

Sandra Garrido Romero

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

Actualmente he sentido un creciente interés por invertir en acciones de bolsa por lo que me estoy formando al respecto. Recabar información de la bolsa de Madrid en tiempo real es algo que me puede ayudar bastante a entender cómo funciona.

El sitio web de la bolsa de Madrid (<http://www.bolsamadrid.es/esp/asp/Portada/Portada.aspx>) contiene bastante información al respecto. Para esta práctica, he seleccionado los datos de las empresas que cotizan en el IBEX 35. Estos datos se encuentran en forma de tabla dentro de la página web (<http://www.bolsamadrid.es/esp/asp/Mercados/Precios.aspx?indice=ESI10000000&punto=indice>) por lo que la búsqueda y extracción de datos es relativamente sencilla.

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

El *dataset* lo voy a llamar preciosAcciones\_IBEX\_35 ya que se van guardando los precios de cada acción de las empresas que forman el IBEX 35.

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* contiene la información de cada una de las empresas que componen el Ibex35. Esta información se actualiza cada cierto tiempo en la página web mencionada en el primer apartado. La información contenida en el dataset cuenta con un registro por cada una de las empresas, además de los datos más importantes sobre las operaciones realizadas a lo largo de cada sesión.

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

La siguiente tabla es la que aparece en la página web y de dónde se han obtenido los datos:

Nombre	Últ.	% Dif.	Máx.	Min.	Volumen	Efectivo (miles €)	Fecha	Hora
▲ ACCIONA	93.0000	0.32	93.8000	92.3500	27.006	2.514.78	07/11/2019	13:07:10
▲ ACERINOX	9.3500	0.86	9.4500	9.3100	731.655	6.882.59	07/11/2019	13:07:35
▼ ACS	36.1400	-1.63	36.9900	36.1400	180.732	6.605.89	07/11/2019	13:07:08
▲ AENA	167.1000	0.42	168.0500	166.3500	16.510	2.762.80	07/11/2019	13:07:29
▲ AMADEUS	69.9600	2.01	71.9200	69.7600	307.918	21.796.17	07/11/2019	13:06:06
▲ ARCELORMIT.	15.4720	6.70	15.7700	15.2000	1.649.595	25.527.81	07/11/2019	13:07:35
▲ BA SABADELL	1.0400	1.07	1.0530	1.0230	11.300.795	11.745.76	07/11/2019	13:07:45
▲ BA SANTANDER	3.8275	1.59	3.8655	3.7730	68.616.380	260.620.16	07/11/2019	13:07:37
▲ BANKIA	1.8395	2.48	1.8545	1.7920	3.669.909	6.707.43	07/11/2019	13:07:44
▲ BANKINTER	6.5020	1.69	6.5640	6.3760	855.177	5.539.67	07/11/2019	13:07:31
▲ BBVA	4.9625	1.19	4.9980	4.8850	6.766.512	33.411.26	07/11/2019	13:07:31
▲ CAIXABANK	2.7820	1.98	2.8140	2.7380	7.533.455	20.931.55	07/11/2019	13:07:37
▼ CELLNEX	37.5400	-0.42	37.8200	37.3000	152.648	5.733.07	07/11/2019	13:07:12
▲ CIE AUTOMOT.	23.1800	0.61	23.4400	23.1200	76.551	1.781.20	07/11/2019	13:05:05
▲ ENAGAS	22.7800	1.24	23.1800	22.7000	574.657	13.140.22	07/11/2019	13:04:38
▲ ENCE	3.9680	5.14	4.0100	3.7680	3.493.074	13.734.87	07/11/2019	13:07:27
▼ ENDESA	23.8700	-0.67	24.1500	23.8600	287.033	6.885.48	07/11/2019	13:07:31
▼ FERROVIAL	26.0100	-1.14	26.4300	25.9900	288.956	7.560.57	07/11/2019	13:07:38
▼ GRIFOLS CLA	28.8000	-1.03	29.2600	28.7400	142.137	4.116.94	07/11/2019	13:06:47
▲ IAG	6.3440	0.06	6.4640	6.3440	1.008.586	6.462.24	07/11/2019	13:06:54
▼ IBERDROLA	8.9520	-1.06	9.0480	8.9520	4.864.178	43.636.60	07/11/2019	13:07:37
▼ INDITEX	28.3700	-0.39	28.7900	28.3300	15.719.436	449.856.91	07/11/2019	13:07:29
▲ INDRAA	9.3050	1.14	9.3500	9.2000	370.977	3.443.90	07/11/2019	13:07:27
INM COLONIAL	11.4700	0.00	11.5300	11.3600	163.835	1.878.54	07/11/2019	13:06:33
▲ MAPFRE	2.5590	0.71	2.5670	2.5400	937.591	2.393.22	07/11/2019	13:07:31
▲ MASMOVIL	21.3800	0.85	21.7200	21.0800	371.786	7.935.92	07/11/2019	13:07:15
▼ MEDIASET	5.5340	-0.25	5.5720	5.5100	203.209	1.125.45	07/11/2019	13:07:18
▲ MELIA HOTELS	7.4350	0.47	7.5350	7.4000	258.803	1.931.80	07/11/2019	13:07:35
▼ MERLIN	13.0600	-0.46	13.1700	13.0300	256.556	3.358.46	07/11/2019	13:03:14
NATURGY	24.4400	0.00	24.6600	24.3900	337.874	8.276.91	07/11/2019	13:05:47
▲ R.E.C.	17.9750	0.08	18.0700	17.9250	556.895	10.025.85	07/11/2019	13:07:15
▲ REPSOL	15.2000	0.46	15.2450	14.9200	2.639.675	40.002.13	07/11/2019	13:07:29
▲ SIEMENS GAME	12.5000	3.91	12.5900	12.0100	1.903.661	23.661.30	07/11/2019	13:06:47
▲ TELEFONICA	6.8180	1.08	6.8250	6.7310	3.310.518	22.477.80	07/11/2019	13:06:50
▲ VISCOFAN	49.1400	0.53	49.3800	48.7000	43.553	2.141.47	07/11/2019	13:07:35

