

# 2025년 데이터 문제해결은행 활용 경진대회 발표

CoxMos / 강규태

## 데이터레시피개발부문

- 신규 창업자와 정책 담당자를 위한  
업종별 지속성장지수 예측 모델



과학기술정보통신부



한국데이터산업진흥원

# 문제 정의 및 배경

## 대학가 상권 변화 속도에 대한 경험적 관찰 및 패턴

“노브랜드 버거(체인점)에서 인형뽑기 매장으로의 교체,  
탕후루 매장의 단기간(약 2년)만에 폐업,  
포토부스(예: 인생네컷) 매장의 급증 후 감소  
등의 사례를 통해 학기마다 대학가 상권이 빠르게  
변화하고 있음을 체감했다.”



유행 기반 업종은 소멸이 빠르다.



식당·프랜차이즈는 비교적 장기간 유지된다.



대체 가능한 업종일수록 경쟁이 치열하다.



**이 변화는 대학가 뿐 아니라 다른 곳에서도 나타나는 특징이지 않을까?**

- 서울 전체 상권에서도 같은 패턴이 나타날까?
- 업종별 지속성장 가능성을 데이터로 예측할 수 있을까?

# 문제 정의 및 배경



## 서울시 상권 폐업률

### 쿠키뉴스 “지원 많아도 폐업률 높아...서울 청년 창업 ‘딜레마’”

노유지 기자 2025.10.18

서울의 청년 창업 기업 5년  
생존율이 약 34.7%에 불과하며,  
30대 이하 청년 대표 기업은 약  
29.4%에 그쳤다.

### 문화일보 “작년 100만명 폐업·서울상가도 두자릿 수 공실률 ... ‘상권 위기’”

이예린 기자 외 3인 2025.06.26

2024년 폐업자 수가 100만 명에  
육박하고,  
음식업·소매업·서비스업 등에서  
폐업이 집중되었으며 상가  
공실률도 증가하고 있다.

### 오마이뉴스 “‘폐업률 160% 충격’ 서울 은평구, 문닫는 가게가 더 많다”

박은미 기자 2025.02.24

서울 은평구에서 창업 대비 폐업  
비율이 160%에 달하며, 신규  
창업보다 폐업이 더 많은 지역  
불균형 현상이 발생하고 있다.

# 데이터 및 접근 방식

## 활용 데이터 서울열린데이터광장 - 서울시 상권분석서비스

**Data1 행정동별 집계**

년분기	행정동_코드	총_유동인구_수	상권_변화_지표	...	월_평균_소득_금액	총_상주인구_수	종사자수_명
20191	11110515	3668225	LH		4015674	13047	5656
20191	11110530	4628956	HH		5208463	9224	NA
⋮							
20252	11110515	3549035	LH		3780222	11328	NA
20252	11110530	4099841	HH		5007768	8992	NA

**Data2 행정동 및 업종별 집계**

년분기	행정동_코드	업종	점포수	개업점포수	폐업점포수	...	당월매출액
20191	11110515	관리서비스	59	7	1		71631292
20191	11110515	교육서비스	49	1	5		62884924
⋮							
20252	11110515	관리서비스	85	3	1		132753531
20252	11110515	교육서비스	57	0	0		250511398

# 데이터 및 접근 방식

## 활용 데이터 서울열린데이터광장 - 서울시 상권분석서비스

	데이터 타입	비고
년분기	Int64	날짜(년도분기)
행정동_코드	Int64	법정동 기준 8자리 코드
총_유동인구_수	Int64	해당 분기 유동인구 합계
상권_변화_지표	String	HH/LH/HL/LL (상권활발도 코드)
월_평균_소득_금액	Int64	거주민 평균 월소득(단위: 원)
총_상주인구_수	Int64	행정동 내 실제 거주 인구 수
종사자수_명	Int64	해당 분기 내 사업체 종사자 수(단위: 명)
⋮		
업종	String	97개의 세부업종을 14개로 분류
점포수	Int64	해당 분기 점포 총 개수(폐·개업 포함)
당월매출액	Int64	해당 분기 총 매출액(단위: 원)

외식업  
 의료서비스  
 여가서비스  
 주거서비스  
 식제품  
 학업품  
 전자/여가

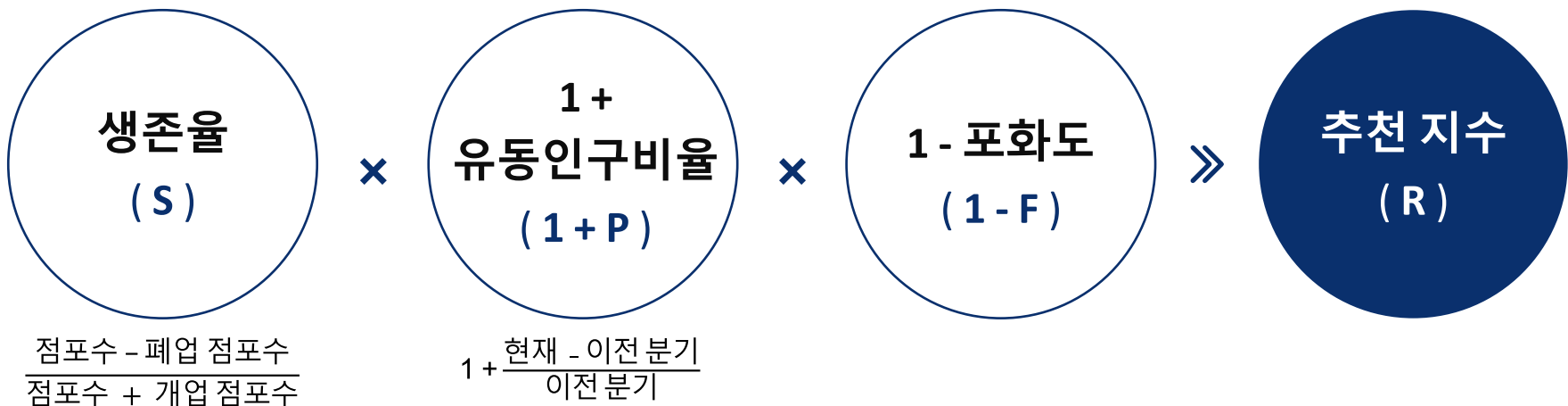
교육서비스  
 법률서비스  
 관리서비스  
 임대서비스  
 의류  
 화장/미  
 운송

# 데이터 및 접근 방식

## 파생 변수 생성: 추천 지수(R)

추천 지수(R)를 구성하는 3가지 핵심 요인

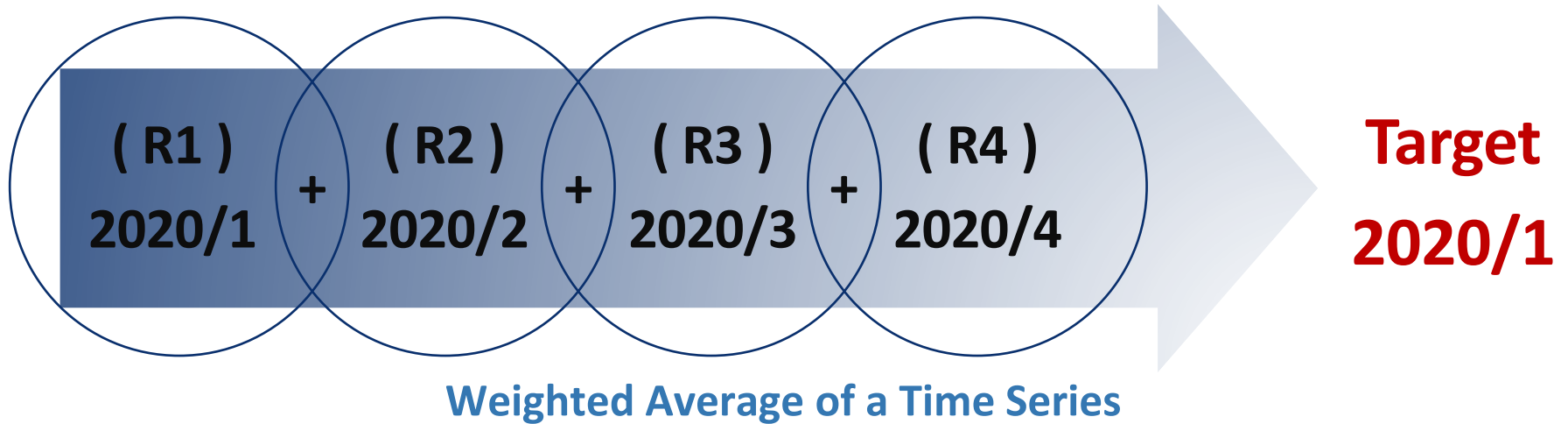
- 01 생존율(S) : 상권의 “지속 가능성” 평가
- 02 유동인구 변화율(P) : “수요 성장 가능성” 분석
- 03 포화도(F) : “경쟁 강도 및 차별화 가능성” 평가



