Advanced Data Journalism Part II: Unstructured Data

Jonathan Soma

Columbia Journalism School

js4571@Columbia.edu

First, we review.

And do some exercises, because watching is so so boring.

REZULTATET E OFERTAVE - AKP SHITJA PERMES LIKUIDIMIT 36

Tabela

Nr.	Njësitë	AKP ID	Çmimi më i Lartë	Ofertuesi	Çmimi
1	Njësia nr.01: Kualiteti Lokali në Gjilan	GJI005	€34,500	L125	€34,500
2	Njësia nr.02: Kualiteti Lokali në Kamenicë	GJI005	€63,555	L50	€63,555
3	Njësia nr.03: Pasuria Bujqësore Toka në Slivovë (Lot A)	GJI011	€31,401	L80	€31,401
4	Njësia nr.04: Pasuria Bujqësore Toka në Kosinë (Lot B)	GJI011	€0	0	€0
5	Njësia nr.05: KB Morava Toka në Livoç të ulët (Lot A)	GJI018	€113,334	L124	€113,334
6	Njësia nr.07: NTSh Sharr Depoja në Kaçanik	GJI105	€32,121	L70	€32,121
7	Njësia nr.08: KB Luboteni Toka dhe Objekti	GJI153	€487,350	L85	€487,350
8	Njësia nr.09: Agrokultura Toka në Gjilan (Lot J)	GJI004	€6,100	L95	€6,100
9	Njësia nr.10: Pasuria Bujqësore Toka në Rahovicë (Lot C)	GJI011	€0	0	€0
10	Njësia nr.11: Pasuria Bujqësore Toka në Pojatë (Lot A)	GJI011	€18,200	L78	€18,200
11	Njësia nr.12: KB Zhegra Toka në Lladovë (Lot B)	GJI024	€0	0	€0
12	Njësia nr.13: Agromorava Toka në Pozharan (Lot D)	GJI035	€31,111	L81	€31,111
13	Njësia nr.14: Agromorava Toka në Gushicë (Lot A)	GJI035	€0	0	€0
14	Njësia nr.16: Agrokultura Toka në Gjilan (Lot M)	GJI004	€21,600	L96	€21,600
15	Njësia nr.17: NSH Produkti Tokë në Dashevc	MIT008	€0	0	€0
16	Njësia nr.18: KB Gllogovci Parcela 00281-0	PRN165	€81,750	L126	€81,750
17	Njësia nr.19: NSH LUX Lokali "Kosovka"	MIT039	€191,000	L15	€191,000
18	Njësia nr.20: NSH LUX Lokali "Elegant"	MIT039	€108,000	L16	€108,000
19	Njësia nr.21: NSH LUX Lokali "Treshnja"	MIT039	€0	0	€0
20	Njësia nr.22: NSH LUX Lokali "Lovac"	MIT039	€14,395	L09	€14,395
21	Njësia nr.23: Konfeksioni Rashka Lokali në Mitrovicë	MIT141	€148,888	L52	€148,888
22	Njësia nr.24: NSH Produkti Tokë në Polac i Vjetër	MIT008	€0	0	€0
23	Njësia nr.25: Agrokultura Furra e Vjetër në Stantërg	MIT002	€0	0	€0
24	Njësia nr.26: Produkti Tokë në Bajë	MIT008	€0	0	€0
25	Njësia nr.27: Produkti Tokë në Suhogërllë	MIT008	€0	0	€0
26	Njësia nr.28: Produkti Tokë në Kllodërmicë	MIT008	€0	0	€0
27	Njësia nr.29: Produkti Tokë në Obri e Poshtme	MIT008	€0	0	€0
28	Njësia nr.30: Elan Parcela 205 ZK Beçuk	MIT012	€0	0	€0
29	Njësia nr.31: Elan Parcela 761 ZK Zhilivodë	MIT012	€0	0	€0
30	Njësia nr.32: Elan Parcela 114-1 ZK Stanovc i Epërm/Përronaj	MIT012	€0	0	€0
31	Njësia nr.33: Elan Parcelat 1041-0 dhe 1045-0 ZK Dumnicë e LLugës/Mesme	MIT012	€0	0	€0

