



데이터 분석을 통한 범죄 건수 예측 및 대책



조장 : 최재명
조원 : 최정안
유수빈
김미나



01

p1

문제 인식


- 분석 배경
- 주제 소개



02

데이터 탐색

- 데이터 선정
- 데이터 분류
- 데이터 목록



03

데이터 분석

- 데이터 전처리
- 모델 설정 및 분석
- 결과 도출



04

Solution

- 문제점 개선 사항

우리의 삶은

안전한가?

p2



우리의 삶은 안전한가?



우리의 삶은 안전한가?

[국감] 수원역, 성범죄 발생 최다...서울역의 3배

▲ 이재형 선임기자 · 2022.10.09 12:15



- 최근 5년간 성범죄, 철도범죄의 36.4% '최다'
- 코레일 전동열차, CCTV 설치율 45% 불과

국회 국토교통위원회 소속 김선교 국회의원(국민의힘, 경기 여주시·영광군)이 한국철도공사(코레일) 및 철도 특별사법경찰대로부터 제출받은 자료에 의하면, 전국 철도범죄 중 가장 많은 범죄유형은 성범죄로, 수원역에서 가장 많은 성범죄가 발생한 것으로 나타났다.

구체적으로 최근 5년여간(2017-2022.8) 철도범죄는 총 12,734건이 발생했고, 성범죄 36.4%(4,631건), 점유이탈물횡령, 재물손괴 등 기타범죄 30.8%(3,921건), 상해·폭행 15.6%(1,987건), 철도 14.4%(1,837건), 철도차량 파손, 승무원 직무 방해 등 철도안전법 위반 7.4%(948건) 순이다.



