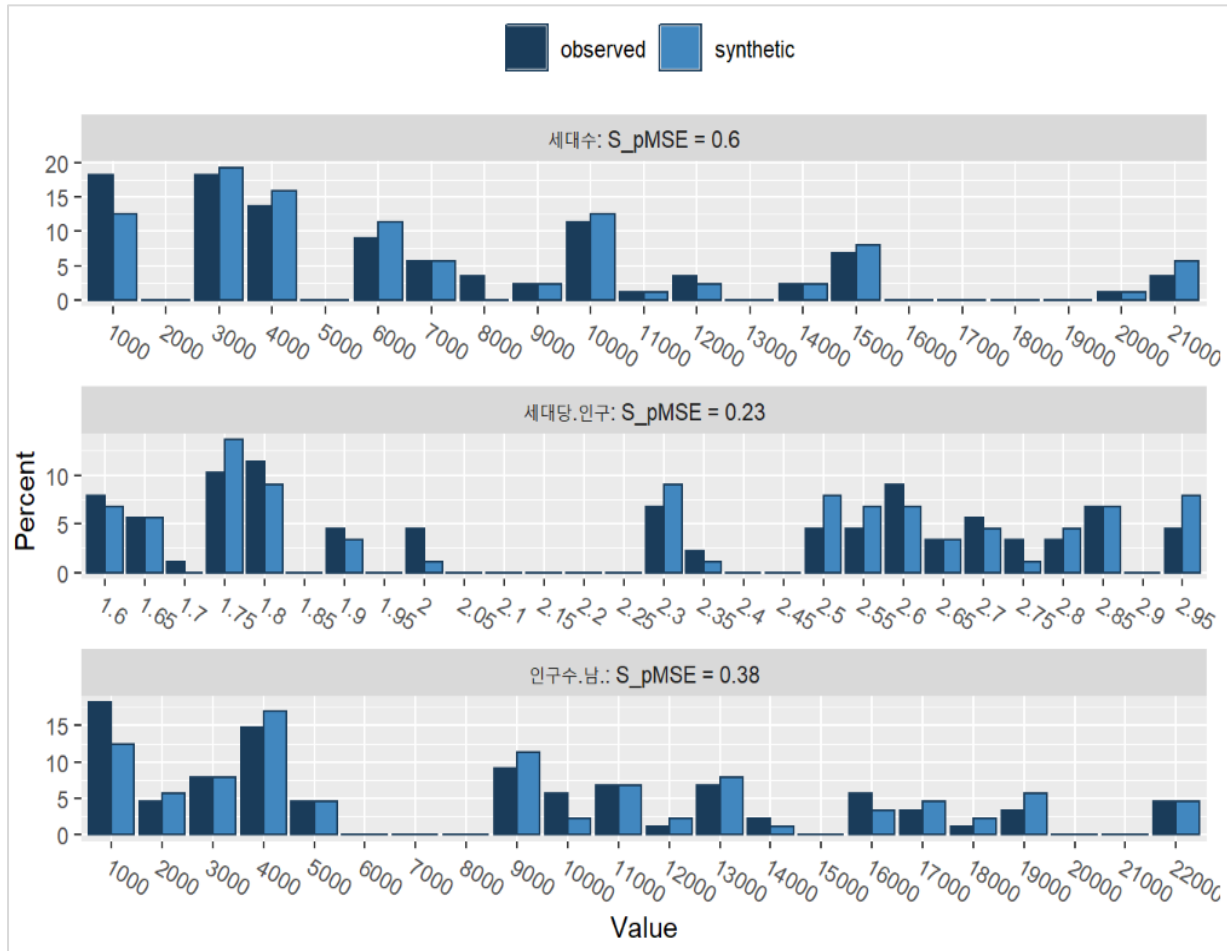
출처 : 개인신용정보 재현데이터 생성, 구영현(2019)

6) synthpop: Generating Synthetic Versions of Sensitive Microdata for Statistical Disclosure Control,
<https://cran.r-project.org/web/packages/synthpop/index.html>

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 세종특별자치시 위험 모형

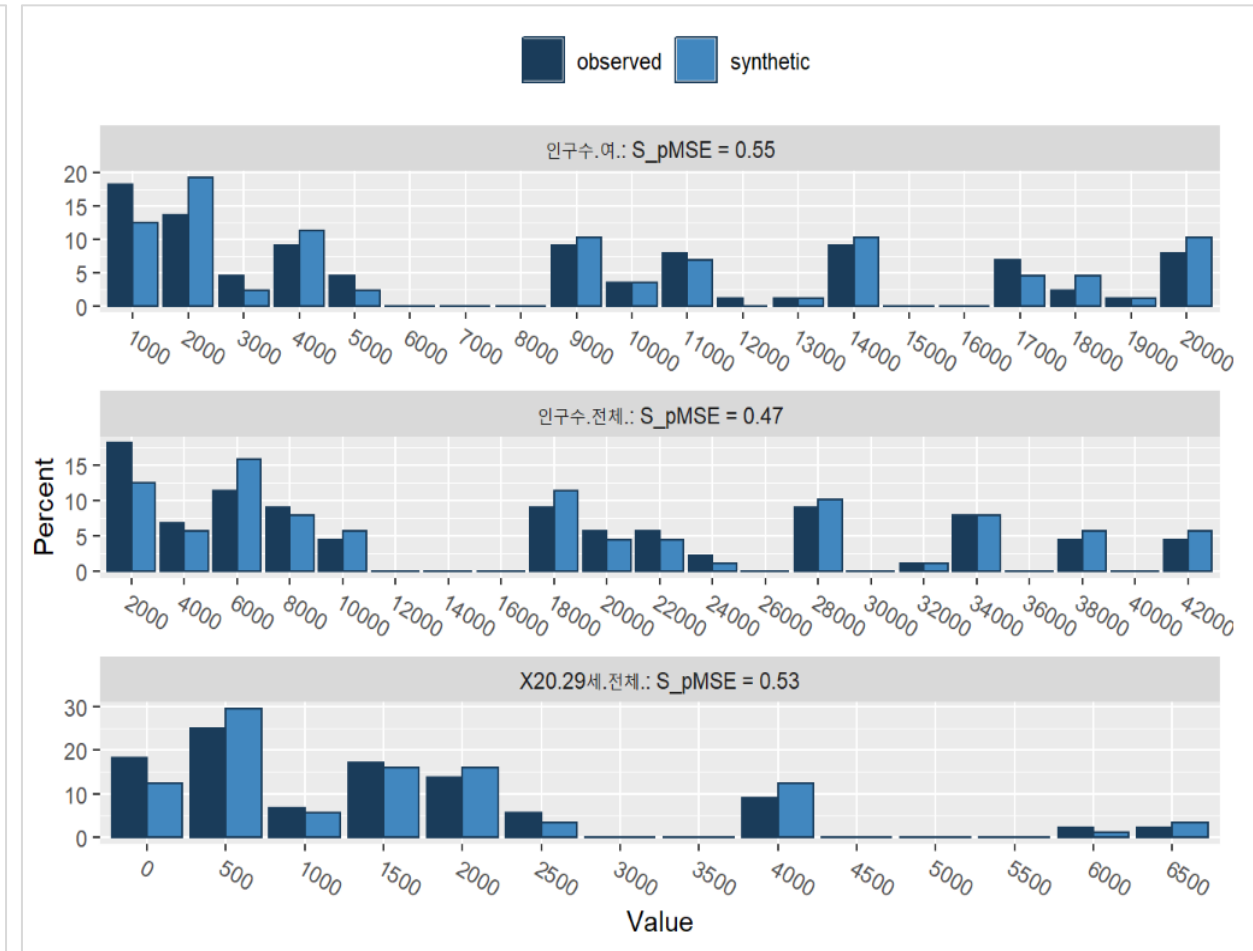
➤ 원본데이터 & 생성된 재현데이터 분포비교(1/2)



“

평가항목 21개로 구성된
세종특별자치시 스코어카드

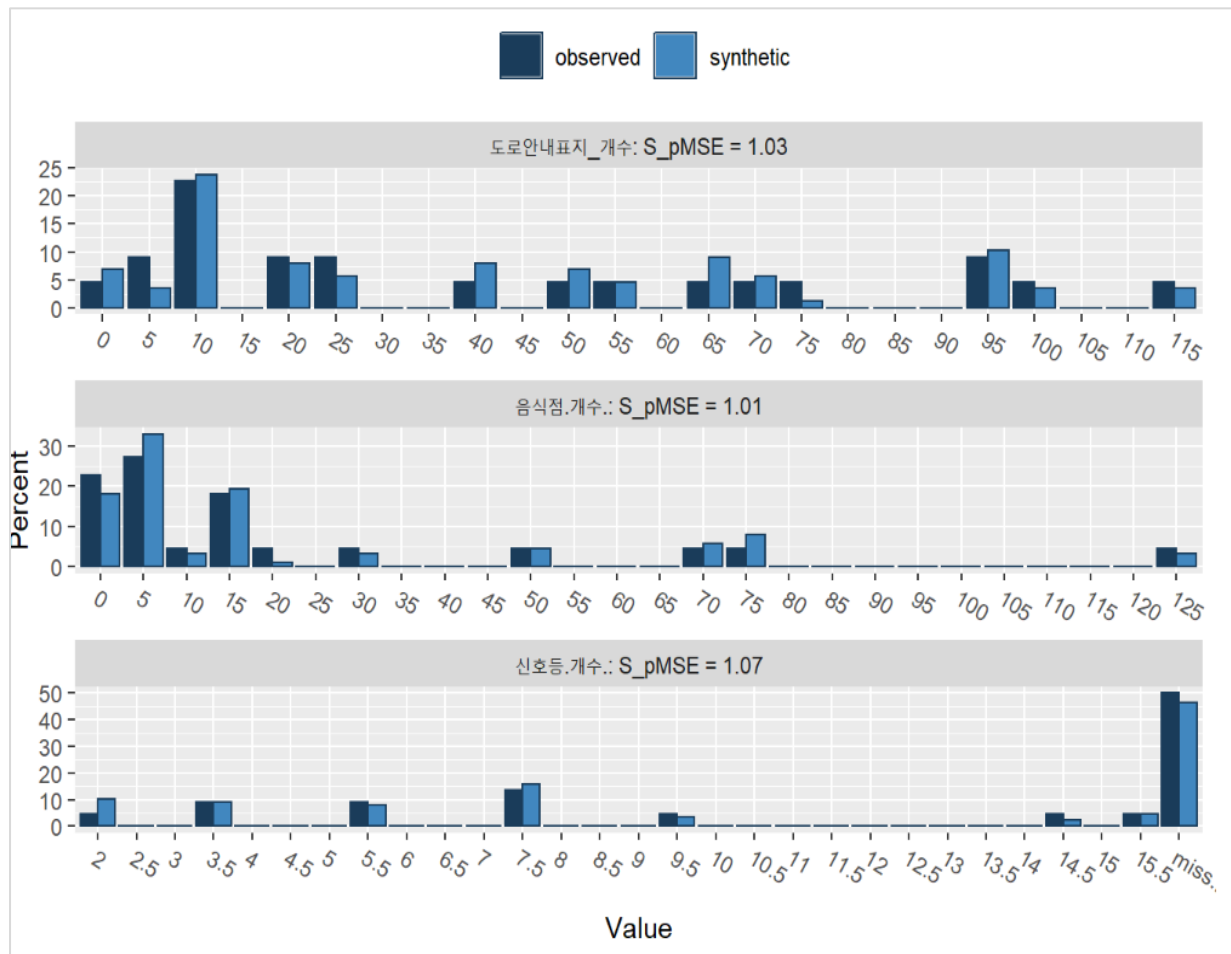
”



