

국내 지역별 수돗물 음용률 분석

16조) 이다빈 | 오피연 | 김준형

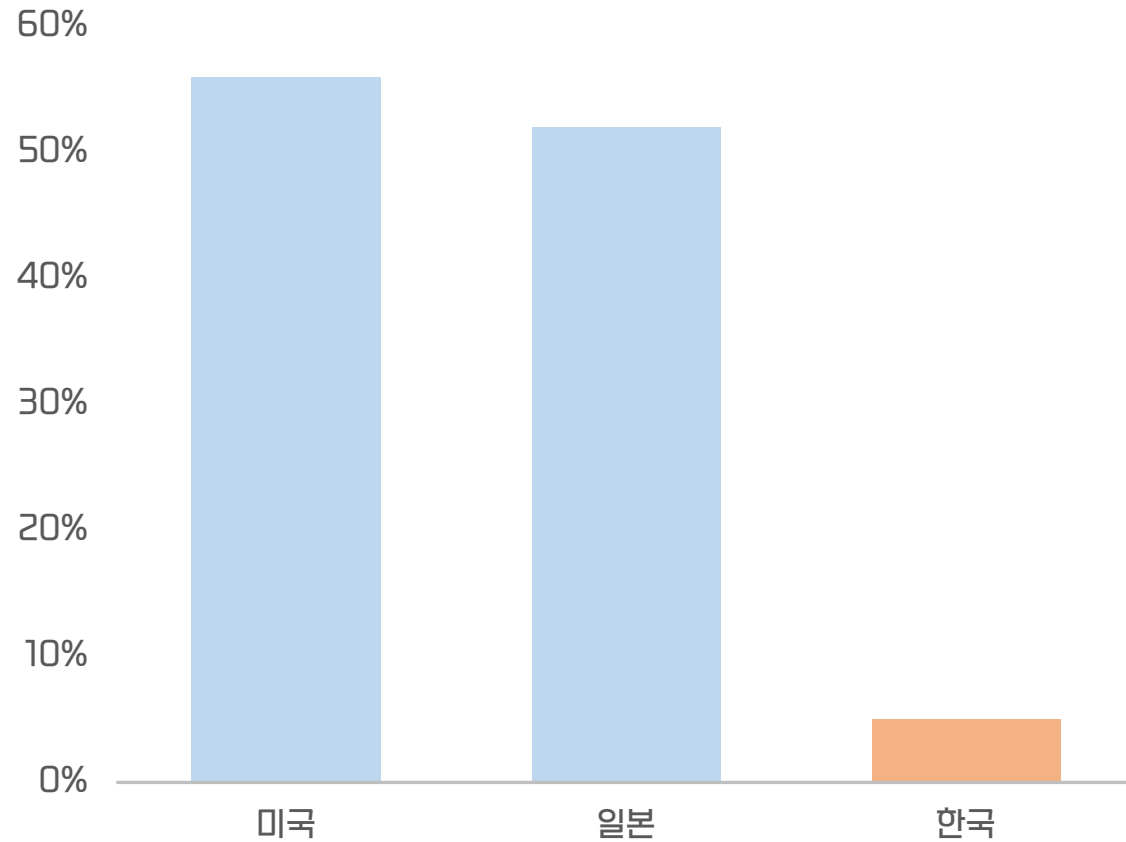


contents

- 01 주제 선정 배경
- 02 데이터 수집/ 분석
- 03 핵심 코드 설명/ 결과
- 04 어려웠던 점/ 참고 자료

1. 주제 선정 배경

국가별 수돗물 음용률



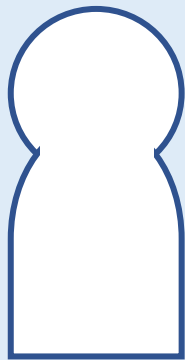
세계 수돗물 품질 순위

순위	국가
1위	핀란드
2위	캐나다
4위	일본
8위	대한민국

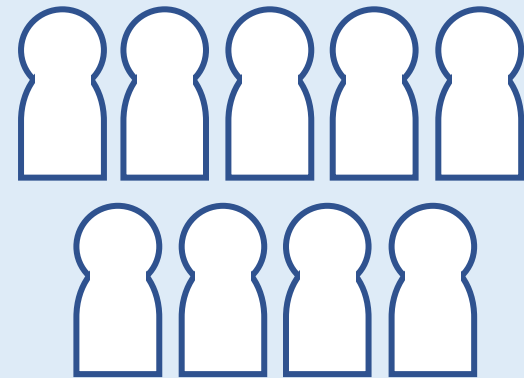
- 우리나라는 세계 수돗물 품질 순위에서 핀란드, 캐나다, 일본에 이어 8위를 차지할 정도로 수돗물 품질이 높게 평가받고 있다.
- 그러나 ‘국가별 수돗물 음용률’에서 우리나라는 미국 56%, 일본 52%에 비해 굉장히 낮은 수치인 5%로, OECD 평균인 50%를 크게 밑도는 것을 알 수 있다.

