

# 제주도 교통사고 발생 예측을 통한 교통 안전 취약 초등학교 검토

범주형자료분석팀



## 목 차

---

1. 분석 배경
2. 데이터 수집 및 전처리
3. 사고 발생 건수 예측 모델링
4. 클러스터링
5. 설문조사 분석
6. 해결방안 제시 및 현행 정책 검토

01

분석 배경

# 01. 분석배경

## | 제주 교통사고 실태

“제주-서귀포 산간도로 내리막 구간  
사고 반복에도 대책은 깜깜”

제주지역 산간도로 내리막 구간에서  
대형 교통사고가 매년 끊이지 않고  
있지만 이에 대한 재발 방지 대책은  
사실상 전무해 사고를 키우고 있다는  
지적이 나오고 있다 ...

한라일보

제주 '관광객·대형차' 교통사고 빈발... 대책은?

제주일보

제주서 해마다 교통사고 4000건 이상 발생

뉴스제주

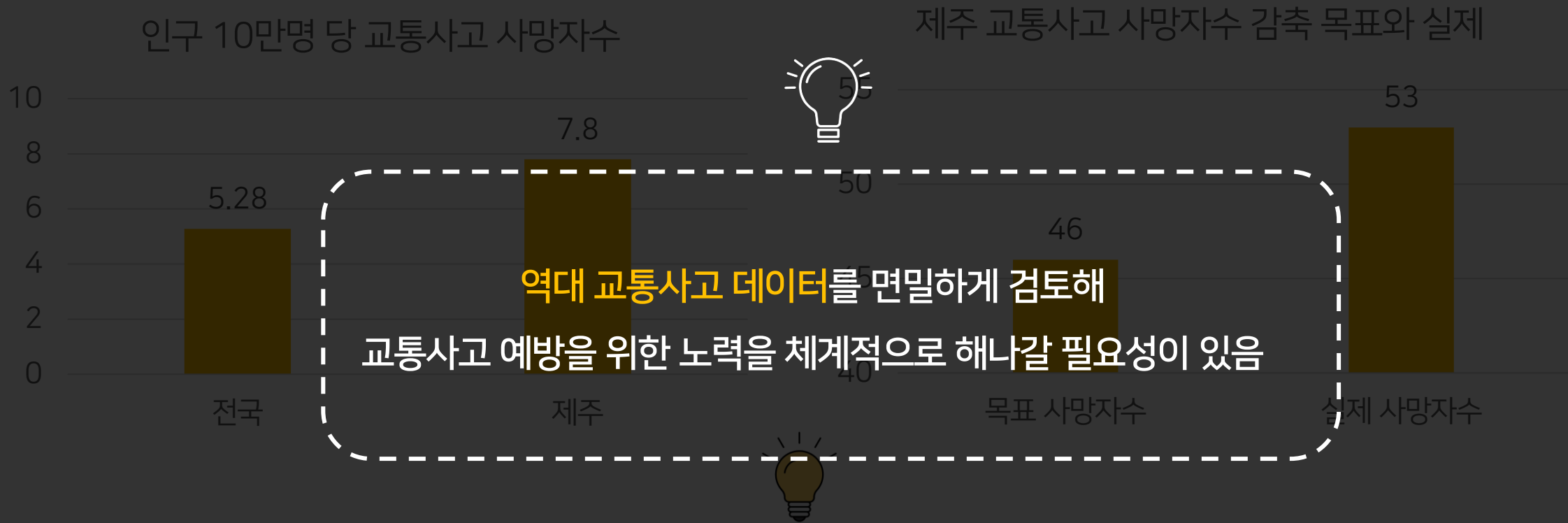
제주 애월 교차로서 3중 추돌 교통사고, 4명 사상



렌터카 이용 건수와 산간도로가 많은 특성으로 인해  
제주도에서는 교통사고가 매해 끊이지 않고 있음

# 01. 분석배경

## | 제주 교통사고 실태



인구 10만명 당 교통사고 사망자수는 제주가 전국 평균보다 높음  
제주 교통사고 사망자수 감축 목표 달성도는 86.8%에 그침

02

## 데이터 수집 및 전처리

## 02. 데이터 수집 및 전처리

### 데이터셋 소개

#### 인구

파일이름	출처	전처리 결과
제주특별자치도_읍면동 단위 내국인 유동인구	공공데이터포털	유동인구

#### 기상 상황

파일이름	출처	전처리 결과
rn_230427	기상청 기상자료개방포털	강수량
STCS_강수일수_MNH	기상청 기상자료개방포털	강수일수
STCS_안개일수_MNH	기상청 기상자료개방포털	안개일수
STCS_눈일수_MNH	기상청 기상자료개방포털	강설일수

#### 도로환경

파일이름	출처	전처리결과
제주특별자치도_교통신호기운영현황	공공데이터포털	교통신호기
지방도 교통량 조사 결과	제주시, 서귀포시	가중교통량
제주시/서귀포시 교 통사고	TAAS	사고발생건수

