

# Anàlisi de l'ús de les Llengües a plataformes de video OTT (Netflix, Disney Plus)

Joan Carles Badia Purroy

Aquests dies ha estat d'actualitat la polèmica sobre el percentatge d'ús del català i llengües minoritàries a les plataformes digitals com ara Netflix i la seva indefensió per part del govern de l'Estat. Vegeu:

<https://www.lavanguardia.com/politica/20211123/7882755/gobierno-obligara-netflix-ofrecer-6-produccion-lenguas-cooficiales.html>

o aquí:

<https://www.naciodigital.cat/noticia/227364/que-implica-el-6-en-catala-a-les-plataformes-per-la-llei-de-audiovisual>

Tanmateix aquests anàlisis em semblen simplistes. Com estan repartit aquests percentatges en les diferents variables? No és el mateix un 6% de pel·lícules que no veu ningú que aquest 6% en pel·lícules mainstream. O no és el mateix un 6% de pel·lícules antigues que noves. A quin públic va dirigit aquest 6%? Son els joves qui haurien de veure aquest contingut en català per a la salut de la llengua. És el públic objectiu d'aquest 6%? Quins percentatges té cada llengua en les plataformes?

Per aquest motiu m'ha semblat interessant aconseguir visibilitzar realment l'ús de les diferents llengües en el contingut de la principal plataforma d'entreteniment, i desenvolupar-lo desplegant en funció d'altres variables importants. Per a fer-ho necessitarem datasets amb variables com ara la popularitat, l'idioma, el país en el que es pot veure, la qualificació d'edat, l'any de la producció etc.. Per a la popularitat cal que el dataset amb dades de netflix estigui combinat amb dades de portals de valoració de films, com ara IMDB o Rotten tomatoes.

Per tal d'aconseguir una visió més global afegiré a la visualització dades de la plataforma Disney Plus. D'aquesta manera podrem comparar l'ús de la llengua entre les dues plataformes tenint en compte la plataforma com a variable categòrica. Per tal d'aconseguir-ho combinaré el dataset de Netflix amb un altre dataset amb dades de Disney Plus, utilitzant com a nexa d'unió un camp que sigui comú a tots ells. El més adient seria utilitzar el codi de imdb. Per a poder fer-ho cada dataset ha de tenir les dades de la plataforma combinades amb la corresponent dada a IMDB. També s'haurà d'unificar els diferents valors de les categories de cada dataset per a que es

puguin combinar. M'hauria agradat afegir altres plataformes però no he trobat cap dataset de Prime Video o HBO Max que hi inclogui la llengua.

Per a la plataforma Netflix les dades seràn procedents del dataset de [kaggle.com](https://www.kaggle.com/ashishgup/netflix-rotten-tomatoes-metacritic-imdb)

<https://www.kaggle.com/ashishgup/netflix-rotten-tomatoes-metacritic-imdb>

Aquest dataset està format per 29 columnes i un total de 15481 registres actualitzat a data 27-04-2021, però no totes ens seran útils o necessàries per a la visualització. El que si que s'haurà de preparar és la manera en la que estan emmagatzemats alguns camps com ara les llengües, el país de cada pel·lícula o el gènere, que es troben en un sol camp separat per comes i s'hauran de separar en diferents registres per a poder-ne fer l'anàlisi.

Per a la plataforma Disney + les dades seràn del dataset <https://www.kaggle.com/unanimad/disney-plus-shows> . Aquest dataset està format per 19 columnes i 993 registres.

Podem trobar per la xarxa diverses visualitzacions sobre netflix i altres plataformes, que van més encarades a comparar diferents variables (nombre de pel·lícules, etc...) per tal d'oferir arguments per tal de triar-ne una o una altra, o per veure quina plataforma té més guanys, però no encarats a la llengua. Per exemple:

<https://www.kaggle.com/niharika41298/netflix-visualizations-recommendation-eda>

<https://towardsdatascience.com/data-analysis-on-ott-platforms-which-service-should-i-choose-8eed953ff7d2>