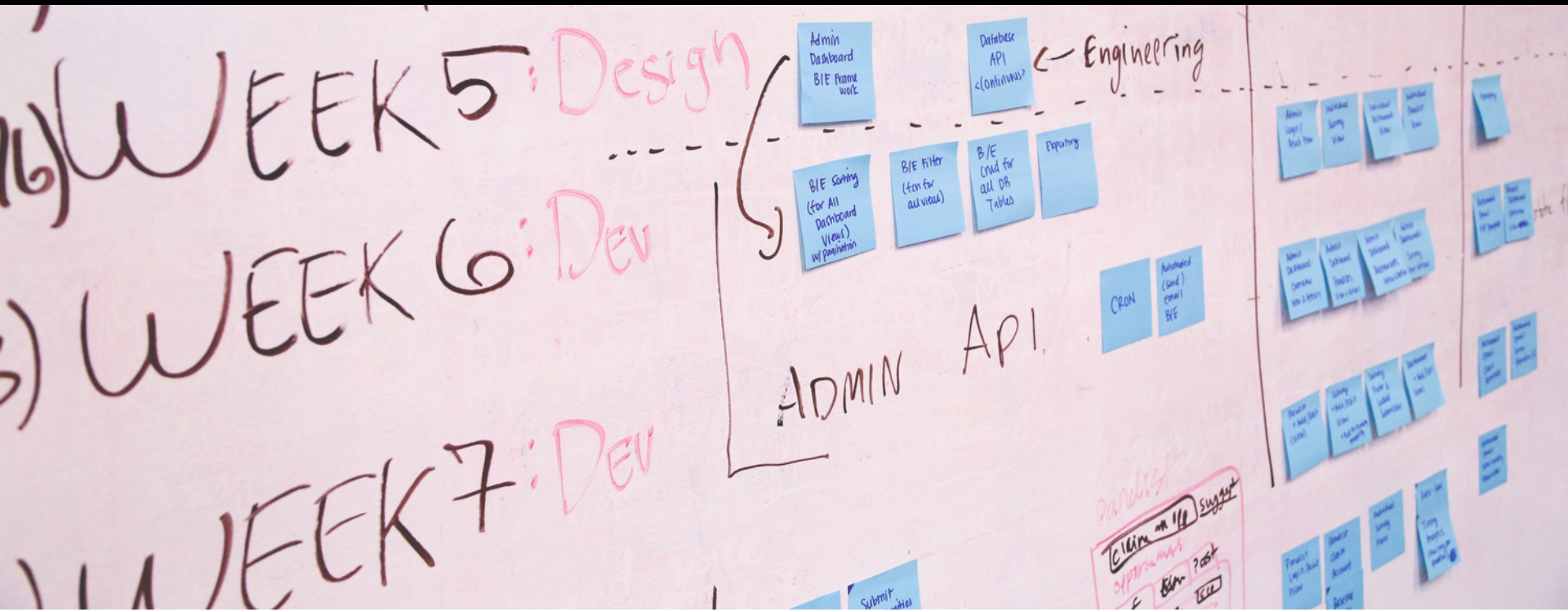


LEARNING GARDEN MAKER WORKSHOP:

WEB SCRAPING IN PARSEHUB FOR NON-CODER

Bernard Suen
Center for Entrepreneurship
Chinese University of Hong Kong

Today's agenda.

1. What is web scraping ?
2. How does it work ?
3. The CRISP-DM model for handling data.
4. Demo of scraping in Python.
5. Scraping in ParseHub.
6. Beyond scraping: Legal & ethical considerations.

What is web scraping?

Web scraping is a process of fetching web pages from a website and extracting specific data from it.

- ✓ Scrap stock prices
- ✓ Scrap product information
- ✓ Scrap race scores
- ✓ Scrap market trend data
- ✓ Scrap demographics and census data
- ✓ Social media and web articles, etc.

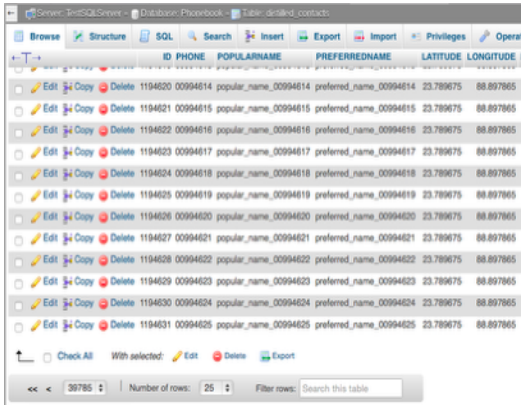
How does it work?

Data

Web Publishing

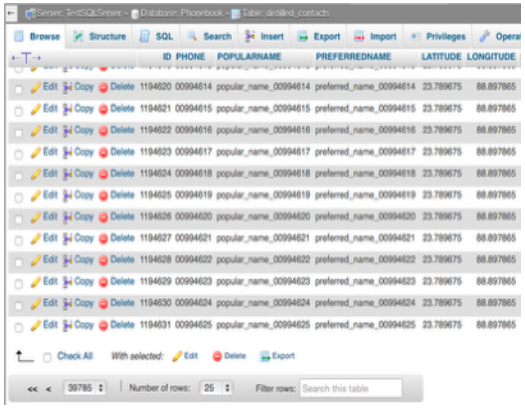
Web

Dynamic Web Building

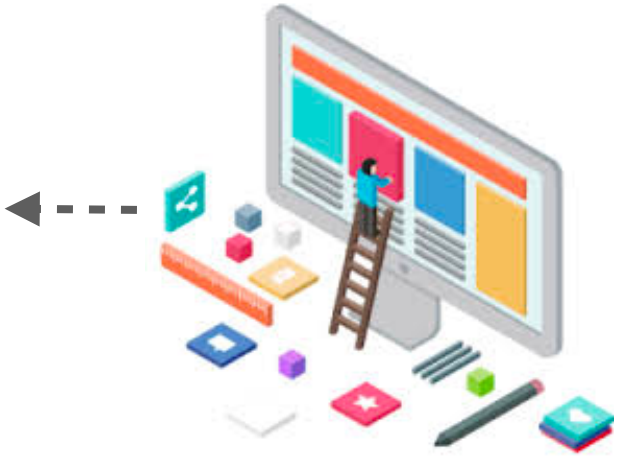


Static Web Building

Data

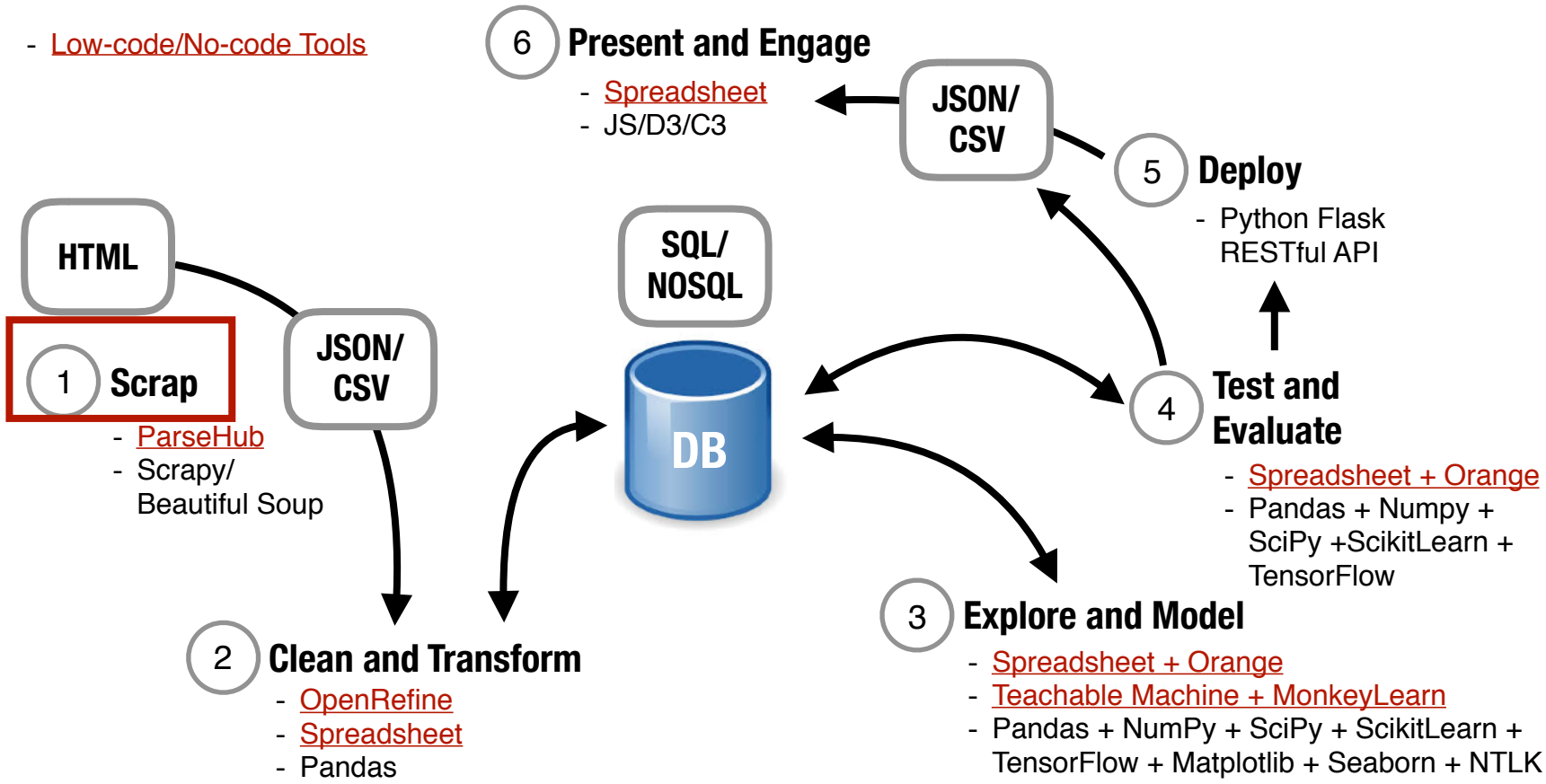


Web Scraping



Web

Scraping is part of a bigger picture.


- Low-code/No-code Tools
6 Present and Engage
- Spreadsheet
- JS/D3/C3
JSON/ CSV
5 Deploy
- Python Flask RESTful API
HTML
SQL/ NOSQL
1 Scrap
- ParseHub
- Scrapy/ Beautiful Soup
JSON/ CSV
DB
4 Test and Evaluate
- Spreadsheet + Orange
- Pandas + Numpy + SciPy +ScikitLearn + TensorFlow
3 Explore and Model
- Spreadsheet + Orange
- Teachable Machine + MonkeyLearn
- Pandas + NumPy + SciPy + ScikitLearn + TensorFlow + Matplotlib + Seaborn + NTLK
2 Clean and Transform
- OpenRefine
- Spreadsheet
- Pandas

Scraping Demo Using Codes and No-code Tools

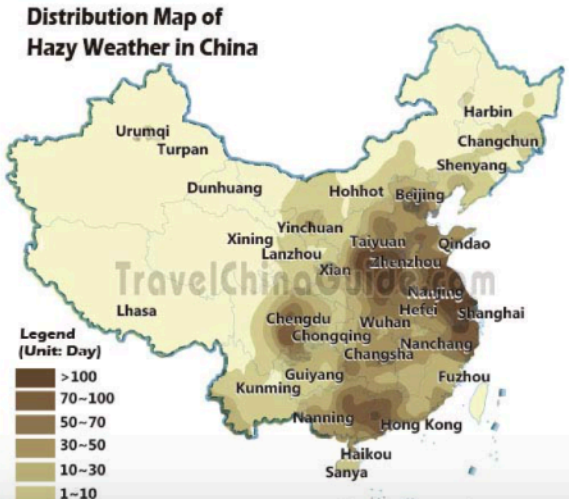
Scraping with Python Beautiful Soup

China Weather
- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December
- Climate Change
- Air Pollution

China Air Pollution

As one of the most popular tourist destinations in the world, China has abundant tourism resources, the vast majority of which are of great renown. However, the number of tourists visiting China declined from 2013, mostly due to its increasingly serious air pollution. Many cities have been adversely affected by hazy weather since the beginning of 2013. This is especially true of central and eastern China's cities such as Tangshan, Jinan, Zhengzhou, and Xi'an for example.

It is known to all that the rapid development of China's economy in the past 30 years resulted in environmental degradation. Nevertheless, the hazy weather records of China in 2013 attracted global attention. It is a pity that as a consequence some foreign residents choose to give up their jobs and life in China.



https://www.travelchinaguide.com/climate/air-pollution.htm

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Presenting China Pollution Figures Using Beautiful Soup, Pandas and Matplotlib

In [2]:

```
from bs4 import BeautifulSoup
import requests
import csv
import pandas as pd
import matplotlib.pyplot as plt

# Fetch URL
html_page = requests.get('https://www.travelchinaguide.com/climate/air-pollution.htm')
# Obtain the entire HTML page
soup = BeautifulSoup(html_page.content,'html.parser')
# Find all the HTML tables
tables = soup.find_all(class_="c_tableX")
# Access the second HTML table (i.e. tables[1] instrad of tables[0]) that contains Air Quality information
# with both table header and table data
table = tables[1] # assign 2nd table in the table list to variable called "table"
table_header = table.find_all('th') # extract table header based on the 'th' tag
# print table header tags and texts
# print(table_header)
header = []
data = []
for th in table.find_all('th'):
    # print header text
    # print(th.text)
    header.append(th.text)
# print list of headers
# print(header)
all_rows = table.find_all("tr") # extract all the table rows
# Enumerate all the rows to extract needed data
# Obtain both index and values using the enumerate function starting at 2nd row
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

```
98 plt.show()
```

```
Air Quality Ranking

    Rank          City Province belongs to  AQI    Air Quality Level PM 2.5  \
0      1         Sanya               Hainan   22            Excellent     10
1      2       Lijiang              Yunnan   24            Excellent     12
2      3          Dali              Yunnan   24            Excellent     15
3      4      Changsha               Hunan   25            Excellent     17
4      5        Haikou              Hainan   28            Excellent     15
5      6       Qingdao            Shandong   29            Excellent     19
6      7       Kunming              Yunnan   33            Excellent     23
7      8         Lhasa               Tibet   34            Excellent     15
8      9        Dalian            Liaoning   43            Excellent     20
9     10     Guangzhou           Guangdong   44            Excellent     30
10    11       Guiyang             Guizhou   45            Excellent     31
11    12   Zhangjiakou               Hebei   49            Excellent     31
12    13      Shenzhen           Guangdong   55                 Good     39
13    14   Zhangjiajie               Hunan   63                 Good     45
14    15        Fuzhou              Fujian   65                 Good     47
15    16       Nanning             Guangxi   68                 Good     49
```

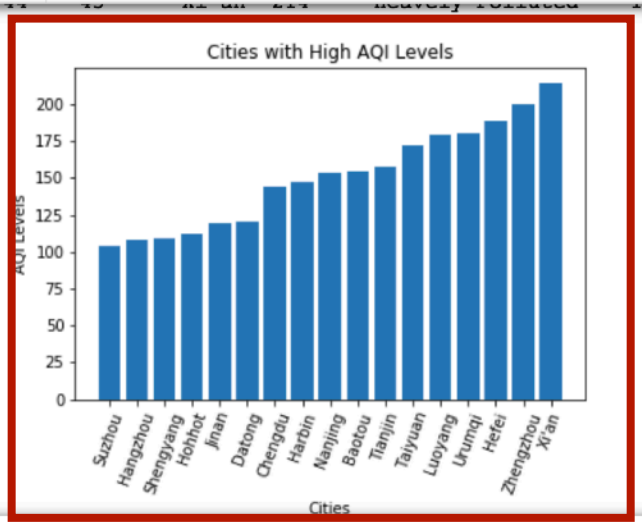
In [3]:

```
1 from matplotlib import pyplot as plt
2 # aqi_lvl = pc['AQI'].tolist()
3 # pm_2_5 = pc['PM 2.5'].tolist()
4 ax = plt.subplot()
5 plt.plot(aqi_lvl, pm_2_5, 'o')
6 plt.ylabel('PM 2.5')
7 plt.xlabel('AQI Levels')
8 plt.show()
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

```
98 plt.show()
```



In [3]:

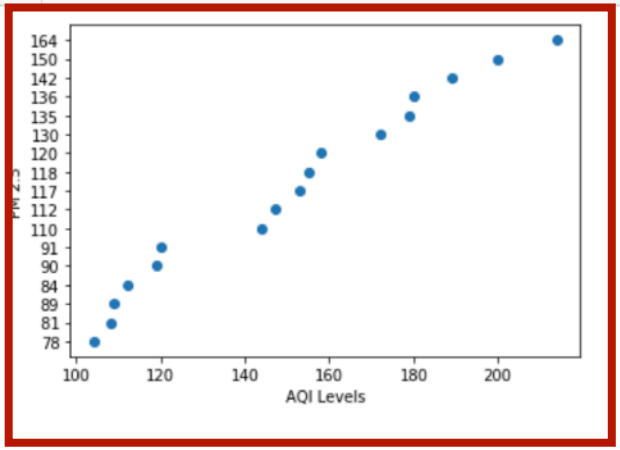
```
1 from matplotlib import pyplot as plt
2 # aqi_lvl = pc['AQI'].tolist()
3 # pm_2_5 = pc['PM 2.5'].tolist()
4 ax = plt.subplot()
5 plt.plot(aqi_lvl, pm_2_5, 'o')
6 plt.ylabel('PM 2.5')
7 plt.xlabel('AQI Levels')
8 plt.show()
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [3]:

```
from matplotlib import pyplot as plt
# aqi_lvl = pc['AQI'].tolist()
# pm_2_5 = pc['PM 2.5'].tolist()
ax = plt.subplot()
plt.plot(aqi_lvl, pm_2_5, 'o')
plt.ylabel('PM 2.5')
plt.xlabel('AQI Levels')
plt.show()
```



Cleaning Data with Python Pandas

Most Basic Data Cleaning Tasks

1. Remove blanks
2. Impute missing entries (need expertise to judge usage instead of relying on statistics alone)
3. Remove duplicates
4. Resolve inconsistent entries
5. Transform incorrect data (e.g. wrong data types such as text instead of numeric and calculations)
6. Rename, combine and split columns for making further data presentation and processing easier

Scraping NBA Data (Original Source:Scraping and Cleaning the NBA Draft by Savvas Tjortjoglou)

In [9]:

```
from urllib.request import urlopen
from bs4 import BeautifulSoup
import pandas as pd
import csv

