

목차

- <Step1.> : 문제정의 및 가설설정
 - [문제정의]
 - [가설설정]
- <Step2.> : 데이터 가공(수집 및 전처리)
 - [데이터 불러오기]
 - [데이터 결합]
 - [데이터 결측값 제거]
 - [데이터 구조 변경]
- <Step3.> : 데이터 모델링 및 시각화
 - [1. 아파트 매매가에 영향을 미치는 요인]
 - [1-1. 건물면적이 넓을수록 아파트 매매가도 높을 것이다 _ ([Implot](#))]
 - [1-2. 건축년도별로 아파트 매매가에 차이가 날 것이다 _ ([heatmap](#))]
 - [1-3. 지역구별로 아파트 매매가에 차이가 날 것이다 _ ([boxplot](#) & [boxenplot](#))]
 - [2. 연도별 아파트 건물면적과 층고]
 - [2-1. 건축년도가 최근일수록 건물면적이 좁을 것이다 _ ([pieplot](#)) & [plot](#)]
 - [2-2. 건축년도가 최근일수록 층고가 높을 것이다 _ ([plot](#))]

<Step1.> : 문제정의 및 가설설정

[문제정의]

1. 아파트 매매가에 영향을 미치는 요인
2. 연도별 아파트 건물면적과 층고

[가설설정]

1. 아파트 매매가에 영향을 미치는 요인
 - 1-1. 건물면적이 넓을수록 아파트 매매가도 높을 것이다.
 - 1-2. 건축년도별로 아파트 매매가에 차이가 날 것이다.
 - 1-3. 지역구별로 아파트 매매가에 차이가 날 것이다.
2. 연도별 아파트 건물면적과 층고
 - 2-1. 건축년도가 최근일수록 건물면적이 좁을 것이다.
 - 2-2. 건축년도가 최근일수록 층고가 높을 것이다.

<Step2.> : 데이터 가공

In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:

```
import platform
```

```

from matplotlib import font_manager, rc
plt.rcParams['axes.unicode_minus'] = False

if platform.system() == 'Darwin':
    rc('font', family='AppleGothic')
elif platform.system() == 'Windows':
    path = "c:/Windows/Fonts/malgun.ttf"
    font_name = font_manager.FontProperties(fname=path).get_name()
    rc('font', family=font_name)
else:
    print('Unknown system... sorry~~~')

```

1. 법정동코드 가공

In [3]:

```
locCode = pd.read_excel('../data/법정동코드.xlsx')
locCode.head()
```

Out[3]:

	법정동코드	시도명	시군구명	읍면동명	동리명	생성일자	말소일자
0	1100000000	서울특별시	NaN	NaN	NaN	19880423	NaN
1	1111000000	서울특별시	종로구	NaN	NaN	19880423	NaN
2	1111010100	서울특별시	종로구	청운동	NaN	19880423	NaN
3	1111010200	서울특별시	종로구	신교동	NaN	19880423	NaN
4	1111010300	서울특별시	종로구	궁정동	NaN	19880423	NaN

In [4]:

```
locCode.shape
```

Out[4]: (20556, 7)

In [5]:

```
# 시도명 서울특별시 외에는 삭제
locCode.drop(locCode[locCode['시도명'] != '서울특별시'].index, inplace=True)
locCode.shape
```

Out[5]: (493, 7)

In [6]:

```
# 법정동 코드 컬럼에서 5자리 추출합니다.
# 예) 서울특별시 종로구 > 11000
locCode['지역코드'] = locCode['법정동코드'] // 100000
locCode.head()
```

Out[6]:

	법정동코드	시도명	시군구명	읍면동명	동리명	생성일자	말소일자	지역코드
0	1100000000	서울특별시	NaN	NaN	NaN	19880423	NaN	11000
1	1111000000	서울특별시	종로구	NaN	NaN	19880423	NaN	11110
2	1111010100	서울특별시	종로구	청운동	NaN	19880423	NaN	11110
3	1111010200	서울특별시	종로구	신교동	NaN	19880423	NaN	11110
4	1111010300	서울특별시	종로구	궁정동	NaN	19880423	NaN	11110

In [7]:

```
# 불필요한 컬럼 제거
loc = locCode.drop(['법정동코드', '동리명', '생성일자', '말소일자'], axis=1)
```

```
loc.head()
```

Out[7]: 시도명 시군구명 읍면동명 지역코드

0	서울특별시	NaN	NaN	11000
1	서울특별시	종로구	NaN	11110
2	서울특별시	종로구	청운동	11110
3	서울특별시	종로구	신교동	11110
4	서울특별시	종로구	궁정동	11110

In [8]:

```
# 구와 동에서 NaN 값 제거
loc.dropna(inplace=True)
loc.head()
```

Out[8]: 시도명 시군구명 읍면동명 지역코드

2	서울특별시	종로구	청운동	11110
3	서울특별시	종로구	신교동	11110
4	서울특별시	종로구	궁정동	11110
5	서울특별시	종로구	효자동	11110
6	서울특별시	종로구	창성동	11110

2. 실거래가 데이터 가공

In [9]:

```
# 실거래가 데이터 불러오기
apt = pd.read_csv('.../data/2020년 서울 아파트 실거래가.csv', encoding="cp949")
apt.head()
```

Out[9]:

	법정동명	건물면적	층정보	건물주용도코드	건물주용도	거래금액	건축년도	아파트명
0	장위동	59.92	5.0	2001	아파트	749000000	0.0	래미안 장위포레카운티
1	길음동	84.77	18.0	2001	아파트	1200000000	0.0	롯데캐슬 클라시아
2	구로동	36.90	1.0	2001	아파트	105000000	1994.0	궁전아트빌라
3	산천동	59.55	2.0	2001	아파트	830000000	2001.0	리버힐삼성
4	이촌동	64.43	5.0	2001	아파트	735000000	1971.0	강변

In [10]:

```
# 불러온 데이터에 NA값이 있는지 조회합니다.
apt.isna().sum()
```

Out[10]:

법정동명	0
건물면적	0
층정보	12163
건물주용도코드	0
건물주용도	0
거래금액	0
건축년도	565

```
아파트명      12163
dtype: int64
```

In [11]: `apt.shape`

Out[11]: (176001, 8)

In [12]: `# 건물주용도 컬럼 제거
apt.drop('건물주용도코드', axis=1, inplace=True)
apt.shape`

Out[12]: (176001, 7)

In [13]: `# 건물주용도 아파트만 남기기
apt.drop(apt[apt['건물주용도'] != '아파트'].index, inplace=True)
apt.head()`

Out[13]:

	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
0	장위동	59.92	5.0	아파트	749000000	0.0	래미안 장위포레카운티
1	길음동	84.77	18.0	아파트	1200000000	0.0	롯데캐슬 클라시아
2	구로동	36.90	1.0	아파트	105000000	1994.0	궁전아트빌라
3	산천동	59.55	2.0	아파트	830000000	2001.0	리버힐삼성
4	이촌동	64.43	5.0	아파트	735000000	1971.0	강변

In [14]: `# 거래금액 억 단위로
apt['거래금액'] = apt['거래금액'] / 100000000
apt.head()`

Out[14]:

	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
0	장위동	59.92	5.0	아파트	7.49	0.0	래미안 장위포레카운티
1	길음동	84.77	18.0	아파트	12.00	0.0	롯데캐슬 클라시아
2	구로동	36.90	1.0	아파트	1.05	1994.0	궁전아트빌라
3	산천동	59.55	2.0	아파트	8.30	2001.0	리버힐삼성
4	이촌동	64.43	5.0	아파트	7.35	1971.0	강변

In [15]: `# 건물면적 평 단위로
apt['건물면적'] = round(apt['건물면적'] * 0.3025)
apt.head()`

Out[15]:

	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
0	장위동	18.0	5.0	아파트	7.49	0.0	래미안 장위포레카운티
1	길음동	26.0	18.0	아파트	12.00	0.0	롯데캐슬 클라시아
2	구로동	11.0	1.0	아파트	1.05	1994.0	궁전아트빌라
3	산천동	18.0	2.0	아파트	8.30	2001.0	리버힐삼성
4	이촌동	19.0	5.0	아파트	7.35	1971.0	강변

In [16]:

```
# 변준영 아파트
# 건축년도 0 인 부분 삭제
bjy_apt = apt.drop(apt[apt['건축년도'] == 0].index)
bjy_apt.head()
```

Out[16]:

	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
2	구로동	11.0	1.0	아파트	1.05	1994.0	궁전아트빌라
3	산천동	18.0	2.0	아파트	8.30	2001.0	리버힐삼성
4	이촌동	19.0	5.0	아파트	7.35	1971.0	강변
5	이태원동	32.0	9.0	아파트	14.30	1982.0	청화1
7	후암동	45.0	4.0	아파트	8.75	2002.0	힐튼빌리지2차

In [17]:

```
# 건축년도 0 -> 2000
apt['건축년도'].replace(0, 2000, inplace=True)
apt.head()
```

Out[17]:

	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
0	장위동	18.0	5.0	아파트	7.49	2000.0	래미안 장위포레카운티
1	길음동	26.0	18.0	아파트	12.00	2000.0	롯데캐슬 클라시아
2	구로동	11.0	1.0	아파트	1.05	1994.0	궁전아트빌라
3	산천동	18.0	2.0	아파트	8.30	2001.0	리버힐삼성
4	이촌동	19.0	5.0	아파트	7.35	1971.0	강변

3. 데이터 프레임 결합

In [18]:

```
loc_apt = pd.merge(loc, apt, left_on='읍면동명', right_on='법정동명')
loc_apt
```

Out[18]:

	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
0	서울특별시	종로구	청운동	11110	청운동	26.0	3.0	아파트	8.95	2000.0	청운현대
1	서울특별시	종로구	청운동	11110	청운동	26.0	3.0	아파트	8.50	2000.0	청운현대
2	서울특별시	종로구	신교동	11110	신교동	41.0	5.0	아파트	10.00	2002.0	신현(102동)
3	서울특별시	종로구	필운동	11110	필운동	33.0	7.0	아파트	9.40	2007.0	신동아블루아광화문의 꿈
4	서울특별시	종로구	필운동	11110	필운동	34.0	6.0	아파트	10.00	2007.0	신동아블루아광화문의 꿈
...
90661	서울특별시	강동구	강일동	11740	강일동	26.0	4.0	아파트	8.20	2009.0	강일리버파크6단지

시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	총정보	건물주용도	거래금액	건축년도	아파트명
90662	서울특별시	강동구	강일동	11740	강일동	26.0	10.0	아파트	7.90	2009.0
90663	서울특별시	강동구	강일동	11740	강일동	18.0	12.0	아파트	8.30	2009.0
90664	서울특별시	강동구	강일동	11740	강일동	18.0	13.0	아파트	7.60	2009.0
90665	서울특별시	강동구	강일동	11740	강일동	35.0	8.0	아파트	12.05	2011.0

90666 rows × 11 columns

<Step3.> : 데이터 모델링 및 시각화

[1. 아파트 매매가에 영향을 미치는 요인]

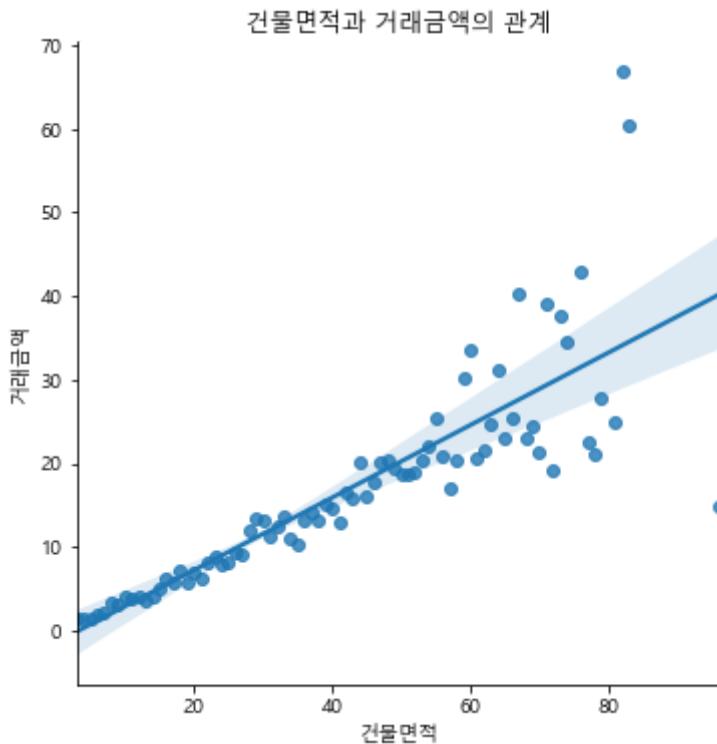
1-1. 건물면적이 넓을수록 아파트 매매가도 높을 것이다. _ (Implot)

```
In [19]: lsst_apt = apt.pivot_table('거래금액', '건물면적').reset_index()
lsst_apt
```

```
Out[19]:   건물면적  거래금액
0           3.0    1.399400
1           4.0    1.409988
2           5.0    1.560451
3           6.0    2.037068
4           7.0    2.281165
...
75          79.0   27.865000
76          81.0   25.000000
77          82.0   67.000000
78          83.0   60.375000
79          96.0   15.000000
```

80 rows × 2 columns

```
In [20]: sns.lmplot(x='건물면적', y='거래금액', data=lsst_apt)
plt.title("건물면적과 거래금액의 관계")
plt.show()
```



분석결과

건물면적이 넓으면 아파트 매매가(거래금액)가 높을 것이라는 가설에 대해 확인했다.

