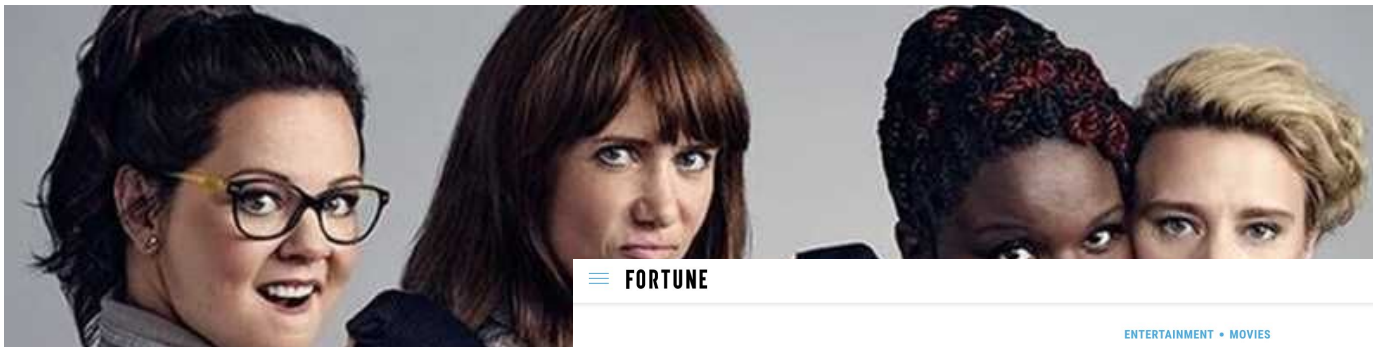


Gender Diversity in Film and Negative Online Backlash

Rachel Franz, Eric Gomez, Jamie Park,
Kate Van Ness, Trent Dillon



FORTUNE

HOME

ENTERTAINMENT • MOVIES

New Female 'Ghostbusters' Trailer Is Bringing Out the Misogynists

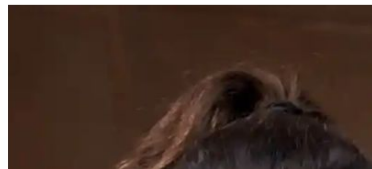


You May Like



Ghostbusters trailer is most disliked in YouTube history

With more than 600,000 thumbs down, Feig's reboot becomes the first streaming channel's 100 least-liked



Sections

The Washington Post
Democracy Dies in Darkness

Sign In

Arts and Entertainment

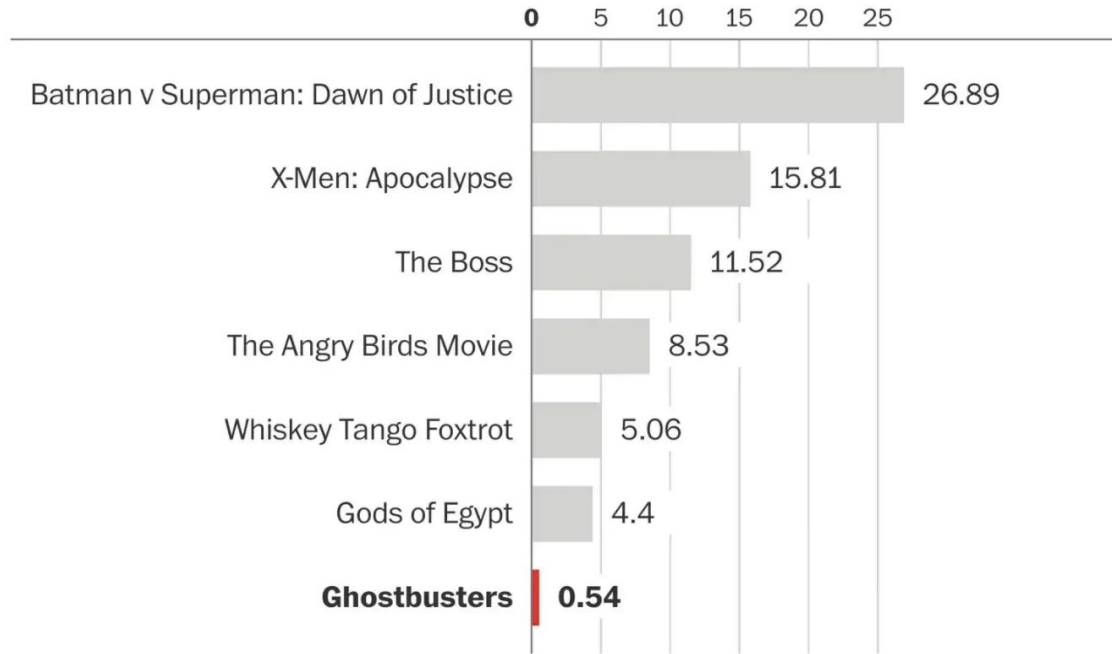
People hate the 'Ghostbusters' trailer, and yes, it's because it stars women

By [Stephanie Merry](#)
March 4, 2016

Most Read [Lifestyle](#)

Ratio of likes to dislikes for selected YouTube movie trailers

A ratio of less than one means there are more dislikes than likes



The data measures the first official theatrical trailers posted at the studio's YouTube channel.
"Batman v Superman" is the Comic-Con trailer.

Source: YouTube

THE WASHINGTON POST / ZACHARY PINCUS-ROTH



+



Are online
commenters biased
against movies that
feature women?

Star Wars: The Last Jedi Controversy



“...politically correct to the point of boredom”

“...they ruined the franchise. It’s basically extreme feminist social justice propaganda”

“...feminism and diversity have ruined Star Wars for me as well as my kids”

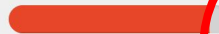
TOMATOMETER ⓘ



91%

Average Rating: 8.1/10
Reviews Counted: 415
Fresh: 377
Rotten: 38

All Critics | Top Critics



Critics Consensus: *Star Wars: The Last Jedi* honors the saga's rich legacy while adding some surprising twists – and delivering all the emotion-rich action fans could hope for.

