

코로나 데이터 웹 데이터 수집

학습내용

- 추가적인 내용을 수집한다.(인구수, 위중증 추가)
- 금일 정보가 아닌 어제기준 완료된 데이터를 가져와 본다.
- 코로나 19(COVID-19) 실시간 상황판
- url : <https://coronaboard.kr/> (<https://coronaboard.kr/>)

In [1]:



```
from IPython.display import display, Image
import os, warnings
import re
warnings.filterwarnings(action='ignore')
```

01 웹 브라우저 띄우기

In [2]:



```
from selenium import webdriver
from bs4 import BeautifulSoup
```

In [3]:



```
driver = webdriver.Chrome('./chromedriver_91')

## https://www.amazon.com/
url = 'https://coronaboard.kr/'
driver.get(url)

import time
time.sleep(3) # 홈페이지 로딩 시간 3초
```

02 더보기 버튼 찾기

In [4]:



```
from selenium.webdriver import ActionChains

some_tag = driver.find_element_by_id('show-more')
ActionChains(driver).move_to_element(some_tag).perform()
some_tag.click()
```

In [5]:



```
some_tag = driver.find_element_by_id('show-more')
ActionChains(driver).move_to_element(some_tag).perform()
some_tag.click()
```

03 추가정보선택

- 콤보박스 : `//*[@id="global-slide"]/div/div[2]/div/div/button/div/div/div`

In [6]:



```
combo_xpath = '//*[@id="global-slide"]/div/div[2]/div/div/button/div/div/div'
some_tag = driver.find_element_by_xpath(combo_xpath)
ActionChains(driver).move_to_element(some_tag).perform()
some_tag.click()
```

- `//*[@id="bs-select-1-5"]`

치료중 추가

In [7]:



```
sel_menu = driver.find_element_by_xpath('//*[@id="bs-select-1-0"]')
sel_menu.click()
```

인구수 추가

In [8]:



```
sel_menu = driver.find_element_by_xpath('//*[@id="bs-select-1-5"]')
sel_menu.click()
```

위중증 추가

In [9]:



```
sel_menu = driver.find_element_by_xpath('//*[@id="bs-select-1-1"]')
sel_menu.click()
```

In [10]:



```
some_tag.click()
```

국가 정보 가져오기

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[2]/a`
- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[2]/a`
- `//*[@id="country-table"]/div/div/table/tbody/tr[4]/td[2]/a`

In [11]:



```
country_xpath = '//*[@id="country-table"]/div/div/table/tbody/tr/td[2]'
sel_country = driver.find_elements_by_xpath(country_xpath)
len(sel_country)
```

Out[11]:

222

In [12]:



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

터키TR
프랑스FR
이란IR
아르헨티나AR
콜롬비아CO
스페인ES
이탈리아IT
독일DE
인도네시아ID
멕시코MX
폴란드PL
남아프리카 공화국ZA
필리핀PH
우크라이나UA
말레이시아MY
페루PE
네덜란드NL
이라크IQ
일본JP*
체코CZ

확진자 정보

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[3]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[3]`
- ..
- `//*[@id="country-table"]/div/div/table/tbody/tr[5]/td[3]`

치료중

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[4]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[3]/td[4]`
- ..

위중증

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[5]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[3]/td[5]`
- ..

사망자 정보

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[6]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[6]`
- ...

완치

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[7]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[7]`
- ..

치명(%)

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[8]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[8]`
- ...

완치(%)

- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[9]`

발생률

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[10]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[10]`
- ...

인구수

- `//*[@id="country-table"]/div/div/table/tbody/tr[1]/td[11]`
- `//*[@id="country-table"]/div/div/table/tbody/tr[2]/td[11]`
- ...

