

조원 소개

양동근



기획, 구현, 발표

주재민



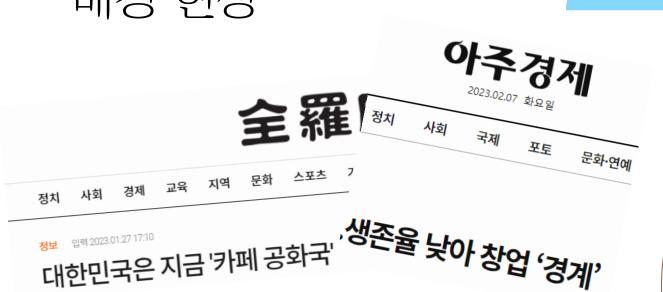
시각화, 전처리

김정한



서칭, 딥러닝 모델학습

배경 현상



우후죽순창업하지만카페폐업률 14%, 치킨집보다 표이

용 김종일기자(idea@sisajournal.com) │ ② 승인 2020.02.11 14:00 │ 및 호수 1581

서울신용보증재단 ── 소상공인 폐업률 ↓



연구 목표

✓ 1. 상권에서 카페의 수는 몇 개가 적정한가

✓ 2. 상권에서 카페의 수는 가까운 미래에 몇 개가 될까?



연구 목표

✔ 기준 = 상권(코드)

✔ 타겟 = 카페 수



선행 연구



- 4. 상권 및 입지분석
 - 커피전문점의 적합 입지는 직장인을 공략 할 수 있는 오 1) 커피전문점 적합입지 피스빌딩 밀집지역, 젊은층의 유동이 많은 대학가, 접근성 이 우수한 역세권 등에 입점하는 것이 좋다. 커피전문점은 20~30대의 여성이 주고객층을 형성하고 있으므로 여자대

소상공인지식

간행물

간행물

제목 커피전문점 창업가이드

- 2) 상권 및 입지 평가방법
- 나. 상권세력의 특성을 파악
- 사. 상권의 발전성을 점검

- 가. 주변지역의 특성을 파악
 - ① 아파트 상권
- 다. 유동 인구의 특성 파악
- 아. 자금에 맞는 입지를 선정
- ② 지하철역 역세권 상권 라. 점포 이용편리성의 파악
- 바. 경쟁점의 강/약점을 조사

학교 주변 상권

마. 교통요인을 점검

④ 오피스 상권

골목상권 내 카페업종 매출액에 영향을 미치는 요인에 관한 연구 factors influencing the sales of cafe in alley tread area godes.

학위논문사항

정실로

2018

서울: 서울과학기술대학교, 2018 학위논문(석사) - 서울과학기술대학교 일반대학원 데이터사이언스학과 2018 EDA 가공 및 데이터 설명



모델 생성 및 검증 시각화

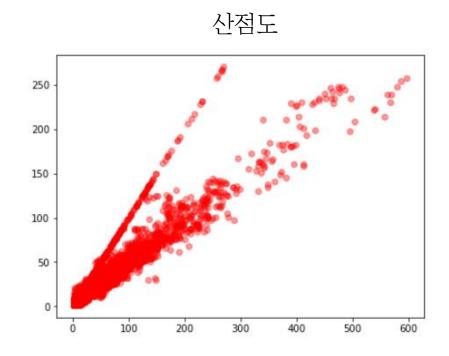


Target = 총 점포 수 (STOR) vs 점포수 (SELNG) (서비스_업종_코드 = 카페)

	상권집계_매출_분기	TBSM_TRDAR_SELNG_QU	76	AGRDE_60_ABOVE_SELNG_CO	연령대_60_이상_매출_건수
	상권집계_매출_분기	TBSM_TRDAR_SELNG_QU	77	STOR_CO	점포_수
	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	1	STDR_YY_CD	기준_년_코드
	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	2	STDR_QU_CD	기준_분기_코드
	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	3	TRDAR_SE_CD	상권_구분_코드
	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	4	TRDAR_CD	상권_코드
	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	5	SVC_INDUTY_CD	서비스_업종_코드
	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	6	STOR_CO	일반_점포_수
	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	7	SIMILR_INDUTY_STOR_CO	총_점포_수
_	상권_점포_개폐업_분기	TBSM_TRDAR_STOR_QU	8	OPBIZ_RT	개업_율



Target = 총 점포 수 (STOR) vs 점포수 (SELNG) (서비스_업종_코드 = 카페)



기술통계량

SELNG['STOR_CO'].describe()

count	18510.000000
mean	23.116207
std	42.297863
min	1.000000
25%	5.000000
50%	9.000000
75%	22.000000
max	598.000000

Name: STOR_CO, dtype: float64

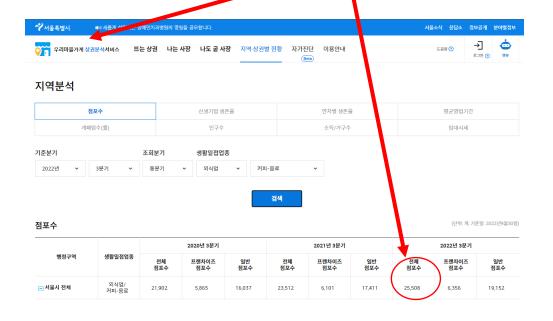
STOR['SIMILR_INDUTY_STOR_CO'].describe()

count	22081.000000
mean	11.994701
std	21.341324
min	0.000000
25%	3.000000
50%	6.000000
75%	12.000000
max	270.000000

Name: SIMILR_INDUTY_STOR_CO, dtype: float64



Target = 총 점포 수 (STOR) vs 점포수 (SELNG)





활용사례(갤러리) 등록

URL 복사

목록 이동

서울시 우리마을가게 상권분석서비스(상권-추정매출)

서울시 골목상권 점포의 추정매출 정보를 제공합니다. (분기별 데이터가 적재됩니다.) 자주하는 질문은 F&Q을 다운받아 확인해 주시기 바랍니다.

데이터의 사이즈가 커 미리보기의 다운로드 서비스가 제대로 작동되지 않을수 있습니다.

구 상권 데이터의 경우 우리마을가게 상권분석 서비스 상권-점포수 데이터를 참조하는 과정에서 발생한 문제입니다



● 수많은 엑셀

- TBSM_TRDAR_APT
- TBSM_TRDAR_BDPOP_QU
- TBSM_TRDAR_BEING_RT_QU
- TBSM_TRDAR_FCLTY
- TBSM_TRDAR_FLPOP_QU
- TBSM_TRDAR_HOUSINCOM_QU
- TBSM_TRDAR_RENT_CURPRC_QU
- TBSM_TRDAR_REPOP_QU
- TBSM_TRDAR_SALE_MT_QU
- 🔚 TBSM_TRDAR_SELNG_QU
- TBSM_TRDAR_STOR_QU
- TBSM_TRDAR_SURV_RT_QU
- **▼** TBSM_TRDAR_WRC_POPLTN

● 타겟 = 카페 수

***STOR(점포):<mark>22081</mark>

SELNG(매출):18510

APT(아파트): 4412

FCLTY(시설):1569

HOUSIN(가구소득):24213

RENT(임대):69674

SALE(영업개월):23360

SURV(신생_생존):22139

BERT(생존율):14384

₩RC(직장인구):24402

FLPOP(유동인구):24937

BDPOP(상존인구):24384

REPOP(상주인구):24382

EDA

전체 살펴보기

✔['상권코드', '연-분기']별

✔ 매우 짧은 시계열

EDA

전체 살펴보기

● 전체 Column 수가 매우 ↑ (over 1000)



행 추출

✓ 1. 대표성

'월 매출액' '영역 면적' '총 유동인구' '총 직장인구' '총 가구수' '상권구분코드' '총 직장인구' ✔ 2. 압축가능성

행 삭제 시 정보 loss가 충분히 적을 여지가 있는



행 추출 - 압축가능성

10대유동인구x15 + 20대유동인구x25 + ··· + 60대이상유동인구x65

유동인구 평균 연령 =

10대유동인구 + 20대유동인구 + … + 60대이상유동인구

추정이익 = log(월 매출액-평균 환산 임대료)



행 추출 - 압축가능성

시간대3인구 + 시간대4인구

카페 피크타임 인구비율 =

시간대1인구 + 시간대2인구+ … + 시간대6인구

교통 = 버스 + 가중치*지하철 + 가중치*철도 +가중치*터미널

쓸 수 있을지 여부는 추후 판단, 일단 만들어보자.



결측치



● 평균으로 대체

● 0으로 대체

총점포수

인구수

시설 수

etc

etc

etc

EDA 스케일링

✔ 전체적으로 우측으로 꼬리가 긴 분포

✔ 이상치 제거?

✔ RobustScaler (꼬리가 긴 분포에 매우 유리)



외부 데이터 활용

✓ 코로나

코로나 거리두기 포장만 됐던 전국 카페 매장 취식 가능해져



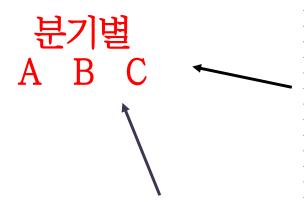
입력: 2021-01-16 11:44 | 수정: 2021-01-16 11:44

✔ 2019 -2022 시계열의 핵심



외부 데이터 활용

✓ 코로나



-			
비율	16.3%	0.0%	0.0%
20년 4분기	38	0	18
비율	41.3%	0.0%	19.6%
21년 1분기	90	0	90
비율	100.0%	0.0%	100.0%
21년 2분기	91	0	91
비율	100.0%	0.0%	100.0%
21년 3분기	0	81	11

2020 6월 단계(수도권만)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
이슈																단계별 리두기															
2020 7월	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30 3	1
단계(수도권만) 이슈																1															
∪π				2.5	이성	라만 남만	<u>카</u> 퍼	 취	식금	지																					_
2020 8월	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30 3	_
단계(수도권만) 이슈	1 사랑제일교회												2											2.5							
Чπ										^[;	5' AI	511	- 보																		_
2020 9월 단계(수도권만)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
이슈	П	Т	П	Т	П		2.5	Т	П													2							Т		
-111																															-
2020 10월	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30 3	1
단계(수도권만)						2															1	<u> </u>									
이슈	_			_	_	_	_	_																					_		_
2020 11월	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
단계(수도권만)									1												1.5						2				
이슈																	3차대유행														
2020 1281	-	_	_		_					4.0	_	단계									금지									20 2	_
2020 12월 단계(수도권만)	1	2	3	2	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19		21	22	23	24	25	26	27	28	29	30 3	1
이슈	П	П			П	Т		П								2.5 5인															
																							_								_
2021 1월	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30 3	1
단계(수도권만)																2.5															
이슈																5인															
2021 2월	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28			
단계(수도권만)	- '	۷	٥	4	٦	U	2.		Э	10	11	12	13	14	13	10	17	10	13	20	21		23	24	23	20	21	20			
이슈														59	21																
																															-



외부 데이터 활용

📠 서울시 우리마을가게 상권분석서비스(상권영역)

✓ 상권코드 ~ 행정동

✓ 주변시설, 아파트row 길이 정상적인 데이터 획득

kai 서울시 우리마을가게 상권분석서비스(상권-집객시설)

📠 서울시 우리마을가게 상권분석서비스(상권배후지-아파트)

***STOR(점포):<mark>22081</mark>

SELNG(매출):18510

APT(아파트):4412

FCLTY(시설):1569-

HOUSIN(제구소득):24213

RENI(임대):69674

SALE(영업개월):23360

SURV(신생_생존):22139

BERT(생존율):14384

₩RC(직장인구):24402

FLPOP(유동인구):24937

BDPOP(상존인구):24384

REPOP(상주인구):24382



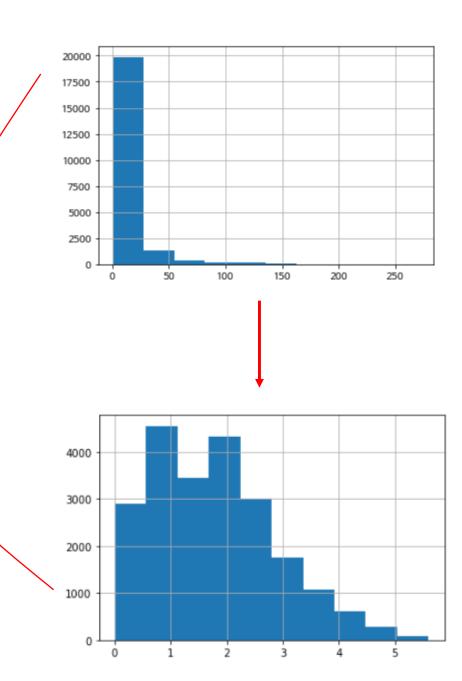
산점도

✔ Correlation의 한계

✔ 일일이 그려보기

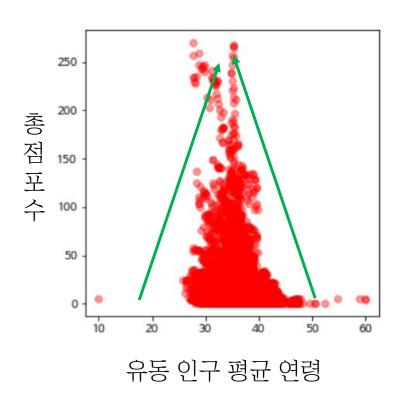
√'총 점포 수' 우측 꼬리

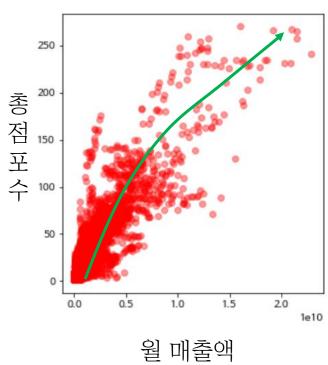
→ LOG scale시도

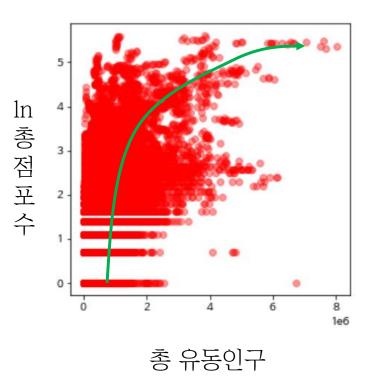




산점도



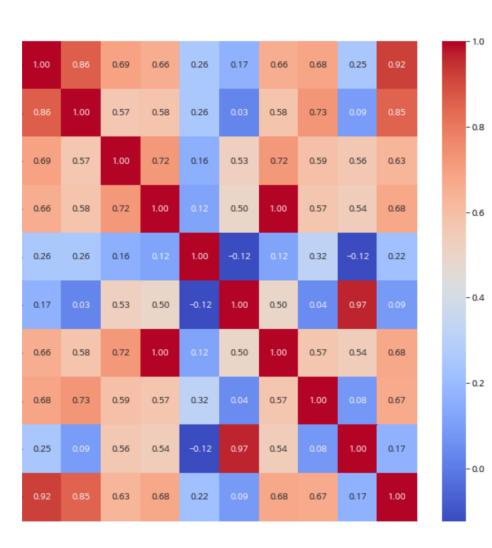






Correlation Heat map

- ✓ Corr
- → 설명변수 간 다중공선성 파악





차원축소

✔ 다중공선성을 줄일 필요

- ✓ Domain knowledge 활용
- ✓ PCA? or Factor analysis?
- ✓ Ridge Lasso

→ 모델 적합 시 판단

EDA 차원축소

✓ 타겟 = 총 점포 수

(총 점포수' = ('프차 수'+ '일반점포 수') → 비율

● SELNG '점포수'

Model 방향성

✔ '상권 구분 코드' → 3가지 (D / R / A)

- ✔ 앞서 살펴본 선행 연구…
- ✔ 동일성 검정(MANOVA분산분석) → 유의한 차이

= 모델을 각각 적합하면 성능이 좋아질 거라는 예상

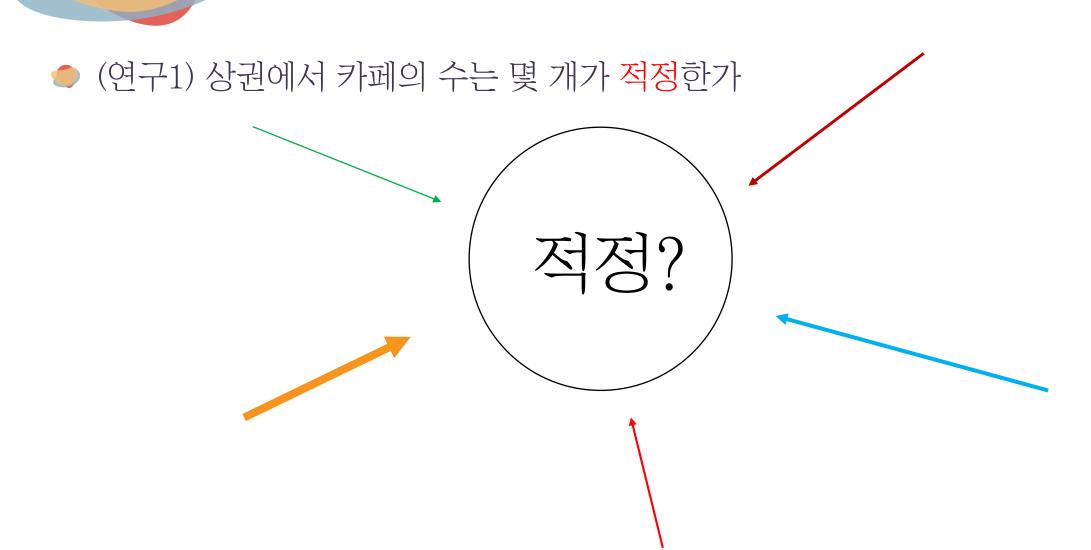
Model

Model -회귀 (연구1)

●(연구1) 상권에서 카페의 수는 몇 개가 적정한가

Model

Model -회귀 (연구1)



Model

Model -회귀 (연구1)

● (연구1) 상권에서 카페의 수는 몇 개가 적정한가

점포당매출 ↑, 폐업률 ↑

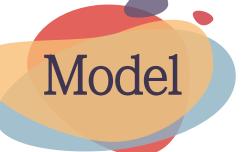
적정?

성장률 ↑, 점포당매출↓

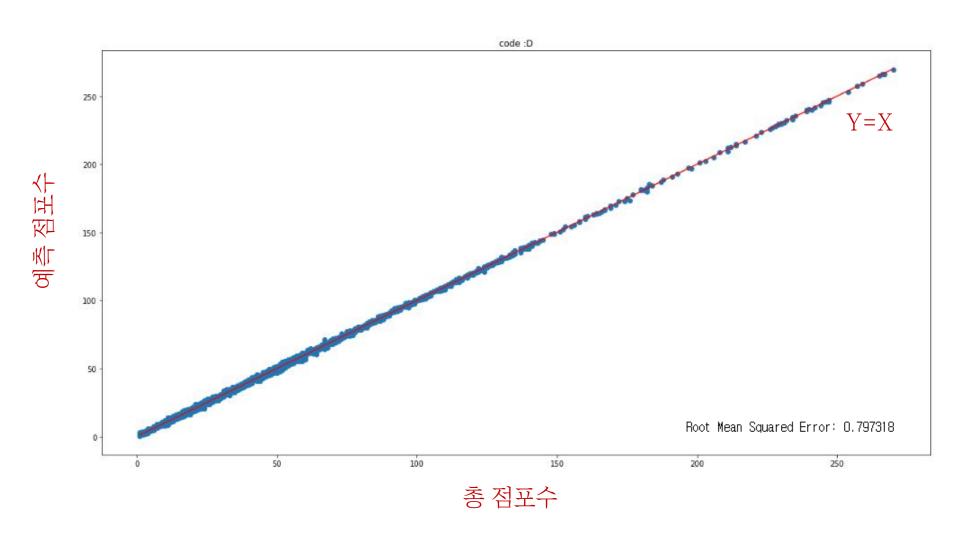
모범적 상권사례? but 상권마다의 특성

● (연구1) 상권에서 카페의 수는 몇 개가 적정한가

- ✓ 전수 데이터 (Data volume ↑)
- ✔ '적정함'의 Pre-모형? → 합쳐진 데이터가 가리키는 방향
- ✔ 어느 features에 가중치를 줄지는 모델이 학습할 것
- ✔ XGB regressor (*boosting 가중치 갱신)

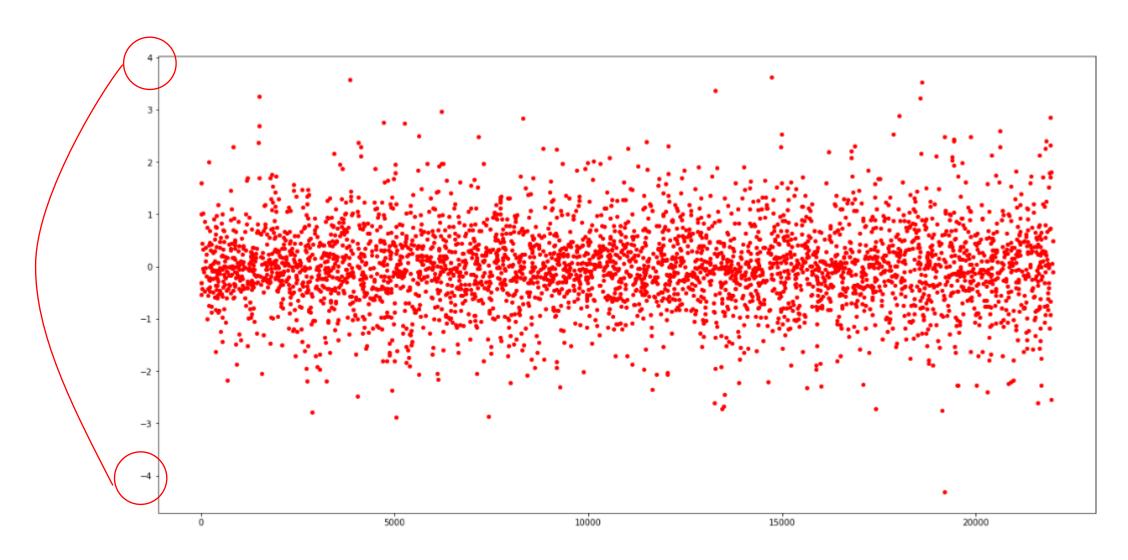


● 상권D = 발달상권 (기본)





● 상권D = 발달상권 (기본)





● 상권D = 발달상권 (기본)

```
def predict_r(df, code, timeord):
    if len(df.loc[(df['상권코드'] == code) & (df['timeord'] == timeord),:]['총점포수'].values)==0:
    #print(code, timeord)
    return None,None,None
else:
    actual_y = df.loc[(df['상권코드'] == code) & (df['timeord'] == timeord),:]['총점포수'].values[0]
    X=df.loc[(df['상권코드'] == code) & (df['timeord'] == timeord)].drop(['총점포수','상권코드'],axis=1)
    X=pcar.transform(X)
    prediction= regr.predict(X)[0]
    print('실제 접포수:' + str(actual_y) + '\mmx텐설 접포수:'+ str(prediction) + '\mmediction + str(actual_y-prediction))
    return actual_y, prediction, actual_y-prediction

predict_r(dfrr, 2130314, 1)
```

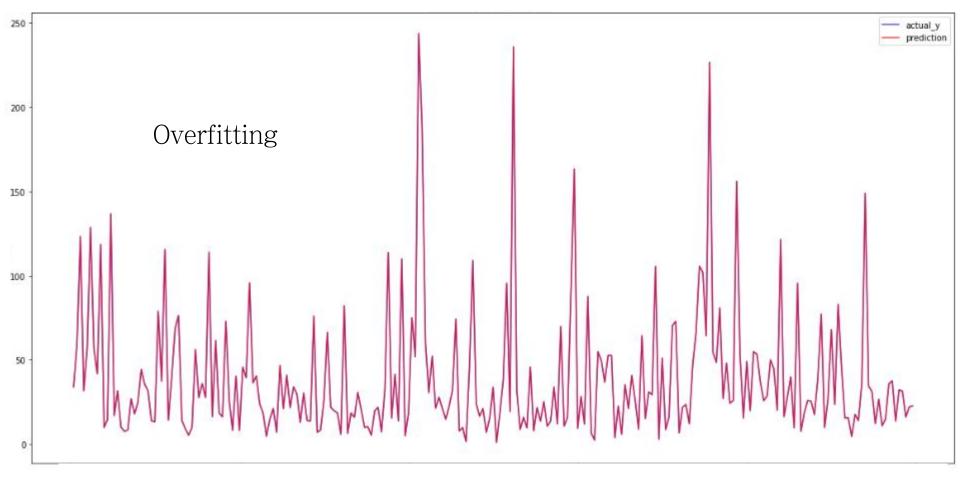
실제 점포수:6.0 포텐셜 점포수:6.2279186

실제-포텐셜:-0.2279186248779297

✔ ['상권코드', '분기']를 입력하면 실제 점포 수와 예측 점포 수를 반환



● 상권D = 발달상권



'상권코드'별 '분기'에 따른 실제점포수와 예측점포수의 '평균'



Parameter 조정

Train-Val-Test

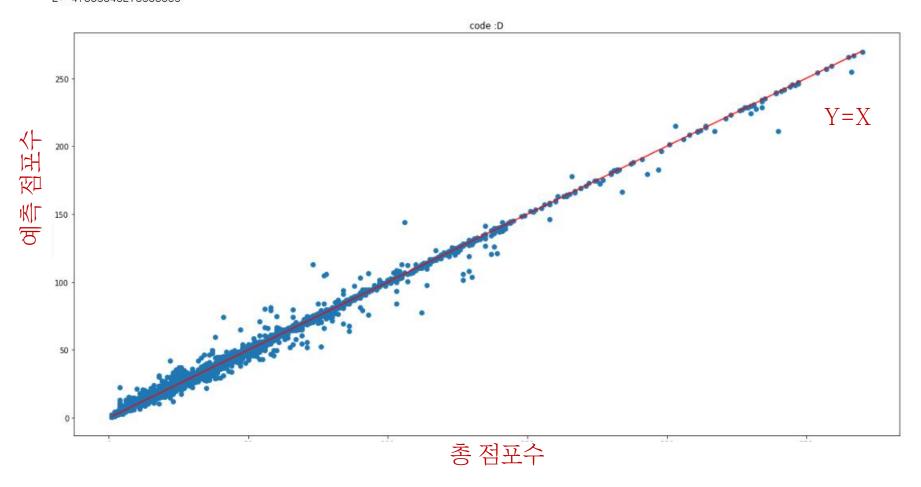
과적합

PCA

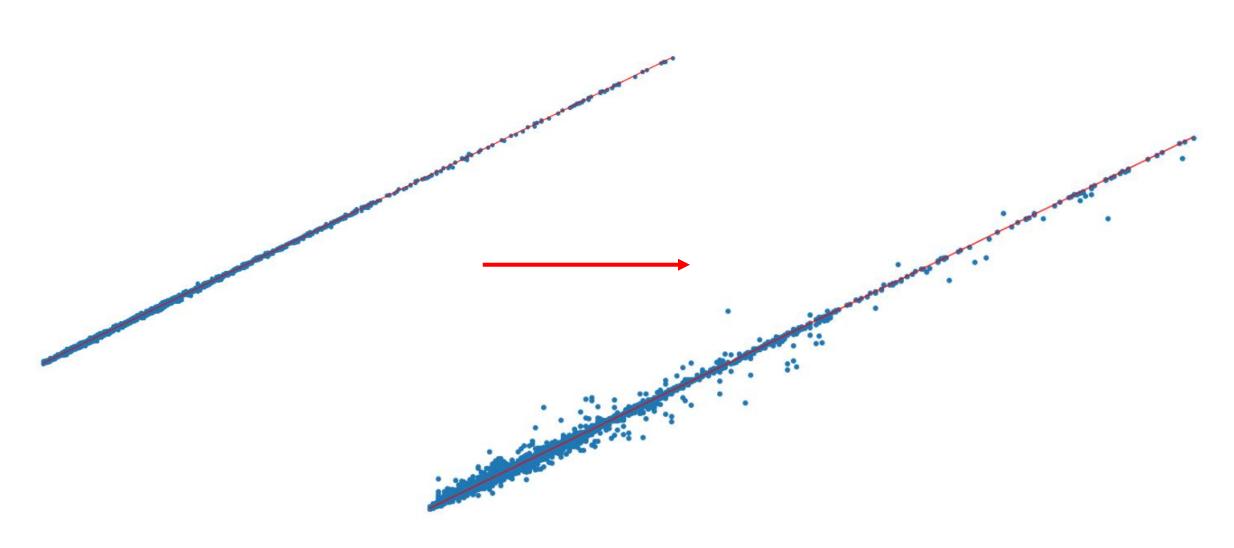
Ridge, Lasso



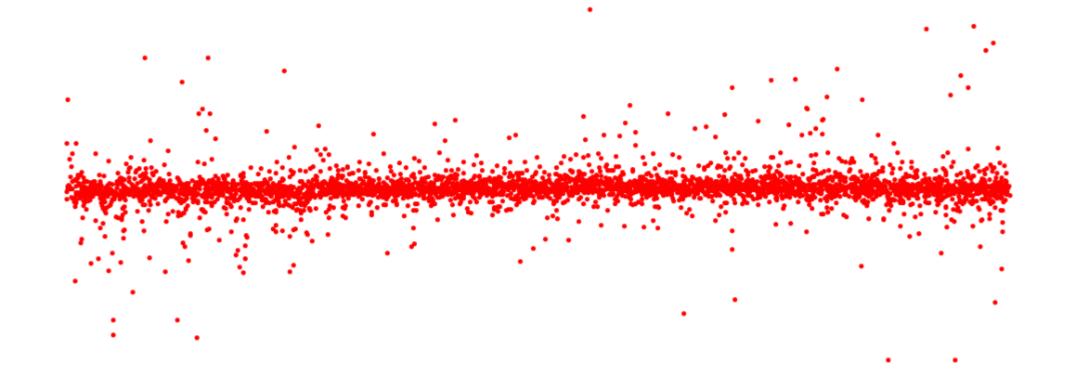
E: 4.039045278965389



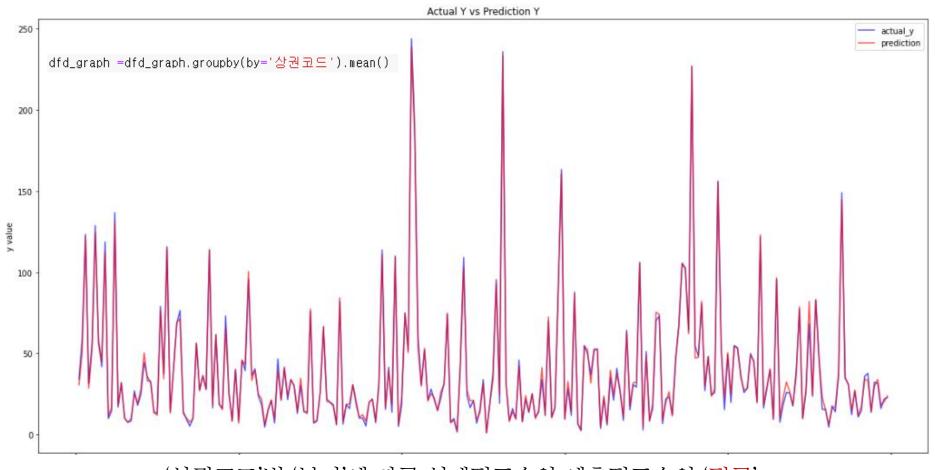










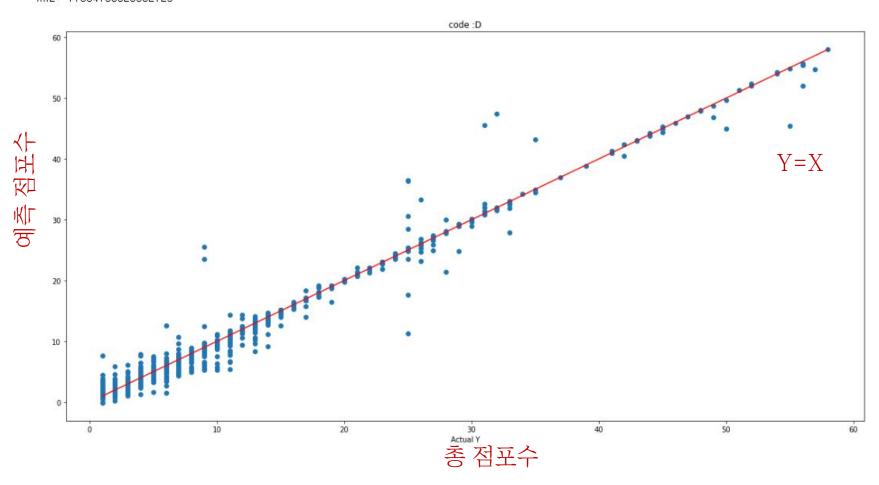


'상권코드'별 '분기'에 따른 실제점포수와 예측점포수의 '평균'



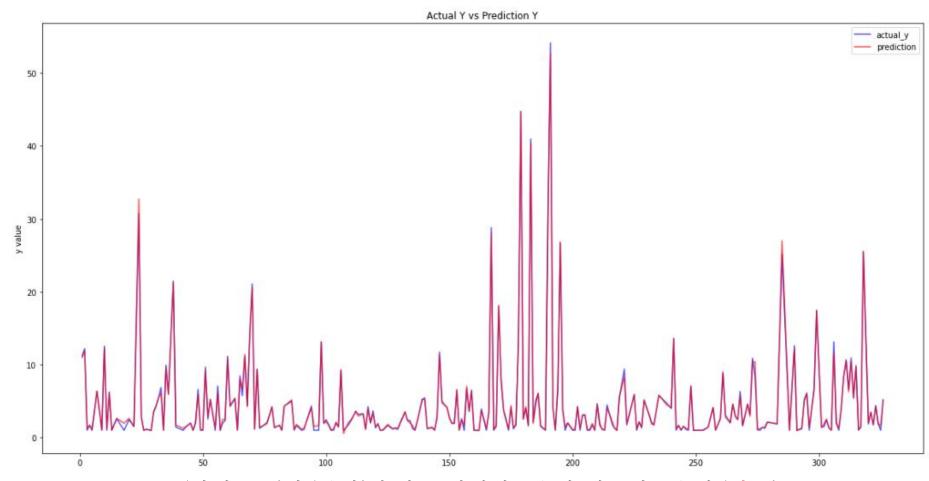
● 상권R = 전통상권

MAE: 1.034736825682728

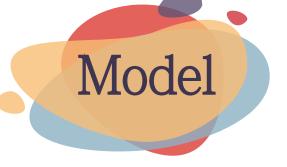




● 상권R = 전통상권

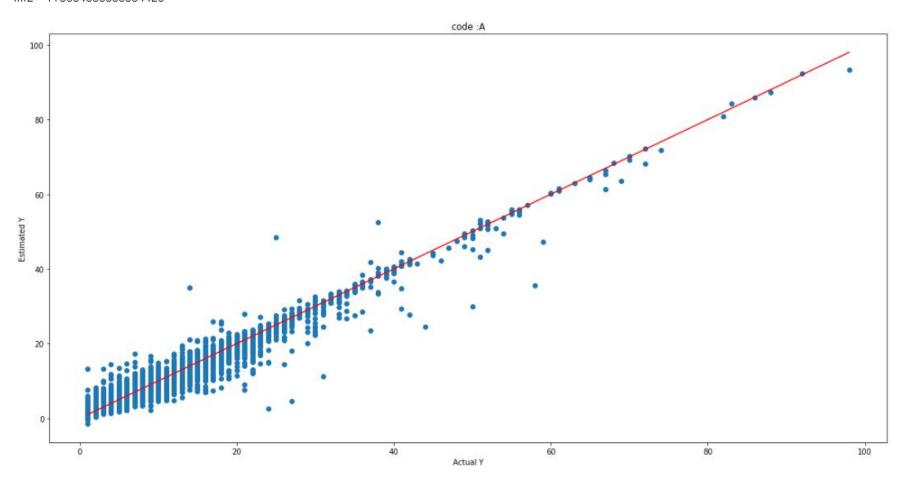


'상권코드'별 '분기'에 따른 실제점포수와 예측점포수의 '평균'



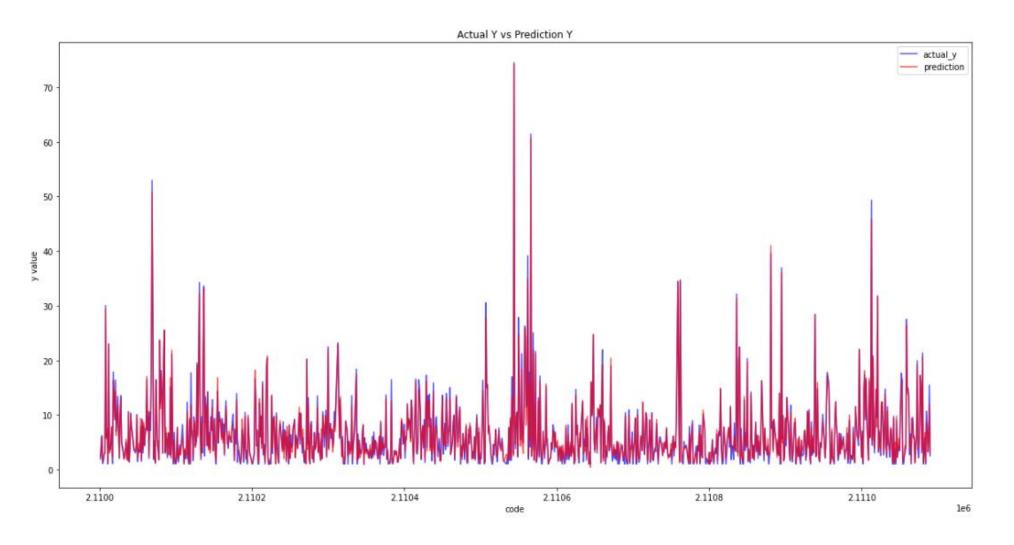
● 상권A = 골목상권

MAE: 1.5954068388304423





● 상권A = 골목상권



Model -회귀 (연구1)

●(연구1) 상권에서 카페의 수는 몇 개가 적정한가



여유 ~ 과밀 시각화(그라데이션)

잔차 = 실제값 - 예측값

기준= 잔차 / 실제값

Rank = '기준'에 대한 Rank

여유 적정 과밀 3중 분류(단색)



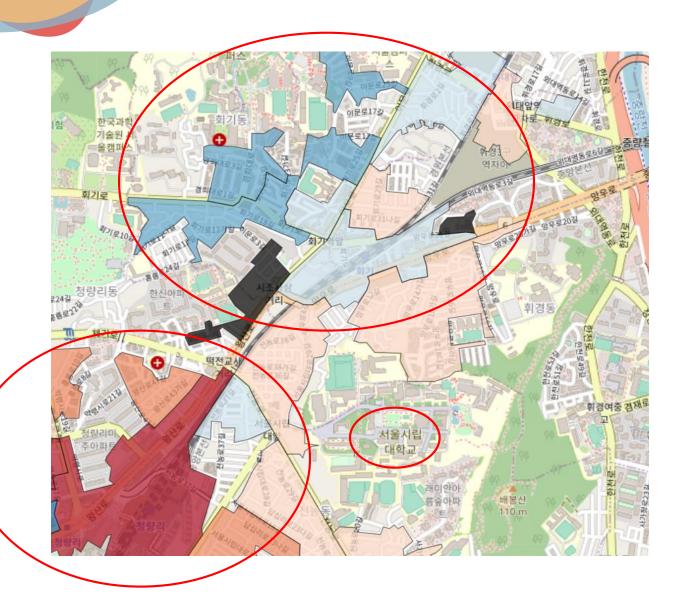


✔ 19년 1분기

✔ 기준: RANK

✓ 파란 계열일 수록 과밀





✔ 19년 1분기

✔ 기준: RANK

✓ 파란 계열일 수록 과밀

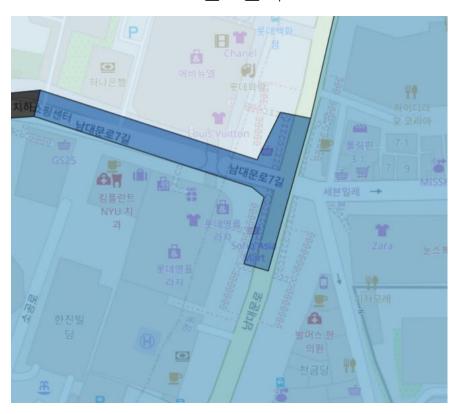
Model -시각화

19년 1분기 과밀 Rank 1등 상권은

현재(22년 3분기) 어떨까…?



19년 1분기



코로나

직격?

22년 3분기





하반기는 다를까… 회복 더딘 명동 상권

기사승인 2022-09-14 06:00:08

f y 7 P N b - + 8



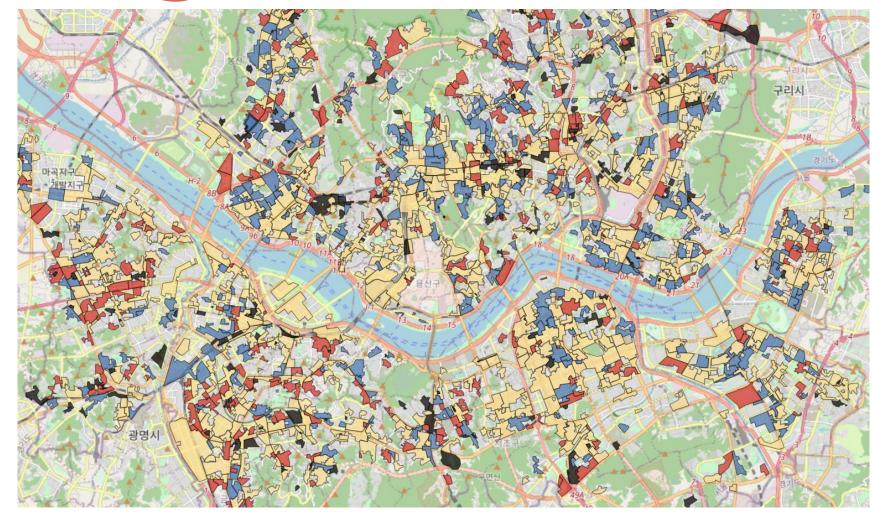


초여름부터 코로나바이러스감염증-19(코로나19) <u>엔데</u>믹 분위기는 명동거리를 <u>빗겨간 모양</u>새다. 거리를 메운 관광객과 대조적으로 텅 빈 상가들이 여전히 자리를 지키면서 반쪽 회복에 그쳤기 때문이다.

13일 오후 2시 서울 중구 명동. 거리에는 내외국인 관광객들의 발걸음이 뒤섞여 다소 활기찬 분위기가 이어졌다. "투 (two), 2000원" 거리에선 넷플릭스 시리즈 '오징어게임'에 등장한 달고나를 구매하는 외국인들이 종종 눈에 띄었다. 화 장품 가게 직원들은 문 앞으로 나와 손님 모시기에 뛰어들었다. 새로 문을 열기 위해 공사를 진행하는 가게도 있었다.

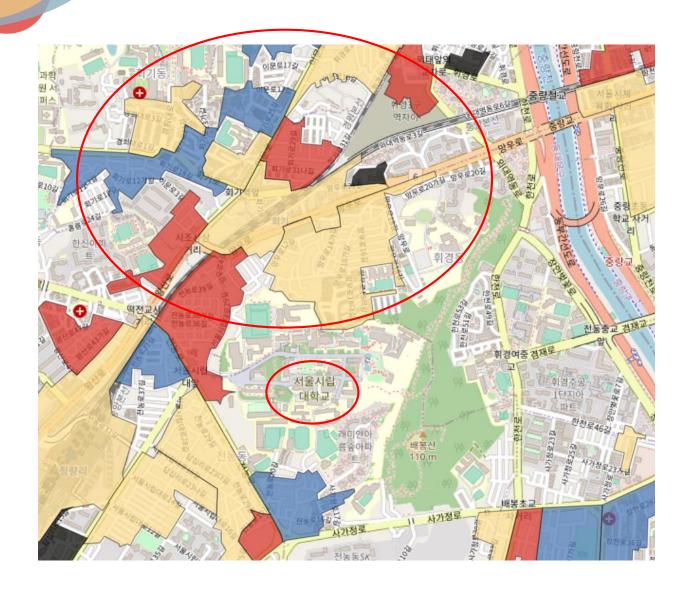
활기찬 분위기와 대조적으로 골목을 채운 상가들은 여전히 텅빈 상태였다. 문을 연 일부 가게들을 제외하고 '영업종료', <u>'임대문의'</u> 딱지가 붙은 상가들이 늘어서 있었다. 메인 거리를 지나 골목으로 진입할수록 빈 상가들을 목격하는 빈도수 가 늘었다. 한 국내 스포츠 브랜드 상가는 '올해 말까지 휴업'이라는 표시를 내걸어두기도 했다.





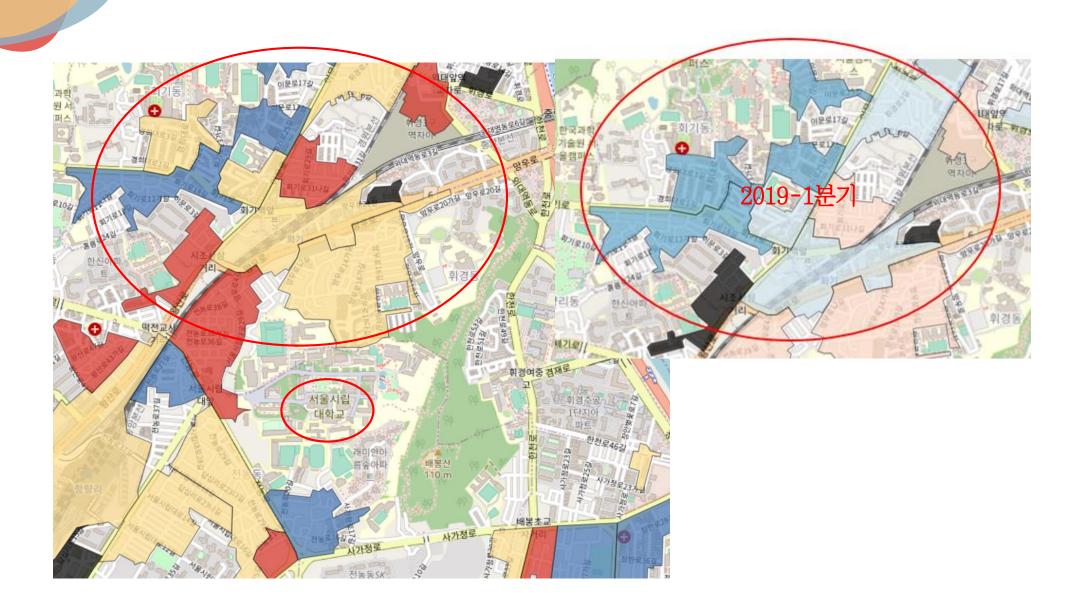
- ✓ 2021년 2분기
- ✓ 기준: 여유 적정 과밀
- ✓ 3중 분류
- ✓(단색)





- ✓ 2021년 2분기
- ✓ 기준: 여유 적정 과밀
- ✓ 청량리~ 회귀
- ✓(단색)

Model -시각화





[대핫상권 르포] ③경희대 상권, 구매력 떨어뜨리는 'NO매력'

경기침체 가운데 상권 변화는 다이내믹

္ 박민규 기자

① 입력 2020.01.22 11:02

Model -시계열 (연구2)

● (연구2) 상권에서 카페의 수는 가까운 미래에 몇 개가 될까?

Model - 시계열 (연구2)

✔['상권코드', '연-분기']

✓ 미래예측 → 미래의_X ?

✔ 비유동적 features → '상권구분코드', '영역 면적'

Model -시계열 (연구2)

✓ 매우 짧은 Time Series (15)

✓ LSTM (Long Short-Term Memory)

✔ Cross-Validation 통한 성능 향상에 집중

Model -시계열 (연구2)

✔ 2022 - 2023 시각화 비교

✓ 2023예측 - (2022+2021).mean 시각화

✔ 성장-하락 지역 예측