> 예측한대로 건물면적이 넓으면 아파트 매매가도 높다.

1-2. 건축년도별로 아파트 매매가에 차이가 날 것이다 _ (heatmap)

```
In [21]: apt_old = loc_apt.replace(0, np.Nan)
apt_old=loc_apt.groupby(['건축년도','시군구명'])[['거래금액']].mean()
apt_old.reset_index(inplace=True)
apt_old
```

Out[21]:

	건축년도	시군구명	거래금액
0	1961.0	구로구	1.715000
1	1965.0	중구	5.500000
2	1966.0	종로구	3.772000
3	1967.0	구로구	1.422500
4	1968.0	중구	2.321667
...
947	2020.0	영등포구	8.630852
948	2020.0	은평구	5.642625
949	2020.0	종로구	23.000000
950	2020.0	중구	2.400000
951	2020.0	중랑구	3.377440

952 rows × 3 columns

최신 건물의 가격평균이 높은가격을 형성 할 것이라고 가정

In [22]:

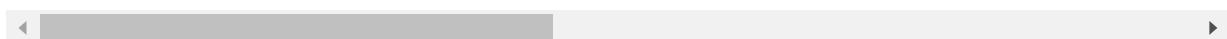
```
old=pd.pivot_table(apt_old,index="건축년도",columns='시군구명', values="거래금액")  
old
```

Out[22]:

시군구명	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원
건축년도									
1961.0	NaN	NaN	NaN	NaN	NaN	NaN	1.715000	NaN	N
1965.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1966.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1967.0	NaN	NaN	NaN	NaN	NaN	NaN	1.422500	NaN	N
1968.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1969.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1970.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1971.0	NaN	NaN	NaN	NaN	2.187917	NaN	1.925000	NaN	N
1972.0	NaN	NaN	NaN	NaN	NaN	NaN	1.848000	NaN	N
1973.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1974.0	9.140000	NaN	NaN	1.700000	2.750000	NaN	NaN	NaN	N
1975.0	NaN	NaN	1.000000	NaN	NaN	NaN	1.850000	1.191786	N
1976.0	31.096159	NaN	NaN	NaN	NaN	3.366250	2.584769	NaN	4.4375
1977.0	22.772727	NaN	NaN	NaN	NaN	NaN	2.706429	NaN	N
1978.0	30.022877	NaN	NaN	NaN	3.940000	20.895208	NaN	NaN	N
1979.0	24.284196	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1980.0	25.078846	17.200000	NaN	NaN	NaN	NaN	NaN	5.710000	N
1981.0	26.107143	NaN	NaN	NaN	NaN	NaN	NaN	5.410102	N
1982.0	27.988704	8.312018	NaN	5.568750	7.205714	NaN	1.527143	4.895000	N
1983.0	23.339235	10.130603	NaN	NaN	3.352500	15.476429	NaN	NaN	6.3279
1984.0	26.842593	10.097113	NaN	NaN	4.739851	NaN	4.173983	3.616667	4.6452
1985.0	22.846533	9.387143	NaN	NaN	3.524750	16.512500	6.296071	NaN	6.0714
1986.0	19.611538	13.242048	NaN	NaN	NaN	NaN	6.868167	5.290357	6.2792
1987.0	16.841217	5.593276	NaN	4.508462	NaN	9.808000	4.999078	NaN	5.3973
1988.0	9.578846	7.239151	NaN	3.341944	4.457568	9.172174	3.518538	NaN	4.8230
1989.0	24.650000	6.618421	NaN	5.943000	NaN	11.761375	5.194831	3.461489	4.3582
1990.0	17.000000	5.927059	3.625909	7.330000	NaN	11.555500	NaN	3.223110	4.1872
1991.0	17.997350	4.397200	4.280420	5.673810	4.343333	8.878824	NaN	5.630577	4.6511
1992.0	10.433730	7.313889	4.631000	5.417192	7.036453	11.218571	5.601831	NaN	4.3489
1993.0	10.561736	6.513750	5.186422	5.861388	5.614155	10.221935	4.961719	2.890000	7.0113

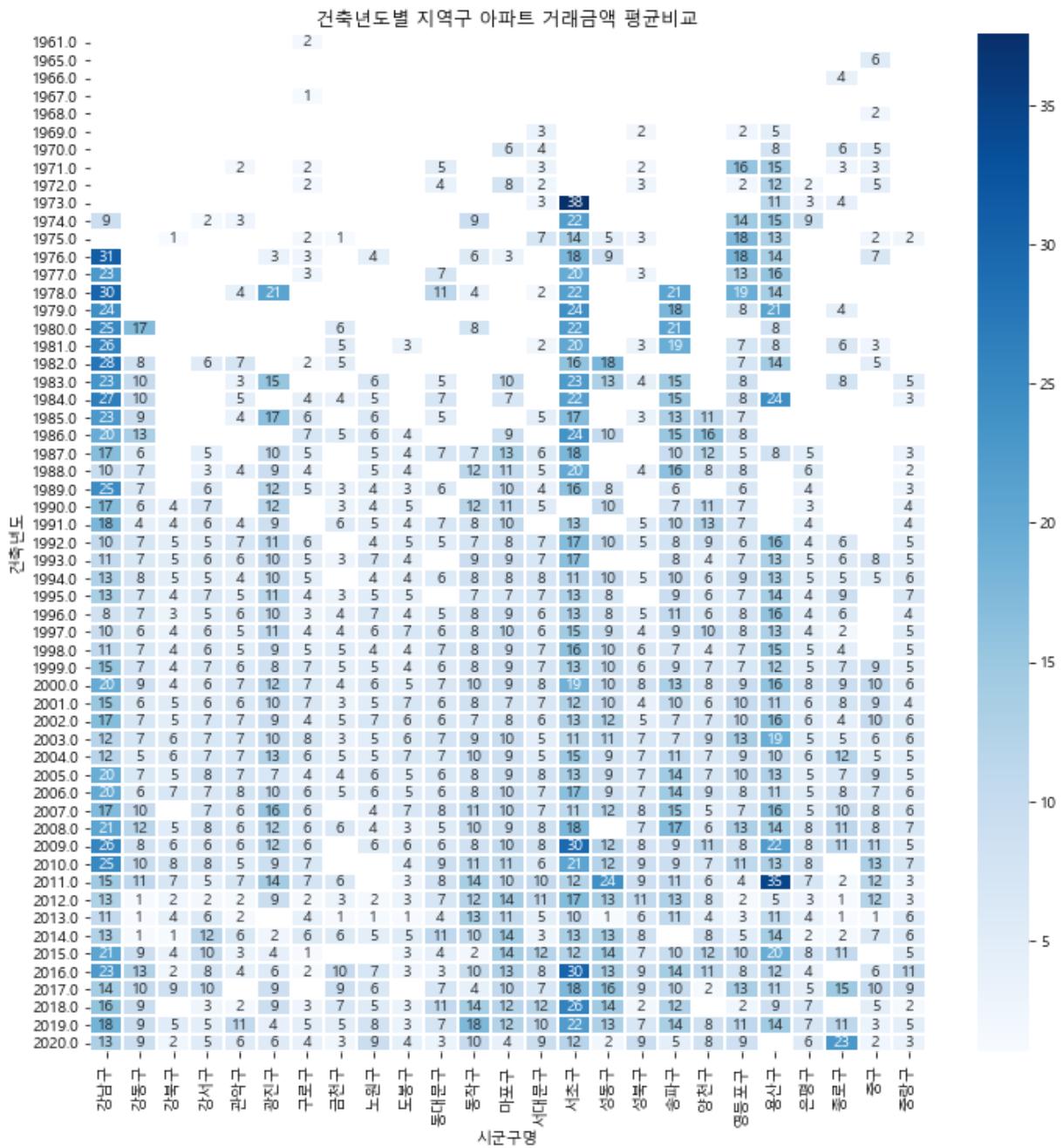
시군구명	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원
건축년도									
1994.0	12.765698	7.533525	4.574375	5.211692	4.273333	9.953333	4.858452	NaN	3.5553
1995.0	13.478667	7.363732	4.268333	6.652527	4.658710	11.467432	4.027831	3.443600	5.1731
1996.0	8.322371	7.213795	3.247000	5.325333	6.115245	10.081547	3.067580	4.106600	6.5214
1997.0	9.571860	5.971402	3.566500	5.652669	4.664750	10.940266	3.609375	3.635000	5.6618
1998.0	10.892617	6.528464	4.422866	6.167289	4.765278	9.256500	4.515938	5.124409	4.3025
1999.0	15.454779	7.459142	3.847973	6.564041	5.796393	8.398611	6.532567	4.540667	5.2294
2000.0	19.762486	8.947747	4.123245	5.516528	7.283938	11.868576	6.542115	4.261492	6.2219
2001.0	15.161875	5.619653	5.426500	5.591667	6.294800	9.581132	6.696912	3.150500	5.2763
2002.0	17.035828	7.067627	4.626774	6.739854	6.686667	9.282909	4.321357	4.925603	6.8005
2003.0	11.928269	7.017598	5.909163	6.753209	7.017108	9.589000	7.699595	3.346389	5.0931
2004.0	11.911965	5.498946	5.643961	6.824278	7.098062	12.565556	6.495537	4.791452	4.8785
2005.0	19.565679	6.537338	4.623108	8.280079	7.263333	6.771852	4.474053	4.115556	6.3866
2006.0	20.109569	5.911063	6.777590	6.653068	7.649184	9.841983	6.251908	4.887708	6.1172
2007.0	17.366849	10.057638	NaN	7.381872	6.292593	15.677451	6.239775	NaN	3.7916
2008.0	20.861402	11.659022	5.189583	8.216151	6.014801	11.600416	5.857568	6.348333	3.9075
2009.0	25.831500	8.164051	5.782963	6.237700	6.408125	11.613409	6.459458	NaN	5.5578
2010.0	25.047917	9.621724	7.608782	8.380400	5.175000	8.805385	6.516771	NaN	N
2011.0	14.635833	10.971832	6.996990	4.505333	7.350000	14.190000	7.449529	6.324000	N
2012.0	12.585000	1.224189	2.353333	1.761739	1.671104	9.393333	1.504770	3.118824	1.6912
2013.0	10.667672	1.163969	4.233333	6.067412	1.550833	NaN	3.974265	1.432362	1.3685
2014.0	12.588034	1.285741	1.397619	11.929787	6.469254	2.295417	5.891512	5.892775	4.7361
2015.0	20.760833	9.015238	3.581111	10.124016	3.026361	4.060000	1.176714	NaN	N
2016.0	22.676829	12.548987	2.292500	7.960208	3.926750	5.772500	1.527083	10.133333	7.0582
2017.0	13.936900	10.302744	8.581250	9.640753	NaN	8.862727	NaN	8.730625	6.2888
2018.0	15.832982	9.327500	NaN	2.731000	1.855714	9.052321	2.849167	6.855000	5.3924
2019.0	18.437492	9.319800	5.157246	4.664194	10.540000	4.083333	4.623386	4.944286	7.7560
2020.0	13.011640	9.120523	2.467451	5.287800	5.571875	5.692000	3.716200	2.545379	9.3553

57 rows × 25 columns



In [23]:

```
import seaborn as sns
plt.figure(figsize=(12, 12))
sns.heatmap(old, annot=True, fmt=".0f", linewidths=1, cmap='Blues')
plt.title('건축년도별 지역구 아파트 거래금액 평균비교')
plt.show()
```



분석 1: 최신 건물보다 오래된 건물의 가격대가 높은것을 확인

- 강남구, 서초구의 경우 오래된 건물의 가격대가 재건축으로 인한 가격 형성으로 예상
 - 용산구의 경우 특정 건축년도 건물의 가격대가 높은 걸 확인하여 추가적인 분석 필요

In [24]:

```
loc_apt0 = loc_apt[loc_apt['시군구명']=='서초구']
loc_apt0_1 = loc_apt0.groupby(['건축년도','법정동명'])[['거래금액']].mean()
loc_apt0_1
```

Out[24]:

거래금액

건축년도 법정동명

1973.0	반포동	37.586389
1974.0	반포동	21.952895
1975.0	방배동	13.775200
1976.0	방배동	17.960625
1977.0	방배동	19.800000

거래금액

건축년도 법정동명

	
2019.0	잠원동	30.111111		
2020.0	방배동	12.850000		
	서초동	5.434417		
	양재동	8.433333		
	잠원동	26.383333		

142 rows × 1 columns

In [25]:

```
loc_apt0_2=pd.pivot_table(loc_apt0_1,index="건축년도",columns='법정동명', values="거래금액")
loc_apt0_2
```

