

**1. Context. Explicar en quin context s'ha recol·lectat la informació. Explicar per què el lloc web triat proporciona aquesta informació.**

El lloc web utilitzat pel desenvolupament d'aquesta pràctica és [www.finanzas.com](http://www.finanzas.com). Aquesta pàgina està operada per la societat INVERSOR EDICIONES S.L. que es dedica a la prestació de serveis financers i d'inversió. El lloc web disposa d'una àrea pública que conté informació econòmica d'interès general com notícies o cotitzacions, i d'una secció privada que requereix registre i ofereix serveis d'assessoria tan econòmica com fiscal, d'anàlisi tècnic, de banca personal o de formació.

Entre la informació que s'hi publica es troben les cotitzacions històriques dels principals indicadors econòmics. Per a dur a terme aquesta pràctica, els indicadors escollits són les cotitzacions de *l'ibex35*, *euro stoxx50*, *dow jones* i la cotització Euro/Dòlar.

La informació obtinguda permetrà l'anàlisi comparat del comportament de les cotitzacions dels indicadors seleccionats, que es poden relacionar amb altres informacions rellevants, que poden condicionar les cotitzacions, com resultats d'eleccions, catàstrofes naturals, declaracions de responsables econòmics, tractats internacionals, ... que es poden determinar a partir de l'anàlisi de l'activitat a les xarxes socials o a partir de pàgines web especialitzades en temes econòmics. Addicionalment, serà possible identificar els dies i els mesos de l'any en que la tendència de les cotitzacions és positiva, quan hi ha un major volum de cotització i si existeix alguna correlació entre els indicadors.

Amb aquesta informació es podrà anar definint i refinant un model que permeti anticipar el comportament d'aquests indicadors, i establir patrons d'actuació i estratègies d'inversió amb una major probabilitat d'èxit.

**2. Definir un títol pel dataset. Triar un títol que sigui descriptiu.**

Cotitzacions anuals de l'€//\$ i dels índexs borsaris espanyol, europeu i americà.

**3. Descripció del dataset. Desenvolupar una descripció breu del conjunt de dades que s'ha extret (és necessari que aquesta descripció tingui sentit amb el títol triat).**

El dataset *Cotitzacions anuals de l'€//\$ i dels índexs borsaris espanyol, europeu i americà* conté els valors diaris, en els quals hi ha hagut activitat econòmica, de l'últim any per a les cotitzacions de *l'ibex35*, *Euro stoxx50*, *Dow Jones* i de l'€//\$. Més concretament es mostren les cotitzacions en el moment d'obertura i el tancament de mercats, les cotitzacions màximes i mínimes diàries, el percentatge de variació respecte el dia anterior i el volum de contractació.

L'objectiu del dataset és disposar de la informació anual dels principals indicadors econòmics per permetre l'anàlisi comparat del comportament de les seves cotitzacions, correlacionar-lo amb altra informació rellevant per tal de poder anticipar el seu comportament i establir una estratègia d'inversió.

Les dades s'obtenen a partir d'un procés de web scraping de les següents pàgines web :

<http://www.finanzas.com/ibex-35/datos-historicos.html>  
<http://www.finanzas.com/divisas/eur-usd/datos-historicos.html>  
<http://www.finanzas.com/euro-stoxx50/datos-historicos.html>  
<http://www.finanzas.com/dow-jones/datos-historicos.html>

Aquestes pàgines mantenen la mateixa estructura, per la qual cosa s'ha implementat una única funció que a partir de la recepció de l'url extreu la informació definida i la inserta en un fitxer CSV. La funció s'ha implementat en Python i s'han utilitzat les llibreries request, BeautifulSoup i csv.

El fitxer CSV fa servir el punt i coma (;) com a separador de camps, i els valors numèrics tenen el punt (.) com el separador de milers i la coma (,) com a separador de decimals.

Els registres del fitxer CSV presenta la següent estructura:

**Tipus** (String) : Identificador del tipus de registre. Permet la selecció dels registres en base al tipus de indicador. Pot prendre els valors `ibex-35`, `euro-stoxx50`, `dow-jones` o `eur-usd`.

