

부동산 커뮤니티 조회수와 주택 거래량의 상관관계

◆ 주택유형별 주택거래현황

- 1) 데이터 전처리
- 2) 2020년 주택거래량 시각화
- 2) 2021년 주택거래량 시각화
- 3) 주택거래량 증감률 시각화

◆ 부동산 커뮤니티 조회수

- 1) 웹 스크래핑
- 2) 데이터 전처리
- 3) 2020년 조회수 합 시각화
- 4) 2021년 조회수 합 시각화
- 5) 조회수 증감률 시각화

◆ 통계분석

- 1) 피어슨 상관계수
- 2) 이상치 제거
- 3) 상관계수 분포 시각화

보드 보기

+ 보기 추가

필터

정렬

Q

...

새로

Schedule&Notice 3

세미 1차프로젝트

1차 세미 프로젝트

4조_프로젝트 주제

+ 새로 만들기

Not Started 0

+ 새로 만들기

In Progress 7

데이터

데이터 전처리

커뮤니티 데이터

웹 스크래핑

데이터 시각화

통계분석

발표자료

+ 새로 만들기

Completed 2

브레인스토밍

데이터 전처리(完)

+ 새로 만들기

데이터 전처리

상태 In Progress

담당자 비어 있음

+ 속성 추가

댓글 추가

재현

은형

선빈

윤민

주택유형별 주택거래현황



KOSIS 국가통계포털

주택유형별 주택거래현황

[주택유형별 주택거래현황 - 데이터 전처리]

원본데이터에서 불필요한 행·열을 삭제하고 필터링을 한 후 `df.drop()`, `df[df['칼럼명'] == 값]`

직관적으로 이해할 수 있도록 칼럼명을 변경함 `df.rename()`

원본데이터

	A	B	C	D	E	F	G	H	I
1	행정구역별	행정구역별	유형별(1)	유형별(2)	2020.01	2020.01	2020.02	2020.02	2020.03
2	행정구역별	행정구역별	유형별(1)	유형별(2)	동(호)수 (면적 (천㎡)	동(호)수 (면적 (천㎡)	동(호)수 (면적 (천㎡)	동(호)수 (면적 (천㎡)	동(호)수 (면적 (천㎡)
3	서울특별시	종로구	주택유형	소계	356	29	299	22	270
4	서울특별시	종로구	주택유형	단독주택	41	4	54	5	48
5	서울특별시	종로구	주택유형	다가구주택	5	1	6	2	5
6	서울특별시	종로구	주택유형	다세대주택	88	5	112	6	91
7	서울특별시	종로구	주택유형	연립주택	28	3	36	3	22
8	서울특별시	종로구	주택유형	아파트	194	16	91	6	104
9	서울특별시	중구	주택유형	소계	317	23	251	18	185
10	서울특별시	중구	주택유형	단독주택	50	5	45	3	29
11	서울특별시	중구	주택유형	다가구주택	2	1	9	2	5
12	서울특별시	중구	주택유형	다세대주택	70	3	66	2	24
13	서울특별시	중구	주택유형	연립주택	9	1	7	0	11
14	서울특별시	중구	주택유형	아파트	186	14	124	10	116
15	서울특별시	용산구	주택유형	소계	958	69	760	54	461
16	서울특별시	용산구	주택유형	단독주택	87	11	69	7	79
17	서울특별시	용산구	주택유형	다가구주택	19	4	13	3	24
18	서울특별시	용산구	주택유형	다세대주택	198	12	200	8	137
19	서울특별시	용산구	주택유형	연립주택	20	2	30	2	35
20	서울특별시	용산구	주택유형	아파트	634	40	448	34	186



행·열삭제및필터링&칼럼명변경

	구	2020.01	2020.02	2020.03	2020.04	2020.05	2020.06	2020.07	2020.08	2020.09	...
1	종로구	356	299	270	173	206	273	500	315	198	...
7	종로구	317	251	185	116	139	247	452	432	550	...
13	용산구	958	760	461	461	1059	598	1030	509	362	...
19	성동구	596	549	465	280	241	542	1808	520	375	...
25	광진구	524	468	506	383	393	556	907	682	901	...
31	동대문구	1058	1322	914	479	553	783	1316	1156	590	...
37	중랑구	552	855	1018	627	614	790	1170	754	798	...
43	성북구	1256	1350	1513	710	696	1117	1758	981	998	...

[주택유형별 주택거래현황 - 데이터 전처리]

구별 데이터로 기준열을 설정하고 데이터프레임을 정렬한 후 `df.index()`

데이터 타입을 실수형으로 바꾸고, 보기 편하도록 행·열을 전치함 `df.astype() / df.T`

기준열 설정 & 데이터 타입 변경

	2020.01	2020.02	2020.03	2020.04	2020.05	2020.06	2020.07	2020.08	2020.09	2020.10	...
구											
종로구	356.0	299.0	270.0	173.0	206.0	273.0	500.0	315.0	198.0	184.0	...
중구	317.0	251.0	185.0	116.0	139.0	247.0	452.0	432.0	550.0	223.0	...
용산구	958.0	760.0	461.0	461.0	1059.0	598.0	1030.0	509.0	362.0	335.0	...
성동구	596.0	549.0	465.0	280.0	241.0	542.0	1808.0	520.0	375.0	196.0	...
광진구	524.0	468.0	506.0	383.0	393.0	556.0	907.0	682.0	901.0	392.0	...
동대문구	1058.0	1322.0	914.0	479.0	553.0	783.0	1316.0	1156.0	590.0	470.0	...
중랑구	552.0	855.0	1018.0	627.0	614.0	790.0	1170.0	754.0	798.0	610.0	...
성북구	1256.0	1350.0	1513.0	710.0	696.0	1117.0	1758.0	981.0	998.0	827.0	...



행·열 전치

구	종로 구	중구	용산 구	성동 구	광진 구	동대 문구	중랑 구	성북 구	강북 구	도봉 구	...
2020.01	356.0	317.0	958.0	596.0	524.0	1058.0	552.0	1256.0	868.0	764.0	...
2020.02	299.0	251.0	760.0	549.0	468.0	1322.0	855.0	1350.0	757.0	728.0	...
2020.03	270.0	185.0	461.0	465.0	506.0	914.0	1018.0	1513.0	837.0	953.0	...
2020.04	173.0	116.0	461.0	280.0	383.0	479.0	627.0	710.0	609.0	513.0	...
2020.05	206.0	139.0	1059.0	241.0	393.0	553.0	614.0	696.0	656.0	559.0	...
2020.06	273.0	247.0	598.0	542.0	556.0	783.0	790.0	1117.0	934.0	1016.0	...
2020.07	500.0	452.0	1030.0	1808.0	907.0	1316.0	1170.0	1758.0	1241.0	1373.0	...
2020.08	315.0	432.0	509.0	520.0	682.0	1156.0	754.0	981.0	655.0	705.0	...
2020.09	198.0	550.0	362.0	375.0	901.0	590.0	798.0	998.0	578.0	588.0	...
2020.10	184.0	223.0	335.0	196.0	392.0	470.0	610.0	827.0	722.0	515.0	...
2020.11	214.0	236.0	484.0	429.0	560.0	373.0	440.0	587.0	466.0	472.0	...
2020.12	296.0	251.0	605.0	643.0	560.0	670.0	704.0	909.0	764.0	772.0	...
2021.01	182.0	177.0	412.0	365.0	448.0	660.0	765.0	1019.0	684.0	771.0	...
2021.02	289.0	184.0	438.0	387.0	471.0	502.0	576.0	1190.0	680.0	879.0	...
2021.03	227.0	129.0	395.0	762.0	498.0	496.0	680.0	688.0	858.0	648.0	...

