

# 자전거 교통사고 원인분석과 대책 마련 보고서

강의명: 빅데이터분석과활용

지도교수: 이근정 교수님

제출일: 2019. 12. 1.



수학과 2016110936 김다운

문헌정보학과 2015110560 원예진

소프트웨어융합학과 2016111571 이희경

## ■ 목차

### I. 제안 배경 및 분석 목적

1. 제안 배경
2. 분석 목적

### II. 자전거 교통사고 요인 분석

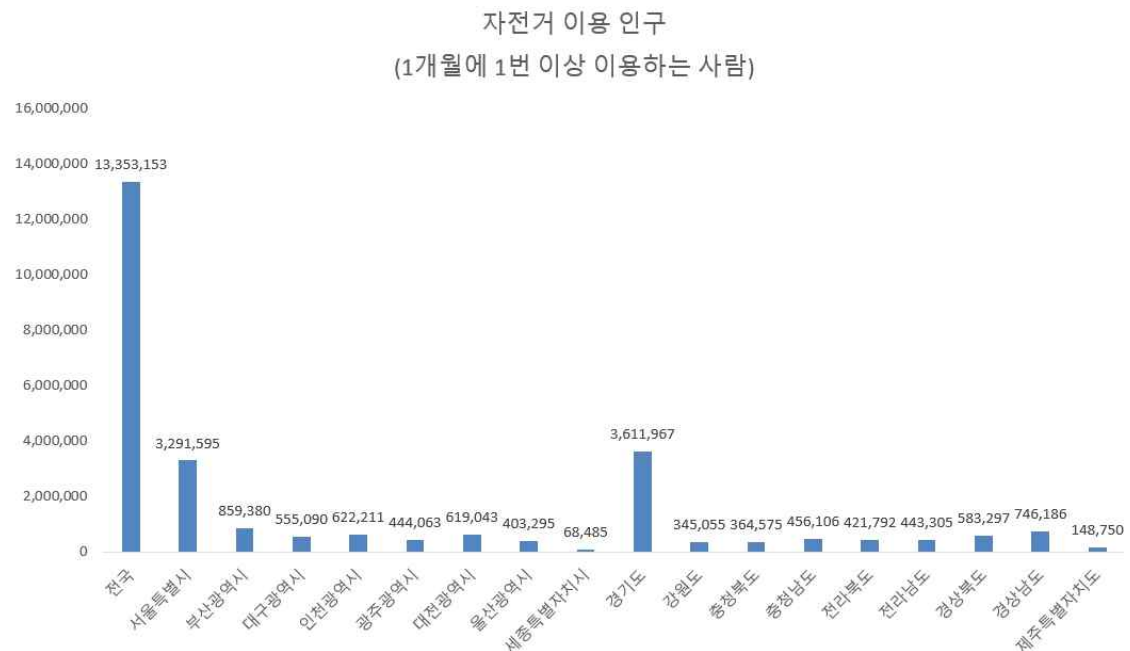
1. 자전거 교통사고 대표 유형 도출 프로세스
2. 활용 데이터
3. 데이터 분석 도구
4. 데이터 분석

### III. 예방 방안 도출

### IV. 참고 문헌 및 데이터 출처

# I. 제안 배경 및 분석 목적

## 1. 제안 배경



KOSIS 국가통계포털의 데이터에 따르면 자전거로 통근·통학을 하는 인구수가 많아지는 추세여서 2000년대에 30만명 대였던 자전거 통근통학 이용자 수가 2010년도 이후에는 약 40만 명 가까이로 약 10만 여명이 늘어났다. 월 1회 이상 이용하는 자전거 이용 인구도 전국적으로 1300만명 이상 있는 등 자전거 이용률이 높아 안전하게 자전거를 타기 위한 기반시

설과 제도의 필요성이 대두되고 있다. 또한 우리나라는 교통혼잡이 심하고 승용차 통행량이 꾸준히 늘어나고 있어 통근·통학 시간에 자전거 이용률을 높여 교통수단 부담률을 나누는 것이 교통혼잡을 완화시키는 대책이 될 수 있다.

## 2. 분석 목적

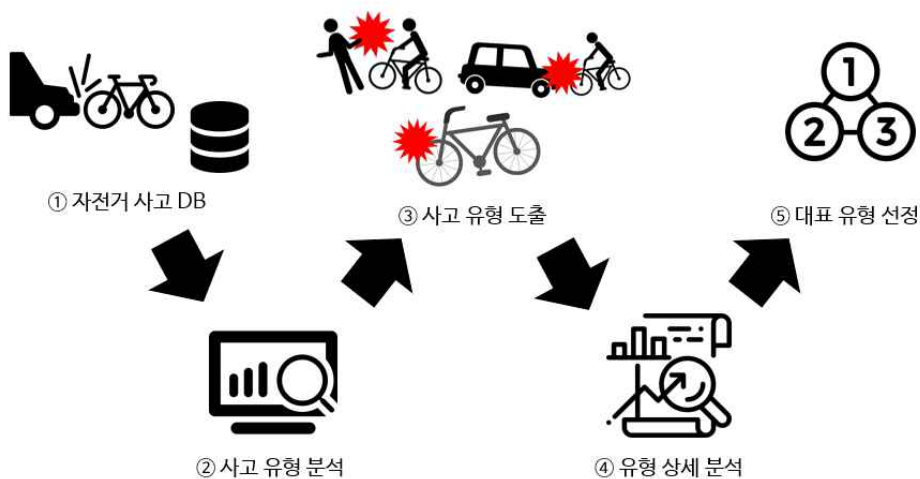
자전거 교통사고와 관련된 데이터를 활용하여 자전거 교통사고 현황과 유형을 분석한다. 이 과정을 통해 도출해낸 특징을 기반으로 주요 사고요인을 알아내어 예방 대책을 제시하는 것이 이번 분석의 목적이다.

## II. 자전거 교통사고 요인 분석

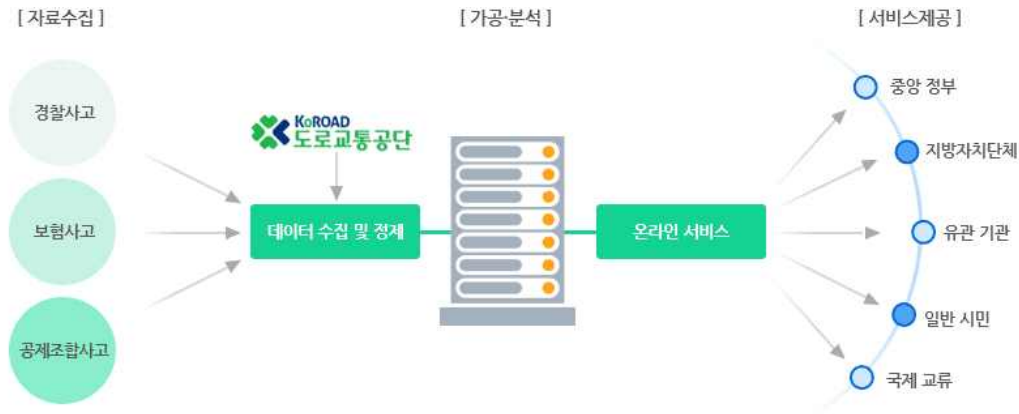
### 1. 자전거 교통사고 대표 유형 도출 프로세스

아래 그림은 자전거 교통사고 데이터베이스로부터 대표 유형 도출까지의 과정을 보여주는 자전거 교통사고 대표 유형을 도출하기 위한 프로세스이다. 우선 사고 유형을 분석하고, 분석한 내용을 바탕으로 전반적인 사고 유형을 도출했다. 그리고 각 사고들을 유형에 맞게 분류하고 분석했다. 마지막으로 여러 변수들을 고려하여 대표 유형을 선정했다.

자전거 교통사고 대표 유형 도출 프로세스



## 2. 활용 데이터



교통사고분석시스템(TAAS)

<TAAS(Traffic Accident Analysis System)> 경찰·보험사·공제조합의 교통사고 자료를 수집, 통합·분석하여 교통안전 정책 수립 등에 활용할 수 있도록 교통사고 정보를 제공하는 시스템이다. 자전거 관련 교통사고의 자료를 얻기 위해 사용하였다.

<census(인구주택총조사)> 연도별 시도의 인구수를 알기 위해 사용하였다.

<KOSIS(국가통계포털)> 자전거를 이용해 통근·통학하는 인구와 자전거 도로 현황/자전거도로 여부에 따른 자전거 교통사고 현황 자료를 사용하였다.

## 3. 데이터 분석 도구



<Excel> 데이터 처리의 가장 기본적이고 대중적인 프로그램으로 쉽고 편리하게 데이터 분석을 할 수 있게 해주는 도구이다.



<Python> 데이터 분석에 특화된 툴로서 데이터 분석을 할 수 있는 다양한 라이브러리를 사용할 수 있다. 그리고 수업시간에 배웠던 내용을 활용 가능하다.

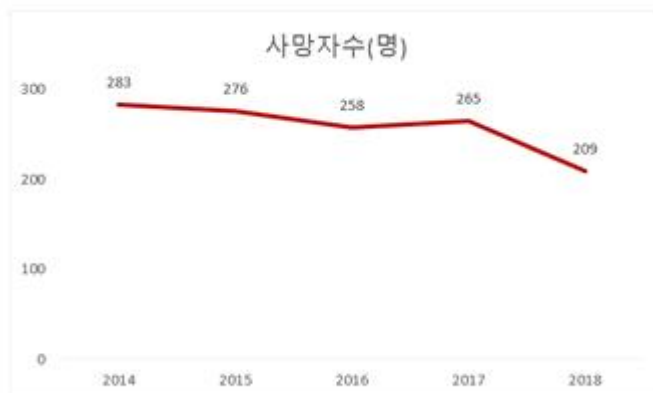
## 4. 데이터 분석

### ① 자전거사고 발생 추세

다음은 자전거 사고 추세를 알기 위해 전체적인 사고건수와 부상자수, 사망자수에 대한 연도별 변화를 꺾은선 그래프로 시각화하였다.

2018년도기준 교통사고 현황(최근5년)

구분	유형	2014년	2015년	2016년	2017년	2018년
자전거사고	발생건수	16,664	17,366	14,937	14,063	11,940
	사망자	283	276	258	265	209
	부상자	17,133	17,905	15,360	14,433	12,258



## ① 자전거사고 발생 추세 예측

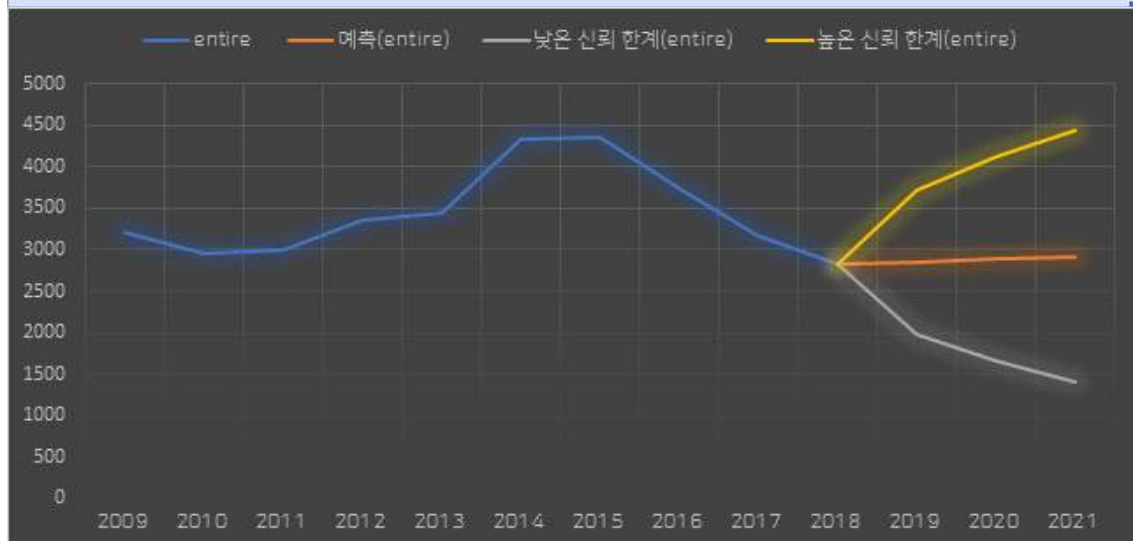
얼핏 단기적으로 보면 사고건수와 부상자수와 사망자수가 줄어드는 것처럼 보인다. 그럼에도 불구하고 사망자 수는 꾸준히 200명을 넘으며, 앞에서 보았듯이 자전거 이용객 수는 꾸준히 늘고 있으므로 이를 주제로 분석을 할 필요성을 느꼈다. 우선 사고건수와 부상자수와 사망자수가 과연 앞으로도 꾸준히 줄어들 것인지 현황을 알기 위해 예측분석을 실시했다.

위 데이터는 2014년도부터 2018년도까지의 짧은 데이터로, 예측분석을 실시하기에 기간이 너무 짧다고 판단하여 더 이전의 데이터들을 추가하기로 했다. 처음엔 긴 연도 범위를 갖는 전국 데이터를 찾지 못해서 우선 서울시 데이터로 2009년도부터 2018년도까지의 예측 그래프를 그렸다. 여기에선 전체 사고건수를 entire로 이름을 바꾸었다.

기간	발생건수	사망자수	부상자수	entire
2009	3055	44	3165	3209
2010	2847	34	2933	2967
2011	2861	19	2980	2999
2012	3225	29	3342	3371
2013	3250	26	3410	3436
2014	4065	37	4295	4332
2015	4062	27	4329	4356
2016	3503	24	3718	3742
2017	2990	30	3143	3173
2018	2680	28	2796	2824

이를 예측분석을 이용하여 3개년 예측값과 낮은 신뢰 한계 예측값, 높은 신뢰 한계 예측값을 그래프로 나타내었다. 낮은 신뢰 한계 예측값은 긍정적인 추세이고, 높은 신뢰 한계 예측값은 부정적인 예측값이다. 우리는 큰 변수가 없으므로 일반 예측값을 사용한다. 이 때 서울의 예측값은 자전거 사고 수가 꾸준히 늘 것이라고 전망한다.

기간	▼ entire ▼	예측(entire)▼	낮은 신뢰 한계(entire)▼	높은 신뢰 한계(entire)▼
2009	3209			
2010	2967			
2011	2999			
2012	3371			
2013	3436			
2014	4332			
2015	4356			
2016	3742			
2017	3173			
2018	2824	2824	2824.00	2824.00
2019		2857.593939	1986.21	3728.98
2020		2891.187879	1659.48	4122.89
2021		2924.781818	1416.01	4433.56



이후, TAAS의 <2007-2015년 시군구별 자전거 교통사고> 데이터를 찾게 되어 연도별 전국 자전거 사고 데이터를 전처리하였다.

구분		2007년			2008년			2009년			2010년			2011년			2012년			2013년			2014년			2015년		
지역		발생건수	사망자수	부상자수	발생건수	사망자수	부상자수	발생건수	사망자수	부상자수	발생건수	사망자수	부상자수	발생건수	사망자수	부상자수	발생건수	사망자수	부상자수	발생건수	사망자수	부상자수	발생건수	사망자수	부상자수	발생건수	사망자수	부상자수
서울	서울	1,862	25	1,908	2,678	29	2,805	3,055	44	3,165	2,847	34	2,993	2,861	19	2,960	3,225	29	3,342	3,250	26	3,410	4,065	37	4,296	4,062	27	4,329
서울	영등포	36	1	38	41	0	42	42	0	43	36	1	36	57	0	57	53	0	55	44	0	44	53	2	56	59	0	62
서울	영등포	18	0	19	23	1	23	34	0	34	43	1	43	44	0	47	30	0	31	26	0	27	31	1	31	39	0	40
서울	영등포	17	1	17	23	1	22	36	0	36	28	1	29	44	0	46	51	0	53	71	0	62	84	0	99	69	1	74
서울	영등포	74	0	76	94	2	97	110	2	113	104	8	102	112	2	113	91	0	101	118	0	130	144	1	153	159	1	176
서울	영등포	111	3	116	135	3	142	212	2	217	167	2	170	163	0	169	184	4	187	223	2	228	236	1	251	253	1	259
서울	영등포	57	1	59	81	0	85	111	1	117	96	0	100	88	0	94	111	1	115	80	0	84	117	1	124	123	2	127
서울	영등포	56	1	56	103	0	107	117	1	123	103	2	104	99	1	104	125	2	124	115	2	120	126	2	128	115	0	122
서울	영등포	90	0	96	143	0	153	188	0	195	133	0	142	114	0	115	150	0	158	117	0	123	135	1	143	139	0	156
서울	영등포	25	0	25	33	1	33	43	1	45	45	2	44	55	0	54	57	0	58	49	0	52	61	1	62	71	0	74
서울	영등포	57	1	57	109	0	113	93	4	96	107	0	111	96	1	100	141	3	145	139	0	153	135	2	141	189	1	193

이 데이터는 2015년에서 끊겨있으므로 전국의 발생건수, 사망자수, 부상자수를 각 지역의 합계를 내어 기존에 가지고 있던 <자전거(가해운전자) 교통사고>와 <자전거(피해운전자) 교통사고>를 합친 데이터와 연결하여 새로운 데이터셋을 만들었다.