Q. 수돗물을 그대로 마시나요?

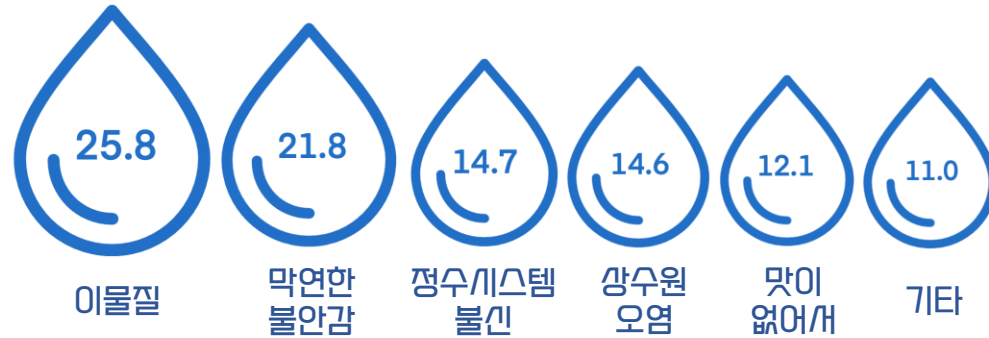
A. 그렇다



A. 아니다
(정수기, 생수)

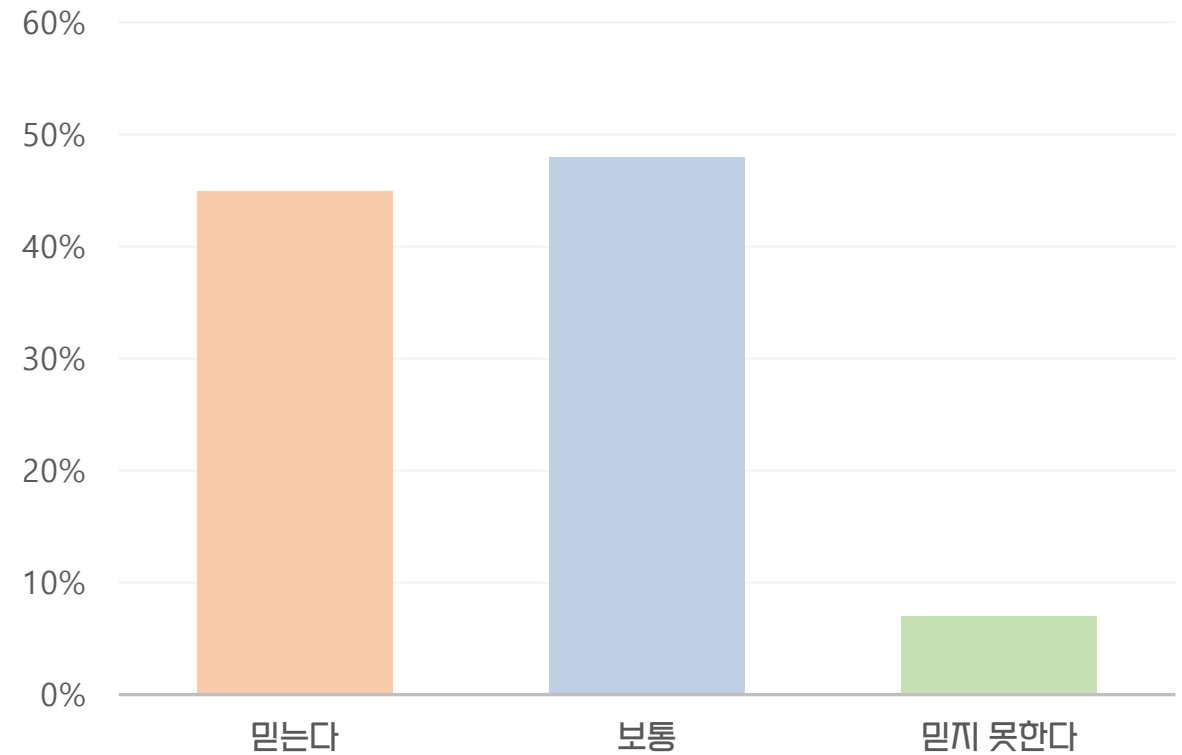


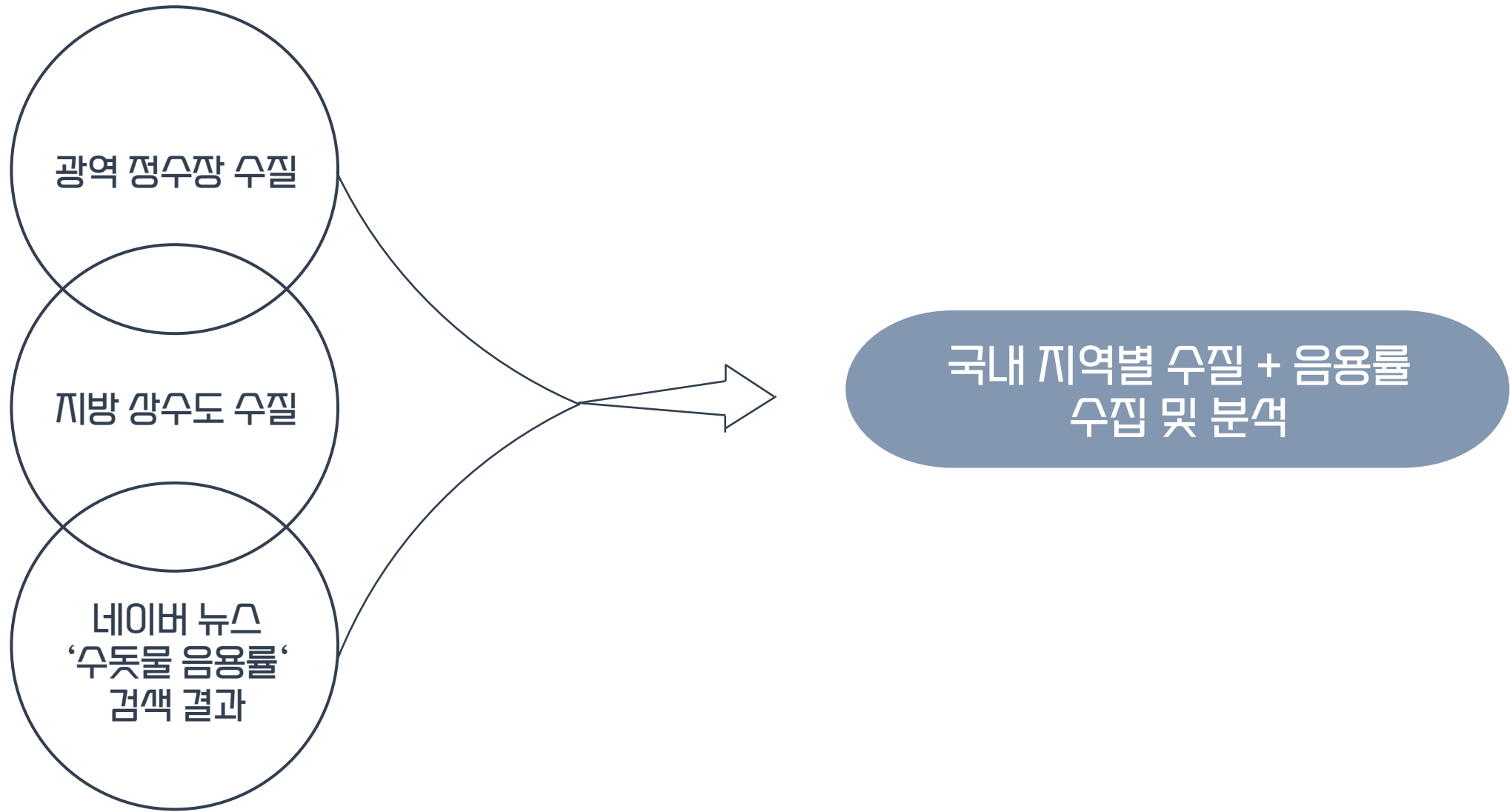
수돗물 안 마시는 이유 (단위: %)