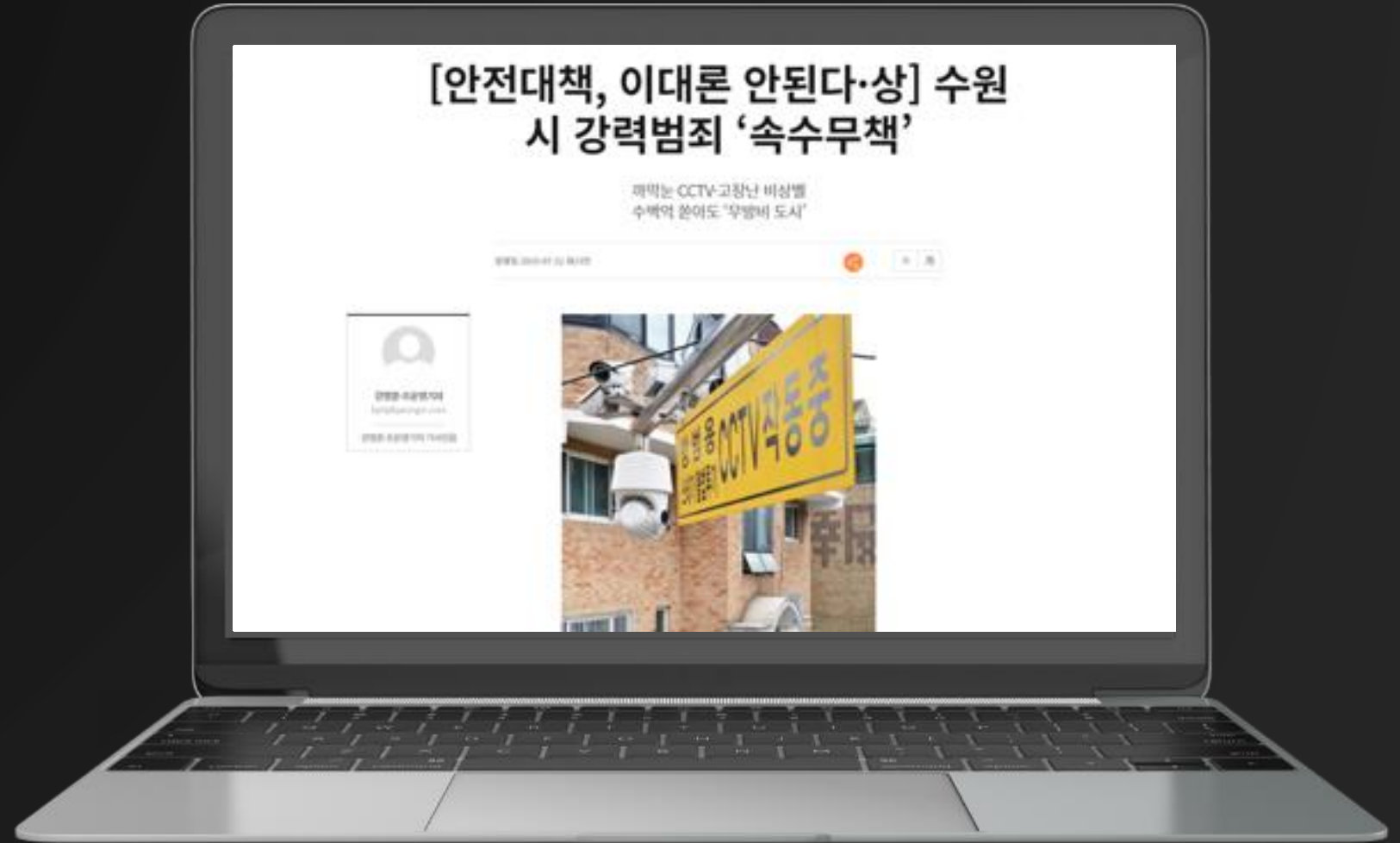


우리의

삶은

안전한가?

p2



우리의 삶은 안전한가?

1인가구 밀집지역, '범죄 발생률' 높다

형사정책연구원 보고서

강도·강제추행 등 5대 범죄 비밀집지역의 2~3배

백미영 기자 desk@lawtimes.co.kr | 입력 : 2018-03-12 오후 3:53:02

f 페이스북

트위터

인쇄

메일보내기

기사스크랩

스크랩 보기

혼밥, 혼술 등 새로운 사회현상으로 이어지고 있는 1인가구의 증가가 지역 범죄발생률에도 영향을 미친다는 연구결과가 나왔다. 1인가구 여부 자체가 범죄피해와 직접적인 연관성을 가지지는 않지만, 1인가구 밀집지역이 비(非)밀집지역보다 범죄발생률이 2~3배 이상 높다는 것이다.

형사정책연구원(원장 김진환)은 최근 이 같은 내용이 담긴 '1인가구 밀집지역의 안전실태와 개선방안 연구(책임연구원 박준휘)' 보고서를 발간했다.

통계청에 따른 1인가구의 표준정의는 혼자서 살림하는 가구, 즉 1인이 독립적으로 취사, 취침 등 생계를 유지하는 가구이다. 법력이나 사실상의 배우자가 존재하는가의 여

우리의 삶은 안전한가?

[1UF]잠 못 드는 1인가구 여성...주거침입 범죄 무방비 노출

아이룸 기자 | 승인 2021.02.19 18:32 | 댓글 0

지난해 우리나라는 1인가구 600만 시대에 진입했다. 이는 전체 가구 중 30%에 달하는 수준이다.

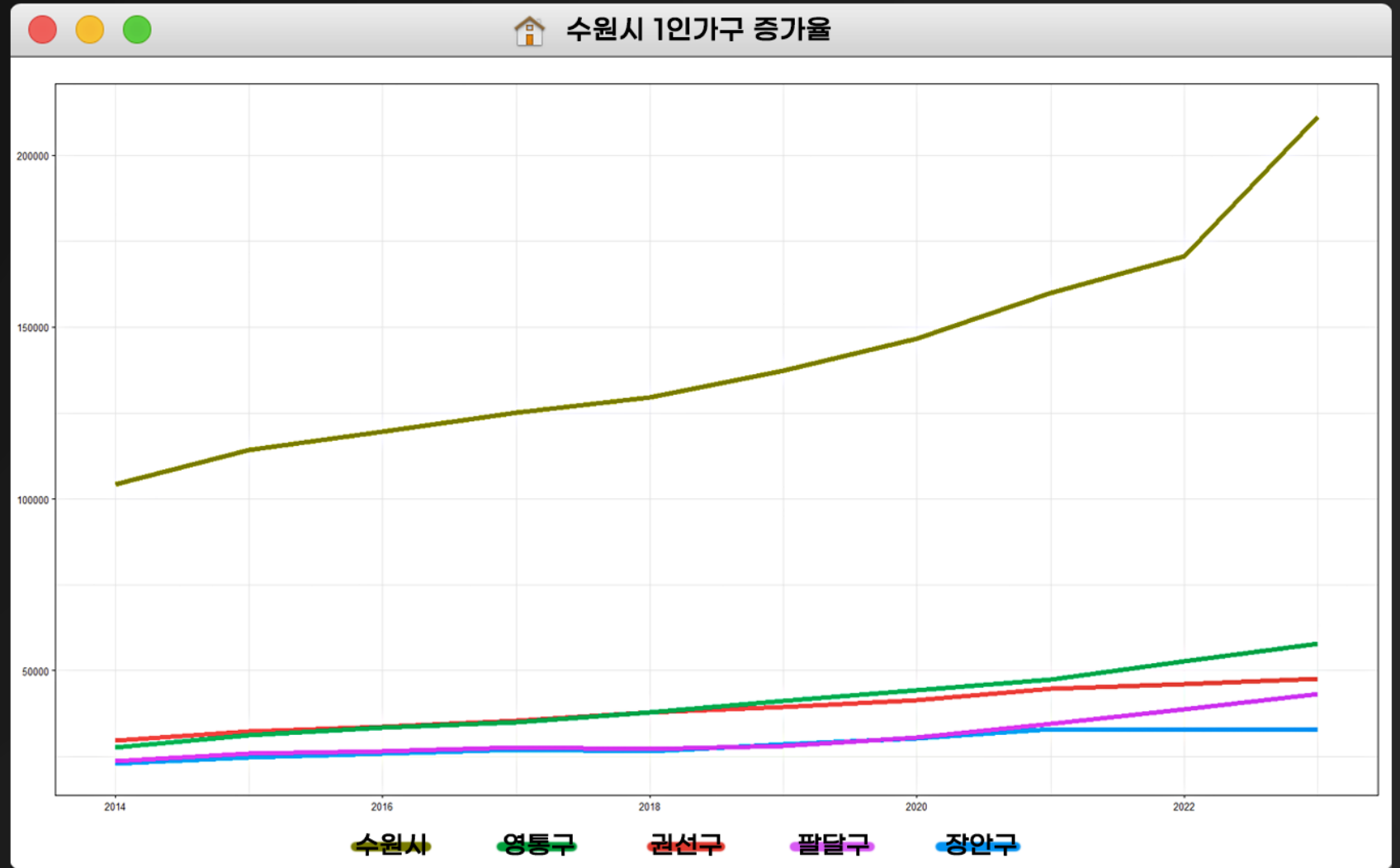
이 중 여성 1인가구는 291만4000가구로, 전체 1인가구의 절반 가량을 차지한다. 늘어나는 여성 1인가구의 수만큼 이들을 상대로 한 성범죄나 주거침입 사례도 끊이지 않고 있어 많은 여성이 불안에 떨고 있다.

