

Métodos Computacionales para las Políticas Públicas

Proyecto Final - Análisis Envolvente de Datos

Santiago Ortiz Ortiz

Importamos las librerías necesarias para desarrollar Web Scraping a la página.

```
In [25]: import pandas as pd
import re
import requests
import time
from urllib.request import urlopen
from bs4 import BeautifulSoup
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import Select
from selenium.webdriver.support.select import Select
from selenium.webdriver.common.action_chains import ActionChains
```

Para ingresar al panel principal.

La página web a la que le vamos a hacer WS es Sistema de Gestión Hospitalaria (SIHO). Url principal: <https://prestadores.minsalud.gov.co/siho/> (<https://prestadores.minsalud.gov.co/siho/>). Está es la URL principal. Sin embargo, cuando uno quiere examinar la página se da cuenta que tiene unas macro que no se muestran de forma tan simple. Por lo tanto, ingresamos a un link secundario que habilita la macro: <https://prestadores.minsalud.gov.co/siho/work.aspx> (<https://prestadores.minsalud.gov.co/siho/work.aspx>)

Una vez acá, el software nos habilita un usuario de invitado. Este ya está digitado, solo hay que darle click en ingresar.

```
In [62]: # Creamos una variable con la URL de la página.
url = "https://prestadores.minsalud.gov.co/siho/work.aspx"

# Con ayuda del WebDriver de Chrome vamos a abrir la página.
browser = webdriver.Chrome(executable_path=r"C:\Users\tatoo\Desktop\PhantomJS\Ch
# A la hora de hacer Web Scraping es importante dejar un periodo de tiempo relat
time.sleep(5)
# Con el .get(url), el programa sabe que tiene que ingresar a esa URL
browser.get(url)
time.sleep(5)
# Le pedimos al programa que encuentre un elemento en el código HTML que se llame
boton_submit = browser.find_element_by_name("btnIngresar").click()
time.sleep(4)
```

