

PRAC 1. WEB SCRAPING

Tipologia i cicle de vida de les dades

Sergi Canet Vela

Curs 2019-2020

Màster de Ciència de Dades / Universitat Oberta de Catalunya

1. Context

Aquest script consulta els preus actualitzats per comprar criptomonedes als diferents Exchanges que ofereixen aquest servei, emmagatzemant el dataframe resultant en un fitxer csv. Les dades s'extreuen del lloc web CoinMarketCap (<https://www.coinmarketcap.com>), el qual es un portal especialitzat en reportar informació actual de les diferents criptomònades del mercat.

2. Títol

"Preus del mercat de criptomonedes a temps real"

3. Descripció

Dataframe actualitzable format pel preu actual als Exchanges principals de fins a 8 criptomonedes escollides per l'usuari.

4. Representació gràfica



Font: <https://hackernoon.com/how-to-pick-the-best-cryptocurrency-exchange-bcabc359a9dd>

5. Contingut

- **exchange:** nom de l'exchange que ofereix alguna de les criptomonedes seleccionades per generar el dataframe.
- **{crypto}-USD:** preu de la criptomoneda en concret en relació al dòlar americà (parella criptomoneda-dòlar). Camp nullable (no tots els exchanges han d'oferir totes les criptomonedes seleccionades). Tindrem des de 1 fins a 8 parelles, segons l'entrada de l'usuari al executar l'script (per defecte si no es passa cap argument es mostren les columnes: bitcoin-USD, ethereum-USD, dash-USD, litecoin-USD, bitcoin-cash-USD).

6. Agraïments

Totes les dades provenen del portal web especialitzat en criptomonedes CoinMarketCap (<https://www.coinmarketcap.com>). El llistat de les criptomonedes disponibles s'extreu del següent enllaç:

- <https://coinmarketcap.com/coins/>

Una vegada l'usuari ha seleccionat quines criptomonedes vol s'obtenen els preus de cada Exchange consultant els enllaços corresponents per a cada cas. Prenent com a exemple el bitcoin, el format d'aquests enllaços es:

- <https://www.coinmarketcap.com/currencies/bitcoin/markets/>

7. Inspiració

Aquest conjunt de dades pot ser utilitzat per analitzar tendències en temps real sobre els diferents Exchanges, permetent als inversors analitzar quins són de forma general els més assequibles per a comprar o bé els més cars per vendre els seus actius i així ajudar a la presa de decisions per fer els moviments corresponents.

També podria servir de base de consulta per un bot d'arbitratge de criptomonedes, ja que l'algorisme es podria basar en consultes periòdiques a aquest dataframe, sempre que es vagi regenerant automàticament executant l'script per obtenir les dades més recents, amb l'objectiu d'automatitzar les decisions que es prendrien manualment de forma molt més lenta. Aquest algorisme encara es podria millorar fent servir eines de Machine Learning per a predir quins Exchanges són més lents en reaccionar a una pujada o baixada dels preus de mercat, de manera que se'n pugui treure més rendiment.

8. Llicència

He escollit fer ús de la llicència permissiva MIT, on s'otorguen els drets de us, copia, modificació, fusió, publicació, distribució, subllicència i venda, sempre que s'inclouï la nota del copyright.

He triat aquesta llicència abans que la CC BY-SA 4.0, ja que les llicències de Creative Commons no estan recomanades per a producte de software, i només seria adequada en cas de que només es publiqui el dataframe. En aquest cas particular, publicar el dataframe de manera independent no té cap sentit, ja que interessa poder-lo regenerar en temps real per obtenir els preus de mercat actualitzats, i per tant, el dataframe sempre ha d'anar lligat al script que el genera.

9. Codi

El codi es pot trobar a GitHub en el següent enllaç, on s'inclou un petit manual amb les indicacions per executar-lo:

- <https://github.com/scanet9/CoinMarketCapScraper>

10. Taula de contribucions

Contribucions	Signa
Recerca prèvia	Sergi Canet
Redacció de les respostes	Sergi Canet
Desenvolupament del codi	Sergi Canet