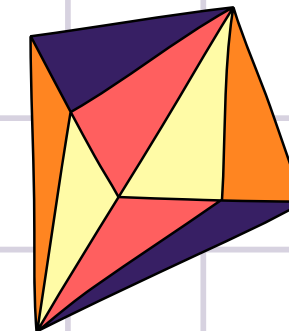
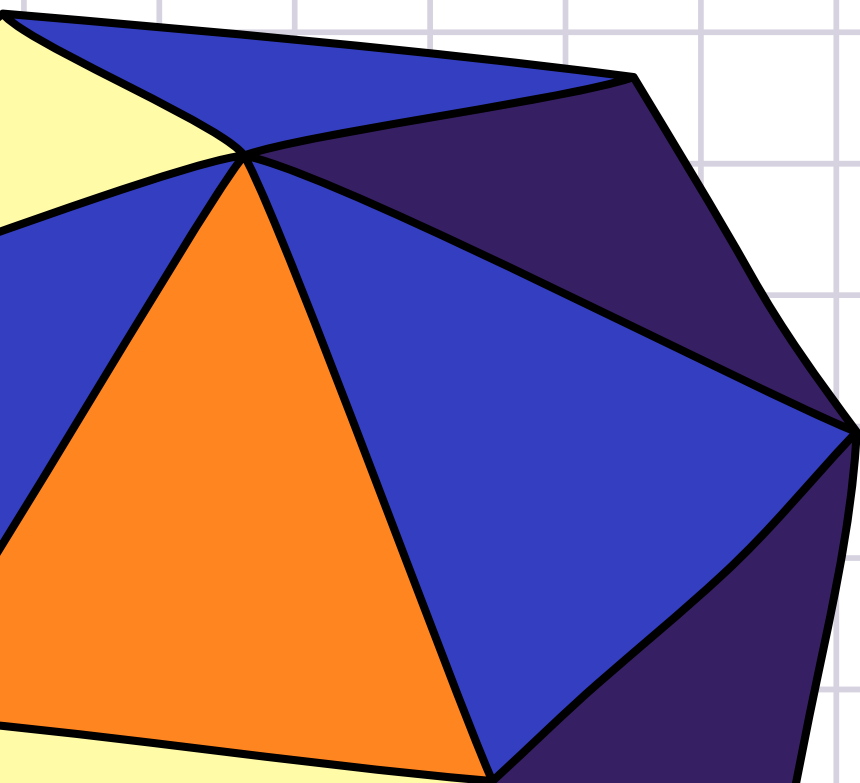
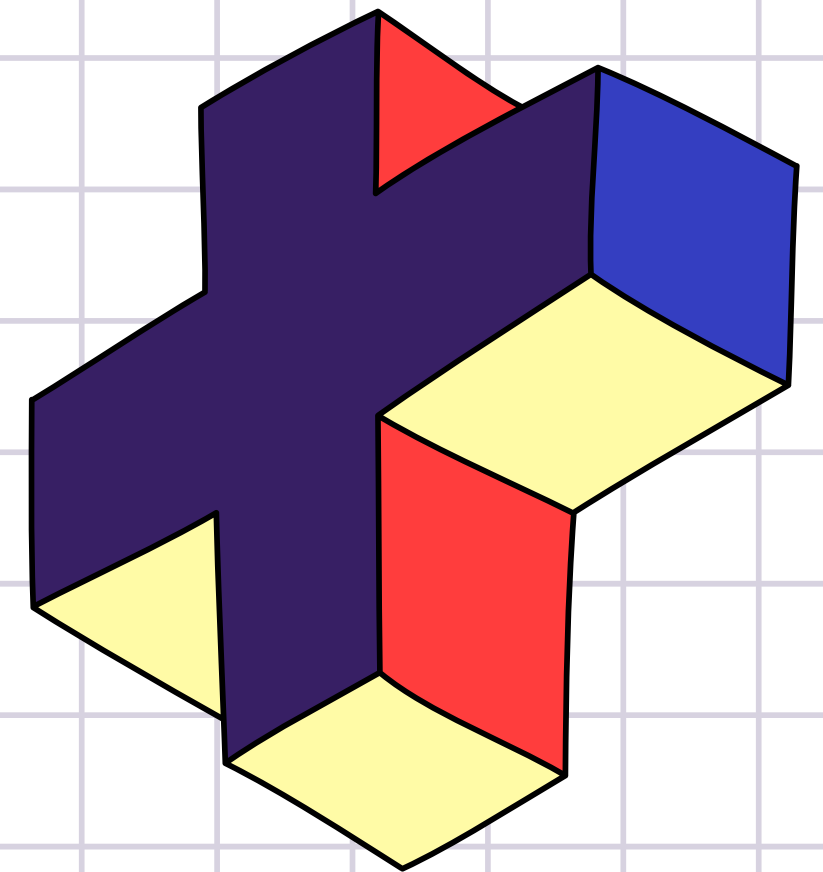


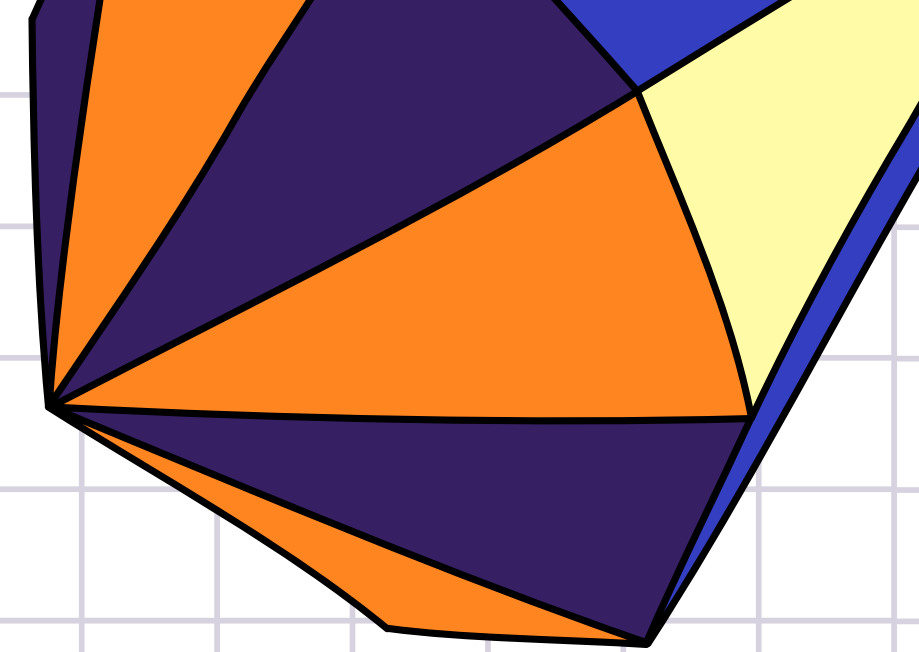
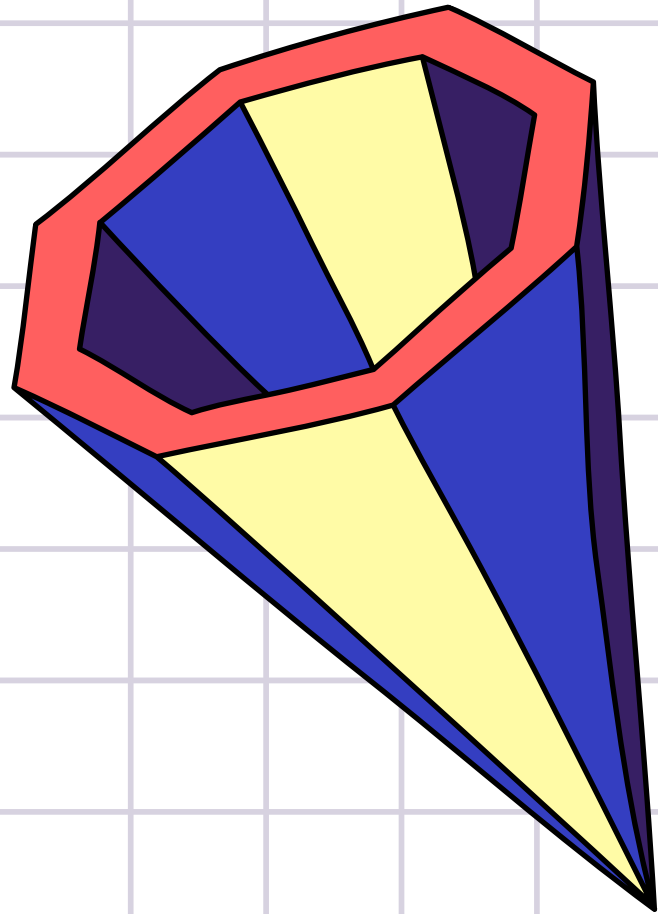
INTRODUÇÃO À CIÊNCIA DE DADOS



