

## **2020 Crawler DataTon**

### **<웹 크롤링을 이용한 코로나가 취미생활과 음식소비 문화에 미친 영향 분석>**

**팀 이름 : 팀원끼리 서로 모르는 팀**

팀원 및 제출자:

202002543 임성묵

202002471 김수빈

202002506 손봉우

제출일 : 2020.12.23

## 목차

- I. 주제설명
- II. 취미생활(키워드) - 코로나가 변화시킨 영화 소비문화와 취미생활
- III. 음식(키워드) - 코로나가 변화시킨 음식 소비문화
- IV. 워드 클라우드(#코로나, #취미, #음식)
- V. 결론 도출

## I. 주제설명

코로나19로 인해 대한민국의 취미생활과 음식 소비 문화에 미친 영향이 무엇인지 크롤링을 이용해서 알아보고자 하였다.

특히 취미생활 키워드와 관련해 코로나가 영화 소비 문화에 미친 영향과 음식 키워드와 관련해 코로나가 음식 소비 문화에 미친 영향을 중점적으로 분석하였다.

## II. 취미생활(키워드) - 코로나가 변화시킨 영화 소비문화와 취미생활

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	온라인 이용관수(월간)		
1		오케이 마담 (OK! MADAM)	2020-08-12	한국	코미디	이철하	엄정화, 박성웅, 이상윤, 배...	100분 28초	214,364		
2		다만 악에서 구하소서 (DELIVER US FROM EVIL)	2020-08-05	한국	범죄	홍원찬	황정민, 이정재, 박정민, 박...	107분 49초	159,605		
3		국제수사 (The Golden Holiday)	2020-09-29	한국	코미디	김봉한	곽도원, 강대명, 김희원, 김...	105분 50초	154,818		
4		오! 나의 (Oh! My Gran)	2020-09-02	한국	코미디 코미디	정세교	나문희, 이희준, 김예은	109분 30초	112,212		
5		무란 (Mulan)	2020-09-17	미국	어드벤처	니키 카로	유역비, 이연걸, 공리, 견자...	115분 9초	106,495		
6		달보 (Pawn)	2020-09-29	한국	드라마	강대규	성동일, 학지관, 김희원, 백...	113분 0초	100,740		
7		반도 (Peninsula)	2020-07-15	한국	액션	연상호	강동원, 이정현	115분 35초	62,998		
8		아웃포스트 (The Outpost)	2020-09-23	미국	전쟁	로드 무리	스콧 이스트우드, 케일럽 ...	123분 22초	56,665		

<크롤링 타겟 사이트:

KOBIS 영화관 입장권 통합전산망; 2019, 2020 기준 박스오피스(1위~25위)>

XML 보건복지부\_코로나19 감염\_현황

활용신청

오류신고 및  
담당자 문의

23
1
관심

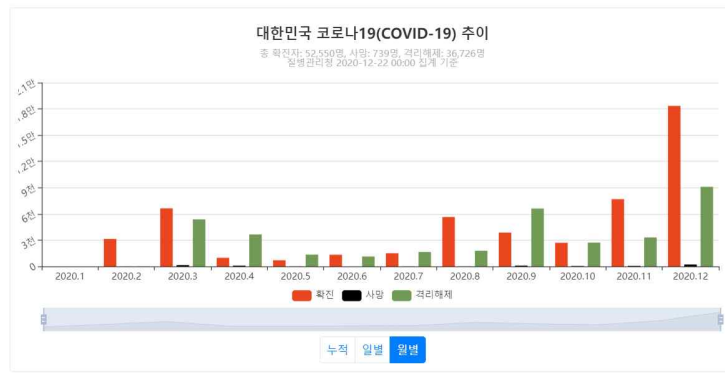
코로나19 감염증으로 인한 일별 확진자, 완치자, 치료중인 환자, 사망자 등에 대한 현황자료(이 제공자료는 관련 발생 상황에 대한 정보를 신속 투명하게 공개하기 위한 것으로, 추가적인 역학조사 결과 등에 따라 수정 및 보완될 수 있음을 알려드립니다.)

### OpenAPI 정보

분류체계	보건 - 보건의료	제공기관	공공데이터활용지원센터
관리부서명		관리부서 전화번호	02-1566-0025
API 유형	REST	데이터포맷	XML
활용신청	3806	키워드	코로나, 호흡기질환, 건강
등록	2020-04-15	수정	2020-12-10
심의유형	개발단계 : 허용 / 운영단계 : 허용		
비용부과유무	무료		
이용허락범위	이용허락범위 제한 없음		
참고문서	01. 보건복지부_OpenAPI 활용가이드_코로나바이러스 감염증_감염현황 조회 서비스_v0.3_(11).docx		

<추가 활용데이터 출처 : Open API; 보건복지부\_코로나19 감염\_현황>

## 국내차트KR



<추가 활용데이터 출처 : [코로나보드](#)>

```
import pandas as pd
import matplotlib.pyplot as plt
# -*-coding:utf-8 -*-
from bs4 import BeautifulSoup
import matplotlib.pyplot as plt
import requests
import sys
import io

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(",","")), 2)

movie_data_ob = [] # movie_data를 저장할 객체 리스트
before = [] # 2020년 전 영화
after = [] # 2020년 이후 영화
#월별로 나눠서 저장할 객체 리스트
y19_m8 = []
y19_m9 = []
y19_m10 = []
y19_m11 = []
y19_m12 = []
y20_m1 = []
y20_m2 = []
y20_m3 = []
y20_m4 = []
y20_m5 = []
y20_m6 = []
y20_m7 = []
```

```

y20_m8 = []
etc = []
# 2020년 기준 조회 1~25위 영화 크롤링
url = 'http://www.kobis.or.kr/kobis/business/stat/online/onlineYearlyBoxRank.do?CSRFToken=fMV3gtByZkRCnU8eIcf'
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)")
#2020년 기준 조회 영화 추가
for i in range(len(title_list)):
    marker = 0
    for j in movie_data_ob:
        if j.title == title_list[i]:
            marker = 1
    if marker == 0 and online_list[i] != 0:
        m1 = movie_data(title_list[i].text, date_list[i].text, offline_list[i].text, online_list[i].text)
        movie_data_ob.append(m1)

# 2019년 기준 조회 1~25위 영화 크롤링
url = 'http://www.kobis.or.kr/kobis/business/stat/online/onlineYearlyBoxRank.do?CSRFToken=fMV3gtByZkRCnU8eIcf'
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)")
#2019년 기준 조회 영화 추가
for i in range(len(title_list)):
    marker = 0
    for j in movie_data_ob:
        if j.title == title_list[i]:
            marker = 1
    if marker == 0 and online_list[i] != 0:
        m1 = movie_data(title_list[i].text, date_list[i].text, offline_list[i].text, online_list[i].text)
        movie_data_ob.append(m1)

```

## <작성한 주요 코드>

### 과정

1. BeautifulSoup, requests 라이브러리를 이용해서 개봉일, 극장누적 관객 수, 온라인 이용건수를 크롤링하였다.
2. 개봉일을 기준으로 월별로 데이터를 분류하였다.(2019년 8월 ~ 2019년 8월 까지 총 12개월)
3.  $movie\_rate = offline / online$   
(offline = 극장누적 관객 수, online = 온라인 이용건수) 온라인 이용건수대비 극장누적 관객 수를 알려주는 "movie\_rate" 라는 지표를 만들었다.  
**#movie\_rate가 0에 가까울수록 극장누적 관객 수 대비 온라인 이용 건수가 많다는 것을 의미한다.**

## <월별 평균 movie\_rate>

```
...
result_ratio = []
result_ratio.append(avg_ratio_y19_m8)
result_ratio.append(avg_ratio_y19_m9)
result_ratio.append(avg_ratio_y19_m10)
result_ratio.append(avg_ratio_y19_m11)
result_ratio.append(avg_ratio_y19_m12)
result_ratio.append(avg_ratio_y20_m1)
result_ratio.append(avg_ratio_y20_m2)
result_ratio.append(avg_ratio_y20_m3)
result_ratio.append(avg_ratio_y20_m4)
result_ratio.append(avg_ratio_y20_m5)
result_ratio.append(avg_ratio_y20_m6)
result_ratio.append(avg_ratio_y20_m7)
result_ratio.append(avg_ratio_y20_m8)

df1 = pd.read_excel('./total_corona_count.xlsx', usecols="A:F")
case = df1['일일확진자수'].tolist()

c_03 = 0
c_04 = 0
c_05 = 0
c_06 = 0
c_07 = 0
c_08 = 0
c_09 = 0
c_10 = 0
c_11 = 0
c_12 = 0

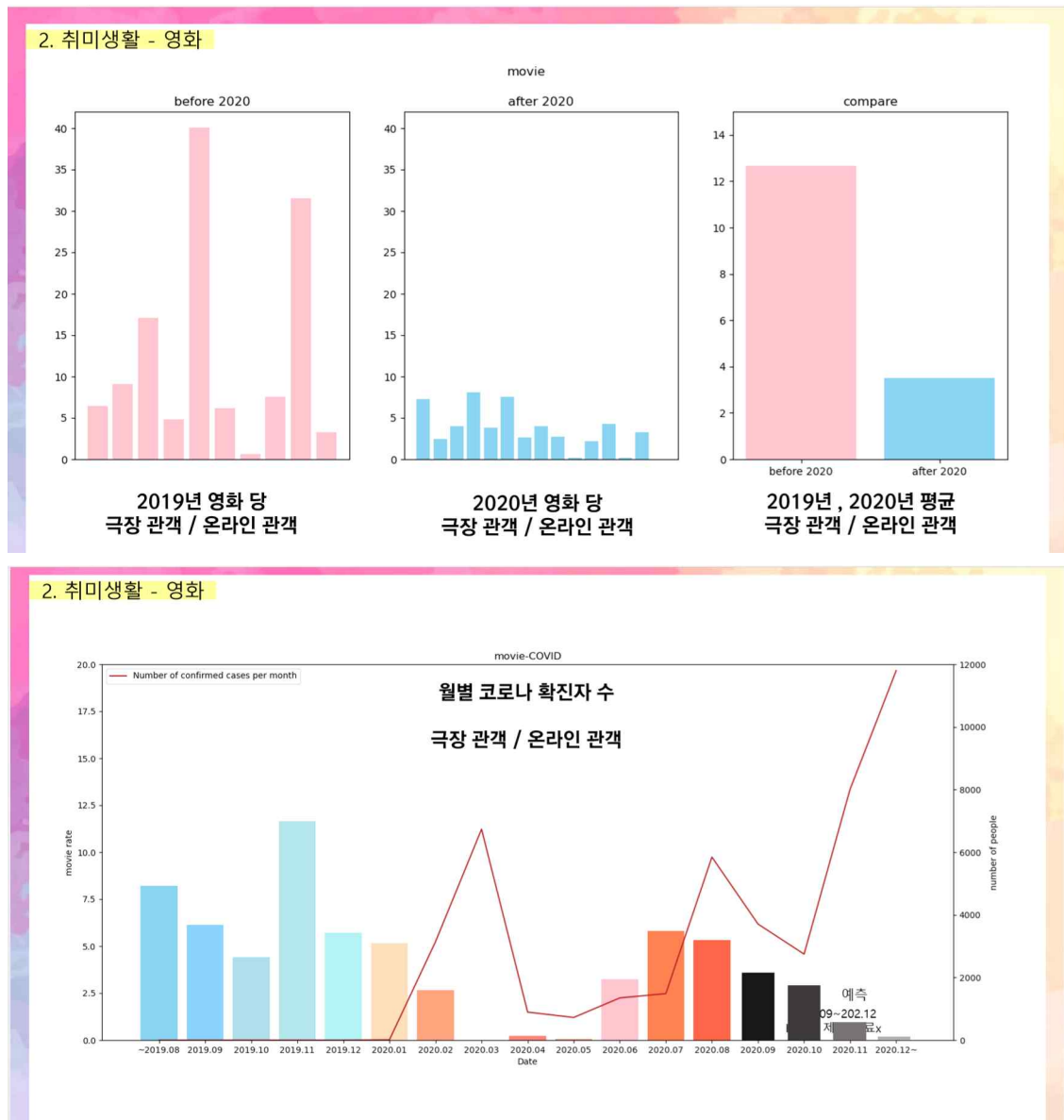
for i in range(len(case)):
    if(i<=26):
        c_03 += case[i]
    elif(i<=57):

C:\Users\tjdan\PycharmProjects\exer\venv\Scripts\python.exe C:/Users/tjdan/PycharmProjects/exer/exer_crawler.py
2019.8 평균 movie_rate : 8.22
2019.9 평균 movie_rate : 6.13
2019.10 평균 movie_rate : 4.41
2019.11 평균 movie_rate : 11.64
2019.12 평균 movie_rate : 5.72
2020.1 평균 movie_rate : 5.17
2020.2 평균 movie_rate : 2.65
2020.3 평균 movie_rate : 0
2020.4 평균 movie_rate : 0.22
2020.5 평균 movie_rate : 0.08
2020.6 평균 movie_rate : 3.25
2020.7 평균 movie_rate : 5.81
2020.8 평균 movie_rate : 5.33
2019.8월 전 기타 평균 movie_rate : 8.22
2020년 전 영화 평균 movie_rate : 10.16
2020년 이후 영화 movie_rate : 3.51

Process finished with exit code 0
```

4. 보건복지부\_코로나19감염\_현황, 코로나보드로부터 코로나19 월별 확진자 수 데이터를 이용하였다.

5. matplotlib, pandas, numpy 라이브러리를 이용해서 위의 모아놓은 데이터들을 여러 기준으로 시각화한 자료를 만들었다.



<월별 코로나 확진자 수와, movie\_rate 그래프>

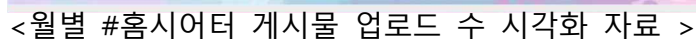
\*선 그래프 : 코로나 확진자 수, 막대 그래프 : movie\_rate(낮을수록 온라인 관객 증가)

- 위의 시각화한 자료들을 보면 알 수 있듯이, 코로나 이전과 이후의 movie\_rate 평균이 3배 이상 차이가 난다. 또한, 코로나 확진자 수가 늘어나면서 뒤따라 movie\_rate도 낮아짐을 알 수 있다. 따라서 확진자 수와 movie\_rate가 상관관계가 있을 가능성을 추론해볼 수 있다.



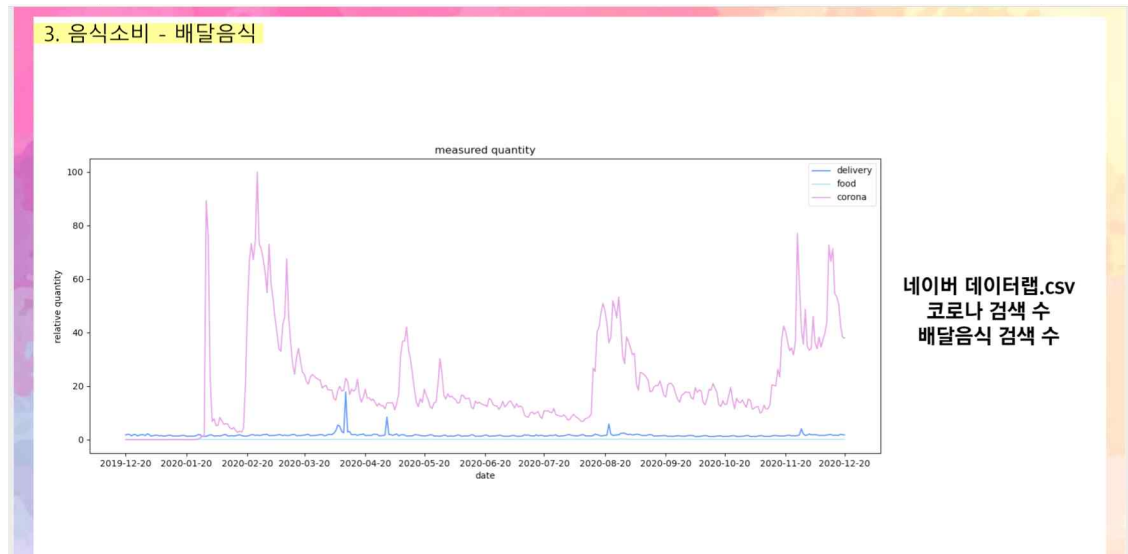
- [illegible]

10. 월별 홈시어터 게시물 업로드 수 시각화 자료를 만들었다.



### III. 음식(키워드) - 코로나가 변화시킨 음식 소비문화

1. 네이버 데이터 랩.csv 파일을 가져와서 pandas, numpy, matplotlib library를 이용해서 코로나와 배달음식 검색수를 그래프로 시각화 하였다.



<위의 네이버 데이터 랩 csv 파일로 시각화한 코로나,배달음식 검색 수 비교 그래프>

\*핑크색 그래프: 코로나 검색 수, 파란색 그래프 : 배달음식 검색 수

-검색량의 양이 상대적으로 차이가 크게 나지만 코로나 검색어 수 그래프를 전체적으로 뒤따라가는 형태를 배달음식 검색 수 그래프가 보이고 있다.

2. 인스타그램에서 #배달음식맛집, #배달의민족리뷰 해시태그 게시물의 날짜를 크롤링했다.(selenium, getpass library이용)

<인스타그램의 해당 해시태그 게시물 업로드 날짜 크롤링과정>

3. 위의 인스타그램 게시물 날짜 데이터를 이용해서 시각화 하였다.

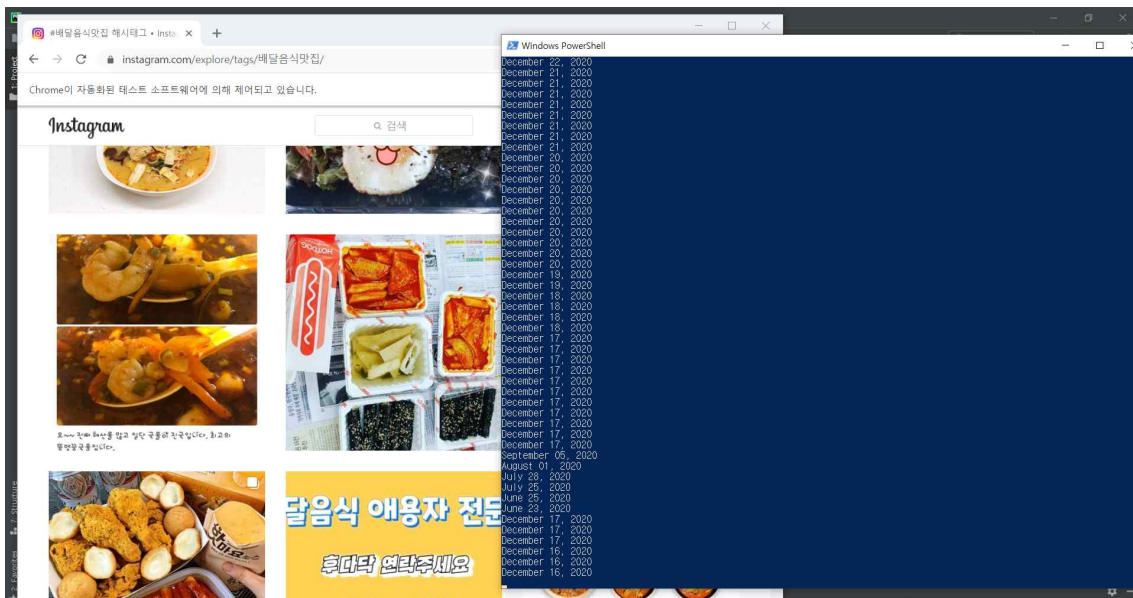
(matplotlib, numpy, pandas library 이용)

<월별 #배달음식맛집, #배달의민족리뷰 업로드 게시물 수 그래프>

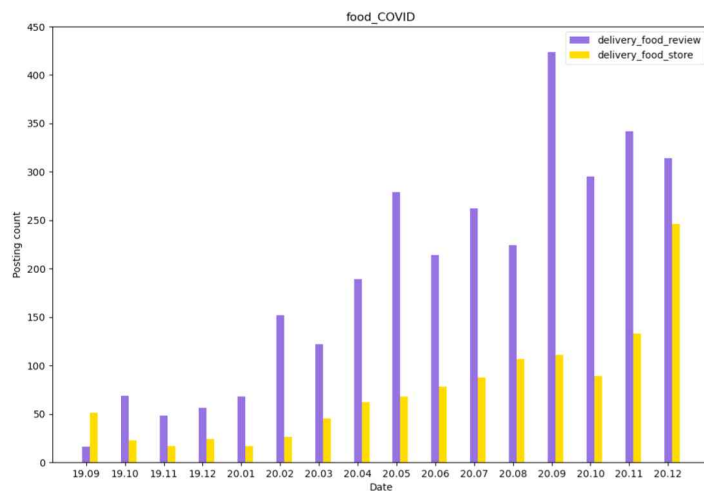
\*보라색 막대 그래프 : 월별 #배달의민족 리뷰 게시물 수,

노란색 막대 그래프 : 월별 #배달음식맛집 게시물 수

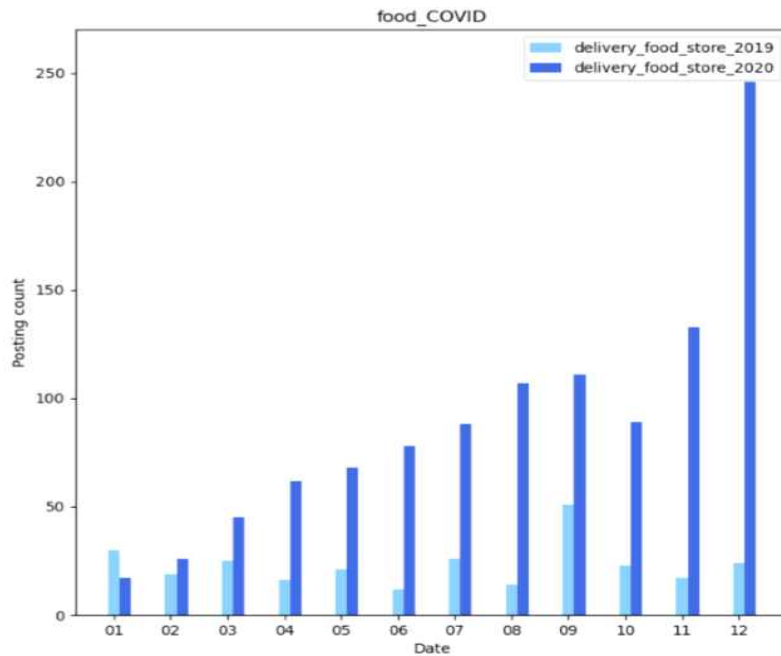
-코로나19가 본격적으로 확산하기 시작한 2020년 1월을 기점으로 두 가지 "배달"을 키워드로 한 게시물의 업로드 수가 증가하고 있음을 알 수 있다.



