

# Pràctica 1

## Membres del grup:

- Aarón Puche Benedito
- Roger Pardell Carrera

## 1. Context

En un moment de pandèmia global, on molts governs han trobat en el confinament i les quarantenes la solució temporal a la situació, l'aturada de les interaccions socials ha provocat que l'oci s'hagi de consumir a casa i en petita escala. Molta gent ha trobat el passatemps perfecte en plataformes de pel·lícules en streaming, les quals proporcionen una quantitat de llargmetratges considerables. El problema sorgeix, masses vegades, en triar quin títol volem veure.

De la necessitat de tenir un document amb els noms de pel·lícules sorgí l'interès per a fer un Web Scraping d'una pàgina web amb aquesta informació: The Movie Database. Aquesta pàgina ens proporciona el conjunt d'informació que ens interessa: títol, any d'estrena, gènere, sinopsis, etc. Així doncs, podem proporcionar un llistat únic de tots els films fins ara rodats.

Tot i que pot semblar un argument banal, aquestes dades poden tenir un important valor per analitzar els gustos dels consumidors i productors al llarg dels anys els quals, barrejats amb elements històrics, es podrien treure conclusions de caràcter sociològics importants.

D'altra banda, aquest conjunt de dades podria ser fàcilment incorporat en aplicacions de caràcter comercial. Podria ajudar-nos a escollir la plataforma de vídeos en streaming que reproduïx més títols del nostre interès, o bé proporcionar les preferències dels consumidors respecte els gèneres o les sinopsis.

## 2. Títol del dataset

Pel·lícules d'acció estrenades entre el 2000 i el 2005.

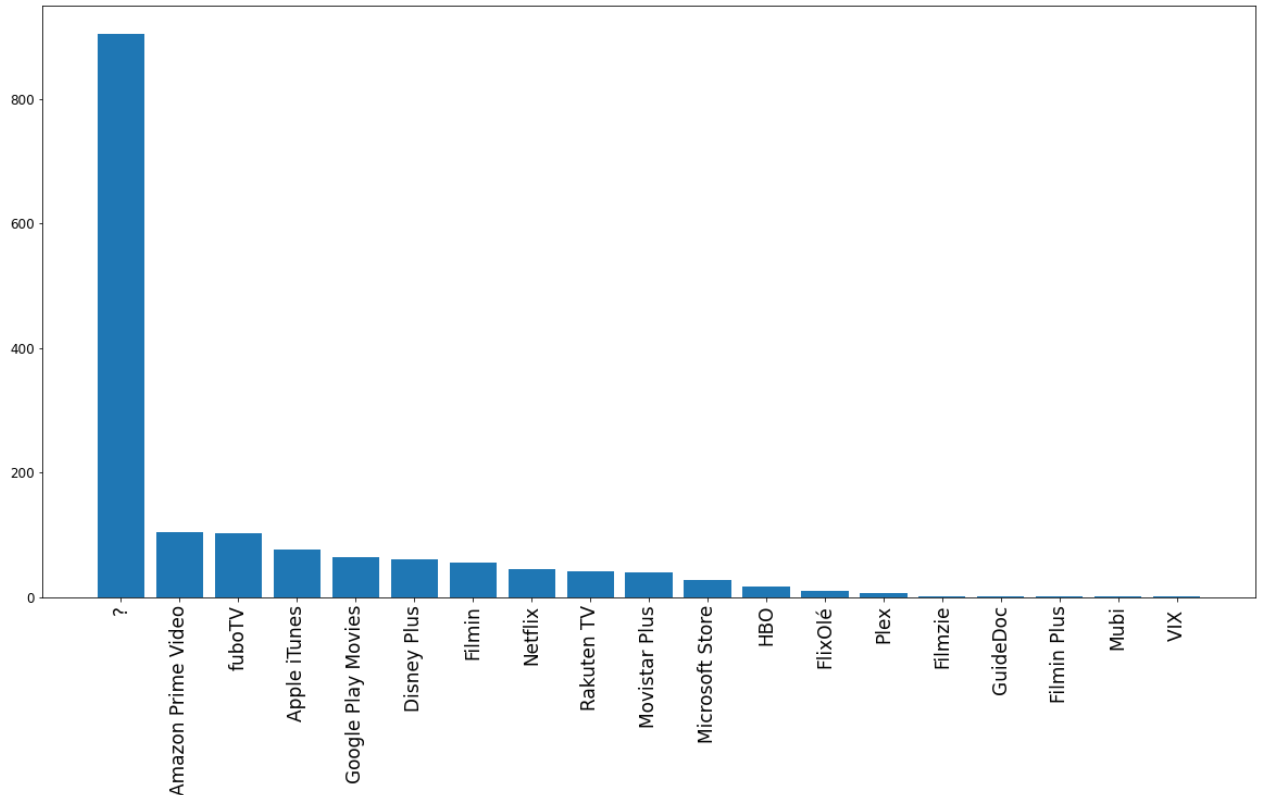
## 3. Descripció del dataset

El data set resultant és compost de XXX observacions, corresponent a totes les pel·lícules d'acció estrenades entre el 2000 i el 2005, i 12 atributs. Aquests atributs són:

- Títol nom de la pel·lícula en castellà d'Espanya.
- Títol original.
- Idioma original.
- Duració, en el format "hores minuts".
- Gèneres, a més del d'acció.
- Directors.
- Sinopsis.
- Data d'estrena.
- Classificació per edat.

- Enllaç a la imatge de cartell.
- Puntuació obtinguda pels visitants de la pàgina web.
- Plataforma de streaming on podem trobar-la.

#### 4. Representació gràfica



En aquest gràfic tenim representada de forma visual la distribució de les pel·lícules segons la plataforma de streaming a la qual pertanyen.

El valor “?” indica que no pertany a cap.

#### 5. Contingut

Com s’ha introduït en l’apartat 3, cada observació consta dels elements més bàsics i importants de cada pel·lícula. Així doncs, podríem classificar títol, títol original, duració, directors, gèneres, sinopsis i enllaç de cartell com a variables alfabètiques; idioma original, plataforma de streaming i classificació per edat com a variables categòriques; data d’estrena com a variable de temps; i puntuació com a variable numèrica.

El període de temps és entre 2000 i 2005.

La recollida de la informació s’ha efectuat amb Python, utilitzant l’eina Selenium. Se li ha donat l’enllaç de la pàgina on s’exposa el conjunt de pel·lícules i, a partir d’allà, s’han seleccionat els filtres. Acte seguit, el codi estén la pàgina fins a carregar el màxim de pel·lícules amb aquelles característiques. Després, es guarda en una llista les

terminacions de totes les extensions de les pel·lícules. Amb el llistat, el codi carrega la pàgina de cada film i n'extreu les dades rellevants per el nostre projecte. Aquestes són emmagatzemades en un fitxer csv.

## 6. Agraïments

Hem decidit extraure informació de la web [The Movie Database](#).

Aquest lloc és un projecte que va començar en 2008 on una comunitat de gent començà a agregar i editar informació sobre pel·lícules i series. Com diuen ells, volien crear una Wikipedia molt especialitzada.

A la seua pagina podem veure alguns dels projectes on s'han utilitzat les seues dades

A partir de les dades d'aquesta web s'han creat algunes aplicacions. Podem veure algunes de les aplicacions més destacades en aquest [enllaç](#). Ens ha cridat l'atenció l'app de [Cinematics](#), disponible a Android. Ens permet veure informació sobre pel·lícules i sèries d'una manera senzilla i molt fluida.

## 7. Inspiració

En aquesta web podem aconseguir un conjunt de dades molt interessant. En permet tant fer anàlisis com crear aplicacions a l'estil de l'app de Cinematics.

Pel que fa a les anàlisis, podríem veure com evolucionen els distints gèneres llarg del temps. Hi ha un gènere que destaca en una franja d'anys concreta? També podem veure la situació de les plataformes streaming. Hi ha alguna plataforma que sigui més predominant? A partir de quin any comencen a tindre un volum significat de pel·lícules?

Per l'altra banda, podem realitzar alguna aplicació de seguiment de pel·lícules. A més, a partir de les dades dels usuaris que utilitzaren la nostra web podríem crear un sistema de recomanació i mostrar de forma més precisa les pel·lícules més interessant que els hi puguin agradar.

## 8. Llicència

Hem decidit que aquest dataset estigui baix la llicència de [Public Domain License](#)

Creguem que és interessant fer anàlisis com veure com evoluciona la producció de pel·lícules al llarg dels anys o analitzar l'impacte de les plataformes de streaming. Per això deixem lliure el dataset per a que, qui vulgui, pugui fer els seus propis anàlisis.

I també, si a partir del dataset algú vol crear la seva pròpia aplicació amb propòsits comercials, serà lliure de fer-ho.

## 9. Codi

Codi en el Jupyter Notebook "Obtenir dataset pel·lícules.ipynb"

## **10. Dataset**

Publicar el dataset en format CSV a Zenodo (obtenció del DOI) amb una breu descripció.