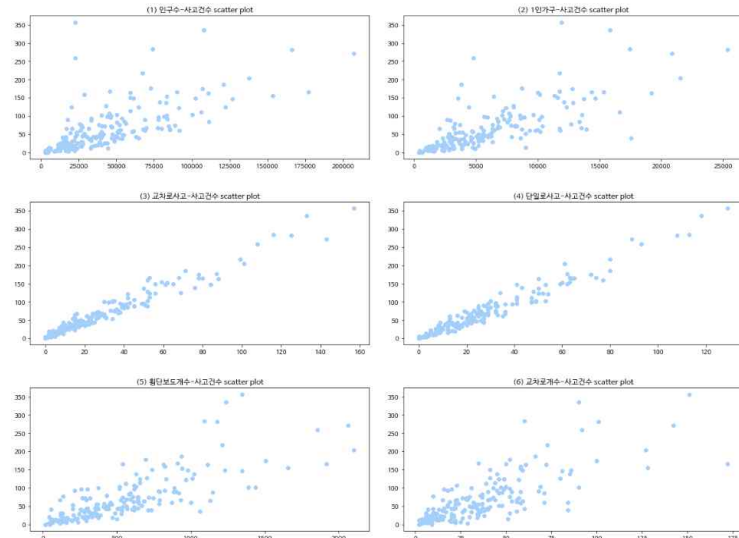


2022 경영경제대학 학술제 참가보고서

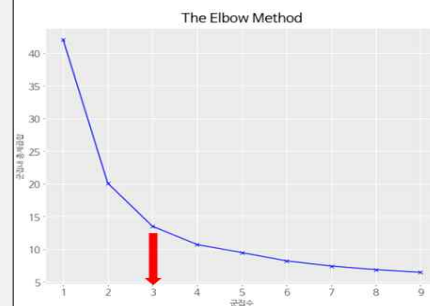
제 목	이륜차 교통사고, 멈춰! (팀명 : 도로시)																																																																																														
신청팀 인적사항	성명	학번	생년월일	연락처	E-mail																																																																																										
	김소은	20196981	2000.03.07.	010-3435-4336	trit1268@naver.com																																																																																										
	김서린	20191376	2000.11.16.	010-6439-6692	tjfls96@naver.com																																																																																										
	서진슬	20192643	2000.01.30.	010-5574-4593	tjwlstmf@naver.com																																																																																										
	장효정	20193521	2000.09.16	010-6456-2495	hyojung000916@daum.net																																																																																										
선정주제	이륜차 사고 다발 지역 예측																																																																																														
주제 선정 이유 (사회적 현상과 관련지어 서술)	<p>코로나19로 인한 언택트 시대에 이륜차 배달 서비스는 폭증하고 있다. 배달 서비스는 “빠름”을 지향하고 운전자는 연락을 받기 위해 도로 위에서 휴대전화 화면을 계속 주시해야 하여 사고의 위험성이 매우 높다. 한국교통안전공단의 이륜차 사고 관련 실태조사에 따르면, 최근 3년간 전체 교통사고 사망자수는 9.7%p 감소했지만, 이륜차 교통사고 건수와 사상자 수가 각각 9.9%p, 12.2%p 증가했다. 전체 교통사고는 줄어드는 추세이지만 이륜차 교통사고는 증가하는 추세임을 알 수 있다. 분석을 통해 이륜차 사고 다발 지역을 예측하고 안전 취약지를 중심으로 배달 서비스 체계 개선을 제안하여 배달 노동자들이 안전하게 일할 수 있는 환경을 조성하고자 한다.</p>																																																																																														
주제 분석 및 연구	<p>1. 데이터 전처리 및 변수 선택</p> <p>1) 데이터 전처리</p> <table><thead><tr><th>자치구</th><th>동</th><th>자치구_동</th><th>경상자수</th><th>중상자수</th><th>사망자수</th><th>부상신고자수</th><th>인구수</th><th>1인가구</th><th>면적</th><th>도로면적</th><th>교차로사고</th><th>단일로사고</th><th>횡단보도개수</th><th>교차로개수</th></tr></thead><tbody><tr><td>0</td><td>강남구 개포동</td><td>강남구 개포동</td><td>30.0</td><td>19.0</td><td>0.0</td><td>5.0</td><td>51732</td><td>3888</td><td>5.27</td><td>1.242666</td><td>34.0</td><td>13.0</td><td>363</td><td>33</td></tr><tr><td>1</td><td>강남구 논현동</td><td>강남구 논현동</td><td>113.0</td><td>40.0</td><td>0.0</td><td>14.0</td><td>45731</td><td>11739</td><td>2.72</td><td>0.641376</td><td>53.0</td><td>74.0</td><td>910</td><td>35</td></tr><tr><td>2</td><td>강남구 대치동</td><td>강남구 대치동</td><td>88.0</td><td>34.0</td><td>1.0</td><td>31.0</td><td>83418</td><td>6426</td><td>3.52</td><td>0.830016</td><td>62.0</td><td>60.0</td><td>940</td><td>58</td></tr><tr><td>3</td><td>강남구 도곡동</td><td>강남구 