

데이터 분석/시각화

# 청와대 국민청원글 키워드 분석

## 3조 - 열정

22.05.02

팀 장 : 강진오

부팀장 : 김미현

# 청와대 국민청원글 키워드 분석

## 가설

국민청원 키워드의 상위 30개 중  
‘아파트’라는 키워드는  
광주의 아파트 붕괴 사고와 관련된 국민청원일 것이다.

# 목차

#INDEX

## STEP1

환경설정

환경설정  
(패키지 설치 등)  
(KoNLP..)

## STEP2

분석 데이터 확보

데이터 생성

## STEP3

Data 전처리

명사추출  
(extractNoun)

불용어 처리  
(gsub 등)

단어별 빈도수세기  
(table)

## STEP4

시각화

시각화  
(wordcloud)

## STEP5

가설확인

가설확인

# 환경설정

## # 01

```
!apt-get update -qq # 설치를 업데이트
!apt-get install fonts-nanum* -qq # 설치한다. fonts-nanum*

import matplotlib.font_manager as fm # 폰트 관련 용도
import matplotlib.pyplot as plt # 그래프 그리는 용도

path = '/usr/share/fonts/truetype/nanum/NanumGothicEco.ttf' # 나눔고딕 폰트 설치 위치
font_name = fm.FontProperties(fname=path, size=10).get_name()
print(font_name)
plt.rc('font', family=font_name)

# 우선 fm._rebuild() 를 해주고 # 폰트 매니저 재빌드가 필요하다.
fm._rebuild()

%matplotlib inline
import matplotlib as mpl # 기본 설정 만지는 용도
import matplotlib.pyplot as plt # 그래프 그리는 용도
import matplotlib.font_manager as fm # 폰트 관련 용도
import numpy as np

path = '/usr/share/fonts/truetype/nanum/NanumGothicEco.ttf' # 나눔고딕 폰트 설치 위치
font_name = fm.FontProperties(fname=path, size=10).get_name()
print(font_name)
plt.rc('font', family=font_name)
```

# 환경설정

## # 02

```
## 음수 표시되도록 설정
mpl.rcParams['axes.unicode_minus'] = False

# 데이터 준비
data = np.random.randint(-200, 100, 50).cumsum()

# 그래프를 그려 한글 확인
plt.plot(range(50), data, 'r')
plt.title('시간별 가격 추이')

# konlpy 설치
pip install konlpy

# 모듈 불러오기
import nltk
import konlpy

from konlpy.tag import Kkma
from konlpy.tag import Hannanum
```

# 분석 데이터 확보

## # 03

```
# 01. 파일 읽기
# 02. 한글 데이터 처리 엔진을 이용해서 konlpy로 명사 추출
# 03. nltk Text 객체 만들기
# - 그래프 그려보기, 분포 확인, 단어 빈도수 확인 ...
# 04. 워드 클라우드
```

```
petition = open("petition.txt").read()
petition[1:1000]
```

청와대 <국민청원> 게시글 제목 Data 확보

# Data 전처리

## # 04

```
# 02. 한글 데이터 처리 엔진을 이용해서 konlpy로 명사 추출
# t = Kkma()
han = Hannanum()
```

```
pet_nouns = han.nouns(petition)
```

```
pet_nouns[0:20], len(pet_nouns)
```

```
# 03. nltk Text 객체 만들기
k = nltk.Text(pet_nouns, name='국민청원')
print(type(k))
print(len(k.tokens))
```

청와대 <국민청원> 게시글 제목 Data 에서 명사 추출

[('축구', 14),  
('반대', 13),  
('요청', 11),  
('고발', 11),  
('처벌', 11),  
('수사', 10),  
('조사', 9),  
('억울', 8),  
('아버지', 8),  
('검사', 8),  
('청와대', 7),  
('병원', 7),  
('저희', 7),  
('사망', 6),  
('허가', 6),  
('운영', 6),  
('요구', 6),  
('대한민국', 6),  
('백신', 6),  
('저', 6),  
('검수완박', 5),  
('설치', 5),  
('것', 5),

