

**TRABAJO PRÁCTICO n** **º1**

Facultad de Ciencias Económicas Universidad de Buenos Aires

**Big data & Machine learning**

**Código de materia/Curso:** Nro 557/2. Prof. Romero, Noelia

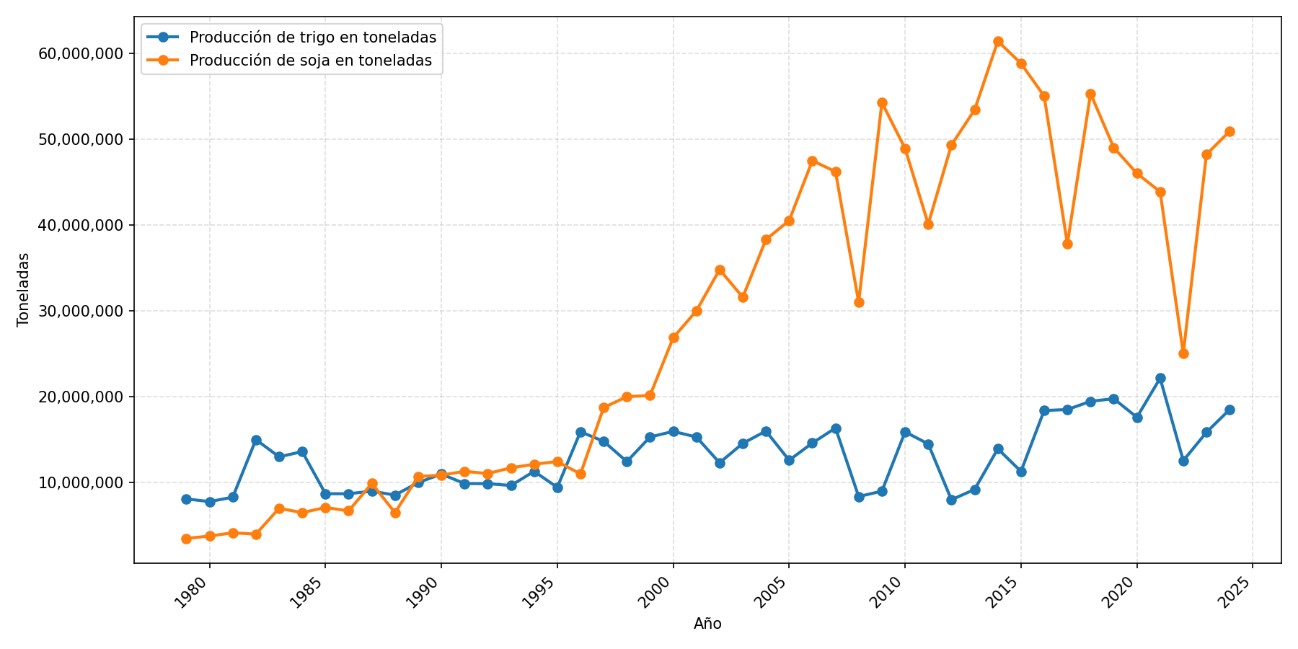
**Integrantes:** Herrero, Facundo (913362), Pougy, Santiago (897575) y da torre, Juan ignacio (910183)

**Jugando con APIs y WebScrapping**

## Parte A

## Ejercicio 1 - Jugando con APIs

Usando la API del Series de Tiempo de la Republica Argentina [link](https://datosgobar.github.io/series-tiempo-ar-api/applications/) , obtener dos series de indicadores del Ministerio de Agricultura, Ganadería y Pesca. En total, dicho Ministerio tiene 72 series anuales y pueden buscar dos series de indicadores de su interés.