## 02. 데이터 수집 및 전처리

### | 설명변수 소개 - 기준

제주특별자치도에서 사고가 발생한 43개 행정동 별로 분리 (읍/면/동)



- ✓ 43개 행정동 중 추자도 위치가 제주와 동떨어져 있음을 파악 후 임의로 제거 → 42개 행정동
- ✓ 원래 행정동보다 세부적인 도로구간 별로 고려하려고 했으나 도로구간 별 데이터가 부재해 대다수의 데이터셋이 행정동별로 수집됨
- ✓ 불가피하게 일차적으로 행정동 단위로 사고다발구역을 선정



## 02. 데이터 수집 및 전처리

### | 최종 train dataset overview

사고 발생건수	행정동	년	월	강수일수	안개일수	눈일수	신호기	유동인구	교통량
3	건입동	2018	1	18	0	12	15	341251.68	178414
6	건입동	2018	2	11	0	8	15	342506.89	178414
7	건입동	2018	3	10	2	0	15	327685.1	178414
...			...		...	...	...	...	...
23	애월읍	2019	11	7	0	0	82	1297213.77	284561
13	애월읍	2019	12	7	0	1	82	1287008.45	284561
21	애월읍	2020	1	10	0	0	82	1303109.06	251408
13	애월읍	2020	2	10	0	4	82	1243326.41	251408
...			...		...	...	...	...	...

# 03

## 사고 발생 건수 예측 모델링

## 04. 사고 발생 건수 예측 모델링

### | 패널 회귀 모형

패널 데이터 (Panel Data)

시계열 자료와 횡단면 자료가 합쳐져 있는 데이터를 의미

n개의 entities로만 구성되어 있는 Cross-section 데이터와 달리

n개의 entities와 시간 T로 구성

시계열 데이터

Entity 1

$t = 1, 2, \dots, T$

횡단면 데이터

Entity 2

$t = 1, 2, \dots, T$

횡단면 데이터

Entity 3

$t = 1, 2, \dots, T$

횡단면 데이터

.....

Entity n

$t = 1, 2, \dots, T$

횡단면 데이터

# 01. Panel Regression

## 년도별 전처리

원데이터의 거주, 근무, 방문 유동인구를 모두 합친 값



사고 발생건수	행정동	년	유동인구
76	건입동	2018	13562.03
73	건입동	2019	13109.78
67	건입동	2020	13191.67
...	...	...	...



사고 발생건수	행정동	년	거주 유동인구	근무 유동인구	방문 유동인구
76	건입동	2018	6222.02	878.68	6461.32
73	건입동	2019	6006.67	811.54	6291.56
67	건입동	2020	6385.87	805.08	13191.67
...	...	...	...	...	...

유동인구 변수를 거주, 근무, 방문 3가지 유형으로 분류하여 사용

# 01. Panel Regression

## 파생변수 생성

1보다 크면 방문 유동인구가 많고

1보다 작으면 거주 + 근무 유동인구가 많다고 해석

행정동	년	방문 인구 비율	유동 인구 성격	연장 당 교통량	흐린날 대비 교통량	운전 가능 인구수
건입동	2018	0.4764	0.9099	6395	625	7668
건입동	2019	0.4799	0.9227	11524	1551	7434
건입동	2020	0.4548	0.8344	10172	1150	7250
...	...	...	...	...	...	...

$$✓ \text{방문인구비율} = \frac{\text{방문 유동인구}}{\text{전체 유동인구}}$$

$$✓ \text{유동인구성격} = \frac{\text{방문 유동인구}}{\text{거주 유동인구} + \text{근무 유동인구}}$$

$$✓ \text{연장 당 교통량} = \frac{\text{교통량}}{\text{연장길이}}$$

$$✓ \text{흐린날 대비 교통량} = \frac{\text{교통량}}{\text{강수일수} + \text{강설일수} + \text{안개일수}}$$

$$✓ \text{운전가능인구수} = 20\text{대} \sim 70\text{대 주민등록인구수}$$

# 01. Panel Regression

## | panel regression 모델링 결과

변수명	Estimate	P-value
Intercept	-21.2935	0.5687
강수일수	0.1370	0.6838
강설일수	-0.1361	0.6710
...	...	...
교통량	57.5095	0.0019
연장길이	45.0273	0.0267
근무 유동인구	110.3706	0.0001
방문인구비율	51.9295	0.0011
연장 당 교통량	-0.0184	0.0001
운전가능인구수	213.6566	<2e-12

min-max scaling 진행 후 모델링

$$Y_{it} = \alpha_i + \beta X_{it} + u_{it}$$



기후상태와 관련된 변수들은 예측에 유의하지 않음

교통량이 증가하면 사고가 날 확률이 57배 증가

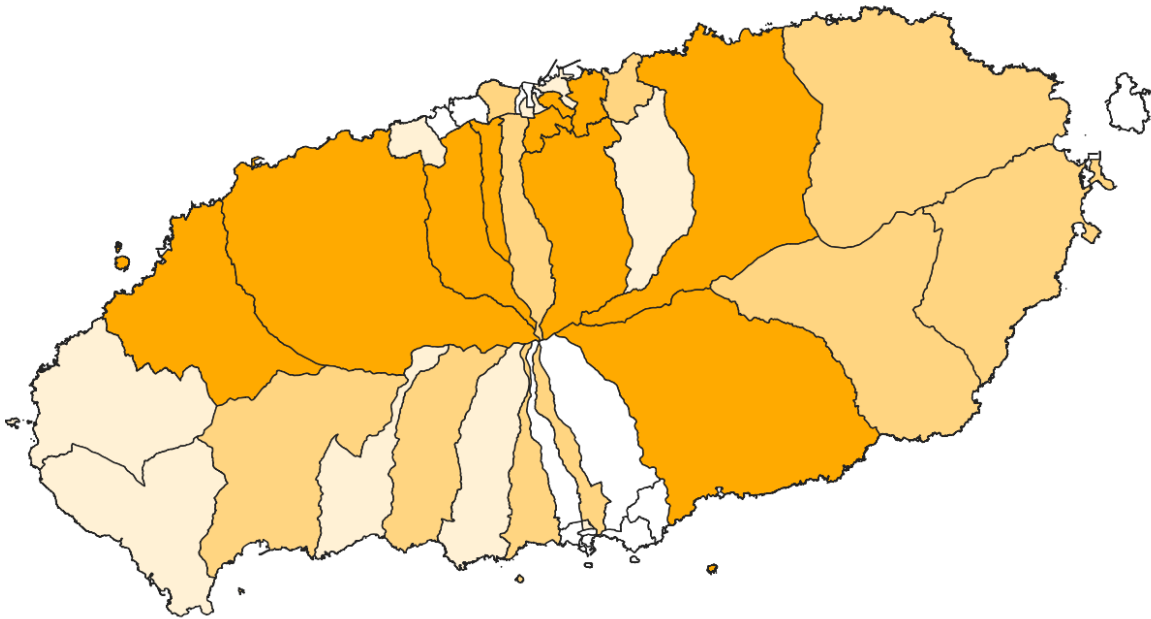
방문인구비율이 증가하면 사고가 날 확률이 51배 증가

연장 당 교통량이 증가하면 사고가 날 확률이 0.01배 감소

# 01. Panel Regression

| panel regression 모델링 결과

Panel Regression 예측 결과



행정동	예측값	실제값
건입동	68.92	82
구좌읍	89.94	102
남원읍	128.32	129
노형동	357.77	356
대륜동	105.67	101
...	...	...
한경면	55.86	61
효돈동	23.39	37

사고 발생 건수가 높은 상위 10개 행정동 중 8개 행정동을 예측!

실제 상위 10위 행정동에 드는 용담2동, 안덕면 대신 남원읍, 화북동으로 예측

# 01. Panel Regression

| panel regression 모델링 결과

## Panel Regression 예측 결과



Panel Regression 예측 결과를 이용해  
사고 발생 예측 건수가 가장 높은 상위 10개 행정동을 선정

노형동 이도2동 연동 애월읍 아라동  
조천읍 일도2동 한림읍 남원읍 화북동

행정동	예측값	실제값
건입동	68.92	82
고잔읍	89.94	102
남원읍	128.32	129
노형동	357.77	356
대륜동	105.67	101
...	...	...
안성면	55.86	이제 절반 왔다..
호도동	23.29	37

사고 발생 건수가 높은 상위 10개 행정동 중 8개 행정동을 예측!  
실제 상위 10위 행정동에 드는 용담2동, 안덕면 대신 남원읍, 화북동으로 예측





04

클러스터링

## 02. 클러스터링

### | 클러스터링 대상 선정

모델링을 이용해 예측한 사고다발 행정동 TOP10에 위치한 초등학교들을 클러스터 대상으로 선정



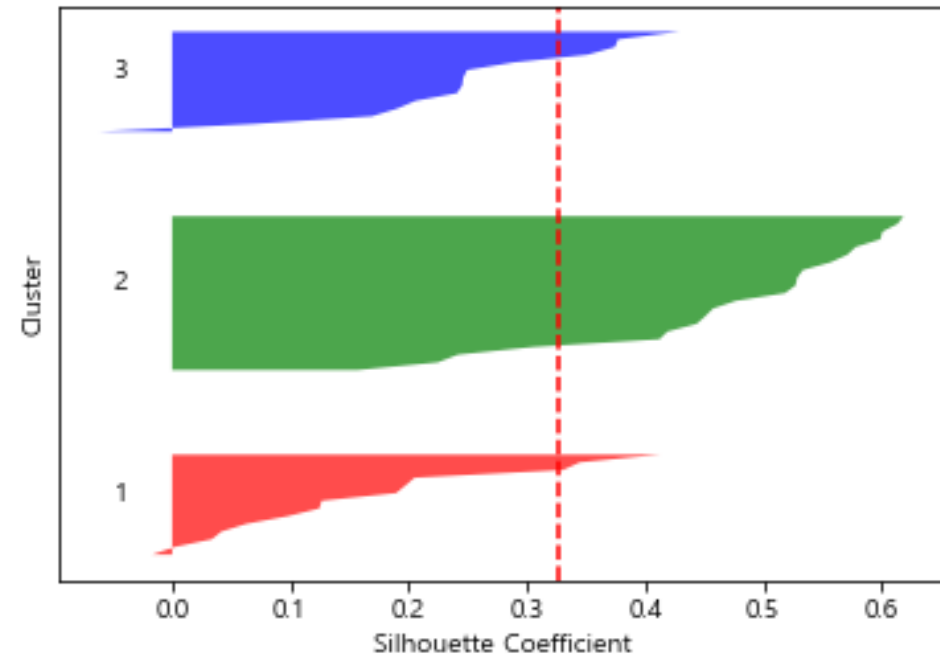
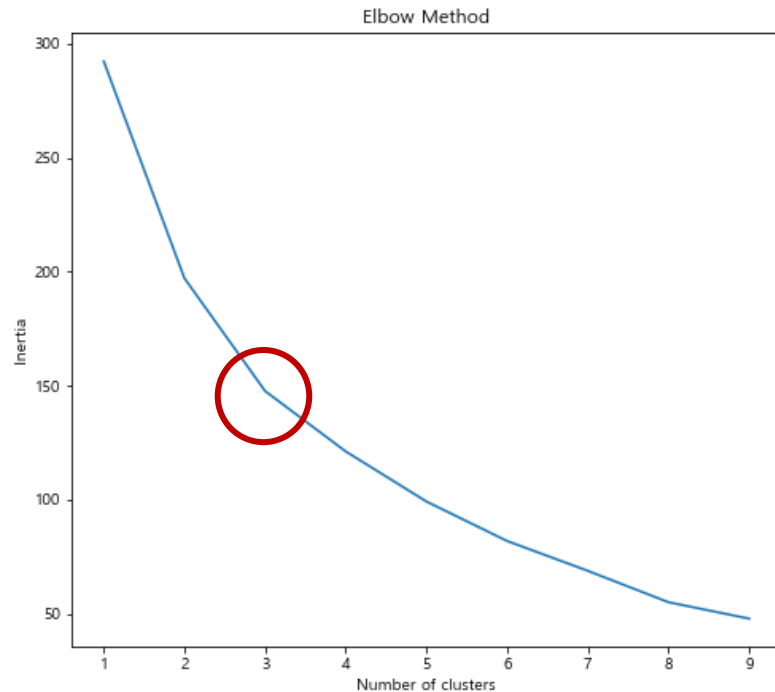
애월읍 노형동 연동 아라동 조천읍 한림읍 이도2동 일도2동 화북동 남원읍  
총 49개 초등학교



초등학교 별로 교통안전을 파악할 수 있는 데이터를 수집해  
클러스터링을 통해 각 초등학교의 교통안전 유형 파악

## 02. 클러스터링

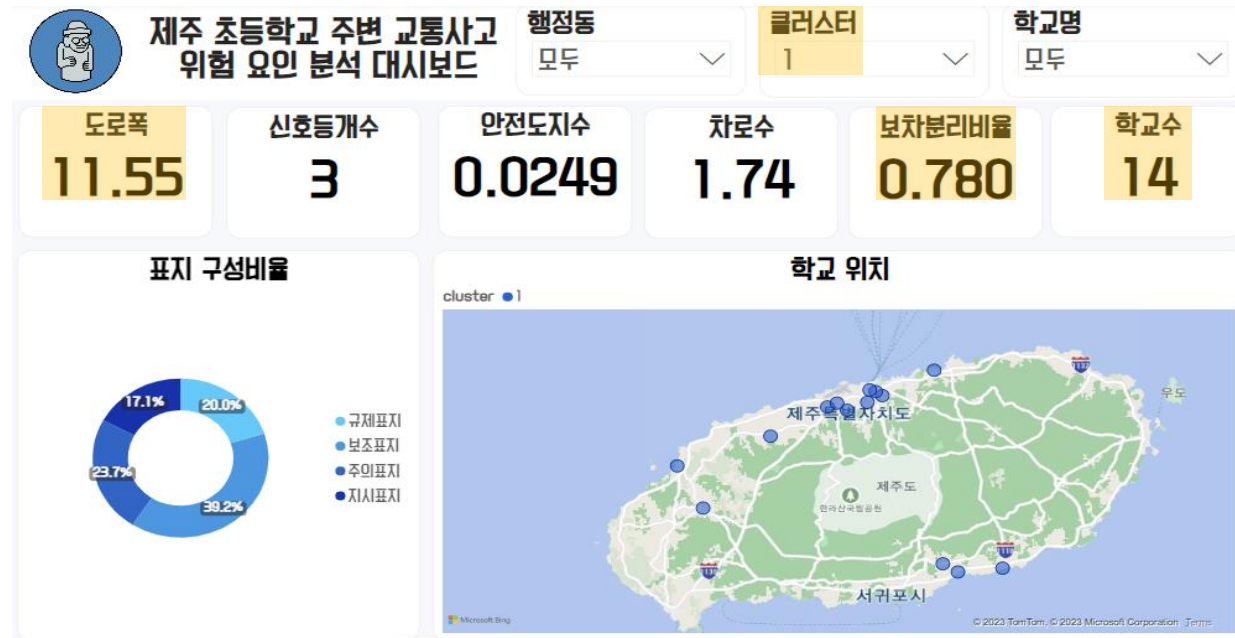
### | K-means 클러스터링



K-means 결과, **최적의 클러스터 개수가 3**으로 선정  
클러스터 개수가 3일 때 **실루엣 계수가 0.4726**으로 도출

## 02. 클러스터링

### 클러스터 1 분석



노형동, 연동 등 제주 **구도심 지역**에 속하는 행정동이 다수 존재  
클러스터 중 평균 **도로폭**이 **가장 좁음**  
**보차분리비율**이 상대적으로 **낮음**

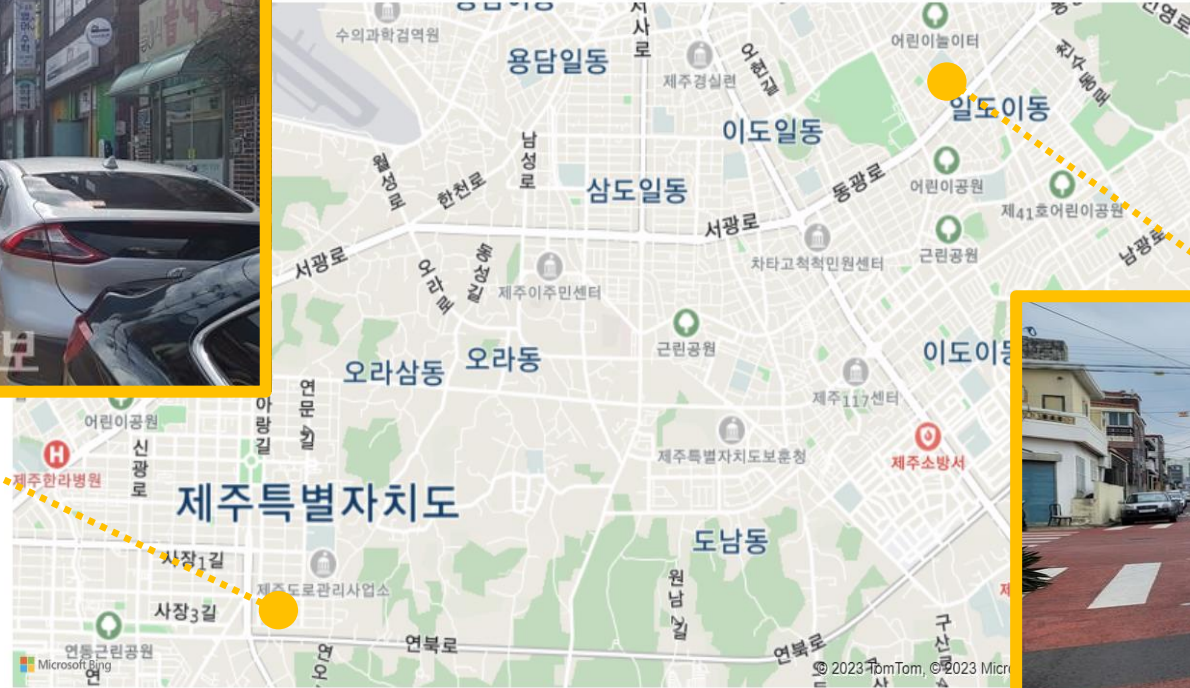
## 02. 클러스터링

### 클러스터 1 분석

신제주초등학교 주변



인화초등학교 주변



도로폭이 좁고 보차분리가 잘 되어있지 않은 것을 확인할 수 있음

교통 인프라가 부족한 것으로 판단



## 02. 클러스터링

### 클러스터2 분석



아라동, 이도2동 등 제주 **신도심 지역**에 속하는 행정동이 다수 존재  
클러스터 중 평균 **도로폭**이 **가장 넓은**



## 02. 클러스터링

### 클러스터2 분석

아라초등학교 주변



이도초등학교 주변



도로폭이 넓고 보차분리가 잘 되어있는 것을 확인할 수 있음

교통 인프라가 우수한 것으로 판단

## 02. 클러스터링

### 클러스터 3 분석



애월읍, 조천읍 등 **교외지역**에 속하는 행정동이 다수 존재  
학생 수와 교통사고 건수가 도심지역에 비해 상대적으로 낮지만  
**학생 수 대비 교통사고 건수가 높음**