도곡동</td><td>52.0</td><td>19.0</td><td>0.0</td><td>8.0</td><td>56269</td><td>3280</td><td>2.04</td><td>0.481032</td><td>36.0</td><td>29.0</td><td>432</td><td>36</td></tr><tr><td>4</td><td>강남구 삼성동</td><td>강남구 삼성동</td><td>53.0</td><td>33.0</td><td>0.0</td><td>13.0</td><td>43968</td><td>5669</td><td>3.18</td><td>0.749844</td><td>32.0</td><td>47.0</td><td>888</td><td>43</td></tr></tbody></table> <p>2020년 이륜차 교통사고 데이터에서 경상자수, 중상자수, 사망자수, 부상신고자수를 법정동 기준으로 정리했다. 인구가 많고, 특히 1인가구가 많은 지역에서 배달 서비스를 많이 이용할 것으로 생각하여 법정동별 인구수와 1인가구수를 정리했다. 교통사고와 밀접한 연관이 있는 도로면적, 횡단보도 개수, 교차로 개수를 법정동별로 정리했다. 도로면적은 데이터가 없어 법정동별 (면적)*(도로율)을 계산하여 정리했다.</p> <p>2) 상관관계 파악</p> <div></div> <p>변수들 사이의 상관계수를 heatmap으로 표현한 결과이다. 건수가 적은 사망자수를 제외한 경상자수, 중상자수, 부상신고자수를 합쳐 ‘사고 건수’라는 새로운 변수를 생성했다. 사고 건수와 인구수, 1인가구수, 교차로사고 건수, 단일로사고 건수, 횡단보도개수, 교차로개수의 상관계수는 순서대로 0.67, 0.74, 0.98, 0.98, 0.79, 0.77로 매우 강한 양의 상관관계를 보이므로 K-means Clustering에 사용할 변수로 선택했다. 그러나 면적, 도로면적은 상관계수가 0.29, 0.28로 사고 건수와 상관이 없기 때문에 변수에서 제외했다.</p>					자치구	동	자치구_동	경상자수	중상자수	사망자수	부상신고자수	인구수	1인가구	면적	도로면적	교차로사고	단일로사고	횡단보도개수	교차로개수	0	강남구 개포동	강남구 개포동	30.0	19.0	0.0	5.0	51732	3888	5.27	1.242666	34.0	13.0	363	33	1	강남구 논현동	강남구 논현동	113.0	40.0	0.0	14.0	45731	11739	2.72	0.641376	53.0	74.0	910	35	2	강남구 대치동	강남구 대치동	88.0	34.0	1.0	31.0	83418	6426	3.52	0.830016	62.0	60.0	940	58	3	강남구 도곡동	강남구 도곡동	52.0	19.0	0.0	8.0	56269	3280	2.04	0.481032	36.0	29.0	432	36	4	강남구 삼성동	강남구 삼성동	53.0	33.0	0.0	13.0	43968	5669	3.18	0.749844	32.0	47.0	888	43
	자치구	동	자치구_동	경상자수	중상자수	사망자수	부상신고자수	인구수	1인가구	면적	도로면적	교차로사고	단일로사고	횡단보도개수	교차로개수																																																																																
	0	강남구 개포동	강남구 개포동	30.0	19.0	0.0	5.0	51732	3888	5.27	1.242666	34.0	13.0	363	33																																																																																
1	강남구 논현동	강남구 논현동	113.0	40.0	0.0	14.0	45731	11739	2.72	0.641376	53.0	74.0	910	35																																																																																	
2	강남구 대치동	강남구 대치동	88.0	34.0	1.0	31.0	83418	6426	3.52	0.830016	62.0	60.0	940	58																																																																																	
3	강남구 도곡동	강남구 도곡동	52.0	19.0	0.0	8.0	56269	3280	2.04	0.481032	36.0	29.0	432	36																																																																																	
4	강남구 삼성동	강남구 삼성동	53.0	33.0	0.0	13.0	43968	5669	3.18	0.749844	32.0	47.0	888	43																																																																																	



위 그림은 사고 건수와 군집화에 사용될 변수들 사이의 관계를 표현한 scatter plot이다. 모두 강한 양의 상관관계가 있는 것을 확인할 수 있다.

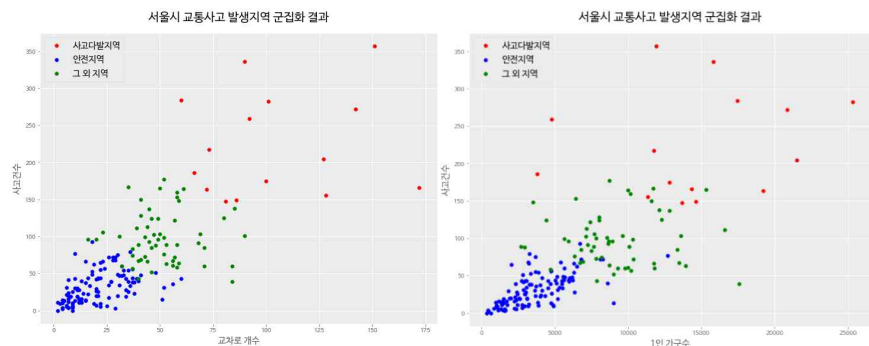
2. K-means Clustering을 통한 이륜차 사고 다발 지역 선정

1) 최적군집수 선정



Elbow Method를 통해 최적 군집 수를 확인했다. 데이터를 3개의 군집으로 나누었을 때의 군집 중심과의 평균 거리가 작아지므로 최적 군집 수를 3개로 정했다.

2) K-means Clustering



3개의 군집으로 군집화한 결과 사고 건수와 교차로 개수, 1인가구수가 주요 기준이 되었다. 앞서 확인했듯이 사고 건수와 교차로 개수, 1인가구수는 아주 강한 상관관계를 보여 적절한 군집화라 판단했다. 각 군집을 ‘사고 다발 지역’, ‘안전지역’, ‘그 외 지역’으로 분류했다. 사고 다발 지역은 사고 건수가 138건 이상, 안전지역은 56건 이하인 지역이다. 이륜차 사고 다발 지역으로 군집화된 ‘강남구 역삼동, 강북구 미아동, 강북구 수유동, 강서구 화곡동, 관악구 신림동, 구로구 구로동, 금천구 독산동, 노원구 상계동, 서초구 서초동, 송파구 잠실동, 양천구 목동, 양천구 신월동, 양천구 신정동, 영등포구 신길동, 중랑구 면목동’ (15개)을 중심으로 이륜차 사고 다발 지역을 예측하고자 한다.

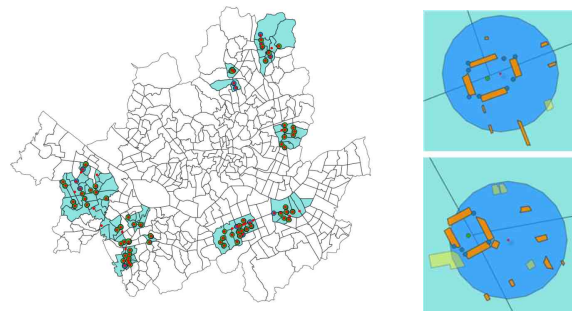
3. GIS기반 공간분석을 통한 이륜차 사고 다발 지역 예측

1) 주요 요인 선정

K-means Clustering을 할 때 기준으로 사용된 교차로 개수와 1인가구수를 공간분석에서도 주요 요인으로 사용했다. 이때 1인가구수는 지도 위에 표시가 어려워 1인가구의 43.9%가 단독주택에 거주하고 단독주택 거주 비율이 전체 인구 대비 높다는 2020 인구주택 총조사의 결과에 따라 단독주택 수로 대체했다. 한국교통안전공단의 이륜차 교통법규 준수 실태조사 결과, 82% 이상이 정지선 위반, 신호 위반에 해당하므로 신호등과 횡단보도 역시 이륜차 교통사고와 밀접한 연관이 있다고 판단했다. 따라서 QGIS를 이용한 공간분석에서는 ‘교차로, 신호등, 횡단보도, 단독주택’을 주요 요인으로 활용했다.

2) 사고 다발 지역의 특징 분석

K-means Clustering을 통해 선정한 행정동을 지도에 표시하고, 한국산업안전보건공단에서 제공하는 ‘이륜차 사고 다발 지역’ 좌표를 그 위에 나타내어 취약지를 파악했다.



사고 건수에 따라 버퍼의 크기를 다르게 설정하여 사고 다발 지역의 특징을 분석했다. 버퍼 크기의 기준은 3-7건 50m, 8-12건 70m, 13-17건 90m, 18-22건 110m, 23-29건 130m, 45건 160m, 75건 200m로 설정했다. 왼쪽의 그림처럼 버퍼를 나타내고, 각 버퍼 내에 있는 교차로(초록), 신호등(파랑), 횡단보도(주황), 단독주택(노랑)를 시각화하여 오른쪽 그림과 같이 표시했다.

버퍼크기	횡단보도수	교차로수	단독주택수	신호등수
50	3.23	0.67	1.72	5.93
70	4.48	0.70	4.93	7.52
90	4.67	1.33	1.00	6.00
110	9.00	1.50	10.50	22.00
130	16.00	2.33	35.33	22.00

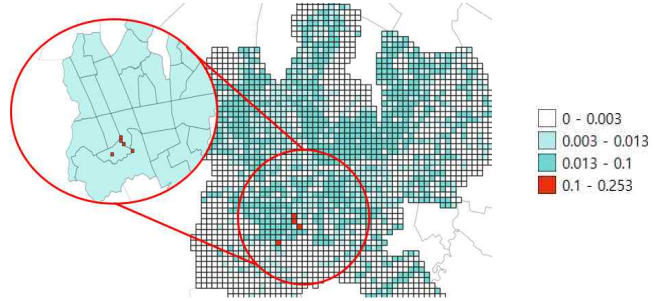
버퍼 크기에 따른 교차로, 신호등, 횡단보도, 단독주택 개수를 파악한 후 평균을 계산한 결과이다. 50m 버퍼가 사고 발생 최소 단위로, 특징을 분석하기에 가장 적합하다고 판단해 버퍼의 지름과 동일한 100*100 그리드를 생성했다.

3) 사고 다발 지역 예측

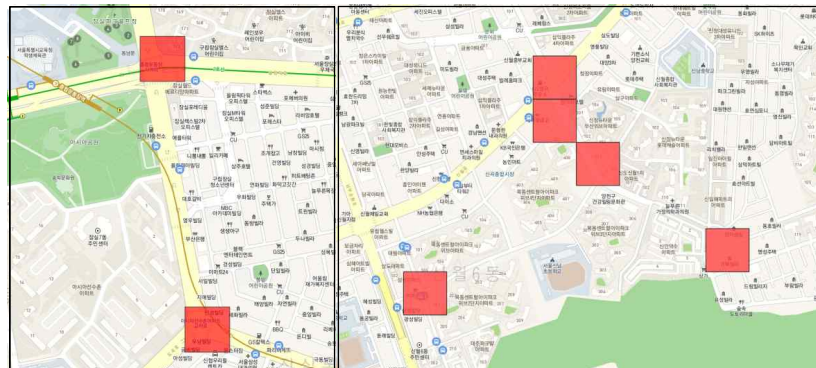
$$f(k) = \frac{x_{k1} + x_{k2} + x_{k3} + x_{k4}}{\sum_{n=1}^p (x_{n1} + x_{n2} + x_{n3} + x_{n4})}$$

$f(k)$ = k번째 그리드의 정규화값
 x_{ij} = i번째 그리드의 특성 j개수
 $(j=1: \text{교차로}, j=2: \text{신호등}, j=3: \text{횡단보도}, j=4: \text{단독주택})$
 p = 해당 행정동의 그리드 개수

각 그리드별로 교차로, 신호등, 횡단보도, 단독주택 개수를 수집한 후, 위의 식에 따라 정규화했다.



정규화된 값을 0.003, 0.013, 0.1을 기준으로 4구간으로 나누어 위 그림(신월6동 예시)과 같이 히트맵으로 표현했다. 값이 클수록 사고가 일어날 위험성이 높다는 의미로 4단계에 해당하는 지역의 위험도가 가장 높으며 4단계에 해당하는 그리드를 빨간색으로 표시했다.



히트맵을 통한 시각화 결과, 최종적으로 ‘상계8동, 잠실7동, 구로1동, 신월6동’ 이 사고가 일어날 가능성이 높은 지역으로 예측되었다.

현실성 및 지속가능성	주제의 현실 적용 가능성	<p>TAAS 교통사고분석시스템에서 실제 2020년도 사고 데이터를 가지고 분석했기 때문에 이륜차 사고 다발 지역의 특징을 알 수 있었다. 이러한 특징을 참고하여 사고를 예방할 수 있는 교통시설물을 설치하거나, 배달 서비스 체계를 개편하는 등의 구체적인 해결 방안 마련이 가능하다. 이는 이륜차 사고 예방에 실질적인 도움이 될 수 있다.</p>
	주제의 발전 가능성 및 타당성	<p>본 연구와 같은 방식으로 최신의 데이터를 가지고 분석한다면 배달 서비스 체계를 개편할 때 최근의 트렌드를 반영할 수 있을 것이다. 주변의 상권, 유형별 주거지역 현황, 배달 서비스 이용 현황 등의 데이터를 추가한다면 더 정교한 분석을 할 수 있을 것이다. 또한, 같은 흐름으로 이륜차뿐만 아니라 다른 교통수단의 사고 취약 지역을 분석할 수도 있고, 사고의 규모로 세분화하여 분석할 수 있다.</p>

창의성 및 전문성	주제에 대한 분석의 차별성 및 독창성	<p>사고는 언제, 어디에서나 일어날 수 있고 이를 예방하기 위한 노력은 중요하다. 단순히 사고 건수가 많은 지역의 단속을 강화하는 것은 일시적인 방안일 뿐, 사고의 근본적인 원인을 개선하는 방안이 아니므로 사고 건수를 줄이지 못한다. 현재 교통사고 관련 서비스는 TAAS의 교통사고 '분석' 시스템일 뿐 '예측' 시스템은 아직 도입되지 않았다. 본 연구는 예측 시스템을 도입할 뿐만 아니라 예측의 근거인 사고의 원인까지 제공하는 시스템을 구축한다. 사고의 범주를 확장하고 AI 기술을 융합하여 실시간으로 정보를 업데이트하면 국민 모두에게 도움이 되는 시스템을 마련할 수 있을 것이다.</p>
	의견 도출의 논리 및 과정	<p>EDA(탐색적 자료 분석)를 통해서 사고 건수와 독립변수들 사이의 상관관계를 파악한다. 사고 건수와의 상관도가 큰 변수를 가장 영향을 많이 주는 변수이므로 정확한 사고 건수 예측이 가능하다.</p> <p>단순히 사고 건수를 기준으로 지역을 군집화할 수 있으나 사고 건수와 상관도가 크게 나온 변수가 많았다. 따라서 복합적인 관계를 고려하기 위해 K-Means Clustering을 통해 사고 건수에 영향을 주는 변수들(교차로 개수, 1인가구수)을 활용해 군집화했다. 그 뒤 사고 건수가 가장 많은 군집을 선택했고, 군집화의 기준이 된 주요 변수는 decision tree를 통해 파악했다.</p> <p>100m*100m의 그리드를 생성하여 각 그리드별로 교차로, 횡단보도, 신호등, 단독주택의 개수를 수집했다. 이후 이를 '하나의 그리드 값 / 해당 행정동 전체의 값'으로 정규화하여 통계량을 계산하였다. 정규화를 진행하여 서로 단위가 다른 변수들의 기준을 동일하게 하였다.</p>
	전공융합 적절성	<p>1) 상관분석(Correlation Analysis) 두 변수 사이에 어떤 선형적 관계가 있는지 분석하는 방법이다. 이때 관계의 강도를 상관관계라고 하며, 피어슨 상관계수 등으로 상관관계를 파악한다.</p> <p>2) K-means Clustering 주어진 데이터를 k개의 클러스터로 묶는 군집화 방법이다. 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작한다.</p> <p>3) Decision Tree 특정 기준에 따라 데이터를 구분하는 모델로, 한 번의 분기 때마다 변수 영역을 두 개로 구분한다. 군집별 특징을 파악할 수 있는 것이 특징이다.</p> <p>위와 같이 자료분석론, 통계프로그래밍, 데이터마이닝 등의 전공과목에서 배운 다양한 분석 방법론을 실제 데이터에 적용하여 분석했다.</p>
의의 및 기대효과		<p>단순히 사고 건수 데이터가 아닌 사고에 영향을 미치는 다양한 변수를 이용해 분석을 진행하였다. 실제 데이터를 활용한 사고 다발 지역을 예측 방법으로 제안함으로써 데이터가 갱신될 때마다 적용 가능하다. 분석 결과로 나온 수치들을 참고하여 사고 다발 지역과 유사한 특징을 보이는 구역을 자치구 단위에서 유심히 관리하여 사고를 예방할 수 있다.</p> <p>이론차 사고는 생명과 직결된 문제이기 때문에 발생하기 전에 예방하는 것이 중요하다. 우발적인 사고를 100% 예측할 수는 없다. 하지만 사고 발생 확률이 높은 지역에 미리 조치를 취함으로써 소중한 생명을 살리고 많은 사회적 비용도 줄일 수 있을 것이다.</p>