

WebScraping con Python 3.6

Práctica 1, Web Scraping

Descripción

El siguiente código en Python 3.6 tiene por objetivo aplicar las técnicas de Web Scraping para extraer datos de la web de la Bolsa de Valores de Lima y generar como resultado un archivo de datos en formato CVS con los datos obtenidos.

Esta práctica pertenece a la asignatura Tipología y ciclo de vida de los datos, correspondiente al Máster en Ciencia de Datos de la Universitat Oberta de Catalunya.

Miembros del equipo

Patricia Reyes Silva

José Pérez Sánchez

Características del Dataset

El conjunto de datos a obtener, representa los datos referidos a las cotizaciones diarias de las acciones e índices cotizados en la Bolsa de Valores Lima de forma diaria.

La práctica se articula con el siguiente esquema:

1 Título: "Valores negociados en al Bolsa de Lima"

2 Subtítulo: "Cotizaciones diarias"

3 Imagen:



4 Contexto y justificación

En la actualidad, las nuevas tecnologías hacen posible acceder a los mercados de valores del mundo simplemente visitando sus respectivas websites, sin movernos de nuestros escritorios. Pero ¿qué pasaría si se requiere conocer los datos de las cotizaciones diarias de más de una bolsa de valores en las que se desea invertir? De la forma manual, se debería visitar diariamente cada una de las websites de interés y recoger manualmente los datos que necesitamos, pero existe también la modalidad automática de extracción de datos, que se vale de la ejecución de scripts para su recopilación. En nuestro contexto, donde los datos se generan a velocidades cada vez mayores (Big Data), todas las herramientas que nos permitan acelerar la recopilación y posterior tratamiento de los mismos son bienvenidas.

La presente práctica tiene como objeto desarrollar un Script en el lenguaje de programación Python que permita la extracción diaria de los datos correspondientes a las cotizaciones de la Bolsa de Valores de Lima, a partir de los datos publicados en su página web.

El uso de lenguaje Python nos parece realmente conveniente para esta tarea, no sólo por su utilidad para realizar Scraping (escarbar) dentro de páginas web, sino también por su enorme potencial para el manejo de grandes conjuntos de datos, para lo cual incluye potentes librerías que facilitan su tratamiento permitiendo la elaboración de distintos modelos de análisis.

5 Contenido

Incluye el conjunto de acciones negociadas en la Bolsa de Lima, para cada valor negociado se obtendría, diariamente, la siguiente información:

Agradecimientos

A la Bolsa de Lima y su infraestructura informática.

Inspiración

El auge de las compañías FinTech y del trading algorítmico, así como la disposición de datos.

Licencia

El código python incluido en los scripts está liberado con licencia GPL3, con permiso de modificación, incluso para uso comercial, distribución, etc., pero con la exención de ninguna responsabilidad por su uso ni ninguna garantía.

Los datos generados mediante este script se liberan bajo licencia "CC BY-SA 4.0 License.", con la siguiente salvaguardia legal:

Los datos obtenidos y almacenados en los ficheros cvs no son necesariamente en tiempo real ni tienen porqué ser totalmente exactos. En este sentido, los autores del script de carga de datos no tendrán ninguna responsabilidad ante cualquier pérdida que pueda tener como consecuencia de utilizar estos datos.

No se aceptará ninguna responsabilidad por cualquier pérdida o menoscabo producido como resultado de la confianza en la información contenida en estos datos, incluidos datos o cotizaciones.

Código

El código esta implementado en el fichero de código python CotizacionesBVL.py

Este código tiene dos funcionalidades diferentes. La principal es la obtención de las cotizaciones diarias de las compañías listada en la Bolsa de Valores de Lima. Para ello basta ejecutar el script sin parámetros

```
python CotizacionesBVL.py
```

Esto generará un fichero en formato CSV con las últimas cotizaciones de las compañías y nombre en formato: CotizacionesDiarias_YYYYMMDD.csv .

La segunda funcionalidad implementada permite obtener las cotizaciones para una empresa determinada, pasando su nemónico o símbolo de cotización y, opcionalmente, un rango de fechas.

Por ejemplo

```
python CotizacionesBVL.py --nemonico BVN --endDate 20180101 --startDate 20140501
```

Para obtener todas las cotizaciones de la empresa BVN, Minera Buenaventura, entre el 01/05/2014 y 01/01/2018, ambos días inclusive si hubo mercado esos días.

Esta opción genera un fichero con nombre en formato CotizacionesDiarias_nemonico.csv, siendo nemónico el correspondiente a la empresa consultada.

Dataset

El dataset incluido en los ficheros CSV generados es idéntico tanto para las cotizaciones diarias de todos los valores negociados en la Bolsa de Lima, como para descargas de cotizaciones de una empresa en un rango de fechas. No obstante, los campos indicados con un asterisco, *, sólo están presentes en el primer caso, siendo un campo vacío en las cotizaciones históricas de una empresa.

- Datos de la cotización

Fecha-Hora	Fecha de Generación del fichero
Fecha cotización	Fecha de cotización de la fila.
Imagen	icono azul si hay subida en el día o rojo en bajada *
Estado	Texto indicando tendencia del día al alza o a la baja *

- Valor negociado

Nombre (Empresa, producto, fondo, índice)

Nemónico Símbolo representativo del valor

Sector (Diversas, Agrario, Industriales, Bancos – financieras, etc..)

Segmento (Clasificación del valor principalmente por su liquidez)

- Cotizaciones

Moneda	Sol / Dólar
Precio Anterior	Precio de cierre día anterior
Fecha Anterior	Fecha de negociación anterior (no necesariamente día anterior)
Apertura	Precio apertura actual.
Última	Precio último del día.
Variación	Variación del precio en % respecto al día anterior, *

- Propuestas

Compra	Precio mayor de Compra en el día *
Venta	Precio menor de Venta en el día *

- Negociación

Número Acciones	Volumen de negociación
Núm. Operaciones	Realizadas en la jornada para este valor
Monto Negocio	Monto/importe negociado en la moneda de cotización

Recursos

- 1.Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- 2.Mitchel, R. (2015). Web Scraping with Python: Collecting Data from the Modern Web. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.
- 3.[Bolsa de Valores de Lima \(BVL\)](#)