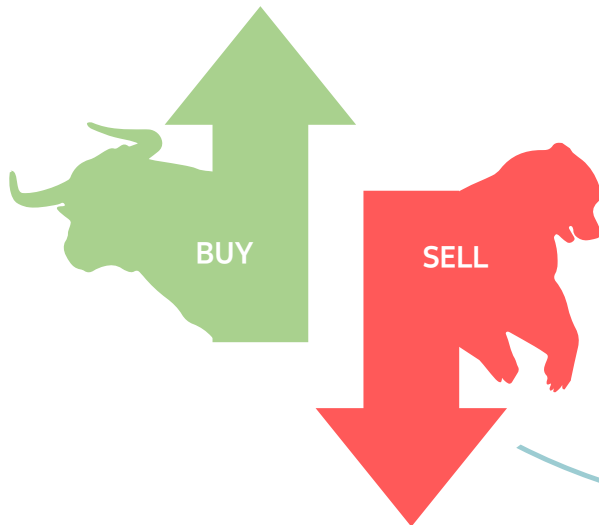# Scraping de Noticias Financieras

Ronald Chester
Nicolás Coddou

Objetivo

Obtener base de datos histórica de titulares de noticias sobre distintos instrumentos financieros.

En un futuro serán procesados con NLP para aportar información a la hora de tomar decisiones de compra o venta .

FINANCIAL TIMES

Digital advertising: Apple takes bigger bite of new business

HTML Websites → Web Scraping → Data

Sentiment Analysis

Positive   Negative   Neutral

BUY

SELL

Se diseñan 3 scripts diferentes, uno para cada uno de los siguientes portales:

FT FINANCIAL TIMES

yahoo! finance

Investing.com y **finviz**

Los CSV resultantes tendrán distintos formatos, ya que responderán a distintas formas de modelar el problema de clasificación.

CSV

Distinto número de filas y columnas

Herramientas:

Otras librerías
utilizadas

NLTK

Se importan las siguientes librerías :

```
import time
from selenium import webdriver
from selenium.webdriver.common.by import By
import os
import pandas as pd
from pandas import Series, DataFrame
from pandas import ExcelWriter
from pandas import ExcelFile
from bs4 import BeautifulSoup
from textblob import TextBlob
import textblob
from textblob import TextBlobstock
from pandas import ExcelFile
from pandas import ExcelWriter
from textblob import TextBlob
from textblob.sentiments import NaiveBayesAnalyzer
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
```

yahoo! finance

Se crea variable driver con objeto webdriver para mediante método GET abrir el Browser Chrome y la página de Yahoo Finance.

```
driver = webdriver.Chrome()

link = 'https://finance.yahoo.com/topic/stock-market-news/'

driver.get(link)
driver.maximize_window()

for i in range(1,15):
    driver.execute_script("window.scrollBy(0, 10000);")
    time.sleep(5)

soup = BeautifulSoup(driver.page_source, 'html.parser')
```

Se pide la palabra clave para buscarla y luego mediante subrutina se seleccionan los títulos con el keywords.

```python
print("Ingrese palabra a scrapear")
nombre = input()

news_list = []

# Encontrar todos los headers en finance.yahoo
for h in soup.findAll('h3',class_='Mb(5px)'):
    news_title = h.findAll()[0].text

    if news_title not in news_list:
        if 'ad' not in news_title:
            news_list.append(news_title)


no_of_news = 0
keyword_list = []
    # Recorre la lista y busca la palabra clave
for i, title in enumerate(news_list):
    text = ''
    if nombre.lower() in title:
        text = ' <------------ Palabra clave'
        no_of_news += 1
        keyword_list.append(title)

    print(i + 1, ':', title, text)

    # Print de Titulos de los artulos que contienen los keywords
print(f'\n--------- Total menciones de "{nombre}" = {no_of_news} ---------')
for i, title in enumerate(keyword_list):
    print(i + 1, ':', title)
```

```
Ingrese palabra a scrapear
stock
1 : Dow Jones Futures: Market Rally Breaks Lower As Jobs Report Intensifies Fed Rate Hike Fears
2 : Bear Market News And How To Handle A Market Correction
3 : 5 Best Chinese Stocks To Buy And Watch: Earnings, New EVs And Warren Buffett Move
4 : GLOBAL MARKETS-Shares tumble, yields jump as data fuel rate angst
5 : Stock Market Today-8/30: Stocks End Lower On Jobs Data, Consumer Confidence Leap
6 : ChargePoint's Earnings Fell Short. The Stock Is Rising Anyway.
7 : Sunak Warns of Risk of Markets Losing Confidence in UK Economy
8 : Stock your tween's school backpack with a hygiene kit—they need it
9 : Chewy stock drops 8% after pet retailer trims outlook for the year  <------------ Palabra clave
10 : Tesla Stock Vs. BYD Stock: EV Giant Dives As Buffett Trims Stake; TSLA Extends Post-Split Slide
11 : Dow ends 300 points lower as stocks retreat for third straight day  <------------ Palabra clave
12 : Enphase: IBD Stock Of The Day Lights Up The Solar Energy Market
13 : US STOCKS-Wall St closes down for 3rd straight session on Fed rate hike worry
14 : CrowdStrike beats Street's earnings view, hikes annual outlook; stock flat after hours  <------------ Palabra clave
15 : HP revenue comes up short because of a soft PC market, sending stock down  <------------ Palabra clave
16 : Chile central bank modifies norms for access, functioning of foreign exchange market
17 : HP Inc.'s Results Disappoint as Soft Market for PCs Takes a Toll
```

