

포스트 코로나 소상공인 개업 :

지표 개발을 통한 리스트 관리

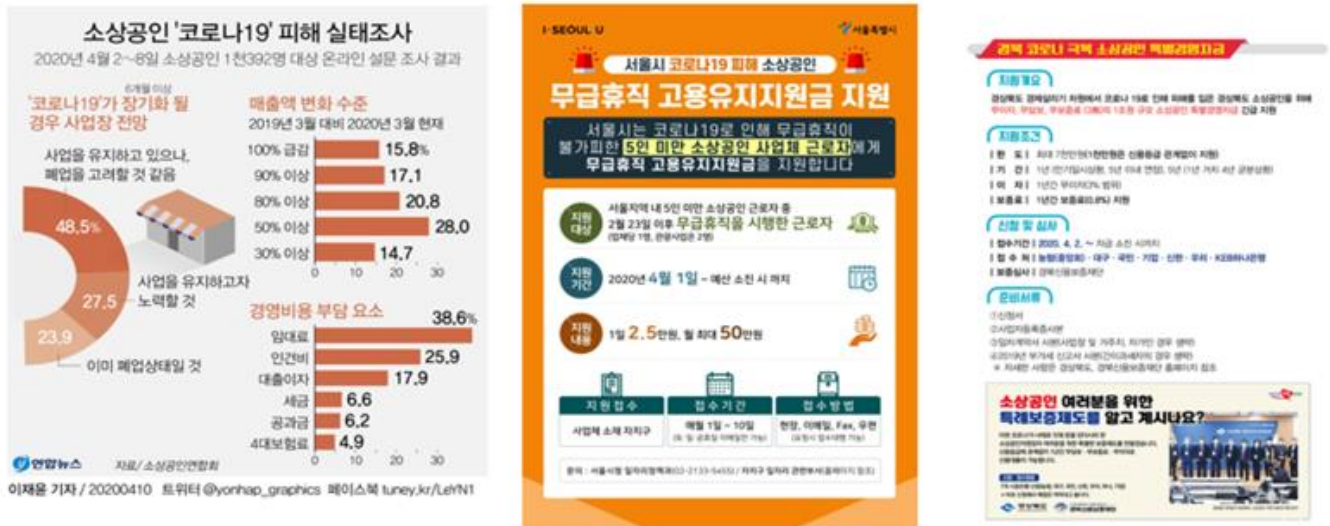
M-Five

강태환 김윤환 박대한 박준민 정규형

1. 주제

포스트코로나 소상공인 개업: 지표 개발을 통한 리스크 관리

2. 분석 배경



<그림1>¹ 소상공인 코로나 19 피해실태 조사 & 정부측 소상공인을 위한 특별 지원 기사자료

2020년 1월 코로나 바이러스 발생 이후 사회 경제 패턴이 급변했다. 이에 많은 기업들이 피해를 입고 있으며, 소상공인의 경우에는 그 정도가 더한 상황이다. 소상공인연합회 자료에 의하면, 코로나 이전에 비해 소상공인들의 매출액이 전반적으로 감소하였으며 절반 이상의 소상공인이 폐업을 고려하고 있다. 국가 및 지자체에서는 다양한 지원을 하고 있음에도 불구하고, 일시적인 지원에 그치는 경우가 많아 장기적인 해결책은 여전히 미지수로 남아 있다.

이에 데이터 분석을 통해 소상공인들의 어려움 해결에 도움이 되고자 한다. 코로나로 인해 변화된 데이터들을 분석하여 서울시 내 점포 영업 리스크를 시각적으로 알 수 있는 새로운 지표를 개발, 최종적으로 지역 및 업종 별 영업 전망을 판단하는 모델을 구축하는 것을 목표로 하였다.

¹ 출처: <https://www.yna.co.kr/view/GYH20200410001600044>, http://www.sbiz.or.kr/mb/mb/local_policy/1216597_1723.jsp, <http://www.gmlbo.net/news/article.html?no=54772%27TcvFyJ%3C%27%22%3EKNOLCo>

3. 데이터 소개

신청 내부데이터			
데이터 명	제공기관	활용여부	활용방안
업종거래특성	BC카드	X	2019년 7월-12월 자료로 자료 집계 시간 범위가 적어 활용하지 않음
업종별매출현황	BC카드	X	
지역별가맹점정보	신한카드	X	가용 정보가 동 별, 업종별 가맹점 수 이외에는 타 데이터와 겹치는 정보가 많아 활용하지 않음
지역별매출및이용고객정보	신한카드	O	서울시 관련 자료만 추출하여 활용 추후 외부데이터와 결합
(코로나19)인구데이터	KT	X	제공기관의 제공 불가 판정
요일별유동인구정보	SKT	O	서울시 동 별 유동인구 정보로 데이터 통합
성연령별유동인구정보	SKT	O	
시간단위별유동인구정보	SKT	O	