('저', 6),  
('검수완박', 5),  
('설치', 5),  
('것', 5),  
('2차', 5),  
('관련', 5),  
('국민', 5),  
('강력처벌', 5),  
('폭행', 5),  
('허용', 5),  
('호소', 4),  
('2022년', 4),  
('필요', 4),  
('개정', 4),  
('피해자', 4),  
('철포', 4),  
('국회', 4),  
('갑질', 4),  
('신상공개', 4),  
('아이들', 4),  
('사건', 4),  
('폐', 4),  
('완화', 4),

('국회', 4),  
('갑질', 4),  
('신상공개', 4),  
('아이들', 4),  
('사건', 4),  
('폐', 4),  
('완화', 4),  
('검찰총장', 4),  
('전', 4),  
('법안', 4),  
('윤석열', 4),  
('장애', 4),  
('내', 4),  
('검찰', 4),  
('남편', 4),  
('철회', 4),  
('요양병원', 3),  
('입법', 3),  
('조치', 3),  
('사고', 3),  
('여러분', 3),  
('행위', 3),  
('소상공', 3),

# Data 전처리

# 05

```
# 불용어 사전
stop_words = ['청원', '수', '코', '의', '주세요', '후']

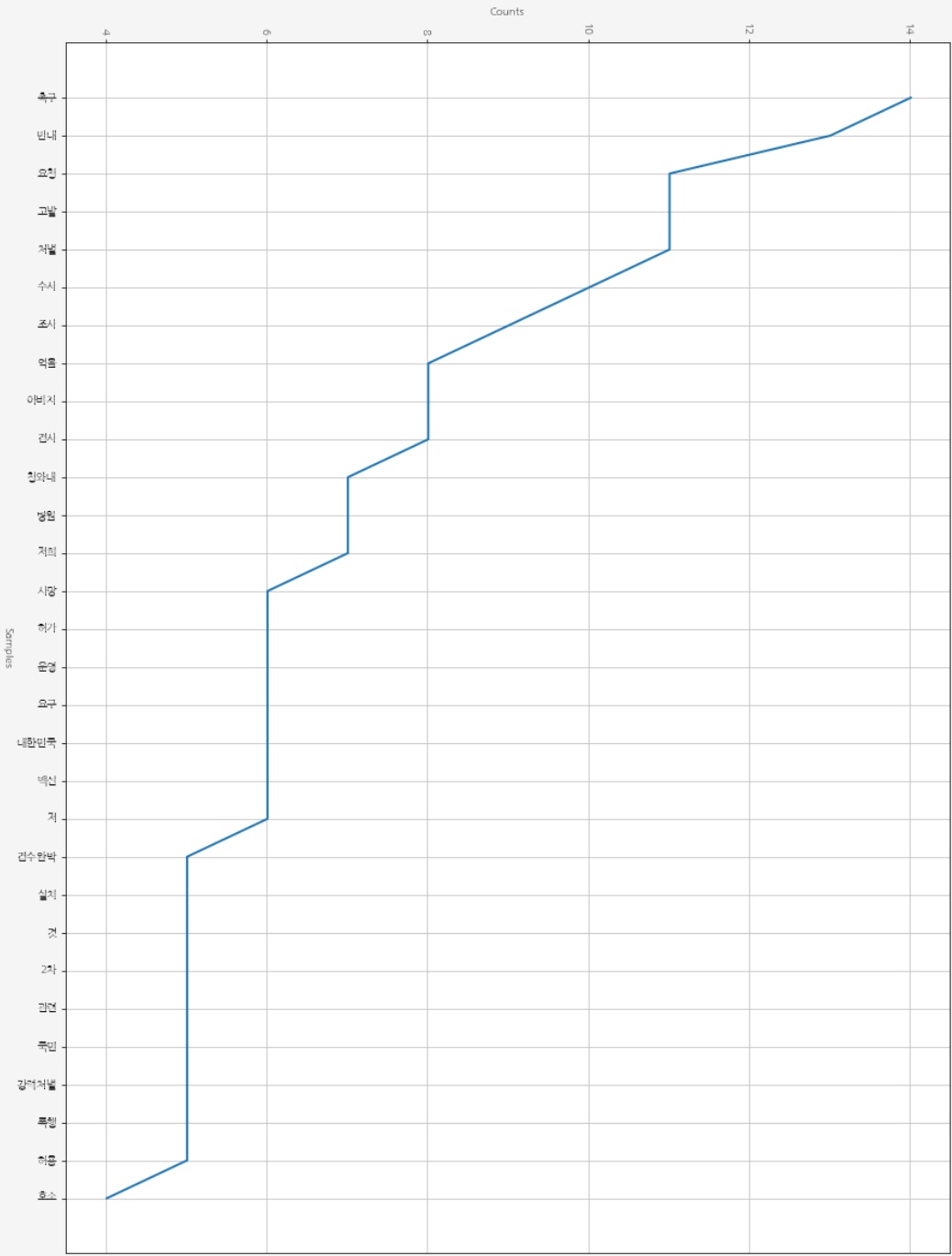
# 불용어를 제외한 리스트
new_k = []
for words in k:
    if words not in stop_words: 불용어 처리
        new_k.append(words)
print(new_k[:20])

# 빈도수 상위 30개 단어 표시
new_ko = nltk.Text(new_k, name='국민청원2')
plt.figure(figsize=(20,15))
new_ko.plot(30)
plt.show()

# 03-1 nltk 객체로를 이용하여 가장 많은 빈도수 데이터 추출
data = new_ko.vocab().most_common(500)
data
```

## 불용어 (stopwords)

- 자연어처리 시 의미가 없는 단어나 조사등을 의미한다.
- 자주 등장하지만 분석을 하는 데에는 거의 기여하는 바가 없는 단어들을 의미.





# 시각화

## # 06

# 모듈 호출

from wordcloud import WordCloud, STOPWORDS

시각화

import numpy as np

from PIL import Image

# - 그래프 그려보기, 분포 확인. 단어 빈도수 확인 ...

# 04. 워드 클라우드

```
wc = WordCloud(background_color='white',      # 배경색
                max_words=500,                # 최대 표시 단어
                contour_width=3,              # 외곽선
                contour_color="steelblue",    # 외곽선 색
                font_path=path)               # 마스크 이미지 정보
wc.generate_from_frequencies(dict(data))
```

plt.figure(figsize=(15,10))

plt.imshow(wc)

plt.axis('off')

plt.show()

# 국민청원 게시물 제목 시각화

```
# 모듈 호출
from wordcloud
```

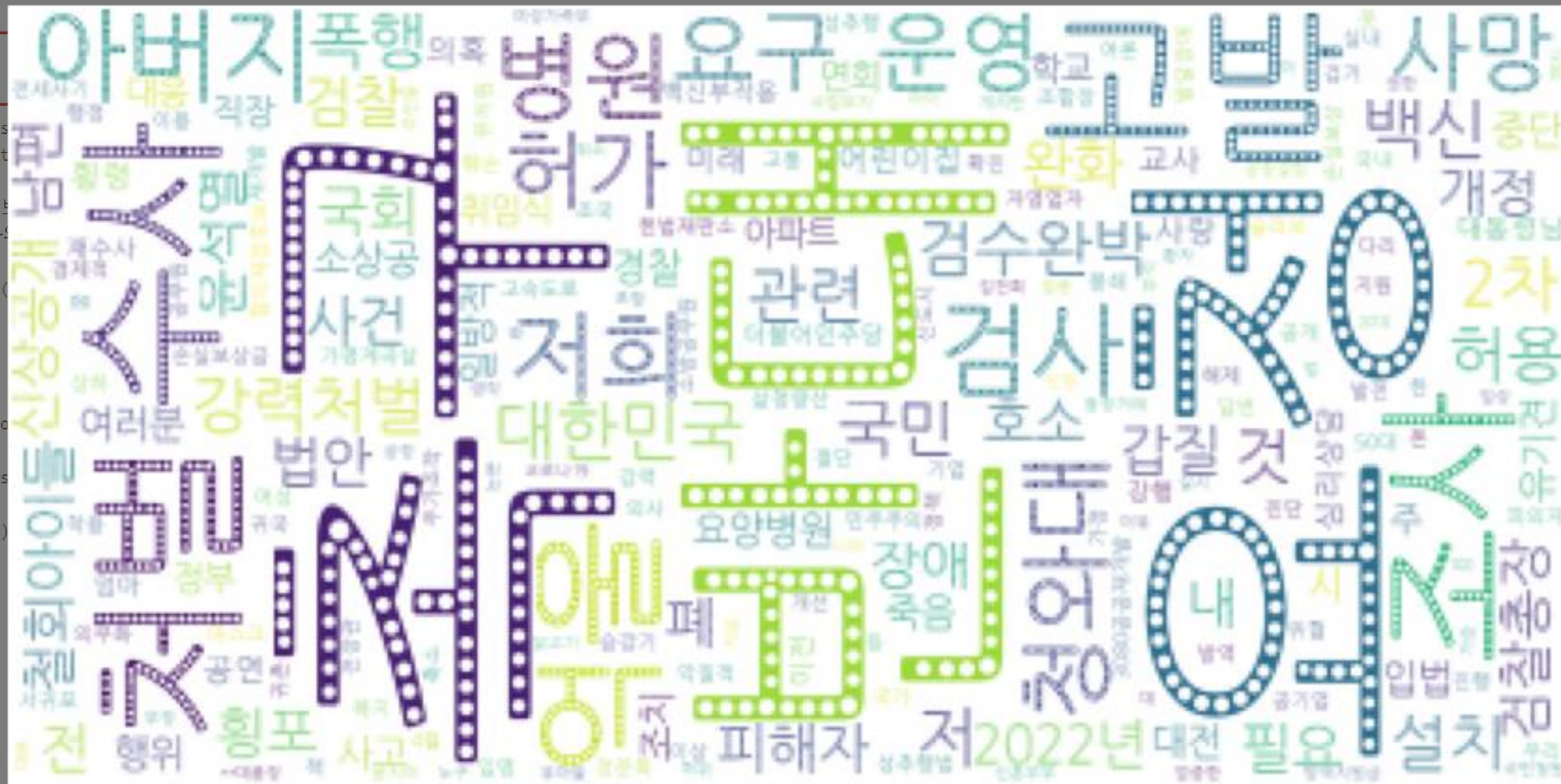
```
import numpy as
from PIL import
```

# - 그래프 그려보

```
wc = WordCloud(
```

```
wc.generate_from
```

```
plt.figure(figsize=(10, 10))
plt.imshow(wc)
plt.axis('off')
plt.show()
```



# 가설확인

# 07

# 모듈 불러오기

```
import selenium
from selenium import webdriver as wd
import pandas as pd
from bs4 import BeautifulSoup
```

‘셀레니엄’ 웹브라우저 환경 설정

# 청와대 검색 페이지

```
driver = wd.Chrome('chromedriver.exe')
url = 'https://www1.president.go.kr/search'
```

청와대 접속

# 검색어 입력 및 검색대상 제목 설정

```
search = '아파트'
driver.find_element_by_xpath('//*[@id="query"]').send_keys(search)
driver.find_element_by_xpath('//*[@id="contents"]/div[1]/div/div[1]/ul/li[2]/label').click()
```

‘아파트’검색

# 날짜 입력

```
start_date = '2022/04/08'
end_date = '2022/05/08'
driver.find_element_by_xpath('//*[@id="start_Date"]').send_keys(start_date)
driver.find_element_by_xpath('//*[@id="end_Date"]').send_keys(end_date)
driver.find_element_by_xpath('//*[@id="login"]/li[2]/input').click()
```

# 국민청원 및 제안 페이지 클릭

```
driver.find_element_by_xpath('//*[@id="contents"]/div[2]/div/ul/li[5]/a').click()
```

# 검색결과 붕괴사고와 관련이 없다. 1차 가설확인

# 그래서 새롭게 ‘백신’이라는 키워드를 선정하여 ‘백신’이라는 키워드는 ‘부작용에 대한 조사를 요청하는 것이다’라는 가서를 세웠다. 2차 가설 설정

Jupyter Notebook

## 가설

국민청원 키워드의 상위 30개 중

‘아파트’라는 키워드는

광주의 아파트 붕괴 사고와 관련된 국민청원일 것이다.



# 가설확인

# 07

# 모듈 불러오기

```
import selenium
from selenium import webdriver as wd
import pandas as pd
from bs4 import BeautifulSoup
```

# 청와대 검색 페이지

```
driver = wd.Chrome('chromedriver.exe')
url = 'https://www1.president.go.kr/search'
driver.get(url)
```

# 검색어 입력 및 검색대상 제목 설정

```
search = '아파트'
driver.find_element_by_xpath('//*[@id="query"]').send_keys(search)
driver.find_element_by_xpath('//*[@id="contents"]/div[1]/div/div[1]/ul/li[2]/label').click()
```

# 날짜 입력

```
start_date = '2022/04/08'
end_date = '2022/05/08'
driver.find_element_by_xpath('//*[@id="start_Date"]').send_keys(start_date)
driver.find_element_by_xpath('//*[@id="end_Date"]').send_keys(end_date)
driver.find_element_by_xpath('//*[@id="login"]/li[2]/input').click()
```

# 국민청원 및 제안 페이지 클릭

```
driver.find_element_by_xpath('//*[@id="contents"]/div[2]/div/ul/li[5]/a').click()
```

# 검색결과 붕괴사고와 관련이 없다. 1차 가설확인

# 그래서 새롭게 '백신'이라는 키워드를 선정하여 '백신'이라는 키워드는 '부작용에 대한 조사를 요청하는 것이다'라는 가서를 세웠다.

## 1차 가설확인

국민청원 키워드의 상위 30개 중

‘아파트’라는 키워드는

광주의 아파트 붕괴 사고와 관련된 국민청원일 것이다.

• 국민청원 및 제안(4)

공공지원 민간 임대 **아파트** 청년특별공급에 대한 민원 및 보상 요청 건

[청원시작 : 2022-04-28] [청원마감 : 2022-05-28] [청원인원 : 293]

공공지원 민간임대주택 공급위치:경기도 평택시 \*\*\*\* \*\* ①공급유형 ▶신혼부부 특별공급 ▶청년 특별공급 ▶일반공급 위 공급유형 中 [청년 특별공급]에 당첨된 임차인으로 거주중에 있습니다. 모집공고문과 계약당시 내용과 현재(2022.04.18.) 재계약 시 임차인 자격조건 유지조건이 변경 되어 혼인신고를 하게 ...

HOME > 국민소통광장 > 국민청원 및 제안

울산 **아파트**분양 거실정면에 송전탑 35m 설치 전국**아파트** 최단거리 송전탑 어디에 하소연 해야하나요?

[청원시작 : 2022-04-26] [청원마감 : 2022-05-26] [청원인원 : 1,184]

울산 \*구 \*동 공공택지조성하에 일반분양 \*\*\*\* 1082세대 (23.11월 입주) A1 국민임대 437세대 A2 공공임대 373세대 C1 공공분양 358세대 \*\*\*\* 입주자공고 21.5월 최대 경쟁률 58:1 => 송전탑이 없었죠. 작은방 큰방 거실 정면에 송전탑 약 35~40m 지점에 탑설치 중도금 입금후 21.11월 송전탑 46m 남쪽 북쪽 ...

HOME > 국민소통광장 > 국민청원 및 제안

입주**아파트** 키친핏미선택시 냉장고장의 일반냉장고와 김치냉장고의 중앙판 제거

[청원시작 : 2022-04-25] [청원마감 : 2022-05-25] [청원인원 : 128]

키친핏은 그냥 부엌모양을 말하는데 키친핏미선택(부엌모양을 안 한다) 세대에게 정해진 부엌공간에 중간판을 굳이 넣어 두 크기에 맞는 일반냉장고와 김치냉고만을 쓰게 하는 것은 억지이고 횡포라 생각합니다. 입주시부터 키친핏미선택세대도 냉장고장에 냉장고 설치 유무를 떠나 자유롭게 쓰고 싶습니다. 유상...

HOME > 국민소통광장 > 국민청원 및 제안

**\*\*시 아파트** 앞 초인접 거리에 건설계획중인 지식산업센터 허가를 막아주세요!! 제발 도와주세요!!

[청원시작 : 2022-04-19] [청원마감 : 2022-05-19] [청원인원 : 294]

**\*\*시 \*\***에 살고 있는 \*\*\*\*\* 아파트주민입니다! 제가 사는 아파트 앞 초인접 거리에 지식산업센터가 들어온다고 합니다! 이 지식산업센터는 저희 아파트 단지 와 마주보고 있고 아파트 1601동 단지를 다 가려버려서 사생활 보호도 전혀 안됩니다. 또한 높은 층고로 인해 저희 아파트 단지보다 높게 지어서 일조권 피...

HOME > 국민소통광장 > 국민청원 및 제안

2차 가설 설정

# 가설확인

# 08

# 청와대 검색 페이지

```
driver = wd.Chrome('chromedriver.exe')
url = 'https://www1.president.go.kr/search'
driver.get(url)
```

# 검색어 입력 및 검색대상 제목 설정

```
search = '백신'
driver.find_element_by_xpath('//*[@id="query"]').send_keys(search)
driver.find_element_by_xpath('//*[@id="contents"]/div[1]/div/div[1]/ul/li[2]/label').click()
```

# 날짜 입력

```
start_date = '2022/04/08'
end_date = '2022/05/08'
driver.find_element_by_xpath('//*[@id="start_Date"]').send_keys(start_date)
driver.find_element_by_xpath('//*[@id="end_Date"]').send_keys(end_date)
driver.find_element_by_xpath('//*[@id="login"]/li[2]/input').click()
```

# 국민청원 및 제안 페이지 클릭

```
driver.find_element_by_xpath('//*[@id="contents"]/div[2]/div/ul/li[5]/a').click()
```

# 검색결과 6개 전부 백신 부작용과 관련있었다.

# 드라이버 현재 페이지의 html 정보 가져오기

```
html = driver.page_source
soup = BeautifulSoup(html, 'lxml')
```

## 2차 가설확인

• 국민청원 및 제안(6)

### 백신부작용으로 여자친구가 다리를 절단했습니다

[청원시작 : 2022-05-04] [청원마감 : 2022-06-03] [청원인원 : 6,985]

제목: 백신 부작용으로 여자친구가 다리를 절단했습니다. 건강하던 30살(93년생) 여성이 백신 부작용으로 추측되는 혈전으로 인해 다리가 절단되었지만 국가에서 인정받지 못해 억울한 사연을 남자친구가 청원합니다. 안녕하세요. 서울에 살고 있는 32살 남자입니다. 지난 3월14일 오후 6시경 결혼을 생각하고 있...

HOME > 국민소통중앙 > 국민청원 및 제안

### 화이자백신 접종으로 인한 회전근계파열에 대한 정부의 책임을 묻고자 합니다.

[청원시작 : 2022-04-26] [청원마감 : 2022-05-26] [청원인원 : 293]

내용제목 그대로 화이자 1~2차 백신 접종 후 신장이 회전근계파열이 발생 하였습니다. 정부는 이와같은 사상의 책임을 묻고자 합니다. 지금 저와 같이 피해를 본 분들이 많다고 들었습니다. 하지만 정부에서는 지금까지 어떠한 대책도 논의 되지 않고, 어떠한 대책도 나오지 않습니다. SIRVA라는 명칭에 대해서 알고 ...

HOME > 국민소통중앙 > 국민청원 및 제안

### 2번사망선고및 백신부작용

[청원시작 : 2022-04-25] [청원마감 : 2022-05-25] [청원인원 : 251]

작년12월달 아버지께서 백신3차맞으시고 6일만에 돌아가셨습니다 아침에갑자기 숨쉬기가 거북하시다 하셔서 제가cpr및 응급구급대원의 빠른처치로 심장은 돌아오셔서 \*\*\*\*\*병원 응급실로 가셨습니다 병원도착후30분후 코로나검사소에서 담당의가 저한테1차사망선고를 하셨습니다 하지만 약5분후 저한테 죽음...

HOME > 국민소통중앙 > 국민청원 및 제안

### 백신 부작용으로 인해 고통받고 있습니다

[청원시작 : 2022-04-20] [청원마감 : 2022-05-20] [청원인원 : 1,423]

본인은 부산 북구에서 2자녀와 부인을 두고 있는 가장이며, 부산 영도 선박회사에서 근무하고 있는 자입니다. 금번 코로나 19로 인하여 국가 코로나 백신을 전 국민에게 예방접종을 하게 되어 본인 또한 그 일정에 따라 1차 2차 백신을 맞았습니다. 그런데 2021.09.24 금요일 오후4경에 모더니 백신 2차접종 후 저녁...

HOME > 국민소통중앙 > 국민청원 및 제안

### 청년의 꿈과 인생을 짓밟은 코로나 백신 화이자접종 후 백혈병

[청원시작 : 2022-04-20] [청원마감 : 2022-05-20] [청원인원 : 614]

근일이 일어 났다! 중국 우한에서 코로나-19 발병! 대한민국은 긴장했다. 얀든, 그나마 안전했던 대한민국에 코로나가 유행처럼 번지기 시작하였고, 대한민국 정부는 백신 구입에 열을 올리며, 실제 처럼 구매하기 시작했다. 그리고는 엄청난 양을 구입하여 노인부터 시작하여 순차적으로 전국민이 예방접종을 하도...

HOME > 국민소통중앙 > 국민청원 및 제안

### \*\*백신으로 중학생 아이 사망

[청원시작 : 2022-04-08] [청원마감 : 2022-05-08] [청원인원 : 3,430]

백신으로 우리아이 죽음으로 만든 문제인 정부 저는 인천 계양구 에서 살고 있으며 2월28일 만13세 사랑으로 키우던 건강한 아이를 잃은 엄마입니다. 하루아침에 멀쩡하던 아들이 잠을 자듯 가버린 그날을 잊을 수 없고 저희 가족은 그 날이후로 아무것도 할 수 없는 고통으로 지내고 있습니다. 저희는 뉴스조차 가...

HOME > 국민소통중앙 > 국민청원 및 제안

# 기타활동

# 09

```
# 청원 제목
soup.find_all('h2', attrs = {'style':"margin-bottom:10px;"})
[title.get_text() for title in soup.find_all('h2', attrs = {'style':"margin-bottom:10px;"})]
```

```
# 청원 제목을 txt파일로 만들어서 워드클라우드 시도
f = open('vaccine.txt', 'w')
f.write("""'백신부작용으로 여자친구가 다리를 절단했습니다',
'화이자백신 접종으로 인한 회전근개파열에 대한 정부의 책임을 묻고자 합니다.',
'2번사망선고및 백신부작용',
'백신 부작용으로 인해 고통받고 있습니다',
'청년의 꿈과 인생을 짓밟은 코로나 백신 화이자접종 후 백혈병',
'**백신으로 중학생 아이 사망'
""")
```

```
# 아래서 부터는 'JVMNotFoundException' 오류가 떠서 해결해볼려 했으나 해결하지 못함
import nltk
!pip install konlpy
!pip install WordCloud

import os
import konlpy

from konlpy.tag import Kkma

from wordcloud import WordCloud

import matplotlib.pyplot as plt

k = Kkma()
```

**Jupyter Notebook**  
에서 워드클라우드 실행 오류

# 기타활동

# 09

## Jupyter Notebook 시각화

# 청원 제목

```
soup.find_all('h2', attrs = {'style':"margin-bottom:10px;"})  
[title.get_text() for title in soup.find_all('h2', attrs = {'style':"margin-bottom:10px;"})]
```

# 청원 제목을 txt파일로 만

```
f = open('vaccine.txt',  
f.write("""'백신부작용으로  
'화이자백신 접종으로 인한  
'2번사망선고및 백신부작용  
'백신 부작용으로 인해 고통  
'청년의 꿈과 인생을 짓밟음  
'**백신으로 중학생 아이  
""")
```

# 아래서 부터는 'JVMNotFo

```
import nltk  
!pip install konlpy  
!pip install WordCloud
```

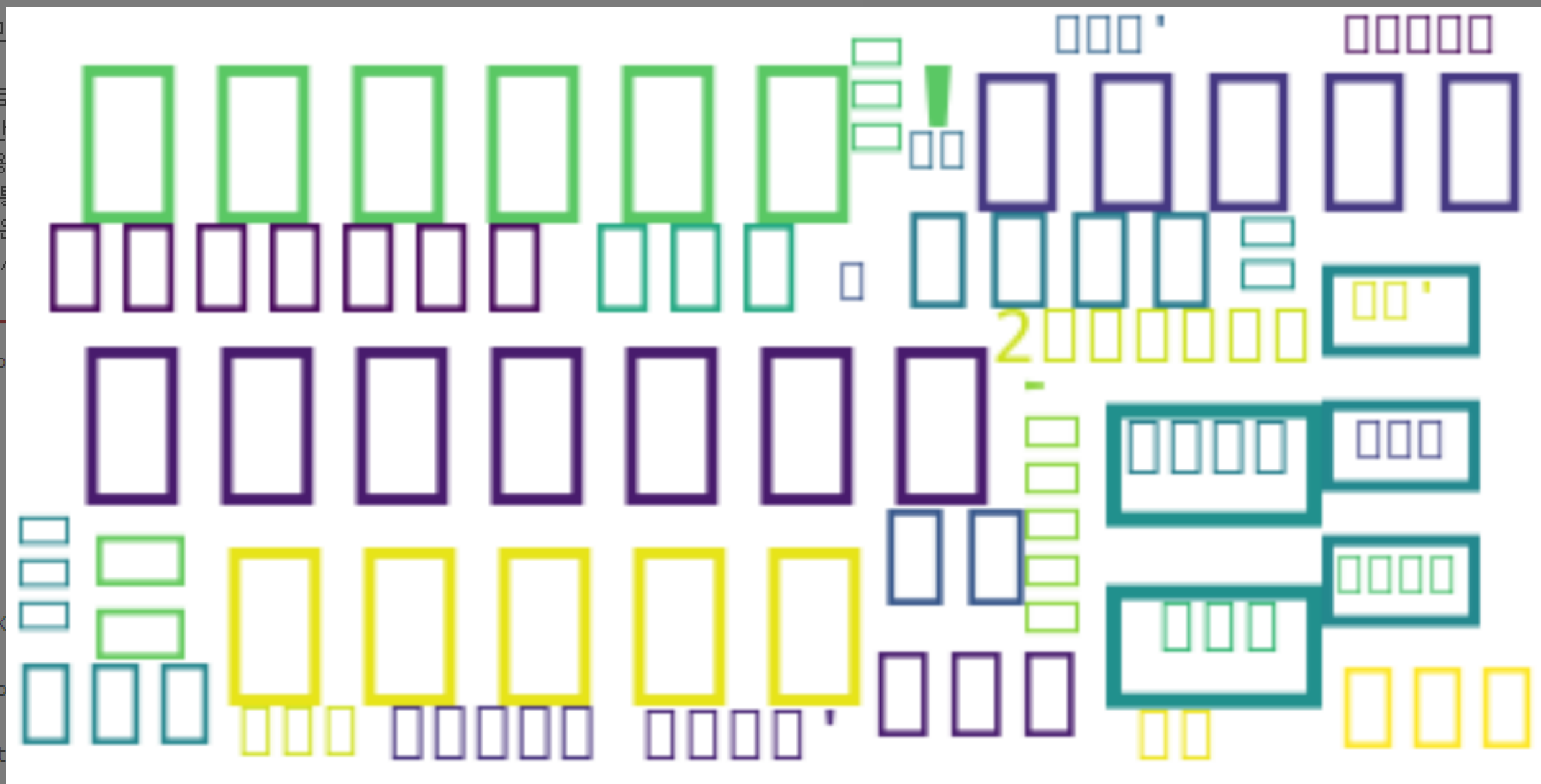
```
import os  
import konlpy
```

```
from konlpy.tag import Kkma
```

```
from wordcloud import WordCloud
```

```
import matplotlib.pyplot
```

```
k = Kkma()
```



3조 - 열정

Thank

팀 장 : 강진오

부팀장 : 김미현

22.05.07