2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 세종특별자치시 위험 모형

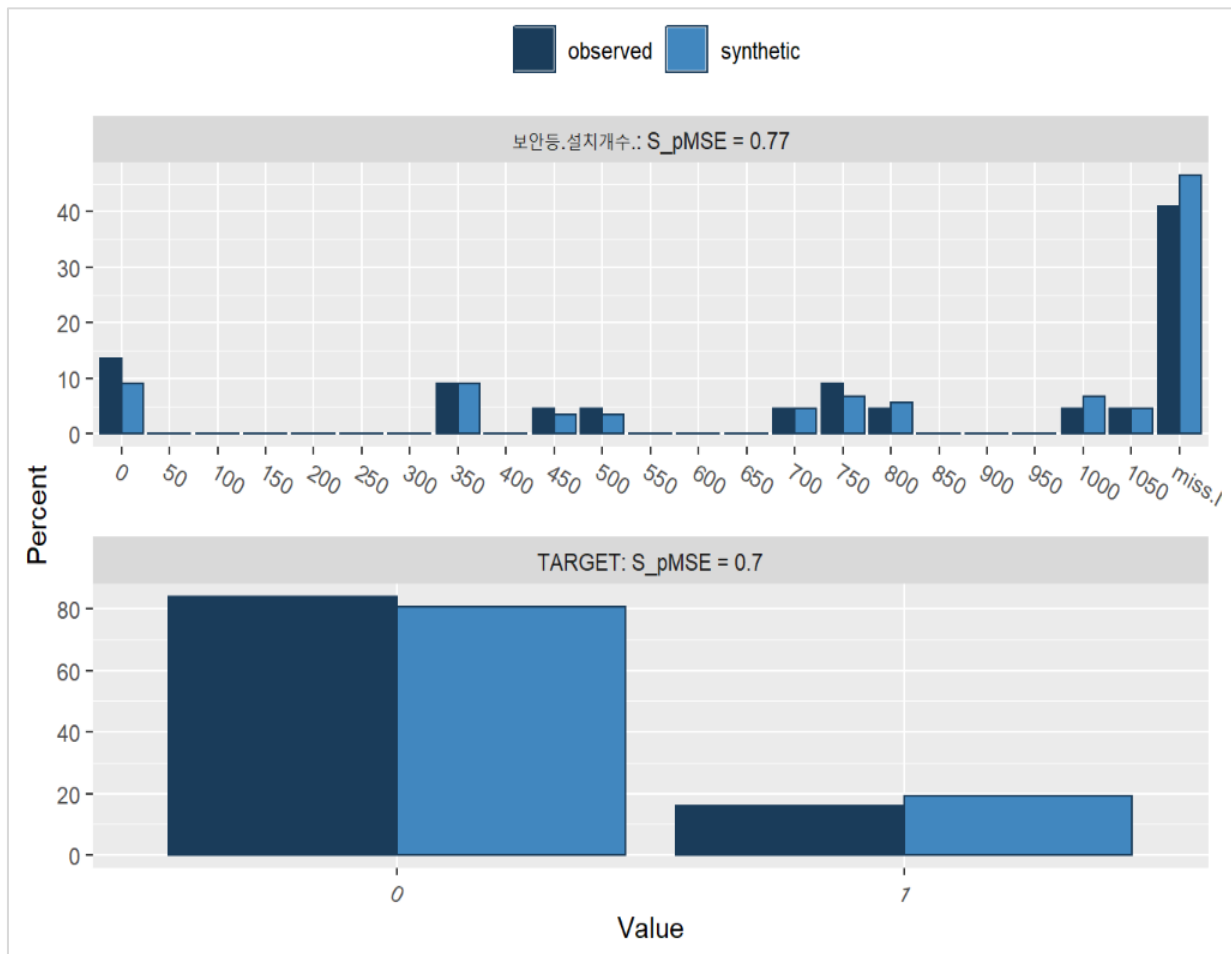
➤ 원본데이터 & 생성된 재현데이터 분포비교 (2/2)



“

평가항목 21개로 구성된
세종특별자치시 스코어카드

”



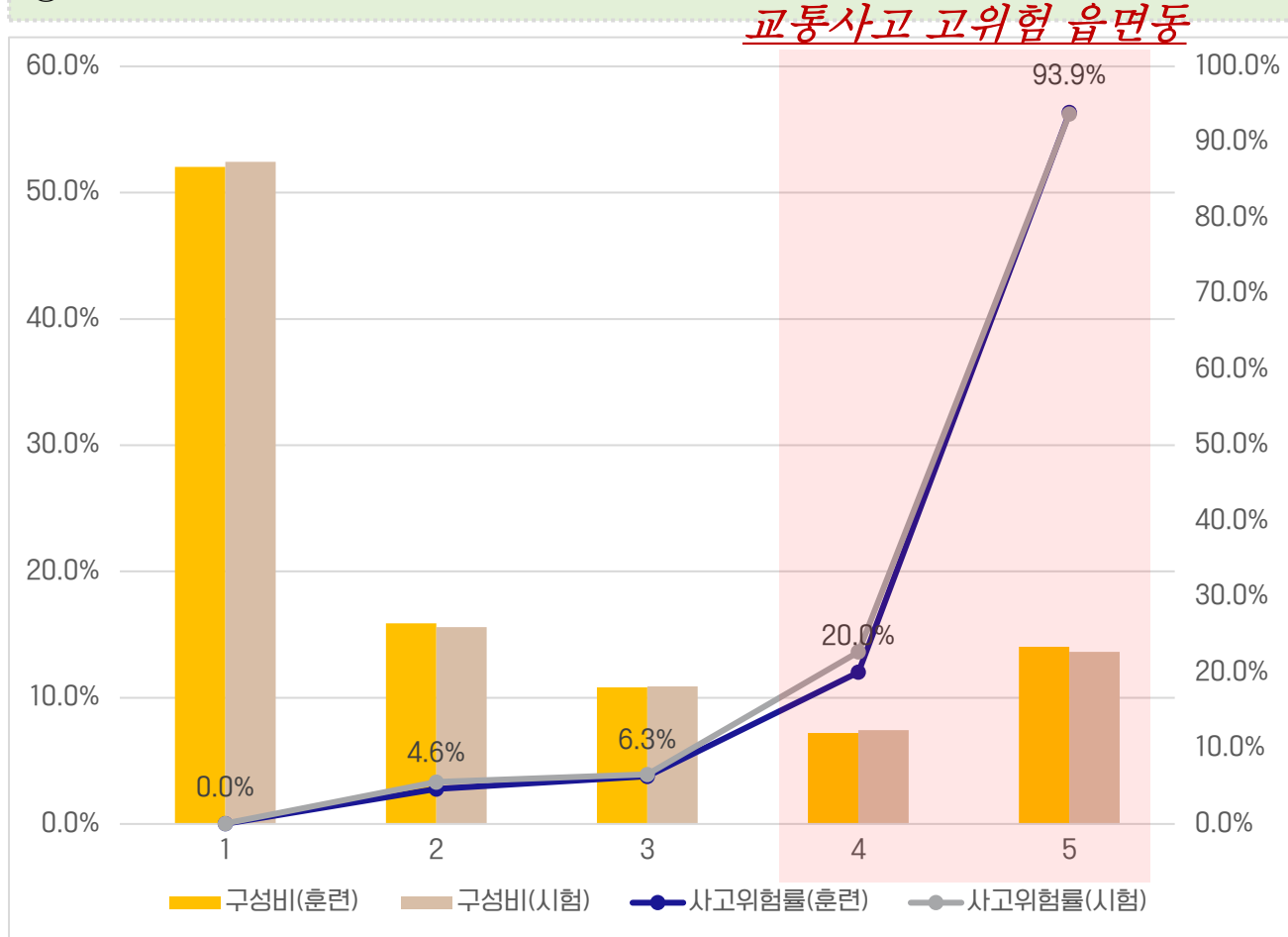
※ 전체 항목 '세종_위험평점_모형정의서_fn.xlsx' 내 '재현데이터' sheet 참조요망

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 세종특별자치시 위험 모형

위험등급 적용결과 요약

① 등급별 사고위험률 역전현상 없음



“ ”

교통 관련 정책 결정을 위해 5등급 시스템 구축

▶ 개발표본 5등급 적용결과(재현데이터)

