

web_scraping2

September 11, 2023

```
[1]: import requests
from bs4 import BeautifulSoup
import pandas as pd
```

```
[43]: uf_url = f'https://www.ibge.gov.br/cidades-e-estados/sp.html'
browsers = {'User-Agent': "Mozilla/5.0 (Windows NT 10.0; Win64; x64)␣
↳AppleWebKit/537.36 \ (KHTML, like Gecko) Chrome / 86.0.4240.198Safari / 537.
↳36"}
page = requests.get(uf_url, headers = browsers)

soup = BeautifulSoup(page.content, 'html.parser')
indicadores = soup.select('.indicador')
```

```
[44]: indicadores
```

```
[44]: [<div class="indicador"><div class="ind-label"><p>Governador</p></div><p
class="ind-value">RODRIGO GARCIA<small> [2019]</small></p></div>,
<div class="indicador"><div class="ind-label"><p>Capital</p></div><p
class="ind-value">São Paulo<small> [2010]</small></p></div>,
<div class="indicador"><div class="ind-label"><p>Gentílico</p></div><p
class="ind-value">paulista</p></div>,
<div class="indicador"><div class="ind-label"><p>Área
Territorial</p></img></div><p class="ind-value">248.219,485 <span
class="indicador-unidade">km²</span><small> [2021]</small></p></div>,
<div class="indicador"><div class="ind-label"><p>População
estimada</p></div><p class="ind-value">46.649.132 <span class="indicador-
unidade">pessoas</span><small> [2021]</small></p></div>,
<div class="indicador"><input class="chart-values" type="hidden" value='["3.37"
,"5.58","9.20","18.50","28.93","36.81","52.28","72.36","102.25","127.11","148.96
","166.25"]' /><input class="chart-labels" type="hidden" value='["1872","1890","1
900","1920","1940","1950","1960","1970","1980","1991","2000","2010"]' /><div
class="ind-label"><p>Densidade
demográfica</p></div><p class="ind-value">166,25 <span class="indicador-
unidade">hab/km²</span><small> [2010]</small></p></div>,</pre>
```

```

<div class="indicador"><input class="chart-values" type="hidden" value='["60756
06","6100512","6025963","5926923","5805590","5673142","5537627","5365948","53436
61","5276326","5367614","5409187","5414208","5396803"]'/><input class="chart-
labels" type="hidden" value='["2008","2009","2010","2011","2012","2013","2014","
2015","2016","2017","2018","2019","2020","2021"]'/><div class="ind-label"><p>Matrículas no
ensino fundamental</p></div><p class="ind-value">5.396.803 <span
class="indicador-unidade">matrículas</span><small> [2021]</small></p></div>,
<div class="indicador"><input class="chart-values" type="hidden"
value='["0.578","0.702","0.783"]'/><input class="chart-labels" type="hidden"
value='["1991","2000","2010"]'/><div class="ind-label"><p>IDH <span>Índice de
desenvolvimento humano</span></p></div><p class="ind-
value">0,783<small> [2010]</small></p></div>,
<div class="indicador"><input class="chart-values" type="hidden" value='["19687
5515.66748","209486514.33074","218966501.13932","218622576.68826","232822496.567
06"]'/><input class="chart-labels" type="hidden"
value='["2013","2014","2015","2016","2017"]'/><div class="ind-label"><p>Receitas
realizadas</p></div><p class="ind-value">232.822.496,57 <span class="indicador-
unidade">R$ (×1000)</span><small> [2017]</small></p></div>,
<div class="indicador"><input class="chart-values" type="hidden" value='["19787
0171.92207","209842014.024","220506971.96347","219266294.16031","231982243.69176
"]'/><input class="chart-labels" type="hidden"
value='["2013","2014","2015","2016","2017"]'/><div class="ind-label"><p>Despesas
empenhadas</p></div><p class="ind-value">231.982.243,69 <span class="indicador-
unidade">R$ (×1000)</span><small> [2017]</small></p></div>,
<div class="indicador"><input class="chart-values" type="hidden" value='["1197"
,"1308","1432","1482","1723","1712","1898","1946","1814","1836"]'/><input
class="chart-labels" type="hidden" value='["2012","2013","2014","2015","2016","2
017","2018","2019","2020","2021"]'/><div class="ind-label"><p>Rendimento mensal
domiciliar per capita</p></div><p class="ind-value">1.836 <span
class="indicador-unidade">R$</span><small> [2021]</small></p></div>,
<div class="indicador"><input class="chart-values" type="hidden" value='["15098
096","16464703","17852829","19139118","20537980","21968359","23286890","24560202
","25718248","26605042","27332101","28138698","29057749","30058975","30778960","
31455473"]'/><input class="chart-labels" type="hidden" value='["2006","2007","20
08","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018","2019"
,"2020","2021"]'/><div class="ind-label"><p>Total de

