

Task: Gather data from IMDB movie ratings for a list of multiple URLs. Many will be duplicates. This is fine.

Step 1:

Scrape IMDB content ratings for several lists of content

1. 175 URLs / videos from here <https://www.imdb.com/search/title?certificates=ID:SU>
 2. 49 URLs / videos from here <https://www.imdb.com/search/title?certificates=ID:21%2B>
 3. 186 URLs / videos from here <https://www.imdb.com/search/title?certificates=ID:13%2B>
 4. 253 URLs / videos from here <https://www.imdb.com/search/title?certificates=ID:17%2B>
 5. Top 200 (first 4 pages) from here <https://www.imdb.com/search/title?certificates=US:R>
 6. Top 200 (first 4 pages) from here https://www.imdb.com/search/title?title_type=feature
- Total of 1063 videos, many will be duplicates

Result:

Get the ID of each content: https://www.imdb.com/title/tt0217869/?ref_=adv_li_tt

Step 2:

Search for the content ID to get the content ratings:

<https://www.imdb.com/title/tt0217869/parentalguide> <— Replace with the ID

Scrape:

- 1) Content certifications:
 - a) This is the list mentioned below in the first image. I.e: ["Argentina:Atp", ... , "United States:PG"]
- 2) User Rating distributions (# None, # mild, # moderate, # Severe) "Sex & Nudity", "Alcohol Drugs and Smoking", "Frightening and intense scenes", "Profanity" (note, you need to click though / grab from the "ipl-vote-button__details" span class)
- 3) User comments

Output required for delivery (Google Spreadsheet):

Content ID	Title	Ratings (all, comma separated)	Sex Nudity None	Sex Nudity mild	Sex nudity moderate	Sex nudity severe	Sex Nudity Comments	[other areas in other columns]
tt54765476	"Title of	"ID:17+","U	0	5	25	5	"Comment	

	Movie"	S:R", etc...					1", "Comment 2"	
All ~								

Images below:

All content certification box to be scraped for each title

Certification	Argentina:Atp	Australia:PG	Belgium:KT/EA	Brazil:Livre	Canada:PG (Alberta)	Canada:PG (British Columbia)	France:Tous publics	Germany:6	Greece:K	Hungary:6	Indonesia:SU	Ireland:PG	Italy:T	Japan:G	Malaysia:P13	Mexico:A	Netherlands:6	Philippines:PG-13	Portugal:M/6	Russia:6+	Singapore:PG	South Korea:All	Spain:A	Sweden:7	Switzerland:8	Taiwan:GP	United Kingdom:PG	United States:PG (certificate #51520)
---------------	---------------	--------------	---------------	--------------	---------------------	------------------------------	---------------------	-----------	----------	-----------	--------------	------------	---------	---------	--------------	----------	---------------	-------------------	--------------	-----------	--------------	-----------------	---------	----------	---------------	-----------	-------------------	---------------------------------------

Below is an example of the box with user ratings and comments.

Sex & Nudity
Mild 10 of 38 found this mild
There is absolutely no sex throughout this entire film. No nudity except Jack-Jack's diaper change.
A brief kiss between 2 characters
Characters of various genders wear spandex leotards (obviously, because it's a superhero movie)
Helen, AKA ElastiGirl wears a rather revealing black-and-grey skin-tight suit throughout most of the film.
Elastigirls suit is very tight as the film focuses on her curves
Helen wears a white bathrobe revealing her legs.

You need to click through to expose the ratings distribution (it is hidden in a div, so can be scraped as well):

Sex & Nudity

Severity? None 17 Mild 10 Moderate 4 Severe 7

There is absolutely no sex throughout this entire film. No nudity except Jack-Jack's diaper change.

All content areas need to be scraped (5 in total):

Alcohol, Drugs & Smoking

Mild 9 of 25 found this mild

Helen and others drink wine and cocktails at parties and the like.

Frightening & Intense Scenes

Mild 11 of 27 found this mild

Scenes of superheroes turned to villains vs. kids are scary to children.

At the beginning of the film, a boy is briefly shown being memory-wiped, and he appears to have been knocked unconscious. Might be slightly worrying to some younger viewers.

[others are not screen shot here, but are listed below again]

“Sex & Nudity”, “Alcohol Drugs and Smoking”, “Frightening and intense scenes”, “Profanity”, “Frightening & Intense Scenes”