

Tipología y ciclo de vida de los datos

PRA1: Extracción de datos históricos de CoinMarketCap.

Autores:
Diego Álvarez Padrón
Kevin Mateo García

1. Contexto:

Que el mundo de las criptomonedas está en auge no es ningún misterio, en este marco, hemos decidido hacer un proyecto de web scraping utilizando la web CoinMarketCap, en concreto, [su espacio dedicado a valores históricos](#).

Hay infinidad de webs que ofrecen este servicio de forma abierta, (Binance, investing.com, cryptodatadownload...), sin embargo, CoinMarketCap es una de las webs de más fiabilidad en cuanto a datos de criptomonedas y además nos presenta la información relativa ordenada en un ranking por capitalización de mercado, hecho interesante a la hora del análisis de datos. En este espacio, encontramos datos desde el año 2013, cuando empezaban a recogerse datos de forma masiva de este tipo de activos. A partir de entonces, la web nos ofrece datos de una gran cantidad de monedas virtuales, según han ido apareciendo y cogiendo fuerza en cuanto a capitalización de mercado. Estos datos son recogidos por semana, habiendo en la mayoría de los meses cuatro tablas con información interesante.

En nuestro dataset generado (almacenado en el csv) encontramos datos desde finales del 2013 hasta octubre de 2021.

2. Título del dataset:

Cryptocurrency_History

3. Descripción del dataset

Nuestro programa se basará en un intervalo de fechas dado y extraerá los datos de las 20 criptomonedas con más capitalización de mercado en dicho intervalo. Estos datos tendrán como columnas:

Rank: Ranking en cuanto a capitalización de mercado en esa fecha.

Nombre: Nombre oficial de la criptomoneda.

Símbolo: Símbolo oficial único de cada criptomoneda.

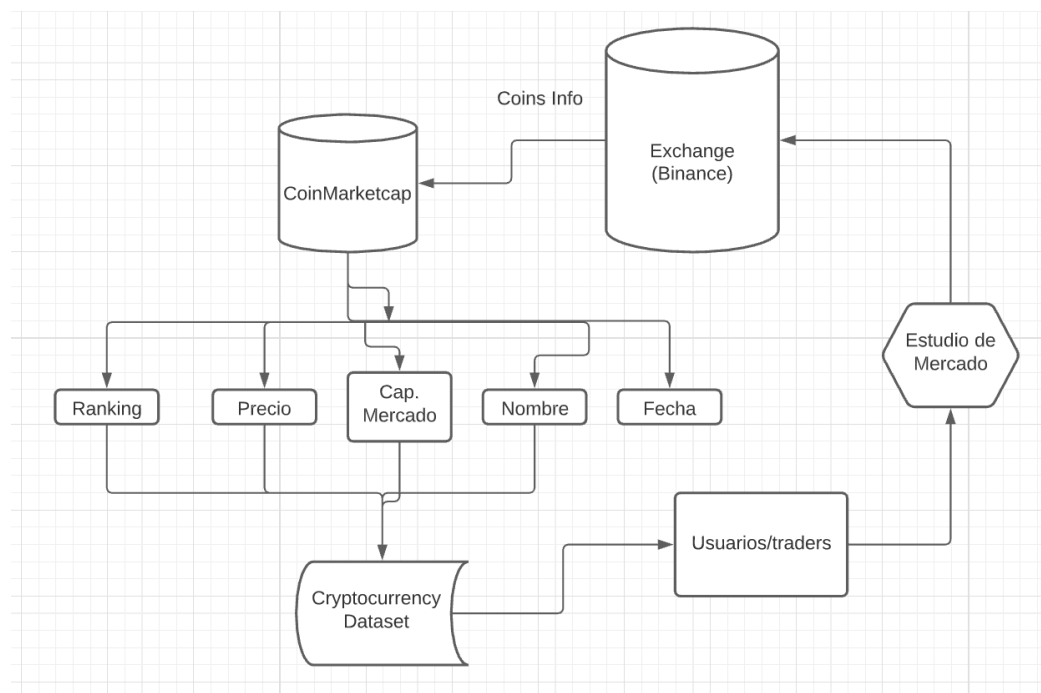
Cap. De Mercado: Es igual al precio por acción multiplicado por el número de acciones en circulación de la criptomoneda.

Precio: Precio de la criptomoneda en dicha fecha.

Cabe destacar que si bien en la web aparecen otros valores, para la actividad que nos ocupa, no serán tenidos en cuenta debido a su falta de interés en el análisis que nos disponemos a hacer. Por ejemplo, se incluyen valores de la variación del precio en las últimas horas, volumen... Teniendo en cuenta que los datos que vamos a extraer serán semanales y la volatilidad que caracteriza el valor de las criptomonedas, estos valores podrán ser despreciables en nuestros análisis.

Si quisiéramos realizar trading en vivo sí que sería interesante abordar un histórico de datos con temporalidad de horas, por ejemplo para la opción de Binance futuros pero es algo que se dejará por ahora a un lado.

4. Representación gráfica



5. Contenido

Nuestro dataset está compuesto por una serie de datos que engloban todo lo relacionado con el precio de las criptomonedas. Todo esto lo hemos agrupado por fechas delimitando los datos entre dos grupos de fechas, desde un momento en el pasado hasta una fecha superior (máximo el día de hoy). Cada fecha engloba el top 20 de criptomonedas con más músculo financiero de ese día ordenandolas de manera

descendente. **Tenemos el ranking que ocupa, nombre, el símbolo, capitalización de mercado y el precio con equivalencia en dólares.**

6. Agradecimientos

Esta obtención de datos ha sido posible gracias a la web <https://coinmarketcap.com/historical> donde se aloja la mayor parte de la información que existe de las criptomonedas, desde proyectos que están en busca de financiación para próximas salidas, eventos de criptomonedas hasta la criptomoneda más antigua. En todo momento hemos tratado con datos públicos y sin manipular, accesibles y comprobables en todo momento.

7. Inspiración

Es interesante tener un histórico amplio del precio de una criptomoneda a 10 años atrás para prever cuál será su posible comportamiento a 10 años vista o para ver monedas que se están iniciando ahora mismo como proyecto, cuál será su valor en 10 años. Con estos datos se puede hacer una estimación del comportamiento que tendrá el mercado y de cómo ha sido la adopción de las criptomonedas por parte de la gran masa de población y como lo será en un futuro.

8. Licencia

Escogemos la licencia CC BY-SA 4.0 License en relación a :

- Promueve el nombre del creador del conjunto de datos generado, se reconoce el trabajo ajeno y en qué medida se ha realizado aportaciones al original.
- Se permite un uso comercial controlado aportando reconocimiento al autor original
- Las contribuciones realizadas deberán publicarse bajo esta licencia.

9. Código

https://github.com/kmateo/crypto_historical_data

10. Dataset

<https://zenodo.org/record/5639736#.YYGgw2DP2bg>

Contribuciones	Firma
Investigación previa	DA,KM
Redacción de las respuestas	DA,KM
Desarrollo del código	DA,KM

- Enlace al vídeo explicativo:

<https://drive.google.com/file/d/13fISwAg1aI2yrwJX3aa3uZVcaXTdaOsR/view?usp=sharing>