

Visit <https://bit.ly/ds-dojo-2024> for materials

Structured data

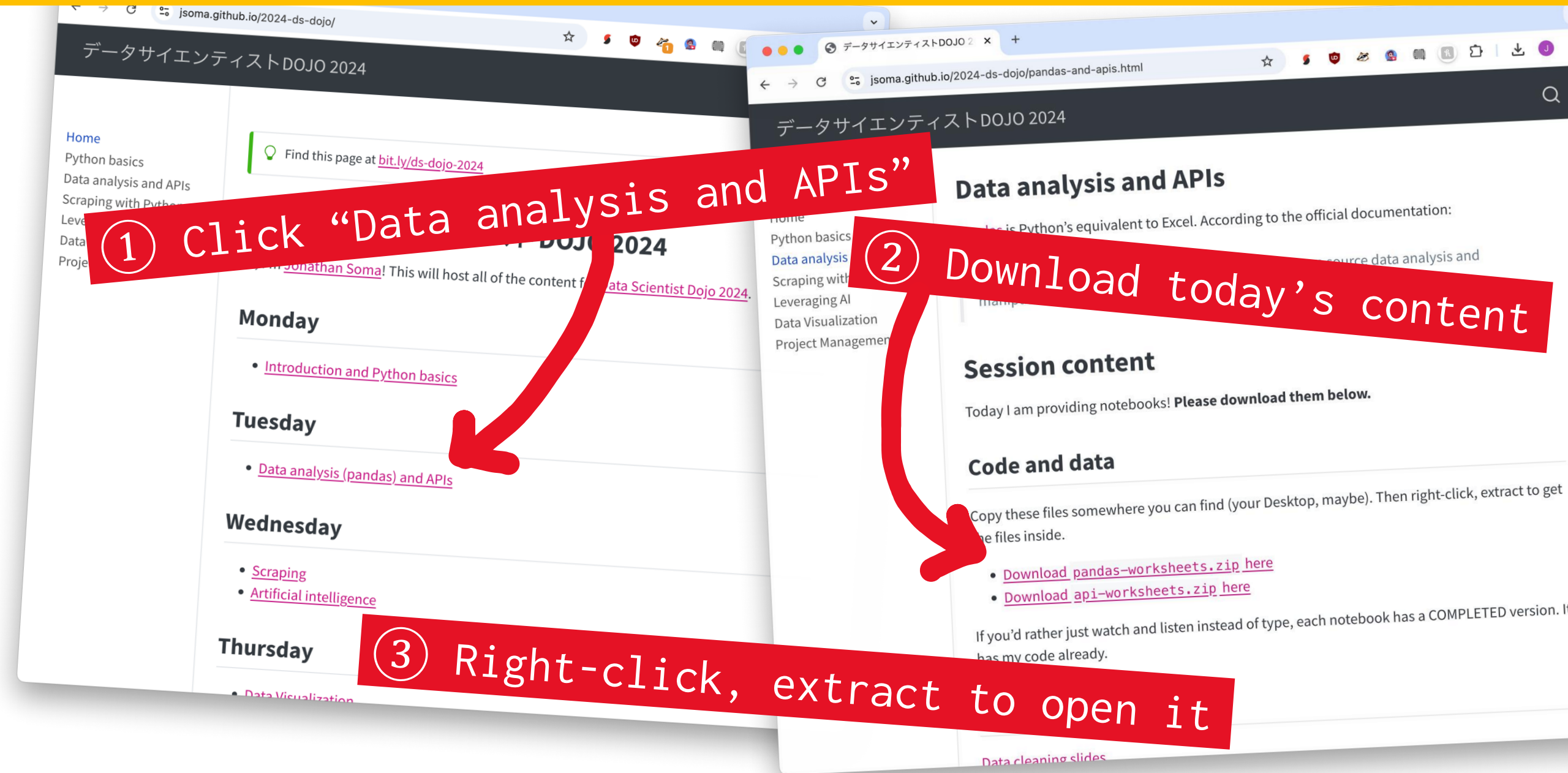
Pandas, CSVs, Excel and APIs

Visit <https://bit.ly/ds-dojo-2024> for materials

① Click “Data analysis and APIs”

