

주제: 서울시 부동산 공공데이터 분석을 통한 전세가 책정 요인 탐구

I. 서론

1. 데이터 소개

대한민국 내 인구밀집도가 가장 높고 주거난이 가장 심각한 서울의 2023 년 [부동산 데이터](#)를 살펴보았다. 해당 데이터는 서울시 부동산 전월세가 정보로 이루어져 있으나, 전세가 책정 요인을 위주로 데이터분석을 진행하기로 했다. 월세가는 소멸되는 금액인 월세와 환급되는 금액인 보증금으로 구성되어 있기 때문에, 하나의 수치로 '가격'을 표시하는 것은 데이터의 부정확한 해석이기 때문이다. 기회가 된다면 보증금에 비해 월세 비중이 높은 부동산의 특징 정도는 분석 가능할 것으로 예상된다.

또한, 데이터 양이 워낙 많기도 하고 부동산 특성상 시세가 빠르게 바뀌기 때문에 2023 년 데이터만을 사용하여 이번 분석을 진행할 것이다.

II. 가설 설정

1. 가설

가격이 비싼 부동산의 특징은 따로 존재하며, 그 특징은 아래와 같다.

- a. 위치가 강남 쪽이다
- b. 층수가 높다
- c. 면적이 넓다
- d. 건축연도가 최근이다
- e. 건물유형은 오피스텔이다
- f. 신규계약이다
- g. 계약기간이 길다

2. 가설검증을 위해 진행할 사항

- a. 데이터 전처리
- b. 시각화
- c. 심층분석 (회귀분석)
- d. 결론 (가설 검정, 분석 인사이트 및 솔루션 제시)

III. 데이터 전처리

1. 불필요한 데이터 제거

먼저 데이터를 불러온 후, 전세와 월세가 아닌 다른 구분이 존재하는지 확인했다. 다른 형태의 부동산계약은 없는 것으로 확인되었고, 전월세구분에서 '전세'값만 살려 데이터프레임을 새로 만들었다. 불필요한 칼럼을 제거했으며, 이에는 접수년도, 전월세구분, 종전임대료, 월세(만원), 본번, 부번, 건물명이 포함되었다. 전월세구분과 월세(만원), 종전임대료 칼럼은 전세 데이터만을 볼 것이기 때문에 모든 행들이 같은 값을 가지고 있어 불필요했기 때문이다. 아울러, 본번, 부번데이터는 이는 정확한 주소에 대한 것이라, 가설 검증에 적합한 데이터가 아니라 삭제처리했다. 자치구와 법정동 정도의 주소정보를 통해 가설을 검증하는 것이 더 적합할 것으로 판단했기 때문이다. 마지막으로, 건물명 칼럼은 실제 건물명이 아니라 "(본번-부번)"의 형식으로 기재되어 있었기에 삭제처리했다. 그 결과, 데이터는 아래와 같았다.

	자치구코드	자치구명	법정동코드	법정동명	지번구분코드	지번구분	층	계약일	임대면적	보증금(만원)	건축년도	건물용도	계약기간	신규계약구분	경신청구권사용	종전보증금
0	11440	마포구	11000	노고산동	1.0	대지	9.0	20230216	33.66	21000	1998.0	오피스텔	23.02~24.02	신규	NaN	0.0
1	11170	용산구	11300	원효로2가	1.0	대지	9.0	20230818	71.67	23900	1990.0	오피스텔	23.09~25.09	신규	NaN	0.0
2	11440	마포구	11000	노고산동	1.0	대지	9.0	20230216	41.04	22000	1998.0	오피스텔	NaN	NaN	NaN	NaN
3	11620	관악구	10300	남현동	1.0	대지	14.0	20230708	28.05	16500	2005.0	오피스텔	23.08~24.08	신규	NaN	0.0
4	11560	영등포구	11700	당산동	1.0	대지	6.0	20230111	14.40	9923	2011.0	오피스텔	23.02~25.02	신규	NaN	0.0
...
262467	11380	은평구	10700	응암동	NaN	NaN	NaN	20231207	82.88	33000	2023.0	단독다가구	23.12~25.12	신규	NaN	0.0
262468	11230	동대문구	10600	장안동	NaN	NaN	NaN	20231207	84.63	20000	1991.0	단독다가구	23.12~27.12	신규	NaN	0.0
262469	11140	중구	12900	남산동2가	NaN	NaN	NaN	20231207	104.85	38000	1997.0	단독다가구	24.01~26.01	신규	NaN	0.0
262470	11110	종로구	10100	청운동	NaN	NaN	NaN	20231207	123.36	60000	1991.0	단독다가구	23.12~25.12	신규	NaN	0.0
262471	11215	광진구	10400	광장동	NaN	NaN	NaN	20231207	138.99	105000	NaN	단독다가구	24.02~26.02	경신	○	110000.0

262472 rows x 16 columns

2. 시계열 데이터 처리

'계약일' 칼럼을 datetime 으로 변환했다. 또한, '계약기간' 칼럼이 '계약시작연월~계약종료연월' 형식으로 되어 있어 이를 분리했다. 계약시작 및 종료 연월에 대한 칼럼을 새로 만든 이유는 계약체결시기에 따른 부동산 가격 변동이 있을 수도 있어 추후 분석해볼 여지가 있기 때문이다. 새로 만든 두 칼럼을 datetime 으로 변환했고, 두 칼럼의 차를 통해 계약기간(월 단위)를 구해 기존의 '계약기간' 칼럼을 대체했다. 이를 통해 계약이 체결된 시기뿐만이 아니라 계약 기간 자체의 영향도 추후 분석할 수 있을 것으로 보인다. 즉, 시계열 데이터까지 처리한 데이터프레임은 아래와 같았다.

	자치구코드	자치구명	법정동코드	법정동명	지번구분코드	지번구분	층	계약일	임대면적	보증금(만원)	건축년도	건물용도	계약기간	신규계약구분	갱신청구권사용	종전보증금	계약시작연월	계약종료연월	
	0	11440	마포구	11000	노고산동	1.0	대지	9.0	2023-02-16	33.66	21000	1998.0	오피스텔	12.0	신규	NaN	0.0	2023-02-01	2024-02-01
	1	11170	용산구	11300	원효로2가	1.0	대지	9.0	2023-08-18	71.67	23900	1990.0	오피스텔	24.0	신규	NaN	0.0	2023-09-01	2025-09-01
	2	11440	마포구	11000	노고산동	1.0	대지	9.0	2023-02-16	41.04	22000	1998.0	오피스텔	NaN	NaN	NaN	NaN	NaT	NaT
	3	11620	관악구	10300	남현동	1.0	대지	14.0	2023-07-08	28.05	16500	2005.0	오피스텔	12.0	신규	NaN	0.0	2023-08-01	2024-08-01
	4	11560	영등포구	11700	당산동	1.0	대지	6.0	2023-01-11	14.40	9923	2011.0	오피스텔	24.0	신규	NaN	0.0	2023-02-01	2025-02-01
	
262467	11380	은평구	10700	을암동	NaN	NaN	NaN	2023-12-07	82.88	33000	2023.0	단독다가구	24.0	신규	NaN	0.0	2023-12-01	2025-12-01	
262468	11230	동대문구	10600	장안동	NaN	NaN	NaN	2023-12-07	84.63	20000	1991.0	단독다가구	48.0	신규	NaN	0.0	2023-12-01	2027-12-01	
262469	11140	종로구	12900	남산동2가	NaN	NaN	NaN	2023-12-07	104.85	38000	1997.0	단독다가구	24.0	신규	NaN	0.0	2024-01-01	2026-01-01	
262470	11110	종로구	10100	청운동	NaN	NaN	NaN	2023-12-07	123.36	60000	1991.0	단독다가구	24.0	신규	NaN	0.0	2023-12-01	2025-12-01	
262471	11215	광진구	10400	광장동	NaN	NaN	NaN	2023-12-07	138.99	105000	NaN	단독다가구	24.0	갱신	○	110000.0	2024-02-01	2026-02-01	

262472 rows × 18 columns

3. 결측값 처리

```

자치구코드      0
자치구명        0
법정동코드      0
법정동명        0
지번구분코드    37079
지번구분        37079
층              37044
계약일          0
임대면적        0
보증금(만원)    0
건축년도        3355
건물용도        0
계약기간        47034
신규계약구분    42374
갱신청구권사용  239704
종전보증금      42375
계약시작연월    47032
계약종료연월    47034
dtype: int64

```

처음 관측된 결측값들은 위와 같았다. 결측값이 나타난 칼럼을 순서대로 처리하고자 했는데, 칼럼 간 관계를 고려해야 했다. '지번구분코드'와 '지번구분'은 같은 정보를 표현하는 변수로, '대지', '산', '블록'의 형태를 나타냈다.

```

지번구분코드
1.0    225215
2.0     149
3.0     29
Name: count, dtype: int64
지번구분
대지    225215
산       149
블록     29
Name: count, dtype: int64

```

관측되는 값의 개수가 같음을 확인했다.