Para que Python sepa que esta en otra pestaña

<https://stackoverflow.com/questions/17676036/python-webdriver-to-handle-pop-up-browser-windows-which-is-not-an-alert> (<https://stackoverflow.com/questions/17676036/python-webdriver-to-handle-pop-up-browser-windows-which-is-not-an-alert>)

```
In [63]: # Fijamos una página principal llamada MAIN
main_window_handle = None
while not main_window_handle:
    main_window_handle = browser.current_window_handle
    browser.find_element_by_link_text('Capacidad Instalada').click()
    signin_window_handle = None
    while not signin_window_handle:
        for handle in browser.window_handles:
            if handle != main_window_handle:
                signin_window_handle = handle
                break
    browser.switch_to.window(signin_window_handle)
```

```
In [64]: browser.page_source # Para comprobar que tenga la nueva forma
```

```

Out[64]: <'html' xmlns=//www.w3.org/1999/xhtml><head>\n\tCapacidad Instalada
a\n</title><link href=" ../includes/appestilos.css" rel="stylesheet"></head>\n<b
ody bgcolor="#ffffff">\n    <form name="aspnetForm" method="post" action=" ../cap
acidadinstaladanacional.aspx?pageTitle=Capacidad+Instalada&pageHlp=%2fSIH0%
2fayudas%2finformes%2fcapacidad.pdf" id="aspnetForm">\n<input type="hidden" nam
e="__VIEWSTATE" id="__VIEWSTATE" value="/wEPDwULLTIwNDI2NTY5MjEPZBYCZg9kFgICAw9
kFggCDw8PFgIeC05hdm1nYXRlVXJsBSMvU0lITY9heXVxYXNwYW5mb3JtZXNvY2FwYWNpZGFkLnBkZm
RkAhEPDxYChGRUZxh0BRNDYXBhY2lkYWQgSW5zdGFsYWRhZGQCEw9kFgQCAQ9kFgJmD2QWAmYPZBYEA
gEPEgQQFRIEMjAwMgQyMDA2BDIwMDQEMjAwNQyMDA2BDIwMDcEMjAwOAQyMDA5BDIwMTAEMjAwMQQy
MDEyBDIwMTMEMjAwNAQyMDE1BDIwMTYEMjAwNWQyMDE4BDIwMTkVEgQyMDAYBDIwMDMEMjAwNAQyMDA
1BDIwMDYEMjAwNWQyMDA4BDIwMDkEMjAwMAQyMDE4BDIwMTIEMjAwMwQyMDE0BDIwMTUEMjAwNgQyMD
E3BDIwMTgEMjAwORQrAxJnZ2dnZ2dnZ2dnZ2dnZ2dnZ2dkZAIDDxAPFgYeDURhdGFUZXh0RmllbGQFB
m5vbWJyZr40RGF0YVZhbHVlRmllbGQFBmNvZGl1bnx4LXyFEYXRhQm91bmRnZBAVJgAiw1hem9uYXNj
QW50aW9xdWl1bHkFyYXVjYTpBcmNoaXBw6lsYWdvIGRlIFNhb1BBbmRyw6lzlCBQcm92aWR1bmNpYSB
5IFNhb1RhIEhndGFsaW5hckF0bM0hbnRyY28fQmFycmFucXVpbgxhIERpc3RyaXRybyBfc3B1Y2lhbA
1Cb2dvdM0hLCBELkMuCEJvbM0tdmFyB0JveFJw6EeQnVlbmF2ZW50dXJhIERpc3RyaXRvIEVzcGVja
WFsBkNhbgGRhcwDYXF1ZXTDoRtDYXJ0YWdlbmEgRGlzdHJpdG8gRXNwZWNPYwWlQ2FzYW5hcmUFQ2F1
Y2EFQ2VzYXIGQ2hvY80zCEPDs3Jkb2JhDEN1bmRpbmFtYXJjYQhHdWfpbs0tYQhHdWf2aWfYzQVIdWl
sYQpMYSBHdWfqaXJhCU1hZ2RhbGVuYQRNZXRhB05hcmNDSw8STm9ydGUgZGUgU2FudGFuZGVyCFB1dH
VtYXlvCFf1aW5kw61vCVJpc2FyYwXkYR1TYW50YSBNYXJ0YSBEBaXN0cm10byBFc3B1Y2lhbA1TYW50Y
W5kZXIFU3VjcmUGVG9saW1hD1ZhbGx1IGRl1bCBdYXVjYQdWYXVw6lzlB1ZpY2hhZGEVJgACOTECMDUC
