

2022 국토교통

데이터 활용 경진대회



임차자 보호를 위한 강동전세 발생 가능지역/주택유형별 분석 및 예측

전세역전팀

박찬호 | 이승수 | 조은정





목 차

INDEX

1. 문제 정의
2. 데이터 수집 및 정제
3. 데이터 분석
4. 대시보드
5. 제안 및 기대효과





1. 문제 정의

임차인의 전세보증금을 전세 사기로부터 피할 수 있도록 전세가율을 파악할 수 있는
강통전세 위험 모니터링 대시보드를 제시합니다.

강통전세란?

전세 계약 만료 시 세입자가
임대인으로부터 전세 보증금을 반환
받지 못할 가능성이 큰 전세 형태

강통전세란?



문제점

임차인이 전세보증금을 돌려받기 어려움

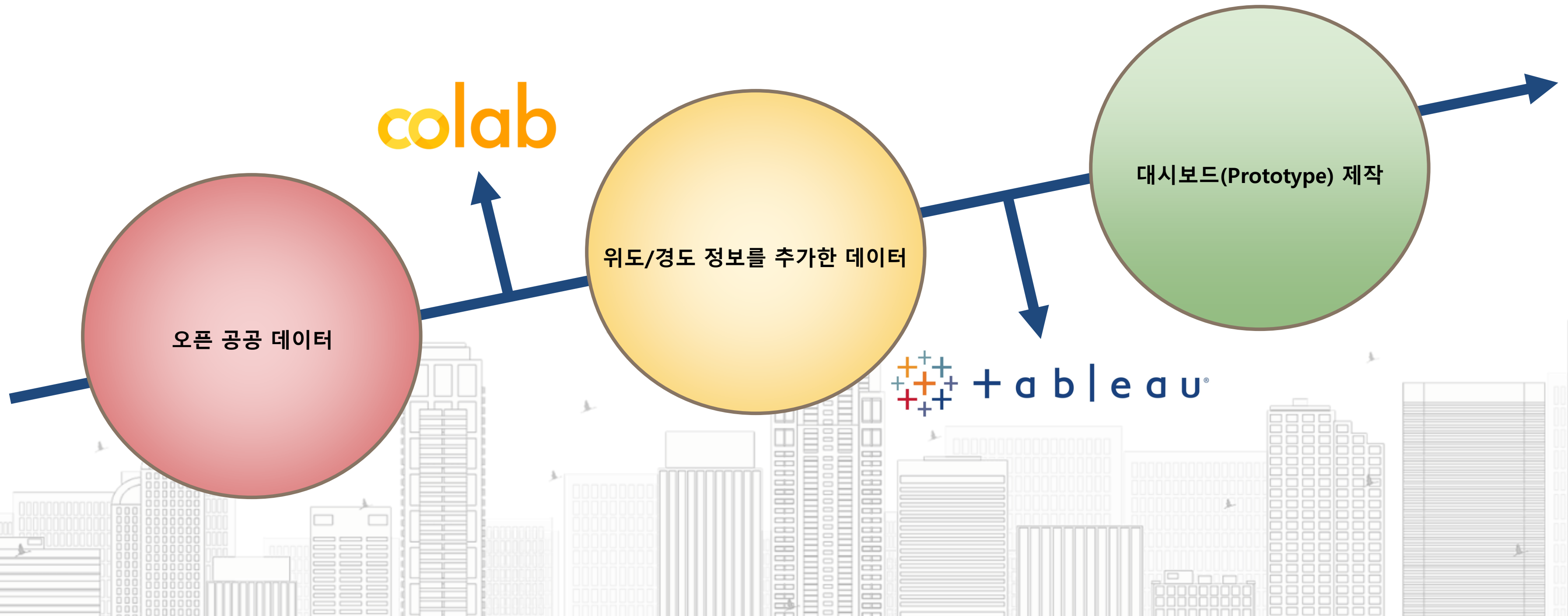


전세역전의 깡통전세 발생가능지역 분석 및 예측

2. 데이터 취합 및 정제

프로젝트 개요

Data



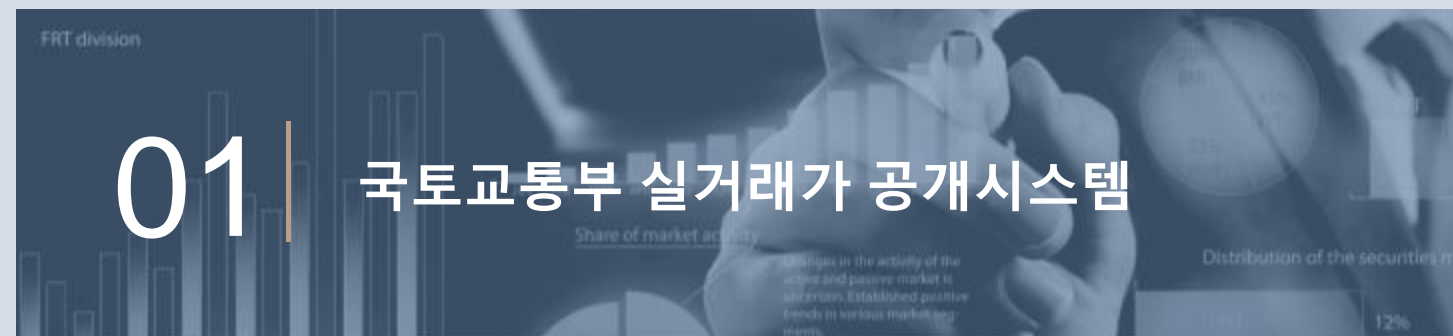


전세역전의 깡통전세 발생가능지역 분석 및 예측

2. 데이터 취합

데이터 구성

Data



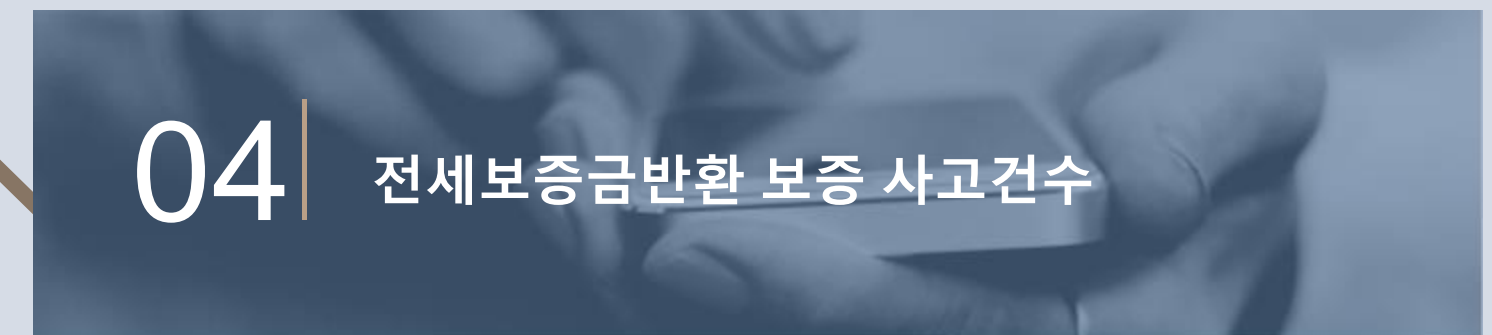
01 | 국토교통부 실거래가 공개시스템



02 | 전세보증금반환 보증 발급건수



03 | 지역별 위도, 경도 데이터



04 | 전세보증금반환 보증 사고건수

가명정보 결합
데이터





전세역전의 강통전세 발생가능지역 분석 및 예측

2 데이터 정제

REFINEMENT

02 전세보증금 발급건수 & 사고건수

- 연도별/지역별/주택유형별로 구분해서 정제 작업

3. 연도/주택유형별로

```

1 data_om = []
2 for i in range(M1): # year
3     for j in range(M2): # class
4         sum = 0
5         for s in range(N):
6             if data[s,0] == data_np_year_type[i][0] and data[s, 1] == data_np_year_type[i][1] and
7                 sum += data[s, 4]
8             else:
9                 continue
10        data_om.append(sum)
11 data_np_om = np.array(data_om)
