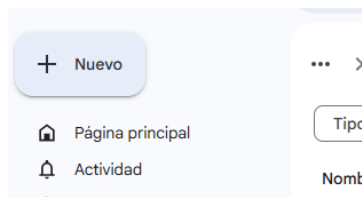


Prácticas Scraping.

1. Entra en la hoja de cálculo de google . Google calc.
 - entra por ejemplo en tu drive , pica en nuevo ----- hoja de cálculo.



Práctica 1: Importación de Tablas Automáticas

Ejemplo: Aprender a extraer bases de datos completas de la Wikipedia sin copiar y pegar manualmente.

1. **En la celda A1:** Pega la URL https://es.wikipedia.org/wiki/Anexo:Países_por_población esta es una página web de Wikipedia con información de los diferentes países.
2. **En la celda A3:** Escribe la siguiente fórmula:
=IMPORTHTML(A1; "table"; 1) con esto extraemos toda la información de la tabla 1 de dicha página web.
3. Cambia el número 1 por el 2 y **ve lo que ocurre.**
4. **Manda captura de pantalla de la información extraída al correo del profesor**

Práctica 2: Extracción de Titulares (XPath de Texto)

Objetivo: Capturar la actualidad de un periódico en tiempo real.

1. **Copia en la casilla A1 la URL de trabajo:** <https://elpais.com/> (o el periódico que prefieras). con esto le decimos la url de la que extraer la información.
2. Copia el **Xpath** para extraer el titular en la casilla A2, el xpath es el código que le dice que parte del código de la página web extraer, por así decirlo la etiqueta a buscar. **//h2**
3. En la casilla A4 copia la siguiente instrucción con la que importan los datos según el xpath de la casilla A2
=IMPORTXML(A1; A2)
4. **Manda captura de pantalla con la información.**

Te aporoto una tabla con diferentes xpath y descripción de lo extraen.

Todos los títulos 1	<code>//h2</code>	Extrae el texto de todos los titulares de noticias.
		Extrae las URLs (links), no el texto.(cuando en el código aparece la arroba significa que trabajamos con páginas web)
Todos los enlaces	<code>//a/@href</code>	
Texto de los enlaces	<code>//a/text()</code>	Extrae solo lo que se puede leer del botón/link.
Por clase específica	<code>//div[@class='precio']</code>	Busca un div que tenga exactamente esa clase.
Por ID único	<code>//*[@id='titular-principal']</code>	La forma más precisa de ir a un dato concreto.
Imágenes	<code>//img/@src</code>	Extrae la URL de todas las imágenes de la web.

Práctica 3: Extracción de Titulares (XPath de Texto).

Extra ahora todos los links de todos los enlaces de la web y manda captura.

Práctica 4: Extracción de Titulares (XPath de Texto).

Extra ahora todos los links de las URLs de todas las imágenes de la web y manda captura.

Pica aquí para un test