

프로젝트 보고서 : 지역별 교통사고 원인 분석 웹 서비스

파이썬 기반 머신러닝 응용 웹 서비스개발 과정

김다인 구동현 민세원 조양훈 최동휘



2020.12.08 ~ 2020.12.24



Contents

- I 주제 선정 배경
- II 프로젝트 개괄
- III 데이터 분석 과정
- IV 웹 개발 과정
- V 결론

I. 주제 선정 배경

교통사고

30%

지난해 어린이 보호구역 교통사고 30%증가
사망자도 2배 늘어

2019년

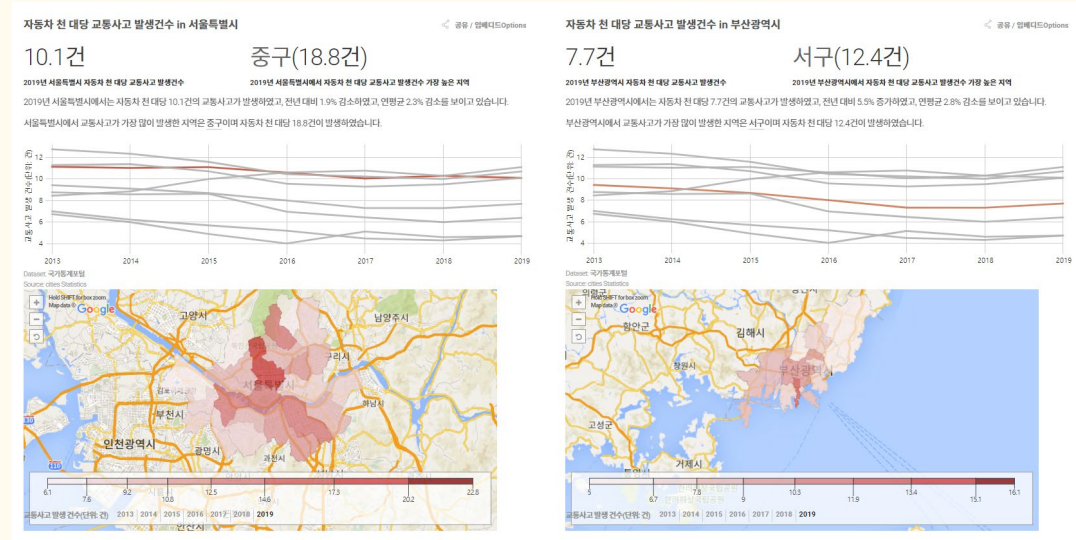
韓, 보행자 사망 OECD 평균보다 3.3배 높아
무단 횡단 · 고령자 사망 비율 높아

20km 제한속도와 관련없이,
다른 요인이 복합적으로 사고에 관여하는가에 대해 의문





→ 사용자 맞춤 교통 웹서비스 구현



<그림> 서울중구와 부산서구의 교통사고 발생건수 비교 (출처:데이터코리아)

사용자가 주로 운전하는 지역과 이동하는 지역의 데이터 분석을 통한 **사고원인 시각화** 및
교통사고 원인과 발생빈도의 연관성 분석을 통한 **주요 교통사고 원인 파악**
공공데이터 포털에서 데이터를 수집, 가공, 분석하여
도로교통사고 예방을 위한 새로운 위험도 평가방법 구현

II. 프로젝트 개괄

- 프로젝트 요약
- 프로젝트 목표
- 업무분장

프로젝트 요약



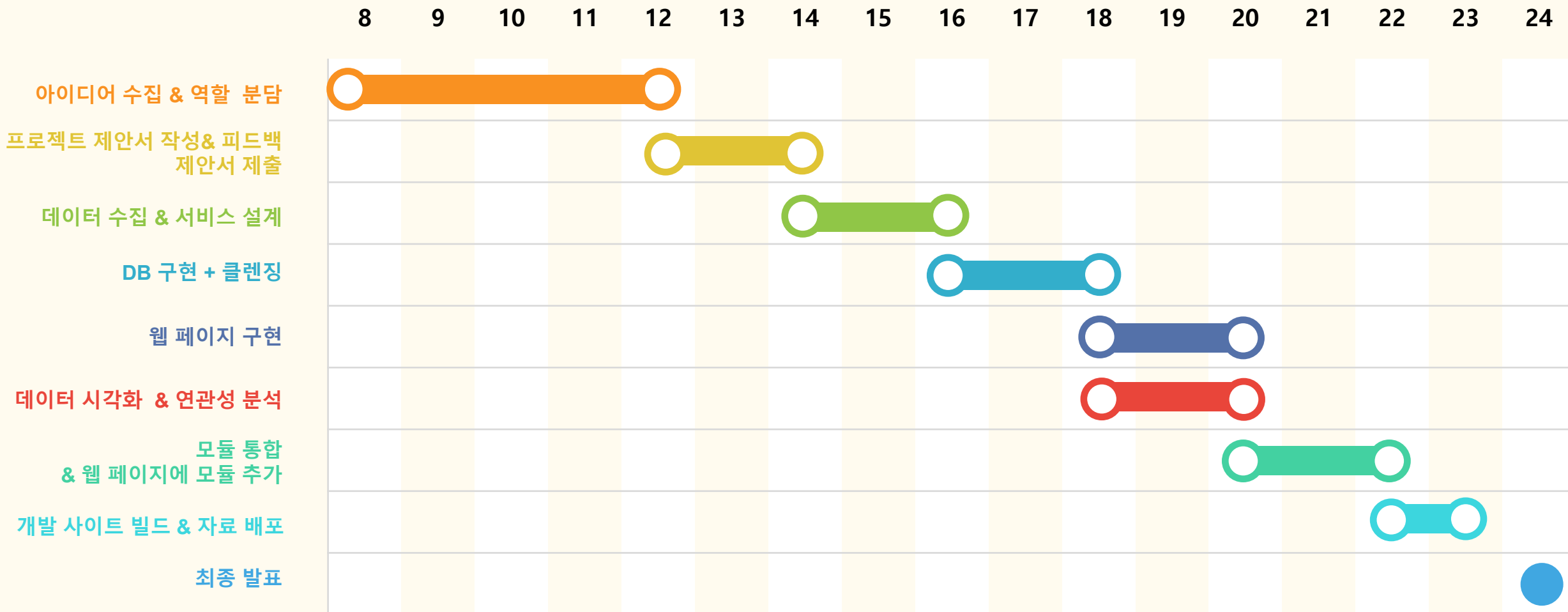
- **주제:** 서울시지역별 교통사고 원인분석 시각화 웹 서비스
- **사용 데이터:** 교통사고, 교통량, 날씨 데이터
- **참여 인원:** 5명(데이터 3팀/웹 2팀)
- **프로젝트 기간:** 2020.12.08~2020.12.24(약 2주)
- **주 사용 언어:** Python
- **주요 프로젝트 내용:** Python 라이브러리를 이용해 데이터 분석 및 시각화/머신러닝 분석한 내용을 웹 서비스로 제공

2020 12



프로젝트 목표

— Gantt 차트



업무 분장

— 데이터 분석/웹 개발



웹 개발

김다인



- 데이터 수집 및 구조 파악
- 풀 스택 웹 구현

웹 개발

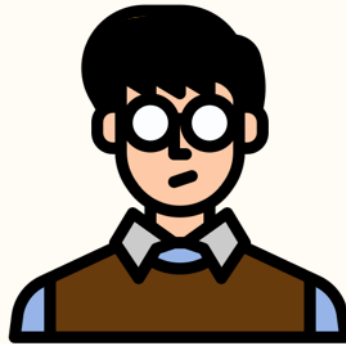
조양훈



- API를 이용한 기능 구현
- 풀 스택 웹 구현

데이터 분석

구동현



- 데이터 수집 및 전처리
- 데이터 분석 및 연관성 분석

데이터 분석

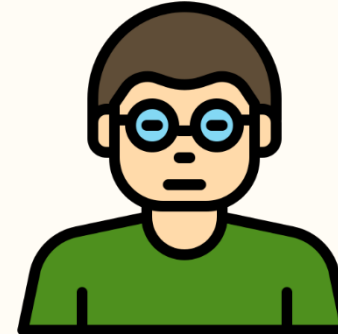
민세원



- 데이터 수집 및 전처리
- 데이터 분석 및 연관성 분석

데이터 분석

최동휘



- 데이터 수집 및 전처리
- 데이터 분석 및 연관성 분석

Ⅲ. 데이터 분석 과정

- 수집데이터
- 기술통계
- 추론통계

수집 데이터

→ 데이터 출처



2017 ~ 2019 서울시 교통사고

교통사고	교통량	날씨
<ul style="list-style-type: none">• 서울시교통사고분석시스템 GIS• 교통사고 정보	<ul style="list-style-type: none">• 서울시교통정보시스템 TOPIS• 지점별, 일자별 교통량	<ul style="list-style-type: none">• 기상청 날씨누리• 평균 기온, 강수량

수집 데이터



1. 서울시 교통사고 데이터 (GIS)

	사고번호	사고일시	요일	시군구	사고내용	사망자수	중상자수	경상자수	부상자수	사고유형	...	기상상태	도로형태	가해운전자종	가해운전자성별	가해운전자연령	가해운전자상해정도	피해운전자종	피해운전자성별	피해운전자연령	피해운전자상해정도
0	2017010100100001	2017년 1월 1일 00시	일요일	서울특별시 용산구 한남동	경상 사고	0	0	5	0	차대차 - 측면 충돌	...	맑음	단일 - 기타	승용	남	58세	상해없음	승용	남	24세	경상
1	2017010100100002	2017년 1월 1일 00시	일요일	서울특별시 강북구 미아동	중상 사고	0	1	6	0	차대차 - 측면 충돌	...	맑음	교차로 - 교차로 안	승용	남	58세	상해없음	승용	남	67세	경상

- 22개 Column

사고번호 / 사고일시 / 요일 / 시군구 / 사고내용 / 사망자수 / ... /
피해운전자차종 / 성별 / 연령 / 상해정도

- 총 11,676개 데이터

수집 데이터



2 서울시 교통량 데이터 (TOPIS)

	일자	요일	지점명	지점번호	구분	방향	0시	1시	2시	3시	...	14시	15시	16시	17시	18시	19시	20시	21시	22시	23시
0	20171001	일	성산로 (금화터널)	A-01	유입	봉원고가차도- >독립문역	641.0	482.0	312	263	...	1675.0	1740.0	1526.0	1647.0	1345.0	1200.0	1089.0	1063.0	853.0	687.0
1	20171002	월	성산로 (금화터널)	A-01	유입	봉원고가차도- >독립문역	428.0	324.0	261	210	...	1687.0	1809.0	1902.0	1842.0	1815.0	1552.0	1374.0	1162.0	973.0	758.0

- 30개 Column :
일자 / 요일 / 지점명 / 지점번호 / 구분 / 방향 / 0시 / ... / 23시
- 월별 데이터 개수 약 8500 개
- 36개의 데이터 파일

수집 데이터



3. 서울시 기상 데이터 (기상청 날씨누리)

[일평균기온(℃)] 108 서울 / 2017년

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
1일	2.7	-4.0	5.0	9.0	20.1	23.3	25.3	28.6	23.3	18.2	14.4	-2.2
2일	5.0	-2.7	1.8	9.0	20.7	20.5	24.3	30.0	23.8	22.0	15.5	1.9
3일	2.0	1.5	3.3	11.4	21.5	20.5	24.7	30.2	23.7	17.6	12.3	4.9
4일	3.9	0.9	5.7	14.1	21.0	21.7	26.7	30.3	24.6	16.7	7.8	-1.2
5일	3.8	1.2	5.2	12.0	20.9	24.2	27.4	31.4	21.9	18.7	8.4	-4.2

