

web_scraping_ibge

September 11, 2023

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

```
[15]: def scraping_uf(uf: str) -> dict:
    uf_url = f'https://www.ibge.gov.br/cidades-e-estados/{uf}.html'
    page = requests.get(uf_url)

    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
```

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

```
[17]: estado
```

```
[17]: {'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]}'
```

```
[18]: for indicador in estado:
    if ']' in estado[indicador]:
```

```

        estado[indicador] = estado[indicador].split(' ')[0][:-8]
    else:
        estado[indicador] = estado[indicador]
estado

```

```

[18]: {'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'}

```

```

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

```

```

[20]: df

```

```

[20]:
Governador      0
Capital         0
Gentílico       0
Área Territorial 0
População estimada 0
Densidade demográfica 0
Matrículas no ensino fundamental 0
IDH Índice de desenvolvimento humano 0
Receitas realizadas 0
Despesas empenhadas 0
Rendimento mensal domiciliar per capita 0
Total de veículos 0

```

```

[ ]:

```