등급	구간	전체건수	비위험지역	위험지역	구성비	사고위험률
1	[678, Inf)	15,472	15,465	7	52.0%	0.0%
2	[619, 678)	4,721	4,503	218	15.9%	4.6%
3	[560, 619)	3,221	3,018	203	10.8%	6.3%
4	[264, 560)	2,140	1,711	429	7.2%	20.0%
5	[-Inf, 264)	4,172	254	3,918	14.0%	93.9%

▶ 검증표본 5등급 적용결과(재현데이터)

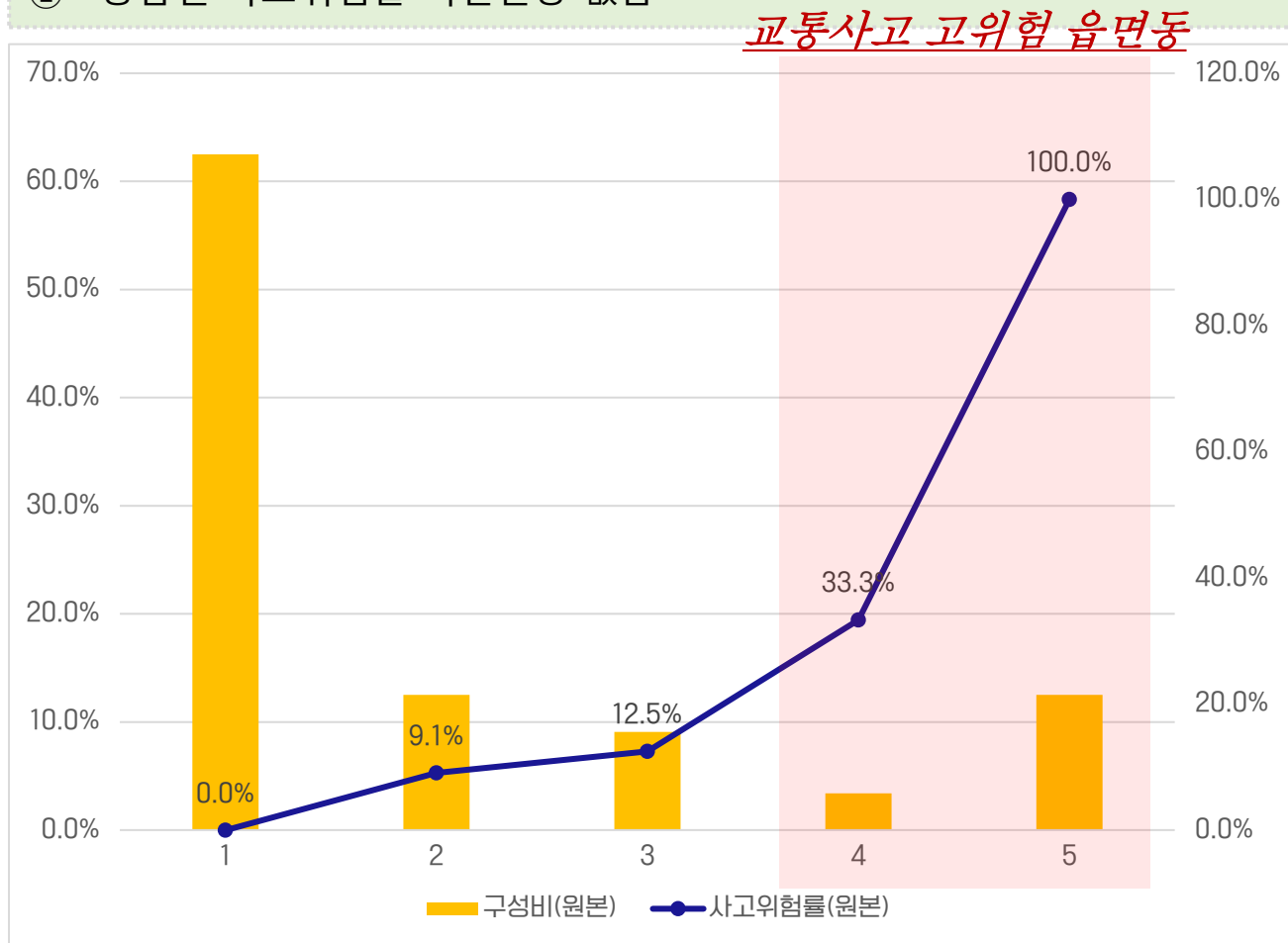
등급	구간	전체건수	비위험지역	위험지역	구성비	사고위험률
1	[678, Inf)	6,680	6,676	4	52.4%	0.1%
2	[619, 678)	1,987	1,877	110	15.6%	5.5%
3	[560, 619)	1,390	1,299	91	10.9%	6.5%
4	[264, 560)	947	732	215	7.4%	22.7%
5	[-Inf, 264)	1,736	109	1,627	13.6%	93.7%

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 세종특별자치시 위험 모형

위험등급 적용결과 요약

① 등급별 사고위험률 역전현상 없음



“

교통 관련 정책 결정을 위해 5등급 시스템 구축

”

▶ 5등급 적용결과(원본데이터)

등급	구간	전체건수	비위험지역	위험지역	구성비	사고위험률
1	[678, Inf)	55	55	-	62.5%	0.0%
2	[619, 678)	11	10	1	12.5%	9.1%
3	[560, 619)	8	7	1	9.1%	12.5%
4	[264, 560)	3	2	1	3.4%	33.3%
5	[-Inf, 264)	11	-	11	12.5%	100.0%

1. 재현데이터를 활용하여 모형개발
2. 원본데이터에 성능검증
3. 유사한 분포가 학습되었음을 확인함

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 충청남도 위험 모형

“

평가항목 11개로 구성된
충청남도 스코어카드

”

✓ 최종 스코어카드 (충청남도)

No	항목명	구간	건수	구성비	비위험지역	위험지역	사고위험률	woe	bin_iv	total_iv	breaks	is_special_values	points
1	X20.29세.여.	[-Inf,1288.666667)	524	90.0%	522	2	0.4%	-1.7856	1.3634	3.1399	1288.666667	FALSE	399
	X20.29세.여.	[1288.666667, Inf)	58	10.0%	47	11	19.0%	2.3267	1.7765	3.1399	Inf	FALSE	-520
2	X100세.이상.남.	[-Inf,0.3333333333)	443	76.1%	440	3	0.7%	-1.2092	0.6560	1.9188	0.3333333333	FALSE	166
	X100세.이상.남.	[0.3333333333,1.3333333333)	113	19.4%	109	4	3.5%	0.4739	0.0550	1.9188	1.3333333333	FALSE	-65
	X100세.이상.남.	[1.3333333333,3)	19	3.3%	16	3	15.8%	2.1050	0.4266	1.9188	3	FALSE	-289
	X100세.이상.남.	[3, Inf)	7	1.2%	4	3	42.9%	3.4912	0.7811	1.9188	Inf	FALSE	-480
3	세대당.인구	[-Inf,2.0433333333)	433	74.4%	431	2	0.5%	-1.5940	0.9622	1.7165	2.0433333333	FALSE	-26
	세대당.인구	[2.0433333333, Inf)	149	25.6%	138	11	7.4%	1.2496	0.7543	1.7165	Inf	FALSE	21

...

10	과속방지턱.개수.	missing	236	40.5%	234	2	0.8%	-0.9832	0.2531	0.3500	missing	TRUE	29
	과속방지턱.개수.	[-Inf,48)	169	29.0%	164	5	3.0%	0.2885	0.0278	0.3500	48	FALSE	-9
	과속방지턱.개수.	[48, Inf)	177	30.4%	171	6	3.4%	0.4290	0.0691	0.3500	Inf	FALSE	-13
11	남녀비율	[-Inf,1.19)	549	94.3%	538	11	2.0%	-0.1110	0.0110	0.1142	1.19	FALSE	28
	남녀비율	[1.19, Inf)	33	5.7%	31	2	6.1%	1.0381	0.1031	0.1142	Inf	FALSE	-257

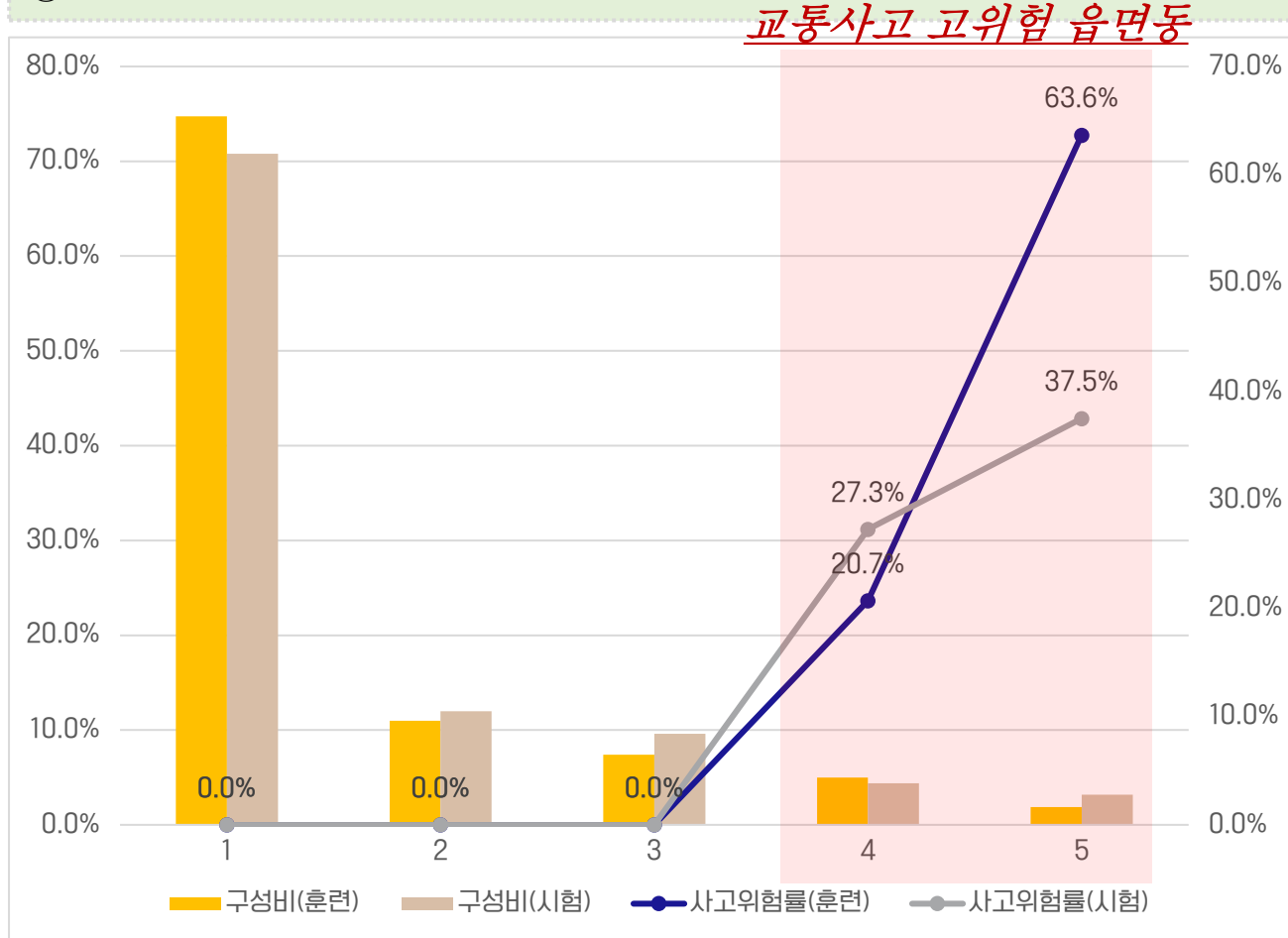
※ 전체 항목 '충남_위험평점_모형정의서_fn.xlsx'내 'scorecard(1)' sheet 참조요망