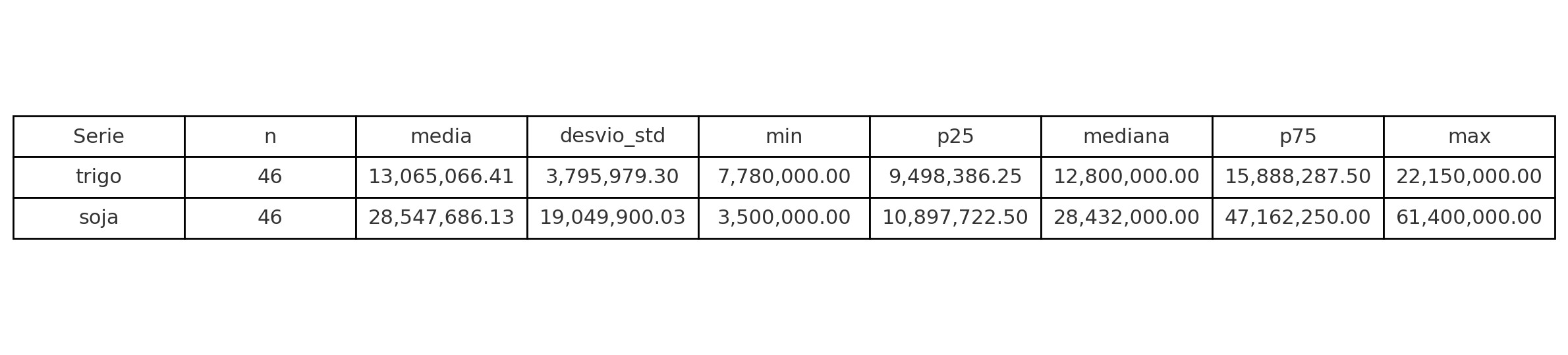
El gráfico muestra la evolución anual (1979–2025) de la producción de trigo

(azul) y soja (naranja) en toneladas. La soja crece con fuerza desde fines de los

’90, alcanza picos >60 Mt a mediados de los 2010, tiene caídas fuertes (2009, 2018 y 2023) y rebota luego hacia ~50 Mt. El trigo se mantiene en niveles menores (8–22 Mt), con tendencia suave al alza y menor volatilidad, con picos cerca de 20–22 Mt alrededor de 2019–2021. En síntesis: la soja creció fuertemente y se consolidó como el principal producto exportador del país, en cambio, el trigo, aunque mantuvo una producción relativamente estable, perdió participación y dejó de ser protagonista desde mediados de los ’90.

## Ejercicio 2 - Repaso de Pandas

Realicen una estadistica descriptiva ambas series de indicadores del Ministerio de Ganaderia y presente en una tabla en su reporte. Dicha tabla debe mostrar: numero de observaciones, media, desvío standard, min, p25, mediana, p75 y max. Comente la tabla en **el reporte**.



La tabla resume 46 años de datos y muestra que la soja produce, en promedio, unas 28,5 Mt, más del doble que el trigo (13,1 Mt). Además, la soja es mucho más volátil: su desvío estándar ronda 19,0 Mt frente a 3,8 Mt del trigo , y su rango intercuartílico es mucho más amplio (36,3 Mt vs 6,4 Mt). En un año típico, la mediana de soja (28,4 Mt) también supera con claridad a la del trigo (12,8 Mt). Los extremos refuerzan esta idea: la soja alcanza picos de 61,4 Mt y pisos de 3,5 Mt, mientras que el trigo se mueve entre 7,8 y 22,2 Mt, evidenciando un comportamiento más estable. En conjunto, la soja creció fuertemente y domina en volumen, pero con marcada variabilidad interanual; el trigo, en cambio, muestra niveles más bajos y un recorrido más parejo.