```

veículos</p></div><p class="ind-value">31.455.473 veículos<small> [2021]</small></p></div>

```
[32]: def scraping_uf(uf: str):
    uf_url = f'https://www.ibge.gov.br/cidades-e-estados/{uf}.html'
    browsers = {'User-Agent': "Mozilla/5.0 (Windows NT 10.0; Win64; x64)␣
↳AppleWebKit/537.36 (KHTML, like Gecko) Chrome / 86.0.4240.198Safari / 537.
↳36"}
    page = requests.get(uf_url, headers = browsers)

    soup = BeautifulSoup(page.content, 'html.parser')
    indicadores = soup.select('.indicador')

    uf_dict = {
        dado.select('.ind-label')[0].text: dado.select('.ind-value')[0].text
        for dado in indicadores
    }

    return uf_dict
```

```
[35]: estado = scraping_uf('sp')
estado
```

```
[35]: {'Governador': 'RODRIGO GARCIA\xa0\xa0\xa0[2019]',
      'Capital': 'São Paulo\xa0\xa0\xa0[2010]',
      'Gentílico': 'paulista',
      'Área Territorial': '248.219,485 km²\xa0\xa0\xa0[2021]',
      'População estimada': '46.649.132 pessoas\xa0\xa0\xa0[2021]',
      'Densidade demográfica': '166,25 hab/km²\xa0\xa0\xa0[2010]',
      'Matrículas no ensino fundamental': '5.396.803 matrículas\xa0\xa0\xa0[2021]',
      'IDH Índice de desenvolvimento humano': '0,783\xa0\xa0\xa0[2010]',
      'Receitas realizadas': '232.822.496,57 R$ (×1000)\xa0\xa0\xa0[2017]',
      'Despesas empenhadas': '231.982.243,69 R$ (×1000)\xa0\xa0\xa0[2017]',
      'Rendimento mensal domiciliar per capita': '1.836 R$\xa0\xa0\xa0[2021]',
      'Total de veículos': '31.455.473 veículos\xa0\xa0\xa0[2021]'}
```

```
[37]: for indicador in estado:
    if ']' in estado[indicador]:
        estado[indicador] = estado[indicador].split(']')[0][: -8]
    else:
        estado[indicador] = estado[indicador]
estado
```

```
[37]: {'Governador': 'RODRIGO GARCIA',
      'Capital': 'São Paulo',
      'Gentílico': 'paulista',
      'Área Territorial': '248.219,485 km²',
```

```
'População estimada': '46.649.132 pessoas',
'Densidade demográfica': '166,25 hab/km²',
'Matrículas no ensino fundamental': '5.396.803 matrículas',
'IDH Índice de desenvolvimento humano': '0,783',
'Receitas realizadas': '232.822.496,57 R$ (×1000)',
'Despesas empenhadas': '231.982.243,69 R$ (×1000)',
'Rendimento mensal domiciliar per capita': '1.836 R$',
'Total de veículos': '31.455.473 veículos'}
```

```
[39]: df = pd.DataFrame(estado.values(), index=estado.keys())
```

```
[49]: df
```

```
[49]:
```

Governador	RODRIGO GARCIA	0
Capital	São Paulo	
Gentílico	paulista	
Área Territorial	248.219,485 km²	
População estimada	46.649.132 pessoas	
Densidade demográfica	166,25 hab/km²	
Matrículas no ensino fundamental	5.396.803 matrículas	
IDH Índice de desenvolvimento humano	0,783	
Receitas realizadas	232.822.496,57 R\$ (×1000)	
Despesas empenhadas	231.982.243,69 R\$ (×1000)	
Rendimento mensal domiciliar per capita	1.836 R\$	
Total de veículos	31.455.473 veículos	

```
[ ]:
```