yahoo! finance

Se recorre la lista con las noticias seleccionadas conjuntamente con el método asociado a TextBlob asignando el sentimiento x noticia.

```
from textblob import TextBlob


news_sentblob = []
for news in keyword_list:
        print(news)
        analysis = TextBlob(news)
        print(analysis.sentiment)
        news_sentblob.append((news,analysis.sentiment[0],
                        analysis.sentiment[1]))
```

	news	polaridad	subjetividad
0	CANADA STOCKS-TSX falls to its lowest in five ...	0.000000	0.000000
1	Nvidia stock hits new 52-week low amid U.S. ex...	0.068182	0.377273
2	Bed Bath & Beyond stock falls again as another...	0.000000	0.000000
3	Nvidia and other chip stocks fall as US restri...	-0.112500	0.237500
4	Jeremy Grantham warns the 'superbubble' in sto...	0.050000	0.700000
5	Global stocks extend falls into new month on d...	0.068182	0.227273
6	The bear market in stocks has a long way to go...	-0.050000	0.400000
7	Toro stock slips after profit beats views but ...	-0.500000	0.500000
8	Am I a fool for keeping my IRA invested in sto...	0.000000	0.000000
9	Ciena stock tumbles after profit and sales mis...	-0.100000	0.400000
10	Signet Jewelers stock surges after profit beat...	0.000000	0.000000
11	What history says about September and the stoc...	0.000000	0.000000
12	Hormel Foods stock drops after profit miss and...	0.000000	0.000000
13	NIO, other China-based EV maker stocks fall af...	-0.125000	0.375000
14	EMERGING MARKETS-Currencies, stocks sell-off a...	0.000000	0.000000
15	Russian stocks ease from recent highs, rouble ...	0.000000	0.125000
16	Australia shares end at 5-week low as miners, ...	0.000000	0.300000

2

En este caso se busca generar un csv que contenga Titulo, Fecha y Cuerpo de cada noticia de un instrumento .

Se usara **Apple Inc.** como ejemplo

Estos son lo resultados de ingresar "apple" en el buscador del Financial Times.

E artículo, una imagen, el resumen y la fecha.

El índice del total de páginas de resultado. Cada una contiene 25 artículos.

*Para poder scrapear los artículos con **Título, Fecha y Cuerpo de la noticia** se necesita principalmente de las siguientes librerias.*
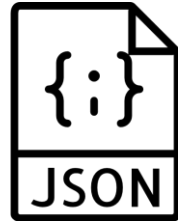
```
Ingrese instrumento: apple
Hay un total de 1375 paginas --> 34375 articulos aprox

Cuantas paginas desea scrapear?: 2
Serian aprox. 50 articulos en 0:32 minutos aprox.
```

BeautifulSoup

{;}
JSON

El script pide ingresar el nombre del instrument a buscar. Se obtiene la cantidad de páginas de resultado y la cantidad de artículos totales de manera aproximada.

Además se pide ingresar cuantas páginas de resultados se desea scrapear.

Con un ciclo For:

Se obtiene el link de cada pagina de resultados al ingresar. El siguiente link se rellena con el instrumento ingresado y el ciclo for ira aumentando la pagina de resultados

```
"https://www.ft.com/search?q={instrumento}&page={pagina}&contentType=article&sort=relevance&expandRefinements=true"
```

Se crea el objeto BeautifulSoup con cada link

```
url = requests.get(url_name)
soup = BeautifulSoup(url.text, "html.parser")
```

FT FINANCIAL TIMES

Dentro de cada página de Resultados...

