

교통사고 유형별 위험구간 예측 및 분석

SA133(김민영, 박연아, 옥수민)



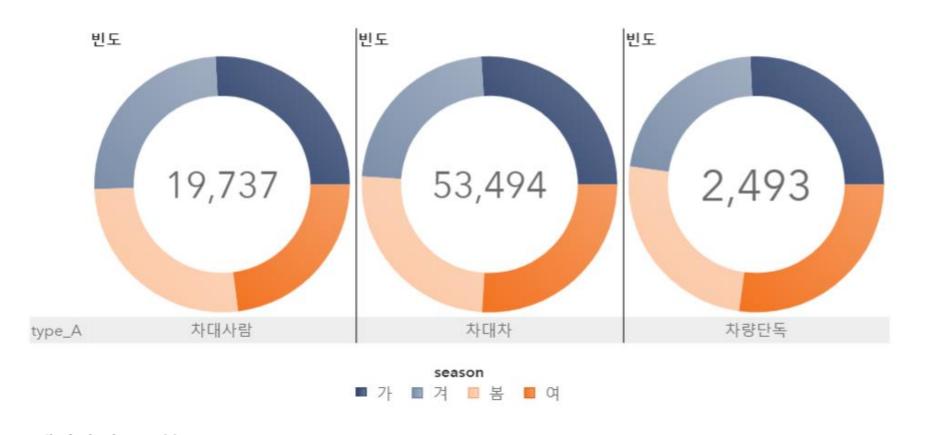






데이터 탐색

- 1. 시간별교통사고탐색
- 2. 장소별교통사고탐색
- 3. 피해자/가해자별교통사고탐색



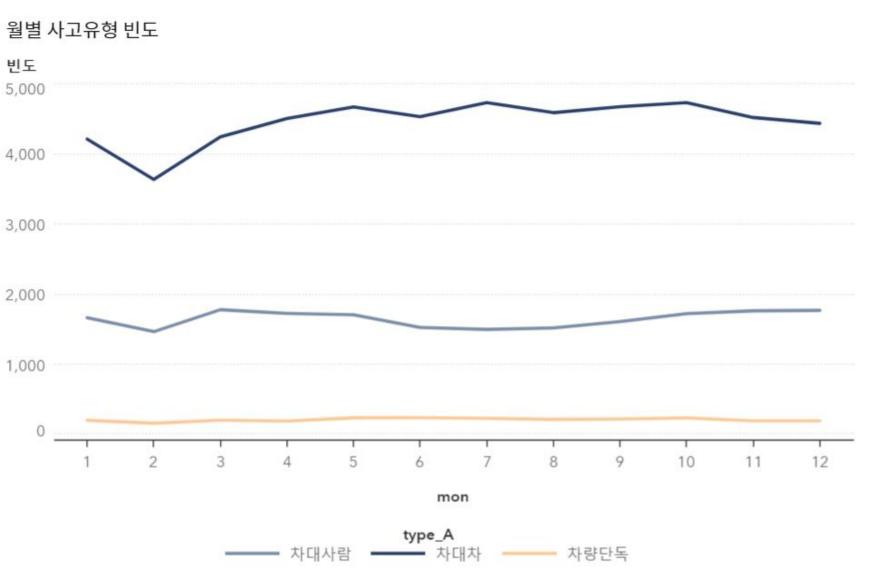
시간별 교통사고

발생 계절별 사고유형

계절별 사고유형

type_A ▲	차대사람	차대차	차량단독	
season ▼	빈도	빈도	빈도	
여	4,539	13,857	677	
봄	5,209	13,421	626	
겨	4,896	12,286	548	
가	5,093	13,930	642	

