

# House Price Prediction

3조 별자  
이승현 박언선 한민규 이광우 정혜윤

# Content

---

- 01. 팀원 소개
- 02. 대회 소개
- 03. Data Description
- 04. Modeling
- 05. 결과
- 06. 경진대회 진행 소감

01

# 팀원 소개

## 이승현 : 봄날의 이승현



### Interested in

- ML Ops

### Introduction

- 컴퓨터공학
- 하루하루 버텨보자!

### Role

- 조장
- 자존감 지킴이!

### In Upstage AI Lab

- 부족한 프로젝트 경험을 쌓고 싶습니다.

# 박언선 : 한잔은 이미 초라해진 나를 위하여



### Interested in

- ML,DL
- MLOps

### Role

- 팀원

### Introduction

산업공학과

Pose detection app(CV,Backend),  
Crime detection(CV),  
42 school P/F 예측(ML)

### In Upstage AI Lab

확실한 진로방향 선택

지식의 깊이

# 한민규 : 조용하고 강한 한민규



### Interested in

- LLM
- Prompt Engineering
- MLOps

### Introduction

- 전기전자공학부

### Role

- 팀원

### In Upstage AI Lab

- AI 분야를 폭 넓게 이해하기
- 많은 프로젝트를 도전해보기

# 이광우 : 해야지 하지 말고 하자



### Interested in

- LLM
- Prompt Engineering
- MLOps

### Introduction

- 신소재공학

### Role

- 팀원

### In Upstage AI Lab

- 혼자서라도 케글에 도전할 수 있는 상태가 되길

# 정혜윤 : 노력의 정혜윤



### Interested in

- NLP
- Human Interaction with AI
- 생성모델의 비윤리적 활용에 대한 방어

### Introduction

- 건축학전공 (학사, 석사)
- AURI '삶의 질 제고를 위한 생활SOC 운영효율화 방안' 연구

### Role

- 팀원

### In Upstage AI Lab

- 공공의 이익을 실현할 수 있는 능력을 키우기
- 꿈 실현을 위해 꼭 필요한 좋은 사람들 만나기!



02

## 대회 소개

# 대회 개요

---

목 적 : 서울시 아파트 실거래가 예측 모델 (regression)

평가지표 : RMSE

대회기간 : 2024. 3. 20. - 2024. 4. 2.

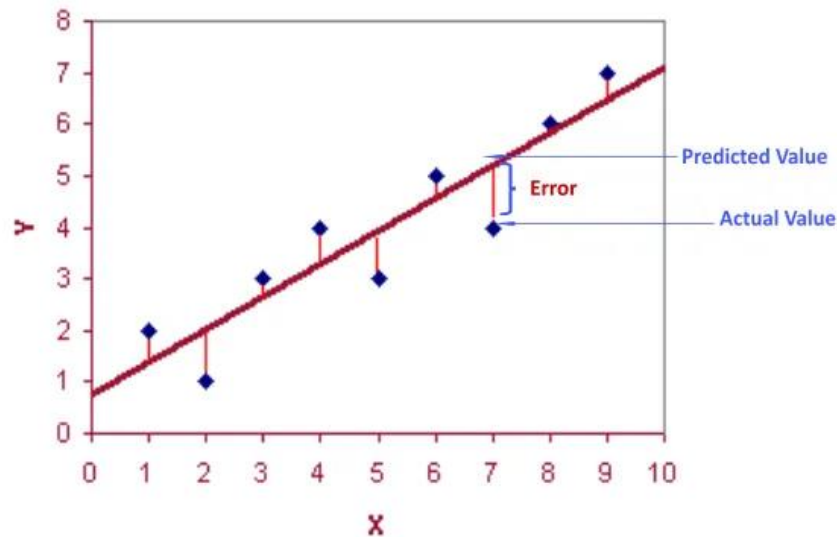
데이터셋 : 국토교통부 실거래가 & 서울시 공공주택 아파트 정보

## RMSE (Root Mean Square Error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2}$$

Where:

- $\Sigma$  is the summation of all values
- $f$  is the predicted value
- $o$  is observed or actual value
- $(f_i - o_i)^2$  are the differences between predicted and observed values and squared
- $N$  is the total sample size



03

# Data Description

## 데이터 설명

---

서울시 아파트의 실거래가에 대한 정보를 포함한 데이터셋으로,

시군구, 아파트명, 전용면적, 계약년월, 계약일, 층, 건축년도, 도로명주소 등의 정보를 포함

**Train data : (1188822, 52)**

**기간 : 2007-01-01 ~ 2023-06-30**

**Test data : (9272, 51)**

**기간 : 2023-07-01 ~ 2023-09-26**

## 데이터의 기초 통계 및 정보 요약

### Train data

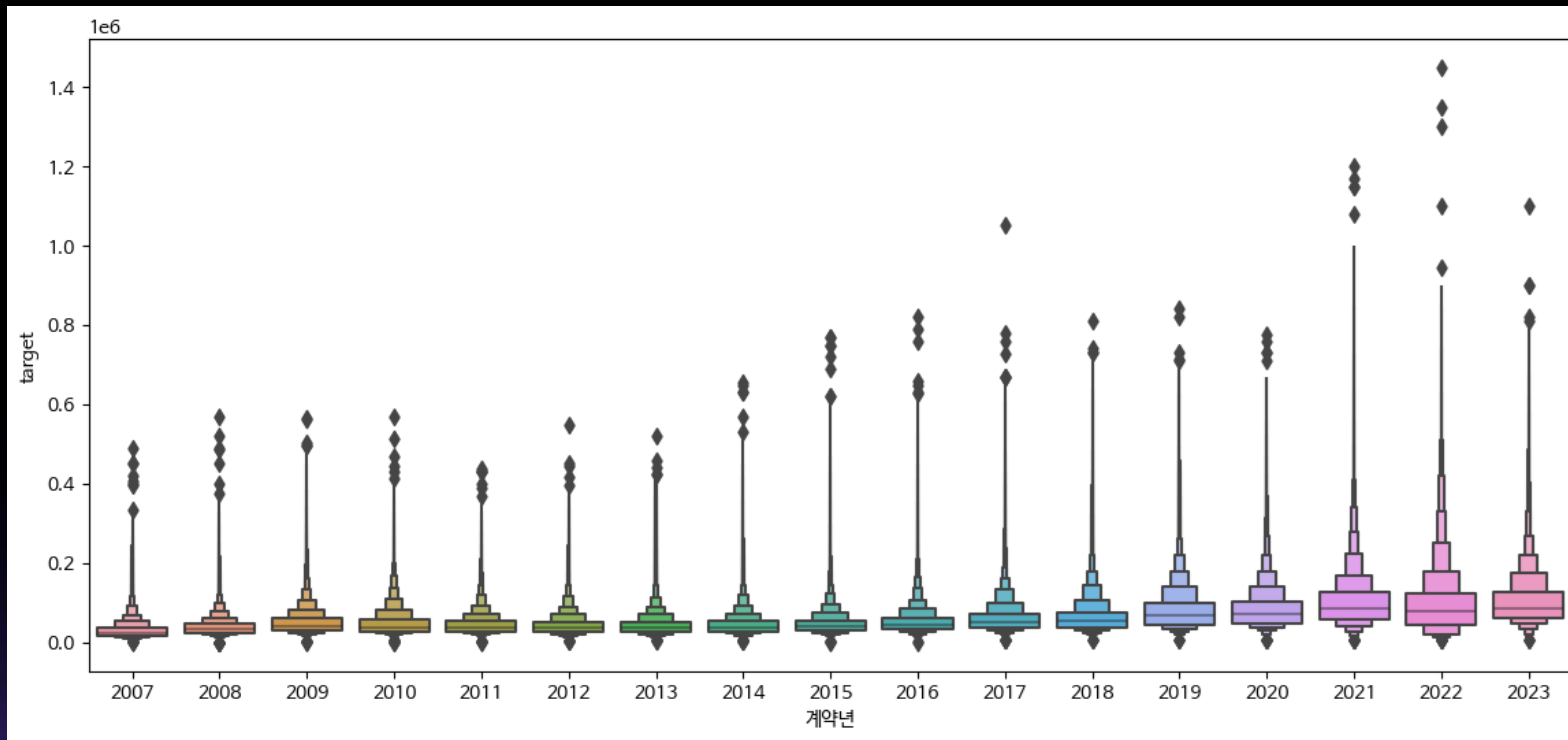
	전용면적(m <sup>2</sup> )	계약년월	층	건축년도	target
mean	77.17	-	8.87	-	57991.53
std	29.36	-	5.98	-	46426.02
min	10.02	2007년 1월	-4	1961	350
25%	59.65	2011년 10월	4	1992	30500
50%	81.88	2015년 7월	8	2000	44800
75%	84.96	2018년 4월	12	2005	69800
max	424.32	2023년 6월	69	2023	1450000

## 데이터의 기초 통계 및 정보 요약

Test data

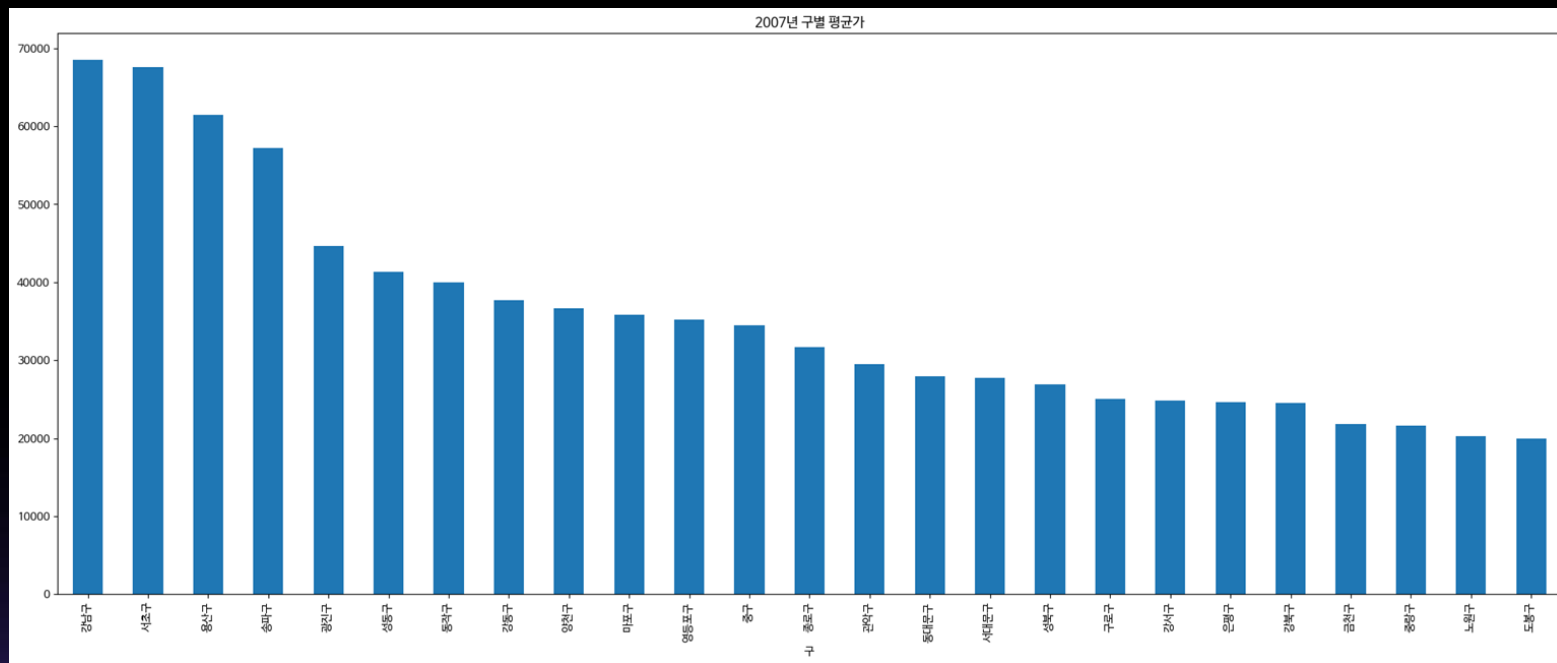
	전용면적(m <sup>2</sup> )	계약년월	층	건축년도
mean	75.41	-	10.02	-
std	29.34	-	6.65	-
min	12.00	2032년 7월	-3	1961
25%	59.70	2023년 7월	5	1996
50%	76.57	2023년 8월	9	2003
75%	84.96	2023년 8월	14	2012
max	301.47	2023년 9월	65	2023