- 수돗물 신뢰 여부 설문에서 믿는다는 의견은 45%, 보통은 48%, 믿지 못한다는 의견은 7%로 보통이 믿는다는 의견보다 많았다.
- 수돗물을 마시지 않는 이유로는 이물질이 25.8%로 가장 높았고 그 뒤로 막연한 불안감, 정수시스템 불신 등이 뒤를 이었다.

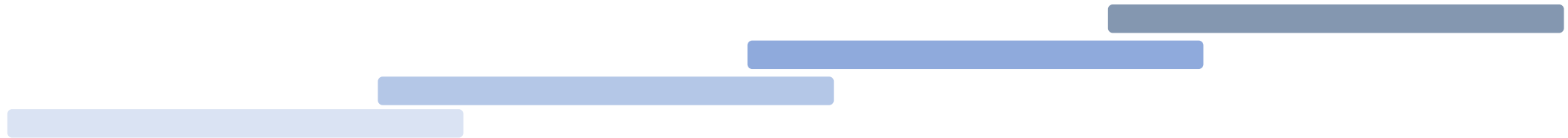
수돗물 신뢰 여부





2. 데이터 수집/ 분석

데이터 수집 + 분석 과정



01. 데이터 수집

- 광역 정수장 수질 데이터
- 지방 상수도 수질 데이터
- 네이버 뉴스 '수돗물 음용률' 기사 데이터

02. 데이터 정리/분석

- 시, 도별로 데이터 정리
- '수돗물 음용률' 관련 기사 본문 키워드 분석

03. 데이터 시각화

- 지역별 수질 + 뉴스 키워드 분석 결과 시각화

04. 결과

- 분석 결과

데이터 수집 + 분석 과정

지방상수도 수질

요청변수(Request Parameter)

항목명(국문)	항목명(영문)	항목크기	항목구분	샘플데이터	
지자체코드	sgccd	5	옵션	48310	지자체코드
정수장코드	sitecd	3	옵션	21002	정수장코드
조회시작일	stdt	8	필수	20190608	조회시작일
조회종료일	eddt	8	필수	20190618	조회종료일

광역정수장 수질

요청변수(Request Parameter)

항목명(국문)	항목명(영문)	항목크기	항목구분	샘플데이터	항목설명
정수장코드	fcode	3	필	A002	정수장 코드 조회 서비스 참조
조회시작일	stdt	10	필	2018-10-01	조회시작일
조회종료일	eddt	10	필	2018-10-05	조회종료일

- 지방상수도 수질은 지자체코드와 정수장코드를, 광역정수장 수질은 정수장코드를 입력한다.
- 수질 데이터와 함께 제공되는 지자체 코드, 정수장코드 데이터와 연결한다.
- 코드를 지역별로 정리하여 지역을 입력 시 코드가 반환되어 해당 지역의 수질을 보여준다.

데이터 수집 + 분석 과정

출력결과(Response Element)

항목명(국문)	항목명(영문)	항목크기	항목구분	
측정시간	measurede	10	필	20141007
맛	item1	3	옵	무미
냄새	item2	3	옵	무취
색도(도)	item3	3	옵	0.2
pH(-)	item4	3	옵	7.0
탁도(NTU)	item5	3	옵	0.06
잔류염소(mg/L)	item6	3	옵	0.71
줄수	numOfRows	5	필	10
페이지번호	pageNo	5	필	1

수질 적합 기준 >

색도	5이하
pH	5.8~8.5
탁도(NTU)	0.5이하
잔류염소	4.0이하

- 지방상수도 수질 데이터, 광역정수장 수질 데이터는 공통적으로 맛, 냄새, 색도, pH, 탁도, 잔류염소를 출력한다.
- 수질 적합 기준은 색도는 5이하, pH는 5.8~8.5사이, 탁도는 0.5이하, 잔류염소는 4.0이하여야 한다.

데이터 수집 + 분석 과정

 | 수돗물 음용률

 파이낸셜뉴스 | 3일 전 | 네이버뉴스

양평군 수도시설 운영관리 '우수'...음용률 향상

3그룹(인구 5만명 이상 20만명 미만 39개 시·군)에 속한 양평군은 올해 평가에서 정수장 실시간 감시-제어 등 모니터링 사업, 정수장 신·증설 사업, 수도시설 개량...



 아시아투데이 | 5일 전

오산시, 2년 연속 상수도행정 최우수 선정

시는 적극적이고 능동적으로 현장을 찾아가는 정책설명회와 고객간담회 실시, 코로나 상황 등을 감안한 수돗물 음용률 제고 위한 온라인 대민홍보강화, 수돗물 유...