## 자전거 교통사고(가해운전자)

시도	2016	2017	2018
서울	1,884	1,656	1,471
부산	249	225	176
대구	501	523	391
인천	165	137	86
광주	128	135	134
대전	117	128	96
울산	115	101	86
세종	6	13	14

시도	2015	2016	2017	2018
합계	11,390	9,700	9,003	7,618
서울	2,299	1,961	1,607	1,447
부산	388	340	360	318
대구	1,077	835	844	668
인천	494	359	349	222
광주	258	240	263	206
대전	485	431	433	311
울산	278	201	199	143
세종	23	31	42	38

시도	2016	2017	2018
서울	3,845	3,263	2,918
부산	589	585	494
대구	1,336	1,367	1,059
인천	524	486	308
광주	368	398	340
대전	548	561	407
울산	316	300	229
세종	37	55	52
경기	3,685	3,295	2,979
강원	415	393	287
충북	570	549	508
충남	402	407	349
전북	552	526	410
전남	396	432	353
경북	1,095	1,063	915
경남	795	810	646
제주	163	172	135
합계	15,636	14,662	12,389

정제한 데이터들을 합치면 아래와 같이 된다.

연도	사고 건수
2007	8721
2008	10848
2009	12532
2010	11259
2011	12121
2012	12908
2013	13316
2014	16664
2015	17366
2016	15636
2017	14662
2018	12389

최종적으로 전처리가 완료된 위 데이터로 예측분석을 실시한 결과는 아래와 같다.

연도	▼ 사고 건수 ▼	예측(사고 건수) ▼	낮은 신뢰 한계(사고 건수) ▼	높은 신뢰 한계(사고 건수) ▼
2007	8721			
2008	10848			
2009	12532			
2010	11259			
2011	12121			
2012	12908			
2013	13316			
2014	16664			
2015	17366			
2016	15636			
2017	14662			
2018	12389	12389	12389.00	12389.00
2019		15939.71661	12148.45	19730.99
2020		16437.57141	12528.70	20346.44
2021		16935.42621	12911.49	20959.36

이를 그래프로 나타내면, 여기에서도 세 갈래의 그래프가 생성된다. 앞의 서울 자전거 사고 예측분석에서와 마찬가지로 높은 신뢰 한계의 예측치는 사고가 급격히 늘어날 것이라고 전망한다. 여기서 눈여겨 볼 점은, 앞의 예측분석과 달리 전국 단위의 자전거 사고는 낮은 신뢰 한계의 예측치조차 사고가 조금씩 늘어날 것이라고 전망했다는 점이다. 우리는 특별한 변수 없이 일반 예측치인 붉은 그래프를 기준으로 볼 것이다. 따라서 장기적인 데이터에 기

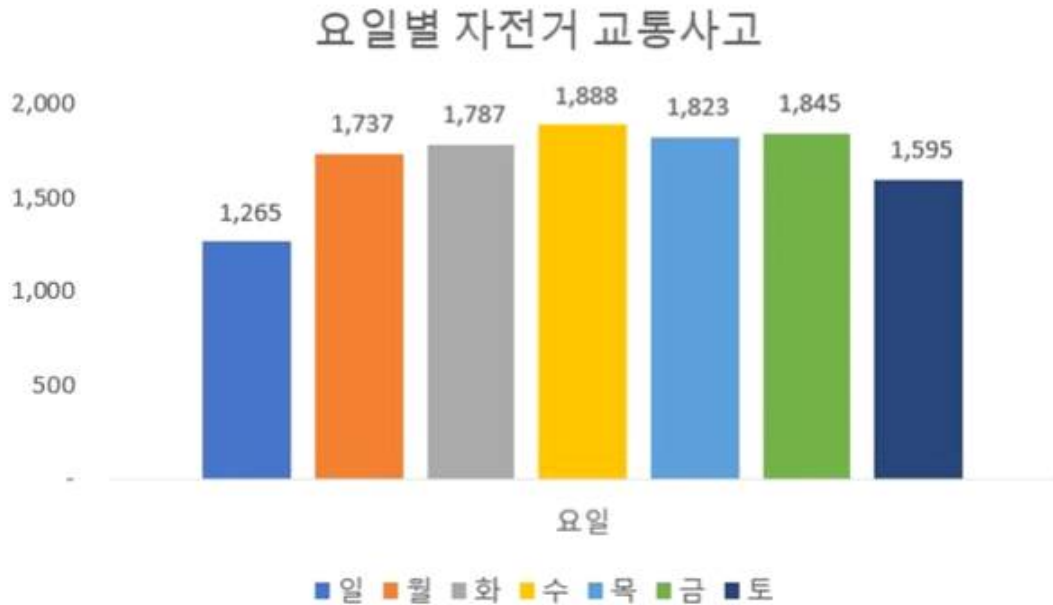
반하여 예측분석을 실시했을 때, 앞으로 자전거 사고는 꾸준히 늘어날 것이라고 볼 수 있다.



## ② 요일별 자전거 교통사고 발생건수

다른 요인들에 대해서도 분석해보았다. 우선 아래의 <자전거 교통사고 통계자료> 데이터의 요일별 로 막대그래프를 나타냈다.

자전거 교통사고 (요일별)						
	발생건수		사망자수		부상자수	
	건수	구성비	명수	구성비	명수	구성비
합계	11,940	100	209	100	12,258	100
일	1,265	10.6	25	12.0	1,319	10.8
월	1,737	14.5	38	18.2	1,748	14.3
화	1,787	15.0	29	13.9	1,847	15.1
수	1,888	15.8	32	15.3	1,936	15.8
목	1,823	15.3	27	12.9	1,873	15.3
금	1,845	15.5	24	11.5	1,892	15.4
토	1,595	13.4	34	16.3	1,643	13.4



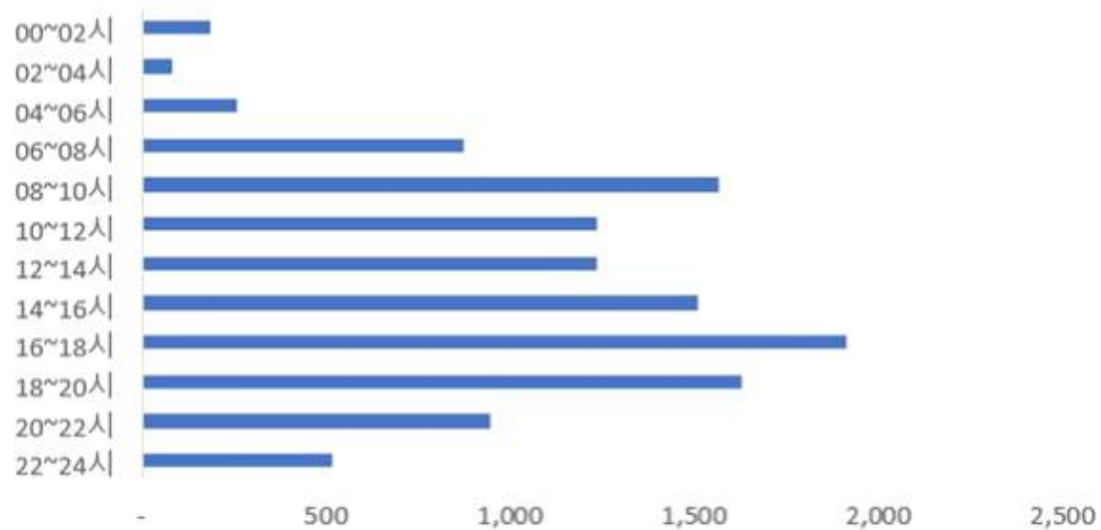
전체적으로 주말을 제외하면 눈에 띄는 특징은 없으나, 제일 사고가 많이 일어나는 요일이 수요일, 금요일, 목요일 순임을 알 수 있었다.

### ③ 시간대별 자전거 교통사고 발생건수

다음은 시간대별 데이터로 가로막대그래프를 나타내었다.

자전거 교통사고 (시간대별)						
	발생건수		사망자수		부상자수	
	건수	구성비	명수	구성비	명수	구성비
합계	11,940	100	209	100	12,258	100
00시~02시	188	1.6	2	1.0	198	1.6
02시~04시	81	0.7	3	1.4	82	0.7
04시~06시	258	2.2	15	7.2	261	2.1
06시~08시	872	7.3	29	13.9	867	7.1
08시~10시	1,564	13.1	22	10.5	1,591	13.0
10시~12시	1,233	10.3	24	11.5	1,261	10.3
12시~14시	1,233	10.3	19	9.1	1,259	10.3
14시~16시	1,509	12.6	26	12.4	1,550	12.6
16시~18시	1,914	16.0	22	10.5	1,995	16.3
18시~20시	1,628	13.6	19	9.1	1,674	13.7
20시~22시	945	7.9	16	7.7	986	8.0
22시~24시	515	4.3	12	5.7	534	4.4

## 시간대별 자전거 교통사고



위의 그래프에서 자전거 사고가 시간대별로는 16~18시, 18~20시, 08~10시 순으로 발생한다는 것을 알 수 있다. 즉, 주로 등·하교, 출퇴근 시간에 많이 발생하고, 저녁시간대 사고가 상대적으로 크다.

### ④ 기상 상태별 교통사고 발생건수

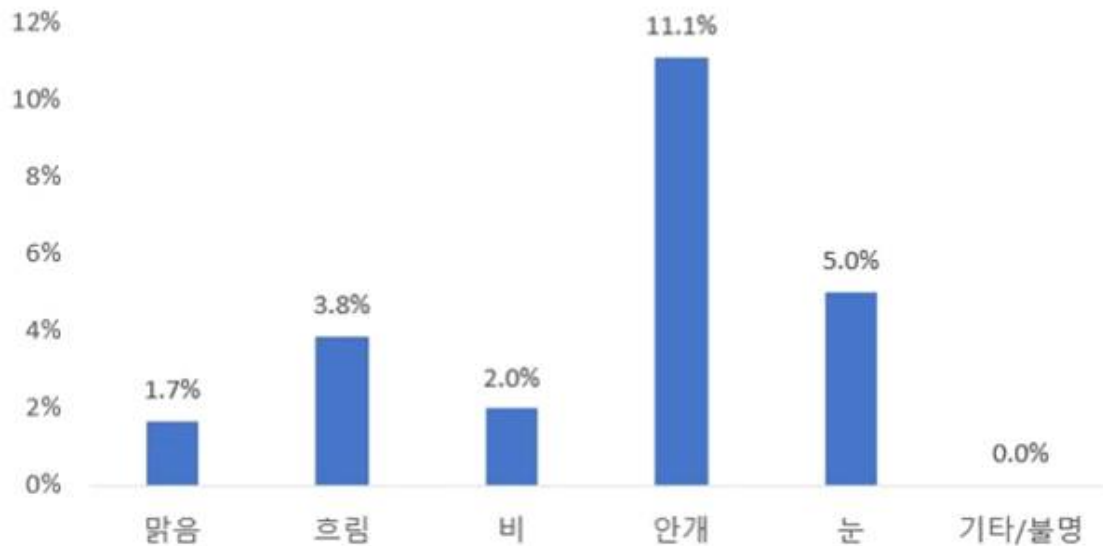
다음은 기상상태별 자전거 교통사고 데이터로 막대그래프를 나타내었다.

#### 자전거 교통사고 (기상상태별)

	발생건수		사망자수			부상자수	
	건수	구성비	명수	구성비	치사율	명수	구성비
합계	11,940	100	209	100	1.8	12,258	100
맑음	11,258	94	189	90	1.7	11,568	94
흐림	312	3	12	6	3.8	314	3
비	302	3	6	3	2.0	307	3
안개	9	0	1	0	11.1	9	0.07
눈	20	0	1	0	5.0	20	0.2
기타/불명	39	0	-	0	0.0	40	0.3



## 기상별 자전거 교통사고 치사율



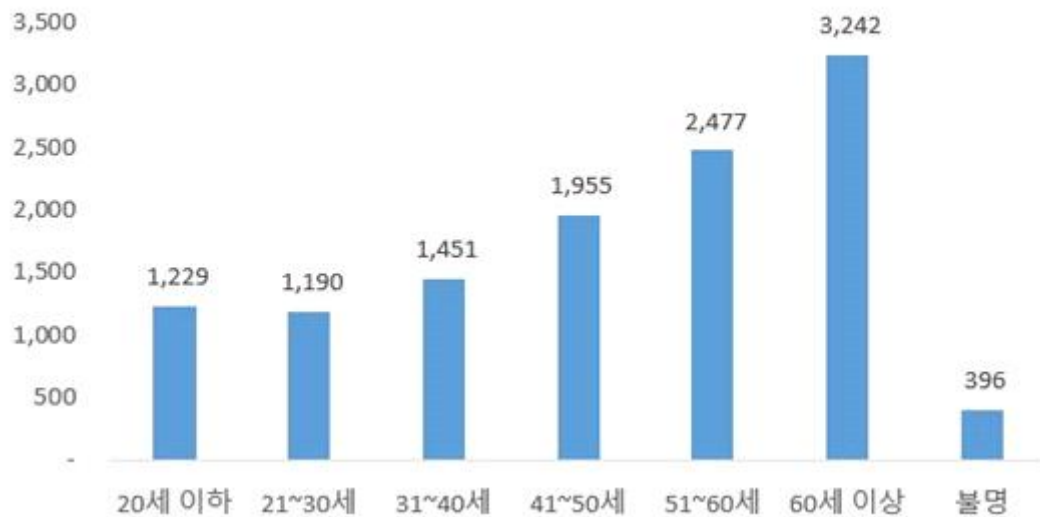
위의 그래프에서 기상상태별 자전거 교통사고 치사율은 안개 낀 날, 눈오는 날, 흐린 날 순으로 발생하는 것을 볼 수 있다.

## ⑤ 연령대별 자전거 교통사고 발생건수

다음은 연령대별 데이터를 막대 그래프로 나타낸 것이다.

자전거 교통사고 (연령대별)						
	발생건수		사망자수		부상자수	
	건수	구성비	명수	구성비	명수	구성비
합계	11,940	100	209	100	12,258	100
20세 이하	1,229	10.3	2	1.0	1,348	11.0
21~30세	1,190	10.0	22	10.5	1,226	10.0
31~40세	1,451	12.2	28	13.4	1,480	12.1
41~50세	1,955	16.4	22	10.5	2,002	16.3
51~60세	2,477	20.7	55	26.3	2,513	20.5
61~64세	960	8.0	17	8.1	976	8.0
65세 이상	2,282	19.1	63	30.1	2,310	18.8
불명	396	3.3	-	0.0	403	3.3

## 연령대별 자전거 교통사고

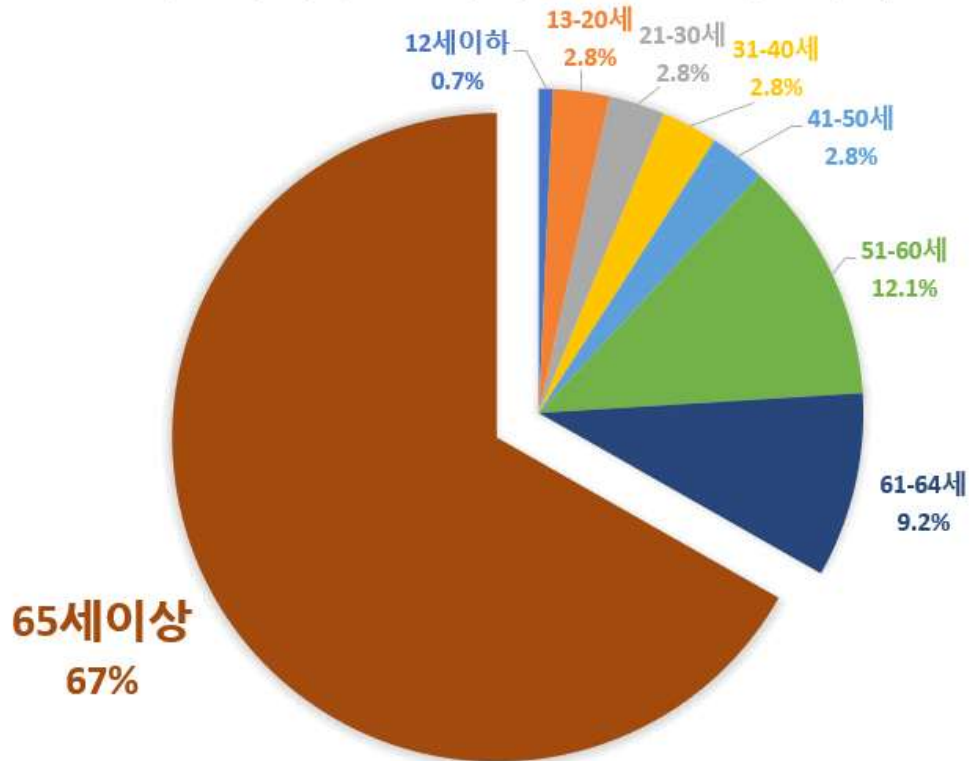


연령대별로는 60대 이상이 가장 많았다. 특히 60대 이상 노년층의 비율은 자전거 사고 피해자 데이터로 비율을 나타낸 원그래프에서도 75% 이상으로 굉장히 높은 비율을 보였다.

