

# 강남구 편의점 매출 예측 서비스

4조 - Convenience Insights  
2024-03-07 ~ 2024-03-29



# 목차

---

## 프로젝트 소개



- 팀원 구성
- 기획 배경
- 웹 서비스 소개
- 개발 환경
- 프로젝트 수행 기간

## 프로젝트 문서



- FLOW CHART
- WBS

## 데이터 탐색



- 데이터 수집
- 데이터 전처리

## 통계적 탐색 분석



- EDA
- 상관분석
- 다중공선성

## 모델링



- 모델 생성
- 모델 학습
- 모델 성능 검증
- 편의점 예상 매출 산출

## 서비스 및 문서



- 서비스 (Streamlit)
- 한계점 / 개선 사항
- 참고 자료
- 부록

# 목차

---

## 프로젝트 소개



- 팀원 구성
- 기획 배경
- 웹 서비스 소개
- 개발 환경
- 프로젝트 수행 기간

## 프로젝트 문서



- FLOW CHART
- WBS

## 데이터 탐색



- 데이터 수집
- 데이터 전처리

## 통계적 탐색 분석



- EDA
- 상관분석
- 다중공선성

## 모델링



- 모델 생성
- 모델 학습
- 모델 성능 검증
- 편의점 예상 매출 산출

## 서비스 및 문서



- 서비스 (Streamlit)
- 한계점 / 개선 사항
- 참고 자료
- 부록

# 1. 프로젝트 소개

## 팀원 구성



최혜린 팀장

- 데이터 전처리
- EDA
- 지도 시각화
- Streamlit 구현



김진아 팀원

- 데이터 전처리
- EDA / 통계 분석
- Machine Learning



송민 팀원

- EDA / 통계 분석
- Machine Learning



송준호 팀원

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



한대희 팀원

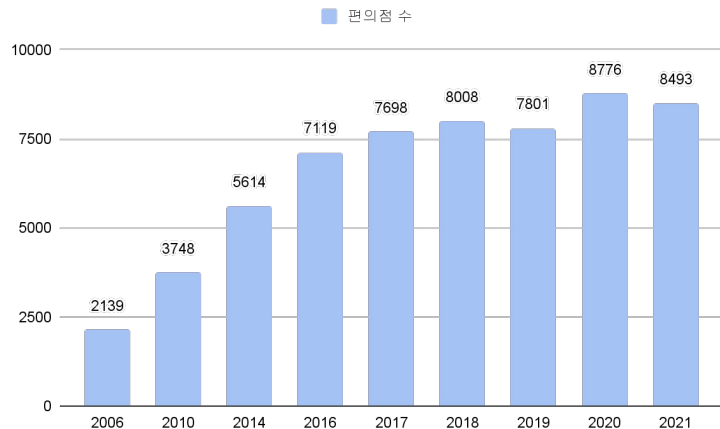
- 통계 분석
- Machine Learning

# 1. 프로젝트 소개

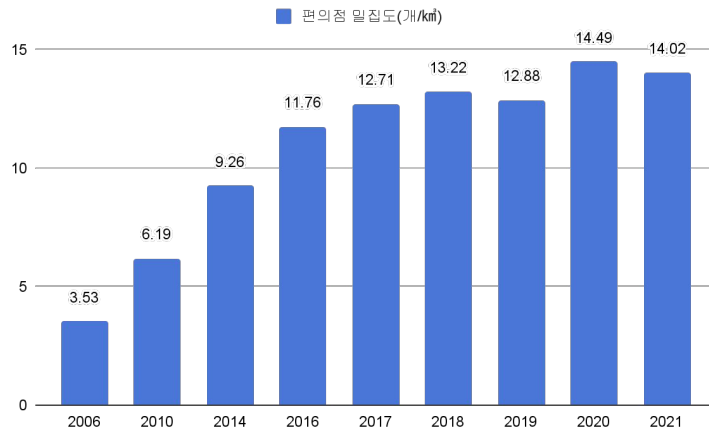
## 기획 배경

- 21년 말 기준 서울의 편의점은 총 **8,493**개로 06년 **2,139**개 대비 약 **4배** 증가
- 1km<sup>2</sup>당 편의점 수를 나타내는 밀집도 또한 06년 3.5개에서 21년 14개로 증가

| 서울 지역 편의점수 |



| 서울 지역 편의점 밀집도 |



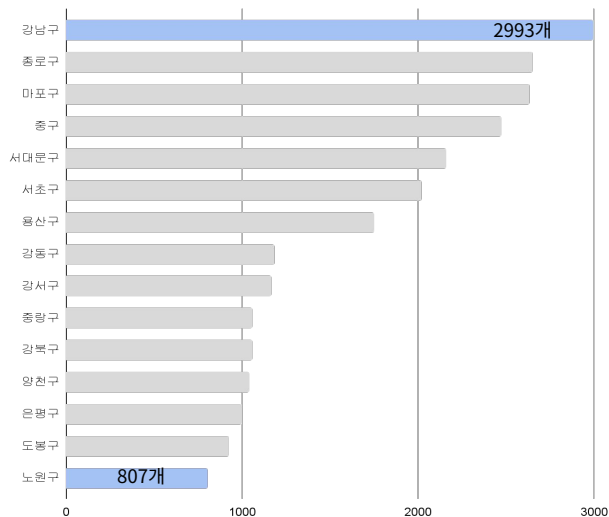
출처: 서울시, 서울 편의점 운영실태 및 현황 분석자료

# 1. 프로젝트 소개

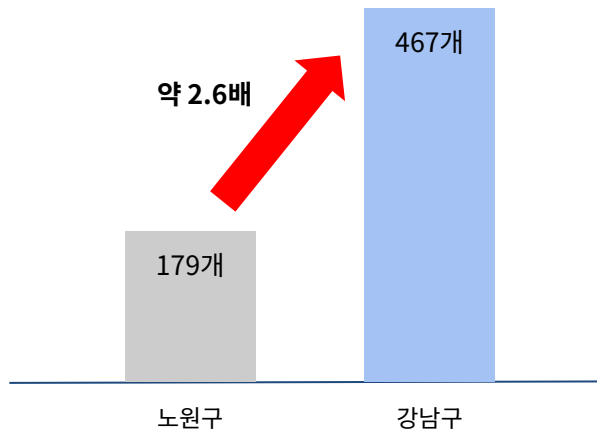
## 기획 배경

- 강남구 - 편의점 467개, 노원구 - 편의점 179개로 약 **2.6배** 차이
- 이는 강남구에 편의점을 창업할 시 경쟁이 더 심한 것으로 해석할 수 있으며, 상권 분석이 필수적임을 시사

| 자치구별 편의시설 수 |



| 노원구와 강남구의 편의점 수 |



출처: 쿠키뉴스, 카페 1950개·편의점 470개...강남구, 서울 최대 '슬세권'

# 1. 프로젝트 소개

## 기획 배경

- 편의점 창업시, 편의점 본사에서 가맹점주에게 **예상 매출액을 알려주는 법**이 존재하지만, 창업 후 **본사가 제시한 매출액**과 **실제 매출액**의 **차이가 커** 가맹점주들이 어려움을 겪는 사례가 발생하기도 함

## | 편의점 창업 시장의 문제점 |

홈 > COVER STORY > 편의점의 역설

편의점 본사가 제시한 **예상매출액**이 과장이라면...

▲ 이지음 기자 | © 일력 2023.02.14 | ☎ 호수 532 | ✉ 댓글 0

HOME > 사회 > 사건/사고

**홈플러스, 예비 편의점주에 예상매출 뺑뚱기**

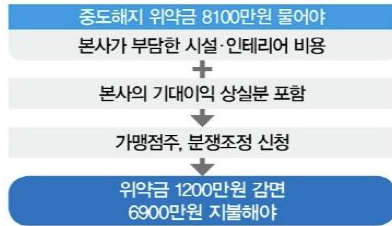
이진원 / yjw@siminilbo.co.kr / 기사승인 : 2017-11-05 16:18:07

### [편의점 분쟁사례]

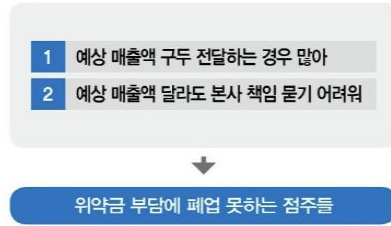


### [폐업하려 했더니...]

[자료 | 경기도 공정거래지원센터]



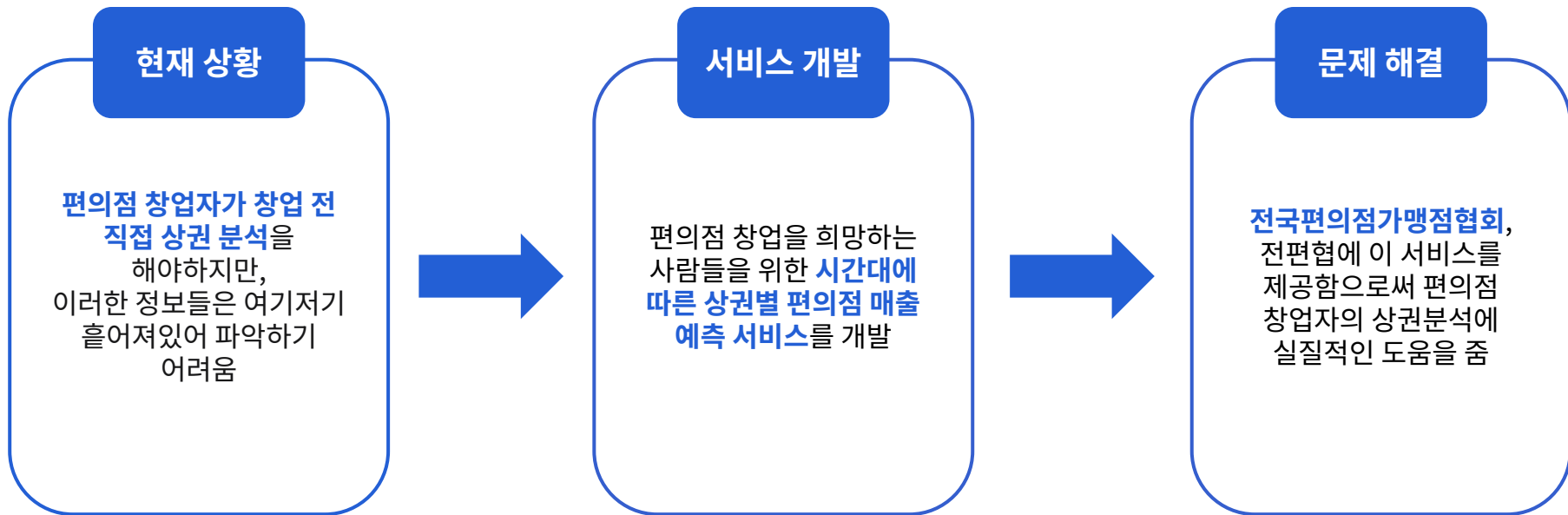
### [숨은 문제점]



출처: 더스쿠프, 편의점 본사가 제시한 ‘예상 매출액’이 과장이라면...  
시민일보, 홈플러스, 예비 편의점주에 예상매출 뺑뚱기

# 1. 프로젝트 소개

기획 배경





# 1. 프로젝트 소개

## 기획 배경

- 전국편의점가맹점협회 사이트에 접속하여 확인한 결과, 경영보고서, 노무관리, 세무정보와 같은 서비스는 존재하지만, **편의점 창업자를 위한 상권분석 서비스**는 없는 것으로 확인

전국편의점가맹점협회

협회소개 ▶ 불공정상담 컨설팅 ▶ 공지뉴스 ▶ 게시판 ▶ 전편협앱 ▶

### 전국 편의점 가맹점 협회

전국 편의점 가맹점 협회는 불공정 제도로 인해 편의점 가맹점주들이 겪는 어려움 해소를 위해 설립되었습니다.  
시대 흐름에 맞는 정책들을 제안하고 실질적인 혜택과 도움이 되도록 노력하겠습니다.

전국편의점가맹점협회

협회소개 ▶ 불공정상담 컨설팅 ▶ 공지뉴스 ▶ 게시판 ▶ 전편협앱 ▶

### 전 편 협 앱

편의점 경영에 특화된 디지털 경영앱

매일매일 돈의 흐름이 보입니다.

매입데이터, 현금영수증  
(세금)계산서를 한눈에!

스마트폰으로 한눈에 매장상황 확인

일바생 관리가 편해져요

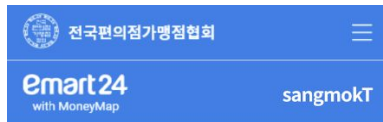
2022년 10월 변경된 근로기준법 개정  
근로계약서  
급여명세서  
출퇴근관리

한눈에 해결!

실수없고 편리한 디지털 세무서비스

세무사와 1:1 독 서비스  
디지털 데이터 기반 실수 차단  
통장 카드 등 데이터 자동전송

전국편의점가맹점협회에서  
점주들의 노하우를 실려 제작하였습니다.



NEW 디지털경영관리 플랫폼 '전편협'... 더보기 >

디지털 경영 관리



경영보고서



노무관리



세무정보



부가서비스



홈



경영보고서



노무관리



세무정보



부가서비스

# 1. 프로젝트 소개

## 웹 서비스 소개

상권별 위치와 평균 매출



강남구  
편의점  
매출 현황

행정동 및 상권별 시간대 평균 매출

시간대별 매출 높은 상권 TOP 5

매출 현황  
순위

매출 예측

시간대별 예상 매출

# 1. 프로젝트 소개

---

개발 환경

개발  
환경



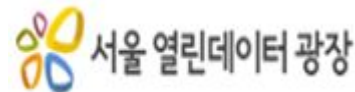
배포



언어

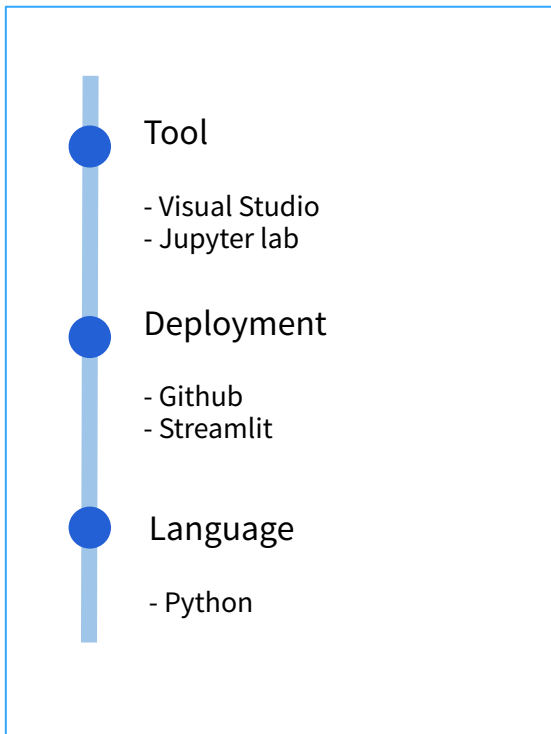


Open  
API



# 1. 프로젝트 소개

## 개발 환경



# 1. 프로젝트 소개

프로젝트 수행 기간



# 목차

---

## 프로젝트 소개



- 팀원 구성
- 기획 배경
- 웹 서비스 소개
- 개발 환경
- 프로젝트 수행 기간

## 프로젝트 문서



- FLOW CHART
- WBS

## 데이터 탐색



- 데이터 수집
- 데이터 전처리

## 통계적 탐색 분석



- EDA
- 상관분석
- 다중공선성

## 모델링



- 모델 생성
- 모델 학습
- 모델 성능 검증
- 편의점 예상 매출 산출

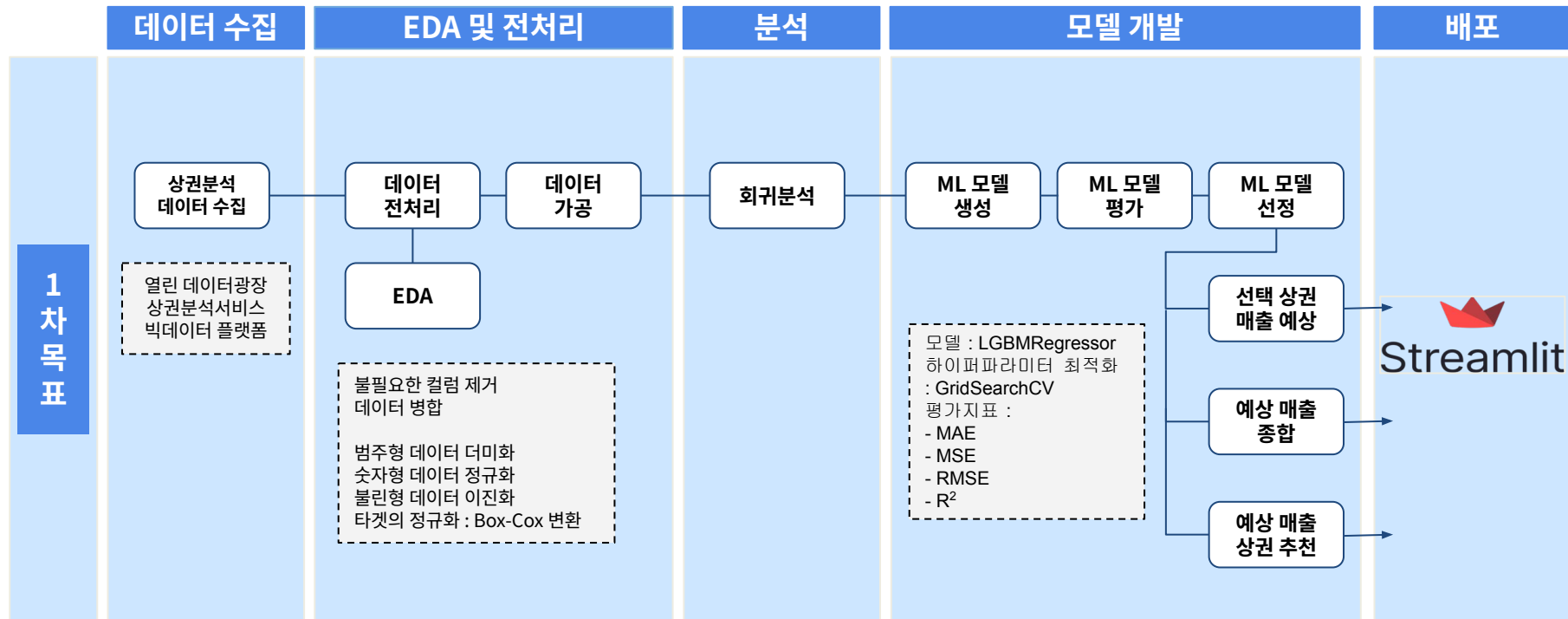
## 서비스 및 문서



- 서비스 (Streamlit)
- 한계점 / 개선 사항
- 참고 자료
- 부록

## 2. 프로젝트 문서

### FLOW CHART



## 2. 프로젝트 문서

### WBS

구분	주요 업무	1주차 (3/06 ~ 3/09)			2주차 (3/10 ~ 3/16)			3주차 (3/17 ~ 3/23)			4주차 (3/24 ~ 3/29)		
기획	주제 선정												
데이터 수집	API 데이터 수집												
데이터 전처리	데이터 형식 통일												
데이터 탐색	EDA												
데이터 시각화	공간 및 지도 시각화												
통계 분석	회귀분석												
모델링	모델 생성 및 평가												
서비스 구현	Streamlit 배포												
최종	프로젝트 발표												



# 목차

## 프로젝트 소개



- 팀원 구성
- 기획 배경
- 웹 서비스 소개
- 개발 환경
- 프로젝트 수행 기간

## 프로젝트 문서



- FLOW CHART
- WBS

## 데이터 탐색



- 데이터 수집
- 데이터 전처리

## 통계적 탐색 분석



- EDA
- 상관분석
- 다중공선성

## 모델링



- 모델 생성
- 모델 학습
- 모델 성능 검증
- 편의점 예상 매출 산출

## 서비스 및 문서



- 서비스 (Streamlit)
- 한계점 / 개선 사항
- 참고 자료
- 부록

# 3. 데이터 탐색

## 데이터 수집

대분류	상권 분석 서비스 데이터							
구분	추정매출	길단위인구	상주인구	직장인구	소득소비	영역	집객시설	점포
상세 내용	<ul style="list-style-type: none"> <li>- 분기 매출 금액</li> <li>- 분기 매출 건수</li> <li>- 시간대 매출 금액</li> <li>- 시간대 매출 건수</li> <li>- 연령대 매출 금액</li> <li>- 연령대 매출 건수</li> <li>- 성별 매출 금액</li> <li>- 성별 매출 건수</li> <li>- 요일 매출 금액</li> <li>- 요일 매출 건수</li> </ul>	<ul style="list-style-type: none"> <li>- 총 유동인구 수</li> <li>- 성별 유동인구 수</li> <li>- 연령 유동인구 수</li> <li>- 시간대 유동인구 수</li> <li>- 요일 유동인구 수</li> </ul>	<ul style="list-style-type: none"> <li>- 총 상주인구 수</li> <li>- 성별 상주인구 수</li> <li>- 연령대 상주인구 수</li> <li>- 총 가구 수</li> <li>- 아파트 가구 수</li> <li>- 비아파트 가구 수</li> </ul>	<ul style="list-style-type: none"> <li>- 총 직장인구 수</li> <li>- 성별 직장인구 수</li> <li>- 연령대 직장인구 수</li> </ul>	<ul style="list-style-type: none"> <li>- 월 평균 소득 금액</li> <li>- 소득 구간</li> <li>- 총 지출 금액</li> <li>- 유형 지출 금액</li> </ul>	<ul style="list-style-type: none"> <li>- X 좌표</li> <li>- Y 좌표</li> <li>- 자치구 코드</li> <li>- 행정동 코드</li> <li>- 면적</li> </ul>	<ul style="list-style-type: none"> <li>- 총 집객시설 수</li> <li>- 유형 집객시설 수</li> </ul>	<ul style="list-style-type: none"> <li>- 개/폐업률</li> <li>- 개/폐업 점포 수</li> <li>- 총 점포 수</li> <li>- 유사 업종 점포 수</li> <li>- 프랜차이즈 점포 수</li> </ul>
출처	서울열린데이터광장							

# 3. 데이터 탐색

## 데이터 전처리

### 1. 강남구, 편의점 데이터만 추출



서비스\_업종\_코드\_명이  
편의점인 데이터 추출



자치구\_코드\_명이  
편의점인 데이터 추출

# 3. 데이터 탐색

데이터 전처리

## 2. 데이터 병합 전 결측치 확인

구분	추정매출	길단위인구	상주인구	직장인구	소득소비	영역	집객시설	점포
결측치	결측치 X	결측치 X	결측치 X	결측치 X	<div>- 월_평균_소득_금액</div> <div>- 소득_구간_코드</div> <div>- 지출_총금액</div> <div>- 식료품_지출_총금액</div> <div>- 의류_신발_지출_총금액</div> <div>- 생활용품_지출_총금액</div> <div>- 의료비_지출_총금액</div> <div>- 교통_지출_총금액</div> <div>- 여가_지출_총금액</div> <div>- 문화_지출_총금액</div> <div>- 교육_지출_총금액</div> <div>- 유흥_지출_총금액</div>	결측치 X	<div>- 관공서_수</div> <div>- 은행_수</div> <div>- 종합병원_수</div> <div>- 일반_병원_수</div> <div>- 약국_수</div> <div>- 유치원_수</div> <div>- 초등학교_수</div> <div>- 고등학교_수</div> <div>- 대학교_수</div> <div>- 백화점_수</div> <div>- 슈퍼마켓_수</div> <div>- 극장_수</div> <div>- 숙박_시설_수</div> <div>- 공항_수</div> <div>- 철도_역_수</div> <div>- 버스_터미널_수</div> <div>- 지하철_역_수</div> <div>- 버스_정거장_수</div>	결측치 X

# 3. 데이터 탐색

데이터 전처리

## 2. 데이터 병합 전 결측치 확인

같은 상권인 다른 분기의  
평균 값으로 대체

구분	추정매출	길단위인구	상주인구	직장인구	소득소비	영역	집객시설	점포
결측치	결측치 X	결측치 X	결측치 X	결측치 X	<ul style="list-style-type: none"> <li>- 월_평균_소득_금액</li> <li>- 소득_구간_코드</li> <li>- 지출_총금액</li> <li>- 식료품_지출_총금액</li> <li>- 의류_신발_지출_총금액</li> <li>- 생활용품_지출_총금액</li> <li>- 의료비_지출_총금액</li> <li>- 교통_지출_총금액</li> <li>- 여가_지출_총금액</li> <li>- 문화_지출_총금액</li> <li>- 교육_지출_총금액</li> <li>- 유흥_지출_총금액</li> </ul>	결측치 X	<ul style="list-style-type: none"> <li>- 관공서_수</li> <li>- 은행_수</li> <li>- 종합병원_수</li> <li>- 일반_병원_수</li> <li>- 약국_수</li> <li>- 유치원_수</li> <li>- 초등학교_수</li> <li>- 고등학교_수</li> <li>- 대학교_수</li> <li>- 백화점_수</li> <li>- 슈퍼마켓_수</li> <li>- 극장_수</li> <li>- 숙박_시설_수</li> <li>- 공항_수</li> <li>- 철도_역_수</li> <li>- 버스_터미널_수</li> <li>- 지하철_역_수</li> <li>- 버스_정거장_수</li> </ul>	결측치 X