[일강수량(mm)] 108 서울 / 2017년

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
1일			5.0	1.0			0.5	2.0		4.0		0.0
2일	0.3		0.5			0.0	92.0			1.5	1.0	
3일							67.5				1.5	4.0
4일							0.5					0.1
5일		0.5	0.5	37.5		0.0		0.0	0.0			0.1

- 일자별 평균기온, 강수량 데이터
- 2017~ 2019년 연도별 데이터
- 12 columns

기술 통계

→ 전처리 과정



서울시 2017년-2019년
3년간 교통사고 데이터 수집



공백 데이터, Nan 값
0으로 변환



구와 기상상태별로
교통사고수를 나타내는
데이터로 변환



필요한 변수 추출

25개 구의 기상상태별 교통사고수
Pie 형태로 출력



기술 통계



→ 전처리 과정: 서울시 교통사고 통계

	사고번호	사고일시	요일	시군구	사고내용	사망자수	중상자수	경상자수	부상신고자수	사고유형	...	기상상태	도로형태	가해운전자종	가해운전자성별	가해운전자연령	가해운전자상해정도	피해운전자종	피해운전자성별	피해운전자연령	피해운전자상해정도
0	2017010100100001	2017년 1월 1일 00시	일요일	서울특별시 용산구 한남동	경상 사고	0	0	5	0	차대차-측면충돌	...	맑음	단일로-기타	승용	남	58세	상해없음	승용	남	24세	경상
1	2017010100100002	2017년 1월 1일 00시	일요일	서울특별시 강북구 미아동	중상 사고	0	1	6	0	차대차-측면충돌	...	맑음	교차로-교차로안	승용	남	58세	상해없음	승용	남	67세	경상

	사고번호	사고일시	요일	시군구	사고내용	사망자수	중상자수	경상자수	부상신고자수	사고유형	...	도로형태	가해운전자종	가해운전자성별	가해운전자연령	가해운전자상해정도	피해운전자종	피해운전자성별	피해운전자연령	피해운전자상해정도	구
0	2017010100100001	2017년 1월 1일 00시	일요일	서울특별시 용산구 한남동	경상 사고	0	0	5	0	차대차-측면충돌	...	단일로-기타	승용	남	58세	상해없음	승용	남	24세	경상	용산
1	2017010100100002	2017년 1월 1일 00시	일요일	서울특별시 강북구 미아동	중상 사고	0	1	6	0	차대차-측면충돌	...	교차로-교차로안	승용	남	58세	상해없음	승용	남	67세	경상	강북

기술 통계

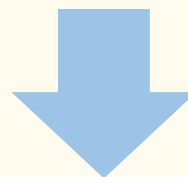


전처리 과정: 서울시 교통량

일자	지점명	구	방향	0시	1시	2시	3시	...	14시	15시	16시	17시	18시	19시	20시	21시	22시	23시
0 20171001	성산로 (금화터널)	아우터	동원고가차도->독립문역	641.0	482.0	312	263	...	1675.0	1740.0	1526.0	1647.0	1345.0	1200.0	1089.0	1063.0	853.0	687.0
1 20171002	성산로 (금화터널)	아우터	동원고가차도->독립문역	428.0	324.0	261	210	...	1687.0	1809.0	1902.0	1842.0	1815.0	1552.0	1374.0	1162.0	973.0	758.0



	지점명	주소
0	성산로(금화터널)	서울 서대문구 신촌동
1	사직로(사직터널)	서울 종로구 사직동
2	자하문로(자하문터널)	서울 종로구 청운동
3	대사관로(삼청터널)	서울 성북구 성북동
4	을곡로(안국역)	서울 종로구 을곡로



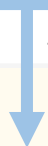
일자	요일	주소	지점명	0시	1시	2시	3시	4시	5시	...	15시	16시	17시	18시	19시	20시	21시	22시	23시	구
0	토	서울 서대문구 신촌동	성산로(금화터널)	893.0	716.0	523.0	483	415.0	604.0	...	1918.0	1892.0	1925.0	1966.0	1453.0	1370.0	1328.0	1183	862.0	서대문구
1	일	서울 서대문구 신촌동	성산로(금화터널)	572.0	452.0	301.0	242	223.0	356.0	...	1723.0	1565.0	1519.0	1287.0	1149.0	972.0	1151.0	877	632.0	서대문구

기술 통계



→ 전처리 과정: 서울시 교통량

Unnamed: 0			0시	1시	2시	4시	5시	6시	7시	8시	9시	...
구 일자												
강 남 구	2017-07-01	6440.352941	1483.411765	1239.588235	1088.117647	941.000000	1110.882353	1432.647059	1593.882353	1984.764706	2255.000000	... 2342.64
	2017-07-02	6235.642857	1293.285714	1084.928571	895.285714	727.642857	827.857143	941.928571	1069.285714	1399.714286	1795.928571	... 2155.57
	2017-07-03	6236.642857	864.785714	709.857143	601.000000	618.571429	1184.428571	2056.642857	2582.214286	2630.285714	2408.857143	... 2345.42
	2017-07-04	6306.400000	1303.000000	1054.933333	867.200000	711.200000	1046.000000	1877.466667	2675.666667	2847.733333	2702.933333	... 2373.13
	2017-07-05	6307.400000	1420.666667	1136.800000	920.866667	752.533333	1145.533333	1991.866667	2612.933333	2802.933333	2680.000000	... 2409.33
...



- 구, 일자별 평균 교통량 계산

기술 통계

→ 전처리 과정: 서울시 날씨



[일평균기온(℃)] 108 서울 / 2017년

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
1일	2.7	-4.0	5.0	9.0	20.1	23.3	25.3	28.6	23.3	18.2	14.4	-2.2
2일	5.0	-2.7	1.8	9.0	20.7	20.5	24.3	30.0	23.8	22.0	15.5	1.9
3일	2.0	1.5	3.3	11.4	21.5	20.5	24.7	30.2	23.7	17.6	12.3	4.9
4일	3.9	0.9	5.7	14.1	21.0	21.7	26.7	30.3	24.6	16.7	7.8	-1.2
5일	3.8	1.2	5.2	12.0	20.9	24.2	27.4	31.4	21.9	18.7	8.4	-4.2



[일강수량(mm)] 108 서울 / 2017년

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
1일			5.0	1.0			0.5	2.0		4.0		0.0
2일	0.3		0.5			0.0	92.0			1.5	1.0	
3일							67.5				1.5	4.0
4일							0.5					0.1
5일		0.5	0.5	37.5		0.0		0.0	0.0			0.1



	일자	평균기온	일강수량	날씨
0	2019-01-01	-5.0	NaN	맑음
1	2019-01-02	-4.9	NaN	맑음
2	2019-01-03	-3.5	NaN	맑음
3	2019-01-04	-1.1	NaN	맑음
4	2019-01-05	-2.8	NaN	맑음
...

- 일자별 평균기온, 강수량, 날씨 데이터 정리

기술 통계



→ 전처리 과정: 서울시 교통량 & 날씨

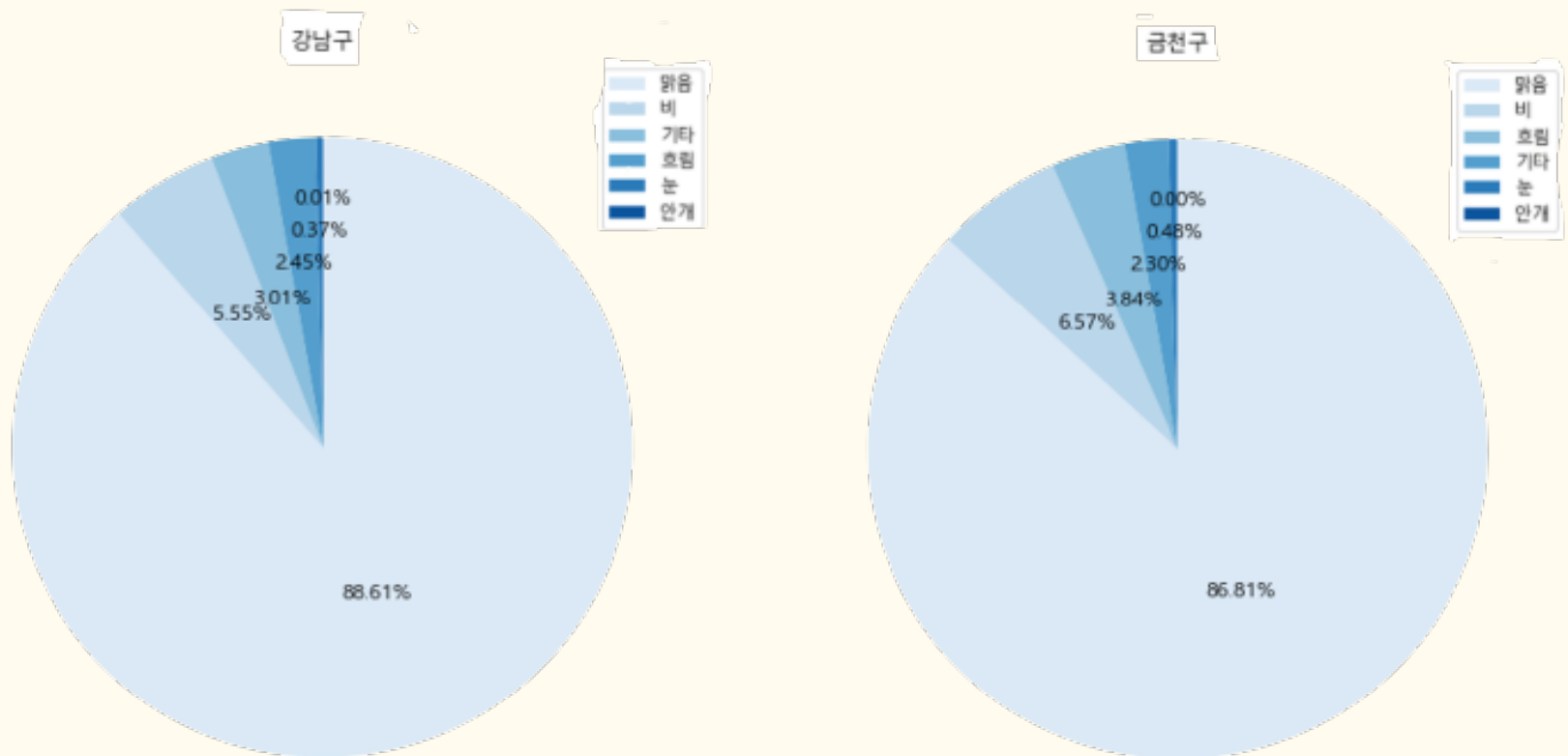
	일자	요일	구	0시	1시	2시	3시	4시	5시	6시	...	16시	17시	18시	19시	20시	21시	22시	23시	평균 기온	날 씨
0	2019-01-02	수	성북구	216.0	137.0	102.0	90.0	233.0	363.0	789	...	1303.0	1461.0	1694.0	1191.0	820.0	779.0	587.0	377.0	-4.9	맑음
1	2019-01-02	수	중랑구	958.0	479.0	296.0	360.0	600.0	1230.0	2289	...	3754.0	4097.0	3924.0	3366.0	2753.0	2372.0	1986.0	1313.0	-4.9	맑음