Can I put this into ChatGPT and ask questions about it?

cheating.csv

Date/Time of Incident	Conduct Charges	Course
8/3/2021	Plagiarism	English 472/Shakespeare Su
9/30/2021	Cheating, Plagiarism	Exercise, Sport and Health E
9/30/2021	Plagiarism	English 111-74-Principles of
9/30/2021	Plagiarism	English 111-74-Principles of
9/30/2021	Plagiarism	English 111-761 Principles of
10/11/2021	Cheating, Plagiarism	Anthropological Sciences
10/11/2021	Cheating, Plagiarism	Anthropological Sciences
10/11/2021	Cheating, Plagiarism	Anthropological Sciences
10/13/2021	Cheating	Health Communication and C
10/13/2021	Cheating	Health Communication and C
10/15/2021	Facilitation	Communication
10/15/2021	Plagiarism	Communication
10/15/2021	Cheating	Principles of Marketing/Marke
10/15/2021	Facilitation	Communication
10/16/2021	Cheating	English 200: Literary Texts a
10/18/2021	Plagiarism	Biology 232/01--Organismal
10/19/2021	Cheating	Introductory Psychology/Psyc

What Python tool do you (and ChatGPT) use to analyze data?


```python
df = pd.read_csv("procurement-results.csv")
df.head()
```

	filename	winner	description	postal address	city	country	
0	b08_njoftim_per_dhenje_te_kontrates (619).pdf	KUVENDI I KOMUNES SE PRISHTINES	NaN	Adresa Postare: TREPQES NR.2	PRISHTINË	Kosovë	
1	b08_njoftim_per_dhenje_te_kontrates (58).pdf	KUVENDI I KOMUNES SE PRISHTINES	" Sanimi dhe asfaltimi i rrugës Fehmi Lladrovc...	Adresa Postare: TREPQES NR.2	PRISHTINË	Kosovë	59
2	b08_njoftim_per_dhenje_te_kontrates_1 (220).pdf	KUVENDI I KOMUNES SE PRISHTINES	Ndertimi i rruges ne Grashtice- Pune shtese	Adresa Postare: TREPQES NR.2	PRISHTINË	Kosovë	4,
3	b08_njoftim_per...	KUVENDI I	Shërbi...				

Did we set up Python on our own computer?

Why should you probably not pay a subscription fee for transcription software?

An exercise

Misinformation on TikTok: Ho

pulitzercenter.org/...

SVG Crowbar   SVG Crowbar   SVG Crowbar   Home   Top Shot   All Bookmarks

Once we had a handful of accounts we wanted to examine further, we opened their profiles on a web browser, scrolled through the page until no more videos were loaded and downloaded the HTML page. We wrote a Python scraper to each account's video urls.

-dlp, we then downloaded each video, its meta-chived them on a local drive (special thanks to rt at *The Washington Post* and Bellingcat for o that library).

Misinformation on TikTok: Ho

pulitzercenter.org/misinformation-tiktok-how-documented-examined-hundreds-videos-differe...

SVG Crowbar   SVG Crowbar   SVG Crowbar   Home   Top Shot   SVG Crowbar 2   popup with tags   Coding   Viz   gamedev   bread

Use yt-dlp and Whisper and Python and Google Colab to download and transcribe a short YouTube video.

'Documented' Examined Hundreds of Videos in Different Languages

ranscription

h videos can be tough, since going through them could take a long time. To help cut down some of the work and also give us an approximate understanding of the work we're looking at, we decided to auto-transcribe the videos we had downloaded using the open-source version of Whisper. This transcription speech recognition model, published by Open AI, is available in various languages. The performance of the model varied from language to language: We tested the library in two languages, Vietnamese and

Author:

Lam Thuy Vo
GRANTEE

Translate page with Google

Adv Data Journalism II: Unstructured Data

Scraping web sites and unstructured data

Making the best of data from the internet and bad, ugly data from PDFs

Links:

- Easy Scraper (modified) Chrome extension for scraping – modified by me t
- Procurement contracts
- Bid document listing (in Albanian, because it has more than the Engli
- Outdoor ads as monuments, sculptures, stunts, installations, statues
  monument's Wikipedia page
- Motrix A great app for downloading things
- Bulk PDF Downloader Chrome extension for downloading all of the PI
- yt-dlp A Python library for downloading videos on the internet
- Whisper A Python library for transcribing audio and video
- Stacher A tool for downloading YouTube videos without code
- Misinformation on TikTok How 'Documented' Examined Hundreds of

Use yt-dlp and Whisper turbo and Python and Google Colab to download and transcribe a short YouTube video.

Scraping web sites and unstructured

Table of Contents

Do I have to be a programmer to analyze data with AI?

An exercise

Let's do this ourselves.

AI Do's and Don't

AI Do's and Don'ts

How can AI help with newsroom tasks?

📦 Data: ai-dos-and-donts-data.zip

📄 Slides: birn-ai-dos-donts.pdf

Links:

- Pristina procurement contract award notices A big long list, in Albanian
- NotebookLM The best possible conversation partner for large numbers
- Claude for Sheets Extension to use AI i
- Claude for Sheets example spreadshee
- Anthropic console Where to get a Clau
  - Claude for Sheets example spreadsheet

Can you discriminate against poor people as a landlord?

MyCity Chatbot Beta

Chatbot Beta

NYC MyCity Official website of the City of New York

MyCity Chatbot Beta

We are continuously working to improve the MyCity Chatbot, which uses business-related NYC agen
AI to answer your questions. Your feedback is invaluable for refinement.

Example Questions

"How do I avoid noise violations and complaints for my construction company?"

"What information can you help me with?"

"How do I apply for the MWBE program?"

Capabilities

Trained to provide you with official NYC Business information.

Will not use the contents of your chat history to learn new information.

Responds to languages required by Local Law 30.

Limitations

May occasionally produce incorrect, harmful or biase

Limited knowledge of the beyond NYC Business top

Trained to decline inappropriate requests.

Terms of Use   Privacy Policy

© 2024 City of New York. All Rights Reserved.

Ask MyCity Chatbot business-related questions only

IMPORTANT: Responses may occasionally produce inaccurate or incomplete content. Validate answers on NYC.gov and MyCity

An exercise

Let's do this ourselves.

Scraping web sites and unstru

Table of Contents

- Useful Links
- Intro
- Using NotebookLM ←
- Using Easy Scraper
- Running code in Colab
- Exercises

Download files: 📦 scraping-web-data.zip

When all you have is a hammer everything looks like a nail.

but there are many things that can be solved with a Python hammer.

0C8EE446-2572-487E-9190-DA2E41B49247.pdf — Nov 2, 2017 at 10:07 AM — 502 KB — PDF Document
0D8822D5-8E33-4601-80C6-4DA76323384A.pdf — Jul 23, 2025 at 2:12 PM — 309 KB — PDF Document
1E15DB7F-47E6-4A49-9B37-0B56BDC0308E.pdf — May 8, 2019 at 3:56 PM — 599 KB — PDF Document
2B8B3DD2-7A05-436F-B676-3ED255C7F788.pdf — May 28, 2025 at 1:50 PM — 181 KB — PDF Document
2F551774-A6CF-4DA4-94C1-8421BB412E8A.pdf — Nov 2, 2017 at 10:07 AM — 86 KB — PDF Document
3E175E76-4F07-4C37-BF67-C310251C5239.pdf — Aug 21, 2019 at 3:34 PM — 55 KB — PDF Document
4B177C46-2CAF-4F17-8F1C-81ADA9F2EA3E.pdf — Jan 1, 2018 at 9:40 AM — 603 KB — PDF Document
5AD5D3DE-5EA7-41AB-AAA9-4210CAA97590.pdf
5AE7104C-8724-490E-907B-871BD6A11C7F.pdf
6B7EFF38-0A3E-43F9-BF89-A817E26CE859.pdf
6CE0359E-6D13-49EC-B2AD-C296A79D48E5.pdf
6D945485-6F65-4925-9374-91B2C70A64E7.pdf
7A770171-5E96-4F8C-991B-A601DE0F40DD.pdf
9A02D67A-C4A9-44D8-87E9-ABC0C8813DFF.pdf
9CA182A2-C4B4-422B-8020-2289CF8B31BB.pdf
9E70B6EF-4DB2-4544-8871-5B38CA828E80.pdf
9F338E1D-B999-4EC6-B10D-C934EFFCD345.pdf
15A8700C-E26C-42A1-AAC9-E03180FE0A5E.pdf
38C3475F-3D0E-4B56-A90B-1B83EB53CBD2.pdf
052C52EF-7825-4C97-A14E-27554DD1938C.pdf
62B523FB-5DF9-478C-97DE-6007F70A8476.pdf
64D0BF1C-B88C-410C-BD57-F5167AEE6BBE.pdf
73C02FA3-EAA1-4403-AE64-97166333FA6F.pdf
73CBC8C0-7EBE-48DD-A558-85E62140489E.pdf

b08_njoftim_per_dhenje_...(65).pdf
b08_njoftim_per_dhenje_...(79).pdf
b08_njoftim_per_dhenje_...(80).pdf
b08_njoftim_per_dhenje_...(95).pdf
b08_njoftim_per_dhenje_...(97).pdf
b08_njoftim_per_dhenje_...(111).pdf
b08_njoftim_per_dhenje_...133).pdf
b08_njoftim_per_dhenje_...170).pdf
b08_njoftim_per_dhenje_...(171).pdf
b08_njoftim_per_dhenje_...176).pdf
b08_njoftim_per_dhenje_...179).pdf
b08_njoftim_per_dhenje_...183).pdf
b08_njoftim_per_dhenje_...194).pdf
b08_njoftim_per_dhenje_...200).pdf
b08_njoftim_per_dhenje_...201).pdf
b08_njoftim_per_dhenje_...202).pdf
b08_njoftim_per_dhenje_...(211).pdf
b08_njoftim_per_dhenje_...212).pdf
b08_njoftim_per_dhenje_...213).pdf
b08_njoftim_per_dhenje_...240).pdf
b08_njoftim_per_dhenje_...255).pdf
b08_njoftim_per_dhenje_...260).pdf
b08_njoftim_per_dhenje_...274).pdf
b08_njoftim_per_dhenje_...284).pdf
b08_njoftim_per_dhenje_...287).pdf
b08_njoftim_per_dhenje_...295).pdf
b08_njoftim_per_dhenje_...307).pdf
b08_njoftim_per_dhenje_...308).pdf

prishtinaonline.com/prokurimi/njoftim-per-dhenjen-e-kontrates

SVG Crowbar | SVG Crowbar | SVG Crowbar | Home | Top Shot | SVG Crowbar 2 | popup with tags | Coding | Viz | gamedev | bread | lede | elrn | make | cat | All Bookmarks

PRISHTINA ONLINE

KRYETARI KUVENDI DOKUMENTE LAJMET QYTETI INVESTIMET KONKURSET KONTAKTI QASJA NË DOKUMENTET ZYRTARE

Shqip English Srpski   Kërko

DEVELOPING TOGETHER PRISHTINA 2016

E-Komuna   Subvencionet në DKRS   eLeja   ePërcjellja   Thirrje Emergjente

Drejtoritë

Administrata
Shëndetësia
Mirëqenia Sociale
Arsimi
Kultura
Sport
Bujqësia
Financa
Prona
Urbanizmi
Investime kapitale dhe menaxhimit të kontratave
Shërbime Publike, Mbrojtje...
Kadastër
Inspekcion
Planifikim strategjik
Parqeve
Siguri dhe Emergjenca
Transformimi