- 전반적으로 차대차의 비율이 가장 크고 차대차 사고는 여름에 가장 큰 빈도를 갖음
- 차대사람 사고는 봄에 가장 큰 빈도를 갖음

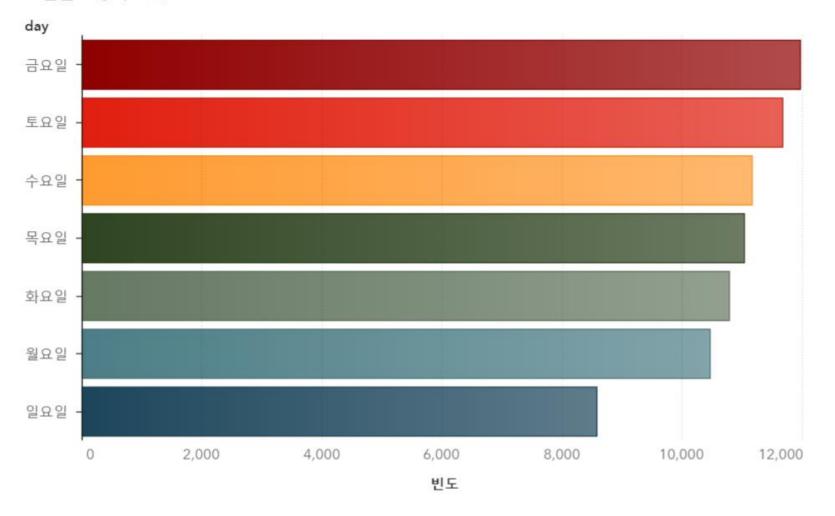


시간별 교통사고

발생 월별 사고유형

• 2월에 사고의 빈도가 줄어들고 그 외에는 전반적으로 비슷한 빈 도를 보임

요일별 교통사고 빈도

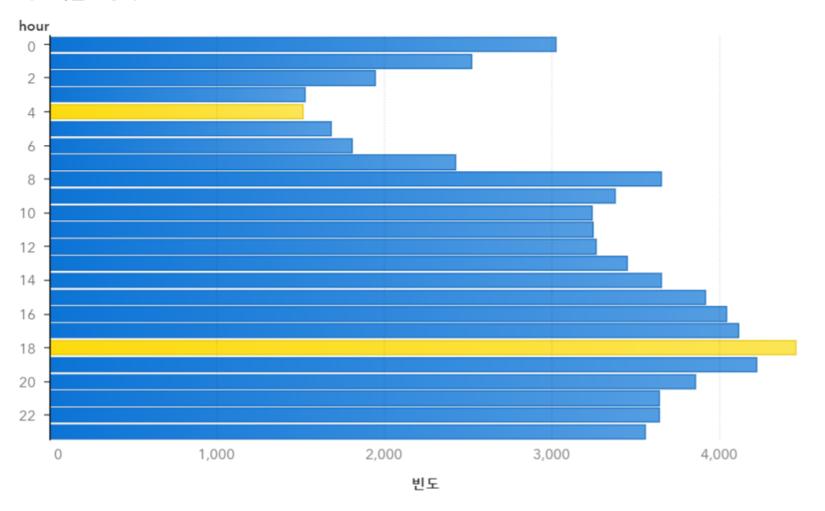


시간별 교통사고

발생 요일별 사고유형

• 요일별로는 금,토,수 순으로 높은 빈도를 보임

시간대별 교통사고 빈도



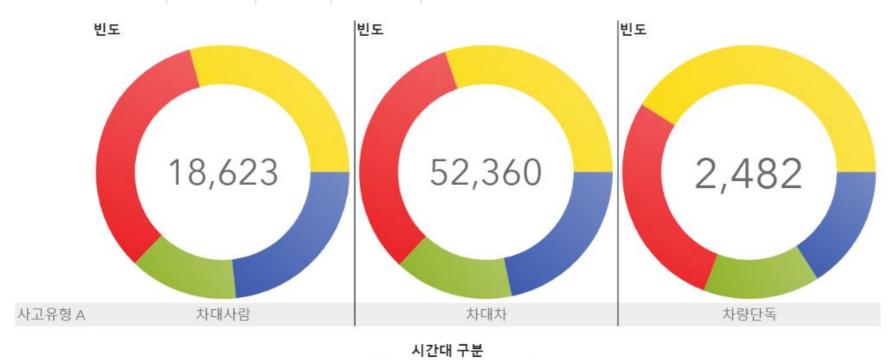
시간별 교통사고

발생 시간별 사고유형

- 새벽 4시에 가장 낮은 사고빈도를 보임
- 저녁 6시에 가장 높은 사고빈도를 보임

시간대별 사고유형

사고유형 A ▲	차대사람	차대차	차량단독	
시간대 구분 ▲	빈도	빈도	빈도	
0‡	5,439	15,821	1,019	
주	6,259	17,199	699	
출	2,583	7,970	369	
퇴	4,342	11,370	395	



시간별 교통사고

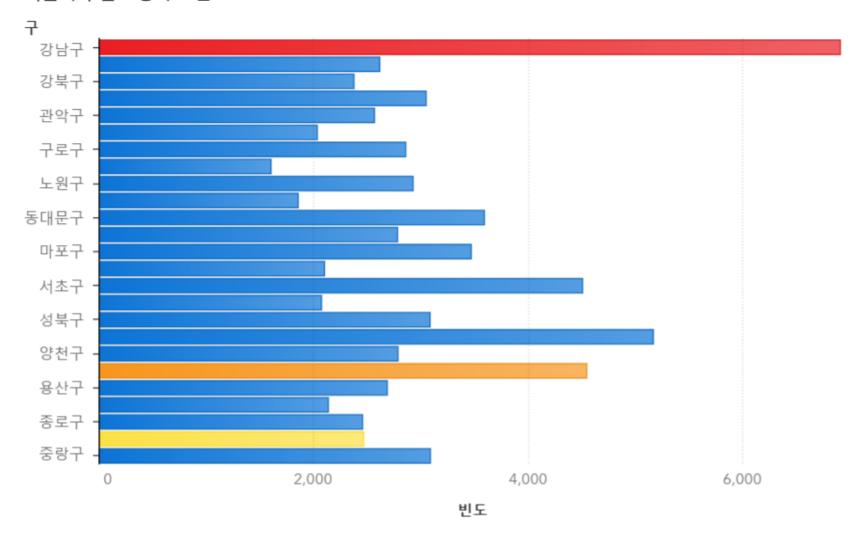
발생 시간대별 사고유형

장소별 교통사고

발생 구별 사고유형

- 강남구가 가장 높은 사고 빈도를 보임
- 금천구가 가장 낮은 사고 빈도를 보임

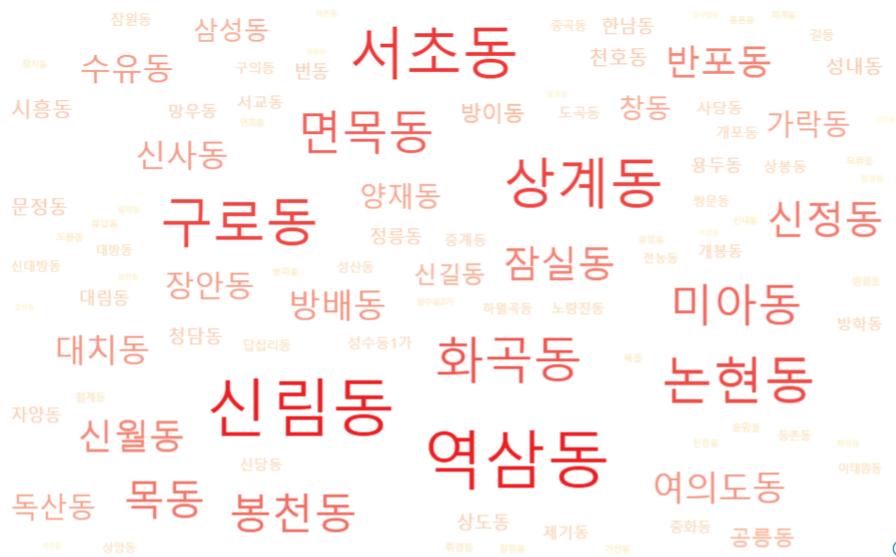
서울시 구별 교통사고 빈도



장소별 교통사고

발생 동별 사고유형

서울시 동별 교통사고 빈도



(i)

피해자/가해자별 법규위반 빈도 교통사고

법규 위반

신호위반 과속

교차로운행방법위반

직진우회전진행방해

보행자보호의무위반

기타

안전운전불이행

불법유턴

중앙선침범

차로위반

안전거리미확보

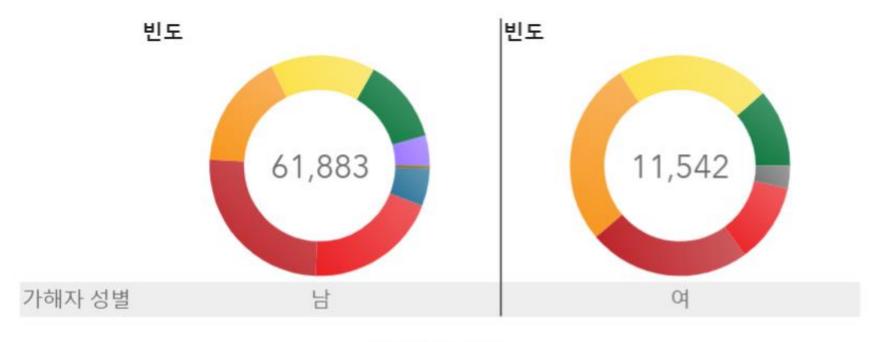
피해자/가해자별 교통사고

성별에 따른 가해자 연령대

- 남성 가해 운전자는
 50대,60대,40대 순으로 높은 빈도를 보임
- 여성 가해 운전자는 40대,50대,30대 순으 로 높은 빈도를 보임

성별에 따른 가해자 연령대

가해자나이 ▲	10대이하	20대	30대	40대	50대	60대	70대	80대이상
att_gender ▲	빈도	빈도	빈도	빈도	빈도	빈도	빈도	빈도
남	2,751	7,666	9,516	10,477	15,644	12,147	3,393	289
여	102	1,319	2,626	3,120	2,750	1,345	262	18





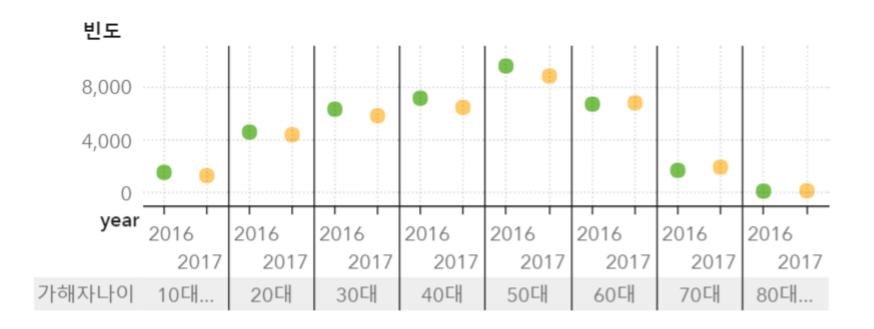
피해자/가해자별 교통사고

연도별 가해자 연령대

2016년, 2017년 모두 50대 가해운전자의 빈도가 가장 높음

연도별 가해자 연령대 빈도 차이

가해자나이 ▲	10대이하	20대	30대	40대	50대	60대	70대	80대이상
year 🔺	빈도							
2016	1,555	4,594	6,318	7,149	9,579	6,699	1,708	140
2017	1,308	4,394	5,825	6,459	8,824	6,799	1,947	167

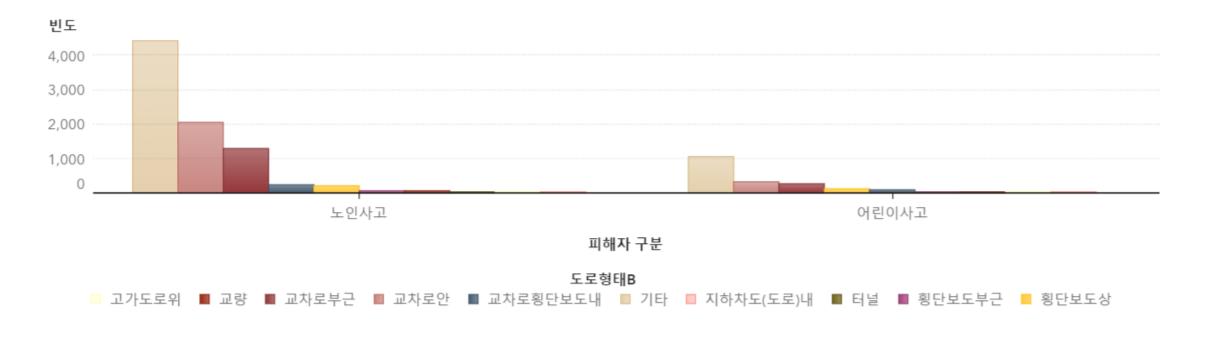


피해자/가해자별 교통사고

피해자별 도로형태

피해자별 교통사고 도로형태

도로형태B ▲	고가도로위	교량	교차로부근	교차로안	교차로횡단보도내	기타	지하차도(도로)내	터널	횡단보도부근	횡단보도상
피해자 구분 ▲	빈도	빈도	빈도	빈도	빈도	빈도	빈도	빈도	빈도	빈도
노인사고	13	66	1,302	2,058	261	4,405	13	19	72	222
어린이사고	1	5	282	337	116	1,055	1		24	135



• 노인과 어린이 피해사고 모두 기타, 교차로안, 교차로부근 순으로 높은 빈도를 보임

데이터 분석

- 1. 데이터 정제
- 2. 데이터모델링

01. 데이터 정제

1) Accident(교통사고 데이터)

결측치/관측치 변환 및 제거

- (1) year, mon, date의 결측치를 time변수를 이용하여 대체
- (2) att_age(가해운전자나이), vic_age(피해운전자나이)에서 기타불명/미분류 값을 99/999로 처리하여, 문자형식변수를 숫자형식으로 바꿔 age1(가해운전자) 생성
- (3) road_typeA(도로형태)-주차장, 미분류, 기타로 된 2863개 관측치 제거 → 도로형태별 사고 원인을 파악하기 위해 제거
- (4) law(법규위반) 79개 관측치 제거 → 교통사고의 주된 원인을 나타내는 법규위반을 알 수 없어서 제거

새로운 변수 생성

- (1) att_car(가해운전자 차종)의 수준을 승용차/승용차 이외로 구분하여 car변수 생성 → 승용차가 압도적으로 관측치가 많아서 다른 수준을 합쳐 비교하기위해
- (2) day(요일)변수를 평일과 주말로 구분하여 dday 변수 생성 →평일 주말별로 교통의 상황이 유의한 차이 확인
- (3) hour(발생시간)을 출근(6-9시),주간(10-16시),퇴근(17-20시),야간(21-5시)으로 hhour(시간대)범주화 →특정시간별로 교통사고의 특징이 예상되어 생성
- (4) mon(발생월)을 봄(3-5월)/여름(6-8월)/가을(9-11월)/겨울(12-2월)으로 season(계절)변수로 범주화 → 계절에 따른 교통사고 특수성을 확인하기위해 생성

01. 데이터 정제

2) DTG

새로운 변수 생성(가속도 변수를 방향에 따라 정확히 보기위해 나눔)

- (1) VX(X축 가속도)가 + 값인 관측치를 VX_f(차량의 전방가속도)를 생성 그 이외 관측치는 VX_f=0으로 둠
- (2) VX(X축 가속도)가 값인 관측치를 VX_b(차량의 후방가속도)를 생성 그 이외 관측치는 VX_b=0으로 둠
- (3) VY(Y축 가속도)가 + 값인 관측치를 VY_r(차량의 우측가속도)를 생성 그 이외 관측치는 VY_r=0으로 둠
- (4) VY(Y축 가속도)가 값인 관측치를 VY_I(차량의 좌측가속도)를 생성 그 이외 관측치는 VY_I=0으로 둠
- (5) VX, VY가 (+,+) 값인 관측치값을 fr(차량의 전방우측가속도)를 생성 그 이외 fr=0으로 둠
- (6) VX, VY가 (+,-) 값인 관측치값을 fl(차량의 전방좌측가속도)를 생성 그 이외 fl=0으로 둠
- (7) VX, VY가 (-,+) 값인 관측치값을 br(차량의 후방좌측가속도)를 생성 그 이외 br=0으로 둠
- (8) VX, VY가 (-,-) 값인 관측치값을 bl(차량의 후방우측가속도)를 생성 그 이외 bl=0으로 둠
- (9) VX와 VY의 절대값을 취한 값의 합의 변수 Y 생성

데이터 분할

RPM 6000이상을 기록한 운전자중 가장 높은 빈도를 가지는 운전자와 일반운전자를 선정하여 데이터 분할

1) 사고유형별 로지스틱 회귀

목적: 교통사고데이터에서 교통사고 유형을 결정하는 요인을 찾기 위해 실시

Y(반응변수): type_A(사고유형)

- 차대 차/ 그 회 (사고 유형 중 차대 차 사고 /차대 사람 사고와 차량 단독 사고를 그 외로 설정)
- 차대 차와 다른 사고의 비율차이가 커 차대 사람, 차량단독을 묶어 이진 로지스틱 회귀 실시

X's(설명변수):

- 고정요인 : road_type_A(도로형태)

- 변동요인 : contents(사고내용), att_car(가해운전자 차량), dday(평일/주말), hhour(시간대),

season(계절), att_age2(가해운전자 연령대)

1) 사고유형별 로지스틱 회귀

단계별 선택(stepwise)후 선별된 X's(설명변수):

- 고정요인 : road_type_A(도로형태)