## 계약년도에 따른 실거래가 분포

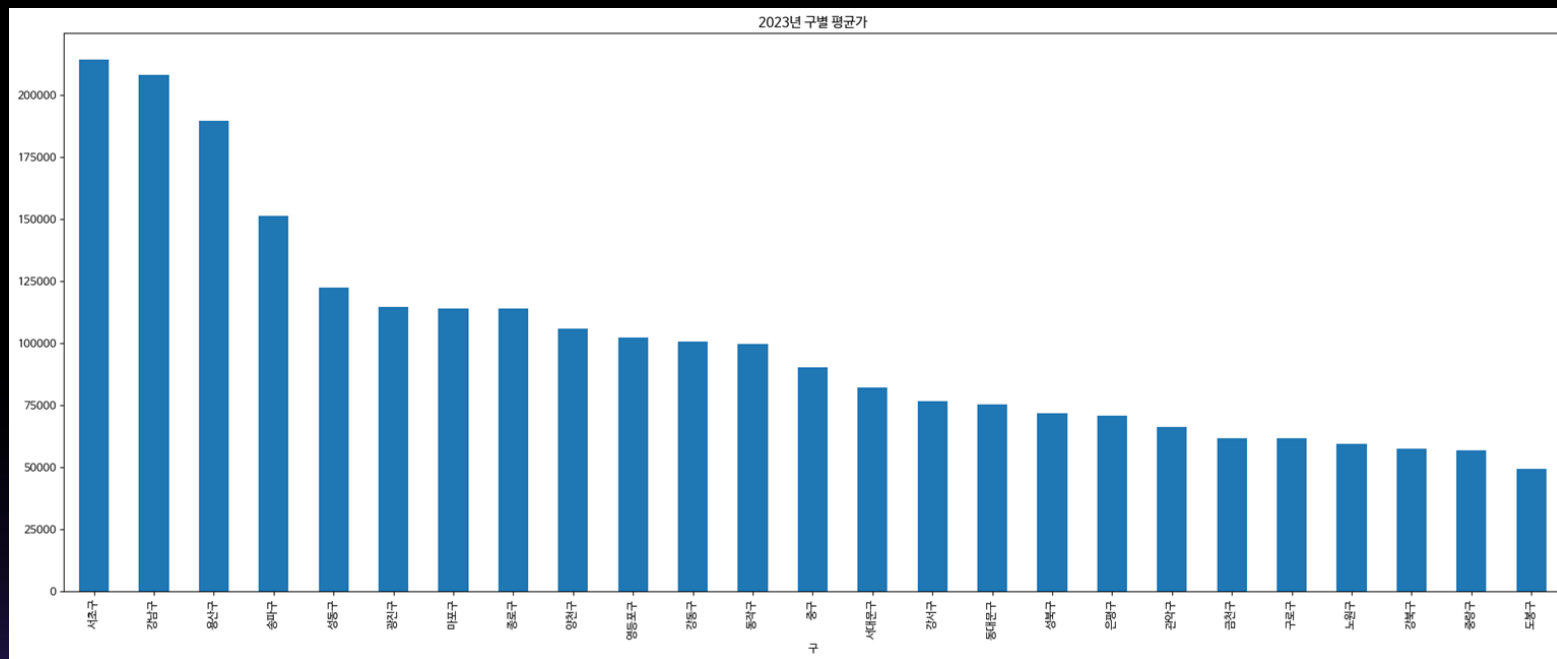




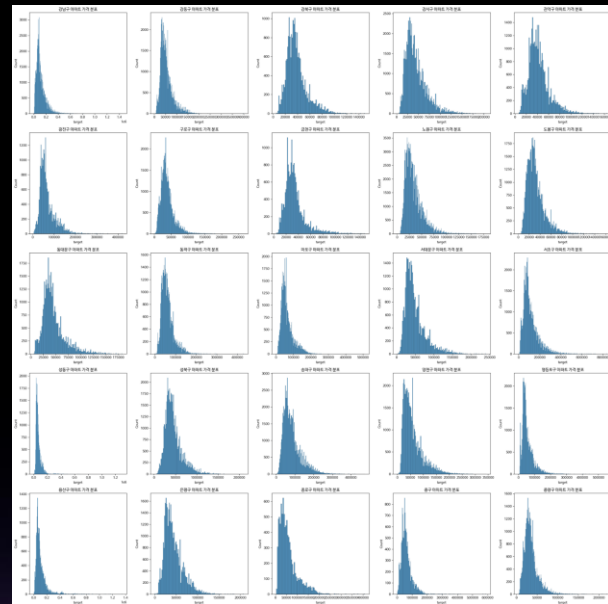
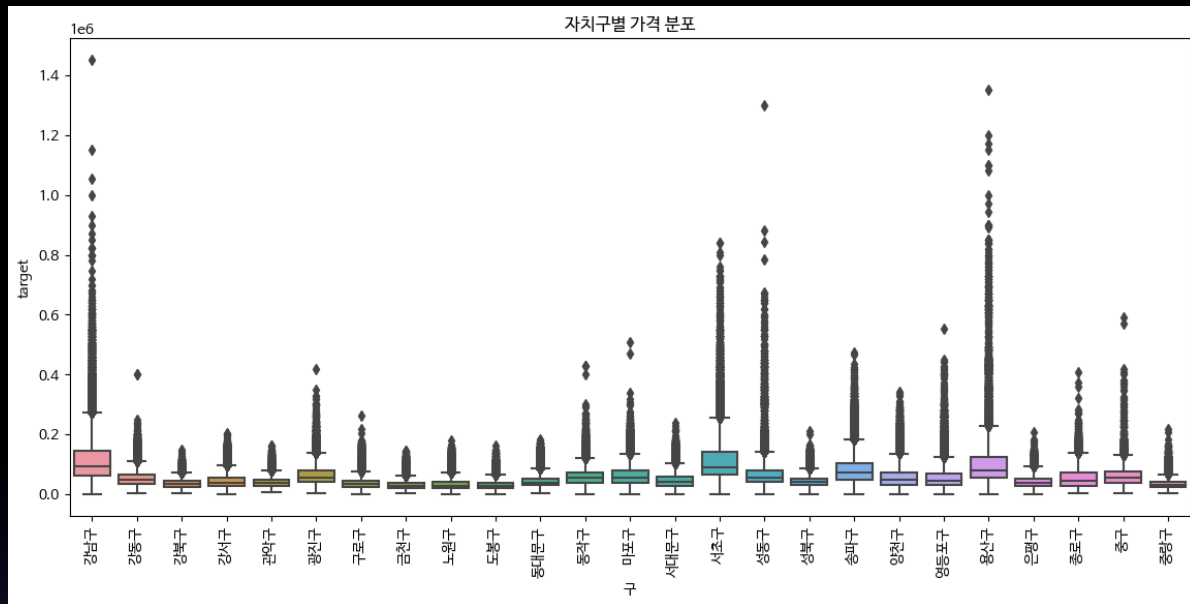
## 2007년에 대한 구별 실거래가 평균가



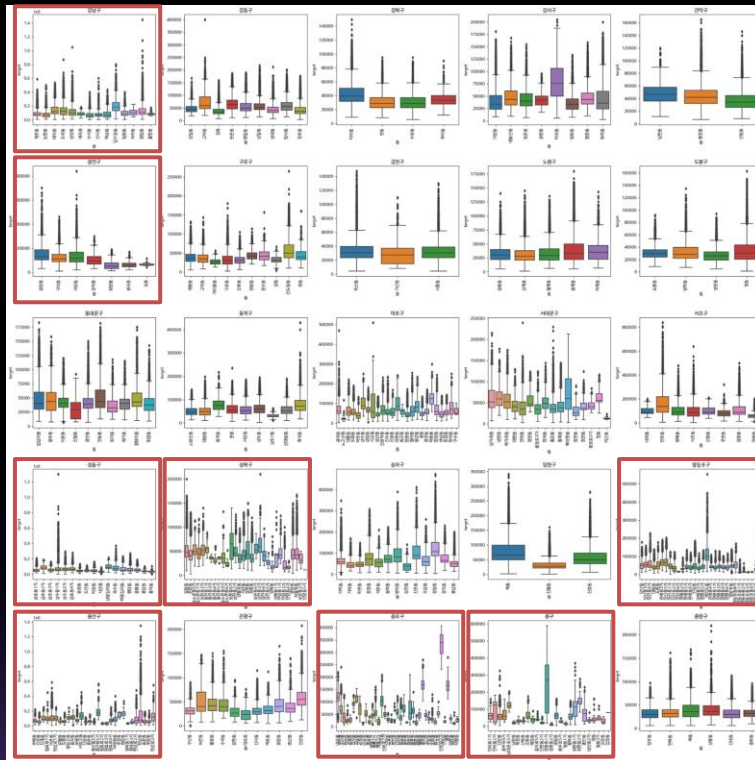
## 2023년에 대한 구별 실거래가 평균가



## 구별 실거래가분포



## 각 구의 동별 실거래가분포



변수별 일일치 비율

변수명	일일치 비율
보타	0.00
금리	0.00
년전	0.00
아파트용	0.75
관리비 상승도	0.75
사용하기가부	0.75
기타/의무 임대/임대=1/2인자	0.75
k-연관력	0.75
k-건설세	0.75
k-건설대수	0.75
k-노동생산	0.75
k-관리비 부과액	0.75
k-세터포인트 (문경영)	0.75
k-관리업식	0.75
k-수정값	0.75
k-주거건물면적	0.75
k-지역면적(세대당)(60㎡~95㎡이하)	0.75
k-전용면적(세대당)(95㎡~135㎡이하)	0.75
단위상충율	0.75
크로Y	0.75
크로X	0.75
k-사용권사명_사용승인일	0.75
주차대수	0.75
건축면적	0.75
k-복도유형	0.75
k-건해번호	0.75
단지승인일	0.75
k-건설대수	0.75
k-단지분류(아파트, 투숙형빌딩 등)	0.75
중개비관리원	0.75
k-건설사(시공사)	0.75
월소비관리원	0.75
k-시행사	0.75
k-팩스번호	0.75
세탁전기제어방법	0.75
고급포장관리번호	0.75
k-통과이지	0.90
단지소개기/Doc	0.90
k-음속강성	1.00
해설서유형	1.00
k-135㎡ 초과	1.00

k-단지분류(아파트, 주상복합등등)	0
단지소개기준clob	0
k-시행사	0
k-주거전용면적	0
k-전용면적별세대현황(60m <sup>2</sup> 이하)	0
k-전용면적별세대현황(60m <sup>2</sup> ~85m <sup>2</sup> 이하)	0
k-85m <sup>2</sup> ~135m <sup>2</sup> 이하	0
k-135m <sup>2</sup> 초과	0
k-홈페이지	0
k-등록일자	0
k-수정일자	0
dtype: int64	

k-세대타입(분양형태)	0
k-관리방식	0
k-난방방식	0
k-전체세대수	0
k-연면적	0
k-주거전용면적	0
k-관리비부과면적	0
k-전용면적별세대현황(60㎡이하)	0
k-전용면적별세대현황(60㎡~85㎡이하)	0
k-85㎡~135㎡이하	0
k-수정일자	0
dtype: int64	

