

EDA

코로나-19 이후의 일상 변화와
문화산업에 대한 시사점
: 재택근무와 OTT 이용 증가를 기준으로

빅데이터분석

7조

2021.05.24

엄선민

차례

- 프로젝트 개요
- 탐구 배경
- 국내 코로나19 확진자 및 이동 추이
- 주요 검색어 비교 및 시각화
- 네이버 블로그 크롤링
- 국내 OTT 사용인구 시각화
- Insights
- 소감문

프로젝트 개요

프로젝트명	코로나19 이후 일상 변화와 문화산업에 대한 시사점		
사업명	재택 시간의 증가 - 원인과 현상: 재택근무와 OTT 사업 발전		
담당업무	텍스트마이닝, 시각화 등		
개발환경	Windows 10		
사용도구	Python, Jupyter, Colab, Tableau, R studio, NAVER API 등		
사용기술	KoNLPY 텍스트 자연어 분석, 정규화 등		
주요 라이브러리	Pandas, Numpy, Seaborn, Matplotlib, Os, Sys, Request, Random,(전처리, 시각화) Nltk, KoNLPY, Pytagcloud, Bs4 등(말뭉치)		

데이터 정보

유형	이름	출처	
정형데이터	Covid 19 - ROK	Data in our world	공공데이터
정형데이터	코로나바이러스감염증-19 환자 현황	중앙방역대책본부	공공데이터
정형데이터	코로나19 이후 일상 생활 변화	한국언론진흥재단	공공데이터
정형데이터	2019~2020 한국 OTT 서비스 이용 경험	미디어통계포털	공공데이터
비정형데이터	구글 이동 추이 보고서	Google	현실 마이닝
비정형데이터	구글 Trends	Google	소셜 마이닝
비정형데이터	네이버 블로그 크롤링	Naver	소셜 마이닝
비정형데이터	네이버 데이터랩	Naver	소셜 마이닝

데이터 전처리 요약

주요 전처리	결측값 처리	이상치 제거
	데이터 클렌징	데이터 스케일링
	피쳐 선택	추출 및 가공
세부 목표		
1	코로나19의 확산세와 이로 인한 일상의 변화 추적을 위한 동시기의 검색어 변동 분석	
2	실제 코로나19 기간 동안 재택 시간의 증감 확인 (동시기의 이동 추이로 본 일상 변화 추적(Google Mobility Index))	
3	네이버 블로그 연관 자연어 분석 - 말뭉치 추출	
4	영화 산업의 부진과 OTT 산업의 성장, 코로나19 확산 사이의 연관성 확인	

탐구 배경

- 감염병 확산을 막기 위해 지난 1년 반 동안 '사회적 거리두기' 정책이 시행되었고, 사람들의 대외활동은 축소되었다.
- 감염병 확산을 방지하기 위한 '집콕' 생활의 증가는 여러 산업에 영향을 끼쳤고, 영화산업은 여행, 연극 등과 더불어 가장 큰 피해를 입은 업종 중 하나이다.
- 재택근무를 시행할 경우 출퇴근 시 및 근무시간 동안 직장 내 감염병 확산을 차단할 수 있어 방역 시스템을 위해 정부 차원에서 크게 장려되었음.
- 영화 산업의 쇠퇴와 OTT 산업의 성장, 코로나 확산은 동시에 이루어졌으며 서로 연관성이 있는지를 확인하고자 한다.
- 많은 사람들은 늘어난 재택 시간동안 OTT 플랫폼을 많이 이용하게 되었다고 응답했다.

Q. 코로나19가 정말 재택근무를 확산시켰을까?

코로나로 직장인 "평균 52일 재택근무 했다"

잡코리아 | 2021-01-25 11:13 | 👁 3,908

코로나로 재택근무 장기화, 직장인 "평균 52일 재택근무 했다"

코로나 사태가 장기화되고 사회적 거리두기 단계가 2.5단계로 격상되며, 직장인들의 재택근무 기간 또한 길어지고 있는 것으로 나타났다. 실제로 직장인들은 코로나 사태 이후 평균 52일간 재택근무를 실시한 것으로 집계돼 눈길을 끌었다. 잡코리아가 직장인 839명을 대상으로 <재택근무 현황> 조사를 실시하고 결과를 발표했다.

이들 직장인들은 코로나 사태 이후 평균 52일간 재택근무를 실시한 것으로 집계됐다. 재택근무를 실시한 기간은 기업별로 상이했는데, 먼저 대기업(61일)과 중견기업(57일)의 경우 평균보다 많은 기간 동안 재택근무를 실시한 것으로 나타났다. 반면 중소기업(48일)과 공기업/공공기관(40일)은 재택근무 실시일이 평균보다 적었다.

- 재택근무가 가능한 산업은 한정되어 있다
 - 영화, 방송, 공연제작업 등은 많은 사람이 한 자리에 모여 협업하는 과정으로 이루어짐 : 재택근무가 어려움
 - 영화 관람 소비가 급감하고 제작환경이 급속도로 악화된 상태에서 해당 산업의 변화 양상을 살펴보는 것은 코로나19 이후 사회를 예측하는 데에 시의성을 제공할 것.

공/감/연/론 뉴시스
NEWSIS

등록 2020-12-12 06:00:00

코로나가 바꾼 영화 산업 지형도...

디즈니 이어 워너브러더스 OTT 강화
韓 승리호·콜·차인표 등 넷플릭스 직행
"영화 제작·투자에도 변화 나타날 것"

 연합뉴스



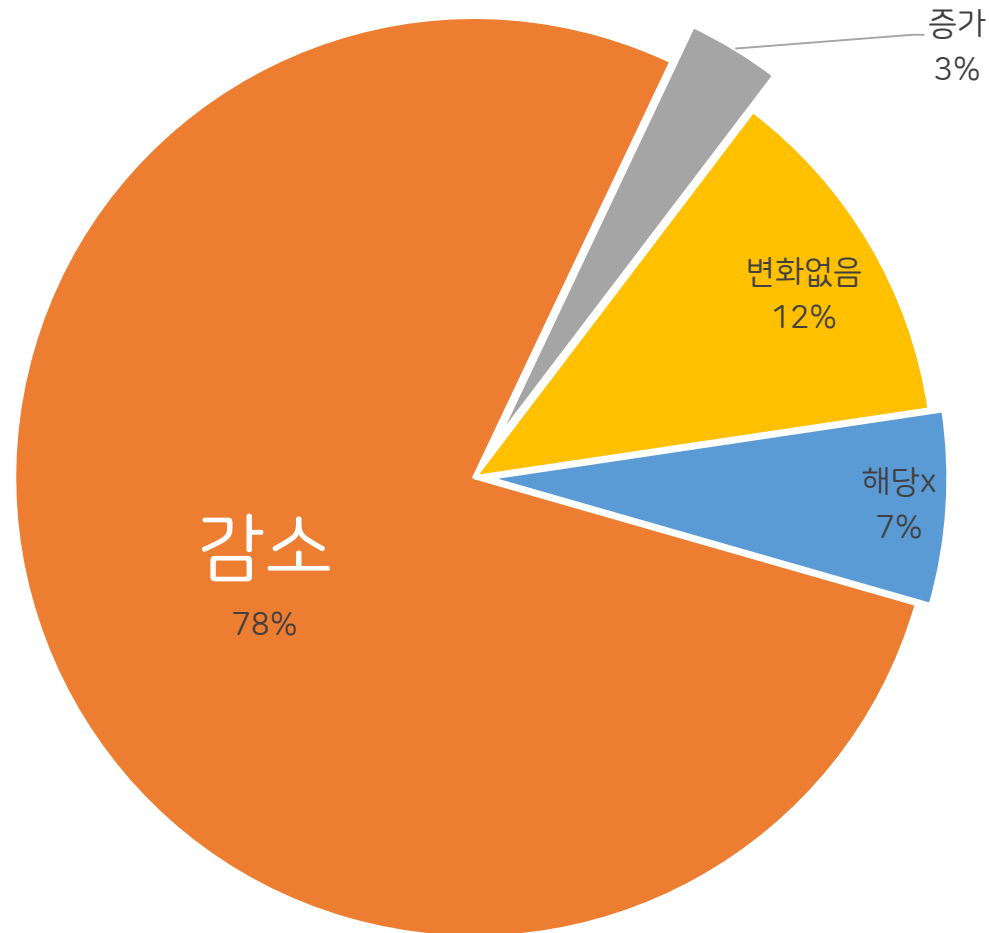
드라마 촬영중단에 휴방까지...방송가 코로나19 초비상(종합)

송고시간 | 2020-08-21 17:24

| 넷플릭스·JTBC 선제대응 이어 모든 현장 긴장...오만석·서이숙은 음성

Q. 코로나19가 정말 재택근무를 확산시켰을까?

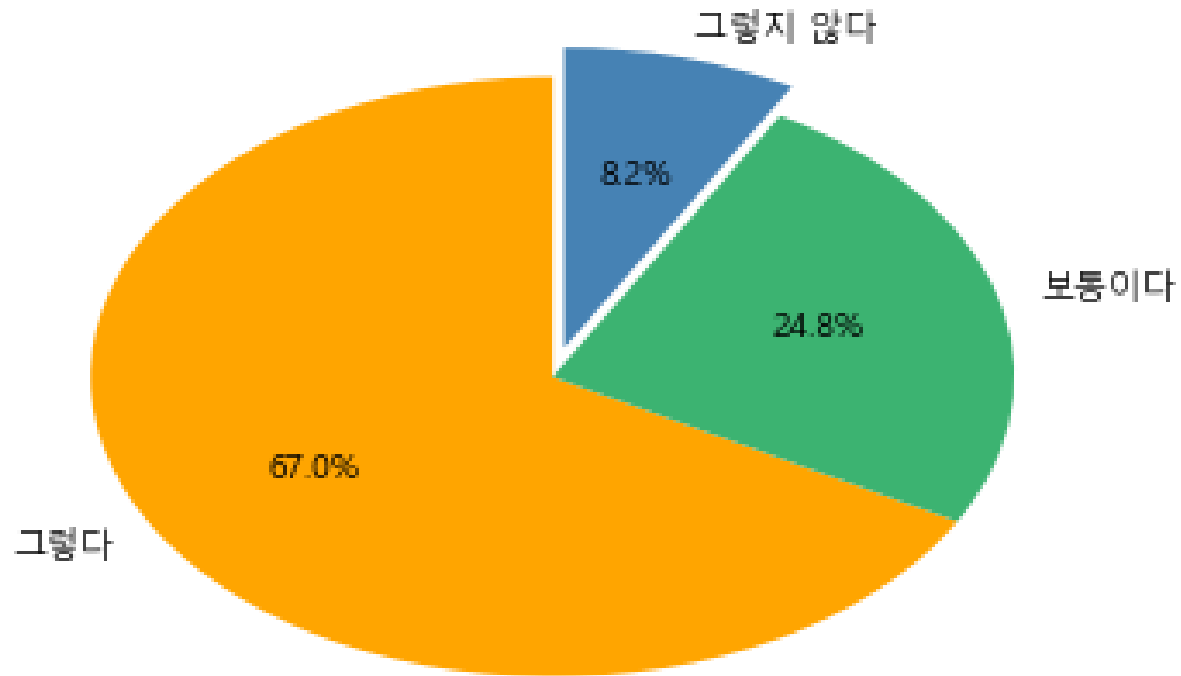
코로나19 이후 공연 · 예술 · 극장 등의 현장 관람 활동 변화



자료 출처 : 한국언론진흥재단 미디어연구센터, [코로나19 이후 국민의 일상 변화]

- 가설 : 코로나19로 인한 외출 감소, 재택 시간 증가는 향후에도 계속되어 영화 산업 등 미디어 활용에 영향을 끼칠 것이다.

집에서 여가를 즐기는 문화가 확산할 것이다

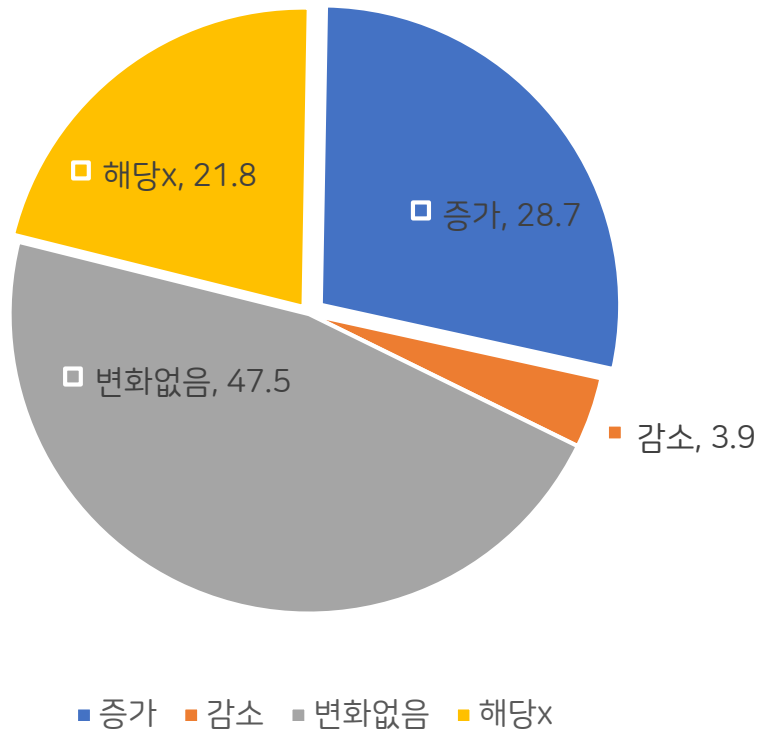


- 본 가설의 입증

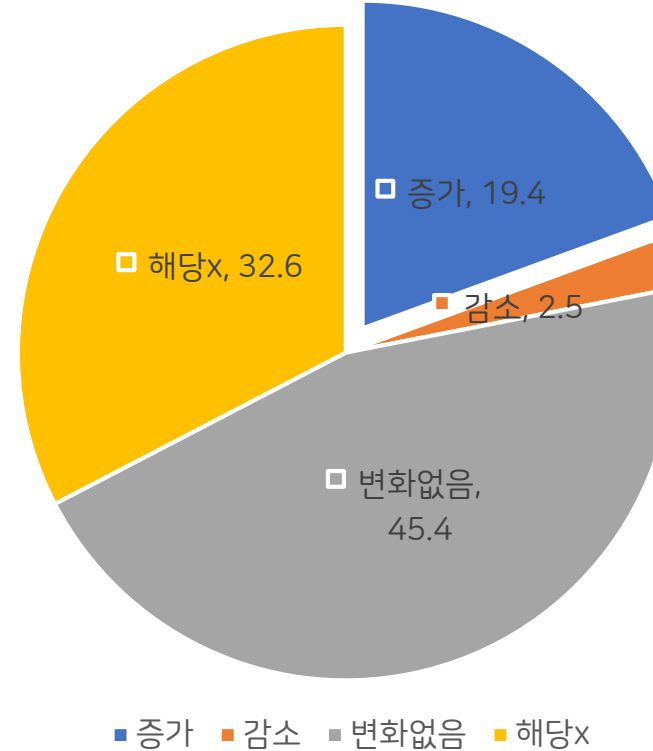
자료 출처 : 한국언론진흥재단 미디어연구센터, [코로나19 이후 국민의 일상 변화]

Q. 코로나19가 정말 재택근무를 확산시켰을까?

남성의 재택근무 여부



여성의 재택근무 여부



국내 코로나19 추이 및 동시기 이동량 시각화

국내 확진자 증감 추이

<https://ourworldindata.org/covid-google-mobility-trends>

```
outbreak=pd.read_csv('../3 - ott/owid-covid-data-korea.csv')
outbreak
#한국 확진자 정보. 20년 1월 20~ 5월 5일. ourworld in data 자료.
```

	iso_code	continent	location	date
0	KOR	Asia	South Korea	2020-01-21
1	KOR	Asia	South Korea	2020-01-22
2	KOR	Asia	South Korea	2020-01-23
3	KOR	Asia	South Korea	2020-01-24
4	KOR	Asia	South Korea	2020-01-25

```
# 필요한 정보만 : 일자, 신규확진자만.
outbreak.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 471 entries, 0 to 470
Data columns (total 59 columns):
 iso_code          471 non-null object
 continent         471 non-null object
 location          471 non-null object
 date              471 non-null object
 total_cases       470 non-null float64
 new_cases         469 non-null float64
```

```
outbreak = outbreak.drop(columns=['iso_code', 'continent', 'location', 'new_cases_smoothed', 'total_deaths', 'new_deaths', 'new_deaths_smoothed', 'total_cases_per_million', 'new_cases_per_million', 'new_cases_smoothed_per_million', 'total_deaths_per_million', 'new_deaths_per_million'])
```

date와 total_cases, new_cases를 제외하고 활용하지 않을 칼럼을 삭제한다.

```
#outbreak.to_csv('../3 - ott/outbreak.csv')
```

	date	total_cases	new_cases
0	2020-01-21	NaN	NaN
1	2020-01-22	1.0	NaN
2	2020-01-23	1.0	0.0
3	2020-01-24	2.0	1.0
4	2020-01-25	2.0	0.0
5	2020-01-26	3.0	1.0
6	2020-01-27	4.0	1.0
465	2021-04-29	122007.0	656.0
465	2021-04-30	122634.0	627.0
466	2021-05-01	123240.0	606.0
467	2021-05-02	123728.0	488.0
468	2021-05-03	124269.0	541.0
469	2021-05-04	124945.0	676.0
470	2021-05-05	125519.0	574.0

471 rows × 3 columns

Google Mobility Report

자료 출처 : <https://www.google.com/covid19/mobility/>



코로나19로 인한 지역사회 이동 추이 변화 확인하기

전 세계가 코로나19 사태에 대응하고 있는 가운데, 공중보건기관에서는 Google 지도와 같은 제품에 사용되는 방식과 동일하게 익명처리되어 집계되는 통계가 코로나19 확산 방지를 위한 중대한 결정을 내리는 데 도움이 된다고 말합니다.

지역사회 이동성 보고서의 목적은 코로나19 확산 방지 정책에 따라 일어나는 변화에 관한 정보를 제공하는 것입니다. 보고서에서는 소매점과 여가 시설, 식료품점과 약국, 공원, 대중교통 정거장, 직장, 주거지 등 여러 카테고리의 장소에 나타나는 시간별, 지역별 이동 추이를 도표로 만들어 보여줍니다.

기준값

이 데이터는 분류된 장소의 방문자(또는 소요 시간)가 기준일과 비교하여 어떻게 변했는지 보여 줍니다. 기준일은 해당 요일의 정상값을 나타냅니다. 기준일은 2020년 1월 3일부터 2월 6일까지 5주 기간의 중앙값입니다.

■ 장소 구분

소매점 및 여가시설
식료품점 및 약국 : 필수 방문지
공원
대중교통 정거장
직장
주거지

■ 시기 구분

2020/02/15 ~ 2021/05/05

이동 추이(mobility)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	country_re	country_re	sub_region	sub_region	metro_area	iso_3166_2	census_fip	place_id	date	retail_and	grocery_ar	parks_per	transit_sta	workplace	residential	percent_change_from	
2	KR	South Korea						ChIJm7oR	2020-02-15	-4	0	3	-6	1	3		
3	KR	South Korea						ChIJm7oR	2020-02-16	-11	0	-31	-15	-3	4		
4	KR	South Korea						ChIJm7oR	2020-02-17	-7	-6	-17	-6	0	5		

```
kr2020=pd.read_csv("../3 - ott/Region_Mobility_Report_CSVs/2020_KR_Region_Mobility_Report.csv")
kr2021=pd.read_csv("../3 - ott/Region_Mobility_Report_CSVs/2021_KR_Region_Mobility_Report.csv")
```

```
kr2020
```

```
kr2020drop=kr2020.drop(['country_region_code', 'country_region', 'sub_region_1', 'sub_region_2',
                        'metro_area', 'iso_3166_2_code', 'census_fips_code', 'place_id'], axis=1)
#불필요한 칼럼 제거
```

```
kr20cut=kr2020drop.loc[:320]
kr20cut #전국 수치만 잘라냄. 2020/2/15~2020/12/31
#date를 인덱스로 옮길 예정 * 21과 합칠 예정
```

원자료가 년도별/지역별(전국-수도권)로 분리되어 있음.
각 칼럼 정리 후 전국 단위 수치만 선별하여 통합한다.

이동량 전처리

```
kr20cut.isnull().sum() # 결측치 유무 확인
```

```
date          0
소매점 및 여가 시설    0
식료품점 및 약국      0
공원          0
대중교통 정거장      0
```

```
kr20cut.describe()
```

#간략한 요약, 중 평균> 직장 -8 소매점 -12 공원 등 야외 +27

316	2020-12-27	-35	-12	-8	-30	-23	11
317	2020-12-28	-13	13	22	-11	-12	9
318	2020-12-29	-18	4	-7	-14	-13	8
319	2020-12-30	-24	2	-20	-21	-16	11
					-16	-22	11

	소매점 및 여가 시설	식료품점 및 약국	공원	대중교통 정거장	직장	주거지
count	321.000000	321.000000	321.000000	321.000000	321.000000	321.000000
mean	-12.934579	5.940810	27.034268	-9.436137	-7.461059	4.333333
std	10.062869	9.800171	28.670031	9.423200	11.234957	3.798574
min	-45.000000	-18.000000	-38.000000	-38.000000	-38.000000	-38.000000
25%	-19.000000	1.000000	6.000000	-15.000000	-15.000000	-15.000000
50%	-11.000000	6.000000	30.000000	-9.000000	-9.000000	-9.000000
75%	-5.000000	10.000000	45.000000	-3.000000	-3.000000	-3.000000
max	15.000000	73.000000	158.000000	16.000000	16.000000	16.000000

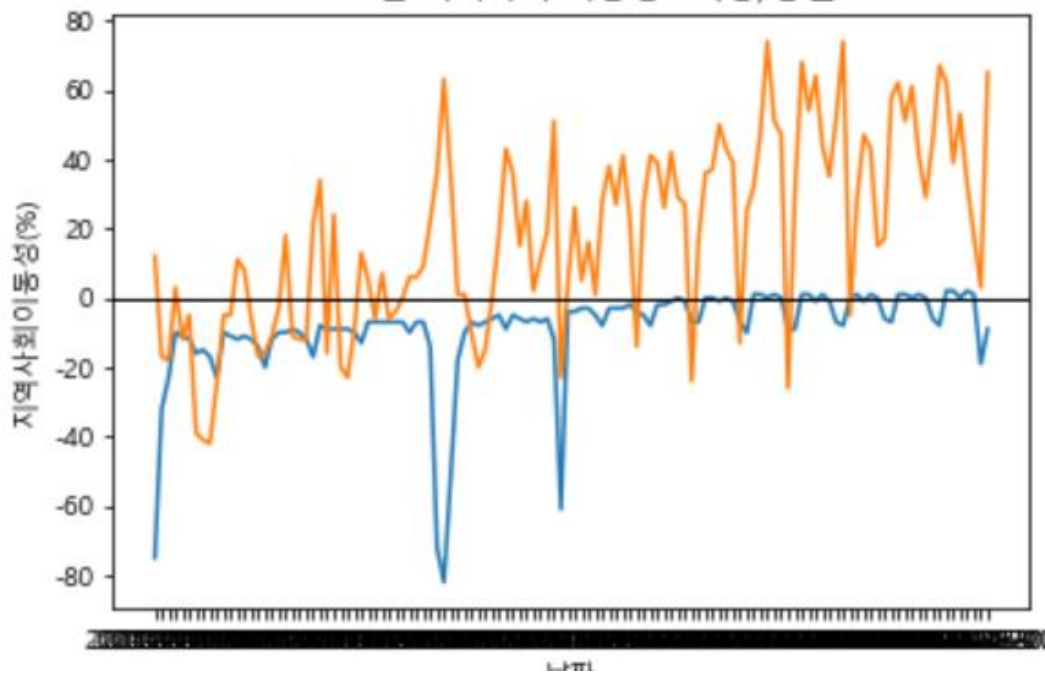
```
kr20cut.columns=['date', '소매점 및 여가 시설',
                  '식료품점 및 약국',
                  '공원',
                  '대중교통 정거장',
                  '직장',
                  '주거지']
```

```
#항목 변경
kr20cut
```



```
plt.plot(kr21cut['date'], kr21cut['직장'], kr21cut['공원'])
plt.axhline(y=0, color='black', linewidth=1)
plt.xlabel('날짜')
plt.ylabel('지역사회이동성(%)')
plt.title('2021년 지역사회 이동성 : 직장/공원')
plt.show()
```

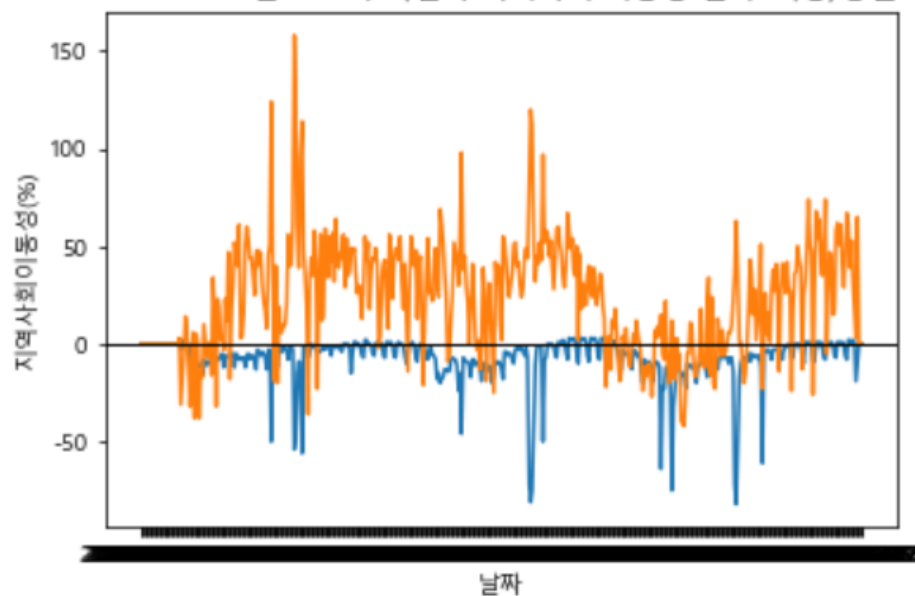
2021년 지역사회 이동성 : 직장/공원



```
plt.plot(covidoutbreak['date'], covidoutbreak['직장'], covidoutbreak['공원'])
plt.xlabel('날짜')
plt.ylabel('지역사회이동성(%)')
plt.title('2020~2021년 코로나 확진과 지역사회 이동성 변화: 직장/공원')
plt.axhline(y=0, color='black', linewidth=1)
plt.show()
```

#y축 신규 확진자 수로 하나 더 설정해야 함

2020~2021년 코로나 확진과 지역사회 이동성 변화: 직장/공원



공원 등 야외로의 이동이 기준치에 비해 크게 증가한 반면
직장으로의 이동은 연중 내내 기준에 비해 감소세를 이어갔다.

확진자수-이동량 정보 통합

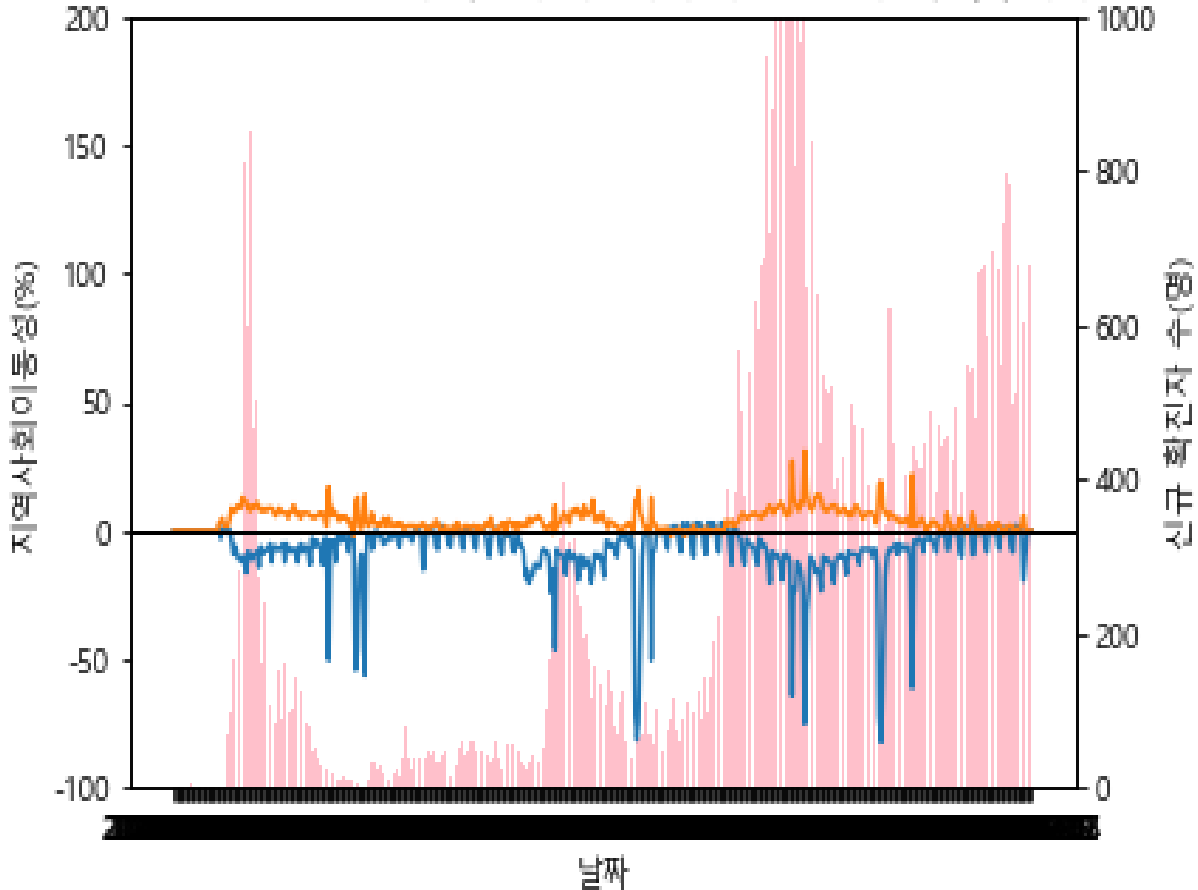
```
covid2021=pd.concat([kr20cut, kr21cut], ignore_index=True)
covid2021
```

	date	total_cases	new_cases	소매점 및 여가 시설	식료품점 및 약국	공원	대중교통 정거장	직장	주거지
0	2020-01-21	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2020-01-22	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	2020-01-23	1.0	0.0	NaN	NaN	NaN	NaN	NaN	NaN
3	2020-01-24	2.0	1.0						
4	2020-01-25	2.0	0.0						
465	2021-04-30	122634.0	627.0	-2.0	24.0	18.0	-2.0	1.0	2.0
466	2021-05-01	123240.0	606.0	-5.0	22.0	3.0	-12.0	-19.0	6.0
467	2021-05-02	123728.0	488.0	6.0	38.0	65.0	1.0	-9.0	1.0
468	2021-05-03	124269.0	541.0	0.0	0.0	0.0	0.0	0.0	0.0
469	2021-05-04	124945.0	676.0	0.0	0.0	0.0	0.0	0.0	0.0

```
covidoutbreak=covidoutbreak.fillna(0)
covidoutbreak
```

공원 등 야외로의 이동이 기준치에 비해 크게 증가한 반면
직장으로의 이동은 연중 내내 기준에 비해 감소세를 이어갔다.

2020~2021년 코로나 확진과 지역사회 이동성 변화: 직장/주거지



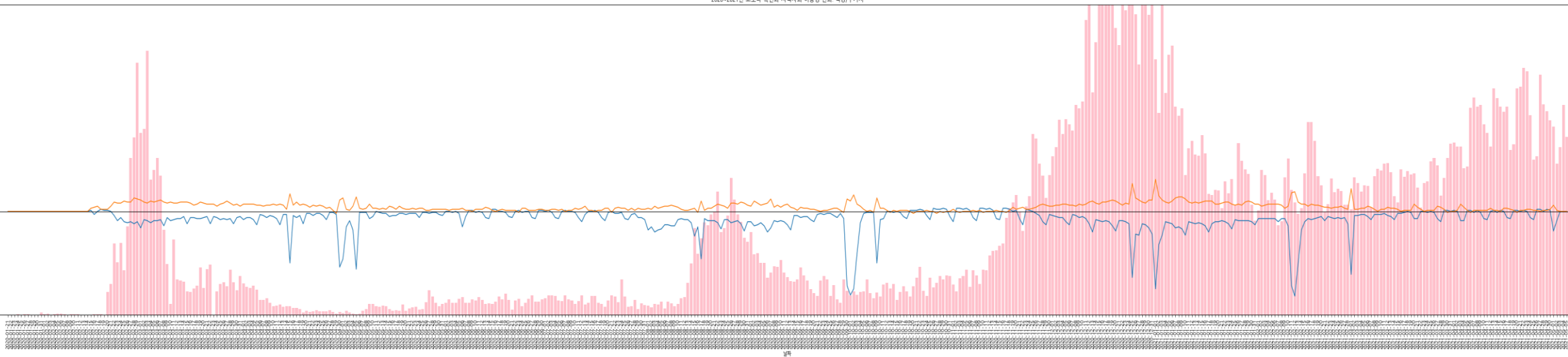
```
fig, ax1 = plt.subplots()
ax1.plot(covidoutbreak['date'], covidoutbreak['직장'], covidoutbreak['주거지'])
ax1.set_xlabel('날짜')
ax1.set_ylim(-100, 200)
plt.ylabel('지역사회이동성(%)')
plt.title('2020~2021년 코로나 확진과 지역사회 이동성 변화: 직장/주거지')
plt.axhline(y=0, color='black', linewidth=1)

ax2 = ax1.twinx()
ax2.bar(covidoutbreak['date'], covidoutbreak['new_cases'], color='pink', label='신규확진자')
ax2.set_ylim(0, 1000)
ax2.set_ylabel('신규 확진자 수(명)')

ax1.set_zorder(ax2.get_zorder() + 10)
ax1.patch.set_visible(False)

plt.show()
```

날짜 x라벨을 공유하는 두 그래프를 그림
: 좌/우 y라벨을
이동 추이 변화/일별신규확진자로 각기 달리 함.
일정한 추세가 있음을 시각적으로 확인 가능.
약 5회의 이동량 급변이 확인됨
주거지로의 이동량 변화는 소폭 상승
직장으로의 이동량은 수 차례 급감



2020-2021 코로나 주요 유행 시기

시기(년/월)	특징	
2020/01/12	국내 첫 확진자 발생	
2020/02~05	1차 대유행	
2020/08~09	2차 대유행	
2020/11~2021/01	3차 대유행	
2021/03~현재	3차 재확산 or 4차 대유행(예상)	

2020-2021 주요 공휴일과의 비교

일자	행사	비고
2020/04/15	21대국회의원선거	
20/05/01-05	노동절~어린이날 연휴	
20/08/15	광복절	
20/09/30-10/03	추석 연휴	
20/12/24-25	크리스마스 연휴	
20/12/31-21/01/01	연말연시	
21/02/10-14	설날 연휴	
21/03/01	삼일절	

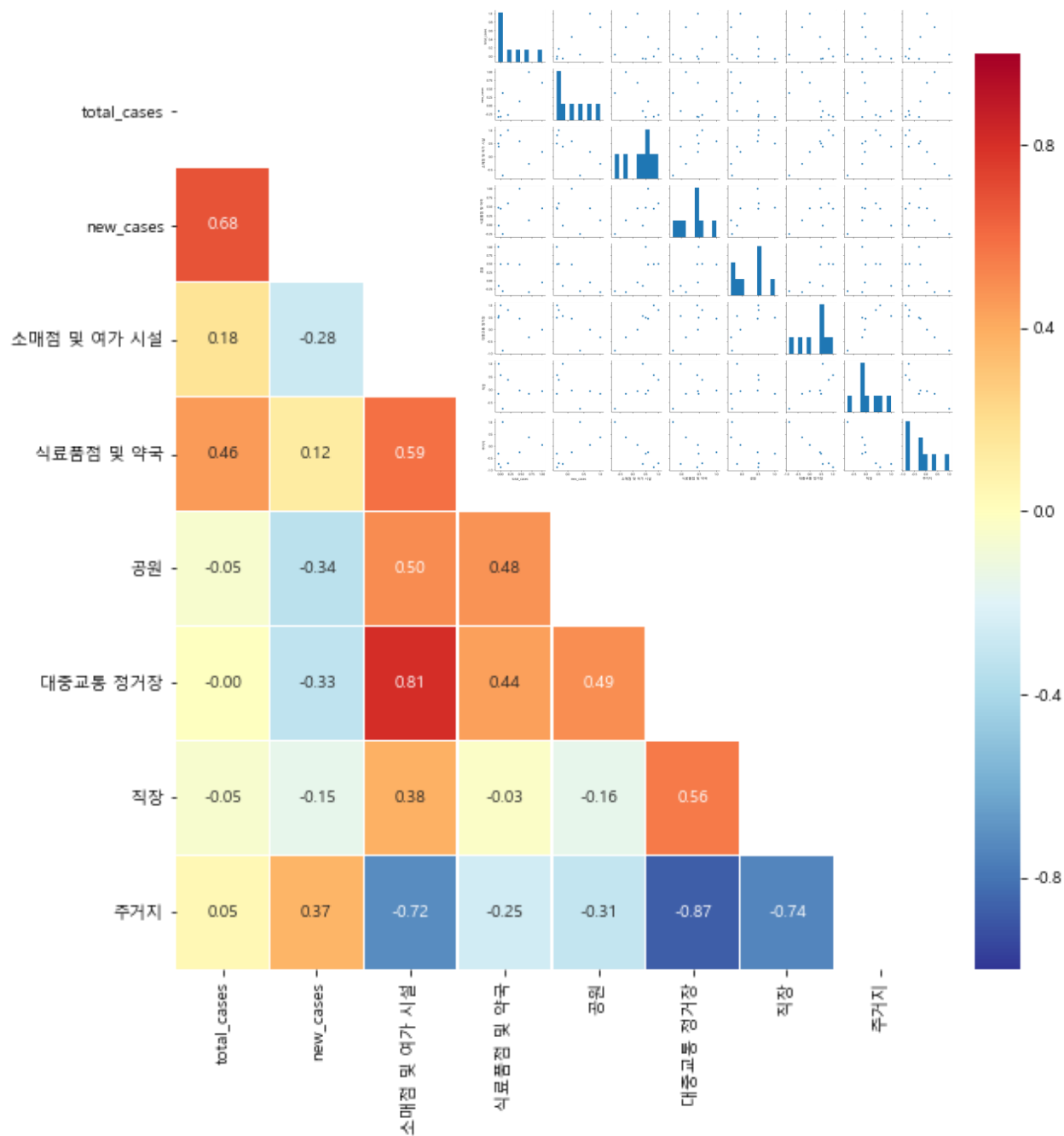
상관관계

양의 상관관계 大 :

- 신규 확진자 Δ = 전체 확진자 Δ
- 여가시설 Δ = 식료품점 Δ , 대중교통 Δ , 공원 Δ
- 직장 Δ = 대중교통 Δ

음의 상관관계 大 :

- 주거지 Δ = 대중교통 ∇ , 여가시설 ∇ , 직장 ∇



```
plt.figure(figsize=(10,10)) # 히트맵 그리기
```

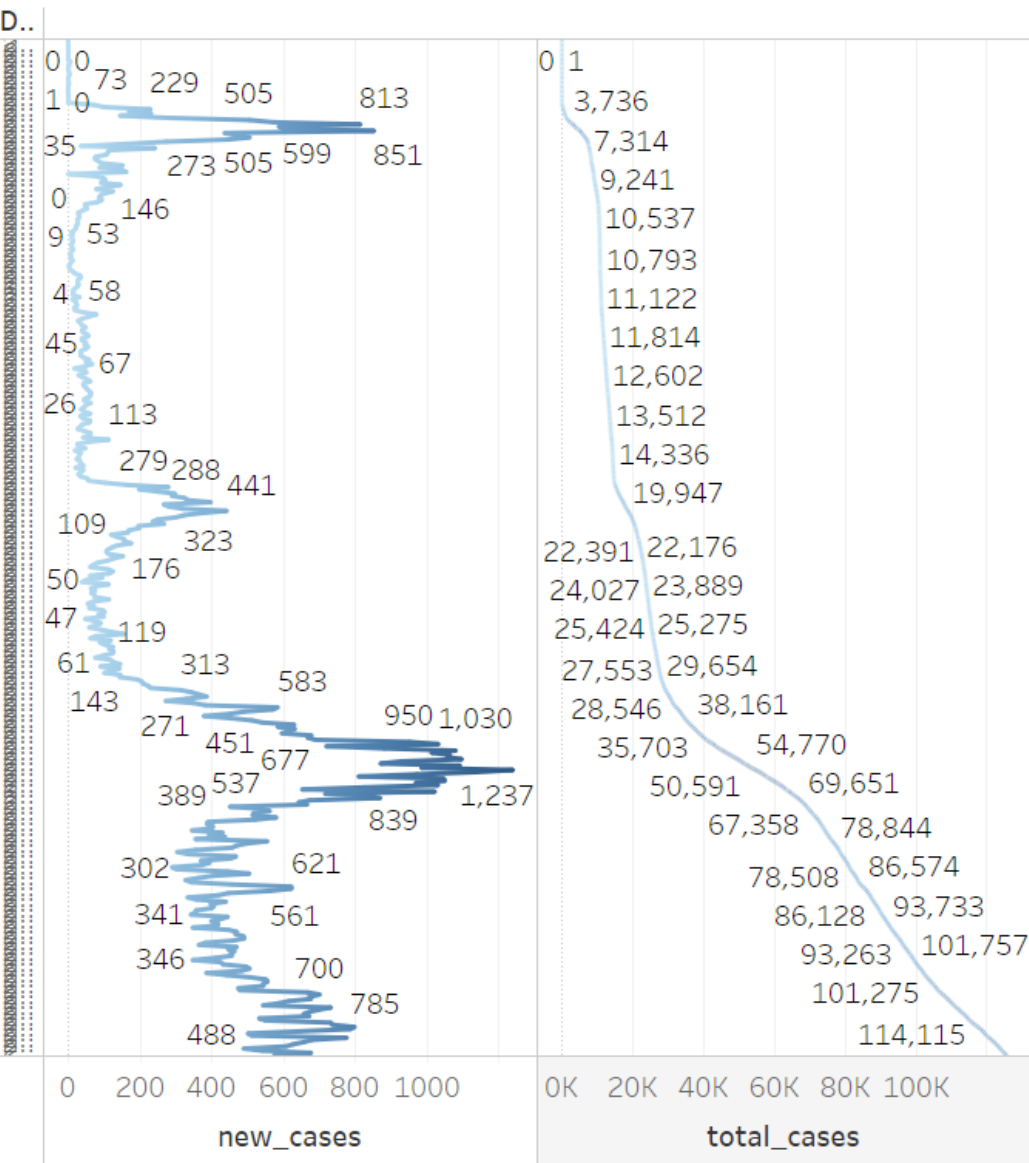
```
# 삼각형 마스크
```

```
mask= np.zeros_like(covidoutbreak.corr(), dtype=np.bool)
mask[np.triu_indices_from(mask)]= True
```

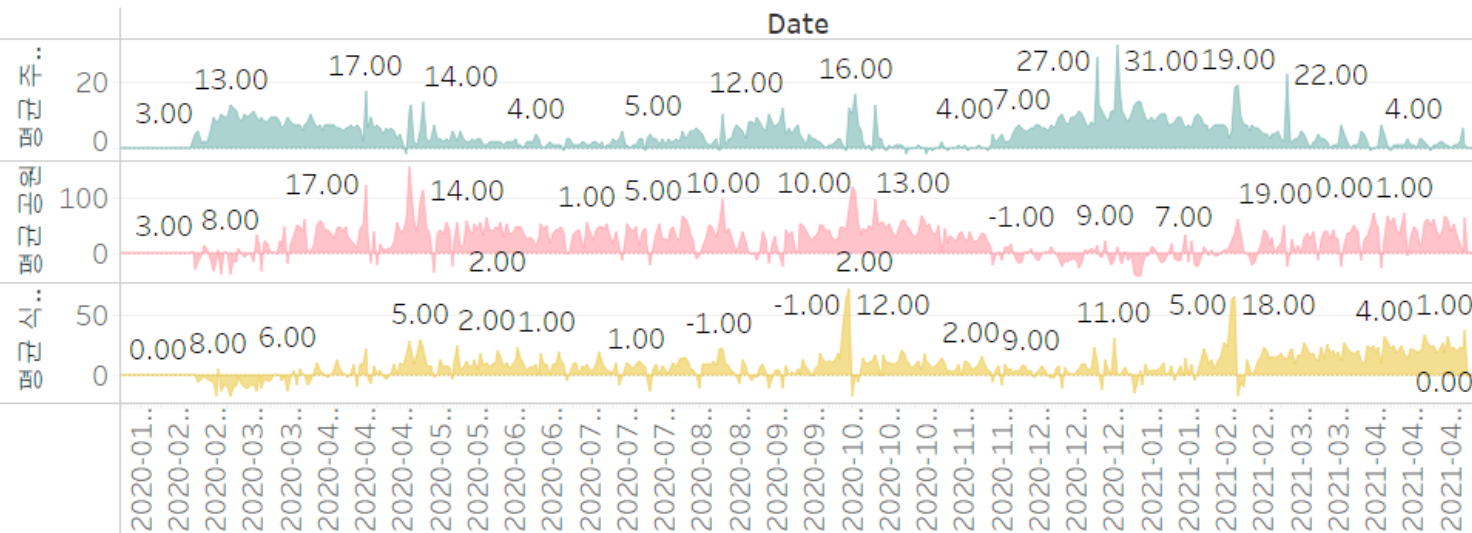
```
sns.heatmap(data=covidoutbreak.corr(),
             cmap='RdYlBu_r',
             annot=True,
             fmt='.2f',
             linewidths=.5,
             vmin= -1, vmax=1,
             mask = mask)
```

Tableau

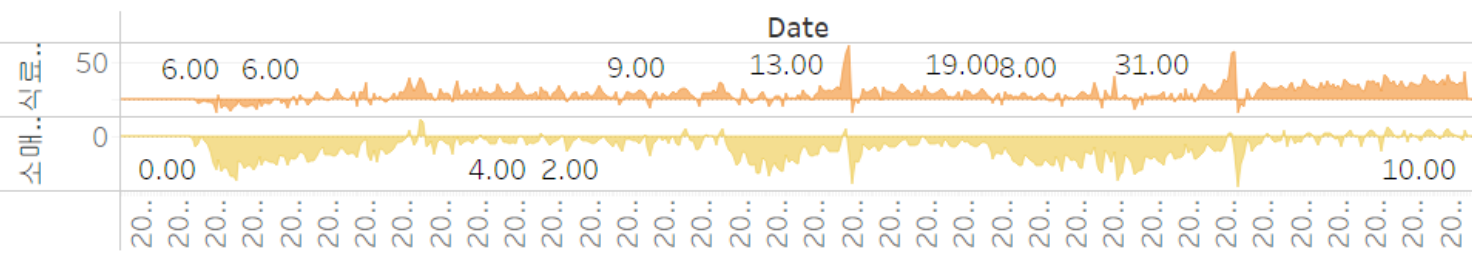
<한국의 COVID19 신규/누적 확진>



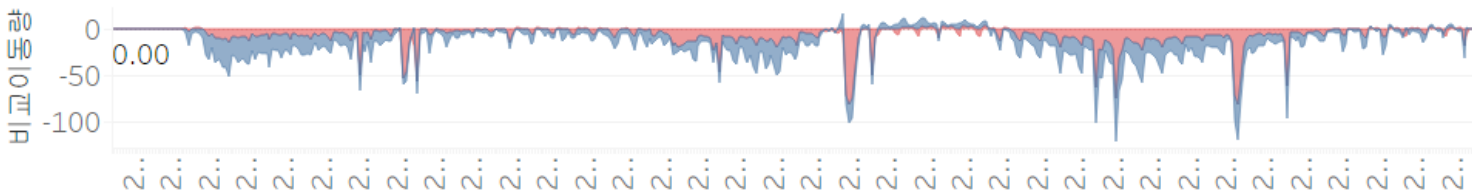
<이동량 변화 : 주거지, 공원, 식료품점 등의 증가-유지>



<식료품점, 약국, 소매점 및 여가시설>



<이동량 감소 : 대중교통 정거장, 직장>

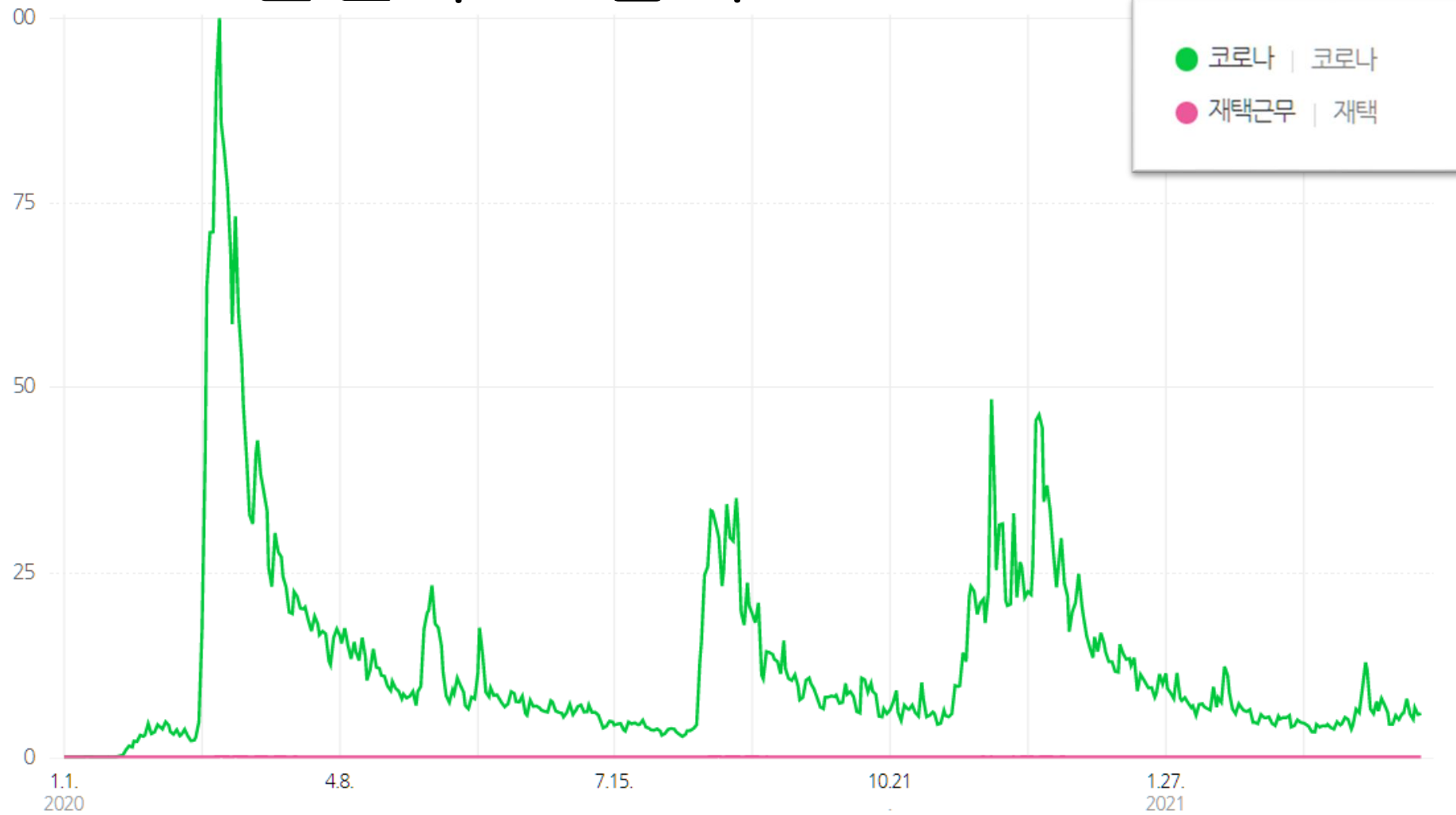


이동 추이를 통한 Insights

- 공원 등 야외 장소(국립공원, 공용 해수욕장, 정박지, 반려견 공원, 광장, 공공 정원)로의 이동성이 평년과 비교해서도 매우 증가한 것이 맞는지의 확인할 필요가 있다.
- 소매점 및 여가시설의 (기준값 대비) 6% 상승 - 세부 장소별 데이터 구분 필요(백화점, 쇼핑센터로의 이동량 증가와는 달리 식당, 놀이공원, 박물관, 도서관, 영화관 등의 장소로는 확연한 이동량 감소가 있었다고 예상)
- 주거지로의 이동 추이 변화가 타국들에 비해 적은데, 이는 미국, 영국 등과는 달리 전면적인 락다운 정책을 시행하지 않은 결과로 보인다.

코로나19 관련 주요 검색어 비교

코로나19 관련 주요 검색어 비교



출처 : 네이버 데이터랩 '코로나', '재택근무' 검색량 그래프



work from home
검색어

stay at home
검색어

home school
검색어

quarantine
검색어

lock down
검색어

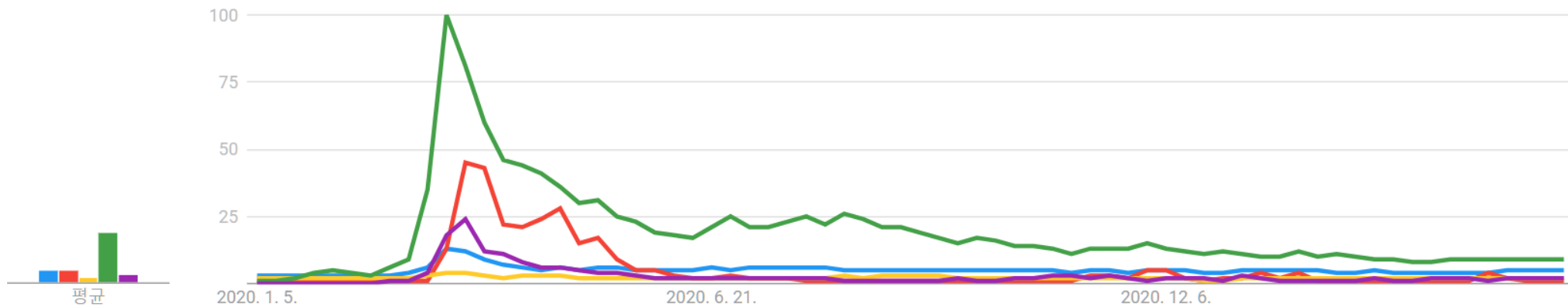
전 세계 ▼

20. 1. 1.~21. 5. 3. ▼

모든 카테고리 ▼

웹 검색 ▼

시간 흐름에 따른 관심도 변화 ?



출처 : Google Trends



재택근무
검색어

언택트
검색어

비대면
검색어

집콕
검색어



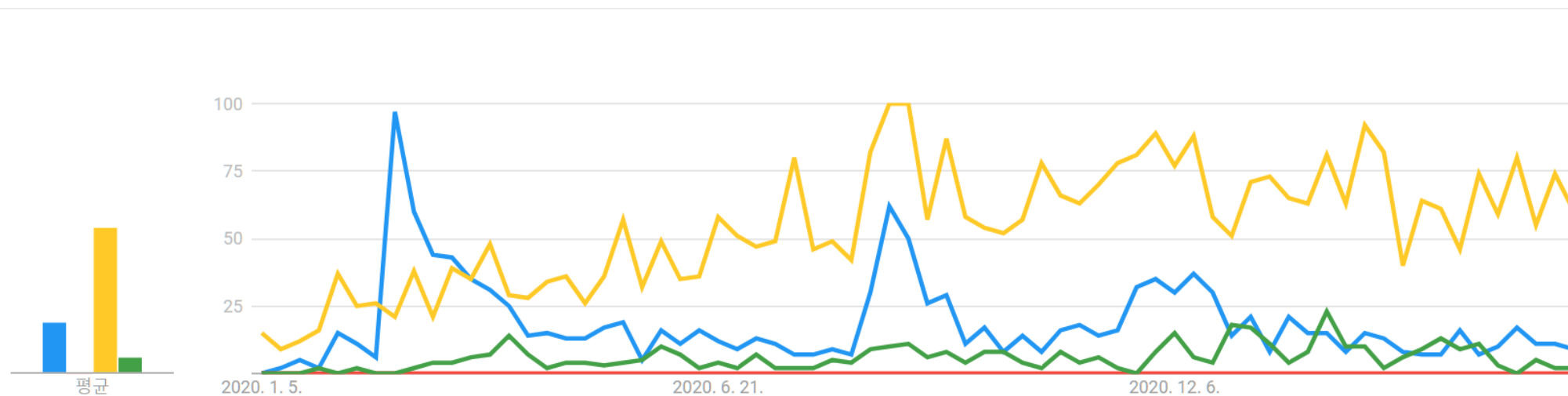
대한민국 ▼

20. 1. 1.~21. 5. 3. ▼

모든 카테고리 ▼

웹 검색 ▼

시간 흐름에 따른 관심도 변화 ?