12 print(data_np_om.reshape(M1, M2))
13 data_np_om = data_np_om.reshape(M1, M2)

```

```

[[ 76 185 73 9787 47 108]
 [ 41 78 35 3893 17 23]
 [ 85 222 60 5830 54 38]
 [ 122 241 66 6386 35 40]
 [ 138 226 64 6702 39 26]
 [ 225 556 93 7272 102 162]
 [ 465 1200 194 7728 184 435]
 [ 540 1649 243 8955 225 606]
 [ 589 2068 283 9289 222 633]
 [ 852 2701 370 13404 277 912]
 [ 1085 3851 594 15696 396 1369]
 [ 1265 3217 575 14651 348 1309]
 [ 1707 4218 805 17390 418 1941]
 [ 2008 5171 751 21640 541 3200]
 [ 2802 7281 1147 23864 717 4259]]

```

Colab

A	B	C	D	E	F	G	H	I	J
Region	2016	2017	2018	2019	2020				
강원도	70	253	582	1549	1695				
경기도	12044	16889	32990	56250	61542				
경상남도	484	1076	2223	5111	5932				
경상북도	307	636	1348	3210	4136				
광주광역시	193	720	1154	2127	2023				
대구광역시	926	1726	2675	3549	3843				
대전광역시	276	535	990	2058	2199				
부산광역시	1331	3265	5976	9499	10480				
서울특별시	3728	11272	25980	43767	48930				
세종특별자치시	74	128	250	595	898				
울산광역시	224	527	1082	1926	2330				
인천광역시	4227	5632	10849	18772	25621				
전라남도	133	150	406	879	909				
전라북도	72	322	662	1467	2138				
제주특별자치도	7	42	110	467	703				
충청남도	269	490	1111	2785	3400				
충청북도	95	255	963	2084	2595				

카카오 API에서 주소 데이터를 바탕으로 위도/경도를 추출하는 코드

임의로 합한 전세/매매 데이터에서 공통되는 주소를 추출해 코드를 실행시켜 저장한 csv 파일 중 일부



전세역전의 갱통전세 발생가능지역 분석 및 예측

3. 데이터 분석

ANALYSIS



전세역전의 강통전세 발생가능지역 분석 및 예측

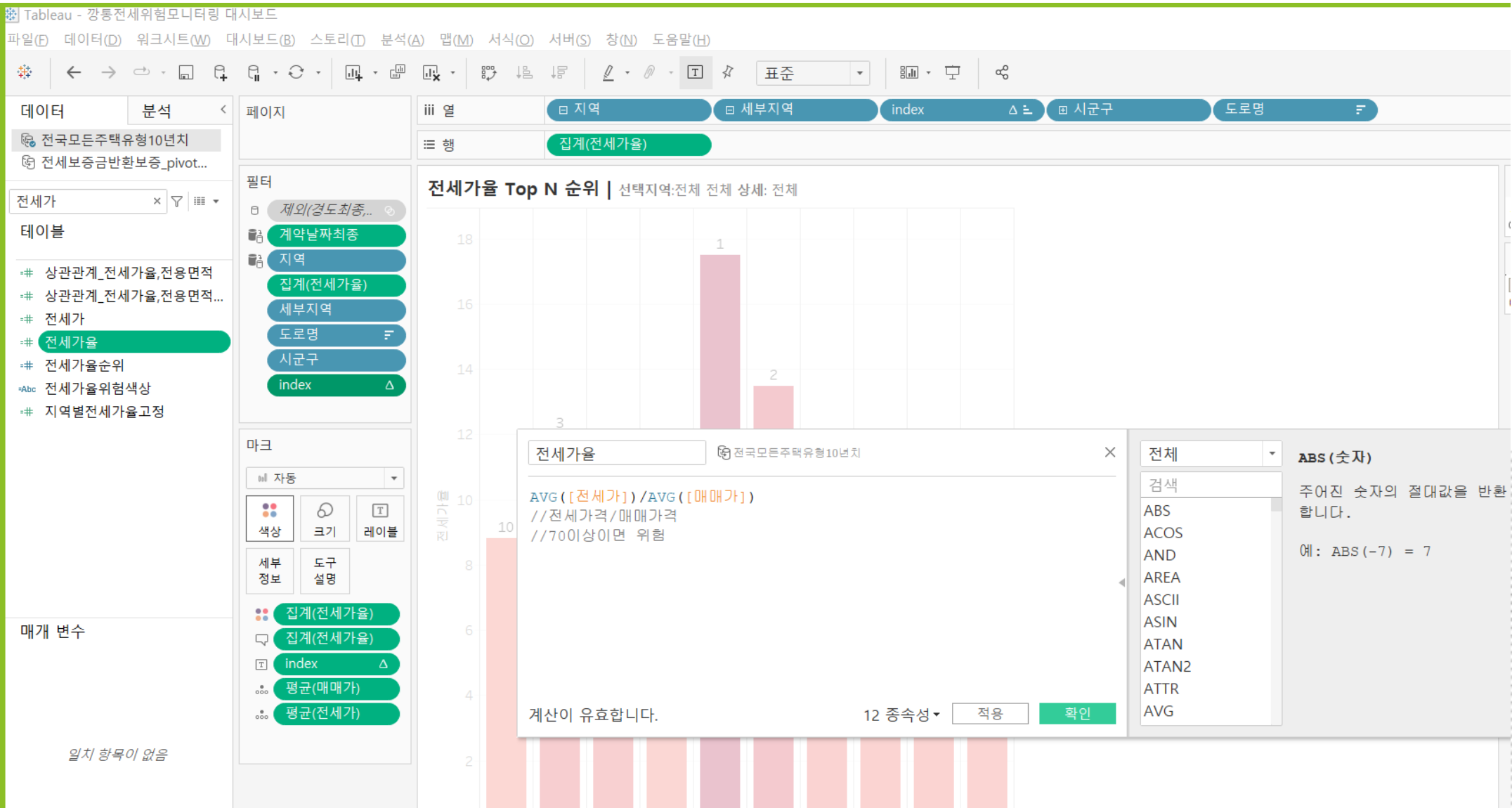
3.1 데이터 분석

데이터 예측 설명

ANALYSIS

- 강통전세 위험률 계산 방법

강통전세위험률(전세가율) = $\frac{\text{전세가격}}{\text{매매가격}} \times 100$





전세역전

전세역전의 강통전세 발생가능지역 분석 및 예측

3.2 데이터 분석

데이터 예측 설명

ANALYSIS

지수 평활법

과거의 모든 자료를 사용하여 평균을 구하면서
최근의 자료에 더 높은 가중치를 부여하는 방법

사용된 오차 함수



RMSE

평균 제곱근 오차는 오차를 제곱한 값의
평균을 낸 것에서 루트를 씌운 편차



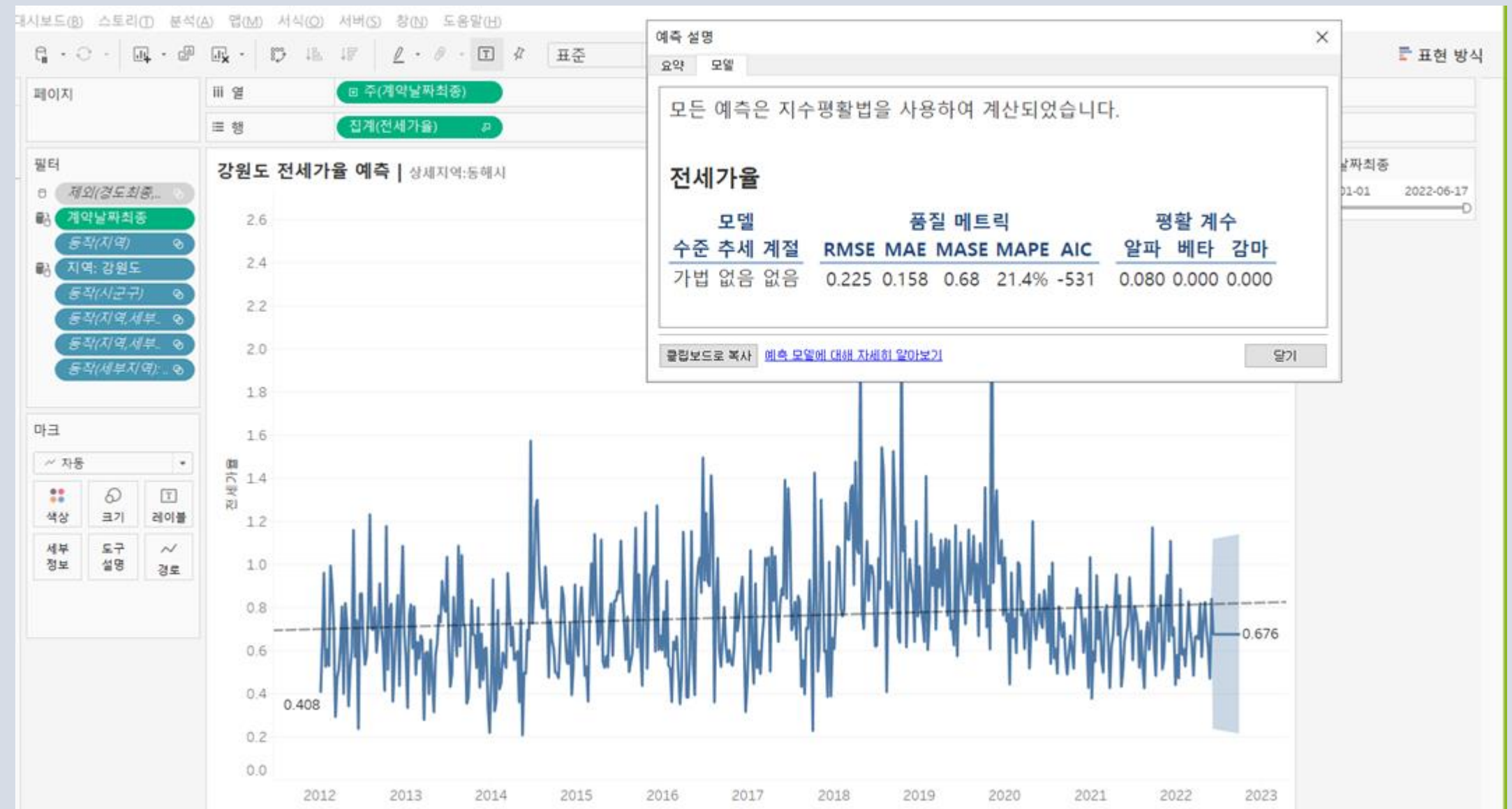
MAE

평균 절대 오차는 동일한 현상을 표현하
는 쌍을 이루는 관측치 간의 오차 척도



MASE

평균 절대 축척 오차는 평균 오차를
배율 계수로 나누어 계산



<강원도 동해시의 전세가율 예측>



전세역전의 강통전세 발생가능지역 분석 및 예측

3.3 데이터 분석

사용한 데이터 분석 방법 - 피어슨 상관 계수를 통한 회귀 모델

ANALYSIS

데이터에서 두 변수의 공분산(covariance)를
각각의 표준 편차의 곱으로 나눈 값

두 변수 X와 Y 간의 선형 상관 관계를 계량화한 수치

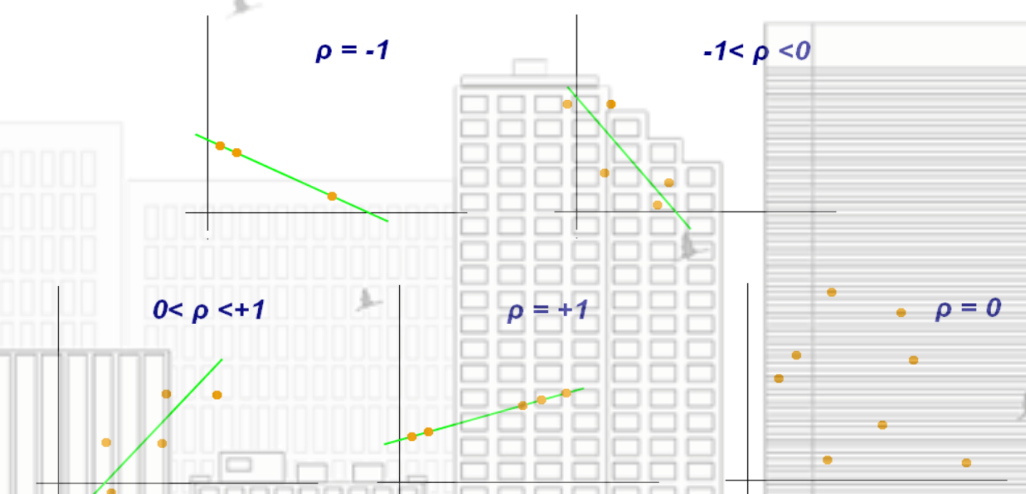
피어슨 상관계수

예측 회귀 모델

상관 계수 절대값이 1에 수렴하면 상관 관계가 높고
0에 수렴한다면 상관 관계가 없다.

$$r_{XY} = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \sqrt{\sum_i^n (Y_i - \bar{Y})^2}}$$

강통 전세와 연관있는 지표를 찾는데 용이!





전세역전의 강통전세 발생가능지역 분석 및 예측

3.3 데이터 분석

사용한 데이터 분석 방법 - 피어슨 상관 계수를 통한 회귀
