Finding & Getting Data - Google Sheet's import HTML

In this video I'm gonna show you a neat trick in Google Sheets to import tables from other pages into the spreadsheet automatically. It's called import HTML.

Let me give you an example. Take a look at the football tables at The Guardian page. We have multiple tables that are representing different championships, football championships around the world. If we wanted to import all of these tables to Google Sheets we might want to do some, maybe select all of the text and then do some copy and pasting. Or, we might want to do automatically the import of a specific table or multiple tables from this page directly to a Google Sheet and we can do just that with importHTML.

Let me show you how it works. So you go to any empty cell and then you type "=importHTML" and then "(". See that "importHTML(" takes three parameters. The first one is the URL of the web page where your table is. The second one is a query, what they call a query. It could be either a table or a list of items. In our case it's going to be a table and it has to be an HTML table. And the last parameter is the position of the table on the web page. So let's go back to the football tables page at The Guardian.

So we have several tables here. And if we wanted to import for example the second table here that represents "La Liga", the Spanish Championship we might want to do this. So we copy the URL that's in the address bar. We can right click and copy or do, control+C, or command+C and then we come back to the Google Sheet and then we need to do quotes. So every time you type in the URL parameter and importHTML, it has to be between quotes and then we do command+C to paste the address and then another quotes and then we use the separator. Notice that the separator is indicated on the two tip that Google Sheets presents to you. Depending on the language settings of your Google account in Google Sheets, this separator could be a different character so if your importHTML malfunction is not working it might be that you're using a different separator as I am. So take a look at that.

So I'm gonna put "," here, and then I'm gonna put the word "table" between quotes because that's what we're importing. And then lastly the position of the table on the web page. Now going back to the example we see that the "La Liga" table is the second table that I see on the page. So, let's type, go back and type the number "2". So it's important that the number is actually not between quotes. The URL parameter is, and the query parameter is, but the position of the table on the web page isn't. So after that, we then close parenthesis and then we hit enter and voila, we have the table that represents the results of the second table on the page at The Guardian.

So what if we wanted to load the the third table? We might just change the number here at the top to number "3", and then it loads the table automatically. Or we might go back to number "1" and then we can just copy the first cell here. So I'm copying it, right click, "Copy" and then I go down here on Row 7 and then I paste and then I change this to number "2". Now importHTML is importing actually two tables the first one and the second one. Notice that when I click on Row 1 A1, the number here is "1" which indicates that it is the first table on the web page. And then when I go to Row 7 it says table number 2. So that's one example. Let me remove this and give you another.

This is the list of highest grossing Indian films. It's a Wikipedia article. See that there are multiple pages on this web page as well. And the web page makes us think that the first table that is showing the list of the highest grossing movies Indian movies is actually the first table here, right?, on the web. So let's, let's take a look at that. We copy the address here at the top "Copy". Then I go back here, empty cell 1, "=importHTML(". Then I quote the address, the URL needs to be between quotes, and then the second parameter, also between quotes, is a "table" and then "," again. And then, because this looks like the first table we're gonna use the number "1" here as well. And then we press enter, see what happens.

It actually didn't load what it looks like to be the first table on the page. That's why H, that's that's because importHTML doesn't actually import the first table that we see. ImportHTML imports the HTML code of the page, and then it loads the first page, the first table that appears in the code. So we need to see actually if it doesn't work, in our case it didn't work. We might want to change a number here really quickly just to test, try and test to see if we can load the table that we're actually seeing. So if we change this to number "2", for the next table in the code then you'll see that it actually loads the table that we're seeing here. So there is a table here on the HTML code that comes before the table that we're seeing. There isn't a structure here that it was configured to be an HTML table but it's not appearing to be an HTML table for us. So you just, you just go there change the number from "1" to "2" and then you load the right table. Let me give you yet another example so that we see how fast importHTML works. Good.

Those are all tables that are you know, small tables like 25 rows, 10 rows. What if we wanted to load a table that has you know, hundreds of rows. Right. Let me let, me show you an example. This is a web page that shows airport, airport codes in Europe. And see that there are hundreds of airports here, all airports in Europe. And this is actually a table. So what if we wanted to import this table automatically to importHTML? So we just copy here the address "Copy", then you come back here, "importHTML" and then we put the URL between quotes. Comma, "table", comma again, table between quotes. And then we put the position of the table on the HTML code and then we hit enter. See how fast it loads, we have actually more than over 200 airports in Europe and it loads very fast.

And the good thing about importHTML is that if I leave this table here as is, with the importHTML function loaded, it'll actually update automatically the table here on the Google Sheet if any changes happen on the website. So, if there's any update here or any update here, or any update here my Google Sheet will be updated automatically as well because importHTML access the website from time to time automatically to load the latest results. Now, some considerations are important. ImportHTML will only import HTML tables, so you need to check out the code to make sure that what you're importing is actually a table. Right. So if we, for example, right click here and do some inspecting to see the HTML code, we will see that there is actually a table tag here on the HTML code that describes this table. Sometimes there are elements on a web page that looks like an HTML table but they are not a table. They might be a Div, they might use a different tag to represent what we see as a table but it's not actually an HTML table. So that's a caveat of importHTML.

The other thing that we need to consider is that the importHTML will only return tables that are visible to Google Sheets. Right. So because Google Sheets will try to access the website, the answer that the website gives to Google Sheets needs to be the HTML code that you see here on your browser. Sometimes some websites will prevent automated

access so Google Sheets is sort of an automated access to the website. So some Web sites will block or prevent the full HTML code to be loaded. So, you might want to pay attention to that too if you get mixed results. So there it is, this is importHTML and this is how we import tables automatically to Google Sheets.