

Visit <https://bit.ly/ds-dojo-2025> for materials

# APIs!

structured data that is NOT  
from a CSV or Excel file

Visit <https://bit.ly/ds-dojo-2025> for materials

# APIs!

structured data that is NOT  
from a CSV or Excel file

 View Slides

 [Download PDF](#)

### Useful Links for this section:

- [Google Colab](#)
- [Inspect Element - Undocumented APIs](#)
- [Using Paginated APIs \(Video\)](#)
- [USGS Earthquake APIs](#)
- [Weather API](#)

Snorlax (Pokémon) - Bulbape...  +

bulbapedia.bulbagarden.net/wiki/Snorlax\_(Pokémon)

This article is about the species. For a specific instance of this species, see [Snorlax \(disambiguation\)](#).

**Snorlax** (Japanese: カビゴン *Kabigon*) is a [Normal-type](#) Poké**mon** introduced in [Generation I](#). It evolves from [Munchlax](#) when leveled up with high friendship. Snorlax can [Gigantamax](#) into **Gigantamax Snorlax** if it has the [Gigantamax Factor](#). In Generations I and II games, their [remakes](#), and [Pokémon X and Y](#), wild Snorlax can be found sleeping in inconvenient locations. In order to catch one, a Trainer must first awaken it with music from either a [Poké Flute](#) or the [Pokégear](#) set to the Poké Flute channel. Snorlax is the [game mascot](#) of [Pokémon Sleep](#).

**Contents** [hide]

- 1 [Biology](#)
  - 1.1 [Forms](#)
  - 1.2 [Evolution](#)
- 2 [Game data](#)
  - 2.1 [NPC appearances](#)
  - 2.2 [Pokédex entries](#)
  - 2.3 [Game locations](#)
    - 2.3.1 [In side games](#)
    - 2.3.2 [In events](#)
      - 2.3.2.1 [Wild Area News](#)
  - 2.4 [Held items](#)
  - 2.5 [Stats](#)
    - 2.5.1 [Base stats](#)
    - 2.5.2 [Pokéathlon stats](#)
  - 2.6 [Type effectiveness](#)
  - 2.7 [Learnset](#)

**Snorlax**  
Sleeping  
Pokémon #0143  
カビゴン  
*Kabigon*



Snorlax



Gigantamax Snorlax

Images on the Bulbagarden Archives



A screenshot of a web browser showing the homepage of PokeAPI. The page has a red header bar with the 'PokeAPI' logo on the left and navigation links for 'Home', 'About', 'API v2', and 'GraphQL' on the right. The main content area has a dark background with the large 'PokeAPI' logo in yellow and blue. Below it, the text 'The RESTful PokéMon API' and 'Serving over 10 billion API calls each month!' is displayed. A green section below contains the text 'All the PokéMon data you'll ever need in one place, easily accessible through a modern free open-source RESTful API.' followed by a blue link 'Check out the docs!'. At the bottom, there is a large white button with the text 'Try it now!'. The browser's address bar shows the URL 'pokeapi.co/api/v2/pokemon/s'.

PokeAPI

Home About API v2 GraphQL

# PokeAPI

The RESTful PokéMon API

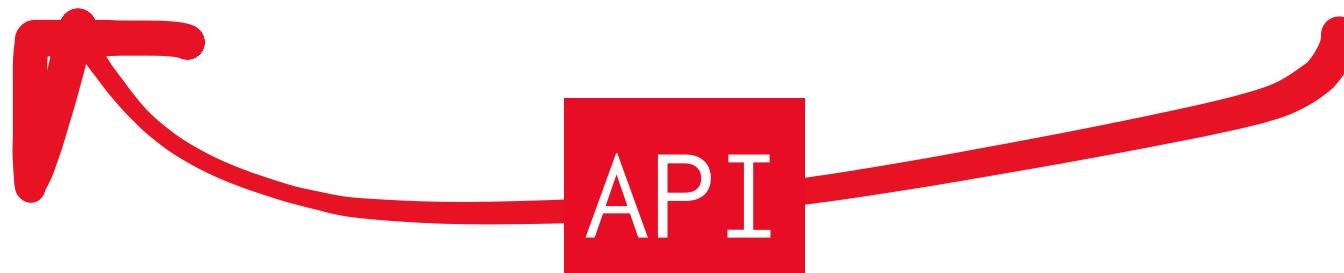
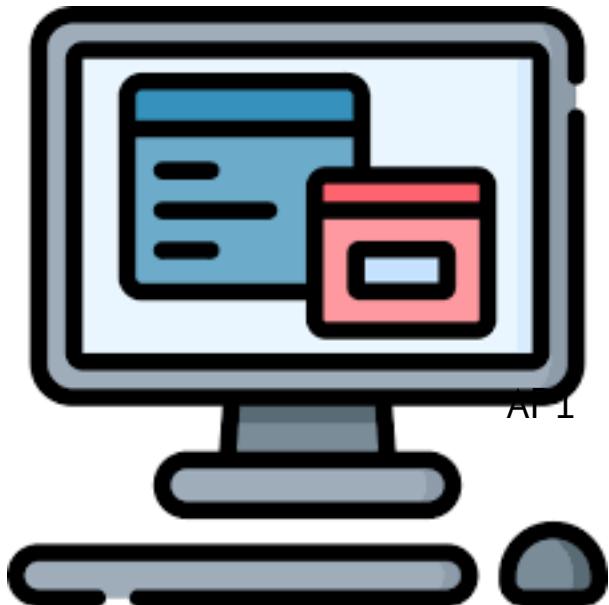
Serving over 10 billion API calls each month!

All the PokéMon data you'll ever need in one place,  
easily accessible through a modern free open-source RESTful API.

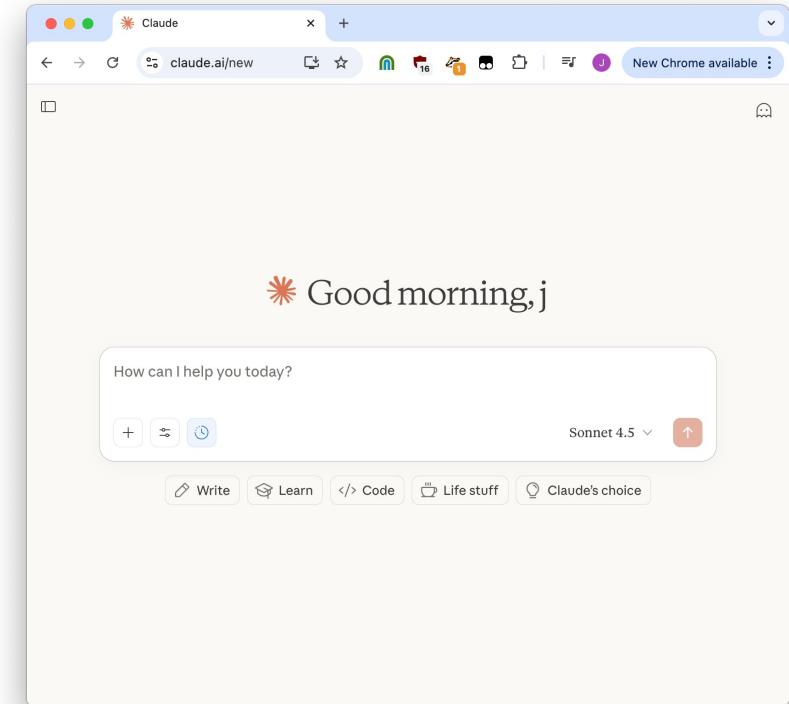
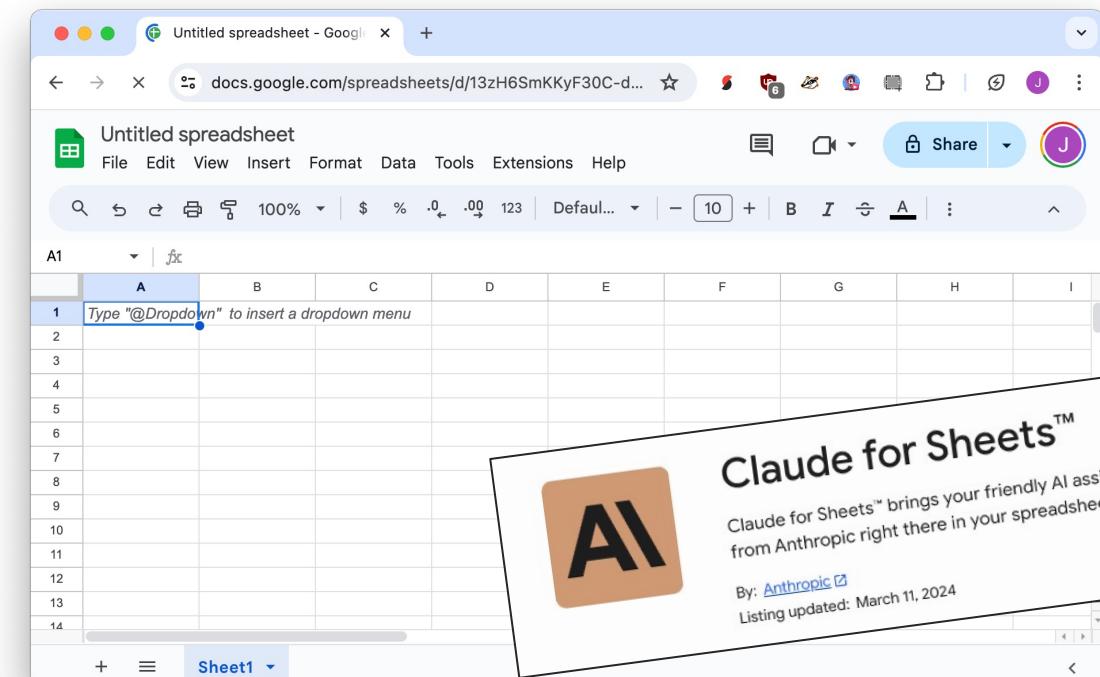
[Check out the docs!](#)

Try it now!

pokeapi.co/api/v2/pokemon/s



API



API

**Key successfully added!**

You must keep a record of the key below as you will not be able to view it again.

sk-ant-api03-Wriod0W3KzsL\_-osV4lnJCMfdZ6...  
nPxe-cpCdewNwV2sRKl\_pfrkj1mQCTqoDcEo1fJX...  
cld7exevD-Q-y5y\_KAAA

API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them. Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically disable any API key that has leaked publicly. Enable tracking to see usage per API key on the [Usage page](#).

