

PRACTICA 1. SCRAPING

Recopilación de un conjunto de valores correspondientes a la bolsa española IBEX35, bolsa estadounidense NASDAQ y fluctuación del euro respecto a las demás divisas utilizando técnicas de scraping

Creado por: Antonio Guzmán y Joaquín Fernández León

09/11/2018

Este conjunto de datos junto una imputación con algoritmos de predicción de series temporales nos pueden ayudar a realizar estudios y poder realizar predicciones futuras que nos indicarán si es interesante comprar o vender de una empresa en concreto en el futuro. Esto es posible porque no solamente tenemos los valores de las empresas sino también la fluctuación de las monedas. De esta manera si un usuario tiene euros y desea realizar una compra o venta en un futuro sobre una empresa que cotiza en dólares necesitará saber no solamente el valor de esa empresa en ese momento futuro sino también conocer el valor que tendrá la divisa en ese momento.

1. Título del dataset. Poned un título que sea descriptivo.

Conjunto de valores correspondientes a la bolsa española IBEX35, bolsa estadounidense NASDAQ y fluctuación del euro respecto a las demás divisas.

2. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de datos por vuestro subtítulo.

Conjunto de valores que permite realizar un estudio del movimiento de tanto las empresas en las bolsas de IBEX35 y NASDAQ cómo la fluctuación en ese mismo instante de las divisas.

3. Imagen. Agregad una imagen que identifique vuestro dataset visualmente



4. Contexto. ¿Cuál es la materia del conjunto de datos?

Con este conjunto de datos podemos realizar predicciones de valores de bolsa en el futuro basándonos en datos y correlaciones entre empresas en el pasado.

5. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?

Nuestros scripts en python están enfocados a crear tres archivos:

- Valores de la bolsa estadounidense NASDAQ
- Valores de la bolsa española IBEX35
- Fluctuación de la divisa euro respecto a las demás

En los dos primeros recogemos los campos:

- **Volumen** : es la cantidad de acciones o contratos negociados en un periodo de tiempo determinado. Por así decirlo por cada compra hay una venta y esto nos servirá como un indicador de compraventas.
- **Nombre** : nombre de la empresa a la que corresponde estos valores.
- **Hora**: Hora correspondiente a la publicación de esos valores en la web.
- **Fecha**: en formato " %Y-%m-%d" para conocer cuándo se ha recogido el dato.
- **Stocks**: muestra el valor que tiene esa empresa en ese momento.

En el último punto recogemos:

- **Currency_1, Currency_2**: nos indica la comparación de la moneda 1 respecto la 2.
- **Value**: es el valor que tiene en ese momento el cambio.
- **Date**: la fecha
- **Hour**: hora a la que se ha mostrado ese dato.

Tabla 1. Ejemplo de scraping de los valores de NASDAQ

	Volumen	Nombre	Hora	Fecha	Stocks
0	23.878.241	ACTIVISION I	19:18	2018-11-09	55,68
1	1.189.934	ADOBE SYSTEM	19:18	2018-11-09	241,78
2	495.885	ALEXION PHAR	19:17	2018-11-09	123,31

Tabla 2. Ejemplo de scraping de los valores de IBEX35

	Volumen	Nombre	Hora	Fecha	Stocks
0	167.223	ACCIONA	17:38	2018-11-09	80,92
1	2.648.010	ACERINOX	17:38	2018-11-09	9,364
2	592.455	ACS	17:38	2018-11-09	34,74

Tabla 3. Ejemplo de scraping de la fluctuación del euro

	CURRENCY_1	CURRENCY_2	VALUE	DATE	HOURL
0	EUR	USD	1,1336	09/11/2018	19:11
1	EUR	GBP	0,8731	09/11/2018	19:11
2	EUR	JPY	128,972	09/11/2018	19:11

Estos datos han sido recogidos de tablas y listas encontradas por la web cada 15 minutos durante una hora para esta práctica. Sin embargo, para realizar un buen estudio y realizar una buena imputación con algoritmos predictivos posteriormente es interesante realizar esta recopilación de datos constantemente e ir ampliando nuestro conjunto de datos con el paso de los meses y años.

6. Agradecimientos. ¿Quién es propietario del conjunto de datos? Includ citas de investigación o análisis anteriores.

La página dónde se ha realizado el scraping es www.infobolsa.es y gracias a ellos hemos podido recopilar esta información tan útil. Sin embargo, el propietario de los datos es la empresa FactSet:

"FactSet Terms of Use The FactSet Research Systems Inc. ("FactSet") on behalf of itself and its affiliates agrees to provide and User agrees to use the FactSet Data described below according to the terms and conditions below."[1]

[1] FACSET. <http://www.infobolsa.es/pdf/facsetDisclaimer.pdf>

7. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?

Este conjunto de datos junto una imputación con algoritmos de predicción de series temporales nos pueden ayudar a realizar estudios y poder realizar predicciones futuras que nos indicarán si es interesante comprar o vender de una empresa en concreto en el futuro. Esto es posible porque no solamente tenemos los valores de las empresas sino también la fluctuación de las monedas. De esta manera si un usuario tiene euros y desea realizar una compra o venta en un futuro sobre una empresa que cotiza en dólares necesitará saber no solamente el valor de esa empresa en ese momento futuro sino también conocer el valor que tendrá la divisa en ese momento.

8. Licencia. Seleccionad una de estas licencias y decid porqué la habéis seleccionado:

o **Other (specified above):** según la empresa FactSet la licencia de la base de datos, la consulta y el software están sujetos a lo siguiente:

"1. License of Databases, Consulting and Software

• Subject to the terms and conditions of this Agreement, FactSet grants to User the limited, nonexclusive, nontransferable rights to use its financial data ("FactSet Data") via this website.

• All proprietary rights, including intellectual property rights, in the FactSet Data will remain property of FactSet."

9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset.

Adjuntamos los archivos *"StocksWebScraping.py"* y *"currencies.py"* con el código pedido en un archivo ZIP. Para ejecutar solo es necesario ejecutar cada *"*.py"* por separado. El primero recoge los valores de las bolsas IBEX35 y NASDAQ; el segundo la fluctuación de las monedas.

También se pueden ver en: ["github.com/joaquinfdez/StocksValuesScraping"](https://github.com/joaquinfdez/StocksValuesScraping)

10. Dataset: Dataset en formato CSV.

Adjuntamos tres CSVs generados por estos programas: *"ibex35.csv"*, *"nasdaq.csv"* y *"currencies_stocks.csv"*.