

Tipologia i cicle de vida de
les dades

Pràctica 1

Marc Trepà

10 d'Octubre 2021



Taula de continguts

Descripció de la Pràctica a realitzar	3
Context	3
Títol	3
Contingut	4
Agraïments	8
Inspiració	8
Llicència	9
Codi	10
Dataset	10
Contribucions	10
Webgrafia	10



Descripció de la Pràctica a realitzar

1. Context

El context per a recuperar **l'informació sobre els canvis de moneda es troba al butlletí oficial** pero alla es publica molta mes informació que no tenen res a veure amb els canvis de moneda. Per tant, per a recuperar a quin butlletí es troba aquesta informació, en **el portal de duana es on es pot trobar l'enllaç**.

A la **web de la duana es recupera la data i l'enllaç del butlletí** on hi ha el canvi de moneda per aquell mes, un cop **s'accedeix al butlletí es recupera el codi de la moneda, la seva descripció i el valor de la moneda**.

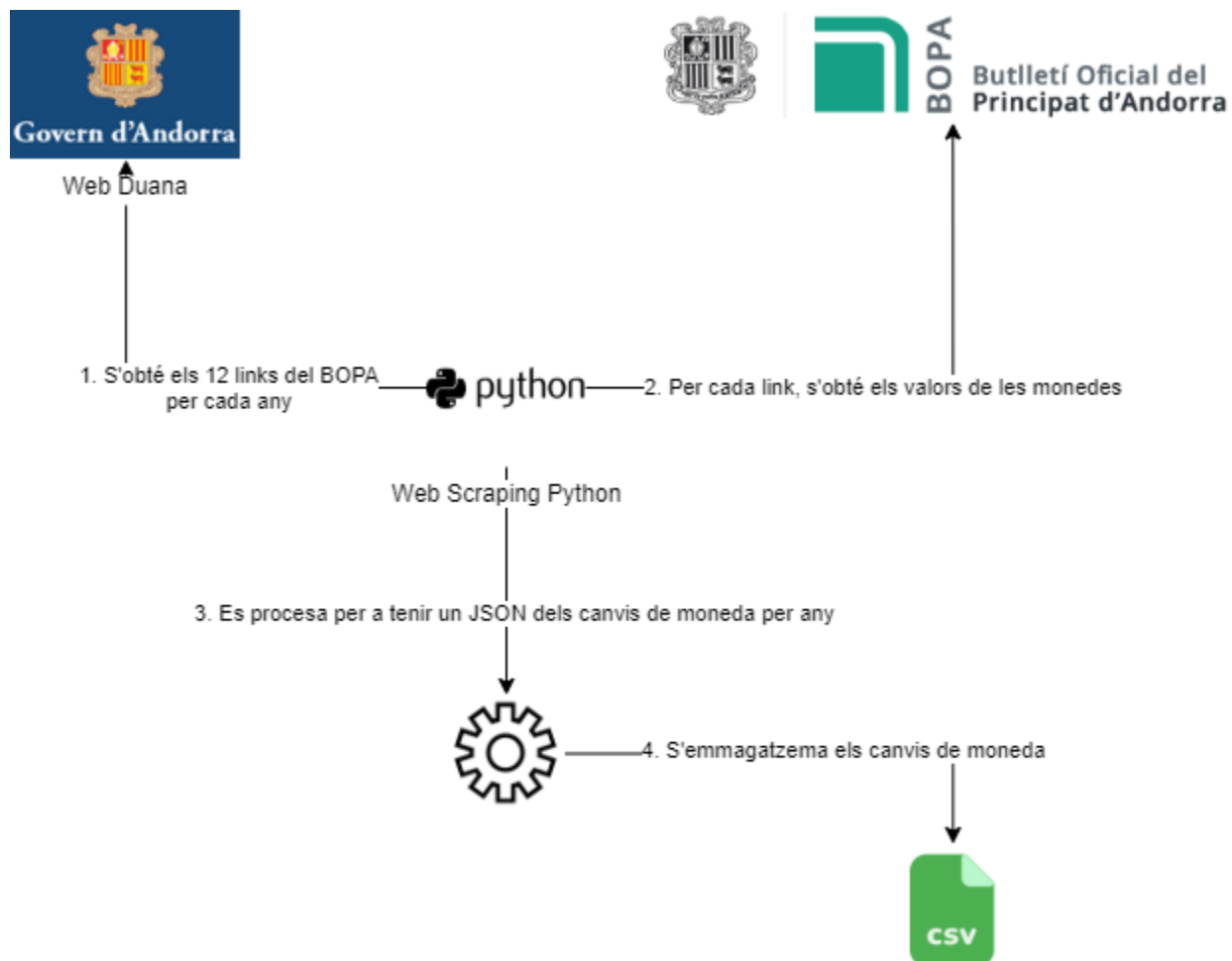
2. Títol

Cotització monetaria.

3. Descripció del dataset

El dataset conté els canvis monetaris de les monedes per any i mes admeses pel Govern d'Andorra.

4. Representació gràfica



5. Contingut

El dataset conté els camps següents:

Camp	Descripció
year	Any de la informació publicada de la moneda, en format YYYY.
month	Mes de la informació publicada de la moneda, en format MM

code	Codi de la moneda, tres caràcters: XXX
description	Descripció del codi de la moneda.
value	Preu de la moneda amb 4 dígitos pels decimals, exemple: 234.5960

El **període de temps de les dades comença a publicar-se l'any 2014** fins l'actualitat on hi ha el canvi de preus de les monedes de cada mes.

El procés per recollir les dades **s'ha de buscar els links al portal de duana** on aquest ens **dona l'enllaç a l'edecte on apareixen els preus** que s'han publicat:

Canvis monetaris

Avis del 23-9-2021 pel qual es publiquen els canvis monetaris mensuals admesos a la cotització oficial al Principat d'Andorra.

El Departament de Tributs i de Fronteres comunica a totes les persones interessades els canvis monetaris mensuals en relació a la cotització oficial al Principat d'Andorra, corresponents al dia 23 de setembre del 2021 i que s'aplicaran a la Duana Andorrana de l'1 al 31 d'octubre del 2021.

Canvis de l'1 al 31 d'octubre del 2021		
GBP	LLIURA ESTERLINA	0,8635
DKK	CORONA DANESA	7,5487
NOK	CORONA NORUEGA	10,2803
SEK	CORONA SUECA	10,3260
CHF	FRANC SUÍS	1,0908
USD	DÒLAR USA	1,1779
CAD	DÒLAR CANADENC	1,5081
JPY	IEN	129,9151
HKD	DÒLAR HONG KONG	9,2529
EUR	EURO	1,0000

Canvis anteriors

— Any 2021

🔗 [Gener 2021](#)

🔗 [Febrer 2021](#)

🔗 [Març 2021](#)

🔗 [Abril 2021](#)

🔗 [Maig 2021](#)

Un cop es recupera el link que ens porta al edicte del mes i any publicat es recupera el preu de la moneda:

Govern

Altres anuncis

Avis

Avis del 21-1-2021 pel qual es publiquen els canvis monetaris m

El Departament de Tributs i de Fronteres comunica a totes les pe
admesos a la cotització oficial en el Principat d'Andorra, correspo
Andorrana de l'1 al 28 de febrer del 2021.

GBP	Lliura esterlina	0,8964
DKK	Corona danesa	7,5509
NOK	Corona noruega	10,8848
SEK	Corona sueca	10,4974
CHF	Franc suís	1,0834
USD	Dòlar USA	1,2205
CAD	Dòlar canadenc	1,5592
JPY	Ien	127,4216
HKD	Dòlar Hong Kong	9,5463

6. Agraïments

La informació del conjunt de dades es propietari del Govern d'Andorra, justament aquesta informació s'ha tret de butlletí oficial, on allà es comenta que es fa públic per al coneixement general.

S'ha fet una busqueda en google per trobar altres anal.lisis anteriors però justament la informació que es publica d'Andorra no apareix res, lo que si que he trobat a la web de data.europa.eu o [kaggle](https://www.kaggle.com) són datasets on es registra el valor de la moneda que representen l'equivalent a un dolar en els exemples trobats, aquesta informació es publicada per eurostat i Federal Reserve's respectivament.