Prokurimi

Departamenti i Prokurimit
Plane dhe Raportet e prokurimit

Departamenti i Prokurimit

Plane dhe Raportet e prokurimit

Njoftimi paraprak

Njoftim për kontratë

Njoftim për dhënien e kontratës

Njoftim për konkurs të projektimit

Njoftim për rezultatet e konkursit të projektimit

Njoftim për anulim

Njoftim për informata shtese

Anulim i dhënies së kontratës

Njoftim për dhënien e kontratës

- Furnizimi , servisimi aparateve kundër zjarrit 25.08.2025
- Furnizim me material higjienik për nxënësit e kom... egjiptian 22.08.2025
- Ndërtimi i tri mbikalimeve të përkohshme në lagje Abëri - (punë shtesë) 21.08.2025
- Intervenime, riparime dhe mirembajtja urgjente d...
- Mirembajtja e aparaturave per printim, kopjim, ska... toner, drum si dhe ngjyra per plotter - Mirmbajtja ... 19.08.2025
- Renovimi i shkollave dhe IEA-së -Gëlqerosja 12.08.2...
- 174 - Prodhimi, furnizimi dhe vendosja e numrave ... banimit në Komunën e Prishtinës. Ri-tender 12.08....
- Revovimi i shkollave ne IEA, Demolimi, furnizimi ... dritareve ne IEAA 06.08.2025
- Instalimi Elektrikes - Blerja, zhvendosja, montimi ... per shkollen Pavarsia 01.08.2025
- Pajisje të ndryshme pune në Drejtorinë e Inspeksio... 01.08.2025
- Organizimi i koncertit Rame Lahaj with friends 31....
- Organizimi i festivalit Sunny Hill 31.07.2025
- 137 Furnizimi dhe montimi i serrave 30.07.2025
- Renovimi i shkollave dhe IEA-Dysheme 30.07.2025

https://prishtinaonline.com/uploads/mirembajtja_e_aparaturave_per_printim_kopjim_skanim_furnizimi_i_tyre_me_toner_drum_si_dhe_ngjyra_per_plotter_-_mirmbajtja_e_e-ki...

---

pak-ks.org/page.aspx?id=1%2C106

SVG Crowbar | SVG Crowbar | SVG Crowbar | Home | Top Shot | SVG Crowbar 2 | popup with tags | Coding | Viz | gamedev | bread | lede | elrn | make | cat | All Bookn

WEBMAIL | SHQIP SRPSKI ENGLISH

AGJENCIA KOSOVARE E PRIVATIZIMIT
KOSOVSKA AGENCIJA ZA PRIVATIZACIJU
PRIVATISATION AGENCY OF KOSOVO

Ballina   Pyetje të shpeshta   AKP   Legjislacioni   Shpallje   Raporte   Lajme   Kontakti

Home / Rezultatet e Ofertimit

Rezultatet e Ofertimit

📄 Shitja - 74

📄 Shitja - 73

📄 Shitja - 72

📄 Shitja - 71

📄 Shitja - 70

📄 Shitja - 69

📄 Shitja - 67

📄 Shitja - 68

|< 1 2 3 4 5 6 7 8 9 10 >|

» SHITJA E ASETEVE NE LIKUIDIM

Shitjet Paraprake të Aseteve
Rezultatet e Ofertimit
Vendimet e Shitjes

Kërkesat kreditore

Qendra e Thirrjeve
+381 (0) 38 500 400

< August 2025 >

Sun	Mon	Tue	Wed	Thu	Fri	Sat
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

Scraping

Is taking data from the web and converting it into structured format
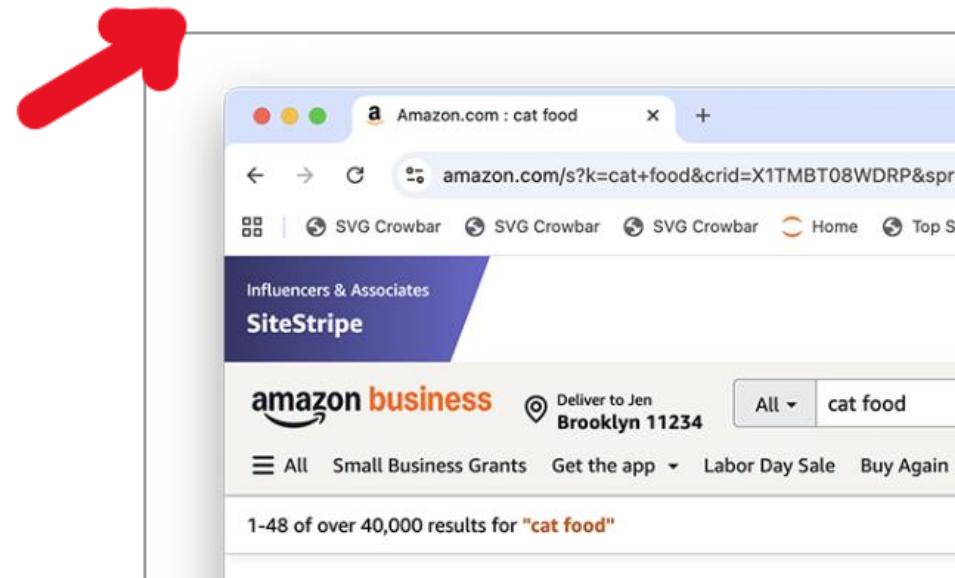
Scraping web sites and unstructured

https://bit.ly/birn-2025-data

Table of Contents

Download files: 📦 scraping-web-data.zip

Using Easy Scraper

After you've installed Easy Scraper, you first need

🚀 Easy Scraper - Click-Friendly Version

The wonderful Easy Scraper doesn't work on all pages, especially specific awful government web pages you are trying to scrape. This is a slightly edited version that should work in more situations.

⬇ Download Extension

📥 Installation Instructions

1. Download the extension using the button above

2. Unzip the downloaded file to a folder on your computer

3. Open Chrome and navigate to `chrome://extensions/`

4. Enable **"Developer mode"** using the toggle in the top right

5. Click **"Load unpacked"**

6. Select the folder where you unzipped the extension

7. The extension is now installed! Look for the Easy Scraper icon in your toolbar

⚠️ **Note:** When the extension uses CDP clicking for restricted pages, you'll see a "Debugger attached" banner. This is normal and required for the enhanced functionality.
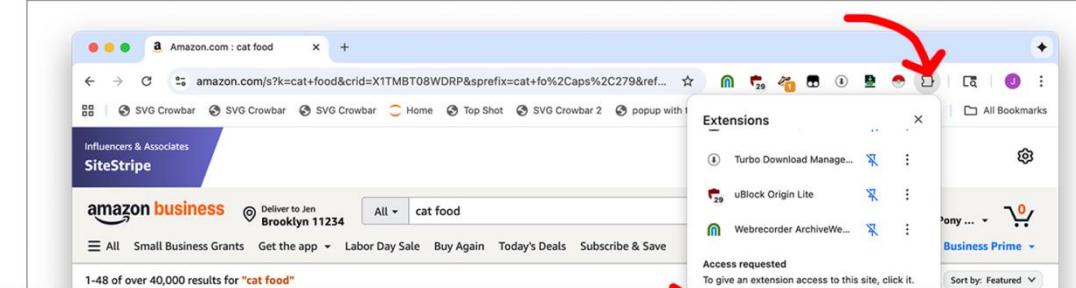
Scraping web sites

← → ↻ ⊙ jsoma.github.io/2025-birn/scraping-web.html ☆

⊞ | 🌐 SVG Crowbar | 🌐 SVG Crowbar | 🌐 SVG Crowbar | 🔴 Home | 🌐 Top Shot | 🌐 SVG Crowbar 2 | 🌐 popup with tags | 📁 Coding | » | 📁 All Bookmarks

Scraping web sites

Useful Links

install ➡

- [Easy Scraper (modified)](#) - Chrome extension for scraping – modified by me to make it work in more cases
- [Procurement contracts](#)
- [Bid document listing](#) - (in Albanian, because it has more than the English version)
- [Outdoor ads as monuments, sculptures, stunts, installations, statues](#) - A resource from the NEWBORN monument's Wikipedia page
- [Motrix](#) - A great app for downloading things
- [Bulk PDF Downloader](#) - Chrome extension for downloading all of the PDFs linked on a page

Scraping is a powerful way to convert web sites into data files that you can analysis with Sheets, Excel, or even ChatGPT.

Using Easy Scraper

After you've installed Easy Scraper, you first need to open it by clicking the extensions button.

How to scrape ➡

	a-size-base-	a-size-small	a-icon-alt		a-size-small	aok-offscreen		a-offscreen	a-price-symb	a-price-whol	a-price-deci	a-price-fract	a-size-small	a-offscreen	a-price	a-icon-text	a-column	a-text-bold	a-button-tex	a-offscreen	a-offscreen
options.																				
options.																				
Tiki Cat Born	4.4	4.4 out of 5 st	https://www.		-452	Price, produc	https://www.	$15.48	$	15	.	48	(	$10.32	$10.32	Today		FREE delivery	Today 5‚ÄÔI	Add to cart	
Purina ONE N	4.8	4.8 out of 5 st	https://www.	(9.2K)		Price, produc	https://www.	$41.78	$	41	.	78	(	$1.90	$1.90	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart	List: $45.96	$45.96
Purina ONE N	4.8	4.8 out of 5 st	https://www.	(7.8K)		Price, produc	https://www.	$17.48	$	17	.	48	(	$17.48	$17.48	Today		FREE delivery	Today 5‚ÄÔI	Add to cart	List: $19.23	$19.23
Purina Fancy	4.7	4.7 out of 5 st	https://www.	(42.3K)		Price, produc	https://www.	$20.97	$	20	.	97	(	$0.29	$0.29	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart	List: $21.36	$21.36
Purina Friskie	4.6	4.6 out of 5 st	https://www.	(16.8K)		Price, product page		$31.20	$	31	.	20	(	$0.14	$0.14	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart		
Meow Mix Ca	4.7	4.7 out of 5 st	https://www.	(47K)		Price, produc	https://www.	$9.29	$	9	.	29	(	$0.09	$0.09	Today		FREE delivery	Today 5‚ÄÔI	Add to cart		
Purina ONE N	4.8	4.8 out of 5 st	https://www.	(13.4K)		Price, produc	https://www.	$34.48	$	34	.	48	(	$34.48	$34.48	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart	List: $37.93	$37.93
Purina Friskie	4.7	4.7 out of 5 st	https://www.	(25.7K)		Price, product page		$17.99	$	17	.	99	(	$0.82	$0.82	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart	List: $18.99	$18.99
Purina Cat C	4.8	4.8 out of 5 st	https://www.	(17.7K)		Price, produc	https://www.	$16.99	$	16	.	99	(	$1.31	$1.31	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart	List: $17.99	$17.99
Purina Friskie	4.7	4.7 out of 5 st	https://www.	(16.9K)		Price, produc	https://www.	$18.68	$	18	.	68	(	$0.14	$0.14	Today		FREE delivery	Today 5‚ÄÔI	Add to cart		
Purina Fancy	4.6	4.6 out of 5 st	https://www.	(10.2K)		Price, product page		$30.58	$	30	.	58	(	$0.42	$0.42	Today		FREE delivery	Today 5‚ÄÔI	Add to cart	List: $33.12	$33.12
Purina ONE S	4.8	4.8 out of 5 st	https://www.	(13.2K)		Price, product page		$34.48	$	34	.	48	(	$2.16	$2.16	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart	List: $37.38	$37.38
Purina Friskie	4.7	4.7 out of 5 st	https://www.	(8.1K)		Price, product page		$18.68	$	18	.	68	(	$0.14	$0.14	Today		FREE delivery	Today 5‚ÄÔI	Add to cart	List: $18.98	$18.98
IAMS Proacti	4.7	4.7 out of 5 st	https://www.	(41.7K)		Price, product page		$16.97	$	16	.	97	(	$2.42	$2.42	Today		FREE delivery	Today 5‚ÄÔI	Add to cart	List: $18.98	$18.98