이러한 현실을 방증하듯 지난해 통계청이 발표한 자료에 따르면 전반적인 사회 안전에 대한 질문에 '불안하다'고 대답한 여성 비율은 35.4%로 남성(27%)보다 높은 것으로 조사됐다. 특히 여성의 절반 이상이 범죄 발생(57%)에 대한 불안을 호소했는데, 여성들이 뽑은 우리 사회의 가장 불안한 요인 역시 범죄 발생(26.1%)인 것으로 나타났다.

그중에서도 강도·강간 등 강력범죄로 이어질 가능성이 큰 주거침입 범죄의 경우 보안이 허술한 다세대 주택이나 원룸에서 주로 발생하는데, 이는 보안 시설이나 경비원, 방범 시설이 갖춰진 아파트에 거주하기 힘든 청년들의 주거 양극화에 따른 결과로도 해석할 수 있다.



우리의 삶은 안전한가?



우리의 삶은 안전한가?

밤에 혼자 동네의 골목길을 걸을 때
범죄 피해에 대한 두려움을 느끼는 정도



4등급



수원시
범죄 안전 등급

55.3%



지역 공공데이터 기반 범죄 발생 위험지역 예측 모델
-전주시를 중심으로-

Crime risk area prediction model based on local public data: Focused on Jeonju

김동완 (Dongwan Kim) | 전북대학교 사회학과 학사과정 | asdf134652@gmail.com

이승재 (Seungjae Lee) | 전북대학교 산업정보시스템공학과 학사과정 | grasio10@naver.com

목 차

1. 서론
2. 이론적 배경
3. 연구방법
4. 연구결과
5. 결론

초록

도시의 성장 및 경제적 수준 향상 등으로 인해 사회는 끊임없이 변하며, 위험사회가 도래하게 되었고 지역사회 불안요인에 범죄는 항상 언급되고 있다. 대부분의 범죄는 무작위적인 분포로 발생하지 않고 공간 및 지역에 따라 일정한 패턴을 보이며 발생한다. 이에 본 연구에서는 전주시의 공공데이터와 범죄 주의 구간 데이터에 활용해 범죄 발생구역의 공간적 특징들을 분석하고, 이에 따른 범죄 발생 위험구역을 기계학습 예측 모델을 통해 구축하고자 하였다.

*키워드: 지역 공공데이터, 공간 빅데이터, 범죄분석, 기계학습, 예측 모델

ABSTRACT

The society is constantly changing due to the growth of cities and the improvement of the economic level, and a risk society has arrived, and crime is always mentioned as a factor of community unrest. Most crimes do not occur in a random distribution, but occur in a certain pattern according to space and region. Therefore, in this study, the spatial characteristics of the crime-prone area were analyzed using public data of Jeonju city and the data of the crime-prone section, and the resulting crime-prone area was constructed through a machine learning prediction model.

* Keywords : Spatial Big Data, Crime Analysis, Machine Learning, Prediction Model



p5



데이터 관련성



Data

논문 검색



데이터 수



정확도



데이터의 양



지표 설정

위해 지표

폭력, 절도, 주취자 수, 유흥업소...

취약 지표

교육시설, 아동/노인/여성 인구

경감 지표

경찰서 및 경찰 공무원 수,
CCTV와 가로등 설치 수



중요도 1

위해 지표

5대 범죄 발생율

취약 지표

1인 가구 수

경감 지표

인구 1000명 당 담당 경찰관 수



중요도 2

위해 지표

인구 1000명 당 주점 수

취약 지표

인구 수

경감 지표

인구 1000명 당 CCTV와 가로등 수

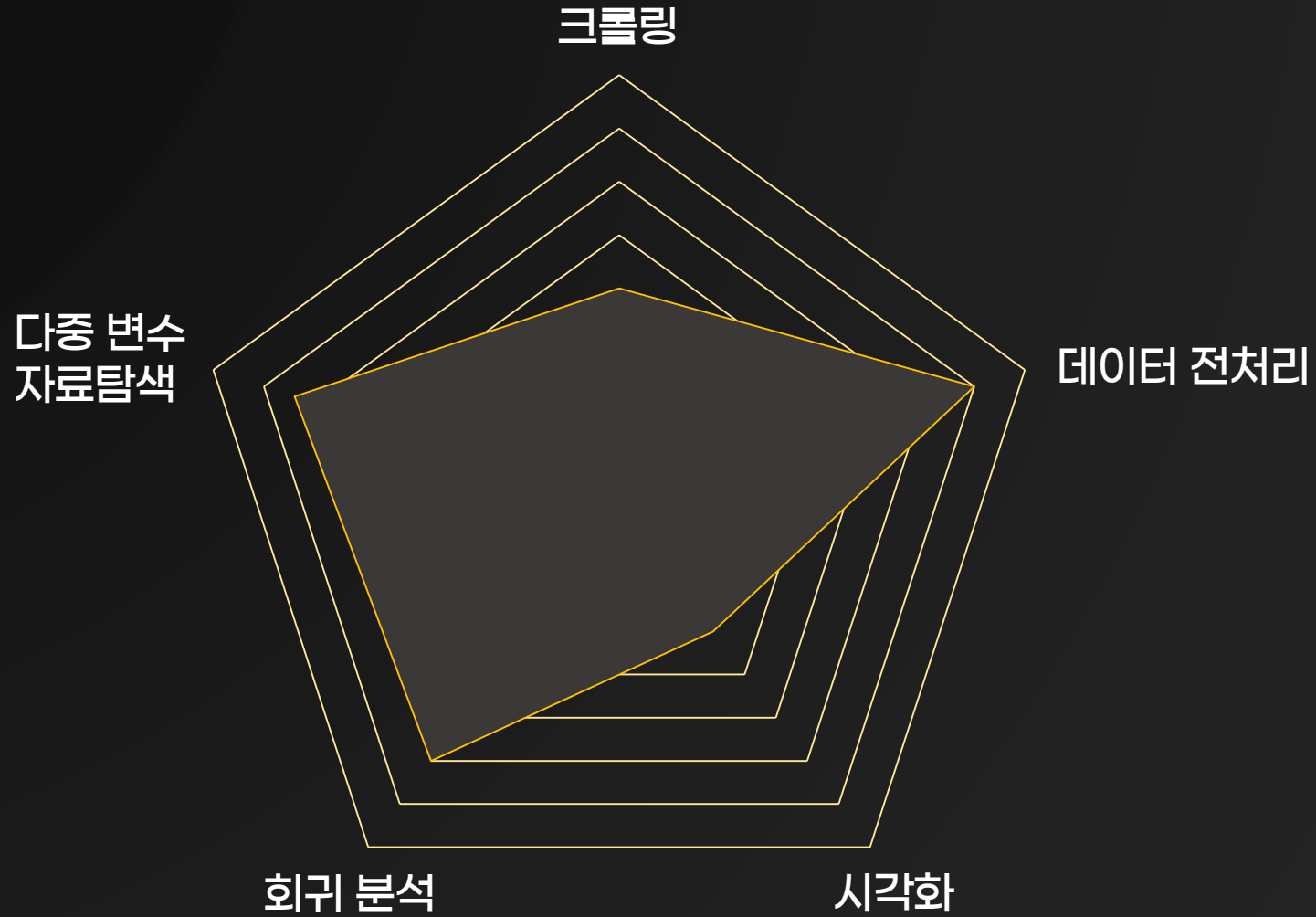


데이터 전처리

project 4



p8



PKG

readxl
ggplot2
rvest
stringr
plotly
wordcloud
wordcloud2
KoNLP
dplyr
foreign
xts
Deducer
leaps