또 다른 범주형 변수인 '신규계약구분'은 '신규'와 '갱신' 두 가지 값을 가질 수 있었다. 결측값을 처리하기 전, '신규계약구분'과 '중전보증금' 칼럼 간 관계가 있을 것 같아 살펴보았다. 신규계약인 경우 중전보증금이 당연히 0 이지만, 갱신계약인 경우에는 중전보증금이 존재하는 것이 상식적이기 때문이다. 데이터 확인 후, '신규'계약은 모두 0 으로 처리했다. 이를 모두 확인한 후, 두 개의 범주형 변수는 최빈값 대체로 결측값을 처리했다. '신규계약구분'에서의 결측값이 모두 '신규'로 대체되었기 때문에, 이에 따른 '중전보증금' 결측값도 모두 0 으로 처리했다. '갱신'계약 중 중전보증금이 결측값인 경우가 없었기에 가능한 처리였다.

수치형 데이터인 '건축년도'와 '층'은 중앙값에 따르기로 했다. 아래 이상치 그래프에서도 확인 가능할 것인데, 평균을 너무 치우쳐지게 하는 이상치들이 섞여 있기에 중앙값을 선택했다. 참고로, interpolate 나 bfill, ffill 등의 보간법을 선택하지 않은 이유는 현재 분석중인 데이터가 시계열 데이터가 아니며, 서로 앞뒤에 위치한 행들 간 관계가 없기 때문이다. KNN 등 보다 정교한 방식으로 결측값을 처리할 수도 있겠으나, 데분기 과정에서 배운 내용에 충실하고자 단순 0, 평균값, 중앙값, 최빈값 대체로 진행하고 있는 것이다.

데이터를 다시 확인해보니 '갱신청구권'은 갱신청구권이 있을 경우 '○'가 기입되어 있으며 나머지 행들은 결측값으로 처리되어 있다. (value_counts 로 확인해보면 결측값이 아닌 경우는 예외없이 '○'값이 기입되어 있다.) 따라서 결측값은 0, '○'이 있는 경우는 1 로 대체되도록 했다.

마지막으로 확인해보니 계약기간에 대한 칼럼들만 결측값이 있다. '계약기간'만 분석할 것이라면 평균 혹은 중앙값 대체가 가능할지도 모르지만, 파생된 두 칼럼 '계약시작연월'과 '계약종료연월'까지 이런 방식으로 처리하기에는 한계가 있어 보였다. 앞으로의 정확한 분석에 더 도움이 되는 방안이라 판단해 결측값들을 드랍했다. 전세 관련 데이터가 26 만개 이상 있으므로, 약 5 만개 정도 되는 결측값들을 드랍하더라도 21 만개가 넘는 데이터가 남아있기 때문에 분석에 큰 문제가 생기지는 않을 것으로 판단된다.

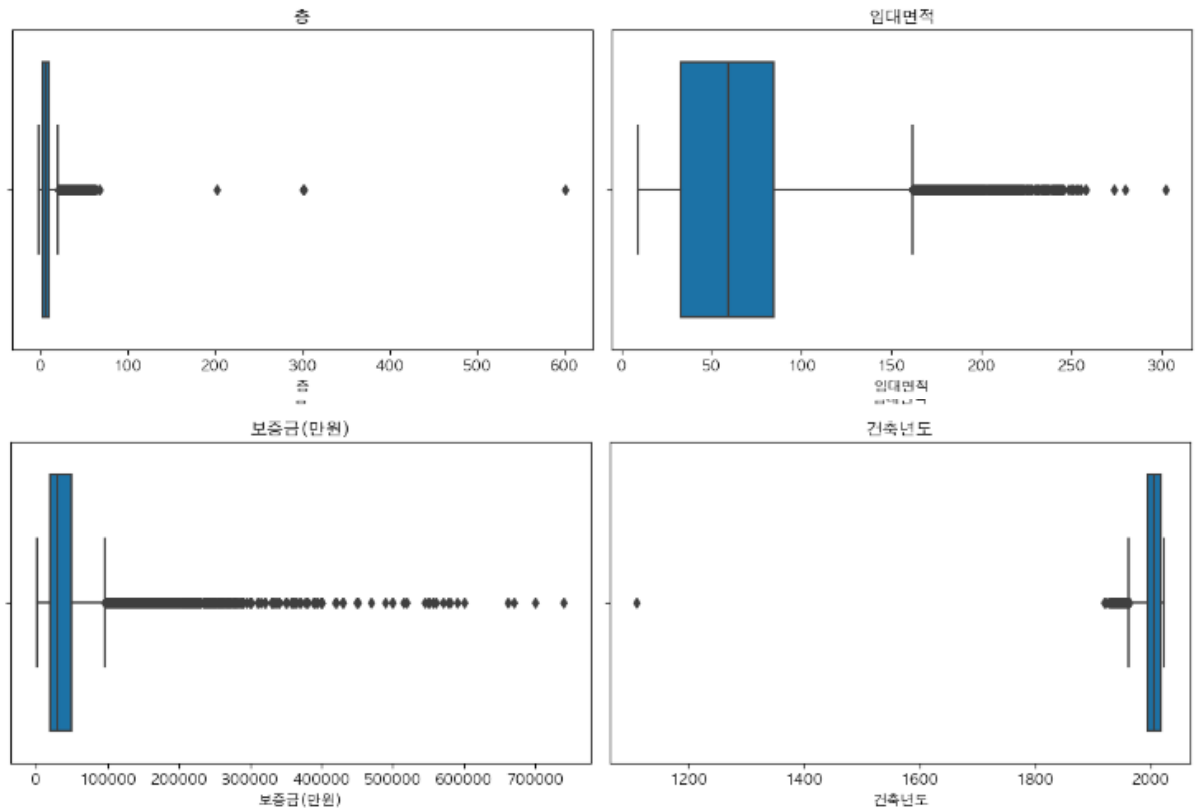
최종적으로 결측값을 처리한 데이터프레임은 다음과 같다.

