WEB SCRAPING WITH PANDAS

Level: Easy Time: Fast

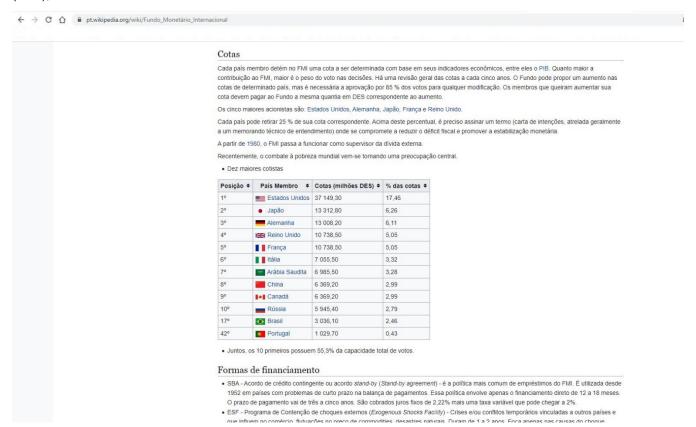
Tools: Python, Pandas

Know how to choose a website

Web scraping with Pandas is very basic. It can be done using the pd.read_html() function, which actually uses beautiful soup, one of the most known libraries for web scraping.

The function will look for tables within the content to bring you the result, but as it is ready for use, there is only so much it will be able to do, thus the method won't do well with complex tables or custom content to be retrieved.

You should aim for simple tables embedded in websites, like this one where you have a content (text), then a table and more content.



The code

The code is really simple.

```
# Web scraping: just use pd.read_html('web address')
page = pd.read_html('https://pt.wikipedia.org/wiki/Fundo_Monet%C3%A1rio_Interna
cional')
```

It will return a list for you

```
type(page)
[OUT]: list
```

If you print it on the screen, you will see that we brought all the content from the page.

```
page
   Ľ→
     0 NaN Esta página cita fontes, mas estas não cobrem ...,
                                             Unnamed: 0
                                                                                               Unnamed: 1
                                                    NaN
                                                                                                      NaN
     1 Mapa-múndi destacando os países-membros do FMI... Mapa-múndi destacando os países-membros do FMI...
                                                                                Organização internacional
                                                   Tipo
                                                Fundação
                                                                            22 de junho de 1944 (77 anos)
                                                   Sede
                                                                          Washington, D.C. Estados Unidos
                                                Membros
                                                                                 Estados da ONU e Kosovo
                                        Línguas oficiais
                                                                               Inglês, francês e espanhol
                                               Filiação
                                                                     Organização das Nações Unidas (ONU)
                                          Diretora-geral
                                                                                   Kristalina Gueorguieva
                                           Sítio oficial
                                                                                              www.imf.org,
          Membros eleitos
     0
                Bélgica Áustria, Bielorrússia, Bélgica, Hungria, Cazaq...
          Países Baixos Bósnia e Herzegovina, Bulgária, Croácia, Chipr...
     1
                  México Costa Rica, El Salvador, Guatemala, Honduras, ...
     3
                  Itália Albânia, Grécia, Itália, Malta, Portugal, San ...
                  Canadá Antígua e Barbuda, Bahamas, Barbados, Belize, ...
              Finlândia Dinamarca, Estônia, Finlândia, Islândia, Letôn...
         Coreia do Sul Austrália, Kiribati, Coreia do Sul, Ilhas Mars...
                   Egito Bahrein, Egito, Iraque, Jordânia, Kuwait, Líba...
                 Malásia Brunei, Camboja, Ilhas Fiji, Indonésia, Laos, ...
     8
               Tanzânia Angola, Botswana, Burundi, Eritreia, Etiópia, ...
     10
                   Suíça Azerbaijão, Quirguistão, Polônia, Sérvia e Mon...
                    Irão Afeganistão, Argélia, Gana, Irã, Marrocos, Paq...
     12
                  Brasil Brasil, Colômbia, República Dominicana, Equado...
                  Índia
                                       Bangladesh, Butão, Índia, Sri Lanka
     13
               Argentina Argentina, Bolívia, Chile, Paraguai, Peru, Uru...
     14
     15 Guiné Equatorial Benim, Burquina Fasso, Camarões, Cabo Verde, C...,
        Posição
                 País Membro Cotas (milhões DES) % das cotas
            1º Estados Unidos 37 149,30
2º Japão 13 312,80
     0
            2º Japão
                      Alemanha
                                        13 008,20
                                                           611
                 Poino Unido
```

Now we need to find the table we want: ten biggest shareholders from IMF.

To do that, we will look at each item. And it happens to be on the index 3.



Shareholders: index 3

page[3]

	Posição	País Membro	Cotas (milhões DES)	% das cotas
0	1°	Estados Unidos	37 149,30	1746
1	2°	Japão	13 312,80	626
2	3°	Alemanha	13 008,20	611
3	4°	Reino Unido	10 738,50	505
4	5°	França	10 738,50	505
5	6°	Itália	7 055,50	332
6	7°	Arábia Saudita	6 985,50	328
7	8°	China	6 369,20	299
8	9°	Canadá	6 369,20	299
9	10°	Rússia	5 945,40	279
10	17°	Brasil	3 036,10	246
11	42°	Portugal	1 029,70	43

Now, just save it in a variable and use for your project df = page[3]

type(df)

pandas.core.frame.DataFrame