

## S12 T02: Tasca de web scraping

```
In [2]: import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
from selenium.webdriver import Firefox
from selenium.webdriver.firefox.options import Options as options
from selenium.webdriver.firefox.service import Service
from selenium import webdriver
from selenium.webdriver.common.by import By
from scrapy import Selector
import scrapy
from scrapy.crawler import CrawlerProcess
```

### Exercici 1. Realitza web scraping d'una pàgina de la borsa de Madrid (<https://www.bolsamadrid.es> (<https://www.bolsamadrid.es/>)) utilitzant BeautifulSoup i Selenium.

Escollim la pàgina que conté l'última valoració de l'Íbex 35. La url seria la següent:

<https://www.bolsamadrid.es/esp/aspx/Mercados/Precios.aspx?indice=ESI100000000> (<https://www.bolsamadrid.es/esp/aspx/Mercados/Precios.aspx?indice=ESI100000000>)

L'inici de la consulta (query) es fa amb el paràmetre `?`; això vol dir que "indice=ESI100000000" està codificat com a "key=value" pairs, i és el resultat de la consulta.

Fem una exploració amb shortcut `ctrl+shift+I` (tinc linux, però també és vàlid per windows), per veure l'estructura amb la consola del navegador (Firefox). Identifiquem html i javascript i els fulls d'estil css a primer cop d'ull.

### Web scraping amb BeautifulSoup

```
In [3]: URL = "https://www.bolsamadrid.es/esp/asp/Mercados/Precios.aspx?indice=ESI100000000"
page = requests.get(URL)

print(page.text)
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-tran
sitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head data-idioma="esp" data-hora-act="Sun, 22 May 2022 15:16:05 GMT" data-app-path="/" data-bolsa="BMadri
d" data-analytics-id="UA-35966870-2"><meta http-equiv="X-UA-Compatible" content="IE=11" /><meta http-equiv
="Content-Type" content="text/html; charset=UTF-8" /><meta id="ctl00_copyright" name="copyright" content="
Copyright © BME 2022" /><title>
    Bolsa de Madrid - Precios de la sesión
</title><link id="ctl00_RSSLink1" rel="alternate" type="application/rss+xml" href="/esp/asp/RSS/RSS.ashx?
feed=Todo" title="Bolsa de Madrid: Todos los contenidos agregados" /><link id="ctl00_RSSLink2" rel="altern
ate" type="application/rss+xml" href="/esp/asp/RSS/RSS.ashx?feed=NotasPrensa" title="Bolsa de Madrid: Not
as de Prensa" /><link id="ctl00_RSSLink3" rel="alternate" type="application/rss+xml" href="/esp/asp/RSS/R
SS.ashx?feed=Regulacion" title="Bolsa de Madrid: Regulación: Circulares e Instrucciones Operativas" /><lin
k id="ctl00_RSSLink4" rel="alternate" type="application/rss+xml" href="/esp/asp/RSS/RSS.ashx?feed=Indice
s" title="Bolsa de Madrid: Avisos Índices IBEX" /><link id="ctl00_RSSLink5" rel="alternate" type="applicat
ion/rss+xml" href="/esp/asp/RSS/RSS.ashx?feed=Revista" title="Bolsa de Madrid: Artículos de la Revista Bo
lsa" /><script src="/js/Master?v=thBxh-xvKtDQpWbHp0k4A3E9sxEywhoh56HgLOsWH3I1"></script>
<link href="/css/css?v=e0iHiLxmqlVKER4v5gQI2K9pdMEvzhww7tRrYIcX2w41" rel="stylesheet"/>
<script src="/asp/Mercados/Precios.js" language="javascript" type="text/javascript"></script></head>
```

Observem la informació de les empreses integrants de l'Íbex-35, diferents menús, capçaleres i peus de pàgina.

Per exemple, **Telefónica** està inserida amb aquesta programació a dins de la taula (veiem també la seva capçalera).

```
<div>
  <table class="TblPort" cellspacing="0" cellpadding="3" id="ctl00_Contenido_tblAcciones" width="
100%">
    <tr align="center">
      <th scope="col">Nombre</th><th scope="col">&#218;lt.</th><th scope="col">% Dif.</th><th
scope="col">M&#225;x.</th><th scope="col">M&#237;n.</th><th scope="col">Volumen</th><th scope="col"
>Efectivo (miles €)</th><th scope="col">Fecha</th><th class="Ult" scope="col">Hora</th>
```

Es pot fer transformar la presentació de la informació amb pàgines web especialitzades com [webformatter](https://webformatter.com/html) (<https://webformatter.com/html>), i visualitzar el codi de forma més intel·ligible.

```
<div class="Menu2 MenuDiv Off" menu="m335"><div class="MenuLink MenuConOp Off"><a href="#" target="
_self">Formación</a></div></div>
<div class="Menu3 MenuSubOp" menu="m335">
<div class="Menu3 MenuDiv Off" menu="m336"><div class="MenuLink Off"><a href="/esp/Inversores/Forma
cion/QueeslaBolsa.aspx" target="_self">¿Qué es la Bolsa?</a></div></div>
<div class="Menu3 MenuDiv Off" menu="m337"><div class="MenuLink Off"><a href="/esp/asp/Inversores/
Formacion/Formacion.aspx?tipo=T" target="_self">Cursos y Masters</a></div></div>
<div class="Menu3 MenuDiv Off" menu="m338"><div class="MenuLink Off"><a href="/esp/Inversores/Forma
cion/Invertir/PerfilInversor.aspx" target="_self">Cómo invertir en Bolsa</a></div></div>
<div class="Menu3 MenuDiv Off" menu="m346"><div class="MenuLink Off"><a href="/esp/Inversores/Forma
cion/Guias/Guias.aspx" target="_self">Guías del inversor</a></div></div>
<div class="Menu3 MenuDiv Off" menu="m347"><div class="MenuLink Off"><a href="/esp/Inversores/Forma
cion/Glosario/A.aspx" target="_self">Glosario</a></div></div>
<div class="Menu3 MenuDiv Off" menu="m375"><div class="MenuLink Off"><a href="/esp/Inversores/Forma
cion/Invertir/FAQS.aspx" target="_self">Preguntas Frecuentes</a></div></div>
</div>
<div class="Menu2 MenuDiv Off" menu="m376"><div class="MenuLink Off"><a href="/esp/Inversores/Merca
doEsp.aspx" target="_self">El Mercado de Valores Español</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m380"><div class="MenuLink Off"><a href="http://biblioteca.bol
sasymercados.es/opac/" target="_blank">Biblioteca BME</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m381"><div class="MenuLink Off"><a href="https://www.bolsasyme
rcados.es/esp/Estudios-Publicaciones/Libros-y-Guias" target="_blank">Libros y guías</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m382"><div class="MenuLink Off"><a href="/esp/asp/Inversores/
Agenda/Agenda.aspx" target="_self">Agenda del Inversor</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m388"><div class="MenuLink Off"><a href="/esp/Inversores/ServI
nversores.aspx" target="_self">Servicios para Inversores</a></div></div>
</div>
<div class="Menu1 MenuDiv Blanco Off" menu="m389"><div class="MenuLink MenuConOp Off"><a href="#" t
arget="_self">PARA EMPRESAS</a></div></div>
<div class="Menu2 MenuSubOp" menu="m389">
<div class="Menu2 MenuDiv Off" menu="m390"><div class="MenuLink Off"><a href="/esp/ParaEmp/ComoCoti
zar/ComoCotizar1.aspx" target="_self">Cómo cotizar en Bolsa</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m391"><div class="MenuLink Off"><a href="/esp/ParaEmp/VenCotiz
ar/VenCotizar1.aspx" target="_self">Ventajas de Cotizar en Bolsa</a></div></div>
```

```
<div class="Menu2 MenuDiv Off" menu="m398"><div class="MenuLink Off"><a href="http://www.bolsasymer  
cados.es/BME_4Companies/esp" target="_blank">BME 4Companies</a></div></div>  
<div class="Menu2 MenuDiv Off" menu="m399"><div class="MenuLink Off"><a href="http://www.entornopre  
mercado.es" target="_blank">Entorno Pre Mercado</a></div></div>  
<div class="Menu2 MenuDiv Off" menu="m400"><div class="MenuLink Off"><a href="/esp/ParaEmp/Eventos-  
Corporativos.aspx" target="_self">Eventos Corporativos</a></div></div>  
</div>  
<div class="Menu1 MenuDiv Blanco Off" menu="m401"><div class="MenuLink MenuConOp Off"><a href="#" t  
arget="_self">PARA MIEMBROS Y PROFESIONALES</a></div></div>  
<div class="Menu2 MenuSubOp" menu="m401">  
<div class="Menu2 MenuDiv Off" menu="m402"><div class="MenuLink Off"><a href="/esp/Miembros/Ibex/IB  
EXContratoLicenciaUso.aspx" target="_self">Contratos IBEX</a></div></div>  
<div class="Menu2 MenuDiv Off" menu="m403"><div class="MenuLink Off"><a href="/esp/Miembros/Ibex/Co  
ntacto.aspx" target="_self">Contacto</a></div></div>  
<div class="Menu2 MenuDiv Off" menu="m404"><div class="MenuLink Off"><a href="javascript:sacaVentan  
a('/docs/SBolsas/InformesSB/listadodevalores.pdf')" target="_self">Listado de Valores</a></div></di  
v>  
<div class="Menu2 MenuDiv Off" menu="m405"><div class="MenuLink MenuConOp Off"><a href="#" target="_  
_self">Modelo de Mercado</a></div></div>  
<div class="Menu3 MenuSubOp" menu="m405">  
<div class="Menu3 MenuDiv Off" menu="m406"><div class="MenuLink Off"><a href="javascript:sacaVentan  
a('/docs/SBolsas/docsSubidos/SIBE/modelomercado.pdf')" target="_self">Acciones y Derechos</a></div>  
</div>  
<div class="Menu3 MenuDiv Off" menu="m407"><div class="MenuLink Off"><a href="javascript:sacaVentan  
a('/docs/SBolsas/docsSubidos/SIBE/modelomercadoETF.pdf')" target="_self">ETFs</a></div></div>  
<div class="Menu3 MenuDiv Off" menu="m408"><div class="MenuLink Off"><a href="javascript:sacaVentan  
a('/docs/SBolsas/docsSubidos/SIBE/mdlmdowarrcertotros.pdf')" target="_self">Warrants</a></div></di  
v>  
</div>  
<div class="Menu2 MenuDiv Off" menu="m409"><div class="MenuLink MenuConOp Off"><a href="/esp/Miembr  
os/AccesoMercado.aspx" target="_self">Acceso al Mercado</a></div></div>  
<div class="Menu3 MenuSubOp" menu="m409">  
<div class="Menu3 MenuDiv Off" menu="m410"><div class="MenuLink Off"><a href="/esp/Miembros/AplicEx  
ternas.aspx" target="_self">Aplicaciones externas</a></div></div>
```

```
</div>
<div class="Menu2 MenuDiv Off" menu="m411"><div class="MenuLink MenuConOp Off"><a href="#" target=
_self">Tarifas</a></div></div>
<div class="Menu3 MenuSubOp" menu="m411">
<div class="Menu3 MenuDiv Off" menu="m412"><div class="MenuLink Off"><a href="/docs/Miembros/tarifa
sSB.pdf" target="_blank">Sociedad de Bolsas</a></div></div>
<div class="Menu3 MenuDiv Off" menu="m413"><div class="MenuLink Off"><a href="/docs/Miembros/Tarifa
sRectoras.pdf" target="_blank">Rectoras</a></div></div>
</div>
<div class="Menu2 MenuDiv Off" menu="m414"><div class="MenuLink Off"><a href="javascript:sacaVentan
a('/docs/SBolsas/InformesSB/miembrosmercado.pdf')" target="_self">Miembros del Mercado</a></d
iv>
<div class="Menu2 MenuDiv Off" menu="m415"><div class="MenuLink Off"><a href="/esp/asp/Regulacion/
Regulacion.aspx?cod=SBolsas" target="_self">Regulación</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m416"><div class="MenuLink Off"><a href="javascript:sacaVentan
a('/docs/Acciones/sect.pdf')" target="_self">Clasificación sectorial</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m417"><div class="MenuLink Off"><a href="/esp/Miembros/Servici
os/MiembrosServicios.aspx" target="_self">Servicios para miembros y profesionales</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m419"><div class="MenuLink Off"><a href="/esp/asp/Miembros/MI
FID-II.aspx" target="_self">MiFID II</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m429"><div class="MenuLink Off"><a href="/esp/Miembros/BMESecu
ritiesSolutions.aspx" target="_self">BME Securities Solutions</a></div></div>
</div>
<div class="Menu1 MenuDiv Blanco Off" menu="m430"><div class="MenuLink MenuConOp Off"><a href="#" t
arget="_self">PARA MEDIOS DE COMUNICACIÓN</a></div></div>
<div class="Menu2 MenuSubOp" menu="m430">
<div class="Menu2 MenuDiv Off" menu="m431"><div class="MenuLink Off"><a href="/esp/asp/Prensa/Sala
Prensa.aspx?tipo=07" target="_self">Notas de Prensa</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m432"><div class="MenuLink Off"><a href="/esp/asp/Prensa/Cont
acte.aspx" target="_self">Contacto de Prensa</a></div></div>
<div class="Menu2 MenuDiv Off" menu="m433"><div class="MenuLink Off"><a href="http://www.bolsasymer
cados.es/esp/Sala-Comunicacion" target="Prensa">Sala de Prensa</a></div></div>
</div>
</div>
```

```
</div>
<script language="javascript">
var MenuIzqLoad_MenusSel = [ "m196","m197","m198" ];
var MenuIzqLoad_MenusDespl = [  ];
MenuIzqLoad(false, 'General');
</script>
</td>

<td class="BaseCont">

<div class="col6"><form method="post" action="/esp/asp/Mercados/Precios.aspx?indice=ESI100000000"
id="aspnetForm">
<input type="hidden" name="__VIEWSTATE" id="__VIEWSTATE" value="jGhsci5wgdIrrZqch+0DS5YQa68TFABnHjX
sAwTRgpNSHLwDJAXmKU3Rh0qwrPlzXhQLWStonLDGb7uw6reUR8L0yJomwY3t1Tuwg6q2gJbH7UdvtqQ7iUGAo4PT+gghNM9wkk
ndV5V/nzhDCKh0tlQaf3WcHu5rE2tdvSCvt5Hs5/4T6HQzG5kTivPJiohWghqreKhzyHANpEosEH8mFfYit2yiRYswXY7qZNfZj
TnfxTCWsSF54qPQTWkbgDv9dmohvw==" />

<input type="hidden" name="__VIEWSTATEGENERATOR" id="__VIEWSTATEGENERATOR" value="DC05C2DC" />
<input type="hidden" name="__EVENTVALIDATION" id="__EVENTVALIDATION" value="fh96g569rPozDIuBsdEXy2s
rM4y79F1nnhPP0jRNp6rK+rnXUgl7vDKjdFYa+AhVI0qhvIQcj9W7IExG7PbszXyJGxBAhH4sXFa6jsoi7A6NciKv+vVlUZ4pnq
Q387tiuLv83xIFrqlnejwkIcF2z/q16gC232d/5pnIIpHI8jNzx0tZqB8Rk0m8pRV63hbvwIZLAVHstitsZkZi+laDrR8X4LBAs
0E4QpPSWcIYzyenrSYeVgWRp5mpFi0AJl9LEvEnSa8s+CnNT1w83DqVkJSPm7rhWwIlHAYuG+0j0MupgegAUwkLS8l3JHPiuxQwG
N/5flR3Tz4R7Xs0zRj9v8EXssgrDv7udYhve09ehHdm02HoTYZDzRFuem8L1hs63qwVt+3InmlvtguwNWMn3iZjfw9T2dB0i6LU
qZdaIXsnkAkBbSKoNDW72GuZxU3tbklRn7Vlt89bpvliu7ur5y5ceBr8jIHVwxs3stp0zZ8NJ5m49y+TfPckMjsEx0MHrdA5e9D
RfK1iwoITwIuyLkJbZJ2xYsqHyKXWPMc6z/ny+yWxcL/Vi8Pn6bBxPKj4dX7xNbWy9PgZRSC9dMYDe2+bxbu9lrYuh+w0lbM3Yz
qv04g4P9r/4mAw24Zvct6oV2dL/bMdtNBBTxy/IUV6qP9R5LBAN8xtcFG0+y8z5n0egmS2GTZHTqoV57r//JGNNaIzDgMGnjk5
enmFM0jab03FVKCa03dN2mnVmbxbpGwrg9YoYQtPMHEGu3mIKSTtmAhK2ywo1D3IogRud0a70GhMXhxTALHdZL3nUKUqbfGcVxd
4uUek4fLpBhHhkg1jqN49Gg==" />
<div class="TituloPag">Precios de la sesión</div>

<table cellpadding="5" cellspacing="2" class="FrmBusq" style="width: 60%;" align="center"><tr>
<td class="Etiqueta">Mercado</td><td class="Campo"> <select name="ctl00$Contenido$SelMercado" id
="SelMercado" onchange="rstSel(0);">
```

```

        <option selected="selected" value="" style="color: #999999;">Seleccione mercado</option>
        <option value="MC">Mercado Continuo</option>
        <option value="ETF">ETFs</option>
        <option value="LT">Latibex</option>
        <option value="GROWTH">BME Growth</option>
        <option value="CMAD">Corros Madrid</option>
        <option value="CBAR">Corros Barcelona</option>
        <option value="CBIL">Corros Bilbao</option>
        <option value="CVAL">Corros Valencia</option>

    </select> </td>
    <td class="Etiqueta">Índice</td><td class="Campo"> <select name="ctl00$Contenido$SelIndice" id="SelIndice" onchange="rstSel(1);">
        <option value="" style="color: #999999;">Seleccione Índice</option>
        <option selected="selected" value="ESI1000000000">IBEX 35</option>
        <option value="ESI0600000000">IBEX MEDIUM CAP</option>
        <option value="ESI2400000000">IBEX SMALL CAP</option>
        <option value="ESI2600000000">IBEX TOP Dividendo</option>
        <option value="ES0S00000901">IBEX 35; Bancos</option>
        <option value="ES0S00000919">IBEX 35; Energía</option>
        <option value="ES0S00000927">IBEX 35; Construcción</option>
        <option value="ES0S00001586">IBEX Gender Equality</option>
        <option value="ES0S00001594">IBEX Gender Equality Total Return</option>
        <option value="ES0S00001602">IBEX Gender Equality Net Return</option>
        <option value="ESI2300000000">FTSE4Good IBEX</option>
        <option value="ESI2000000000">FTSE Latibex All Share</option>
        <option value="ESI2200000000">FTSE Latibex TOP</option>
        <option value="ESI2500000000">FTSE Latibex BRASIL</option>

    </select> </td>
    <td class="Etiqueta">Sector</td><td class="Campo"> <select name="ctl00$Contenido$SelSector" id="SelSector" onchange="rstSel(2);">
        <option value="" style="color: #999999;">Seleccione sector</option>
        <option value="01">Petróleo y Energía</option>

