

# 제1회 금융 빅데이터 활용 아이디어 공모전 제안서

## - 편의점 상권 분석 및 추천 시스템

팀명: 무르무르

팀 구성원: 김민석, 권지혜, 이은진(고려대학교)

### 1. 아이디어 소개 및 배경

안녕하십니까? 저희는 '편의점 상권 분석 및 추천 시스템' 아이디어를 제안하게 된 무르무르 팀입니다. 최근 편의점 폐업률은 눈에 띄게 증가하고 있습니다. 2019년 2분기 기준 서울시 편의점 5년 생존률은 34.7%이며, 폐업률 또한 3.5%로 증가하는 추세입니다. 저희 팀은 이러한 사회적 현상에 문제의식을 가지고, 새로이 편의점을 개업하는 점주들에게 도움을 주고자 빅데이터를 편의점 상권분석에 활용하여 추천 시스템을 만들어 보고자 합니다.

기존 편의점 점주들은 본사가 추천하는 상권을 기준으로 위치를 결정해 왔습니다. 하지만 상권이 발달해 있다는 이유로 편의점이 과도 밀집된 지역에 개업을 하기도 했으며, 데이터에 기반하지 않은 채로 편의점 위치를 정하기도 했습니다. 이는 편의점수가 무분별하게 늘어나 편의점 폐업률이 늘어나는 결과를 초래했습니다. 상권을 제대로 분석하여 높은 수익률이 예측되는 위치를 점주들에게 추천해 준다면 편의점 폐업률을 낮출 수 있을 것이라는 생각에 위와 같은 서비스를 기획하게 되었습니다.

### 2. 빅데이터 분석 과정

#### (1) 데이터 선정 및 기초 EDA(시각화)

저희는 BC카드에서 제공한 데이터와 "우리마을가게 상권분석서비스(공공데이터)", 각종 금융 수치를 활용하여 상권 분석을 진행하였습니다. 데이터 분석 절차는 다음과 같습니다.

우선, 편의점의 매출에 영향을 미칠 요소들과 총 매출액을 계산해 줄 변수를 설정하였습니다.

1	동별 임대료	공공데이터 활용
2	주위 편의점 점포 개수	
3	주위 슈퍼마켓 점포 개수	
4	주위 유동인구	BC카드 제공, KT데이터 활용
5	소비자 밀집도	BC카드 제공, 망고플레이트 데이터 활용
6	총 매출액	BC카드 제공, BC카드 데이터 활용

[표 1: 변수 설명 및 출처]

임대료는 소상공인의 사업에 가장 중요한 부분 중 하나입니다. 편의점의 경우도 마찬가지입니다.

다. 편의점은 대부분의 점포가 상가 1층, 접근성이 좋은 곳에 위치하고 있습니다. 좋은 위치는 높은 임대료를 수반하며, 많은 점주들이 임대료를 부담인 요소로 꼽고 있습니다. 따라서 임대료를 변수 중 하나로 선정하였으며, 2018년 2분기 데이터를 기준으로 합니다.

**주위 편의점 점포 개수**는 상권에 경쟁자가 얼마나 많은지 파악할 수 있는 좋은 지표입니다. 최근 편의점 본사들이 편의점 밀집 지역에 지점을 내는 것을 허용하여 상권 내 편의점 치킨게임 현상이 일어나기도 한 만큼, 주위에 수요에 비해 많은 편의점 수가 있는지 파악하는 것은 신규 편의점 점주들에게 중요한 요소라 보았습니다. 이 데이터의 경우 2019년 2분기 데이터를 기준으로 합니다.