# 데이터 전처리

## X, Y 좌표

- 서울시 공공주택 아파트 정보가 없는 경우 X, Y 좌표가 모두 결측
- 지하철과 버스정류장 정보를 활용하기 위해서는 X, Y 좌표를 채워야 함
- 네이버지도 API를 활용하여 결측 X, Y 좌표를 채움
- 아파트 X, Y좌표와 버스정류장 혹은 지하철까지의 거리가 일정 거리 이하인 경우 역세권으로 설정하여 역세권여부 변수를 추가



scipy.spatial.cKDTree

```
class scipy.spatial.cKDTree(data, leafsize=16, compact_nodes=True,  
copy_data=False, balanced_tree=True, boxsize=None)
```

## 그 외 불필요한 컬럼

- 해제사유발생일(결측 99.5%), 등기신청일자(결측 98.5%), 거래유형(결측 96.3%), 중개사 소재지(결측 96.6%) 등은 결측치도 너무 높고 실거래가에 영향을 미치지 않는 변수들로 판단되어 사용하지 않기로 함

- ```
df apt val
```

| 06_부속표 3 |      |      |              |                             |              |     |      |        |       |     |      |             |       |        |             |             |         |      |      |
|----------|------|------|--------------|-----------------------------|--------------|-----|------|--------|-------|-----|------|-------------|-------|--------|-------------|-------------|---------|------|------|
| 0.0s     |      |      |              |                             |              |     |      |        |       |     |      |             |       |        |             |             |         |      |      |
| 기준연도     | 기준점  | 법정공로 | 도로명주소        | 시도                          | 시군구          | 읍면  | 동리   | 특수지점코드 | 분면    | 본면  | 특수지역 | 단위저장        | 통행    | 요금     | 전통시장<br>별치자 | 공식가격        | 단위코트    | 통로   | 호로   |
| 0        | 2022 | 1    | 1.111010e+09 | 서울특별시 중랑구 지하<br>모로36길 16-14 | 서울특별시<br>중랑구 | NaN | 청운동  | 0.0    | 1.0   | 0.0 | NaN  | 청운역상<br>별치자 | 1     | 111    | 187,490     | 989000000.0 | 3.0     | 1.0  | 1.0  |
| 1        | 2022 | 1    | 1.111010e+09 | 서울특별시 중랑구 지하<br>모로36길 16-14 | 서울특별시<br>중랑구 | NaN | 청운동  | 0.0    | 1.0   | 0.0 | NaN  | 청운역상<br>별치자 | 1     | 112    | 187,490     | 989000000.0 | 3.0     | 1.0  | 2.0  |
| 2        | 2022 | 1    | 1.111010e+09 | 서울특별시 중랑구 지하<br>모로36길 16-14 | 서울특별시<br>중랑구 | NaN | 청운동  | 0.0    | 1.0   | 0.0 | NaN  | 청운역상<br>별치자 | 1     | 121    | 157,250     | 926000000.0 | 3.0     | 1.0  | 3.0  |
| 3        | 2022 | 1    | 1.111010e+09 | 서울특별시 중랑구 지하<br>모로36길 16-14 | 서울특별시<br>중랑구 | NaN | 청운동  | 0.0    | 1.0   | 0.0 | NaN  | 청운역상<br>별치자 | 1     | 122    | 157,250     | 926000000.0 | 3.0     | 1.0  | 4.0  |
| 4        | 2022 | 1    | 1.111010e+09 | 서울특별시 중랑구 지하<br>모로36길 16-14 | 서울특별시<br>중랑구 | NaN | 청운동  | 0.0    | 1.0   | 0.0 | NaN  | 청운역상<br>별치자 | 1     | 131    | 152,330     | 905000000.0 | 3.0     | 1.0  | 5.0  |
| -        | -    | -    | -            | -                           | -            | -   | -    | -      | -     | -   | -    | -           | -     | -      | -           | -           | -       | -    | -    |
| 525204   | 2022 | 1    | 1.129014e+09 | 서울특별시 성북구 오래<br>선로 46       | 서울특별시<br>성북구 | NaN | 하월곡동 | 0.0    | 222.0 | 0.0 | NaN  | 월곡두산<br>입비  | 113.0 | 1601.0 | 114,752     | 713000000.0 | 52690.0 | 13.0 | 14.0 |
| 525205   | 2022 | 1    | 1.129014e+09 | 서울특별시 성북구 오래<br>선로 46       | 서울특별시<br>성북구 | NaN | 하월곡동 | 0.0    | 222.0 | 0.0 | NaN  | 월곡두산<br>입비  | 113.0 | 1602.0 | 114,752     | 713000000.0 | 52690.0 | 13.0 | 15.0 |
| 525206   | 2022 | 1    | 1.129014e+09 | 서울특별시 성북구 오래<br>선로 46       | 서울특별시<br>성북구 | NaN | 하월곡동 | 0.0    | 222.0 | 0.0 | NaN  | 월곡두산<br>입비  | 113.0 | 1701.0 | 114,752     | 713000000.0 | 52690.0 | 13.0 | 16.0 |
| 525207   | 2022 | 1    | 1.129014e+09 | 서울특별시 성북구 오래<br>선로 46       | 서울특별시<br>성북구 | NaN | 하월곡동 | 0.0    | 222.0 | 0.0 | NaN  | 월곡두산<br>입비  | 113.0 | 1702.0 | 114,752     | 713000000.0 | 52690.0 | 13.0 | 17.0 |



### 지하철 역세권, 버스정류장세권

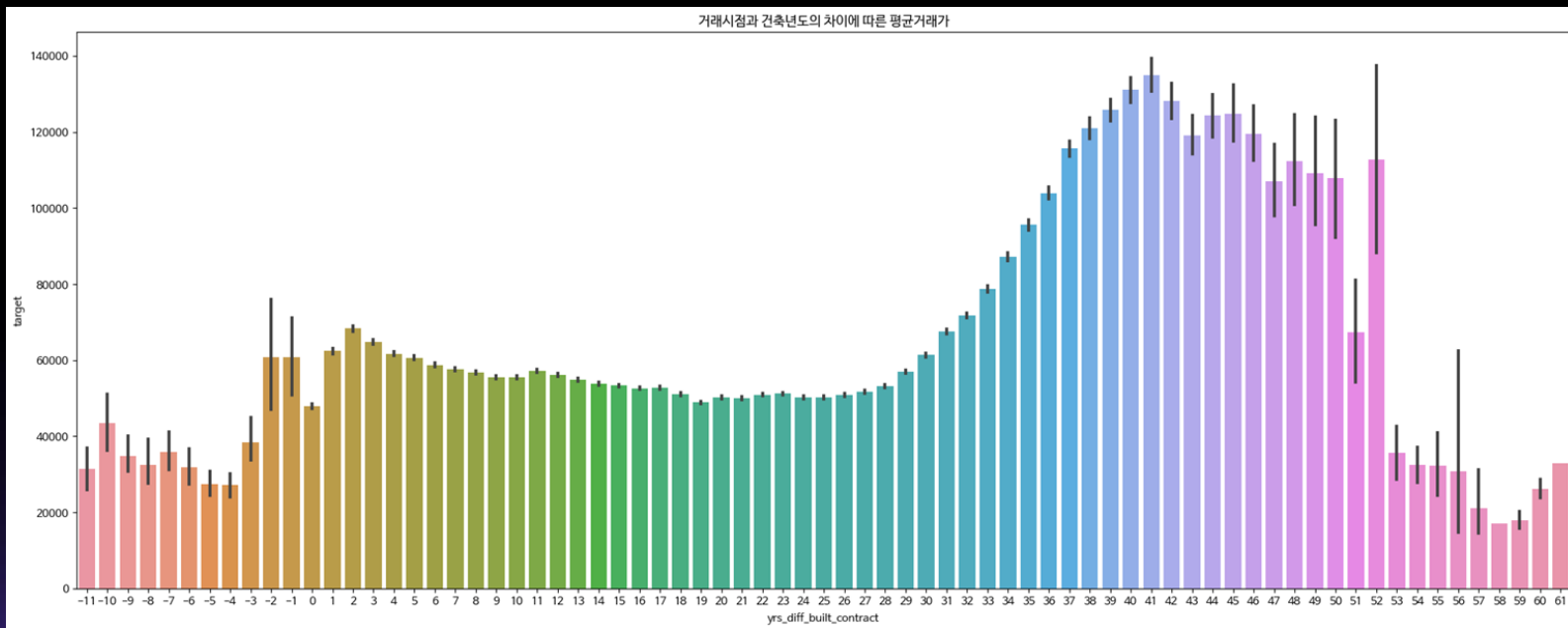
대회에 제공된 지하철과 버스정류장 데이터를 이용하여 지하철역은 1km, 버스 정류장은 500m 내에 있으면 역세권 true를 반환. 그러나 feature importance에서 중요도가 낮고, 실제 적용 후 제출한 결과 리더보드 상에서도 점수가 개선되지 않았음. 가격예측에 큰 영향을 안 미칠것이라 판단, drop하기로 결정.



# Feature Engineering

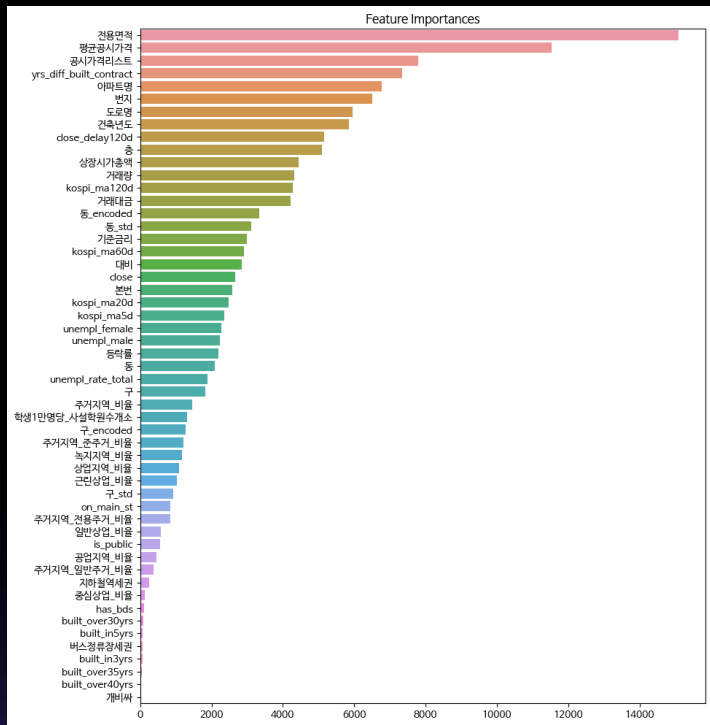
## Yrs\_diff = (계약년 - 건축년도)

- 단순히 건축년도나 계약년도보다는 계약 시점에 아파트가 얼마나 오래되었는가 중요하다 판단
- 준공 후 30년 이상이 되어야 재건축이 가능하기 때문에 yrs\_diff 28이상부터 41까지 가파르게 평균가가 오르는 것을 확인