[주택유형별 주택거래현황 - 데이터 전처리]

피벗테이블을 활용하여 년도별 데이터를 생성한 후 증감률을 계산하여 칼럼을 추가함 $\text{증감률} = (\text{21년도주택거래량} - \text{20년도주택거래량}) / (\text{20년도주택거래량})$

json 파일과 choropleth 를 그려 시각화를 진행하고, 구글맵스 라이브러리를 활용하여 마커를 생성함 `import json / import folium / from folium.features import DivIcon`

피벗테이블&증감률칼럼추가

	2020년도	2021년도	증감률
구			
강남구	16097.0	14298.0	-11.175996
강동구	16559.0	10395.0	-37.224470
강북구	9087.0	6856.0	-24.551557
강서구	16214.0	11557.0	-28.722092
관악구	9517.0	6797.0	-28.580435
광진구	6832.0	6305.0	-7.713700
구로구	10790.0	7770.0	-27.988879
금천구	6383.0	5619.0	-11.969293
노원구	15439.0	7448.0	-51.758534
도봉구	8958.0	7799.0	-12.938156
동대문구	9684.0	5663.0	-41.522098

json파일활용& folium choropleth 생성

```
{'type': 'FeatureCollection',
 'features': [{'type': 'Feature',
  'properties': {'SIG_CD': '11320',
    'SIG_KOR_NM': '도봉구',
    'SIG_ENG_NM': 'Dobong-gu',
    'ESRI_PK': 0,
    'SHAPE_AREA': 0.00211,
    'SHAPE_LEN': 0.239901},
  'seoul_municipalities_geo.json'}
```

```
import folium
from folium.features import DivIcon

center = [37.5502, 126.982]
map = folium.Map(center, zoom_start=11, tiles='cartodbdark_matter')
# #map

map.choropleth(geo_data=json_file,
  data=df2['2020년'],
  columns=[df2.index, df2['2020년']],
  fill_color = 'BuPu',
  key_on = feature.properties.SIG_KOR_NM)
```

구글맵스 위치데이터활용&마커생성

```
{'종로구청': (37.5734684, 126.978984),
 '중구청': (37.5637584, 126.9975517),
 '용산구청': (37.532454, 126.9902471),
 '성동구청': (37.5634092, 127.0369449),
 '광진구청': (37.5385333, 127.0823772),
 '동대문구청': (37.5744155, 127.0397427),
 '중랑구청': (37.60630460000001, 127.0931523),
 '성북구청': (37.589366, 127.016743),
 '강북구청': (37.6397767, 127.0255184),
 '도봉구청': (37.6687738, 127.0470706),
```

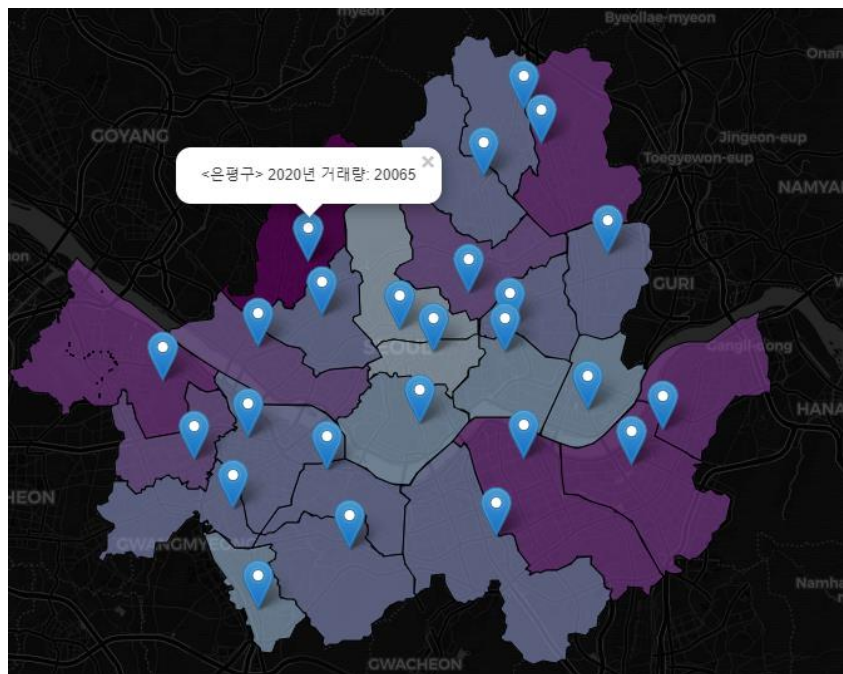
```
for name, coordinate in coordinate_list.items():
  folium.Marker([coordinate[0], coordinate[1]],
    popup = '<' + name[:1] + '>' \
      + '\n' + '2020년 거래량:' + '\n' \
      + str(pivot_df.loc[name[:1], '2020년도'])
    ).add_to(map)
```

[2020년 주택유형별 주택거래현황 시각화]

2020년 기준, 은평구의 주택거래량은 20,065건으로 가장 거래량이 많음

이에 반해, 종로구와 중구는 각각 3,284건/2,626건으로 거래량이 상대적으로 저조함

2020년 주택거래량 시각화



지역별 거래량(내림차순)

구	2020년도	2021년도	증감률
은평구	20065.0	14771.0	-26.384251
송파구	17055.0	13124.0	-23.048959
강동구	16559.0	10395.0	-37.224470
강서구	16214.0	11557.0	-28.722092
강남구	16097.0	14298.0	-11.175996
노원구	15439.0	7448.0	-51.758534
양천구	12811.0	7613.0	-40.574506
성북구	12702.0	8743.0	-31.168320
마포구	11689.0	7246.0	-38.010095
서대문구	11483.0	5231.0	-54.445702

지역별 거래량(오름차순)

구	2020년도	2021년도	증감률
종로구	3284.0	2770.0	-15.651644
중구	3399.0	2626.0	-22.741983
금천구	6383.0	5619.0	-11.969293
성동구	6644.0	3909.0	-41.164961
광진구	6832.0	6305.0	-7.713700
용산구	7622.0	4977.0	-34.702178
중랑구	8932.0	7543.0	-15.550828
도봉구	8958.0	7799.0	-12.938156
강북구	9087.0	6856.0	-24.551557
관악구	9517.0	6797.0	-28.580435

[2021년 주택유형별 주택거래현황 시각화]

2021년 기준, 은평구의 주택거래량은 14,771건으로 작년에 이어 가장 거래량이 많음
이에 반해, 종로구와 중구는 각각 3,284건/3,399건으로 거래량이 상대적으로 저조함

2021년 주택거래량 시각화



지역별 거래량(내림차순)

	2020년도	2021년도	증감률
구			
은평구	20065.0	14771.0	-26.384251
강남구	16097.0	14298.0	-11.175996
송파구	17055.0	13124.0	-23.048959
강서구	16214.0	11557.0	-28.722092
서초구	10764.0	11249.0	4.505760
강동구	16559.0	10395.0	-37.224470
성북구	12702.0	8743.0	-31.168320
도봉구	8958.0	7799.0	-12.938156
구로구	10790.0	7770.0	-27.988879
양천구	12811.0	7613.0	-40.574506

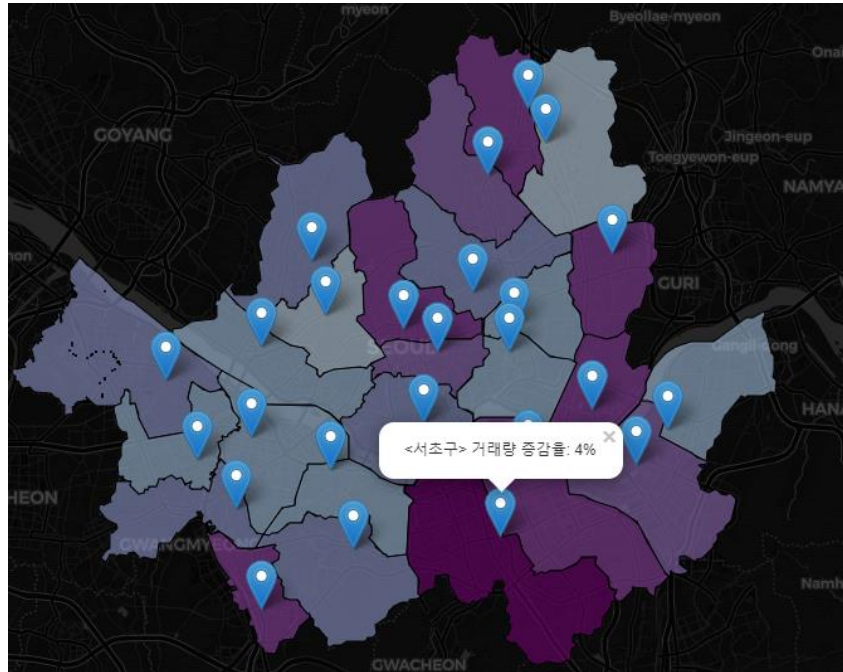
지역별 거래량(오름차순)

	2020년도	2021년도	증감률
구			
중구	3399.0	2626.0	-22.741983
종로구	3284.0	2770.0	-15.651644
성동구	6644.0	3909.0	-41.164961
용산구	7622.0	4977.0	-34.702178
서대문구	11483.0	5231.0	-54.445702
금천구	6383.0	5619.0	-11.969293
동대문구	9684.0	5663.0	-41.522098
영등포구	10895.0	6072.0	-44.268013
광진구	6832.0	6305.0	-7.713700
동작구	10990.0	6622.0	-39.745223

[주택유형별 주택거래현황 증감률 시각화]

2020년 대비, 서초구/광진구/강남구는 타지역 대비 매매량이 유지가 됨
이에 반해, 서대문구와 노원구의 매매량은 각각 약 -54%, -51% 하락함

2020년 주택거래량 시각화



지역별 거래량(내림차순)

구	2020년도	2021년도	증감률
서초구	10764.0	11249.0	4.505760
광진구	6832.0	6305.0	-7.713700
강남구	16097.0	14298.0	-11.175996
금천구	6383.0	5619.0	-11.969293
도봉구	8958.0	7799.0	-12.938156
중랑구	8932.0	7543.0	-15.550828
종로구	3284.0	2770.0	-15.651644
중구	3399.0	2626.0	-22.741983
송파구	17055.0	13124.0	-23.048959
강북구	9087.0	6856.0	-24.551557

지역별 거래량(오름차순)

구	2020년도	2021년도	증감률
서대문구	11483.0	5231.0	-54.445702
노원구	15439.0	7448.0	-51.758534
영등포구	10895.0	6072.0	-44.268013
동대문구	9684.0	5663.0	-41.522098
성동구	6644.0	3909.0	-41.164961
양천구	12811.0	7613.0	-40.574506
동작구	10990.0	6622.0	-39.745223
마포구	11689.0	7246.0	-38.010095
강동구	16559.0	10395.0	-37.224470
용산구	7622.0	4977.0	-34.702178

부동산 커뮤니티 조회수



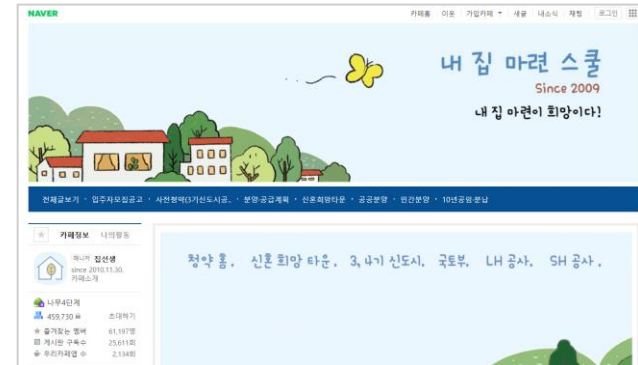
부동산스터디

카페회원수 1,826,371명



아름다운내집갯기

카페회원수 1,106,491명



내집마련스쿨

카페회원수 459,730명

[웹 스크래핑 - 스크래핑]

Selenium의 Chrome webdriver를 설치하여 진행

```
1 from webdriver_manager.chrome import ChromeDriverManager
2 from selenium.webdriver.chrome.service import Service
3 from selenium import webdriver
4
5 from urllib.request import urlopen
6 from bs4 import BeautifulSoup
7
8 import numpy as np
9 import pandas as pd
10 import time
11 import re

1 service = Service(executable_path=ChromeDriverManager().install())
2 driver = webdriver.Chrome(service=service)
```

크롬 드라이버 다운 로드 : <https://chromedriver.chromium.org/downloads>

셀레니움 드라이버 설치 참고 <https://greeksharifa.github.io/references/2020/10/30/python-selenium-usage/>

[웹 스크래핑 - 스크래핑]

딕셔너리에 구별 menuid를 저장하여 사용

네이버카페의 경우 <iframe> 안에 있어 접근 후 카테고리 내 태그의 정보를 가져올 수 있음

```
# 각 구별 url을 불러오기 위한 키값
```

```
gu_dict = { '강남구':160, '강동구':162, '강북구':149, '강서구':156, '관악구':180,  
            '광진구':205, '구로구':157, '금천구':189, '노원구':151, '도봉구':207,  
            '동대문구':150, '동작구':159, '마포구':203, '서대문구':155, '서초구':179,  
            '성동구':152, '성북구':148, '송파구':161, '양천구':202, '영등포구':158,  
            '용산구':154, '은평구':141, '종로구':142, '중구':143, '중랑구':153 }  
}
```

```
# 각 구를 대상으로 반복
```

```
for key in gu_dict:
```

```
# 페이지를 넘겨가며 스크래핑 반복
```

```
while True:
```

```
    url = 'https://cafe.naver.com/rainup?iframe_url=/ArticleList.nhn%3Fsearch.  
    driver.get(url)
```

```
# 네이버 카페 본문 접근
```

```
driver.switch_to.frame('cafe_main')
```

```
# 페이지 +1
```

```
page += 1
```

[웹 스크래핑 - 스크래핑]

네이버카페의 경우 각 카테고리 <iframe>에 있음 -> iframe에 들어가지 않을 경우 데이터를 불러올 수 없음

따라서 <iframe>에 들어가지 않은 아래의 경우 데이터가 없어 get_text()를 사용할 수 없다고 나와 있음

In [38]:

```
1 # 서울시 '구'별 페이지 html 불러오기
2 web.find('div',{'class':'list-style'}).get_text()
3
```

AttributeError Traceback (most recent call last)

<ipython-input-38-e743f8b47193> in <module>

```
1 # 서울시 '구'별 페이지 html 불러오기
----> 2 web.find('div',{'class':'list-style'}).get_text()
```

AttributeError: 'NoneType' object has no attribute 'get_text'

위의 Attribute Error를 해결하기 위해서 <iframe>으로 들어가는 코드를 추가

추가 코드 : driver.switch_to.frame('café_main')

[웹 스크래핑 - 스크래핑]

정규 표현식을 작성하여 2020년, 2021년에 작성된 게시글을 가져옴
이 조건을 이용하여 다음 구로 넘어가 스크래핑을 진행할 수 있도록 함

```
# 정규표현식 작성
```

```
date_match1 = re.compile("^(202)[0-1]\.\d{2}") #2020-2021 검사
```

```
date_match2 = re.compile("^(2019)\.\d{2}") #2019 검사
```

```
# 정규표현식 검사 ( 페이지를 넘기기 위한 )
```

```
for day in days:
```

```
    if date_match1.search(day): # 2020.01. ~ 2021.12.  
        include1.append(day)
```

```
for day in days:
```

```
    if date_match2.search(day): # 2019.01  
        include2.append(day)
```

[웹 스크래핑 - 데이터 전처리]

이 프로젝트에서 필요한 데이터는 구별 게시물 수, 각 게시물의 조회수이고,
전처리한 데이터를 월별, 연별로 분석할 것이므로 이에 맞도록 데이터를 수정

```
# '강남구' df에 구별 조회수 df join
for data in gu_list:
    df = pd.read_excel(data+'.xlsx')
    df2 = df.iloc[:,0:2] 날짜데이터에서 '일' 데이터를 제외
    df2.rename(columns={'게시물': data }, inplace=True)
    df2 = df2.set_index('월')
    df_gu = df_gu.join(df2) 25개구의 데이터를 하나의 데이터프레임으로 join
```

구별 조회수, 게시물 수 전처리 진행

[웹 스크래핑 - 데이터 전처리]

이 프로젝트에서 필요한 데이터는 구별 게시물 수, 각 게시물의 조회수이고,
전처리한 데이터를 월별, 연별로 분석할 것이므로 이에 맞도록 데이터를 수정

```
df_gu.to_excel('구별 조회수 합산.xlsx') # 구별 조회수 데이터
조회수_변화.to_excel('커뮤니티 조회수 증감률.xlsx') # 구별 조회수 증감률 데이터
게시물수_변화.to_excel('구별 게시물수.xlsx') # 구별 게시물수 데이터
```

	A	B	C	D	E	F
1	월	강남구	강동구	강북구	강서구	관악구
2	2020.01	68	26	5	56	9
3	2020.02	54	26	2	88	18
4	2020.03	59	22	3	136	10
5	2020.04	73	30	4	33	3
6	2020.05	42	19	4	53	4
7	2020.06	51	74	6	69	6
8	2020.07	89	41	17	49	4
9	2020.08	51	52	8	38	5
10	2020.09	37	43	5	41	5
11	2020.1	28	38	6	41	8
12	2020.11	15	46	1	13	5
13	2020.12	14	58	4	28	12
14	2021.01	21	30	8	17	6

구별 게시물수.xlsx

	A	B	C	D	E	F
1	월	강남구	강동구	강북구	강서구	관악구
2	2020.01	26283	20409	6459	43389	3781
3	2020.02	20561	16538	633	63455	10304
4	2020.03	27494	22820	1763	83854	10120
5	2020.04	19888	21676	828	10701	2316
6	2020.05	13681	12945	2074	24176	2325
7	2020.06	22687	61917	2000	29364	5175
8	2020.07	36332	35711	5925	24861	1603
9	2020.08	17558	34121	3790	25870	1196
10	2020.09	8033	20918	3310	11410	2029
11	2020.1	10686	28541	4243	15444	3270
12	2020.11	5405	47003	0	5807	2911
13	2020.12	4913	71718	5165	17266	4512
14	2021.01	12170	32543	5519	7451	1634

구별 조회수 합산.xlsx

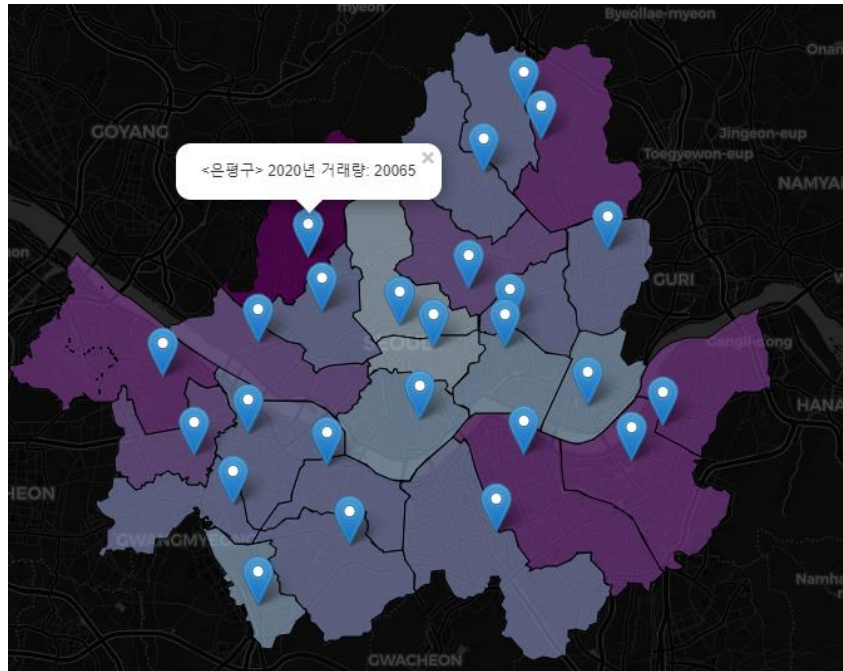
	A	B	C	D
1		2020년도	2021년도	증감률
2	강남구	4338.806	5754.758	32.63461
3	강동구	9721.915	11472.39	18.00547
4	강북구	6737.496	5605.892	-16.7956
5	강서구	6224.913	7371.681	18.42223
6	관악구	6633.006	4746.981	-28.4339
7	광진구	6894.986	6333.079	-8.14951
8	구로구	14915.4	12007.05	-19.4989
9	금천구	8540.739	5862.281	-31.361
10	노원구	8965.817	5783.138	-35.4979
11	도봉구	7499.33	5615.85	-25.1153
12	동대문구	5074.172	3758.06	-25.9375
13	동작구	7978.114	5363.202	-32.7761

커뮤니티 조회수 증감률.xlsx

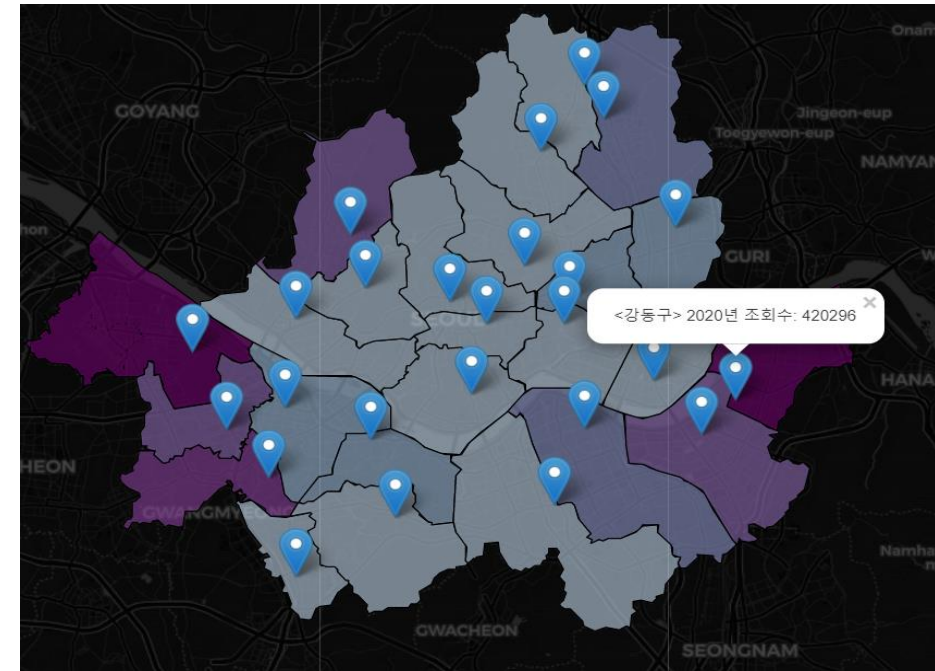
[2020년 조회수 합 시각화]

2020년 기준, 거래량이 높았던 강남3구 및 강서지역의 조회수가 대체적으로 높았음
실 거래량이 많은 강북지역은 상대적으로 조회수가 낮은 편에 속했음

2020년 주택거래량 시각화



2020년 조회수 합 시각화



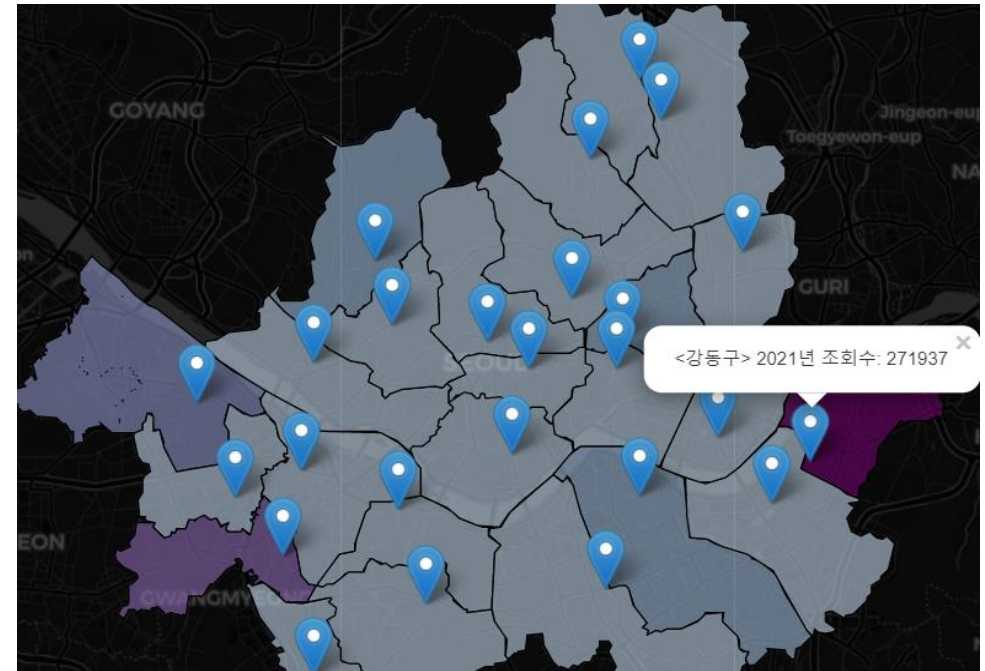
[2021년 조회수 합 시각화]

2021년 기준, 강남3구 지역, 강서, 강북 지역은 여전히 거래량이 높은 수준
반면 조회수 측면에서, 특정 3개구가 이상점 수준으로 높은 것을 확인할 수 있음

2021년 주택거래량 시각화



2021년 조회수 합 시각화

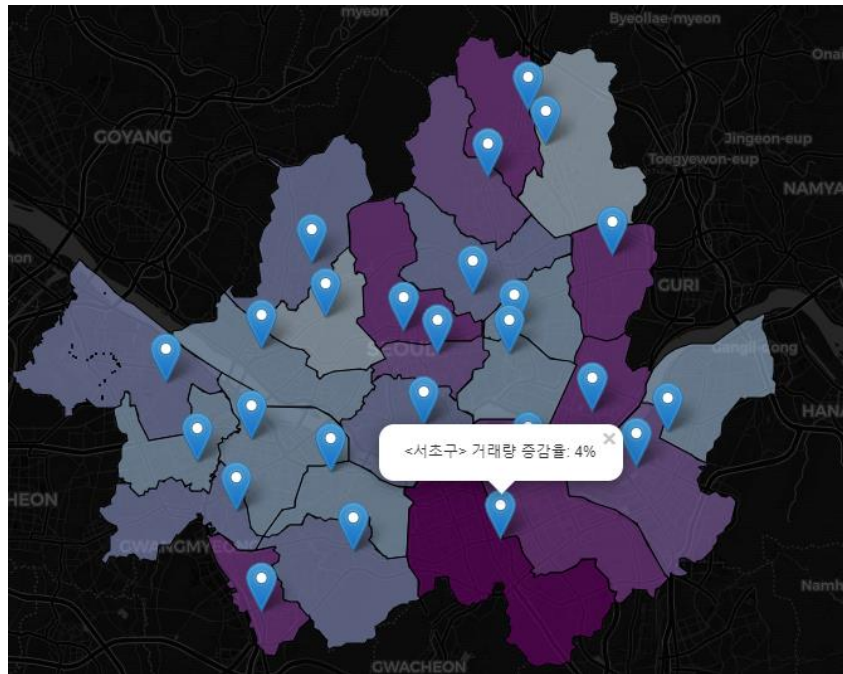


[2020 - 2021년 조회수 증감률 시각화]

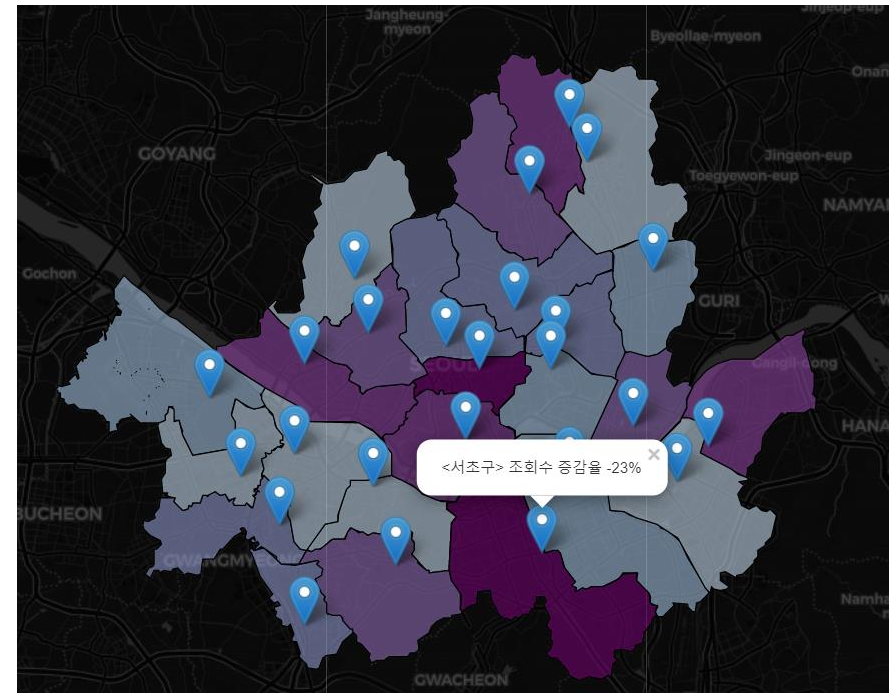
거래량과 마찬가지로, 2020년대비 2021년의 조회수는 전반적으로 떨어짐