AUDIENCE SCORE ⓘ



45%
liked it

Average Rating: 2.8/5
User Ratings: 203,456

Theoretical Motivation

Threat to status

“Their problem isn’t *Star Wars*, their problem is being threatened” -- J.J. Abrams

High status groups (Whites, men, etc.) feel threatened by changes in the status quo/group hierarchy (Scheepers & Ellemers, 2005; Outten et al., 2011; Craig et al., 2014)

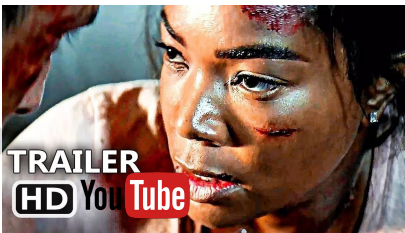
Social norms/expectations of gender roles

Women in leading roles may violate cultural norms of women as passive and not agentic (Eagly & Karau, 2002; Rudman & Glick, 2001)

Online disinhibition effect

People feel more comfortable behaving in typically unacceptable ways because the Internet provides a perceived shield of anonymity and psychological distance (Diener et al., 1976; Suler, 2004)

Theoretical Motivation



Threat to status



“Their problem isn’t *Star Wars*, their problem is being threatened” -- J.J. Abrams

High status groups (Whites, men, etc.) feel threatened by women in the status quo/group hierarchy (Scheepers & Ellemers, 2005; Outten et al., 2004)

Social norms/expectations of gender



Women in leading roles may violate cultural norms of women as passive and not agentic (Eagly & Karau, 2002; Rudman & Glick, 2001)



Online disinhibition effect

Are online commenters biased against movies that feature women?

People feel more comfortable behaving in typically unacceptable ways because the Internet provides a perceived shield of anonymity and psychological distance (Diener et al., 1976; Suler, 2004)

Component Overview: Initial Dataset



- 5000 movies
- ID, title, cast list with genders

| Cast Edit | | | |
|---|-------------------------------------|-----|--|
| Cast overview, first billed only: | | | |
|  | Lizzy Caplan | ... | Marlena Diamond |
|  | Jessica Lucas | ... | Lily Ford |
|  | T.J. Miller | ... | Hudson 'Hud' Platt |
|  | Michael Stahl-David | ... | Rob Hawkins |
|  | Mike Vogel | ... | Jason Hawkins |
|  | Odette Annable | ... | Beth McIntyre (as Odette Yustman) |
|  | Anjul Nigam | ... | Bodega Cashier |
|  | Margot Farley | ... | Jenn |

Component Overview: Gender Diversity Score



Gender
Diversity
Score (GDS)

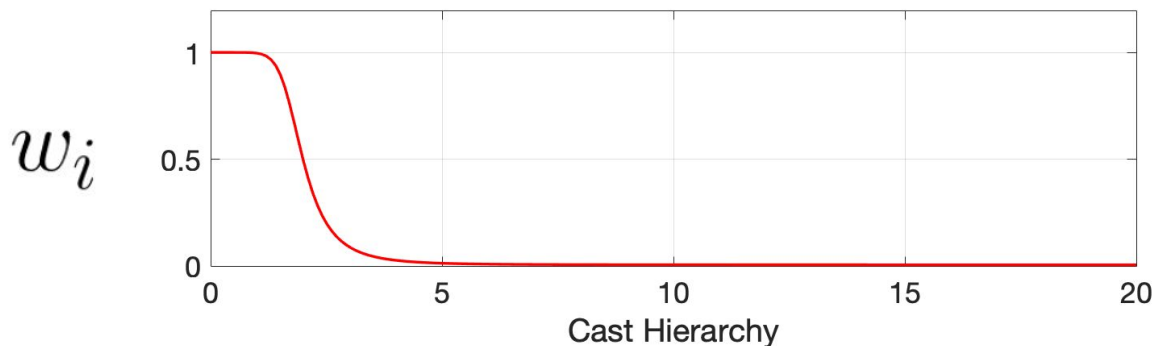
- 5000 movies
- ID, title, cast list with genders
- Calculate GDS from cast list for each movie

dual purpose:

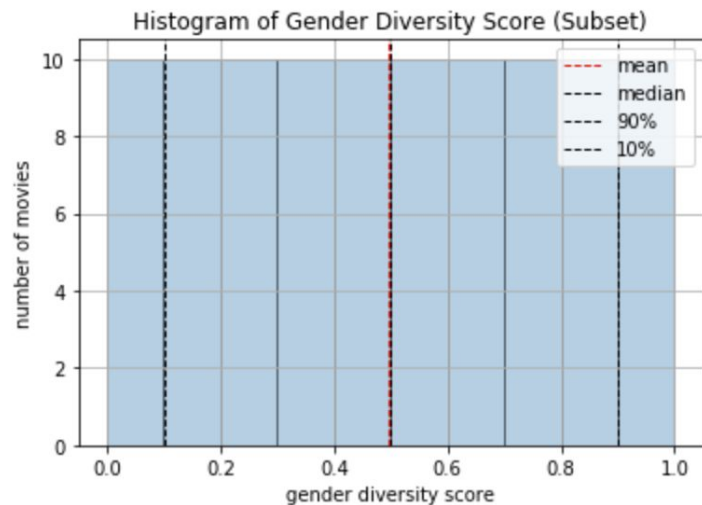
- 1. quantify diversity**
- 2. select subset**

$$= \frac{1}{n} \sum_{i=1}^n \alpha(G_i) * w_i \quad \alpha = \begin{cases} 0 & G_i = M \\ 1 & G_i = F \end{cases}$$

