

편의점 운영 시간대 선택을 도와 운영 부담 감소를 위한

# 강남구 지역 상권 기반 시간대별 편의점 매출 예측

Value Trackers

김신우, 김성원, 류민승, 한규동, 최수민

# 목차

강남구 지역 상권 기반  
시간대별 편의점 매출 예측

## 01 주제 선정 배경

- 주제 선정 배경
- 목적 및 필요성
- 분석 대상 선정
- 프로젝트 목표 설정

## 02 데이터 분석

- 분석 방법 설정
- 데이터 수집
- 데이터 전처리

## 03 모델링

- 상권 분류
- 최종 변수 정리
- 예측 모델 소개 및 구현
- 모델 성능 평가 지표
- Feature Engineering
- 예측 모델 성능 평가

## 04 결과 정리

- Streamlit 구현
- 기대효과 / 한계점
- 프로젝트 일정
- 데이터 정의서
- 사용 프로그램

# 팀원 소개

Value Trackers

팀장

**한규동**

# 모델링 # EDA  
#Geopy시각화  
# Streamlit

<https://bmil2011s.tistory.com/>

팀원

**김성원**

# Streamlit # EDA  
# 모델링 # 군집화

<https://castle1data.tistory.com/>

팀원

**김신우**

# Streamlit # EDA  
# 모델링 # 크롤링

<https://datapiLOTS.tistory.com>

팀원

**류민승**

# 모델링 # EDA  
# Feature Engineering  
# PPT 제작

<https://everyday-joyful.tistory.com>

팀원

**최수민**

# PPT 지원

# 주제 선정 배경 | 편의점 매장운영비용 증가에 따른 자영업자 부담 증가

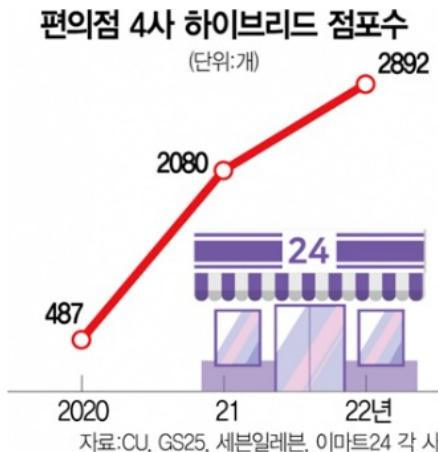
## 01. 편의점 관리비 & 인건비 부담 증가

- 전기요금 등 기본 관리비 증가
- 편의점 평균운영경비 비중은 **인건비가 58.7%**로 가장 높음
- 체감 인건비 인상은 **최저임금 상승률(16.9%)**보다 높음
- 야간 인건비 및 주휴수당** 등으로 24시간 영업 부담이 심함



## 02. 편의점 경쟁 과열

- 편의점 업계의 전체 매출과 점포 수, 판매액의 급성장 속 **과밀 경쟁**  
→ 지역 점포당 매출은 감소
- 생활밀접소매업종(30개) 중 **편의점의 폐업 수**가 가장 많음  
(2019년 7월~2022년 3분기 사업자등록 기준)



■ 2022년 3분기 서울시에선...  
편의점 339곳 개점, 324곳 폐점

### 업체 간 경쟁 과열

- 영업권 보장의 하점
- 구속력 없는 자율규약

왜 유독 편의점 폐업 많나

# 목적 및 필요성

| 실효성 있는 매출액 예상을 통한 매장 운영 지원 필요

## 24시간 계약을 유도하는 편의점 브랜드

**Nice to CU**

담보	3,600만원	5,000만원
수입	매출이익 중	<b>24시간: 80%</b> 19시간: 75%
		<b>24시간: 70%</b> 19시간: 65%

**GS25**

계약 조건	최종 경영주 수익 배분율		최대 73% (특약 83%)	최대 67%	최대 48%
	경영주 수익 배분율	66%	60%	41%	
	<b>24시간 영업 장려금</b>	<b>5%</b>	<b>5%</b>	<b>5%</b>	
	영업 활성화 장려금	최대 2%	최대 2%	최대 2%	
	계약 기간	5년	4년	4년	

**24시간 운영 혜택**

- 인테리어 비용을 추가 지원
- 24시간 영업장려금 지원
- 매출이익 중 수입 비율 증가

## 매장 운영 효율성 증가 및 손익 판단을 위한 입지 기반 매출 정보 제공 필요

[유통핫이슈] 편의점 야간영업 딜레마..‘하이브리드 매장’이 답이다

김성아 기자 | 승인 2021.07.05 02:05 | 의견 0



편의점의 야간영업은 코로나19 이전부터 업계 ‘딜레마’로 불려왔다.

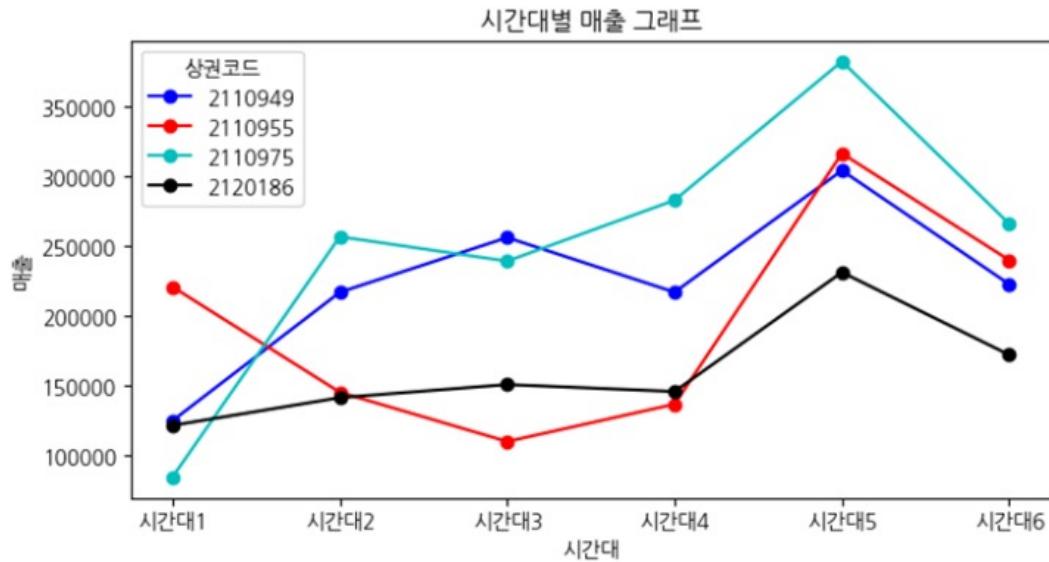
코로나19 이전에는 심야시간 유동인구가 적어지는 주택가·지방 소도시 위주 점포에 적용됐으나 코로나 시국 거리두기 적용으로 수도권 점포도 적용 대상이 됐다.