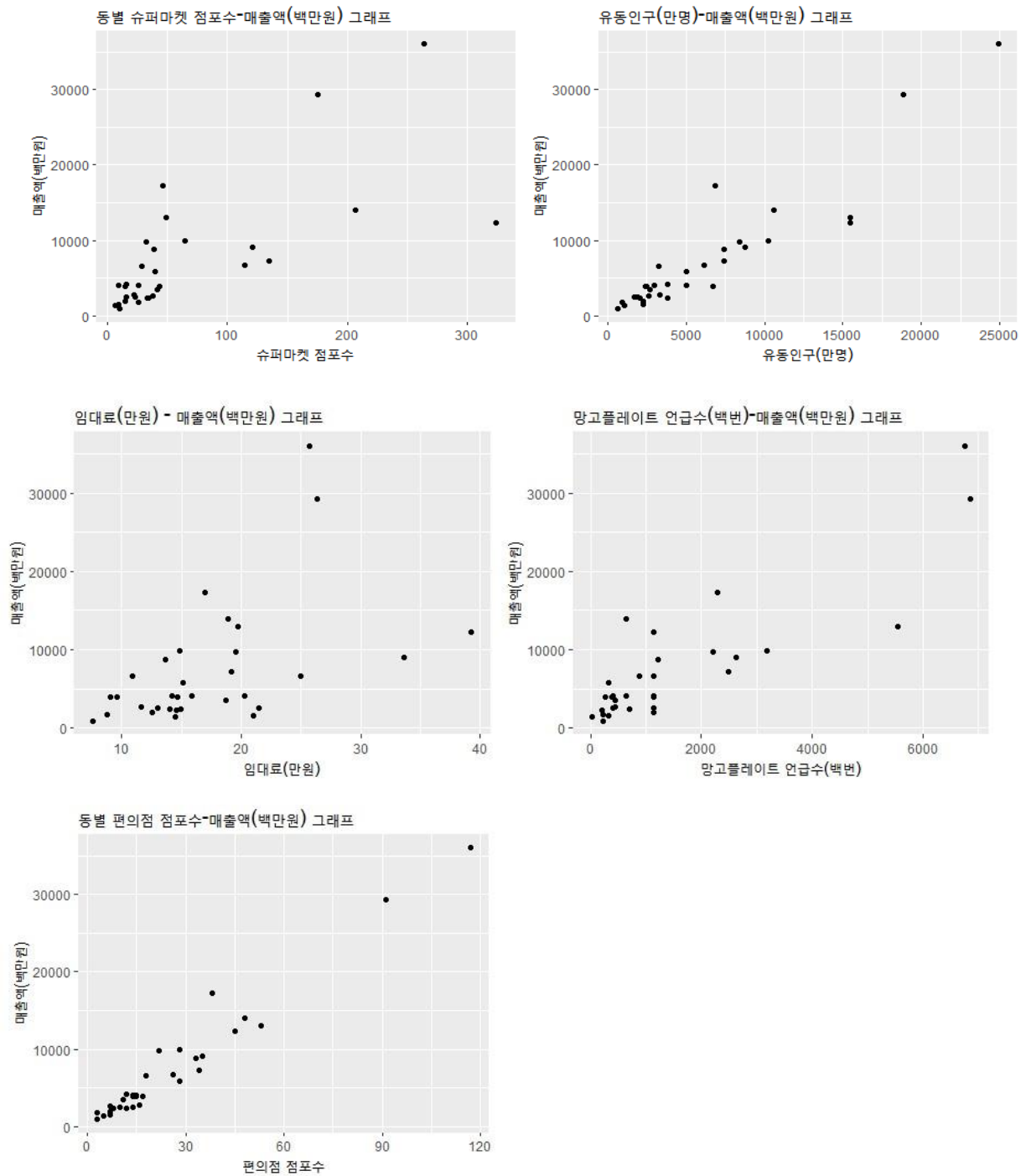
**슈퍼마켓**은 편의점의 대표적인 대체재입니다. 이에 슈퍼마켓이 과도하게 많은 상권 또한 피해야 하므로, 이 변수를 포함하였으며, 2019년 2분기 데이터를 사용했습니다.