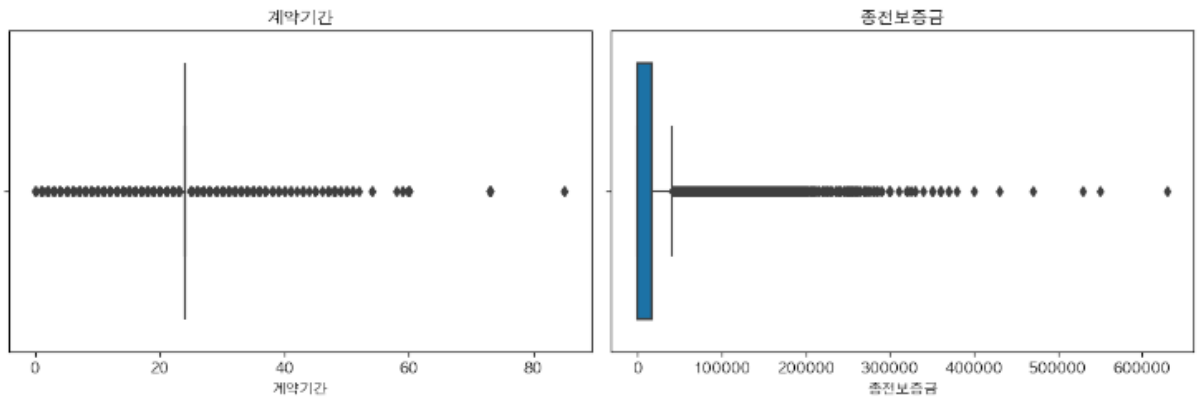
	자치구코드	자치구명	법정동코드	법정동명	지번구분코드	지번구분	층	계약일	임대면적	보증금(만원)	건축년도	건물용도	계약기간	신규계약구분	갱신청구권사용	중전보증금	계약시작연월	계약종료연월
0	11440	마포구	11000	노고산동	1.0	대지	9.0	2023-02-16	33.66	21000	1998.0	오피스텔	12.0	신규	0	0.0	2023-02-01	2024-02-01
1	11170	용산구	11300	원효로2가	1.0	대지	9.0	2023-08-18	71.67	23900	1990.0	오피스텔	24.0	신규	0	0.0	2023-09-01	2025-09-01
3	11620	관악구	10300	남현동	1.0	대지	14.0	2023-07-08	28.05	16500	2005.0	오피스텔	12.0	신규	0	0.0	2023-08-01	2024-08-01
4	11560	영등포구	11700	당산동	1.0	대지	6.0	2023-01-11	14.40	9923	2011.0	오피스텔	24.0	신규	0	0.0	2023-02-01	2025-02-01
5	11530	구로구	10200	구로동	1.0	대지	5.0	2023-03-19	16.80	12000	2006.0	오피스텔	24.0	갱신	0	12000.0	2023-04-01	2025-04-01
...
262467	11380	은평구	10700	응암동	1.0	대지	6.0	2023-12-07	82.88	33000	2023.0	단독다가구	24.0	신규	0	0.0	2023-12-01	2025-12-01
262468	11230	동대문구	10600	장안동	1.0	대지	6.0	2023-12-07	84.63	20000	1991.0	단독다가구	48.0	신규	0	0.0	2023-12-01	2027-12-01
262469	11140	중구	12900	남산동2가	1.0	대지	6.0	2023-12-07	104.85	38000	1997.0	단독다가구	24.0	신규	0	0.0	2024-01-01	2026-01-01
262470	11110	종로구	10100	청운동	1.0	대지	6.0	2023-12-07	123.36	60000	1991.0	단독다가구	24.0	신규	0	0.0	2023-12-01	2025-12-01
262471	11215	광진구	10400	광장동	1.0	대지	6.0	2023-12-07	138.99	105000	2005.0	단독다가구	24.0	갱신	1	110000.0	2024-02-01	2026-02-01

215438 rows x 18 columns

4. 이상치 처리

이상치의 존재를 확인해보기 위해 연속변수 모두 박스플롯으로 시각화했다.





역시 확인해보니 총이 600 개가 넘는 경우, 건축년도가 1200 년 이전인 경우와 같이 상식을 벗어나는 이상치들이 보였다. 앞으로 데이터를 분석하고, 칼럼과 전세가 간 비교를 진행해야 하기 때문에 이상치는 버리기로 했다. 다만, '보증금(만원)'과 '중전보증금'은 가설에서 살펴보고자 하는 전세가 그 자체이기 때문에 따로 이상치 처리를 하지 않았다. 유독 높은 전세가를 가진 부동산에 대해서도 따로 분석을 하는 것이 본 연구의 취지와 맞다고 생각했기 때문이다.

	자치구코드	자치구명	법정동코드	법정동명	지번구분코드	지번구분	층	계약일	임대연적	보증금(만원)	건축년도	건물용도	계약기간	신규계약구분	갱신청구권사용	중전보증금	계약시작연월	계약종료연월
1	11170	용산구	11300	원효로2가	1.0	대지	9.0	2023-08-18	71.67	23900	1990.0	오피스텔	24.0	신규	0	0.0	2023-09-01	2025-09-01
4	11560	영등포구	11700	당산동	1.0	대지	6.0	2023-01-11	14.40	9923	2011.0	오피스텔	24.0	신규	0	0.0	2023-02-01	2025-02-01
5	11530	구로구	10200	구로동	1.0	대지	5.0	2023-03-19	16.80	12000	2006.0	오피스텔	24.0	갱신	0	12000.0	2023-04-01	2025-04-01
6	11560	영등포구	11700	당산동	1.0	대지	6.0	2023-01-30	14.88	10500	2011.0	오피스텔	24.0	신규	0	0.0	2023-03-01	2025-03-01
7	11260	종로구	10400	목동	1.0	대지	3.0	2023-05-15	84.71	46000	2003.0	아파트	24.0	신규	0	0.0	2023-07-01	2025-07-01
...
262466	11290	성북구	13500	종암동	1.0	대지	6.0	2023-12-07	81.99	19000	1995.0	단독다가구	24.0	신규	0	0.0	2024-01-01	2026-01-01
262467	11380	은평구	10700	응암동	1.0	대지	6.0	2023-12-07	82.88	33000	2023.0	단독다가구	24.0	신규	0	0.0	2023-12-01	2025-12-01
262469	11140	중구	12900	남산동2가	1.0	대지	6.0	2023-12-07	104.85	38000	1997.0	단독다가구	24.0	신규	0	0.0	2024-01-01	2026-01-01
262470	11110	종로구	10100	청운동	1.0	대지	6.0	2023-12-07	123.36	60000	1991.0	단독다가구	24.0	신규	0	0.0	2023-12-01	2025-12-01
262471	11215	광진구	10400	광장동	1.0	대지	6.0	2023-12-07	138.99	105000	2005.0	단독다가구	24.0	갱신	1	110000.0	2024-02-01	2026-02-01

197490 rows x 18 columns

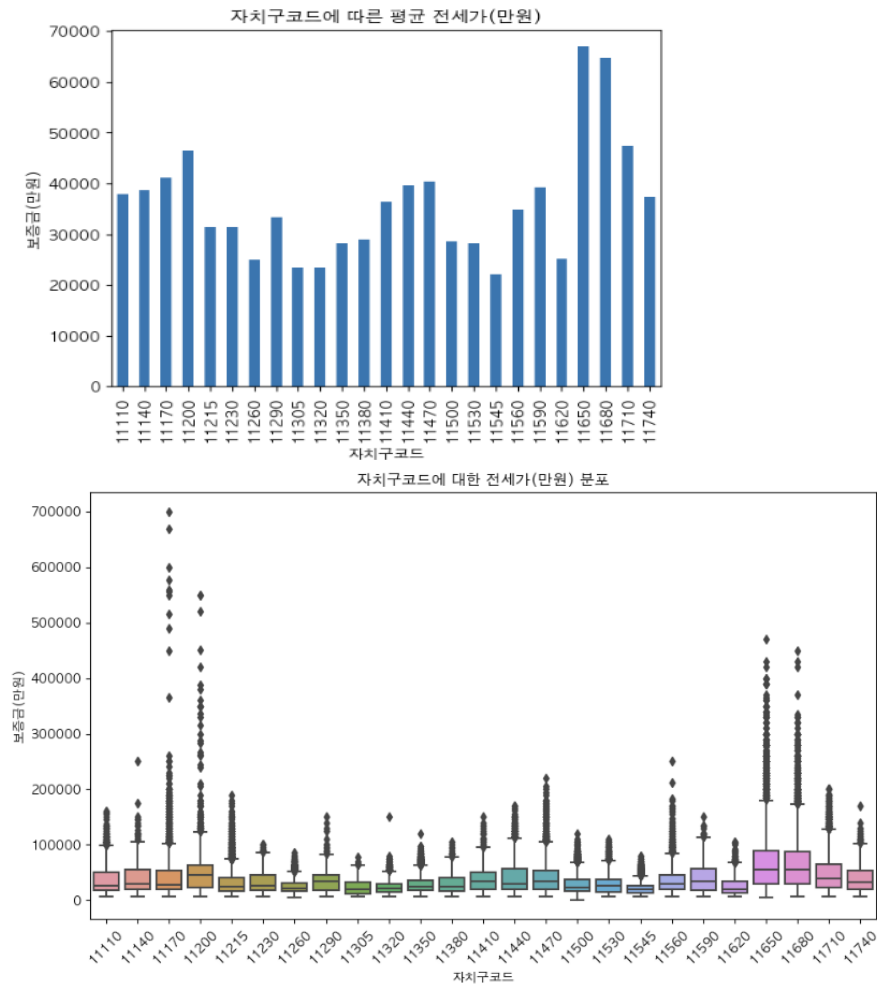
따라서, 최종적으로 전처리를 마친 데이터는 위와 같다.

