

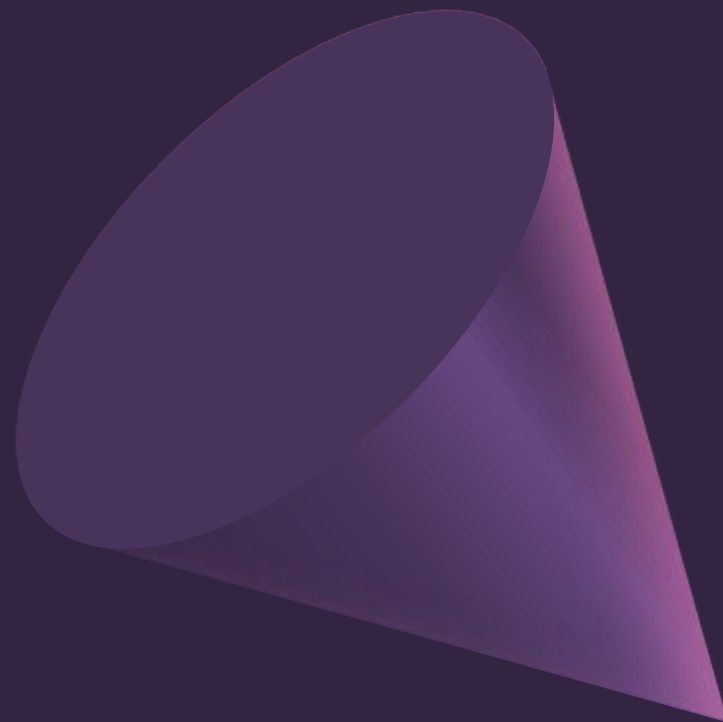
21/01/18

웹크롤링을

이용하여  
해외지수-국내증시  
시각화 자료  
만들고,  
예측하기

김해민

배경설명



2020년에 진행했던  
프로젝트로써,  
2022년에 재진행하고,  
업그레이드하고자 함

## 최근조회종목

MYSTOCK

알테오젠

221,200 ▼3.41%

진에어

11,050 ▼5.96%

기업은행

8,470 ▼4.19%

한진칼

92,300 ▼6.77%

넥슨지터

8,840 ▼5.05%

## 주요뉴스

더보기

[외환브리핑] 코로나19 재확산 우려 부각...원·달러...  
 [원자재시황] 국제유가, 코로나19 재확산 우려에 ...  
 [유럽증시] 美 연준 부정적 경기 회복 전망에 줄줄...  
 [굿모닝 증시] 단기간에 오른 증시...美 대선 후보...  
 뉴욕 증시, 3월 바닥을 다시 테스트할까 [김현석...  
 '오늘만큼 올랐나'...美다우지수 폭락, 국내증시는...

## TOP종목

더보기

거래상위	상한가	하한가	시가총액 상위
KODEX 200선물인버스2X	5,515	▲265	+5.05%
KODEX 레버리지	12,030	▼690	-5.42%
KODEX 인버스	6,245	▲175	+2.88%
현대약품	7,510	▲300	+4.16%
제넨바이오	2,480	▲75	+3.12%
KODEX 코스닥150선물...	6,155	▲145	+2.41%

## 수익률 상위펀드 3개월수익률 순 | 2020.06.12기준

더보기

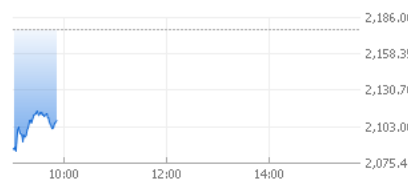
국내펀드	해외펀드	브릭스	삼성그룹주
------	------	-----	-------

- 미래에셋TIGER의로기증권상장지수투자신탁(주식)
- KBKBSTAR코스닥150선물레버리지증권상장지수투자신탁(주..
- 키움KOSF코스닥150선물레버리지증권상장지수투자신탁[..

## 오늘의증시

실시간 2020.06.12 09:51 장중

코스피 2,106.57 ▼70.21 -3.23%



개인 +1,185 외국인 -1,338 기관 +385 (역월)  
 ↑ 6 ▲ 54 - 7 ▼ 838 ↓ 0

코스닥 732.35 ▼24.71 -3.26%

코스피 200 279.12 ▼9.50 -3.29%

## 업종상위

더보기

- 건강관리업체맞.. 서울리거 독십자랩셀  
+0.89% +9.51% +1.69%
- 제약 일양약품우 부광약품  
+0.74% +18.15% +8.56%
- 생명과학도구맞.. 랩지노믹스 휴마시스  
-0.30% +4.48% +3.69%

## 테마상위

더보기

- 코로나19(진단/.. 수전텍 부광약품  
-0.39% +8.88% +8.56%
- 메르스 코로나.. 오공 경남바이오..  
-0.67% +13.06% +10.45%

네이버 해피빈 공감가게

청소년들을  
도와주는  
디자인 브랜드



## 해외증시

더보기

다우산업(06.11)	25,128.17 ▼1,861.82
나스닥(06.11)	9,492.73 ▼527.62
홍콩H(06.11)	9,944.60 ▼198.88
상해종합(06.11)	2,920.90 ▼22.86
니케이225(06.12)	22,031.38 ▼441.53

## 인기 검색종목

더보기

- 삼성전자 52,200 ▼2,100
- 셀트리온 296,000 ▲8,000
- 삼성바이오로.. 772,000 ▲25,000
- 이지바이오 49,150 ▲5,900
- 부광약품 40,600 ▲3,200

이토록 부담없는 혜택  
24개월 이자율 0%  
렉서스 어셀 케이 프로모션



해외 주요지수

현지시간 기준

국가명	지수명	현재가	전일대비	등락률	등락률 그래프	시간
미국	다우 산업	25,128.17	▼ 1,861.82	-6.90%	<div></div>	06.11 17:13
미국	다우 운송	8,839.36	▼ 739.71	-7.72%	<div></div>	06.11 17:13
미국	나스닥 종합	9,492.73	▼ 527.62	-5.27%	<div></div>	06.11 16:01
미국	나스닥 100	9,588.48	▼ 505.78	-5.01%	<div></div>	06.11 16:01
미국	S&P 500	3,002.10	▼ 188.04	-5.89%	<div></div>	06.11 17:13
미국	필라델피아 반도체	1,876.66	▼ 128.00	-6.39%	<div></div>	06.11 16:01
브라질	브라질 BOVESPA	94,685.98	▼ 2,060.57	-2.13%	<div></div>	06.10
중국	상해종합	2,920.90	▼ 22.856	-0.78%	<div></div>	06.11 15:01
중국	상해 A	3,061.65	▼ 23.96	-0.78%	<div></div>	06.11 15:01
중국	상해 B	207.22	▼ 1.513	-0.72%	<div></div>	06.11 15:00
일본	니케이225	22,012.38	▼ 460.53	-2.05%	<div></div>	06.12 09:39
홍콩	항셍	24,480.15	▼ 569.58	-2.27%	<div></div>	06.11 16:08
홍콩	항셍 차이나기업(H)	9,944.60	▼ 198.88	-1.96%	<div></div>	06.11 16:08
홍콩	항셍 차이나대기업(B)	3,836.09	▼ 93.81	-2.39%	<div></div>	06.11 16:08
대만	대만 가권	11,535.77	▼ 184.39	-1.57%	<div></div>	06.11 13:33
인도	인도 SENSEX	33,538.37	▼ 708.68	-2.07%	<div></div>	06.11 16:10
말레이시아	말레이시아 KLCI	1,557.25	▼ 18.02	-1.14%	<div></div>	06.11 16:50
인도네시아	인도네시아 IDX종합	4,854.75	▼ 65.93	-1.34%	<div></div>	06.11
영국	영국 FTSE 100	6,076.70	▼ 252.43	-3.99%	<div></div>	06.11 16:35
프랑스	프랑스 CAC 40	4,815.60	▼ 237.82	-4.71%	<div></div>	06.11 17:35
독일	독일 DAX30	11,970.29	▼ 559.87	-4.47%	<div></div>	06.11 17:37
유럽	유로스톡스 50	3,144.57	▼ 149.14	-4.53%	<div></div>	06.11 17:50
러시아	러시아 RTS	1,238.11	▼ 39.47	-3.09%	<div></div>	06.11 18:39
이탈리아	이탈리아 FTSE MIB	18,806.86	▼ 951.15	-4.81%	<div></div>	06.11 17:22