구분별(1)	구분별(2)	2017			
		발생 (건)		사망자 (명)	
		소계	구성비 (%)	소계	구성비 (%)
가해운전자	소계	5,659	100	126	100
	12세이하	571	10.1	2	1.6
	13-20세	1,029	18.2	6	4.8
	21-30세	524	9.3	4	3.2
	31-40세	401	7.1	4	3.2
	41-50세	560	9.9	9	7.1
	51-60세	842	14.9	20	15.9
	61-64세	364	6.4	13	10.3
	65세이상	1,340	23.7	68	54
	불명	28	0.5	0	0
피해운전자	소계	9,003	100	141	100
	12세이하	954	10.6	1	0.7
	13-20세	1,221	13.6	4	2.8
	21-30세	823	9.1	4	2.8



## 자전거 피해운전자 사고 연령별 사망자 비율

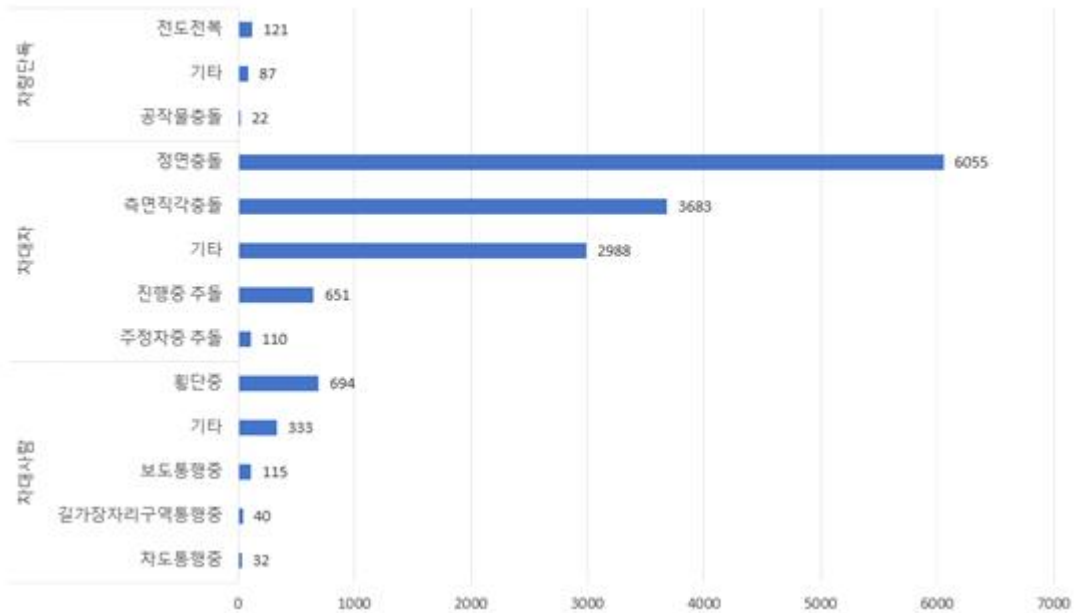


### ㉔ 사고유형별 자전거 교통사고 발생

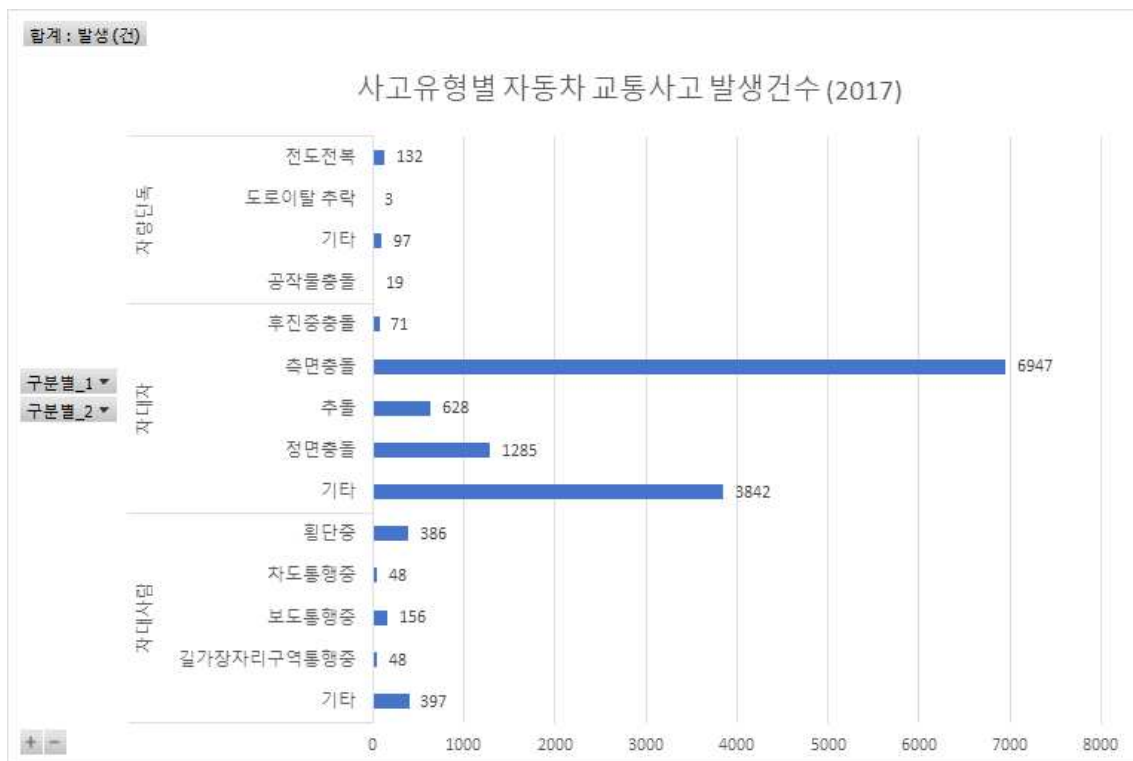
다음은 KOSIS의 <사고유형별 자전거 교통사고 현황> 데이터를 막대그래프로 나타낸 것이다.

구분별_1	구분별_2	occur	die	hurt
합계	소계	14,937	258	15,360
차대사람	소계	1,214	6	1,332
차대사람	횡단중	694	3	768
차대사람	차도통행중	32	0	36
차대사람	길가장자리구역통행중	40	0	43
차대사람	보도통행중	115	1	119
차대사람	기타	333	2	366
차대차	소계	13,487	225	13,819
차대차	정면충돌	6,055	70	6,277

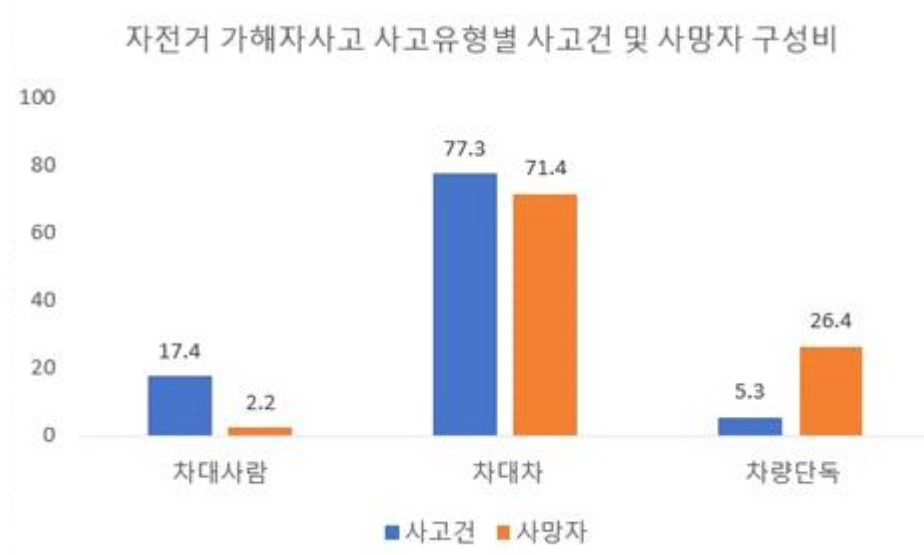
사고유형별 자전거 교통사고 발생건수(2016)



구분별_1	구분별_2	발생 (건)	사망자 (명)	부상자 (명)
차대사람	횡단중	386	3	419
차대사람	차도통행중	48	0	57
차대사람	길가장자리구역통행중	48	0	51
차대사람	보도통행중	156	1	164
차대사람	기타	397	6	420
차대차	정면충돌	1,285	13	1,363
차대차	측면충돌	6,947	123	7,078
차대차	추돌	628	38	630
차대차	후진중충돌	71	1	71
차대차	기타	3,842	49	3,953
차량단독	공작물충돌	19	7	13
차량단독	도로이탈 추락	3	2	1
차량단독	전도전복	132	12	121
차량단독	기타	97	8	90



위 시각화 그래프에서 도로 대부분이 측면직각충돌 위험에 노출되어 있으므로 상충지점을 최소화시켜야 됨을 알 수 있다. 자전거 가해자사고를 차대사람, 차대차, 차량단독의 분류 기준으로 취합하여 사망자 구성비를 나타낸 시각화 그래프는 다음과 같다.

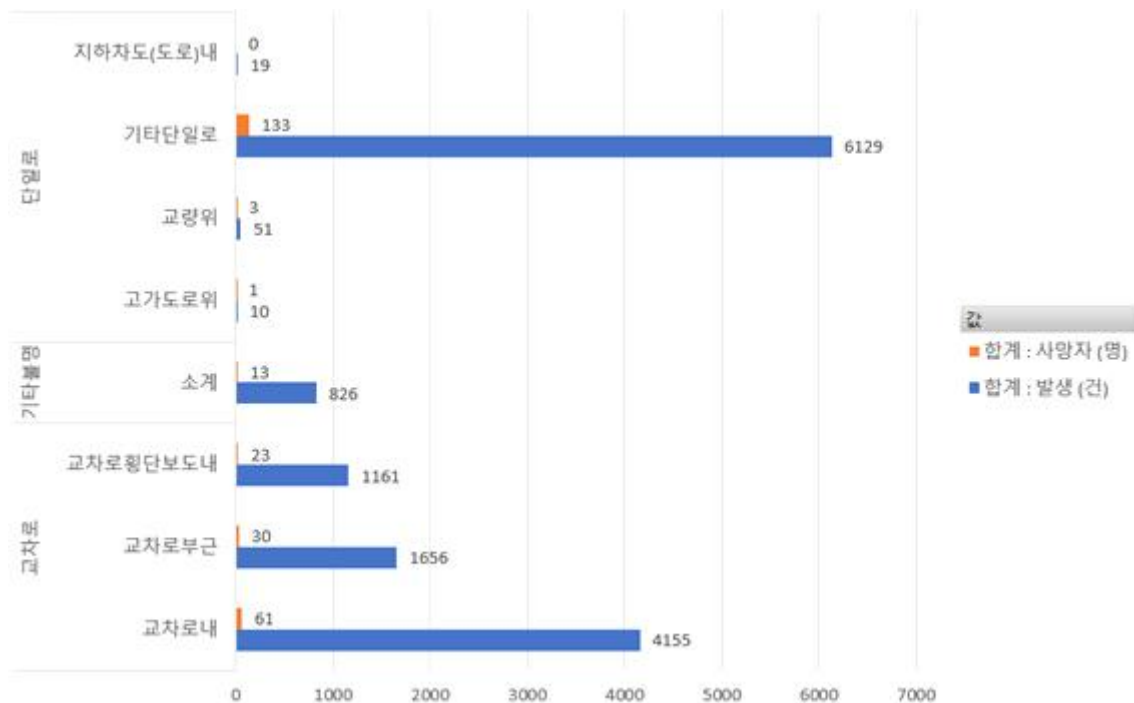


여기에서 자전거는 '차'로 분류되고, 차대차사고(77.3%)가 가장 많았다.

⑦ 도로형태별 자전거 사고 발생

다음은 도로형태별 자전거 사고 데이터를 막대그래프로 나타낸 것이다.

구분별(1)	구분별(2)	2017	
		발생 (건)	사망자 (명)
합계	소계	14,063	265
단일로	소계	6,222	137
	터널안	13	0
	교량위	51	3
	고가도로위	10	1
	지하차도(도로)내	19	0
	기타단일로	6,129	133
교차로	소계	6,972	114
	교차로내	4,155	61
	교차로횡단보도내	1,161	23
	교차로부근	1,656	30
주차장	소계	42	0
철길건널목	소계	1	1
기타불명	소계	826	13



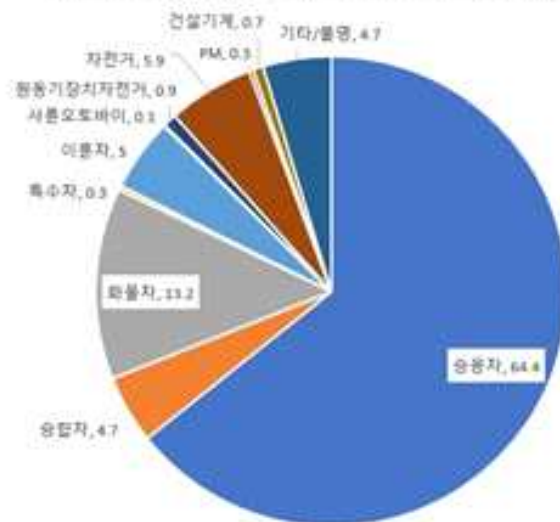
분석 결과, 주행 시간과 길이가 긴 일반적인 주행로(기타단일로)에서 발생한 자전거사고가 약 43%인 반면, 약 50%의 사고가 교차로 및 횡단보도와 그 부근에서 발생했다. 또한 교차로와 횡단보도 및 그 부근의 사고가 절반을 넘게 차지하고 있었다. 현재 자전거도로가 교차로와 횡단보도 부근에서의 안전을 고려하지 못한 경우가 많은데, 자전거도로가 단절되고 자전거횡단보도가 설치되지 않은 곳이 많으며, 통행 안내가 부족하여 어디로 가야 할지 알 수 없는 경우가 많다. 이는 자동차와 자전거가 상충하는 교차로와 횡단보도 구간이 주요 위험 구간이며 안전개선의 중요도가 높음을 시사한다.

## ⑧ 차종별 자전거 교통사고 발생

다음은 TAAS의 <자전거사고 차종별 현황(2018)> 데이터를 원그래프로 시각화한 것이다.

		사고건수		사망자수		부상자수	
가해운전자	피해운전자						
자전거	합계	4771	100	91	100	5041	100
	승용차	2239	46.9	37	40.7	2279	45.2
	승합차	236	4.9	3	3.3	247	4.9
	화물차	414	8.7	16	17.6	403	8
	특수차	20	0.4	0	0	20	0.4
	이륜차	226	4.7	2	2.2	278	5.5
	원동기장치자전거	34	0.7	0	0	41	0.8
	자전거	449	9.4	3	3.3	556	11
	PM	2	0	0	0	3	0.1
	건설기계	22	0.5	4	4.4	18	0.4
	농기계	2	0	0	0	2	0
	보행자	829	17.4	2	2.2	920	18.3
	기타/불명	298	6.2	24	26.4	274	5.4
합계	자전거	7618	100	121	100	7773	100
승용차		4903	64.4	66	54.5	4908	63.1
승합차		355	4.7	9	7.4	357	4.6
화물차		1003	13.2	33	27.3	982	12.6
특수차		25	0.3	1	0.8	24	0.3
이륜차		380	5	0	0	435	5.6
사륜오토바이		8	0.1	0	0	11	0.1
원동기장치자전거		67	0.9	0	0	73	0.9
자전거		449	5.9	3	2.5	556	7.2
PM		20	0.3	0	0	21	0.3
건설기계		51	0.7	9	7.4	42	0.5
기타/불명		357	4.7	0	0	364	4.7

자전거사고 차종별 현황 - 자전거가 피해자인 경우



자전거사고 차종별 현황 - 자전거가 가해자인 경우



자전거가 가해자인 사고에서는 자전거:승용차(46.9%), 자전거:보행자(17.4%), 자전거:자전거(9.4%) 등의 순으로 많았고, 자전거가 피해자인 사고에서는 승용차:자전거(64.4%), 화물차:자전거(13.2%), 자전거:자전거(5.9%) 등의 순으로 많이 발생하였다. 자동차와 자전거의 안전한 공존을 위한 제도와 인식의 부족으로 차 대 자전거 사고가 많이 발생하고 있다. 이는 차량과 자전거 간에 상호 안전을 고려한 통행이 이루어지지 못하고 있음을 나타내며, 특히, 자동차 운전자의 자전거 보호가 부족하기 때문으로 판단된다.