# 3. 데이터 탐색

데이터 전처리

## 2. 데이터 병합 전 결측치 확인

결측치, 0으로 대체

구분	추정매출	길단위인구	상주인구	직장인구	소득소비	영역	집객시설	점포
결측치	결측치 X	결측치 X	결측치 X	결측치 X	<ul style="list-style-type: none"> <li>- 월_평균_소득_금액</li> <li>- 소득_구간_코드</li> <li>- 지출_총금액</li> <li>- 식료품_지출_총금액</li> <li>- 의류_신발_지출_총금액</li> <li>- 생활용품_지출_총금액</li> <li>- 의료비_지출_총금액</li> <li>- 교통_지출_총금액</li> <li>- 여가_지출_총금액</li> <li>- 문화_지출_총금액</li> <li>- 교육_지출_총금액</li> <li>- 유흥_지출_총금액</li> </ul>	결측치 X	<ul style="list-style-type: none"> <li>- 관공서_수</li> <li>- 은행_수</li> <li>- 종합병원_수</li> <li>- 일반_병원_수</li> <li>- 약국_수</li> <li>- 유치원_수</li> <li>- 초등학교_수</li> <li>- 고등학교_수</li> <li>- 대학교_수</li> <li>- 백화점_수</li> <li>- 슈퍼마켓_수</li> <li>- 극장_수</li> <li>- 숙박_시설_수</li> <li>- 공항_수</li> <li>- 철도_역_수</li> <li>- 버스_터미널_수</li> <li>- 지하철_역_수</li> <li>- 버스_정거장_수</li> </ul>	결측치 X

# 3. 데이터 탐색

데이터 전처리

## 3. 데이터 병합

구분	공통 컬럼
추정매출	<ul style="list-style-type: none"><li>- 기준_년분기_코드</li><li>- 상권_구분_코드</li><li>- 상권_구분_코드_명</li><li>- 상권_코드</li><li>- 상권_코드_명</li></ul>
길단위인구	
상주인구	
직장인구	
소득소비	
집객시설	
점포	

공통 컬럼을 기준으로  
1차 병합



강남구 위치 정보  
2차 병합

### 3. 데이터 탐색

데이터 전처리

#### 4. 컬럼 변경

기존 컬럼
기준_년분기_코드



변경 컬럼
기준_년 기준_분기

기존 컬럼
시간대_xx~xx_매출_금액



변경 컬럼
시간대 시간대_매출금액

기준_년	기준_분기
2021	1
2021	1
2021	1
2021	1
2021	1

시간대	시간대_매출금액
00~06	377166450.0
06~11	222467605.0
11~14	192457360.0
14~17	230188421.0
17~21	531497598.0



### 3. 데이터 탐색

데이터 전처리

#### 5. 불필요한 컬럼 제거

불필요한 컬럼
- 매출 건수 관련 컬럼 - 성별 관련 컬럼 - 요일 관련 컬럼 . . . - 유형별 집객시설 수 - 유형별 지출 총금액

불필요한 컬럼 제거



최종 컬럼	
• 시간대_매출금액	• 월_평균_소득_금액
• 기준_년	• 지출_총금액
• 기준_분기	• 유사_업종_점포_수
• 상권_구분_코드_명	• 개업_점포_수
• 상권_코드_명	• 폐업_점포_수
• 행정동_코드_명	
• 시간대	
• 시간대_유동인구_수	
• 총_직장_인구_수	
• 총_상주인구_수	
• 총_가구_수	
• 집객시설_수	

# 목차

---

## 프로젝트 소개



- 팀원 구성
- 기획 배경
- 웹 서비스 소개
- 개발 환경
- 프로젝트 수행 기간

## 프로젝트 문서



- FLOW CHART
- WBS

## 데이터 탐색



- 데이터 수집
- 데이터 전처리

## 통계적 탐색 분석



- EDA
- 상관분석
- 다중공선성

## 모델링



- 모델 생성
- 모델 학습
- 모델 성능 검증
- 편의점 예상 매출 산출

## 서비스 및 문서



- 서비스 (Streamlit)
- 한계점 / 개선 사항
- 참고 자료
- 부록

## 4. 통계적 탐색 분석

최종 컬럼 확인

종속변수

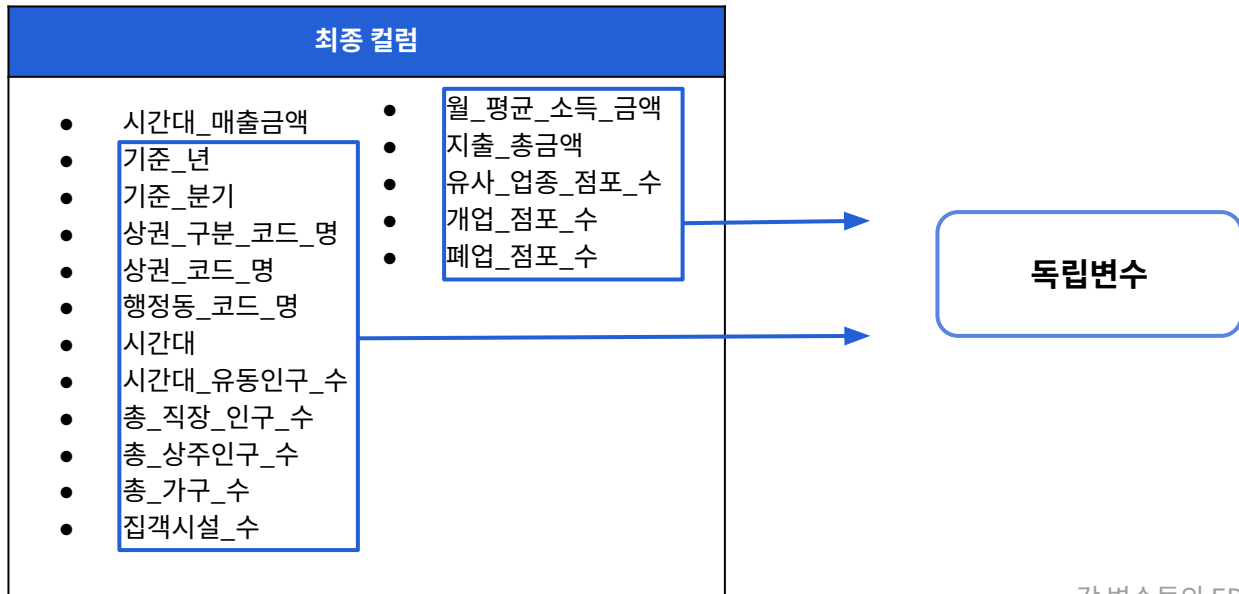
최종 컬럼

- 시간대\_매출금액
- 기준\_년
- 기준\_분기
- 상권\_구분\_코드\_명
- 상권\_코드\_명
- 행정동\_코드\_명
- 시간대
- 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

각 변수들의 EDA는  
부록에서 확인 가능

## 4. 통계적 탐색 분석

최종 컬럼 확인



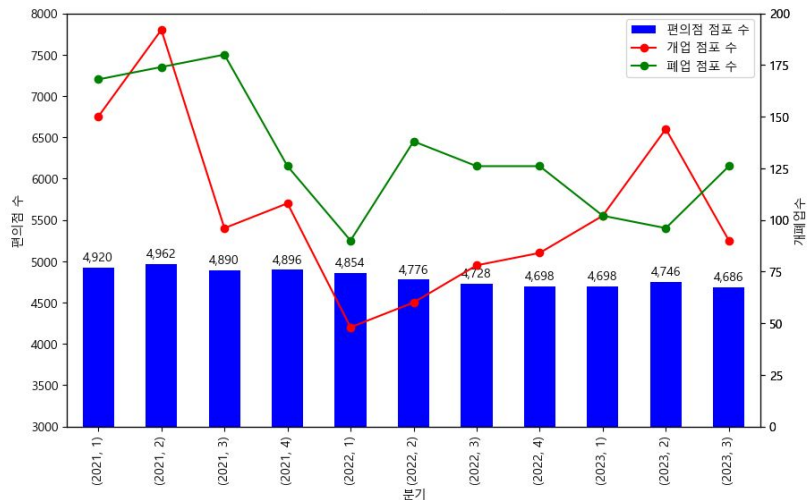
각 변수들의 EDA는  
부록에서 확인 가능

## 4. 통계적 탐색 분석

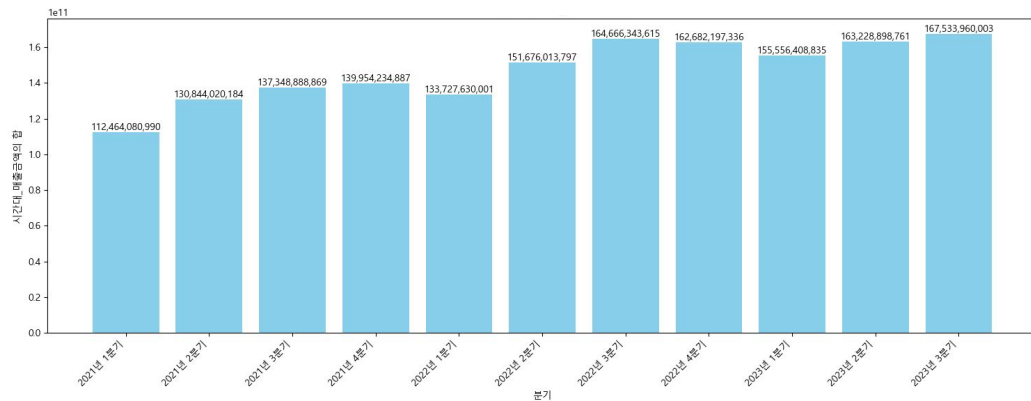
### EDA

- 강남구는 대부분 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 |



| 년도 및 분기별 매출금액 |

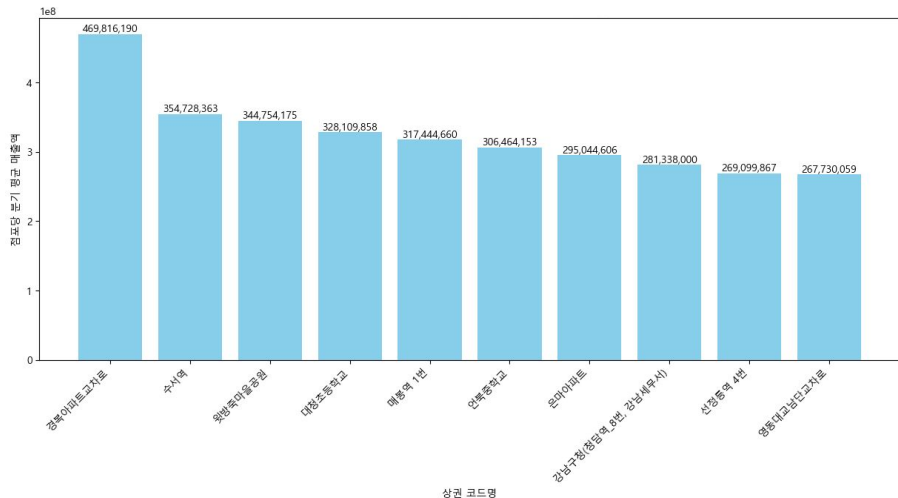


## 4. 통계적 탐색 분석

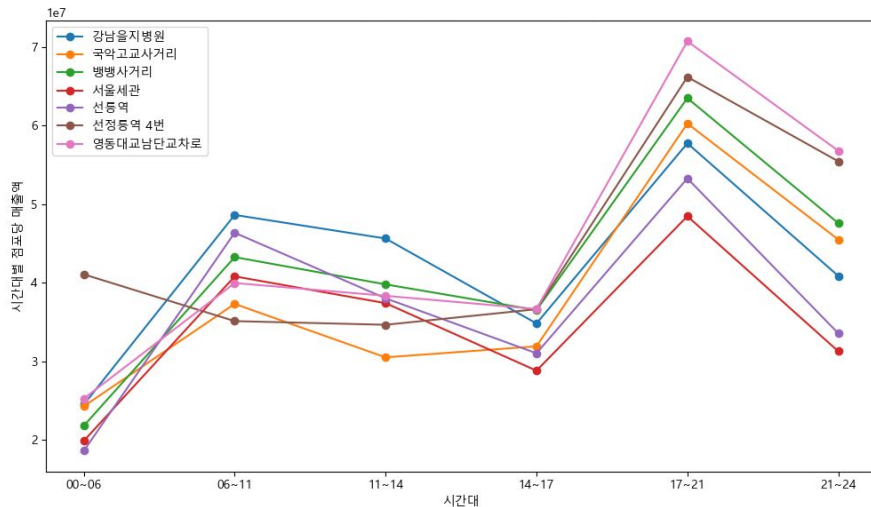
### EDA

- 상권에 따라 매출액이 다름
- 임의의 상권을 추출했을 때, 시간대별 매출 추이가 비슷하나, 상권별로 조금씩 다른 매출 추이를 보임
- 상권에 따라 시간대별 매출에 영향을 끼치는 요인이 다를 것으로 예상할 수 있음