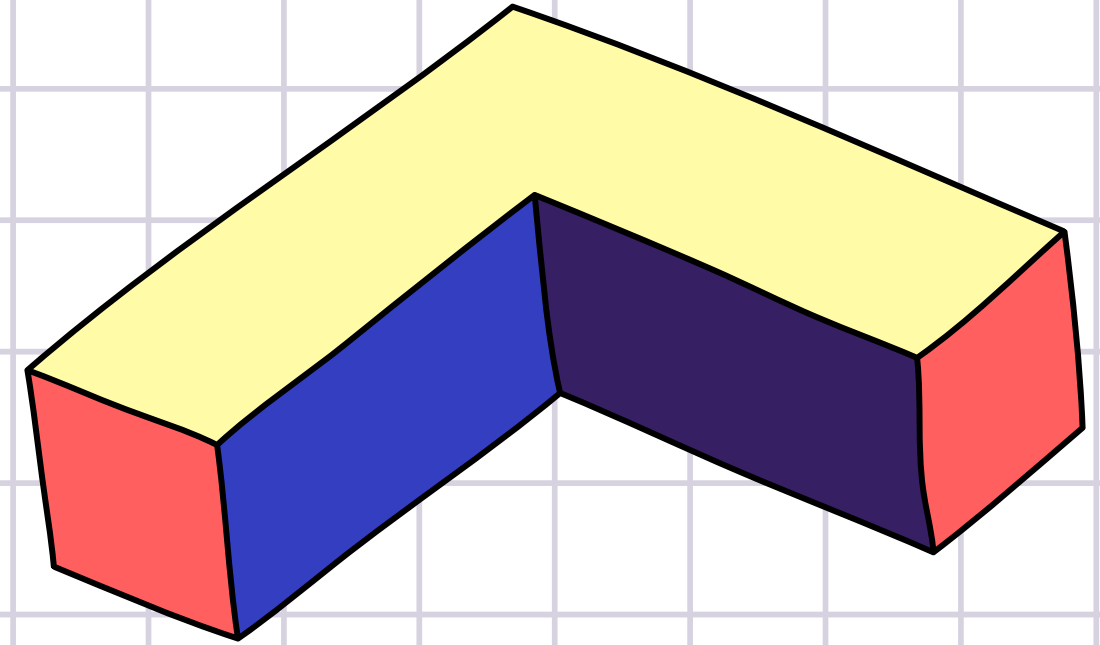
02

INTRODUÇÃO AO WEBSCRAPING E APIS





O QUE É WEBSCRAPING?



TÉCNICA PARA EXTRAIR

grandes quantidades de dados de websites. A maneira como essa informação é extraída varia de acordo com o tipo de website e sua estrutura.

OS SCRIPTS DE WEBSCRAPING

podem navegar automaticamente em várias páginas de um website, coletar e organizar as informações necessárias.

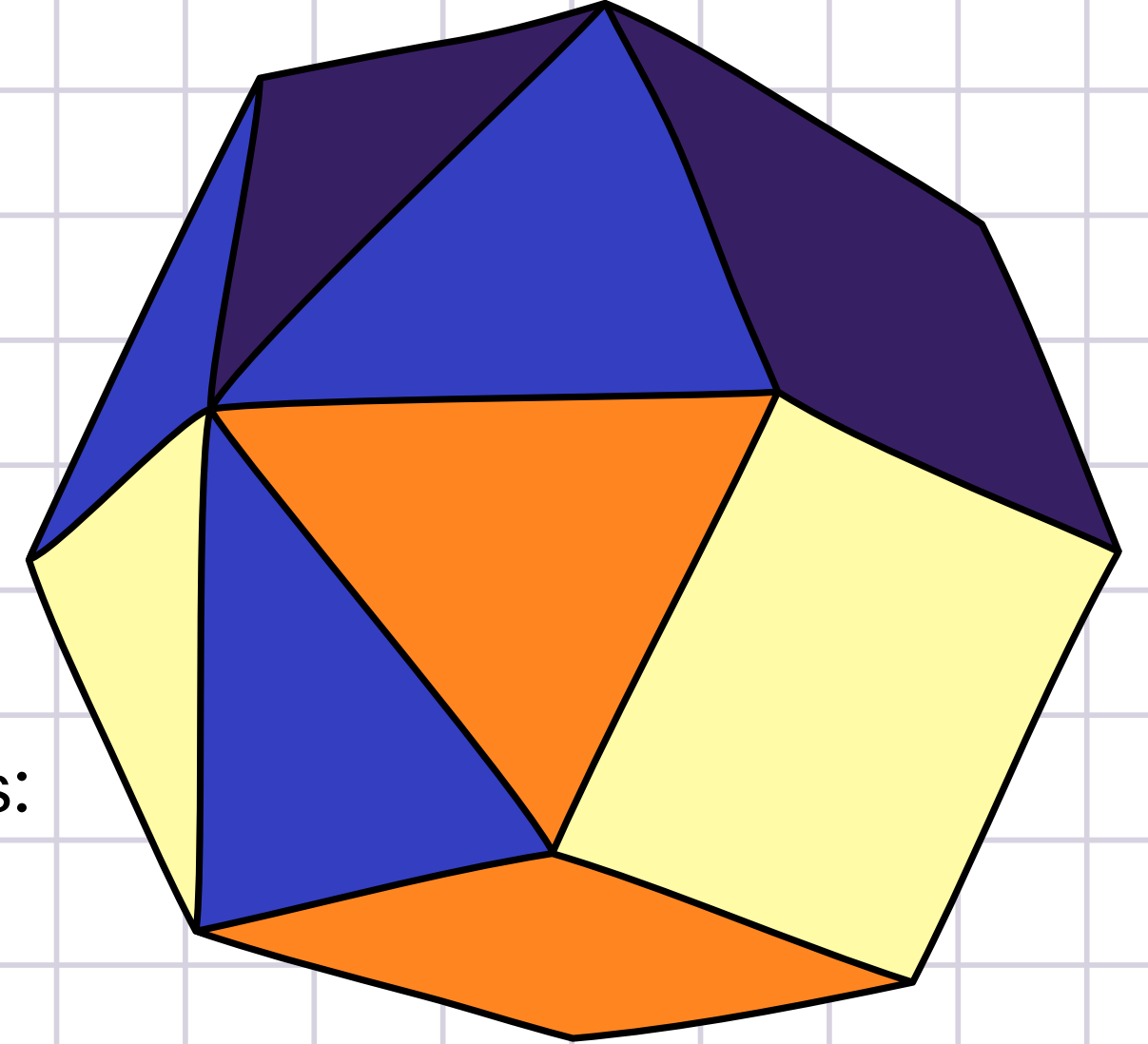
A PRÁTICA DE WEBSCRAPING

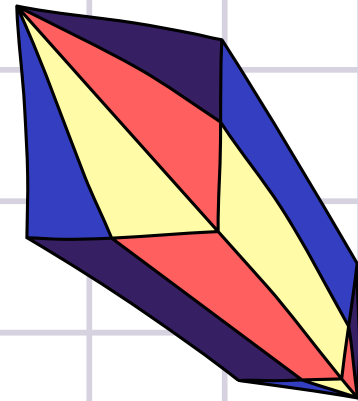
deve ser feita respeitando os termos de serviço do site e as leis de privacidade e proteção de dados.