**Data** (Date) : Data de la jornada de cotització. La data presenta el format DD/MM/AAAA.

**Apertura** (float) : Valor de cotització a la apertura dels mercats.

**Tancament** (float) : Valor de cotització al tancament dels mercats.

**Percentatge** (float) : Percentatge de variació de la cotització durant la jornada.

**Màxim** (float) : Valor de cotització màxim durant la jornada.

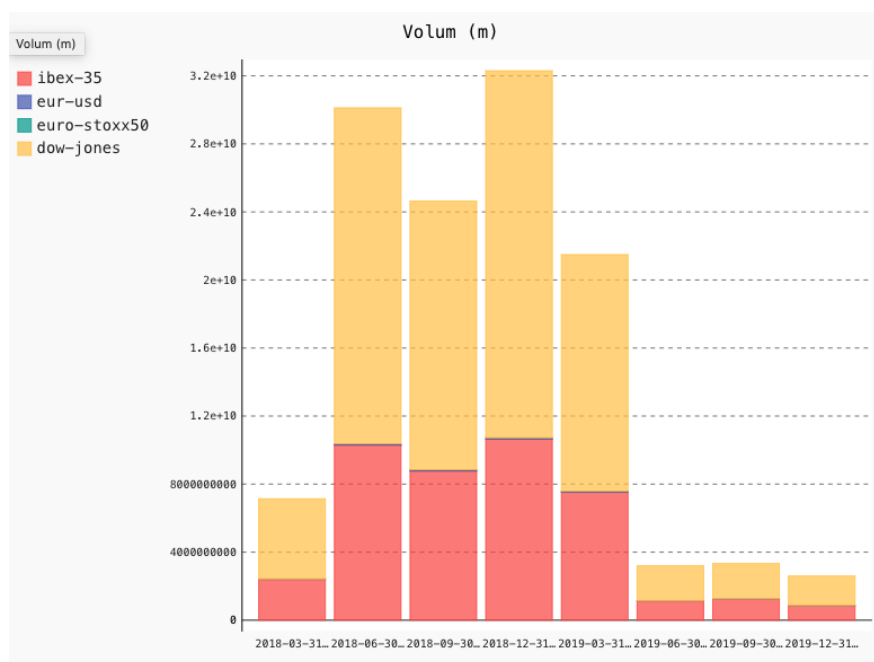
**Mínim** (float) : Valor de cotització mínim durant la jornada.

**Volum** (float) : Volum de contractació durant la jornada en € o \$.

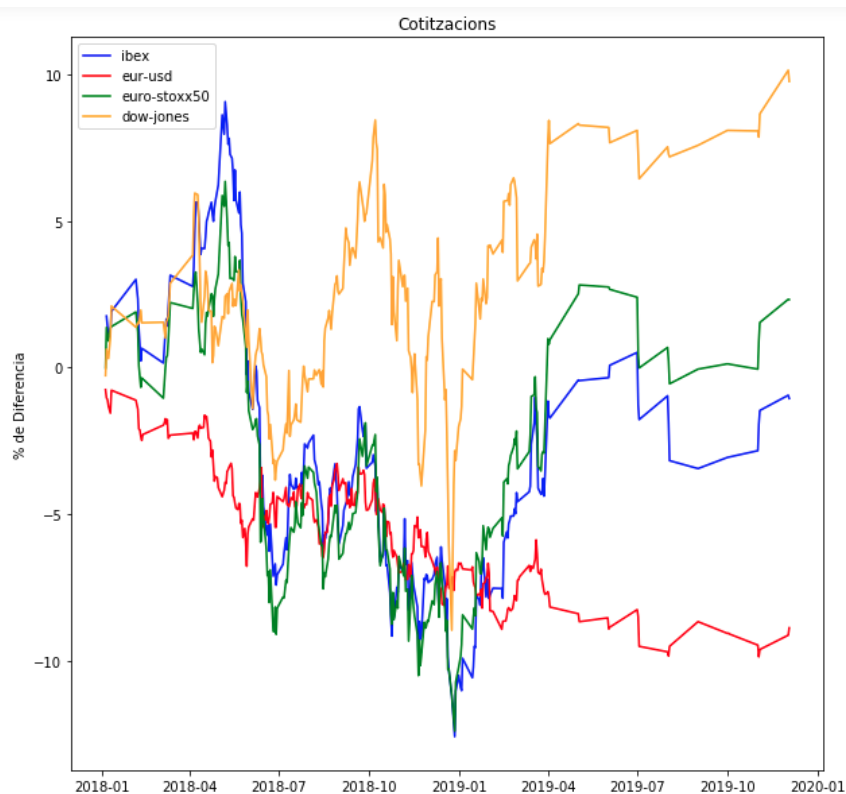
Les dades obtingudes es registren en el format que s'obtenen i no s'en fa cap tractament ni filtratge en fase de captura. Les dades són consistents i únicament és reb el marcador -- quan no hi ha dades per el camp rebut.

En la recuperació de les dades del dataset a partir del fitxer CSV cal tenir en compte el format de la data, el tipus de separador de milers i decimals, i fer el tractament necessari pel cas de dades no disponibles (--)

#### 4. Representació gràfica



Total de volum de contractació quadrimestral per a cadascuna de les cotitzacions



Evolució percentual de les diferents cotitzacions

**5. Contingut. Explicar els camps que inclou el dataset, el període de temps de les dades i com s'ha recollit.**

Per a cada registre del conjunt de dades es disposa dels següents atributs :

- Tipus : Tipus de cotització (ibex-35, euro-stoxx50, dow-jones o eur-usd)
- Data : Data de la jornada de cotització
- Apertura : Valor de cotització a la apertura dels mercats
- Tancament : Valor de cotització al tancament dels mercats
- Percentatge : Variació en % de la cotització durant la jornada
- Màxim : Valor de cotització màxim durant la jornada
- Mínim : Valor de cotització mínim durant la jornada
- Volum : Volum de contractació durant la jornada

Pel que fa a la temporalitat del conjunt de dades, anteriorment s'ha comentat que aquest conjunt conté dades diàries dels dies amb activitat econòmica del darrer any.

