

Python & Web Scraping: Transformando Dados Coletados em Vantagem Competitiva

Rodrigo Garcia Brunini

**Cientista de Dados
Engenheiro PhD
Desenvolvedor de Projetos**

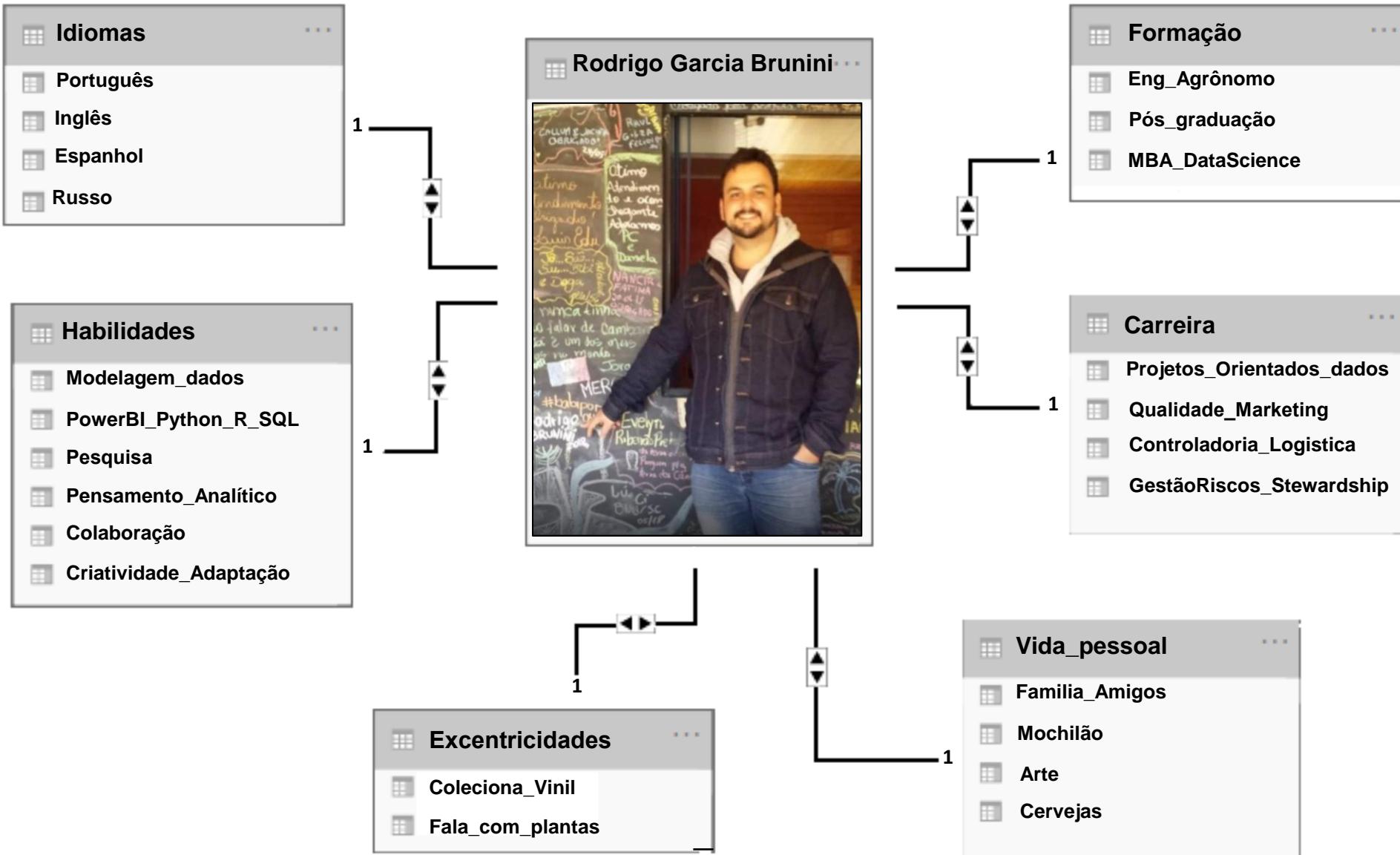
São Paulo - SP, 2026

Mr. Bean - 1990



Quem sou eu?

São Paulo - SP, Janeiro de 2026



O que vamos aprender hoje (OBJETIVOS)



Compreender Web Scraping

Entender os fundamentos técnicos e o papel estratégico do web scraping no ecossistema de ciência de dados modern.



Dominar Ferramentas Python

Bibliotecas essenciais como requests, BeautifulSoup e Selenium para coletar dados de forma eficiente.



Aspectos Legais e Éticos

Entender os fundamentos técnicos e o papel estratégico do web scraping no ecossistema de ciência de dados.

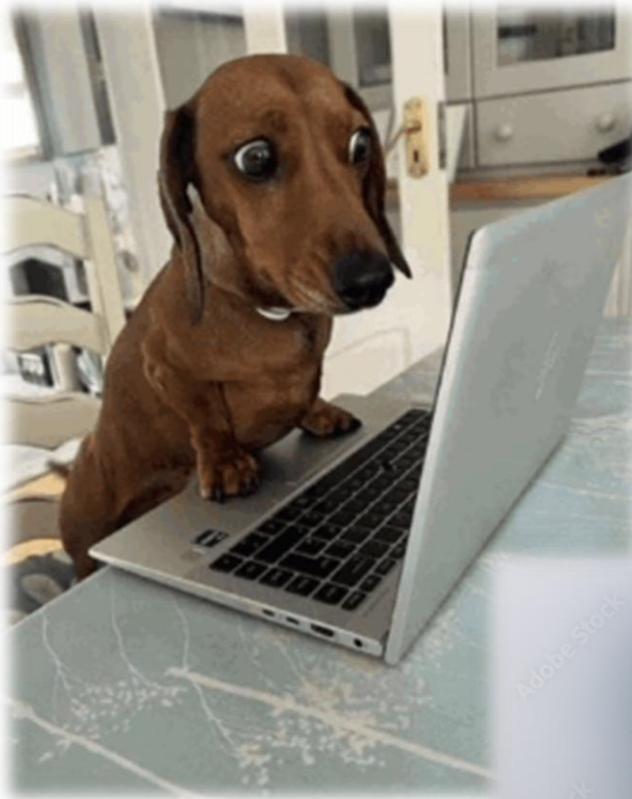


Gerar Vantagem Competitiva

Transformar dados coletados em insights acionáveis que impulsionam decisões estratégicas de negócio.

Para quem é o Web Scraping?

Iniciantes



- **Analistas de negócios buscando automatizar coleta de dados.**
- **Profissionais de marketing interessados em monitoramento de mercado.**
- **Estudantes de ciência de dados iniciando projetos práticos.**
- **Empreendedores querendo entender sua concorrência.**

Avançados



- **Cientistas de dados expandindo seu portfolio.**
- **Desenvolvedores Python implementando soluções de ETL.**
- **Pesquisadores acadêmicos coletando dados para estudos.**
- **Consultores estratégicos agregando valor aos clientes.**

Agenda: Nossa Jornada

01

Fundamentos & Contexto

Introdução aos dados, python e web scraping, importância estratégica e casos de uso reais no mercado.

04

Demonstrações Práticas

Ambiente dev, código de coleta, documentação, tratamento e análise de dados reais.

02

Aspectos Técnicos

Bibliotecas Python, estrutura de projetos e fluxo de trabalho de coleta de dados.

05

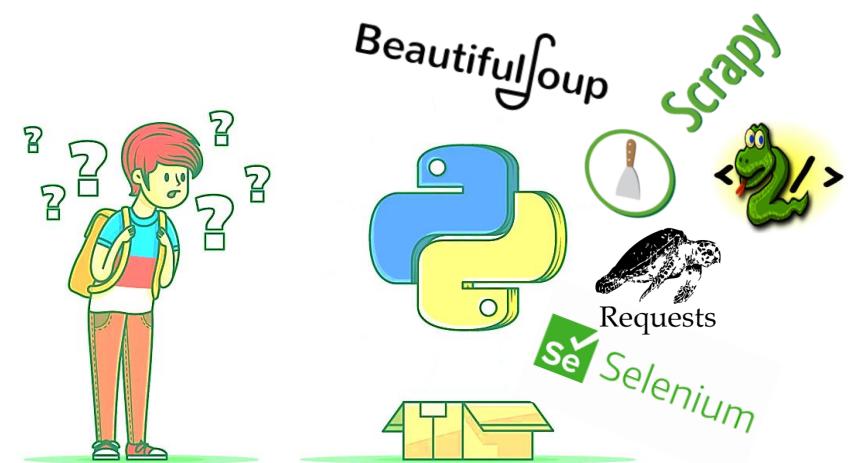
Insights & Dúvidas

Bate Papo, Dúvidas e dicas.

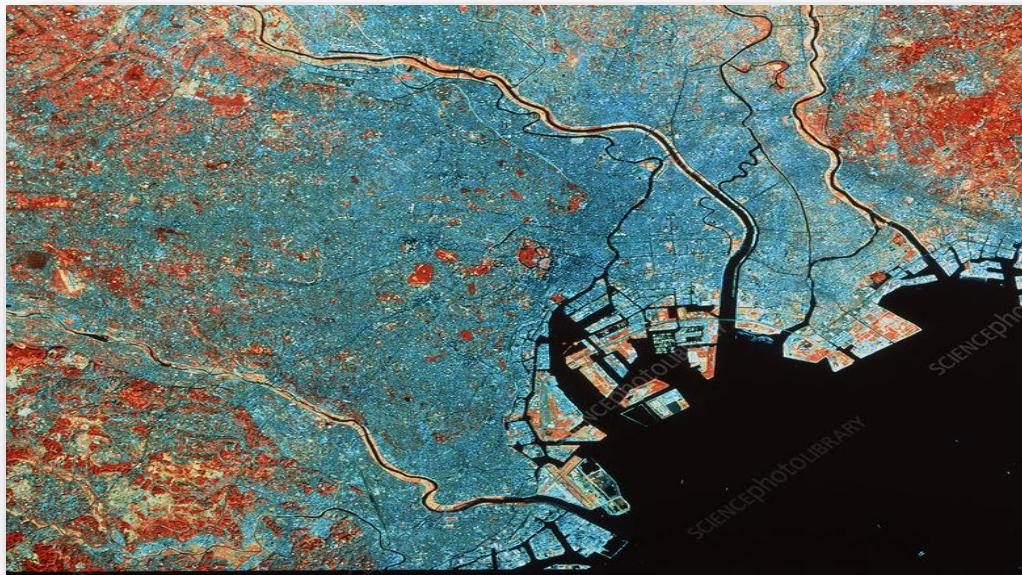
03

Legal & Ética

LGPD, responsabilidades legais e boas práticas para coleta responsável.