```



```

        <option value="02">Mat.Basicos, Industria y Construcci&#243;n</option>
        <option value="03">Bienes de Consumo</option>
        <option value="04">Servicios de Consumo</option>
        <option value="05">Servicios Financieros</option>
        <option value="06">Tecnolog&#237;a y Telecomunicaciones</option>
        <option value="07">Servicios Inmobiliarios</option>

</select> </td>
    <td class="Buscar">
        <input type="submit" name="ctl00$Contenido$Consultar" value="Consultar" onclick="return chkSel
();" id="ctl00_Contenido_Consultar" />
    </td>
</tr></table>
<br /><br />

<div class="LinkAntSig">

<div class="Ctr">IBEX 35®</div>
<div class="finFila"></div>
</div>

<div>
    <table class="TblPort" cellpadding="3" cellspacing="0" align="Center" id="ctl00_Contenido_tblÍndice" width="80%" style="margin-bottom: 20px;">
        <tr align="center">
            <th scope="col">Nombre</th><th scope="col">Anterior</th><th scope="col">&#218;ltimo</th>
            <th scope="col">% Dif.</th><th scope="col">M&#225;ximo</th><th scope="col">M&#237;nimo</th><th scope="col">Fecha</th><th scope="col">Hora</th><th class="Ult" scope="col">% Dif.<br/>Año 2022</th>
        </tr><tr align="right">
            <td class="DifFlSb" align="left">IBEX 35&#174;</td><td>8.406,00</td><td>8.484,50</td><td class="DifClSb">0,93</td><td>8.546,90</td><td>8.456,10</td><td align="center">20/05/2022</td><td align="center">17:38:00</td><td class="DifClBj">-2,63</td>
        </tr>

```

```
</table>
</div>

<div>
  <table class="TblPort" cellspacing="0" cellpadding="3" id="ctl00_Contenido_tblAcciones" width="
100%">
    <tr align="center">
      <th scope="col">Nombre</th><th scope="col">&#218;lt.</th><th scope="col">% Dif.</th><th
scope="col">M&#225;x.</th><th scope="col">M&#237;n.</th><th scope="col">Volumen</th><th scope="col"
>Efectivo (miles €)</th><th scope="col">Fecha</th><th class="Ult" scope="col">Hora</th>
    </tr><tr align="right">
      <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES012
5220311">ACCIONA</a></td><td>183,1000</td><td class="DifClSb">0,83</td><td>185,3000</td><td>182,900
0</td><td>89.921</td><td>16.536,77</td><td align="center">20/05/2022</td><td class="Ult" align="cen
ter">Cierre</td>
    </tr><tr align="right">
      <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES013
2105018">ACERINOX</a></td><td>10,6500</td><td class="DifClBj">-0,70</td><td>10,9500</td><td>10,6250
</td><td>877.025</td><td>9.480,32</td><td align="center">20/05/2022</td><td class="Ult" align="cent
er">Cierre</td>
    </tr><tr align="right">
      <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES016
7050915">ACS</a></td><td>25,4300</td><td class="DifClSb">1,07</td><td>25,8400</td><td>25,3300</td>
<td>624.557</td><td>15.948,53</td><td align="center">20/05/2022</td><td class="Ult" align="center">
Cierre</td>
    </tr><tr align="right">
      <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES010
5046009">AENA</a></td><td>135,8000</td><td class="DifClSb">1,46</td><td>136,3000</td><td>133,4000</
td><td>100.816</td><td>14.174,97</td><td align="center">20/05/2022</td><td class="Ult" align="cente
r">Cierre</td>
    </tr><tr align="right">
      <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES015
7097017">ALMIRALL</a></td><td>10,5800</td><td class="DifClSb">1,73</td><td>10,6000</td><td>10,3700
</td><td>346.430</td><td>3.642,88</td><td align="center">20/05/2022</td><td class="Ult" align="cent
```

```

er">Cierre</td>
    </tr><tr align="right">
        <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0109067019">AMADEUS</a></td><td>59,1000</td><td class="DifClSb">0,96</td><td>60,0400</td><td>58,7600</td><td>516.071</td><td>30.610,30</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
    </tr><tr align="right">
        <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=LU1598757687">ARCELORMIT.</a></td><td>27,4000</td><td class="DifClSb">1,03</td><td>28,0450</td><td>27,3300</td><td>615.965</td><td>17.103,20</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
    </tr><tr align="right">
        <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0113900J37">B.SANTANDER</a></td><td>2,7440</td><td class="DifClSb">0,13</td><td>2,7905</td><td>2,7320</td><td>40.067.111</td><td>110.999,29</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
    </tr><tr align="right">
        <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0113860A34">BA.SABADELL</a></td><td>0,7314</td><td class="DifClBj">-0,57</td><td>0,7504</td><td>0,7314</td><td>22.843.011</td><td>16.918,40</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
    </tr><tr align="right">
        <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0113679I37">BANKINTER</a></td><td>5,5640</td><td class="DifClBj">-0,68</td><td>5,6780</td><td>5,5560</td><td>2.052.978</td><td>11.519,40</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
    </tr><tr align="right">
        <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0113211835">BBVA</a></td><td>4,7475</td><td class="DifClBj">-0,40</td><td>4,8285</td><td>4,7355</td><td>24.941.682</td><td>119.088,61</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
    </tr><tr align="right">
        <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0140609019">CAIXABANK</a></td><td>3,1200</td><td class="DifClBj">-0,16</td><td>3,1840</td><td>3,1090</td>