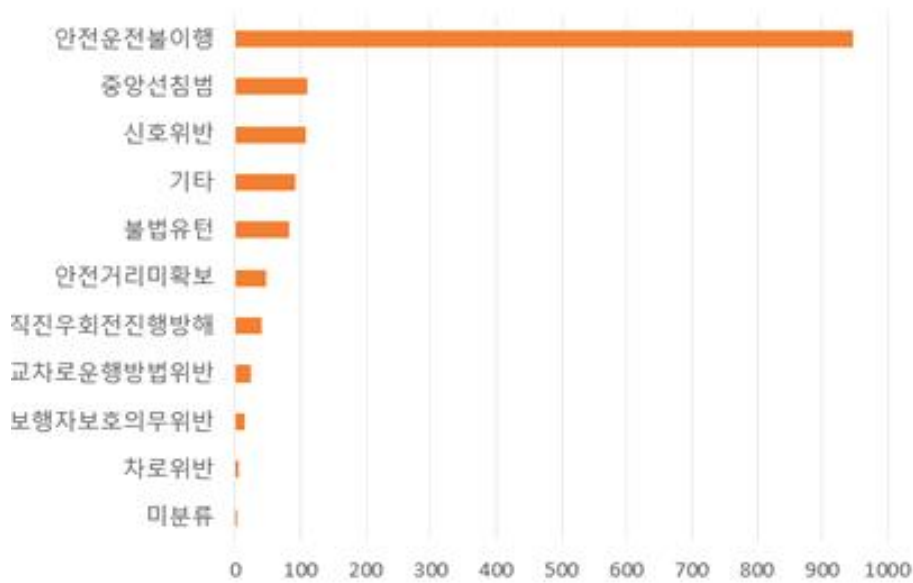
## ⑨ 법규 위반 유형별 자전거 교통사고 발생

다음은 자전거 사고 법규 위반 유형 데이터를 막대 그래프로 시각화한 것이다.

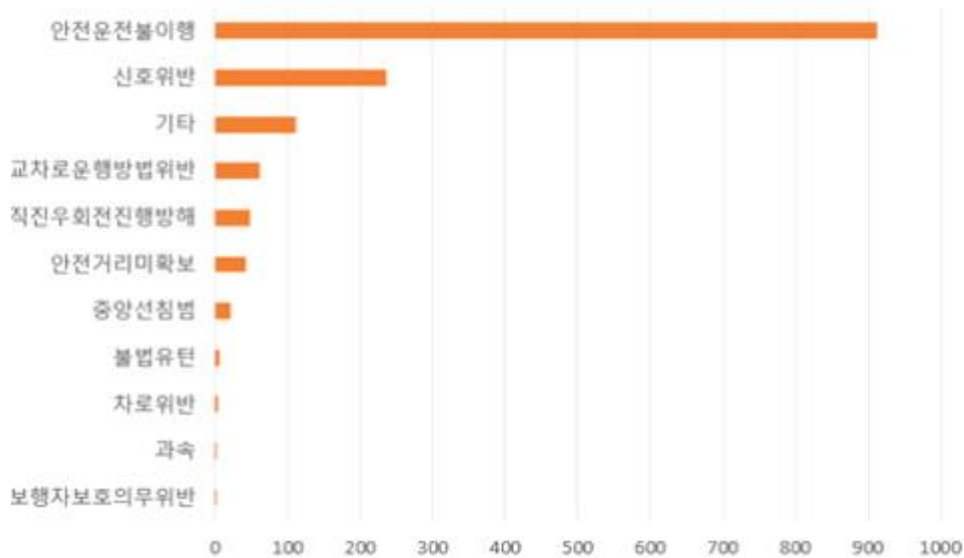
	발생 (건)	구성비 (%)
중앙선 침범	374	7.8
신호위반	366	7.7
안전거리 미확보	172	3.6
안전운전 의무 불이행	3,029	63.5
교차로 통행방법 위반	172	3.6
보행자 보호의무 위반	41	0.9
기타	617	12.9



### 자전거 가해 운전자 법규 위반



### 자전거 피해 운전자 법규 위반 사항

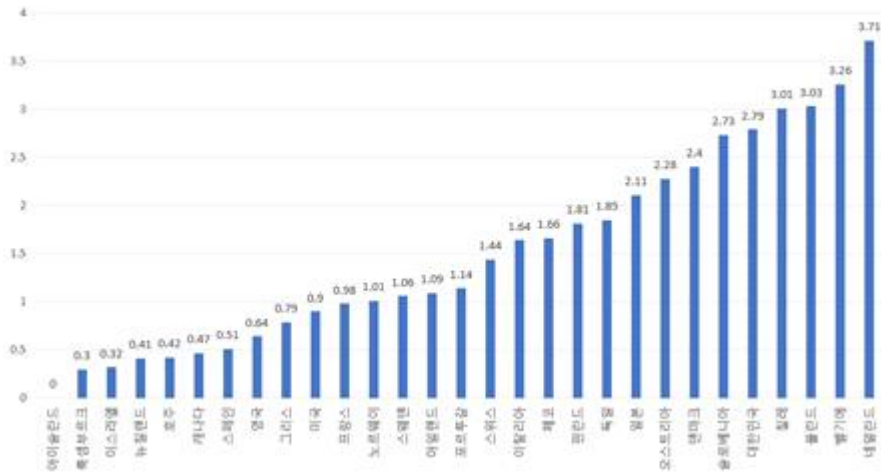


가해 운전자와 피해 운전자의 경우 모두 '안전운전불이행'이 가장 많은 원인을 차지했다. 자전거 주행 시 반드시 안전 운전 할 수 있도록 사전 안내가 필요할 것이다. 향후 관련 기관에서 교통사고 데이터 수집 시 상세 위반 내용에 대한 정보 수집을 하면 더욱 구체적인 교통사고 원인 분석과 그에 따른 사전 예방 조치가 가능할 것이다.

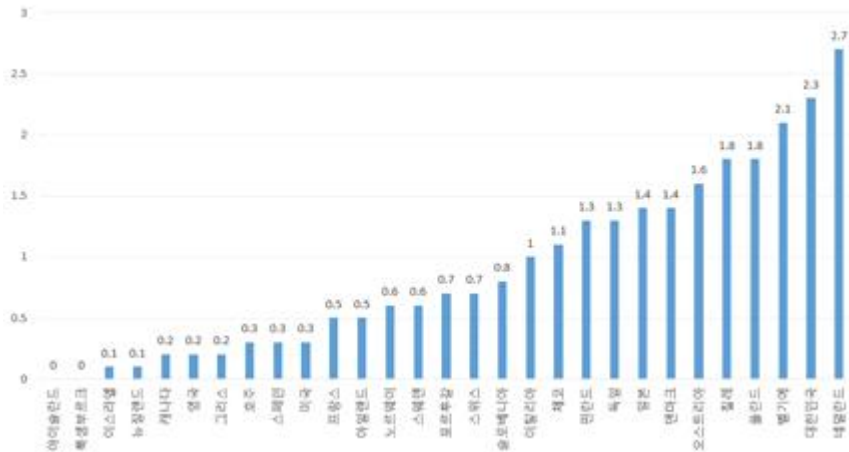
다음은 OECD 회원국 자전거 교통사고를 비교한 데이터를 막대그래프로 한 것이다.



인구 10만 명당 자전거 승차 중 사망자 수

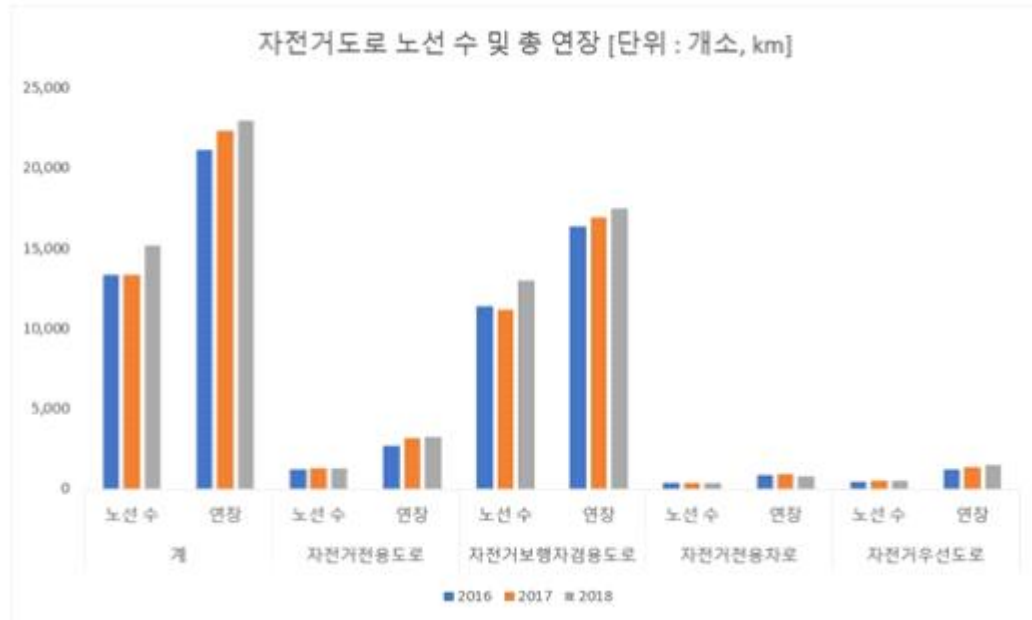


65세 이상 노인 인구 10만 명당 자전거 승차 중 사망자 수



우리나라의 인구 10만 명당 자전거 승차 중 사망자 수는 OECD 회원국 중 다섯 번째로 많은 2.79명이었다. 우리나라의 65세 이상 노인 인구 10만 명당 자전거 승차 중 사망자수는 OECD 회원국 중 두 번째로 많은 2.3명이었다.

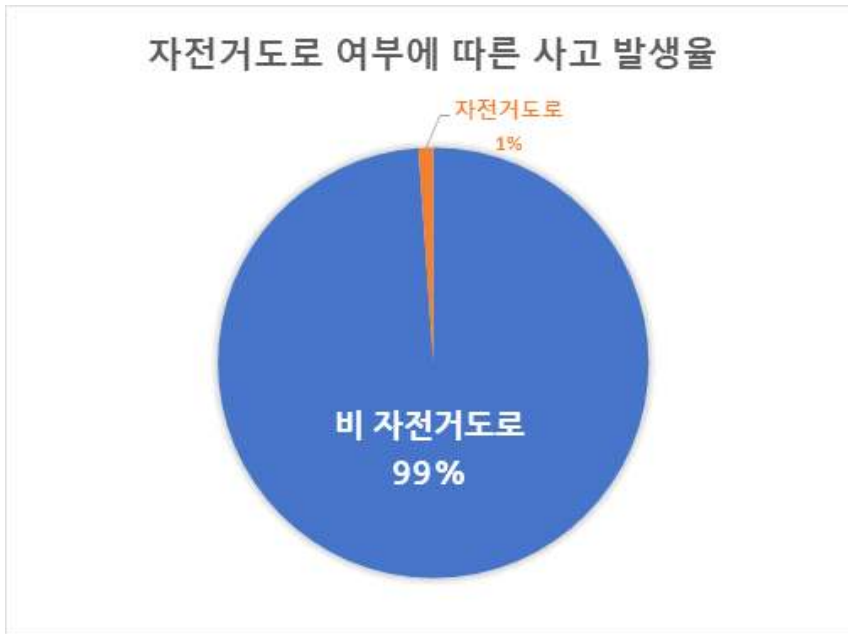
다음은 연도별 자전거도로 연장 및 노선 수 데이터를 막대 그래프화하였다.



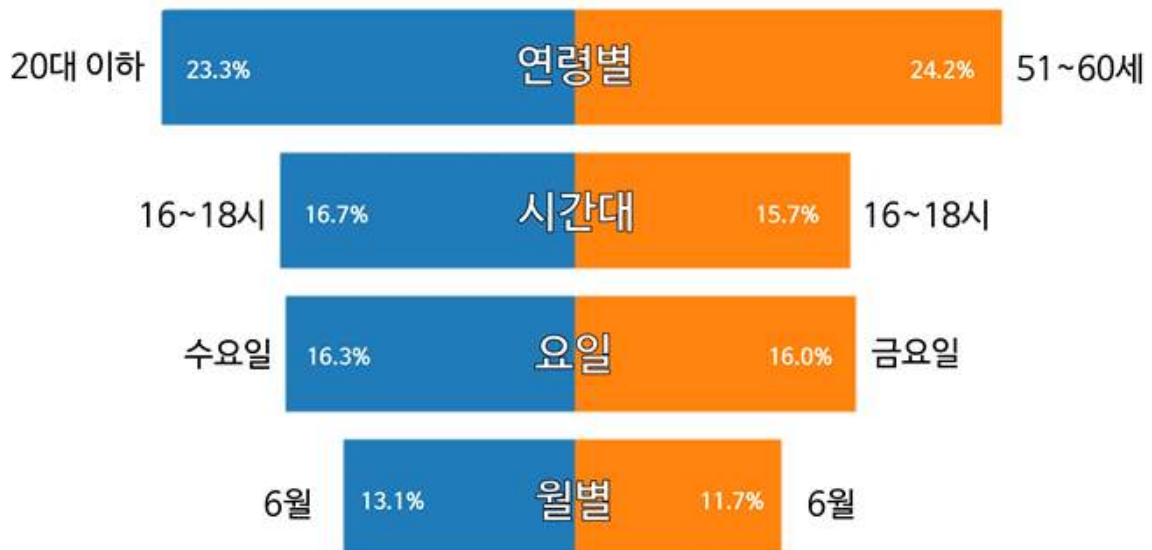
국내 자전거 이용 활성화 정책에 따라 자전거 노선 및 자전거 교통 인프라가 계속 확장되는 추세이다. 2018년 국내 자전거도로 총 연장은 약 2만 3000km로 2016년 약 2만 1000km에 비해 약 8.6% 증가하였으며, 2018년 노선 수는 약 15,000개로 2016년 약 13,000개소에 비해 약 13% 증가하였다.

다음은 자전거도로 여부에 따른 자전거 교통사고 현황이다.

구분별	2017	
	발생 (건)	사망자 (명)
	구성비 (%)	구성비 (%)
비 자전거도로	98.9	100
자전거도로	1.1	0
합계	100	100
	발생 (건)	사망자 (명)
비 자전거도로	98.9	100
자전거도로	1.1	0
합계	100	100



자전거도로에서의 사고 건수는 158건이고, 비 자전거도로에서는 13,905건이었다. 자전거(가해자, 피해자) 교통사고의 특성을 정리하면 다음과 같았다.

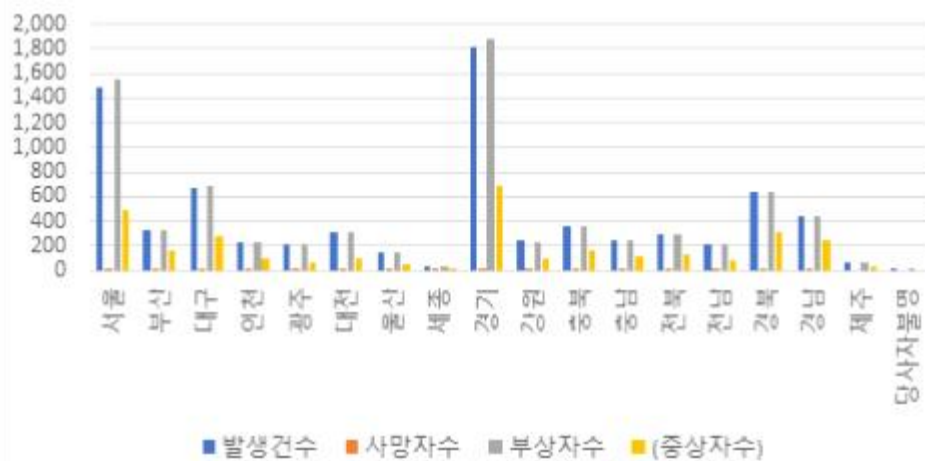


전국 단위의 사고 데이터를 분석한 뒤, 지역별 자전거 사고 발생 현황을 알기 위해 도로교통공단에서 제공하는 <전국 교통사고다발지역> 데이터를 얻었다. 아래 그래프와 같이 당연히 사고가 많이 일어나는 지역은 인구가 많은 서울과 경기도 중심이었기에, 자전거 사고의 특성을 파악하기 위해서 인구 대비 자전거 사고 발생건수비율을 알고자 하였다. 하지만 해당 데이터가 없어 TAAS에서 제공하는 <자전거(피해운전자) 교통사고>와 <자전거(가해운전자) 교통사고> 데이터를 활용하여 새로이 데이터를 만들었다. 아래는 전처리 전 자전거 교통사고 피해자&가해자 데이터와 이를 시각화한 그래프들이다.

## 자전거(피해운전자) 교통사고

시도	발생건수	사망자수	부상자수	(중상자수)
서울	1,490	11	1,559	496
부산	324	4	330	157
대구	677	7	690	277
인천	226	3	230	94

도별 자전거사고(피해자) 정보



## 자전거(가해운전자) 교통사고

시도	발생건수	사망자수	부상자수	(중상자수)
서울	1,880	18	1,988	373
부산	265	1	277	65
대구	552	2	577	119
인천	194	3	201	40
광주	180	4	193	44



그리고 census에서 제공하는 <2018년 인구주택 총조사 등록센서스 방식 집계결과 부록표 (배포용)> 데이터에서 시도별로 다시 정리했다.

1. 시도별 인구, 가구 및 주택						
(단위: 명, 가구, 호, %)						
시 도	인 구	구성비	가 구	구성비	주 택	구성비
전 국	51,629,512	100.0	20,499,543	100.0	17,633,327	100.0
읍 부	4,984,023	9.7	1,942,637	9.5	1,862,018	10.6
면 부	4,730,290	9.2	2,030,579	9.9	2,066,655	11.7
동 부	41,915,199	81.2	16,526,327	80.6	13,704,654	77.7
수도권	25,713,241	49.8	10,037,990	49.0	8,061,137	45.7
특광역시	22,913,539	44.4	9,205,293	44.9	7,406,048	42.0
서울특별시	9,673,936	18.7	3,981,741	19.4	2,894,078	16.4
부산광역시	3,395,278	6.6	1,378,164	6.7	1,220,782	6.9

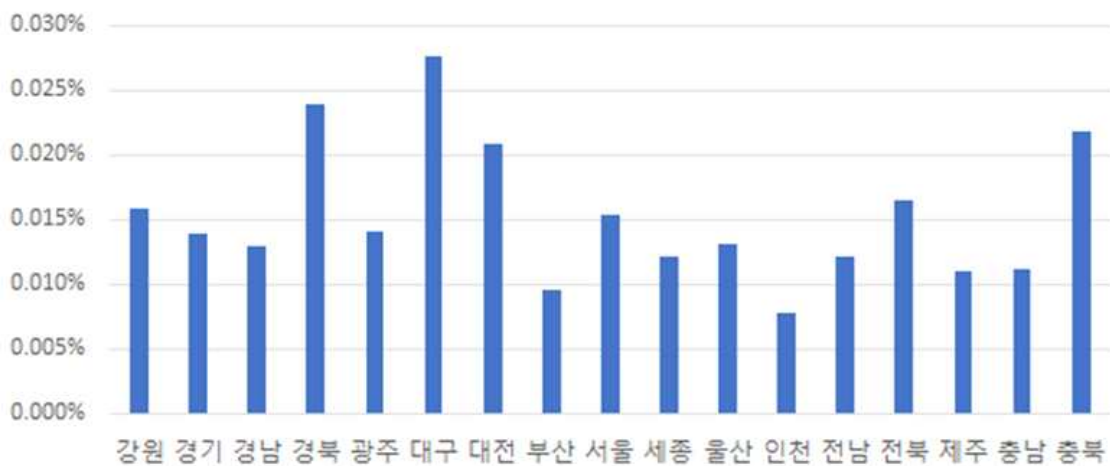
앞에서 처리한 인구 데이터를 이용, 시도별 사고 데이터에  $\frac{\text{발생건수}}{\text{인구}}$ 를 함수 기능으로 계산하여 열을 추가했다. 순서대로 전처리된 피해자, 가해자 데이터이다.

시도	발생건수	사망자수	부상자수	(중상자수)	인구	발생건수/인구
서울	1,490	11	1,559	496	9,673,936	0.000154022
부산	324	4	330	157	3,395,278	9.54266E-05
대구	677	7	690	277	2,444,412	0.000276958
인천	226	3	230	94	2,936,117	7.69724E-05
광주	209	5	210	66	1,490,092	0.00014026
대전	315	4	319	102	1,511,214	0.000208442
울산	151	7	148	57	1,150,116	0.000131291

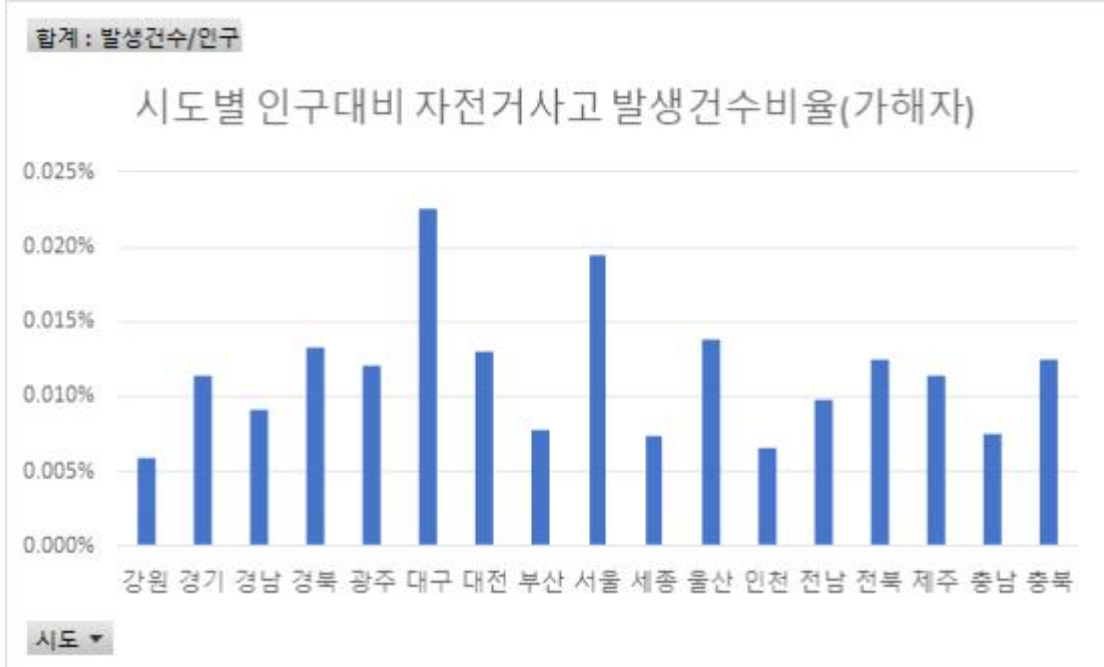
시도	발생건수	사망자수	부상자수	(중상자수)	인구	발생건수/인구
서울	1,880	18	1,988	373	9,673,936	0.000194337
부산	265	1	277	65	3,395,278	7.80496E-05
대구	552	2	577	119	2,444,412	0.000225821
인천	194	3	201	40	2,936,117	6.60737E-05
광주	180	4	193	44	1,490,092	0.000120798

우선 전처리 된 피해자 데이터에서, 시도와 발생건수/인구 열로 시도별 인구대비 자전거 사고 발생건수비율 그래프를 그렸다. 여기서 대구가 인구대비 자전거사고가 가장 많이 발생한 지역임을 알 수 있었다.

시도별 인구대비 자전거사고 발생건수비율(피해자)



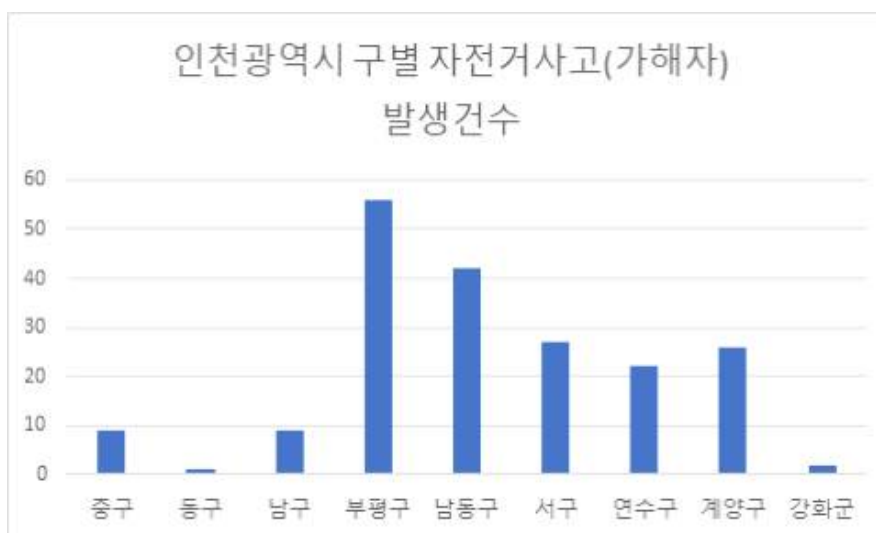
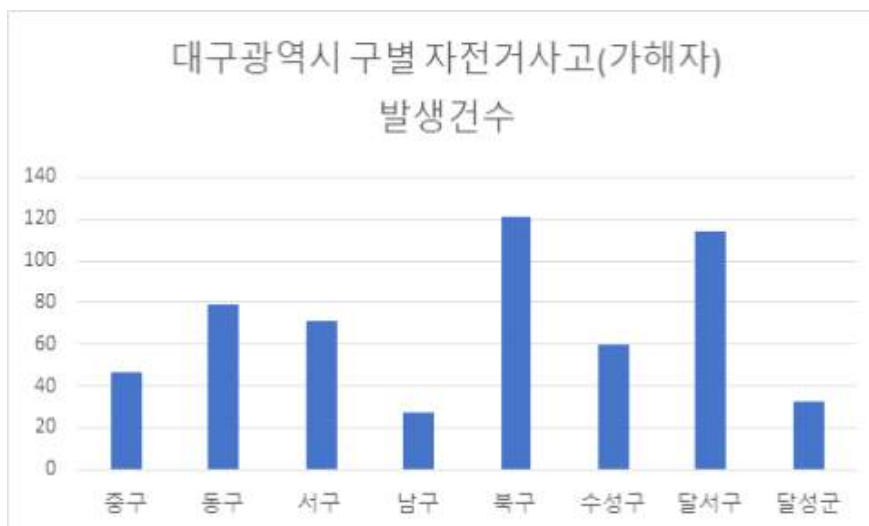
가해자 데이터 또한 같은 방식으로 나타내었을 때, 대구가 인구대비 자전거사고 발생비율이 가장 높은 지역이었다.



시도별 자전거사고 발생건수를 분석한 뒤, 시 단위로 구별 자전거사고 발생건수들을 따로 분석해보았으나 여러 시를 막대그래프로 나타내 본 결과, 이 이후로는 지도 위로 시각화하여 분석하는 편이 더 이해에 도움이 될 것이라 판단했다.









자전거 사고를 시각적으로 이해하기 쉽도록 표현하기 위해 위치정보를 기반으로 시각화해주는 파이썬의 folium 라이브러리를 사용하였다. 이후 진행되는 히트맵 시각화들은 기본적으로 도로교통공단에서 제공하는 <전국 교통사고 다발지역> 데이터를 지역별로 나누어 위도, 경도 데이터들을 이용했다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
사고지역관리번호	사고년도	사고유형구분	위치코드	시도시군구	사고지역위칭	발생건수	사상자수	사망자수	중상자수	경상자수	부상자수	위도	경도
2019038	2018	자전거	11110001	서울특별시	서울특별시 종로구 종로6가(출원지문사거리 부근)	4	5	0	1	1	3	37.57123	127.00917
2019038	2018	자전거	11170001	서울특별시	서울특별시 용산구 서빙고동(반포대교북단교차로 부근)	4	5	0	2	2	1	37.51928	126.99438
2019038	2018	자전거	11170002	서울특별시	서울특별시 용산구 이촌동(이촌한강공원 부근)	4	6	0	1	3	2	37.51754	126.96881
2019038	2018	자전거	11200001	서울특별시	서울특별시 성동구 용답동(덕현해밀아파트 부근)	5	6	0	3	2	1	37.56473	127.05346
2019038	2018	자전거	11200002	서울특별시	서울특별시 성동구 성수동1가(성원중학교 부근)	5	5	0	2	3	0	37.53982	127.05146
2019038	2018	자전거	11200003	서울특별시	서울특별시 성동구 용답동(성동구립용답도서관 부근)	4	4	0	3	1	0	37.56753	127.05121
2019038	2018	자전거	11200004	서울특별시	서울특별시 성동구 하왕십리동(세경교회 부근)	4	4	0	2	2	0	37.5696	127.0303
2019038	2018	자전거	11200005	서울특별시	서울특별시 성동구 용답동(서울교통공사사거리 부근)	4	4	0	1	2	1	37.56254	127.05806

우선, 앞의 그래프에서 대구가 인구대비 사고 발생율이 제일 높았으므로 대구의 heatmap을 그렸다. 우선 <전국 교통사고 다발지역>데이터를 전처리하여 1219daegu.xlsx로 저장하여 사용한다. 심각성 제고를 위한 변수로 새로이

$$dangerrate = \frac{(\text{중상자수}) + (\text{사망자수})}{(\text{사상자수}) + (\text{사망자수}) + (\text{중상자수})} \text{를 생성하였다.}$$

year	위치코	시도시군구명	발생건	사상자	사망자	중상자	위도	경도	dangerrate
2012	27200001	대구광역시 남구	4	10	0	4	35.84491	128.5906	0.4
2012	27200002	대구광역시 남구	5	6	0	1	35.85372	128.5901	0.166666667
2012	27290001	대구광역시 달서구	4	4	0	1	35.81469	128.5348	0.25
2012	27290002	대구광역시 달서구	4	4	0	1	35.82653	128.5477	0.25
2012	27290003	대구광역시 달서구	4	4	0	2	35.83824	128.5529	0.5
2012	27290004	대구광역시 달서구	4	4	0	2	35.83718	128.5484	0.5
2012	27290005	대구광역시 달서구	4	4	0	2	35.84959	128.4917	0.5
2012	27290006	대구광역시 달서구	4	4	0	3	35.84367	128.5302	0.75
2012	27290007	대구광역시 달서구	4	4	0	3	35.80843	128.547	0.75

다음은 데이터를 주피터 노트북으로 가져온 화면이다.

```
In [87]: import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns
import folium

import platform

%matplotlib inline

df = pd.read_excel('1219daegu.xlsx')
df.head()
```

Out[87]:

	year	위치코드	시도시군구 명	발생 건수	사상 자수	사망 자수	중상 자수	위도	경도	dangerrate
0	2012	27200001	대구광역시 남구	4	10	0	4	35.844914	128.590553	0.400000
1	2012	27200002	대구광역시 남구	5	6	0	1	35.853724	128.590113	0.166667
2	2012	27290001	대구광역시 달서구	4	4	0	1	35.814688	128.534829	0.250000
3	2012	27290002	대구광역시 달서구	4	4	0	1	35.826531	128.547714	0.250000
4	2012	27290003	대구광역시 달서구	4	4	0	2	35.838245	128.552858	0.500000

그리고 대구의 중상 내지 사망 사고 비율을 heat map으로 나타내었다. 범위가 넓기에 대구광역시를 화면에 꽉 채워 나타내기 위해 위도와 경도의 평균을 지도의 첫 구현 위치로 설정하고, zoom의 정도를 12로 설정하였다.

# 대구의 중상~사망 사고 비율 나타낸 히트맵

```
from folium.plugins import HeatMap
```

# define centre of map

```
TOR_COORDINATES = (df['위도'].mean(), df['경도'].mean())
```

# subset to match subset of locations

```
MAX_RECORDS = 60000
```

# create empty map zoomed in on Daegu

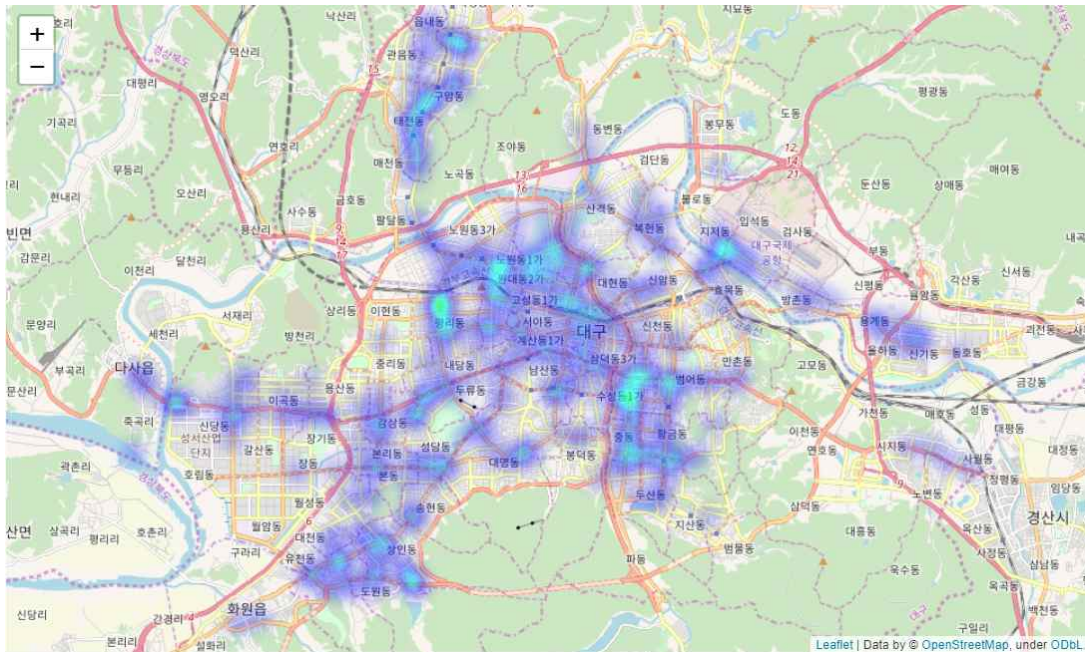
```
map_tor = folium.Map(location=TOR_COORDINATES, zoom_start=12)
```

# define heat map

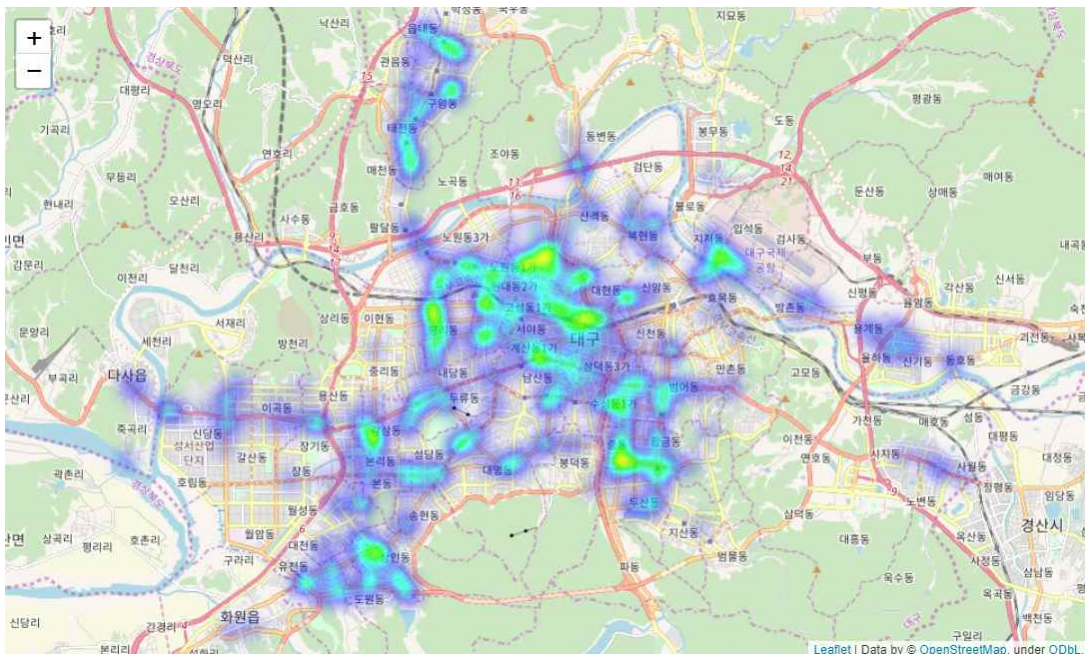
```
HeatMap(data=df[['위도', '경도', 'dangerrate']].groupby(['위도', '경도']).sum().reset_index().values.tolist(),
        radius=8, max_zoom=13).add_to(map_tor)
```

```
display(map_tor)
```

그 결과 다음과 같은 heat map이 구현된다. dangerrate가 높을수록 해당 사고 발생 지역이 밝은 색으로 표시되는 것을 확인할 수 있다.



dangerrate는 전체 사고를 대표할 수 없으므로 전체 사고 건수에 관한 히트맵도 그려준다. 따라서 앞의 코드 중 HeatMap data의 데이터 프레임에서 dangerrate만 빼주었다. 그러면 다음과 같은 그림을 얻는다.



이는 2012년도~2018년도의 모든 사고 발생 데이터를 나타낸 heat map 시각화 결과이므로, 연도별 사고 발생 현황을 알기 위해 따로 시각화하여 gif 파일로 나타내었다.

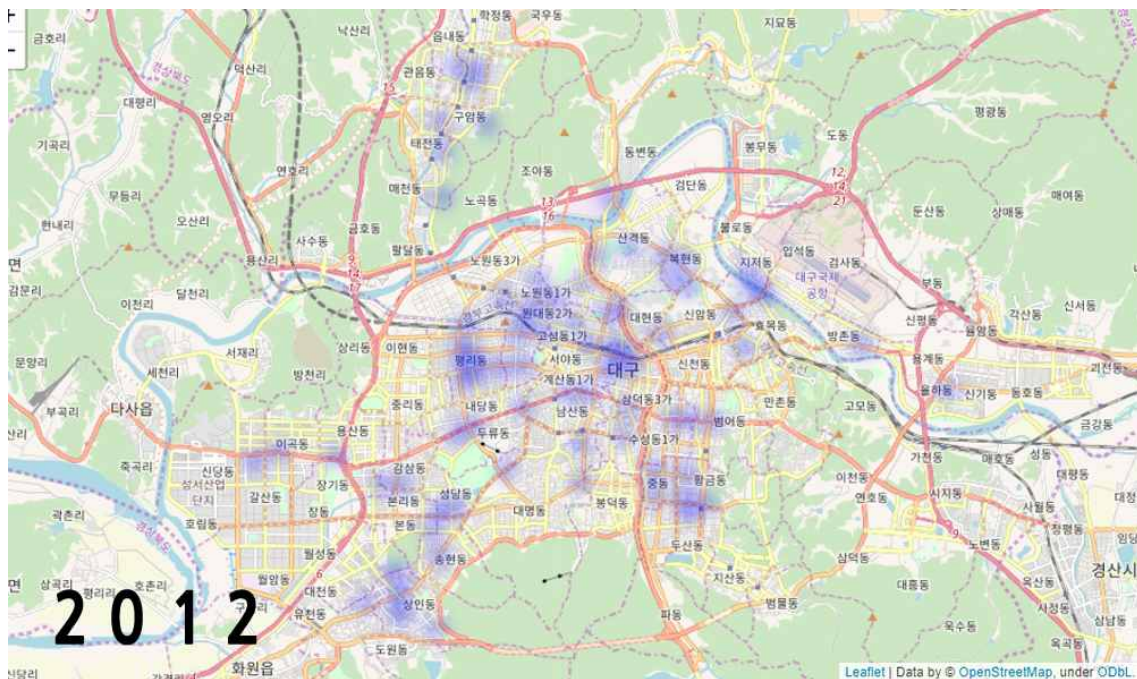
HeatMapWithTime을 이용하면 주피터 노트북 내에서 손쉽게 스크롤을 이용하여 연도별로 구현된 heat map을 나타낼 수 있지만, 발표 특성상 문서나 ppt의 자료로 쓰기에는 gif 파일이 적합할 것 같아 따로 연도별로 전처리하여 나타내었다.

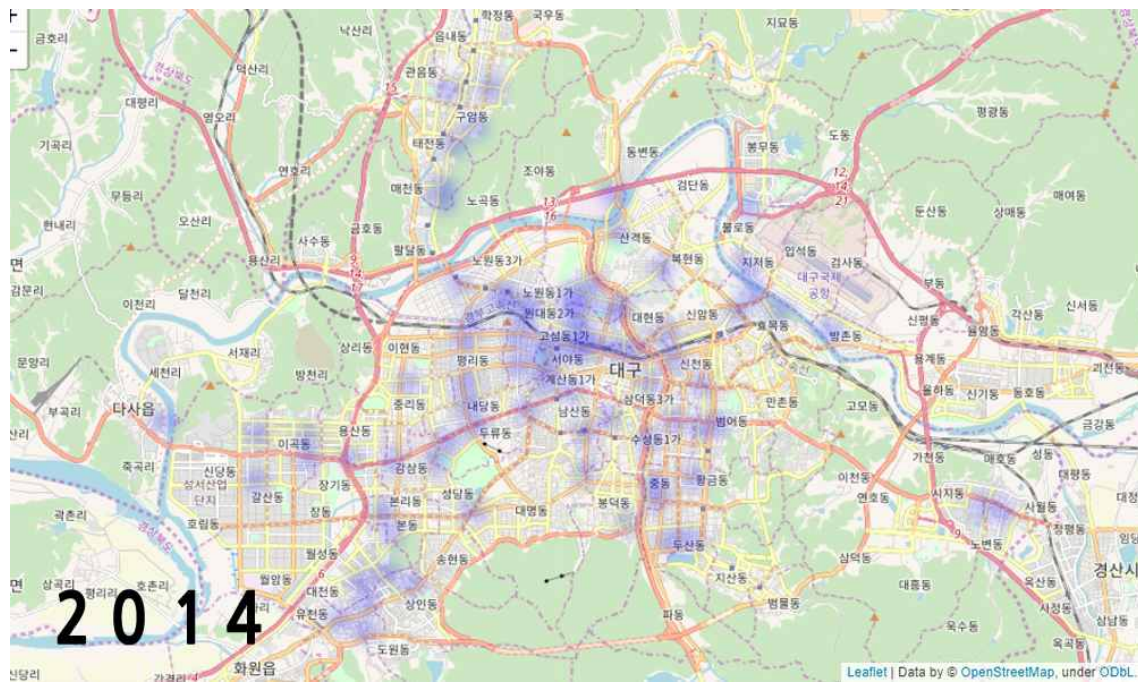
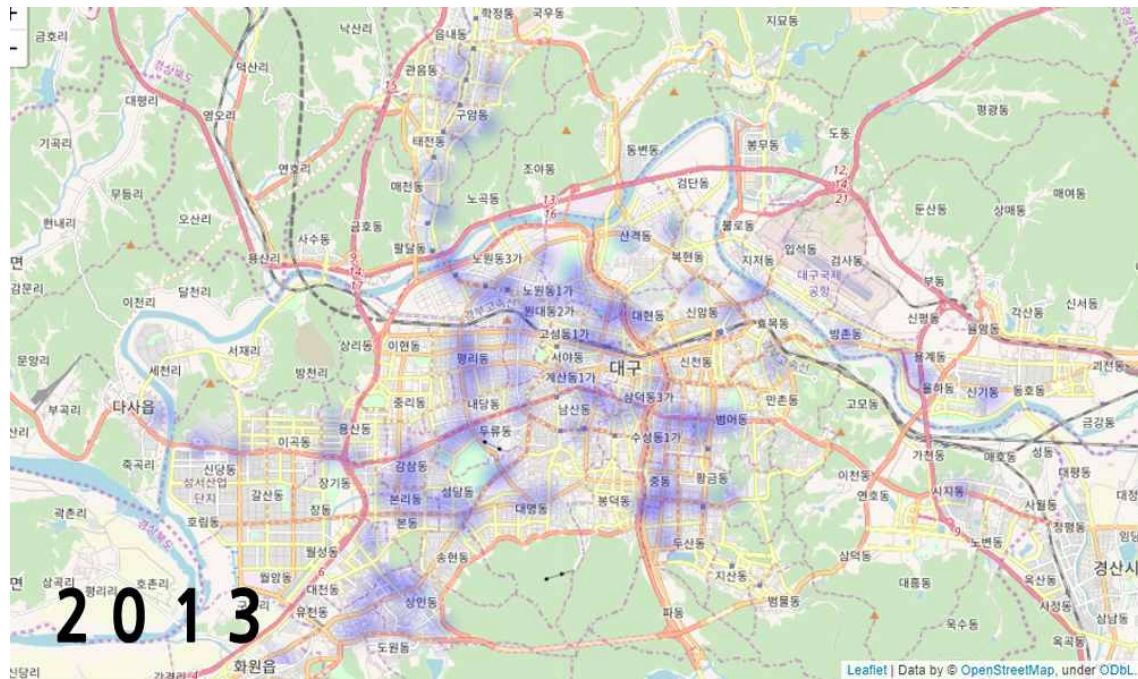


연도별로 나타낼 때 지도의 중심을 현재 위도와 경도의 평균으로 했으므로 조금씩 변화가 있어 중심의 위치를 적당한 값으로 고정시켰다.

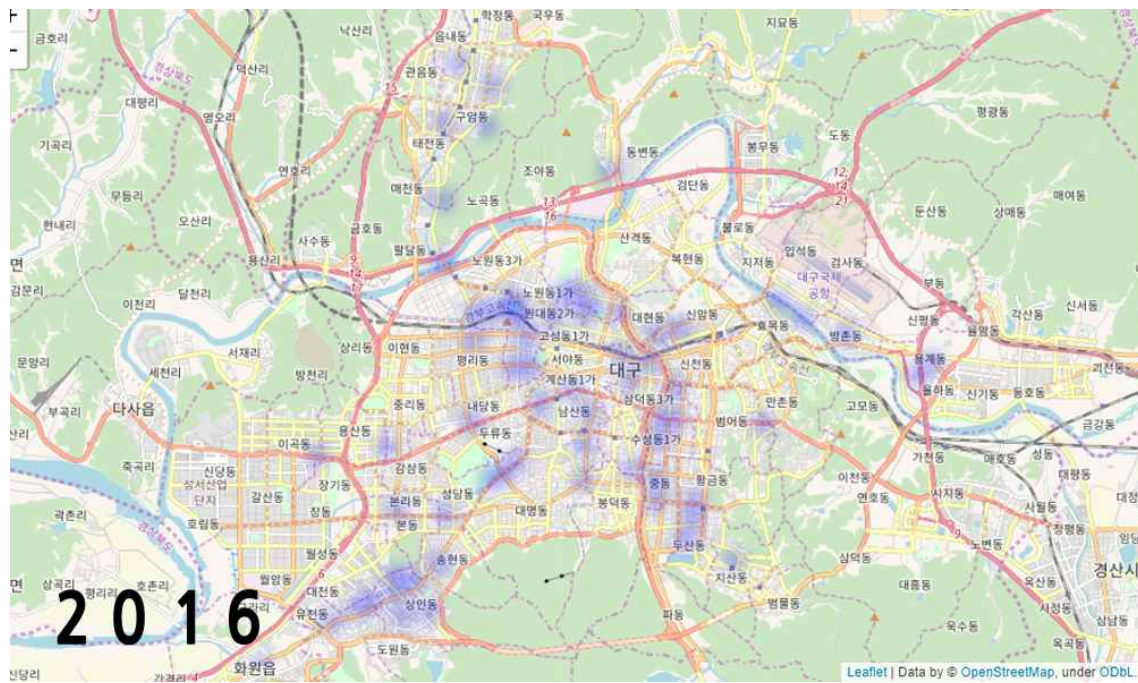
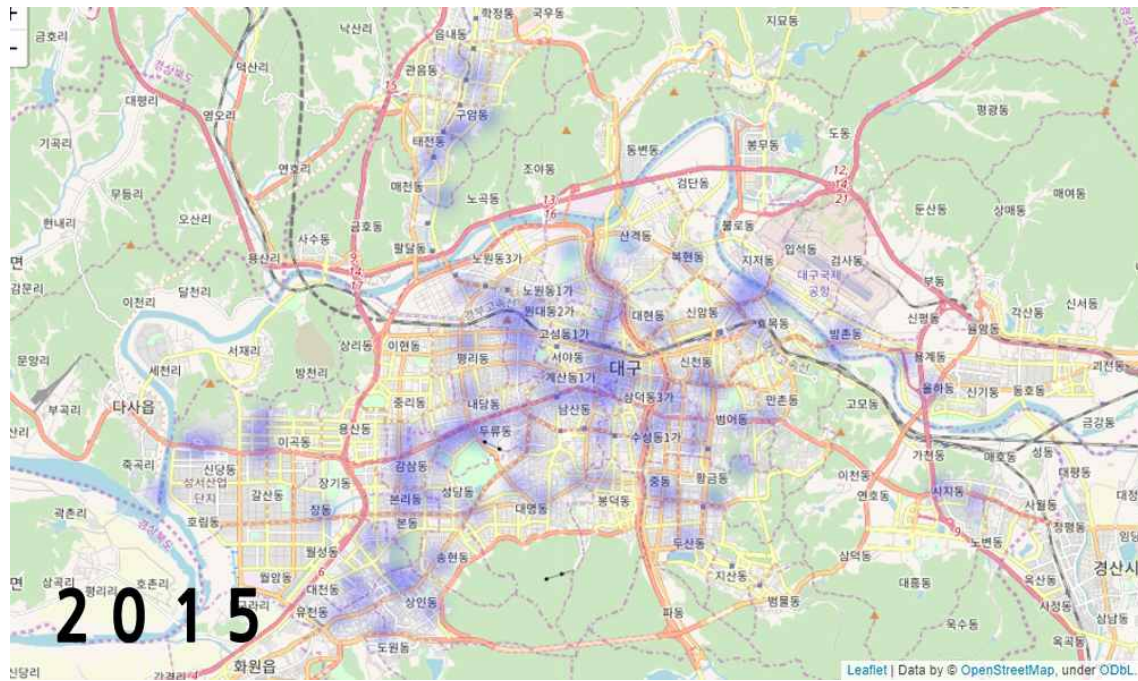
```
# 2012년
from folium.plugins import HeatMap
df = pd.read_excel('2012.xlsx')
# define centre of map
TOR_COORDINATES = (35.86, 128.583275222)
# subset to match subset of locations
MAX_RECORDS = 60000
# create empty map zoomed in on Daegu
map_tor = folium.Map(location=TOR_COORDINATES, zoom_start=12)
# define heat map
HeatMap(data=df[['위도', '경도']].groupby(['위도', '경도']).sum().reset_index().values.tolist(), radius=8, max_zoom=13).add_to(map_tor)
display(map_tor)
```

같은 방식으로 2018년도까지의 자료를 시각화한 결과는 다음과 같다. 좌측 하단의 연도를 나타내는 숫자는 따로 이미지 편집 툴을 이용하여 넣었다.