<표1> 신청 내부 데이터

반입 외부 데이터			
데이터 명	제공기관	활용여부	활용방안
서울시 우리마을가게 상권분석서비스(상권영역)	서울 열린데이터 광장	O	상권 데이터로 열병합 하여 활용
서울시 우리마을가게 상권분석서비스(상권-점포)		O	
서울시 우리마을가게 상권분석서비스(상권-추정매출)		O	
서울시 우리마을가게 상권분석서비스(상권-추정매출)_2019		O	
서울시 우리마을가게 상권분석서비스(자치구별 상권변화지표)		O	
서울시 주민등록인구 (동별) 통계		O	
일반음식점_인허가데이터	LOCALDATA	O	데이터 통합 과정에서 행정동, 법정동 매칭에 활용
행정동 법정동 코드	행정안전부	O	
서울시 유동인구 데이터	SKT bigdatahub	O	내부 데이터의 유동인구 정보 데이터와 병합

<표2> 반입 외부 데이터

신청한 내부데이터와 반입한 외부 데이터, 이에 대한 활용여부와 방안은 <표1>, <표2>와 같다. 여러 데이터를 활용하여 최종 데이터를 구축한 후 모델링 및 분석에 활용하였다.

4. 데이터 통합

분기	법정동	20대남자유동인구	20대여자유동인구	...	60대여자유동인구	00시유동인구	...	23시유동인구	월요일유동인구	...	일요일유동인구
2019년	법정동	성연령별 유동인구 정보 (내부데이터)				시간단위별유동인구정보 (내부데이터)			요일별유동인구정보 (내부데이터)		
2020년		2020년 유동인구 정보									

<표1> 통합된 인구 데이터

분기	동	업종코드	월요일매출	...	일요일매출	00시매출	...	23시매출	20대매출	...	60대매출	남성매출	여성매출	주중매출	주말매출	점포수	...	개업률	폐업률
2019년	법정동		지역별매출및이용고객정보 (내부데이터)																
2020년	행정동		상권데이터 (외부 데이터)																

<표2> 통합된 상권데이터

분기	동	업종코드	유동인구관련정보	매출정보	개폐업정보	그룹정보
2019년	행정동	업종코드	인구 데이터	소비자데이터 + 상권데이터		클러스터링 결과
2020년						

<표3> 통합된 최종 데이터

<표1>은 통합된 인구 데이터 개요이다. 내부데이터의 2019년 유동인구정보와 외부데이터의 2020년 인구정보를 통합하였다. 2019년 4월 - 2020년 6월 까지의 자료를 3개월 단위로 2-6분기로 나누어 총 467개의 법정동이 각각 분기별로 6개의 행을 갖는다. 2020년 유동인구 정보는 구별 정보만 제공하기 때문에 층화추출(stratified sampling) 기법을 이용해 동 별 인구로 분할한 데이터를 추가 생성하여 활용하였다.

<표2>는 통합된 상권데이터의 개요이다. 상권데이터(외부데이터)를 열병합 하여 하나의 데이터로 구축한 후 지역별매출이용고객정보(내부데이터)와 행병합하여 완성하였다. '분기'의 경우 <표1>의 진행방향과 동일하며, '동'의 경우 내부데이터의 경우 460여개의 법정동, 외부데이터의 경우 325개의 행정동으로 표기되어 있다. 업종코드는 업종중분류코드로 (동, 업종) 조합에 해당하는 행이 분기별로 6개의 행

<표3>은 최종 데이터의 개요이다. '분기'는 위와 동일한 처리방식을 사용했으며 '동'의 경우 법정동, 행정동의 1:1, 1:N, N:1의 관계를 고려해 행정동 기준 데이터프레임을 완성하였다. '업종 코드'의 경우 내부에서 제공하는 업종중분류코드를 기준으로 빈도수를 고려함과 동시에 소상공인 관련 코드 위주로 최종 업종코드 7개를 사용하였다. (한식, 중식, 일식, 카페 및 베이커리, 식료품, 유흥주점, 편의점) '그룹 정보'는 추후 모델링을 통해 (분기, 행정동, 업종코드) 조합을 13개의 그룹으로 클러스터링한 그룹 정보이다.

