

오산시 오색시장 유동인구 주요 요인 분석 및 시계열 모델링

김시연*

*한신대학교 컴퓨터공학부
pennya6@hs.ac.kr

Distribution of Major Factors of floating population and TimeSeries Modeling in Osan Osaek market

Kim siyeon

*Department of Computer Engineering, Hanshin University

요 약

오산시 오색시장은 시장과 상인의 역사를 바탕으로 다양한 문화 체험 콘텐츠를 결합하여 활기찬 젊음의 생기를 느낄 수 있는 대표 시장이다. 전통시장의 활성화를 위해 오색시장 유동인구에 영향을 주는 요인들을 분석하고자 한다. 대형마트 운영일, 주말, 인구수, 날씨, 코로나 단계 정도, 장날 등의 데이터를 추가적으로 수집하여 주 요인을 분석하였다. 이후 시계열 모델링을 통해 유동인구를 예측하여 추후에 유동인구 예측에 도움을 주고자 한다.

1. 서 론

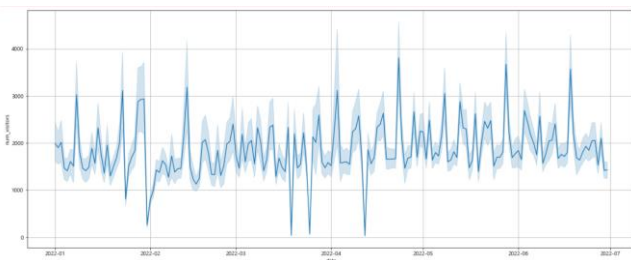
오산시 중앙동에 위치한 오색시장은 1792년 화성궐리지 등장 이후 2015년 문화관광형 시장에 선정될 만큼 대표적인 시장이다. 오색시장은 5개의 즐거움을 주는 길을 통해 다양한 매력을 줌으로써 시민들에게 큰 사랑을 받고 있다.

하지만 재래시장의 선호도가 줄어들고 있는 요즘 재래시장을 이용하는 유동인구의 주요 요인을 파악하는 것은 큰 쟁점이다.

2. 선행 연구

2.1 22년 상반기 유동인구 시각화

오산시 오색시장 22년도 상반기 유동인구를 월별로 시각화한 결과이다. 아래 (그림 1)을 통해 확인할 수 있다.



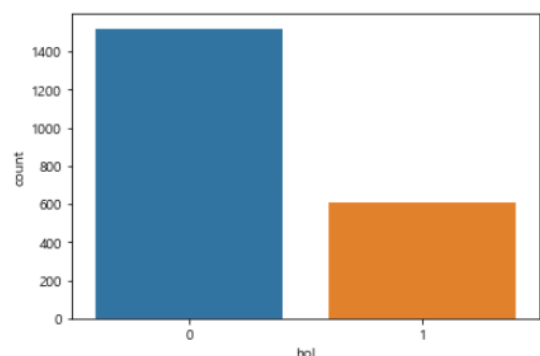
(그림 1) 오산시 오색시장 유동인구 시각화

2.2 새로운 feature

보다 정확한 데이터 분석을 위해 새로운 feature들을 추가하였다. 주말여부, 장날여부, 오산시 인구수, 월별 코로나 확진자 수, 대형마트 휴무일, 날씨, 강수 여부를 새롭게 추가하였다.

2.3 가설 1) 주말여부와 오색시장 유동인구는 연관성이 있다.

시장을 이용하는 고객들이 주말인 토, 일에 이용할 확률이 높을 것이라는 가설을 토대로 데이터를 생성하여 분석을 진행하였다. 주말이면 1, 아니면 0 으로 범주형 데이터로 생성하였다. 아래 (그림 2)을 통해 비율을 확인할 수 있다.