El fet que l'estructura de les dades és la mateixa per a tots els indicadors, a facilitat la seva recollida ja que ha permès utilitzar les mateixes funcions de *python* per als diferents *urls*.

Aquestes dades s'han recollit mitjançant tècniques de web scraping sobre la taula de valors històrics anuals per cada un dels indicadors descrits. Posteriorment, s'ha generat un fitxer CSV amb el camps i contingut especificats.

**6. Agraïments. Presentar el propietari del conjunt de dades. És necessari incloure, a cites de recerca o anàlisis anteriors (si n'hi ha).**

Com s'ha explicat en el primer apartat, el lloc web utilitzat pel desenvolupament d'aquesta pràctica és [www.finanzas.com](http://www.finanzas.com). Aquesta pàgina està operada per la societat INVERSOR EDICIONES S.L. Aquesta societat es dedica a la prestació de serveis financers i d'inversió.

A part d'aquesta pàgina web, per tal d'implementar tècniques de web scraping en *python* s'han utilitzat els apunts d'aquesta unitat que es troben el campus virtuals així com s'ha consultat diferents entrades de la pàgina web per a consultes informàtiques [www.stackoverflow.com](http://www.stackoverflow.com) o també s'ha utilitzat com a model l'exemple que es mostra en el següent enllaç <https://kite.com/python/examples/4420/beautifulsoup-parse-an-html-table-and-write-to-a-csv>.

En la mateixa pàgina web d'on hem obtingut les dades hi ha un apartat amb estudis tècnics d'aquestes, entre d'altres estudis, en aquesta secció hi podem trobar rankings i valors per a les diferents empreses que participen en cada mercat borsari, per a cada mercat, quina decisió sobre el tipus d'acció (compra o venda) és la més adequada per a dur-se a terme en el moment actual, els preus objectius a aconseguir, la rendibilitat obtinguda, etc.

**7. Inspiració. Explicar per què és interessant aquest conjunt de dades i quines preguntes es pretenen respondre.**

El motiu de la selecció d'aquesta pàgina per a fer web scraping és que els dos integrants del grup han mostrat un interès comú pels diferents mercats de valors.

Per altre banda, tal i com s'ha mencionat anteriorment, aquest dataset conté les dades històriques dels principals índexs borsaris d'Espanya, Europa i Estats units, així com l'evolució de la cotització €/\$. Aquest fet ens permet plantejar, entre moltes altres, les següents qüestions :

- Analitzar l'evolució d'un índex respecte el comportament dels altres.
- Analitzar l'evolució de la cotització de l'€/€ respecte al comportament dels índexs borsaris.
- Períodes de l'any on hi ha major tendència de creixement o decreixement.
- Períodes de l'any on hi ha major volum de contractació.
- Comparació del volum de contractació respecte a l'evolució dels índexs (quan creix la contractació creixen els índexs, ...)
- Dies de la setmana més proclius al creixement o decreixement dels índexs

Com hem citat anteriorment, en la pàgina web d'on hem obtingut les dades, es realitza un petit estudi en que es contesta les preguntes de si és millor moment per a comprar o per a vendre, o bé s'obté la rendibilitat obtinguda de la jornada.

L'enllaç [http://www.expansion.com/mercados/cotizaciones/indices/ibex35\\_I.B.html](http://www.expansion.com/mercados/cotizaciones/indices/ibex35_I.B.html) conté un altre exemple en el que es realitza un breu estudi a un conjunt de dades molt semblant a l'escollit per a fer aquesta pràctica. En aquest estudi es mostra un gràfic de l'evolució temporal del valor de l'Ibex 35, els nivells mínims i màxims anuals, la revalorització anual o bé el PER anual.

## 8. Llicència.

La llicència seleccionada és la GNU General Public License v.3. Es tracta d'una llicència de drets d'autor, orientada a programari lliure i codi obert, que permet la copia, distribució i modificació del codi. El programari sota aquesta llicència és lliure però qualsevol modificació o transformació s'ha de continuar distribuint amb la mateixa llicència GPL (copyleft).

La llicència GPL no permet la distribució d'executables sense la possibilitat d'obtenir el seu codi font de forma gratuïta, requereix que la llicència i el copyright s'han d'incloure amb el codi i s'han de documentar els canvis realitzats.

La llicència GNU General Public License v.3 contempla una limitació de la responsabilitat de l'emissor del programari i no contempla cap tipus de garantia.

És la llicència lliure de programari més utilitzada i Linux és el cas de programari sota llicència GPL amb més difusió.

Contribucions	Signa
Recerca prèvia	CR, PC
Redacció de les respostes	CR, PC
Desenvolupament codi	CR, PC