Asymmetrical Sigmoidal Weighting

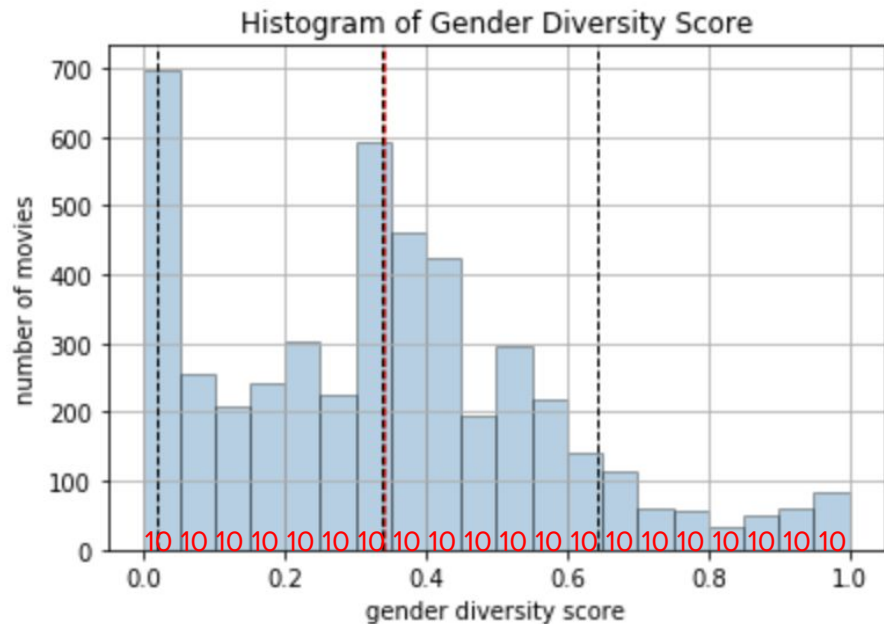


Component Overview: Gender Diversity Score

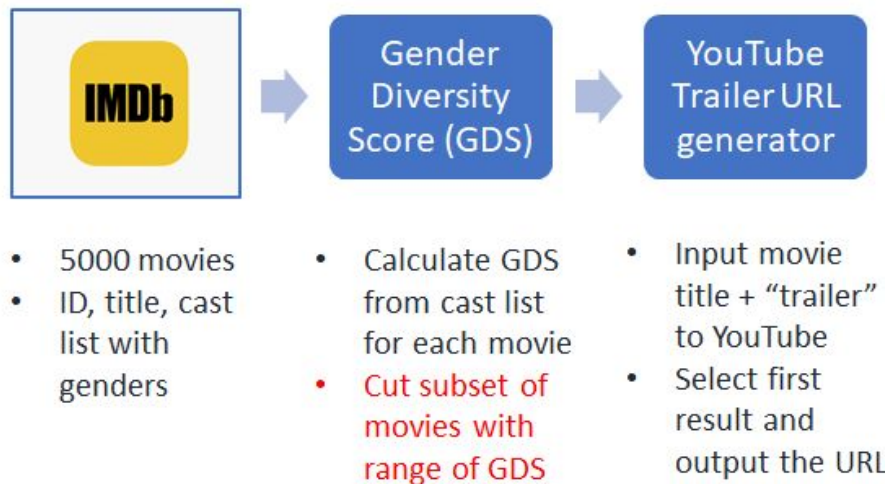


dual purpose:

1. *quantify diversity*
2. ***select subset***



Component Overview: Youtube URL Extraction



Component Overview: Youtube URL Extraction

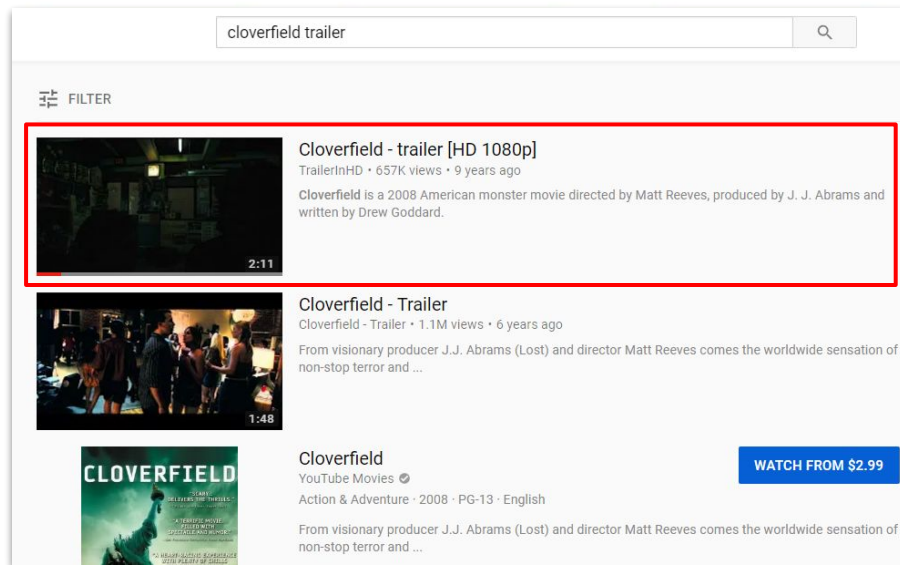


Gender
Diversity
Score (GDS)



YouTube
Trailer URL
generator

- 5000 movies
- ID, title, cast list with genders
- Calculate GDS from cast list for each movie
- **Cut subset of movies with range of GDS**
- Input movie title + "trailer" to YouTube
- Select first result and output the URL



