

서울시 유사지역 내 소비행태 비교분석 및 미비 업종 진단

최다운¹, 심현지², 조수환³

염정⁴, 김태원⁵, 조민주⁶, 김진⁷

¹상명대학교 경영학부 학부생

²상명대학교 지능·데이터융합학부 학부생

³상명대학교 융합전자공학과 학부생

⁴서울여자대학교 언론영상학부 학부생

⁵세종대학교 데이터사이언스학과 학부생

⁶상명대학교 경영학부 학부생

⁷상명대학교 빅데이터융합전공 교수

choid0717@naver.com, tlaguswl21@naver.com, shan6517@naver.com,

ybt642537@gmail.com, 001002k@gmail.com, mocarain113@gmail.com, jinkim@smu.ac.kr

Consumption behavior in similar areas in Seoul Comparative analysis and diagnosis of deficient industries

Daun Choi¹, Hyunji Shim², Suhwan Jo³

Jung Yum⁴, Taewon Kim⁵, Minjoo Cho⁶, Jin Kim⁷

¹Faculty of Business Administration, Sangmyung University

²Faculty of Artificial Intelligence and Data Engineering, Sangmyung University

³ Dept. of Convergence Electronic Engineering, Sangmyung University

⁴Dept of Journalism and Media, Seoul Women's University

⁵Data Science Department, Sejong University

⁶Faculty of Business Administration, Sangmyung University

⁷Big Data Convergence Major, Sangmyung University

요 약

본 연구는 최근 서울시의 모든 지역이 과밀억제권역에 해당되는 문제를 해결하는 방법을 분석하였다. 서울시 행정동별 소득수준, 사업체 통계, 거주인구 데이터, 매출액, 대중교통 승차 승객 수 등의 변수를 활용하여 2번의 클러스터링을 거쳐 진행하였다. 2번에 걸쳐 클러스터링 된 군집별 특징을 바탕으로 서울시 내 미비 업종 진단을 제안한다.

1. 서론

최근 업종 과밀 문제가 대두되고 있다. 특히 서울시의 경우 모든 지역이 과밀억제권역에 해당한다. 더 이상 규제만이 해결책이 될 수 없다고 생각했기에 서울시 내에 신규/업종 변경 창업자에게 새로운 대안을 제시하는 방안을 생각해 보았다. [1]본 연구에서는 현재 제공하고 있는 서비스나 데이터들은 단순한 집계데이터나 지역별 포화지수만을 보여줄 뿐 업종별 상대적 비율까지는 보여주지 않는다는 한계점을 확인하였다. 이에 따라 정확한 상권특징 파악을 통해 특정 업종의 과밀도화를 방지할 뿐만 아니라 상대적으로 부족한 업종을 추천해 줄 수 있을 것으로 예상된다.

2. 분석 과정

2-1. 변수 선택

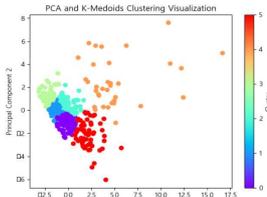
서울시 내국인 신한카드 소비 데이터, 서울시 행정동별 소득 수준 통계 등의 데이터를 421개의 행정동 기준으로 맞추어 전처리 하였다. 전처리 후 분야별 데이터로 구분하고 비율 및 총계 처리하여 변수로 사용하였다.

2-2. 11개 요인 선택 병합

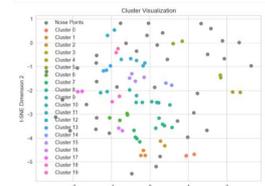
본격적인 분석 전 [2] 차원 축소를 통해 상관관계를 시각화하였으며, 산업 관련 요인과 취약, 인구 관련 요인이 가장 영향력이 높은 변수임을 알게 되었다. 본 연구에서는 수집한 데이터 중 산업과 인구 관련 데이터를 찾았고, 추세 파악을 위한 총합 값을

변수로 사용하였다.

3. 모델링



(그림1) 1차 클러스터링



(그림2) 2차 클러스터링

3-1. 1차 클러스터링

1차 군집화의 경우 K-means, GMM(Gaussian mixture model) 방법 모두 시도해 보았다. gmm의 경우, 시각화하였을 때 군집 별로 겹치는 부분이 많았고, 실루엣 점수가 낮아 최종 기각하였다. 최종 선택한 k-means의 경우 엘보우 포인트와 실루엣값을 모두 고려하여 실루엣 계수가 높고, 경사가 급변하는 지점인 6을 최적의 클러스터 수로 도출하였다. 이후 각 클러스터의 중심값에 가장 가까운 대푯값 5개 지역을 산출하였으며, 군집 중심 해석을 통해 1차 군집 특성을 도출하였다(그림1). cluster 0은 가장 많은 인구가 거주하는 상업지구 단지, 1은 유동성이 높은 핫플레이스, 2는 지역 매출액이 가장 낮은 군집, 3은 평균 소득 수준, 사업체 수 모두 높은 상업지구, 4는 소규모 지역 특성 업무지구, 5는 비교적 최근 개발된 주거지역 상권으로 판단하였다.

3-2. 2차 클러스터링

1차 클러스터링에서 사용한 총합 변수를 제외한 해당 세부 변수 59개(연령대별, 업종별, 시간대별 등)를 선정하여 2차 클러스터링을 진행하였다(그림2). 2차 클러스터링을 진행하기 위해 서울시 공공데이터의 상권변화 지표와 군집 특성을 참고하여 6개의 군집 중 한 군집으로 선택하였고, 가장 낮은 매출액을 보이지만 지역주민들의 소득 수준은 다른 군집 평균과 비슷한 2 군집을 활성화 가능성이 크다고 판단하여 선택하였다.

t-sne 모델을 사용하여 차원 축소를 진행하였다. 혼란도와 최대반복 회수, KL divergence 함께 고려하여 파라미터 결정하였다. 이후, 노이즈 값이 확인 할 수 있고 core point끼리의 세부 군집에 용이한 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)을 활용하여 2차 군집화를 진행하였고, EPS(epsilon)를 시각화하여 최적의 EPS를 확인함과 동시에, minimum sample 수의 경우 최대한 작게 잡아 가까운 것들만 군집되도록 분석을

진행하였다. 그 결과 총 20개의 군집을 도출할 수 있었다. 최종적으로 도출한 2차 클러스터링 결과에 신한카드 소비데이터를 활용해 군집 별 소비패턴을 확인하였고, 이를 통해 같은 군집 내 소비패턴 확인 및 과일, 미비 업종을 판단할 수 있었다.

4. 결론 및 정책 제언

2번의 클러스터링을 통해 유사지역을 분류하였고, 이를 통해 지역별 미비 및 과일 업종을 도출하였다. 이를 바탕으로 한 지역 맞춤 정책을 제안한다.

4-1. 창업 업종 추천 서비스 제공

창업을 원하는 지역과 업종을 선택하면 유사지역 대비 점포가 얼마나 밀집되어 있는지 보여주며, 해당 지역에 미비 업종과 과일 업종 상위 3개를 각각 보여준다. 또한, 시각화한 군집 유형별 소비 특징 자료들을 제공하여, 수치화된 데이터를 이해하기 어려운 초보 소상공인에게 직관적인 서비스 제공한다. 아래 그림과 같이(그림3), 2차 클러스터링 중 0 군집의 경우 두동의 매출이 비슷하지만, 삼청동의 매출이 비교적 높고, 같은 군집인 가회동을 활성화하기 위해서는 커피전문점과 패스트푸드점 매출을 강화해야 함을 도출하였다. 추가적으로, Xgboost 모델을 활용하여 업종만 선택하면 소비패턴 기반 과일지역을 선정해 알려주는 서비스도 제공할 수 있다.



(그림3) 클러스터 0 결과 분석

4-2. 한계 및 추후 연구

수집한 데이터별 기간이 상이하야 2015~2020년 데이터들을 추출하여 분석을 진행하였다는 한계점이 존재한다. 그러나 더욱 다양한 소비데이터 수집이 가능하다면 더욱 유의미한 결과가 나올 것으로 예상된다.

참고문헌

[1] 이승재 외 1인, 빅데이터 분석기법을 활용한 서울시 취약상권 지역 개선방안 연구, 서울디지털재단, 72-75, 2020.12.31.
 [2] 이승주, 주성분 분석을 이용한 빅데이터 분석, 한국지능시스템학회 논문지, 2-3, 2015.12
 1) 이 논문은 2023년도 과학기술정보통신부의 재원으로 한국데이터산업진흥원-2023 데이터청년캠퍼스 지원을 받았습니다.