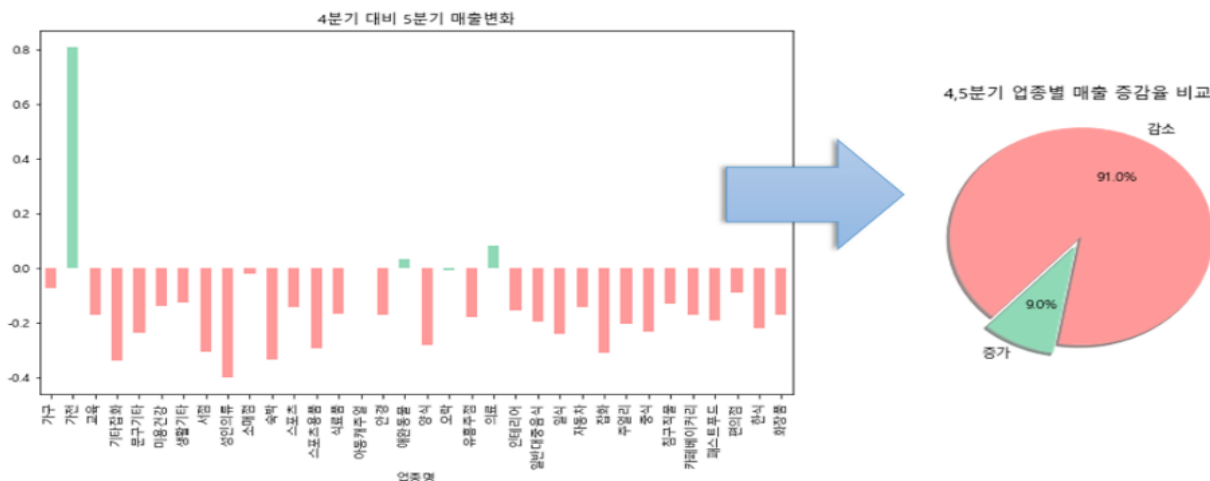
통합된 최종 데이터를 바탕으로 EDA를 진행하여 코로나 전후의 특징을 파악하고자 하였다.

The figure consists of three bar charts, each representing a different day of the week (Friday, Sunday, and Saturday). Each chart shows the number of flights (CNT) for five different categories (2, 3, 4, 5, 6). The y-axis represents the count, ranging from 0 to 100,000 for Friday and Sunday, and 0 to 80,000 for Saturday. The legend indicates that the bars represent FRI_FLTNG_PLTN_CNT, SUN_FLTNG_PLTN_CNT, and SATR_FLTNG_PLTN_CNT.

Category	FRI_FLTNG_PLTN_CNT	SUN_FLTNG_PLTN_CNT	SATR_FLTNG_PLTN_CNT
2	~98,000	~65,000	~81,000
3	~86,000	~60,000	~71,000
4	~91,000	~65,000	~82,000
5	~36,000	~36,000	~34,000
6	~37,000	~33,000	~34,000

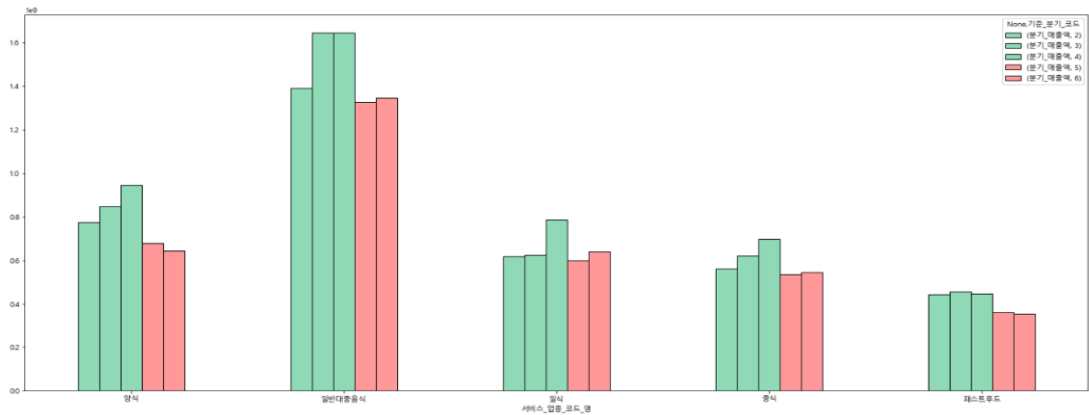
코로나 유행 이전인 4분기까지와 유행 이후 분기 간 극명한 차이가 존재한다. 2-4분기에 비해 5-6분기에서 유동인구가 급감하였다. 해당 요일뿐만 아니라 요일별, 성별, 연령별, 시간대별 유동인구가 모든 행정동에서 같은 패턴을 보였다.

2) 업종별 정보



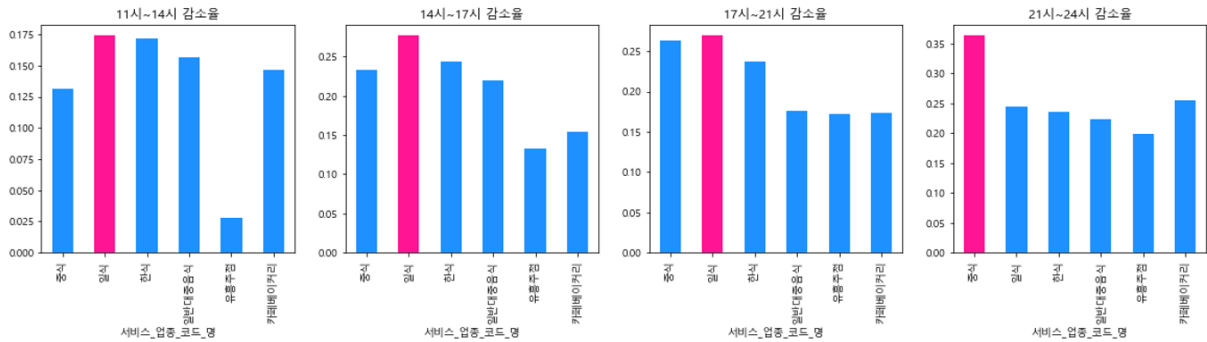
<그림3> 4분기 대비 5분기 매출변화 & 4.5분기 업종별 매출 증감률 비교

인구정보의 동향과 동일하게, 가전과 의료를 제외한 모든 업종에서 매출이 감소하였다. 가전의 경우 코로나의 영향으로 집에 머무는 시간이 증가하고 보복소비가 늘어 매출이 증가한 것으로 보이며, 의료의 경우 바이러스 유행의 직접적인 영향으로 인한 매출 증가로 보인다. 하지만 이를 제외한 업종에서 감소를 보였다는 결과는 소상공인의 영업 어려움을 보여준다.

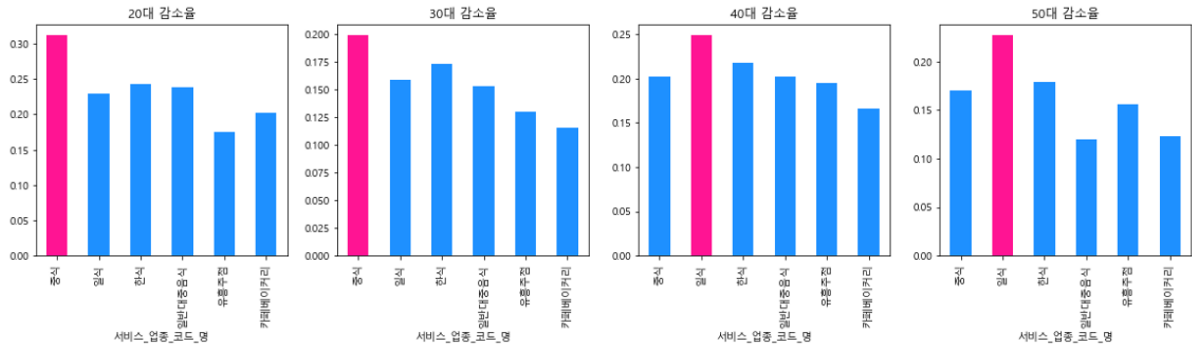


<그림4> 2-6분기 요식업계별 매출액 (코로나 후 분기 붉은색 표기)

<그림4>에서도 알 수 있듯이 요식업계에서 코로나 이후 대체적으로 매출은 크게 감소했으며 일반대용음식에서 더욱 두드러지게 나타난다.



<그림5> 시간대별 요식업계 매출 감소율



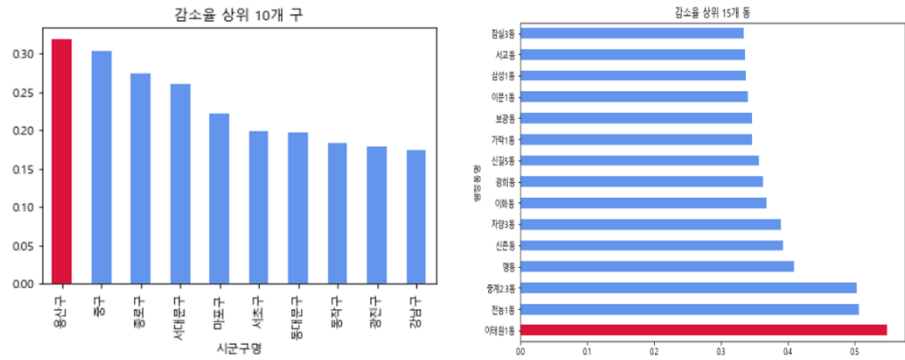
<그림6> 연령대별 요식업계 매출 감소율

<그림5>를 통해 21시 이전까지 일식의 매출 감소가 가장 크게 나타나고 있음을 알 수 있다. 21시 이후에는 중식 매출이 약 35%로 크게 감소하였다. 유흥주점은 저녁 시간대 이후 시간 감소율이 크게 나타나고 있다.

<그림6>에서도 알 수 있듯이, 주 소비를 이루는 연령별 요식업의 매출 감소율은 2-30대와 4-50대가 비슷한 양상을 보였으며, 2-30대의 경우 중식에서, 4-50대에서는 일식에서 매출액이 가장 많이 감소하

였다.

3) 지역별 정보



<그림7> 매출 감소율 상위 10개 구 & 매출 감소율 상위 15개 동

대부분의 지역에서 매출 감소를 보였다. 그 중에서 코로나의 지속적 영향이 큰 지역을 파악하고자 하였다. 4분기 대비 6분기 감소율을 도출하여 기준 상위 10개구, 15개동의 현황을 파악, 용산구와 이태원1동이 상위 구와 동으로 나타났다.

6. 모델링

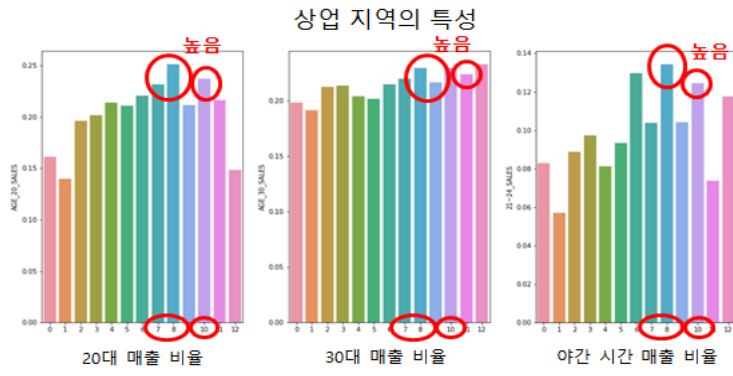
초기 모델은 타겟변수를 선정하여 지도 학습 모델을 구축, Feature Importance를 통해 적절한 지표를 개발하고자 하였다. 타겟변수로는 폐업률, 해당 분기 매출액, 이전 분기 매출액, 전 분기 대비 매출액 증감률 등을 고려하였다. 하지만 폐업률의 경우 중앙값이 0일 정도로 right-skewed되어 있다는 점, 경제의 후행 지표로서 5분기, 6분기에 대한 시장 정보가 반영이 안 되어 있다는 점, 매출 관련 변수는 재료비, 운영비, 임대료 등을 고려한 순이익 정보가 배제되어 있다는 점 등이 타겟 변수로서의 적절성이 떨어진다고 판단하였다.

따라서 (행정동, 업종) 조합을 그룹화 하기 위한 비지도 학습의 Clustering을 진행하였다. K-Means와 DBSCAN, Hierarchical clustering 중 작동 원리와 성능 등을 고려하여 최종적으로 K-Means 모델을 사용하였다. 코로나19 유행 전, 후의 특성을 파악하기 위해 코로나 유행 전을 1번 그룹부터 13번 그룹까지, 코로나 시대 후를 a번 그룹부터 m그룹까지 총 26개의 군집으로 나누었다.

7. 분석결과

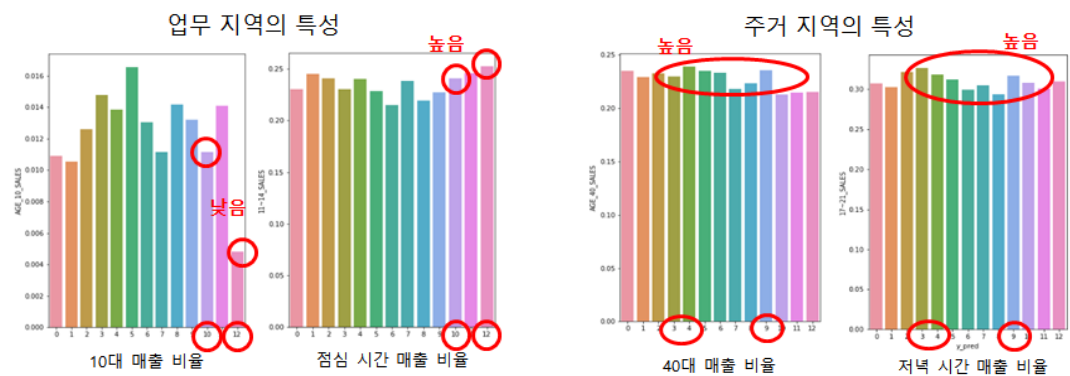
(1) 군집화 결과

Clustering 그룹의 특성을 지역별, 업종별로 살펴보았다. Clustering 결과 파악을 위해 행정동을 크게 주거지역, 상업지역, 업무지역으로 나눈 뒤, 같은 특성의 동들이 비슷한 군집으로 묶였는지 살펴보았다.



<그림8> Clustering 결과에 따른 상업지역의 특성

신촌동, 서교동(홍대)과 같은 상업 지역들은 코로나 전에는 7, 8, 10번 그룹에, 코로나 이후에는 i, j, k 그룹에 속했다. <그림8>을 통해 해당 그룹들은 공통적으로 20대, 30대의 매출 비율이 높으며, 야간 시간의 매출 비율이 높은 특성을 가지고 있는 것을 확인할 수 있었다.



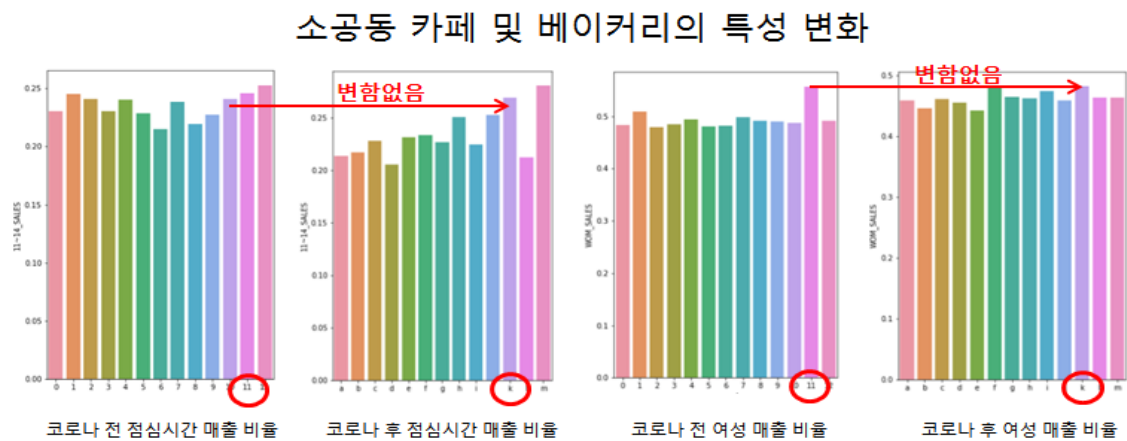
<그림9> Clustering 결과에 따른 업무지역과 주거지역의 특성

소공동, 삼성동 등의 업무 지역들은 코로나 전에는 주로 10, 11, 12번 그룹에, 코로나 이후에는 h, m 그룹에 속했다. 해당 그룹은 공통적으로 10대의 매출 비율이 낮으며, 점심 시간의 매출 비율이 높은 특성을 가지고 있는 것을 확인할 수 있었다.

옥수동, 아현동 등의 주거 지역들은 코로나 전에는 주로 3, 4, 9번 그룹에, 코로나 이후에는 e, g그룹에 속했다. 해당 그룹은 공통적으로 40대의 매출 비율이 높으며, 저녁 시간의 매출 비율이 높은 특성을 가지고 있는 것을 확인할 수 있었다. 이는 가족 단위의 저녁 외출과 가족의 가장이 지불하는 형태로 예상된다.

(2) 코로나 전후 패턴 변화

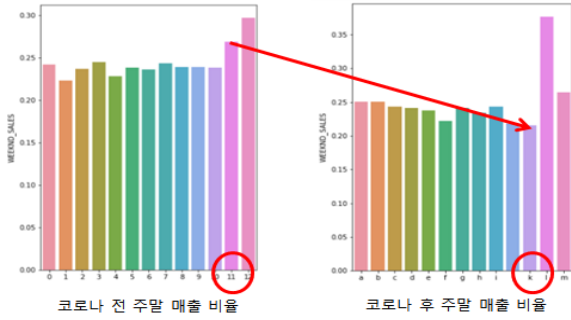
Clustering 결과 도출 후, 코로나19 유행 전후의 지역별, 업종별 특성 변화가 있는지 살펴보았다.



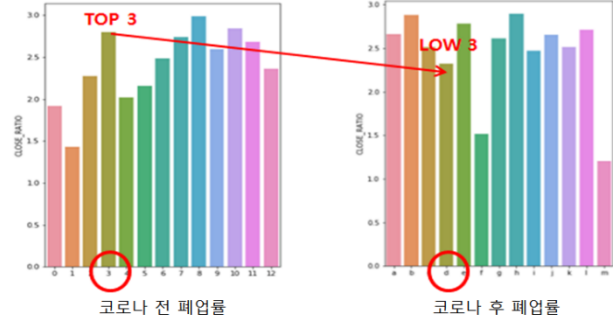
<그림10> 소공동 카페 및 베이커리의 특성 변화

소공동(업무지역) 카페 및 베이커리의 경우, 코로나 전 11번 그룹, 코로나 후 k그룹에 속했다. 11번 그룹과 k그룹 모두 11~14시 매출 비율이 높았고, 여성 매출 비율이 높다는 공통된 특징이 있다. 코로나 시대 전후 특성 변화가 없는 것을 확인할 수 있다.

신촌동, 서교동 한식의 특성 변화



옥수동 유흥주점의 특성 변화



<그림11> 신촌동, 서교동 한식의 특성 변화/ 옥수동 유흥주점의 특성 변화

신촌동, 서교동(상업지역) 한식의 경우, 코로나 전 11번 그룹, 코로나 후 k그룹에 속했다. 11번 그룹은 주말의 매출 비율이 높은 특징을 갖지만, k그룹은 주말 매출 비율이 낮은 것을 확인할 수 있다.

옥수동(주거지역)에서 유흥 주점의 경우 코로나 전 3번 그룹, 코로나 후 d그룹에 속했다. 3번 그룹의 폐업률은 높은 편에 속하는 반면, d그룹은 폐업률에서 뚜렷한 특징을 보이지 않았다.

위와 같이 해당지역과 업종의 코로나 전후 그룹변화와 특성을 도출하여, 코로나 전후 사회경제적 패턴을 파악할 수 있다.

8. 활용방안

클러스터링 결과를 바탕으로 동, 업종 별로 영업 위험 점수를 도출하였다. 코로나 이후의 달라진 패턴을 적용하여 위험 점수를 생성하고, 이를 통해 새로운 창업자나 기존 영업자가 쉽게 활용할 수 있도록 가시화하였다.