- 변동요인 : contents(사고내용),

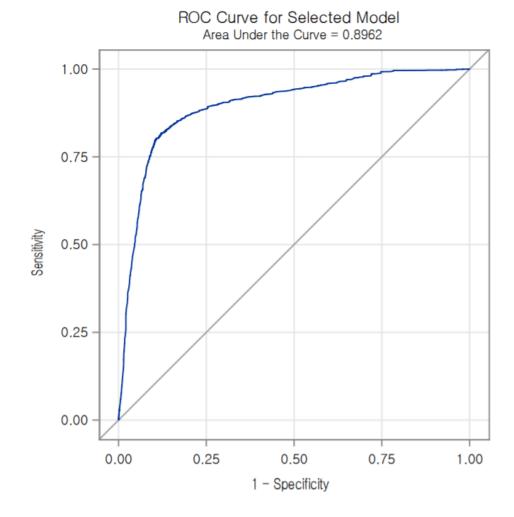
att_car(가해운전자 차량)

dday(평일/주말)

hhour(시간대)

season(계절)

att_age2(가해운전자 연령대)



Analysis of Maximum Likelihood Estimates											
Parameter	_	DF	Estimate	Standard Error	Wald	Dr. ChiCa					
		1	0.5520	0.0355	Chi-Square 241.2010	Pr > ChiSq <.0001					
Intercept	경상사고	1	0.3320	0.0333	1943.9781	<.0001					
	부상신고사고	1	0.7992	0.0181	340.2382	<.0001					
contents											
Contents	사망사고	1	-1.0346	0.0855	146.3888	<.0001					
Contents	중상사고	0	0.6106	0.0476	4242.0502						
road_type_A	교차로	1	0.6196	0.0176	1243.0602	<.0001					
road_type_A	단일로	0	0								
Att_Car	승용차 이외	1	-0.3565	0.0185	372.0333	<.0001					
Att_Car	승용차	0	0								
Dday	주말	1	0.0739	0.0193	14.7004	0.0001					
Dday	평일	0	0								
Hhour	야간	1	-0.0148	0.0236	0.3904	0.5321					
Hhour	주간	1	0.0485	0.0231	4.3996	0.0359					
Hhour	출근시간	1	0.1340	0.0285	22.0990	<.0001					
Hhour	퇴근시간	0	0								
Season	가을	1	-0.0847	0.0238	12.6804	0.0004					
Season	겨울	1	-0.1771	0.0242	53.3618	<.0001					
Season	봄	1	-0.1462	0.0237	37.9097	<.0001					
Season	여름	0	0								
Att_Age2	노년	1	-0.1663	0.0433	14.7544	0.0001					
Att_age2	장년	1	-0.3023	0.0256	139.6160	<.0001					
Att_Age2	중년	1	-0.2332	0.0263	78.6573	<.0001					
Att_age2	초년	0	0								

contents 경상사고

road_type_A 교차로

Att_car 승용차

Dday 주말

Hhour 출근시간

Season 여름

Att_age2 초년

차대차 >

차량단독 차대사람



1) 사고유형별 로지스틱 회귀

X's(설명변수) 수준해석:

<season(계절)>

여름일 때 차대차 사고의 비율이 가장 높고 겨울일 때 가장 낮다.

겨울에 눈이 오거나 땅이 얼었을 경우에 오히려 운전자들이 조심스러운 운전태도를 보여 안전운전불이행으로 일어 나는 차대차 사고보다는 미끄러운 노면상태에 의한 차량단독 사고나 차대사람사고가 날 확률이 클 것으로 추측

<hhour(시간대)>

출근시간일 때가 퇴근시간일 때보다 차대차 사고가 날 확률이 그 외의 사고가 날 확률보다 크다.

출근시간에는 회사밀집구역에 차가 몰려서 접촉사고나 충돌사고등 차대차 사고가 날 확률이 높다고 추측할 수 있고, 이에 반해 주거지역으로 흩어지는 퇴근시간에는 비교적 차대차 사고가 날 확률이 크지않을 것으로 예측

<att_age2(가해운전자 연령대)>

초년일 때 차대차 사고를 낼 확률이 가장 높고 장년일 때 가장 낮다.

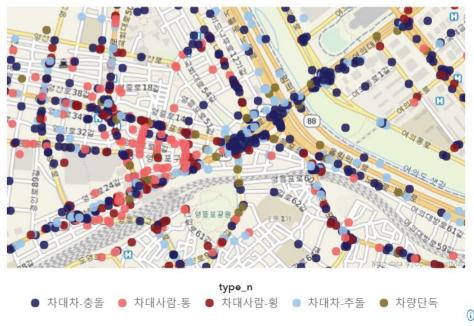
중장년은 기본적으로 빈도가 많기 때문에 차대차 사고의 비율이 그 외에 비해 크게 높지는 않지만 교통상황에 미숙한 초년층과 신체적 요인이 영향을 줄 수 있는 노년층이 차대차 사고를 낼 확률을 크게 높이는 것으로 보임. 특히 초년층은 차량단독이나 차대사람사고보다는 가벼운 차대차 사고를 많이 일으키는 것으로 추측

강남구_사고유형별 위치 시각화

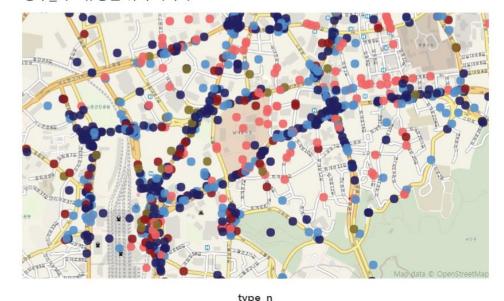


- 차대차 사고는 주로 대로변에서 많이 일어남
 →특히 충돌사고는 교차로 지점에 집중
- 차대사람사고는 길 교차지점이나 대로변 교차로에서 많이 일어남

영등포구 사고유형별 위치 시각화



중구_사고유형별 위치 시각화



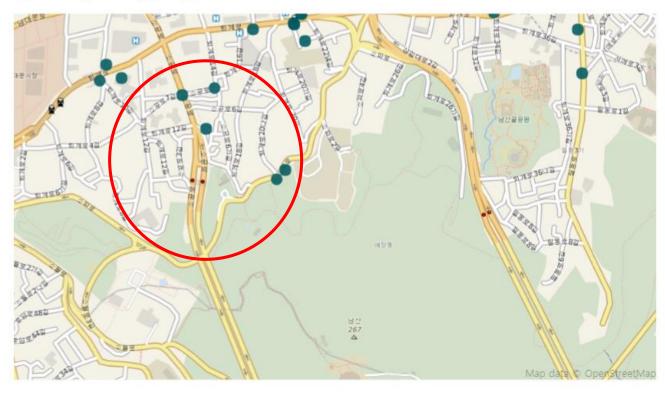
중구



- 중구 남산 터널 중 남산 3호 터널보다 남산 1호 터널에서 교통사고가 몰려 있다
- → 요금소 별 차량 속도에 영향을 주는 요인이 있을 것이라 예상 → 시설물데이터 확인

중구

남산 1호/3호 터널 요금소 CCTV

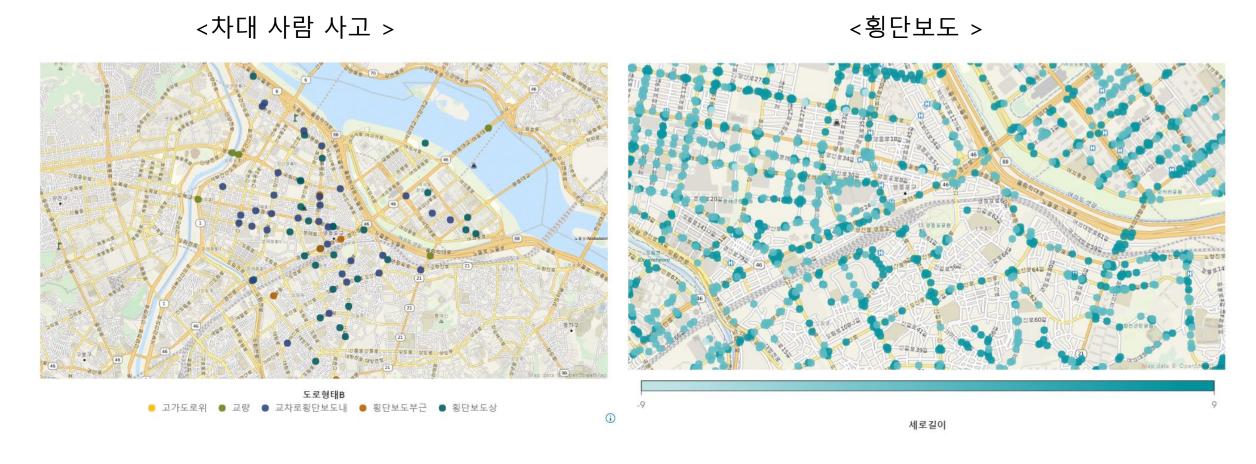


설치목적구분

● 교통단속

→ 같은 요금소가 위치해 있지만 남산 1호 터널근처 CCTV가 없고 그에 비해 3호 터널은 근처 CCTV 있다

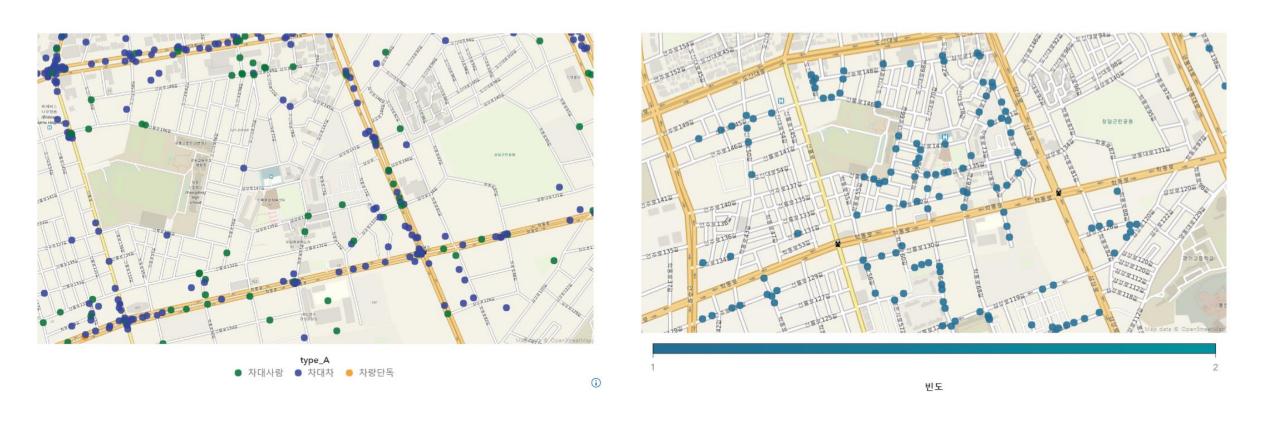
영등포



- 차대 사람 사고는 횡단보도 세로길이가 짧은 구간에서 많이 발생
- → 횡단보도가 짧은 지역은 무단횡단 할 가능성과 신호등이 없는 경우가 있음

강남

<과속방지턱 >



- 강남구 역근처 사거리 대로변에 차대차 사고가 집중되어 있음
- → 과속 방지턱이 있는 곳이 사고가 적으나 대로변에 과속 방지턱 설치는 위험
- → 신호위반 단속 카메라 설치가 보다 효율적

2) DTG 위험 운전자 로지스틱 회귀분석(16년도 6월)

목적: DTG(운행기록)데이터에서 위험운전행동을 하는 운전자 결정하는 요인을 찾기 위해 실시

Y(반응변수): danger(rpm기준으로 위험운전자/일반운전자) 1, 0으로 줌

- 위험운전자 1명 운행기록+ 일반운전자 1명 운행기록

X's(설명변수):

- break(브레이크 신호), VX_f(전방가속도), VX_b(후방가속도), VY_r(우측가속도), VX_l(좌측가속도)
- fr(전방우측가속도), fl(전방좌측가속도), br(후방우측가속도), bl(후방좌측가속도)

2) DTG 위험 운전자 로지스틱 회귀분석

Analysis of Maximum Likelihood Estimates										
				Standard	Wald		Standardize			
Parameter		DF	Estimate	Error	Chi-Square	Pr > ChiSq	d Estimate			
Intercept		1	-4.9441	0.1110	1982.4375	<.0001				
break	0	1	0.6726	0.0313	461.1712	<.0001	0.1849			
break	1	0	0							
vx_f		1	1.0532	0.0300	1234.3497	<.0001	0.4751			
vy_r		1	-1.4949	0.0442	1141.4243	<.0001	-0.4702			
vy_l		1	0.4460	0.0342	170.1686	<.0001	0.1412			
fr		1	0.3537	0.0212	279.1924	<.0001	0.6197			
fl		1	-0.4730	0.0208	518.2138	<.0001	-0.6037			

break=0 브레이크를 **밟지 않을수록** VX_f 전방가속도가 **클수록** VX_l 단순좌측가속도 **크지 않을수록** fr 전방우측가속도 **클수록**

VX_R 단순우측가속도가 **크지 않을수록** fr 전방좌측가속도 **클수록**

위험운전자일 확률이 커짐

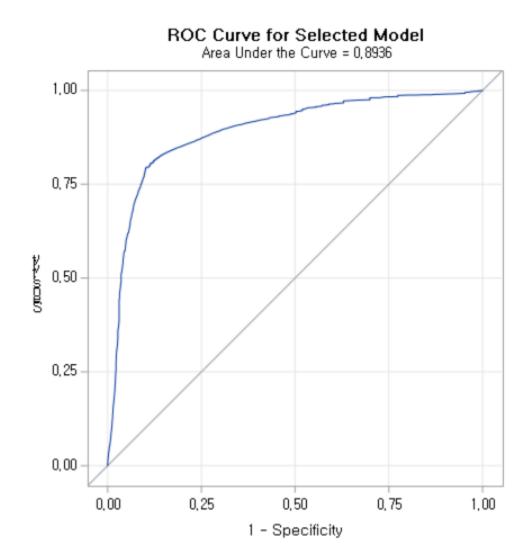
2) DTG 위험 운전자 로지스틱 회귀분석

X's(설명변수) 수준해석:

<VX_f(전방가속도)> 전방가속도가 클수록 위험운전자일 확률이 높은 것으로 보 아 위험운전자를 구분하는 유의한 변수임을 확인

<fr(전방우측), fl(전방좌측)> 좌우방향성을 갖는 전방가속도가 커지면 위험 운전자일 확률이 높음

→ 차선변경을 하거나 직진하는 단일로에서 크게 가속을 하는 행동은 운전자의 태도를 결정하는 중 요한 요인



해결 방안

- 1. 교차로 차대차 교통사고
- 2. 차대사람 교통사고
- 3. 지역별 사고다발구역 교통사고

01. 해결 방안

차대차 교차로 교통사고 분석

문제점: 충돌사고 많이 일어남

- 교차로 진입우선권에 대한 정보제공 미흡
- 신호등 설치 위치 및 설치 형태의 문제점
- 교차로 차선 구별 상태 미흡

해결방안:

- 신호등 밑 표지판 대신 신속한 정보 인지를 위한 전광판 설치
- 꼬리물기나 과속방지를 위한 단속 카메라 설치

01. 해결 방안

차대 사람 교통사고 분석

문제점:

- 횡단 중 사고 발생 경우 보행자 무단횡단
- 좁은길 차량과 보행자가 동시에 통행하는 경우 충돌
- 야간에 운전자가 보행자를 인지하지 못하고 충돌

해결방안:

- 교통사고가 잦은 지역에 횡단보도의 폭을 넓힘
- 인도와 차도 구별하는 도로 분리대 설치하고 폭도 다른 구역보다 넓힘
- 횡단보도와 정지선의 거리를 넓힘
- 야간 통행자들을 위해 신호등 불빛에 LED기능 추가
- 이동차량이 많고 통행자도 많은 복잡한 도로에 대각선 횡단보도 설치

01. 해결 방안

지역별 사고 다발구역 교통사고 분석

중구: 남산 1호 터널과 서울역 앞 4차선 도로

- 남산 1호 터널은 근처 과속 감지 CCTV설치
- 서울역 4차선 도로 정확한 교통정보를 위해 보다 정교한 표지판 설치

영등포: 올림픽대로와 영등포로 교차지점

- 차대 사람 사고가 횡단보도 길이가 짧은 곳에 많이 발생
- 올림픽 대로와 일반도로 교차지점은 속도 제한 표지판 설치

강남: 역삼동 강남역, 역삼역 근처 사거리

- 역 근처 사거리는 유동인구가 많은 지역임에 불구하고 횡단 보도 부족
- 차량 속도 제한을 위해서 규칙적 횡단보도 설치

Appendix

- /*전체데이터 파생변수 생성*/
- data SA133.acc_new;
- set SA133.accident new;
- if att_car='승용' then car='yes';
- else car='no';/*승용차의 비율이 압도적으로 많았고 사고유형에서 승용차와 다른차 들간의 유의한 차이를 파악하기위해 생성*/
- if day='월요일' or day='화요일' or day='수요일' or day='목요일' or day='금요일' then dday='평일';
- else dday='주말';/*주말과 평일의 교통량과 차량이동시간 등 여러가지 차이를 보기 위해 생성*/
- if 6<=hour<=9 then hhour='출';
- else if 10<=hour<=16 then hhour='주';
- else if 17<=hour<=20 then hhour='퇴';
- else hhour='야';/*출퇴근 시간의 교통량이 몰리기 때문에 유의한 차이의 파악하기위 해 생성*/
- if 2<mon<6 then month='봄';
- else if 5<mon<9 then month='여름';
- else if 8<mon<12 then month='가을';
- else month='겨울';
- if road_type_A='교차로' or road_type_A='단일로';
- if law='미분류' then delete;
- run;

/*나이*/
/* 가해자 나이 파생변수 */
data SA133.acc_new1;
set SA133.acc_new;
if att_age<20 then age1='10대이하';
else if 19<att_age<30 then age1='20대';
else if 29<att_age<40 then age1='30대';
else if 39<att_age<50 then age1='40대';
else if 49<att_age<60 then age1='50대';
else if 59<att_age<70 then age1='60대';
else if 69<att_age<80 then age1='70대';
else if 79<att_age<99 then age1='80대이상';
else if 98<att_age then age1='기타';
run;

run;

/* 피해자 나이 파생변수 */
data SA133.acc_new1;
set SA133.acc_new1;
if vic_age<20 then age2='10대이하';
else if 19
vic_age<30 then age2='20대';
else if 29
vic_age<40 then age2='30대';
else if 39
vic_age<50 then age2='40대';
else if 49
vic_age<60 then age2='50대';
else if 59
vic_age<70 then age2='60대';
else if 69
vic_age<80 then age2='70대';
else if 79
else if 79
else if 98
else if 98

- /*사고 유형 파생변수 생성한 전체데이터 (type_new)*/
- data soo.acc_new1_t;
- set soo.acc_new1;
- if type='차대차 정면충돌' or type='차대차 측면직각충돌' or type='차대차 측면충돌' then type_n='차대차-충돌';
- else if type='차대차 추돌' or type='차대차 추돌 진행중' or type='차대차 후진중충돌' then type_n='차대차-추돌';
- else if type='차대차 기타' then type_n='차대차-기타';
- else if type='차대사람 횡단중' then type_n='차대사람-횡단';
- else if type='차대사람 기타' or type='차대사람 길가장자리구역통행중' or type='차대사람 보도통행중' or type='차대사람 차도통행중' then type_n='차대사람-통행';
- else type_n='차량단독';
- run;
- data soo.acc new1 t;
- set soo.acc new1;
- if type='차대차 정면충돌' or type='차대차 측면직각충돌' or type='차대차 측면충돌' or type='차대차 후진중충돌' then type_n='차대차-충돌';
- else if type='차대차 추홀' or type='차대차 추돌 진행중' then type_n='차대차-추돌';
- else if type='차대차 기타' then type_n='차대차-기타';
- else if type_A='차대사람' then type_n='차대사람';
- else type_n='차량단독';
- run;

- /* 독립성 검정 */
- PROC FREQ data=SA133.acc_new1;
- TABLE age1*law/ chisq expected stdres crosslist norow nocol nopercent;
- run;
- PROC FREQ data=SA133.acc new1;
- TABLE age1*road_type_a/ chisq expected stdres crosslist norow nocol nopercent;
- run;
- PROC FREQ data=SA133.acc_new1;
- TABLE age1*hhour/ chisq expected stdres crosslist norow nocol nopercent;
- run;

- /* 로지스틱 회귀분석을 위한 파생변수 생성 */
- /*교통사고데이터 사고유형(type_A) 파생변수*/
- data SA133.acc_new1;
- set SA133.acc_new1;
- if type_a='차대차' then type_aa='차대차';
- else type_aa='그 외';
- if age1='기타' then delete;
- run;
- /*교통사고데이터 연령(age2) 파생변수2*/
- data SA133.acc_new1;
- set SA133.acc_new1;
- if age1='80대이상' or age1='70대' then age2='노년';
- else if age1='60대' or age1='50대' then age2='장년';
- else if age1='40대' or age1='30대' then age2='중년';
- else if age1='20대' or age1='10대이하' then age2='초년';
- run;

- /* 이진 로지스틱 회귀분석 실시 */
- ods noproctitle;
- ods graphics / imagemap=on;
- proc logistic data=MYCAS.ACC_NEW1 plots=(roc);
- class contents road_type_A car dday hhour month age2 / param=glm;
- model type_aa(event='차대차')=contents road_type_A car dday hhour month age2 /
- link=logit selection=stepwise slentry=0.05 slstay=0.05 hierarchy=single
- technique=fisher;
- run;