지도에서 보이는 것처럼, 거래량 증감율과 조회수 증감율의 추이가 전반적으로 비슷한 것을 알 수 있음

거래량증감율시작화



조회수증감율시작화



통계분석

거래량데이터: 2개년간 월별, 년별 거래량

조회수 데이터: 2개년간 월별, 년별 조회수와 게시글 수

→

시간을 월별/년별로 쪼개고

조회수를 단순 합 조회수, 게시글 당 평균 조회수로 쪼개서 인사이트 도출

1. 2020 / 2021년 조회수 합 대비 거래량
2. 2020 / 2021년 평균조회수 대비 거래량
3. 2020 / 2021년 월별 조회수 합 대비 거래량

```
from scipy import stats
```

[2020 조회수 합 대비 거래량]

2020년의 조회수 총합대비거래량은, 상관계수 0.74로 강한상관을 보여줌.

관련코드

```
# 2020년의 조회수 합산와 거래량의 상관계수

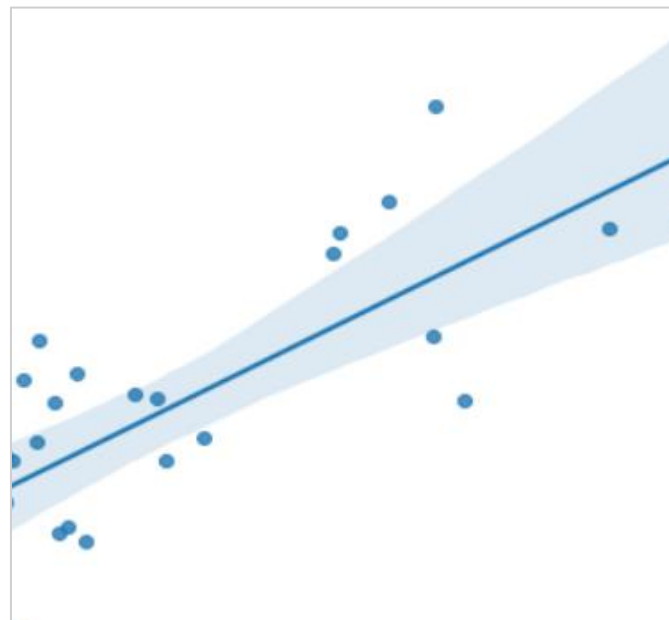
x1 = year_sum_data['2020년도'] # 조회수
x1.index.name = '구'
x2 = df2['2020년'].copy() # 거래량
x2.sort_index(inplace = True)

corr_frame = pd.DataFrame({'x': x1, 'y' : x2})

plt.figure(figsize=(10,10))
sns.lmplot(x = 'x', y = 'y', data = corr_frame, ci = 95)
plt.show()

stats.pearsonr(x1, x2)
```

시각화와상관계수(피어슨)



(0.7458767808274971, 1.8702291559923833e-05)

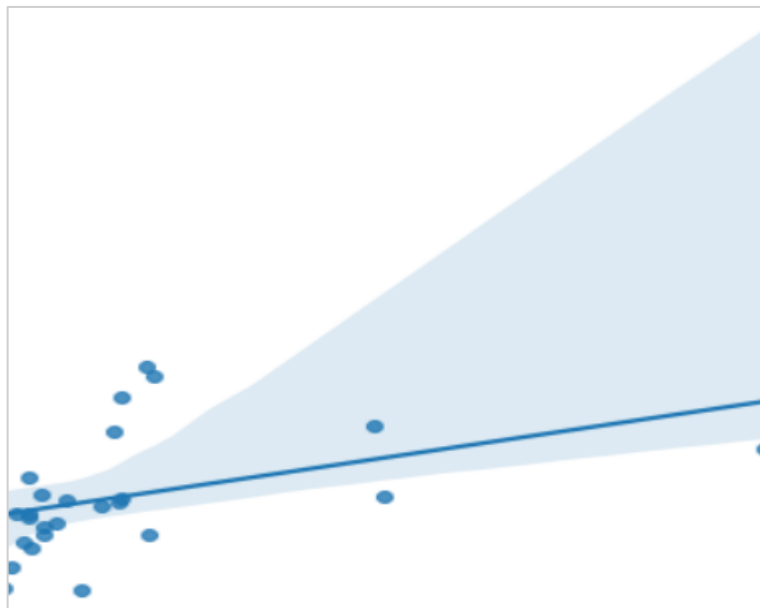
[2021 조화수 합 대비 거래량]

2021년의 데이터에서, 유독 3개의 구(강동, 구로, 강서)의 조화수가 높았음

1.5*IQR 방식으로 이상치를 제거하였을 때, 상관계수와 P-value가 크게 개선됨

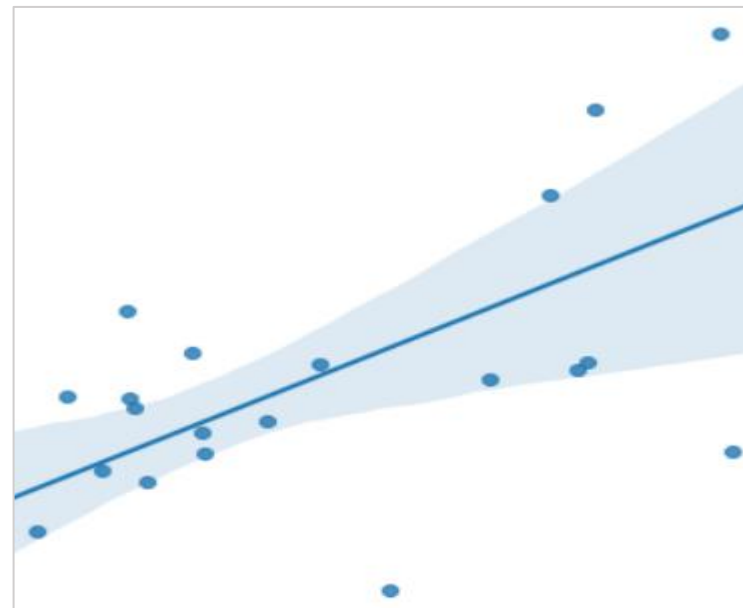
이상치 제거 전, 후 모두 양의 상관 관계를 가짐.

이상치 제거 전 시각화, 상관계수



(0.4116164120149343, 0.04091865423041306)

이상치 제거 후 시각화, 상관계수



(0.7458767808274971, 1.8702291559923833e-05)

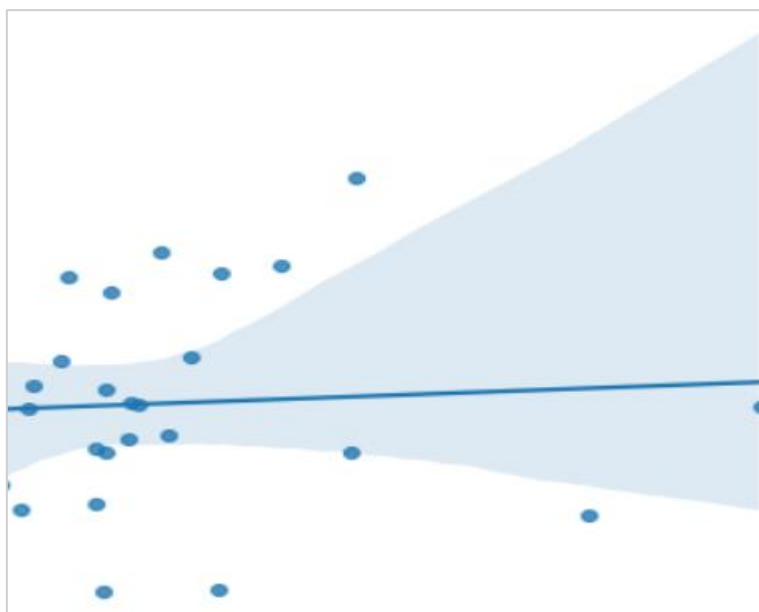
[2020 평균 조회수 대비 거래량]

평균 조회수는 게시글 당 조회수를 적절히 합하고 나눠 구하였음

이상치제거 전, 후 모두 5% 유의수준을 벗어나는 P-value를 가짐

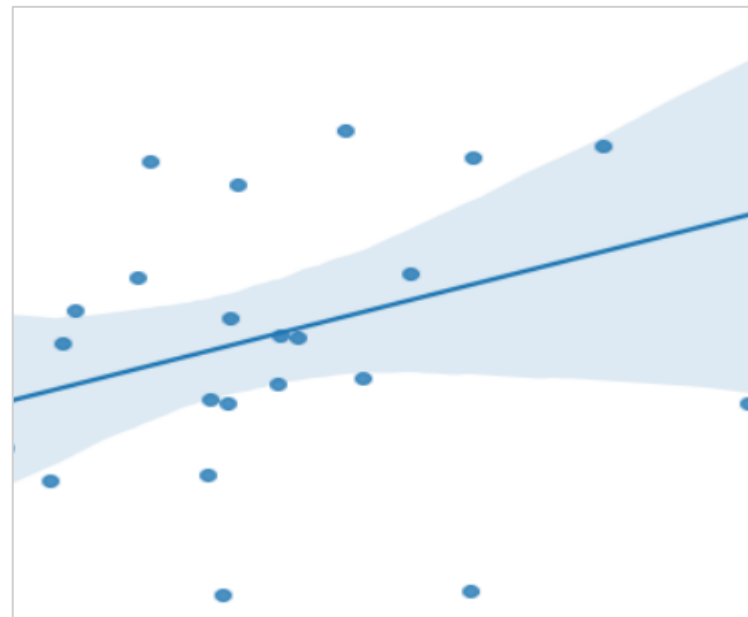
따라서, 명백히 양의 상관을 가지는 가치 판단의 영역

이상치제거전 시각화, 상관계수



(0.060450463199745216, 0.7740812608908041)

이상치제거후 시각화, 상관계수



(0.35494944484434415, 0.09651817508032985)

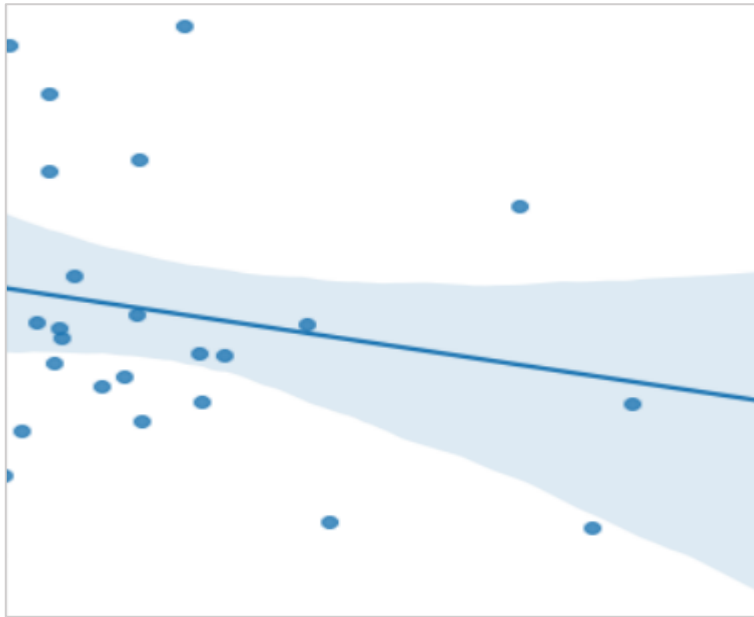
[2021 평균 조회수 대비 거래량]

평균 조회수는 게시글 당 조회수를 적절히 합하고 나눠 구하였음

이상치제거의 효과가 작았고, 앞의 결과와 다르게 오히려 음의 상관을 보여줌

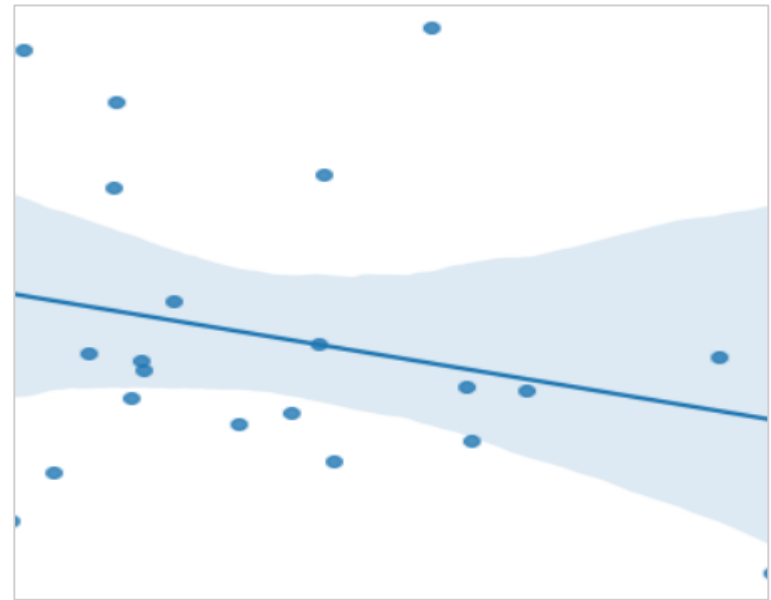
P-value가 꽤 높은 수준이라, 음의 상관을 일반화하기는 어려울 것으로 판단됨

이상치제거전시각화,상관계수



$(-0.23488508223052215, 0.2583916422760082)$

이상치제거후시각화,상관계수



$(-0.2424347838585333, 0.28966978189468834)$

[2020 - 2021 월별 조회수 합 대비 거래량]