```

```
td><td>18.086.447</td><td>56.753,73</td><td align="center">20/05/2022</td><td class="Ult" align="ce
nter">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES010
5066007">CELLNEX</a></td><td>40,4200</td><td class="DifClSb">2,20</td><td>40,8900</td><td>39,4100</
td><td>934.479</td><td>37.760,74</td><td align="center">20/05/2022</td><td class="Ult" align="cente
r">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES010
5630315">CIE AUTOMOT.</a></td><td>21,3000</td><td class="DifClSb">0,95</td><td>21,7000</td><td>21,1
400</td><td>258.060</td><td>5.525,16</td><td align="center">20/05/2022</td><td class="Ult" align="c
enter">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES013
0960018">ENAGAS</a></td><td>21,4700</td><td class="DifClSb">0,28</td><td>21,6300</td><td>21,2900</t
d><td>831.194</td><td>17.823,64</td><td align="center">20/05/2022</td><td class="Ult" align="cente
r">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES013
0670112">ENDESA</a></td><td>20,6700</td><td class="DifClSb">0,05</td><td>20,9200</td><td>20,5800</t
d><td>902.924</td><td>18.693,33</td><td align="center">20/05/2022</td><td class="Ult" align="cente
r">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES011
8900010">FERROVIAL</a></td><td>23,3100</td><td class="DifClSb">0,56</td><td>23,6400</td><td>23,1100
</td><td>791.476</td><td>18.479,13</td><td align="center">20/05/2022</td><td class="Ult" align="cen
ter">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES013
7650018">FLUIDRA</a></td><td>22,7800</td><td class="DifClSb">1,24</td><td>23,2400</td><td>22,5400</
td><td>310.475</td><td>7.128,03</td><td align="center">20/05/2022</td><td class="Ult" align="cente
r">Cierre</td>
</tr><tr align="right">
  <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES017
```

```
1996087">GRIFOLS CL.A</a></td><td>19,8800</td><td class="DifClBj">-0,15</td><td>20,1800</td><td>19,
6800</td><td>779.006</td><td>15.476,23</td><td align="center">20/05/2022</td><td class="Ult" align
="center">Cierre</td>
</tr><tr align="right">
<td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES017
7542018">IAG</a></td><td>1,4790</td><td class="DifClSb">2,21</td><td>1,5015</td><td>1,4540</td><td>
14.614.243</td><td>21.631,29</td><td align="center">20/05/2022</td><td class="Ult" align="center">C
ierre</td>
</tr><tr align="right">
<td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES014
4580Y14">IBERDROLA</a></td><td>11,1000</td><td class="DifClSb">3,11</td><td>11,1800</td><td>10,8600
</td><td>22.219.036</td><td>241.268,15</td><td align="center">20/05/2022</td><td class="Ult" align
="center">Cierre</td>
</tr><tr align="right">
<td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES014
8396007">INDITEX</a></td><td>20,4200</td><td class="DifClSb">0,54</td><td>20,7100</td><td>20,0600</
td><td>3.627.429</td><td>74.407,17</td><td align="center">20/05/2022</td><td class="Ult" align="cen
ter">Cierre</td>
</tr><tr align="right">
<td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES011
8594417">INDRA A</a></td><td>9,6150</td><td class="DifClSb">1,10</td><td>9,7700</td><td>9,4950</td>
<td>629.913</td><td>6.081,13</td><td align="center">20/05/2022</td><td class="Ult" align="center">C
ierre</td>
</tr><tr align="right">
<td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES013
9140174">INM.COLONIAL</a></td><td>7,6800</td><td class="DifClSb">0,92</td><td>7,7150</td><td>7,6000
</td><td>571.711</td><td>4.387,52</td><td align="center">20/05/2022</td><td class="Ult" align="cent
er">Cierre</td>
</tr><tr align="right">
<td class="DifFlBj" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES012
4244E34">MAPFRE</a></td><td>1,7730</td><td class="DifClBj">-0,11</td><td>1,8090</td><td>1,7730</td>
<td>4.653.735</td><td>8.316,98</td><td align="center">20/05/2022</td><td class="Ult" align="center"
>Cierre</td>
</tr><tr align="right">
```

```
<td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES0176252718">MELIA HOTELS</a></td><td>6,9650</td><td class="DifClSb">0,94</td><td>7,0700</td><td>6,9250</td><td>719.012</td><td>5.023,26</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES0105025003">MERLIN</a></td><td>10,3400</td><td class="DifClSb">0,98</td><td>10,5000</td><td>10,1500</td><td>841.403</td><td>8.740,20</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES0116870314">NATURGY</a></td><td>27,6800</td><td class="DifClSb">0,58</td><td>28,0200</td><td>27,5600</td><td>337.127</td><td>9.346,64</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES0169501022">PHARMA MAR</a></td><td>70,1000</td><td class="DifClSb">1,18</td><td>70,8000</td><td>69,2400</td><td>48.329</td><td>3.380,44</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES0173093024">R.E.C.</a></td><td>19,5250</td><td class="DifClSb">0,08</td><td>19,6100</td><td>19,4050</td><td>921.833</td><td>17.988,95</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
</tr><tr align="right">
  <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES0173516115">REPSOL</a></td><td>14,4100</td><td class="DifClBj">-0,21</td><td>14,7500</td><td>14,4050</td><td>6.414.647</td><td>92.102,10</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
</tr><tr align="right">
  <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/fichaValor.aspx?ISIN=ES0157261019">ROVI</a></td><td>59,9000</td><td class="DifClSb">4,08</td><td>60,6000</td><td>58,1500</td><td>131.909</td><td>7.877,93</td><td align="center">20/05/2022</td><td class="Ult" align="center">Cierre</td>
```

```

        </tr><tr align="right">
            <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES014
3416115">SIEMENS GAME</a></td><td>16,7450</td><td class="DifClSb">1,12</td><td>16,9300</td><td>16,4
550</td><td>2.112.904</td><td>35.052,29</td><td align="center">20/05/2022</td><td class="Ult" align
="center">Cierre</td>
        </tr><tr align="right">
            <td class="DifFlBj" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES016
5386014">SOLARIA</a></td><td>21,8600</td><td class="DifClBj">-3,70</td><td>22,7600</td><td>21,7100
</td><td>614.090</td><td>13.580,97</td><td align="center">20/05/2022</td><td class="Ult" align="cen
ter">Cierre</td>
        </tr><tr align="right">
            <td class="DifFlSb" align="left"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES017
8430E18">TELEFONICA</a></td><td>4,6920</td><td class="DifClSb">1,10</td><td>4,7080</td><td>4,6310</
td><td>10.879.203</td><td>50.798,39</td><td align="center">20/05/2022</td><td class="Ult" align="ce
nter">Cierre</td>
        </tr>
    </table>
</div>

```

```
<div class="Nota">
```

```
    Información diferida 15 minutos.<br />
```

```
<u>Precios</u> expresados en euros.<br />
```

```
<u>Efectivo</u> expresado en miles de euros.<br />
```

```
El <u>volumen y efectivo</u> para cada valor incluye todas las operaciones realizadas hasta el cier
re de la sesión de contratación. El <u>volumen y efectivo total</u> incluyendo las operaciones espe
ciales realizadas después del cierre de la sesión está disponible en la consulta histórica.<br />
```

```
</div>
```

```
</form></div>
```

```

        <div class="finFila"></div>
    </td>
</tr><tr class="PrePie">
    <td class="BaseIzq noimpr"></td>
    <td class="BaseSep noimpr"></td>
    <td colspan="2"><div id="capaIconosRRSS" class="noimpr">
        Síguenos en:<br />
        <ul><li class="mclic" ><a href="https://twitter.com/GrupoBME" target="_blank">&nbsp;<im
g src="/images/Twitter.png" border="0" align="absmiddle" alt="Twitter" />&nbsp;</a></li><li class="
mclic"><a href="https://www.flickr.com/photos/archivobme/" target="_blank">&nbsp;&nbsp;</a></li><li class="mclic"><a href
="https://www.linkedin.com/company/grupo-bme" target="_blank">&nbsp;&nbsp;</a></li><li class="mclic"><a href="https://ww
w.youtube.com/user/GrupoBME?ob=0&feature=results_main" target="_blank">&nbsp;&nbsp;</a></li></ul>
    </div></td>
    <td class="BaseSep noimpr"></td>
    <td class="BaseDcha noimpr"></td>
</tr><tr id="Pie" valign="top">
    <td class="PieIzq noimpr"></td>
    <td class="PieFondo BaseSep noimpr"></td>
    <td colspan="2" class="PieFondo"><div>
        <div id="PieOps" class="noimpr"><ul><li class="mclic"><a href="/esp/asp/Market-Status
/Market-Status.aspx" target="_self">&nbsp;Estado del Mercado&nbsp;</a></li><li class="mclic"><a hr
ef="/esp/asp/Comun/MapaWeb.aspx" target="_self">&nbsp;Mapa web&nbsp;</a></li><li class="mclic"><a
href="javascript:Disclaimer('esp')" target="_self">&nbsp;Disclaimer&nbsp;</a></li><li class="mclic
k"><a href="/esp/privacidad/Politica-de-Privacidad" target="_self">&nbsp;Política de Privacidad&nbs
p;</a></li><li class="mclic"><a href="/esp/privacidad/Politica-de-cookies" target="_self">&nbsp;Po
lítica de cookies&nbsp;</a></li><li class="mclic"><a href="/esp/asp/Privacidad/Aviso-Legal.aspx?C
odBolsa=BMadrid" target="_self">&nbsp;Aviso legal&nbsp;</a></li></ul></div>
        <div id="PieWebs" class="noimpr"><form id="formWeb">
            <select name="ctl00$webs" id="ctl00_webs">
<option value="">Webs del grupo BME</option>

```



```
<option value="https://www.bolsasymercados.es">BME</option>
```

Creem un BeautifulSoup object, amb "html.parser" com a analitzador (sintàctic) del contingut de la pàgina.

```
In [4]: soup = BeautifulSoup(page.content, "html.parser")
```

```
In [5]: print(soup)
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

```
<html xmlns="http://www.w3.org/1999/xhtml">
```

```
<head data-analytics-id="UA-35966870-2" data-app-path="/" data-bolsa="BMadrid" data-hora-act="Sun, 22 May 2022 15:16:05 GMT" data-idioma="esp"><meta content="IE=11" http-equiv="X-UA-Compatible"/><meta content="text/html; charset=utf-8" http-equiv="Content-Type"/><meta content="Copyright © BME 2022" id="ctl00_copyright" name="copyright"/><title>
```

```
Bolsa de Madrid - Precios de la sesión
```

```
</title><link href="/esp/asp/RSS/RSS.ashx?feed=Todo" id="ctl00_RSSLink1" rel="alternate" title="Bolsa de Madrid: Todos los contenidos agregados" type="application/rss+xml"/><link href="/esp/asp/RSS/RSS.ashx?feed=NotasPrensa" id="ctl00_RSSLink2" rel="alternate" title="Bolsa de Madrid: Notas de Prensa" type="application/rss+xml"/><link href="/esp/asp/RSS/RSS.ashx?feed=Regulacion" id="ctl00_RSSLink3" rel="alternate" title="Bolsa de Madrid: Regulación: Circulares e Instrucciones Operativas" type="application/rss+xml"/><link href="/esp/asp/RSS/RSS.ashx?feed=Indices" id="ctl00_RSSLink4" rel="alternate" title="Bolsa de Madrid: Avisos Índices IBEX" type="application/rss+xml"/><link href="/esp/asp/RSS/RSS.ashx?feed=Revista" id="ctl00_RSSLink5" rel="alternate" title="Bolsa de Madrid: Artículos de la Revista Bolsa" type="application/rss+xml"/><script src="/js/Master?v=thBxh-xvKtDQpWbHp0k4A3E9sxEywhoh56HgLOsWH3I1"></script>
```

```
<link href="/css/css?v=e0iHiLxmqlVKER4v5gQI2K9pdMEvzhww7tRrYIcX2w41" rel="stylesheet"/>
```

```
<script language="javascript" src="/js/Menu.js?v=thBxh-xvKtDQpWbHp0k4A3E9sxEywhoh56HgLOsWH3I1"></script></head>
```

Veiem un resultat molt semblant al aconseguit amb [webformatter].

Volem llistar les empreses de l'Íbex-35, les quals estan tabulades sota els següents elements:

```
div
table class="TblPort"
id="ctl00_Contenido_tblÍndice"
id="ctl00_Contenido_tblAcciones"
<th scope="col">Nombre</th>
```

```
In [6]: results = soup.find(id='ctl00_Contenido_tblAcciones')
```

```
In [7]: print(results)
```

```
<table cellpadding="3" cellspacing="0" class="TblPort" id="ctl00_Contenido_tblAcciones" width="100%">
<tr align="center">
<th scope="col">Nombre</th><th scope="col">Últ.</th><th scope="col">% Dif.</th><th scope="col">Máx.</th><th scope="col">Mín.</th><th scope="col">Volumen</th><th scope="col">Efectivo (miles €)</th><th scope="col">Fecha</th><th class="Ult" scope="col">Hora</th>
</tr><tr align="right">
<td align="left" class="DifFlSb"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0125220311">ACCIONA</a></td><td>183,1000</td><td class="DifClSb">0,83</td><td>185,3000</td><td>182,9000</td><td>89.921</td><td>16.536,77</td><td align="center">20/05/2022</td><td align="center" class="Ult">Cierre</td>
</tr><tr align="right">
<td align="left" class="DifFlBj"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0132105018">ACERINOX</a></td><td>10,6500</td><td class="DifClBj">-0,70</td><td>10,9500</td><td>10,6250</td><td>877.025</td><td>9.480,32</td><td align="center">20/05/2022</td><td align="center" class="Ult">Cierre</td>
</tr><tr align="right">
<td align="left" class="DifFlSb"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0167050915">ACS</a></td><td>25,4300</td><td class="DifClSb">1,07</td><td>25,8400</td><td>25,3300</td><td>624.557</td><td>15.948,53</td><td align="center">20/05/2022</td><td align="center" class="Ult">Cierre</td>
</tr><tr align="right">
<td align="left" class="DifFlSb"><a href="/esp/asp/empresas/FichaValor.aspx?ISIN=ES0105046009">AENA</a></td><td>125,0000</td><td class="DifClSb">1,40</td><td>126,0000</td><td>122,4000</td><td>100,010</td><td>14
```

```
In [8]: print(results.prettify())
```

```
<table cellpadding="3" cellspacing="0" class="TblPort" id="ctl00_Contenido_tblAcciones" width="100%">
  <tr align="center">
    <th scope="col">
      Nombre
    </th>
    <th scope="col">
      Últ.
    </th>
    <th scope="col">
      % Dif.
    </th>
    <th scope="col">
      Máx.
    </th>
    <th scope="col">
      Mín.
    </th>
    <th scope="col">
```

```
In [9]: #Observem les columnes de la capçalera i les llistem
ibex_elements = results.find_all("th")
```

```
In [10]: print(ibex_elements)

[<th scope="col">Nombre</th>, <th scope="col">Últ.</th>, <th scope="col">% Dif.</th>, <th scope="col">Má
x.</th>, <th scope="col">Mín.</th>, <th scope="col">Volumen</th>, <th scope="col">Efectivo (miles €)</th>,
<th scope="col">Fecha</th>, <th class="Ult" scope="col">Hora</th>]
```

```
In [11]: #Extraiem els títols de les columnes
index_col=[]

for element in results.find_all("th"):
    index_col.append(element.get_text())

print(index_col)

['Nombre', 'Últ.', '% Dif.', 'Máx.', 'Mín.', 'Volumen', 'Efectivo (miles €)', 'Fecha', 'Hora']
```

```
In [12]: files = []
for element in results.find_all("td"):
```

```

files.append(element.text)

print(files)
['ACCIONA', '183,1000', '0,83', '185,3000', '182,9000', '89.921', '16.536,77', '20/05/2022', 'Cierre', 'AC
ERINOX', '10,6500', '-0,70', '10,9500', '10,6250', '877.025', '9.480,32', '20/05/2022', 'Cierre', 'ACS', '
25,4300', '1,07', '25,8400', '25,3300', '624.557', '15.948,53', '20/05/2022', 'Cierre', 'AENA', '135,8000
', '1,46', '136,3000', '133,4000', '100.816', '14.174,97', '20/05/2022', 'Cierre', 'ALMIRALL', '10,5800',
'1,73', '10,6000', '10,3700', '346.430', '3.642,88', '20/05/2022', 'Cierre', 'AMADEUS', '59,1000', '0,96',
'60,0400', '58,7600', '516.071', '30.610,30', '20/05/2022', 'Cierre', 'ARCELORMIT.', '27,4000', '1,03', '2
8,0450', '27,3300', '615.965', '17.103,20', '20/05/2022', 'Cierre', 'B.SANTANDER', '2,7440', '0,13', '2,79
05', '2,7320', '40.067.111', '110.999,29', '20/05/2022', 'Cierre', 'BA.SABADELL', '0,7314', '-0,57', '0,75
04', '0,7314', '22.843.011', '16.918,40', '20/05/2022', 'Cierre', 'BANKINTER', '5,5640', '-0,68', '5,6780
', '5,5560', '2.052.978', '11.519,40', '20/05/2022', 'Cierre', 'BBVA', '4,7475', '-0,40', '4,8285', '4,735
5', '24.941.682', '119.088,61', '20/05/2022', 'Cierre', 'CAIXABANK', '3,1200', '-0,16', '3,1840', '3,1090
', '18.086.447', '56.753,73', '20/05/2022', 'Cierre', 'CELLNEX', '40,4200', '2,20', '40,8900', '39,4100',
'934.479', '37.760,74', '20/05/2022', 'Cierre', 'CIE AUTOMOT.', '21,3000', '0,95', '21,7000', '21,1400', '
258.060', '5.525,16', '20/05/2022', 'Cierre', 'ENAGAS', '21,4700', '0,28', '21,6300', '21,2900', '831.194
', '17.823,64', '20/05/2022', 'Cierre', 'ENDESA', '20,6700', '0,05', '20,9200', '20,5800', '902.924', '18.
693,33', '20/05/2022', 'Cierre', 'FERROVIAL', '23,3100', '0,56', '23,6400', '23,1100', '791.476', '18.479,
13', '20/05/2022', 'Cierre', 'FLUIDRA', '22,7800', '1,24', '23,2400', '22,5400', '310.475', '7.128,03', '2
0/05/2022', 'Cierre', 'GRIFOLS CL.A', '19,8800', '-0,15', '20,1800', '19,6800', '779.006', '15.476,23', '2
0/05/2022', 'Cierre', 'IAG', '1,4790', '2,21', '1,5015', '1,4540', '14.614.243', '21.631,29', '20/05/2022
', 'Cierre', 'IBERDROLA', '11,1000', '3,11', '11,1800', '10,8600', '22.219.036', '241.268,15', '20/05/2022
', 'Cierre', 'INDITEX', '20,4200', '0,54', '20,7100', '20,0600', '3.627.429', '74.407,17', '20/05/2022', '
Cierre', 'INDRA A', '9,6150', '1,10', '9,7700', '9,4950', '629.913', '6.081,13', '20/05/2022', 'Cierre', '
INM.COLONIAL', '7,6800', '0,92', '7,7150', '7,6000', '571.711', '4.387,52', '20/05/2022', 'Cierre', 'MAPFR
E', '1,7730', '-0,11', '1,8090', '1,7730', '4.653.735', '8.316,98', '20/05/2022', 'Cierre', 'MELIA HOTELS
', '6,9650', '0,94', '7,0700', '6,9250', '719.012', '5.023,26', '20/05/2022', 'Cierre', 'MERLIN', '10,3400
', '0,98', '10,5000', '10,1500', '841.403', '8.740,20', '20/05/2022', 'Cierre', 'NATURGY', '27,6800', '0,5
8', '28,0200', '27,5600', '337.127', '9.346,64', '20/05/2022', 'Cierre', 'PHARMA MAR', '70,1000', '1,18',
'70,8000', '69,2400', '48.329', '3.380,44', '20/05/2022', 'Cierre', 'R.E.C.', '19,5250', '0,08', '19,6100
', '19,4050', '921.833', '17.988,95', '20/05/2022', 'Cierre', 'REPSOL', '14,4100', '-0,21', '14,7500', '1
4,4050', '6.414.647', '92.102,10', '20/05/2022', 'Cierre', 'ROVI', '59,9000', '4,08', '60,6000', '58,1500
', '131.909', '7.877,93', '20/05/2022', 'Cierre', 'SIEMENS GAME', '16,7450', '1,12', '16,9300', '16,4550',
'2.112.904', '35.052,29', '20/05/2022', 'Cierre', 'SOLARIA', '21,8600', '-3,70', '22,7600', '21,7100', '61
4.090', '13.580,97', '20/05/2022', 'Cierre', 'TELEFONICA', '4,6920', '1,10', '4,7080', '4,6310', '10.879.2
03', '50.798,39', '20/05/2022', 'Cierre']

```

Es pot comprovar que al llistat "files" tenim un canvi cada 9 dades, que equival a les 9 columnes de l'índex. Ara hem d'extreure aquesta informació per crear un dataframe de 9 columnes (atributs), i afegir l'índex a posteriori.

```
dades={}
for i in range(9):
    dades[i]=files[i::9]

print(dades)
```

[illegible]

```
In [14]: #Creem el dataframe i li afegim el títol de les columnes
ibex_df = pd.DataFrame(dades)
ibex_df.columns = index_col
```

```
In [15]: ibex_df
```

```
Out[15]:
```

|    | Nombre       | Últ.     | % Dif. | Máx.     | Mín.     | Volumen    | Efectivo (miles €) | Fecha      | Hora   |
|----|--------------|----------|--------|----------|----------|------------|--------------------|------------|--------|
| 0  | ACCIONA      | 183,1000 | 0,83   | 185,3000 | 182,9000 | 89.921     | 16.536,77          | 20/05/2022 | Cierre |
| 1  | ACERINOX     | 10,6500  | -0,70  | 10,9500  | 10,6250  | 877.025    | 9.480,32           | 20/05/2022 | Cierre |
| 2  | ACS          | 25,4300  | 1,07   | 25,8400  | 25,3300  | 624.557    | 15.948,53          | 20/05/2022 | Cierre |
| 3  | AENA         | 135,8000 | 1,46   | 136,3000 | 133,4000 | 100.816    | 14.174,97          | 20/05/2022 | Cierre |
| 4  | ALMIRALL     | 10,5800  | 1,73   | 10,6000  | 10,3700  | 346.430    | 3.642,88           | 20/05/2022 | Cierre |
| 5  | AMADEUS      | 59,1000  | 0,96   | 60,0400  | 58,7600  | 516.071    | 30.610,30          | 20/05/2022 | Cierre |
| 6  | ARCELORMIT.  | 27,4000  | 1,03   | 28,0450  | 27,3300  | 615.965    | 17.103,20          | 20/05/2022 | Cierre |
| 7  | B.SANTANDER  | 2,7440   | 0,13   | 2,7905   | 2,7320   | 40.067.111 | 110.999,29         | 20/05/2022 | Cierre |
| 8  | BA.SABADELL  | 0,7314   | -0,57  | 0,7504   | 0,7314   | 22.843.011 | 16.918,40          | 20/05/2022 | Cierre |
| 9  | BANKINTER    | 5,5640   | -0,68  | 5,6780   | 5,5560   | 2.052.978  | 11.519,40          | 20/05/2022 | Cierre |
| 10 | BBVA         | 4,7475   | -0,40  | 4,8285   | 4,7355   | 24.941.682 | 119.088,61         | 20/05/2022 | Cierre |
| 11 | CAIXABANK    | 3,1200   | -0,16  | 3,1840   | 3,1090   | 18.086.447 | 56.753,73          | 20/05/2022 | Cierre |
| 12 | CELLNEX      | 40,4200  | 2,20   | 40,8900  | 39,4100  | 934.479    | 37.760,74          | 20/05/2022 | Cierre |
| 13 | CIE AUTOMOT. | 21,3000  | 0,95   | 21,7000  | 21,1400  | 258.060    | 5.525,16           | 20/05/2022 | Cierre |
| 14 | ENAGAS       | 21,4700  | 0,28   | 21,6300  | 21,2900  | 831.194    | 17.823,64          | 20/05/2022 | Cierre |
| 15 | ENDESA       | 20,6700  | 0,05   | 20,9200  | 20,5800  | 902.924    | 18.693,33          | 20/05/2022 | Cierre |
| 16 | FERROVIAL    | 23,3100  | 0,56   | 23,6400  | 23,1100  | 791.476    | 18.479,13          | 20/05/2022 | Cierre |
| 17 | FLUIDRA      | 22,7800  | 1,24   | 23,2400  | 22,5400  | 310.475    | 7.128,03           | 20/05/2022 | Cierre |
| 18 | GRIFOLS CL.A | 19,8800  | -0,15  | 20,1800  | 19,6800  | 779.006    | 15.476,23          | 20/05/2022 | Cierre |
| 19 | IAG          | 1,4790   | 2,21   | 1,5015   | 1,4540   | 14.614.243 | 21.631,29          | 20/05/2022 | Cierre |

|    | Nombre       | Últ.    | % Dif. | Máx.    | Mín.    | Volumen    | Efectivo (miles €) | Fecha      | Hora   |
|----|--------------|---------|--------|---------|---------|------------|--------------------|------------|--------|
| 20 | IBERDROLA    | 11,1000 | 3,11   | 11,1800 | 10,8600 | 22.219.036 | 241.268,15         | 20/05/2022 | Cierre |
| 21 | INDITEX      | 20,4200 | 0,54   | 20,7100 | 20,0600 | 3.627.429  | 74.407,17          | 20/05/2022 | Cierre |
| 22 | INDRA A      | 9,6150  | 1,10   | 9,7700  | 9,4950  | 629.913    | 6.081,13           | 20/05/2022 | Cierre |
| 23 | INM.COLONIAL | 7,6800  | 0,92   | 7,7150  | 7,6000  | 571.711    | 4.387,52           | 20/05/2022 | Cierre |
| 24 | MAPFRE       | 1,7730  | -0,11  | 1,8090  | 1,7730  | 4.653.735  | 8.316,98           | 20/05/2022 | Cierre |
| 25 | MELIA HOTELS | 6,9650  | 0,94   | 7,0700  | 6,9250  | 719.012    | 5.023,26           | 20/05/2022 | Cierre |
| 26 | MERLIN       | 10,3400 | 0,98   | 10,5000 | 10,1500 | 841.403    | 8.740,20           | 20/05/2022 | Cierre |
| 27 | NATURGY      | 27,6800 | 0,58   | 28,0200 | 27,5600 | 337.127    | 9.346,64           | 20/05/2022 | Cierre |
| 28 | PHARMA MAR   | 70,1000 | 1,18   | 70,8000 | 69,2400 | 48.329     | 3.380,44           | 20/05/2022 | Cierre |
| 29 | R.E.C.       | 19,5250 | 0,08   | 19,6100 | 19,4050 | 921.833    | 17.988,95          | 20/05/2022 | Cierre |
| 30 | REPSOL       | 14,4100 | -0,21  | 14,7500 | 14,4050 | 6.414.647  | 92.102,10          | 20/05/2022 | Cierre |
| 31 | ROVI         | 59,9000 | 4,08   | 60,6000 | 58,1500 | 131.909    | 7.877,93           | 20/05/2022 | Cierre |
| 32 | SIEMENS GAME | 16,7450 | 1,12   | 16,9300 | 16,4550 | 2.112.904  | 35.052,29          | 20/05/2022 | Cierre |
| 33 | SOLARIA      | 21,8600 | -3,70  | 22,7600 | 21,7100 | 614.090    | 13.580,97          | 20/05/2022 | Cierre |

```
In [16]: #Guardem el resultat
ibex_df.to_csv("//home/rusi/Escritorio/rubenIT/DataSources/ibex35_20220222.csv")
```

## Web scraping amb Selenium

```
In [17]: #comprovo la versió Selenium instal·lada.
import selenium
selenium.__version__
```

```
Out[17]: '4.1.5'
```

```
In [18]: opts = options()
opts.add_argument("--headless")
opts.headless # Operating in headless mode
#assert opts.add_argument("--headless")
```

```
browser = Firefox(options=opts)
browser = Firefox(executable_path = '/usr/local/bin/geckodriver')
url = 'https://www.bolsamadrid.es'
browser.get(url)
```

-----  
WebDriverException Traceback (most recent call last)

/tmp/ipykernel\_25892/1324337208.py in <module>

3 opts.headless # Operating in headless mode

4 #assert opts.add\_argument("--headless")

----> 5 browser = Firefox(options=opts)

6 browser = Firefox(executable\_path = '/usr/local/bin/geckodriver')

7 url = 'https://www.bolsamadrid.es'

~/anaconda3/lib/python3.9/site-packages/selenium/webdriver/firefox/webdriver.py in \_\_init\_\_(self, firefox\_profile, firefox\_binary, capabilities, proxy, executable\_path, options, service\_log\_path, service\_args, service, desired\_capabilities, log\_path, keep\_alive)

173 service\_args=service\_args,

174 log\_path=service\_log\_path)

--> 175 self.service.start()

176

177 executor = FirefoxRemoteConnection(

~/anaconda3/lib/python3.9/site-packages/selenium/webdriver/common/service.py in start(self)

96 count = 0

97 while True:

---> 98 self.assert\_process\_still\_running()

99 if self.is\_connectable():

100 break

~/anaconda3/lib/python3.9/site-packages/selenium/webdriver/common/service.py in assert\_process\_still\_running(self)

108 return\_code = self.process.poll()

109 if return\_code:

--> 110 raise WebDriverException(

111 'Service %s unexpectedly exited. Status code was: %s'

112 % (self.path, return\_code))

WebDriverException: Message: Service geckodriver unexpectedly exited. Status code was: 1



Després de molts intents, no puc fer servir Selenium, em dona un error. Tanmateix, trobo finalment la **solució** a:

[https://www.youtube.com/watch?v=l\\_4Tmb2arVg](https://www.youtube.com/watch?v=l_4Tmb2arVg) ([https://www.youtube.com/watch?v=l\\_4Tmb2arVg](https://www.youtube.com/watch?v=l_4Tmb2arVg))

<https://pythonshowcase.com/question/webdriverexception-selenium-common-exceptions-webdriverexception-message-service-webdrivers-geckodriver-unexpectedly-exited-status-code-was-1> (<https://pythonshowcase.com/question/webdriverexception-selenium-common-exceptions-webdriverexception-message-service-webdrivers-geckodriver-unexpectedly-exited-status-code-was-1>)

```
In [19]: driver_path = '/usr/local/bin/geckodriver'
driver = Firefox(service=Service(driver_path))
driver.get("https://www.bolsamadrid.es")
```

El menú "Acciones" es troba a:

```
<div class="seg"><a href="/?id=esp" target="_self">Inicio</a> / <a href="#" target="_self">MERCADOS
Y COTIZACIONES</a> / <a href="/esp/asp/Mercados/Precios.aspx?indice=ESI100000000" target="_self">A
cciones</a> / <span class="segUlt">Precios de la sesión &gt;</span></div>
```

```
In [1]: ###Exemple de la solució a l'error. No l'executem, el deixem a mode de consulta
#elem = driver.find_element_by_link_text('Acciones')
#elem.send_keys("pycon")
#elem.send_keys(Keys.RETURN)
#assert "No results found." not in driver.page_source
#driver.close()
```

**Molt important: cal clicar al link del navegador-selenium abans de seguir els passos!!**

```
In [20]: # busquem el link on tenim el string "Acciones" i cliquem
links = driver.find_elements_by_link_text('Acciones')
links[0].click
```

```
/tmp/ipykernel_25892/1063298664.py:2: DeprecationWarning: find_elements_by_link_text is deprecated. Please
use find_elements(by=By.LINK_TEXT, value=text) instead
  links = driver.find_elements_by_link_text('Acciones')
```

Out[20]:

<bound method WebElement.click of <selenium.webdriver.remote.webelement.WebElement (session="b9e2f365-e4bf

```
In [21]: # un cop a la nova finestra busquem els valors de la taula
results = driver.find_element_by_id('ctl00_Contenido_tblAcciones')
```

/tmp/ipykernel\_25892/180685495.py:2: DeprecationWarning: find\_element\_by\_\* commands are deprecated. Please use find\_element() instead  
results = driver.find\_element\_by\_id('ctl00\_Contenido\_tblAcciones')

```
In [22]: #Extraiem els títols de les columnes
index_col=[]

for element in results.find_elements_by_tag_name("th"):
    index_col.append(element.text)

print(index_col)
```

['Nombre', 'Últ.', '% Dif.', 'Máx.', 'Mín.', 'Volumen', 'Efectivo (miles €)', 'Fecha', 'Hora']

/tmp/ipykernel\_25892/690601767.py:4: DeprecationWarning: find\_elements\_by\_tag\_name is deprecated. Please use find\_elements(by=By.TAG\_NAME, value=name) instead  
for element in results.find\_elements\_by\_tag\_name("th"):

```
In [24]: files = []
for element in results.find_elements_by_tag_name("td"):
    files.append(element.text)

print(files)
```

/tmp/ipykernel\_25892/3486389068.py:2: DeprecationWarning: find\_elements\_by\_tag\_name is deprecated. Please use find\_elements(by=By.TAG\_NAME, value=name) instead  
for element in results.find\_elements\_by\_tag\_name("td"):