url = "http://www.basketball-reference.com/draft/NBA_2014.html"
html = urlopen(url)
soup = BeautifulSoup(html,'html.parser')
column_headers = []
for th in soup.findAll('tr',limit=2)[1].findAll('th'):
    column_headers.append(th.getText())
data_rows = soup.findAll('tr')[2:]  # skip the first 2 header rows
player_data = []
for i in range(len(data_rows)):  # for each table row
    player_row = []
    # for each table data element from each table row
    for th in data_rows[i].findAll('th',limit=1):
        player_row.append(th.getText())
    for td in data_rows[i].findAll('td'):
        # get the text content and append to the player_row
        player_row.append(td.getText())
    # then append each pick/player to the player_data matrix
    player_data.append(player_row)
df = pd.DataFrame(player_data, columns=column_headers)
df.to_csv('projects/nba.csv', sep='\t', encoding='utf-8')
# print(df.shape)
# pd.set_option('display.max_rows', 62)
```

Draft History | Draft Years ▾

			Round 2			Totals					Shooting			Per Game					
Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%	3P%	FT%	MP	PTS	TRB	AST	WS	WS
31	31	MIL	Damien Inglis		1	20	156	36	31	10	.351	.231	.875	7.8	1.8	1.6	0.5	-0.1	
32	32	PHI	K.J. McDaniels	Clemson	3	148	2092	782	324	95	.412	.290	.776	14.1	5.3	2.2	0.6	1.3	
33	33	CLE	Joe Harris	Virginia	6	331	8037	3452	1040	535	.479	.426	.779	24.3	10.4	3.1	1.6	14.7	
34	34	NYK	Cleanthony Early	Wichita State	2	56	801	241	123	42	.346	.263	.750	14.3	4.3	2.2	0.8	-0.6	
35	35	UTA	Jarnell Stokes	Tennessee	3	28	151	67	40	7	.581		.531	5.4	2.4	1.4	0.3	0.6	
36	36	MIL	Johnny O'Bryant	LSU	4	147	1684	509	352	69	.402	.360	.663	11.5	3.5	2.4	0.5	0.2	
37	37	TOR	DeAndre Daniels	UConn															
38	38	DET	Spencer Dinwiddie	Colorado	6	317	8162	4103	875	1578	.410	.318	.793	25.7	12.9	2.8	5.0	18.3	
39	39	PHI	Jerami Grant	Syracuse	6	454	11125	4220	1748	484	.465	.347	.669	24.5	9.3	3.9	1.1	22.6	
40	40	MIN	Glenn Robinson	Michigan	6	281	4930	1683	731	218	.459	.373	.768	17.5	6.0	2.6	0.8	7.6	
41	41	DEN	Nikola Jokić		5	381	11054	6462	3657	2098	.524	.338	.826	29.0	17.0	9.6	5.5	48.7	
42	42	HOU	Nick Johnson	Arizona	1	28	262	74	39	11	.347	.238	.680	9.4	2.6	1.4	0.4	-0.1	
43	43	ATL	Edy Tavares		2	13	101	33	32	4	.625		.273	7.8	2.5	2.5	0.3	0.2	
44	44	MIN	Markel Brown	Oklahoma State	3	113	1794	584	234	132	.380	.295	.781	15.9	5.2	2.1	1.2	1.1	
45	45	CHH	Dwight Powell	Stanford	6	371	6956	2867	1711	364	.563	.293	.739	18.7	7.7	4.6	1.0	26.0	

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

23	24	24	CHH	Shabazz Napier	UConn	6	340	5876	2382	645	...	.346	.815	17.3	7.0	1.9	2.5	8.3	.068	-0.6	2.1
25	26	26	MIA	P.J. Hairston	UNC	2	111	2000	664	266	...	.295	.810	18.0	6.0	2.4	0.5	0.9	.021	-3.9	-1.0
27	28	28	LAC	C.J. Wilcox	Washington	3	66	376	132	31	...	.333	.813	5.7	2.0	0.5	0.5	0.1	.007	-4.3	-0.2
28	29	29	OKC	Josh Huestis	Stanford	3	76	1068	187	180	...	.312	.240	14.1	2.5	2.4	0.3	0.7	.030	-3.5	-0.4
29	30	30	SAS	Kyle Anderson	UCLA	6	359	7022	1936	1428	...	.307	.698	19.6	5.4	4.0	2.0	17.6	.120	1.5	6.2
30		None	None	None	None	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None	None
31	Rk	None	None	None	None	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None	None
33	32	32	PHI	K.J. McDaniels	Clemson	3	148	2092	782	324	...	.290	.776	14.1	5.3	2.2	0.6	1.3	.030	-2.8	-0.4
34	33	33	CLE	Joe Harris	Virginia	6	325	7860	3332	1014	...	.423	.786	24.2	10.3	3.1	1.6	14.2	.087	-0.8	2.3
35	34	34	NYK	Cleanthony Early	Wichita State	2	56	801	241	123	...	.263	.750	14.3	4.3	2.2	0.8	-0.6	-.039	-5.9	-0.8
36	35	35	UTA	Jarnell Stokes	Tennessee	3	28	151	67	40	...		.531	5.4	2.4	1.4	0.3	0.6	.202	0.1	0.1
37	36	36	MIL	Johnny O'Bryant	LSU	4	147	1684	509	352	...	.360	.663	11.5	3.5	2.4	0.5	0.2	.006	-6.2	-1.8
38	37	37	TOR	DeAndre Daniels	UConn						...										
39	38	38	DET	Spencer Dinwiddie	Colorado	6	317	8162	4103	875	...	.318	.793	25.7	12.9	2.8	5.0	18.6	.109	0.3	4.8
40	39	39	PHI	Jerami Grant	Syracuse	6	447	10909	4111	1727	...	.348	.666	24.4	9.2	3.9	1.1	22.3	.098	-0.9	3.1
41	40	40	MIN	Glenn Robinson	Michigan	6	279	4887	1664	721	...	.372	.769	17.5	6.0	2.6	0.8	7.5	.074	-1.5	0.6
43	42	42	HOU	Nick Johnson	Arizona	1	28	262	74	39	...	.238	.680	9.4	2.6	1.4	0.4	-0.1	-.020	-6.5	-0.3
45	44	44	MIN	Markel Brown	Oklahoma State	3	113	1794	584	234	...	.295	.781	15.9	5.2	2.1	1.2	1.1	.029	-2.9	-0.4
46	45	45	CHH	Dwight Powell	Stanford	6	371	6956	2867	1711	...	.293	.739	18.7	7.7	4.6	1.0	26.1	.180	1.5	6.1
47	46	46	WAS	Jordan Clarkson	Missouri	6	445	12054	6573	1404	...	.341	.818	27.1	14.8	3.2	2.6	14.0	.056	-0.7	3.9

In [41]:
```
1 df.dropna(thresh=4)
```

Out[41]:

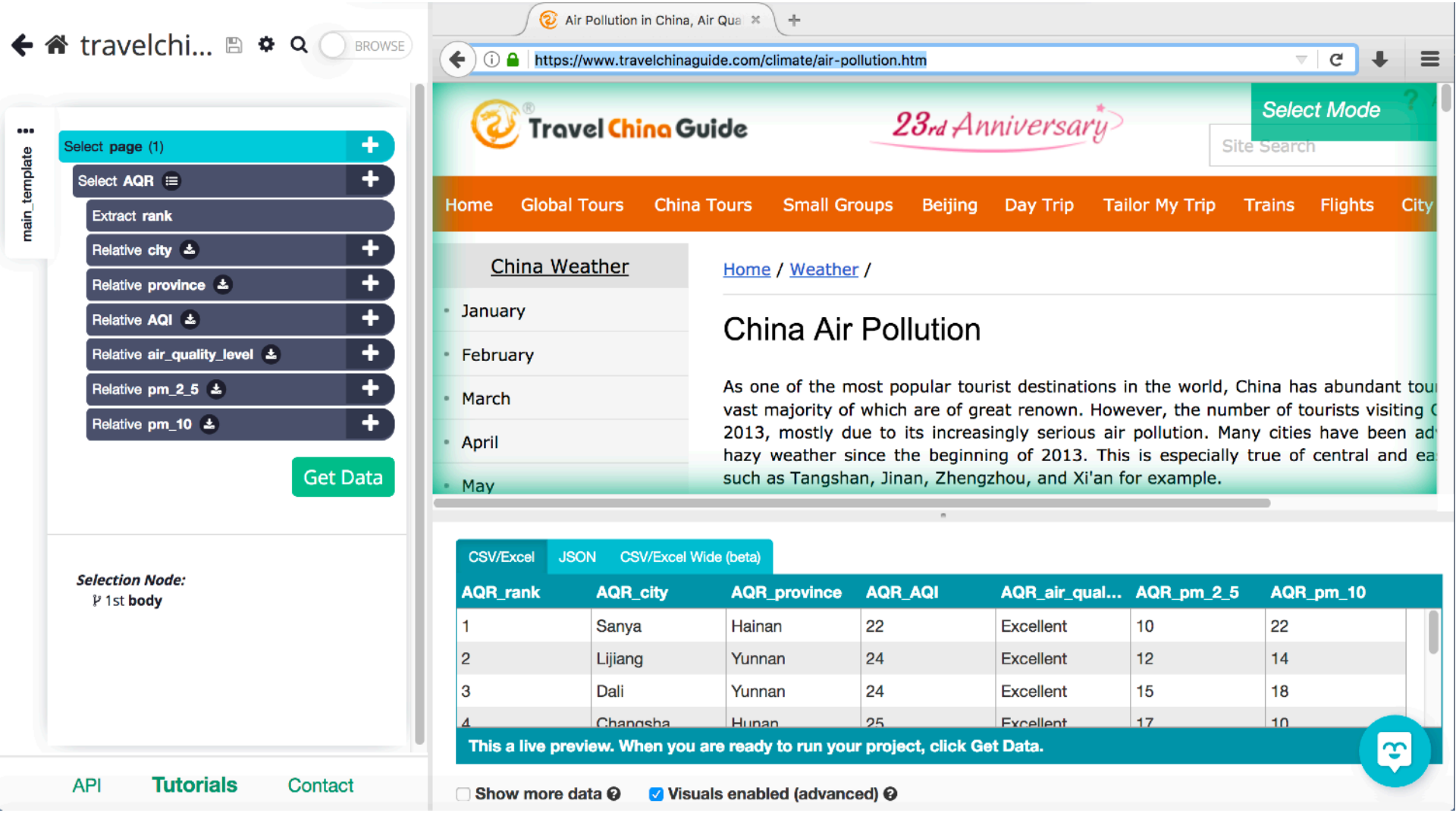
	Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	...	3P%	FT%	MP	PTS	TRB	AST	WS	WS_per_48	BPM	VORP
0	1	1	CLE	Andrew Wiggins	Kansas	6	454.0	16242	8943	1977	...	.332	.732	35.8	19.7	4.4	2.3	15.0	.044	-1.9	0.2
1	2	2	MIL	Jabari Parker	Duke	6	285.0	8258	4255	1633	...	.324	.740	29.0	14.9	5.7	2.1	13.2	.077	-0.8	2.4
2	3	3	PHI	Joel Embiid	Kansas	4	209.0	6358	5005	2397	...	.319	.793	30.4	23.9	11.5	3.1	22.8	.172	4.3	10.1
3	4	4	ORL	Aaron Gordon	Arizona	6	403.0	11517	5143	2587	...	.319	.701	28.6	12.8	6.4	2.4	21.7	.090	0.0	5.7
5	6	6	BOS	Marcus Smart	Oklahoma State	6	401.0	11607	3950	1430	...	.318	.775	28.9	9.9	3.6	4.1	20.7	.086	-0.3	5.0
6	7	7	LAL	Julius Randle	Kentucky	6	375.0	10934	6032	3375	...	.295	.724	29.2	16.1	9.0	2.8	20.5	.090	-0.5	4.1
7	8	8	SAC	Nik Stauskas	Michigan	5	335.0	6662	2272	688	...	.353	.814	19.9	6.8	2.1	1.5	3.8	.028	-3.0	-1.7
8	9	9	CHH	Noah Vonleh	Indiana	6	335.0	5672	1660	1741	...	.310	.691	16.9	5.0	5.2	0.8	9.3	.079	-2.4	-0.5
9	10	10	PHI	Elfrid Payton	LA-Lafayette	6	387.0	11350	4247	1703	...	.289	.625	29.3	11.0	4.4	6.6	14.9	.063	-0.5	4.3
10	11	11	DEN	Doug McDermott	Creighton	6	410.0	8170	3369	900	...	.412	.825	19.9	8.2	2.2	0.8	13.9	.082	-2.0	-0.1
12	13	13	MIN	Zach LaVine	UCLA	6	353.0	10857	6240	1279	...	.375	.819	30.8	17.7	3.6	3.6	12.0	.053	-0.6	3.8
13	14	14	PHO	T.J. Warren	NC State	6	328.0	9436	5080	1358	...	.361	.779	28.8	15.5	4.1	1.2	19.5	.099	-0.2	4.3
14	15	15	ATL	Adreian Payne	Michigan State	4	107.0	1403	429	315	...	.254	.680	13.1	4.0	2.9	0.6	-0.6	-.019	-5.9	-1.4
16	17	17	BOS	James Young	Kentucky	4	95.0	812	219	96	...	.277	.563	8.5	2.3	1.0	0.3	0.8	.046	-3.4	-0.3
17	18	18	PHO	Tyler Ennis	Syracuse	4	186.0	2336	779	250	...	.317	.768	12.6	4.2	1.3	1.9	1.3	.026	-3.4	-0.8
18	19	19	CHI	Gary Harris	Michigan State	6	368.0	10663	4459	961	...	.360	.803	29.0	12.1	2.6	2.1	17.9	.081	-0.9	2.9
20	21	21	OKC	Mitch McGary	Michigan	2	52.0	557	227	183	...	.000	.580	10.7	4.4	3.5	0.3	1.3	.108	-2.6	-0.1
21	22	22	MEM	Jordan Adams	UCLA	2	32.0	263	101	30	...	.385	.607	8.2	3.2	0.9	0.6	0.4	.070	1.7	0.2
22	23	23	UTA	Rodney Hood	Duke	6	341.0	9326	4247	980	...	.372	.839	27.3	12.5	2.9	1.9	17.1	.088	-0.5	3.4
23	24	24	CHH	Shabazz Napier	UConn	6	345.0	5986	2433	657	...	.345	.815	17.4	7.1	1.9	2.5	8.1	.065	-0.7	2.0

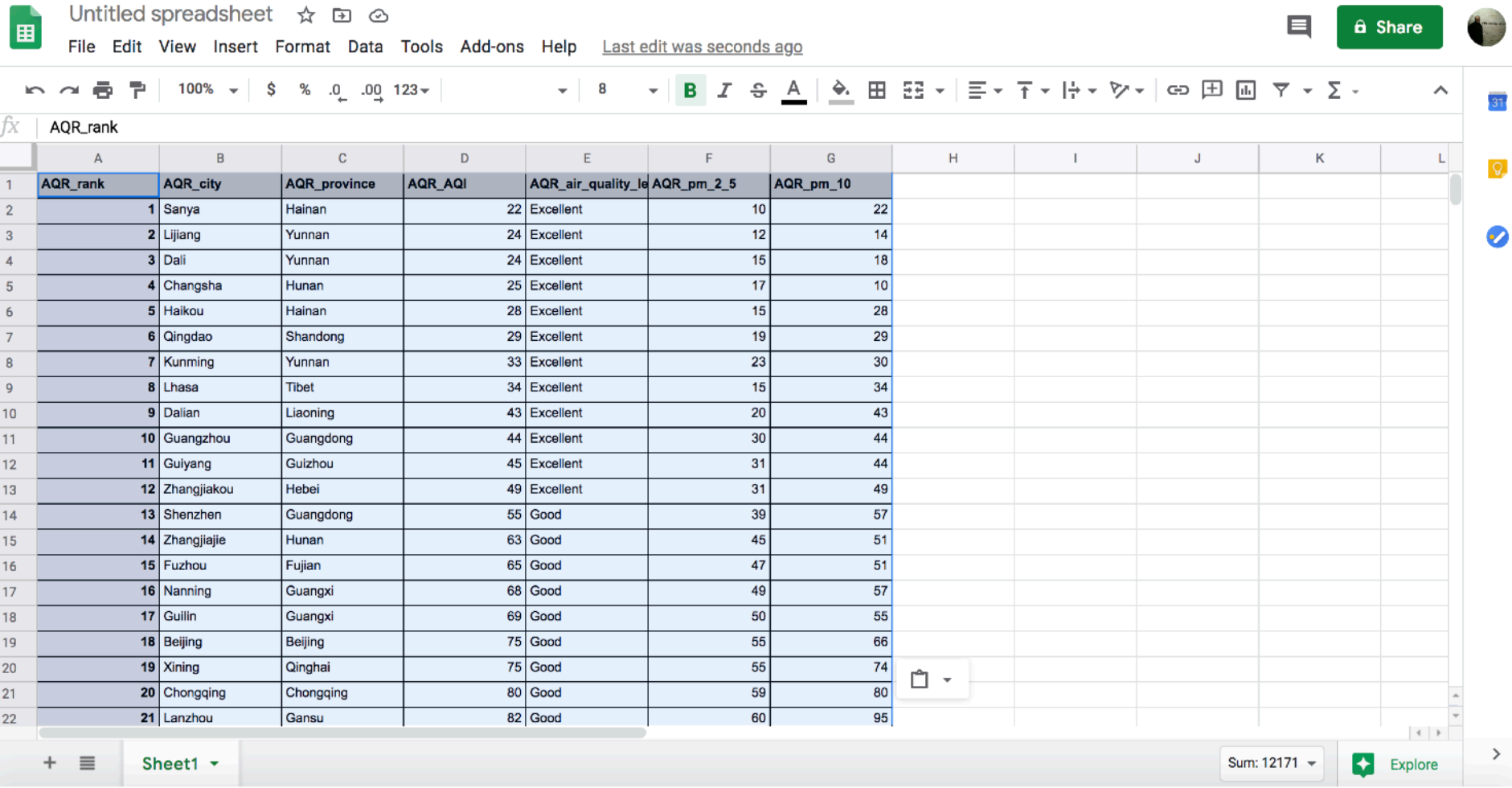
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

23	24	24	CHH	Shabazz Napier	UConn	6	345.0	5986	2433	657	...	.345	.815	17.4	7.1	1.9	2.5	8.1	.065	-0.7	2.0
25	26	26	MIA	P.J. Hairston	UNC	2	111.0	2000	664	266	...	.295	.810	18.0	6.0	2.4	0.5	0.9	.021	-3.9	-1.0
27	28	28	LAC	C.J. Wilcox	Washington	3	66.0	376	132	31	...	.333	.813	5.7	2.0	0.5	0.5	0.1	.007	-4.3	-0.2
28	29	29	OKC	Josh Huestis	Stanford	3	76.0	1068	187	180	...	.312	.240	14.1	2.5	2.4	0.3	0.7	.030	-3.5	-0.4
29	30	30	SAS	Kyle Anderson	UCLA	6	367.0	7212	1991	1456	...	.311	.699	19.7	5.4	4.0	2.0	17.9	.119	1.4	6.3
33	32	32	PHI	K.J. McDaniels	Clemson	3	148.0	2092	782	324	...	.290	.776	14.1	5.3	2.2	0.6	1.3	.030	-2.8	-0.4
34	33	33	CLE	Joe Harris	Virginia	6	331.0	8037	3452	1040	...	.426	.779	24.3	10.4	3.1	1.6	14.7	.088	-0.7	2.5
35	34	34	NYK	Cleanthony Early	Wichita State	2	56.0	801	241	123	...	.263	.750	14.3	4.3	2.2	0.8	-0.6	-.039	-5.9	-0.8
36	35	35	UTA	Jarnell Stokes	Tennessee	3	28.0	151	67	40	...	NaN	.531	5.4	2.4	1.4	0.3	0.6	.202	0.1	0.1
37	36	36	MIL	Johnny O'Bryant	LSU	4	147.0	1684	509	352	...	.360	.663	11.5	3.5	2.4	0.5	0.2	.006	-6.2	-1.8
39	38	38	DET	Spencer Dinwiddie	Colorado	6	317.0	8162	4103	875	...	.318	.793	25.7	12.9	2.8	5.0	18.3	.108	0.4	4.8
40	39	39	PHI	Jerami Grant	Syracuse	6	454.0	11125	4220	1748	...	.347	.669	24.5	9.3	3.9	1.1	22.6	.098	-0.9	3.1
41	40	40	MIN	Glenn Robinson	Michigan	6	281.0	4930	1683	731	...	.373	.768	17.5	6.0	2.6	0.8	7.6	.074	-1.5	0.6
43	42	42	HOU	Nick Johnson	Arizona	1	28.0	262	74	39	...	.238	.680	9.4	2.6	1.4	0.4	-0.1	-.020	-6.5	-0.3
45	44	44	MIN	Markel Brown	Oklahoma State	3	113.0	1794	584	234	...	.295	.781	15.9	5.2	2.1	1.2	1.1	.029	-2.9	-0.4
46	45	45	CHH	Dwight Powell	Stanford	6	371.0	6956	2867	1711	...	.293	.739	18.7	7.7	4.6	1.0	26.0	.179	1.4	6.1
47	46	46	WAS	Jordan Clarkson	Missouri	6	453.0	12232	6699	1424	...	.342	.818	27.0	14.8	3.1	2.6	13.9	.055	-0.7	3.9
48	47	47	PHI	Russ Smith	Louisville	2	27.0	131	53	15	...	.188	.769	4.9	2.0	0.6	0.7	-0.2	-.068	-5.1	-0.1
49	48	48	MIL	Lamar Patterson	Pitt	2	40.0	435	93	57	...	.236	.720	10.9	2.3	1.4	1.1	0.1	.006	-5.7	-0.4
50	49	49	CHI	Cameron Bairstow	New Mexico	2	36.0	167	44	36	...	.200	.846	4.6	1.2	1.0	0.2	-0.1	-.021	-8.1	-0.3
56	55	55	MIA	Semaj Christon	Xavier	1	64.0	973	183	87	...	.190	.548	15.2	2.9	1.4	2.0	0.1	.006	-5.1	-0.8
57	56	56	DEN	Devyn Marble	Iowa	2	44.0	457	97	69	...	.222	.375	10.4	2.2	1.6	0.7	-0.3	-.028	-5.5	-0.4
59	58	58	SAS	Jordan McRae	Tennessee	4	123.0	1696	846	225	...	.355	.772	13.8	6.9	1.8	1.4	1.7	.048	-2.1	0.0

Scraping with ParseHub


travelchi...
BROWSE
main_template
Select page (1)
Select AQR
Extract rank
Relative city
Relative province
Relative AQI
Relative air_quality_level
Relative pm_2_5
Relative pm_10
Get Data
Selection Node:
1st body
API
Tutorials
Contact
Air Pollution in China, Air Qua
https://www.travelchinaguide.com/climate/air-pollution.htm
Travel China Guide
23rd Anniversary
Select Mode
Site Search
Home
Global Tours
China Tours
Small Groups
Beijing
Day Trip
Tailor My Trip
Trains
Flights
City
China Weather
January
February
March
April
May
Home / Weather /
China Air Pollution
As one of the most popular tourist destinations in the world, China has abundant tou
vast majority of which are of great renown. However, the number of tourists visiting
2013, mostly due to its increasingly serious air pollution. Many cities have been ad
hazy weather since the beginning of 2013. This is especially true of central and ea
such as Tangshan, Jinan, Zhengzhou, and Xi'an for example.
CSV/Excel
JSON
CSV/Excel Wide (beta)
AQR_rank
AQR_city
AQR_province
AQR_AQI
AQR_air_qual...
AQR_pm_2_5
AQR_pm_10
1
Sanya
Hainan
22
Excellent
10
22
2
Lijiang
Yunnan
24
Excellent
12
14
3
Dali
Yunnan
24
Excellent
15
18
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)



AQR_rank	AQR_city	AQR_province	AQR_AQI	AQR_air_quality_le	AQR_pm_2_5	AQR_pm_10
1	Sanya	Hainan	22	Excellent	10	22
2	Lijiang	Yunnan	24	Excellent	12	14
3	Dali	Yunnan	24	Excellent	15	18
4	Changsha	Hunan	25	Excellent	17	10
5	Haikou	Hainan	28	Excellent	15	28
6	Qingdao	Shandong	29	Excellent	19	29
7	Kunming	Yunnan	33	Excellent	23	30
8	Lhasa	Tibet	34	Excellent	15	34
9	Dalian	Liaoning	43	Excellent	20	43
10	Guangzhou	Guangdong	44	Excellent	30	44
11	Guiyang	Guizhou	45	Excellent	31	44
12	Zhangjiakou	Hebei	49	Excellent	31	49
13	Shenzhen	Guangdong	55	Good	39	57
14	Zhangjiajie	Hunan	63	Good	45	51
15	Fuzhou	Fujian	65	Good	47	51
16	Nanning	Guangxi	68	Good	49	57
17	Guilin	Guangxi	69	Good	50	55
18	Beijing	Beijing	75	Good	55	66
19	Xining	Qinghai	75	Good	55	74
20	Chongqing	Chongqing	80	Good	59	80
21	Lanzhou	Gansu	82	Good	60	95

Untitled spreadsheet

File Edit View Insert Format Data Tools Add-ons Help Last edit was seconds ago

Share

100% $ % .0 .00 123 Default (Ari... 10

	A	B	D
1	AQR_rank	AQR_city	AQR_AQI
2	1	Sanya	22
3	2	Lijiang	24
4	3	Dali	24
5	4	Changsha	25
6	5	Haikou	28
7	6	Qingdao	29
8	7	Kunming	33
9	8	Lhasa	34
10	9	Dalian	43
11	10	Guangzhou	44
12	11	Guiyang	45
13	12	Zhangjiakou	49
14	13	Shenzhen	55
15	14	Zhangjiajie	63
16	15	Fuzhou	65
17	16	Nanning	68
18	17	Guilin	69
19	18	Beijing	75
20	19	Xining	75
21	20	Chongqing	80
22	21	Lanzhou	82

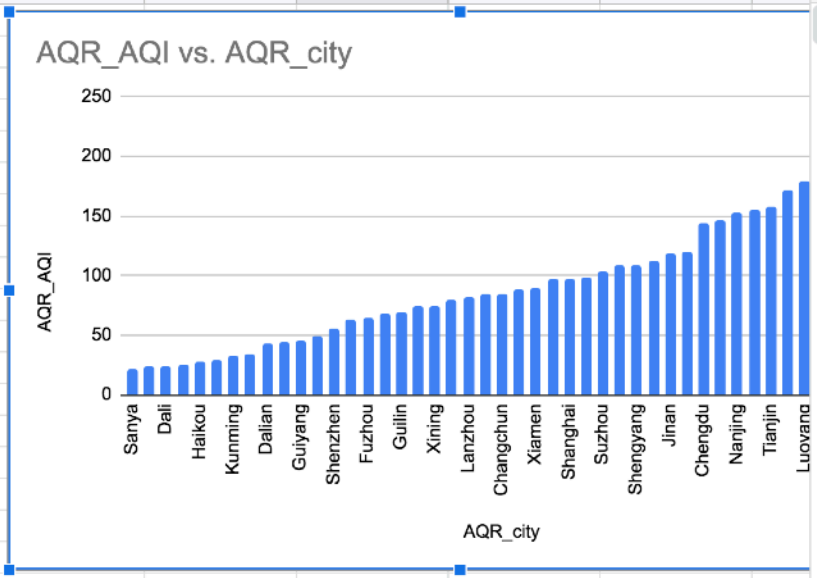


Chart editor

Setup Customize

Chart type

Column chart

Stacking

None

Data range

B1:D46

X-axis

AQR_city

Aggregate

Series

123 AQR_AQI

Sheet1

Explore

Untitled spreadsheet

Saved to Drive

File Edit View Insert Format Data Tools Add-ons Help Last edit was seconds ago

Share

100% $ % .0 .00 123 Default (Ari... 10

	B	D	F
1	AQR_city	AQR_AQI	AQR_pm_2_5
2	Sanya	22	10
3	Lijiang	24	12
4	Dali	24	15
5	Changsha	25	17
6	Haikou	28	15
7	Qingdao	29	19
8	Kunming	33	23
9	Lhasa	34	15
10	Dalian	43	20
11	Guangzhou	44	30
12	Guiyang	45	31
13	Zhangjiakou	49	31
14	Shenzhen	55	39
15	Zhangjiajie	63	45
16	Fuzhou	65	47
17	Nanning	68	49
18	Guilin	69	50
19	Beijing	75	55
20	Xining	75	55
21	Chongqing	80	59
22	Lanzhou	82	60

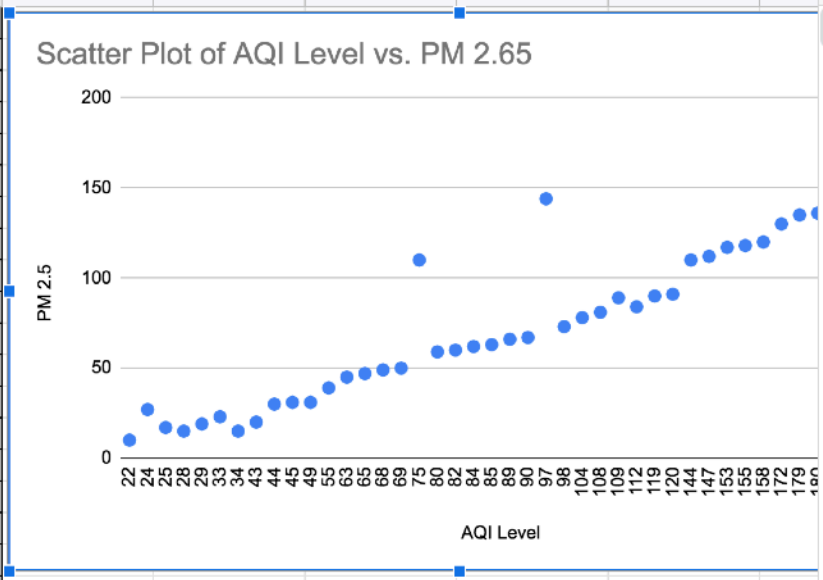


Chart editor

Setup Customize

Chart type

Scatter chart

Data range

D1:F46

X-axis

D1:D46

Aggregate

Series Sum

F1:F46 Sum

Add Series

Include hidden / filtered data

Sheet1

Explore

Untitled spreadsheet

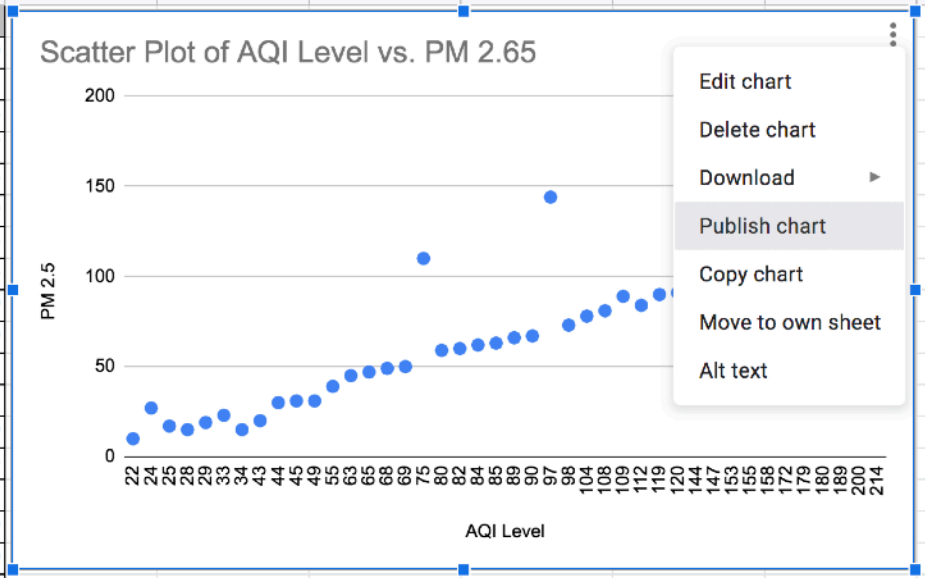
File Edit View Insert Format Data Tools Add-ons Help Last edit was 3 hours ago

Share

100% $ % .0 .00 123 Helvetica ... 8

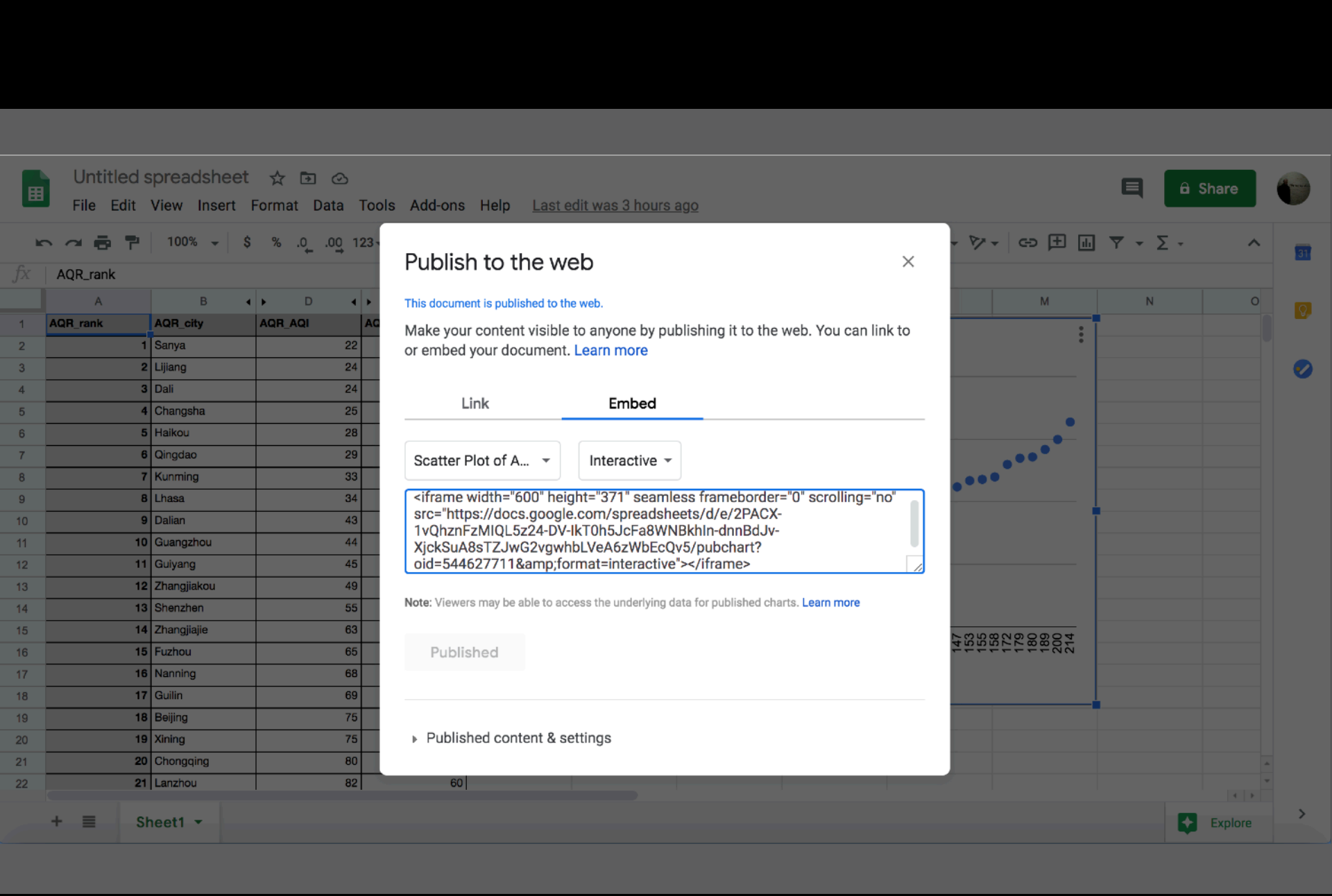
fx AQR_rank

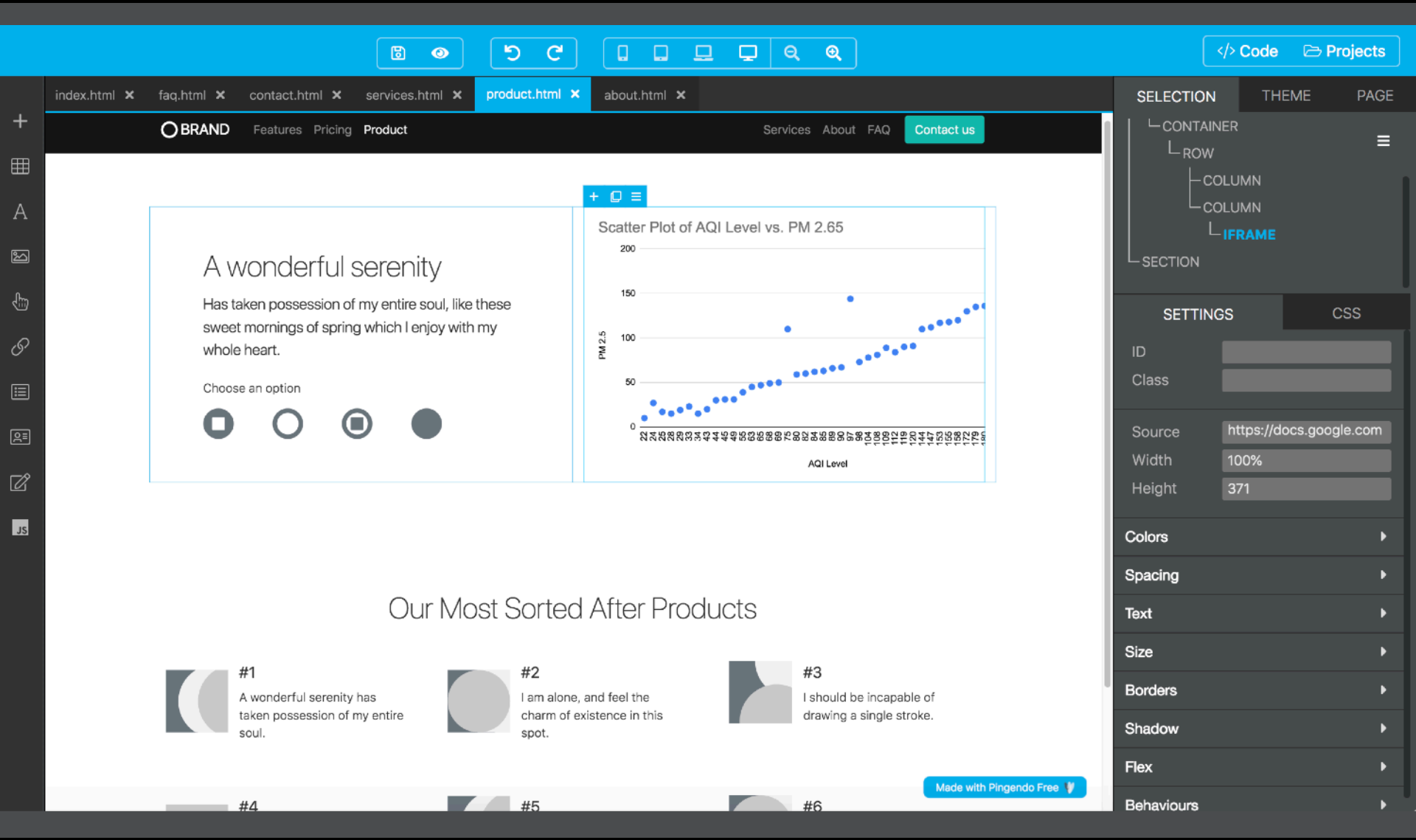
	A	B	D	F
1	AQR_rank	AQR_city	AQR_AQI	AQR_pm_2_5
2	1	Sanya	22	10
3	2	Lijiang	24	12
4	3	Dali	24	15
5	4	Changsha	25	17
6	5	Haikou	28	15
7	6	Qingdao	29	19
8	7	Kunming	33	23
9	8	Lhasa	34	15
10	9	Dalian	43	20
11	10	Guangzhou	44	30
12	11	Guiyang	45	31
13	12	Zhangjiakou	49	31
14	13	Shenzhen	55	39
15	14	Zhangjiajie	63	45
16	15	Fuzhou	65	47
17	16	Nanning	68	49
18	17	Guilin	69	50
19	18	Beijing	75	55
20	19	Xining	75	55
21	20	Chongqing	80	59
22	21	Lanzhou	82	60



Sheet1

Explore

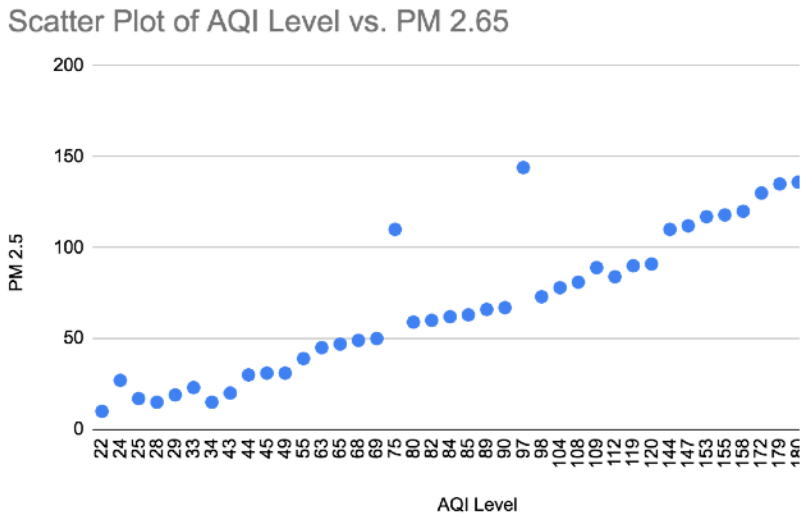

Untitled spreadsheet
File Edit View Insert Format Data Tools Add-ons Help Last edit was 3 hours ago
Share
AQR_rank
Publish to the web
This document is published to the web.
Make your content visible to anyone by publishing it to the web. You can link to or embed your document. Learn more
Link
Embed
Scatter Plot of A...
Interactive
<iframe width="600" height="371" seamless frameborder="0" scrolling="no" src="https://docs.google.com/spreadsheets/d/e/2PACX-1vQhznFzMIQL5z24-DV-IkT0h5JcFa8WNBkhIn-dnnBdJv-XjckSuA8sTZJwG2vgwhbLVeA6zWbEcQv5/pubchart?oid=544627711&format=interactive"></iframe>
Note: Viewers may be able to access the underlying data for published charts. Learn more