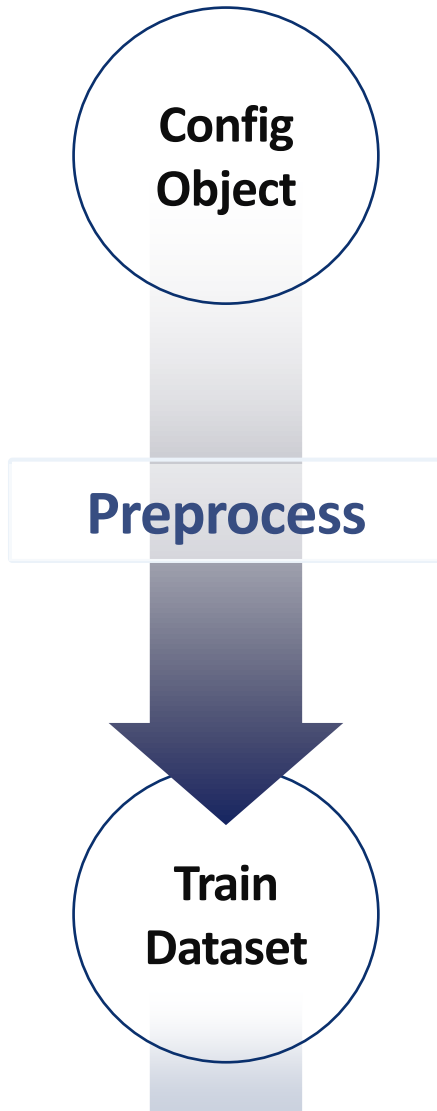
# 데이터 및 접근 방식

## Target 변수 설정: 성장 지수(T)

2020년도 1분기의 성장지수



# 분석 또는 설계 과정



## Main Script

**Config > Preprocess > Training > Prediction > Visualization**

└─ (Data\_vals, Model\_vals, Learning\_vals, Visual\_vals)

## Preprocess



### Data Cleaning

- 탐색적 데이터 분석(EDA, Exploratory Data Analysis)
- 결측치와 이상치 탐지 및 처리



### Feature Engineering

- 파생 변수 생성, 성장 지속 지수 설계 및 추출

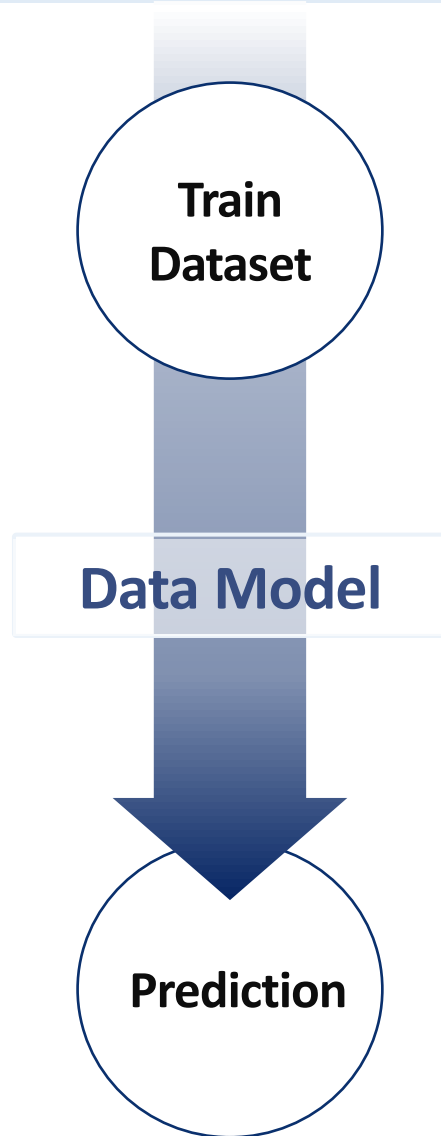


### Data Transformation

- 상관관계 기반 데이터 정제, 데이터 스케일링, 누설 변수 제거



# 분석 또는 설계 과정



## Training / Prediction



### Modeling

- Light GBM: 업종별 모델 개별 선언 및 성능 개선 디버깅 진행  
(ACC 91% OVER)



## Visualization



### Logger

- 데이터 기록: Accuracy, Feature Importance, Total Data Matrics

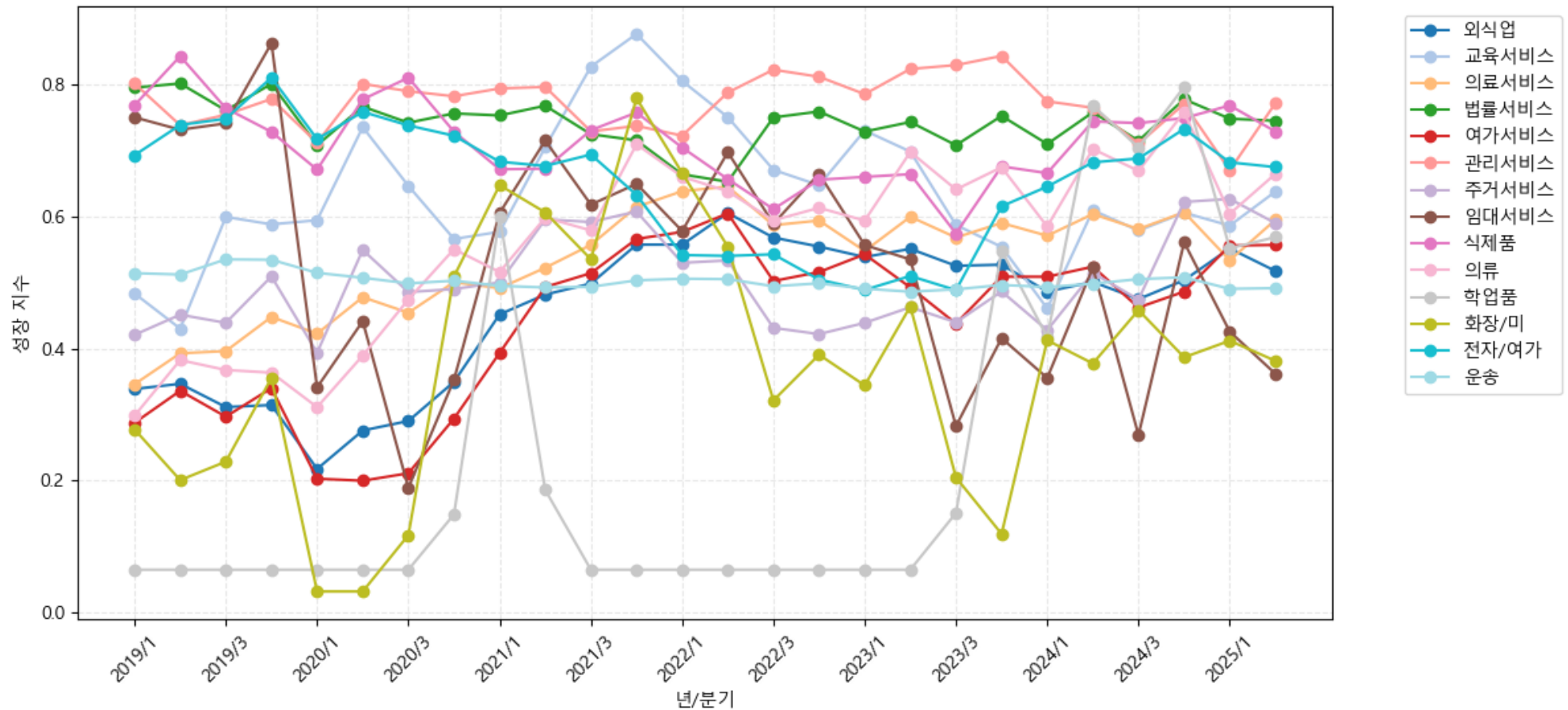


### Graph

- 시각화: Feature Graph, Trend chart, Geospatial heatmap

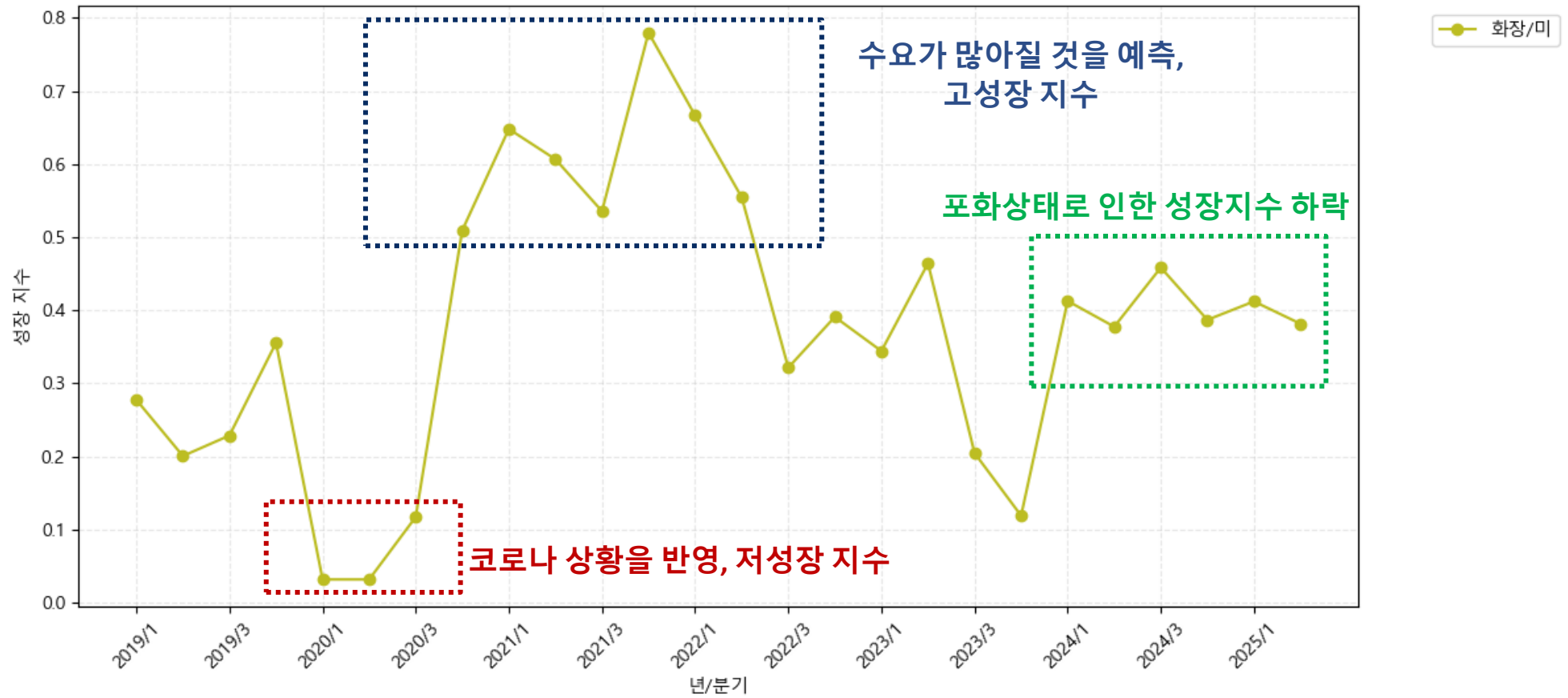
# 주요 결과

## 동대문구 회기동 업종별 성장지수 추세



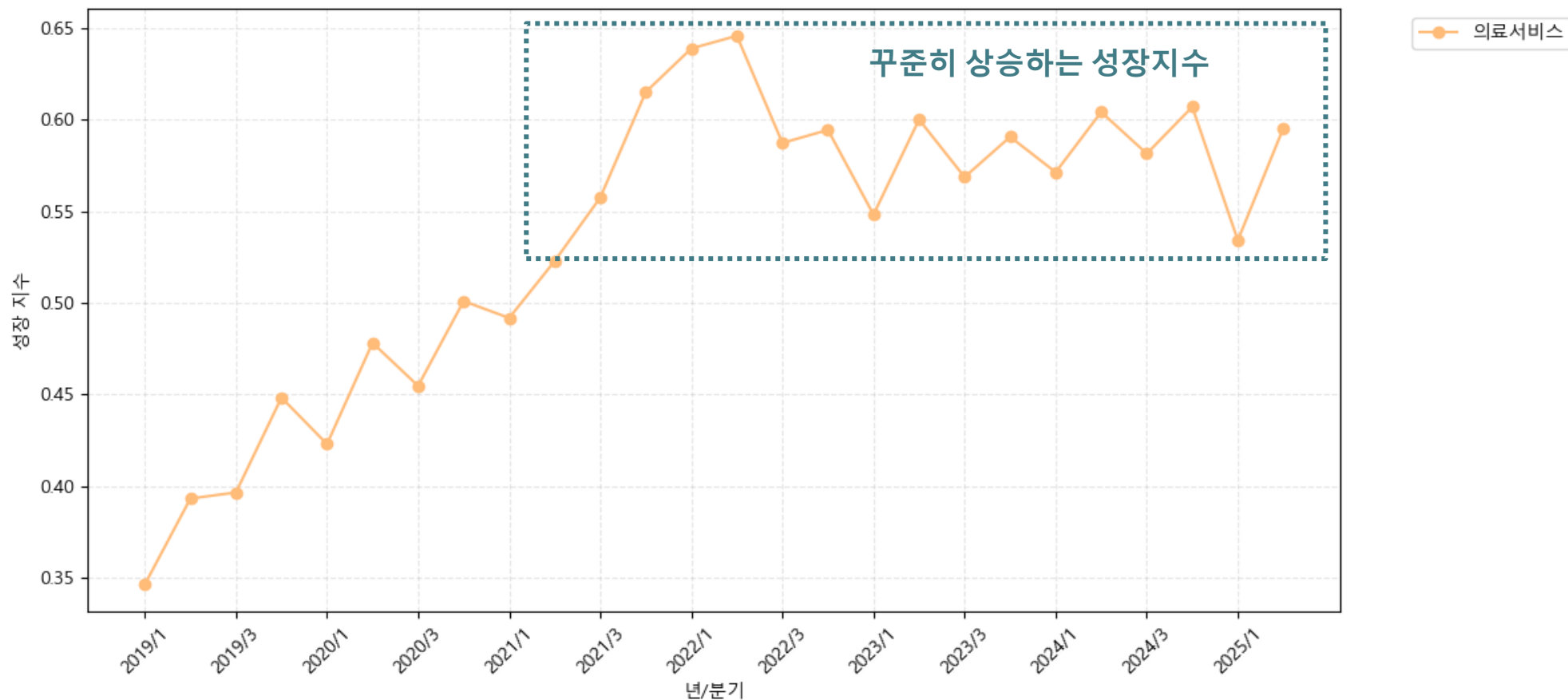
# 주요 결과

## 동대문구 회기동 성장지수 추세 – “화장/미”



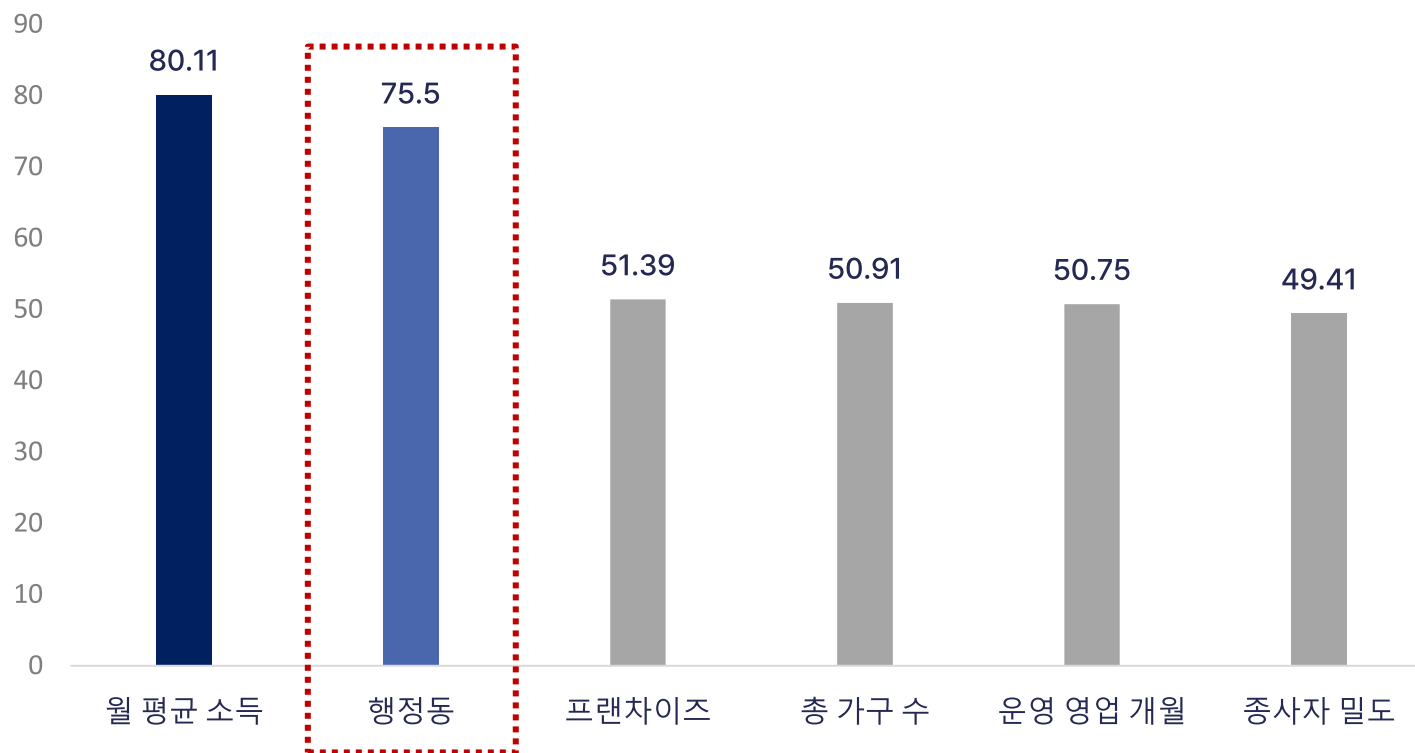
# 주요 결과

## 동대문구 회기동 성장지수 추세 – “의료서비스”



# 주요 결과

## 의료 서비스 | Feature Importance



경희대학교, 경희의료원

# 적용 및 확산 가능성

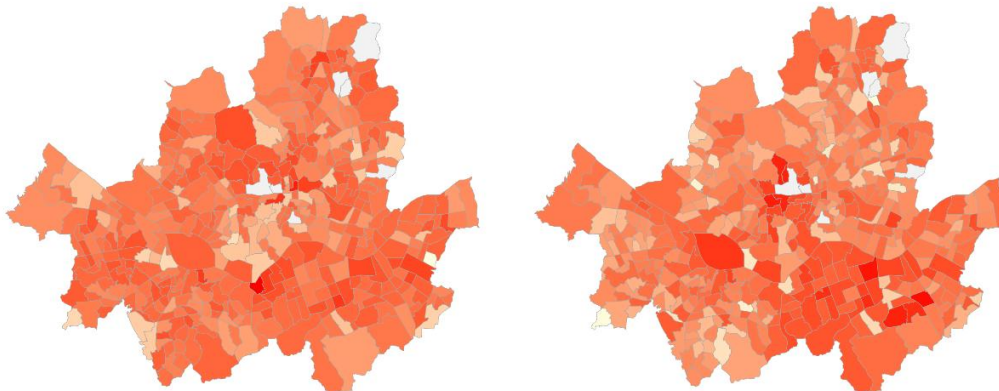
년분기	행정동 코드	지역	외식업	교육 서비스	...	운송
20251	11680660	강남구_개포1동	0.22802	0.36642		0.90961
20251	11680670	강남구_개포2동	0.80623	0.74738		0.67860
⋮						
20252	11680660	강남구_개포1동	0.20154	0.32913		0.90684



서울시 행정동 단위 업종별 지속성장지수 예측 단계구분도

교육서비스

외식업



업종별 지속성장지수 

## 분석 도구로써 뛰어난 확장성

지리 기반 시각화로 생존율·요인을 파악하고, 유연한 모델로 타 지역·장기 모니터링까지 확장 가능하다.

## 폐점 위험 예상 지역 파악

분석 결과를 기반으로 폐점 위험이 높은 구역을 조기에 식별해 선제적 대응이 가능하다.

## 창업 계층 임대료 지원 및 홍보 강화

위험 지역에 대한 맞춤형 임대료 지원·홍보 전략을 마련해 상권 안정화와 소상공인 보호에 기여한다.

## 리스크 평가 모델 적용 가능

창업 대출·점포 보험 등 금융상품의 리스크 평가와 데이터 기반 지표로 활용할 수 있다.

# 결론 및 제언

## 후속 연구의 필요성 및 지속 활용 방안

### 온라인 기반 정성적 데이터 반영



x(구 트위터)  
트렌드.인스타그램  
해시태그를 활용한  
단기 유행 키워드 추출  
및 성장지수 반영



리뷰 데이터의 긍·부정  
분석 및 업종별 핵심  
평가 요소 반영  
(맛·친절·성적 향상 등)



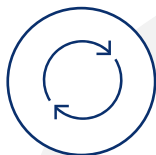
인스타그램 홍보  
게시물 기반 초기·지속  
홍보 활동 분석 및  
지속가능성 변수 반영

# 결론 및 제언

## 후속 연구의 필요성 및 지속 활용 방안

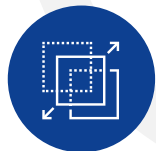
### 지속적인 업데이트 체계 구축

분기별로 축적되는 데이터로 모델을 지속 업데이트하면, 변화 추세와 위험 신호를 상시 파악하는 모니터링 시스템으로 확장 가능



### 업종 및 상권 분류 정규화

현재 임의 기준으로 그룹화된 업종·상권을 세분화·재구조화 함으로써 예측의 현실성과 활용도를 높일 수 있음



### 전략적인 의사결정 지원 모델

현실적인 성장 지속 지표를 바탕으로 창업·정책·투자 등 실무에서의 전략적 의사결정을 돕는 기반 제공



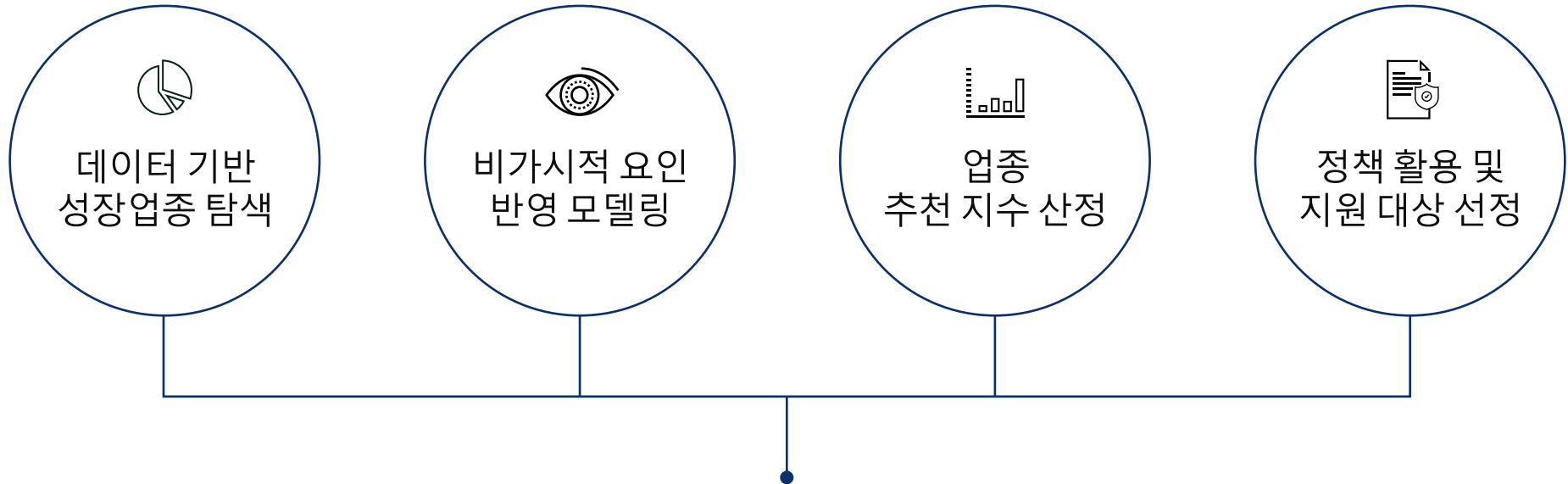
### 온라인 기반 트렌드 반영 모델링

SNS·검색량 등 온라인 데이터를 활용해 유행 업종의 민감도를 측정하고, 단기 트렌드 업종의 지속 전략을 제안하는 확장 연구



# 결론 및 제언

## 성장 업종 예측 모델의 주요 시사점



데이터 기반으로 비가시적 요인을 반영해  
미래 성장 업종을 추정·추천하는 분석 모델

# 감사합니다



과학기술정보통신부



한국데이터산업진흥원