

데이터 분석 이론 및 실습 프로젝트 보고서

눈, 비, 사고



백석대학교

2022.06. 담당교수 : 강윤희 교수님

백석대학교

컴퓨터공학부

멀티미디어학과

by 20172543 우정호

area 데이터 분석

about 강수, 교통사고

1. 서론

2. 본론

3. 결론

4. 마무리

프로젝트 보고서

◎ 1. 서론

○ 1.1 개요

교통사고, 그 중에서도 강수 시 시야 미확보, 수막현상 등을 원인으로 발생하는 교통사고의 위험성은 이전부터 강조해왔으며 주의를 주고 있지만 우천 시 발생하는 교통사고의 치사율은 맑은 날에 비해 여전히 37.5% 높다. 경기도에서 발생하는 사망교통사고와 강수 간의 상관관계를 규명하기 위한 가설의 설정과 검증의 과정을 통해 프로젝트를 진행하며 사고가 다발하는 구역에 관한 확장분석을 함께 진행한다. 이는 강수를 원인으로 발생하는 사망교통사고에 대비하여 적절한 조치를 취하였는지, 그렇지 않다고 한다면 도로보수, 단속 강화 등의 조치를 취하여야할 지역을 특정하기 위함이다.

○ 1.2 가설

□ 가설(1). 강수날에 일반 차량보다 화물차의 사고가 증가할 것이다.

- 화물차는 기본적으로 일반 소/중형차량과 비교해 무거워 즉각적인 조작이 어렵고 차고가 높아 근거리 시야 확보에 불리하다. 강수날에 비 혹은 눈으로 인한 수막현상 발생시 그 위험도가 더욱 증가 할 것이고 시야확보 역시 더욱 어려움이 있을 것이기 때문에 강수날에 화물차의 사망사고 비율은 증가할 것이고 일반차량의 사망사고 비율은 감소 할 것이다.

□ 가설(2). 강수량이 높을수록 사고가 많이 발생할 것이다.

- 강수량이 많아질수록, 수막현상으로 인한 미끄러짐 사고가 발생할 확률이 높아져 더 많은 사고가 발생할 것이다.

□ 가설(3). 강수날 교통사고는 야간에 더 많이 발생했을 것이다.

- 강수가 있으면 주간에도 시야 확보가 쉽지 않으나 야간에는 시야를 확보가 더욱 쉽지 않기 때문에 강수날에는 야간에 더 많은 사고가 발생 할 것이다.

□ 가설(4). 강수날 교통사고는 안전운전을 하지 않아서 발생한 사고가 가장 많을 것이다.

- 강수날에는 시야 확보가 맑은날에 비해 어렵고 수막현상으로 인해 브레이크가 잘 듣지 않을 것이기 때문에 안전운전을 하지 않아서 발생하는 사고가 많을 것이다.

○ 1.3 확장분석

사고가 다발하는 구역을 지도상에 표시하는 경우, 육안으로 표시된 마커가 밀집된 장소를 특정할 수 있습니다. 이를 통해 해당 구역 및 지점은 우천 시 사고가 다발하는 장소라는 결론에 도달할 수 있으며 해당 방법으로 2016년과 2020년의 자료를 비교 분석하는 경우 여전히 사고가 다발하는 위험 지역과 새롭게 사고가 다발하는 구역을 특정할 수 있습니다.

◎ 2. 본론

○ 2.1 데이터 수집

• 활용자료

1. ASOS 종관 기상 관측 데이터[기상청], 기상자료개방포털(<https://data.kma.go.kr/>)
2. 사망교통사고 데이터, [도로교통공단], 공공데이터포털(<https://www.data.go.kr/>)

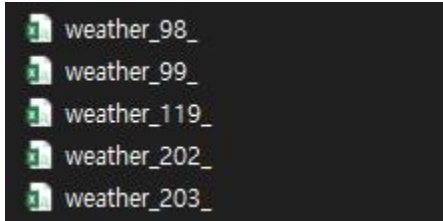
• ASOS 종관 기상 관측 데이터 수집 방법

- 지역과 일자 파라미터를 알맞게 설정하고 데이터 분석에 필요한 정보만을 http 방식으로 openAPI 인증키를 이용하여 데이터를 JSON 형식으로 추출한 후, CSV 형식으로 저장함

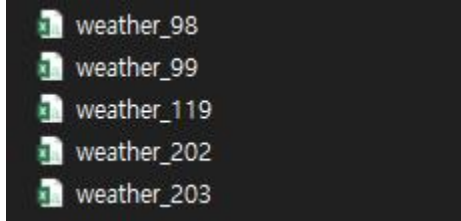
- 수도권기상청에서 기상관측을 주관하는 8개 지역 중 특별시인 서울과 광역시인 인천, 외곽지역인 백령도를 제외한 동두천, 파주, 수원, 양평, 이천 등 5개 지역의 기상 데이터를 수집

• ASOS 종관 기상 관측 데이터 수집 결과

(2016년)



(2020년)



(추출한 CSV)

tm	stnNm	sumRn
2016-01-01	수원	
2016-01-02	수원	
2016-01-03	수원	

- 사망교통사고 데이터 수집 방법
- 지역과 일자 파라미터를 알맞게 설정하고 데이터 분석에 필요한 정보만을 http 방식으로 openAPI 인증키를 이용하여 데이터를 JSON 형식으로 추출한 후, CSV 형식으로 저장함
- 자료의 신뢰성과 분석의 양질을 높이기 위해 수집 지역이었던 5개 지역과 각각 인접해 있는 가평, 고양, 용인, 포천, 화성 등 추가 5개 지역을 포함하여 총 10개 지역의 사고 데이터를 수집