모델

ANALYSIS

```

1 val_list = []
2 val_list2 = []
3 for element in np.unique(year1//100):
4     idx_ = df[df['Unnamed: 6']//100 == element].index
5     idx_2 = df2[df2['Unnamed: 7']//100 == element].index
6     li = []
7     for i in idx_:
8         li.append(hi[i].strip().split(','))
9
10    val_li = []
11    for i in range(len(li)):
12        value = ''
13        for j in range(len(li[i])):
14            value += li[i][j]
15        val_li.append(value)
16
17    print(val_li)

```



✓ [57] 1 df100.corr() # 아파트

	보증사고금액	전세가율(평균)	전세가율(중앙값)
보증사고금액	1.000000	0.384525	0.413293
전세가율(평균)	0.384525	1.000000	0.982383
전세가율(중앙값)	0.413293	0.982383	1.000000

✓ [58] 1 df101.corr() # 연립, 다세대

	보증사고금액	전세가율(평균)	전세가율(중앙값)
보증사고금액	1.000000	0.994909	0.976149
전세가율(평균)	0.994909	1.000000	0.990445
전세가율(중앙값)	0.976149	0.990445	1.000000

✓ [59] 1 df102.corr() # 오피스텔

	보증사고금액	전세가율(평균)	전세가율(중앙값)
보증사고금액	1.000000	0.768276	0.844563
전세가율(평균)	0.768276	1.000000	0.909594
전세가율(중앙값)	0.844563	0.909594	1.000000

1 df103.corr() # 단독, 다가구

	보증사고금액	전세가율(평균)	전세가율(중앙값)
보증사고금액	1.000000	0.622133	0.736321
전세가율(평균)	0.622133	1.000000	0.985522
전세가율(중앙값)	0.736321	0.985522	1.000000