Se busca cada Artículo

```python
for article in soup.find_all("div", class_="o-teaser__content"):

    # Se obtiene Link de articulo
    content = article.find("a", class_ = "js-teaser-heading-link", href=True)
    link = "https://www.ft.com" + content["href"]
    links.append(link)

    # Se entra en el link
    url = requests.get(link)
    soup = BeautifulSoup(url.text, "html.parser")
```

Se obtiene link para ingresar a cada artículo

Se entra a cada link

FT FINANCIAL TIMES

Al entrar en cualquier artículo del FT aparecerá el pop-up de subscripción impidiendo leer el artículo. ❗

Es posible **saltar** este Paywall utilizando la data en formato **JSON** contenida dentro del HTML de cada artículo.

Technology sector

Save to myFT

Group of Apple employees pushes back against return-to-office order

...not hampering their work. n Sla s, the internal messaging platform used by Apple, more than 1 000 Appl employees have joined the group "Remote Work...

AUGUST 22, 2022

Become an FT subscriber to read:

" Group of Apple employees pushes back against return-to-office order "

Make informed decisions with the FT

Keep abreast of significant corporate, financial and political developments around the world. Stay informed and spot emerging risks and opportunities with independent global reporting, expert commentary and analysis you can trust.

Choose your subscription

Trial	Digital	ePaper	GROUP SUBSCRIPTION Team or Enterprise
Try full digital access and see why over 1 million readers subscribe to the FT	Be informed with the essential news and opinion	An easy-to-navigate digital replica of the print edition	Premium FT.com access for multiple users, with integrations & admin tools
$1 for 4 weeks	**$40 per month**	**$20 for 3 months**	**Pay based on use**
Then $69 per month New customers only Cancel anytime during your trial	OR **$375 for 1 year** BEST VALUE - SAVE 20%	INTRODUCTORY OFFER Then $24.75 every 3 months	Does my organisation subscribe?
Select	Select	Select	Get Started
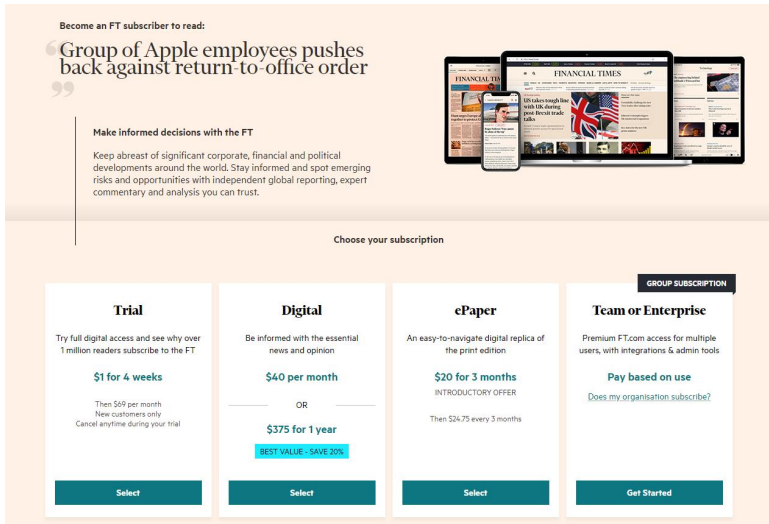
FT FINANCIAL TIMES

Dentro del HTML de cada artículo existe un elemento `<script>` dentro del `<head>`.

HTML Websites

```
<!DOCTYPE html>
<html lang="en-GB" class="js enhanced js-focus-visible" data-o-component="o-typography" style="overflow-x:hidden;background-color:#fff1e5;color:#33302e" data-js-focus-visible>
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <title>Subscribe to read | Financial Times</title>
    <meta name="description" content="News, analysis and comment from the Financial Times, the world's leading global business publication">
    <meta name="robots" content="index,follow,max-snippet:200,max-image-preview:large">
    <meta name="google-site-verification" content="4-t8sFaPvpO5FH_Gnw1dkM28CQepjzo8UjjAkdDflTw">
    <script async src="https://www.googletagmanager.com/gtm.js?id=GTM-NWQJW68"></script>
    <script type="application/ld+json">…</script>
    <script type="application/ld+json">…</script>
    <meta property="fb:pages" content="8860325749">
    <meta property="twitter:site" content="@FinancialTimes">
    <meta name="apple-itunes-app" content="app-id=1200842933">
    <link rel="icon" type="image/svg+xml" href="https://www.ft.com/__origami/service/image/v2/images/raw/ftlogo-v1%3Abr
```
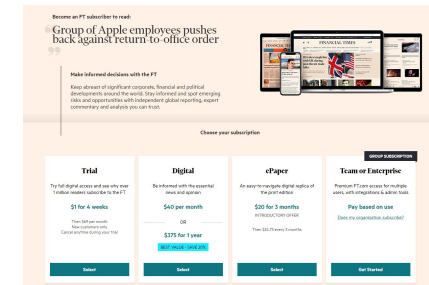
Dentro de este se encuetra **TODA** la informacioón de la noticia en formato `JSON`.


```
<script type="application/ld+json">
```

{"@context":"http://schema.org","@type":"NewsArticle","headline":"Group of Apple employees pushes back against return-to-office order","datePublished":"2022-08-22T04:00:03.841Z","dateModified":"2022-08-22T04:00:03.841Z","isAccessibleForFree":"False","isPartOf":{"@type":["CreativeWork","Product"],"name":"Financial Times","productID":"ft.com:subscribed"},"image":{"@type":"ImageObject","@context":"http://schema.org":"https://d1e00ek4ebabms.cloudfront.net/production/c749b1ba-79f0-42d5-9889-3727ca18e   .png","width":2506,"height":1410},"author":{"@type":"Person","@context":"http://schema.org","na   trick McGee","url":"https://www.ft.com/patrick-mcgee","height":"1.01492537 smoots",   :{"@type":"Organization","@context":"http://schema.   name":"Financial Times","legalName":"The Financial Times Ltd.","logo":{"@type":"ImageObject   ntext":"http://schema.org","url":"https://im.ft-static.com/m/img/masthead_main.jpg","width":435,   ":36},"url":"https://www.ft.com","sameAs":["https://www.twitter.com/FT","https://www.face   om/financialtimes","https://www.linkedin.com/company/financial-times","https://www.youtube.com/user/FinancialTimesVideos","https://www.instagram.com/financialtimes"]}},"articleBody":"Apple's headquarters in Cupertino, California [https://d1e00ek4ebabms.cloudfront.net/production/c749b1ba-79f0-42d5-3727ca18e4da.png]\nApple has told emp oyees working in and around its Cupertino headquarters that\nthey must return to the office for three days a week from next month @ Sam\nHall/BloombergApple employees are pushing back against the iPhone   er's call\nfor workers to return to the office next month, arguing that they have shown\nthey can perform "exceptional work" during two-plus years of flexible\narrangements.\n\nApple Together   roup of workers that formed last year when offices around the\nglobe were forced to work remotely because of the pandemic, began circulating a\npetition internally on Sunday, demanding "location fle   work".\n\nThe petition, seen by the Financial Times, is a response to an order from chief\nexecutive Tim Cook last week telling employees in and around the Cupertino\nheadquarters that they must   turn to the office three days a week from\nSeptember 5. Cook said he wanted to preserve the "in-person collaboration that\nis so essential to our culture".\n\nApple Together counters that a "unifor   date from senior leadership" fails to\nrespect the "many compelling reasons" why some employees are "happier and more\nproductive" working outside of traditional office arrangements. \n\nThe grou   demanding the company allow employees to work with their "immediate\nmanager" to decide their working arrangements, and that they should not be\nsubject to "high-level approvals" and "complex procedures" or have to provide\nprivate information. \n\nA corporate employee within hardware engineering in Cupertino who is helping to\norganise the petition told the Financial Times that Apple Together intended to\ncollect signatures this week before verifying and sending the results to\nexecutives.\n\n"At this juncture we will not be releasing any specific names of individuals\npublicly or to exec leadership to protect our colleagues, especially in light of\nretail union busting and recent reports of allegations of retaliation from HR,"\nthis person said.\n\nApple declined to comment. \n\nSilicon Valley companies, including Facebook and Google, allowed engineers to\nstay at home when Covid-19 forced people to work remotely in March 2020. In some\ncases employees were allowed to relocate to other parts of the country without\nit affecting their salaries. \n\nWhether work should now return to pre-Covid norms has become a contentious\nissue, with some companies touting their flexible policies to lure and retain\ntalent.\n\nLast year Spotify introduced a "Work from Anywhere" policy, saying it would\nsupport work-life balance by giving employees "the freedom to choose" where to\nwork.\n\nRECOMMENDED\n * Working from home debates must include what 'home' actually is\n\nTesla boss Elon Musk, by contrast, told workers that "remote work is no longer\nacceptable". Except for "particularly exceptional contributors", he wrote to\nemployees in June, it will be assumed that workers who fail to show up in person\n"have resigned".\n\nCook has not been as forceful as Musk — in March he acknowledged a return to the\noffice might be "an unsettling change" for some — but since June 2021 he has\nrepeatedly tried to get workers back to the office, only to have the plans\ndelayed by new waves in Covid cases.\n\nApple has thrived during the pandemic period, with its market valuation roughly\ndoubling from $1.4tn in February 2020 to $2.8tn today. Some employees argue that\nproves that the lack of in-office culture is not hampering their work. \n\nOn Slack, the internal messaging platform used by Apple, more than 10,000 Apple\nemployees have joined the group "Remote Work Advocacy". And on Blind, the\nanonymous messaging platform for tech employees, return-to-work discussions are\namong the most frequent and popular issues among the iPhone maker's employees.\n\nIn May, a prominent machine learning computer scientist, Ian Goodfellow, left\nApple for Google sibling DeepMind, reportedly telling colleagues that Apple's\nreturn-to-work policy was one of the main reasons he left. \n\nThe outspokenness of some Apple employees appears to have had some impact. I   June, Cook had asked workers to come back to the office on Mondays, Tuesdays and\nThursdays. However, in last week's memo the policy was relaxed to Tuesdays and\nThursdays, plus a third day that will   by individual teams. \n\n"We believe that Apple should encourage, not prohibit, flexible work to build a\nmore diverse and successful company where we can feel comfortable to 'think\ndif   ,"\nthe Apple petition said.","wordCount":734,"publisher":{"@type":"Organization","@context":"http://schema.org","name":"Financial Times","legalName":"The Financial Times Ltd.","logo":{"@type":"ImageObject","@context":"http://schema.org","url":"https://im.ft-static.com/m/img/masthead_main.jpg","width":435,"height":36},"url":"https://www.ft.com","sameAs":["https://www.twitter.com/FT","https://www.facebook.com/financialtimes","https://www.linkedin.com/company/financial-times","https://www.youtube.com/user/FinancialTimesVideos","https://www.instagram.com/financialtimes"]}}

```
</script>
```

