

Web Scraping: Valores históricos de precios de criptomonedas

Rafael López García y Carlos Luis Gento de Celis

01/11/2021

Índice

1	Descripción del dataset	2
2	Contexto	2
3	Representación gráfica.	2
4	Contenido	3
5	Agradecimientos	3
6	Inspiración	3
7	Licencia	4
8	Código y Dataset	5
9	Tabla de contribuciones	5

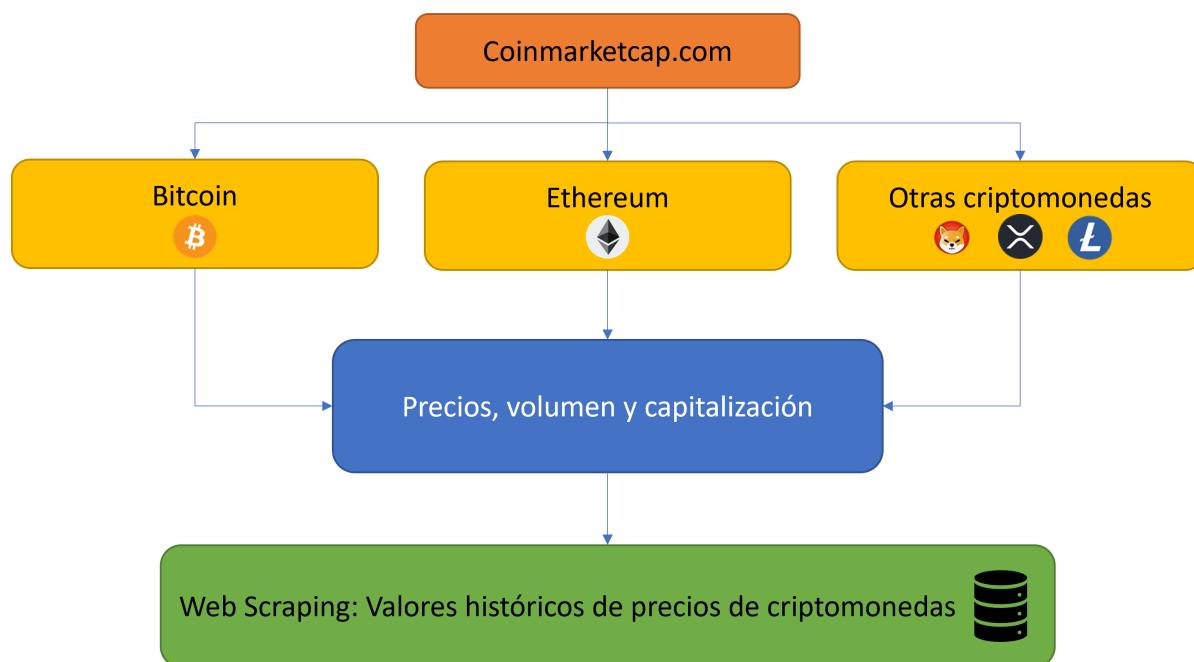
1 Descripción del dataset

El conjunto de datos generado con esta actividad contiene información diaria sobre el comportamiento de precios, volumen intercambiado y valor total de capitalización de más de 6000 criptomonedas durante año previo a la creación de este dataset. Entre las criptomonedas incluidas se encuentran las más conocidas, como Bitcoin, Ethereum o Litecoin.

2 Contexto

La información contenida en este conjunto de datos se ha extraído del sitio web *coinmarketcap.com*, que contiene información diaria sobre precios de criptactivos en circulación desde abril de 2013 hasta la actualidad. Se trata de uno de los sitios de referencia a la hora de seguir la evolución del valor de las criptomonedas, y es frecuentemente citado por investigadores, instituciones y prensa.

3 Representación gráfica.



4 Contenido

El dataset que aquí se presenta contiene información diaria de 6336 criptomonedas entre los días 31 de octubre de 2020 y 29 de octubre de 2021.

Para cada criptomoneda se recogen los siguientes campos:

- **symbol:** Código identificador de la criptomoneda
- **date:** Fecha de la sesión
- **open:** precio de la criptomoneda en el momento de la apertura de la sesión diaria
- **high:** precio máximo alcanzado durante la sesión diaria
- **low:** precio mínimo alcanzado durante la sesión diaria
- **close:** precio de la criptomoneda en el momento de cierre de la sesión diaria
- **volumen:** Volumen en dólares de las criptomonedas intercambiadas durante la sesión diaria
- **marketcap:** Valor de capitalización de la criptomoneda.

Tal y como se ha comentado anteriormente, los datos se han extraído del sitio web coinmarketcap.com. Para ello se han utilizado técnicas de Web Scraping y el lenguaje Python para extraer la información que estaba alojada en las páginas HTML.

5 Agradecimientos

Los datos para crear este conjunto de datos se han extraído de la web *coinmarketcap.com*. La información contenida en esta web se ha utilizado para realizar numerosos proyectos de recolección de información de precios de criptomonedas, como por ejemplo *Cryptocurrency Historical Prices* en *kaggle.com* o en *Cryptocurrency Price by Day/Hr (2013 - 2018) (w/Bitcoin)* en *data.world*. También se ha utilizado para realizar análisis y estudios, por ejemplo sobre la evolución de precios de criptomonedas (*Lansky 2016*) o para el desarrollo de modelos de predicción (*Cryptocurrency Predictions with ARIMA*, *kaggle*).

Para llevar a cabo la recolección de estos datos, se han tenido en cuenta los *términos de uso* publicados en *coinmarketcap.com*, y la información contenida en su fichero *robots.txt*, que permite el webcrawling de la web excepto en ciertos directorios a los cuales no hemos accedido a la hora de recopilar los precios históricos de las criptomonedas.

6 Inspiración

Las criptomonedas han generado mucho interés desde que la primera de ellas, el Bitcoin, se presentó en 2008 como un instrumento de pagos electrónicos entre particulares que no necesitaba a un tercer agente de confianza (generalmente alguna institución financiera) que procesase y verificase dichos pagos (*Nakamoto, 2008*). Sin embargo, tal y como expuso

el Nobel de Economía *Paul Krugman en 2013*, su éxito como moneda dependía tanto de ser aceptado como medio de cambio, como en ser considerado como un “almacén de valor” razonablemente estable (*store of value*).

El Bitcoin no ha tenido mucho éxito para ser aceptado como medio de cambio (la única jurisdicción que ha adoptado el Bitcoin como moneda de curso legal es El Salvador, el *7 de septiembre de 2021*), pero el crecimiento continuado de su valor en el largo plazo (*Baur and Dimpfl, 2021*) y el de otras criptomonedas como *Ethereum* ha hecho que aparezcan muchos defensores de las criptomonedas como “almacenes de valor”. Esto también parece reflejarse en el comportamiento de los inversores en criptomonedas, ya que muchos tienden a mantener sus inversiones en ellas (*Auer and Tercero-Lucas, 2021*). Sin embargo, la excesiva volatilidad de sus precios en el corto plazo, y en particular el del Bitcoin, (*Baur and Dimpfl, 2021*), generan muchas dudas y rechazo con respecto a que puedan considerarse tanto como medio de cambio como unidad contable.

Con este dataset se podrían responder las siguientes preguntas:

- ¿Cuánto ha variado el precio del Bitcoin en las últimas semanas, meses o años? ¿Y cuánto lo ha hecho el del resto de criptomonedas?
- ¿Hay patrones de comportamiento comunes y periódicos en los precios de las criptomonedas, como por ejemplo efectos estacionales?
- ¿Cuál es la criptomoneda que ha presentado mayor estabilidad en su precio durante el último año?
- ¿Cuál es la criptomoneda que ha presentado un mayor aumento en su precio durante los últimos meses?
- ¿Qué Criptomonedas están son las que han presentado mayor volatilidad en sus precios durante las últimas semanas? ¿Y los últimos meses?
- ¿Hay alguna criptomoneda cuyo precio sea razonablemente estable como para poder ser considerada “almacén de valor”?

Con respecto a los proyectos y análisis previos mencionados anteriormente en el apartado Agradecimientos, este dataset contiene información más cercana al momento actual, y por tanto puede contribuir a perfeccionar dichos análisis.

7 Licencia

La licencia seleccionada para la publicación de este conjunto de datos es **CCO: Public Domain License**. Las razones que han llevado a la elección de esta licencia son que hemos encontrado otros repositorios y proyectos con conjuntos de datos obtenidos de la misma fuente que utilizan dicha licencia (*Cryptocurrency Historical Prices* o *Cryptocurrency Price by Day/Hr (2013 - 2018) (w/Bitcoin)*, mencionados anteriormente) y que incluye términos que nos parecen adecuados con el trabajo realizado:

- Los autores han dedicado la obra al dominio público, mediante la renuncia a todos sus derechos a la obra bajo las leyes de derechos autorales en todo el mundo, incluyendo todos los derechos conexos y afines, en la medida permitida por la ley.

- Se puede copiar, modificar, distribuir e interpretar la obra, incluso para propósitos comerciales, sin pedir permiso.

8 Código y Dataset

El código fuente y el dataset pueden encontrarse en el repositorio Git accesible a través de este enlace: *Git*.

También se puede acceder al dataset en Zenodo a través del siguiente enlace: *Zenodo*.

9 Tabla de contribuciones

Contribuciones	Firma
Investigación previa	R.L.G y C.G.C
Redacción de las respuestas	R.L.G y C.G.C
Desarrollo del código	R.L.G y C.G.C