NAME	SECRET KEY	TRACKING	CREATED	LAST USED	PERMISSIONS
Secret key	sk-...MJ0b	+ Enable	Mar 6, 2023	Apr 13, 2023	All
Secret key	sk-...e67A	+ Enable	Mar 18, 2023	Apr 19, 2023	All
Secret key	sk-...1PWe	+ Enable	Mar 18, 2023	Mar 18, 2023	All
Secret key	sk-...M1LX	+ Enable	Mar 19, 2023	Mar 19, 2023	All
Secret key	sk-...2ABE	+ Enable	Mar 23, 2023	Never	All

API keys

Close

Anthropic Console

ANTHROPIC

Your Organization Members & Invites Plans & Billing API Keys Logs Usage

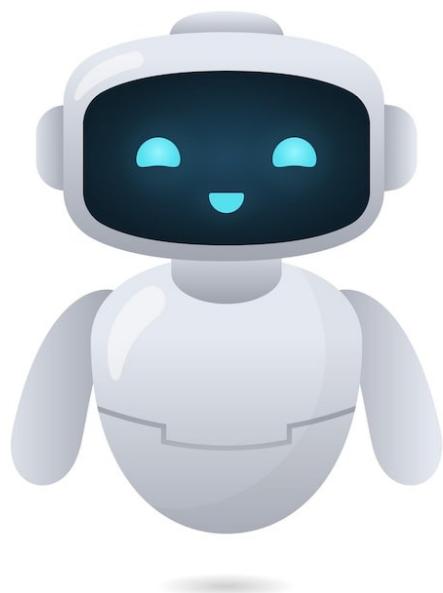
API Keys

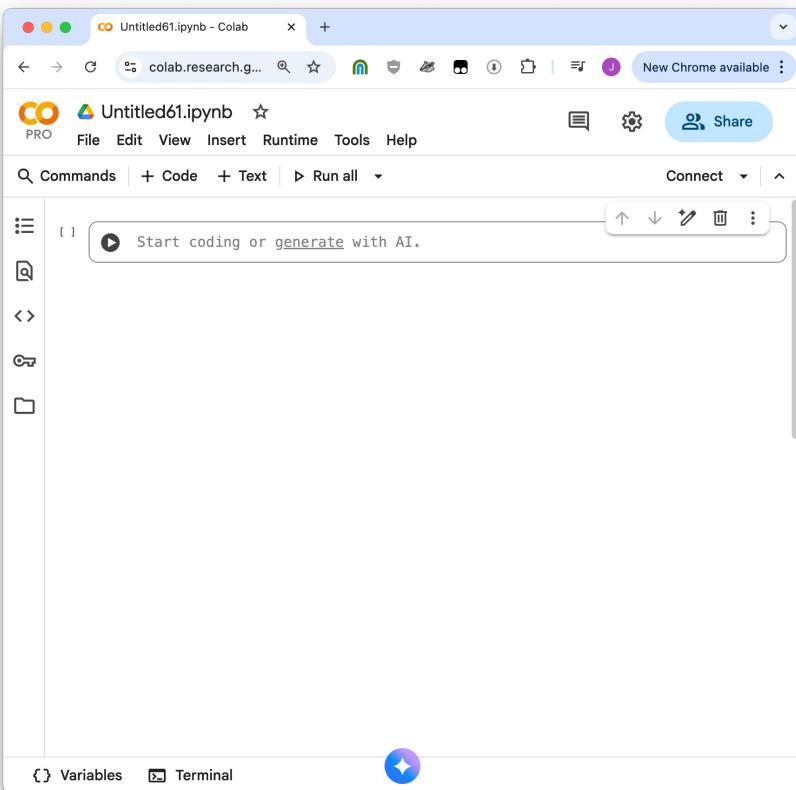
NAME KEY CREATED AT UPDATED AT ...

claude for sheets sk-ant-api03-1Sk...iAAA Apr 9, 2024 ...

sample class sk-ant-api03-g0n...rgAA Apr 10, 2024 ...

Create Key





API

data

API

# Why not just a CSV?

The screenshot shows the homepage of the Tokyo Open Data portal ([portal.data.metro.tokyo.lg.jp](http://portal.data.metro.tokyo.lg.jp)). The page has a dark blue header with the Tokyo Open Data logo and navigation links for searching data, APIs, and case studies. Below the header is a large white search bar with a placeholder "公開中のデータを検索する" and a "データを探す" button. The main content area features the "TOKYO OPEN DATA" logo and a sub-headline "東京都 オープンデータカタログサイト". It includes a call-to-action for a survey and two buttons for requesting data or seeing its applications. At the bottom are links for the community, hackathon, and roundtable.

portal.data.metro.tokyo.lg.jp

TOKYO OPEN DATA  
東京都 オープンデータカタログサイト

サイト内検索 | 文字サイズ: 小 中 大 都庁総合トップ

→ 一覧からデータを探す → APIを探す → データ活用事例を見る → データで見る東京

公開中のデータを検索する

オープデータは様々な用途で活用することができます。

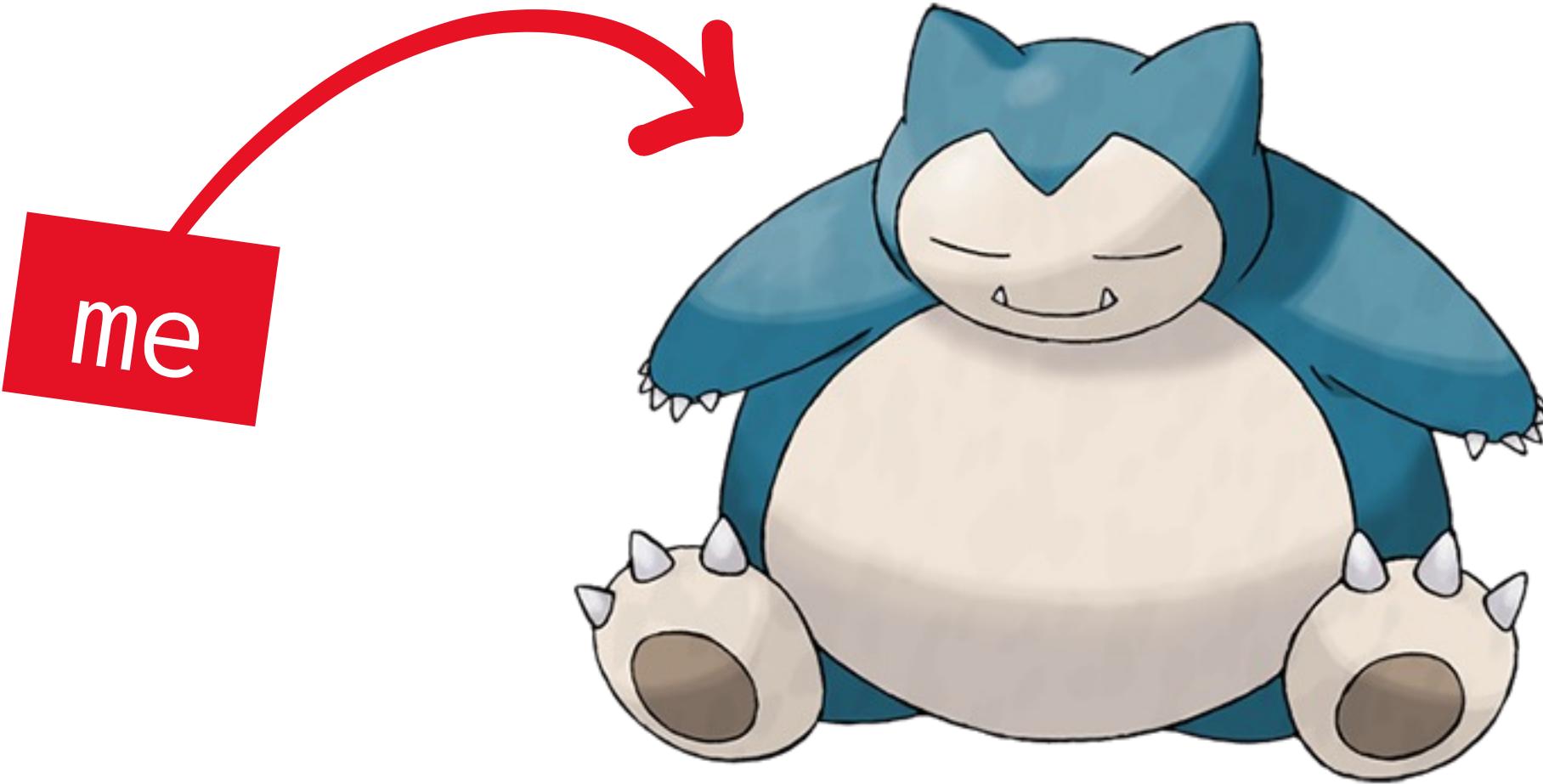
- 都や区市町村に関する様々な情報を調べる
- 都政に関するデータを確認する
- アプリやサービス開発に活用する

→ データをリクエストする → データを活用された方へ

東京都オープンデータコミュニティ 都知事杯オープンデータ・ハッカソン 東京都オープンデータ・ラウンドテーブル

# 2025

JANUARY							FEBRUARY							MARCH							APRIL							
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	
29	30	31	1	2	3		26	27	28	29	30	1		23	24	25	26	27	28	1	30	31	1	2	3	4	5	
5	6	7	8	9	10	11	2	3	4	5	6	8		2	3	4	5	6	7	8	6	7	8	9	10	11	12	
12	13	14	15	16	17	18	9	10	11	12	13	14	15	9	10	11	12	13	14	15	13	14	15	16	17	18	19	
19	20	21	22	23	24	25	16	17	18	19	20	21	22		23	24	25	26	27	28	29	20	21	22	23	24	25	26
26	27	28	29	30	31	1	23	24	25	26	27	28	1	23	24	25	26	27	28	29	27	28	29	30	1	2	3	
2	3	4	5	6	7	8	2	3	4	5	6	7	8	30	31	1	2	3	4	5		4	5	6	7	8	9	10
MAY							JUNE							JULY							AUGUST							
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	
27	28	29	30	1	2	3	1	2	3	4	5	6	7	29	30	1	2	3	4	5	27	28	29	30	31	1	2	
4	5	6	7	8	9	10	8	9	10	11	12	13	14	6	7	8	9	10	11	12	3	4	5	6	7	8	9	
11	12	13	14	15	16	17	15	16	17	18	19	20	21	13	14	15	16	17	18	19	10	11	12	13	14	15	16	
18	19	20	21	22	23	24	22	23	24	25	26	27	28	20	21	22	23	24	25	26	17	18	19	20	21	22	23	
25	26	27	28	29	30	31	29	30	1	2	3	4	5	29	30	1	2	3	4	5	1	1	2	3	4	5	6	
1	2	3	4	5	6	7																						
NOVEMBER							DECEMBER																					
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	
28	29	30	1	2	3	4	26	27	28	29	30	1	2	26	27	28	29	30	1	30	1	2	3	4	5	6		
5	6	7	8	9	10	11	5	6	7	8	9	10	11	2	3	4	5	6	7	8	7	8	9	10	11	12	13	
14	15	16	17	18	19	20	12	13	14	15	16	17	18	9	10	11	12	13	14	15	14	15	16	17	18	19	20	
21	22	23	24	25	26	27	19	20	21	22	23	24	25	16	17	18	19	20	21	22	21	22	23	24	25	26	27	
28	29	30	1	2	3	4	26	27	28	29	30	31	1	23	24	25	26	27	28	29	28	29	30	31	1	2	3	
5	6	7	8	9	10	11	2	3	4	5	6	7	8	30	1	2	3	4	5	6	4	5	6	7	8	9	10	



me

Google Search x +

google.com/search?q=snorlax&oq=snorlax&gs\_lcrp=EgZjaHJvbWUyDwgAEEUYORIRAhBABiKBTIMCAEQABhDGIAEGloFMg0IAhAAGJECGIAEGloFMgwIAxAA... ⋮

snorlax X Microphone Image Search

Google J

All Images Shopping Videos News Maps Web More Tools

Bulbapedia  
[https://bulbapedia.bulbagarden.net/wiki/Snorlax\\_\(Pok%C3%A9mon\)](https://bulbapedia.bulbagarden.net/wiki/Snorlax_(Pok%C3%A9mon)) ⋮

