Milestone3
최종 프로젝트 보고서

2조

60181655 이서윤

60191434 배정연

60201699 이한별

60201672 문인배

명지대학교

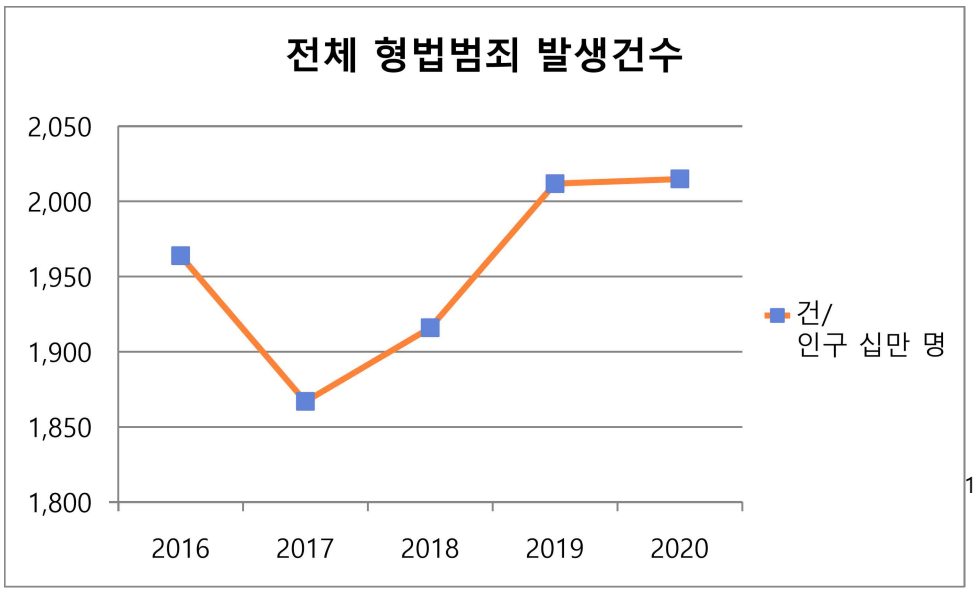
2022.12.7

비즈니스모델링2	**Milestone 1**	**프로젝트 계획서**
전공	**조명**	
응용소프트웨어전공	2조	

<제목 차례>

1. 서론

1.1. 주제 선정 배경



[1][2]

최근 코로나 및 불안정한 대내외적 사회분위기에 편승한 범죄가 증가하고 있다. 과거에는 언론에 보도되어도 대수롭지 않게 넘겼던 사건도 이제는 자신과 연관될 수도 있다고 걱정하는 세태가 되었다. 또한, '마계 인천', '갱의 도시 부산' 등 특정 지역이 범죄로 인해 반복해서 언론에 보도되어 실제 지역 특성이나 인구 규모와는 관련없이 부정적인 선입견이 씌워진 경우도 있다. 이러한 선입견을 통해 해당 지역이 앞으로도 범죄가 다수 발생할 것이라고 무조건적인 추정을 하는 경우도 증가하였고, 이와 더불어 특정인의 말 한마디가 작용하며 '이부망천' 등 해당 지역의 이미지에 또다른 악영향을 주는 경우도 생겨났다.

부정적인 선입견에 더불어, 우리가 통상적으로 범죄와 관련있다고 생각하는 여러 요소들이 존재하는데, 그 예시로는 '치안수준', '해당 지역의 소득수준', '외국인 거주 비율' 등이 있을 수 있다.

1.2. 프로젝트 의의

주제 선정 배경에서 살펴보았듯이, 실제 지역적 특성과는 관계없이 적용된 부정적 선입견을 타파하고, 소득수준 등 범죄와 관련있다고 보여지는 여러 요소들이 실제로 영향을 미치는지 알아보고자 하며, 더 나아가 수집된 자료 및 범죄와의 관련성이 나타난 일부 요소를 바탕으로 각 지역별 범죄 발생건수, 특히 강력범죄에 대한 발생건수를 예측해보고자 한다.

이 프로젝트에서는 당해 프로젝트의 주제 선경 배경과 같이 시군구별로 강력범죄발생건수를 살펴보아 실제로 범죄율이 높은지를 살펴본 뒤, 이를 통해 앞으로 각 시군구별로 범죄율 추이가 어떠할지 몇 가지 변수를 통해 살펴볼 예정이다.

---

1) 출처: 범죄율, 국가지표체계

2) 범죄율 공식은 다음과 같다: $\text{형법범죄율} = \frac{\text{형법범죄발생건수}}{\text{해당지역총인구}} \times 100000$

2. 이론적 배경

2.1. 선행연구 고찰

2.1.1. 범죄 발생에 영향을 주는 지역적 요소에 대한 통상적 고찰

우선, 지역별 범죄 발생건수를 예측하기 위해 어떤 지역적 요소가 범죄 발생과 관련이 있는지를 살펴보아야한다. 이 단계에서는 통상적으로 범죄에 영향을 줄 수 있는 요인을 살펴볼 것이다. 이후, 이러한 요인이 지역적 특성으로 반영된 데이터를 통해 각 요소가 범죄 발생과 밀접한 상관관계를 가지고 있는지를 살펴볼 것이다.

경제적 요인은 전통적으로 범죄 발생을 유발하는 요소로써 고려되어왔다. 특히, 해당 지역의 소득수준이 비교적 낮거나, 지역 경제에 큰 역할을 하지 못하는 비청장년층의 비율이 높다면, 이 지역민들은 대체로 빈부격차 혹은 상대적 박탈감을 더 잘 느낄 수 있는 환경에 노출되어있다고 볼 수 있으며, 이런 상황은 이들을 범죄로 이끌 가능성이 높다고 볼 수 있겠다.

지역 인구 구조는 범죄 발생 요소 중 하나로써 꾸준히 거론되는 요소이다. 이 중 외국인 거주 비율이 범죄 발생에 영향을 줄 수 있다고 보인다. 근래들어 국내 외국인 유입이 상당히 활발해졌고, 마찬가지로 외국인 범죄도 증가하는 추세에 있다. 어떤 논문에 따르면, 체류 외국인이 1.7배 증가하는 동안 외국인 범죄는 2.36배 증가하였다[3]는 연구결과도 존재하여, 외국인 비율과 범죄 발생이 무관하지는 않다고 판단할 수 있다. 또한, 해당 지역에 유동인구가 많이 상주한다면, 범죄 발생 후 인파에 묻혀서 범죄 현장에서 멀리 도피하려는 범죄자 특성상 유동인구가 많은 곳이 범죄 발생을 유도할 수 있을 것이라고 보인다. 범죄학에서 범죄율을 분석하는데 주목하는 성비, 연령대 등도 원인이 될 수 있다.

또한, 해당 지역의 치안수준을 들 수 있는 데, 예를 들어 경찰관 배치 수 및 실제 관서 수 또한 범죄 발생을 억제하는 요소로써 고려해볼 수 있다. 상식적으로 치안수준이 높은 곳에서는 그 자체로써 범죄 발생을 억제하는 수단이 될 수 있기 때문이다.

마지막으로, 지역적 특성을 고려해볼 수 있다. 어떤 지역이 인구밀도가 높은 경우, 사람과 사람 간의 거리가 타 지역에 비해 상대적으로 가까워서 이러한 주거환경이 이웃간의 마찰 등으로 스트레스를 유발하거나 절도와 같은 범죄 행위를 쉽게 저지를 수 있는 환경이 될 수 있다. 또한, 지역민의 생활 만족도에는 해당 지역의 안정적인 주거환경 및 녹지 혹은 노후화된 시설의 비율이 연관이 있을 수 있다. 특히, 비교적 녹지가 적은 지역에 거주하는 사람이 그렇지 않은 사람보다 더 높은 공격성을 띤다는 연구결과도 있다.[4] 또한, 상업지구가 많은 지역은 그 특성상 잦은 전입/출이 발생하여 혼잡한 분위기를 조성할 수 있으며, 유동인구가 많아져 크고 작은 범죄가 발생하는 원인을 제공할 수 있다. 이러한 지역에서의 거주는 경제적 요인과 비슷하게 삶의 만족도를 저하시키고 상대적 박탈감을 증가시키므로써 범죄발생의 요인이 될 수 있다.

3) 인용 논문: 한국의 강력 범죄 발생 추이 및 통제 요인 연구, 고려대학교, 권태연 등 2명, 2016.11.5.
4) 인용 논문: 도시녹지와 옥외범죄율 간의 상관관계 연구, 한국조경학회지, 47(1) : 49~56, 김영제, 2019.2.

2.1.2. 결론

정리하자면, 우리는 통상적으로 다음의 지역별 요소들이 어떤 지역의 범죄 발생 건수에 영향을 준다고 보았다.

① 지역별 소득수준
② 지역별 청장년층 거주 수준
③ 지역별 외국인 거주 수준
④ 지역별 성비
⑤ 지역별 치안수준
⑥ 지역별 전입출 수준
⑦ 지역별 상업구역 면적
⑧ 지역별 녹지 면적

3. 프로젝트 설계

3.1. 데이터 소개

3.1.1. 데이터 출처

프로젝트에 사용할 데이터는 모두 통계청[5], 혹은 이를 바탕으로 한 분석자료나 통계청에 준하는 공신력을 가진 기관에서 제공하는 데이터를 기반으로 한다.

3.1.2. 사용 데이터

2.1.2.에 기반하여, 이 프로젝트에서는 다음과 같은 데이터를 사용 및 조합할 예정이다.

순번	대분류	데이터명	데이터 주기	출처
1	치안수준	경찰청_전국 경찰서별 강력범죄 발생 현황	년	공공데이터
2	소득수준	시·군·구별_근로소득_연말정산_신고현황_주소지	년	KOSIS
3	인구구조	시군구_경제활동인구_총괄	년	
4	인구구조	시군구별_외국인주민_현황	월	
5	인구구조	남녀성비(시도/시/군/구)	월	
6	치안수준	경찰청_경찰관서 위치 주소 현황	년	
7	치안수준	경찰청_전국 경찰서별 경찰관 현황	년	공공데이터
8	지역특성	용도지역_시군구_도시지역	년	KOSIS
9	지역특성	시군구별_이동자수	년	
10	지역특성	용도지역_시군구_도시지역	년	
11	지역특성	행정구역분류 총괄표	수시	통계청

3.1.3. 프로젝트 적용 데이터

순번	예측변수명	예측변수 설명	출처 데이터명
1	d20_col_code	시도시군구_Code	행정구역분류 총괄표
2	d20_col_name	시도시군구_Name	행정구역분류 총괄표
3	d20_col_1	범죄발생(건)	경찰청_전국 경찰서별 강력범죄 발생 현황
4	d20_col_2	경찰관서수(개)	경찰청_경찰관서 위치 주소 현황
5	d20_col_3	급여액(1백만원)	시·군·구별_근로소득_연말정산_신고현황_주소지
6	d20_col_4	주거지역(㎡)	용도지역_시군구_도시지역
7	d20_col_5	상업지역(㎡)	용도지역_시군구_도시지역
8	d20_col_6	공업지역(㎡)	용도지역_시군구_도시지역
9	d20_col_7	녹지지역(㎡)	용도지역_시군구_도시지역
10	d20_col_8	미지정지역(㎡)	용도지역_시군구_도시지역
11	d20_col_9	남자수(명)	남녀성비(시도/시/군/구)
12	d20_col_10	청장년층 수(명)	시군구_경제활동인구_총괄
13	d20_col_11	거주외국인수(명)	시군구별_외국인주민_현황
14	d20_col_12	순이동(명)	시군구별_이동자수
15	d20_col_13	경찰관수(명)	경찰청_전국 경찰서별 경찰관 현황

5) URL: kosis.kr

3.1.4. 원본 데이터 구조 예시

예)

1) 범죄발생 지역

수록기간: 년 2014 ~ 2020 / 자료갱신일: 2021-12-20 / 주석정보

시점 | 증감/증감률 | 행렬전환 | 열고정해제

발생지역별(1)	범죄별(1)	2018	2019	2020
계	범죄발생총건수(A) (건)	1,738,190	1,767,684	1,714,579
	인구(B) (명)	51,826,059	51,849,861	51,829,023
	A/B x 100,000 (건/10만명)	3,353.9	3,409.2	3,308.1
서울	범죄발생총건수(A) (건)	341,288	340,504	318,320
	인구(B) (명)	9,765,623	9,729,107	9,668,465
	A/B x 100,000 (건/10만명)	3,494.8	3,499.8	3,292.4
부산	범죄발생총건수(A) (건)	126,230	128,725	127,673
	인구(B) (명)	3,441,453	3,413,841	3,391,946
	A/B x 100,000 (건/10만명)	3,667.9	3,770.7	3,764
대구	범죄발생총건수(A) (건)	79,489	82,905	79,182
	인구(B) (명)	2,461,769	2,438,031	2,418,346
	A/B x 100,000 (건/10만명)	3,228.9	3,400.5	3,274.2
인천	범죄발생총건수(A) (건)	94,605	101,289	96,340
	인구(B) (명)	2,954,642	2,957,026	2,942,828
	A/B x 100,000 (건/10만명)	3,201.9	3,425.4	3,273.7
광주	범죄발생총건수(A) (건)	51,223	52,827	49,869
	인구(B) (명)	1,459,336	1,456,468	1,450,062
	A/B x 100,000 (건/10만명)	3,510	3,627.1	3,439.1
대전	범죄발생총건수(A) (건)	48,184	47,929	46,960
	인구(B) (명)	1,489,936	1,474,870	1,463,882
	A/B x 100,000 (건/10만명)	3,234	3,249.7	3,207.9
울산	범죄발생총건수(A) (건)	25,323	25,814	24,911
	인구(B) (명)	1,155,623	1,148,019	1,136,017
	A/B x 100,000 (건/10만명)	2,191.3	2,248.6	2,192.8
세종	범죄발생총건수(A) (건)	5,331	6,300	6,651
	인구(B) (명)	314,126	340,575	355,831
	A/B x 100,000 (건/10만명)	1,697.1	1,849.8	1,869.1
경기	범죄발생총건수(A) (건)	429,953	433,498	429,487
	인구(B) (명)	13,077,153	13,239,666	13,427,014

예시1) 범죄발생지역 데이터

경찰서	경찰관
서울중부	497
서울종로	620
서울남대문	472
서울서대문	648
서울혜화	468
서울용산	663
서울성북	518
서울동대문	765
서울마포	766

예시2) 경찰청_전국 경찰서별 경찰관 현황

3.1.5. 프로젝트 적용 데이터 예시

예) 프로젝트 적용 데이터

d20_col_cd	d20_col_name	d20_col_1	d20_col_2	d20_col_3	d20_col_4	d20_col_5	d20_col_6	d20_col_7	d20_col_8	d20_col_9	d20_col_10	d20_col_11	d20_col_12	d20_col_13
11010	서울특별시 종로구	2910	23	15754700	9830838	2937285	0	11204384	0	72635	99443	24201	-1622	1111
11020	서울특별시 중구	3225	17	7865918	6037439	3911718	0	25135	0	61222	85116	24248	-956	994
11030	서울특별시 용산구	2739	10	3109922	11337245	1557493	0	8975262	0	110722	154226	36413	1297	690
11040	서울특별시 성동구	2247	10	9287969	9947485	473145	2051234	4332406	0	143387	206711	23580	-7943	623
11050	서울특별시 광진구	3346	12	6961669	11598584	201658	0	5274873	0	167310	251393	42903	-4777	717
11060	서울특별시 동대문구	3240	12	5629534	13039923	828159	0	383081	0	169769	238146	41102	-3227	812
11070	서울특별시 중랑구	3543	9	6079037	10821733	357606	0	7352187	0	195165	274638	17564	-1998	749
11080	서울특별시 성북구	2425	12	3103521	17791331	397229	0	6432862	0	210810	303471	29595	-5278	924
11090	서울특별시 강북구	2611	9	7363982	10979002	272237	0	12384780	0	150143	206220	12611	-5031	660
11100	서울특별시 도봉구	2097	9	3824449	8710578	267953	1473803	10387632	0	158620	220647	8048	-7457	557
11110	서울특별시 노원구	3574	9	4500363	13929759	590480	0	21026587	0	252383	361577	13537	-9531	796
11120	서울특별시 은평구	3219	11	6853020	15376440	510002	0	13902080	0	230279	325340	15622	199	890
11130	서울특별시 서대문구	2374	9	7581482	15439244	283546	0	1963055	0	149360	217752	30516	2925	670
11140	서울특별시 마포구	3340	9	5132923	13383946	940109	0	9559911	0	175150	263518	27535	-2452	875
11150	서울특별시 양천구	3107	9	12489338	12536636	796830	92872	4043275	0	222824	313141	14666	-3940	750
11160	서울특별시 강서구	4150	11	5643902	14718309	1324125	2920235	22500605	0	280034	408695	22005	-12375	924
11170	서울특별시 구로구	3978	9	6381557	10269448	513589	4195729	5157562	0	199673	279606	100542	-3139	727
11180	서울특별시 금천구	2481	6	14594561	5845488	149622	4121945	2895630	0	117226	165150	59776	-1210	558
11190	서울특별시 영등포구	4902	11	8272740	8013840	2607260	5024926	8708804	0	187647	261623	101558	10783	995
11200	서울특별시 동작구	2974	8	7824542	13790984	344623	0	2252275	0	189152	275008	34619	-5154	658
11210	서울특별시 관악구	4858	10	6122572	15322132	395303	0	13846047	0	248339	362863	53119	-4925	904
11220	서울특별시 서초구	4205	13	6510425	18892364	1320064	0	26687459	0	203359	282871	16785	-6419	1074
11230	서울특별시 강남구	6648	16	2628977	24188974	1677285	0	13631009	0	257999	364545	20672	-6257	1491
11240	서울특별시 송파구	5123	11	2293687	20962824	2363811	0	10530667	0	322299	463077	23866	-9529	1040
11250	서울특별시 강동구	3600	10	4560624	12965432	680697	0	10910074	0	226216	315438	15814	23166	742
21010	부산광역시 중구	1331	6	612248	895634	1883211	191057	393204	1103840	20282	26058	4655	-99	317
21020	부산광역시 서구	1041	5	3509702	4382073	1103317	872926	8012630	14991826	52341	66836	5126	585	347
21030	부산광역시 동구	1379	6	4531302	2521908	2495193	1721887	3090992	1763325	43509	55417	5239	1230	402
21040	부산광역시 영도구	981	5	6793310	4813394	797831	1811599	7603070	41139788	55846	70861	5094	-2479	361
21050	부산광역시 부산진구	5494	10	687250	11070225	4841137	0	13755174	0	173920	242231	7841	3023	851
21060	부산광역시 동래구	2350	7	6116418	9460208	1159033	154027	5924309	0	131735	178971	3587	-248	531

예) 프로젝트 적용 데이터 변수 설명 데이터

	A	B
1	var_name	var_description
2	d20_col_code	시도시군구_Code
3	d20_col_name	시도시군구_Name
4	d20_col_1	범죄발생(건)
5	d20_col_2	경찰관서수(개)
6	d20_col_3	급여액(1백만원)
7	d20_col_4	주거지역(㎡)
8	d20_col_5	상업지역(㎡)
9	d20_col_6	공업지역(㎡)
10	d20_col_7	녹지지역(㎡)
11	d20_col_8	미지정지역(㎡)
12	d20_col_9	남자수(명)
13	d20_col_10	청장년층 수(명)
14	d20_col_11	거주외국인수(명)
15	d20_col_12	순이동(명)
16	d20_col_13	경찰관수(명)

3.2. 데이터 정제 및 전처리

3.2.1. 데이터 단위

이 프로젝트에서 사용할 데이터는 '시군구'를 기본 단위로 한다. 이때의 '시군구'란, 단체장을 선거로 선출하는 '자치단체'를 기준으로 한다. 이는 일반구와 같이 자치단체가 아닌 경우는 해당 지역단체의 상위 자치단체에 포함시켰다는 뜻이다. 예를 들어, '경기도 성남시 분당구'나 '경기도 성남시 수정구'처럼 도청 예하 일반구와 같이 자치단체가 아닌 경우는 상위 자치단체인 '경기도 성남시'에 통합하여 취급하였다. 마지막으로, 단순히 지역명을 ID로 사용하게 되는 경우, 오타 등의 문제가 발생할 수 있으므로, 대한민국 정부에서 각 지역별로 부여한 '행정코드'를 통해 시군구를 구분하였다.

3.2.2. 경찰관서 취급 기준

이 프로젝트에서 취급하는 경찰관서는 '경찰청', '경찰서', '지구대', '파출소'이다. 치안센터는 그 특성상 상주인원이 항상 존재하는 것도 아니며, 업무 수준도 일반 민원수준에 그치며, 현황 관리도 탄력적으로 하고 있다는 점으로 인해 실제적인 치안능력을 제공한다고 보기 어려워 이 프로젝트에서 치안센터는 제외하였다.

또한, 치안 데이터를 지역 특성으로 반영하는 과정에서 우리는 경찰청 및 지역구와 관계없는 경찰관서(ex. 고속도로 순찰대)를 지역 특성으로 반영하기 위해, 해당 관서의 소재지를 통해 관서를 반영하였다. 실제로도 경찰청은 지역본부의 역할 뿐만 아니라, 중범죄를 직접 수사 및 검거하기 때문이다. 그러나, 이는 사건의 중경에 따라 관할이 달라지는 것이지 경찰청이 반드시 경찰서보다 뛰어나다는 것은 아니므로 지역별 경찰관서 수에는 별도의 가중치를 부여하지 않았다.

3.2.3. 경찰관 수 및 강력범죄 발생 건수 취급 기준

보안의 문제로 인해 경찰서 예하 지구대 및 파출소의 정확한 경찰관 수 및 강력범죄 발생 건수에 대한 데이터를 확보할 수 없었다. 예를 들어, 강원도 양양군의 경우, 강원속초경찰서에서 해당 지역의 지구대/파출소를 관리하므로 강원도 양양군의 경찰관 수 및 강력범죄 발생 건수가 별도 제공되지 않는다.

이 프로젝트에서는 이러한 한계점을 극복하기 위해 지구대와 파출소 비율을 3:1로 취급하여 경찰서의 경찰관 수 및 강력범죄 건수에서 분리하여 해당 지역에 할당하였다. 예를 들어, 강원도 양양군의 경우, 해당 지역에 위치한 경찰관서는 파출소 4개와 지구대 1개가 있다. 우리는 해당 경찰관서를 관리하는 속초경찰서의 데이터에서 $\frac{6}{13}$을 곱한 값을 강원도 양양군의 데이터로 취급하였고, 해당 데이터를 제외한 데이터를 강원도 속초시의 데이터로 취급하였다. 단, 전라남도 신안군에 한하여 타 지역의 파출소보다도 규모가 작은 파출소가 너무 많아, 이 지역에 속한 파출소 15개 중 4개만 파출소로 인정하여 가중치를 $\frac{4}{28}$로 부여하였다.

3.2.4. 신설/개명 경찰청

2020년 기준으로 '울산북부경찰서', '태안경찰서', '남양주북부경찰서'가 신설되었다. 또한, '화

성동부경찰서'가 '오산경찰서'로 개명하였다. 특히, 남양주북부경찰서는 2020년 기준으로 신설된지 얼마 되지 않았으며(2020년 12월 23일), 남양주남부경찰서와 같이 남양주시를 담당하므로 이 프로젝트에서는 둘을 합쳐 '남양주경찰서'라는 하나의 데이터로 취급하였다.

3.2.5. 시군구 정보 불일치 등 결측치

간혹, 먼 과거부터 축적된 데이터는 과거의 시군구 정보를 삭제할 수 없어 해당 행이 그대로 남아있는 경우가 있다. 가령, 경기도 이천군, 포천군은 각각 1996년 3월, 2003년 10월에 자치시로 승격되었지만, '시군구별_이동자수' 데이터에서는 해당 행이 남아있다. 이런 경우, 2020년 기준으로도 해당 행들은 아무런 값도 없으므로 데이터 수집 단계에서 제거하였다.

3.2.6. 정규화

```
library(caret)
norm.values <- preProcess(district.df[, ], method = c("center", "scale"))
norm.district.df[, ] <- predict(norm.values, district.df[, ])
```

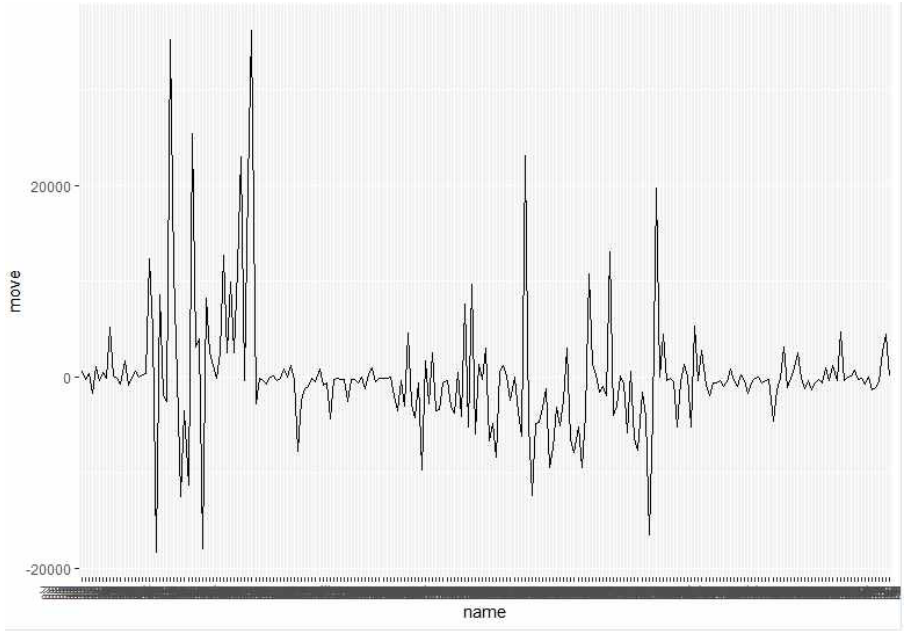
▲ 정규화 과정

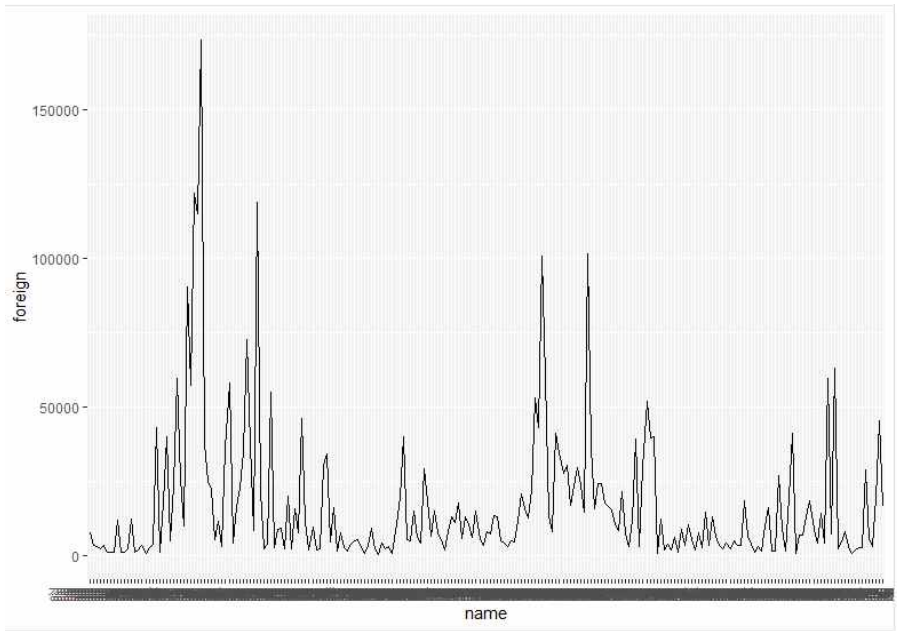
데이터의 단위가 모두 다르므로 정규화가 반드시 필요하였다. 예를 들어, 급여액 데이터는 대체로 6자리 이상의 숫자로 이뤄져있으나, 경찰관서수 데이터는 3자릿수를 넘는 경우가 없었다. 따라서, 이러한 단위의 간극을 좁히기 위해 'Centering' 및 'Scaling'을 통해 모든 변수를 일정 범위 내의 데이터로 조정하였다.

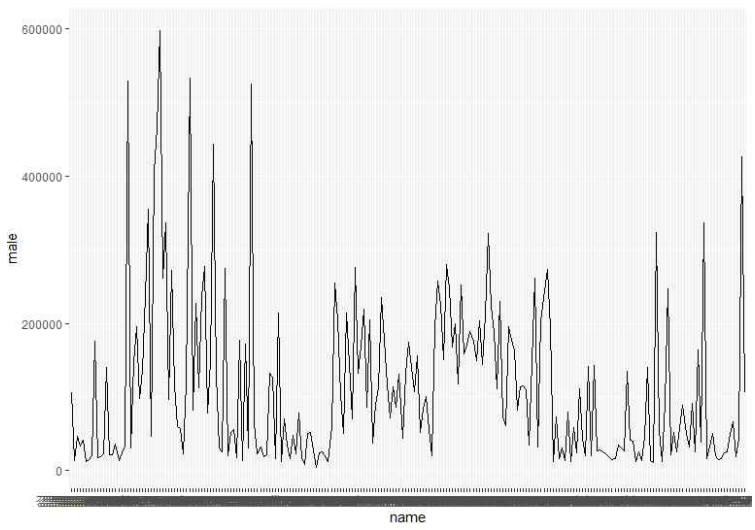
3.2.7. 이상치, 결측치 처리

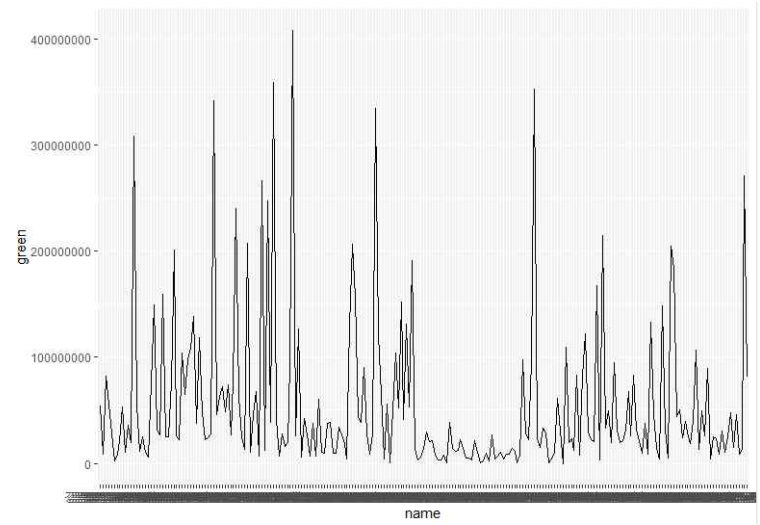
앞서 언급했듯 이전 행정정보로 인해 결측치가 발생한 경우가 있었지만, 그 외의 결측치는 없었다. 이상치 또한 지역적인 특성으로 남기는 것이 좋다는 의견을 반영하여 별도의 처리를 하지 않았다.

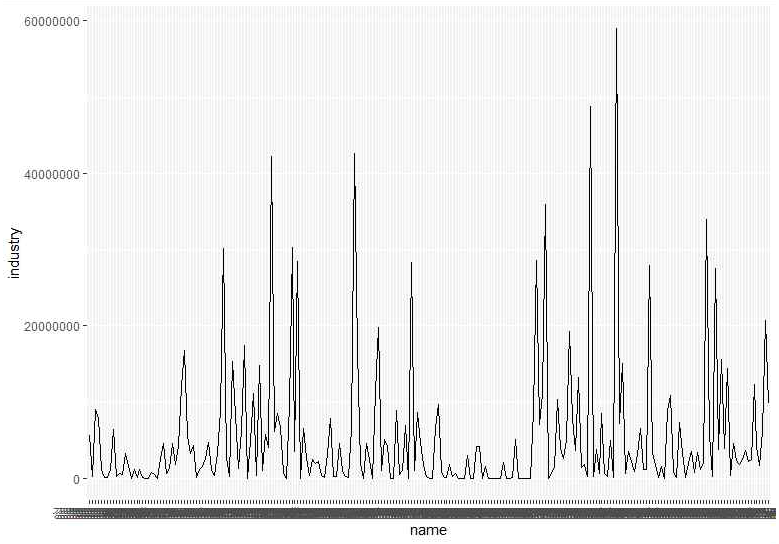
3.3. 데이터 개요

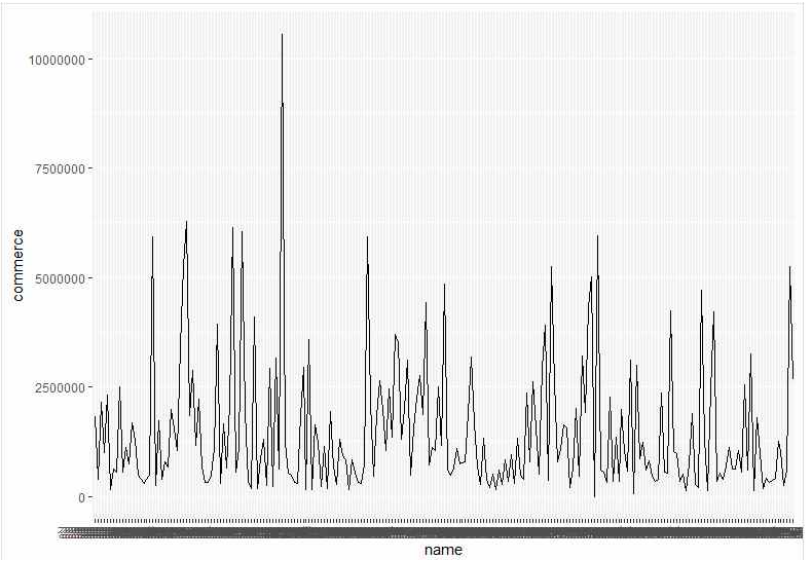

20000
0
-20000
move
name

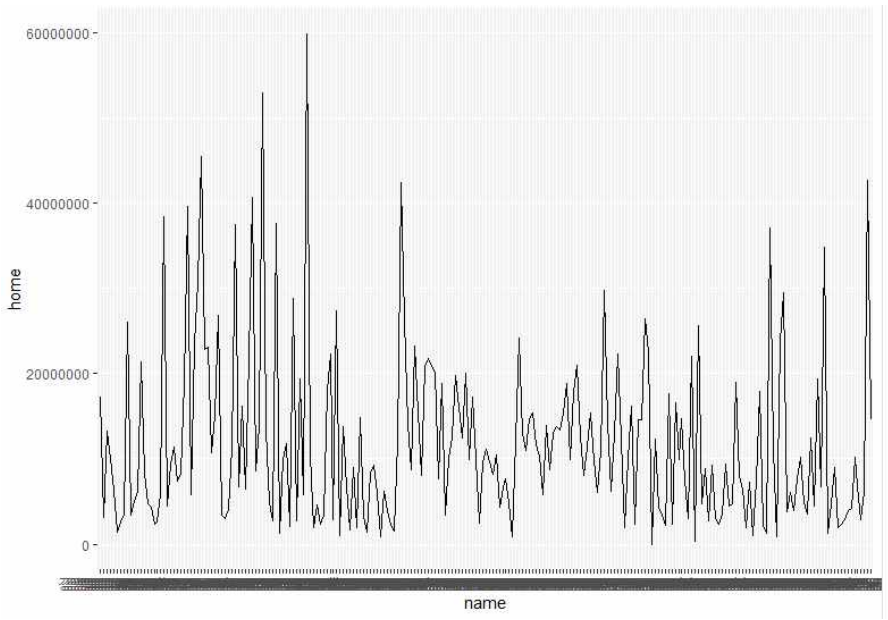

150000
100000
50000
0
foreign
name

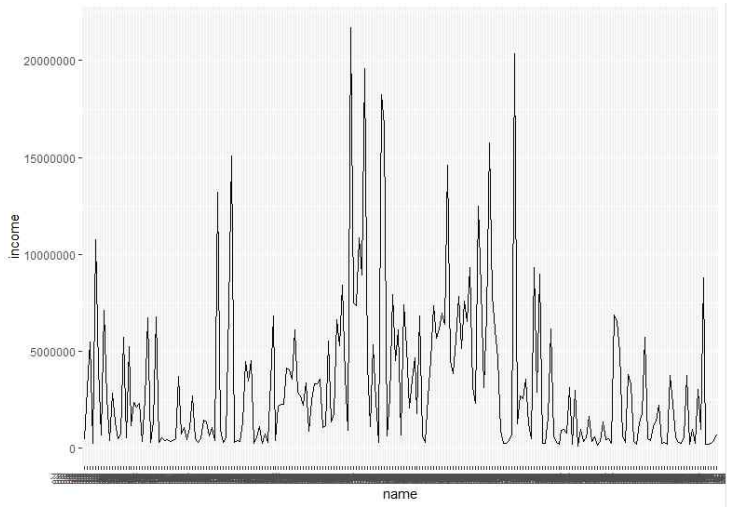

600000
400000
200000
0
male
name

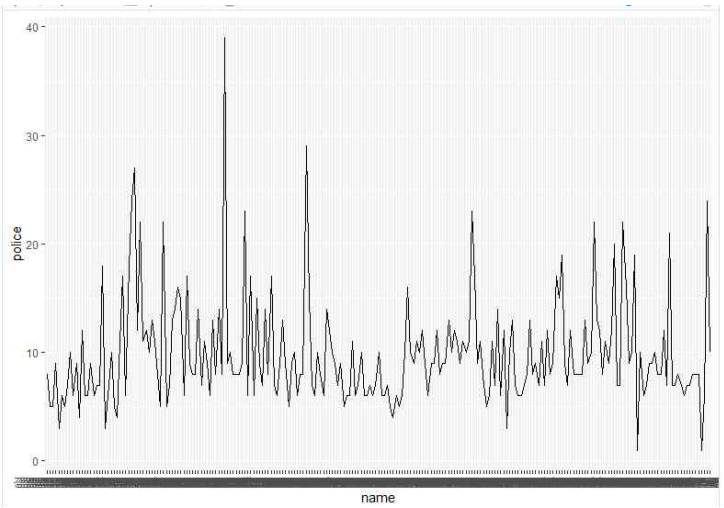

400000000
300000000
200000000
100000000
0
green
name

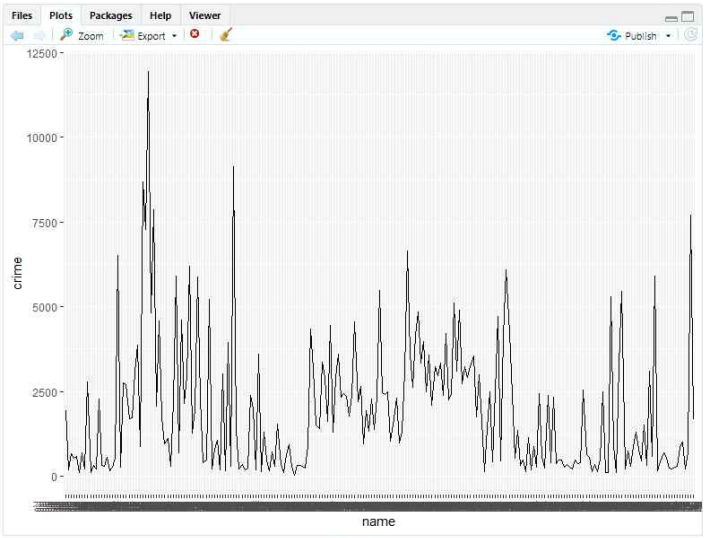

60000000
40000000
20000000
0
industry
name

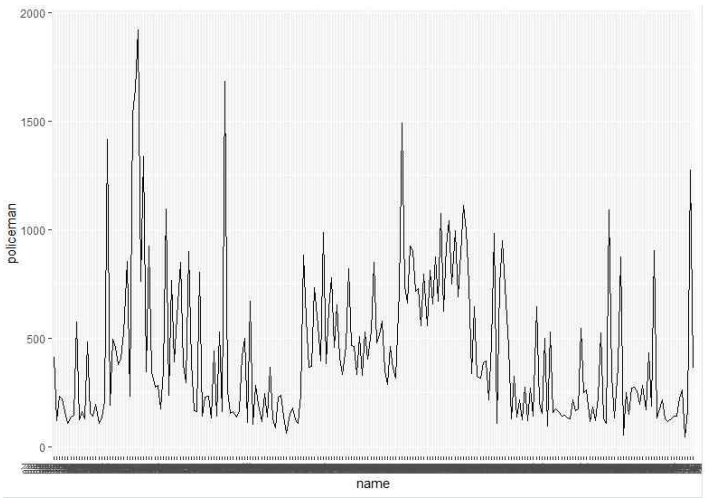

10000000
7500000
5000000
2500000
0
commerce
name


60000000
40000000
20000000
0
home
name


20000000
15000000
10000000
5000000
0
income
name


40
30
20
10
0
police
name


Files
Plots
Packages
Help
Viewer
Zoom
Export
Publish
12500
10000
7500
5000
2500
0
crime
name


2000
1500
1000
500
0
policeman
name

3.4. 모델링

3.4.1. 개요

이 프로젝트에서는 특정한 지역 데이터를 통해 해당 지역의 범죄발생건수를 예측하는 것이 목표이다. 이를 위해서는 'Prediction'에 해당하는 알고리즘을 적용하기로 하였고, 이에 따라 2 가지 알고리즘(Multiple Regression Analysis, K-Nearest Neighbor)을 적용하였다.

3.4.2. Multiple Regression Analysis

다항식의 계수를 통해 종속변수를 예측하는 알고리즘이다.

```
Call:
lm(formula = d20_col_1 ~ ., data = train.crimperdistrict.df)

Residuals:
     Min       1Q   Median       3Q      Max
-0.71806 -0.08862 -0.01692  0.07025  0.68989

Coefficients:
             Estimate Std. Error t value             Pr(>|t|)
(Intercept) -0.001172   0.014901  -0.079              0.93740
d20_col_2   -0.046520   0.028169  -1.651              0.10076
d20_col_3   -0.029082   0.016973  -1.713              0.08872 .
d20_col_4    0.093425   0.051229   1.824              0.07022 .
