

Desafío con Regex

El siguiente bloque de texto HTML contiene información sobre varios libros disponibles en una tienda. Nuestro objetivo es extraer información clave utilizando **expresiones regulares** y organizarla en una tabla. La información a extraer incluye:

1. **Título del libro:** El texto que se encuentra dentro de las etiquetas ``.
2. **Autor del libro:** El texto que se encuentra dentro de las etiquetas ``.
3. **Precio:** La cantidad que aparece dentro de las etiquetas ``, precedida por el símbolo de dólar \$.
4. **Unidades en stock:** El número que aparece después de "unidades en stock" en el texto fuerte ``.
5. **Correo electrónico:** La dirección de correo que se encuentra dentro de las etiquetas `<a>` con el atributo `href` que comienza con `mailto:`.
6. **Número de teléfono:** El número que se menciona al principio de la frase que contiene la palabra "llama al".
7. **Información de contacto en redes sociales:** Los nombres de usuario que aparecen precedidos por @.

A continuación se muestra un ejemplo de cómo podrías escribir un script en Python utilizando **expresiones regulares** para extraer esta información:

```
import re

# Texto HTML que contiene la información de los libros
texto_html = """
<html>
  <head>
    <title>Catálogo de Libros</title>
  </head>
  <body>
    <h1>Bienvenidos a nuestra tienda de libros</h1>
    <p>Estamos emocionados de presentar una amplia variedad de
libros.
    Algunos de nuestros títulos más populares incluyen:</p>
    <div class="producto">
      <p>He estado leyendo un libro fascinante llamado
<strong>El Camino del Artista</strong>, de <em>Julia Cameron</em>.
      Te lo recomiendo mucho, realmente es transformador.
      Por cierto, puedes comprarlo por solo <span>$15.99</span>
si te interesa.
      Recuerda que hay <strong>34 unidades en
stock</strong>.</p>
      <p>Si deseas más información, contacta a <a
href="mailto:julia@example.com">julia@example.com</a> o llama al +34
612 345 678.
```

```
        ¡No te olvides de seguir a @juliacameron en redes
sociales!</p>
    </div>
    <div class="producto">
        <p>En esta sección, también ofrecemos <strong>Cien años de
soledad</strong> de <em>Gabriel García Márquez</em>.
        Es un clásico, aunque el precio de <span>$12.50</span>
podría parecer un poco alto.
        Aún así, tenemos <strong>50 unidades en stock</strong>,
así que no dudes en preguntar.
        Puedes enviar un correo a <a
href="mailto:gabriel@example.com">gabriel@example.com</a> o marcar +57
123 456 7890 para más detalles.</p>
        <p>Además, ¡sigue a @gabrielgarciamarquez en Twitter para
actualizaciones!</p>
    </div>
    <div class="producto">
        <p>Si te gustan las historias conmovedoras, deberías leer
<strong>El Principito</strong>. Escrito por
        <em>Antoine de Saint-Exupéry</em>, este libro es ideal
para reflexionar sobre la vida. El precio es de solo
        <span>$10.00</span>. Actualmente, tenemos <strong>120
unidades en stock</strong>, ¡así que no te lo pierdas!</p>
        <p>Para más información, contáctame a <a
href="mailto:antoine@example.com">antoine@example.com</a> o al +33 1
23 45 67 89. No te olvides de seguir a @saint_exupery para más
contenido interesante.</p>
    </div>
    <div class="producto">
        <p>He escuchado que <strong>La sombra del viento</strong>
de <em>Carlos Ruiz Zafón</em> es un gran libro,
        y su precio es de <span>$13.75</span>. Estamos emocionados
de tener <strong>78 unidades en stock</strong>.
        La historia es cautivadora, puedes pedir más detalles al
<a href="mailto:carlos@example.com">carlos@example.com</a>
        o al +34 987 654 321.</p>
        <p>No olvides seguir a @carlosruizzafon en Instagram, ¡él
comparte mucho contenido interesante!</p>
    </div>
    <div class="producto">
        <p>Finalmente, para aquellos interesados en estrategias,
<strong>El arte de la guerra</strong> de <em>Sun Tzu</em>
        es imprescindible. A solo <span>$9.99</span> y con
<strong>25 unidades en stock</strong>, es una gran oportunidad.
        Contacta a <a
href="mailto:sun@example.com">sun@example.com</a> o llama al +86 10
1234 5678 para más información.</p>
        <p>¡Sigue a @suntzu para obtener más consejos
estratégicos.</p>
```

```
        </div>
        <p>¡Gracias por visitar nuestra tienda de libros! No olvides
que también tenemos promociones especiales, así que estate atento.
Si necesitas ayuda, estamos aquí para ti. ¡Que tengas un gran
día!</p>
    </body>
</html>
"""
```

```
# Patrón para el título del libro
```

```
patron_titulo = r"<strong>(?!\\d+\\s+unidades)(.*?)</strong>"
titulos_libros = re.findall(patron_titulo, texto_html)
titulos_libros
```

```
['El Camino del Artista',
 'Cien años de soledad',
 'El Principito',
 'La sombra del viento',
 'El arte de la guerra']
```

```
# Patrón para el autor del libro
```

```
patron_autor = r"<em>(.*?)</em>"
autores_libros = re.findall(patron_autor, texto_html)
autores_libros
```

```
['Julia Cameron',
 'Gabriel García Márquez',
 'Antoine de Saint-Exupéry',
 'Carlos Ruiz Zafón',
 'Sun Tzu']
```

```
# Patrón para el precio
```

```
patron_precio = r"\\$(\\d+\\.\\d{2})"
precios_libros = re.findall(patron_precio, texto_html)
precios_libros
```

```
['15.99', '12.50', '10.00', '13.75', '9.99']
```

```
# Patrón para las unidades en stock
```

```
patron_stock = r"<strong>(\\d+)\\s+unidades en stock</strong>"
unidades_stock = re.findall(patron_stock, texto_html)
unidades_stock
```

```
['34', '50', '120', '78', '25']
```

```
# Patrón para el correo electrónico
```

```
patron_email = r"mailto:(.*?)\\"
correos_electronicos = re.findall(patron_email, texto_html)
correos_electronicos
```

```

['julia@example.com',
 'gabriel@example.com',
 'antoine@example.com',
 'carlos@example.com',
 'sun@example.com']

# Patrón para el número de teléfono
patron_telefono = r"\+[\\d\\s]+[\\d]"
telefonos = re.findall(patron_telefono, texto_html)
telefonos

['+34 612 345 678',
 '+57 123 456 7890',
 '+33 1 23 45 67 89',
 '+34 987 654 321',
 '+86 10 1234 5678']

# Patrón para los usuarios en redes sociales (precedidos por @)
patron_redes = r"(?!\\w)(@\\w+)"
usuarios_redes = re.findall(patron_redes, texto_html)
usuarios_redes

['@juliacameron',
 '@gabrielgarciamarquez',
 '@saint_exupery',
 '@carlosruizzafon',
 '@suntzu']

import pandas as pd

# Crear el diccionario con las listas proporcionadas
data = {
    "Titulo del libro": titulos_libros,
    "Autor": autores_libros,
    "Precio": precios_libros,
    "Unidades en stock": unidades_stock,
    "Correo electronico": correos_electronicos,
    "Telefono": telefonos,
    "Usuario en redes sociales": usuarios_redes
}

# Crear DataFrame
tabla = pd.DataFrame(data)

# Mostrar DataFrame
tabla