Esta data en fomato JSON es la información que el Financial Times comparte con Google para que sus artículos puedan aparecer en los resultados de las busquedas.

Entonces, para cada artículo:

1. Se entra en el link.
2. Se obtiene cada "application/ld+json".

   (El cual contiene toda la data de la noticia)

3. Se obtiene el body, la fecha y el título.

1
```python
# Se entra en el link
url = requests.get(link)
soup = BeautifulSoup(url.text, "html.parser")
```

2
```python
# Se obtiene json (que es la forma que tiene ft para el paywall). Este json contiene todo lo relacionado con la noticia
data = [json.loads(i.string) for i in soup.find_all("script", type="application/ld+json")]

body = ""
```

3
```python
# Se obtienen las partes especficias del json
for i in data:
    try:
        body = i["articleBody"]
        fecha = i["datePublished"]
        titulo = i["headline"]
    except KeyError:
        pass
```

FT FINANCIAL TIMES

```
Ingrese instrumento: apple
Hay un total de 1375 paginas --> 34375 articulos aprox

Cuantas paginas de resultados desea scrapear?: 40
Serian aprox. 1000 articulos en 10:40 minutos aprox.

Desea Continuar? [y]/n: y

Paginas de Resultados Scrapeadas: 100%|██████████| 40/40 [10:18<00:00, 15.46s/it]Listo!
Se obtuvieron 1000 articulos
```


HTML Websites → Web Scraping → Data

apple_40.csv

1000 filas x 4 Columnas

Origen de archivo: 65001: Unicode (UTF-8)
Delimitador: Coma

Titulo	Fecha	Link	Body
Vietnam to start making Apple Watch and MacBook	24-08-2022	https://www.ft.com/content/fcd59b6d-0be5-47b3-859...	Apple Watch models are displayed at a store in Singapo...
Apple's return-to-office order sparks anxiety among tec...	24-08-2022	https://www.ft.com/content/9f3e0159-f0e2-4a33-aeb...	When Tim Cook sent his workforce home in March 202...
The trickle-down effect of empty rivers and reservoirs	22-08-2022	https://www.ft.com/content/763a5853-f073-43f0-bce...	Crab catching on the dried out Jialing river, a major trib...
Group of Apple employees pushes back against return-t...	22-08-2022	https://www.ft.com/content/e40ba5a2-9dbd-4f88-87f...	Apple's headquarters in Cupertino, California [https://d...
Droughts grip large parts of the globe	22-08-2022	https://www.ft.com/content/c736f20a-af54-4246-aa40...	This is an audio transcript of the FT News Briefing podc...
Apple warns users to install software updates to preven...	19-08-2022	https://www.ft.com/content/46fe652a-e1bd-48dd-978...	Apple iPhone 13s [https://d1e00ek4ebabms.cloudfront...
Vietnam gets a bigger bite of Apple, Nintendo heirs put...	18-08-2022	https://www.ft.com/content/5b30433c-62cc-4f16-858...	Vietnamese woman walks past an Apple store [https://...
Apple tells employees to return to office 3 days a week...	16-08-2022	https://www.ft.com/content/b8080070-4c41-4193-838...	Tim Cook [https://d1e00ek4ebabms.cloudfront.net/pro...
Warren Buffett's Berkshire Hathaway makes big bet on...	15-08-2022	https://www.ft.com/content/eba73448-5065-446f-be0...	Warren Buffett [https://d1e00ek4ebabms.cloudfront.n...
Digital advertising: Apple takes bigger bite of new busin...	15-08-2022	https://www.ft.com/content/b0ae756a-89b7-486e-857...	A smartphone screen showing the Disney+ logo [https:/...
Bad Sisters, Apple TV Plus review — siblings turn murde...	15-08-2022	https://www.ft.com/content/c1afefaa-2e81-4e3a-938d...	Five women and a young girl stand in a church pew dre...
Taiwan security officials want Foxconn to drop stake in...	10-08-2022	https://www.ft.com/content/866467a4-6627-4356-ac4...	Workers on the production line at a Foxconn plant in G...
Investors divided over how long Big Tech rally will last	09-08-2022	https://www.ft.com/content/50c439e6-a038-40c6-a33...	Shares of leading US technology companies have rebou...
Small businesses count cost of Apple's privacy changes	09-08-2022	https://www.ft.com/content/1959b06d-0a6e-4c37-952...	Shoppers on Bond Street in London look at their mobile...
Apple under scrutiny for mishandling misconduct claims	08-08-2022	https://www.ft.com/content/4525f453-8312-4831-932...	This is an audio transcript of the FT News Briefing podc...
Five Days at Memorial, Apple TV Plus review — one of t...	08-08-2022	https://www.ft.com/content/1b37e295-c879-436c-b43...	Four people in medical scrubs look out the window, the...
Taiwanese Apple supplier battles activists over $4bn ca...	07-08-2022	https://www.ft.com/content/cc440b45-2440-4021-9d8...	An employee walks past a logo of Catcher Technology a...
Luck — Apple takes on Pixar with tale of leprechauns a...	04-08-2022	https://www.ft.com/content/be829013-02b4-4b18-a9...	In a scene from an animated film, a dragon looks down...
The women calling out Apple's handling of misconduct...	04-08-2022	https://www.ft.com/content/96160847-af3f-44b6-812...	Megan Mohr was five years into her Apple career when...

CSV

apple_40.csv

1000 filas x 4 Columnas

Origen de archivo
65001: Unicode (UTF-8)

Delimitador
Coma

Título	Fecha	Link	Body
Vietnam to start making Apple Watch and MacBook	24-08-2022	https://www.ft.com/content/fcd59b6d-0be5-47b3-859...	Apple Watch models are displayed at a store in Singapo...
Apple's return-to-office order sparks anxiety among tec...	24-08-2022	https://www.ft.com/content/9f3e0159-f0e2-4a33-aeb...	When Tim Cook sent his workforce home in March 202...
The trickle-down effect of empty rivers and reservoirs	22-08-2022	https://www.ft.com/content/763a5853-f073-43f0-bce...	Crab catching on the dried out Jialing river, a major trib...
Group of Apple employees pushes back against return-t...	22-08-2022	https://www.ft.com/content/e40ba5a2-9dbd-4f88-87f...	Apple's headquarters in Cupertino, California [https://d...
Droughts grip large parts of the globe	22-08-2022	https://www.ft.com/content/c736f20a-af54-4246-aa40...	This is an audio transcript of the FT News Briefing podc...
Apple warns users to install software updates to preven...	19-08-2022	https://www.ft.com/content/46fe652a-e1bd-48dd-978...	Apple iPhone 13s [https://d1e00ek4ebabms.cloudfront...
Vietnam gets a bigger bite of Apple, Nintendo heirs put...	18-08-2022	https://www.ft.com/content/5b30433c-62cc-4f16-858...	Vietnamese woman walks past an Apple store [https://...
Apple tells employees to return to office 3 days a week...	16-08-2022	https://www.ft.com/content/b8080070-4c41-4193-838...	Tim Cook [https://d1e00ek4ebabms.cloudfront.net/pro...
Warren Buffett's Berkshire Hathaway makes big bet on...	15-08-2022	https://www.ft.com/content/eba73448-5065-446f-be0...	Warren Buffett [https://d1e00ek4ebabms.cloudfront.n...
Digital advertising: Apple takes bigger bite of new busin...	15-08-2022	https://www.ft.com/content/b0ae756a-89b7-486e-857...	A smartphone screen showing the Disney+ logo [https:/...
Bad Sisters, Apple TV Plus review — siblings turn murde...	15-08-2022	https://www.ft.com/content/c1afefaa-2e81-4e3a-938d...	Five women and a young girl stand in a church pew dre...
Taiwan security officials want Foxconn to drop stake in...	10-08-2022	https://www.ft.com/content/866467a4-6627-4356-ac4...	Workers on the production line at a Foxconn plant in G...
Investors divided over how long Big Tech rally will last	09-08-2022	https://www.ft.com/content/50c439e6-a038-40c6-a33...	Shares of leading US technology companies have rebou...
Small businesses count cost of Apple's privacy changes	09-08-2022	https://www.ft.com/content/1959b06d-0a6e-4c37-952...	Shoppers on Bond Street in London look at their mobile...
Apple under scrutiny for mishandling misconduct claims	08-08-2022	https://www.ft.com/content/4525f453-8312-4831-932...	This is an audio transcript of the FT News Briefing podc...
Five Days at Memorial, Apple TV Plus review — one of t...	08-08-2022	https://www.ft.com/content/1b37e295-c879-436c-b43...	Four people in medical scrubs look out the window, the...
Taiwanese Apple supplier battles activists over $4bn ca...	07-08-2022	https://www.ft.com/content/cc440b45-2440-4021-9d8...	An employee walks past a logo of Catcher Technology a...
Luck — Apple takes on Pixar with tale of leprechauns a...	04-08-2022	https://www.ft.com/content/be829013-02b4-4b18-a9...	In a scene from an animated film, a dragon looks down...
The women calling out Apple's handling of misconduct...	04-08-2022	https://www.ft.com/content/96160847-af3f-44b6-812...	Megan Mohr was five years into her Apple career when...