d20_col_5    0.051299   0.028192   1.820              0.07084 .
d20_col_6   -0.036752   0.019120  -1.922              0.05650 .
d20_col_7   -0.011563   0.023257  -0.497              0.61980
d20_col_9    0.486993   0.490702   0.992              0.32260
d20_col_10  -0.199934   0.485922  -0.411              0.68134
d20_col_11   0.156979   0.023447   6.695       0.000000000418 ***
d20_col_12  -0.065745   0.021410  -3.071              0.00254 **
d20_col_13   0.525223   0.051535  10.192 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.186 on 148 degrees of freedom
Multiple R-squared:  0.9685,    Adjusted R-squared:  0.9662
F-statistic:   414 on 11 and 148 DF,  p-value: < 0.00000000000000022
```

▲ Multiple Regression Analysis Model 생성 모습

3.4.3. K-Nearest Neighbor

레코드가 가진 독립변수를 통해 해당 레코드와 가장 근접한 레코드들의 종속변수를 통해 해당 레코드의 종속변수를 예측하는 알고리즘이다.

3.4.4. 모델 평가

① 모델링 결과

Multiple Regression Analysis

Train Set

	ME	RMSE	MAE	MPE	MAPE
Test set	0.00000000000000001994932	0.1788969	0.1230258	-15.72918	63.29408

Validation Set

	ME	RMSE	MAE	MPE	MAPE
Test set	-0.003890885	0.258606	0.1635281	-52.00016	98.30585

K-Nearest Neighbor

	k	RMSE
1	1	702.6741
2	2	907.7167
3	3	908.6556
4	4	874.2070
5	5	944.0663

② 평가

두 알고리즘 중 하나를 선택한다면, RMSE가 작은 Mulitple Regression Analysis 기반 Model을 선택하는 것이 가장 이상적이라고 본다.

4. 결론

지역별 변수를 통해 강력범죄를 예측하는 모델을 만들어 보았다. 경찰관수(d20_col_13)이나 거주 외국인수(d20_col_11)과 같이 통념처럼 강력범죄에 일조하는 변수도 있는가 하면, 소득수준(급여액, d20_col_3)과 같이 영향이 있다고 보기 힘든 경우도 있었다. 이처럼 강력범죄는 단순히 지역적인 특성만을 따른다고 보기에는 어려운 경향이 있다는 것을 알 수 있다.