

2022년 하계

경진대회

홈쇼

5639190 통계학전공 최지우

5515842 경영정보학과 김웅기

목차

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점

01

권리금이란?

01 권리금이란?

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점

권리금이란?

권리금이란?

- 임대차 목적물인 상가건물에서 영업을 하는 사람 또는 영업을 하려는 사람이 영업시설, 비품, 거래처, 신용, 영업상의 노하우, 상가건물의 위치에 따른 영업상의 이점 등 유형·무형의 재산적 가치의 양도 또는 이용대가로서 임대인, 임차인에게 보증금과 차임 이외에 지급하는 금전 등의 대가를 의미한다.



01 권리금이란?

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점

권리금의 종류

1. 바닥 권리금
점포의 위치 자체의 이점으로 발생하는 권리금.
2. 영업 권리금
기존의 거래처·고객의인수, 영업 노하우의 이전, 기존영업의 신용 등으로 발생하는 권리금.
3. 시설 권리금
기존 상가 점포에 설치되어 있는 인테리어나 비품을 인수하면서 발생하는 권리금



02

데이터 수집 및 탐색

- 01. 분석 목적
- 02. 데이터 수집
- 03. 데이터 탐색

02 분석 목적

분석 목적

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점

1. 상점 매매에 코로나가 미친 영향

2. 권리금 결정 요인 분석

02 데이터 수집

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점

데이터 수집

1. 공공데이터포털(www.data.go.kr)
 - 전국 초중고등학교 위치 표준 데이터
 - 도시철도 역사정보 표준 데이터
 - 연도별, 월별 소비자물가 데이터
2. 코레일(<https://info.korail.com>)
 - 도시철도별 승하차인원 데이터
3. 중앙 방역 대책본부(www.ncov.mohw.go.kr)
 - 코로나 바이러스 감염증-19 (6.28일 기준)

02 데이터 탐색

데이터 탐색

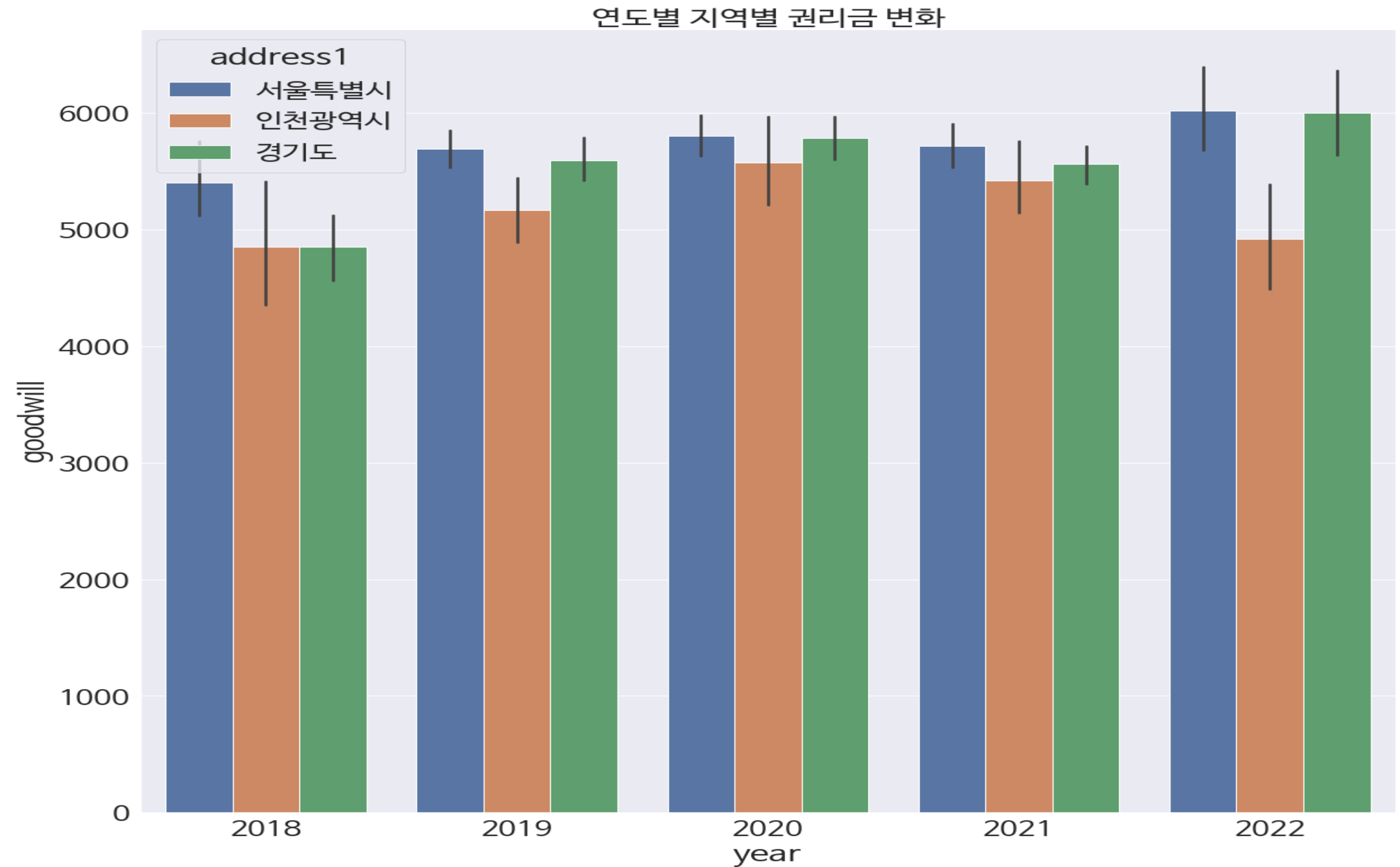
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점



02 데이터 탐색

데이터 탐색

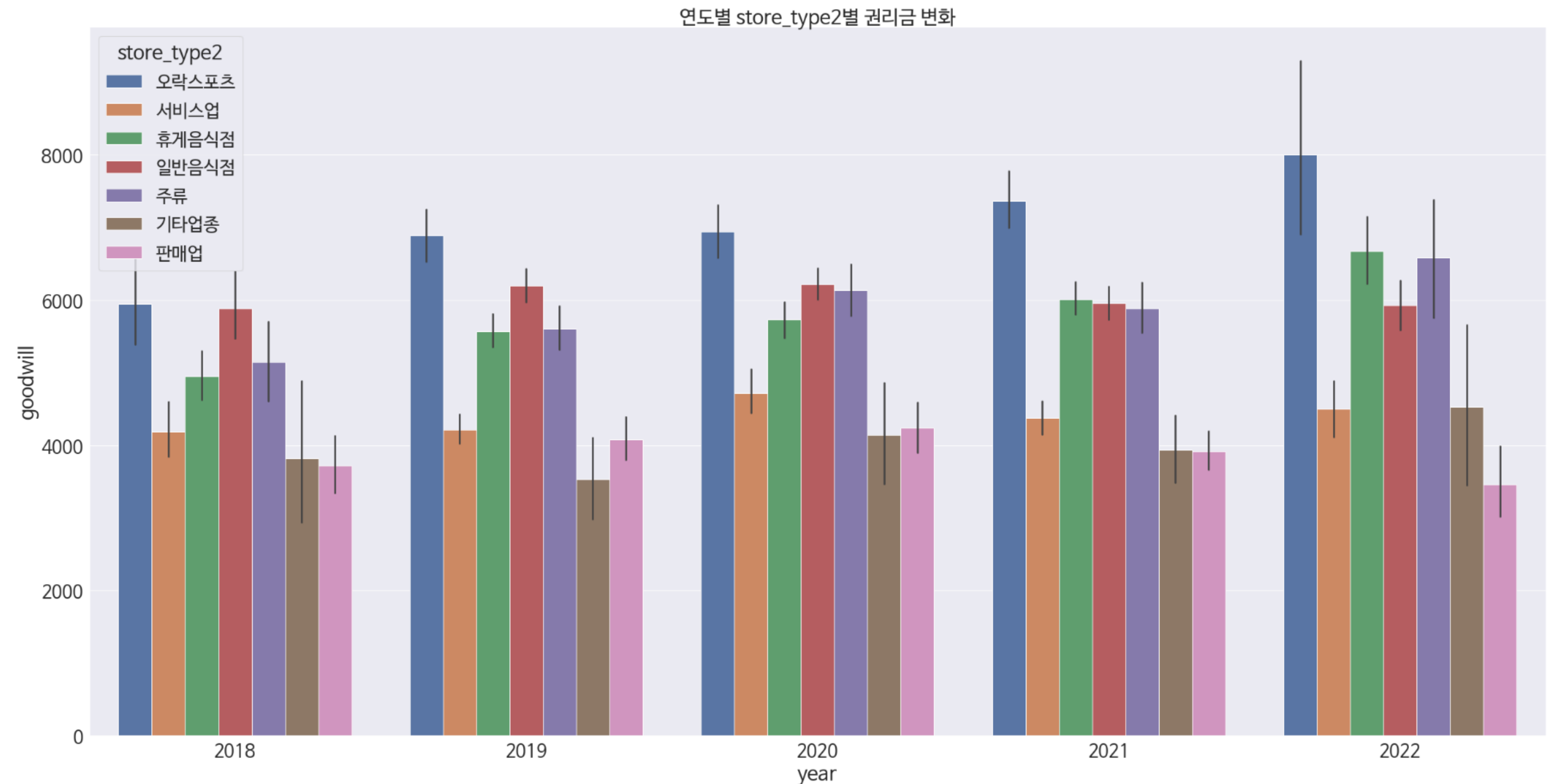
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

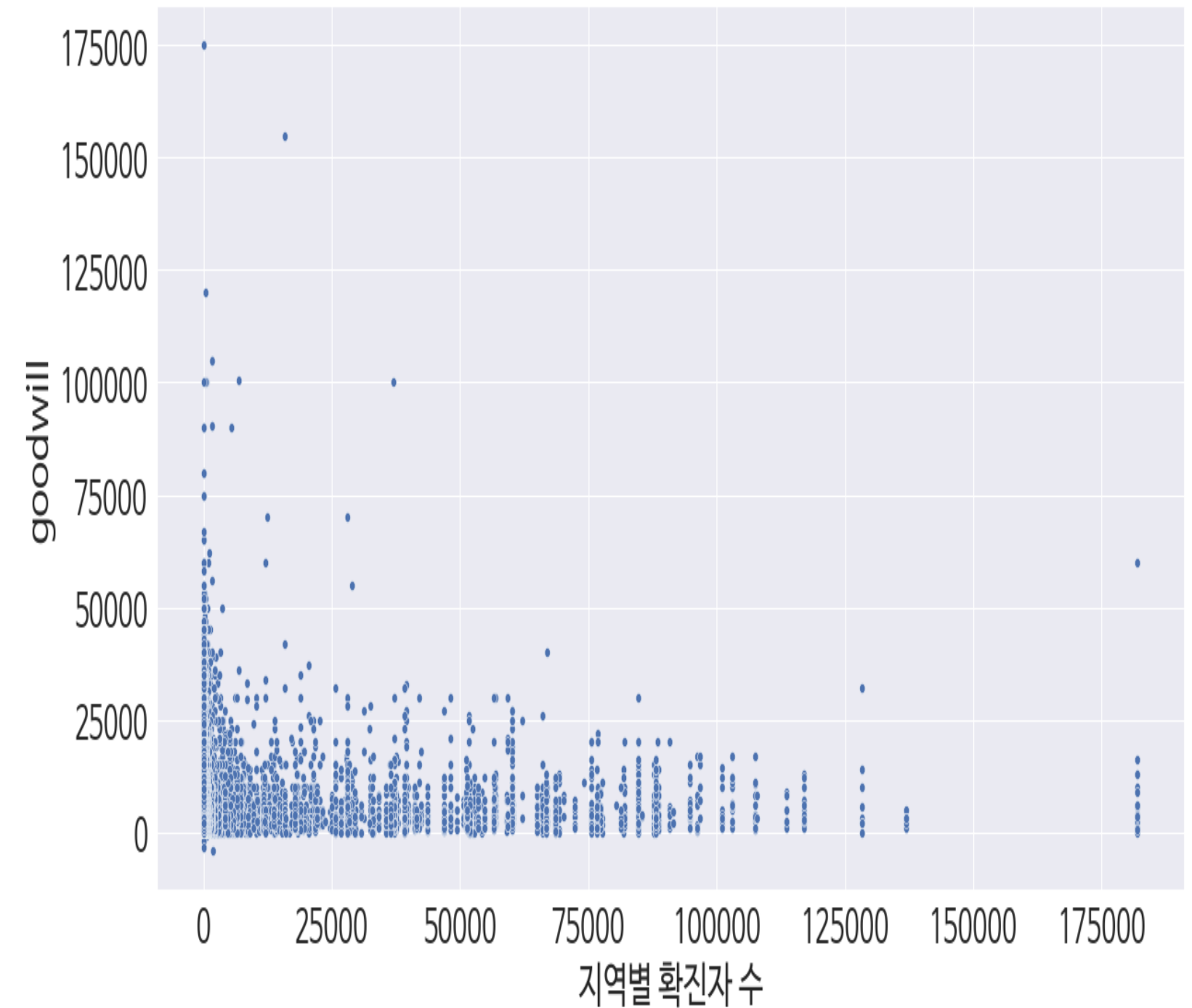
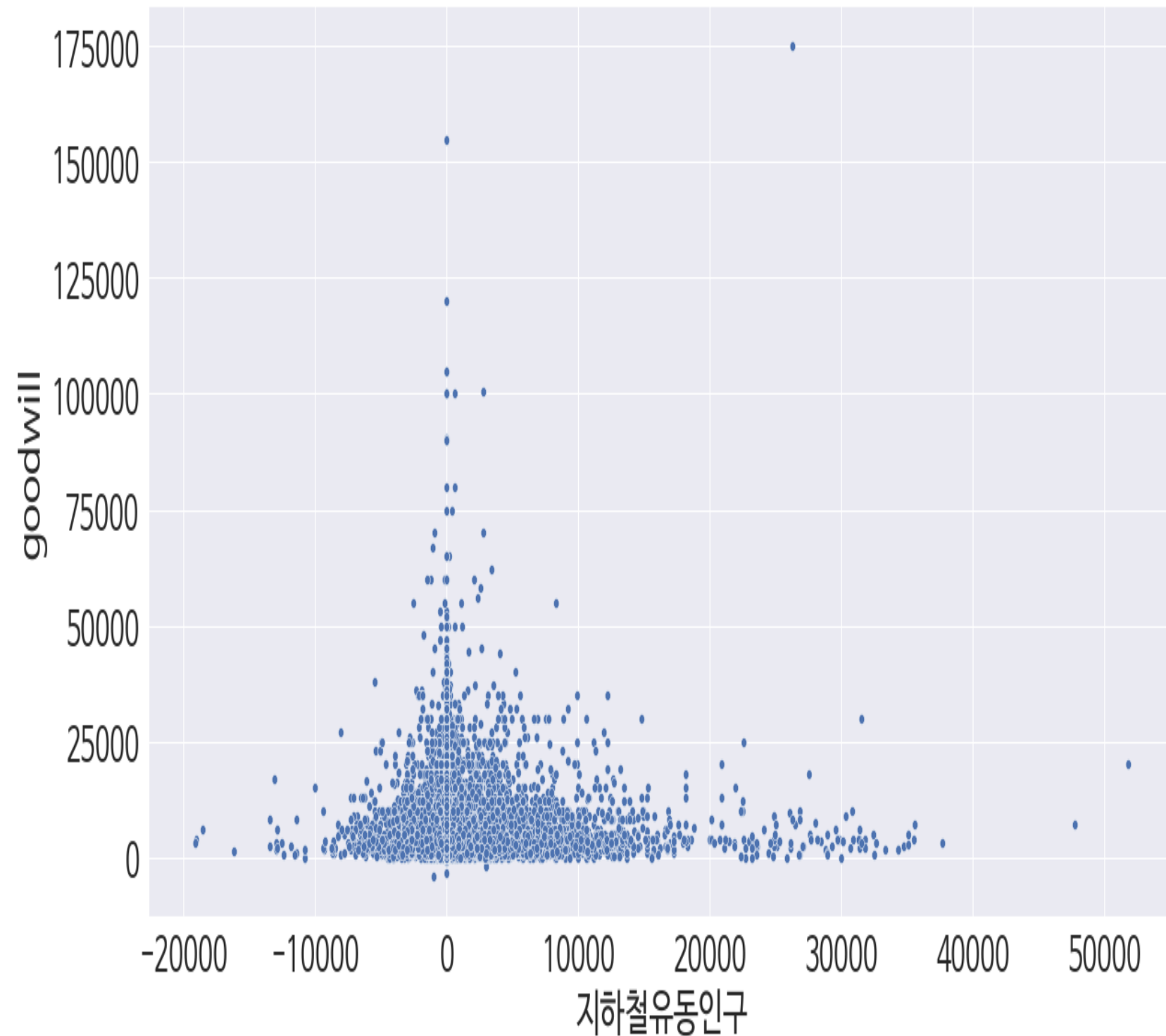
04 모델

05 결과 및 시사점



02 데이터 탐색

데이터 탐색



03

전처리

- 01. 변수 생성
- 02. 이상치 확인
- 03. 스케일링
- 04. 상관계수 확인
- 05. 다중 공선성 확인

03 변수 생성

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

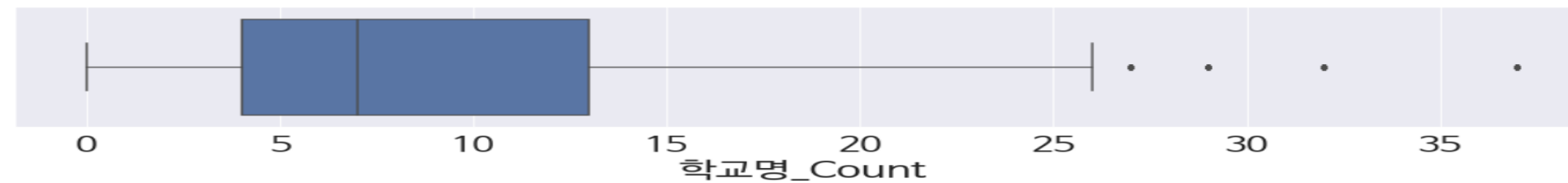
05 결과 및 시사점

변수 생성

- 각 거래별 점포의 순수 매매 금액을 파악하기 위해 sale_price칼럼에서 권리금을 의미하는 goodwill을 뺀 후 price 칼럼으로 저장
- 면적별 권리금을 파악하기 위해 권리금을 의미하는 goodwill을 면적을 의미하는 area로 나누어 rate 칼럼으로 저장
- store_type을 대분류별로 분류하기 위해 데이터 제공 사이트 정보를 바탕으로 대분류 후 store_type2 칼럼으로 저장
- 연도, 월, 일별 권리금 추세 파악과 면적당 권리금, 층별 권리금 정보를 확인하기 위해 contract_date 칼럼에서 year, month, day 칼럼으로 나누어 저장
- 각 시도별, 시군구별, 동별 권리금, 면적당 권리금, 층별 권리금 파악을 위해 address 데이터를 시도별로 나눈 후 address1, 2, 3으로 저장
- 각 지역별 승하차 인원의 차이를 알아보기 위해 하차인원 - 승차인원의 값을 지하철 유동인구 칼럼으로 저장

03 이상치 확인

이상치 확인



- 변수마다 이상치가 존재하지만 실제 데이터이기 때문에 제거하지 않기로 함
- 변수마다 단위가 다르기 때문에 스케일링이 필요함

03 스케일링

스케일링

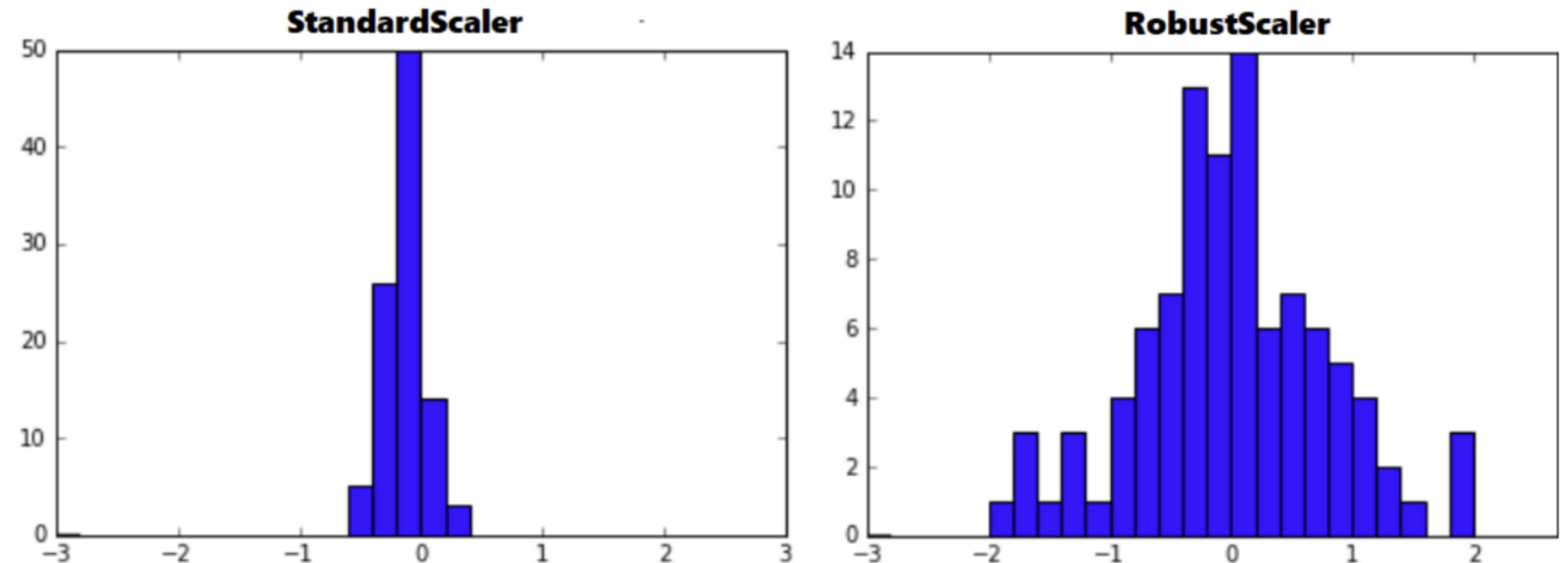
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점



아웃라이어의 영향을 최소화할 수 있는 Scale기법

Robust Scaler 사용

03 상관계수 확인

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점

상관계수 확인



증감여부와 각 변수와의 선형적 관계 파악

03 다중공선성 확인

다중공선성 확인

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델

05 결과 및 시사점

	3	4	6	7	5	0	2	1
VIF Factor	1.462657	1.380336	1.318827	1.211837	1.102224	1.052771	1.03804	1.014684
features	소비자물가지수	지역별 확진자 수	역사명_Count	지하철유동인구	학교명_Count	level	rate	price

VIF (Variance Inflation Factor)의 값이 10 이상이면 일반적으로 다중 공선성 문제가 크다고 판단

04

모델링

- 01. LSTM
- 02. RandomForest
- 03. GradientBoost
- 04. LightGBM
- 05. XGBOOST

04 모델링

모델링

01 권리금이란?

02 데이터 수집 및 탐색

- 종속변수는 goodwill, 독립변수는 rate, price 등 8개이며 36919 row로 구성됨

03 전처리

04 모델링

- train과 test를 7:3 비율로 나눔

05 결과 및 시사점

04 LSTM

LSTM

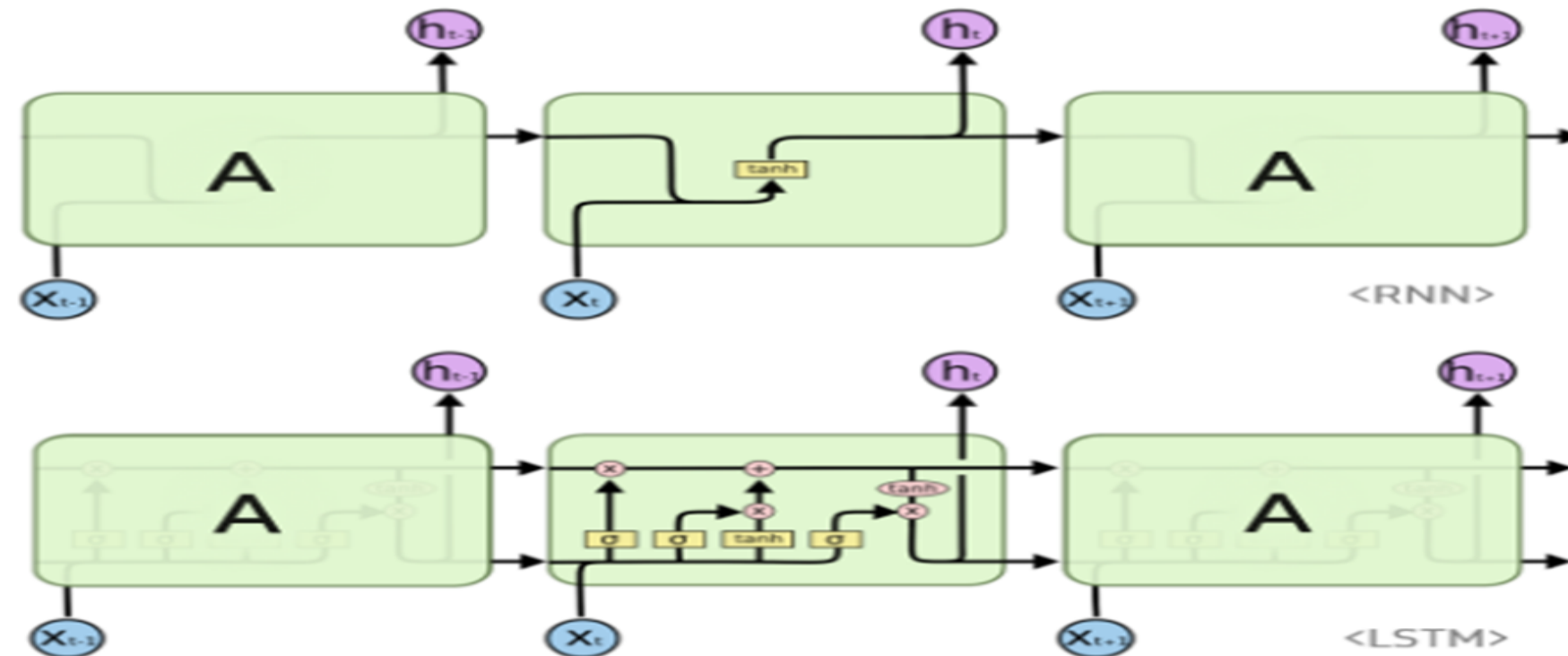
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점



- LSTM은 순차적으로 등장하는 데이터 처리에 적합한 모델인 RNN을 보완하여 만들어진 모델
- RNN의 vanishing gradient problem문제를 극복하기 위해 RNN의 hidden state에 cell-state를 추가한 형태이다.

04 LSTM

LSTM

- 모델 요약

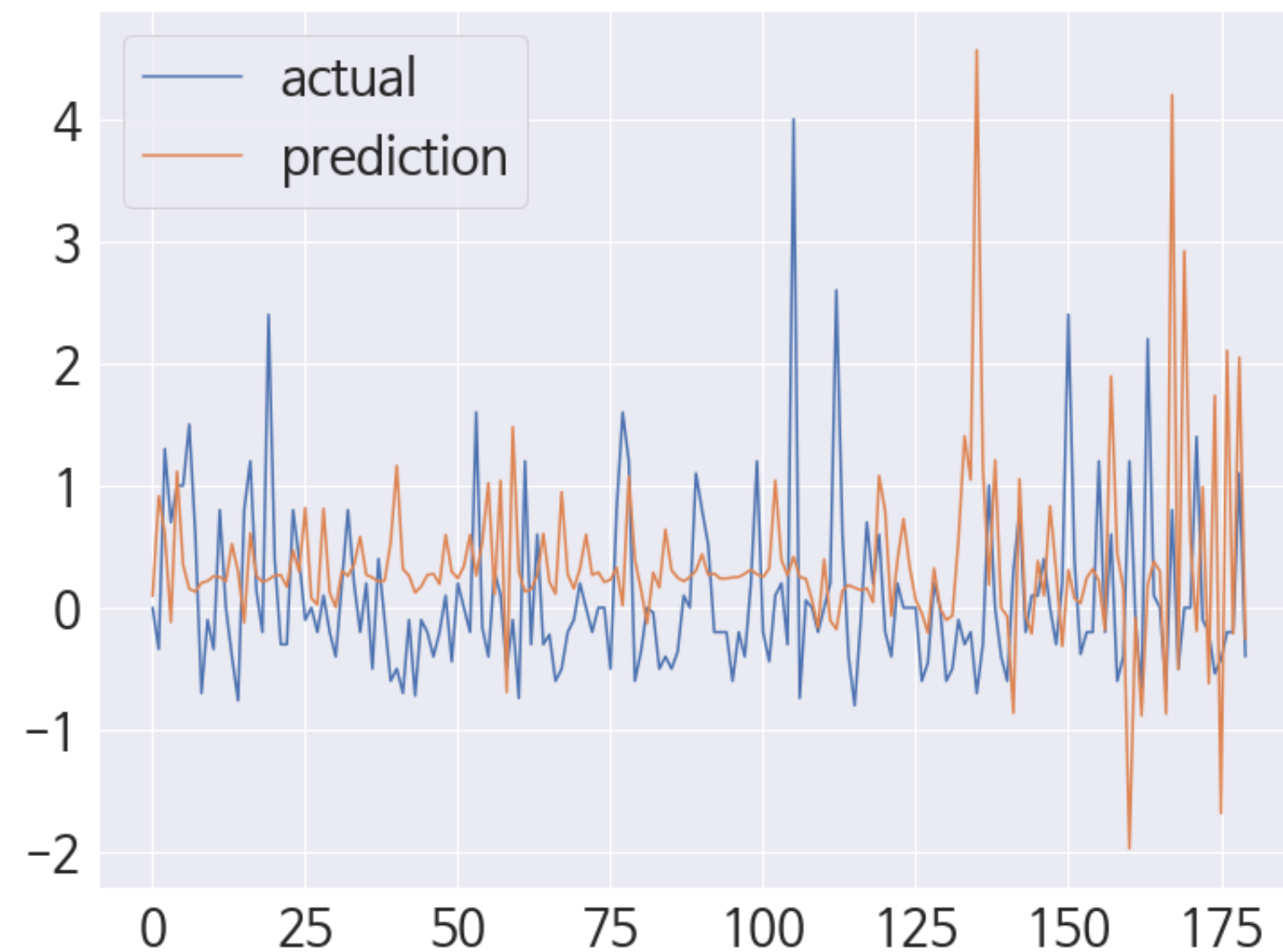
Model: "sequential_9"

Layer (type)	Output Shape	Param #
lstm_9 (LSTM)	(None, 20)	2320
dense_15 (Dense)	(None, 10)	210
dense_16 (Dense)	(None, 1)	11

Total params: 2,541

Trainable params: 2,541

Non-trainable params: 0



MAE : 0.7084, MSE : 1.015

04 RandomForest

RandomForest

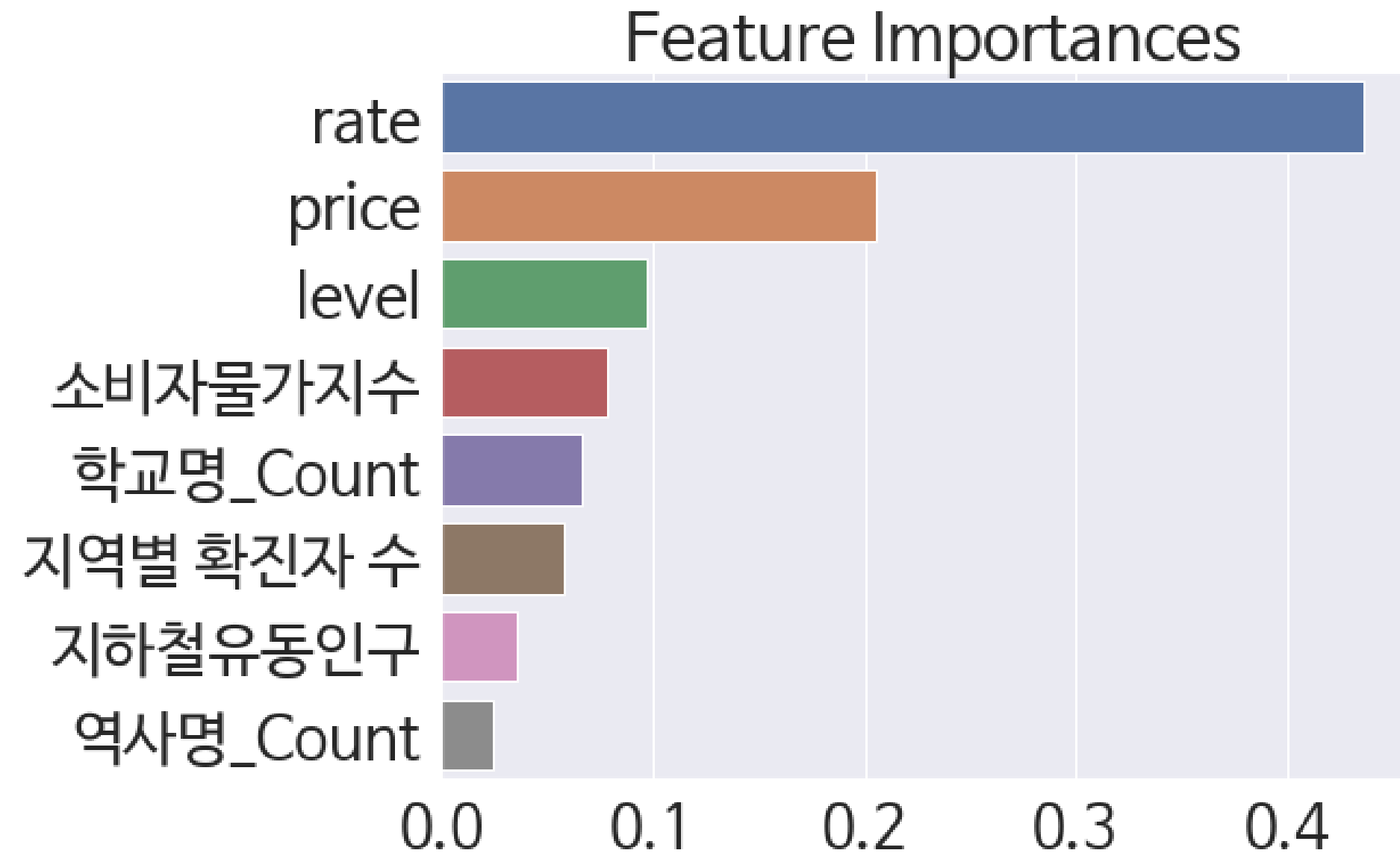
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점



train R^2 : 0.934

train mse: 0.092

test R^2 : 0.492

test mse : 0.840

04 GradientBoost

GradientBoost

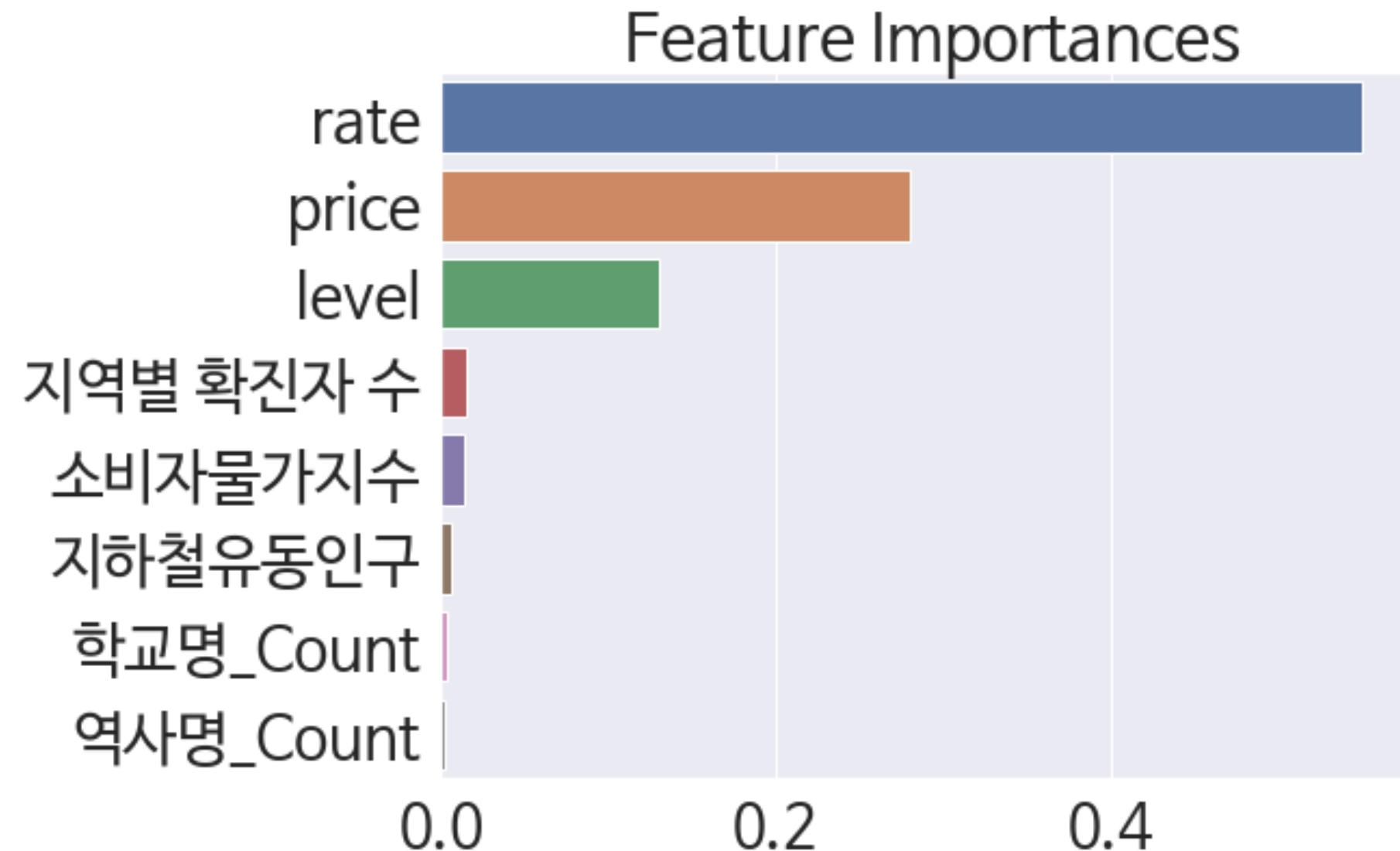
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점



train R^2 : 0.602

train mse: 0.557

test R^2 : 0.504

test mse : 0.820

04 LightGBM

LightGBM

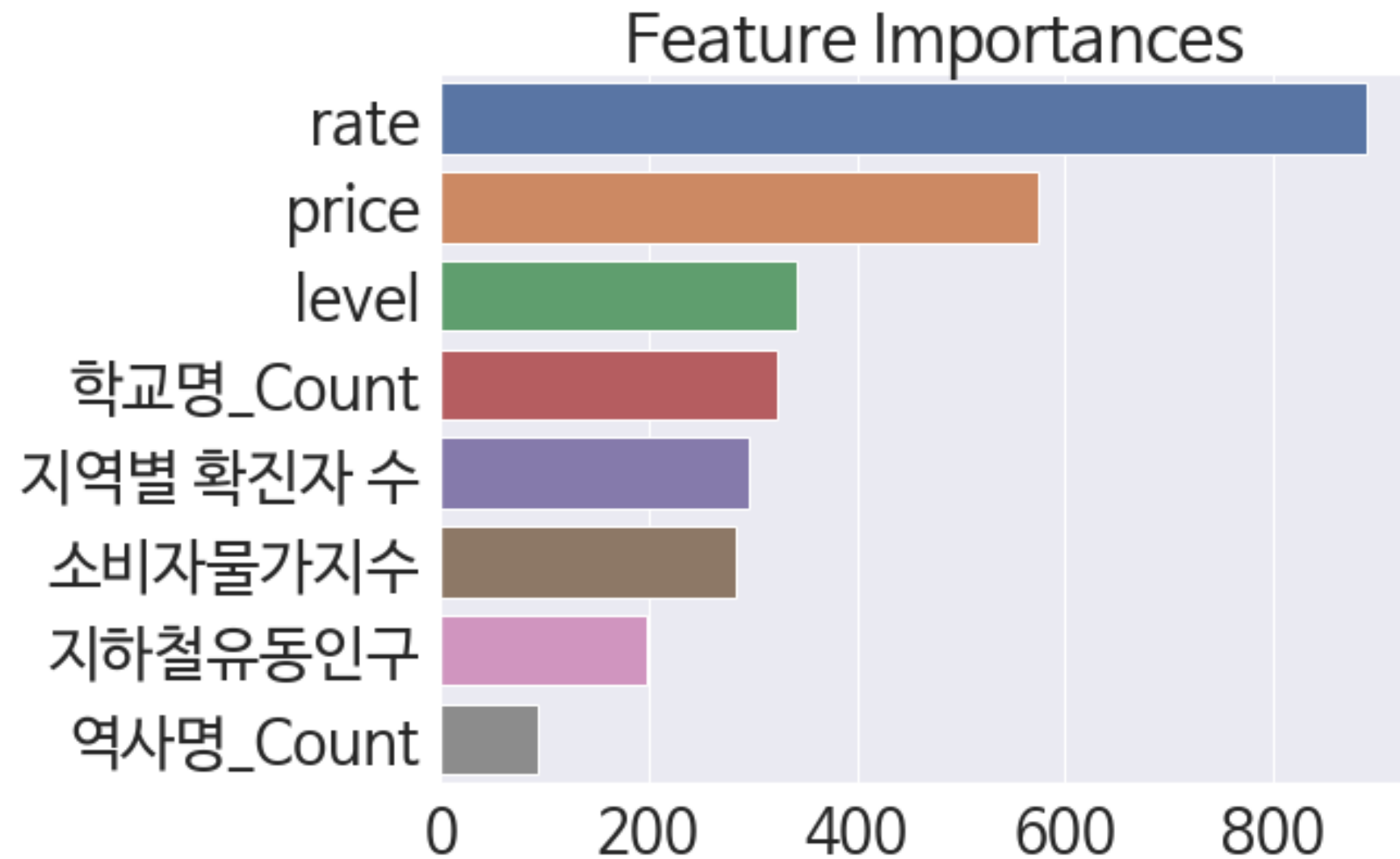
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점



train R^2 : 0.682

train mse: 0.473

test R^2 : 0.494

test mse : 0.837

04 XGBOOST

XGBOOST

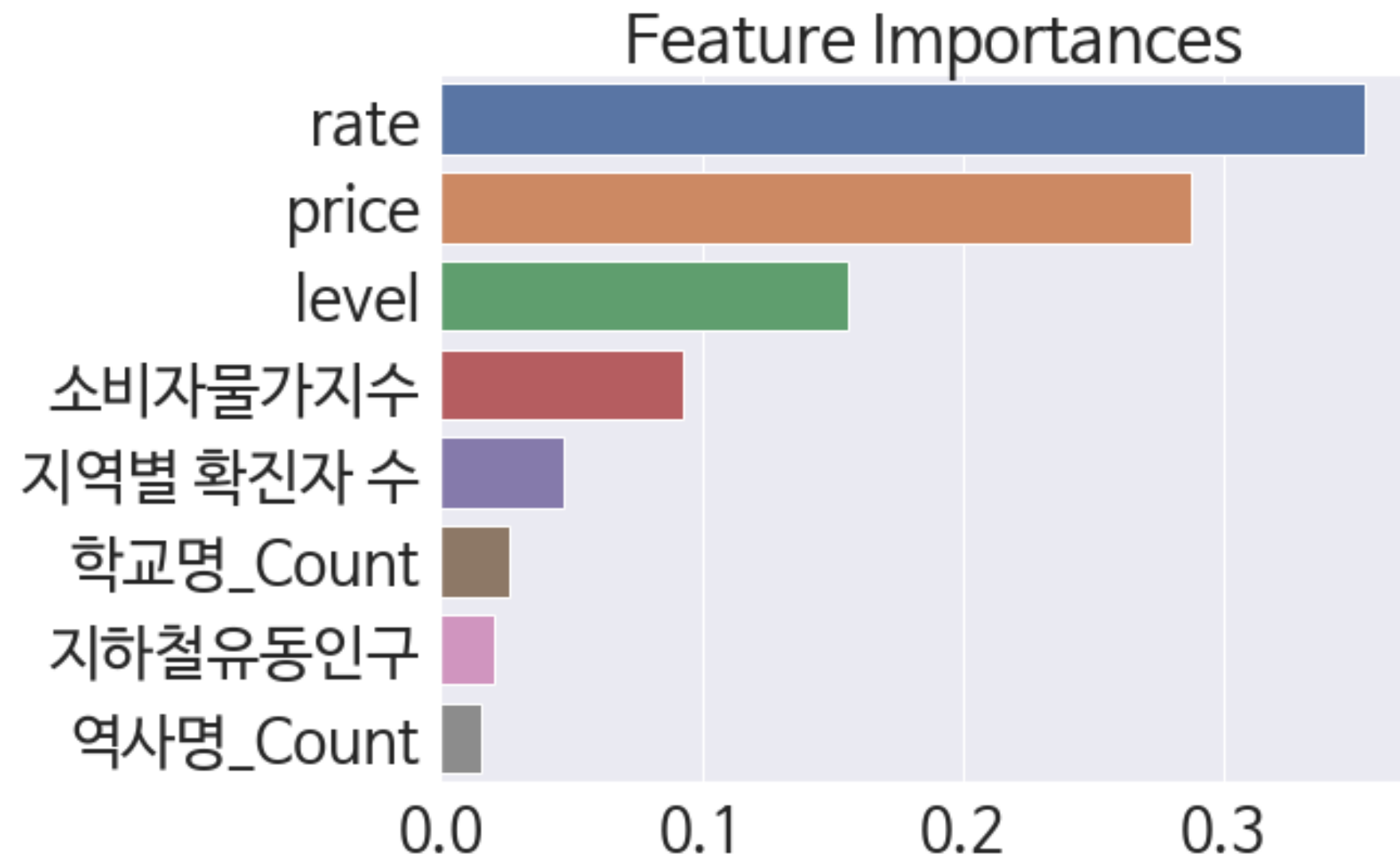
01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점



train R^2 : 0.593

train mse: 0.570

test R^2 : 0.505

test mse : 0.818

04 모델링

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점

모델링 변수 중요도 결과

	RF	GBT	LGBM	XGBOOST	합계
Price	2	2	2	2	8
Rate	1	1	1	1	4
Level	3	3	3	3	12
소비자물가지수	4	5	6	4	19
학교명_Count	5	7	4	6	22
역사명_Count	8	8	8	8	32
유동인구	7	6	7	7	27
지역별 확진자 수	6	4	5	5	20

05

결과 및 시사점

- 01. 분석 결과
- 02. 시사점

05 분석 결과

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점

분석 결과

모델링 결과를 바탕으로 제공된 원본 데이터에서 파생되거나 원본 데이터에 존재하는 변수들이 권리금에 대체적으로 많은 영향을 주는 변수라고 할 수 있다.

rate 변수는 면적에 따른 권리금을 의미하는데, 권리금을 결정하는 데에 면적이 가장 큰 영향을 줬다고 해석할 수 있다. 그 다음으로는 총 가격에서 권리금을 뺀 price 변수인 것을 확인할 수 있다.

05 시사점

01 권리금이란?

02 데이터 수집 및 탐색

03 전처리

04 모델링

05 결과 및 시사점

시사점

제공된 데이터에서 정확한 주소를 알 수 있다면 유동인구 파악과 더불어 더 정확한 데이터 분석이 가능했을 것이며, 쓸 수 있는 변수도 다양했을 것이다.

공공데이터를 포함한 대부분의 데이터가 최신화가 잘 이루어지지 않아 분석이 가능한 범위가 제한적이었다. 만약 데이터의 최신화가 잘 이루어진 데이터를 구할 수 있다면 좋은 분석이 가능할 것이다.

제공된 데이터의 주소와 수집한 데이터의 주소가 맞지 않아 전처리를 하는데에 많은 어려움이 있었다. 도로명주소와 지번주소가 같이 있었다면 지역별 분석을 하는데에 있어 많은 도움이 될 것이다.

감사합니다