Published
Published content & settings
Sheet1
Explore


Code
Projects
index.html
faq.html
contact.html
services.html
product.html
about.html
SELECTION
THEME
PAGE
CONTAINER
ROW
COLUMN
COLUMN
IFRAME
SECTION
SETTINGS
CSS
ID
Class
Source
https://docs.google.com
Width
100%
Height
371
Colors
Spacing
Text
Size
Borders
Shadow
Flex
Behaviours
BRAND
Features
Pricing
Product
Services
About
FAQ
Contact us
A wonderful serenity
Has taken possession of my entire soul, like these sweet mornings of spring which I enjoy with my whole heart.
Choose an option
Scatter Plot of AQI Level vs. PM 2.65
PM 2.5
AQI Level
Our Most Sorted After Products
#1
A wonderful serenity has taken possession of my entire soul.
#2
I am alone, and feel the charm of existence in this spot.
#3
I should be incapable of drawing a single stroke.
#4
#5
#6
Made with Pingendo Free

A wonderful serenity

Has taken possession of my entire soul, like these sweet mornings of spring which I enjoy with my whole heart.

Choose an option



Clean up data in OpenRefine.

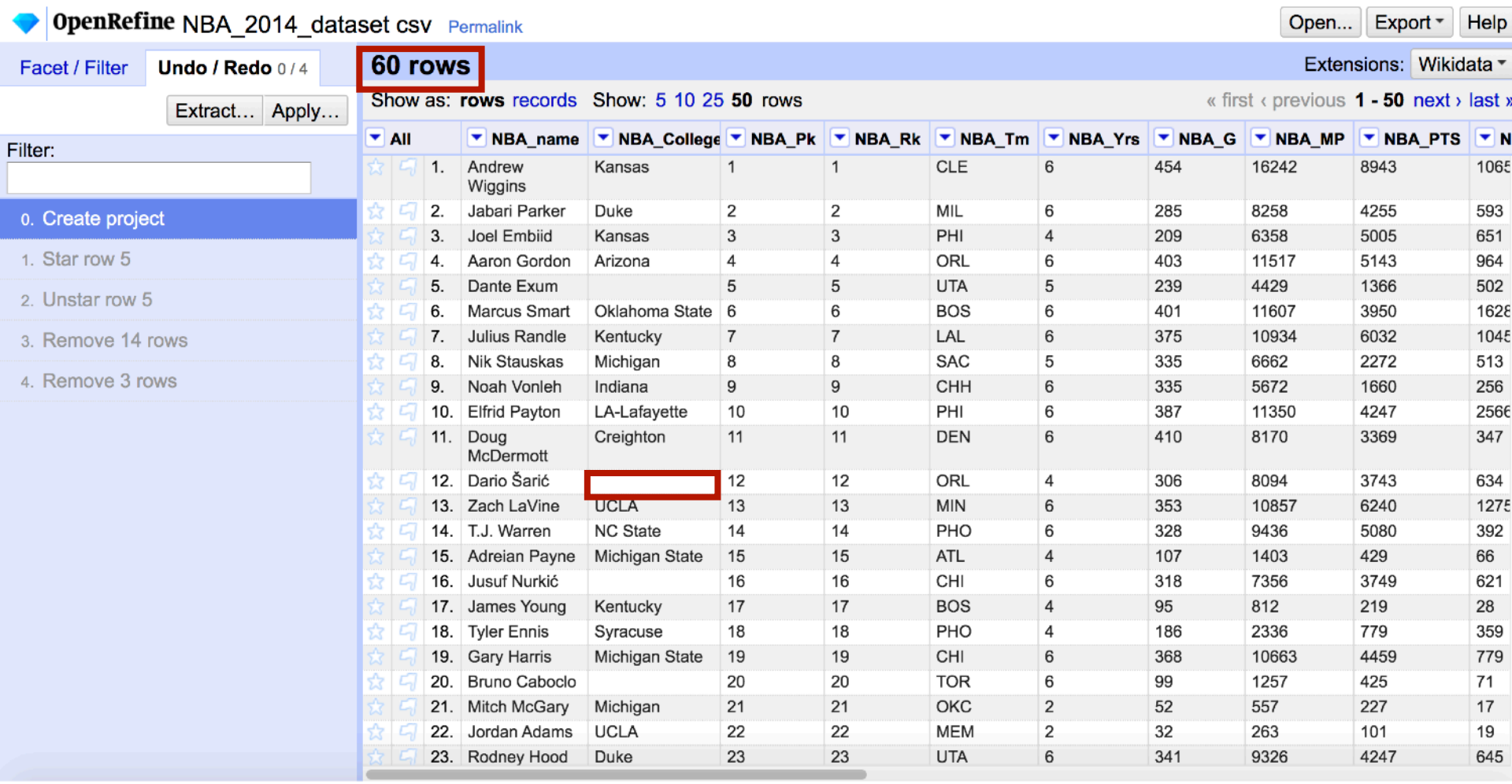
Most Basic Data Cleaning Tasks

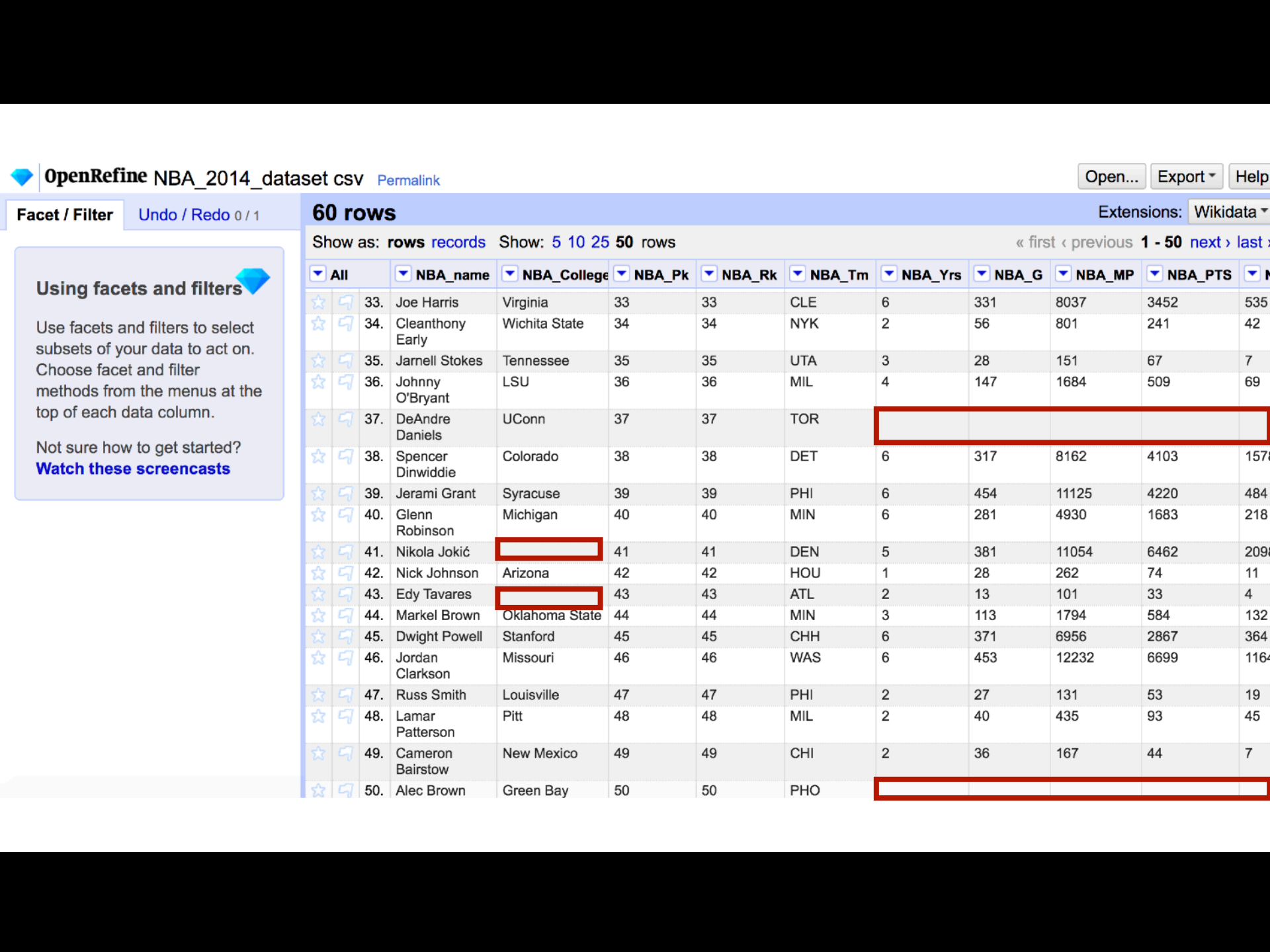
1. Remove blanks
2. Impute missing entries (need expertise to judge usage instead of relying on statistics alone)
3. Remove duplicates
4. Resolve inconsistent entries
5. Transform incorrect data (e.g. wrong data types such as text instead of numeric and calculations)
6. Rename, combine and split columns for making further data presentation and processing easier

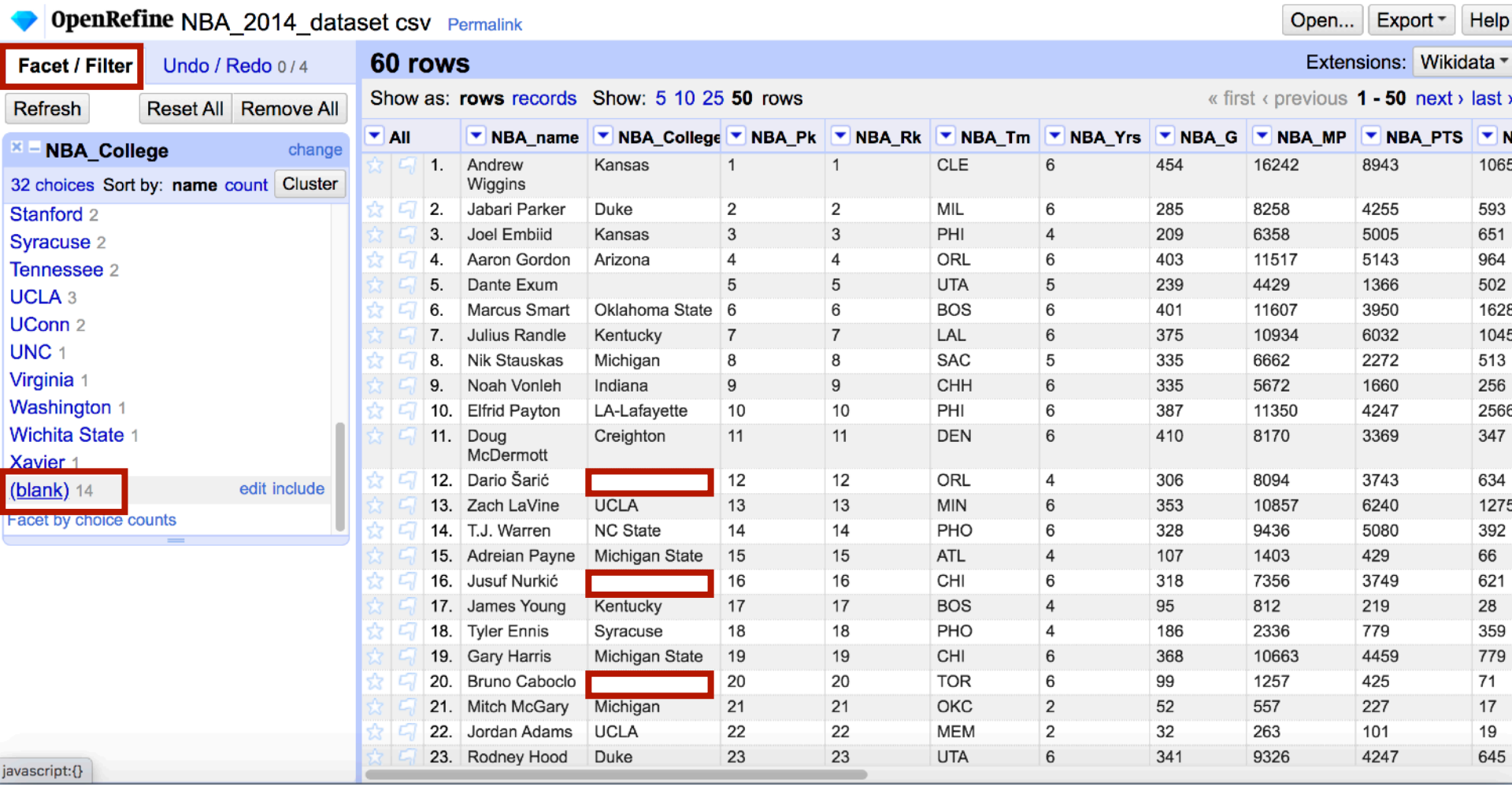
NBA_name	NBA_College	NBA_Pk	NBA_Rk	NBA_Tm	NBA_Yrs	NBA_G	NBA_MP	NBA_PTS	NBA_AST	NBA_FG	NBA_THREE_P	NBA_
Andrew Wiggins	Kansas	1	1	CLE	6	454	16242	8943	1065	0.441	0.332	0.7
Jabari Parker	Duke	2	2	MIL	6	285	8258	4255	593	0.493	0.324	0.7
Joel Embiid	Kansas	3	3	PHI	4	209	6358	5005	651	0.480	0.319	0.7
Aaron Gordon	Arizona	4	4	ORL	6	403	11517	5143	964	0.448	0.319	0.7
Dante Exum		5	5	UTA	5	239	4429	1366	502	0.407	0.308	0.7
Marcus Smart	Oklahoma State	6	6	BOS	6	401	11607	3950	1628	0.373	0.318	0.7
Julius Randle	Kentucky	7	7	LAL	6	375	10934	6032	1045	0.493	0.295	0.7
Nik Stauskas	Michigan	8	8	SAC	5	335	6662	2272	513	0.389	0.353	0.8
Noah Vonleh	Indiana	9	9	CHH	6	335	5672	1660	256	0.460	0.310	0.6
Elfrid Payton	LA-Lafayette	10	10	PHI	6	387	11350	4247	2566	0.452	0.289	0.6
Doug McDermott	Creighton	11	11	DEN	6	410	8170	3369	347	0.465	0.412	0.8
Dario Šarić		12	12	ORL	4	306	8094	3743	634	0.441	0.358	0.8
Zach LaVine	UCLA	13	13	MIN	6	353	10857	6240	1275	0.447	0.375	0.8
T.J. Warren	NC State	14	14	PHO	6	328	9436	5080	392	0.507	0.361	0.7
Adreian Payne	Michigan State	15	15	ATL	4	107	1403	429	66	0.406	0.254	0.6
Jusuf Nurkić		16	16	CHI	6	318	7356	3749	621	0.492	0.096	0.6
James Young	Kentucky	17	17	BOS	4	95	812	219	28	0.367	0.277	0.5
Tyler Ennis	Syracuse	18	18	PHO	4	186	2336	779	359	0.419	0.317	0.7
Gary Harris	Michigan State	19	19	CHI	6	368	10663	4459	779	0.454	0.360	0.8
Bruno Caboclo		20	20	TOR	6	99	1257	425	71	0.400	0.316	0.8
Mitch McGary	Michigan	21	21	OKC	2	52	557	227	17	0.527	0.000	0.5
Jordan Adams	UCLA	22	22	MEM	2	32	263	101	19	0.402	0.385	0.6
Rodney Hood	Duke	23	23	UTA	6	341	9326	4247	645	0.426	0.372	0.8

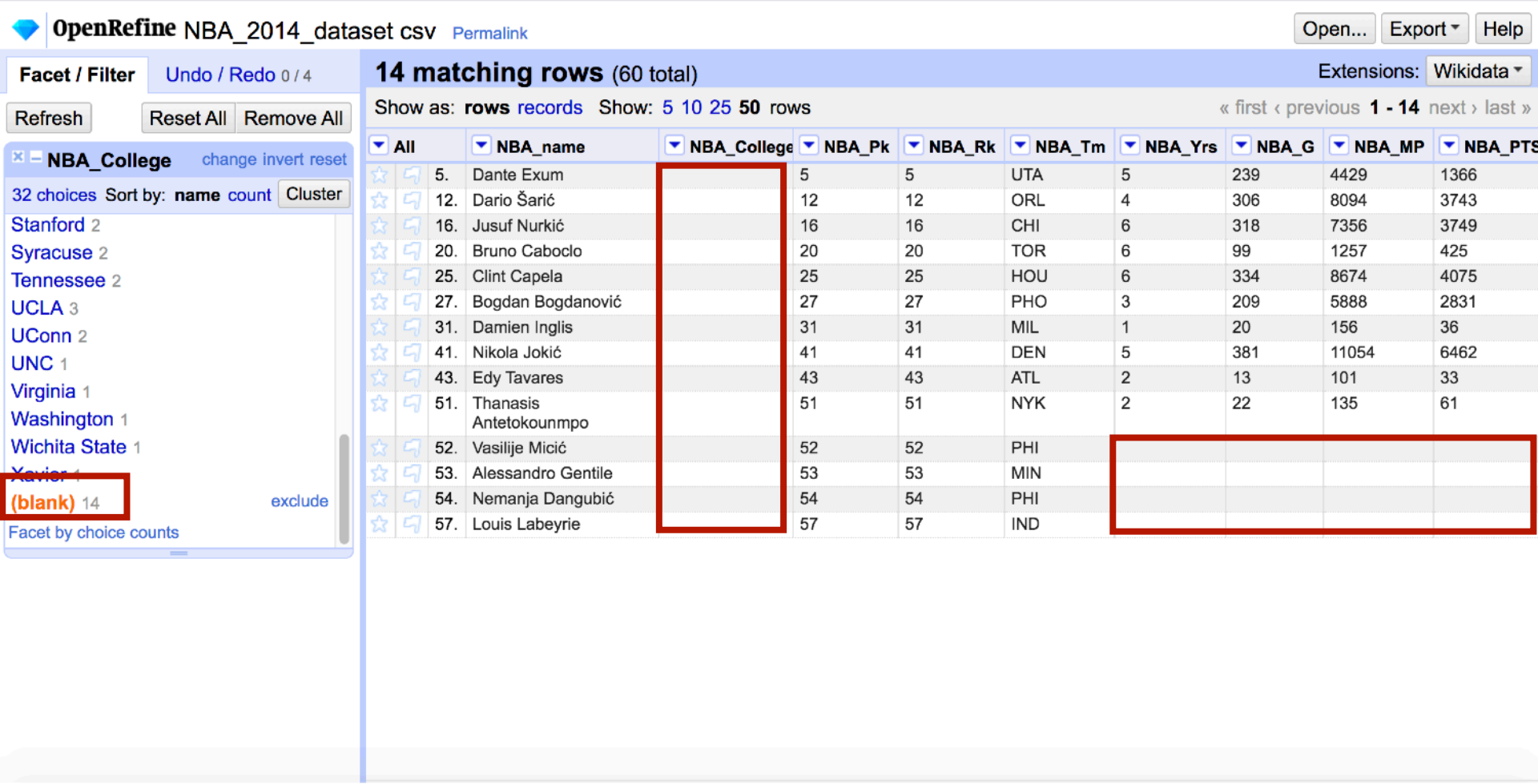
Sheet 1

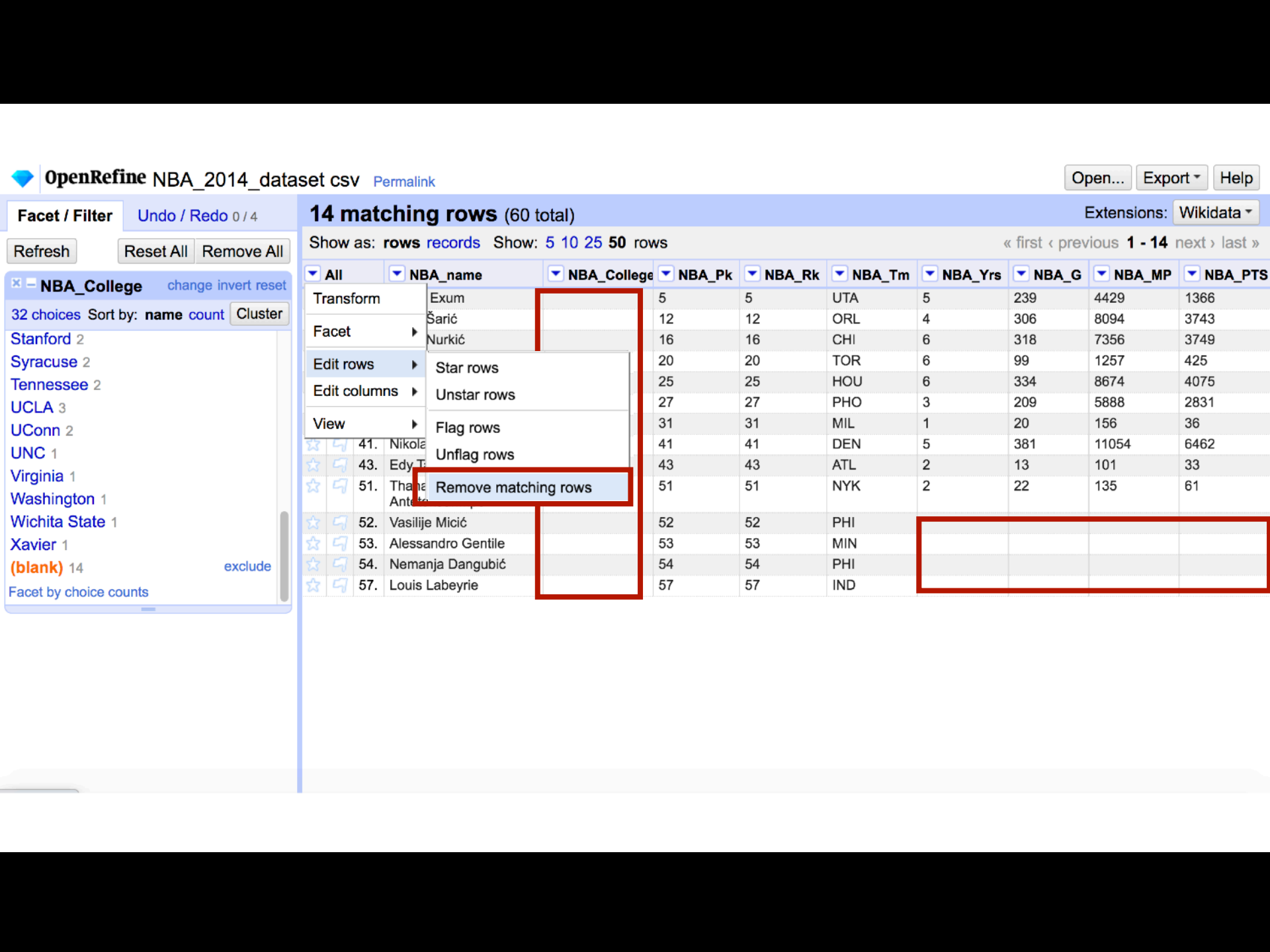
DeAndre Daniels	UConn	37	37	TOR								
Spencer Dinwiddie	Colorado	38	38	DET	6	317	8162	4103	1578	0.410	0.318	0.7
Jerami Grant	Syracuse	39	39	PHI	6	454	11125	4220	484	0.465	0.347	0.6
Glenn Robinson	Michigan	40	40	MIN	6	281	4930	1683	218	0.459	0.373	0.7
Nikola Jokić		41	41	DEN	5	381	11054	6462	2098	0.524	0.338	0.8
Nick Johnson	Arizona	42	42	HOU	1	28	262	74	11	0.347	0.238	0.6
Edy Tavares		43	43	ATL	2	13	101	33	4	0.625		0.2
Markel Brown	Oklahoma State	44	44	MIN	3	113	1794	584	132	0.380	0.295	0.7
Dwight Powell	Stanford	45	45	CHH	6	371	6956	2867	364	0.563	0.293	0.7
Jordan Clarkson	Missouri	46	46	WAS	6	453	12232	6699	1164	0.446	0.342	0.8
Russ Smith	Louisville	47	47	PHI	2	27	131	53	19	0.319	0.188	0.7
Lamar Patterson	Pitt	48	48	MIL	2	40	435	93	45	0.326	0.236	0.7
Cameron Bairstow	New Mexico	49	49	CHI	2	36	167	44	7	0.296	0.200	0.8
Alec Brown	Green Bay	50	50	PHO								
Thanasis Antetokounmpo		51	51	NYK	2	22	135	61	15	0.519	0.000	0.4
Vasilije Micić		52	52	PHI								
Alessandro Gentile		53	53	MIN								
Nemanja Dangubić		54	54	PHI								
Semaj Christon	Xavier	55	55	MIA	1	64	973	183	130	0.345	0.190	0.5
Devyn Marble	Iowa	56	56	DEN	2	44	457	97	29	0.304	0.222	0.3
Louis Labeyrie		57	57	IND								
Jordan McRae	Tennessee	58	58	SAS	4	123	1696	846	167	0.417	0.355	0.7
Xavier Thames	San Diego State	59	59	TOR								
Cory Jefferson	Baylor	60	60	SAS	2	58	581	205	16	0.444	0.125	0.5

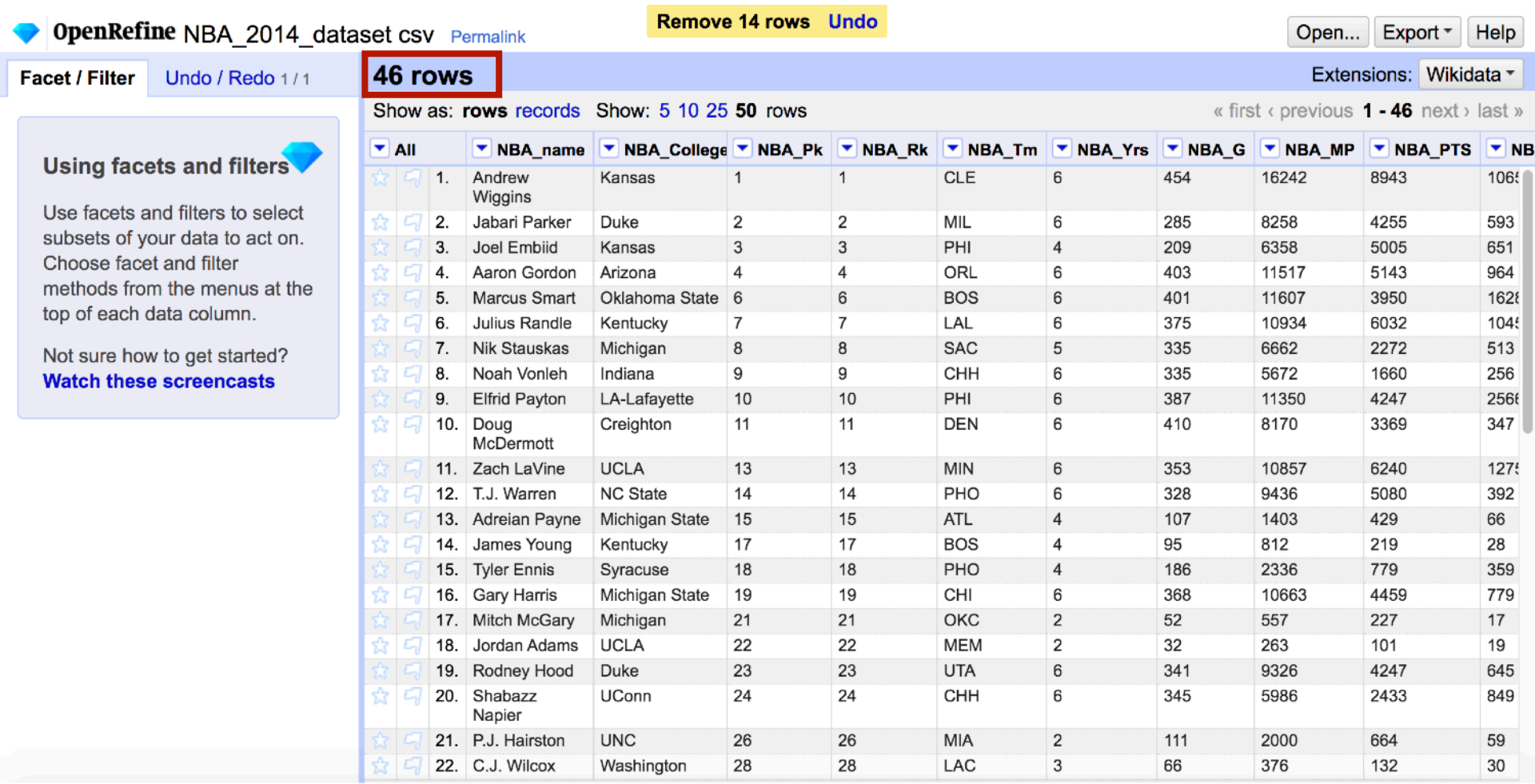

OpenRefine NBA_2014_dataset csv Permalink
Open... Export Help
Facet / Filter
Undo / Redo 0 / 4
Extract... Apply...
Filter:
0. Create project
1. Star row 5
2. Unstar row 5
3. Remove 14 rows
4. Remove 3 rows
60 rows
Extensions: Wikidata
Show as: rows records Show: 5 10 25 50 rows
« first ‹ previous 1 - 50 next › last »
All
NBA_name
NBA_College
NBA_Pk
NBA_Rk
NBA_Tm
NBA_Yrs
NBA_G
NBA_MP
NBA_PTS
1. Andrew Wiggins Kansas 1 1 CLE 6 454 16242 8943 1065
2. Jabari Parker Duke 2 2 MIL 6 285 8258 4255 593
3. Joel Embiid Kansas 3 3 PHI 4 209 6358 5005 651
4. Aaron Gordon Arizona 4 4 ORL 6 403 11517 5143 964
5. Dante Exum 5 5 UTA 5 239 4429 1366 502
6. Marcus Smart Oklahoma State 6 6 BOS 6 401 11607 3950 1628
7. Julius Randle Kentucky 7 7 LAL 6 375 10934 6032 1045
8. Nik Stauskas Michigan 8 8 SAC 5 335 6662 2272 513
9. Noah Vonleh Indiana 9 9 CHH 6 335 5672 1660 256
10. Elfrid Payton LA-Lafayette 10 10 PHI 6 387 11350 4247 2566
11. Doug McDermott Creighton 11 11 DEN 6 410 8170 3369 347
12. Dario Šarić 12 12 ORL 4 306 8094 3743 634
13. Zach LaVine UCLA 13 13 MIN 6 353 10857 6240 1275
14. T.J. Warren NC State 14 14 PHO 6 328 9436 5080 392
15. Adreian Payne Michigan State 15 15 ATL 4 107 1403 429 66
16. Jusuf Nurkić 16 16 CHI 6 318 7356 3749 621
17. James Young Kentucky 17 17 BOS 4 95 812 219 28
18. Tyler Ennis Syracuse 18 18 PHO 4 186 2336 779 359
19. Gary Harris Michigan State 19 19 CHI 6 368 10663 4459 779
20. Bruno Caboclo 20 20 TOR 6 99 1257 425 71
21. Mitch McGary Michigan 21 21 OKC 2 52 557 227 17
22. Jordan Adams UCLA 22 22 MEM 2 32 263 101 19
23. Rodney Hood Duke 23 23 UTA 6 341 9326 4247 645


OpenRefine NBA_2014_dataset csv Permalink
Open... Export Help
Facet / Filter Undo / Redo 0 / 1
60 rows
Extensions: Wikidata
Show as: rows records Show: 5 10 25 50 rows
« first ‹ previous 1 - 50 next › last
Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started?
Watch these screencasts
All
NBA_name
NBA_College
NBA_Pk
NBA_Rk
NBA_Tm
NBA_Yrs
NBA_G
NBA_MP
NBA_PTS
33. Joe Harris Virginia 33 33 CLE 6 331 8037 3452 535
34. Cleanthony Early Wichita State 34 34 NYK 2 56 801 241 42
35. Jarnell Stokes Tennessee 35 35 UTA 3 28 151 67 7
36. Johnny O'Bryant LSU 36 36 MIL 4 147 1684 509 69
37. DeAndre Daniels UConn 37 37 TOR
38. Spencer Dinwiddie Colorado 38 38 DET 6 317 8162 4103 157
39. Jerami Grant Syracuse 39 39 PHI 6 454 11125 4220 484
40. Glenn Robinson Michigan 40 40 MIN 6 281 4930 1683 218
41. Nikola Jokić 41 41 DEN 5 381 11054 6462 209
42. Nick Johnson Arizona 42 42 HOU 1 28 262 74 11
43. Edy Tavares 43 43 ATL 2 13 101 33 4
44. Markel Brown Oklahoma State 44 44 MIN 3 113 1794 584 132
45. Dwight Powell Stanford 45 45 CHH 6 371 6956 2867 364
46. Jordan Clarkson Missouri 46 46 WAS 6 453 12232 6699 116
47. Russ Smith Louisville 47 47 PHI 2 27 131 53 19
48. Lamar Patterson Pitt 48 48 MIL 2 40 435 93 45
49. Cameron Bairstow New Mexico 49 49 CHI 2 36 167 44 7
50. Alec Brown Green Bay 50 50 PHO

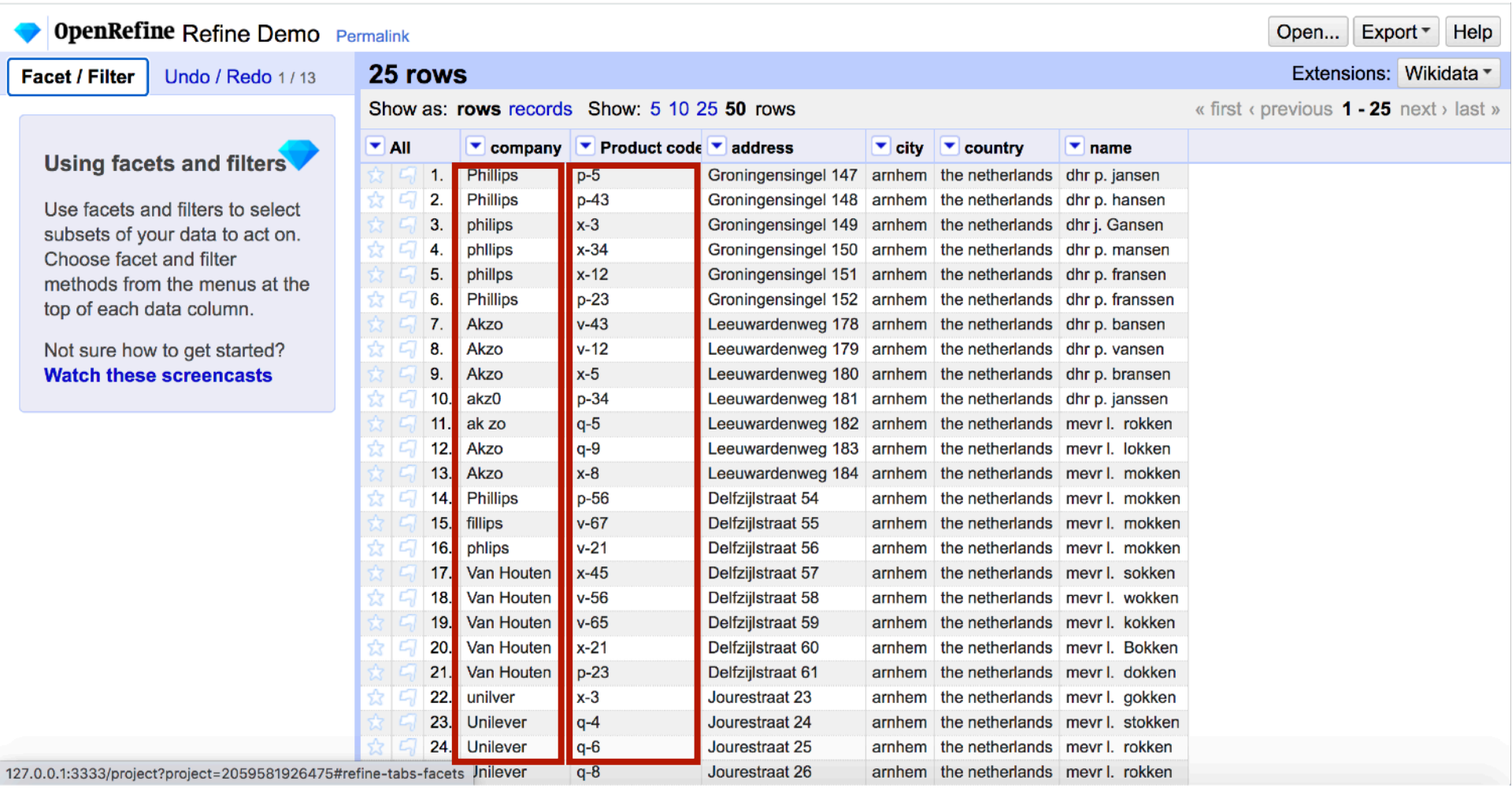

OpenRefine NBA_2014_dataset csv
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 0 / 4
Refresh
Reset All
Remove All
NBA_College
change
32 choices Sort by: name count
Cluster
Stanford 2
Syracuse 2
Tennessee 2
UCLA 3
UConn 2
UNC 1
Virginia 1
Washington 1
Wichita State 1
Xavier 1
(blank) 14
edit include
Facet by choice counts
60 rows
Extensions: Wikidata
Show as: rows records
Show: 5 10 25 50 rows
« first ‹ previous 1 - 50 next › last »
All
NBA_name
NBA_College
NBA_Pk
NBA_Rk
NBA_Tm
NBA_Yrs
NBA_G
NBA_MP
NBA_PTS
1. Andrew Wiggins Kansas 1 1 CLE 6 454 16242 8943 1065
2. Jabari Parker Duke 2 2 MIL 6 285 8258 4255 593
3. Joel Embiid Kansas 3 3 PHI 4 209 6358 5005 651
4. Aaron Gordon Arizona 4 4 ORL 6 403 11517 5143 964
5. Dante Exum 5 5 UTA 5 239 4429 1366 502
6. Marcus Smart Oklahoma State 6 6 BOS 6 401 11607 3950 1628
7. Julius Randle Kentucky 7 7 LAL 6 375 10934 6032 1045
8. Nik Stauskas Michigan 8 8 SAC 5 335 6662 2272 513
9. Noah Vonleh Indiana 9 9 CHH 6 335 5672 1660 256
10. Elfrid Payton LA-Lafayette 10 10 PHI 6 387 11350 4247 2566
11. Doug McDermott Creighton 11 11 DEN 6 410 8170 3369 347
12. Dario Šarić 12 12 ORL 4 306 8094 3743 634
13. Zach LaVine UCLA 13 13 MIN 6 353 10857 6240 1275
14. T.J. Warren NC State 14 14 PHO 6 328 9436 5080 392
15. Adreian Payne Michigan State 15 15 ATL 4 107 1403 429 66
16. Jusuf Nurkić 16 16 CHI 6 318 7356 3749 621
17. James Young Kentucky 17 17 BOS 4 95 812 219 28
18. Tyler Ennis Syracuse 18 18 PHO 4 186 2336 779 359
19. Gary Harris Michigan State 19 19 CHI 6 368 10663 4459 779
20. Bruno Caboclo 20 20 TOR 6 99 1257 425 71
21. Mitch McGary Michigan 21 21 OKC 2 52 557 227 17
22. Jordan Adams UCLA 22 22 MEM 2 32 263 101 19
23. Rodney Hood Duke 23 23 UTA 6 341 9326 4247 645
javascript:{}

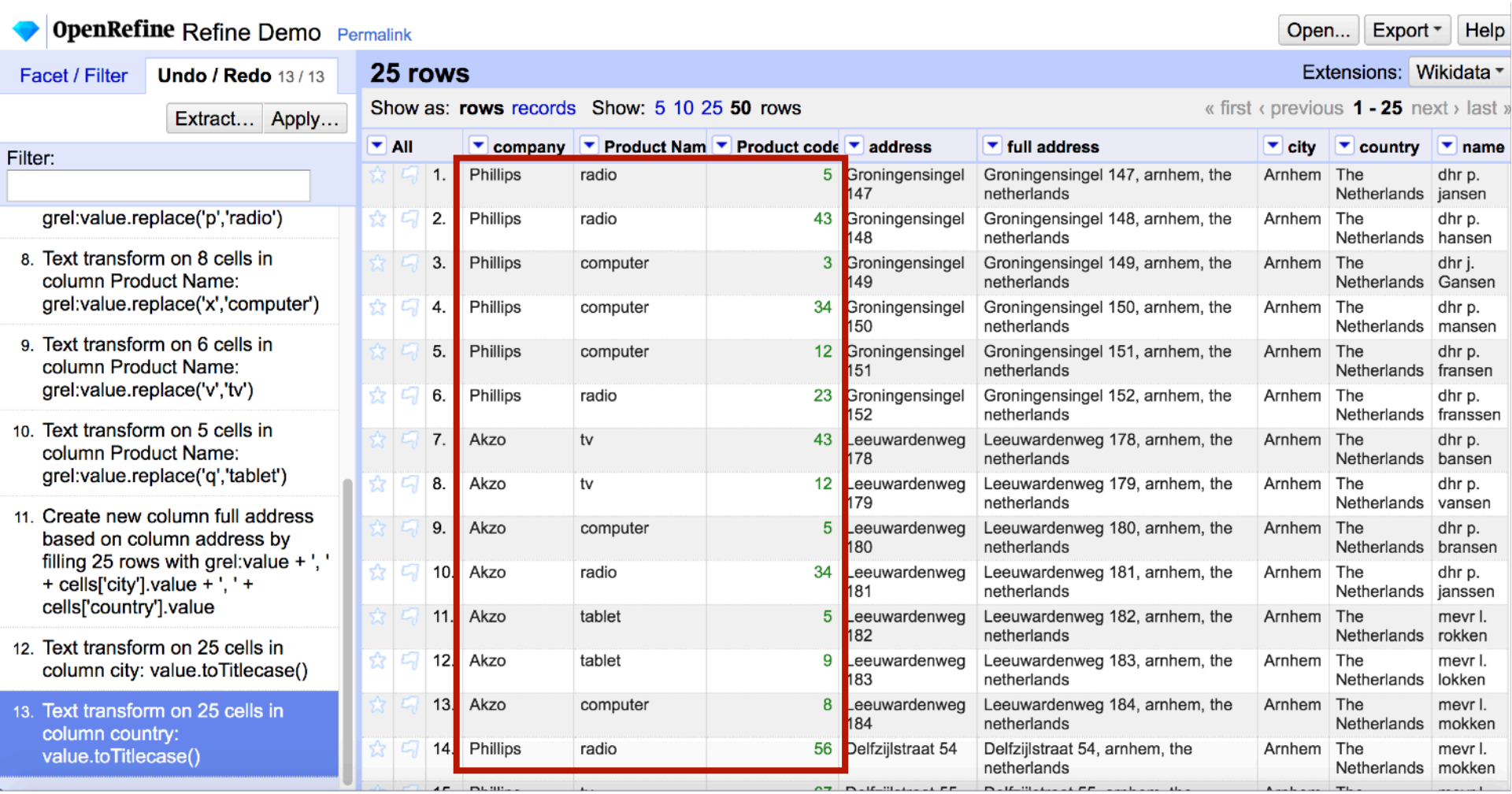

OpenRefine NBA_2014_dataset csv Permalink
Open... Export Help
Facet / Filter Undo / Redo 0 / 4
Refresh Reset All Remove All
NBA_College change invert reset
32 choices Sort by: name count Cluster
Stanford 2
Syracuse 2
Tennessee 2
UCLA 3
UConn 2
UNC 1
Virginia 1
Washington 1
Wichita State 1
(blank) 14 exclude
Facet by choice counts
14 matching rows (60 total)
Extensions: Wikidata
Show as: rows records Show: 5 10 25 50 rows
« first ‹ previous 1 - 14 next › last »
All NBA_name NBA_College NBA_Pk NBA_Rk NBA_Tm NBA_Yrs NBA_G NBA_MP NBA_PTS
5. Dante Exum 5 5 UTA 5 239 4429 1366
12. Dario Šarić 12 12 ORL 4 306 8094 3743
16. Jusuf Nurkić 16 16 CHI 6 318 7356 3749
20. Bruno Caboclo 20 20 TOR 6 99 1257 425
25. Clint Capela 25 25 HOU 6 334 8674 4075
27. Bogdan Bogdanović 27 27 PHO 3 209 5888 2831
31. Damien Inglis 31 31 MIL 1 20 156 36
41. Nikola Jokić 41 41 DEN 5 381 11054 6462
43. Edy Tavares 43 43 ATL 2 13 101 33
51. Thanasis Antetokounmpo 51 51 NYK 2 22 135 61
52. Vasilije Micić 52 52 PHI
53. Alessandro Gentile 53 53 MIN
54. Nemanja Dangubić 54 54 PHI
57. Louis Labeyrie 57 57 IND

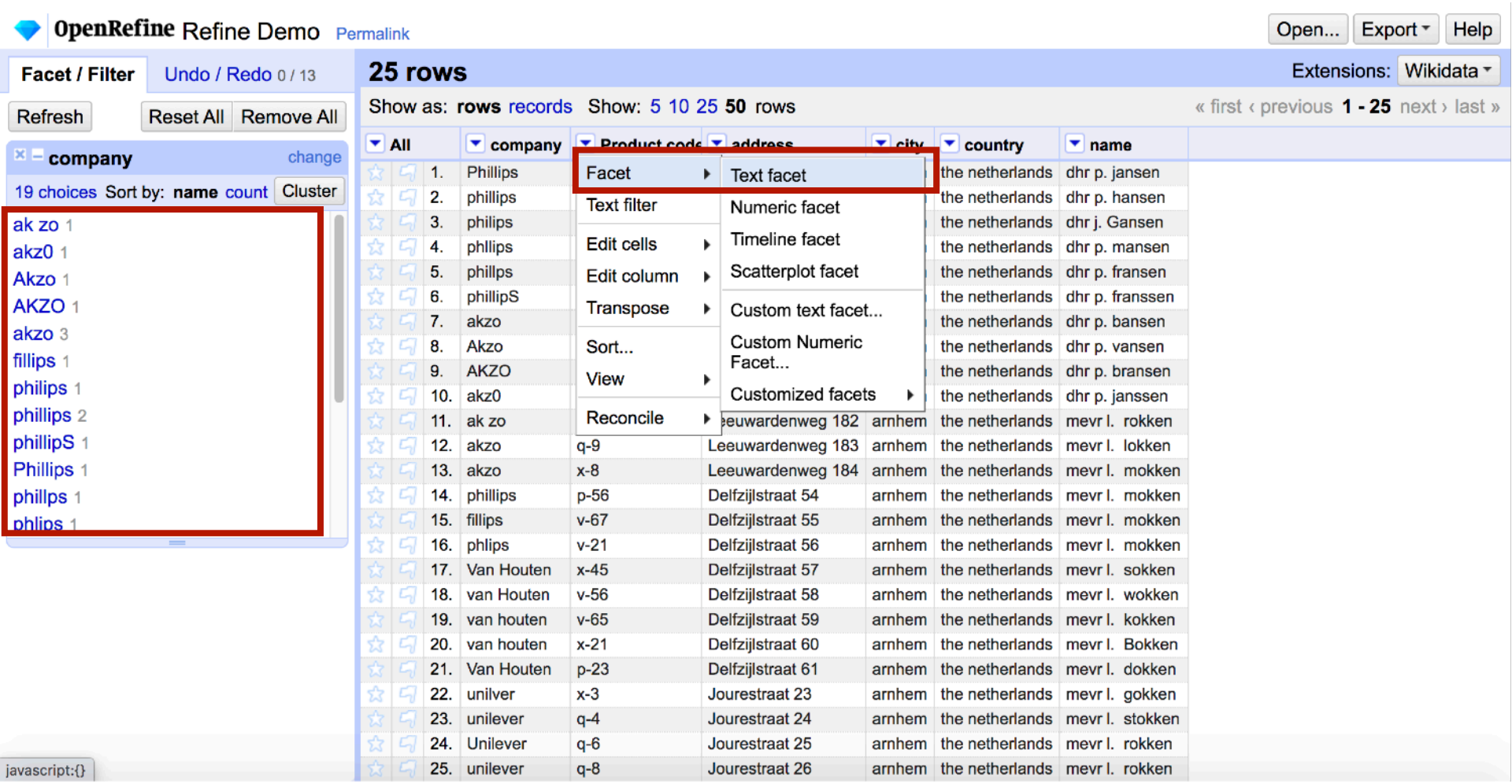

OpenRefine NBA_2014_dataset csv
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 0 / 4
Refresh
Reset All
Remove All
NBA_College
change invert reset
32 choices Sort by: name count
Cluster
Stanford 2
Syracuse 2
Tennessee 2
UCLA 3
UConn 2
UNC 1
Virginia 1
Washington 1
Wichita State 1
Xavier 1
(blank) 14
exclude
Facet by choice counts
14 matching rows (60 total)
Extensions: Wikidata
Show as: rows records
Show: 5 10 25 50 rows
« first ‹ previous 1 - 14 next › last »
All
NBA_name
NBA_College
NBA_Pk
NBA_Rk
NBA_Tm
NBA_Yrs
NBA_G
NBA_MP
NBA_PTS
Transform
Facet
Edit rows
Edit columns
View
Star rows
Unstar rows
Flag rows
Unflag rows
Remove matching rows
Exum 5 5 UTA 5 239 4429 1366
Šarić 12 12 ORL 4 306 8094 3743
Nurkić 16 16 CHI 6 318 7356 3749
20 20 TOR 6 99 1257 425
25 25 HOU 6 334 8674 4075
27 27 PHO 3 209 5888 2831
31 31 MIL 1 20 156 36
41. Nikola 41 41 DEN 5 381 11054 6462
43. Edy T 43 43 ATL 2 13 101 33
51. Tha Ant 51 51 NYK 2 22 135 61
52. Vasilije Micić 52 52 PHI
53. Alessandro Gentile 53 53 MIN
54. Nemanja Dangubić 54 54 PHI
57. Louis Labeyrie 57 57 IND

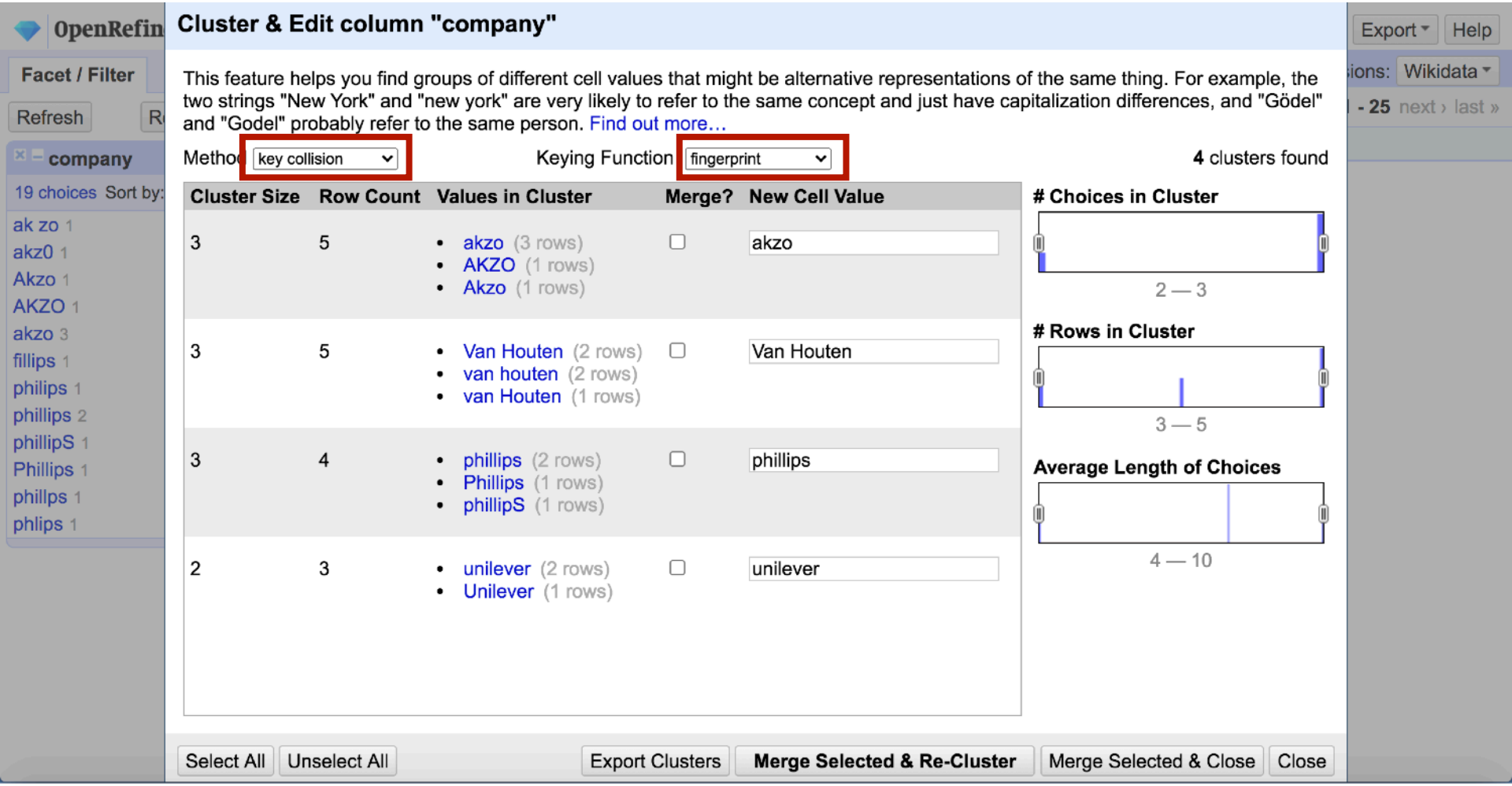

OpenRefine NBA_2014_dataset csv Permalink
Remove 14 rows Undo
Open... Export Help
Facet / Filter Undo / Redo 1 / 1
46 rows
Extensions: Wikidata
Show as: rows records Show: 5 10 25 50 rows
« first ‹ previous 1 - 46 next › last »
Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started?
Watch these screencasts
All
NBA_name
NBA_College
NBA_Pk
NBA_Rk
NBA_Tm
NBA_Yrs
NBA_G
NBA_MP
NBA_PTS
1. Andrew Wiggins Kansas 1 1 CLE 6 454 16242 8943
2. Jabari Parker Duke 2 2 MIL 6 285 8258 4255 593
3. Joel Embiid Kansas 3 3 PHI 4 209 6358 5005 651
4. Aaron Gordon Arizona 4 4 ORL 6 403 11517 5143 964
5. Marcus Smart Oklahoma State 6 6 BOS 6 401 11607 3950
6. Julius Randle Kentucky 7 7 LAL 6 375 10934 6032
7. Nik Stauskas Michigan 8 8 SAC 5 335 6662 2272 513
8. Noah Vonleh Indiana 9 9 CHH 6 335 5672 1660 256
9. Elfrid Payton LA-Lafayette 10 10 PHI 6 387 11350 4247
10. Doug McDermott Creighton 11 11 DEN 6 410 8170 3369 347
11. Zach LaVine UCLA 13 13 MIN 6 353 10857 6240
12. T.J. Warren NC State 14 14 PHO 6 328 9436 5080 392
13. Adreian Payne Michigan State 15 15 ATL 4 107 1403 429 66
14. James Young Kentucky 17 17 BOS 4 95 812 219 28
15. Tyler Ennis Syracuse 18 18 PHO 4 186 2336 779 359
16. Gary Harris Michigan State 19 19 CHI 6 368 10663 4459 779
17. Mitch McGary Michigan 21 21 OKC 2 52 557 227 17
18. Jordan Adams UCLA 22 22 MEM 2 32 263 101 19
19. Rodney Hood Duke 23 23 UTA 6 341 9326 4247 645
20. Shabazz Napier UConn 24 24 CHH 6 345 5986 2433 849
21. P.J. Hairston UNC 26 26 MIA 2 111 2000 664 59
22. C.J. Wilcox Washington 28 28 LAC 3 66 376 132 30

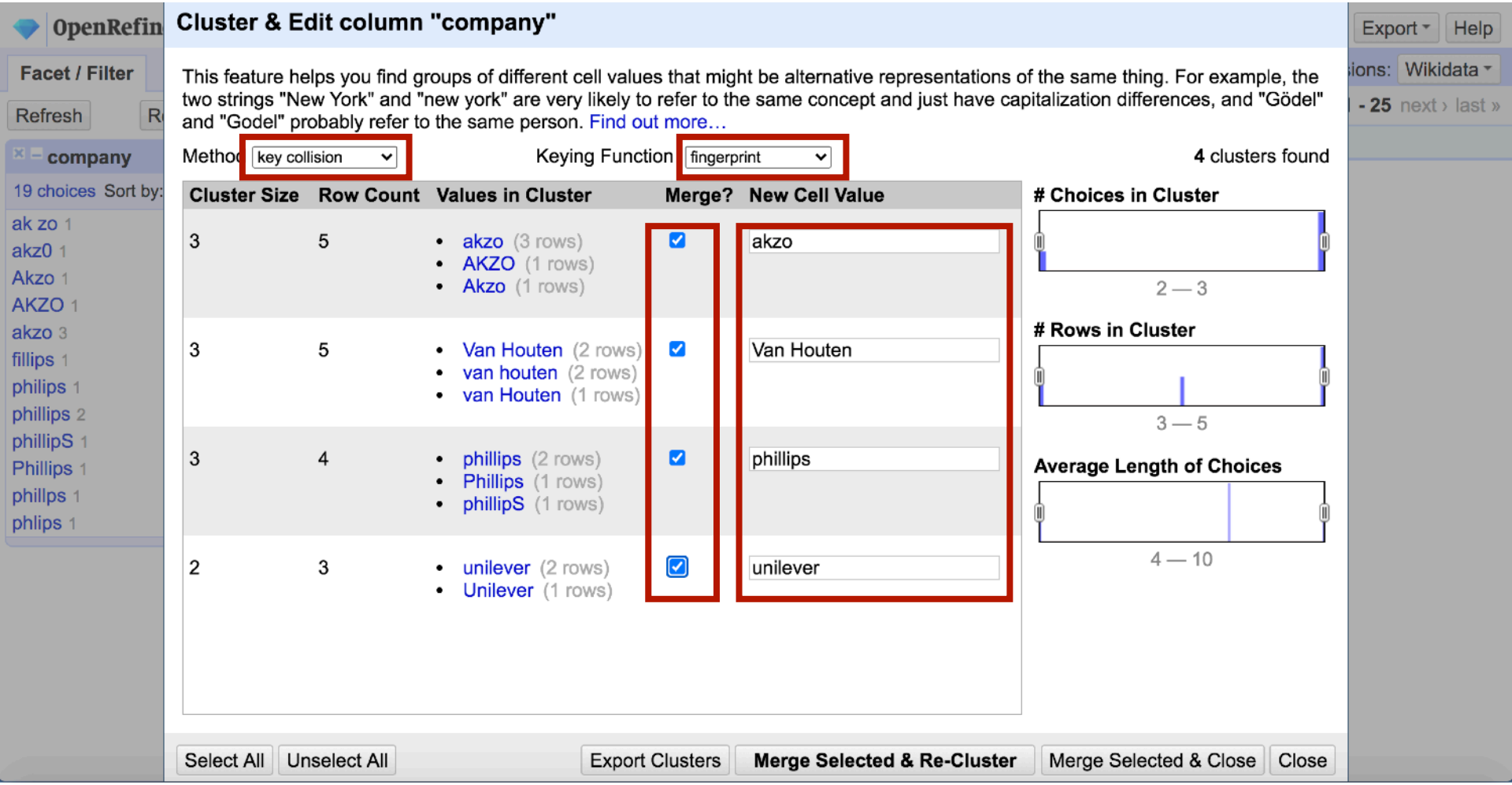
Another example.

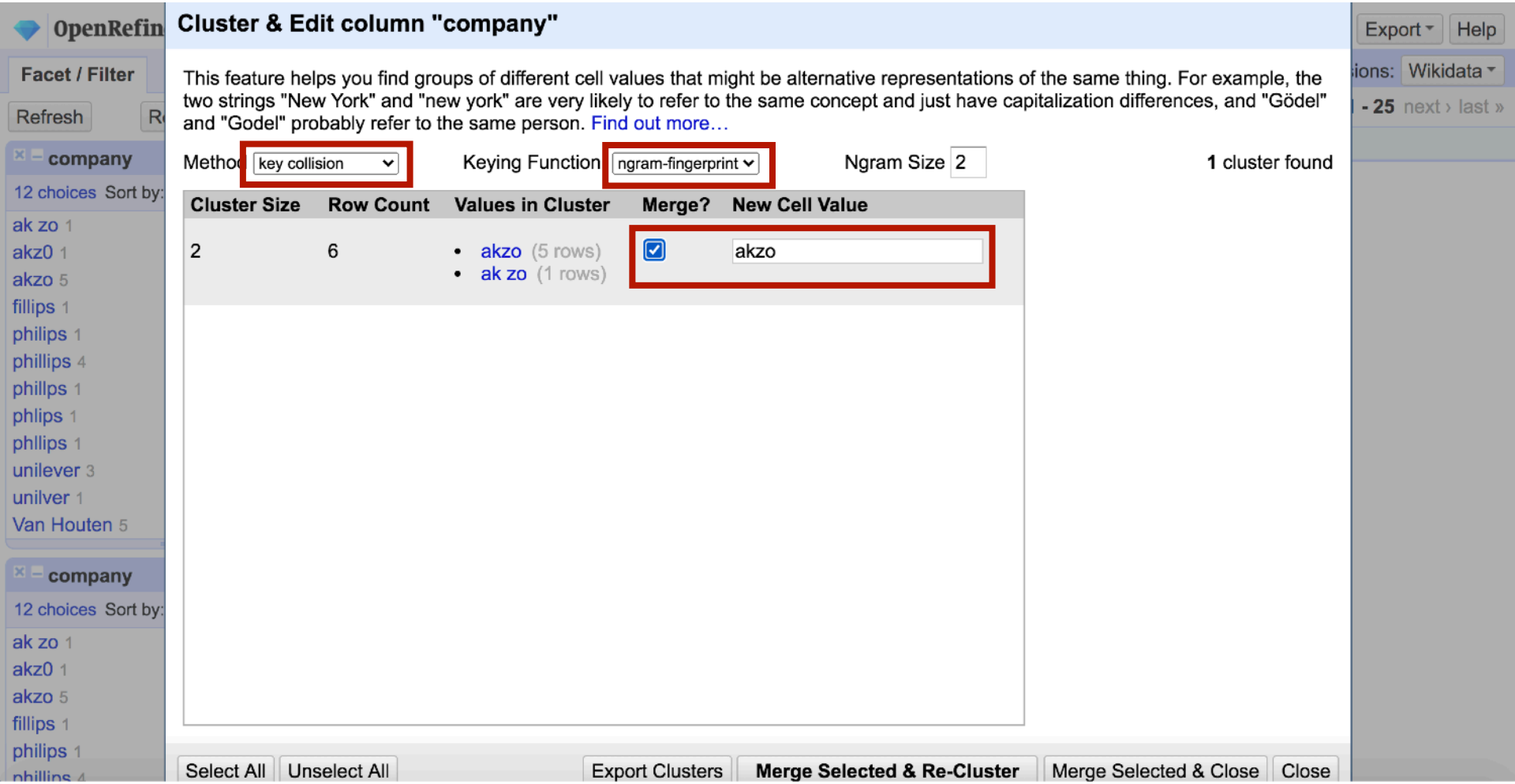

OpenRefine Refine Demo
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 1 / 13
25 rows
Extensions: Wikidata
Show as: rows records Show: 5 10 25 50 rows
« first ‹ previous 1 - 25 next › last »
Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started?
Watch these screencasts
All
company
Product code
address
city
country
name
1. Phillips p-5 Groningensingel 147 arnhem the netherlands dhr p. jansen
2. Phillips p-43 Groningensingel 148 arnhem the netherlands dhr p. hansen
3. philips x-3 Groningensingel 149 arnhem the netherlands dhr j. Gansen
4. phllips x-34 Groningensingel 150 arnhem the netherlands dhr p. mansen
5. phillps x-12 Groningensingel 151 arnhem the netherlands dhr p. fransen
6. Phillips p-23 Groningensingel 152 arnhem the netherlands dhr p. franssen
7. Akzo v-43 Leeuwardenweg 178 arnhem the netherlands dhr p. bansen
8. Akzo v-12 Leeuwardenweg 179 arnhem the netherlands dhr p. vansen
9. Akzo x-5 Leeuwardenweg 180 arnhem the netherlands dhr p. bransen
10. akz0 p-34 Leeuwardenweg 181 arnhem the netherlands dhr p. janssen
11. ak zo q-5 Leeuwardenweg 182 arnhem the netherlands mevr l. rokken
