

팀원끼리 모르는 팀



Crawler Dataton

김수빈 임성묵 손봉우

INDEX

1. 주제에 관하여

2. 취미생활 - 영화

3. 음식소비 - 배달음식

4. 결론 도출

INDEX

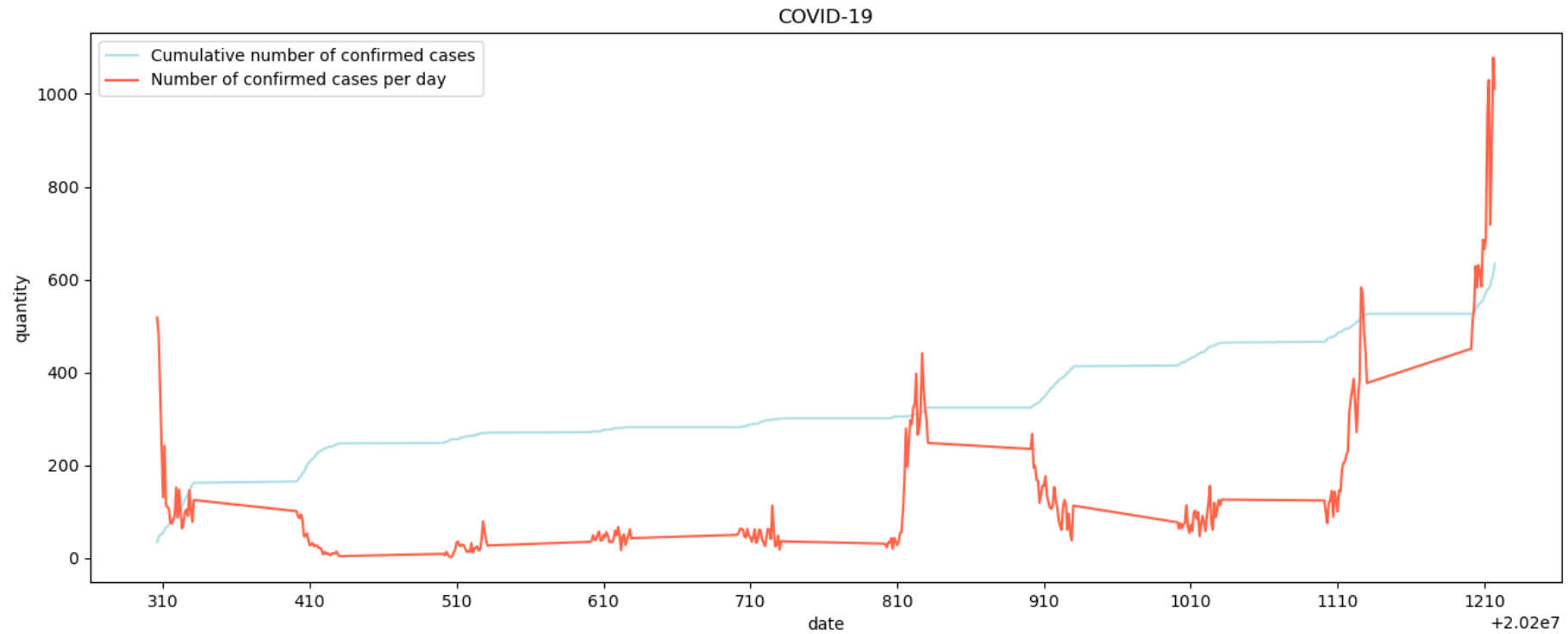
1. 주제에 관하여

2. 취미생활 - 영화

3. 음식소비 - 배달음식

4. 결론 도출

1. 주제에 관하여



보건복지부_코로나19감염_현황.csv
일일 코로나 확진자 수
누적 코로나 사망자 수



웹 크롤링을 이용한 코로나가 취미생활과 음식소비 문화에 미친 영향 분석



INDEX

1. 주제에 관하여

2. 취미생활 - 영화

3. 음식소비 - 배달음식

4. 결론 도출

2. 취미생활 - 영화

```
# -*-coding:utf-8 -*-
import ...
sys.stdout = io.TextIOWrapper(sys.stdout.detach(), encoding='utf-8')
sys.stderr = io.TextIOWrapper(sys.stderr.detach(), encoding='utf-8')

class movie_data:
    def __init__(self, title, date, offline, online):
        self.title = title
        self.date = date
        self.offline = offline
        self.online = online
        self.ratio = round(int(offline.replace(",","."))/int(online.replace(",",".")))

url = 'http://www.kobis.or.kr/kobis/business/stat/online/onlineYearly8c
res = requests.get(url)
html = res.text
soup = BeautifulSoup(html, 'html.parser')
text = soup.text.strip()
title_list = soup.select("#table > tbody > tr > td > a > span")
date_list = soup.select("#table > tbody > tr > td:nth-child(4)")
offline_list = soup.select("#table > tbody > tr > td:nth-child(10)")
online_list = soup.select("#table > tbody > tr > td:nth-child(11)")
movie_data_ob = []

before = [] #2020년 전 영화
after = [] #2020년 이후 영화

for i in range(len(title_list)):
    m1 = movie_data(title_list[i].text, date_list[i].text, offline_list[i].text, online_list[i].text)
    movie_data_ob.append(m1)
for i in movie_data_ob:
    year = i.date[0:4]
    if int(year) < 2020:
        before.append(i)
    else:
        after.append(i)
```

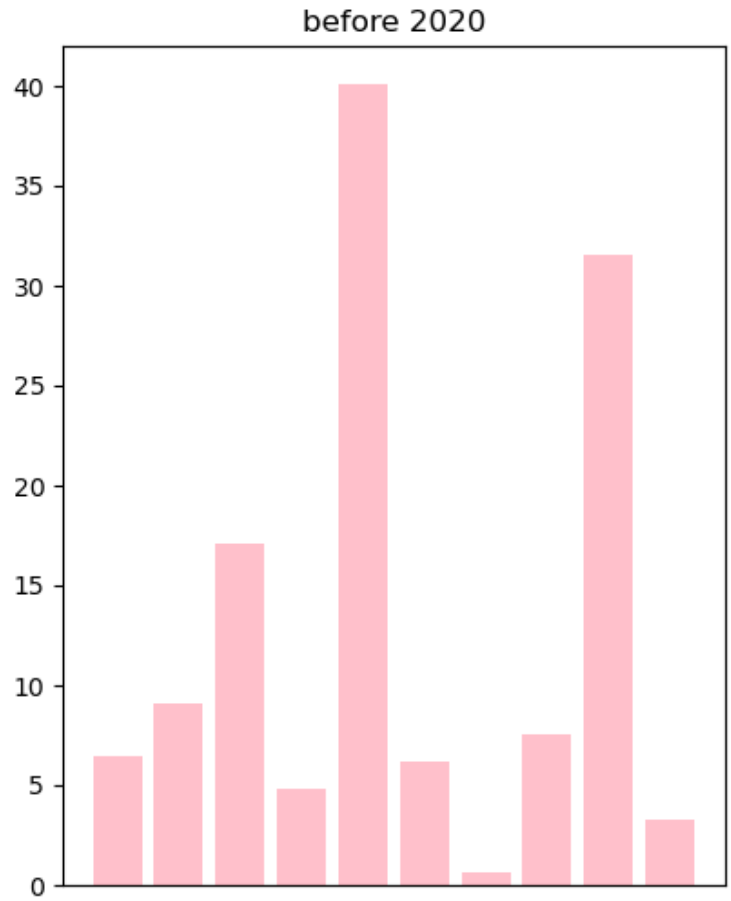
1 ~ 25 위		26 ~ 50 위	51 ~ 75 위	76 ~ 100 위	101 ~ 125 위	126 ~ 150 위	151 ~ 175 위	176 ~ 200 위	201 ~ 225 위	226 ~ 250 위	251 ~ 275 위	276 ~ 300 위
순위	영화명		개봉일	제작국가	장르	감독	배우	상영시간	극장 누적 관객수	온라인 이용건수(연 간)		
1		기생충 (PARASITE)	2019-05-30	한국	드라마	봉준호	송강호,이선균,조여정,최...	131분 39초	10,313,145	1,602,824		
2		백두산 (ASHFALL)	2019-12-19	한국	어드벤처	이해준,김병서	이병헌,하정우,마동석,전...	128분 0초	8,252,909	908,321		
3		겨울왕국 2 (Frozen 2)	2019-11-21	미국	애니메이션	크리스 벅,제니퍼 리	크리스틴 벨,이디나 멘젤...	103분 4초	13,747,792	804,764		
4		시동 (START-UP)	2019-12-18	한국	드라마	최정열	마동석,박정민,장혜인,염...	101분 48초	3,318,479	692,852		
5		남산의 부장들 (The Man Standing Next)	2020-01-22	한국	드라마	우민호	이병헌,이성민,곽도원,이...	113분 38초	4,750,345	656,967		
6		정직한 후보 (HONEST CANDIDATE)	2020-02-12	한국	코미디	장유정	라미란,김무열,나문희,윤...	104분 30초	1,537,491	624,264		

KOBIS(영화관입장권통합전산망)

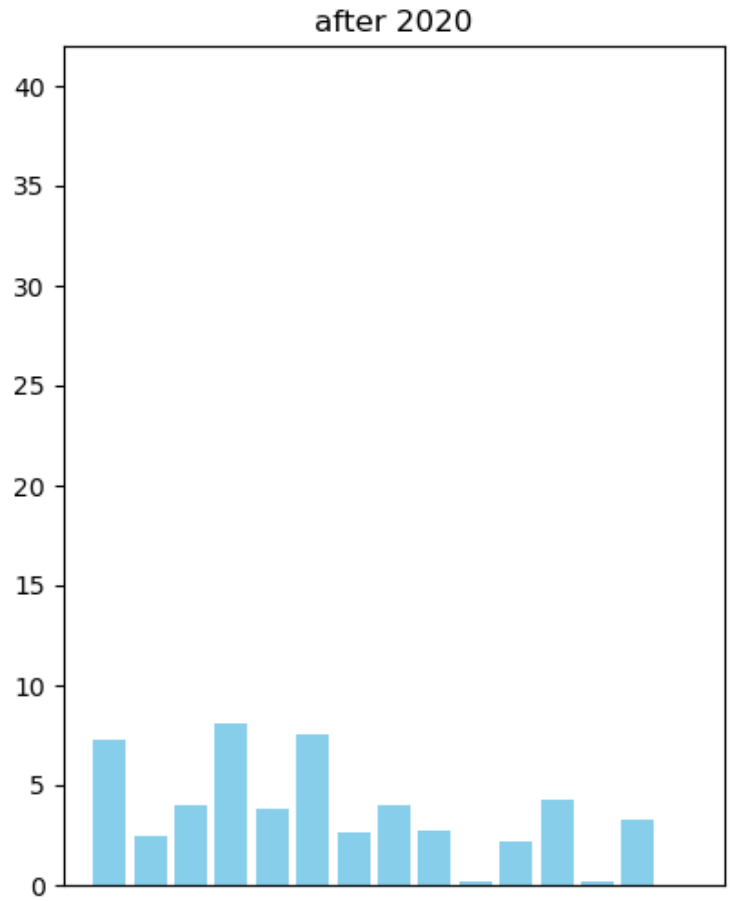
극장 누적 관객수, 온라인 이용건수 크롤링

2. 취미생활 - 영화

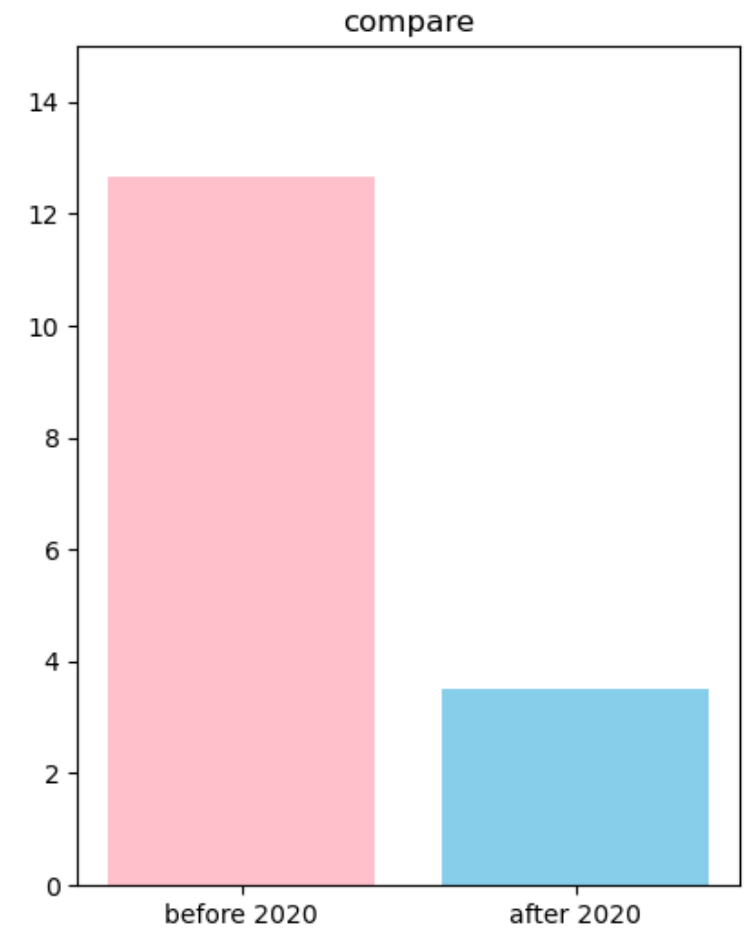
movie



2019년 영화 당
극장 관객 / 온라인 관객

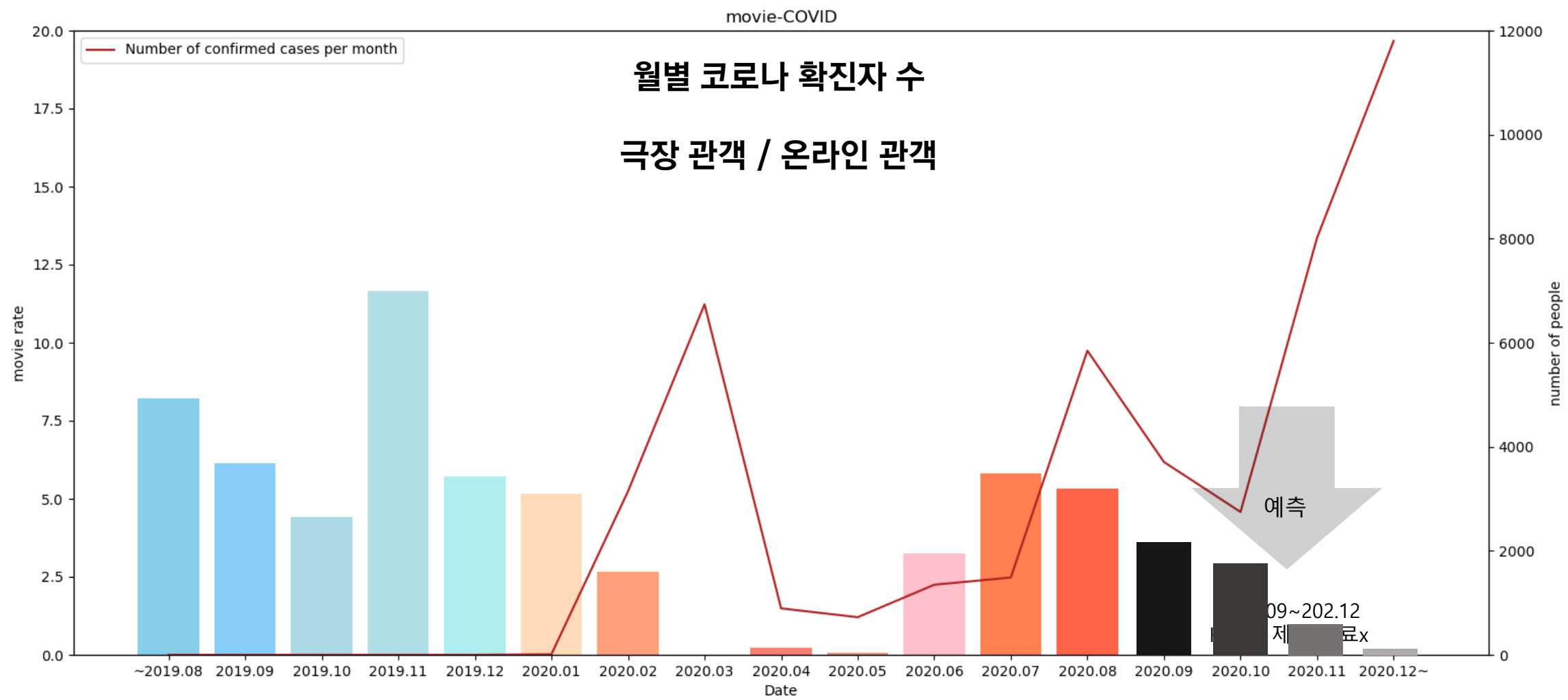


2020년 영화 당
극장 관객 / 온라인 관객



2019년 , 2020년 평균
극장 관객 / 온라인 관객

2. 취미생활 - 영화



2. 취미생활 - 영화

```
BASE_URL = 'https://www.instagram.com/'
SEARCH_URL = BASE_URL + 'explore/tags/'

options = webdriver.ChromeOptions()
#options.add_argument('headless')
options.add_argument('window-size=1920x1080')
options.add_argument("disable-gpu")
options.add_argument("disable-infobars")
options.add_argument("--disable-extensions")
options.add_argument("user-agent=Mozilla/5.0 (Macintosh; Intel Mac OS X 10_12_6) AppleWebKit/537.36 (KHTML, like Gecko)")
options.add_experimental_option("excludeSwitches", ["enable-logging"])
driver = webdriver.Chrome('chromedriver.exe', options=options)
driver.get(BASE_URL)
sleep(3)

def login():
    input_id = driver.find_elements_by_css_selector('input._2hvTZ.pexuQ.zyHYP')[0] #아이디 입력받는거 선택
    input_id.clear()
    input_id.send_keys(email)
    input_pw = driver.find_elements_by_css_selector('input._2hvTZ.pexuQ.zyHYP')[1] #비밀번호 입력받는거 선택
    input_pw.clear()
    input_pw.send_keys(pwd) #비밀번호 전송
    input_pw.submit() #로그인
    sleep(4) #로딩

def search(query):
    driver.get(SEARCH_URL + query) #query를 검색한 사이트로 이동
    sleep(4) #로

def crawl():
    srcList = [] #중복 방지 리스트

    open('result.txt', 'w')._result_file_clear

    for i in range(30000): #대충 스크롤 30000번 하겠다는 뜻
        imgList = driver.find_elements_by_css_selector('img.FFVAD') #find_img_by_selector

        dateList = ''

        for j in range(len(imgList)):
            alt = str(imgList[j].get_attribute('alt')) #날짜가 포함된 alt
```

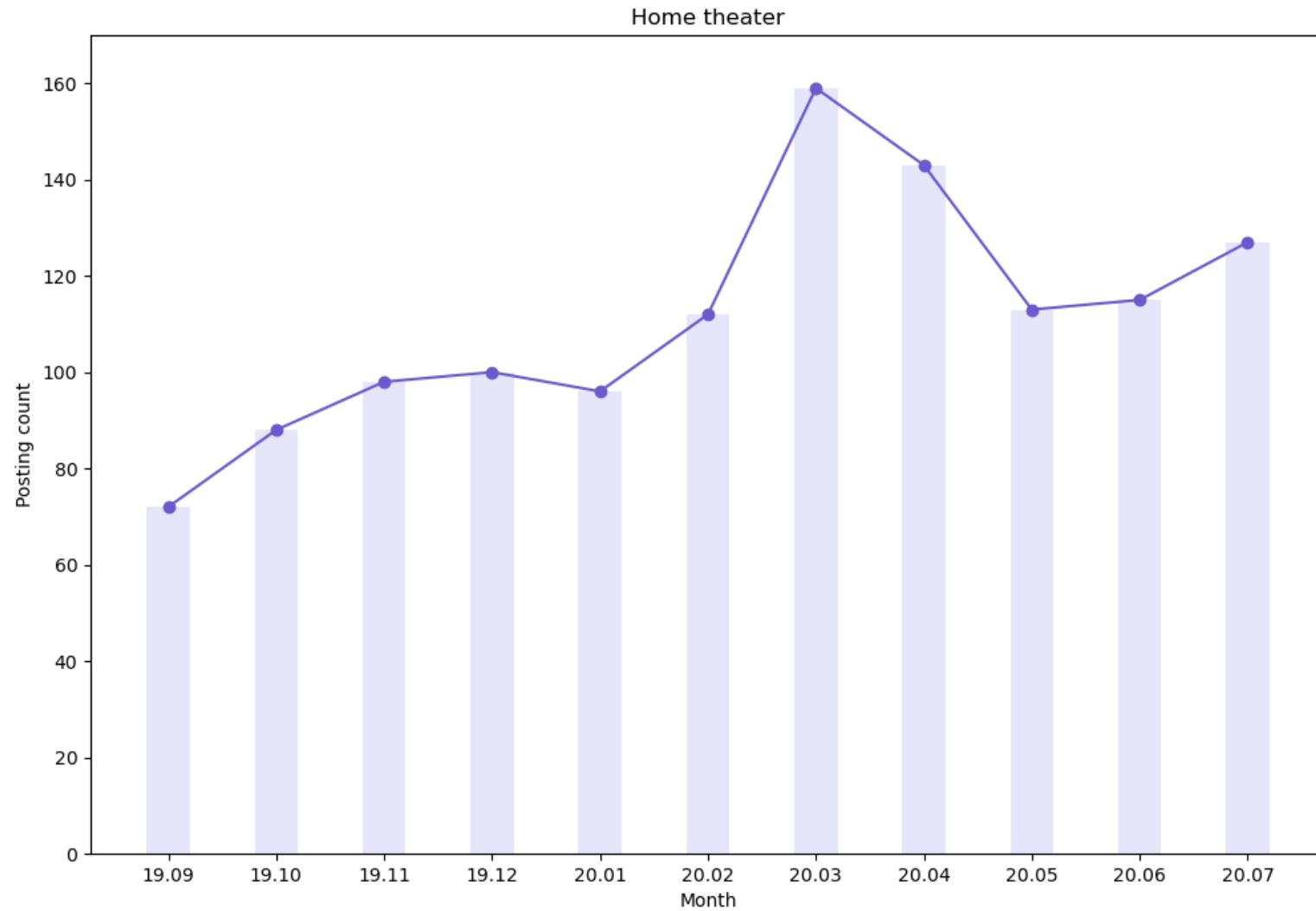
August 20, 2020
August 20, 2020
August 20, 2020
August 20, 2020
August 19, 2020
August 19, 2020
August 19, 2020
August 19, 2020
August 19, 2020
August 19, 2020
August 19, 2020
August 19, 2020
August 19, 2020
August 18, 2020
August 18, 2020
August 17, 2020
August 17, 2020
August 17, 2020
August 17, 2020
August 16, 2020
August 15, 2020
August 15, 2020
August 14, 2020
August 14, 2020
August 14, 2020
August 14, 2020
August 14, 2020

인스타그램

#홈시어터 해시태그

게시물 업로드 날짜 크롤링

2. 취미생활 - 영화



인스타그램

#홈시어터 해시태그

월별 게시물 업로드 수

INDEX

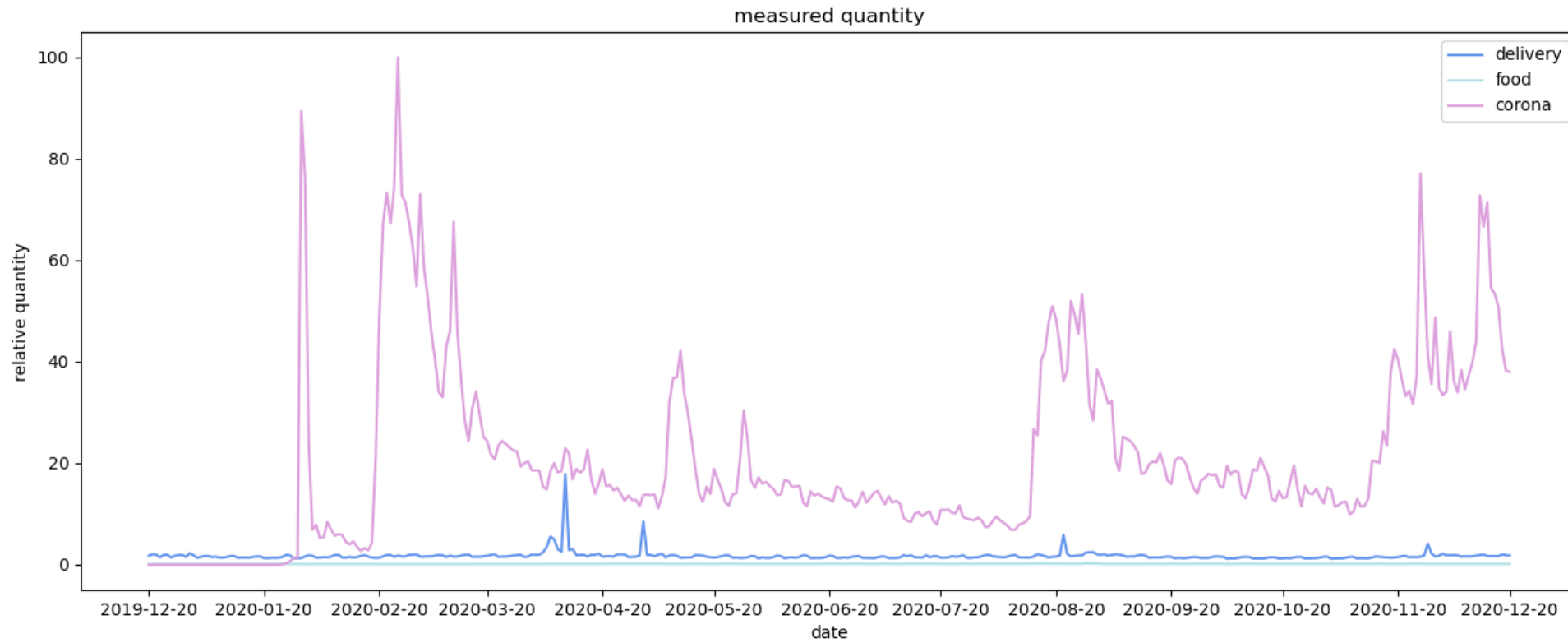
1. 주제에 관하여

2. 취미생활 - 영화

3. 음식소비 - 배달음식

4. 결론 도출

3. 음식소비 - 배달음식



네이버 데이터랩.csv
코로나 검색 수
배달음식 검색 수

3. 음식소비 - 배달음식

모름지기 먹스타그램 맛스타그램 인스타푸드 푸드스타그램 음식
slime clay eating sound clayart clayslime satisfying crushing clips
라면 라면스타그램 부대찌개 다이어트 요리스타그램 음식 먹방 술
1일1주인 ramen ajigen misoramen gourmet yummy 맛소스 맛소
먹스타그램 먹팔 맛팔 음식 존맛탱 언어유평판 선팔하듯맛팔 좋아이
랍스터 랍스타그램 랍스타그램 자축합스타그램 자축합스타그램 랍스타그램이
서울 seoul korea koreablog korean 대한민국 밤 दोराा 음식 food
foodbloggersofinstagram foodstagram berlinfoodie berlinfoodie
식사 한끼메뉴 건강식 간단식 간편요리 간편요리 요리스타그램 쿡
홍대 연세대 이대 명지대 마칭 배달의민족 배민라이더스 요기요홀
음식 냉장고파먹기 냉장고스토브 반찬 저녁 집밥 신혼집 저녁밥상 한
코של 샐러드 닭가슴살 다이어트 식단 따뜻한밥상 미역국 소고기미
음식 냉장고파먹기 냉장고스토브 반찬 저녁 집밥 신혼집 저녁밥상 한
음식 음식스타그램 먹스타그램 먹스타그램 먹스타그램 팔 맛집
맛집 음식 음식스타그램 와플 커피 먹수곳 서울맛집 여행 여행자

```

BASE_URL = 'https://www.instagram.com/'
SEARCH_URL = BASE_URL + 'explore/tags/'

options = webdriver.ChromeOptions()
#options.add_argument('headless')
options.add_argument('window-size=1920x1080')
options.add_argument("disable-gpu")
options.add_argument("disable-infobars")
options.add_argument("--disable-extensions")
options.add_argument('user-agent=Mozilla/5.0 (Macintosh; Intel Mac OS X 10_12_6) AppleWebKit/537.36 (KHTML, like Gecko)')
options.add_experimental_option("excludeSwitches", [{"enableLogging"}])
driver = webdriver.Chrome('chromedriver.exe', options=options)
driver.get(BASE_URL)
sleep(3)

def login():
    input_id = driver.find_elements_by_css_selector('input._2hvTZ.peguQ.zyHYP')[0]
    input_id.clear()
    input_id.send_keys(email)
    input_pw = driver.find_elements_by_css_selector('input._2hvTZ.peguQ.zyHYP')[1]
    input_pw.clear()
    input_pw.send_keys(pwd)
    input_pw.submit()
    sleep(4)

def search(query):
    driver.get(SEARCH_URL + query) #query를 검색할 사이트로 이동
    sleep(4) #로딩

def crawl():
    srcList = [] #중복 방지 리스트
    open('result.txt', 'w') #result file clear

    for i in range(30000): #대충 스크롤 30000번 하겠다는 뜻
        imgList = driver.find_elements_by_css_selector('img.FFVAD') #find img by sele
        dateList = ''

        for j in range(len(imgList)):
            alt = str(imgList[j].get_attribute('alt')) #알트값가 포함된 alt

def crawl():
    open('hashtag.txt', 'w')

    first = WebDriverWait(driver, EC.visibility_of_element_located, (By.CSS_SELECTOR, 'img.FFVAD'), 10)
    first.click()
    sleep(1)

    for i in range(5000):

        line = ''
        for j in range(10):

            content = WebDriverWait(driver, EC.visibility_of_element_located, (By.CSS_SELECTOR, 'img.FFVAD'), 10)
            tags = re.findall(r'#[a-z0-9]+', content)

```

[illegible]

인스타그램

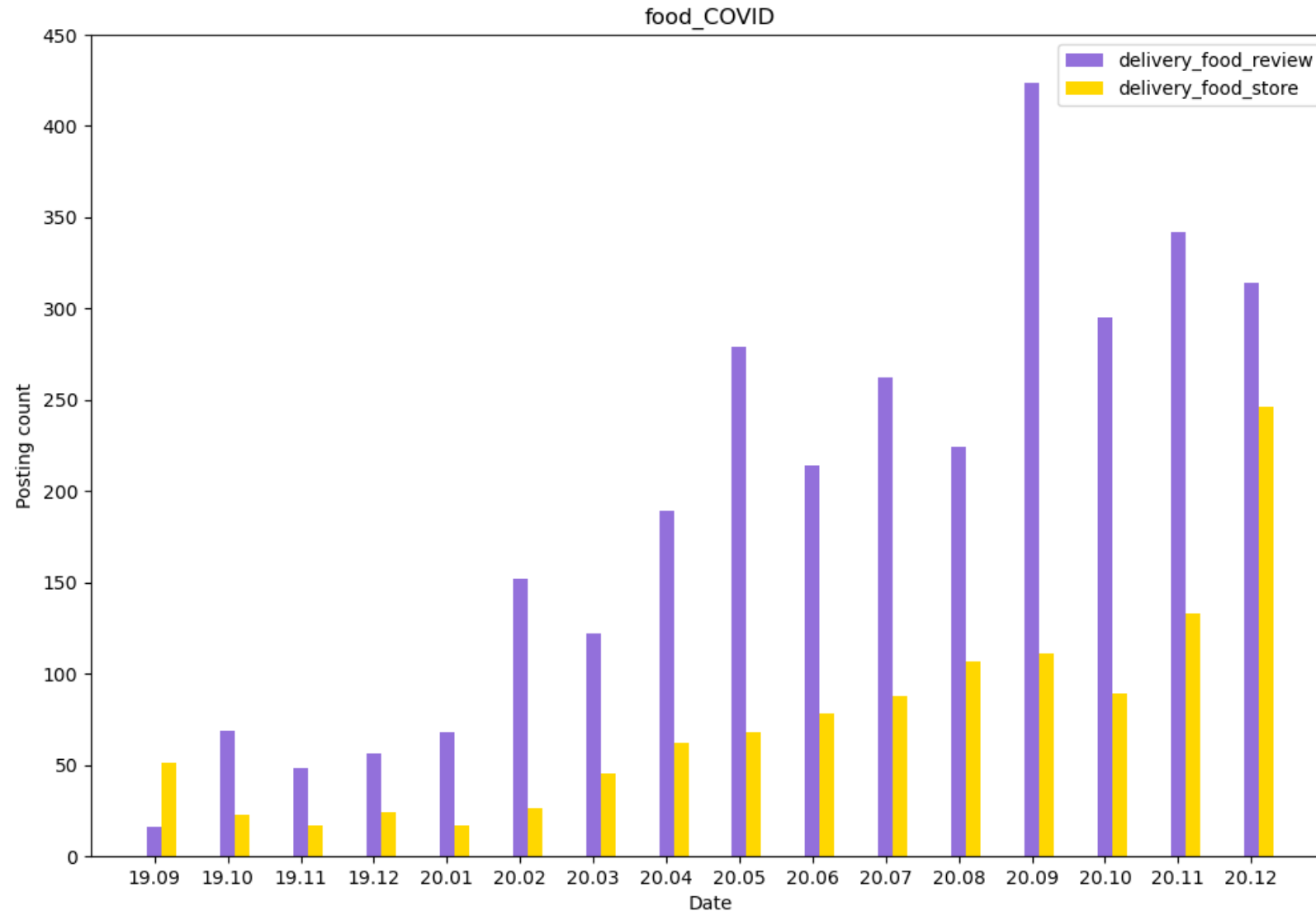
#배달음식맛집 해시태그

#배달의민족리뷰 해시태그

게시물 업로드 날짜 크롤링

게시물 해시태그 크롤링

3. 음식소비 - 배달음식



인스타그램

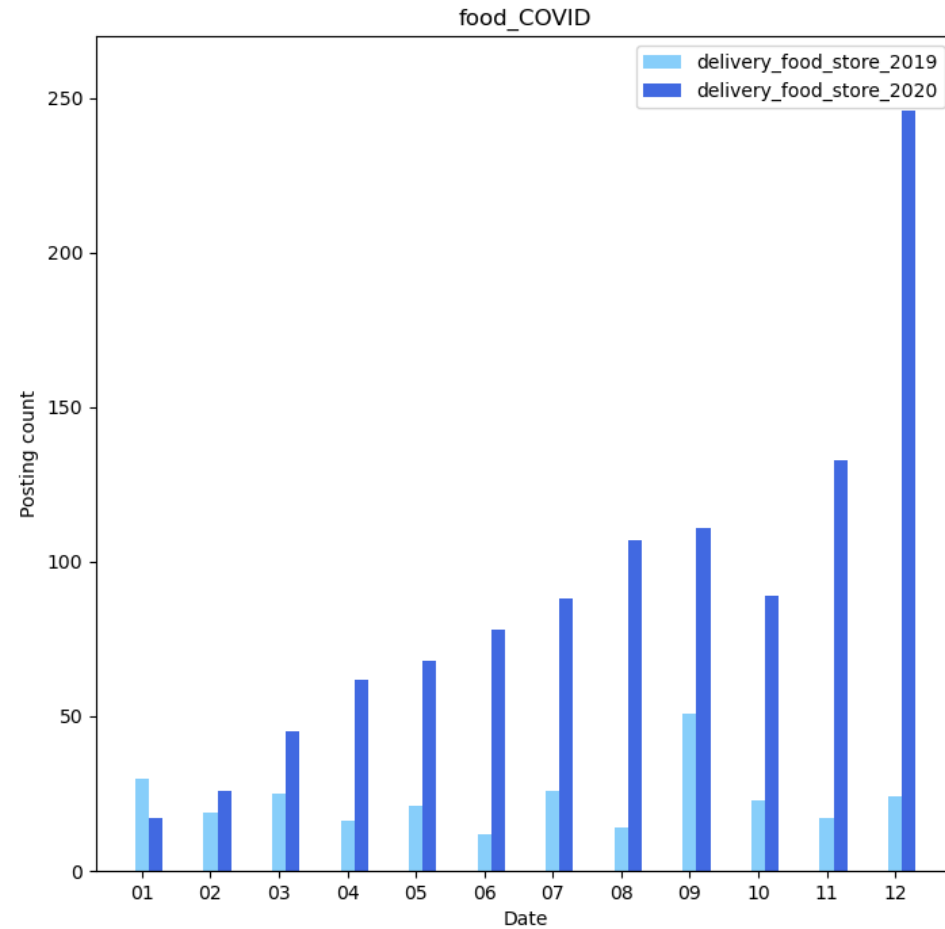
#배달음식맛집 해시태그

#배달의민족리뷰 해시태그

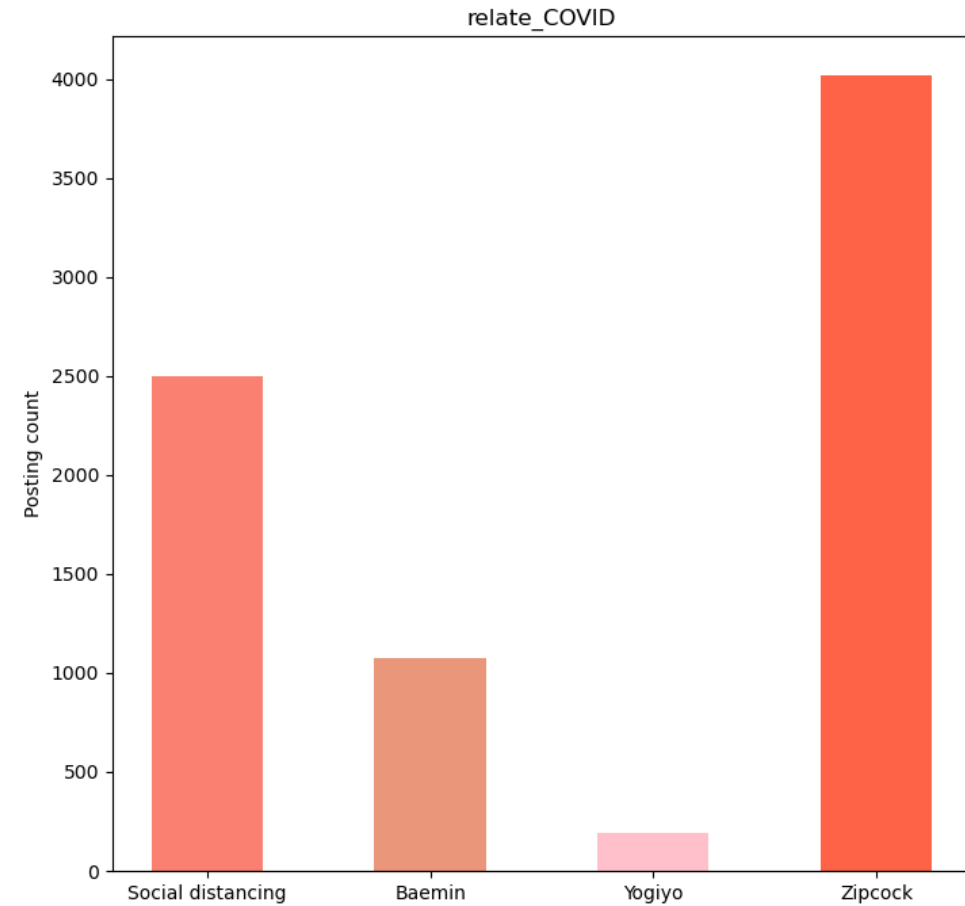
월별 게시물 수 비교

3. 음식소비 - 배달음식

Instagram



인스타그램
#배달음식맛집 해시태그
2019, 2020 비교



인스타그램
#사회적 거리두기 #요기요 #집콕 #배민
2020.12.22~20 약 3일간 게시물 개수

3. 음식소비 - 배달음식

```
import requests
from bs4 import BeautifulSoup

# 아래 주소가 메인페이지 내부에서 호출되는 실시간 검색어 데이터를 넘겨주는 주소
# requests.get("주소").json() 을 하면 데이터를 json 형태로 받아올 수 있습니다.
# 아래 주소를 직접 브라우저에서 접속해보시기 바랍니다.

json = requests.get('https://www.naver.com/srchrank?frm=main').json()

# json 데이터에서 "data" 항목의 값을 추출
ranks = json.get("data")

# 해당 값은 리스트 형태로 제공되기에 리스트만큼 반복
for r in ranks:
    # 각 데이터는 rank, keyword, keyword_synonyms
    rank = r.get("rank")
    keyword = r.get("keyword")
    print(rank, keyword)

headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, li
html = requests.get('https://datalab.naver.com/keyword/realtimelist.naver?where=main', headers=h

parsed = BeautifulSoup(html.content, 'html.parser')
data1 = parsed.findAll('span', 'item_title')

for item in data1:
    print(item.get_text())
```

- 1 김예령
- 2 윤석민
- 3 송영규
- 4 1g전자우
- 5 정경심
- 6 2021 수능 만점자
- 7 nba
- 8 신한대학교
- 9 수능 성적 발표
- 10 한국교육과정평가원
- 11 비긴어게인
- 12 신박한 정리 송영규
- 13 펜트하우스 17회
- 14 석경에이티
- 15 지놈앤컴퍼니
- 16 문희준
- 17 5인이상 집합금지 식당
- 18 인터넷시몬선교센터
- 19 정용진
- 20 나인테크

네이버 실시간 검색어
웹 크롤링

INDEX

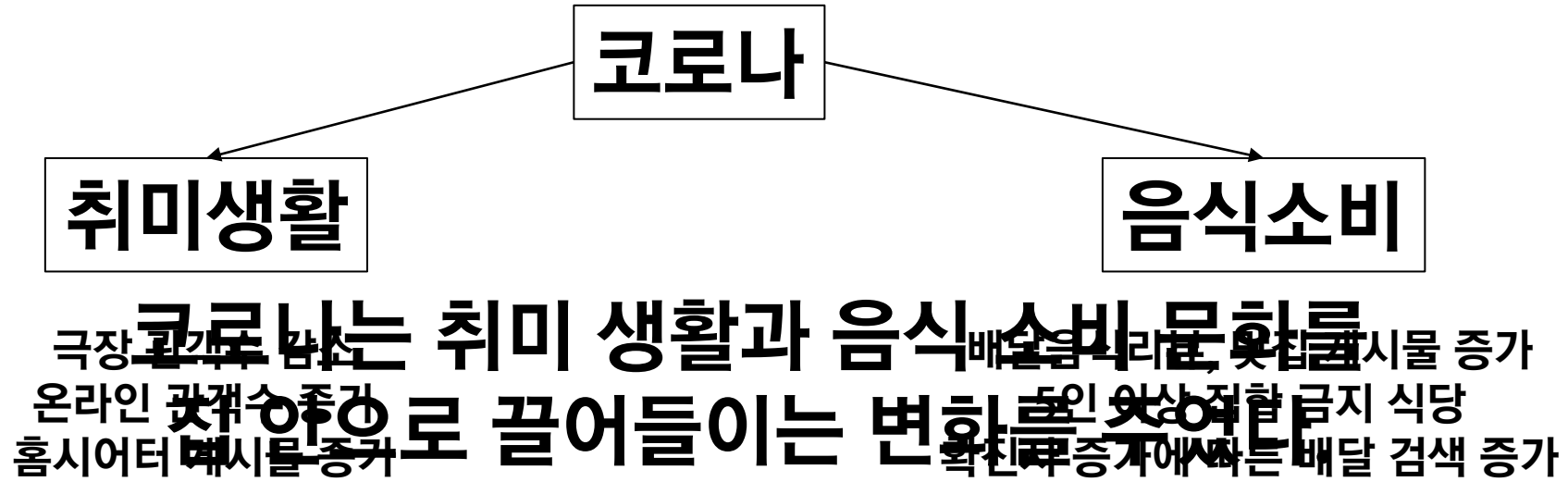
1. 주제에 관하여

2. 취미생활 - 영화

3. 음식소비 - 배달음식

4. 결론 도출

4. 결론 도출



⇒ 코로나는 외부 취미 생활도 내부 취미 생활 형태로 바꿨다.

⇒ 코로나는 배달 음식 섭취 증가로 이끌었고 음식점 운영에도 영향을 주었다.

팀원끼리 모르는 팀



김수빈 임성묵 손봉우