

## Data

- Introduce to the data  
我們要使用[kaggle連結](#)中的credit.csv,keywords.csv,links.csv,movies\_metadata.csv及rating.csv。全部csv皆來自於IMDB這個網路電影資料庫，包含了自2017七月以前的電影作品
- Where is the data from  
資料是kaggle網站上的資料[kaggle連結](#)，來自於IMDB這個網路電影資料庫
- What do the columns and tables mean  
credit.csv :此csv用來記錄每部電影的相關人員  
cast:演員資料，含有姓名,飾演角色等基礎資訊  
crew:工作人員資料，含有姓名,職位等資訊  
id:primary key，每一部電影都會有一個獨特的id  
=====  
keywords.csv :用來記錄每部電影的關鍵字  
id:primary key，每一部電影都會有一個獨特的id  
keywords:即是電影的所有的關鍵字 EX:Toy  
=====  
links.csv :紀錄每個id對英的imdbid及temdbid  
movieid:即是credit及keyword中的id， foreign key  
imdbid:在imdb網站中的id  
temdbid:在temdb網站中的id  
=====  
movies\_meatdata.csv :用來記錄每部電影的詳細資料  
adult:true代表著成人片， false則反，只有二分法的分級且true佔極少量  
belongs\_to\_collection:有續集或其他篇章的才有，記錄了該作品的系列 EX:toy story collection'  
budgets:即為預算(美金)  
genres:代表電影的分類 EX:Animation,Comedy  
homepage:電影的網站  
id:即foreign key  
imbd\_id:在imdb網站中的id  
original\_language:原片的語言  
original\_title:原片的標題  
ovewrview:電影大綱  
popularity:就是人氣  
poster\_path:電影海報的網址  
production\_companies:製作公司  
production\_countries:製作國  
release\_date:上映日  
revenue:收益(美金)  
runtime:影片時長(min)  
spoke\_languages:所說的語言(可有複數個)  
status:電影的狀態(是否上架之類的

tagline:標語

title:標題

video:在imbd中是否有影片存在

vote\_average:評分平均

vote\_count:總投票數

=====

ratings.csv :各用戶針對不同電影的評分

userId:使用者id

movieId:所評分的電影id

rating:評分, 從零到五且精度為小數點一位

timestamp:評分當下的時間戳

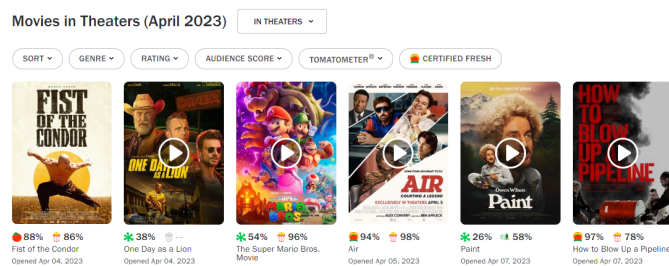
- Data link

採用的資料庫連結:<https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>

## Application Design

- Main idea

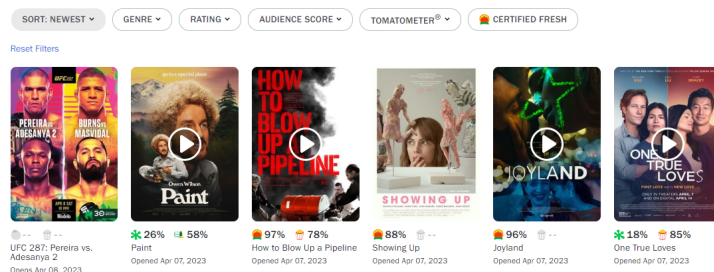
我們希望透過這個網站和資料庫, 來達成篩選或推薦電影的目的, 因此我們希望使用者在瀏覽網站時可以透過我們給定的標籤來篩選或是找尋想看的電影, 並且如果能補上2017至今的檔案, 可能還可以更進一步地做到一些本周上映、或是本周熱映的功能來供使用者搜尋喜愛的院線片, 如<sup>1</sup>[rotten tomatoes](https://www.rottentomatoes.com/):



- Functionally

蒐集kaggle中各個類型movies.csv, 我們可以透過底下的分類, 並且將其做一個排序, 根據使用者想查詢的優先順序來做篩選, 更進一步地, 我們或許還可以從關鍵字去做搜尋

例如: 上映時間、類型, 如下圖:



GENRE

×

ACTION

ADVENTURE

ANIMATION

ANIME

BIOGRAPHY

COMEDY

CRIME

DOCUMENTARY

CLEAR ALL

APPLY

Popular TV Show and Movie Genres

COMEDY

SCI-FI

HORROR

ROMANCE

ACTION

THRILLER

DRAMA

MYSTERY

CRIME

ANIMATION

ADVENTURE

FANTASY

COMEDY-ROMANCE

ACTION-COMEDY

SUPERHERO

另外我們在rating.csv檔中已蒐集了一部份IMDb和GroupLens使用者的評價，而使用者在網站上也能增添對於某部電影的評價，並且我們也會將其評價(星數)作為一筆新的資料加入我們的rating.csv中讓使用者能在觀賞完電影後能夠將反饋傳至網站。

## Work Plan

Time schedule:

	建立資料 庫伺服器(XAMP)	建立web 的伺服器(python)	資料庫 管理(mysql)	
近期				遠期

而我們會先建立好相關的資料庫，爾後再建立web並套用application

Discussion: 我們使用Discord來討論，但是Discord的群組對話似乎沒有辦法直接公開的樣子

Repo: <https://github.com/Duck997/2023-Spring-NYCU-Intro-To-DB>