Out[25]:

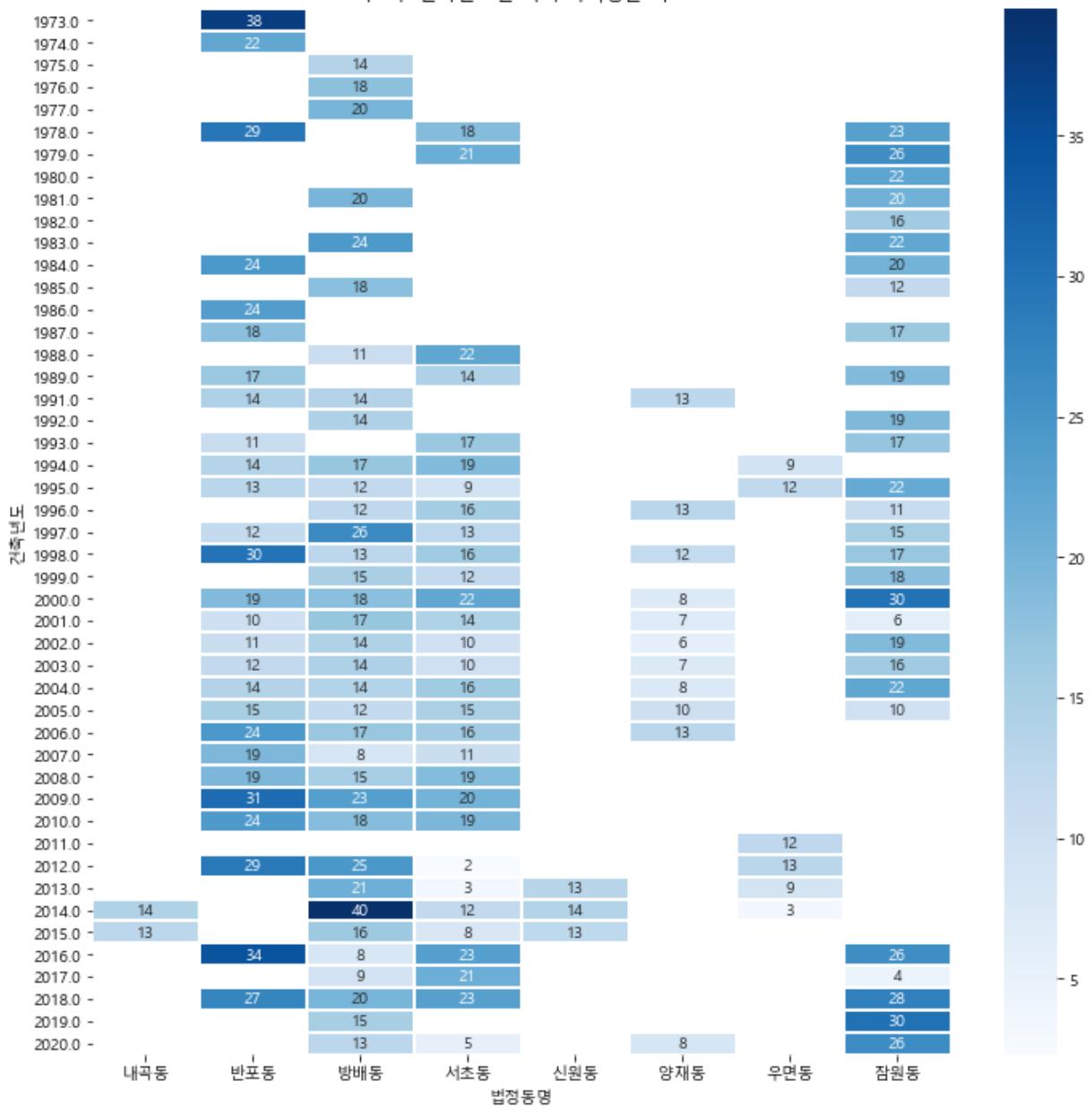
법정동명	내곡동	반포동	방배동	서초동	신원동	양재동	우면동	잠원동
건축년도								
1973.0	NaN	37.586389	NaN	NaN	NaN	NaN	NaN	NaN
1974.0	NaN	21.952895	NaN	NaN	NaN	NaN	NaN	NaN
1975.0	NaN	NaN	13.775200	NaN	NaN	NaN	NaN	NaN
1976.0	NaN	NaN	17.960625	NaN	NaN	NaN	NaN	NaN
1977.0	NaN	NaN	19.800000	NaN	NaN	NaN	NaN	NaN
1978.0	NaN	29.348500	NaN	18.475714	NaN	NaN	NaN	23.159848
1979.0	NaN	NaN	NaN	21.020000	NaN	NaN	NaN	25.988889
1980.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	22.424286
1981.0	NaN	NaN	19.718182	NaN	NaN	NaN	NaN	20.396250
1982.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	15.929500
1983.0	NaN	NaN	24.300000	NaN	NaN	NaN	NaN	22.090625
1984.0	NaN	24.500000	NaN	NaN	NaN	NaN	NaN	20.095000
1985.0	NaN	NaN	18.004000	NaN	NaN	NaN	NaN	12.000000
1986.0	NaN	23.696429	NaN	NaN	NaN	NaN	NaN	NaN
1987.0	NaN	18.094643	NaN	NaN	NaN	NaN	NaN	16.541667
1988.0	NaN	NaN	10.860588	22.316176	NaN	NaN	NaN	NaN
1989.0	NaN	16.665000	NaN	14.301875	NaN	NaN	NaN	18.983333
1991.0	NaN	14.250000	13.777500	NaN	NaN	12.757500	NaN	NaN
1992.0	NaN	NaN	14.253684	NaN	NaN	NaN	NaN	19.033793
1993.0	NaN	11.000000	NaN	16.555000	NaN	NaN	NaN	17.020833
1994.0	NaN	13.632222	16.865000	18.833333	NaN	NaN	9.380370	NaN

법정동명	내곡동	반포동	방배동	서초동	신원동	양재동	우면동	잠원동
건축년도								
1995.0	NaN	12.991667	12.000000	9.337500	NaN	NaN	12.241212	21.562500
1996.0	NaN	NaN	12.391667	15.520000	NaN	13.162500	NaN	11.270000
1997.0	NaN	11.800000	26.500000	12.695000	NaN	NaN	NaN	15.454000
1998.0	NaN	29.800000	12.858182	15.939500	NaN	11.800000	NaN	16.710870
1999.0	NaN	NaN	14.838077	11.938437	NaN	NaN	NaN	18.233333
2000.0	NaN	18.910000	18.138889	22.036667	NaN	7.519483	NaN	30.000000
2001.0	NaN	10.032308	16.851410	14.122059	NaN	7.197222	NaN	6.060000
2002.0	NaN	10.799375	14.163000	9.874262	NaN	5.862614	NaN	18.989079
2003.0	NaN	11.910000	14.253060	10.071521	NaN	7.383571	NaN	16.083333
2004.0	NaN	13.683333	13.824187	15.998611	NaN	7.631818	NaN	22.286667
2005.0	NaN	15.141667	11.850745	14.664390	NaN	10.316667	NaN	9.930000
2006.0	NaN	24.389286	16.786364	16.101840	NaN	12.912500	NaN	NaN
2007.0	NaN	19.175000	8.420833	10.800000	NaN	NaN	NaN	NaN
2008.0	NaN	19.366667	15.333333	18.700000	NaN	NaN	NaN	NaN
2009.0	NaN	30.868295	23.371429	20.045385	NaN	NaN	NaN	NaN
2010.0	NaN	24.184706	18.380625	19.310417	NaN	NaN	NaN	NaN
2011.0	NaN	NaN	NaN	NaN	NaN	NaN	12.350000	NaN
2012.0	NaN	29.061111	24.650000	2.300000	NaN	NaN	12.910000	NaN
2013.0	NaN	NaN	20.706818	3.174722	13.333333	NaN	9.071994	NaN
2014.0	14.188254	NaN	39.500000	12.139474	13.901034	NaN	3.307500	NaN
2015.0	12.762500	NaN	16.300000	7.983333	12.522143	NaN	NaN	NaN
2016.0	NaN	34.312698	8.212500	23.427273	NaN	NaN	NaN	26.108696
2017.0	NaN	NaN	9.130000	20.800000	NaN	NaN	NaN	4.500000
2018.0	NaN	27.310182	19.819231	23.127500	NaN	NaN	NaN	28.098333
2019.0	NaN	NaN	15.158990	NaN	NaN	NaN	NaN	30.111111
2020.0	NaN	NaN	12.850000	5.434417	NaN	8.433333	NaN	26.383333

In [26]:

```
plt.figure(figsize=(12, 12))
sns.heatmap(loc_apt0_2, annot=True, fmt=".0f", linewidths=1, cmap='Blues')
plt.title('서초구 건축년도별 지역 가격평균 비교')
plt.show()
```

서초구 건축년도별 지역 가격평균 비교



분석 2 : 오래된 건물의 높은 가격형성의 이유를 재건축으로 예상, 표본으로 서초구 선별하여 동별 분석

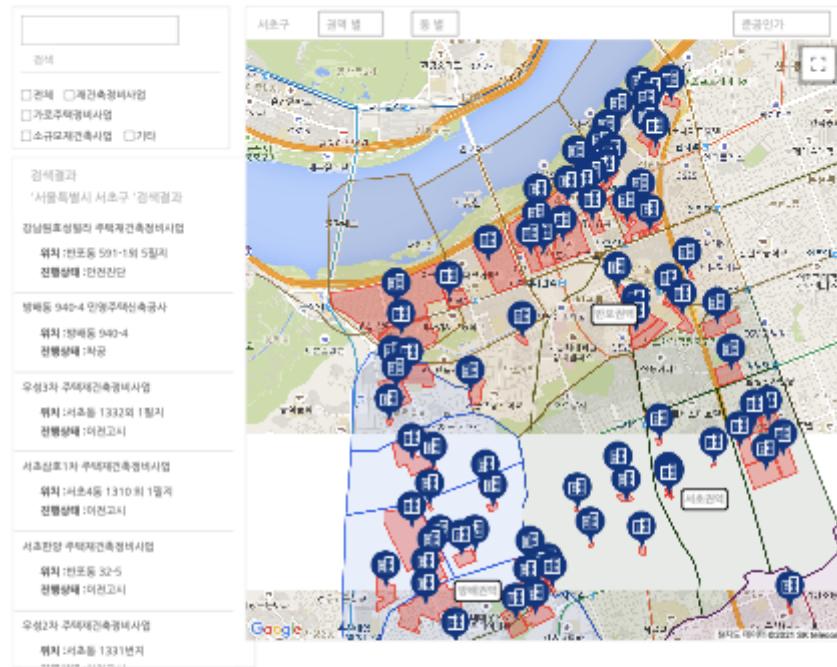
- 분석과 같이 반포동, 잠원동, 방배동의 경우 재건축으로 인한 높은 가격 형성을 확인 할 수 있었다. (사진첨부)

(사진1:서초구 공동주택&재건축 정보포털)

2021. 5. 21.

서초구 공동주택 & 재건축 정보포털

사업 현황



사이트맵



(06750) 서울특별시 서초구 낭류순환로 258(서초동) Tel. 02-2155-6114 (120) 다산증선아트 펜션 02-2155-6100-3) 이안·금종월(당직실)
COPYRIGHTS 2015 SEOCHO SEOUL KOREA, ALL RIGHT RESERVED.

https://www.seocho.go.kr/renewal/main_main.jsp

1/1

In [27]:

```
loc_apt1 = loc_apt[loc_apt['시군구명']=='용산구']
loc_apt1_1 = loc_apt1.groupby(['건축년도','법정동명'])[['거래금액']].mean()
loc_apt1_1
```

Out[27]:

거래금액

건축년도	법정동명	거래금액
1969.0	동자동	4.898333
	이태원동	4.420000
	효창동	4.550000
1970.0	원효로2가	4.210000
	이촌동	8.800000

...

거래금액

건축년도	법정동명	거래금액
2017.0	한남동	8.238652
2018.0	문배동	4.592500
	원효로1가	4.400000
	효창동	14.260000
2019.0	효창동	14.456000

127 rows × 1 columns

```
In [28]: loc_apt1_2=pd.pivot_table(loc_apt1_1,index="건축년도",columns='법정동명', values="거래금액")
loc_apt1_2
```

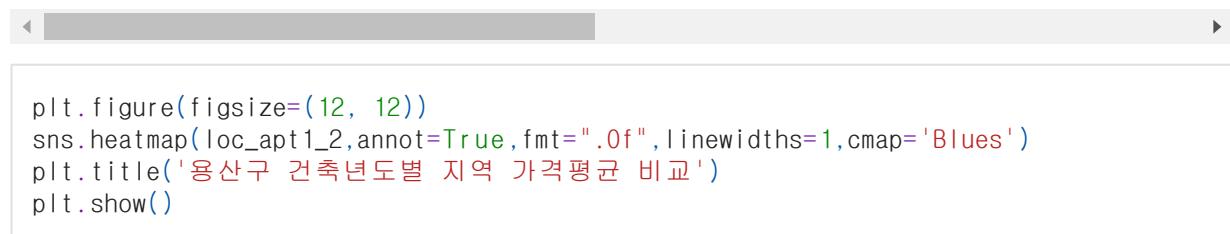
```
Out[28]:
```

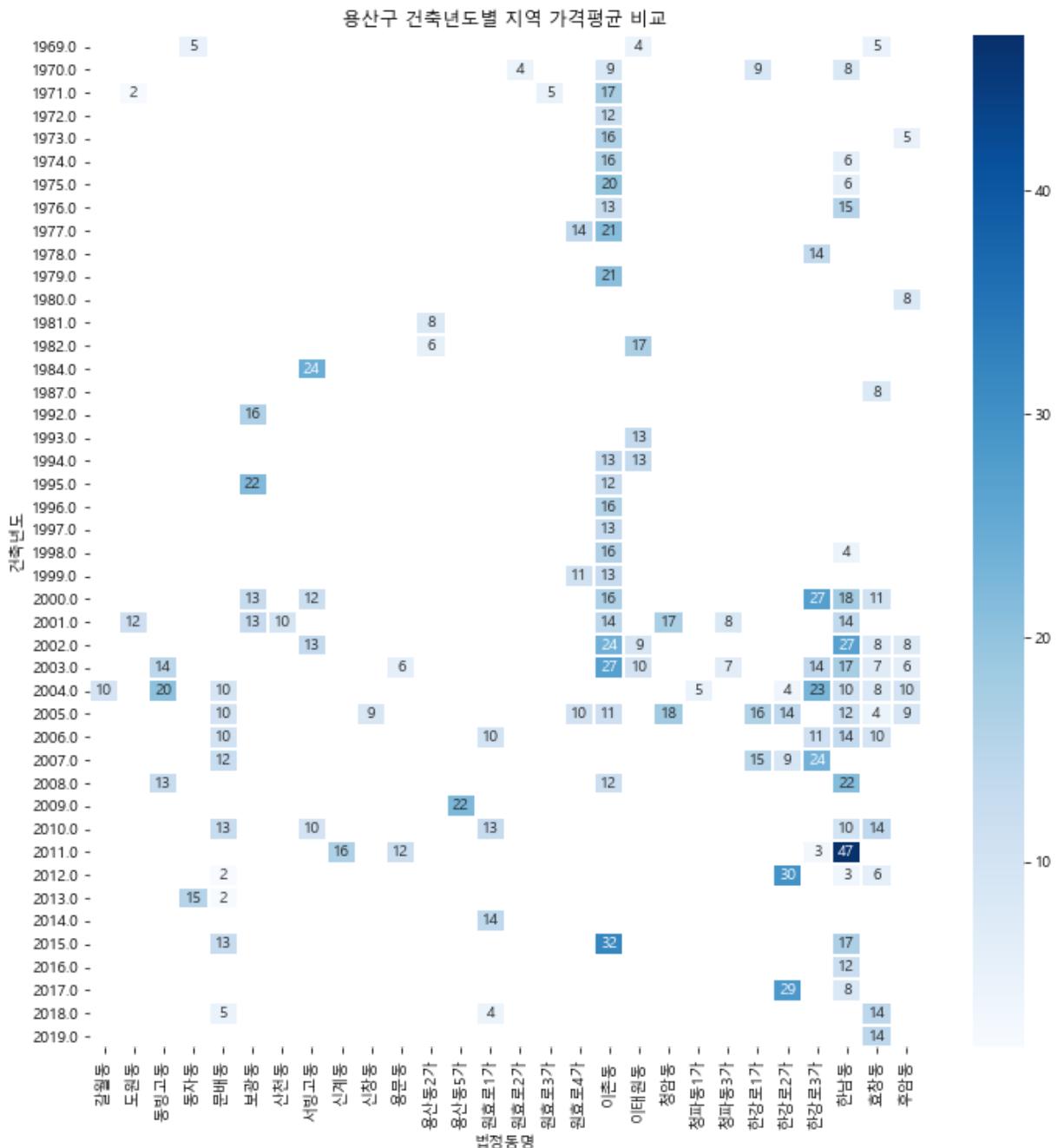
법정동명	갈월동	도원동	동빙고동	동자동	문배동	보광동	산천동	서빙고동	신계동	신창동
------	-----	-----	------	-----	-----	-----	-----	------	-----	-----

건축년도	갈월동	도원동	동빙고동	동자동	문배동	보광동	산천동	서빙고동	신계동	신창동
1969.0	NaN	NaN	NaN	4.898333	NaN	NaN	NaN	NaN	NaN	NaN
1970.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1971.0	NaN	1.870000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1972.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1973.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1974.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1975.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1976.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1977.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1978.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1979.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1980.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1981.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1982.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1984.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	24.388393	NaN	NaN
1987.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1992.0	NaN	NaN	NaN	NaN	NaN	16.403125	NaN	NaN	NaN	NaN
1993.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1994.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1995.0	NaN	NaN	NaN	NaN	NaN	22.000000	NaN	NaN	NaN	NaN
1996.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1997.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

법정동명	갈월동	도원동	동빙고동	동자동	문배동	보광동	산천동	서빙고동	신계동	신창동
건축년도										
1998.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1999.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2000.0	NaN	NaN	NaN	NaN	NaN	12.887500	NaN	11.656250	NaN	NaN
2001.0	NaN	11.503831	NaN	NaN	NaN	13.150000	9.575417	NaN	NaN	NaN
2002.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	13.350000	NaN	NaN
2003.0	NaN	NaN	14.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2004.0	9.75	NaN	20.5	NaN	9.590909	NaN	NaN	NaN	NaN	NaN
2005.0	NaN	NaN	NaN	NaN	10.118125	NaN	NaN	NaN	NaN	8.618750
2006.0	NaN	NaN	NaN	NaN	10.420625	NaN	NaN	NaN	NaN	NaN
2007.0	NaN	NaN	NaN	NaN	12.395000	NaN	NaN	NaN	NaN	NaN
2008.0	NaN	NaN	12.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2009.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2010.0	NaN	NaN	NaN	NaN	12.615455	NaN	NaN	10.100000	NaN	NaN
2011.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	16.4675	NaN
2012.0	NaN	NaN	NaN	NaN	1.686000	NaN	NaN	NaN	NaN	NaN
2013.0	NaN	NaN	NaN	15.393333	1.946429	NaN	NaN	NaN	NaN	NaN
2014.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2015.0	NaN	NaN	NaN	NaN	12.567857	NaN	NaN	NaN	NaN	NaN
2016.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2017.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2018.0	NaN	NaN	NaN	NaN	4.592500	NaN	NaN	NaN	NaN	NaN
2019.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

44 rows × 28 columns





분석 3: 용산구의 경우 특정 건축년도 건물의 가격대가 높은걸 확인하여 추가적인 분석 실시

- 해당 특정 건축년도의 경우 한남동의 매물로 확인 자세한 분석이 필요함.

In [30]:

```
loc_apt2 = loc_apt[loc_apt['법정동명']=='한남동']  
loc_apt2
```

Out [30]:

시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명	
3172	서울특별시	용산구	한남동	11170	한남동	26.0	1.0	아파트	11.00	2000.0	효성빌리지
3173	서울특별시	용산구	한남동	11170	한남동	69.0	5.0	아파트	24.40	2000.0	형우베스트빌3차
3174	서울특별시	용산구	한남동	11170	한남동	18.0	8.0	아파트	10.40	2004.0	한남동리첸시아

	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
3175	서울특별시	용산구	한남동	11170	한남동	18.0	1.0	아파트	9.50	1970.0	시범
3176	서울특별시	용산구	한남동	11170	한남동	18.0	7.0	아파트	9.00	2001.0	리버탑
...
3358	서울특별시	용산구	한남동	11170	한남동	63.0	5.0	아파트	28.50	2003.0	힐탑트레저1동
3359	서울특별시	용산구	한남동	11170	한남동	53.0	15.0	아파트	16.50	2001.0	리버티하우스
3360	서울특별시	용산구	한남동	11170	한남동	14.0	5.0	아파트	8.80	2017.0	한남아이파크애비뉴
3361	서울특별시	용산구	한남동	11170	한남동	23.0	3.0	아파트	16.20	1976.0	장미1
3362	서울특별시	용산구	한남동	11170	한남동	39.0	12.0	아파트	18.45	2003.0	한남힐스테이트

191 rows × 11 columns

In [31]:

```
loc_apt2_1 = loc_apt2 .groupby(['건물면적','아파트명'])[['거래금액']].mean()
loc_apt2_1
```

Out[31]:

거래금액

건물면적	아파트명	거래금액
5.0	월쏘나슈트	2.819383
6.0	월쏘나슈트	3.282050
7.0	월쏘나슈트	3.928250
8.0	월쏘나슈트	4.476800
9.0	한성1	3.885000
	현대안성타워	2.750000
10.0	월쏘나슈트	5.406750
	한남동리첸시아	5.483333
	한성1	3.900000
11.0	한남동리첸시아	5.790000
12.0	현대안성타워	3.700000
14.0	성아1	4.150000
	한남동리첸시아	8.475000
	한남아이파크애비뉴	7.845000
15.0	한남아이파크애비뉴	8.491714
	현대안성타워	5.300000
16.0	장미1	12.990000

거래금액

건물면적	아파트명	거래금액
	한남동리첸시아	7.850000
18.0	리버탑	9.500000
	시범	8.983333
	태성1	6.233333
	한남더힐	21.759091
	한남동동원베네스트	8.500000
	한남동리첸시아	9.966667
	한성1	7.000000
19.0	아일랜드캐슬	6.166667
20.0	한남해피트리	11.887500
21.0	시범	9.300000
22.0	한남동리첸시아	10.705000
23.0	장미1	16.600000
25.0	오페라하우스3차	8.900000
	한남동동원베네스트	12.000000
26.0	대림아르빌	11.700000
	대성 이태리하우스	13.500000
	리버탑	11.725000
	한남힐스테이트	14.386667
	효성빌리지	12.950000
28.0	오페라하우스WEST	18.400000
35.0	한성1	10.800000
38.0	한남힐스테이트	17.000000
39.0	한남힐스테이트	18.987500
42.0	신성미소시티	13.500000
46.0	한남힐스테이트	21.000000
52.0	트원빌 에이동	21.500000
53.0	리버티하우스	18.000000
54.0	한남더힐	41.700000
57.0	한남동UNVILLAGE루하우스	16.700000
60.0	현대하이페리온	26.633333
63.0	한남더힐	44.058333
	힐탑트레저1동	28.500000
68.0	현대하이페리온	31.300000
69.0	현대하이페리온II	34.300000

거래금액

건물면적		아파트명													
		형우베스트빌3차		22.700000											
70.0		힐탑트레져1동		32.100000											
71.0		한남더힐		51.221429											
73.0		한남더힐		68.380000											
74.0		한남더힐		77.500000											

In [32]:

```
loc_apt2_2=pd.pivot_table(loc_apt2_1, index="건물면적", columns='아파트명', values="거래금액")
loc_apt2_2
```

Out[32]:

아파트명	대림아르빌	대성이태리하우스	리버타입	리버티하우스	성아1	시범	신성미소시티	아일랜드캐슬	오페라하우스3차	오페라하우스WEST	한남아이파크애비뉴	한남해피트리
건물면적												
5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14.0	NaN	NaN	NaN	NaN	4.15	NaN	NaN	NaN	NaN	NaN	7.845000	NaN
15.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8.491714	NaN
16.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18.0	NaN	NaN	9.500	NaN	NaN	8.983333	NaN	NaN	NaN	NaN	NaN	NaN
19.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6.166667	NaN	NaN	NaN	NaN
20.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	11.8875
21.0	NaN	NaN	NaN	NaN	NaN	9.300000	NaN	NaN	NaN	NaN	NaN	NaN
22.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8.9	NaN	NaN	NaN
26.0	11.7	13.5	11.725	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	18.4	NaN	NaN
35.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
38.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

아파트명	대림아르빌	대성이태리하우스	리버탑	리버티하우스	성아1	시범	신성미소시티	아일랜드캐슬	오페라하우스3차	오페라하우스WEST	한남아이파크애비뉴	한남해피트리	
건물면적													
39.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
42.0	NaN	NaN	NaN	NaN	NaN	NaN	13.5	NaN	NaN	NaN	...	NaN	NaN
46.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
52.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
53.0	NaN	NaN	NaN	18.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
54.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
57.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
60.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
63.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
68.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
69.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
70.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
71.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
73.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
74.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	