월별 상관관계수가 일정하지 않음

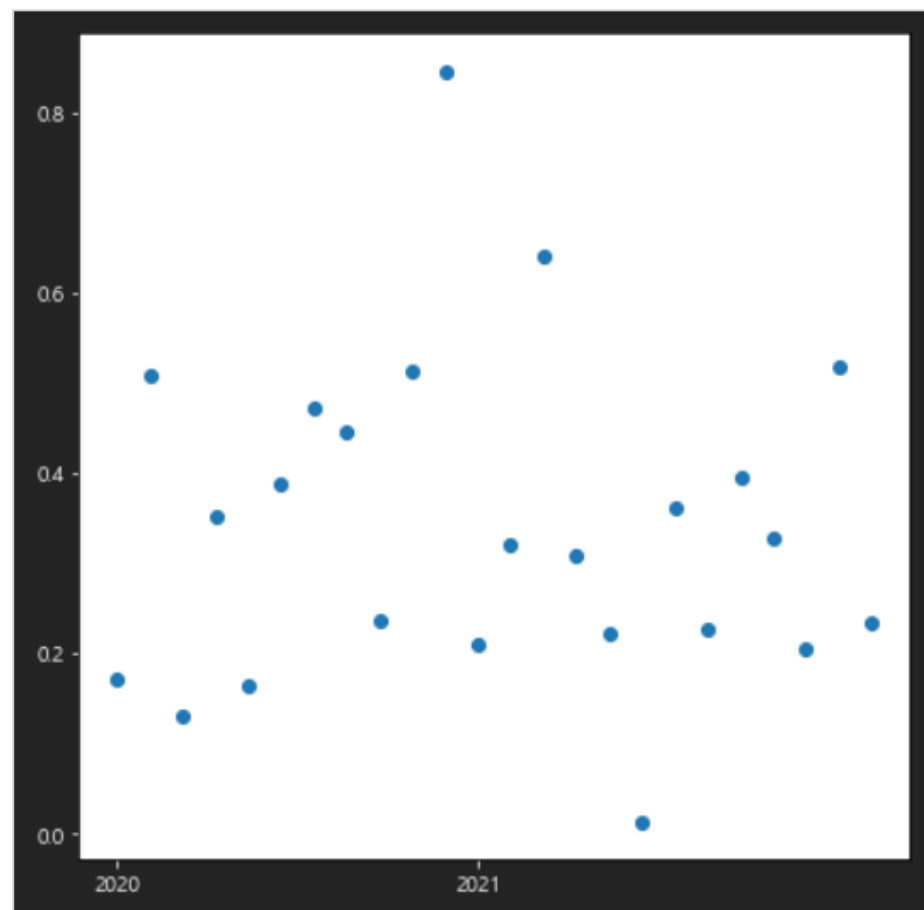
관련코드

```
month_corr_list = []
for month in range(1, 13, 1):
    x1 = month_sum_data[float('2020.{}'.format(str(month).zfill(2)))]
    x1.index.name = '구'
    x2 = df2['2020.{}'.format(str(month).zfill(2))].copy()
    x2.sort_index(inplace = True)
    month_corr_list.append(stats.pearsonr(x1, x2))

for month in range(1, 13, 1):
    x1 = month_sum_data[float('2021.{}'.format(str(month).zfill(2)))]
    x1.index.name = '구'
    x2 = df2['2021.{}'.format(str(month).zfill(2))].copy()
    x2.sort_index(inplace = True)
    month_corr_list.append(stats.pearsonr(x1, x2))

corr_list = []
p_list = []
for corr, p in month_corr_list:
    corr_list.append(corr)
    p_list.append(p)
```

시간에따른 월별상관관계수의흐름



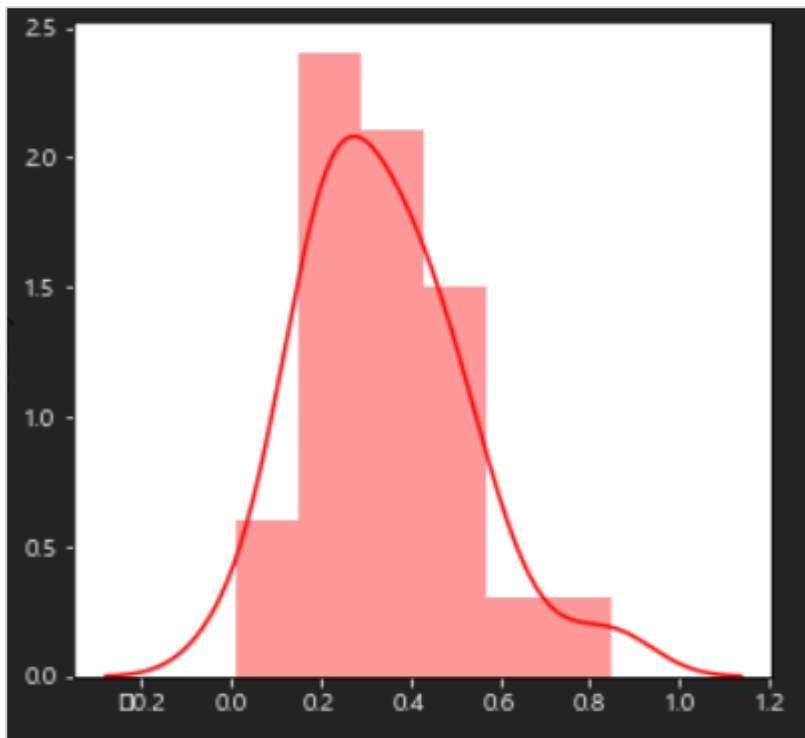
[2020 - 2021 월별 조회수 합 대비 거래량]

24개월 월별 상관관계 분포를 보면, 주로 0.2~0.6의양의상관관계를 보여줌

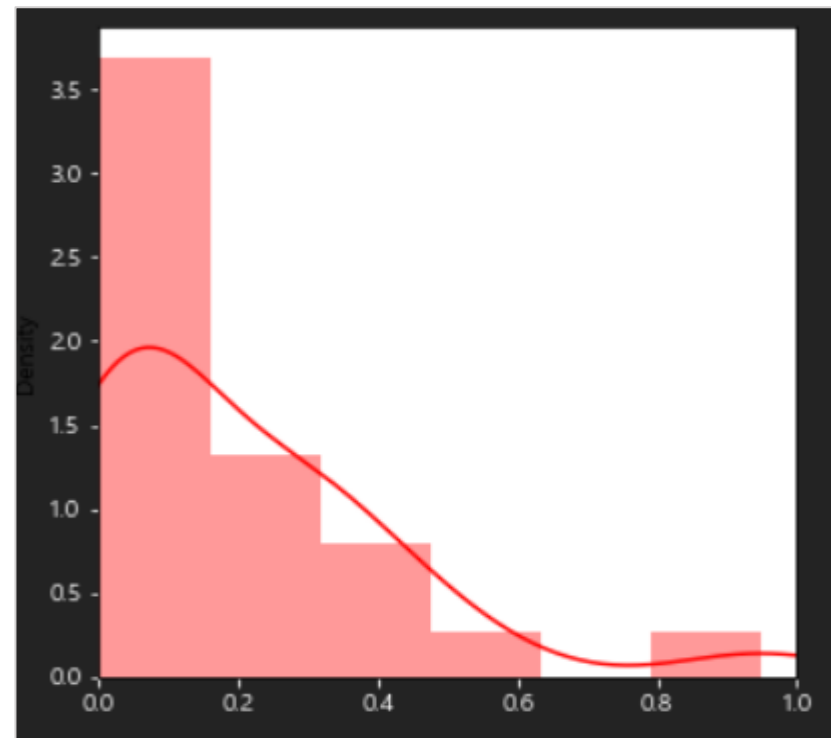
P-value 측면에서, 유의수준 5%보다 작은 값은 24개월 중 7개월 뿐

:결론적으로, 월별 조회수와 거래량은 주로 약한~중간 정도의 상관을 띠지만, 신뢰하기는 어렵다

24개월의 월별상관계수분포



24개월의 월별 P-value 분포



E.O.D