주요 연관 단어 검색량 비교

- 목표: 코로나19 방지를 위해 대두된 정책 중 일상생활에 영향을 크게 미친 정도를 파악
- 수단: 주요 포털 검색량 비교
- 기간: 2020/01/01~ 2021/05/05, 주 단위
- 결과: 포괄적 방역 지침과 세부 정책 간 관심도 차이가 나타남.
(사회적 거리두기 > 비대면 > 자가격리 > 재택근무/
Quarantine > Stay at Home > Work from Home)

주요 연관 단어 검색량 시각화

```
timeline1=pd.read_csv('../3 - ott/multitimeline1.csv')
timeline1
```

	주	재택근무	Zoom 회의	원격수업	자가격리	사회적거리두기
0	2020-01-05	1	0	0	0	0
1	2020-01-12	1	0	0	0	0
2	2020-01-19	2	0	0	0	0
3	2020-01-26	1	0	0	2	0
4	2020-02-02	2	0	0	6	0
5	2020-02-09	2	0	0	7	0
6	2020-02-16	6	0	0	12	0
7	2020-02-23	35	3	0	34	1
8	2020-03-01	17	1	0	19	4
9	2020-03-08	22	3	0	16	3
10	2020-03-15	13	4	0	12	5
11	2020-03-22	13	4	0	19	14
12	2020-03-29	8	4	0	27	28
13	2020-04-05	10	8	0	30	27

```
timeline1.info()
# 68주 동안의 검색량 변화를 볼 수 있음.
```

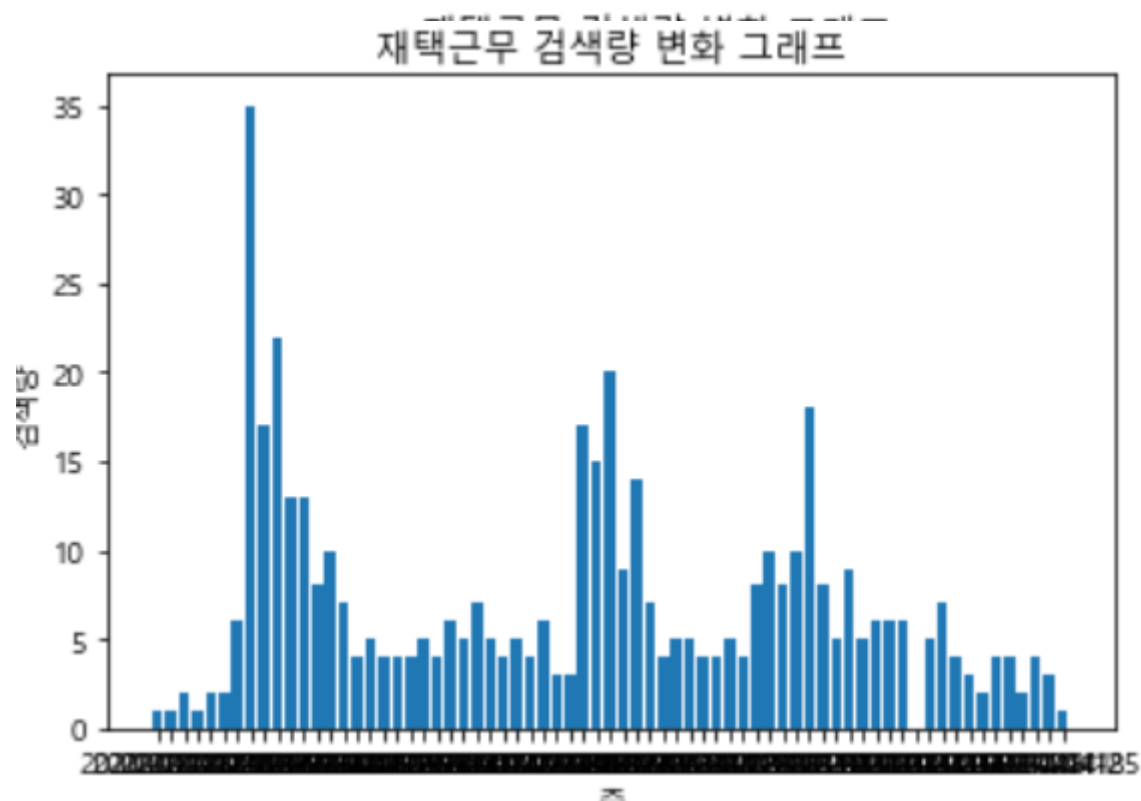
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 69 entries, 0 to 68
Data columns (total 6 columns):
주                69 non-null object
재택근무          69 non-null int64
Zoom 회의         69 non-null int64
원격수업          69 non-null int64
자가격리          69 non-null int64
사회적거리두기    69 non-null int64
dtypes: int64(5), object(1)
memory usage: 3.3+ KB
```

```
timeline1.describe()
```

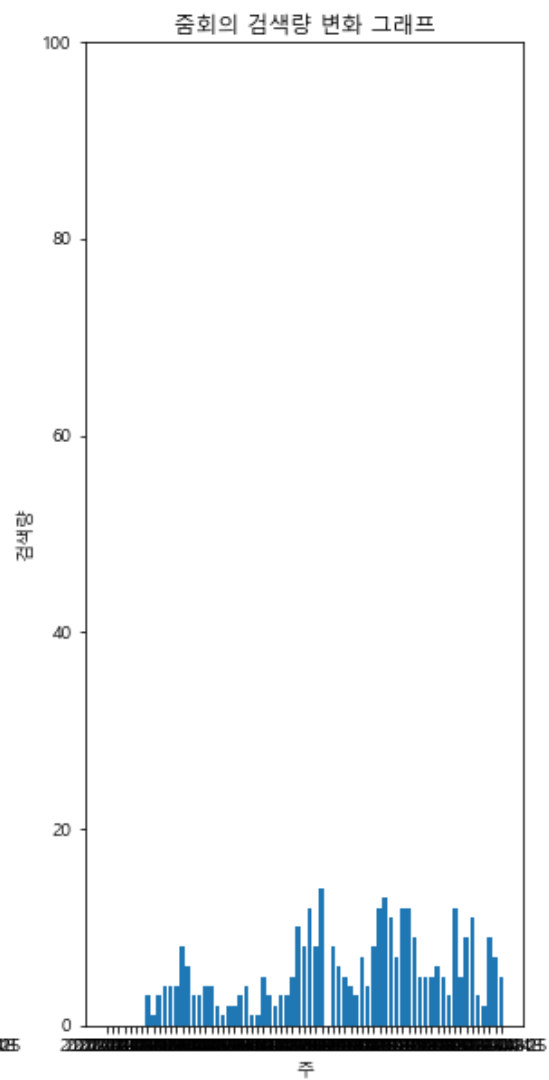
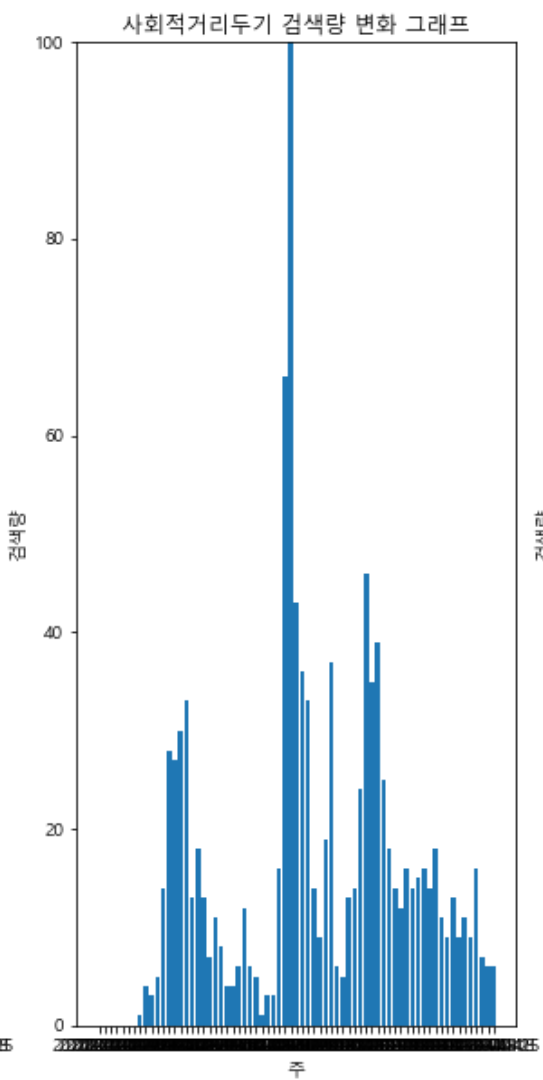
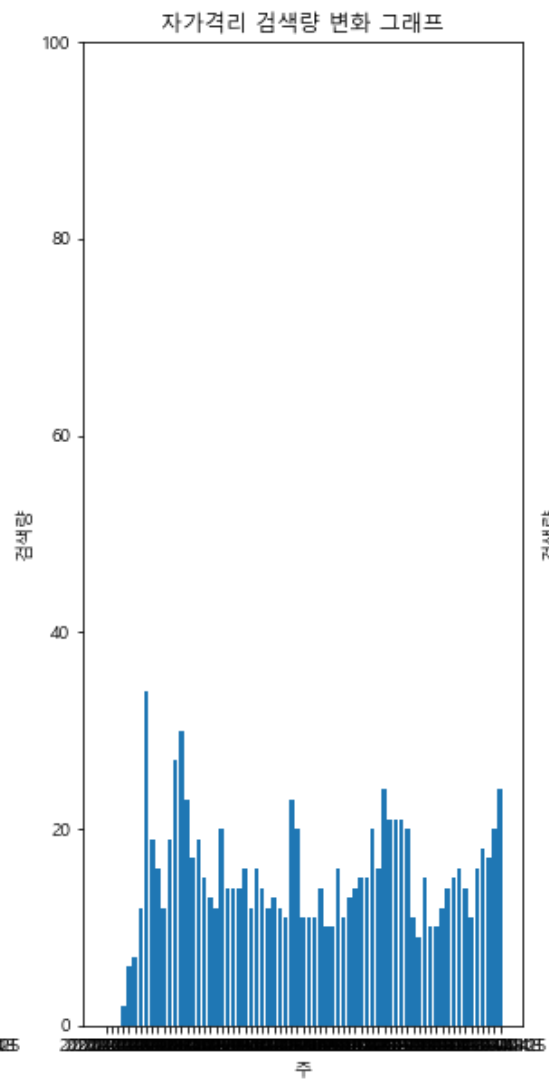
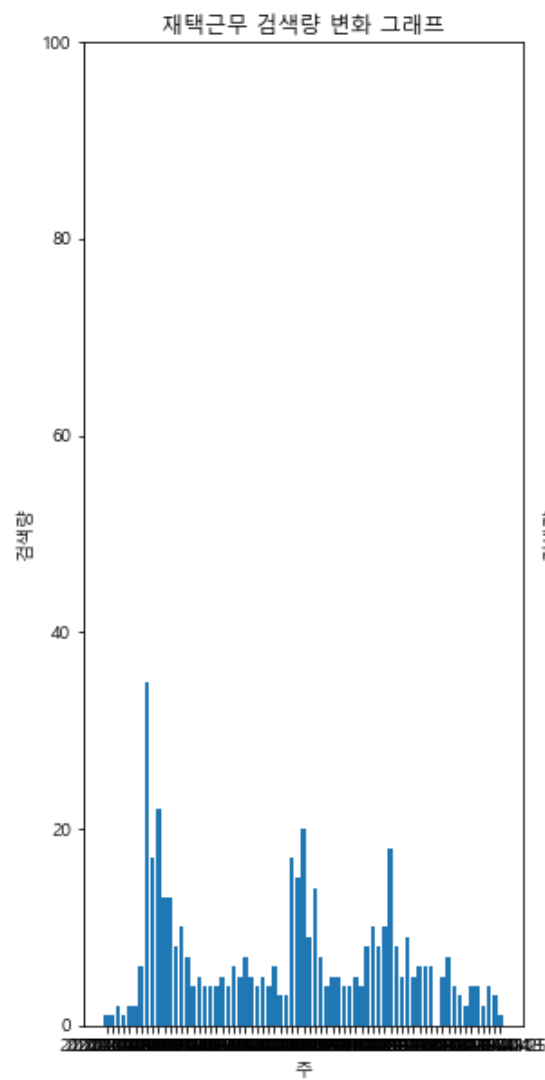
	재택근무	Zoom 회의	원격수업	자가격리	사회적거리두기
count	69.000000	69.000000	69.0	69.000000	69.000000
mean	6.797101	5.057971	0.0	14.782609	15.550725
std	5.799543	3.749055	0.0	6.232905	16.601325
min	0.000000	0.000000	0.0	0.000000	0.000000
25%	4.000000	3.000000	0.0	11.000000	5.000000
50%	5.000000	4.000000	0.0	14.000000	12.000000
75%	8.000000	8.000000	0.0	19.000000	18.000000
max	35.000000	14.000000	0.0	34.000000	100.000000

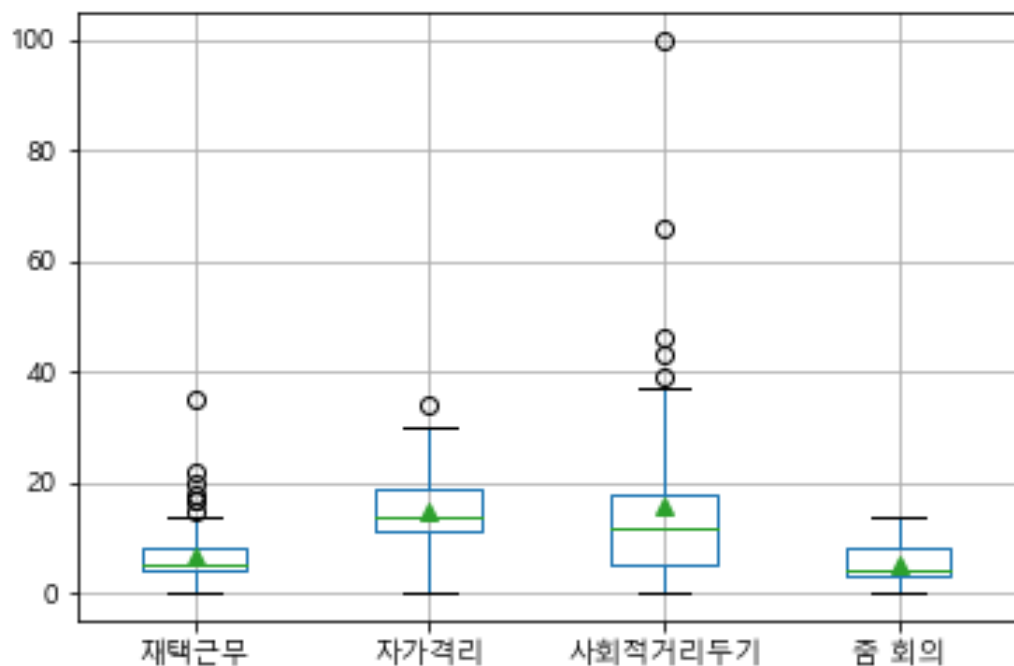
	주	재택근무	Zoom 회의	원격수업	자가격리	사회적거리두기
0	2020-01-05	1	0	0	0	0
1	2020-01-12	1	0	0	0	0
2	2020-01-19	2	0	0	0	0
3	2020-01-26	1	0	0	2	0
4	2020-02-02	2	0	0	6	0
5	2020-02-09	2	0	0	7	0
6	2020-02-16	6	0	0	12	0
7	2020-02-23	35	3	0	34	1
8	2020-03-01	17	1	0	19	4
9	2020-03-08	22	3	0	16	3
10	2020-03-15	13	4	0	12	5
11	2020-03-22	13	4	0	19	14
12	2020-03-29	8	4	0	27	28
13	2020-04-05	10	8	0	30	27
14	2020-04-12	7	6	0	23	30
15	2020-04-19	4	3	0	17	33
16	2020-04-26	5	3	0	19	13
17	2020-05-03	4	4	0	15	18
18	2020-05-10	4	4	0	12	12

