

Pràctica 1

1. Context

Explicar en quin context s'ha recollit la informació. Explicar per què el lloc web triat proporciona aquesta informació.

En aquesta pràctica recollirem la informació de la pàgina web <https://scryfall.com/>. Aquesta pàgina web recopila el llistat de totes les cartes de *Magic: The Gathering* (MTG) publicades, juntament amb el preu actual. MTG és un joc de cartes col·leccionables publicat per primera vegada l'any 1993, en que dos jugadors s'enfronten l'un a l'altre amb les seves cartes. Periòdicament s'afegeixen noves cartes al jocs, i actualment existeixen més de 20,000 cartes diferents. És per això que els cercadors com Scryfall són molt populars entre els jugadors, ja que els permeten aplicar una gran varietat de filtres i obtenir un llistat de les cartes que els poden interessar. A més, Scryfall està integrat amb diverses tendes de compra-venta de cartes, pel que proporciona també el preu de mercat actual de les cartes.

2. Títol

Definir un títol que sigui descriptiu pel dataset.

MTG Price Set

3. Descripció del dataset

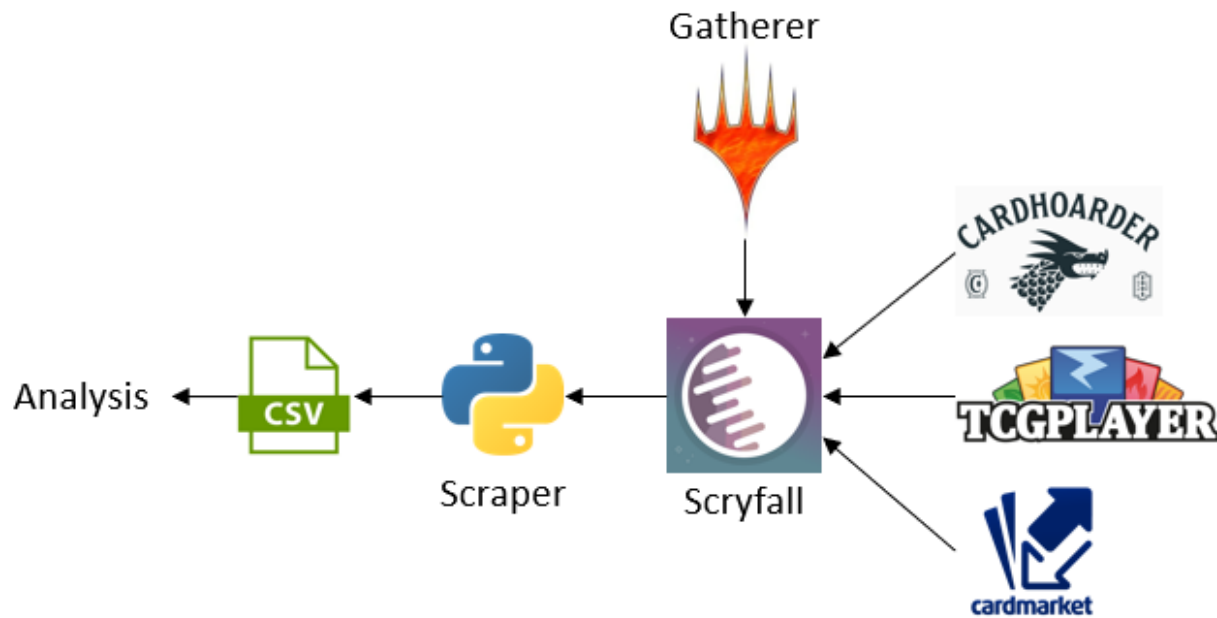
Desenvolupar una descripció breu del conjunt de dades que s'ha extret. És necessari que aquesta descripció tingui sentit amb el títol escollit.

El dataset que volem obtenir és un llistat de totes les cartes de MTG disponibles juntament amb el seu preu. De cada carta ens interessa saber alguns dels seus atributs, com el cost o l'any de publicació. Referent al preu, ens interessen tres valors: el preu en USD al mercat dels Estats Units, el preu en EUR del mercat Europeu, i el preu en TIX, pel mercat online. TIX és la moneda de canvi en l'aplicació online de MTG, on $1 \text{ TIX} \approx 1 \text{ USD}$.

Scryfall obté el llistat de cartes del [Gatherer](#), que és la base de dades oficial de MTG i s'actualitza regularment cada vegada que surten noves cartes al mercat. El preu l'obté de tres llocs diferents: [Cardmarket](#) pel mercat europeu, [TCGPlayer](#) pel mercat nord-americà, i [Cardhoarder](#) pel mercat online.

4. Representació gràfica

Dibuixar un esquema o diagrama que identifiqui el dataset visualment i el projecte escollit.



