

# PRAC 1

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

Aquest dataset tracta dels videojocs. En ell trobareu videojocs de totes les generacions de totes les plataformes. He triat mobygames com a font per a la descarrega, ja que la he trobat molt completa, amb una gran quantitat de dades que podria aprofitar per a respondre varies preguntes sobre aquest context..

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

Llistat de videojocs multigeneracional

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).

Aquest dataset consta d'un llistat de videojocs, amb diferents característiques des de 1951 fins a 2020. Algunes de les dades no són videojocs, sinó que són complements per a videojocs, o edicions especials, aquestes dades s'eliminaran en la neteja de dades. Les principals característiques seran el nom, la data, el gènere i les plataformes.

4. Representació gràfica. Presentar una imatge o esquema que identifiqui el dataset visualment



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

Tenim un total de 114712 registres, i per cada registre tenim valors possibles a 24 columnes diferents. Entre parèntesis indicaré el nombre de valors nulls de cada atribut. El 5 atributs que estan en negreta seran el valor principals del dataset. La resta, es veurà a la hora de realitzar la neteja de dades, a veure quins ens quedem, quins s'eliminen, o si combinem alguns atributs.

- **GameTitle** (1): Nom del videojoc
- **Genre** (67): Gènere del videojoc
- **Platform** (52076): Plataforma del videojoc (nomes una)
- **Platforms** (62638) Plataformes del videojoc (quan es multiplataforma)
- **Released** (1): Data de sortida
- Published by (6932): Editora del videojoc
- Developed by (46304): Desenvolupador del joc
- Add-on (88717): Indica si es un complement per a un videojoc
- Gameplay (57539): Característiques del joc
- Official Site (86736): Nom de la pagina oficial
- Perspective (45287): Perspectiva
- Setting (84194): Ambientació
- Vehicular (105938): Si un joc de vehicles, indica el tipus de vehicle
- Narrative (104000): Tipus de narrativa
- Pacing (99808): Ritme del joc
- Special Edition (112248): Indica si es una edicio especial
- ESRB Rating (104769): Classificació per edat ESRB
- Interface (92893): Tipus d'interfície
- Misc (103138): Varies característiques: regional, llicencia...
- Visual (85918): Visualització
- Art (106981): Estil de disseny del joc
- Sport (108816): En els joc esportius, indica el esport
- Educational (111895): En els jocs educatius, indica la branca
- Amazon Rating (114710): Classificació per edat amazon

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

Agraïments especials a Cloudflare Inc., propietari de la plana web des de la que he realitzat la recol·lecció de dades mobygames.com. Per a realitzar això, he utilitzat tècniques de web escraping amb Python tenint com a referencia els llibres “Web Scraping with Python, Packt Publising” de Lawson, R i “Web scraping, Oberta UOC Publishing” de Laia Subirats Maté i Mireia Calvo González..

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

Me he decidit a realitzar aquest dataset, per la meva afició als videojocs de qualsevol plataforma, ja sigui antiga o d'última generació. Gracies a aquest dataset, podrem estudiar la evolució dintre del mons dels videojocs com per exemple quina consola a sigut la mes longeva o com a anat modificant-se el gènere mes usat al llarg dels anys.

8. Llicència. Seleccionar una d'aquestes llicències pel dataset resultant i explicar 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

He seleccionat la llicencia CC BY-NC-SA 4.0 la qual especifica que es pot utilitzar les dades, però no es poden utilitzar ni les dades originals ni les derivades per fines comercials, i la distribució de les possibles obres derivades hauran de ser sota la mateixa llicencia.