Component Overview: Comment Scraper



- 5000 movies
- ID, title, cast list with genders

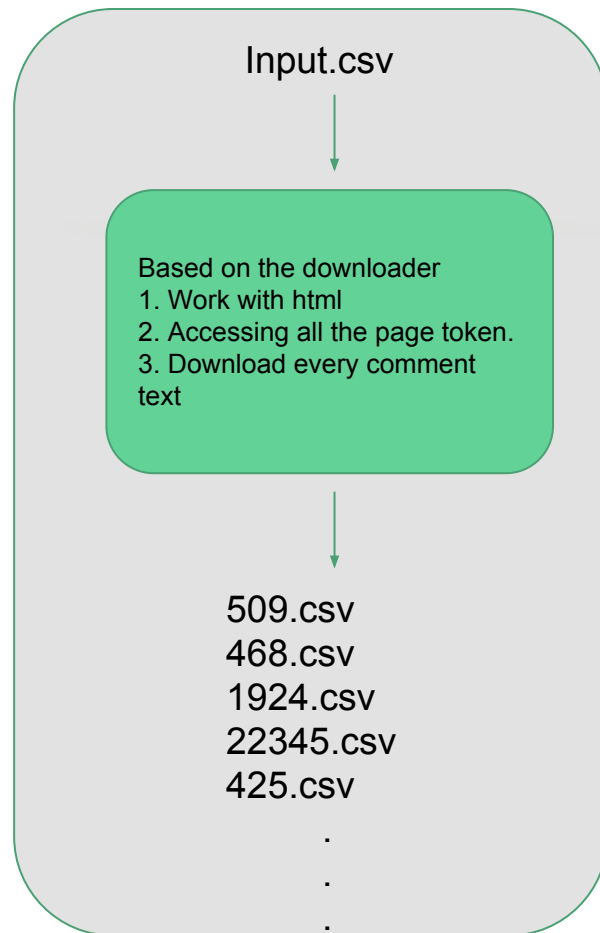
- Calculate GDS from cast list for each movie
- **Cut subset of movies with range of GDS**

- Input movie title + "trailer" to YouTube
- Select first result and output the URL

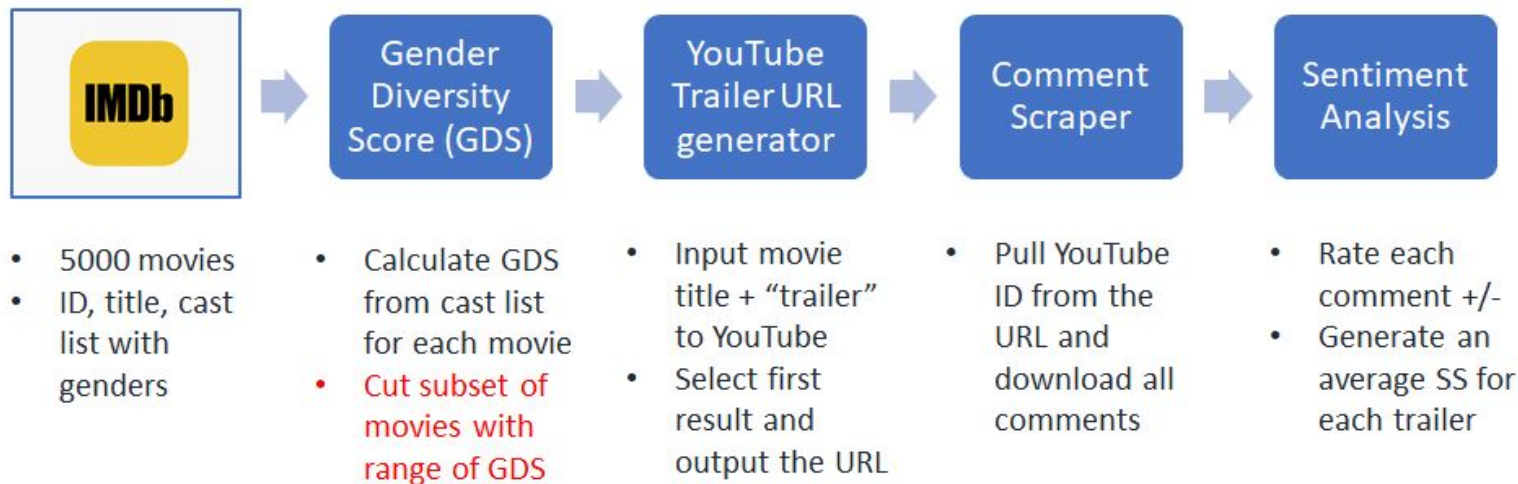
- Pull YouTube ID from the URL and download all comments

Reference :

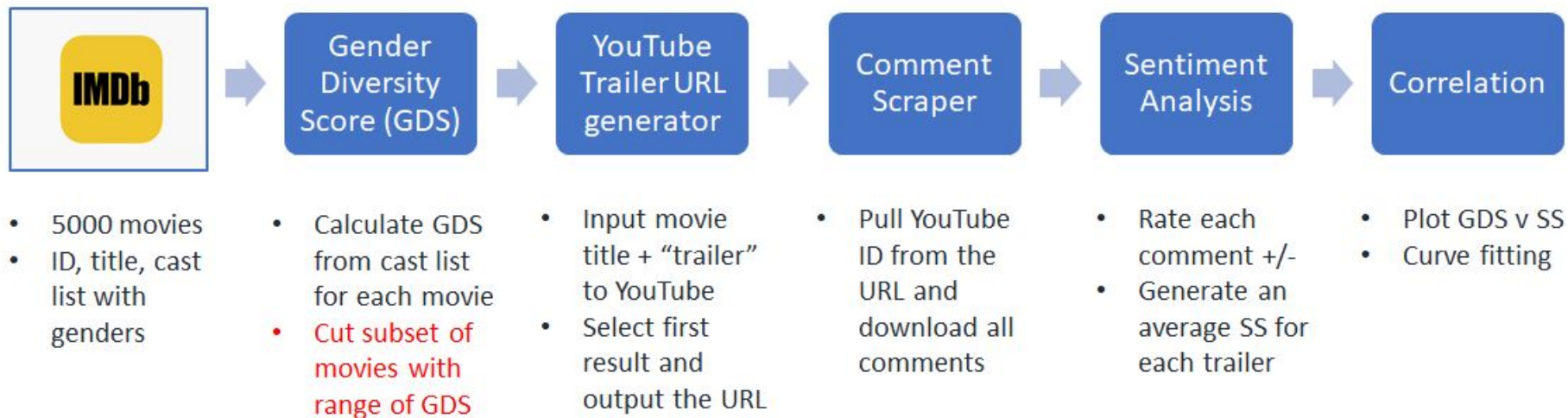
<https://github.com/egbertbouman/youtube-comment-downloader>



Component Overview: Sentiment Analysis



Component Overview: Correlation Analysis



Use Cases and More

Primary Use Case: Exploring gender representation in film and how people online receive gender diverse movies

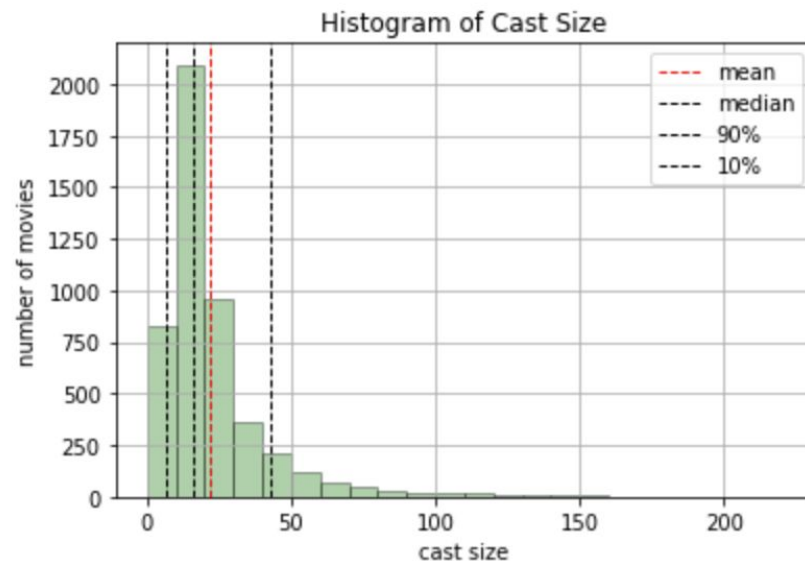
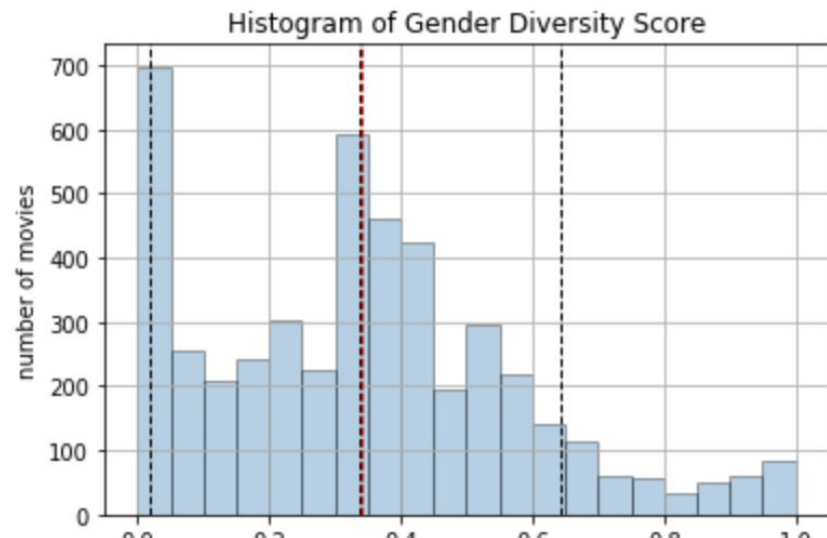
- Interaction:
 - Download visual output files to see results

Secondary Use Case: Specific components could be used for understanding how specific content is received online via sentiment analyses on YouTube comments:

- Create new queries for YouTube URLs
- Interaction: using comment downloader and sentiment analysis

Diversity Score and Subselection of Data

$$\text{score} = \frac{1}{n} \sum_{i=1}^n \alpha_i(G_i) \cdot \omega_i$$



Demo

Sentiment Score

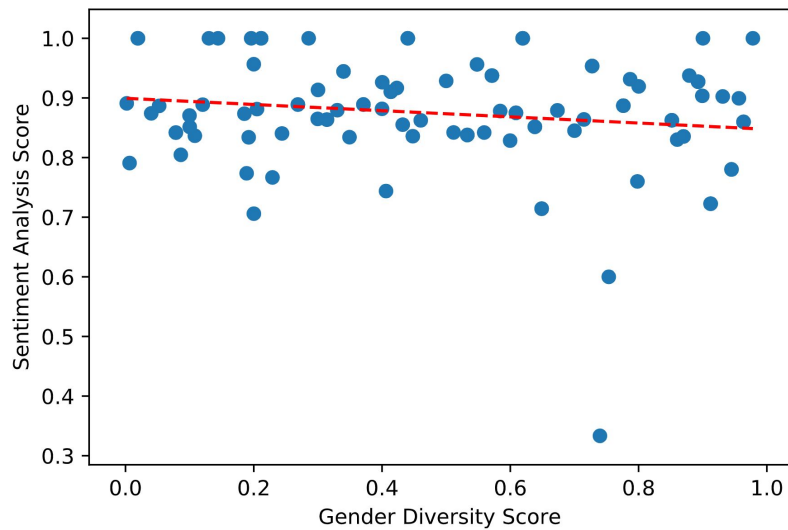
Tried 3 different techniques

1. Nltk VADER pre-trained classifier (.56 accuracy)
 - a. Trained on Twitter posts
2. Naive Bayes on a hand-labelled dataset (.82 accuracy)
 - a. Probably overfitting training data
3. TextBlob's pre-trained classifier (.72 accuracy)

Sentiment Score

- Chose Textblob because could achieve high accuracy without hand labelling data
- Labelled comments from 100 movies as 1 (for positive) or 0 (for negative)
- Averaged the labels for a sentiment score between 0 and 1
- Added sentiment score to the created dataframe

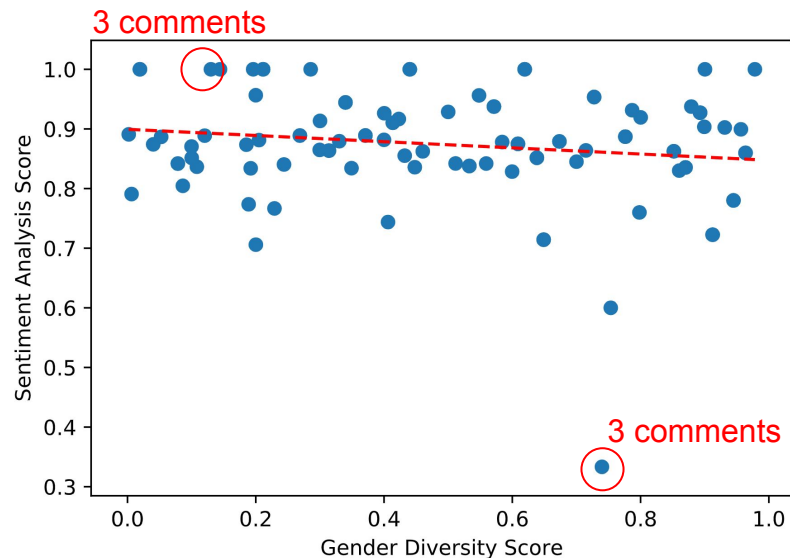
Results



- *Original objective:* Identify overall change in sentiment for more gender diverse films

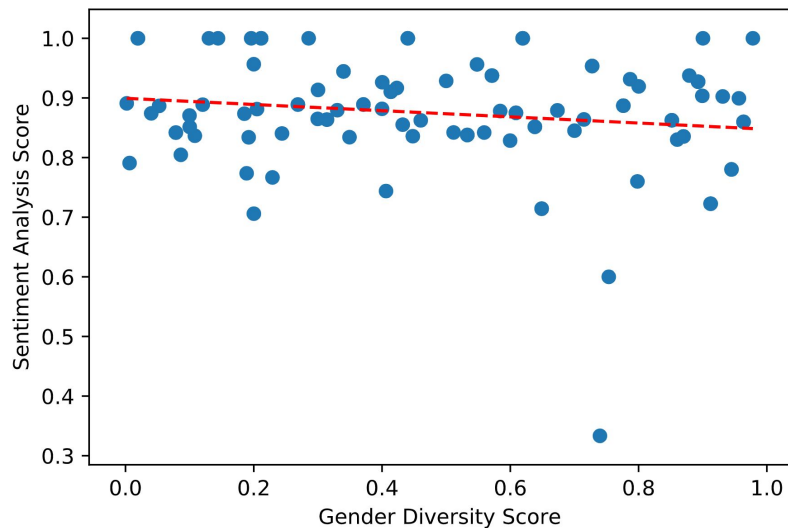


Results



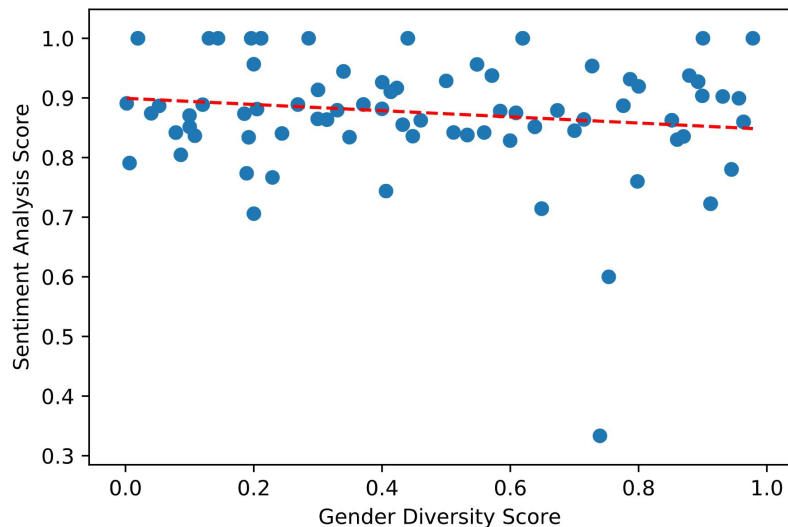
- *Original objective:* Identify overall change in sentiment for more gender diverse films
- *Outcome:* Slight decrease in sentiment dominated by films with too few comments

Results



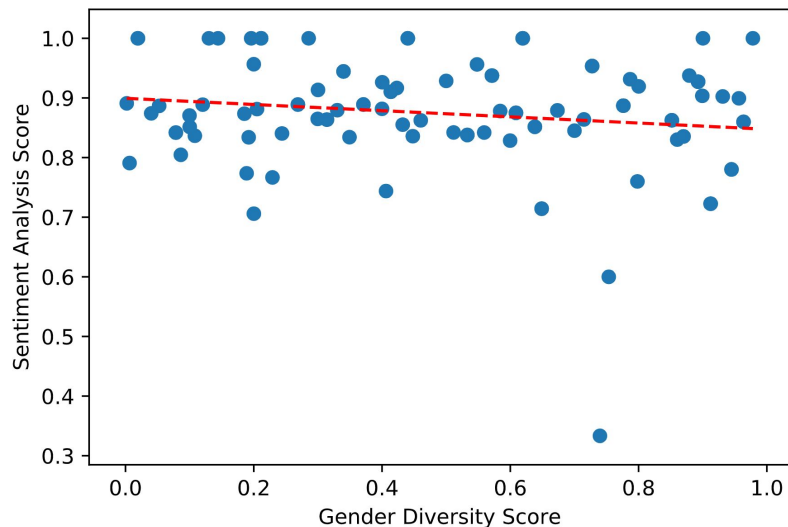
- *Original objective*: Identify overall change in sentiment for more gender diverse films
- *Outcome*: Slight decrease in sentiment dominated by films with too few comments
- *What happened?*

Results



- *Original objective*: Identify overall change in sentiment for more gender diverse films
- *Outcome*: Slight decrease in sentiment dominated by films with too few comments
- *What happened?*
 - Half of all movies with < 60 comments

Results



- *Original objective*: Identify overall change in sentiment for more gender diverse films
- *Outcome*: Slight decrease in sentiment dominated by films with too few comments
- *What happened?*
 - Half of all movies with < 60 comments
 - Automated YouTube URL selection tricky for small, independent films

Discussion

Do movies with more women receive more negative online backlash?

- According to our results, no. However, our analysis has limitations
 - small and limited sample size
 - unable to iterate subselection
- The data also has limitations
 - unable to assess popularity of movie via IMDB
 - not inclusive to trans actors and actresses
 - additional datasets clearly needed
- General conclusion
 - we think our work is a good starting point and framework, but more specific work and data is needed to make confident claims on the question at hand

Conclusion

Benefits:

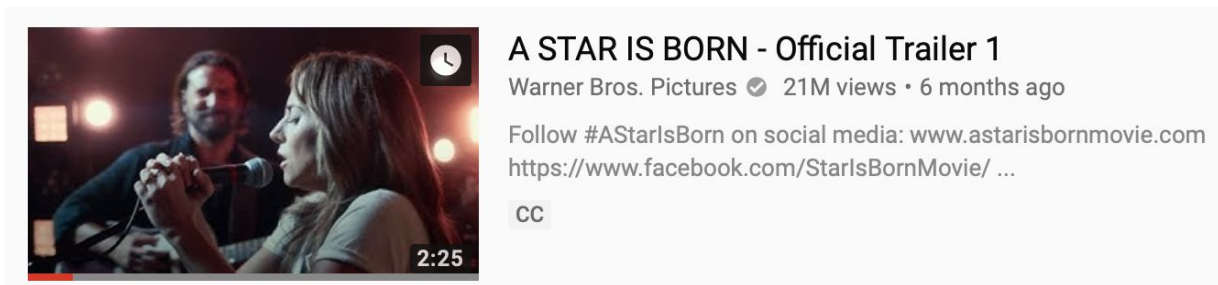
- Qualitative observation of how gender diversity in film is likely (or not) to affect online response
- Independent component use:
 - analyze your own subset of movies or one individual movie and compare its score to trends
 - analyze online response to content other than film

Future directions:

- Execute using entire IMDb dataset
- Remove films with < 100 comments

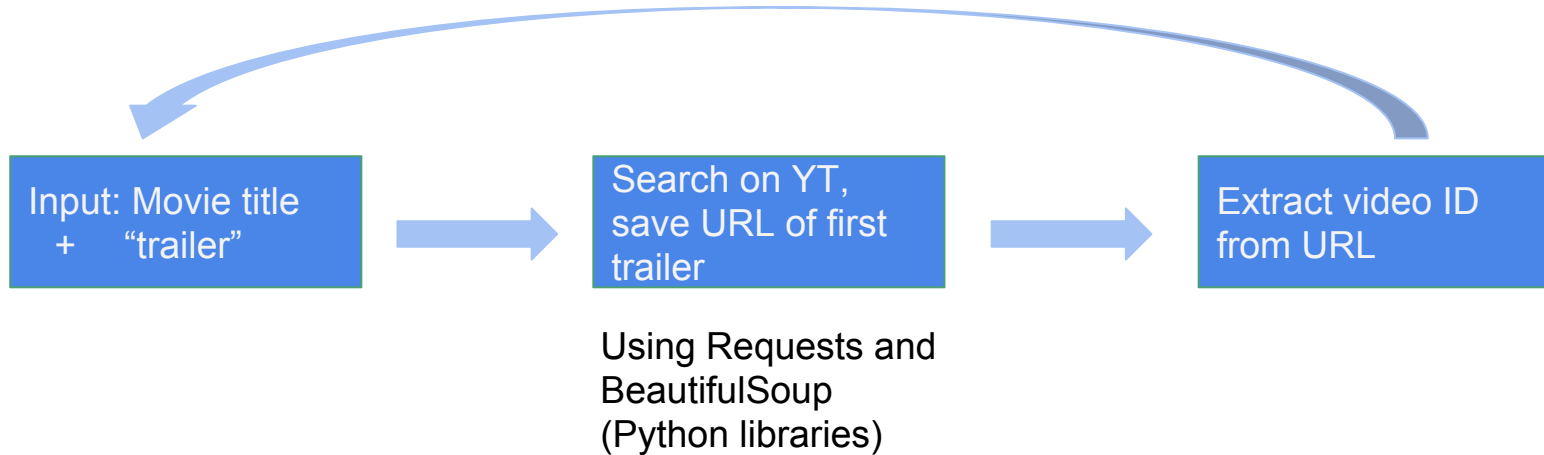
Video ID Extraction

Comment extraction program we found requires input of YT video IDs

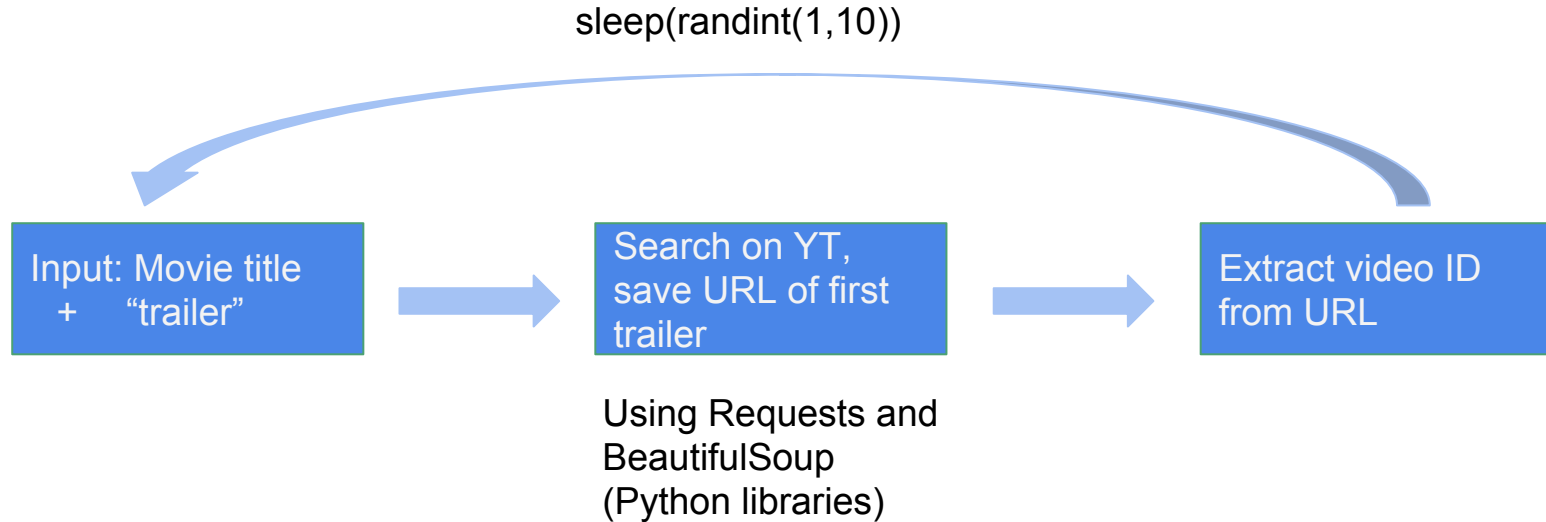


<https://www.youtube.com/watch?v=nSbzyEJ8X9E&t>

Video ID Extraction

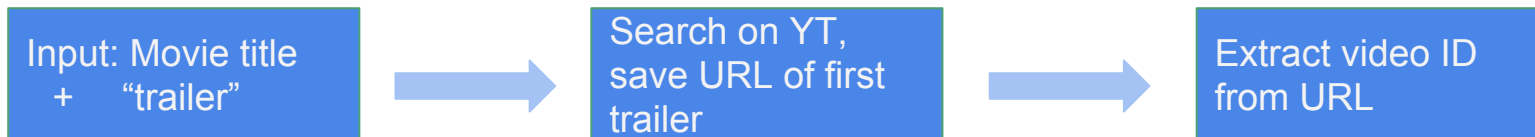


Video ID Extraction



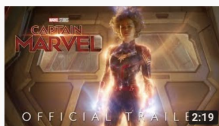
Video ID Extraction

`sleep(randint(1,10))`



a star is born trailer

FILTER



Marvel Studios

Ad Marvel Entertainment • 9.6M views
Captain Marvel - March 8 Samuel L. Jackson is Nick Fury.



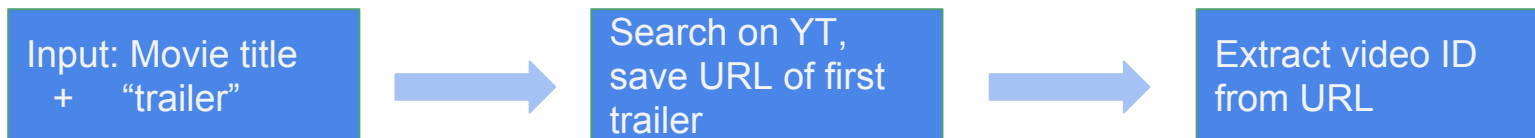
A STAR IS BORN - Official Trailer 1

Warner Bros. Pictures • 21M views • 6 months ago
Follow #AStarIsBorn on social media: www.astarishornmovie.com
<https://www.facebook.com/StarIsBornMovie/> ...
CC

Using Requests and
BeautifulSoup
(Python libraries)

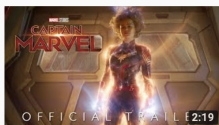
Video ID Extraction

`sleep(randint(1,10))`



a star is born trailer

FILTER



Marvel Studios

Ad Marvel Entertainment • 9.6M views

Captain Marvel - March 8 Samuel L. Jackson is Nick Fury.

A STAR IS BORN - Official Trailer 1

Warner Bros. Pictures • 21M views • 6 months ago

Follow #AStarIsBorn on social media: www.astarishornmovie.com
<https://www.facebook.com/StarIsBornMovie/> ...

CC

Using Requests and
BeautifulSoup
(Python libraries)

All Comments

Comments are disabled for this video.