| 2021년 ~ 2023년 3분기 상권별 평균 매출액 |



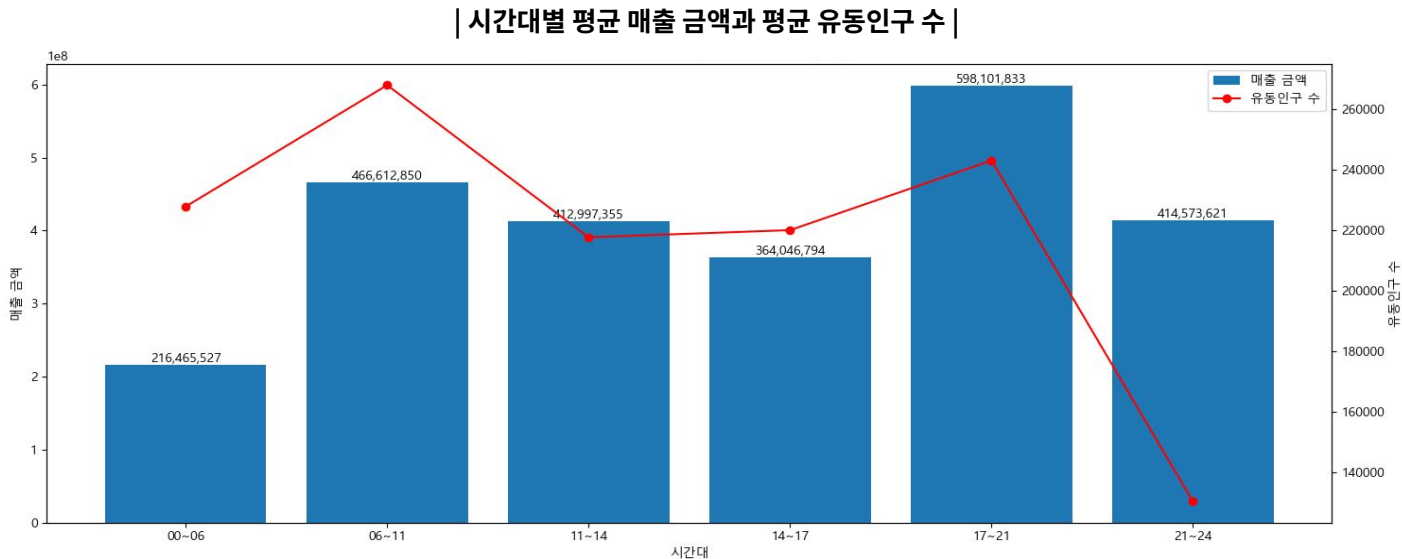
| 상권에 따른 시간대별 점포당 평균 매출액 |



## 4. 통계적 탐색 분석

### 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



## 4. 통계적 탐색 분석

### 선행연구 기반 컬럼 추가

- 선행연구에 기반하여, '편의점 밀도', '연령별 상주인구 비중', '연령별 직장인구 비중' 컬럼 추가

#### 기존 컬럼

- |              |              |
|--------------|--------------|
| • 기준_년       | • 월_평균_소득_금액 |
| • 기준_분기      | • 지출_총금액     |
| • 상권_구분_코드_명 | • 유사_업종_점포_수 |
| • 상권_코드_명    | • 개업_점포_수    |
| • 행정동_코드_명   | • 폐업_점포_수    |
| • 시간대        |              |
| • 시간대_매출금액   |              |
| • 시간대_유동인구_수 |              |
| • 총_직장_인구_수  |              |
| • 총_상주인구_수   |              |
| • 총_가구_수     |              |
| • 집객시설_수     |              |

#### 최종 컬럼

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

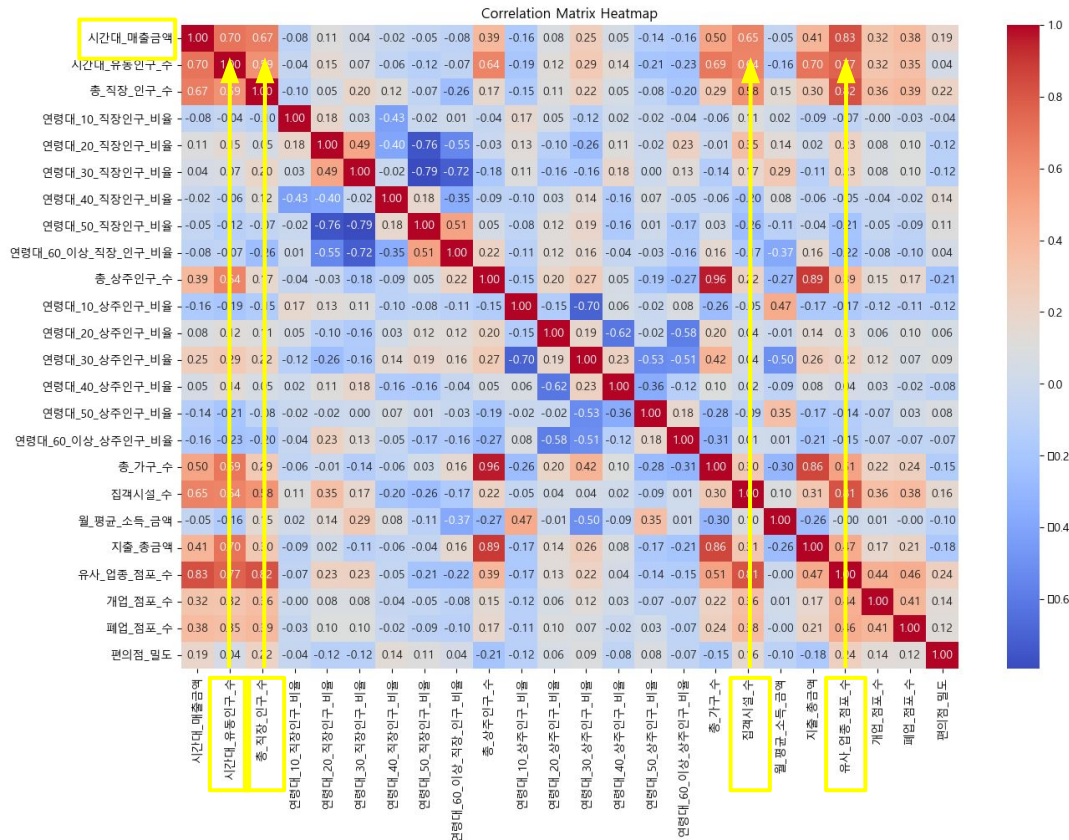


# 4. 통계적 탐색 분석

## 상관분석

- 유사\_업종\_점포\_수 - 0.83
- 시간대\_유동인구\_수 - 0.70
- 총\_직장\_인구\_수 - 0.67
- 집객시설\_수 - 0.65

=> 해당 변수들이 시간대\_매출금액에 비교적 많은 영향을 끼칠 것으로 예상



## 4. 통계적 탐색 분석

### 다중공선성

- 독립변수들 간의 VIF 확인
- 연령별\_직장인구\_비율, 연령별\_상주인구\_비율, 총\_상주인구\_수, 총\_가구\_수, 유사\_업종\_점포\_수의 VIF가 10이상으로 높음

=> 트리 기반의 모델을 활용하여 다중공선성 문제를 해결할 것으로 기대

	VIF Factor	features
0	4.457923	시간대_유동인구_수
1	5.303403	총_직장_인구_수
2	inf	연령대_10_직장인구_비율
3	inf	연령대_20_직장인구_비율
4	inf	연령대_30_직장인구_비율
5	inf	연령대_40_직장인구_비율
6	inf	연령대_50_직장인구_비율
7	inf	연령대_60_이상_직장_인구_비율
8	35.345023	총_상주인구_수

9	inf	연령대_10_상주인구_비율
10	inf	연령대_20_상주인구_비율
11	inf	연령대_30_상주인구_비율
12	inf	연령대_40_상주인구_비율
13	inf	연령대_50_상주인구_비율
14	inf	연령대_60_이상_상주인구_비율
15	32.874815	총_가구_수
16	5.411849	집객시설_수
17	2.586857	월_평균_소득_금액
18	6.784772	지출_총금액
19	15.814548	유사_업종_점포_수
20	1.357144	개업_점포_수
21	1.417630	폐업_점포_수
22	1.574667	편의점_밀도

# 목차

---

## 프로젝트 소개



- 팀원 구성
- 기획 배경
- 웹 서비스 소개
- 개발 환경
- 프로젝트 수행 기간

## 프로젝트 문서



- FLOW CHART
- WBS

## 데이터 탐색



- 데이터 수집
- 데이터 전처리

## 통계적 탐색 분석



- EDA
- 상관분석
- 다중공선성

## 모델링



- 모델 생성
- 모델 학습
- 모델 성능 검증
- 편의점 예상 매출 산출

## 서비스 및 문서



- 서비스 (Streamlit)
- 한계점 / 개선 사항
- 참고 자료
- 부록

# 5. 모델링

---

## 시계열 분석 (Time Series Analysis)

- 데이터  $y$ 의 현재와 과거 즉 **시간의 차이에 따라 나타나는 변화**를 분석 예) 주가 예측, 매출 예측, 기온 예측 등
- 일반적으로 추세(trend), 순환(cycle), 계절(seasonal) 성분, 불규칙(irregular) 또는 우연(random) 성분 등으로 구성되며, 이 성분들에 의해 변동된다고 가정
- 데이터 변동 패턴 관찰 → 추세, 계절, 불규칙 성분으로 세분화 → 미래 관측값 예측(지수 모델링 기법/ARIMA 기법)

## 5. 모델링

### 시계열 분석 (Time Series Analysis)

**추세(trend):** 관측값의 전반적 상승 / 하락 경향을 나타내는 성분

**순환(cycle):** 주기적인 변화를 가지긴 하나, 계절에 의한 것이 아닌 주기가 긴 경우를 나타내는 성분

**계절(seasonal):** 특정 주기에 따라 일정하게 변동할 때 설명되는 요인을 나타내는 성분

**불규칙(irregular):** 특정한 규칙으로 설명할 수 없는 오차를 나타내는 성분

**우연(random):** 시간에 따른 규칙적인 움직임과 무관하게 랜덤한 원인을 나타내는 성분

# 5. 모델링

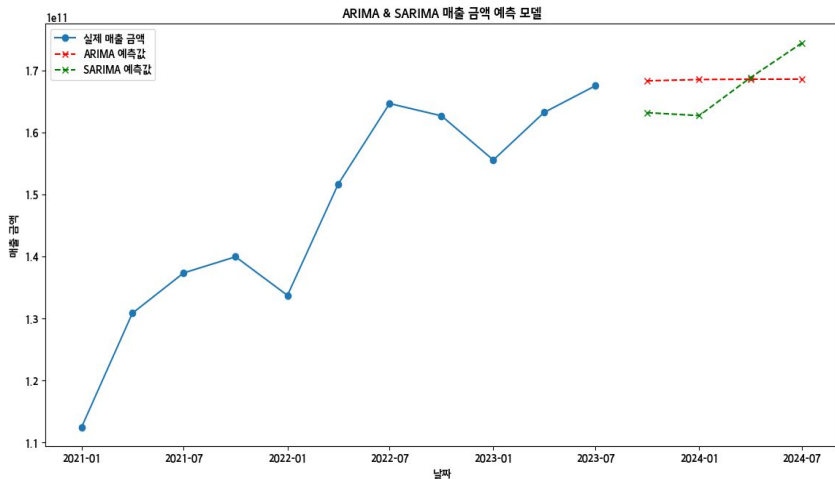
## ARIMA & SARIMA

### ARIMA (AutoRegressive Integrated Moving Average)

- 비계절적 데이터의 추세와 변동성을 모델링하는 데 적합
- 시계열의 비정상성을 처리하기 위해 **차분(differencing)**을 사용

### SARIMA (Seasonal ARIMA)

- ARIMA의 확장 버전으로, 계절성 패턴을 추가적으로 모델링할 수 있음
- **계절적 차분(seasonal differencing)**을 통해 시계열의 계절성을 처리