12. Akzo q-9 Leeuwardenweg 183 arnhem the netherlands mevr l. lokken
13. Akzo x-8 Leeuwardenweg 184 arnhem the netherlands mevr l. mokken
14. Phillips p-56 Delfzijlstraat 54 arnhem the netherlands mevr l. mokken
15. fillips v-67 Delfzijlstraat 55 arnhem the netherlands mevr l. mokken
16. phlips v-21 Delfzijlstraat 56 arnhem the netherlands mevr l. mokken
17. Van Houten x-45 Delfzijlstraat 57 arnhem the netherlands mevr l. sokken
18. Van Houten v-56 Delfzijlstraat 58 arnhem the netherlands mevr l. wokken
19. Van Houten v-65 Delfzijlstraat 59 arnhem the netherlands mevr l. kokken
20. Van Houten x-21 Delfzijlstraat 60 arnhem the netherlands mevr l. Bokken
21. Van Houten p-23 Delfzijlstraat 61 arnhem the netherlands mevr l. dokken
22. unilver x-3 Jourestraat 23 arnhem the netherlands mevr l. gokken
23. Unilever q-4 Jourestraat 24 arnhem the netherlands mevr l. stokken
24. Unilever q-6 Jourestraat 25 arnhem the netherlands mevr l. rokken
Unilever q-8 Jourestraat 26 arnhem the netherlands mevr l. rokken
127.0.0.1:3333/project?project=2059581926475#refine-tabs-facets

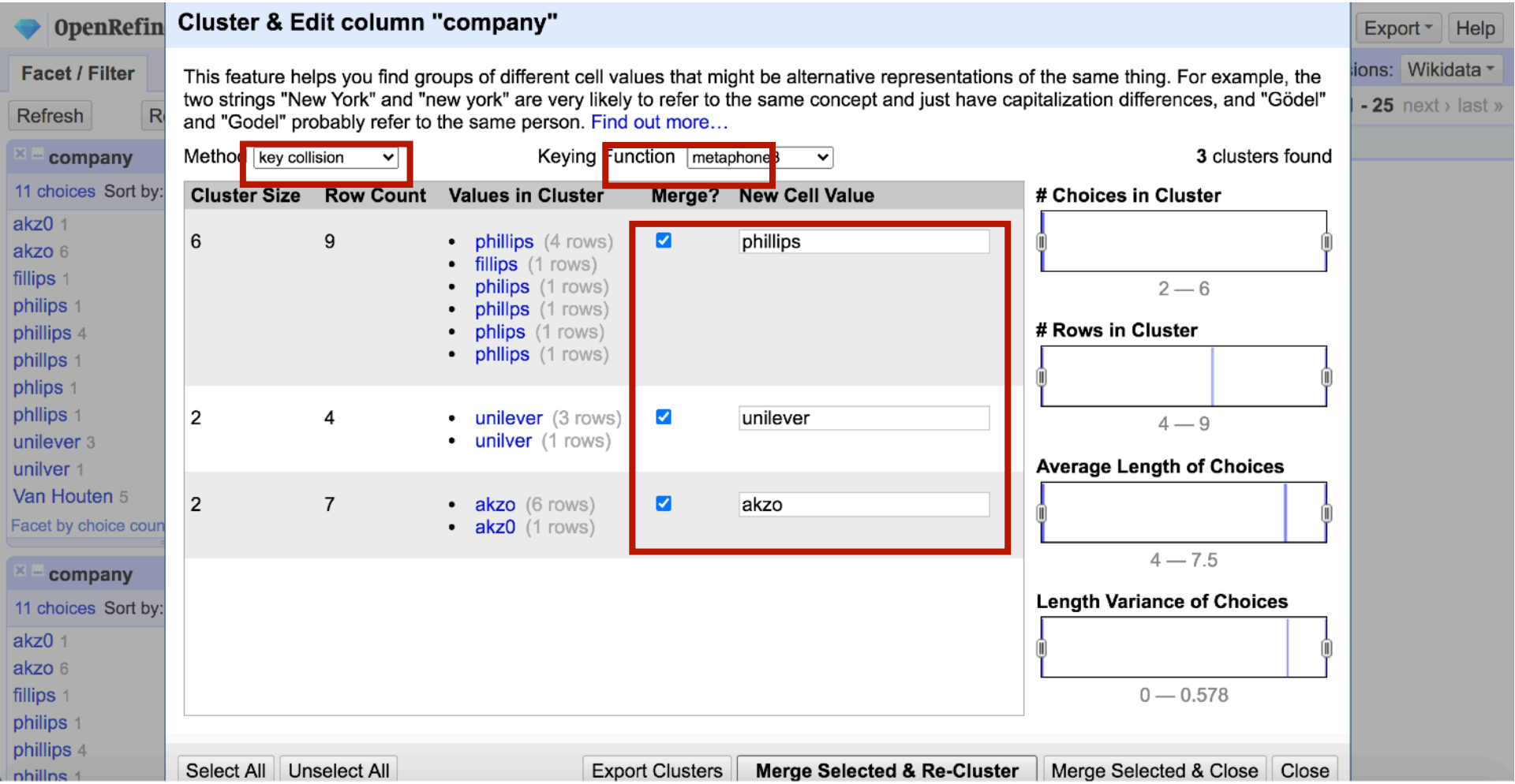

OpenRefine Refine Demo
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 13 / 13
Extract...
Apply...
Filter:
grel:value.replace('p','radio')
8. Text transform on 8 cells in column Product Name: grel:value.replace('x','computer')
9. Text transform on 6 cells in column Product Name: grel:value.replace('v','tv')
10. Text transform on 5 cells in column Product Name: grel:value.replace('q','tablet')
11. Create new column full address based on column address by filling 25 rows with grel:value + ', ' + cells['city'].value + ', ' + cells['country'].value
12. Text transform on 25 cells in column city: value.toTitlecase()
13. Text transform on 25 cells in column country: value.toTitlecase()
25 rows
Extensions: Wikidata
Show as: rows records Show: 5 10 25 50 rows
« first ‹ previous 1 - 25 next › last »
All
company
Product Nam
Product code
address
full address
city
country
name
1. Phillips radio 5 Groningensingel 147 Groningensingel 147, arnhem, the netherlands Arnhem The Netherlands dhr p. jansen
2. Phillips radio 43 Groningensingel 148 Groningensingel 148, arnhem, the netherlands Arnhem The Netherlands dhr p. hansen
3. Phillips computer 3 Groningensingel 149 Groningensingel 149, arnhem, the netherlands Arnhem The Netherlands dhr j. Gansen
4. Phillips computer 34 Groningensingel 150 Groningensingel 150, arnhem, the netherlands Arnhem The Netherlands dhr p. mansen
5. Phillips computer 12 Groningensingel 151 Groningensingel 151, arnhem, the netherlands Arnhem The Netherlands dhr p. fransen
6. Phillips radio 23 Groningensingel 152 Groningensingel 152, arnhem, the netherlands Arnhem The Netherlands dhr p. franssen
7. Akzo tv 43 Leeuwardenweg 178 Leeuwardenweg 178, arnhem, the netherlands Arnhem The Netherlands dhr p. bansen
8. Akzo tv 12 Leeuwardenweg 179 Leeuwardenweg 179, arnhem, the netherlands Arnhem The Netherlands dhr p. vansen
9. Akzo computer 5 Leeuwardenweg 180 Leeuwardenweg 180, arnhem, the netherlands Arnhem The Netherlands dhr p. bransen
10. Akzo radio 34 Leeuwardenweg 181 Leeuwardenweg 181, arnhem, the netherlands Arnhem The Netherlands dhr p. janssen
11. Akzo tablet 5 Leeuwardenweg 182 Leeuwardenweg 182, arnhem, the netherlands Arnhem The Netherlands mevr l. rokken
12. Akzo tablet 9 Leeuwardenweg 183 Leeuwardenweg 183, arnhem, the netherlands Arnhem The Netherlands mevr l. lokken
13. Akzo computer 8 Leeuwardenweg 184 Leeuwardenweg 184, arnhem, the netherlands Arnhem The Netherlands mevr l. mokken
14. Phillips radio 56 Delfzijlstraat 54 Delfzijlstraat 54, arnhem, the netherlands Arnhem The Netherlands mevr l. mokken

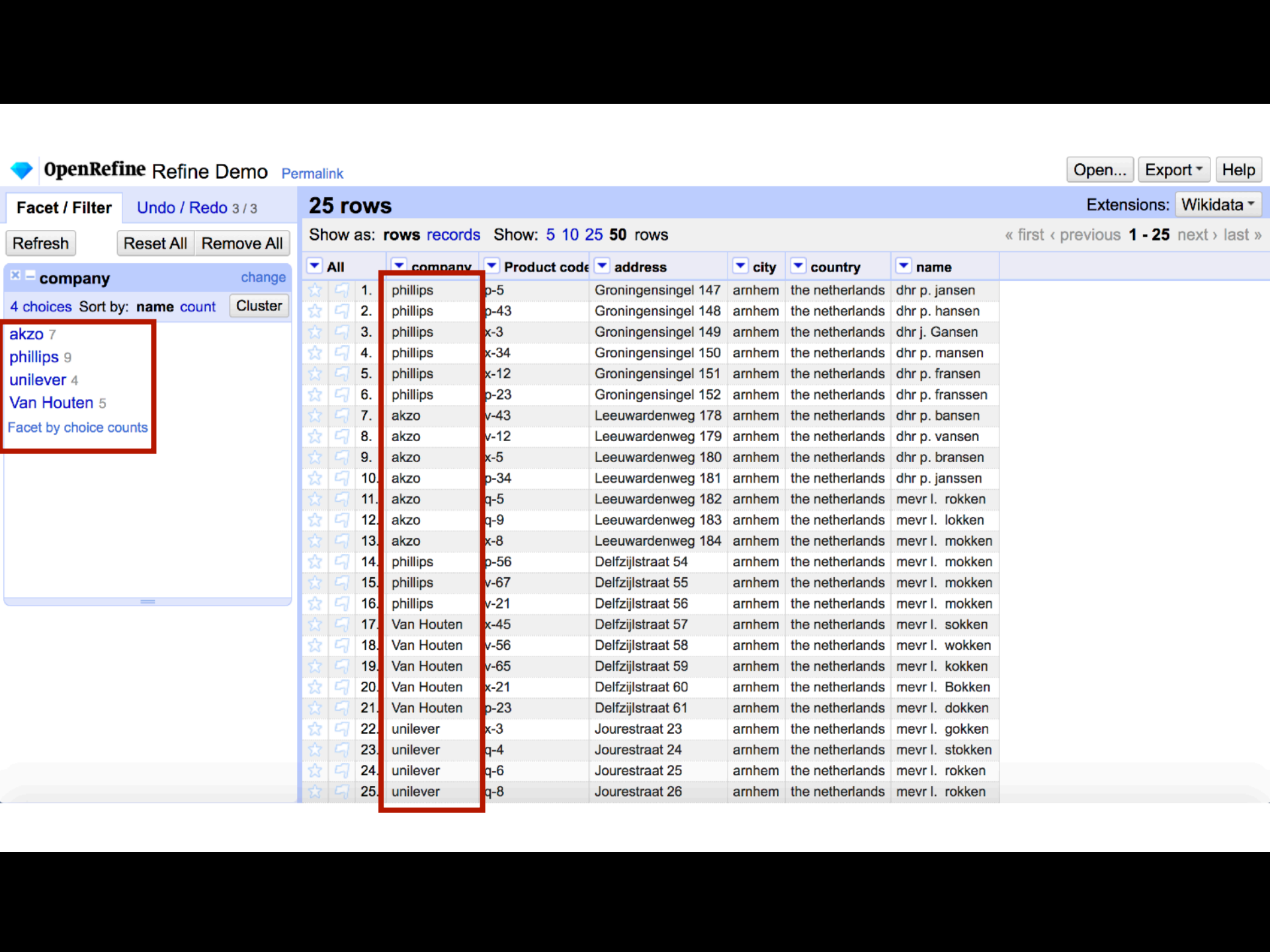

OpenRefine Refine Demo
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 0 / 13
25 rows
Extensions:
Wikidata
Refresh
Reset All
Remove All
Show as: rows records
Show: 5 10 25 50 rows
« first ‹ previous 1 - 25 next › last »
company
change
19 choices Sort by: name count
Cluster
ak zo 1
akz0 1
Akzo 1
AKZO 1
akzo 3
fillips 1
philips 1
phillips 2
phillipS 1
Phillips 1
phillps 1
phlips 1
All
company
Product code
address
city
country
name
Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile
Text facet
Numeric facet
Timeline facet
Scatterplot facet
Custom text facet...
Custom Numeric Facet...
Customized facets
1. Phillips
2. phillips
3. philips
4. phllips
5. phillps
6. phillipS
7. akzo
8. Akzo
9. AKZO
10. akz0
11. ak zo
12. akzo q-9 Leeuwardenweg 183 arnhem the netherlands mevr l. lokken
13. akzo x-8 Leeuwardenweg 184 arnhem the netherlands mevr l. mokken
14. phillips p-56 Delfzijlstraat 54 arnhem the netherlands mevr l. mokken
15. fillips v-67 Delfzijlstraat 55 arnhem the netherlands mevr l. mokken
16. phlips v-21 Delfzijlstraat 56 arnhem the netherlands mevr l. mokken
17. Van Houten x-45 Delfzijlstraat 57 arnhem the netherlands mevr l. sokken
18. van Houten v-56 Delfzijlstraat 58 arnhem the netherlands mevr l. wokken
19. van houten v-65 Delfzijlstraat 59 arnhem the netherlands mevr l. kokken
20. van houten x-21 Delfzijlstraat 60 arnhem the netherlands mevr l. Bokken
21. Van Houten p-23 Delfzijlstraat 61 arnhem the netherlands mevr l. dokken
22. unilver x-3 Jourestraat 23 arnhem the netherlands mevr l. gokken
23. unilever q-4 Jourestraat 24 arnhem the netherlands mevr l. stokken
24. Unilever q-6 Jourestraat 25 arnhem the netherlands mevr l. rokken
25. unilever q-8 Jourestraat 26 arnhem the netherlands mevr l. rokken
javascript:{}

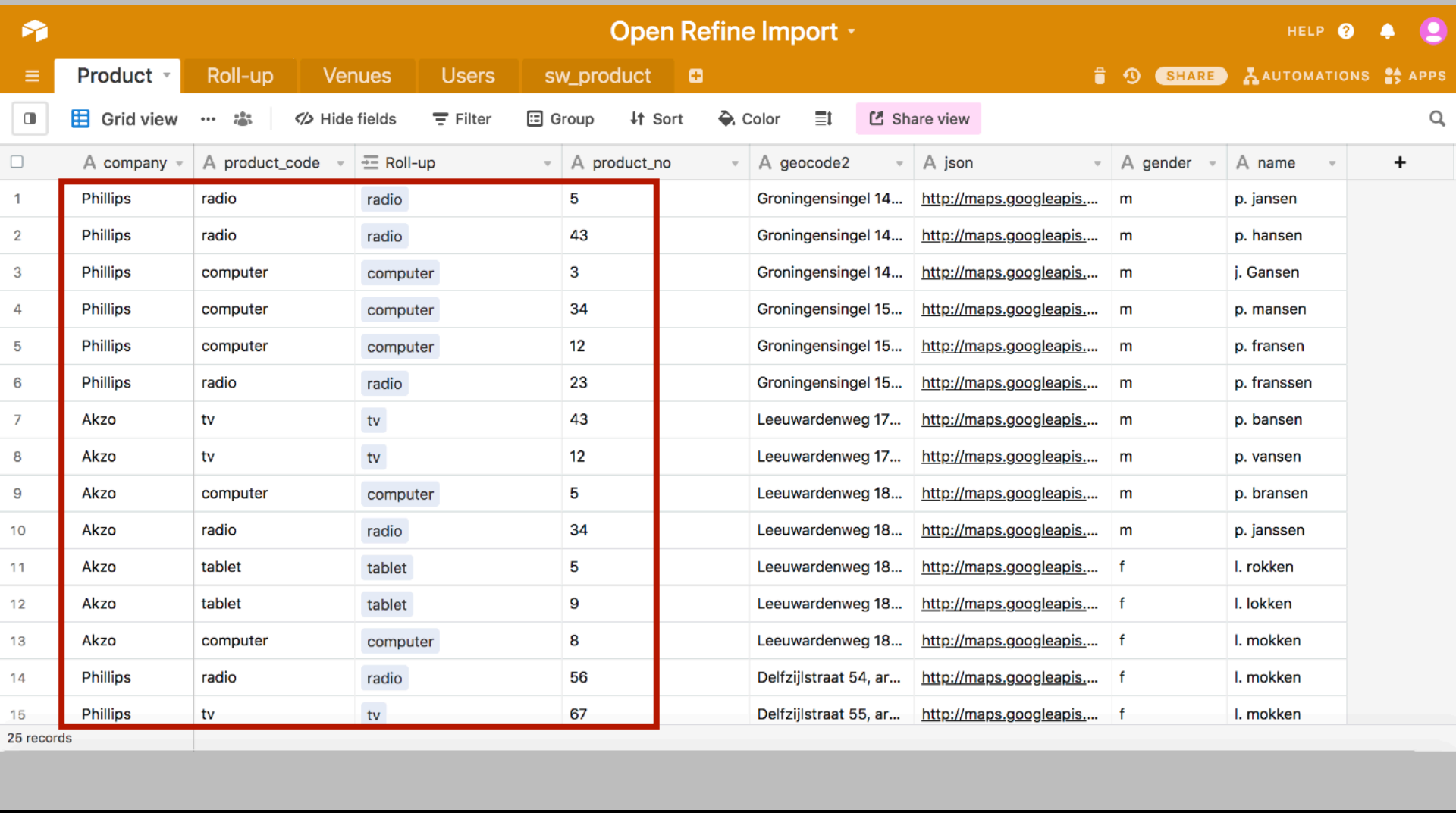

OpenRefin
Facet / Filter
Refresh
company
19 choices Sort by:
ak zo 1
akz0 1
Akzo 1
AKZO 1
akzo 3
fillips 1
philips 1
phillips 2
phillipS 1
Phillips 1
phillps 1
phlips 1
Export
Help
Wikidata
- 25 next › last »
Cluster & Edit column "company"
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more…
Method key collision
Keying Function fingerprint
4 clusters found
Cluster Size
Row Count
Values in Cluster
Merge?
New Cell Value
3 5 akzo (3 rows) AKZO (1 rows) Akzo (1 rows) akzo
3 5 Van Houten (2 rows) van houten (2 rows) van Houten (1 rows) Van Houten
3 4 phillips (2 rows) Phillips (1 rows) phillipS (1 rows) phillips
2 3 unilever (2 rows) Unilever (1 rows) unilever
Choices in Cluster
2 — 3
Rows in Cluster
3 — 5
Average Length of Choices
4 — 10
Select All
Unselect All
Export Clusters
Merge Selected & Re-Cluster
Merge Selected & Close
Close


OpenRefin
Facet / Filter
Refresh
company
19 choices Sort by:
ak zo 1
akz0 1
Akzo 1
AKZO 1
akzo 3
fillips 1
philips 1
phillips 2
phillipS 1
Phillips 1
phillps 1
phlips 1
Cluster & Edit column "company"
Export
Help
Wikidata
- 25 next › last »
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more…
Method key collision
Keying Function fingerprint
4 clusters found
Cluster Size
Row Count
Values in Cluster
Merge?
New Cell Value
3
5
akzo (3 rows)
AKZO (1 rows)
Akzo (1 rows)
akzo
3
5
Van Houten (2 rows)
van houten (2 rows)
van Houten (1 rows)
Van Houten
3
4
phillips (2 rows)
Phillips (1 rows)
phillipS (1 rows)
phillips
2
3
unilever (2 rows)
Unilever (1 rows)
unilever
Choices in Cluster
2 — 3
Rows in Cluster
3 — 5
Average Length of Choices
4 — 10
Select All
Unselect All
Export Clusters
Merge Selected & Re-Cluster
Merge Selected & Close
Close


OpenRefin
Facet / Filter
Refresh
company
12 choices Sort by:
ak zo 1
akz0 1
akzo 5
fillips 1
philips 1
phillips 4
phillps 1
phlips 1
phllips 1
unilever 3
unilver 1
Van Houten 5
company
12 choices Sort by:
ak zo 1
akz0 1
akzo 5
fillips 1
philips 1
Export
Help
Wikidata
- 25 next › last »
Cluster & Edit column "company"
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more…
Metho
key collision
Keying Function
ngram-fingerprint
Ngram Size 2
1 cluster found
Cluster Size
Row Count
Values in Cluster
Merge?
New Cell Value
2
6
akzo (5 rows)
ak zo (1 rows)
akzo
Select All
Unselect All
Export Clusters
Merge Selected & Re-Cluster
Merge Selected & Close
Close

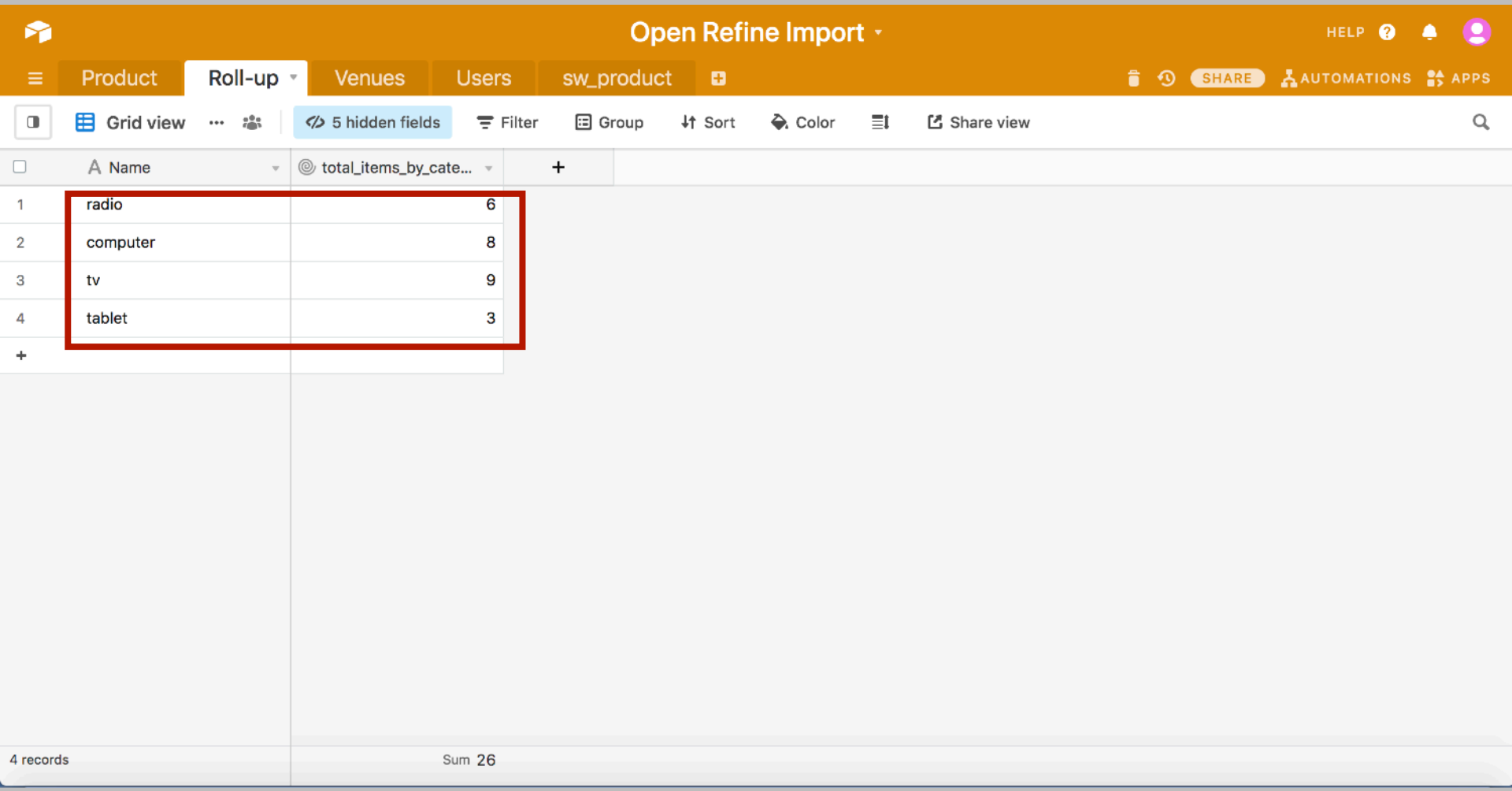

OpenRefin
Facet / Filter
Refresh
company
11 choices Sort by:
akz0 1
akzo 6
fillips 1
philips 1
phillips 4
phillps 1
phlips 1
phllips 1
unilever 3
unilver 1
Van Houten 5
Facet by choice coun
company
11 choices Sort by:
akz0 1
akzo 6
fillips 1
philips 1
phillips 4
Export
Help
ions:
Wikidata
- 25 next › last »
Cluster & Edit column "company"
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more…
Method key collision
Keying Function metaphone3
3 clusters found
Cluster Size
Row Count
Values in Cluster
Merge?
New Cell Value
6
9
phillips (4 rows)
fillips (1 rows)
philips (1 rows)
phillps (1 rows)
phlips (1 rows)
phllips (1 rows)
phillips
2
4
unilever (3 rows)
unilver (1 rows)
unilever
2
7
akzo (6 rows)
akz0 (1 rows)
akzo
Choices in Cluster
2 — 6
Rows in Cluster
4 — 9
Average Length of Choices
4 — 7.5
Length Variance of Choices
0 — 0.578
Select All
Unselect All
Export Clusters
Merge Selected & Re-Cluster
Merge Selected & Close
Close

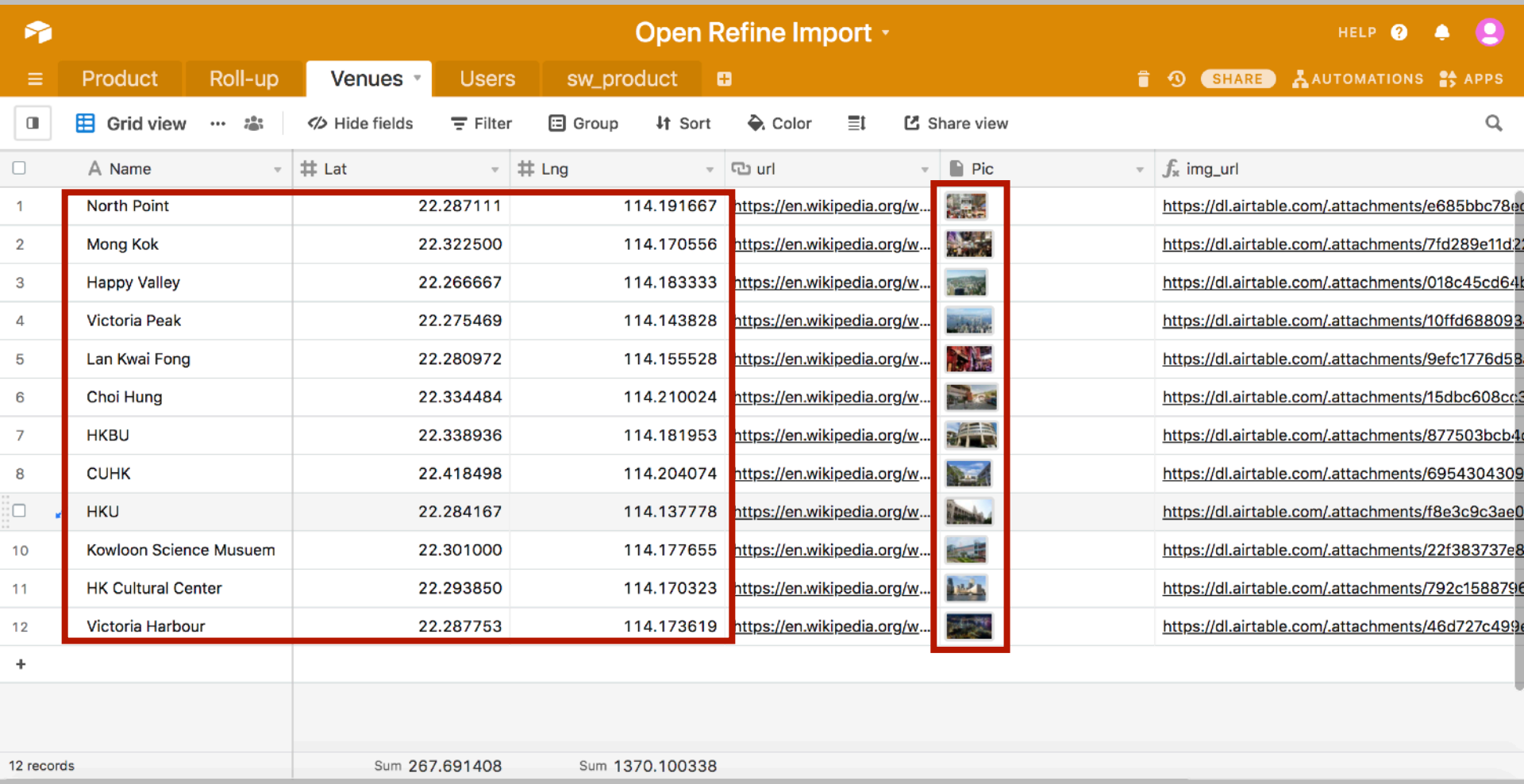

OpenRefine Refine Demo
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 3 / 3
Refresh
Reset All
Remove All
company
change
4 choices Sort by: name count
Cluster
akzo 7
phillips 9
unilever 4
Van Houten 5
Facet by choice counts
25 rows
Extensions: Wikidata
Show as: rows records Show: 5 10 25 50 rows
« first ‹ previous 1 - 25 next › last »
All
company
Product code
address
city
country
name
1. phillips p-5 Groningensingel 147 arnhem the netherlands dhr p. jansen
2. phillips p-43 Groningensingel 148 arnhem the netherlands dhr p. hansen
3. phillips x-3 Groningensingel 149 arnhem the netherlands dhr j. Gansen
4. phillips x-34 Groningensingel 150 arnhem the netherlands dhr p. mansen
5. phillips x-12 Groningensingel 151 arnhem the netherlands dhr p. fransen
6. phillips p-23 Groningensingel 152 arnhem the netherlands dhr p. franssen
7. akzo v-43 Leeuwardenweg 178 arnhem the netherlands dhr p. bansen
8. akzo v-12 Leeuwardenweg 179 arnhem the netherlands dhr p. vansen
9. akzo x-5 Leeuwardenweg 180 arnhem the netherlands dhr p. bransen
10. akzo p-34 Leeuwardenweg 181 arnhem the netherlands dhr p. janssen
11. akzo q-5 Leeuwardenweg 182 arnhem the netherlands mevr l. rokken
12. akzo q-9 Leeuwardenweg 183 arnhem the netherlands mevr l. lokken
13. akzo x-8 Leeuwardenweg 184 arnhem the netherlands mevr l. mokken
14. phillips p-56 Delfzijlstraat 54 arnhem the netherlands mevr l. mokken
15. phillips v-67 Delfzijlstraat 55 arnhem the netherlands mevr l. mokken
16. phillips v-21 Delfzijlstraat 56 arnhem the netherlands mevr l. mokken
17. Van Houten x-45 Delfzijlstraat 57 arnhem the netherlands mevr l. sokken
18. Van Houten v-56 Delfzijlstraat 58 arnhem the netherlands mevr l. wokken
19. Van Houten v-65 Delfzijlstraat 59 arnhem the netherlands mevr l. kokken
20. Van Houten x-21 Delfzijlstraat 60 arnhem the netherlands mevr l. Bokken
21. Van Houten p-23 Delfzijlstraat 61 arnhem the netherlands mevr l. dokken
22. unilever x-3 Jourestraat 23 arnhem the netherlands mevr l. gokken
23. unilever q-4 Jourestraat 24 arnhem the netherlands mevr l. stokken
24. unilever q-6 Jourestraat 25 arnhem the netherlands mevr l. rokken
25. unilever q-8 Jourestraat 26 arnhem the netherlands mevr l. rokken

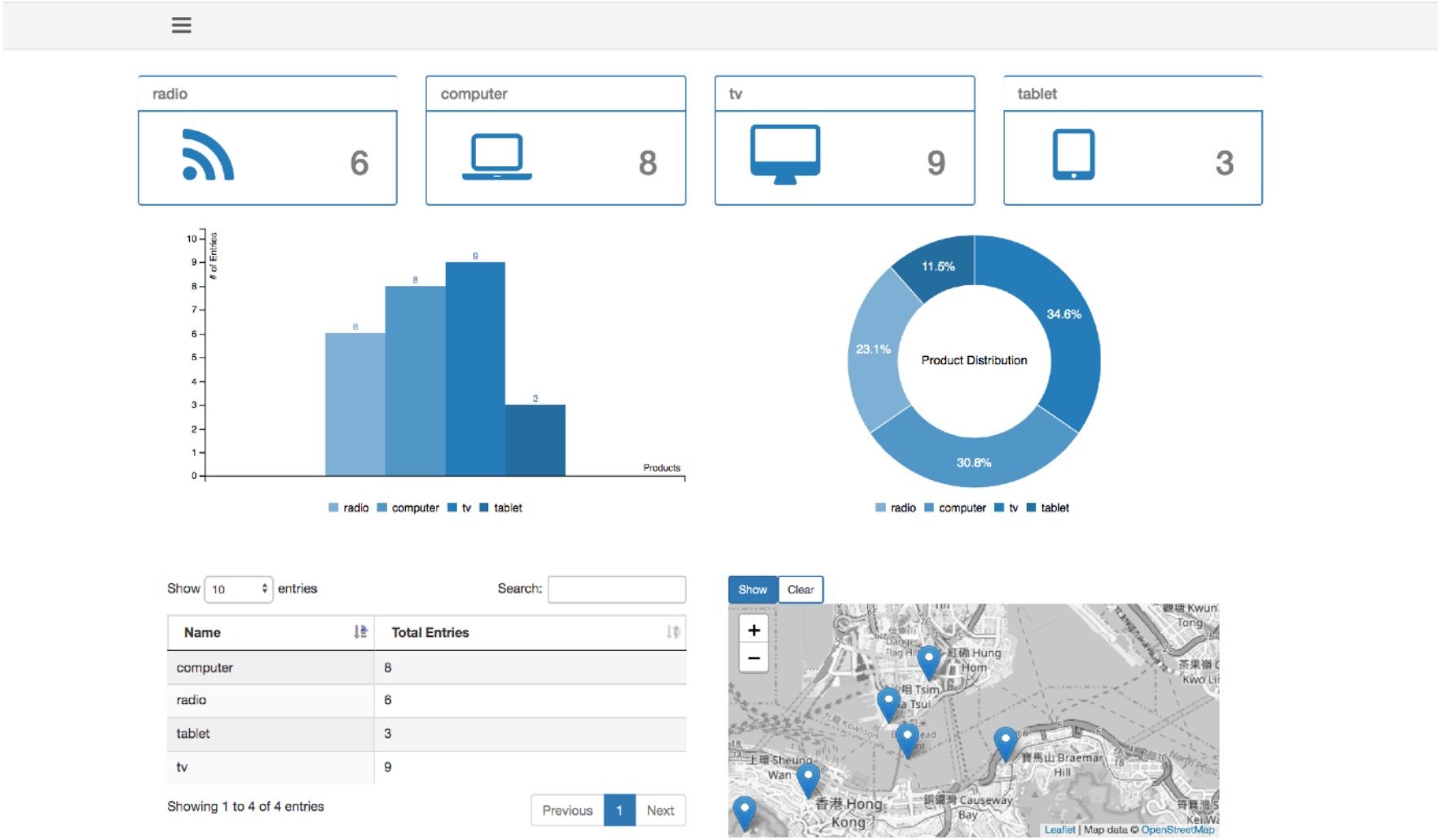


	company	product_code	Roll-up	product_no	geocode2	json	gender	name
1	Phillips	radio	radio	5	Groningensingel 14...	http://maps.googleapis....	m	p. jansen
2	Phillips	radio	radio	43	Groningensingel 14...	http://maps.googleapis....	m	p. hansen
3	Phillips	computer	computer	3	Groningensingel 14...	http://maps.googleapis....	m	j. Gansen
4	Phillips	computer	computer	34	Groningensingel 15...	http://maps.googleapis....	m	p. mansen
5	Phillips	computer	computer	12	Groningensingel 15...	http://maps.googleapis....	m	p. fransen
6	Phillips	radio	radio	23	Groningensingel 15...	http://maps.googleapis....	m	p. franssen
7	Akzo	tv	tv	43	Leeuwardenweg 17...	http://maps.googleapis....	m	p. bansen
8	Akzo	tv	tv	12	Leeuwardenweg 17...	http://maps.googleapis....	m	p. vansen
9	Akzo	computer	computer	5	Leeuwardenweg 18...	http://maps.googleapis....	m	p. bransen
10	Akzo	radio	radio	34	Leeuwardenweg 18...	http://maps.googleapis....	m	p. janssen
11	Akzo	tablet	tablet	5	Leeuwardenweg 18...	http://maps.googleapis....	f	l. rokken
12	Akzo	tablet	tablet	9	Leeuwardenweg 18...	http://maps.googleapis....	f	l. lokken
13	Akzo	computer	computer	8	Leeuwardenweg 18...	http://maps.googleapis....	f	l. mokken
14	Phillips	radio	radio	56	Delfzijlstraat 54, ar...	http://maps.googleapis....	f	l. mokken
15	Phillips	tv	tv	67	Delfzijlstraat 55, ar...	http://maps.googleapis....	f	l. mokken

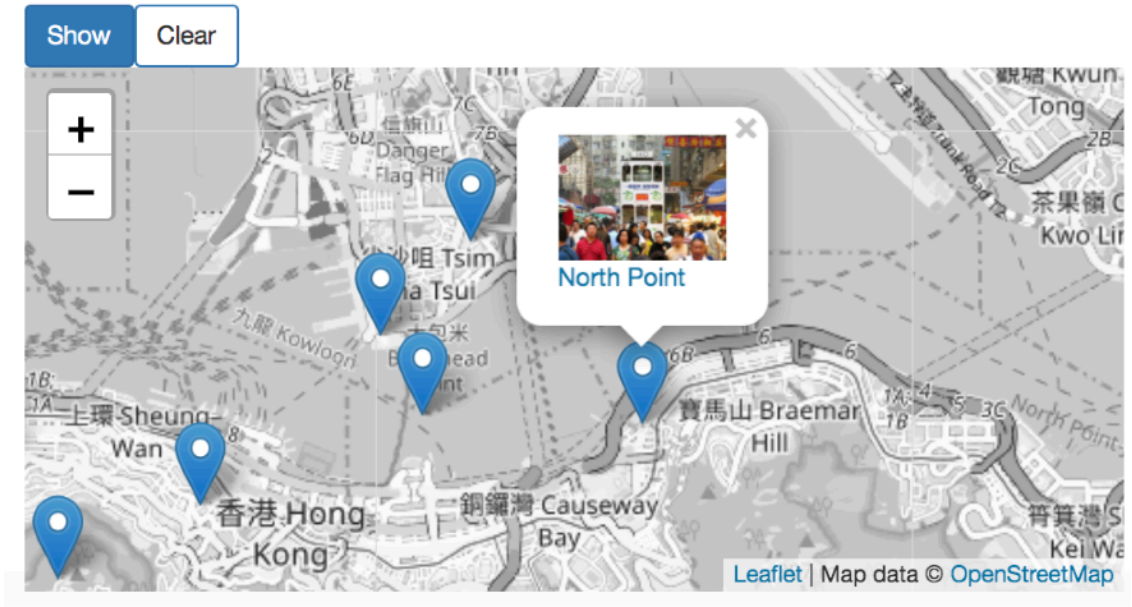
25 records


Open Refine Import
HELP
Product
Roll-up
Venues
Users
sw_product
SHARE
AUTOMATIONS
APPS
Grid view
5 hidden fields
Filter
Group
Sort
Color
Share view
Name
total_items_by_cate...
1 radio 6
2 computer 8
3 tv 9
4 tablet 3
4 records
Sum 26


Open Refine Import
HELP
Product
Roll-up
Venues
Users
sw_product
SHARE
AUTOMATIONS
APPS
Grid view
Hide fields
Filter
Group
Sort
Color
Share view
Name
Lat
Lng
url
Pic
img_url
1 North Point 22.287111 114.191667 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/e685bbc78e
2 Mong Kok 22.322500 114.170556 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/7fd289e11d
3 Happy Valley 22.266667 114.183333 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/018c45cd64
4 Victoria Peak 22.275469 114.143828 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/10ffd688093
5 Lan Kwai Fong 22.280972 114.155528 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/9efc1776d58
6 Choi Hung 22.334484 114.210024 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/15dbc608cc
7 HKBU 22.338936 114.181953 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/877503bcb4
8 CUHK 22.418498 114.204074 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/6954304309
HKU 22.284167 114.137778 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/f8e3c9c3ae0
10 Kowloon Science Musuem 22.301000 114.177655 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/22f383737e8
11 HK Cultural Center 22.293850 114.170323 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/792c158879
12 Victoria Harbour 22.287753 114.173619 https://en.wikipedia.org/w... https://dl.airtable.com/.attachments/46d727c499
12 records
Sum 267.691408
Sum 1370.100338

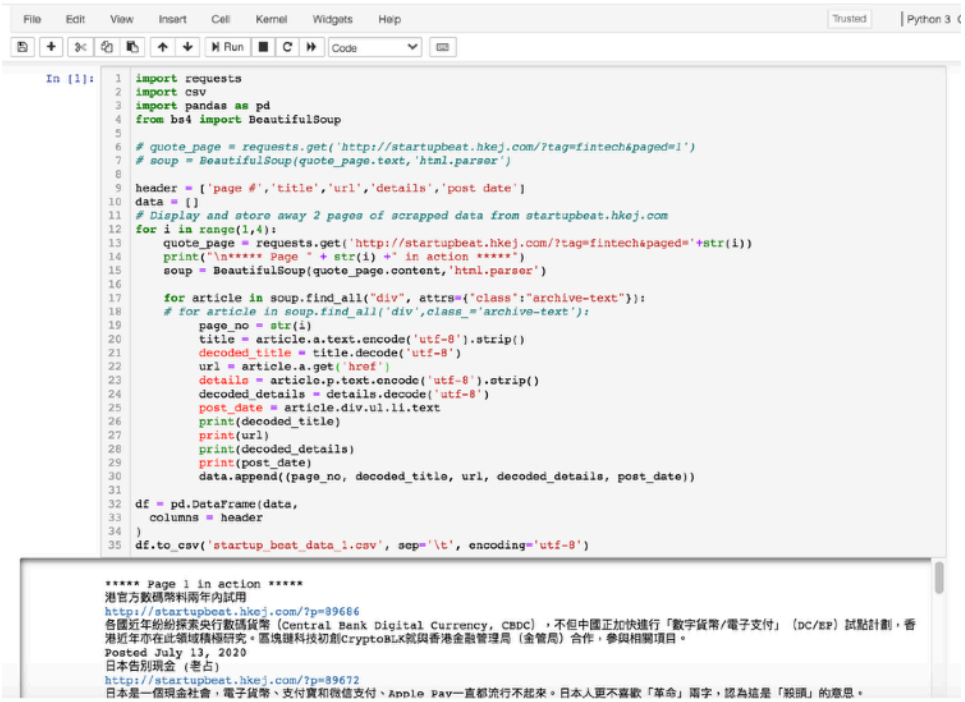


https://python30.cuhkcfe.io/



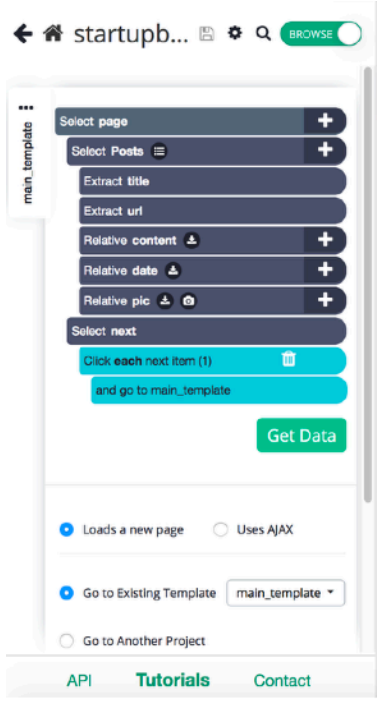
Geospatial Data Visualization

Web Scrapping



