



대우건설 Project 부동산 가격 예측

경기도 시도별 신축 아파트 평당가 예측 모델

PREFLAT

팀원 소개 & INDEX

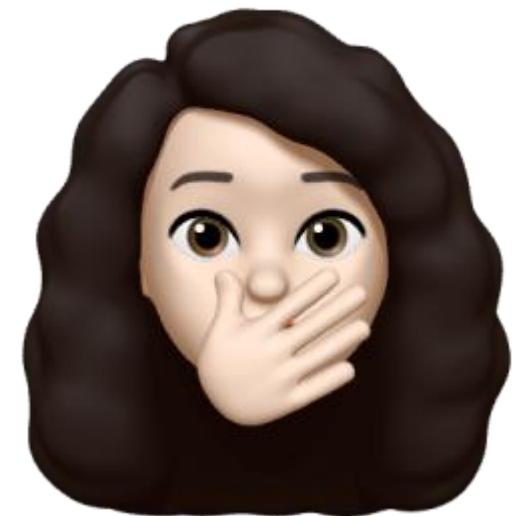
<조장>



김지호



성치웅



강유민



남윤종



이난영



주제 선정 배경



데이터 소개



가격 예측 모델



활용방안

미래 부동산 가격 예측의 어려움

집값 떨어진다던 한국부동산원... 올해 전망 않기로 결정 2021.01.05

집값 전망 눈감은 한국부동산원, 정부 통계기관 맞나 2022.01.21

연기 거듭되는 부동산원 하반기 전망 발표... 속사정은? 2022.08.15

코로나19 확산, 전쟁 등 예상치 못한 변동 요인으로
인해 부동산 가격 등락 예측에 부담 가중

PREFLAT

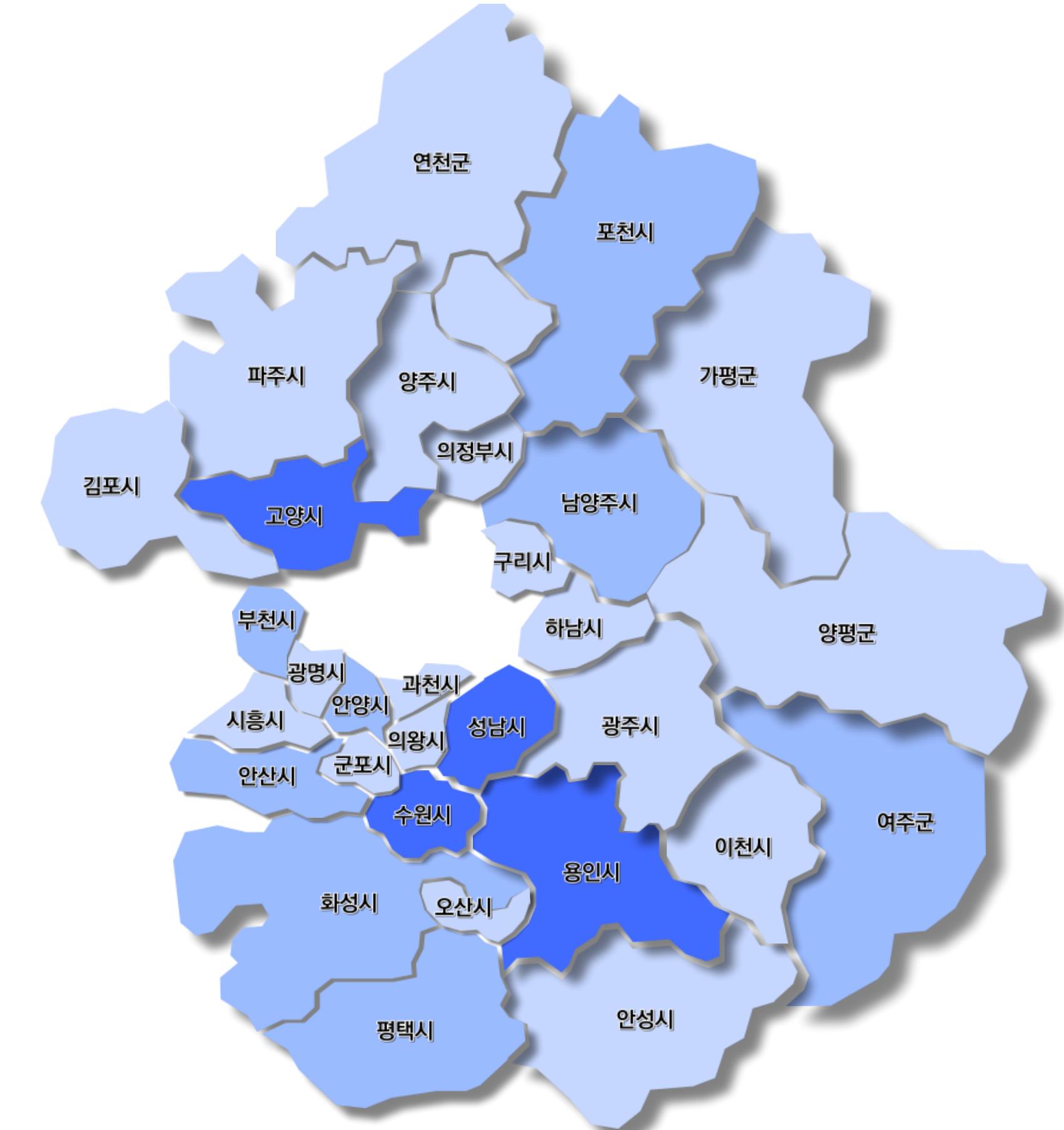
01. 주제 선정

경기도를 선택한 이유

경기도 5년 이내
신축 아파트 평당가 예측



지리적으로 서울특별시와 인천광역시를 둘러싸고 있으며
해당 두 개 특별, 광역시와 함께 대한민국의 수도권을
형성하고 있는 광역자치단체

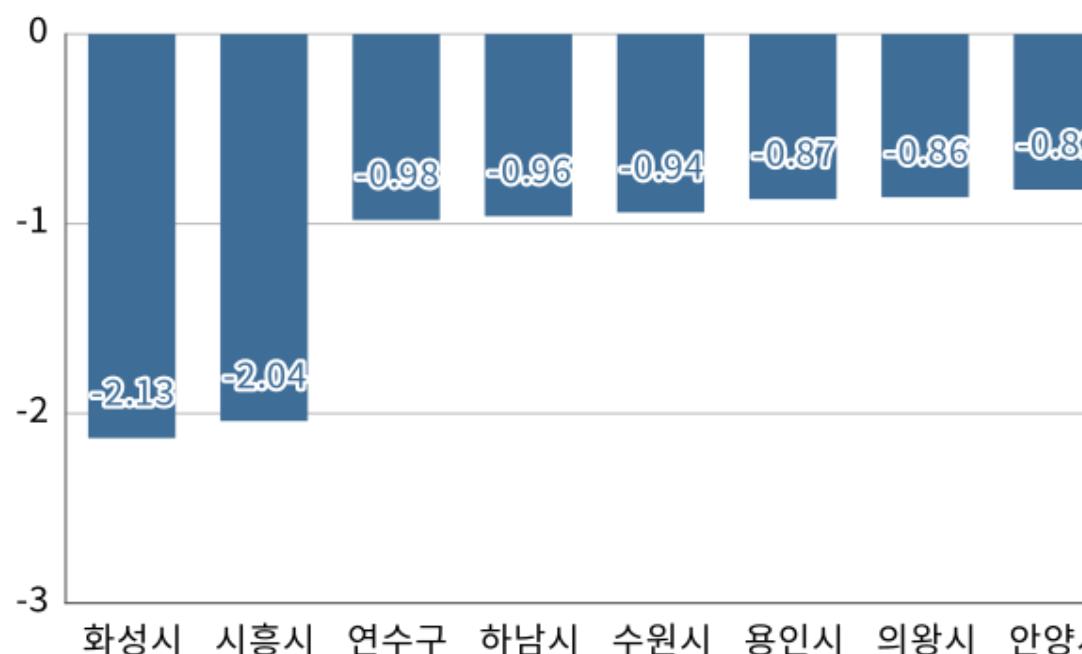


PREFLAT

01. 주제 선정

경기도를 선택한 이유

현상1. 다양한 요인에 따른 경기도 주요도시 아파트 하락세



■ 2022년 들어 집값 하락률 높은 지역 순위

(단위 : %, 자료 : 한국부동산원)



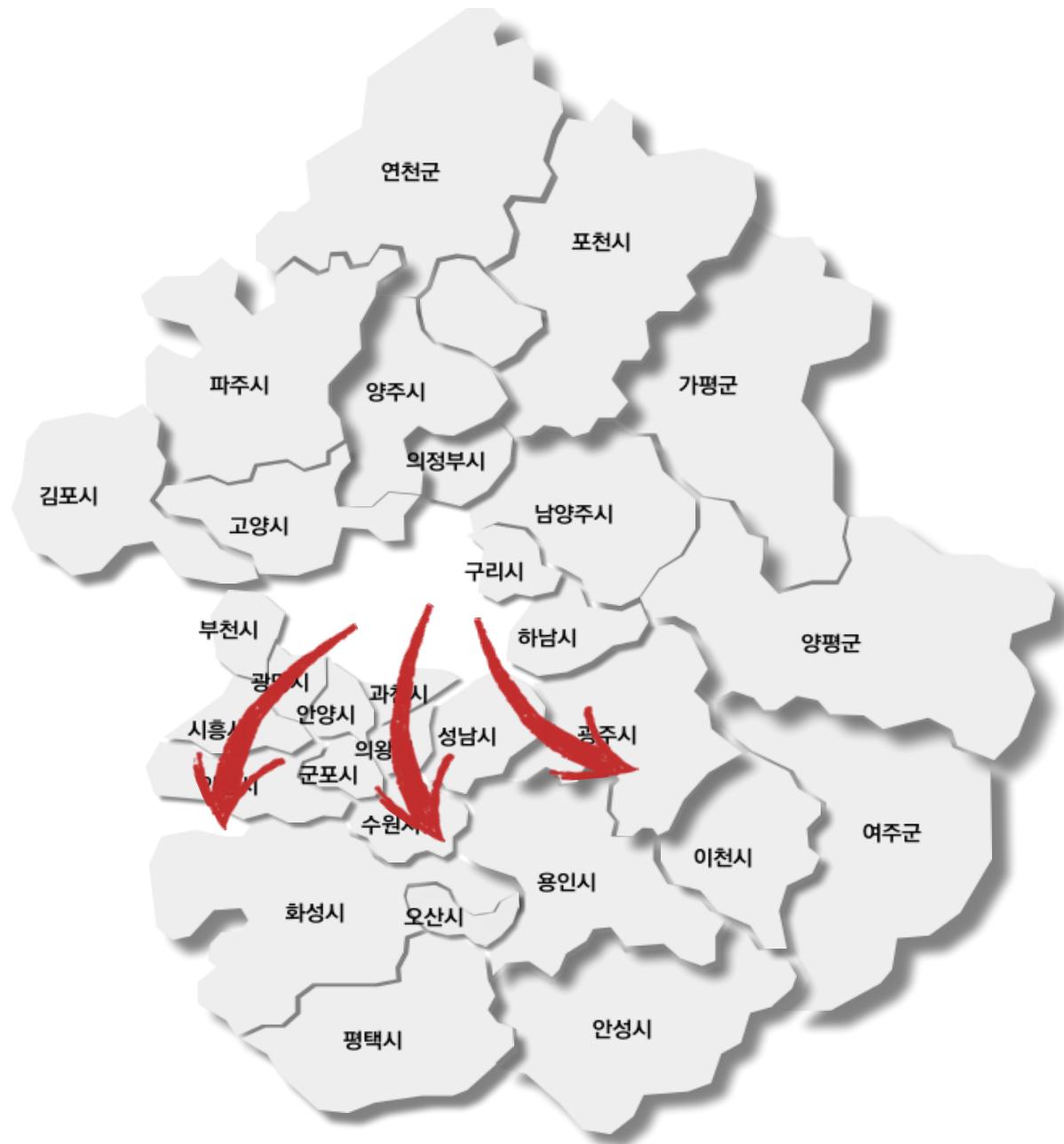
의왕, 용인, 화성, 수원 등 경기 남부권 집값 하락세...
금리 인상에 따른 거래 절벽이 심화하면서 매수세 하락

