





- ☑ 위키피디아의 '올림픽 메달 집계' 데이터 내려 받아 특정 테이블 활용
- ☑ 하계 올림픽 메달 집계를 위한 데이터프레임 생성
- ☑ 특정 메달 순서로 정렬하고 '대한민국'이 있는 행 검색



학습목표

- ☆ 위키피디아의 '올림픽 메달 집계' 전체 페이지에서 데이터 읽어와 특정 데이터프레임을준비할 수 있다.
- ☑ 하계 올림픽 메달 집계를 위한 데이터프레임을 생성할 수 있다.
- ☑ 특정 메달 순서로 정렬하고 '대한민국'이 있는 행 검색할 수 있다.

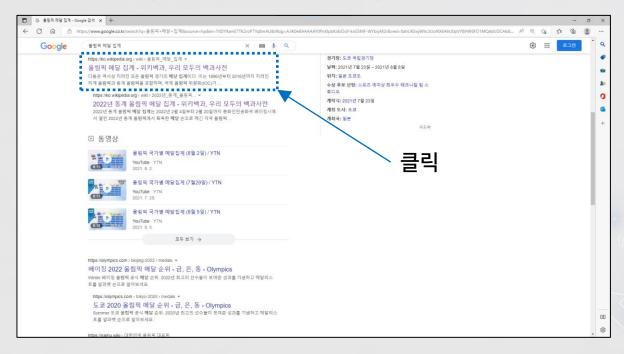
LESSON 01

인터넷 자료를 🥨 DataFrame에 저장





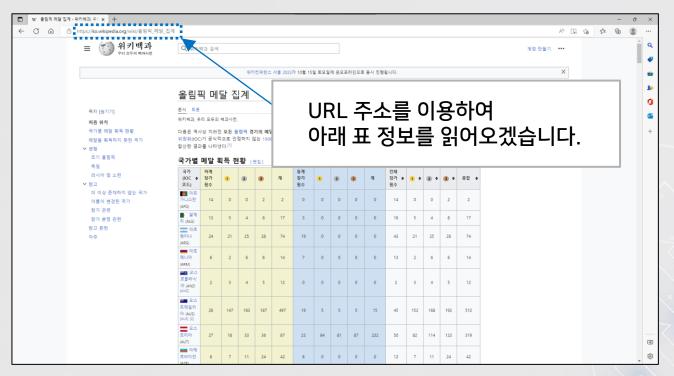
- → 올림픽 메달 기록 데이터 (1/2)
- ❤️ pandas 라이브러리 실습을 위해 올림픽 메달 기록 데이터를 활용하겠습니다.
- ② Google에서 "올림픽 메달 집계"라고 검색 → "올림픽 메달 집계 위키백과, 우리 모두의 백과사전" 검색 결과 클릭





→ 올림픽 메달 기록 데이터 (2/2)





인데 인터넷 자료를 DataFrame에 저장



→ pandas를 이용하여 URL로부터 데이터 읽어 오기

```
import pandas as pd
df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84')
print(df)
           국가 (IOC 코드) 하계 참가 횟수
                                    Unnamed: 2
         아프가니스탄 (AFG)
           알제리 (ALG)
          아르헨티니 (ARG)
          아르메니아 (ARM
    오스트랄라시아 (ANZ) [ANZ]
                                                     대괄호 []가 있습니다!
                                              그 뜻은 자료형이 리스트라는 거겠죠?
                                                  df의 구성을 살펴보겠습니다.
    러시아 출신 올림픽 선수 (OAR)
151
        혼성 (ZZX) [ZZX]
                                    8
152
                 총합
                                  5116
                                                    343U 13000
                                                전체 참가 횟수 ₩
    동계 참가 횟수 Unnamed: 7 Unnamed: 8 Unnamed:
                                                43
                                                13
```

인터넷 자료를 DataFrame에 저장



→ DataFrame의 리스트



UI 인터넷 자료를 DataFrame에 저장



→ 데이터 프레임 df[0] 살펴보기

import pandas as pd df = pd.read html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD %EB%A9%94%EB%8B%AC %EC%A7%91%EA%B3%84') df [0] 하 전 체 국가 Unnamed: Unnamed: Unnamed: Unnamed: Unnamed: Unnamed: (IOC 코 드) 아프가니 스탄 0 2 0 14 0 0 2 (AFG) 5 4 0 5 4 21 21 25 0 28 74 2 6 0 2 6 14 0 13 14 아 (ARM) 오스트랄 라시아 3 (ANZ) [ANZ]

df[0]에 우리가 원하는 데이터가 담겨있네요!