넥슨지티 8,860 ▼ 450

하이트진로 38,850 ▼ 300

삼성전자 51,900 ▼ 2,400

지니뮤직 3,055 ▼ 245

삼성 레버... 345 ▼ 55

메가엠디 4,060 ▲ 180

해외 인기검색

- 다우 산업 -6.90%
- 나스닥 종합 -5.27%
- 니케이225 -2.05%
- 독일 DAX30 -4.47%
- 상해종합 -0.78%

투자 가이드

- 해외증시 거래시간
- 해외 휴장일

3,002.10

전일대비 ▼188.04 (-5.89%)

전일 3,190.14

고가 3,123.53

52주 최고 3,393.52

시가 3,123.53

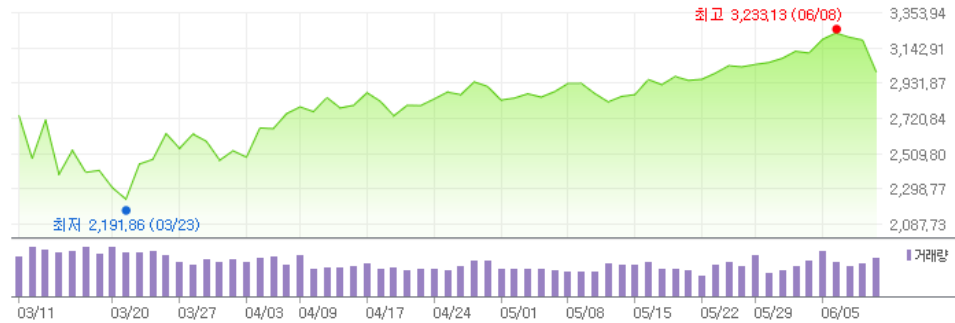
저가 2,999.49

52주 최저 2,191.86



선차트 3개월 1년 3년 5년 10년

봉차트 일봉 주봉 월봉



#### 관련정보

거래소 Standard & Poor's Corp Indices  
기준통화 USD  
거래시간 23:30~06:00  
(한국시간 기준) 서머타임 22:30~05:00

#### ! 지수 설명

미국의 국제신용평가기관인 S&P에서 작성하는 주가 지수로, 뉴욕증권거래소에 상장된 500개의 우량종목으로 이루어진 지수

네이버 지식백과

#### 일별 시세

일자	종가	전일대비	시가	고가	저가
2020.06.11	3,002.10	▼188.04	3,123.53	3,123.53	2,999.49
2020.06.10	3,190.14	▼17.04	3,213.42	3,223.27	3,181.49
2020.06.09	3,207.18	▼25.21	3,213.32	3,222.71	3,193.11
2020.06.08	3,232.39	▲38.46	3,199.92	3,233.13	3,196.00
2020.06.05	3,193.93	▲81.58	3,163.84	3,211.72	3,163.84
2020.06.04	3,112.35	▼10.52	3,111.56	3,128.91	3,090.41
2020.06.03	3,122.87	▲42.05	3,098.90	3,130.94	3,098.90
2020.06.02	3,080.82	▲25.09	3,064.78	3,081.07	3,051.64
2020.06.01	3,055.73	▲11.42	3,038.78	3,062.18	3,031.54
2020.05.29	3,044.31	▲14.58	3,025.17	3,049.17	2,998.61

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 다음 >

#### 최근조회

#### MY STOCK

알테오젠	219,700	▼9,300
진에어	11,000	▼750
기업은행	8,360	▼480
한진칼	91,000	▼8,000
넥슨지티	8,850	▼460
하이트진로	38,850	▼300
삼성전자	51,900	▼2,400
지니뮤직	3,055	▼245
삼성 레버...	350	▼50
메가엠디	4,060	▲180

◀ ▶

## 과제 전 목표



# 기술설명

## -웹크롤링-





금융 홈

국내증시

해외증시

시장지표

펀드

리서치

뉴스

MY

## 국내증시

## 주요시세정보

코스피 | 코스닥 | 선물

코스피200 | 코넥스

시가총액 | 배당

업종 | 테마 | 그룹사

ETF | ETN

상승 | 보합 | 하락

상한가 | 하한가

급등 | 급락

거래상위 | 급증 | 급감

투자자별매매동향

외국인매매 | 기관매매

프로그램매매동향

증시자금동향

신규상장

외국인보유

장외시세

IPO

## 투자자보호

관리종목

거래정지종목

시장정보종목

## 조건검색

골든크로스 | 겹상승

이격도과열 | 투심과열

상대강도과열

## 기업 전자공시

## 공매도 거래 현황

금융홈 &gt; 국내증시 &gt; 주요시세정보 &gt; 코스피200

## 코스피200

실시간 2020.06.17 12:18 장중



코스피200	280.62	상한종목수	0
전일대비	▼ 1.97	상승종목수	44
등락률	-0.70%	하한종목수	0
장중최고	284.90	하락종목수	150
장중최저	278.33	보합종목수	6
52주최고	307.54	거래량(천주)	231,557
52주최저	196.27	거래대금(백만)	7,749,528

## 시간별 시세

체결시각	체결가	전일비	변동량(천주)	거래량(천주)	거래대금(백만)
12:15	280.78	▼ 1.81	16	229,302	7,696,858
12:14	280.63	▼ 1.96	1,088	229,286	7,696,270
12:13	280.73	▼ 1.86	892	228,198	7,679,301
12:12	280.72	▼ 1.87	1,450	227,306	7,663,205
12:11	280.96	▼ 1.63	2,125	225,856	7,640,981
12:10	280.61	▼ 1.98	374	223,731	7,607,530

## 일별 시세

날짜	체결가	전일비	등락률	거래량(천주)	거래대금(백만)
2020.06.17	280.78	▼ 1.81	-0.64%	229,302	7,696,858
2020.06.16	282.59	▲ 14.64	+5.46%	247,894	9,831,669
2020.06.15	267.95	▼ 13.83	-4.91%	262,444	12,965,653
2020.06.12	281.78	▼ 6.84	-2.37%	232,833	10,950,830
2020.06.11	288.62	▼ 3.28	-1.12%	275,904	11,835,398
2020.06.10	291.90	▲ 0.58	+0.20%	229,067	8,568,457

## 최근조회

## MY STOCK

삼성중공우	744,000	▲
SK증권우	6,960	▲
삼성물산	120,500	▼
KODEX 200선	5,380	▲
씨젠	108,100	▼
알테오젠	227,000	▼
진에어	10,850	▼
기업은행	8,710	▼
한진칼	86,300	▼
넥슨지티	9,380	▼

## 인기검색어

더보기

1 빅텍	8,800	▲
2 삼성중공우	744,000	▲
3 두산인프라코어	8,400	▲
4 삼성전자	51,900	▼
5 셀트리온	287,000	▼

## 가장 많이 본 뉴스

더보기

- \* 무상증자 공시에 개미를 우르...
- \* [뉴욕마감] 美소비 잠깐 부활...
- [외환브리핑] 무력도발 나선 ...
- [김현석의 월스트리트나우] ...
- [유럽종시] 단돈 5파운드 짜리 ...