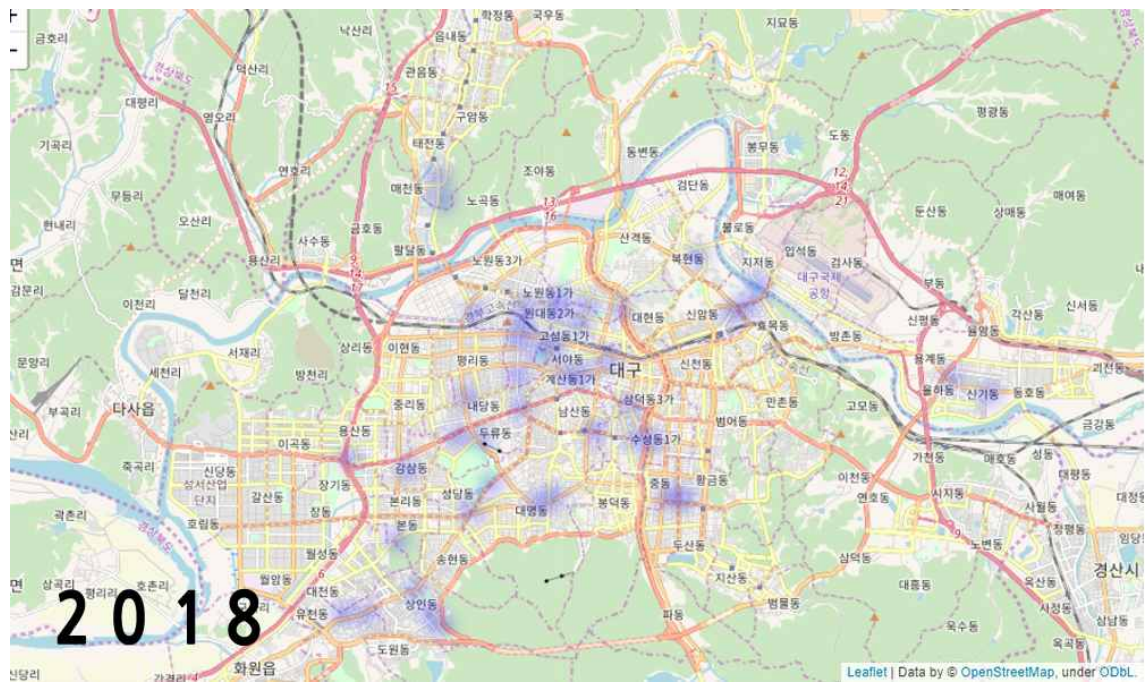
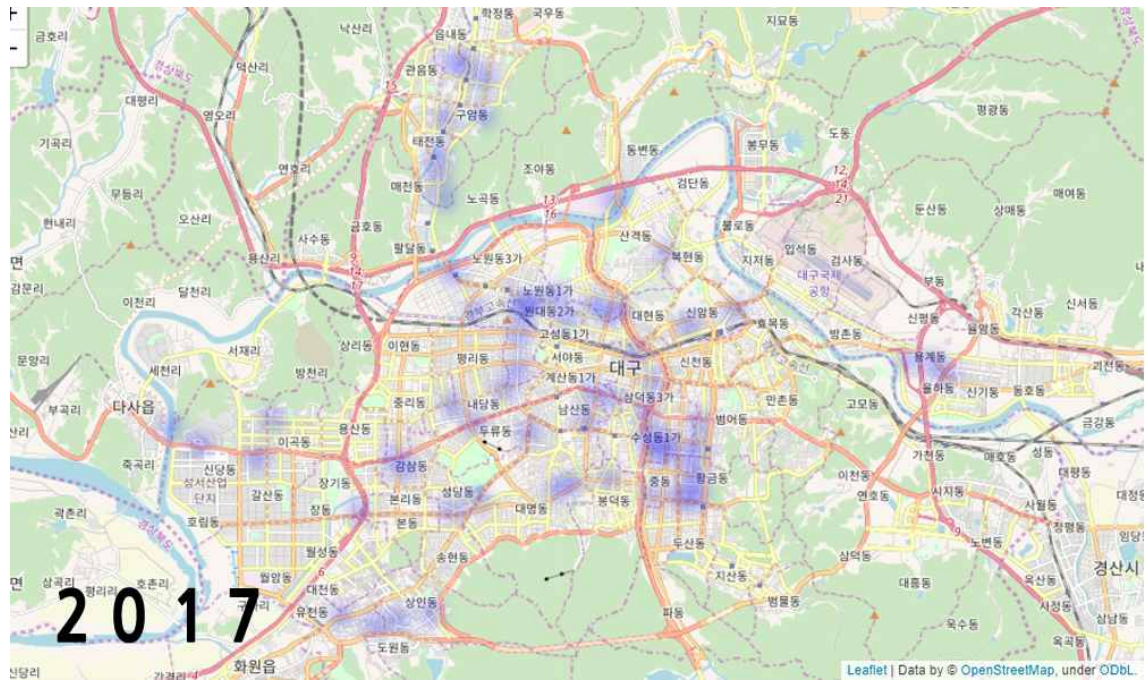






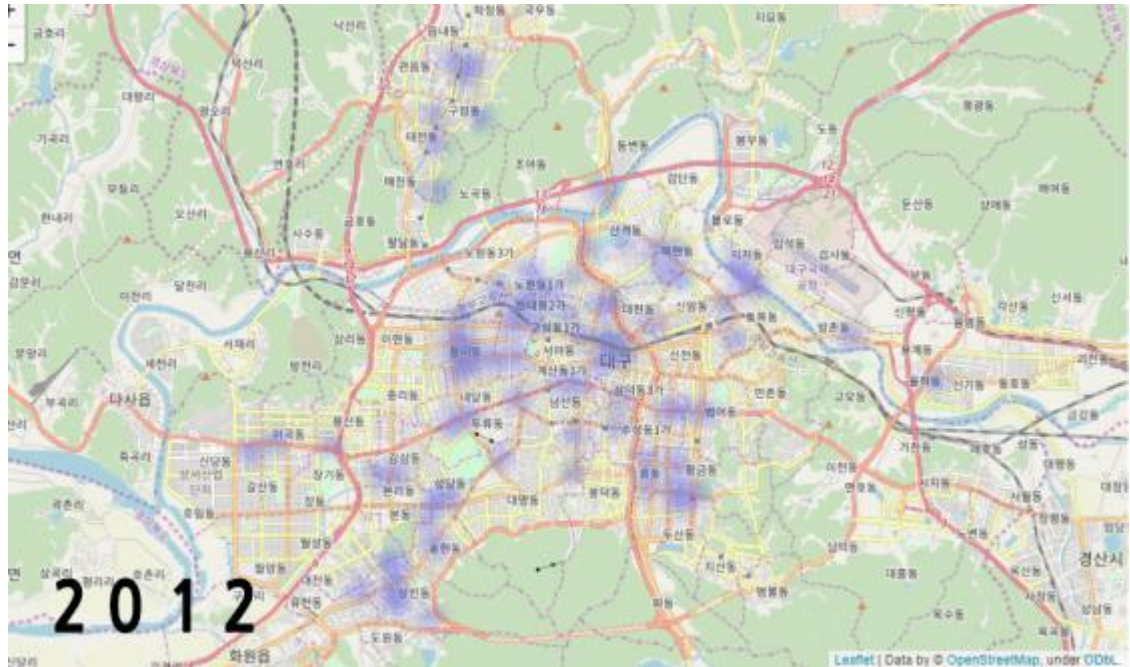




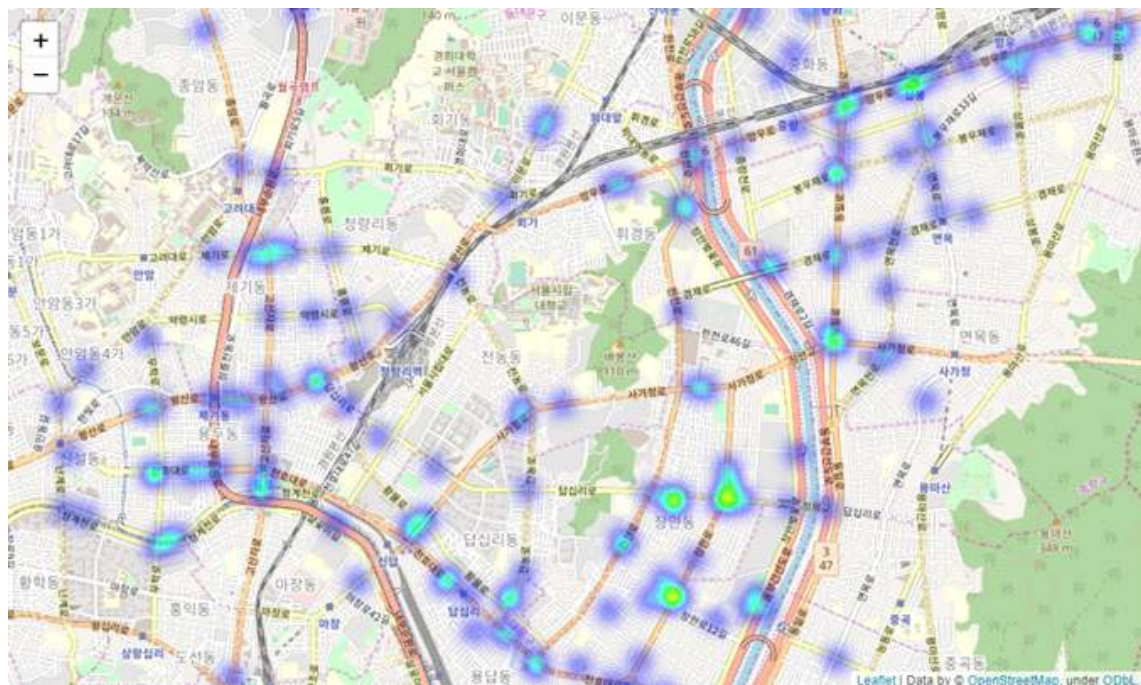




이를 영상 편집 툴을 이용하여 gif로 변환한 것이 아래와 같다.



여기에서 우리는 자전거 교통사고 다발지역과 발생지점을 파악할 수 있다. 범위를 전국으로 늘리면 다른 지역 또한 확인 가능하다. 다음은 청량리~회기 부근의 자전거 사고를 나타낸 것이다. 앞에서 분석한 바와 같이 교차로와 횡단보도 및 그 부근의 사고가 많은 것을 대략적으로 확인 가능하다.





## <로지스틱 회귀분석>

본 데이터 분석에서는 TAAS(Traffic Accident Analysis System) 교통사고분석시스템의 2018년 서울특별시에서 일어난 자전거 가해운전자, 피해운전자 데이터를 활용하였다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
사고번호	발생일자	발생요일	발생시간	사고내용	사망자수	중상자수	경상자수	부상전교자수	사고유형	행위유형	노면상태	기상상태	도로형태	가해운전자 차종	가해운전자 성별	가해운전자 연령	가해운전자 생년월도	피해운전자 차종	피해운전자 성별	피해운전자 연령	피해운전자 생년월도
2018년 1월 1일 08:41	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	44세	1969.01	승용	남	42세	1968.01
2018년 1월 1일 09:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 자전거	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	60세	1958.01	승용	남	79세	1959.01
2018년 1월 1일 09:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	여	56세	1962.01	승용	여	64세	1963.01
2018년 1월 1일 10:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	61세	1957.01	승용	남	17세	1999.01
2018년 1월 1일 11:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	여	36세	1982.01	승용	여	22세	1996.01
2018년 1월 1일 12:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	52세	1966.01	승용	남	62세	1963.01
2018년 1월 1일 13:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	여	48세	1970.01	승용	여	51세	1969.01
2018년 1월 1일 14:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	여	53세	1965.01	승용	여	84세	1954.01
2018년 1월 1일 15:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	48세	1970.01	승용	남	11세	1997.01
2018년 1월 1일 16:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	62세	1956.01	승용	남	23세	1995.01
2018년 1월 1일 17:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	71세	1947.01	승용	남	16세	1998.01
2018년 1월 1일 18:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	63세	1955.01	승용	남	48세	1969.01
2018년 1월 1일 19:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	51세	1967.01	승용	남	55세	1962.01
2018년 1월 1일 20:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	여	39세	1979.01	승용	여	39세	1979.01
2018년 1월 1일 21:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	67세	1951.01	승용	남	78세	1949.01
2018년 1월 1일 22:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	49세	1969.01	승용	남	48세	1970.01
2018년 1월 1일 23:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	48세	1970.01	승용	남	88세	1930.01
2018년 1월 2일 00:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	17세	1997.01	승용	남	32세	1985.01
2018년 1월 2일 01:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	51세	1966.01	승용	남	23세	1995.01
2018년 1월 2일 02:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	23세	1995.01	승용	남	57세	1961.01
2018년 1월 2일 03:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	36세	1982.01	승용	남	60세	1958.01
2018년 1월 2일 04:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	여	29세	1989.01	승용	여	26세	1992.01
2018년 1월 2일 05:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	67세	1953.01	승용	남	44세	1974.01
2018년 1월 2일 06:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	69세	1951.01	승용	남	62세	1956.01
2018년 1월 2일 07:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	남	63세	1956.01	승용	남	59세	1959.01
2018년 1월 2일 08:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	63세	1956.01	승용	남	37세	1981.01
2018년 1월 2일 09:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	22세	1996.01	승용	남	58세	1960.01
2018년 1월 2일 10:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	62세	1956.01	승용	남	77세	1941.01
2018년 1월 2일 11:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	62세	1956.01	승용	남	59세	1959.01
2018년 1월 2일 12:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	여	29세	1989.01	승용	여	70세	1948.01
2018년 1월 2일 13:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	여	40세	1978.01	승용	여	60세	1958.01
2018년 1월 2일 14:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 기타	신호무인	보통 - 건조	맑음	인월로 - 기타	승용	여	42세	1976.01	승용	여	21세	1997.01
2018년 1월 2일 15:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	43세	1975.01	승용	남	59세	1960.01
2018년 1월 2일 16:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	48세	1970.01	승용	남	22세	1996.01
2018년 1월 2일 17:51	월요일	서울특별시 양천구 목동	경상사고	0	0	1	0	0	자전거 - 승용	신호무인	보통 - 건조	맑음	교차로 - 교차로부근	승용	남	48세	1970.01	승용	남	58세	1960.01

자전거 사고에 영향을 미치는 독립변수들은 다음과 같다.

변수명	변수 설명
성별	남성-1, 여성-0
연령	연령
법률 위반	교통법규 위반 유-1, 무-0
도로종류	교차로-1, 기타-0
날씨	맑음-0, 기타-1
시간	주간-0, 야간-1
노면상태	건조-0, 기타-1

자전거 사고의 심각도를 나타내는 종속변수는 사망자나 중상자가 있으면 '심각도 높음-1'으로, 나머지 경우를 '심각도 낮음-1'으로 분류했다.

분석에 적절하게 데이터 전처리를 실시하였다.

```
import csv
f = open('attacker.csv')
data = csv.reader(f)
next(data)

severity=[] #사망자나 중상자가 있으면 심각도 높음-1, 아니면 심각도 낮음-0

for row in data:
    if int(row[5])+int(row[6]) >= 1:
        severity.append(1)
    else:
        severity.append(0)
```

row[5]는 사망자 수고, row[6]은 중상자 수이다. int()를 사용해 정수형으로 바꿨다.

```
sex=[] #자전거 이용자의 성별(남-1, 여-0)
for row in data:
    if row[15] == "남":
        sex.append(1)
    else:
        sex.append(0)
```

```
sex=[] #자전거 이용자의 성별(남-1, 여-0)
for row in data:
    if row[19] == "남":
        sex.append(1)
    else:
        sex.append(0)
```

자전거 가해운전자 데이터

자전거 피해운전자 데이터

row[15]는 자전거 가해운전자의 성별이다. row[19]는 자전거 피해운전자의 성별이다. 남이면 1로, 여이면 0으로 분류했다.

```
age=[] #자전거 이용자의 연령
for row in data:
    number = int( row[16].replace("세", ""))
    age.append(number)
```

row[16]은 자전거 이용자의 연령이다. 세를 빈칸으로 바꾸기 위해 replace() 함수를 사용했다. 그리고 int()를 사용해 정수형으로 바꿨다.

```
law_violation=[] #자전거 이용자의 교통법규 위반 유.무(유-1, 무-0)
for row in data:
    if row[10] in ("교차로운행방범위반", "보행자보호의무위반", "불법유턴", "신호위반", "중앙선침범", "직진우회전진행방해", "차로위반"):
        law_violation.append(1)
    else:
        law_violation.append(0)
```

row[10]은 법규위반 내용이다. row[10]에 교차로운행방범위반, 보행자보호의무위반, 불법유턴, 신호위반, 중앙선침범, 직진우회전진행방해, 차로위반이 있으면 1로, 그렇지 않으면 0으로 분류했다.

```
road_type=[] #도로종류 (교차로-1, 나머지-0)
```

```
for row in data:
    if "교차로" in row[13]:
        road_type.append(1)
    else:
        road_type.append(0)
```

row[13]은 도로종류이다. 도로종류가 교차로이면 1로, 그렇지 않으면 0으로 분류했다.

```
weather=[] #(맑음-0, 나머지-1)
```

```
for row in data:
    if row[12] == "맑음":
        weather.append(0)
    else:
        weather.append(1)
```

row[12]는 날씨이다. 날씨가 맑으면 0으로, 그렇지 않으면 1로 분류했다.

```
time=[] #(주간(오전7시~오후6시)-0, 야간(오후7시~오전6시)-1)
```

```
for row in data:
    hour = row[1].split(' ')[3].replace("시", "")
    if hour in ("19", "20", "21", "22", "23", "00", "01", "02", "03", "04", "05", "06"):
        time.append(1)
    else:
        time.append(0)
```

row[1]은 사고 발생일시이다. row[1]의 형태는 'ㅇㅇㅇㅇ년 ㅇ월 ㅇ일 ㅇㅇ시'여서 공백으로 split()를 사용해 문자열을 분리했다. row[1][0]은 ㅇㅇㅇㅇ년, row[1][1]은 ㅇ월, row[1][2]는 ㅇ일, row[1][3]은 ㅇㅇ시이다. 오후 7시에서 오전 6시를 야간으로 분류했고, 오전 7시에서 오후 6시는 주간으로 분류했다.

```
roadbed=[] #노면상태(건조-0, 기타-1)
```

```
for row in data:
    if "건조" in row[11]:
        roadbed.append(0)
    else:
        roadbed.append(1)
```

row[11]은 노면상태이다. 노면상태가 건조이면 0으로, 그렇지 않으면 1로 분류했다.