```
In [1]:
import requests
import csv
import pandas as pd
from bs4 import BeautifulSoup

# quote_page = requests.get('http://startupbeat.hkej.com/?tag=fintech&paged=1')
# soup = BeautifulSoup(quote_page.text,'html.parser')

header = ['page #','title','url','details','post date']
data = []
# Display and store away 2 pages of scrapped data from startupbeat.hkej.com
for i in range(1,4):
    quote_page = requests.get('http://startupbeat.hkej.com/?tag=fintech&paged='+str(i))
    print("\n***** Page " + str(i) +" in action *****")
    soup = BeautifulSoup(quote_page.content,'html.parser')

    for article in soup.find_all("div", attrs={"class":"archive-text"}):
    # for article in soup.find_all('div',class_='archive-text'):
        page_no = str(i)
        title = article.a.text.encode('utf-8').strip()
        decoded_title = title.decode('utf-8')
        url = article.a.get('href')
        details = article.p.text.encode('utf-8').strip()
        decoded_details = details.decode('utf-8')
        post_date = article.div.ul.li.text
        print(decoded_title)
        print(url)
        print(decoded_details)
        print(post_date)
        data.append((page_no, decoded_title, url, decoded_details, post_date))

df = pd.DataFrame(data,
  columns = header
)
df.to_csv('startup_beat_data_1.csv', sep='\t', encoding='utf-8')
```

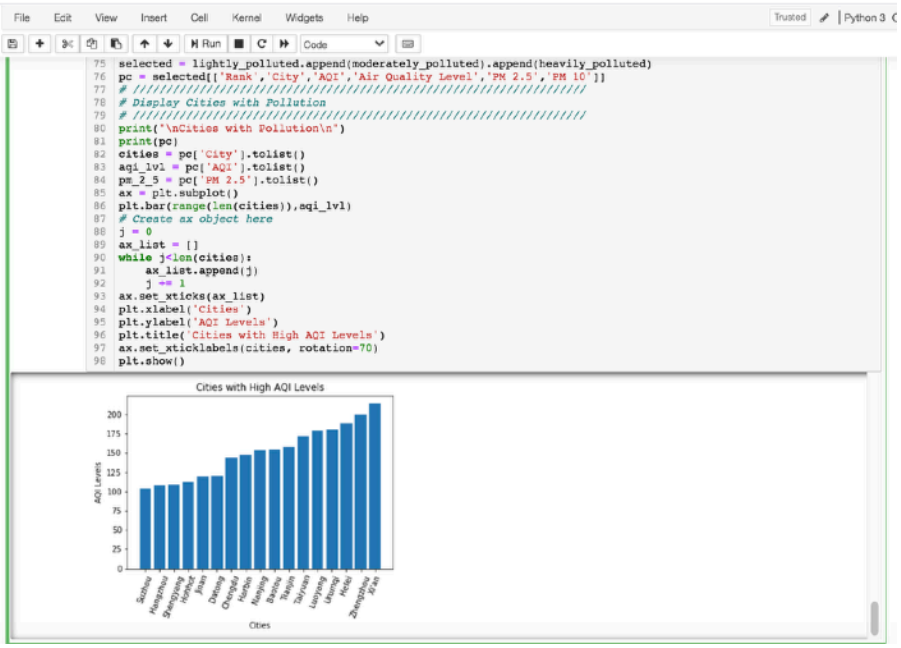
```
***** Page 1 in action *****
港官方數碼幣料兩年內試用
http://startupbeat.hkej.com/?p=89686
各國近年紛紛探索央行數碼貨幣（Central Bank Digital Currency，CBDC），不但中國正加快進行「數字貨幣/電子支付」（DC/EP）試點計劃，香港近年亦在此領域積極研究。區塊鏈科技初創CryptoBLK就與香港金融管理局（金管局）合作，參與相關項目。
Posted July 13, 2020
日本告別現金（老古）
http://startupbeat.hkej.com/?p=89672
日本是一個現金社會，電子貨幣、支付寶和微信支付、Apple Pay一直都流行不起來。日本人更不喜歡「革命」兩字，認為這是「殺頭」的意思。
```

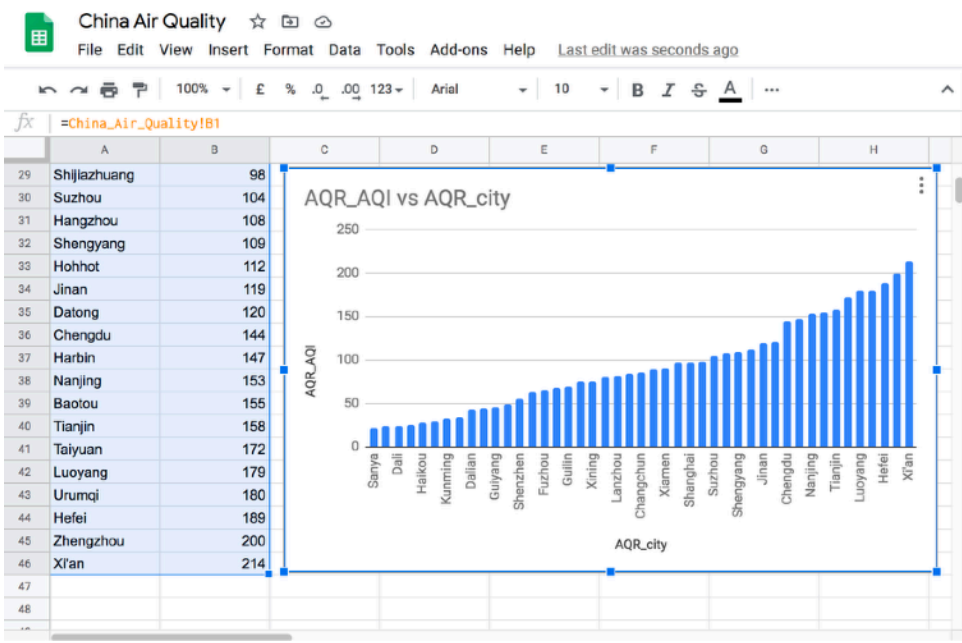


Programmer Track

Non-Programmer Track

Making Charts





	A	B
29	Shijiazhuang	98
30	Suzhou	104
31	Hangzhou	108
32	Shengyang	109
33	Hohhot	112
34	Jinan	119
35	Datong	120
36	Chengdu	144
37	Harbin	147
38	Nanjing	153
39	Baotou	155
40	Tianjin	158
41	Taiyuan	172
42	Luoyang	179
43	Urumqi	180
44	Hefei	189
45	Zhengzhou	200
46	Xi'an	214

Programmer Track

Non-Programmer Track

Tutorials on OpenRefine

Tutorial 1
https://programminghistorian.org/en/lessons/cleaning-data-with-openrefine#getting-started-installing-openrefine-and-importing-data

Practice Dataset 1
https://programminghistorian.org/assets/phm-collection.tsv

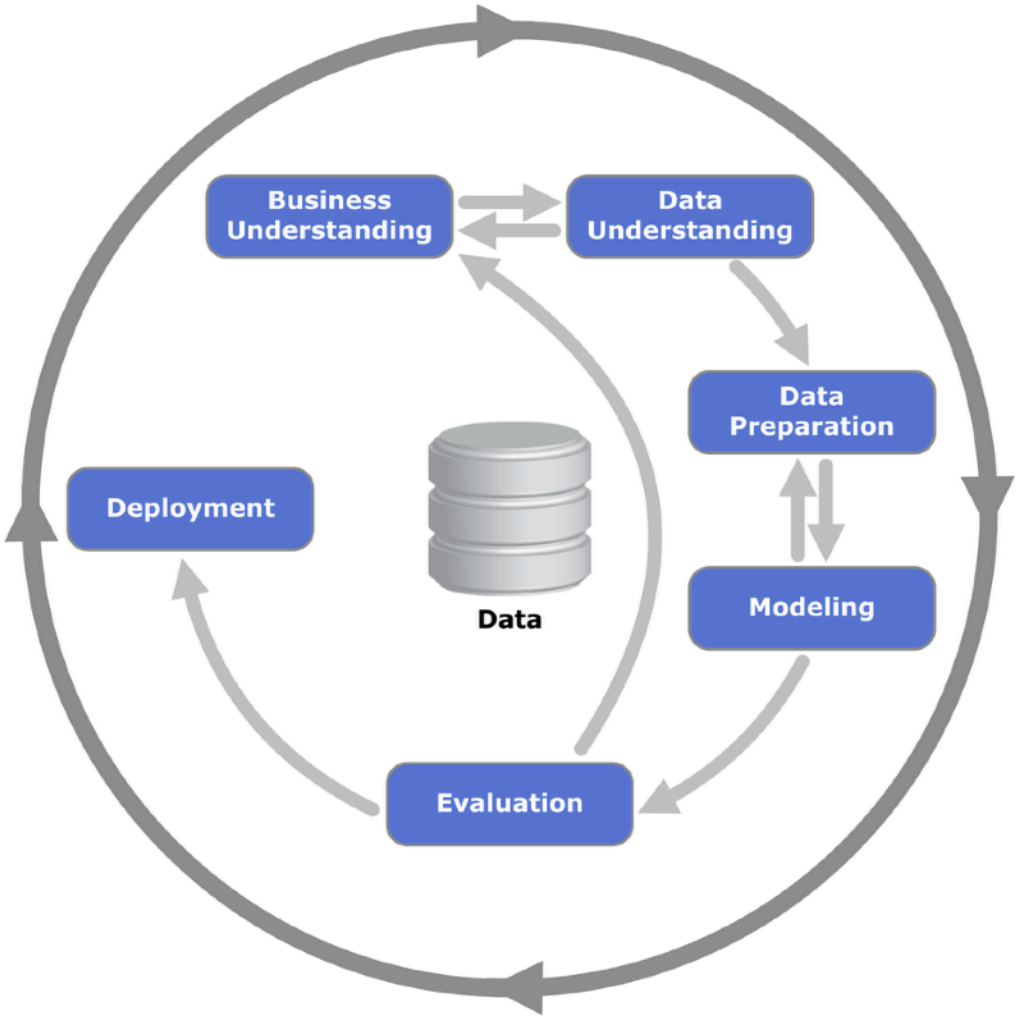
Tutorial 2

http://d3-media.blogspot.hk/2013/11/how-to-refine-your-data.html

Practice Dataset 2

https://docs.google.com/spreadsheets/d/171GoZt1fF3E1e79gb_HAh2uImq111DscD2_dCtjkNVw/edit?copiedFromTrash#gid=0

The CRISP-DM Model



CRoss Industry Standard Process for Data Mining

Harvard Business Review

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

From the October 2012 Issue

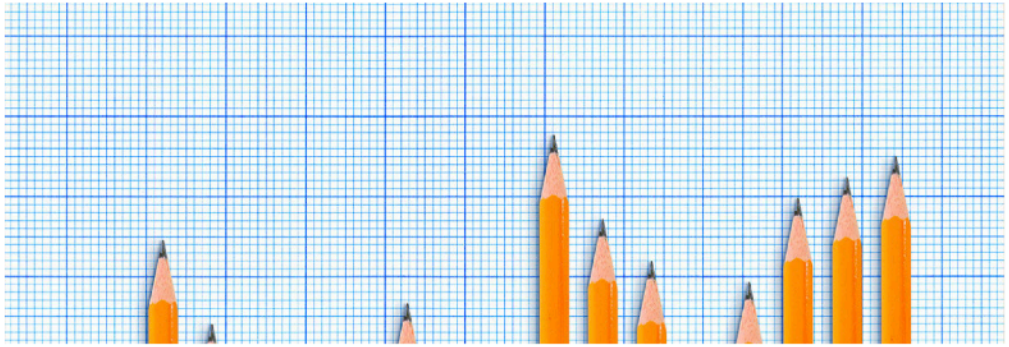
https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

ANALYTICS

What Data Scientists Really Do, According to 35 Data Scientists

by Hugo Bowne-Anderson

August 15, 2018



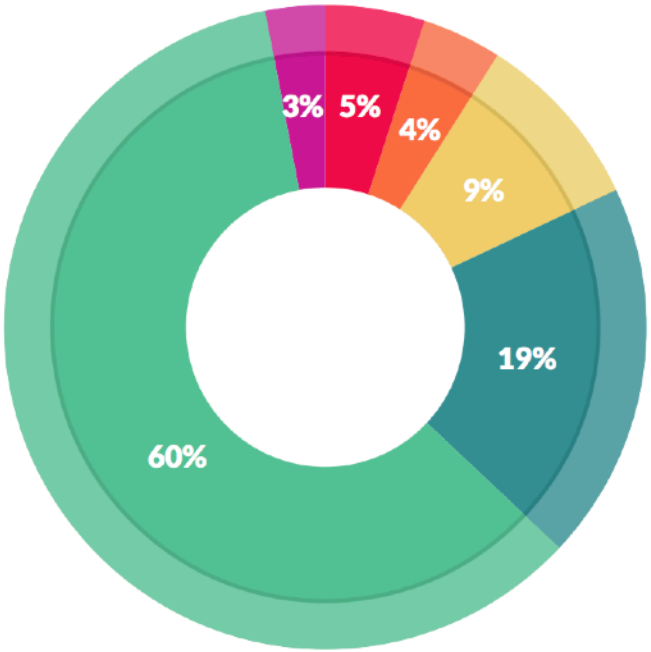
https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists?referral=03758&cm_vc=rr_item_page.top_right

Source: By Hugo Bowen-Anderson August 15, 2018 Harvard Business Review

What data scientists do. We now know how data science works, at least in the tech industry. First, data scientists lay a solid data foundation in order to perform robust analytics. Then they use online experiments, among other methods, to achieve sustainable growth. Finally, they build machine learning pipelines and personalized data products to better understand their business and customers and to make better decisions. In other words, in tech, data science is about infrastructure, testing, machine learning for decision making, and data products.

How a Data Scientist Spends Their Day

Here's where the popular view of data scientists diverges pretty significantly from reality. Generally, we think of data scientists building algorithms, exploring data, and doing predictive analysis. That's actually not what they spend most of their time doing, however.

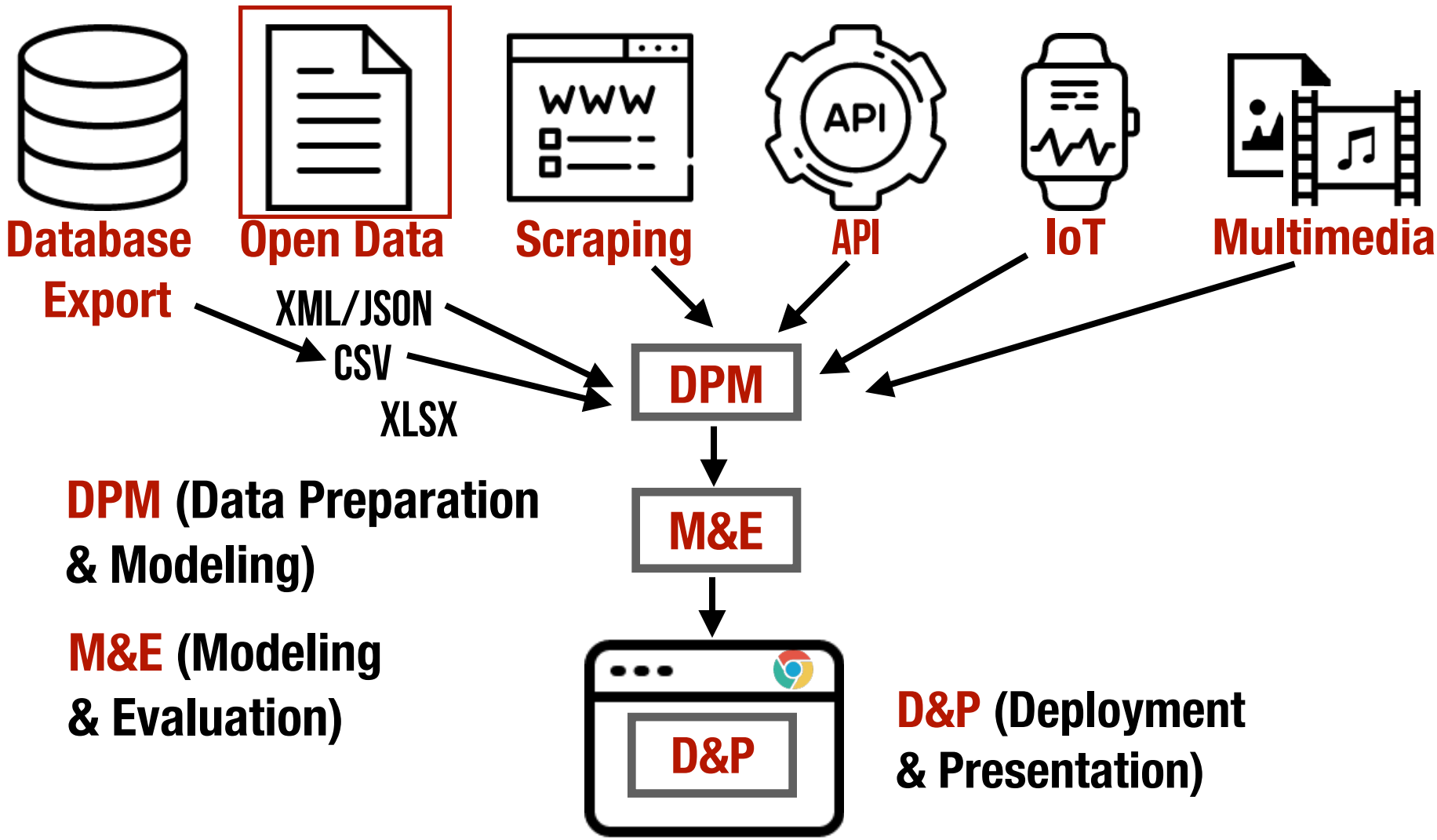


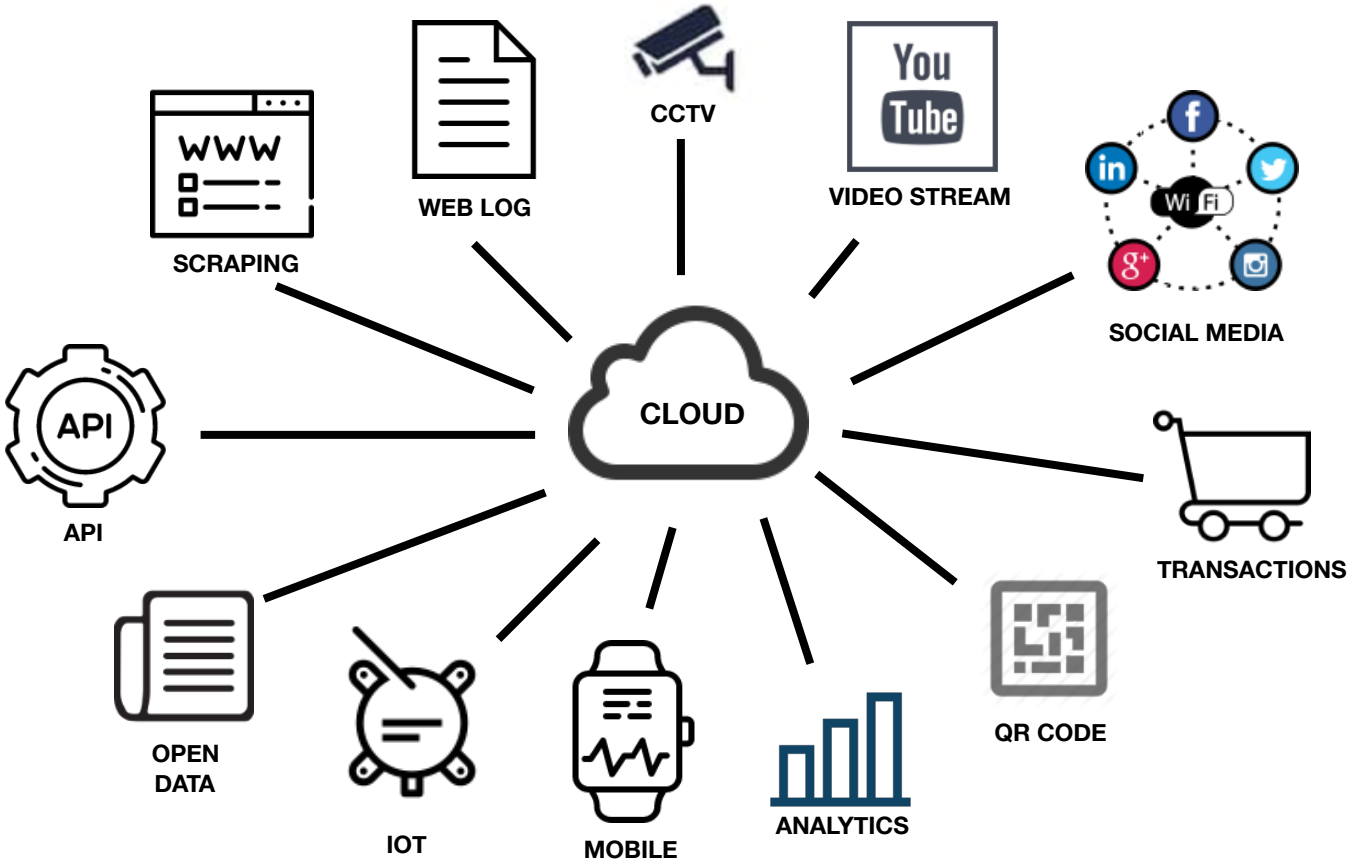
What data scientists spend the most time doing

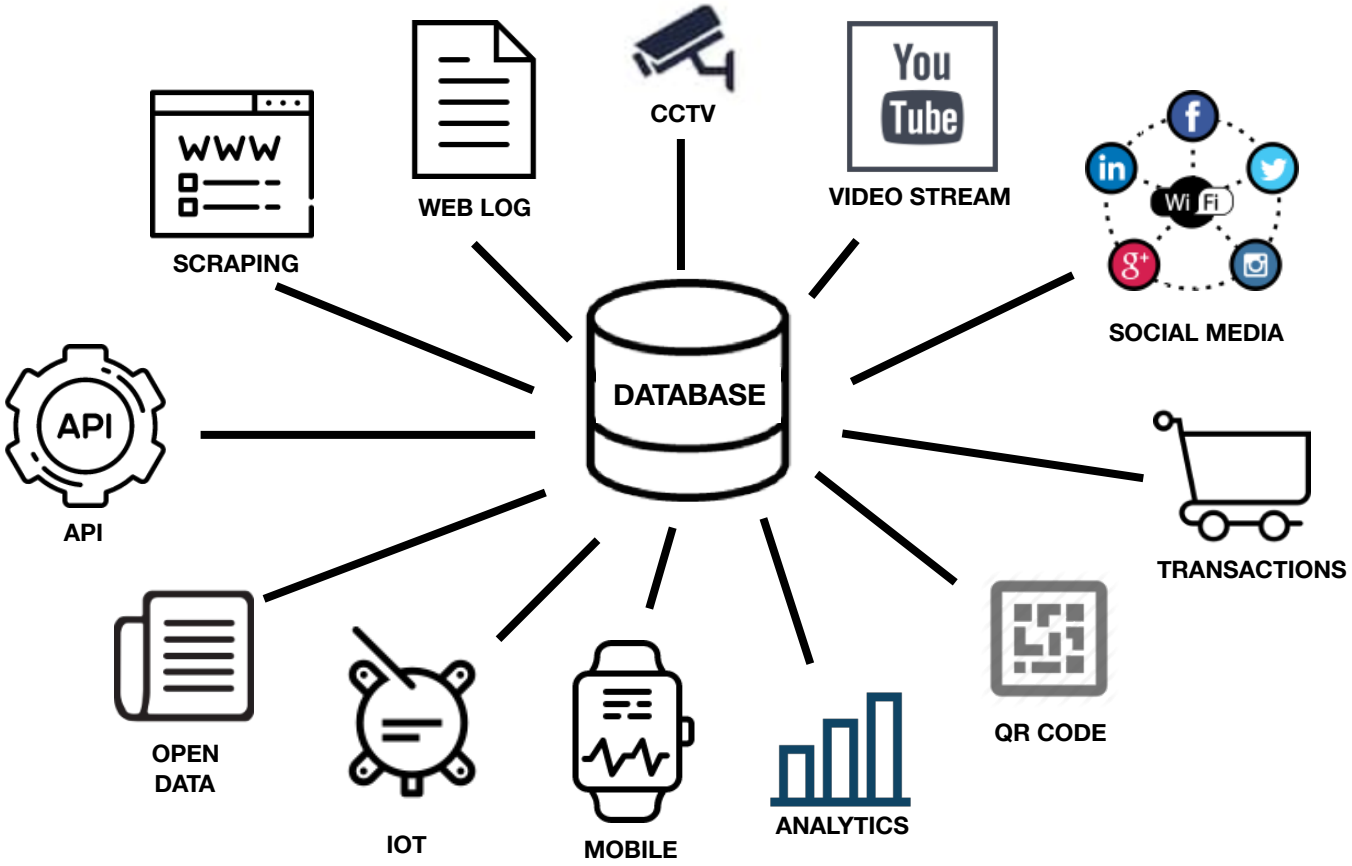
- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

Source: Data Science 2016 Report by CrowdFlower

Other sources of data besides scraping.


Database Export
Open Data
Scraping
API
IoT
Multimedia
XML/JSON
CSV
XLSX
DPM
M&E
D&P
DPM (Data Preparation & Modeling)
M&E (Modeling & Evaluation)
D&P (Deployment & Presentation)


WWW
SCRAPING
WEB LOG
CCTV
You
Tube
VIDEO STREAM
Wi Fi
SOCIAL MEDIA
API
API
CLOUD
TRANSACTIONS
OPEN
DATA
IOT
MOBILE
ANALYTICS
QR CODE


WWW
SCRAPING
WEB LOG
CCTV
You
Tube
VIDEO STREAM
Wi Fi
SOCIAL MEDIA
API
API
DATABASE
TRANSACTIONS
OPEN
DATA
IOT
MOBILE
ANALYTICS
QR CODE

NYC | OpenData 311 | Search all NYC.gov websites

NYC OpenData

Home Data About Learn Alerts Contact Us Blog

Open Data for All New Yorkers

Where can you find public Wi-Fi in your neighborhood? What kind of tree is in front of your office? Learn about where you live, work, eat, shop and play using NYC Open Data.

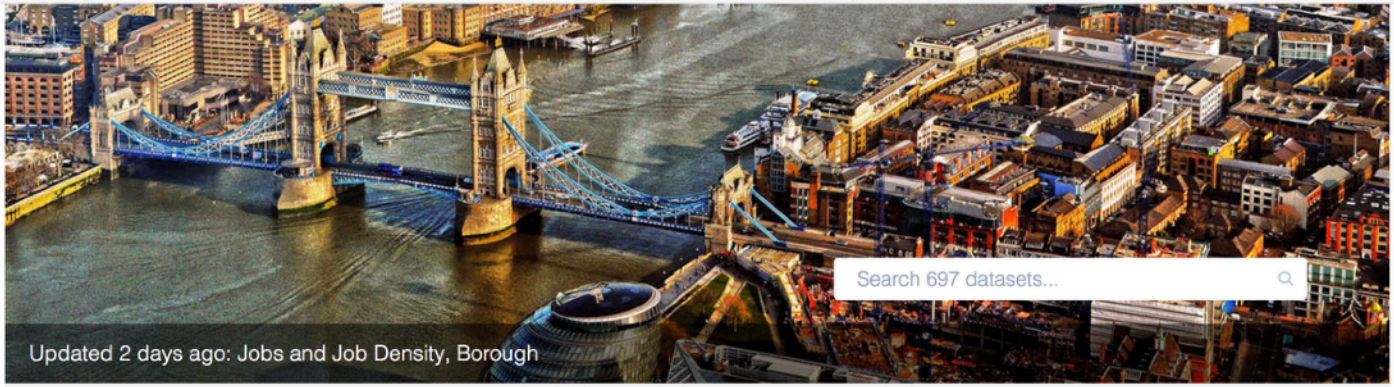
Search Open Data for things like 311, Buildings, Crime

MAYOR OF LONDON LONDONASSEMBLY

LONDON DATASTORE

Register an Account Login

Blog Data Apps & Analysis Developers Boroughs City Data Strategy More



JOBS AND ECONOMY TRANSPORT ENVIRONMENT COMMUNITY SAFETY HOUSING COMMUNITIES HEALTH LONDON AS A WORLD CITY GLA PERFORMANCE

Click on a circle to see more...

Welcome to the Datastore

The London Datastore is a free and open data-sharing portal where anyone can access data relating to the capital. Whether you're a citizen, business owner, researcher or developer, the site provides over 700 datasets to help you understand the city and develop solutions to London's problems Please do have a look around, and let us

Office of the Government Chief Information Officer
The Government of the Hong Kong Special Administrative Region

[Hong Kong
Smart City Blueprint]

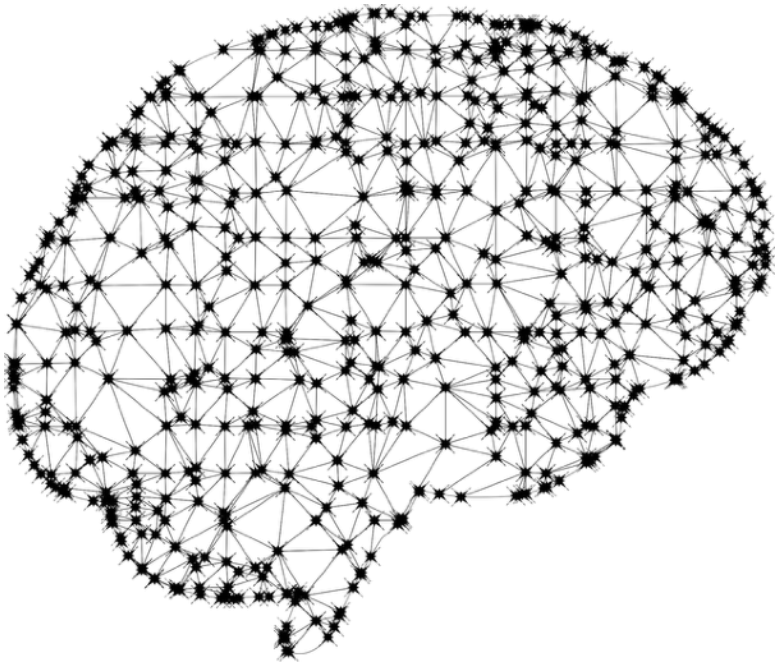
Home | Vision & Mission | Development Plans | Open e-Blueprint | Download Blueprint | 繁 簡 A A A

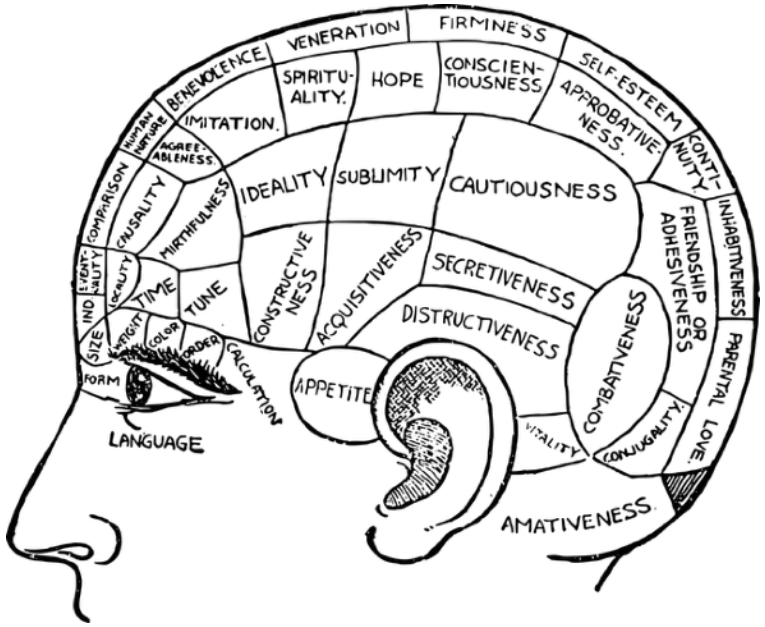
Open Data

Home > Open Data

Smart Mobility | Smart Living | Smart Environment | Smart People | Smart Government | Smart Economy | Open Data

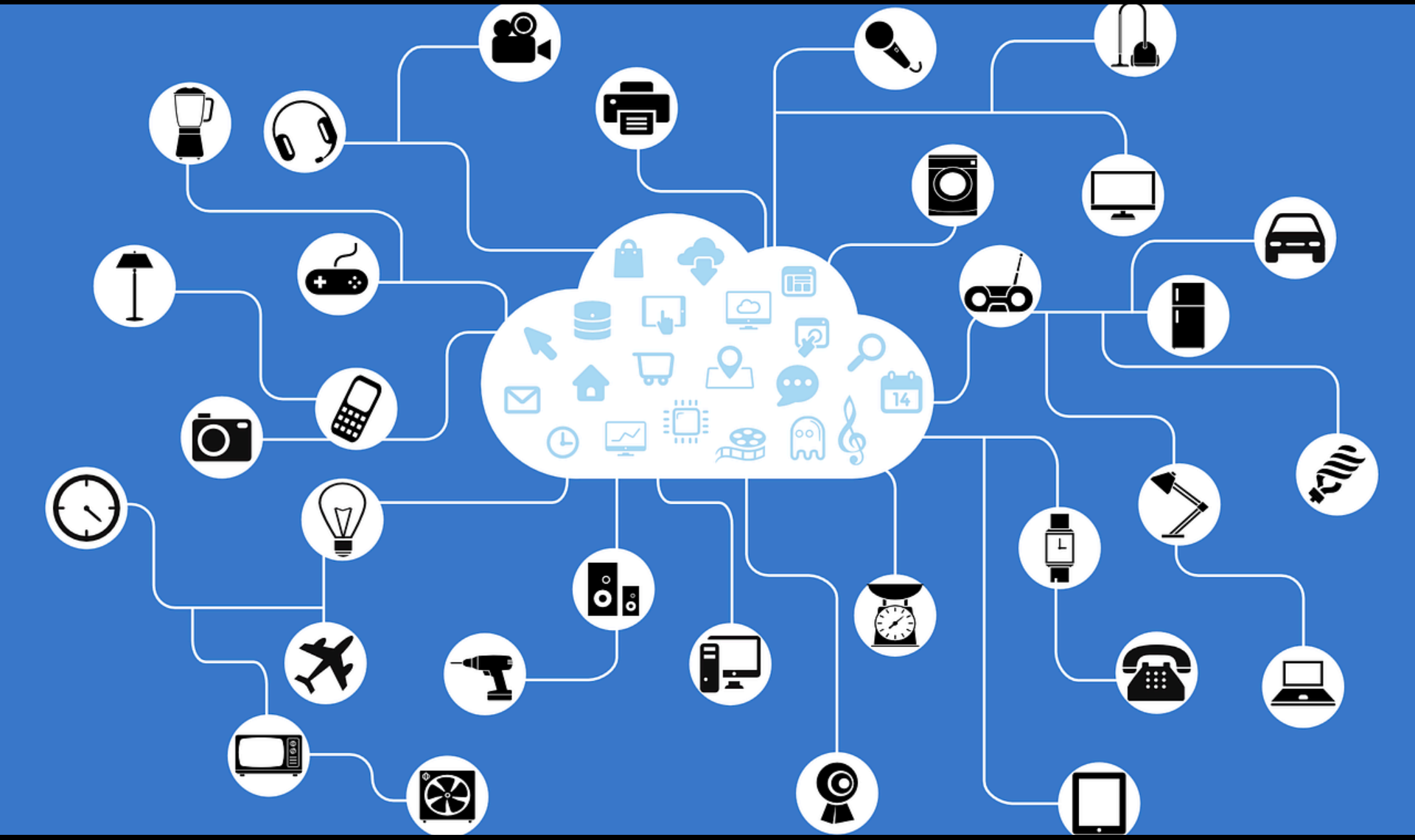
Digital infrastructure for 21st century living.

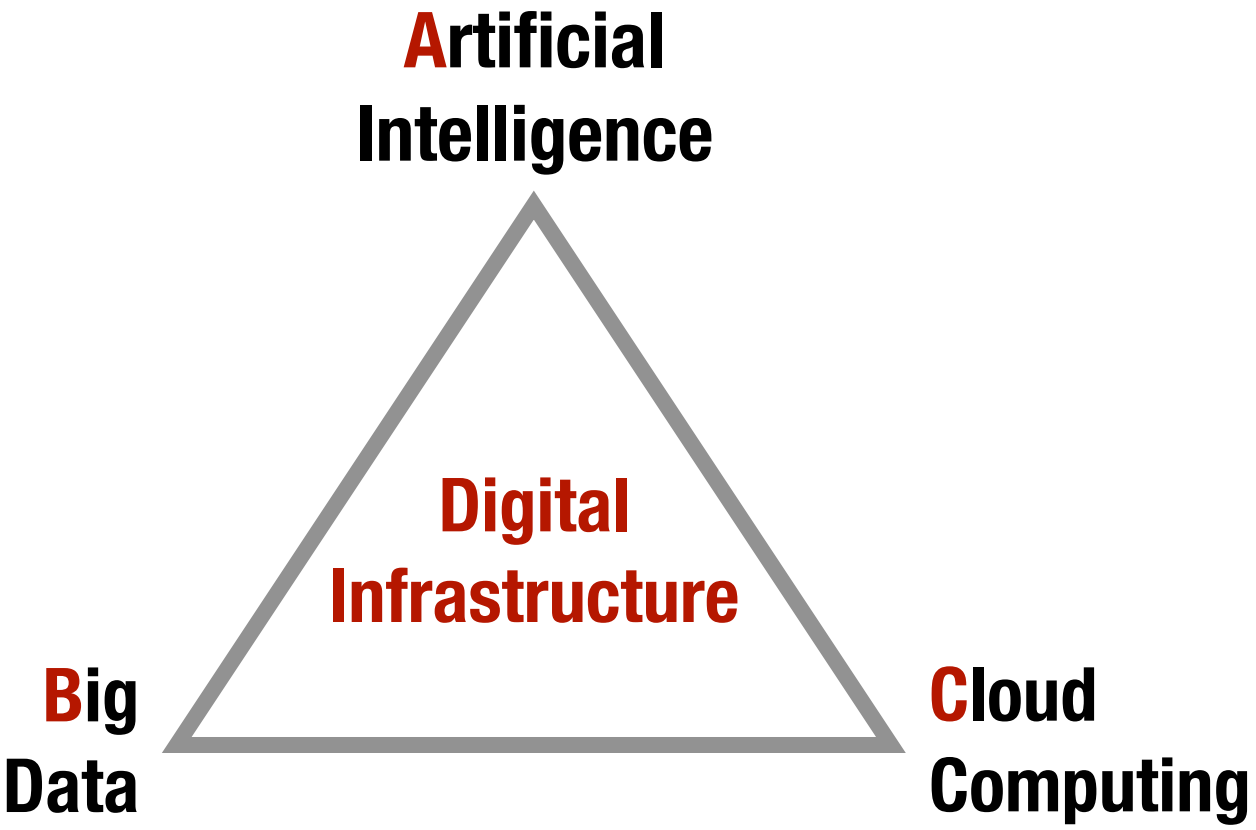


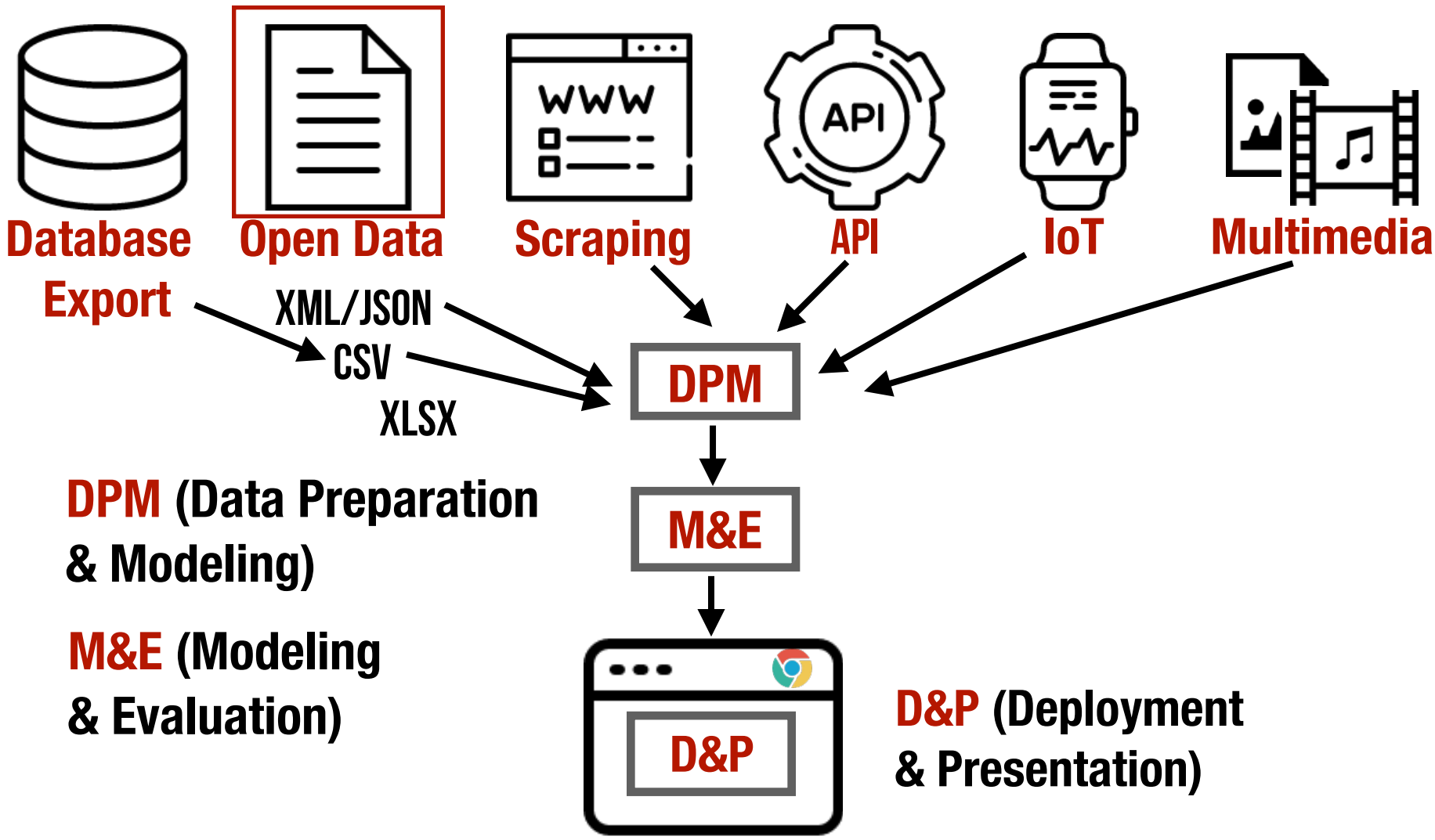


ARTIFICIAL INTELLIGENCE


BIG DATA

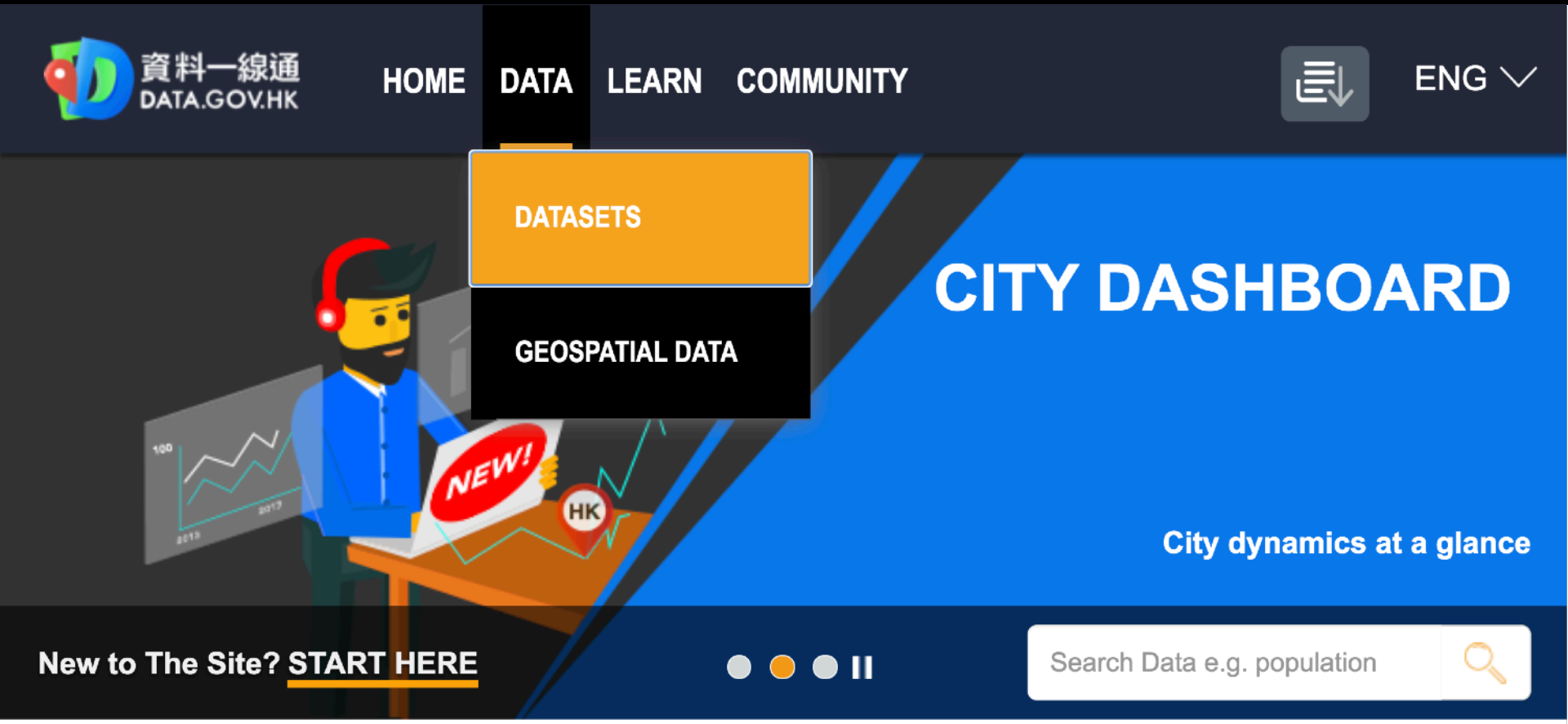



Artificial
Intelligence
Digital
Infrastructure
Big
Data
Cloud
Computing


Database Export
Open Data
Scraping
API
IoT
Multimedia
XML/JSON
CSV
XLSX
DPM
M&E
D&P
DPM (Data Preparation & Modeling)
M&E (Modeling & Evaluation)
D&P (Deployment & Presentation)

香港智慧城市

Hong Kong Smart City 藍圖 Blueprint



https://data.gov.hk/en-data/dataset/hk-lcsd-facility-facility-bkbc

- Convert JSON to CSV (https://codebeautify.org/jsonviewer)
- Convert XML to CSV (https://www.convertcsv.com/xml-to-csv.htm)
- Import CSV into Excel or Numbers

Basketball Courts (Free Outdoor Pitches/Courts)

LEISURE AND CULTURAL SERVICES DEPARTMENT | Recreation and Culture

UPDATE FREQUENCY: **AS AND WHEN NEW FACILITY IS ADDED OR AMENDMENT IS MADE**

Location of Basketball Courts (Free Outdoor Pitches/Courts)

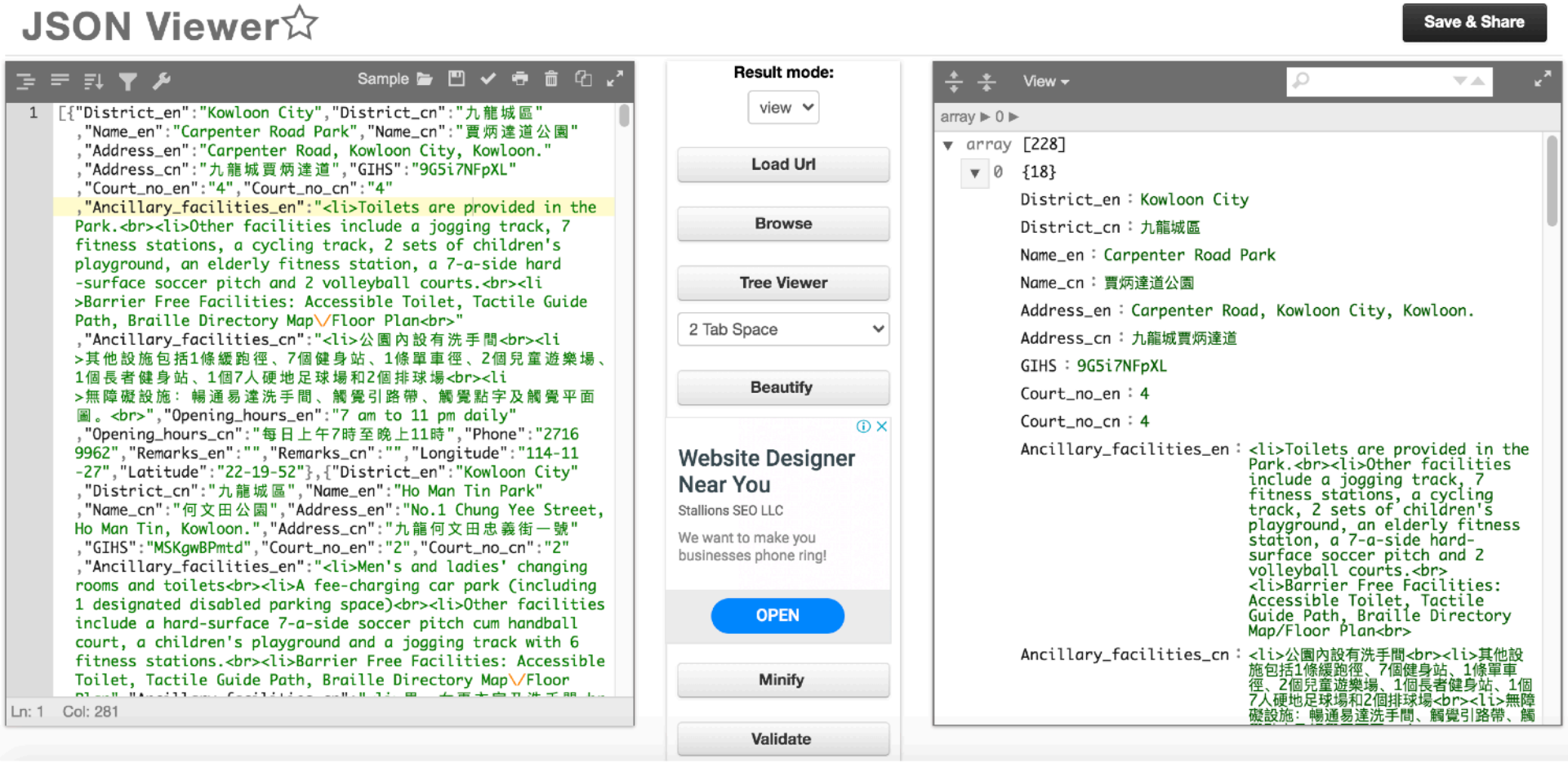
Data Dictionar(ies) : https://www.lcsd.gov.hk/datagovhk/facility/facility-bkbc_data_dictionary.pdf

---

1 JSON file(s)

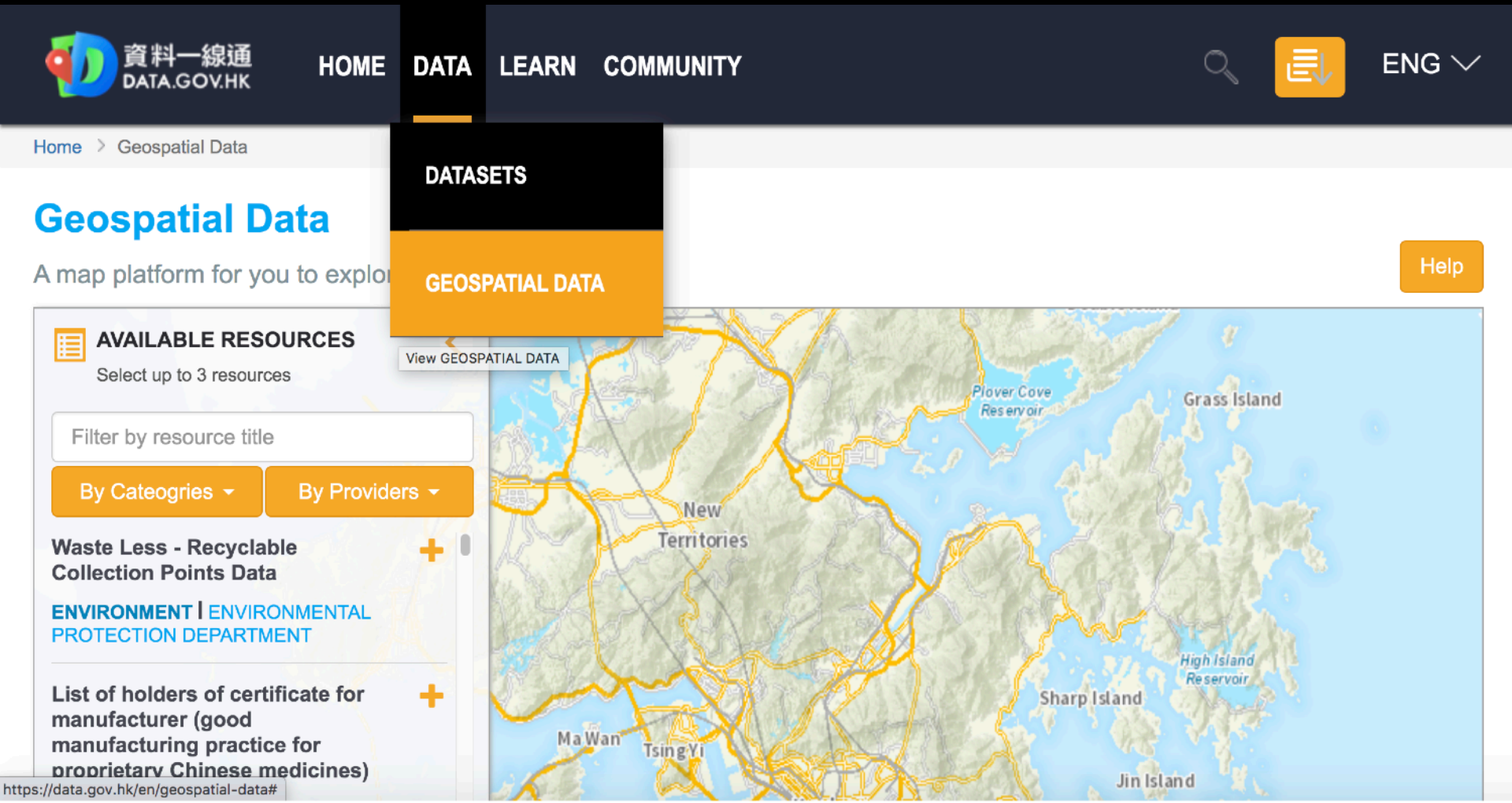
Add All to Queue

[{"District_en":"Kowloon City","District_cn":"九龍城區","Name_en":"Carpenter Road Park","Name_cn":"賈炳達道公園","Address_en":"Carpenter Road, Kowloon City, Kowloon.","Address_cn":"九龍城賈炳達道","GIHS":"9G5i7NFpXL","Court_no_en":"4","Court_no_cn":"4","Ancillary_facilities_en":"<li>Toilets are provided in the Park.
Other facilities include a jogging track, 7 fitness stations, a cycling track, 2 sets of children's playground, an elderly fitness station, a 7-a-side hard-surface soccer pitch and 2 volleyball courts.
Barrier Free Facilities: Accessible Toilet, Tactile Guide Path, Braille Directory Map\/Floor Plan
","Ancillary_facilities_cn":"<li>公園內設有洗手間
其他設施包括1條緩跑徑、7個健身站、1條單車徑、2個兒童遊樂場、1個長者健身站、1個7人硬地足球場和2個排球場
無障礙設施：暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面圖。
","Opening_hours_en":"7 am to 11 pm daily","Opening_hours_cn":"每日上午7時至晚上11時","Phone":"2716 9962","Remarks_en":"","Remarks_cn":"","Longitude":"114-11-27","Latitude":"22-19-52"},{"District_en":"Kowloon City","District_cn":"九龍城區","Name_en":"Ho Man Tin Park","Name_cn":"何文田公園","Address_en":"No.1 Chung Yee Street, Ho Man Tin, Kowloon.","Address_cn":"九龍何文田忠義街一號","GIHS":"MSKgwBPmtd","Court_no_en":"2","Court_no_cn":"2","Ancillary_facilities_en":"<li>Men's and ladies' changing rooms and toilets
A fee-charging car park (including 1 designated disabled parking space)
Other facilities include a hard-surface 7-a-side soccer pitch cum handball court, a children's playground and a jogging track with 6 fitness stations.
Barrier Free Facilities: Accessible Toilet, Tactile Guide Path, Braille Directory Map\/Floor Plan","Ancillary_facilities_cn":"<li>男、女更衣室及洗手間
1個收費停車場 (設有1個殘疾人士專用車位)
其他設施包括1個硬地7人足球場兼手球場、1個兒童遊樂場、1條緩跑徑和6個健身站
無障礙設施：暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面圖","Opening_hours_en":"7 am to 11 pm daily","Opening_hours_cn":"每日上午7時至晚上11時","Phone":"2762 7837","Remarks_en":"","Remarks_cn":"","Longitude":"114-10-50","Latitude":"22-18-44"},{"District_en":"Kowloon City","District_cn":"九龍城區","Name_en":"Hoi Sham Park","Name_cn":"海心公園","Address_en":"Yuk Yat Street, Tokwawan, Kowloon.","Address_cn":"九龍土瓜灣旭日街","GIHS":"nZ2deVMDpF","Court_no_en":"1","Court_no_cn":"1","Ancillary_facilities_en":"<li>Men’s and ladies’ toilets
Other facilities include 2 hard-surface 5-a-side soccer, a children's playground and elderly fitness equipment.
Barrier Free Facilities: Accessible Toilet, Tactile Guide Path, Braille Directory Map\/Floor Plan","Ancillary_facilities_cn":"<li>男、女洗手間
其他設施包括2個硬地5人足球場、兒童遊樂場及長者健體設施
無障礙設施：暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面圖","Opening_hours_en":"7 am to 11 pm daily","Opening_hours_cn":"每日上午7時至晚上11時","Phone":"2334 3576 \/ 2762 2083","Remarks_en":"","Remarks_cn":"","Longitude":"114-11-30","Latitude":"22-18-54"},{"District_en":"Kowloon City","District_cn":"九龍城區","Name_en":"Junction Road Park","Name_cn":"聯合道公



https://codebeautify.org/jsonviewer

District_en	District_cn	Name_en	Name_cn	Address_en
Kowloon City	九龍城區	Carpenter Road Park	賈炳達道公園	Carpenter Road, Kowloon City, Kowloon.
Kowloon City	九龍城區	Ho Man Tin Park	何文田公園	No.1 Chung Yee Street, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Hoi Sham Park	海心公園	Yuk Yat Street, Tokwawan, Kowloon.
Kowloon City	九龍城區	Junction Road Park	聯合道公園	195 Junction Road, Kowloon City.
Kowloon City	九龍城區	Kam Shing Road Recreation Ground	金城道遊樂場	Kam Shing Road, Kowloon.
Kowloon City	九龍城區	Kau Pui Lung Road Playground	靠背壟道遊樂場	Kau Pui Lung Road, Kowloon.
Kowloon City	九龍城區	Kent Road Garden	根德道花園	Kent Road, Kowloon.
Kowloon City	九龍城區	King Wan Street Playground	景雲街遊樂場	King Wan Street, To Kwa Wan, Kowloon.
Kowloon City	九龍城區	King's Park High Level Service Reservoir Playground	京士柏上配水庫遊樂場	Chung Hau Street, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Kowloon Tsai Park	九龍仔公園	13 Inverness Road, Kowloon City, Kowloon.
Kowloon City	九龍城區	Lung Cheung Road Playground	龍翔道遊樂場	Beacon Hill Road, Kowloon.
Kowloon City	九龍城區	Ma Tau Wai Road Playground	馬頭圍道遊樂場	Ma Tau Wai Road,Tokwawan, Kowloon.
Kowloon City	九龍城區	Oxford Road Playground	牛津道遊樂場	Oxford Road, Kowloon.
Kowloon City	九龍城區	Peace Avenue Playground	太平道遊樂場	Peace Avenue, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Perth Street Sports Ground	巴富街運動場	Shek Ku Street , Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Pui Ching Road Playground	培正道遊樂場	Pui Ching Road, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Rutland Quadrant Children's Playground	律倫街兒童遊樂場	Rutland Quadrant, Kowloon.
Kowloon City	九龍城區	Sung Wong Toi Playground	宋王臺遊樂場	Sung Wong Toi Road, Kowloon.
Kowloon City	九龍城區	Tai Wan Road Playground	大環道遊樂場	Tai Wan Road, Hung Hom, Kowloon.
Kowloon City	九龍城區	Tai Wan Shan Park	大環山公園	Wan Hoi Street, Hung Hom, Kowloon.
Kowloon City	九龍城區	To Kwa Wan Recreation Ground	土瓜灣遊樂場	66 Ha Heung Road, To Kwa Wan, Kowloon



https://data.gov.hk/en/geospatial-data

資料一線通 DATA.GOV.HK

HOME DATA LEARN COMMUNITY ENG

Home > Datasets > FEHD facility and service locations > FEHD facility and service locations (English)

FEHD facility and service locations (English)

FOOD AND ENVIRONMENTAL HYGIENE DEPARTMENT | Health | XML | API Available

LAST UPDATED ON: **11/09/2020**
UPDATE FREQUENCY: **AS AND WHEN NECESSARY**

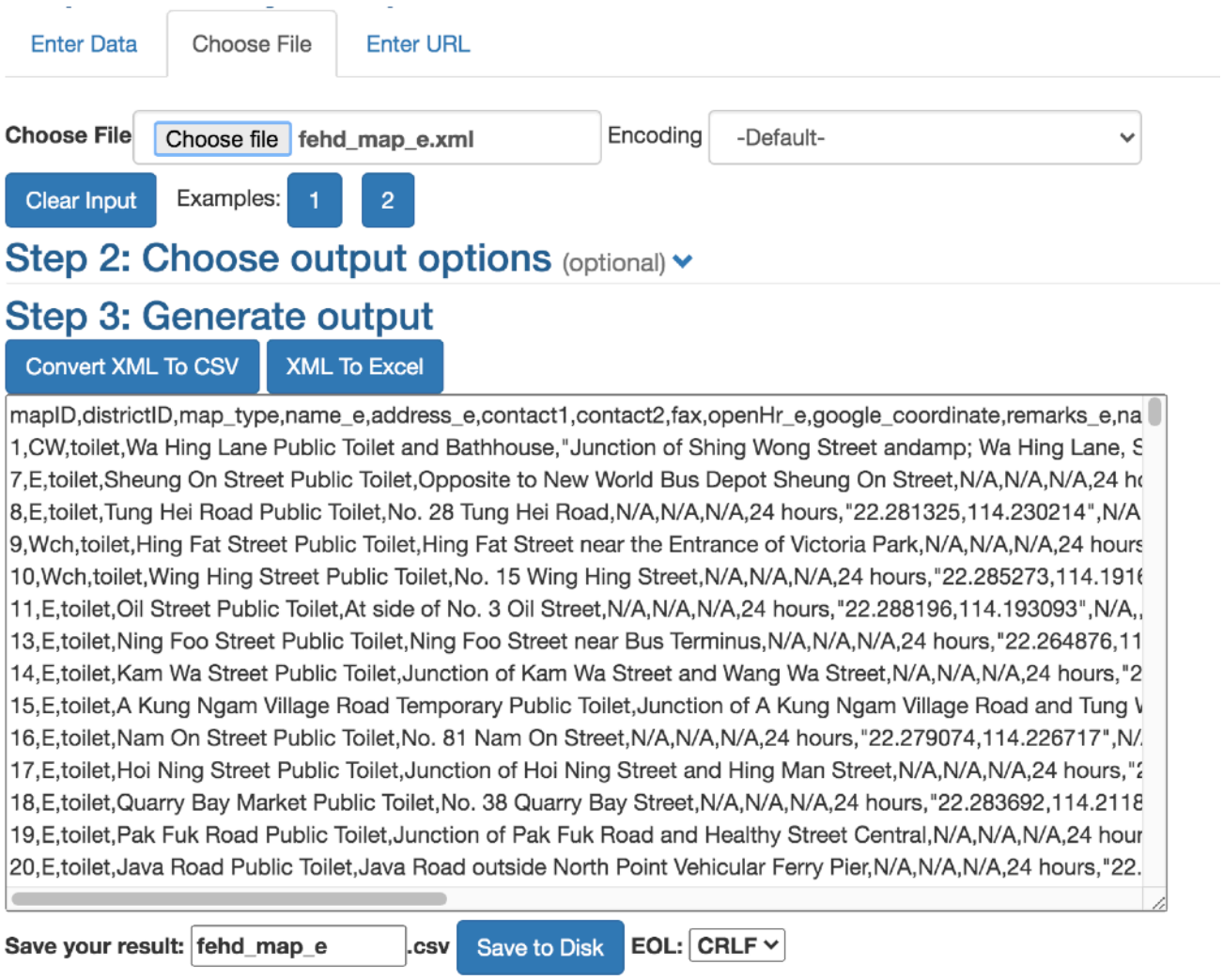
Locations and details about different facilities and services provided by the Food and Environmental Hygiene Department.

This dataset provides geo-referenced information (including district, type of facility/service, name, address, contact numbers, opening hours and coordinates) about different types of facilities/services provided by the Food and Environmental Hygiene Department (including public toilets, public refuse collection points, public markets, cooked food markets, etc).

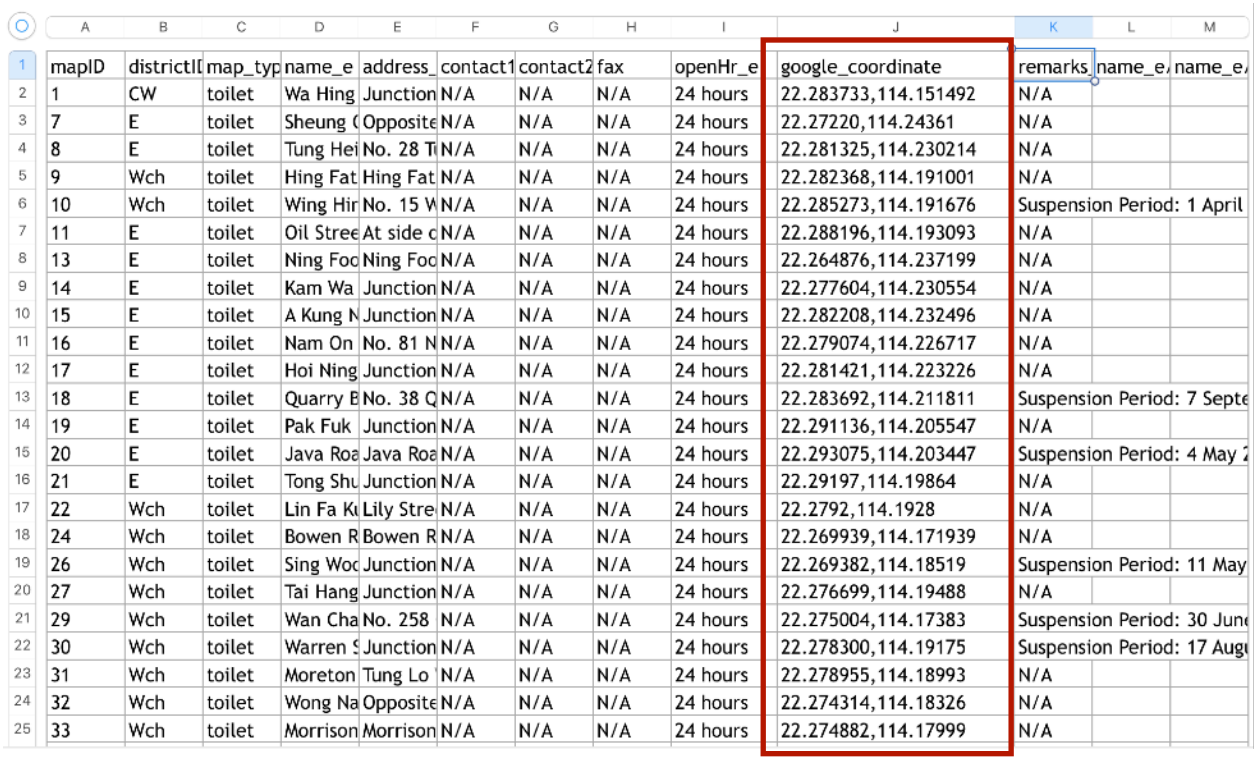
URL: http://www.fehd.gov.hk/english/map/fehd_map_e.xml

Download Add to Queue

https://data.gov.hk/en-data/dataset/hk-fehd-fehdlocatn-fehd-facility-and-service-locations



https://www.convertcsv.com/xml-to-csv.htm



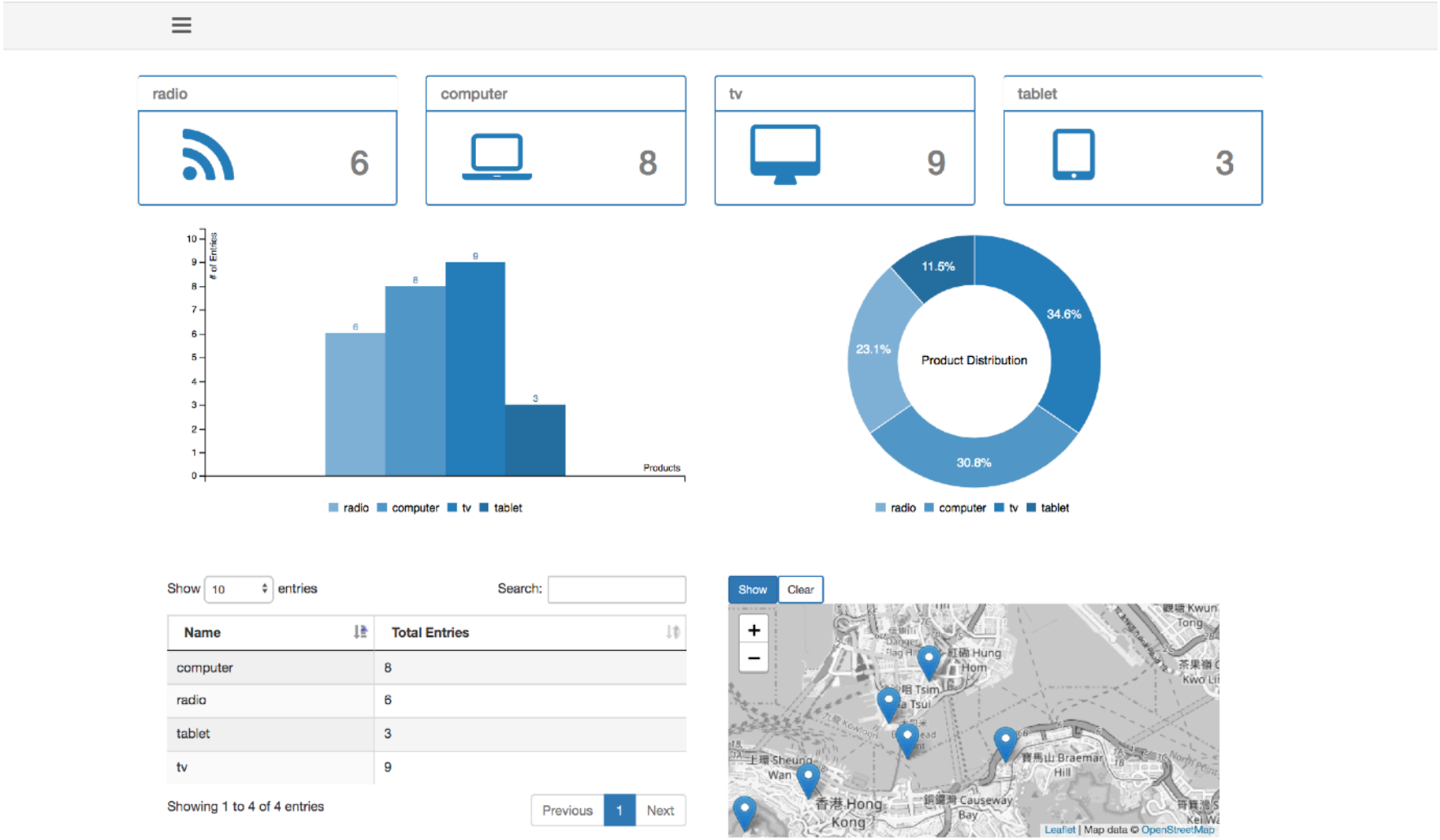
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	mapID	districtID	map_typ	name_e	address_	contact1	contact2	fax	openHr_e	google_coordinate	remarks_	name_e	name_e
2	1	CW	toilet	Wa Hing	Junction	N/A	N/A	N/A	24 hours	22.283733,114.151492	N/A		
3	7	E	toilet	Sheung (	Opposite	N/A	N/A	N/A	24 hours	22.27220,114.24361	N/A		
4	8	E	toilet	Tung Hei	No. 28 T	N/A	N/A	N/A	24 hours	22.281325,114.230214	N/A		
5	9	Wch	toilet	Hing Fat	Hing Fat	N/A	N/A	N/A	24 hours	22.282368,114.191001	N/A		
6	10	Wch	toilet	Wing Hir	No. 15 W	N/A	N/A	N/A	24 hours	22.285273,114.191676	Suspension Period: 1 April		
7	11	E	toilet	Oil Stree	At side c	N/A	N/A	N/A	24 hours	22.288196,114.193093	N/A		
8	13	E	toilet	Ning Foo	Ning Foo	N/A	N/A	N/A	24 hours	22.264876,114.237199	N/A		
9	14	E	toilet	Kam Wa	Junction	N/A	N/A	N/A	24 hours	22.277604,114.230554	N/A		
10	15	E	toilet	A Kung N	Junction	N/A	N/A	N/A	24 hours	22.282208,114.232496	N/A		
11	16	E	toilet	Nam On	No. 81 N	N/A	N/A	N/A	24 hours	22.279074,114.226717	N/A		
12	17	E	toilet	Hoi Ning	Junction	N/A	N/A	N/A	24 hours	22.281421,114.223226	N/A		