```

	Titulo del libro	Autor	Precio	Unidades en stock
0	El Camino del Artista	Julia Cameron	15.99	

1	Cien años de soledad	Gabriel García Márquez	12.50
50			
2	El Principito	Antoine de Saint-Exupéry	10.00
120			
3	La sombra del viento	Carlos Ruiz Zafón	13.75
78			
4	El arte de la guerra	Sun Tzu	9.99
25			
	Correo electronico	Telefono	Usuario en redes sociales
0	julia@example.com	+34 612 345 678	@juliacameron
1	gabriel@example.com	+57 123 456 7890	@gabrielgarciamarquez
2	antoine@example.com	+33 1 23 45 67 89	@saint_exupery
3	carlos@example.com	+34 987 654 321	@carlosruizzafon
4	sun@example.com	+86 10 1234 5678	@suntzu

Desafío con Regex (Problemas en el Texto)

El siguiente bloque de texto HTML contiene información sobre varios libros disponibles en una tienda. Nuestro objetivo es extraer información clave utilizando **expresiones regulares** y organizarla en una tabla. La información a extraer incluye:

1. **Título del libro:** El texto que se encuentra dentro de las etiquetas ``.
2. **Autor del libro:** El texto que se encuentra dentro de las etiquetas ``.
3. **Precio:** La cantidad que aparece dentro de las etiquetas ``, precedida por el símbolo de dólar \$.
4. **Unidades en stock:** El número que aparece después de "unidades en stock" en el texto fuerte ``.
5. **Correo electrónico:** La dirección de correo que se encuentra dentro de las etiquetas `<a>` con el atributo `href` que comienza con `mailto:`.
6. **Número de teléfono:** El número que se menciona al principio de la frase que contiene la palabra "llama al".
7. **Información de contacto en redes sociales:** Los nombres de usuario que aparecen precedidos por @.

A continuación se muestra un ejemplo de cómo podrías escribir un script en Python utilizando **expresiones regulares** para extraer esta información:

```
import re

# Texto HTML que contiene la información de los libros
texto_html = """
<html>
<head>
<title>Catálogo de Libros</title>
</head>