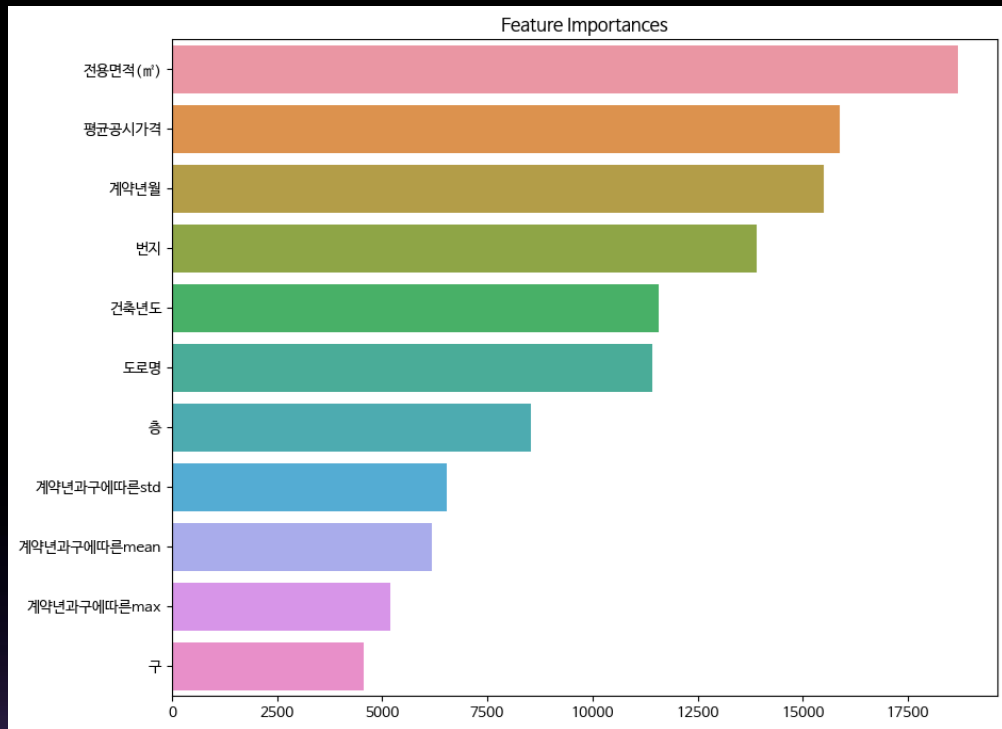


# Feature Select

RMSE : 17642.1889 → 12607.6082



RMSE : 15222.3181 → 13346.8945



04

# Modeling

# Model Select

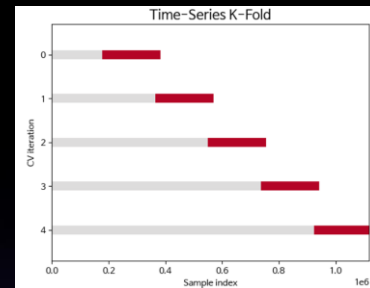
## Baseline

### Features

|   | 시군구           | 번지    | 본번    | 아파트명   | 전용면적(㎡) | 계약년월   | 층 | 건축년도 | 도로명   | target |
|---|---------------|-------|-------|--------|---------|--------|---|------|-------|--------|
| 0 | 서울특별시 강남구 개포동 | 658-1 | 658.0 | 개포6차우성 | 79.97   | 201712 | 3 | 1987 | 언주로 3 | 124000 |

### Dataset split

|             |                                               |                              |                   |
|-------------|-----------------------------------------------|------------------------------|-------------------|
| 분할 방식       | Hold out random<br>(test_size=0.2, 0.1, 0.01) | Hold out time split<br>(8:2) | Time series split |
| Public rmse | 18410                                         | 51968                        | 74542             |



### Optimize Hyperparameter- Hold out random (test\_size=0.2)

```
▼ LGBMRegressor
LGBMRegressor(bagging_fraction=0.9, feature_fraction=0.8, max_depth=32,
               n_estimators=2000, num_leaves=60, verbosity=-1)
```

# Hyperparameter Tuning

## Optuna를 이용한 하이퍼파라미터 튜닝

```
best_model = optimize_hyperparameters(n_trials=100)

[92] ✓ 91m 18.6s

... [I 2024-03-29 06:08:47.232] A new study created in memory with name: no-name-6c971f5c-41ad-4746-b0fe-d6104a31f3e3
Training until validation scores don't improve for 50 rounds
[100] training's rmse: 6826.27 training's l2: 4.6598e+07 valid_1's rmse: 7273.64 valid_1's l2: 5.29058e+07
[200] training's rmse: 5978 training's l2: 3.57365e+07 valid_1's rmse: 6596.51 valid_1's l2: 4.3514e+07
[300] training's rmse: 5552.44 training's l2: 3.08296e+07 valid_1's rmse: 6312.03 valid_1's l2: 3.98418e+07
[400] training's rmse: 5254.72 training's l2: 2.76121e+07 valid_1's rmse: 6140.8 valid_1's l2: 3.77094e+07
[500] training's rmse: 5027.56 training's l2: 2.52763e+07 valid_1's rmse: 6020.36 valid_1's l2: 3.62447e+07
[600] training's rmse: 4848.59 training's l2: 2.35088e+07 valid_1's rmse: 5936.97 valid_1's l2: 3.52476e+07
[700] training's rmse: 4702.92 training's l2: 2.21175e+07 valid_1's rmse: 5876.67 valid_1's l2: 3.45352e+07
[800] training's rmse: 4580.74 training's l2: 2.09832e+07 valid_1's rmse: 5828.32 valid_1's l2: 3.39693e+07
[900] training's rmse: 4466.16 training's l2: 1.99465e+07 valid_1's rmse: 5790.81 valid_1's l2: 3.35334e+07
[1000] training's rmse: 4360.42 training's l2: 1.90133e+07 valid_1's rmse: 5762.74 valid_1's l2: 3.32092e+07
[1100] training's rmse: 4268.11 training's l2: 1.82168e+07 valid_1's rmse: 5736.92 valid_1's l2: 3.29123e+07
[1200] training's rmse: 4186.44 training's l2: 1.75263e+07 valid_1's rmse: 5718.05 valid_1's l2: 3.2696e+07
[1300] training's rmse: 4110.71 training's l2: 1.68979e+07 valid_1's rmse: 5702.93 valid_1's l2: 3.25234e+07
[1400] training's rmse: 4041.91 training's l2: 1.63371e+07 valid_1's rmse: 5689.71 valid_1's l2: 3.23728e+07
[1500] training's rmse: 3978.46 training's l2: 1.58282e+07 valid_1's rmse: 5674.08 valid_1's l2: 3.21952e+07
[1600] training's rmse: 3923.28 training's l2: 1.53921e+07 valid_1's rmse: 5669.09 valid_1's l2: 3.21386e+07
[1700] training's rmse: 3866.6 training's l2: 1.49506e+07 valid_1's rmse: 5661.8 valid_1's l2: 3.2056e+07
[1800] training's rmse: 3814.02 training's l2: 1.45467e+07 valid_1's rmse: 5650.81 valid_1's l2: 3.19316e+07
[1900] training's rmse: 3760.69 training's l2: 1.41428e+07 valid_1's rmse: 5644.2 valid_1's l2: 3.1857e+07
[2000] training's rmse: 3708.56 training's l2: 1.37534e+07 valid_1's rmse: 5641.07 valid_1's l2: 3.18216e+07
[2100] training's rmse: 3663.53 training's l2: 1.34215e+07 valid_1's rmse: 5636.92 valid_1's l2: 3.17749e+07
[2200] training's rmse: 3619.21 training's l2: 1.30987e+07 valid_1's rmse: 5635 valid_1's l2: 3.17532e+07
```

```
best_model.get_params()

1] ✓ 0.0s

{'boosting_type': 'gbdt',
 'class_weight': None,
 'colsample_bytree': 1.0,
 'importance_type': 'split',
 'learning_rate': 0.1,
 'max_depth': 49,
 'min_child_samples': 233,
 'min_child_weight': 0.001,
 'min_split_gain': 0.0,
 'n_estimators': 4368,
 'n_jobs': None,
 'num_leaves': 177,
 'objective': None,
 'random_state': None,
 'reg_alpha': 0.0,
 'reg_lambda': 0.0,
 'subsample': 1.0,
 'subsample_for_bin': 200000,
 'subsample_freq': 0,
 'feature_fraction': 0.6691188003121078,
 'bagging_fraction': 0.9858111931327013,
 'lambda_l1': 1.7965786613884013e-05,
 'lambda_l2': 1.5944039983833933e-08}
```

## 추가한 Features (Feature engineering)

### 구, 동단위 기술 통계값

| 계약년과구에따른std  | 계약년과구에따른mean | 계약년과구에따른max |
|--------------|--------------|-------------|
| 60408.755329 | 127433.7879  | 1053000.0   |

| 계약년과동에따른std  | 계약년과동에따른mean  | 계약년과동에따른max |
|--------------|---------------|-------------|
| 32126.755612 | 108070.249556 | 260000.0    |

### 공시가격 데이터

| 평균 공시가격       | 공시가격리스트                                           |
|---------------|---------------------------------------------------|
| 120960.000000 | [125300.0, 125300.0, 125300.0, 125300.0, 10390... |
| 120960.000000 | [125300.0, 125300.0, 125300.0, 125300.0, 10390... |
| 120960.000000 | [125300.0, 125300.0, 125300.0, 125300.0, 10390... |
| 120960.000000 | [125300.0, 125300.0, 125300.0, 125300.0, 10390... |
| 120960.000000 | [125300.0, 125300.0, 125300.0, 125300.0, 10390... |
| ...           | ...                                               |
| 39021.326676  | [42800.0, 41800.0, 44000.0, 43100.0, 44700.0, ... |

### 지하철 역세권, 버스정류장세권

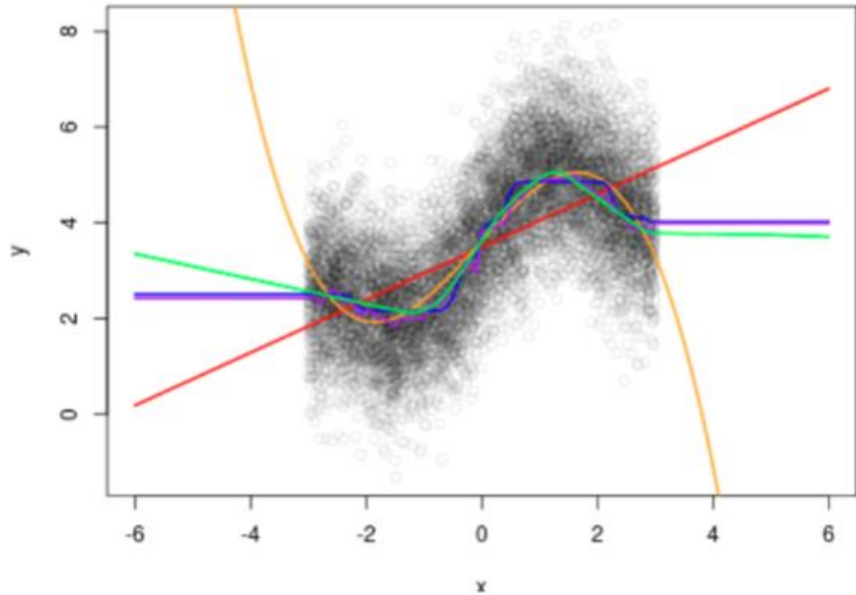


| 번지      | 전용면적(m²) | 계약년월   | 층 | 건축년도 | 도로명   | target   | is_test | 구   | 평균공시가격   | 계약년과동에따른std  | 계약년과동에따른mean  | 계약년과동에따른max |
|---------|----------|--------|---|------|-------|----------|---------|-----|----------|--------------|---------------|-------------|
| 0 658-1 | 79.97    | 201712 | 3 | 1987 | 연주로 3 | 124000.0 | 0       | 강남구 | 120960.0 | 32126.755612 | 108070.249556 | 260000.0    |

## 높은 가격을 예측하는 모델의 필요성

### 내삽과 외삽

- 내삽: 주위에 데이터가 많을 때, 결과 값을 예측하는 것
- 외삽: 대부분의 데이터와 동떨어진 점에서 결과 값을 예측하는 것



- 선형 모형  $y = b_0 + b_1x$
- 3차 다항 모형  $y = b_0 + b_1x + b_2x^2 + b_3x^3$
- 랜덤 포레스트
- 그래디언트 부스팅 머신(xgboost)
- 인공신경망

Max 를 올리는 것이 핵심인가.