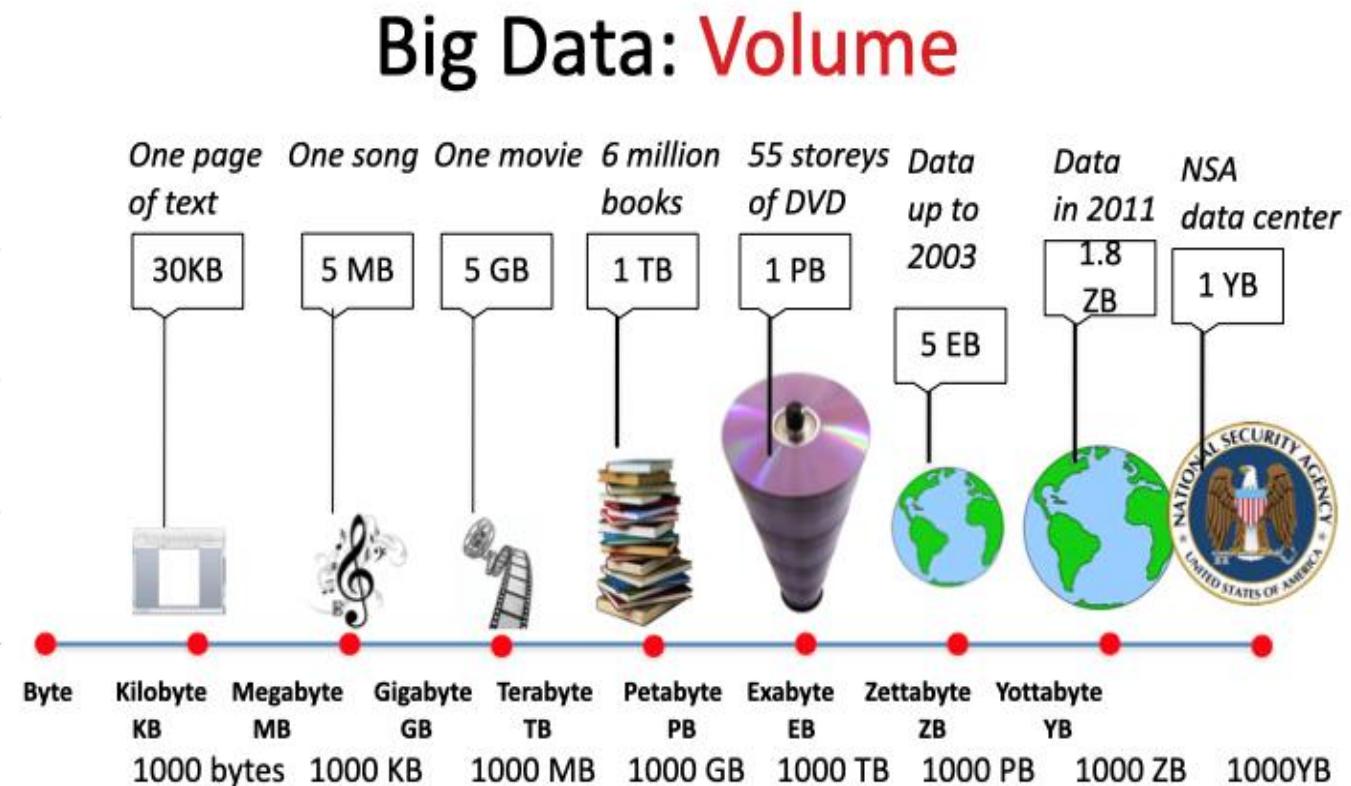
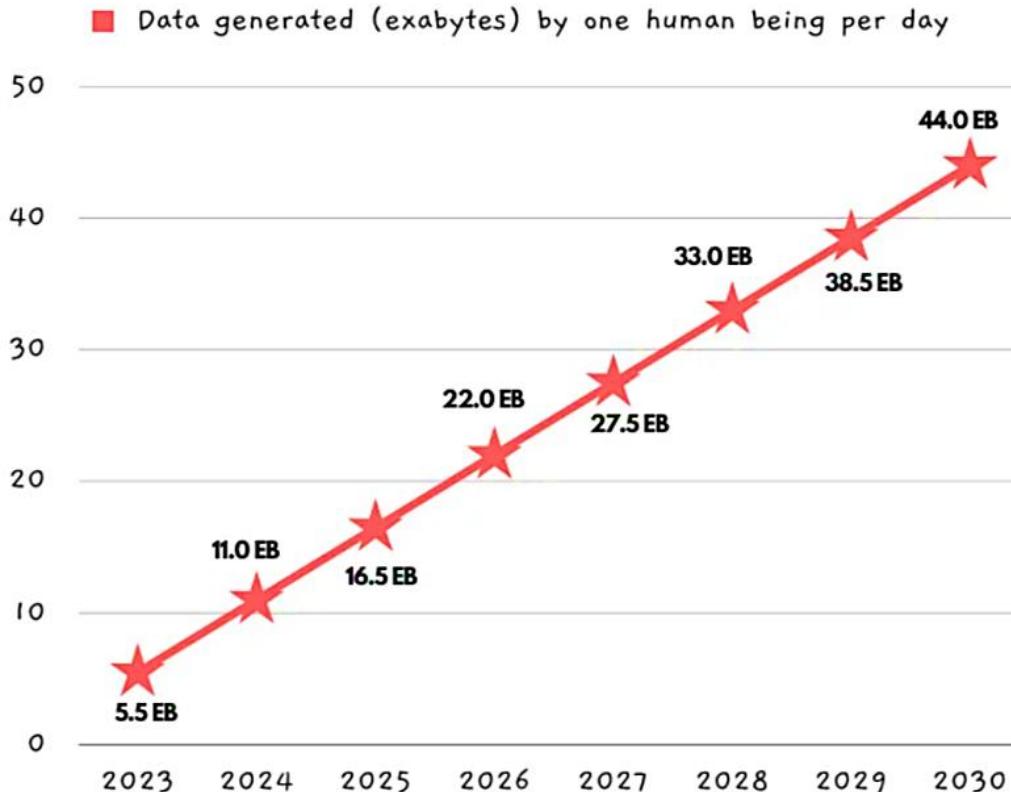


Dados - Porque?



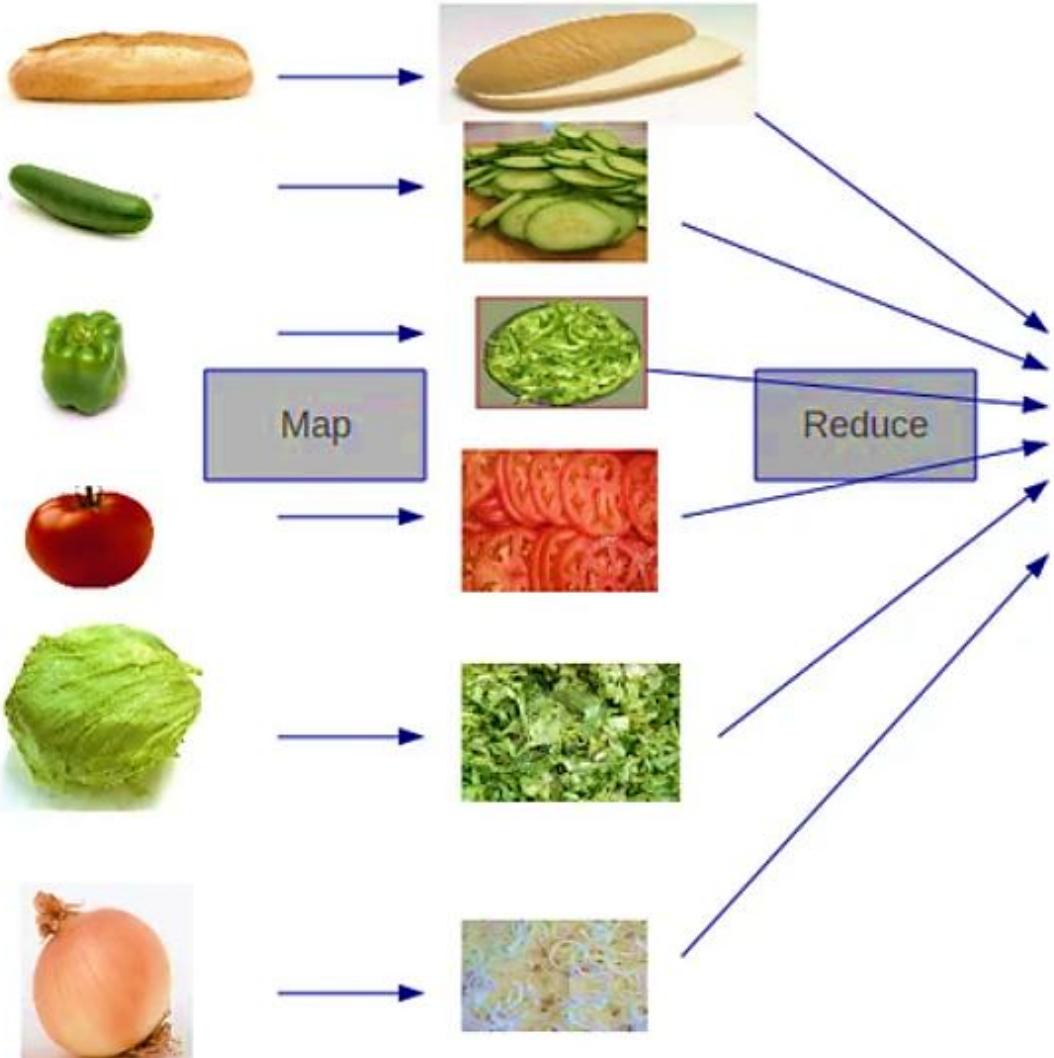
Dados – Big Data

Em média, uma pessoa irá gerar 1,7 MB de dados por segundo até 2025

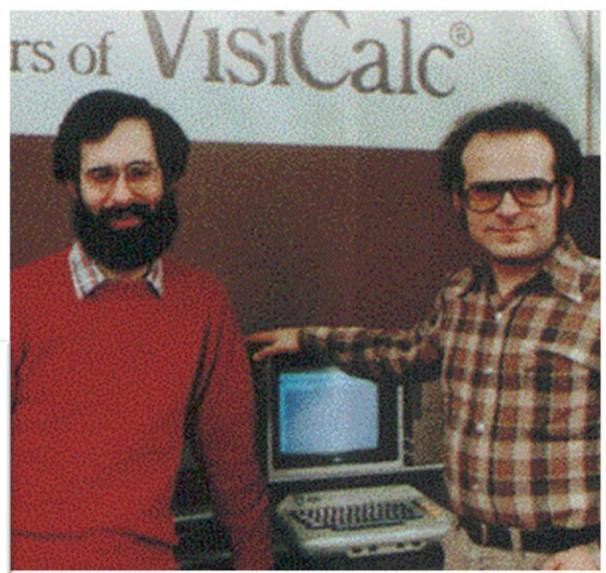


Dados - Big Data

Resumindo o que as empresas querem:

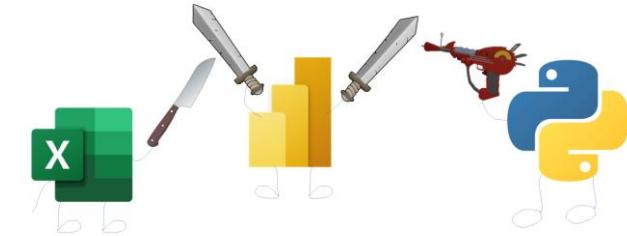
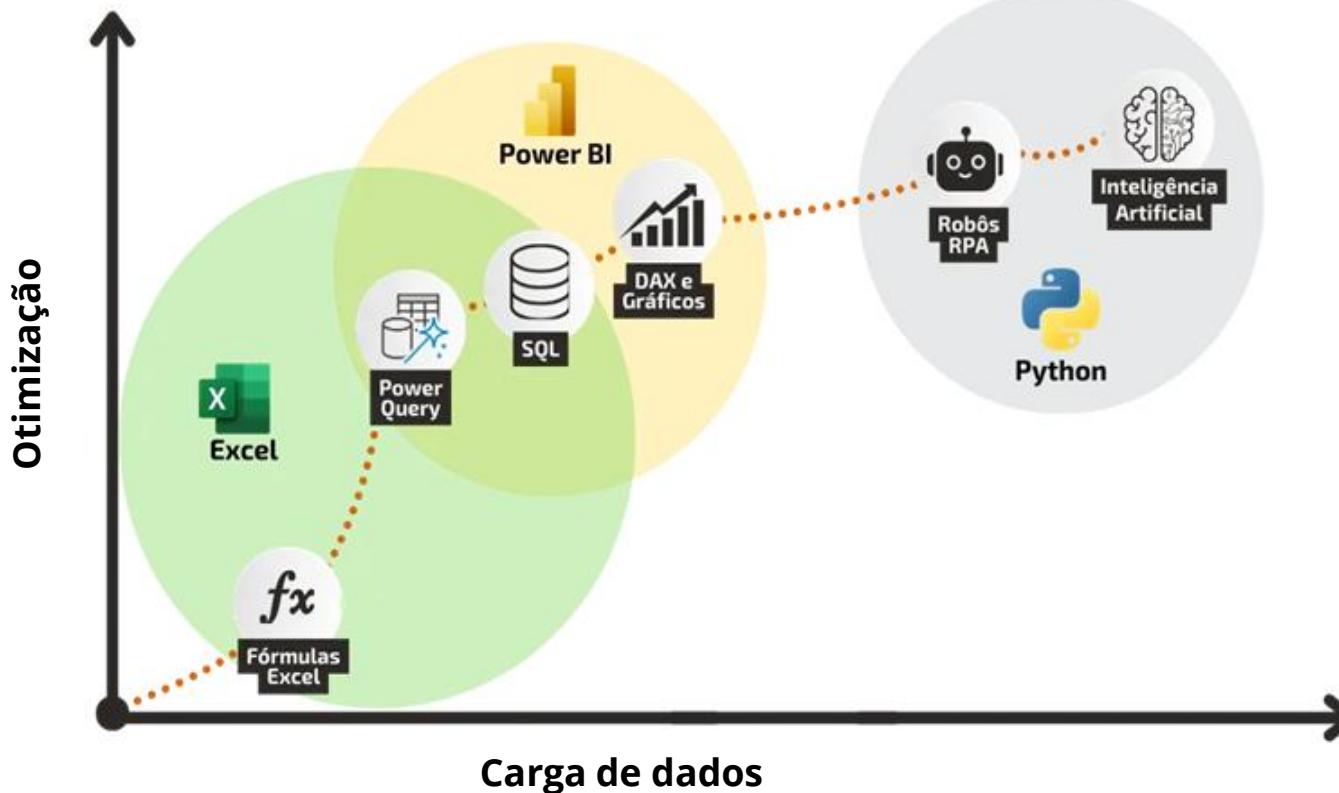


Como é de verdade



Dados - Big Data

Existe melhor ferramenta para trabalhar com dados?



Python: O que é?

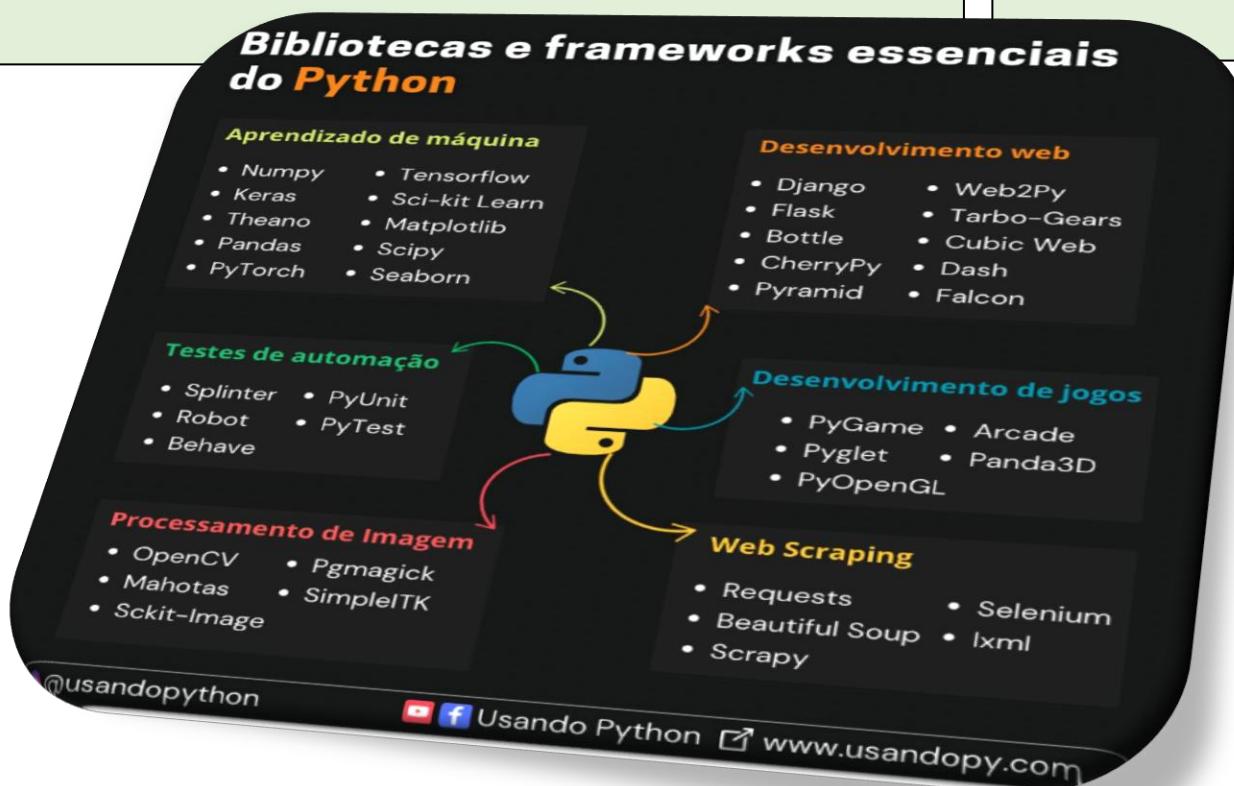
Python é uma linguagem de programação de alto nível, interpretada e de propósito geral, conhecida por sua sintaxe clara e legível.

Criada por Guido van Rossum em 1991, tornou-se uma das linguagens mais populares do mundo.

<https://www.python.org/>

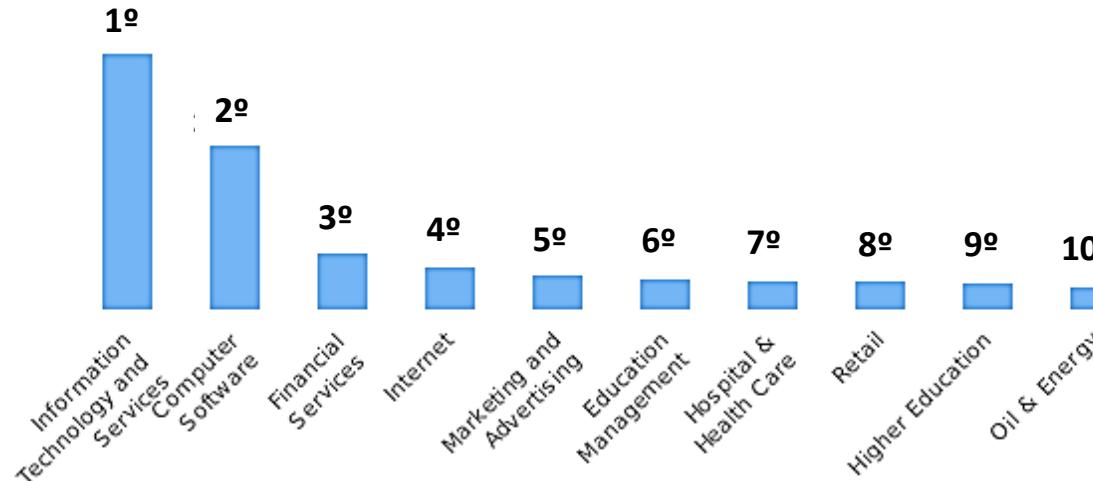
Porque usar?

Potencial de uso no dia-a-dia das rotinas de trabalho: dados, banco de dados, ciência de dados, automatizações, programação e estatísticas. Seus ganhos são quase imediatos em produtividade e eficiência.

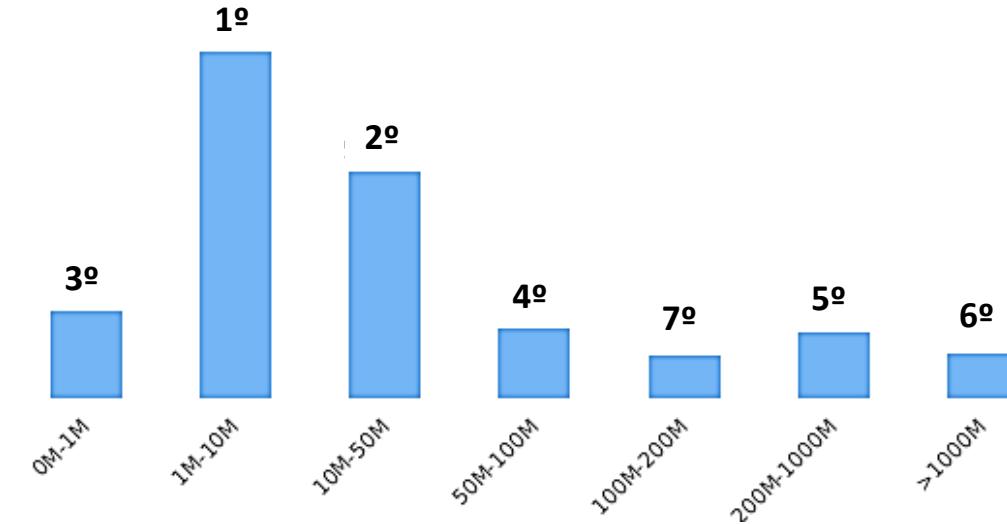


Uso do Python: Milhares de soluções em todos os campos das ciências

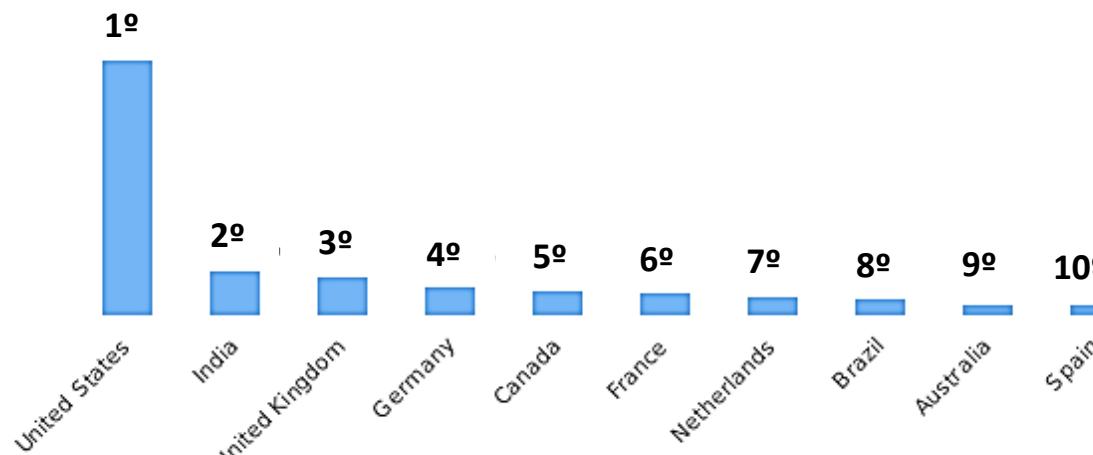
Distribuição de empresas que usam Python por setor



Uso de Python vs. Tamanho da Companhia (USD)



Distribuição de países que usam Python

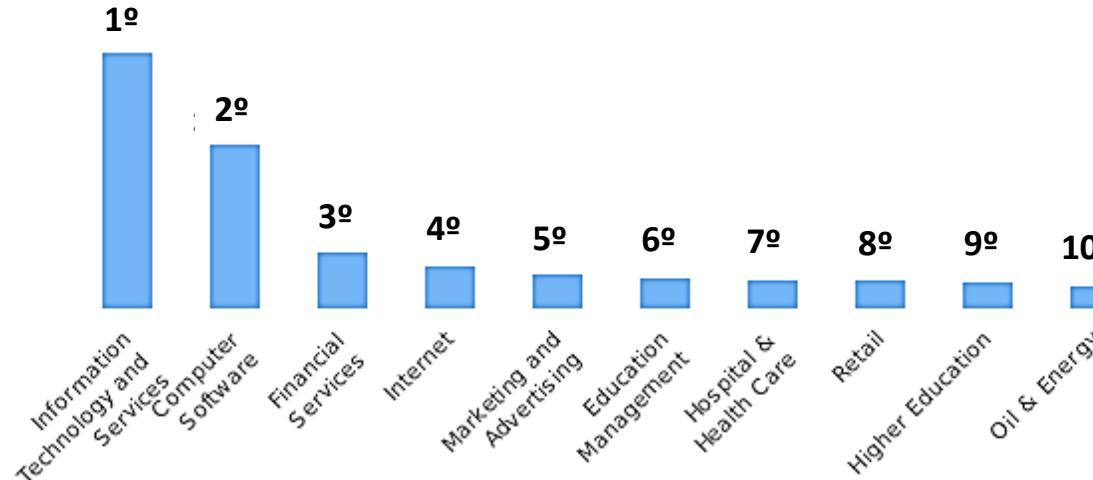


powered by enlyft.com

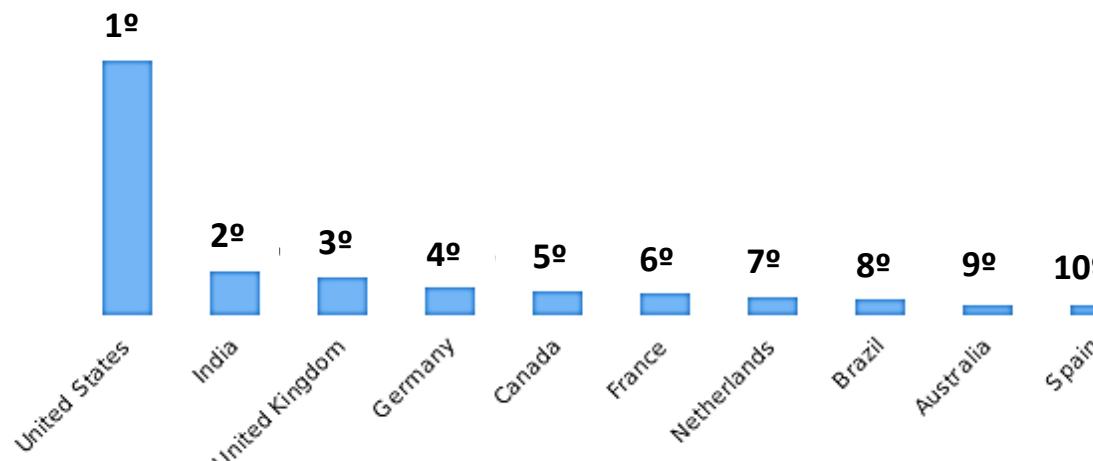
Top Companhias que usam Python



Distribuição de empresas que usam Python por setor



Distribuição de países que usam Python



Introdução ao Web Scraping



O Que é Web Scraping?

Web scraping é a técnica automatizada de extração de dados de websites, convertendo informações não estruturadas em formatos organizados e analisáveis. No ecossistema de ciência de dados, funciona como uma ponte essencial entre a web e seus pipelines analíticos.

Papel no Ecossistema de Dados

- **Fonte primária de dados quando APIs não estão disponíveis.**
- **Componente crítico em pipelines de ETL modernos.**
- **Habilitador de análises competitivas em tempo real.**
- **Ferramenta democrática de acesso a informações públicas.**



Web Scraping vs API



- API (Interface de Programação de Aplicações) é como uma porta organizada que um site ou serviço oferece, fornecendo dados já organizados e pensados para serem consumidos por programas, é mais seguro, estável e eficiente.
- Web scraping obtém dados “na ponta” da página, que podem mudar de formato a qualquer momento e exigem mais tratamento.

Dicas:

- Use API sempre que houver disponível (vai poupar tempo e reduzir erros);
- Recorra ao web scraping quando não houver API ou quando a informação só estiver disponível diretamente na página, lembrando que scraping pode ter limitações legais e técnicas (ex.: bloquear acessos ou exigir respeito a robots.txt).

dadosabertos.bcb.gov.br

Última atualização

Estatísticas do Pix

Estatísticas relativas a chaves Pix cadastradas no Diretório de Identificadores de Contas Transacionais (DICT), transações liquidadas no Sistema de Pagamentos Instantâneos (SPI)...

[HTML](#) [API](#) [OData](#) [JSON](#)

Estatísticas do SPI - Sistema de Pagamentos Instantâneos

Estatísticas das movimentações financeiras transitadas no SPI (Sistema de Pagamentos Instantâneos) processadas por meio de lançamentos nas contas PI mantidas pelos participantes...

[HTML](#) [API](#) [OData](#) [JSON](#)

SFN - BCO SAFRA S.A.

Conjunto que lista todas as informações desta instituição do Sistema Financeiro Nacional (SFN) que foram disponibilizadas no formato de dados abertos por determinação do Banco...

[HTML](#) [API](#) [JSON](#)

Informações Classificadas e Desclassificadas do BCB

Róis de Informações classificadas e desclassificadas ao amparo da LAI (Lei 12.527, de 18 de novembro de 2011) no âmbito do Banco Central do Brasil. O conjunto de dados contém as...

[CSV](#)

Títulos da carteira do BC para operações conjugadas com dealers

Os dados referem-se ao fechamento do dia de referência e contemplam títulos que podem ser utilizados em operações conjugadas com dealers. Essas operações funcionam de forma...

[HTML](#) [API](#) [OData](#) [JSON](#)

SFN - BANCO CENTRAL DO BRASIL

Popular

Cartões de crédito ativos

Conceito: Quantidade de cartões de crédito ativos. São considerados cartões ativos os cartões com histórico de transações nos últimos doze meses anteriores, ao fim de cada...

[JSON](#) [CSV](#) [wsdl](#) [HTML](#)



[dados](#) [Perguntas Frequentes](#) [Contato](#) [Sobre o Portal](#)

[Home](#) > Organizações > BCB/Decem > Estatísticas do Pix > **Estatísticas de Fraude no Pix (MED)**

[Ir para recurso](#)

Estatísticas de Fraude no Pix (MED)

URL: https://olinda.bcb.gov.br/olinda/servico/Pix_DadosAbertos/versao/v1/aplicacao#/recursos/EstatisticasFraudesPix

Este recurso reúne informações periódicas sobre transações Pix contestadas pelos usuários pagadores via Mecanismo Especial de Devolução (MED) previsto no Art. 41-B, inciso I, do Regulamento Anexo à Resolução BCB nº 1, de 12 de agosto de 2020, abrangendo tanto os volumes quanto os valores envolvidos em situações de suspeita de fraude. Dados serão publicados 30 dias após o término do mês.

Este recurso permite a construção da URL, com a manipulação de seus parâmetros, assim como a visualização dos dados da API.



Endereço padrão:

[https://olinda.bcb.gov.br/olinda/servico/Pix_DadosAbertos/versao/v1/odata/EstatisticasFraudesPix?\\$format=json&\[Outros Parâmetros\]](https://olinda.bcb.gov.br/olinda/servico/Pix_DadosAbertos/versao/v1/odata/EstatisticasFraudesPix?$format=json&[Outros Parâmetros])

Os parâmetros disponíveis podem ser visualizados na [documentação](#)

Formatos de Retorno:

Por padrão os dados podem ser retornados em 4 formatos:

- json (Padrão)
- xml
- text/csv
- text/html

robots.txt - O que é?

É um arquivo de texto na raiz do site (ex.: <https://site.com/robots.txt>) que orienta robôs sobre quais partes do site podem ou não acessar.

É uma convenção de cortesia, não uma barreira técnica.

Diretivas comuns

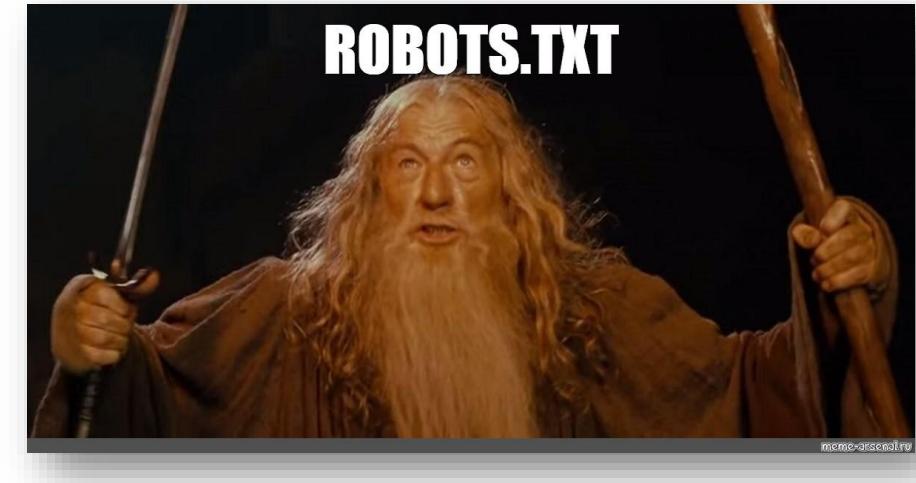
- User-agent: especifica a qual robô a regra se aplica (* = todos).
- Disallow: /caminho/ não acessar esse caminho.
- Allow: /caminho/ permite acesso a um subcaminho (usado junto com Disallow).
- Sitemap: <https://site.com/sitemap.xml> informa onde está o sitemap.

Exemplos rápidos

- Permitir tudo: User-agent: * Disallow:
- Bloquear tudo: User-agent: * Disallow: /
- Bloquear uma pasta: User-agent: * Disallow: /admin/

Limitações e boas práticas

- É informativo e voluntário (robôs maliciosos podem ignorá-lo).
- **Não proteja dados sensíveis (use autenticação/controle de servidor para isso).**
- Antes de raspar, verifique robots.txt e respeite as regras; faça requisições com ritmo (pausas) para não sobrecarregar o servidor.
- Checar programaticamente (rápido)
- Exemplo em Python: `import requests print(requests.get("https://www.exemplo.com/robots.txt").text)`



robots.txt - Exemplo

The screenshot shows a browser window displaying the robots.txt file for the IBGE website (ibge.gov.br/robots.txt). The robots.txt file content is as follows:

```
User-agent: *
Disallow: /administrator/
Disallow: /bin/
Disallow: /cache/
Disallow: /cli/
Disallow: /components/
Disallow: /homolog/
Disallow: /includes/
Disallow: /installation/
Disallow: /language/
Disallow: /layouts/
Disallow: /libraries/
Disallow: /logs/
Disallow: /modules/
Disallow: /plugins/
Disallow: /tmp/
Disallow: /templates/
Disallow: /cgi-bin/
Allow: /templates/*.css
Allow: /templates/*.js
Allow: /modules/mod_estatisticas_site/*
Allow: /modules/mod_mapgeo/*.css
Allow: /modules/mod_mensagem_portal/*.css
Allow: /modules/mod_popclock/*.css
Allow: /modules/mod_ppp/*.css
Allow: /modules/mod_questionario_censo/*
Allow: /modules/mod_quick_facts/*.css
Allow: /modules/mod_rbmc_download/*.css
Allow: /modules/mod_rbmc_estacoes/*.css
Allow: /modules/mod_rbmc_analise_dados/*
Allow: /modules/mod_redes_sociais/*.css
Allow: /modules/mod_resultado_censo_agro/*
Allow: /modules/mod_rmpg_download/*.css
Allow: /modules/mod_usodaterra/*.css
Allow: /modules/mod_saladeimprensa/*.css
Allow: /modules/mod_valores_nutricionais/*
Allow: /modules/mod_estatisticas_site/*
Allow: /modules/mod_mapgeo/*.js
Allow: /modules/mod_mensagem_portal/*.js
Allow: /modules/mod_popclock/*.js
Allow: /modules/mod_ppp/*.js
Allow: /modules/mod_questionario_censo/*
Allow: /modules/mod_quick_facts/*.js
Allow: /modules/mod_rbmc_download/*.js
Allow: /modules/mod_rbmc_estacoes/*.js
Allow: /modules/mod_rbmc_analise_dados/*
Allow: /modules/mod_redes_sociais/*.js
```

The page also features a large image of police officers in riot gear with the text "CYBER ATTACK" and "CYBER SECURITY". A caption below the image states "Maioria das Empresas".

Fluxo de Processamento do Web Scraping



Inteligência Competitiva

Monitore preços, estratégias e movimentos da concorrência em tempo real. Empresas que dominam esta prática conseguem antecipar tendências de mercado e ajustar suas estratégias antes dos competidores.

Automação e Eficiência

Elimine coletas manuais que consomem horas de trabalho. Um script bem desenvolvido pode executar em minutos o que uma equipe levaria dias para compilar, liberando recursos para análise de valor agregado.

Tomada de Decisão Baseada em Dados

Transforme dados públicos dispersos em insights concentrados. Desde análise de sentimento em reviews até tendências de consumo, o scraping fornece o combustível para decisões estratégicas informadas.

Criação de Bases de Dados Proprietárias

Construa ativos de dados únicos que se tornam diferenciais competitivos. Empresas inovadoras criam datasets exclusivos combinando fontes públicas de formas que seus concorrentes não imaginaram.



Requisições HTTP

Biblioteca `requests` envia requisições GET/POST para servidores web, simulando navegação



Parsing HTML/XML

`BeautifulSoup` e `lxml` navegam pela estrutura DOM, extraindo elementos específicos



Automação Avançada

`Selenium` controla navegadores reais para sites com JavaScript dinâmico



Bibliotecas mais usadas nos projetos de Web Scraping

requests: Comunicação HTTP simplificada

BeautifulSoup4: Parser HTML/XML intuitivo

lxml: Parser de alta performance

Selenium: Automação de navegador

pandas: Estruturação e análise de dados

scrapy: Framework completo para projetos complexos

Bibliotecas para Python



Requests

(*requests 2.32.5 - <https://pypi.org/project/requests/>*)

- *Usada para fazer requisições HTTP (GET, POST, etc.)*
- *Baixa páginas HTML, APIs e arquivos*
- *Simples, rápida e muito popular*



BeautifulSoup

(*bs4 - <https://beautiful-soup-4.readthedocs.io/en/latest/>*)

- *Usada para extrair dados do HTML*
- *Faz o parse da página e permite navegar pelas tags*
- *Muito fácil de usar para páginas estáticas*
- *Funciona junto com Requests*



Selenium

(<https://selenium-python.readthedocs.io/>)

- *Automatiza navegadores (Chrome, Firefox, Edge)*
- *Usada para páginas com JavaScript pesado*
- *Simula ações humanas (cliques, rolagem, formulários)*
- *Ideal para sites dinâmicos*



Scrapy

(<https://docs.scrapy.org/en/latest/>)

- *Framework completo de web scraping*
- *Gerencia requisições, filas, paralelismo e pipelines*
- *Ideal para projetos grandes e crawlers*
- *Muito performático e escalável*



(<https://pypi.org/project/lxml/>)

- *Parser de HTML e XML muito rápido*
- *Permite usar XPath e CSS selectors*
- *Ideal para grandes volumes de dados*
- *Mais performática que BeautifulSoup*



Planejamento & Identificação

Defina objetivos claros, identifique fontes de dados, analise a estrutura HTML dos sites-alvo e planeje a estratégia de navegação. Documente os elementos CSS/XPath necessários.



Coleta de Dados

Implemente scripts de requisição respeitando robots.txt e rate limits. Configure tratamento de erros, retentativas e logging. Considere uso de proxies para maior robustez.



Limpeza e Validação

Remova duplicatas, trate valores nulos, normalize formatos e valide tipos de dados. Aplique regex para extração de padrões e transformações necessárias.



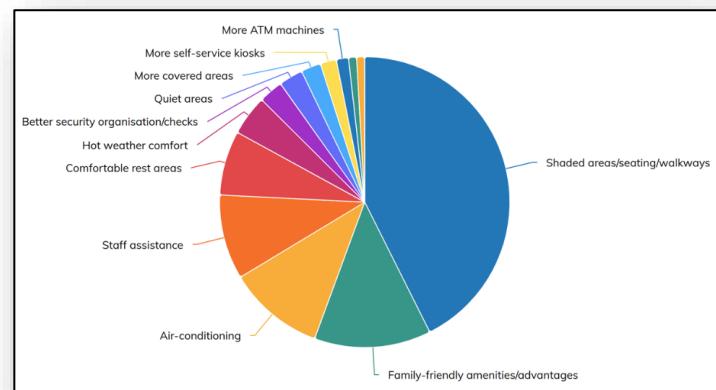
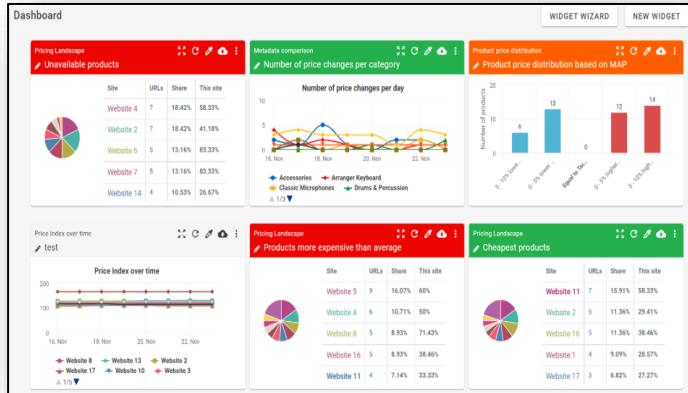
Armazenamento

Escolha o formato adequado: CSV para simplicidade, JSON para hierarquias, SQL para relacionamentos, ou NoSQL para flexibilidade. Implemente versionamento dos dados.



Análise e Insights

Explore os dados com pandas, gere visualizações significativas, identifique padrões e anomalias. Transforme números em narrativas açãoáveis para stakeholders.



E-commerce & Varejo

- Monitoramento dinâmico de preços
- Análise de disponibilidade de produtos
- Tracking de reviews e sentimento

Inteligência Competitiva

- Mapeamento de estratégias concorrentes
- Análise de posicionamento de mercado
- Identificação de novos entrantes

Pesquisa de Mercado

- Coleta de dados demográficos
- Tendências de comportamento do consumidor
- Análise de segmentação

Suporte a Business Intelligence

Alimentação automatizada de dashboards executivos com KPIs de mercado, permitindo tomadas de decisão em tempo real baseadas em dados externos atualizados.

Geração de Bases Proprietárias

Criação de datasets únicos combinando múltiplas fontes públicas, gerando ativos de informação que se tornam diferenciais competitivos sustentáveis.

LGPD e Aspectos Legais

Lei Geral de Proteção de Dados no Contexto do Web Scraping

A LGPD (Lei nº 13.709/2018) estabelece regras sobre coleta, armazenamento e tratamento de dados pessoais no Brasil. No contexto de web scraping, é fundamental compreender que a lei se aplica quando há coleta de informações que possam identificar pessoas naturais.

1

Dados Públicos vs. Dados Pessoais

Nem todo dado disponível publicamente pode ser coletado livremente. Informações como CPF, endereços residenciais e dados de contato pessoal estão protegidos pela LGPD, mesmo se acessíveis online.

2

Web Scraping é Crime?

Web scraping em si não é crime. Torna-se ilegal quando viola direitos autorais, invade sistemas protegidos, coleta dados pessoais sem base legal, ou desrespeita os Termos de Uso do site que podem ter força contratual.

3

Responsabilidades do Operador

Quem coleta dados é responsável pela finalidade, segurança e uso adequado. Deve haver base legal (como legítimo interesse ou consentimento), transparência sobre o uso e medidas de segurança para proteger os dados coletados.



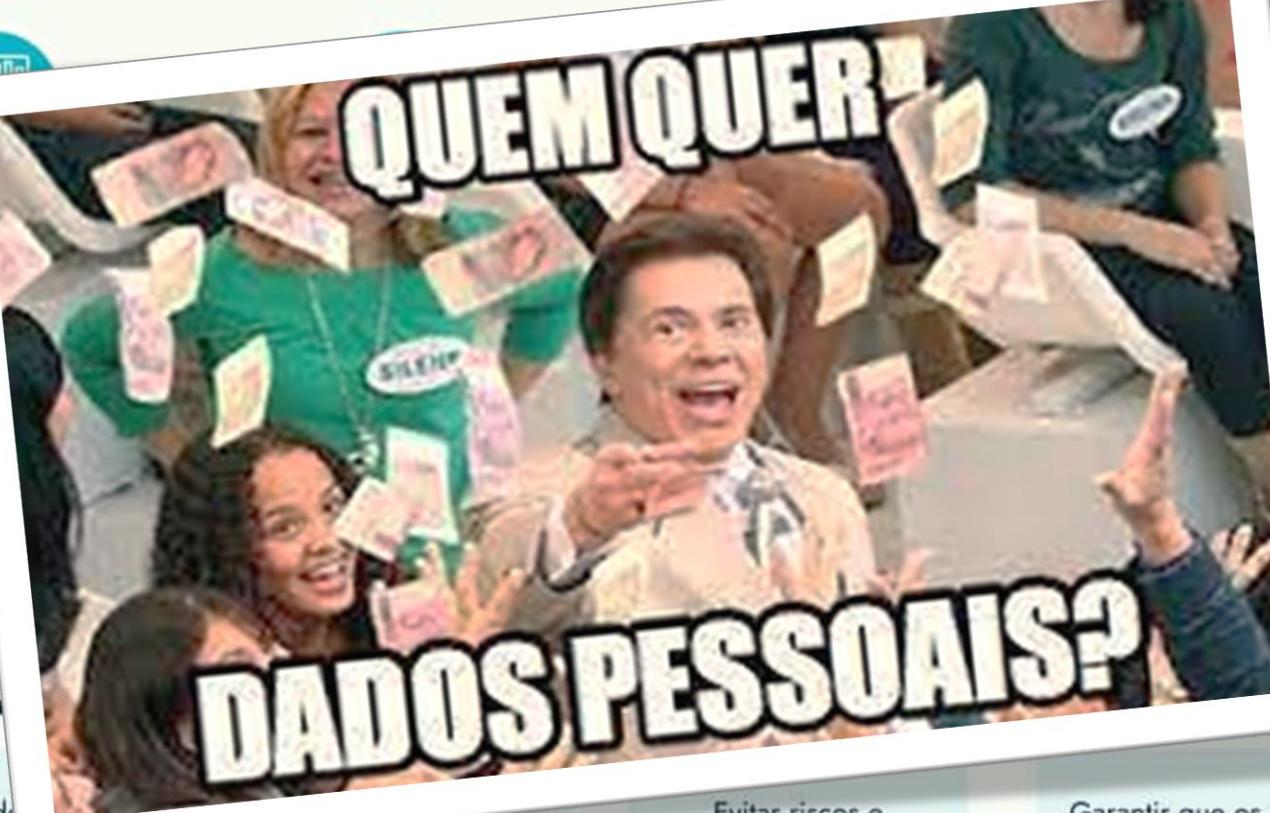
Diretrizes Práticas de Conformidade

Evite coletar dados pessoais sensíveis • Respeite robots.txt • Documente a finalidade e base legal • Implemente anonimização quando possível • Mantenha logs de acesso • Consulte jurídico em casos de dúvida.



LGPD e Aspectos Legais

Lei Geral de Proteção de Dados no Contexto do Web Scraping



Fin

Usar
com ob
e le



Transpa

Explicar de
clara como os
dados são usados.

Proteger os dados
contra vazamentos e
acessos indevidos.

Evitar riscos e
danos no tratamento
das informações.

Garantir que os dados
não sejam usados para
fins discriminatórios.



Qualidade dos dados

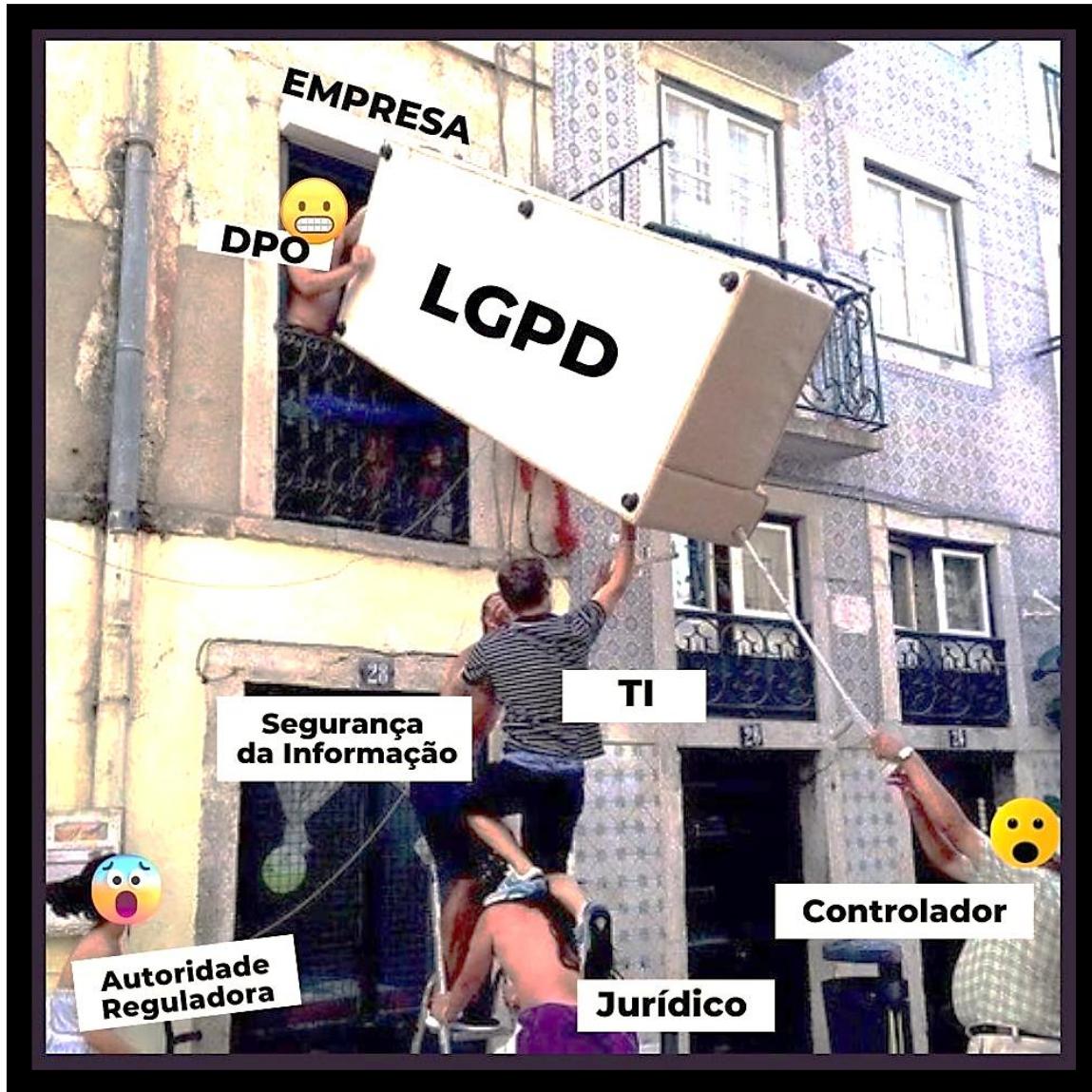
Manter as informações
corretas, atualizadas
e relevantes.



Responsabilização e prestação de contas

Demonstrar que a
empresa cumpre as
regras da LGPD.

LGPD e Aspectos Legais



**As empresas estão
preparadas?**

Importante:

- Avise seu superior e verifique as normas da empresa!
- Documente seu projeto!
- Sempre verifique os robots.txt e os termos de uso do site antes de fazer o web scraping.
- Prefira APIs oficiais quando existirem são estáveis e respeitam limites de acesso.

Prática



Objetivo para esta aula:

- Capturar dados públicos do IBGE sobre municípios (nome, código IBGE, população estimada, UF) e salvar em CSV para análise da **Quantidade Populacional por Município**.
- Mostrar passo a passo desde instalar bibliotecas até ter scripts funcionais e documentados.
- Montar um relatório simples mas funcional.

Por que esse exemplo?

- IBGE tem dados públicos e exemplos fáceis de entender. Além disso, mostra a diferença entre usar uma API e fazer scraping de HTML.

Metodologia (o que vamos usar e por quê):

- Primeira escolha: usar a API pública do IBGE — simples, rápida, confiável.
- Alternativa (quando não existe API): pegar a página HTML e extrair os dados com web scraping.
- Linguagem: Python, por ser simples e ter boas bibliotecas (requests, BeautifulSoup, pandas).

Base de Municípios IBGE

<https://servicodados.ibge.gov.br/api/docs/localidades#api-Municipios-municipiosGet>

The screenshot shows the Swagger API documentation for the IBGE Municipal Database. The left sidebar contains navigation links for various administrative divisions: AGLOMERACAOURBANA, DISTRITOS, MESORREGIOES, MICRORREGIOES, MUNICIPIOS, and PAISES. The main content area is titled 'Municípios' and describes the 'GET /localidades/municipios' endpoint. It includes a table of query parameters with detailed descriptions for 'orderBy' and 'view'. Below this, the 'Responses' section shows the schema for a successful 200 status code, which is an array of municipality objects.

Municípios

Obtém o conjunto de municípios do Brasil

GET

<https://servicodados.ibge.gov.br/api/v1/localidades/municipios>

Parâmetros

Name	Description
orderBy	▼ string Por padrão, os municípios são ordenados pela propriedade <code>id</code> , que corresponde ao respectivo identificador do município. Se desejar ordenar alfabeticamente pelo nome do município, use o parâmetro <code>orderBy</code> com o valor <code>nome</code> https://servicodados.ibge.gov.br/api/v1/localidades/municipios?orderBy=nome Obtém os municípios ordenados alfabeticamente pelo nome.
view	▼ string Modo de visualização. Por padrão, a resposta JSON é configurada para refletir a hierarquia à qual pertencem as localidades. A outra opção é configurar o parâmetro <code>view</code> com o valor <code>nivelado</code> , que faz com que as localidades sejam renderizadas no mesmo nível https://servicodados.ibge.gov.br/api/v1/localidades/municipios?view=nivelado

Responses

Status: 200 - Um array de municípios

Schema

```
[  
  {  
    id: number  
    nome: string  
    microrregiao: {  
      id: number  
      nome: string  
      mesorregiao: []  
    }  
  }  
]
```

Base de Municípios IBGE

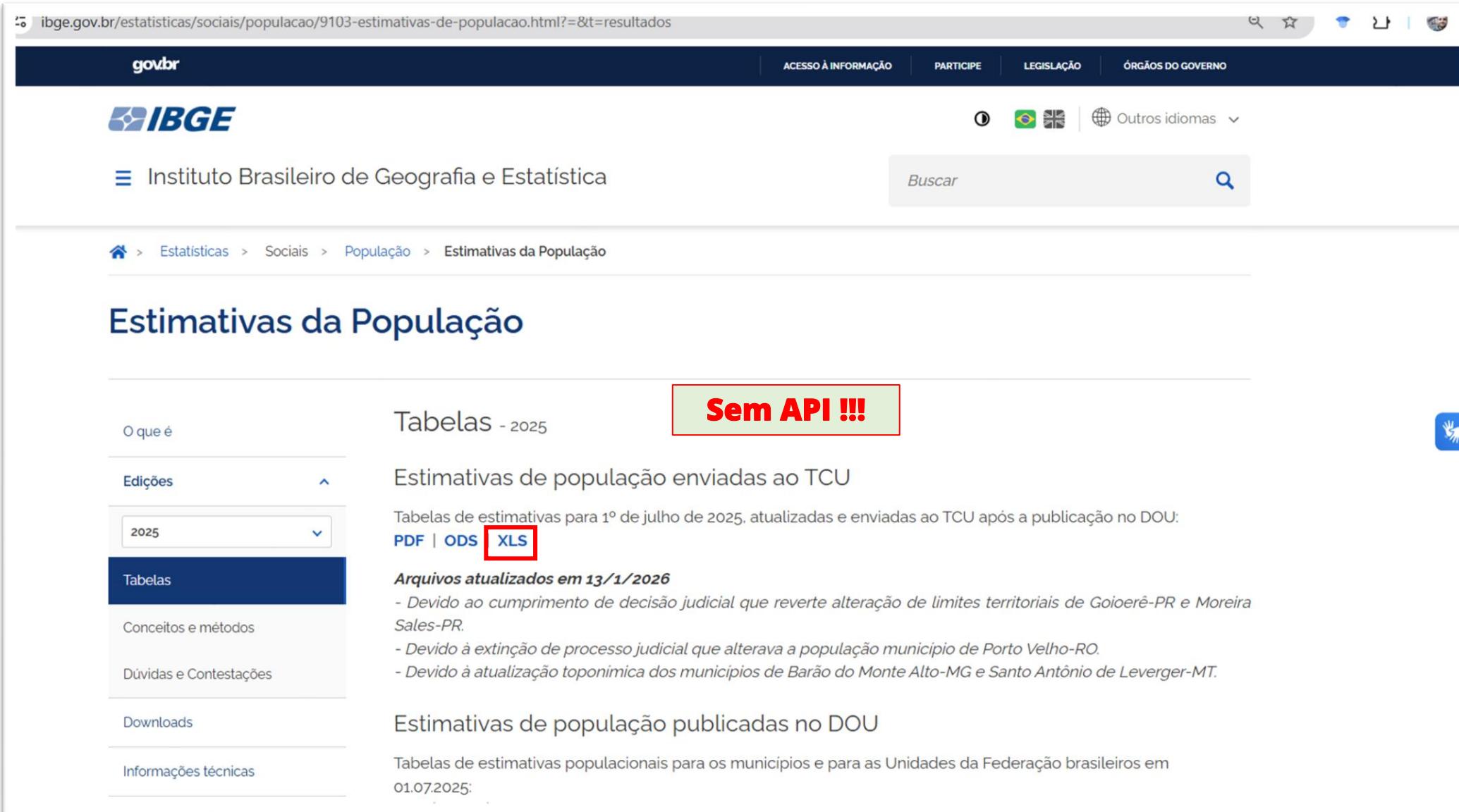
<https://servicodados.ibge.gov.br/api/v1/localidades/municipios>



The screenshot shows a browser window with the URL <https://servicodados.ibge.gov.br/api/v1/localidades/municipios>. The page title is "Ilos de formatação". The main content area displays a large JSON object representing the IBGE Municipal Database. The JSON structure includes various properties such as 'id', 'nome', 'sigla', 'UF', 'regiao', 'regiao-intermediaria', 'regiao-imediata', and 'mesorregiao'. The 'regiao' field contains detailed information about the municipality's location, including 'microrregiao', 'regiao': 'N', 'regiao': 'Norte', and 'regiao': 'Rondônia'. The 'regiao-intermediaria' and 'regiao-imediata' fields further divide the region into smaller administrative units. The 'mesorregiao' field provides a broader regional context. The JSON object is very long, spanning most of the page height.

Base de População IBGE

<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>



The screenshot shows the official website of the Instituto Brasileiro de Geografia e Estatística (IBGE). The URL in the address bar is <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>. The page title is "Estimativas da População". On the left, there's a sidebar with links for "O que é", "Edições" (with "2025" selected), "Tabelas" (which is highlighted in blue), "Conceitos e métodos", "Dúvidas e Contestações", "Downloads", and "Informações técnicas". The main content area starts with a red box containing the text "Sem API !!!". Below it, there's a section titled "Estimativas de população enviadas ao TCU" with download links for "PDF", "ODS", and "XLS" (the "XLS" link is highlighted with a red box). A note says "Arquivos atualizados em 13/1/2026" followed by a list of three updates. Further down, there's another section titled "Estimativas de população publicadas no DOU" with a note about tables for municipalities and Unidades da Federação.

ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados

gov.br

IBGE

Instituto Brasileiro de Geografia e Estatística

ACESSO À INFORMAÇÃO | PARTICIPE | LEGISLAÇÃO | ÓRGÃOS DO GOVERNO

Outros idiomas

Buscar

Estatísticas > Sociais > População > Estimativas da População

Estimativas da População

O que é

Edições

2025

Tabelas

Conceitos e métodos

Dúvidas e Contestações

Downloads

Informações técnicas

Tabelas - 2025

Sem API !!!

Estimativas de população enviadas ao TCU

Tabelas de estimativas para 1º de julho de 2025, atualizadas e enviadas ao TCU após a publicação no DOU:

[PDF](#) | [ODS](#) **XLS**

Arquivos atualizados em 13/1/2026

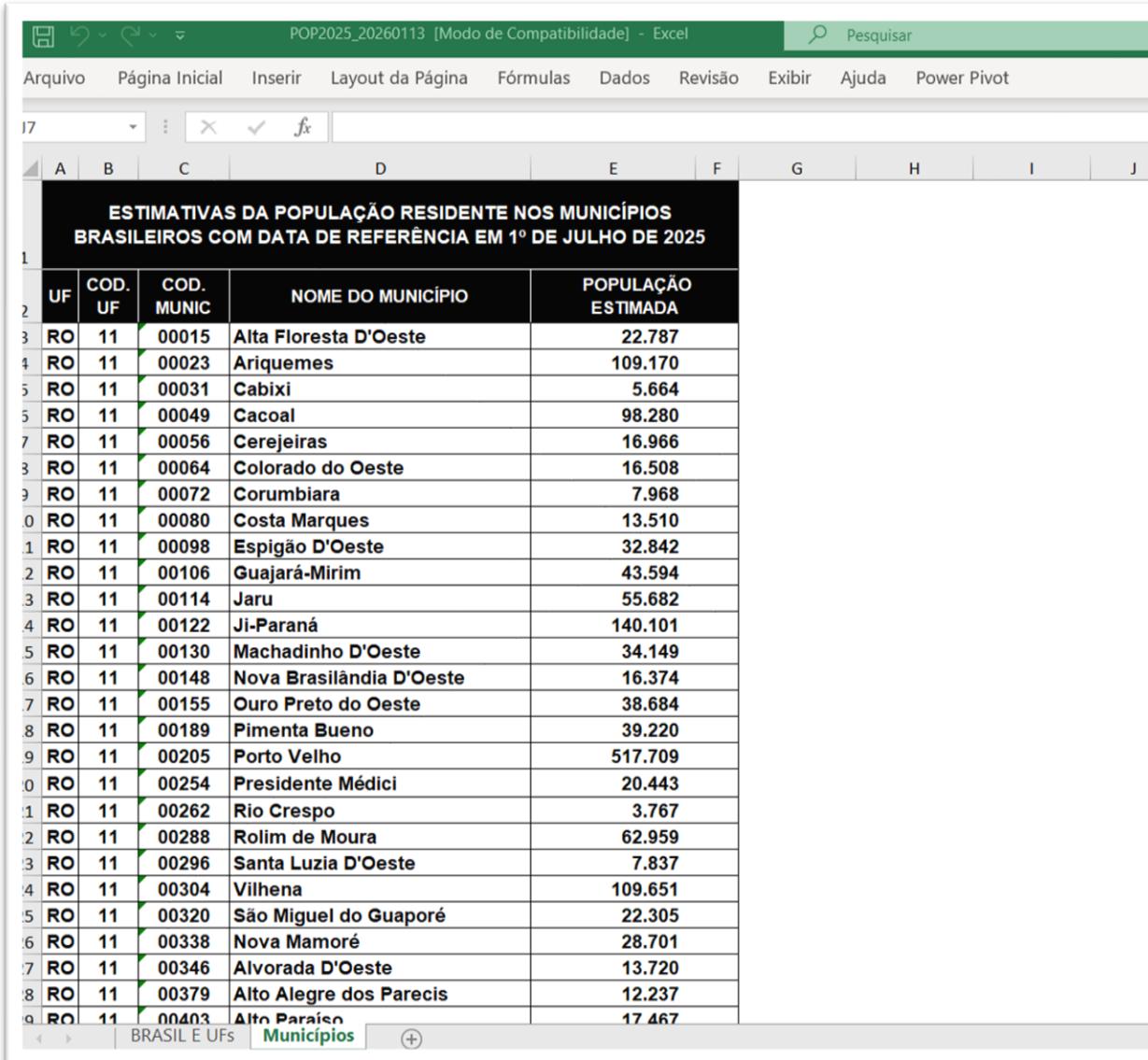
- Devido ao cumprimento de decisão judicial que reverte alteração de limites territoriais de Goioerê-PR e Moreira Sales-PR.
- Devido à extinção de processo judicial que alterava a população município de Porto Velho-RO.
- Devido à atualização topográfica dos municípios de Barão do Monte Alto-MG e Santo Antônio de Leverger-MT.

Estimativas de população publicadas no DOU

Tabelas de estimativas populacionais para os municípios e para as Unidades da Federação brasileiros em 01.07.2025:

Base de População IBGE

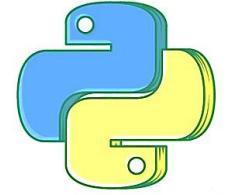
<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>



The screenshot shows an Excel spreadsheet with the following data:

ESTIMATIVAS DA POPULAÇÃO RESIDENTE NOS MUNICÍPIOS BRASILEIROS COM DATA DE REFERÊNCIA EM 1º DE JULHO DE 2025				
UF	COD. UF	COD. MUNIC	NOME DO MUNICÍPIO	POPULAÇÃO ESTIMADA
RO	11	00015	Alta Floresta D'Oeste	22.787
RO	11	00023	Ariquemes	109.170
RO	11	00031	Cabixi	5.664
RO	11	00049	Cacoal	98.280
RO	11	00056	Cerejeiras	16.966
RO	11	00064	Colorado do Oeste	16.508
RO	11	00072	Corumbiara	7.968
RO	11	00080	Costa Marques	13.510
RO	11	00098	Espigão D'Oeste	32.842
RO	11	00106	Guajará-Mirim	43.594
RO	11	00114	Jaru	55.682
RO	11	00122	Ji-Paraná	140.101
RO	11	00130	Machadinho D'Oeste	34.149
RO	11	00148	Nova Brasilândia D'Oeste	16.374
RO	11	00155	Ouro Preto do Oeste	38.684
RO	11	00189	Pimenta Bueno	39.220
RO	11	00205	Porto Velho	517.709
RO	11	00254	Presidente Médici	20.443
RO	11	00262	Rio Crespo	3.767
RO	11	00288	Rolim de Moura	62.959
RO	11	00296	Santa Luzia D'Oeste	7.837
RO	11	00304	Vilhena	109.651
RO	11	00320	São Miguel do Guaporé	22.305
RO	11	00338	Nova Mamoré	28.701
RO	11	00346	Alvorada D'Oeste	13.720
RO	11	00379	Alto Alegre dos Parecis	12.237
RO	11	00403	Alto Paraíso	17.467

MÃO NA MASSA



BeautifulSoup



Requests



Selenium

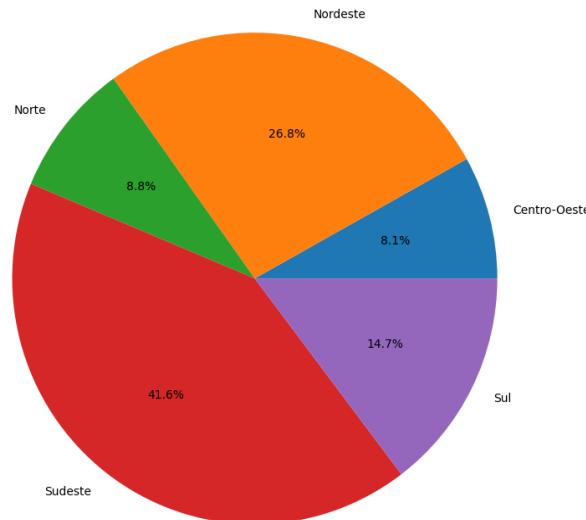


Situação Populacional dos Municípios e Regiões Brasileiras (2025)

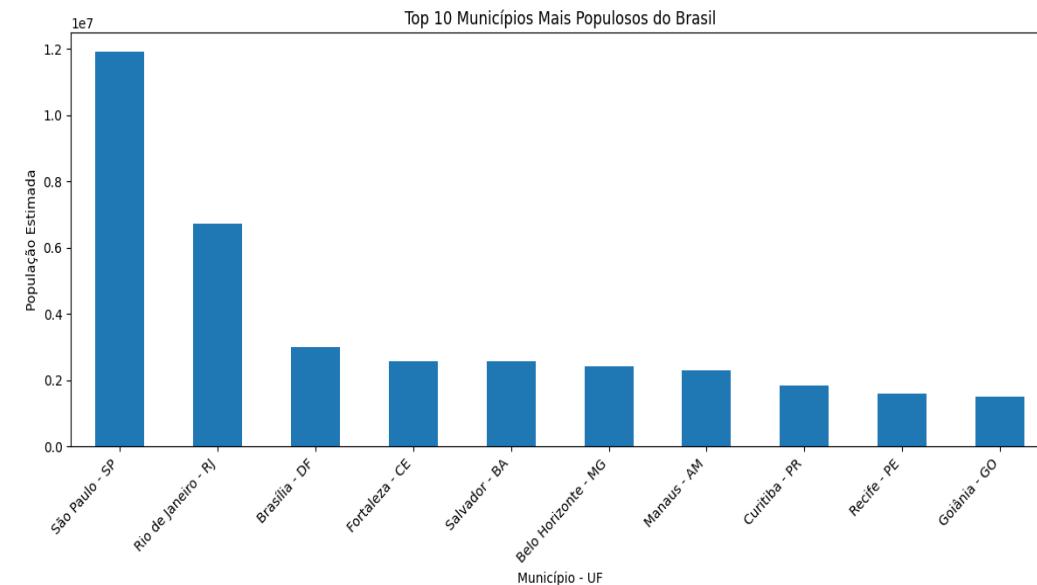
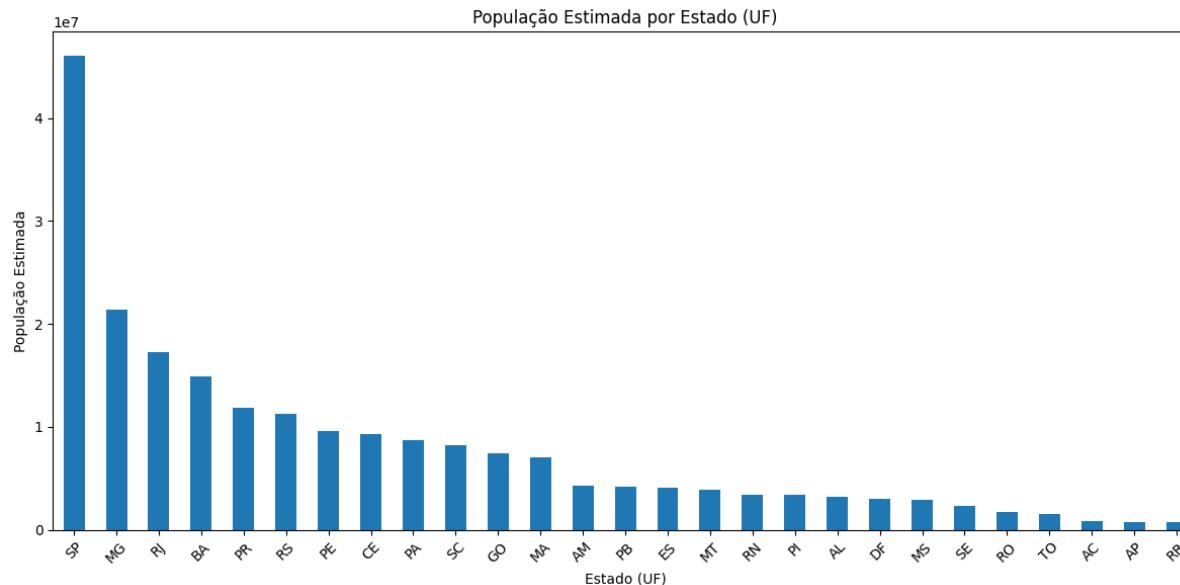
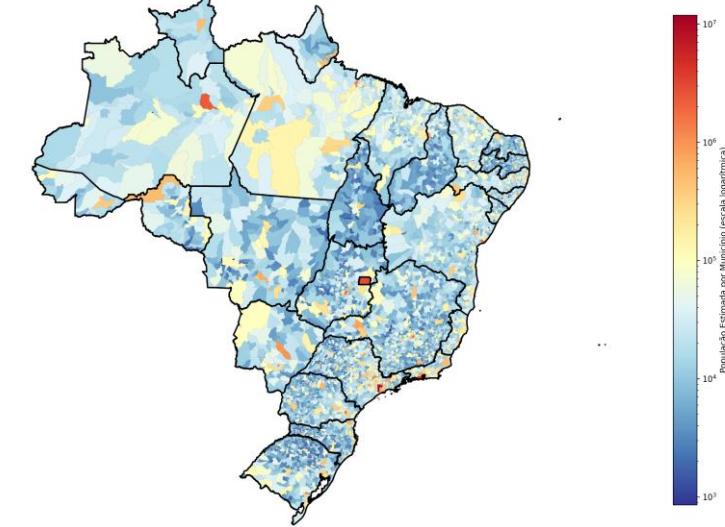
Relatório estatístico descritivo da população dos municípios brasileiros (2025).

Métrica	Valor
Total de Municípios	55.710
População Média	3.830.929
População Mediana	113.550
Desvio Padrão	21.741.951
Mínimo	8.560
Máximo	119.049.610
1º Quartil	53.480
3º Quartil	252.990
Valores Ausentes	-
Outliers Detectados	6.170

Distribuição da População por Região

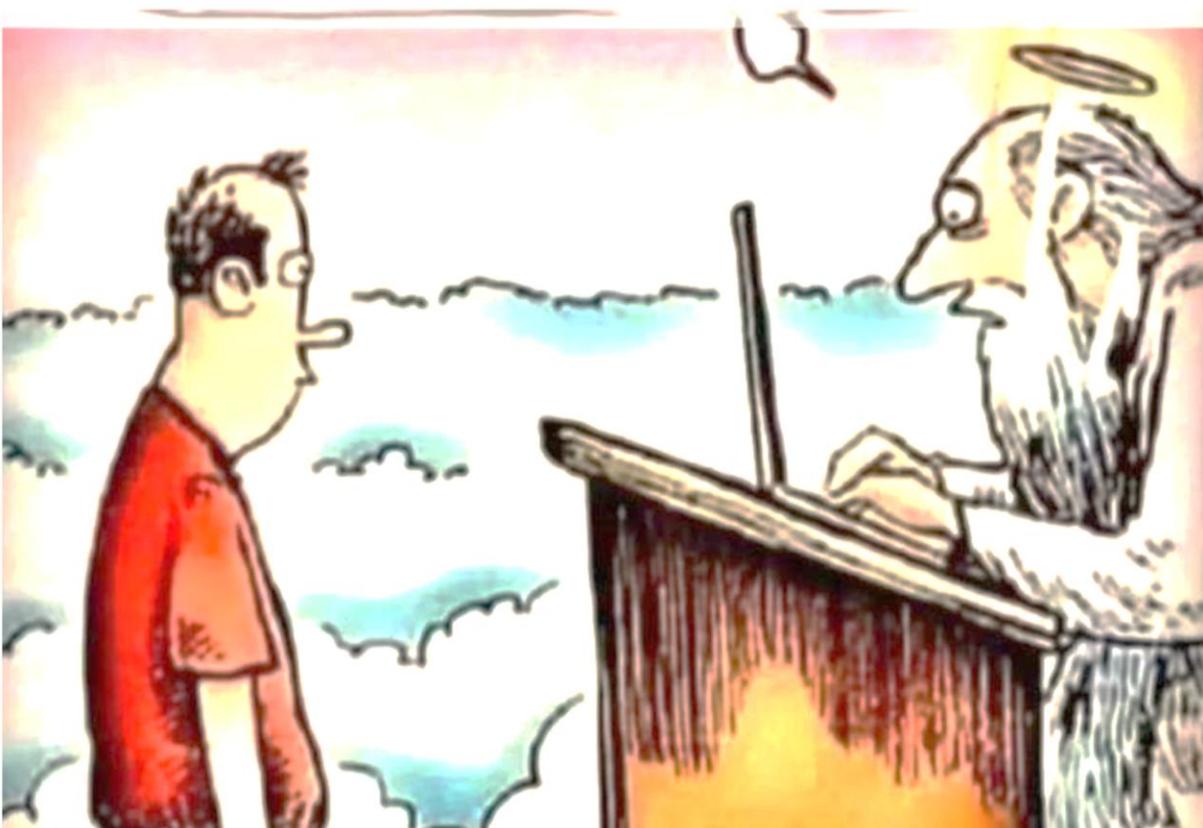


Mapa do Brasil — População Municipal com Divisas Estaduais



Estudem! Estudem! Estudem!

**Aqui diz que vc estudou análise de dados
por 10 anos e grava vídeos dizendo:
Seja um Analista de Dados em 7 dias!**



E o ChatGPT? Quando Usar?

Dicas de quando Usar:

- Geração de conteúdo não sensível (posts, e-mails, rascunhos)
- Brainstorming e ideação (assuntos, títulos, estratégias, fluxograma)
- Automação de tarefas repetitivas (resumos, templates, formatação)
- Assistência a programação (verificação de erros, documentação)
- Estudos

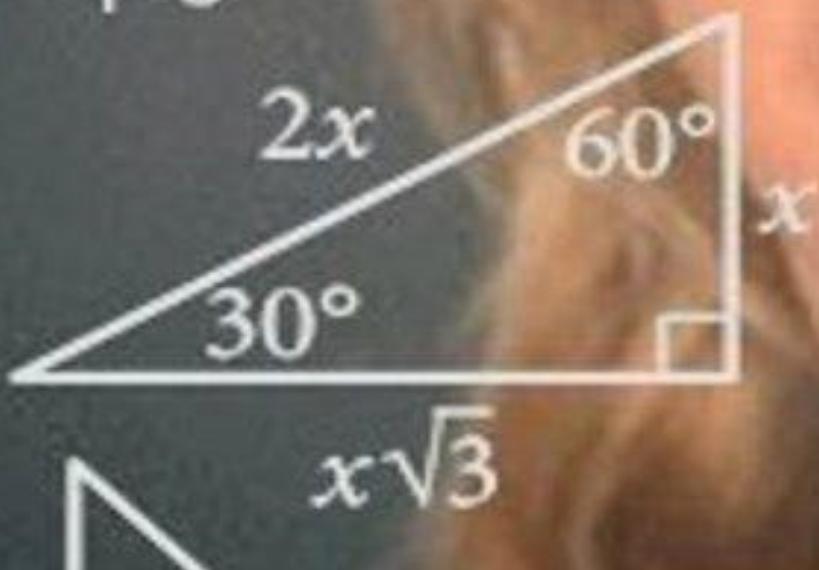
Quando não usar / ter cuidado:

- Decisões críticas sem revisão humana (**Alucinações, Histórico viciado**)
- Dados sensíveis ou pessoais sem proteção adequada
- Situações que exigem compreensão profunda do contexto cultural ou emocional
- Quando a veracidade absoluta das informações for imprescindível
- Uso pessoal no trabalho (**Assédio, conteúdo impróprio**)



DÚVIDAS?

\sin	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$
\cos	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$
\tan	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$



$$\int \sin x dx = -\cos x + C$$

$$\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C,$$

$$\int \operatorname{tg} x dx = -\ln |\cos x| + C$$

$$\int \frac{dx}{\sin x} = \ln \left| \operatorname{tg} \frac{x}{2} \right| + C$$

$$\int \frac{dx}{a^2 + x^2} = \frac{1}{a} \operatorname{arctg} \frac{x}{a} + C$$

Referências Bibliográficas

Livros:

Fávero, L. P.; Belfiore, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS®, Stata®, R® e Python.** GEN LTC, 2024.

McKinney, W. (2017). **Python for Data Analysis: Data Wrangling with pandas, NumPy, and Python.** O'Reilly Media

Mitchell, R. (2015). **Web Scraping with Python: Collecting More Data from the Modern Web.** O'Reilly Media.

Artigos:

Mumuni, A.; Mumuni, F. Automated data processing and feature engineering for deep learning and big data applications: a survey. **Journal of Information and Intelligence**, v. 3, n. 2, p. 113-153, 2025.

Rapôso, C. F. L. Desafios e Conformidade de Cloud Computing com a LGPD: Uma Revisão Sistemática. **Editora Impacto Científico**, p. 35-46, 2025.

Documentação Oficial:

Scrapy: <https://docs.scrapy.org/>

Beautiful Soup: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

Requests: <https://docs.python-requests.org/>

Selenium: <https://www.selenium.dev/documentation>

Python: <https://playwright.dev/python/>

pandas: <https://pandas.pydata.org/>

OBRIGADO!

Rodrigo Garcia Brunini
Eng. Agrônomo PhD.
MBA – Data Science and Analytics
Desenvolver de Projetos

<https://www.linkedin.com/in/rnbrunini>

"No futuro, o pensamento estatístico será tão necessário para a cidadania e liberdade como saber ler e escrever."

- H.G. Wells.