### 모델 성능

ARIMA - MAE 21,367,509,120.59 / RMSE 43,353,668,254.71

SARIMA - MAE 43,774,410,318.39 / RMSE 77,763,589,777.93

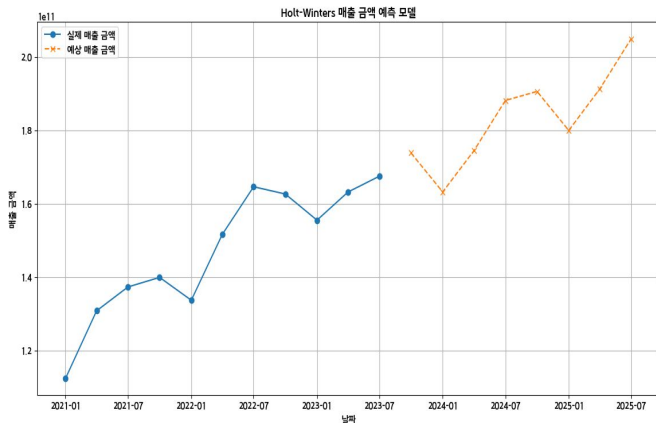
ARIMA 모델이 SARIMA 모델보다 낮은 평균 절대 오차(MAE)와 평균 제곱근 오차 (RMSE)를 보여줌

오차범위가 상당히 넓으므로 다른 모델 고려

# 5. 모델링

## Holt-Winters 모형 시계열 예측

- 시계열 데이터를 분석하고 예측하는 데 사용되는 통계적 기법
- 시계열 데이터에 존재할 수 있는 세 가지 주요 구성 요소인 **레벨(level)**, **추세(trend)**, **계절성(seasonality)**을 모두 고려하여 미래의 데이터 포인트를 예측
- 주기적인 패턴이나 계절적 변동성이 중요한 역할을 하는 시계열 데이터의 단기 예측에 유용



### 모델 성능

Mean Absolute Error (MAE) : 7,317,570,194.14

Root Mean Squared Error (RMSE) : 8,237,463,143.96

오차범위가 상당히 넓으므로 다른 모델 고려

- 편의성을 위해 각 분기의 첫 번째 달을 가정
- Holt-Winters 모형 적합/미래 예측

## 5. 모델링

### 최종 모델 선택

- 시계열, RandomForest, XGBoost, LightGBM 등 다양한 머신러닝 알고리즘을 적용한 결과, **LightGBM** 모델이 가장 좋은 성능을 보임

모델 알고리즘	데이터 재가공	Feature Engineering	교차 검증	하이퍼 파라미터 튜닝	MAE	MSE	RMSE	결정 계수
LGBM	범주형 변수 더미화 / 숫자형 변수 Standard Scaler 종속변수 Box-cox변환	연령별 직장인구 비율 연령별 상주인구 비율 편의점 밀도		num_leaves': [25, 30, 35], 'learning_rate': [0.12, 0.13, 0.14], 'n_estimators': [375, 400, 425]	2,136,328,813,899,580	26930386.26	46220437.19	<b>0.9885977 983</b>



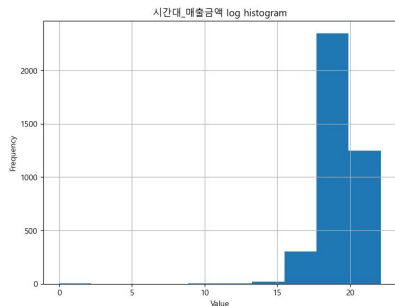
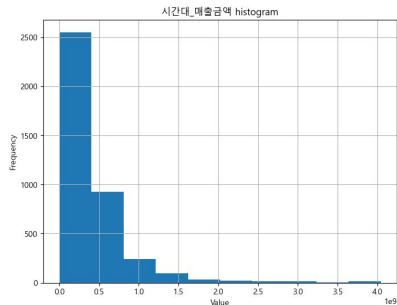
# 5. 모델링

## 종속변수 Scaling

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

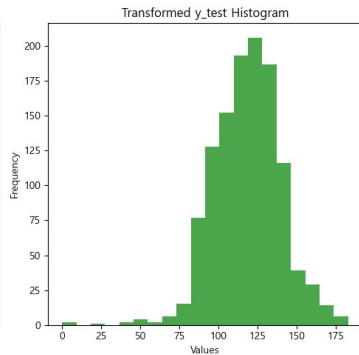
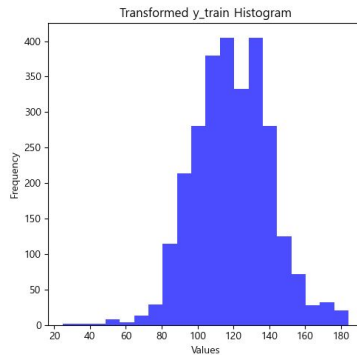
### Log 변환

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



### Box-cox 변환

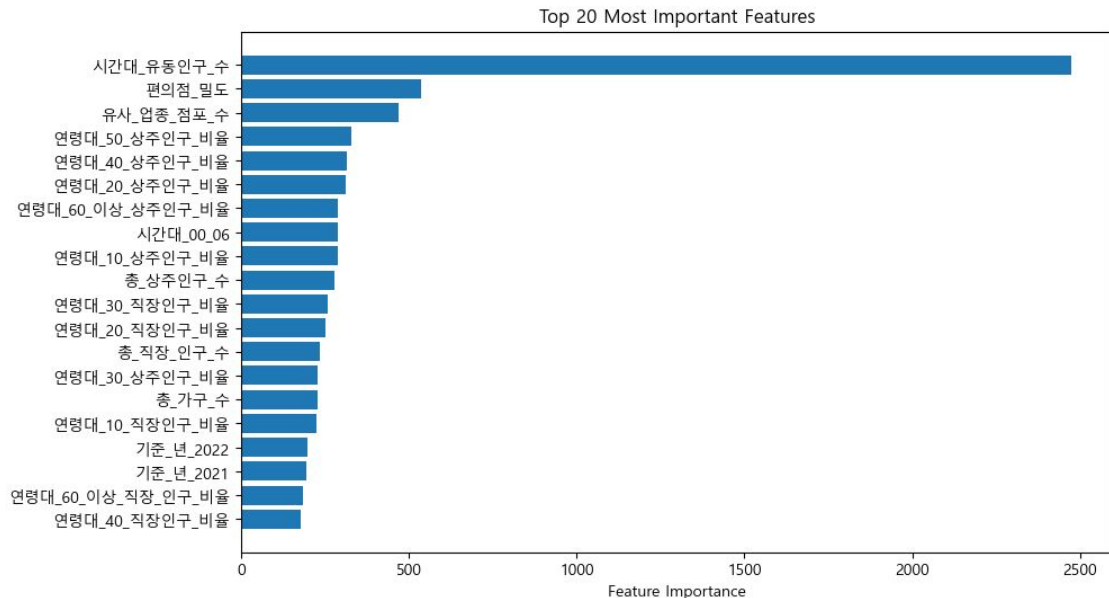
- 분포의 skew를 해결해 줄 수 있는 Box-cox 변환 시도
- train 데이터 세트와 test 데이터 세트의 종속변수가 정규분포와 유사한 형태로 변환 됨



# 5. 모델링

## Feature Importance

- 상관분석에서 높은 상관 관계를 보였던 [시간대\_유동인구\_수]가 가장 중요한 피처임을 확인

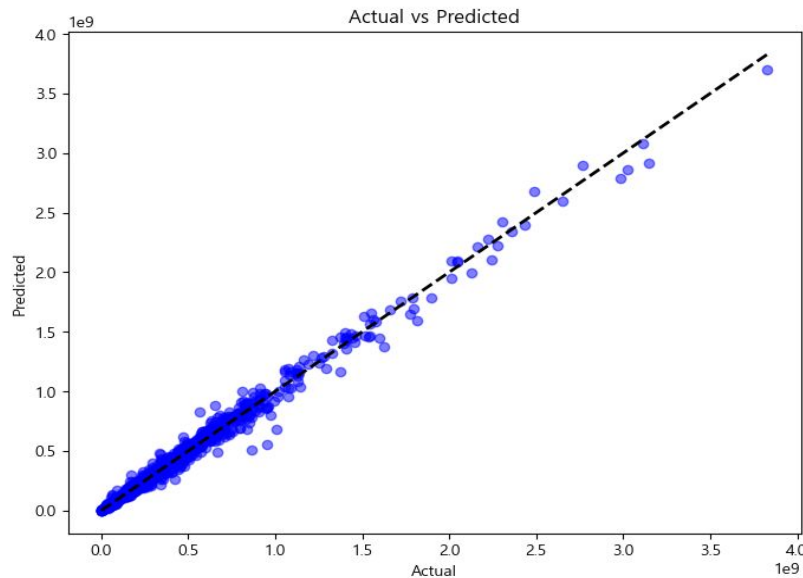


# 5. 모델링

## 실제값 vs 예측값

- 실제값과 예측값을 시각화로 나타냈을 때,  $y=x$  을 기준으로 거의 균등하게 분포함을 확인

=> 이는 모델의 데이터를 잘 예측한다는 것을 의미



# 목차

---

## 프로젝트 소개



- 팀원 구성
- 기획 배경
- 웹 서비스 소개
- 개발 환경
- 프로젝트 수행 기간

## 프로젝트 문서



- FLOW CHART
- WBS

## 데이터 탐색



- 데이터 수집
- 데이터 전처리

## 통계적 탐색 분석



- EDA
- 상관분석
- 다중공선성

## 모델링



- 모델 생성
- 모델 학습
- 모델 성능 검증
- 편의점 예상 매출 산출

## 서비스 및 문서



- 서비스 (Streamlit)
- 한계점 / 개선 사항
- 참고 자료
- 부록

# 6. 서비스 및 문서

## 서비스 (Streamlit)

### 메뉴 구성

#### 메뉴

메뉴를 선택  
하세요:

홈

강남구 편의점 분포  
현황

강남구 편의점 매출  
현황

매출 현황 순위

매출 예측 모델링

### Streamlit - HOME

#### 강남구 편의점 매출 예측



이 앱은 편의점 예비  
창업자들을 위한 강남구  
지역의 편의점 시간대별  
매출 예측 서비스입니다.

이 앱이 여러분들께  
조금이나마 도움이 되기를  
바랍니다.



### Streamlit - EDA

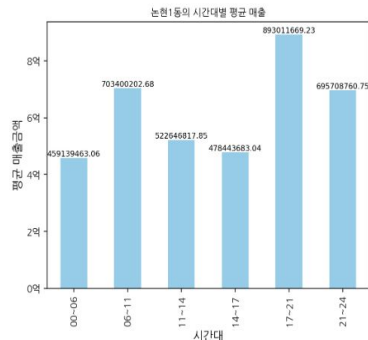
#### 강남구 편의점 매출 현황

발장동 코드명 선택:

논현1동

선택된 발장동 코드명: 논현1동

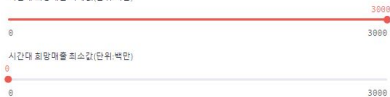
논현1동의 시간대별 평균 매출



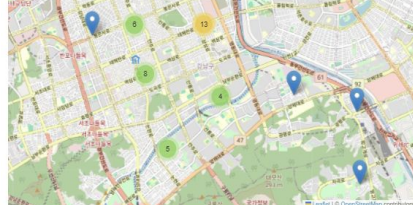
#### 강남구 편의점 예상매출 종합

예상 매출 상한수치: 예상 매출 하한수치

시간대 최상매출 최대값(단위: 백만)



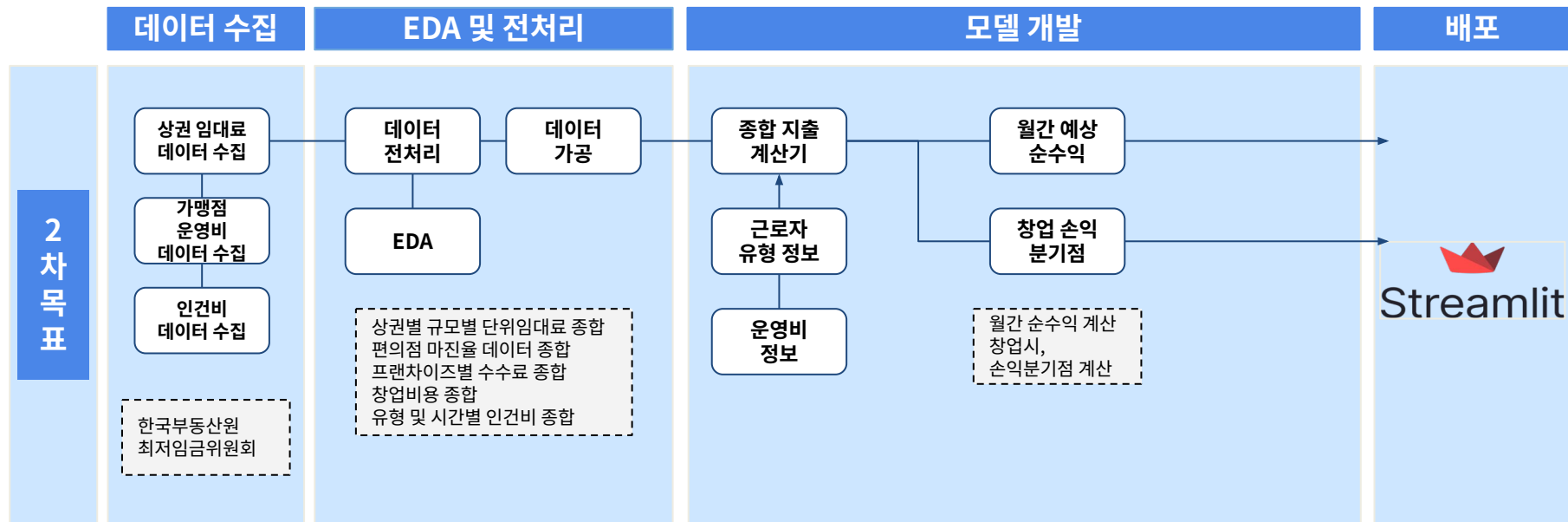
시간대 최하매출 최소값(단위: 백만)