ODECODgCMDgCMDkCMTECMTCMTUCNzcCMTcCMTgCMTQCOCODUCTKCMjACMjcCMjMCMjUCOTQCOTUCNDE
CNDQCNDcCNTACNTICNTQCOCODYCNjMCNjYCNdGcNjgCNzACNzMCNzYCOTcCOTkUKwMmZ2dnZ2dnZ2dnZ2
dnZ2dnZ2dnZ2dnZ2dnZ2dnZ2dnZ2dnZ2dnZ2dkZAIDD2QWBMYPZBYCZg9kFgICAQ8PFgIfAQUITmFja
W9uYwXkZAIBD2QWAmYPZBYCAGEPpCsACwEADxYIHgheYXRhS2V5cxYAHgtfIUl0ZW1Db3VudGYeCVBh
Z2VDb3VudAIBHhVfIURhdGFtB3VyY2VJdGVtQ291bnRmZGQCA9kFgJmD2QWAgIBDzwrAASBAa8WCB8
FFgAfBmYfBwIBHwhmZGQCiw8PFgIfAAUjL1NjSE8vYXl1ZGFzL2luZm9ybWVzL2NhcGFjaWRhZC5wZG
ZkZBgBBR5fX0Nvb1RyB2xzUmVxdWlyZVBvc3RCYWNrS2V5X18WBAUQX2N0bDA6aWJ0dWV2b0hkcgURX
2N0bDA6aWJ0dXNjYXJIZHIFEF9jdGwwOml1TnVldm9GdHIFEFV9jdGwwOml1QnVzY2FyRnRyQlJ5dStL
kVqx+BxVoqWdISJBhVfVzeMx73MWG3zgTXg=">\n\n<input type="hidden" name="__VIEWSTAT
EGENERATOR" id="__VIEWSTATEGENERATOR" value="CDDFEAAA">\n<input type="hidden" n
ame="__EVENTVALIDATION" id="__EVENTVALIDATION" value="/wEdAD2AYkbKfXVxMjLZ6wCjB
v079f091K5A3efx5e+2zoL0c8MnqKGR0rClTz0V6RZiWxig9AnBtjjkg/VEAneC+d99h9tMfU8e9v1W
Yo15vWdJI6JRYX1r1HLM+1P+bJpa0Q6UtlKt+I6s0dFKPk1bR4R9/JIRgPbk0YKa3eoc1l4hPutj5xR
1YFL/Uf8mgGvAvIZpdd4Y9hshvYycAsBkL5AK5qJlx/+p1ToCkMvNhef0vf09LlGgOE17GJZyGi1Xb
q7XLALwelQXA89io1tXN2xhYq3aYptZkHdp2HTQZR20rxPibIic1qF/zAvkz9LJob1CSS9ilBgrPzOw
a1KkkIDH0gUP3Ou716ueOS5Mx07R2aiCu1MlbsxPslWNU04Q00XbmQLAuYoDFu+QjGv9Tw7Bs/H9Mzv
Yxe4fd6/b3uqlbFjP/RfzZ9mH8XqnyTiIGq0zwSpDegbgEWThrj900Tr6u93Bo00Jr2aHejoJ6d4Wu/
OxG9jepid/hwWS86FL2GurwaY3N3EdD7Hz01Y4ZL4XzXFNzxPfsYY7jSjxzGGkU3bxOSbX/eTuvnN65
FKn9HLp6Tt2BaDM85pGPnzn3vWmSglbwXg1YXZev2Jj6nF8xXNjbuMoPoMGZ+1ZAtDHC1NyW5q8+xkg
Obj2PT+A8/kG7J4zUbTeSDrveEUKejh8YuZnV0N1cuwh1kNAK5kYtOY8xxPXqUF+0k4n7+ojQcpzGqJ
xLyMt6HJwJ+yv3chGaSOOTJzaDJ0BhQg3ZgJLrP9b28sRPTcicSwS2pI1x2DR3zgTXyrfd+3Kg9dp+F
DhLp4u6c2yyQwM9wxCOG0Urz4yzYEvxRQ88w70507NHtre8ibj2hq0tbFvtLTZYIEb8V6k0bzbodaS8
ZAQLsJMGYK0/04WEbbjLOYwfMJ8+1/Grh3rSE8y6796+XW6xgkS6R21s+4MIfqh2Ny1lFSNuWK/I0aB
mZYY70goYRWeZhuavhykLZkQpRHPnx23aEcjwdIH8McVR7IkkVGzWwscz0KX0aoy6Yntlw4BMz/6Z
ECXQ8q6VHCoeAsGM7si4C8zHXnyoKE3dIQ6sjXAtUXWHEapxT8b/bSh3UVUdFkPtCBUGIFYSXRGMD02
kJxMCYjm62ZJ2SpAsXwQlSdFux//YaDilxkHBMTr3NSPFVClw7BX5ZepAbZMn2a1ANAUM7xaUamB/x
bo0yDJ4Le1FsIlyyki/LdPAD09gTTjgrQDJLLdkwMEVh/tVnmh0BccF18HpD/9nnThZoc1US6pj5Krf
ZvkjYC8r20ztDaf21L374PbDqY0Zb1WNjz4jIb62YPZMpDAz8mi6Io/2sugnYY=">\n    <div>\n
<table border="0" cellpadding="0" cellspacing="0">\n        <tbody><tr>\n
<td>\n                </td>\n                <td>\n                    </td>\n                <td>\n                    <input type="im
age" name="__ctl0_ibNuevoHdr" id="__ctl0_ibNuevoHdr" src=" ../images/nuevo.jpg" al
t="Nuevo" align="top" style="border-width:0px;"></td>\n            <td>\n