## Ejercicio 3 - Practicando con Matplotlib

Figura 1. Producción anual de Trigo y Soja en Argentina (1979–2025) – estilo *pyplot*

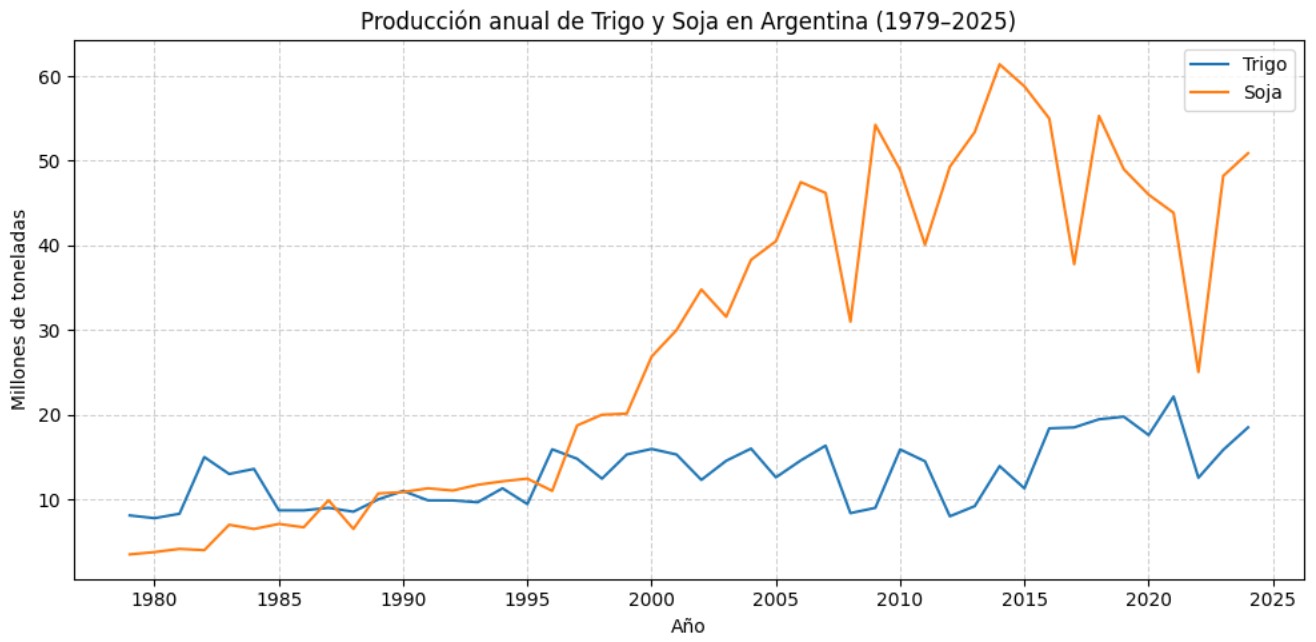
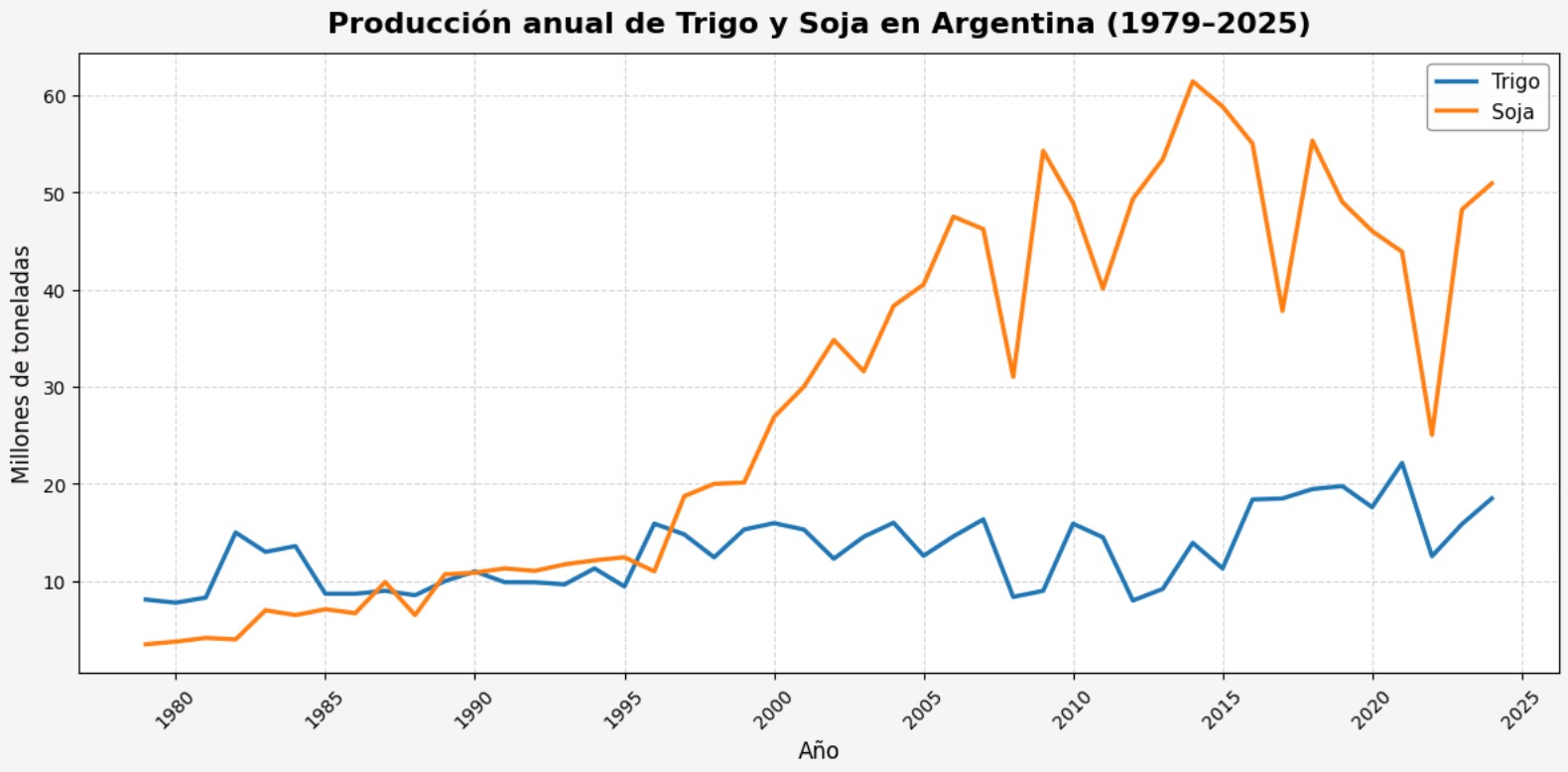


Figura 2. Producción anual de Trigo y Soja en Argentina (1979–2025) – estilo *orientado a objetos*



**Descripción:**

Las figuras muestran la evolución de la producción anual de trigo y soja en Argentina entre 1979 y 2025. La **soja** evidencia un crecimiento sostenido a partir de fines de los años noventa, alcanzando máximos que superan las 60 millones de toneladas y registrando descensos marcados en años críticos como 2009, 2018 y 2023. La **producción de trigo**, en cambio, se mantiene en un rango más acotado —entre 8 y 22 millones de toneladas— con una trayectoria estable y un repunte moderado hacia el período 2019–2021.

En conjunto, los gráficos refuerzan la idea de que la soja se consolidó como el principal cultivo del país, con gran dinamismo pero también con alta volatilidad interanual, mientras que el trigo preserva un recorrido más regular y de menor magnitud relativa.

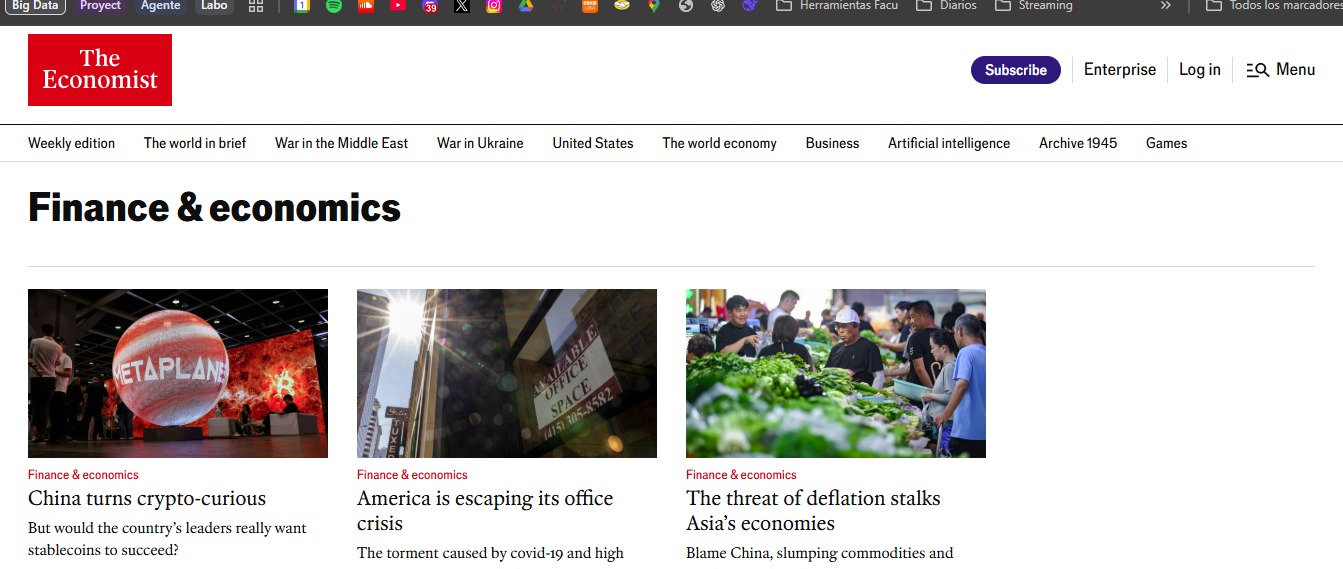
## Parte B

## Ejercicio 4 – Webscrapping

Para el siguiente ejercicio tomamos la decisión de “scrappear” dos diarios económicos. Uno local y otro extranjero. Por un lado, ámbito Financiero y por el otro The economist.



1) Adjuntamos portadas de los diarios al día de la fecha 2/9/25. Con el propósito de contrastar los títulos obtenidos con webscrapping.



Output del código, están las salidas (dataframes) tanto de manera concatenada como por separado:

|  |  |  |  |
| --- | --- | --- | --- |
| **direccion** | **titulo** | **link** | **title** |
| https://www.ambito.com/finanzas/dolar-libre-flotacion-cual-es-el-poder-fuego-del-tesoro-intervenir-y-cuanto-vendio-el-anuncio-n6185708 | Dólar sin libre flotación: cuál es el poder de fuego del Tesoro para intervenir y cuánto vendió tras el anuncio | https://www.economist.com/finance-and-economics/2025/09/02/china-turns-crypto-curious | China turns crypto-curious |
| https://www.ambito.com/finanzas/los-bonos-dolares-caen-el-wall-street-y-los-activos-argentinos-sufren-la-previa-electoral-n6185535 | Los bonos en dólares se hundieron hasta 3% tras el anuncio de intervención del Tesoro en el mercado cambiario | https://www.economist.com/finance-and-economics/2025/09/01/america-is-escaping-its-office-crisis | America is escaping its office crisis |
| https://www.ambito.com/finanzas/el-dolar-oficial-cayo-11-luego-la-decision-del-gobierno-intervenir-la-plaza-cambiaria-n6185573 | El dólar oficial cayó $11 luego de la decisión del Gobierno de intervenir en la plaza cambiaria | https://www.economist.com/finance-and-economics/2025/09/01/the-threat-of-deflation-stalks-asias-economies | The threat of deflation stalks Asia’s economies |
| https://www.ambito.com/bullrich-comparo-la-filtracion-audios-el-caso-maldonado-es-un-ataque-directo-la-democraciaargentina-n6185820 | Bullrich comparó la filtración de audios con el caso Maldonado: "Es un ataque directo a la democracia argentina" | https://www.economist.com/finance-and-economics/2025/08/28/trumps-interest-rate-crusade-will-be-self-defeating | Trump’s interest-rate crusade will be self-defeating |
| https://www.ambito.com/politica/crisis-el-gobierno-la-trama-coimas-genero-paralisis-y-desconfianza-mutua-dias-la-eleccion-n6185643 | Crisis política en el Gobierno: la trama de coimas generó parálisis y desconfianza mutua a días de la elección | https://www.economist.com/finance-and-economics/2025/08/28/gambling-or-investing-in-america-the-line-is-increasingly-blurred | Gambling or investing? In America, the line is increasingly blurred |
| https://www.ambito.com/economia/la-recaudacion-asociada-la-actividad-economica-anoto-agosto-su-primera-caida-2025-n6185797 | La recaudación asociada a la actividad económica anotó en agosto su primera caída de 2025 | https://www.economist.com/finance-and-economics/2025/08/28/how-trumps-war-on-the-federal-reserve-could-do-serious-damage | How Trump’s war on the Federal Reserve could do serious damage |
| https://www.ambito.com/politica/cristina-kirchner-cuestiono-javier-milei-las-ultimas-medidas-economicas-terminaste-quemar-los-libritos-la-escuelaaustriaca-n6185784 | Cristina cuestionó a Milei por las últimas medidas económicas: "Terminaste de quemar los libritos de la escuela Austríaca" | https://www.economist.com/finance-and-economics/2025/08/27/assessing-the-case-against-lisa-cook | Assessing the case against Lisa Cook |
| https://www.ambito.com/economia/empresarios-preocupados-la-tasa-interes-y-la-caida-la-actividad-n6185777 | Empresarios preocupados por la tasa de interés y la caída de la actividad | https://www.economist.com/finance-and-economics/2025/08/27/why-you-should-buy-your-employers-shares | Why you should buy your employer’s |

A la hora de hacer el *web scraping* hubo muchas diferencias en la obtención de la información de los títulos del diario estadounidense. En cambio, en los títulos del diario local apareció una pequeña traba relacionada con la protección de información por parte del sitio.  
La manera de hacer *scraping* fue muy diferente en ambos casos. Mientras que en el diario local utilizamos herramientas iguales a las vistas en clase, en el extranjero tuvimos que recurrir a una librería distinta y al uso de RSS, algo común en medios de comunicación grandes. Un RSS significa *“Really Simple Syndication”*. Es un formato XML que los medios usan para publicar sus artículos de manera estandarizada. A través de él obtenemos el título del artículo, el link, la fecha y el resumen (a veces también imágenes).

XML, o *Lenguaje de Marcado Extensible*, es un formato de texto diseñado para describir y estructurar datos de forma que puedan ser fácilmente compartidos entre diferentes sistemas de computación, como sitios web y bases de datos. En este caso, el método de obtención de la información no lo hicimos con *BeautifulSoup* como en Ámbito, sino que usamos una librería que interpreta XML y extrae la información de manera similar, llamada “Feedparser”. Una vez obtenida, la pasamos a un *DataFrame*.

En lo que respecta a la obtención de información de los títulos de Ámbito, no fue difícil, ya que seguimos el mismo procedimiento del tutorial dado por el curso. El desafío estuvo en el acceso al *scraping*: tuvimos que recurrir al uso de un *User-Agent*, que sirve para que la página de Ámbito no detecte que la consulta está siendo ejecutada por una máquina, dicho en criollo.  
  
Un *User-Agent* (agente de usuario) es un software que se identifica ante un servidor web para solicitar información, transmitiendo detalles sobre el cliente que hace la petición, como el nombre del navegador, el sistema operativo y el dispositivo.

En conclusión, este apartado fue interesante de ejecutar y, como detallamos, nos encontramos con obstáculos diferentes: por un lado, el acceso al *scraping* (Ámbito), y por el otro, la recolección de información (The Economist). Para más detalles, en el archivo.ipynb está explicado el proceso completo.