

라즈베리분석단

(LH) 지식산업센터 공실률 최소화를 위한 해결방안 및 분석결과 도출

2단계 타원형 EFA 및 TSRI를 활용한 신도시 공실률 최저지점 예측

목차

01

분석 개요

- 분석 배경
- 분석 프로세스

02

성남시 현황분석

03

성남시 주요 요인 도출

- 2단계 타원형 EFA
- TSRI

04

신도시 적용

- 균등분할
- 부분요인배치
- 최적위치 및 면적산출

05

신도시 입지/면적 추천

- 제조업 입지·면적 추천
- 서비스업 입지·면적 추천
- 기대효과 및 추가 개선점

01. 분석 개요

- 분석 배경
- 분석 프로세스

분석 배경

입지가 좋은 곳, 상권이 좋은 곳은 무조건 선호?

'공실 지옥'이 된 지산

아파트형 공장은 말 그대로 아파트 같은 건물에 공장이 들어선 형태입니다. 작은 공장만 들어갈 수 있으니, 자연스럽게 제조업처럼 넓은 공간이 필요한 업종은 제한됐습니다. 하지만 수도권엔 안성맞춤이었습니다. 수도권은 어차피 대형 공장에 대한 규제가 심했기 때문에 지산이 좋은 대체제로 각광 받았습니다.

일반 오피스 건물과 다른 입지조건도 지산의 매력으로 꼽혔습니다. 3종일반주거지역이나 준주거지역에 지산을 건설할 수 있기 때문인데요. 지역 입장에서는 활용도가 낮은 땅에 일자리가 생기는 건물을 지을 수 있다는 점이 매력적으로 다가왔습니다. 신도시 역세권에 지산이 많이 들어선 이유입니다.

하지만 문제가 있었습니다. 지산이 들어선 곳들이 주거 입지는 비교적 좋지만, 원래 기업들이 있던 자리는 아니었습니다. 지산이 들어서는 규모와 속도에 비해 입주할 기업은 턱없이 부족했습니다. 공실률이 높아질 수밖에 없었던 것이죠. 경기도 신도시 중에선 고양시 향동지구, 하남시 미사강변도시 등이 이런 경우에 해당됐습니다. 남양주 다산신도시도 주거환경은 좋지만, 과잉공급된 지산은 공실 문제에 허덕이고 있습니다. 평택 고덕신도시도 상황은 비슷합니다. 고덕은 삼성전자와 연관된 중소 기업이 입주할 거란 기대감에 많은 지산이 건설됐지만, 지금은 '공실 지옥'으로 불립니다.



MBC뉴스. "'공실 지옥'된 '지식산업센터'..현장 돌아보다 '경악' [뉴스.zip/MBC뉴스]." MBCNEWS, 2024년 1월 30일

한동인. "'공실 지옥' 된 지식산업센터...왜?" 뉴스토마토, 2024년 10월 18일.

입지가 좋은 곳,
상권이 좋은 곳은

무조정답은 NO !!!

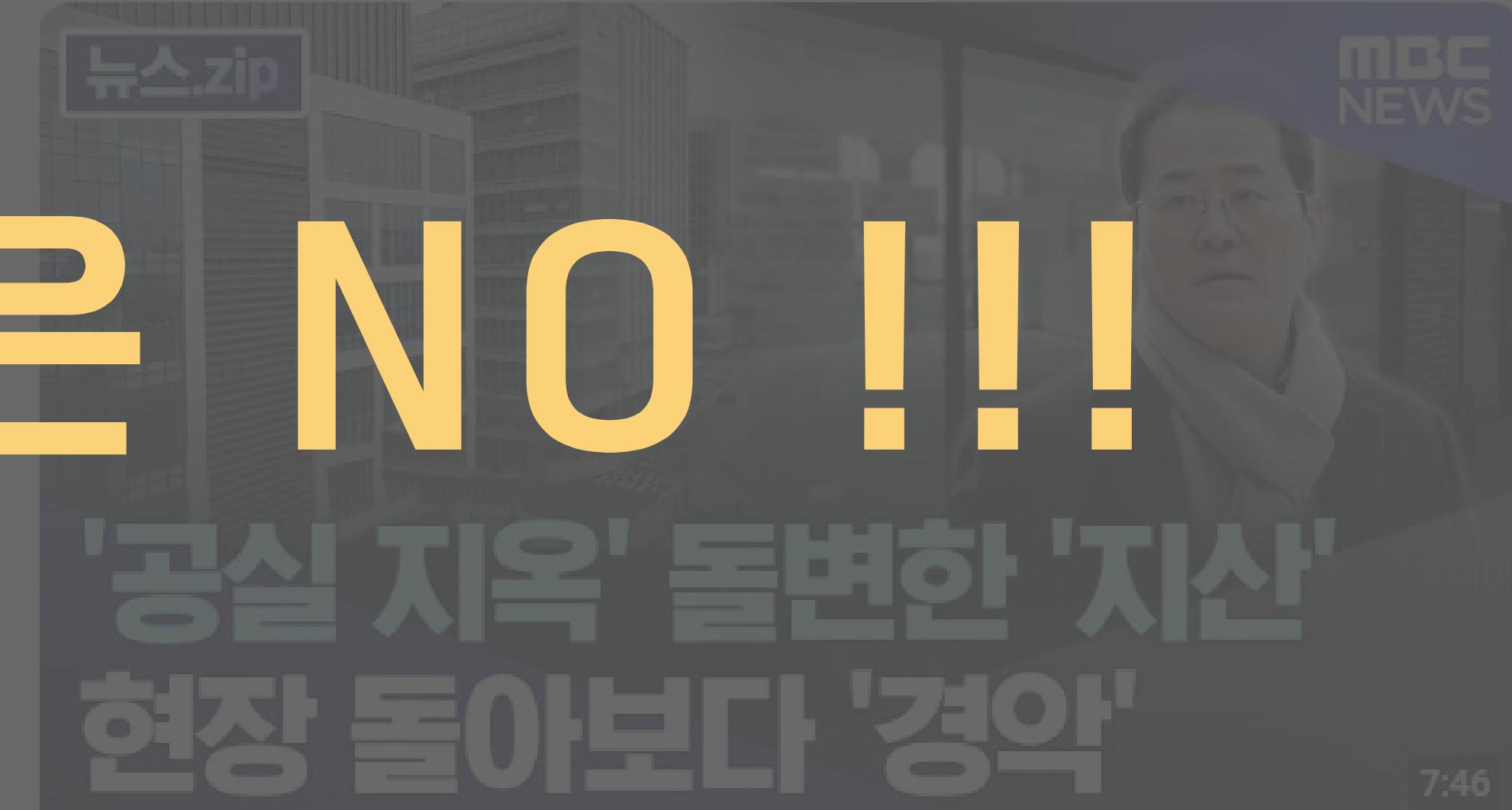
'공실 지옥'이 된 지산

아파트형 공장은 말 그대로 아파트처럼 들어선 형식으로 들어갈
있으니, 자연스럽게 제조업처럼 넓은 공간이 필요한 업종은 제한됐습니다. 하지만 수도권엔 안성맞
춤이었습니다. 수도권은 어차피 대형 공장에 대한 규제가 심했기 때문에 지산이 좋은 대체제로 각광
받았습니다.

일반 오피스 건물과 다른 입지조건도 지산의 매력으로 꼽혔습니다. 3층일반주거지역이나 준주거지
역에 지산을 건설할 수 있기 때문인데요. 지역 입장에서는 활용도가 낮은 땅에 일자리가 생기는 건
물을 지을 수 있다는 점이 매력적으로 다가왔습니다. 신도시 역세권에 지산이 많이 들어선 이유입니
다.

하지만 문제가 있었습니다. 지산이 들어선 곳들이 주거 입지는 비교적 좋지만, 원래 기업들이 있던
자리는 아니었습니다. 지산이 들어서는 규모와 속도에 비해 입주할 기업은 턱없이 부족했습니다. 공
실률이 높아질 수밖에 없었던 것이죠. 경기도 신도시 중에선 고양시 향동지구, 하남시 미사강변도시
등이 이런 경우에 해당됐습니다. 남양주 다산신도시도 주거환경은 좋지만, 과잉공급된 지산은 공실
문제에 허덕이고 있습니다. 평택 고덕신도시도 상황은 비슷합니다. 고덕은 삼성전자와 연관된 중소
기업이 입주할 거란 기대감에 많은 지산이 건설됐지만, 지금은 '공실 지옥'으로 불립니다.

한동인. "'공실 지옥' 된 지식산업센터...왜?" 뉴스토마토, 2024년 10월 18일.



MBC뉴스. "'공실 지옥'된 '지식산업센터'..현장 돌아보다 '경악' [뉴스.zip/MBC뉴스]." MBCNEWS,
2024년 1월 30일

하지만 문제가 있었습니다. 지산이 들어선 곳들이 주거 입지는 비교적 좋지만, 원래 기업들이 있던 자리는 아니었습니다. 지산이 들어서는 규모와 속도에 비해 입주할 기업은 턱없이 부족했습니다. 공실률이 높아질 수밖에 없었던 것이죠. 경기도 신도시 중에선 고양시 향동지구, 하남시 미사강변도시 등이 이런 경우에 해당됐습니다. 남양주 다산신도시도 주거환경은 좋지만, 과잉공급된 지산은 공실 문제에 허덕이고 있습니다. 평택 고덕신도시도 상황은 비슷합니다. 고덕은 삼성전자와 연관된 중소 기업이 입주할 거란 기대감에 많은 지산이 건설됐지만, 지금은 '공실 지옥'으로 불립니다.

현실은 지산이 기업 입지가 아닌 주거 입지를 기준으로 들어와있어 수요층이 만족하지 못하는 상황

하지만 문제가 있습니다. 기산이 들어선 곳들이 주거 입지는 비교적 좋지만, 원래 기업들이 있던
자리는 아니었습니다. 기산이 들어서는 규모가 커지면서 기업들은 대부분 이전했습니다. 공
실률이 높아질 수밖에 없었던 것이죠. 경기도 신도시 중에선 고양시 향동지구, 하남시 미사강변도시
등이 이런 경우를 많습니다. 남양주 하산신도시는 환경을 뛰어넘는 고급 지산입니다.
단지에 입주하고자 하는 분들이 많았지만, 경쟁률은 삼성전자와 연세대학교 등
기업이 입주할 거란 기대감에 많은 지산이 건설됐지만, 지금은 '공실 지옥'으로 불립니다.

**그렇다면 해답은 하나,
수요층의 니즈를 파악하라!**

수요층들은 무조건 좋은 입지만 바라보지 않는다.

그렇다면 그들이 진짜 원하는 것은 무엇일까?

**주거 입지를 기준으로 들어와있어
수요층이 만족하지 못하는 상황**

분석 프로세스

데이터 수집

COMPAS 기본제공데이터

성남시 유동인구 데이터
성남시 카드매출 데이터
성남시 상권정보 데이터
성남시 상가개폐업 데이터
성남시 표제부 데이터
성남시 격자데이터
....

V-WORLD DATA 외부 수집 데이터

성남시 실거래가 데이터
경기도 버스정류장 데이터
지하철역 데이터
공시지가 데이터
경기도 상권정보데이터
....

데이터 전처리

전처리 및 격자매핑

- 100m격자 > 400m 격자변환
 - 자체 알고리즘 이용
 - 교통 지수 산출
(최소이동시간, 교통수용율)
 - 400m 격자별 데이터 매핑
(상권, 교통,)
 - 표제부 데이터 EFA

EDA 및 현황분석

데이터 별 회귀분석, EDA
(종속변수 공실률)

데이터 분석 및 적용

2단계 타원형 EFA

- 자체 KMO 최적 알고리즘 및 전처리 프로세스 활용
- 2단계 타원형 EFA진행
- 설명가능한 잠재요인 도출

TSRI 및 신도시 적용

- TSRI, CF등 분석 진행
- 도시내 복잡한 상호작용 분석
- 균등분할, 부분요인배치 등 활용하여 신도시 완벽적용

기대효과

공실률 해결방안 도출

기존 사회이론을
수학적으로 재확인
공실률 원인 파악 및 해결방안 도출

타 도시 적용

분석결과 타도시에 적용
지속가능한 지산 계획 수립,



02. 성남시 현황분석

1. 성남시 지식산업센터 공실률 현황

◦ 데이터 처리

- 지식산업센터 데이터를 좌표(Point)로 변환
- 시계열 데이터 정리 및 결측치 보정

◦ 격자 생성 및 지식산업센터 위치 매핑

- 100m 격자를 기반으로 400m 격자 생성
- 지식산업센터와 100m, 400m 공간 조인

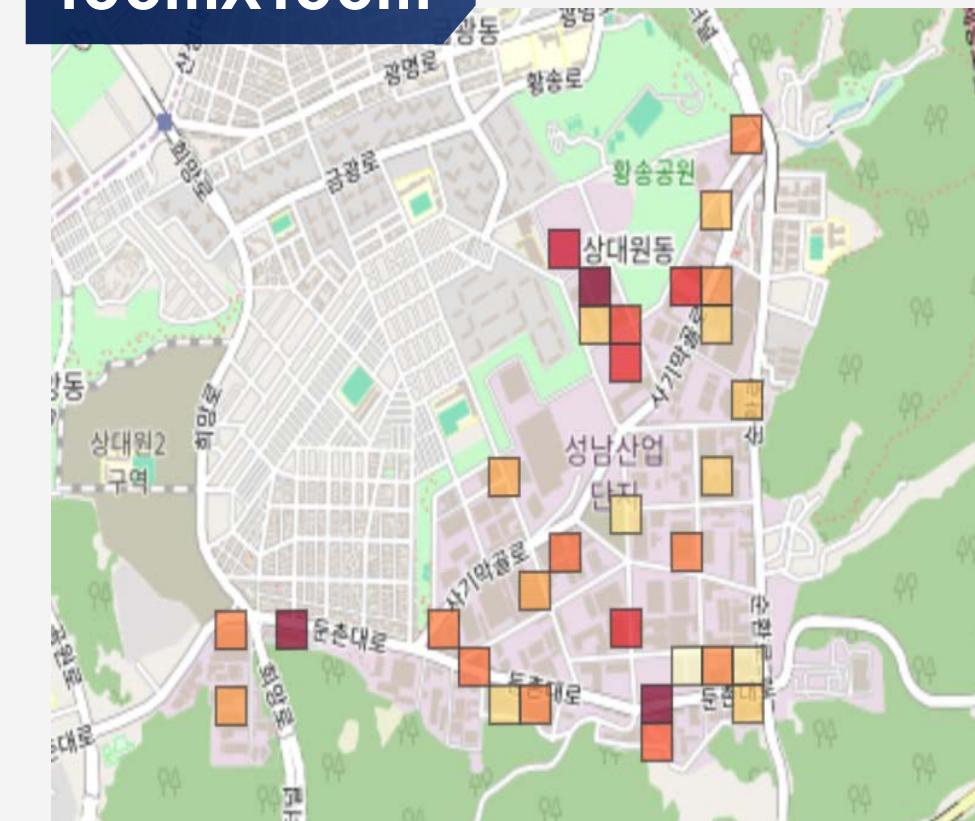
◦ 공실률 계산 및 시각화

$$\text{Vacancy Rate} = 1 - (\text{cpn_in} / \text{tot_cpn})$$

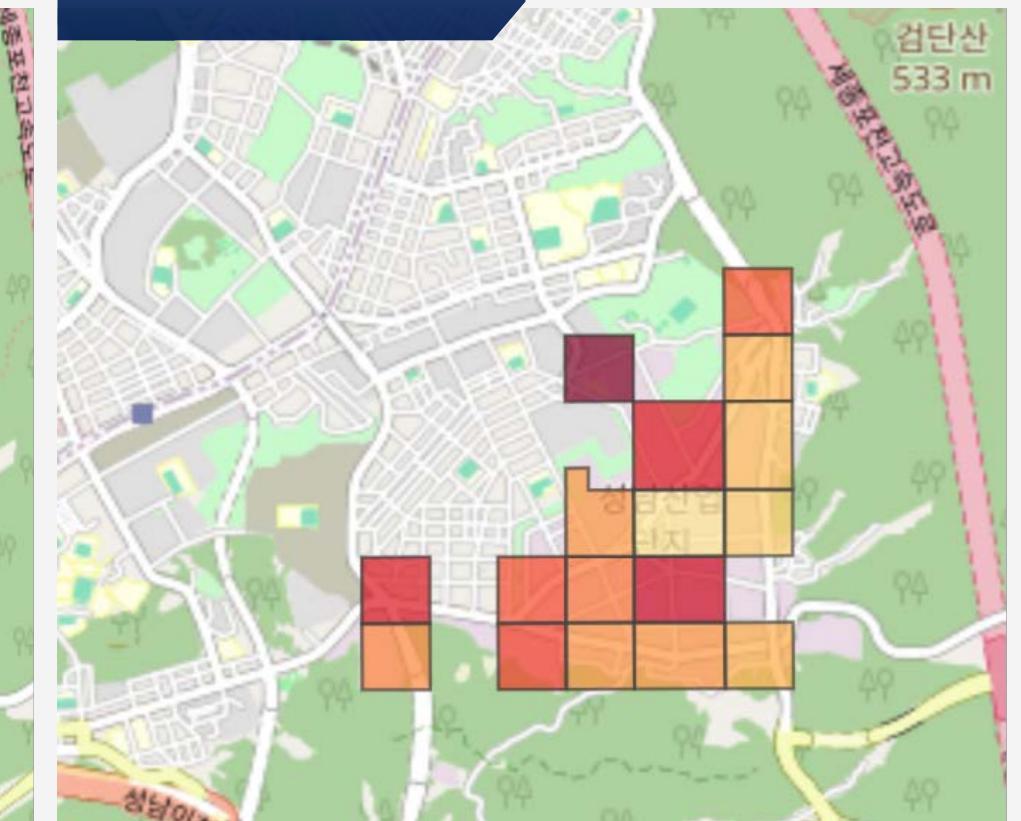
- 격자별 공실률 계산
- 공실률이 낮을수록 노란색, 높을수록 빨간색으로 표시

100m, 400m 격자의 2023년 공실률 지도 시각화

100mX100m



400mX400m



- 지리상으로 공실률 정도를 확인

2. 건물 연면적 - 지식산업센터(제공) 전처리

1단계

2단계

3단계

업종 분류 및 격자 데이터 매칭

표준분류산업코드를 기반으로

상권 정보의 업종을

지식기반 제조업 / 지식기반 서비스업
으로 분류

각 상권의 위치 정보를 기반으로
공간 데이터 생성 후 성남시 400m
격자 데이터와 공간 조인하여 매칭

격자별 제조업 비율 계산

각 격자 내 제조업 포인트 수와
전체 상권 포인트 수를 집계하여
제조업 비율 산출 후,
계산된 제조업 비율을
격자 데이터에 반영

주요 업종 칼럼 생성

주요 업종 결정 기준

제조업 비율 50% 이상 : 제조업
제조업 비율 50% 미만 : 서비스업

각 지식산업센터 포인트가 속한 격자의
주요 업종 정보 추가



2. 건물 연면적 - 분석 결과

각 주요 업종(제조업, 서비스업)을 그룹화하여 회귀분석 진행

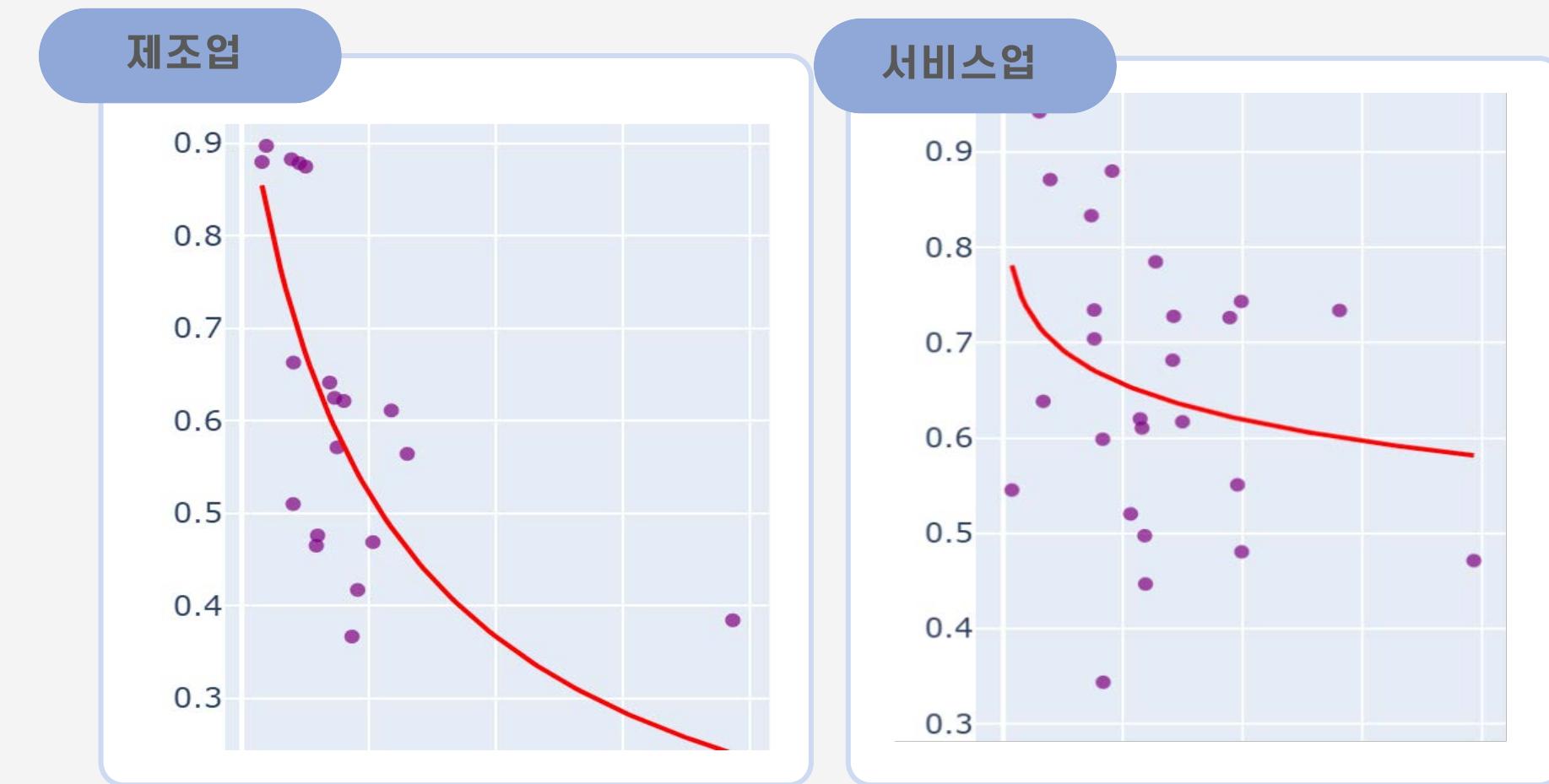
	R^2	p-value
제조업	0.354	0.0056
서비스업	0.048	0.281