- 운영 시간 선택지 : 24h / 19h / other
- 브랜드에서 제공하는 예상 매출액  
→ 구체적 자료에 근거하지 않은 대략적인 예상 매출액을 구두로 전달

# 분석 대상 선정 | 강남구 선정

## 01. 상권 별로 상이한 시간대별 매출액

- 서울시 상권 중, 임의의 4개 상권을 특정하여 확인



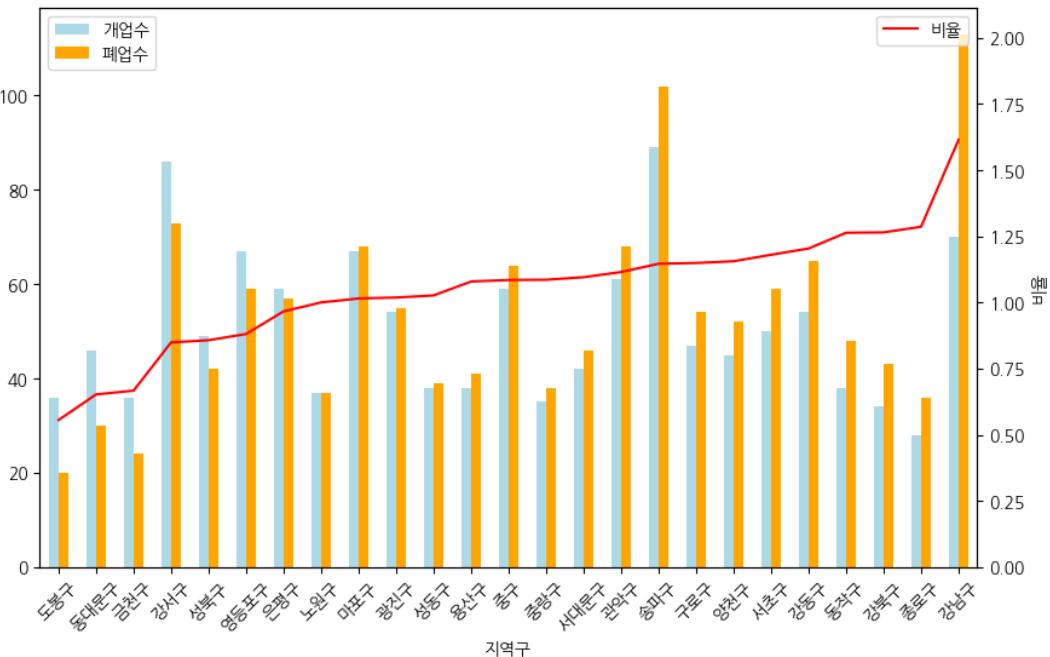
분석  
결과

상권 별로 시간대별 매출에 통계적 차이가 있음을 확인

- F-통계량 (F-Statistics): 163.56706049040335
- P-값 (P-Value): 1.526317887214294e-162

## 02. 경쟁은 가장 심하고, 운영은 어려운 강남구

- 시단위 급에 육박하는 강남구의 편의점 수 편의점 포화상태
- 서울시 지역구 별 개폐업수 1위
- 평균 영업기간
- 면적당 매출액



- 서울시 상권 중, 임의의 4개 상권을 특정하여 확인



시단원군에 윤반한 강남구의 폐이전 승 폐이전 포화산터

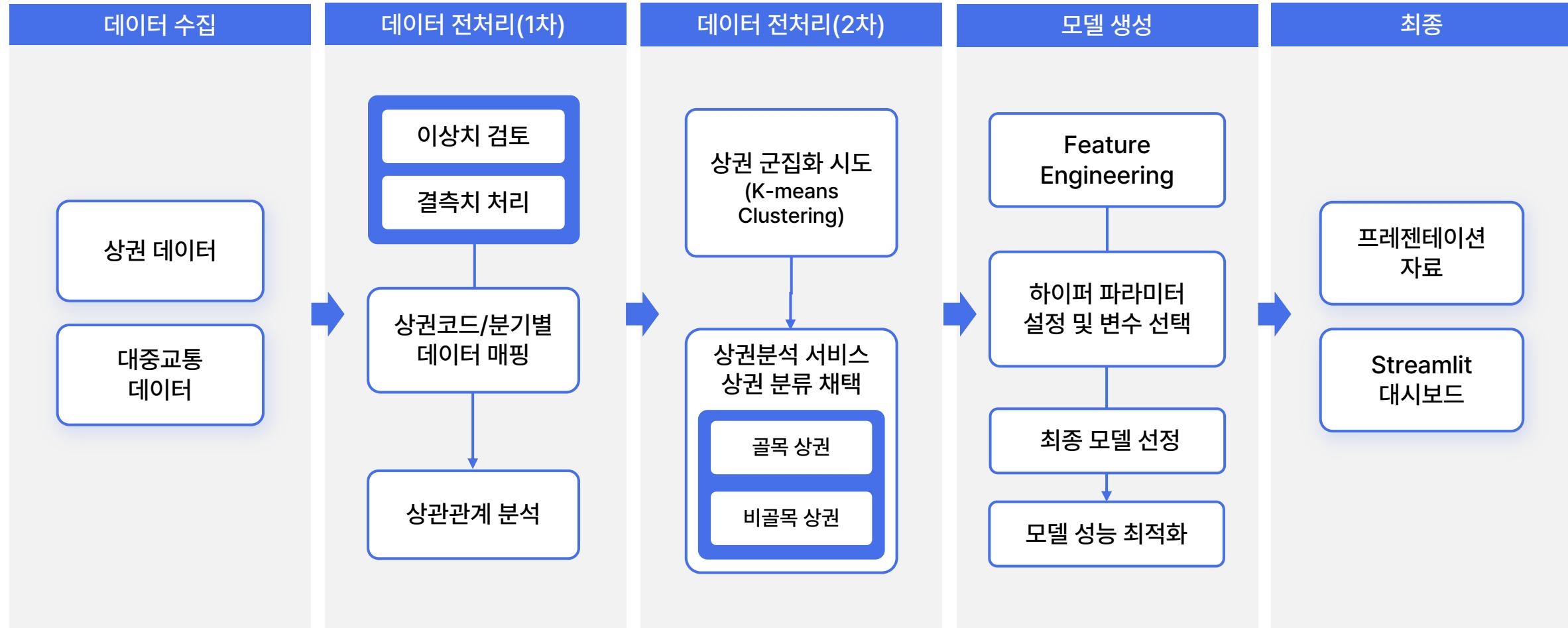
상업 1위  
업수 1위

# 강남구 상권 내 편의점 시간대 별 매출 예측 서비스

## - 편의점 점주의 입지별 효율적인 운영 시간대 선정을 위한 -

# 분석 방법

## | 데이터 분석 프로세스



# 데이터 수집 | 주요 독립변수 데이터 수집

대분류	구분	상세내용	출처
상권분석 서비스 데이터	추정 매출	<ul style="list-style-type: none"> <li>• 분기별 매출 금액</li> <li>• 시간대별 매출 금액</li> <li>• 점포수</li> </ul>	서울열린데이터광장
	아파트	<ul style="list-style-type: none"> <li>• 아파트 단지 수</li> <li>• 가격 별로 세대 수</li> </ul>	서울열린데이터광장
	직장 인구	<ul style="list-style-type: none"> <li>• 총 직장 인구 수</li> </ul>	서울열린데이터광장
	생활 인구	<ul style="list-style-type: none"> <li>• 총 생활 인구 수</li> <li>• 시간대 생활 인구 수</li> </ul>	서울열린데이터광장
	집객 시설	<ul style="list-style-type: none"> <li>• 총 집객 시설 수</li> <li>• 집객 시설 수(관광서, 종합병원, 교육기관 등)</li> </ul>	서울열린데이터광장
	상주 인구	<ul style="list-style-type: none"> <li>• 총 가구 수</li> <li>• 총 상주 인구 수</li> <li>• 아파트 가구 수</li> <li>• 비아파트 가구 수</li> </ul>	서울열린데이터광장
	상권 위치 정보	<ul style="list-style-type: none"> <li>• X좌표, Y좌표(EPSG:5181)</li> </ul>	서울열린데이터광장
대중교통 데이터	버스 정류장	<ul style="list-style-type: none"> <li>• 정류장별 시간대별 승하차 승객 수</li> <li>• 정류장 위치 정보</li> </ul>	서울열린데이터광장
	지하철 역	<ul style="list-style-type: none"> <li>• 역 별 시간대별 승하차 승객 수</li> <li>• 역 위치 정보</li> </ul>	서울열린데이터광장