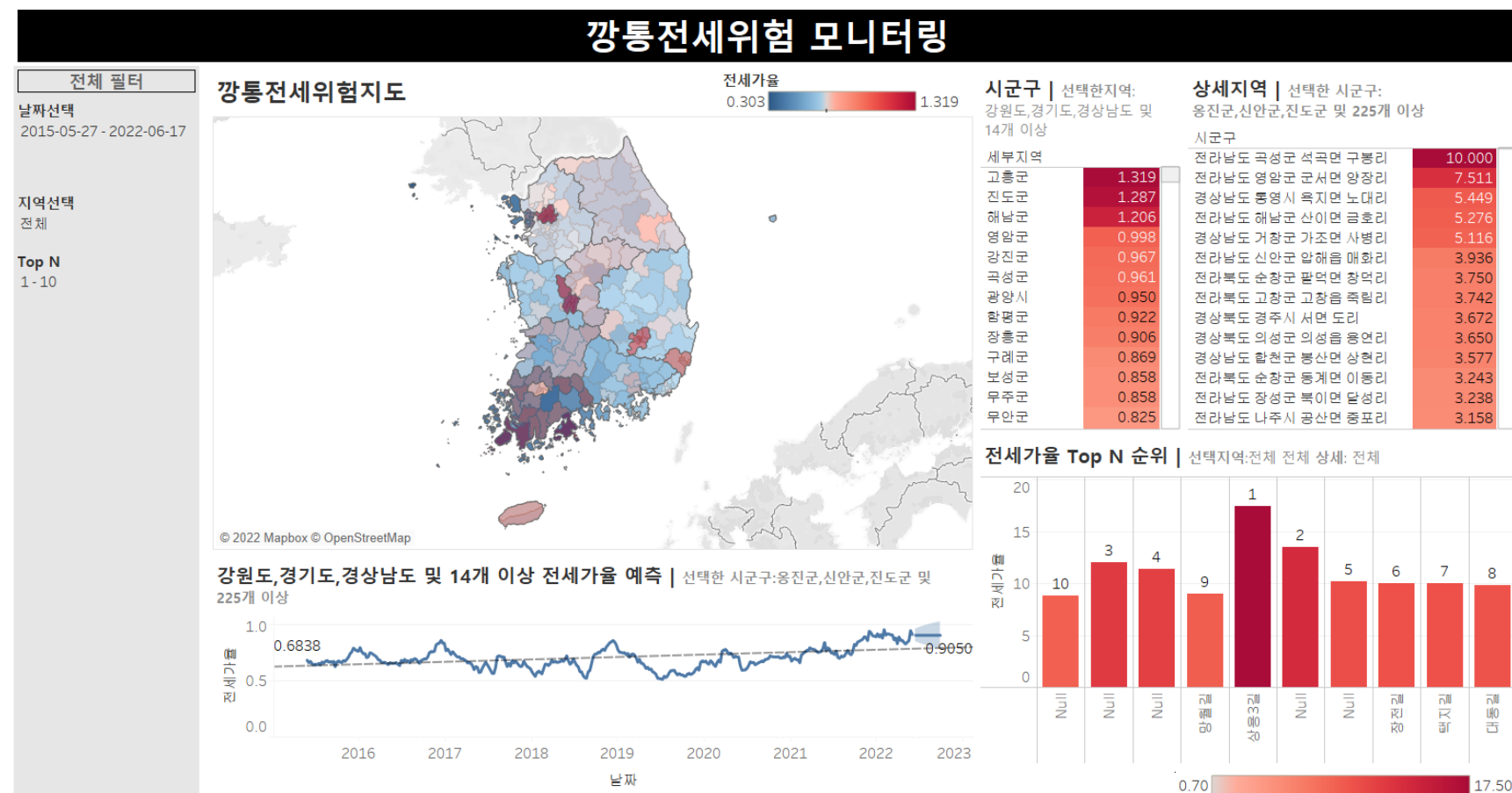
전세역전

전세역전의 강통전세 발생가능지역 분석 및 예측

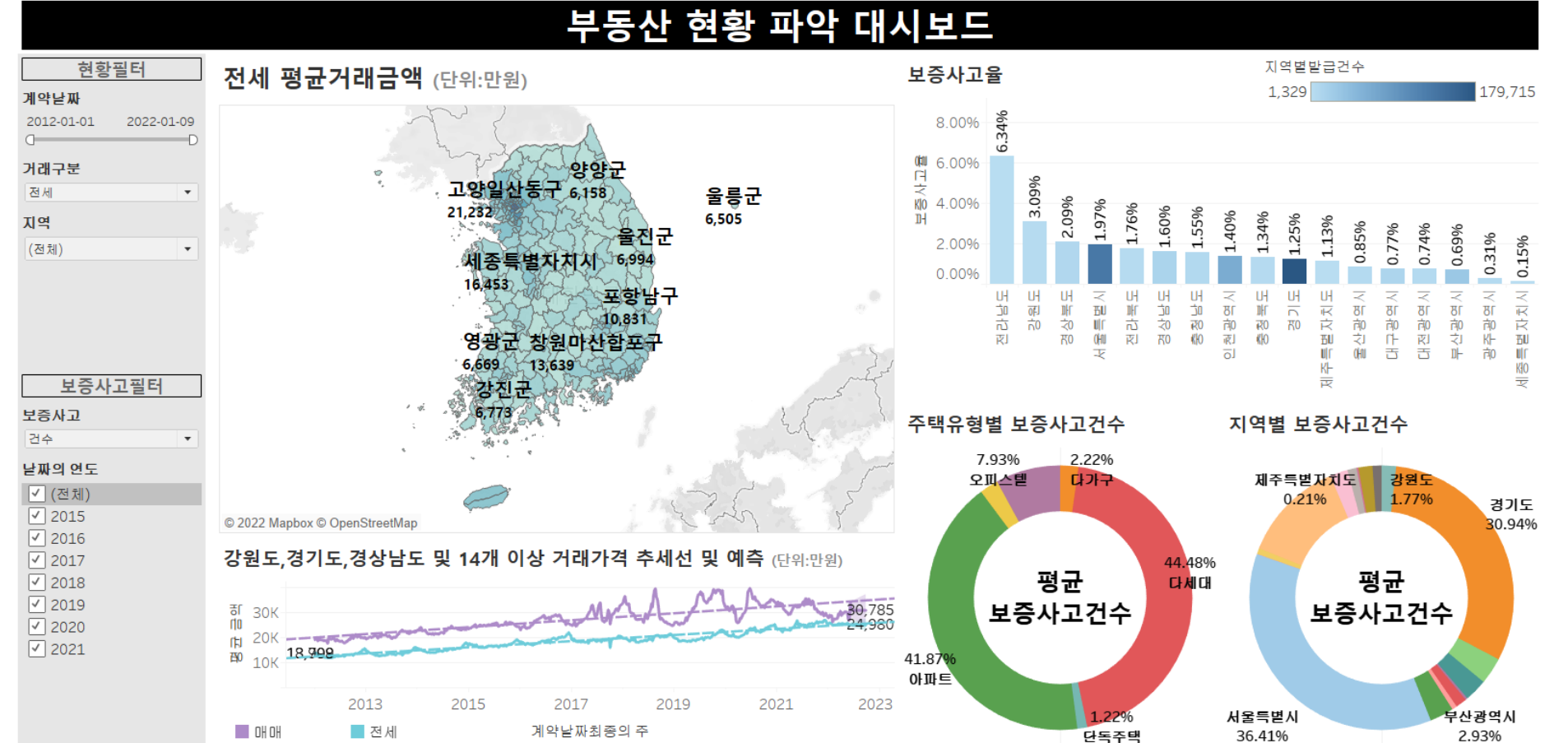
4. 대시보드

DASHBOARD

강통전세위험 모니터링



부동산 현황 파악 대시보드



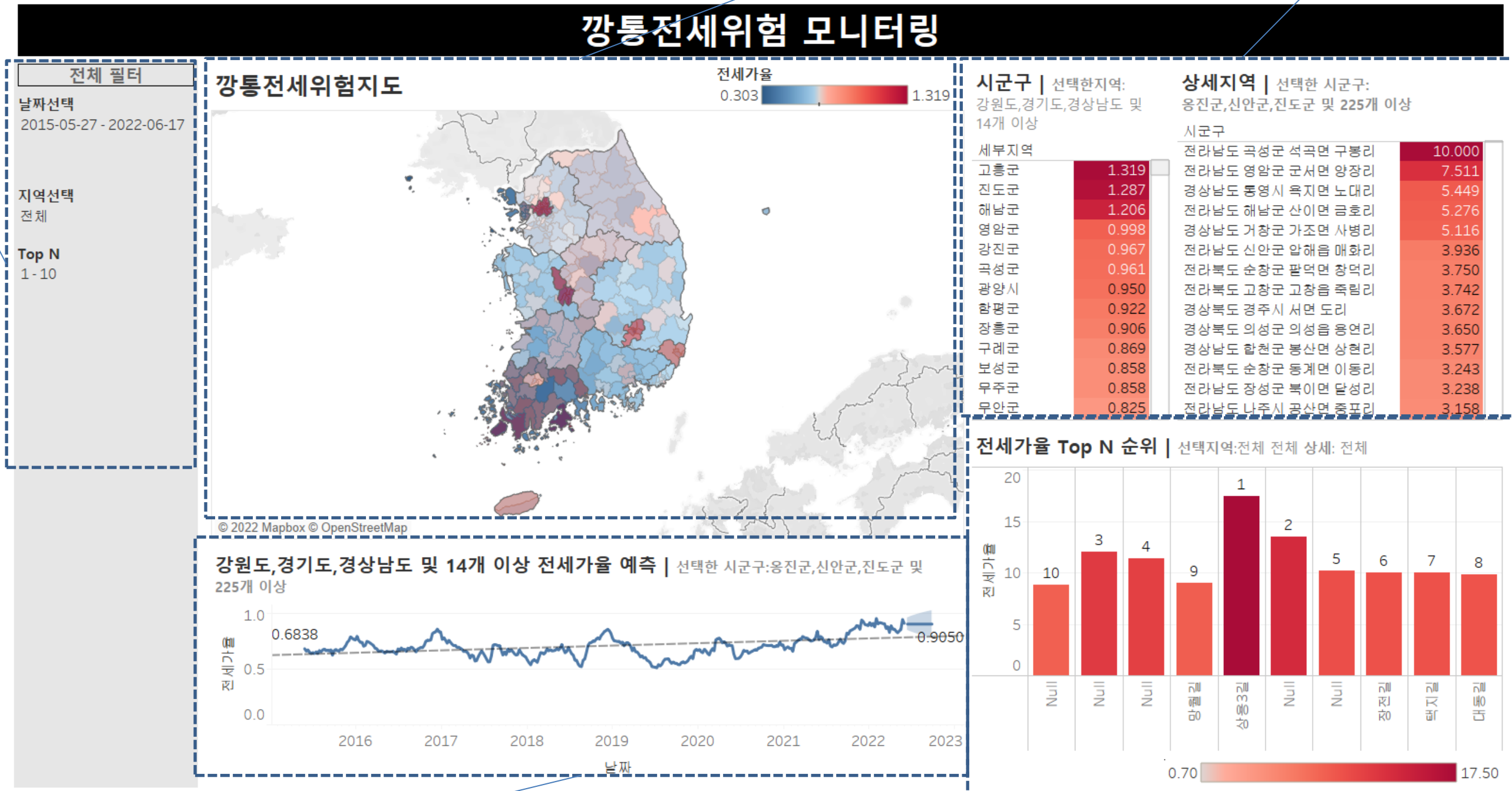


강통전세위험 모니터링 대시보드 동작 방식 설명

지도의 지역을 선택하면
다른 차트에 모두 영향을
주는 필터 역할도 겸한다.

지역 선택 > 시군구 단위에서 전세가율
표시 > 시군구 선택 > 상세지역
단위에서 전세가율 표시

전체 차트에 해당하는
필터
Ex) 날짜 및 지역을
선택할 수 있다.



전세가율 Top N 순위에
영향을 주는 필터 설정에
따라 상위 N개의
전세가율이 높은
상세주소를 보여준다.

선택한 날짜 범위에 따라 전세가율 예측을 보여준다.
마찬가지로 필터 및 오른쪽의 상세지역에서 지역을 선택하면 해당 지역의 전세가율의 흐름과 예측을 보여준다.

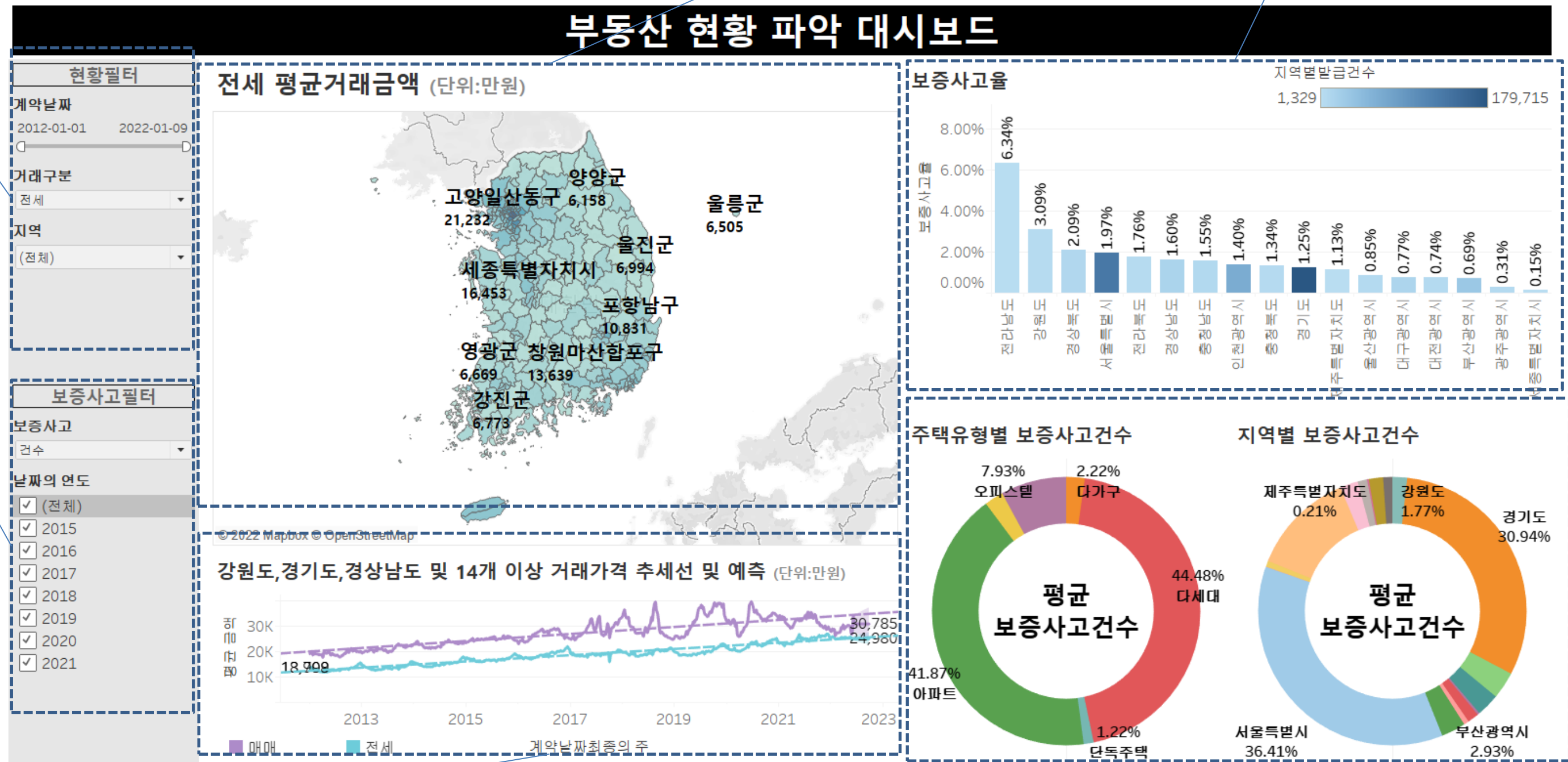


전세역전

부동산 현황 파악 대시보드 동작 방식 설명

거래구분 필터에 따라 전 지역의 평균 거래 금액을 지도로 파악할 수 있다.

각 지역별 보증사고건수/보증발급건수를 백분율 값으로 나타낸 차트



좌측의 차트들에 영향을 주는 필터 이다.

우측의 차트들에 영향을 주는 필터이다.

선택한 날짜 범위에 따라 매매가 및 전세가의 추세를 알 수 있다. 선택한 날짜 범위를 바탕으로 가격을 예측한다.



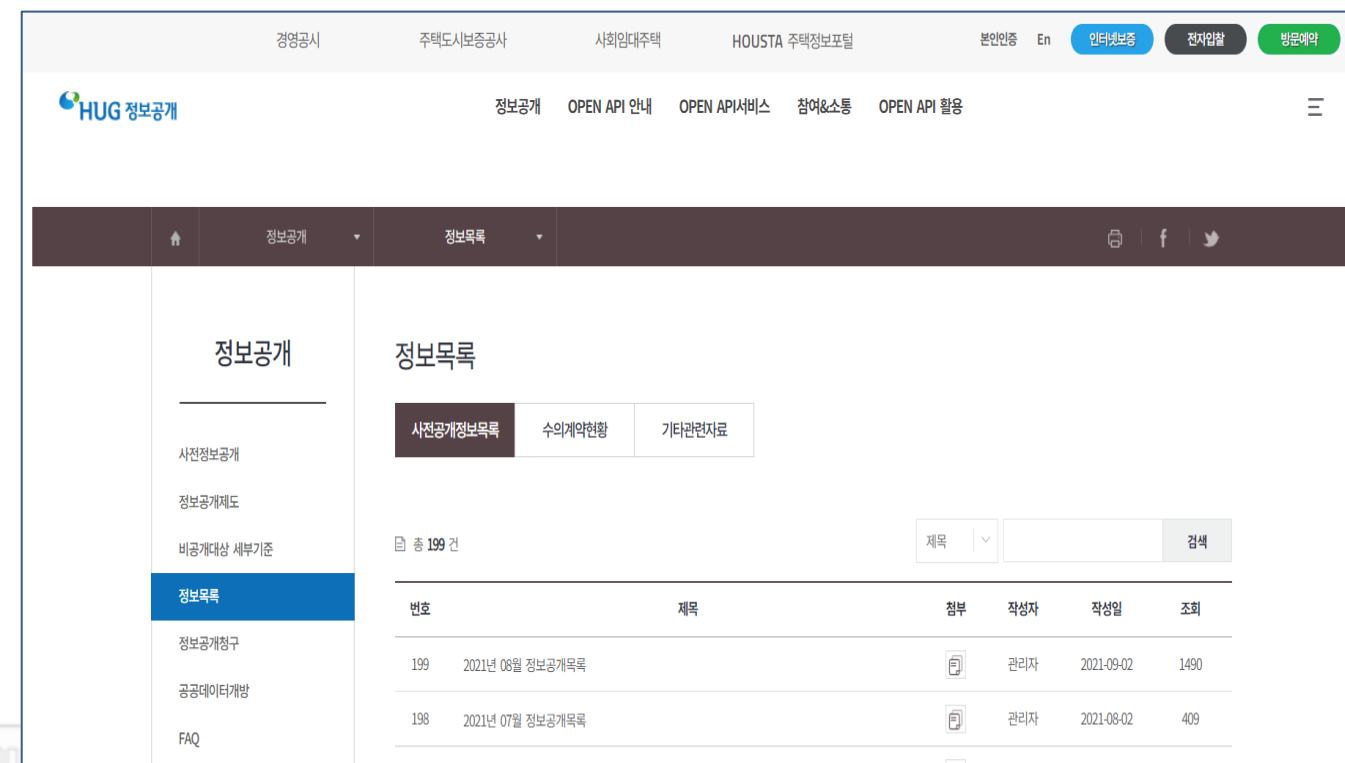
전세역전의 강통전세 발생가능지역 분석 및 예측

5. 제안 및 기대효과

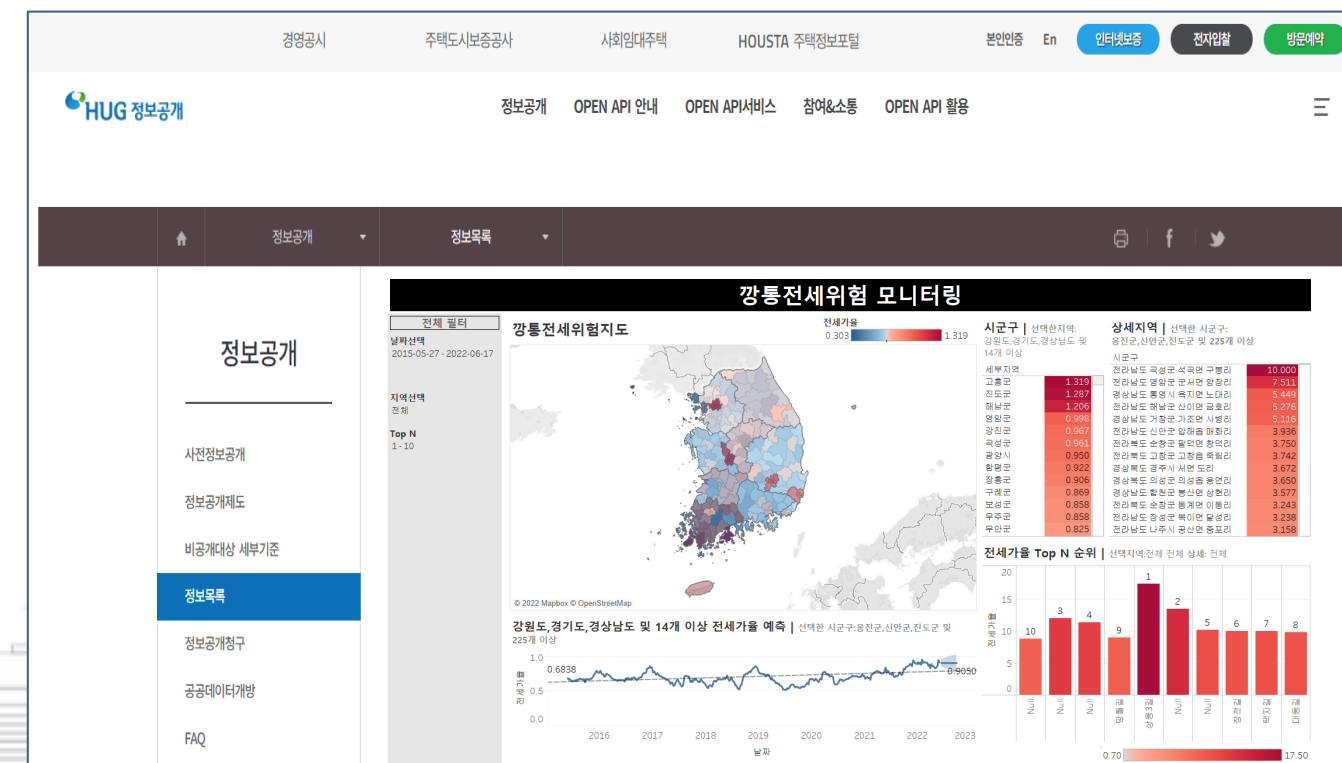
PROPOSE

직관적인 대시보드 게시

AS-IS



TO-BE



전세역전

2022 국토교통 데이터 활용 경진대회

임차자 보호를 위한 강통전세 발생 가능지역 및 주택유형별 예측분석

감사합니다.

