

TIPOLOGIA I CICLE DE VIDA DE LES DADES

Pràctica 1: Web Scraping

DOCUMENT ADJUNT A LA PRÀCTICA

Realitzat per Cristian Alarcón Sanabria i Daniel Vilaseca Miguel

Professor Sergio Trilles Oliver

Curs 2019-2020 segon semestre

1. CONTEXT

La pràctica es realitza en el context de trobar una pàgina web amb informació actualitzada i fiable sobre algun producte o eina que s'utilitzi en el dia a dia.

En el nostre treball hem recol·lectat informació sobre els tipus de canvi de diferents divises.

Amb el codi que hem presentat, executant-lo un cop al dia, es pot anar generant un csv que contingui els valors (actual del moment de l'execució del programa, màxim diari i mínim diari) del tipus de canvi de diferents parells de divises.

Per la pràctica presentem el dataset amb les dades del dia 12/04 i 13/04, però es podria ampliar i acabar tenint un csv amb les dades de molts més dies.

La informació del dataset s'ha extret del lloc web: www.finanzas.com/divisas/ on podem trobar diferent informació sobre el tipus de canvi de diferents parells de divises i la seva evolució.

S'ha escollit aquesta pàgina per a poder utilitzar un codi una mica més complex i poder utilitzar una spider per fer web scraping, ja que altres pàgines amb valors de divises o eren molt senzilles o massa complexes, creiem que en aquesta en concret tenim un terme mig on ens trobem còmodes per a la realització de la pràctica.

2. TÍTOL

Com a títol pel data set escollim TdC-Divises-per-dia. El títol descriu les dades que podem trobar al dataset, el tipus de canvi de diferents parells de divises en diferents dies.

3. DESCRIPCIÓ DEL DATASET

El dataset és un conjunt de dades que representa el canvi de divises de la relació de diferents monedes (normalment relacionades amb l'euro o el dòlar), on cada dia s'afegeixen dades.

4. REPRESENTACIÓ GRÀFICA

Aquesta és la informació a la plana web de la que extraiem les dades:

<div>EURO DOLAR USA OTRAS DIVISAS</div>						
Cruce	Último	Diferencia	Dif.%	Máximo	Mínimo	Hora
EUR/USD	1,0913	-0,0019	-0,17	1,0914	1,0912	22:10
EUR/GBP	0,8722	-0,0052	-0,59	0,8722	0,8721	22:10
EUR/JPY	117,5950	-0,8450	-0,71	117,6300	117,5600	22:10
EUR/CHF	1,0556	-0,0004	-0,04	1,0557	1,0555	22:10
EUR/ARS	71,1590	0,6120	0,87	71,1610	71,1580	21:16
EUR/AUD	1,7079	-0,0177	-1,02	1,7080	1,7078	22:10

En canvi al nostre dataset les dades queden així, on podem veure que ens hem quedat amb les dades més importants i s'ha extret el nom complet de les divises entrant a cada moneda mitjançant l'spider de python, a més s'afegeix la data del dia en que s'executa el programa.