# 데이터 전처리 | 상권 데이터 병합 및 강남구 추출

## 01. 데이터 병합

- 연도와 분기, 상권 기본정보를 공통 컬럼으로 채택
- 서울시 상권 상세 데이터와 위치 정보를 병합

공통 컬럼	
• 기준_년도	• 상권_구분_코드_명
• 기준_분기	• 상권_코드
상권 정보 데이터	
• 추정 매출	• 생활인구
• 상주 인구	• 직장인구
• 아파트	• 집객 시설
위치	
• 시군구_코드	• 위도
	• 경도

## 02. 강남구 추출

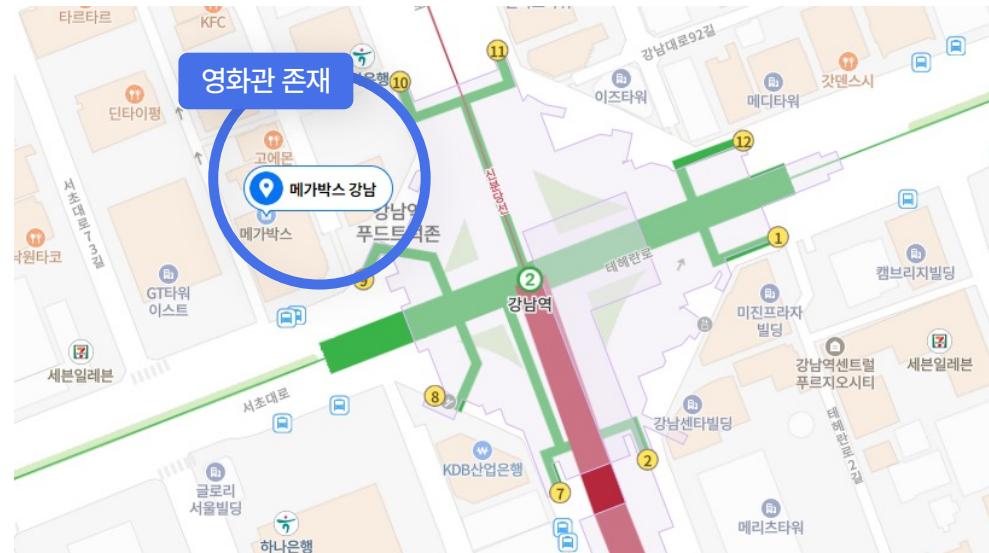
- 공통 컬럼으로 1차 병합 후
- 위치 정보를 병합 → 강남구 데이터 추출

매출	기준_년_코드	상권_구분_코드_명	상권_코드	상권_코드_명
125163.1502	2020	골목상권	2110949	한남IC
94688.22344	2020	골목상권	2110951	신사은행나무공원
177736.4615	2020	골목상권	2110952	논현목련공원
34587.75458	2020	골목상권	2110954	국기원
221061.4832	2020	골목상권	2110955	논현초등학교
156661.2839	2020	골목상권	2110957	언북중학교
114302.1758	2020	골목상권	2110958	역삼역 4번

# 데이터 전처리 | 상권 내 집객 시설 데이터 일부 삭제

## 측정 오류가 심한 집객 시설 데이터 삭제

- ▶ 강남역에 위치한 대표적 집객 시설 다수가 공공데이터에  
집계되지 않은 경우 → 해당 시설 삭제



[강남역 상권 예시]

→ 역 근처에 존재하는 영화관이 데이터 상에는 없음 확인

### 제거된 집객시설 Feature

- 제거 된 집객 시설 Feature
- 종합 병원 수
- 일반 병원 수
- 약국
- 영화관
- 유치원
- 초등학교
- 중학교
- 고등학교
- 대학교 등.. 14개 항목

### 분석에 활용할 집객시설 Feature

- 총 집객 시설 수
- 관공서 수
- 은행 수
- 백화점 수
- 숙박시설 수

# 데이터 전처리 | 상권 데이터 결측치 처리

## 01. 데이터별 결측치 확인

아파트 데이터 결측치	
• 아파트_단지_수	68
• 아파트_가격_1억_미만_세대_수	68
• 아파트_가격_1억_세대_수	68
• 아파트_가격_2억_세대_수	68
• 아파트_가격_3억_세대_수	68
• 아파트_가격_4억_세대_수	68
• 아파트_가격_5억_세대_수	68
• 아파트_가격_6억_세대_수	68

## 02. 결측치 상권 확인 및 처리

아파트 결측치 포함 상권	
• 한국교통안전공단 강남자동차검사소	
• 코엑스	
• 수서역	
• 강남개포시장	
• 강남 마이스 관광특구	
• 대왕초등학교(세곡동사거리서남측)	

이전 연도/분기 데이터는 존재  
휴먼 에러로 인한 누락 값으로 판단  
**“과거 데이터”로 대체**

집객 시설 데이터 결측치	
• 집객시설_수	61
• 관공서_수	61
• 은행_수	61
• 백화점_수	61
• 숙박_시설_수	61

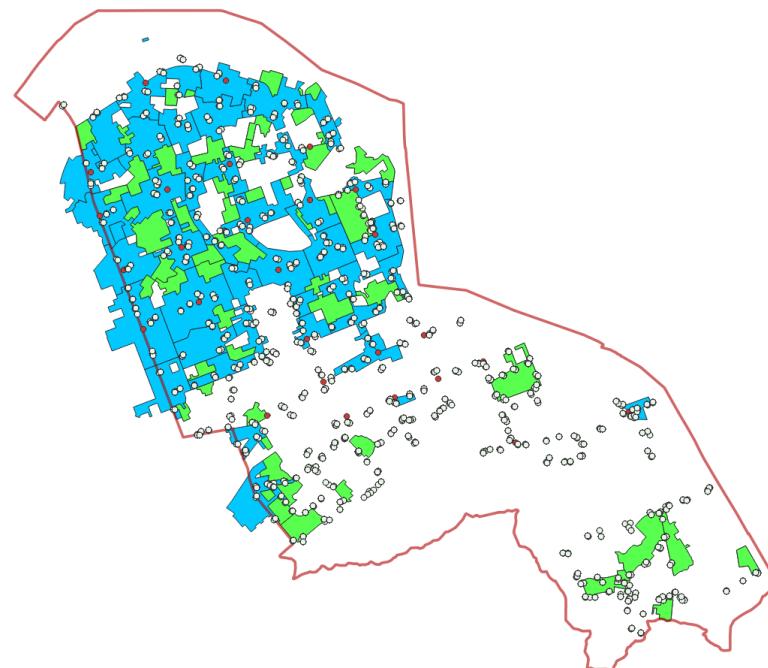
집객시설 결측치 포함 상권	
• 논현로18길	
• 논현목련공원	
• 역삼월드메르디앙아파트	
• 언주역 6번	
• 언주중학교	
• 청담근린공원	
• 국기원	

수집된 데이터 **기간 내 결측치** 존재  
개별 상권 데이터 확인 후  
**“0”으로 대체**

# 데이터 전처리 | 대중교통 데이터 추출 및 병합

## 01. 상권영역 폴리곤 데이터 기반 영역 표시

- 상권영역 폴리곤 데이터를 이용, QGIS로 상권영역 표시
- 상권 영역 기준, 버스 정류장은 흰색 점/지하철은 적색 점으로 표시
- 지하철역과 버스정류장 위치데이터는 위도와 경도 이용