- 일자, 구, 시간에 따른 교통량과 날씨 데이터

- 29 columns

기술 통계

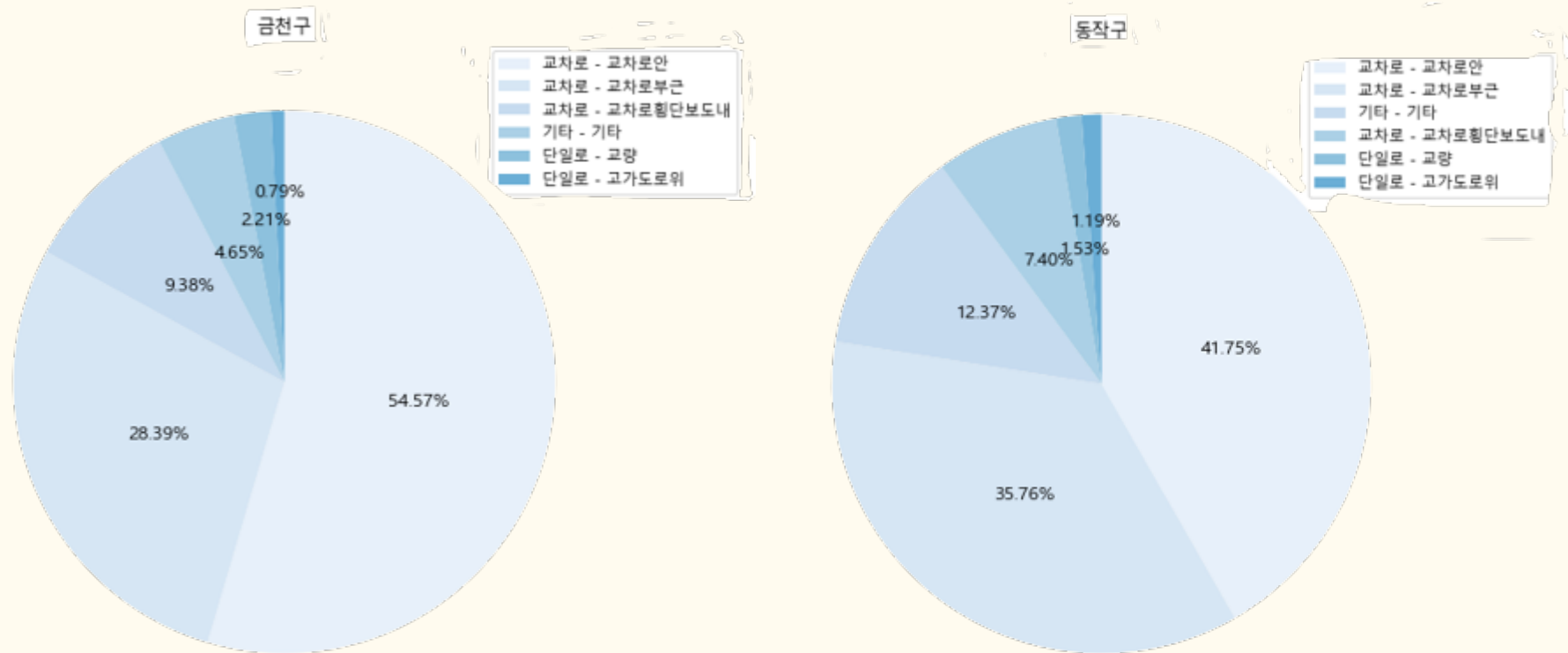
→ 시각화 결과(Pie): 날씨별 교통사고 비율



날씨별 교통사고 비율 (강남구, 금천구)

기술 통계

→ 시각화 결과(Pie): 도로형태별 교통사고 비율

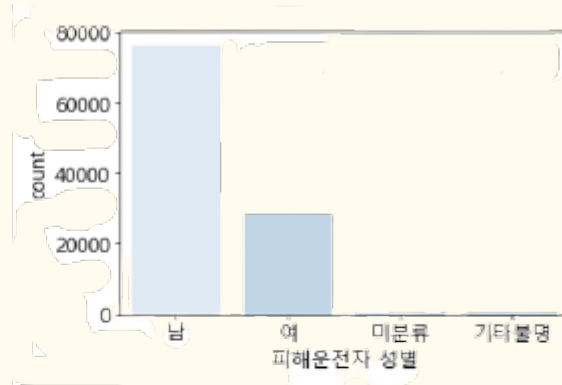
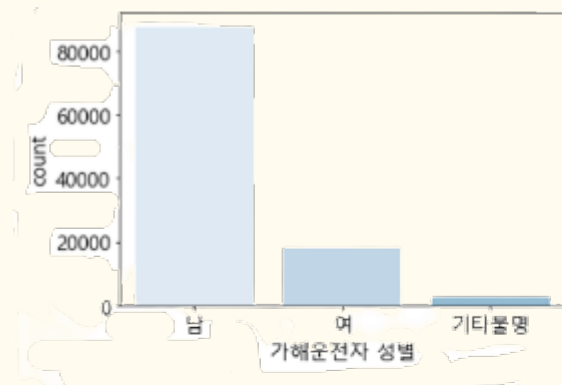


도로형태별 교통사고 비율 (금천구, 동작구)

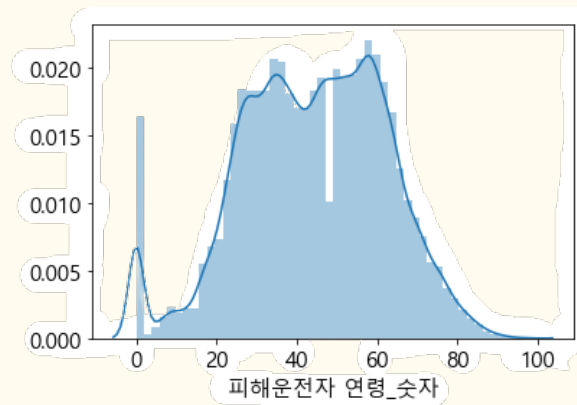
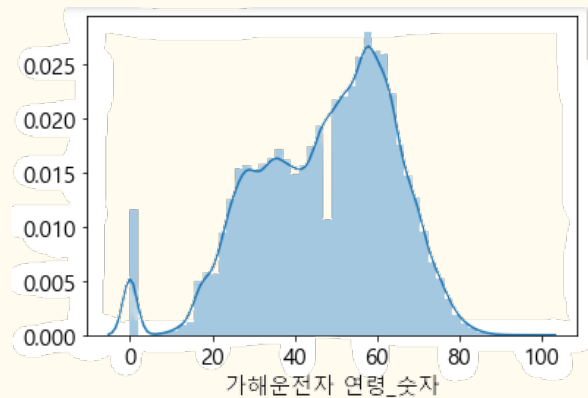
기술 통계



시각화 결과(Bar): 가해운전자 & 피해운전자



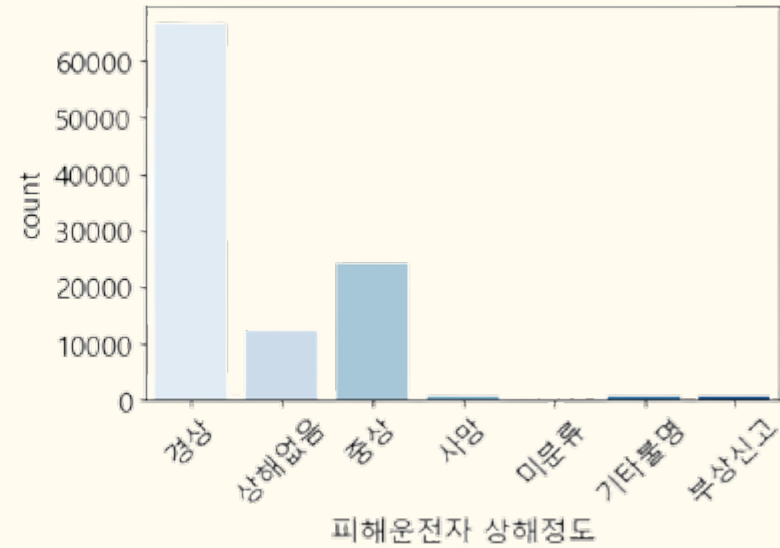
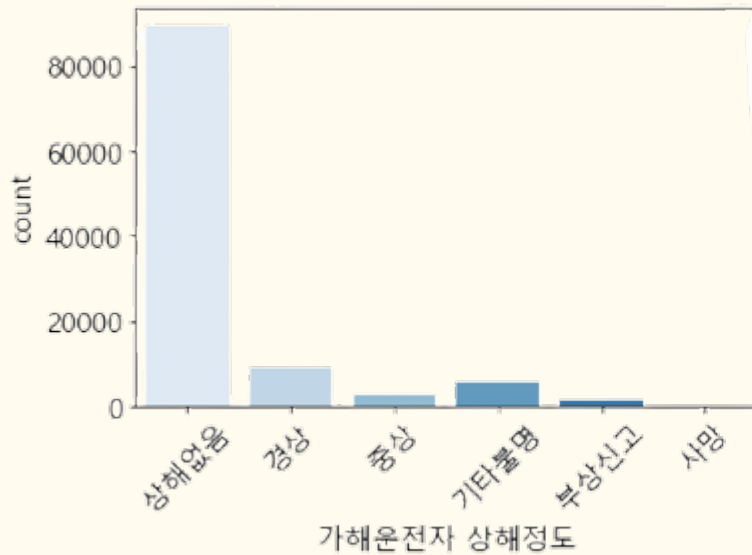
- 가해운전자/피해운전자 :
두 가지 경우 모두 남자의 성별이
압도적으로 많음



- 가해 운전자 : 60대, 70대
- 피해 운전자 : 20대~70대 고르게 분포

기술 통계

→ 시각화 결과(Bar): 상해정도

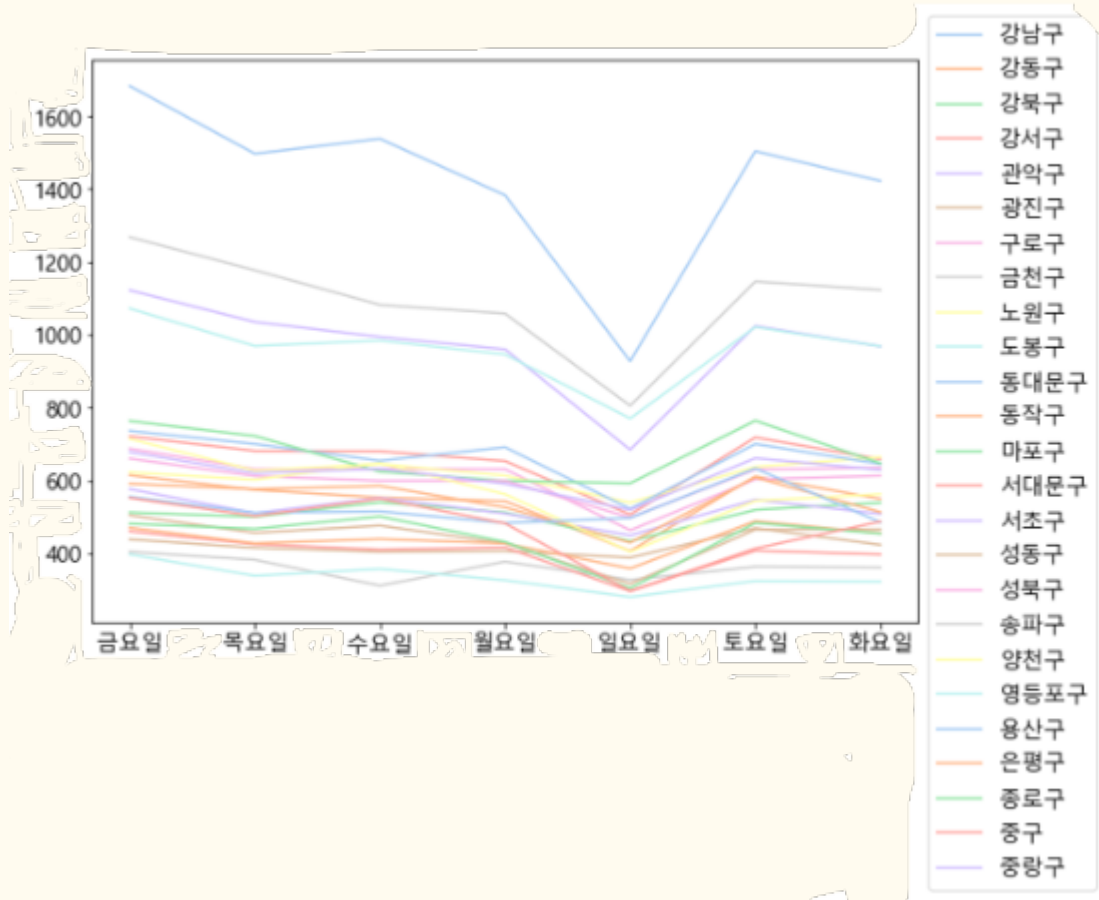


- 가해운전자
 - '상해없음' 빈도가 가장 높음

- 피해운전자
 - '경상' 빈도가 가장 높음
 - '중상'이 두번째로 높음

기술 통계

→ 시각화 결과(Line)



2017년 + 2018년 + 2019년

총 3년간의 실제 사고 건수



요일 별 시각화

추론 통계

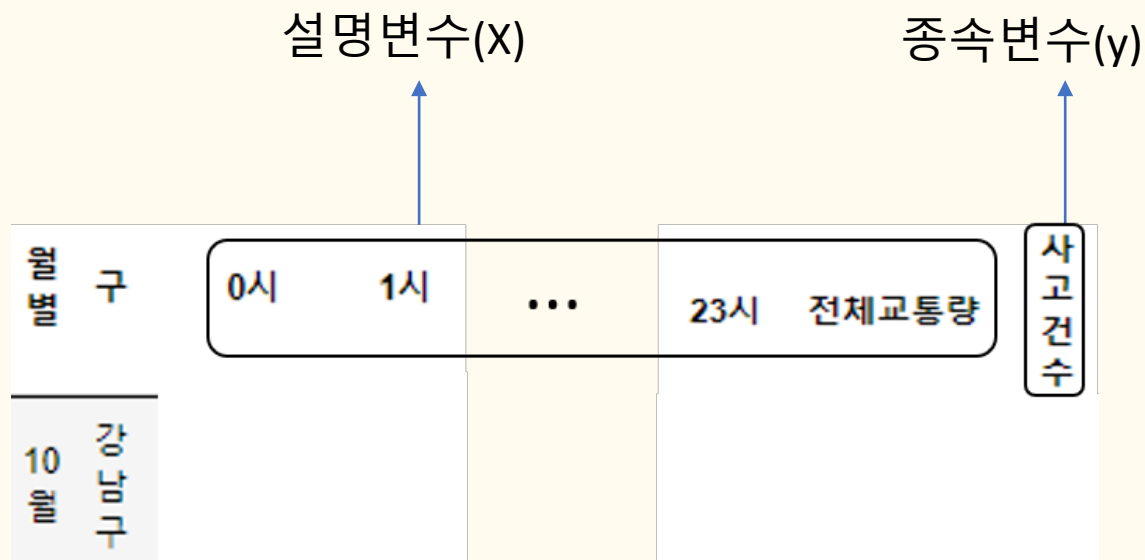
→ 3단계분석



추론 통계

→ 분석 과정

1. 월 별로 구의 24시간 교통량과 교통사고량 형태의 데이터로 전처리



- 10월 강남구의 07시 교통량/10월 강남구의 교통사고 수



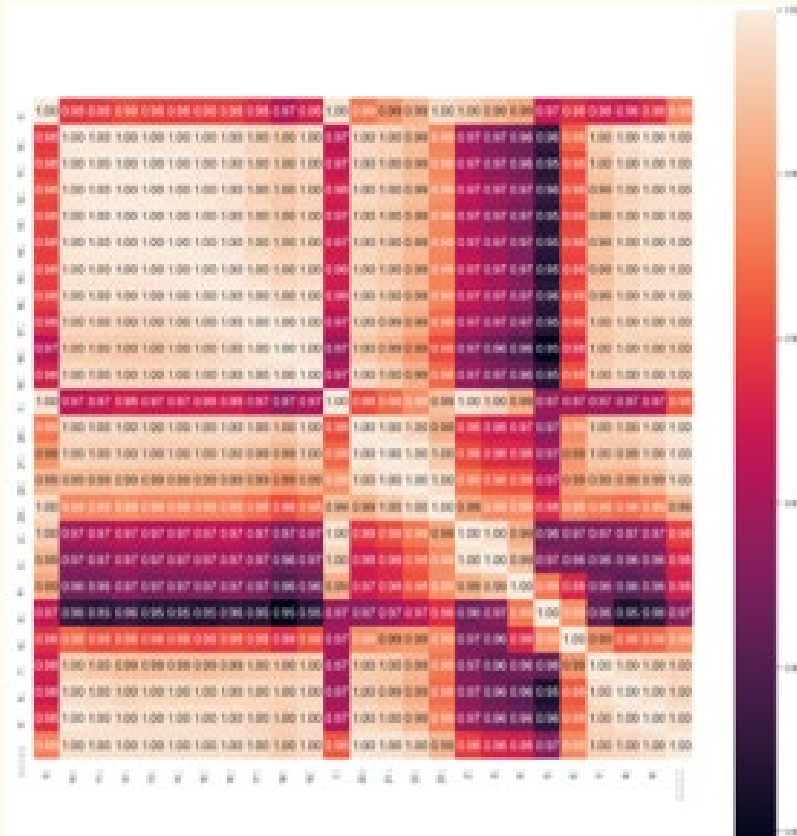
추론 통계

분석 과정

1. PCA 주성분 분석을 통한 2020년 교통사고 예측



1. 월 별로 구의 24시간 교통량과 교통사고량 형태의 데이터로 전처리



- 밝을수록 상관관계 높음

- x_i 간 상관관계 매우 높음

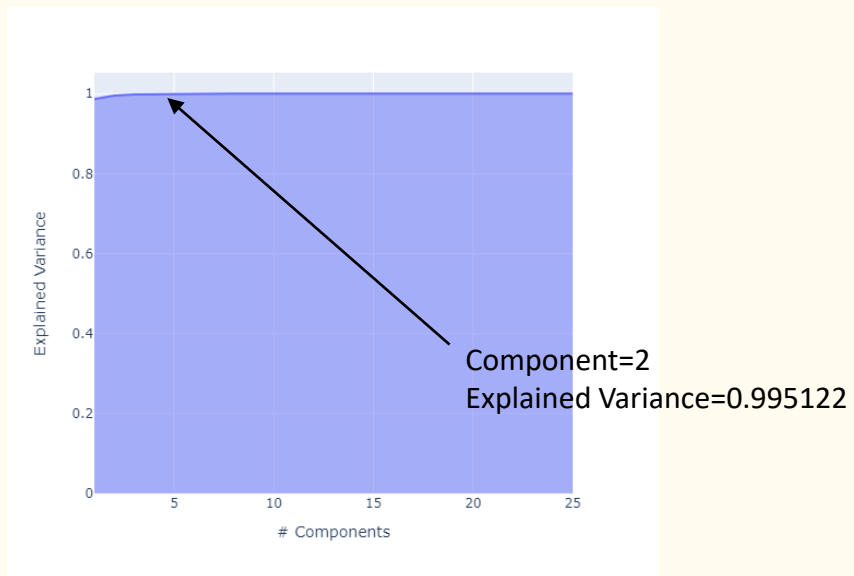
추론 통계

분석 과정

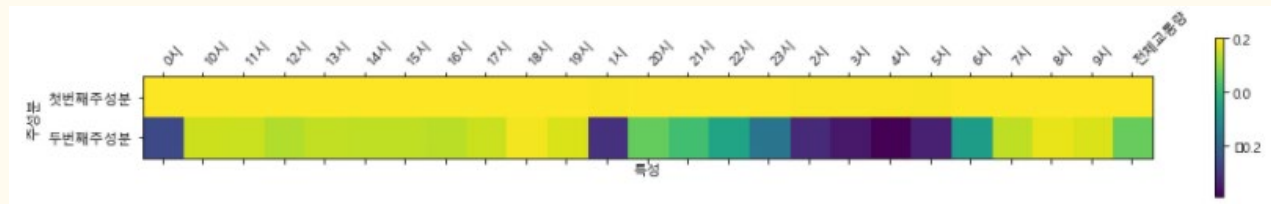
1. PCA 주성분 분석을 통한 2020년 교통사고 예측



2. Standard Scaler 사용한 데이터 변환



3. PCA에 fit



- X 변수 모두 중요한 성분

- 모든 요인으로 적절한 component 개수 선정
- PCA의 요소는 2개로 설정

추론 통계

분석 결과

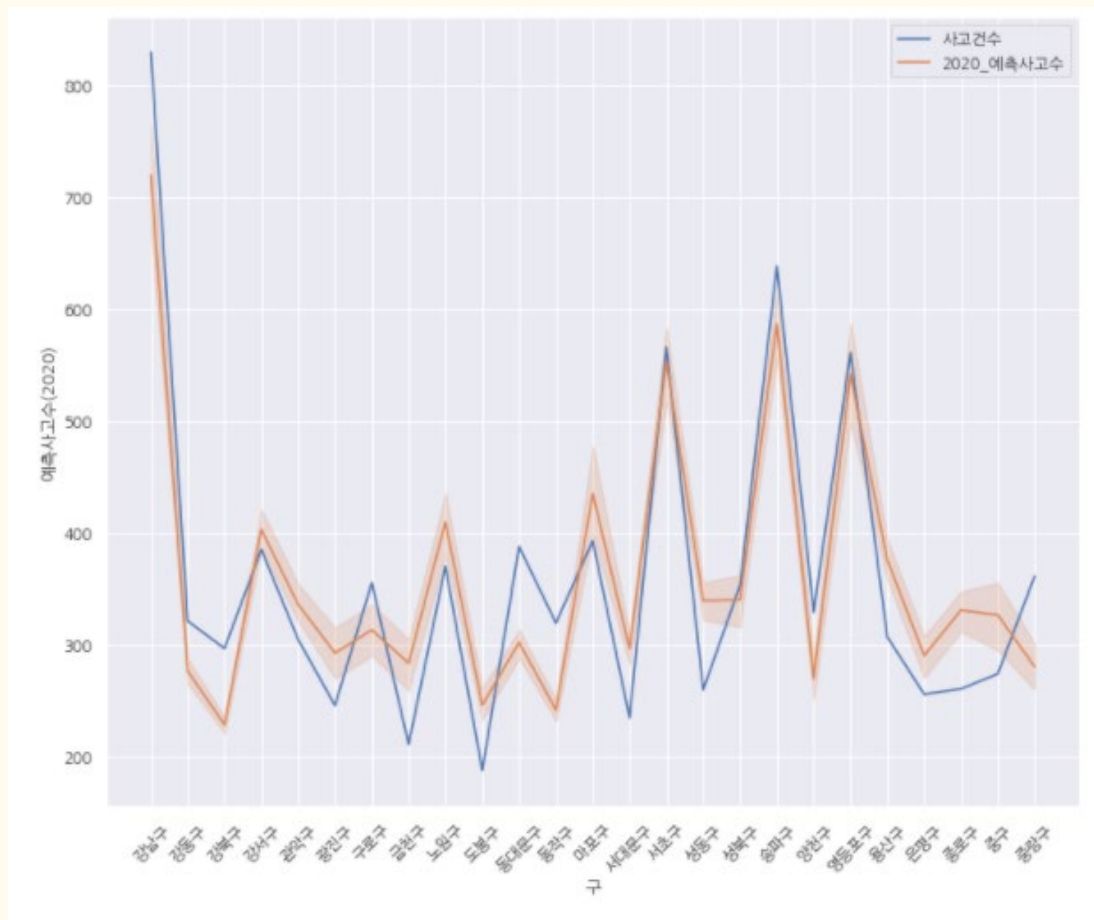
1. PCA 주성분 분석을 통한 2020년 교통사고 예측



VIF Factor	features
0	11600.7 0시
1	38416.3 10시
2	54399.5 11시
3	39443.4 12시
4	43118.9 13시
5	37636.6 14시
6	39803.5 15시
7	35614.4 16시
8	41810.6 17시
9	36565.1 18시
10	30435.9 19시
11	12230.3 1시
12	28899.2 20시
13	47144.2 21시
14	25595.6 22시
15	24287.8 23시
16	13515.6 2시
17	4935.1 3시
18	4884.0 4시
19	7417.9 5시
20	22419.5 6시
21	32651.3 7시
22	32440.2 8시
23	46759.7 9시
24	11806916.2 전체교통량



VIF Factor	features
0	5121.5 0시
1	11625.5 10시
2	11565.4 11시
3	8782.8 12시
4	11459.1 13시
5	5191.1 14시
6	8986.4 15시
7	8020.9 16시
8	4869.5 17시
9	4453.7 18시
10	3500.7 19시
11	8092.7 1시
12	3446.0 20시
13	11896.9 21시
14	12430.1 22시
15	6054.2 23시
16	8704.3 2시
17	3648.4 3시
18	1783.7 4시
19	611.0 5시
20	1267.3 6시
21	2668.9 7시
22	3699.7 8시
23	7305.3 9시

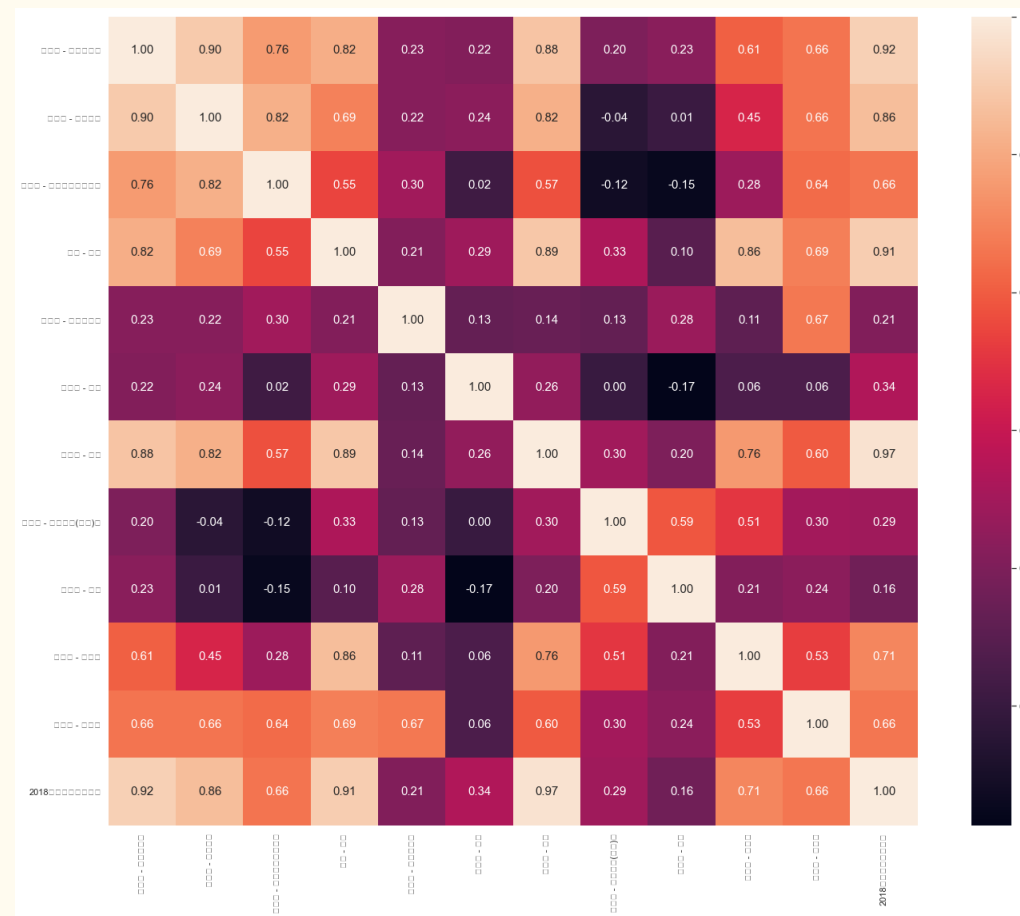
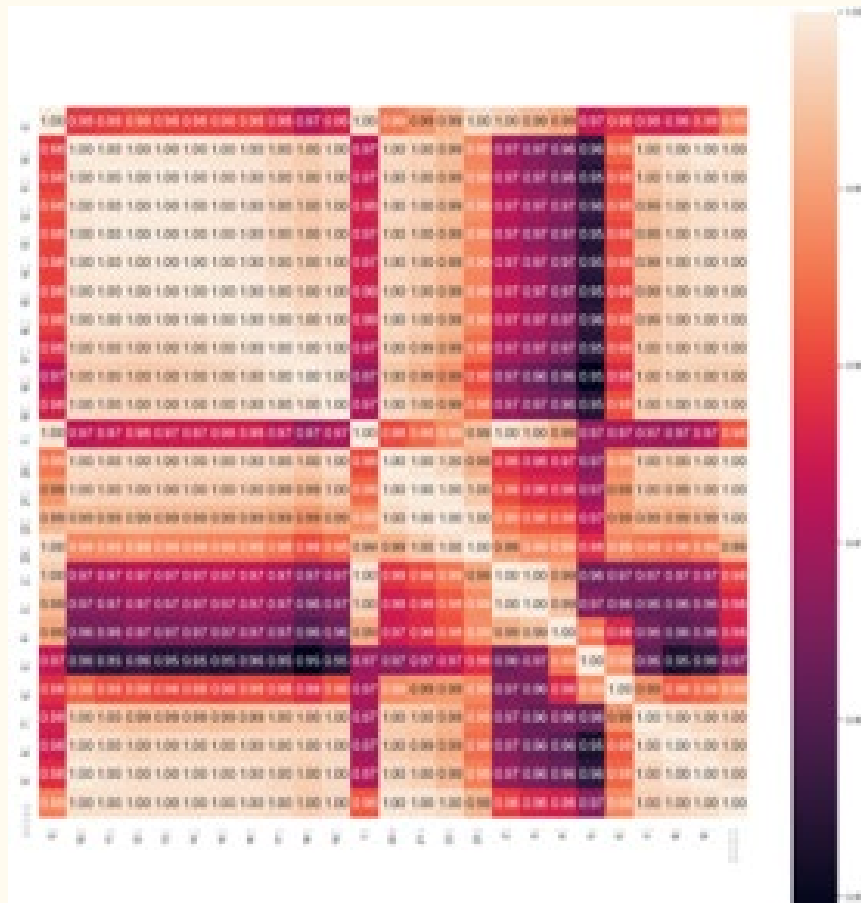


2017년 서울시 전체 교통량 & 교통사고 데이터

추론 통계

분석 과정

2. VIF를 통한 2020년 교통사고 예측



1) 월별 구의 24시간 교통량과 교통사고량

2) 구별 도로형태와 교통사고량

추론 통계

분석 결과

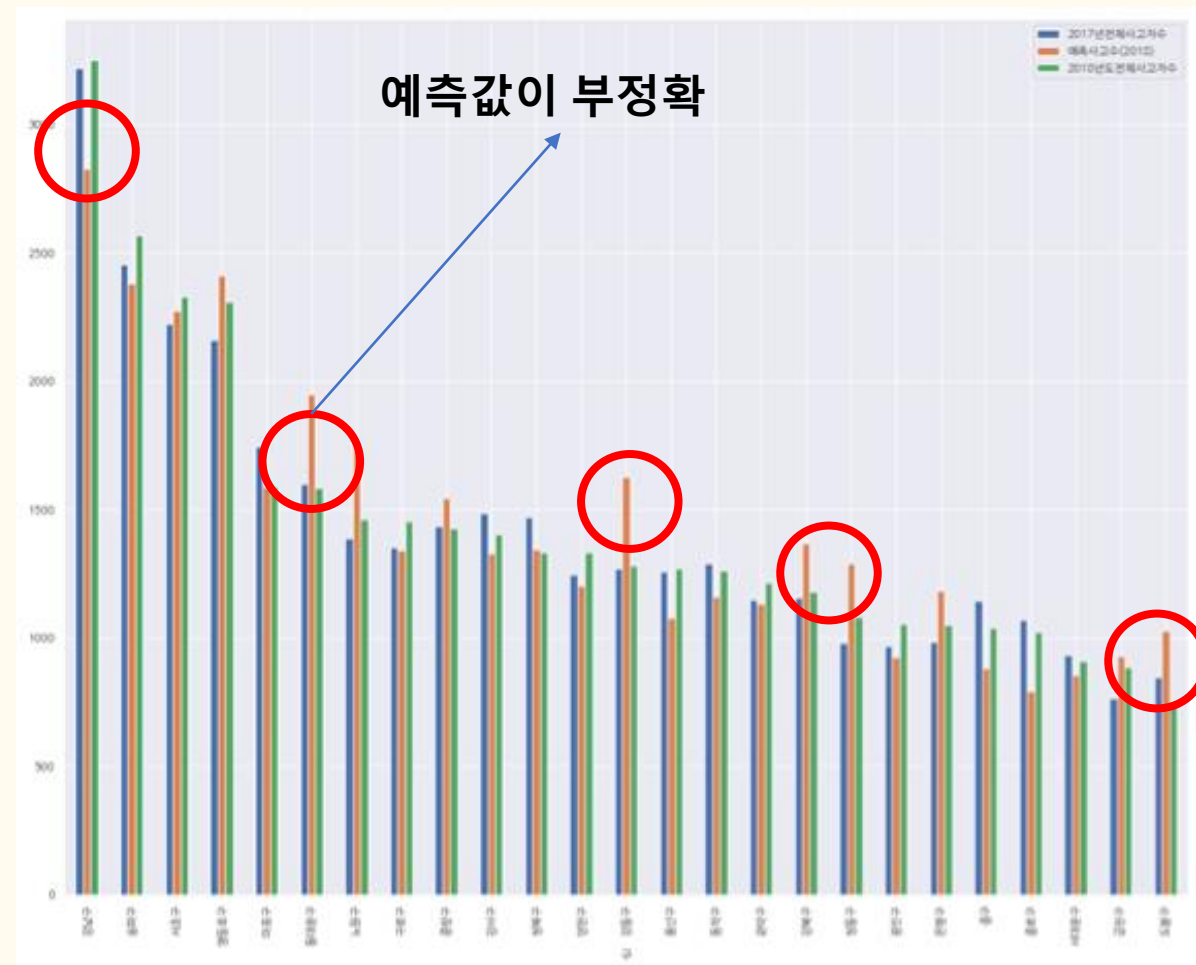
2. VIF를 통한 2020년 교통사고 예측



VIF Factor		features
0	77.9	2017년전체사고자수
1	26.1	교차로 - 교차로부근
2	7.1	교차로 - 교차로안
3	3.7	교차로 - 교차로횡단보도내
4	7.2	기타 - 기타
5	1.9	단일로 - 고가도로위
6	2.2	단일로 - 교량
7	24.3	단일로 - 기타
8	3.3	단일로 - 지하차도(도로)내
9	1.5	단일로 - 터널
10	2.1	미분류 - 미분류
11	4.5	주차장 - 주차장

같은 분류 내 높은 VIF제거
전체 중 가장 높고 낮은 VIF제거

VIF Factor		features
0	2.2	교차로 - 교차로안
1	1.2	단일로 - 교량
2	2.3	주차장 - 주차장
3	2.5	기타 - 기타



2017~2019년 교통사고 데이터

추론 통계

분석 결과

2. VIF를 통한 2020년 교통사고 예측

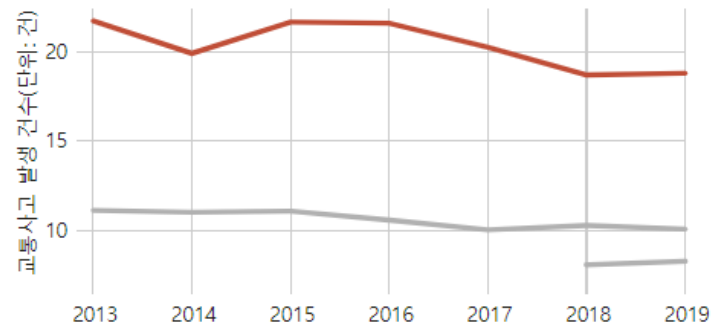


자동차 천 대당 교통사고 발생건수

18.8건

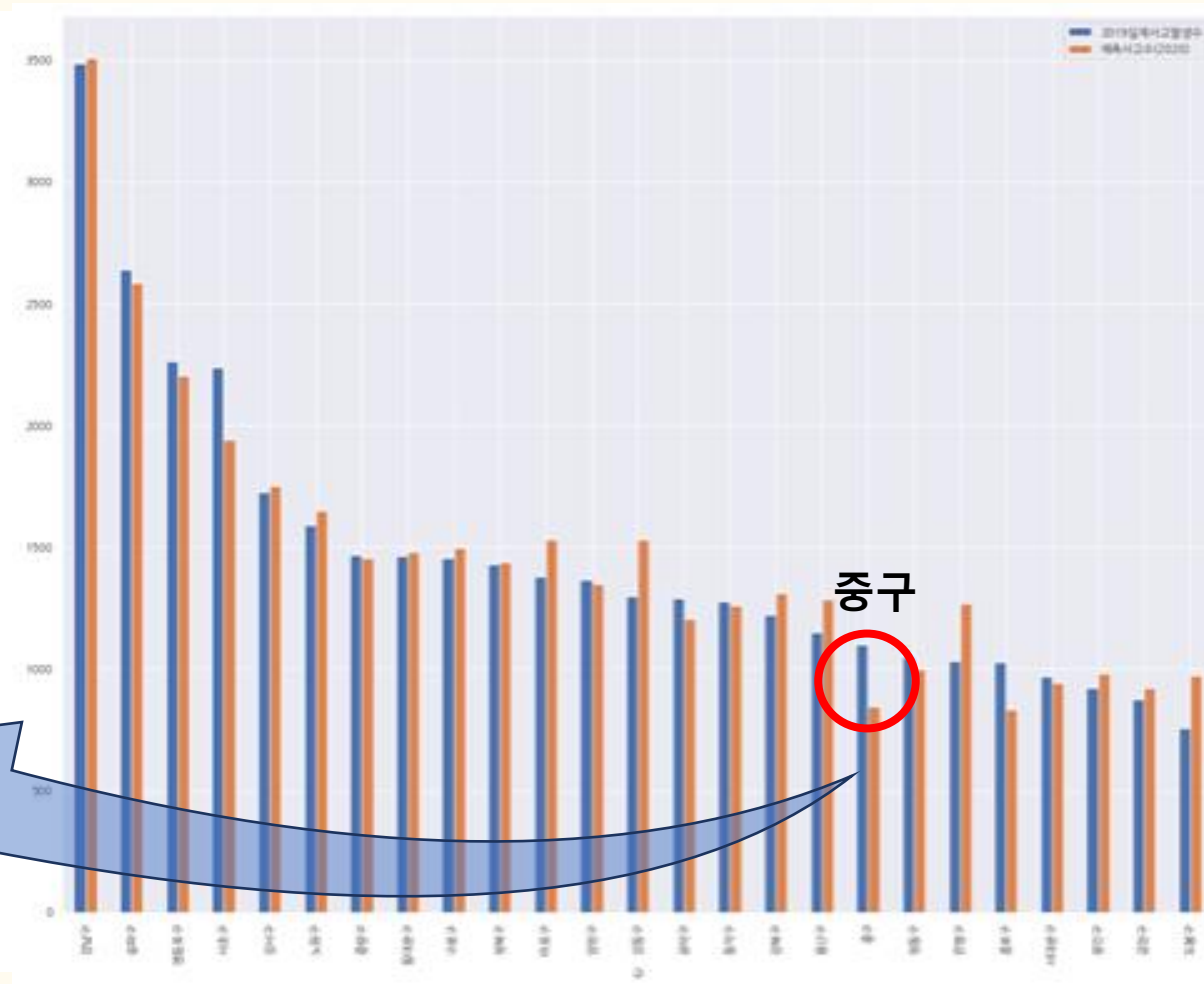
2019년 중구 자동차 천 대당 교통사고 발생건수

2019년 중구에서는 자동차 천 대당 18.8건의 교통사고가 발생하였고, 전년 대비 0.5% 증가하였고, 연평균 3.4% 감소를 보이고 있습니다.



Dataset: 국가통계포털
Source: cities Statistics

3.4%초과



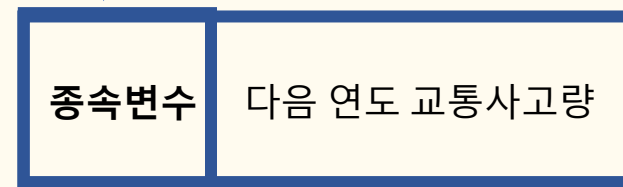
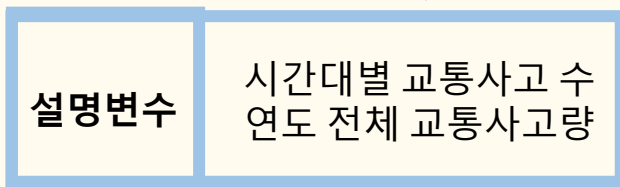
추론 통계

3. StandardScaler 사용 교통사고수 예측



→ 분석 과정 : 데이터 전처리 (시간대* 별 교통사고량)

년도	구	늦은오후	새벽	오전	오후	2018_총_사고수	2019_총_사고수	
0	2019년	강남구	623	528	922	1413	3248	3486
1	2019년	강동구	239	146	370	542	1283	1297



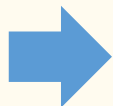
추론 통계

분석 결과

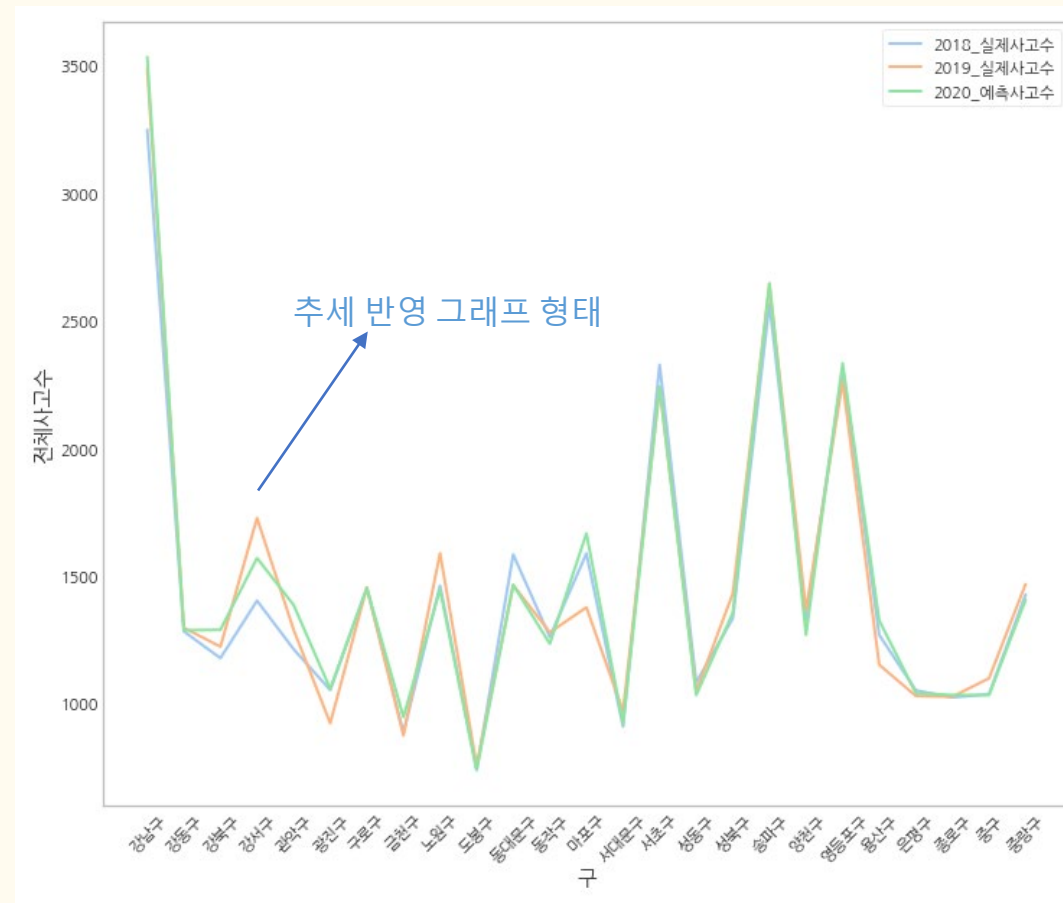
3. StandardScaler 사용 교통사고수 예측



	VIF Factor	features
0	33.7	2018_총_사고수
1	27.8	늦은오후
2	12.2	새벽
3	45.4	오전
4	62.2	오후



	VIF Factor	features
0	24.5	늦은오후
1	11.3	새벽
2	19.4	오전
3	33.6	2018_총_사고수



IV. 웹 개발 과정

- 사용기술
- 웹구조도
- 구현화면

사용 기술



→ 사용 언어/웹 프레임워크/API/라이브러리

언어

Python, CSS, JavaScript



웹 프레임워크

Flask



파이썬 라이브러리

Flask, Pandas, Numpy,
Matplotlib, Seaborn



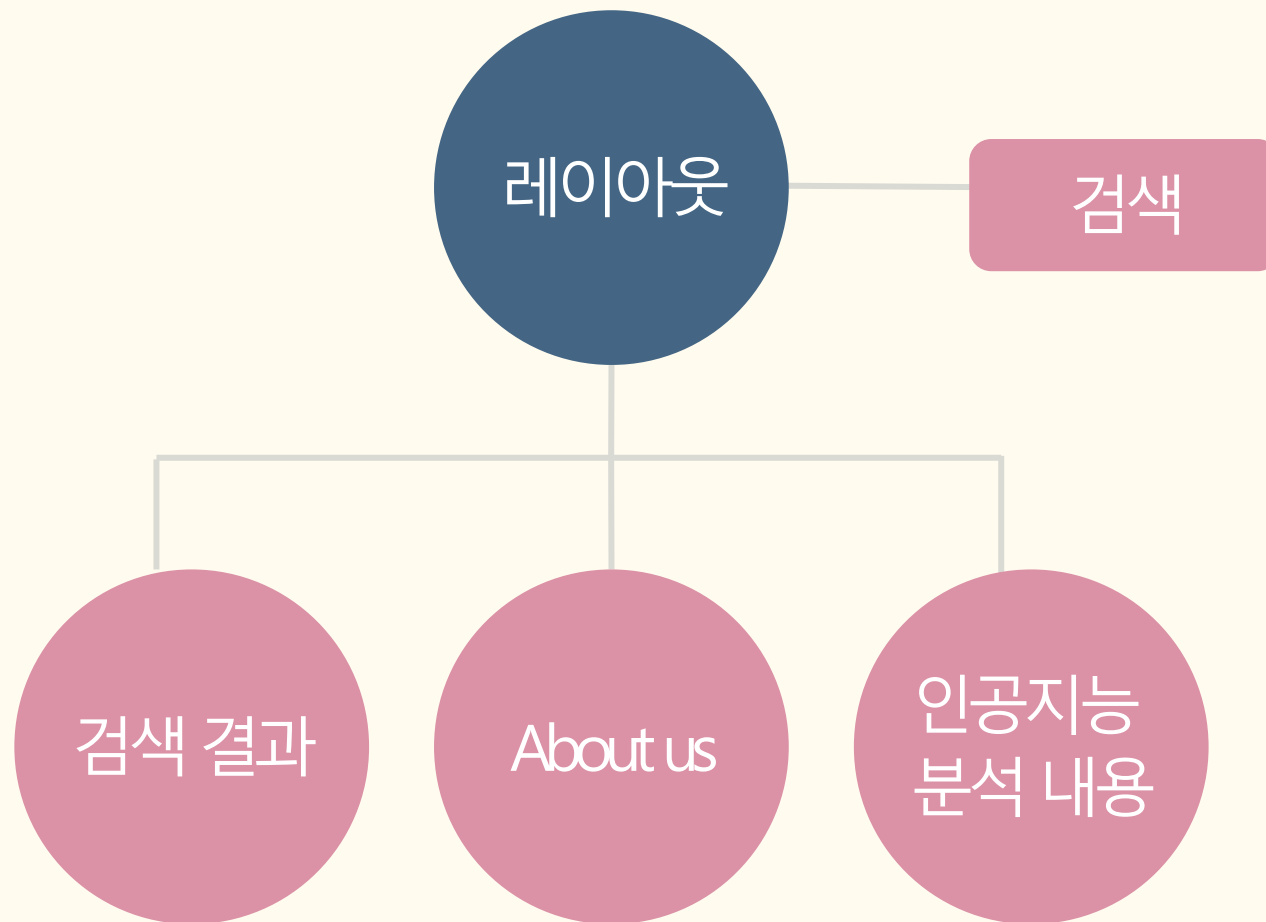
API

날씨 API
OpenWeatherMap



웹 구조도

→ 정보 구조도



웹 구조도

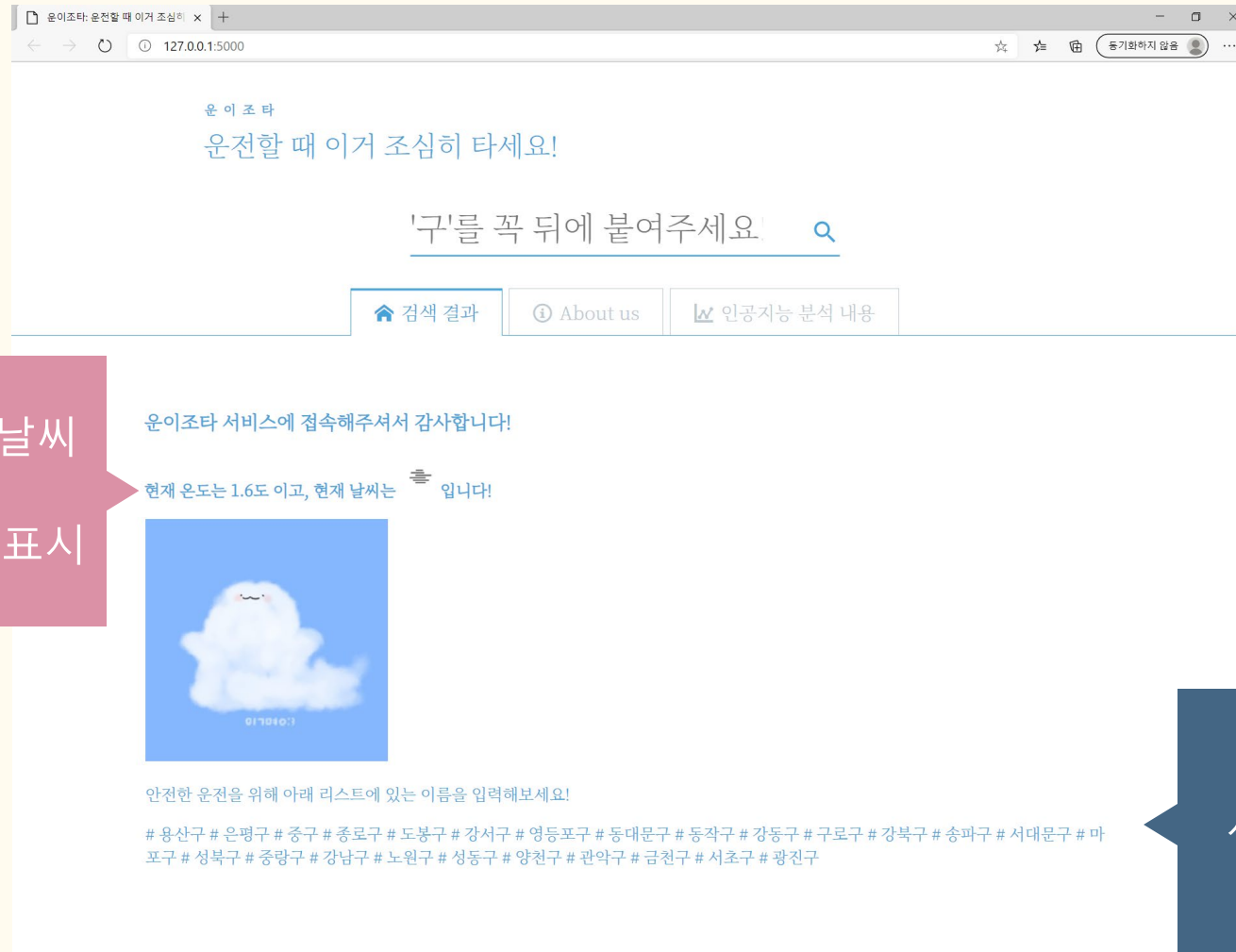
→ 파일 구조도



GoodLuck			비고
파일	main.py		메인 파일
	visualization.py		데이터 시각화에 필요한 함수 모듈
	서울 통합 교통사고.csv		데이터 분석에 필요한 csv 파일
폴더	templates	main.html	각 탭의 내용이 들어가는 파일
		layout.html	헤더와 탭 등 반복되는 레이아웃 파일
	static	icomoon	아이콘 폰트에 필요한 폴더
		Img	이미지 폴더
		responsive_weather.js	날씨를 API로 받아오기 위한 JS파일
		search.css / search.js	검색 애니메이션 기능에 필요한 CSS/JS 파일
		style.css	웹의 전반적인 디자인에 필요한 CSS파일
		Tabs.js	Tab 기능에 필요한 JS파일

구현 화면

메인 화면 - 검색 전



Jinja2 템플릿 엔진을
이용하여 헤더 상속

OpenWeatherMap 날씨
API를 이용하여
당일 서울의 날씨를 표시

데이터를 바탕으로
시각화 가능한 구들을
유동적으로 표시

구현 화면

검색



운이조타: 운전할 때 이거 조심해

127.0.0.1:5000

동기화하지 않음

운 이 조 타

'구'를 꼭 뒤에 붙여주세요!

▼ 🔍

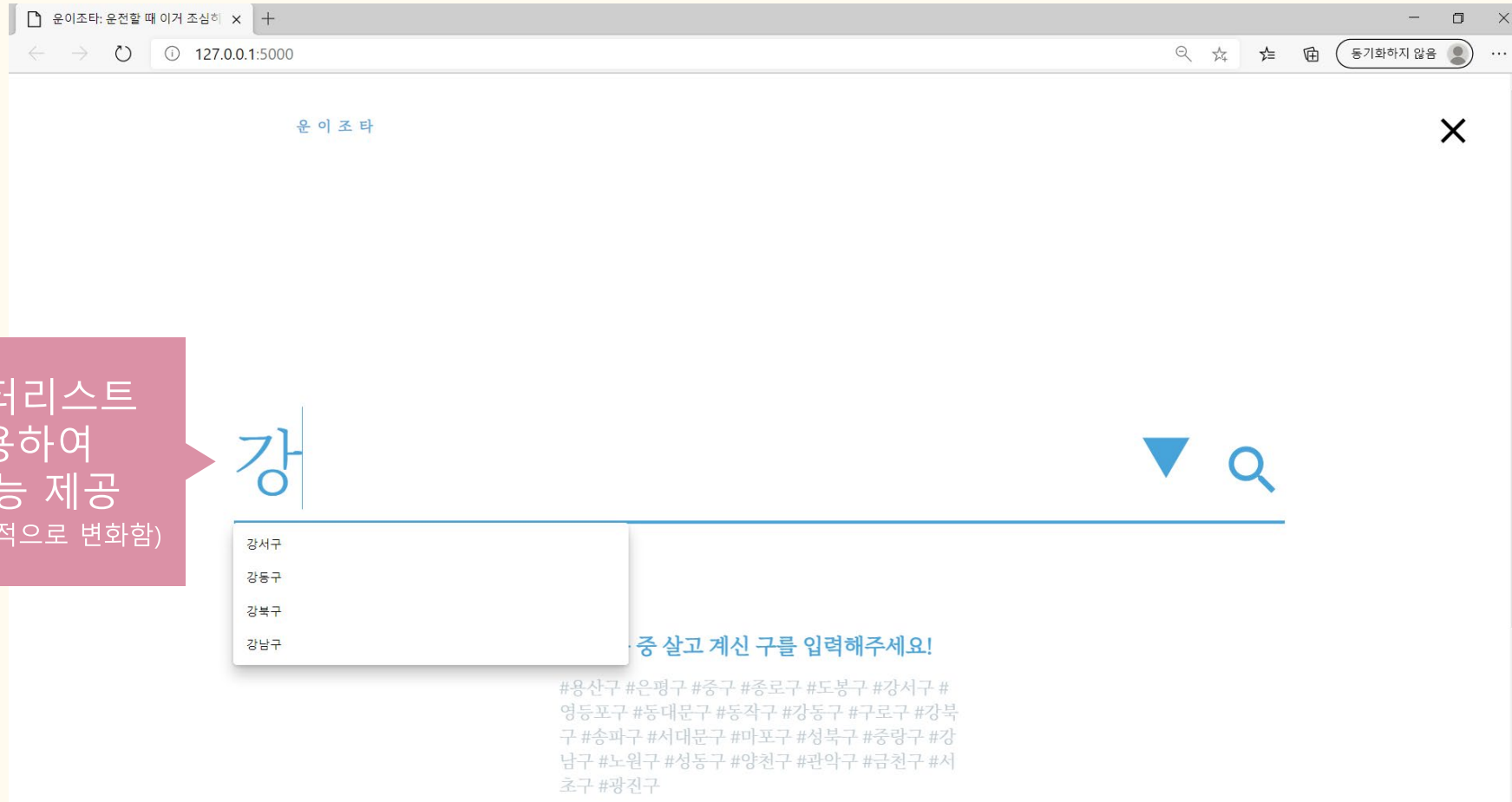
데이터를 바탕으로
시각화 가능한 구들을
유동적으로 표시

→ 다음 중 살고 계신 구를 입력해주세요!

#용산구 #은평구 #중구 #종로구 #도봉구 #강서구 #영등포구 #동대문구 #동작구 #강동구 #구로구 #강북구 #송파구 #서대문구 #마포구 #성북구 #중랑구 #강남구 #노원구 #성동구 #양천구 #관악구 #금천구 #서초구 #광진구

구현 화면

→ 검색 - 자동 완성



HTML5 데이터리스트
기능을 이용하여
자동완성 기능 제공
(데이터 리스트는 유동적으로 변화함)

구현 화면

→ 검색 예시: '강남구'



각 구별 데이터 분석
내용 시각화

구현 화면



→ 검색 예시: '처인구' - 서비스 하지 않는 지역

문이조타: 운전할 때 이거 조심하 x

127.0.0.1:5000/?search=처인구

문 이 조 타
운전할 때 이거 조심하 타세요!