IV. 시각화

이제부터는 앞에서 말한 가설을 증명하기 전에 보증금과 다른 컬럼 간의 관계를 시각화를 통해 확인해 볼 것이다.

1. 지역자치구에 따른 전세가

먼저 지역에 따른 전세가를 보증금의 평균값을 통해 시각화해보겠다.

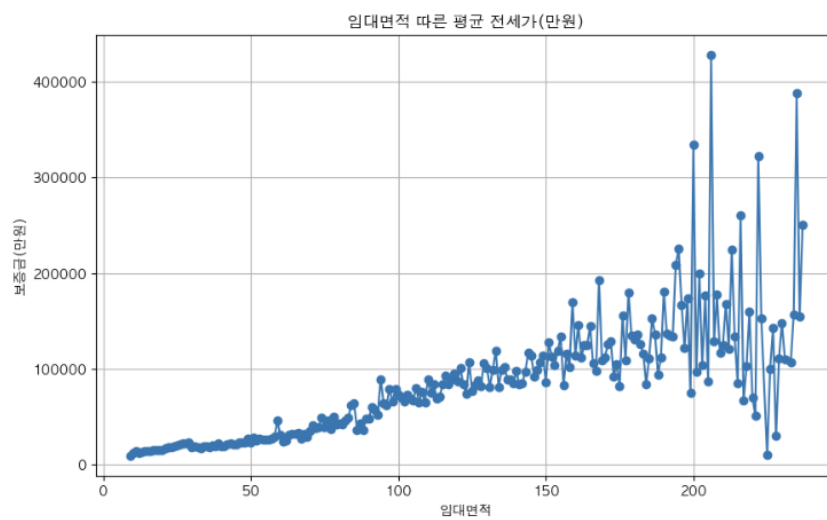
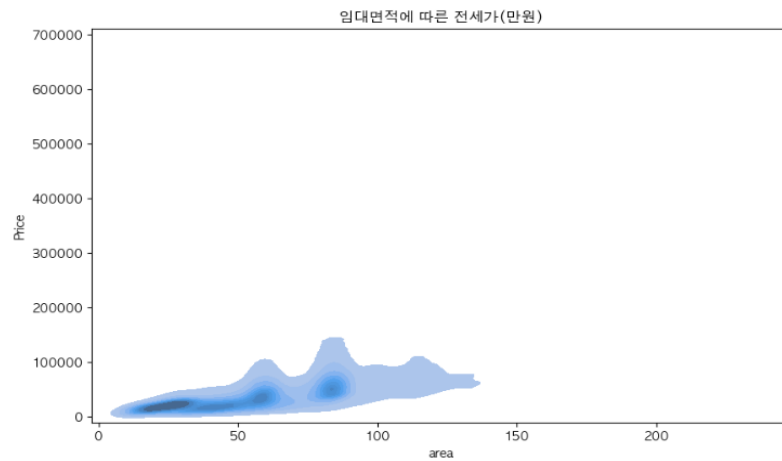


위 그래프를 분석해 보았을때 11650(서초구),11680(강남구)가 평균적으로 높은 보증금을 가지는 것으로 확인됐다.

추가로 위 박스 플롯과 히스토그램을 비교해 보았을때 11170(용산구)는 평균적인 집값은 다른 지역들과 비슷하나 이상치의 개수가 다른 지역보다 특히 많아 히스토그램에서의 가격 평균 값이 높게 표시된 것으로 보인다.

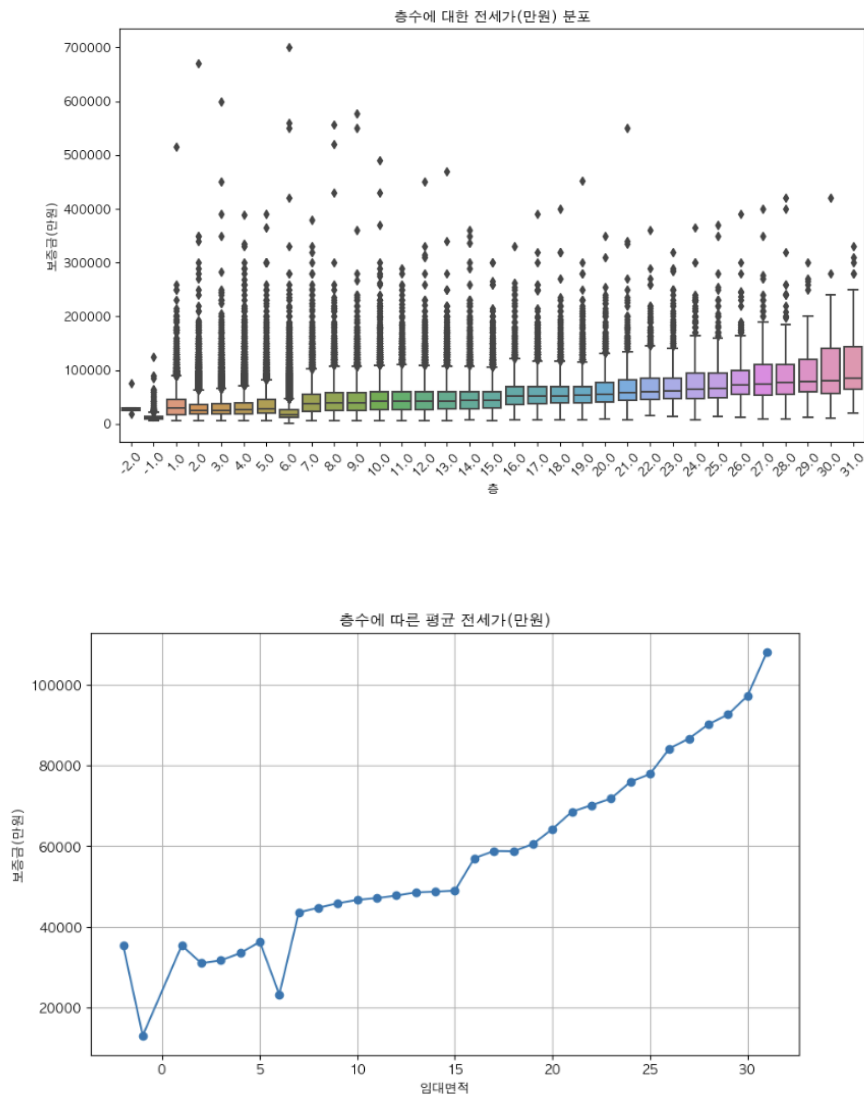
2. 임대 면적에 따른 전세가

그다음으로는 임대 면적에 따른 전세가를 비교해 보겠다.



위 그래프를 분석해 보았을 때 임대면적이 약 150 이하 인 집은 임대면적과 전세가 사이의 유의미한 관계가 있는 것으로 보인다.