**Snorlax (Pokémon) - Bulbapedia**

Snorlax is a huge, bipedal, dark blue-green mammalian Pokémon with a cream-colored face, belly, and feet.

Ash's Snorlax III Menu Pokémon Sleep

People also ask :

What is Snorlax in Japanese? ⌄

Why is Snorlax sleepy? ⌄

What type of animal is a Snorlax? ⌄

What is Snorlax power? ⌄

Feedback

Pokemon.com  
<https://www.pokemon.com/pokedex/snorlax> ⋮

**Snorlax | Pokédex**

This gluttonous Pokémon eats constantly, apart from when it's asleep. It devours nearly 900 pounds of food per day.

**Snorlax :**



More images

Snorlax, known in Japan as Kabigon, is a Pokémon species in Nintendo and Game Freak's Pokémon franchise.

[Wikipedia](#)

**Type:** Normal

**Weakness:** Fighting

**Ability (hidden):** Gluttony

**Abilities:** Thick Fat, Immunity

**Evolves from:** Munchlax

**Category:** Sleeping

**Weight:** 460 kg

People also search for

Snorlax (Pokémon) - Bulbape... +

bulbapedia.bulbagarden.net/wiki/Snorlax\_(Pokémon)

BULBAGARDEN News Bulbapedia Forums Discord More Editors Account Search Bulbapedia

article discussion view source history tools in other languages

In commemoration of Nintendo's 135<sup>th</sup> anniversary, September 20 to 30 is NIWA Cross-Wiki Week. [Click here](#) for event details and the possibility to win a \$20/€20 Nintendo eShop gift card.  
Please remember to follow the [manual of style](#) and [code of conduct](#) at all times.  
Check the [Bulbagarden home page](#) for up-to-date Pokémon news and discuss it on the [forums](#) or in the [Bulbagarden Discord server](#).

## Snorlax (Pokémon)

For Pokémon GO information on this species, see [the game's section](#).

←  #0142: Aerodactyl Pokémon #0144: Articuno →

This article is about the species. For a specific instance of this species, see [Snorlax \(disambiguation\)](#).

**Snorlax** (Japanese: カビゴン *Kabigon*) is a Normal-type Pokémon introduced in [Generation I](#). It evolves from [Munchlax](#) when leveled up with high friendship. Snorlax can [Gigantamax](#) into [Gigantamax Snorlax](#) if it has the [Gigantamax Factor](#). In Generations I and II games, their remakes, and [Pokémon X and Y](#), wild Snorlax can be found sleeping in inconvenient locations. In order to catch one, a Trainer must first awaken it with music from either a [Poké Flute](#) or the [Pokégear](#) set to the Poké Flute channel. Snorlax is the game mascot of [Pokémon Sleep](#).

**Contents [hide]**

- 1 Biology
  - 1.1 Forms
  - 1.2 Evolution
- 2 Game data
  - 2.1 NPC appearances
  - 2.2 Pokédex entries
  - 2.3 Game locations
    - 2.3.1 In side games
    - 2.3.2 In events
      - 2.3.2.1 Wild Area News
  - 2.4 Held items
  - 2.5 Stats
    - 2.5.1 Base stats
    - 2.5.2 Pokéathlon stats
  - 2.6 Type effectiveness
  - 2.7 Learnset

**Snorlax** #0143  
Sleeping Pokémon Kabigon



Snorlax

**Gigantamax Snorlax**

Images on the Bulbagarden Archives

Snorlax (Pokémon) - Bulbape... X +

bulbapedia.bulbagarden.net/wiki/Snorlax\_(Pokémon)

2.5.1 Base Stats  
2.5.2 Pokéathlon stats  
2.6 Type effectiveness  
2.7 Learnset  
2.7.1 By leveling up  
2.7.2 By TM  
2.7.3 By breeding  
2.7.4 By a prior Evolution  
2.8 Side game data  
2.9 Form data  
2.9.1 Gigantamax  
2.10 Evolution data  
2.11 Sprites

3 In animation  
3.1 Main series  
3.1.1 Major appearances  
3.1.1.1 Ash's Snorlax  
3.1.1.2 Other  
3.1.2 Minor appearances  
3.1.3 Pokédex entries  
3.2 Pokémon Origins  
3.3 Pokémon Masters Animated Trailer  
3.4 GOTCHA!  
3.5 Pokémon Evolutions  
3.6 The Adventures of Snorlax & Cubone

4 In the manga  
4.1 Ash & Pikachu  
4.2 The Electric Tale of Pikachu  
4.3 Pokémon Adventures  
4.4 Pokémon Battle Frontier  
4.5 Pokémon: Yeah! I Got Pokémon!  
4.6 Pokémon Mystery Dungeon: Ginji's Rescue Team  
4.7 Pokémon Journeys  
4.8 Pokémon Pocket Monsters  
4.9 Pokémon Ruby-Sapphire  
4.10 Snorlax's Dream Gourmet

Gigantamax Snorlax

Images on the Bulbagarden Archives

Type	Normal						
Abilities	Gluttony Hidden Ability Immunity or Thick Fat						
Gender ratio	87.5% male, 12.5% female						
Catch rate	25 (7.7%)						
Breeding	Egg Group: Monster Hatch time: 40 cycles						
Height	6'11" (Snorlax) 114'10"+ (Gigantamax Snorlax)						
Weight	2.1 m (Snorlax) 35.0+ m (Gigantamax Snorlax) 1014.1 lbs. (Snorlax) ??? lbs. (Gigantamax Snorlax) 460.0 kg (Snorlax) ??? kg (Gigantamax Snorlax)						
Base experience yield	154 (Gen. I-IV) 189 (V+)						
Leveling rate	Slow						
EV yield	Total: 2 <table border="1"><tbody><tr><td>2 (HP)</td><td>0 (Atk)</td><td>0 (Def)</td><td>0 (Sp.Atk)</td><td>0 (Sp.Def)</td><td>0 (Speed)</td></tr></tbody></table>	2 (HP)	0 (Atk)	0 (Def)	0 (Sp.Atk)	0 (Sp.Def)	0 (Speed)
2 (HP)	0 (Atk)	0 (Def)	0 (Sp.Atk)	0 (Sp.Def)	0 (Speed)		
Shape							
Footprint							
Pokédex color	Black						
Base friendship	70						

# Structured vs unstructured data

- Easy for **computers** to understand and process
- CSV files + XML, JSON, Excel, etc etc etc
- Easy for **humans** to understand and process
- Websites, books, articles, etc etc etc

# APIs

## Application Programming Interface

# APIS

The image shows a screenshot of the PokeAPI website. At the top, there's a pink header bar with the word "APIS" in white. Below it is a red navigation bar with links for "Home", "About", "API v2", and "GraphQL v1beta". The main content area has a dark background. On the left, a red banner contains the text "Structured data" in white. A large red curved arrow points from this banner towards the bottom right. In the center, the PokeAPI logo is displayed with the text "The RESTful Pokémon API" and "Serving over 2.5 billion API calls each month!". To the right, a yellow banner displays the URL "https://pokeapi.co". At the bottom, there's a light green section with the text "All the Pokémon data you'll ever need in one place, easily accessible through a modern free open-source RESTful API." followed by a blue link "Check out the docs!". The bottom-most part of the page features a white search bar with the URL "https://pokeapi.co/api/v2/", the search term "pokemon/ditto", a magnifying glass icon, and a blue "Submit" button.

https://pokeapi.co/api/v2/

pokemon/ditto

Submit

earthquake.usgs.gov/fdsnws/ x +

earthquake.usgs.gov/fdsnws/event/1/query?format=geojson&starttime=2014-01-01&endtime=2014-01-02

Pretty-print

```
{  
  "type": "FeatureCollection",  
  "metadata": {  
    "generated": 1727127591000,  
    "url": "https://earthquake.usgs.gov/fdsnws/event/1/query?format=geojson&starttime=2014-01-01&endtime=2014-01-02",  
    "title": "USGS Earthquakes",  
    "status": 200,  
    "api": "1.14.1",  
    "count": 325  
  },  
  "features": [  
    {  
      "type": "Feature",  
      "properties": {  
        "mag": 1.29,  
        "place": "10km SSW of Idyllwild, CA",  
        "time": 1388620296020,  
        "updated": 1457728844428,  
        "tz": null,  
        "url": "https://earthquake.usgs.gov/earthquakes/eventpage/ci11408890",  
        "detail": "https://earthquake.usgs.gov/fdsnws/event/1/query?eventid=ci11408890&format=geojson",  
        "felt": null,  
        "cdi": null,  
        "mmi": null,  
        "alert": null,  
        "status": "reviewed",  
        "tsunami": 0,  
        "sig": 26,  
        "net": "ci",  
        "code": "11408890",  
        "ids": ",ci11408890,",  
        "sources": ",ci,",  
        "types": ",cap,focal-mechanism,nearby-cities,origin,phase-data,scitech-link,",  
        "nst": 39,  
        "dmin": 0.06729,  
        "rms": 0.09,  
        "gap": 51,  
        "magType": "ml",  
        "type": "earthquake",  
        "title": "M 1.3 - 10km SSW of Idyllwild, CA"  
      },  
      "geometry": {  
        "type": "Point",  
        "coordinates": [-116.7776667, 33.6633333, 11.008]  
      },  
      "id": "ci11408890"  
    },  
    {  
      "type": "Feature",  
      "properties": {  
        "mag": 1.1,  
        "place": "Central Alaska",  
        "time": 1388620046501,  
        "updated": 1689800179099,  
        "tz": null,  
        "url": "https://earthquake.usgs.gov/earthquakes/eventpage/ak01421ig3u",  
        "detail": "https://earthquake.usgs.gov/fdsnws/event/1/query?eventid=ak01421ig3u&format=geojson"  
      }  
    }  
  ]  
}
```

API Documentation - Earthquake Catalog

earthquake.usgs.gov/fdsnws/event/1/



Earthquake Hazards Program

← Feeds and Notifications

Real-time Notifications

Earthquake Notification Service

Tweet Earthquake Dispatch

Real-time Feeds

ATOM

KML

Spreadsheet

QuakeML

GeoJSON Summary

GeoJSON Detail

For Developers

API Documentation - EQ Catalog

Change Log

Feed Lifecycle Policy

Developer's Corner

# API Documentation - Earthquake Catalog

This is an implementation of the [FDSN Event Web Service Specification](#), and allows custom searches for earthquake information using a variety of parameters.

**i** Please note that automated applications should use [Real-time GeoJSON Feeds](#) for displaying earthquake information whenever possible, as they will have the best performance and availability for that type of information.

**URL**

<https://earthquake.usgs.gov/fdsnws/event/1/>

request known enumerated parameter values for the interface.

- <https://earthquake.usgs.gov/fdsnws/event/1/application.json>

**application.wadl**

request WADL for the interface.

- <https://earthquake.usgs.gov/fdsnws/event/1/application.wadl>

**catalogs**

request available catalogs.

- <https://earthquake.usgs.gov/fdsnws/event/1/catalogs>

**contributors**

Let's try this out

- Open up the **Earthquakes** notebook.
- First, I'll work on it myself

## Your mission

- Find the URL for 2.5+ earthquakes for the past 30 days
- Give the URL to Gemini/Colab and **put them on a map**
- With a circle size by how strong the earthquakes were

## Your mission

- Visit [weatherapi.com](https://weatherapi.com)
- Make an account using “Sign up”
- Copy your API key
- (we'll use it in the next step)

# Your mission

- Click the “API Explorer” link
- Enter your API key
- Select “Forecast”
- Look for Tokyo, three day forecast
- Use the URL with Colab to graph the hourly temperatures over the next three days, with daily high/low highlighted

Visit <https://bit.ly/ds-dojo-2025> for materials

# Scraping!

stealing data from the  
internet for fun

The New York Times

U.S. INTERNATIONAL CANADA ESPAÑOL 中文

GIVE THE TIMES Account Nasdaq -0.47% ↓

ober 16, 2025

U.S. World Business Arts Lifestyle Opinion Audio Games Cooking Wirecutter The Athletic

LIVE N.Y.C. Mayoral Debate 37m ago Trump Administration 59m ago

John Bolton Is Indicted Over  
ing of Classified  
ation

John, a Trump aide turned critic, is part  
of presidential foes to become  
rial targets. But his case gained  
um in the Biden administration.

updates >

Indictment

ia-Ukraine War Photos Russian Provocations in Europe Ukraine's Weapons The Changing Business of War

ump Plans to Meet With Putin to  
About a Cease-Fire in Ukraine

ident Trump had grown increasingly  
strated with President Vladimir Putin for  
gg on Russia's three-year war on  
aine.

IN READ

Russia Is Arming Drones With North  
European Cluster Weapons, Report Says

MIN READ

Standing Up for the Beauty of a Good Bloody Mary

It's scorned for its association with brunch and bartenders often are reluctant to mix one. But a good Bloody Mary is great, our critic writes.

5 MIN READ

What Do You Give Someone Who's Just Gotten Engaged?

He's an N.B.A. Superstar. But Even His Coach Has 'No Clue' How Tall He Is.

FROM THE ATHLETIC

5 MIN READ

Book1 - Excel

File Home Insert Page Layout Formulas Data Review

Clipboard

Font

Alignment

A1

A B C D E F G

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

Sheet1

Ready

A large red arrow points from the New York Times homepage towards the top-left cell (A1) of the Excel spreadsheet.

Visit <https://bit.ly/ds-dojo-2025> for materials

# HTML

the language that all web  
sites are built with

```
<h1>This is a headline</h1>
<h2>This is a smaller headline</h2>
<h3>And an even smaller headline</h3>
```

# **This is a headline**

## **This is a smaller headline**

### **And an even smaller headline**

```
<h1>This is a headline</h1>
<h2>This is a smaller headline</h2>
<h3>And an even smaller headline</h3>
<p>This is a paragraph</p>

<p>This is a paragraph with
<a href="google.com">a link</a></p>
<div>This is ANYTHING</div>
<div>Anything! Anything in the world</div>
```

# This is a headline

## This is a smaller headline

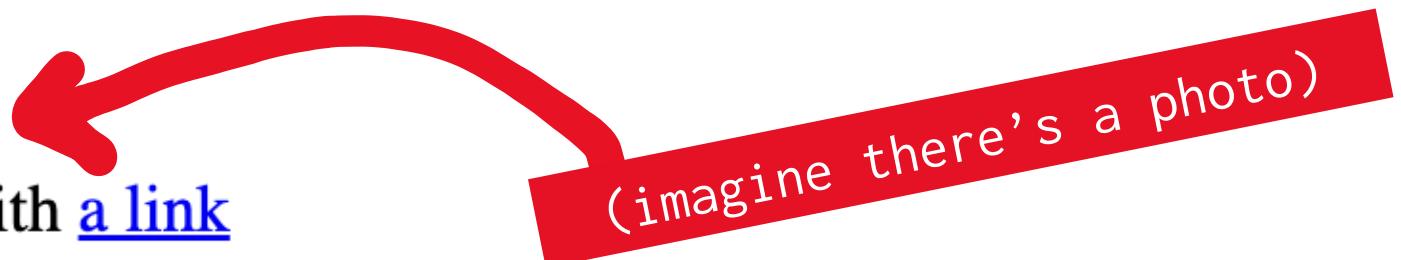
### And an even smaller headline

This is a paragraph

This is a paragraph with [a link](#)

This is ANYTHING

Anything! Anything in the world



(imagine there's a photo)

<h1>An incredible story</h1>

<p>By J. Soma</p>

<p>This is the start of the story.</p>

<p>It is an amazing story!</p>

<p>”It’s incredible,” said the source.</p>

<p>An editor agreed: “it’s true!”</p>

<p>Additional reporting by Mulberry the fat  
mean cat</p>

<p>Edit: An earlier version said “Jonathan”  
Soma</p>

# An incredible story

By J. Soma

This is the start of the story.

It is an amazing story!

"It's incredible," said the source.

An editor agreed: "it's true!"

Additional reporting by Mulberry the fat mean cat

Edit: An earlier version said "Jonathan" Soma

<h1>An incredible story</h1>

<p>By J. Soma</p>

<p>This is the start of the story.</p>

<p>It is an amazing story!</p>

<p>”It’s incredible,” said the source.</p>

<p>An editor agreed: “it’s true!”</p>

<p>Additional reporting by Mulberry the fat  
mean cat</p>

<p>Edit: An earlier version said “Jonathan”  
Soma</p>

```
<h1>An incredible story</h1>
<p id="byline">By J. Soma</p>
<p>This is the start of the story.</p>
<p>It is an amazing story!</p>
<p>"It's incredible," said the source.</p>
<p>An editor agreed: "it's true!"</p>
<p class="additional">Additional reporting by
Mulberry the fat mean cat</p>
<p class="correction">Edit: An earlier version
said "Jonathan" Soma</p>
```

# An incredible story

By J. Soma

This is the start of the story.

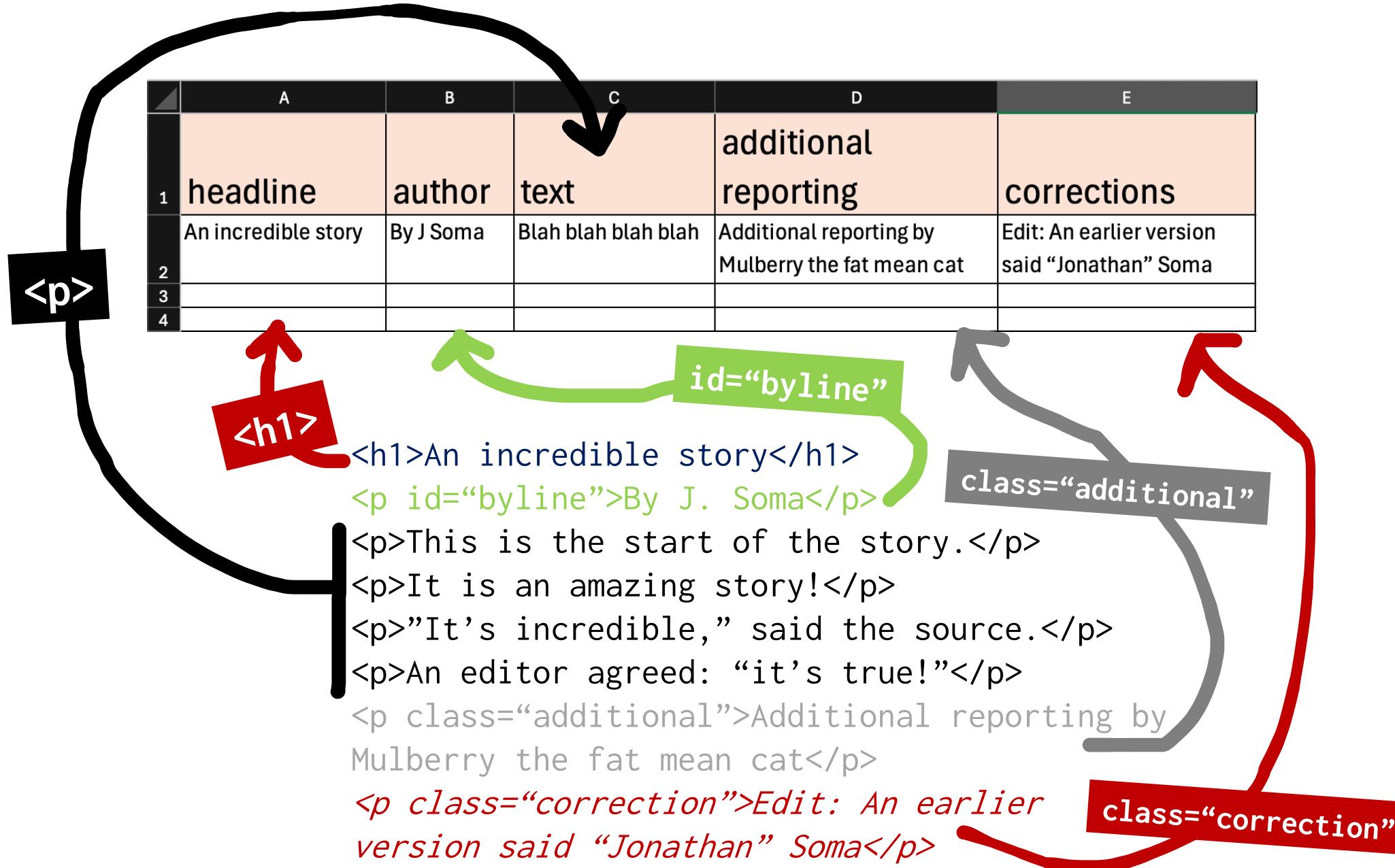
It is an amazing story!