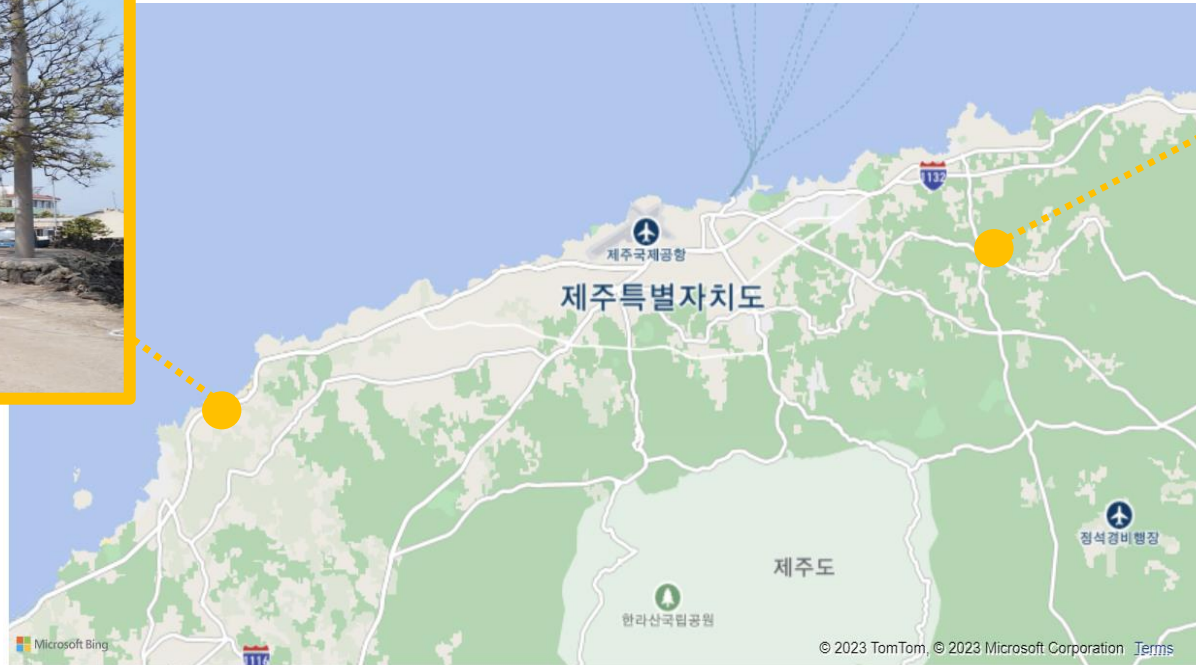
## 02. 클러스터링

### 클러스터 3 분석

귀덕초등학교 주변



대흘초등학교 주변



전반적인 **교통 인프라**는 **평균적**임

**안전도** 지수가 가장 **우수**함

05

설문조사 분석

### 03. 설문조사 분석

#### | 설문조사 분석 목적

제주도 전체 초등학생들을 대상으로 진행한 **설문조사** 데이터를 활용해  
**초등학교 교통안전**에 대한 **학생들의 인식**을 **확인**



**요인 분석**을 통해 도출된 요인들이 각각 어떤 정보를 제공하는지 분석  
각 요인별로 3개의 클러스터 간의 **요인 점수**를 **비교**해  
해결방안을 도출하기 위한 정보로 활용



## 03. 설문조사 분석

### | 요인분석

#### 요인분석

여러 변수들 간의 공분산과 상관관계 등을 이용하여 변수들 간의 **상호관계**를 분석하고  
결과를 토대로 문항과 변수들 간의 **상관성**을 파악해  
여러 변수들이 지닌 정보를 **공통된 요인**으로 묶어서 나타내는 방법

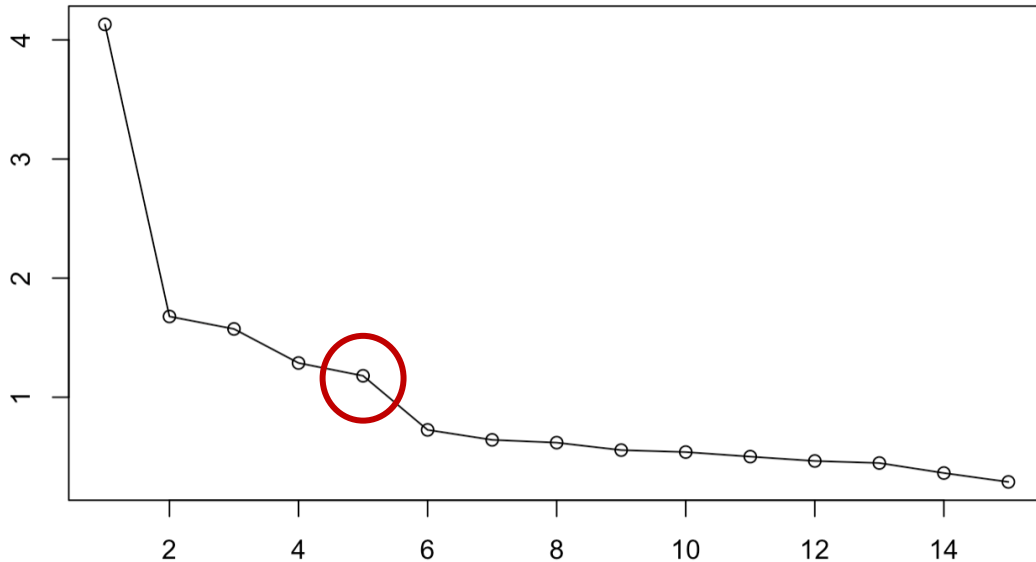


만약 1번 ~ 10번 질문을 요인분석을 진행해  
1번, 2번, 3번이 **하나의 요인**으로 묶인다면  
3개의 질문이 **높은 상관성**을 갖고 있다고 판단

