

국민 청원 데이터 가져오기

학습 목표

- 국민 청원 사이트의 정보를 가져와보기
 - URL : '<https://www1.president.go.kr/petitions>' (<https://www1.president.go.kr/petitions>)

In [109]:

```
from selenium import webdriver

path = '/Users/toto/Documents/GitHub/PythonBasic/part03_1A_web/class01_02_selenium/unit02_15_국민청원'
driver = webdriver.Chrome(path + "/chromedriver_mac_100")
```

In [110]:

```
url = 'https://www1.president.go.kr/petitions'
# url 열기
driver.get(url)
```

In [111]:

```
from selenium.webdriver.common.by import By
```

In [112]:

```
xpath1 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[1]'
xpath2 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[2]'
xpath3 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[3]'
xpath4 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[4]'

sel_category = driver.find_elements(By.XPATH, xpath1)
sel_title = driver.find_elements(By.XPATH, xpath2)
sel_date = driver.find_elements(By.XPATH, xpath3)
sel_peoples = driver.find_elements(By.XPATH, xpath4)
```

In [113]:

```
category = []
title = []
date = []
peoples = []

for one in sel_category:
    category.append(one.text)

for one in sel_title:
    title.append(one.text)

for one in sel_date:
    date.append(one.text)

for one in sel_peoples:
    peoples.append(one.text)

print(category)
print(title)
print(date)
print(peoples)
```

```
['정치개혁', '정치개혁', '보건복지', '교통/건축/국토', '교통/건축/국토', '행정', '문화/예술/체육/언론']
['국회의 입법권 완전 박탈을 요구합니다!!', '검수완박 입법 반대합니다', '박신부작용', '가덕도신공항의 예비타당성조사 면제를 반대합니다.', '거실앞에 실외기가 160개 믿어지십니까?', '***** 조합장등 수사를 빨리 진행해주시고 엄벌에 처해주세요. 조합원들이 고통으로 기다리다 죽어갑니다.', '대한민국에서 유일한 국민들의 힐링문화공간 지켜주세요']
['2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22']
['291명', '1,612명', '197명', '816명', '178명', '230명', '260명']
```

In [114]:

```
# 다른 방법
dat_info = []

for info1 in [sel_category, sel_title, sel_date, sel_peoples]:
    all_data = []
    for one in info1:
        all_data.append(one.text)

    dat_info.append(all_data)

print(dat_info[0])
print(dat_info[1])
print(dat_info[2])
print(dat_info[3])
```

```
['정치개혁', '정치개혁', '보건복지', '교통/건축/국토', '교통/건축/국토', '행정', '문화/예술/체육/언론']
['국회의 입법권 완전 박탈을 요구합니다!!', '검수완박 입법 반대합니다', '박신부작용', '가덕도신공항의 예비타당성조사 면제를 반대합니다.', '거실앞에 실외기가 160개 믿어지 십니까?', '***** 조합장등 수사를 빨리 진행해주시고 엄벌에 처해주세요. 조합원들이 명령으로 기다리다 죽어갑니다.', '대한민국에서 유일한 국민들의 힐링문화공간 지켜주세요']
['2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22', '2022-05-22']
['291명', '1,612명', '197명', '816명', '178명', '230명', '260명']
```

여러 페이지 가져오기

In [115]:

```
# 2페이지 //*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[2]
# 3페이지 //*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[3]
# ...
# 10페이지 //*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[10]
num = 2
page_xpath = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[%d]'
print(page_xpath)
```

```
//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[2]
```

In [116]:

```
import time
```

In [117]:

```
# 페이지 확인
for i in range(2,5):
    xpath = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[%d]' % i
    sel_btn = driver.find_element(By.XPATH, xpath)
    sel_btn.click()
    time.sleep(3)
```

In [118]:

```
xpath1 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[1]/a[2]'
xpath2 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[2]/div[2]/a[2]'
xpath3 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[3]/div[2]/a[2]'
xpath4 = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li/div/div[4]/div[2]/a[2]'
```

In [119]:

```
# 10페이지의 정보 가져오기
# # //*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[2]
# # //*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[3]
# # //*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[1]/div/div[1]
# for i in range(2,11):
#     xpath = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[1]/a[%d]' % i
#     sel_btn = driver.find_element(By.XPATH, xpath)
#     sel_btn.click()

#     time.sleep(3)

#     sel_category = driver.find_elements(By.XPATH, xpath1)
#     sel_title = driver.find_elements(By.XPATH, xpath2)
#     sel_date = driver.find_elements(By.XPATH, xpath3)
#     sel_peoples = driver.find_elements(By.XPATH, xpath4)

#     dat_info = []

#     for info1 in [sel_category, sel_title, sel_date, sel_peoples]:
#         all_data = []
#         for one in info1:
#             all_data.append(one.text)

#     dat_info.append(all_data)

#     print(dat_info[0])
#     print(dat_info[1])
#     print(dat_info[2])
#     print(dat_info[3])
#     print()
```

In [123]:

```
### 10페이지 이상인 경우(50페이지 가져오기)
cnt = 1
f_cnt = 1

dat_cat = []
dat_title = []
dat_date = []
dat_peoples = []

for i in range(2,20):
    if i==11: # NEXT 버튼 # 10페이지의 NEXT버튼
        cnt = 1
        f_cnt = 2
        xpath = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[%d]/a'
    elif i==21 or i==31 or i==41: # 20페이지, 30페이지 이후의 NEXT버튼
        cnt = 1
        f_cnt = 3
        xpath = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[%d]/a'
    else:
        cnt +=1
        if f_cnt==2 or f_cnt==3: # 20페이지 이상은 f_cnt가 2로 동일
            f_cnt=2
        xpath = '//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[4]/div/div[%d]/a[%d]'

    sel_btn = driver.find_element(By.XPATH, xpath)
    sel_btn.click()

    time.sleep(3)

    sel_category = driver.find_elements(By.XPATH, xpath1)
    sel_title = driver.find_elements(By.XPATH, xpath2)
    sel_date = driver.find_elements(By.XPATH, xpath3)
    sel_peoples = driver.find_elements(By.XPATH, xpath4)

    dat_info = []
    for info1 in [sel_category, sel_title, sel_date, sel_peoples]:
        all_data = []
        for one in info1:
            all_data.append(one.text)

        dat_info.append(all_data)

    dat_cat.extend(dat_info[0])
    dat_title.extend(dat_info[1])
    dat_date.extend(dat_info[2])
    dat_peoples.extend(dat_info[3])

    print(dat_info[0])
    print(dat_info[1])
    print(dat_info[2])
    print(dat_info[3])
    print()
```

['행정', '육아/교육', '행정', '기타', '교통/건축/국토', '안전/환경', '인권/성평
등']
['소상공인존 도와주세요 소상공인 손실보상금 문제', '놀권리를 침해하는 주말 학원보
충을 반대합니다.', '고 ***군 의문사 검찰전면 재수사촉구', '가평계곡살인 ***와 ***

를 법정최고형을 촉구합니다', '발전 공기업인 ****발전의 중소기업에 대한 갑질을 멈춰주세요', '50대 의사 살인 후 발에 시신유기한 40대 여성 신상공개 촉구합니다', '청와대는 말씀하신 답변을 지켜 주십시오!']
['2022-05-22', '2022-05-22', '2022-05-21', '2022-05-21', '2022-05-21', '2022-05-21', '2022-05-21']
['120명', '112명', '605명', '747명', '411명', '350명', '448명']

['외교/통일/국방', '보건복지', '인권/성평등', '교통/건축/국토', '반려동물', '반려동물', '보건복지']
['무능력하고 무책임한 상해 영사관을 고발합니다', '****병원의 직무과실로 인한 코로나 조치 미비로 인한 아버지의 억울한 죽음을 호소합니다', '억울하게 성추행범으로 몰린 중학교 2학년 아들을 도와주세요', '불필요한 구간 단속', '유기견, 유기묘 안락사를 멈춰주세요!', '제주도 강아지를 묶에 땅에 묻어 유기한 유기범을 검거하고 처벌해주세요.', '백신 부작용으로 인해 고통받고 있습니다']
['2022-05-21', '2022-05-21', '2022-05-21', '2022-05-21', '2022-05-21', '2022-05-21', '2022-05-21']

csv파일 또는 xlsx파일로 출력

In [124]:

```
import pandas as pd

dat_dic = {"분류":dat_cat, "제목":dat_title,
           "청원 만료일":dat_date, "참여인원":dat_peoples}

dat_df = pd.DataFrame(dat_dic)
print("데이터 행열 :", dat_df.shape)
dat_df.head()
```

데이터 행열 : (126, 4)

Out[124]:

	분류	제목	청원 만료일	참여인원
0	행정	소상공인존 도와주세요 소상공인 손실보상금 문제	2022-05-22	120명
1	육아/교육	놀권리를 침해하는 주말 학원보충을 반대합니다.	2022-05-22	112명
2	행정	고 ***군 의문사 검찰전면 재수사촉구	2022-05-21	605명
3	기타	가평계곡살인 ***와 ***를 법정최고형을 촉구합니다	2022-05-21	747명
4	교통/건축/국토	발전 공기업인 ****발전의 중소기업에 대한 갑질을 멈춰주세요	2022-05-21	411명

In [125]:

```
from datetime import datetime

date_now = datetime.today().strftime('%Y%m%d')
file_name = "national_petition_" + date_now
dat_df.to_csv(file_name + ".csv")
dat_df.to_excel(file_name + ".xlsx")
```

In []:

