



Universitat Oberta
de Catalunya

Práctica Web Scraping

AUTOR: **JOSÉ ANTONIO GONZÁLEZ CONSTANZA**

DNI: **07484042A**

CORREO: **jgonzalezcon@uoc.edu**

Introducción

La práctica de Web Scraping de la asignatura “Tipología y ciclo de vida de los datos”, perteneciente al Máster en Ciencia de Datos de la Universitat Oberta de Catalunya, se ha desarrollado de manera individual por **José Antonio González Constanza**. En toda su consecución, se ha tratado de realizar un trabajo de gran calidad que permitiera desarrollar todas las competencias y objetivos recogidos en la práctica.

Descripción de la práctica a realizar

Tal como viene enunciado en la misma, se procede a describir y explicar cada uno de los apartados:

1.- Contexto.

El objetivo se ha centrado en recopilar información de Yahoo Finance, en concreto las valoraciones de stock con mejores subidas en precio, los llamados “gainers”. Mi propósito se centro en comprobar la posibilidad de poder hacer scraping en una web como Yahoo, concretamente en el área de finanzas, para confirmar la posibilidad de tomar datos financieros de stock de una web de gran relevancia en dicho campo.

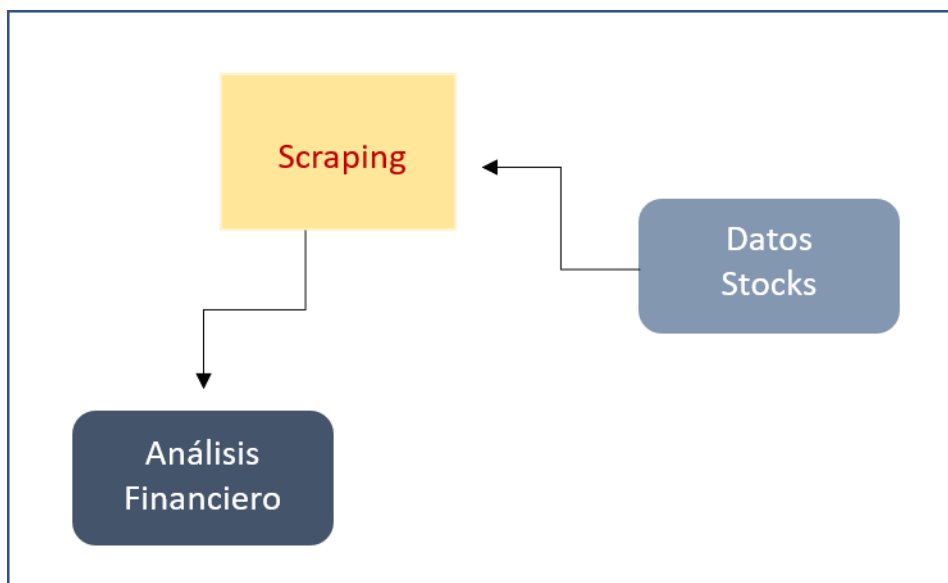
2.- Título

“INTRADAY PRICE GAINERS”

3.- Descripción del dataset

El dataset obtenido incluye los 25 mejores stocks (gainers) desde el ultimo cierre de los mercados y sus características. Es decir, las mejores subidas en precio acumulado de los diferentes stocks que componen la bolsa o stock market en EE. UU.

4.- Representación gráfica.



5.- Contenido

Symbol: El símbolo que representa al stock.

Name: Nombre de la empresa que cotiza en el stock market.

Price (Intraday): Precio actual del stock o acción.

Change: Cambio porcentual producido en el precio del stock desde el último cierre.

%Change: Cambio porcentual acumulado en el precio del stock.

Volume: Volumen total negociado del stock.

Avg Vol (3 month): Volumen total negociado en promedio de los últimos 3 meses.

Market Cap: Capitalización bursatil (Market capitalization). El total valor en dólares de todas las acciones de una empresa.

PE Ratio (TTM): Price to earnings ratio. Una de las más populares ratios usados por los inversores para determinar el valor justo de un stock. Es decir, el precio que los inversores estarían dispuestos a pagar por los beneficios de la compañía.

6.- Agradecimientos

Agradecimiento a la web de Yahoo principalmente en el área financiera, donde disponen de un extenso conjunto de datos para realizar todo tipo de análisis financiero.

Comprobamos la exclusión de determinados directorios y alguna página.

```
User-agent: *
Sitemap: https://finance.yahoo.com/sitemap_en-us_desktop_index.xml
Sitemap: https://finance.yahoo.com/sitemaps/finance-sitemap_index_US_en-US.xml.gz
Sitemap: https://finance.yahoo.com/sitemaps/finance-sitemap_googlenewsindex_US_en-US.xml.gz
Disallow: /r/
Disallow: /__rapidworker-1.2.js
Disallow: /__blank
Disallow: /_td_api
Disallow: /_remote
User-agent:googlebot
Disallow: /m/

User-agent:googlebot-news
Disallow: /m/
```

Considero que no existen repercusiones a nivel legal pues estamos realizando scraping de información pública. Por ello, se puede establecer que un uso moderado en la obtención de los datos considerados en el proyecto, no deberían producir ningún tipo de problema ético o legal.

7.- Inspiración

Los datos de interés nos permiten usar precios de stocks en tiempo real, para la automatización parcial de análisis de activos financieros. Es decir, la búsqueda de información financiera para la toma de decisiones en todo tipo de operaciones relacionadas con el trading. Es verdad, que los datos proporcionados, tan solo nos permiten conocer aquellos activos cuyas subidas acumuladas en precio han sido más altas, pero podemos extrapolar dichos procedimientos a otros tipos de datos, también financieros, que nos permitan automatizar tareas y con ello, realizar análisis para la toma de decisiones.

8.- Licencia

Los datos obtenidos son mostrados de forma publica sin necesidad de restricciones de acceso, por lo que considero su uso de dominio público. Por tanto, podemos considerar una licencia **Released Under CC0: Public Domain License**.

9.- Código y vídeo

<https://github.com/jgonzalezcon/Scraping-Financial-Data>

10.- Dataset

El dataset se trata de un archivo .csv donde se recogen los 25 mejores stocks en cuanto a subidas acumuladas de sus precios. Incorpora todas las características o atributos que explican los movimientos de los precios intraday como volumen de negociación sobre el activo, nombre de la empresa, ratios, etc..

[illegible]

DOI:

10.5281/zenodo.5651077

Markdown:

```
[![DOI](https://zenodo.org/badge/DOI/10.5281/zenodo.5651077.svg)](https://doi.org/10.5281/zenodo.5651077)
```

reStructuredText:

```
.. image:: https://zenodo.org/badge/DOI/10.5281/zenodo.5651077.svg
   :target: https://doi.org/10.5281/zenodo.5651077
```

HTML:

```
<a href="https://doi.org/10.5281/zenodo.5651077"></a>
```

Image URL:

<https://zenodo.org/badge/DOI/10.5281/zenodo.5651077.svg>

Target URL:

<https://doi.org/10.5281/zenodo.5651077>

Recursos

- Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC.
- Masip, D. (2019) El lenguaje Python. Editorial UOC.
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Simon Munzert, Christian Rubba, Peter Meißner, Dominic Nyhuis. (2015).
- Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.
- Tutorial de Github <https://guides.github.com/activities/hello-world>.