오산시, 2년 연속 상수도행정 최우수도시 선정 | 이뉴스투데이 | 5일 전
오산시, 환경부 평가서 2년연속 상수도행정 최우수도시 선정 | 뉴스핌 | 5일 전
오산시, 2년 연속 상수도행정 '최우수도시' 선정...2000만원 ... | 아주경제 | 4일 전
오산시, 환경부 수도 운영·관리 2년 연속 최우수 | NSP통신 | 4일 전

관련뉴스 12건 전체보기 >

 한국강사신문 | 3일 전

다큐온, 공공 음용수 '수돗물 드시나요?'

그래서 우리나라 수돗물 음용률은 34%. 끓여서 먹거나 차로 마시는 경우를 다 포함한 수치로 선진국의 수돗물 음용률에 비해 턱없이 낮다. 우리는 왜 수돗물을 믿...



```
{
  "cnt": 1,
  "description": "시는 <b>수돗물 음용률</b> 제고를 위한 온라인 홍보강화, <b>수돗물</b>
  "link": "http://www.sisajournal.com/news/articleView.html?idxno=230057",
  "org_link": "http://www.sisajournal.com/news/articleView.html?idxno=230057",
  "pDate": "2021-12-20 17:00:00",
  "title": "[오산24시] '탄소중립도시 오산!', 신재생에너지 확대 보급 모색한다"
},
```

- 지역 입력 시 해당 지역 관련 기사를 보여주기 위해 '지역명' + '수돗물 음용률' 검색 시 나오는 기사 리스트 크롤링
- 크롤링한 기사 데이터에서 링크를 가져와 각각의 기사 제목, 본문, 요약내용 크롤링

데이터 수집 + 분석 과정

[[('수돗물', '음용률', '온라인', '홍보', '강화', '수돗물', '민관', '협력', '수시', '평가', '연속', '최우수상', '차지도', '사업', '운영', '관리', '실태', '평가', '연속', '최수물', '유충', '대응', '전국', '우수', '사례', '소개'], ['점검', '우수', '기관', '선정', '수돗물', '안심', '확인', '아일보', '양평'], ['경기도', '오산시', '환경부', '주관', '독물', '음용률', '온라인', '홍보', '자체', '평가', '점수보', '강화', '수돗물', '유충', '대응', '전국', '우수', '기부', '시작', '소재', '어린이집'], ['수도', '사업', '운최우수', '경기도', '오산시', '환경부', '감안', '수돗물', '관', '교체', '수돗물', '공급', '스마트', '관망', '관리'], ['위해', '스마트', '관망', '관리', '통해', '수돗물', '생산산'], ['경기도', '여주시', '시장', '항진', '지난', '여주시과정', '수질', '수량', '실시간', '확인', '수질', '이상'], ['위해', '수돗물', '생산', '공급', '과정', '수질', '수량',

[(0, '0.022*관리" + 0.020*수돗물" + 0.016*스마트" + 0.014*사업" + 0.012*0.011*음용률" + 0.010*경기도" + 0.009*시민" + 0.009*여주시" + 0.009*수도 + 0.018*수돗물" + 0.018*사업" + 0.018*평가" + 0.013*스마트" + 0.013*운영0*관망" + 0.009*오산시" + 0.009*시민" + 0.009*음용률" + 0.007*실태" + 0.017*평가" + 0.015*스마트" + 0.015*사업" + 0.014*오산시" + 0.012*상수도"*전국" + 0.009*시민" + 0.008*복지" + 0.007*구축" + 0.007*음용률"), (3, *사업" + 0.014*수도" + 0.013*상수도" + 0.012*스마트" + 0.012*운영" + 0.0(질" + 0.008*여주시" + 0.008*환경부" + 0.007*실태"))]

- 수돗물 음용률과 관련된 토픽 분석을 위해 크롤링한 기사의 제목, 본문, 요약내용 저장 후 명사 추출 후 벡터화
- 설정한 토픽의 개수에 따라 토픽 분석 결과 출력 + 분석 결과 시각화

3. 핵심 코드 설명/ 결과

핵심 코드 설명

getTextFromNews.py

```
def countHangul(text):  
    #텍스트로부터 한글을 읽는데, 단어를 개수로 셈(정규식사용)  
    hanCount=len(re.findall(u'[\u3130-\u318F\uAC00-\uD7A3]+',text))  
    return hanCount
```

```
def getTextFromLink(url):  
    print(url)  
    try:  
        html=urllib.request.urlopen(url)  
    except HTTPError as e:  
        err=e.read()  
        code=e.getcode()  
        return None  
    soupNews = BeautifulSoup(html, 'html.parser')  
    #print(soupNews.prettify())  
    tag_tbody = soupNews.find('tbody')  
  
    #tag_br=soupNews.select_one("br").parent.get_text().strip()  
    |  
    tag_p=soupNews.find_all("p")  
  
    #print("#ntag_p")  
    #print(tag_p)  
    #print("#ntag_br")  
    #print(tag_br)  
    return tag_p
```

- def countHangul(text)

뉴스기사의 url링크를 받아와서 리퀘스트 요청 후
<p>태그를 기준으로 기사 본문내용을 읽어옴

- getTextFromLink(url)

%s_naver_news.json형식의 해당 검색어로 작성
된 뉴스 기사를 읽어온 후 getTextFromLink()를
호출하여 본문 기사를 읽어와서 result에 저장

- getTextFromNewsJson(srcText)

뉴스의 제목과, 요약, 추출한 본문의 내용을
_TEXT.json 파일명으로 저장합니다.