3. 층에 따른 전세가

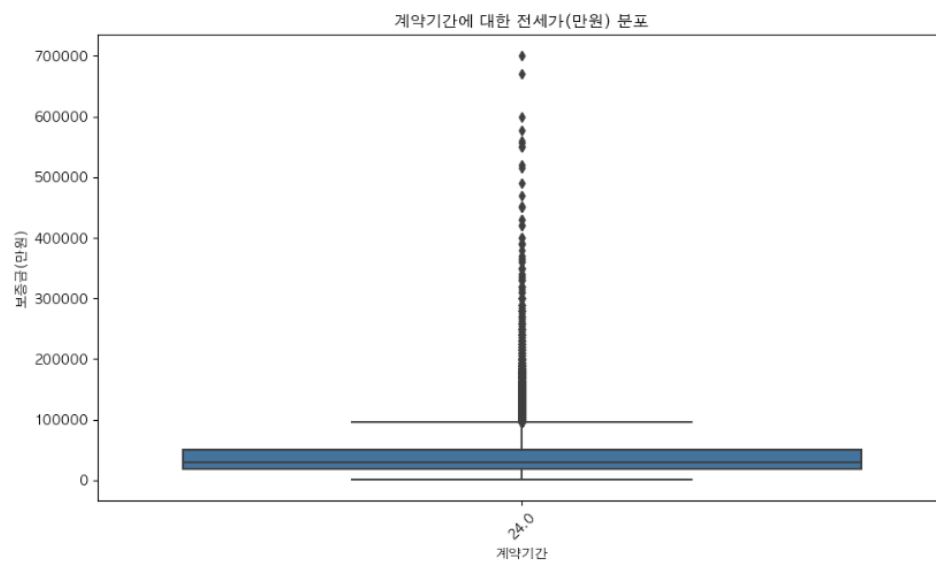
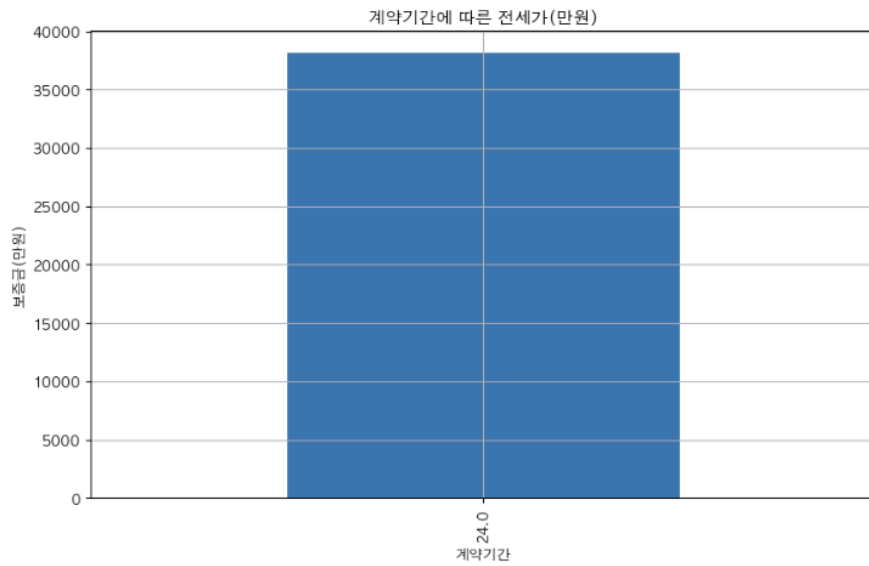


그다음으로는 집의 층수와 전세가 사이의 관계를 분석해 보았다.

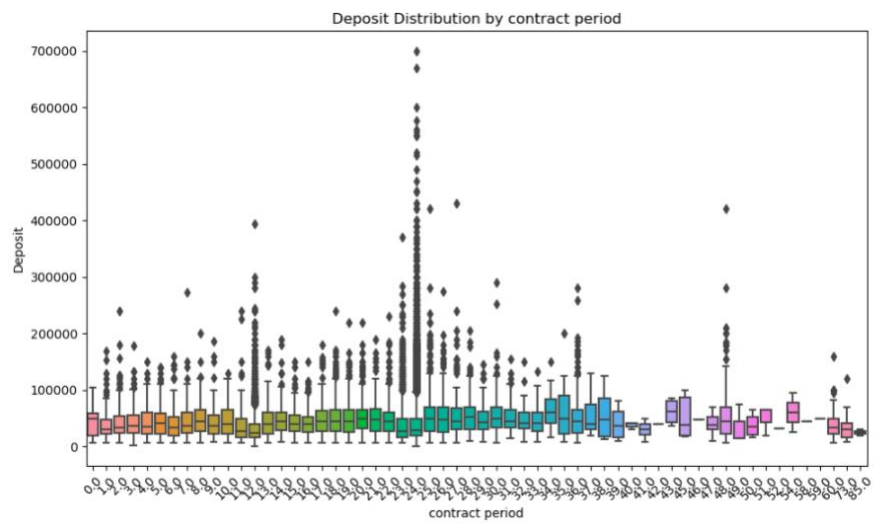
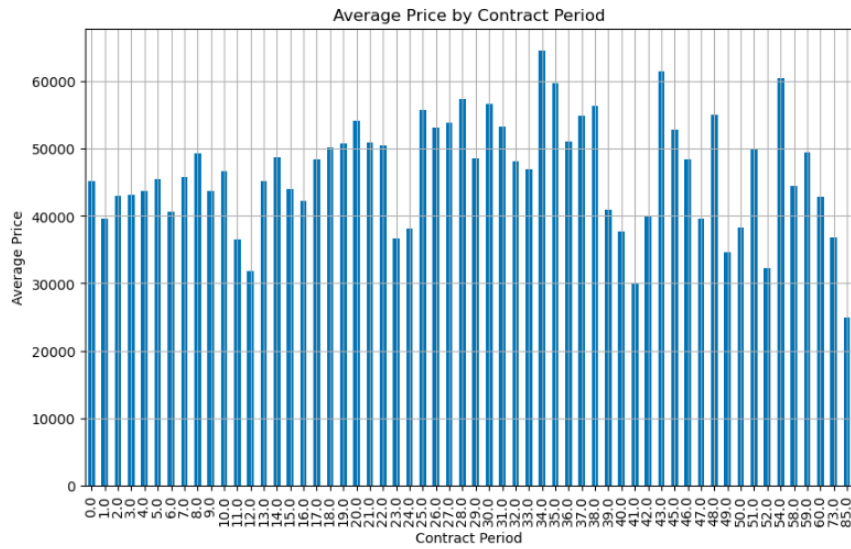
집의 층수와 보증금의 관계를 살펴보면 10 층 전까지는 비례관계를 확인할 수 없으나 15 층 이상의 집에 대해서는 층수와 집값 간의 비례관계를 확인할 수 있다.

4. 계약 기간에 따른 전세가

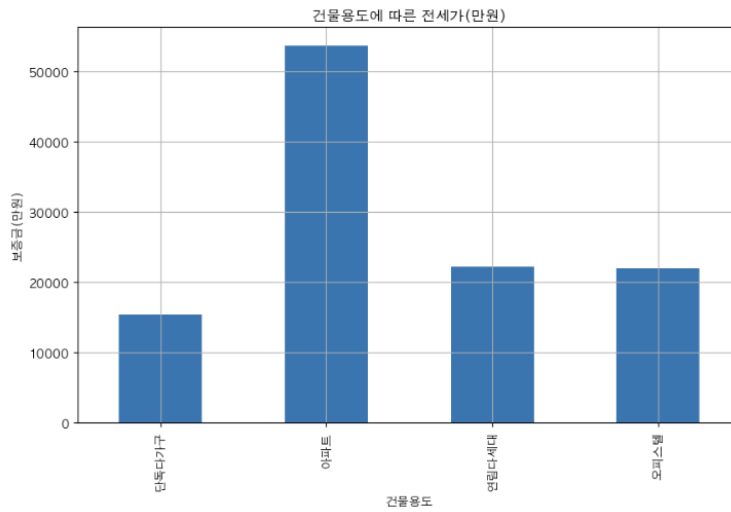
그 다음은 계약기간과 전세가 사아 관계를 시각화 해보았으나. 이상치를 제거하니 24 개월 계약에 해당되는 데이터만 남아 계약기간의 데이터가 분포되어 있지 않아 관계를 확인하기 어려웠다.

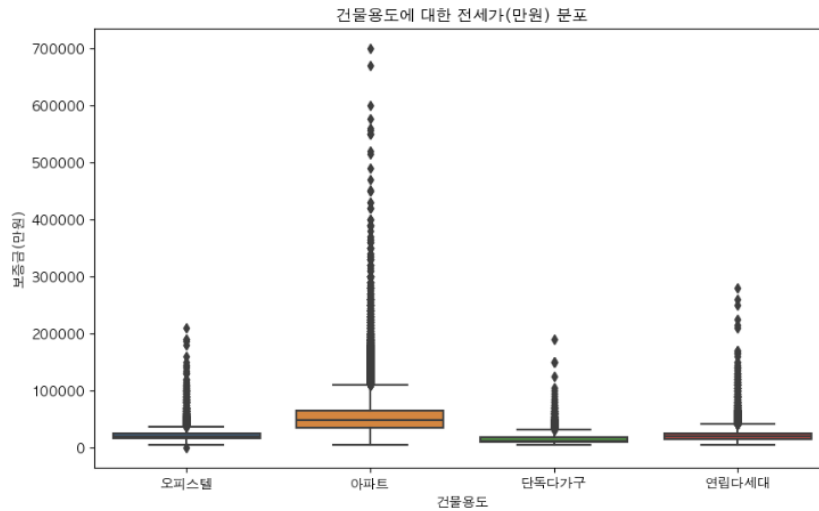


그래서 이상치를 제거하지 않은 데이터를 사용하여 계약 기간과 집값의 관계를 시각화해보았다. 하지만 두 컬럼 사이에 어떠한 유의미한 관계가 있다고 보긴 어려운 결과가 확인되었다.



5. 건물 용도에 따른 전세가

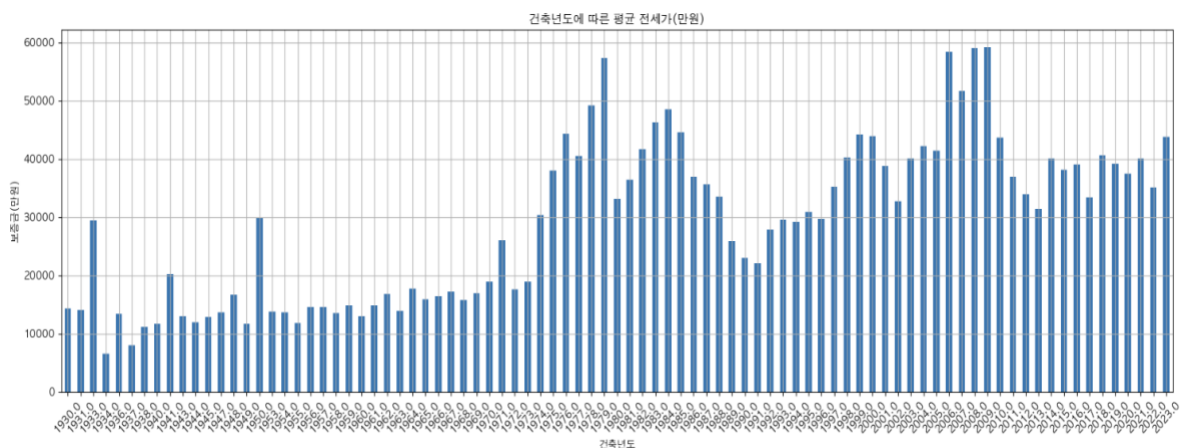


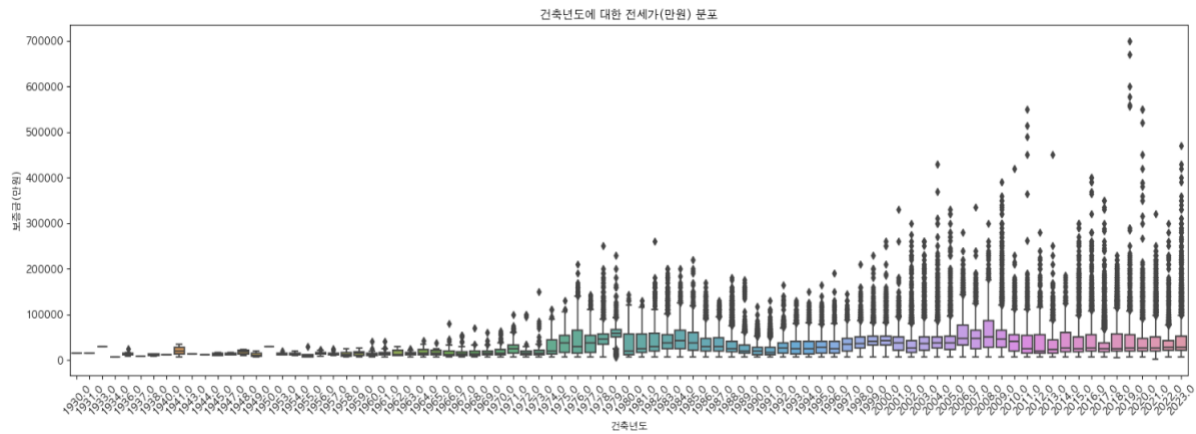


건물 용도에 따른 전세가를 보았을 때는 보증금 평균값의 차이는 크지 않으나 위 박스플롯 그래프와 히스토그램을 비교해 보았을 때 아파트의 경우 이상치의 분포가 다른 집의 형태보다 확장되어 있어서 히스토그램에서의 평균값은 다른 형태의 집의 보증금과 많이 차이가 나는 것으로 보인다.

6. 건축 연도에 따른 전세가

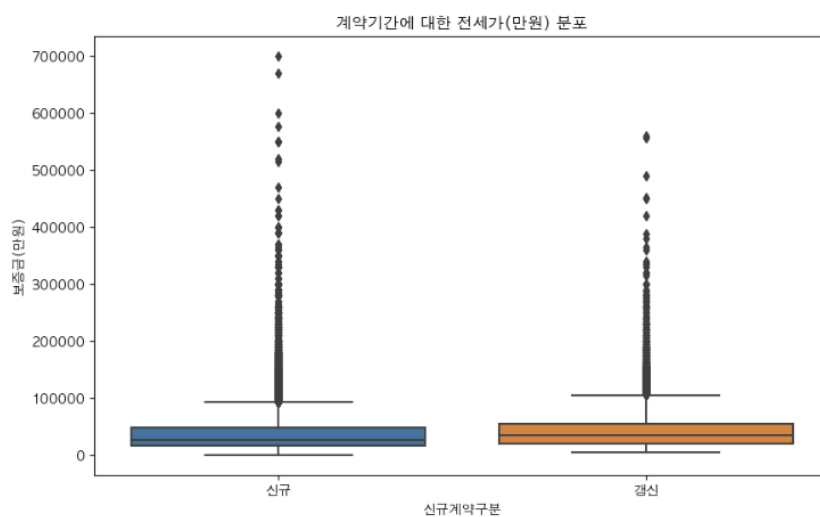
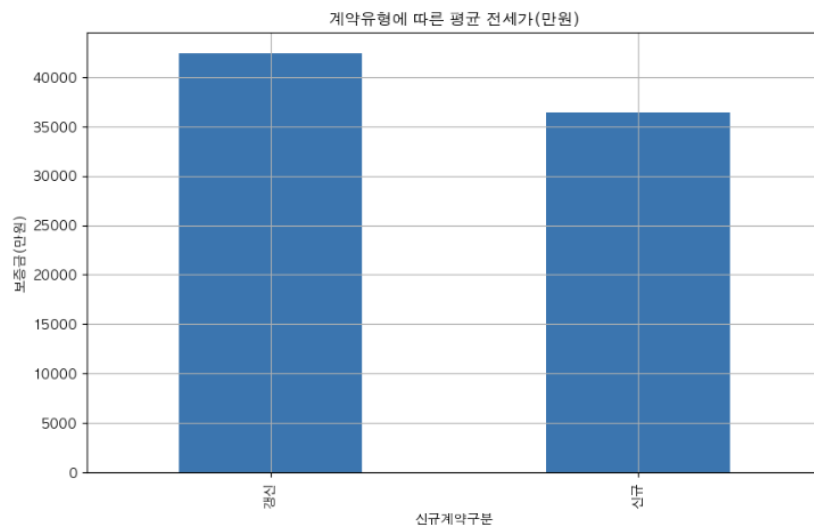
건축연도에 따른 집값의 분포를 확인해 보겠다. 아래 그래프를 보면 대략 1970 년대를 기준으로 가격의 평균값이 차이가 보이지만 그 이상의 관계를 확인하기는 어려워 보인다. 히스토그램에서는 건축연도와 가격이 양의 관계를 가지는 것처럼 보이나, 박스플롯 그래프를 통해 이는 특정 이상치에 의한 것임을 확인할 수 있다.





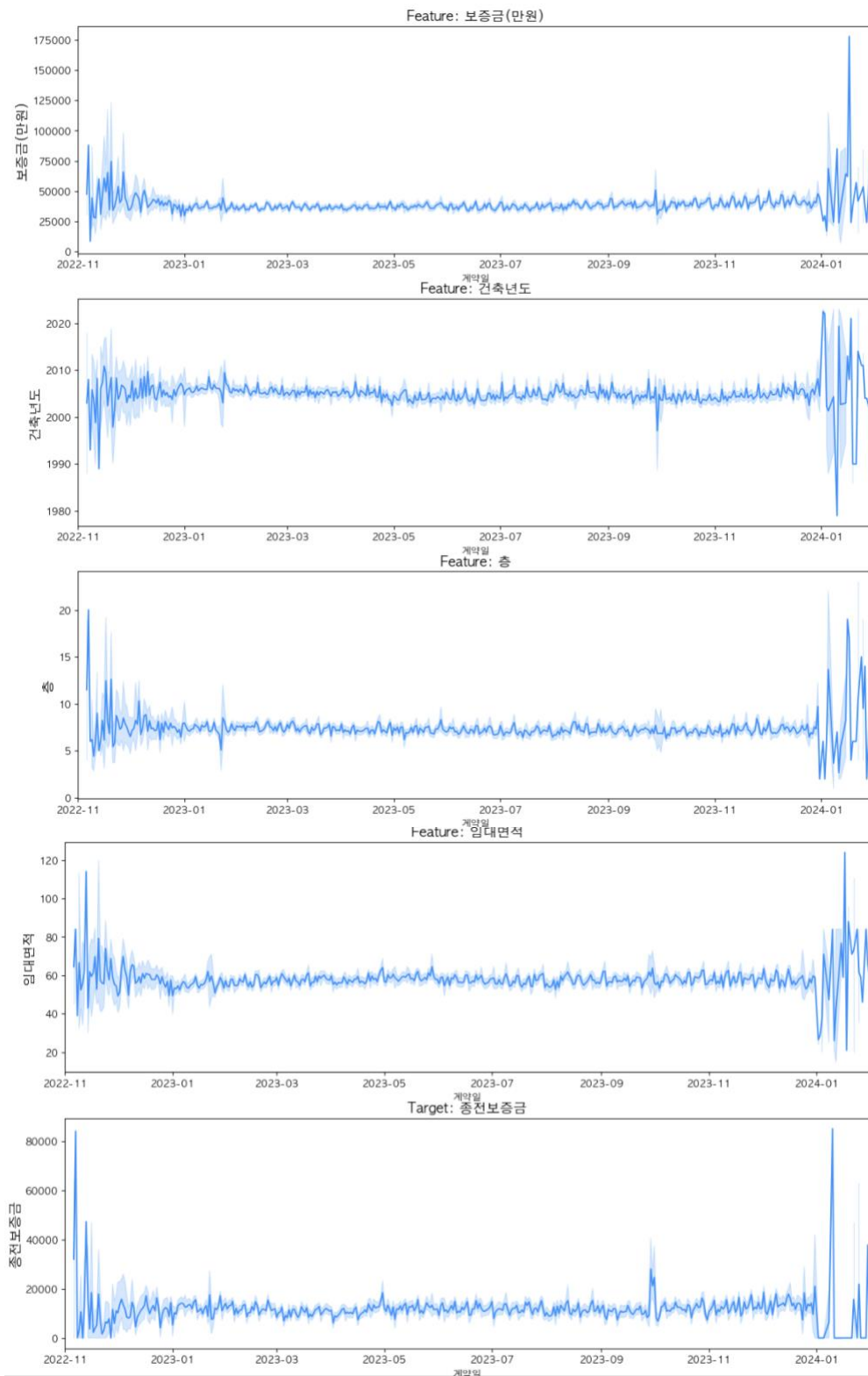
7. 계약 유형에 따른 전세가

아래의 그래프를 보면 신규 계약의 집과 계약을 갱신하는 집의 보증금 평균값이 아주 미세하게 차이를 보이는 것을 확인할 수 있다.



V. 심층 분석

1. 시계열 분석; 계약일에 따른 전세가, 건축년도, 층, 임대면적, 종전 보증금

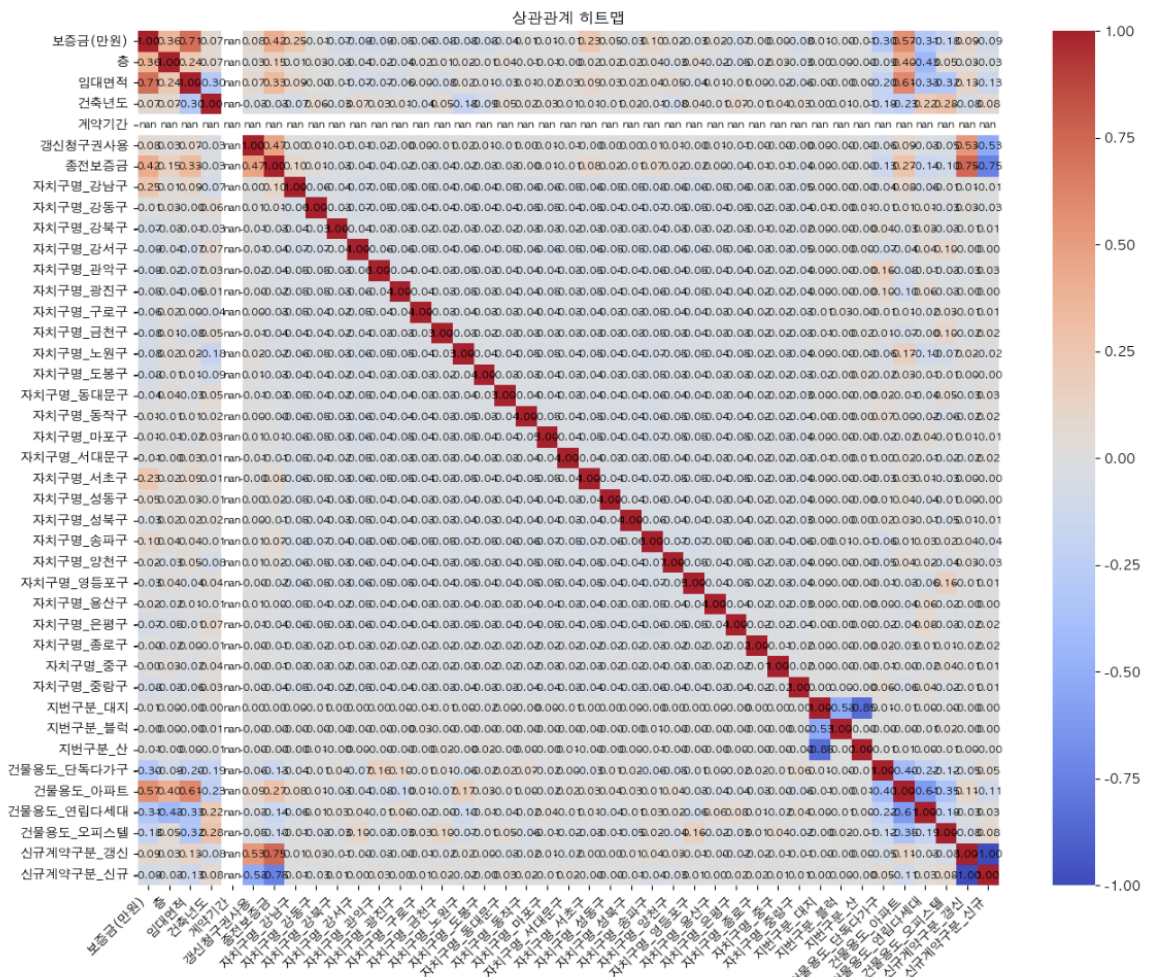


주로 연말, 연초에 활발한 부동산 거래가 이루어지는 것을 알 수 있었으며, 거래량이 활발한 시기에는 대체로 높은 전세가로 거래되었다.

2. 상관계수 분석; 칼럼과의 관계

‘자치구코드’, ‘법정동코드’, ‘법정동명’, ‘지번구분코드’, ‘지번구분’, ‘계약일’, ‘계약시작연월’, ‘계약종료연월’ 등 가설검증에 필요없다고 판단된 컬럼을 날린 후, 원-핫 인코딩을 진행하였다.

이후, 상관관계에 대한 분석을 히트맵으로 시각화하였다.



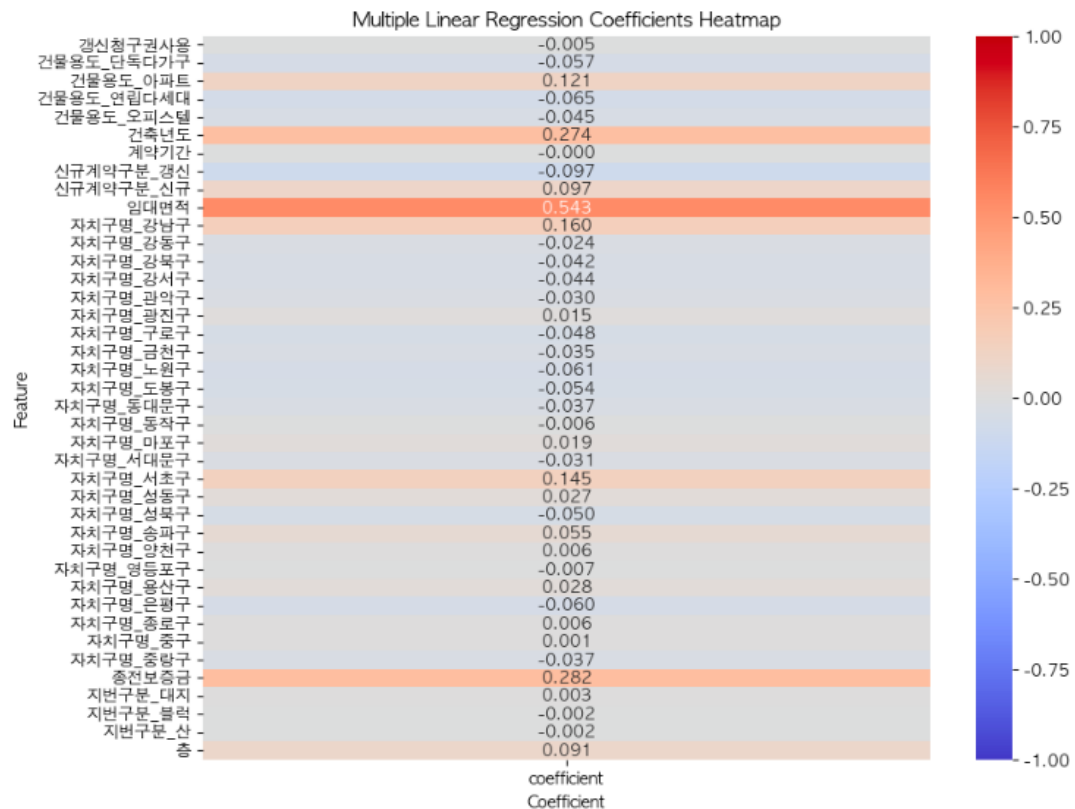
다음은 '보증금(만원)' 과 상관관계가 높다고 판단되는 컬럼이다.

- a. 층 0.363431
b. 임대면적 0.712078
c. 종전보증금 0.420855
d. 자치구명_강남구 0.246167
e. 자치구명_서초구 0.231204
f. 건물용도_아파트 0.572068

따라서, 전세가는 층, 임대면적, 중전보증금에 영향을 받으며, 해당 a,b,c 의 값이 클수록 전세가가 높다. 지역구로는 강남구, 서초구가 전세가가 높았고, 건물용도는 아파트가 전세가가 높다.

추가적으로 변수를 스케일링해본 후 회귀모델을 진행한 결과, 비슷한 결론에 다다를 수 있었다.

OLS Regression Results			
Dep. Variable:	보증금(만원)	R-squared (uncentered):	0.770
Model:	OLS	Adj. R-squared (uncentered):	0.770
Method:	Least Squares	F-statistic:	1.889e+04
Date:	Sun, 25 Feb 2024	Prob (F-statistic):	0.00
Time:	14:30:31	Log-Likelihood:	-1.3509e+05
No. Observations:	197490	AIC:	2.702e+05
Df Residuals:	197455	BIC:	2.706e+05
Df Model:	35		
Covariance Type:	nonrobust		



이때에도 임대면적과 종전보증금이 높게 나타났고, 건축년도 같은 경우가 새로이 관측된 유의미한 변수라고 할 수 있다.

VI. 결론

1. 가설 검정 및 분석 인사이트

전세가가 높은 부동산의 공통적 특성이 따로 존재하는 것으로 확인했고, 가설이 참으로 드러났다. 구체적인 변수들 역시 가설에서 예측했던 바와 매우 비슷했다. 그 중 통계적으로 가장 유의미한 것은 면적, 층, 종전보증금 등으로 드러났다. 이외에도 강남서초 전세가가 비교적 높은 것도 유의하여 볼 만하다. 또한, 단독다가구, 연립다세대, 오피스텔보다 아파트의 전세가가 비교적 높았다.

2. 솔루션

‘서울시 주거난’에 관련된 솔루션을 제시하기 위해서 높은 전세가의 요인들을 분석해왔다. 특히 높은 집값에 큰 영향을 받을 사회초년생을 생각하면, 면적이 비교적 작더라도 감당 가능한 수준의 전세가를 책정하는 방향으로 부동산 시장이 변화하는 솔루션이 있을 것으로 판단된다. 건물유형 역시 면적이 클 확률이 높고 가격 편차가 큰 아파트 형태에 비해 오피스텔 형태가 더 적합할 것으로 사료된다. 이외에도 특정 자치구(강남구, 서초구)에서의 높은 집값이 확인되는 것으로 보아, 서울시 내에서도 지역 간 편차를 줄이기 위한 노력이 필요해 보인다. 여기서 말하는 편차는 본 연구에서 확인한 층, 건축년도 등의 요인을 뜻하는 것이 아니다. 부동산의 가격 책정에는 교통, 인프라 등 다양한 요소가 있기 때문에 이런 부분들을 행정적 차원에서도 지원하여 편차를 줄이는 것이 바람직해 보인다.

본 연구는 2023 년 서울시 부동산 데이터로 진행되었기 때문에, 조금 더 다양한 연도 혹은 지역 데이터 분석을 통해 더 넓은 범위에서도 분석내용이 일치하는지 확인해본 후, 솔루션 내용을 적용하는 것을 고려해도 좋을 것이다. 이외에도 월세 형태의 부동산 계약도 보증금과 월세 간 관계를 정확히 파악하여 분석해본다면 더 많은 인사이트를 얻을 수 있을 것이다.