데이터 전처리

project 4



p9

데이터셋									
	A	B	C	D	E	F	G	H	I
1	year	location	crime	people	cctv	steerlight	police	rate police	rate_crime
2	2000	수원시	21155	948065	643	7828	964	10.17	223
3	2001	수원시	20718	974866	682	8395	1014	10.40	213
4	2002	수원시	20281	1019711	701	9624	1063	10.42	199
5	2003	수원시	19843	1032944	722	10536	1113	10.78	192
6	2004	수원시	19406	1033829	752	11655	1162	11.24	188
7	2005	수원시	18968	1045587	814	12571	1212	11.59	181
8	2006	수원시	18531	1068906	871	13318	1261	11.80	173
9	2007	수원시	18093	1067702	921	14574	1311	12.28	169
10	2008	수원시	17656	1067425	940	15156	1360	12.74	165
11	2009	수원시	17218	1073149	1186	15715	1410	13.14	160
12	2010	수원시	16781	1077535	1370	16190	1459	13.54	156
13	2011	수원시	16343	1088489	1668	16380	1484	13.63	150
14	2012	수원시	15906	1120258	2377	16970	1512	13.50	142
15	2013	수원시	14441	1178509	3101	17992	1618	13.73	123
16	2014	수원시	14782	1209169	3481	21697	1680	13.89	122
17	2015	수원시	15299	1221975	4114	22967	1760	14.40	125
18	2016	수원시	14157	1231224	5780	23438	1777	14.43	115
19	2017	수원시	14104	1240480	7094	23933	1826	14.72	114
20	2018	수원시	14085	1242212	7742	24118	1862	14.99	113
21	2019	수원시	13680	1235022	9356	24305	1878	15.21	111
22	2020	수원시	12504	1221913	11479	24578	1921	15.72	102
23	2021	수원시	10415	1216958	12110	24598	1920	15.78	86
24	2022	수원시	9623	1225051	12890	33328	1976	16.13	79
25	2023	수원시	9186	1225761	12949	33516	2017	16.46	75
26	2014	장안구	2542	307151	787	4013	322	10.48	83
27	2014	권선구	4439	347928	915	8579	453	13.02	128
28	2014	팔달구	5064	217229	772	2545	564	25.96	233
29	2014	영통구	2737	336861	1007	6560	280	8.31	81
30	2015	장안구	2797	305166	959	4049	330	10.81	92
31	2015	권선구	4429	360138	1044	8767	471	13.08	123
32	2015	팔달구	5242	213938	1087	3002	582	27.20	245
33	2015	영통구	2831	342733	1024	7149	297	8.67	83
34	2016	장안구	2537	302819	1464	4131	334	11.03	84
35	2016	권선구	4097	367584	1402	9061	507	13.79	111

범죄 건수?

음의 상관 관계

- CCTV, 가로등, 경찰관 수, 1인 가구 수

양의 상관 관계

- 주점 수, 학교 수, 학생 수

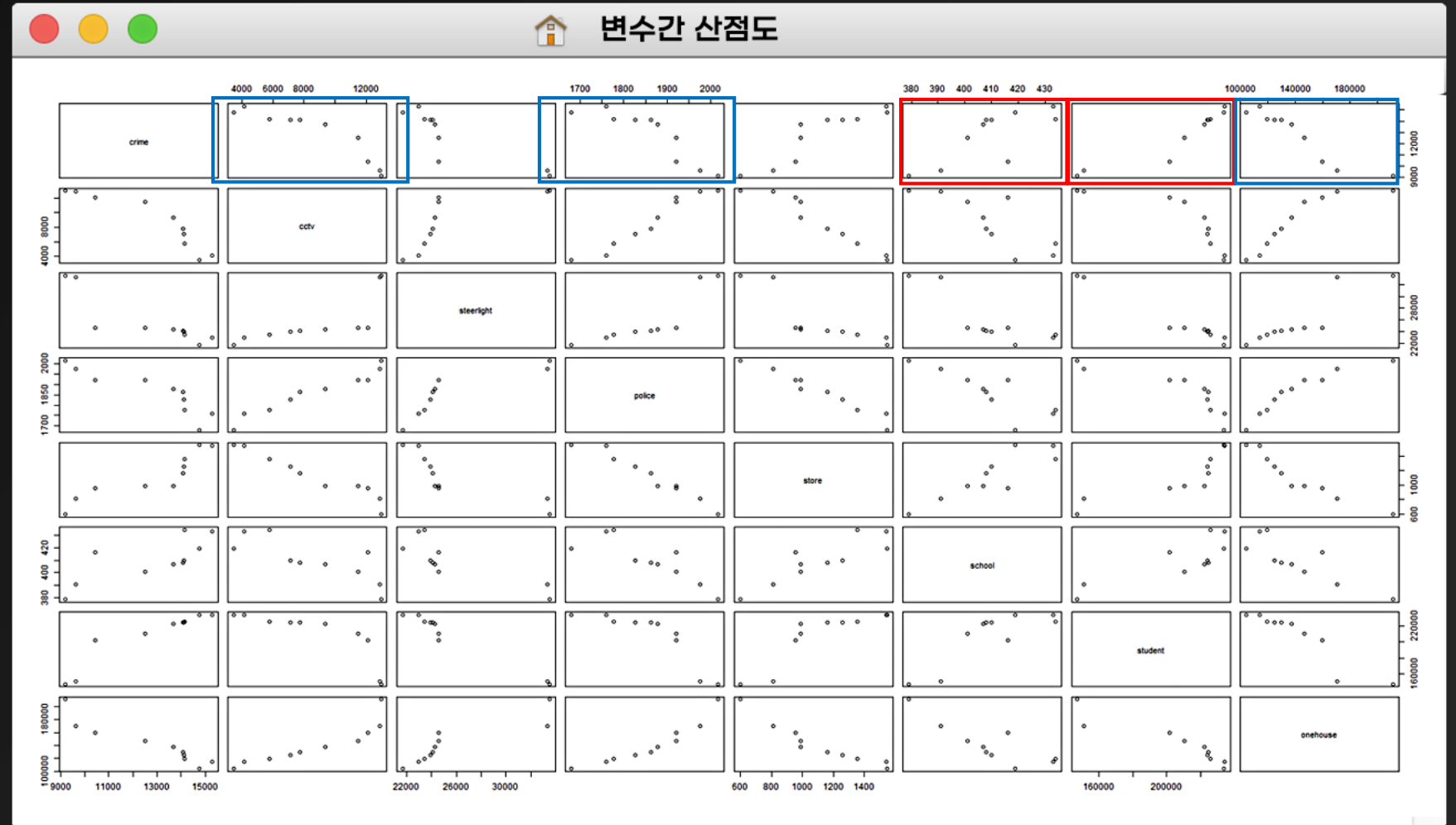
p10

	crime	people	cctv	steerlight	police	store	school	student	onehouse
crime	1.0000	0.1824	-0.9160	-0.8645	-0.8848	0.9029	0.7639	0.9381	-0.9390
people	0.1824	1.0000	0.0328	0.0222	0.1806	-0.1301	-0.1052	0.1021	-0.0266
cctv	-0.9160	0.0328	1.0000	0.7529	0.9690	-0.9621	-0.7938	-0.8156	0.8898
steerlight	-0.8645	0.0222	0.7529	1.0000	0.8190	-0.8251	-0.8207	-0.9818	0.8905
police	-0.8848	0.1806	0.9690	0.8190	1.0000	-0.9744	-0.8305	-0.8434	0.9272
store	0.9029	-0.1301	-0.9621	-0.8251	-0.9744	1.0000	0.8740	0.8584	-0.9448
school	0.7639	-0.1052	-0.7938	-0.8207	-0.8305	0.8740	1.0000	0.8218	-0.8365
student	0.9381	0.1021	-0.8156	-0.9818	-0.8434	0.8584	0.8218	1.0000	-0.9275
onehouse	-0.9390	-0.0266	0.8898	0.8905	0.9272	-0.9448	-0.8365	-0.9275	1.0000

산점도

변수 선택

- CCTV, 경찰관 수, 학교 수, 학생수, 1인 가구 수



최종 회귀식

범죄건수 = $77.17 \times [\text{학교 수}] - 0.04 \times [\text{CCTV 개수}] - 9.4 \times [\text{경찰관 수}]$

```
회귀분석

Residuals:
      Min       1Q   Median       3Q      Max
-96.598 -26.221  -8.388  40.382 111.201

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
cctv      -0.3950      0.1079   -3.660  0.00807 **
police    -9.3952      2.7163   -3.459  0.01056 *
school    77.1723      9.7553    7.911 9.79e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 71.85 on 7 degrees of freedom
Multiple R-squared:  0.9997, Adjusted R-squared:  0.9996
F-statistic: 8351 on 3 and 7 DF, p-value: 8.479e-13
```

23년 범죄 건수 분석

약 9,360건 예상

2021~2023 년도 데이터를 통한 범죄예측					
year	location	crime	pred	lwr	upr
2021	수원 시	10415	11429.07	10685.1	12173.05
2022	수원 시	9623	9353.2	8560.9	10145.5
2023	수원 시	<NA>	9362.26	8439.72	10284.81

실제값(crime)

예측값(pred)

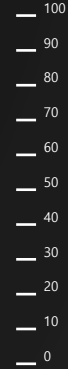
95% 신뢰하한(lwr)

95% 신뢰상한(upr)



수원시 경찰관 수

21년 기준 1,920명



최종 회귀식

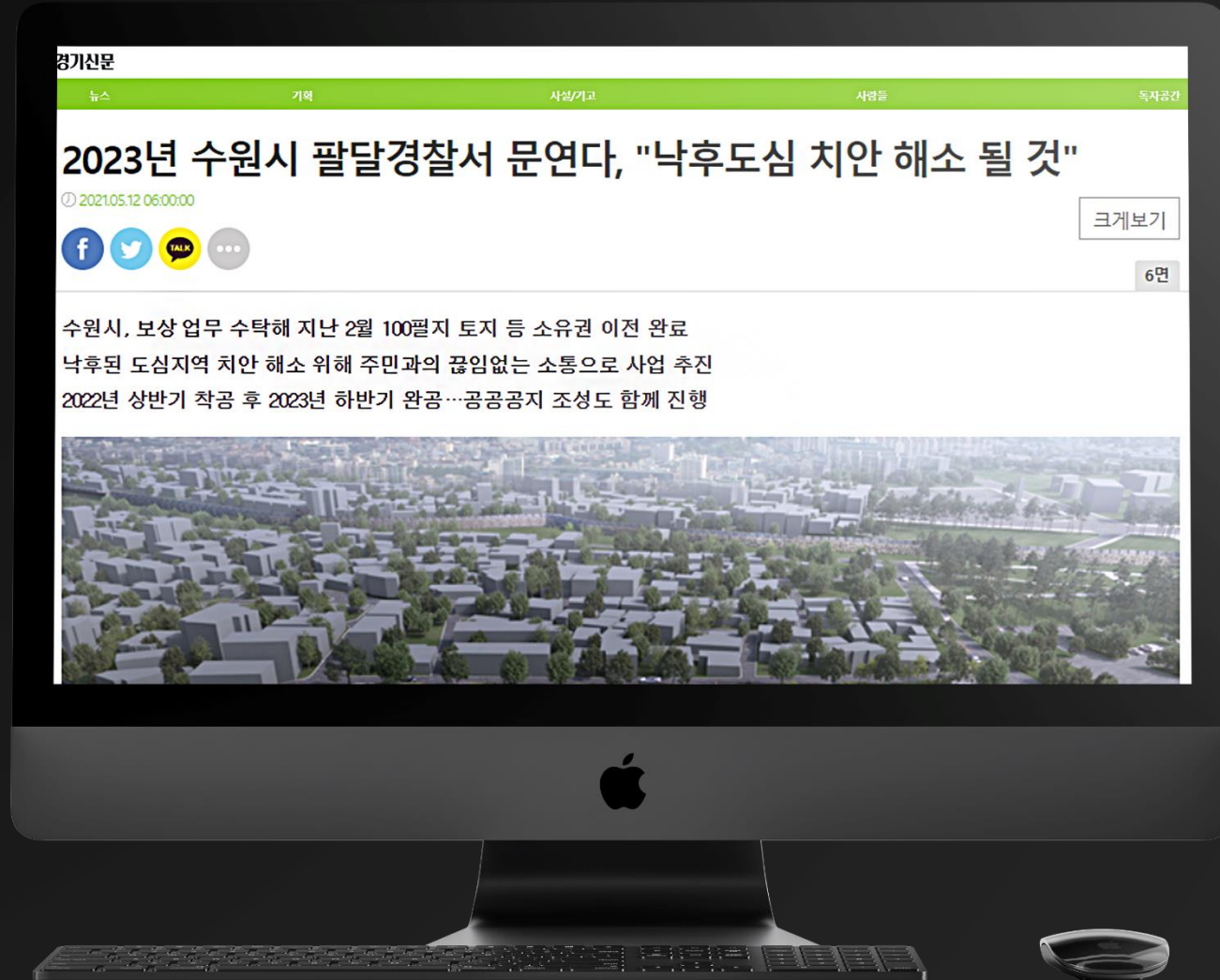
범죄건수 = $77.17 \times [\text{학교 수}] -$
 $0.04 \times [\text{CCTV 개수}] - 9.4 \times [\text{경찰관 수}]$



취약지역 문제점 개선

project name in eng

p15



문제점 1

‘구’별로 범죄율이 상이하

- 수집 변수들 외에도 다른 변수가 있을 것이라 추측
 - 논문 검색: 집값, 소득이 나옴

문제점 2

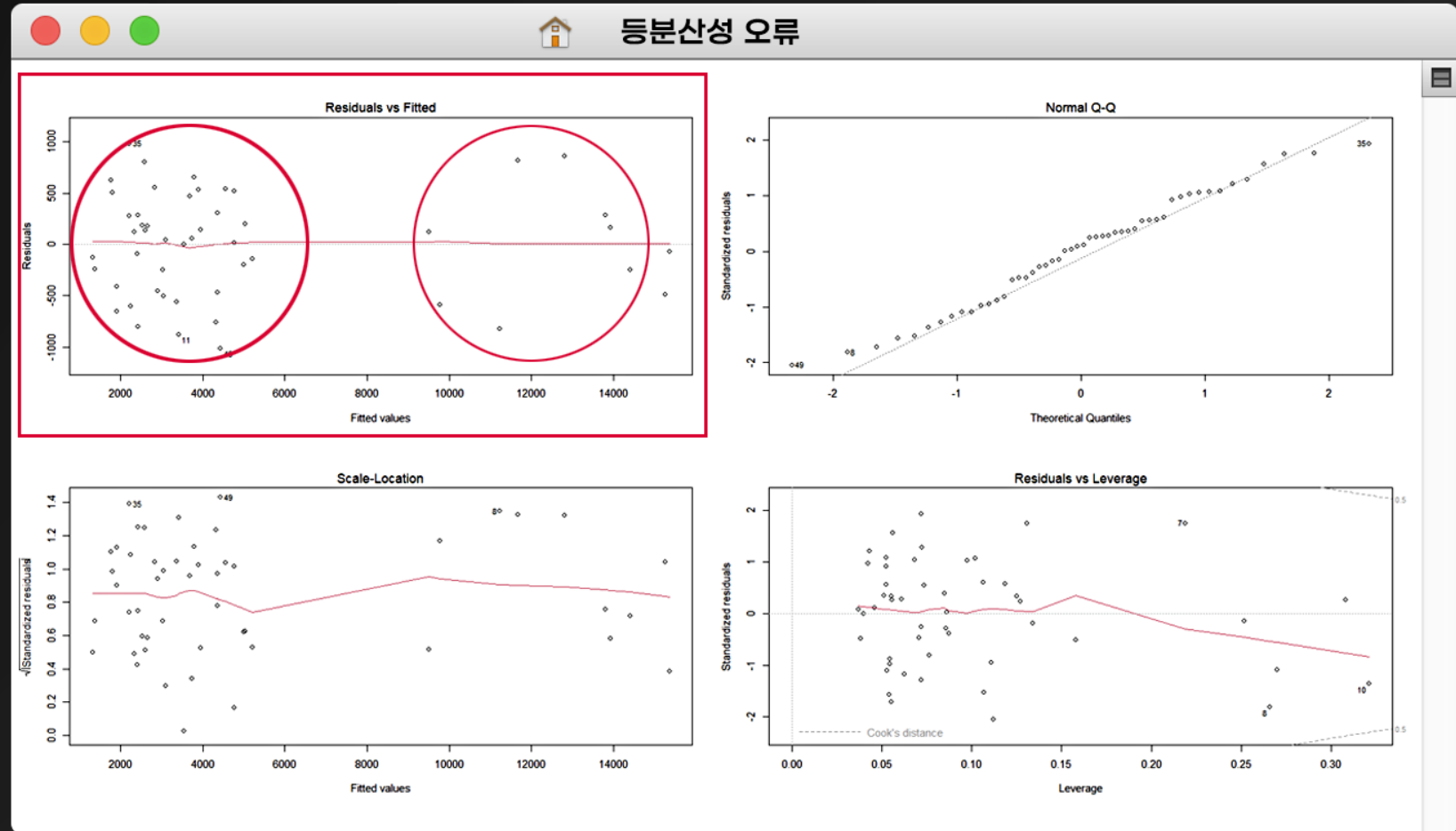
수원시의 비교 대상이 종속변수인 구가 되어버림

- 등분산성 오류
 - 비교 대상을 다른 인구가 비슷한 대도시로 해야 함

문제점 3

변수 간 가중치 미설정

- 행정안전부 데이터와 다른



Q&A



연세 **IT** 미래교육원
YONSAI IT EDUCATION CENTER

수집한 데이터 출처

경찰청 (<https://www.police.go.kr/index.do>)

수원남부경찰서 (<https://www.ggpolicy.go.kr/swnb/index.do>)

수원중부경찰서 (<https://www.ggpolicy.go.kr/swjb/index.do>)

수원서부경찰서 (<https://www.ggpolicy.go.kr/swsb/index.do>)

통계청 (<https://kostat.go.kr/ansk/>)

통계지리정보서비스 (<https://sgis.kostat.go.kr/view/index>)

국가통계포털 (<https://kosis.kr/>)

수원시 빅데이터포털 (<https://data.suwon.go.kr/>)

공공데이터포털 (<https://www.data.go.kr/>)

교육통계서비스 (<https://kess.kedi.re.kr/index>)

DBPIA (<https://www.dbpia.co.kr>)



조원 별 한줄 소감

project name in eng

p19

최재명



실력이 부족해서 아쉬웠습니다

오후 2시 12분

최정안



마무리 잘 해서 다행입니다

오후 2시 12분

유수빈



좋은 경험이었습니다

오후 2시 13분

김미나



좋은 기회였다고 생각합니다

오후 2시 14분

감사합니다



연세 **IT** 미래교육원
YONSAI IT EDUCATION CENTER