[21] 민터넷 자료를 DataFrame에 저장

인덱스



→ 인덱스(Index)와 열 이름(Column Name) 확인

	impo	rt pandas as pd															
	df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84') df[0]																
		국가 (IOC 코 드)	하 계 참 가 횟 수	Unnamed: 2	Unnamed: 3	Unnamed: 4	계	동 계 참 가 횟 수	Unnamed: 7	Unnamed: 8	Unnamed: 9	계.1	전 체 참 가 횟 수	Unnamed: 12	Unnamed: 13	Unnamed: 14	총합
	0	아프가니 스탄 (AFG)	14	0	0	2	2	0	0	0	0	0	14	0	0	2	2
	1	알제리 (ALG)	13	5	4	8	17	3	0	0	0	0	16	5	4	8	17
	2	아르헨티 나 (ARG)	24	21	25	28	74	19	0	0	0	0	43	21	25	28	74
	- 3	아르메니 아 (ARM)	6	2	6	6	14	7	0	0	0	0	13	2	6	6	14
***************************************	4	오스트랄 라시아 (ANZ) [ANZ]	2	3	4	5	12	0	0	0	0	0	2	3	4	5	12
L																	

열 이름

UI 인터넷 자료를 DataFrame에 저장



→ 인덱스를 '국가 (IOC 코드)'로 변경하기

import pandas as pd df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84') df2 = df[0].set_index('국가 (10C 코드)') df2 Unnamed: Unnamed: Unnamed: Unnamed: Unnamed: Unnamed: Unnamed: Unnamed: Unnamed: 국가 (IOC 코드) 아프가니스 0 0 탄 (AFG) 알제리 13 5 0 0 17 (ALG) 아르헨티나 21 74 19 25 0 0 21 28 74 (ARG) 아르메니아 0 2 0 13 6 14 (ARM) 오스트랄라 3 12 0 0 0 12 시아 (ANZ) [ANZ]

인덱스가 "국가 (IOC 코드)"로 변경되었습니다.

인터넷 자료를 DataFrame에 저장



→ 하계 정보만 추출하기

```
import pandas as pd
df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84')
df2 = df[0].set index('국가 (100 코드)')
summer = df2.iloc[:, :5]
# iloc은 integer location의 약어로, 데이터 프레임의 행이나 열의 순서를 나타내는 정수로 특정 값을 추출합니다.
                        하계 참가 횟수 Unnamed: 2 Unnamed: 3 Unnamed: 4
            국가 (IOC 코드)
        아프가니스탄 (AFG)
                                14
             알제리 (ALG)
                                13
                                           5
                                                                    17
          아르헨티나 (ARG)
                                24
                                          21
                                6
                                           2
          아르메니아 (ARM)
  오스트랄라시아 (ANZ) [ANZ]
           독립 (IOA) [IOA]
        독립 참가 (IOP) [IOP]
                                           0
 러시아 출신 올림픽 선수 (OAR)
                                0
                                           0
          혼성 (ZZX) [ZZX]
                                3
                                           8
                   총합
                                28
                                        5116
                                                  5082
                                                            5490 15688
153 rows × 5 columns
```

iloc() 함수를 활용하여 하계(Summer) 정보만 추출하였습니다.

UI 인터넷 자료를 DataFrame에 저장



→ 컬럼 이름 설정하기

```
import pandas as pd
df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84')
df2 = df[0].set index('국가 (100 코드)')
summer = df2.iloc[:, :5]
summer.columns = ['경기수', '금', '은', '동', '합계']
                        경기수
                                              합계
                                        동
            국가 (IOC 코드)
        아프가니스탄 (AFG)
             알제리 (ALG)
          아르헨티나 (ARG)
                                                                            2
         아르메니아 (ARM)
  오스트랄라시아 (ANZ) [ANZ]
           독립 (IOA) [IOA]
        독립 참가 (IOP) [IOP]
 러시아 출신 올림픽 선수 (OAR)
          혼성 (ZZX) [ZZX]
                           28 5116 5082 5490 15688
153 rows × 5 columns
```

데이터 프레임의 columns에 컬럼 이름을 설정하였습니다.

민터넷 자료를 DataFrame에 저장



→ 내림차순으로 정렬하기

```
import pandas as pd
df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84')
df2 = df[0].set index('국가 (100 코드)')
summer = df2.iloc[:, :5]
summer.columns = ['경기수', '금', '은', '동', '합계']
summer = summer.sort_values('금', ascending=False)
summer
               국가 (IOC 코드)
                               28 5116 5082 5490 15688
                               27 1022 795 706 2523
     미국 (USA) [P] [Q] [R] [Z] [F]
             소련 (URS) [URS]
                                9 395 319 296 1010
          영국 (GBR) [GBR] [Z]
                               28 263 295 293
     중화인민공화국 (CHN) [CHN]
                               10 224 167 155
                 레바논 (LIB)
 세르비아 몬테네그로 (SCG) [SCG]
               지부티 (DJI) [B]
               키프로스 (CYP)
            아프가니스탄 (AFG)
                               14 0 0 2 2
153 rows × 5 columns
```

sort_values() 함수를 이용하면 원하는 열을 기준으로 데이터 순서를 정렬할 수 있습니다.

인터넷 자료를 DataFrame에 저장



⊸ 엑셀 파일로 저장하기

import pandas as pd

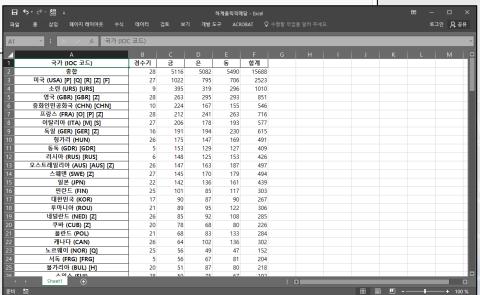
df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84')

df2 = df[0].set_index('국가 (100 코드)')

summer = df2.iloc[:, :5]

summer - columns = ['경기수', '금', '은', '동', '합계'] summer = summer.sort_values('금', ascending=False) summer.to_excel('하계올림픽메달.xlsx')

to_excel() 함수를 이용하여 작업했던 "summer" 데이터 프레임을 엑셀 파일로 저장합니다.



LESSON 02

pandas의 (DataFrame 햄 검색



☑근 pandas의 DataFrame 행 검색



→ 엑셀 파일 읽어 오기 index_col=0

```
import pandas as pd
  df = pd.read_excel('하계올림픽메달.xlsx', index_col=0)
  df.head()
✓ 0.0s
                                                                       Python
                                               동
                         경기수
                                   금
                                                    합계
          국가 (IOC 코드)
                   총합
                                5116
                                       5082
                                             5490
                                                   15688
 미국 (USA) [P] [Q] [R] [Z] [F]
                                 1022
                                                   2523
                                        795
                                              706
          소련 (URS) [URS]
                                  395
                                        319
                                             296
                                                    1010
       영국 (GBR) [GBR] [Z]
                            28
                                  263
                                             293
                                                    851
                                        295
중화인민공화국 (CHN) [CHN]
                             10
                                  224
                                        167
                                              155
                                                     546
```

🛂 pandas의 DataFrame 행 검색



→ 엑셀 파일 읽어 오기

import pandas as pd df = pd.read_excel('하계올림픽메달.xlsx') df.head() 국가 (IOC 코드) 경기수 합계 28 5116 5082 5490 15688 미국 (USA) [P] [Q] [R] [Z] [F] 27 1022 795 706 2523 소련 (URS) [URS] 319 296 1010 영국 (GBR) [GBR] [Z] 293 851 295 4 중화인민공화국 (CHN) [CHN] 10 224 167 155 546

☑근 pandas의 DataFrame 행 검색



→ '국가 (IOC 코드)' 열에서 '대한민국 (KOR)' 행 검색 출력

```
df['국가 (IOC 코드)'] == '대한민국 (KOR)'
✓ 0.0s
      False
     False
     False
     False
     False
     False
148
     False
     False
     False
     False
Name: 국가 (IOC 코드), Length: 153, dtype: bool
  df[df['국가 (IOC 코드)'] == '대한민국 (KOR)']
✓ 0.0s
    국가 (IOC 코드) 경기수 금 은 동 합계
15 대한민국 (KOR)
                   17 90 87 90 267
  df.loc[df['국가 (IOC 코드)'] == '대한민국 (KOR)']
 ✓ 0.0s
    국가 (IOC 코드) 경기수 금 은 동
   대한민국 (KOR)
                   17 90 87 90 267
```

☑근 pandas의 DataFrame 행 검색



→ '국가 (IOC 코드)' 열에서 '대한민국 (KOR)' 행 검색 출력

```
df['국가 (IOC 코드)'].isin(['대한민국 (KOR)'])
✓ 0.0s
      False
      False
      False
      False
      False
     False
     False
149
     False
     False
     False
Name: 국가 (IOC 코드), Length: 153, dtype: bool
  df[df['국가 (IOC 코드)'].isin(['대한민국 (KOR)'])]
✓ 0.0s
    국가 (IOC 코드) 경기수 금 은 동
15 대한민국 (KOR)
                   17 90 87 90 267
  df.loc[df['국가 (IOC 코드)'].isin(['대한민국 (KOR)'])]
✓ 0.0s
    국가 (IOC 코드) 경기수 금 은 동
    대한민국 (KOR)
                   17 90 87 90 267
```

Series.isin([item1, item2, ···]) 시리즈의 각 요소가 전달된 값 시퀀스의 요소와 <mark>정확히 일치</mark>하는지 여부를 보여주는 부울 시리즈를 반환



→ '국가 (IOC 코드)' 열에서 '대한민국' 등 부분 문자열이 있는 행 검색 출력

```
df[df['국가 (IOC 코드)'].str.contains('대한민국')]
✓ 0.0s
   국가 (IOC 코드) 경기수 금 은 동 합계
   대한민국 (KOR)
                  17 90
                        87 90
                               267
 df[df['국가 (IOC 코드)'].str.contains('KOR')]
✓ 0.0s
   국가 (IOC 코드) 경기수 금 은 동
   대한민국 (KOR)
                  17 90 87 90
                               267
 df[df['국가 (IOC 코드)'].str.contains('러시아')]
✓ 0.0s
              국가 (IOC 코드) 경기수 금 은 동
                                            합계
10
            러시아 (RUS) [RUS]
                             6 148 125
                                        153
                                             426
        러시아 제국 (RU1) [RU1]
86
    러시아 출신 올림픽 선수 (OAR)
```

SUMMARY

학습정긴





•••

- 위키피디다의 '올림픽 메달 집계' 전체 페이지에서 데이터 읽어와 특정 데이터프레임을 준비
- 하계 올림픽 메달 집계를 위한 데이터프레임을 생성

```
import pandas as pd

df = pd.read_html('https://ko.wikipedia.org/wiki/%EC%98%AC%EB%A6%BC%ED%94%BD_%EB%A9%94%EB%8B%AC_%EC%A7%91%EA%B3%84')

df2 = df[0].set_index('국가 (IOC 코드)')

summer = df2.iloc[:, :5]
summer.columns = ['경기수', '금', '은', '참계']
summer = summer.sort_values('금', ascending=False)
summer
```

🧓 '국가 (IOC 코드)' 열에서 '대한민국' 등 부분 문자열이 있는 행 검색

```
df[df['국가·(IOC·코드)'].str.contains('대한민국')]

✓ 0.0s

국가(IOC 코드) 경기수 금 은 동 합계

15 대한민국(KOR) 17 90 87 90 267
```