getTopicLDA.py

```
def okt_tokenizer(text):
    tokens = okt.morphs(text)
    return tokens

-----
def getTopicModelLDA(srcText):
    file_name=srcText
    warnings.filterwarnings(action='ignore')
    okt = Okt()
    with open('./'+file_name+'_TEXT.json', encoding='utf8') as j_f:
        data = json.load(j_f)

    data_title = []
    data_description = []
    data_text=[]

    for item in data:
        data_title.append(item['title'])
        data_description.append(item['description'])
        data_text.append(item['TEXT'])

    for d in description:
        all_noun_tk.append(okt.nouns(d)) #형태소가 명사인 것만 추출
    for d in title:
        all_noun_tk.append(okt.nouns(d)) #형태소가 명사인 것만 추출
    for d in text:
        all_noun_tk.append(okt.nouns(d)) #형태소가 명사인 것만 추출
```

```
#LDA 모델의 입력 벡터 생성
#1 - 단어사전생성
dictionary = corpora.Dictionary(all_noun_tk2)
print(dictionary[1]) #작업 확인용 출력

#2 - 단어와 출현빈도(count)의 코퍼스 생성
corpus = [dictionary.doc2bow(word) for word in all_noun_tk2]
print(corpus) #작업 확인용 출력

#LDA 모델 생성 및 훈련
k = 4 #토픽의 개수 설정
lda_model = gensim.models.Ldamulticore.LdaMulticore(corpus, iterations = 12,
                                                    num_topics = k, id2word = dictionary,
                                                    passes = 1, workers = 10)

print(lda_model.print_topics(num_topics = k, num_words = 15))

lda_vis = pyLDavis.gensim_models.prepare(lda_model, corpus, dictionary)
pyLDavis.display(lda_vis)
pyLDavis.save_html(lda_vis, './'+file_name+"_vis.html")
```

- okt_tokenizer(text)

konlpy를 사용, 한글을 추출

- getTopicModelLDA(srcText)

srcText+"_TEXT.json"파일을 읽어와서 형태소가 명사인것만 추출.
그 후 1.단어사전생성, 2.단어와 출현빈도의 코퍼스생성 3.LDA모델
을 생성후 훈련하고 해당 내용을 _vis.html파일로 보기 쉽게 저장

핵심 코드 설명

```
def getWaterPurificationService(code):
    jsonResult = []

    jsonData = getWaterPurificationItem(code)
    print(jsonData)
    if (jsonData['response']['header']['resultMsg'] == 'NORMAL SERVICE.'):
        if (jsonData['response']['body']['totalCount'] != 0):
            taste = jsonData['response']['body']['items']['item']['item1']
            smell = jsonData['response']['body']['items']['item']['item2']
            color = jsonData['response']['body']['items']['item']['item3']
            ph = jsonData['response']['body']['items']['item']['item4']
            ntu = jsonData['response']['body']['items']['item']['item5']
            chlorine = jsonData['response']['body']['items']['item']['item6']

            jsonResult.append({'맛': taste, '냄새': smell, '색도': color, 'pH': ph, '탁도': ntu, '잔류염소': chlorine})

        else:
            print("데이터가 없습니다.")
            return None

    print(jsonResult)
    return jsonResult
```

#지역명(name)으로 정수장 코드 반환

```
def findCode(name):
    dic = {"수도권":["A005","A020","A028","A038","A040","A041","A051","A056","A064","A090"],
          "서울시":["A005","A020","A028","A038","A040","A041","A051","A056","A064","A090"],
          "경기도":["A005","A020","A028","A038","A040","A041","A051","A056","A064","A090"],
          "강원도":["A039","A072"],
          "충청북도":["A061","A063"],
          "충청남도":["A007","A015","A030","A037","A042","A060"],
          "전라북도":["A004","A014","A023","A033","A036"],
          "전라남도":["A018","A021","A029","A065","A071"],
          "경상북도":["A002","A012","A055","A069"],
          "경상남도":["A013","A026","A027","A034","A045","A047"],
          "울산광역시":["A050","A057"]}

    return dic[name]
```

핵심 코드 설명

```
def getPurificationData(srcText):  
    jsonResult = []  
    code = []  
    #name = input('지역 입력>') #지역명 입력  
    name=srcText  
    code = findCode(name)      #지역명 -> 코드로 변경 후 저장  
    print(code)  
  
    for c in code:  
        jsonResult.append(getWaterPurificationService(c))  
  
    #데이터가 있으면 저장  
    if (jsonResult != None):  
        with open('./%s.json' % (name), 'w', encoding = 'utf8') as outfile:  
            jsonFile = json.dumps(jsonResult, indent=4, sort_keys = True, ensure_ascii = False)  
            outfile.write(jsonFile)  
            print("저장 완료")  
        return jsonResult  
    return False
```

핵심 코드 설명

핵심 코드 설명

결과

```
if __name__ == '__main__':  
    main()
```

도시를 입력하세요(ex: 경기도): 경기도

경기도 수돗물 음용률

['A005', 'A020', 'A028', 'A038', 'A040', 'A041', 'A051', 'A056', 'A064', 'A090']

[2021-12-22 14:32:50.894230] Url Request Successs

{'response': {'header': {'resultCode': '00', 'resultMsg': 'NORMAL SERVICE.'}, 'body': {'items': {'item': {'item1': '무미', 'item2': '무취', 'item3': 0, 'item4': 7.84, 'item5': 0.059, 'item6': 0.83, 'mesurede': 20211201}}, 'numOfRows': 10, 'pageNo': 1, 'totalCount': 1}}}

[{'맛': '무미', '냄새': '무취', '색도': 0, 'pH': 7.84, '탁도': 0.059, '잔류염소': 0.83}]

[2021-12-22 14:32:54.211593] Url Request Successs

{'response': {'header': {'resultCode': '00', 'resultMsg': 'NORMAL SERVICE.'}, 'body': {'items': {'item': {'item1': '무미', 'item2': '무취', 'item3': 0, 'item4': 7.76, 'item5': 0.068, 'item6': 0.87, 'mesurede': 20211201}}, 'numOfRows': 10, 'pageNo': 1, 'totalCount': 1}}}

[{'맛': '무미', '냄새': '무취', '색도': 0, 'pH': 7.76, '탁도': 0.068, '잔류염소': 0.87}]

[2021-12-22 14:32:57.317545] Url Request Successs

{'response': {'header': {'resultCode': '00', 'resultMsg': 'NORMAL SERVICE.'}, 'body': {'items': {'item': {'item1': '무미', 'item2': '무취', 'item3': 0, 'item4': 7.78, 'item5': 0.049, 'item6': 0.82, 'mesurede': 20211201}}, 'numOfRows': 10, 'pageNo': 1, 'totalCount': 1}}}

[{'맛': '무미', '냄새': '무취', '색도': 0, 'pH': 7.78, '탁도': 0.049, '잔류염소': 0.82}]

[2021-12-22 14:33:00.462791] Url Request Successs

{'response': {'header': {'resultCode': '00', 'resultMsg': 'NORMAL SERVICE.'}, 'body': {'items': {'item': {'item1': '무미', 'item2': '무취', 'item3': 0, 'item4': '7.50', 'item5': 0.045, 'item6': '0.80', 'mesurede': 20211201}}, 'numOfRows': 10, 'pageNo': 1, 'totalCount': 1}}}

[{'맛': '무미', '냄새': '무취', '색도': 0, 'pH': '7.50', '탁도': 0.045, '잔류염소': '0.80'}]

[2021-12-22 14:33:04.053957] Url Request Successs

{'response': {'header': {'resultCode': '00', 'resultMsg': 'NORMAL SERVICE.'}, 'body': {'items': {'item': {'item1': '무미', 'item2': '무취', 'item3': 0, 'item4': 7.38, 'item5': 0.028, 'item6': 0.98, 'mesurede': 20211201}}, 'numOfRows': 10, 'pageNo': 1, 'totalCount': 1}}}

결과

```
if __name__ == '__main__':
    main()

2': '무취', 'item3': '불검출', 'item4': 7.41, 'item5': 0.073, 'item6': 0.82, 'mesurede': 20211201}}, 'numOfRows': 10, 'pageNo': 1,
'totalCount': 1}}]
[{'맛': '무미', '냄새': '무취', '색도': '불검출', 'pH': 7.41, '탁도': 0.073, '잔류염소': 0.82}]
저장 완료
광역정수장완료=====
[2021-12-22 14:33:20.836247] Url Request Success
[2021-12-22 14:33:21.140169] Url Request Success
[2021-12-22 14:33:21.431557] Url Request Success
[2021-12-22 14:33:21.637022] Url Request Success
[2021-12-22 14:33:21.753109] Url Request Success
전체 검색 : 331 건
가져온 데이터 : 331 건
경기도 수돗물 음용률_naver_news.json SAVED
뉴스크롤링완료=====
{'cnt': 1, 'description': "시는 <b>수돗물 음용률</b> 제고를 위한 온라인 홍보강화, <b>수돗물</b> 유출대응 전국우수사례 소개 등 시책
을 마련했다.... <b>경기도</b> '민관협력 우수시군 평가'에서도 2년 연속 최우수상을뽐내2008차지했다. 광상욱 시장은 &quot;코로나19가...
", 'link': 'http://www.sisajournal.com/news/articleView.html?idxno=230057', 'org_link': 'http://www.sisajournal.com/news/articleVi
ew.html?idxno=230057', 'pDate': '2021-12-20 17:00:00', 'title': "[오산24시] '탄소중립도시 오산!', 신재생에너지 확대 보급 모색한
다"}
{'cnt': 2, 'description': "<b>경기도</b> 오산시는 환경부에서 실시한'2021년수돗물수질조사'에서'과리신대평가'에서도'2년 연속 최우
```

결과

```
if __name__ == '__main__':  
    main()
```

Some characters could not be decoded, and were replaced with REPLACEMENT CHARACTER.

<http://www.shinailbo.co.kr/news/articleView.html?idxno=1486811>

<http://www.newsway.co.kr/news/view?tp=1&ud=2021110516464049702>

<http://www.shinailbo.co.kr/news/articleView.html?idxno=1479357>

<http://www.knnews.co.kr/news/articleView.php?idxno=1362742>

https://science.ytn.co.kr/program/program_view.php?s_mcd=0082&s_hcd=&key=202110271720510695

제목, 요약, 타이틀, json 완료

본문텍스트, 요약, 타이틀출완료=====

['수돗물', '음용률', '온라인', '홍보', '강화', '수돗물', '유충', '대응', '전국', '우수', '사례', '소개', '시책', '마련', '경기도', '민관', '협력', '수시', '평가', '연속', '최우수상', '차지', '박상욱', '시장', '코로나'], ['경기도', '오산시', '환경부', '실시', '수도', '사업', '운영', '관리', '실태', '평가', '연속', '최우수', '기관', '감안', '수돗물', '음용률', '온라인', '홍보', '강화', '수돗물', '유충', '대응', '전국', '우수', '사례', '소개'], ['경기도', '양평군', '환경부', '주관', '수도', '시설', '운영', '관리', '실태', '점검', '우수', '기관', '선정', '수돗물', '안심', '확인', '수돗물', '음용률', '향상', '친환경', '정부', '정책', '이행', '평가', '신아일보', '양평'], ['경기도', '오산시', '환경부', '주관', '수도', '사업', '운영', '관리', '실태', '평가', '연속', '최우수', '기관', '독물', '음용률', '온라인', '홍보', '자체', '평가', '점수', '수돗물', '공급', '과정'], ['감안', '수돗물', '음용률', '온라인', '홍보', '강화', '수돗물', '유충', '대응', '전국', '우수', '사례', '소개', '지난', '건설', '협회', '경기도', '오산시', '운영', '위원회', '기부', '시작', '소재', '어린이집'], ['수도', '사업', '운영', '관리', '실태', '평가', '연속', '최우수', '행안부', '공기업', '평가', '최우수', '경기도', '오산시', '환경부', '감안', '수돗물', '음용률', '온라인', '홍보', '강화', '우수', '시책', '시행', '충청남도', '수돗물', '공급', '수돗물', '음용률', '관리', '경기도', '양주시', '양주시', '시립', '안