편의점의 주 소비층은 '지나가다' 들린 사람에 있습니다. 따라서 주위 **유동인구**가 편의점 상권 파악에 중요한 요소가 될 것이라 판단하였고, 변수로 설정하였습니다. 유동인구 데이터는 BC카드에서 제공한 통신사 KT 데이터를 사용하였으며, 통신사 KT의 점유율(26.3%)을 가중치로 두어 전체 유동 인구수를 추측하였습니다.

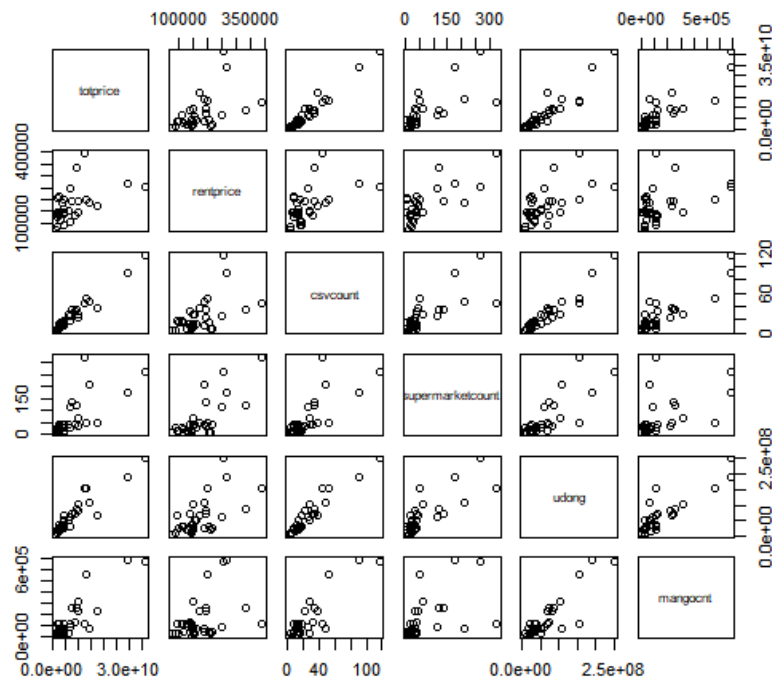
**소비자 밀집도**는 해당 지역에 소비자가 얼마나 몰려 있는지 보여주는 지표입니다. 이는 BC카드에서 제공한 망고플레이트 데이터를 활용하였습니다. 망고플레이트 사용자 밀집도 데이터에는 날짜, 위도, 경도, 활동 횟수가 포함되어 있습니다. 데이터 내 정확한 행정동이 주어지지 않아 Geocoder-Xr(주소 좌표 변환 툴)을 활용해 각 위도, 경도에 해당하는 주소를 변환했고, 해당 주소로부터 행정동을 추출했습니다. 이를 통해 활동 횟수가 일어난 위치(행정동)를 파악했고, 이를 동별로 모두 합하여 밀집도를 계산할 수 있었습니다.

마지막으로 **총 매출액**은 BC카드에서 제공한 BC카드 소비액 데이터를 활용하였습니다. 기존의 데이터를 동별로 합산한 후, BC카드의 신용판매 및 금융점유율(15.9)을 가중치로 하여 총 소비량을 추정하였습니다.

매출액에 영향을 미치는 요소들을 뽑은 후, 이를 그래프로 나타내어 상관관계를 파악해 보기도 하였습니다. 대체적으로 저희 팀에서 추측한 것과 일치하는 형태를 보여주었습니다.



또한, 전체적인 계수간 분포를 보기 위해 매출량-각 변수들 간의 종합 산점도도 그려보았습니다.