```
plt.title('재택근무 검색량 변화 그래프')
plt.xlabel('주')
plt.ylabel('검색량')
plt.bar(timeline1['주'], timeline1['재택근무'])
plt.show()
```



```
plt.figure(figsize=[20,10])
plt.subplot(141)
plt.title('재택근무 검색량 변화 그래프')
plt.xlabel('주')
plt.ylabel('검색량')
plt.ylim(0,100)
plt.bar(timeline1['주'], timeline1['재택근무'])
plt.subplot(142)
plt.title('자가격리 검색량 변화 그래프')
plt.xlabel('주')
plt.ylabel('검색량')
plt.ylim(0,100)
plt.bar(timeline1['주'], timeline1['자가격리'])
plt.subplot(143)
plt.title('사회적거리두기 검색량 변화 그래프')
plt.xlabel('주')
plt.ylabel('검색량')
plt.ylim(0,100)
plt.bar(timeline1['주'], timeline1['사회적거리두기'])
plt.subplot(144)
plt.title('Zoom 회의 검색량 변화 그래프')
plt.xlabel('주')
plt.ylabel('검색량')
plt.ylim(0,100)
plt.bar(timeline1['주'], timeline1['Zoom 회의'])
plt.show()
```





```

timeline1.boxplot(['재택근무',
                  '자가격리',
                  '사회적거리두기',
                  '줌 회의'],
                  grid=True,
                  showmeans=True
                  )

```

'재택근무'에 대한 반응은 다른 검색어들에 비해 한정적이었음을 볼 수 있음. 가장 많은 영향력을 지닌 검색어는 '사회적 거리두기'가 가장 포괄적으로 장시간 많이 검색되었다.

'사회적거리두기'와 '자가격리'의 평균은 유사하나 최대값 등에서 큰 차이를 보임.

재택근무 관련 네이버 블로그 크롤링

네이버 블로그 크롤링

- 목표: 재택근무와 관련한 네이버 블로그 검색 결과를
- 수단: 네이버 블로그 내 “재택근무”로 검색, 결과를 말뭉치 방식으로 시각화
- 기간: 2020/01/01~ 2021/05/23
- 결과: 최근 1년 간 블로그 유저들은 ‘재택근무’는 ‘코로나/ 일상/ 직장인 / 오피스’ 등과 연관지어 관심사를 가지고 있음을 확인 가능.

네이버 블로그 검색 api 연결

```
In [35]: client_id=""  
client_secret=""
```

```
In [66]: encText = urllib.parse.quote("재택근무 -알바")  
url="https://openapi.naver.com/v1/search/blog.json?query="+encText  
url
```

```
Out [66]: 'https://openapi.naver.com/v1/search/blog.json?query=%EC%9E%AC%ED%83%9D%EA%B7%BC%EB%AC%B4%20-%EC%95%8C%EB%B0%94'
```

```
In [67]: request=urllib.request.Request(url)  
request.add_header("X-Naver-Client-Id", client_id)  
request.add_header("X-Naver-Client-Secret", client_secret)
```

```
In [68]: response=urllib.request.urlopen(request)  
response.getcode()
```

```
Out [68]: 200
```

검색 url 생성
특정 단어 제외 설정(불용어)
Json형식으로 결과 저장

```
def create_url(api_node, search_text, start_num, disp_num):  
    base="https://openapi.naver.com/v1/search"  
    node="/" + api_node + ".json"  
    param_query="?query="+urllib.parse.quote(search_text)  
    param_start("&start="+str(start_num)  
    parma_disp("&display="+str(disp_num)  
    return base+node+ param_query+param_start+parma_disp
```

```
client_id=""  
client_secret=""
```

```
url=create_url("blog", "재택근무 -알바 -부업 -주식 -코인  
-추천 -광고 -주부 -단기알바 -고액 -프리미엄 -고수익 -아르  
바이트 -대출 -복면가왕 -매매", 1,100)
```

```
request=urllib.request.Request(url)  
request.add_header("X-Naver-Client-Id", client_id)  
request.add_header("X-Naver-Client-Secret", client_secret)
```

```
response=urllib.request.urlopen(request)  
#response.getcode()  
#response.read()  
content=response.read().decode('utf-8')  
content
```

```
"lastBuildDate": "Fri, 14 May 2021 16:21:06 +0900",  
"total": 304627, "start": 1, "display": 100, "items": [  
  {"title": "2021년 2차 <b>재택근무</b> 종합컨설팅  
사업 수요기업 모집...", "link": "https://blog.naver.com/kotera?Redirect=Log&logNo=22231646386",  
  "description": "원격근무 방식도 포함 ○ 운영규정 마련  
인사·노무관리 체계 구축 ○ <b>재택근무</b> 관련 노  
적 쟁점사항 자문 ○ <b>재택근무</b> 친화적 조직문화
```

개선

```
result=json.loads(content)  
result
```

```
{  
  'lastBuildDate': 'Fri, 14 May 2021 16:21:06 +0900',  
  'total': 304627,  
  'start': 1,  
  'display': 100,  
  'items': [  
    {'title': '2021년 2차 <b>재택근무</b> 종합컨  
      'link': 'https://blog.naver.com/kotera?Redirect=Log&  
      'description': '원격근무 방식도 포함 ○ 운영규정 마련  
<b>재택근무</b> 친화적 조직문화 개선 교육 ○ <b>재택근무</b>  
      'bloggername': '정책자금에 관한 모든 것',  
      'bloggerlink': 'https://blog.naver.com/kotera',  
      'postdate': '20210420'},  
    {'title': '<b>재택근무</b>용 사무의자 시디즈 비교 후기  
      'link': 'https://blog.naver.com/shineway?Redirect=Lo  
      'description': '조정으로 <b>재택근무</b> 일수도 대폭  
      다행인지 허먼밀러 에어론에 대한 아쉬움이 좀 더 줄었다 을  
      'bloggername': 'shineway',  
      'bloggerlink': 'https://blog.naver.com/shineway',  
      'postdate': '20210420'}  
  ]  
}
```

```
pd.DataFrame({'제목':titles, '설명':descriptions, '작성일':postdates})
```

앞서 반복 검색한 결과에서 태그를 제외하여 데이터프레임 형태로 저장.
말뭉치 corpus 형태로 변환 출력.

		설명	작성일
0	2021년 2차 재택근무 종합컨설팅 지원사업 수요기업 모집...	원격근무 방식도 포함 ◦ 운영규정 마련 등 인사·노무관리 체계 구축 ◦ 재택근...	20210420
1	재택근무용 사무의자 시디즈 비교 후기 : T20 (탭 플러스)...	조정으로 재택근무 일수도 대폭 줄었고 요새 확진자가 암만 늘어나도 재택...	20210428
2	재택근무, 회사에 가지 않아도 돼요! 줌화상회의, 줌미팅	사람인에서 기업 355개사를 대상으로 한 조사에서는 재택근무를 실시하는...	20210511

	제목	설명	작성일
0	2021년 2차 재택근무 종합컨설팅 지원 사업 수요기업 모집...	원격근무 방식도 포함 ◦ 운영규정 마련 등 인사·노무관 리 체계 구축 ◦ 재택근무 관...	20210420
1	재택근무용 사무의자 시디즈 비교 후기 : T20 (탭 플러스)...	조정으로 재택근무 일수도 대폭 줄었고 요새 확진자가 암만 늘어나도 재택을 늘릴 기미...	20210428
2	재택근무, 회사에 가지 않아도 돼요! 줌 화상회의, 줌미팅	사람인에서 기업 355개사를 대상으로 한 조사에서는 재택근무를 실시하는 109개의 ...	20210511

```
title_corpus = "".join(result_search['제목'].tolist())
description_corpus = "".join(result_search['설명'].tolist())
print(title_corpus)
```

2021년 2차 재택근무 종합컨설팅 지원사업 수요기업 모집... 재택근무용 사무의자 시디즈
비교 후기 : T20 (탭 플러스)... 재택근무, 회사에 가지 않아도 돼요! 줌화상회의, 줌
미팅재택근무 무엇이든 물어보세요!(카드뉴스로 보는 Q&A)21年4月.日本) 일본 일상/재
택근무자의 점심/금융성향테스트복면가왕 재택근무 vs 피맥 정체 누구?인강용 노트북 문

말뭉치에서 명사 단위 분리, 한글자 단어 및 불용어 제거

```
from konlpy.tag import Okt
from collections import Counter
```

```
nouns_tagger = Okt()
nouns = nouns_tagger.nouns(description_corpus)
count = Counter(nouns)
```

count

```
Counter({'격': 21,
        '근무': 145,
        '방식': 7,
        '포함': 7,
        '운영': 9,
        '규정': 2,
        '마련': 2,
        '등': 26,
```

```
# 한글자 키워드를 제거합니다.
remove_char_counter = Counter({x : co
print(remove_char_counter)
```

```
Counter({'재택근무': 285, '코로나': 3
14, '일기': 14, '오피스': 12, '복면':
```

```
# 한국어 약식 불용어사전 예시 파일입니다. 출처 - (https://www.ranks.nl/stopwords
rean)
```

```
korean_stopwords_path = '../korean_stopwords.txt'
```

```
# 텍스트 파일을 엽니다.
```

```
with open(korean_stopwords_path, encoding='utf-8') as f:
    stopwords = f.readlines()
stopwords = [x.strip() for x in stopwords]
print(stopwords[:10])
```

```
['아', '휴', '아이구', '아이쿠', '아이고', '어', '나', '우리', '저희', '따라']
```

```
# 네이버 블로그 페이지에 맞는 불용어를 추가합니다.
```

```
naver_blog_stopwords = ['19금', '거나', '때문', '재택근무', '재택', '근무', '복면']
for stopword in naver_blog_stopwords:
    stopwords.append(stopword)
```

가장 출현 빈도수가 높은 100개의 단어를 선정합니다.