2 세부 수행 내용

② 교통사고 위험 평점화 모형 개발 - 충청남도 위험 모형

위험등급 적용결과 요약

① 등급별 사고위험률 역전현상 없음



“ 교통 관련 정책 결정을 위해 5등급 시스템 구축 ”

▶ 개발표본 5등급 적용결과

등급	구간	전체건수	비위험지역	위험지역	구성비	사고위험률
1	[-Inf, 425)	11	4	7	1.9%	63.6%
2	[425, 621)	29	23	6	5.0%	20.7%
3	[621, 817)	43	43	-	7.4%	0.0%
4	[817, 1013)	64	64	-	11.0%	0.0%
5	[1013, Inf)	435	435	-	74.7%	0.0%

▶ 검증표본 5등급 적용결과

등급	구간	전체건수	비위험지역	위험지역	구성비	사고위험률
1	[-Inf, 425)	8	5	3	3.2%	37.5%
2	[425, 621)	11	8	3	4.4%	27.3%
3	[621, 817)	24	24	-	9.6%	0.0%
4	[817, 1013)	30	30	-	12.0%	0.0%
5	[1013, Inf)	177	177	-	70.8%	0.0%

2 세부 수행 내용

③ 대시보드 - 구성요소 설명

“

분석결과 활용가능한 시도별 대시보드 개발

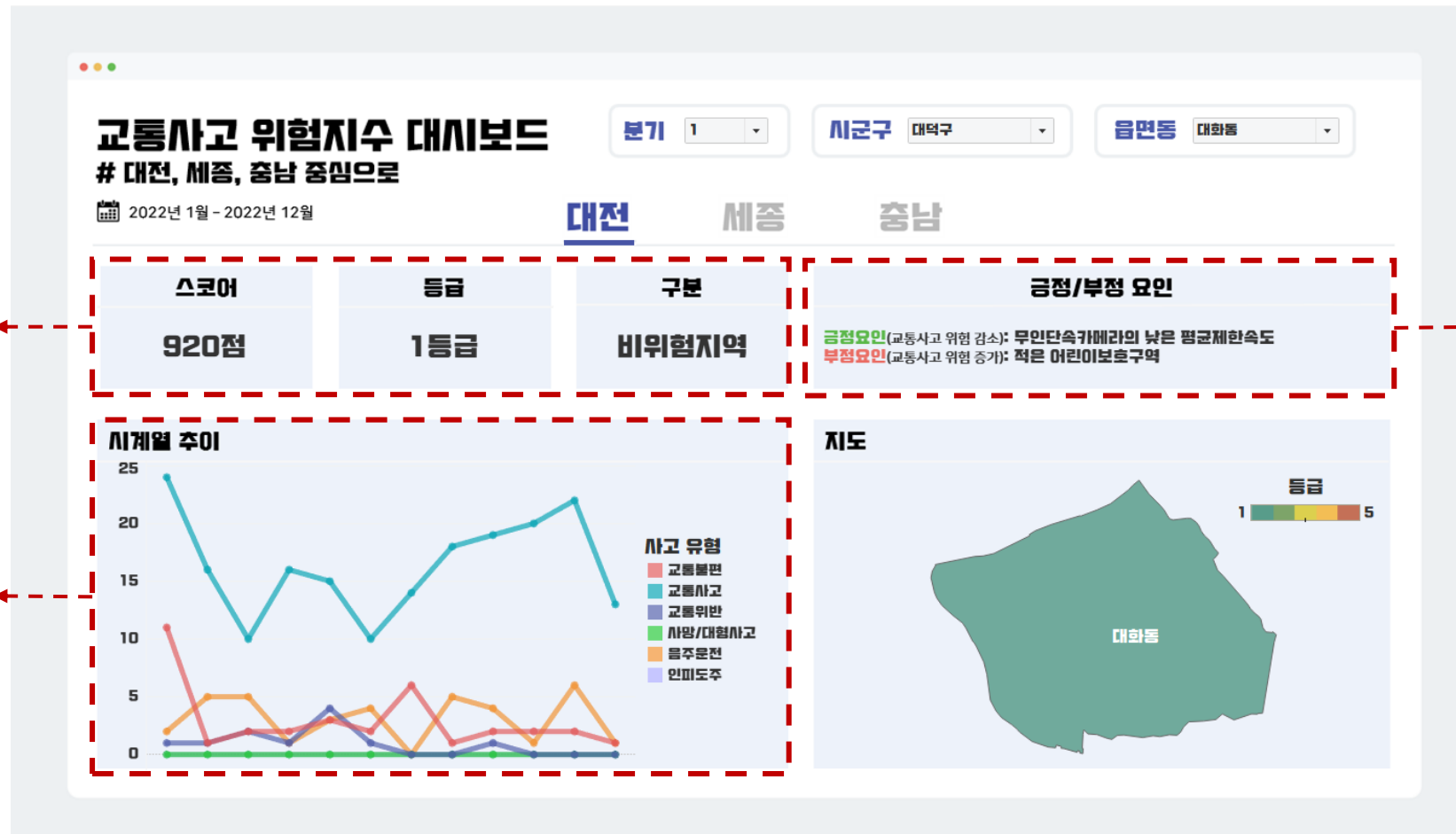
”

✓ 시각화 tool(=테블로) 활용하여 분석결과 대시보드 구현 (해당 링크를 통해 접속하면 직접 시연 가능하도록 구현함)

: https://public.tableau.com/views/_16763648445880/sheet0?:language=ko-KR&publish=yes&:display_count=n&:origin=viz_share_link