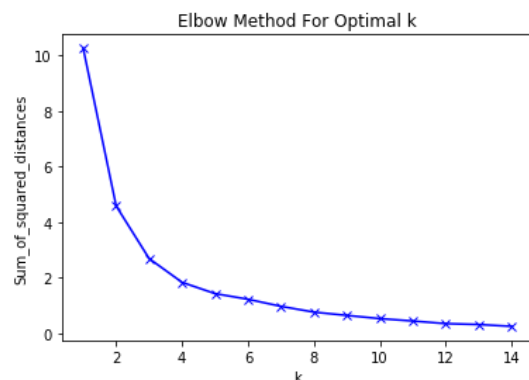


각 변수들은 대체적으로 정비례하는 성향을 가지고 있음을 파악할 수 있었습니다. 변수간 상관이 높다면 회귀모델을 적합했을 때 다중공선성(Multicollinearity) 문제가 생길 수 있으므로, 당초 계획했던 회귀분석 예측 모델 대신 클러스터링을 통한 분석을 진행하기로 하였습니다.

## (2) K-MEANS 클러스터링을 이용하여 동별 특성 파악

앞서 언급한 변수들 중, 주변 편의점과 슈퍼마켓 개수, 임대료(A)는 높을수록 새로이 편의점을 창업하려는 점주들에게 좋지 않습니다. 반면 유동인구와 소비자밀집도, 총매출액(B)은 높을수록 편의점을 개업하기 적합하다고 볼 수 있습니다. 유사한 특성을 가진 동끼리 클러스터로 묶은 후, 각 클러스터별로 A요소, B요소의 상대적 순위를 비교한다면 최적 위치 추천 서비스가 가능해질 것이라 보고, 클러스터링을 실시하였습니다.

Elbow Method를 사용하여 분석한 결과(우측 그래프 참조), 클러스터가 4개일 때 가장 적절하게 군집화를 진행할 수 있다는 결론을 얻게 되었습니다. 이에 따라 4개의 클러스터로 전체 데이터 클러스터링을 진행해 보았으며, 각 클러스터별로 변수 합계를 구해 보았습니다.



Clusters	B			A		
	유동인구	맛집밀집도	총매출	임대료	편의점밀집도	슈퍼마켓밀집도
0	27040831.22	54790.05556	2780832840	142424.6667	10.55555556	22.16666667
1	218761965.8	680453	32645333264	260308.5	104	219.5
2	120887816.5	240405.75	14135352604	237307	46	156.5
3	70816211.5	175688.125	7965932208	189906	28	71.875

[표 2: 4개의 클러스터에 대한 각 변수 평균]

하지만 클러스터를 4개로 구성하니 A요소가 낮고 B요소가 높은, 최적의 편의점 입지 추천 구역을 얻어내기에는 결과가 다소 넓고 애매하였습니다. 좋은 입지와 좋지 않은 입지를 유의미하게 구분할 수 있는 결과가 나오지 않았기에 클러스터 개수를 늘려 진행해 보기로 하였습니다. 클러스터 개수를 늘려가며 결과를 분석해 본 결과, 10개의 클러스터로 분석을 진행하는 것이 가장 적합하다고 판단하였습니다.

Clusters	B			A		
	유동인구	맛집밀집도	총매출	임대료	편의점밀집도	슈퍼마켓밀집도
0	260741798.5	62903.42857	3933894671	146204.4286	14	23.71428571
1	188522003.8	684730	29269237044	263898	91	175
2	347750060.8	230703.25	9365363143	204237.5	29.5	64.25
3	68440593.16	228649	17270405572	170041	38	46
4	71478490.49	38323.4	1505257233	128817.8	5	13.4
5	249001927.8	676176	36021429484	256719	117	264
6	309292851.7	334443.5	12632922579	295023	49	186.5
7	218779631.2	120673	6566501274	175574.5	26.5	79.5
8	154514673	59046.66667	2498573711	149354	11.16666667	27.66666667
9	105817821.3	64087	14005159686	189141	48	207

[표 3: 10개의 클러스터에 대한 변수 평균]

위와 같이 10개로 클러스터링 한 결과, 유의미한 결과를 얻을 수 있었습니다. 다른 클러스터에 비해 확연히 A요소가 낮고 B요소가 높은 0번 클러스터가 가장 추천할 만 한 입지라 판단하였습니다. 저희 팀은 이 원리에 착안하여, 다양한 지역의 클러스터 중 해당 지역이 어느 클러스터에 속하며, 이는 상대적으로 좋은 입지인지, 나쁜 입지인지 정보를 제공한 후 가장 좋은 입지까지 추천해주는 시스템을 구현하였습니다.

### 3. 서비스 구현

서비스는 다음과 같은 형태로 구현될 것입니다.

#### (1) 원하는 지역 범위를 선택한다.

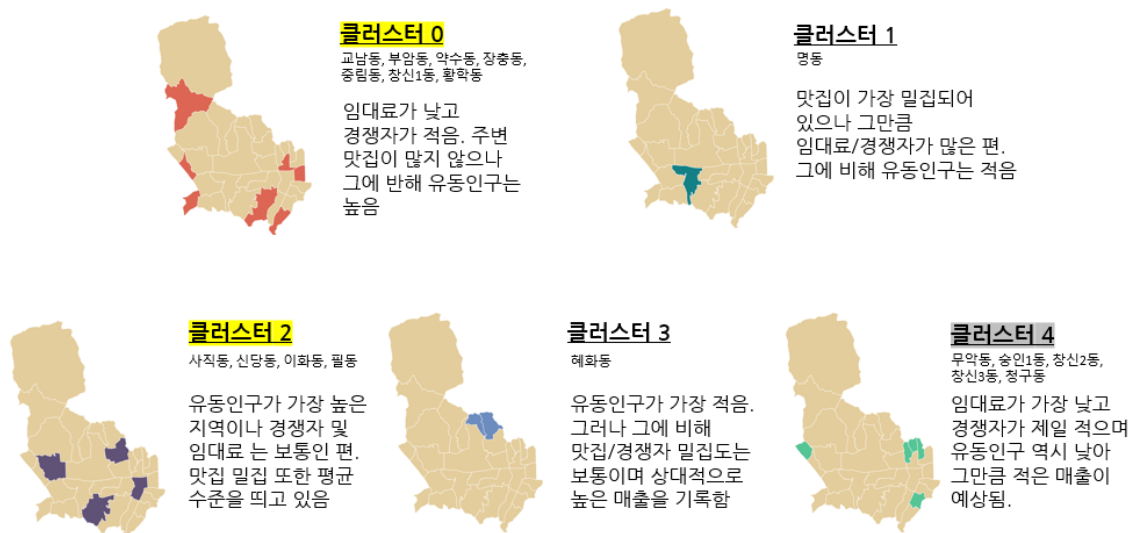
예비 창업자가 창업을 원하는 대략적인 범위를 설정하게 합니다. 예를 들어, 서울 중구와 종로구를 선택하면 이에 맞는 상권 데이터를 불러오게 됩니다. 상권 데이터에는 편의점 매출액, 주위 편의점 수, 주위 슈퍼마켓 수, 임대료, 유동인구수가 포함됩니다.

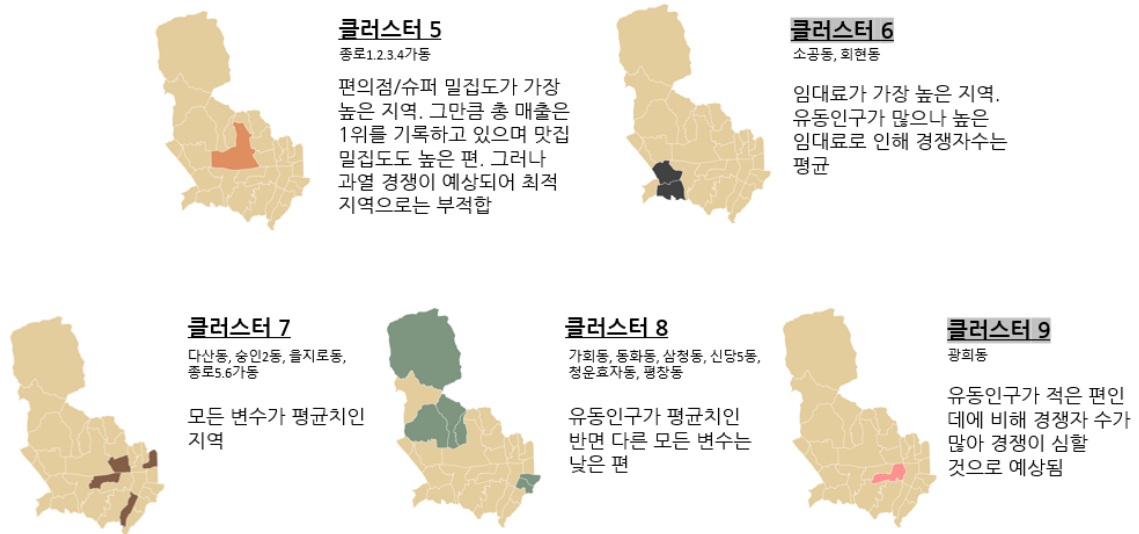
#### (2) 데이터 분석을 진행한다.

지역범위를 선택하는 순간, 지역에 포함되는 상권 분석이 진행됩니다. 유의미한 상권 분석 추천이 가능하도록 클러스터 수를 최적화 한 후, 비슷한 값을 가진 상권끼리 자동으로 클러스터링이 진행됩니다. 추천을 결정하는 기준은 각 변수들이며, A요소(상권에 마이너스 요인)와 B요인(상권에 플러스 요인)으로 구분됩니다.

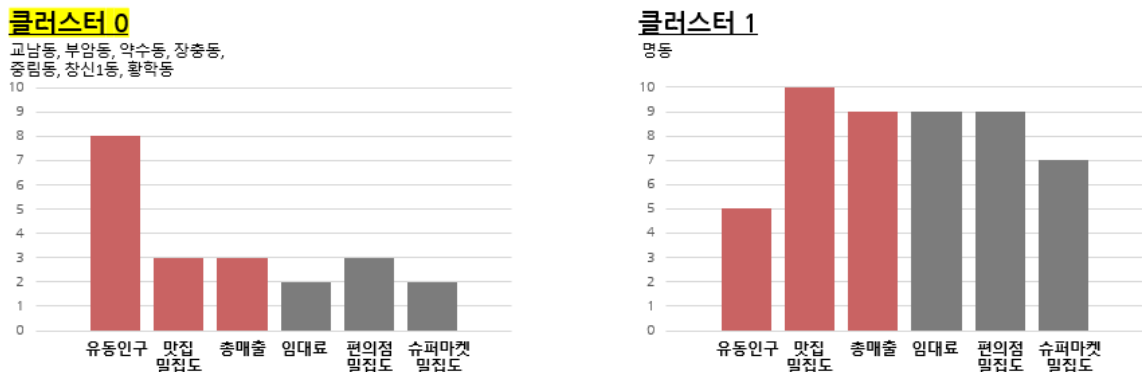
#### (3) 추천 결과를 출력한다.

분석 결과를 가지고, 본 시스템은 A요소의 값이 적으며 B요소의 값이 큰 지역을 순서대로 추천해 줍니다. 또한 예비 창업자가 궁금증을 가지는 상권을 클릭하면 자동으로 분석 결과를 시각화하여, 데이터를 보는데 능숙하지 않은 창업자도 쉽게 상권의 장단을 파악할 수 있도록 해 줍니다.





[ 분석 결과 예시: 종로구, 중구를 지역 범위로 설정하였을 때 ]



[ 분석 결과 예시: 각 클러스터를 클릭하였을 때 볼 수 있는 그래프 ]

## 4. 본 서비스의 함의 및 더 나아갈 점

### (1) 본 서비스의 함의

본 서비스는 데이터를 분석하여 상권을 정하는 데 어려움을 느끼는 예비 편의점 창업자들을 위해 만들어졌습니다. 만약 서비스가 편의점 회사에 도입된다면, 창업자들은 회사의 추천에 신뢰를 가지고 새로 점포를 낼 수 있을 것입니다. 공공 서비스로 만들어진다면 일반인들도 쉽게 본 서비스에 접근하여 입지를 조사할 수 있습니다. 특히 입지 추천 분석 결과가 보기 쉽게 그림과 그래프로 이뤄져 있기 때문에, 직관적인 이해가 가능할 것이라 보입니다.

또한 이미 편의점을 운영하고 있는 점주들이 직접 자신의 가게 상권을 분석해 볼 수도 있습니다.

다. 본 시스템은 상권 별 변수값을 그래프로 나타내 주기 때문에, 편의점 수익이 좋지 않다면 분석 결과를 통해 어떤 부분에서 다른 상권보다 미흡한지, 어떤 부분에서 장점을 가지고 있는지 파악할 수 있습니다. 이는 자신의 점포를 객관적으로 이해하는 데 도움을 줄 수 있을 것입니다.

## (2) 더 나아갈 점

### 가. 상권을 보다 더 잘게 쪼갤 필요가 있다.

본 서비스를 구현할 때, 본디 목적은 블록별로 상권을 구별하는 것이었습니다. 하지만 블록 별 데이터를 구하기가 현실적으로 쉽지 않아 동 별 데이터로 구현해 보았습니다. 이 아이디어가 현실화된다면 데이터를 보다 더 세부적으로 수집하여 상권을 작게 구별할 필요가 있으며, 그렇다면 추천 서비스의 질은 더욱 높아질 것입니다.

### 나. 변수의 개수를 지속적인 기계학습을 통해 늘려간다.

현재 변수 개수는 6개로, 상권을 분석하는데 유용한 정보들을 모아 두었습니다. 하지만 편의점 최적 상권을 예측할 수 있는 더욱 다양한 변수들이 존재할 것이라 생각합니다. 따라서, 지속적으로 편의점 상권과 관련된 코드를 시스템에 학습시키고 보완하여 서비스의 질을 높여 갈 것입니다.

### 다. 타 업종 상권분석으로 서비스의 범위를 넓힌다.

편의점 상권 분석에서 나아가, 다른 업종의 상권을 분석하는 서비스도 내고 싶습니다. 시스템의 구조는 동일하고, 시스템 내에 들어가는 변수만 바꾸어 주면 타 상권도 충분히 분석이 가능합니다. 장기적으로는 다양한 업종의 상권을 분석하는 종합 상권 분석 시스템으로의 발전도 기대해 볼 수 있습니다.

## 5. 마무리하며

학교 근처 너무 많은 편의점이 생겨 상권이 과밀화되고, 급기야 자주 가던 편의점이 문을 닫는 것을 보고 많은 안타까움을 느꼈습니다. 이 아이디어는 문을 닫기 전날, 그 편의점을 찾은 한 팀원과 사장님과의 대화에서 시작되었습니다. “당연히 학교 근처이니 자취생이 많고, 잘 될 줄 알았다” 는 말씀을 듣고, 일반인이 데이터를 객관적으로 분석하기는 정말 어렵고, 이에 기반하여 의사 결정을 내리기는 더욱 어렵겠다는 생각을 가지게 되었습니다. 이 시스템이 꼭 상용화되어 데이터에 대한 지식이 없는 일반인도 최적의 상권을 분석할 수 있길 바랍니다. 감사합니다.