```
ranked_tags = remove_char_counter.most_common(100)
```

pytagcloud로 출력할 100개의 단어를 입력합니다. 단어 출력의 최대 크기는 80으로 제한합니다.

```
taglist = pytagcloud.make_tags(ranked_tags, maxsize=80)
```

pytagcloud 이미지를 생성합니다. 폰트는 나눔 고딕을 사용합니다.

```
pytagcloud.create_tag_image(taglist, 'wordcloud.jpg', size=(500, 800), fontname='NanumGothic', rectangular=False)
```

최다빈도 단어부터 워드클라우드 형태로 제작

```
nouns_tagger = Okt()
```

```
nouns = nouns_tagger.nouns(title_corpus)
```

```
count = Counter(nouns)
```

```
remove_char_counter = Counter({x : count[x] for x in count if len(x) > 1})
```

```
remove_char_counter = Counter({x : remove_char_counter[x] for x in count if x not in stopwords})
```

```
ranked_tags = remove_char_counter.most_common(100)
```

```
taglist = pytagcloud.make_tags(ranked_tags, maxsize=80)
```

```
pytagcloud.create_tag_image(taglist, 'title_wordcloud.jpg', size=(600, 600), fontname='NanumGothic', rectangular=False)
```

```
Image(filename='title_wordcloud.jpg')
```


WordCloud 출력



코로나	33
일상	29
근무	18
재택	17
시대	14
일기	14
오피스	12
왕	11
홈	11
피맥	10
복면	10
용	10
필수	9
모니터	9

재태근무



2019~2020 국내 OTT 서비스 이용 시각화

2019~2020 한국 OTT 서비스 이용 경험

```
ott=pd.read_csv('../3 - ott/OTT_exp_kor.csv') # 미디어통계포털, 2019
~2020 한국 OTT 서비스 이용 경험
```

구분별		구분별(2)	2019	2019.1	2019.2	2020										
구분별(1)		구분별(2)	사례 수 (건)	이용경험 있음 (%)	이용경험 없음 (%)	사례 수 (건)										
0	구분별(1)	구분별(2)	사례 수 (건)	이용경험 있음 (%)	이용경험 없음 (%)	사례 수 (건)										
1	전체	소계	10864	41.0	59.0	10302	72.2	27.8								
2	성별	남	5401	44.0	56.0	5146										
3	성별	여	5463	38.1	61.9	5156										
4	연령	만10대미만	346	23.3	76.7	258										
5	연령	만10-19세	1036	48.4	51.6	984										
6	연령	만20-29세	1564	65.4	34.6	1440										

		남	여	9	19	29	39	49	59	69	70
2019 이용경험 유		41.0	44.0	38.1	23.3	48.4	65.4	62.0	47.2	35.5	19.8
2019 이용경험 무		59.0	56.0	61.9	76.7	51.6	34.6	38.0	52.8	64.5	80.2

```
plt.pie(ott2019['남'],
        labels=['유', '무'],
        autopct='%1.1f%%')
plt.title("2019년 남성(5401명) 중 OTT 서비스 이용 경험", fontdict=
{'weight':'bold', 'size':18})
plt.show()
```

```
ott=ott.transpose()
```

성별/연령 추출

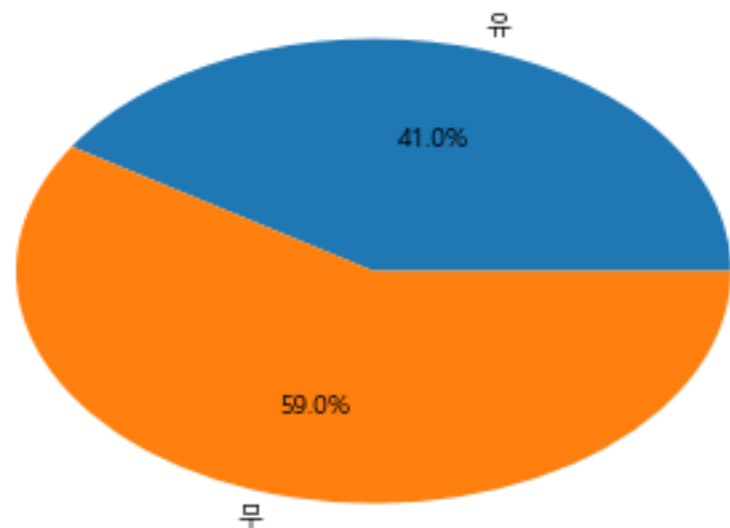
```
ott1=ott.iloc[:, [1,2,3,4,5,6,7,8,9,10]]
```

19/20년 구분 추출

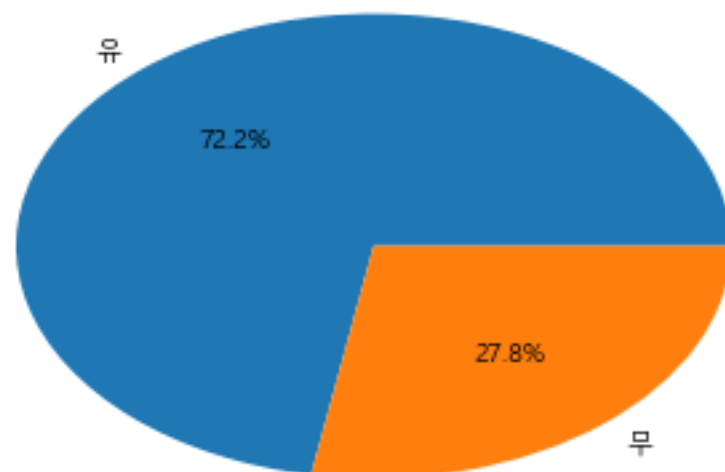
```
plt.pie(ott2019['남'],
        labels=['유', '무'],
        autopct='%1.1f%%')
plt.title("2019년 남성(5401명) 중 OTT 서비스 이용 경험", fontdict=
{'weight':'bold', 'size':18})
plt.show()
```

#2019년 남성(5401명) 중 OTT 서비스 이용 확률

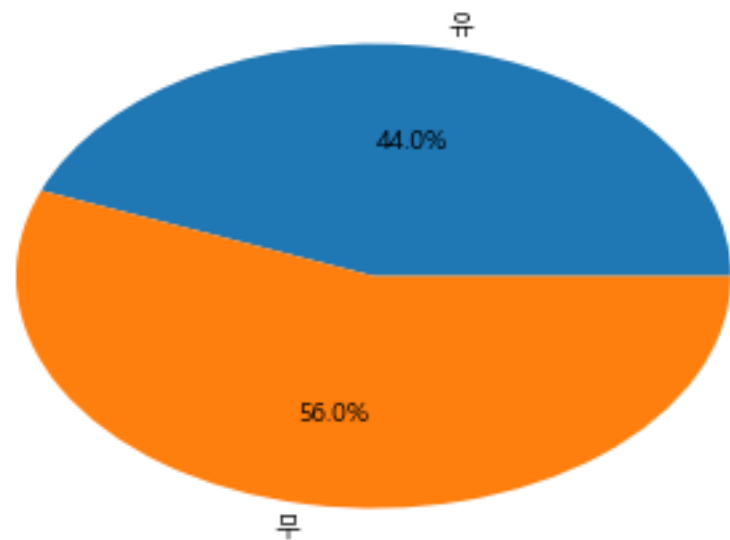
2019년 남성(5401명) 중 OTT 서비스 이용 경험



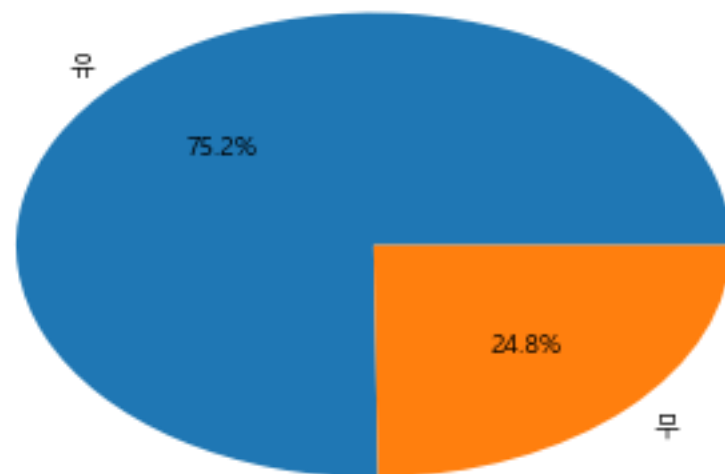
2020년 남성(5146명) 중 OTT 서비스 이용 경험



2019년 여성(5463명) 중 OTT 서비스 이용 경험



2020년 여성(5156명) 중 OTT 서비스 이용 경험



고려해야 할 점:

- 동시기 확산세는 존재했으나, 실제 변화의 요인으로 직접 연결될 수 있는지는 탐구해야 함.
- OTT 사업은 2010년대부터 이후 계속 성장세에 있었음.
- 개봉 영화의 극장 상영과 대형 마케팅이 어려워진 상황에서, OTT 플랫폼이 새로운 돌파구가 될 것이라는 예측은 계속되었으나 선후관계, 상관관계의 확인은 필요하다.

Insights

- 언론 등에서 코로나19로 인해 재택근무가 활성화되었다고 보도하고 있지만, 실제로 재택근무가 사회 전반에 시행되었다기에는 무리가 있다.
- 시기적으로 감염병 방지 차원에서 일부 기간동안 이동을 자제시킨 영향을 찾아볼 수 있으나, 그보다는 명절, 공휴일 등 계절 요소가 더 크게 작동하였다.
- 주요 검색어를 분석함으로써 실제로 코로나19 시기동안 많은 사람들의 관심사가 어디를 향해 있었는지 확인할 수 있었다. 사회 전반적으로는 '사회적 거리두기'를 포함한 일상 생활의 변화에 대한 관심이 특히 높았다.
- 말뭉치 검색을 통해 '재택근무'는 일부 산업 사무직에 한정되어 이루어진 변화이자 새로운 경험으로 기록할 만한 변화상으로 기록되고 있음을 볼 수 있었다.
- 향후 OTT 이용 인구 증가가 미칠 영향에 대해서는 후속탐구가 필요하다.

소감문

- 금번 프로젝트를 통해 탐색적 데이터 분석(EDA) 과정을 체험할 수 있었다. 다양한 시각화 툴을 이용해 보는 과정은 흥미로웠고, 공공데이터를 사용한 만큼 나름의 사회적 함의를 찾아보게 되며 데이터분석에 대한 관심을 더욱 키워나가게 되었다. 파이썬 라이브러리를 다양하게 사용하면서 전처리 과정 전반을 손에 익혀가는 데에 보람이 있었다.
- 특히 아쉬움이 남는 점은 머신러닝(ML)을 수행할 수 있을 만한 데이터를 확보하는 데에 실패했다는 점이다. 어떤 데이터를 선별하여 사용할 것인가에 대한 고민과 탐색이 중요함을 다시금 깨닫는 기회가 되었다.
- 텍스트분석으로

감사합니다

Thank you for listening.