처인구

검색 후 아이콘
변화

🔍 검색 결과 About us 인공지능 분석 내용

:3

아직 서비스를 하지 않고 있는 지역이거나 잘못 된 이름을 입력하신 것 같습니다. 아래 리스트에 있는 이름만 입력해주세요!

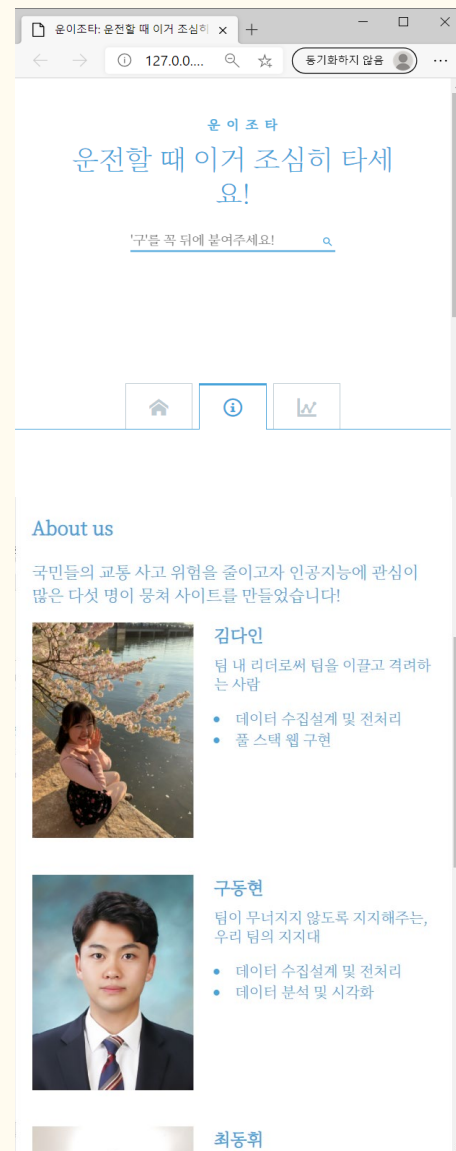
#용산구 #은평구 #중구 #종로구 #도봉구 #강서구 #영등포구 #동대문구 #동작구 #강동구 #구로구 #강북구 #송파구 #서대문구 #마포구 #성북구 #중랑구 #강남구 #노원구 #성동구 #양천구 #관악구 #금천구 #서초구 #광진구

데이터를 바탕으로
시각화 가능한 구들을
유동적으로 표시

구현 화면

About us

웹 크기에 따른 반응형
웹 사이트

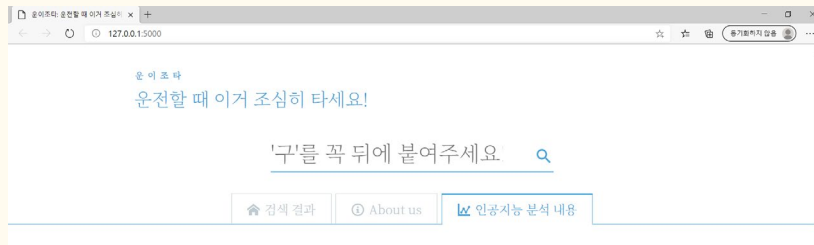


탭 클릭을 통한 페이지
이동



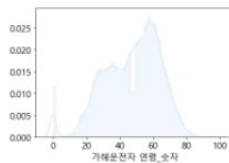
구현 화면

인공지능 분석 내용



인공지능 분석

서울시 교통사고 데이터를 2017년부터 2019년까지 분석해보았습니다.



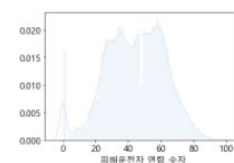
그래프1

2017년 ~ 2019년(3년)

서울시 교통사고 가해 운전자 연령

*0세: nan, 공백 데이터의 합

*98세: 98세+98세 이상



그래프2

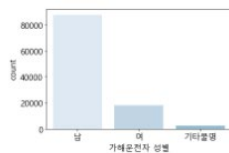
2017년 ~ 2019년(3년)

서울시 교통사고 피해 운전자 연령

*0세: nan, 공백 데이터의 합

*98세: 98세+98세 이상

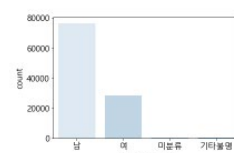
가해 운전자 60대, 70대
피해 운전자 20대 ~ 70대까지 고르게 분포



그래프3

2017년 ~ 2019년(3년)

서울시 교통사고 가해 운전자 성별 분포

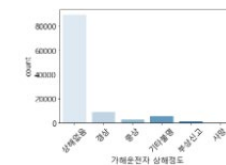


그래프4

2017년 ~ 2019년(3년)

서울시 교통사고 피해 운전자 성별 분포

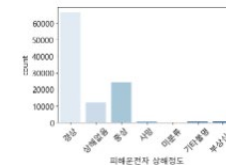
가해운전자/피해운전자
두 가지 경우 모두
남자의 성별이
압도적으로 많음



그래프5

2017년 ~ 2019년(3년)

서울시 교통사고 가해 운전자 상해정도



그래프6

2017년 ~ 2019년(3년)

서울시 교통사고 피해 운전자 상해정도

가해운전자

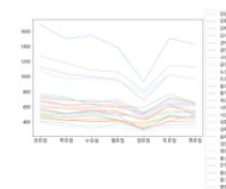
"상대없음"이 독보적으로

높은 수치를 보임

피해운전자

"경상"은 독보적으로 높은 수치를 보임

"중상" 역시 적은 수치는 아님



그래프7

2017년 ~ 2019년(3년)

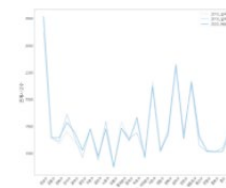
서울시 구의 요일 별 실제 사고 건수

2017년 + 2018년 + 2019년

총 3년간의 실제 사고 건수

<=

요일 별 시각화



그래프8

서울시 구의 시간대 별 사고 건수로

2020년도 사고 건수 예측하기

2018년도 구 별 전체_사고 건수

+

2019년도 구 별 전체_사고 건수

=

2020년도 구 별 사고 건수 예측

데이터 분석 내용 그래프
이미지와 분석 내용들

V. 결론

- 최종 결론
- 한계
- 향후 과제

최종 결론

→ 웹 개발 팀



바로 웹 작업에 착수하면 될 것이라 생각 **BUT** “어떤 데이터를 어떻게 분석해서 어떤 결과를 낼 것인지” 구상 과정이 선행 되어야 웹 개발이 가능 => 데이터 분석 팀과 협업하여 데이터 분석 과정 및 웹 UI 결정

데이터 분석 팀이 보내준 코드를 그대로 붙여넣기 하면 될 것이라 생각 **BUT** 웹 서비스에 맞게 코드 최적화가 요구됨 + 빅데이터로 인한 느린 웹 로딩 속도 문제 => 차차 해결이 요구됨



데이터 분석 팀 →

교통사고에 영향을 주는 요인을 **날씨, 시간, 도로형태**로 설정

=>약 10만건의 공공데이터 직접 전처리 및 분석

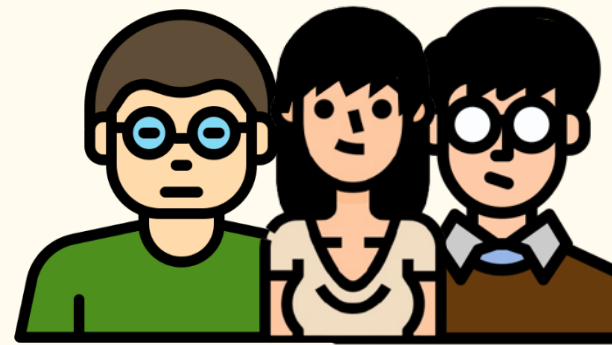
=>데이터 시각화, 선형회귀분석, PCA 주성분 분석 등을 활용

날씨 데이터=> 시각화 =>전반적으로 잘 진행됨

시간, 도로형태 데이터=> 미래 교통사고수 예측=>예상치 못한 문제 많이 발생

하지만 결과적으로,

standard scaler로 데이터 변환=> 선형회귀분석에 적합시키고 교통사고 수 예측 성공



한계

— 웹 개발 팀

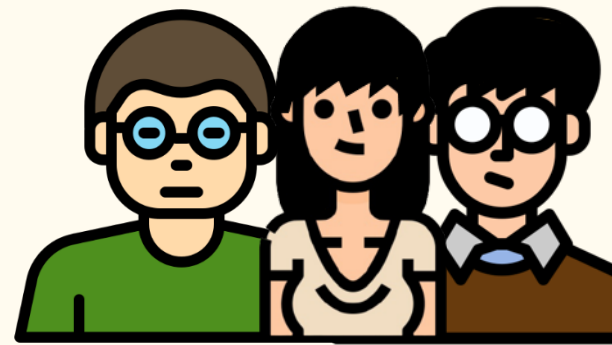


- 검색 시 구를 빼고 검색해도 검색되게 하려 했으나 기간 안 제출이 불가할 것 같아 제외
- 웹 배포를 위해 AWS에 서버를 올리는 것까지는 성공했으나 도메인에 올리는 부분에서 에러 발생
- 데이터 과다로 인해 그래프와 범례가 간혹 제대로 로딩되지 않음



데이터 분석 팀 —

- 날씨, 교통상태 등등 다양한 변수 => 사고원인에 존재하는 주성분을 찾기 힘들
- VIF => 사고 원인은 복합적으로 존재=> 10이내여야 하나 20을 넘는 수치들
- VIF 낮추는 것 시도 BUT 예측치 error범위에서 많은 차이를 보임
- 이를 확인하기 위해서 다양한 방법으로 데이터 전처리와 분석을 진행
- => 주성분 요소를 PCA로도 확실히 뽑아내지 못했으며, kNN알고리즘, 군집화 머신러닝 또한 유효한 결과를 확인 할 수 없었음
- 지도 학습, 비지도 학습의 차이점을 알기 위해서 데이터를 분석하고 비교
- => 결과적으로 선형회귀분석으로 예측하는 지도학습을 선택



향후 과제

— 웹 개발 팀

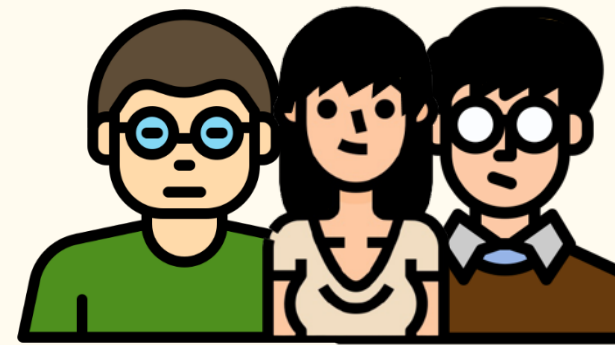


- '강남구' 대신 '강남'으로도 검색 가능하게 하기
- 웹 배포
- 데이터 과다로 인한 느린 로딩 속도와 에러 해결하기
- 데이터베이스에 데이터 업로드 해서 불러오기
- 웹 새로 고침 시 매번 데이터 새로 불러오는 문제 해결



데이터 분석 팀 —

- 웹 서비스 속도향상을 위한 데이터 전처리 과정 최소화
- 맞춤 데이터 불러오기
- VIF값을 낮게 가져오면서 정확도 높여서 예측하기
- 실시간으로 들어오는 데이터를 받아서 지역별 교통사고수 예상하기
- 서울시 이외의 지역 맞춤별 데이터 제공



감사합니다