COMO RASPAR?

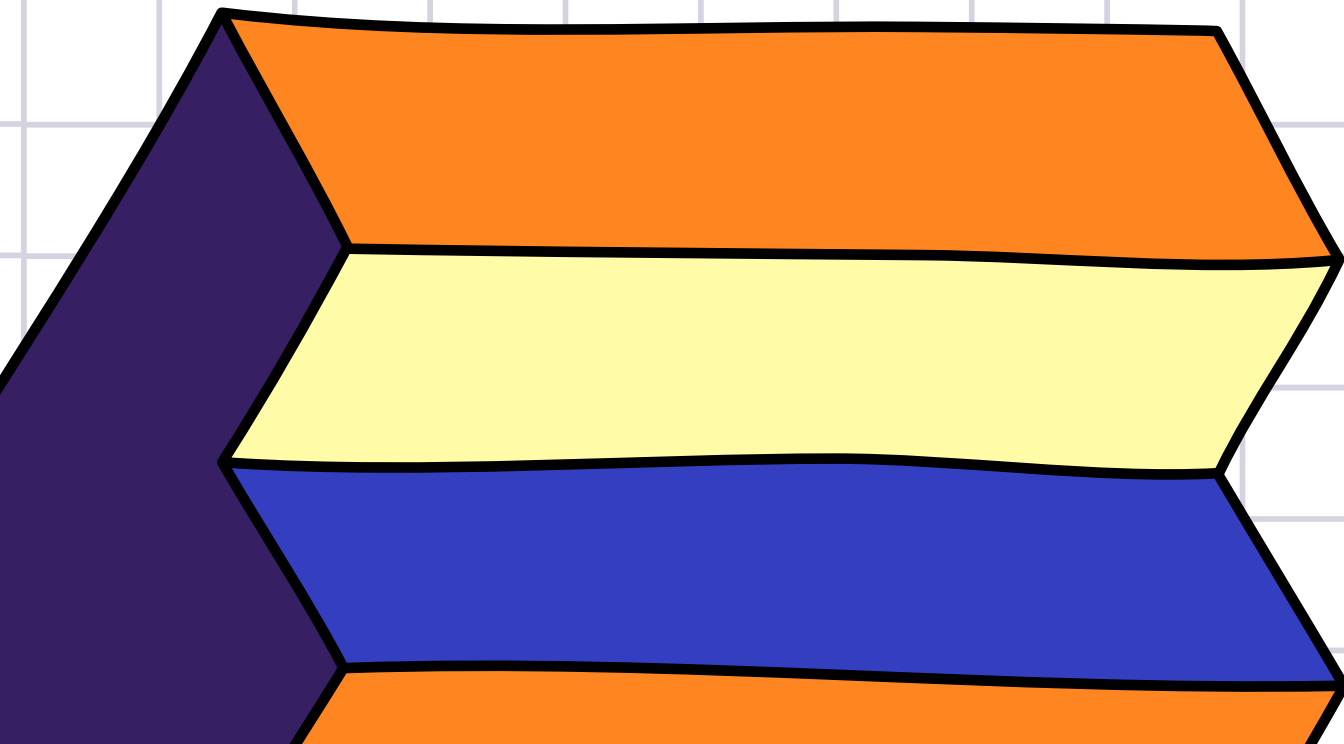
Existem diversas bibliotecas que permitem a raspagem de dados:

- **requests**
- **BeautifulSoup4**
- **Selenium**
- **Playwright**
- **Scrapy**





REQUESTS & BEAUTIFULSOUP4



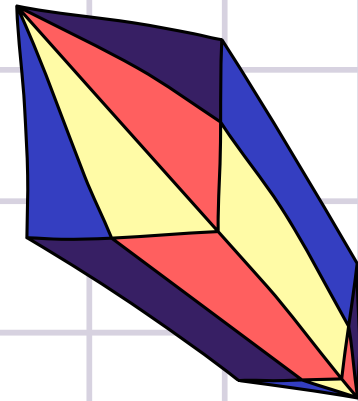
- **REQUESTS:**

- permite fazer solicitações à sites, obtendo o conteúdo

- **BEAUTIFULSOUP4:**

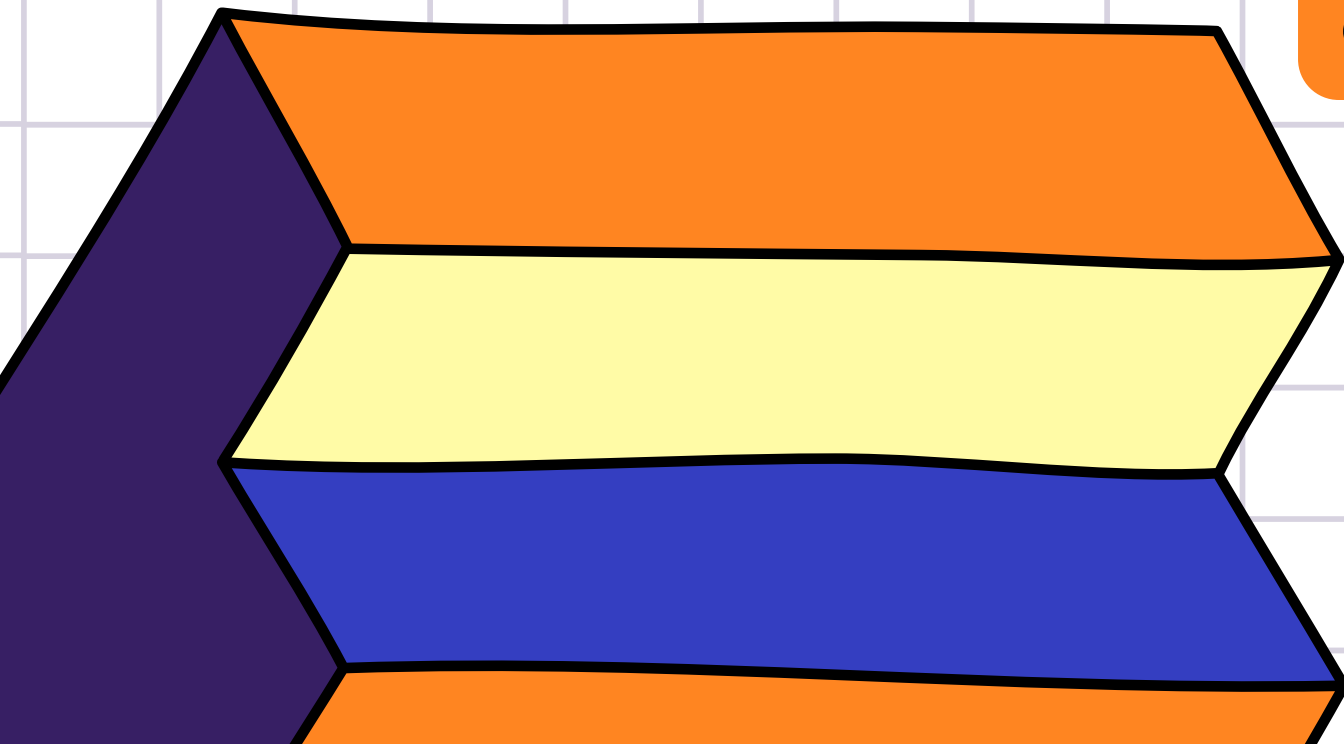
- processa o conteúdo, navegando pelo html e obtendo os objetos desejados

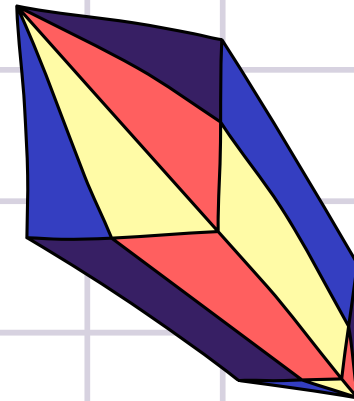
A página da Web vira um objeto no Python.



REQUESTS

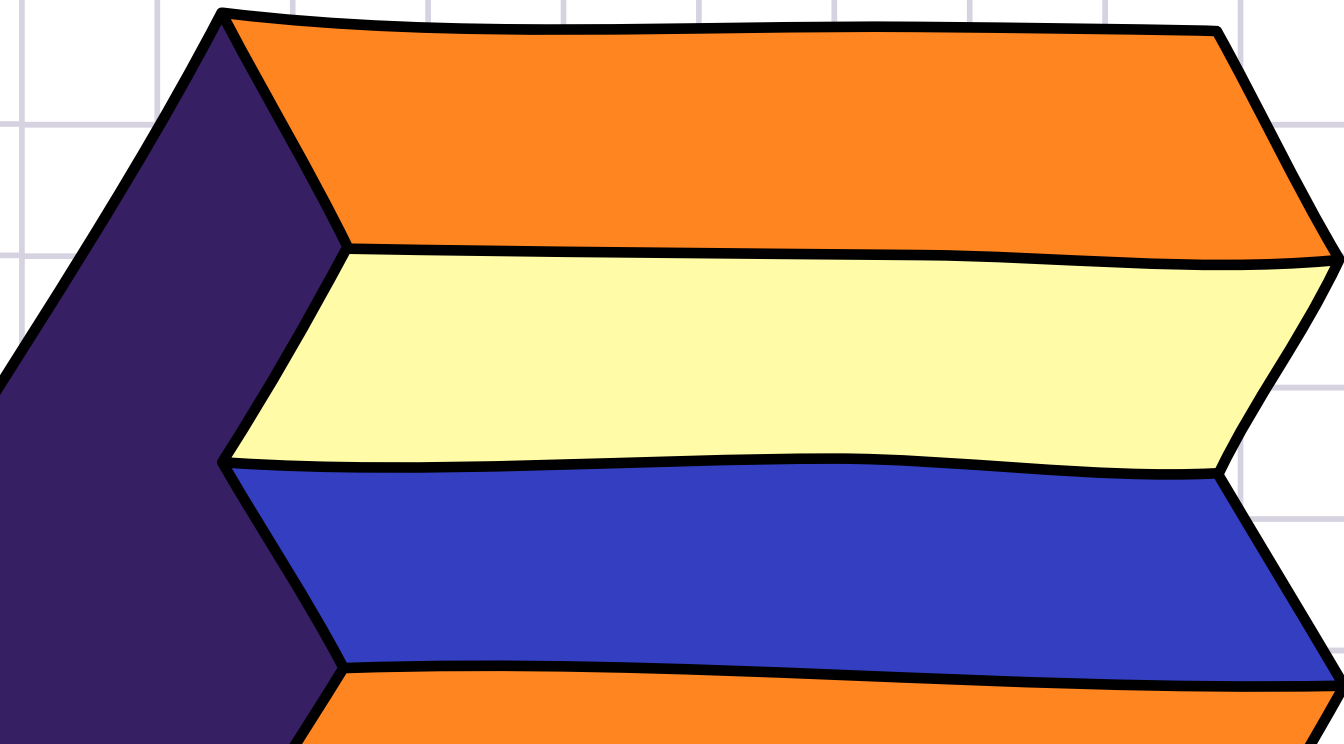
```
import requests  
  
url = 'https://www.brasilparalelo.com.br/search?query=lula'  
  
response = requests.get(url)  
  
content = response.content
```



