

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

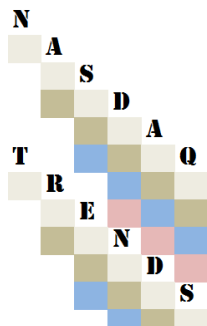
NASDAQ TRENDS. Monitorización de volúmenes en el NASDAQ

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

Obtención de datos para la búsqueda de desviaciones respecto al volumen normal durante los días anteriores en el mercado NASDAQ100. Mediante un diseño modular se ofrece la posibilidad de detectar desviaciones respecto a la tendencia general, con el objetivo de poder detectarlas a tiempo y obtener beneficios.

La entrega incluye la obtención de los datos procedentes de tres fuentes, Google, Yahoo y Nasdaq, pero se pueden añadir nuevas fuente fácilmente.

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



(Imagen libre de derechos, hecha por mi)

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

El contexto sería la búsqueda de la maximización de los beneficios (o minimización de las pérdidas) mediante la inversión en bolsa.

Sin duda existen muchos métodos más complejos para decidir compras y ventas, pero el nivel de variación de movimientos respecto a la media de los meses anteriores me parece un indicativo que podría ser válido.

La materia serían los movimientos de las acciones, a nivel diario y a nivel promedio de los últimos meses.

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

Se incluye el momento de toma de la muestra y, para cada uno de las páginas de las que se obtienen datos, la hora de los datos, el nivel actual de movimiento de acciones y el nivel promedio de movimientos que le constan a la página.

Si bien se ha tomado varias medidas, el objetivo de este scrapper sería poder detectar acciones que tengan un nivel de negociación en lo que va de jornada superior al que tienen como media. Eso daría una 'alarma' al usuario, que podría detectar movimientos a los que podría apuntarse.

La recogida de datos se ha hecho durante varios días, tanto con la bolsa cerrada como abierta, con el objetivo de poder limpiar los datos y poder controlar posibles excepciones.

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

Como agradecimiento, las fuentes de donde tomo los datos, Google, Yahoo y Nasdaq. Y doble agradecimiento a Google, con cuyo GoogleSheets y su función importXML desarrollé un versión anterior que ha resultado de mucha utilidad para esta práctica.

En este caso la propiedad de los datos es de NASDAQ (National Association of Securities Dealers Automated Quotation), que es la compañía que monitoriza los movimientos y, por tanto, tiene constancia de las fluctuaciones. De acuerdo con la Wikipedia (<https://es.wikipedia.org/wiki/NASDAQ>) Nasdaq como bolsa es propiedad de la compañía NASDAQ Stock Market.

Para hacernos una idea, a nivel español tenemos la Bolsa de Madrid. La empresa que gestiona la Bolsa de Madrid es Bolsas y Mercados Españoles (que sería el equivalente a Nasdaq Stock Market). Y decide generar un índice usando las 35 empresas más capitalizadas, el IBEX-35, (que sería el equivalente al NASDAQ100).

Para más información, <https://www.bolsasymercados.es/esp/Sobre-BME/Que-es>

Por tanto, el conjunto de datos final es de NASDAQ... ¿seguro?

Pues no. Nasdaq habrá licenciado el derecho de acceso a los datos a terceros, en este caso a Google, Yahoo y la propia web de Nasdaq... Por tanto los datos también podría considerarse suyos.

También podrían ser del autor (Rodrigo Minguito Linaje), por haberlos recopilado, modificado, limpiado y agregado.

Otra de las posibilidades es que pudieran llegar a ser propiedad de la UOC, pero esto queda descartado ya que no se trata de una beca.

En cualquier caso, dejaría el análisis legal a un abogado especializado o, al menos, alguien con Curso de Propiedad Intelectual

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

La información obtenida no responde a ningún interés altruista, de beneficio a la comunidad. Simplemente busca la posibilidad de maximizar el rendimiento del capital intentando encontrar patrones en los movimientos bursátiles.

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

No sé si los datos son de mi propiedad (punto 6), por lo que no puedo licenciarlos. De hecho, es muy posible que su obtención no sea del todo legal, ya que alguna de las propias páginas web de donde vienen indican que se debe obtener su permiso expreso para disponer de ellos.

En el supuesto de tener la propiedad de los datos y no estar sujeto a ninguna licencia anterior superior (cosa dudosa), elegiría CC0, ya que indica que no mantengo ningún derecho y, realmente no “siento” que los datos sean de mi propiedad ni merezca ningún reconocimiento por ellos.

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

(adjunto)

10. Dataset: Dataset en formato CSV

(adjunto)