## 02. 상권 내 대중교통 정보 컬럼 추가

- 해당 상권 영역에 포함된 버스 정류장 수 & 지하철 역 수 컬럼 추가
- 포함된 버스 정류장 종류 & 지하철 역 종류 컬럼 추가

기준_년 _코드	기준_ 분기_ 코드	상권_코드	폴리곤_좌표	포함된 _버정 _수	포함 된_역 _수	버스정류장들	지하 철역 들
620	2022	1	POLYGON ((203334.626 443760.667300001, 203351...))	0	0	[ ]	[ ]
621	2022	1	POLYGON ((202191.9843 446916.686100001, 202201...))	0	0	[ ]	[ ]
622	2022	2	POLYGON ((202191.9843 446916.686100001, 202201...))	0	0	[ ]	[ ]
623	2022	1	POLYGON ((205881.8329 445001.2619, 205888.196 ...))	5	0	[23245, 23248, 23435, 23509, 23511]	[ ]
624	2022	1	POLYGON ((204412.5002 446860.439400001, 204436...))	4	0	[23174, 23176, 23184, 23186]	[ ]
...	...	...	...	...	...	...	...
909	2022	4	POLYGON ((205341.3474 442789.2896, 205364.8504...))	0	0	[ ]	[ ]
910	2022	4	POLYGON ((204332.5735 445272.9542, 204438.4928...))	6	1	[23214, 23218, 23241, 23258, 23259, 23276]	[선 릉]
911	2022	4	POLYGON ((202790.0829 447564.47650000104, 2027...))	6	1	[23103, 23105, 23106, 23149, 23429, 23430]	[입구 점]
912	2022	4	POLYGON ((202063.9767 447018.8023000097, 2020...))	2	0	[23101, 23108]	[ ]
913	2022	4	POLYGON ((203483.0631 444271.242000001, 203491...))	7	0	[23292, 23293, 23302, 23303, 23304, 23313, 23512]	[ ]

# 데이터 전처리 | 대중교통 데이터 추출 및 병합

## 01. 상권 영역 내 지하철 이용 승하차 승객 수(총계)

- 연도 / 분기 / 지하철역 기준
- 각 시간대에 지하철을 이용한 승하차 승객 수를 총합

			연도	분기	지하철역	00-06시	06-11시	11-14시	14-17시	17-21시	21-24시
0	2020	1	강남	152555	3719525	1984625	2369429	4565122	1863535		
1	2020	1	강남구청	27841	1351101	465065	544235	1227153	349236		
2	2020	1	개포동	4454	162279	75881	80158	146614	56165		
3	2020	1	구룡	2358	70733	29613	39200	51238	14470		
4	2020	1	논현	30177	1073607	326798	401962	999650	326825		
...	...	...	...	...	...	...	...	...	...	...	...
319	2022	4	일원	16404	587731	293259	343524	411463	111671		
320	2022	4	청담	29607	1310937	386226	485790	1015878	234064		
321	2022	4	학동	19463	1509700	401817	513143	1259980	261640		
322	2022	4	학여울	3731	127566	76648	91934	106777	24630		
323	2022	4	한티	13161	737687	416398	469708	813595	270734		

## 02. 상권 영역 내 버스정류장 이용 승하차 승객 수(총계)

- 연도 / 분기 / 버스정류장 기준
- 각 시간대에 버스정류장을 이용한 승하차 승객 수를 총합

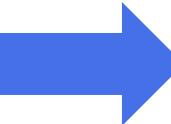
			연도	분기	버스정류장	00~06_승하차_승객수	06~11_승하차_승객수	11~14_승하차_승객수	14~17_승하차_승객수	17~21_승하차_승객수	21~24_승하차_승객수
0	2020	1	23101			6672	71579	49177	57856	81299	31310
1	2020	2	23101			8033	73779	48388	60335	86690	33037
2	2020	3	23101			7756	74680	49755	58587	83374	33121
3	2020	4	23101			6377	73417	51018	59961	77762	26662
4	2020	1	23102			4348	42466	33203	48549	87372	38859
...	...	...	...	...	...	...	...	...	...	...	...
6220	2022	4	23994			5737	57677	23457	29475	38924	13017
6221	2022	1	23995			665	17806	8226	9174	12517	5088
6222	2022	2	23995			802	23953	9273	11761	15561	6048
6223	2022	3	23995			657	23896	10403	12483	16820	6208
6224	2022	4	23995			721	26423	10947	13817	17724	6570

# 데이터 전처리 | 대중교통 데이터 결측치 처리

## 01. 지하철 데이터

- 지하철 역 컬럼 내,  
‘양재’ 역 데이터만 없음(누락값으로 추정)
- 지하철역 위도 경도 데이터  
→ 결측치 없음
- 지하철역 이용 승하차 승객 수  
→ 결측치 없음

양재역은 3개년  
데이터 모두 없음  
**“누락값 제외”**



### 지하철 최종 데이터

- 지하철역 (이름)
- 지하철역 위치  
(위도, 경도)
- 시간대별 지하철 승하차 승객수
- 상권 내 포함된 지하철 역 개수
- 상권 내 포함된 지하철 역 종류

## 02. 버스 데이터

- 다년도 데이터를 수집했으므로  
과거 좌표 및 실제 지도와 교차 비교를  
통해 열린데이터 광장에서 제공한  
데이터를 모두 정제함
- 그럼에도 불구하고 가상의 정류장 하나가  
지명 등 임의로 파악하기도 모호한  
경우가 있었음

결측값이  
다른 데이터들에  
비해 현저히 작음  
**“결측치 삭제”**



### 버스 최종 데이터

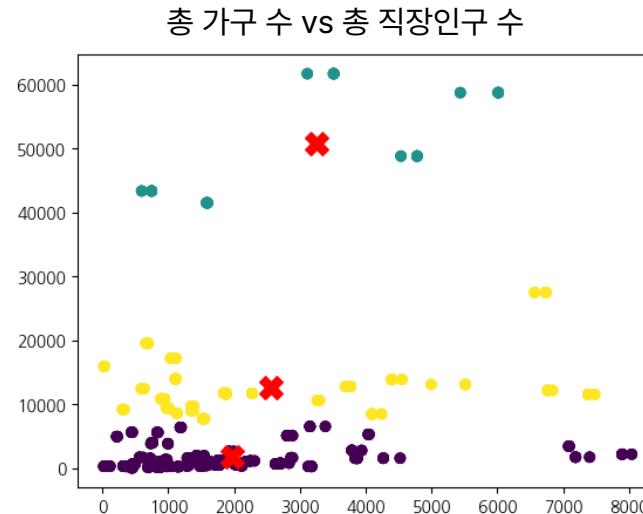
- 버스정류장 (코드)
- 버스정류장 위치  
(위도, 경도)
- 시간대별 버스 승하차 승객수
- 상권 내 포함된 버스정류장 개수
- 상권 내 포함된 버스정류장 종류

# 모델링 | 상권 분류

## 상권 분류 가설 설정 및 K-means Clustering 시도

가설 1

상권 분류에 따라 총 가구 수 대비  
직장인구 수에 차이가 존재한다

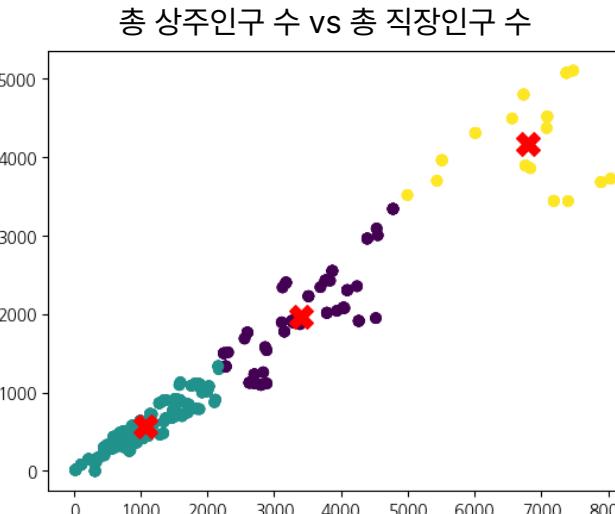


"0.32"

실루엣 계수

가설 2

상권 분류에 따라 총 상주 인구 수 대비  
직장인구 수에 차이가 존재한다



"0.29"

실루엣 계수

뚜렷한  
군집화 결과  
없음

▶ 이 외 여러 가설을 통해 2개의 변수 무작위 조합으로 군집 분류를 시도했지만, 실루엣 계수 0.5 이상의 유의미한 분류를 찾을 수 없었음

# 모델링 | 상권 분류

## 서울시 상권분석 시스템에서 사용한 상권 분류 활용



# 모델링 | 최종 변수 정리

## 파생변수 생성 전, 전체 Feature 정리

대중교통 데이터	<ul style="list-style-type: none"> <li>시간대_버스_승하차 승객 수</li> <li>시간대_지하철_승하차 승객 수</li> <li>버스정류장_수</li> <li>지하철역_수</li> </ul>	기본 인구 데이터	<ul style="list-style-type: none"> <li>총 상주인구 수</li> <li>총 가구 수</li> <li>총_직장인구_수</li> <li>총_생활인구_수</li> <li>시간대_생활인구_수</li> </ul>	아파트 세대수 데이터	<ul style="list-style-type: none"> <li>아파트_단지_수</li> <li>아파트_가격_1_억_미만_세대_수</li> <li>아파트_가격_1_억_세대_수</li> <li>아파트_가격_2_억_세대_수</li> <li>아파트_가격_3_억_세대_수</li> <li>아파트_가격_4_억_세대_수</li> <li>아파트_가격_5_억_세대_수</li> <li>아파트_가격_6_억_이상_세대_수</li> </ul>
기타 데이터	<ul style="list-style-type: none"> <li>점포 수</li> <li>Area</li> </ul>	요일별 생활인구 세분화	<ul style="list-style-type: none"> <li>월요일_생활인구_수</li> <li>화요일_생활인구_수</li> <li>수요일_생활인구_수</li> <li>목요일_생활인구_수</li> <li>금요일_생활인구_수</li> <li>토요일_생활인구_수</li> <li>일요일_생활인구_수</li> </ul>	배후지 아파트 세대수 데이터	<p>(골목상권만 해당)</p> <ul style="list-style-type: none"> <li>배후지_아파트_단지_수</li> <li>배후지_ 아파트_가격_1_억_미만_세대_수</li> <li>배후지_ 아파트_가격_1_억_세대_수</li> <li>배후지_ 아파트_가격_2_억_세대_수</li> <li>배후지_ 아파트_가격_3_억_세대_수</li> <li>배후지_ 아파트_가격_4_억_세대_수</li> <li>배후지_ 아파트_가격_5_억_세대_수</li> <li>배후지_ 아파트_가격_6_억_이상_세대_수</li> </ul>
집객시설 데이터	<ul style="list-style-type: none"> <li>집객시설_수</li> <li>관광서_수</li> <li>은행_수</li> <li>백화점_수</li> <li>숙박_시설_수</li> </ul>	연령대별 생활인구 세분화	<ul style="list-style-type: none"> <li>연령대_10_생활인구_수</li> <li>연령대_20_생활인구_수</li> <li>연령대_30_생활인구_수</li> <li>연령대_40_생활인구_수</li> <li>연령대_50_생활인구_수</li> <li>연령대_60_이상_생활인구_수</li> </ul>		

# 모델링 | 예측 모델 소개 및 구현

## 01. Machine Learning 모델



- 학습 시간이 적은 편이며, 예측 오류 손실을 최소화 할 수 있어 선택
- 수치형 feature 뿐만 아니라 카테고리형 feature의 자동변환과 최적 분할로 인코딩을 별도 진행하지 않아도 되어 간편
- 대용량 데이터에 대한 뛰어난 예측 성능과 병렬 컴퓨팅 기능
- 단, 본 프로젝트에 적용시 전체 데이터 세트 개수가 적은 편이라 과적합 발생에 대한 우려가 있었음

## 02. 과적합 방지 및 예측 성능 향상 기술

### K-fold Cross Validation

- 데이터를 k개로 분할한 뒤, k-1개를 학습용 데이터 세트로, 1개를 평가용 데이터 세트로 사용하는 K-fold 교차 검증을 사용
- 초기 k 값을 5로 설정했으나, 데이터 세트의 개수를 고려하여 과적합 방지를 위해 k 값을 10으로 변경하여 모델링 수행

### Hyper Parameter Setting with RandomSearch

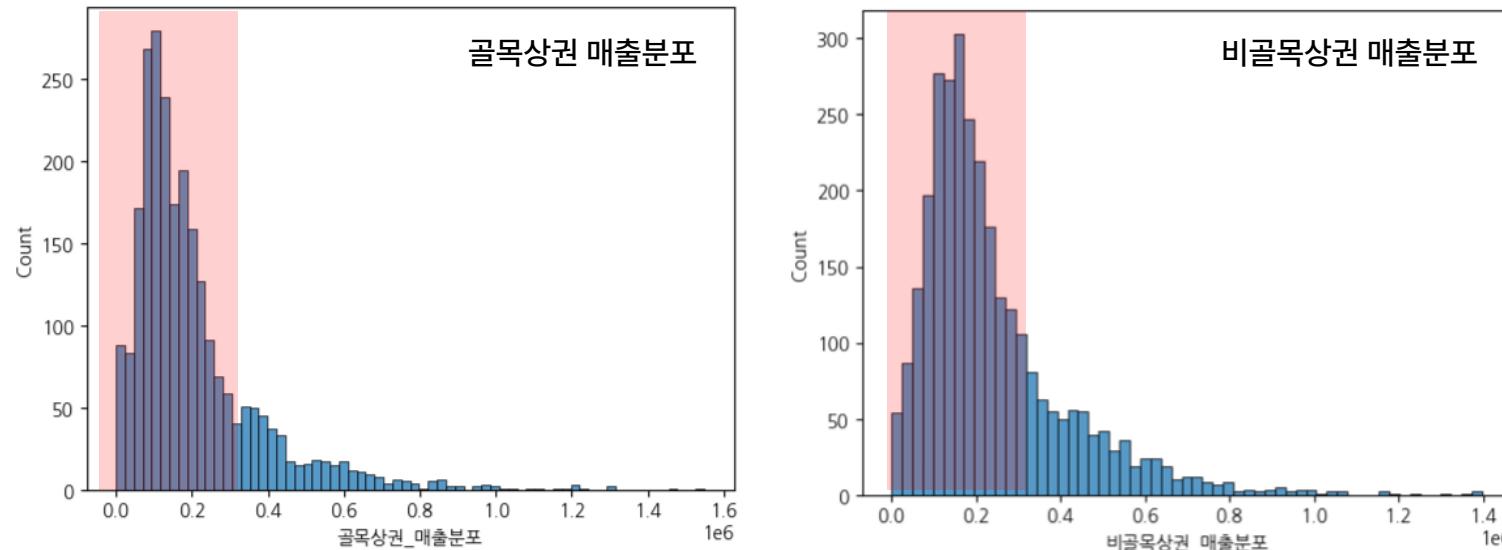
- 과적합을 방지하고 최적의 모델 성능을 위한 하이퍼 파라미터 튜닝 진행
- 모델 내부에서 최적의 하이퍼 파라미터 조합을 찾기 위한 RandomSearch 튜닝 선택
- `num_leaves, learning_rate, n_estimators, early_stopping_rounds`

# 모델링 | 모델 성능 평가 지표 선정

## 균일하지 않은 분포의 종속변수 예측을 위해 RMSE 선정

- 종속변수의 분포(왜도 약 2.5)가 고르지 않아, RMSE로 모델 성능 평가 지표를 선정
- RMSE(Root Mean Squared Error)는 MSE의 제곱근으로, 예측값과 실제값 사이의 오차의 제곱에 루트를 씌운 값
- 오차의 제곱인 MSE에 비해, 이상치에 덜 민감한 편

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N (pred_i - target_i)^2}$$

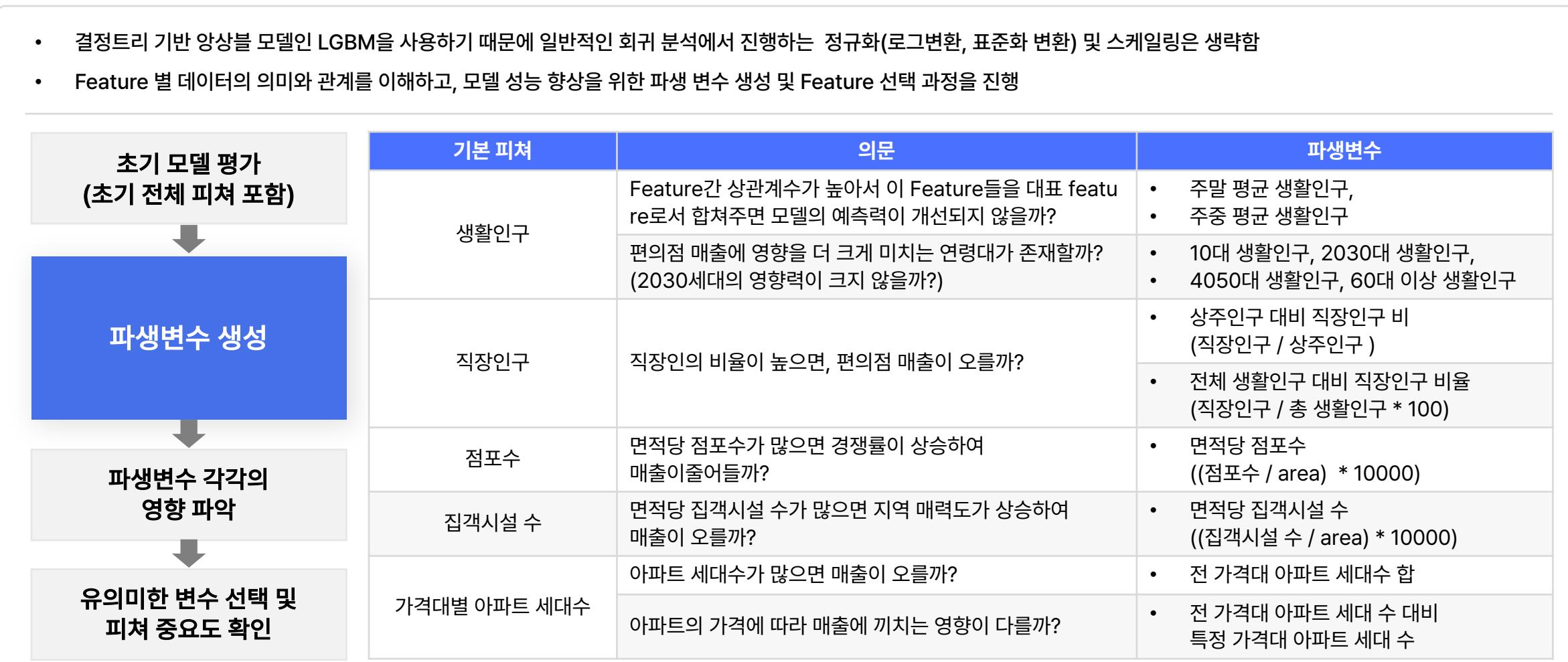


균일하지 않은  
분포로 MAE 대신  
**RMSE 사용**

# 모델링 | Feature Engineering

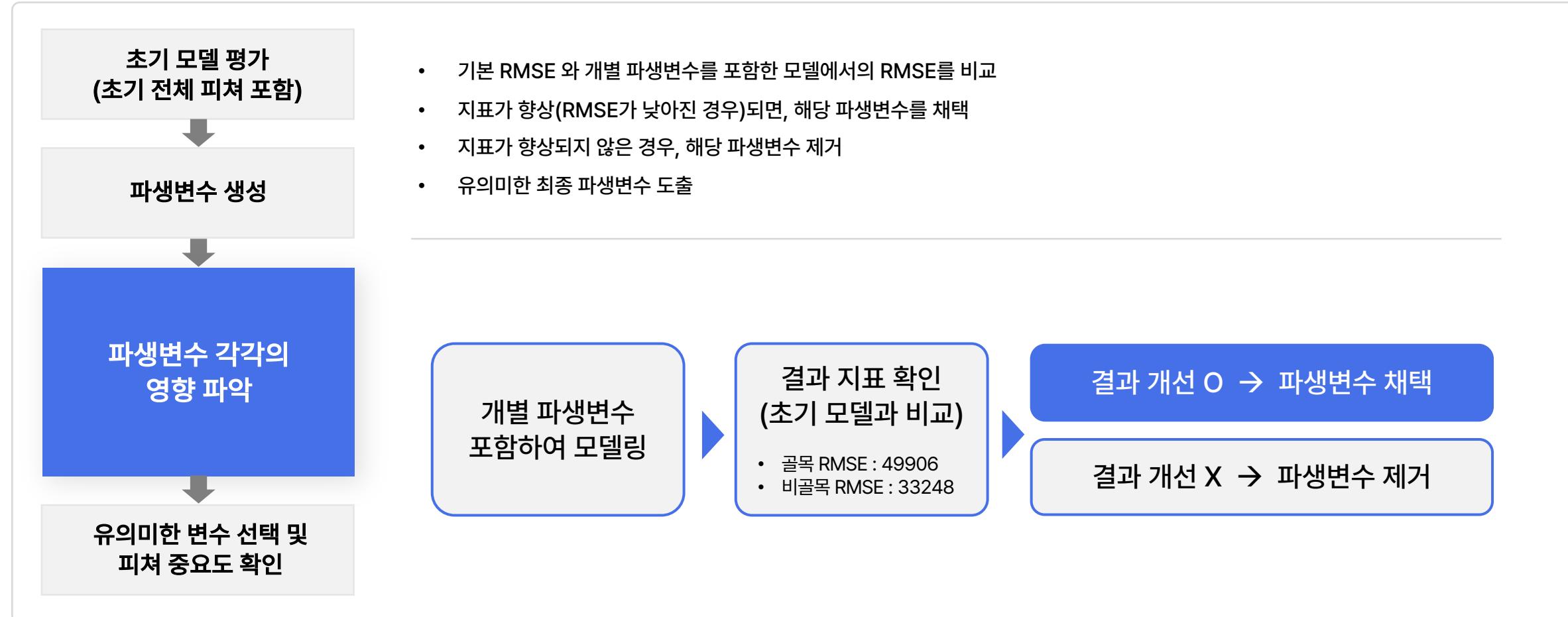
## 초기 모델 평가 및 파생변수 생성

- 결정트리 기반 양상을 모델인 LGBM을 사용하기 때문에 일반적인 회귀 분석에서 진행하는 정규화(로그변환, 표준화 변환) 및 스케일링은 생략함
- Feature 별 데이터의 의미와 관계를 이해하고, 모델 성능 향상을 위한 파생 변수 생성 및 Feature 선택 과정을 진행



# 모델링 | Feature Engineering

## 모델링에 사용할 파생변수 채택 : 전진선택법



# 모델링 | Feature Engineering

## 모델링에 사용할 최종 변수 선택(기본변수 + 파생변수) : 후진제거법



# 모델링 | 예측 모델 성능 평가 종합

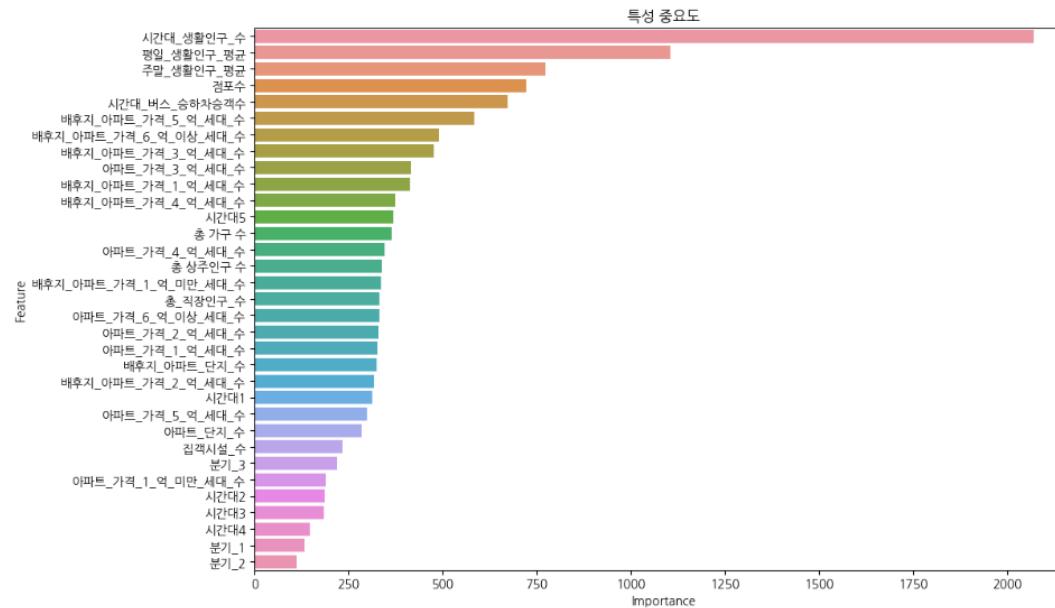
## 01. 골목상권

초기 RMSE  
42666

최종 RMSE  
41951

최적 하이퍼파라미터:

```
'objective': 'regression', 'num_leaves': 59, 'n_estimators': 253, 'metric': 'mse', 'learning_rate': 0.05, 'early_stopping_rounds': 41
```



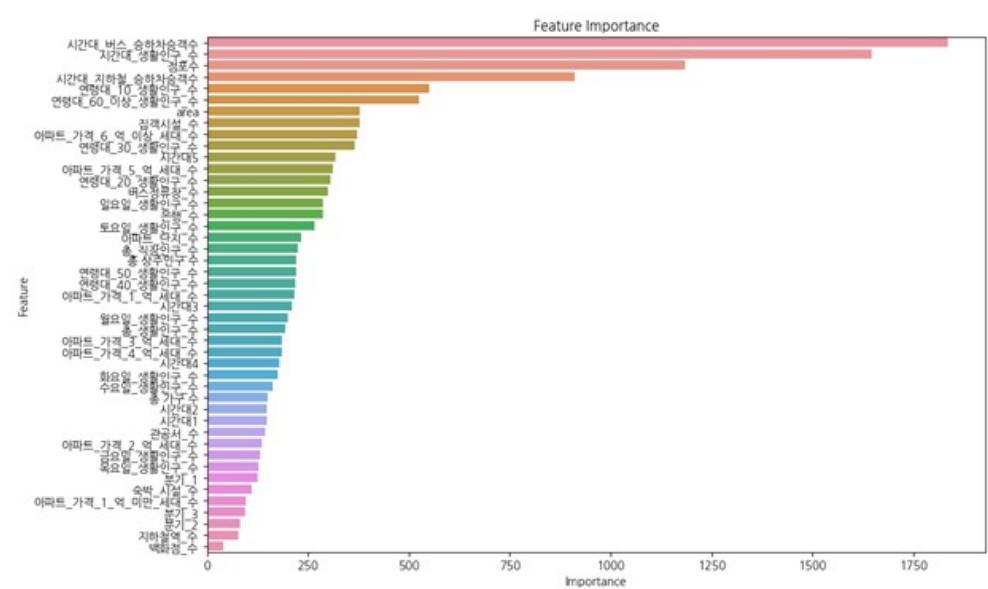
## 02. 비골목상권

초기 RMSE  
42666

최종 RMSE  
33248

최적 하이퍼파라미터:

```
'objective': 'regression', 'num_leaves': 59, 'n_estimators': 253, 'metric': 'mse', 'learning_rate': 0.05, 'early_stopping_rounds': 41
```



# Streamlit 구현

## Home : 서비스 안내

**Home**

**main page**

**sub page**

**강남구 편의점 매출 예측 서비스**

본 서비스는 공공데이터를 기반으로 강남구의 편의점 매출에 영향을 미치는 다양한 지표를 분석하여 시간대별 매출 예측을 제공합니다.

**MainPage**

**SubPage**

시간대\_생활인구\_수:  
61351.46

평일\_생활인구\_평균:  
54461.95

주말\_생활인구\_평균:  
47899.00

비고 : 매출 예측에 사용되는 데이터는 모두 2022년 분기별 최신 데이터를 활용하였습니다.

[종합] 시간대별 예상 매출

소개 : 상권별 편의점 매출 예측 데이터에 대한 시각화 자료를 확인할 수 있습니다.

사용법 : Side bar에서 상권과 분기를 선택 후 Main에서 결과 확인합니다.

비고 : 슬라이더의 기본값은 각 상권 및 분기별 최신 데이터로 설정되어 있습니다.

[세부] 매출 요인 조절 & 예측

소개 : 매출과 관련된 변수를 직접 조절하여 예상 매출을 확인할 수 있습니다.

예상 사용 시나리오 : 이번에 예측하고자 하는 편의점 근처에 새로운 지하철 라인이 개통한다면, 지하철 승하차 승객수를 조절한 후 그 결과를 확인해 볼 수 있습니다.

소개 : 매출과 관련된 변수를 직접 조절하여 예상 매출을 확인할 수 있습니다.

사용법

👉 Side bar에서 상권과 분기를 선택하면, 상권별로 조절할 수 있는 변수가 Main에 표시됩니다.

👉 변수들을 슬라이더로 조절한 후에 "예측 버튼"을 누르면 선택한 변수로 예상 매출액을 확인할 수 있습니다.

## Main Page : 상권 별 시간대 매출 확인

**Home**

**main page**

**sub page**

상권을 선택하세요  
한남IC

분기를 선택하세요!  
1분기

**강남구 편의점 매출 예측 서비스**

최측 사이드바에서 상권과 분기를 선택하면, 시간대별 예상 매출을 확인하실 수 있습니다  
하단 지도에서 상권의 영역을 확인해보세요! ↗

한남IC 1분기 편의점 시간대별 예상 매출

👉 어느 시간대에 매출이 가장 높은지 확인해보세요!

700k  
600k  
500k  
400k  
300k  
200k  
100k  
0

시간대	예상 매출
09 ~ 06	330,186
06 ~ 11	534,469
11 ~ 14	539,398
14 ~ 17	523,891
17 ~ 21	701,436
21 ~ 24	501,199

👉 시간대 17 ~ 21 의 매출이 가장 높습니다!

<https://procpspredictor.streamlit.app/>

# Streamlit 구현

## Sub Page : 변수 조절 및 매출 시뮬레이션

Home

main page

sub page

상권을 선택하세요

역삼역 4번

시간대를 선택하세요

시간대1(00 ~ 06)

분기를 선택하세요!

1분기

강남구 편의점 매출 예측 서비스 - 상세

👉 좌측 side bar에서 상권/시간대/분기를 선택해주세요!

👉 slide를 조정한 후, 예측하기 버튼을 눌러주세요!

점포수:  
3.00

1.00

총 상주인구 수:  
861.00

317.00

총 가구 수:  
544.00

0.00

예측하기

예측 결과

역삼역 4번의 1분기 시간대1(00 ~ 06) 예상 매출은 104,604원입니다.

# 기대효과 / 한계점

## 영업 시간대별 매출 예측으로 매출 안정화와 비용 절감 도모



### 기대 효과

#### 영업 시간대별 매출 예측

- 매장 운영 시간의 효과적 선택
- 비용 절감 및 수익 증진 도모

#### 효율적 자원 배분

- 근로자 파트타임 시간 조절 등 인력 계획
- 상품 발주 시기 조정 등 재고 관리

#### 창업 시 입점 지역 선택

- 신규 창업 상권 선택 시, 예상 비용과 매출을 비교하여 입점 지역 선택에 도움

### 한계점

#### 개별 편의점 매출 정보 수집 불가

- 상권 내 편의점 매출을 평균하여 예측
- 개별 편의점의 매출 예측 정확도 하락 우려

#### 상권 구분 기준 선택 불가

- 서울시 상권 분석 서비스에서 제공한 상권 영역을 기반으로 처리

#### 1시간 단위 시간대 선택 불가

- 시간별 매출이 상권 분석 서비스에서 지정한 6개 구간으로 설정됨
- 개별 시간대에 대한 예측 정확도 하락 우려

# 일정 | 프로젝트 추진 일정

대분류	구분	1주차 (9/6 ~ 9/10)	2주차 (9/11~9/17)	3주차 (9/18~9/25)
기획	기획 상세	■		
데이터 수집	골목상권 및 대중교통 데이터 수집	■ ■		
데이터 전처리	EDA(변수 특징 확인)		■ ■ ■	
	상권 기능별 분류		■ ■ ■	
	기능별 변수에 대한 상관 분석		■ ■ ■	
	최종 기능별 변수 선정 및 변수 변환		■ ■	
모델링 및 평가	골목상권 모델링 / 비골목 상권 모델링		■ ■ ■	
대시 보드 제작	Streamlit 대시보드 구현			■ ■ ■ ■
PPT 제작	프레젠테이션 자료 제작			■ ■ ■ ■

# 데이터 정의서

소분류	컬럼명	데이터 타입	NULL	비고
인구 데이터	총 상주인구 수	int	비허용	서울시 공간정보담당관에서 주민등록주소 기반으로 작성된 상주인구 DB에서 산출하는 정보
	총 가구 수	Int	비허용	상권 내에 거주하고 있는 가구 수의 합
	총 직장인구 수	int	비허용	상권 내에 직장을 가지고 있는 인구 수의 합
	총 생활인구 수	int	비허용	서울시가 보유한 빅데이터와 KT의 통신데이터로 특정한 '특정 시점'에 '서울의 특정 지역'에 '존재'하는 모든 인구
	시간대 생활인구 수	int	비허용	00~06, 06~11, 11~14, 14~17, 17~21, 21~24 으로 각 시간대가 설정됨.
	요일별 생활인구 수	int	비허용	요일 별 상권내에서 관측되는 인구의 합
	연령대별 생활인구 수	int	비허용	10대, 20대, 30대, 40대, 50대, 60대 이상으로 나뉨.
아파트 데이터	아파트 단지 수	int	비허용	서울시 공간정보담당관에서 제공된 아파트 DB 기반 정보
	배후지 아파트 단지 수	int	비허용	상권 기준 200m 외곽까지의 아파트 단지 수
	아파트 가격별 세대수	int	비허용	서울시 공간정보담당관에서 제공된 아파트 DB(아파트 세대수) 및 상주인구 DB(전체 세대수) 기반 정보
	배후지 아파트 가격별 세대수	int	비허용	상권 기준 200m 외곽까지의 아파트 세대 수
집객시설 데이터	집객시설 수	int	비허용	서울시 소재 관공서, 금융기관, 병원, 학교, 문화관광(영화관), 숙박시설 등 인구집중 유발시설 정보
	각 집객시설별 수	int	비허용	상권 내 집객시설 수의 합

# 데이터 정의서

소분류	컬럼명	데이터 타입	NULL	비고
점포 수 데이터	점포 수	int	비허용	사업자등록번호 기반 서울시 소재 사업체 정보
면적 데이터	Area	int	비허용	상권 면적
추정 매출액 데이터	시간대별 매출	int	비허용	3개 카드사의 카드승인금액을 서울시의 보정비율로 나누어 100개 생활밀접업종 매출액을 추정
대중교통 데이터	버스 정류장 수	int	비허용	상권 내 버스정류장 수
기타	시간대별 버스 승하차 승객 수	int	비허용	상권 내 버스정류장의 승하차 승객 수 총합
	지하철역 수	int	비허용	상권 내 지하철 수
	시간대별 지하철 승하차 승객 수	int	비허용	상권 내 지하철역의 승하차 승객 수 총합
	기준 년 코드	int	비허용	2020~2022
	상권 구분 코드 명	object	비허용	발달상권/골목상권/전통시장/관광특구
	상권 코드	int	비허용	상권 별 고유 코드값
	상권 명	object	비허용	간단하게 상권을 설명할 수 있는 상권의 이름

- 배후지 : 상권 영역 테두리 바깥으로 반경 200m 내
- 보정비율 : 연별, 자치구, 행정동, 상권별, 업종별 보정비율
- 카드승인금액 : 월별, 자치구, 행정동, 상권별, 업종별 카드승인금액

# 참고 문헌

---

- 이철환(2021). 편의점 상권 추정과 매출 예측에 관한 연구(건국대학교)
- 이임동, 이찬호, 강상목(2010). 편의점 매출에 영향을 미치는 입지 요인에 대한 실증연구(부산대학교)
- 김현철, 이승일(2019). 서울시 골목상권 매출액에 영향을 미치는 요인 연구(서울도시연구 제 20권 1호)
- 강필수(2022). 빅데이터 기반 상권 정보 한눈에.. 서울시 상권분석서비스 개편 (이코노미스트 기사)
- 송주환, 김영훈(2022). 자영업 점포의 상권 탐색 및 유형 분류 : 청주시 음식점을 사례로. 한국지역지리학회 학술대회 발표집, 개최지.
- 김영재, 박인권(2023). 빅데이터를 활용한 젠트리피케이션 상권의 장소성 분류와 특성 분석 -서울시 14개 주요상권을 중심으로-
- 오윤서, 김동환, 송지현(2022). GIS 기반 로컬상권 형성에 따른 점포집적 특성 - 서울시 성수동 지역의 음식점을 중심으로 (한국실내디자인학회 학술대회 논문집, 개최지)
- 신기동, 김태영, 김현순, 박태원, 유민태, 한영숙(2022). 시흥시 골목상권 현황 분석 및 활성화 방향(경기연구원)
- 이금숙(1995). 지역 접근성 측정을 위한 일반모형. 응용지리

# 사용 프로그램



# 사용 프로그램

Thank you

Q & A