제조업

- 건물 면적이 커질수록 공실률이 유의하게 감소
- 실제 공실률 낮아지는 경향 확인

서비스업

- 면적 증대와 공실률 간 유의한 관계를 찾지 못함
- 공실률에 영향을 주는 다른 요인이 더 중요한 것으로 추정



- 업종별로 공실률에 영향을 미치는 요인이 상이함을 보여주며, 투자나 정책 결정 시 이와 같은 차이를 고려할 필요가 있음을 시사
- 다만, 산포도의 분포를 보아 단순 통계적 요인이 아닌 제조업이 선호하는 입지, 혹은 다른 요인이 작용할 가능성도 높아보임

3. 거주인구, 유동인구(제공) - 전처리

1단계

데이터 격자 매핑

유동인구

위·경도 정보를 이용하여
성남시 400m 격자 데이터와 매핑

거주인구

성남시 100m 격자 데이터와 병합하여
폴리곤 중심점을 이용해 위경도 추출한 후
위·경도 정보를 이용하여 성남시 400m
격자 데이터와 매핑

2단계

인구 데이터 집계 및 데이터 결합

유동인구

요일별, 시간대별, 성·연령별 유동인구
데이터를 400m 격자 단위로 집계

거주인구

5km 격자 데이터 생성 후 5km 격자 단위로 모
든 거주인구 데이터 합산

데이터를 400m 격자 기준으로 통일한 뒤,
공실률·유동인구·거주인구·업종 분류 정보를
한 테이블에 결합

3단계

주요활동인구 컬럼 생성 - 유동인구

시간대별 유동인구

시간 칼럼을 7~10시 '출퇴근 시간' /
17~20시 '일반 시간'으로 라벨링 후
격자별 유동인구 평균·합계 산출

요일별 유동인구

요일 칼럼을 월~금 '주중' / 토~일 '주말'
로 라벨링 후 격자별 유동인구 평균·합계 산출

3. 거주인구, 유동인구(제공) - 분석 결과(1)

인구와 공실률 회귀분석

- 거주인구 총합과 유동인구 총합을 로그 값 변환하여 독립변수로 회귀분석 실행

(1) 모델 요약 지표

지표	값/결과
R^2	0.089
p-value	0.134

- R^2이 0.089로 낮고, p-value=0.134로 유의미한 모델이라고 보기 어려움
- 전체적으로 공실률을 설명하는 데에 거주인구와 유동인구가 큰 설명력을 보이지 않음

(2) 추정된 회귀계수

변수	p-value
intercept	0.000
log_total_res (거주인구 로그)	0.438
log_total_flow (유동인구 로그)	0.072

- Intercept가 유일하게 통계적으로 유의 ($p<0.01$)하나, 절편만으로는 공실률에 대한 의미 있는 해석이 어려움
- log_total_flow(유동인구)는 $p=0.072$ 로 “10% 수준”에서 약간의 유의성이 보이지만, 5% 이내로는 유의하지 않음.
- log_total_res(거주인구)는 $p=0.438$ 로 통계적으로 유의하지 않음

(3) Cook's Distance, 표준화 잔차 등 (요약)

Obs(예시)	dfb_Intercept	dfb_log_total_res	dfb_log_total_flow	cooks_d	standard_resid
377	0.488845	-0.323813	-0.338453	0.097787	0.097787
500	-0.292958	0.190737	0.205402	0.036412	-1.035376

- 최대 Cook's Distance가 약 0.10 미만 수준으로, 특정 관측치가 극단적
- 일부 관측치(예: 377, 500)에서 표준화 잔차가 ± 12 정도 분포하나, 일반적으로 ± 23 이상이면 강한 이상치로 의심됨을 고려하면 크게 문제될 수준은 아님.

3. 거주인구, 유동인구(제공) - 분석 결과(2)

주활동인구지표와 공실률 회귀분석

- 주요업종별 다양한 주활동인구 지표(로그 변환)를 독립변수로 하여 공실률(로짓 변환) 간의 단순 회귀분석 실행

주요 거주인구: 공실률에 가장 영향을 줄 가능성이 높은 20~40대 거주인구만 합산
주요 유동인구: 공실률에 가장 영향을 줄 가능성이 높은 20~40대 유동인구만 합산

주활동인구 지표

주활동인구_기본: 주요 거주인구+주요 유동인구

주활동인구_출퇴근: 주요 거주인구+출퇴근 시간대 주요 유동인구

주활동인구_일반: 주요 거주인구+일반 시간대 주요 유동인구

주활동인구_주중: 주요 거주인구+주중 주요 유동인구

주활동인구_주말: 주요 거주인구+주말 주요 유동인구

frac_주중: 주중 유동인구/전체 유동인구

frac_주말: 주말 유동인구/전체 유동인구

주활동인구_주중_출퇴근: 주활동인구_출퇴근*frac_주중

주활동인구_주중_일반: 주활동인구_일반*frac_주중

주활동인구_주말_출퇴근: 주활동인구_출퇴근*frac_주말

주활동인구_주말_일반: 주활동인구_일반*frac_주말

유의미한 결과 요약 지표

주요업종	pop_measure	corr	coef_pop	p_pop	r_squared
지식기반서비스	기본	0.3073	0.285179	0.014994	0.222465
지식기반서비스	출퇴근	0.3066	0.298448	0.013042	0.230561
지식기반서비스	일반	0.2867	0.262878	0.022153	0.199543
지식기반서비스	주중	0.3126	0.237488	0.026869	0.188080
지식기반서비스	주중_출퇴근	0.3164	0.317920	0.008461	0.255330
지식기반서비스	주중_일반	0.2966	0.283471	0.014667	0.223746

- 위 결과, 지식기반서비스업에서 '인구 지표'와 공실률 간에 약 0.19~0.26 정도의 설명력(R^2)이 나오고, 인구 지표 계수(coef_pop)도 양(+)으로 유의하게 나타남
- 즉, 이들 특정 시간대/유형(기본, 출퇴근, 일반, 주중, 주중_출퇴근, 주중_일반)으로 측정한 주요 거주인구가 많을수록, 해당 지식기반서비스의 공실률이 (단순회귀 기준) 증가

4. 층별개요(제공) - 전처리 및 분석결과

업종 그룹화(제조업·서비스업) 후 공실률과 복지 주요시설 개수 간 관계 회귀분석

◦ 위치 정보 추가

- 연속지적과 병합하여 폴리곤 중심점을 이용해 위치 정보 (lon,lat) 추출

분석 결과

대부분 통계적으로 유의미한 결과($p < 0.05$)가 나오지 않음

◦ 격자 공간 조인

- 위도,경도가 추가된 데이터의 lon,lat컬럼을 이용해 Point 생성
- 각 Point와 400m 격자 데이터를 공간 조인

◦ 복지/주요시설 개수

- '음식','체력','보육','어린이','육아','라운지','휴게'라는 단어가 들어가면 복지 주요시설로 간주하고 카운트

5. 대중교통 접근성 분석 - 전처리

1. 위치정보 추가

- 1) 시작점(start point) : 지식산업센터
- 2) 목적지(end point) : 강남역
- 3) 대중교통 정보 : 버스정류장, 지하철역



이동시간 계산을 위해
각각에 Point 객체 부여

2. 최소이동시간

지하철만 이용

- 1) 걷기 : 거리/보행속도*60
- 2) 지하철 : 대기시간+(거리/지하철속도*60)

버스 >> 지하철 환승

- 1) 걷기 : 거리/보행속도*60
- 2) 버스 : 대기시간+(거리/버스 속도*60)
- 3) 지하철 : 대기시간+(거리/지하철속도*60)

버스만 이용

- 1) 걷기 : 거리/보행속도*60
- 2) 버스 : 대기시간+(거리/버스 속도*60)



각 시나리오에서 계산된 이동시간 중 최소값 선택

최소이동시간

3. 교통수용률 계산

$$\text{버스수용률 } y = 1.8429 e^{-0.0033x}$$

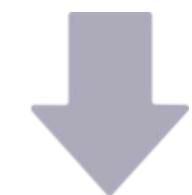
$$\text{지하철수용률 } y = 1.8424 e^{-0.0047x}$$

지하철과 버스의 서비스권역 비교 및 이용자들의 도보거리 추정,
김경환 외 4인, 2010

출발점에서 가장 가까운
3개의 지하철역과 버스정류장 선택



선택된 3개의 지하철역 및 버스정류장
각각에 대해 수용률 수식을 적용하고 그 값을 합산



총 교통수용율 = 지하철 수용율 + 버스 수용율

총 교통수용율

5. 대중교통 접근성 분석 - 분석 결과

지산 업종별 공실률과 교통접근성 관계 분석

$$\text{logit}(y) = \beta + \beta_1 \log(\text{최소이동시간}) + \beta_2 \log(\text{교통수용율}) + \epsilon$$

구분	제조업 (N=20)	지식산업 (N=26)
설명력(R2)	0.289	0.157
모형 유의성(F-test p)	0.0553	0.139
상관계수 : 최소이동시간(log) vs 공실률	0.3572	-
상관계수 : 교통수용율(log) vs 공실률	0.0927	-
회귀계수 : 최소이동시간(log)	-11.5226 (p=0.031)	3.7779 (p=0.056)
회귀계수 : 교통수용율(log)	0.8324 (p=0.459)	0.6915 (p=0.376)

회귀분석 진행

종속변수 : 공실률 로짓변환값

독립변수 : 최소이동시간과 교통수용율의 로그 변환 값

1) 설명력, 모형 유의성

- 제조업과 지식산업 각각 경계에 있거나, 유의하지 않음

2) 회귀계수

- 교통수용율은 제조업과 지식산업 둘 다 유의미하지 않다고 판단
- 제조업의 경우, 최소이동시간 유의함 + 음의 관계를 가짐
- 지식산업의 경우, 최소이동시간 유의함 + 양의 관계를 가짐

6. 실거래가(외부) - 전처리

1단계

2단계

3단계

4단계

데이터 결합

- 아래 기재된 외부 데이터를 결합 후, 계약 연월 기준으로 정렬

외부 데이터

- <2개년치 데이터 사용 : 총 12개>
- 2023~2024 상업업무 성남시.csv
 - 2023~2024 공장창고 성남시.csv

▲경기부동산포털

데이터 매칭

- 도로명 주소(도로명, rd_addr)을 기준으로 데이터 매칭
- 매칭 후, 각 업무공장 데이터에 지식산업센터 이름 (kig_ids_ct_nm), 위치정보 (lon,lat) 결과 추가 + 매칭되지 않은 행 제거

데이터 격자 매핑

- 매칭된 업무공장 데이터에 위,경도 컬럼 부여하여 Point 객체 생성
- 이후 성남시 400m 격자 데이터와 공간조인

데이터 정제 및 변환

- 거래금액(만원) 컬럼값 float 형으로 변환
- 격자 값 중 NULL값인 행 제거
- 2023 공실률 지표를 상하위 30% 그룹으로 분류

6. 실거래가(외부) - 분석 결과(1)

업종 그룹화(제조업, 서비스업) 후 공실률과 거래금액(실거래가)간 회귀분석

제조업과 서비스업 모두
통계적으로 유의미한 관계가 나타나지 않음

- 공실률은 임대료, 건물 상태, 교통 접근성 등 여러 요인의 영향을 받기 때문에, 단순히 거래금액만으로 설명하기 어려움
- 실거래가(거래금액)"가 공실률에 영향을 미치지 않는다고 단정할 수는 없으나, 분석 자체에서는 단독 변수로서는 유의미한 설명력을 보이지 않음

업종 그룹화(제조업, 서비스업) 후 교통요인과 거래금액(실거래가) 간 회귀분석

여러 유형과 업종에서
교통수용율과 이동시간은 거래금액과
유의미한 관계가 보이지 않음

- 대부분의 모델에서 **p-value > 0.05 (유의X)**
- 모든 모델에서 **R^2값 0.01~0.07 수준** (매우 낮은 설명력)
- 교통 변수 외에 입지, 상권 특성, 기업 규모, 주변 인프라 등 다른 변수 추가 필요

6. 실거래가(외부) - 분석결과(2)

거래금액·교통수용률 상호작용이 공실률에 미치는 영향

- 주요업종과 상/하위30% 그룹별 회귀분석 실행

업종	그룹	R^2	p-value	유의미한 변수
제조업	하위 30%	0.451	0.908	없음
제조업	상위 30%	0.349	0.794	없음
지식기반 서비스업	하위 30%	0.976	0.059	X1, X2, X2_2, X1_2, X1_X2
지식기반 서비스업	상위 30%	0.756	0.503	없음

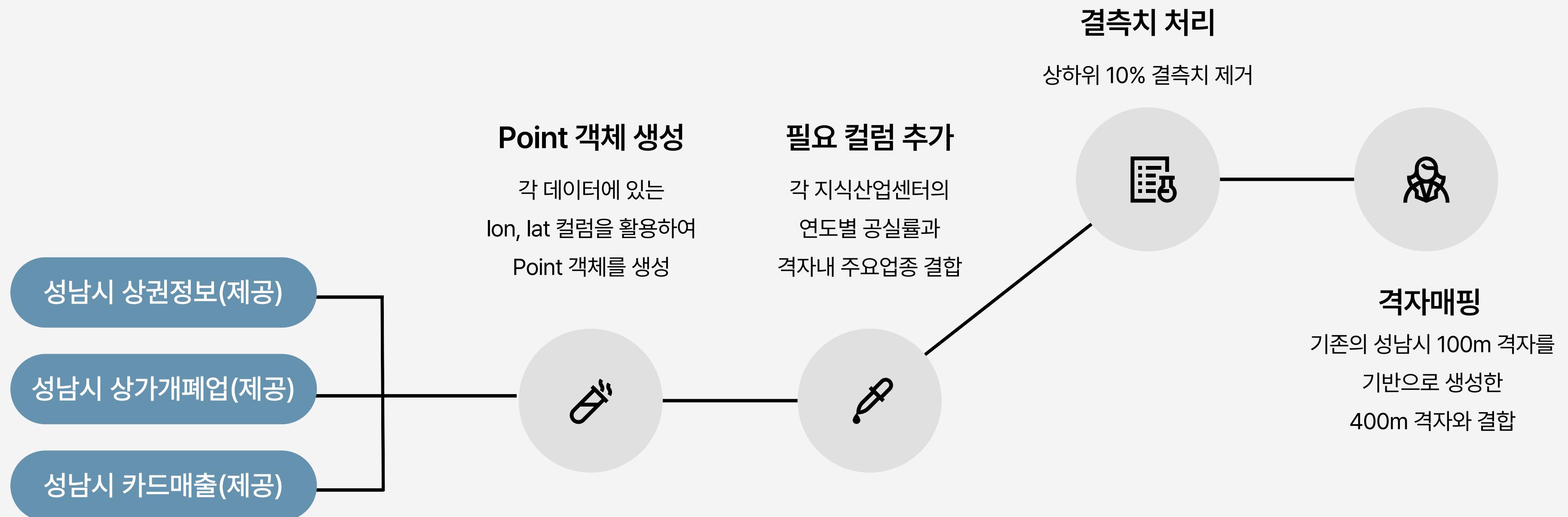
1) 제조업

- 교통 관련 변수가 공실률과 유의미한 관계가 없음
- 단순한 교통 변수만으로는 공실률을 설명하기 어려움
- 입지 특성, 임대료 수준, 상권 특성 등 추가 변수 필요

2) 지식기반서비스

- 하위 30% 의 경우 유의미하다는 결과 확인
 - R-squared = 0.976
 - P(F-statistic) = 0.0592
- 거래금액, 교통수용율, 교통수용율의 이차항, 교통수용율의 비선형 효과가 공실률 예측에 중요한 요소
- 비선형성과 상호작용 효과가 중요함을 도출

7. 상업 관련 분석 - 공통 전처리



7-1. 상업 관련 분석 - 상권정보(제공)

격자/업종별 상가개수와 공실률 간 회귀분석 진행

제조업

- 일부 업종(자동차 수리 및 세차, 세탁 등)이 미세하게 공실률을 증가시키는 효과 확인
- 유의한 음의 효과는 '운동/경기용품소매'에서만 확인
- 제조업의 경우 단순히 주변 상권 분포만으로 공실률을 설명하기는 어려움을 시사

지식기반 서비스업

- 업종에 따라 공실률이 크게 달라짐
- 공실률을 유의하게 증가시키는 업종 : 부페, 대행업, 개인 및 가정용품 수리 등
- 공실률을 유의하게 증가시키는 업종 : 별식/퓨전요리, 자동차/이륜차, 양식 등

7-2. 상업 관련 분석 - 상가개폐업(제공)

✓ 공실률&생존기간 시계열 분석

성남시 400M 격자별
주요 업종별로 데이터 분리 후
업종별로 공실률과 생존기간 관계 분석

앞선 공통 전처리 단계에서 생존기간 생성 과정 추가

생존기간 = 폐업일(cls_date) - 개업일(lcpmt_dt)
현재 시점(current_date)과 이전 시점(prev_date)을 고려하여
조정된 생존기간을 반환



회귀분석 실행
종속변수 : 조정생존율 로짓 변환 값
독립변수 : 각 시점별 공실률 로그 변환 값

1

제조업

- 공실률이 생존율에 약한 영향을 미침
- 영향은 업종별로 차이가 존재
- 특히 '일반게임제공업'과 '종합여행업'은 공실률 증가시 생존 가능성이 낮아짐

2

서비스업

- 공실률과 생존율 간의 연관성이 거의 없음
- 일부 업종에서만 미미한 영향이 존재

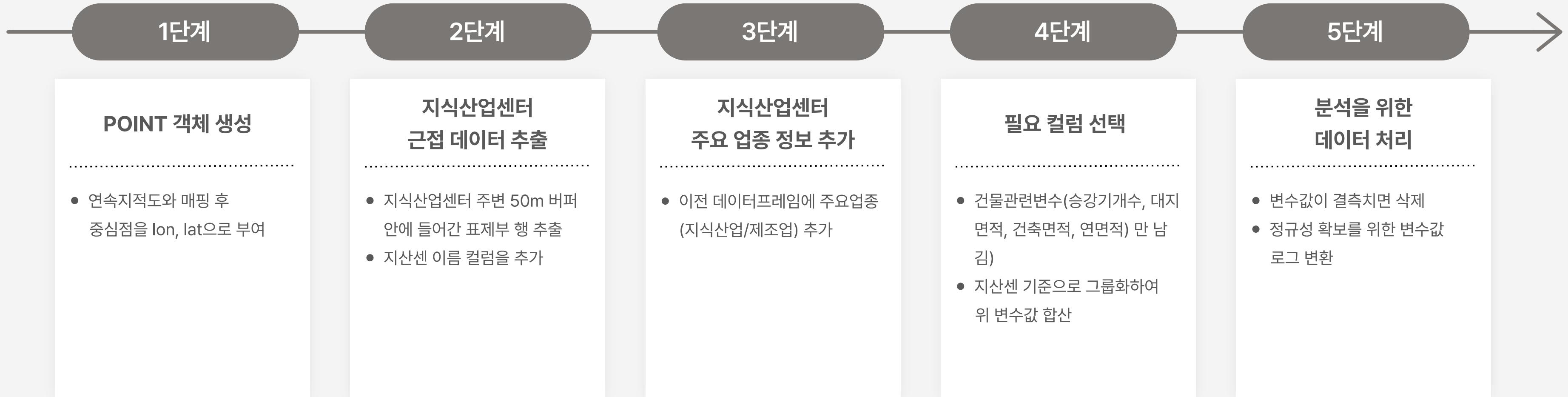
7-3. 상업 관련 분석 - 카드매출(제공)

월별 카드매출과 공실률 시각화 - 추이 확인

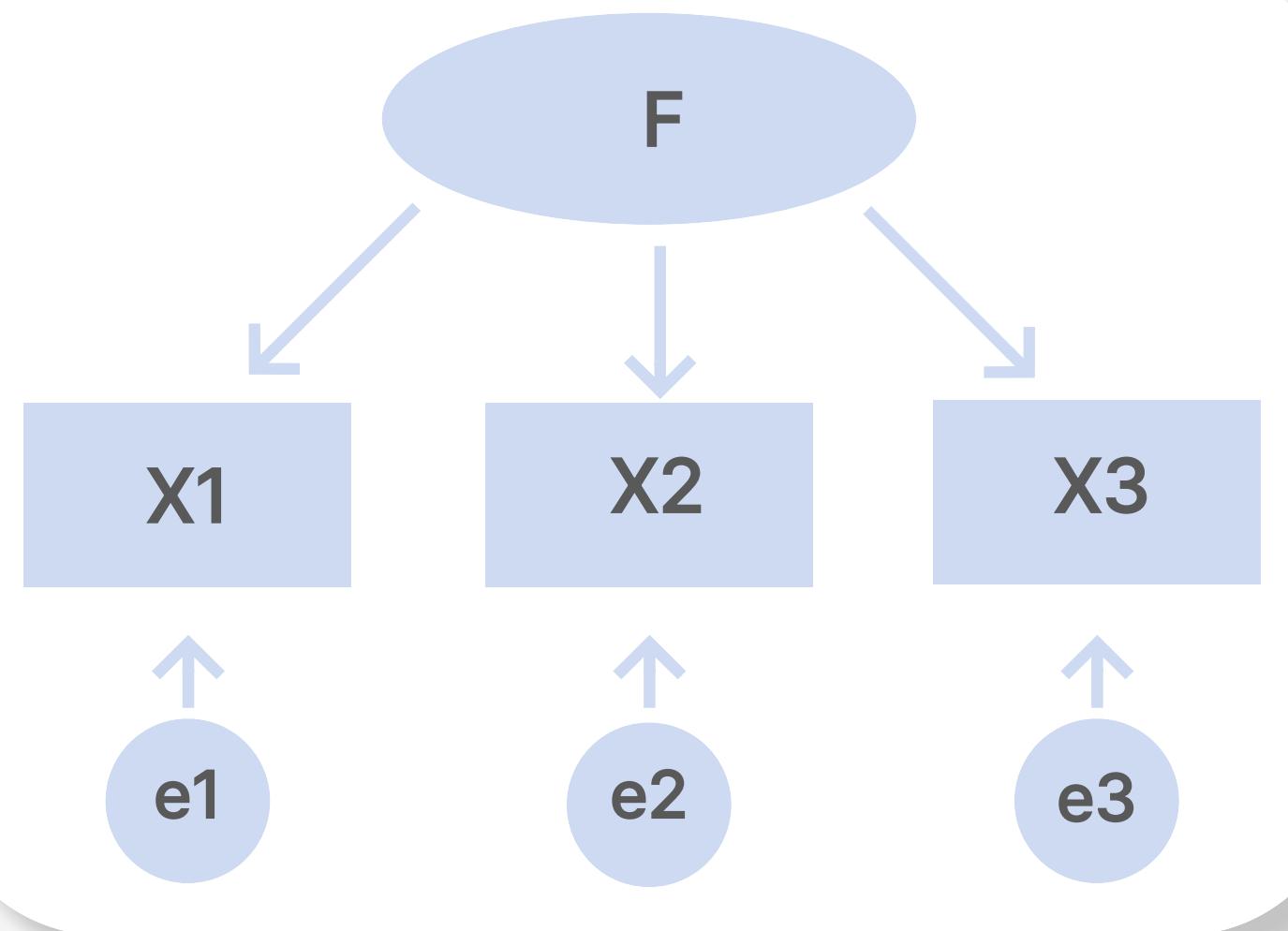


카드매출과 공실률간의 관계는 현실적으로 의미를 부여하기 어렵고,
단독변수만으로 해석이 어려움을 확인

8. 표제부(제공) EFA 분석 - 전처리



8. 표제부(제공) EFA 분석 - 분석 진행



EFA(탐색적 요인분석)이란?

- 여러 개의 변수들 속에서 공통된 특성을 가진 숨겨진 요인을 찾아내는 통계 기법
 - 데이터 차원을 축소하여 변수 간의 관계를 이해하는데 사용
-
- 서로 관련 있는 변수들을 하나의 요인으로 묶어 건물 특성(예: 면적, 승강기 개수 등)과 같은 데이터를 하나의 공통된 요인으로 변환

8. 표제부(제공) EFA 분석 - 분석 진행



PCA VS EFA

$$\Sigma_X \approx Q \Lambda_D Q^T$$

PCA) 단순히 전체 잔차를 가지고 요인 추출

$$\Sigma_X \approx \Lambda \Lambda^T + \Psi$$

EFA) 변수별 고유분산까지 고려하여 요인 추출

PCA와 같은
순수 차원축소 방법보다

EFA와 같은
측정모델 성격을 가진것이 적합!

$$\begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & \cdots & \cdots & 1 \end{pmatrix} \approx \Lambda \Lambda^T + \Psi.$$

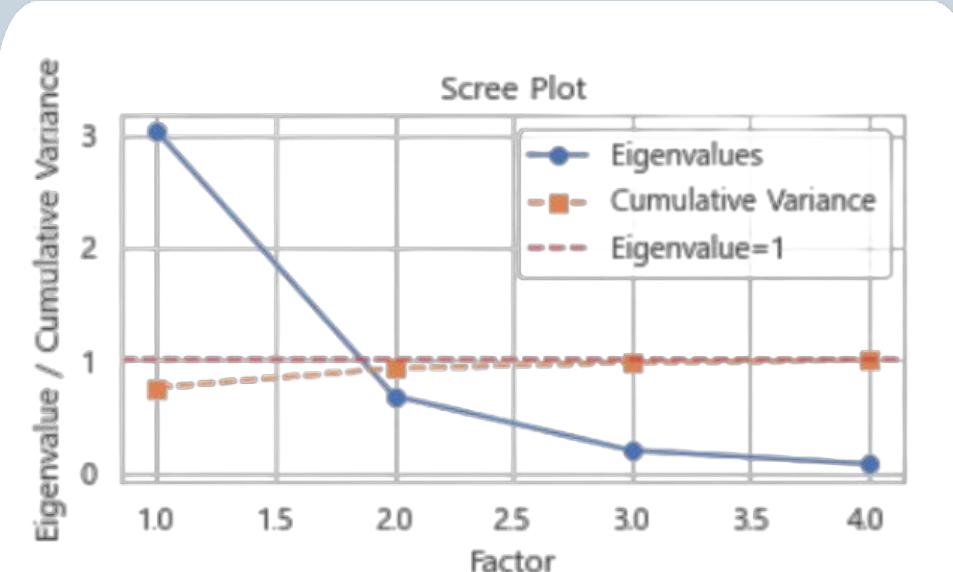
상관행렬

각 변수간의 상관관계 네트워크를
요약과, 공통과 개별특수성 설명

8. 표제부(제공) EFA 분석 - 분석 결과

Scree Plot

요인 분석에서 고유 값의 크기를 시각화하여 적절한 요인 개수를 결정하는 그래프



첫 번째 요인이
데이터의 대부분을 설명

이후 급격히 감소하는 패턴

요인 분석 시 요인 개수를
1~2개 정도로 설정하는 것이 적절

- 요인 개수를 1개로 설정하고 EFA 수행
- 요인 로딩(Factor Loadings)을 계산하여 각 변수와 요인 간의 관계 확인
- 로딩 값이 절대값 0.4 이상이면 해당 요인과 강한 연관

공통성(Communality) 분석

- 각 변수가 요인으로 얼마나 잘 설명되는가를 나타내는 값
- 공통성 < 0.2 일 경우, 요인으로 설명되기 어려운 변수로 판단

신뢰도(Cronbach's α) 분석

- 값이 0.8 이상일 경우 매우 신뢰할 수 있는 변수라고 판단

변수명	요인 로딩 값	해석
ride_use_elvt_cnt	-0.642	건물의 승강기 개수가 요인과 부적 상관
plat_area	-0.908	건물의 대지 면적이 요인과 강한 부적 상관
arch_area	-0.962	건물의 건축 면적이 요인과 매우 강한 부적 상관
tot_area	-0.931	건물의 연면적이 요인과 강한 부적 상관

전체 Cronbach's $\alpha = 0.871$

=> 이 요인(Factor1)에 포함된 변수들은 일관성이 높아 신뢰할 수 있음

건물 면적의 요인 로딩은 모두 강한 음수(-)값

'건물 규모가 작은 경우 높은 값을 갖고, 큰 경우 낮은 값을 갖는 요인'

03. 성남시 주요 요인 도출

- 2단계 타원형 EFA
- TSRI

2단계 타원형 EFA

1. 차원축소 데이터 선정

- ① 카드 매출 데이터
- ② 성연령별·시간대별·요일별 유동인구 데이터
- ③ 상권 정보 데이터



KMO 최적 알고리즘

- 각 EFA 수행시, 앞선 데이터에서 KMO 값이 0.7 이상인 요인 식별
- 내부 변수의 개수가 많아 설명력이 높은 요인을 우선적으로 탐색하는 알고리즘을 적용하여 최적 요인을 도출할 수 있음

2. 데이터 전처리

- 독립변수 > 로그변환

카드 매출 데이터

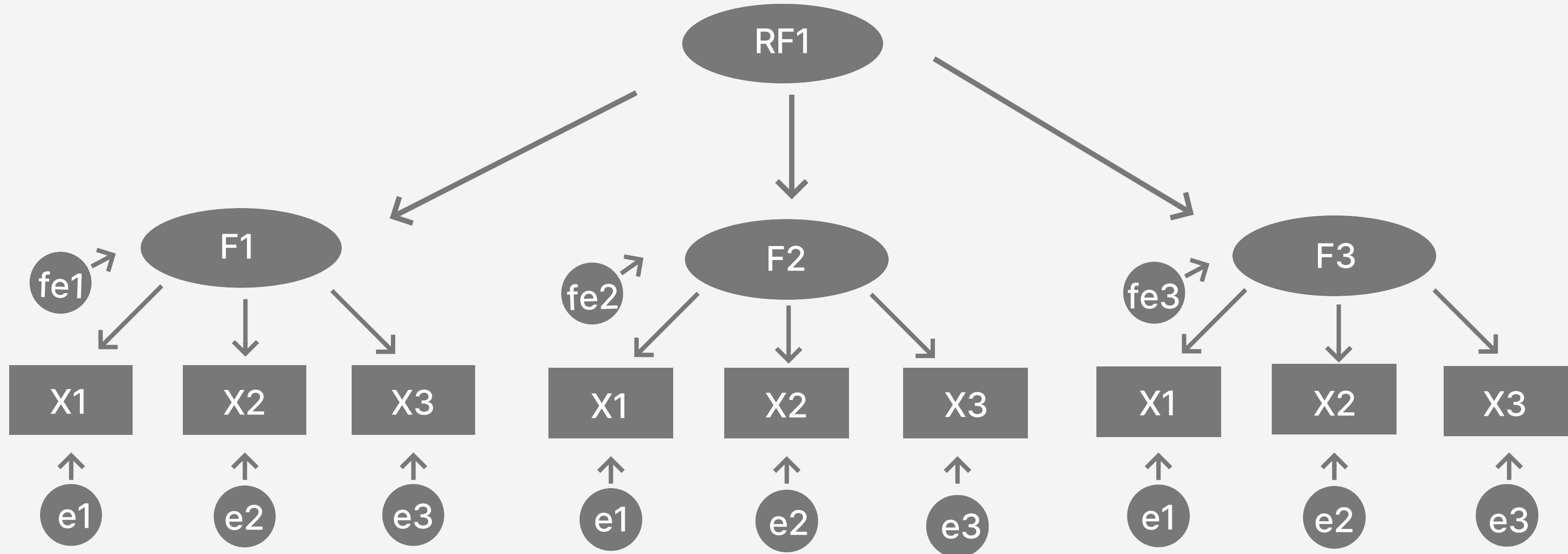
$$Y = \log(X) = (\log X_1, \dots, \log X_p)$$
$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left(-\frac{1}{2}(y - \mu)^T \Sigma^{-1} (y - \mu)\right).$$

데이터 특징

- 양의 값만 존재
- 한쪽 꼬리가 매우 긴(스큐) 형태로 분포
- 정규성 가정을 직접 적용하기 어려움
- 분산이 커지거나 이상치(outlier)가 발생하기 쉬움
- 로그변환 사용시, 근사 정규성을 얻게되며 동시에 유사 타원형분포로 취급이 가능하고 EFA의 타당성 확보

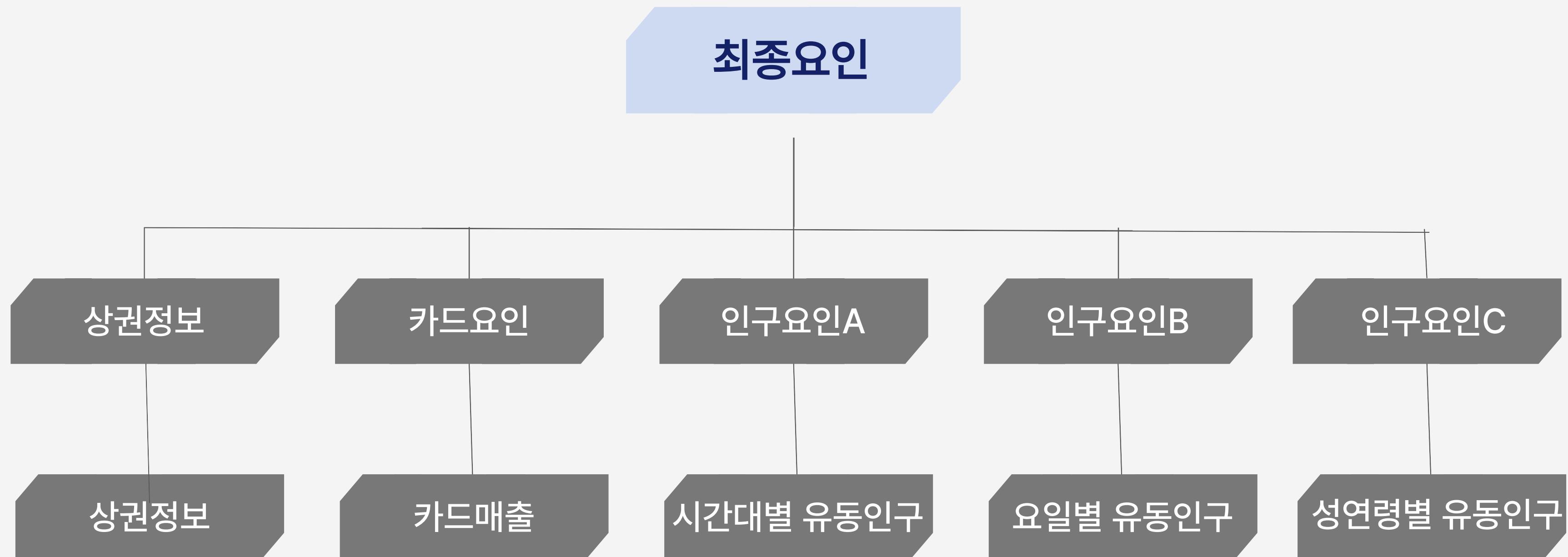
2단계 타원형 EFA

- 2단계 탐색적 요인분석(EFA) 모형 도식화



2단계 타원형 EFA

- 2단계 타원형 EFA 데이터 구조



2단계 타원형 EFA - 요인추출 기법 선택

구분	PCA	PAF (주축요인법) 	ML (최대우도법)
목적(초점)	전체 분산 최대 설명 (차원 축소)	공통 분산만 추출 (EFA 본 취지)	잠재요인 모형에 대한 최대우도추정 (정규성 가정)
모형적 해석	고유오차 ϵ_i 명시적 분리 X	각 변수 고유오차(특수성) 분리 O	각 변수 고유오차(특수성) 분리 O + 통계적 적합도 검증
수학적 구조	$R = QDQ^T$, 상위 고유값 k	$R_{modified} \approx \Lambda\Lambda^T$ (반복 대각 치환)	$\max_{(\Lambda, \Psi)} L(\Lambda, \Psi) \text{ s.t. } \Sigma = \Lambda\Lambda^T + \Psi$
기하학적 의미	전체 분산기준 측정 정확, 잔차를 통째로 처리	공통상관'만으로 부분공간 형성, 각 변수의 고유분산은 대각(Ψ)로 분리	정규성하에서 Σ 를 $\Lambda\Lambda^T + \Psi$ 형태로 분해, 가능도 극대화
장점	계산빠름	EFA본질(공통성 VS 특수성) 강조, 정규성 가정 덜 엄격	통계적 엄밀성, 공통성/특수성 분리
분포/가정	가정 없음(단순 선형대수적)	정규성 엄격하지 않음	다면량 정규성 가정 (자료가 이를 만족해야 이상적)
EFA 적합도	낮음	높음 (공통성 중심)	높음 (단, 분포가정 충족 시)

2단계 타원형 EFA

회전방법 : Oblimin

직각회전(Orthogonal Rotation)

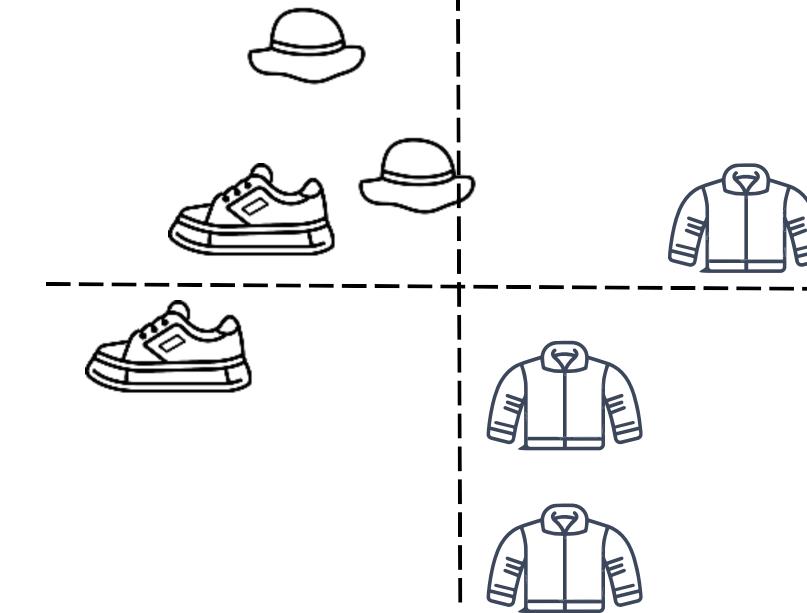
- 요인 간 상관관계가 0이라고 가정
- 각 요인을 독립적인 요소로 해석할 수 있도록 변환하는 방식



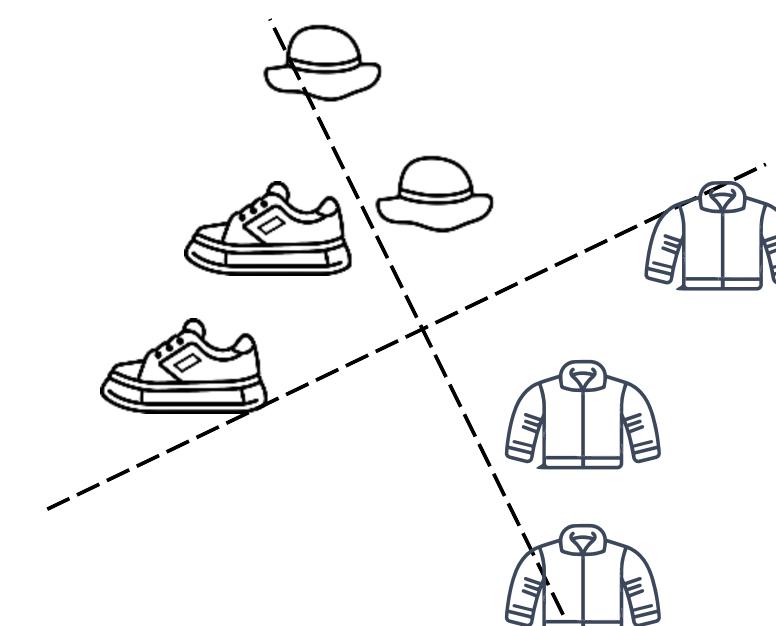
하지만 현실에서는 카드매출, 상권정보, 유동인구 간에
상관관계가 전혀 없다고 가정하기 어려움

현실을 보다 자연스럽게 모델링하기 위해서는
요인 간 상관을 허용하는 사각회전이 더욱 적합!

직각회전



사각회전



2단계 타원형 EFA - 1단계 EFA

3. 상권 이름(개수)와 공실률 회귀분석

$$S = \{ x_j \mid p\text{-value}(\hat{\beta}_j) < \alpha \}$$

- 회귀분석을 통해 공실률에 유의미하게 영향을 주는 상권 변수 확인
- 해당 변수들을 모아 **KMO 최적 알고리즘** 과 **EFA**를 적용하여, **상권 잠재요인을 도출**

4. 카드매출 데이터 EFA

$$U = \{ F_i \mid |\text{corr}(F_i, x'_j)| \geq \rho_0 \text{ for some } j \} \cup \{ x'_j \mid |\text{corr}(F_i, x'_j)| \geq \rho_0 \text{ for some } i \}$$

- 앞서 도출된 상권요인과 카드매출 지표 간 상관관계를 확인한 뒤, 상관이 확인된 변수들을 합쳐 **재차 EFA 수행**
- 상권 활성도와 함께 작동하는 **추가 잠재요인** **도출**

2단계 타원형 EFA - 유동인구 EFA 및 2단계 EFA

5. 성연령별·시간대별·요일별 유동인구 EFA

각 지표들의 잠재구조를 파악하고
이를 통해 “유동인구 요인” 도출

6. 2단계 EFA

- 앞서 도출된 상권정보, 카드매출, 3개 유동인구 요인을 확인한 뒤, **KMO 최적 알고리즘** 수행 후 **재차 EFA** 수행
- 이를 통해 해당 지역을 설명할 **최종 잠재요인** 도출

평가 방법

1단계	KMO, 바틀렛, 크로노스 알파
2단계	KMO, 바틀렛, 합성신뢰도

- 1단계의 경우, 일반적으로 KMO, Bartlett 검정, Cronbach's α 등을 사용
- 2단계의 경우, 최종 요인들이 “진짜 제대로 측정” 되는지를 가중치(요인 적재치)와 오차까지 고려해 더욱 정밀한 검증 필요

합성신뢰도(CR)

- 문항마다 다른 적재치를 반영하여 실제 측정구조를 정확히 평가
- 요인 간 내적 합치도, 타당성을 높이는데 도움이 됨
- 최근 EFA 및 SEM 등 연구 전반에서 최종 단계에서의 요인 타당성 검증을 위해 합성신뢰도 사용 권장

2단계 타원형 EFA - 결과

데이터	추출요인개수 (생존요인수)	Bartlett's Test	생존요인 크로노스 알파	합성신뢰도 (CR)
상권	3(2)	통과	0.85이상	0.9이상
카드매출	3(2)	통과	0.9이상	0.95이상
성연령별 유동인구	2(2)	통과	0.95이상	0.94이상
시간대별 유동인구	3(3)	통과	0.95이상	0.95이상
요일별 유동인구	2(1)	통과	0.95이상	0.95이
최종 (2단계) EFA	3(3)	통과	CR로 측정	0.7~0.95

크로노스 알파나 합성신뢰도가 낮은 경우 요인제외
(만족시에만 생존요인으로 취급)

2단계 타원형 EFA - 결과

요인	주요특징	도메인 지식	활용 방안
상권 요인1	취미, 기념품등 기타 업종	관광객이나 여가·취미 소비를 겨냥한 업종이 많음	지자체·관광공사가 협력해 도시 브랜드 가치를 높이는 사업 전략 수립에 활용
상권 요인2	외식, 유통등 생활 서비스형 업종	생활 밀접형 서비스, 기업 근로자 서비스 제공 업종, 근무지·주거지 주변에서 반복 소비가 발생하므로 유동인구의 업무·생활 패턴에 영향 받음	지역 주민의 생활 밀착 소비 행태 분석을 통해 상권 경쟁력 강화 및 기업 선호 입지 구축 활용
카드 요인1	외식·유흥 업종 집중	생활 서비스(레스토랑, 카페, 주점 등)가 중심이 됨 외식 문화나 유흥지구의 활성도 확인	근로자, 도시 소비 트렌드 및 프로모션 전략 수립에 유용
카드 요인2	의료·기타 업종 특성	병원, 약국 등 의료 관련 업종 혹은 기타 소매업종이 둑임	지역 내 의료·상점 인프라 수준, 다양성 관련 전략 수립에 유용
시간대 유동인구 요인1	아침~낮 시간대 활동인구 밀집도가 높음	직장인·학생 중심 이동 패턴일 가능성	오전·낮 시간대 프로모션, 점심식사 특화 운영 전략 등에 활용
최종 요인2	상권요인2, 카드요인1 등 주로 적재 상권(외식·유흥) + 통합 매출 특성	상권 클러스터로 해석 가능	기업 입지 선호도, 대중교통 전략수립등 다양한 방안으로 활용 가능

2단계 타원형 EFA

단순 OLS/구분	설명력	유의변수	해석	상관관계/구분	유의 상관변수	상관계수	해석
전체	20.3%	최종요인2: $p \approx 0.07$ 계수 방향 (음)	상권 발달 지역 일수록 기업·시설 유입이 어느 정도 활발함 시사	전체	최종요인2 vs 공실률	-0.38	전반적으로 상권이 발달된 곳에서는 공실이 줄어드는 음(-) 상관 관계
제조업	47%	교통수용율: $p \approx 0.057$ 계수 방향 (양)	대형 교통 인프라가 잘 확보된 지역이라도 제조업 입주가 활성화되지 않을 수 있음을 시사	제조업	유의한 상관관계X		
지식기반서비스	56%	- final_efa_Factor2 : $p=0.005$ (음) - price_per_area_log : $p \approx 0.052$ (양)	지식기반기업은 주변 인프라(외식·유통 등)를 중요시 하지만 지나치게 높은 가격은 입주 저해로 이어질 수 있음	지식기반서비스	최종요인2 VS 공실률	-0.63	지식기반서비스업은 주변 상권 인프라를 선호하는 경향이 뚜렷함

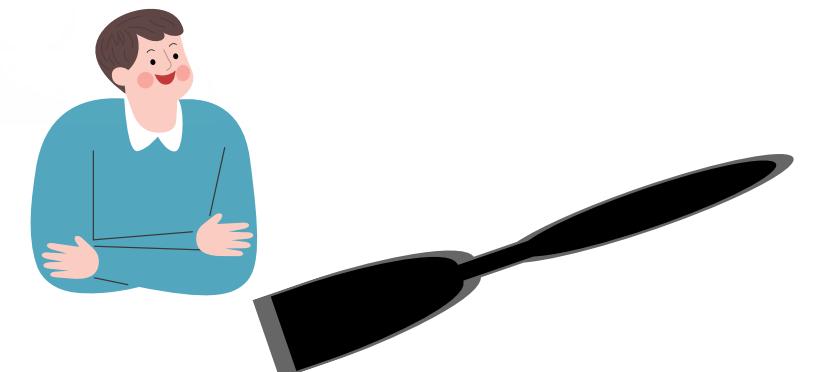
내생성 문제

- 부동산 경제학과 자연공실률 이론 등 다양한 연구 및 학문에서 공실률과 임대료가 상호작용함은 이미 알려진 사실
- 수학적으로는 내생성 관계라고 하며 이를 무시할 경우 계수(β) 추정치의 편향·불일관성, 정책적 해석 오류 등 문제 발생

- 일반 OLS는 $\hat{\beta} = (X^T X)^{-1} X^T y$ 이므로 회귀계수에 대응하는 y 를 X 의 열공간에 직교투영한 벡터
- OLS의 수학적 조건은 $X^T (y - X\hat{\beta}) = 0 \iff X^T e = 0$. 즉, 독립변수와 오차항의 비상관성을 전제

- 공실률과 임대료는 독립변수가 오차항과 상관되어 있는 상황이므로 직교 조건이 깨짐
- X 에 대해 오차벡터가 수직이 아님
- 즉, 투영하게 되지만 그 방향이 틀리게되며 회귀계수가 편향된 값으로 나오게 되며, 정사영 자체가 왜곡, 이론적 근거가 약해짐

이미 부동산 경제학 등으로
공실률과 임대료의 상호작용이 알려진 상황에서
내생성을 고려하지 않는 것은 무리가 있음



수직으로 있어야할 조명이 기울어짐

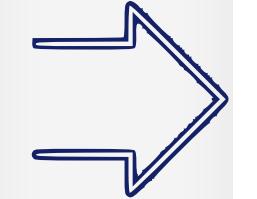
◦ TSRI란?

- 최근 논문들에서 활발히 사용되는 기법으로 Control Function 접근법과 유사한 아이디어이지만, 2단계에서 비선형 추정을 적용하고자 할 때 사용
- 일반 2SLS는 2단계 모형이 선형이라는 전제하에 일관성을 확보하지만 도시데이터에서는 비선형 형태를 보이는 경우 존재
이때 TSRI를 적용하면, 내생 변수를 통제하면서도 비선형 모형을 적용하여 더 현실적이고 탄력적인 추정을 수행 가능

방법	개념/절차	장점	단점
2단계추정	1단계: 내생변수를 도구변수로 회귀해 예측값 도출 2단계: 예측값을 본 모형에 대체하여 추정	비교적 간단하고 이해하기 쉬움 선형 모형에서 가장 널리 사용	비선형 모형에 직접 적용하기 곤란
제어함수 접근	1단계: 내생변수에 대한 회귀 후 잔차 추출 2단계: 본 모형에 잔차를 추가 변수로 포함	내생성 구조(오차항과의 상관) 직접 통제	잔차를 잘못 추정하면 편향 가능성O
TSRI	비선형 모형에서 2단계 추정 시 잔차를 직접 포함	- 의료/도시경제 분야 등에서 내생적 선택을 다룰 때 자주 사용 - 비선형 내생성 처리에 적합	구현난이도 높음
GMM	모멘트 조건(예: EZ·오차] $=0$ 등)을 활용하여 IV를 일반화해 추정	모멘트 조건 활용으로 다양한 상황(패널, 다중 IV) 적용 가능 효율성 개선 가능	모멘트 조건 설정이 어렵고 표본이 작으면 추정 불안정

TSRI 1단계

단순 선형회귀가 아닌 **랜덤포레스트** 수행
→ 복잡한 **비선형 관계** 추정하는데 유리



- (1) 학습데이터를 이용해 랜덤포레스트 모델을 훈련
- (2) 예측치와 잔차를 각각 구함

TSRI 2단계

1 단계에서 얻어낸 잔차를 삽입하여
2단계에서 **가우시안 프로세스 회귀** 사용
→ Y를 더욱 **탄력적으로 모델링**, 데이터 패턴을
매끄럽게 추정하고 불확실성(분산)까지 함의
더 풍부한 해석 제공

TSRI - 결과 (전체 업종 결과 해석)

모형	R ²	유의 변수/계수 부호	해석/시사점
제어함수 접근	R ² = 0.336 p값 유의	• 합산이동시간(-, p<<0.05) • price_per_area_log 유의 아님	이동시간이 길수록 (대중교통 서울 이동시간) 공실률이 오히려 낮아짐 → 업종·지역 특성상 대중교통 이동시간 상관 없이 수요가 충분한 것으로 해석
TSRI	R ² = 0.4830 RMSE= 0.1337 MAE= 0.1119	개별 변수 유의성은 낮으나 SHAP 결과 여러 특성 분산	- 가격보다는 입지·건물요인이 복합적으로 작용 - 예측력은 중간 정도 수준
2단계 추정	R ² = -1.55 등(음수) or 추정치 유의성 미흡	유의한 변수 거의 없음	의미 X
GMM	R ² =0.321	합산이동시간(-, p≈0.01) price_per_area_log(음, p≈0.10)	이동시간이 길수록 공실률↓, 결과 일관 가격 영향은 통계적 유의성 낮음

TSRI - 결과 (제조업 결과 해석)

모형	R^2	유의 변수/계수 부호	해석/시사점
제어함수 접근	$R^2= 0.532$ p값 유의	합산이동시간(-, $p\approx 0.04\sim 0.08$) final_efa_Factor1(+, $p=0.02$)	<ul style="list-style-type: none"> 외곽(서울과 대중교통 이동시간)임에도 공실률 제조·물류 등 외곽 입지를 선호할 가능성 주변 유동인구(Factor1)도 공실에 유의한 영향
TSRI	RMSE= 0.1355 MAE=0.1106 $R^2= 0.5706$	<ul style="list-style-type: none"> 변동 요인 분산(SHAP) 주로 합산이동시간/일부 건물, 유동인구 요인 	<ul style="list-style-type: none"> IV2SLS보다 예측력 안정적 외곽 선호 및 주변 유동인구(Factor1) 반영
2단계 추정	$R^2= 0.456$	final_efa_Factor1($p\approx 0.018$) 합산이동시간(음, $p\approx 0.08$)	<ul style="list-style-type: none"> 제조업에서 도구변수 모형은 어느 정도 의미 가격은 영향 미미
GMM	$R^2= 0.459$	합산이동시간(-, $p=0.034$) price_per_area_log(+), 유의성 낮음	<ul style="list-style-type: none"> 외곽 입지라도 공실이 낮은 현상 재확인 <ul style="list-style-type: none"> 가격 영향은 통계적으로 크지 않음

TSRI - 결과 (서비스업 결과 해석)

모형	R^2	유의 변수/계수 부호	해석/시사점
제어함수 접근	$R^2 = 0.596$ p값 유의	final_efa_Factor2(+, p=0.004) int_price_per_area_log_final_efa_Factor2(-, p=0.004)	<ul style="list-style-type: none"> 단순 상권클러스터 (Factor2) 높으면 공실 그러나 가격이 함께 오르면 (기업클러스터가 형성되어 있으면) 그 효과가 일부 상쇄 초기 기업클러스터 형성시 공실률이 낮게 형성
TSRI	RMSE= 0.1262 MAE= 0.1061 $R^2 = 0.4172$	서울과 대중교통 접근성은 의미 없고 Factor2/상호작용이 중요한 편	GPR로 비선형적 관계 부분 포착 기업클러스터와 상권클러스터의 균형점이 중요 TSRI 접근이 이 상호작용을 더 정확히 포착
2단계 추정	$R^2 = 0.5447$	대부분 추정계수 통계적 유의성 낮음	의미 X
GMM	$R^2 = 0.5447$	final_efa_Factor2(+), 상권클러스터, 가격(기업클러스터 형성, 선호도)(-)이나 p값 0.19~0.20 수준(유의성 미달)	상호작용 포착을 하지 못하며, 클러스터 개념 설명 불가

TSRI - 분석 총결론

업종별로 나누어 지식산업센터의 위치 선정
교통, 공시지가와 상권클러스터에 기업이 선호할 적정구간을 찾아 기업이 선호하게 만들고,
업종별 기업클러스터를 빠르게 형성시켜야 공실률을 최소화

Bartik(2021): 단순 유흥/학원 상권은 자족기능이 부족해, 기업 입지가 기피될 가능성이 크다고 지적 본 결과가 이를 지지: “유흥/학원 편중 상권은 전체 공실률에 악영향”. 그러나 기업클러스터가 형성되면 “생활형 인프라”로 작동해, 그 표본만 보면 공실률이 낮아지는 상반 결과가 동시에 존재. 첨단·지식 분야가 무조건 도심 고가 입지를 선호한다는 일반론과 달리, “저렴한 임대 공간을 찾아 교외로 나가는 현상”도 빈번하다고 주장 결과적으로, 비싼 지역이 공실률라는 전통적 패턴과 달리, 지식기반서비스 한정에선 임대료가 상한선(Threshold)을 넘으면 공실 증가.

Florida(2020): “창의·지식집약 산업은 부동산 비용 효율에 민감해, 중간수준 임대료가 형성된 지역에서 더 많이 성장한다” 너무 비싸면, 창업·벤처 기업은 외곽이나 인근 임대료가 낮은 곳으로 이전(공실률)이 발생.

Glaeser(2012): “도심 공동화” 가능성을 언급했으며, 재개발이 지연된 곳은 공실이 누적될 수 있다 마찬가지로 초기에 클러스터 형성이 제대로 이뤄 지지 않으면 공동화 상태가 지속될수 있음 또한 “고학력·전문직 클러스터는 주변의 학습·취미·외식 문화가 도시 내 집적을 더 가속화한다”고 봄.

Roth & Sheppard(2019): “지식기반 근로자들은 사무실 인근의 문화·교육·취미시설이 풍부할수록 ‘직장만족도’가 상승해 임차 유지율이 높아진다”고 분석.

“ 기존 사회과학이론 수학적으로 재확인 ”

04. 신도시 적용

인천 계양구를 중심으로

- 균등분할
- 부분요인배치

신도시 적용

◦ 신도시 전처리

1. CAD 데이터 변환&격자 결합

- CAD 파일을 외부 CAD 프로그램을 활용해 DXF 형식으로 변환
- WGS84 위경도 좌표계(EPSG:4326)로 변환
- CAD의 텍스트 정보를 CSV로 저장
- 이를 계양구 100m 격자 데이터와 결합

2. 격자별 평균 공시지가 산출

- 이미 확보해 둔 계양구 공시지가 파일을 이용하여, 100M 격자 단위로 평균 공시지가 산출

3. 최소이동시간·교통수용율 산출

- 성남시와 마찬가지로 최소이동시간(강남 까지 걸리는 대중교통이동시간)과 교통수용율 산출

4. EFA 기반 상권 클러스터링

- 탐색적 요인분석(EFA)으로 도출된 요인적 재량을 활용해 상권·소비·유동 등을 대표하는 요인점수 계산
- 이를 통해 계양구 상권클러스터(factor2)를 구하고 가격, 교통, 격자위치, 블록명 등과 함께 csv로 저장

◦ 신도시 사용 데이터 파일

상권 클러스터 지수

- 계양구 카드 매출 데이터
- 계양구 성연령별·시간대별·요일별 유동인구 데이터
- 계양구 상권정보 데이터

격자별 공시지가(가격)

- 국토교통부_표준지공시지가_20240101.csv
- 230619_2023년도 표준지 공공데이터포털 자료_황민경(GS-1425).csv
- LSMD_CONT_LDREG_28245_202501.shp

격자별 교통 점수

- (참고)3-3. 231127 인천계양 지구단위계획구역 결정도(2차).dxf
- 서울교통공사_1_8호선 역사 좌표(위경도) 정보_20241031.csv
 - 국토교통부_전국 버스정류장 위치정보_20241028.csv

신도시 적용

◦ 상권 클러스터 지수 산출

1. 격자 단위 데이터 산출

- 계양구 카드 매출 데이터
- 계양구 성연령별·시간대별·요일별 유동인구 데이터
- 계양구 상권정보 데이터



격자 단위로 정리하여 집계
분석을 위한 주요 지표를 추출

2. 요인적재값 기반 업종 필터링

- 성남시 EFA에서 계산된 요인적재값을 활용하여 업종 필터링
- 유사한 요인값을 가지는 동일·유사 업종만 선별
- 예) 성남시에서 '카페·베이커리' 중심 요인이 도출되었다면, F&B(음식) 업종에 집중하여 필터링

3. 요인적재값 가중합을 활용한 상권 분석

- 필터링된 업종의 상권, 카드, 유동인구 데이터를 요인적재값으로 가중 합산
- 가중합 결과를 활용해 최종 상권 클러스터 도출
- 격자별 상권 특성을 분석하여 공간적 분포 파악

신도시 적용 - 최적점 선정

1. 분석 구간 설정

- 데이터의 극단값(상·하위 10%)을 제외하고, 중간 80% 구간(10%~90%)을 집중 분석
- 4차원 공간에서 10%~90% 구간을 선택하면, 이는 4차원 직육면체 형태로 정의
- 각 축을 N등분하면, 전체적으로 $(N+1)^4$ 개의 격자점 생성

2. 공실률 최소화 분석

- 2단계 모델의 예측값을 활용하여, 부분요인배치 실시
- 상권클러스터와 가격의 상호작용을 기반으로 주요 변수(price, factor2, 건물특성, 교통특성) 조정
- 다양한 조합에서 공실률 예측값을 비교하여 최소 공실률을 유도하는 최적 조합을 찾고, 상호작용 효과 해석

3. 최적점 적용

- 부분인자설계를 통해 81개의 최적 조합 도출
- 해당 최적점을 계양구 데이터에 적용하여 분석 진행

신도시 적용 - 스케일 조정

도시 간 지표 스케일 차이 조정

문제점

- 성남시와 계양구의 변수 범위 다름 (예: 공시지가와 실거래가 차이)
- 두 도시의 데이터 분포(하이퍼박스)가 겹치지 않으면 직접적인 거리·유사도 비교 어려움

$$[x_{1,\min}, x_{1,\max}] \times [x_{2,\min}, x_{2,\max}] \times \cdots \times [x_{n,\min}, x_{n,\max}]$$

성남시 데이터가 놓이는 하이퍼박스

$$[x'_{1,\min}, x'_{1,\max}] \times [x'_{2,\min}, x'_{2,\max}] \times \cdots \times [x'_{n,\min}, x'_{n,\max}]$$

계양구 데이터가 놓이는 하이퍼박스



해결 방안

- 각 변수 x_j 를 $[0,1]$ 범위로 정규화(스케일링)하여 아핀 변환 적용
- 이를 통해 서로 다른 도시의 데이터를 동일한 기준에서 비교 가능

신도시 적용

계양구 최적 위치 선정

$$d_{L1}(x, y) = \sum_{j=1}^n |x_j - y_j| \quad (x, y \in [0, 1]^n)$$

- 계양구 데이터를 스케일링한 후, 최적점과의 맨해튼 거리 계산
- 거리 값 $\{d_1, \dots, d_m\}$ 중 가장 작은 5개 지점 선택
- 해당 계양구 주변 환경과(가격, 상권클러스터, 교통) **가장 유사한 조건**에서 **공실률이 최소가 되었던 스텝 5개**를 얻게 됨
- 이를 바탕으로 격자별 최적 면적 점수를 산출하고 **추천 위치(격자)** 선정

계양구 최적 면적 선정

$$\bar{M} = \frac{\sum_{k=1}^5 w_{i_k} M_{i_k}}{\sum_{k=1}^5 w_{i_k}}, w_{i_k} = \frac{1}{d_{i_k} + \varepsilon}$$

- 선택된 5개 지점의 근방 해들을 가중 평균하여 최적 값 도출
- 면적 역변환(면적요인 \rightarrow 실제면적)을 적용하여 **최적 면적** 산출

05. 신도시 입지/면적 추천

인천 계양구를 중심으로

-
- 제조업 입지·면적 추천
 - 서비스업 입지·면적 추천
 - 기대효과 및 추가 개선점

결과1 - 지식기반 제조업



추천 위치: 산업5
추천 연면적: 약 30,000m²

1. 유동인구 증가 → 공실률 상승 가능성

- 인구 밀집도가 높을수록 제조업 건물의 공실이 증가할 가능성을 시사
- 해당 지역은 오피스텔 규제가 있어 과밀 가능성이 낮으며, 균린공원 등 근로자 편의성이 높음

2. 서울 접근성 → 제조업 선호 증가

- 적절한 대중교통 접근성을 갖춰 투기세력 유입과 임대료 급등 가능성이 낮음
- 제조업 기업들이 더욱 선호하는 지역이 됨

3. 넓은 부지 면적 → 제조업 최적 조건

- 충분한 부지 면적은 생산·물류·R&D 시설 운영의 핵심 요인으로, 제조업 기업의 입지 선호도를 높임
- 특히 산업4 지역과 달리 창고 면적 제한이 없어 제조업 유치 가능성이 높음

4. 제조업 클러스터 효과

- 인접한 산업4 지역과의 기업 밀집도가 증가하면, 협업 시너지 및 집적 이익 발생
- 이를 통해 지역 내 제조업 생태계 조성에 유리한 환경을 제공

5. 오피스텔 제한 지역 → 안정적인 제조업 입지

- 오피스텔 입주 제한 구역에서는 부동산 가격 급등 가능성이 낮음
- 제조 기업들이 안정적으로 입지할 수 있는 유리한 환경을 조성

결과2 - 지식기반 서비스업



추천 위치: 자족31
추천 연면적: 약 25,000m²

1. 상권 클러스터와 지식기반서비스업의 상호작용

- 상권 클러스터(Factor2)가 높은 지역에 입지할 경우, 초기 입주 효과가 뚜렷하고 지역 가치 상승에 기여
- 단, 가격이 지나치게 높으면 일부 상호작용 효과가 약화될 가능성이 있음
- 해당 입지는 **발달한 상권 대비 상대적으로 낮은 가격을 유지**

2. 자족31의 클러스터 효과 및 입지 조건

- 도시형공장, 벤처기업, 소프트웨어진흥시설 등을 주용도로 허용
- 전시장, 회의장, 창고시설 등 부용도 시설도 수용 가능하여 기업 및 상권 집적 가능성 높음
- 지식기반서비스업의 클러스터 효과를 극대화할 수 있는 구조**

3. 자족31·32·33의 연계성

- 유사 업종과 서비스가 한 곳에 집중된 클러스터 효과가 공실률에 큰 영향을 미침
- 자족31과 인접한 자족32·33이 동일 용도를 포함하고 있어, 지식기반서비스업의 집적에 유리

4. 기업 집적 및 산업 특화 환경 조성

- 다양한 기업이 입주할 수 있는 구조로, 지식기반산업과 지원 시설이 함께 조성
- 기업 간 협업 및 네트워크 구축에 최적화된 환경 제공**

5. 주변 환경과 기업 만족도

- 공원 및 커뮤니티 용지 인접, 종사자 만족도를 높여 임차 유지를 향상
- 이는 기업 클러스터 형성에 유리한 조건으로 작용

결론

기대효과

1

- 정교한 통계 기법을 적용함으로써, 지식산업센터 관련 문제를 보다 과학적이고 체계적으로 접근 및 해결 가능

2

- 수학적 분석과 도메인 지식을 균형있게 고려하여 한쪽에 치우치지 않고 현실적인 해결 방안 도출 가능

3

- 도시의 복잡한 상호작용을 설명할 수 있는 구체적 지표를 제시하여, 실무 적용에 용이하고 직관적으로 이해할 수 있는 결과물 도출 가능

4

- 실제 규제사항 및 용도계획을 충분히 반영함으로써, 이론과 실제를 아우르는 현실적인 방안 마련 가능

추가 개선점

1

- 하이퍼 파라미터 조정, 시간대 조정 등 다양한 보완을 통해 보다 풍부하고 현실적인 분석 기대됨
- SHAP등 추가 시각화를 도입하고 반복적으로 개선함으로서 더욱 직관적인 결과 도출

2

- 추가적인 데이터 확보를 통해 더 폭넓은 분석이 가능해질 것으로 예상

3

- 시계열 분석과 공간 분석 기법을 도입하면 다양한 해석과 심층 분석이 이루어짐

4

- 전국 단위로 표본을 확장하여 분석 결과의 일반화 및 다양한 상황에 대응할 수 있는 이론 마련 가능



감사합니다.

참고문헌

- Roth, A., & Sheppard, S. (2019). Workforce Cultural Amenities and Regional Growth. *Journal of Urban Economics*, 115(3), 45–67.
- Glaeser, E. L. (2012). *Triumph of the City*. Macmillan.
- Florida, R. (2020). *The Rise of the Creative Class Revisited*. Basic Books.
- Bartik, T. J. (2021). *Local Economic Development Policies*. Upjohn Institute for Employment Research.
- Eppli, M. J., & Tu, C. C. (1999). Valuing the New Urbanism: The Impact of the New Urbanism on Prices of Single-Family Homes. *Urban Land Institute*.
- Comparing the performance of two-stage residual inclusion method
- 2SLS vs 2SRI: Appropriate Methods for Rare Outcome
- Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, 27(3), 531-543.
- Rosen, H. S., & Smith, K. (1983). Estimating the natural vacancy rate: A theoretical and empirical analysis. *Journal of Real Estate Research*, 1(2), 123–145.
- Wheaton, W. C., & Torto, R. G. (1988). Dynamics of office market vacancies and rent adjustments. *Real Estate Economics*, 16(3), 367–389.
- Shilling, J. D., Sirmans, C. F., & Corgel, J. B. (1992). Office market dynamics: Vacancy and rent relationships in commercial real estate. *Journal of Real Estate Finance*, 10(2), 175–193.
- 김성혜, & 고현림. (2016). 산업단지 외부 지식산업센터의 임대료 결정요인에 관한 연구: 서울 성수동 지식산업센터를 중심으로, 주거환경(한국주거환경학회논문집), Vol.14 No.2 [2016], 201-218. 2016
- 김대봉, "환경 및 물리적 서비스특성이 지식산업센터의 가격에 미치는 영향에 관한 연구" *한국주거환경학회* 13 (13): 253-263, 2015

활용 공공데이터 출처

- 공공데이터 포털 (계양구 지하철역, 버스정류장 , 공시지가, 상권정보)
URL: <https://www.data.go.kr/>(<https://www.data.go.kr/>)
- 국토교통부 브이월드 디지털 트윈국토(SHP파일)
URL: https://www.vworld.kr/dtmk/dtmk_ntads_s002.do?dsId=30563[https://www.vworld.kr/dtmk/dtmk_ntads_s002.do?dsId=30563]
- 국토교통부 부동산 실거래가 공개시스템(성남시 실거래가 파일)
URL: <https://rt.molit.go.kr/>[<https://rt.molit.go.kr/>)