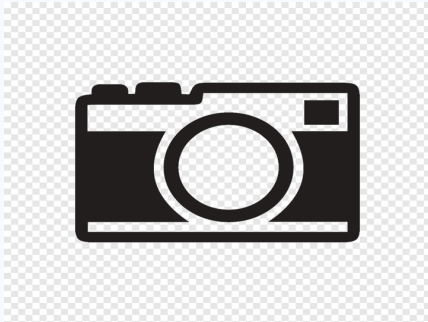


예상 매출액 추정 보고서

2024-06-05 16:17:31

예상 연간 매출액은

0,000,000,000원 ~
0,000,000,000원입니다.



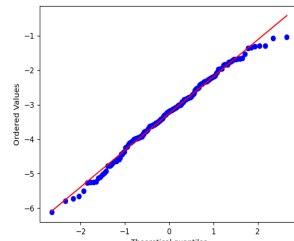
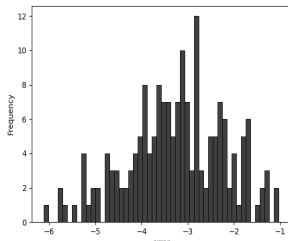
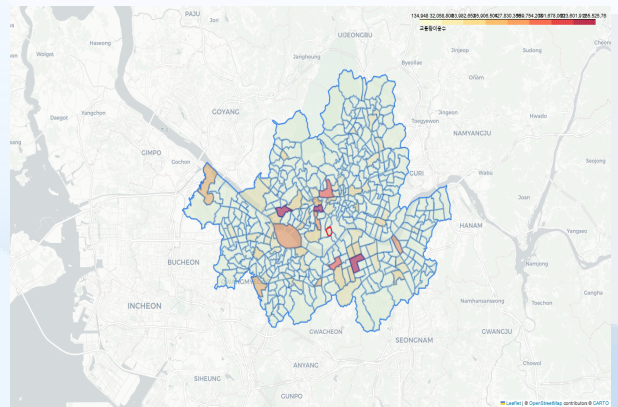
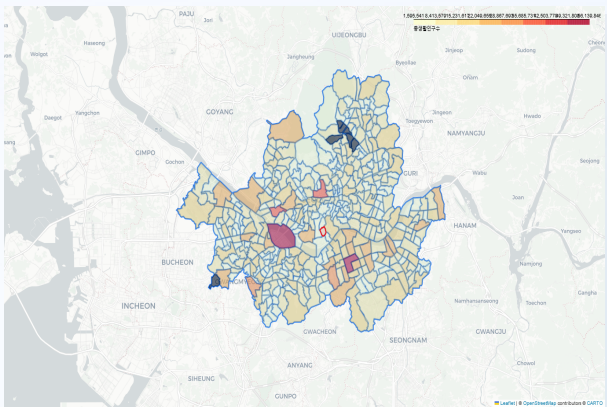
본 분석 과제는 랜덤 포레스트(Random Forest) 기법을 사용하여 전년 특정 행정동 점포당 예상 매출액을 예측함. 랜덤포레스트기법은 앙상블 학습 기법 중 하나로, 다수의 의사 결정 트리(Decision Trees)를 사용하여 예측 성능을 향상시키는 방법임. 그리고 사용자가 선택한 행정동의 과거 연도별 매출액 데이터 분포를 통해 당해 예상 매출액을 회귀예측하여, 랜덤포레스트 기법을 통해 생성한 예측값과 비교 실시. 서비스의 신뢰성을 확보하기 위해, 비교 결과 두 예측값의 차이가 크지 않은 경우에만 대하여 예상 매출액 추정 서비스를 제공함.

예측 변수의 선정은 결과 변수에 유의미한 영향을 줄 것으로 예상되는 행정동별 지하철/버스 승하차인원, 생활인구, 인당 연간 총 소득 금액 및 지출 총금액, 점포수를 선정. 유사한 주제의 선행 연구를 고려하여, 연구 지역의 금융기관의 수가 패스트푸드 매출액에 유의미한 영향을 미친다는 결론 반영, 금융기관 소재 예측 변수 포함.



행정동코드	버스교통량	지하철교통량	금융기관	생활인구	인당 연간 총 소득 금액	지출 총금액	음식 지출 총금액	패스트푸드 점포수
-------	-------	--------	------	------	---------------	--------	-----------	-----------

결과 변수인 전년 예상 매출액과 가장 높은 피어슨-상관 계수를 갖는 예측 변수인 교통량, 생활인구를 시각화한 HeatMap은 다음과 같음.



오차의 분포는 positive-skewed 분포이며, 이를 로그변환을 통해 정규분포에 가깝게 나타낸 모습은 좌측 사진과 같음. 오차 분포의 95% 신뢰구간을 구하기 위해 정규성을 검증한 후, 이를 다시 역변환하여 각 구간의 끝값을 도출함. 정규성 검증 결과, 정성적 검증 수단인 QQ Plot은 우측 사진과 같으며, 정량적 검증 수단인 Shapiro-wilk 검증 결과는 다음과 같음.

Statistic : 0.9901507353817449
p-value : 0.26189148637806126