위험 평점 모형
적용 결과값

사고 유형별
시계열 추이



등급산출시 작용한
긍·부정 요인을
표시함

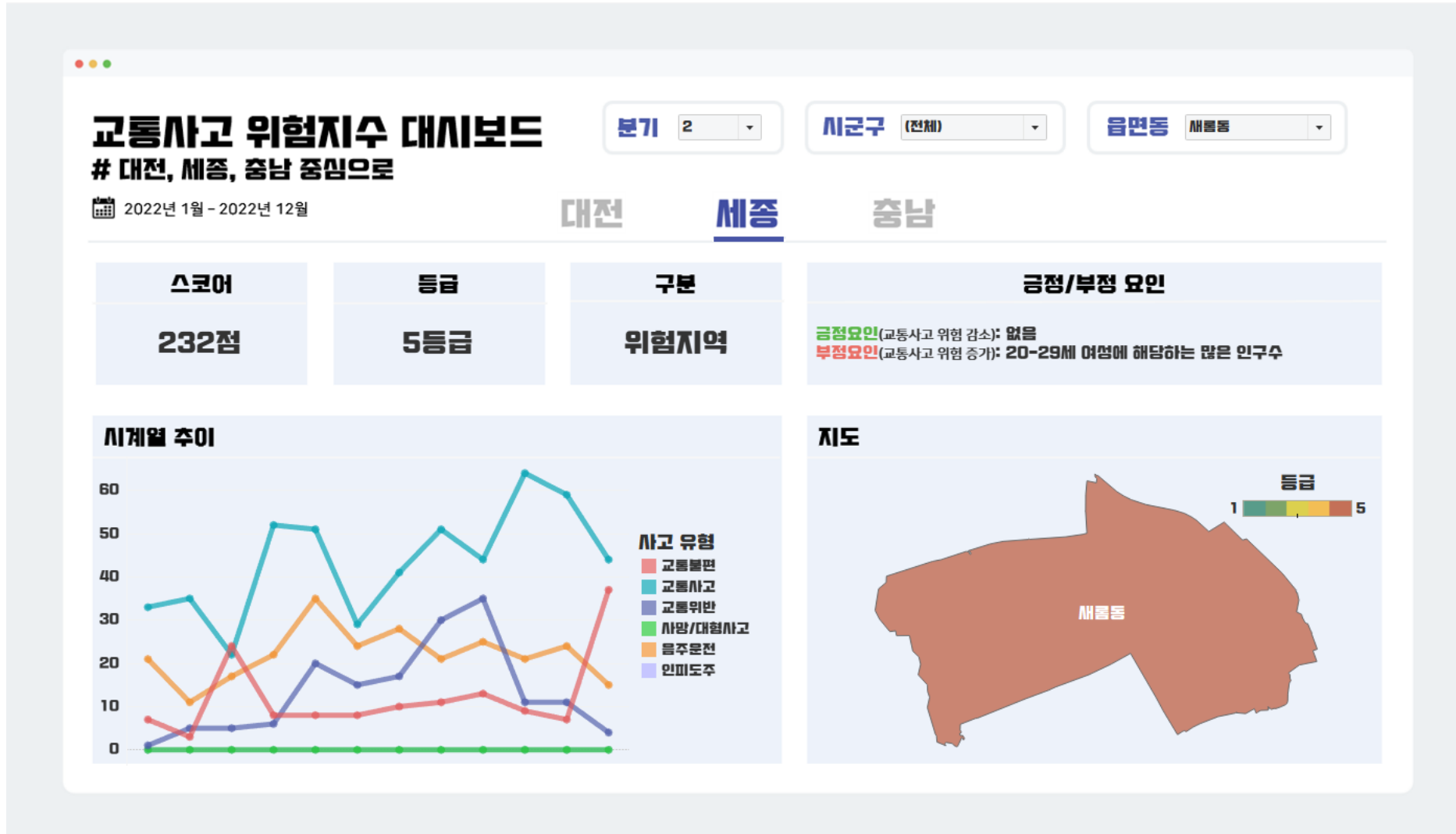
2 세부 수행 내용

③ 대시보드 - 세종특별자치시

“

분석결과 활용가능한 시도별 대시보드 개발

”



2 세부 수행 내용

③ 대시보드 - 충청남도

“

분석결과 활용가능한 시도별 대시보드 개발

”



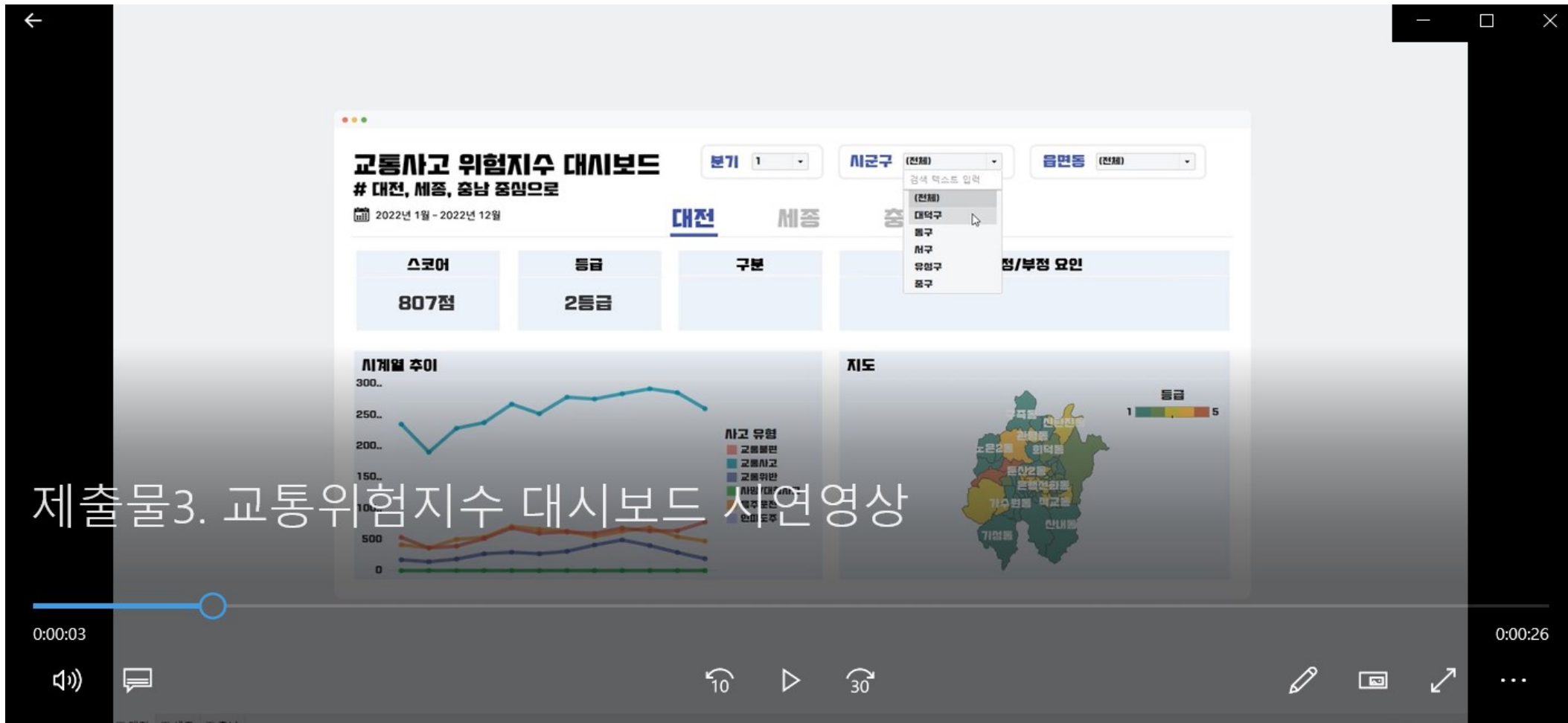
2 세부 수행 내용

③ 대시보드 - 시연영상

“

분석결과 활용가능한 시도별 대시보드 개발

”



※ '제출물3. 교통위험지수 대시보드 시연영상.mp4' 참조요망



경찰대학
KOREAN NATIONAL POLICE UNIVERSITY



대전·세종·충남
지역혁신플랫폼



3

수행 성과

- ① 활용 방안
- ② 기대효과
- ③ 건의 사항

3 수행 성과

① 활용 방안 - 교통 관련 정책 결정

“ 위험 등급별 교통정책 가이드 제시 ”

등급 세분화의 필요성

● 등급별 활동

대전 2등급 사고위험률 2.4%

세종 2등급 사고위험률 9.1%

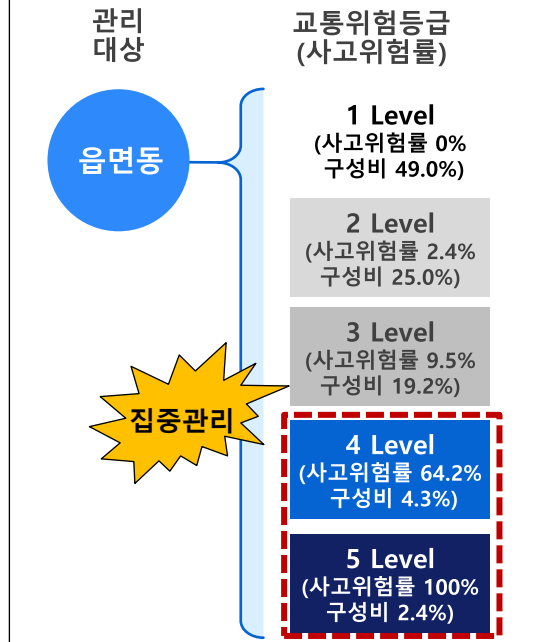
→ 차별화된 등급별 활동 필요

충남 2등급 사고위험률 0%

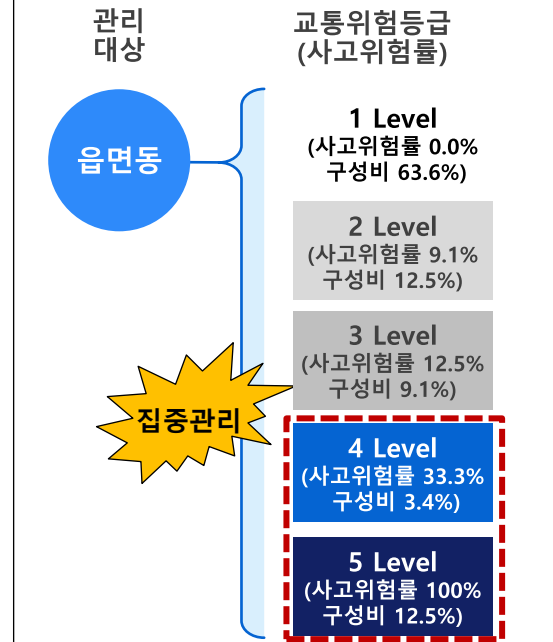
1) 대전, 충남 : 1~2등급 읍면동 현황유지, 지속적 모니터링
3등급 읍면동 교통 시설물 점검
4~5등급의 읍면동 교통 시설물 재정비, 교통 사고 위험 캠페인 실시 등을 통한 집중관리

2) 세종 : 1등급 읍면동 현황유지, 지속적 모니터링
2~3등급 읍면동 교통 시설물 점검.
4~5등급의 읍면동 시설물 재정비, 교통사고 위험 캠페인 실시 등을 통한 집중관리
→ 예) 교통위험등급 1등급(평균 사고위험률 0%)
→ 교통위험등급 5등급(평균 사고위험률 100%)

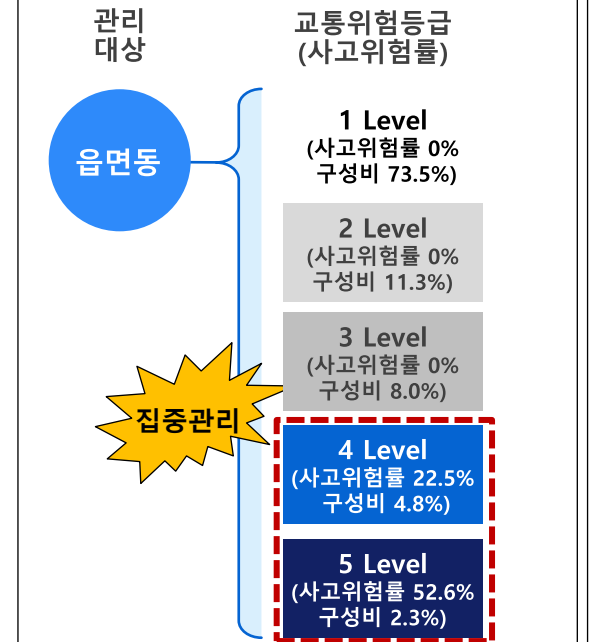
대전 교통 위험 5등급 체계



세종 교통 위험 5등급 체계



충남 교통 위험 5등급 체계



대전 등급별 교통정책 가이드

1 Level	현황 유지
2 Level	모니터링
3 Level	교통 시설물 점검
4 Level	교통사고 위험 캠페인
5 Level	시설물 재정비

세종 등급별 교통정책 가이드

1 Level	현황 유지, 모니터링
2 Level	교통 시설물 점검
3 Level	교통 시설물 점검
4 Level	교통사고 위험 캠페인
5 Level	시설물 재정비

충남 등급별 교통정책 가이드

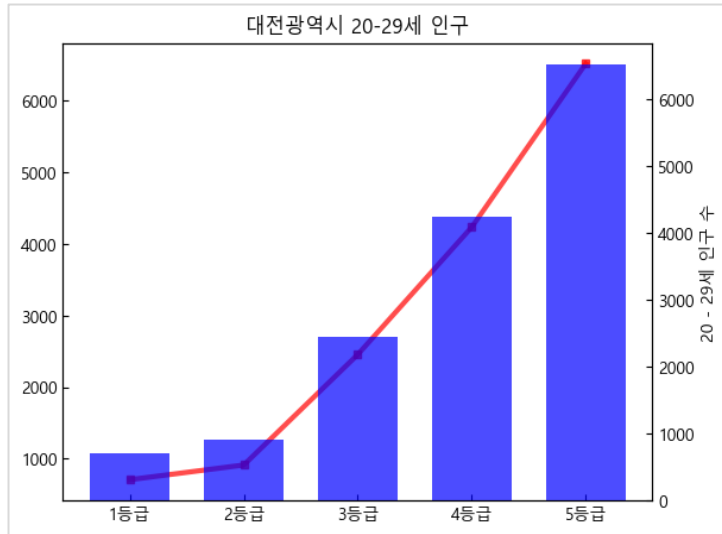
1 Level	현황 유지
2 Level	모니터링
3 Level	교통 시설물 점검
4 Level	교통사고 위험 캠페인
5 Level	시설물 재정비

3 수행 성과

① 활용 방안 - 정책 결정시 고려 항목 도출, 연령 관련 항목(1/2)

➤ Y축 : 분석항목 실제값

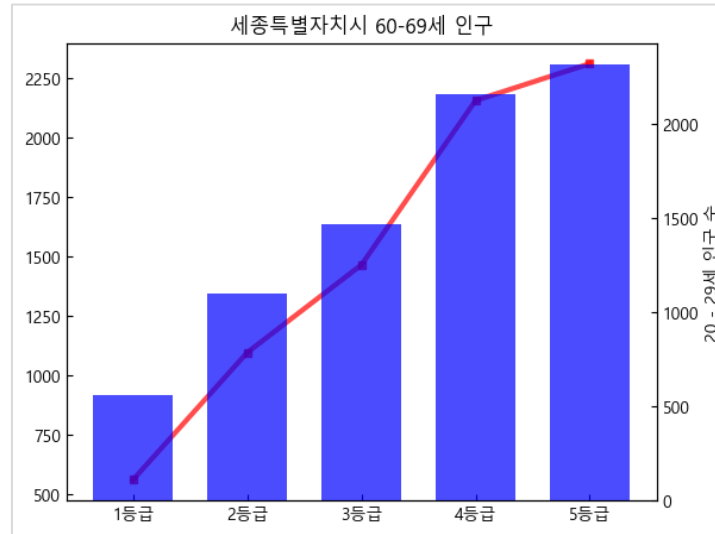
< 대전광역시 >



20-29세 인구가 많을수록
교통사고 위험등급 ↑

➔ 변화가 주변 교통경찰
인력 강화

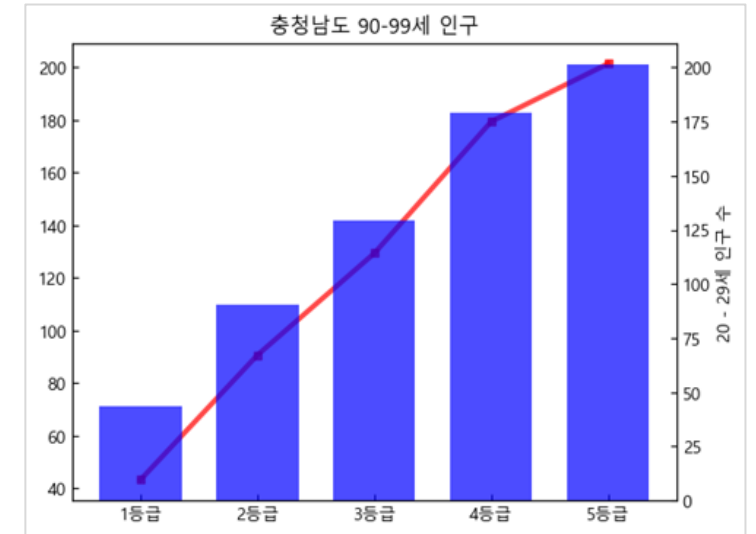
< 세종특별자치시 >



60-69세 인구가 많을수록
교통사고 위험등급 ↑

➔ 노년층을 위한
교통 안전 시설물 증설

< 충청남도 >



90-99세 인구가 많을수록
교통사고 위험등급 ↑

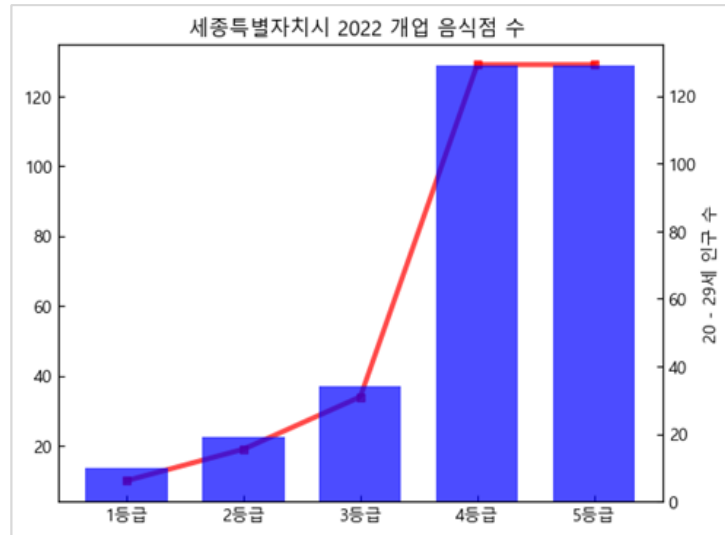
➔ 노년층을 위한
교통 안전 시설물 증설

3 수행 성과

① 활용 방안 - 정책 결정시 고려 항목 도출, 시설물 & 기타 관련 항목 (2/2)