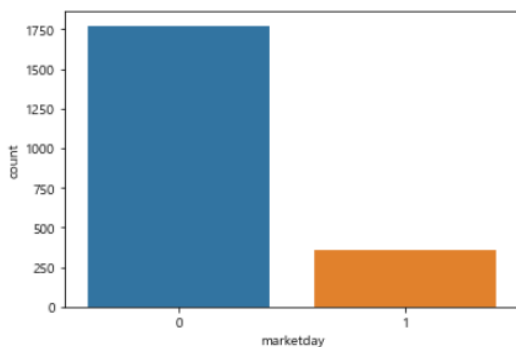


(그림 2) 주말여부 시각화

ttest 결과 -16.2325 로 대립가설이 맞음을 확인할 수 있다.

2.4 가설 2) 대형마트 휴무일과 오색시장 유동인구는 관계가 있다.

대형마트 휴무 시 재래시장 이용 확률이 더 높을 것이라는 대립가설을 토대로 데이터 분석을 실시하였다. 오산시 오색시장은 3,8,13,18,23,28 일이 장날임을 확인하였다. 이를 토대로 장날이면 1 아니면 0으로 범주형 데이터를 생성하였다. 아래 (그림 3)을 통해 비율을 확인할 수 있다.



(그림 3) 장날여부 시각화

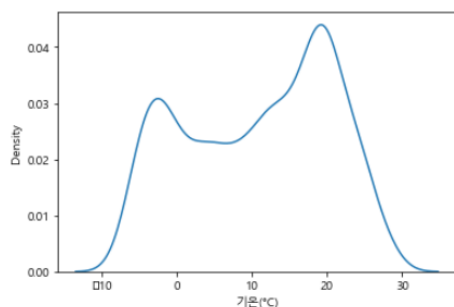
ttest 결과 -17.2620 로 대립가설이 맞음을 확인할 수 있다.

2.5 가설 3) 기온과 오색시장 유동인구는 관계가 있다.

재래시장은 실외라는 특징이 있다. 따라서 기온이 너무 춥거나 더운 경우 이용객이 감소할 것이라는 대립가설을 세웠다. 기온과 지면 온도 데이터를 월별로 시각화 한 결과는 (그림 4)를 통해 확인할 수 있다.



(그림 4) 월별 기온 추세 시각화



(그림 5) 기온 히스토그램

기온과 유동인구 상관분석 결과 0.16으로 약한 상관관계임을 확인할 수 있다.

3. 모델링 및 결과

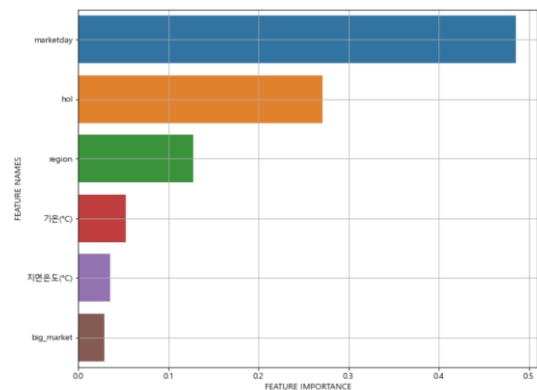
앞에서 진행한 데이터 분석을 토대로 이후 모델링을 진행하였다. LinearRegression, KNN, SVR, 의사결정트리, 랜덤포레스트, XGBoost, ARIMA, SARIMA, SARIMAX 모델링을 진행하였고 결과는 다음과 같다.

모델	MAPE	R2 score
LinearRegression	2.75	-0.08
KNN	0.17	-0.88
SVR	0.72	-0.03
의사결정트리	0.71	0.21
랜덤포레스트	0.72	0.25
XGBoost	0.06	0.80
SARIMAX	0.11	

(표 1) 모델링 결과

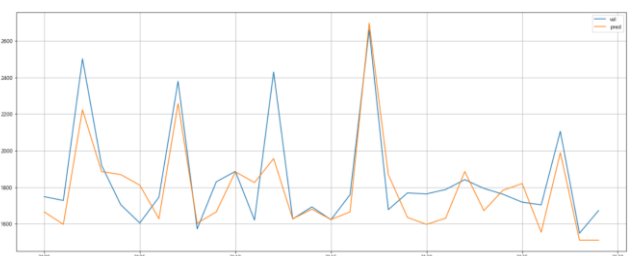
모델링 결과 XGBoost와 SARIMAX 모델이 높은 정확도임을 확인할 수 있다.

(그림 6)은 XGBoost 변수 중요도이다. 새롭게 생성한 변수들의 중요도가 높음을 확인할 수 있다.



(그림 6) Xgboost 변수 중요도

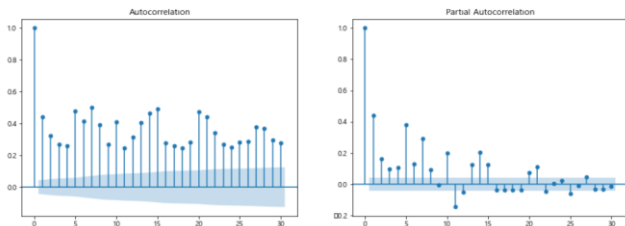
(그림 7)은 XGBoost 예측 시각화 그래프이다. 추세는 비슷하지만 최소, 최대 데이터 예측 부분에서 오차가 큼을 확인할 수 있다.



(그림 7) Xgboost 예측 시각화

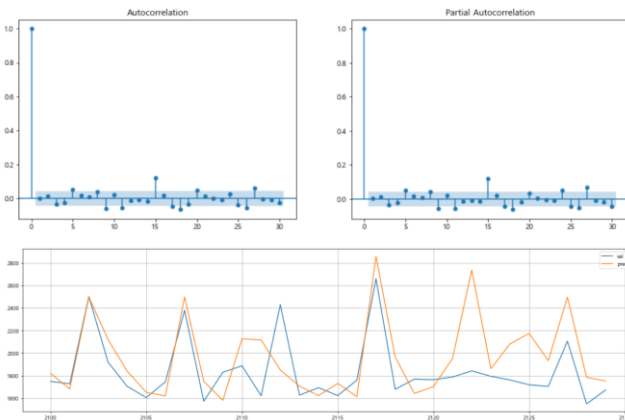
따라서 추세와 최소, 최대 오차를 줄일 수 있는 SARIMAX 모델을 사용하였다. (그림 8)은 y_train에 대한 자기상관성 확인 결과이다. 정규성 검정 0.00746, 정상성

검정 0.54662로 미적합하다. 또한 잔차분석시 7일의 패턴이 있음을 확인 할 수 있다.



(그림 8) y_{train} 의 잔차분석 시각화

SARIMAX $p=4, d=1, q=4, P=2, D=1, Q=3, m=7$ 로 학습을 하였다. 각 값은 하이퍼파라미터 튜닝을 통해 찾은 값이다. 그 결과 $aic = 32553$, $mape=0.1166$ 이다. 잔차분석 결과는 (그림 9)을 통해 확인할 수 있다.



(그림 9) 잔차분석 및 예측 데이터 시각화

4. 결론

위 연구를 통해 오색시장 유동인구 주요요인을 분석하였다. 주말여부, 장날여부가 큰 영향을 미침을 확인할 수 있었다. 모델링 결과 높은 정확도의 모델링 결과를 얻지 못했다. 하지만 추세와 최소, 최대를 줄일 수 있었던 전통적인 시계열 모델인 SARIMAX가 적절하다고 판단하였다.

추후에 다양한 데이터 추가하여 모델링한다면 보다 높은 정확도를 기대할 수 있을 것으로 예상된다.

참 고 문 헌

- [1] "화성밸리지" 오산시 오색시장 홈페이지
<http://5colormarket.com/bbs/page.php?hid=history>