13	18	E	toilet	Quarry B	No. 38 Q	N/A	N/A	N/A	24 hours	22.283692,114.211811	Suspension Period: 7 Septe		
14	19	E	toilet	Pak Fuk	Junction	N/A	N/A	N/A	24 hours	22.291136,114.205547	N/A		
15	20	E	toilet	Java Roa	Java Roa	N/A	N/A	N/A	24 hours	22.293075,114.203447	Suspension Period: 4 May 2		
16	21	E	toilet	Tong Shu	Junction	N/A	N/A	N/A	24 hours	22.29197,114.19864	N/A		
17	22	Wch	toilet	Lin Fa Ku	Lily Stre	N/A	N/A	N/A	24 hours	22.2792,114.1928	N/A		
18	24	Wch	toilet	Bowen R	Bowen R	N/A	N/A	N/A	24 hours	22.269939,114.171939	N/A		
19	26	Wch	toilet	Sing Woo	Junction	N/A	N/A	N/A	24 hours	22.269382,114.18519	Suspension Period: 11 May		
20	27	Wch	toilet	Tai Hang	Junction	N/A	N/A	N/A	24 hours	22.276699,114.19488	N/A		
21	29	Wch	toilet	Wan Cha	No. 258	N/A	N/A	N/A	24 hours	22.275004,114.17383	Suspension Period: 30 June		
22	30	Wch	toilet	Warren S	Junction	N/A	N/A	N/A	24 hours	22.278300,114.19175	Suspension Period: 17 Augu		
23	31	Wch	toilet	Moreton	Tung Lo	N/A	N/A	N/A	24 hours	22.278955,114.18993	N/A		
24	32	Wch	toilet	Wong Na	Opposite	N/A	N/A	N/A	24 hours	22.274314,114.18326	N/A		
25	33	Wch	toilet	Morrison	Morrison	N/A	N/A	N/A	24 hours	22.274882,114.17999	N/A		

_ID	__text			
CW	Central and Western			
E	Eastern			
S	Southern			
Wch	Wan Chai			
Is	Islands			
YT	Yau Tsim			
MK	Mong Kok			
SSP	Sham Shui Po			
KC	Kowloon City			
WTS	Wong Tai Sin			
KT	Kwun Tong			
TW	Tsuen Wan			
KwT	Kwai Tsing			
N	North			
TP	Tai Po			
SK	Sai Kung			
ST	Sha Tin			
TM	Tuen Mun			
YL	Yuen Long			

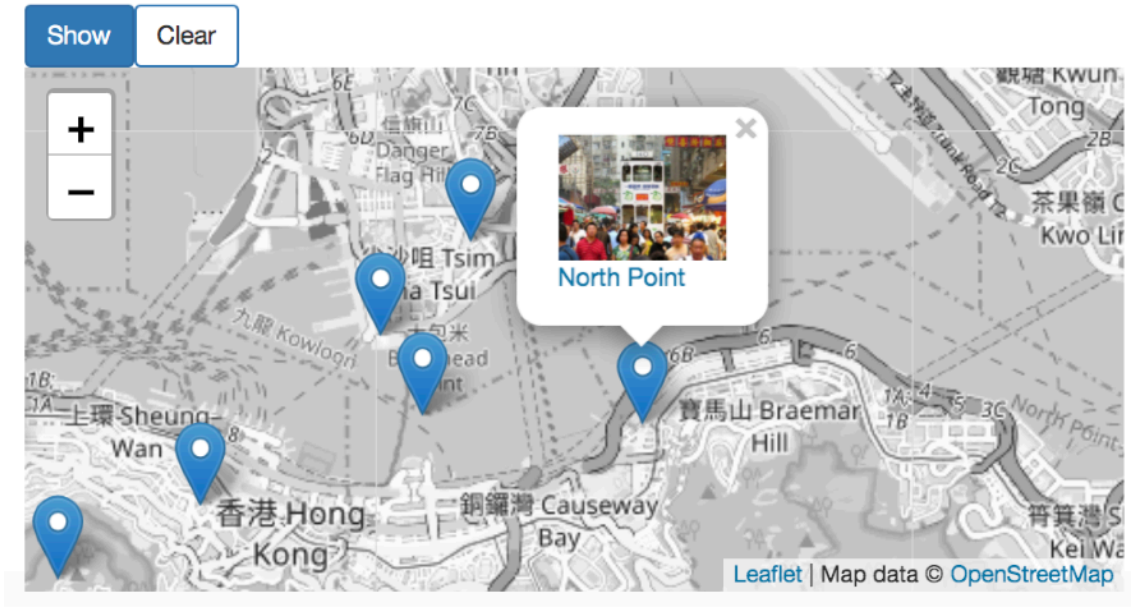
Product | Roll-up | Venues | Users | sw_product

Grid view | Hide fields | Filter | Group | Sort | ...

	Name	Lat	Lng	url	Pic	img_url
1	North Point	22.287111	114.191667	https://en.wikipedia.org/...		https://dl.airtab
2	Mong Kok	22.322500	114.170556	https://en.wikipedia.org/...		https://dl.airtab
3	Happy Valley	22.266667	114.183333	https://en.wikipedia.org/...		https://dl.airtab
4	Victoria Peak	22.275469	114.143828	https://en.wikipedia.org/...		https://dl.airtab
5	Lan Kwai Fong	22.280972	114.155528	https://en.wikipedia.org/...		https://dl.airtab
6	Choi Hung	22.334484	114.210024	https://en.wikipedia.org/...		https://dl.airtab
7	HKBU	22.338936	114.181953	https://en.wikipedia.org/...		https://dl.airtab
8	CUHK	22.418498	114.204074	https://en.wikipedia.org/...		https://dl.airtab
9	HKU	22.284167	114.137778	https://en.wikipedia.org/...		https://dl.airtab
10	HK Science Musuem	22.301000	114.177655	https://en.wikipedia.org/...		https://dl.airtab
11	HK Cultural Center	22.293850	114.170323	https://en.wikipedia.org/...		https://dl.airtab
12	Victoria Harbour	22.287753	114.173619	https://en.wikipedia.org/...		https://dl.airtab



https://python30.cuhkcfe.io/



Geospatial Data Visualization

資料一線通 DATA.GOV.HK

HOME DATA LEARN COMMUNITY

ENG

CITY DASHBOARD

ANNUAL OPEN DATA PLANS

FEEDBACK

APPLICATIONS

1st Hand HK Property Data

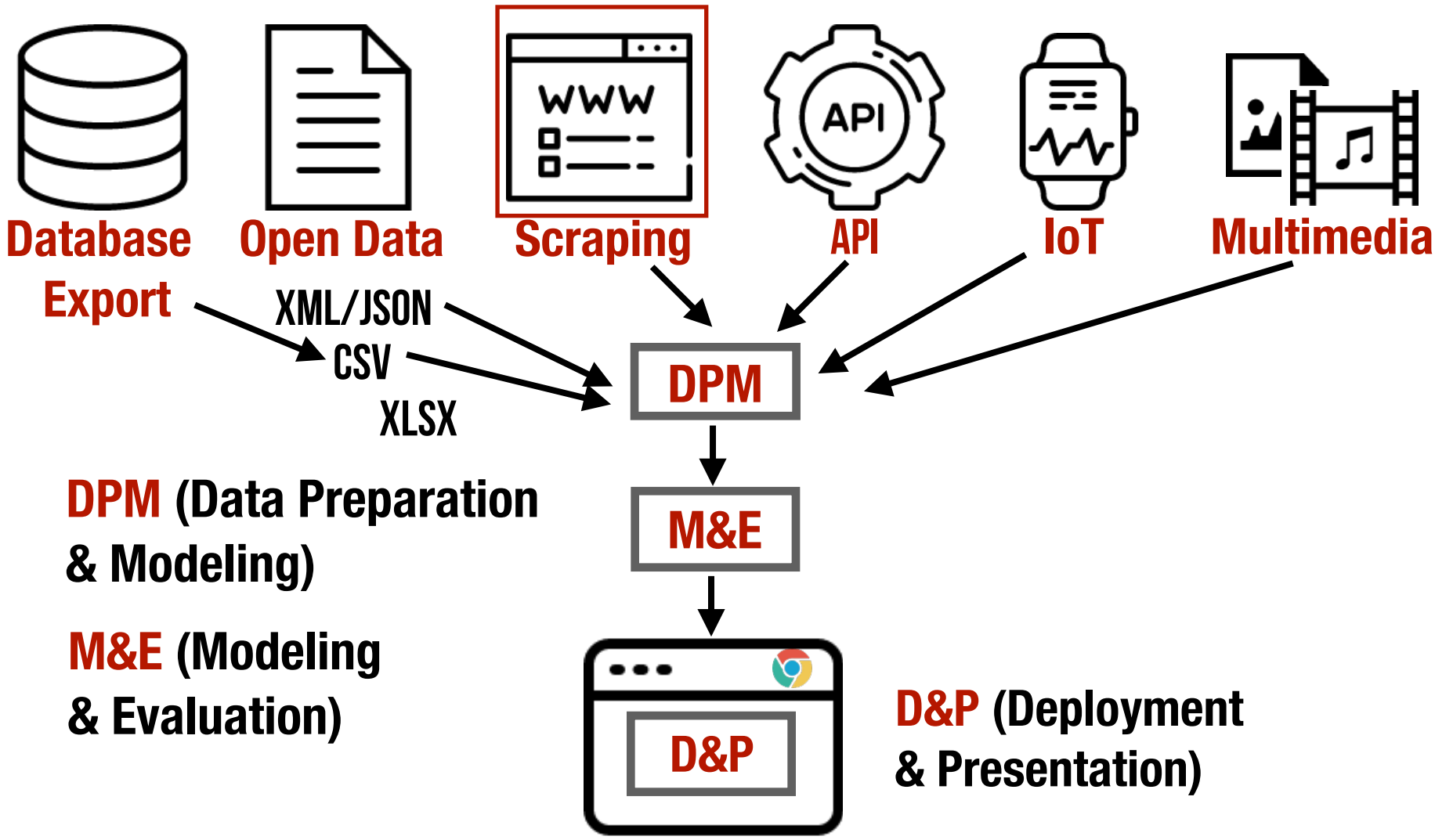
The application provides the latest property news and property data on the first hand Hong Kong residential property market.

28Hse
informa
Kong p
includir
listings by agents and landlords, property

provides information for the Hong Kong Property Market, including latest property listings by agents and landlord, property...

Geospatial Data Visualization

https://data.gov.hk/en/application


Database Export
Open Data
Scraping
API
IoT
Multimedia
XML/JSON
CSV
XLSX
DPM
M&E
D&P
DPM (Data Preparation & Modeling)
M&E (Modeling & Evaluation)
D&P (Deployment & Presentation)

parsehub

Features Download Pricing Help Blog Log in Sign up

A free web scraper that is easy to use

ParseHub is a free and powerful web scraping tool. With our advanced web scraper, extracting data is as easy as clicking on the data you need.

Download ParseHub for Free

Open a website

Download our desktop app. Choose a site to scrape data from.

Click to select data

Get data from multiple pages. Interact with AJAX, forms, dropdowns, etc.

Download results

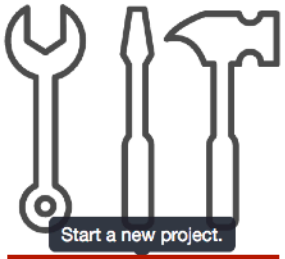
Access data via JSON, Excel and API. Data is collected by our servers.

Scraping single page table.

bernard@intechni...

- Projects
- Runs
- My Account
- Integrations
- Plans & Billing
- Tutorials
- Documentation
- API
- Contact
- Log Out

Recent projects



Start a new project.

+ New Project

covid19 hkej.com Project

travelchinaguide.com Proj...

startupbeat.hkej.com Proj...

basketball-reference.com ...

+ See more

Recent runs

Interactive Tutorials

- Learn the Basics (8 min)
- Select and Download Data (3 min)
- Group data with Relative Select (3 min)
- Click and Navigate to Links (4 min)

Written Tutorials

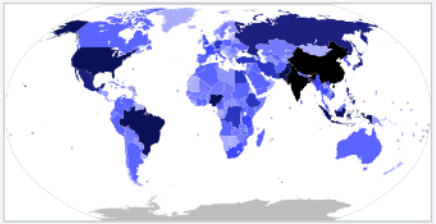
1. Parsehub 101
2. Pagination ('next' page buttons)
3. Scrape product details
4. Scrape leads from directories

List of countries and dependencies by population

From Wikipedia, the free encyclopedia

This is a **list of countries and dependencies by population**. It includes sovereign states, inhabited dependent territories and, in some cases, constituent countries of sovereign states, with inclusion within the list being primarily based on the ISO standard ISO 3166-1. For instance, the United Kingdom is considered as a single entity, while the constituent countries of the Kingdom of the Netherlands are considered separately. In addition, this list includes certain states with limited recognition not found in ISO 3166-1.

Also given in percent is each country's population compared with the world population, which the United Nations estimates at 7.82 billion as of today.

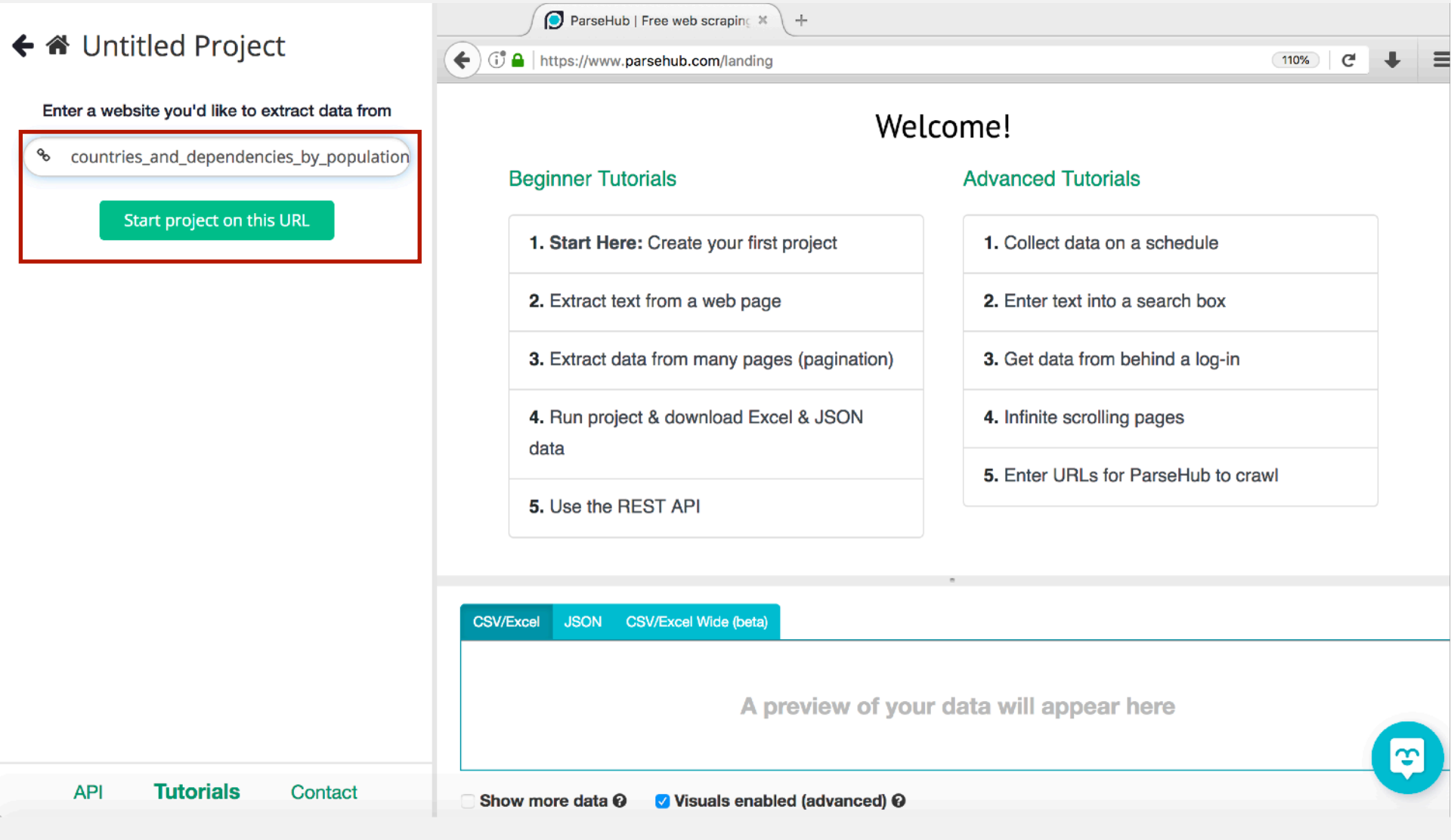


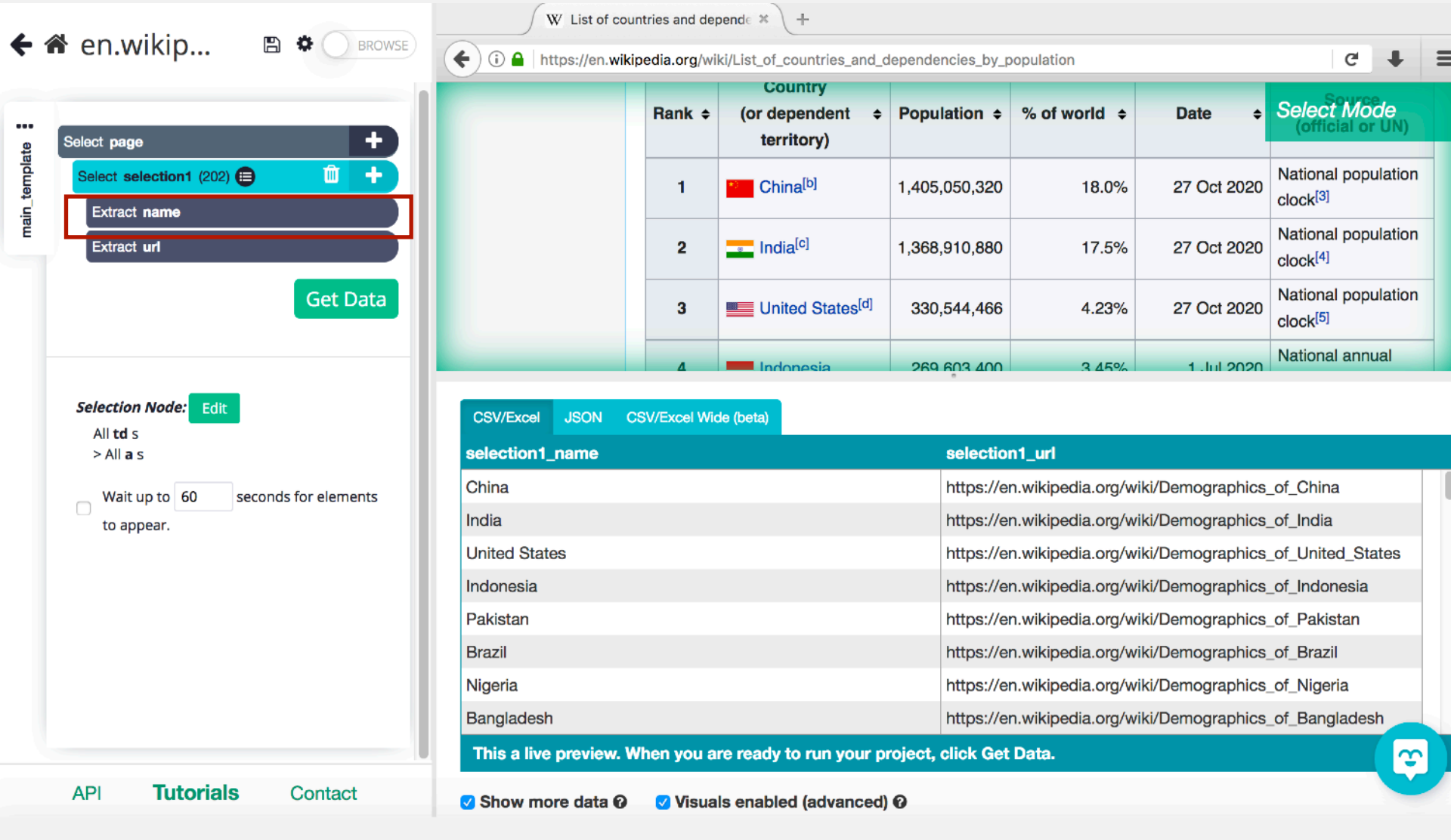
Map of countries and territories by population in 2019

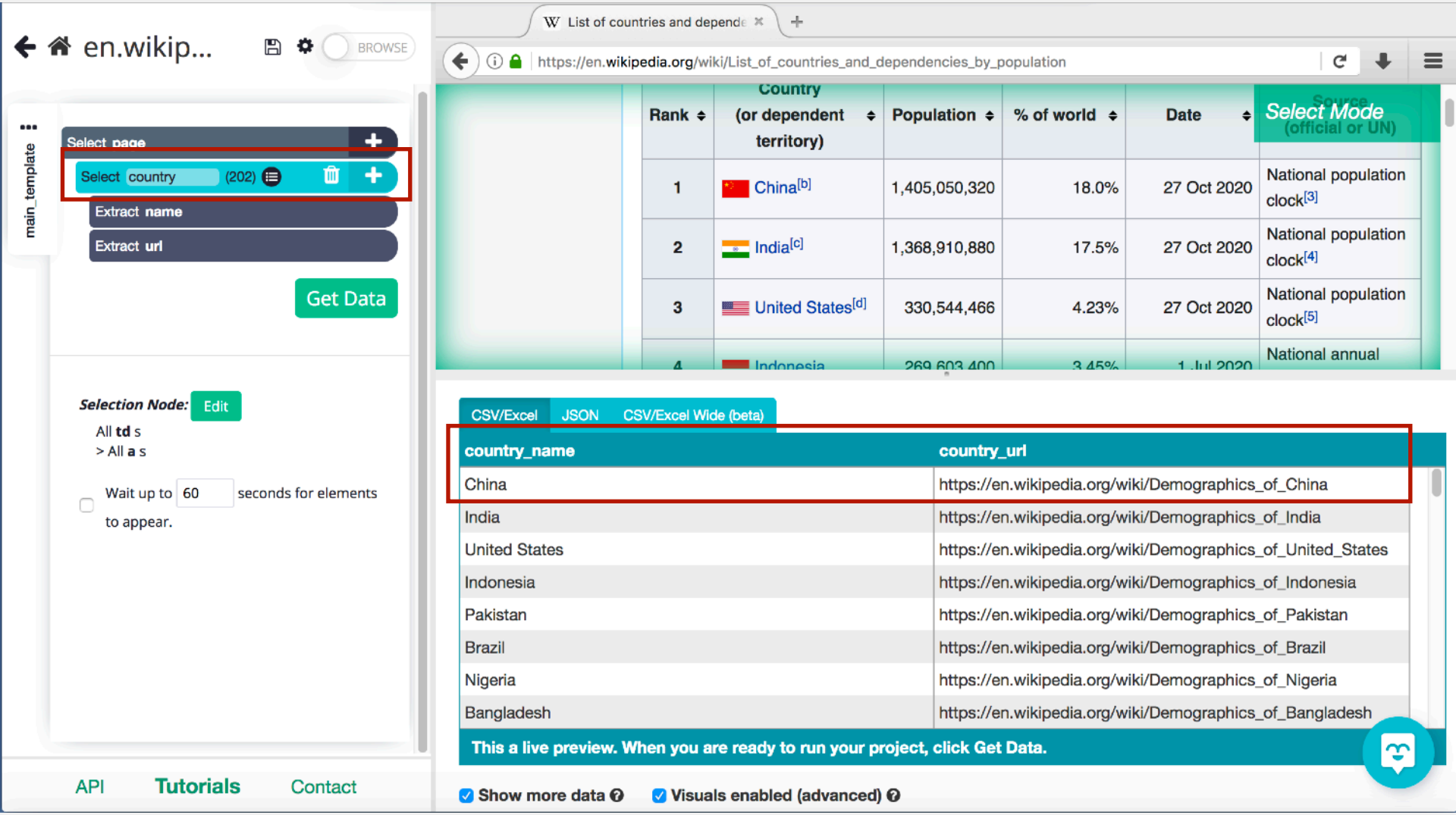
A cartogram of the world population in 2018

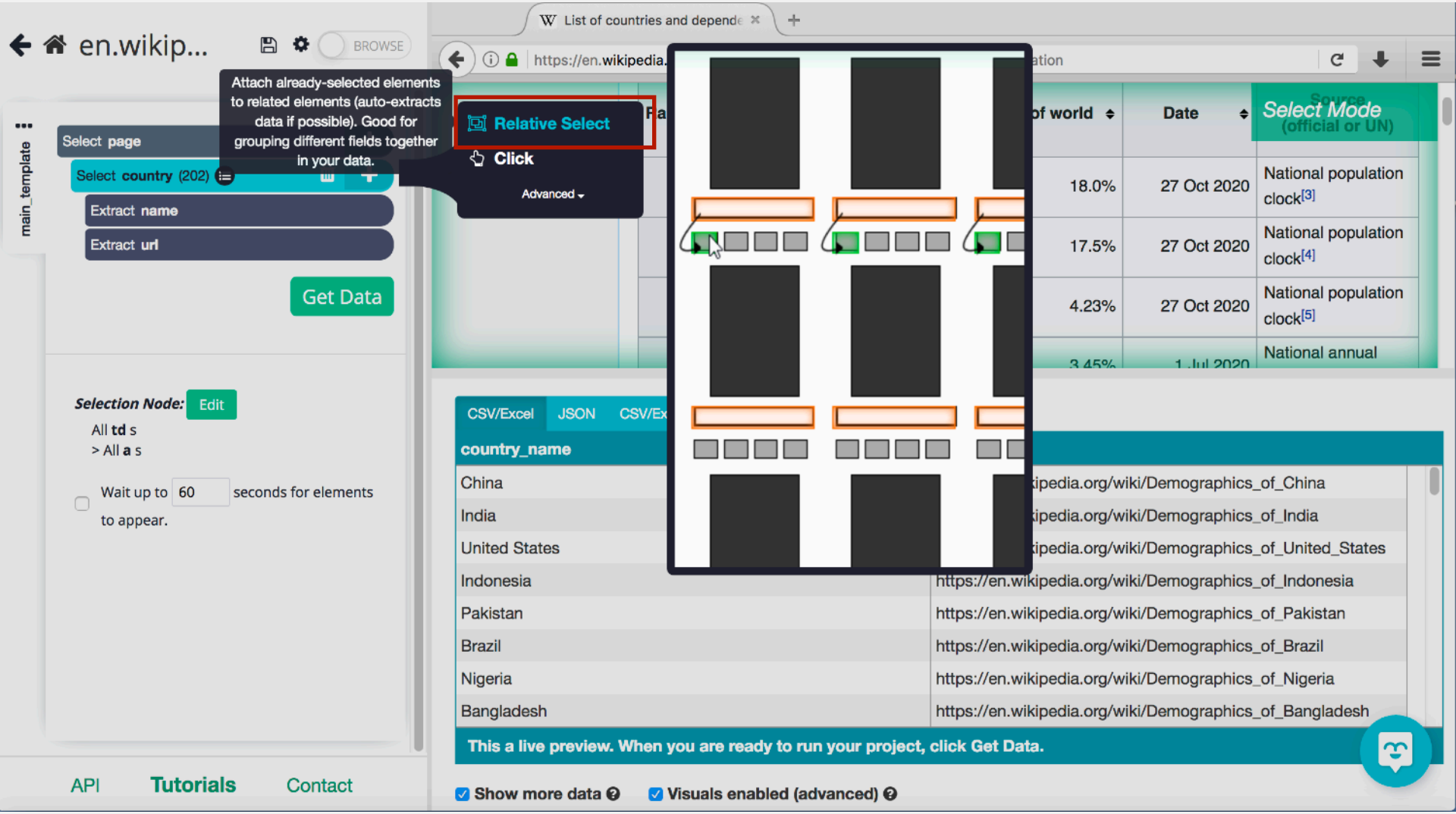
Contents [hide]

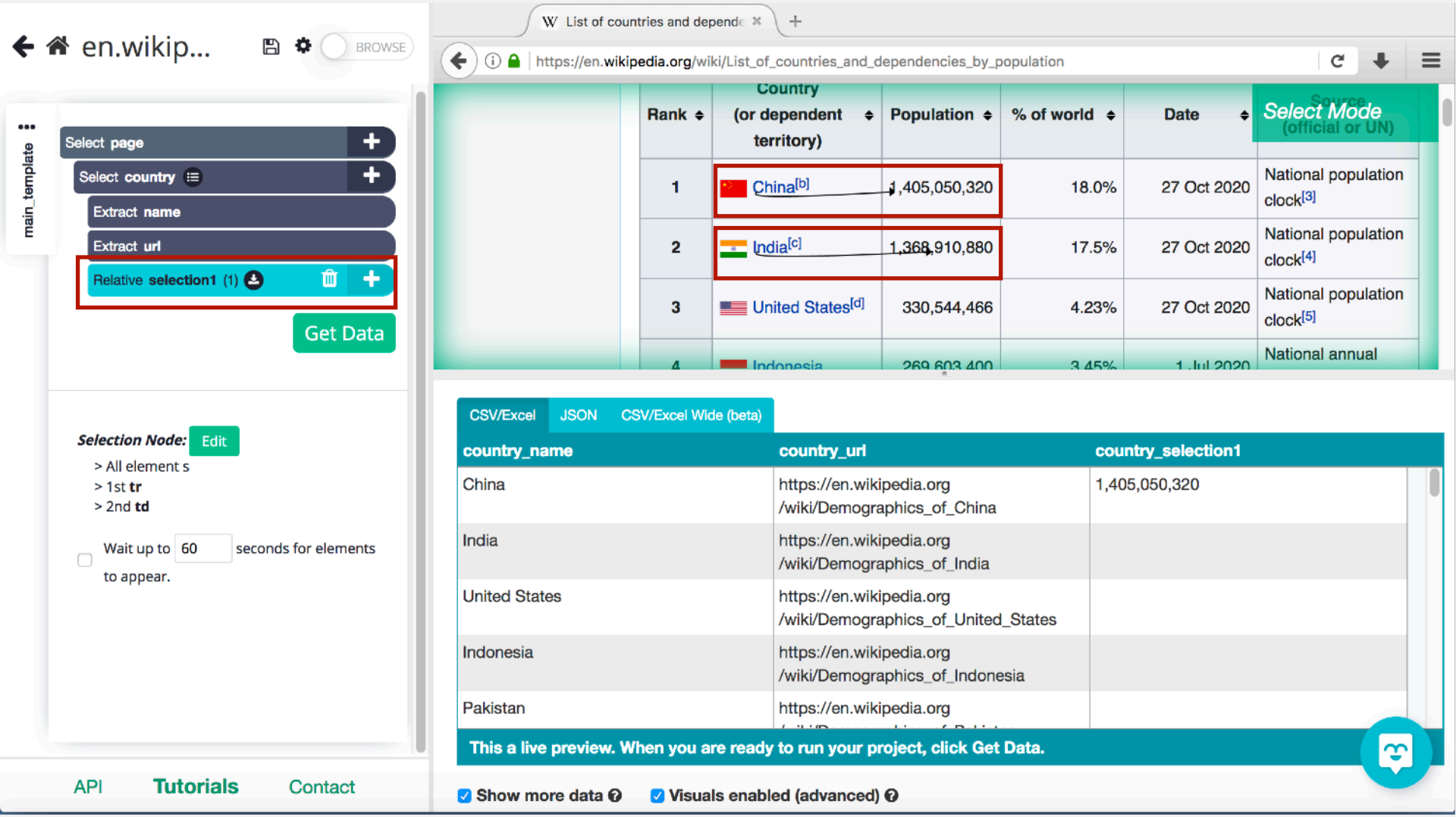
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population

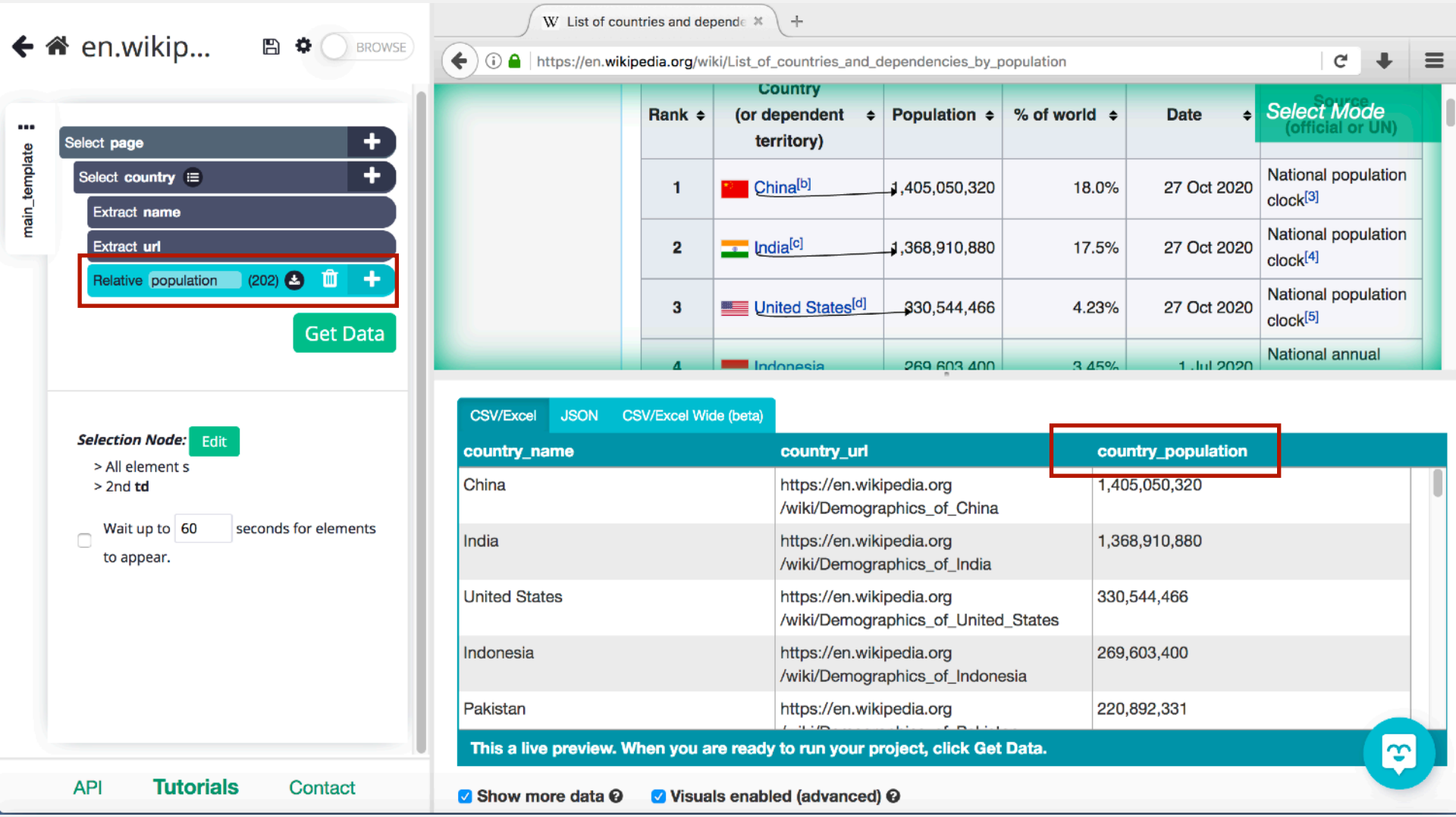

Untitled Project
Enter a website you'd like to extract data from
countries_and_dependencies_by_population
Start project on this URL
API
Tutorials
Contact
ParseHub | Free web scraping
https://www.parsehub.com/landing
110%
Welcome!
Beginner Tutorials
1. Start Here: Create your first project
2. Extract text from a web page
3. Extract data from many pages (pagination)
4. Run project & download Excel & JSON data
5. Use the REST API
Advanced Tutorials
1. Collect data on a schedule
2. Enter text into a search box
3. Get data from behind a log-in
4. Infinite scrolling pages
5. Enter URLs for ParseHub to crawl
CSV/Excel
JSON
CSV/Excel Wide (beta)
A preview of your data will appear here
Show more data
Visuals enabled (advanced)

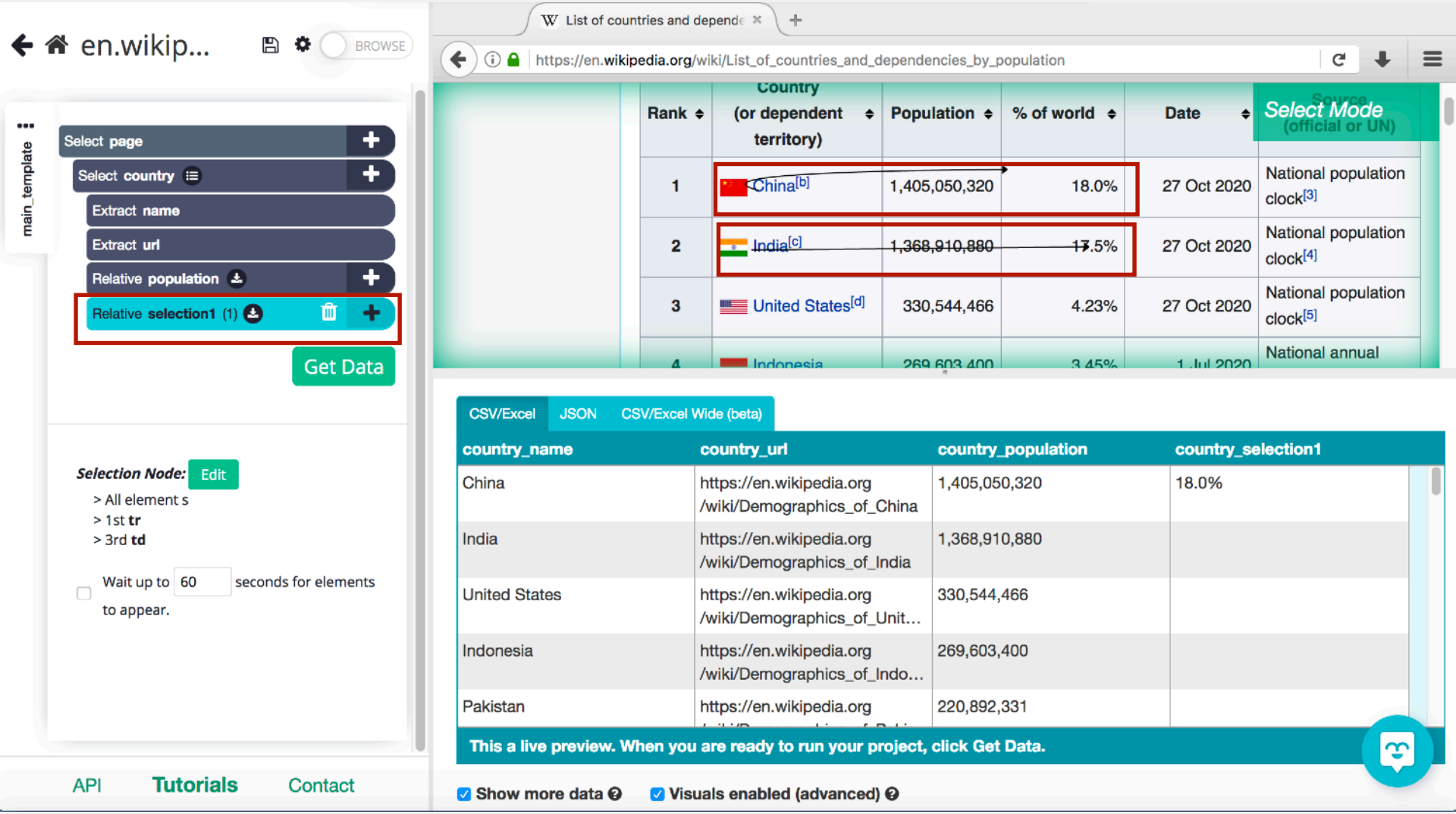

en.wikip...
BROWSE
main_template
Select page
Select selection1 (202)
Extract name
Extract url
Get Data
Selection Node:
Edit
All td s
> All a s
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depend
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Source (official or UN)
Select Mode
1 China[b] 1,405,050,320 18.0% 27 Oct 2020 National population clock[3]
2 India[c] 1,368,910,880 17.5% 27 Oct 2020 National population clock[4]
3 United States[d] 330,544,466 4.23% 27 Oct 2020 National population clock[5]
4 Indonesia 269,603,400 3.45% 1 Jul 2020 National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
selection1_name
selection1_url
China https://en.wikipedia.org/wiki/Demographics_of_China
India https://en.wikipedia.org/wiki/Demographics_of_India
United States https://en.wikipedia.org/wiki/Demographics_of_United_States
Indonesia https://en.wikipedia.org/wiki/Demographics_of_Indonesia
Pakistan https://en.wikipedia.org/wiki/Demographics_of_Pakistan
Brazil https://en.wikipedia.org/wiki/Demographics_of_Brazil
Nigeria https://en.wikipedia.org/wiki/Demographics_of_Nigeria
Bangladesh https://en.wikipedia.org/wiki/Demographics_of_Bangladesh
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)

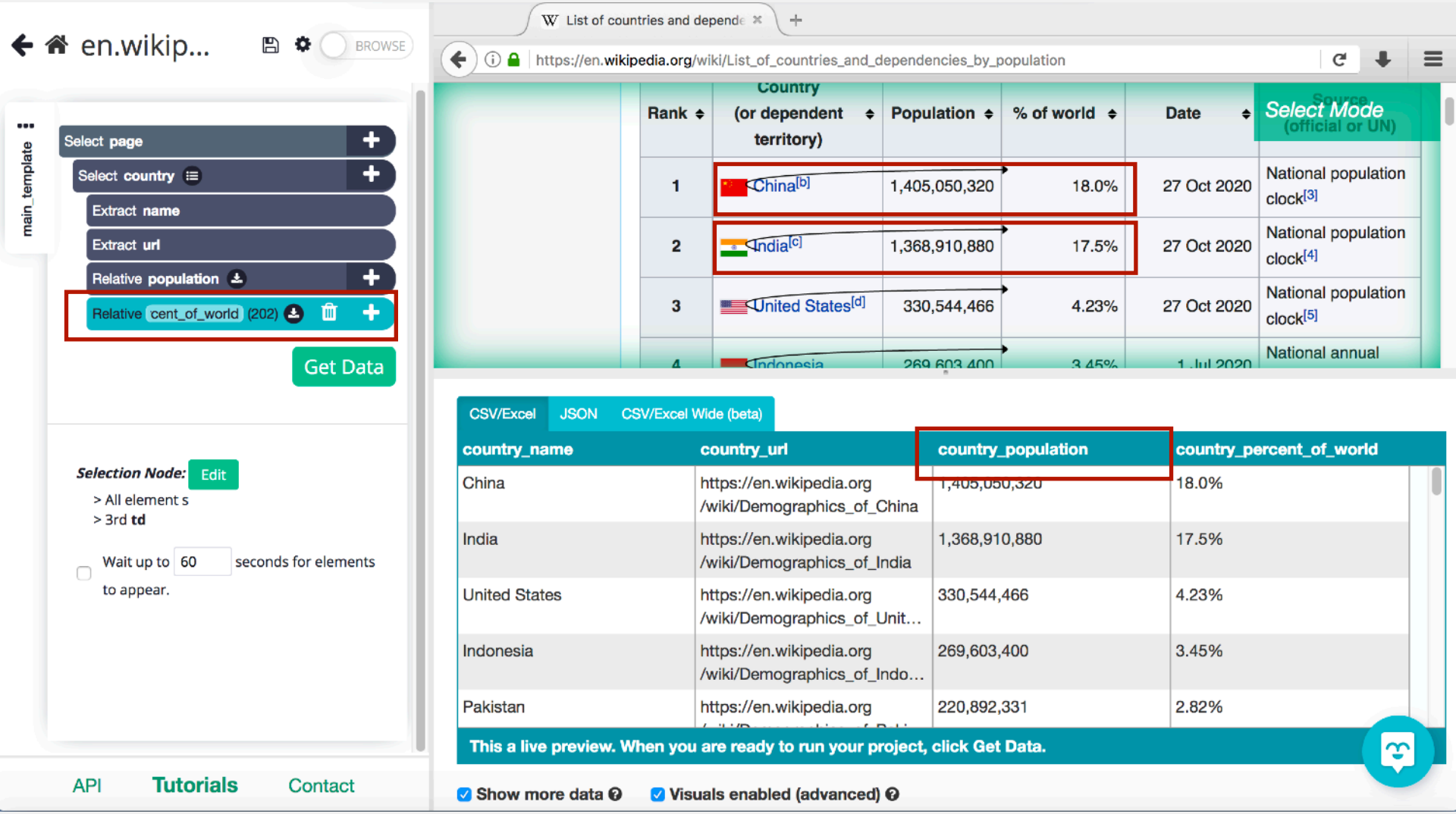

en.wikip...
BROWSE
main_template
Select page
Select country (202)
Extract name
Extract url
Get Data
Selection Node:
Edit
All td s
> All a s
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depende
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Source (official or UN)
Select Mode
1 China[b] 1,405,050,320 18.0% 27 Oct 2020 National population clock[3]
2 India[c] 1,368,910,880 17.5% 27 Oct 2020 National population clock[4]
3 United States[d] 330,544,466 4.23% 27 Oct 2020 National population clock[5]
4 Indonesia 269,603,400 3.45% 1 Jul 2020 National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
country_name
country_url
China https://en.wikipedia.org/wiki/Demographics_of_China
India https://en.wikipedia.org/wiki/Demographics_of_India
United States https://en.wikipedia.org/wiki/Demographics_of_United_States
Indonesia https://en.wikipedia.org/wiki/Demographics_of_Indonesia
Pakistan https://en.wikipedia.org/wiki/Demographics_of_Pakistan
Brazil https://en.wikipedia.org/wiki/Demographics_of_Brazil
Nigeria https://en.wikipedia.org/wiki/Demographics_of_Nigeria
Bangladesh https://en.wikipedia.org/wiki/Demographics_of_Bangladesh
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)

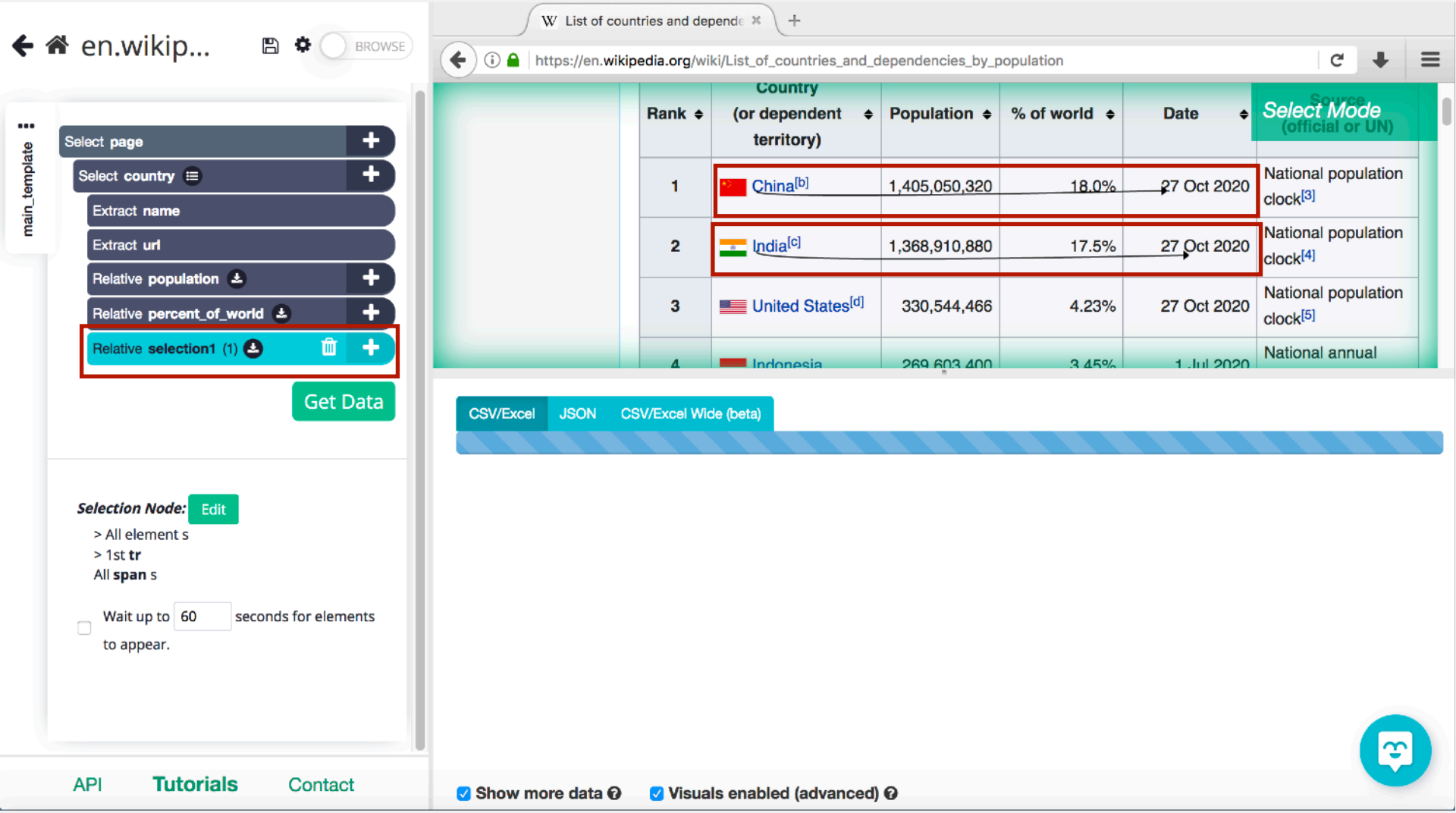

en.wikip...
BROWSE
main_template
Attach already-selected elements to related elements (auto-extracts data if possible). Good for grouping different fields together in your data.
Relative Select
Click
Advanced
Select page
Select country (202)
Extract name
Extract url
Get Data
Selection Node:
Edit
All td s
> All a s
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depende
https://en.wikipedia.
of world
Date
Select Mode
(official or UN)
18.0%
27 Oct 2020
National population clock[3]
17.5%
27 Oct 2020
National population clock[4]
4.23%
27 Oct 2020
National population clock[5]
National annual
CSV/Excel
JSON
country_name
China
India
United States
Indonesia
Pakistan
Brazil
Nigeria
Bangladesh
https://en.wikipedia.org/wiki/Demographics_of_Indonesia
https://en.wikipedia.org/wiki/Demographics_of_Pakistan
https://en.wikipedia.org/wiki/Demographics_of_Brazil
https://en.wikipedia.org/wiki/Demographics_of_Nigeria
https://en.wikipedia.org/wiki/Demographics_of_Bangladesh
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)

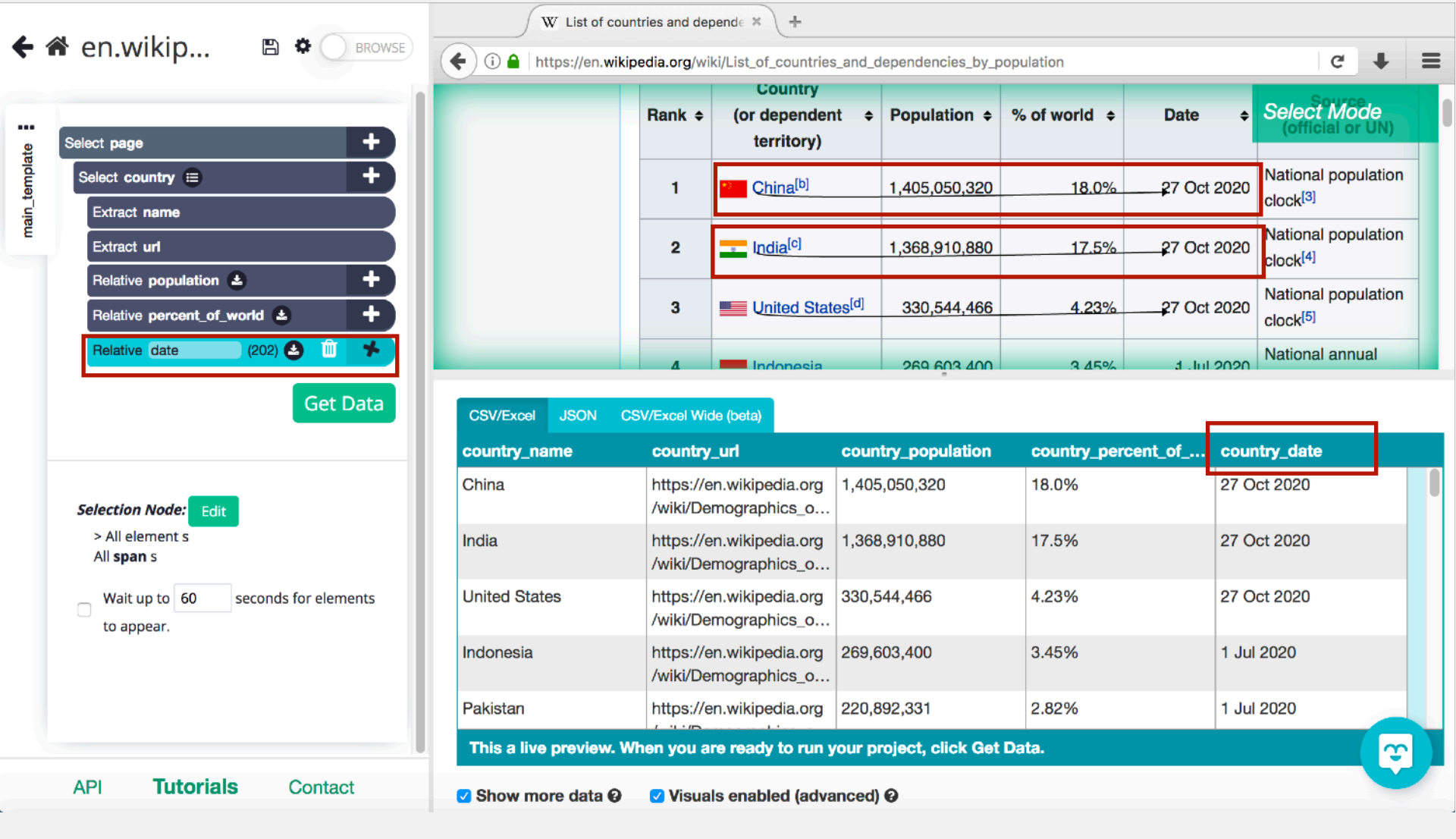

en.wikip...
BROWSE
main_template
Select page
Select country
Extract name
Extract url
Relative selection1 (1)
Get Data
Selection Node:
Edit
> All element s
> 1st tr
> 2nd td
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depend
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Select Mode
(official or UN)
1 China[b] 1,405,050,320 18.0% 27 Oct 2020 National population clock[3]
2 India[c] 1,368,910,880 17.5% 27 Oct 2020 National population clock[4]
3 United States[d] 330,544,466 4.23% 27 Oct 2020 National population clock[5]
4 Indonesia 269,603,400 3.45% 1 Jul 2020 National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
country_name
country_url
country_selection1
China
https://en.wikipedia.org/wiki/Demographics_of_China
1,405,050,320
India
https://en.wikipedia.org/wiki/Demographics_of_India
United States
https://en.wikipedia.org/wiki/Demographics_of_United_States
Indonesia
https://en.wikipedia.org/wiki/Demographics_of_Indonesia
Pakistan
https://en.wikipedia.org
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)


en.wikip...
BROWSE
main_template
Select page
Select country
Extract name
Extract url
Relative population (202)
Get Data
Selection Node:
Edit
> All element s
> 2nd td
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depende
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Source (official or UN)
Select Mode
1 China[b] 1,405,050,320 18.0% 27 Oct 2020 National population clock[3]
2 India[c] 1,368,910,880 17.5% 27 Oct 2020 National population clock[4]
3 United States[d] 330,544,466 4.23% 27 Oct 2020 National population clock[5]
4 Indonesia 269,603,400 3.45% 1 Jul 2020 National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
country_name
country_url
country_population
China https://en.wikipedia.org/wiki/Demographics_of_China 1,405,050,320
India https://en.wikipedia.org/wiki/Demographics_of_India 1,368,910,880
United States https://en.wikipedia.org/wiki/Demographics_of_United_States 330,544,466
Indonesia https://en.wikipedia.org/wiki/Demographics_of_Indonesia 269,603,400
Pakistan https://en.wikipedia.org 220,892,331
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)


en.wikip...
BROWSE
main_template
Select page
Select country
Extract name
Extract url
Relative population
Relative selection1 (1)
Get Data
Selection Node:
Edit
> All element s
> 1st tr
> 3rd td
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depend
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Source (official or UN)
Select Mode
1 China[b] 1,405,050,320 18.0% 27 Oct 2020 National population clock[3]
2 India[c] 1,368,910,880 17.5% 27 Oct 2020 National population clock[4]
3 United States[d] 330,544,466 4.23% 27 Oct 2020 National population clock[5]
4 Indonesia 269,603,400 3.45% 1 Jul 2020 National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
country_name
country_url
country_population
country_selection1
China https://en.wikipedia.org/wiki/Demographics_of_China 1,405,050,320 18.0%
India https://en.wikipedia.org/wiki/Demographics_of_India 1,368,910,880
United States https://en.wikipedia.org/wiki/Demographics_of_Unit... 330,544,466
Indonesia https://en.wikipedia.org/wiki/Demographics_of_Indo... 269,603,400
Pakistan https://en.wikipedia.org 220,892,331
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)


en.wikip...
BROWSE
main_template
Select page
Select country
Extract name
Extract url
Relative population
Relative cent_of_world (202)
Get Data
Selection Node:
Edit
> All element s
> 3rd td
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depende
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Source (official or UN)
Select Mode
1 China[b] 1,405,050,320 18.0% 27 Oct 2020 National population clock[3]
2 India[c] 1,368,910,880 17.5% 27 Oct 2020 National population clock[4]
3 United States[d] 330,544,466 4.23% 27 Oct 2020 National population clock[5]
4 Indonesia 269,603,400 3.45% 1 Jul 2020 National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
country_name
country_url
country_population
country_percent_of_world
China https://en.wikipedia.org/wiki/Demographics_of_China 1,405,050,320 18.0%
India https://en.wikipedia.org/wiki/Demographics_of_India 1,368,910,880 17.5%
United States https://en.wikipedia.org/wiki/Demographics_of_Unit... 330,544,466 4.23%
Indonesia https://en.wikipedia.org/wiki/Demographics_of_Indo... 269,603,400 3.45%
Pakistan https://en.wikipedia.org 220,892,331 2.82%
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)


en.wikip...
BROWSE
main_template
Select page
Select country
Extract name
Extract url
Relative population
Relative percent_of_world
Relative selection1 (1)
Get Data
Selection Node:
Edit
> All element s
> 1st tr
All span s
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depende
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Select Mode
(official or UN)
1
China[b]
1,405,050,320
18.0%
27 Oct 2020
National population clock[3]
2
India[c]
1,368,910,880
17.5%
27 Oct 2020
National population clock[4]
3
United States[d]
330,544,466
4.23%
27 Oct 2020
National population clock[5]
CSV/Excel
JSON
CSV/Excel Wide (beta)
Show more data
Visuals enabled (advanced)

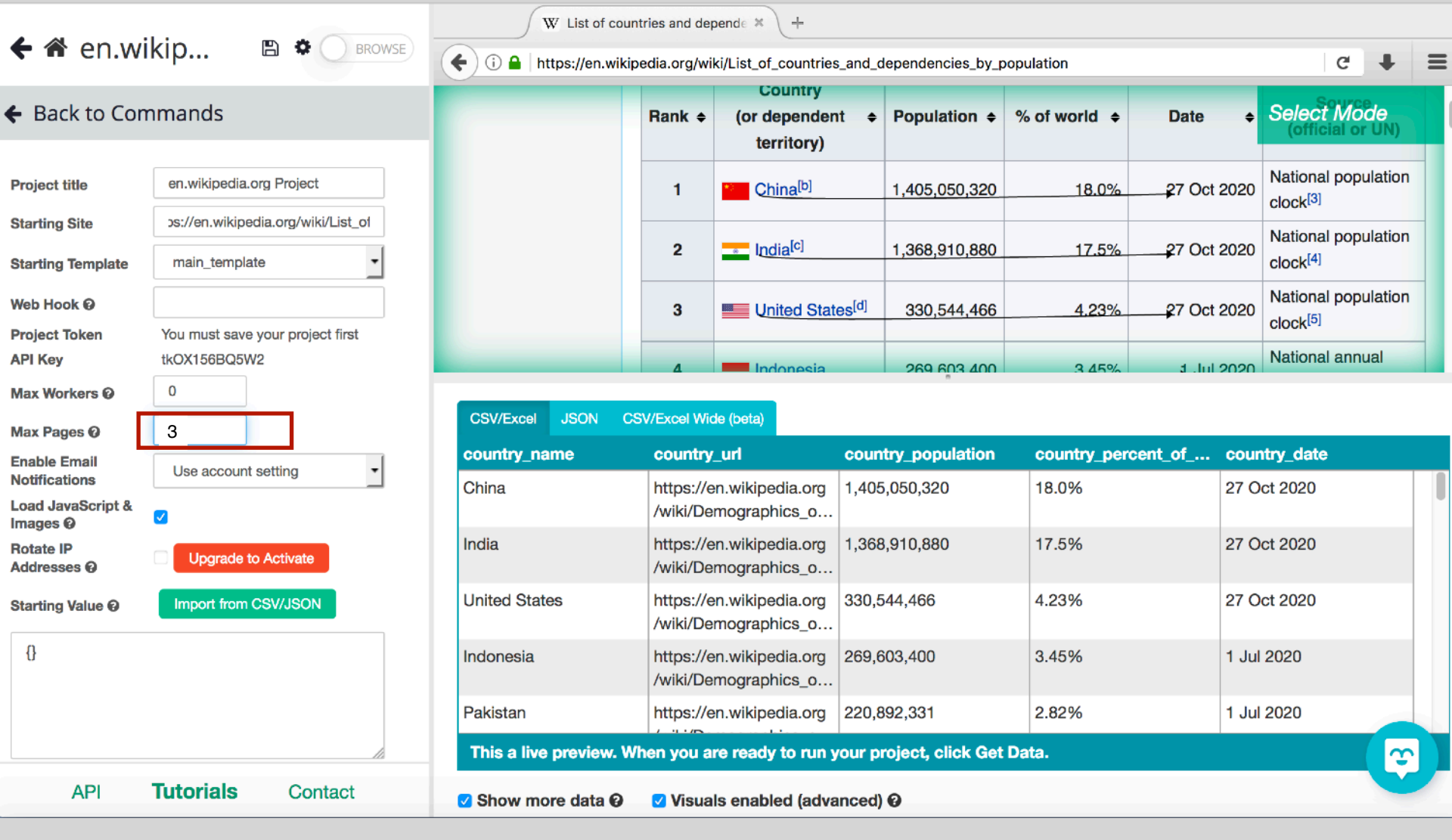

en.wikip...
BROWSE
main_template
Select page
Select country
Extract name
Extract url
Relative population
Relative percent_of_world
Relative date (202)
Get Data
Selection Node: Edit
> All element s
All span s
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
List of countries and depende
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Select Mode
Source (official or UN)
1 China[b] 1,405,050,320 18.0% 27 Oct 2020 National population clock[3]
2 India[c] 1,368,910,880 17.5% 27 Oct 2020 National population clock[4]
3 United States[d] 330,544,466 4.23% 27 Oct 2020 National population clock[5]
4 Indonesia 269,603,400 3.45% 1 Jul 2020 National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
country_name
country_url
country_population
country_percent_of_...
country_date
China https://en.wikipedia.org/wiki/Demographics_o... 1,405,050,320 18.0% 27 Oct 2020
India https://en.wikipedia.org/wiki/Demographics_o... 1,368,910,880 17.5% 27 Oct 2020
United States https://en.wikipedia.org/wiki/Demographics_o... 330,544,466 4.23% 27 Oct 2020
Indonesia https://en.wikipedia.org/wiki/Demographics_o... 269,603,400 3.45% 1 Jul 2020
Pakistan https://en.wikipedia.org 220,892,331 2.82% 1 Jul 2020
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)

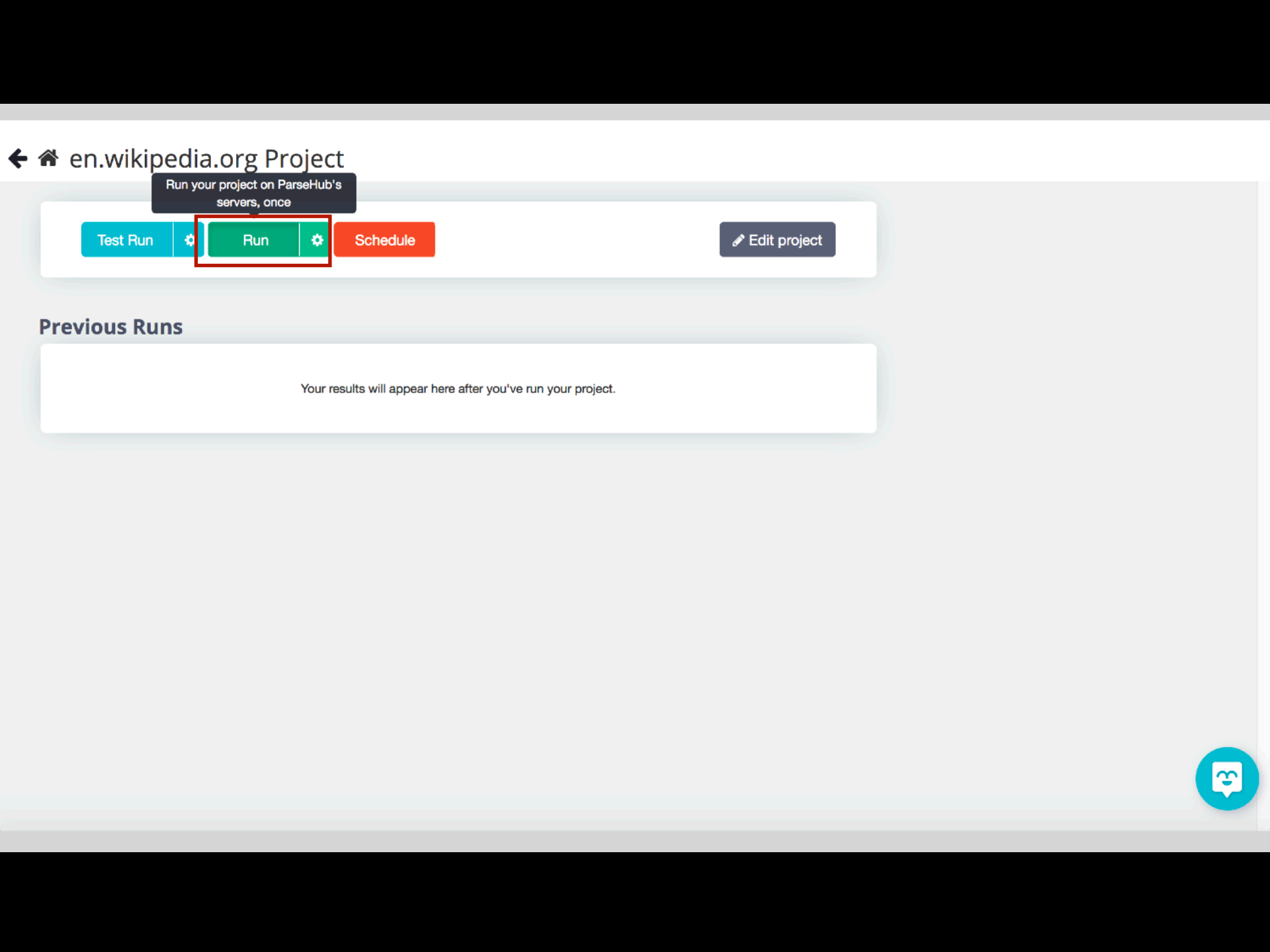
en.wikipedia.org Project

Test Run | Run | Schedule | Edit project

Previous Runs

Your results will appear here after you've run your project.


en.wikip...
BROWSE
Back to Commands
Project title
en.wikipedia.org Project
Starting Site
Starting Template
main_template
Web Hook
Project Token
You must save your project first
API Key
tkOX156BQ5W2
Max Workers
0
Max Pages
3
Enable Email Notifications
Use account setting
Load JavaScript & Images
Rotate IP Addresses
Upgrade to Activate
Starting Value
Import from CSV/JSON
{}
API
Tutorials
Contact
List of countries and depende
https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
Rank
Country (or dependent territory)
Population
% of world
Date
Source (official or UN)
Select Mode
1
China[b]
1,405,050,320
18.0%
27 Oct 2020
National population clock[3]
2
India[c]
1,368,910,880
17.5%
27 Oct 2020
National population clock[4]
3
United States[d]
330,544,466
4.23%
27 Oct 2020
National population clock[5]
4
Indonesia
269,603,400
3.45%
1 Jul 2020
National annual
CSV/Excel
JSON
CSV/Excel Wide (beta)
country_name
country_url
country_population
country_percent_of_...
country_date
China
https://en.wikipedia.org/wiki/Demographics_o...
1,405,050,320
18.0%
27 Oct 2020
India
https://en.wikipedia.org/wiki/Demographics_o...
1,368,910,880
17.5%
27 Oct 2020
United States
https://en.wikipedia.org/wiki/Demographics_o...
330,544,466
4.23%
27 Oct 2020
Indonesia
https://en.wikipedia.org/wiki/Demographics_o...
269,603,400
3.45%
1 Jul 2020
Pakistan
https://en.wikipedia.org
220,892,331
2.82%
1 Jul 2020
This a live preview. When you are ready to run your project, click Get Data.
Show more data
Visuals enabled (advanced)


en.wikipedia.org Project
Run your project on ParseHub's servers, once
Test Run
Run
Schedule
Edit project
Previous Runs
Your results will appear here after you've run your project.

en.wikipedia.org Project

Edit project

Data is being collected. Please wait.

Starting up...

Refreshing status in **2** second(s). Refresh now

Download Data

CSV/Excel | JSON | API

Cancel Run

All dates and times are in UTC +0000.

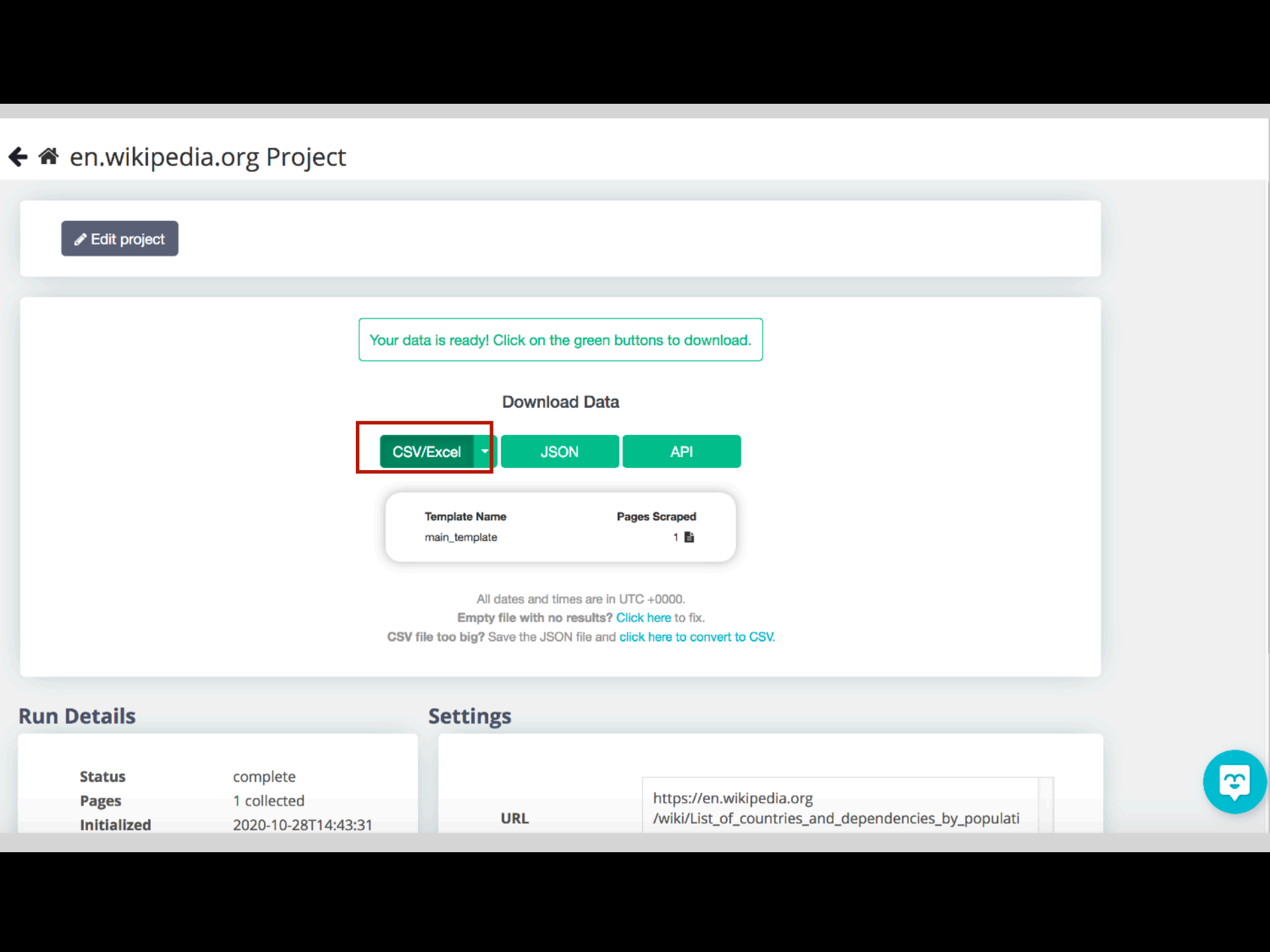
Empty file with no results? Click here to fix.

CSV file too big? Save the JSON file and click here to convert to CSV.

Run Details

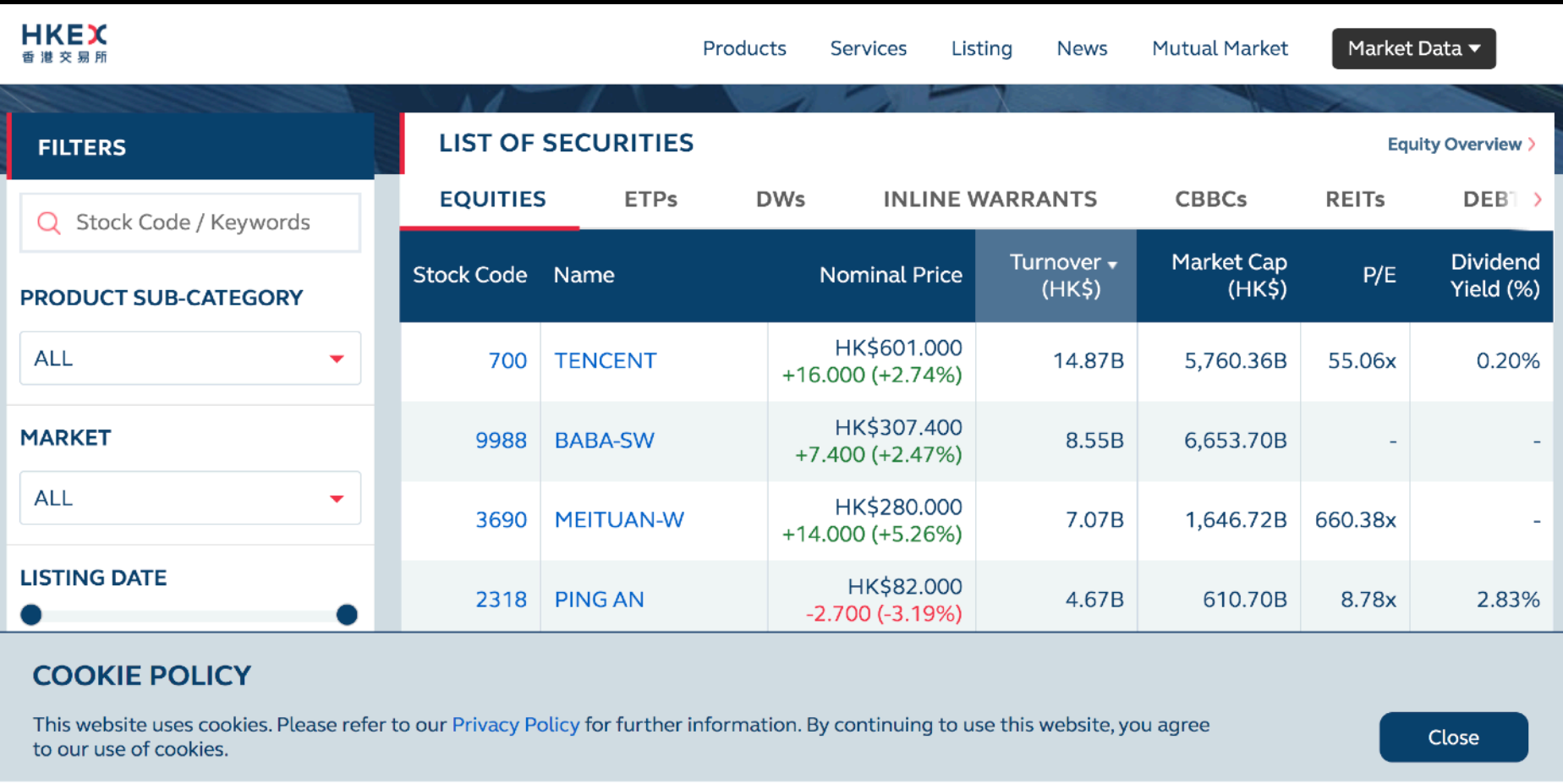
Settings

Started job


en.wikipedia.org Project
Edit project
Your data is ready! Click on the green buttons to download.
Download Data
CSV/Excel
JSON
API
Template Name
Pages Scraped
main_template
1
All dates and times are in UTC +0000.
Empty file with no results? Click here to fix.
CSV file too big? Save the JSON file and click here to convert to CSV.
Run Details
Settings
Status
complete
Pages
1 collected
Initialized
2020-10-28T14:43:31
URL
https://en.wikipedia.org
/wiki/List_of_countries_and_dependencies_by_populati

country_name	country_url	country_population	country_percent_of_world	country
China	https://en.wikipedia.org/wiki/Demographics_of_China	1,405,050,320	18.0%	27 Oct 2
India	https://en.wikipedia.org/wiki/Demographics_of_India	1,368,910,880	17.5%	27 Oct 2
United States	https://en.wikipedia.org/wiki/Demographics_of_United_States	330,544,466	4.23%	27 Oct 2
Indonesia	https://en.wikipedia.org/wiki/Demographics_of_Indonesia	269,603,400	3.45%	1 Jul 202
Pakistan	https://en.wikipedia.org/wiki/Demographics_of_Pakistan	220,892,331	2.82%	1 Jul 202
Brazil	https://en.wikipedia.org/wiki/Demographics_of_Brazil	212,254,617	2.71%	27 Oct 2
Nigeria	https://en.wikipedia.org/wiki/Demographics_of_Nigeria	206,139,587	2.64%	1 Jul 202
Bangladesh	https://en.wikipedia.org/wiki/Demographics_of_Bangladesh	169,542,128	2.17%	27 Oct 2
Russia	https://en.wikipedia.org/wiki/Demographics_of_Russia	146,748,590	1.88%	1 Jan 20
Mexico	https://en.wikipedia.org/wiki/Demographics_of_Mexico	127,792,286	1.63%	1 Jul 202
Japan	https://en.wikipedia.org/wiki/Demographics_of_Japan	125,880,000	1.61%	1 Oct 20
Philippines	https://en.wikipedia.org/wiki/Demographics_of_Philippines	109,349,790	1.40%	27 Oct 2
DR Congo	https://en.wikipedia.org/wiki/Demographics_of_Democratic_Republic_of_the_Congo	101,935,800	1.30%	1 Jul 202
Egypt	https://en.wikipedia.org/wiki/Demographics_of_Egypt	101,100,567	1.29%	27 Oct 2
Ethiopia	https://en.wikipedia.org/wiki/Demographics_of_Ethiopia	100,829,000	1.29%	1 Jul 202
Vietnam	https://en.wikipedia.org/wiki/Demographics_of_Vietnam	96,483,981	1.23%	1 Jul 20
Iran	https://en.wikipedia.org/wiki/Demographics_of_Iran	83,898,529	1.07%	27 Oct 2
Turkey	https://en.wikipedia.org/wiki/Demographics_of_Turkey	83,154,997	1.06%	31 Dec 2
Germany	https://en.wikipedia.org/wiki/Demographics_of_Germany	83,122,889	1.06%	30 Jun 2
France	https://en.wikipedia.org/wiki/Demographics_of_France	67,132,000	0.858%	1 Sep 20

Exercise time.



https://www.hkex.com.hk/Market-Data/Securities-Prices/Equities?sc_lang=en

Scraping multiple pages.

My Account | Login/ Register | English

Search entire store here...

CATALOG

My Cart HK$0.00

NEW IN | WOMEN | MEN | KIDS | FOOTWEAR | APPAREL | ACCESSORIES | BRANDS | SALE | KAKAO FRIENDS X CATALOG

Home ▸ Footwear

SHOP BY

CATEGORY

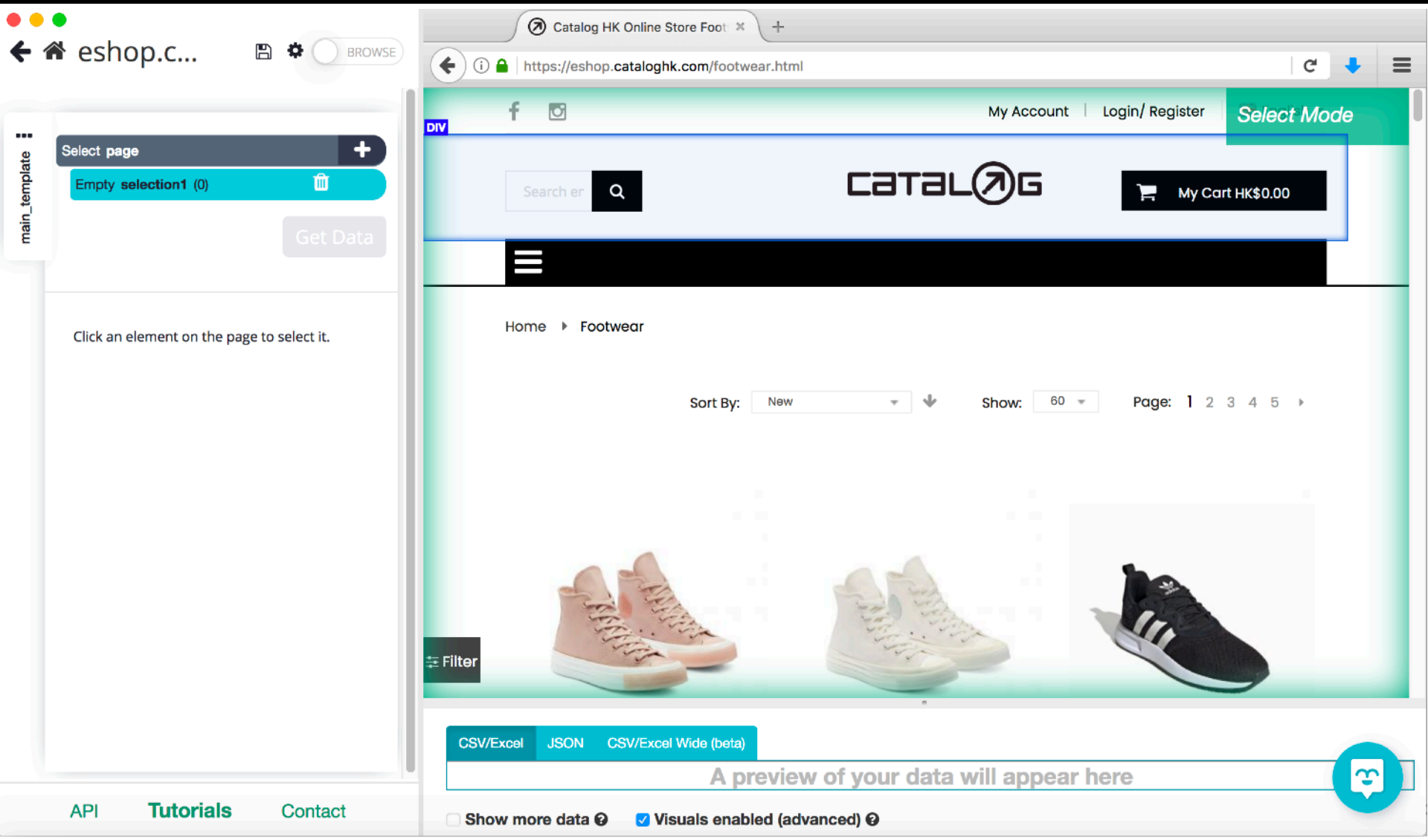
Sneakers	406
Casual Footwear	262
Sandals	133

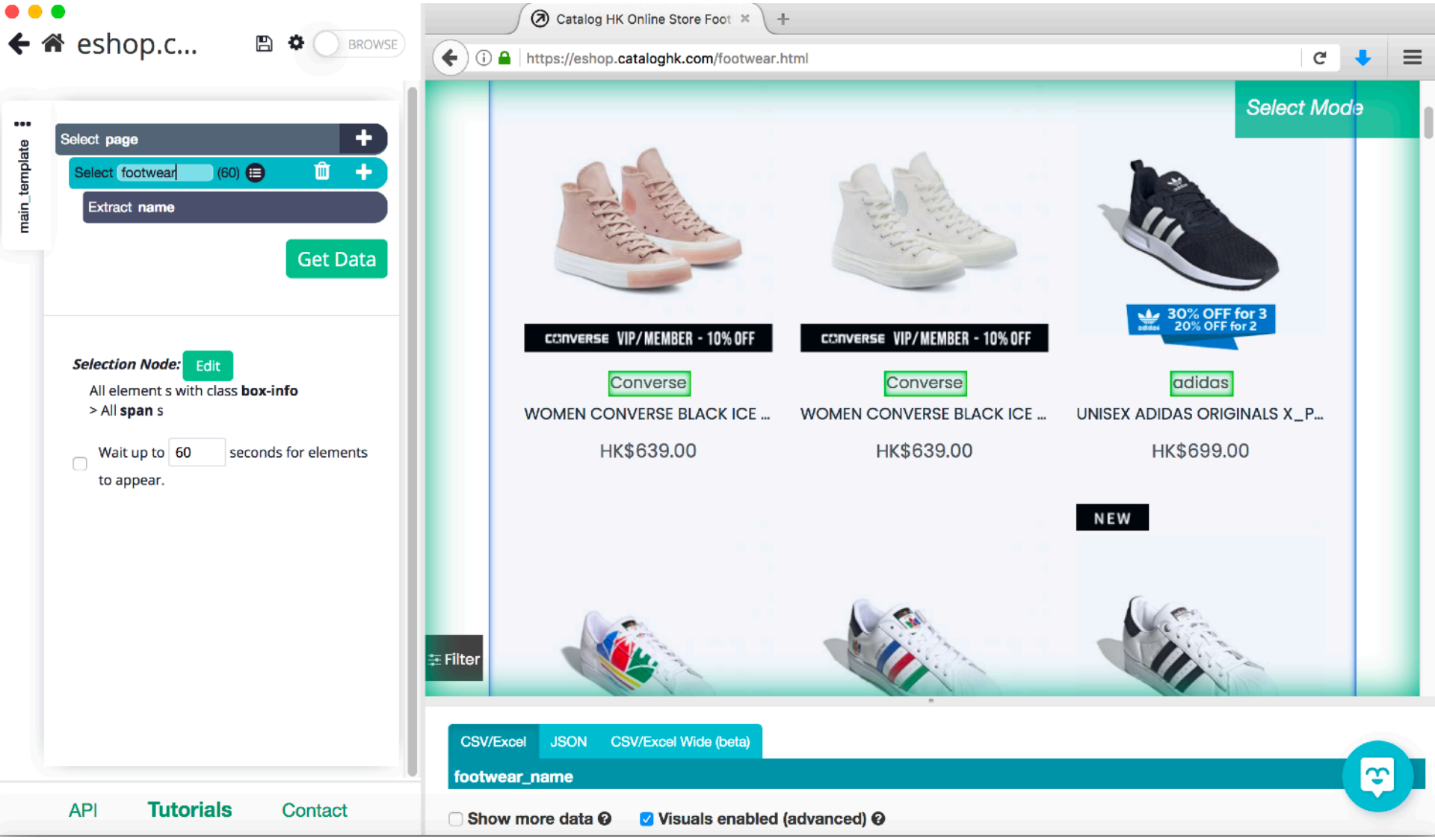
BRANDS

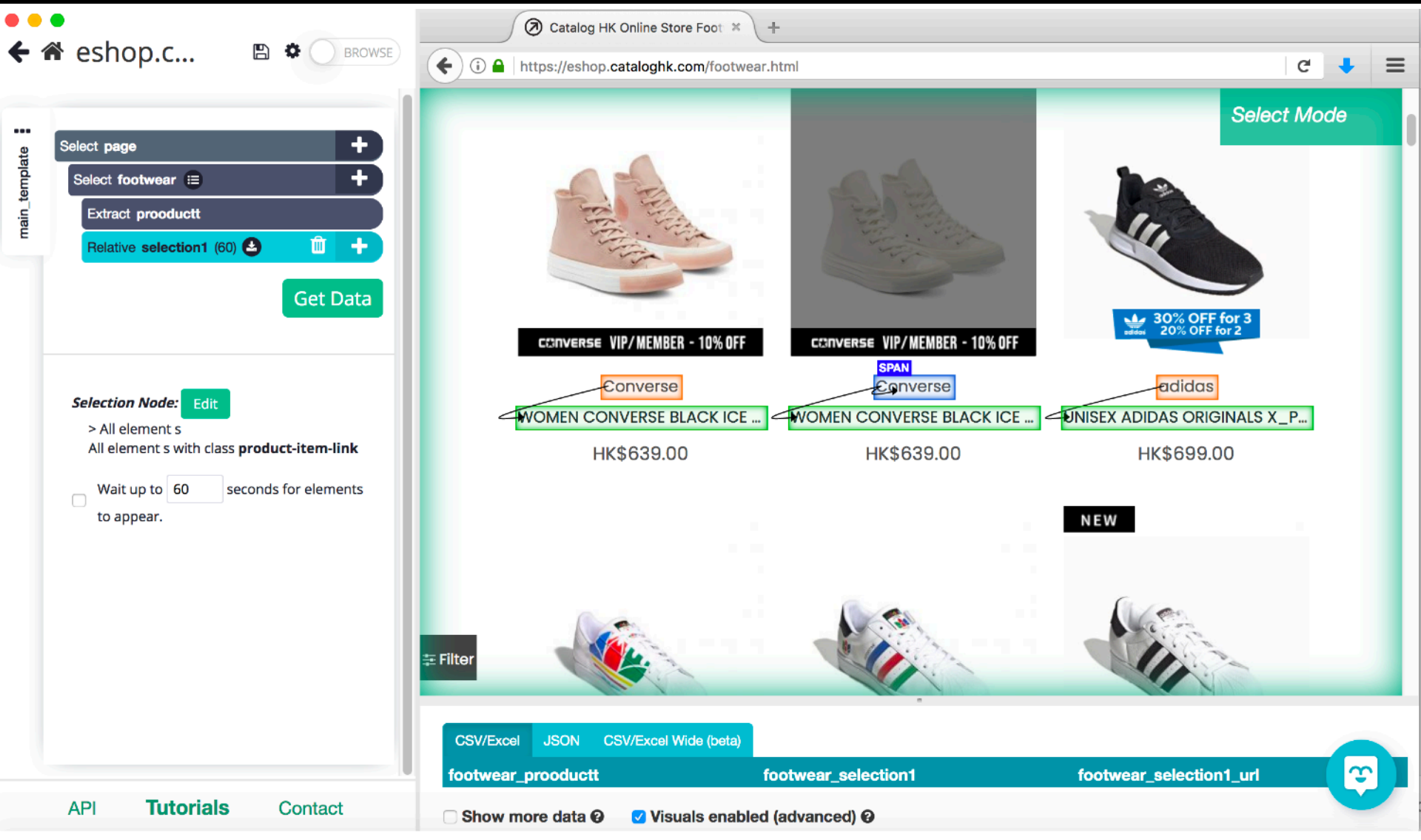
adidas Age Asics Birkenstock BT21 MEETS CATALOG Converse Crocs Dr. Martens Havaianas Jason Markk Keen

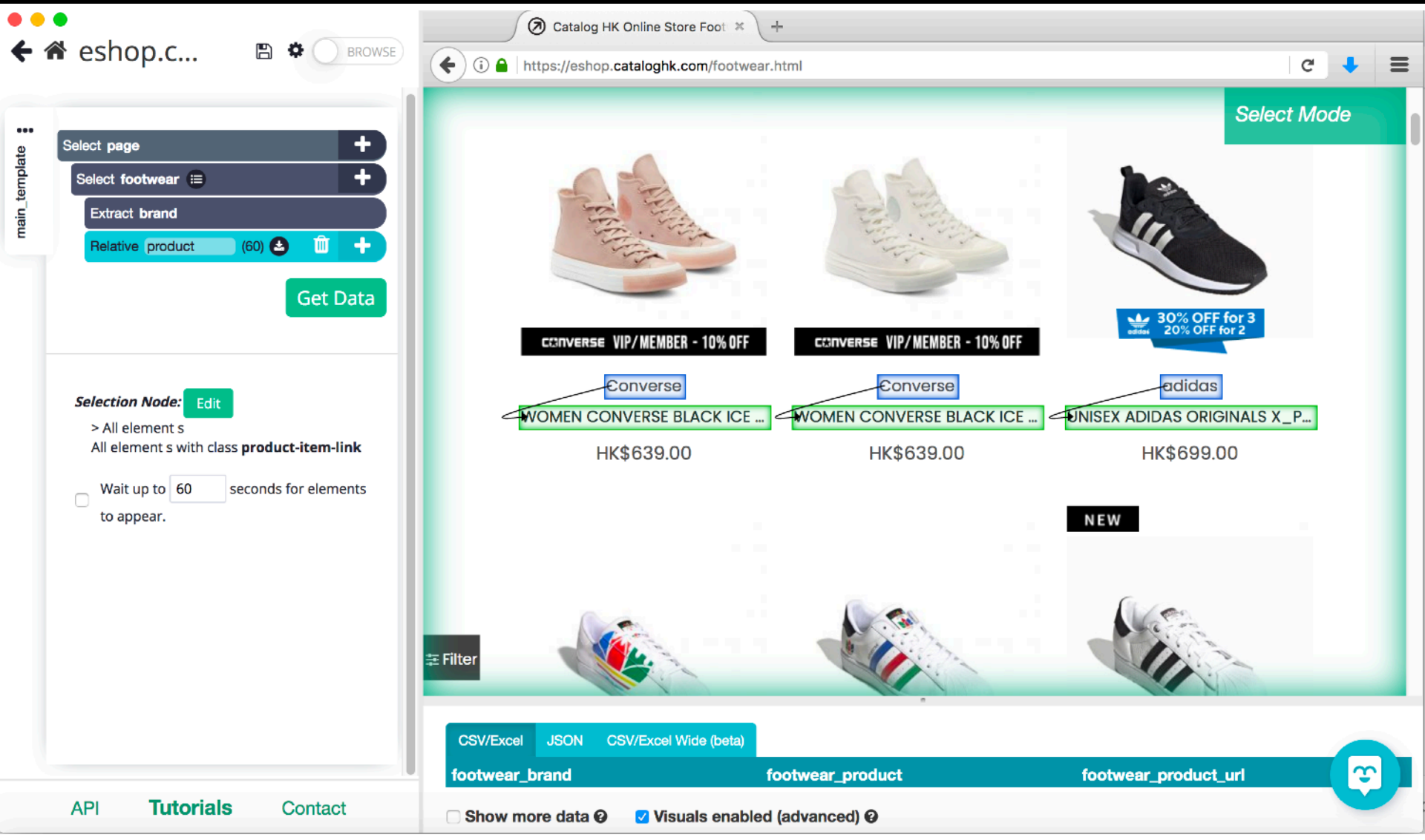
Sort By: New ↓ Show: 60 Page: 1 2 3 4 5 ▸

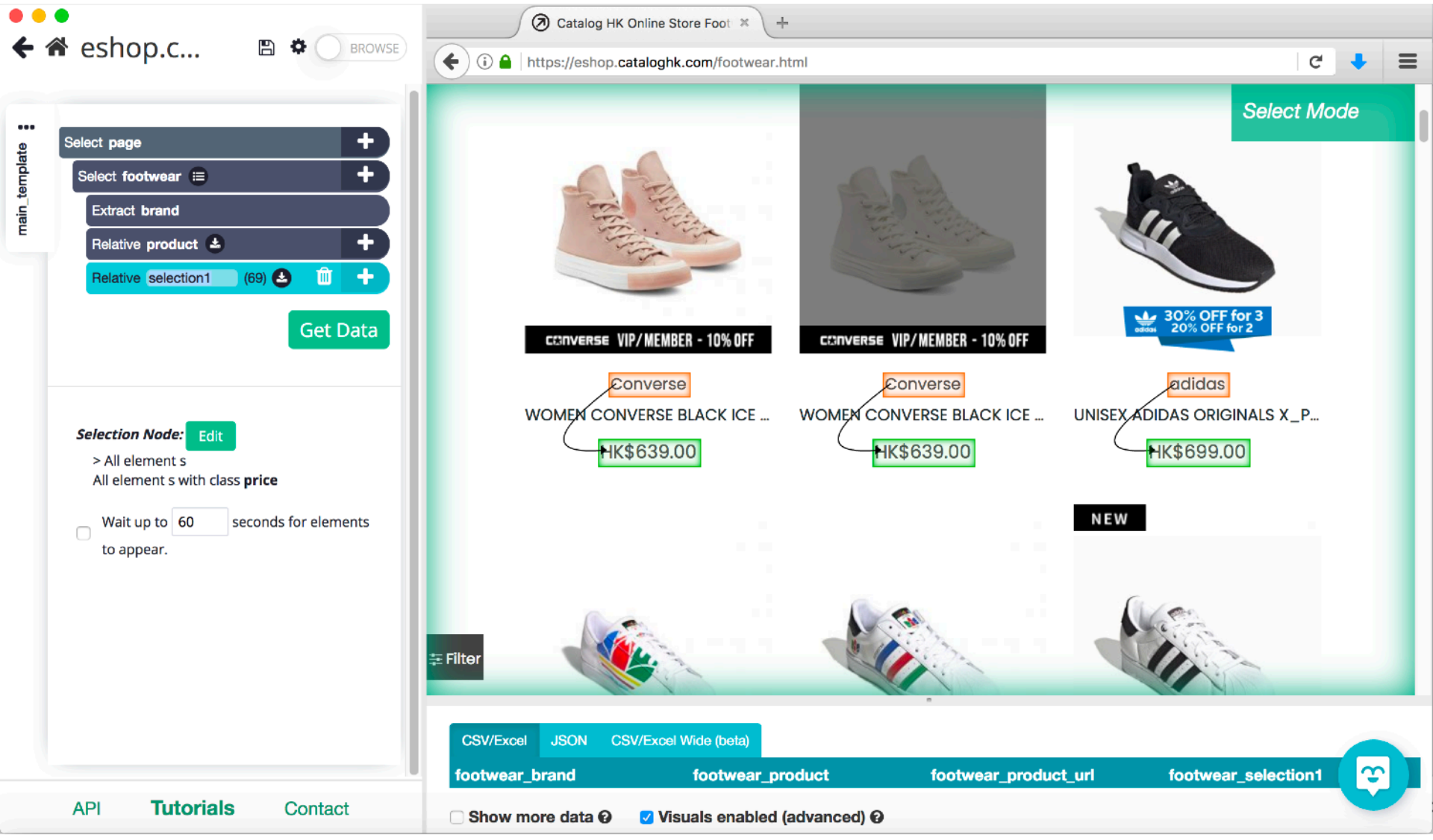
https://eshop.cataloghk.com/footwear.html

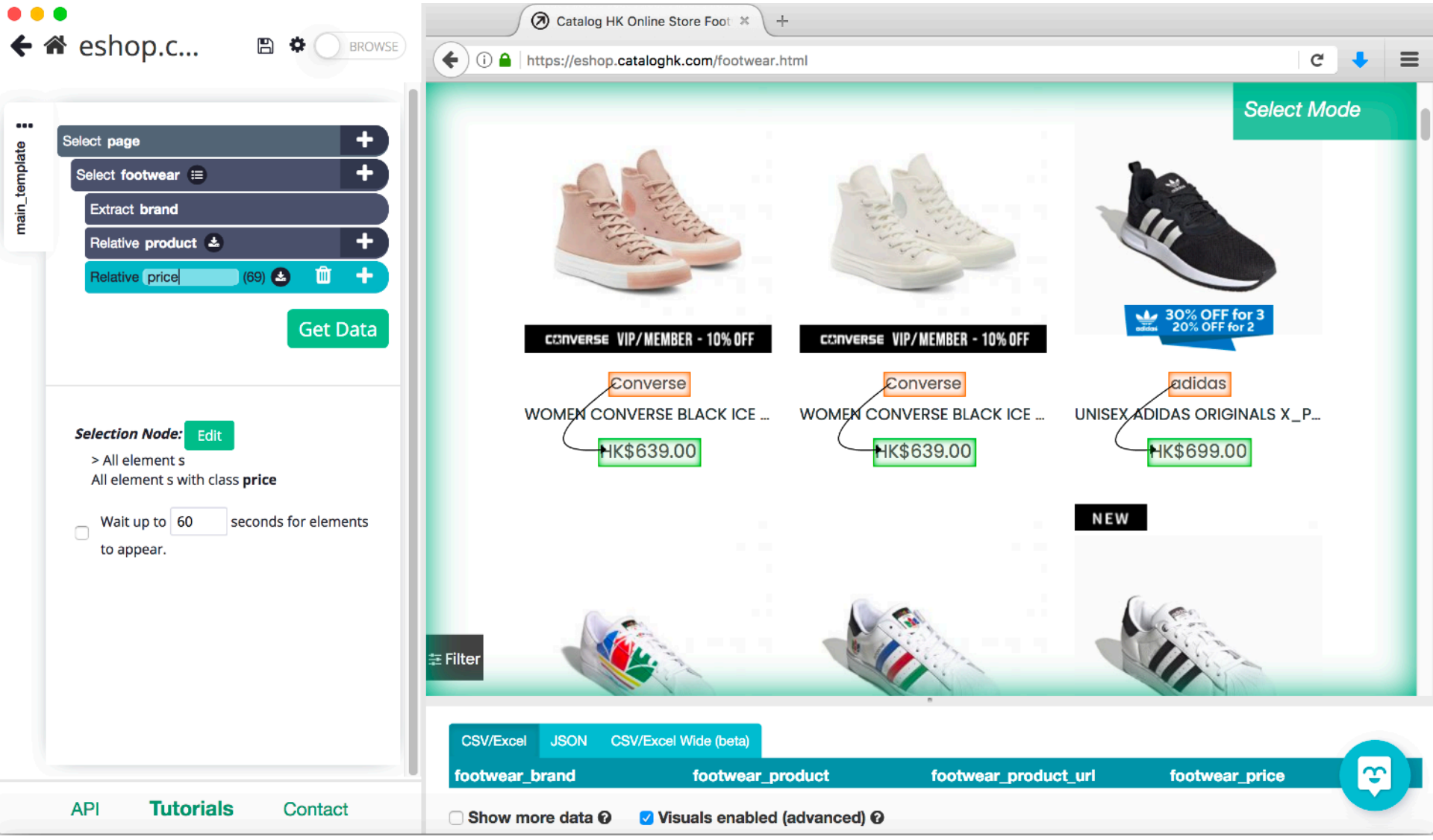

eshop.c...
BROWSE
main_template
Select page
Empty selection1 (0)
Get Data
Click an element on the page to select it.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
My Account
Login/ Register
Select Mode
DIV
Search er
CATALOG
My Cart HK$0.00
Home
Footwear
Sort By:
New
Show:
60
Page: 1 2 3 4 5
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
A preview of your data will appear here
Show more data
Visuals enabled (advanced)

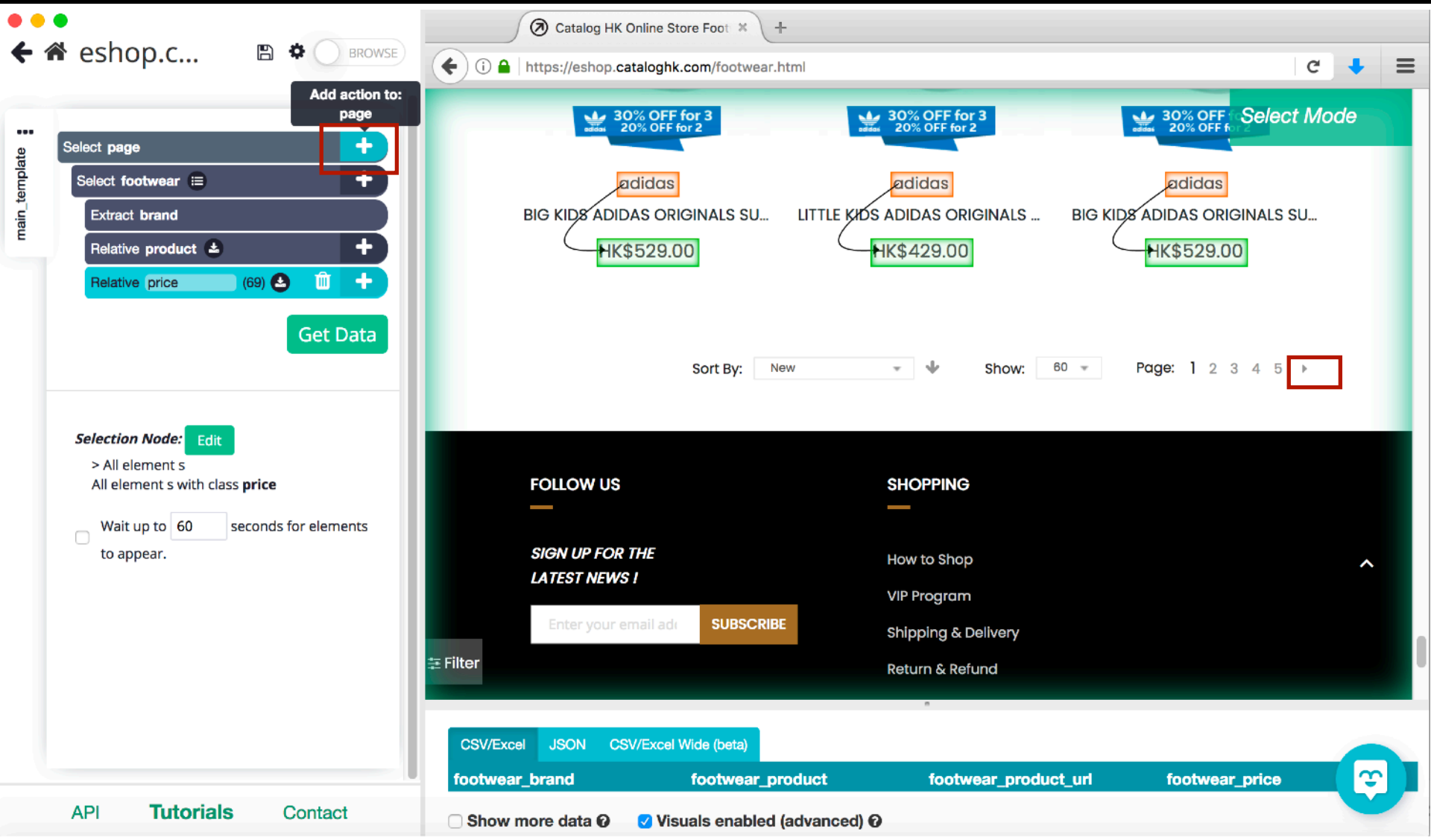

eshop.c...
BROWSE
main_template
Select page
Select footwear (60)
Extract name
Get Data
Selection Node:
Edit
All element s with class box-info
> All span s
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
Select Mode
CONVERSE VIP/MEMBER - 10% OFF
Converse
WOMEN CONVERSE BLACK ICE ...
HK$639.00
CONVERSE VIP/MEMBER - 10% OFF
Converse
WOMEN CONVERSE BLACK ICE ...
HK$639.00
30% OFF for 3
20% OFF for 2
adidas
UNISEX ADIDAS ORIGINALS X_P...
HK$699.00
NEW
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_name
Show more data
Visuals enabled (advanced)

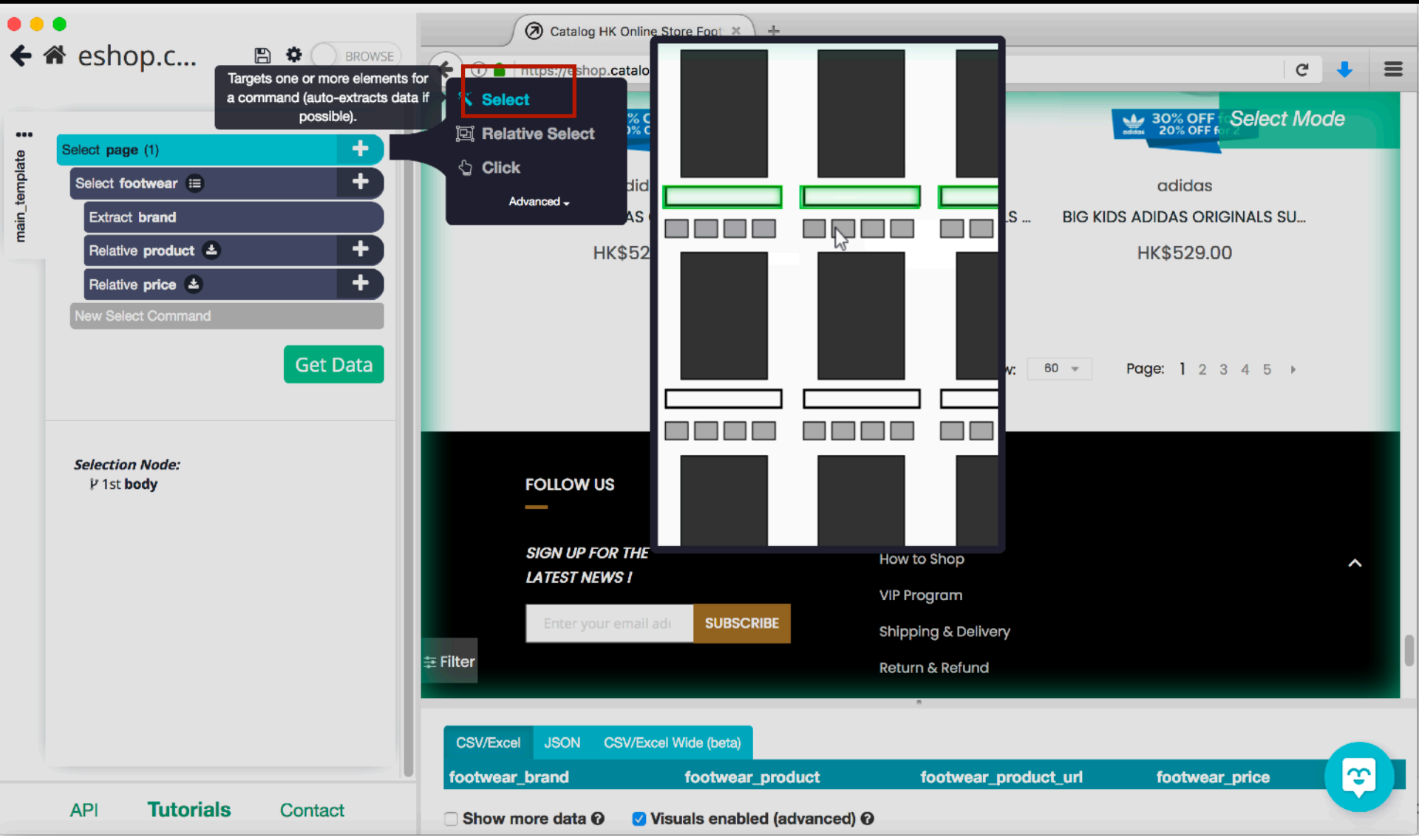

eshop.c...
BROWSE
main_template
Select page
Select footwear
Extract prooductt
Relative selection1 (60)
Get Data
Selection Node:
Edit
> All element s
All element s with class product-item-link
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
Select Mode
CONVERSE VIP/MEMBER - 10% OFF
CONVERSE VIP/MEMBER - 10% OFF
30% OFF for 3
20% OFF for 2
SPAN
Converse
Converse
adidas
WOMEN CONVERSE BLACK ICE ...
WOMEN CONVERSE BLACK ICE ...
UNISEX ADIDAS ORIGINALS X_P...
HK$639.00
HK$639.00
HK$699.00
NEW
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_prooductt
footwear_selection1
footwear_selection1_url
Show more data
Visuals enabled (advanced)

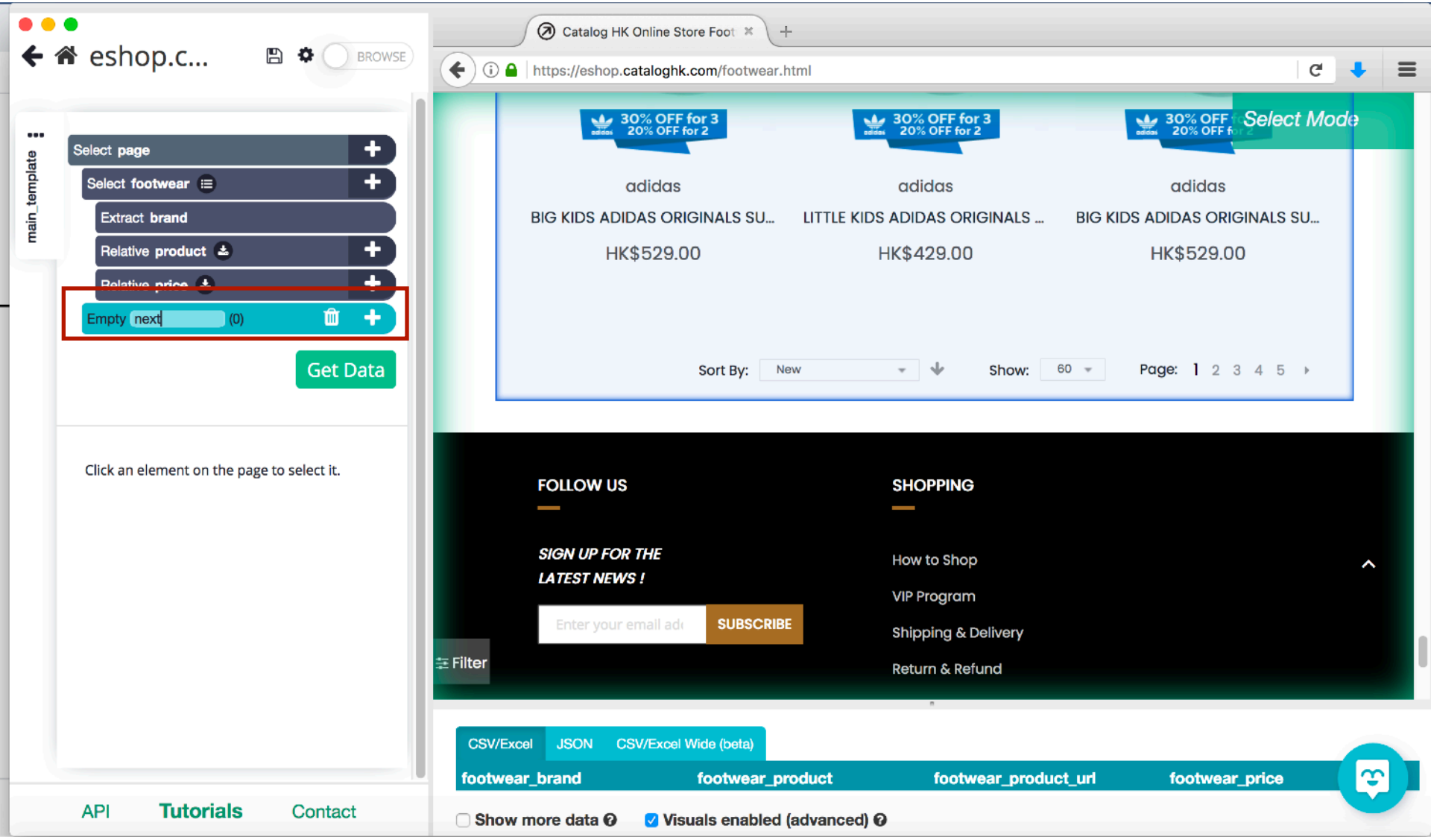

eshop.c...
BROWSE
main_template
Select page
Select footwear
Extract brand
Relative product (60)
Get Data
Selection Node:
Edit
> All element s
All element s with class product-item-link
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
Select Mode
CONVERSE VIP/MEMBER - 10% OFF
CONVERSE VIP/MEMBER - 10% OFF
30% OFF for 3
20% OFF for 2
Converse
Converse
adidas
WOMEN CONVERSE BLACK ICE ...
WOMEN CONVERSE BLACK ICE ...
UNISEX ADIDAS ORIGINALS X_P...
HK$639.00
HK$639.00
HK$699.00
NEW
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
Show more data
Visuals enabled (advanced)

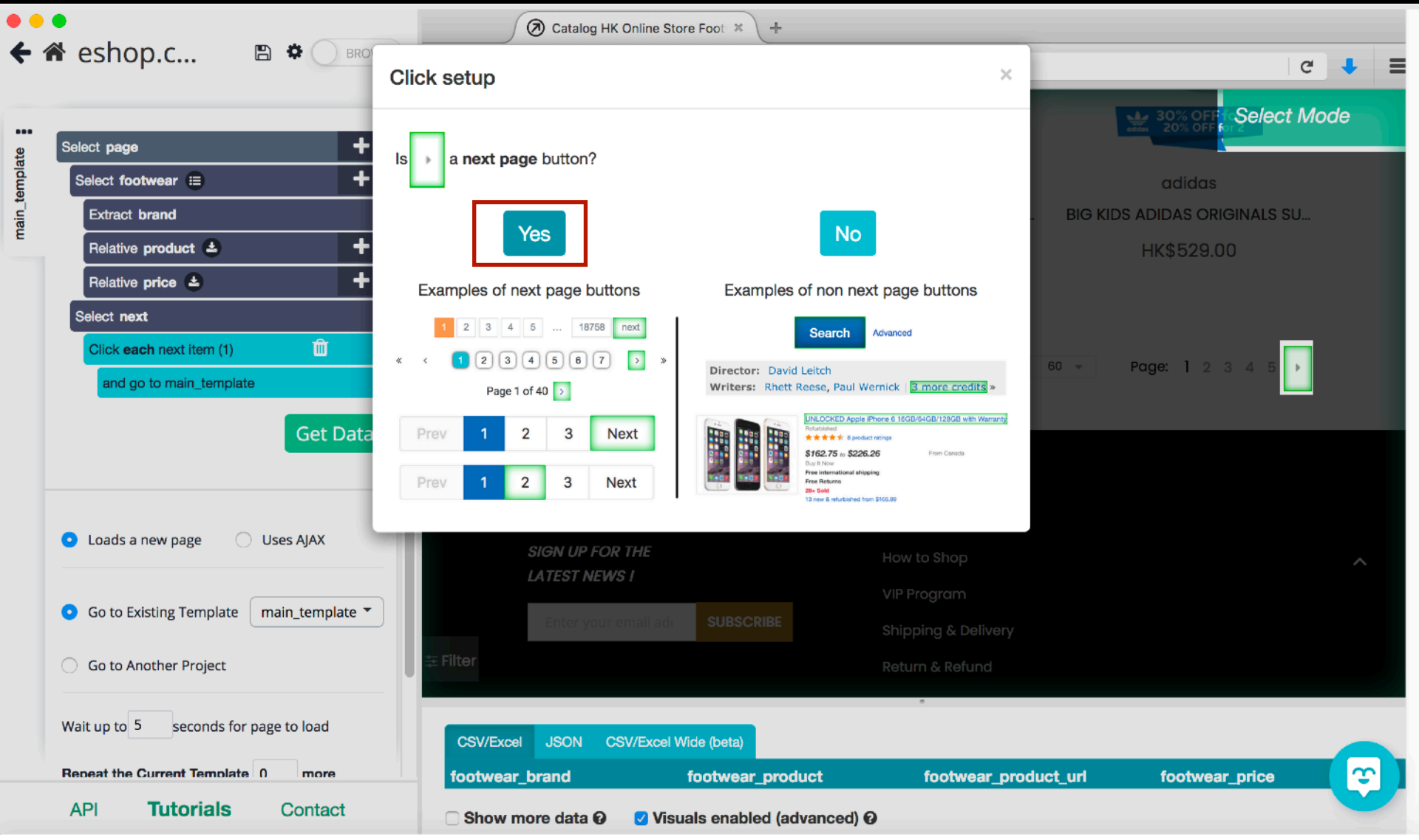

eshop.c...
BROWSE
main_template
Select page
Select footwear
Extract brand
Relative product
Relative selection1 (69)
Get Data
Selection Node:
Edit
> All element s
All element s with class price
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
Select Mode
CONVERSE VIP/MEMBER - 10% OFF
CONVERSE VIP/MEMBER - 10% OFF
30% OFF for 3
20% OFF for 2
Converse
WOMEN CONVERSE BLACK ICE ...
HK$639.00
Converse
WOMEN CONVERSE BLACK ICE ...
HK$639.00
adidas
UNISEX ADIDAS ORIGINALS X_P...
HK$699.00
NEW
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
footwear_selection1
Show more data
Visuals enabled (advanced)

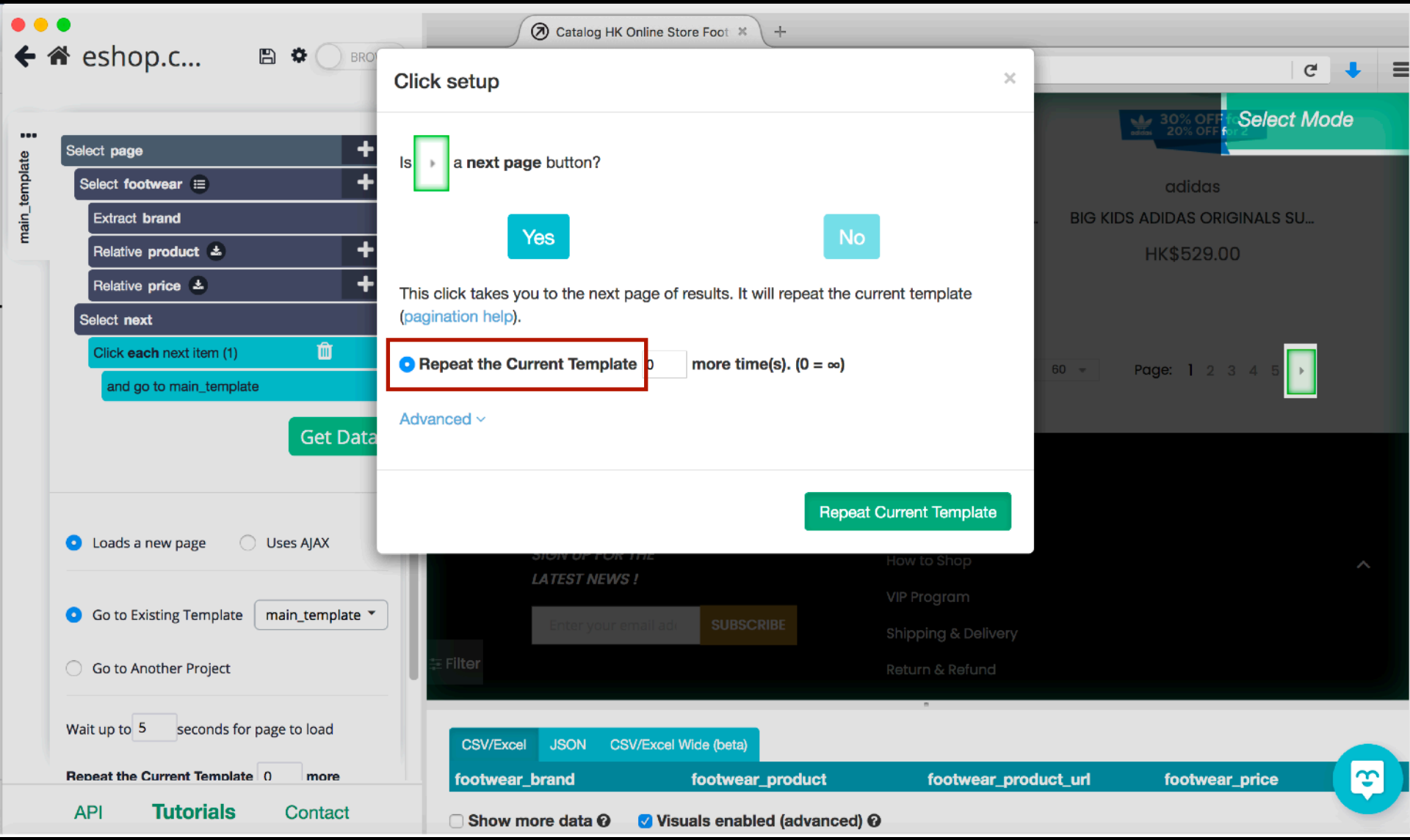

eshop.c...
BROWSE
main_template
Select page
Select footwear
Extract brand
Relative product
Relative price (69)
Get Data
Selection Node:
Edit
> All element s
All element s with class price
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
Select Mode
CONVERSE VIP/MEMBER - 10% OFF
CONVERSE VIP/MEMBER - 10% OFF
30% OFF for 3
20% OFF for 2
Converse
WOMEN CONVERSE BLACK ICE ...
HK$639.00
Converse
WOMEN CONVERSE BLACK ICE ...
HK$639.00
adidas
UNISEX ADIDAS ORIGINALS X_P...
HK$699.00
NEW
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
footwear_price
Show more data
Visuals enabled (advanced)

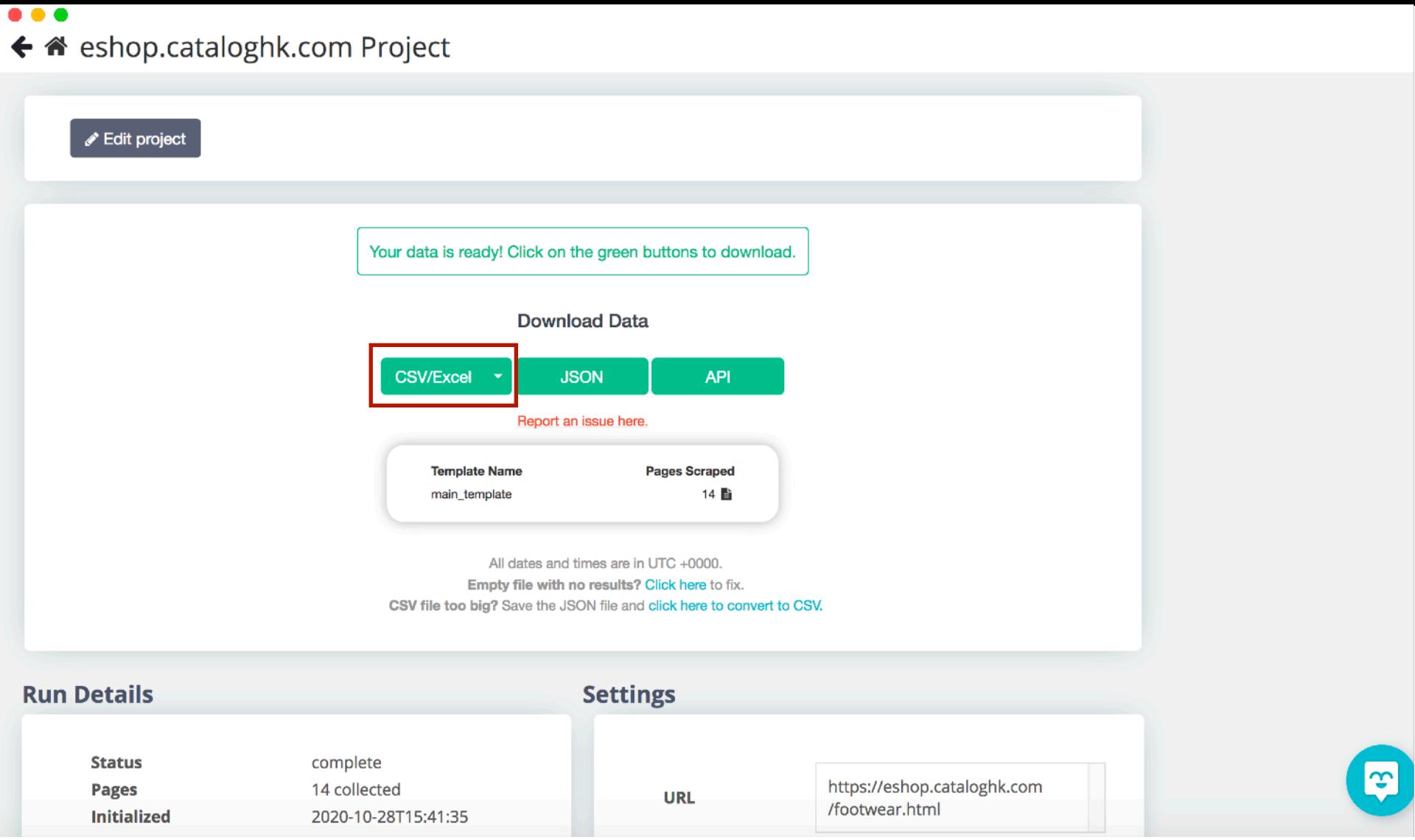

eshop.c...
BROWSE
Add action to: page
main_template
Select page
Select footwear
Extract brand
Relative product
Relative price (69)
Get Data
Selection Node:
Edit
> All element s
All element s with class price
Wait up to 60 seconds for elements to appear.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
30% OFF for 3 20% OFF for 2
adidas
BIG KIDS ADIDAS ORIGINALS SU...
HK$529.00
LITTLE KIDS ADIDAS ORIGINALS ...
HK$429.00
Select Mode
Sort By: New
Show: 60
Page: 1 2 3 4 5
FOLLOW US
SIGN UP FOR THE LATEST NEWS !
Enter your email ad
SUBSCRIBE
SHOPPING
How to Shop
VIP Program
Shipping & Delivery
Return & Refund
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
footwear_price
Show more data
Visuals enabled (advanced)


eshop.c...
BROWSE
Targets one or more elements for a command (auto-extracts data if possible).
Select
Relative Select
Click
Advanced
main_template
Select page (1)
Select footwear
Extract brand
Relative product
Relative price
New Select Command
Get Data
Selection Node:
1st body
Catalog HK Online Store Foot
Select Mode
30% OFF
20% OFF
adidas
BIG KIDS ADIDAS ORIGINALS SU...
HK$529.00
Page: 1 2 3 4 5
FOLLOW US
SIGN UP FOR THE LATEST NEWS !
Enter your email ad
SUBSCRIBE
How to Shop
VIP Program
Shipping & Delivery
Return & Refund
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
footwear_price
Show more data
Visuals enabled (advanced)
API
Tutorials
Contact


eshop.c...
BROWSE
main_template
Select page
Select footwear
Extract brand
Relative product
Relative price
Empty next (0)
Get Data
Click an element on the page to select it.
API
Tutorials
Contact
Catalog HK Online Store Foot
https://eshop.cataloghk.com/footwear.html
Select Mode
30% OFF for 3
20% OFF for 2
adidas
BIG KIDS ADIDAS ORIGINALS SU...
HK$529.00
adidas
LITTLE KIDS ADIDAS ORIGINALS ...
HK$429.00
adidas
BIG KIDS ADIDAS ORIGINALS SU...
HK$529.00
Sort By: New
Show: 60
Page: 1 2 3 4 5
FOLLOW US
SIGN UP FOR THE LATEST NEWS !
Enter your email ad
SUBSCRIBE
SHOPPING
How to Shop
VIP Program
Shipping & Delivery
Return & Refund
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
footwear_price
Show more data
Visuals enabled (advanced)


eshop.c...
main_template
Select page
Select footwear
Extract brand
Relative product
Relative price
Select next
Click each next item (1)
and go to main_template
Get Data
Loads a new page
Uses AJAX
Go to Existing Template
main_template
Go to Another Project
Wait up to 5 seconds for page to load
Repeat the Current Template 0 more
API
Tutorials
Contact
Catalog HK Online Store Foot
Click setup
Is a next page button?
Yes
No
Examples of next page buttons
Examples of non next page buttons
Page 1 of 40
Prev 1 2 3 Next
Search
Advanced
Director: David Leitch
Writers: Rhett Reese, Paul Wernick
3 more credits
Select Mode
adidas
BIG KIDS ADIDAS ORIGINALS SU...
HK$529.00
Page: 1 2 3 4 5
SIGN UP FOR THE LATEST NEWS !
SUBSCRIBE
How to Shop
VIP Program
Shipping & Delivery
Return & Refund
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
footwear_price
Show more data
Visuals enabled (advanced)


eshop.c...
main_template
Select page
Select footwear
Extract brand
Relative product
Relative price
Select next
Click each next item (1)
and go to main_template
Get Data
Loads a new page
Uses AJAX
Go to Existing Template
main_template
Go to Another Project
Wait up to 5 seconds for page to load
Repeat the Current Template 0 more
API
Tutorials
Contact
Catalog HK Online Store Foot
Click setup
Is a next page button?
Yes
No
This click takes you to the next page of results. It will repeat the current template (pagination help).
Repeat the Current Template 0 more time(s). (0 = ∞)
Advanced
Repeat Current Template
Select Mode
adidas
BIG KIDS ADIDAS ORIGINALS SU...
HK$529.00
Page: 1 2 3 4 5
LATEST NEWS !
SUBSCRIBE
How to Shop
VIP Program
Shipping & Delivery
Return & Refund
Filter
CSV/Excel
JSON
CSV/Excel Wide (beta)
footwear_brand
footwear_product
footwear_product_url
footwear_price
Show more data
Visuals enabled (advanced)

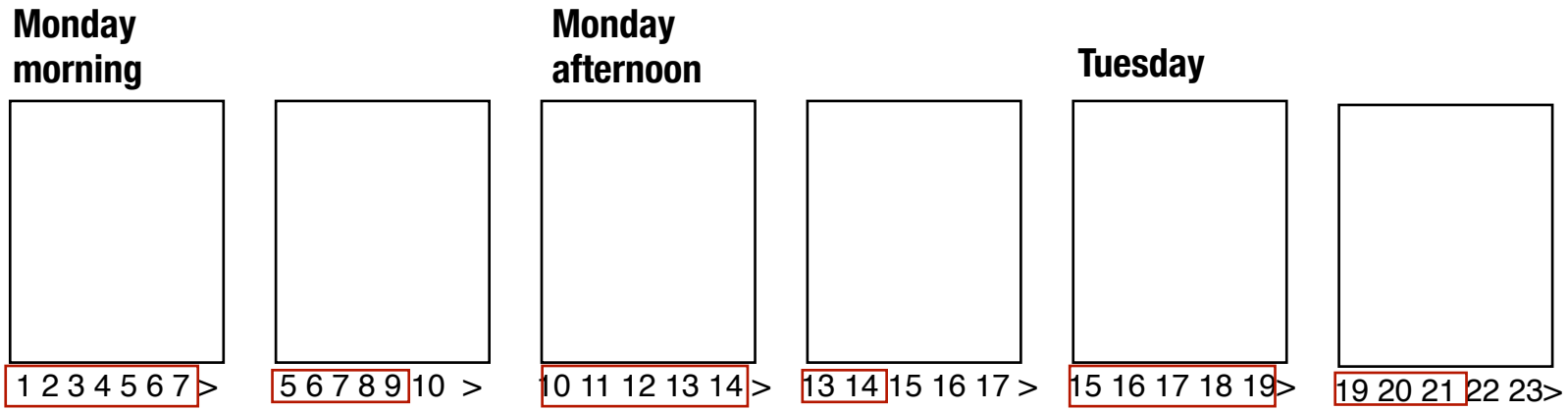

eshop.cataloghk.com Project
Edit project
Your data is ready! Click on the green buttons to download.
Download Data
CSV/Excel
JSON
API
Report an issue here.
Template Name
Pages Scraped
main_template
14
All dates and times are in UTC +0000.
Empty file with no results? Click here to fix.
CSV file too big? Save the JSON file and click here to convert to CSV.
Run Details
Status
complete
Pages
14 collected
Initialized
2020-10-28T15:41:35
Settings
URL
https://eshop.cataloghk.com
/footwear.html

Converse	WOMEN CONVERSE BLACK ICE CHUCK 70 HI TOP	https://eshop.cataloghk.com/150569539c.html	H
Converse	WOMEN CONVERSE BLACK ICE CHUCK 70 HI TOP	https://eshop.cataloghk.com/150569540c.html	H
adidas	UNISEX ADIDAS ORIGINALS X_PLR S	https://eshop.cataloghk.com/108fy2852.html	H
adidas	UNISEX ADIDAS ORIGINALS SUPERSTAR 2020 PURE	https://eshop.cataloghk.com/108fu9519.html	H
adidas	UNISEX ADIDAS ORIGINALS SUPERSTAR 2020	https://eshop.cataloghk.com/108fu9521.html	H
adidas	WOMEN ADIDAS ORIGINALS SUPERSTAR	https://eshop.cataloghk.com/108h69025.html	H
adidas	UNISEX ADIDAS ORIGINALS SUPERSTAR 2020	https://eshop.cataloghk.com/108fw5388.html	H
Converse	UNISEX CONVERSE CHUCK 70 HIGH TOP	https://eshop.cataloghk.com/150169336c.html	H
Converse	WOMEN CONVERSE CHUCK TAYLOR ALL STAR LIFT LEATHER HIGH TOP	https://eshop.cataloghk.com/150561676c.html	H
Converse	WOMEN CONVERSE CHUCK TAYLOR ALL STAR LIFT LEATHER HIGH TOP	https://eshop.cataloghk.com/150561675c.html	H
Converse	UNISEX CONVERSE SEASONAL COLOUR VINTAGE CANVAS CHUCK 70 HIGH TOP	https://eshop.cataloghk.com/150169341c.html	H
Converse	UNISEX CONVERSE VINTAGE CANVAS CHUCK 70 HIGH TOP	https://eshop.cataloghk.com/150169342c.html	H
Converse	UNISEX CONVERSE SEASONAL COLOUR VINTAGE CANVAS CHUCK 70 LOW TOP	https://eshop.cataloghk.com/150169343c.html	H
Converse	UNISEX CONVERSE SEASONAL COLOUR VINTAGE CANVAS CHUCK 70 LOW TOP	https://eshop.cataloghk.com/150169344c.html	H
Keen	WOMEN KEEN UNEEK SNK	https://eshop.cataloghk.com/1kn1023508.html	HK
Age	UNISEX AGE CUT KILLERWHALE	https://eshop.cataloghk.com/1abgrctwhbk011.html	H
Keen	WOMEN KEEN UNEEK SNK	https://eshop.cataloghk.com/1kn1023506.html	HK
PUMA	WOMEN PUMA CILIA LUX	https://eshop.cataloghk.com/1pu37028215.html	H
Age	UNISEX AGE TOP REFLECTIVE AURORA	https://eshop.cataloghk.com/1abrftoprnbk013.html	H
Keen	MEN KEEN UNEEK SNK	https://eshop.cataloghk.com/1kn1023500.html	HK
Age	UNISEX AGE CUT MA-1	https://eshop.cataloghk.com/1abmactolbk011.html	H
Age	UNISEX AGE CUT REFLECTIVE AURORA	https://eshop.cataloghk.com/1abrfctrnbk011.html	H

Sheet 1

	A	B	C
	PUMA	WOMEN PUMA ROMA AMOR HERITAGE WNS	https://eshop.cataloghk.com/1bd3709470f.html
	adidas	UNISEX ADIDAS ORIGINALS OZWEEGO	https://eshop.cataloghk.com/108ee7002.html
793	STARE	WOMEN STARE M4807 ORIGINAL	https://eshop.cataloghk.com/1tsstu91lo006bg.html
794	BT21 MEETS CA	WOMEN BT21 MEETS CATALOG COOKY SLIP ON	https://eshop.cataloghk.com/1b1sjdf19bt132pn.html
795	VANS	UNISEX VANS BRUSHED TWILL STYLE 36	https://eshop.cataloghk.com/105vn0a3dz3vlq.html
796	adidas	WOMEN ADIDAS ORIGINALS CONTINENTAL 80	https://eshop.cataloghk.com/108eg4592.html
797	Reebok	LITTLE KIDS REEBOK VERSA PUMP FURY	https://eshop.cataloghk.com/101dv8544.html
798	Reebok	UNISEX REEBOK CLASSIC SLIDE	https://eshop.cataloghk.com/101dv3698.html
799	Reebok	UNISEX REEBOK CLASSIC SLIDE	https://eshop.cataloghk.com/101dv3697.html
800	Converse	UNISEX CONVERSE CHUCK TAYLOR ALL STAR LOW TOP	https://eshop.cataloghk.com/150m9007c.html
801	New Balance	UNISEX NEW BALANCE 200 SLIDES	https://eshop.cataloghk.com/112smf200b1.html
802	Reebok	BIG KIDS REEBOK ROYAL RIPPLE WHITE/PURPLE	https://eshop.cataloghk.com/101dv4329.html
803	Reebok	BIG KIDS REEBOK ROYAL RIPPLE WHITE/PINK	https://eshop.cataloghk.com/101dv4328.html
804	adidas	UNISEX ADIDAS ORIGINALS U_PATH RUN BLACK	https://eshop.cataloghk.com/108g27636.html
805	adidas	WOMEN ADIDAS ORIGINALS SWIFT RUN W WHITE	https://eshop.cataloghk.com/108d96647.html
806	VANS	LITTLE KIDS VANS OLD SKOOL BLACK AND WHITE CHECKERBOARD	https://eshop.cataloghk.com/105vn000ex8bww.html
807	Converse	BIG KIDS CONVERSE STAR COURT 2V OX WHITE	https://eshop.cataloghk.com/150762864c.html
808	VANS	UNISEX VANS STYLE 36 DECON SF	https://eshop.cataloghk.com/105vn0a3mvlqc5.html
809	Converse	UNISEX CONVERSE CHUCK 70 HI	https://eshop.cataloghk.com/150162052c.html
810	adidas	MEN ADIDAS ORIGINALS STAN SMITH CF WHITE	https://eshop.cataloghk.com/108s75187.html
811	Jason Markk	JASON MARKK ESSENTIAL KIT	https://eshop.cataloghk.com/3ja003500.html

Periodic scraping in smaller batches.


Monday
morning
Monday
afternoon
Tuesday
1 2 3 4 5 6 7 >
5 6 7 8 9 10 >
10 11 12 13 14 >
13 14 15 16 17 >
15 16 17 18 19 >
19 20 21 22 23 >

Exercise time.

登入 註冊 | 訂閱 / 續訂

信報財經新聞

熱門：港區國安法 新冠肺炎　搜尋：搶奪中環人 ESG報道系列

搜尋

齊心抗疫，優惠再延長！

信健康 StartUpBeat EJFQ 港股360 信報月刊 優雅生活 ejinsight 信報通識 信報出版 好·食·玩

主頁 即時新聞 今日信報 港股360 地產投資 財富管理 專題 信報視頻

最新 新聞專題 人物訪問 生活休閒 信報教育 活動專輯 全部

瘟疫大流行 >全球疫情發展

南韓增36宗新冠肺炎確診 逾半屬境外輸入

南韓中央防疫對策本部稱，當地新增36宗感染新型冠狀病毒肺炎的確診病例，累計近1.43萬宗。另外，當地新增1例死 ...

2020年7月31日

全球確診人數超過1745萬 美國達463萬

據Worldometer統計，截至香港時間周五早上8時，全球確診人數突破1745萬人，增加27.7

瘟疫大流行最新發展

Test Ru...

BROWSE

main_template

Select page

Select article (19)

Extract title

Extract url

Relative story

Relative post_date

Select next

Click each next item

and go to main_template

Selection Node: Edit

2nd a
> All element s

Wait up to 60 seconds for elements to appear.

API Tutorials Contact

全球疫情發展 - 信報網站 h | 全球疫情發展 - 信報網站 h | Test Run | 全球疫情發展 - 信報網站 h

https://www1.hkej.com/features/topic/tag/瘟疫大流行最新發展

2020年7月30日

Select Mode

更多內容

全球確診人數突破1700萬
美國逾456萬

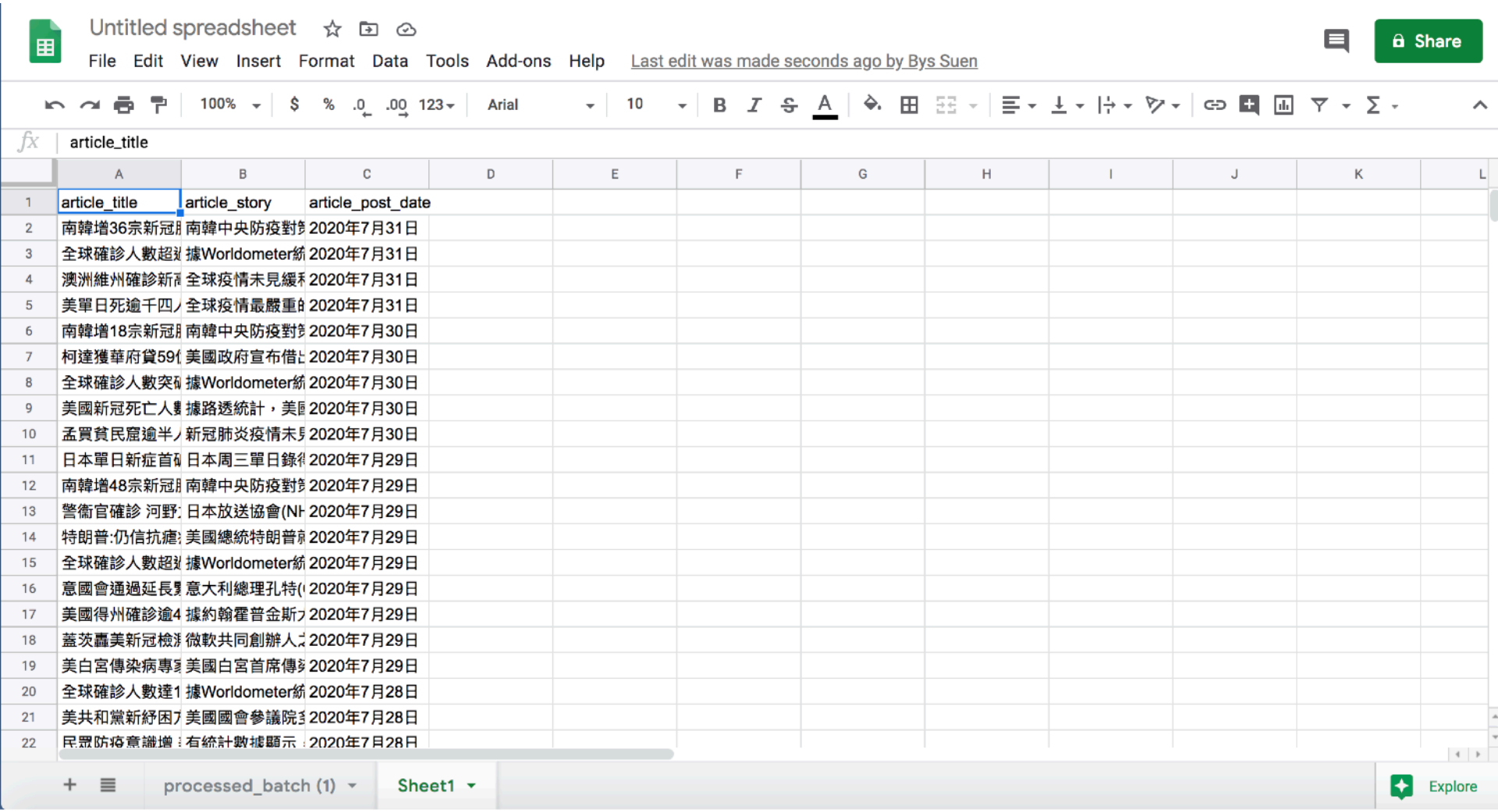
據Worldometer統計，截至香港時間周四早上

CSV/Excel JSON

article_title	article_url	article_story	article_story_url	article_post_date
南韓增36宗新冠肺炎確診 逾半屬境外輸入	https://www1.hkej.com /features/article?q= %23%E7%98%9F %E7%96%AB%E5 %A4%A7%E6%B5 %81%E8%A1%8C %E6%9C%80%E6 %96%B0%E7%99 %BC%E5%B1%95...	南韓中央防疫對策本部稱，當地新增36宗感染新型冠狀病毒肺炎的確診病例，累計近1.43萬宗。另外，當地新增1例死 ...	https://www1.hkej.com /features/article?q= %23%E7%98%9F %E7%96%AB%E5 %A4%A7%E6%B5 %81%E8%A1%8C %E6%9C%80%E6 %96%B0%E7%99 %BC%E5%B1%95...	2020年7月31日
全球確診人數超過1745萬 美國達463萬	https://www1.hkej.com /features/article?q= %23%E7%98%9F %E7%96%AB%E5	據Worldometer統計，截至香港時間周五早上8時，全球確診人數突破1745萬人，增加27.7萬人，死亡	https://www1.hkej.com /features/article?q= %23%E7%98%9F %E7%96%AB%E5	2020年7月31日

This a live preview. When you are ready to run your project, click Get Data.

Show more data Visuals enabled (advanced)


Untitled spreadsheet
File Edit View Insert Format Data Tools Add-ons Help
Last edit was made seconds ago by Bys Suen
Share
article_title
article_title article_story article_post_date
processed_batch (1)
Sheet1
Explore

Legal and Ethical Considerations.

scraping and how they play out.

Is Web Scraping Legal?

Your first thought might be to look at the legal side of things.

The truth is that the legality of web scraping is still relatively up in the air.

Meaning that there are currently no specific laws that refer to the legality of web scraping. So it is neither legal or illegal.

https://www.parsehub.com/blog/web-scraping-ethical/

Scraping Publicly Available Information

Another factor to keep in mind is the type of data you'd be scraping. In our case, we always refer to publicly available data.

This is data that has been made public by the owner of said data. Private and leaked information is not considered publicly available information.

Ethics in Web Scraping

James Densmore Jul 23, 2017 · 3 min read

We all scrape web data. Well, those of us who work with data do. Data scientists, marketers, data journalists, and the data curious alike. Lately, I've been thinking more about the ethics of the practice and have been dissatisfied by the lack of consensus on the topic.

Let me be clear that I'm talking **ethics** not the law. The law in regards to scraping web data is complex, fuzzy and ripe for reform, but that's another matter. It's not that no one is thinking, or writing, about the ethics in scraping but rather that both those scraping and those being scraped can't

https://towardsdatascience.com/ethics-in-web-scraping-b96b18136f01

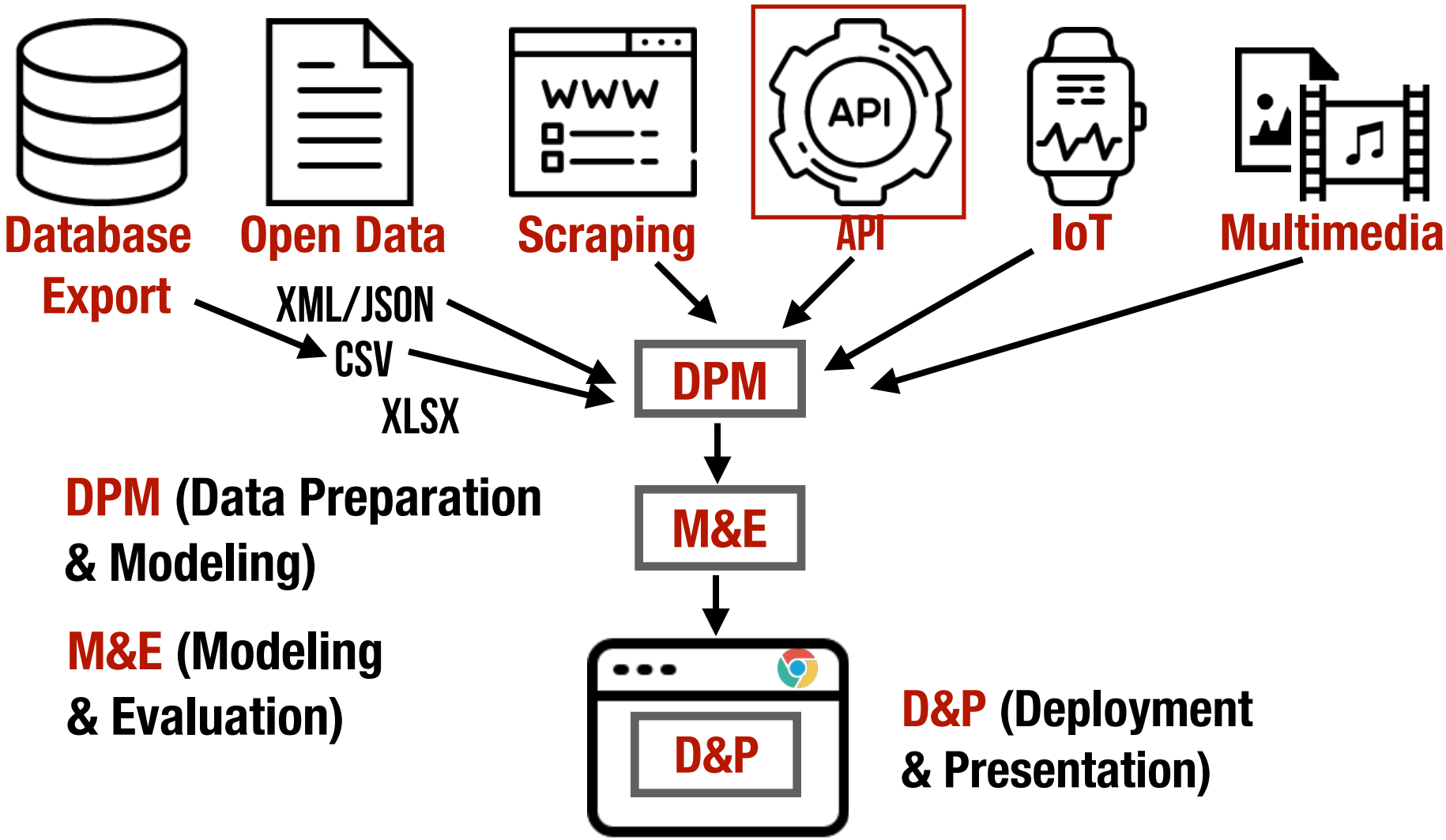
WRITTEN BY
James Densmore
Data Science and Data Engineering Consultant at Data Liftoff
https://www.dataliftoff.com

Follow

The Ethical Scraper

I, the web scraper will live by the following principles:

- If you have a public API that provides the data I'm looking for, I'll use it and avoid scraping all together.
- I will always provide a User Agent string that makes my intentions clear and provides a way for you to contact me with questions or concerns.
- I will request data at a reasonable rate. I will strive to never be confused for a DDoS attack.
- I will only save the data I absolutely need from your page. If all I need it OpenGraph meta-data, that's all I'll keep.
- I will respect any content I do keep. I'll never pass it off as my own.
- I will look for ways to return value to you. Maybe I can drive some (real) traffic to your site or credit you in an article or post.
- I will respond in a timely fashion to your outreach and work with you towards a resolution.
- I will scrape for the **purpose of creating new value from the data**, not to duplicate it.


Database Export
Open Data
Scraping
API
IoT
Multimedia
XML/JSON
CSV
XLSX
DPM
M&E
D&P
DPM (Data Preparation & Modeling)
M&E (Modeling & Evaluation)
D&P (Deployment & Presentation)

Harvard Business Review

GROWTH STRATEGY

The Strategic Value of APIs

by Bala Iyer and Mohan Subramaniam

JANUARY 07, 2015

McKinsey&Company
Digital McKinsey

How We Help Clients | Our Insights | Careers | Our People | Contact Us

API經濟來了：從開放創新到Open API

by 《數位時代》整合行銷部 2014.11.04

McKinsey&Company
Digital McKinsey

How We Help Clients | Our Insights | Careers | Our People | Contact Us

Deloitte. Insights

Realizing the economic value of APIs

Download the report →

Get insights via email →

Organization and governance

Successful API development requires cross-functional decision-making and oversight. Traditionally, IT departments have developed APIs with little or no input from other functions within the organization. Proactive API-centric organizations generally engage individuals from the relevant lines of business in the API development cycle to provide needed guidance around the right level of desired functionality and customer experience.

These organizations also involve their procurement and legal departments in monitoring intellectual property, developing SLAs and contracting. Marketing professionals play an important role in addressing the branding and promotion of APIs, and focusing on the attraction and recruitment of partners and external developers. Risk managers evaluate potential impacts of security breaches as well as unintended use of the APIs by legitimate parties.

Skills and capabilities

The API economy has implications for both an organization's skills and cultural mindset. An API-centric business requires new capabilities, including product management, data science and intellectual property management, as well as a culture that stresses innovation and resource reuse. For example, the product management skills needed to oversee the entire lifecycle of API management are often different from the technical capabilities needed to architect and code the actual APIs themselves.

One way organizations have set out to address these issues is the use of innovation teams centered on creating an API-centric transformation. These businesses create physical and virtual spaces that bring together architects, developers, product managers and business leaders to rapidly exchange ideas and interact with one another. They provide education on API development, develop use-case opportunities, and leverage internal and external collaborative events to promote API value and usage.

APIs drive digital transformation[4]

International Airlines Group (IAG), one of the world's largest airline groups, needed to transform itself into an agile, digital organization. The company was facing competitive investments in innovation and digital technologies, as well as fear of disruption from a potential "uberization" of the airline industry. IAG set up a digital business team to oversee a digital transformation and connectivity program, with APIs as the underlying driver.

This resulted in closer collaboration between the business and IT functions, connecting operations from the front- to the back-office via APIs and reducing much of the complexity caused by disjointed silos. The airline began a "Connectivity Program" to extend the digital mindset across the company, with a focus on timesaving techniques and tools to help airline crews do their jobs.

The Airbnb Story

API partner

Request access

Connect to our API.
Connect to millions of travelers on Airbnb.

Request access

Reach more guests

Reach millions of new guests by connecting with our global community.

Integrate and build

Set up your integration to give clients full control over how they want to list on Airbnb.

Access support

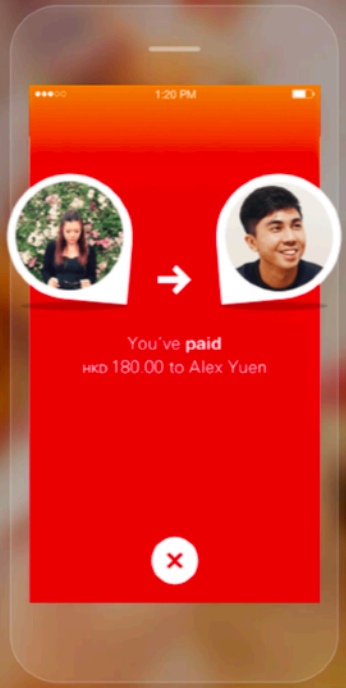
Explore in-depth technical documentation and get partner manager support.

AAABNB

PayMe | from HSBC

Personal | Business | What's new | Need help? | Download PayMe | English

Pay anyone with any bank



Pay your friends with just a few taps, regardless of which Hong Kong local bank they use and without ever asking for an account number.

How it works

Discover more

Google Cloud Platform

Prediction API

Natural Language API

Translation API

Vision API

PayMe Marketplace API
PayMe • paymeapi
Documentation Inspector

Download
API Blueprint

INTRODUCTION

API Capabilities and Structure

Getting Started and Connection Details

Service URLs

Credit Card Numbers for Testing Purposes

PayMe Marketplace API

INTRODUCTION

Our API is designed to allow platforms to offer a full payment solution as part of their product.

Using the API you can:

微信支付 WeChat Pay

Home

Documents

WeChat payment API development documents, Ch

Quick Pay

Buyers can present the pay code, vendor will scan the code to finish the transaction.

Alibaba Open Platform

Home Resources

Platform

Tmall.hk

API for Alibaba

Sourcing API： Click here

WholeSale API:Click here

Please visit dc.cuhkcfe.io
for more information.

Thank you for your time!