---

RMSE 17000 대

|       | target       |
|-------|--------------|
| count | 9.272000e+03 |
| mean  | 1.070703e+05 |
| std   | 7.676328e+04 |
| min   | 7.885000e+03 |
| 25%   | 6.286500e+04 |
| 50%   | 8.754050e+04 |
| 75%   | 1.263052e+05 |
| max   | 1.095998e+06 |



RMSE 15000 대

|       | target       |
|-------|--------------|
| count | 9.272000e+03 |
| mean  | 1.079307e+05 |
| std   | 7.701048e+04 |
| min   | 8.919000e+03 |
| 25%   | 6.316725e+04 |
| 50%   | 8.824300e+04 |
| 75%   | 1.279645e+05 |
| max   | 1.279440e+06 |



## Various Models with Trial and Error

### 예측 성능 하락의 원인 분석

- 가장 예측성능이 좋은 모델에서 잔차의 절댓값이 50000 이상인 데이터들을 세부적으로 분석해 봄.

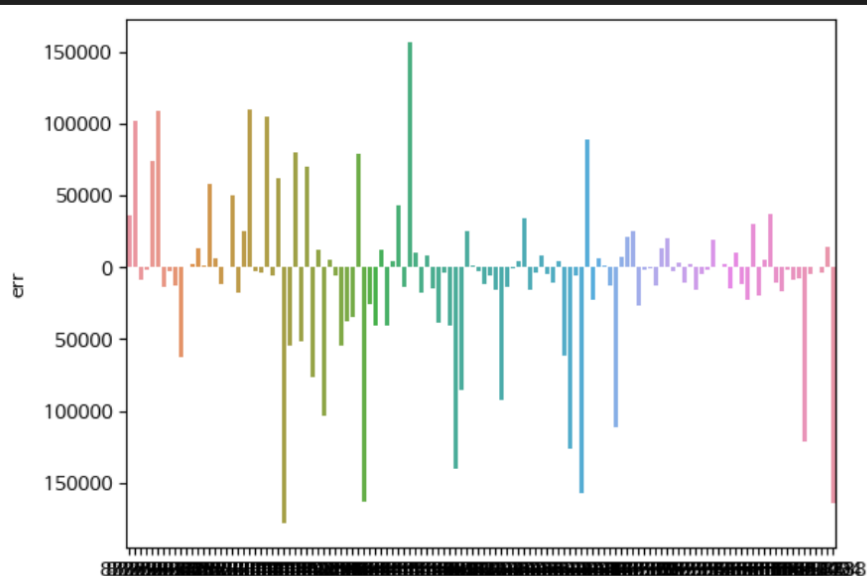
```
err_['아파트명'].value_counts()
```

✓ 0.0s

|          |    |
|----------|----|
| 한남더힐     | 30 |
| 갤러리아포레   | 8  |
| 타워팰리스2   | 6  |
| 아이파크     | 5  |
| 상지리츠빌카일룸 | 5  |

|         |   |
|---------|---|
| ..      |   |
| 잠실시그마타워 | 1 |
| 르네상스빌   | 1 |
| 그랜빌캐슬   | 1 |
| 현대3차    | 1 |
| 준아트빌    | 1 |

Name: 아파트명, Length: 123, dtype: int64



## 멘토님께 받은 피드백 및 의견

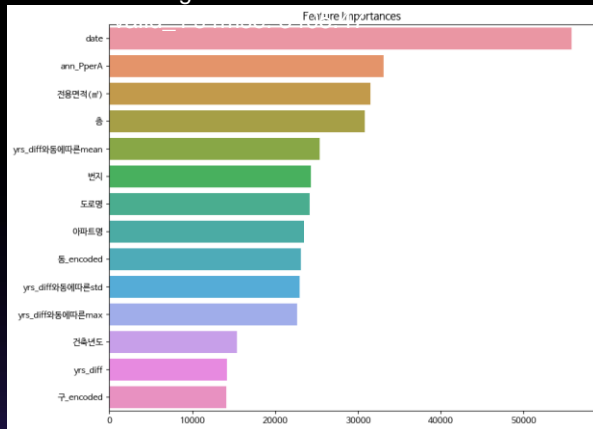
비싼 아파트를 잘 예측해야 RMSE가 낮게 나오므로 두 개의 모델로 분리하여 예측하는 것을 추천  
EDA 결과 전용면적(㎡)을 기준으로 60미만, 60이상~135미만, 135이상으로 나누어 모델링

→ 135㎡이상의 아파트 데이터가 적어 예측 성능저하

→ 135㎡이상의 경우만 5-fold CV를 해보았으나 fold 별 val rmse가 6575.63~8860.78

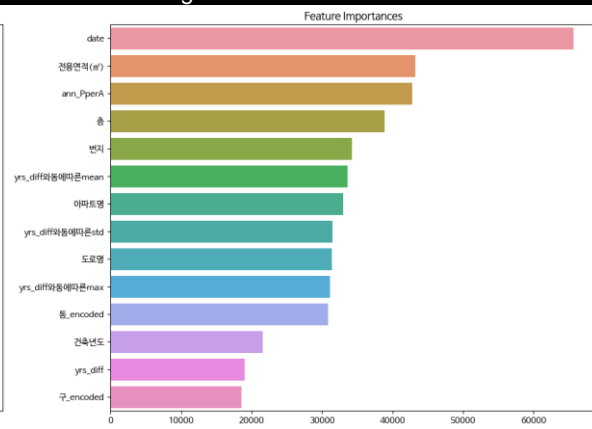
gbm0

● training's rmse: 1554.89



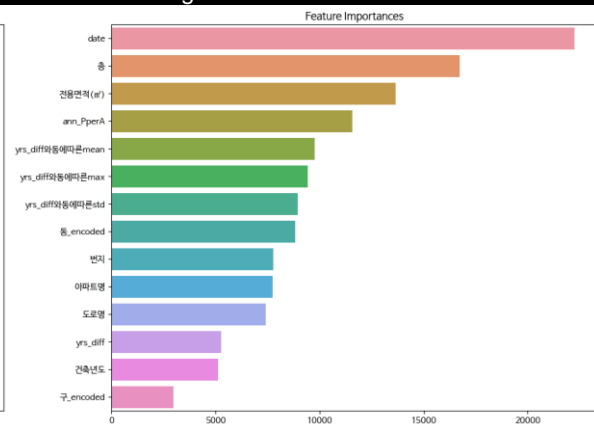
gbm60

● training's rmse: 2529.51



gbm135

● training's rmse: 3073.22



# 50억 이상 아파트에 대해 분석

`df[df['아파트명']=='PH129']`

✓ 0.0s Python

|        | 번<br>지 | 아파<br>트명 | 전용<br>면적<br>(㎡) | 계약년<br>월 | 층  | 건축<br>년도 | 도로<br>명     | target    | is_test | 평균공시가격        | 동   | 계약년과동<br>에따른std | 계약년과동<br>에따른mean | 계약년과<br>동에따른<br>max | yrs_diff | Close_de |
|--------|--------|----------|-----------------|----------|----|----------|-------------|-----------|---------|---------------|-----|-----------------|------------------|---------------------|----------|----------|
| 224566 | 129    | PH129    | 273.96          | 202111   | 10 | 2020     | 압구정로 79길 88 | 929070.0  | 0       | 854851.724138 | 청담동 | 199738.445661   | 277167.682927    | 1150000.0           | 1        |          |
| 224567 | 129    | PH129    | 273.96          | 202204   | 16 | 2020     | 압구정로 79길 88 | 1450000.0 | 0       | 854851.724138 | 청담동 | 230216.769107   | 249733.258065    | 1450000.0           | 2        |          |
| 241238 | 129    | PH129    | 273.96          | 202103   | 14 | 2020     | 압구정로 79길 88 | 1150000.0 | 0       | 854851.724138 | 청담동 | 199738.445661   | 277167.682927    | 1150000.0           | 1        |          |
| 241239 | 129    | PH129    | 273.96          | 202107   | 6  | 2020     | 압구정로 79길 88 | 1000000.0 | 0       | 854851.724138 | 청담동 | 199738.445661   | 277167.682927    | 1150000.0           | 1        |          |

# 50억 이상 아파트에 대해 분석

df[df['아파트명']=='파크한남']

✓ 0.0s Python

| 번호      | 아파트명 | 전용면적 (㎡) | 계약년월   | 층 | 건축년도 | 도로명          | target    | is_test | 평균공시가격        | 동   | 계약년과동에따른 std  | 계약년과동에따른 mean | 계약년과동에따른 max | yrs_diff | Close_delay120d |
|---------|------|----------|--------|---|------|--------------|-----------|---------|---------------|-----|---------------|---------------|--------------|----------|-----------------|
| 236660  | 파크한남 | 268.95   | 202111 | 3 | 2020 | 유엔빌리지길 80-1  | 1150000.0 | 0       | 324238.544892 | 한남동 | 284088.039234 | 297951.973684 | 1200000.0    | 1        | 3221.870        |
| 236661  | 파크한남 | 268.67   | 202111 | 4 | 2020 | 유엔빌리지길 80-1  | 1170000.0 | 0       | 478113.931889 | 한남동 | 284088.039234 | 297951.973684 | 1200000.0    | 1        | 3247.430        |
| 236662  | 파크한남 | 268.95   | 202112 | 2 | 2020 | 유엔빌리지길 80-1  | 1200000.0 | 0       | 631989.318885 | 한남동 | 284088.039234 | 297951.973684 | 1200000.0    | 1        | 3267.930        |
| 236663  | 파크한남 | 268.67   | 202204 | 5 | 2020 | 유엔빌리지길 80-36 | 1350000.0 | 0       | 785864.705882 | 한남동 | 354143.419439 | 416847.500000 | 1350000.0    | 2        | 3197.565        |
| 288789  | 파크한남 | 268.67   | 202108 | 3 | 2020 | 유엔빌리지길 80-1  | 1000000.0 | 0       | 151718.305454 | 한남동 | 284088.039234 | 297951.973684 | 1200000.0    | 1        | 3100.580        |
| 288790  | 파크한남 | 268.67   | 202109 | 2 | 2020 | 유엔빌리지길 80-1  | 1080000.0 | 0       | 113647.137223 | 한남동 | 284088.039234 | 297951.973684 | 1200000.0    | 1        | 3035.460        |
| 1127415 | 파크한남 | 268.67   | 202308 | 4 | 2020 | 유엔빌리지길 80-36 | NaN       | 1       | 785864.705882 | 한남동 | 279436.203653 | 334323.529412 | 1100000.0    | 3        | 2417.680        |

# 50억 이상 아파트에 대해 분석

파크한남 = test에 1개잇음 / 유엔빌리지길 80-1 / 가격 말안됨. 공시가격 15억인데 100억에 거래됨. 애는 따로 처리해줘야할 듯.

파크한남 = 유엔빌리지길 80-36

아크로 서울 포레스트 = test에 있음 / 왕십리로 83-21 / 대체로 가격 비쌌 60억 이상~ / 이거 260평짜리가 130억인데, test에 260평짜리 없음 그냥 지우기 ㅋㅋ 233420  
르가든더맨한남 = 잇음 / 한남대로10길 16 / 한남동 / 90억짜리 아파트가 있는데 6층이 2개 2층이 1개 2층이름 이상함 가격.

한남더힐 = 대체로 20억 이상하는 아파트 ㅇㅇ

아이파크 = 잇음 / 영동대로 640 / 삼성동 / 많은 아파트가 있음;

래미안 퍼스티지 잇음. / 반포대로 275/ 내비두기 많네.

갤러리아 포레 잇음. / 서울숲2길 32-14 / 내비두기 많음.

마크힐스이스트윙 = 없음 / 영동대로142길 19 / 청담동 / 애는 딱히 튀지않은 적당히 삼.

마크힐스웨스트윙 = 없음 / 영동대로142길 21 / 청담동 / 애는 1.5배 이상 비싸게 사는 경향이 있는데, 넘어가겠음

상지카일름 = 없음 / 도산대로101길 38 / 청담동 / 애 1.5배 이상 비싸게 사는 친구 잇음.

상지리츠빌카일름 = 없음 / 남부순환로373길 23 / 도곡동 / 23년에 1.5배 이상 비싸게 산거 잇음.

상지리츠빌카일름 = 없음 / 삼성로 636 / 삼성동 / 22년에 1.5배 이상 비싸게 삼.

상지리츠빌장충동카일름 = 없음 / 동호로20나길 30 / 장충동1가 / 23년에 거의 2배 비싸게삼, 2017년에 슬슬 비싸게 사기 시작함.

상지리츠빌카일름2차 = 없음 / 압구정로79길 87 / 애 거의 23년에 90억 가까이사는데... 청담동이 비싼 동네는 맞지만 이걸 너무비산데...? outlier라고 판단됨.

상지리츠빌카일름3차= 없음 / 압구정로79길 72 / 평수가 달라서 싸게산것 처럼 느껴지지만, 비싸게삼 근데 그중 394190 애 투자의신임. 싸게삼.

압구정하이츠파크 = 없음 / 압구정로42길 78 / 신사동 / 23년에 영끌해서 비싸게사네... ?

빌폴라리스 =없음 / 압구정로79길 76 / 청담동 / 23년 영끌해서 거의 2배가까이 비싸게 삼.

이그니마빌 = 없음 / 압구정로62길 17-4 / 청담동 / 애 23년에 매우싸게삼 !!! 이런친구를 삭제해야함 이게 맞음.

이그니마빌2 = 없음 / 도산대로61길 19 / 청담동 / 2019년에 싸게 산 친구 있는데, 지워야함 내가 생각했을 때.

청담어퍼하우스 = 없음 / 학동로81길 28 / 청담동 / 2020년에 정가에 산 사람이 있음 .

레베빌 = 없음 / 동광로39길 18 / 반포동 / 23년도에 3배 뛰기힘. 말안됨. 20년도에 2배뛰기힘.

반포주공1단지 = 없음 / 신반포로 45 / 반포동 / 18년도에 3배뛰기힘 .

떼레라하우스 = 없음 / 잠원로 221-31 / 잠원동 / 21년에 정가에 산애 있음 없애야함.

힐데스하임빌라 = 없음 / 남부순환로377길 26 / 도곡동 / 21년 19년에 정가에 산애 있음

삼성동해안하우스2 = 없음 / 영동대로128길 45 / 삼성동 / 여기는 꽤 괜찮은데?

아펠바움 = 없음 / 영동대로114길 72 / 삼성동 / 공시가격보다 2018년부터 비싸게 사기 시작함. 아주 좋음

연세리버빌3 = 없음 / 선릉로190길 84 / 청담동 / 2019년에 공시가격과 거의 비슷하게 산친구 있음 투자의 신임. 없애야함.

효성빌라청담101 = 없음 / 압구정로71길 28 / 청담동 / 아주 비싸게삼 70억 거래는 없애도... 될듯?

래미안첼리투스 = 없음 / 이촌로 310 / 이촌동 / 115개 인걸 봐서 2015년 건축 2018년 분양? 그리고 적당한 거래고 23년에 2배뛰워기. 적당함.

로덴하우스이스트빌리지 = 없음 / 남부순환로378길 9 / 도곡동 / 2020년에 공시가랑 비슷하게 산거 있음 투자의신임 없애야함.

로덴하우스웨스트빌리지 = 없음 / 남부순환로378길 11 / 도곡동 / 5 index clear

미성1차 = 없음 / 압구정로11길 17 / 압구정동 / 58 내비둡시다.

한양8차 = 없음 / 압구정로 403 / 압구정동 / 23 / 여기 19년도에 싸게사는 투자의 신들잇음. 이거 없애야하고 . 여길보면 15~ 17년도에 엄청 싸게삼 noise임.

갤러리아포레 = 없음 / 왕십리로 85 / 성수동 1가 / 30 index / 애 지워야할듯? 이유는 2015년~ 2016년사이에 많음. 이집 사람들 좀버 잘하네.

논현라폴리움 = 없음 / 학동로46길 32 / 논현동 / 8index / 17년에 좀 싸게사네? 투자의신. 17년도를 없앨까 고민하는 분기점.

ph129 = 없음 / index로 없애버림 / 보지도마셈 ㅋㅋ /

# 50억 이상 아파트에 대해 분석

```
df[df['도로명']=='한남대로10길 16']
```

✓ 0.0s

Python

|         | 번<br>지 | 아파트명         | 전용면적<br>(㎡) | 계약년<br>월 | 층 | 건축년<br>도 | 도로명            | target   | is_test | 평균공시가<br>격 | 동       | 계약년과동에따른<br>std | 계약년과동에따른<br>mean | 계약년과동에따른<br>max | yrs_diff | Close_delay120d |
|---------|--------|--------------|-------------|----------|---|----------|----------------|----------|---------|------------|---------|-----------------|------------------|-----------------|----------|-----------------|
| 221950  | 386    | 르가든더메인<br>한남 | 227.37      | 202306   | 2 | 2019     | 한남대로10<br>길 16 | 465000.0 | 0       | 429625.0   | 한남<br>동 | 279436.203653   | 334323.529412    | 1100000.0       | 4        | 2313.69         |
| 236650  | 386    | 르가든더메인<br>한남 | 225.41      | 202205   | 6 | 2019     | 한남대로10<br>길 16 | 900000.0 | 0       | 429625.0   | 한남<br>동 | 354143.419439   | 416847.500000    | 1350000.0       | 3        | 2839.01         |
| 236651  | 386    | 르가든더메인<br>한남 | 269.12      | 202206   | 2 | 2019     | 한남대로10<br>길 16 | 900000.0 | 0       | 429625.0   | 한남<br>동 | 354143.419439   | 416847.500000    | 1350000.0       | 3        | 2973.25         |
| 288737  | 386    | 르가든더메인<br>한남 | 238.38      | 202105   | 3 | 2019     | 한남대로10<br>길 16 | 450000.0 | 0       | 429625.0   | 한남<br>동 | 284088.039234   | 297951.973684    | 1200000.0       | 2        | 2601.54         |
| 288738  | 386    | 르가든더메인<br>한남 | 227.37      | 202105   | 2 | 2019     | 한남대로10<br>길 16 | 400000.0 | 0       | 429625.0   | 한남<br>동 | 284088.039234   | 297951.973684    | 1200000.0       | 2        | 2625.91         |
| 288739  | 386    | 르가든더메인<br>한남 | 228.34      | 202109   | 6 | 2019     | 한남대로10<br>길 16 | 970000.0 | 0       | 429625.0   | 한남<br>동 | 284088.039234   | 297951.973684    | 1200000.0       | 2        | 3041.01         |
| 1127412 | 386    | 르가든더메인<br>한남 | 237.52      | 202307   | 3 | 2019     | 한남대로10<br>길 16 | NaN      | 1       | 429625.0   | 한남<br>동 | 279436.203653   | 334323.529412    | 1100000.0       | 4        | 2359.53         |
| 1127413 | 386    | 르가든더메인<br>한남 | 239.67      | 202307   | 4 | 2019     | 한남대로10<br>길 16 | NaN      | 1       | 429625.0   | 한남<br>동 | 279436.203653   | 334323.529412    | 1100000.0       | 4        | 2484.02         |

# 50억 이상 아파트에 대해 분석

df[df['아파트명']=='아크로서울포레스트']

✓ 0.0s Python

|         | 번지      | 아파트명      | 전용면적<br>(㎡) | 계약년<br>월 | 층  | 건축년<br>도 | 도로명        | target    | is test | 평균공시가격        | 동      | 계약년과동에따른<br>std | 계약년과동에따른<br>mean | 계약년과동에따른<br>max | yrs_diff | Close_delay120d |
|---------|---------|-----------|-------------|----------|----|----------|------------|-----------|---------|---------------|--------|-----------------|------------------|-----------------|----------|-----------------|
| 216384  | 685-700 | 아크로서울포레스트 | 159.603     | 202303   | 27 | 2020     | 왕십리로 83-21 | 620000.0  | 0       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2261.925        |
| 216385  | 685-700 | 아크로서울포레스트 | 159.603     | 202304   | 45 | 2020     | 왕십리로 83-21 | 675000.0  | 0       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2288.780        |
| 216386  | 685-700 | 아크로서울포레스트 | 164.856     | 202304   | 3  | 2020     | 왕십리로 83-21 | 575000.0  | 0       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2348.430        |
| 233419  | 685-700 | 아크로서울포레스트 | 165.182     | 202204   | 6  | 2020     | 왕십리로 83-21 | 620000.0  | 0       | 423578.928571 | 성수동 1가 | 262281.567769   | 300048.809524    | 1300000.0       | 2        | 3029.040        |
| 233420  | 685-700 | 아크로서울포레스트 | 264.546     | 202209   | 47 | 2020     | 왕십리로 83-21 | 1300000.0 | 0       | 423578.928571 | 성수동 1가 | 262281.567769   | 300048.809524    | 1300000.0       | 2        | 2695.860        |
| 276041  | 685-700 | 아크로서울포레스트 | 159.603     | 202102   | 43 | 2020     | 왕십리로 83-21 | 560000.0  | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2432.350        |
| 276042  | 685-700 | 아크로서울포레스트 | 159.603     | 202102   | 43 | 2020     | 왕십리로 83-21 | 560000.0  | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2432.350        |
| 276043  | 685-700 | 아크로서울포레스트 | 164.856     | 202102   | 11 | 2020     | 왕십리로 83-21 | 350000.0  | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2437.530        |
| 276044  | 685-700 | 아크로서울포레스트 | 159.603     | 202103   | 34 | 2020     | 왕십리로 83-21 | 530000.0  | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2406.170        |
| 276045  | 685-700 | 아크로서울포레스트 | 159.603     | 202103   | 34 | 2020     | 왕십리로 83-21 | 530000.0  | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2406.170        |
| 276046  | 685-700 | 아크로서울포레스트 | 97.485      | 202103   | 7  | 2020     | 왕십리로 83-21 | 280000.0  | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2332.590        |
| 276047  | 685-700 | 아크로서울포레스트 | 198.219     | 202104   | 38 | 2020     | 왕십리로 83-21 | 552000.0  | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2345.260        |
| 1124878 | 685-700 | 아크로서울포레스트 | 198.219     | 202307   | 40 | 2020     | 왕십리로 83-21 | NaN       | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2359.530        |
| 1124879 | 685-700 | 아크로서울포레스트 | 95.810      | 202307   | 4  | 2020     | 왕십리로 83-21 | NaN       | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2399.860        |
| 1124880 | 685-700 | 아크로서울포레스트 | 159.603     | 202307   | 23 | 2020     | 왕십리로 83-21 | NaN       | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2419.475        |
| 1124881 | 685-700 | 아크로서울포레스트 | 198.219     | 202308   | 41 | 2020     | 왕십리로 83-21 | NaN       | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2439.090        |

## 50억 이상 아파트에 대해 분석

---

아크로서울포레스트 = 233420, 276043,  
르가든더메인한남 = 236651,  
상지리츠빌카일룸3차 = 394194,  
이그나빌 = 203593,  
이그나빌2 = 394206,  
로덴하우스이스트빌리지 = 296155,  
PH129 = 224566, 224567, 241238, 241239



# VVIP 저격 모델 만들기.

```
exp_dong = ['반포동', '한남동', '청담동', '압구정동', '삼성동', '도곡동', '성수동1가']
```

Pythor

```
df['exp'] = df['동'].isin(exp_dong)*1
```

```
df['m135'] = (df['전용면적(m')']>135) *1
```

```
df['exp2'] = df['exp']+df['m135']
```

```
df
```

Pythor

|         | 번지    | 전용면적<br>(m) | 계약년<br>월 | 층   | 건축년<br>도 | 도로명           | target   | is_test | 평균공시가격        | 동   | 계약년과동에따른<br>std | 계약년과동에따른<br>mean | 계약년과동에따른<br>max | yrs_diff | Close_delay120d | exp | m135 | exp2 |
|---------|-------|-------------|----------|-----|----------|---------------|----------|---------|---------------|-----|-----------------|------------------|-----------------|----------|-----------------|-----|------|------|
| 0       | 658-1 | 79.97       | 201712   | 3   | 1987     | 연주로 3         | 124000.0 | 0       | 120960.000000 | 개포동 | 32126.755612    | 108070.249556    | 260000.0        | 30       | 2372.640000     | 0   | 0    | 0    |
| 1       | 658-1 | 79.97       | 201712   | 4   | 1987     | 연주로 3         | 123500.0 | 0       | 120960.000000 | 개포동 | 32126.755612    | 108070.249556    | 260000.0        | 30       | 2382.560000     | 0   | 0    | 0    |
| 2       | 658-1 | 54.98       | 201712   | 5   | 1987     | 연주로 3         | 91500.0  | 0       | 120960.000000 | 개포동 | 32126.755612    | 108070.249556    | 260000.0        | 30       | 2394.480000     | 0   | 0    | 0    |
| 3       | 658-1 | 79.97       | 201801   | 4   | 1987     | 연주로 3         | 130000.0 | 0       | 120960.000000 | 개포동 | 43598.058342    | 142378.337531    | 260000.0        | 31       | 2388.350000     | 0   | 0    | 0    |
| 4       | 658-1 | 79.97       | 201801   | 2   | 1987     | 연주로 3         | 117000.0 | 0       | 120960.000000 | 개포동 | 43598.058342    | 142378.337531    | 260000.0        | 31       | 2382.100000     | 0   | 0    | 0    |
| ...     | ...   | ...         | ...      | ... | ...      | ...           | ...      | ...     | ...           | ... | ...             | ...              | ...             | ...      | ...             | ... | ...  | ...  |
| 1128089 | 816   | 84.65       | 202307   | 13  | 2014     | 신내역로1<br>길 85 | NaN      | 1       | 39021.326676  | 신내동 | 11341.218441    | 54184.649123     | 85000.0         | 9        | 2428.570000     | 0   | 0    | 0    |
| 1128090 | 816   | 84.62       | 202307   | 12  | 2014     | 신내역로1<br>길 85 | NaN      | 1       | 39021.326676  | 신내동 | 11341.218441    | 54184.649123     | 85000.0         | 9        | 2425.080000     | 0   | 0    | 0    |
| 1128091 | 816   | 101.65      | 202308   | 12  | 2014     | 신내역로1<br>길 85 | NaN      | 1       | 39021.326676  | 신내동 | 11341.218441    | 54184.649123     | 85000.0         | 9        | 2420.253333     | 0   | 0    | 0    |
| 1128092 | 816   | 84.94       | 202309   | 18  | 2014     | 신내역로1<br>길 85 | NaN      | 1       | 39021.326676  | 신내동 | 11341.218441    | 54184.649123     | 85000.0         | 9        | 2415.426667     | 0   | 0    | 0    |
| 1128093 | 816   | 84.65       | 202309   | 13  | 2014     | 신내역로1<br>길 85 | NaN      | 1       | 39021.326676  | 신내동 | 11341.218441    | 54184.649123     | 85000.0         | 9        | 2410.600000     | 0   | 0    | 0    |

645997 rows x 18 columns

# VVIP 저격 모델 만들기.

```
df_train0 = df_train[df_train['exp2']!=2]

df_train_exp = df_train[(df_train['exp2']==2)]
```

df\_train\_exp

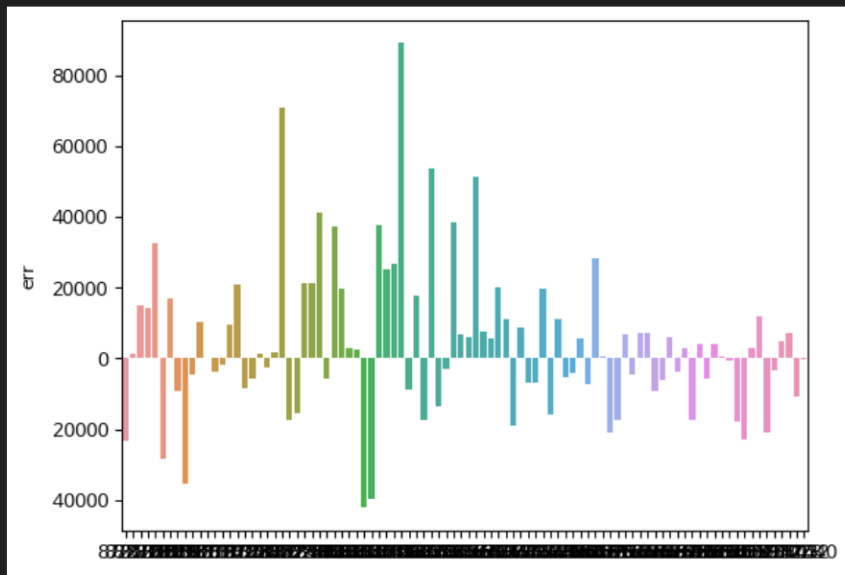
|        | 번지   | 전용면적(m²) | 계약년월     | 층      | 건축년도 | 도로명  | target   | 평균공시가격        | 동             | 계약년과동에따른std  | 계약년과동에따른mean  | 계약년과동에따른max   | yrs_diff | exp2 |     |
|--------|------|----------|----------|--------|------|------|----------|---------------|---------------|--------------|---------------|---------------|----------|------|-----|
|        | 542  | 4059     | 176.9924 | 201712 | 17   | 2006 | 4866     | 240000.0      | 194236.642239 | 62           | 51482.392380  | 135377.500000 | 492000.0 | 11   | 2   |
|        | 622  | 4059     | 138.3156 | 201808 | 4    | 2006 | 4866     | 260000.0      | 194236.642239 | 62           | 64740.609694  | 162540.491311 | 500000.0 | 12   | 2   |
|        | 654  | 6238     | 142.7700 | 201711 | 1    | 1978 | 1598     | 143000.0      | 143337.500000 | 62           | 51482.392380  | 135377.500000 | 492000.0 | 39   | 2   |
|        | 659  | 6238     | 142.7700 | 201801 | 6    | 1978 | 1598     | 149000.0      | 143337.500000 | 62           | 64740.609694  | 162540.491311 | 500000.0 | 40   | 2   |
|        | 661  | 6238     | 142.7700 | 201801 | 6    | 1978 | 1598     | 158000.0      | 143337.500000 | 62           | 64740.609694  | 162540.491311 | 500000.0 | 40   | 2   |
|        | ...  | ...      | ...      | ...    | ...  | ...  | ...      | ...           | ...           | ...          | ...           | ...           | ...      | ...  | ... |
| 807664 | 1292 | 155.3800 | 201505   | 18     | 2012 | 5489 | 188000.0 | 227025.692695 | 111           | 61736.026445 | 138614.054126 | 360000.0      | 3        | 2    |     |
| 807669 | 1292 | 155.3800 | 201506   | 3      | 2012 | 5489 | 184000.0 | 227025.692695 | 111           | 61736.026445 | 138614.054126 | 360000.0      | 3        | 2    |     |
| 807674 | 1292 | 155.9500 | 201508   | 19     | 2012 | 5489 | 204000.0 | 227025.692695 | 111           | 61736.026445 | 138614.054126 | 360000.0      | 3        | 2    |     |
| 807676 | 1292 | 155.9500 | 201509   | 4      | 2012 | 5489 | 188000.0 | 227025.692695 | 111           | 61736.026445 | 138614.054126 | 360000.0      | 3        | 2    |     |
| 807677 | 1292 | 155.9500 | 201509   | 23     | 2012 | 5489 | 197000.0 | 227025.692695 | 111           | 61736.026445 | 138614.054126 | 360000.0      | 3        | 2    |     |

5601 rows x 14 columns

## VVIP 저격 모델 만들기. (학습 후)

```
wt=train_val_exp[train_val_exp['아파트명']=='한남더힐']  
  
sns.barplot(data=wt, x=wt.index, y='err')
```

<Axes: ylabel='err'>



# 각자 모델 비교해보기

```
df_test[cols].sort_values(by='target_han',ascending=False).head(30)
```

|      | 아파트명      | 전용면적(m²) | 계약년월   | 층  | 동     | target_un | target_han | target_sh |
|------|-----------|----------|--------|----|-------|-----------|------------|-----------|
| 8593 | 파크크한남     | 268.670  | 202308 | 4  | 한남동   | 1152078   | 1297434    | 1267351   |
| 8285 | 한남더힐      | 240.305  | 202308 | -1 | 한남동   | 993837    | 1083572    | 1044829   |
| 8591 | 르가든더메인한남  | 239.670  | 202307 | 4  | 한남동   | 711644    | 971661     | 702758    |
| 8590 | 르가든더메인한남  | 237.520  | 202307 | 3  | 한남동   | 642381    | 917480     | 610231    |
| 8282 | 한남더힐      | 208.478  | 202307 | 2  | 한남동   | 789562    | 866517     | 794579    |
| 6059 | 아크로서울포레스트 | 198.219  | 202308 | 41 | 성수동1가 | 846379    | 850942     | 656164    |
| 6056 | 아크로서울포레스트 | 198.219  | 202307 | 40 | 성수동1가 | 837852    | 846514     | 651657    |
| 6042 | 갤러리아포레    | 241.930  | 202308 | 27 | 성수동1가 | 848102    | 839621     | 705119    |
| 5009 | 반포자이      | 244.628  | 202307 | 23 | 반포동   | 735900    | 757175     | 691986    |
| 5013 | 반포자이      | 244.972  | 202308 | 23 | 반포동   | 735900    | 757175     | 720339    |
| 6041 | 갤러리아포레    | 217.860  | 202308 | 36 | 성수동1가 | 688067    | 745028     | 633242    |
| 255  | 타워팰리스1    | 301.470  | 202307 | 55 | 도곡동   | 694526    | 737833     | 702344    |
| 256  | 타워팰리스1    | 244.660  | 202308 | 50 | 도곡동   | 674884    | 691285     | 603446    |
| 5024 | 반포자이      | 244.543  | 202309 | 2  | 반포동   | 603144    | 679902     | 625505    |
| 262  | 타워팰리스2    | 243.140  | 202308 | 53 | 도곡동   | 633759    | 642771     | 574320    |
| 521  | 신현대11차    | 183.410  | 202307 | 9  | 압구정동  | 626558    | 637810     | 611848    |
| 5025 | 반포자이      | 216.490  | 202309 | 17 | 반포동   | 602777    | 627885     | 602858    |
| 6058 | 아크로서울포레스트 | 159.603  | 202307 | 23 | 성수동1가 | 615435    | 626707     | 619230    |
| 264  | 타워팰리스3    | 244.224  | 202308 | 62 | 도곡동   | 611005    | 615421     | 541629    |
| 4965 | 래미안퍼스티지   | 168.650  | 202307 | 28 | 반포동   | 598135    | 613999     | 570675    |
| 4983 | 래미안퍼스티지   | 169.310  | 202309 | 12 | 반포동   | 592766    | 611030     | 528186    |
| 5861 | 트라움하우스3   | 273.860  | 202307 | 8  | 서초동   | 613004    | 609686     | 576474    |
| 4974 | 래미안퍼스티지   | 168.650  | 202308 | 21 | 반포동   | 595660    | 604027     | 537140    |
| 5495 | 반포주공1단지   | 140.130  | 202307 | 4  | 반포동   | 565762    | 599814     | 622219    |



# 각자 모델 비교해보기.

|  | 번지      | 아파트명               | 전용면적<br>(㎡) | 계약년<br>월 | 층  | 건축년<br>도 | 도로명        | target   | is_test | 평균공시가격        | 동      | 계약년과동에따른<br>std | 계약년과동에따른<br>mean | 계약년과동에따른<br>max | yrs_diff | Close_delay120d | e |
|--|---------|--------------------|-------------|----------|----|----------|------------|----------|---------|---------------|--------|-----------------|------------------|-----------------|----------|-----------------|---|
|  | 216384  | 685-700 아크로서울 포레스트 | 159.603     | 202303   | 27 | 2020     | 왕십리로 83-21 | 620000.0 | 0       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2261.925        |   |
|  | 216385  | 685-700 아크로서울 포레스트 | 159.603     | 202304   | 45 | 2020     | 왕십리로 83-21 | 675000.0 | 0       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2288.780        |   |
|  | 216386  | 685-700 아크로서울 포레스트 | 164.856     | 202304   | 3  | 2020     | 왕십리로 83-21 | 575000.0 | 0       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2348.430        |   |
|  | 233419  | 685-700 아크로서울 포레스트 | 165.182     | 202204   | 6  | 2020     | 왕십리로 83-21 | 620000.0 | 0       | 423578.928571 | 성수동 1가 | 262281.567769   | 300048.809524    | 1300000.0       | 2        | 3029.040        |   |
|  | 276041  | 685-700 아크로서울 포레스트 | 159.603     | 202102   | 43 | 2020     | 왕십리로 83-21 | 560000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2432.350        |   |
|  | 276042  | 685-700 아크로서울 포레스트 | 159.603     | 202102   | 43 | 2020     | 왕십리로 83-21 | 560000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2432.350        |   |
|  | 276044  | 685-700 아크로서울 포레스트 | 159.603     | 202103   | 34 | 2020     | 왕십리로 83-21 | 530000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2406.170        |   |
|  | 276045  | 685-700 아크로서울 포레스트 | 159.603     | 202103   | 34 | 2020     | 왕십리로 83-21 | 530000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2406.170        |   |
|  | 276046  | 685-700 아크로서울 포레스트 | 97.485      | 202103   | 7  | 2020     | 왕십리로 83-21 | 280000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2332.590        |   |
|  | 276047  | 685-700 아크로서울 포레스트 | 198.219     | 202104   | 38 | 2020     | 왕십리로 83-21 | 552000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2345.260        |   |
|  | 276048  | 685-700 아크로서울 포레스트 | 200.749     | 202105   | 16 | 2020     | 왕십리로 83-21 | 580000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2543.030        |   |
|  | 276049  | 685-700 아크로서울 포레스트 | 202.620     | 202105   | 13 | 2020     | 왕십리로 83-21 | 590000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2609.465        |   |
|  | 276050  | 685-700 아크로서울 포레스트 | 200.749     | 202105   | 11 | 2020     | 왕십리로 83-21 | 600000.0 | 0       | 423578.928571 | 성수동 1가 | 156469.247475   | 234488.855422    | 845000.0        | 1        | 2675.900        |   |
|  | 1124878 | 685-700 아크로서울 포레스트 | 198.219     | 202307   | 40 | 2020     | 왕십리로 83-21 | NaN      | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2359.530        |   |
|  | 1124879 | 685-700 아크로서울 포레스트 | 95.810      | 202307   | 4  | 2020     | 왕십리로 83-21 | NaN      | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2399.860        |   |
|  | 1124880 | 685-700 아크로서울 포레스트 | 159.603     | 202307   | 23 | 2020     | 왕십리로 83-21 | NaN      | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2419.475        |   |
|  | 1124881 | 685-700 아크로서울 포레스트 | 198.219     | 202308   | 41 | 2020     | 왕십리로 83-21 | NaN      | 1       | 423578.928571 | 성수동 1가 | 131574.802869   | 185585.714286    | 675000.0        | 3        | 2439.090        |   |

# 결과 평균내기.

```
df_test['mean'] = (df_test['target_han'] + df_test['target_un'] + df_test['target_sh']) / 3
```

Python

```
df_test
```

Python

|   | index   | 번지    | 아파트명    | 전용면적 (㎡) | 계약년월   | 층  | 건축년도 | 도로명    | target | 평균공시가격        | ... | 계약년과<br>동에따른<br>max | yrs_diff | Close_delay120d | exp2 | target1 | target_un | target_han | target_sh | target_best | mean          |
|---|---------|-------|---------|----------|--------|----|------|--------|--------|---------------|-----|---------------------|----------|-----------------|------|---------|-----------|------------|-----------|-------------|---------------|
| 0 | 1118822 | 658-1 | 개포6차우성  | 79.9700  | 202307 | 5  | 1987 | 안주로3   | NaN    | 120960.000000 | ... | 370000.0            | 36       | 2449.800000     | 0    | 180258  | 180258    | 183984     | 193610    | 187124      | 185950.666667 |
| 1 | 1118823 | 651-1 | 개포더샵트리에 | 108.2017 | 202308 | 10 | 2021 | 개포로311 | NaN    | 155775.000000 | ... | 370000.0            | 2        | 2465.100000     | 0    | 262650  | 262650    | 257254     | 262267    | 289902      | 260723.666667 |
| 2 | 1118824 | 652   | 개포우성3차  | 161.0000 | 202307 | 15 | 1984 | 개포로307 | NaN    | 182537.777778 | ... | 370000.0            | 39       | 2480.400000     | 1    | 300967  | 300967    | 305897     | 275190    | 305133      | 294018.000000 |

## 결과 평균내기.

```
han_ = final[final['전용면적(m²)']>=240]
han_
```

|      | 아파트명    | 전용면적(m²) | 계약년월   | 층  | 동     | target_un | target_han | target_sh |
|------|---------|----------|--------|----|-------|-----------|------------|-----------|
| 8593 | 파크한남    | 268.670  | 202308 | 4  | 한남동   | 1152078   | 1297434    | 1267351   |
| 8285 | 한남더힐    | 240.305  | 202308 | -1 | 한남동   | 993837    | 1083572    | 1044829   |
| 6042 | 갤러리아포레  | 241.930  | 202308 | 27 | 성수동1가 | 848102    | 839621     | 705119    |
| 5009 | 반포자이    | 244.628  | 202307 | 23 | 반포동   | 735900    | 757175     | 691986    |
| 5013 | 반포자이    | 244.972  | 202308 | 23 | 반포동   | 735900    | 757175     | 720339    |
| 255  | 타워팰리스1  | 301.470  | 202307 | 55 | 도곡동   | 694526    | 737833     | 702344    |
| 256  | 타워팰리스1  | 244.660  | 202308 | 50 | 도곡동   | 674884    | 691285     | 603446    |
| 5024 | 반포자이    | 244.543  | 202309 | 2  | 반포동   | 603144    | 679902     | 625505    |
| 262  | 타워팰리스2  | 243.140  | 202308 | 53 | 도곡동   | 633759    | 642771     | 574320    |
| 264  | 타워팰리스3  | 244.224  | 202308 | 62 | 도곡동   | 611005    | 615421     | 541629    |
| 5861 | 트라움하우스3 | 273.860  | 202307 | 8  | 서초동   | 613004    | 609686     | 576474    |

```
sh_ = final[final['전용면적(m²)']<240]
```

sh\_

|      | 아파트명      | 전용면적(m²) | 계약년월   | 층  | 동     | target_un | target_han | target_sh |
|------|-----------|----------|--------|----|-------|-----------|------------|-----------|
| 8591 | 르가든더메인한남  | 239.670  | 202307 | 4  | 한남동   | 711644    | 971661     | 702758    |
| 8590 | 르가든더메인한남  | 237.520  | 202307 | 3  | 한남동   | 642381    | 917480     | 610231    |
| 8282 | 한남더힐      | 208.478  | 202307 | 2  | 한남동   | 789562    | 866517     | 794579    |
| 6059 | 아크로서울포레스트 | 198.219  | 202308 | 41 | 성수동1가 | 846379    | 850942     | 656164    |
| 6056 | 아크로서울포레스트 | 198.219  | 202307 | 40 | 성수동1가 | 837852    | 846514     | 651657    |
| 6041 | 갤러리아포레    | 217.860  | 202308 | 36 | 성수동1가 | 688067    | 745028     | 633242    |
| 521  | 신현대11차    | 183.410  | 202307 | 9  | 압구정동  | 626558    | 637810     | 611848    |
| 5025 | 반포자이      | 216.490  | 202309 | 17 | 반포동   | 602777    | 627885     | 602858    |
| 6058 | 아크로서울포레스트 | 159.603  | 202307 | 23 | 성수동1가 | 615435    | 626707     | 619230    |
| 4965 | 래미안퍼스티지   | 168.650  | 202307 | 28 | 반포동   | 598135    | 613999     | 570675    |
| 4983 | 래미안퍼스티지   | 169.310  | 202309 | 12 | 반포동   | 592766    | 611030     | 528186    |



05

# 결과

## 최종 순위 및 평가지표 결과

### Public LB

|   |       |           |            |    |     |
|---|-------|-----------|------------|----|-----|
| 1 | 3조 별자 | H 언선 이 민규 | 14691.1409 | 86 | 2h  |
| 2 | 7조    | 주형 송연     | 14793.8293 | 49 | 14h |
| 3 | ML_8조 | 종완 소미     | 15234.6503 | 53 | 1d  |
| 4 | ML_6조 | 경태 H 도후   | 16621.8992 | 58 | 2d  |
| 5 | ML_4조 | 수영 현지     | 16714.4593 | 57 | 1h  |

### Private LB

|   |       |           |            |    |    |
|---|-------|-----------|------------|----|----|
| 1 | ML_1조 | 재민 ha 병원  | 10988.6858 | 60 | 3h |
| 2 | 7조    | 주형 송연     | 12206.3539 | 49 | 1h |
| 3 | 3조 별자 | H 언선 이 민규 | 12614.1573 | 86 | 3h |
| 4 | ML_8조 | 종완 소미     | 12713.1653 | 53 | 1d |
| 5 | ML_6조 | 경태 H 도후   | 13331.0750 | 58 | 7h |

06

# 그룹 스터디 진행 소감

## 그룹 스터디 진행 소감

이승현

너무 재밌었습니다. 나의 장점과 단점을 찾을 수 있던 좋은 계기 였습니다.  
단점은 모델링과 이론이 약하니까 이 부분을 채워 놓자.

박언선

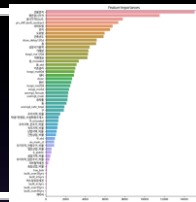
해볼 수 있는걸 다 해보고 재미있었습니다. 좋은 팀원들과 함께해서 많이 배웠습니다.

이광우

###

한민규

팀원분들 덕에 많이 웃었고, 부족한 부분을 채울 수 있었습니다.  
(퍼블릭 스코어를 맹신하고 혜윤님의 도메인 지식과 빼곡한 features 를 반영하지 못한게 아쉽습니다.)



정혜윤

티키타카란 이런 것이구나!

# Q&A

---

감사합니다.

—