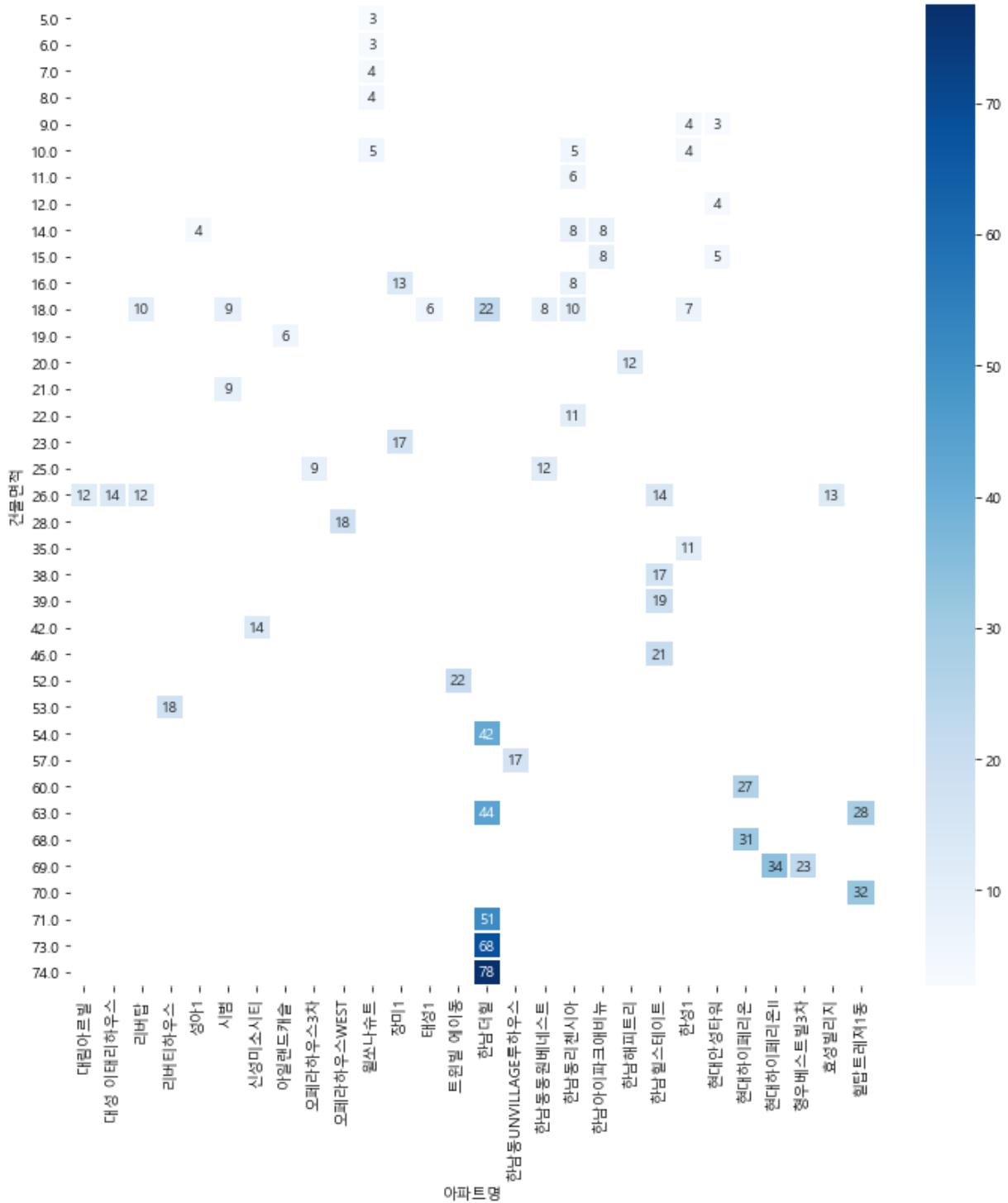
37 rows × 28 columns



In [33]:

```
plt.figure(figsize=(12, 12))
sns.heatmap(loc_apt2_2, annot=True, fmt=".0f", linewidths=1, cmap='Blues')
plt.title('한남동 건물면적별 아파트 가격평균 비교')
plt.show()
```

한남동 건물면적별 아파트 가격평균 비교



분석 4 : 용산구, 한남동의 특정 매물 분석

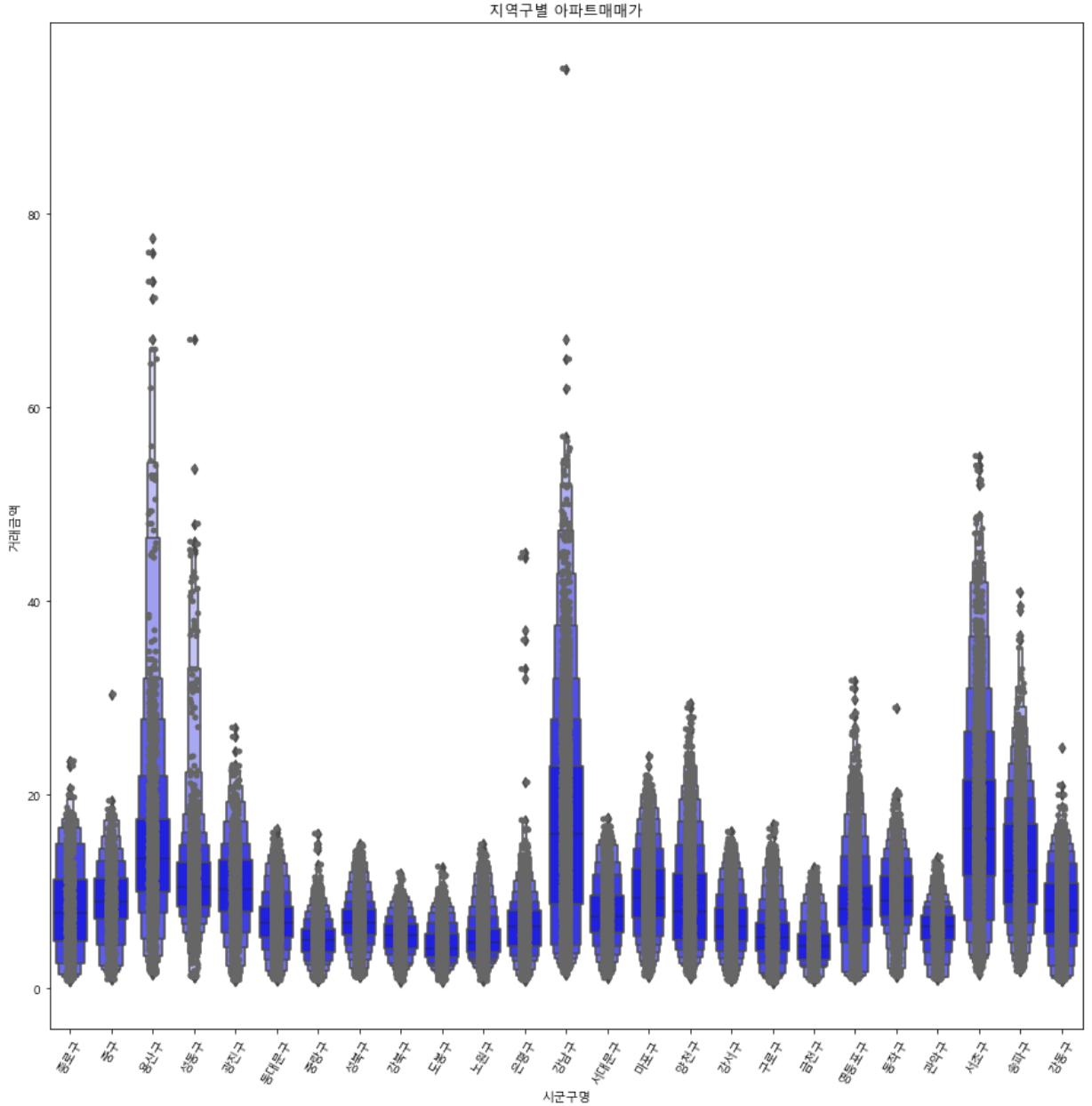
- 해당 매물의 경우 고가의 한남더 힐 가격으로 확인

1-3. 지역구별로 아파트 매매가에 차이가 날 것이다 _ (boxplot & boxenplot)

우선 단순히 지역구별 아파트매매가에 대해 확인했다.

```
In [34]: plt.figure(figsize=(15, 15))
plt.xticks(rotation=60)
plt.title('지역구별 아파트매매가')
sns.boxenplot(x="시군구명", y="거래금액",
               color="b",
               scale="linear",
```

```
data=loc_apt)
sns.stripplot(x='시군구명', y='거래금액', data=loc_apt, jitter=True, color="0.4")
plt.show()
```



이상치에 대해 확인하기 위해 50억이 넘어가는 레코드에 대해 확인했다.

```
In [35]: loc_apt[loc_apt['거래금액'] > 50]
```

Out [35]:

	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	총정보	건물주용도	거래금액	건축년도	아파트명
3179	서울특별시	용산구	한남동	11170	한남동	71.0	6.0	아파트	56.00	2011.0	한남더힐
3185	서울특별시	용산구	한남동	11170	한남동	73.0	1.0	아파트	71.30	2011.0	한남더힐
3186	서울특별시	용산구	한남동	11170	한남동	73.0	3.0	아파트	62.00	2011.0	한남더힐

	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
3220	서울특별시	용산구	한남동	11170	한남동	71.0	5.0	아파트	52.50	2011.0	한남더힐
3226	서울특별시	용산구	한남동	11170	한남동	73.0	5.0	아파트	66.00	2011.0	한남더힐
3227	서울특별시	용산구	한남동	11170	한남동	73.0	4.0	아파트	64.50	2011.0	한남더힐
3231	서울특별시	용산구	한남동	11170	한남동	73.0	-1.0	아파트	73.00	2011.0	한남더힐
3238	서울특별시	용산구	한남동	11170	한남동	71.0	8.0	아파트	53.00	2011.0	한남더힐
3252	서울특별시	용산구	한남동	11170	한남동	73.0	2.0	아파트	67.00	2011.0	한남더힐
3258	서울특별시	용산구	한남동	11170	한남동	71.0	6.0	아파트	52.70	2011.0	한남더힐
3262	서울특별시	용산구	한남동	11170	한남동	71.0	11.0	아파트	54.00	2011.0	한남더힐
3269	서울특별시	용산구	한남동	11170	한남동	73.0	3.0	아파트	66.00	2011.0	한남더힐
3275	서울특별시	용산구	한남동	11170	한남동	71.0	8.0	아파트	53.00	2011.0	한남더힐
3282	서울특별시	용산구	한남동	11170	한남동	71.0	3.0	아파트	54.50	2011.0	한남더힐
3286	서울특별시	용산구	한남동	11170	한남동	73.0	1.0	아파트	65.00	2011.0	한남더힐
3289	서울특별시	용산구	한남동	11170	한남동	73.0	3.0	아파트	73.00	2011.0	한남더힐
3296	서울특별시	용산구	한남동	11170	한남동	71.0	1.0	아파트	50.50	2011.0	한남더힐
3297	서울특별시	용산구	한남동	11170	한남동	74.0	1.0	아파트	77.50	2011.0	한남더힐
3348	서울특별시	용산구	한남동	11170	한남동	73.0	-1.0	아파트	76.00	2011.0	한남더힐

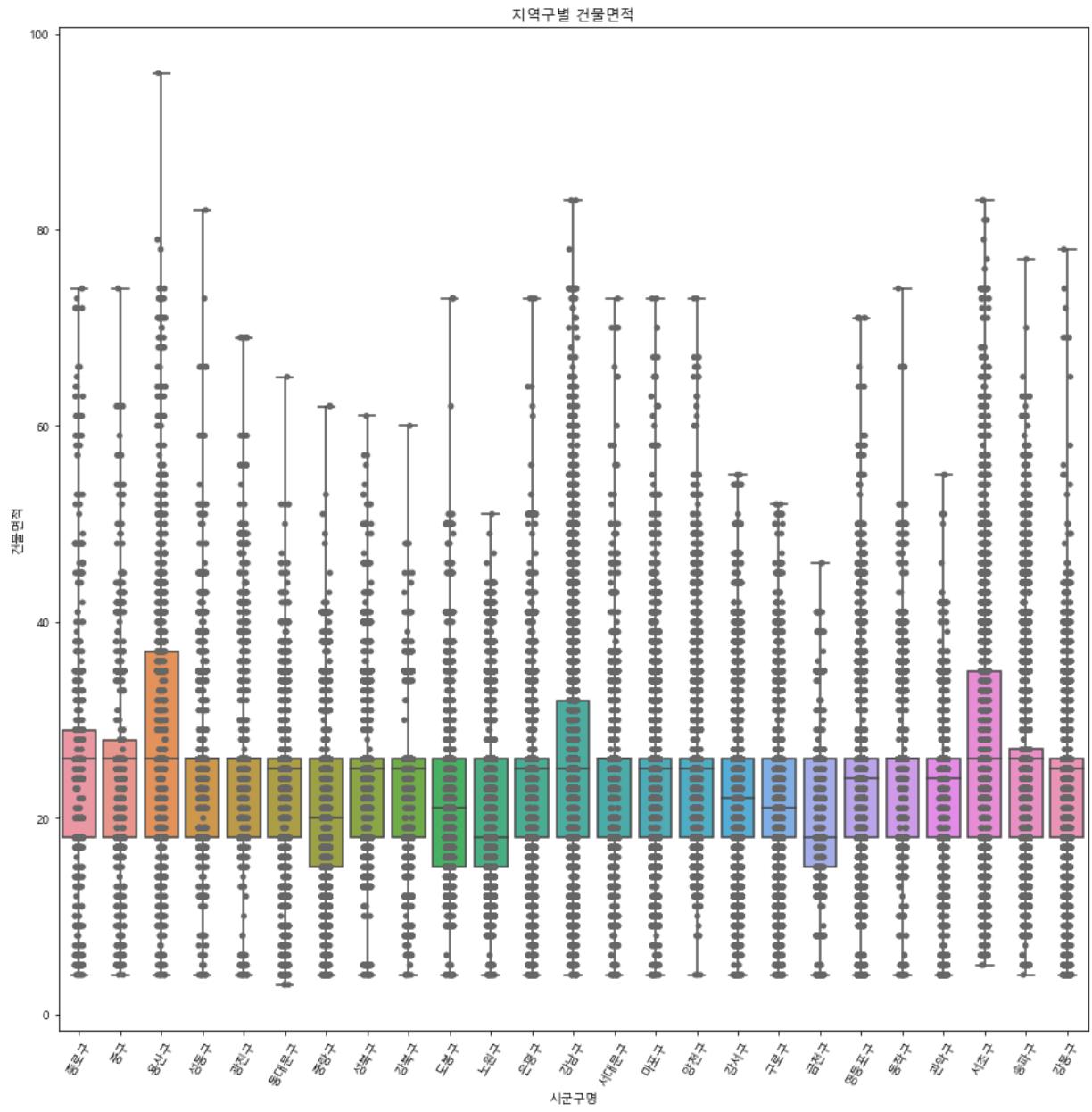
	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
6438	서울특별시	성동구	성수동1가	11200	성수동1가	73.0	32.0	아파트	53.70	2011.0	갤러리아 포레
6477	서울특별시	성동구	성수동1가	11200	성수동1가	82.0	44.0	아파트	67.00	2011.0	갤러리아 포레
75420	서울특별시	서초구	반포동	11650	반포동	51.0	11.0	아파트	54.00	2016.0	아크로리버파크
75495	서울특별시	서초구	반포동	11650	반포동	47.0	7.0	아파트	52.50	2016.0	아크로리버파크
75757	서울특별시	서초구	반포동	11650	반포동	54.0	31.0	아파트	55.00	2016.0	아크로리버파크
75887	서울특별시	서초구	반포동	11650	반포동	47.0	11.0	아파트	53.50	2016.0	아크로리버파크
75974	서울특별시	서초구	반포동	11650	반포동	47.0	25.0	아파트	54.00	2016.0	아크로리버파크
76743	서울특별시	서초구	서초동	11650	서초동	83.0	8.0	아파트	52.00	2002.0	트라움하우스3
78232	서울특별시	강남구	청담동	11680	청담동	61.0	5.0	아파트	55.80	2019.0	청담 어퍼하우스
78237	서울특별시	강남구	청담동	11680	청담동	83.0	12.0	아파트	95.00	2000.0	더펜트하우스청담
78252	서울특별시	강남구	청담동	11680	청담동	69.0	5.0	아파트	62.00	2019.0	효성빌라 청담101(A동)
78259	서울특별시	강남구	청담동	11680	청담동	61.0	3.0	아파트	53.00	2019.0	청담 어퍼하우스
78269	서울특별시	강남구	청담동	11680	청담동	60.0	5.0	아파트	52.00	2019.0	청담 어퍼하우스
78317	서울특별시	강남구	삼성동	11680	삼성동	59.0	23.0	아파트	57.00	2004.0	아이파크
78515	서울특별시	강남구	삼성동	11680	삼성동	59.0	7.0	아파트	55.00	2004.0	아이파크
79397	서울특별시	강남구	압구정동	11680	압구정동	59.0	10.0	아파트	51.75	1976.0	현대1차 (12,13,21,22,31,32,33 동)

	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명
79411	서울특별시	강남구	압구정동	11680	압구정동	60.0	12.0	아파트	52.00	1979.0	현대7차 (73~77,82,85동)
79418	서울특별시	강남구	압구정동	11680	압구정동	74.0	5.0	아파트	65.00	1979.0	현대7차 (73~77,82,85동)
79471	서울특별시	강남구	압구정동	11680	압구정동	74.0	9.0	아파트	67.00	1979.0	현대7차 (73~77,82,85동)
79527	서울특별시	강남구	압구정동	11680	압구정동	60.0	14.0	아파트	52.00	1976.0	현대2차 (10,11,20,23,24,25동)
79559	서울특별시	강남구	압구정동	11680	압구정동	62.0	15.0	아파트	54.00	1984.0	한양8(81동한개동)
79568	서울특별시	강남구	압구정동	11680	압구정동	60.0	12.0	아파트	52.00	1979.0	현대7차 (73~77,82,85동)
80551	서울특별시	강남구	도곡동	11680	도곡동	83.0	19.0	아파트	54.50	2009.0	로덴하우스 웨스트 빌리지
80589	서울특별시	강남구	도곡동	11680	도곡동	74.0	19.0	아파트	50.50	2009.0	로덴하우스 이스트 빌리지
80805	서울특별시	강남구	도곡동	11680	도곡동	71.0	67.0	아파트	54.25	2004.0	타워팰리스3
80888	서울특별시	강남구	도곡동	11680	도곡동	74.0	47.0	아파트	53.50	2002.0	타워팰리스1
80916	서울특별시	강남구	도곡동	11680	도곡동	64.0	4.0	아파트	56.54	2018.0	상지리츠빌 카일룸
80976	서울특별시	강남구	도곡동	11680	도곡동	67.0	55.0	아파트	55.50	2002.0	타워팰리스1

이 시점에 건물면적에 대한 통제가 필요하다고 생각했다. 실제로 확인해보니 건물면적이 굉장히 다양하게 나왔다.

In [36]:

```
plt.figure(figsize=(15, 15))
plt.xticks(rotation=60)
plt.title('지역구별 건물면적')
sns.boxplot(x='시군구명', y='건물면적', data=loc_apt, whis=np.inf)
sns.stripplot(x='시군구명', y='건물면적', data=loc_apt, jitter=True, color="0.4")
plt.show()
```



In [37]:

```
loc_apt = loc_apt
loc_apt['평당가'] = loc_apt['거래금액'] / loc_apt['건물면적']
loc_apt
```

Out[37]:

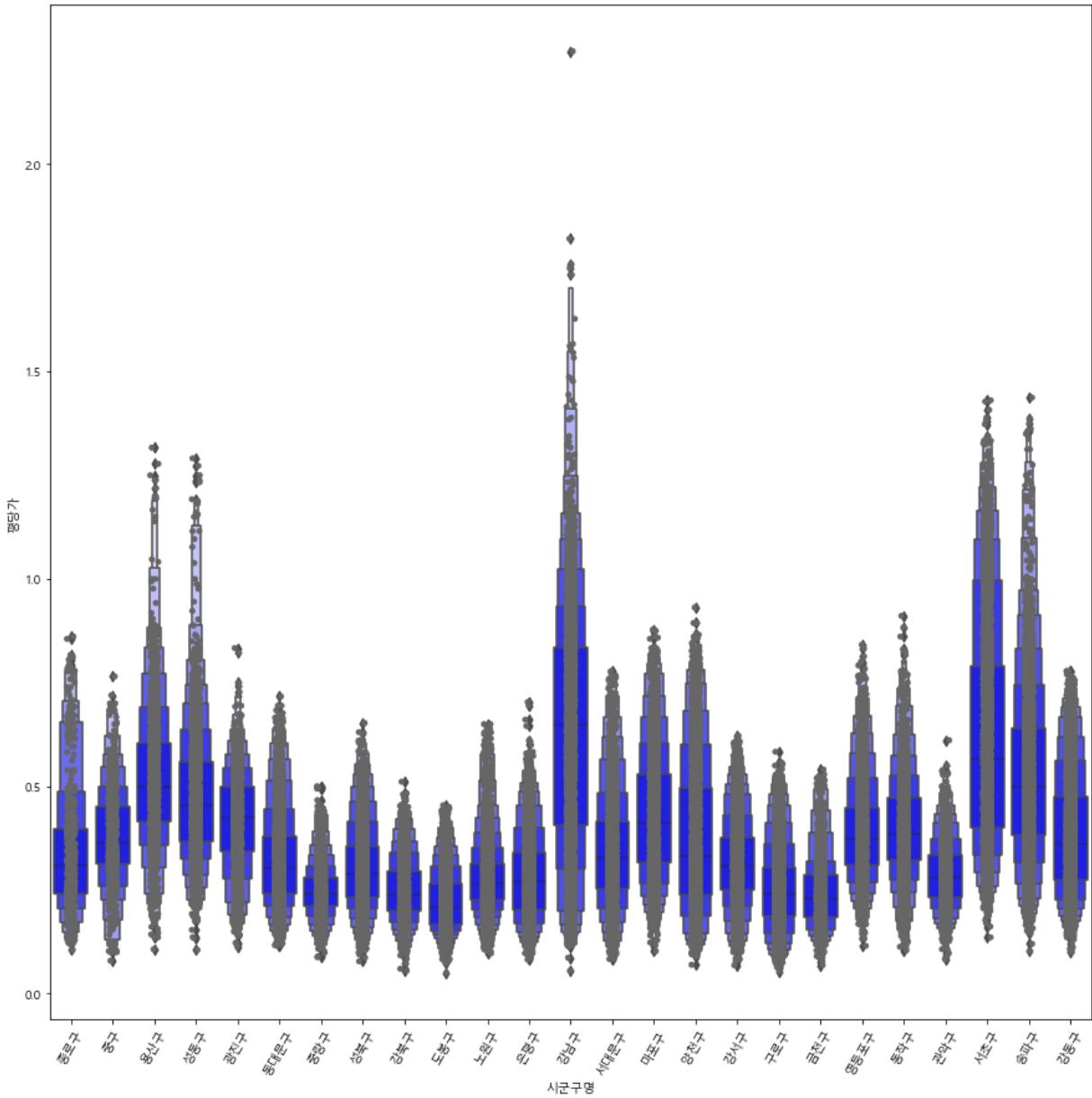
	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	총정보	건물주용도	거래금액	건축년도	아파트명	평당가
0	서울특별시	종로구	청운동	11110	청운동	26.0	3.0	아파트	8.95	2000.0	청운현대	0.344231
1	서울특별시	종로구	청운동	11110	청운동	26.0	3.0	아파트	8.50	2000.0	청운현대	0.326923
2	서울특별시	종로구	신교동	11110	신교동	41.0	5.0	아파트	10.00	2002.0	신현(102동)	0.243902
3	서울특별시	종로구	필운동	11110	필운동	33.0	7.0	아파트	9.40	2007.0	신동아블루아광화문의꿈	0.284848

	시도명	시군구명	읍면동명	지역코드	법정동명	건물면적	층정보	건물주용도	거래금액	건축년도	아파트명	평당가
4	서울특별시	종로구	필운동	11110	필운동	34.0	6.0	아파트	10.00	2007.0	신동아블루아광화문의꿈	0.294118
...
90661	서울특별시	강동구	강일동	11740	강일동	26.0	4.0	아파트	8.20	2009.0	강일리버파크6단지	0.315385
90662	서울특별시	강동구	강일동	11740	강일동	26.0	10.0	아파트	7.90	2009.0	강일리버파크3단지	0.303846
90663	서울특별시	강동구	강일동	11740	강일동	18.0	12.0	아파트	8.30	2009.0	강일리버파크3단지	0.461111
90664	서울특별시	강동구	강일동	11740	강일동	18.0	13.0	아파트	7.60	2009.0	강일리버파크7단지	0.422222
90665	서울특별시	강동구	강일동	11740	강일동	35.0	8.0	아파트	12.05	2011.0	고덕리엔파크2단지	0.344286

90666 rows × 12 columns

In [38]:

```
plt.figure(figsize=(15, 15))
plt.xticks(rotation=60)
plt.title('지역구별 평당가')
sns.boxenplot(x="시군구명", y="평당가",
               color="b",
               scale="linear",
               data=loc_apt)
sns.stripplot(x='시군구명', y='평당가', data=loc_apt, jitter=True, color="0.4")
plt.show()
```



분석결과

분명 지역구별로 아파트 매매가에 차이가 있는 것은 맞으나 지역구 단위보다 더 세세하게 나눠야 할 필요가 있었다.

- > 분석결과 : 지역구별 아파트 매매가의 차이를 확인했다. `barplot`으로 봤을 때의 이상치는 아래와 같은 이유가 있었다.
 - 희소성이 강한 조망권 프리미엄의 이유로 서울에서 최고가에 거래된 아파트의 대부분이 한강변에 몰려있다.(서초구)
 - 강남은 기존의 입지와 학군지 진입 수요차이가 가장 높아 거래 가격이 높다.
 - 그 뒤로 용산구와 종로구는 서울의 중심과 같아서 어디로든 왕래가 편하고 유동인구가 많아 거래가격이 높다.
 - 역세권과 비역세권의 차이로 분양권 프리미엄이 크게 달라져 지역간의 가격 차이가 크다.
 - 사업체가 많은 지역과 그렇지 않은 지역에서의 소득격차가 있기 때문에 아파트 매매가에도 영향을 미친다.
 - 서울 중심에 비해 외곽의 거래가는 상대적으로 낮음을 알 수 있다.