CSV

El largo de cada csv dependerá de cuantas páginas se deseen scrapear. La idea al final sería concatenar los distintos csv de todos los instrumentos.

Aún se encuentra en revisión el como se aplicará NLP a este csv.

FT FINANCIAL TIMES

3

Para estos dos portales hay menos avance. Se están probando otras formas de consolidar el CSV en base a revisiones bilbiograficas.

Estos dos portales contienen varios indicadores sobre los instrumentos además de una sección de noticias sobre estos.

La idea es poder juntar la información de ambas y generar un CSV que contenga al menos 25 titulares por día como columnas y una columna final booleana de 0 o 1 que representa si la acción subió o no.

finviz

Finviz se puede trabajar facimente con :

BeautifulSoup

Column1	Column2
Fechas	Titular
['Aug-23-22', '11:43PM']	Dow Jones Futures Fall: Market Rally Struggling; Apple,...
['08:46PM']	Sony Unveils High-End PS5 Controller With Parts You Ca...
['07:08PM']	Dow Jones Futures: Market Rally Struggling; Apple, The...
['04:37PM']	Apple employees push back against return-to-office ma...
['03:29PM']	Apple (AAPL) Provides Self-Service Overhaul for Mac N...
['02:47PM']	Historic hearing aid ruling could help 'those four-out-of-...
['02:33PM']	5 Dow Jones Stocks To Buy And Watch In August: Chevr...
['11:49AM']	Apple to compete with Xiaomi, Samsung in Indias smart...
['11:31AM']	Chip Giant Nvidia's Stock Slump Fails to Remove Its Pric...
['10:46AM']	Apple plans to produce iPhone 14 in India: Report
['10:21AM']	These Investment Wines Have Outperformed Apple, Fo...
['10:19AM']	Alphabet (GOOGL) Boosts Fitness Tracking With Latest...
['10:17AM']	Apple Sees This Business Reaching $10 Billion Soon
['09:45AM']	3 Things About Unity Software That Smart Investors Kn...
['09:40AM']	Apple Stock Just Had Its Worst 2 Days Since June. Why I...
['07:03AM']	Engineer admits stealing Apple Car trade secrets before...
['06:46AM']	Apple to Cut New iPhone 14 Production Lag Between In...
['06:06AM']	2 Simple Reasons Why Apple's Next iPhone Could Be a...
['06:00AM']	Down Between 51% and 78%: 3 Nasdaq Growth Stocks...

```python
# Ingreso de URL
url = "https://finviz.com/quote.ashx?t="


instrumento = input("Ingresar instrumento en minuscula: ")
instrumentoUpper = instrumento.upper()


numCSV = input("Ingresar numero de CSV de ese instrumento: ")


# ==========================================
# Se genera url segun instrumento
url_name = url + instrumentoUpper


# Creacion objeto bs4#
# Se define objeto BeautifulSup
url = requests.get(url_name, headers={'User-Agent': 'Mozilla/5.0'})
soup = BeautifulSoup(url.content, "lxml")
soup.prettify


# Se encuentra tablon de noticias
tabla = soup.find(id="news-table")


# ==========================================
###### Generacion de DF ######
fechas = []
titulares = []


for articulo in  tabla.find_all("tr"): # para cada fila del tablon de noticias
    fecha = articulo.find("td").string.split()
    titular = articulo.find("a").string

    fechas.append(fecha)
    titulares.append(titular)


diccDF = {"Fechas": fechas, "Titular": titulares}


paraCSV = pd.DataFrame(diccDF)
```

