



MEALKIT

밀키트 매장 최적 입지 선정



양충모¹, 이희준², 윤수영², 정태민², 김경민², 차원식²

1. 조장, 2. 조원





CONTENTS

01

분석 배경

- 주제 선정
- 분석 필요성
- 분석 전략

02

데이터와 전처리

- 자료 출처 및 세부 내용
- 전처리

03

데이터 분석

- 모델 선정
- 회귀 모델

04

결론

- 최종 입지 선정
- 한계 및 향후 연구 방향



01

분석 배경

주제 선정 | 분석 필요성 | 분석 전략

01 분석 배경 _ 주제 선정



밀키트란? 손질된 식재료와 양념, 조리법을 세트로 제공하는 제품

장점 : 손쉬운 조리가 가능해 시간이 절약되고, 외식이나 배달음식보다 가격이 저렴

밀키트 구매 전세계 1위

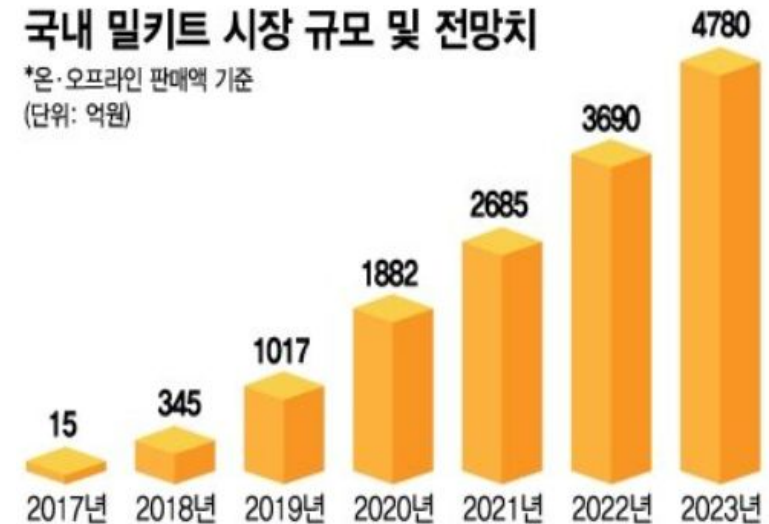


출처 : 스마트 투데이

밀키트 시장성 증가

국내 밀키트 시장 규모 및 전망치

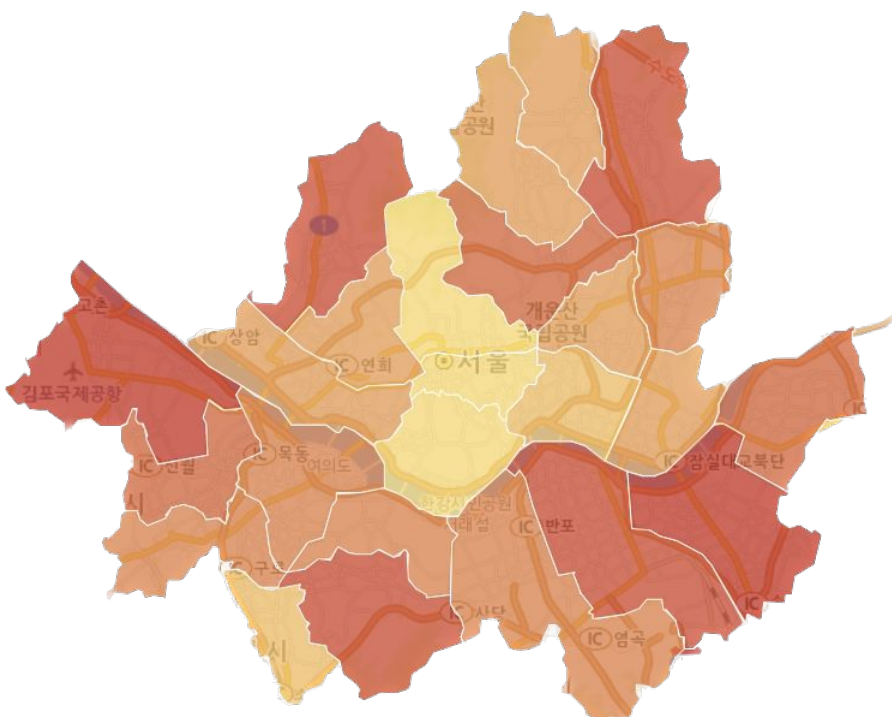
*온·오프라인 판매액 기준
(단위: 억원)



출처 : 유로미터

국민들이 밀키트를 선호하는 이유는 전세계에서 가장 높은 **도시화율**과 **3고 (고물가, 고금리, 고부채)** 등이 주요 요인

01 분석 배경 _ 분석 필요성 : 인구수와 매장 분포



서울시 인구수 분포



서울시 밀키트 매장 위치 분포

→ 인구수와 밀키트 매장 분포는 **비례하지 않음**

01 분석 배경 _ 분석 필요성 : 인구수와 방문자 수



담곡 길음점 밀키트
★4.47/5 · 방문자리뷰 904 · 블로그리뷰 15

인구수 : 53,989



NBM 신정역점 밀키트
방문자리뷰 52 · 블로그리뷰 2

인구수 : 173,388

→ 인구수와 매장 방문자 수는
비례하지 않음



장사가 잘되는 매장의 기준은?

여러 요인들을 수집해 분석한다면 잘되는 매장의 공통점 파악 + 기계학습
커져가는 밀키트 산업에서 **최적의 입지조건**을 분석할 수 있음

01 분석 배경 _ 분석 전략



서울시 밀키트 점포
183개 대상



1인당 평균 방문횟수가 높을수록
매출이 높은 점포로 생각

밀키트의 선호와 구매의 증가 → 시장성의 증가
밀키트의 입지분석 판단 내렸으나
인구수와 매장분포 및 방문자 수는 비례X → 관련성 없다 판단



장사가 잘 되는 매장의 기준을 다시 고려함



여러 영향요인들을 모으고 추출해 최적 입지 요건을 제시하는 것으로 주제 선정
(조사범위는 서울특별시로 제한)



02

데이터와 전처리

자료 출처 및 세부 내용 | 데이터 전처리

02 데이터와 전처리 _ 자료 수집

분류	변수명	사용 용도	자료 출처
지역	행정동 ID, 시, 구, 동	ID	KaKao API
1인당 방문수	방문자 ID, 방문수	target	웹 크롤링
교통량	버스 승객 수	밀키트 점포에 대한 접근성	Open API
사업체	사업체 수, 종사자 수, 평균종사자	회사원들의 밀키트 수요	서울 열린데이터 광장
성 연령별 인구수	~20세, 21세~40세, 41세~70세, 남성 총 인구, 여성 총 인구, 총 인구	성, 연령별 밀키트 수요	서울 열린데이터 광장
세대종류	1인 가구, 2인 가구, 3~5인 가구, 6인~ 가구	가구유형별 밀키트 수요	서울 열린데이터 광장
비만도	비만도 분포	과체중인 사람과 밀키트 수요의 관계	서울 열린데이터 광장

분류	변수명	사용 용도	자료 출처
배달금액	금액 총수	배달수요와 밀키트 수요와의 관계	서울 열린데이터 광장
외국인 수	외국인 수	외국인의 밀키트 수요	서울 열린데이터 광장
경쟁업체	인근 편의점 매장 수, 인근 대형마트 매장 수, 인근 음식점 매장 수	주변 인프라가 밀키트 점포에 미치는 영향	Open API
주거 유형	아파트, 일반단독주택, 다가구주택, 연립주택, 다세대주택	주거 유형별 밀키트 수요	서울 열린데이터 광장
공시지가	23년 공시지가(원/m ²)	공시지가와 밀키트 수요의 관계	서울 열린데이터 광장

02 데이터와 전처리 _ 자료 수집 : 웹 크롤링



담곡 월계점 밀키트

★ 4.52/5 · 방문자리뷰 487 · 블로그리뷰 16



→ Selenium을 활용해 밀키트 매장별, 고객별 누적 방문 수를 추출
이때 방문 기준은 리뷰 수로 정함

02 데이터와 전처리 _ 자료 수집 : 웹 크롤링



	매장명	손님	방문횟수
0	담곡 마들역점	후후훗23	2
1	담곡 마들역점	Joe6484	1
2	담곡 마들역점	레베캉	1
3	담곡 마들역점	13****	2
4	담곡 마들역점	아낌20	2