② Download today's content

③ Right-click, extract to open it

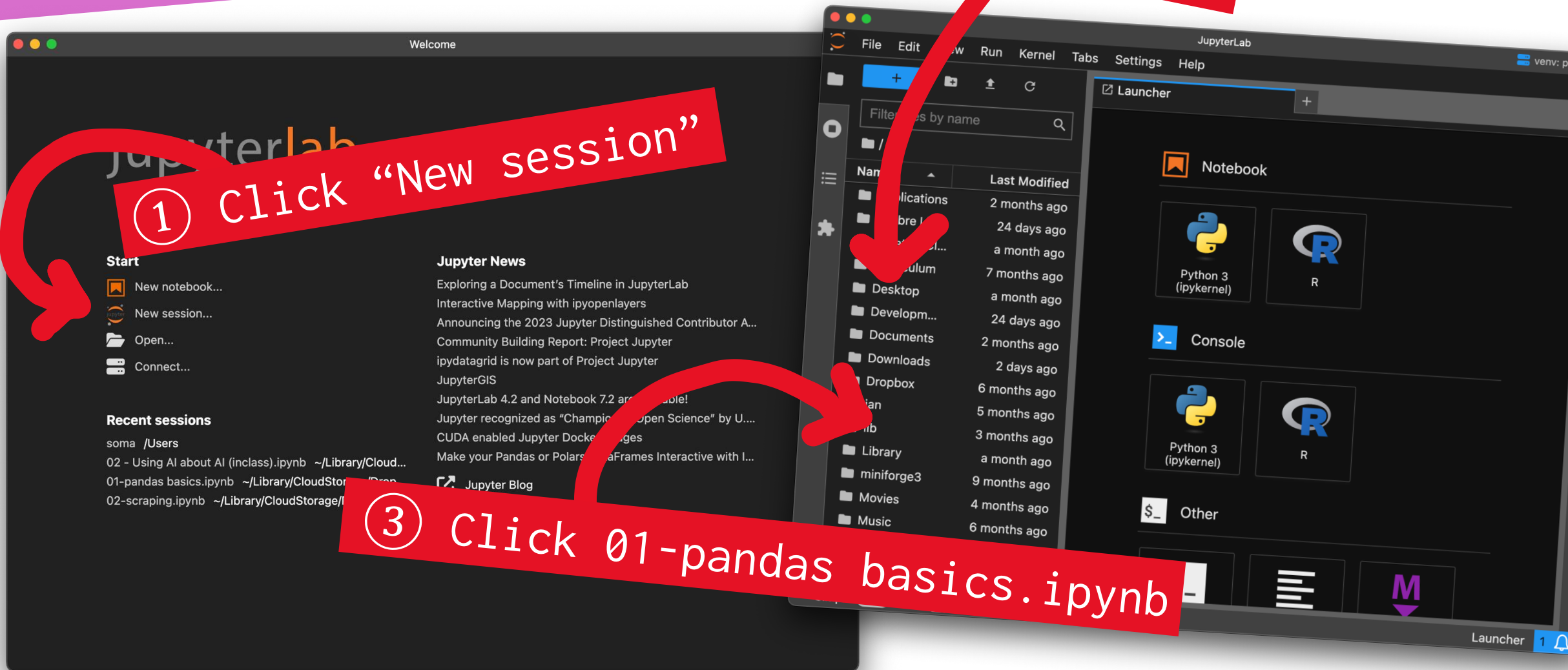


Open 01-pandas basics.ipynb

② Browse

① Click "New session"

③ Click 01-pandas basics.ipynb



Now let's do a
little pandas

Basic data cleaning

replace and split are
99% of the game

66 years old

replace "years old"
with... nothing!

~~66 years old~~

convert to number

66

j soma, js4571@columbia.edu, @dangerscarf

split on “,”

j soma | js4571@columbia.edu | @dangerscarf

take second piece

js4571@columbia.edu

j soma, js4571@columbia.edu, @dangerscarf

j soma | js4571@columbia.edu | @dangerscarf

split

take second

js4571@columbia.edu

split on @

replace with nothing

~~js4571@columbia.edu~~

take second

js4571 | columbia.edu

columbia.edu

I have text like "j soma, js4571@columbia.edu, @dangerscarf" in a Google Sheet and I want to split it by ", ", take the second part, then replace the "@columbia.edu" with nothing. Just give me the formula.

**just ask!
then TEST IT**



You can use the following Google Sheets formula to achieve this:

plaintext

Copy code

```
=SUBSTITUTE(INDEX(SPLIT(A1, ", "), 2), "@columbia.edu", "")
```

Explanation:

1. ``SPLIT(A1, ", ")`` splits the text in cell ``A1`` by ", ".
2. ``INDEX(SPLIT(A1, ", "), 2)`` selects the second part of the split text.
3. ``SUBSTITUTE(INDEX(SPLIT(A1, ", "), 2), "@columbia.edu", "")`` removes the "@columbia.edu" part from the selected text.




I have text like "j soma, js4571@columbia.edu, @dangerscarf" in a Google Sheet and I want to extract the "@columbia.edu" part. Write me a formula to do this. Only show me the formula.

**just ask!
then TEST IT**



excel

 Copy code

```
=RIGHT(A1,LEN(A1)-FIND("@",A1)+1)
```



Feel free to reach out at 555-555-5555



555-555-5555

**just ask!
then TEST IT**



I have text like "j soma, js4571@columbia.edu, @dangerscarf" in a pandas dataframe column named "details" and I want to split it by ",", take the second part, then replace the "@columbia.edu" with nothing. Try not to use .apply. Don't create the df, just give me the code.

< 6/8 >



Certainly! Here's the code to achieve that:

python

Copy code

```
# Split the 'details' column by ",", " and take the second part
df['email'] = df['details'].str.split(',', ' ').str[1]

# Replace "@columbia.edu" with nothing
df['email'] = df['email'].str.replace('@columbia.edu', '', regex=False)
```

This will split the column and then replace the domain part as required.



just ask!
then TEST IT

I have text like "j soma, js4571@columbia.edu, @dangerscarf" in a pandas dataframe column named "details" and I want to extract the "@columbia.edu" part. Try not to use .apply. Don't create the df, just give me the code.



You can use the `str.extract` method to extract the "@columbia.edu" part from the "details" column. Here's the code to do that:

python

 Copy code

```
df['extracted'] = df['details'].str.extract(r'(\b@\w+\b)\.edu\b')
```



Regular expressions

regex is the other
other 1%

66 years old

hug what we want
(capture group)



(66) years old

match anything



(.*) years old

I am 66 years old

hug what we want



I am (66) years old

match anything



I am (.*) years old

Contact me at 555-555-5555

hug what we want

Contact me at (.*)

but maybe not!

Contact me at 555-555-5555 or
js4571@columbia.edu

Contact me at (\d\d\d-\d\d\d-\d\d\d\d)

just numbers

Contact me at 555-555-5555



maybe we don't
need an anchor?

(\d\d\d-\d\d\d-\d\d\d\d)

Contact me at 5555555555



(\d\d\d-?\d\d\d-?\d\d\d\d)

Contact me at 555+555+5555



any character is fine!
(and also optional)

(\d\d\d.\?\d\d\d.\?\d\d\d\d)

I have a table that looks like this. Write the SIMPLEST POSSIBLE regex to just extract the values for "Deforested area" in km. Maybe anchor around the year at the start of each line?

just ask!
then TEST IT

Year	Deforestation of the Amazon rainforest	Deforested area	Accumulated loss
2005	10,311 km2 (7,341 sq mi)	332,470 km2 (128,370 sq mi)	
2006	10,311 km2 (5,516 sq mi)	346,756 km2 (133,883 sq mi)	
2007	10,311 km2 (4,498 sq mi)	358,407 km2 (138,382 sq mi)	
2008	10,311 km2 (4,985 sq mi)	371,318 km2 (143,367 sq mi)	
2009	7,464 km2 (2,882 sq mi)	378,782 km2 (146,249 sq mi)	
2010	7,000 km2 (2,700 sq mi)	385,782 km2 (148,951 sq mi)	
2011	6,418 km2 (2,478 sq mi)	392,200 km2 (151,400 sq mi)	
2012	4,571 km2 (1,765 sq mi)	396,771 km2 (153,194 sq mi)	
2013	5,891 km2 (2,275 sq mi)	402,662 km2 (155,469 sq mi)	
2014	4,848 km2 (1,872 sq mi)	407,510 km2 (157,340 sq mi)	

< 5/5 >



To extract the "Deforested area" values in km² using a regex, you can use the following simple pattern:

SCSS

Copy code

```
^\d{4}\s+(\d{1,5})\s+km2
```

Here's a breakdown of the regex:

- `^` asserts the position at the start of the line.`
- `\d{4}` matches exactly four digits (the year).`
- `\s+` matches one or more whitespace characters.`
- `(\d{1,5})` captures between 1 and 5 digits (the deforested area value) in a capturing group.`
- `\s+` matches one or more whitespace characters.`
- `km2` matches the literal string "km2".`

This pattern will capture the deforested area in km² from each line.

🔊 📄 ↺ 🔒 ⚙️



Now let's do a
little more pandas

Visit <https://bit.ly/ds-dojo-2024> for materials

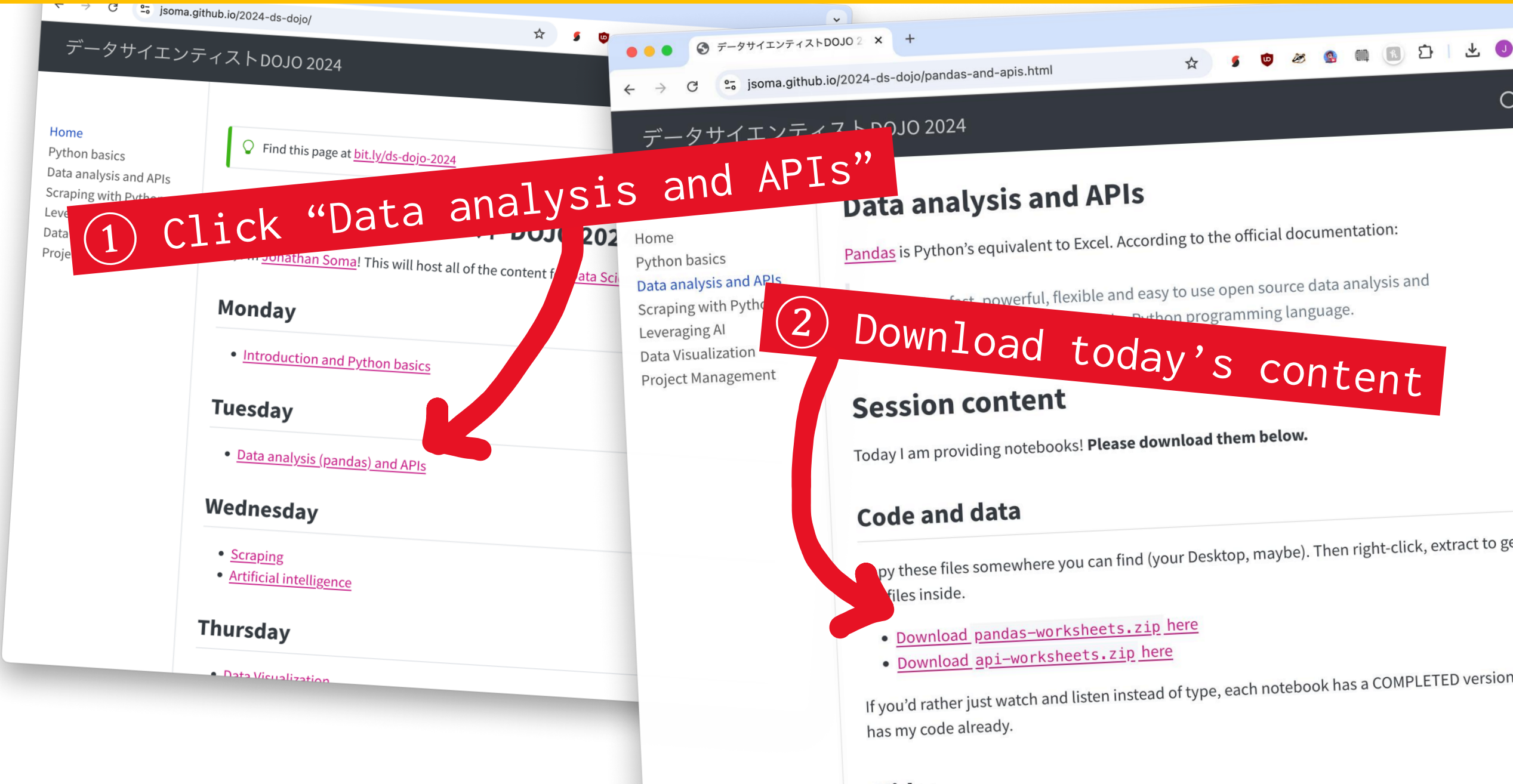
APIs

structured data from the
internet

Visit <https://bit.ly/ds-dojo-2024> for materials

① Click “Data analysis and APIs”

② Download today's content



Does anyone have a
favorite pokemon?

This is an important question

snorlax - Google Search


google.com/search?q=snorlax&oq=snorlax&gs_lcrp=EgZjaHJvbWUyDwgAEEUYORiRAhiABBiKBTIMCAEQABhDGIAEGloFMg0IAhAAGJECGIAEGloFMgwIAxAA...

Google

snorlax

Tools

All Images Shopping Videos News Maps Web More

Bulbapedia

https://bulbapedia.bulbagarden.net/wiki/Snorlax_(Po...

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



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) - Bulbapedia

bulbapedia.bulbagarden.net/wiki/Snorlax_(Pokémon)

BULBAGARDEN

NewsBulbapediaForumsDiscordMore

EditorsAccountSearch Bulbapedia

articlediscussionview sourcehistorytoolsin other languages


In commemoration of Nintendo's 135th 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

Sleeping Pokémon


カビゴン

Kabigon

#0143



Snorlax



Gigantamax Snorlax

[Images on the Bulbagarden Archives](#)




- 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

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

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

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.

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/>

application.json

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



APIs

Structured data



The RESTful Pokémon API

Serving over 2.5 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!

`https://pokeapi.co/api/v2/` `pokemon/ditto`



Submit

jupyterlab

Start



New notebook...



New session...



Open...



Connect...

Recent sessions

02 - Using AI about AI (inclass).ipynb ~/Library/Cloud...

01-pandas basics.ipynb ~/Library/CloudStorage/Drop...

02-scraping.ipynb ~/Library/CloudStorage/Dropbox/...

Jupyter News

JupyterGIS

JupyterLab 4.2 an

Jupyter recognize

CUDA enabled Ju

Make your Pandas

JupyterLab Deskt

Python environme

JupyterLab 4.1 an

Curvenote joins th

Python Environme

Jupyter Blog

Open JupyterLab

doc - JupyterLab

venv: python

Untitled2.ipynb

File Edit View Run Kernel Tabs Settings Help

Notebook Python 3 (ipykernel)

```
[ ]: from playwright.async_api import async_playwright

# Start Playwright
playwright = await async_playwright().start()
browser = await playwright.chromium.launch(headless=False)
page = await browser.new_page()

# Navigate to the URL
await page.goto("https://www.tdlr.texas.gov/LicenseSearch/")

[ ]: import pandas as pd

# Wait for the results to be visible after form submission
await page.wait_for_selector("table")

# Scrape table rows
rows = await page.query_selector_all("table tr")

data = []
for row in rows:
    cells = await row.query_selector_all("td")
    if len(cells) > 0:
        license_name = await cells[0].inner_text()
        expiration_date = await cells[1].inner_text()
        license_holder = await cells[2].inner_text()

df = pd.DataFrame(data)
df.head()
```

Visit <https://bit.ly/ds-dojo-2024> for materials

APIs

structured data from the
internet