

Unveiling the Media Tapestry: A Comprehensive Analysis of Movie Coverage

Michael Almanza, Tristan Leclair-Vani,
Jean-Alexandre Nolin, Juan Rodriguez Bracho

McGill University
845 Sherbrooke St,
Montréal, Qc, Canada

Abstract

In the dynamic realm of the film industry, media coverage serves as a powerful force in shaping audience perceptions, influencing box office outcomes, and contributing to the cultural discourse surrounding movies. This paper presents an exploration of recent movie coverage, drawing insights from a dataset sourced from newsapi.org. Our selected movie span genres and budgets, offering a comprehensive view of media coverage.

Introduction

Understanding media dynamics

Movies are not just cultural artifacts; they are phenomena that generate conversations, controversies, and critiques across various media platforms. This study undertakes a meticulous examination of media coverage, shedding light on movies that dominate headlines and those relegated to the sidelines. As we embark on this exploration, we aim to answer these questions: How did media coverage vary across different movies? Which aspect of the movies were discussed in these articles?

Media Coverage Reign: "Killers of the Flower Moon" top the media

Our investigation unravels a hierarchy in media coverage, with certain movies rising to the top of the conversation in the cinematic narrative. Notably, "Killers of the Flower Moon" emerges from the pack, representing a strong 17.6% of all references to movies in our collected data. The success in media of this film, compared to its co-releases, becomes a central theme in our analysis, and variations in coverage and subjects hint at the importance placed upon this film.

The Dataset: A Cross-Section of Cinematic Diversity

Drawing insights from a diverse dataset sourced from newsapi.org, our movie selection spans genres and budgets, offering a comprehensive view of media dynamics. However, it is the disproportionate attention received by "Killers of the Flower Moon" that piques our interest and forms the crux of our exploration.

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Methodological Rigor: Decoding Media Dynamics

Our methodology employs rigorous open coding, a complete typology, and TF-IDF analysis to decode the intricacies of media dynamics. The findings illuminate the multifaceted relationship between movies and media, unraveling the factors that propel certain films to the forefront while others linger in the shadows.

Unveiling Trends: Media Interest in "Killers of the Flower Moon"

As we navigate the media landscape, our study sheds light on specific aspects that captivate media attention. Contrary to expectations, our findings suggest that "Killers of the Flower Moon" receives comparatively less coverage regarding its box office performance while capturing significant attention in domains such as awards and streaming. This paradoxical trend prompts a deeper reflection on the nuanced ways in which media shapes the narrative surrounding movies.

In essence, this paper invites readers to explore the intricate tapestry of media coverage, where narratives are crafted, reputations are forged, and the ebb and flow of coverage collectively shape the cinematic consciousness.

Data

Data Source

We collected a total of 490 valid articles in English from newsapi.org about the following movies:

- Renaissance: a film by Beyonce
- Godzilla Minus One
- Napoleon
- Wish
- The Hunger Games: The Ballad of Songbirds & Snakes
- Thanksgiving
- Trolls Band Together
- The Marvels
- Five Nights at Freddy's
- After Death
- Killers of the Flower Moon

For the movies Wish and Thanksgiving, we decided to add an additional keyword to the query since these movie titles were common words that would have resulted in unrelated articles, given the holiday season in which we conducted our data collection. The final query of keywords to collect the data was the following:

```
1 "Renaissance: a film by beyonce" OR
  "Godzilla Minus One" OR "Napoleon
  " OR "(Wish AND Disney)" OR "The
  Hunger Games: The Ballad of
  Songbirds & Snakes" OR "Trolls
  Band Together" OR "(Thanksgiving
  AND Sony Pictures)" OR "The
  Marvels" OR "Five Nights at
  Freddy's" OR "After Death" OR "
  Killers of the Flower Moon"
```

Data Description

Variables That being said, each article has the following important attributes:

- Source: The source of the article
- Author: The author of the article
- Title: The title of the article
- Description: A short description of the article
- URL: The URL of the article
- Published At: The date the article was published
- Content: The content of the article

Data Size Our query returned a total 4451 articles, but unfortunately, due to free API constraints, we were only able to extract 490 total.

Data Quality and Missing Values

Missing Values An initial exploration of the dataset showed that there was only a single missing articles, which was marked as [Removed] by the API. This article was removed from the dataset.

Data Cleaning There also wasn't much data cleaning to be done, as the API returned a clean dataset. More data cleaning was done post-coding, as we will discuss in the next section.

Data Exploration

A small preliminary exploration of the dataset was done to get a better understanding of the data. The following are some of the insights we found:

Article Sources We found that the dataset contained articles from a total of 95 different sources. The following is a list of the top 10 sources:

- /FILM: 64
- Deadline: 43
- Screen Rant: 40
- IndieWire: 35
- The A.V. Club: 25
- ComingSoon.net: 21

- Just Jared: 18
- Forbes: 16
- Livejournal.com: 15
- Vulture: 11

Titles and Descriptions Titles and Descriptions varied in length, with averages of 76 and 186 characters, respectively. A more extensive examination was done later on, including running TF-IDF on the title and description to find important terms in relation to topics and movies.

Methods

Movie Choice

The movies we analysed had at least one weekend with a top 5 domestic (US+Canada) gross in the last 7 weekends, from the weekend of October 20th up to and including December 1st, according to the-numbers.com. October 20th was the first weekend in wide-release for Killers of the Flower Moon. We excluded re-releases to enable a fair comparison between movies that were released in the same time period. This consistent choice of movie allowed our group to remove our personal bias when choosing movies to analyse. Although this method might not consider movies in limited release or films that did very poorly at the box office, this resulted in a list of movies that were financially relevant and trending in the media, allowing the team to collect more articles. The articles were collected on December 10th.

Open Coding

After collecting the movie articles, we proceeded with two rounds of open coding. We randomly sampled 200 articles from our dataset and assigned each member of our team to manually assign a topic and take note of any movie mentioned to a fourth of that sample. We decided to manually take note of each movie mentioned to reduce the chances of the title appearing in an article without really talking about the movie, as might be case with Napoleon. We annotated the articles based on their titles, descriptions, and the first sentences of their content. After annotating these 200 articles for a first time, we met up together and went over our typology. We reduced our number of topics from 44 to 7, with somewhat clear definitions. For the second round of open coding, we assigned each fourth of the previous samples to another member of the team and reassigned topics to each article. After the second annotation, we met again and talked about the topics that needed a more rigorous definition. These two rounds of open coding provided us with a complete and thorough typology that will be presented in the following section of the report.

After the typology was defined, we randomly assigned two coders to each article that was not yet annotated. Although randomly assigning coders might result in an unbalance in the number of articles assigned to each coder, based on the law of large numbers this should result in a pretty low variance in the number of articles assigned to each coder. After the remaining articles were annotated, each disagreement was further examined and resolved by the coders not assigned to the article in question.

After the typology was defined, we collected the remaining 290 articles and assigned two coders to each article in the following manner: we divided the articles in groups of 50, and assigned a unique pairing of coders to each group. Although this resulted in an imbalance of number of articles assigned to each coder, the difference was small enough to introduce an insignificant amount of bias. Since they were not shuffled, our coder assignment method did introduce some systematic bias, as we are unsure in what orders the API returns the articles. In retrospect, before splitting the remaining articles in group, we should've randomly shuffled the articles to reduce systematic bias. One possible impact of the systematic bias is that some coder might have annotated articles from the specific time period, depending in what order the API returns its query result. Nonetheless, after the remaining articles were annotated, each disagreement was further examined and resolved by the coders not assigned to the article in question.

Results

Typology

The first part of our typology involves coding the title:

Movie

1. **Movie Title:** if the title was mentioned explicitly, we code the title as the title of the movie. It must be clear that the movie is the subject of the article. A negative example would be an article discussing Napoleon the historical figure, unrelated to Ridley Scott's film "Napoleon".
2. **Not Mentioned:** if there is no mention of the title of one of the movies we considered, but the article is relevant to movies in general, then the article falls into this category. A positive edge case would be an article related to movies that mentions Beyoncé, but not Renaissance, the title of her movie and so we can't be sure given the limited information returned by the newsapi.
3. **Not Relevant:** if the article is unrelated to movies, it is tagged with this title. A positive example of this would be an article discussing Napoleon's hat, with no mention of Ridley Scott's movie.

The second part of our typology consisted of the following topics and only applied to articles whose title was a valid movie from our list:

Topics

1. **Box Office:** this includes any article that talks about the performance of the movie, the box office totals, projected opening weekend. It also includes articles that talk about the target audience for the film. It is denoted as Box in our coding and results.
2. **Streaming/home-viewing:** this includes articles that mention the movie and a streaming service or physical media (blu-ray, dvd, etc). In particular, an article discussing the views of the movie on a streaming service or the release of the movie on a streaming service/to physical media is a positive example of Streaming. The sale of a movie to a streaming service or the sales of physical media are also considered to be within the Streaming topic. It is denoted

as Streaming in our coding and results. An edge case that is a negative example is an article that discusses Five Nights at Freddy's opening weekend box office despite the film's day and date release on a streaming service. This is a box office example since although it mentions streaming, the focus is on the film's box office success.

3. **Awards:** this includes articles that mention any aspect of the film has won or could be in contention for an award. Awards ceremonies are included if the film in question is mentioned to have won/contended for an award. It is denoted as Awards in our coding and results. A negative example here is articles that talk about Robert de Niro's speech controversy. If an award related to Killers of the Flower Moon is not explicitly mentioned, this is not an awards example.
4. **Production:** this includes articles that focus on special effects, practical effects, set design, direction, cinematography, props, behind the scenes and casting. Any other aspect of production that is the focus of the article should also be included. It is denoted as Prod in our coding and results.
5. **Plot and Performances:** this includes articles that focus on aspects related to the script of the film. This encapsulates the discussion of scenes and acting performances, as a performance acts out the script. Any review of the film which does not mention specifically an aspect of production is included in this category, as we assume that it will review the plot. It is denoted as Plot in our coding and results. A positive example would be an article discussing a change in a scene of Killers of the Flower Moon. We also found reviews often talked about aspects of the plot or actor's performances, so it lead us to make this category.
6. **Promotion:** this includes articles that are related to the premiere of the film, such as mentions of the red carpet or talking about the movie opening soon. Also included in this topic are interviews with people that worked on the film unless they involved a specific aspect of production or the plot and performances. This includes podcasts and talk shows where people are talking about the movie, but not as a review, as these people are promoting the movie. Another part of this topic are articles which cover toys and merchandise related to the movie. It is denoted as Promo in our coding and results.
7. **Movie only mentioned:** this includes articles which are not talking about a particular aspect of the movie, despite having mentioned the film title. This would include articles that focus on some aspect of the actors/director and only mention the title of the movie as part of their introduction. It is denoted as NR (Not Relevant) in our coding and as "Movie only mentioned" in our graphs. A negative edge case for this topic would be an article discussing the passing of someone who worked on the score of Killers of the Flower Moon. The article's main focus is on the person, however the score is an important part of the article, leading to a classification of production.

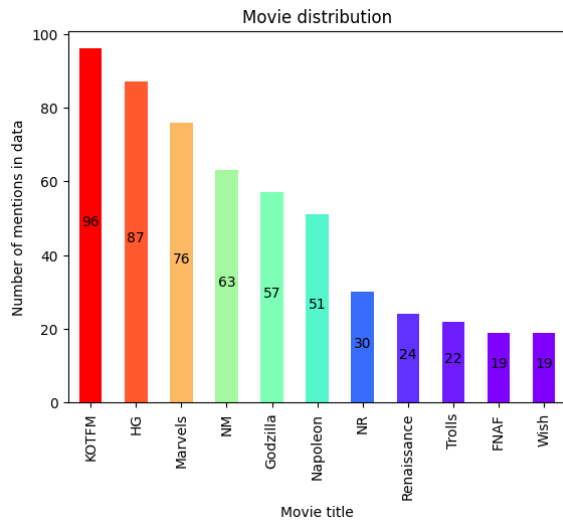


Figure 1: Distribution of the movie mentions in all the articles collected

Movie Coverage

The movie with the most mentions was Killers of the Flower Moon, with 96 mentions, followed by Hunger Games: The Ballad of Songbirds and Snakes (87 mentions), The Marvels (76 mentions), Godzilla Minus One (57 mentions), Napoleon (51 mentions), Renaissance: A Film by Beyoncé (24 mentions), Trolls Band Together (22 mentions), Five Nights at Freddy’s (19 mentions), and Wish (19 mentions). 63 articles were related to movies but the specific movies in question could not be determined from the title and description alone. Lastly, 30 articles were completely unrelated to movies.

Topic Distribution

Regarding the distribution of topics, the most common topic in the articles was "Box Office" (36.4%), closely followed by "Plot and Performances" (26.6%). The rest of the topics in decreasing order were "Promotion" (13.1%), "Streaming/home-viewing" (6.4%), "Awards" (5.8%), and "Production" (4.9%). 6.9% of the articles mentioned a movie but did not fit into any particular category.

Killers of the Flower Moon

Pertaining to the aspects of Killers of the Flower Moon that were covered in the media, the most common topic was "Plot and Performances" (25.0%) followed by "Awards" (22.9%). The following topics were "Streaming" (15.6%), "Promotion" (8.3%), "Production" (6.2%), and "Box Office" (4.2%) being the least covered topic. Out of all the articles that mentioned Killers of the Flower Moon, 17.7% of them only mentioned the movie without falling into any of the categories.

Top 10 TF-IDF Words

Below are the top 10 words by TF-IDF for each topic of our typology. For each article, the title and description were used

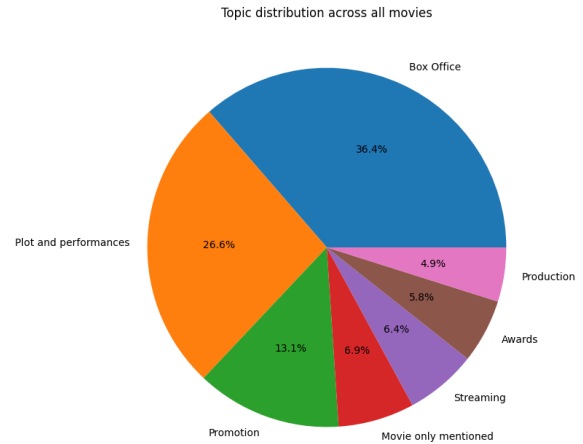


Figure 2: Topic distribution over all articles collected

as the text of the document.

1. Box Office: "box", "office", "weekend", "million", "hunger", "games", "marvels", "napoleon", "wish", "ballad"
2. Streaming: "watch", "streaming", "digital", "apple", "available", "home", "moon", "killers", "flower", "december"
3. Awards: "award", "awards", "flower", "killers", "moon", "critics", "list", "oppenheimer", "board", "scorsese"
4. Production: "music", "director", "band", "moon", "killers", "flower", "trolls", "design", "hunger", "games"
5. Plot and Performances: "godzilla", "marvels", "minus", "review", "movie", "scene", "marvel", "napoleon", "flower", "moon"
6. Promotion: "premiere", "hunger", "games", "ballad", "songbirds", "snakes", "zegler", "beyonce", "hunter", "godzilla"
7. Movie only Mentioned: "de", "speech", "robert", "niro", "gotham", "robbie", "trump", "leonardo", "dicaprio", "awards"

Discussion

Relative Coverage by the media of Killers of the Flower Moon

Killers of the Flower Moon represented about 17.6% of all the movie articles we collected and is ahead of the next most popular movie by 9 mentions. From Figure 1, we can observe that the frequencies of mentions by movie decrease relatively linearly (by 6-13 mentions) until a heavy drop between Napoleon and Renaissance (almost 20 mentions), at which point the frequencies decrease very lightly. This suggests there is a remarkable difference between the top

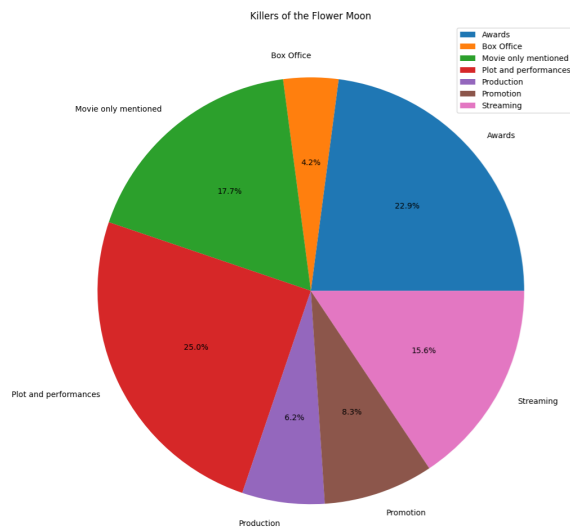


Figure 3: Topics of the articles where Killers of the Flower Moon is mentioned

6 and the bottom 4 movies when it comes to media coverage. Moreover, the linear decrease in media coverage from the first to the sixth movie indicates a strong hierarchy relating to which movies are more trendy (with Killers of the Flower Moon at the top), as opposed to a flat distribution where movies get about the same amount of coverage.

Aspects of Killers of the Flower Moon covered by the media

The media seems to not be interested at all about Killers of the Flower Moon's box office performance: compared to the total topic distribution, less articles talked about Killers of the Flower Moon's box office performance than the average, at 36.4% against 4.2%. This is surprising as Killers of the Flower currently sits at the 35th spot for top grossing movie for the year just behind Trolls Band Together and The Marvels, respectively at 28th and 29th, per the-numbers.com. On the other hand, more articles covered the movie's accolades it received. Compared to the average of 5.8% for the Awards topic, Killers of the Flower Moon racked up a staggering 22.9% of the articles mentioning some kind of awards. This is evident from the fact that the production has won 17 awards so far, with many more for which the movie is currently nominated. Furthermore, people seem to want to know when the movie will be available on streaming, as it also has a big difference with the average (6.4% vs 15.6%).

Robert de Niro's speech While accepting an award, Killers of the Moon actor Robert de Niro made a speech about his disagreements with former president Donald Trump. Some media coverage that mentioned the movie only did so while introducing the actor, before explaining the event. This explains why the "Movie only mentioned" topic is also overrepresented as those articles were assigned that topic. Compared to the average of 6.9% for that topic,

Killers of the Flower Moon racked up 17.7%.

Under representation of some movies

One thing to note about our data is that we failed to collect even a single article related to the movies "After Death" and "Thanksgiving", which may have created some overrepresentation of our other movies. This lack of articles for these two movies may be due to issues with the keywords we used. As mentioned previously, "Thanksgiving" had to be paired with an additional keyword, which may have limited the collected articles too much. As for "After Death", this movie grossed considerably less at the box office than the other movies, and it was not shown in theaters in December, so it may simply not have been a relevant enough movie to have substantial media coverage at the time of our data collection.

TF-IDF

Term frequency - inverse document frequency (TF-IDF) puts a smaller emphasis on words that appear frequently throughout the entire set of articles. This is because the idf term is inversely proportional to the number of articles the word appears in, with a score of 0 if it appears in all articles. In this way, it provides us with the terms which are most relevant to each topic.

For Box Office, we get terms which we would expect, but give us no insights into movies like "box", "office", "week-end" and "million". However, these do serve to confirm that TF-IDF is working as expected. The movie titles that appear are Hunger Games, Marvels, Napoleon, and Wish in that order, which indicates that the box office was a large portion of the news cycle for these films. The absence of Killers of the Flower Moon indicates that relative to other movies released in the same time period, its box office is not covered as much.

For the topics of Streaming and Awards, Killers of the Flower Moon is the only film from our list that appears in the TF-IDF top 10. This shows there is a strong interest in being able to watch Killers of the Flower Moon from home and shows that the film has resonated with critics, having more articles focusing on awards the film has won or could win relative to other movies.

In terms of Production, Killers of the flower moon also has high relative coverage as it appears tied for first with Trolls Band Together. This is a somewhat peculiar result as "band" appears before "moon", "killers", and "flower", which is then followed by "trolls". Upon further inspection, it seems that "band" being such a popular term for Production was most likely due to the band NSYNC being closely associated to the movie through their connection to Justin Timberlake, who worked on multiple scores and soundtracks for the franchise over the years, starting in 2016 with the release of the song "Can't Stop the Feeling!" along with the first installation of the series, Trolls.

In the Plot and Performances topic, Killers of the Flower Moon falls behind Godzilla Minus One, The Marvels, and Napoleon. This suggests that there was relatively more reviews and articles breaking down specific scenes for these movies than Killers of the Flower Moon. This corresponds

with what was seen in the data annotation phase, as there were many articles that focused on end credit scenes for The Marvels and Godzilla.

Killers of the Flower Moon does not appear at all the TF-IDF top 10 of the Promotion category. 5 of these entries belong to the Hunger Games: the Ballad of Songbirds and Snakes, owing to its lengthy title. Additionally, "zegler" and "hunter" are parts of names of the actors from the Hunger Games movie. This makes it clear that Hunger Games had a quite strong relative coverage of promotion compared to other films. Beyoncé and Godzilla also crack the top 10, reinforcing that Promotion was not a strong topic for Killers of the Flower Moon.

Finally, the Movie only Mentioned topic is dominated by words related to Robert de Niro's awards speech at the Gotham awards, where his comments about Trump were censored. There were several articles with this as their subject, and we felt it was not relevant to the film for which he won the award, Killers of the Flower Moon. In most of the articles, it was not mentioned what award was given or if the film had received an award, so it did not meet our typology's criteria for falling under the Awards topic. The articles also appeared to focus on the censoring of de Niro as opposed to anything related to the movie. tF-IDF seems to underscore this definition as no word from the title of the film was among the top 10 for the topic. There were also a few articles about Leonardo DiCaprio that were unrelated to Killers of the Flower Moon which were coded as Movie only Mentioned, explaining his appearance in the top 10 TF-IDF.

Overall, TF-IDF suggests the stronger topics of relative coverage for Killers of the Flower Moon were Streaming, Awards, and Production. In figure 3, streaming and awards make up a larger percentage than the average topic distribution in figure 2. Promotion and Box Office were the weakest for the film, which aligns with figure 3, where they represent a small proportion of the topics related to Killers of the Flower Moon. So, TF-IDF is supported by the results found using proportions.

Group Member contribution

- Michael Almanza: Code for TF-IDF, results and discussion for TF-IDF, typology section of report, process and retrieval for choosing which movies to include.
- Tristan Leclair-Vani: All other scripts for gathering, processing and analysing the data, including data exploration notebooks and creation of graphs. Also tested the code.
- Jean-Alexandre Nolin: Methods, Discussions for the topics focused on by the media, Formatting, Introduction, Bug finder of tested code.
- Juan Rodriguez Bracho: Results, Discussions
- Together: coding, open coding, typology refining
- Github Repository