데이터 프레임은 분석을 수행할 때 가장 빈번히 사용되는 핵심적인 형식으로서, 숫자, 문자 등 다양한 데이터를 하나의 테이블에 담을 수 있는 자료구조이다. 분석을 위해 리스트를 데이터프레임으로 변환했다.

```
df = pd.DataFrame(data={"severity":severity, "sex":sex, "age":age, "law_violation":law_violation, "road_type":road_type, "weather":weather, "time":time, "roadbed":roadbed})
print(df)
```

	severity	sex	age	law_violation	road_type	weather	time	roadbed
0	1	0	48	1	0	0	0	0
1	0	1	67	0	0	0	0	0
2	0	0	72	0	1	0	0	0
3	0	1	33	1	1	0	0	0
4	0	0	24	1	1	0	1	0
5	1	1	62	1	0	0	0	0
6	1	0	57	0	0	0	0	0
7	1	1	23	0	0	0	1	0
8	0	1	57	0	1	0	1	0
9	0	1	71	1	1	0	1	0
10	0	0	60	1	1	0	0	0
11	1	0	51	1	1	0	0	0
12	0	1	52	0	0	0	0	0
13	0	0	61	0	0	0	0	0
14	0	1	59	0	0	0	1	0
15	0	1	76	0	1	0	1	0
16	0	1	56	0	0	0	0	0
17	0	0	63	0	0	0	0	1
18	0	1	14	0	0	0	0	0
19	0	0	55	0	0	0	0	0
20	0	1	55	0	0	0	0	0
21	0	1	56	1	1	0	1	0
22	1	1	22	0	1	0	1	0
23	1	0	52	1	0	0	0	0
24	1	1	66	1	1	0	0	0
25	0	1	41	0	0	0	1	0
26	1	0	69	0	1	0	0	0
27	0	1	86	1	1	1	0	1
28	0	1	44	0	0	0	0	0

로지스틱 회귀 분석을 적용하기에 앞서 다음과 같이 모델을 만들 training set과 모델을 평가할 test set을 7:3 비율로 나눴다.

```
#독립변수와 종속변수를 지정해주고 훈련용 데이터와 검증용 데이터를 7:3으로 분할해준다.
x = df[['sex', 'age', 'law_violation', 'road_type', 'weather', 'time', 'roadbed']] #변수 지정
y = df[['severity']]
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=0)
```

자전거 가해운전자 데이터에서 학습용 데이터셋 정확도는 0.75로 나왔고, 검증용 데이터셋 정확도 또한 0.75로 나왔다. 자전거 피해운전자 데이터에서 학습용 데이터셋 정확도는 0.63로 나왔고, 검증용 데이터셋 정확도는 0.69로 나왔다.

```
from sklearn.linear_model import LogisticRegression

log = LogisticRegression() #로지스틱 회귀분석 시행

log.fit(x_train, y_train) #모델의 정확도 확인
print('학습용 데이터셋 정확도 : %.2f' % log.score(x_train, y_train))
print('검증용 데이터셋 정확도 : %.2f' % log.score(x_test, y_test))
```

학습용 데이터셋 정확도 : 0.75  
검증용 데이터셋 정확도 : 0.75

자전거 가해운전자 데이터

```
from sklearn.linear_model import LogisticRegression

log = LogisticRegression() #로지스틱 회귀분석 시행

log.fit(x_train, y_train) #모델의 정확도 확인
print('학습용 데이터셋 정확도 : %.2f' % log.score(x_train, y_train))
print('검증용 데이터셋 정확도 : %.2f' % log.score(x_test, y_test))
```

학습용 데이터셋 정확도 : 0.63  
검증용 데이터셋 정확도 : 0.69

자전거 피해운전자 데이터

조금 더 자세한 평가를 위해 classification\_report 모듈을 이용하였다. 이 모듈은 모형 성능평가를 위한 모듈로 정밀도(precision), 재현율(recall), F1-score, support를 구해준다.

```
from sklearn.metrics import classification_report
y_pred=log.predict(x_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.75	1.00	0.86	327
1	0.00	0.00	0.00	110
accuracy			0.75	437
macro avg	0.37	0.50	0.43	437
weighted avg	0.56	0.75	0.64	437

자전거 가해운전자 데이터

```
from sklearn.metrics import classification_report
y_pred=log.predict(x_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.71	0.95	0.81	305
1	0.42	0.08	0.14	130
accuracy			0.69	435
macro avg	0.57	0.52	0.48	435
weighted avg	0.62	0.69	0.61	435

자전거 피해운전자 데이터

모델의 회귀계수와 오즈비를 구해 독립변수가 분류 결정에 미치는 영향의 정도를 알아보기 위해 다른 방식의 로지스틱회귀분석을 진행했다. logit이라는 변수에 종속변수로 하는 데이터를 넣고, 앞서 넣은 독립변수 x값도 입력한 뒤 적합 시켜줬다.

```
import statsmodels.api as sm
logit = sm.Logit(df[['severity']],x) #로지스틱 회귀분석 시행
result = logit.fit()
```

```
Optimization terminated successfully.
Current function value: 0.570209
Iterations 5
```

각 회귀계수가 사고 심각도에 어떻게 영향을 미치는지 확인했다. 먼저, 유의확률을 봤다. 자전거 가해운전자 데이터에서는 성별, 나이, 도로형태만 유의수준 95%에서 종속변수에 영향을 주는 유의한 변수임을 알 수 있다. 자전거 피해운전자 데이터에서는 성별, 시간만 유의수준 95%에서 종속변수에 영향을 주는 유의한 변수임을 알 수 있다.

Coef.(편회귀계수)는 한 독립변수 외의 다른 독립변수들이 모두 일정한(조정된, 통제된) 상태에서 한 독립변수가 1 단위 변할 때 종속변수가 얼마나 변하는 가를 알려준다. Coef. 부호를 통해 종속변수에 미치는 영향의 방향을 파악할 수 있다. Coef.의 값이 양수이면 그 컬럼의 값이 커질수록 종속변수가 TRUE일 확률이 높아진다는 뜻이고, 반대로 Coef.의 값이 음수이면 그 컬럼의 값이 커질수록 종속변수가 FALSE일 확률이 높아진다는 뜻이다.

자전거 가해운전자 데이터에서 Coef.의 부호만을 봤을 때, 법규위반을 할수록, 날씨가 맑지 않을수록, 시간대가 야간일 경우 사고 심각도에 큰 영향을 미치는 것을 알 수 있다. 자전거 피해운전자 데이터에서 Coef.의 부호만을 봤을 때, 나이가 많을수록, 법규위반을 할수록, 노면상태가 건조하지 않을수록, 사고 심각도에 큰 영향을 미치는 것을 알 수 있다.

result.summary2()							result.summary2()						
Model:		Logit	Pseudo R-squared:		-0.018		Model:		Logit	Pseudo R-squared:		-0.009	
Dependent Variable:		severity	AIC:		1673.3094		Dependent Variable:		severity	AIC:		1897.0002	
Date:		2019-12-01 01:09	BIC:		1710.2887		Date:		2019-12-01 02:03	BIC:		1933.9410	
No. Observations:		1455	Log-Likelihood:		-829.65		No. Observations:		1447	Log-Likelihood:		-941.50	
Df Model:		6	LL-Null:		-815.16		Df Model:		6	LL-Null:		-933.44	
Df Residuals:		1448	LLR p-value:		1.0000		Df Residuals:		1440	LLR p-value:		1.0000	
Converged:		1.0000	Scale:		1.0000		Converged:		1.0000	Scale:		1.0000	
No. Iterations:		5.0000						No. Iterations:		5.0000			
	Coef.	Std.Err.	z	P> z	[0.025	0.975]		Coef.	Std.Err.	z	P> z	[0.025	0.975]
sex	-0.5978	0.1093	-5.4711	0.0000	-0.8119	-0.3836	sex	-0.5843	0.1016	-5.7493	0.0000	-0.7835	-0.3851
age	-0.0077	0.0020	-3.8798	0.0001	-0.0117	-0.0038	age	0.0022	0.0018	1.2183	0.2231	-0.0014	0.0058
law_violation	0.0609	0.1483	0.4108	0.6812	-0.2297	0.3515	law_violation	0.0699	0.1310	0.5334	0.5938	-0.1869	0.3266
road_type	-0.5903	0.1302	-4.5344	0.0000	-0.8454	-0.3351	road_type	-0.1823	0.1132	-1.6099	0.1074	-0.4041	0.0396
weather	0.0901	0.2729	0.3303	0.7412	-0.4448	0.6250	weather	-0.3041	0.2610	-1.1651	0.2440	-0.8157	0.2075
time	0.0448	0.1318	0.3398	0.7340	-0.2136	0.3032	time	-0.3287	0.1176	-2.7952	0.0052	-0.5592	-0.0982
roadbed	-0.1780	0.3095	-0.5752	0.5651	-0.7845	0.4285	roadbed	0.0351	0.3010	0.1165	0.9072	-0.5549	0.6251

자전거 가해운전자 데이터

자전거 피해운전자 데이터

오즈 비(Odds Ratio)를 출력했다. Odds Ratio란 Odds의 비율이다. Odds란 성공/실패와 같이 상호 배타적이며 전체를 이루고 있는 것들의 비율을 의미한다. 오즈비가 1을 기준으로 큰지 작은지를 파악하여 종속변수에 미치는 영향의 방향을 파악할 수 있다. 독립변수가 2개 이상 있을 때는 다른 독립변수를 일정한 값으로 고정한 경우의 오즈비로 해석된다. 아무런 관계가 없을 때 오즈비는 1이다. 1에서 멀리 떨어질수록 종속변수와의 관계가 강하다는 뜻이다. 즉, 종속변수 여부에 큰 영향을 준다는 뜻이다.

자전거 가해운전자 데이터에서 나이, 법규위반, 날씨, 시간 변수의 경우, 1에 가까운 값으로 나타난다. 따라서 나이, 법규위반, 날씨, 시간 변수는 별다른 영향을 주지 않았음(관계없음)을 알 수 있다. 반면, 성별, 도로형태 변수는 1과 떨어져 있으므로 사고 심각도에 영향을 미쳤음을 알 수 있다.

자전거 피해운전자 데이터에서 나이, 법규위반, 도로형태 변수의 경우, 1에 가까운 값으로 나타난다. 따라서 나이, 법규위반, 도로형태 변수는 별다른 영향을 주지 않았음(관계없음)을 알 수 있다. 반면, 성별 변수는 1과 떨어져 있으므로 사고 심각도에 영향을 미쳤음을 알 수 있다.

np.exp(result.params)		np.exp(result.params)	
sex	0.550042	sex	0.557500
age	0.992284	age	1.002220
law_violation	1.062796	law_violation	1.072366
road_type	0.554164	road_type	0.833388
weather	1.094330	weather	0.737794
time	1.045816	time	0.719846
roadbed	0.836931	roadbed	1.035695
dtype: float64		dtype: float64	

자전거 가해운전자 데이터

자전거 피해운전자 데이터



## Ⅲ. 예방 방안 도출

지금까지 한 데이터 분석을 바탕으로 '자전거 안전사고 예방 방안'을 도출하였다.

### 1) 우선적으로 개인의 안전의식 제고가 필요

앞에서 분석한 바와 같이 '자전거 가해/피해 운전자의 법규위반 사항' 중 64.2%로 가장 많이 차지하는 '안전운전 의무 불이행'에 집중하였다. '안전운전 의무 불이행' 사항에는 1)운전 중 휴대전화 사용, 2)전방 주시 태만, 3)운전 미숙, 4)안전모 미착용 등이 있다. 중앙선 침범, 신호위반, 과속과 같은 의도적 위반행위가 아닌 운전 집중하지 못하는 안전 불감증의 결과로 일어나는 사고라 할 수 있다. 자전거 운전자의 법규위반이 사고로 이어지는 경우가 대부분을 차지하는 만큼 자전거 이용자의 안전 교육과 안전의식 제고를 위한 홍보가 필요하다 할 수 있습니다.

### 2) 안전정보를 즉각적으로 제공할 수 있는 매체 제공

GPS 기능을 활용하여 최근 7년간 수집된 사고데이터에서 얻어낸 사고 다발지역에서 그 구간에서 주의해야 되는 사고 유형을 공공자전거의 디스플레이 패널이나 실행한 App에서 경고문을 띄우고 승용차와 충돌할 가능성이 높은 교차로와 보행자와 충돌할 가능성이 높은 단일로에서 전방을 주시하고, 과속을 하기 쉬운 구간에서 과속을 주의하라는 등의 안전정보를 제공하는 것이 운전자의 위험을 인지하게 하여 사고 예방을 하는데 도움을 줄 수 있다.



<따릉이의 디스플레이 패널>



<자전거 라이딩 어플(Open Rider)>

→GPS 기능을 이용하여 현재  
속도와 위치를 제공한다.



### 3) 자전거 사고 다발지역에 알맞은 안전 시설 설치

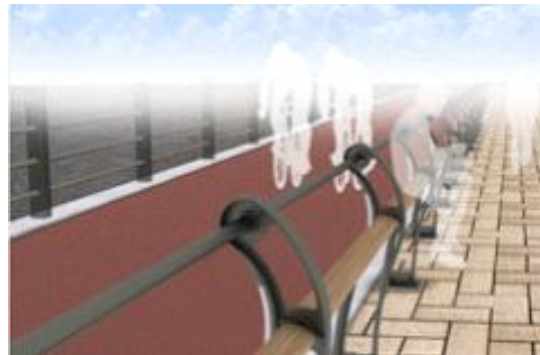


<[대림] 가로공원사거리>

→ 승용차와 충돌사고가 많이 발생하는 교차로에 자전거 전용 횡단보도를 설치하는 것이 사고 예방에 도움을 준다.



<[여의도] 한강 자전거길>



→ 자전거 도로와 인도, 자전거 도로와 일반 도로의 경계가 명확하지 않은 구간에 자전거 전용 난간을 설치하여 승용차와 보행자와의 충돌을 방지하고 자전거가 속도를 내는 구간에 과속방지턱을 설치하여 과속으로 인한 충돌을 방지한다.

#### 4) '자전거 도로 시급 지역'에 자전거 도로 설치

자전거 교통사고 다발지역, 승용차와 측면직각충돌이 일어나기 쉬운 교차로 등 사고예방 효과가 큰 지역을 선정하여 우선 구축하는 것이 자전거 사고 발생을 예방하는데 도움을 줄 수 있다.



<자전거 도로가 없는 방면에서만 사고가 발생한 경우>

## IV. 참고 문헌 및 데이터 출처

- 권성대, 김윤미, 김재곤, 하태준. (2015). 자전거 사고예측모형 개발 및 개선방안 제시에 관한 연구. 대한토목학회논문집, 35(5), 1135-1146.
- 김장락. (2018). 시군구 단위 자전거 헬멧 착용률과 자전거 사고율의 연관성. 보건교육건강증진학회지, 35(3), 15-24.
- 신혜정, 홍지연, 이수범. (2012). 자전거 사고 심각도에 영향을 미치는 요인 연구. 교통연구, 19(1), 67-78.
- 오주택, 김응철, 김도훈. (2007). 도시부 교차로에서의 자전거사고 심각도 분석에 관한 연구. 대한토목학회논문집 D, 27(4D), 389-395.
- 이성규, 박정희, 강대룡. (2018). 자전거도로 관리체계 구축방안. [CODIL] 건설보고서, 2018(5).
- 이유식, 황윤형, 성광모, 김성진, 이정욱. (2016). 교차로내 자전거 교통사고 유형 분석에 관한 연구. 한국자동차공학회 추계학술대회 및 전시회, 794-797.
- 최승희, 이구연. (2016). 자전거 교통 사고 현황 및 예측 분석. 전자공학회논문지, 53(9), 89-96.
- 홍종선, 김명진. (2010). 자전거 교통사고의 통계분석 및 활용. 한국데이터정보과학회지, 21(6), 1081-1090.

TAAS 교통사고분석시스템 <http://taas.koroad.or.kr/>

KOSIS 국가통계포털 [kosis.kr](http://kosis.kr)

인구주택총조사 <http://www.census.go.kr/>