

Localizando e obtendo os dados - Comando importHTML no Google Sheets

Neste vídeo, eu vou mostrar um truque legal no Google Sheets para importar tabelas de outras páginas para a planilha automaticamente. É o comando importHTML. Deixe-me dar um exemplo.

Dê uma olhada nas tabelas de futebol na página do jornal The Guardian. Temos várias tabelas que estão representando diferentes campeonatos de futebol ao redor do mundo. Se quiséssemos importar todas estas tabelas para o Google Sheets, podemos talvez selecionar todo o texto e, em seguida, copiar e colar. Ou, podemos querer fazer automaticamente a importação de uma tabela específica ou de várias tabelas desta página diretamente para uma planilha do Google e podemos fazer exatamente isso com importHTML. Deixe-me mostrar como funciona.

Você vai a qualquer célula vazia e digita =importHTML e depois (. Veja que importHTML usa três parâmetros. O primeiro é a URL da página web onde está sua tabela. A segunda é uma busca, o que chamam de busca, poderia ser de uma tabela ou de uma lista de itens. No nosso caso, vai ser uma tabela e ela tem que ser uma tabela HTML. E o último parâmetro é a posição da tabela na página web.

Então, vamos voltar para a página das tabelas de futebol no Guardian. Temos várias tabelas aqui. E se quiséssemos importar, por exemplo, a segunda tabela aqui, que representa "La Liga", o Campeonato Espanhol, podemos querer fazer assim: copiamos a URL que está na barra de endereço. Podemos clicar com o botão direito e copiar ou fazer ctrl + C, ou command + C e, em seguida, voltamos para a planilha do Google e aí precisamos usar aspas. Então cada vez que você inserir o parâmetro de URL no importHTML, ele tem de estar entre aspas e depois fazemos command + V ou ctrl + V para colar o endereço e depois outra aspa e então usamos o separador.

Perceba que o separador é indicado nas duas dicas que a planilha do Google apresenta a você. Dependendo das configurações de idioma de sua conta do Google, no Google Sheets, este separador pode ser um caractere diferente. Portanto, se o seu comando importHTML não está funcionando, pode ser que você esteja usando um separador diferente do que eu estou usando. Então dê uma olhada nisso.

Então eu vou colocar uma vírgula aqui, e então eu vou colocar a palavra "table" entre aspas, porque é isso que estamos importando. E então, finalmente, a posição da tabela na página web. Agora, voltando ao exemplo, vemos que a tabela "La Liga" é a segunda tabela que eu vejo na página. Então, vamos digitar, vamos voltar e digitar o número 2.

É importante que o número não esteja entre aspas. O parâmetro URL está, e o parâmetro de busca está, mas a posição da tabela na página web não está entre aspas. Então, depois disso, fechamos o parêntese e apertamos Enter e voilà, temos a tabela que representa os resultados da segunda tabela na página do Guardian.

E se quiséssemos carregar a terceira tabela? Nós só precisaríamos alterar o número aqui no topo para o número 3 e ele carrega a tabela automaticamente. Ou podemos voltar ao número 1 e podemos simplesmente copiar a primeira célula aqui. Então, eu estou copiando aqui, clico com o botão direito, "Copiar" e depois desço aqui na linha 7 e então eu coloco e mudo aqui para o número 2. Agora o importHTML está importando duas tabelas, a primeira e a segunda. Observe que quando eu clico na célula A1, o número aqui é 1, o que indica que é a primeira tabela na página web. E então quando eu vou para a linha 7, aqui diz tabela número 2. Então este é um exemplo. Deixe-me remover isso e dar outro exemplo.

Esta é a lista dos filmes indianos de maior bilheteria. É um artigo da Wikipedia. Veja que existem várias tabelas nesta página web também. Ea página web nos faz pensar que a primeira tabela que está mostrando a lista dos filmes indianos de maior bilheteria é realmente a primeira tabela aqui, certo?, na página web. Então vamos dar uma olhada nisso. Nós copiamos o endereço aqui no topo, "Copiar". Então eu volto aqui, célula vazia 1, digito =importHTML(e coloco aspas, pois o endereço, a URL precisa estar entre aspas, e em seguida, o segundo parâmetro, também entre aspas, "table" e depois vírgula de novo. E então, porque essa parece a primeira tabela, vamos usar o número 1 aqui também. E, então, apertamos Enter, e veja o que acontece.

Ele na verdade não carregou o que parece ser a primeira tabela na página. Isso porque importHTML na verdade não importa a primeira tabela que nós vemos. ImportHTML importa o código HTML da página e carrega a primeira tabela que aparece no código. Então, precisamos ver, se não funcionar, e no nosso caso não funcionou, podemos querer mudar o número aqui rapidamente, apenas para testar, experimentar e testar para ver se podemos carregar a tabela que estamos realmente vendo. Então, se mudamos isso para o número 2, para a próxima tabela no código, então você vai ver que ele carrega a tabela que estamos vendo aqui. Portanto, há uma tabela aqui no código HTML que vem antes da tabela que estamos vendo. Há uma estrutura aqui que foi configurada para ser uma tabela HTML, mas não está parecendo ser uma tabela HTML para nós. Então você só vai lá e muda o número de 1 para 2 e aí você carrega a tabela certa.

Deixe-me dar outro exemplo para que a gente veja o quão rápido importHTML funciona. Bom, essas tabelas todas são pequenas, com 25 linhas, 10 linhas. E se quiséssemos carregar uma tabela que tem centenas de linhas? Certo, deixe-me mostrar um exemplo.

Esta é uma página web que mostra códigos de aeroportos na Europa. E veja que existem centenas de aeroportos aqui, todos os aeroportos na Europa. E esta é de fato uma tabela. E se quiséssemos importar esta tabela automaticamente para importHTML? Então, nós apenas copiamos aqui o endereço, "Copiar", então você volta aqui, =importHTML(e então colocamos a URL entre aspas. Vírgula, "table", vírgula novamente, table entre aspas. E então nós colocamos a posição da tabela no código HTML e apertamos Enter. Veja o quão rápido carrega, temos mais de 200 aeroportos na Europa e ele carrega muito rápido.

E a coisa boa do importHTML é que se eu deixar esta tabela aqui como ela está, com a função importHTML ativada, ele vai atualizar automaticamente a tabela aqui na planilha do Google se fizerem qualquer mudança no site. Então, se houver qualquer atualização aqui,

ou aqui, ou aqui, minha planilha do Google será atualizada automaticamente também porque o importHTML acessa o site de tempos em tempos automaticamente para carregar os resultados mais recentes.

Agora, algumas considerações são importantes. ImportHTML só vai importar tabelas HTML, portanto você precisa checar o código para se certificar de que o que você está importando é realmente uma tabela, certo? Então, se nós, por exemplo, clicamos com o botão direito aqui e fazemos uma inspeção para ver o código HTML, veremos que há de fato uma tag "table" aqui no código HTML que descreve esta tabela. Às vezes, há elementos em uma página web que parecem uma tabela HTML, mas não são uma tabela. Eles podem ser um Div, eles podem usar uma tag diferente para representar o que vemos como uma tabela mas não é realmente uma tabela HTML. Então, esse é um alerta sobre o importHTML.

Outra coisa que precisamos considerar é que o importHTML só vai retornar tabelas que são visíveis para o Google Sheets, certo? Então, como o Google Sheets vai tentar acessar o site, a resposta que o site dá ao Google Sheets precisa ser o código HTML que você vê aqui no seu navegador. Às vezes, alguns sites vão impedir o acesso automatizado, e o Google Sheets é uma espécie de acesso automatizado ao site. Então alguns sites vão bloquear ou impedir que o código HTML completo seja carregado. Então, é bom você prestar atenção a isso também se você conseguir resultados incompletos.

Então é isso, este é o importHTML e é assim que importamos tabelas automaticamente para o Google Sheets.