Target

= 1인당 평균 방문수
= 총 방문수 / ID수



target	
variable	value
NBM 구로궁동점	1.863636
NBM 독바위점	1.080000
NBM 밀키트 금천점	0.000001
NBM 성신여대점	1.295455
NBM 신월점	1.000000
...	...
팔도만찬 만리점	1.166667
팔도만찬 방학역점	1.000000
팔도만찬 신정네거리점	1.000000
팔도만찬 흥제점	1.000000
홈즈앤쿡 쌍문점	1.391304
184 rows × 2 columns	

02 데이터와 전처리 _ 전처리 : Scaling, 상관관계 분석, PCA feature selection



01

Standard Scaling

$$z = \frac{X - \mu}{\sigma}$$

02

상관관계 분석
다중공선성 제거

03

PCA

연관성이 있어 보이는
Feature 통합

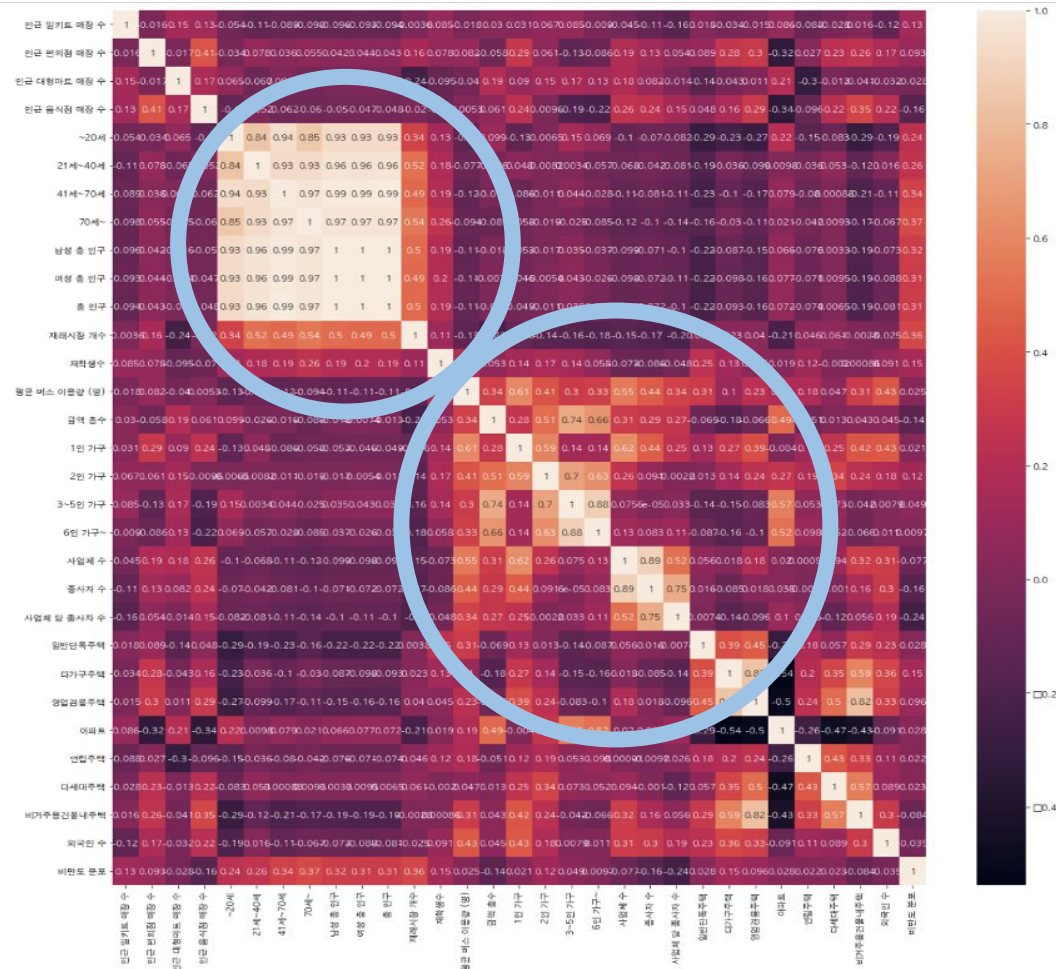
서울시 밀키트 점포는 183개로 데이터가 적지만, Feature는 32개로 과적합이 될 가능성 있음

Standard Scaling 수행 후, 상관관계와 PCA를 통해 Feature간 종속성을 해소

02 데이터와 전처리 _ 전처리 : 상관관계 분석



상관관계가 0.7 이상인 Feature들은 다중공선성이 있다고 판단해 통합



01

사업체 수

사업체 수, 종사자 수, 사업체당 종사자 수

02

총인구수

성별 인구수, 연령별 인구수, 총인구수

03

1~2인 가구, 3인 가구 이상

1인 가구, 2인 가구, 3~5인 가구, 6인 가구 이상

04

아파트, 일반주택

아파트, 일반 단독주택, 다가구주택, 연립주택, 다세대주택

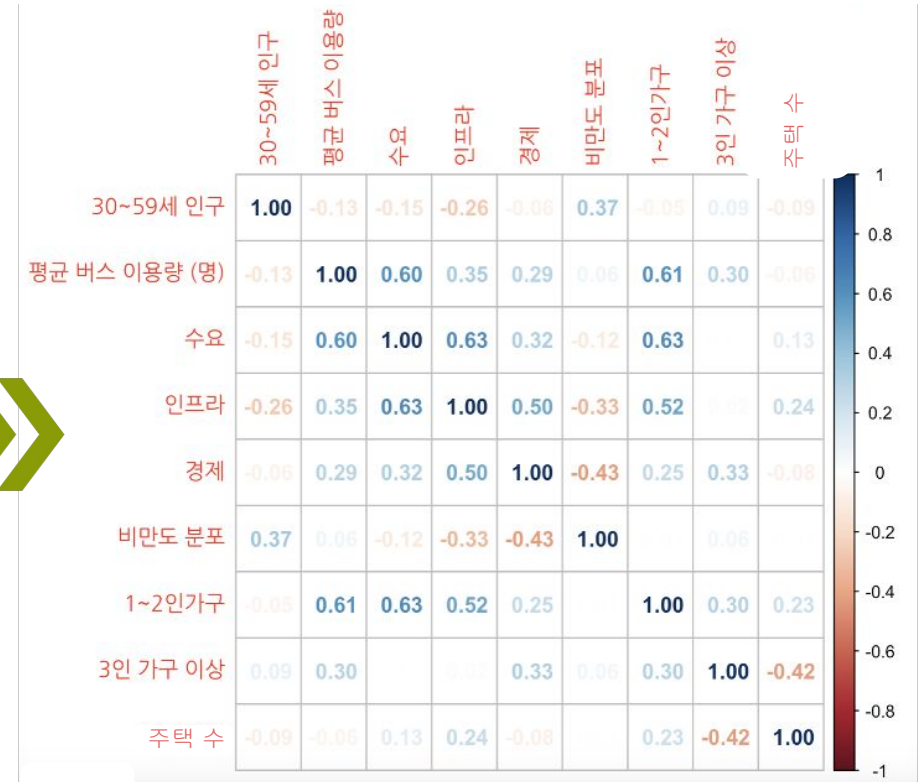
02 데이터와 전처리 _ 전처리 : PCA를 통해 feature 간소화



20개의 Feature를 비슷한 카테고리 별로 묶어서 PCA하여 최종 9개의 Feature 생성

최종 변수명

분류	자료명
인구수	30~59세 인구
교통량	평균 버스 이용량(명)
수요량	사업체 수, 외국인 수
인프라	인근 음식점 매장 수, 인근 편의점 수, 인근 대형마트 수
경제	23년 공시지가(원/m), 배달 금액 총수
비만도 분포	비만도 분포
1-2인 가구	1~2인 가구 수
3인 가구 이상	3인 가구 이상
주택 수	아파트 수, 일반 주택 수