”It’s incredible,” said the source.

An editor agreed: “it’s true!”

Additional reporting by  
Mulberry the fat mean cat

*Edit: An earlier version said “Jonathan”  
Soma*



Let's see it on  
the internet!

BBC Home - Breaking News, +

bbc.com

Home News Sport Business Innovation Culture Travel Earth Video Live

Register Sign In

# What will be a row in our spreadsheet?

 **LIVE VP pick Walz to speak as Democrats deploy star guests**

The Minnesota governor will run alongside presidential nominee Kamala Harris for November's US election.

 **Three things the Democrats have avoided so far at the DNC**

What they have tried to avoid says as much about their weaknesses as what they choose to highlight, writes Anthony Zurcher.

2 hrs ago | US & Canada

- Day three of Democratic Convention features party's rising stars
- Obamas mock Trump over 'black jobs' and crowd sizes

 **Divers find five bodies in wreck of Sicily yacht**

 **Gaza nurse says whole family, including quadruplets, killed in air strike**

5 hrs ago | Middle East & Africa

 **Andrew Tate held overnight after police raid homes in Romania**

The internet personality and his brother have been remanded in custody as part of a probe into new allegations.

44 mins ago | Europe

 **Ancient ocean of magma found on Moon south pole**

The findings are from India's historic Chandrayaan-3 mission that landed on the Moon's south pole.

6 hrs ago | Science & Environment

► **German Navy blasts out Darth Vader theme on Thames**

BBC Home - Breaking News, +

bbc.com

Home News Sport Business Innovation Culture Travel Earth Video Live

Register Sign In



**LIVE VP pick Walz to speak as Democrats deploy star guests**

The Minnesota governor will run alongside presidential nominee Kamala Harris for November's US election.



**Three things the Democrats have avoided so far at the DNC**

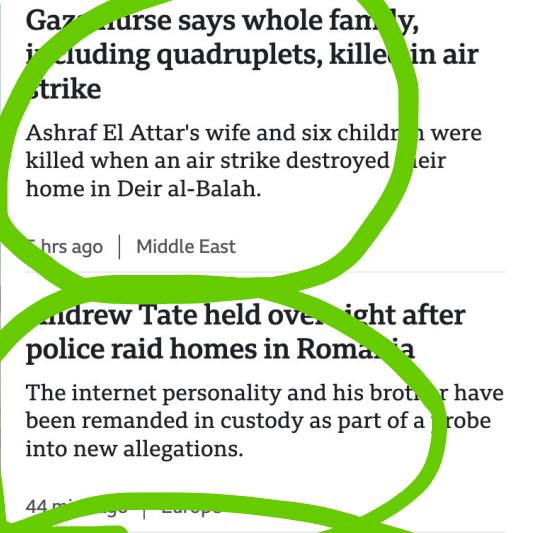
What they have tried to avoid says as much about their weaknesses as what they choose to highlight, writes Anthony Zurcher.

2 hrs ago | US & Canada

- Day three of Democratic Convention features party's rising stars
- Obamas mock Trump over 'black jobs' and crowd sizes



**Divers find five bodies in wreck of Saily yacht**



**Gaza nurse says whole family, including quadruplets, killed in air strike**

Ashraf El Attar's wife and six children were killed when an air strike destroyed their home in Deir al-Balah.

5 hrs ago | Middle East



**Andrew Tate held overnight after police raid homes in Romania**

The internet personality and his brother have been remanded in custody as part of a probe into new allegations.

44 mins ago | Europe



**Ancient ocean of magma found on Moon south pole**

The findings are from India's historic Chandrayaan-3 mission that landed on the Moon's south pole.

6 hrs ago | Science & Environment

► German Navy blasts out Darth Vader theme on Thames

What will be a column of data?

## Andrew Tate held overnight after police raid homes in Romania

The internet personality and his brother have been remanded in custody as part of a probe into new allegations.

44 mins ago

| Europe



## [L] **LIVE** VP pick Walz to speak as Democrats deploy star guests

[R] The Minnesota governor will run alongside presidential nominee Kamala Harris for November's US election.

Data sets - Open data

data.gov.rs/sr/datasets/

# Data sets

2435 sets in total

Download data→

What will be a row in our spreadsheet?

Schedule according to... Organizations ...

 **CATALOG OF CONTAMINATED LOCATIONS**

The cadastre of contaminated sites represents a set of data on endangered, polluted and degraded lands and it is an integral part of the Land Information System managed by the Environmental Protection Agency. State bodies, i.e. organizations, bodies of autonomous provinces, local self-government units...

 **Soil condition in urban areas**

The data show the degree of soil endangerment from chemical pollution in urban areas in the period 2005-2023. year. Soil testing was carried out at locations in the immediate vicinity of traffic roads, industrial zones, near landfills, water supply sources, educational institutions, recreation...

 **Realization regular program**

A set of data on the performed services of the Animal Hygiene Program.

resources 6  
use 0  
supporting 0

<https://data.gov.rs/sr/datasets/stanje-zemljshta-u-urbanim-sredinama/> training providers

resources 1  
use 0

Data sets - Open data

data.gov.rs/sr/datasets/

# Data sets

2435 sets in total [Examples of use→](#) [Download data→](#)

Data search...

Schedule according to... Organizations Labels Formats ...

 **CATALOG OF CONTAMINATED LOCATIONS**  
The cadastre of contaminated sites represents a set of data on endangered, polluted and degraded lands and it is an integral part of the Land Information System managed by the Environmental Protection Agency. State bodies, i.e. organizations, bodies of autonomous provinces, local self-government units...  
use 0 supporting 0

 **Soil condition in urban areas**  
The data show the degree of soil endangerment from chemical pollution in urban areas in the period 2005-2023. year. Soil testing was carried out at locations in the immediate vicinity of traffic roads, industrial zones, near landfills, water supply sources, educational institutions, recreation...  
resources 22 use 0 supporting 0

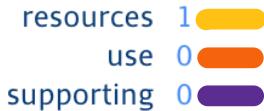
 **Realization regular program**  
A set of data on the performed services of the Animal Hygiene Program.  
resources 6 use 0 supporting 0

<https://data.gov.rs/sr/datasets/stanje-zemljišta-u-urbanim-sredinama/> training providers  
resources 1 use 0



## CATALOG OF CONTAMINATED LOCATIONS

The cadastre of contaminated sites represents a set of data on endangered, polluted and degraded lands and it is an integral part of the Land Information System managed by the Environmental Protection Agency. State bodies, i.e. organizations, bodies of autonomous provinces, local self-government units...



What will be a column of data?

Visit <https://bit.ly/birn-data>, click **scraping**

Computers don't see like us  
(usually), we need to see

HTML code

# Right click (or command+click) and Inspect

Data.gov.rs

Претрага података...

Распоред према... Организације Ознаке Формати ...

ресурса 1  
употреба 0  
подржавања 0

ресурса 22  
употреба 0  
подржавања 0

ресурса 6  
употреба 0  
подржавања 0

ресурса 1  
употреба 0  
подржавања 0

Look Up "КОНТАМИНИРАНИХ"  
Copy  
Copy Link to Highlight  
Search Google for "КОНТАМИНИРАНИХ"  
Print...  
Open in Reading Mode  
Translate Selection to English  
Block element...  
**Inspect**  
Speech  
Services

Катастар контаминираних ло  
земљиштима и он је саставни  
животне средине. Државни о  
локалне самоуправе...

Стање земљишта у урбани  
Подаци показују степен угрож  
2005-2023. године. Испитива  
саобраћајница, индустријских  
установа, рекреац...

Редовни програм Реализација  
Скуп података о извршеним услугама Програма делатности зоохигијене.

Евиденција акредитованих спроводилаца обука НАЈУ  
Евиденција акредитованих спроводилаца обука НАЈУ За потребе акредитације спроводилаца програма  
обуке, Национална академија, по правилу, два пута годишње објављује јавни позив за поједине  
области стручног усавршавања. Јавни позив објављује се на интернет страници Националне академије,  
министарства и

# Find the HTML for our “row”

Претрага



## КАТАСТАР КОНТАМИНИРАНИХ ЛОКАЦИЈА

Катастар контаминираних локација представља скуп података о угроженим, загађеним и деградираним земљиштима и он је саставни део Информационог система земљишта који води Агенција за заштиту животне средине. Државни органи, односно организације, органи аутономних покрајина, јединице локалне самоуправе...