## 6. 서비스 및 문서

한계점 / 개선 사항

- 인건비 및 가맹점 운영비를 계산하여 월 순수익 예측 모델을 구현하려고 하였으나, 시간적 이유로 구현하지 못함



## 6. 서비스 및 문서

---

### 참고 문헌

#### • 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…, 더스쿠프, 2023.02.14  
<https://www.thescoop.co.kr/news/articleView.html?idxno=56799>
- 3) 이진원, 홈플러스, 예비 편의점주에 예상매출 뺑뚱기, 시민일보, 2017.11.05  
<https://www.siminilbo.co.kr/news/articleView.html?idxno=537797>

## 6. 서비스 및 문서

부록 - 서비스

페이지 이동 사이드 바

- 메뉴를 선택하여 원하는  
페이지로 이동

메뉴

☞ 메뉴를 선택  
하세요:

🏠 홈

📄 강남구 편의점 분포  
현황

📈 강남구 편의점 매출  
현황

📊 매출 현황 순위

📊 매출 예측 모델링

강남구 편의점 매출 예측 



이 앱은 편의점 예비  
창업자들을 위한 강남구  
지역의 편의점 시간대별  
매출 예측 서비스입니다.

이 앱이 여러분들께  
조금이나마 도움이 되기를  
바랍니다.





## 6. 서비스 및 문서

부록 - 서비스

페이지 이동 사이드 바

- 메뉴를 선택하여 원하는  
페이지로 이동

홈 화면

- 서비스의 목적 확인

메뉴

☞ 메뉴를 선택  
하세요:

🏠 홈

📄 강남구 편의점 분포  
현황

📊 강남구 편의점 매출  
현황

📈 매출 현황 순위

📊 매출 예측 모델링

강남구 편의점 매출 예측 



이 앱은 편의점 예비  
창업자들을 위한 강남구  
지역의 편의점 시간대별  
매출 예측 서비스입니다.

이 앱이 여러분들께  
조금이나마 도움이 되기를  
바랍니다.





## 6. 서비스 및 문서

부록 - 서비스

### 지도 시각화

- 원하는 상권의  
분포 현황 확인

### 세부 정보 확인

- 사용자가 선택한 상권의  
평균 매출 금액 확인

메뉴

☞ 메뉴를 선택  
하세요:

🏠 홈

📊 강남구 편의점 분  
포 현황

📍 강남구 편의점 매출  
현황

📊 매출 현황 순위

📊 매출 예측 모델링

### 강남구 편의점 분포 현황 🗺️ (2021년 1분기 ~ 2023년 3분기)

궁금한 상권을 선택하세요 📍



## 6. 서비스 및 문서

부록 - 서비스

선택 박스

- 원하는 행정동 선택

메뉴

☞ 메뉴를 선택  
하세요:

☰ 홈

☰ 강남구 편의점 분포  
현황

📈 강남구 편의점 매  
출 현황

📊 매출 현황 순위

🔮 매출 예측 모델링

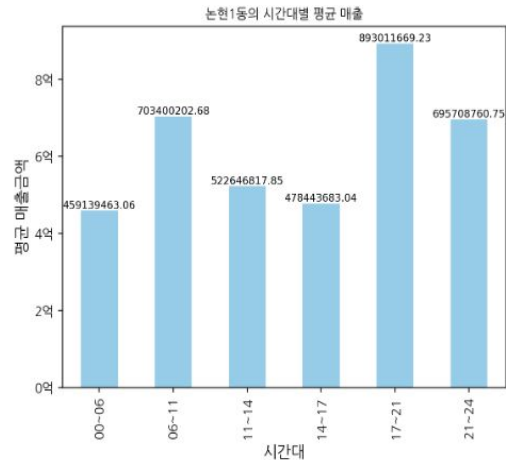
강남구 편의점 매출 현황 📊

행정동 코드명 선택:

논현1동

선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출



## 6. 서비스 및 문서

부록 - 서비스

### 셀렉트 박스

- 원하는 행정동 선택

### 막대 그래프

- 사용자가 선택한 행정동의  
시간대별 평균 매출 확인

### 메뉴

☞ 메뉴를 선택  
하세요:

☰ 홈

☰ 강남구 편의점 분포  
현황

📈 강남구 편의점 매  
출 현황

📊 매출 현황 순위

📈 매출 예측 모델링

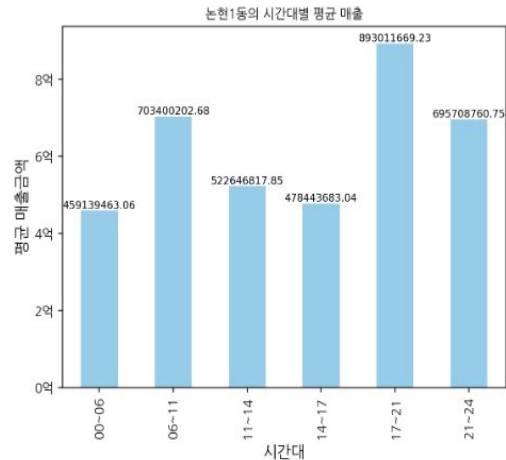
### 강남구 편의점 매출 현황

행정동 코드명 선택:

논현1동

선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출



## 6. 서비스 및 문서

### 부록 - 서비스

#### 선택 박스

- 위에서 선택한 행정동에  
해당하는 상권 선택

#### 메뉴

☞ 메뉴를 선택  
하세요:

☰ 홈

☐ 강남구 편의점 분포  
현황

📍 강남구 편의점 매  
출 현황

📊 매출 현황 순위

📈 매출 예측 모델링

행정동 코드명 선택:

논현1동

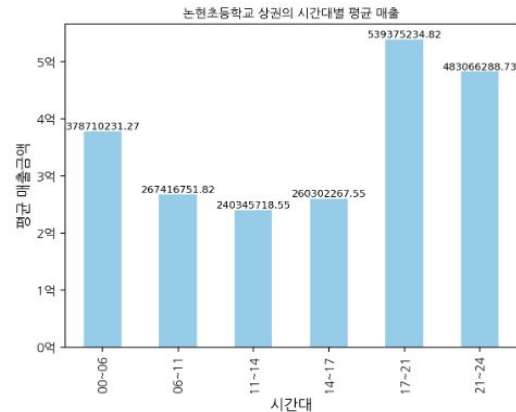
선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출

논현1동에 대한 상권 코드명 선택:

논현초등학교

논현초등학교 상권의 시간대별 평균 매출



## 6. 서비스 및 문서

### 부록 - 서비스

#### 선택 박스

- 위에서 선택한 행정동에  
해당하는 상권 선택

#### 막대 그래프

- 사용자가 선택한 상권의  
시간대별 평균 매출 확인

#### 메뉴

☞ 메뉴를 선택  
하세요:

☰ 홈

☐ 강남구 편의점 분포  
현황

📍 강남구 편의점 매  
출 현황

📊 매출 현황 순위

📈 매출 예측 모델링

행정동 코드명 선택:

논현1동

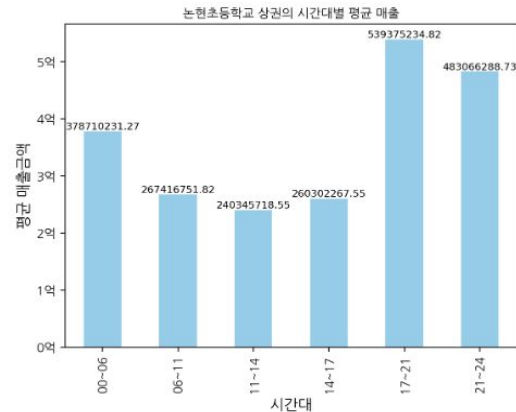
선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출

논현1동에 대한 상권 코드명 선택:

논현초등학교

논현초등학교 상권의 시간대별 평균 매출



## 6. 서비스 및 문서

부록 - 서비스

라디오 버튼

- 원하는 시간대 선택

메뉴

☰ 메뉴를 선택  
하세요:

☰ 홈

☰ 강남구 편의점 분포  
현황

☰ 강남구 편의점 매출  
현황

☑ 매출 현황 순위

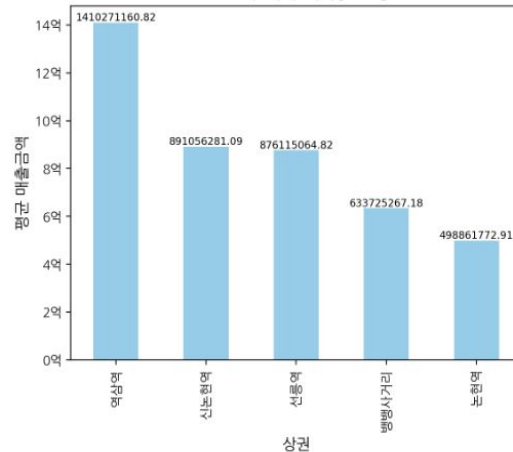
☑ 매출 예측 모델링

매출 현황 순위 💰

시간대를 선택하세요:

- ☒ 00:00 ~ 06:00
- ☐ 06:00 ~ 11:00
- ☐ 11:00 ~ 14:00
- ☐ 14:00 ~ 17:00
- ☐ 17:00 ~ 21:00
- ☐ 21:00 ~ 24:00

00:00 ~ 06:00 시간대 매출이 가장 높은 상권 TOP5





## 6. 서비스 및 문서

부록 - 서비스

라디오 버튼

- 원하는 시간대 선택

막대 그래프

- 사용자가 선택한 시간대에  
가장 높은 매출을 기록한 상권  
TOP5의 시간대별 매출 확인

메뉴

☞ 메뉴를 선택  
하세요:

☰ 홈

☐ 강남구 편의점 분포  
현황

📈 강남구 편의점 매출  
현황

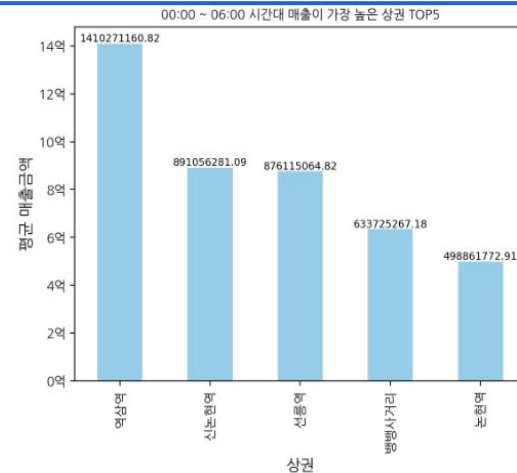
📊 매출 현황 순위

📈 매출 예측 모델링

매출 현황 순위 💰

시간대를 선택하세요:

- ☒ 00:00 ~ 06:00
- ☐ 06:00 ~ 11:00
- ☐ 11:00 ~ 14:00
- ☐ 14:00 ~ 17:00
- ☐ 17:00 ~ 21:00
- ☐ 21:00 ~ 24:00



## 6. 서비스 및 문서

### 부록 - 서비스

라디오 버튼

- 원하는 유형 선택

메뉴

☞ 메뉴를 선택  
하세요:

☰ 홈

☰ 강남구 편의점 분포  
현황

☰ 강남구 편의점 매출  
현황

☰ 매출 현황 순위

🔴 매출 예측 모델링

유형을 선택하세요

☒ 선택 상권 예상 매출

☐ 예상 매출 종합

☐ 예상 매출 상권 추천

행정동 선택

삼성1동

상권 선택

강남 마이스 관광특구

분기 선택

1분기

### 강남구 편의점 예상매출 종합 🧠

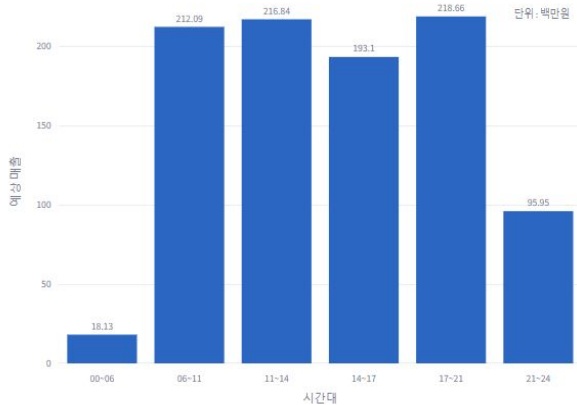
시간대별 예상매출 비교    시간대별 예상매출 상권 지도

행정동: 삼성1동

상권: 강남 마이스 관광특구

1분기 시간대별 예상매출 비교 표: (단위:백만원)

1분기 시간대별 예상 매출 비교



## 6. 서비스 및 문서

### 부록 - 서비스

#### 라디오 버튼

- 원하는 유형 선택

#### 셀렉트 박스

- 원하는 행정동, 상권, 분기 선택

#### 메뉴

☞ 메뉴를 선택  
하세요:

☰ 톨

☐ 강남구 편의점 분포  
현황

☐ 강남구 편의점 매출  
현황

☐ 매출 현황 순위

🔴 매출 예측 모델링

유형을 선택하세요

- ☒ 선택 상권 예상 매출
- ☐ 예상 매출 종합
- ☐ 예상 매출 상권 추천

행정동 선택

삼성1동

상권 선택

강남 마이스 관광특구

분기 선택

1분기

## 강남구 편의점 예상매출 종합 🧠

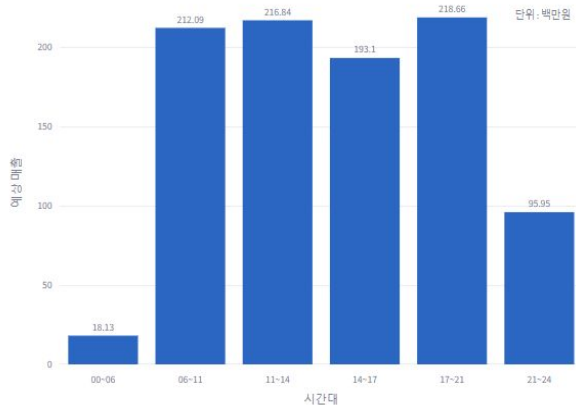
시간대별 예상매출 비교    시간대별 예상매출 상권 지도

행정동: 삼성1동

상권: 강남 마이스 관광특구

1분기 시간대별 예상매출 비교 표: (단위: 백만원)

1분기 시간대별 예상 매출 비교



## 6. 서비스 및 문서

### 부록 - 서비스

#### 라디오 버튼

- 원하는 유형 선택

#### 셀렉트 박스

- 원하는 행정동, 상권, 분기 선택

#### 막대 그래프

- 사용자 선택 조건에 따른  
시간대별 예상 매출 확인

#### 메뉴

☞ 메뉴를 선택  
하세요:

☰ 톨

☐ 강남구 편의점 본포  
현황

☐ 강남구 편의점 매출  
현황

☐ 매출 현황 순위

🔴 매출 예측 모델링

유클을 선택하세요

- ☒ 선택 상권 예상 매출
- ☐ 예상 매출 종합
- ☐ 예상 매출 상권 추천

행정동 선택

삼성1동

상권 선택

강남 마이스 관광특구

분기 선택

1분기

### 강남구 편의점 예상매출 종합 🧠

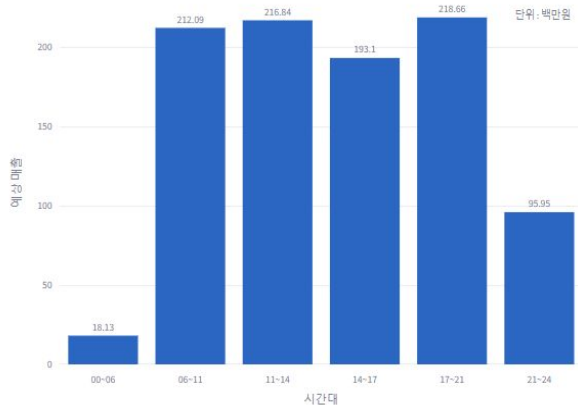
시간대별 예상매출 비교    시간대별 예상매출 상권 지도

행정동: 삼성1동

상권: 강남 마이스 관광특구

1분기 시간대별 예상매출 비교 표: (단위: 백만원)

1분기 시간대별 예상매출 비교



## 6. 서비스 및 문서

### 부록 - 서비스

탭

- 원하는 탭으로 이동

메뉴

☑ 메뉴를 선택  
하세요:

간

☐ 강남구 편의점 분포  
현황

☐ 강남구 편의점 매출  
현황

☐ 매출 현황 순위

🔴 매출 예측 모델링

유틸리티 선택하세요

- ☒ 선택 상권 예상 매출
- ☐ 예상 매출 종합
- ☐ 예상 매출 상권 추천

행정동 선택

삼성1동

상권 선택

강남 마이스관광특구

분기 선택

1분기

## 강남구 편의점 예상매출 종합 🗨️

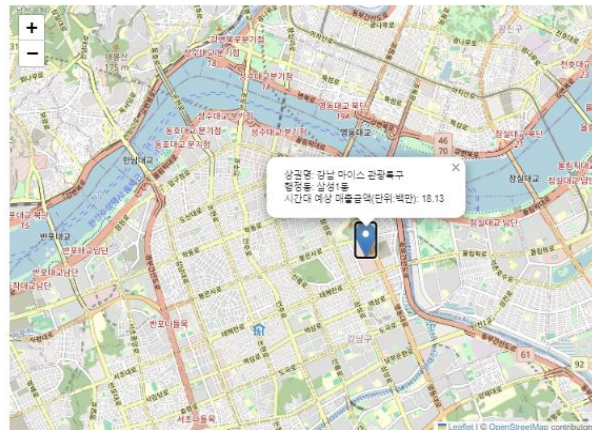
시간대별 예상매출 비교 시간대별 예상매출 상권 지도

시간대 선택

00시~06시

1분기, 00시~06시

삼성1동 [강남 마이스관광특구] 예상매출



## 6. 서비스 및 문서

### 부록 - 서비스

#### 탭

- 원하는 탭으로 이동

#### 셀렉트 박스

- 원하는 시간대 선택

#### 메뉴

☞ 메뉴를 선택  
하세요:

간식

☞ 강남구 편의점 분포  
현황

☞ 강남구 편의점 매출  
현황

☞ 매출전환 순위

🔴 매출 예측 모델링

유틸리티 선택하세요

- ☒ 선택 상권 예상 매출
- ☐ 예상 매출 종합
- ☐ 예상 매출 상권 추천

행정동 선택

삼성1동

상권 선택

강남 마이스관광특구

분기 선택

1분기

## 강남구 편의점 예상매출 종합 🗨️

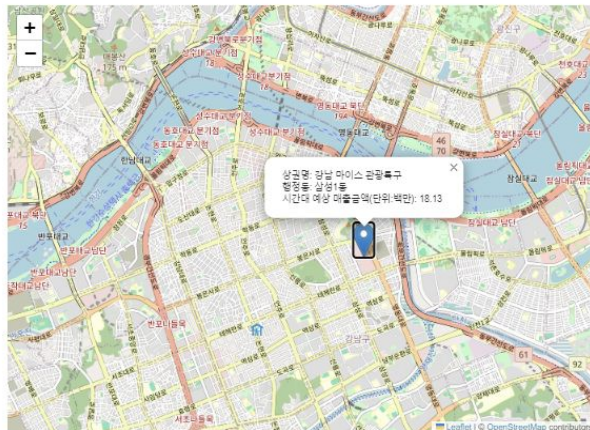
시간대별 예상매출 비교 시간대별 예상매출 상권 지도

시간대 선택

00시~06시

1분기, 00시~06시

삼성1동 [강남 마이스관광특구] 예상매출



## 6. 서비스 및 문서

### 부록 - 서비스

#### 탭

- 원하는 탭으로 이동

#### 셀렉트 박스

- 원하는 시간대 선택

#### 세부 정보 확인

- 사용자가 선택한 시간대의  
예상 매출 금액 확인

#### 메뉴

☞ 메뉴를 선택  
하세요:

간식

☞ 강남구 편의점 분포  
현황

☞ 강남구 편의점 매출  
현황

☞ 매출현황 순위

🔴 매출 예측 모델링

유틸리티 선택하세요

- ☒ 선택 상권 예상 매출
- ☐ 예상 매출 종합
- ☐ 예상 매출 상권 추천

행정동 선택

삼성1동

상권 선택

강남 마이스관광특구

분기 선택

1분기

## 강남구 편의점 예상매출 종합 🗨️

시간대별 예상매출 비교 시간대별 예상매출 상권 지도

시간대 선택

00시~06시

1분기, 00시~06시

삼성1동 [강남 마이스관광특구] 예상매출







## 6. 서비스 및 문서

### 부록 - 서비스

#### 탭

- 상권별 시간대 / 년 / 월  
예상 매출 각각 확인 가능

#### 세부 정보 확인

- 상권별 시간대 / 년 / 월  
예상 매출을 지도 시각화를  
통해 각각 확인 가능

#### 메뉴

☞ 메뉴를 선택  
하세요:

☰ 홈

☰ 강남구 편의점 본포  
현황

☑ 강남구 편의점 매출  
현황

☑ 매출전환순위

🔴 매출 예측 모델링

유행을 선택하세요

☐ 선택 상권 예상 매출

☒ 예상 매출 종합

☐ 예상 매출 상권 추천

### 강남구 편의점 예상매출 종합 📍

분기별, 시간대별 예상매출종합 예상 연 매출종합 예상 월 매출종합

예상 연 매출종합



## 6. 서비스 및 문서

### 부록 - 서비스

#### 슬라이더

- 사용자가 희망하는 매출의  
최소값과 최대값을 입력

#### 메뉴

☞ 메뉴를  
선택하세요:

☰ 홈

☒ 강남구 편의점  
분포 현황

☒ 강남구 편의점  
매출 현황

☒ 매출 현황 순  
위

● 매출 예측 모  
델링

유형을 선택하세요

☐ 선택 상권 예상 매출

☐ 예상 매출 종합

☒ 예상 매출 상권 추천

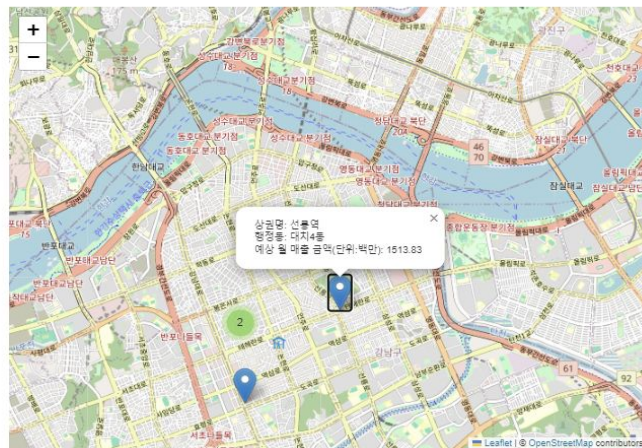
#### 강남구 편의점 예상매출 종합 🗨️

예상 매출 상권 추천 예상 순 수익

월 희망매출 최대값(단위:백만)



월 희망매출 최소값(단위:백만)



## 6. 서비스 및 문서

### 부록 - 서비스

#### 슬라이더

- 사용자가 희망하는 매출의  
최소값과 최대값을 입력

#### 세부 정보 확인

- 사용자 입력값을 만족하는  
상권의 정보 지도에 표시

#### 메뉴

☞ 메뉴를  
선택하세요:

☰ 홈

☒ 강남구 편의점  
분포 현황

☒ 강남구 편의점  
매출 현황

☒ 매출 현황 순  
위

● 매출 예측 모  
델링

유형을 선택하세요

☐ 선택 상권 예상 매출

☐ 예상 매출 종합

☒ 예상 매출 상권 추천

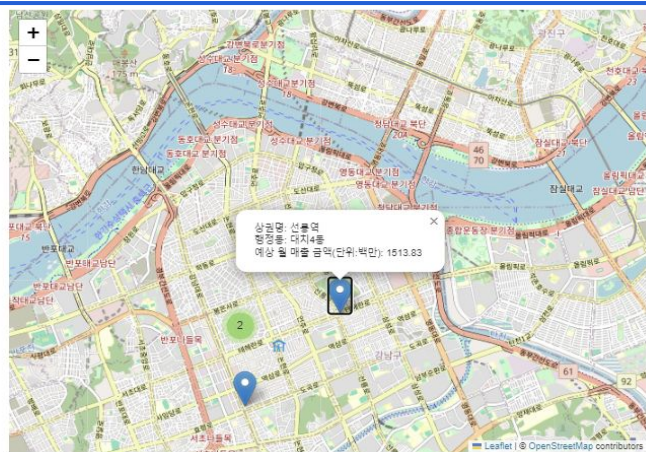
### 강남구 편의점 예상매출 종합 🗨

예상 매출 상권 추천 예상 순 수익

월 희망매출 최대값(단위:백만)