최종으로 선정된 feature는 상관관계 0.7 이하로 다중공선성이 없다고 판단



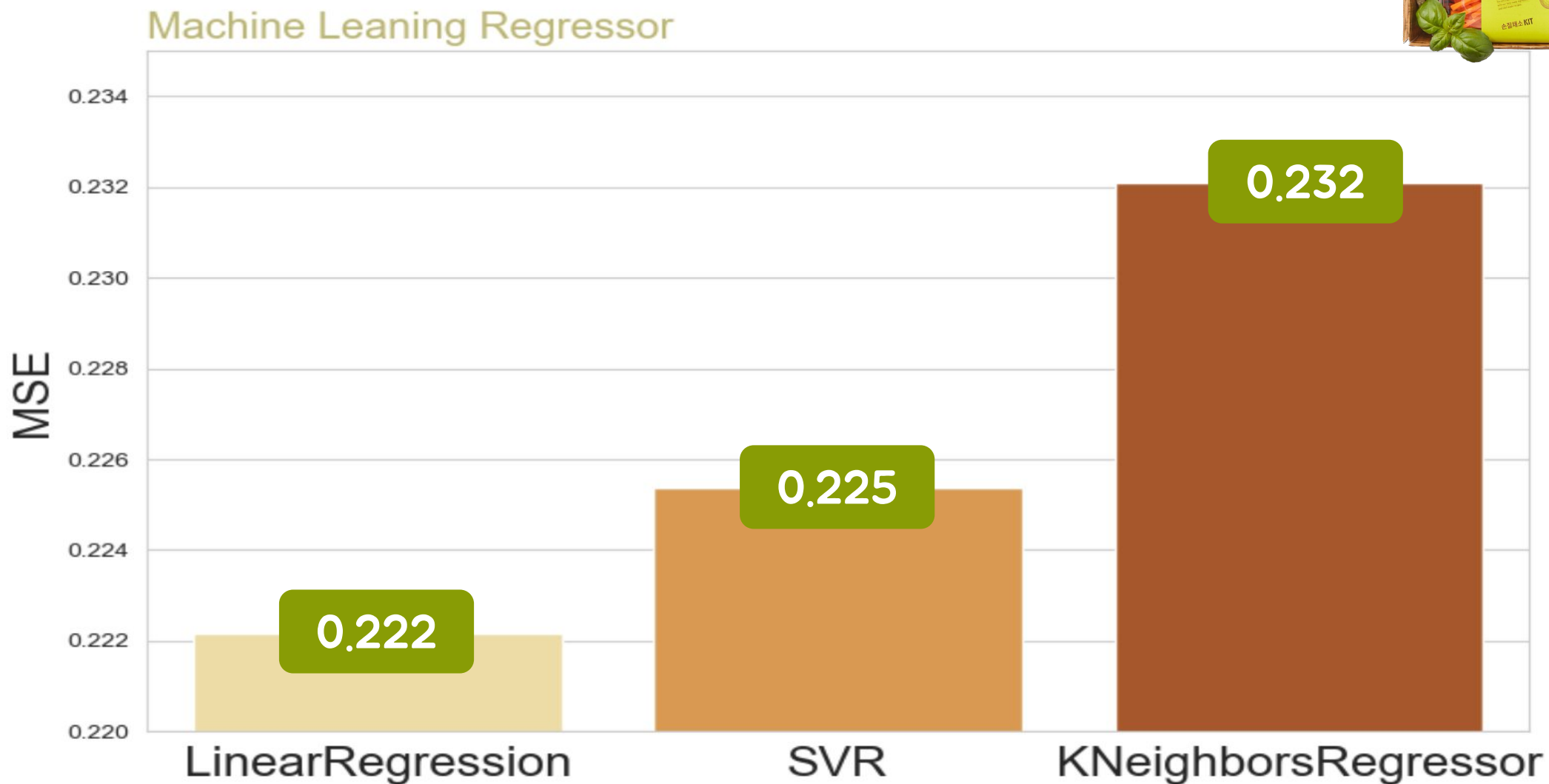
03

데이터 분석

모델 선정 | 회귀 모델



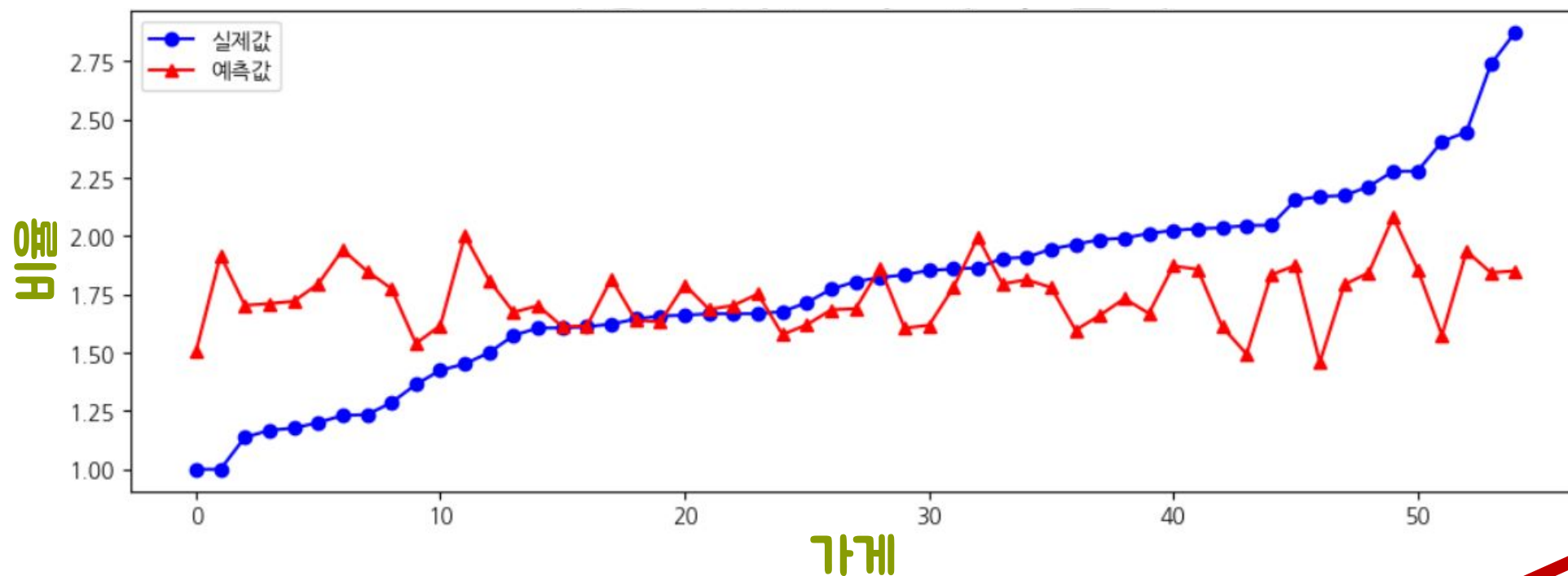
03 데이터 분석 _ 머신러닝 : 회귀 모델



03 데이터 분석 _ 모델 선정 : 다중회귀분석



다중회귀분석 예측 결과



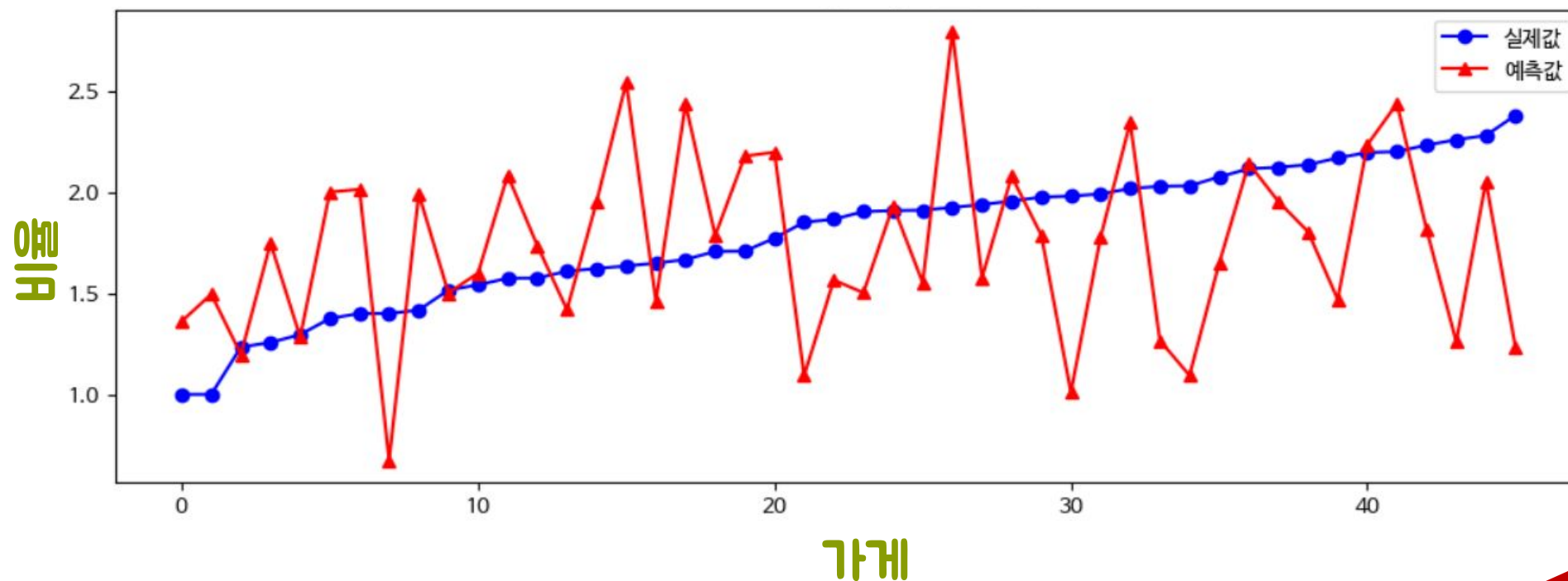
훈련 데이터 점수 : 0.0967
검증 데이터 점수 : -0.0847
MSE : 0.222



03 데이터 분석 _ 모델 선정 : SVR



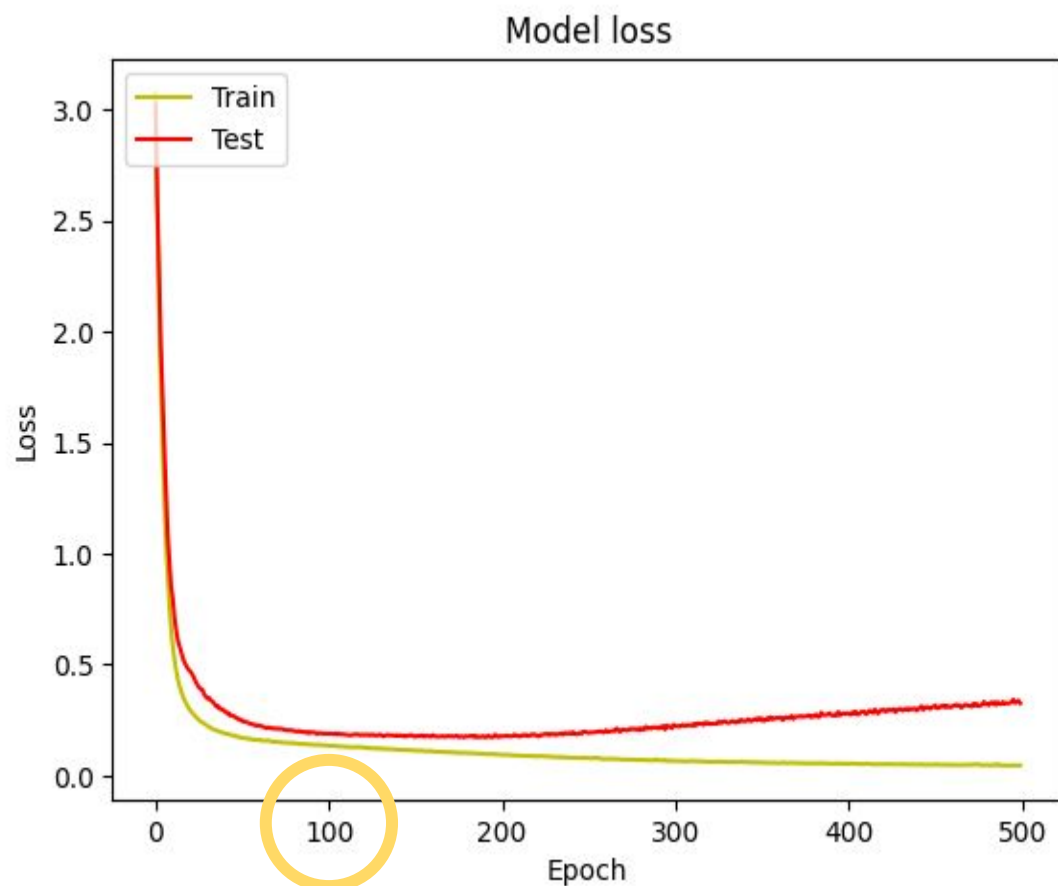
SVR 예측 결과



훈련 데이터 점수 : 0.8617
검증 데이터 점수 : -1.3557
MSE : 0.225



03 데이터 분석 _ 딥러닝



번호	예측 값	실제값
98	2.082180	1.424242
97	1.849665	1.666667
161	1.966502	1.984375
69	1.427585	1.500000
173	2.011337	2.045714
45	1.733741	1.821618
16	2.607004	1.772152
51	1.617770	1.000000
35	1.385016	2.045045
82	1.699690	1.666667

최대 오차
0.834852
Loss : 0.1417





04

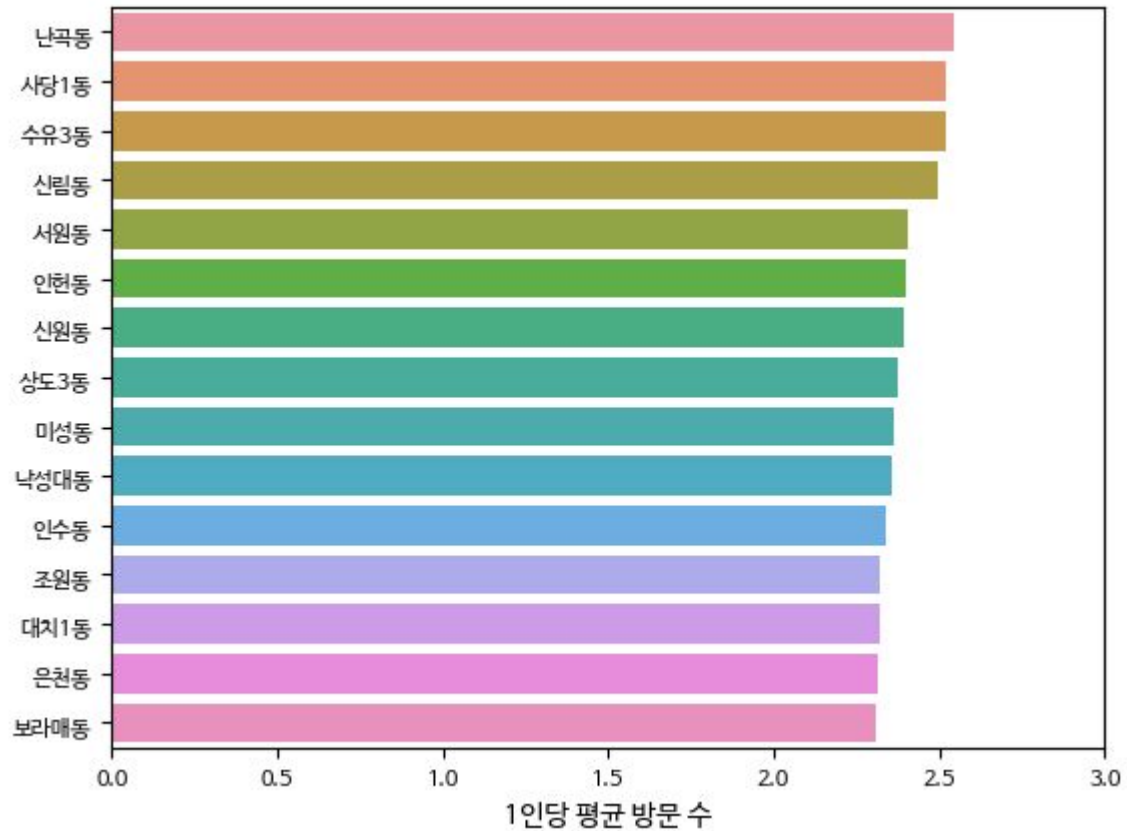
결론

최종 입지 선정 | 한계 및 향후 연구계획

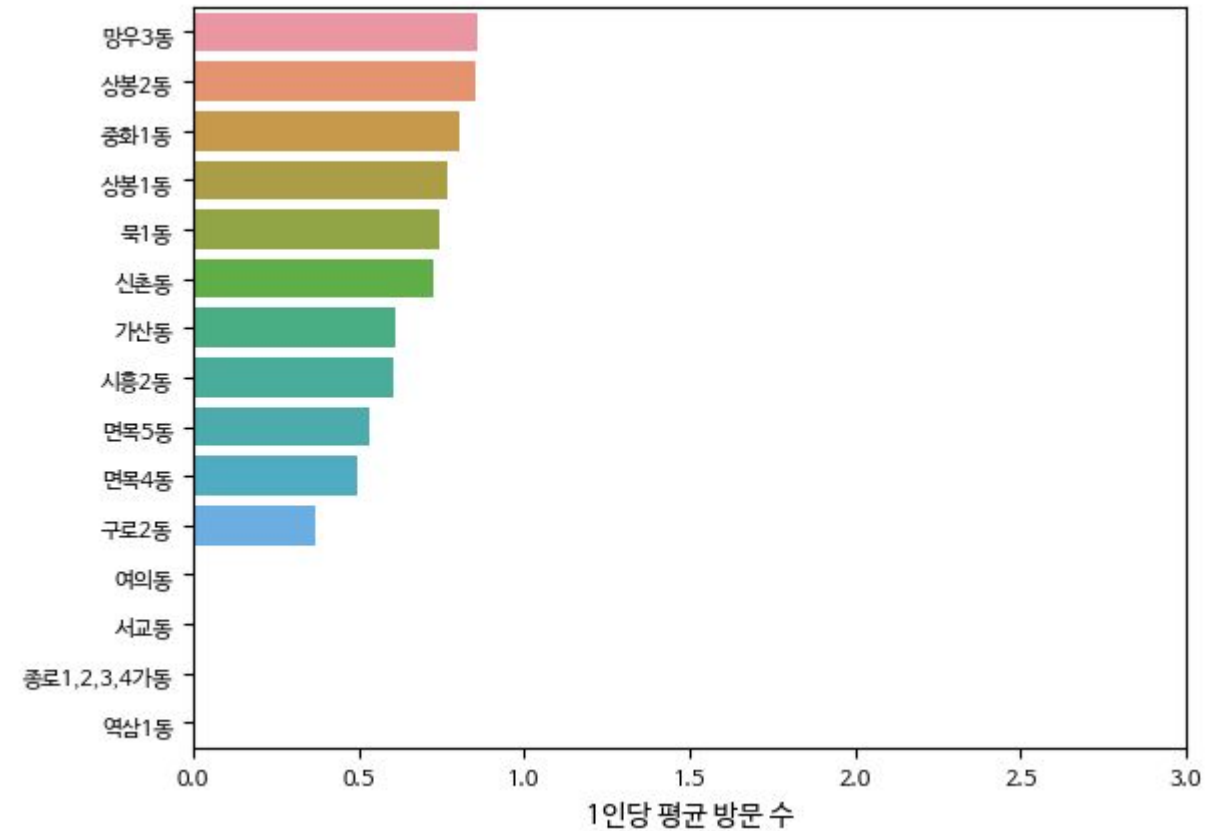
04 결론 _ 최종 입지 선정



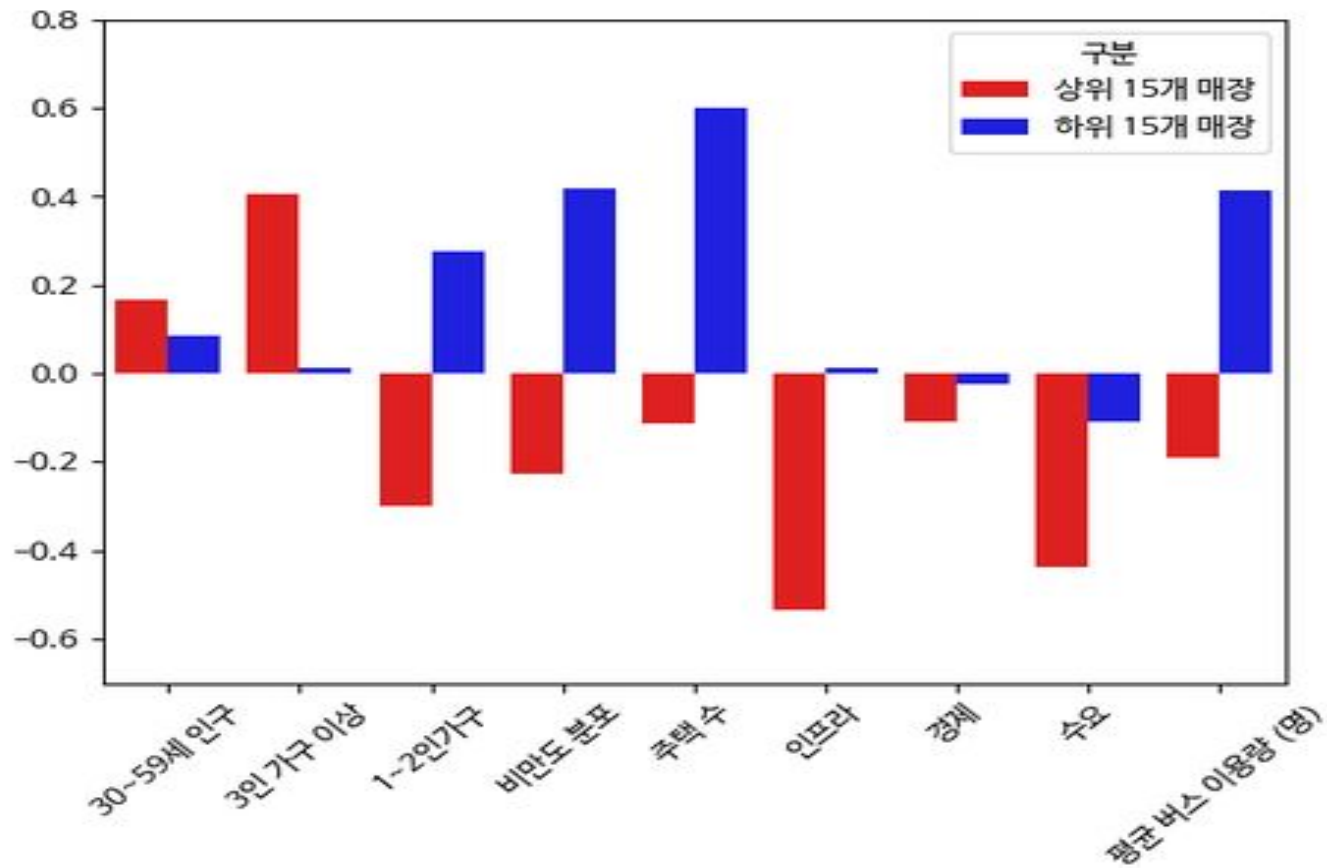
밀키트 입지요인 분석 상위 15곳 선정



밀키트 입지요인 분석 하위 15곳 선정



04 결론 _ 최종 입지 선정



최종 입지 선정을 위한 주요 지표



인프라

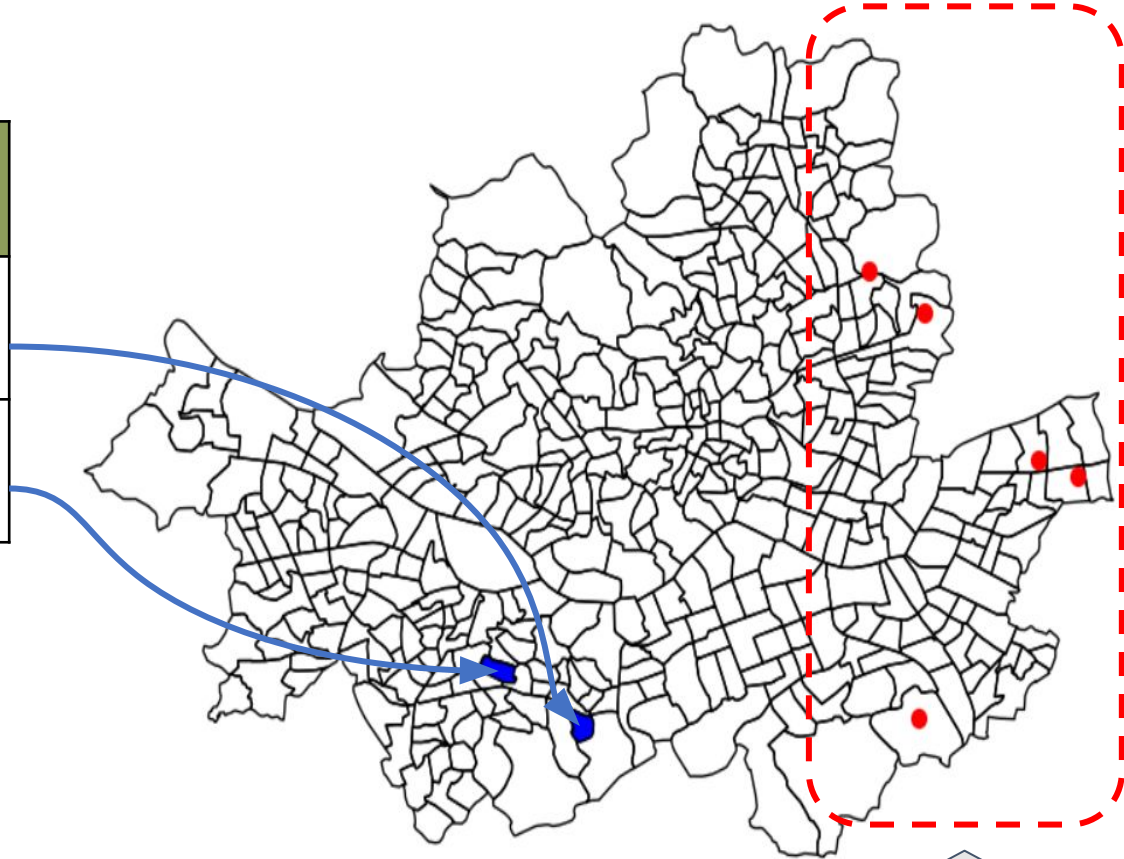


평균 버스 이용량

04 결론 _ 최종 입지 선정



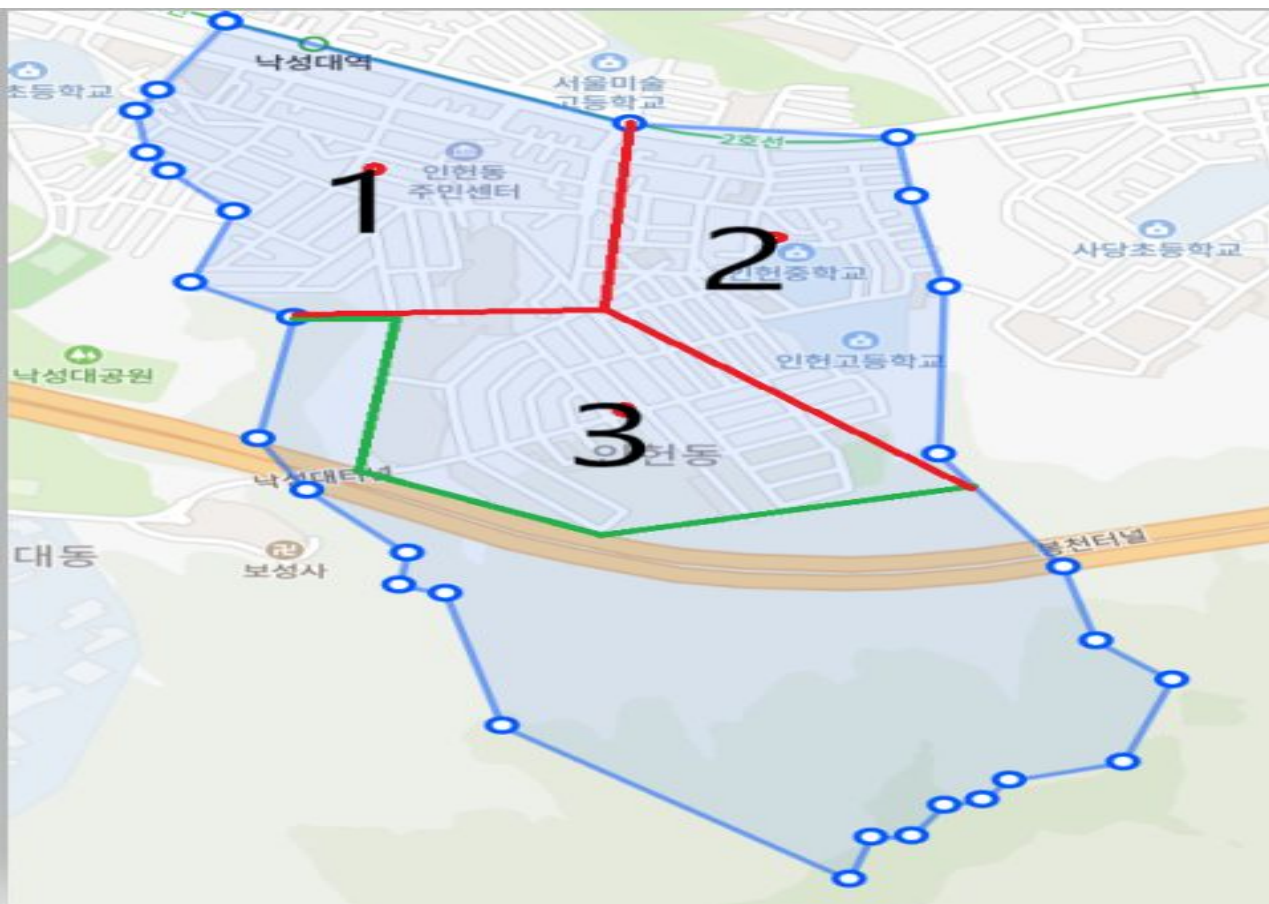
	평균 버스 이용량(명)	인프라
인현동	-0.467519595	0.131015668
보라매동	-0.282257932	0.192040454



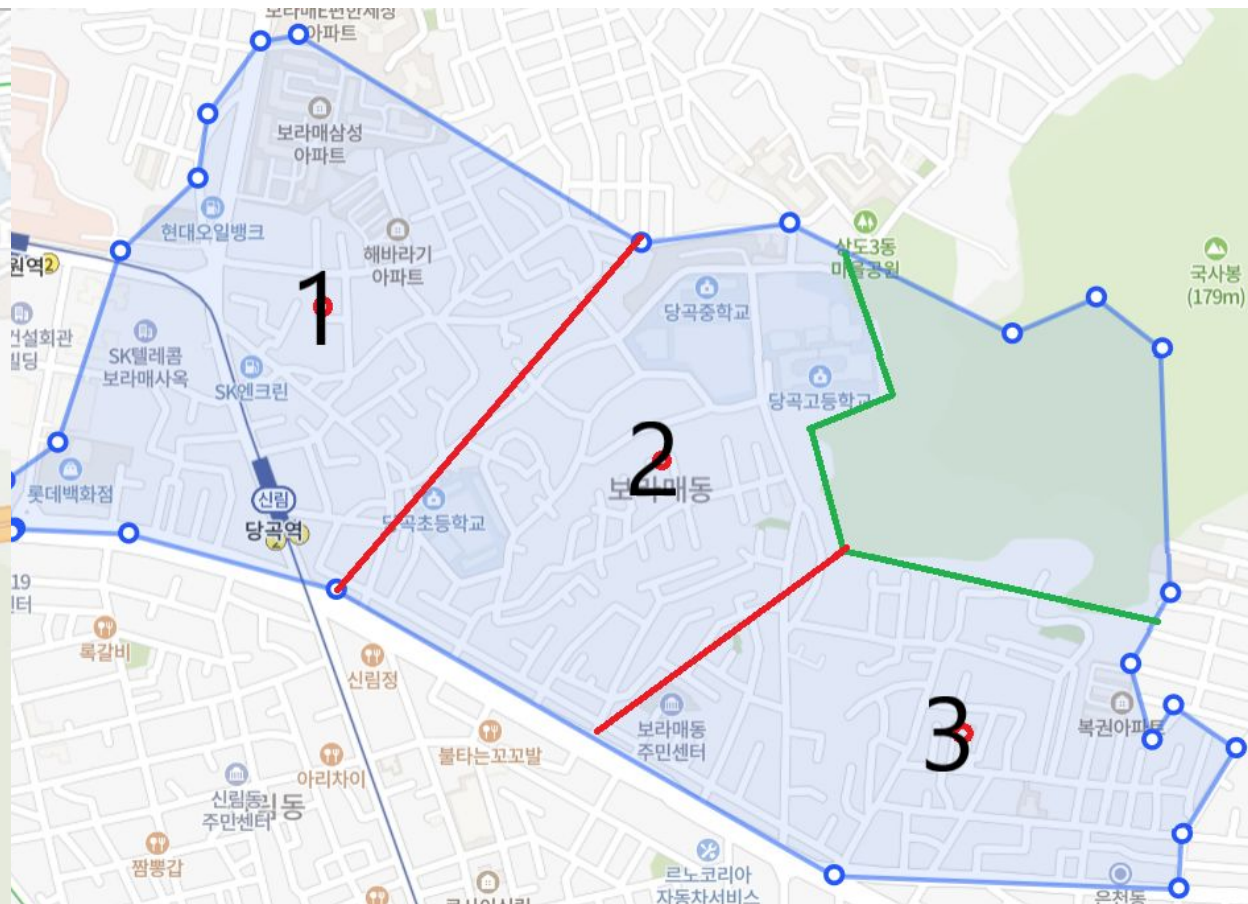
훈련 데이터 우수 점포

04 결론 _ 최종 입지 선정

인현동



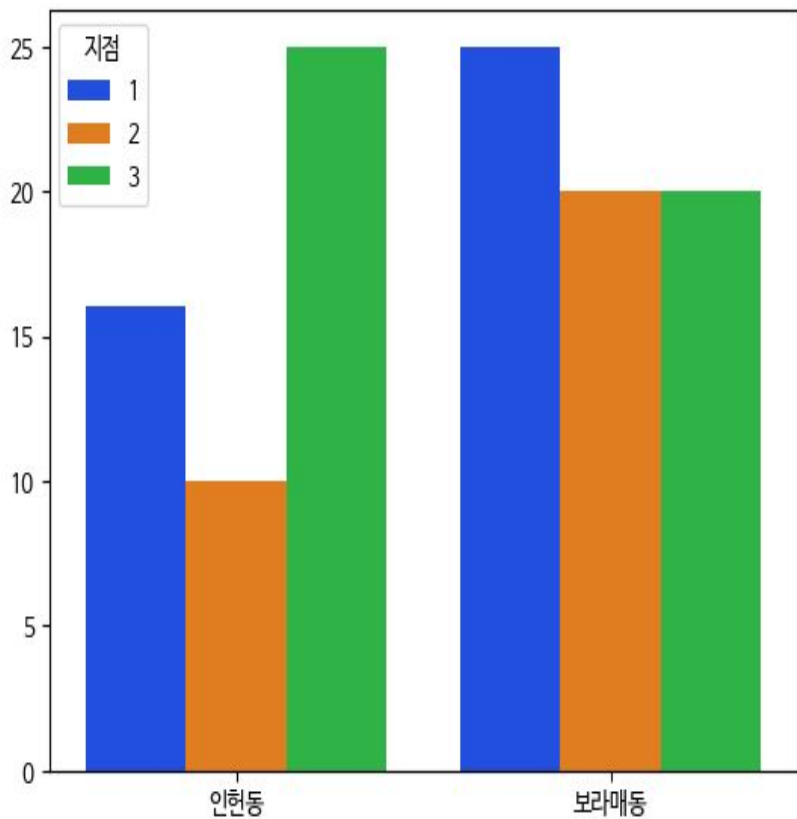
보라매동



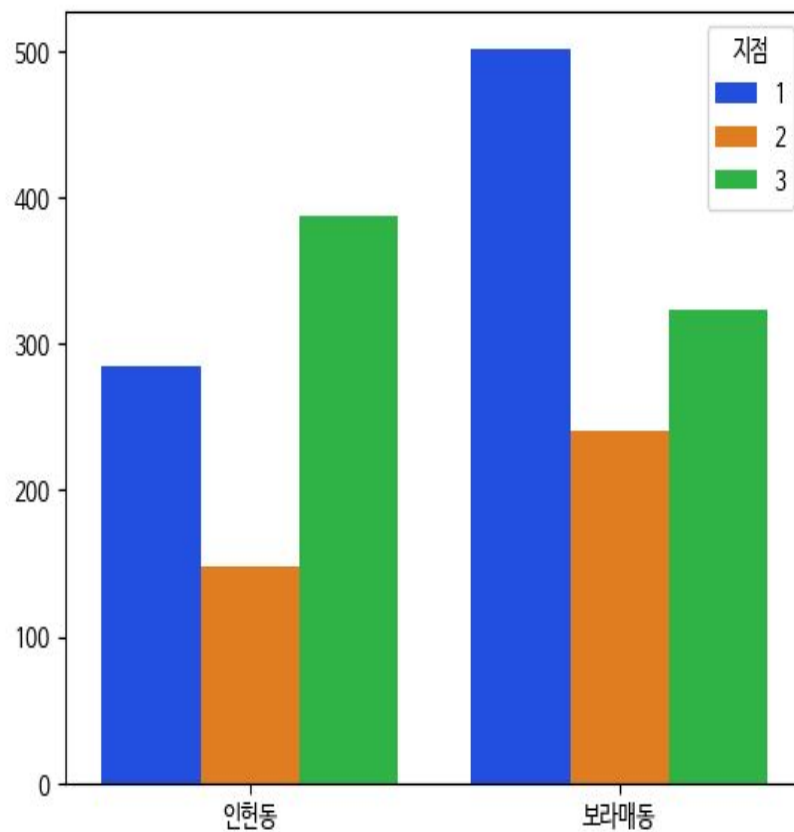
04 결론 _ 최종 입지 선정



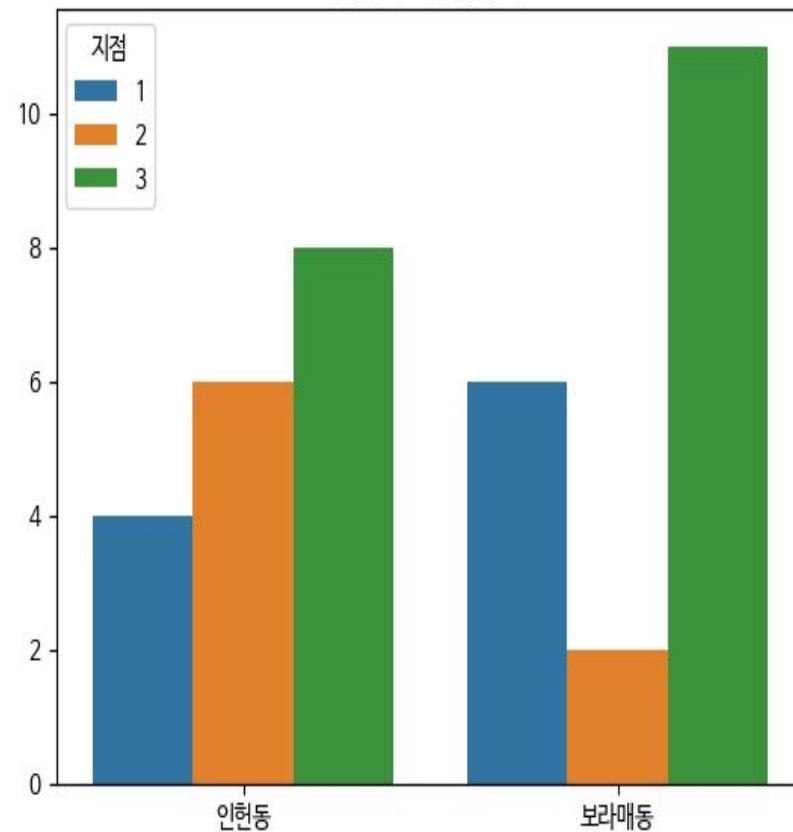
인근 편의점 개수



인근 음식점 개수



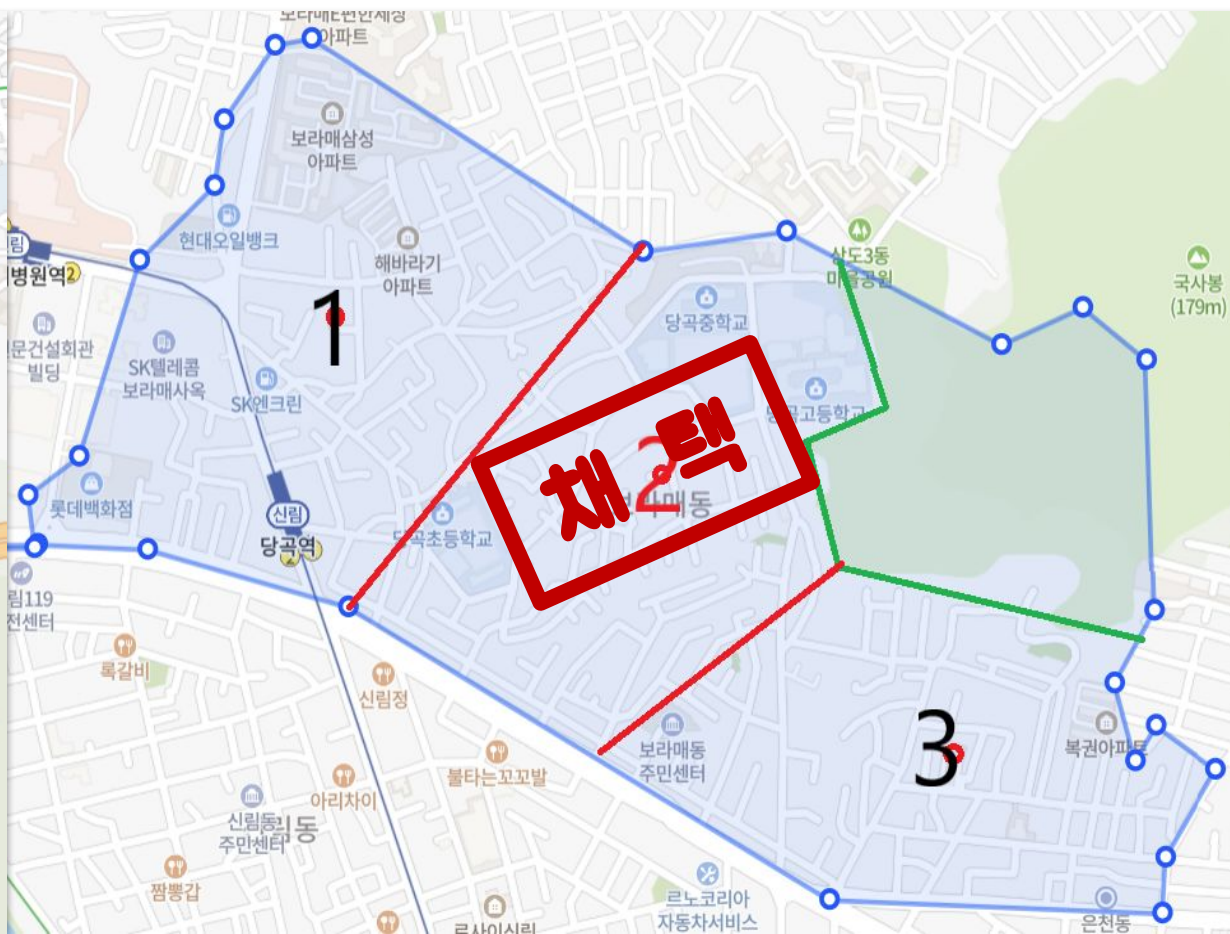
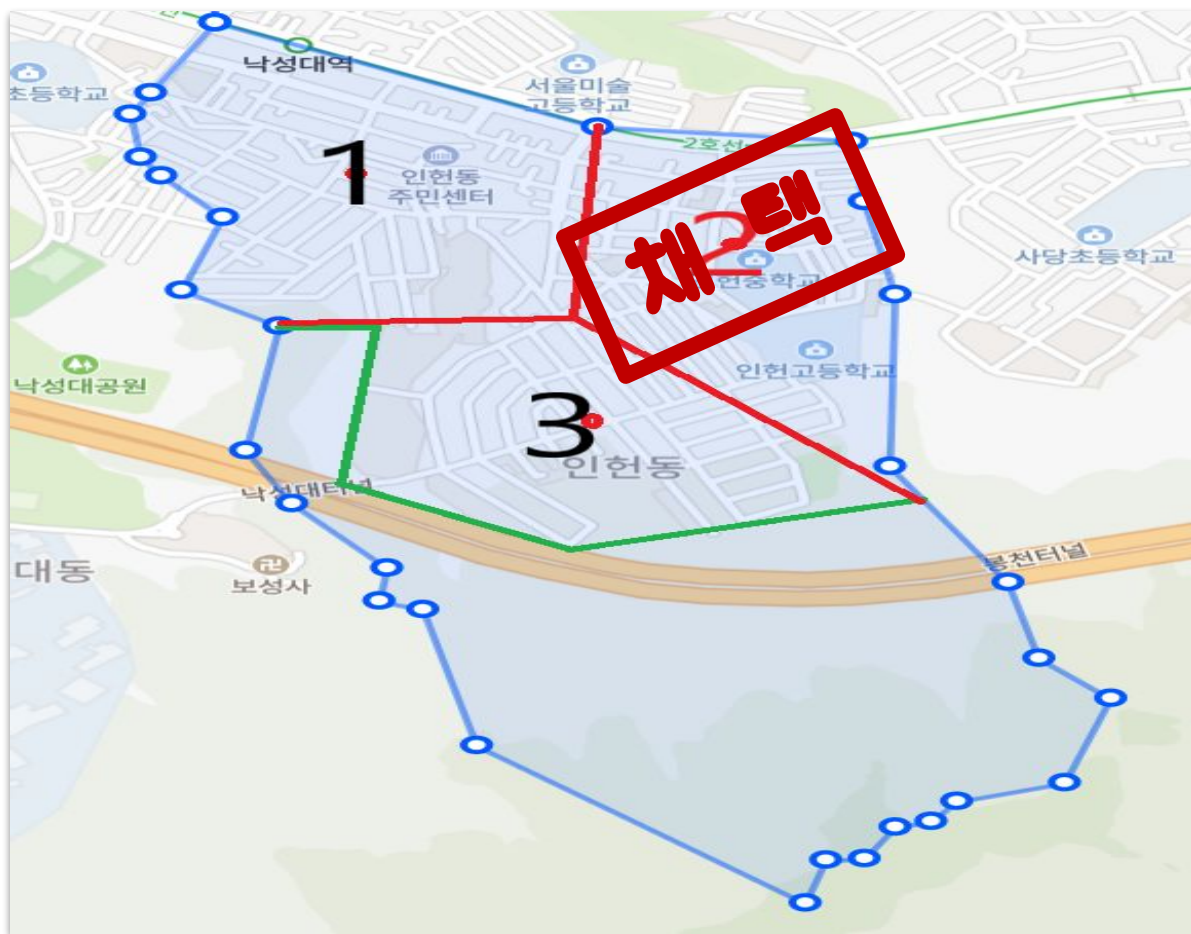
인근 버스 정류장 수



04 결론 _ 최종 입지 선정

인현동

보라매동



04 결론 _ 활용 방안

지역별로 모델 수립이 가능



- 서울시를 벗어나 비슷한 지역 특징을 가진 다른 도시에도 적용 가능 (부산, 인천 등)
- 대량의 데이터가 구축이 된다면 전국적으로도 활용 가능

다른 도메인에 적용 가능



- 유사한 특성을 무인 카페, 아이스크림 매장 등에도 활용 가능



04 결론 _ 한계 및 보완점

한계

- 각 점포의 상세한 매출액 데이터에 접근 불가하였음
→ **타겟 선정에 주관성 들어감**
- 밀키트 산업이 코로나19 이후 신흥산업이기 때문에
활용할 수 있는 데이터 표본이 적음
→ **학습부족**
- 최신화된 데이터의 부재(인구수, 교통량, 밀키트점포 등)
→ **평균치나, 과거 데이터로 추정할 수 밖에 없음**

향후 연구 시, 보완할 사항

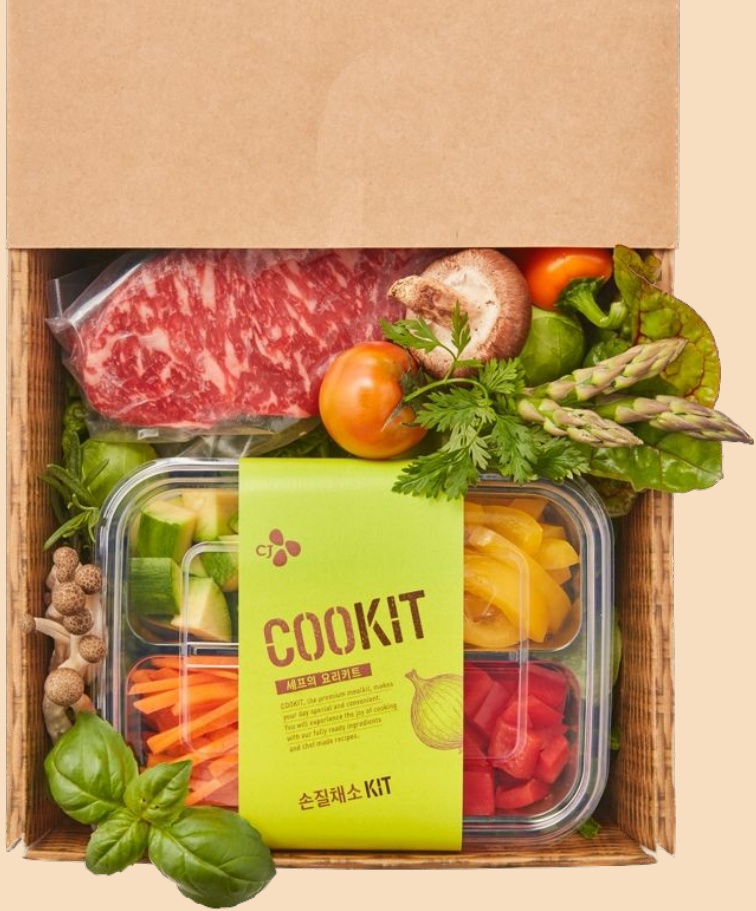
- 실제 매출액 데이터를 활용하여 분석한다면,
본 연구보다 상대적으로 더 정밀한 예측이 가능할 것
- 향후 밀키트 점포가 더 생겨, 활용할 표본이
증가한다면
훈련 모델의 성능 향상을 기대할 수 있음





Q & A





THANK YOU