Investing.com

El caso de Investing es más complejo ya que require de Selenium para pasar un pop-up de subscripción.

Además require utilizar el buscador para obtener el link de cada búsqueda de instrumento

Selenium

Investing.com

AAPL Historical Data

Se obtiene la data histórica:

Time Frame:

Daily ▾   ⬇ Download Data   07/31/2022 - 08/31/2022

Date ⌄	Price ⌄	Open ⌄	High ⌄	Low ⌄	Vol. ⌄	Change % ⌄
08/31/2022	157.22	160.12	160.56	157.19	85.29M	-1.06%
08/30/2022	158.91	162.13	162.56	157.72	74.79M	-1.53%
08/29/2022	161.38	161.15	162.90	159.82	73.31M	-1.37%
08/26/2022	163.62	170.57	171.05	163.56	78.96M	-3.77%
08/25/2022	170.03	168.78	170.14	168.35	51.22M	+1.49%
08/24/2022	167.53	167.32	168.11	166.24	53.84M	+0.18%
08/23/2022	167.23	167.08	168.71	166.65	54.15M	-0.20%

```python
# ========= Data Historica =========

browser.find_element_by_link_text("Historical Data").click()

fecha = pd.to_datetime(browser.find_element_by_xpath('//td[@class="datatable_cell__3gwri font-bold"]').text)


precioFinal = float(browser.find_element_by_xpath('//td[@class="datatable_cell__3gwri datatable_cell--align-end__Wua8C datatable_cell--down__2CL8n"]').text)

volumen =  browser.find_element_by_xpath('//td[@data-test="volume-cell"]').text
volumen = float(volumen.replace("M", ""))

deltaPorcentual = browser.find_element_by_xpath('//td[@class="datatable_cell__3gwri datatable_cell--align-end__Wua8C datatable_cell--down__2CL8n font-bold"]').text
deltaPorcentual = deltaPorcentual[deltaPorcentual.find("(")+1:deltaPorcentual.rfind(")")]
deltaPorcentual = float(deltaPorcentual.replace("%", ""))
```

Investing.com

El problema con este portal es que obtener la información de cada noticia tomaría demasiado tiempo con selenium.

Existe un problema al utilizar BeautifulSoup ya que no se ha encontrado forma de pasar el pop-up de subscripción.

La metodología a seguir esta en revisión.

An Ominous Month Looms For Battered Nasdaq Bulls

By Bloomberg - 15 hours ago | 💬 1

(Bloomberg) -- The second half of August has been bruising for technology stocks, but those hoping for a respite from the declines shouldn't relax just yet: September is just around the corner. The...

Taiwan's TSMC progressing well with Arizona chip plant, governor says

By Reuters - 16 hours ago

TAIPEI (Reuters) -Taiwanese chip maker TSMC is making "excellent" progress building its new plant in Arizona, the governor of the U.S. state said on Wednesday, going on to praise his state's role in...

Foreign business travel missing ingredient for Irish hotel recovery - Dalata

By Reuters - 17 hours ago

By Padraic Halpin DUBLIN (Reuters) - Executives at Ireland's large hub of multinational companies are still only going on a small fraction of the foreign business trips they made before the COVID-19...

Reef's Highly Anticipated Reef Card is Officially Available for Crypto Holders

By DailyCoin - 17 hours ago

Reef, a layer 1 Substrate based blockchain for DeFi, NFTs, and gaming, announced today its highly anticipated Reef Card is now available for people in the United Kingdom and Europe. Reef has...

Apple chief privacy officer set to leave company for law firm -Bloomberg News

By Reuters - Aug 30, 2022 | 💬 5

(Reuters) - Apple Inc (NASDAQ:AAPL)'s Chief Privacy Officer Jane Horvath will be leaving the company soon to work at a law firm, Bloomberg News reported on Tuesday citing people familiar with the...

1 2 3 4 5 6 7 8 9 10   Next ➡