

Miembros del grupo: Víctor Camí Núñez

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

Evolución con el tiempo de los índices bursátiles.

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

Adquisición de los índices bursátiles de principales bolsas mundiales a lo largo de un período de tiempo predeterminado.

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



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

El dataset contiene los índices bursátiles de las principales bolsas de Europa, Estados Unidos, Latinoamérica y Asia.

Estos índices bursátiles varían de forma constante debido a la evolución de los mercados o incluso de noticias que aparecen en los medios de comunicación.

Un caso reciente en España es la sentencia que parece que obligará a los bancos a pagar los impuestos de las hipotecas, esto ha provocado que el Ibex35 lleve perdiendo puntos, en especial los bancos, desde hace una semana.

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

Los campos que incluye el dataset son:

- Continente: nombre del continente al cual pertenece la bolsa de la que se obtiene el índice bursátil.
- Nombre: nombre de la bolsa de la cual se obtiene el índice bursátil.
- Tendencia: indicador que marca si la bolsa está subiendo o bajando. Dicho indicador no se obtiene del web scraping sino que toma el valor 'Sube' o 'Baja' dependiendo de si el campo 'Variación puntos' es positivo ('Sube') o negativo ('Baja').
- Puntos: puntos de cotización de la bolsa actualmente.
- Variación (%): variación porcentual de los puntos de la bolsa con respecto al valor anterior de los puntos. Dicho campo puede ser tanto positivo como negativo.
- Variación puntos: variación de los puntos de cotización actuales con respecto a los anteriores. Dicho campo puede ser tanto positivo como negativo.
- Puntos Anterior: último valor de la cotización del cierre anterior.
- Hora: hora de actualización de los índices bursátiles. Dicho campo puede no estar informado si los datos se obtienen de un día en el cual alguna de las bolsas no está abierta, por ejemplo, un fin de semana.
- Fecha: fecha del último día en el cual se han obtenido los datos. La fecha de este campo puede no ser la fecha de hoy debido a que hay días en los que las bolsas cierran. Dicho campo siempre está informado.
- Fecha descarga: fecha en la que se realiza la obtención de los datos. Este dato no se obtiene de realizar web scraping sino que se pone en cada registro para dejar constancia de cuando se obtuvo.

Estos datos se han recogido mediante web scraping de la página <https://www.eleconomista.es/indices-mundiales/>.

El programa de Python permite escoger el número de veces que se quieren obtener los datos y el tiempo entre cada obtención de datos.

Para poder ver la variación de los índices se ha decidido que se obtengan los datos una vez cada hora durante 10 horas.

Además de generar este dataset, se han descargándolos gráficos con la evolución de las siguientes bolsas en un día en concreto:

- DOW JONES

- IBEX 35
- IGBC
- IPC
- NASDAQ

Hay que tener en cuenta que estos gráficos van cambiando con el tiempo por lo que es conveniente intentar realizar la descarga cuando supongamos que ya han cerrado las bolsas. Soy consciente que las imágenes no tienen gran calidad y son algo pequeñas pero son las que había en la página web.

Estos gráficos tienen en el eje de las X las diferentes horas del día y en el eje de las Y tienen el valor de los puntos de dicha bolsa.

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

Los datos de los índices bursátiles se han obtenido del periódico digital el Economista, les agradezco que no hayan bloqueado dicha página en la que tienen los índices bursátiles en su fichero de robots.txt.

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

Vivimos en un mundo muy cambiante y la bolsa no es una excepción. Cualquier sentencia, ley nueva o noticia que circula por Internet, las redes sociales o los medios de comunicación pueden afectar, ya sean ciertas o no dichas noticias, a los mercados y por ende a las bolsas mundiales.

Un ejemplo de ello, como ya se ha comentado, es la reacción a la baja que están teniendo el Ibex35, especialmente los bancos, en respuesta a la sentencia judicial referente a quien ha de pagar los impuestos de las hipotecas.

El uso de este dataset se englobaría en un hipotético proyecto más complejo. Dicho proyecto intentaría en primer lugar relacionar las subidas o bajadas de las bolsas con las principales noticias económicas.

Una vez establecida dicha relación, se podría generar el modelo obtenido para la toma de decisiones sobre la compra o venta de acciones en función de las noticias económicas porque permitiría predecir si dichas noticias van a afectar positivamente o negativamente.

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

- Released Under CC0: Public Domain License
- Released Under CC BY-NC-SA 4.0 License
- Released Under CC BY-SA 4.0 License
- Database released under Open Database License, individual contents under Database Contents License
- Other (specified above)

○ Unknown License

Para el conjunto de datos escogería la licencia: Released Under CC BY-NC-SA 4.0 License. Las características de este tipo de licencia son:

- Se pueden copiar y distribuir los datos en cualquier medio.
- Se ha de reconocer al propietario de los datos tanto si se emplean para uso no comercial como si se copian, adaptan o modifican.
- Además, si una persona modifica los datos, estos están bajo el mismo tipo de licencia por lo que no se permitiría el uso comercial de los datos derivados de cualquier transformación de los originales.
- No permite el uso comercial de los datos

Se ha escogido este tipo de licencia es que no permite el uso comercial de los datos. Dato que los datos se van a emplear en un proyecto que tiene la finalidad ya explicada es necesario que nadie más los utilice para sacar beneficio de ningún tipo.

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.

El programa que genera el dataset se llama generador-dataset.py, dicho código se ha escrito para Python 2.7 pero se ha ejecutado también en 3 de Python instalada en mi ordenador sin problemas. No obstante se recomienda el uso de la versión 2.7 para evitar cualquier tipo de problemática asociada a incompatibilidades dado que hay diferentes versiones 3 de Python.

10. Dataset: Dataset en formato CSV

El fichero del dataset tiene el nombre *datos_indices(fecha_actual).csv* donde *fecha_actual* se ha de substituir por la fecha en la cual se realiza la ejecución del fichero.

Dicho fichero contiene una cabecera con el nombre de los campos y los propios campos separados por ‘,’.

Referencia externas empleadas en la realización de esta práctica

- <https://stackoverflow.com/questions/8286352/how-to-save-an-image-locally-using-python-whose-url-address-i-already-know>
- <https://www.youtube.com/watch?v=ng2o98k983k>