```

```
<body>
<h1>Bienvenidos a nuestra tienda de libros</h1>
<p>Estamos emocionados de presentar una amplia variedad de libros.
Algunos de nuestros títulos más populares incluyen:</p>
<div class="producto">
<p>He estado leyendo un libro fascinante llamado <strong>El
Camino del Artista</strong>, de <em>Julia Cameron</em>. Te lo
recomiendo mucho, realmente es transformador.
Por cierto, puedes comprarlo por solo <span>$15.99</span> si
te interesa.
Recuerda que hay <strong>34 unidades en stock</strong>.</p>
<p>Si deseas más información, contacta a <a
href="mailto:julia@example.com">julia@example.com</a> o llama al +34
612 345 678.
¡No te olvides de seguir a @juliacameron en redes
sociales!</p>
</div>
<div class="producto">
<p>En esta sección, también ofrecemos <strong>Cien años de
soledad</strong> de <em>Gabriel García Márquez</em>.
Es un clásico, aunque el precio de <span>$12.50</span> podría
parecer un poco alto.
Aún así, tenemos <strong>50 unidades en stock</strong>, así
que no dudes en preguntar.
Puedes enviar un correo a <a
href="mailto:gabriel@example.com">gabriel@example.com</a> o marcar +57
123 456 7890 para más detalles.</p>
<p>Además, ¡sigue a @gabrielgarciamarquez en Twitter para
actualizaciones!</p>
</div>
<div class="producto">
<p>Si te gustan las historias conmovedoras, deberías leer
<strong>El Principito</strong>. Escrito por <em>Antoine de Saint-
Exupéry</em>,
este libro es ideal para reflexionar sobre la vida. El precio
es de solo <span>$10.00</span>.
Actualmente, tenemos <strong>120 unidades en stock</strong>,
¡así que no te lo pierdas!</p>
<p>Para más información, contáctame a <a
href="mailto:antoine@example.com">antoine@example.com</a> o al +33 1
23 45 67 89.
No te olvides de seguir a @saint_exupery para más contenido
interesante.</p>
</div>
<div class="producto">
<p>He escuchado que <strong>La sombra del viento</strong> de
<em>Carlos Ruiz Zafón</em> es un gran libro,
y su precio es de <span>$13.75</span>. Estamos emocionados de
tener <strong>78 unidades en stock</strong>.
```

```
La historia es cautivadora, puedes pedir más detalles al <a
href="mailto:carlos@example.com">carlos@example.com</a> o al +34 987
654 321.</p>
```

```
<p>No olvides seguir a @carlosruizzafon en Instagram, ¡él
comparte mucho contenido interesante!</p>
```

```
</div>
```

```
<div class="producto">
```

```
<p>Finalmente, para aquellos interesados en estrategias,
<strong>El arte de la guerra</strong> de <em>Sun Tzu</em> es
imprescindible.
```

```
A solo <span>$9.99</span> y con <strong>25 unidades en
stock</strong>, es una gran oportunidad.
```

```
  Contacta a <a
href="mailto:sun@example.com">sun@example.com</a> o llama al +86 10
1234 5678 para más información.</p>
```

```
<p>Sigue a @suntzu para obtener más consejos estratégicos.</p>
```

```
</div>
```

```
<p>¡Gracias por visitar nuestra tienda de libros! No olvides que
también tenemos promociones especiales, así que estate atento. Si
necesitas ayuda, estamos aquí para ti. ¡Que tengas un gran día!</p>
```

```
</body>
```

```
</html>
```

```
"""
```

```
def rsl(lista):
```

```
    return [i.replace('\n', ' ') for i in lista]
```

```
# Patrón para el título del libro
```

```
patron_titulo = r"<strong>(?!\\d+\\s+unidades)([<]*)</strong>"
```

```
titulos_libros = rsl(re.findall(patron_titulo, texto_html, re.DOTALL))
```

```
titulos_libros
```

```
['El Camino del Artista',
 'Cien años de soledad',
 'El Principito',
 'La sombra del viento',
 'El arte de la guerra']
```

```
# Patrón para el autor del libro
```

```
patron_autor = r"<em>(.*?)</em>"
```

```
autores_libros = rsl(re.findall(patron_autor, texto_html, re.DOTALL))
```

```
autores_libros
```

```
['Julia Cameron',
 'Gabriel García Márquez',
 'Antoine de Saint- Exupéry',
 'Carlos Ruiz Zafón',
 'Sun Tzu']
```

```
# Patrón para el precio
```

```
patron_precio = r"<span>\$(\\d+\\.\\d{2})</span>"
```

```
precios_libros = rsl(re.findall(patron_precio, texto_html, re.DOTALL))
precios_libros
```

```
['15.99', '12.50', '10.00', '13.75', '9.99']
```

```
# Patrón para las unidades en stock
```

```
patron_stock = r"<strong>(\d+)\s+unidades\s+en\s+stock</strong>"
unidades_stock = rsl(re.findall(patron_stock, texto_html, re.DOTALL))
unidades_stock
```

```
['34', '50', '120', '78', '25']
```

```
# Patrón para correos electrónicos
```

```
patron_correos = r"mailto:(.*?)\"
correos_electronicos = rsl(re.findall(patron_correos, texto_html,
re.DOTALL))
correos_electronicos
```

```
['julia@example.com',
'gabriel@example.com',
'antoine@example.com',
'carlos@example.com',
'sun@example.com']
```

```
# Patrón para el número de teléfono
```

```
patron_telefono = r"\+[\d\s]+[\d]"
telefonos = rsl(re.findall(patron_telefono, texto_html, re.DOTALL))
telefonos
```

```
['+34 612 345 678',
'+57 123 456 7890',
'+33 1 23 45 67 89',
'+34 987 654 321',
'+86 10 1234 5678']
```

```
# Patrón para los usuarios en redes sociales (precedidos por @)
```

```
patron_redes = r"(?!\\w)(@\\w+)"
usuarios_redes = rsl(re.findall(patron_redes, texto_html, re.DOTALL))
usuarios_redes
```

```
['@juliacameron',
'@gabrielgarciamarquez',
'@saint_exupery',
'@carlosruizzafon',
'@suntzu']
```

```
import pandas as pd
```

```
# Crear el diccionario con las listas proporcionadas
```

```
data = {
    "Titulo del libro": titulos_libros,
    "Autor": autores_libros,
```



```

    "Precio": precios_libros,
    "Unidades en stock": unidades_stock,
    "Correo electronico": correos_electronicos,
    "Telefono": telefonos,
    "Usuario en redes sociales": usuarios_redes
}

# Crear DataFrame
tabla = pd.DataFrame(data)

# Mostrar DataFrame
tabla

```

	Titulo del libro	Autor	Precio	Unidades en stock \
0	El Camino del Artista	Julia Cameron	15.99	34
1	Cien años de soledad	Gabriel García Márquez	12.50	50
2	El Principito	Antoine de Saint- Exupéry	10.00	120
3	La sombra del viento	Carlos Ruiz Zafón	13.75	78
4	El arte de la guerra	Sun Tzu	9.99	25

	Correo electronico	Telefono	Usuario en redes sociales
0	julia@example.com	+34 612 345 678	@juliacameron
1	gabriel@example.com	+57 123 456 7890	@gabrielgarciamarquez
2	antoine@example.com	+33 1 23 45 67 89	@saint_exupery
3	carlos@example.com	+34 987 654 321	@carlosruizzafon
4	sun@example.com	+86 10 1234 5678	@suntzu

Desafío con Regex

El siguiente bloque de texto HTML contiene información sobre varios dispositivos electrónicos disponibles en una tienda. Nuestro objetivo es extraer información clave utilizando **expresiones regulares** y organizarla en una tabla. La información a extraer incluye:

1. **Nombre del dispositivo:** El texto que se encuentra dentro de las etiquetas ``, seguido de una etiqueta ``.
2. **Marca del dispositivo:** El texto que se encuentra dentro de las etiquetas ``.
3. **Precio:** La cantidad que aparece dentro de las etiquetas ``, precedida por el símbolo de dólar \$.
4. **Unidades en stock:** El número que aparece después de "unidades en stock" en el texto fuerte ``.
5. **Correo electrónico:** La dirección de correo que se encuentra dentro de las etiquetas `<a>` con el atributo `href` que comienza con `mailto:`.

6. **Número de teléfono:** El número que se menciona en el formato internacional, precedido por el símbolo +.
7. **Información de contacto en redes sociales:** Los nombres de usuario que aparecen precedidos por @.
8. **Código del producto:** El texto que aparece después de "Código del producto:" en el texto fuerte .

Utiliza expresiones regulares en Python para capturar esta información y construir una tabla organizada.

```
import re

# Texto HTML que contiene la información de los dispositivos
texto_html = """
<html>
<head>
  <title>Catálogo de Dispositivos Electrónicos</title>
</head>
<body>
  <h1>Bienvenidos a nuestra tienda de dispositivos electrónicos</h1>
  <p>Estamos emocionados de presentar una amplia variedad de
productos. Algunos de nuestros artículos más populares incluyen:</p>

  <div class="producto">
    <p>He estado usando un dispositivo increíble llamado
<strong>Smartwatch Pro</strong>, <em>Gadget Corp</em>.
    Te lo recomiendo mucho, realmente es transformador. Por
cierto, puedes comprarlo por solo <span>$199.99</span> si te interesa.

    Recuerda que hay <strong>50 unidades en stock</strong>. Código
del producto: <strong>COD-001</strong>.</p>
    <p>Si deseas más información, contacta a <a
href="mailto:support@gadgetcorp.com">support@gadgetcorp.com</a> o
llama al +34 612 345 678.
    ¡No te olvides de seguir a @gadgetcorp en redes sociales!</p>
  </div>

  <div class="producto">
    <p>En esta sección, también ofrecemos <strong>Cámara Ultra
HD</strong>, <em>Camera Inc</em>.
    Es un clásico, aunque el precio de <span>$299.00</span> podría
parecer un poco alto.
    Aún así, tenemos <strong>20 unidades en stock</strong>. Código
del producto: <strong>COD-002</strong>.</p>
    <p>Puedes enviar un correo a <a
href="mailto:sales@camerainc.com">sales@camerainc.com</a> o marcar +57
123 456 7890 para más detalles.</p>
    <p>Además, ¡sigue a @camera_inc en Twitter para
actualizaciones!</p>
  </div>
</body>
</html>
"""
```

```
<div class="producto">
  <p>Si te gustan los gadgets innovadores, deberías leer sobre
<strong>Auriculares Inalámbricos</strong>, <em>SoundWave</em>.
  Este producto es ideal para disfrutar de la música. El precio
es de solo <span>$89.99</span>.
  Actualmente, tenemos <strong>100 unidades en stock</strong>.
Código del producto: <strong>COD-003</strong>.</p>
  <p>Para más información, contáctame a <a
href="mailto:info@soundwave.com">info@soundwave.com</a> o al +33 1 23
45 67 89.
  No te olvides de seguir a @soundwave para más contenido
interesante.</p>
</div>
```

```
<div class="producto">
  <p>He escuchado que <strong>Portátil Ultra Ligero</strong>,
<em>TechZone</em> es un gran dispositivo,
  y su precio es de <span>$799.99</span>. Estamos emocionados de
tener <strong>30 unidades en stock</strong>.
  Código del producto: <strong>COD-004</strong>.</p>
  <p>La historia es cautivadora, puedes pedir más detalles al <a
href="mailto:contact@techzone.com">contact@techzone.com</a> o al +34
987 654 321.</p>
  <p>No olvides seguir a @techzone en Instagram, ¡ellos
comparten mucho contenido interesante!</p>
</div>
```

```
<div class="producto">
  <p>Finalmente, para aquellos interesados en videojuegos,
<strong>Consola de Juegos Avanzada</strong>, <em>GameSphere</em> es
imprescindible.
  A solo <span>$499.99</span> y con <strong>25 unidades en
stock</strong>. Código del producto: <strong>COD-005</strong>.</p>
  <p>Contacta a <a
href="mailto:support@gamesphere.com">support@gamesphere.com</a> o
llama al +86 10 1234 5678 para más información.</p>
  <p>Sigue a @gamesphere para obtener más consejos
estratégicos.</p>
</div>
```

```
<div class="producto">
  <p>¿Buscas una excelente laptop? <strong>Gamer Laptop
X</strong>, <em>UltraTech</em> tiene un rendimiento impresionante.
  Su precio es de <span>$1,199.00</span> y tenemos <strong>15
unidades en stock</strong>. Código del producto:
<strong>COD-006</strong>.</p>
  <p>Escribe a <a
href="mailto:info@ultratech.com">info@ultratech.com</a> o llama al +34
912 345 678 para más información.</p>
```

<p>También puedes seguir a @ultratech en redes sociales.</p>
</div>

<div class="producto">

<p>Para quienes buscan una pantalla grande, recomendamos el Monitor 4K, Visionary.

Con un precio de \$499.99 y 40 unidades en stock. Código del producto: COD-007.</p>

<p>Contáctanos a través de support@visionary.com o llama al +34 987 654 321.</p>

<p>Sigue a @visionary en Twitter para más actualizaciones.</p>
</div>

<div class="producto">

<p>La Tablet Pro, TabletMaker es perfecta para el trabajo y el ocio.

A solo \$349.99, tenemos 60 unidades en stock. Código del producto: COD-008.</p>

<p>Envía un correo a info@tabletmaker.com o llama al +34 654 321 987.</p>

<p>No olvides seguir a @tabletmaker en Instagram.</p>
</div>

<div class="producto">

<p>Para los amantes de la música, los Altavoces Bluetooth, SoundBlast son ideales.

A un precio de \$89.99 y 100 unidades en stock. Código del producto: COD-009.</p>

<p>Puedes contactarnos a support@soundblast.com o al +34 321 654 987.</p>

<p>Sigue a @soundblast en redes sociales.</p>
</div>

<div class="producto">

<p>Por último, los Relojes Inteligentes, TimeSmart son perfectos para los entusiastas de la tecnología.

Con un precio de \$149.99 y 75 unidades en stock. Código del producto: COD-010.</p>

<p>Para consultas, escribe a contact@timesmart.com o llama al +34 765 432 109.</p>

<p>No olvides seguir a @timesmart para más novedades.</p>
</div>

<p>¡Gracias por visitar nuestra tienda de dispositivos electrónicos! No olvides que también tenemos promociones especiales,

```
así que estate atento. Si necesitas ayuda, estamos aquí para ti. ¡Que
tengas un gran día!</p>
</body>
</html>
"""
```

```
# Patrón para el nombre del dispositivo
```

```
patron_nombre = r"<strong>(.*?)</strong>,\s+<em>"
nombres = re.findall(patron_nombre, texto_html)
nombres
```

```
['Smartwatch Pro',
 'Cámara Ultra HD',
 'Auriculares Inalámbricos',
 'Portátil Ultra Ligero',
 'Consola de Juegos Avanzada',
 'Gamer Laptop X',
 'Monitor 4K',
 'Tablet Pro',
 'Altavoces Bluetooth',
 'Relojes Inteligentes']
```

```
# Patrón para la marca del dispositivo
```

```
patron_marca = r"<em>(.*?)</em>"
marcas = re.findall(patron_marca, texto_html)
marcas
```

```
['Gadget Corp',
 'Camera Inc',
 'SoundWave',
 'TechZone',
 'GameSphere',
 'UltraTech',
 'Visionary',
 'TabletMaker',
 'SoundBlast',
 'TimeSmart']
```

```
# Patrón para el precio del dispositivo
```

```
patron_precio = r"\$(\d+(?:,\d+)*(?:\.\d{2}))"
precios = re.findall(patron_precio, texto_html)
precios
```

```
['199.99',
 '299.00',
 '89.99',
 '799.99',
 '499.99',
 '1,199.00',
 '499.99',
 '349.99',
```

```
'89.99',  
'149.99']
```

Patrón para las unidades en stock

```
patron_stock = r"<strong>(\d+) unidades en stock</strong>"
```

```
stocks = re.findall(patron_stock, texto_html)
```

```
stocks
```

```
['50', '20', '100', '30', '25', '15', '40', '60', '100', '75']
```

Patrón para extraer correos electrónicos

```
patron_email = r"mailto:(.*?)\""
```

```
correos = re.findall(patron_email, texto_html)
```

```
correos
```

```
['support@gadgetcorp.com',  
'sales@camerainc.com',  
'info@soundwave.com',  
'contact@techzone.com',  
'support@gamesphere.com',  
'info@ultratech.com',  
'support@visionary.com',  
'info@tabletmaker.com',  
'support@soundblast.com',  
'contact@timesmart.com']
```

Patrón para el número de teléfono en formato internacional

```
patron_telefono = r"\+[\\d\\s]+[\\d]"
```

```
telefonos = re.findall(patron_telefono, texto_html)
```

```
telefonos
```

```
['+34 612 345 678',  
'+57 123 456 7890',  
'+33 1 23 45 67 89',  
'+34 987 654 321',  
'+86 10 1234 5678',  
'+34 912 345 678',  
'+34 987 654 321',  
'+34 654 321 987',  
'+34 321 654 987',  
'+34 765 432 109']
```

Patrón para los usuarios en redes sociales

```
patron_redes = r"(?!\\w)(@\\w+)"
```

```
usuarios_redes = re.findall(patron_redes, texto_html)
```

```
usuarios_redes
```

```
['@gadgetcorp',  
'@camera_inc',  
'@soundwave',  
'@techzone',
```

```
'@gamesphere',  
'@ultratech',  
'@visionary',  
'@tabletmaker',  
'@soundblast',  
'@timesmart']
```

Patrón para el código del producto

```
patron_codigo = r"<strong>(COD-\d+)</strong>"  
codigos = re.findall(patron_codigo, texto_html)  
codigos
```

```
['COD-001',  
'COD-002',  
'COD-003',  
'COD-004',  
'COD-005',  
'COD-006',  
'COD-007',  
'COD-008',  
'COD-009',  
'COD-010']
```

```
import pandas as pd
```

Crear un diccionario con la información extraída

```
data = {  
    "Nombre del dispositivo": nombres,  
    "Marca": marcas,  
    "Precio": precios,  
    "Unidades en stock": stocks,  
    "Correo electronico": correos,  
    "Numero de telefono": telefonos,  
    "Usuario en redes sociales": usuarios_redes,  
    "Codigo del producto": codigos  
}
```

Crear un DataFrame a partir del diccionario

```
tabla = pd.DataFrame(data)
```

Mostrar la tabla resultante

```
tabla
```

	Nombre del dispositivo	Marca	Precio	Unidades en stock
0	Smartwatch Pro	Gadget Corp	199.99	50
1	Cámara Ultra HD	Camera Inc	299.00	20
2	Auriculares Inalámbricos	SoundWave	89.99	100

3	Portátil Ultra Ligero	TechZone	799.99	30
4	Consola de Juegos Avanzada	GameSphere	499.99	25
5	Gamer Laptop X	UltraTech	1,199.00	15
6	Monitor 4K	Visionary	499.99	40
7	Tablet Pro	TabletMaker	349.99	60
8	Altavoces Bluetooth	SoundBlast	89.99	100
9	Relojes Inteligentes	TimeSmart	149.99	75
	Correo electronico	Numero de telefono	Usuario en redes sociales	
\				
0	support@gadgetcorp.com	+34 612 345 678	@gadgetcorp	
1	sales@camerainc.com	+57 123 456 7890	@camera_inc	
2	info@soundwave.com	+33 1 23 45 67 89	@soundwave	
3	contact@techzone.com	+34 987 654 321	@techzone	
4	support@gamesphere.com	+86 10 1234 5678	@gamesphere	
5	info@ultratech.com	+34 912 345 678	@ultratech	
6	support@visionary.com	+34 987 654 321	@visionary	
7	info@tabletmaker.com	+34 654 321 987	@tabletmaker	
8	support@soundblast.com	+34 321 654 987	@soundblast	
9	contact@timesmart.com	+34 765 432 109	@timesmart	
	Codigo del producto			
0	COD-001			
1	COD-002			
2	COD-003			
3	COD-004			
4	COD-005			
5	COD-006			
6	COD-007			
7	COD-008			
8	COD-009			
9	COD-010			

Desafío con Regex

El siguiente bloque de texto contiene información sobre varios eventos que se llevarán a cabo en diferentes lugares. Nuestro objetivo es extraer información clave utilizando **expresiones regulares** y organizarla en una tabla. La información a extraer incluye:

1. **Nombre del evento:** El texto que se encuentra entre las etiquetas `<h2>`.
2. **Fecha del evento:** La fecha que se menciona en el formato DD/MM/AAAA dentro de las etiquetas `<date>`.
3. **Hora de inicio:** La hora que se menciona en el formato HH:MM dentro de las etiquetas `<time>`.
4. **Lugar:** El texto que aparece dentro de las etiquetas `<location>`.
5. **Precio:** El monto que aparece precedido por el símbolo de dólar \$ dentro de las etiquetas `<cost>`.
6. **Organizador:** El nombre que se encuentra dentro de las etiquetas `<organizer>`.
7. **Número de contacto:** El número de teléfono en formato internacional precedido por + dentro de las etiquetas `<contact>`.
8. **Sitio web:** La URL que se menciona dentro de las etiquetas `<url>`.

Utiliza expresiones regulares en Python para capturar esta información y construir una tabla organizada.

```
import re

# Texto HTML que contiene la información de los eventos
texto_html = """
<html>
<head>
  <title>Calendario de Eventos</title>
</head>
<body>
  <h1>Eventos Destacados</h1>
  <p>Explora nuestra variedad de eventos programados a lo largo del
año.</p>

  <div class="evento">
    <p>Asiste al <h2>Festival de Música</h2>, <em>Sonidos del
Mundo</em>.
    Disfruta de grandes artistas por solo <cost>$50.00</cost>.
    Hay <strong>100 unidades en stock</strong>. Código del evento:
<strong>EVT-001</strong>.</p>
    <p>Contacta a <a
href="mailto:info@sonidosdelmundo.com">info@sonidosdelmundo.com</a> o
llama al <contact>+34 123 456 789</contact>.</p>
    <p>Sigue a @sonidosdelmundo para más novedades.</p>
    <date>15/05/2024</date>
    <time>18:00</time>
    <location>Auditorio Nacional</location>
  </div>
</body>
</html>
"""
```

```
<url>www.sonidosdelmundo.com</url>
</div>

<div class="evento">
  <p>No te pierdas la <h2>Carrera de Maratón</h2>, <em>Correr es
Vida</em>.
  Participa por solo <cost>$30.00</cost>.
  Quedan <strong>200 unidades en stock</strong>. Código del
evento: <strong>EVT-002</strong>.</p>
  <p>Puedes enviar un correo a <a
href="mailto:contacto@correresvida.com">contacto@correresvida.com</a>
o llamar al <contact>+34 987 654 321</contact>.</p>
  <p>Sigue a @correresvida para actualizaciones.</p>
  <date>20/06/2024</date>
  <time>09:00</time>
  <location>Parque Central</location>
  <url>www.correresvida.com</url>
</div>

<div class="evento">
  <p>Ven al <h2>Simposio de Tecnología</h2>,
<em>Innovadores</em>.
  A un costo de <cost>$100.00</cost>.
  Disponibles <strong>150 unidades en stock</strong>. Código del
evento: <strong>EVT-003</strong>.</p>
  <p>Para más información, contacta a <a
href="mailto:info@innovadores.com">info@innovadores.com</a> o al
<contact>+34 321 654 987</contact>.</p>
  <p>Sigue a @innovadores en redes sociales.</p>
  <date>10/07/2024</date>
  <time>10:00</time>
  <location>Centro de Convenciones</location>
  <url>www.innovadores.com</url>
</div>

<div class="evento">
  <p>Participa en la <h2>Exposición de Arte</h2>, <em>Artistas
Unidos</em>.
  El precio es de <cost>$25.00</cost>.
  Tenemos <strong>50 unidades en stock</strong>. Código del
evento: <strong>EVT-004</strong>.</p>
  <p>Contacta a <a
href="mailto:info@artistasunidos.com">info@artistasunidos.com</a> o
llama al <contact>+34 987 654 321</contact>.</p>
  <p>No olvides seguir a @artistasunidos en Instagram.</p>
  <date>25/08/2024</date>
  <time>11:00</time>
  <location>Galería de Arte Moderno</location>
  <url>www.artistasunidos.com</url>
</div>
```

```
<div class="evento">
  <p>Únete al <h2>Congreso de Salud</h2>, <em>Salud y
Bienestar</em>.
  A solo <cost>$75.00</cost>.
  Tenemos <strong>80 unidades en stock</strong>. Código del
evento: <strong>EVT-005</strong>.</p>
  <p>Puedes escribir a <a
href="mailto:contacto@saludybienestar.com">contacto@saludybienestar.co
m</a> o al <contact>+34 654 321 987</contact>.</p>
  <p>Sigue a @saludybienestar para más información.</p>
  <date>30/09/2024</date>
  <time>15:00</time>
  <location>Centro de Salud y Bienestar</location>
  <url>www.saludybienestar.com</url>
</div>
```

```
<div class="evento">
  <p>Disfruta de la <h2>Feria del Libro</h2>, <em>Lector en
Acción</em>.
  A un costo de <cost>$10.00</cost>.
  Quedan <strong>500 unidades en stock</strong>. Código del
evento: <strong>EVT-006</strong>.</p>
  <p>Escribe a <a
href="mailto:info@lectoraccion.com">info@lectoraccion.com</a> o llama
al <contact>+34 654 321 000</contact>.</p>
  <p>Más información en @lectoraccion.</p>
  <date>05/10/2024</date>
  <time>10:00</time>
  <location>Biblioteca Municipal</location>
  <url>www.lectoraccion.com</url>
</div>
```

```
<div class="evento">
  <p>Asiste a la <h2>Conferencia sobre Innovación</h2>,
<em>Futuro Brillante</em>.
  Participa por <cost>$60.00</cost>.
  Disponibles <strong>300 unidades en stock</strong>. Código del
evento: <strong>EVT-007</strong>.</p>
  <p>Puedes contactarnos a <a
href="mailto:info@futurobrillante.com">info@futurobrillante.com</a> o
al <contact>+34 987 123 456</contact>.</p>
  <p>Sigue a @futurobrillante para actualizaciones.</p>
  <date>15/11/2024</date>
  <time>14:00</time>
  <location>Auditorio Principal</location>
  <url>www.futurobrillante.com</url>
</div>
```

```
<div class="evento">
```

```
<p>Únete al <h2>Seminario de Marketing Digital</h2>,
<em>Marketing al Día</em>.
A solo <cost>$45.00</cost>.
Hay <strong>250 unidades en stock</strong>. Código del evento:
<strong>EVT-008</strong>.</p>
<p>Contacta a <a
href="mailto:info@marketingaldia.com">info@marketingaldia.com</a> o
llama al <contact>+34 321 987 654</contact>.</p>
<p>No olvides seguir a @marketingaldia en redes sociales.</p>
<date>12/12/2024</date>
<time>16:00</time>
<location>Centro de Formación</location>
<url>www.marketingaldia.com</url>
</div>

<div class="evento">
  <p>No te pierdas la <h2>Encuentro de Fotografía</h2>,
  <em>Captura el Momento</em>.
  A un costo de <cost>$35.00</cost>.
  Quedan <strong>150 unidades en stock</strong>. Código del
  evento: <strong>EVT-009</strong>.</p>
  <p>Puedes contactarnos a <a
href="mailto:info@capturaelmomento.com">info@capturaelmomento.com</a>
o al <contact>+34 999 654 321</contact>.</p>
  <p>Sigue a @capturaelmomento para más información.</p>
  <date>20/01/2025</date>
  <time>10:30</time>
  <location>Estudio Fotográfico</location>
  <url>www.capturaelmomento.com</url>
</div>

<div class="evento">
  <p>Participa en la <h2>Convención de Videojuegos</h2>,
  <em>Gamer Fest</em>.
  El precio es de <cost>$80.00</cost>.
  Hay <strong>200 unidades en stock</strong>. Código del evento:
  <strong>EVT-010</strong>.</p>
  <p>Escribe a <a
href="mailto:info@gamerfest.com">info@gamerfest.com</a> o llama al
<contact>+34 888 888 888</contact>.</p>
  <p>No olvides seguir a @gamerfest en redes sociales.</p>
  <date>01/02/2025</date>
  <time>11:00</time>
  <location>Centro de Exposiciones</location>
  <url>www.gamerfest.com</url>
</div>
</body>
</html>
"""
```

```

# Patrón para el nombre del evento
patron_evento = r"<h2>(.*?)</h2>"
eventos = re.findall(patron_evento, texto_html)
eventos

['Festival de Música',
 'Carrera de Maratón',
 'Simposio de Tecnología',
 'Exposición de Arte',
 'Congreso de Salud',
 'Feria del Libro',
 'Conferencia sobre Innovación',
 'Seminario de Marketing Digital',
 'Encuentro de Fotografía',
 'Convención de Videojuegos']

# Patrón para la fecha del evento (DD/MM/AAAA)
patron_fecha = r"<date>(\d{2}/\d{2}/\d{4})</date>"
fechas = re.findall(patron_fecha, texto_html)
fechas

['15/05/2024',
 '20/06/2024',
 '10/07/2024',
 '25/08/2024',
 '30/09/2024',
 '05/10/2024',
 '15/11/2024',
 '12/12/2024',
 '20/01/2025',
 '01/02/2025']

# Patrón para la hora de inicio (HH:MM)
patron_hora = r"<time>(\d{2}:\d{2})</time>"
horas = re.findall(patron_hora, texto_html)
horas

['18:00',
 '09:00',
 '10:00',
 '11:00',
 '15:00',
 '10:00',
 '14:00',
 '16:00',
 '10:30',
 '11:00']

# Patrón para el lugar
patron_lugar = r"<location>(.*?)</location>"

```

```

lugares = re.findall(patron_lugar, texto_html)
lugares

['Auditorio Nacional',
 'Parque Central',
 'Centro de Convenciones',
 'Galería de Arte Moderno',
 'Centro de Salud y Bienestar',
 'Biblioteca Municipal',
 'Auditorio Principal',
 'Centro de Formación',
 'Estudio Fotográfico',
 'Centro de Exposiciones']

# Patrón para el precio (puede incluir comas en los miles)
patronPrecio = r"<cost>\$(\d+(?:,\d+)*(?:\.\d{2}))</cost>"
precios = re.findall(patronPrecio, texto_html)
precios

['50.00',
 '30.00',
 '100.00',
 '25.00',
 '75.00',
 '10.00',
 '60.00',
 '45.00',
 '35.00',
 '80.00']

# Patrón para el organizador
patronOrganizador = r"<em>(.*?)</em>"
organizadores = re.findall(patronOrganizador, texto_html)
organizadores

['Sonidos del Mundo',
 'Correr es Vida',
 'Innovadores',
 'Artistas Unidos',
 'Salud y Bienestar',
 'Lector en Acción',
 'Futuro Brillante',
 'Marketing al Día',
 'Captura el Momento',
 'Gamer Fest']

# Patrón para el número de contacto (en formato internacional)
patronContacto = r"<contact>(\+[\\d\\s]+[\\d])</contact>"
contactos = re.findall(patronContacto, texto_html)
contactos

```

```
[ '+34 123 456 789',
  '+34 987 654 321',
  '+34 321 654 987',
  '+34 987 654 321',
  '+34 654 321 987',
  '+34 654 321 000',
  '+34 987 123 456',
  '+34 321 987 654',
  '+34 999 654 321',
  '+34 888 888 888']
```

Patrón para el sitio web

```
patron_url = r"<url>(.*?)</url>"
```

```
urls = re.findall(patron_url, texto_html)
```

```
urls
```

```
['www.sonidosdelmundo.com',
 'www.correresvida.com',
 'www.innovadores.com',
 'www.artistasunidos.com',
 'www.saludybienestar.com',
 'www.lectoraccion.com',
 'www.futurobrillante.com',
 'www.marketingaldia.com',
 'www.capturaelmomento.com',
 'www.gamerfest.com']
```

```
import pandas as pd
```

Crear un diccionario con los datos

```
data = {
    "Nombre del evento": eventos,
    "Fecha del evento": fechas,
    "Hora de inicio": horas,
    "Lugar": lugares,
    "Precio": precios,
    "Organizador": organizadores,
    "Numero de contacto": contactos,
    "Sitio Web": urls
}
```

Crear un DataFrame con los datos

```
tabla_eventos = pd.DataFrame(data)
```

Mostrar la tabla resultante

```
tabla_eventos
```

	Nombre del evento	Fecha del evento	Hora de inicio	\
0	Festival de Música	15/05/2024	18:00	
1	Carrera de Maratón	20/06/2024	09:00	
2	Simposio de Tecnología	10/07/2024	10:00	

3	Exposición de Arte	25/08/2024	11:00
4	Congreso de Salud	30/09/2024	15:00
5	Feria del Libro	05/10/2024	10:00
6	Conferencia sobre Innovación	15/11/2024	14:00
7	Seminario de Marketing Digital	12/12/2024	16:00
8	Encuentro de Fotografía	20/01/2025	10:30
9	Convención de Videojuegos	01/02/2025	11:00

	Lugar	Precio	Organizador	Numero de contacto \
0	Auditorio Nacional	50.00	Sonidos del Mundo	+34 123 456 789
1	Parque Central	30.00	Correr es Vida	+34 987 654 321
2	Centro de Convenciones	100.00	Innovadores	+34 321 654 987
3	Galería de Arte Moderno	25.00	Artistas Unidos	+34 987 654 321
4	Centro de Salud y Bienestar	75.00	Salud y Bienestar	+34 654 321 987
5	Biblioteca Municipal	10.00	Lector en Acción	+34 654 321 000
6	Auditorio Principal	60.00	Futuro Brillante	+34 987 123 456
7	Centro de Formación	45.00	Marketing al Día	+34 321 987 654
8	Estudio Fotográfico	35.00	Captura el Momento	+34 999 654 321
9	Centro de Exposiciones	80.00	Gamer Fest	+34 888 888 888

	Sitio Web
0	www.sonidosdelmundo.com
1	www.correresvida.com
2	www.innovadores.com
3	www.artistasunidos.com
4	www.saludybienestar.com
5	www.lectoraccion.com
6	www.futurobrillante.com
7	www.marketingaldia.com
8	www.capturaelmomento.com
9	www.gamerfest.com