결과

```
if __name__ == '__main__':  
    main()
```

```
(1), (643, 1), (644, 1), (645, 5), (646, 1), (647, 1), (648, 1), (649, 1), (650, 1), (651, 1), (652, 1), (653, 3), (654, 2), (655,  
2), (656, 1), (657, 1), (658, 1), (659, 1), (660, 1), (661, 1), (662, 1), (663, 1), (664, 1), (665, 2), (666, 1), (667, 1), (668,  
1), (669, 4), (670, 1), (671, 1), (672, 1), (673, 2), (674, 1), (675, 1), (676, 1), (677, 1), (678, 1), (679, 2), (680, 1), (681,  
1), (682, 1), (683, 1), (684, 1), (685, 1), (686, 2), (687, 1), (688, 1), (689, 1), (690, 1), (691, 1), (692, 1), (693, 1), (694,  
2), (695, 1), (696, 1), (697, 1), (698, 1), (699, 2), (700, 1), (701, 1), (702, 1), (703, 1), (704, 1), (705, 2), (706, 1), (707,  
1), (708, 2), (709, 1), (710, 1), (711, 1), (712, 1), (713, 1), (714, 1), (715, 1), (716, 2), (717, 1), (718, 1), (719, 1), (720,  
1), (721, 1), (722, 2), (723, 1), (724, 1), (725, 2), (726, 1)], [])
```

```
[(0, '0.022*"관리" + 0.020*"수돗물" + 0.016*"스마트" + 0.014*"사업" + 0.012*"수질" + 0.012*"평가" + 0.011*"관망" + 0.011*"상수도"  
+ 0.011*"음용률" + 0.010*"경기도" + 0.009*"시민" + 0.009*"여주시" + 0.009*"수도" + 0.008*"오산시" + 0.007*"공급"'), (1, '0.027*"관  
리" + 0.018*"수돗물" + 0.018*"사업" + 0.018*"평가" + 0.013*"스마트" + 0.013*"운영" + 0.012*"상수도" + 0.011*"수도" + 0.010*"수질"  
+ 0.010*"관망" + 0.009*"오산시" + 0.009*"시민" + 0.009*"음용률" + 0.007*"실태" + 0.007*"환경부"'), (2, '0.022*"관리" + 0.021*"수돗  
물" + 0.017*"평가" + 0.015*"스마트" + 0.015*"사업" + 0.014*"오산시" + 0.012*"상수도" + 0.012*"운영" + 0.011*"수도" + 0.009*"관망"  
+ 0.009*"전국" + 0.009*"시민" + 0.008*"복지" + 0.007*"구축" + 0.007*"음용률"'), (3, '0.022*"관리" + 0.021*"수돗물" + 0.021*"평가"  
+ 0.019*"사업" + 0.014*"수도" + 0.013*"상수도" + 0.012*"스마트" + 0.012*"운영" + 0.010*"오산시" + 0.010*"시민" + 0.010*"관망" + 0.  
008*"수질" + 0.008*"여주시" + 0.008*"환경부" + 0.007*"실태"')]
```

LDA토픽모델 완료=====

[2021-12-22 14:42:16.333414] URL Request Success

지자체명을 입력하세요>>

결과

```
if __name__ == '__main__':  
    main()
```

고령군 0.010* 남양주시 0.000* 고령군 0.000* 양평군 0.001* 양평군

LDA토픽모델 완료=====

[2021-12-22 14:36:11.253041] URL Request Success

거제시

고령군

광주시

단양군

동두천시

봉화군

사천시

양주시

예천군

완도군

진도군

청송군

통영시

파주시

함평군

지자체명을 입력하세요>>

결과

```
if __name__ == '__main__':  
    main()
```

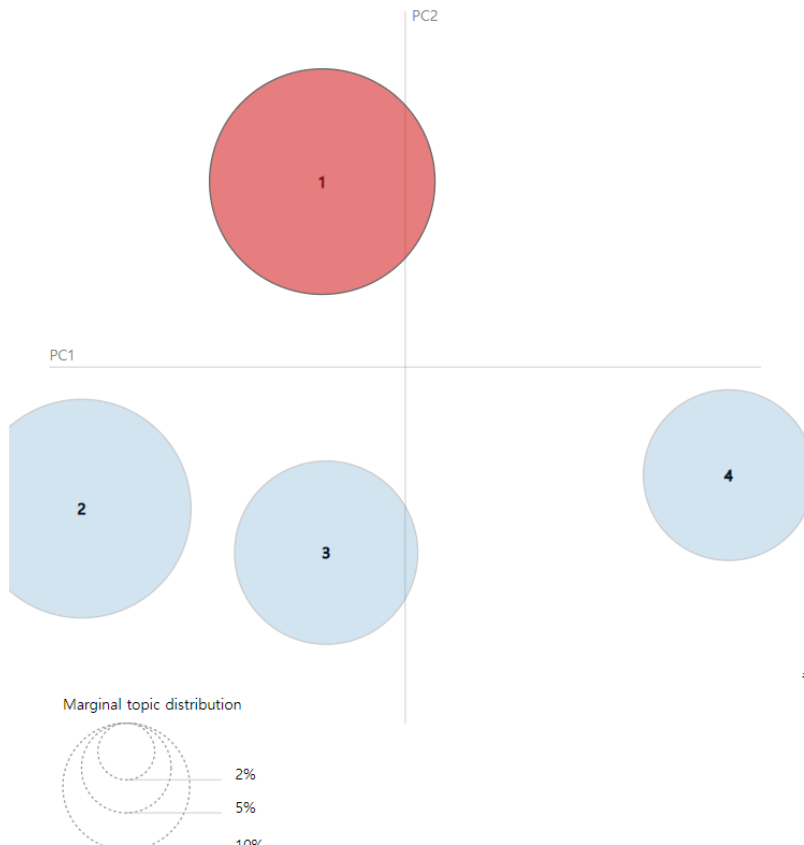
```
문산군  
청송군  
통영시  
파주시  
함평군  
지자체명을 입력하세요>>파주시  
[2021-12-22 14:36:39.605729] URL Request Success  
[2021-12-22 14:36:39.904357] URL Request Success  
문산정수장  
정수장명을 입력하세요>>문산정수장  
[2021-12-22 14:36:44.352564] URL Request Success  
1. 일일 수질정보  
2. 주간 수질정보  
3. 월간 수질정보  
4. 상수원수 수질정보  
원하는 정보를 선택하십시오>>1  
시작 날짜(yyyyMMdd)>>20200101  
끝 날짜(yyyyMMdd)>>20200107  
[2021-12-22 14:36:55.510060] URL Request Success  
수질정보Saved
```

결과

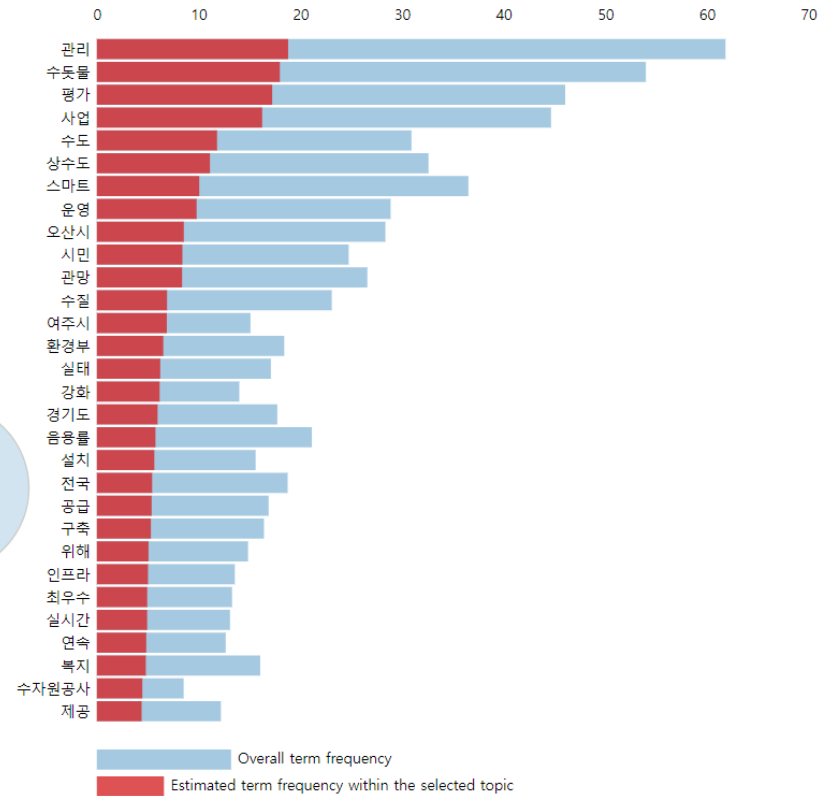
Selected Topic:

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (31.5% of tokens)



1. $\text{saliency}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al (2012)
2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

4. 어려웠던 점/ 참고 자료

어려웠던 점

1) 팀프로젝트로 인한 버전 차이&import의 문제

Ex) pyLDAvis.gensim -> pyLDAvis.gensim_model 명명 변경

2) 공용api/사용 관련 문제

- api를 제공하는 서버에서 문제 발생 빈번
- 서버문제로 인한 url요청 무산
- 첫 url요청은 가능하나 데이터를 요청하는 과정에서 빈번한 오류 발생
- 일일요청건수가 초과하는 경우

어려웠던 점

```
SystemError                                Traceback (most recent call last)
<ipython-input-1-4e24e7d26684> in <module>
      1 from konlpy.tag import Okt
----> 2 okt = Okt()
      3

SystemError: java.nio.file.InvalidPathException: Illegal char <*>
```

3) Konlpy같은 특정 모듈 설치의 어려움

문제> konlpy 설치가 됐음에도 okt 모듈을 찾지 못함.

해결방안> 자바 경로의 중복 문제.

시스템 환경변수에서 JAVA_HOME을 제외한 모든 jdk 경로를 삭제.

Jdk 경로를 삭제하지 않을 시 konlpy를 재설치해도 정상 작동을 하지 않음.

jvm.py안의 java sources에 *를 지우고, Twitter(Okt)의 마지막에도 *가 존재하는데 이 *가 자바 중복 경로 문제로 생성된다고 생각.

어려웠던 점

4) 뉴스기사 본문을 크롤링할 때의 문제

특정한 뉴스 페이지들의 자동화된 데이터 크롤링 차단

해결방안 > 사용자 agency를 구글로 설정하여 들어감

5) 데이터 크롤링 + 분석 시 걸리는 소요 시간

- 모델을 통하여 학습을 시킬 경우에 해당 시간이 상당히 오래 소요됨
- 컴퓨터 사양의 문제로 데이터를 조금 축소하여 진행

참고자료

활용 데이터

- 한국수자원공사_지방상수도 수질

<https://www.data.go.kr/data/3072596/openapi.do>

- 한국수자원_광역정수장 수질

<https://www.data.go.kr/data/3072578/openapi.do>

뉴스

- https://www.hani.co.kr/arti/society/society_general/901677.html
- <http://www.jjan.kr/news/articleView.html?idxno=1142948>
- <http://www.segye.com/newsView/20211007519972?OutUrl=naver>