```
[ 'ACCIONA', '183,1000', '0,83', '185,3000', '182,9000', '89.921', '16.536,77', '20/05/2022', 'Cierre', 'AC
ERINOX', '10,6500', '-0,70', '10,9500', '10,6250', '877.025', '9.480,32', '20/05/2022', 'Cierre', 'ACS', '
25,4300', '1,07', '25,8400', '25,3300', '624.557', '15.948,53', '20/05/2022', 'Cierre', 'AENA', '135,8000
', '1,46', '136,3000', '133,4000', '100.816', '14.174,97', '20/05/2022', 'Cierre', 'ALMIRALL', '10,5800',
'1,73', '10,6000', '10,3700', '346.430', '3.642,88', '20/05/2022', 'Cierre', 'AMADEUS', '59,1000', '0,96',
'60,0400', '58,7600', '516.071', '30.610,30', '20/05/2022', 'Cierre', 'ARCELORMIT.', '27,4000', '1,03', '2
8,0450', '27,3300', '615.965', '17.103,20', '20/05/2022', 'Cierre', 'B.SANTANDER', '2,7440', '0,13', '2,79
05', '2,7320', '40.067.111', '110.999,29', '20/05/2022', 'Cierre', 'BA.SABADELL', '0,7314', '-0,57', '0,75
04', '0,7314', '22.843.011', '16.918,40', '20/05/2022', 'Cierre', 'BANKINTER', '5,5640', '-0,68', '5,6780
', '5,5560', '2.052.978', '11.519,40', '20/05/2022', 'Cierre', 'BBVA', '4,7475', '-0,40', '4,8285', '4,735
5', '24.941.682', '119.088,61', '20/05/2022', 'Cierre', 'CAIXABANK', '3,1200', '-0,16', '3,1840', '3,1090
', '18.086.447', '56.753,73', '20/05/2022', 'Cierre', 'CELLNEX', '40,4200', '2,20', '40,8900', '39,4100',
'934.479', '37.760,74', '20/05/2022', 'Cierre', 'CIE AUTOMOT.', '21,3000', '0,95', '21,7000', '21,1400', '
258.060', '5.525,16', '20/05/2022', 'Cierre', 'ENAGAS', '21,4700', '0,28', '21,6300', '21,2900', '831.194
', '17.823,64', '20/05/2022', 'Cierre', 'ENDESA', '20,6700', '0,05', '20,9200', '20,5800', '902.924', '18.
693,33', '20/05/2022', 'Cierre', 'FERROVIAL', '23,3100', '0,56', '23,6400', '23,1100', '791.476', '18.479,
13', '20/05/2022', 'Cierre', 'FLUIDRA', '22,7800', '1,24', '23,2400', '22,5400', '310.475', '7.128,03', '2
0/05/2022', 'Cierre', 'GRIFOLS CL.A', '19,8800', '-0,15', '20,1800', '19,6800', '779.006', '15.476,23', '2
0/05/2022', 'Cierre', 'IAG', '1,4790', '2,21', '1,5015', '1,4540', '14.614.243', '21.631,29', '20/05/2022
', 'Cierre', 'IBERDROLA', '11,1000', '3,11', '11,1800', '10,8600', '22.219.036', '241.268,15', '20/05/2022
', 'Cierre', 'INDITEX', '20,4200', '0,54', '20,7100', '20,0600', '3.627.429', '74.407,17', '20/05/2022', '
Cierre', 'INDRA A', '9,6150', '1,10', '9,7700', '9,4950', '629.913', '6.081,13', '20/05/2022', 'Cierre', '
INM.COLONIAL', '7,6800', '0,92', '7,7150', '7,6000', '571.711', '4.387,52', '20/05/2022', 'Cierre', 'MAPFR
E', '1,7730', '-0,11', '1,8090', '1,7730', '4.653.735', '8.316,98', '20/05/2022', 'Cierre', 'MELIA HOTELS
', '6,9650', '0,94', '7,0700', '6,9250', '719.012', '5.023,26', '20/05/2022', 'Cierre', 'MERLIN', '10,3400
', '0,98', '10.5000', '10.1500', '841.403', '8.740,20', '20/05/2022', 'Cierre', 'NATURGY', '27.6800', '0.5
```

```
In [25]: dades={}
         for i in range(9):
             dades[i]=files[i::9]

         print(dades)
```

```
{0: ['ACCIONA', 'ACERINOX', 'ACS', 'AENA', 'ALMIRALL', 'AMADEUS', 'ARCELORMIT.', 'B.SANTANDER', 'BA.SABADELL', 'BANKINTER', 'BBVA', 'CAIXABANK', 'CELLNEX', 'CIE AUTOMOT.', 'ENAGAS', 'ENDESA', 'FERROVIAL', 'FLUIDRA', 'GRIFOLS CL.A', 'IAG', 'IBERDROLA', 'INDITEX', 'INDRA A', 'INM.COLONIAL', 'MAPFRE', 'MELIA HOTELS', 'MERLIN', 'NATURGY', 'PHARMA MAR', 'R.E.C.', 'REPSOL', 'ROVI', 'SIEMENS GAME', 'SOLARIA', 'TELEFONICA'], 1: ['183,1000', '10,6500', '25,4300', '135,8000', '10,5800', '59,1000', '27,4000', '2,7440', '0,7314', '5,5640', '4,7475', '3,1200', '40,4200', '21,3000', '21,4700', '20,6700', '23,3100', '22,7800', '19,8800', '1,4790', '11,1000', '20,4200', '9,6150', '7,6800', '1,7730', '6,9650', '10,3400', '27,6800', '70,1000', '19,5250', '14,4100', '59,9000', '16,7450', '21,8600', '4,6920'], 2: ['0,83', '-0,70', '1,07', '1,46', '1,73', '0,96', '1,03', '0,13', '-0,57', '-0,68', '-0,40', '-0,16', '2,20', '0,95', '0,28', '0,05', '0,56', '1,24', '-0,15', '2,21', '3,11', '0,54', '1,10', '0,92', '-0,11', '0,94', '0,98', '0,58', '1,18', '0,08', '-0,21', '4,08', '1,12', '-3,70', '1,10'], 3: ['185,3000', '10,9500', '25,8400', '136,3000', '10,6000', '60,0400', '28,0450', '2,7905', '0,7504', '5,6780', '4,8285', '3,1840', '40,8900', '21,7000', '21,6300', '20,9200', '23,6400', '23,2400', '20,1800', '1,5015', '11,1800', '20,7100', '9,7700', '7,7150', '1,8090', '7,0700', '10,5000', '28,0200', '70,8000', '19,6100', '14,7500', '60,6000', '16,9300', '22,7600', '4,7080'], 4: ['182,9000', '10,6250', '25,3300', '133,4000', '10,3700', '58,7600', '27,3300', '2,7320', '0,7314', '5,5560', '4,7355', '3,1090', '39,4100', '21,1400', '21,2900', '20,5800', '23,1100', '22,5400', '19,6800', '1,4540', '10,8600', '20,0600', '9,4950', '7,6000', '1,7730', '6,9250', '10,1500', '27,5600', '69,2400', '19,4050', '14,4050', '58,1500', '16,4550', '21,7100', '4,6310'], 5: ['89.921', '877.025', '624.557', '100.816', '346.430', '516.071', '615.965', '40.067.111', '22.843.011', '2.052.978', '24.941.682', '18.086.447', '934.479', '258.060', '831.194', '902.924', '791.476', '310.475', '779.006', '14.614.243', '22.219.036', '3.627.429', '629.913', '571.711', '4.653.735', '719.012', '841.403', '337.127', '48.329', '921.833', '6.414.647', '131.909', '2.112.904', '614.090', '10.879.203'], 6: ['16.536,77', '9.480,32', '15.948,53', '14.174,97', '3.642,88', '30.610,30', '17.103,20', '110.999,29', '16.918,40', '11.519,40', '119.088,61', '56.753,73', '37.760,74', '5.525,16', '17.823,64', '18.693,33', '18.479,13', '7.128,03', '15.476,23', '21.631,29', '241.268,15', '74.407,17', '6.081,13', '4.387,52', '8.316,98', '5.023,26', '8.740,20', '9.346,64', '3.380,44', '17.988,95', '92.102,10', '7.877,93', '35.052,29', '13.580,97', '50.798,39'], 7: ['20/05/2022', '20/05/2022
```

In [26]: *#Creem el dataframe i li afegim el títol de les columnes*

```
ibex_df = pd.DataFrame(dades)
ibex_df.columns = index_col
```

In [27]: ibex\_df

Out[27]:

|   | Nombre   | Últ.     | % Dif. | Máx.     | Mín.     | Volumen | Efectivo (miles €) | Fecha      | Hora   |
|---|----------|----------|--------|----------|----------|---------|--------------------|------------|--------|
| 0 | ACCIONA  | 183,1000 | 0,83   | 185,3000 | 182,9000 | 89.921  | 16.536,77          | 20/05/2022 | Cierre |
| 1 | ACERINOX | 10,6500  | -0,70  | 10,9500  | 10,6250  | 877.025 | 9.480,32           | 20/05/2022 | Cierre |
| 2 | ACS      | 25,4300  | 1,07   | 25,8400  | 25,3300  | 624.557 | 15.948,53          | 20/05/2022 | Cierre |
| 3 | AENA     | 135,8000 | 1,46   | 136,3000 | 133,4000 | 100.816 | 14.174,97          | 20/05/2022 | Cierre |

|    | Nombre       | Últ.    | % Dif. | Máx.    | Mín.    | Volumen    | Efectivo (miles €) | Fecha      | Hora   |
|----|--------------|---------|--------|---------|---------|------------|--------------------|------------|--------|
| 4  | ALMIRALL     | 10,5800 | 1,73   | 10,6000 | 10,3700 | 346.430    | 3.642,88           | 20/05/2022 | Cierre |
| 5  | AMADEUS      | 59,1000 | 0,96   | 60,0400 | 58,7600 | 516.071    | 30.610,30          | 20/05/2022 | Cierre |
| 6  | ARCELORMIT.  | 27,4000 | 1,03   | 28,0450 | 27,3300 | 615.965    | 17.103,20          | 20/05/2022 | Cierre |
| 7  | B.SANTANDER  | 2,7440  | 0,13   | 2,7905  | 2,7320  | 40.067.111 | 110.999,29         | 20/05/2022 | Cierre |
| 8  | BA.SABADELL  | 0,7314  | -0,57  | 0,7504  | 0,7314  | 22.843.011 | 16.918,40          | 20/05/2022 | Cierre |
| 9  | BANKINTER    | 5,5640  | -0,68  | 5,6780  | 5,5560  | 2.052.978  | 11.519,40          | 20/05/2022 | Cierre |
| 10 | BBVA         | 4,7475  | -0,40  | 4,8285  | 4,7355  | 24.941.682 | 119.088,61         | 20/05/2022 | Cierre |
| 11 | CAIXABANK    | 3,1200  | -0,16  | 3,1840  | 3,1090  | 18.086.447 | 56.753,73          | 20/05/2022 | Cierre |
| 12 | CELLNEX      | 40,4200 | 2,20   | 40,8900 | 39,4100 | 934.479    | 37.760,74          | 20/05/2022 | Cierre |
| 13 | CIE AUTOMOT. | 21,3000 | 0,95   | 21,7000 | 21,1400 | 258.060    | 5.525,16           | 20/05/2022 | Cierre |
| 14 | ENAGAS       | 21,4700 | 0,28   | 21,6300 | 21,2900 | 831.194    | 17.823,64          | 20/05/2022 | Cierre |
| 15 | ENDESA       | 20,6700 | 0,05   | 20,9200 | 20,5800 | 902.924    | 18.693,33          | 20/05/2022 | Cierre |
| 16 | FERROVIAL    | 23,3100 | 0,56   | 23,6400 | 23,1100 | 791.476    | 18.479,13          | 20/05/2022 | Cierre |
| 17 | FLUIDRA      | 22,7800 | 1,24   | 23,2400 | 22,5400 | 310.475    | 7.128,03           | 20/05/2022 | Cierre |
| 18 | GRIFOLS CL.A | 19,8800 | -0,15  | 20,1800 | 19,6800 | 779.006    | 15.476,23          | 20/05/2022 | Cierre |
| 19 | IAG          | 1,4790  | 2,21   | 1,5015  | 1,4540  | 14.614.243 | 21.631,29          | 20/05/2022 | Cierre |
| 20 | IBERDROLA    | 11,1000 | 3,11   | 11,1800 | 10,8600 | 22.219.036 | 241.268,15         | 20/05/2022 | Cierre |
| 21 | INDITEX      | 20,4200 | 0,54   | 20,7100 | 20,0600 | 3.627.429  | 74.407,17          | 20/05/2022 | Cierre |
| 22 | INDRA A      | 9,6150  | 1,10   | 9,7700  | 9,4950  | 629.913    | 6.081,13           | 20/05/2022 | Cierre |
| 23 | INM.COLONIAL | 7,6800  | 0,92   | 7,7150  | 7,6000  | 571.711    | 4.387,52           | 20/05/2022 | Cierre |
| 24 | MAPFRE       | 1,7730  | -0,11  | 1,8090  | 1,7730  | 4.653.735  | 8.316,98           | 20/05/2022 | Cierre |
| 25 | MELIA HOTELS | 6,9650  | 0,94   | 7,0700  | 6,9250  | 719.012    | 5.023,26           | 20/05/2022 | Cierre |
| 26 | MERLIN       | 10,3400 | 0,98   | 10,5000 | 10,1500 | 841.403    | 8.740,20           | 20/05/2022 | Cierre |
| 27 | NATURGY      | 27,6800 | 0,58   | 28,0200 | 27,5600 | 337.127    | 9.346,64           | 20/05/2022 | Cierre |
| 28 | PHARMA MAR   | 70,1000 | 1,18   | 70,8000 | 69,2400 | 48.329     | 3.380,44           | 20/05/2022 | Cierre |

|    | Nombre       | Últ.    | % Dif. | Máx.    | Mín.    | Volumen   | Efectivo (miles €) | Fecha      | Hora   |
|----|--------------|---------|--------|---------|---------|-----------|--------------------|------------|--------|
| 29 | R.E.C.       | 19,5250 | 0,08   | 19,6100 | 19,4050 | 921.833   | 17.988,95          | 20/05/2022 | Cierre |
| 30 | REPSOL       | 14,4100 | -0,21  | 14,7500 | 14,4050 | 6.414.647 | 92.102,10          | 20/05/2022 | Cierre |
| 31 | ROVI         | 59,9000 | 4,08   | 60,6000 | 58,1500 | 131.909   | 7.877,93           | 20/05/2022 | Cierre |
| 32 | SIEMENS GAME | 16,7450 | 1,12   | 16,9300 | 16,4550 | 2.112.904 | 35.052,29          | 20/05/2022 | Cierre |

```
In [28]: #Guardem el resultat
ibex_df.to_csv("//home/rusi/Escritorio/rubenIT/DataSources/ibex35_20220222_selenium.csv")
```

## Exercici 2. Documenta en un word el teu conjunt de dades generat amb la informació que tenen els diferents arxius de Kaggle.

```
In [29]: ibex_df02=pd.read_csv("//home/rusi/Escritorio/rubenIT/DataSources/ibex35_20220222.csv")
```

```
In [30]: ibex_df02.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            35 non-null    int64
1   Nombre                35 non-null    object
2   Últ.                  35 non-null    object
3   % Dif.                35 non-null    object
4   Máx.                  35 non-null    object
5   Mín.                  35 non-null    object
6   Volumen                35 non-null    object
7   Efectivo (miles €)    35 non-null    object
8   Fecha                 35 non-null    object
9   Hora                  35 non-null    object
dtypes: int64(1), object(9)
memory usage: 2.9+ KB
```

In [31]: `ibex_df02.head()`

Out[31]:

|   | Unnamed: 0 | Nombre   | Últ.     | % Dif. | Máx.     | Mín.     | Volumen | Efectivo (miles €) | Fecha      | Hora   |
|---|------------|----------|----------|--------|----------|----------|---------|--------------------|------------|--------|
| 0 | 0          | ACCIONA  | 183,1000 | 0,83   | 185,3000 | 182,9000 | 89.921  | 16.536,77          | 20/05/2022 | Cierre |
| 1 | 1          | ACERINOX | 10,6500  | -0,70  | 10,9500  | 10,6250  | 877.025 | 9.480,32           | 20/05/2022 | Cierre |
| 2 | 2          | ACS      | 25,4300  | 1,07   | 25,8400  | 25,3300  | 624.557 | 15.948,53          | 20/05/2022 | Cierre |
| 3 | 3          | AENA     | 135,8000 | 1,46   | 136,3000 | 133,4000 | 100.816 | 14.174,97          | 20/05/2022 | Cierre |
| 4 | 4          | ALMIRALL | 10,5800  | 1,73   | 10,6000  | 10,3700  | 346.430 | 3.642,88           | 20/05/2022 | Cierre |

In [32]:

```

ibex_df02['Últ.'] =ibex_df02['Últ.'].str.replace(',','').astype(float)
ibex_df02['% Dif.'] =ibex_df02['% Dif.'].str.replace(',','').astype(float)
ibex_df02['Máx.'] =ibex_df02['Máx.'].str.replace(',','').astype(float)
ibex_df02['Mín.'] =ibex_df02['Mín.'].str.replace(',','').astype(float)
ibex_df02['Efectivo (miles €)'] =ibex_df02['Efectivo (miles €)'].str.replace('.', '').str.replace(',','').as

```

/tmp/ipykernel\_25892/3383750530.py:5: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will \*not\* be treated as literal strings when regex=True.

```

ibex_df02['Efectivo (miles €)'] =ibex_df02['Efectivo (miles €)'].str.replace('.', '').str.replace(
',','').astype(float)

```

In [33]: `ibex_df03=ibex_df02.drop("Unnamed: 0",axis=1)`

In [34]: `ibex_df03`

Out[34]:

|   | Nombre   | Últ.     | % Dif. | Máx.     | Mín.     | Volumen | Efectivo (miles €) | Fecha      | Hora   |
|---|----------|----------|--------|----------|----------|---------|--------------------|------------|--------|
| 0 | ACCIONA  | 183.1000 | 0.83   | 185.3000 | 182.9000 | 89.921  | 16536.77           | 20/05/2022 | Cierre |
| 1 | ACERINOX | 10.6500  | -0.70  | 10.9500  | 10.6250  | 877.025 | 9480.32            | 20/05/2022 | Cierre |
| 2 | ACS      | 25.4300  | 1.07   | 25.8400  | 25.3300  | 624.557 | 15948.53           | 20/05/2022 | Cierre |
| 3 | AENA     | 135.8000 | 1.46   | 136.3000 | 133.4000 | 100.816 | 14174.97           | 20/05/2022 | Cierre |
| 4 | ALMIRALL | 10.5800  | 1.73   | 10.6000  | 10.3700  | 346.430 | 3642.88            | 20/05/2022 | Cierre |
| 5 | AMADEUS  | 59.1000  | 0.96   | 60.0400  | 58.7600  | 516.071 | 30610.30           | 20/05/2022 | Cierre |

|    | Nombre       | Últ.    | % Dif. | Máx.    | Mín.    | Volumen    | Efectivo (miles €) | Fecha      | Hora   |
|----|--------------|---------|--------|---------|---------|------------|--------------------|------------|--------|
| 6  | ARCELORMIT.  | 27.4000 | 1.03   | 28.0450 | 27.3300 | 615.965    | 17103.20           | 20/05/2022 | Cierre |
| 7  | B.SANTANDER  | 2.7440  | 0.13   | 2.7905  | 2.7320  | 40.067.111 | 110999.29          | 20/05/2022 | Cierre |
| 8  | BA.SABADELL  | 0.7314  | -0.57  | 0.7504  | 0.7314  | 22.843.011 | 16918.40           | 20/05/2022 | Cierre |
| 9  | BANKINTER    | 5.5640  | -0.68  | 5.6780  | 5.5560  | 2.052.978  | 11519.40           | 20/05/2022 | Cierre |
| 10 | BBVA         | 4.7475  | -0.40  | 4.8285  | 4.7355  | 24.941.682 | 119088.61          | 20/05/2022 | Cierre |
| 11 | CAIXABANK    | 3.1200  | -0.16  | 3.1840  | 3.1090  | 18.086.447 | 56753.73           | 20/05/2022 | Cierre |
| 12 | CELLNEX      | 40.4200 | 2.20   | 40.8900 | 39.4100 | 934.479    | 37760.74           | 20/05/2022 | Cierre |
| 13 | CIE AUTOMOT. | 21.3000 | 0.95   | 21.7000 | 21.1400 | 258.060    | 5525.16            | 20/05/2022 | Cierre |
| 14 | ENAGAS       | 21.4700 | 0.28   | 21.6300 | 21.2900 | 831.194    | 17823.64           | 20/05/2022 | Cierre |
| 15 | ENDESA       | 20.6700 | 0.05   | 20.9200 | 20.5800 | 902.924    | 18693.33           | 20/05/2022 | Cierre |
| 16 | FERROVIAL    | 23.3100 | 0.56   | 23.6400 | 23.1100 | 791.476    | 18479.13           | 20/05/2022 | Cierre |
| 17 | FLUIDRA      | 22.7800 | 1.24   | 23.2400 | 22.5400 | 310.475    | 7128.03            | 20/05/2022 | Cierre |
| 18 | GRIFOLS CL.A | 19.8800 | -0.15  | 20.1800 | 19.6800 | 779.006    | 15476.23           | 20/05/2022 | Cierre |
| 19 | IAG          | 1.4790  | 2.21   | 1.5015  | 1.4540  | 14.614.243 | 21631.29           | 20/05/2022 | Cierre |
| 20 | IBERDROLA    | 11.1000 | 3.11   | 11.1800 | 10.8600 | 22.219.036 | 241268.15          | 20/05/2022 | Cierre |
| 21 | INDITEX      | 20.4200 | 0.54   | 20.7100 | 20.0600 | 3.627.429  | 74407.17           | 20/05/2022 | Cierre |
| 22 | INDRA A      | 9.6150  | 1.10   | 9.7700  | 9.4950  | 629.913    | 6081.13            | 20/05/2022 | Cierre |
| 23 | INM.COLONIAL | 7.6800  | 0.92   | 7.7150  | 7.6000  | 571.711    | 4387.52            | 20/05/2022 | Cierre |
| 24 | MAPFRE       | 1.7730  | -0.11  | 1.8090  | 1.7730  | 4.653.735  | 8316.98            | 20/05/2022 | Cierre |
| 25 | MELIA HOTELS | 6.9650  | 0.94   | 7.0700  | 6.9250  | 719.012    | 5023.26            | 20/05/2022 | Cierre |
| 26 | MERLIN       | 10.3400 | 0.98   | 10.5000 | 10.1500 | 841.403    | 8740.20            | 20/05/2022 | Cierre |
| 27 | NATURGY      | 27.6800 | 0.58   | 28.0200 | 27.5600 | 337.127    | 9346.64            | 20/05/2022 | Cierre |
| 28 | PHARMA MAR   | 70.1000 | 1.18   | 70.8000 | 69.2400 | 48.329     | 3380.44            | 20/05/2022 | Cierre |
| 29 | R.E.C.       | 19.5250 | 0.08   | 19.6100 | 19.4050 | 921.833    | 17988.95           | 20/05/2022 | Cierre |
| 30 | REPSOL       | 14.4100 | -0.21  | 14.7500 | 14.4050 | 6.414.647  | 92102.10           | 20/05/2022 | Cierre |



|    | Nombre       | Últ.    | % Dif. | Máx.    | Mín.    | Volumen   | Efectivo (miles €) | Fecha      | Hora   |
|----|--------------|---------|--------|---------|---------|-----------|--------------------|------------|--------|
| 31 | ROVI         | 59.9000 | 4.08   | 60.6000 | 58.1500 | 131.909   | 7877.93            | 20/05/2022 | Cierre |
| 32 | SIEMENS GAME | 16.7450 | 1.12   | 16.9300 | 16.4550 | 2.112.904 | 35052.29           | 20/05/2022 | Cierre |

In [35]: `ibex_df03.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Nombre                35 non-null    object
1   Últ.                  35 non-null    float64
2   % Dif.                35 non-null    float64
3   Máx.                  35 non-null    float64
4   Mín.                  35 non-null    float64
5   Volumen                35 non-null    object
6   Efectivo (miles €)    35 non-null    float64
7   Fecha                 35 non-null    object
8   Hora                  35 non-null    object
dtypes: float64(5), object(4)
memory usage: 2.6+ KB
```

Pugem el csv a Kaggle (és de google, aplicatiu web que no cal instal·lar, i permet visualitzar dades o fer-ho servir de notebook). Ho documentem en un doc de libroffice, tal i com es demana.

### Exercici 3. Tria una página web que tu vulguis i realitza web scraping mitjançant la llibreria Scrapy.

In [36]: `url="https://cadenaser.com/"`  
`page02 = requests.get(url).content`

In [37]: `sel = Selector(text = page02)`

Busquem la class que conté enllaços de notícies, i en concret les de "Deportes". A continuació, un extracte:

```

<div class="mnu" id="mnu"><div class="c-mnu"><a href="/actualidad/" title="Actualidad">Actualidad</a><ul class="n1" id="/actualidad"><li><a href="/actualidad/" title="Actualidad">Actualidad</a></li><li><a href="/actualidad/politica/" title="Pol

id="/deportes/futbol"><button>Volver</button><p><a href="/deportes/futbol/" title="F\xc3\xbatbol">F\xc3\xbatbol</a></p><ul><li><a href="/deportes/futbol/liga-santander/" title="Liga Santander">Liga Santander</a></li><li><a href="/deportes/futbol/liga-smartbank/" title="Liga Smartbank">Liga Smartbank</a></li><li><a href="/deportes/futbol/champions-league/" title="Champions League">Champions League</a></li><li><a href="/deportes/futbol/europa-league/" title="Europa League">Europa League</a></li></ul></div></li><li><a href="/deportes/motociclismo/" title="Motociclismo">Motociclismo</a></li><li><a href="/deportes/formula-1/" title="F\xc3\xb3rmula 1">F\xc3\xb3rmula 1</a></li><li><a href="/deportes/baloncesto/" title="Baloncesto">Baloncesto</a></li><li><a href="/deportes/tenis/" title="Tenis">Tenis</a></li><li><a href="/deportes/mas-deportes/" title="M\xc3\xals deportes">M\xc3\xals deportes</a></li></ul>

```

```

In [40]: # seleccionem la divisió dels continguts principals
links = sel.xpath ('//div[@class="mnu"]//a[@href]').extract()
links

```

Out[40]:

```
['/actualidad/',  
 '/actualidad/',  
 '/actualidad/politica/',  
 '/actualidad/tribunales/',  
 '/actualidad/internacional/',  
 '/actualidad/sociedad/',  
 '/actualidad/economia-y-negocios/',  
 '/actualidad/ciencia-y-tecnologia/',  
 '/deportes/',  
 '/deportes/',  
 '/deportes/futbol/',  
 '/deportes/futbol/liga-santander/',  
 '/deportes/futbol/liga-smartbank/',  
 '/deportes/futbol/champions-league/',  
 '/deportes/futbol/europa-league/',  
 '/deportes/motociclismo/',  
 '/deportes/formula-1/',  
 '/deportes/baloncesto/',  
 '/deportes/tenis/',  
 '/deportes/mas-deportes/',  
 '/humor/',  
 '/ocio-y-cultura/',  
 '/ocio-y-cultura/',  
 '/ocio-y-cultura/ocio/',  
 '/ocio-y-cultura/cultura-y-sociedad/',  
 '/ocio-y-cultura/gastro/',  
 '/ocio-y-cultura/gente/',  
 '/ocio-y-cultura/cine-y-tv/',  
 '/ocio-y-cultura/historia/',  
 '/ocio-y-cultura/musica/',  
 '/ocio-y-cultura/salud-y-bienestar/',  
 '/opinion/',  
 '/ccaa/andalucia/',  
 ..
```

```
In [42]: outcome = []  
for link in links:  
    #busquem aquells links on apareixi el string "deportes"  
    if 'deportes/futbol/' in link:  
        outcome.append(link)  
        break
```

```
In [43]: outcome
```

```
Out[43]: ['/deportes/futbol/']
```

```
In [44]: # anem a la secció "deportes"
page03 = requests.get(url+outcome[0]).content
```

```
In [77]: # parsegem el contingut amb la llibreria Selector de scrapy
sel02 = Selector(text = page03)
```

```
In [78]: sel02
```

```
Out[78]: <Selector xpath=None data='<html lang="es"><head><meta charset="...">
```

Fem un extracte de la informació continguda:

```
<div class="ele-4"
```

```
<div class="s-s">
```

```
<article class="sc t-s">
```

```
<figure>
```

```
<a
```

```
href="/nacional/2022/05/21/la-segunda-division-se-resue
```

```
lve-los-resultados-en-directo-cadena-ser/"
```

```
title="El Almer\x3\xada no logra el ascenso directo tr
```

```
as empatar contra el Alcorc\x3\xb3n en la jornada de la salvaci\x3\xb3n para M\x3\xallaga y Spor
ting"
```

```
>
```

```
In [117]: # volem trobar tots els tags "articles" i el seu text associat
news_headlines = sel02.xpath('//article/div/h2/a/text()').extract()
```

```
# volem trobar tots els tags "articles" i el seu text associat
#news_headlines = sel.xpath('//article/div/h2/a/text()').extract()
```

```
In [118]: news_headlines[:20]
```

```
Out[118]: []
```

```
In [98]: sel02.xpath("//div").extract()
```

```
Out[98]: ['<div id="fusion-app"><script type="application/ld+json">[{"@context":"https://schema.org", "@type":"ItemList", "url":"https://cadenaser.com", "itemListElement":[{"@type":"ListItem", "position":1, "url":"https://cadenaser.com/nacional/2022/05/22/la-lucha-por-el-descenso-en-directo-laliga-se-resuelve-en-carrusel-deportivo-cadena-ser/"}, {"@type":"ListItem", "position":2, "url":"https://cadenaser.com/nacional/2022/05/21/ultima-hora-de-mbappe-y-el-real-madrid-en-directo-cadena-ser/"}, {"@type":"ListItem", "position":3, "url":"https://cadenaser.com/nacional/2022/05/22/me-parece-el-mayor-gatillazo-de-la-historia-del-futbol-cadena-ser/"}, {"@type":"ListItem", "position":4, "url":"https://cadenaser.com/nacional/2022/05/22/ander-herrera-explica-en-carrusel-como-se-entero-el-vestuario-del-psg-de-la-renovacion-de-mbappe-cadena-ser/"}, {"@type":"ListItem", "position":5, "url":"https://cadenaser.com/nacional/2022/05/22/manu-carreno-ni-el-real-madrid-ni-ningun-club-de-futbol-puede-competir-contr-todo-un-estado-como-catar-cadena-ser/"}, {"@type":"ListItem", "position":6, "url":"https://cadenaser.com/nacional/2022/05/21/la-segunda-division-se-resuelve-los-resultados-en-directo-cadena-ser/"}, {"@type":"ListItem", "position":7, "url":"https://cadenaser.com/nacional/2022/05/21/kylian-mbappe-anuncia-su-renovacion-con-el-psg-hasta-2025-cadena-ser/"}, {"@type":"ListItem", "position":8, "url":"https://cadenaser.com/nacional/2022/05/21/asi-fue-el-show-de-la-renovacion-de-mbappe-en-el-parque-de-los-principes-cadena-ser/"}, {"@type":"ListItem", "position":9, "url":"https://cadenaser.com/nacional/2022/05/21/laliga-denunciara-al-psg-ante-la-uefa-tras-la-renovacion-de-mbappe-cadena-ser/"}, {"@type":"ListItem", "position":10, "url":"https://cadenaser.com/nacional/2022/05/21/barcelona-olympique-de-lyon-la-final-de-la-champions-league-en-directo-cadena-ser/"}, {"@type":"ListItem", "position":11, "url":"https://cadenaser.com/nacional/2022/05/21/tebas-lo-que-va-a-hacer-el-psg-renovando-a-mbappe-es-un-insulto-al-futbol-cadena-ser/"}, {"@type":"ListItem", "position":12, "url":"https://cadenaser.com/nacional/2022/05/21/tebas-lo-que-va-a-hacer-el-psg-renovando-a-mbappe-es-un-insulto-al-futbol-cadena-ser/"}], [{"@type":"Image", "url":"https://cadenaser.com/images/2022/05/22/la-lucha-por-el-descenso-en-directo-laliga-se-resuelve-en-carrusel-deportivo-cadena-ser/"}]]
```

```
In [104]: #extraiem informació de div class="ele-4", que pertany a Deportes
sel03=sel02.xpath("//div[@class='ele-4']").extract()
print(sel03)
```

```
[ '<div class="ele-4"><div class="s-s"><article class="sc t-s    "><figure><a href="/nacional/2022/05/21/ul
tima-hora-de-mbappe-y-el-real-madrid-en-directo-cadena-ser/" title="Mbappé, en directo | El jugador se que
da en el PSG hasta 2025"><picture><head><meta charset="utf-8"><meta http-equiv="X-UA-Compatible" content="ie=edge"><meta h
ttp-equiv="Content-Type" content="text/html; charset=utf-8"><meta name="viewport" content="width=device-wi
dth, initial-scale=1"><meta name="referrer" content="no-referrer-when-downgrade"><meta name="lang" content
="es-ES"><meta name="robots" content="index, follow, noarchive"><meta name="robots" content="max-video-pre
view:3"><meta name="robots" content="max-image-preview:large"><meta name="format-detection" content="addre
ss=no;email=no,telephone=no"><meta name="robots" content="max-snippet:-1"><meta name="organization" conten
t="Cadena SER."><meta property="og:image" content="https://cadenaser.com/pf/resources/dist/img/ser-cll.sv
g?d=217"><meta property="og:image:width" content="300"><meta property="og:image:height" content="300"><met
a property="og:type" content="website"><meta property="og:site_name" content="cadena SER"><meta property="
fb:app_id" content="229574677178613"><meta property="fb:pages" content="15658775846,32974073100,1411837632
66,297798423468,260097275665,154038834626728,126239340759913,152716074759665,286241832727,108754747127102
1,397748630349260,260807093963798,218355098183367,192630137538936,209200415772655,9060610628506,1594098194
2404,130512786996530,1930066643884232,767055920005326,965792190174136,1414168792217838,124134164273538,107
895409321804,168883333672674,503744749685810,263858326757,258738666530,319811467873,222909084390267,943306
019036011,534829370024405,175022845914285,597264423810487,105594696642793,426479974076402,417737195051925,
1586904454928885,161976080524709,202089616485001,391992824157969,275214532511758,119523764775530,141266639
267590,198050246911,278472948668,394007950541,1571792933045547,157613800916527,586964561374374,8724060227
2,1766650093650479,1884462045171442,192810794093275,154946021197794,154185008018557,345200755667666,838686
046216716,178334868928834,230508573734771 ,188156367978437,337959822922824,839188389472025 ,18685351468776
5,301327016550101,180577101005735,152206074731276,122521500362562,206124002747207,267172000762,15206252700
```

```
In [114]: #If you want to get all div descendant of this html
sel02.xpath("/html//div").extract()
```

```
Out[114]:
```

```
[ '<div id="fusion-app"><script type="application/ld+json">[{"@context":"https://schema.org", "@type":"ItemList", "url":"https://cadenaser.com", "itemListElement":[{"@type":"ListItem", "position":1, "url":"https://cadenaser.com/nacional/2022/05/22/la-lucha-por-el-descenso-en-directo-laliga-se-resuelve-en-carrusel-deportivo-cadena-ser/"}], [{"@type":"ListItem", "position":2, "url":"https://cadenaser.com/nacional/2022/05/21/ultima-hora-de-mbappe-y-el-real-madrid-en-directo-cadena-ser/"}], [{"@type":"ListItem", "position":3, "url":"https://cadenaser.com/nacional/2022/05/22/me-parece-el-mayor-gatillazo-de-la-historia-del-futbol-cadena-ser/"}], [{"@type":"ListItem", "position":4, "url":"https://cadenaser.com/nacional/2022/05/22/ander-herrera-explica-en-carrusel-como-se-entero-el-vestuario-del-psg-de-la-renovacion-de-mbappe-cadena-ser/"}], [{"@type":"ListItem", "position":5, "url":"https://cadenaser.com/nacional/2022/05/22/manu-carreno-ni-el-real-madrid-ni-ningun-club-de-futbol-puede-competir-contratodo-un-estado-como-catar-cadena-ser/"}], [{"@type":"ListItem", "position":6, "url":"https://cadenaser.com/nacional/2022/05/21/la-segunda-division-se-resuelve-los-resultados-en-directo-cadena-ser/"}], [{"@type":"ListItem", "position":7, "url":"https://cadenaser.com/nacional/2022/05/21/kylian-mbappe-anuncia-su-renovacion-con-el-psg-hasta-2025-cadena-ser/"}], [{"@type":"ListItem", "position":8, "url":"https://cadenaser.com/nacional/2022/05/21/asi-fue-el-show-de-la-renovacion-de-mbappe-en-el-parque-de-los-principales"}]]</script></div>
```

In [ ]: