

코로나-19 전후에 따른 서울시 유동인구, 카드소비 데이터 관계분석

나형선(HyungSun Na)*, 김진우(JinWoo Kim)*, 안진현(Jinhyun Ahn)**, 전대성(Daesung Jun)***, 임동혁(Dong-Hyuk Im)****

*광운대학교 인공지능융합학과

*e-mail : nayosk@kw.ac.kr

**제주대학교 경영정보학과

**e-mail : jha@jejunu.ac.kr

***전주대학교 행정학과

***e-mail : daesungj@jj.ac.kr

****광운대학교 정보융합학부

****e-mail : dhim@kw.ac.kr

Analyzing Relationships Between Floating Population and Card Consuming Data in Seoul Before and After COVID-19

HyungSun Na*, JinWoo Kim *, Jinhyun Ahn**, Daesung Jun***, Dong-Hyuk Im****

*Dept. of Applied Artificial Intelligence, Kwang-Woon University

**Dept. of Management Information Systems, Jeju National University

***Dept. of Public Administration, Junju University

****School of Information Convergence, Kwang-Woon University

요 약

COVID-19 가 장기간 지속됨에 따라 시민들의 생활패턴, 생계 등에 큰 영향을 미치고 있다. 본 논문에서는 서울시의 유동인구 및 카드 소비데이터를 이용하여 COVID-19 가 어떤 영향을 미쳤는지 알아보려 한다. 이를 분석하기 위하여 대용량 데이터인 2019 년 1 월 ~6 월 그리고 2020 년 1 월 ~ 6 월 서울시 유동인구 및 카드 소비 데이터를 이용하였다. 서울시를 각 행정구로 나누어 이에 관련한 유동인구의 변화, 유동인구의 증감의 변화를 추정하고 마찬가지로 카드소비데이터의 증감의 변화를 추정하여 서울시 여러 행정구의 유동인구, 카드데이터 두 데이터 간의 연관 관계를 다방면으로 분석하여 엄밀한 인구 밀집도 분석으로 COVID-19 가 서울 지역경제에 미친 영향에 대하여 실증분석을 진행하였다.

1. 서론

2003 년 사스, 2015 년 메르스 이후 2019 년 신종 코로나 바이러스가 전 세계로 확산 되면서 감염자가 늘어나고 있다. 앞서 언급된 다른 전염병에 비해 치사율은 낮은 편이지만 전염성이 더 강해 2021 년 3 월 현재 전세계적으로 약 1 억 2 천명의 누적 확진자와 약 270 만명의 사망자가 발생하였다. 이러한 사태가 사스, 메르스 때와는 달리 장기간 지속됨에 따라 사회적으로는 생활 패턴부터 크게는 생계에 직결되는 산업까지 시민들의 생활 전반에 영향을 미치고 있다. 앞으로도 제 2 의 코로나 바이러스와 같은 전염병의 등장을 간과할 수 없게 되면서 이로 인해 벌어지는 시민들의 경제적 생존에 대응하기 위한 정책들이 필

요하다고 판단되고 있다.

본 논문에서는 이러한 상황에서 코로나 바이러스 확산이 지역경제에 미친 영향에 대한 실증분석을 위해 서울시 유동인구 데이터 분석 결과를 토대로 카드 소비 데이터와의 관계 분석을 진행하였다. 비슷한 연구로 [1] 에서는 유입인구에 대한 분석이 이루어졌으나, 본 논문에서는 유동인구와 카드소비 데이터에 대한 관계분석이 이루어졌다는 차이점을 가진다. 또한 [2] 에서는 서울시 외국인 밀집지역에 대한 집중적인 분석을 진행하였고, 본 논문에서는 서울시 전체의 유동인구를 분석했다는 차이점을 두고 있다. [3] 에서는 수원시 카드소비 데이터에 대한 분석이 이루어졌으나 본 논문에서는 서울시 카드데이터와 유동인구 데이터에 대한 관계분석이 이루어졌다는 차이점을 가진다.

본 연구를 통해 코로나 바이러스 확산에 취약한 업종과 그렇지 않은 업종에 대한 구분을 도출해 낼 수 있고, 그에 따른 정책의 우선순위를 정하는데 도움이 될 수 있을 것이다. 뿐만 아니라 지역경제활성화를 위한 정부의 긴급재난지원금 등의 효과와 긴급재난지원금 효과에 영향을 미치는 요인들을 도출함으로써 전염병 확산에 따른 지역경제 활성화를 위한 효율적인 지방정부 역할에 대한 시사점을 제시할 수 있을 것이다.

2. 데이터 전처리

본 논문에서 분석을 위해 2 가지 유형의 데이터를 사용하였다. 첫번째는 서울시 50 셀별 유동인구 데이터, 두번째는 서울시 집계구별 카드소비 데이터이다. 코로나 전후 비교를 위해 두 데이터 모두 2019 년 1~6 월, 2020 년 1~6 월의 데이터를 사용하였다.¹

여기서 유동인구 데이터란 주 체류지가 아닌 곳에서 이동중인 인구, 해당 지역에서 일시적으로 체류하거나 이동 중인 인구를 나타낸다.

2.1 데이터 클리닝

카드데이터를 월별로 업종명을 기준으로 통합하여 분석한 결과 몇몇 업종에 결측치가 존재하였다. 본 논문에서는 2019 년과 2020 년이 비교가 되어야 하기 때문에 결측치가 존재하는 업종명들은 삭제하였다.

2.2 데이터 축소

카드데이터의 경우 칼럼에 성별, 연령대가 존재하면서 한 줄에 표현될 수 있는 데이터들이 여러 개의 레코드로 분리되어 있었다. 효율적인 데이터 분석을 위해 성별과 연령대를 묶어 각각 칼럼으로 추가하여 데이터의 크기를 감소시켰다.

2.3 데이터 통합

유동인구 데이터의 경우 10 일 단위로 파일이 저장되어 있어 카드데이터와 통일하기 위해 한달 단위로 파일을 통합하였다.

2.4 데이터 변환

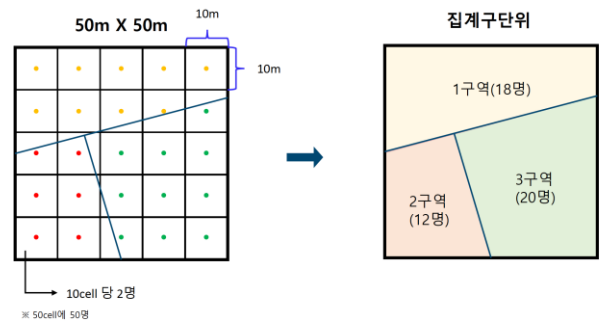
2.4.1 단위 면적 통일

유동인구 데이터와 카드데이터의 관계 분석을 위해서는 데이터들의 단위 면적을 통일시킬 필요가 있다. 유동인구, 유입인구의 단위는 50cell 단위이며 카드소비 데이터는 집계구 단위를 사용하기 때문에 두 데이터의 단위 면적을 집계구 단위로 통일하는 매핑 작업을 진행하였다.



(그림 1) 집계구단위 50cell 단위

그림 1 은 소비데이터의 집계구단위와 유동인구 데이터의 50cell 단위를 시각화 한 것이다. 그림 1 을 참고하면, 50cell 단위의 경우 집계구에 비해 큰 단위이고 집계구의 크기가 다양함으로 오차범위가 생기기 때문에 50cell 단위를 10cell 로 분할하는 작업을 진행하였다.



(그림 2) 50cell 집계구단위로 매핑 예시

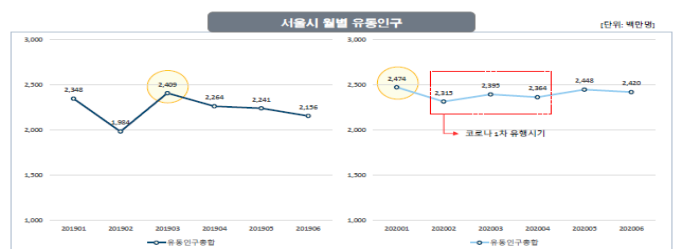
그림 2 는 50cell 단위를 10cell 단위로 분할한후 집계구 단위에 매핑하는 작업을 시각화 한 것이다. 그림 2 처럼 50cell 을 25 등분하여 10cell 로 분할한 후 유동인구 수를 25 로 나눈 후, 10cell 의 중심좌표를 해당 좌표가 포함되는 집계구에 넣어주어 50cell 좌표를 집계구 단위로 변환하였다.

2.4.2 시간대 통일

카드 데이터의 경우 시간대가 새벽, 오전, 점심, 오후, 저녁, 심야 총 6 단위로 이루어져 있고 유동인구 데이터는 한시간 단위로 이루어져 있기 때문에 더 큰 범위의 카드데이터에 유동인구의 시간대를 맞출 필요가 있어 시간대별로 통합하는 과정을 진행하였다.

3. 분석결과

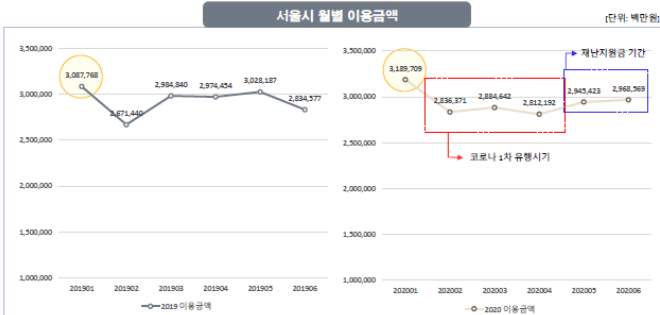
3.1 기초 통계분석



(그림 3) 서울시 월별 유동인구 데이터

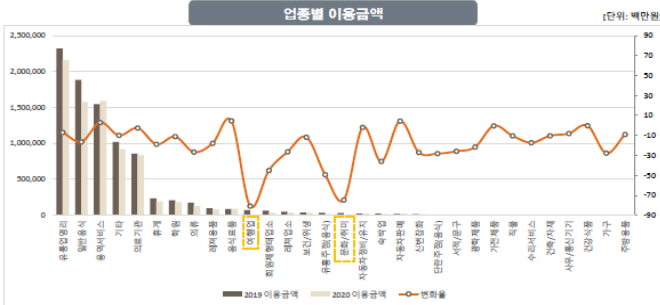
¹ 유동인구 및 카드 소비 데이터는 2020 년 하반기 AI 데이터 가공 바우처 지원 사업에 의해 제공됨.

그림 3 의 서울시 월별 유동인구 데이터를 보면 코로나 1 차 유행시기인 2020 년도 2, 3, 4 월이 2019 년도 2, 3, 4 월에 비해 오히려 조금 증가하는 추세 보이며 서울시 내에서의 유동인구 데이터는 코로나의 영향을 거의 받지 않는 것으로 보인다.



(그림 4) 서울시 월별 카드 소비 데이터

그림 4 의 서울시 월별 카드 소비 데이터를 보면 코로나 1 차 유행시기인 2, 3, 4 월에 코로나의 영향으로 소비가 줄어든 것으로 판단되고 1 차 재난지원금의 영향으로 5, 6 월에 카드 소비데이터가 증가하였고 6 월의 경우는 전년도와 비교해서도 이용금액이 증가하였다.

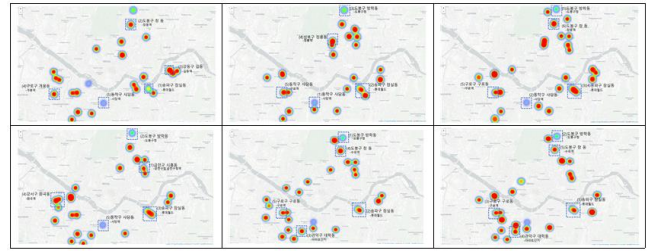


(그림 5) 서울시 월별 업종당 카드 소비 데이터

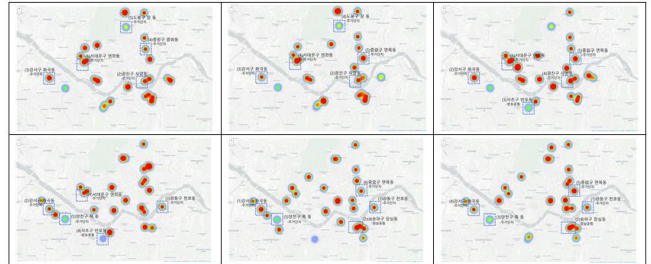
그림 5 의 서울시 월별 업종당 카드 소비 데이터를 보면 여행업과 문화/취미 업종이 코로나의 영향으로 크게 감소한 것을 볼 수 있다.

3.2 유동인구와 코로나 확진자수 관계 분석

2019 년 유동인구 데이터와 2020 년의 유동인구 데이터의 차이점을 확인하기 위해 집계구별로 2019 년 대비 유동인구가 증가(그림 6), 감소한(그림 7) 상위 집계구 30 개를 히트맵을 통해 분석하였다. 집계구가 많이 모여있는 지역일수록 히트맵의 붉은색 강도가 강하게 나타나며 집계구가 적은 지역은 옅은 색이 나타난다.

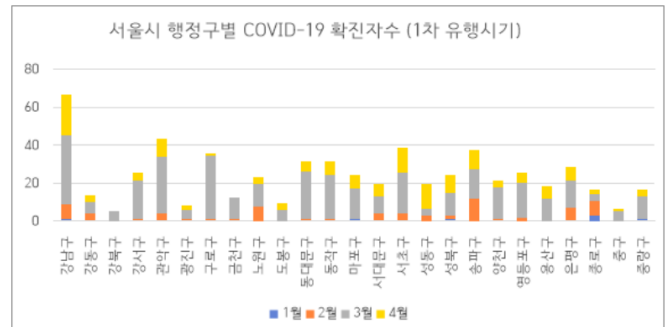


(그림 6) 서울시 유동인구 감소 상위 집계구 히트맵



(그림 7) 서울시 유동인구 증가 상위 집계구 히트맵

분석 결과 유동인구 증가는 송파구, 강서구, 관악구, 강남구 등의 한강 이남에서 2019 년 대비 증가한 것으로 나타나고, 유동인구 감소는 한강 이북 지역에서 나타났다. 이러한 결과를 보면 사회적 거리두기에 따른 영향은 강북지역에 크게 나타난 것으로 보인다.

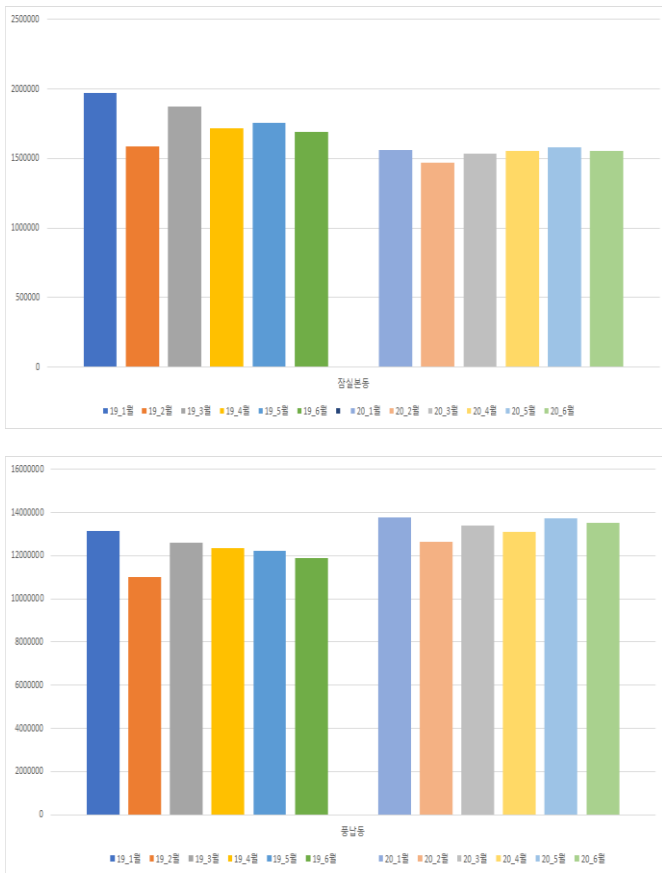


(그림 8) COVID-19 1 차 유행시기 행정구별 확진자수

그림 8 의 COVID-19 1 차 유행시기인 2020 년도 1~4 월의 확진자 수를 분석하면 유동인구가 증가하는 강남 지역의 확진자 수가 더 많게 나타나는 것을 볼 수 있다.

본 연구에서는 유동인구와 카드 소비 데이터의 연관성을 찾고자 유동 인구의 변화가 가장 크면서 1 차 대유행시기에 확진자 수가 4 번째로 많았던 송파구의 유동 인구와 카드 소비 데이터를 관계 분석하였다.

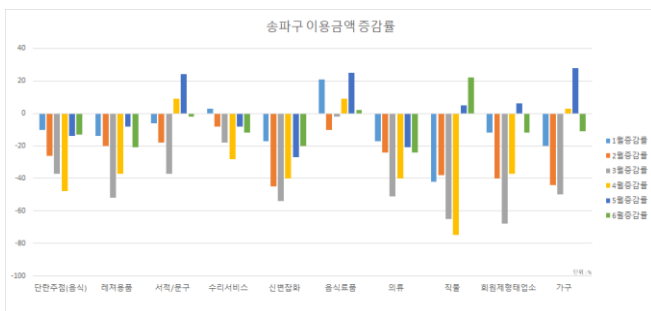
3.3 유동인구와 카드소비 데이터 관계분석



(그림 9) 송파구 유동인구 분석결과

그림 9의 송파구 유동인구 데이터 분석결과로 좌측 그래프는 송파구의 유동인구가 감소하는 지역 잠실동(본동), 오류동, 거여동 중 잠실동(본동)이며 반대로 유동인구가 증가하는 우측 그래프 지역은 풍납동, 오금동, 장지동 중 풍납동이다. 이러한 분석 결과 토대로 송파구의 카드 소비 데이터의 가맹점 위치를 구체적으로 분석한 결과 마찬가지로 잠실동의 카드 이용 건수는 눈에 띄게 감소하였으며, 풍납동의 경우는 이용건수가 늘어났다.

3.4 긴급재난 지원금 효과 분석



(그림 10) 송파구 카드 데이터 이용금액 증감률

그림 10은 긴급재난 지원금 이전과 이후로 분석하여 긴급재난 지원금의 효과가 나타난 것으로 보이는 업종이다. 음식점, 레저용품, 서적/문구, 수리서비스, 신변잡화, 의류, 직물 등이 재난 지원금의 효과를 본 것으로 보인다. 마지막으로 숙박업, 유흥주점, 자동차 정비/유지, 학원, 휴게의 경우는 재난지원금의 효과를 받지 못한 것으로 분석되었다.

4. 결론

빅데이터 모델링을 활용하여 COVID19 감염병 확산이 지역경제에 미친 영향에 대한 실증분석을 통해 유동인구 데이터 기반 카드 소비 매출액 분석을 통해 지방정부의 지역경제 활성화 방안에 대한 정책적 시사점을 제시해 줄 수 있을 것이다. 뿐만 아니라 유동인구 데이터 및 소비 카드 매출액 분석을 활용한 보다 엄밀한 인구 밀집도 분석을 통해 사전적으로 확산 위험 밀집도가 높은 지역과 업종을 찾아냄으로써 선제적으로 COVID19 n 차 감염 확산 방지에 대한 노력을 통해 지역경제에 미치는 부정적 영향을 최소화하기 위한 시사점을 제시해 줄 수 있을 것이다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2021-2018-0-01417). 또한 한국 데이터산업진흥원의 2020년 하반기 AI 데이터 가공 바우처 사업의 지원을 받아 수행된 연구임.

참고문헌

- [1] 진주혜, 성병찬. (2020). 코로나-19에 따른 서울시 생활인구 변화와 동별 반응 차이 분석. 응용통계연구, 33(6), 697-712.
- [2] 허정원, 장주영. (2020). 코로나 19 확산시기 서울시 외국인 밀집지역의 지역특성과 생활인구 변화. 공간과 사회, 73(0), 99-137.
- [3] 이성호, 최석환. (2020). 신용카드 빅데이터를 활용한 코로나 19 발생의 지역상권 영향 분석 수원시를 사례로. 공간과 사회, 73(0), 167-208.