ресурса 1  
употреба 0  
подржавања 0



## Стање земљишта у урбаним срединама

Подаци показују степен угрожености земљишта од хемијског загађења у урбаним срединама у периоду 2005-2023. године. Испитивање земљишта вршено је на локацијама у непосредној близини прометних саобраћајница, индустријских близини депонија, изворишта водоснабдевања, педагошких установа, рекреац



Mouse/click  
around

```

Elements Console Sources Network Performance > × 7 5 | ⚙ : X
preview_url: "/catalog/preview/?url=https://opendata.stat.gov.rs/datasets/dolasti-turista-godishnji-podatsi", "published": "2024-03-20T15:56:43.643000", "schema": {}, "title": "\u0414\u043e\u043b\u0430\u0441\u0446\u0438 \u0442\u0443\u0440\u044f\u043f\u043e\u0434\u043d\u0430\u0446\u0438", "type": "api", "url": "https://opendata.stat.gov.rs/data/WcfJsonRestService.Service1.svc/dataset/2201IND01/1/csv"}], "slug": "dolasti-turista-godishnji-podatsi", "spatial": null, "tags": ["statistika", "ugostiteljstvo-i-turizam"], "temporal_coverage": null, "title": "\u0414\u043e\u043b\u0430\u0441\u0446\u0438 \u0442\u0443\u0440\u044f\u043f\u043e\u0434\u043d\u0430\u0446\u0438", "uri": "https://data.gov.rs/api/1/datasets/dolasti-turista-godishnji-podatsi/"}] data-v-2e2e7c>
  <ul data-v-2e2e7c class=>
    <li data-v-2e2e7c>
      <a class="unstyled w-10 block" data-v-2e2e7c>
        <article class="dataset-card height-auto dataset-search-result py-xs" badges="object Object">
          created_at="2024-05-29T13:28:17.964000" extras=[object Object]
          frequency_date="2024-05-29T00:00:00" id="66571151a561adbb2191d89e"
          last_modified="2024-06-06T08:22:33.848000" last_update="2024-05-29T00:00:00"
          license="public_domain" page="https://data.gov.rs/sr/datasets/katastar-kontaminiranikh-lokatsija/" slug="katastar-kontaminiranikh-lokatsija" tags uri="https://data.gov.rs/api/1/datasets/katastar-kontaminiranikh-lokatsija/" data-v-c9e660 data-v-2e2e7c>
          ><div class="card-logo search-logo-card" data-v-c9e660>...</div> flex
          <div class="card-data search-logo-card" data-v-c9e660>
            <h4 class="card-title" data-v-c9e660>КАТАСТАР КОНТАМИНИРАНИХ ЛОКАЦИЈА</h4>
            <div class="card-description mt-xs" data-v-c9e660>...</div> == $0
          </div>
          <ul class="card-hover-data" data-v-c9e660>...</ul> flex
        </article>
      </a>
    </li>
  </ul>

```

METHOD ONE

rhythm.less:66

Скупови података - Отворе +

# Find the HTML for our “row”

Click this

Search + click here

Elements Console Sources Network Performance > X 7 5 J : X

```
preview_url: "/tabular/preview/?url=https://opendata.stat.gov.rs/datasets/dolastsi-turista-godishnji-podatsi", "published": "2024-03-20T15:56:43.643000", "schema": {}, "title": "\u0414\u043e\u043b\u0430\u0441\u0446\u0438 \u0442\u0443\u0440\u0438\u0441\u0442\u0430 - \u0433\u043e\u0434\u0438\u0448\u045a\u0438 \u043f\u043e\u0434\u0430\u0446\u0438", "type": "api", "url": "https://opendata.stat.gov.rs/WcfJsonRestService.Service1.svc/dataset/2201IND01/1/csv"}], "slug": "dolastsi-turista-godishnji-podatsi", "spatial": null, "tags": ["statistika", "ugostiteljstvo-i-turizam"], "temporal_coverage": null, "title": "\u0414\u043e\u043b\u0430\u0441\u0446\u0438 \u0442\u0443\u0440\u0438\u0441\u0442\u0430 - \u0433\u043e\u0434\u0438\u0448\u045a\u0438 \u043f\u043e\u0434\u0430\u0446\u0438", "uri": "https://data.gov.rs/api/1/datasets/dolastsi-turista-godishnji-podatsi/"}}] data-v-2e2e7c>
```

▼<ul data-v-2e2e7c class=>

▼<li data-v-2e2e7c>

▼<a class="unstyled w-10 block" data-v-2e2e7c>

▼<article class="dataset-card height-auto dataset-search-result py-xs" badges created\_at="2024-05-29T13:28:17.964000" extras=[object Object]" frequency\_date="2024-05-29T00:00:00" id="66571151a561adbb2191d89e" last\_modified="2024-06-06T08:22:33.848000" last\_update="2024-05-29T00:00:00" license="public\_domain" page="https://data.gov.rs/sr/datasets/katastar-kontaminiranih-lokatsija/" slug="katastar-kontaminiranih-lokatsija" tags uri="https://data.gov.rs/api/1/datasets/katastar-kontaminiranih-lokatsija/" data-v-c9e660>

...</a>

class="card-logo search-logo-card" data-v-c9e660>...</div> flex

class="card-data search-logo-card" data-v-c9e660>

class="card-title" data-v-c9e660>КАТАСТАР КОНТАМИНИРАНИХ ЛОКАЦИЈА</h4>

div class="card-description mt-xs" data-v-c9e660> == \$0

"Катастар контаминираних локација представља скуп података о угроженим, загађеним и деградираним земљиштима и он је саставни део Информационог система земљишта који води Агенција за заштиту животне средине. Државни органи, односно организације, органи аутономних покрајина, јединице локалне самоуправе...

ресурса 1  
употреба 0  
подржавања 0

Стање земљишта у урбаним срединама

Подаци показују степен угрожености земљишта од хемијског загађења у урбаним срединама у периоду 2005-2023. године. Испитивање земљишта вршено је на локацијама у непосредној близини прометних саобраћајница, индустријских близини депонија, изворишта водоснабдевања, педагошких установа, рекреац

## METHOD TWO

amazon.com  
search results

Sales at Lidl

List of amphibians  
from Germany's Red  
List Center

Headlines from  
Nikkei Asia

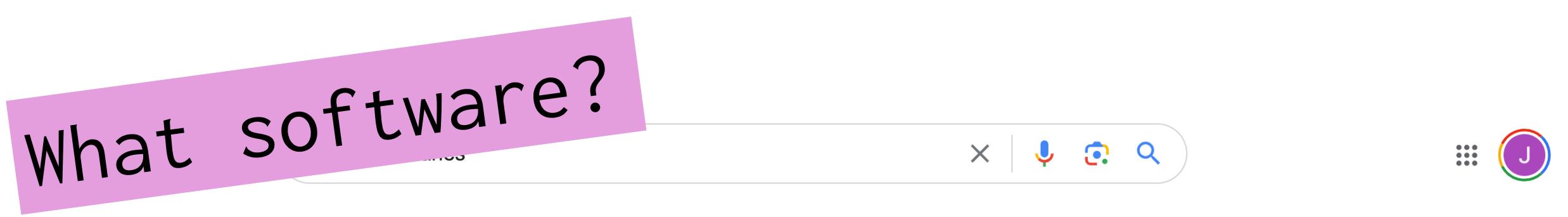
Now try it.  
Anywhere!!!!

School Board Minutes  
from Grand Island Public  
Schools in Nebraska

datasets on  
data.gov.ro

Wikipedia's  
Fictional Big  
Cats

Scraping is just  
connecting to some software  
to do that for us!  
in your spreadsheet



All Images Videos News Books Web Finance

Tools



## Scraping libraries

From sources across the web



Selenium



Lxml



Playwright



ZenRows S.L.



Beautiful Soup



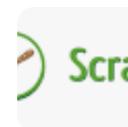
Requests



Puppeteer



Cheerio



Scrapy



MechanicalSoup



Urllib3



[\[–\]](#) **Swingbiter**  **70 points** 2 years ago

## Learn the basic html elements that build up a website.

[\[-\] \*\*coventous\*\*](#)  **22 points** 2 years ago

I recommend checking

W 1 point 2 years ago

[–] Nerdvana

Look into learning and understanding Docker, so Google for that. On a bonus tip, there's (official?) Chrome/Chrome/etc) which made me very happy (I like Docker!)

For more information about the study, please contact Dr. John Smith at (555) 123-4567 or via email at [john.smith@researchinstitute.org](mailto:john.smith@researchinstitute.org).

from - data - response

**[-] luizv4z**  **12 points** 2 years ago

From my own research

My own research, run away from Selenium. The right direction is CDP (Chrome Developers Protocol), I could scrap Facebook without getting banned.

```
2 all_links = soup.findAll(name=
```

~~Do python on them until~~

**Bea**  [-] riisen  3 points

for small projects use built-in tools, its am

Beginner projects go with scrapy, its an

for bigger projects  
add save save-RES repo

[permalink](#) [source](#) [embed](#) [save](#)

F.S. permit, unless

[permalink](#) [source](#) [embed](#) [save](#) [save-RES](#) [report abuse](#)

[–] ned334  5 points 2 years ago

Google "Selenium find\_element(By.XPATH, '/XPATH/')

All elements have an XPath that you can copy from chrome by Inspect -> right click on code block -> copy full Xpath.

# Scraping solved

[permalink](#) [source](#) [embed](#) [save](#) [save-RES](#) [report](#) [reply](#)

[permalink](#) [source](#) [embed](#) [save](#) [save-RES](#) [report](#) [reply](#) [hide child comments](#)

Selenium is garbage!

...but BeautifulSoup  
can't scrape all sites.

## Top discounts for you

Page 1 of 3



Continental Grand Prix  
5000 Rennrad Faltreifen //  
28-622 (700x28C)

85

400+ viewed in past month

\$66.00

FREE Delivery Friday,  
Aug 23



Wink Super X-Large Ultra  
Thin Lubricated Latex  
Condoms, Premium Latex  
for Smooth and Natural...

76

100+ viewed in past month

\$14.99 (\$0.30/Count)

FREE Delivery Friday,  
Aug 23



Assortment Rubber Ducks  
in Bulk, 50-Pack Assorted  
Mini Duckies Toy for  
Ducking Cruise Ships, 2"

328

5K+ viewed in past month

\$24.99

Overnight by 8:00  
AM



50 Pack Rubber Ducks in  
Bulk, Jeep Ducks for  
Ducking, Assorted Rubber  
Ducks Jeep Ducking, Bab...

569

4K+ viewed in past month

\$21.89

Overnight by 8:00  
AM



ArtCreativity Assorted  
Rubber Ducks Jeep Ducking  
(100Pack) - Rubber Ducki...

982

700+ viewed in past month

**20% off** **Limited time deal**

\$39.98

List: \$49.99

Today by 10:00 PM



Kicko Assorted Rubber  
Ducks with Mesh Bag - 50  
Ducklings, 2 Inch – Jeep  
Ducks for Kids, Baby Bath...

2,543

2K+ viewed in past month

\$28.99

FREE Delivery  
Saturday, Aug 24



Glitter Rubber Ducks in  
Bulk - (Pack of 50) Assorted  
2-inch Duck Toys for Baby  
Shower Rubber Duckies,...

308

1K+ viewed in past month

\$26.79

Today by 10:00 PM



## Bulk savings to consider

Page 1 of 7



Chochkees Asso

Ducks Toy Duckies for Kids  
and Toddlers, Bath Birthday

Baby Showers Classroom,...

1,155



Rubber Ducks Jeep Ducking  
(50 Pack) - Rubber Duckie...

100+ viewed in past month

982



Set, Mini Colorful Rubber  
Duckies Bath Toy for  
Child,Float & Squeak Tiny...

763



Bulk, Jeep Ducks for  
Ducking, Assorted Rubber  
Ducks Jeep Ducking, Bab...

569



Ducks: Fun Unique Military-  
Inspired Bath Toys for Jeep  
Ducking or Play - 2 inches

800+ viewed in past month

569



Bulk, Jeep Ducks for  
Ducking, Assorted Rubber  
Ducks for Jeeps, Bath Toy...

569



Glitter Rubber Duck Toy  
Assortment Duckies for  
Kids, Bath Birthday Gifts...

1,094

# Some sites need interaction





## TEXAS DEPARTMENT OF LICENSING &amp; REGULATION

## TDLR License Data Search (Active Licenses only)

[Search Help](#) | [Download License files](#) | [Download Other](#) | [Questions/Comments](#)

Inquire by License Type	Inquire by License #
Choose One (Optional)	<input type="text"/> (Numeric only)
Inquire by Expiration Date	
<input type="text"/> (mmddyyyy)	
Inquire by Name (Last, First) or by Business Name	
<input type="text"/>	
Inquire by Location (City)	
Choose One (Optional)	Type the first letter to scroll down.
Inquire by County	
Choose One (Optional)	Type the first letter to scroll down.
Inquire by Zip Code	
<input type="text"/>	
<input type="button" value="Search"/> <input type="button" value="Reset"/>	

If license not found, please contact Customer Service at 800-803-9202

Data last updated: 8/21/2024 06:01

[Bookmark This Page](#)

Some sites need interaction

Fast and reliable end-to-end 🎉

playwright.dev/python/

Playwright for Python Docs API Python ▾ Community

GitHub Discord Slack Search

# Playwright enables reliable end-to-end testing for modern web apps.

GET STARTED ⚡ Star 64k+

I love Playwright!



Playwright is perfect!

*But!* It's new, so

ChatGPT isn't perfect

at it. *But!* We try.

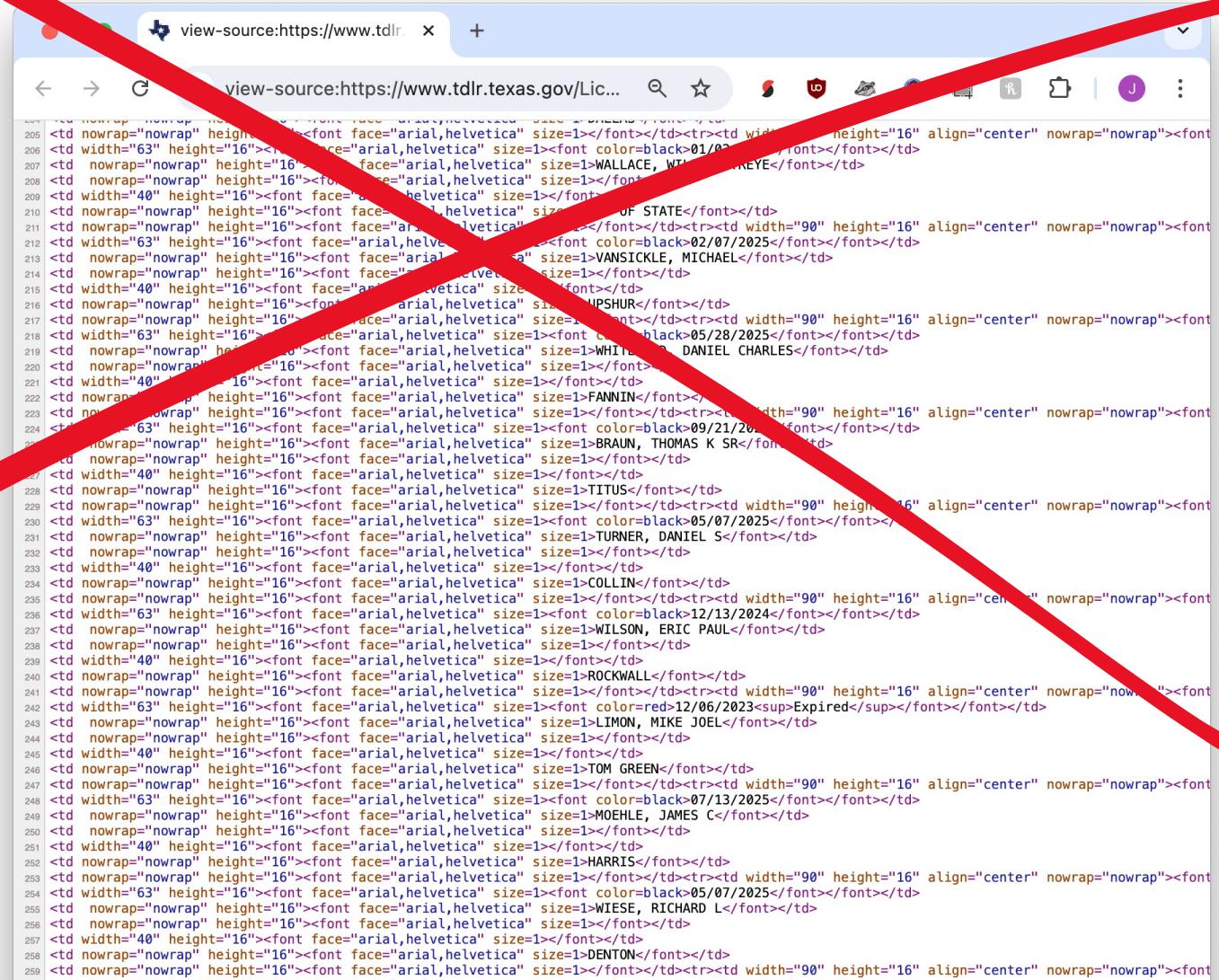
Playwright + ChatGPT +  
pasting samples of HTML

=

infinite scrapers

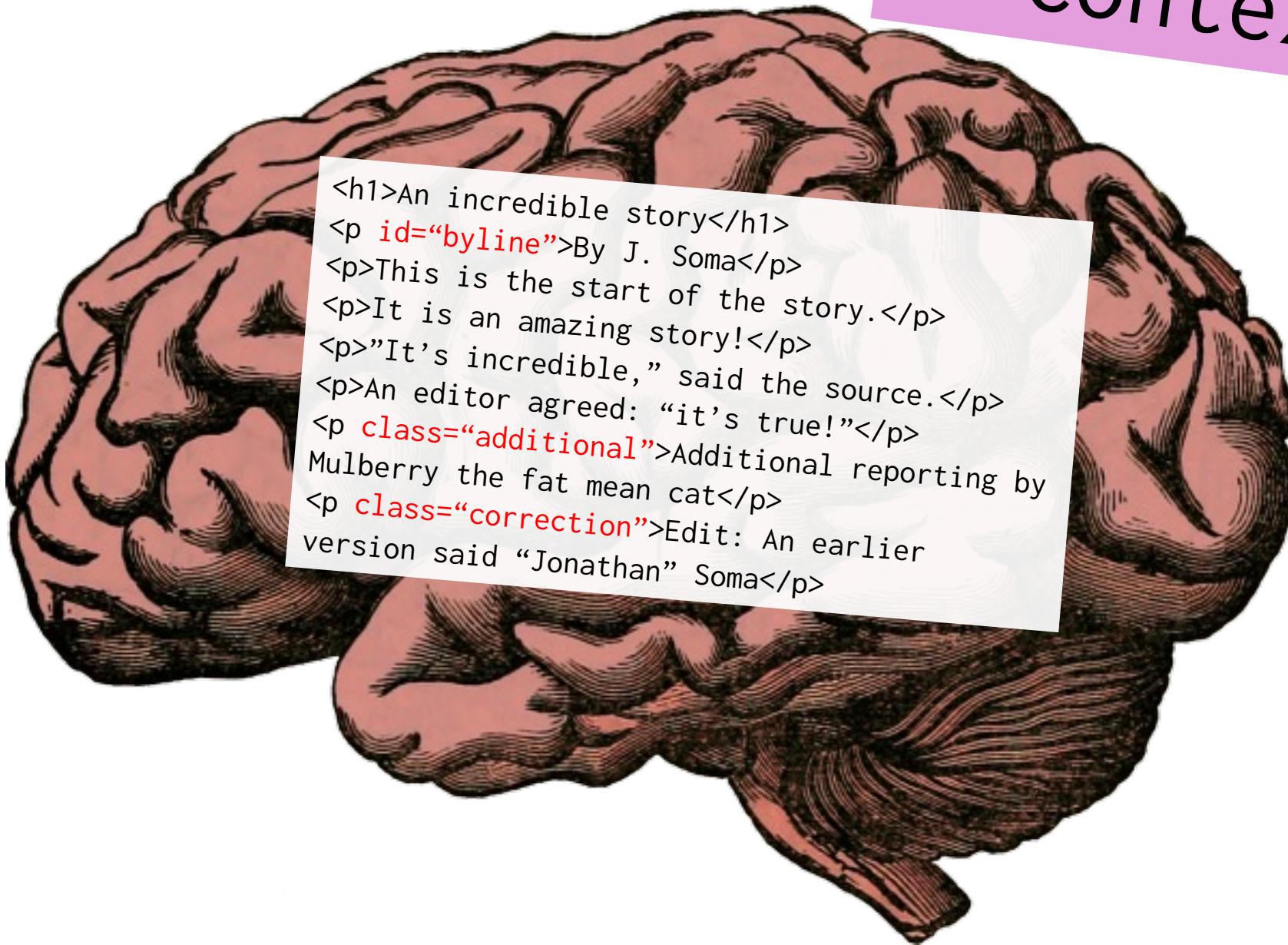


“Write a scraper for these search results:”