Els passos que s'han tingut en compte per actuar dintre dels termes legals son:

- Les publicacions tant de la web de duana com la del bopa no s'esta accedint a traves de cap login i per tant no s'han acceptat cap condició o termes legals de que no es pugui aplicar el web scraping i les webs estan plubliques per a que tothom tingui accés.
- S'ha analitzat el fitxer de robots.txt i del sitemap.xml per identificar si allà on s'accedeix es un lloc privat o restringit i no s'ha trobat cap impediment per accedir. L'únic que al rastregar els bots dels buscadors la web del BOPA no esta permesa, pero publicament a traves d'enllaços es pot accedir per tant no hi ha cap infracció de drets d'autor.
- S'ha revisat les webs per si hi ha alguna API per poder obtenir la informació i evitar realitzar el proces de web scraping.
- S'ha treballat amb la manera de no sobrecarregar els servidors ni de accedir a servidors que no es te acces, utilitzant solament la informació que esta publicada aixi obtenint-la d'una manera justa.
 - Per la part del desenvolupament s'ha realitzat una petició per a obtenir el HTML i a partir d'aquest s'ha anat treballant el procés.


7. Inspiració

S'ha cregut interessant **agrupar els valors** de cada moneda durant un llarg període de temps **en un dataset** perquè així es més facil poder analitzar les variacions de les monedes ja que **actualment la informació esta dispersada en diferents pàgines web** i ara esta en un sol lloc amb un fitxer CSV per a poder treballar-ho.

Tenir aquest dataset **es pot analitzar el mercat de quines monedes tenen més demanda o oferta** ja que els preus pujaran o baixaran i pots preveure si és millor intercanviar-ho a Euros o no

utilitzant machine learning, també es pot analitzar **quines monedes es van devaluar durant el temps del covid-19**.

Si es compara amb els datasets que s'han trobat, **l'estructura dels fitxers** amb el dataset del **kaggle utilitza el CSV** però en l'altre, el **eurostat, treballa amb XML**. Pel que fa **els camps el dataset de kaggle cada fila es la data** i les columnes ens indiquen el tipus de moneda que es:

Time Series	AUSTRALIA - AU...	EURO AREA - EU...	NEW ZEALAND - ...	UNITED KINGDO...
Date	AUD/USD	EUR/USD	NZD/USD	GBP/USD
 3Jan00 31Dec19	ND 4% 1.3943 0% Other (5011) 96%	ND 4% 0.8898 0% Other (5010) 96%	ND 4% 1.4599 0% Other (5010) 96%	ND 4% 0.64 0% Other (5005) 96%
2000-01-03	1.5172	0.9847	1.9033	0.6146
2000-01-04	1.5239	0.97	1.9238	0.6109
2000-01-05	1.5267	0.9676	1.9339	0.6092
2000-01-06	1.5291	0.9686	1.9436	0.607
2000-01-07	1.5272	0.9714	1.938	0.6104
2000-01-10	1.5242	0.9754	1.935	0.6107
2000-01-11	1.5209	0.9688	1.9365	0.6068
2000-01-12	1.5202	0.9727	1.9365	0.6073

El cas de l'eurostats cada fila ens indica la moneda i el mes que pertany igual com s'ha generat el nostre dataset i finalment el format són molt semblant al que s'ha generat, l'únic que el meu cas la data s'ha separat per any i mes.

8. Llicència

Les dades que hi ha al dataset s'han tret de manera legal i la informació es pot trobar a internet publicada, així que amb la llicència de **Released Under CC0: Public Domain License** ens serveix a part la informació es pot copiar, modificar i distribuir sense permís, ja que el dataset es l'agrupació de lo que està publicat al butlletí.

9. Codi

El codi per la utilització del web Scraping s'ha realitzat en python i està pujat al [GitHub](#) on hi ha una breu explicació dels fitxers i com executar-lo. El desenvolupament s'ha treballat amb el IDE PyCharm.

10. Dataset

Un cop s'ha generat el CSV aquest també està al repository del Github i s'ha pujat també al Zenodo a la següent ruta: <https://zenodo.org/record/5590287#.YXGcHIVBzIU>

Contribucions

Contribucions	Signatura
Investigació prèvia	MTP
Redacció de les respostes	MTP
Desenvolupament del codi	MTP


Webgrafia

Pregunta 8. <https://creativecommons.org/share-your-work/public-domain/cc0/>

Pregunta 8. <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Pregunta 9. <https://www.geeksforgeeks.org/ssl-certificate-verification-python-requests/>

Pregunta 9. https://linuxhint.com/find_children_nodes_beautiful_soup/



Pregunta 9.

<https://stackoverflow.com/questions/27981545/suppress-insecurerequestwarning-unverified-http-s-request-is-being-made-in-pytho>

Pregunta 9. <https://www.programcreek.com/python/example/93085/whois.whois>

