

서울시 소비자의 특성에 따른 요식업종 선호도 파악 및 생활지역권 내의 상권 분석을 통한 창업 가이드 제공

CUA이 4기 금융B팀

윤다인(소프트웨어학부), 최은서(소프트웨어학부), 허인(응용통계학과)

[요약] 본 프로젝트는 요식업 분야의 자영업자에게 창업 가이드를 제공하는 것을 목표로 한다. 서울특별시 내의 성별과 연령에 따른 요식업에 대한 선호도 및 상권, 생활인구, 매출건수를 분석함으로써 창업 시 성공 가능성(매출이 척도가 된다)이 높은 생활지역권, 시간대, 업종을 제시하고자 한다. 더 나아가, 손쉬운 가이드 이용을 위해 여러 시각화 결과를 제시한다.

1. 서론

국토연구원이 2021년 6월 4일 발표한 ‘코로나19 유행 1년간의 서울시 지역경제 변화와 시사점’ 자료에 따르면, 2020년 4분기 요식업 매출이 전년 4분기 매출보다 28%이상 감소하였다. 긴급재난지원금이 지급된 시기에 매출이 소폭 상승하는 상가도 존재하였으나, 대부분의 매출은 하락세를 이어갔다. 하지만, 요식업의 매출 하락폭은 지역 및 소비자의 특성에 따라 달라졌다.

요식업 매출은 영업 시간에도 영향을 받는다. SBS에서 현재 방영하는 프로그램 ‘백종원의 골목식당’에서 백종원 대표는 여러 식당의 문제점을 파악하고 해결 방안을 제시한다. 해당 프로그램에서 백종원 대표는 공통적으로 ‘타임마케팅’의 중요성을 이야기한다. 24시간을 식당에서 영업하는 것 보다 시간대별로 세분화하여 집중 영업 시간, 즉 영업 및 매출의 건수, 금액을 증가시키는 시간을 알고 이외에 시간은 영업준비나 휴식에 사용하는 것이 장기적으로는 음식의 맛을 보전할 수 있어 장기적으로 식당에 이롭다는 것이다.

본 프로젝트에서는 ‘서울시’로 지역을 한정하며, 도시관리 계획을 구체적으로 제시하기 위한 지침인 ‘지역생활권’을 사용하여 서울시를 생활권 단위로 분류함으로써, 특정 지역의 상권분석을 가능하게 하였다.

본 프로젝트에서는 다음 3가지의 데이터를 활용하였다. ①[KB 국민카드 지역별 음식점 세부 업종 집계] 데이터의 날짜, 성별, 연령별 요식업 매출건수를 이용하여 코로나19 전후 성별, 연령별 선호업종을 파악했다. ②[공공데이터포털 - 소상공인시장진흥공단 상가(상권) 정보] 데이터를 이용해 생활권의 요식업종별 상가 개수를 파악함으로써 특정 지역의 동일 업종 현황을 확인하였다. ③[서울 열린데이터 광장 - 서울생활인구] 데이터를 활용해 코로나19 전후 특정 생활권의 예비소비자를 분석

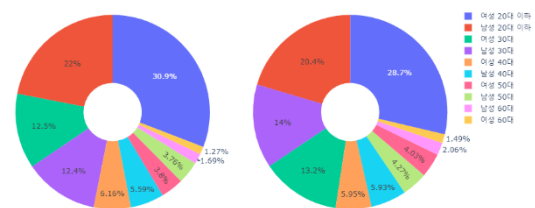
하였다. 생활인구는 특정 시점에 각 공간에 있는 인구를 추정해 놓은 데이터로 해당 생활지역권 내에서 언제든지 식사할 수 있는 소비자로 판단하였다. 이후 시계열 분석을 통해 지역생활권 내 예비소비자의 성별 및 연령에 따른 요식업종의 매출건수를 예측함으로써 지역생활권-시간대별 추천업종과 지역생활권-업종별 추천 영업시간을 알 수 있는 가이드를 제공하였다.

본 프로젝트는 코로나19 전후 요식업 분야를 비교하고, 그에 따른 가이드를 제공함으로써 요식업 분야 창업 혹은 기존의 요식업 분야 개편에 도움이 되어 코로나19로 인하여 침체된 자영업 및 경제 전반의 활성화에 기여할 것으로 예상된다.

2. 본론

1) 데이터 전처리 및 시각화

각 성별, 연령에 따라 선호하는 업종을 분석했다. 1차로 마케팅업종 중분류를 기준 업종으로 삼아 분석을 진행하고, 2차로 더 좁은 범위인 마케팅업종 소분류를 기준으로 삼아 성별, 연령별 선호 업종을 파악하였다. 이를 선그래프와 막대그래프, 파이차트 3가지의 시각화 도구를 이용하여 나타냈으며, 아래 그림은 1차 분석 중 파이차트로 나타낸 결과를 제시한 것이다.

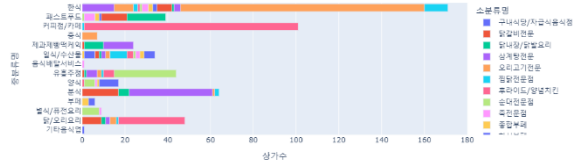


<그림1> 코로나19 전(왼쪽)과 후(오른쪽) 양식(마케팅업종 중분류)에 대한 성별 연령별 선호도

특정 업종에 대한 성별, 연령별 선호도는 시기에 따라 큰 차이가 존재하지 않으며, 양식의 경우 성별에 관계없이 20대 이하가 가장 선호한다는 것을 확인할 수 있다.

이후 116개의 생활권 중분류/소분류에 따른 업종별 상가의 개수를 파악하였다. 앞서 제시한 3개의 데이터의 지역 분할 기준이 다르고(행정동, 법정동의 차이) ‘동’을 분류 기준으로 할 경우 대상 지역의 개수가 많아지는 점, ‘구’를 분류 기준으로 할

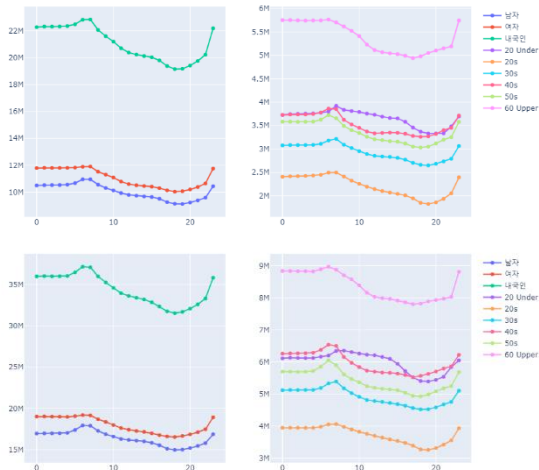
경우 대상 지역 수가 작아지는 문제가 발생하여 행정동을 지역생활권 단위로 묶어 합 116개의 지역생활권이 도출되었으며, 이를 분석 대상 지역으로 삼았다.



<그림2> 개포, 일원 생활권 내 업종별 상가 개수

위 그림은 개포, 일원 생활권 내에 있는 업종별 상가 개수를 나타낸 것이다. 한식을 운영하는 상가 개수가 다른 업종에 비해 많다는 것을 한 눈에 파악할 수 있으며, 그 중에서도 오리고기전문 상가의 개수가 많다는 것을 알 수 있다. 지역생활권 내 업종별 상가 개수를 파악함으로써 향후 창업 가이드 제공 시 상가의 개수와 매출건수의 비례 관계를 파악하고자 하였다.

다음으로 지역생활권 내 시간에 따른 성별, 연령별 생활인구수를 파악하였다. 하루를 4개의 시간대, 새벽(0~6시), 아침(6~12시), 점심(12~18시), 저녁(18~24시)로 나누어 세부적으로 소비자를 분석함으로써 향후 요식업종 및 업종 영업시간에 대한 정확한 추천이 가능하도록 하였다. 코로나19의 영향을 파악하기 위하여 코로나19 전후를 나누어 분석을 진행하였다.



<그림3> 코로나19 전(위)과 후(아래) 개포, 일원 생활지역 내 시간에 따른 성별, 연령별 생활인구

분석 결과 개포, 일원 생활권의 경우 시간대에 따른 성별, 연령별 생활인구수는 코로나19 전과 비슷한 양상을 보인다. 코로나19로 인한 거리두기 등 방역지침에도 불구하고 코로나19 이후 전체 생활인구수가 증가한 것으로 보아 해당 생활권은 코로나19로 인해 지역 경제에 타격을 받지 않았음을 확인할 수 있다.

2) 행렬 분해를 통한 협업 필터링

협업 필터링을 기반으로, 행렬 분해 기법을 활용

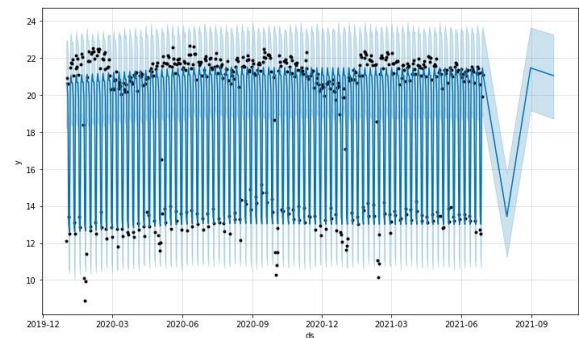
해 생활지역권-성별연령 요인 행렬과 요식업종-성별연령 요인 행렬의 전치 행렬(즉, 성별연령-요식업종 행렬)로 분해된 데이터 세트를 다시 내적 곱으로 결합하였다.

생활지역권-성별연령 요인 행렬을 도출하기 위하여 시계열 분석을 진행하였다. Facebook이 제공하는 Prophet 라이브러리를 활용하여 년월일에 따른 각 성별연령의 생활인구수를 시계열 분석을 통해 예측하였다.

<표1> 가락 생활지역권 내 남자20대 이하 생활인구수 예측 결과 (단위: 만 명)

ds	yhat	yhat_lower	yhat_upper
2021-06-29	21.459616	19.041010	23.666428
2021-06-30	21.051511	18.764383	23.462325
2021-07-31	13.432051	11.250938	15.748942
2021-08-31	21.458394	19.169269	23.635351
2021-09-30	21.042409	18.723242	23.256738

ds(년월일)에 따른 yhat(생활인구수 예측값)과 yhat_lower(생활인구수 하한선), yhat_upper(생활인구수 상한선)를 확인하였다. 향후 3개월의 생활인구수를 예측하여 예측결과의 평균을 가락 생활지역권 내 남자20대 이하의 내달 생활인구수로 확정하였다.



<그림4> 가락 생활지역권의 남자20대이하 생활인구수 예측 결과 (단위: 만 명)

예측 결과를 그래프 추이로 확인하였다. x축은 년월을 나타낸 것이고, y축은 생활인구수(단위: 만 명)를 나타낸 것이다. 산점도가 더 이상 나타나지 않는 이후의 그래프가 예측 결과를 나타낸 것이며, 진한 파란색의 선은 yhat을, 파란색 선의 위아래의 하늘색 음영은 yhat_lower와 yhat_upper의 차이를 보여준다.

시계열 모델 평가를 위해 기존 데이터셋의 마지막 1개월을 제외한 데이터를 train 데이터로, 마지막 1개월을 test 데이터로 활용하였다. 절대 오차 평균(MAE)을 이용해 모델 평가를 진행하였고 모델 평가 결과 모델이 적합하다고 판단해 최종적으로 생활지역권-성별연령별 요인 행렬을 도출하였다.

다음으로 성별연령-요식업종 행렬을 도출하였다. [KB 국민카드] 데이터와 [공공데이터포털] 데이터에서 제공하는 요식업종의 범위가 달라 요식업종의 종류 및 범위를 [공공데이터포털] 데이터에 맞게 통일하였고, 이에 따라 성별연령-요식업종 행렬을 도출하였다. 이를 앞서 도출한 생활지역권-성별연령별 요인 행렬과 내적곱을 하여 각 생활지역권별로

선호되는 요식업종(매출건수로 파악)을 도출하였다. 이때, 지역생활권 내에서 특정 업종의 상가 개수가 많을 경우 소비자가 분산되어 매출이 감소할 수 있다는 점을 감안하여 지역 내 해당 업종의 상가 수로 나누어 최종적으로 한 지역 내 해당 업종 상가 1개 기준 평균 매출건수를 도출하였다.

3) 지역생활권별 요식업종 가이드 제공

분석 결과를 바탕으로 지역생활권에 따라 요식업종 및 영업시간 추천을 포함하는 가이드를 생성하였다.

지역생활권별 요식업종 가이드

1. 지역생활권이란?
 서울시 지역을 생활권 단위로 나누어 도시관리 계획을 구체적으로 제시하기 위한 공간 관리 지침

**분석하고 싶은 지역생활권을 선택하세요

지역생활권

영업시간별

업종별

지역생활권별 요식업종 가이드

1. 지역생활권이란?
 서울시 지역을 생활권 단위로 나누어 도시관리 계획을 구체적으로 제시하기 위한 공간 관리 지침

지역생활권

영업시간별

업종별

2. 가좌지역 생활권 상권분석
 행정구역 : 북가좌2동, 남가좌2동, 북가좌동, 남가좌동

<그림5> 생활지역별별 요식업종 가이드 화면: 지역생활권 및 업종 선택 전(위)과 후(아래)

지역생활권이 생소한 사용자를 위해 ①지역생활권 정의 ②선택한 지역생활권에 어떤 행동동이 속하는지 나타내었다. 지역생활권 선택 시 가로 누적 그래프를 이용한 상가수(그림2)와 코로나 전후 영업시간에 따른 성별 연령별 생활인구수와 비율을 직관적으로 파악할 수 있게 그래프(그림3)로 나타내었다. 영업 시간은 전체, 아침, 점심, 저녁, 새벽 중 선택 가능하다. ③코로나19 전후 성별, 연령별 업종에 대한 선호도와 선택 업종의 성별, 연령별 선호 비율을 매출건수를 이용해서 파악한 뒤 이를 각각 막대그래프, 선그래프, 원 그래프(그림1와 다수)로 표현하였다. 마지막으로 ④지역생활권-영업시간별 추천 업종과 지역생활권-업종별 추천 영업시간을 제공하였다.

3. 결 론

본 프로젝트는 서울특별시 내 지역생활권을 분석 대상 지역으로 삼아, 성별과 연령에 따른 업종 선호도와 생활인구수, 상가의 개수를 통해 상권을 분석하였다. 분석 결과, 업종별 선호도는 코로나19 전후 시기와 관계없이 큰 차이를 보이지 않았으며, 생활지역내 업종별 선호도를 확인한 결과 다수의 지역에서 한식에 대한 선호도가 압도적으로 높은 것을 파악할 수 있었다. 예비소비자의 척도로 사용한 생활인구수는 지역생활권에 따라 코로나19 전후로 증가하거나 감소하는 등 매우 상이한 양상을 보였다. 이를 바탕으로 요식업에 종사하는 자영업자 및 창업가를 위한 가이드를 제공하였다. 특정 지역생활권 내에서 매출건수가 높을 것으로 예상되는 업종

및 영업시간을 추천하였기에 매우 유의미한 결과라고 판단된다.

시계열 분석 및 모델 생성 과정에서 생활인구수 예측값의 하한선과 상한선의 차이를 더욱 작게 만들었다면 현재의 연구 결과보다 더욱 구체적이고 정확한 분석 결과가 도출되었을 것이라고 예상된다. 또한, 요식업 가이드를 홈페이지, 사이트 구축을 통해 제공하였다면 프로젝트의 취지에 맞게 사용자의 편리성에 더욱 부합했을 것이다.

본 프로젝트에는 116개의 지역생활권과 성별, 연령, 매출건수, 생활인구수, 상가 개수 등 다양한 요인들이 분석 대상이 되었기에 분석 및 시각화 결과의 양이 매우 방대하다. 이를 소논문에 모두 담을 수 없는 점을 감안하여, 본 소논문에는 예시가 될 수 있는 분석 결과만 추출하여 설명하였으며 그 외의 모든 분석 및 시각화 결과는 CUA이 github repository에서 확인 가능하다.

참고 문헌

- “코로나19 유행 1년간의 서울시 지역경제 변화와 시사점”, 국토연구원, 2021.06.04
- “지금 아니면 못 사요”... ‘타임 마케팅’ 통하다, 2019.02.07, 한경,
<https://www.hankyung.com/economy/article/201902077030i>
- 서울생활권계획,
<https://planning.seoul.go.kr/plan/main.do>
- 파이썬 머신러닝 완벽 가이드, p.570~578 (잠재 요인 협업 필터링)
- “Time Series Forecasting - ARIMA vs Prophet”, 2020.01.15, Krish Hariharan,
<https://medium.com/analytics-vidhya/time-series-forecasting-arima-vs-prophet-5015928e402a>
- Chart Events, <https://plotly.com/python/chart-events/>