

## **Joc de dades: LPs publicats 2020 (12/04)**

Sol Balsells Mejía  
Màster en Ciència de Dades  
13 d'abril 2020

### **Context**

Aquest joc de dades s'ha construït per a resoldre la Pràctica 1 de l'assignatura *Tipologia i Cicle de Vida de les Dades*. Mitjançant tècniques de web scraping, s'ha analitzat el web [albumoftheyear.org](http://albumoftheyear.org) per identificar quins són els treballs musicals amb format LP que s'han publicat des que va iniciar-se el 2020 fins a la data de publicació del joc de dades.

### **Descripció del dataset i imatge**

El joc de dades *LPs publicats el 2020 (12/04)* conté dades sobre tots els treballs amb aquest format que s'han publicat aquest any, segons el contingut del web Album Of The Year actualitzat el 12 d'abril de 2020. Aquest web és una pàgina musical on els usuaris afegeixen els treballs musicals que es publiquen i s'hi afegeix diferent informació sobre aquests treballs, a més de valoracions dels usuaris i de crítiques professionals.

Com únicament s'ha utilitzat el web Album Of The Year com a font de dades, n'utilitzem el logotip (afegint-hi l'any) com a imatge del joc de dades.



## **Contingut**

El web Album Of The Year té un cercador on es permet buscar informació sobre els treballs publicats durant un any concret. El codi utilitza el cercador de treballs de 2020 per analitzar tots els llançaments d'aquest any i, per a cada un d'aquests, n'analitza el format per identificar si es tracta d'un LP.

En cas que el format del treball sigui LP, es recopila informació sobre aquest treball. Els camps seleccionats amb informació sobre el treball són l'autor, el títol d'aquest, la data de publicació, el gènere musical i la discogràfica que l'ha publicat. A més, s'ha inclòs al joc de dades informació sobre l'opinió dels mitjans musicals (nota mitjana i nombre de crítiques) i sobre l'opinió dels usuaris del web (nota mitjana i nombre de valoracions).

## **Agraïments**

Les dades han estat recopilades del web Album Of The Year ([albumoftheyear.org](http://albumoftheyear.org)).

## **Inspiració**

El joc de dades conté informació sobre els discs publicats en el que va de 2020, així que podria utilitzar-se per estudiar quins han estat els LPs millor valorats segons diferents paràmetres (gènere musical, data de publicació, discogràfica, usuaris/crítica especialitzada).

El joc de dades no és gaire complex perquè la seva construcció és costosa des del punt de vista computacional. No obstant, la informació del web Album Of The Year podria donar lloc a anàlisis bastant interessants si, a partir del codi de web scraping utilitzat, s'ampliés la informació que es recopilar. Entre algunes de les aplicacions potencials del joc de dades, es podria estudiar les valoracions que fa cada mitjà especialitzat per estudiar similituds entre aquests mitjans; es podria estudiar les valoracions individuals de cada usuari per a generar recomanacions musicals; es podria analitzar l'evolució temporal del nombre de publicacions al llarg de l'any 2020 i comparar-la amb la d'altres anys per veure l'impacte que ha tingut el Covid-19 en aquesta faceta de la indústria musical...

En resum, el joc de dades presentat mostra únicament una part de la informació que es podria extreure d'aquest web. S'ha optat per a presentar aquesta versió més simplificada perquè el joc de dades no s'utilitzarà per a cap anàlisi posterior i, per tant, el cost computacional que tindria obtenir una versió més complexa no està justificat, però a partir del codi presentat es podria ampliar fàcilment el joc de dades.

## **Llicència**

La llicència triada és CC BY-NC-SA 4.0.

S'ha triat, principalment, perquè considero que es tracta d'un codi amb bastant potencial però que necessita ampliar-se per adaptar-se a les necessitats que tinguin l'usuari que l'utilitzi. Aquesta llicència permet que el codi s'utilitzi i es modifiqui lliurement sempre que es mantinguin els reconeixements necessaris.

Per altra banda, no sé quin ús comercial es podria extreure del joc de dades, així que podria optar per la llicència que en permet l'ús, però com la llicència inclou també l'apartat SA (ShareAlike), que demana que la llicència no es canviï encara que es modifiqui el codi, he optat per triar aquesta opció per protegir els interessos de possibles persones que modifiquessin el codi.