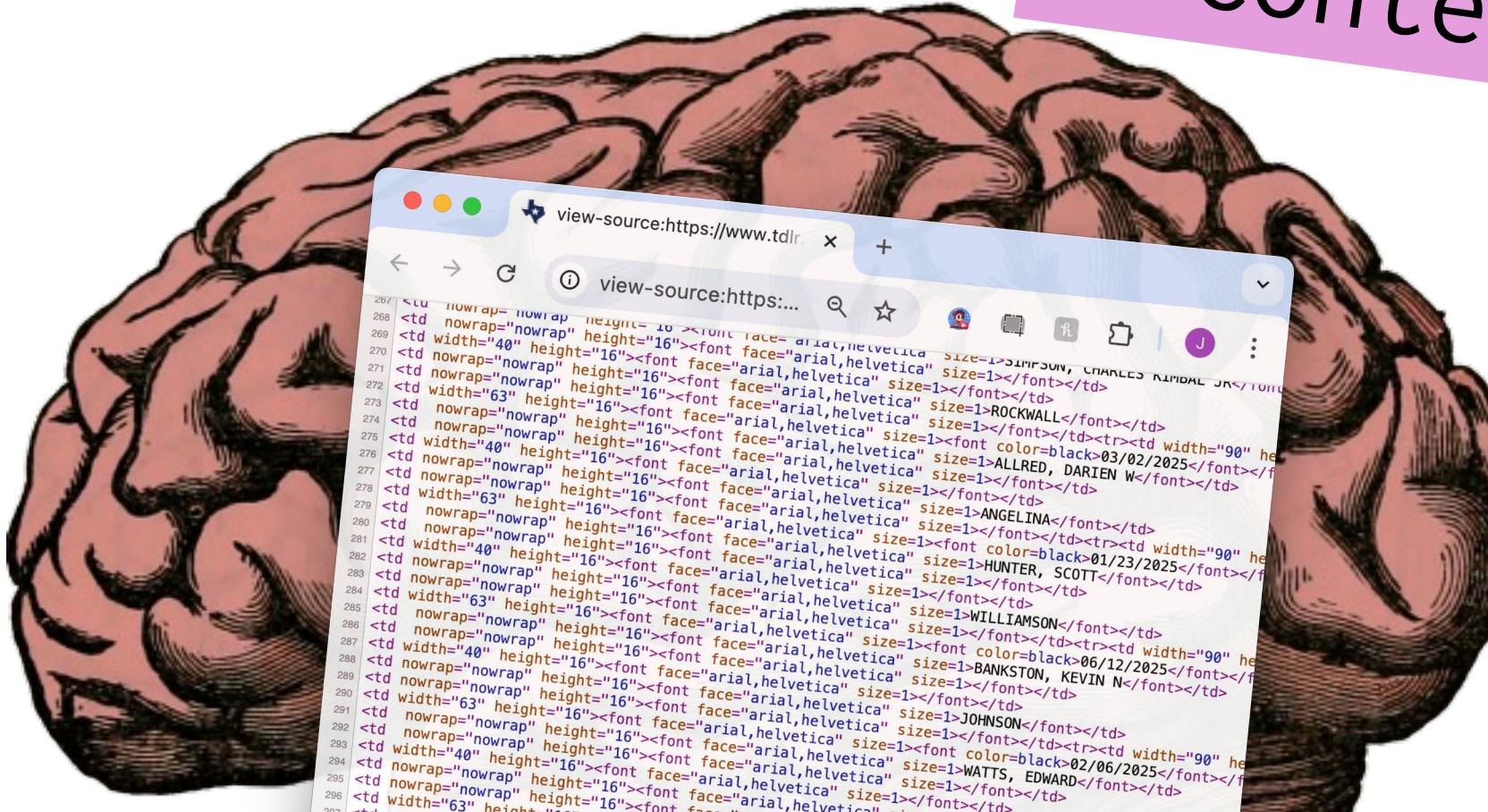


```
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td><r...>
205 <td width="63" height="16"><font face="arial,helvetica" size=1>01/02/2025</font></td>
206 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WALLACE, WT</font></td>
207 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>REYE</font></td>
208 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
209 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
210 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>OF STATE</font></td>
211 <td width="63" height="16"><font face="arial,helvetica" size=1>02/07/2025</font></td>
212 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>VANSICKLE, MICHAEL</font></td>
213 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
214 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
215 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WUPSHUR</font></td>
216 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>05/28/2025</font></td>
217 <td width="63" height="16"><font face="arial,helvetica" size=1>WHITE, D DANIEL CHARLES</font></td>
218 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
219 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
220 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>FANNIN</font></td>
221 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>09/21/2025</font></td>
222 <td width="63" height="16"><font face="arial,helvetica" size=1>BRAUN, THOMAS K SR</font></td>
223 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
224 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
225 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>TITUS</font></td>
226 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>05/07/2025</font></td>
227 <td width="63" height="16"><font face="arial,helvetica" size=1>TURNER, DANIEL S</font></td>
228 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
229 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
230 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>COLLIN</font></td>
231 <td width="63" height="16"><font face="arial,helvetica" size=1>12/13/2024</font></td>
232 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WILSON, ERIC PAUL</font></td>
233 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
234 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROCKWALL</font></td>
235 <td width="63" height="16"><font face="arial,helvetica" size=1>12/06/2023</font></td>
236 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1><sup>Expired</sup></font></td>
237 <td width="63" height="16"><font face="arial,helvetica" size=1>LIMON, MIKE JOEL</font></td>
238 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
239 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
240 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>TOM GREEN</font></td>
241 <td width="63" height="16"><font face="arial,helvetica" size=1>07/13/2025</font></td>
242 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1><font color=red>05/07/2025</font></font></td>
243 <td width="63" height="16"><font face="arial,helvetica" size=1>MOEHLER, JAMES C</font></td>
244 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
245 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
246 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
247 <td width="63" height="16"><font face="arial,helvetica" size=1></font></td>
248 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>05/07/2025</font></td>
249 <td width="63" height="16"><font face="arial,helvetica" size=1>HARRIS</font></td>
250 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
251 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
252 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WIESE, RICHARD L</font></td>
253 <td width="63" height="16"><font face="arial,helvetica" size=1></font></td>
254 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>05/07/2025</font></td>
255 <td width="63" height="16"><font face="arial,helvetica" size=1>DENTON</font></td>
256 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
257 <td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
258 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
```

# “Context window”



# “Context window”



```
267 <tr>
268 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>CHARLES KIMBALL JR</font></td>
269 <td width="40" height="16"><font face="arial,helvetica" size=1>ROCKWALL</font></td>
270 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>03/02/2025</font></td>
271 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ALLRED, DARIEN W</font></td>
272 <td width="63" height="16"><font face="arial,helvetica" size=1>ANGELINA</font></td>
273 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>01/23/2025</font></td>
274 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>HUNTER, SCOTT</font></td>
275 <td width="40" height="16"><font face="arial,helvetica" size=1>WILLIAMSON</font></td>
276 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>JOHNSON</font></td>
277 <td width="63" height="16"><font face="arial,helvetica" size=1>06/12/2025</font></td>
278 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>BANKSTON, KEVIN N</font></td>
279 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WATTS, EDWARD</font></td>
280 <td width="40" height="16"><font face="arial,helvetica" size=1>BASTROP</font></td>
281 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>02/06/2025</font></td>
282 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>SKELTON, ROBERT WAYNE</font></td>
283 <td width="63" height="16"><font face="arial,helvetica" size=1>05/07/2025</font></td>
284 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>HARRIS</font></td>
285 <td width="40" height="16"><font face="arial,helvetica" size=1>05/21/2025</font></td>
286 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>HORTON, IRVIN E JR</font></td>
287 <td width="63" height="16"><font face="arial,helvetica" size=1>MICHAEL JAY</font></td>
288 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>OUT OF STATE</font></td>
289 <td width="40" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
290 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
291 <td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
292 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
293 <td width="40" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
294 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
295 <td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
296 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
297 <td width="40" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
298 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
299 <td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
300 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
301 <td width="40" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
302 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
303 <td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
304 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
305 <td width="40" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
306 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
307 <td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
308 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
309 <td width="40" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
310 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
311 <td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
312 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
313 <td width="40" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
314 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
315 <td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
```

Visit <https://bit.ly/birn-data>, click **scraping**

## What we might need:

- Start URL
- Forms to fill out (*optional*)
- “Rows” of our spreadsheet
- Pagination/“next” pages (*optional*)
- A **magical prompt**

<https://bit.ly/ds-dojo-25>

Open a new Colab  
notebook and let's  
start scraping!

National Azabu / VEGETABLE/野菜

Registration 会員登録 Favorites お気に入り Login ログイン Cart 買い物カゴ

VEGETABLE/野菜 Enter keywords

top | All items | PRODUCE / 野菜・果物 | VEGETABLE/野菜

▶ SALE / お買い得商品  
▶ SEASONAL ITEMS / 季節商品  
▶ NATIONAL AZABU BRAND / ナショナル麻布オリジナル  
▶ INTERNATIONAL FOOD / 各国の食材  
▶ BAKERY / ベーカリー<sub>▶ PRODUCE / 野菜・果物  
▶ VEGETABLE/野菜  
▶ FRUITS/果物  
▶ MEAT / 精肉  
▶ SEAFOOD / 鮮魚  
▶ DAIRY / 乳製品・チーズ・卵  
▶ DELI / デリカ  
▶ PANTRY / 食品  
▶ FROZEN FOOD / 冷凍食品  
▶ SNACKS & SWEETS / お菓子類  
▶ BEVERAGE / 飲料  
▶ ALCOHOL / 酒類  
▶ HOUSEHOLD ESSENTIALS / 雑貨  
▶ BABY / ベビー用品</sub>

## PRODUCE / 野菜・果物 | VEGETABLE/野菜

Found 167 items.

Sort Number Recommended おすすめ順 15

1 2 3 4 5 Next

HALLOWEEN PUMPKIN ( DOMESTIC )  
¥3,580 (8%tax)  
(Tax ¥265)※  
**Limited items 限定品**

HALLOWEEN PUMPKIN S SIZE  
¥1,480 (8%tax)  
(Tax ¥109)※  
**Limited items 限定品**

MATSUTAKE MUSHROOMS  
¥2,980 (8%tax)  
(Tax ¥220)※  
在庫：残りあと 1

BOILED CORN  
¥698 (8%tax)  
(Tax ¥51)※  
**PERISHABLE 冷蔵**

CHESTNUTS

WASABI ( JAPANESE HOURSERADISH )

YUZU CITRON

BROCCOLI

# Playwright gets blocked!

Camoufox solves it,  
but ChatGPT and AI  
tools are very very  
very very very bad  
at it.

The screenshot shows a web browser window displaying the Camoufox documentation at [camoufox.com](https://camoufox.com). The page has a light blue header with the title "Introduction | Camoufox". Below the header is a navigation bar with links to GitHub, Issues, Discussions, PyPi, and Donate. A search bar is also present. The main content area features a sidebar with a navigation menu. The menu includes sections like "Introduction", "Features List", "Stealth Performance", "Python Interface", "Installation", "Usage", "GeoIP & Proxy Support", "Main World Execution", "Remote Server", "Virtual Display", "BrowserForge Integration", "Passing Config", "Fingerprint Injection", "Development", "Liability Disclaimer", and "Donate". The "Introduction" section is currently selected. To the right of the sidebar is the main content area, which contains a large image of a lizard logo with the word "Camoufox" in blue. Below the image is the text "An open source anti-detect browser with robust fingerprint injection." followed by a GitHub trending badge. At the bottom of the content area is a "Warning" box stating "Camoufox is in active development! Releases are available, but may not be suitable for production use." On the far right, there is a "CONTENTS" sidebar with links to "Features" and "Design and Implementation".