웹크롤링

국내증시  
koshi200

## 웹크롤링

### 국내증시 kospi200

#### 일별 시세

날짜	체결가	전일비	등락률	거래량(천주)	거래대금(백만)
2020.06.17	280.78	▼ 1.81	-0.64%	229,302	7,696,858
2020.06.16	282.59	▲ 14.64	+5.46%	247,894	9,831,669
2020.06.15	267.95	▼ 13.83	-4.91%	262,444	12,965,653
2020.06.12	281.78	▼ 6.84	-2.37%	232,833	10,950,830
2020.06.11	288.62	▼ 3.28	-1.12%	275,904	11,835,398
2020.06.10	291.90	▲ 0.58	+0.20%	229,067	8,568,457

1 | 2 | 3 | 4 | 5

#국내 코스피 지수 You, 4 hours ago • 1

def domesitc(index\_cd, last\_page): #bs4 이용해서 가져옴

historical\_prices = dict()

for page\_n in range(1, last\_page+1):

naver\_index = f"https://finance.naver.com/sise/sise\_index\_day.nhn?code={index\_cd}&page="

# print(naver\_index)

source = urlopen(naver\_index).read()

source = bs4.BeautifulSoup(source, "html.parser")

dates = source.find\_all('td', class\_='date') # <td class="date"> 태그에서 날짜 수집

prices = source.find\_all('td', class\_='number\_1') # <td class="number\_1"> 태그에서 price

for n in range(len(dates)):

## 코스피200



코스피200	282.04	상한종목수	0
전일대비	▼ 0.55	상승종목수	51
등락률	-0.19%	하한종목수	0
장중최고	284.90	하락종목수	144
장중최저	278.33	보합종목수	5

뒤로(B) Alt+왼쪽 화살표  
앞으로(F) Alt+오른쪽 화살표  
새로고침(R) Ctrl+R

스크랩북에 추가  
북마크에 추가  
다른 이름으로 저장(A)... Ctrl+S

☐ 스페이스 열기  
☐ 모바일창에서 열기  
☐ 사이드바에서 열기

플러그인 호환 모드  
인쇄(P) Ctrl+P  
전송(C)...

페이지 소스 보기(O)

프레임 소스 보기(V)

프레임 새로고침(F)

검사(N)

Ctrl+U

Ctrl+Shift+J

### 시간별 시세

체결시간	체결가
11:41	279.85
11:40	279.70
11:39	279.44
11:38	279.37
11:37	279.54
11:36	279.37

### 일별 시세

날짜	체결가
2020.05.14	253.65
2020.05.13	255.85
2020.05.12	253.37
2020.05.11	254.95
2020.05.08	256.62
2020.05.07	254.46

« 맨앞 1 2 3 4 5 6 7 8 9 10 다음 » 맨뒤 »

# 웹크롤링

해외증시  
S&P500-미국  
NIKKEI255-일본  
Shanghai-중국

삼성증권우	744,000	↑
SK증권우	6,960	↑
삼성물산	121,500	↓
KODEX 200선	5,345	▲
씨젠	108,100	↓
알테오젠	226,900	↓
진에어	10,850	↓
기업은행	8,710	↓
한진칼	86,600	↓
넥슨지티	9,420	↓

### 인기검색어

빅텍	8,780	▲
삼성증권우	744,000	↑
두산인프라코어	8,460	▲
삼성전자	52,000	↓
셀트리온	289,500	↓

### 가장 많이 본 뉴스

- \* 무상증자 공시에 재미들 우르...
- \* [뉴욕마감] 美소비 깜짝 부활...
- \* [외환브리핑] 무력도발 나선 ...
- \* [김현석의 월스트리트나우] '...
- \* [유럽증시] 단돈 5파운드자리 ...

# S&P 500 SPI@SPX

2020.06.16 17:19 현지시간 기준 | 10분 지연제공

**3,124.74**

전일대비 ▲58.15(+1.90%)

전일 3,066.59

고가 3,153.45

52주 최고 3,393.52

시가 3,131.00

저가 3,076.06

52주 최저 2,191.86



선차트 3개월 1년 3년 5년 10년

봉차트 일봉 주봉 월봉



- 뒤로(B) Alt+왼쪽 화살표
- 앞으로(F) Alt+오른쪽 화살표
- 새로고침(R) Ctrl+R
- 스크랩북에 추가
- 북마크에 추가
- 다른 이름으로 저장(A)...
- Ctrl+S
- 스페이스 열기
- 모바일창에서 열기
- Ctrl+S



## 관련정보

거래소 Standard & Poor's Corp  
Indices  
기준통화 USD  
거래시간 23:30~06:00  
(한국시간 기준) 서머타임 22:30~05:00