• 사망교통사고 데이터 수집 결과

(2016년)

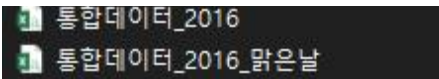
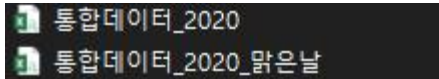
(2020년)

(추출한
CSV)

ocrrnc_dt	dth_dnv_c	dght_cd	injsps_cnt	acc_ty_lcla	acc_ty_mls	acc_ty_cd	aslt_vtr_cd	road_frm_l	road_frm_r	wrngdo_is	dmge_isrt	dmge_isrt	lo_crd	la_crd
2020010111	1	1	1	3	32	34	5	1	5	11 ##	##		127.626	37.10341
2020010319	1	2	1	1	15	5	5	2	8	1	12	12	127.4038	37.25528
2020010519	1	2	1	3	32	35	1	1	5	1 ##	##		127.5212	37.10322
2020011120	1	2	1	1	11	1	5	2	8	1	12	12	127.4331	37.28082

○ 2.2 데이터 전처리

- 기상 데이터와 사고 데이터의 통합
- 기상 데이터에서 강수가 있었던 날의 데이터만을 추출하고, 해당 강수날의 일자를 이용하여 사고 데이터 중 강수가 발생하였던 일자의 데이터만을 분류, CSV로 저장 시 강수 데이터 중 일자와 지역명, 강수량 항목에 데이터 분석 시 필요한 사고 데이터 항목을 추가한 통합데이터를 저장
- 해당 과정을 2016년과 2020년 데이터를 대상으로 진행

(2016년)		(2020년)	
---------	---	---------	---

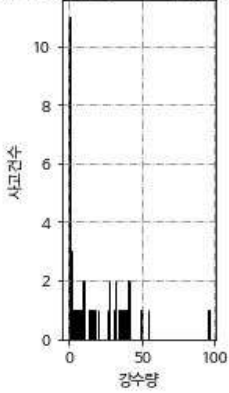
날짜	지역명	강수량	사상자수	주요구분	노부상자수	사고유형A	사고유형B	사고유형C	가해자법도	로형태도	로형태C	가해당사자	피해당사자	경도	위도
0 20160101 수원		0	1	2	1	2	23 Z1		5	1	5	1	3	127.023	37.23059
1 20160101 파주		0	1	1	1	2	22	22	4	1	5	8	3	126.8135	37.74139
2 20160105 파주		0	1	2	4	2	22	22	2	1	5	1	1	126.8788	37.63173
3 20160105 파주		0	1	2	2	2	22	22	3	2	6	1	3	126.746	37.66205
4 20160106 이천		0	1	1	1	3	34	31	5	1	2	7 ##		127.457	37.33464
5 20160106 이천		0	1	2	4	2	22	22	2	1	5	1	2	127.2032	37.14944
6 20160107 양평		0	1	2	1	1	11	1	5	1	5	3	12	127.3584	37.81495

날짜	지역명	강수량	사량자수	주야구분	부상자수	사고유형C	사고유형중	사고유형중	사고유형중	가해자법	도로형태	도로형태	가해당사자	피해당사자	경도	위도
0 20200106	이천	4.4	1		2	1	2	23	23	5	1	5	5	3	127.1141	37.28406
1 2020106	동두천	2.5	1		2	1	1	13	3	5	1	5	1	12	127.3141	38.00234
2 20200107	동두천	49.1	1		1	2	1	11	1	7	1	5	1	12	127.2744	38.09136
3 20200119	수원	0.4	1		2	4	2	24	25	2	1	5	1	1	126.9524	37.19507
4 20200119	양평	0.7	1		1	1	3	31	32	5	2	8	3	##	127.5168	37.81089
5 20200127	이천	2.1	1		2	1	1	11	1	5	2	6	1	12	127.1667	37.24338

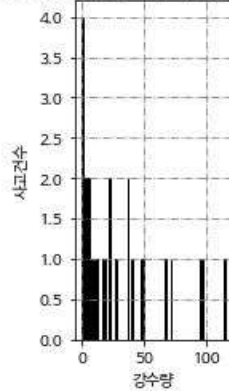
<2016년 강수량-사망자수 산포도>

- 2016년과 2020년 강수량-사고발생횟수 그래프를 살펴보면 앞서 말했던 것과 같이, 강수량이 적을수록 많은 사고가 발생했음을 확인 할 수 있음
- 요약: ① 강수량이 많은날 발생한 사고의 경우 많은 사망자가 나오는 사고로 이어지는 것은 아님 ② 사고는 강수량이 적은 날에 집중적으로 발생하였음 ③ 강수량이 적을수록 사고횟수가 많아지는 경향을 보임