5. Contingut

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

Pel dataset inclourem totes les cartes d'aquelles col·leccions publicades en anglès i amb més de 100 cartes. Només ens interessen les cartes de col·leccions publicades perquè el preu de les col·leccions no publicades són especulatius. A més, filtrem per col·leccions amb més de 100 cartes i publicades en anglès per eliminar col·leccions promocionals, que solen contenir cartes existents, però amb un preu més elevat que l'habitual degut a la seva exclusivitat.

Aquests són els camps del nostre dataset:

- **set_abbreviation**: Identificador de la col·lecció on es troba la carta.
- **set_card_count**: Nombre de cartes en la col·lecció.
- **set_release_date**: Data de publicació de la col·lecció.
- **card_number**: Nombre de la carta dins de la col·lecció.
- **card_name**: Nom de la carta.
- **card_mana_cost**: Cost de jugar de la carta, en la notació tradicional de MTG.
- **card_type**: Tipus de carta.
- **card_rarity**: Raresa de la carta.
- **card_artist**: Dibuixant de la carta.
- **card_price_usd**: Preu en USD de la carta, segons TCGPlayer.
- **card_price_eur**: Preu en EUR de la carta, segons Cardmarket.
- **card_price_tix**: Preu en TIX de la carta, segons Cardhoarder.

6. Agraïments

Presentar el propietari del conjunt de dades. És necessari incloure cites d'anàlisis anteriors o, en cas de no haver-n'hi, justificar aquesta cerca amb anàlisis similars. Justificar quins passos s'han seguit per actuar d'acord amb els principis ètics i legals en el context del projecte.

Wizards of the Coast, el propietari del joc de MTG, permet l'ús de les seva IP per fans sempre i quan no se'n faci un mal ús. Trobem les guies d'ús aquí: <https://company.wizards.com/en/legal/fancontentpolicy>

Scryfall també permet l'ús de les seves dades. Té una API que permet consultar les dades, però que hem omès per realitzar aquesta pràctica. No obstant, ens hem adherit als límits de peticions que es demanen de no fer més de 10 peticions per segon, afegint un sleep al nostre codi. Aquí estan les seves guies d'ús: <https://scryfall.com/docs/api>.

7. Inspiració

Explicar per què és interessant aquest conjunt de dades i quines preguntes es pretenen respondre. És necessari comparar amb les anàlisis anteriors presentades a l'apartat 6.

El preu de les cartes normalment està lligat a la seva disponibilitat/rarsa i a l'ús que es fa de la carta. Per exemple, una carta que tots els jugadors necessitin per jugar serà molt més cara que una carta que ningú usi. Si a més, aquesta carta és difícil d'aconseguir, el seu preu pot créixer exponencialment. Un dels casos en que s'intenti treure profit a les cartes és durant el que s'anomena pre-venta: la venda de cartes que encara no han publicat i per tant, encara no es poden emprar. En aquest moment els jugadors i les tendes especulen el potencial d'una carta i intenten treure'n profit comprant aquelles cartes que creuen que seran molt demandades pel menor preu. Amb aquest conjunt de dades es podria realitzar un model de regressió o classificació (si convertim els preus en rangs de valors) que ens permetés veure si, a partir d'un conjunt d'atributs de la carta, aquesta podria arribar a tenir un preu alt o baix, i per tant, si ens interessa comprar-la a un preu baix durant la pre-venta (si està disponible), o si és millor esperar a que baixi de preu un cop hagi estat publicada.

El nostre dataset presenta un conjunt de dades molt més simplificat per poder realitzar models simples de classificació i regressió sobre el preu de les cartes que els conjunts de dades complets existents.

8. Llicència

Seleccionar una d'aquestes llicències pel dataset resultant i justificar el motiu de la seva selecció:

- *Released Under CC0: Public Domain License*
- *Released Under CC BY-NC-SA 4.0 License*
- *Released Under CC BY-SA 4.0 License*
- *Database released under Open Database License, individual contents under Database Contents License*
- *Other (specified above)*
- *Unknown License*

Com el conjunt de dades que estem utilitzant pertany a Wizards of the Coast, no podem publicar les dades sense cap tipus de llicència, ja que infringirien les normes d'ús estipulades per ells. En aquest cas, triarem *Released Under CC BY-NC-SA 4.0 License*, ja que evita l'ús comercial de les dades, mentre que en segueix permetent la seva distribució.

9. Codi

Adjuntar al repositori Git el codi amb el qual s'ha generat el dataset, preferiblement en Python o, alternativament, en R.

<https://github.com/rodesdecarro/scryfall-web-scraper>

10. Dataset

Publicar el dataset obtingut en format CSV a Zenodo amb una breu descripció. Obtenir i adjuntar l'enllaç del DOI.

<https://doi.org/10.5281/zenodo.6449920>

11. Vídeo

S'ha de lliurar un vídeo explicatiu de la pràctica on cadascun dels integrants del grup expliqui amb les seves pròpies paraules tant les respostes del projecte com el codi utilitzat per a dur a terme l'extracció. El vídeo ha de ser enviat a través d'un enllaç a Google Drive que heu de proporcionar, juntament amb l'enllaç al repositori Git, al moment de lliurar la pràctica.