### 3. 음식소비 - 배달음식



인스타그램  
#배달음식맛집 해시태그  
#배달의민족리뷰 해시태그  
월별 게시물 수 비교



### 인스타그램 #배달음식맛집 해시태그 2019, 2020 비교

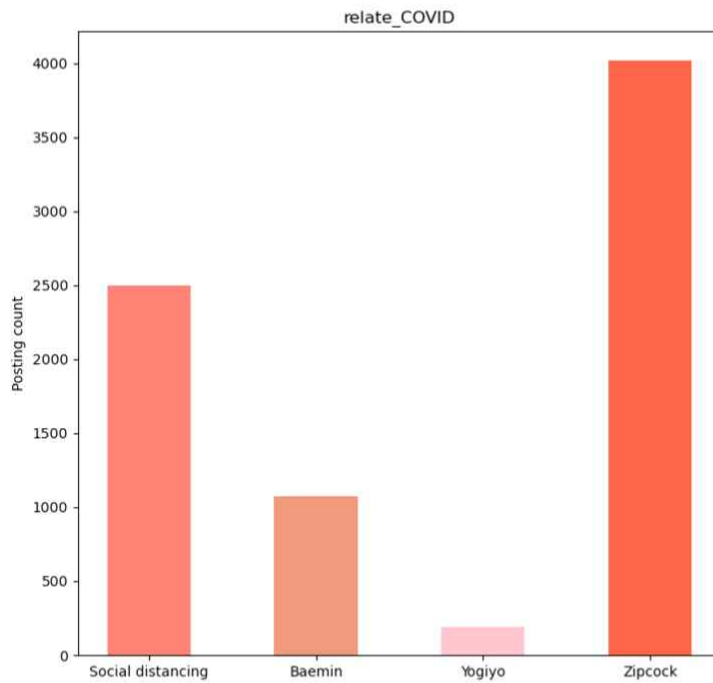
4. 다른 해시태그도 크롤링하여 시각화 하였다.

<2019,2020 월별 #배달음식맛집 게시물 업로드 수 그래프>

\* 하늘색 막대 그래프 : 2019년 기준 #배달음식맛집 게시물 수,

파란색 막대 그래프 : 2020년 기준 #배달음식맛집 게시물 수

-2019년과 달리 2020년에는 코로나의 확산세가 이어짐에 따라 #배달음식맛집 게시물의 수도 증가함을 알 수 있다.



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

<2020.12.22. ~ 21 기준 #사회적거리두기, #배민, #요기요, #집콕 게시물 수 그래프>

- \* 왼쪽 막대 그래프부터 #사회적거리두기, #배민, #요기요, #집콕 게시물 수
- 3일간 #사회적거리두기, #배민, #요기요, #집콕 게시물 수가 많음을 알 수 있다.

## 5. 네이버 실시간 검색어를 크롤링해서 특정 키워드가 있는 검색어를 추출

```
#-*-coding:utf-8-*-
from bs4 import BeautifulSoup
import requests
import sys
import io
#
sys.stdout = io.TextIOWrapper(sys.stdout.detach(), encoding='utf-8')
sys.stderr = io.TextIOWrapper(sys.stderr.detach(), encoding='utf-8')

url = 'https://search.naver.com/search.naver?where=nexearch&sm=top_hyt&fbm=1&ie=utf8&query=%EC%8B%A4%EA%B2%80'
res = requests.get(url)
html = res.text
soup = BeautifulSoup(html, 'html.parser')
k = soup.select("#sub_pack > section.sc_new.sp_nkeyword.type_news > div > div.hot_keyword.srch.aside.news_tab")
now = []
for i in k:
    now.append(i.text)
for i in now:
    if "식당" in i:
        print(i)
    elif "코로나" in i:
        print(i)
    elif "배달" in i:
        print(i)
    elif "집콕" in i:
        print(i)
    elif "홈시어터" in i:
        print(i)
    elif "맛집" in i:
        print(i)
    elif "거리두기" in i:
        print(i)
    elif "확진" in i:
        print(i)
```

- 1 김예형
- 2 윤석민
- 3 송영규
- 4 lg전자우
- 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 지능앤컴퍼니
- 16 문희준
- 17 5인이상 집합금지 식당
- 18 인터넷시온선교센터
- 19 정용진
- 20 나인테크

### <네이버 크롤링 코드>

-5인이상 집합금지가 발표되고 "5인이상 집합금지 식당" 실시간 검색어에 오르면서 사람들이 음식 소비 형태에 대해 관심과 걱정이 증가했다고 추론할 수 있다.

#### IV. 워드 클라우드(#코로나, #취미, #음식)



주제  
웹 크롤링을 이용한 코로나가  
취미생활과 음식소비 문화에  
미친 영향 분석

코로나와 취미, 음식 세 가지 키워드를 중심으로 워드 클라우드를 하였다. 세 가지 키워드와 관련된 것들을 한눈에 알아 볼 수 있다.

## V. 결론 도출

취미생활 부분에서 코로나19 발생 시기와 맞물려서 극장 관객수 감소, 온라인 관객 수 증가, #홈시어터 게시물 증가라는 결과로 이어졌다. 또한 코로나 확진자수 그래프와 영화 관람객 수 비율 그래프가 비슷한 형태를 띄었고 이는 코로나가 영화 소비 문화에 많은 영향을 주었다고 볼 수 있다. 특히 외부 취미생활 중 하나인 영화 관람을 내부 취미 생활 형태로 변하는데 많은 영향을 주었다.

음식소비 부분에서도 마찬가지로 코로나19 발생 이후 코로나의 검색 수와 배달음식 검색 수 그래프가 비슷한 형태를 보여 상관관계를 유추할 수 있었다. 또한 인스타그램에서 #배달음식맛집 게시물수 월별 그래프를 보면, 뚜렷한 변화가 없는 2019년과는 다르게 코로나 이후인 2020년에는 꾸준히 우상향하는 그래프를 볼 수 있었다. 또한 다른 해시태그 게시물 데이터를 이용해 나타낸 지표도 배달음식에 대한 관심의 증가를 뒷받침하였다.

음식소비 부분에서도 역시 코로나19가 사람들의 배달음식 섭취 증가를 이끌었다고 볼 수 있다. 음식소비도 전체적으로 외부에서 내부로 변화하도록 코로나가 영향을 미쳤다고 생각한다.

**따라서 코로나19는 취미생활과 음식소비 문화를 집안으로 끌어들이는 변화를 주었다.**



**발표 영상 Link:**

<https://www.youtube.com/watch?v=RxBiaD8EzU>

**출처 모음:**

**1. 사이트**

\* < [KOBIS영화관입장권통합전산망](#) >:

<http://www.kobis.or.kr/kobis/business/stat/online/onlineYearlyBoxRank.do?CSRFToken=Z7fUb58TtbipiegE89UbK29aHnINhIAfpcgVI2WpbY&curPage=1&loadEnd=0&searchType=search&sSearchYearFrom=2020&sMultiMovieYn=>

\* < Open API; [보건복지부\\_코로나19감염\\_현황](#) >:

<https://www.data.go.kr/tcs/dss/selectApiDataDetailView.do?publicDataPk=15043376>

\* < [코로나보드](#) >:

<https://coronaboard.kr/>

\* < 네이버 데이터 랩 >:

<https://datalab.naver.com/>

\* < 인스타그램 >:

<https://www.instagram.com/>

**2. 파이썬 라이브러리**

numpy, matplotlib, pandas, BeautifulSoup4, requests, getpass, selenium