

# **Tipologia I Cicle De Vida De Les Dades**

---

## **PRAC 1: Web Scraping**

---

**Estanislau Trepap**

## 1. Títol i subtítol.

Per a al conjunt de dades generat, s'ha escollit el següent títol/subtítol:

**Historic Crypto-Trading Dataset: Capitalitzacions de mercat històriques de les 25 principals criptodivises.**

## 2. Descripció i context.

El conjunt de dades que s'ha generat per aquesta PRAC, conté les capitalitzacions mercantils històriques dels últims cinc anys per a les 25 principals criptodivises, en termes de valoració mercantil (al canvi, en dòlars americans), a dia d'avui.

El conjunt de dades conté dades diàries històriques dels diferents indicadors mercantils per a les 25 principals criptodivises, a dia d'avui, dels últims cinc anys: Des del 28 d'abril de 2013 fins al 15 d'Abril de 2018. Els indicadors principals que es presenten en el conjunt de dades són: *capitalització mercantil, valoració a l'inici i al final de la jornada mercantil, valoracions màximes i mínimes diàries, i volum diari de mercat.*

## 3. Imatge identificativa.



*Il·lustració 1: «Criptodivisa» i «Criptomercats» (Bitcoin). Autor: David McBee. Llicència: CC-CO*

## 4. Contingut.

El conjunt de dades generat amb l'script «*scraper*» implementat per aquesta PRAC conté les valoritzacions mercantils per a les 25 criptodivises més importants. Les dades estan compreses des del 28 d'abril de 2013 fins el 15 d'abril de 2018, cobrint, gairebé, cinc anys complets de dades. La data d'inici és la data des que els autors de la web de *Cryptocurrency Market Capitalizations* (<https://coinmarketcap.com>), porten recollint aquesta informació.

El conjunt de està compres de 23219 registres (files) i 8 atributs en cada registre, ocupant en disc, aproximadament, 1,7MB. El fitxer que es presenta es troba en format de valors separats per comes, o CSV, utilitzant la coma, «,», com a separador.

Els atributs presents en el «*dataset*» corresponen als diferents valors de la valoració mercantil de cada dia, en l'interval de dades especificat anteriorment, juntament amb el nom de la criptodivisa corresponent. Aquests atributs són:

- **Name:** Denominació de la criptodivisa.
- **Date:** Data a la que el registre fa referència. Es tracta d'un dia en format AAAA-MM-DD.
- **Open:** Valoració, en dòlars, de la criptodivisa en el moment de l'apertura (o inici del dia).
- **High:** Valoració màxima, al canvi, a la que ha arribat la criptodivisa en el dia en qüestió.
- **Low:** Valoració mínima, al canvi, a la que ha arribat la criptodivisa en el dia en qüestió.
- **Close:** Valoració de la criptodivisa al final del dia (de la jornada mercantil).
- **Volume:** Valoració total del volum monetari que s'ha mogut durant el dia en qüestió.
- **Market Cap:** Valor de la capitalització de mercat (*preu \* volum circulant*) de la criptodivisa en la data del registre.

Tots els atributs monetaris es troben en dòlars americans.

Les criptodivises amb dades presents en el conjunt de dades, es a dir, les vint-i-cinc més importants en termes de capitalització mercantil, són les següents: *bitcoin*, *ethereum*, *ripple*, *bitcoin-cash*, *litecoin*, *cardano*, *stellar*, *iota*, *neo*, *monero*, *nem*, *dash*, *ethereum-classic*, *qtum*, *verge*, *lisk*, *bitcoin-gold*, *zcash*, *nano*, *bytecoin-bcn*, *steem*, *wanchain*, *siacoin*, *bitshares* i, finalment, *dogecoin*.

## 5. Agraïments.

S'agraeix principalment als autors del lloc web *Cryptocurrency Market Capitalizations* (<https://coinmarketcap.com/>), per la seva tasca en mostrar i recollir les dades de capitalització de mercat històriques i oferir-les sense cap cost.

## 6. Inspiració.

Un tema tant actual com el de les criptomonedes i el «*blockchain*», pot ésser d'interès per a individus i empreses ben diverses. El conjunt de dades que hem presentat i generat per a la realització d'aquesta PRAC pot resultar interessant per a realitzar estudis de l'evolució de les valoracions mercantils de les diferents criptodivises, trobar oportunitats d'arbitratge, utilitzar-se com a punt de partida per veure si factors externs o successos d'actualitat han contribuït en l'evolució de dites valoracions tant a l'alça com a la baixa, descobrir patrons en les evolucions mercantils, entre molts d'altres.

## 7. Llicència.

La llicència escollida per al conjunt de dades d'aquesta PRAC, el codi font i els seus continguts és:

**CC BY-SA 4.0 License.**

Aquesta llicència permet a altres persones i/o entitats remesclar, retocar i crear a partir d'aquesta obra, inclús amb fins comercials, sempre i quan se t'atorgui el reconeixement oportú i al mateix temps llicenciïn les seves noves creacions sota les mateixes condicions.

S'ha escollit aquesta llicència per la seva permissivitat i perquè, fins i tot, en permet derivacions comercials. Això és útil per si una empresa està interessada en realitzar algun estudi i/o realitzar algun obra derivada de les dades aquí recollides. També trobem interessant que les obres que se'n derivin, donin crèdit als continguts originals, així com que ho llicenciïn aquestes obres derivades sota les mateixes condicions.

## 8. Codi i conjunt de dades generat.

El codi font implementat per a la generació del conjut de dades d'aquesta PRAC, es troba en el repositori de *GitHub*: <https://github.com/etrepat/tcvd-prac1>, acompanyant a aquest document.

## Recursos.

Material de l'assignatura.

Documentació online del llenguatge Python: <https://docs.python.org/3/>

Documentació online de la llibreria Pandas: <https://pandas.pydata.org/>

**Lawson, R.** (2015). *Web Scraping with Python*. Packt Publishing Ltd.

**Mitchell, R.** (2015). *Web Scraping with Python: Collecting Data from the Modern Web*. O'Reilly Media, Inc.

Cryptocurrency Market Capitalizations (<https://coinmarketcap.com/>)