```

ks/Desktop/MCPP/MCPP_santiago.ortiz/Proyecto Final/Proyecto final.ipynb 4/

Capacidad instalada	
Concepto	Cantidad
Nacional	
Capacidad instalada por nivel	
Nivel	Concepto
Cantidad	Porcentaje
Promedio	
Grabar	
Nuevo	
Buscar	
Eliminar	
Ayuda	

Lo que hacen los dos códigos siguientes es:

- Encontrar las casillas año y departamento. Como ambas son elementos de multiple opción este desplegará unos valores.
- Año: 2002 - 2019 (Sin embargo, no hay datos para el 2019)
- Depto: Amazonas - Vaupés (+Bogotá DC.)

```
In [65]: # El espacio del año tiene un nombre como el siguiente:
cod_año = browser.find_element_by_name("_ctl0:ContentPlaceHolder1:ddano")
for option in cod_año.find_elements_by_tag_name('option'):
    if option.text == '2003':
        option.click()
        break
```

```
In [91]: # El espacio del departamento tiene un nombre como el siguiente:
cod_depto = browser.find_element_by_name("_ctl0:ContentPlaceHolder1:dddepa_codigo")
for option in cod_depto.find_elements_by_tag_name('option'):
    if option.text == 'Caldas':
        option.click() # select() in earlier versions of webdriver
    break
```

```
In [92]: # Después de seleccionar Los departamentos Y LOS AÑOS LE DAMOS CLICK a Buscar
        boton_buscar = browser.find_element_by_name("_ctl0:ibBuscarFtr")
        boton_buscar.click()
```

```
In [93]: html = browser.execute_script("return document.getElementsByTagName('html')[0].innerHTML")
soup = BeautifulSoup(html, "html.parser")
```

```
In [94]: re.findall('<td align="center">(\d*)</td><td>(.*?)</td><td align="right">(\d*)</td>')
Out[94]:
```

```
Out[94]: [('1', 'Camas de hospitalización', '678'),
          ('2', 'Camas de hospitalización', '167'),
          ('3', 'Camas de hospitalización', '287'),
          ('1', 'Camas de observación', '74'),
          ('2', 'Camas de observación', '39'),
          ('3', 'Camas de observación', '10'),
          ('1', 'Consultorios de consulta externa', '165'),
          ('2', 'Consultorios de consulta externa', '44'),
          ('3', 'Consultorios de consulta externa', '15'),
          ('1', 'Consultorios en el servicio de urgencias', '33'),
          ('2', 'Consultorios en el servicio de urgencias', '14'),
          ('3', 'Consultorios en el servicio de urgencias', '4'),
          ('1', 'Salas de quirófanos', '12'),
          ('2', 'Salas de quirófanos', '6'),
          ('3', 'Salas de quirófanos', '5'),
          ('1', 'Mesas de partos', '35'),
          ('2', 'Mesas de partos', '7'),
          ('3', 'Mesas de partos', '0'),
          ('1', 'Número de unidades de odontología', '79'),
          ('2', 'Número de unidades de odontología', '13'),
          ('3', 'Número de unidades de odontología', '0')]
```

```
In [95]: agregar = re.findall('<td align="center">(\d*)</td><td>(.*?)</td><td align="right">(\d*)</td></tr>')
```

Ideas para la base de datos:

- <https://thispointer.com/python-pandas-how-to-add-rows-in-a-dataframe-using-dataframe-append-loc-iloc/> (<https://thispointer.com/python-pandas-how-to-add-rows-in-a-dataframe-using-dataframe-append-loc-iloc/>)

```
In [84]: df_base = pd.DataFrame(columns=['Nivel', 'Concepto', 'Cantidad', 'Departamento',
df_base
```

Out[84]:

Nivel	Concepto	Cantidad	Departamento	Año	Clase
-------	----------	----------	--------------	-----	-------

```
In [96]: agregar_base = pd.DataFrame(agregar, columns = ['Nivel', 'Concepto', 'Cantidad'])
agregar_base
```

Out[96]:

	Nivel	Concepto	Cantidad
0	1	Camas de hospitalización	678
1	2	Camas de hospitalización	167
2	3	Camas de hospitalización	287
3	1	Camas de observación	74
4	2	Camas de observación	39
5	3	Camas de observación	10
6	1	Consultorios de consulta externa	165
7	2	Consultorios de consulta externa	44
8	3	Consultorios de consulta externa	15
9	1	Consultorios en el servicio de urgencias	33
10	2	Consultorios en el servicio de urgencias	14
11	3	Consultorios en el servicio de urgencias	4
12	1	Salas de quirófanos	12
13	2	Salas de quirófanos	6
14	3	Salas de quirófanos	5
15	1	Mesas de partos	35
16	2	Mesas de partos	7
17	3	Mesas de partos	0
18	1	Número de unidades de odontología	79
19	2	Número de unidades de odontología	13
20	3	Número de unidades de odontología	0

```
In [97]: len(agregar_base["Nivel"]) # Esta siempre aparece, para ver cuantos datos son
```

Out[97]: 21

[illegible]


```
In [99]: agregar_base["Departamento"] = depto
         agregar_base
```

Out[99]:

	Nivel	Concepto	Cantidad	Departamento
0	1	Camas de hospitalización	678	Caldas
1	2	Camas de hospitalización	167	Caldas
2	3	Camas de hospitalización	287	Caldas
3	1	Camas de observación	74	Caldas
4	2	Camas de observación	39	Caldas
5	3	Camas de observación	10	Caldas
6	1	Consultorios de consulta externa	165	Caldas
7	2	Consultorios de consulta externa	44	Caldas
8	3	Consultorios de consulta externa	15	Caldas
9	1	Consultorios en el servicio de urgencias	33	Caldas
10	2	Consultorios en el servicio de urgencias	14	Caldas
11	3	Consultorios en el servicio de urgencias	4	Caldas
12	1	Salas de quirófanos	12	Caldas
13	2	Salas de quirófanos	6	Caldas
14	3	Salas de quirófanos	5	Caldas
15	1	Mesas de partos	35	Caldas
16	2	Mesas de partos	7	Caldas
17	3	Mesas de partos	0	Caldas
18	1	Número de unidades de odontología	79	Caldas
19	2	Número de unidades de odontología	13	Caldas
20	3	Número de unidades de odontología	0	Caldas

```
In [100]: df_base = df_base.append(agregar_base, ignore_index=True, sort=False)
```

In [101]: df_base

Out[101]:

	Nivel	Concepto	Cantidad	Departamento	Año	Clase
0	1	Camas de hospitalización	253	Bolívar	NaN	NaN
1	2	Camas de hospitalización	459	Bolívar	NaN	NaN
2	1	Camas de observación	199	Bolívar	NaN	NaN
3	2	Camas de observación	65	Bolívar	NaN	NaN
4	1	Consultorios de consulta externa	227	Bolívar	NaN	NaN
5	2	Consultorios de consulta externa	45	Bolívar	NaN	NaN
6	1	Consultorios en el servicio de urgencias	59	Bolívar	NaN	NaN
7	2	Consultorios en el servicio de urgencias	11	Bolívar	NaN	NaN
8	1	Salas de quirófanos	16	Bolívar	NaN	NaN
9	2	Salas de quirófanos	12	Bolívar	NaN	NaN
10	1	Mesas de partos	81	Bolívar	NaN	NaN
11	2	Mesas de partos	9	Bolívar	NaN	NaN
12	1	Número de unidades de odontología	116	Bolívar	NaN	NaN
13	2	Número de unidades de odontología	8	Bolívar	NaN	NaN
14	1	Camas de hospitalización	678	Caldas	NaN	NaN
15	2	Camas de hospitalización	167	Caldas	NaN	NaN
16	3	Camas de hospitalización	287	Caldas	NaN	NaN
17	1	Camas de observación	74	Caldas	NaN	NaN
18	2	Camas de observación	39	Caldas	NaN	NaN
19	3	Camas de observación	10	Caldas	NaN	NaN
20	1	Consultorios de consulta externa	165	Caldas	NaN	NaN
21	2	Consultorios de consulta externa	44	Caldas	NaN	NaN
22	3	Consultorios de consulta externa	15	Caldas	NaN	NaN
23	1	Consultorios en el servicio de urgencias	33	Caldas	NaN	NaN
24	2	Consultorios en el servicio de urgencias	14	Caldas	NaN	NaN
25	3	Consultorios en el servicio de urgencias	4	Caldas	NaN	NaN
26	1	Salas de quirófanos	12	Caldas	NaN	NaN
27	2	Salas de quirófanos	6	Caldas	NaN	NaN
28	3	Salas de quirófanos	5	Caldas	NaN	NaN
29	1	Mesas de partos	35	Caldas	NaN	NaN
30	2	Mesas de partos	7	Caldas	NaN	NaN
31	3	Mesas de partos	0	Caldas	NaN	NaN
32	1	Número de unidades de odontología	79	Caldas	NaN	NaN
33	2	Número de unidades de odontología	13	Caldas	NaN	NaN

Nivel		Concepto	Cantidad	Departamento	Año	Clase
34	3	Número de unidades de odontología	0	Caldas	NaN	NaN

```
In [60]: base = base.append(prueba, columns = ['Nivel', 'Concepto', 'Cantidad'])
```

```
-----
TypeError                                Traceback (most recent call last)
<ipython-input-60-87a83dc0f7a7> in <module>
----> 1 base = base.append(prueba, columns = ['Nivel', 'Concepto', 'Cantidad'])

TypeError: append() got an unexpected keyword argument 'columns'
```

Hasta acá funciona melo!, pero hay que pensar en una forma de hacer que Python sepa que va a trabajar en la otra ventana

- Revisar: <https://stackoverflow.com/questions/32858434/switch-to-popup-in-python-using-selenium> (<https://stackoverflow.com/questions/32858434/switch-to-popup-in-python-using-selenium>)

```
In [53]: vacia = []
         if len(vacia) == 0:
             vacia = [(0,0,0,0)]
         vacia
```

```
Out[53]: [(0, 0, 0, 0)]
```

```
In [ ]:
```