## 자주 묻는 질문

미국의 국제 신용평가기관인 S&P에서 작성하는 주가 지수로, 뉴욕증권거래소에 상장된 500개의 우량종목으로 이루어진 지수

네이버 지식백과

## 일별 시세

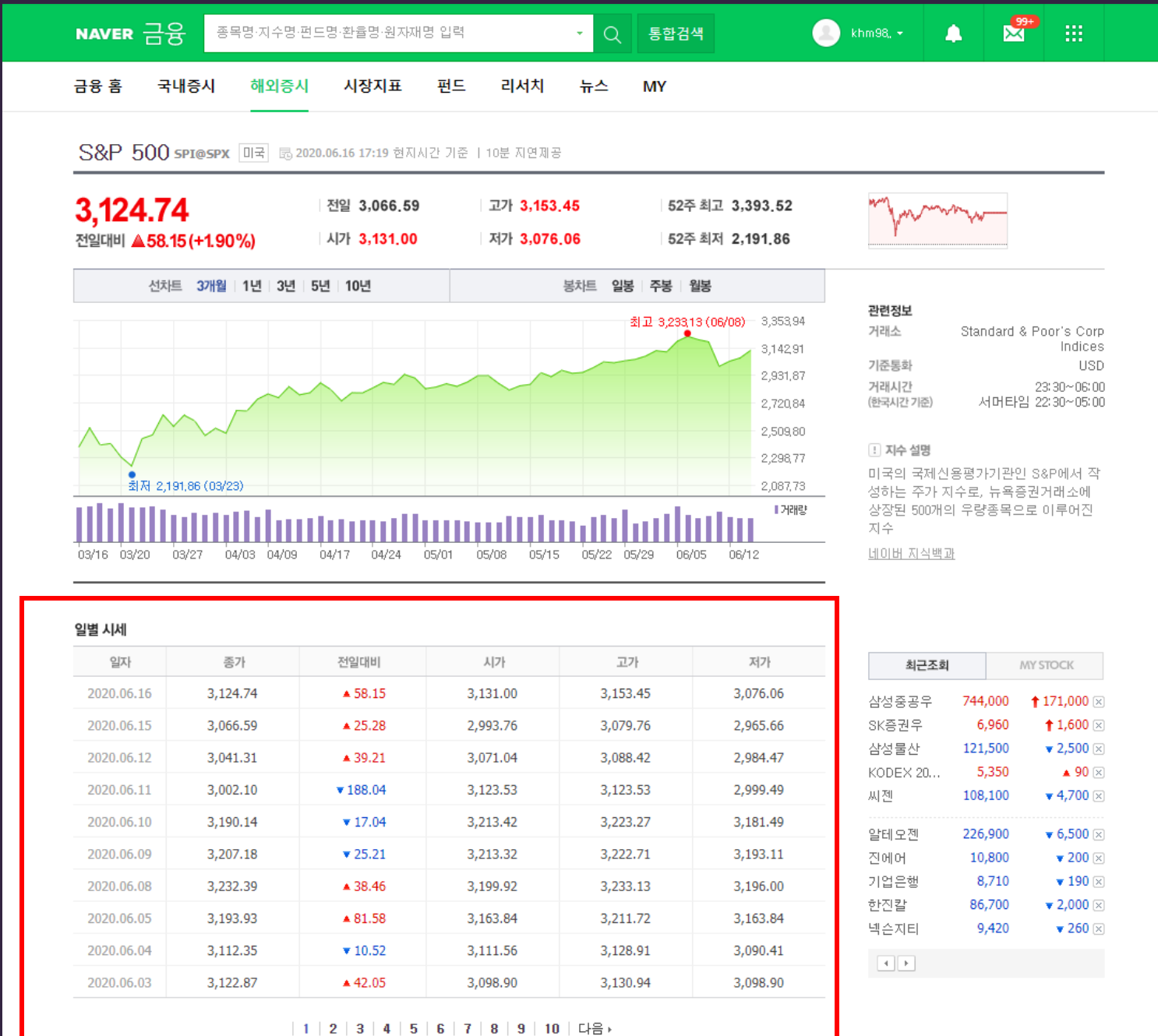
일자	저가	고가	시가	종가	전일
2020.06.16	3,076.06	3,153.45	3,131.00	3,124.74	▲58.15(+1.90%)
2020.06.15	2,965.66	3,076.06	3,066.59	3,066.59	
2020.06.12	2,984.47	3,076.06	3,066.59	3,066.59	
2020.06.11	2,999.49	3,076.06	3,066.59	3,066.59	
2020.06.10	3,181.49	3,076.06	3,066.59	3,066.59	
2020.06.09	3,207.18	3,222.71	3,213.32	3,222.71	▼25.21
2020.06.08	3,232.39	3,233.13	3,199.92	3,233.13	▲38.46
2020.06.05	3,193.93	3,163.84	3,163.84	3,163.84	▲81.58
2020.06.04	3,112.35	3,128.91	3,111.56	3,128.91	▼10.52
2020.06.03	3,122.87	3,130.94	3,098.90	3,130.94	▲42.05

1 2 3 4 5 6 7 8 9 10 다음

# 웹크롤링

해외증시  
S&P500-미국  
NIKKEI255-일본  
Shanghai-중국

최근조회	MY STOCK
삼성증권우	744,000 ↑171,000
SK증권우	6,960 ↑1,600
삼성물산	121,500 ▼2,500
KODEX 20...	5,325 ▲65
씨젠	107,300 ▼5,500
알테오젠	233,900 ▲500
진에어	10,850 ▼150
기업은행	8,720 ▼180
한진칼	86,300 ▼2,400
넥슨지티	9,400 ▼280



# 웹크롤링

해외증시  
S&P500-미국  
NIKKEI255-일본  
Shanghai-중국

## 일별 시세

일자	증가	전일대비	시가	고가	저가
2020.04.20	2,823.16	▼ 51.40	2,845.62	2,868.98	2,820.43
2020.04.17	2,874.56	▲ 75.01	2,842.43	2,879.22	2,830.88
2020.04.16	2,799.55	▲ 16.19	2,799.34	2,806.51	2,764.32
2020.04.15	2,783.36	▼ 62.70	2,795.64	2,801.88	2,761.54
2020.04.14	2,846.06	▲ 84.43	2,805.10	2,851.85	2,805.10
2020.04.13	2,761.63	▼ 28.19	2,782.46	2,782.46	2,721.17
2020.04.09	2,789.82	▲ 39.84	2,776.99	2,818.57	2,762.36
2020.04.08	2,749.98	▲ 90.57	2,685.00	2,760.75	2,663.30
2020.04.07	2,659.41	▼ 4.27	2,738.65	2,756.89	2,657.67
2020.04.06	2,663.68	▲ 175.03	2,578.28	2,676.85	2,574.57

| 1 | 2 | 3 |

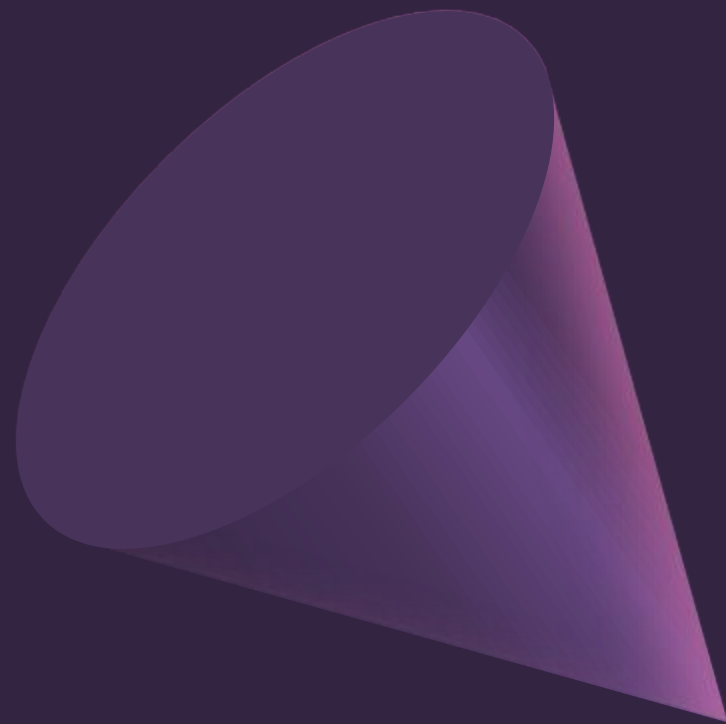
## 웹크롤링

해외증시  
S&P500-미국  
NIKKEI255-일본  
Shanghai-중국

```
def foreign(symbol, last_page):
    historical_prices = dict()
    for page_n in range(1, last_page + 1):
        url = f"https://finance.naver.com/world/worldDayListJson.nhn?symbol={symbol}&fdtc=0"
        # print(url)
        resp = requests.get(url, params={"page": page_n}, headers={'User-agent': 'Mozilla/5.0'})
        soup = resp.json()
        # print(soup)
        # print(type(soup))
        for n in range(len(soup)):
            date = pd.to_datetime(soup[n]["xymd"]).date()
            price = float(soup[n]["clos"])
```