2016년 강수량-사고발생 상관관계

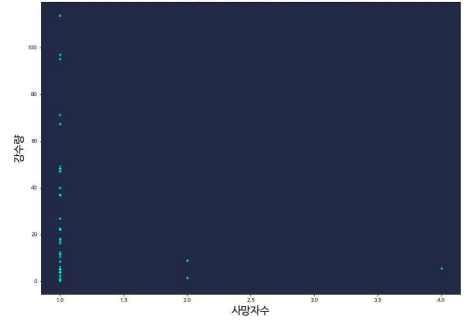


2020년 강수량-사고발생 상관관계



<2016년과 2020년 강수량-사고횟수 그래프>

2020년 강수량-사망자수 상관관계



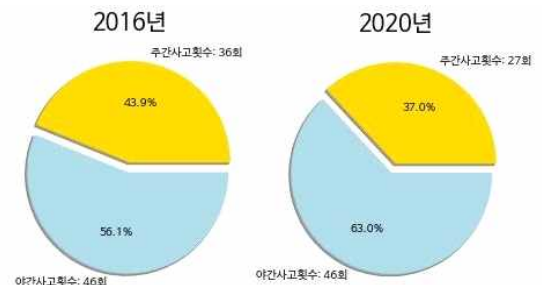
<2020년 강수량-사망자수 산포도>

■ 가설(3). 강수날 교통사고는 야간에 더 많이 발생할 것이다.

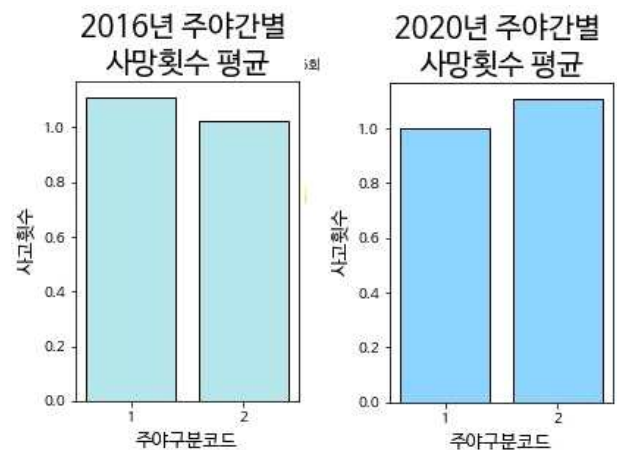
- 강수날 주야간에 발생하는 사고의 분석을 위해 주야간 별 사고횟수와 사망자수의 평균을 이용하여 그래프를 출력, 비교 분석을 진행하였음
- 2016년과 2020년 주야간 사고횟수 비교를 통해 두 해 모두 강수날에 주간보다 야간에 더 많은 사고가 발생했음을 확인할 수 있음
- 2016년과 비교해 2020년에는 야간사고횟수는 동일한 반면, 주간사고횟수는 감소한 것을 확인할 수 있음
- 2016년 주야간별 사망자수 평균은 주간이 약1.1명, 야간이 약 1명으로 주간 평균이 더 높은 것을 알 수 있음
- 2020년 주야간별 사망자수 평균은 주간이 약 1명, 야간이 약1.1명으로 야간 평균이 더 높은 것을 알 수 있음
- 요약: ① 강수날에는 주간보다 야간에 더 많은 사고가 일어났음을 확인하였음 ② 강수날 발생하는 주간사고의 횟수가 줄어들고 있음을 확인하였음 ③ 주야간별 사망자수 평균은 큰 차이가 없으나 이전과 비교해 주간의 사망자수 평균은 줄어들고 야간의 사망자수 평균은 늘어나고 있음을 확인하였음

주야코드	주야
DGHT_CD	DGHT_NM
1	주간
2	야간

<주야코드 표>



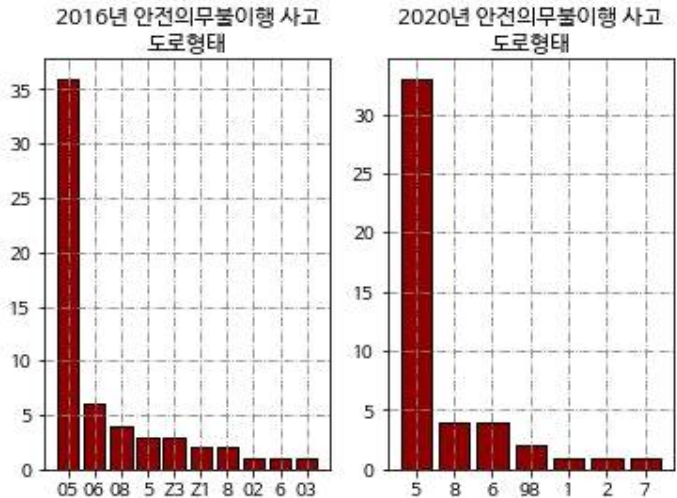
<2016년과 2020년 주야간 사고횟수 비교>



<2016년과 2020년 주야간별 사망자수 평균>

■ 가설(4). 강수날 교통사고 안전운전을 하지 않아서 발생한 사고가 가장 많을 것이다.

- 가설의 검증을 위하여 가해자법규위반 항목을 시각화하여 분석을 진행하였으며, 사고 중 가장 많은 법규위반항목에 해당되는 항목에 대해 어떠한 도로형태에서 사고가 발생하는지를 분석하였음
- 2016년 법규위반 항목을 확인하면 가해자법규위반5 즉, 안전운전 의무 불이행 항목이 가장 많은 비중을 차지함을 알 수 있었음
- 2020년 법규위반 항목 역시 가해자법규위반5 즉, 안전운전 의무 불이행 항목이 가장 많은 비중을 차지함을 알 수 있었음
- 2016년과 2020년 모두 2순위의 항목이 가해자법규위반 3에 해당되는 신호위반 항목임을 알 수 있었음
- 강수날 발생한 안전운전 의무 불이행 항목에 해당되는 사고 가운데 2016년과 2020년 모두 도로형태5 즉, 기타 단일로에서 발생한 사고가 가장 많았음을 알 수 있었음
- 강수날 발생한 안전운전 의무 불이행 항목에 해당되는 사고 가운데 2016년과 2020년 모두 그 순위는 다르지만 도로형태6,8 즉 교차로 내와 교차로 부근에서의 사고가 각각 2,3순위를 기록하였음을 알 수 있음
- 요약: ① 강수날 발생한 사망사고는 안전운전 의무 불이행 항목을 위반하여 발생한 사고가 가장 많았음 ② 강수날 발생한 사망사고 가운데 신호위반이 원인으로 발생한 사고 역시 많았음을 확인 할 수 있음 ③ 강수날 발생한 안전운전 의무 불이행 사망사고는 기타 단일로 부근에서 가장 많이 발생하였고 교차로 내와 근처에서 발생한 사고가 그 뒤를 이었음을 확인함



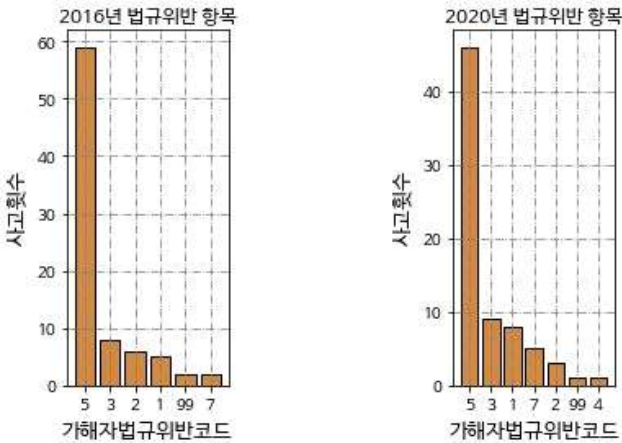
<안전의무 불이행 항목 위반 사고 도로형태 비교>

가해자법규위반코드	가해자법규위반
ASLT_VTR_CD	ASLT_VTR_NM
01	과속
02	중앙선 침범
03	신호위반
04	안전거리 미확보
05	안전운전 의무 불이행
06	교차로 통행방법 위반
07	보행자 보호의무 위반
99	기타

<가해자법규위반코드 표>

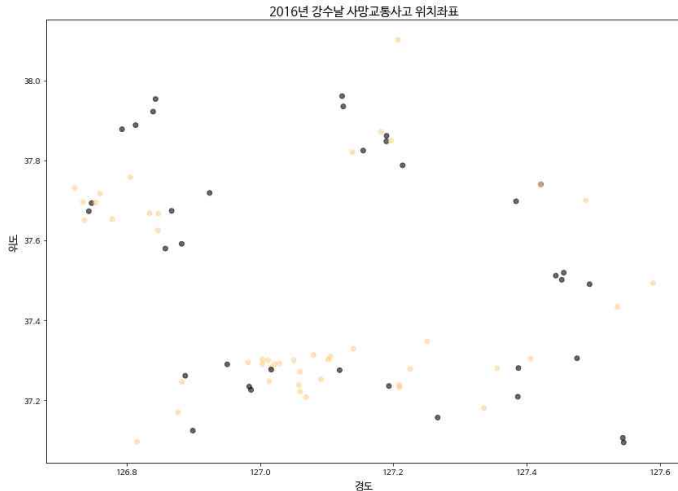
도로형태코드_대분류코드	도로형태코드	도로형태_대분류	도로형태
ROAD_FRM_LCLAS_CD	ROAD_FRM_CD	ROAD_FRM_LCLAS_NM	ROAD_FRM_NM
01	01	단일로	터널안
01	02	단일로	교량위
01	03	단일로	고가도로위
01	04	단일로	지하차도(도로)내
01	05	단일로	기타단일로
01	Z1	단일로	횡단보도상
01	Z2	단일로	횡단보도부근
02	06	교차로	교차로내
02	07	교차로	교차로횡단보도내
02	08	교차로	교차로부근
04	10	철길건널목	철길건널목
05	98	기타	기타
99	99	불명	불명
Z3	Z3	기타/불명	기타/불명
##	##	없음	없음

<도로형태코드 표>

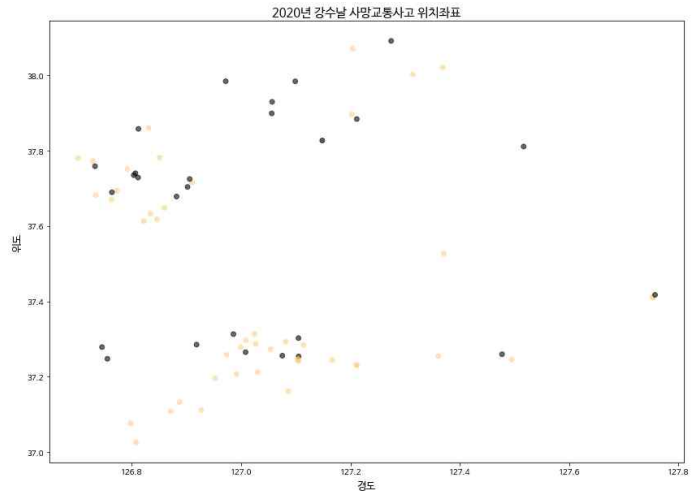


<2016년과 2020년 법규위반 항목 비교>

○ 2.4 데이터 분석 - 확장 분석

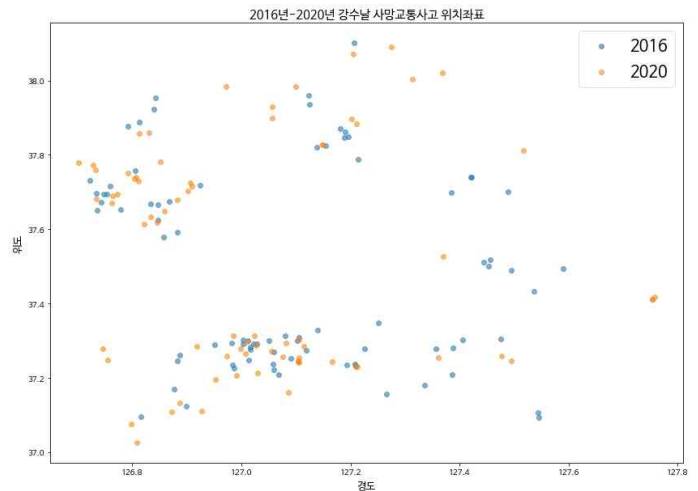


<2016년 강수날 사망교통사고 위치좌표>

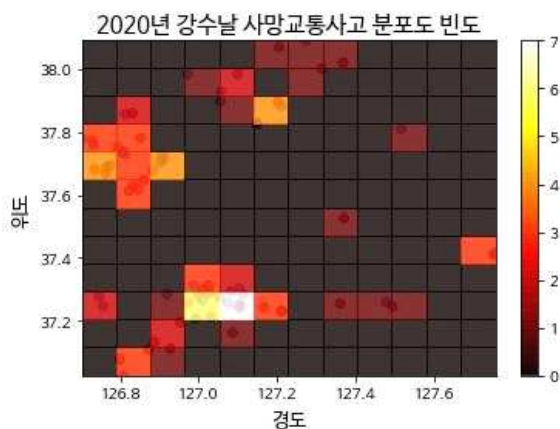
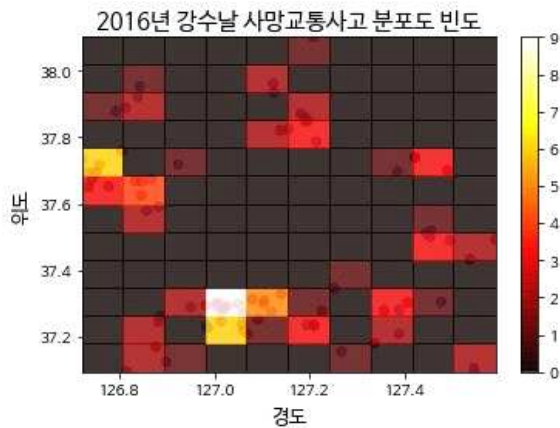


<2020년 강수날 사망교통사고 위치좌표>

- 1차적으로 2016년과 2020년 강수날에 발생한 사망교통사고의 위치좌표를 산포도를 이용해 시각화를 진행함
- 2차적으로 2016년과 2020년의 산포도를 동시에 출력하였고 육안으로 확인한 결과 37.2~37.3, 126.9~127.2 지역과 37.65~37.8, 126.7~126.9 지역, 2개의 지역이 2016년에서 2020년으로 시간이 경과했음에도 강수날에 사고가 다발하는 위험구역으로 확인됨

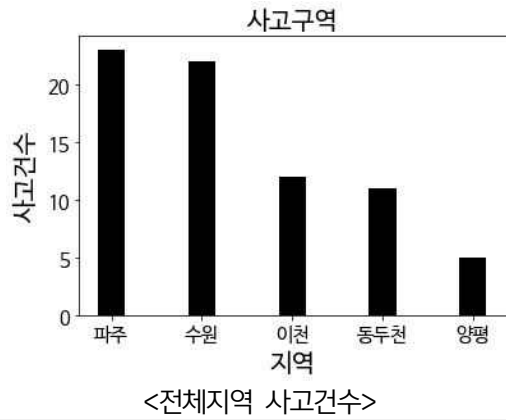


<2016년과 2020년 강수날 사망교통사고 위치좌표>



<2016년 사고 위치좌표 분포도(위)와 2020년 사고 위치좌표 분포도(아래)>

- 앞선 과정을 통하여 강수날 발생한 사망교통사고의 위치좌표를 출력하여 시간이 경과했음에도 여전히 사고가 다발하는 구역을 육안으로 확인하였으나 이는 지역을 특정하기에는 어려움이 있음. 따라서 경도와 위도를 기준으로 빈도를 측정하는 분포도를 시각화하여 보다 세부적으로 지역을 특정
- 위의 2개의 분포도를 통해 <37.2~37.3, 126.9~127.2> 지역(이상 위험지역1)과 <37.65~37.8, 126.7~126.9> 지역(이상 위험지역2), 2개의 지역이 위험구역으로 확인됨
- 특정한 위험지역의 경도와 위도 데이터를 이용하여 해당 범위 이내에서 발생한 사고 데이터 정보만을 따로 추출하여 새로운 데이터프레임을 구축

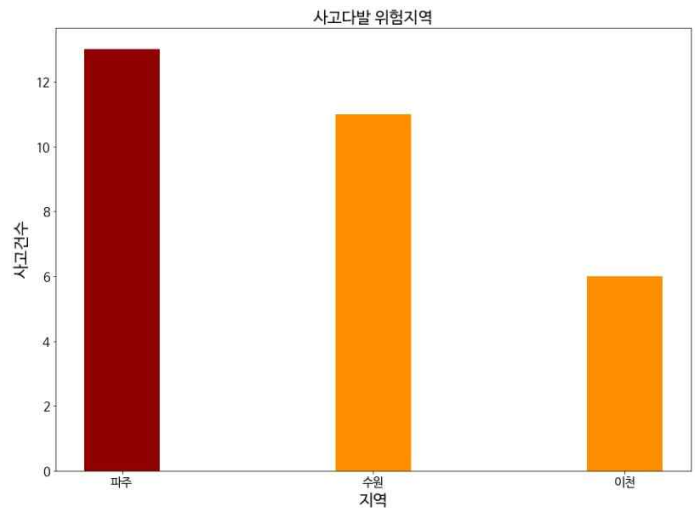


01	11	01	차대사람	횡단중	횡단중
01	12	02	차대사람	차도통행중	차도통행중
01	13	03	차대사람	길가장자리구역 통행중	길가장자리구역통행중
01	14	04	차대사람	보도통행중	보도통행중
01	15	05	차대사람	기타	기타
02	21	21	차대차	정면충돌	정면충돌
02	22	22	차대차	측면충돌	측면충돌
02	23	23	차대차	추돌	추돌
02	23	21	차대차	추돌	진행중 추돌
02	23	22	차대차	추돌	주정차중 충돌
02	24	25	차대차	기타	기타

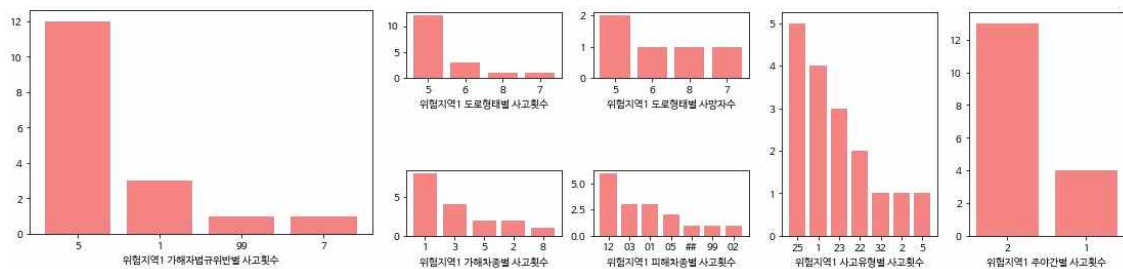
<사고유형 코드 표>

01	승용차
02	승합차
03	화물차
04	특수차
05	어플차
06	사륜오토바이(ATV)
07	원동기장치자전거
08	자전거
09	개인형이동수단(PM)
10	건설기계
11	농기계
12	보행자

<가해/피해당사자차종별 코드 표>

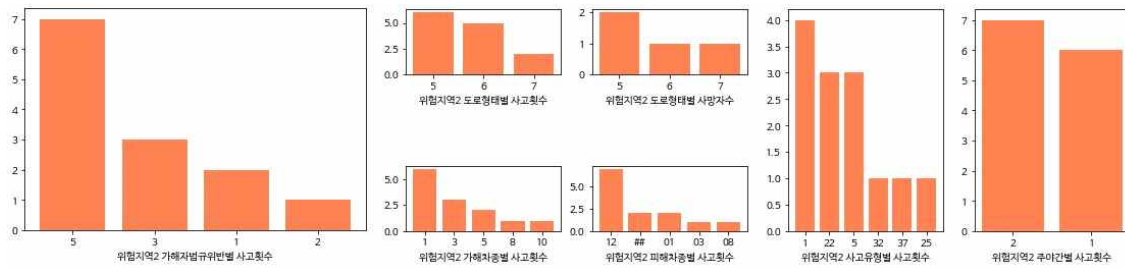


<위험지역 사고 건수>



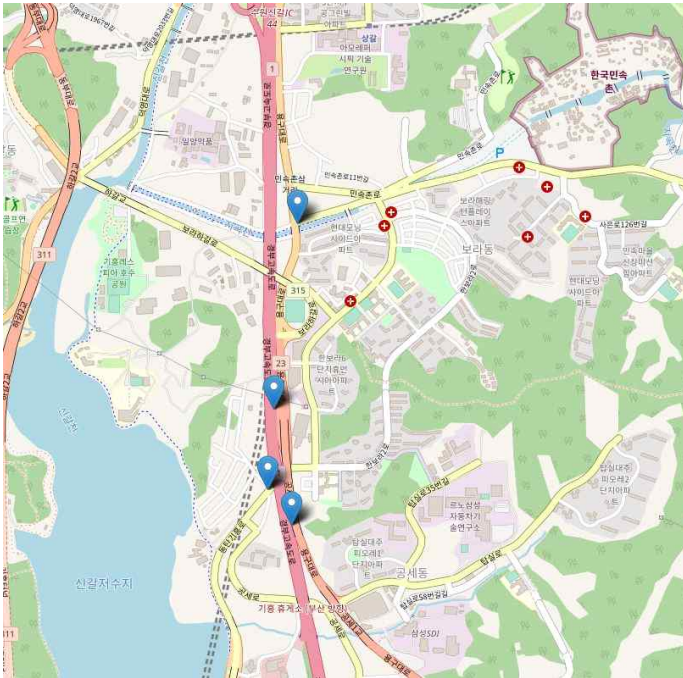
<위험지역1>

- 2020년 강수날 전지역에서 발생한 사망교통사고 횟수의 통계를 보면, 파주/수원/이천/동두천/양평 순으로 사고가 발생함
- 2020년 강수날 위험지역에서 발생한 사망교통사고 횟수의 통계를 보면, 파주/수원/이천 순으로 각각 13건, 11건, 6건의 사고가 발생함
- 데이터분석을 통하여 특정한 위험지역 내에서의 사고는 모두 파주(위험지역2), 수원, 이천(이상 위험지역2)에서 발생하였으며 해당 지역들은 전 지역 사고횟수 통계에서도 각각 1,2,3순위를 기록한 지역으로 특정한 위험지역의 신뢰성을 높여주는 자료라고 생각되어짐

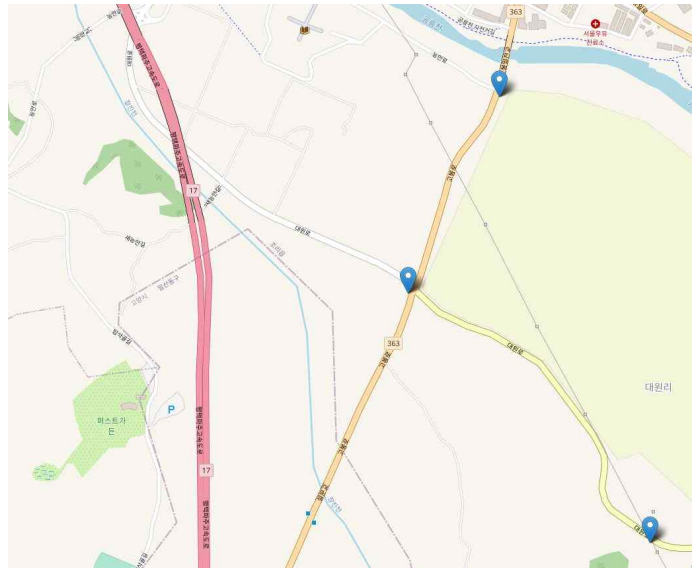


<위험지역2>

- 가해자법규위반
 - 위험지역1과 2 모두 가해자법규5 즉, 안전운전 의무 불이행 항목에 해당하는 사고가 가장 많았음
- 도로형태
 - 위험지역1과 2 모두 도로형태5 즉, 기타 단일로에서 사고가 가장 많이 발생하였음
 - 위험지역1과 2 모두 도로형태5 즉, 기타 단일로에서 가장 많이 사망자가 발생하였음
 - 위험지역에서는 기타 단일로에서 가장 많은 사고 및 사망자의 발생이 일어남을 확인 가능
- 가해/피해차종
 - 위험지역1과 2 모두 가해차종은 차종1 즉, 승용차가 가장 많은 사고횟수를 기록함
 - 위험지역1과 2 모두 피해차종은 차종12 즉, 보행자 가장 많은 사고횟수를 기록함
 - 위험지역에서는 승용차가 가장 많은 가해를 하였고 보행자가 가장 많은 피해를 당하였다는 것을 확인함
- 사고유형
 - 위험지역1에서는 사고유형25 즉, 차대차 간 기타 이유로 발생한 사고의 횟수가 가장 많았음
 - 위험지역2에서는 사고유형1 즉, 차대사람 간 횡단중 발생한 사고의 횟수가 가장 많았음
- 주야간
 - 위험지역1과 2 모두 주간보다 야간에 발생한 사고가 많았음



<색출 지역1, 경부고속도로>



<색출지역2, 고봉로>

- 위험지역 위치좌표에 해당하는 지점을 실제 지도 위에 표시하여 사고가 발생한 지점의 실제 위치가 어디인지 표기하여 위험구역을 보다 정밀하게 관찰 할 수 있었음
- 위험지역 내에서도 위험지역1의 경부고속도로와 위험지역2의 고봉로가 사고가 다발하는 구간임을 확인할 수 있었음

◎ 3. 결론

○ 3.1 가설 결론

(1) 가설1. 강수날에 일반 차량보다 화물차의 사고가 증가할 것이다.

- 일반차량(승용차/승합차)의 경우 맑은날에 비해 강수날 사망사고비율이 비슷하거나 상승하는 경향이 있지만 화물차량의 경우 맑은날에 비해 강수날 사망사고비율이 하락하는 경향이 있음
- "강수날에 일반차량보다 화물차량의 사고가 증가할 것이다."라는 가설은 일반차량의 사망사고비율이 상승폭을 그리고 화물차량의 사망사고비율이 하락폭을 그리며 잘못된 가설이라는 것을 확인하였음

>>> 가설 : ×

(2) 가설2. 강수량이 높을수록 사고가 많이 발생할 것이다.

- 강수량이 많은날 발생한 사고가 많은 사망자가 나오는 사고로 직결되는 것은 아닌것으로 확인되었음
- 강수날 발생한 사고는 강수량이 적은 날에 집중적으로 발생한 경향이 있음을 확인함
- "강수량이 높을수록 사고가 많이 발생할 것이다."라는 가설은 사고가 강수량이 적은 날에 집중적으로 발생하였다는 것으로 잘못된 가설이라는 것을 확인하였음. 다만, 분석 과정에서 강수량이 적을수록 사고횟수는 증가하는 반비례 관계에 있음을 확인하였고 이는 많은 양의 강수시, 운전자가 경각식을 지니고 운전을 하는 반면, 소량의 강수시, 방심하는 경향이 있을 것이기 때문이라고 추측함

>>> 가설 : ×

(3) 강수날 교통사고는 야간에 더 많이 발생할 것이다.

- 강수날에는 주간보다 야간에 더 많은 사고가 일어나며 주간에 발생하는 사고의 경우 그 횟수가 줄어들고 있음을 확인하였음
- "강수날 교통사고는 야간에 더 많이 발생할 것이다."라는 가설은 사실인 것으로 확인하였음. 추가로, 분석 과정에서 주야간별 사망자수의 평균은 주간의 경우 줄어들고 야간의 경우에 늘어나고 있음을 확인함

>>> 가설 : ○

(4) 가설4. 강수날 교통사고는 안전운전을 하지 않아서 발생한 사고가 가장 많을 것이다.

- 강수날 발생한 사망교통사고는 안전운전 의무 불이행 항목을 위반하여 발생한 사고가 1위, 신호위반이 원인으로 발생한 사고가 2위인 것으로 확인됨
- "강수날 교통사고는 안전운전을 하지 않아서 발생한 사고가 가장 많을 것이다."라는 가설에서 말하는 안전운전은 안전운전 의무 불이행 항목에 부합하는 가설이므로 사실인 것으로 확인하였음. 추가로, 해당 항목의 위반으로 발생한 사고는 기타 단일로 부근에서 가장 많이 발생하였고 교차로 내와 교차로 부근에서 역시 다수 발생하였음을 확인함

>>> 가설 : ○

○ 3.2 확장 분석 결론

- 시간이 경과했음에도 강수날 사망교통사고가 지속적으로 발생하는 지역을 특정해 위험지역으로 분류함
- 위험지역1[수원(화성), 이천(용인)] : 경도: 37.2~37.3, 위도: 126.9~127.2
- 위험지역2[파주(고양)] : 경도 37.65~37.8, 위도: 126.7~126.9
- 위험지역1과 위험지역2에서 2020년에 발생한 사고의 횟수는 각각 17건과 13건으로 확인되었음
- 경기도내 사고발생횟수 순위는 파주, 수원, 이천 순으로 이는 위험지역내 사고발생횟수 순위와 동일함. 자료의 일관성은 분석의 신뢰성을 높여주는 근거로 작용
- 위험지역에서는 야간에 승용차와 보행자 간, 기타 단일로에서 가해측의 안전운전 의무 불이행으로 발생하는 사고가 가장 일반적이었음
- 위험지역1에서는 차대차 간 기타 이유로 발생하는 사고가 가장 많았으며 위험지역2에서는 차대사람 간 횡단 중 발생하는 사고의 유형이 가장 많았음
- 위험지역 안에서도 밀접해 있거나 같은 노상에 위치한 구역을 색출 할 수 있었는데, 위험지역1의 기흥휴게소 ~ 수원 신갈IC, 경부고속도로[용인]와 위험지역2의 설문교 ~ 봉일천교, 고봉로[파주]가 사고가 다발하는 구간임을 확인할 수 있었음.
- 위험지역과 위험지역내 추가 색출 구간에는 다음과 같은 조치가 이루어져야 함
- 위험지역은 시군구 규모에서 강수 시 발생하는 사망교통사고에 대한 추가적인 조치가 이루어질 필요가 있음
- 위험지역 중에서 사고가 다발하는 특정 구간은 시군구 규모에서 도로 보수 혹은 단속 강화 등의 집중적인 조치가 이루어질 필요가 있음

㉠ 4. 마무리

○ 4.1 개선점 및 소감

팀 프로젝트를 마무리하며 개인적으로 느끼는 보완점은 첫째, 더 많은 기간의 데이터를 수집하여 분석의 근거와 신뢰성을 높이지 못한 부분이 아쉽습니다. 프로젝트 초반부에 더 긴 기간의 데이터 수집을 계획으로 설정하였더라면 이행 했을 것인데, 프로젝트 중반부를 넘어서 깨달은 부분이 아쉽고 계획 단계의 중요성을 실감하였습니다. 둘째, 코딩 실력이 부족하여 아쉬웠습니다. 이번 프로젝트에 삽입된 파일이 아닌 다른 데이터를 사용하더라도 코드가 원활하게 작동할 수 있도록 하나의 '플랫폼'과 같이 프로젝트 코드를 구성 했으면 했습니다. 또한, 데이터프레임을 구성할 때에 더 간결한 문장을 사용하여 전체 코드 길이를 줄이고, 더 원활한 시각화 통계 처리를 통해 제가 말하고자 하는 바를 정확히 전달했으면 하는 아쉬움이 남습니다.

프로젝트를 마무리하고 보니 부족해보이는 부분도 많고 아쉬움도 많이 남지만 데이터 분석의 전반적인 기초와 시각화 등을 배우고 실습하는 과정에서 데이터 분석이 매우 흥미로운 분야라는 생각을 많이 했습니다. 그래서 프로젝트를 붙들고 인터넷과 강의를 찾아보며 노력하는 그 과정이 재미있었다고 생각합니다. 한 학기 동안 강의를 비롯하여 많은 것을 열성적으로 알려주신 교수님께 감사드리고, 데이터 학문에 관심이 있는 만큼 이번 프로젝트와 강의를 발판 삼아 더 열심히 노력하는 전공자가 되겠습니다. 감사합니다.