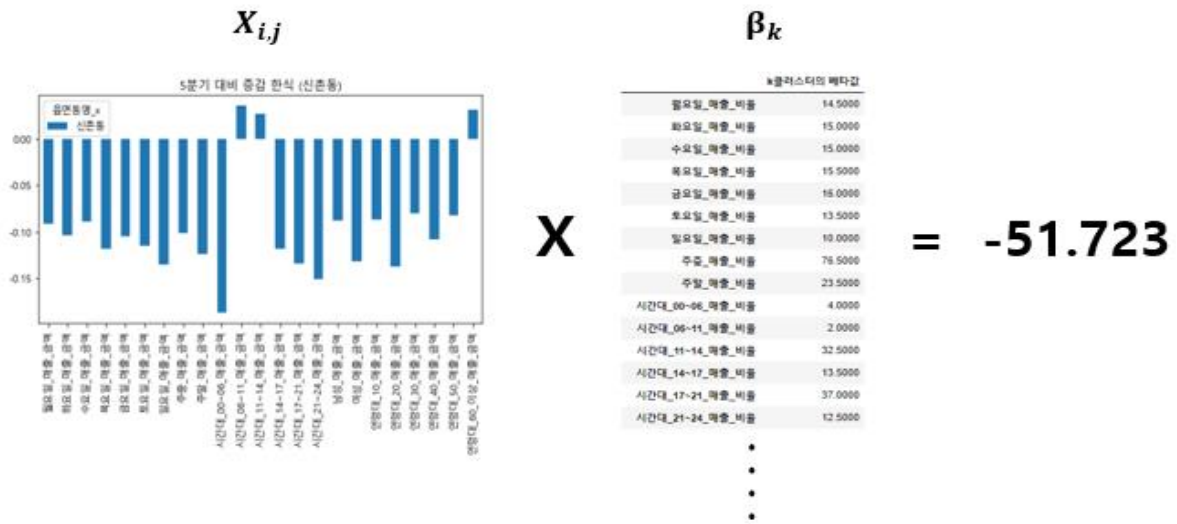
$$\text{위험점수} = \sum X_{i,j} \beta_k$$

$$i = \text{신촌동, 소공동} \dots \text{옥수동} \quad j = \text{한식, 중식, \dots, 편의점} \quad k = a, \dots, m$$

$$X_{i,j} : i \text{ 동, } j \text{ 업종의 5 분기 대비 6 분기 증감률}$$

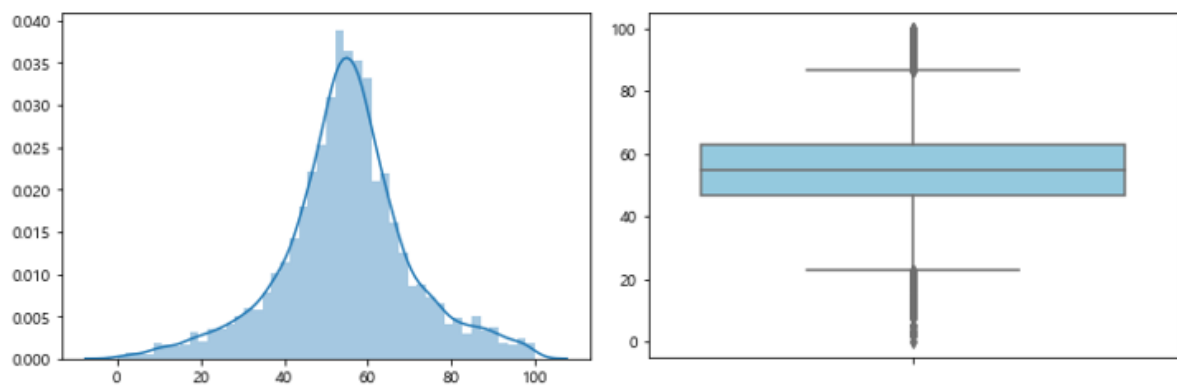
$$\beta_k : k \text{ 그룹에 속한 데이터의 가중치 값,}$$

(해당 변수의 값이 가지는 비중으로 이루어진 벡터)



<그림 12> 신촌동 한식업종의 위험점수 산출 과정

<그림 12>와 같이 (신촌동, 한식)에 해당하는 각 변수들의 증감률(X)을 산출한 후, 해당하는 Cluster 의 변수들의 가중치(β) ((ex)월요일 매출 비율은 전체 요일에서 14%를 차지)를 이용해 최종 위험 점수를 도출하였다.



<그림 12> 최종점수의 분포

위와 같이 도출된 최종점수를 0-100 점 범위로 스케일링한 후 <그림 12>와 같은 분포를 얻을 수 있었다. 행정동과 업종별로 위험 점수를 산출하여 영업 위험정도를 파악하고, 소상공인이 포스트 코로나 시대를 대비하여 영업 지속 여부, 창업 여부를 결정하는데 도움을 줄 수 있을 것이다

9. 결론

데이터 안심구역에서 제공하는 내부 데이터와 외부 공공 데이터를 활용하여 코로나 전후 패턴변화를 파악하였다. Clustering 기법을 이용하여 행정동, 업종별로 그룹화한 뒤 각 그룹의 특성, 코로나 전후의 그룹변화 파악을 통해 코로나로 인한 변화 양상을 확인할 수 있었다. 더불어, 변화를 고려한 지표 개발을 통해 최종적으로 소상공인의 개업 및 영업 리스크 관리에 도움을 줄 수 있는 방안을 모색하였다.