

Práctica 1 M2.851 - Tipología y ciclo de vida de los datos

Autores:

Waziri Ajibola Lawal
David Fernández González

Índice

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.	2
2. Definir un título para el dataset. Elegir un título que sea descriptivo.	2
3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título Elegido).....	3
4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.....	4
5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.	5
6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).	6
7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.....	6
8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:	6
9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.	7
10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.....	7

Práctica 1 M2.851 - Tipología y ciclo de vida de los datos

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

Hemos realizado un pequeño proyecto usando Python, para recoger automáticamente información sobre las empresas del Ibex 35. Entre la información recogida mostramos la última transacción, el volumen, el máximo y el mínimo diario. Además, recolectamos también un histórico desde enero del 2020, mostrando el valor a cierre, el cambio, cambio %, máximo y mínimo y el volumen. También mostramos un gráfico con la cotización desde enero 2020 hasta noviembre,

Con todos estos datos podemos ver la evolución de las empresas del Ibex 35. Podemos ver cómo ha afectado el Covid-19 a la evolución de las cotizaciones.

Vemos por ejemplo cómo impactó a IAG



2. Definir un título para el dataset. Elegir un título que sea descriptivo.

Títulos en Zenodo:

1. Historical quotation of Ibex 35 companies from January 2020 to November 2020

Gráficos con el histórico de cotizaciones de las empresas del Ibex 35 desde Enero 2020 hasta Noviembre 2020

2. Listed values of Ibex 35 companies from January 2020 to November 2020

Datos de las compañías del Ibex 35 desde Enero 2020 hasta Noviembre 2020:

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título Elegido).

El dataset muestra los valores de cotización de las empresas del Ibex 35 desde Enero de 2020 hasta Noviembre del mismo año.

El dataset incluye un fichero que contendrán información de las empresas del Ibex 35 (ticker, nombre, última transacción, volumen, máximo diario y mínimo diario). También incluye un conjunto de ficheros con un histórico de cotizaciones de cada empresa (día, cierre, cambio, cambio%, máximo, mínimo y volumen). Y, para completar la información visual, se incluye un conjunto de imágenes para cada empresa (histórico de cotizaciones)

Práctica 1 M2.851 - Tipología y ciclo de vida de los datos

4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.

DÍA	CIERRE	CAMBIO	CAMBIO%	MÁXIMO	MÍNIMO	VOLUMEN
6/11/2020	89.75 €	3.05	3.29%	95.60 €	88.65 €	161175
5/11/2020	92.80 €	1	1.09%	93.90 €	91.65 €	67307
4/11/2020	91.80 €	0.15	0.16%	92.15 €	88.40 €	98194
3/11/2020	91.95 €	3.25	3.66%	91.95 €	89.60 €	72924
2/11/2020	88.70 €	2	2.31%	89.40 €	85.70 €	97353
30-10-20	86.70 €	0.2	0.23%	86.70 €	85.15 €	137172
29-10-20	86.90 €	3	3.34%	91.10 €	86.70 €	139018
28-10-20	89.90 €	2.15	2.34%	91.60 €	89.30 €	201994
27-10-20	92.05 €	3	3.16%	95.25 €	91.70 €	73222
26-10-20	95.05 €	1.05	1.09%	96.70 €	94.65 €	46314
23-10-20	96.10 €	1.95	2.07%	96.30 €	93.95 €	76997
22-10-20	94.15 €	1.2	1.26%	95.10 €	93.40 €	59041
21-10-20	95.35 €	2	2.05%	98.75 €	95.00 €	81650
20-10-20	97.35 €	0.15	0.15%	98.25 €	96.30 €	52781
19-10-20	97.20 €	0.2	0.21%	98.75 €	97.20 €	58663
16-10-20	97.40 €	0.85	0.88%	97.60 €	96.10 €	97327
15-10-20	96.55 €	1.05	1.08%	97.40 €	95.65 €	80398
14-10-20	97.60 €	0.5	0.51%	98.05 €	95.65 €	93656
13-10-20	97.10 €	1.35	1.37%	98.40 €	96.65 €	54427



5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

Los campos que incluye el datasets son:

- **Día:** fecha en la que se calcula los valores de cotización
- **Cierre:** última cotización que registró durante el día en el mercado bursátil
- **Cambio:** variación en el valor de cotización durante el día en el mercado bursátil
- **Cambio %:** porcentaje de variación en el valor de cotización durante el día en el mercado bursátil
- **Máximo:** precio máximo de cotización que ha alcanzado durante el día en el mercado bursátil
- **Mínimo:** precio mínimo de cotización durante el día en el mercado bursátil
- **Volumen:** cantidad de títulos negociados durante el día en el mercado bursátil

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

El propietario de los datos que se hemos utilizado es www.bolsa.es. Nos ofrece la posibilidad de poder recolectar datos para crear un dataset de los valores de cotización de las empresas del Ibex 35.

Además, hemos consultado el siguiente material:

- Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC.
- Masip, D. (2010). El lenguaje Python. Editorial UOC.
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- BeautifulSoup
- Selenium

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

Hemos pensado que ver la evolución de las cotizaciones de las empresas del Ibex 35 sería un buen enfoque para la realización de la práctica. Se puede ver cómo ha afectado la crisis provocada por el Covid-19 a todas estas empresas.

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

Entendemos que cualquiera puede acceder y recolectar los datos en la página web www.bolsa.es pero debido a que son datos sensibles y reales de empresas de la bolsa consideramos que no se puede hacer uso comercial de los datos sin los consentimientos de estas empresas.

Por lo tanto, la licencia para nuestro dataset es “**Released Under CC BY-NC-SA 4.0 License**”. Se puede compartir y adaptar el material siempre que se reconozca la autoría del documento, indicar los cambios que se han realizado, además de no hacer un uso comercial de estos.

9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

El código fuente se encuentra aquí

- https://github.com/fernandezgonzalezd/ciclo_vida_datos

10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

Los DOI de publicación en Zenodo son:

- 10.5281/zenodo.4262291 *Historical quotation of Ibex 35 companies from January 2020 to November 2020*

Los histogramas muestran los valores de las compañías del Ibex35 desde enero del 2020 a noviembre 2020

- 10.5281/zenodo.4262200 *Listed values of Ibex 35 companies from January 2020 to November 2020*

Para cada empresa se listan los valores diarios de cierre, cambio, cambio %, máximo, mínimo y volumen

Práctica 1 M2.851 - Tipología y ciclo de vida de los datos

Contribuciones	Firmas
Investigación previa	WAjibolaL,DFdezGlez
Redacción de las respuestas	WAjibolaL,DFdezGlez
Desarrollo código	WAjibolaL,DFdezGlez