➤ Y축 : 분석항목 실제값

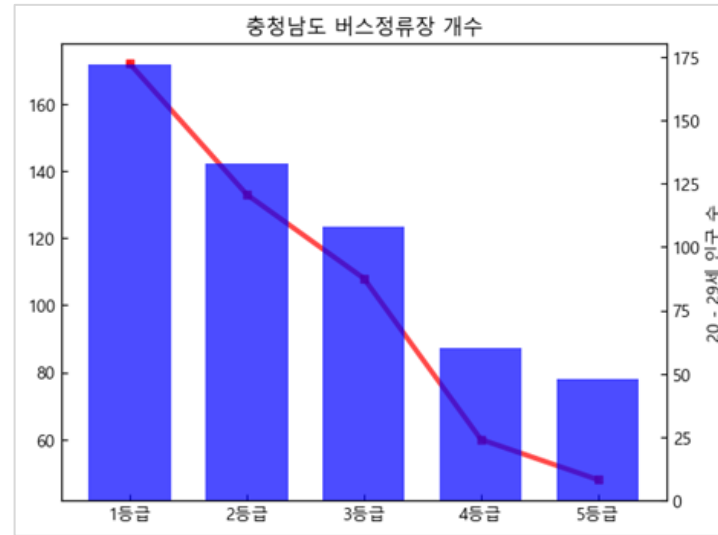
< 세종특별자치시 >



2022년 개업 음식점 수가 많을수록
교통사고 위험등급 ↑

➔ 변화가 주변 교통경찰
인력 강화

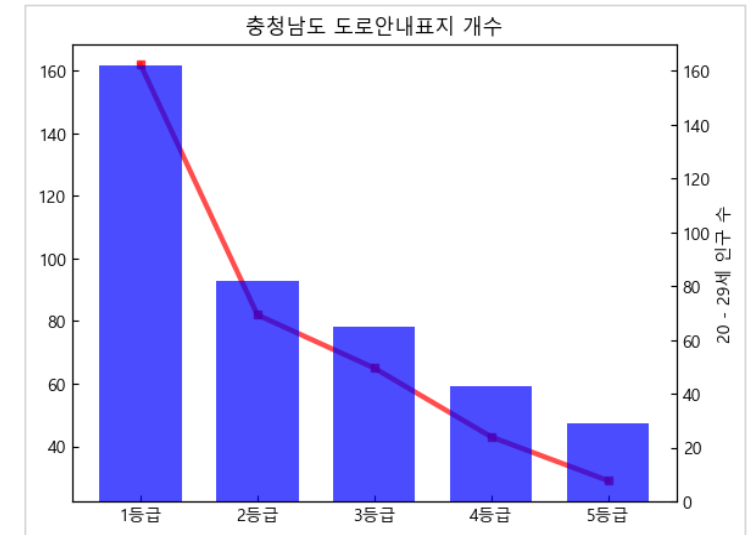
< 충청남도 >



버스 정류장 개수가 많을수록
교통사고 위험등급 ↓

➔ 시설물 재정비

< 충청남도 >



도로 안내 표지 개수가 많을수록
교통사고 위험등급 ↓

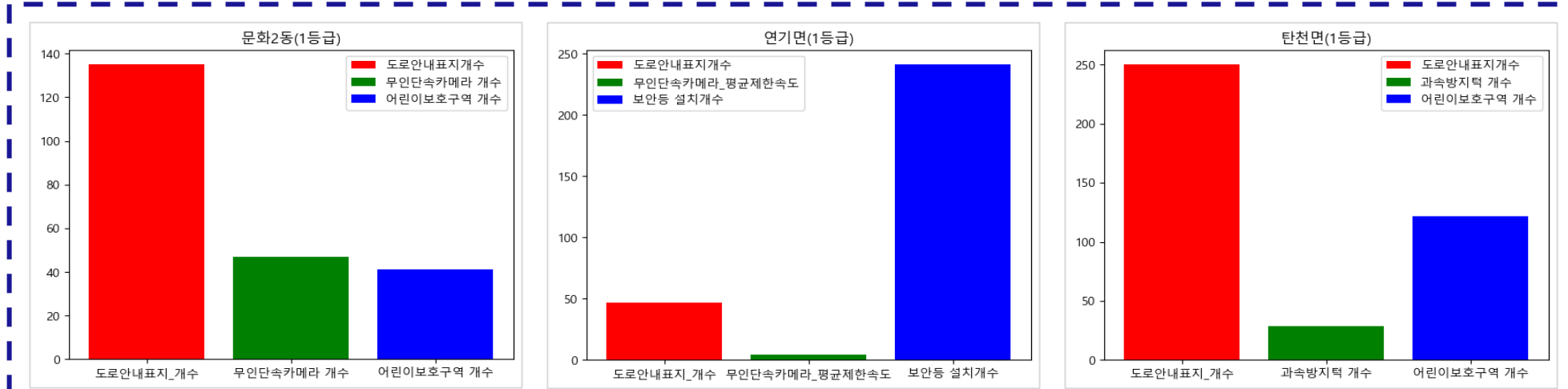
➔ 시설물 재정비

3 수행 성과

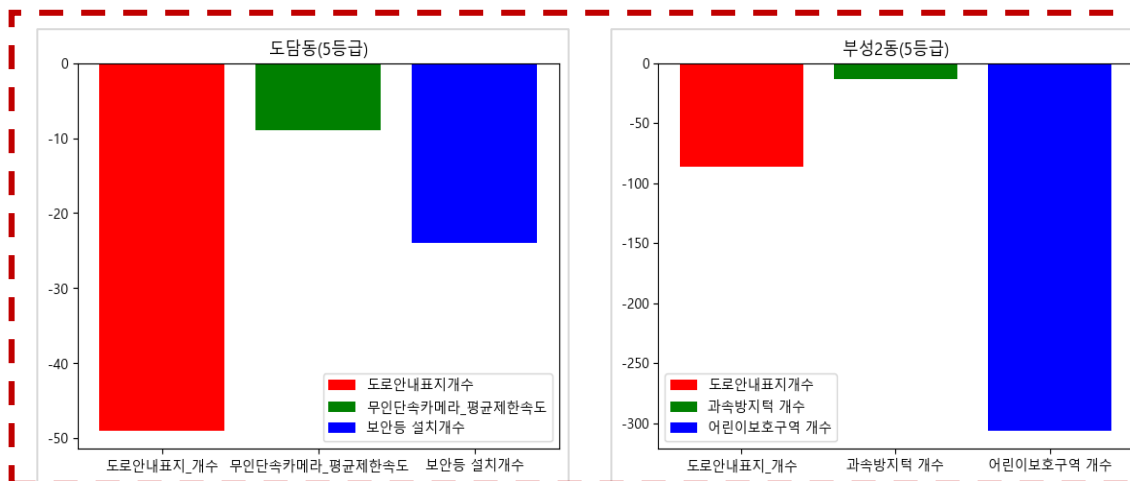
① 활용 방안 - 시설물 재정비 필요성

➤ 교통 안전 관련 시설물 재정비가 필요한 지역들의 스코어 취득 & 등급 현황

1등급:
스코어 양수



5등급:
스코어 음수



관측결과 :

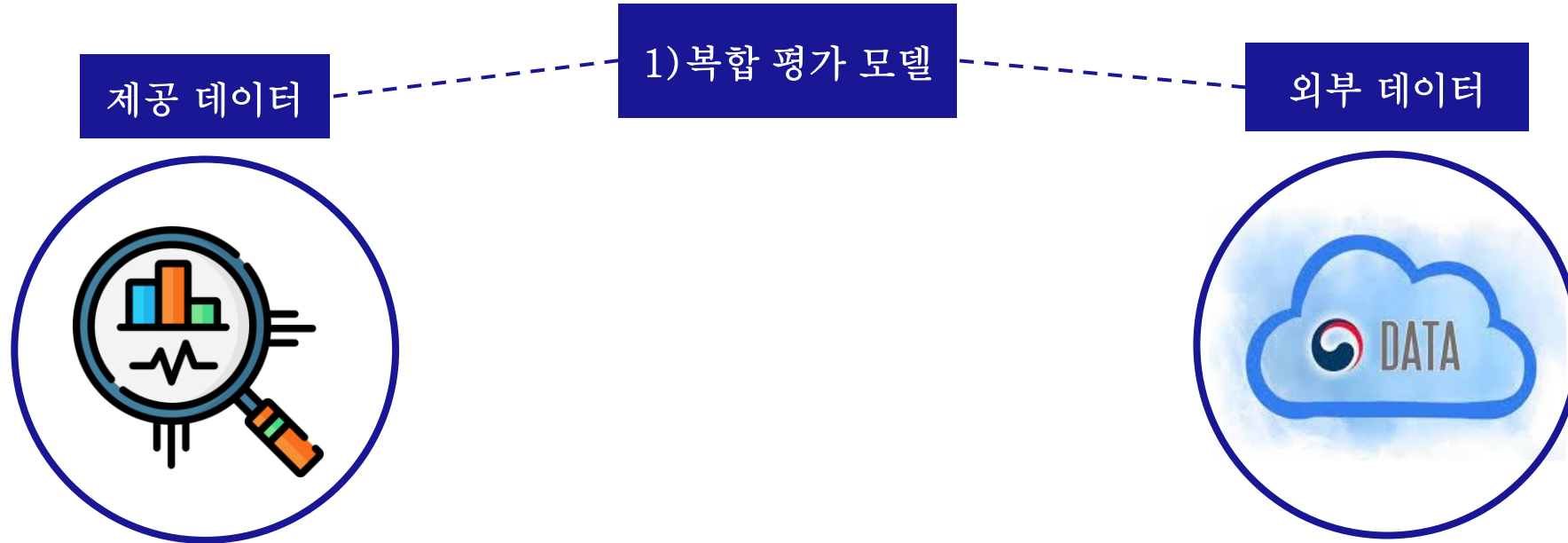
위험등급이 5등급인 읍면동은 1등급 대비 스코어 음수값을 취득함

→ 시설물 재정비를 통해 등급 상승 필요

3 수행 성과

② 기대 효과

□ **교통사고 위험률 복합 평가 모델**을 통한 읍면동별 교통사고 위험지역 확인



- 112 신고 데이터를 활용하여 읍면동 교통사고 위험률 산출
- 외부 데이터 구매 비용 절감

- 인구, 보안등, 신호등 등 공개된 외부 데이터 활용
- 다양한 변수를 고려한 모델 고도화

제공받은 교통사고 데이터와 공공데이터 등 외부데이터의 결합을 통한
복합 교통사고 위험률 평가 모델 개발

3 수행 성과

② 기대 효과

□ 교통사고 위험률 복합 평가 모델을 통한 읍면동별 교통사고 위험지역 확인



- 교통사고 예방 및 관련 정책 제안을 위한 데이터 분석 진행
→ 읍면동별 특성을 반영한 평가모델 구축
- 교통사고 고위험 지역 → 개발한 모델 통해 선별

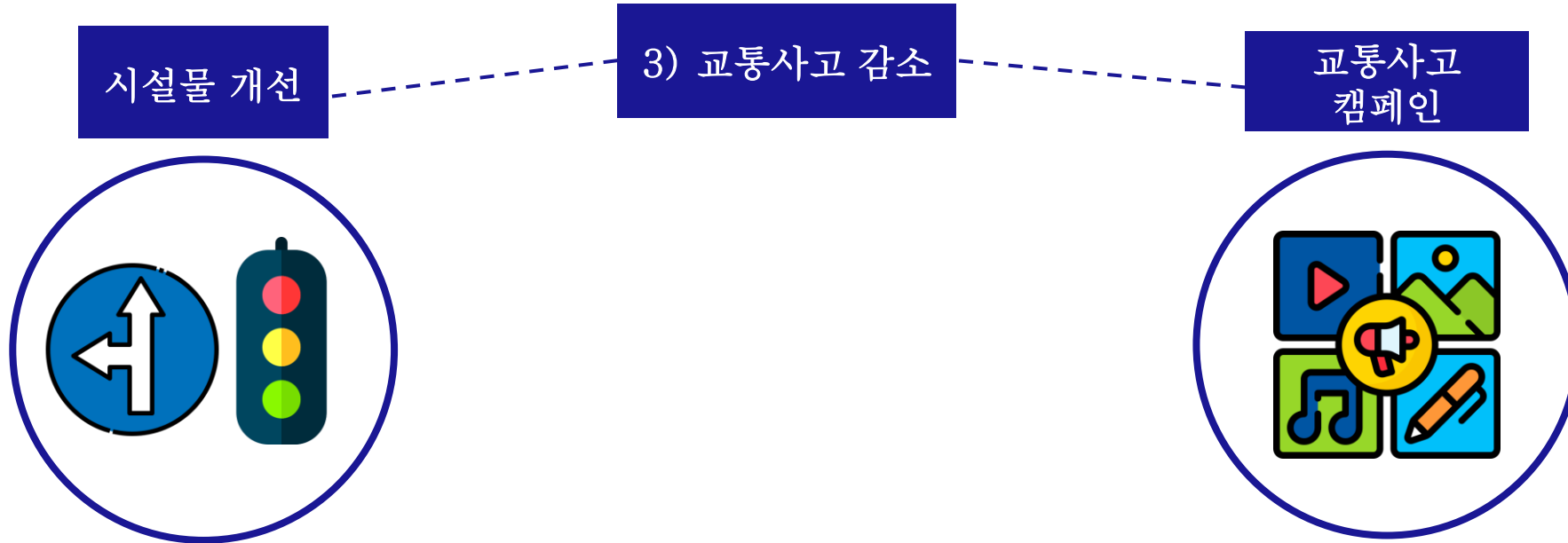
- 스코어카드 대시보드 개발(시각화)
 - a. 분기별, 읍면동별 교통사고 위험등급 조회
 - b. 교통 사건 유형별 시계열 추세 그래프 조회
 - c. 교통사고 위험률 긍정, 부정 요인 조회

교통사고 위험 지역 선별 모델 개발과 대시보드 개발을 통한 읍면동별 교통사고 위험지역 확인

3 수행 성과

② 기대 효과

□ 교통사고 감소 효과



- 교통사고 고위험 읍면동 교통 시설물 개선
-> 교통사고 감소 효과

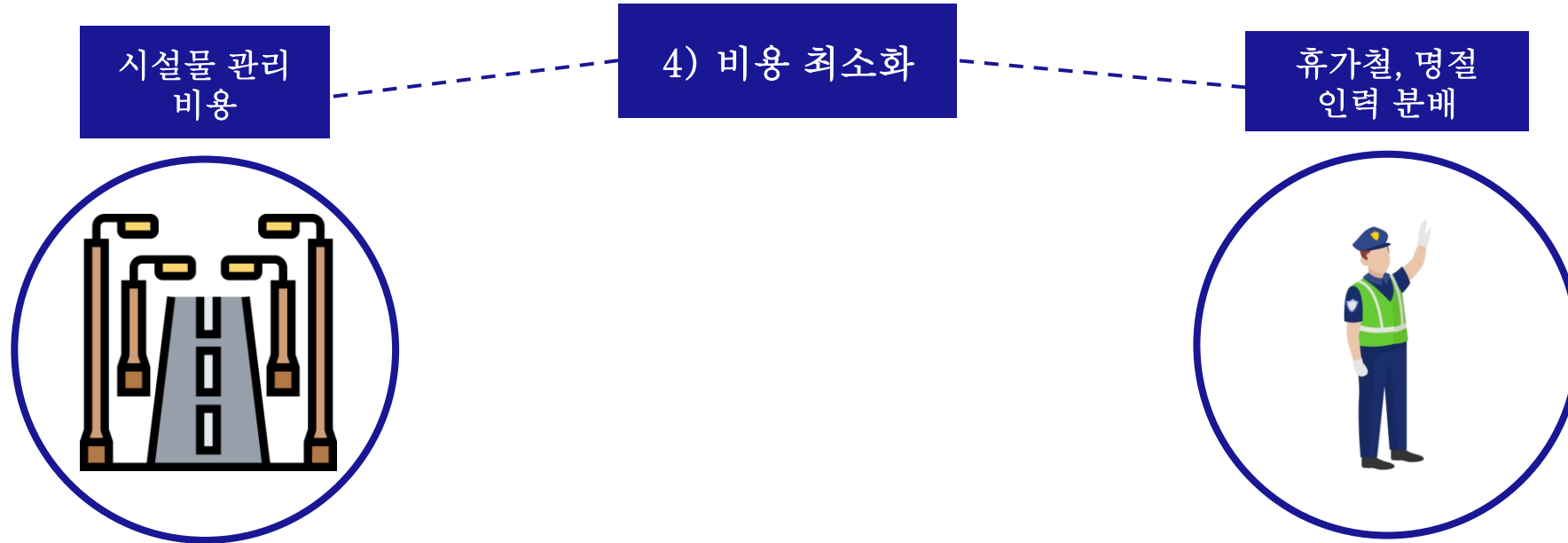
- 고위험 읍면동 대상 교통사고 캠페인 진행
-> 교통사고 감소 효과

교통사고 고위험 읍면동 시설물 개선, 교통사고 캠페인을 통한 교통사고 감소 효과

3 수행 성과

② 기대 효과

□ 사회적 비용 최소화



- 교통사고 고위험 읍면동 대상 집중 시설물 점검 및 관리
-> 합리적인 예산 집행, 관리비용 최소화

- 휴가철, 명절 등 유입인구를 고려한 인력 배치
-> 인력 분산으로 인한 자원 낭비 최소화

교통사고 고위험 읍면동 시설물 관리비용 최소화, 휴가철 명절 인력 낭비 최소화

3 수행 성과

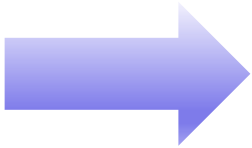
③ 건의 사항

□ 모형 고도화 & 기타사항 개선을 위한 추가 제언

순번	구분	항목	개선방안
1	데이터 확보	시간대별 읍면동 유동인구 데이터 필요	통신사와 제휴하여 데이터 확보
2		시간대별 신고장소 날씨 데이터 필요	추가 수집
3		시간대별 신고장소 작업현황 데이터 필요 (ex) 도로 청소차 작업중, 제초 작업중 등)	추가 수집
4		충청남도 통합 데이터 필요	충청남도 시군구별 데이터 통합, 확보 가능(데이터가 없는 시군구는 분석 불가능)
5		공공데이터 전국 표준 데이터에 누락 건 존재	개별 지역 데이터를 합쳐 데이터 확보 가능
6		데이터 건수 미비(ex : 세종특별시 과속방지턱 조화 12건)	전국 표준 데이터 업데이트 -> 데이터 확보 가능
7		읍면동 단위로 데이터가 존재하지 않음(시군구 단위로 존재)	위도, 경도 좌표를 이용한 'Reverse Geocoding' 을 통해 데이터 확보 가능
8	데이터 품질개선	112 신고 데이터와 공공데이터 매핑 어려움	통합 데이터 제공(=112 신고 데이터 + 기타 공공데이터) (벤더사 : 대전,세종,충남 지역혁신플랫폼)
9		112 신고 데이터 품질개선 : HPPN_PNU_ADDR(발생지점) - 대부분 법정동 값, 그러나 '신탄진휴게소, 충남대학교' 등 법정동 외 값 다수 존재	형식 통일하여 품질개선
10		교통 관련 데이터 최신화 되어있지 않음	데이터 업로드 관련 담당자 & 정책 확정
11	연구	교통 관련 신고유형별 심각정도 파악 필요	산정방식 추가 연구 진행

“ ”

데이터 확보 및 품질개선을 통해 모형 고도화 및 추가 분석 가능



참고 문헌

- S. Xie. scorecard: Credit Risk Scorecard, 2020. URL <https://CRAN.R-project.org/package=scorecard>. R package version 0.3.0. [p1]
- Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke Creation of Synthetic Data in R. *Journal of Statistical Software*, 74(11), 1–26. <https://doi.org/10.18637/jss.v074.i11>
- Elbow Point Detection Library, <https://kneed.readthedocs.io/en/stable/>
- synthpop: Generating Synthetic Versions of Sensitive Microdata for Statistical Disclosure Control, <https://cran.r-project.org/web/packages/synthpop/index.html>
- 윤현석. (2021). 교통사고 발생 요인 분석을 통한 교통사고 예방 및 감소 방안: 광주광역시 2020 교통사고분석시스템 자료를 중심으로. *인문사회* 21, 12(6), 2731–2744.
- 서지민 and 이수기. (2016). 서울시 보행자 교통사고에 영향을 미치는 물리적 환경요인에 관한 연구 : 2014 TAAS 자료를 중심으로. *국토계획*, 51(3), 197–216.
- 윤운진(2017). 교통 빅데이터를 활용한 교통사고 다발지역 선정 및 분류 기술 개발 연구. 과학기술정보통신부 Ministry of Science and ICT
- 이정원 and 이충호. (2019). 공공데이터를 활용한 노인교통사고 발생유형 분석연구. *산업융합연구*, 17(4), 53–58.
- 박준태, 홍지연 and 이수범. (2010). 기상특성에 따른 교통사고 안전성 평가지표 개발 (고속도로를 대상으로). *대한교통학회지*, 28(1), 157–163.
- 구영현.(2019). 개인신용정보 재현데이터 생성. 국가통계방법론심포지엄
- 강갑생, ‘GDP 3%’ 40조 길바닥서 날린다…한국인 삶 옥죄는 ‘이것’ , 뉴스원샷, 2021.05.08, <https://www.joongang.co.kr/article/24052724#home>
- 김남주, 코로나19 시국에도 교통사고로 인한 사회적 비용은 오히려 늘어, 교통일보, 2022.06.23, <http://www.tdaily.co.kr/news/view.php?idx=39972>
- 윤해리, 사회적 거리두기 해제 뒤 보행자 교통사고 27% 증가, YTN, 2022.05.12, https://www.ytn.co.kr/_ln/0103_202205121648435457
- 김기룡, 충남지역 교통시설물 개선 지점...교통사고 15~42% ↓, 신아일보, 2023.02.07, <http://www.shinailbo.co.kr/news/articleView.html?idxno=1659071>
- 이창우, 전남 교통사고 사망자 200명 '역대 최저'...감소 인원 전국 1위, 뉴시스, 2023.01.29, https://newsis.com/view/?id=NISX20230128_0002172732&cID=10201&pID=10200
- 김잔디, "어르신 운전면허 강화했더니 교통사고↓", 연합뉴스, 2023.01.29, <https://www.yna.co.kr/view/AKR20230128036700004?input=1195m>
- 이소현, 설 연휴 교통사고 56.7% 감소..."평온한 명절 유지", 이데일리, 2023.01.24, <https://www.edaily.co.kr/news/read?newsId=01751526635481000&mediaCodeNo=257&OutLnkChk=Y>



감사합니다