



HANKUK UNIVERSITY
OF FOREIGN STUDIES

서울 대학 상권의 분석 :회귀 분석을 통한 인사이트

안우혁 김예슬 이태녕

Background

코로나 팬데믹 이후 대학 상권은 다양한 변화가 나타났다. 대면 수업 재개로 인해 유동인구가 급증하며 상권 활성화가 나타났지만, 음식점은 줄어듦과 술집은 늘어난 것이 주목되었다.

이러한 변화에 대한 원인과 영향은 여전히 명확하지 않으며, 대학 상권과 업종 간의 관계를 명확히 이해가 필요하다.

연구의 목표는 이러한 대학 상권의 특성과 업종별 변화에 대한 인사이트를 제공하여 지역 상권과 외식업종에 대한 이해를 증진하고 지역 상권의 발전을 지원하며 창업자와 자영업자에게 유용한 정보를 제공하는 것이다.

Objectives

1. 대학 상권의 변화 이해: 대학 상권의 음식점 및 술집 변화에 대한 이유와 영향을 명확히 이해한다.
2. 업종별 분포 분석: 음식점과 술집의 밀도 변화에 초점을 맞춰, 대학 상권의 업종별 분포 변화를 분석한다.
3. 지역 상권 발전 지원: 연구 결과를 통해 대학 상권과 지역 상권의 발전을 지원하고 창업자 및 자영업자에게 유용한 정보와 제언을 제공한다.

Method

기존의 상권매출에 영향을 미치는 변수들의 관계성에 대한 구체적인 정보를 얻고자 회귀분석을 선택하였다. 이 방법은 다양한 변수 간의 관계성과 각 변수가 종속변수인 상권매출에 미치는 영향을 추정하는 데 유용하다.

밀집이용시설로부터의 거리,최근의 트렌드,업종의 밀집도에 따른 변수를 추가하여 영향력을 분석 또한 대형상권과 소규모상권을 구분하기 위한 지시변수를 추가하여 종합적인 분석을 수행하였다. 이를 통해 상권매출에 영향을 미치는 새로운 관점을 제시하고자 한다.

1.데이터 수집과 전처리

-범주형 자료로 주어진 서비스_업종_코드를 one-hot-encoding하여 전처리

(‘서울열린데이터광장’에서 제공하는 ‘서울시상권분석서비스’를 참고하여 외식업종을 분류하였다.)

-대학정문의 gps좌표에서 거리가 1000m 이내인 매출 gps좌표만을 획득하여 데이터셋 처리

(WGS84 좌표계 기준, Haversin 공식을 사용하여 gps좌표사이 거리 계산)

2.회귀 분석 모델 구축

-다중회귀분석

수집한 데이터를 기반으로 변수를 설정

상권 매출에 영향을 분석할 때 근처역, 근처대학의 수와 재학생수에 따른 상권분류, 카페수와 음식점수 등 여러 요인들을 독립변수로 두고 매출(종속변수) 예측

3.모델 평가

Adjusted R-squared값을 기준으로 모델을 선택

4.모델 개선

전체 Feature set에서 p-value가 가장 큰 변수부터 제거하는 후진 제거법을 사용하여 예측성능을 개선

Elastic Net regularization을 사용하여 구해진 feature에 대해 과적합을 방지

Results

회귀분석 결과 p-value 값으로 종속변수의 유의미성 판단 및 대학상권과 주변 환경과의 상관 분석

(i) 근처 역의 개수: 대학 상권의 형성에 근처 역의 개수가 주요한 영향을 미친다. 역이 존재하는지 여부보다는 근처 역의 개수가 대학 상권과 규모의 독립성을 더 명확하게 보여준다.

(ii) 학교 규모와의 상관관계: 학생 수와 주점 수 사이에 기대했던 상관관계가 나타나지 않으며, 대학의 규모가 대학 상권의 주점 수에 큰 영향을 미치지 않음을 나타낸다.

(iii) 카페와 음식점의 관계: 카페 수와 주점 수의 상관관계를 입증하지 못했다. 반면, 음식점 수와 주점 수 사이에는 상당한 관련성이 있다.

(iv) 대학의 위치와 도심화의 영향: k-Means 클러스터링을 통해 대학들을 지리적으로 그룹화하면, 그룹별로 매출이 상이하므로 대학그룹의 영향력과 대학상권별 특성을 알 수있다. 비교적 매출에 영향력이 높은 그룹과 그렇지 않은 그룹을 분리하여서 대학생 의존도가 높은 상권과 그렇지 않은 상권을 특정해내었다.

Conclusions

우리 연구는 대학 상권 내 상권을 분석하였다. 근처역의분포와는 연관성을 찾아볼 수있었고, 도심화된 상권이라는 특성이 대학상권을 분류하는 또다른 변수임도 찾을 수 있었다. 단순 잠재고객으로 판단하였던 재학생수의 경우도 상권에는 유의미한 변수가 아닌것으로 보인다. 학생들의 등하교 데이터가 추가로 주어진다면 더 자세한 분석이 가능할 것이다. 현재는 제한된 정보접근으로 서울시의 그룹별 대학상권을 분석하였지만, 전국적인 개별 대학에 대한 분석을 실행한다면 위에서 추정한 변수에 관해 더욱 세부적인 결과를 얻을 수있을 것으로 기대된다.