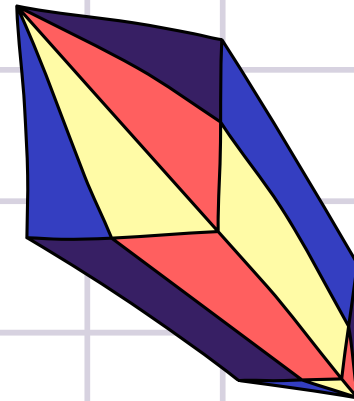


BEAUTIFUL SOUP4

```
from bs4 import BeautifulSoup  
soup = BeautifulSoup(content, 'lxml')  
soup
```



ESTRUTURA HTML

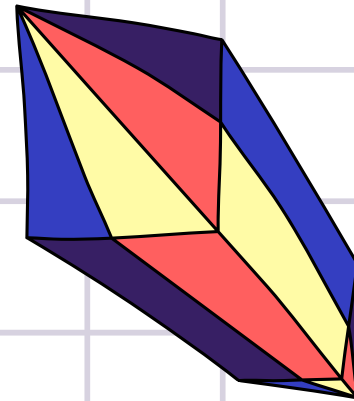


```
<!DOCTYPE html>
<html>

  <head>
    <title>Page Title</title>
  </head>

  <body>
    <h2>Heading Content</h2>
    <p>Paragraph Content</p>
  </body>

</html>
```



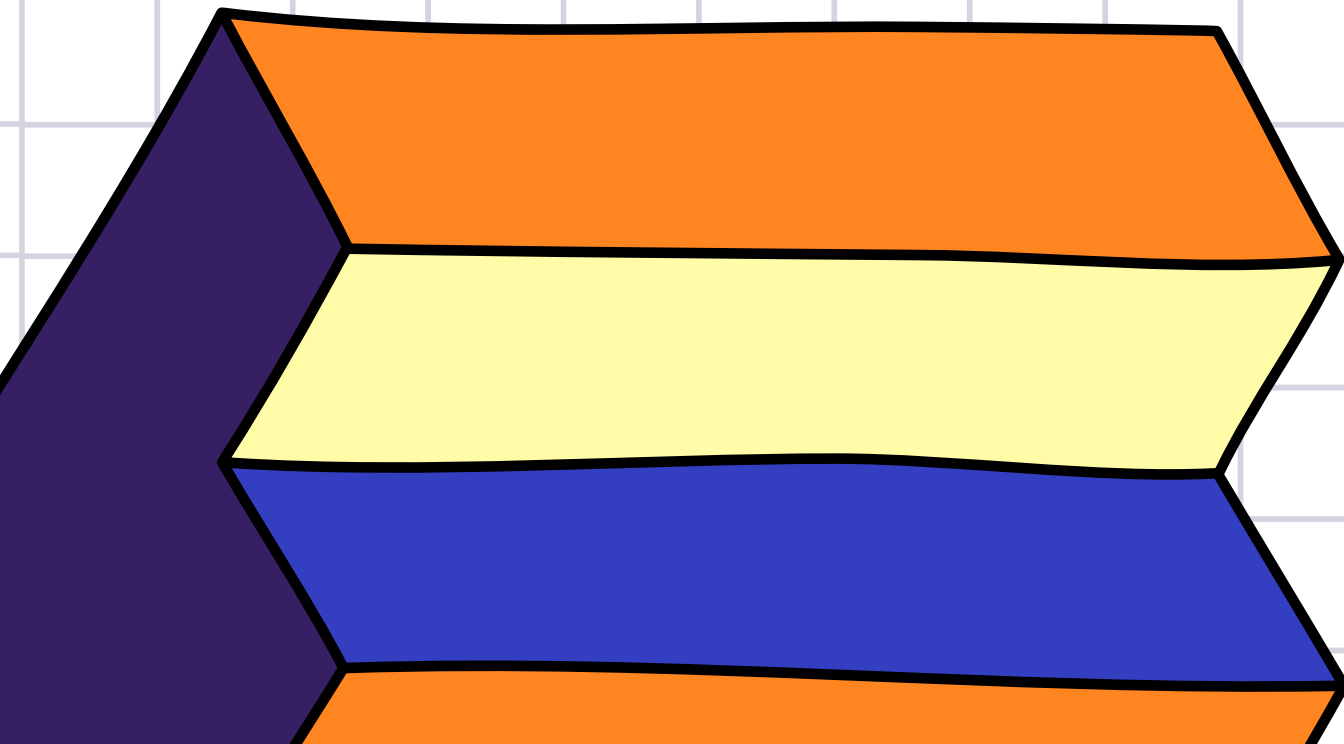
ESTRUTURA HTML

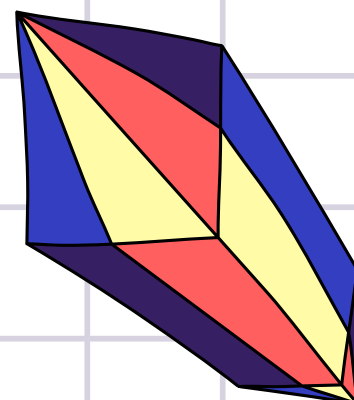
`<div>`

``

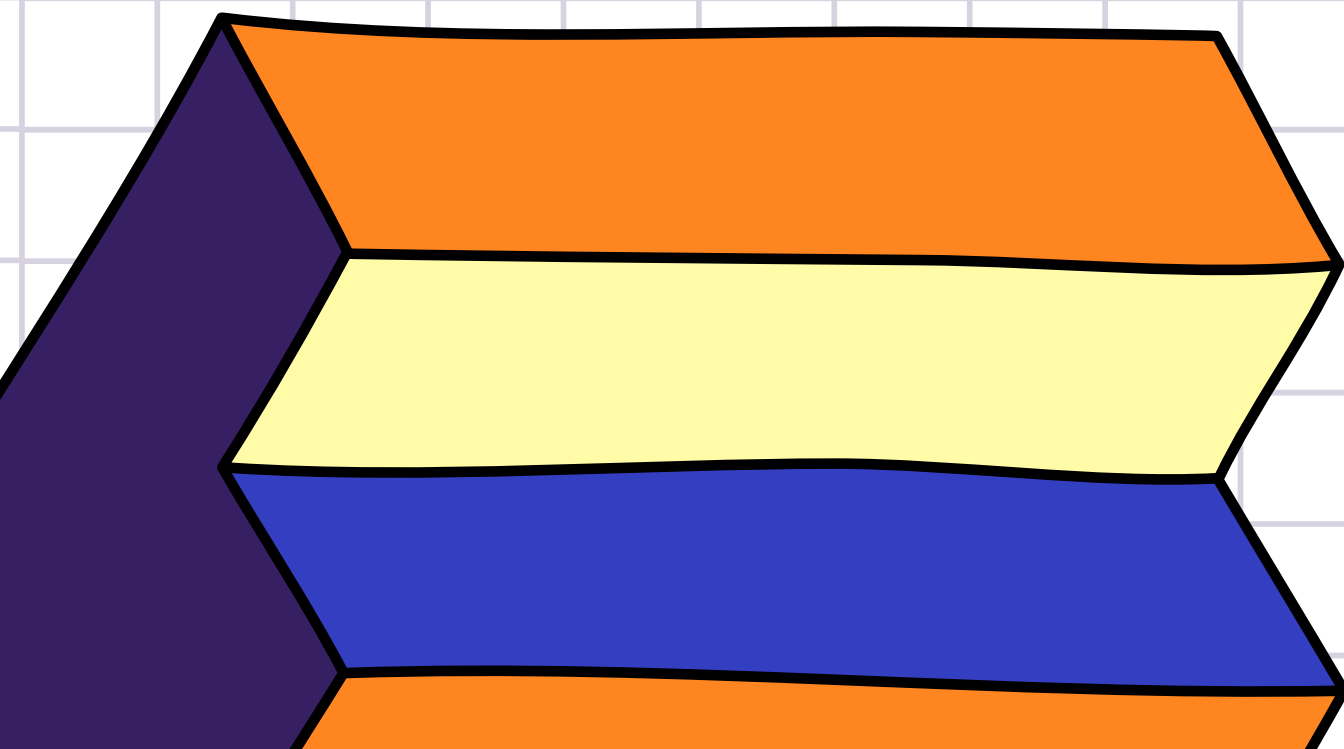
``

`<a>`





ONDE APRENDER MAIS HTML?



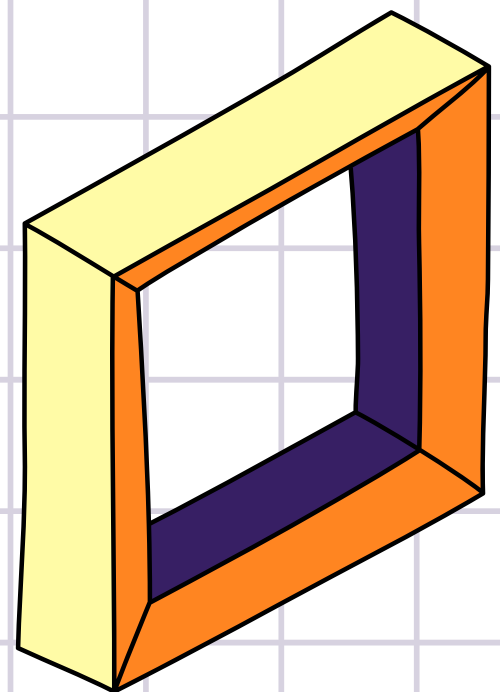
VAMOS AO COLAB!