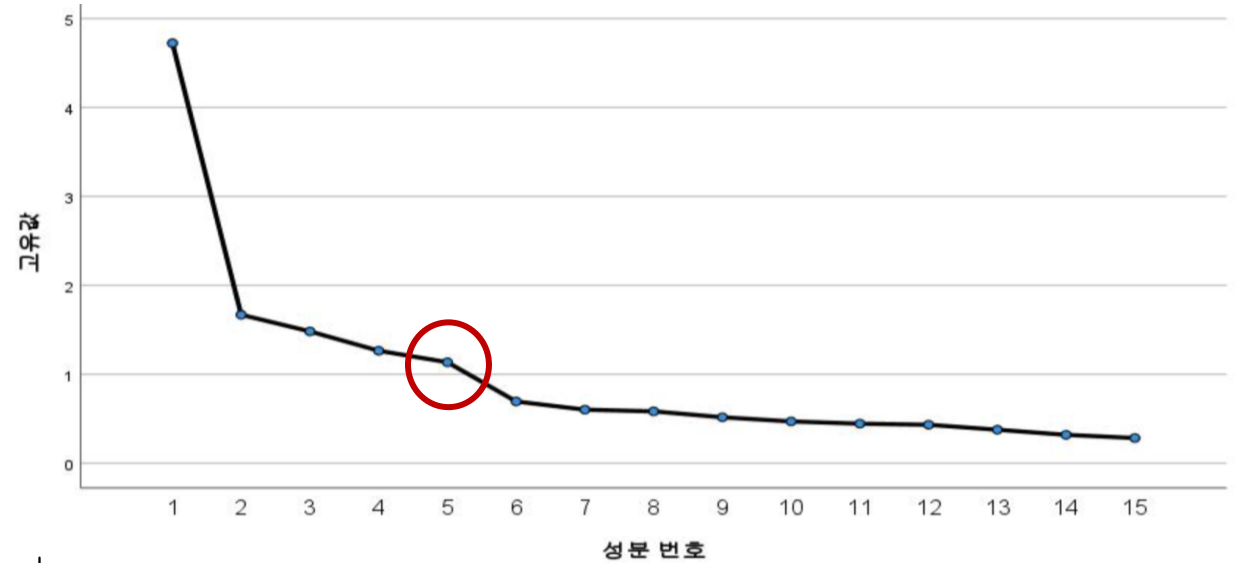
### 03. 설문조사 분석

#### | 요인분석 결과

R Scree Plot



SPSS Scree Plot



일반적으로 고유값이 1보다 크면 하나의 요인이 변수 1개 이상의 분산을 설명함을 의미  
Scree plot 결과를 통해 총 5개의 요인으로 15개의 항목을 요약 가능하다고 판단

### 03. 설문조사 분석

#### | 요인분석 결과

도출된 5개의 요인에 포함된 각 질문들의 **공통점**을 파악해 키워드 부여

#### 안전에 관한 전반적인 인식

##### 요인 1

집에서 학교로 갈 때 교통사고나 안전사고 위험은  
어느 정도라고 생각하나요?

표시한 길에서 교통사고와 안전사고 위험은  
어느 정도라고 생각하나요?

#### 초등학생들의 차도 관련 위험도

##### 요인 2

차가 다닐 수 있는 골목길에서 숨바꼭질 등의 놀이를 했다.

차가 다니는 도로 옆에서 친구들과 놀이를 했다.

주차된 차 근처에서 놀이를 했다.

### 03. 설문조사 분석

#### | 요인분석 결과

도출된 5개의 요인에 포함된 각 질문들의 공통점을 파악해 키워드 부여

#### 초등학생들의 횡단 시 위험도

(신호등, 횡단보도 등 안전시설물에 대한 인식)

요인 3
찾길 건너편에서 친구가 불렀을 때 바로 친구에게 건너갔다.
늦어서 횡단보도를 뛰어서 건넜다.
초록불로 바뀌자마자 바로 건넜다.
근처 횡단보도에서 초록 신호등이 깜빡일 때 횡단보도로 뛰어가서 건넜다.

#### 초등학생들의 대중교통

#### 이용 시 위험도

요인 4
타는 버스를 놓칠까봐 찾길을 바로 건넜다.
버스에서 내릴 때 급하게 내렸다.
급하게 버스에서 내려 길 반대편으로 바로 건넜다.

#### 초등학생들의 자전거

#### 이용 시 위험도

요인 5
자전거 도로가 아닌 곳에서 자전거를 탔다.
헬멧을 벗고 자전거를 탔다.
자전거를 탄 채로 횡단보도를 건넜다.

### 03. 설문조사 분석

#### | ANOVA

학생들의 요인 별 요인 점수를 사용해  
각 요인별로 3개의 **클러스터** 간에 유의미한 **중요도 차이**가 존재하는지 **ANOVA**를 통해 분석



각 클러스터 별로 1~3번 클러스터의 학생들의 5개의 요인에 대한  
각각의 중요도 지표 **평균값** 간에 **차이**가 **존재**하는지 검정



**가장 높은 평균**을 갖고 있는 **클러스터**가 해당 요인에  
**가장 높은 위험도**를 가진다고 판단

인코딩 시, 위험도가 높은 항목일 수록 높은 점수가 부여되도록 진행했기 때문!



### 03. 설문조사 분석

#### 설문조사 분석 결과

##### ANOVA 결과 해석

결론	
요인	요주의 클러스터
안전 전반적 인식도	클러스터 1
차도 관련 위험도	클러스터 1
횡단 시 위험도	클러스터 1
대중교통 이용 시 위험도	클러스터 2



설문조사 **요인 분석**과 **ANOVA**를 통해  
도출한 결론을 반영하여 각 **클러스터**에 맞는 **해결방안** 제시!

### 03. 설문조사 분석

#### | 명목형 자료 분석

#### 클러스터 1

일어날 뻔했거나 목격한 어린이 교통사고 또는 안전사고 유형	
인도를 건던 중 교통사고	440
일반 횡단보도를 건너던 중 교통사고	40
차길을 건던 중 교통사고	16
교차로 횡단보도를 건너던 중 교통사고	11
무단횡단 중 교통사고	9
길 가장자리를 건던 중 교통사고	8

집에서 학교 (학교에서 집) 로 가는 길에 위험요소	
신호등이 없음	91
골목 등 좁은 길에서 갑자기 차가 튀어나옴	38
차가 다니는 길가로만 다닐 수 있음	36
인도에 차가 서 있음	32
횡단보도 초록불 신호가 너무 짧음	17
길가에 차들이 많이 세워져 있어 앞을 가림	16

클러스터 1 대부분의 학생들이 인도에서의 위험을 인식

### 03. 설문조사 분석

#### | 명목형 자료 분석

#### 클러스터 2

교통사고와 안전사고를 막기 위해 필요하다고 생각하는 시설물이 있다면 추가로 적어주세요	
신호등	23
횡단보도	13
인도	7
안내표지판	3
CCTV	2
기타	5

집에서 학교 (학교에서 집)로 가는 길에 위험요소	
신호등이 없음	156
길가에 차들이 많이 세워져 있어 앞을 가림	60
인도에 차가 서 있음	44
횡단보도 초록불 신호가 너무 짧음	36
골목 등 좁은 길에서 갑자기 차가 튀어나옴	32
...	...

클러스터 2 대부분의 학생들이 **신호등 부재**의 위험을 인식  
실제로 신호등에 대한 필요성을 느낌

### 03. 설문조사 분석

#### | 명목형 자료 분석

#### 클러스터 3

집에서 학교로 갈 때 어떻게 가는지 선택해주세요	
걸어서	563
부모님 자동차	220
자전거	40
버스	16
어린이 통학버스	5
기타	6

일어날 뻔했거나 목격한 어린이 교통사고 또는 안전사고 유형	
인도를 걷던 중 교통사고	720
일반 횡단보도를 건너던 중 교통사고	58
차길을 걷던 중 교통사고	25
교차로 횡단보도를 건너던 중 교통사고	17
무단횡단 중 교통사고	17
길 가장자리를 걷던 중 교통사고	8

클러스터 3에서 보행으로 통학하는 초등학생 수가 많다는 점을 고려했을 때  
보행자 친화적 교통 시설물의 부족을 추측해볼 수 있음

# 06

## 해결방안 제시 및 현행 정책 검토

## 04. 해결방안 제시 및 현행 정책 검토

### | 클러스터 별 해결방안 제시

클러스터링과 요인분석에서 도출한 특징으로 각 클러스터에 알맞은 해결안 제시

클러스터	해결방안
클러스터1	교통 인프라 수준 향상 주차장 증설, 주정차 단속 강화 횡단보도 체계 개선
클러스터2	교통 안전 인식 개선 신호등 체계 개선 중앙분리대
클러스터3	보차 분리 교통표지 증설

## 04. 해결방안 제시 및 현행 정책 검토

| 2023년 기준 정책검토



클러스터링 및 요인분석을 진행한 결과,  
하나의 정책이 제주도 전역에서 통일적으로 이뤄지는 것이 아니라  
**클러스터별로 지역에 맞는 맞춤형 정책**이 나타나는 것이 좋을 것이라고 예상

# 감사합니다

진짜 안녕 ~

