

# **Powerhouses of Cinema:**

Uncovering an AAA title

August 14, 2021 | Mikee Sevilla

## **Production of films is a key investment made by production and entertainment studios.**

Due to cost, time, and effort required in producing a film, production decisions must be guided by sound and rational bases to ensure the sustainability of the investment.

In the uncertainty of today's markets, minimizing risk and maximizing opportunities are crucial.

**How can production studios use metadata and audience perception to minimize business risk moving forward?**

**A**

What are the trends?  
**(Data and Analytics)**

**B**

What will be the hits?  
**(Modelling)**

**C**

What do we need to understand?  
**(Social Listening)**

<b>Titles</b>	Titles in IMDb (not limited to movies)	<b>8,152,534</b> (5.1 million are TV episodes, 800k are shorts)
<b>Names</b>	Persons in IMDb	<b>11,129,900</b>
<b>Ratings</b>	IMDb rating and votes information for titles	<b>11,782,82</b>
<b>Principals</b>	Principal cast/crew for titles	<b>46,122,476</b>
<b>Crew</b>	Director and writer information for all the titles in IMDb	<b>8,152,534</b>
<b>Episode</b>	TV episode information	<b>5,970,241</b>



**584,500**  
movies (all-time)

# The Movies Dataset

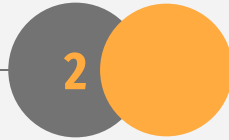


**Titles - Movie  
Data (580k)**



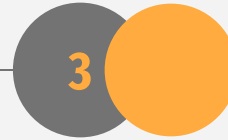
tconst  
Titles  
titleType  
isAdult  
startYear  
runtimeMinutes  
genres

**Merge: Ratings  
Data (266k)**



tconst  
ratings  
numVotes

**Merge: Names  
Data**



nconst  
primaryName  
primaryProfession  
knownForTitles

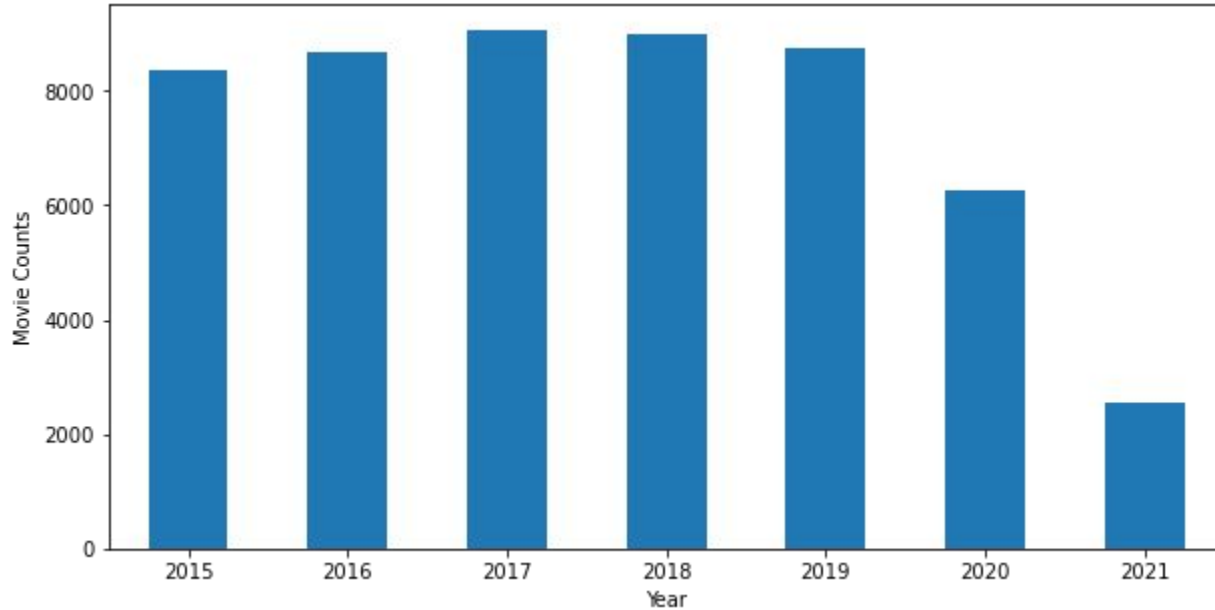
**Merge:  
Principals Data**



tconst  
nconst  
ordering  
category  
job  
characters

# Exploring **IMDB**

IMDB catalogued more than **8,000 new movie entries per year** pre-pandemic.

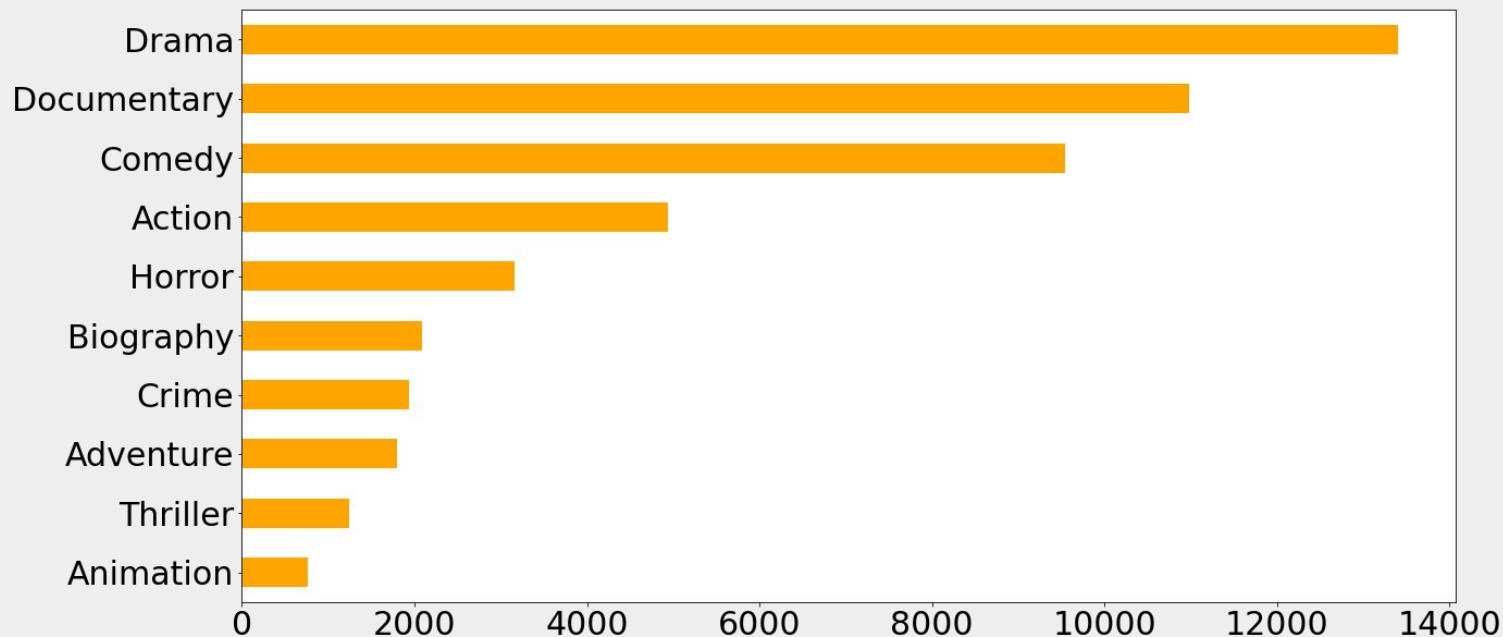


IMDB catalogued **9,065** movies in 2018.

The pandemic's effect on the movie industry is also seen in the **decrease of entries** in 2020 and 2021.

# 26 genres

with **Drama** as the most common genre in the last 5 years.





## Highest Rated Movie in 2018

Average Rating

10.0

1. **The Outsiders**  
(Documentary)
2. **Are We Making a Film?  
Behind the Scenes: Follow  
the Dead**  
(Documentary)
3. **Visconti's Eagles**  
(Documentary)

User Votes

904,079

**Avengers (2018)**  
(Action, Adventure, Sci-Fi)

# Most Popular Actor in 2018

(based on Actor Share)

## Robert Downey Jr.

### Tony Stark, Avengers (2018)

(Action, Adventure, Sci-Fi)

#### What is Actor Share?

Actor Share is **Total Votes divided by Ordering** value:

E.g. Robert Downey Jr. is the first principal cast of the Avengers: Infinity War

Total Votes of Avengers = 904,079, Ordering = 1

Actor Share for Robert Downey Jr: 90

E.g. Michael B. Jordan. is the second principal cast of Black Panther

Total Votes of Black Panther = 664,935, Ordering = 2

Actor Share for Michael B. Jordan: 332,467.5

Actor/Actress	Actor Share	Rank
Robert Downey Jr.	904,079.0	1
Chadwick Boseman	664,935.0	2
Chris Hemsworth	524,142.5	3
Ryan Reynolds	508,645.0	4
Rami Malek	478,611.0	5
Michael B. Jordan	462,886.5	6
Jason Momoa	438,482.0	7
Viggo Mortensen	416,597.0	8
Shameik Moore	414,942.0	9
Tye Sheridan	392,364.0	10

## **Longest Runtime**

(Aggregate - Trilogy)

**558**

minutes

**The Lord of the Rings**

**(2001, 2002, 2003)**

(Adventure, Drama, Fantasy)

## **Longest runtime**

(Single Movie)

**51,420**

minutes

**Logistics (2012)**

(Documentary)

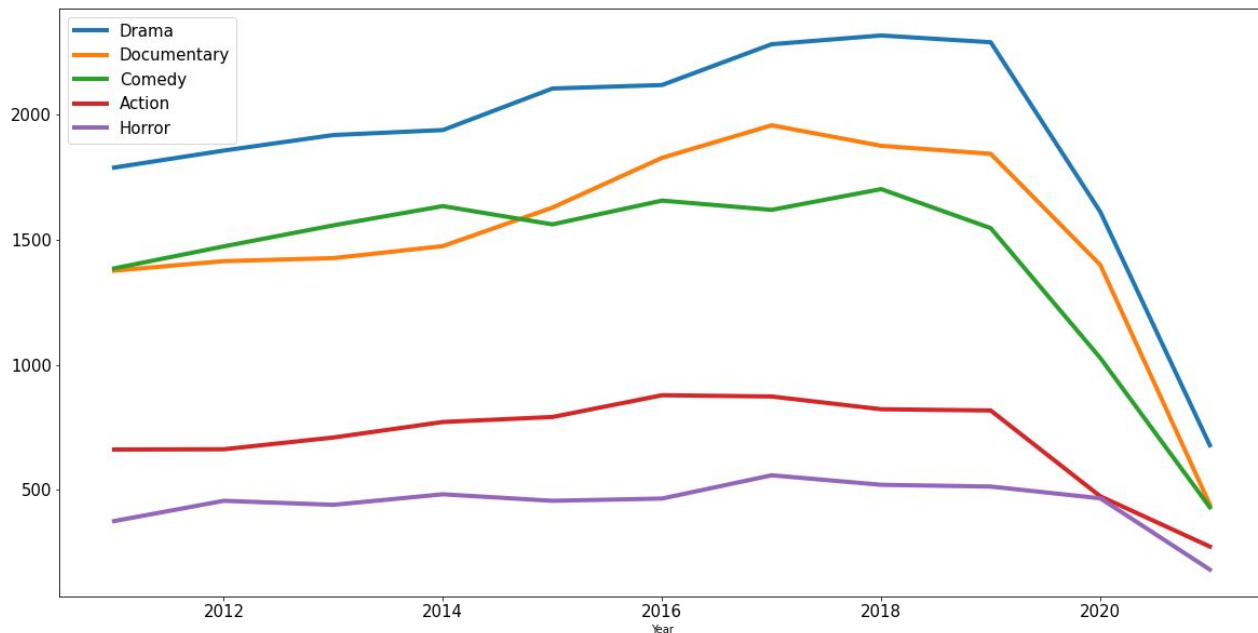
Now let's look at the **sentiment over time**

| **production**

| **critics**

| **audience**

# Production prefers to create **drama**



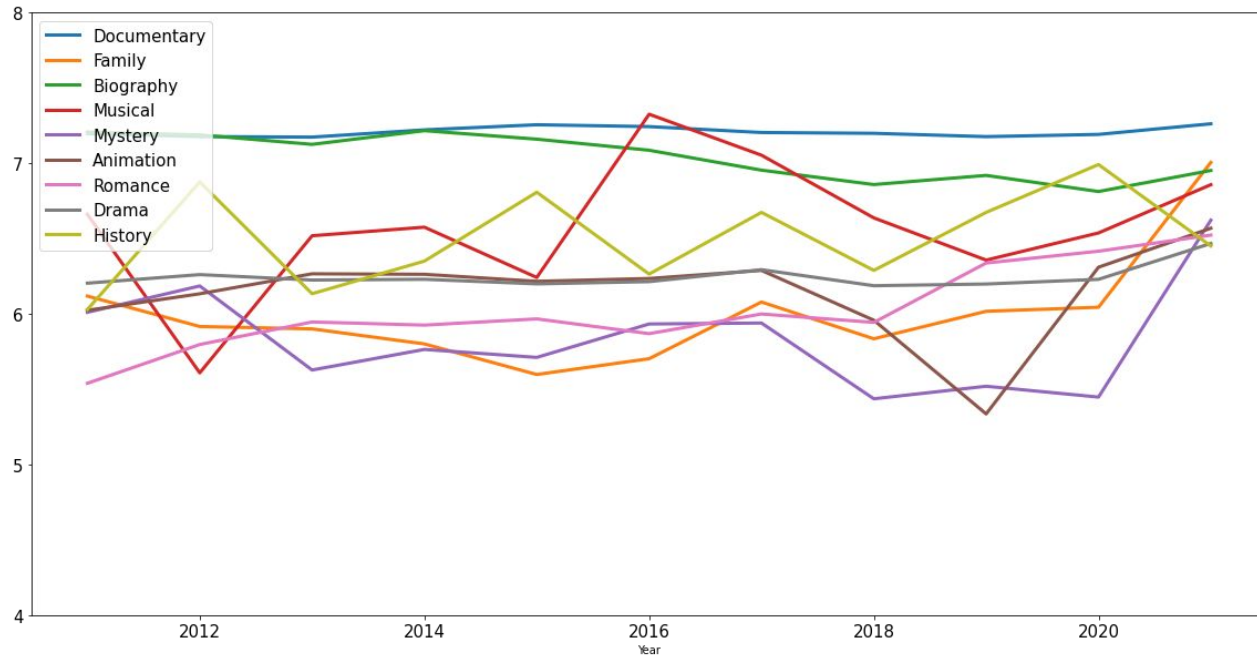
\*first mentioned genre

In the last **10 years**, drama has reigned supreme

Especially **steep increase in around 2013**, with over 500 more drama titles than 2012

Top dramas in 2012 include:  
Locke

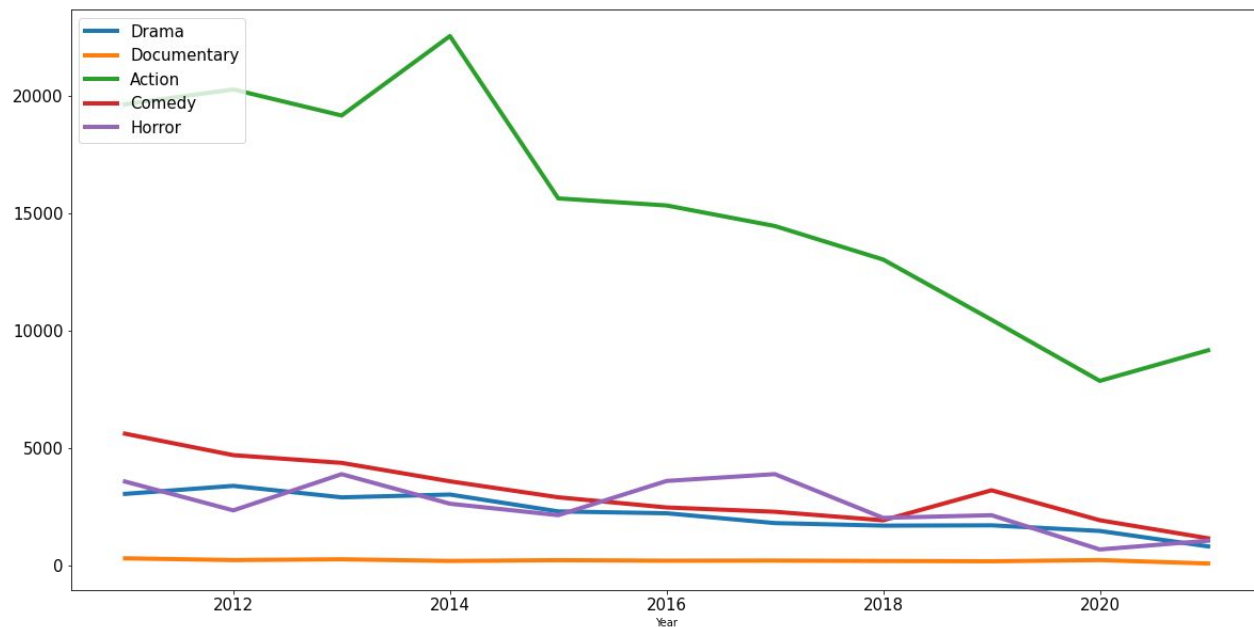
while **critics** hail **documentaries**.



Documentaries receive

However, the top 10 genres do not show much variation in terms of ratings

However, **audiences** crave for **action**.



In the last decade, action movies have been receiving the highest votes in IMDB.

Action movies with the most votes in 2014 include “Guardians of the Galaxy”, “Captain America: The Winter Soldier”, and “X-Men: Days of Future Past”

**What makes a hit?**



# IMDB's numeric indicators

## Number of Votes

- Has a robust **correlation (0.62)** with Box Office Earnings (n=3,000)<sup>1</sup>
- **Highly Skewed** to the left (0.53)
- IMDb publishes weighted vote averages rather than raw data averages. All votes received by users, not all votes have the same impact (or 'weight') on the final rating.

## Average Rating

- Average Rating has a **medium correlation** to (0.37) to Box Office Earnings
- However, **winnings in awards shows** tend to extend the lifespan of movies and help them penetrate more markets - the "Oscar Halo"<sup>2</sup>  
(e.g. Parasite)

1. Supplementary dataset downloaded from: Retrieved from: TMDB Box Office Prediction (2019). Kaggle. Retrieved from: TMDB Box Office Prediction | Kaggle  
2. Deuchert, Eva, Kossi Adjahmah, and Florian Pauly. 2005. "For Oscar Glory or Oscar Money?" Journal of Cultural Economics 29:159–76.

## Clustering on **Number of Votes** and **Average Rating**

(K-Means)

	Cluster 1	Cluster 2	Cluster 3
No. of movies	166,367	99,860	640
Percentage of movies	62%	37%	0.02%
Average Number of Votes	2832	950	528,183
Average Rating	6.95	4.76	7.52

silhouette score at 0.53 (n=3)

## Clustering on **Estimated Number of Votes Per Year** and **Average Rating** (K-Means)

	Cluster 1	Cluster 2	Cluster 3
No. of movies	144,233	94,820	548
Percentage of movies	60%	39%	0.2%
Average Number of Votes Per Year	308.49	123.32	62,025.81
Average Rating	6.98	4.86	7.34

silhouette score at 0.53 (n=3)

## Determining an AAA Title

	1Q	Median	Mean	3Q
Estimated Votes Per Year	38,009	50,289	62,025	72,698
Average Rating	6.80	7.4	7.34	8.0
No. of Identified AAA Films (All-Time)	332	169	114	51
No. of Identified AAA Films (2020)	16	8	6	1

# Proposed AAA Film Metric

(based on 1st Quartile values of clustering)

<b>Estimated Votes Per Year</b>	> 38,009
<b>Average Rating</b>	>= 6.80
<b>No. of Identified AAA Films (All-Time)</b>	332
<b>No. of Identified AAA Films (2020)</b>	16

# The AAA movie

1	Development
2	Pre-Production
3	Production
4	Principal Photography
5	Wrap
6	Post-Production
7	Distribution

# Two paths to an AAA title

## Commissioning a Title

- Begins at Development
- Concepts, script, casting, are all up for decision

## Scaling International Distribution

- Begins after Post-Production
- Titles may have been already screened in special venues
- Non-mainstream film productions

## Features in Real-Life Film Industry

1. “Hit” / “AAA Movie”
2. Box Office Performance
3. Critical Reception
4. Genre
5. Runtime Minutes
6. Sequel or Not
7. Top 100 Actor/Actress
8. Is an Adult Film

## Proxy Features in IMDB Dataset

1. Proposed AAA Metric - Target
2. Vote Count Per Year
3. Average Rating
4. Top 5 Genres
5. Runtime Minutes
6. Sequel or Not
7. Top 100 Actor/Actress
8. Is an Adult Film



# Predicting the Hits

# Predicting the Hits

**Target Variable**

hit\_label

**Accuracy**

**Features**

**False Positive**

# Model Comparison

Model	Logistic Regression	Decision Tree
<b>Features:</b> Runtime, top5genre, averageRating, top10Actors, isAdult, isSequel, top100Actors		
<b>Recall</b>	68%	6%
<b>Accuracy</b>	81%	99%
<b>False Positive Rate</b>	18%	0.2%

# Recommendations

- Presence of high-profile actors
  - Corollary: High budget
- IP-based films (e.g. comic superheroes) do well, and so do their sequels
  - Supports corollary of higher budget - funding is needed to cover acquisition costs of IP and movie rights
- Top 5 Genres: Action, Crime, Mystery, Horror, Adventure
- Consider distribution of non-mainstream movies with high critic regard

**Getting the complete picture**

# What do we need to understand?

Listening to the most important stakeholder: **our audience**

A

Does this type of plot arc keep audiences engaged?

B

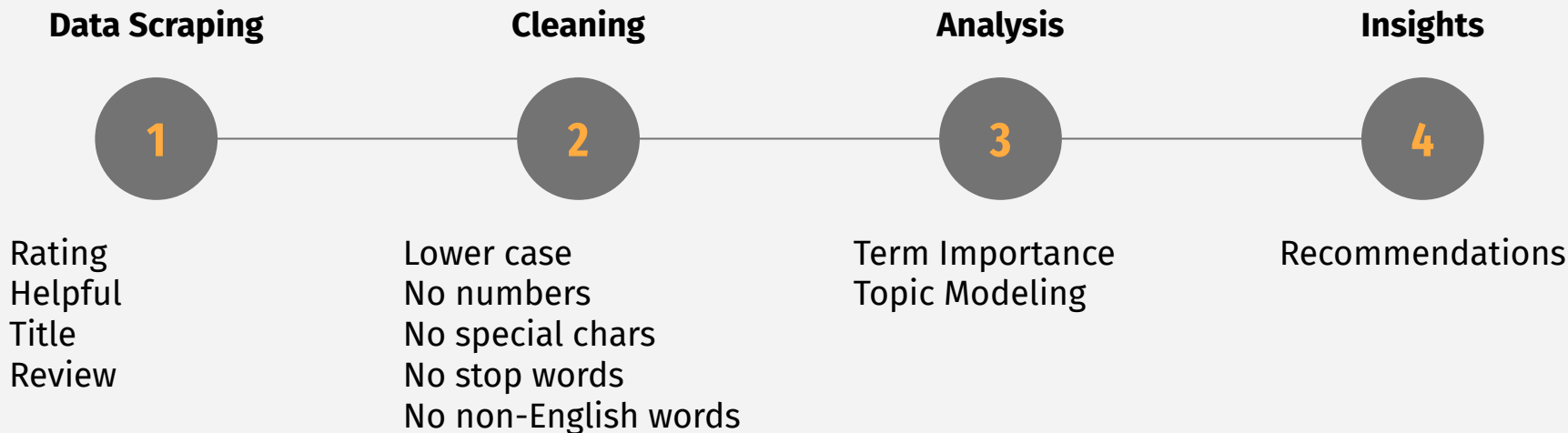
What are the overarching feelings of audiences towards the film?

C

How is character development perceived?



# Movie Reviews Dataset<sup>1</sup>



1. Supplementary dataset downloaded from: Pal, A. et. al (2019). IMDB Movie Reviews Dataset. Retrieved from: <https://ieee-dataport.org/open-access/imdb-movie-reviews-dataset>

# Listening to audiences

Avengers: Endgame has  
**8723 reviews.**



## What did audiences think of Avengers: Endgame?

- Emotion-related words frequently were mentioned "**speechless**", "**emotion**", "**joy**".
- Viewers **tended heavily compare** the present movie to preceding ones ("**infinity\_war**")
  - May indicate a need for **more evaluation when producing new entries** for multi-part franchises

### Rush of Emotion

"**feel**", "**amazing**", "**joy**", "**emotional**"

### Attachment to characters

"**hero**", "**hulk**", "**captain**", "**emotional**"

### Time travel

"**time**", "**time\_travel**", "**past**", "**back**"

The above topics support Marvel's push for **multiple spin-offs for their wildly popular characters**

**Time-travel** remains to be an intriguing plotline for fans

## What topics did they talk about?

Topic modeling through  
Latent Dirichlet  
Allocation (LDA)

# **Recommendations for Action genre based on Avengers' Endgame Reviews**

- Character development is highly valued and may even be of more importance than the portrayal of action, even in action films
- The rush of action films may be heightened by attachments to well-developed characters

# General Recommendations

- Consider **other sources** of audience sentiment and opinion such as discussion forums, reddit, twitter, tumblr, etc
- Evaluate entities in reviews such as character names to identify archetypes of **fan-favorites** and hated characters
- Other helpful features
  - Budget
  - Streaming statistics
- Explore the use of Metaflow library by Netflix