APIS

APPLICATION PROGRAMMING INTERFACE

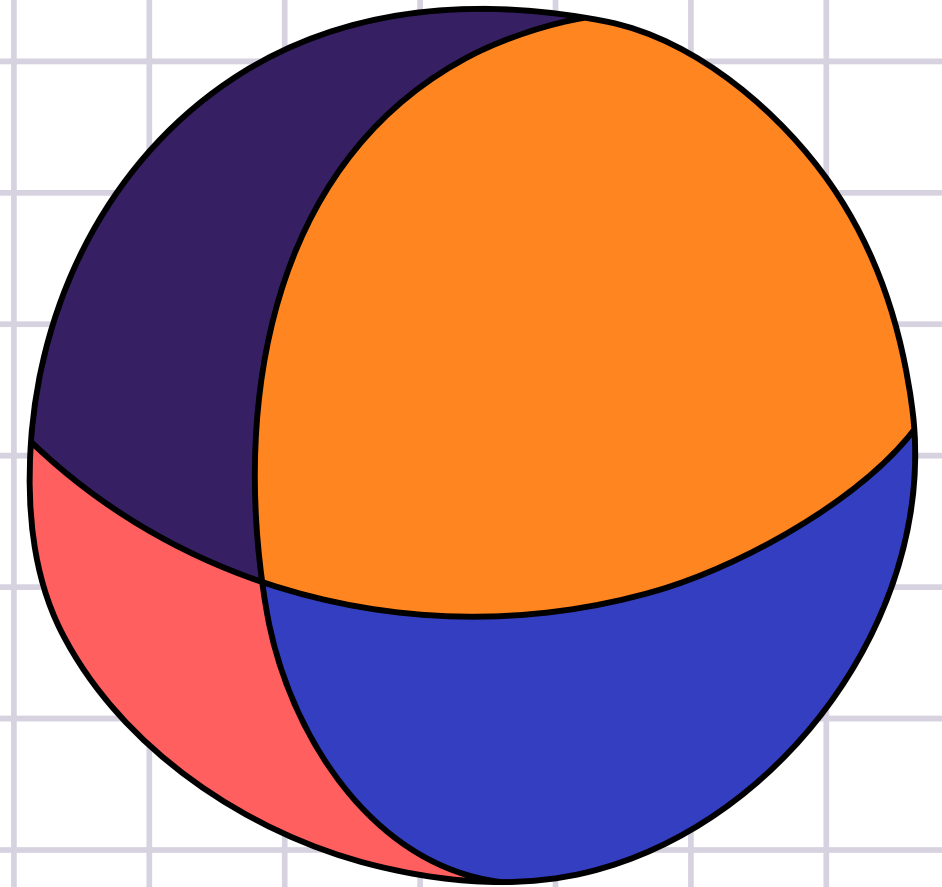
- É a maneira com que um software se comunicar com outro
- Em geral, alguns sistemas possuem APIs para obter dados da Web: Câmara dos Deputados, Yahoo Finance, electionsBR, IPEADData, Spotify, Twitter, CrowdTangle, dentre outros

**UTILIZAR APIS EM PYTHON É MUITO
SIMILAR AO QUE FIZEMOS COM O
BRASILPARALELO!**



API DA CÂMARA

<https://dadosabertos.camara.leg.br/swagger/api.html>



- **Lista de deputados em atividade:**

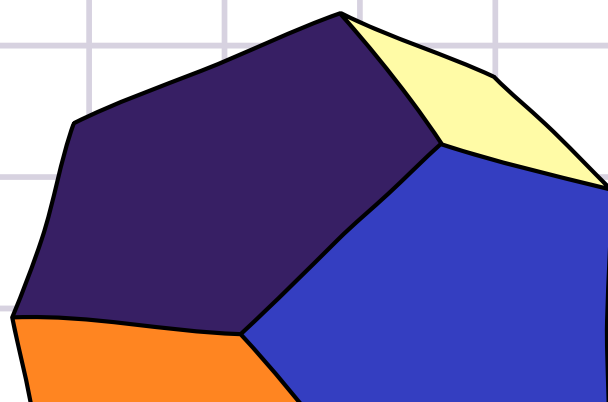
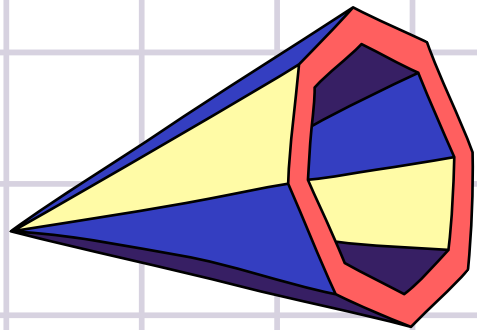
- <https://dadosabertos.camara.leg.br/api/v2/deputados?ordem=ASC&ordenarPor=nome>

- **Lista de PLs:**

- <https://dadosabertos.camara.leg.br/api/v2/proposicoes?ordem=ASC&ordenarPor=id>

- **Frentes parlamentares:**

- <https://dadosabertos.camara.leg.br/api/v2/frentes>



VAMOS AO COLAB!