

강남구 편의점
매출 예측 서비스



Convenience Insights

Contents

01 프로젝트 소개

- 팀원 소개
- 기획 배경

02 데이터 분석

- 데이터 수집
- 데이터 전처리
- EDA

03 모델링

- 종속변수 Scailing
- 예측모델 및 피처중요도
- 실제값 vs 예측값

04 서비스 및 문서

- Streamlit 소개
- WBS
- Work-Flow
- 참고문헌 및 분석도구

팀원소개



최혜린 팀장

- 데이터 전처리
- EDA
- 지도 시각화
- Streamlit 구현



김진아 팀원

- 데이터 전처리
- EDA / 통계 분석
- Machine Learning
- Streamlit 구현



송 민 팀원

- EDA / 통계 분석
- Machine Learning



송준호 팀원

- EDA / 통계 분석
- Machine Learning
- 지도 시각화
- Streamlit 구현



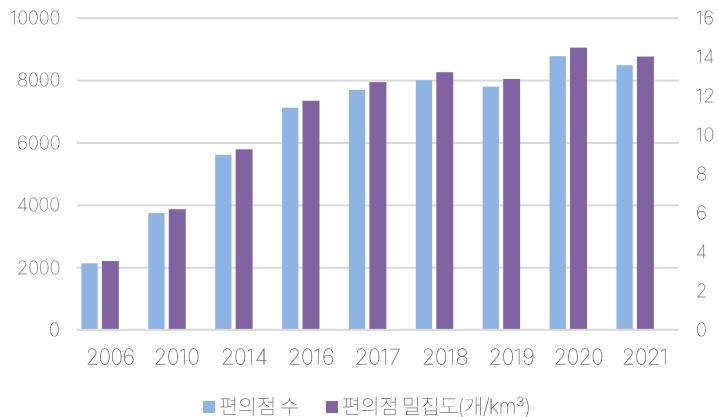
한대희 팀원

- 통계 분석
- Machine Learning

기획배경

- 편의점 창업 수요의 증가로 서울지역 편의점 수가 15년새 4배 증가
- 편의점 창업시, 편의점 본사에서 가맹점주에게 예상 매출액을 알려주는 법이 존재하지만 창업 후 본사가 제시한 매출액과 실제 매출액의 차이가 커 가맹점주들이 어려움을 겪는 사례가 발생

| 서울 지역 편의점수와 밀집도 |



| 편의점 창업 시장 문제점 |

A screenshot of a news article from 'Homyplus' (홈플러스) discussing franchise issues. The article highlights a discrepancy between expected sales and actual sales, leading to difficulties for franchisees. It includes quotes from franchisees and franchisees, and a summary of the problem points.

■ 편의점 분쟁사례

- “일매출 약 130만원 나올 것”
가맹본부
- “실제 일매출 70만원에 그쳐”
가맹점주 A씨

■ 페업하려 했더니…

- 중도해지 위약금 8100만원 물어야!
- 본사가 부담한 시설·인테리어 비용
- 본사의 기대이익 상실분 포함
- 가맹점주, 분쟁조정 신청
- 위약금 1200만원 감면 6900만원 지불해야

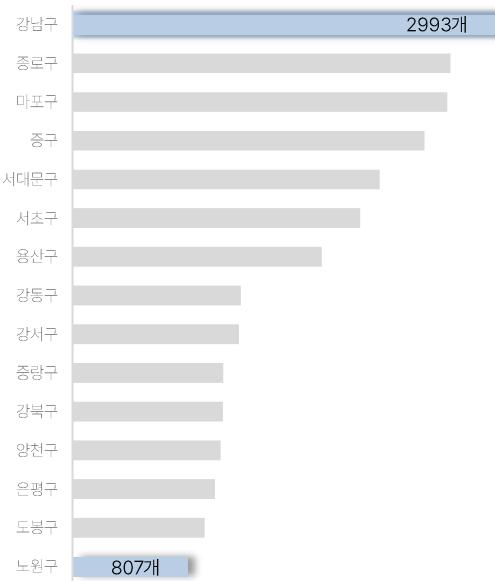
■ 숨은 문제점

- 예상 매출액 구두 전달하는 경우 많아
- 예상 매출액 달라도 본사 책임 물기 어려워

기획배경

- 서울시에서 강남구와 노원구의 편의시설 수의 차이가 매우 크며, 편의점 개수 또한 약 2.6배로 큰 차이가 남
- 이는 강남구에 편의점 창업 시 경쟁이 더 심한 것으로 해석할 수 있으며, 상권 분석이 필수적임을 시사

| 자치구별 편의시설 수 |

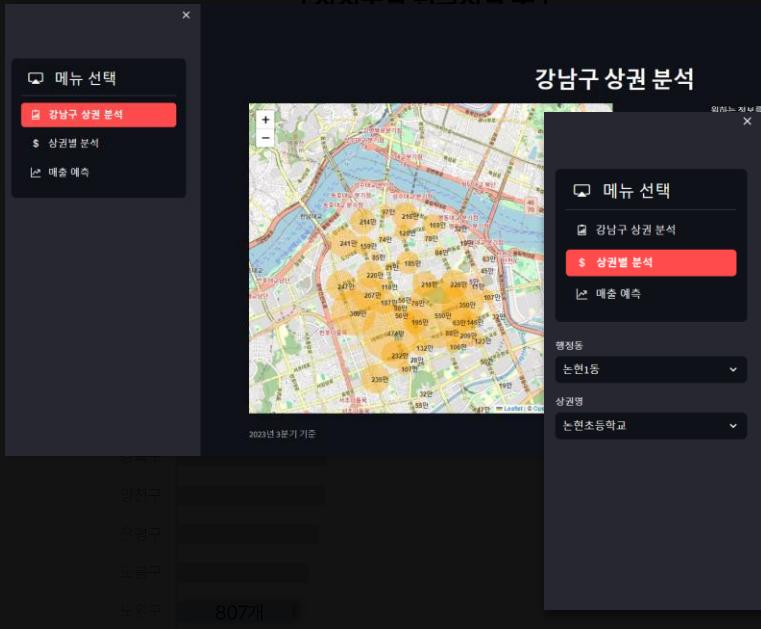


| 노원구와 강남구의 편의점 수 |



강남구 상권 분석 및 편의점 매출 예측 서비스

- 서울시에서 강남구와 노원구의 편의점 매출 예측 서비스
 - 이는 강남구에 편의점 창업을 희망하는 사람들이 손쉽게 정보를 얻을 수 있는



데이터 수집

대분류	서울열린데이터광장 상권 분석 서비스 데이터							
구분	영역	길단위인구	상주인구	소득소비	직장인구	집객시설	추정매출	점포
상세 내용	<ul style="list-style-type: none"> - 상권_구분_코드 - 상권_구분_코드_명 - 상권_코드 - 상권_코드_명 - 자치구_코드 - 자치구_코드_명 - 행정동_코드 - 행정동_코드_명 - 영역_면적 	<ul style="list-style-type: none"> - 기준_년분기_코드 - 상권_구분_코드 - 상권_구분_코드_명 - 상권_코드 - 상권_코드_명 - 시간대별 유동인구수 - 요일별 유동인구수 	<ul style="list-style-type: none"> - 기준_년분기_코드 - 상권_구분_코드 - 상권_구분_코드_명 - 상권_코드 - 상권_코드_명 - 총_상주인구_수 - 총_가구_수 - 연령별_상주인구_수 	<ul style="list-style-type: none"> - 기준_년분기_코드 - 상권_구분_코드 - 상권_구분_코드_명 - 상권_코드 - 상권_코드_명 - 월_평균_소득_금액 - 소득_구간_코드 - 지출_총금액 	<ul style="list-style-type: none"> - 기준_년분기_코드 - 상권_구분_코드 - 상권_구분_코드_명 - 상권_코드 - 상권_코드_명 - 총_직장_인구_수 - 상권_코드_명 - 연령별_직장_인구_수 		<ul style="list-style-type: none"> - 기준_년분기_코드 - 상권_구분_코드 - 상권_구분_코드_명 - 상권_코드 - 상권_코드_명 - 서비스_업종_코드 - 서비스_업종_코드_명 - 서비스_업종_코드_명 - 유사_입종_점포_수 - 개업_점포_수 - 폐업_점포_수 	<ul style="list-style-type: none"> - 기준_년분기_코드 - 상권_구분_코드 - 상권_구분_코드_명 - 상권_코드 - 상권_코드_명 - 서비스_업종_코드 - 서비스_업종_코드_명 - 서비스_업종_코드_명 - 유사_입종_점포_수 - 개업_점포_수 - 폐업_점포_수

데이터 전처리

데이터 병합							
영역	길단위인구	상주인구	소득소비	직장인구	집객시설	추정매출	점포
상권_구분_코드	기준_년분기_코드						
상권_구분_코드_명	상권_구분_코드						
상권_코드_명	상권_구분_코드_명						
상권_코드_명	상권_코드						
자치구_코드	상권_코드_명						
자치구_코드_명	시간대	총_상주인구_수	월_평균_소득_금액	총_직장_인구_수	집객시설_수	서비스_업종_코드	서비스_업종_코드
...

필요한 컬럼 추출							
•	자치구_코드_명 == 강남구						
•	서비스_업종_코드_명 == 편의점						

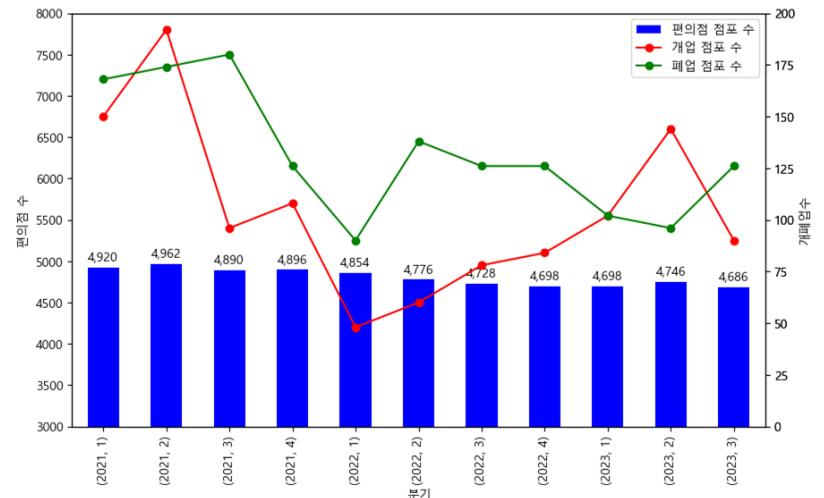


결측값 처리							
집객시설 수							
•	집객시설 수가 null 인 값은 0으로 대체						
월 평균 소득 금액							
•	동일한 행정동 내 상권들의 평균값으로 대체						
지출 총 금액							
•	동일한 행정동 내 상권들의 평균값으로 대체						

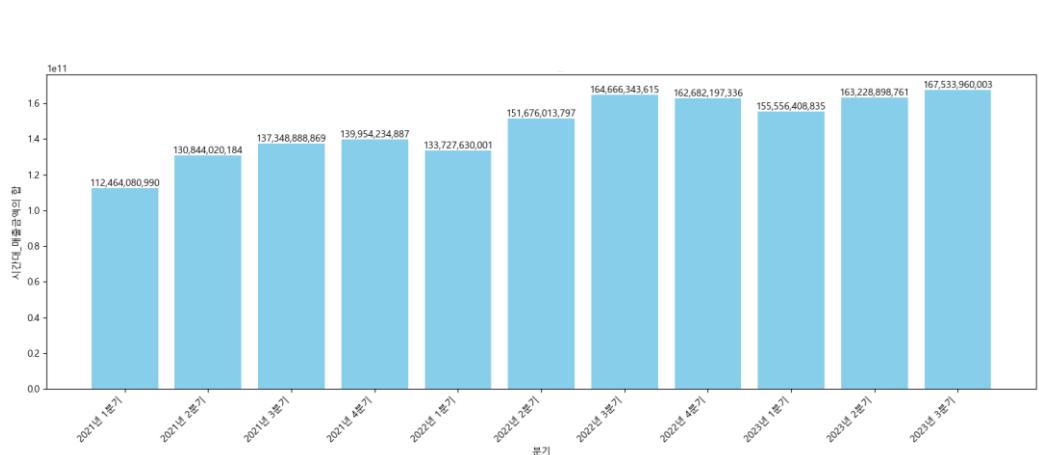
탐색적 통계분석(EDA)

- 분기별로 보았을 때 편의점 개업 점포 수보다 폐업 점포 수가 더 많음
- 편의점 수는 조금씩 줄어드는 추이를 보이나, 거의 비슷한 수준 기록 중
- 반면에 매출 금액은 조금씩 증가하는 추이 기록

| 분기별 편의점 수와 개폐업수 |



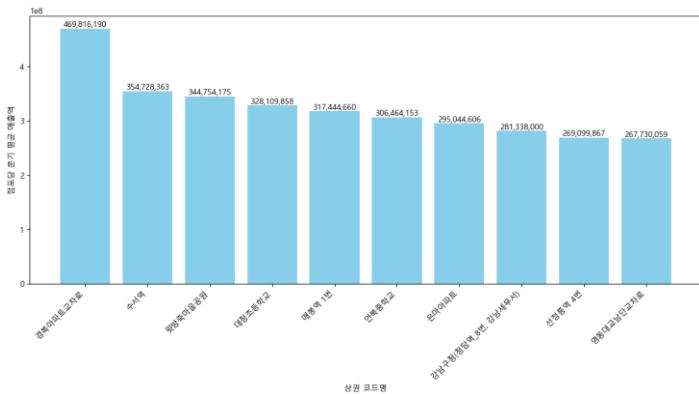
| 연도 및 분기별 매출금액 |



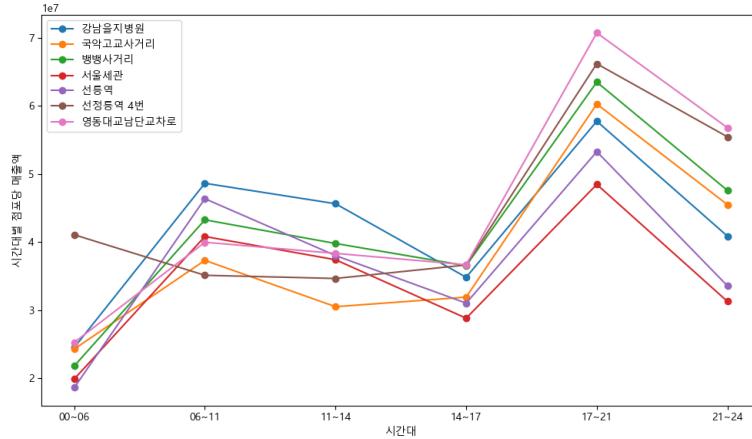
탐색적 통계분석(EDA)

- 상권에 따라 매출액에 차이가 있음
- 임의의 상권을 추출했을 때, 시간대별 매출 추이가 비슷하나, 상권별로 조금씩 다른 매출 추이를 보임

| 상권별 평균 매출액 |



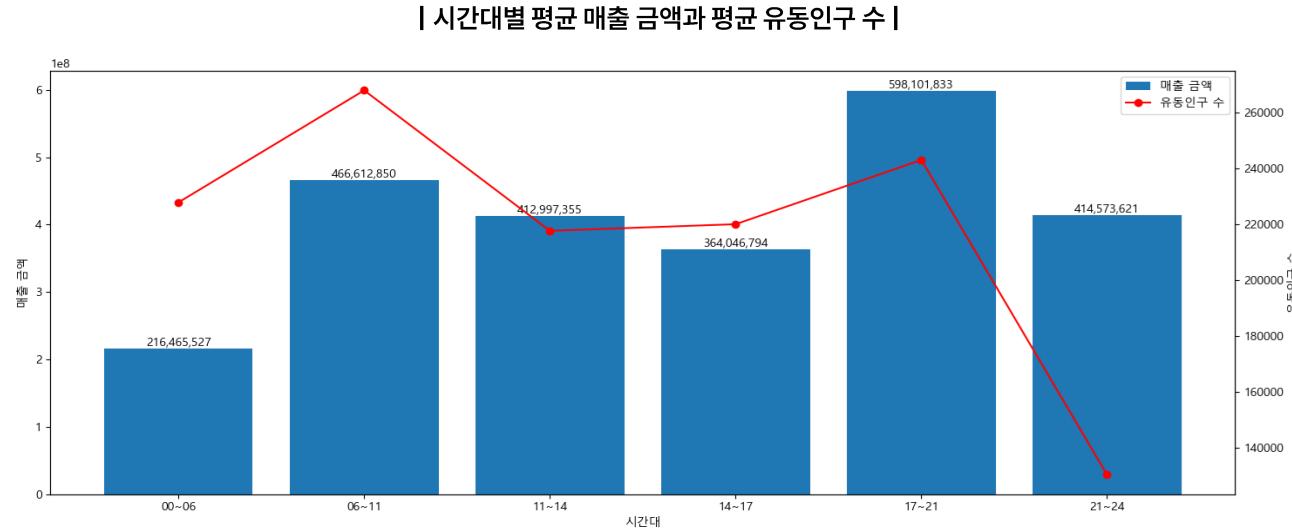
| 상권에 따른 시간대별 점포당 평균 매출액 |



❖ 상권에 따라 시간대별 매출에 영향을 끼치는 요인이 다를 것으로 예상

탐색적 통계분석(EDA)

- 출퇴근 시간에 높은 매출을 기록함
- 유동인구가 많을 때 매출이 증가하는 추이를 보이지만, 21~24 시간대에서는 유동인구가 크게 줄어듦에도 불구하고 출퇴근 시간 다음으로 많은 매출을 기록함



❖ 유동인구 수 외에도 다른 변수들이 시간대 매출에 영향을 끼칠 것으로 예상할 수 있음

탐색적 통계분석(EDA)

- 선행연구에 기반하여, 기존 컬럼에서 '편의점 밀도', '연령별 상주인구 비율', '연령별 직장인구 비율' 파생변수 생성

최종 컬럼

- 기준_년
- 기준_분기
- 상권_구분_코드_명
- 상권_코드_명
- 행정동_코드_명
- 시간대
- 시간대_매출금액
- 시간대_유동인구_수
- 총_직장_인구_수
- 총_상주인구_수
- 총_가구_수
- 집객시설_수
- 월_평균_소득_금액
- 지출_총금액
- 유사_업종_점포_수
- 개업_점포_수
- 폐업_점포_수
- 편의점 밀도
- 연령별 상주인구 비중
- 연령별 직장인구 비중

편의점 밀도

유사_업종_점포_수 / 영역_면적

연령별 상주인구 비율

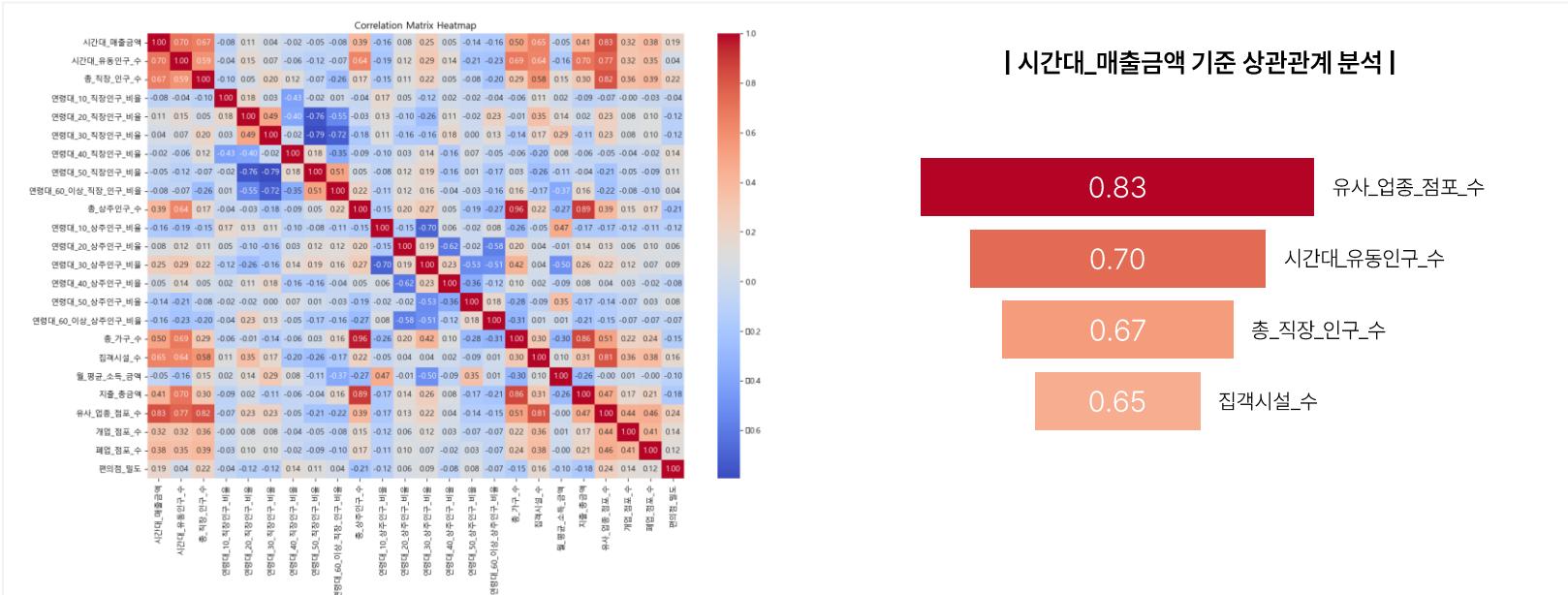
각 연령별 상주인구 수/ 전체 상주인구 수

연령별 직장인구 비율

각 연령별 직장인구 수/ 전체 직장인구 수

탐색적 통계분석(EDA)

상관분석



❖ 점포수, 유동인구, 직장인구, 집객시설 수가 시간대_매출금액에 비교적 많은 영향을 끼칠 것으로 예상

탐색적 통계분석(EDA)

다중공선성 확인

VIF Factor	features	VIF Factor	features		
0	4.457923	시간대_유동인구_수	13	inf	연령대_50_상주인구_비율
1	5.303403	총_직장_인구_수	14	inf	연령대_60_이상_상주인구_비율
2	inf	연령대_10_직장인구_비율	15	32.874815	총_가구_수
3	inf	연령대_20_직장인구_비율	16	5.411849	집객시설_수
4	inf	연령대_30_직장인구_비율	17	2.586857	월_평균_소득_금액
5	inf	연령대_40_직장인구_비율	18	6.784772	지출_총금액
6	inf	연령대_50_직장인구_비율	19	15.814548	유사_업종_점포_수
7	inf	연령대_60_이상_직장_인구_비율	20	1.357144	개업_점포_수
8	35.345023	총_상주인구_수	21	1.417630	폐업_점포_수
9	inf	연령대_10_상주인구_비율	22	1.574667	편의점_밀도
10	inf	연령대_20_상주인구_비율			
11	inf	연령대_30_상주인구_비율			
12	inf	연령대_40_상주인구_비율			

독립변수들 간의 다중공선성이 10 이상인 변수

- 연령별_직장인구_비율
- 연령별_상주인구_비율
- 총_상주인구_수
- 총_가구_수
- 유사_업종_점포_수

☞ 모델링 시 다중공선성 문제에 영향을 덜 받는 트리기반 모델을 사용

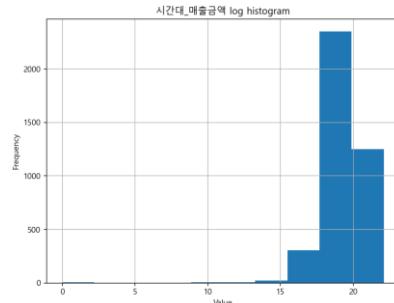
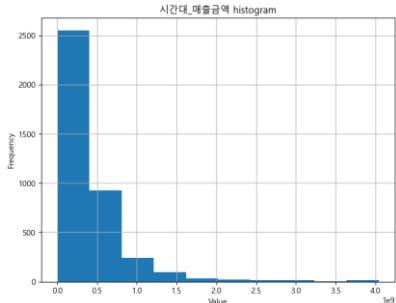
모델링

증속변수 Scaling

- Box-cox 변환은 데이터의 분포를 변환하여 skewness를 조절하는 방법
- 데이터 분포가 정규화를 이루지 않는 모형에서 정규화하는데 많이 사용됨

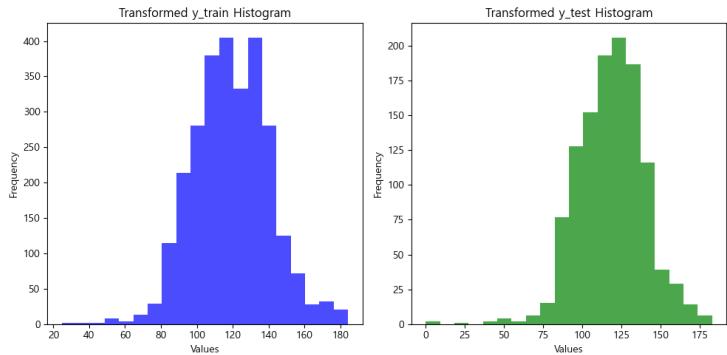
Log 변환

- 종속변수가 왼쪽으로 skew되어있어, 정규분포로 변환하기 위해 log변환 시도
- log 변환 이후, 오른쪽으로 skew되고 분포 문제가 해결되지 않음



Box-cox 변환

- 분포의 skew를 해결해 줄 수 있는 Box-cox 변환 시도
- train 데이터 세트와 test 데이터 세트의 종속변수가 정규분포와 유사한 형태로 변환됨



모델링

예측 모델 및 피처중요도

예측 모델



- 다양한 모델으로 모델링해 본 결과, LGBM에서 가장 우수한 점수 기록하여 해당 모델 선정
- GridsearchCV 방식으로 최적의 하이퍼파라미터 튜닝 진행

초기 점수

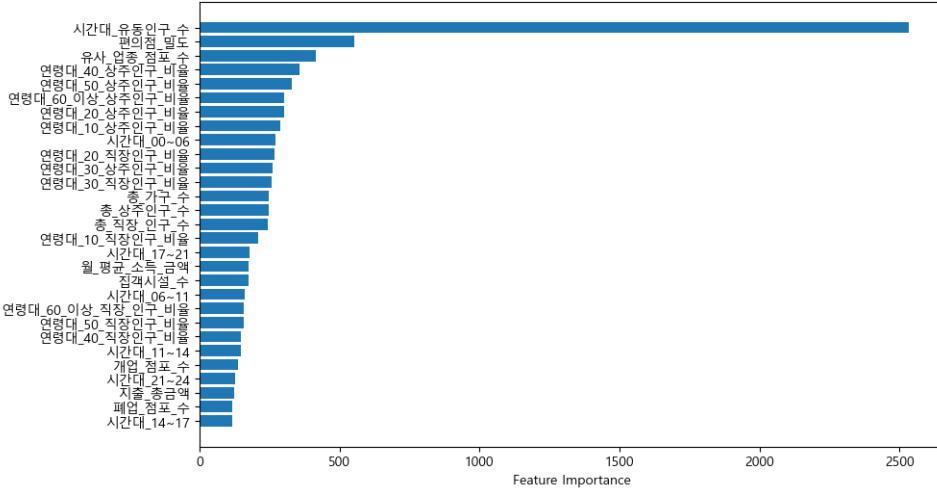
결정계수 - 0.9894
RMSE - 44617455

낮은 FI 제거 후

결정계수 - 0.9890
RMSE - 45499194

- Feature Importance가 가장 낮은 '폐업_점포_수' 컬럼 제거 후 RMSE와 결정계수가 개선되지 않아 기존 변수들을 최종 변수로 결정

피처중요도

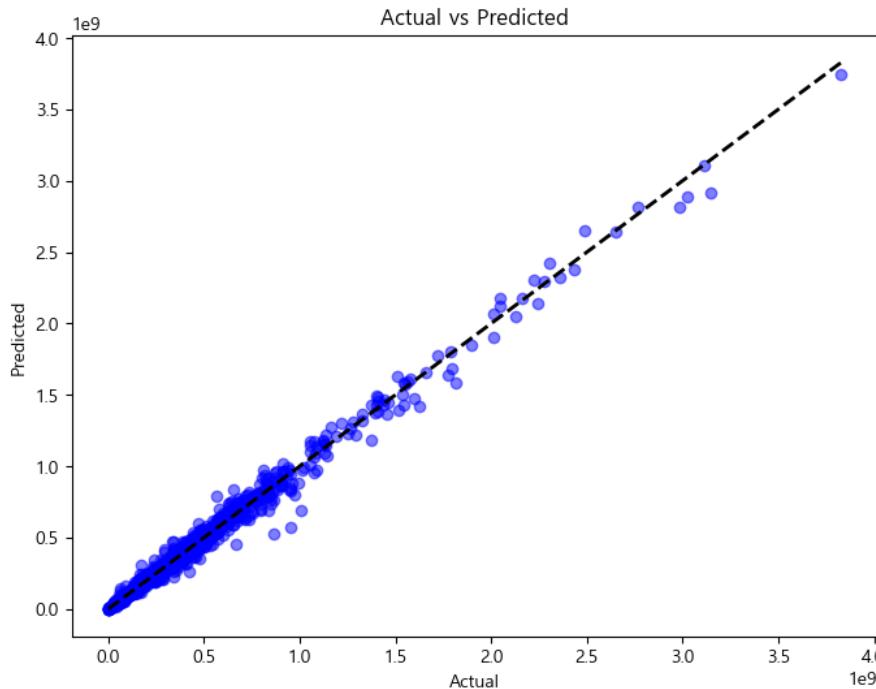


- '시간대_유동인구_수'가 가장 높은 피처 중요도를 보임
- 상관분석 시 높은 점수를 기록했던 유사_업종_점포_수, '시간대_유동인구_수' 피처에서 높은 중요도를 보임

모델링

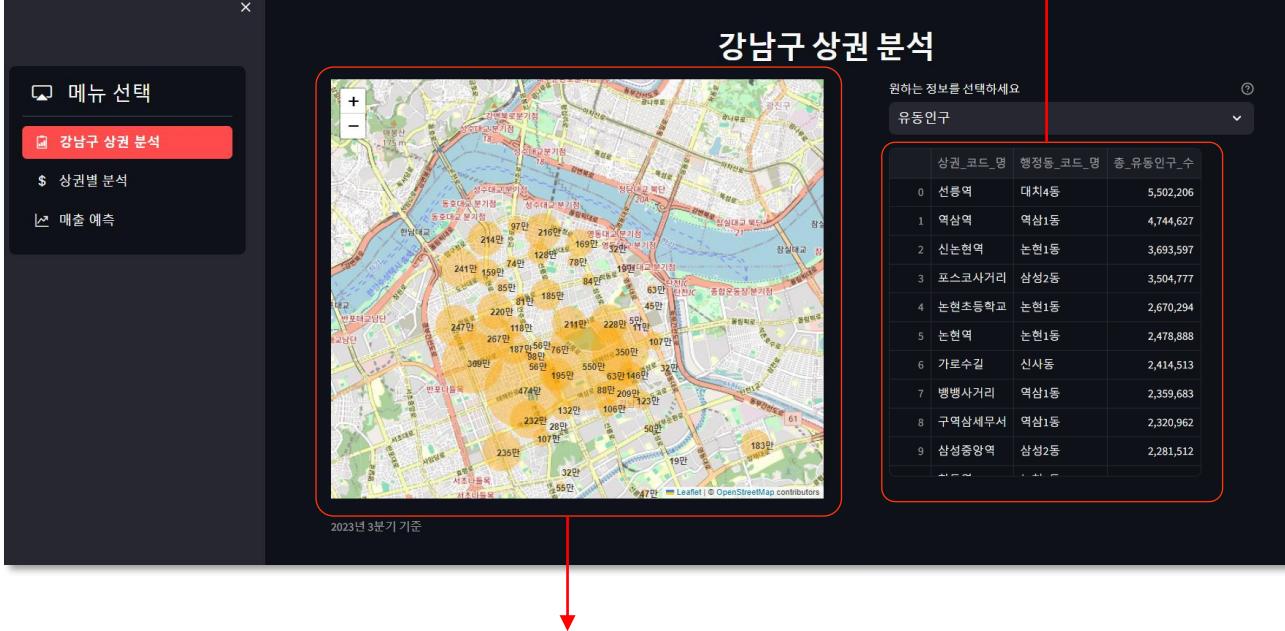
실제값 vs 예측값

- 모델이 예측한 값과 실제 값 사이의 관계를 시각적으로 표현했을 때, 모델이 예측한 값과 실제 값이 비슷한 것으로 확인 됨



Streamlit 소개

강남구 상권 분석 Page



- 선택한 정보에 대한 데이터가 높은 상권순으로 데이터 확인

- 상권위치별로 정보에 대한 데이터를 원의 크기로 나타내어 시각적으로 확인

Streamlit 소개

상권별 분석 Page

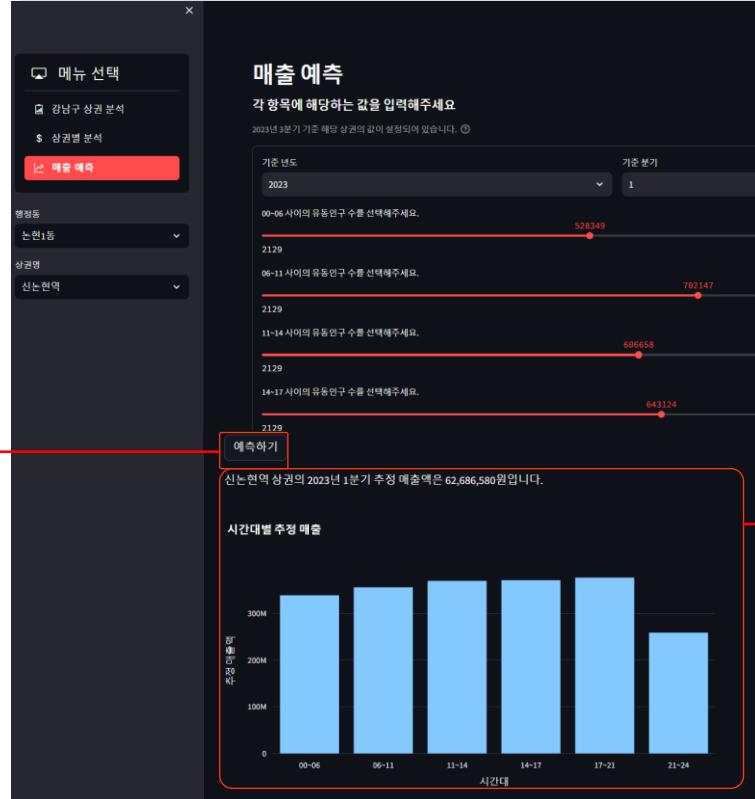
- 원하는 정보를 템 하면 선택한 상권에 대한 자세한 데이터 확인 가능



- 확인하고 싶은
상권 선택

Streamlit 소개

상권별 분석 Page



- 예측하고 싶은 상권에 대한 정보 입력 후 예측하기

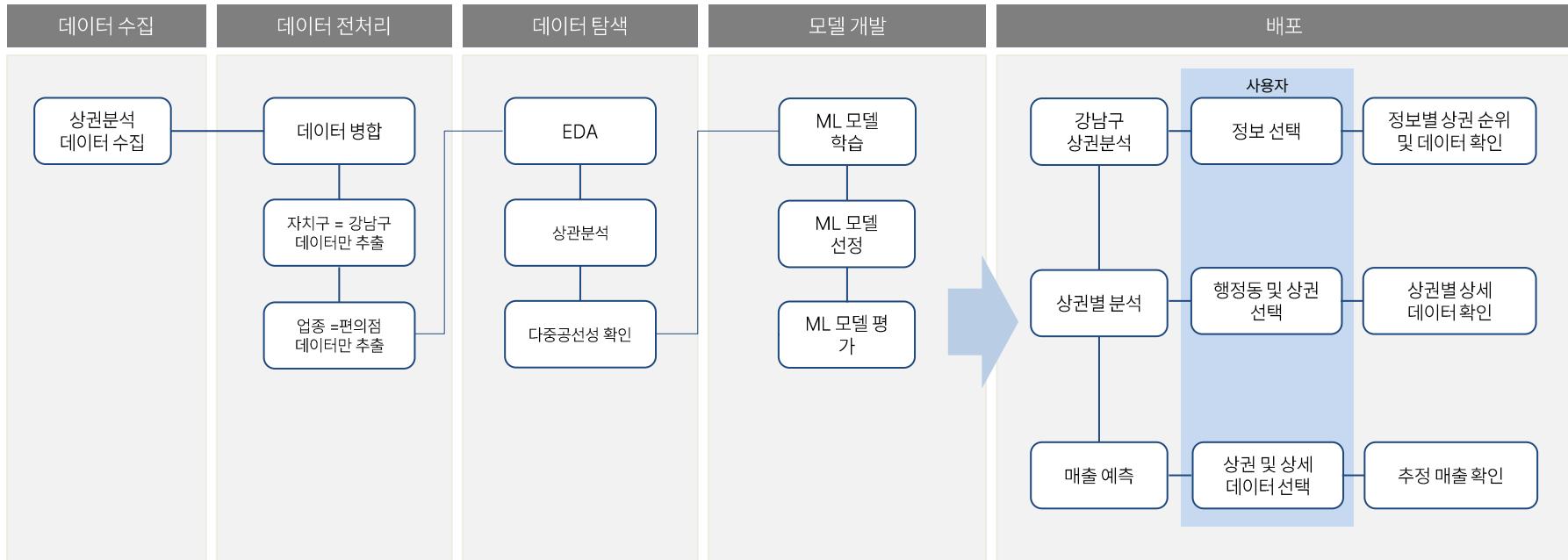


- 해당 상권의 총 추정 매출액 확인
- 시간대별 추정 매출액 확인



WBS

Work-Flow



참고문헌 및 분석도구

논문

- 1) 김현철, 이승일, 2019, "서울시 골목상권 매출액에 영향을 미치는 요인에 관한 연구", 「서울도시연구」, 제 20권 제 1호
- 2) 황규성, 2014, "편의점 입지선정시 매출에 영향을 미치는 요인분석"
- 3) 김미성, 2020, "서울시 상권 데이터의 시각화에 기반한 매출액 예측"
- 4) 이철환, 2012, "편의점의 상권 추정과 매출 예측에 관한 연구"
- 5) 김동명, 2020, "시스템 다이내믹스를 활용한 편의점 특정 상품 매출 분석 및 예측"
- 6) 이임동, 이찬호, 강상목, 2010, "편의점 매출에 영향을 미치는 입지요인에 대한 실증연구"

뉴스 기사

- 1) 송금종, 카페 1950개·편의점 470개...강남구, 서울 최대 '슬세권', 쿠키뉴스, 2023.12.24
<https://www.kukinews.com/newsView/kuk202312140078>
- 2) 이지원, 편의점 본사가 제시한 '예상 매출액'이 과장이라면..., 더스coop, 2023.02.14
<https://www.thescoop.co.kr/news/articleView.html?idxno=56799>
- 3) 이진원, 홈플러스, 예비 편의점주에 예상매출 뻥튀기, 시민일보, 2017.11.05
<https://www.siminilbo.co.kr/news/articleView.html?idxno=537797>

분석 도구



THANK YOU
