

Power Query (M): Web scraping using Web.BrowserContents and Html.Table



www.vistacompany.ir

flexyourdata.com | youtube.com/@flexyourdata | linkedin.com/in/owenhprice

We can easily extract data from a webpage Using Get data>Web

From Web

Basic Advanced

URL
wikipedia.org/wiki/List_of_countries_by_total_health_expenditure_per_capita

OK

Navigator

Display Options

- HTML Tables [3]
 - OECD countries only Total health spending per capita in PPP international dolla
 - Total health spending per capita in PPP international dollars Not inflation-adj
 - Table 1
- Suggested Tables [5]
 - Table 2
 - Table 3
 - Table 4
 - Table 5
 - Table 6
- Text [2]
 - HTML Code
 - Displayed Text

If the page includes `<table>` elements, they are listed under HTML tables

```
><table class="wikitable sortable static-row-numbers plainrowheaders srn-white-background jquery-tablesorter" border="1" style="text-align:right;">...</table> == $0
```

`table.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background.jquery...`

415.53 × 1198.56

sortable. This allows ran...

lth in LOCATION" links.

OECD countries only. Total health spending per capita in PPP international dollars. Not inflation-adjusted. Costa Rica joined OECD in 2021.^{[1][2][4][5]}

	Location	2018	2019	2020	2021
1	Australia *	5,194	5,130	5,627	
2	Austria *	5,519	5,624	5,883	6,693
3	Belgium *	5,315	5,353	5,407	
4	Canada *	5,308	5,190	5,828	5,905
5	Chile *	2,281	2,297	2,413	2,608
6	Colombia *	1,192	1,293	1,336	
7	Costa Rica *	1,530	1,556	1,618	
8	Czech Republic *	3,129	3,272	3,805	
9	Denmark *	5,307	5,360	5,694	6,384

- HTML Tables [3]
- OECD countries only Total health spending per capita in PPP international dolla
- Total health spending per capita in PPP international dollars Not inflation-adj
- Table 1

PQ will also try to detect other structured data

Here, the table-of-contents is an unordered (bulleted) list - ``. PQ has suggested this as **Table 2**

```
<ul>
  <li class="toclevel-1 tocsection-1">
    <a href="#Organization for Economic Co-operation and Development">...</a>
    <ul>
      <li class="toclevel-2 tocsection-2">
        <a href="#Table">...</a>
      </li>
      <li class="toclevel-2 tocsection-3">
        <a href="#Bar_charts">...</a>
      </li>
    </ul>
  </li>
  <li class="toclevel-1 tocsection-4">
    <a href="#World Health Organization">...</a>
    <ul>
      <li class="toclevel-2 tocsection-5">...</li>
    </ul>
  </li>
  <li class="toclevel-1 tocsection-6">
    <a href="#See_also">
      <span class="tocnumber">3</span>
      <span class="toctext">See also</span>
    </a>
  </li>
  <li class="toclevel-1 tocsection-7">
    <a href="#Notes_and_references">
      <span class="tocnumber">4</span>
      <span class="toctext">Notes and r...
    </a>
  </li>
</ul>
```

ul 359.3 × 154.17 Contents [hide]

- 1 Organization for Economic Co-operation and Development
 - 1.1 Table
 - 1.2 Bar charts
- 2 World Health Organization
 - 2.1 Table 2
- 3 See also
- 4 Notes and references

Display Options

- HTML Tables [3]
 - OECD countries only...
 - Total health spending...
 - Table 1
- Suggested Tables [5]
 - Table 2
 - Table 3

Table 2

Column1	Column2	Column3
Organization for Economic Co-operation and Development	edit	null
Table	edit	[1]
Bar charts	edit	[1]
World Health Organization	edit	[6]
Table 2	edit	[6]
See also	edit	null
Notes and references	edit	null

When we load a table, PQ generates some code

Web.BrowserContents downloads the HTML from the page

```
1 let
2   Source = Web.BrowserContents("https://en.wikipedia.org/wiki/List_of_countries_by_total_health_expenditure_per_capita"),
3   htmlTable
4     = Html.Table(
5       Source,
6       {
7         {"Column1", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(1)"},
8         {"Column2", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(2)"},
9         {"Column3", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(3)"},
10        {"Column4", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(4)"},
11        {"Column5", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(5)"}
12      },
13      [RowSelector="TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR"]
14    )
15 in
16   htmlTable
```

Html.Table extracts data from the downloaded HTML

If the page has a lot of content, we can choose to ask `Web.BrowserContents` to wait for it to load completely before downloading

The first parameter is the URL of the page

```
1 let
2   Source = Web.BrowserContents(
3     "https://en.wikipedia.org/wiki/List_of_countries_by_total_health_expenditure_per_capita",
4     [WaitFor = [Timeout = #duration(0,0,0,45)]]
5   )
6 in
7   Source
```

The second (optional) parameter can be a record with one field – `WaitFor`, which is itself a record with one or two fields – either `Timeout` (duration) or `Selector` (text) or both

There are two options for WaitFor

WaitFor is a record with one or two fields – either Timeout (duration) or Selector (text) or both

 Selector must be a valid CSS selector text.

See [CSS Selectors Reference \(w3schools.com\)](https://www.w3schools.com/css/css_selectors.asp) for details

```
1 let
2   Source = Web.BrowserContents(
3     "https://en.wikipedia.org/wiki/List_of_countries_by_total_health_expenditure_per_capita",
4     [WaitFor = [Timeout = #duration(0,0,0,45)]]
5   )
6 in
7   Source
```

Give the function 45 seconds to download the complete HTML of the page

```
1 let
2   Source = Web.BrowserContents(
3     "https://en.wikipedia.org/wiki/List_of_countries_by_total_health_expenditure_per_capita",
4     [WaitFor = [Selector = ".table"]]
5   )
6 in
7   Source
```

Wait for the element with class attribute "table" before finishing the download

Html.Table extracts information from valid HTML text

The first parameter of `Html.Table` is a valid HTML text, often (but not always) derived from `Web.BrowserContents`

```
3
4
5
6
7
8
9
10
11
12
13
14
htmlTable
= Html.Table(
  Source,
  {
    {"Column1", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(1)"},
    {"Column2", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(2)"},
    {"Column3", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(3)"},
    {"Column4", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(4)"},
    {"Column5", "TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR > :nth-child(5)"}
  },
  [RowSelector="TABLE.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR"]
)
```

The second parameter is a list of lists.

Each sub-list takes the form

```
{
  column name,
  CSS selector,
  (optional) function to process the element
  attributes
}
```

The 3rd param, which is optional on one-column tables, is a one-field record – `RowSelector`, which is a CSS Selector used to identify each row in the table

RowSelector tells Html.Table which elements to use

```
[RowSelector="TABLE .wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background:nth-child(13) > * > TR"]
```

Select all `<TR>` elements whose parent is any element - `*` - which has a parent `<table>` element with class attribute

wikitable sortable static-row-numbers plainrowheaders srn-white-background which is the 13th child of its parent

Parent

13th child of parent

Table rows `<tr>`

```
<div class="mw-parser-output">
  <p class="mw-empty-elt"> </p>
  <div class="thumb tright">...</div>
  <div class="thumb tright">...</div>
  <p>...</p>
  <p>...</p>
  <p>...</p>
  <p>...</p>
  <div id="toc" class="toc" role="navigation" aria-labelledby="mw-toc-heading">...</div>
  <h2>...</h2>
  <h3>...</h3>
  <dl>...</dl>
  <p>...</p>
  <table class="wikitable sortable static-row-numbers plainrowheaders srn-white-background jquery-tablesorter" border
    <caption>...</caption>
    <thead>...</thead>
    <tbody>
      <tr>
        <td style="text-align:left">
          <span class="flagicon" style="display:inline-block;width:25px;text-align:left;">...</span>
          &nbsp;
          <a href="/wiki/Healthcare in Australia" class="mw-redirect" title="Healthcare in Australia">Australia *</a>
        </td>
        <td>5,194 </td>
        <td>5,130 </td>
        <td>5,627 </td>
        <td> </td>
      </tr>
    </tbody>
  </table>
  ...</tr>
```

OK, take a breath...

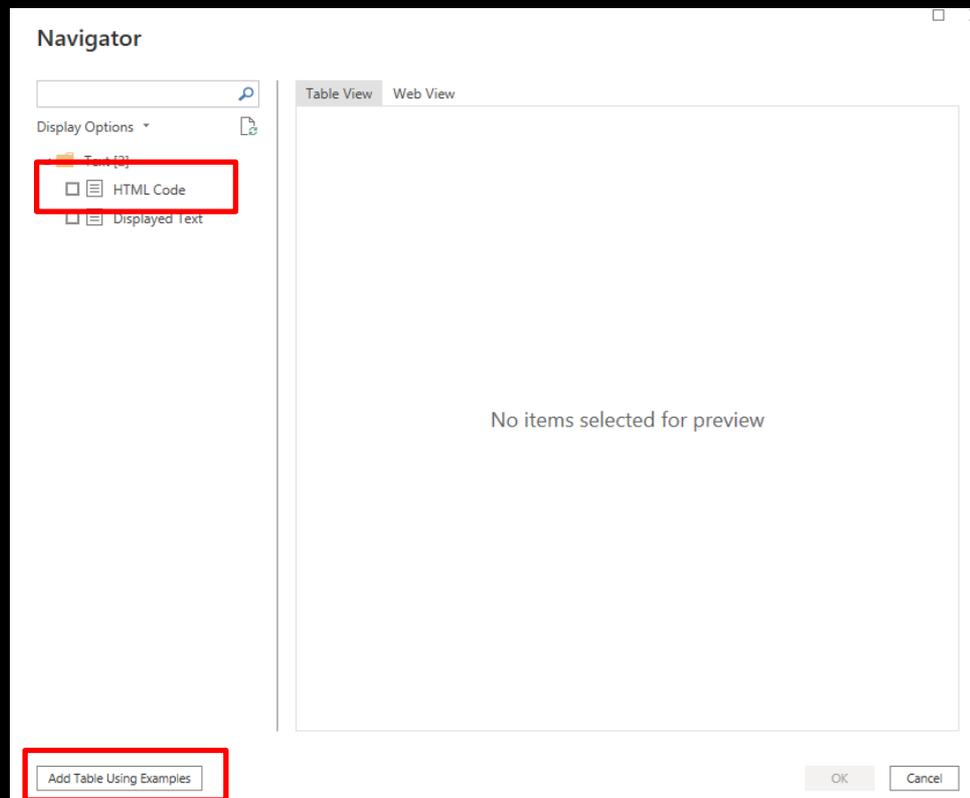
Most of the time, Power Query is smart enough to find the correct elements for you and you won't need to worry manually typing in these CSS Selectors

HUGE RELIEF.

But what if all you see in the Navigator pane is this?

Power Query is great, but sometimes the website you want to scrape isn't so easily parsed.

Sometimes no tables can be found.



We can either Add Table Using Examples or load the HTML code and build Html.Table by hand

1. 'Add Table Using Examples'

i) try to find the right elements

With 'Add Table Using Examples', we try typing in a column and Power Query will attempt to locate an element in the page that contains what we've typed.

Add Table Using Examples

Customer Reviews
★★★★☆ & Up
★★★★☆ & Up
★★★★☆ & Up
★★★★☆ & Up

TV Show
Doctor Who

Actor
Peter Davison
David Tennant
Janet Fielding
Sylvester McCoy
Sophie Aldred
Colin Baker
David Bradley

Director
Peter Moffatt
Richard Martin
Christopher Barry
Andrew Morgan
Pennant Roberts

RESULTS

Amazon's Choice



Sponsored
Doctor Who: The Complete Thirteenth Series - Flux
★★★★☆ ~ 897
DVD
\$18.59
Save \$3.72 with coupon
prime
FREE delivery Wed, Jan 4 on \$25 of items shipped by Amazon
Or fastest delivery Tue, Jan 3

- Starring: Various
- Directed by: Various



Sponsored
Doctor Who: The Complete David Tennant Collection (Blu-ray)
★★★★☆ ~ 4,667
Blu-ray
\$26.99 \$44.98
Save \$4.20 with coupon
FREE delivery

- Starring: Various
- Directed by: Various

Column1	
Flux	
2	⌘ "doctor who dvd" ...8.59 \$3.72 coupon applied at checkoutSave \$3.72 with coupon
3	🌐 /Doctor-Who-Complete-Thirteenth-Flux/dp/B09KP9NKNB/ref=sr_1_5?ke...IN543NilsInFzcCl6jQuNzkrfQ%3D&sprefix=%2Caps%2C110&sr=8-5
4	🌐 /Doctor-Who-Complete-Thirteenth-Flux/dp/B09KPBYF22/ref=ice_ac_b...rds=doctor+who+ dvd&qid=1672347922&sprefix=%2Caps%2C110&sr=8-5
5	🌐 /Doctor-Who-Complete-Thirteenth-Flux/dp/B09KPBYF22/ref=sr_1_5?keywords=doctor+who+ dvd&qid=1672347922&sprefix=%2Caps%2C110&sr=8-5
6	🌐 /Doctor-Who-Complete-Thirteenth-Flux/dp/B09KPBYF22/ref=sr_1_5?ke...vd&qid=1672347922&sprefix=%2Caps%2C110&sr=8-5#customerReviews
7	🌐 /Doctor-Who-Complete-Thirteenth-Flux/dp/B09KPBYF22/ref=sxin_15_a...efix=%2Caps%2C110&sr=1-1-7d9bfb42-6e38-4445-b604-42cab39e191b
8	🌐 /Doctor-Who-Complete-Thirteenth-Flux/dp/B09KPBYF22/ref=sxin_15_a...0&sr=1-1-7d9bfb42-6e38-4445-b604-42cab39e191b#customerReviews
9	🌐 /gp/slredirect/picassoRedirect.html/ref=pa_sp_atf_aps_sr_pg1_1?i...47922%26sprefix%3D%252Caps%252C110%26sr%3D8-1-spons%26psc%3D1
10	🌐 /gp/slredirect/picassoRedirect.html/ref=pa_sp_atf_aps_sr_pg1_1?i...3D%252Caps%252C110%26sr%3D8-1-spons%26psc%3D1#customerReviews
+	⌘ Add to ListAdd to List ShareShareReportReportYour feedback is i...st delivery Tue, Jan 3 Starring: Various Directed by: Various
⌘	⌘ Add to ListAdd to List ShareShareReportReportYour feedback is i...8.59 \$3.72 coupon applied at checkoutSave \$3.72 with coupon
⌘	⌘ Amazon's Choicefor doctor who dvd Doctor Who: The Complete Thir... \$25 shipped by Amazon.Starring: Various Directed by: Various
⌘	⌘ Doctor Who: The Complete Thirteenth Series - Flux

We select the item in the list that best matches what we're looking for

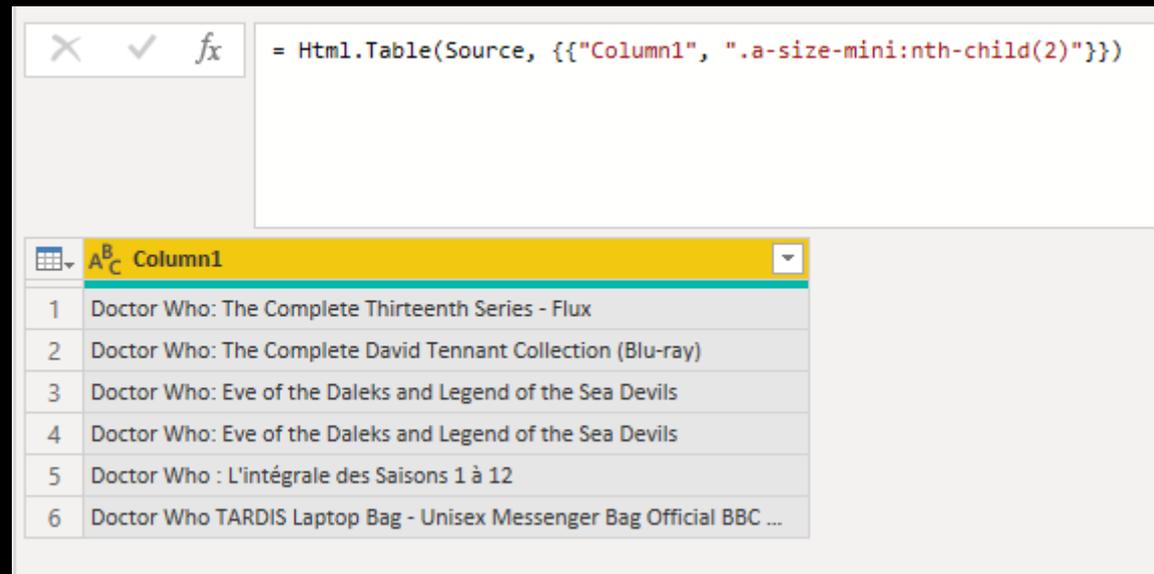
1. 'Add Table Using Examples'

ii) review the results

We can then review the selector and the results

In this case, "Add Tables Using Examples" has used the wrong selector.

The list of titles is too short.



The screenshot shows a web browser's developer console. At the top, there is a red error message: `= Html.Table(Source, {{"Column1", ".a-size-mini:nth-child(2)"}})`. Below the error, a table is displayed with the following data:

	Column1
1	Doctor Who: The Complete Thirteenth Series - Flux
2	Doctor Who: The Complete David Tennant Collection (Blu-ray)
3	Doctor Who: Eve of the Daleks and Legend of the Sea Devils
4	Doctor Who: Eve of the Daleks and Legend of the Sea Devils
5	Doctor Who : L'intégrale des Saisons 1 à 12
6	Doctor Who TARDIS Laptop Bag - Unisex Messenger Bag Official BBC ...

2. Build Html.Table manually

i) find the correct element using "developer tools"

If we use the browser's developer tools, we can review the elements on-screen and select the element whose class attribute we want

`span.a-size-medium.a-color-base.a-text-normal`

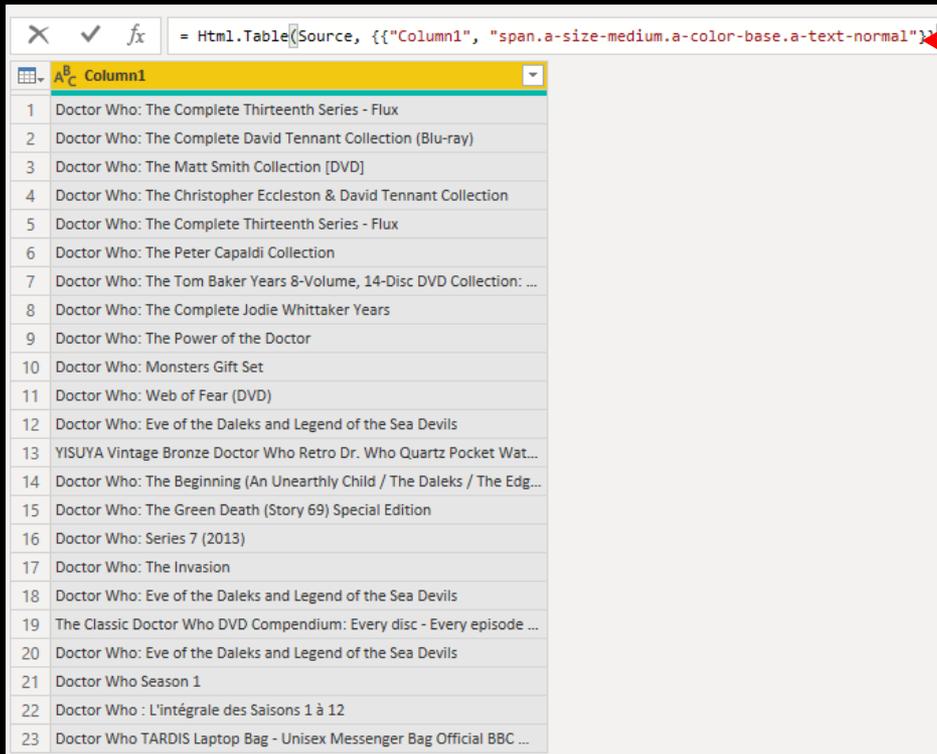
The screenshot shows the Amazon product page for 'Doctor Who: The Complete Thirteenth Series - Flux'. A browser developer tools overlay is visible, showing the selected element's classes: `span.a-size-medium.a-color-base.a-text-normal`. The product details include a 4.5-star rating from 897 reviews, a price of \$18.59 with a \$3.72 coupon, and Prime delivery by Saturday, December 31. The developer tools also show the element's dimensions (415.95 x 22) and accessibility information.

The screenshot shows the browser's developer tools with the HTML structure of the product page. A red arrow points from the text above to the specific `span` element in the HTML tree. The HTML structure includes various classes and attributes, such as `sg-col`, `a-section`, and `a-link-normal`. The selected element is `Doctor Who: The Complete Thirteenth Series - Flux`.

2. Build `Html.Table` manually

ii) use it as the CSS Selector for the table column

```
span.a-size-medium.a-color-base.a-text-normal
```



The screenshot shows a data table with a yellow header row. The header row has a column labeled "Column1". The table contains 23 rows of data, all of which are extracted from the "Column1" of the source table. A red arrow points from the CSS selector in the formula bar to the "Column1" header in the table.

	Column1
1	Doctor Who: The Complete Thirteenth Series - Flux
2	Doctor Who: The Complete David Tennant Collection (Blu-ray)
3	Doctor Who: The Matt Smith Collection [DVD]
4	Doctor Who: The Christopher Eccleston & David Tennant Collection
5	Doctor Who: The Complete Thirteenth Series - Flux
6	Doctor Who: The Peter Capaldi Collection
7	Doctor Who: The Tom Baker Years 8-Volume, 14-Disc DVD Collection: ...
8	Doctor Who: The Complete Jodie Whittaker Years
9	Doctor Who: The Power of the Doctor
10	Doctor Who: Monsters Gift Set
11	Doctor Who: Web of Fear (DVD)
12	Doctor Who: Eve of the Daleks and Legend of the Sea Devils
13	YISUYA Vintage Bronze Doctor Who Retro Dr. Who Quartz Pocket Wat...
14	Doctor Who: The Beginning (An Unearthly Child / The Daleks / The Edg...
15	Doctor Who: The Green Death (Story 69) Special Edition
16	Doctor Who: Series 7 (2013)
17	Doctor Who: The Invasion
18	Doctor Who: Eve of the Daleks and Legend of the Sea Devils
19	The Classic Doctor Who DVD Compendium: Every disc - Every episode ...
20	Doctor Who: Eve of the Daleks and Legend of the Sea Devils
21	Doctor Who Season 1
22	Doctor Who : L'intégrale des Saisons 1 à 12
23	Doctor Who TARDIS Laptop Bag - Unisex Messenger Bag Official BBC ...

By editing the CSS Selector for column1, We've extracted the title from each matching element

2. 'Build Html.Table manually

iii) add additional columns

Each page element has a tag and usually a class attribute. Understanding CSS Selectors is critical to successful web scraping

A ^B _C Title	A ^B _C Price whole	A ^B _C Price fraction
1	null	null
2	null	null
3	null	null
4 Doctor Who: The Complete Thirteenth Series - Flux	18.	59
5 Doctor Who: The Complete David Tennant Collection (Blu-ray)	26.	99
6 Doctor Who: The Matt Smith Collection [DVD]	19.	67
7 Doctor Who: The Christopher Eccleston & David Tennant Colle...	18.	99
8 Doctor Who: The Complete Thirteenth Series - Flux	18.	59
9 Doctor Who: The Peter Capaldi Collection	19.	70
10 Doctor Who: The Tom Baker Years 8-Volume, 14-Disc DVD Coll...	49.	99
11 Doctor Who: The Complete Jodie Whittaker Years	34.	99
12 Doctor Who: The Power of the Doctor	19.	98
13 Doctor Who: Monsters Gift Set	28.	64
14 Doctor Who: Web of Fear (DVD)	19.	99
15 Doctor Who: Eve of the Daleks and Legend of the Sea Devils	11.	96
16 Doctor Who: The Beginning (An Uneearthly Child / The Daleks / ...	19.	99
17 Doctor Who: The Green Death (Story 69) Special Edition	9.	87
18 Doctor Who: Series 7 (2013)	36.	88
19 Doctor Who: The Invasion	22.	98
20 Doctor Who: Eve of the Daleks and Legend of the Sea Devils	14.	99
21 The Classic Doctor Who DVD Compendium: Every disc - Every ...	20.	86
22 Doctor Who: Eve of the Daleks and Legend of the Sea Devils	11.	96
23 Doctor Who Season 1	1.	99
24 Doctor Who : L'intégrale des Saisons 1 à 12	276.	98
25 Doctor Who TARDIS Laptop Bag - Unisex Messenger Bag Offici...	69.	99

Advanced Editor

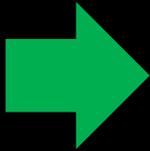
Table 11

```
let
  Source = Web.BrowserContents("https://www.amazon.com/s?k=doctor+who+dvd&prefi
  doctorWhoTitles
    = Html.Table(
      Source,
      {
        {"Title", "span.a-size-medium.a-color-base.a-text-normal"},
        {"Price whole", ".a-price-whole"},
        {"Price fraction", ".a-price-fraction"}
      },
      [RowSelector = "div.a-section.a-spacing-small.a-spacing-top-small"]
    )
in
  doctorWhoTitles
```

If we have more than one column, we must also have a RowSelector

Web scraping with Power Query - Takeaways

1. In Power Query, most web-scraping can be done with very little effort using Get Data>Web
2. In some cases, you may need to inspect the HTML using your browser's developer tools
3. A good understanding of CSS Selectors will help you be more successful with parsing HTML. [CSS Selectors Reference \(w3schools.com\)](https://www.w3schools.com/css/css-selector-reference.php) is a good resource for this



!! Be sure of the legality of scraping data from the websites you want to interact with

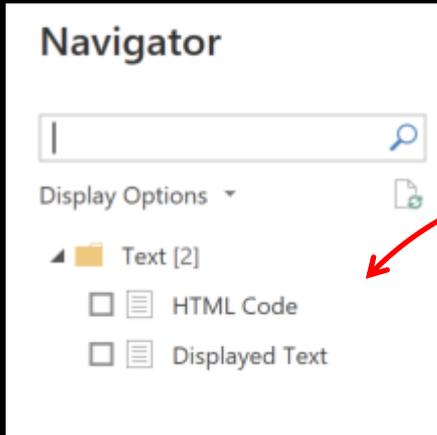
For an example
walkthrough,
continue reading →

Shopee is a popular online supermarket in several Asian countries

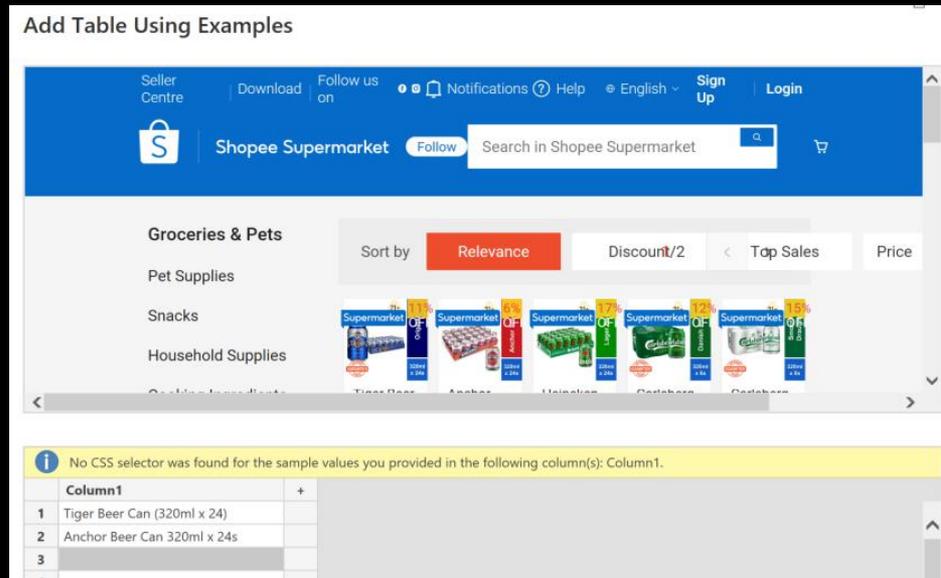
This search for alcoholic beverages has 25 results on the first page. When initially loading, 15 are visible in the window.

The screenshot displays the Shopee Supermarket interface. At the top, there's a navigation bar with links for Seller Centre, Download, and social media. The main header features the Shopee logo, 'Shopee Supermarket', and a search bar. Below the header, a sidebar on the left lists categories like Groceries & Pets, Pet Supplies, Snacks, Household Supplies, Cooking Ingredients, Coffee & Tea, and a 'SEARCH FILTER' section. The search filter includes 'Brand' (Somersby, Carlsberg, Tiger, 1664) and 'by category' (Beer & Cider (35), Others (4)). The main content area shows a grid of 25 product listings, each with an image, brand name, quantity, price, and discount tags. The products include Tiger Beer Can, Anchor Beer Can, Heineken Lager Beer, Carlsberg Danish Pilsner Beer, Carlsberg Smooth Draught Beer, Heineken Football Jersey, Tiger Beer Can (320ml x 6), 1664 Blanc Beer Can, Connor's Stout, Somersby Watermelon Can, Somersby Apple Cider Can, Somersby Blackberry Cider Can, Carlsberg Special Brew Strong Lager Beer Can, and Connor's Stout Forge.

I tried to get the brand details from the page



The Navigator showed me there were no HTML tables and no Suggested tables



“Add Table Using Examples” couldn’t find a CSS Selector for the brand



I used the developer tools in my browser to inspect the HTML of the page

```
560 <div class="s3NT03">
561   <div class="col-xs-2-4 sjMKCq" data-sqe="item"><a data-sqe="link"
562     href="/Tiger-Beer-Can-(320ml-x-24)-i.52784309.5574327192">
563     <div class="VJclHx">
564       <div class="_00bcqe">
565 > <div style="pointer-events: none;">...
593 </div>
594 <div class="_6LWLGE _3Qqe0p">
595   <div class="dRp1b8">
596     <div class="LsVFHr yL4BVF HwcBH9">Tiger Beer Can (320ml x 24)
597     </div>
598   </div>
```



I found the text “Tiger Beer Can (320ml x 24)” in the div element with class attribute

LsVFHr yL4BVF HwcBH9

The CSS Selector

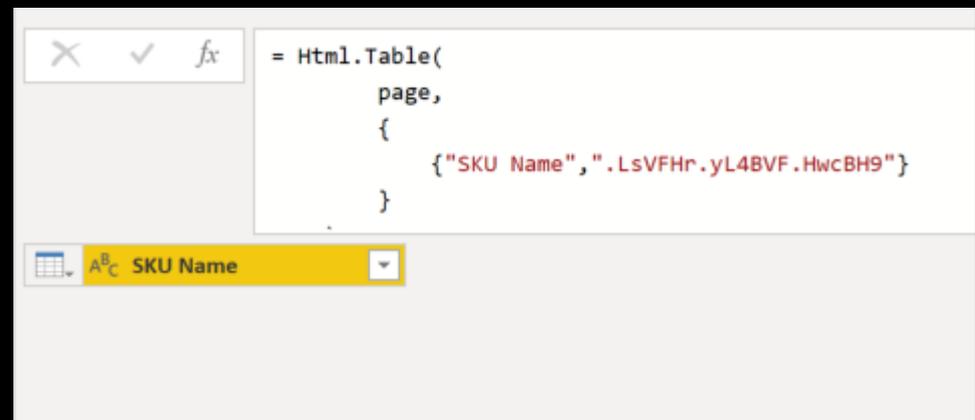
“.LsVFHr.yL4BVF.HwcBH9”

will find all such elements on the page

I used the Selector to try to extract just the brand name from the page

```
1 let
2   page = Web.BrowserContents("https://shopee.com.my/supermarket/Alcoholic-Beverages-cat.11000003.11000114"),
3   htmlTable = Html.Table(
4     page,
5     {
6       {"SKU Name", ".LsVFHr.yL4BVF.HwcBH9"}
7     }
8   )
9 in
10  htmlTable
```

The first attempt returned no data!



I added a Timeout to Web.BrowserContents to give the page more time to fill the elements

```
1 let
2   page = Web.BrowserContents(
3     "https://shopee.com.my/supermarket/Alcoholic-Beverages-cat.11000003.11000114",
4     [WaitFor = [Timeout = #duration(0,0,0,30)]]
5   ),
6   htmlTable = Html.Table(
7     page,
8     {
9       {"SKU Name", ".LsVFHr.yL4BVF.HwcBH9"}
10    }
11  )
12 in
13   htmlTable
```



	SKU Name
1	Tiger Beer Can (320ml x 24)
2	Anchor Beer Can 320ml x 24s
3	Heineken Lager Beer 320ml x 24s
4	Carlsberg Danish Pilsner Beer Can (320ml x 6 P...
5	Carlsberg Smooth Draught Beer Can (320ml x 6)
6	Heineken Football Jersey - Random [Gift Rede...
7	Tiger Beer Can (320ml x 6)
8	1664 Blanc Beer Can 320ml x 24s
9	Connor's 320ml x 24s
10	Somersby Watermelon Can (320ml x 4's)
11	Somersby Apple Cider Can 320ml x 4s
12	Somersby Apple Cider Can (320ml x 24 Cans)
13	Somersby Blackberry Cider Can 320ml x 4s
14	Carlsberg Special Brew Strong Lager Beer Can ...
15	Connor's 320ml x 4s

Even though there were 25 brands on the page, the query only returned 15

I found that the page had a script to fill the first 15 brands when the page loads. The rest are filled when the window is scrolled!

```
653 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
739 </a></div>
740 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
826 </a></div>
827 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
918 </a></div>
919 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1010 </a></div>
1011 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1090 </a></div>
1091 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1177 </a></div>
1178 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1269 </a></div>
1270 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1360 </a></div>
1361 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1434 </a></div>
1435 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1508 </a></div>
1509 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1600 </a></div>
1601 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1674 </a></div>
1675 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1766 </a></div>
1767 > <div class="col-xs-2-4 sJMKCq" data-sqe="item"><a data-sqe="link"...
1839 </a></div>
1840 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
1864 </div>
1865 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
1889 </div>
1890 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
1914 </div>
1915 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
1939 </div>
1940 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
1964 </div>
1965 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
1989 </div>
1990 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
2014 </div>
2015 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
2039 </div>
2040 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
2064 </div>
2065 > <div class="col-xs-2-4 sJMKCq" data-sqe="item">...
2000 </div>
```

The first 15 brands were populated with data

The remaining 10 were populated with placeholders and didn't have all the child elements

I changed `Html.Table` to extract the JSON from those elements

```
1 let
2   Source = Web.BrowserContents(
3     "https://shopee.com.my/supermarket/Alcoholic-Beverages-cat.11000003.11000114?",
4     [WaitFor=[Timeout=#duration(0, 0, 0, 30)]]
5   ),
6
7   /*
8   The data for the elements are in a JSON document in a script element.
9   the script element has a [TextContent] field which represents the element's value.
10  This value is the JSON for all products on the page, plus a few extra rows we don't need.
11  so if that JSON contains the word Product, then keep it, otherwise put null
12  */
13  GetProductJSON = Html.Table(
14    Source,
15    [{"Column1", "script[type=""application/ld+json""]",
16     each if Text.Contains([TextContent], "Product") then [TextContent] else null}]
17  ),
18  --
```

This CSS selector means
*select the elements with a
script tag and an attribute
called type whose value is
"application/ld+json"*

If the `[TextContent]` attribute
of the returned element
contains the word `Product`,
then put the whole JSON,
otherwise put null

	Column1
1	null
2	null
3	null
4	{"@context":"http://schema.org", "@type":"Product", "name":"Tiger Beer Can (320ml x 24)", "description":"","url":"htt...
5	{"@context":"http://schema.org", "@type":"Product", "name":"1664 Blanc Beer Can 320ml x 4s", "description":"","url": "...
6	{"@context":"http://schema.org", "@type":"Product", "name":"Anchor Beer Can 320ml x 24s", "description":"","url": "...
7	{"@context":"http://schema.org", "@type":"Product", "name":"Heineken Lager Beer 320ml x 24s", "description":"","url": "...
8	{"@context":"http://schema.org", "@type":"Product", "name":"Carlsberg Danish Pilsner Beer Can (320ml x 6 Pcs)", "de...
9	{"@context":"http://schema.org", "@type":"Product", "name":"Carlsberg Smooth Draught Beer Can (320ml x 6)", "desc...
10	{"@context":"http://schema.org", "@type":"Product", "name":"Heineken Football Jersey - Random [Gift Redemption]"...
11	{"@context":"http://schema.org", "@type":"Product", "name":"Tiger Beer Can (320ml x 6)", "description":"","url":"http...
12	{"@context":"http://schema.org", "@type":"Product", "name":"1664 Blanc Beer Can 320ml x 24s", "description":"","url": "...
13	{"@context":"http://schema.org", "@type":"Product", "name":"Connor's 320ml x 24s", "description":"","url":"https://s...

Then it was just a matter of extracting the data using UI-equivalent commands

```
19 RemoveNullRows = Table.SelectRows(GetProductJSON,each [Column1]<>null),
20
21 ConvertJSONDocumentToRecord = Table.TransformColumns(RemoveNullRows,{},Json.Document),
22
23 AsTable = Table.FromRecords(ConvertJSONDocumentToRecord[Column1]?),
24
25 RemoveUnwantedColumns = Table.RemoveColumns(AsTable,{"@context", "@type", "description", "brand"}),
26
27 ExpandRatingInfo = Table.ExpandRecordColumn(
28     RemoveUnwantedColumns,
29     "aggregateRating",
30     {"bestRating", "worstRating", "ratingCount", "ratingValue"},
31     {"aggregateRating.bestRating", "aggregateRating.worstRating", "aggregateRating.ratingCount", "aggregateRating.ratingValue"}
32 ),
33
34 ExpandOffers = Table.ExpandRecordColumn(
35     ExpandRatingInfo,
36     "offers",
37     {"price", "priceCurrency"},
38     {"offers.price", "offers.priceCurrency"}
39 )
40 in
41 ExpandOffers
```

The result was all 25 brands from the first page of search results

```
= Table.ExpandRecordColumn(ExpandRatingInfo, "offers", {"price", "priceCurrency"}, {"offers.price", "offers.priceCurrency"})
```

	name	url	productID	image
1	Tiger Beer Can (320ml x 24)	https://shopee.com.my/Tiger-Beer-Can-(320ml-x-24)-i.52784309.5574...	5574327192	https://cf.sh
2	1664 Blanc Beer Can 320ml x 4s	https://shopee.com.my/1664-Blanc-Beer-Can-320ml-x-4s-i.52784309....	3590229276	https://cf.sh
3	Anchor Beer Can 320ml x 24s	https://shopee.com.my/Anchor-Beer-Can-320ml-x-24s-i.52784309.40...	4074317213	https://cf.sh
4	Heineken Lager Beer 320ml x 24s	https://shopee.com.my/Heineken-Lager-Beer-320ml-x-24s-i.52784309...	4090136933	https://cf.sh
5	Carlsberg Danish Pilsner Beer Can (320ml x 6 Pcs)	https://shopee.com.my/Carlsberg-Danish-Pilsner-Beer-Can-(320ml-x-6...	8159733462	https://cf.sh
6	Carlsberg Smooth Draught Beer Can (320ml x 6)	https://shopee.com.my/Carlsberg-Smooth-Draught-Beer-Can-(320ml-...	10106941593	https://cf.sh
7	Heineken Football Jersey - Random [Gift Redemption]	https://shopee.com.my/Heineken-Football-Jersey-Random-Gift-Rede...	9859679100	https://cf.sh
8	Tiger Beer Can (320ml x 6)	https://shopee.com.my/Tiger-Beer-Can-(320ml-x-6)-i.52784309.74743...	7474323865	https://cf.sh
9	1664 Blanc Beer Can 320ml x 24s	https://shopee.com.my/1664-Blanc-Beer-Can-320ml-x-24s-i.52784309...	3192452867	https://cf.sh
10	Connor's 320ml x 24s	https://shopee.com.my/Connor's-320ml-x-24s-i.52784309.5390150484	5390150484	https://cf.sh
11	Somersby Watermelon Can (320ml x 4's)	https://shopee.com.my/Somersby-Watermelon-Can-(320ml-x-4's)-i.52...	6190148984	https://cf.sh
12	Somersby Apple Cider Can 320ml x 4s	https://shopee.com.my/Somersby-Apple-Cider-Can-320ml-x-4s-i.5278...	11707267207	https://cf.sh
13	Somersby Apple Cider Can (320ml x 24 Cans)	https://shopee.com.my/Somersby-Apple-Cider-Can-(320ml-x-24-Cans)...	11707267129	https://cf.sh
14	Somersby Blackberry Cider Can 320ml x 4s	https://shopee.com.my/Somersby-Blackberry-Cider-Can-320ml-x-4s-i....	11607267223	https://cf.sh
15	Carlsberg Special Brew Strong Lager Beer Can (320ml x 4 Pcs)	https://shopee.com.my/Carlsberg-Special-Brew-Strong-Lager-Beer-Ca...	10306941624	https://cf.sh
16	Connor's 320ml x 4s	https://shopee.com.my/Connor's-320ml-x-4s-i.52784309.4790150476	4790150476	https://cf.sh
17	Skol Beer Can 320ml x 4s	https://shopee.com.my/Skol-Beer-Can-320ml-x-4s-i.52784309.119072...	11907267173	https://cf.sh
18	1664 Rosé Beer Can 320ml x 4s	https://shopee.com.my/1664-Rosé-Beer-Can-320ml-x-4s-i.52784309....	13141600622	https://cf.sh
19	Asahi Beer Can 320ml x 24s	https://shopee.com.my/Asahi-Beer-Can-320ml-x-24s-i.52784309.6390...	6390148978	https://cf.sh
20	Skol Beer Can 320ml x 24s	https://shopee.com.my/Skol-Beer-Can-320ml-x-24s-i.52784309.11907...	11907267055	https://cf.sh
21	Heineken Lager Beer 320ml x 4s	https://shopee.com.my/Heineken-Lager-Beer-320ml-x-4s-i.52784309....	3290165103	https://cf.sh
22	Asahi Beer Can 320ml x 4s	https://shopee.com.my/Asahi-Beer-Can-320ml-x-4s-i.52784309.86601...	8660110010	https://cf.sh
23	Carlsberg Special Brew Strong Lager Beer Can 320ml x 24s	https://shopee.com.my/Carlsberg-Special-Brew-Strong-Lager-Beer-Ca...	8259733468	https://cf.sh
	Somersby Blackberry Cider Can (320ml x 24 Cans)	https://shopee.com.my/Somersby-Blackberry-Cider-Can-(320ml-x-24-...	9160260817	https://cf.sh
	Somersby Watermelon Can (320ml x 24 Cans)	https://shopee.com.my/Somersby-Watermelon-Can-(320ml-x-24-Can...	5990148999	https://cf.sh

