

빅데이터 분석 및 시각화 개론

최종 보고서

19조

박영현 윤신웅 이동현

Data: 웹 크롤링

우리는 도로교통공단(TAAS)에서 제공하는 교통사고분석시스템을 이용하여 데이터를 수집하였다.

전체

검색

모두닫기

모두열기

초기화

공통메뉴

교통안전

교통사고(경찰DB)

교통사고(통합DB)

사고일반

합계

교통사고

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

교통사고(경찰DB)

교통사고(통합DB)

교통사고일반

이 사이트에서는 왼쪽 드롭다운 리스트에서 원하는 분류항목을 선택하고 위의 달력 아이콘을 클릭하여 원하는 기간을 설정한 뒤 데이터를 조회한 후 엑셀파일로 다운로드 받을 수 있도록 서비스한다. 그러나 이 데이터를 가지고 분석하기에는 몇 가지 문제가 있었다. 먼저 여러 분야가 한꺼번에 정리된 데이터가 없었다. TAAS에서는 요일별, 시간대별, 사고유형별, 법규위반별 등 다양한 분야 중 두 분야를 쌍을 지어 정리된 데이터를 제공하고 있었는데, 모든 분야에 대해 쌍이 지어지지 않아 분석에 제한이 있었다. (예를 들어 시간대별-차종별 데이터는 TAAS에서 제공되지 않았음) 두 번째로, 데이터가 Tidy Data 형태가 아니었다. 우리는 이 문제점들을 해결하기 위해 데이터들을 날짜에 따라 분류시킨 새로운 데이터를 만들기로 하였다. 일주일 단위로 데이터를 다운받는 코드를 만들고 데이터를 이어 붙여 3년간의(2014~2016) 데이터를 만들었다. 이 과정을 거치면 A와 C가 연결된 데이터가 없더라도 (A,B)쌍, (B,C)쌍 데이터가 있으면 이 둘을 합쳐 A와 C를 함께 분석할 수 있게 된다.

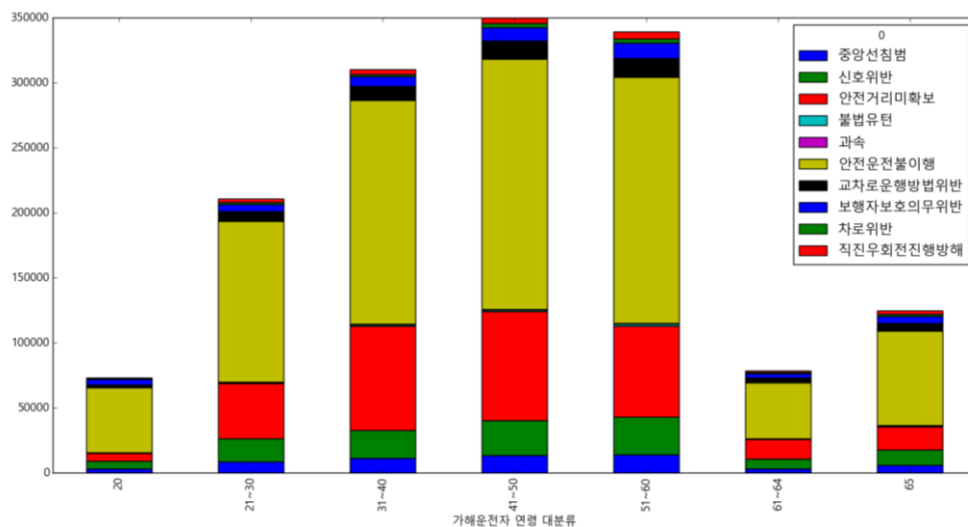
법규구분	기준구분	합계	일요일	월요일	화요일	수요일	목요일	금요일	토요일
합계	발생건수	18,561	1,788	2,874	2,570	2,761	2,771	3,043	2,754
	구성비(%)	100.0	9.6	15.5	13.8	14.9	14.9	16.4	14.8
	사망자수	102	7	16	12	19	18	10	20
	구성비(%)	100.0	6.9	15.7	11.8	18.6	17.6	9.8	19.6
	부상자수	29,085	3,226	4,236	3,821	4,230	4,182	4,518	4,872
	구성비(%)	100.0	11.09	14.56	13.14	14.54	14.38	15.53	16.75
	중상자수	2,846	306	403	376	437	396	445	483
	구성비(%)	100.0	10.8	14.2	13.2	15.4	13.9	15.6	17.0
	경상자수	15,581	1,691	2,253	2,043	2,233	2,278	2,413	2,670
	구성비(%)	100.0	10.9	14.5	13.1	14.3	14.6	15.5	17.1
	루상신고자수	10,658	1,229	1,580	1,402	1,560	1,508	1,660	1,719
	구성비(%)	100.0	11.5	14.8	13.2	14.6	14.1	15.6	16.1
중앙선침범	발생건수	372	49	61	52	50	43	60	57
	구성비(%)	100.0	13.2	16.4	14.0	13.4	11.6	16.1	15.3
	사망자수	8	2	0	1	2	1	0	2
	구성비(%)	100.0	25.0	0.0	12.5	25.0	12.5	0.0	25.0
	부상자수	573	82	90	86	72	65	92	86
	구성비(%)	100.0	14.31	15.71	15.01	12.57	11.34	16.06	15.01
	중상자수	128	17	14	19	11	17	20	30
	구성비(%)	100.0	13.3	10.9	14.8	8.6	13.3	15.6	23.4
	경상자수	362	57	51	55	49	40	62	48
	구성비(%)	100.0	15.7	14.1	15.2	13.5	11.0	17.1	13.3
	루상신고자수	83	8	25	12	12	8	10	8
	구성비(%)	100.0	9.6	30.1	14.5	14.5	9.6	12.0	9.6
신호위반	발생건수	758	82	116	91	120	133	102	114
	구성비(%)	100.0	10.8	15.3	12.0	15.8	17.5	13.5	15.0
	사망자수	6	0	0	0	1	2	2	1
	구성비(%)	100.0	0.0	0.0	0.0	16.7	33.3	33.3	16.7
	부상자수	1,270	164	176	159	198	224	141	208
	구성비(%)	100.0	12.91	13.86	12.52	15.59	17.64	11.10	16.38
	중상자수	257	33	44	29	54	41	21	35
	구성비(%)	100.0	12.8	17.1	11.3	21.0	16.0	8.2	13.6
	경상자수	801	106	95	105	120	152	92	131
	구성비(%)	100.0	13.2	11.9	13.1	15.0	19.0	11.5	16.4
	루상신고자수	212	25	37	25	24	31	28	42
	구성비(%)	100.0	11.8	17.5	11.8	11.3	14.6	13.2	19.8

위 데이터는 TAAS에서 받은 요일별-법규위반별 Raw Data이고 아래 데이터는 Tidy Data로 정리한 데이터이다. 모든 분야의 데이터를 다음과 같은 형태로 만들어 저장한 뒤 사용했다.

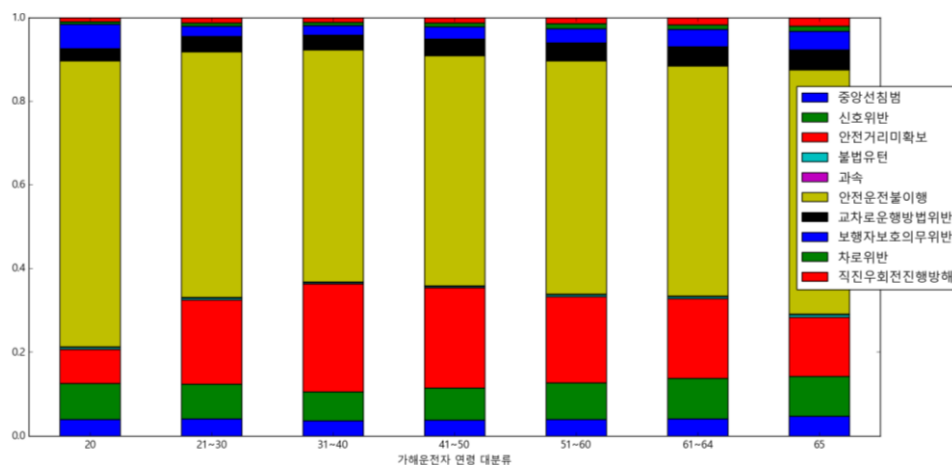
날짜	요일구분	중앙선침범	신호위반	전거리미확	불법유턴	과속	전운전불이	로운행방법	자보호의무	차로위반	우회전진행	기타	미분류
20160101	금요일	50	102	199	6	1	527	28	16	7	14	688	461
20160102	토요일	39	105	256	3	-	545	50	23	9	17	764	524
20160103	일요일	46	85	181	4	3	470	26	18	8	12	668	408
20160104	월요일	42	112	291	5	-	682	44	43	11	14	1040	621
20160105	화요일	60	103	271	11	2	615	53	46	14	17	962	614
20160106	수요일	45	111	294	5	2	619	45	42	14	15	857	589
20160107	목요일	43	116	300	6	2	678	55	56	9	13	964	635
20160108	금요일	44	127	344	3	2	717	57	47	17	20	1098	763
20160109	토요일	48	110	290	4	1	703	41	34	13	16	1019	676
20160110	일요일	42	78	193	4	-	509	40	17	4	13	600	380
20160111	월요일	39	103	316	4	-	682	66	36	17	24	1007	659
20160112	화요일	62	113	306	6	2	730	54	44	15	27	1077	681
20160113	수요일	78	113	575	9	1	952	46	41	6	10	1593	1064
20160114	목요일	80	123	375	4	1	781	69	47	9	18	1190	830
20160115	금요일	69	131	422	8	-	790	50	56	17	21	1252	875
20160116	토요일	72	164	329	8	-	713	55	33	14	21	1029	676
20160117	일요일	47	88	219	3	2	605	42	27	10	17	731	439
20160118	월요일	77	132	470	10	2	861	52	65	13	15	1466	993
20160119	화요일	53	115	368	3	1	777	53	48	19	14	1247	846
20160120	수요일	52	138	395	11	4	727	43	48	12	20	1110	744
20160121	목요일	53	130	349	3	3	722	41	45	10	19	1074	772
20160122	금요일	48	132	348	-	-	761	51	53	21	24	1134	724
20160123	토요일	54	96	337	8	1	749	41	37	10	14	1091	724
20160124	일요일	41	98	223	10	2	553	33	25	6	12	651	435
20160125	월요일	64	174	412	6	2	774	41	46	16	18	1231	886

Analysis 1: 두 필드 사이의 관계 분석

두 분야의 관계를 분석하기 위해 Stacked Bar Plot을 사용하였다. 가해운전자 연령 - 법규위반 데이터를 예로 들어 설명하면, 먼저 가해운전자연령대에 따른 사고 발생건수 Bar Plot을 그리고 그 안에서 어떤 법규위반에 의해 사고가 발생했는지를 Stacked bar로 나타낸 것이다.

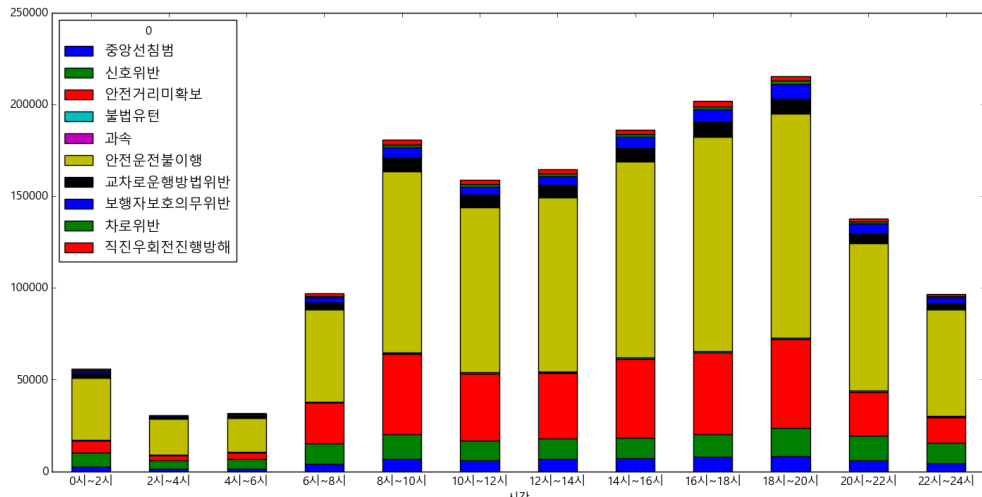


Bar Plot은 위에서부터 직진우회전진행방해, 차로위반, 보행자보호의무위반.... 순서로 구성되어있다. 위 그래프를 보면 30대~50대에서 가장 사고가 많이 발생했고 대부분 안전운전불이행과 안전거리미확보에 의한 사고였다는 것을 알 수 있다. 다만 나이에 따라 법규위반의 비율이 변화는 것을 눈으로 쉽게 확인할 수 없어, 각각의 법규위반의 퍼센트 값에 대해서 그래프를 새로 그려보았다.

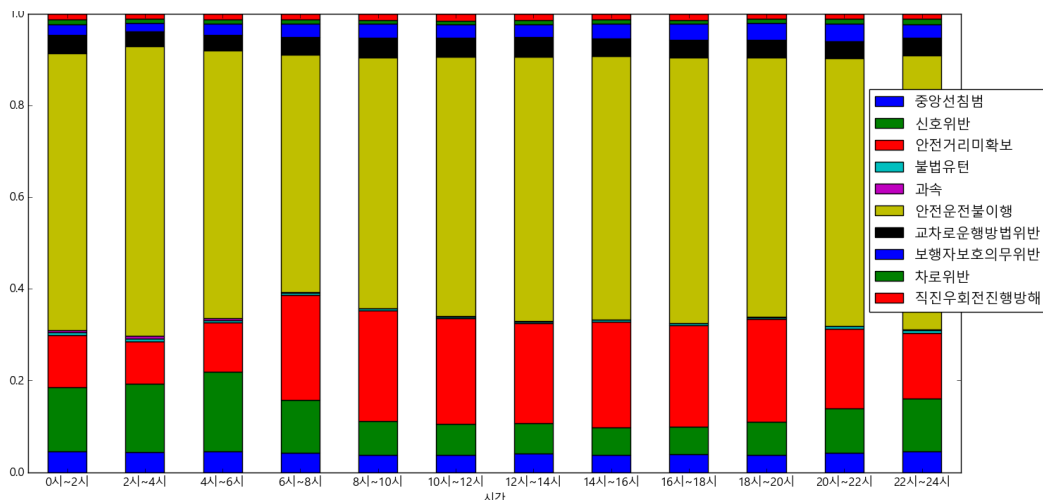


확인 결과 30대까지는 안전거리미확보 사고가 증가했다가, 그 이후부터는 안전운전

불이행 사고가 증가하는 경향을 확인할 수 있었다. 40대까지 보행자보호의무위반, 신호위반 사고가 감소했다가 그 이후로 다시 증가하는 모습도 확인할 수 있었다. 위와 같이 웹크롤링으로 수집한 모든 데이터 상에 대해서 막대그래프를 그려보았고, 두 feature 간의 관계를 분석할 수 있었다.



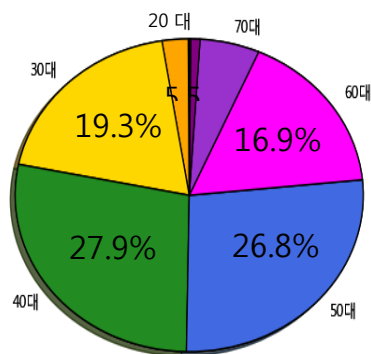
그래프 분석 결과, 출근이 시작되는 6시부터 10시까지 사고가 크게 증가하였다. 점심 시간에 부근에 사고가 잠깐 감소하였으나 퇴근시간까지 꾸준히 증가하여 18시~20 사이에 정점을 찍는 모습을 보였다. 법규위반은 안전거리미확보와 안전운전불이행이 제일 많았다.



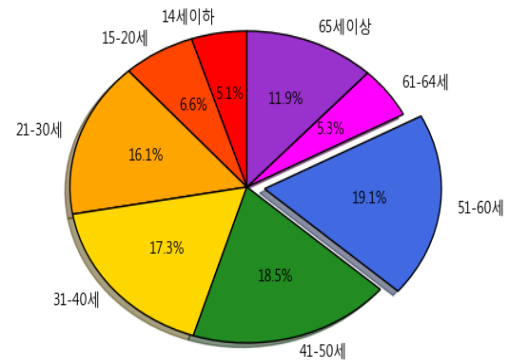
퍼센트 그래프로 분석해본 결과 밤사이에 신호위반사고와 과속사고가 증가하고 안전거리미확보 사고가 감소하는 것을 확인할 수 있었다. 밤에 절대적인 교통량이 감소하기 때문으로 보인다. 반대로 낮 시간대에는 신호위반 사고가 감소하고 안전거리미확보사고가 크게 증가하는 모습을 보였다.

Analysis 2: 파이차트 분석

우선, 연령별로 면허 소지자와 사상자를 비교하여 보았다. 서울특별시 기준으로 40대와 50대가 27.9%와 26.8%로 가장 면허 소지자의 비율이 높았다. 연령 별 사상자는 교통사고를 낸 쪽이든 당한 쪽이든 다친 사람의 비율을 나타낸 지표인데, 50대의 경우가 19.1%, 40대가 18.5%로 가장 높은 비율을 차지하였다. 20대의 경우가 면허 소지자는 적은 반면 사상자 비율은 다른 연령층과 비슷함을 확인할 수 있다. 이는 교통사고를 내서 다친 것 보다는 타인의 사고에 의해 다친 경우가 더 많거나, 운전 경력이 적어 사고를 많이 냈던 것으로 추측할 수 있다.

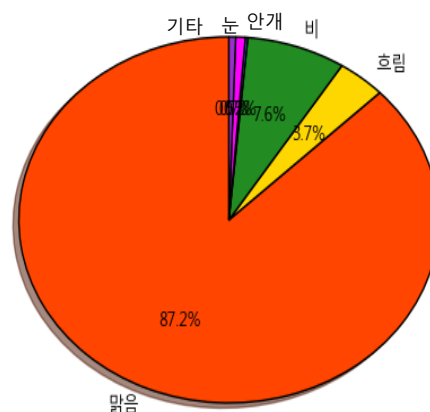


연령 - 면허 소지자 비율(서울)



사상자 - 연령 별 비율

기상에 대해서도 파이차트 분석을 통해 사고 비율을 나타내어 보았는데, 전체 사고 수의 87.2%를 맑음이 차지하고 있었다. 이는 원래 맑은 날씨가 많기 때문에 나온 결과이다. 이를 바탕으로 하여 이후 기상과 연관된 분석을 할 때 맑음이 대부분을 차지하고 있다는 점을 염두하고 분석을 진행하였다.

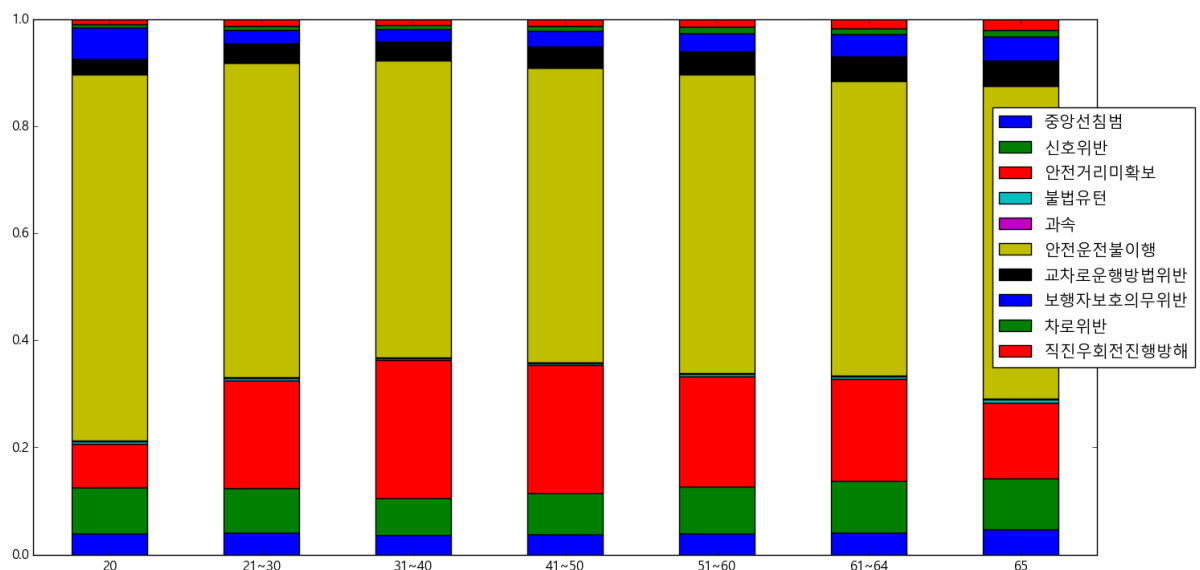


기상 - 사고비율

Analysis 3: 산점도 및 상관계수 분석

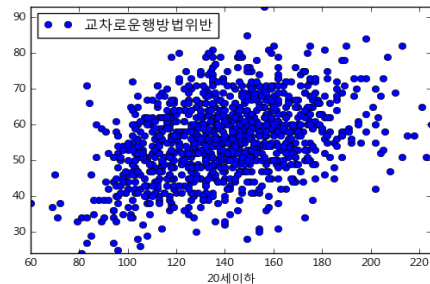
처음 분석을 계획하였을 때, 산점도와 상관계수 분석을 통해 두 필드 간 연관성이 있는지 분석하는 것을 목표로 하였었다. 산점도를 찍을 때는 X축과 Y축에 서로 다른 두 가지 분야에서 한 가지의 column씩을 가져와서, 어느 날 X에서의 사고 수와 Y에서의 사고수를 한 점으로 나타내는 식으로 하여 플롯하였다. 각 column에 대해 짝지어 종류별로 scatter plot 한 후 상관계수 분석을 진행해 보고, 그 중 계수가 큰 값들에 대해 자세히 살펴보았다.

연령 - 법규위반의 경우 모든 연령대에 걸쳐 안전운전 불이행과 안전거리 미확보의 상관계수가 가장 높게 나왔다. 플롯된 그래프를 보았을 때 상당히 깔끔한 직선 모양인데, 이를 자세히 보면 위의 두 가지 법규위반은 굉장히 범용적이기 때문인 것으로 확인할 수 있다.

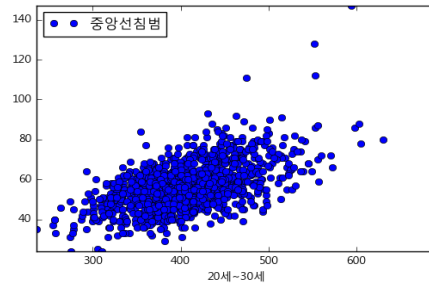


모든 연령에 대해서 대부분의 법규위반은 안전운전불이행과 안전거리 미확보였기 때문에, 이 두 가지에 대해서만 상관계수 분석을 하는 것은 좋은 결과를 얻지 못할 것이라고 판단하였다. 따라서 연령 - 법규위반의 상관계수 분석에서 두 가지 법규위반 요소는 제외하고 가장 계수가 높은 것을 가져왔다. 그 결과, 20세 이하와 60대 이상은 교차로운행위반과 의 상관계수가 가장 높았고, 30대~50대에는 보행자보호위반의 상관계수가 가장 높았던 것으로 확인되었다. 20대의 경우는 중앙선침범과의 상관계수가 가장 높았다. 연령 - 법규위반 그래프를 보았을 때 확인할 수 있는 것은, 우리

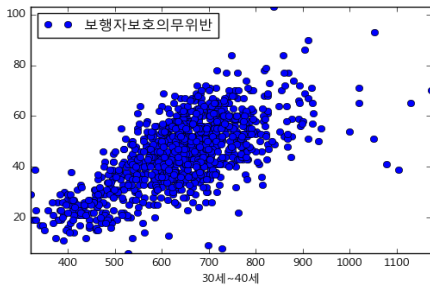
분석에서 그 분야의 사고수가 많으면 전체적으로 상관계수 또한 올라가는 양상을 보였다는 것이다. 가해자비율이 높은 30~50대에서 모든 법규위반에 대해 대체적으로 다른 연령대보다 높은 상관계수를 보임을 알 수 있었다. 그렇기 때문에 연령-법규위반 전체에서 상관계수가 가장 높은 몇 가지만을 분석하지 않고 연령대별로 높은 값을 가지는 법규위반에 대해 그래프 분석을 하게 되었다.



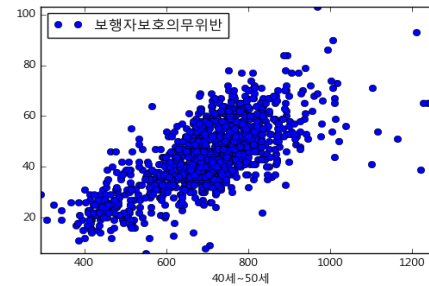
20 세이하 - 교차로운행방위반 : 0.383



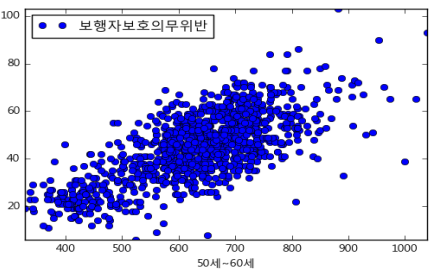
20 세~30 세 - 중앙선침범 : 0.564



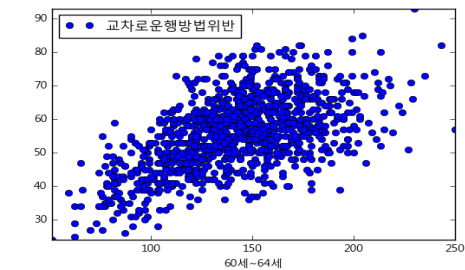
30 세~40 세 - 보행자보호의무위반 : 0.678



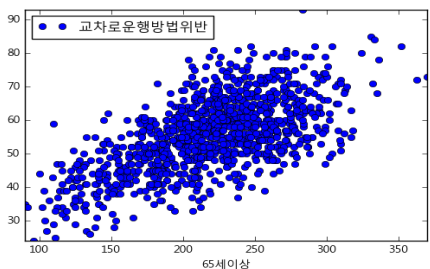
40 세~50 세 - 보행자보호의무위반 : 0.719



50 세~60 세 - 보행자보호의무위반 : 0.698



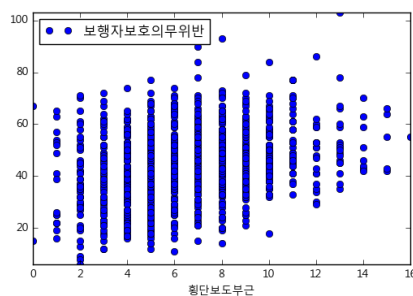
60 세~64 세 - 교차로운행방위반: 0.557



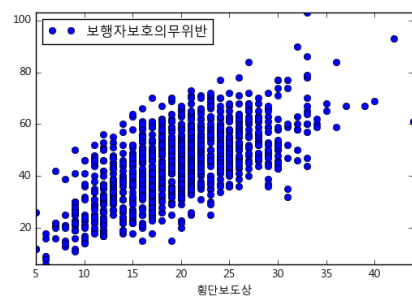
65 세이상 - 교차로운행방위반 : 0.626

다음으로는, 관련이 있어 보이는 분야인 도로형태 - 법규위반에 대한 산점도 분석을 하였다. 연령 -법규위반과 마찬가지로, 사고수의 비율이 높은 교차로안 과 교차로 부

근 같은 곳은 안전운전불이행과의 상관계수가 가장 높았다. 주목할만한 점은 횡단보도상과 횡단보도부근에 있다. 이 두 장소에 대해서도 안전운전불이행의 사고비율이 훨씬 높지만, 상관계수 분석에서는 보행자보호의무위반과의 계수가 각각 0.656, 0.290으로 더 높았던 것을 확인할 수 있다. 실제로 횡단보도 쪽에서 나는 사고의 다수는 보행자보호의무 위반과 밀접한 연관이 있고, 이것이 분석에서 잘 나타난 것으로 보아 이러한 분석 방식이 어느 정도 의미를 가짐을 시사한다. 생각보다 특이했던 부분은 횡단보도상에서의 사고 수와 상관계수가 모두 횡단보도 부근에서보다 훨씬 많았다는 것인데, 이를 통해 횡단보도에서 안전 확보가 필요함을 확인할 수 있다.

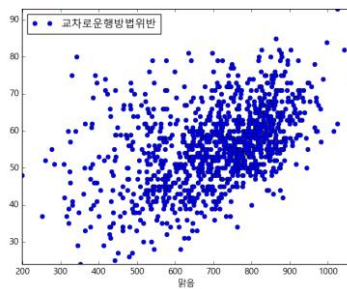


횡단보도부근 -
보행자보호의무위반 : 0.290

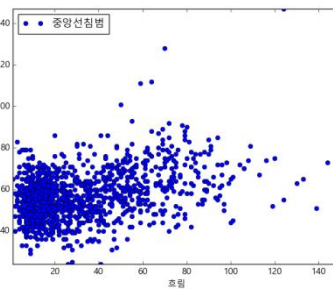


횡단보도상 -
보행자보호의무위반 : 0.656

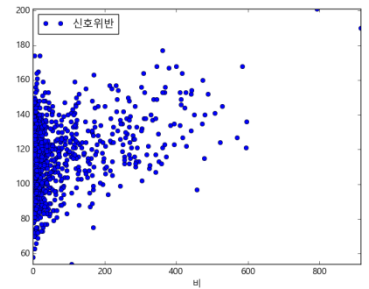
기상 - 법규위반에 대해서도 산점도를 그려보았다. 아래 그래프는 기상 별로 상관계수가 가장 높았던 법규위반 요소를 가져온 것이다.



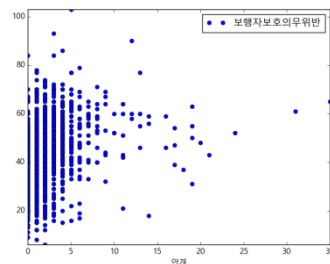
맑음 - 교차로운행방법위반:
0.453



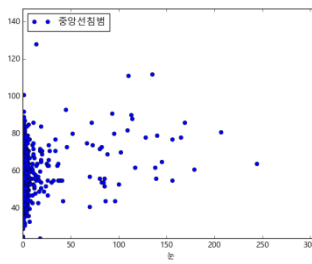
흐림 - 중앙선침범: 0.366



비 - 신호위반: 0.367



안개 - 보행자보호의무위반:
0.231

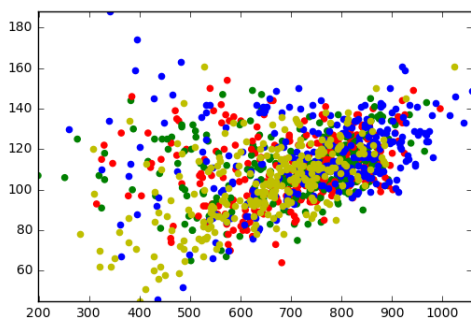


눈 - 중앙선침범: 0.296

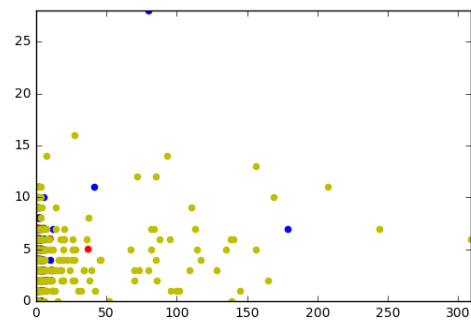
Analysis 3-1: 계절 별 산점도 분석

두 필드의 사고수를 각각 변수로 갖는 산점도를 찍어 보았는데, 조금 더 생각을 해 보니 눈 등의 날씨 필드에 대해서는 눈이 오지 않는 날이 많은 등의 문제가 있었다. 그래서 눈, 비(특히 눈) 자주 오는 계절이 따로 있는 우리나라의 특성을 고려하여 계절 별로 다른 색의 점을 찍어보기로 하였다.

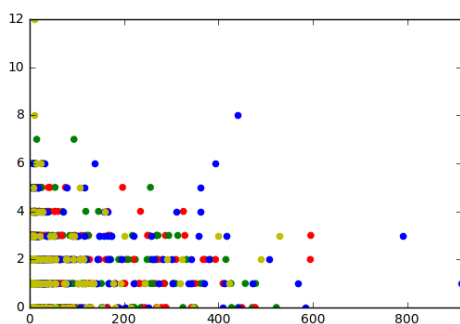
3년치 데이터이기 때문에 계절을 나누는 조건문은 '2014년 봄 기간||2015년 봄||2016년 봄'과 같은 식으로 구성하였다. 그리고 나뉜 각 데이터 프레임을 한 figure에 다른 색으로 그래프를 그렸다. 그래프는 다음과 같았다(x축, y축), (봄, 여름, 가을, 겨울 순서대로 빨강, 녹색, 파랑, 노랑).



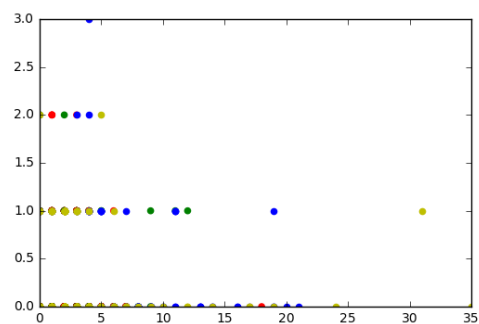
(맑음, 교차로부근)



(눈, 교량)



(비, 터널)



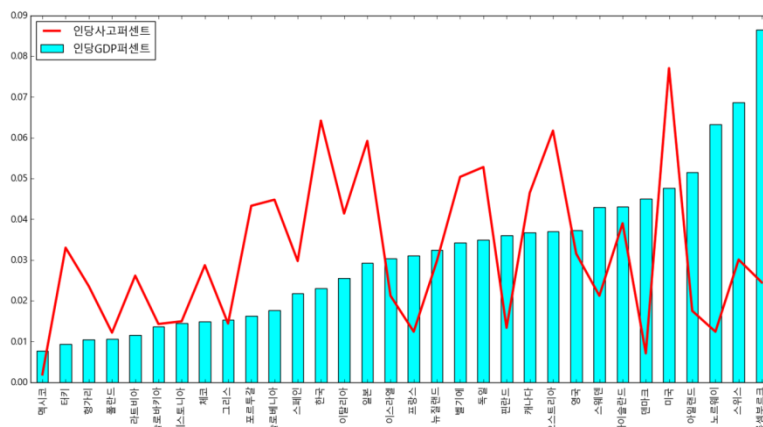
(안개, 철길건널목)

눈의 경우는 일어난 사고가 압도적으로 겨울의 색깔을 하고 있어 성공적인 분석 방법이었지만, 다른 날씨의 경우 계절에 따라 큰 차이를 보이지 못하여 아쉬웠다. 한 가지 공통점이라면 겨울은 대체적으로 원점에 가까운, 즉 적은 사고수를 보였다.

Extra analysis 1: OECD 국가의 GDP와 행복지수가 사고에 미치는 영향 분석

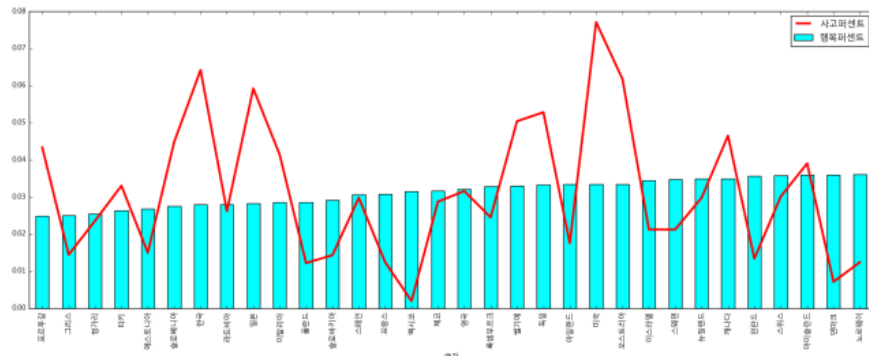
지금까지의 분석은 국내 일어났던 사고의 통계에 대해서만 다루었는데, 이 파트에서는 통계 포털사이트에서 OECD 국가에 대한 총 GDP와 UN에서 제공하는 세계 행복지수, 그리고 인구수 데이터를 가져와 GDP나 행복지수가 사고에 영향을 미치는가에 대해 분석하였다.

GDP는 총 인구수에 대한 GDP를 제공하고, 사고 또한 발생한 총 사고 수 데이터를 가져온 것이기 때문에 인구수가 많다면 그에 비례하여 두 값이 모두 높을 것이다. 그렇기 때문에 인구수 데이터를 추가로 가져와서 각각을 인구수로 나누어 1인당 GDP와 1인당 평균 사고수를 구하였고, 이를 그래프로 나타내 보았다.



붉은 꺾은선 그래프가 인당 사고의 비율을 나타낸 것이고, 하늘색 막대그래프가 국가별 인당 GDP이다. GDP에 대해서 오름차순으로 정렬하고 플롯해 보았을 때, 사고수와 인당 GDP에서 큰 연관성은 찾지 못하였다.

행복지수와 사고에 대해서도 같은 방식으로 분석을 해 보았는데, 행복지수의 경우는 인구수에 비례하여 증가하는 값이 아니므로 그대로 사용하였다.

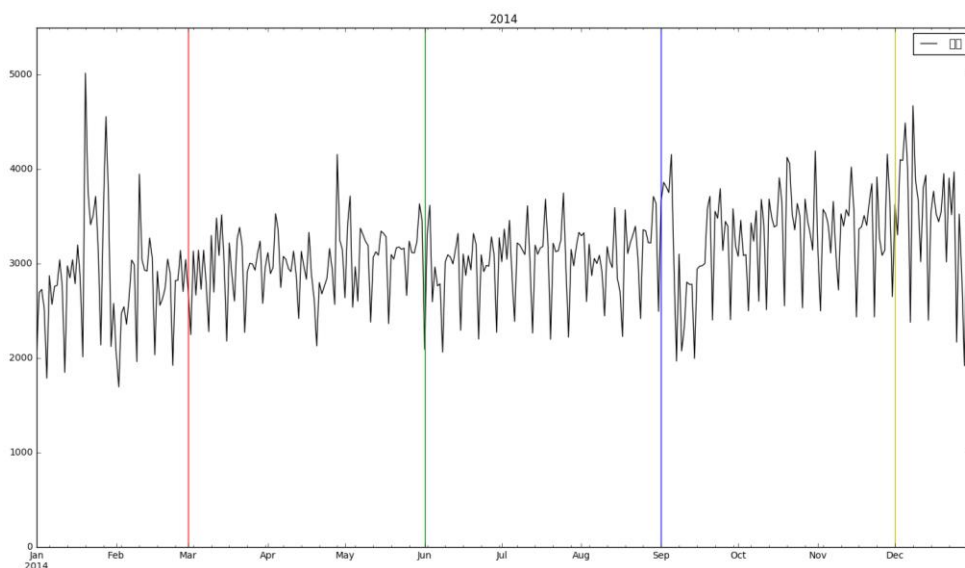


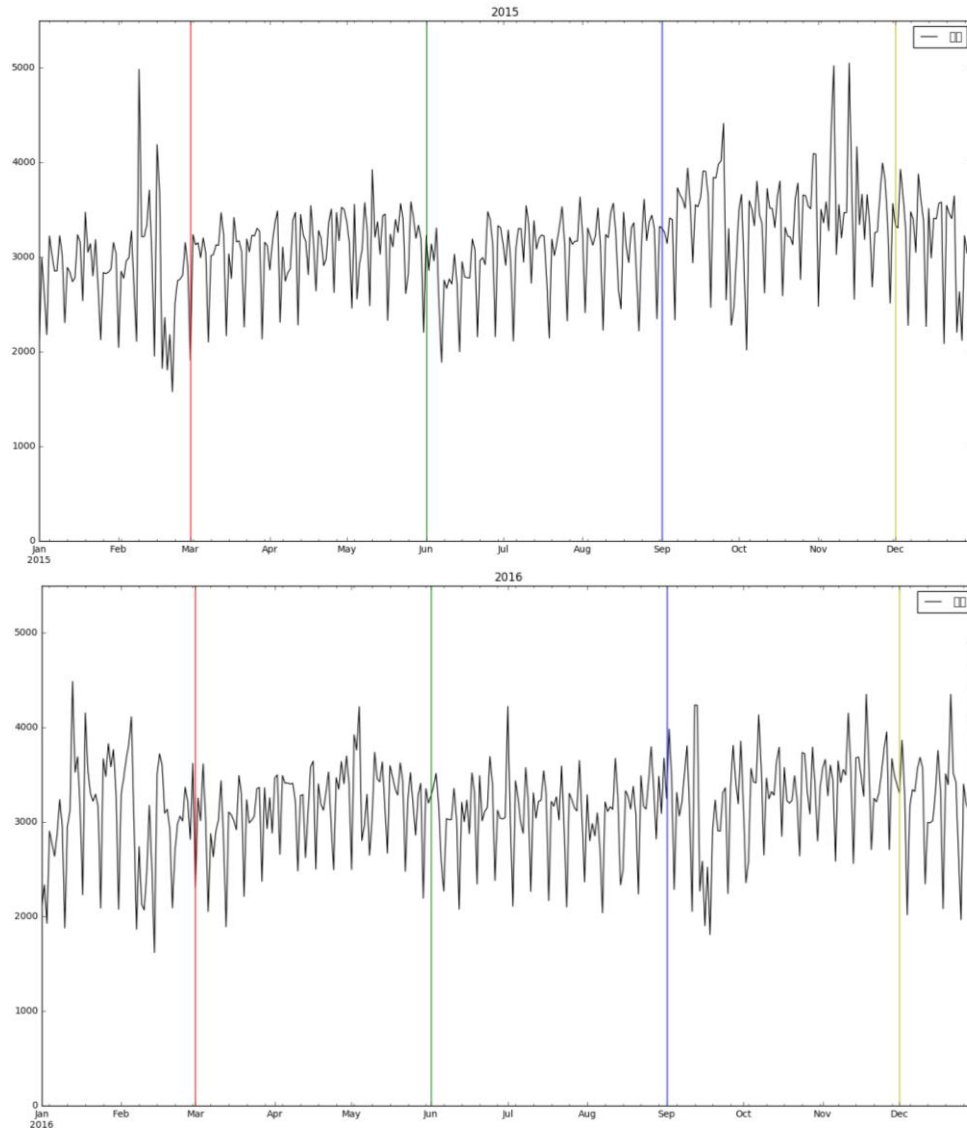
위 그래프는 행복지수를 오름차순으로 정렬한 후 플롯한 그래프이다. 마찬가지로 붉은색 꺾은선 그래프는 사고를, 하늘색 막대그래프는 행복지수를 나타낸다. 이 분석의 결과 행복지수 또한 사고 발생 비율과는 큰 연관성을 보이지 않았음을 확인할 수 있었다.

그러나 이 두 가지 분석은 OECD 국가에 대해서만 진행한 것이기 때문에, 이에 포함되지 않는 전세계 국가들을 대상으로 같은 분석을 해 본다면 보다 의미 있는 결과를 가질 수도 있을 것이다.

Extra analysis 2: 1년 전체 경향성 분석

처음의 목적대로 1년 단위의 사고 경향성을 알아보기 위해 요일 별 시간 별 사고 수 데이터의 하루 사고 수 총계 데이터를 이용하여 1년 단위 그래프를 3개년을 그렸다.





선은 각각 계절을 구분하기 위한 선이다. 생각보다 너무 사고 수가 들쭉날쭉 해서 전체적인 경향성을 파악하기는 힘들었다. 다만 대략 2천~4천 건이 꾸준히 일어나는 것을 알 수 있었고, 높이 솟은 부분들은 대부분 명절과 그 근처임을 확인하였다. 세 계절 모두 겨울의 평균 사고 수는 줄어들지만 일년 중 가장 사고가 많은 날도 겨울에 있음을 알 수 있었다.

Conclusion & Discussion

분석을 통해 몇 가지 사실을 확인할 수 있었다. 첫 번째로 과속, 신호 위반에 의한 사고의 경우 낮에 비해 밤에 증가하는 것을 알 수 있었다. 이는 낮보다 밤의 교통량

이 현저하게 낮아 과속이 쉽고 신호 위반이 잘 단속이 되지 않기 때문이라고 생각하였다. 그래서 우리는 이 사고들에 대해 수를 줄이기 위해선 단속을 조금 더 강화하는 방안을 제안하고 싶다.

두 번째로 연령에 따라 사고 시 위반하는 법규가 조금씩 다르다는 것을 알 수 있었다. 크게 젊은 층과 장년층이 다른 점을 보여주었다. 이에 대한 대책으로는 특정 연령층이 면허 갱신을 위해서 조금 더 잘 위반하는 법규에 대해서는 교육을 강화하는 방법이 있겠다.

세 번째로, 횡단보도 근처보다 횡단보도상에서 일어나는 사고가 많았다는 것이다. 이는 횡단보도 근처에 신호체계가 복잡한 것과 동시에 사람들이 급한 일 등의 상황에서 횡단보도에 대한 안전의식이 조금 부족하기 때문이라고 생각했다 또한 횡단보도가 그렇게 안전하지 않다는 점도 원인이 되는 것 같아 이를 개선하기 위해 횡단보도의 구조를 개선하고, 횡단보도 안전 의식을 강화하기 위한 캠페인이나 운전자 교육을 이수하는 것이 좋을 것이다.