1	Divises	Tipus de canvi	Tipus màxim	Tipus mínim	Data
2	['EURO/DOLAR SINGAPUR']	['1,5458']	['1,5493']	['1,5438']	12/04/2020
3	['DOLAR NEOZELAND/DOLAR USA']	['0,6078']	['0,6105']	['0,6058']	12/04/2020
4	['DOLAR AUSTRALIA/DOLAR USA']	['0,6351']	['0,6371']	['0,6312']	12/04/2020
5	['LIBRA ESTERLINA/DOLAR USA']	['1,2455']	['1,2489']	['1,2446']	12/04/2020
6	['DOLAR US/RAND SUDAFRICANO']	['17,9587']	['18,0608']	['17,9405']	12/04/2020
7	['DOLAR USA/DOLAR SINGAPUR']	['1,4135']	['1,4176']	['1,4121']	12/04/2020
8	['DOLAR USA/CORONA SUECA']	['9,9237']	['9,9701']	['9,9007']	12/04/2020
9	['DOLAR USA/CORONA NORUEGA']	['10,1863']	['10,2490']	['10,1836']	12/04/2020
10	['DOLAR USA/PESO MEXICANO']	['23,3450']	['23,7895']	['23,2972']	12/04/2020
11	['DOLAR USA/DOLAR HONG KONG']	['7,7515']	['7,7538']	['7,7511']	12/04/2020
12	['DOLAR USA/CORONA DANESA']	['6,8227']	['6,8363']	['6,8144']	12/04/2020
13	['DOLAR USA/YUAN CHINO']	['7,0324']	['7,0469']	['7,0257']	12/04/2020
14	['DOLAR USA/DOLAR CANADIENS']	['1,3961']	['1,4011']	['1,3946']	12/04/2020
15	['DOLAR USA/REAL BRASILEÏO']	['5,0943']	['5,1078']	['5,0943']	12/04/2020
16	['DOLAR USA/FRANCO SUÏZO']	['0,9664']	['0,9681']	['0,9640']	12/04/2020
17	['DOLAR USA/YEN JAPONES']	['108,4340']	['108,6000']	['108,3250']	12/04/2020
18	['DOLAR USA/EURO']	['0,9144']	['0,9159']	['0,9128']	12/04/2020
19	['EURO/RAND SUDAFRICANO']	['19,6483']	['19,7597']	['19,6344']	12/04/2020
20	['EURO/CORONA SUECA']	['10,8511']	['10,8897']	['10,8410']	12/04/2020

5. CONTINGUT

El conjunt de dades es compon de la variable "Divises" (primera columna) que ens indica les dues divises comparades amb el valor de la primera en comparació amb la segona. La variable "Tipus de canvi", que és una extracció del moment en que s'ha executat el codi del canvi de moneda a la plana web. La variable "Tipus màxim" ens mostra el màxim valor de canvi que ha tingut la primera moneda respecte a la segona al llarg del dia. La variable "Tipus mínim" ens mostra el contrari que la variable anterior, el mínim diari de la primera moneda respecte a la segona. I per últim la variable "Data" ens diu el dia en que s'han extret les dades.

El dataset inclou les dades dels dies 12 i 13 d'abril.

6. AGRAÏMENTS

La nostra font d'informació ha estat la plana web finanzas.com (propietari del conjunt de dades), propietat de Inversor Ediciones, SL, inscrita en el registre Mercantil de Madrid, basada en les dades públiques dels valors de les divises.

Els recursos d'informació que hem fet servir per respondre la pràctica (a més dels materials de l'aula virtual) són els següents:

<https://www.youtube.com/watch?reload=9&v=E5cSNSeBhjw>

Curs Web Scraping in Python de <https://www.datacamp.com/>

7. INSPIRACIÓ

En el nostre cas el conjunt de dades és interessant per poder fer una valoració ràpida del valor de les diferents monedes, ja sigui per a planificar un viatge on es necessiti de moneda estrangera, per a valorar compres internacionals o per a valorar el grau d'afectació a un país de possibles crisis econòmiques.

Aquestes dades podrien interessar a una empresa que hagi d'utilitzar moneda estrangera (exportar o importar productes).

També es podria fer servir el conjunt de dades per analitzar l'evolució del tipus de canvi de diferents monedes al llarg del temps i veure tendències.

8. LLICÈNCIA

El valor de les divises és de domini públic, per tant aquestes dades operen sota la llicència Released Under CC0: Public Domain License.

9. CODI

El codi està adjuntat a Github, l'arxiu té el nom "Tdc-divises-Finanzas.py".

10. DATASET

El csv "TdC-Divises-per-dia" està adjuntat a Github.

També està adjuntat a Zenodo: <https://doi.org/10.5281/zenodo.3750134>

Taula de contribucions

Contribucions	Signa
Recerca prèvia	DV, CA
Redacció de les respostes	DV, CA
Desenvolupament codi	DV, CA

Tots els integrants del grup han participat en totes les fases de la pràctica.