특히, 수도권광역급행철도(GTX)-신분당선 등 철도 교통 호재로 가격이 급등했던 경기 남부권 집값 하락 추세

PREFLAT

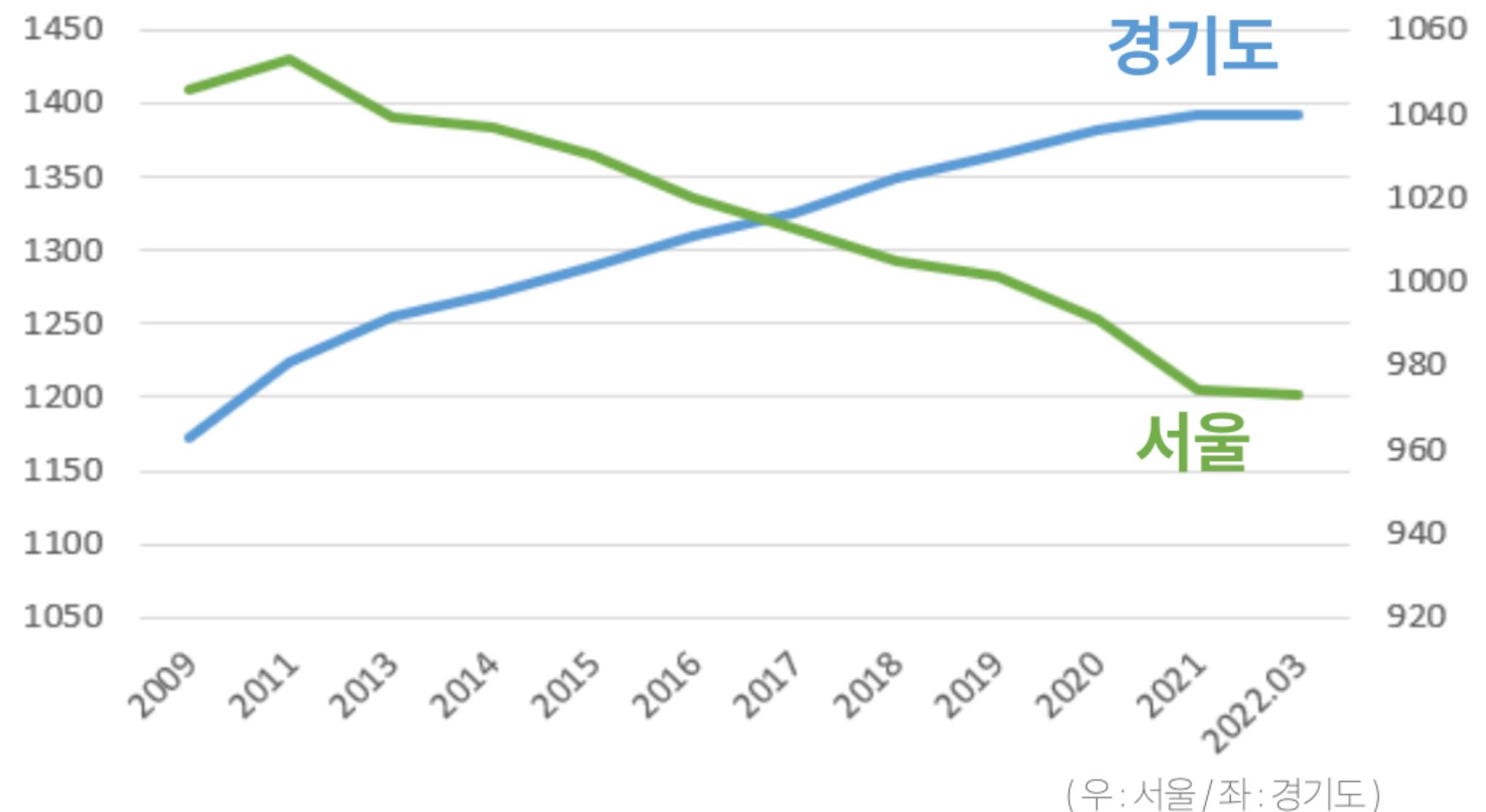
01. 주제 선정

경기도를 선택한 이유



현상2. 서울에서 경기도로의
인구유출

늘어나는 경기도 인구, 줄어드는 서울 인구



PREFLAT

01. 주제 선정

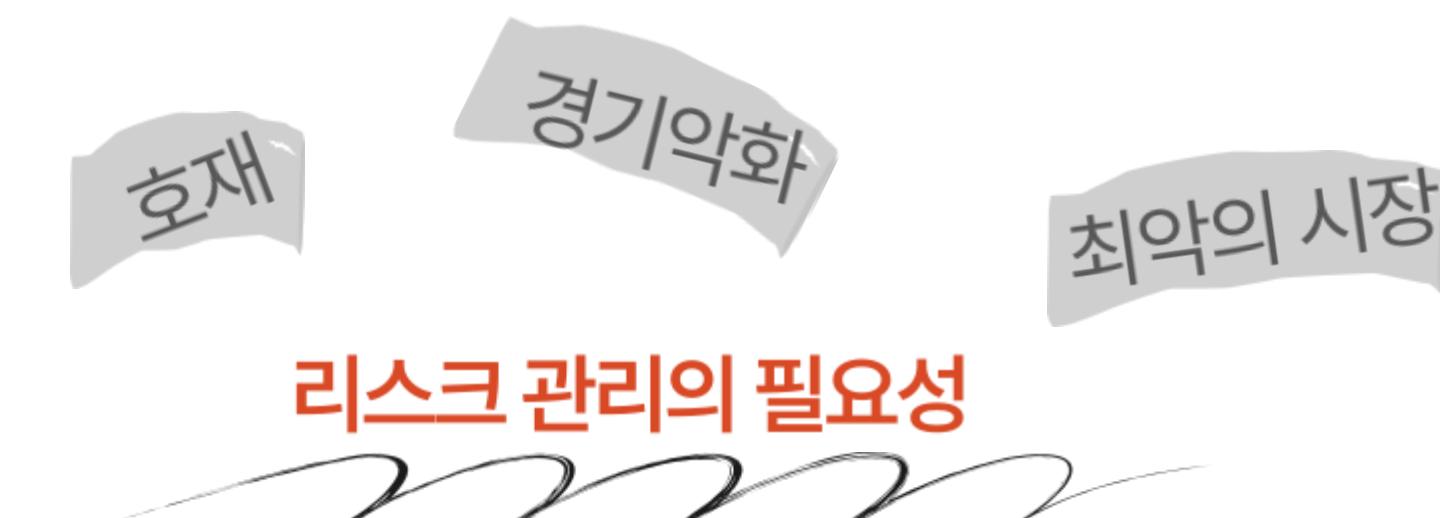
경기도를 선택한 이유



현상3. 가파르게 늘어나는 경기도 아파트 입주물량

입주 쏟아지는 경기도...
집값 하락세 가팔라지나 - 2022/08/17

"새 아파트 전세 사실 분 없나요"...
성남 수원 '입주폭탄'에 집주인 잠 못든다
-2022/08/22



PREFLAT



01. 변수 소개

종속 변수
: 5년 이내 각 시별 평당가

독립변수



| 연령평균 | 가구수 | 사설학원수 | 학업 성취도 | 매매회전율 | 미분양수 | 입주물량 | 강남접근성 점수1 | 강남접근성 점수2 | GTX | 주택가격 전망CSI |
|--------------|---------------|-------|-----------------|-----------|--------|--------|-----------|-----------|------------|--------------|
| 아파트 연령 평균 | 2인 ~4인 가구수 | 시별 개수 | 국가 기준 학업 성취도 | 매매량 / 세대수 | 미분양 | 입주 물량 | 지하철 기준 | 자동차 기준 | 예정 도시 | 소비자 심리 지수 |
| 신도시 | LTV | UIG | 출퇴근량 | 대기업수 | 매수우위지수 | 가계대출금리 | 미국금리종가 | EPU | 선도 50지수 | |
| 신도시 유무 | 규제 유무 | 선행 지수 | 서울 기준 | 대기업 | 소비자 지수 | 금리지표 | 금리 지표 | 불확실성지수 | 상위 50개 아파트 | |

PREFLAT

01. 변수 소개



01 / 학군

- 사설 학원수
- 학업 성취도

02 / 일자리

- 대기업 수

03 / 교통

- 강남 접근성 점수1 - 지하철 기준
- 강남 접근성 점수2 - 자동차 기준
- 출퇴근량

04 / 호재성

- GTX
- 신도시

05 / 금리

- 가계 대출 금리
- 미국 금리 종가

06 / 정책

- UIG
- EPU
- LTV

07 / 심리 지수

- 주택가격 전망 CSI
- 매수우위지수
- KB 선도 50 지수

PREFLAT

01. 변수 소개

독립 변수 추가 소개



01 / 강남 접근성 점수 1 (지하철 기준)

- 30분 이내 5점
- 50분 이내 4점
- 1시간 이내 3점
- 1시간 이상 2점
- 역세권 無 1점

02 / 강남 접근성 점수 2 (자동차 기준)

- 40분 이내 3점
- 60분 이내 2점
- 60분 이상 1점

03 / 신도시

| 금리 | 신도시 | 신도시 有 | 신도시 無 |
|------|-----|-------|-------|
| 2-3% | 5 | 4 | |
| 4-5% | 3 | 2 | |
| 6-7% | 1 | 0 | |

PREFLAT

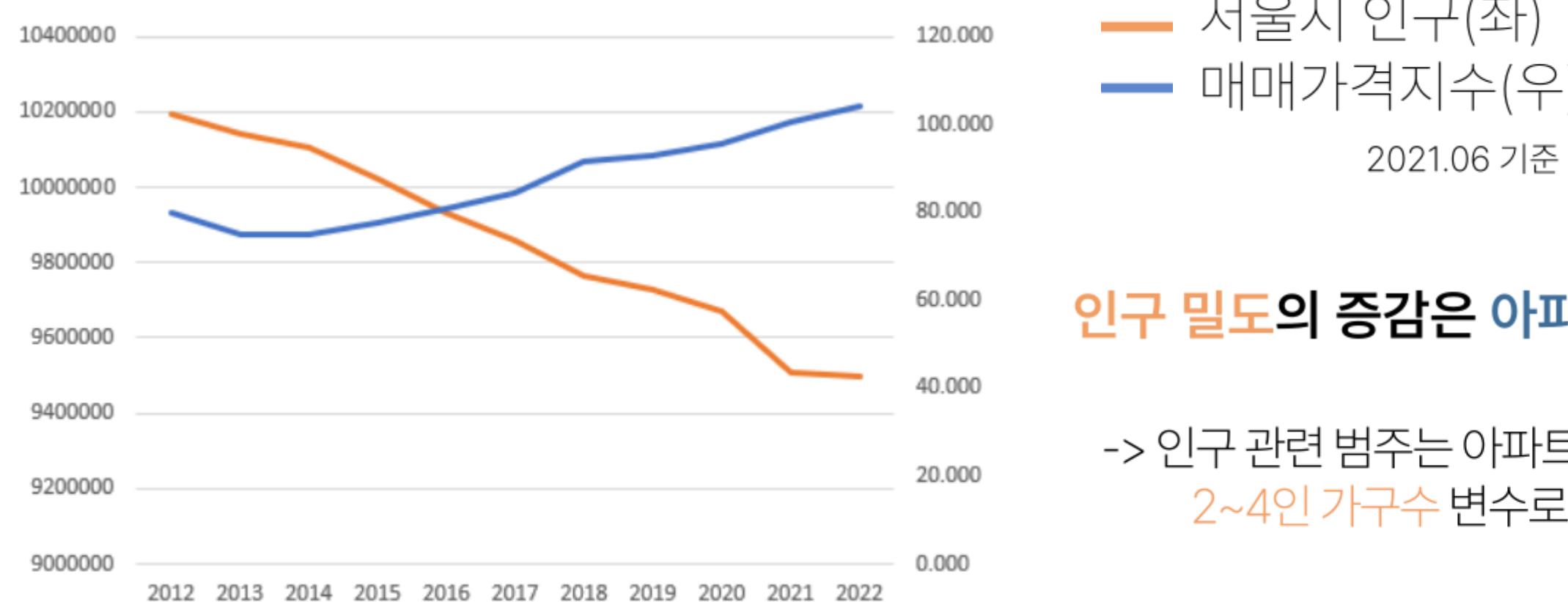


01. 변수 소개

아파트 가격 변동에 영향을 주는 요인

01 / 인구와 주택의 관계성

인구 밀도가 높아질 수록 아파트의 가격이 증가한다는
인과관계를 단언할 수 없음



서울시 인구(좌)
매매가격지수(우)

2021.06 기준

인구 밀도의 증감은 아파트 가격에 큰 영향을 끼치지 않음

-> 인구 관련 범주는 아파트 구매력이 높은
2~4인 가구수 변수로 대체

PREFLAT

01. 변수 소개

아파트 가격 변동에 영향을 주는 요인



02 / 학군

- 아이들의 교육 환경을 조성하기 위해 학군이 좋은 아파트의 선호도가 증가
- > 학업 성취도, 사설학원수 변수

03 / 일자리

- 청년취업율, 경제활동인구는 아파트 구매자 표본의 대표성을 띄기 어려움
- > 대기업 수

PREFLAT

01. 변수 소개

아파트 가격 변동에 영향을 주는 요인



04 / 금리

- **미국 금리 종가**: 국회예산정책처에 따르면, 우리나라 집값과 금리는 미국 10년물 금리를 후행
- **가계 대출 금리**: 3개의 금리 중 가장 대표성을 띠며 상관성이 높은 금리를 변수로 사용

05 / 정책

- **UIG**: 뉴욕 연방준비은행에서 발표하는 향후 인플레이션을 추정해볼 수 있는 데이터
(Underlying Inflation Gauge)
- **EPU**: 경제 불확실성 지수로, 뉴스기사의 텍스트 데이터를 분석해 경제 흐름을 파악
(Economic Policy Uncertainty)
- **LTV**: 주택담보대출 비율 (Loan to Value)

PREFLAT

01. 변수 소개

아파트 가격 변동에 영향을 주는 요인

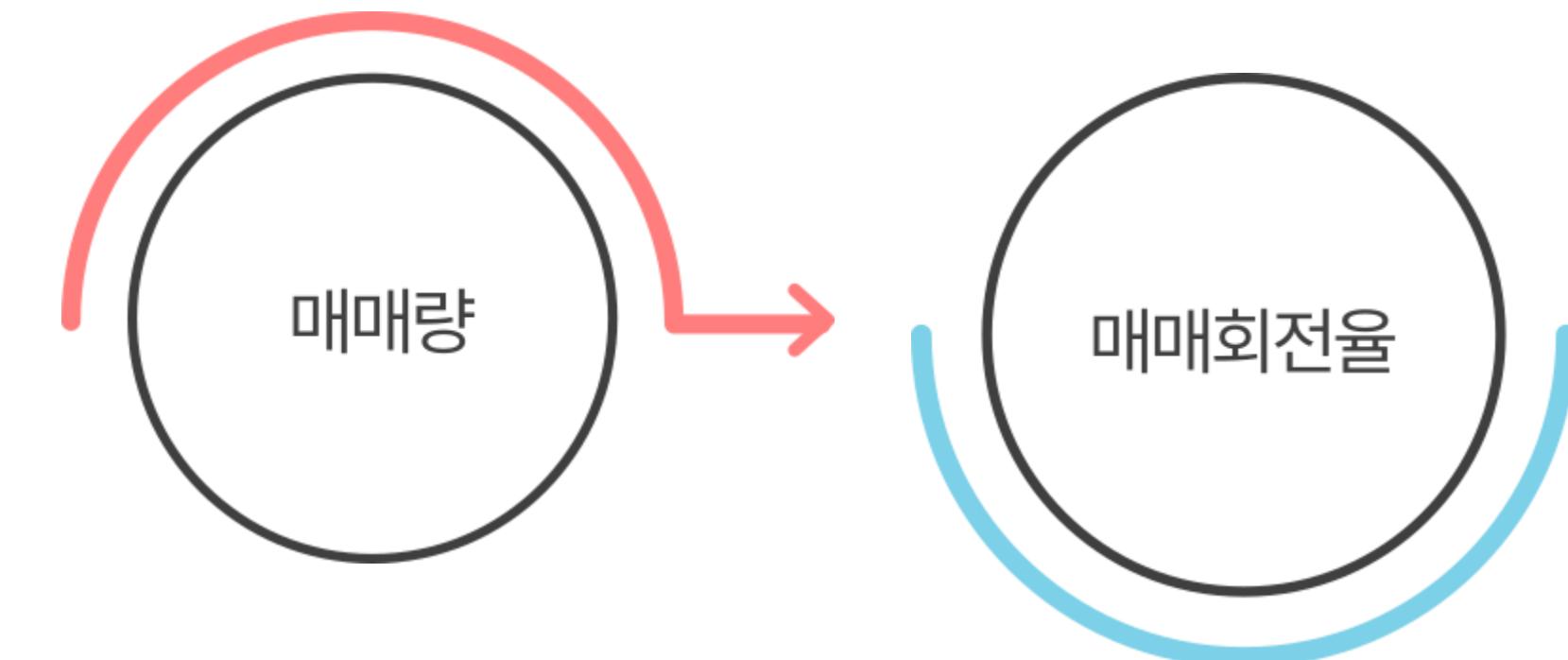


06 / 심리

- KB 선도 50 지수: 시가총액 상위 50개 단지를 선정하여 시가총액의 지수와 변동률을 파악

07 / 매매회전율

- 매매량은 아파트 가구수가 많을 수록 높을 수 있기 때문에 아파트 매매의 회전율이 아파트의 인기를 나타내기에 더 적합한 지표로 작용 가격 변동에 대한 설명력이 더 높음



PREFLAT

02. 데이터 전처리



01 / NULL 값 처리

- 평당가 : 변화의 평균을 단계적으로 추가
 - 도시: 경기도 전체에서 연천, 포천, 과천을 제외한 28개 도시
 - 입주물량: 연데이터를 월별 데이터로 나누기 위해 12로 나눔
 - 학원수: 1년 동일하게 기입
 - 학업성취도: 2022년 8월 기준으로 동일하게 설정

-> 신축 아파트 부족으로 인해 거래량 결측치가 많아 분석에 의미가 저하될 것으로 예상

| | | | | | | | | |
|------|------|------|-----|------|------|------|------|--------|
| | | | | | | | | 16,332 |
| 1362 | 1362 | 1361 | ... | 1361 | 1361 | 1361 | 1361 | |

PREFLAT

02. 데이터 전처리



02 / 데이터 타입 변경

| 연령 평균 | int | GTX | category |
|-------|-----|-------------------------|----------|
| 가구수 | int | LTV | category |
| 사설학원수 | int | 강남 접근성 점수 1 (지하철 기준) | int |
| 입주물량 | int | 강남 접근성 점수 2 (자동차 기준) | int |

03 / one-hot encoding

피처 값의 유형에 따라 새로운 피처를 추가해 고유값에 해당하는 컬럼에 1을 표시하고 나머지 컬럼에는 0을 표시하는 방식으로, 문자열 값들을 숫자형으로 인코딩하는 작업

| 수원시 | GTX_0 | GTX_1 | LTV_0 | LTV_1 |
|---------|-------|-------|-------|-------|
| 2018.01 | 0 | 1 | 0 | 1 |
| 2021.01 | 1 | 0 | 1 | 0 |

PREFLAT

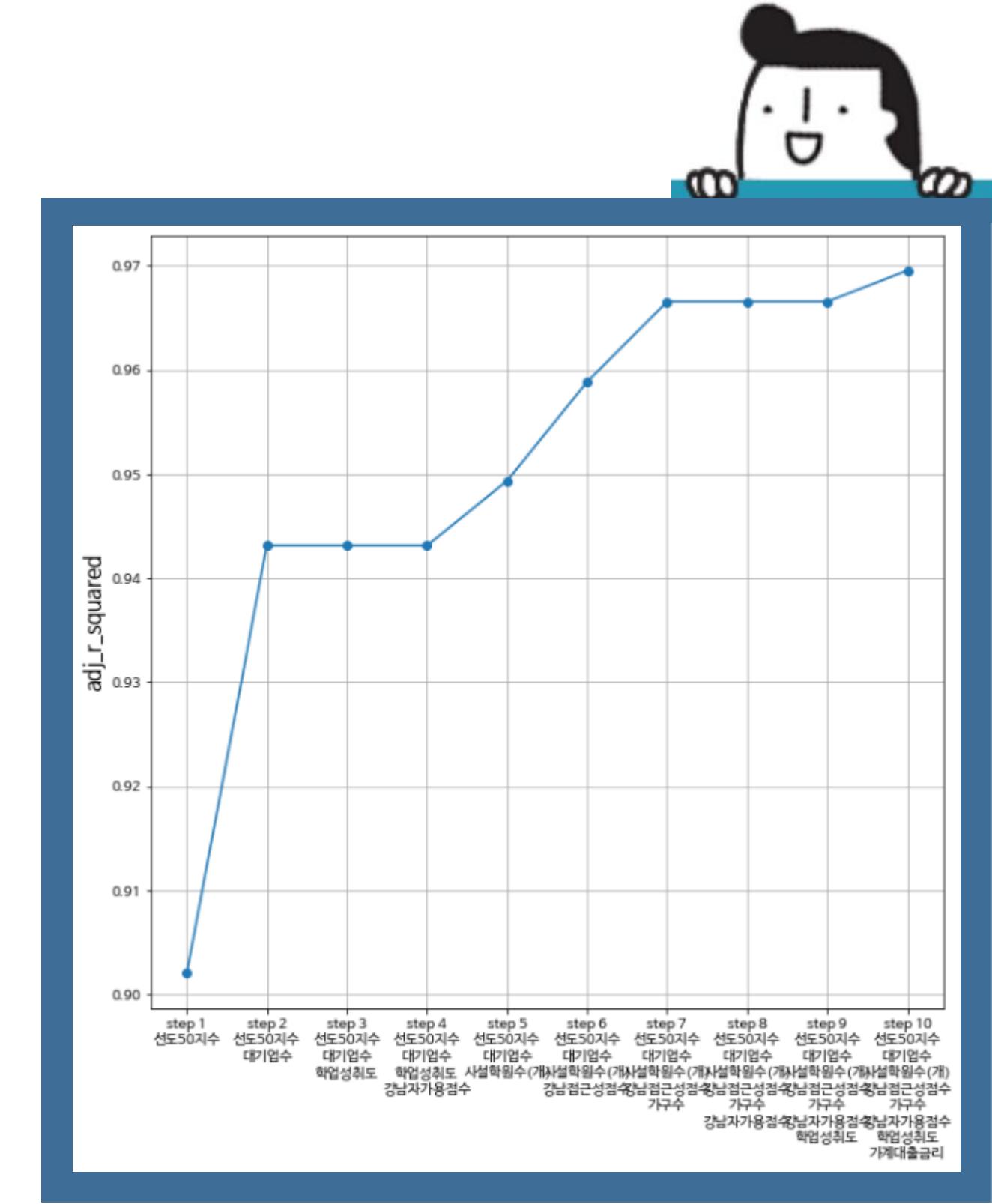
01. 변수 선택

도시별 변수 선택

변수 선택 기준 _ forward selection

| | |
|------|---|
| 김포시 | UIG, 미분양수, 연령평균, 입주물량, 출퇴근량, 가구수, 사설학원수 |
| 고양시 | 입주물량, epu, 연령평균, 미분양수 |
| 남양주시 | 입주물량, UIG, 매매회전율, 미분양수, epu |
| 하남시 | 입주물량, 사설학원수, epu, 가구수, 출퇴근량 |
| 성남시 | UIG, 가계대출금리, 매수우위지수, 사설학원수, 출퇴근량, 입주물량 |
| 수원시 | 사설학원수, 대기업수, 미분양수, epu, '입주물량' |
| 용인시 | 대기업수, 입주물량, 매수우위지수, 미분양수, 출퇴근량 |
| 화성시 | 대기업수, 미분양수, 미국금리종가, UIG, 출퇴근량, 입주물량 |

p-value 0.05 변수 3개 이하인 도시는 p-value 상향 조정



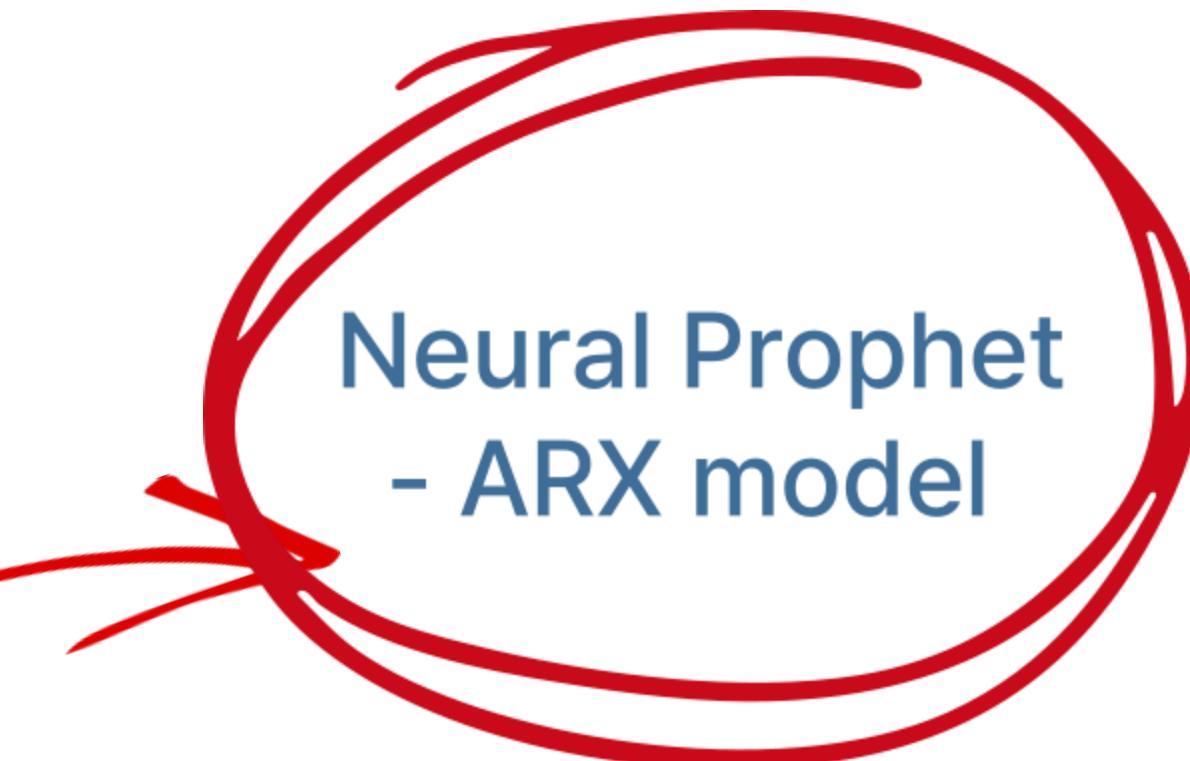
부동산 가격 예측 모델

PREFLAT MODEL

모델 선택



| MODEL | MAE |
|-------------------|--------|
| linear regression | 1357.9 |
| XGBoost | 160.4 |
| Random Forest | 134.6 |
| ARIMAX | 356.6 |
| Neural Prophet | 150. |



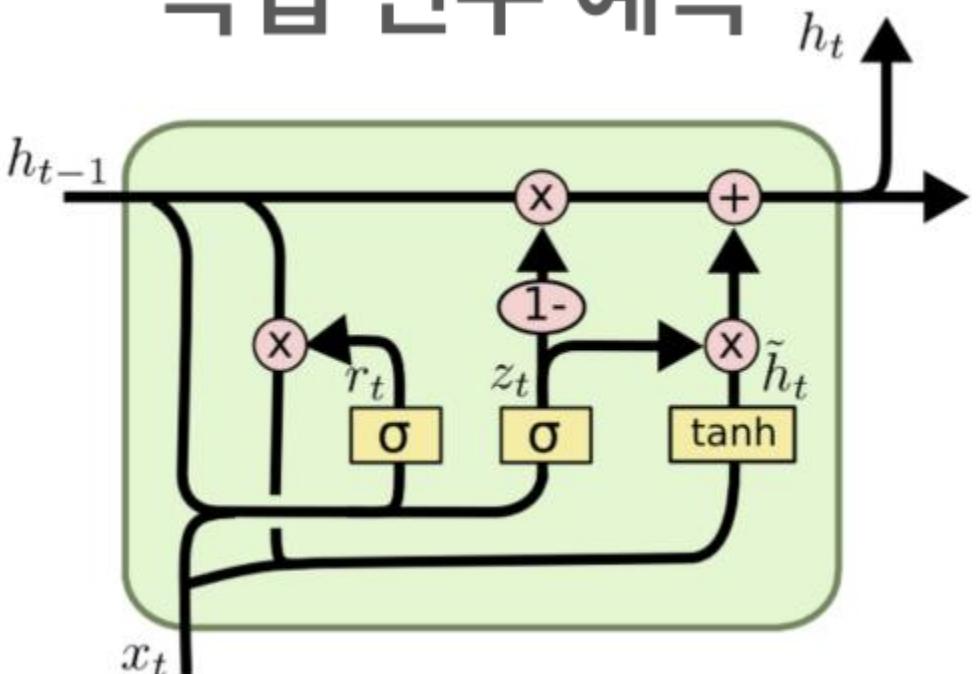
[남양주시, 2020-01-01 기준 test 값 예측]

PREFLAT

01. 모델링 과정

PREFLAT MODELING

LSTM을 통한
독립 변수 예측



LSTM은 장기 종속성을 학습하도록 고안된 RNN의 특수한 버전으로, RNN은 단순한 신경망 layer 한 층인 반면에, LSTM은 다수의 layer가 서로 상호작용 하도록 설계

PROPHET

예측 변수를 통한
미래 평당가 예측

변수 선택

2022년06월
~2024년 05월 예측값

부동산 가격 예측 모델

PREFLAT MODEL

시계열 분석이란,

주가, 에너지 사용량 등 시간적인 속성을
가지고 있는 데이터를 분석, 예측하는 것을 의미



비선형적인 요소 및 연속함수의 특
징 또는 패턴 변동성이 큰 부분을
반영하기 어려웠던 단점

prophet의 단점 보완
우수한 성능
해석 가능성
설정 및 사용 용의성

부동산 가격 예측 모델

PREFLAT MODEL



자기 회귀 과정에 독립 변수를 추가한 모델

2-lag 된 X 변수, Y 변수를 설명변수로 추가

ARX auto - regression + X

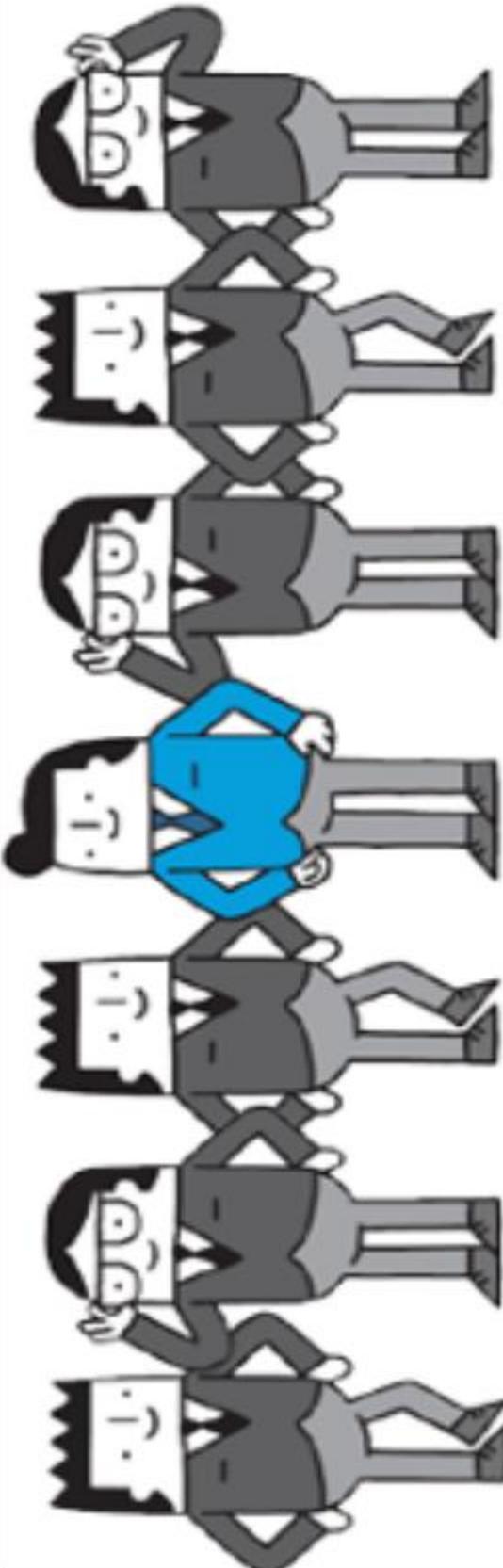
| ds | y | y_diff | y_diff2 |
|----|------|-----------|------------|
| | 입주물량 | 입주물량_diff | 입주물량_diff2 |

PREFLAT

03. neural prophet ARX - model

파라미터 설정

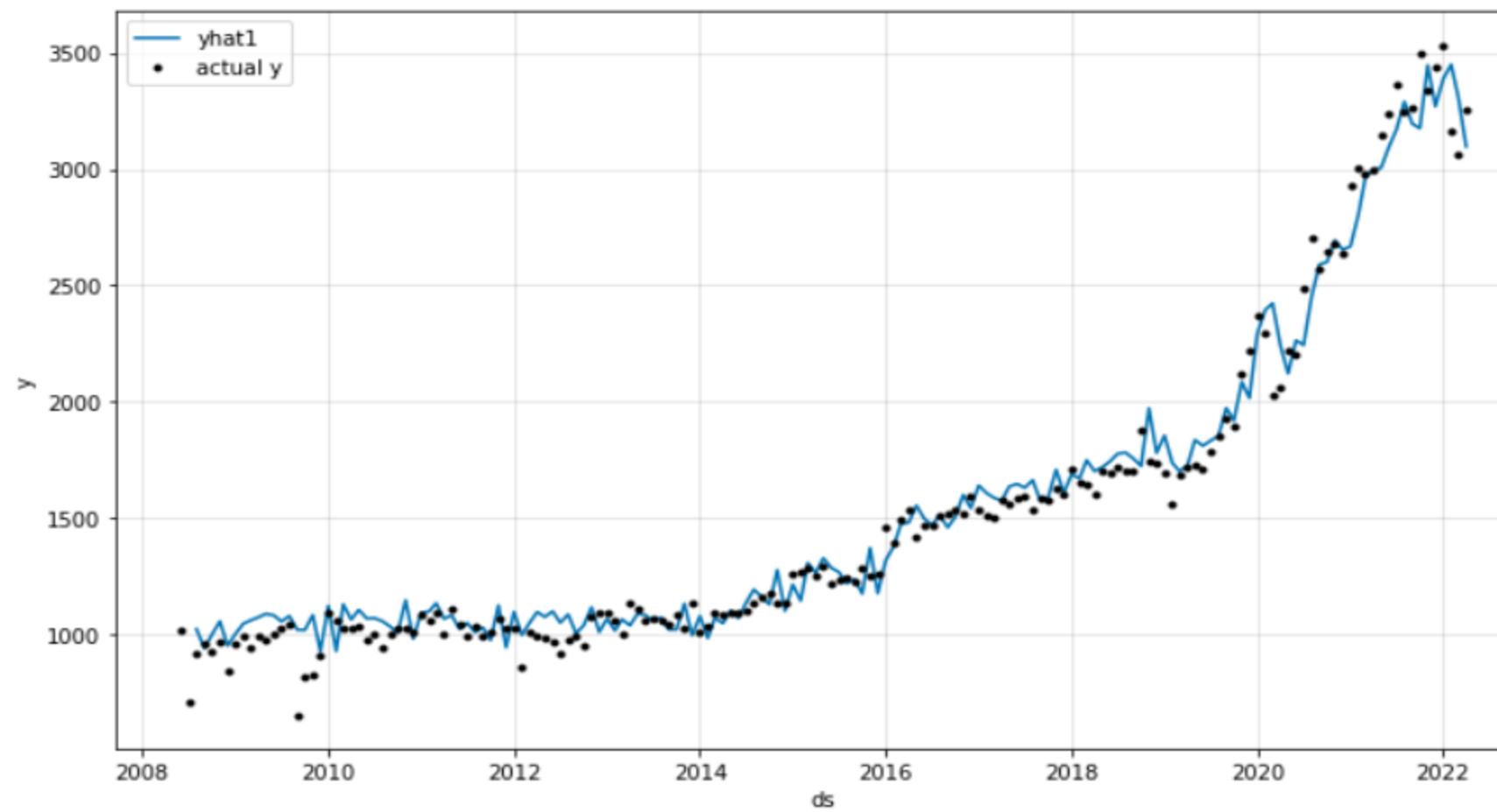
| | | | | | |
|-------------|---------|-------|---------------|------|---------|
| growth | off | 추세 | learning_rate | 0.1 | 학습률 |
| seasonality | True | 계절성 | n_lags | 2 | 시차 |
| batch_size | 64 ~128 | 학습량 | hidden_layers | 4 | 은닉층 |
| epochs | 100~200 | 학습사이즈 | hidden | 8~12 | 은닉층 차원수 |



PREFLAT

05. 모델 검증

train / test



[남양주시 train / test 2020.01 기준]

MAE 150.0

| 날짜 | 실제 평당가 |
|------------|-----------|
| 2022-06-01 | 3102.9794 |
| 2022-07-01 | 2930.7642 |

오차 범위 내에 6월과 7월이 예측

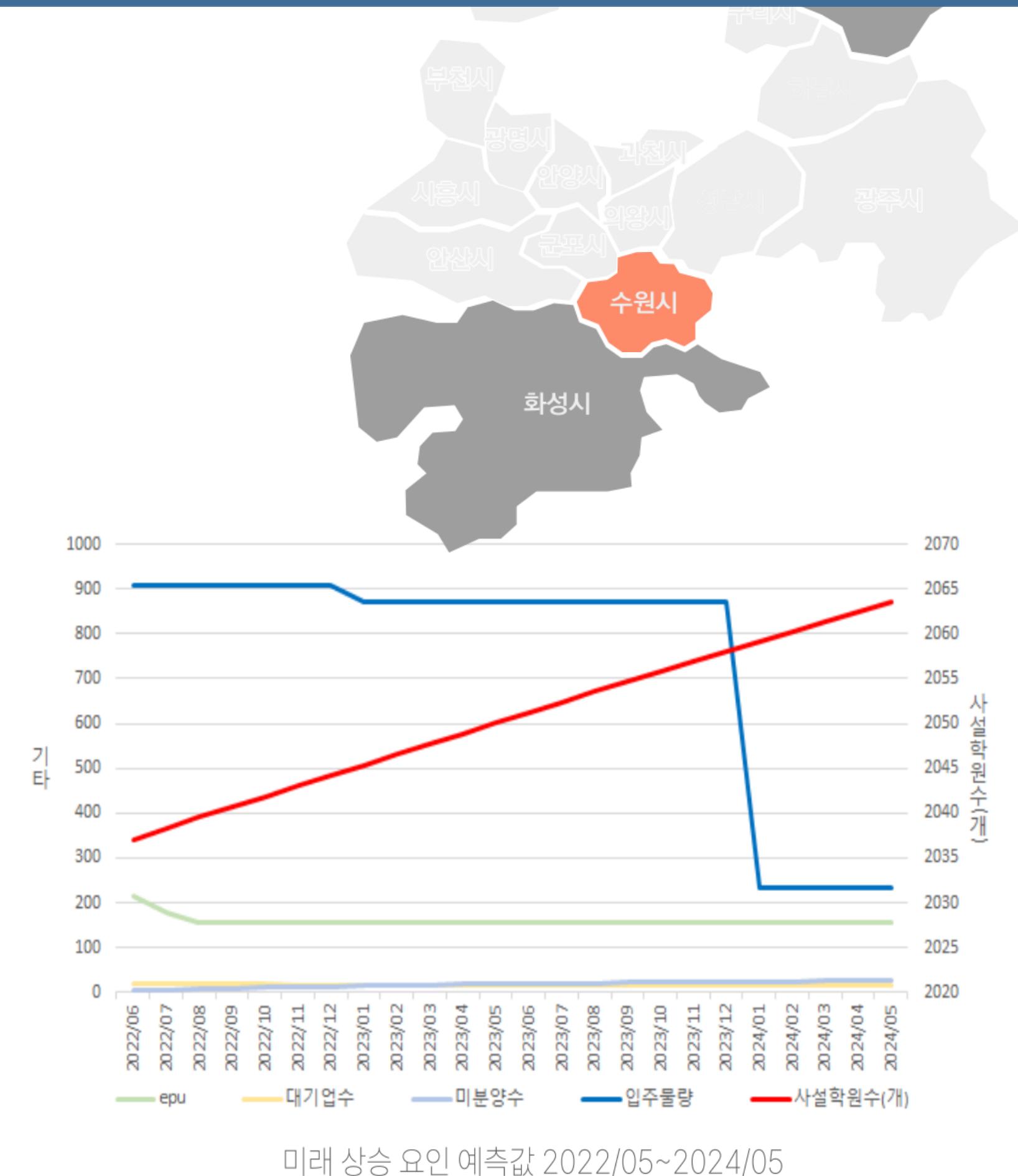
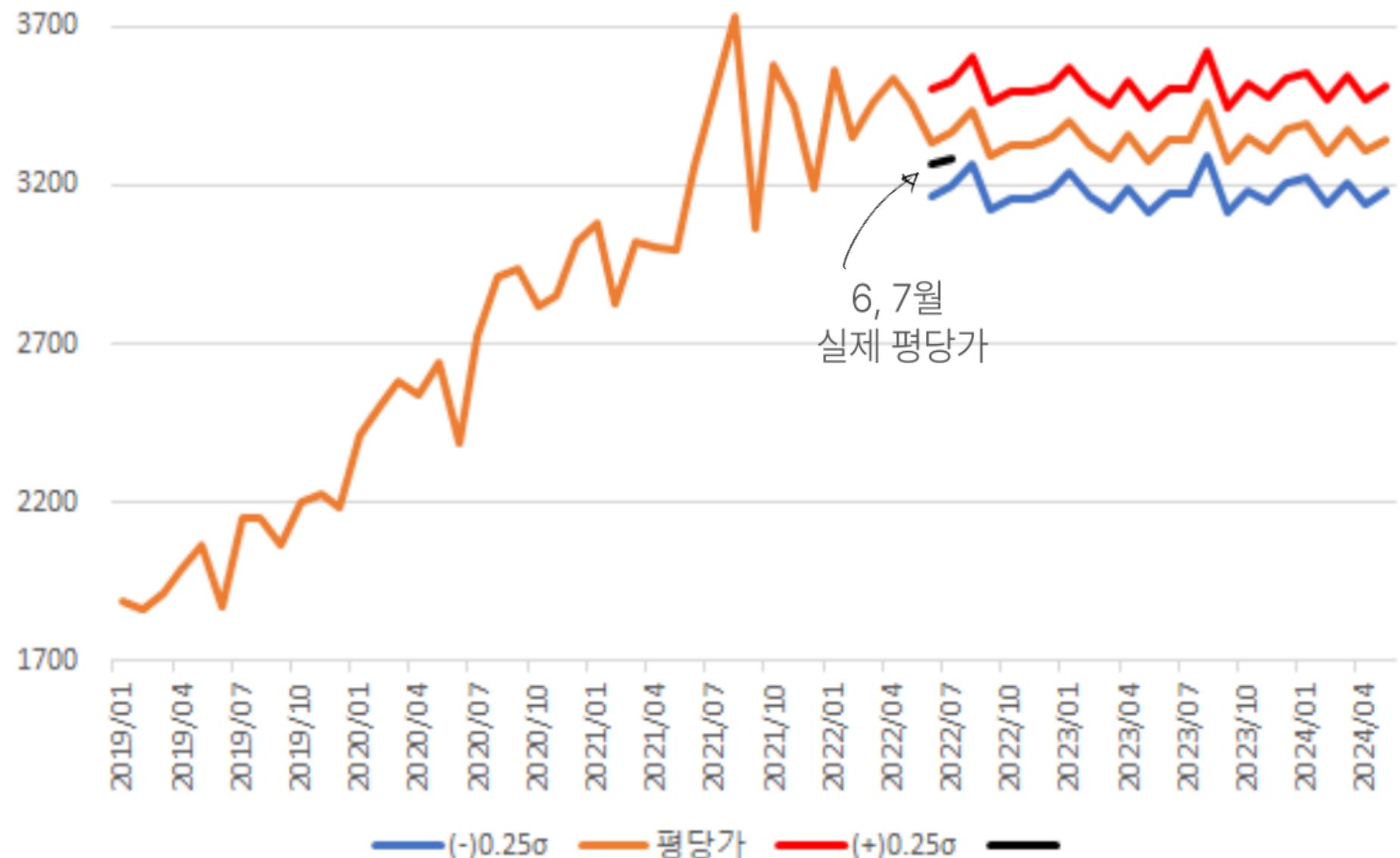
오차범위는 0.25σ 를 기준

PREFLAT

01. 활용 방안

가격 변동의 예측 결과

수원시 상승유지지역

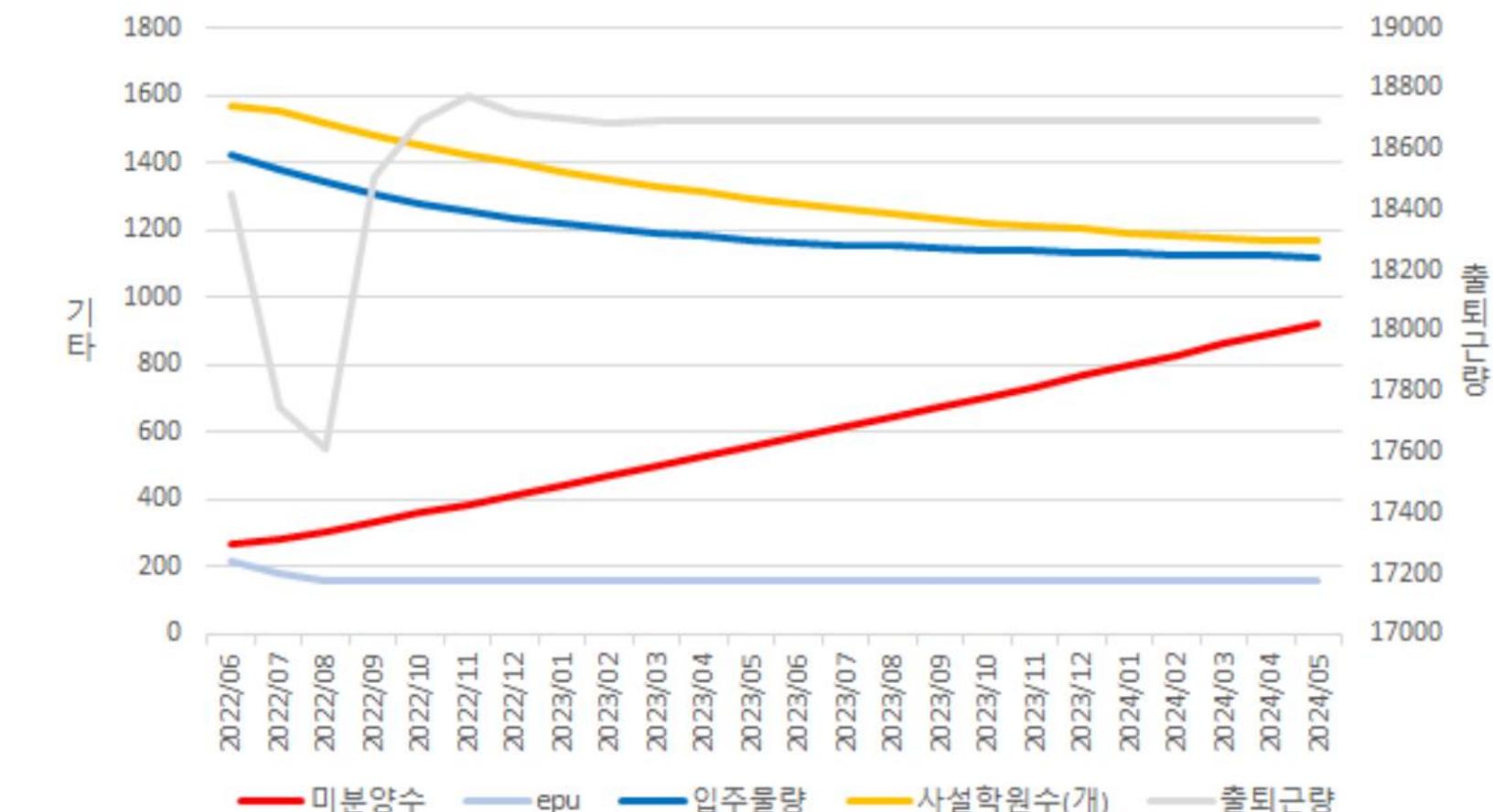
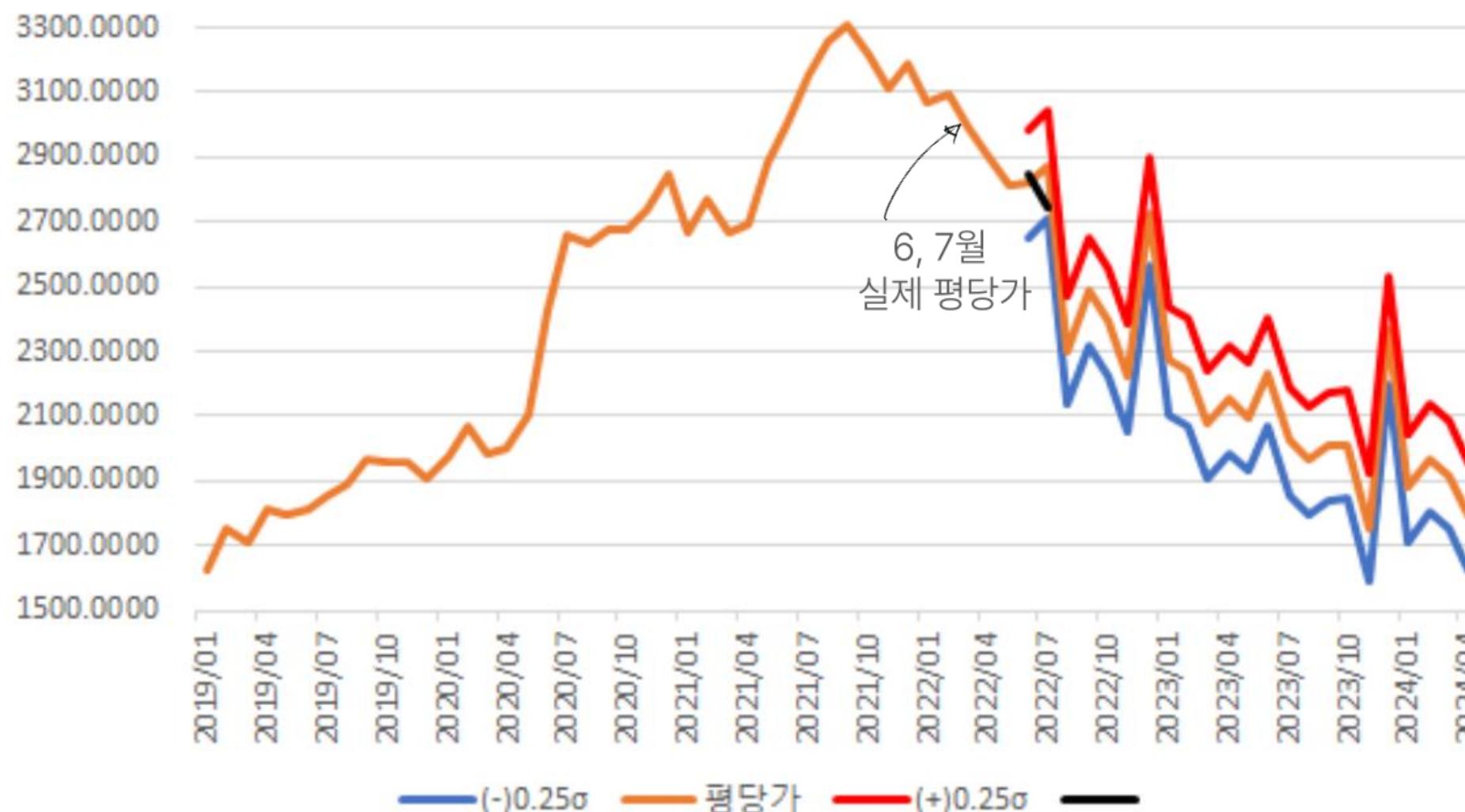


PREFLAT

01. 활용 방안

가격 변동의 예측 결과

화성시 하락예상지역

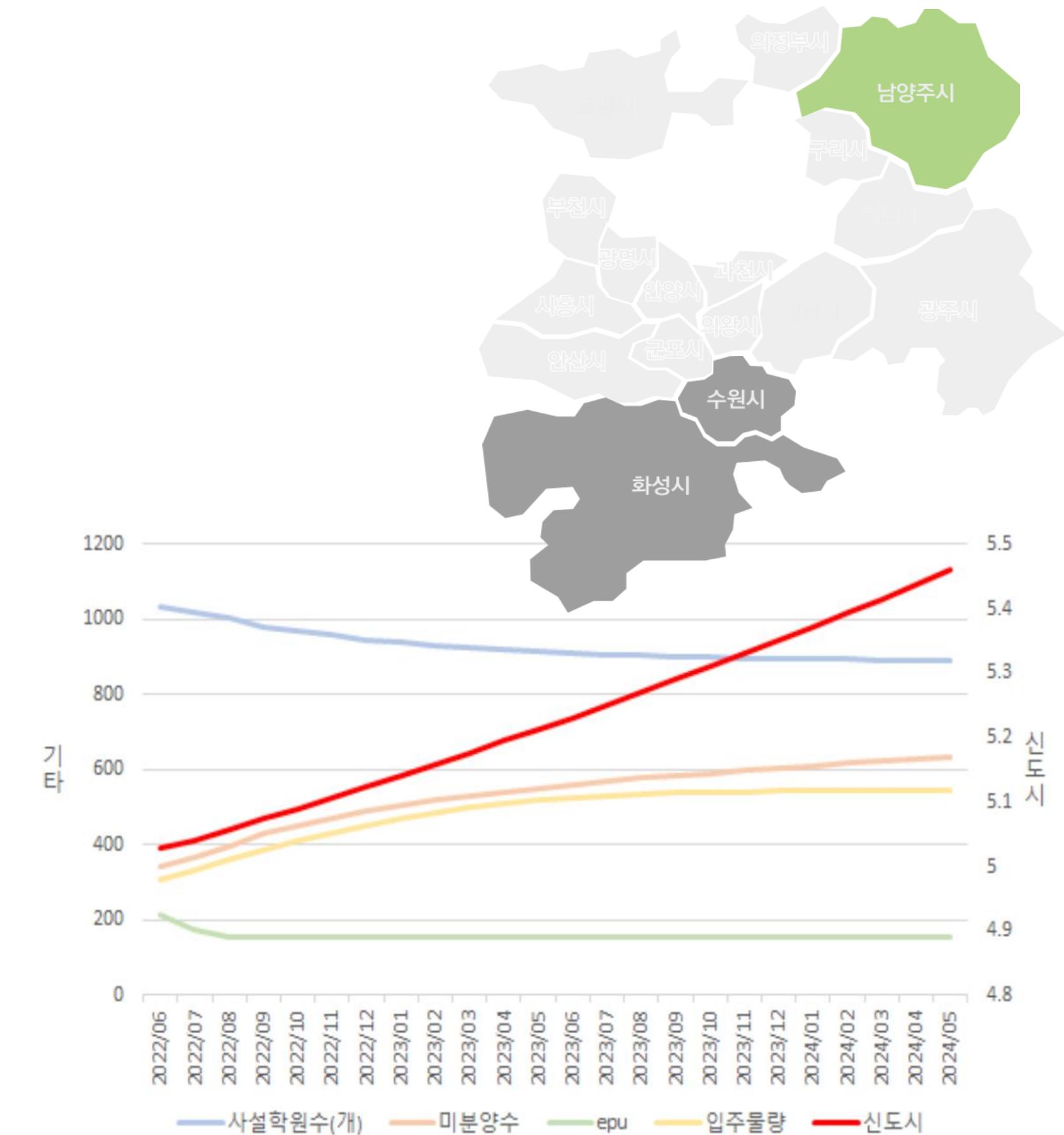
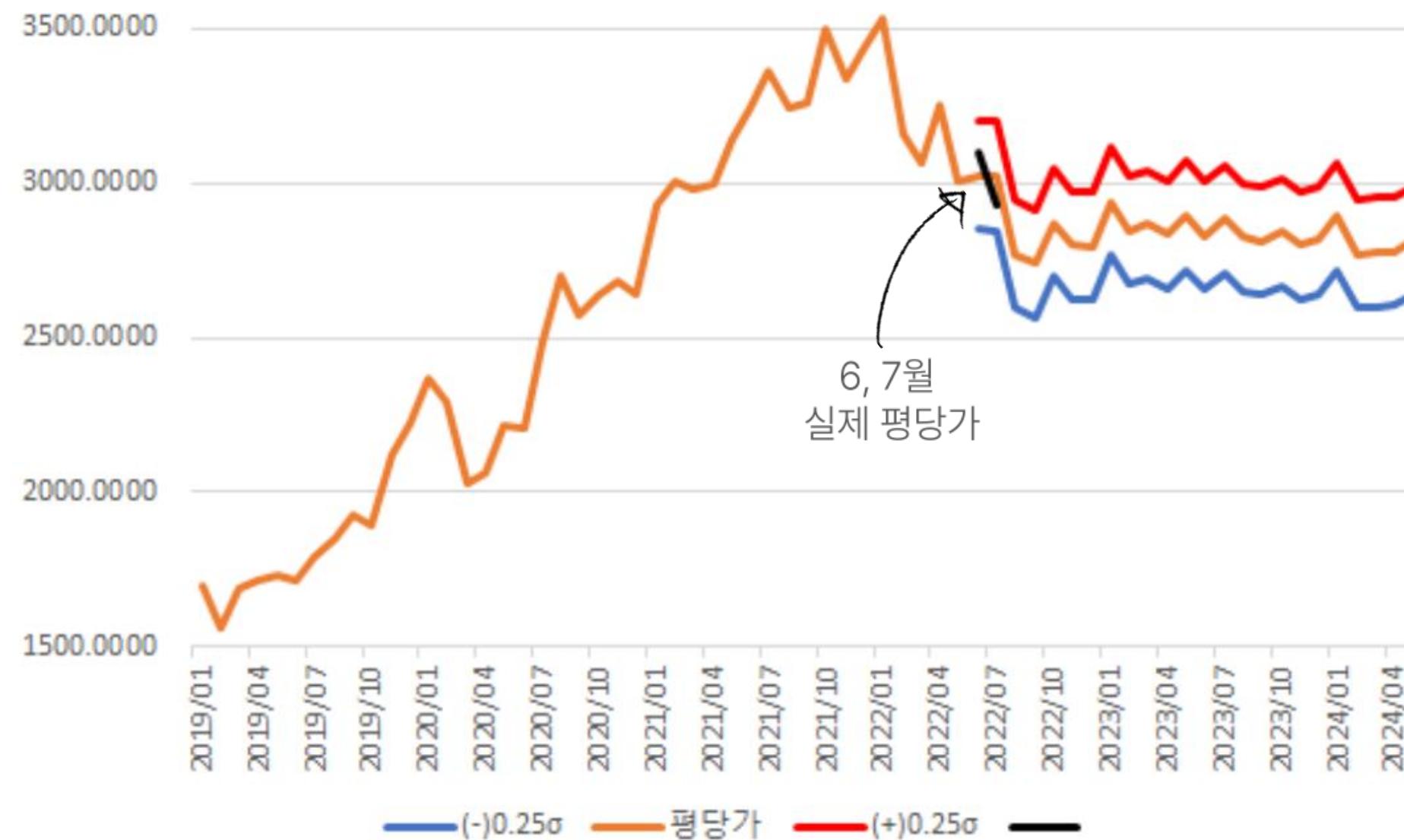


PREFLAT

01. 활용 방안

가격 변동의 예측 결과

남양주시 유망지역



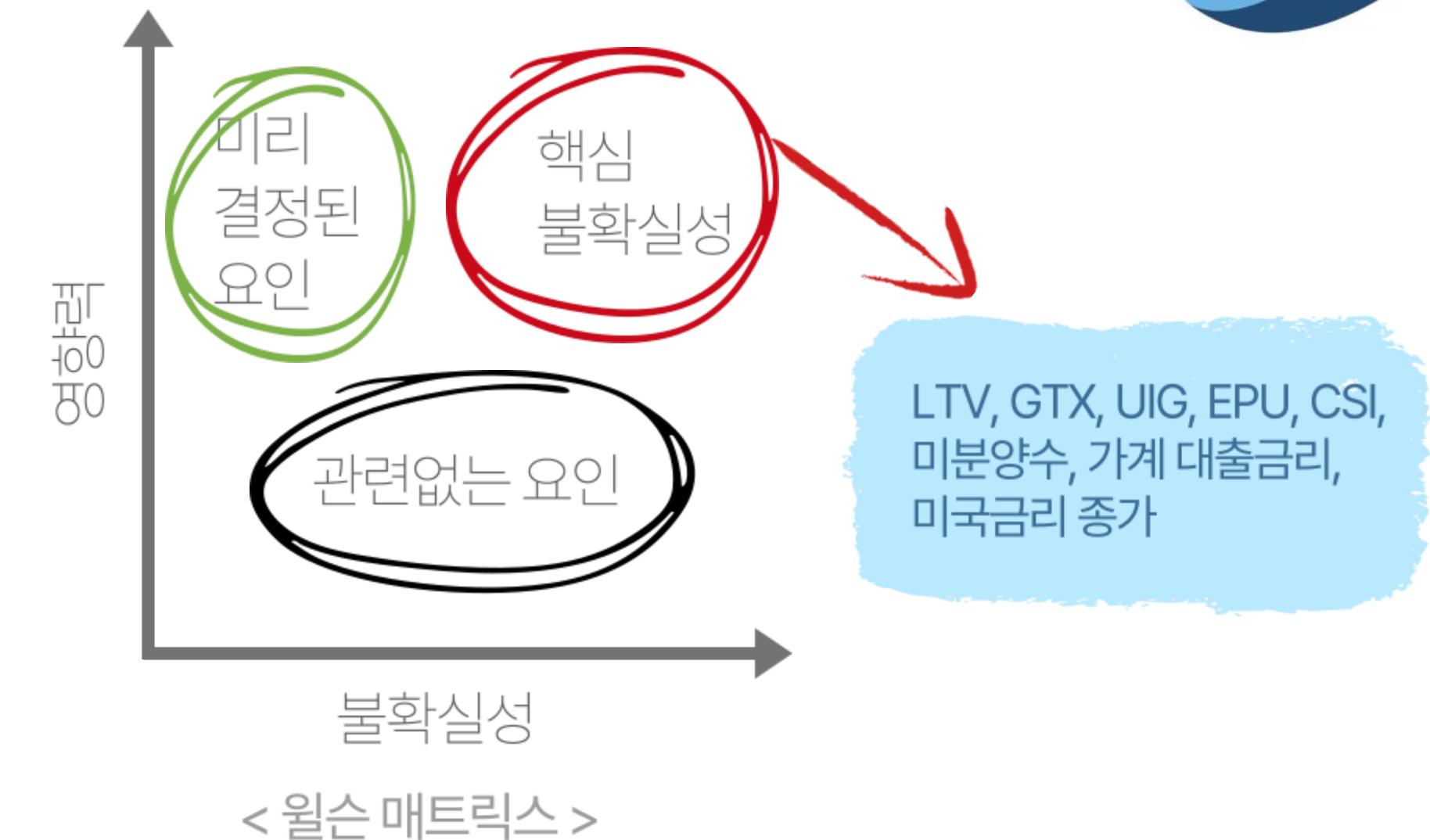
PREFLAT

02. 시나리오 분석

가격 변동의 세가지 시나리오

시나리오 분석이란,

급변하는 미래 불확실성에 대응하기 위해 단정적 미래예측이 아닌 다양한 미래사회 모습을 파악하고 비전, 목표 및 전략을 수립하는 시나리오 기반의 전략적 미래예측



PREFLAT

02. 시나리오 분석

탐색적 기법을 통한 세가지 시나리오



시나리오 1) 건설사 최대 실적

경제 회복이 빨라 시장 상황이 개선되는 경우

시나리오 2)

이전의 추세를 반영하는 경우

시나리오 3) 건설사 손실 감당 임계치

부동산 시장이 지속적으로 안좋은 경우

미래는 본질적으로 불확실하기 때문에 시나리오 중 가능성성이 높은 시나리오를 판단하는 것은 의미가 없으며 모든 시나리오를 '동등'하게 고려



PREFLAT

03. 기대효과

예측을 통해 얻을 수 있는 효과



1. 과학적 의사결정

현재
데이터 기반 →

미래 예측
데이터 기반

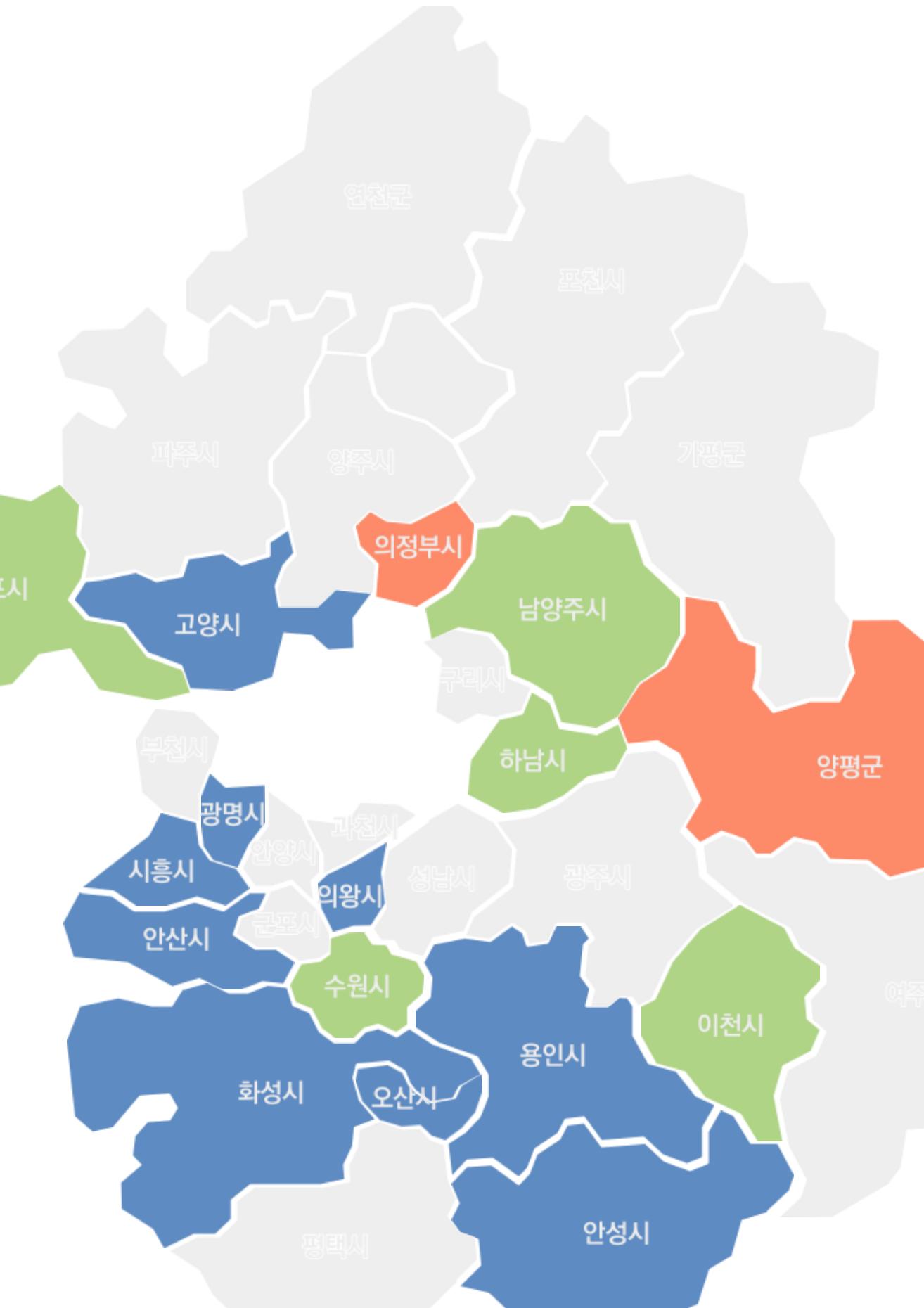
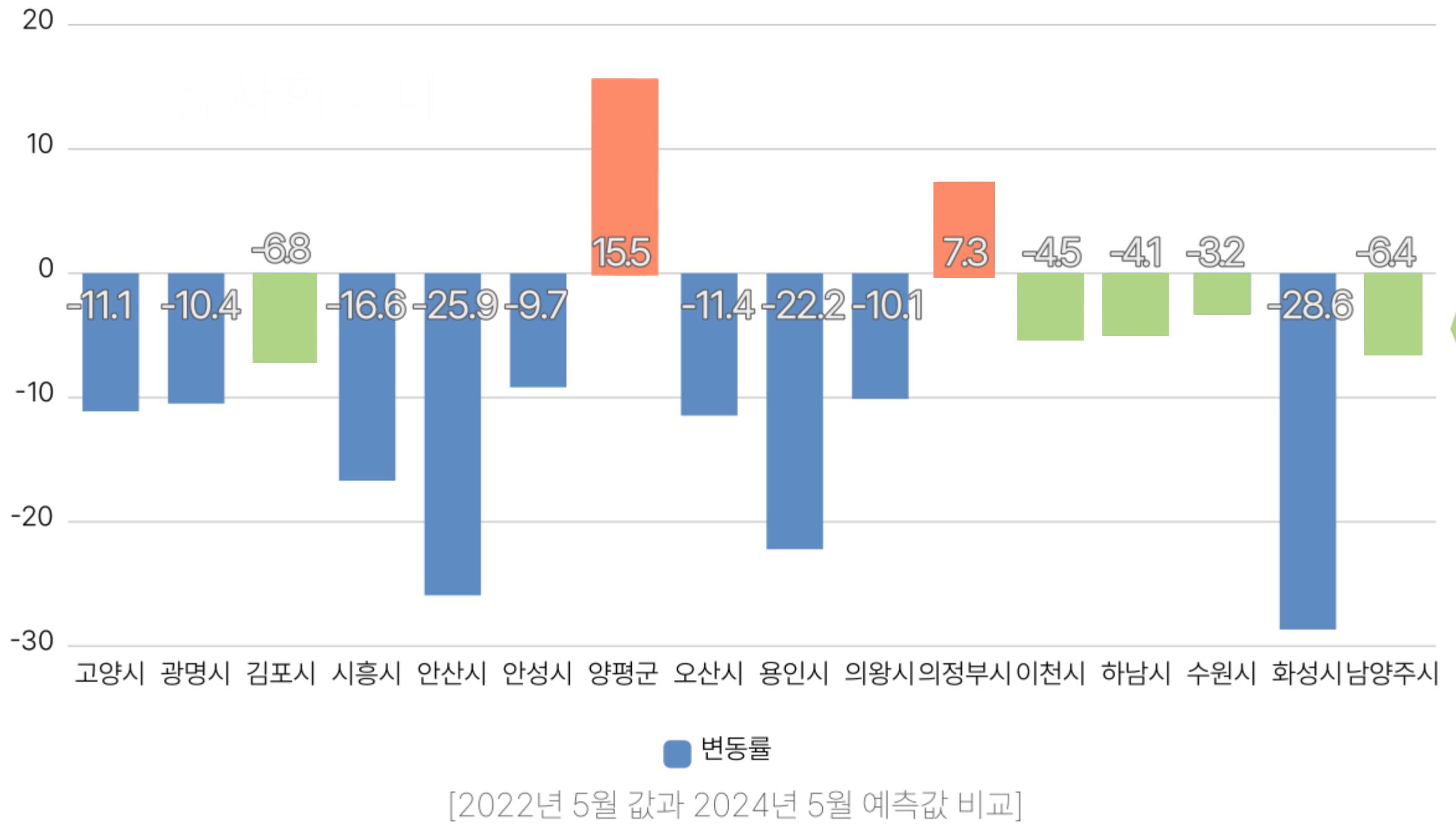
현재 건설사는 FS로 수익성을 판단
but 수주에서 분양까지 최대 2-3년 소요된다는 한계
미래 데이터 기반 사업진행을 통해 리스크 감소에 도움

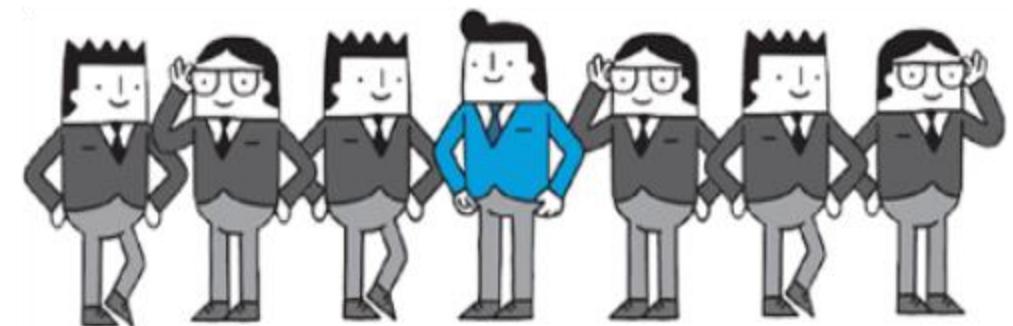
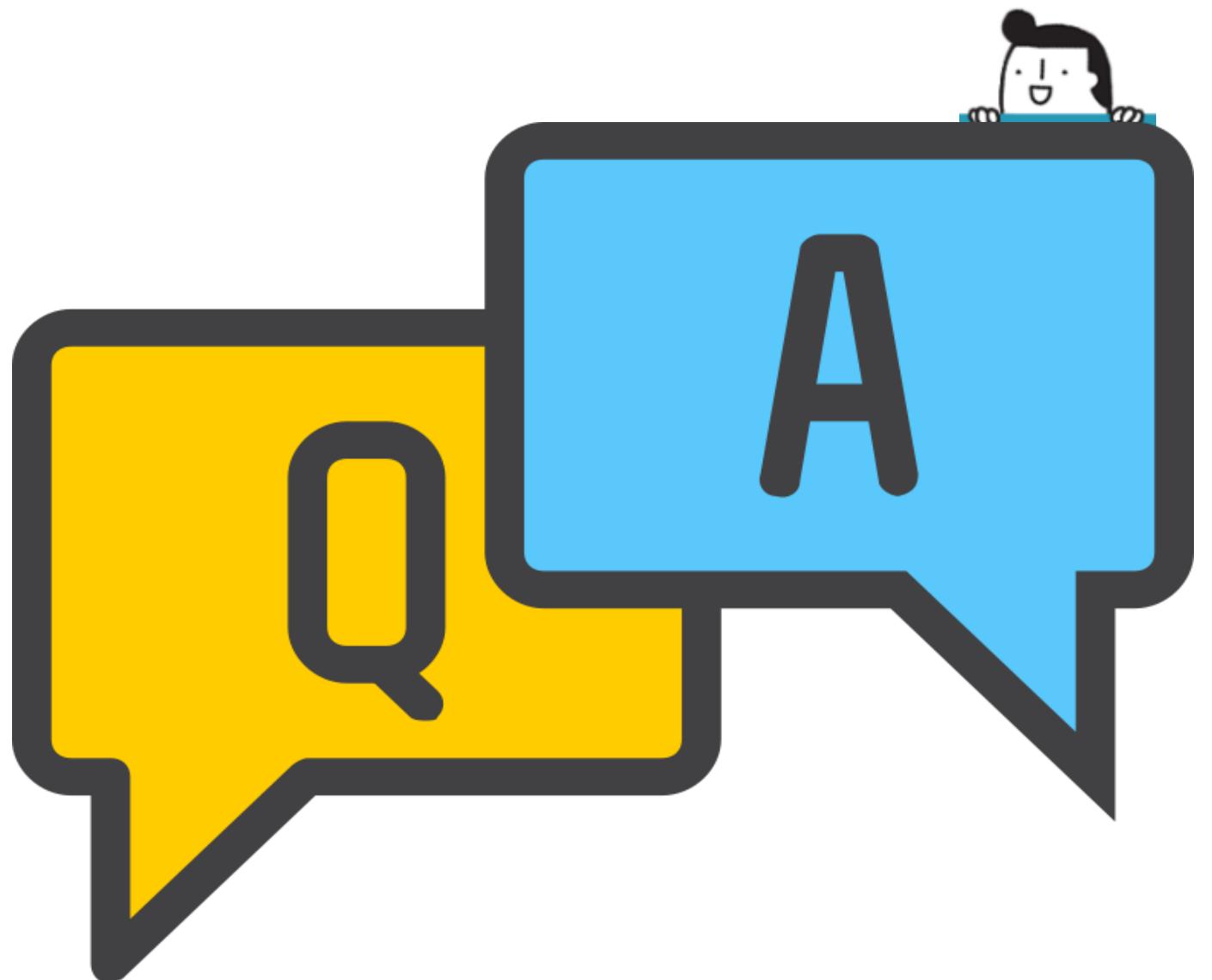
2. 기업의 리스크 관리에 기여

건설사가 시장 변동에 대비 가능

→ 급변하는 미래의
불확실성에 대한
리스크 감소







출처

책

빅데이터로 예측하는 대한민국 부동산의 미래 (저자 : 조영광)

논문

기계학습 알고리즘을 활용한 지역 별 아파트 실거래가격지수 예측모델 비교: LIME 해석력 검증

<https://koreascience.kr/article/JAKO202029660099787.pdf>

기계학습을 이용한 아파트 매매가격 예측 연구 : 한국 아파트의 내·외적 데이터 수집과 가격 예측 중심으로

https://manuscriptlink-society-file.s3-ap-northeast-1.amazonaws.com/kips/conference/2020fall/presentation/KIPS_C2020B0235.pdf

기계 학습을 이용한 공동주택 가격 추정: 서울 강남구를 사례로* http://www.kreaa.or.kr/data/vol24-1/24_01_05.pdf

머신러닝을 이용한 부동산 지수 예측 모델 비교. <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10521014>

SOM과 LSTM을 활용한 지역기반의 부동산 가격 예측* <http://koreascience.kr/article/JAKO202121061531451.pdf>

시나리오 분석 참고 논문

<https://scienceon.kisti.re.kr/srch/selectPORSrchReport.do?cn=DI200900084>

데이터출처

연령평균- <http://rtdown.molit.go.kr/> // 가구수(2-4인) - 통계청 KOSIS https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1JC1505

사설학원수-통계청 KOSIS / https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1YL21191&conn_path=I2

매매회전율 1)경기도 공동주택 현황/경기 데이터드림 <https://data.gg.go.kr/portal/data/service/selectServicePage.do?page=1&sortColumn=&sortDirection=&infId=VUPYJVKMYEYIKOQDILSR30099546&infSeq=1&searchWord=%EC%95%84%ED%8C%8C%ED%8A%B8>
2)매매거래량-한국부동산원 통계정보시스템 <https://www.reb.or.kr/r-one/cm/cntnts/cntntsView.do?mi=10113&cntntsId=1409>

미분양수-대우건설 // 입주물량-대우건설

강남접근성점수1-서울교통공사. <http://www.seoulmetro.co.kr/kr/cyberStation.do>. // 강남접근성점수2-네이버 지도. <https://map.naver.com/v5>

GTx - <https://www.gg.go.kr/contents/contents.do?cidx=497&menuId=1850>

주택가격전망 CSI -한국은행 // 신도시-국토교통부. <https://www.xn--3-3u6ey6lv7rsa.kr/kor/CMS/Contents/Contents.do?mCode=MN037>

LTV-https://www.kacpta.or.kr/stock/stock_board.asp // UIG-뉴욕 연방준비은행

<https://www.newyorkfed.org/research/policy/underlying-inflation-gauge>

출퇴근량-통계청 KOSIS // https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1PA1503&conn_path=I2

대기업수-통계청 KOSIS. https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_2BE0002

매수우위지수- <https://data.kbland.kr/kbstats/psychology-of-housing-market>

가계대출금리 - 한국은행 // 미국금리종가 - <https://kr.investing.com/rates-bonds/u.s.-10-year-bond-yield>

EPU - https://www.policyuncertainty.com/korea_monthly.html // kb 선도지수 - 한국은행

Thank you

감사합니다