건축/재개발 기대감 상당한 한남, 성동, 광진구 신흥 부촌으로 떠올라

한강변을 따라 들어선 3.3㎡당 1억원을 넘긴 아파트를 찾기란 생각보다 쉬운 일이다. 최근 서울 아파트 가격을 견인하는 주요 키워드 역시 '재개발'과 '한강 조망권' 아파트다.

실제 한강 이남의 강남구 '암구정신현대' 전용 84㎡는 지난해 2월 34억 9천만원에 거래를 완료했다. 또한, 서초구 '아크로리버파크' 전용 84㎡ 역시 지난해 37억 2천만원에 거래됐다. 한강 이북에 자리한 한강 조망권 아파트도 인기몰이 중이다. 성동구 '트리마제' 전용 84㎡타입은 지난 1월 30억에 거래를 마쳤다. 재건축 기대감이 높은 용산구 '한강맨션' 전용 89㎡타입은 지난 2월에 26억원에 거래되면서 매서운 상승세를 입증했다.

출처 :

<http://www.ksilbo.co.kr/news/articleView.html?idxno=900393>

◇서울시 25개 자치구 중 유일하게 거래량 늘어난 강남구

3일 서울부동산정보광장에 따르면 4월 강남구 아파트 거래 수는 총 146건이었다. 지난 3월(137건)보다 소폭 증가했다. 4월 서울 아파트 거래량(3014건)이 3월(4415건)보다 31.7% 감소하는 와중에 유일하게 늘었다. 강남구와 함께 '강남 3구'로 불리는 서초구와 송파구와도 양상이 달랐다. 지난 4월 서초구의 아파트 거래량은 115건에서 92건으로, 송파구의 아파트 거래량은 148건에서 131건으로 거래가 줄었다.

거래가 이뤄진 물건들은 '사교육 1번지'라고 불리는 대치동에 물려있었다. 지난 4월 강남구 대치동 은마아파트 전용면적 76.79㎡ 7층은 17억4500만원에 거래됐다. 같은 면적 같은 층 매물은 지난해 12월 21억5000만원에 거래됐다. 4억원 정도 가격이 떨어진 셈이다. 최근 호가는 18억5000만~19억원대로 다시 올랐다. 강남구 대치동의 한보 미도맨션2차 전용면적 126.33㎡도 지난달 27억5000만원에 거래됐다. 지난 2월 동일 면적 주택의 실거래가보다 1억원 넘게 떨어졌다.

◇"학군지 진입 수요가 차이 만들었다"

출처 :

<https://m.post.naver.com/viewer/postView.nhn?volumeNo=28439115&memberNo=28162671>

정부의 부동산 규제 속에서도 지하철역과 인접한 '역세권' 아파트의 인기는 꾸준한 흐름을 나타내고 있다. 같은 지역이라도 역과의 거리에 따라 적게는 수천만원, 많게는 억 단위의 가격 상승 차이를 보이는가 하면, 신규 분양 단지의 경우 주변 '역' 유무에 따라 분양권 프리미엄도 크게 칼리는 것으로 드러났다.

특히 수도권 역세권 아파트와 비역세권 아파트의 집값 상승 차이는 두드러지게 나타났다. 실제 KB부동산시세에 따르면, 7호선 청담역과 인접한 청담현대 1차(1982년 12월 입주)의 전용면적 84m²는 지난 1년 동안(2017년 12월~2018년 12월) 4억5,000만원(10억원→14억5,000만원) 상승했다.

반면 청담역과 다소 거리가 먼 삼성청담공원아파트(1999년 8월 입주) 전용면적 84m²는 같은 기간 3억원(10억5,000만원→13억5,000만원), 대우유로카운티(2003년 10월 입주) 전용면적 105m²는 2억5,000만원(11억7,500만원→14억2,500만원) 올랐다. 역과의 거리가 벌어질 때마다 프리미엄 차이는 '역대'에 달했다.

역 거리에 따라 집값이 달라지다보니 분양권에는 높은 웃돈이 붙어 거래되고 있다. 국토교통부 아파트실거래가시스템을 살펴본 결과, 오는 6월 입주를 앞둔 성북역 역세권 아파트인 성북역 롯데캐슬 골드타운의 전용면적 84.91m² 분양권은 지난 3월, 분양가(5억3,000여만원) 보다 3억원 가량 오른 8억3,809만원(27층)에 거래될 정도로 높은 프리미엄을 형성하고 있다. 반면 역과 멀리 떨어진 비역세권 아파트는 프리미엄 없이 분양가 수준에 머무르고 있는 것으로 나타났다.

출처 :

<https://www.sedaily.com/NewsView/1VJCTFCVKW>

서울시 자치구별 사업체수 및 종사자수(서울시 통계)

사업체와 종사자가 가장 많은 자치구는 강남구이다.

자치구	사업체수	종사자수
강 남 구	142,746	1,388,272
강 등 구	59,544	291,980
강 북 구	37,836	144,836
강 서 구	77,048	496,408
관 악 구	52,210	227,888
광 진 구	49,070	255,758
구 로 구	77,106	436,140
금 천 구	64,944	468,024
노 월 구	53,232	237,426
도봉 구	37,788	148,628
동 대문 구	63,756	281,496
동 작 구	39,832	225,128
마 포 구	74,464	510,304
서 대문 구	39,592	218,980
서 초 구	93,080	877,466
성 풍 구	55,736	349,638
성 북 구	46,574	216,402
송 파 구	95,544	685,706
양 천 구	52,252	246,622
영 등 포 구	85,178	735,114
용 산 구	41,626	282,432
은 관 구	49,746	180,098
종로 구	79,904	530,034
중 구	121,914	781,060
중랑 구	56,048	206,032
합계	823,385	5,210,936

출처 : <https://invisiblacity.tistory.com/1522>

2. 연도별 아파트 건물면적과 층고

2-1. 건축년도가 최근일수록 건물면적이 좁을 것이다_ (pieplot & plot)

1, 2인 가구가 늘어나고 있음을 확인할 필요가 있어 외부 데이터를 가져왔다.

In [39]:

```
import pandas as pd
fam = pd.read_csv('.../data/가구원수.txt', sep="Wt", header=1)
fam
```

Out [39] :

	기간	구분	일반가구 수	1인	2인	3인	4인	5인	6인	7인 이 상	평균가 구원수
0	2000	합계	3,085,936	502,245	524,663	670,538	989,621	312,946	63,668	22,255	3.1
1	2000	종로구	56,284	12,784	10,593	11,384	14,259	5,447	1,286	531	2.9
2	2000	중구	44,461	10,353	8,562	9,119	11,116	4,173	839	299	2.9
3	2000	용산구	74,981	15,233	14,824	16,127	19,781	6,969	1,458	589	2.9
4	2000	성동구	102,285	17,205	18,657	22,370	30,598	10,238	2,364	853	3.1
...
203	2019	관악구	242,154	119,863	55,423	34,875	24,563	5,806	1,283	341	2.0
204	2019	서초구	155,377	42,667	38,892	35,327	29,850	6,704	1,503	434	2.5
205	2019	강남구	206,524	66,073	51,560	43,435	35,800	7,671	1,564	421	2.4
206	2019	송파구	252,212	68,861	65,402	57,821	47,607	10,014	2,038	469	2.5
207	2019	강동구	162,976	45,188	43,297	36,735	29,461	6,606	1,333	356	2.5

208 rows × 11 columns

In [40] :

```
fam = fam[fam['구분'] == '합계']
fam
```

Out [40] :

	기간	구 분	일반가구 수	1인	2인	3인	4인	5인	6인	7인 이 상	평균 가구 원수
0	2000	합계	3,085,936	502,245	524,663	670,538	989,621	312,946	63,668	22,255	3.1
26	2005	합계	3,309,890	675,739	670,455	732,043	917,243	248,450	49,781	16,179	2.9
52	2010	합계	3,504,297	854,606	781,527	788,115	807,836	209,061	47,659	15,493	2.7
78	2015	합계	3,784,490	1,115,744	930,467	817,440	701,945	169,436	38,547	10,911	2.5
104	2016	합계	3,784,705	1,138,860	931,262	816,946	686,469	163,555	37,153	10,460	2.5
130	2017	합계	3,813,260	1,180,540	956,978	813,629	662,920	154,992	34,649	9,552	2.4
156	2018	합계	3,839,766	1,229,421	977,469	807,545	637,385	147,231	32,059	8,656	2.4
182	2019	합계	3,896,389	1,299,787	1,005,065	801,819	614,026	138,596	29,336	7,760	2.3

```
In [41]: # 콤마 없애기
for i in range(2, 10):
    fam.iloc[:, i] = fam.iloc[:, i].str.replace(',', '').astype(int)
fam
```

C:\Users\Wowen\Anaconda3\lib\site-packages\pandas\core\indexing.py:1676: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
self._setitem_single_column(ilocs[0], value, pi)
```

Out[41]:

	기간	구분	일반가구 수	1인	2인	3인	4인	5인	6인	7인 이상	평균가구 원수
0	2000	합계	3085936	502245	524663	670538	989621	312946	63668	22255	3.1
26	2005	합계	3309890	675739	670455	732043	917243	248450	49781	16179	2.9
52	2010	합계	3504297	854606	781527	788115	807836	209061	47659	15493	2.7
78	2015	합계	3784490	1115744	930467	817440	701945	169436	38547	10911	2.5
104	2016	합계	3784705	1138860	931262	816946	686469	163555	37153	10460	2.5
130	2017	합계	3813260	1180540	956978	813629	662920	154992	34649	9552	2.4
156	2018	합계	3839766	1229421	977469	807545	637385	147231	32059	8656	2.4
182	2019	합계	3896389	1299787	1005065	801819	614026	138596	29336	7760	2.3

In [42]:

```
# 5인 이상으로 둑음
fam['5인 이상'] = fam[['5인', '6인', '7인 이상']].sum(axis=1)
fam
```

<ipython-input-42-5c44314aa91c>:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
fam['5인 이상'] = fam[['5인', '6인', '7인 이상']].sum(axis=1)
```

Out[42]:

	기간	구분	일반가구수	1인	2인	3인	4인	5인	6인	7인 이상	평균가구원수	5인 이상
0	2000	합계	3085936	502245	524663	670538	989621	312946	63668	22255	3.1	398869
26	2005	합계	3309890	675739	670455	732043	917243	248450	49781	16179	2.9	314410

기간	구분	일반가구수								평균가구원수	5인 이상	
			1인	2인	3인	4인	5인	6인	7인 이상			
52	2010	합계	3504297	854606	781527	788115	807836	209061	47659	15493	2.7	272213
78	2015	합계	3784490	1115744	930467	817440	701945	169436	38547	10911	2.5	218894
104	2016	합계	3784705	1138860	931262	816946	686469	163555	37153	10460	2.5	211168
130	2017	합계	3813260	1180540	956978	813629	662920	154992	34649	9552	2.4	199193
156	2018	합계	3839766	1229421	977469	807545	637385	147231	32059	8656	2.4	187946
182	2019	합계	3896389	1299787	1005065	801819	614026	138596	29336	7760	2.3	175692

In [43]:

```
# 빠른 나머지 컬럼 삭제
fam.drop(['5인', '6인', '7인 이상', '평균가구원수'], axis=1, inplace=True)
fam
```

C:\Users\Wowen\Anaconda3\lib\site-packages\pandas\core\frame.py:4308: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
 return super().drop()

Out[43]:

	기간	구분	일반가구수	1인	2인	3인	4인	5인 이상
0	2000	합계	3085936	502245	524663	670538	989621	398869
26	2005	합계	3309890	675739	670455	732043	917243	314410
52	2010	합계	3504297	854606	781527	788115	807836	272213
78	2015	합계	3784490	1115744	930467	817440	701945	218894
104	2016	합계	3784705	1138860	931262	816946	686469	211168
130	2017	합계	3813260	1180540	956978	813629	662920	199193
156	2018	합계	3839766	1229421	977469	807545	637385	187946
182	2019	합계	3896389	1299787	1005065	801819	614026	175692

In [44]:

```
labels = fam.columns[3:8]
explode = [0.05, 0.05, 0.05, 0.05, 0.05]
colors = ['#ffff00', '#66ff33', '#cc33ff', '#3366ff', '#993300']

fig = plt.figure(figsize=(15, 10))
for i in range(4):
    plt.subplot(40 + i + 1)
    plt.title(str(fam.iat[i, 0]) + "년도")
    plt.pie(fam.iloc[i, 3:8],
            labels=labels,
```

```

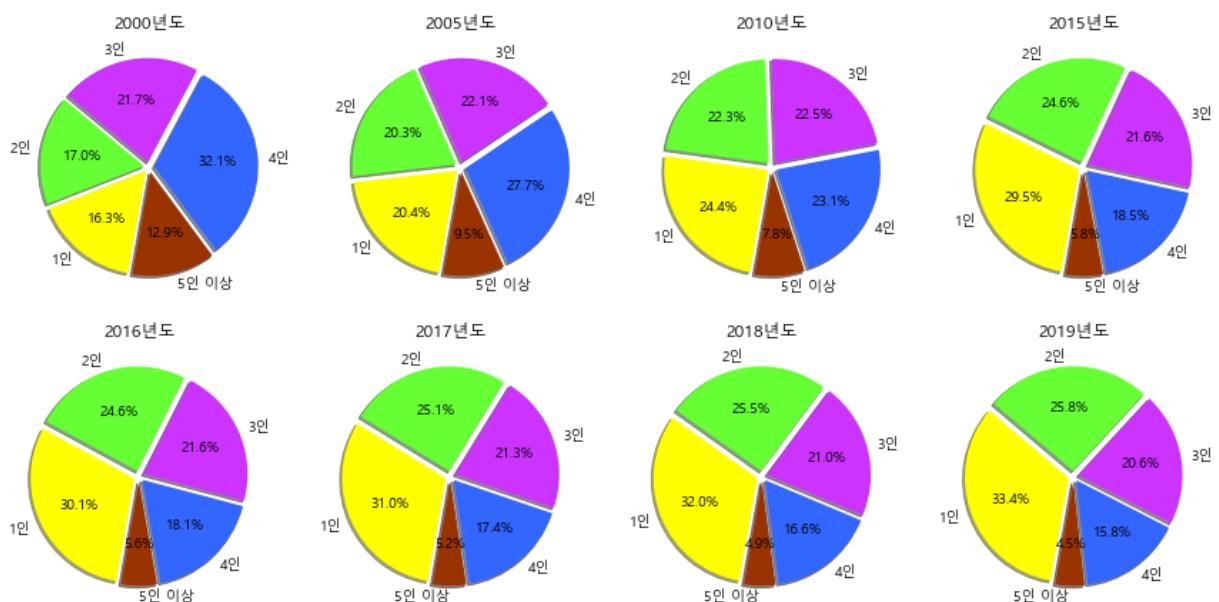
autopct='%.1f%%',
startangle=260,
counterclock=False,
explode=explode,
shadow=True,
colors=colors)

plt.show()

fig = plt.figure(figsize=(15, 10))
for i in range(4, 8):
    plt.subplot(240 + i - 3)
    plt.title(str(fam.iat[i, 0]) + "년도")
    plt.pie(fam.iloc[i, 3:8],
            labels=labels,
            autopct='%.1f%%',
            startangle=260,
            counterclock=False,
            explode=explode,
            shadow=True,
            colors=colors)

plt.show()

```



분석결과

시간이 지남에 따라 1,2인 가구는 점점 늘고, 4인 이상 가구는 점점 줄어드는 것을 확인 할 수 있다.

위 결과로 미루어 봤을 때 오래된 아파트일수록 면적이 넓을 것으로 예상했다.

In [45]:

```
# 년도별 평균 건물면적
df2_1 = bby_apt.groupby(['건축년도'])[['건물면적']].mean()
df2_1.head()
```

Out [45]:

건축년도	건물면적
1961.0	15.444444
1965.0	14.000000

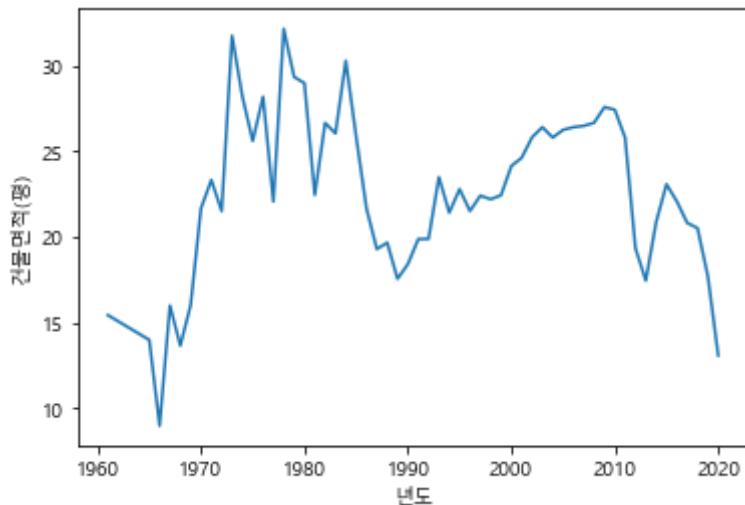
건물면적

건축년도

1966.0	9.000000
1967.0	16.000000
1968.0	13.666667

In [46]:

```
# 서울 전지역 년도별 건물면적
plt.plot(df2_1)
plt.xlabel('년도')
plt.ylabel('건물면적(평)')
plt.show()
```



분석결과

상관관계 파악이 어렵지만 가구원수가 급격히 줄어든 게 건물면적에 분명한 영향을 끼쳤음을 알 수 있다.

> 2000~2020년도의 데이터를 봤을 때 1, 2인 가구가 급격히 늘어난 최근에 들어서는 건물면적이 줄어드는 것을 확인할 수 있다.

2-2. 건축년도가 최근일수록 층고가 높을 것이다 _ (plot)

In [47]:

```
df3_1 = bgy_apartment.groupby(['건축년도'])[['층정보']].mean()
df3_1.head()
```

Out[47]:

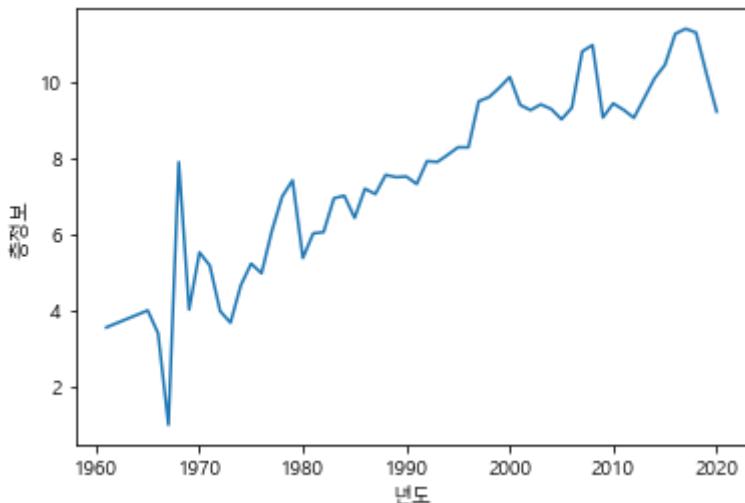
층정보

건축년도

1961.0	3.555556
1965.0	4.000000
1966.0	3.400000
1967.0	1.000000
1968.0	7.888889

In [48]:

```
plt.plot(df3_1)
plt.xlabel('년도')
plt.ylabel('총정보')
plt.show()
```



분석결과

최신 아파트일수록 총수가 높아지는 것을 확인 할 수 있다.

Feedback

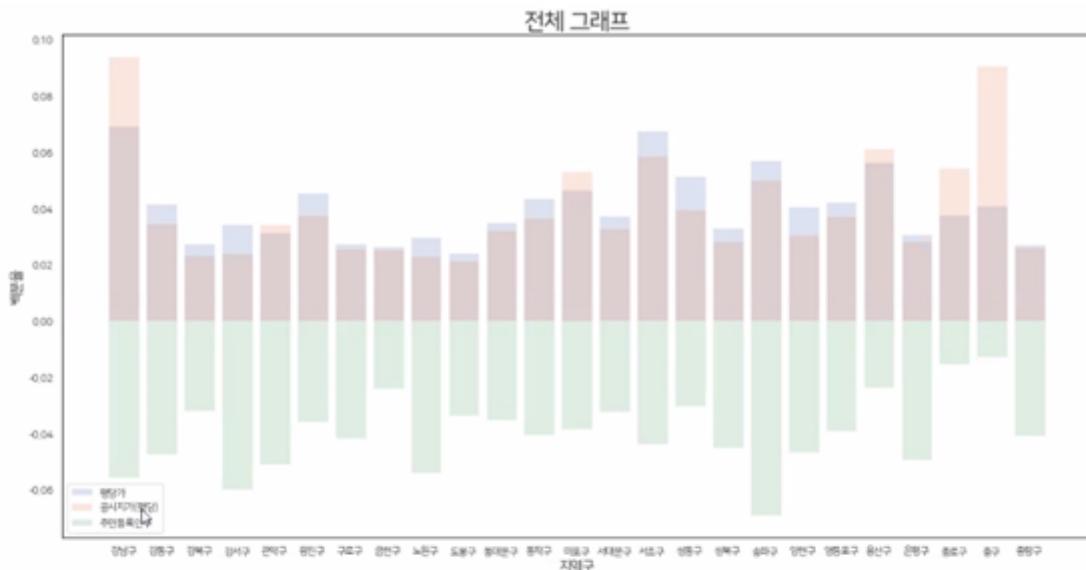
- 이미지, 분석결과 잘 활용했다. 하지만, 마크다운은 살짝 아쉬웠음
- 건축년도별 아파트매매가 분석 디테일하게 잘했다.
- 가구원수 데이터 좋았다. 파이차트를 활용해 시각화와 함께 수치를 보여주는 것 좋았다.

아쉬운 점

- 브랜드 선호도에 대해 전처리가 어려워서 마무리하지 못한 점이 아쉽다.
- 건축년도별 건물면적에 영향을 끼친 다른 요인을 파악하지 못한 점이 아쉬웠다.

다른 팀 잘한 거

공시지가 사용



분석결과

공시지가와 평당가는 어느정도 관련성이 있는 것을 알 수 있지만 몇 부분의 의외인 부분이 있는데 바로 종구, 종로구, 강남구이다. 이 부문은 평당가보다 공시지가의 가격이 월등히 높은데 이 지역들은 주거지역보다 상권으로 가치가 매겨져 아파트 거래가격인 평당가보다 공시지가가 높게 나왔다고 생각할 수 있다.

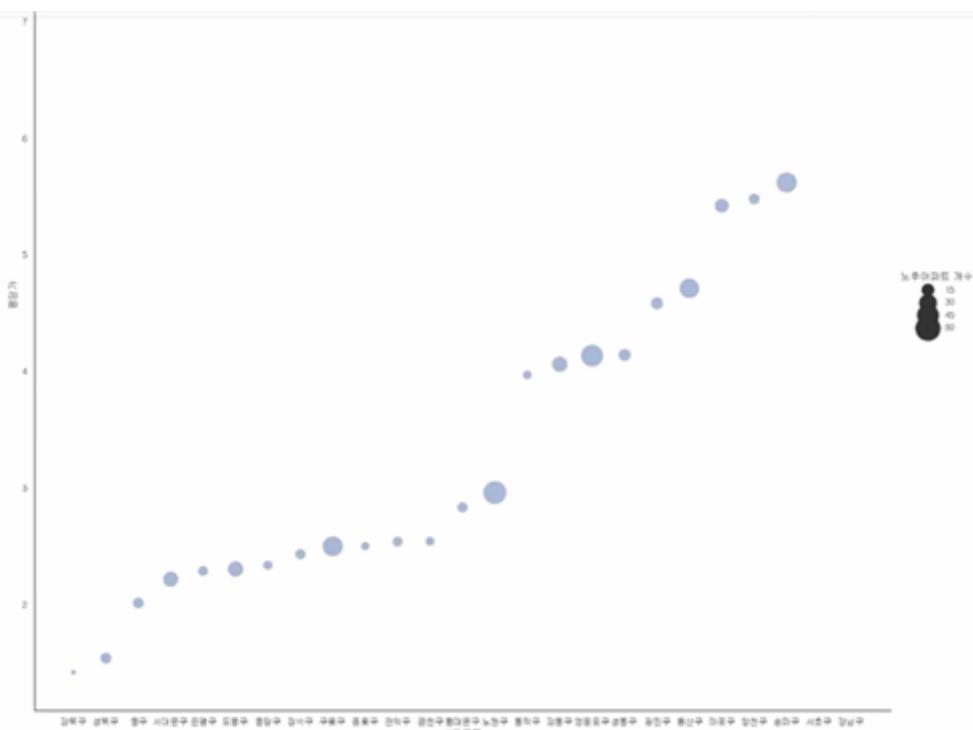
또한 당연히 비활수록 적게 살고 저렴할수록 더 많이 거주할세라고 생각했는데 결과를 보면 그렇지 않을수 있다.

데이터 전처리 과정 요약

데이터 전처리

- 1) 주거형태 아파트만 한정
 - 2) 건물연적을 평으로 수정
 - 3) 건축년도 0.0년 2020년으로 수정
 - 4) 법정동코드는 [시군구명_읍면동명]만 사용

지역구+평당가+노후아파트



구획한

```
# 1층 / 저층(2~4) / 중층(5~9) / 고층(10~14) / 초고층(그이상)
def floor(x):
    str = x
    if x < 0:
        str = '0_지하'
    elif x == 1:
        str = '1_1층'
    elif x<= 4:
        str= '2_저층'
    elif x<= 9:
        str= '3_중층'
    elif x<= 14:
        str = '4_고층'
    else:
        str ='5_초고층'
    return str

df['층구분']= df['층정보'].map(floor)
df
```