어제 오늘 선택

In [13]:



```
# 어제 : //*[@id="global-slide"]/div/div[2]/ul/li[2]/a
flag_today = False # False : 어제, Today : 오늘
```

In [14]:



```
# 어제일 경우에만 버튼을 선택
if flag_today==False:
    sel_yesterday = driver.find_element_by_xpath(' //*[@id="global-slide"]/div/div[2]/ul/li[2]/a')
    sel_yesterday.click()
```

전체 데이터 가져오기

In [15]:



```
all_data = []
for i in range(2, 12):
    tmp = '//*[@id="country-table"]/div/div/table/tbody/tr/td[{}]'.format(i)
    # print(tmp)
    sel_ele = driver.find_elements_by_xpath(tmp)

    column_data = []
    for one in sel_ele:
        # print(one.text)
        column_data.append(one.text)

    print( len(sel_ele) )
    all_data.append(column_data)
    print(column_data)
    print()
```

all_data

222

['미국US', '인도IN', '브라질BR', '영국GB', '러시아RU', '터키TR', '프랑스FR', '이란IR', '아르헨티나AR', '콜롬비아CO', '스페인ES', '이탈리아IT', '독일DE', '인도네시아ID', '멕시코MX', '폴란드PL', '남아프리카 공화국ZA', '필리핀PH', '우크라이나UA', '말레이시아MY', '페루PE', '네덜란드NL', '이라크IQ', '일본JP*', '체코CZ', '칠레CL', '캐나다CA', '태국TH', '방글라데시BD', '이스라엘IL', '파키스탄PK', '벨기에BE', '루마니아RO', '스웨덴SE', '포르투갈PT', '세르비아RS', '모로코MA', '카자흐스탄KZ', '쿠바CU', '스위스CH', '헝가리HU', '요르단JO', '네팔NP', '베트남VN', '오스트리아AT', '아랍에미리트AE', '튀니지TN', '그리스GR', '레바논LB', '조지아GE', '과테말라GT', '사우디아라비아SA', '벨라루스BY', '코스타리카CR', '스리랑카LK', '에콰도르EC', '볼리비아BO', '불가리아BG', '아제르바이잔AZ', '파나마PA', '미얀마MM', '파라과이PY', '쿠웨이트KW', '슬로바키아SK', '크로아티아HR', '팔레스타인PS', '우루과이UY', '아일랜드IE', '베네수엘라VE', '온두라스HN', '덴마크DK', '도미니카 공화국DO', '에티오피아ET', '리비아LY', '리투아니아LT', '대한민국KR', '오만OM', '이집트EG', '몽골MN', '슬로베니아SI', '몰도바MD', '바레인BH', '아르메니아AM', '케냐KE', '카타르QA', '보스니아 헤르체고비나BA', '잠비아ZM', '나이지리아NG', '알제리DZ', '북마케도니아MK', '노르웨이NO', '키르기스스탄KG', '보츠와나BW', '우즈베키스탄UZ', '알바니아AL', '라트비아LV', '에스토니아EE', '아프가니스탄AF', '모잠비크MZ', '핀란드FI', '몬테네그로ME', '짐바브웨ZW', '나미비아NA', '가나GH', '우간다UG', '키프로스CY', '캄보디아KH', '코소보XK', '엘살바도르SV', '호주AU', '르완

In [18]:



```
print(len(all_data) )
```

In [19]:



```
for i in range(len(all_data)):
    print(len( all_data[i]) )
```

222
222
222
222
222
222
222
222
222
222

In [32]:



```
import pandas as pd

dict_dat = { "국가":all_data[0],
             "확진자":all_data[1],
             "치료중":all_data[2],
             "위중증":all_data[3],
             "사망자":all_data[4],
             "완치":all_data[5],
             "치명(%)":all_data[6],
             "완치(%)":all_data[7],
             "발생률":all_data[8],
             "인구수":all_data[9]
           }

dat = pd.DataFrame(dict_dat)
dat
```

Out[32]:

	국 가	확진자	치료중	위중증	사망자	
0	미 국 US	44,055,692\n(+113,357)	9,823,512\n(-14,871)	21,296\n(-523)	711,222\n(+2,103)	33,520,958\
1	인 도 IN	33,718,984\n(+4,743)	284,976\n(-4,806)	8,944	447,828\n(+129)	32,986,18
2	브 라 질 BR	21,381,790\n(+15,395)	403,027\n(-7,475)	8,318	595,520\n(+818)	20,383,243
3	영 국 GB	7,736,235\n(+34,520)	1,344,510\n(-1,886)	831\n(-47)	136,375\n(+167)	6,255,350
4	러 시 아 RU	7,487,138\n(+22,430)	626,809\n(+3,117)	2,300	206,388\n(+857)	6,653,941
...
217	마 셜 제 도 MH	4	0	N/A	0	
218	바 누 아 투 VU	4	0	N/A	1	

국 가	확진자	치료중	위중증	사망자
219 사 모 아 WS	3	0	N/A	0
220 세 인 트 헬 레 나 SH	2	0	N/A	0
221 미 크 로 네 시 아 FM	1	0	N/A	0

222 rows × 10 columns

주어진 데이터를 좀 더 사용하기 쉽게 전처리한다.

In [33]:



```
### 확진자를 총 확진자와, 일일 확진자로
dat['치료중합계'] = dat['치료중'].str.split('Wn').str[0]
dat['치료중1일'] = dat['치료중'].str.split('Wn').str[1]

dat['위중증합계'] = dat['위중증'].str.split('Wn').str[0]
dat['위중증1일'] = dat['위중증'].str.split('Wn').str[1]

dat['확진자합계'] = dat['확진자'].str.split('Wn').str[0]
dat['확진자1일'] = dat['확진자'].str.split('Wn').str[1]

dat['사망자합계'] = dat['사망자'].str.split('Wn').str[0]
dat['사망자1일'] = dat['사망자'].str.split('Wn').str[1]

dat['완치합계'] = dat['완치'].str.split('Wn').str[0]
dat['완치1일'] = dat['완치'].str.split('Wn').str[1]

dat = dat.drop([ '치료중', '위중증', '확진자', '사망자', '완치'], axis=1)

dat.head()
```

Out[33]:

	국가	치명률 (%)	완치률 (%)	발생률	인구수	치료중합계	치료중1일	위중증합계	위중증1일	확진자합계	확진자1일
0	미국 US	1.6	76.1	133,098	331,002,651	9,823,512	(-14,871)	21,296	(-523)	44,055,692	(+113,300)
1	인도 IN	1.3	97.8	24,434	1,380,004,385	284,976	(-4,806)	8,944	NaN	33,718,984	(+4,700)
2	브라질 BR	2.8	95.3	100,592	212,559,417	403,027	(-7,475)	8,318	NaN	21,381,790	(+15,300)
3	영국 GB	1.8	80.9	113,959	67,886,011	1,344,510	(-1,886)	831	(-47)	7,736,235	(+34,500)
4	러시아 RU	2.8	88.9	51,305	145,934,462	626,809	(+3,117)	2,300	NaN	7,487,138	(+22,400)

In [34]:



```
dat['발생률'] = dat['발생률'].str.replace(pat=r'[,]', repl=r'', regex=True)
dat['인구수'] = dat['인구수'].str.replace(pat=r'[,]', repl=r'', regex=True)

dat['치료중합계'] = dat['치료중합계'].str.replace(pat=r'[,]', repl=r'', regex=True)
dat['치료중1일'] = dat['치료중1일'].str.replace(pat=r'[,()]', repl=r'', regex=True)

dat['위중증합계'] = dat['위중증합계'].str.replace(pat=r'[,]', repl=r'', regex=True)
dat['위중증1일'] = dat['위중증1일'].str.replace(pat=r'[,()]', repl=r'', regex=True)

dat['확진자합계'] = dat['확진자합계'].str.replace(pat=r'[,]', repl=r'', regex=True)
dat['확진자1일'] = dat['확진자1일'].str.replace(pat=r'[,()]', repl=r'', regex=True)

dat['사망자합계'] = dat['사망자합계'].str.replace(pat=r'[,]', repl=r'', regex=True)
dat['사망자1일'] = dat['사망자1일'].str.replace(pat=r'[,()]', repl=r'', regex=True)

dat['완치합계'] = dat['완치합계'].str.replace(pat=r'[,]', repl=r'', regex=True)
dat['완치1일'] = dat['완치1일'].str.replace(pat=r'[,()]', repl=r'', regex=True)
dat.head()
```

Out[34]:

	국 가	치 명 (%)	완치 (%)	발생률	인구수	치료중 합계	치료중 1일	위중 증합 계	위중증1 일	확진자합 계	확진자1 일	사망자1 일
0	미 국 US	1.6	76.1	133098	331002651	9823512	-14871	21296	+113357	44055692	+113357	7
1	인 도 IN	1.3	97.8	24434	1380004385	284976	-4806	8944	+4743	33718984	+4743	4
2	브 라 질 BR	2.8	95.3	100592	212559417	403027	-7475	8318	+15395	21381790	+15395	5
3	영 국 GB	1.8	80.9	113959	67886011	1344510	-1886	831	+34520	7736235	+34520	1
4	러 시 아 RU	2.8	88.9	51305	145934462	626809	+3117	2300	+22430	7487138	+22430	2

In [35]:



```
from datetime import datetime
from datetime import date
import datetime
import os

today = datetime.date.today()
yesterday = today - datetime.timedelta(days=1)

# print(today)
# print(yesterday)
if flag_today == True:
    file_make_time = today
else:
    file_make_time = yesterday

print( file_make_time )

print( os.getcwd() )
path_dir = os.getcwd() + "WWdataWW"
path_file = path_dir + str(file_make_time)
print( path_dir, path_file, sep="Wn" )
```

2021-09-29

C:\Users\Wtoto\Documents\WGithub\Wcorona_analysis

C:\Users\Wtoto\Documents\WGithub\Wcorona_analysis\Wdata\W

C:\Users\Wtoto\Documents\WGithub\Wcorona_analysis\Wdata\W2021-09-29

In [36]:



```
dat.to_csv(path_file + "_corona.csv", index=False)
dat.to_excel(path_file + "_corona.xlsx", index=False)
os.listdir( path_dir )
```

Out[36]:

```
['2021-08-07_corona.csv',
'2021-08-07_corona.xlsx',
'2021-09-19_corona.csv',
'2021-09-19_corona.xlsx',
'2021-09-20_corona.csv',
'2021-09-20_corona.xlsx',
'2021-09-28_corona.csv',
'2021-09-28_corona.xlsx',
'2021-09-29_corona.csv',
'2021-09-29_corona.xlsx',
'20210808_16datamerge.csv',
'20210808_16datamerge.xlsx',
'20210808_16_today_corona.csv',
'20210808_16_today_corona.xlsx',
'20210808_16_vaccine_bloomberg.csv',
'20210808_16_vaccine_bloomberg.xlsx',
'20210920_00_vaccine_bloomberg.csv',
'20210920_00_vaccine_bloomberg.xlsx',
'20210921_00_datamerge.csv',
'20210921_00_datamerge.xlsx',
'20210929_00_vaccine_bloomberg.csv',
'20210929_00_vaccine_bloomberg.xlsx',
'20210929_01_datamerge.csv',
'20210929_01_datamerge.xlsx',
'20210929_14_vaccine_bloomberg.csv',
'20210929_14_vaccine_bloomberg.xlsx',
'20210930_00_datamerge.csv',
'20210930_00_datamerge.xlsx',
'20210930_00_vaccine_bloomberg.csv',
'20210930_00_vaccine_bloomberg.xlsx',
'country.csv',
'country.xlsx']
```

추가 과제

- 위증증 - 총 합계와 일일 합계 나누기
- 확진자/총인구 값 구하기
- history
 - 2021.09.30 version 1.xx
- 출처를 밝히시고 위의 내용에 대해 자유롭게 사용 가능합니다.

In []:



