Búsqueda y obtención de datos – Comando importHTML de Google Sheets

En este video, les mostraré un truco genial en Google Sheets para importar tablas de otras páginas a la hoja de cálculo automáticamente. Es el comando importHTML. Déjenme darles un ejemplo.

Echen un vistazo a las tablas de fútbol en la página del periódico The Guardian. Tenemos varias tablas que representan diferentes campeonatos, campeonatos de fútbol en todo el mundo. Si quisiéramos importar todas estas tablas a Google Sheets quizás podríamos seleccionar todo el texto y luego copiarlo y pegarlo. O bien, es posible que queramos importar automáticamente una tabla específica o varias tablas de esta página directamente a una Google Sheet, y podemos hacerlo con el comando importHTML. Déjenme mostrarles cómo funciona.

Ustedes van a cualquier celda vacía y escriben =importHTML y luego un paréntesis "(". Vean que =importHTML(tiene tres parámetros. El primero es la URL de la página web donde está su tabla. El segundo es una búsqueda, que ellos llaman una búsqueda, podría ser un tabla o una lista de elementos. En nuestro caso, será una tabla y debe ser una tabla HTML. Y el último parámetro es la posición de la tabla en la página web.

Así que volvamos a la página de tablas de fútbol de The Guardian. Tenemos varias tablas aquí. Y si quisiéramos importar, por ejemplo, la segunda tabla aquí, que representa "La Liga", el campeonato de España, podríamos querer hacer esto: copiamos la URL que está en la barra de direcciones. Podemos hacer clic derecho y copiar o hacer ctrl + C, o command + C, luego volvemos a la hoja de cálculo de Google y luego necesitamos usar comillas. Por lo tanto, cada vez que ingrese el parámetro URL en importHTML, debe estar entre comillas y luego hacemos command + V para pegar la dirección y luego otras comillas y luego usamos el separador.

Noten que el separador se indica en las dos muestras que Google Sheets les presenta. Dependiendo de la configuración de idioma de su cuenta de Google, en Google Sheets este separador podría ser un carácter diferente, por lo que si el comando importHTML no está funcionando, es posible que estén utilizando un separador diferente como yo. Así que échenle un vistazo a eso.

Entonces pondré una coma aquí, y luego pondré la palabra "table" entre comillas, porque eso es lo que estamos importando. Y luego, finalmente, la posición de la tabla en la página web. Ahora, volviendo al ejemplo, vemos que la tabla de "La Liga" es la segunda tabla que veo en la página. Así que escribamos, regresemos y escribamos el número 2.

Es importante que el número NO esté entre comillas. El parámetro URL sí lo está, y el parámetro de búsqueda también, pero la posición de la tabla en la página web no está entre comillas. Entonces, después de eso, cerramos el paréntesis y presionamos Enter y *voilá*, tenemos la tabla que representa los resultados de la segunda tabla en la página The Guardian.

¿Y si quisiéramos cargar la tercera tabla? Solo necesitamos cambiar el número aquí en la parte superior al número 3 y este carga la tabla automáticamente. O podemos volver al número 1 y simplemente podemos copiar la primera celda aquí. La estoy copiando aquí, haciendo clic derecho en "copiar" y luego bajo aquí en la fila 7 y luego pego y cambio esto por el número 2. Ahora importHTML está importando dos tablas, la primera y la segunda. Noten que cuando hago clic en la Fila 1 A1, el número aquí es 1, lo que indica que es la primera tabla en la página web. Y luego cuando voy a la Fila 7 dice tabla número 2. Ese es un ejemplo. Permítanme eliminar esto y darles otro.

Esta es la lista de las películas indias más taquilleras. Es un artículo de Wikipedia. Vean que hay varias tablas en esta página web. Y la página web nos hace pensar que la primera tabla que muestra la lista de las películas indias más taquilleras es realmente la primera tabla aquí, ¿correcto?, en la página web. Así que echemos un vistazo a esto. Copiamos la dirección aquí en la parte superior, "Copiar". Luego vuelvo aquí, a la celda vacía 1, escribo =importHTML(, luego pongo comillas, porque la dirección, la URL necesita estar entre comillas, y luego el segundo parámetro, también entre comillas "table" y luego coma de nuevo [","]. Luego, porque esta parece la primera tabla, usemos el número 1 aquí también, y luego presione Enter, y veamos qué sucede.

En realidad, no cargó lo que parece ser la primera tabla de la página. Eso es porque importHTML de hecho no importa la primera tabla que vemos. ImportHTML importa el código HTML de la página y carga la primera tabla que aparece en el código. Por lo tanto, tenemos que ver, si no funciona, y en nuestro caso no funcionó, es posible que queramos cambiar el número aquí rápidamente, solo para probar, experimentar y probar para ver si podemos cargar la tabla que de hecho estamos viendo. Entonces, si cambiamos esto al número 2, para la siguiente tabla en el código, verán que carga la tabla que estamos viendo aquí. Entonces, hay una tabla aquí en el código HTML que viene antes de la tabla que estamos viendo. Aquí hay una estructura que se ha configurado para ser una tabla HTML, pero no nos parece una tabla HTML. Así que simplemente vayan allí y cambien el número de 1 a 2 y luego carguen la tabla correcta.

Déjenme darles otro ejemplo para que podamos ver qué tan rápido funciona importHTML. Bueno, estas tablas son todas pequeñas, 25 filas, 10 filas. ¿Qué pasaría si quisiéramos cargar una tabla que tiene cientos de filas? Bien, déjenme mostrarles un ejemplo.

Esta es una página web que muestra códigos de aeropuertos en Europa. Y vean que hay cientos de aeropuertos aquí, todos los aeropuertos en Europa. Y esto es de hecho una tabla. ¿Qué pasaría si quisiéramos importar esta tabla automáticamente a importHTML? Entonces simplemente copiamos la dirección aquí, "Copiar", luego regresamos aquí, =importHTML(luego incluimos la URL entre comillas. Coma, "table", coma otra vez, table entre comillas. Y luego colocamos la posición de la tabla en el código HTML y presionamos Enter. Vean qué tan rápido se carga, tenemos más de 200 aeropuertos en Europa y se carga muy rápido.

Y lo bueno de importHTML es que si dejo esta tabla aquí como está, con la función importHTML habilitada, actualizará automáticamente la tabla aquí en la hoja de cálculo de Google si realizan algún cambio en el sitio. Entonces, si hay alguna actualización aquí, aquí o aquí, mi Google Sheet también se actualizará automáticamente porque importHTML accede al sitio de vez en cuando automáticamente para cargar los últimos resultados.

Ahora, algunas consideraciones son importantes. ImportHTML solo importará tablas HTML, por lo que deben verificar el código para asegurarse de que lo que están importando es de hecho una tabla, ¿correcto? Entonces, si, por ejemplo, hacemos clic derecho aquí y hacemos una inspección para ver el código HTML, veremos que en realidad hay una etiqueta de "table" aquí en el código HTML que describe esto como una tabla. A veces hay elementos en una página que parecen una tabla HTML, pero no son una tabla. Pueden ser un Div, pueden usar una etiqueta diferente para representar lo que vemos como una tabla, pero en realidad no es una tabla HTML. Entonces esta es una advertencia sobre importHTML.

Otra cosa que debemos considerar es que importHTML solo devolverá tablas que sean visibles para Google Sheets, ¿verdad? Entonces, como Google Sheets intentará acceder al sitio, la respuesta que el sitio le da a Google Sheets debe ser el código HTML que ven aquí en su navegador. A veces, algunos sitios evitarán el acceso automatizado, y Google Sheets es un tipo de acceso automatizado al sitio. Entonces algunos sitios bloquearán o evitarán que se cargue el código HTML completo. Por lo tanto, también deben prestar atención a esto si obtienen resultados incompletos.

Así que eso es todo, esto es importHTML y así es como automáticamente importamos tablas en Google Sheets.