Purina ONE N	4.8	4.8 out of 5 st	https://www.	(8.1K)		Price, produc	https://www.	$17.48	$	17	.	48	(	$2.50	$2.50	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart		
Purina ONE N	4.7	4.7 out of 5 st	https://www.	(5.4K)		Price, produc	https://www.	$33.48	$	33	.	48	(	$33.48	$33.48	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart	List: $35.74	$35.74
Purina Fancy	4.7	4.7 out of 5 st	https://www.	(44.6K)		Price, product page		$23.39	$	23	.	39	(	$0.26	$0.26	Today		FREE delivery	Today 5‚ÄÔI	Add to cart	List: $26.70	$26.70
Purina Friskie	4.7	4.7 out of 5 st	https://www.	(13.1K)		Price, pro																
Hill's Science	4.7	4.7 out of 5 st	https://www.	(14.1K)		Price, pro																
Purina Fancy	4.6	4.6 out of 5 st	https://www.	(6.4K)		Price, pro																
Purina Fancy	4.7	4.7 out of 5 st	https://www.		-557	Price, pro																
Purina Friskie	4.7	4.7 out of 5 st	https://www.	(11.6K)		Price, pro																
Purina Pro Pl	4.7	4.7 out of 5 st	https://www.	(10.8K)		Price, pro																
Purina Kitten	4.8	4.8 out of 5 st	https://www.	(15.9K)		Price, pro																
Purina ONE H	4.8	4.8 out of 5 st	https://www.	(12.3K)		Price, pro																
Purina Friskie	4.7	4.7 out of 5 st	https://www.	(33.5K)		Price, pro																
Purina ONE N	4.7	4.7 out of 5 st	https://www.	(12.4K)		Price, pro																
Hill's Science	4.7	4.7 out of 5 st	https://www.	(5K)		Price, pro																
Hill's Science	4.7	4.7 out of 5 st	https://www.	(4.8K)		Price, pro																
Fancy Feast	4.5	4.5 out of 5 st	https://www.		-796	Price, pro																
IAMS Proacti	4.8	4.8 out of 5 st	https://www.	(23.3K)		Price, pro																
Purina Fancy	4.7	4.7 out of 5 st	https://www.	(2K)		Price, product page		$21.12	$	21	.	12	(	$0.29	$0.29	Today		FREE delivery	Today 5‚ÄÔI	Add to cart		
Solid Gold Hi	4.6	4.6 out of 5 st	https://www.	(8.9K)		Price, product page		$43.69	$	43	.	69	(	$3.64	$3.64	Today		FREE delivery	Today 5‚ÄÔI	Add to cart	List: $45.99	$45.99
Purina ONE N	4.5	4.5 out of 5 st	https://www.	(6.9K)		Price, produc	https://www.	$28.36	$	28	.	36	(	$0.39	$0.39	Tomorrow		FREE delivery	Tomorrow, Ai	Add to cart		

(overlay text: "but we don't have any more ChatGPT power :(")

An exercise

Let's do this ourselves.

Scraping web sites and unstru

Table of Contents

- Useful Links

- Intro

- Using NotebookLM

- Using Easy Scraper

- Running code in Colab

- Exercises

Download files: 📦 scraping-web-data.zip

An exercise (Part 2)

Let's analyze the results

Scraping web sites and unstru

Table of Contents

Download files: 📦 scraping-web-data.zip

Okay we know how to do everything now

We are geniuses????

Exercises

Let's do even more.

Scraping web sites and unstr

Table of Contents

- Useful Links

- Intro

- Using NotebookLM

- Using Easy Scraper

- Running code in Colab

- Exercises

Download files: 📦 scraping-web-data.zip