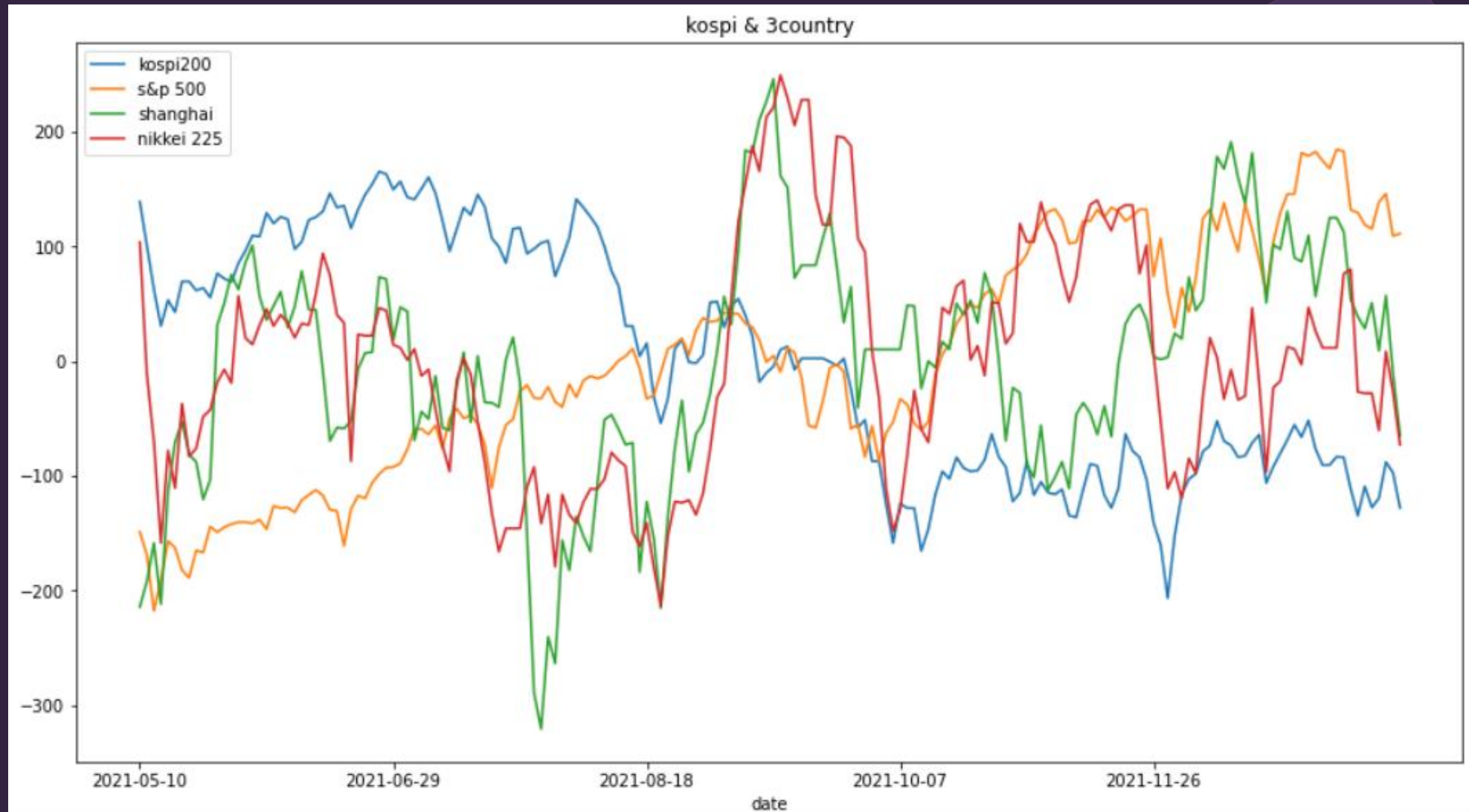
# 기술설명

## -시각화-





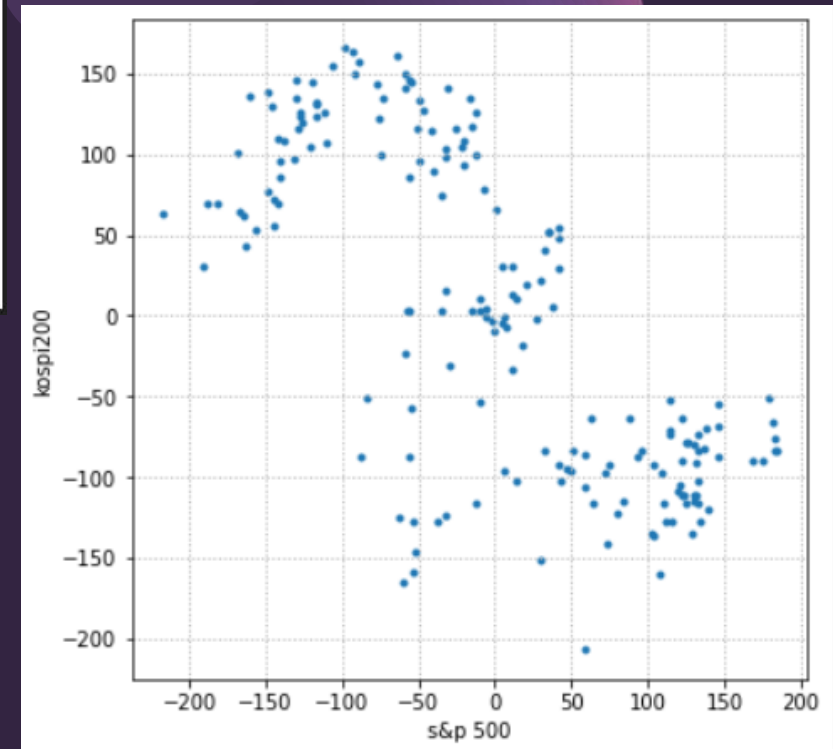
## 시각화-4개국



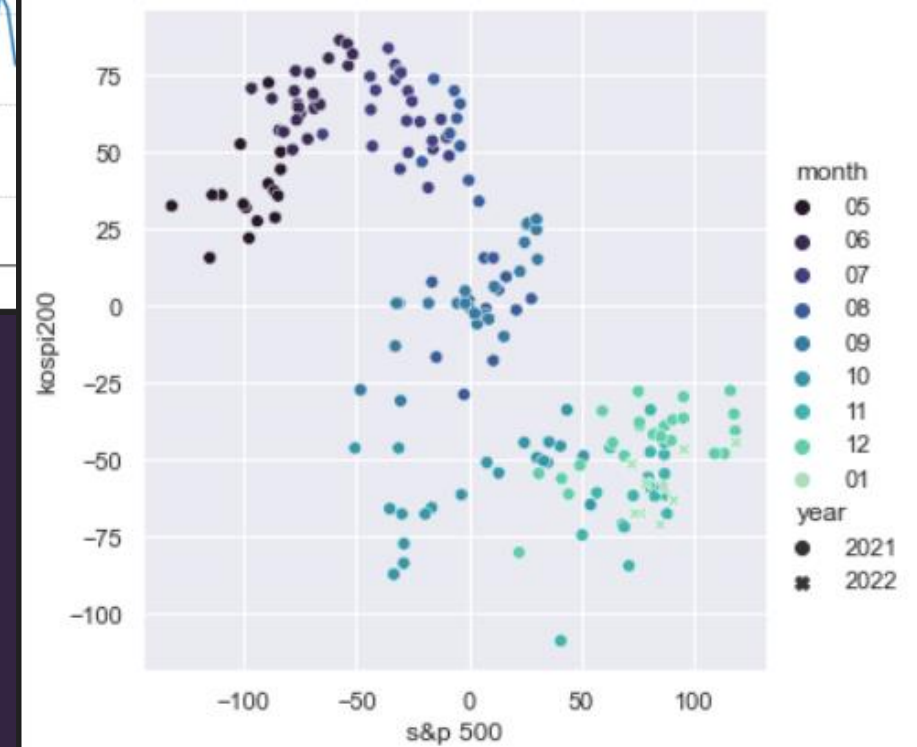
```
price_df.loc[:,["snp500","nikkei225","shanghai"]]=scaler.fit_transform(price_df.loc[:,["snp500","nikkei225","shanghai"]])  
price_df.loc[:,["kospi200"]]=scaler.fit_transform(price_df.loc[:,["kospi200"]])*100
```



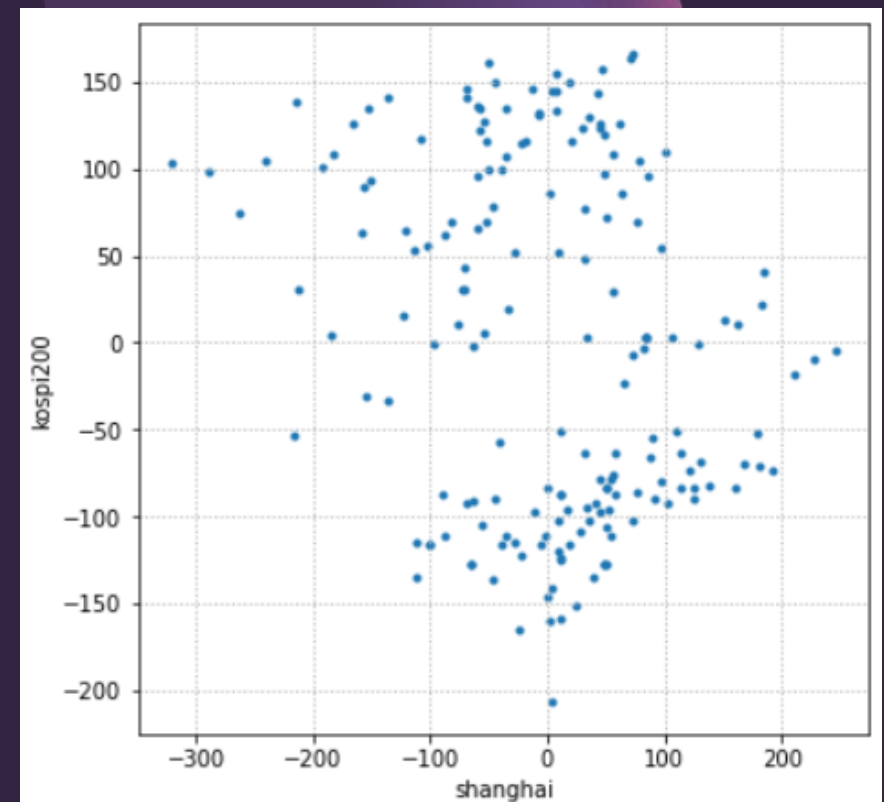
# 시각화-한/미



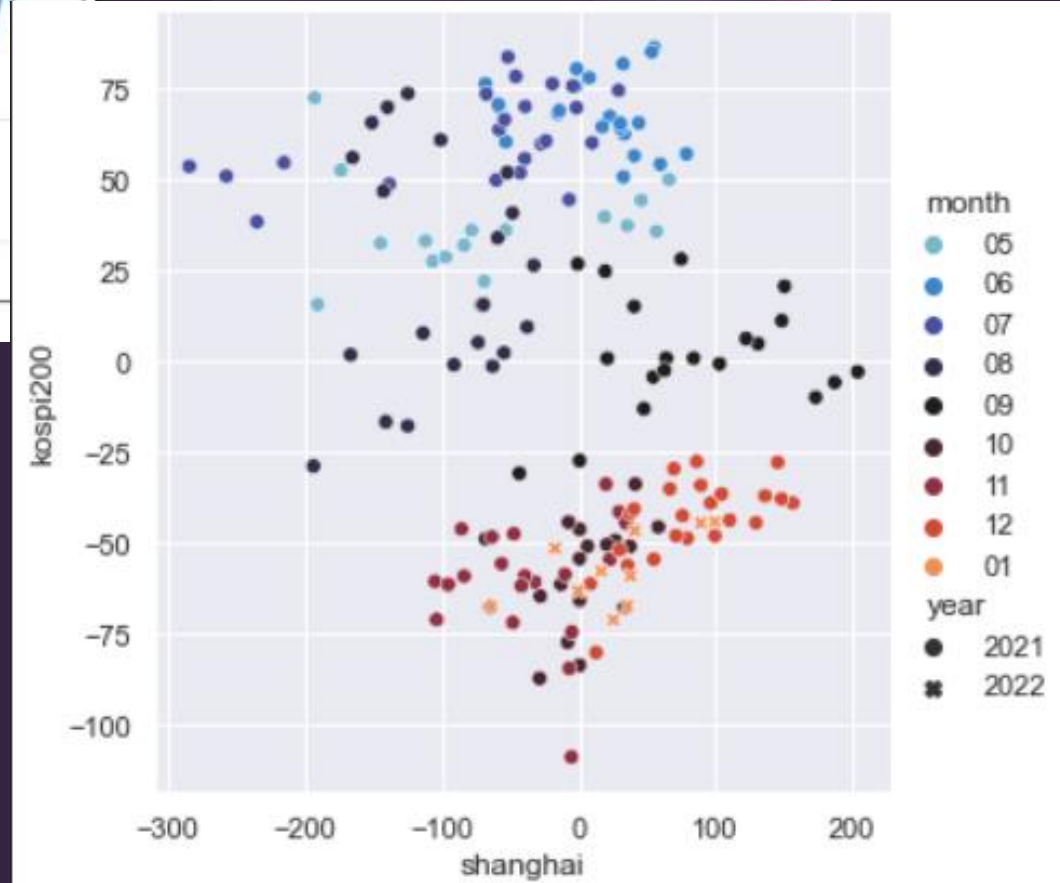
# 시각화-한/미



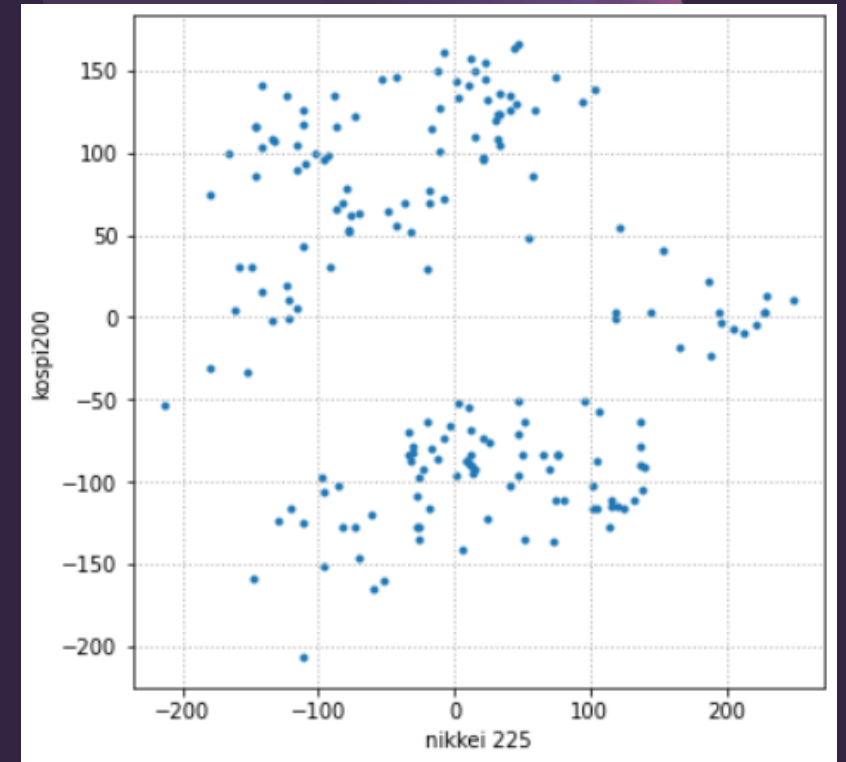
# 시각화-한/중



# 시각화-한/중

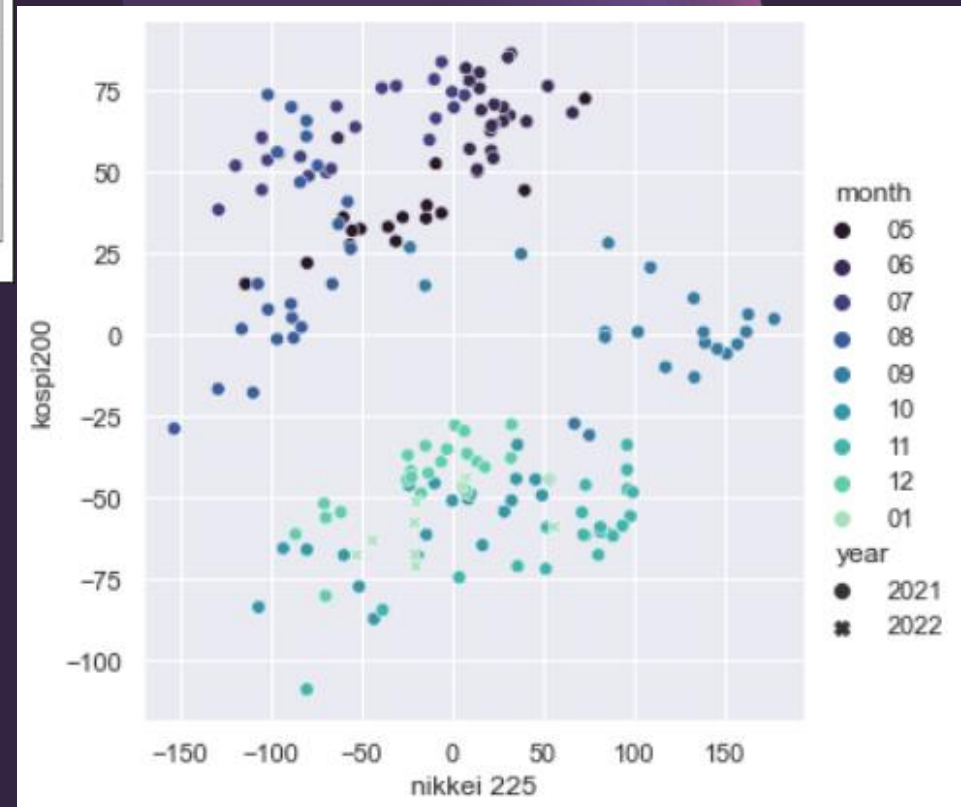


# 시각화-한/일





# 시각화-한/일



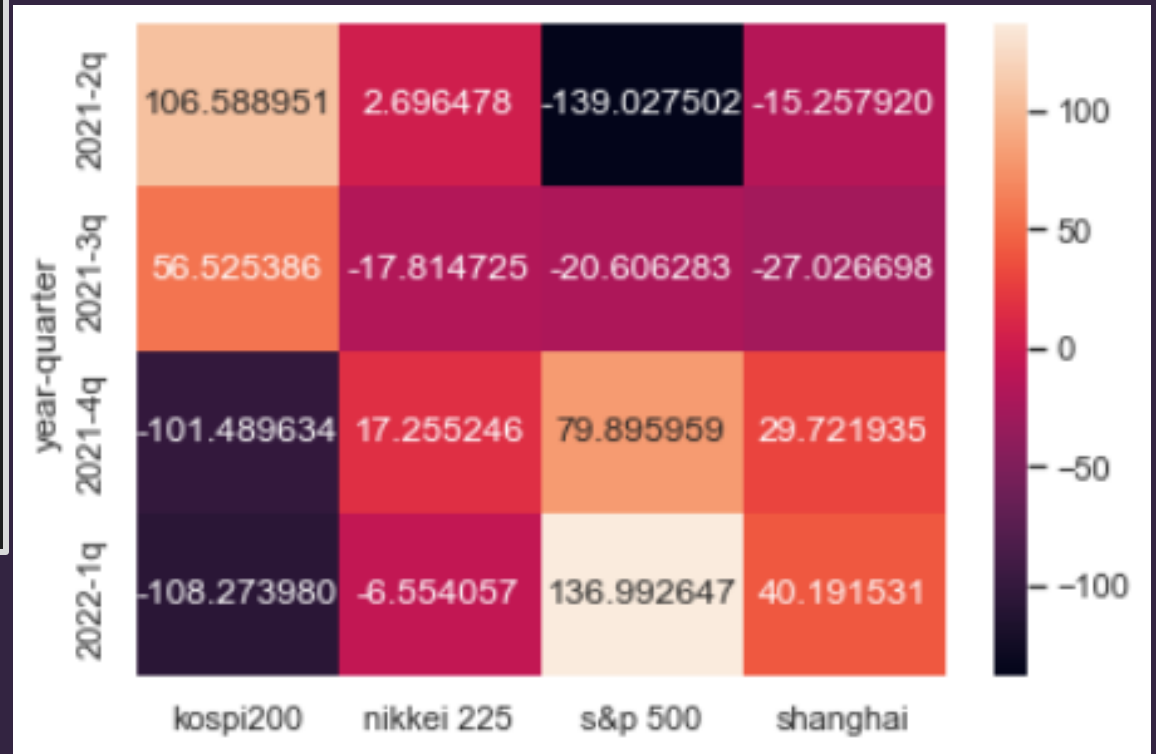
# 시각화 -피벗/히트맵

```
pivot = pd.pivot_table(price_df, index=["year", "quarter"])

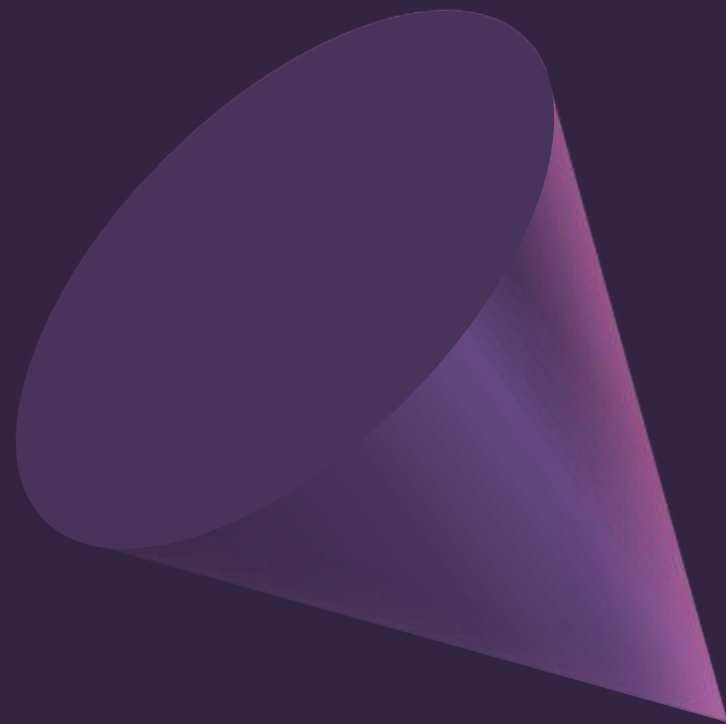
print(pivot)

sns.heatmap(pivot, annot=True, fmt='f')
plt.show()
```

```
(615, 5)
year quarter      kospi200      nikkei225      shanghai      snp500
2017 3q      85.240383 -150.634827  136.308816 -162.295111
      4q      157.693306  41.440931  146.278758  -97.523180
2018 1q      126.625158  47.614274  141.306201  -35.814986
      2q       94.231147  43.831563   36.178583  -51.272024
      3q       23.438768  63.970534  -87.523055   15.978416
      4q      -76.736528   0.622045 -153.721634  -62.089580
2019 1q      -43.799874  -55.875209  -71.388497  -40.167403
      2q      -54.916818  -25.457398   9.057110   30.389698
      3q      -93.653598  -37.642383  -26.845093   68.464715
      4q      -30.296984  108.756793  -17.532057  137.342178
2020 1q      -65.152815  -15.956905  -22.935770  100.381343
      2q     -112.787964  -89.956286  -49.698637   47.992254
```



# 기술설명 -코드-





```
def split(datasets,timesteps):  
    x_values=list()  
    y_values=list()  
    for i in range(len(datasets)-timesteps):#10-6  
        x=datasets[i:i+timesteps,1:]#0  
        y=datasets[i+timesteps,0]  
        x_values.append(x)  
        y_values.append(y)  
    return np.array(x_values),np.array(y_values)
```

```
from sklearn.preprocessing import RobustScaler

scaler = RobustScaler()
x_train=scaler.fit_transform(x_train)
x_test=scaler.transform(x_test)
x_pre=scaler.transform(x_pre)
```

## 소스코드 -모델(XGB)

```
model = XGBRegressor(n_estimators=2000)

#트레이닝

model.fit(x_train,y_train,eval_metric="rmse",eval_set=((x_train,y_train),(x_test,y_test)))

#테스트

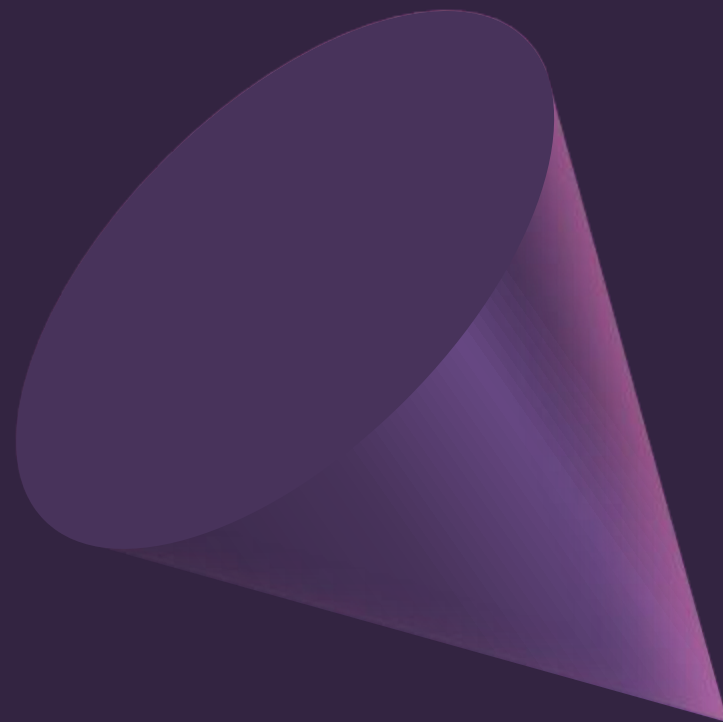
y_pre=model.predict(x_test)
y_pre = y_pre.reshape(-1,)

print(__file__)

print(f"r2:{r2_score(y_test,y_pre)}")
# print(f"x test.shape:{x_test.shape}")
```

DNN 모델로 진행을 같이 했으나, 유의미한 결과가 나오지 않아, 유의미한 결과가 나온 XGB로 1차 진행할 예정입니다.

결과



## 결과 -3개국

#2022

# # xgbregressor

# r2:0.9597776560881905

# y\_test:[396.31 411.5 394.82 414.54 388.47 431.38 387.52 397.68 430.15 39

# 390.39 433.26 434.73 432.63 395.01 417.8 410.38 393.74 428.56 424.74]

# y\_pre:[395.23767 410.71042 392.70322 416.73672 391.85153 426.7297 389.62

# 394.37204 421.3623 409.0211 393.6113 430.54108 430.2827 ]

R2:0.959

# 결과 -3개국

```
#2022
```

```
# # xgbregressor
```

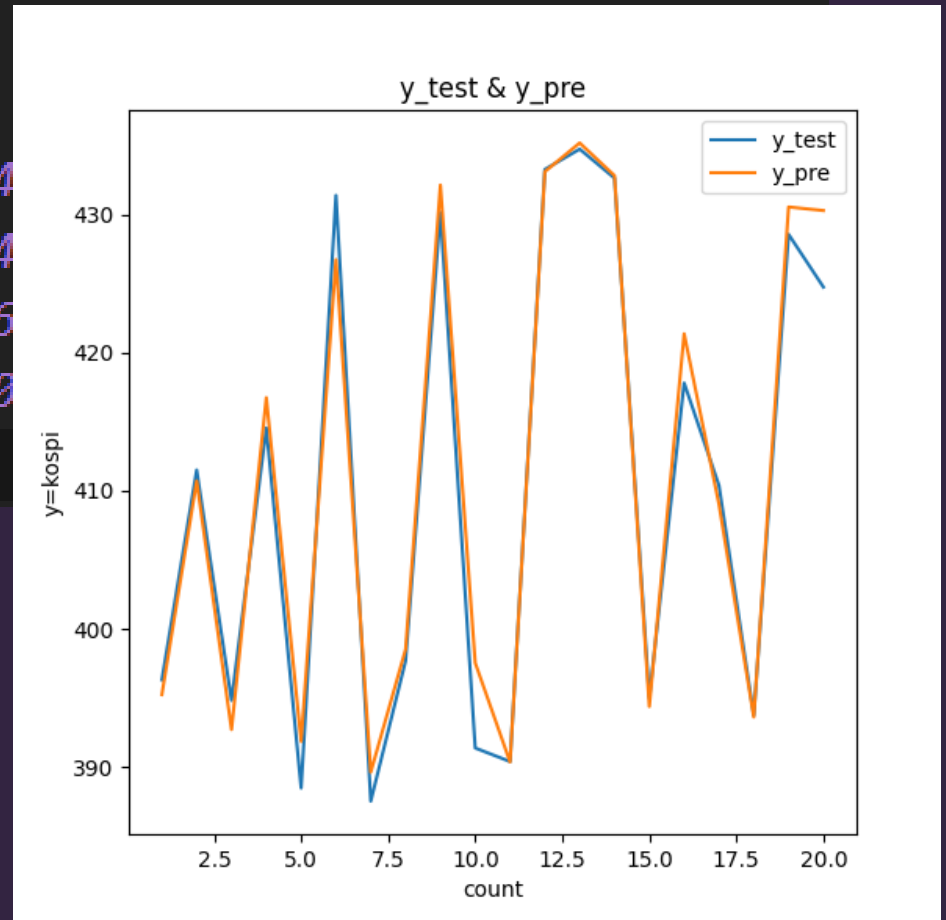
```
# r2:0.9597776560881905
```

```
# y_test:[396.31 411.5 394.82 414.54 388.47 4
```

```
# 390.39 433.26 434.73 432.63 395.01 417.8 4
```

```
# y_pre:[395.23767 410.71042 392.70322 416.736
```

```
# 394.37204 421.3623 409.0211 393.6113 430
```



감사합니다