

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 [3]:

```
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")
```

- 에러 발생 시,
 - 에러 메시지 : 개발자를 확인할 수 없기 때문에 'chromedriver' 을 열수 없습니다.
- 해결
 - xattr -d com.apple.quarantine chromedriver

In [4]:

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

답변 대기 중인 청원 가져오기

- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[1]/div/div[1]
- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[2]/div/div[1]
- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[3]/div/div[1]
- ...

In [6]:

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

In [8]:

```
xpath = '//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li/div/div[1]'
sel_category = driver.find_elements(By.XPATH, xpath)
sel_category
```

Out[8]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="77ccb9595030fb05a3fb6c798b0eadc6", element="50d92ed8-ebaa-4254-90d9-fa92502c0f4a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="77ccb9595030fb05a3fb6c798b0eadc6", element="6d3d4d9c-2165-457c-9460-051f85910a02")>,
 <selenium.webdriver.remote.webelement.WebElement (session="77ccb9595030fb05a3fb6c798b0eadc6", element="691c6a67-244a-4da2-bc17-d238ec93a42d")>,
 <selenium.webdriver.remote.webelement.WebElement (session="77ccb9595030fb05a3fb6c798b0eadc6", element="43948fea-28e9-4d94-b120-ca1ae26bf6e0")>,
 <selenium.webdriver.remote.webelement.WebElement (session="77ccb9595030fb05a3fb6c798b0eadc6", element="1b613976-53cf-4efe-8863-ff167a9e9b3c")>,
 <selenium.webdriver.remote.webelement.WebElement (session="77ccb9595030fb05a3fb6c798b0eadc6", element="5ed0fac6-5a7e-4105-a35e-e0eda6268b6c")>,
 <selenium.webdriver.remote.webelement.WebElement (session="77ccb9595030fb05a3fb6c798b0eadc6", element="0d0eff9d-82cc-4952-a77b-50d23cce430a")>]
```

In [9]:

```
for one in sel_category:
    print(one.text)
```

정치개혁
정치개혁
정치개혁
외교/통일/국방
기타
보건복지
정치개혁

답변 대기 중인 청원 - 제목, 청원 만료일, 참여인원

제목

- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[1]/div/div[2]/a
- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[2]/div/div[2]/a
- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[3]/div/div[2]/a
- ..

청원 만료일

- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[1]/div/div[3]
- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[2]/div/div[3]
- //*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[3]/div/div[3]
- ..

참여 인원

- `//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[1]/div/div[4]`
- `//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[2]/div/div[4]`
- `//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li[3]/div/div[4]`
- ..

In [20]:

```
### 분류, 제목, 청원 만료일, 참여 인원
```

In [11]:

```
xpath1 = '//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li/div/div[1]'
xpath2 = '//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li/div/div[2]'
xpath3 = '//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li/div/div[3]'
xpath4 = '//*[@id="cont_view"]/div[2]/div/div/div[1]/div[6]/div/div[2]/div[2]/ul/li/div/div[4]'
```

In [12]:

```
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 [14]:

```
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', '정치개혁', '외교/통일/국방', '기타', '보건복지', '정치개혁']
['윤석열 당선자의 대장동 부산저축은행 부실수사 봐주기 의혹과 김건희의 주가조작 실체의 진상조사 확인을 위한 청원입니다.', '대구 동구 부정선거 및 선거관리위원회의 관리 부실로 대선 재투표를 희망합니다.', '이명박 전 대통령 사면을 반대합니다.', '윤석열 당선인 직무실 만들고자, 국가안전 중추인 국방부를 강압 이전하여, 국민의 혈세 수천억을 날리는 것을 막아주시요', '문재인 대통령님 사랑합니다.', '의료 민영화의 첫 걸음이 될 제주 영리병원을 국가가 매수해주십시오.', '제 20대 대통령의 직무실 이전에 반대합니다.']
['22.04.10', '22.04.09', '22.04.14', '22.04.16', '22.04.09', '22.04.13', '22.04.16']
['548,720명', '265,060명', '355,501명', '544,898명', '287,985명', '220,047명', '212,122명']
```

In [17]:

```
import pandas as pd
```

In [18]:

```
df = { "분류":category, "제목":title, "청원 만료일":date, "청원 참여":peoples}
dat_df = pd.DataFrame(df)
dat_df.to_csv("답변대기중청원.csv", index=False)
```

전체 목록 정보 가져오기

분류, 제목, 청원 만료일, 참여 인원

- 분류
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[1]/div/div[1]`
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[2]/div/div[1]`
 - ...
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[7]/div/div[1]`
- 제목

- `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[1]/div/div[2]/a`
- `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[2]/div/div[2]/a`
- ...
- `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[7]/div/div[2]/a`
- 청원 만료일
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[1]/div/div[3]`
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[2]/div/div[3]`
 - ...
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[7]/div/div[3]`
- 참여 인원
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[1]/div/div[4]`
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[2]/div/div[4]`
 - ...
 - `//*[@id="cont_view"]/div[2]/div/div/div[2]/div[2]/div[4]/div/div[2]/div[2]/ul/li[7]/div/div[4]`

In [22]:

```

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 [23]:

```
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 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']

['279명', '1,578명', '188명', '781명', '172명', '226명', '254명']