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

Por cada una de las empresas se guarda el nombre, el último valor que ha tenido la acción, el porcentaje de la diferencia entre subidas y bajadas, el máximo y el mínimo valor que ha tenido la acción, el volumen de operaciones realizadas, el efectivo en miles de euros de las operaciones realizadas, la fecha y la hora en las que se han recogido los datos.

El período de tiempo de los datos es de un día ya que el proceso está pensado para realizarse al finalizar la sesión de banca diaria. Aunque se podría realizar en cualquier momento para obtener los datos lo más actualizados posible.

Los datos se han recogido mediante la técnica de *Web Scraping* utilizando *Python* como lenguaje de programación. Primero, se ha recuperado la información de la página web mediante la librería *Requests*. Después, se parsea la información con la librería *BeautifulSoup*. Una vez parseada, se obtiene la tabla en la que se encuentran los datos que se quieren obtener. A continuación, se realiza un bucle que recorre tanto las filas como las columnas para ir guardando en variables los datos de cada una de las empresas en un archivo CSV.

**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 es la sociedad del grupo BME que es la responsable de la página web y de toda la recogida y tratamiento de datos.

No ha sido necesaria una investigación previa ya que en la página web de la bolsa de Madrid aparece la política de privacidad

(<http://www.bolsamadrid.es/esp/privacidad/Politica-de-Privacidad>).

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

Este conjunto de datos es bastante interesante ya que muestra cómo varían las acciones y las operaciones que se realizan en bolsa a lo largo de un día o varios. Permite obtener los datos necesarios para poder realizar predicciones sobre la subida o bajada de acciones. Además, pueden realizarse gráficas con todos los datos recogidos que muestre por cada una de las empresas que tendencias tiene a lo largo de un mismo día o de varios, dependiendo de cuando se ejecute el *script*.

Las preguntas que pretende responder son aquellas derivadas de la observación de la bolsa de Madrid. Este *dataset* puede servir de bastante ayuda a personas que trabajen en un ámbito económico y necesiten estar al día de las operaciones del IBEX 35, sin necesidad de buscar en internet. Simplemente ejecutando el *script* pueden tener un archivo Excel todos los datos necesarios.

No sólo sirve para personas que se dediquen a economía. Cualquier persona que tenga interés en inversiones en bolsa puede tomar el conjunto de datos como referente para predecir el nivel de las acciones como he mencionado anteriormente.

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

- 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

La licencia escogida es CC0 (*Public Domain License*). Los datos que he obtenido son de una fuente pública que puede utilizar cualquier persona y no les he realizado ninguna modificación. Debido a ello el conjunto de datos puede distribuirse libremente, copiarse y utilizarse para fines comerciales. Incluso, los datos pueden utilizarse para crear datos derivados de ellos. En definitiva, el *dataset* es de dominio público.

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

El código utilizado ha sido *Python*. El archivo se encuentra dentro del repositorio *GitHub*.

10. **Dataset.** Presentar el *dataset* en formato CSV.

El *dataset* se ha exportado a un archivo CSV que se encuentra dentro del repositorio *GitHub*.

Contribuciones	Firma
Investigación previa	SGR
Redacción de las respuestas	SGR
Desarrollo código	SGR