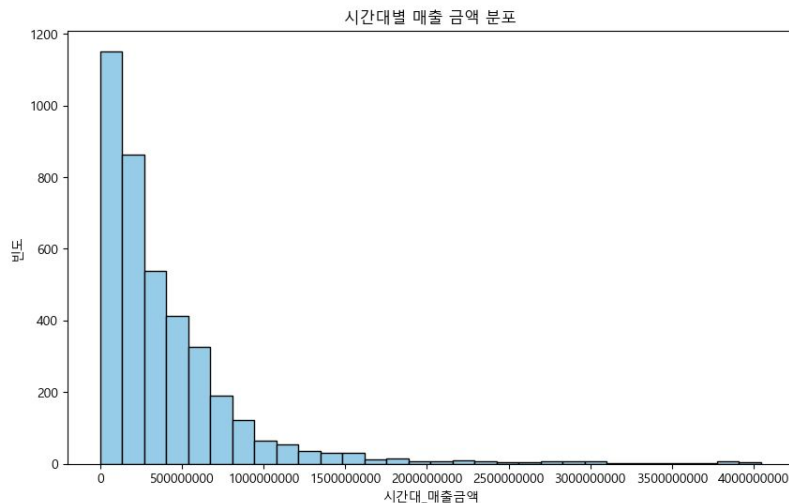
월 희망매출 최소값(단위:백만)



## 6. 서비스 및 문서

부록 - EDA

종속변수 - 연속형



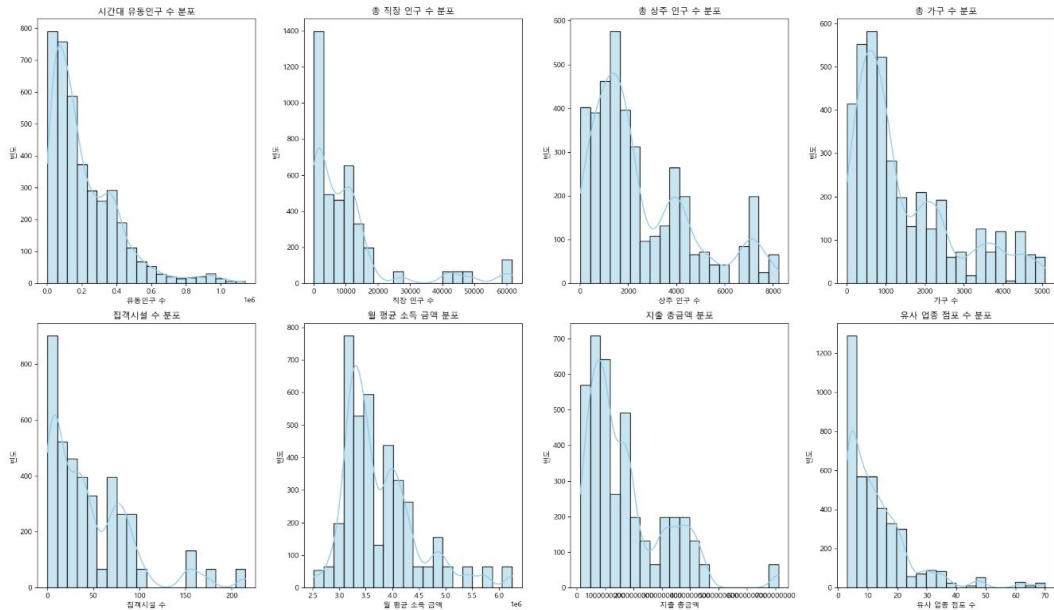
최종 컬럼

- 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

# 6. 서비스 및 문서

## 부록 - EDA

독립변수 - 연속형



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

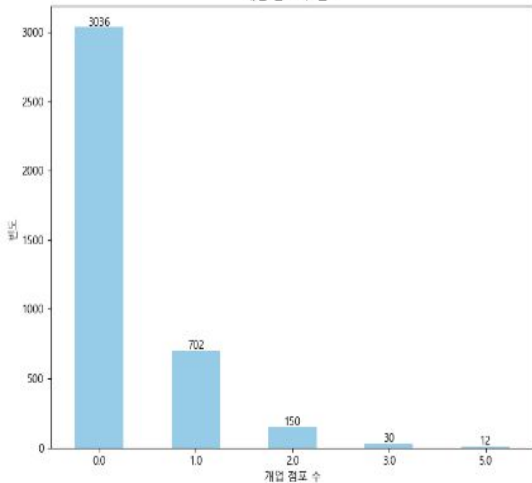
- 월\_평균\_소득\_금액
- 지출\_총금액
- 유사\_업종\_점포\_수
- 개업\_점포\_수
- 폐업\_점포\_수

## 6. 서비스 및 문서

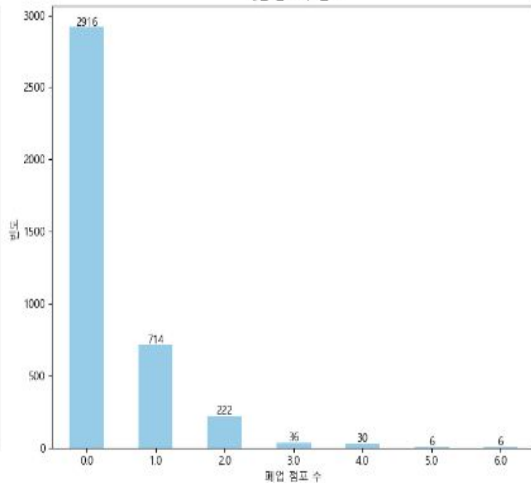
부록 - EDA

독립변수 - 연속형  
But, 범주형 변수의 분포와 유사함

개업 점포 수 분포



폐업 점포 수 분포



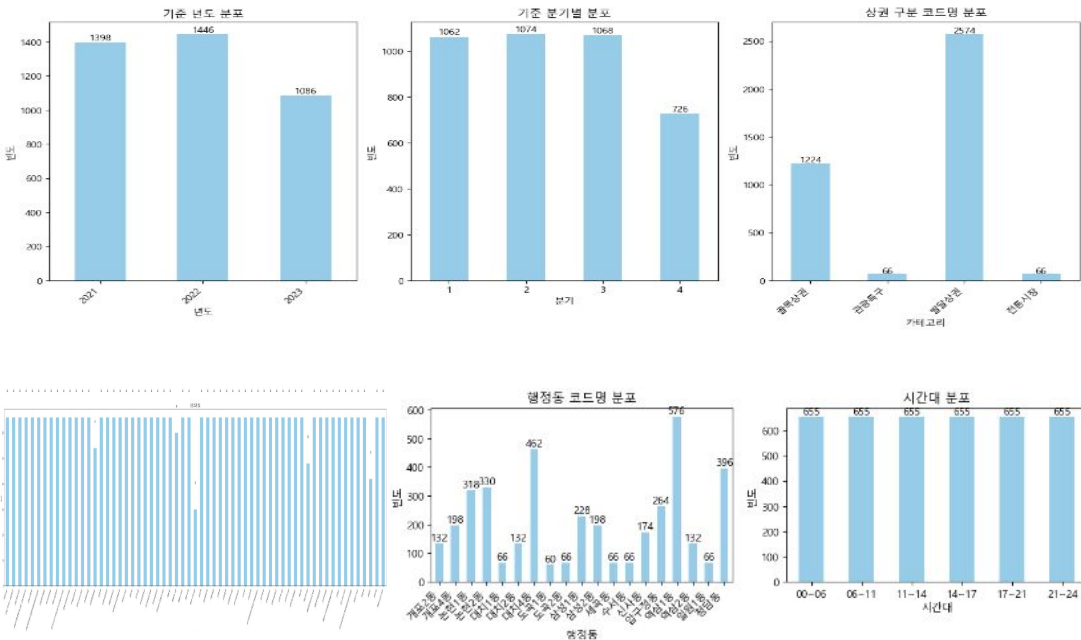
### 최종 컬럼

- 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

# 6. 서비스 및 문서

## 부록 - EDA

### 독립변수 - 범주형



### 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

## 6. 서비스 및 문서

### 부록 - 모델링

모델 알고리즘	데이터 재가공	Feature Engineering	교차 검증	하이퍼 파라미터 튜닝	MAE	MSE	RMSE	결정 계수
RandomForest	StandardScaler	연령별 직장인구 비율					67898424.45	<b>0.9648454212</b>
RandomForest	StandardScaler	인구수 대비 가구수 비율					66123873.23	<b>0.9754612381</b>
RandomForest	OneHotEncoder	연령별 직장인구 비율					66731521.15	<b>0.9648153123</b>
RandomForest	OneHotEncoder	인구수 대비 가구수 비율					66197315.94	<b>0.9618132176</b>



## 6. 서비스 및 문서

### 부록 - 모델링

모델 알고리즘	데이터 재가공	Feature Engineering	교차 검증	하이퍼 파라미터 튜닝	MAE	MSE	RMSE	결정 계수
RandomForest	StandardScaler: 수치형 데이터 OneHotEncoder: 범주형 데이터	인구 대비 가구 수의 비율	KFold				69963979.84	0.9798526552
RandomForest	StandardScaler: 수치형 데이터 OneHotEncoder: 범주형 데이터	인구 대비 가구 수의 비율	KFold	{'n_estimators': 500, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': 50, 'bootstrap': True}			69628037.42	0.9800303186
RandomForest + GradientBoosting	StandardScaler: 수치형 데이터 OneHotEncoder: 범주형 데이터	StandardScaler: 수치형 데이터 OneHotEncoder: 범주형 데이터						0.9773915991

## 6. 서비스 및 문서

### 부록 - 모델링

모델 알고리즘	데이터 재가공	Feature Engineering	교차 검증	하이퍼 파라미터 튜닝	MAE	MSE	RMSE	결정 계수
XGBoost	StandardScaler: 수치형 데이터 OneHotEncoder: 범주형 데이터	인구 대비 가구 수의 비율, 주말 대비 평일 유동인구 비율	KFold				60577427.09	<b>0.9849553596</b>
XGBoost	StandardScaler: 수치형 데이터 OneHotEncoder: 범주형 데이터	인구 대비 가구 수의 비율, 주말 대비 평일 유동인구 비율	KFold	{'subsample': 0.6, 'n_estimators': 500, 'min_child_weight': 1, 'max_depth': 7, 'learning_rate': 0.05, 'colsample_bytree': 0.8}			57367659.29	<b>0.9864766996</b>
XGBoost	시간대_매출금액', '평일_유동인구_수_평균', '주말_유동인구_수_평균', '총_상주인구_수', '상권_코드_명', '행정동_코드_명' 제거	인구 대비 가구 수의 비율					75272170.99	<b>0.9812817025</b>

## 6. 서비스 및 문서

### 부록 - 모델링

모델 알고리즘	데이터 재가공	Feature Engineering	교차 검증	하이퍼 파라미터 튜닝	MAE	MSE	RMSE	결정 계수
LGBM	범주형 변수 더미화 / 숫자형 변수 Standard Scaler 종속변수 Box-cox변환	연령별 직장인구 비율 연령별 상주인구 비율 편의점 밀도		num_leaves=31, learning_rate=0.1, n_estimators=100	3,518,915,356, 168,770	33206245.03	59617671	<b>0.9812185361</b>
LGBM	범주형 변수 더미화 / 숫자형 변수 Standard Scaler 종속변수 Box-cox변환	연령별 직장인구 비율 연령별 상주인구 비율 편의점 밀도		GridSearchCV 'num_leaves': [15, 31, 50], 'learning_rate': [0.05, 0.1, 0.2], 'n_estimators': [50, 100, 200]	2,283,010,610, 939,760	27001065.95	47780860	<b>0.9878149154</b>
LGBM	범주형 변수 더미화 / 숫자형 변수 Standard Scaler 종속변수 Box-cox변환	연령별 직장인구 비율 연령별 상주인구 비율 편의점 밀도		num_leaves': [15, 31, 50], 'learning_rate': [0.1, 0.15, 0.2], 'n_estimators': [200, 300, 400]	2,289,286,910, 788,070	27355461.16	47846493	<b>0.987781417</b>



# THANK YOU

---