Beyond the Screen: Analyzing the Unique Success of the Barbie Movie in 2023

Marie-Elise Latorre[1], Yianni Zavaliagkos[1],
Taha Rhaouti[1], Stephanie Fontaine[1],

[1]McGill University
marie-elise.latorre@mail.mcgill.ca, yianni.zavaliagkos@mail.mcgill.ca, ahmed.rhaouti@mail.mcgill.ca,
stephanie.fontaine2@mail.mcgill.ca

Abstract

This project examines the media coverage of movies released in Summer 2023, with a focus on *Barbie's* exceptional performance and visibility. In our study, we utilized over 500 news articles, performed TF-IDF and sentiment analysis, and reviewed daily article releases. The study explores the correlation between a movie's thematic elements, such as fashion and cultural commentary, and its gross income. *Barbie's* success is attributed to its cultural icon status, trendsetting in fashion, and engagement in social media dynamics. Comparative analysis of other contemporary films, including *Oppenheimer* and *Mission Impossible*, highlights the unique factors contributing to *Barbie's* media dominance, offering insights into the impact of branding, narrative appeal, and societal trends on a movie's market success.

Introduction

In 2023, the cinematic landscape witnessed a remarkable phenomenon with the release of the Barbie movie, a film that not only revitalized a beloved global icon but also challenged societal norms. Directed by Greta Gerwig, *Barbie's* transition from an iconic toy to a big-screen sensation represents a significant cultural milestone. This movie was controversial from the beginning, given the doll's polarizing stature in the pop-cultural imagination - an emblem of both feminist empowerment and hyper-corporate materialism (Taste of Cinema, 2023).

Despite this reputation, *Barbie* transcended its commercial origins, becoming a subversive commentary on gender norms, capitalism, and self-worth. The film's clever script navigated complex themes with humor and insight, making it a box-office success and a critical piece. Furthermore, the movie's depiction of Barbieland, a utopian society led by *Barbies*, and its exploration of real-world societal dynamics, highlighted the intricate relationship between gender roles and societal expectations (New Statesman, 2023).

The release of the *Barbie* movie in theaters on July 21, 2023, marked a significant moment in the brand's history, offering a modern reinterpretation of *Barbie's* legacy. This report presents a comprehensive analysis of the media coverage of the *Barbie* movie. Our key findings reveal that the movie achieved substantial commercial success. The media narrative surrounding the *Barbie* movie demonstrates the importance of representation, brand identity, and overall positive sentiment are correlated with higher gross margin profit.

Furthermore, our comparative analysis with other films that were released at the same time highlights how movies can benefit from each other as well as a correlation between public reception and overall sentiment of media coverage. This can further be seen in the keywords appearing in TF-IDF analysis such as "Barbenheimer" relating *Oppenheimer* and *Barbie* to each other.

This report aims to provide insights into the ongoing legacy of *Barbie*, its media coverage, and overall sentiment related to the gross margin profit of Hollywood productions in 2023.

Data

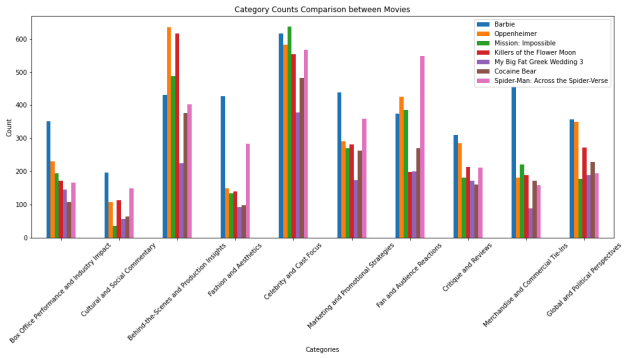
Dataset Composition and Sampling Methodology

We examined movies released in 2023, querying the GNews API (Abdullah, 2023) for articles published one month before and one month after each movie's release. Our queries used specific keywords: the movie title, the main starring actor, and the director. After querying, we filtered for unique articles to eliminate duplicates. We also filtered for articles in English, ensuring language consistency, and performed content quality checks to exclude brief or less credible sources. Additionally, we implemented a relevance scoring system based on keyword density and prioritized diverse sources to avoid source bias. This approach allowed us to capture the natural media attention each film received, providing an authentic reflection of their real-world impact and popularity in the media landscape.

Data Preparation

After querying and filtering the articles, we categorized them by extracting the full text of the article. We conducted an open coding of ~200 articles to create our ten categories. We then expanded our coding to the full dataset, allowing articles to belong to multiple categories if applicable, based on the keywords found within the text. This method provided a structured yet flexible approach to understand the thematic distribution of our dataset.

Figure 1: Bar chart showing a visual representation of the numbers presented in **Table 1**



Methods and Results

Our goal was to look for the representation of the Barbie movie in the media when compared to other movies. We used many techniques including distributional analytics, TF-IDF, and sentiment analysis amongst other data science techniques to answer our question.

Categorical Distribution

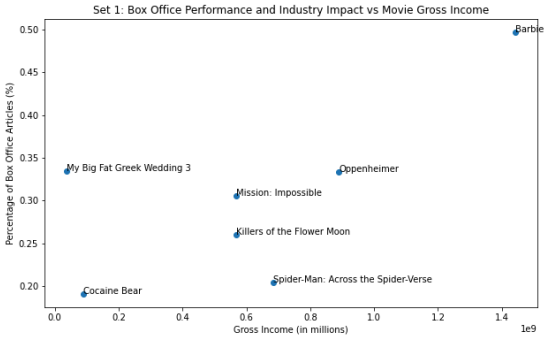
First, we explored the distributions of each category relative to the gross income for each movie. To normalize our results, we compared gross income to the proportional number of articles instead of the raw number to account for the different numbers of total articles as can be seen in Table 1 (or visually in Figure 1).

Table 1: Table representing the proportion of articles for each category by movie.

Movie Title	Total # Articles	Box Office Performance and Industry Impact	Cultural and Social Commentary	Behind-the-Scenes and Production Insights	Fashion and Aesthetics	Celebrity and Cast Focus	Marketing and Promotional Strategies	Fan and Audience Reactions	Critique and Reviews	Merchandise and Commercial Tie-Ins	Global and Political Perspectives
Barbie	714	0.50	0.33	0.60	0.65	0.84	0.65	0.53	0.46	0.61	0.49
Oppenheimer	698	0.31	0.15	0.85	0.27	0.80	0.38	0.60	0.41	0.27	0.54
Mission: Impossible	667	0.29	0.06	0.70	0.18	0.94	0.39	0.59	0.26	0.33	0.24
Killers of the Flower Moon	690	0.23	0.17	0.85	0.19	0.82	0.39	0.26	0.31	0.25	0.37
My Big Fat Greek Wedding 3	464	0.30	0.09	0.38	0.16	0.79	0.32	0.42	0.35	0.22	0.38
Cocaine Bear	621	0.19	0.11	0.63	0.15	0.77	0.46	0.41	0.23	0.29	0.36
Spider-Man: Across the Spider-Verse	874	0.18	0.15	0.44	0.32	0.67	0.47	0.60	0.22	0.19	0.26

After plotting our results in Figure 2, we found that the number of articles talking about the box office performance strongly correlates with the gross income of the movie.

Figure 2: Scatter plot representing gross income vs the proportion of 'Box Office Performance and Industry Impact' articles

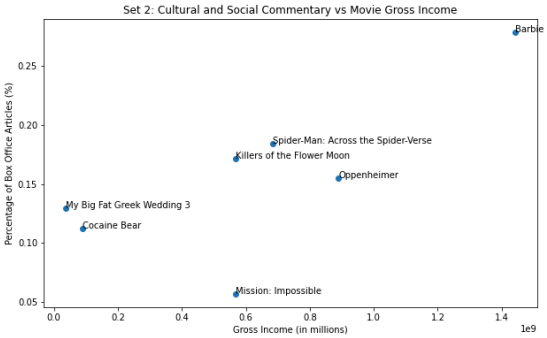


In Figure 2, we manually looked into the articles of the high-box-office-article movies and noticed that *Barbie* garnered a lot of media attention about it being the highest grossing movie of the year. We believe the fact that was the most financially successful is the reason it has significantly more box office attention than a similar movie like *Oppenheimer*. Conversely, *My Big Fat Greek Wedding 3 (MBFGW)* attracted the media for the opposite reason. The articles tended to focus more on the reasons behind the movie's underperformance. This intense media scrutiny at both ends of the spectrum underscores how box office performance, whether extraordinarily good or bad, remains a focal point of interest and discussion in the film industry.

Another notable point on the graph is the movie *Cocaine Bear (CB)*. *CB* garnered similar financial success to *MBFGW*. We believe this relationship has to do with expectations versus the results of a movie. *CB* is a slapstick comedy of an exaggerated true story about a bear who found a lost package of cocaine and went on a killing spree, while *MBFGW* is a three-quel of a popular movie from the early 2000s. For the box office articles about *CB*, we found that authors were not very surprised with the results of gross income, hence, garnering less media attention.

In Figure 3, we explored a second category, Cultural and Social Commentary. We saw more influential movies made more money than less influential movies.

Figure 3: A graph showing gross income vs proportion of articles in the cultural and social commentary categories



Films such as *Spider-Man: Into the Spider-Verse,* which incorporate representation, have the potential to empower viewers by offering diverse and relatable characters. Positive coverage discussing representation can influence audiences to engage with a movie. When movies authentically portray diverse cultures, perspectives, or societal issues, they resonate strongly with audiences who see themselves represented on screen (Shabo, 2023). This resonance often translates into increased audience engagement and a broader demographic appeal, consequently contributing to a film's success and higher gross income. However, *Mission Impossible 7* emerged as an outlier in Figure 3. This is probably because it is a popular action movie where the emphasis is not on cultural representations. Rather, it is based on the immersive action sequences and stunts performed by Tom Cruise.

Release Date Histograms

By examining when a movie is first shown in theaters, we can gain a better understanding of the movie's cultural relevance and market strategy as well as uncover insights into pre-release anticipation, on-release attention, and post-release engagement.

Figure 4: Histograms representing the distribution of articles released for *Barbie* and *My Big Fat Greek Wedding*. The graphs are centered on the release date of the movie.

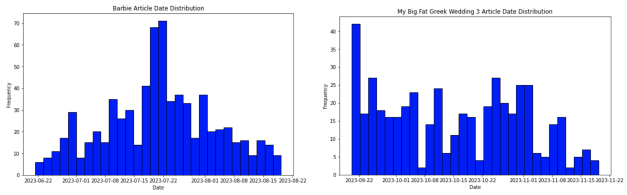
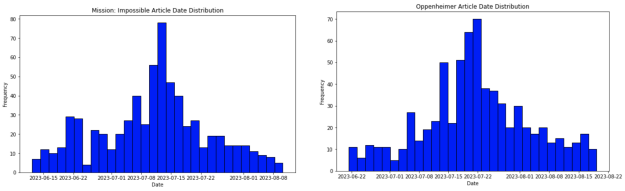


Figure 4 highlights the impact of release dates on media coverage for *Barbie* on the left and *MBFGW* on the right. The high article per day frequency before the release of *Barbie* indicates an increasing amount of anticipation leading up to the launch. This was followed by a spike in attention on the release date. After that, the movie had high audience retention, encouraging the media to continue its coverage post-release.

On the other hand, *MBFGW* has a left-skewed histogram. It depicts a strong pre-release interest, likely due to effective marketing and having a pre-existing audience. However, the post-release decline indicates a challenge in maintaining audience engagement. Furthermore, we manually looked at some of these articles after the release date and saw a generally negative sentiment toward the movie. This may suggest a decreased interest in the movie across the board.

Figure 5: Histograms representing the distribution of articles released for *Mission Impossible 7 (left) and Oppenheimer* (right). The graphs are centered on the release date of the movie.



The histograms for *Mission Impossible 7 and Oppenheimer* have a similar shape to *Barbie*. This is likely because they were released around the same time of year and they had large marketing budgets.

One notable aspect of the *Mission Impossible 7* graph is that there was a dip in media coverage around the time of the *Oppenheimer* and *Barbie* release. This is explained in the gross income of the movies. *Barbie* and *Oppenheimer* were more successful, causing the media's eyes to shift away from *Mission Impossible 7*. In addition, the unimodal and symmetric histograms for *Barbie*, *Oppenheimer*, and *Mission Impossible 7*, all released in the summer, suggest that summer months are advantageous for movie releases, likely due to higher audience availability and interest. In contrast, *My Big Fat Greek Wedding*'s fall release shows a decline in post-release interest, possibly due to the resumption of regular activities and decreased leisure time among audiences. Despite not performing as well as *Barbie* and *Oppenheimer*, *Mission Impossible* still outperformed *My Big Fat Greek Wedding*, reinforcing that summer releases may have a strategic advantage.

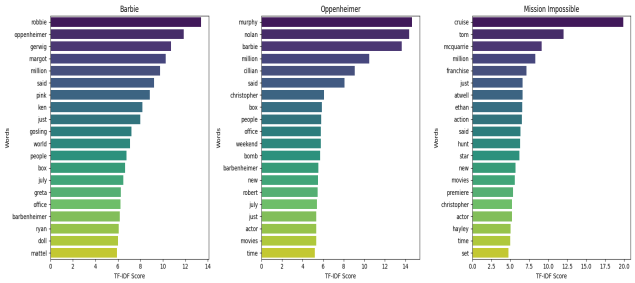
TF-IDF Text Analysis

Utilizing the Scikit-learn Python library (scikit-learn, 2019), we calculated the Term Frequency - Inverse Document Frequency (TF-IDF) for each article. TF-IDF is a metric that increases in value proportionally to the number of times a word appears in a certain document and

decreases proportionally to the log frequency of that word in the corpus. TF-IDF effectively balances the weighting, accounting for the fact that certain words are inherently more common than others.

We started by grouping our articles by movie. For this section, we decided to look specifically at *Barbie*, *Mission Impossible*, and *Oppenheimer* since these movies were all released within the same month and were compared extensively to each other at the time.

Figure 6: Bar charts representing TF-IDF scores for *Barbie*, *Mission Impossible*, and *Oppenheimer*.



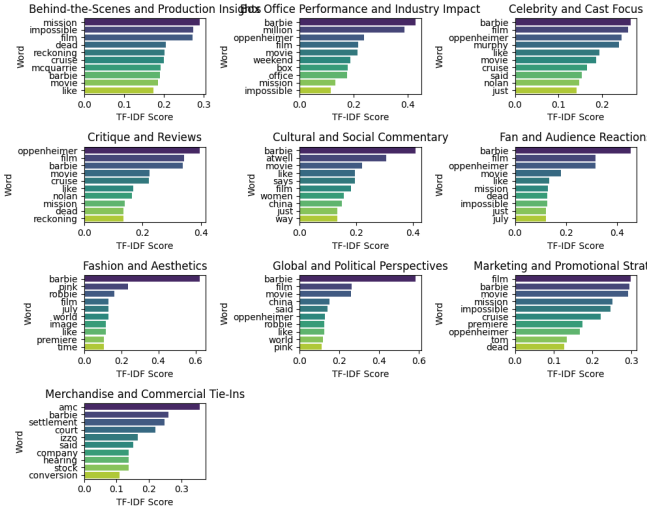
For the results in Figure 6, we made sure to exclude words in each movie title because their TF score is very high, but their IDF is underrepresented. For example, articles about *Barbie* will not include the word "mission" and thus the word will be pushed to the top of the TF-IDF score for *Mission Impossible*. The same pattern can be seen for the actors of each respective movie.

Both *Barbie* and *Oppenheimer* have a high TF-IDF score for the opposite movie title. This not only indicates a correlation between their online presence and social media reactions but also a direct linking of the two in pop culture. The connection of the two sparked the creation of a new term "Barbenheimer," also seen in both charts. This could have sparked curiosity or amusement among social media users, fostering discussions that indirectly promoted both movies and possibly influenced their box office performance through increased online buzz and intrigue.

In addition to the direct linkage between *Barbie* and *Oppenheimer* in media discussions, the TF-IDF results suggest a broader thematic and cultural interplay between these movies. The high scores for keywords related to the main actors in *Barbie* indicate a media focus on celebrity influence, possibly contributing to the film's widespread appeal. In contrast, *Mission Impossible 7* and *Oppenheimer* seem to be more focused on their respective themes and narrative elements, as indicated by the prevalence of specific keywords related to their titles and content. This distinction might reflect different audience engagement strategies, where *Barbie* leverages star power, while the others focus on story and cinematic experience.

Similarly, we explored TF-IDF scores when grouping by category. We expect that the movie titles here will be the most relevant words because they have a high TF, but only appear in the documents about that movie i.e. low IDF. This will give us a clear visualization of which movie leads each category.

Figure 7: Bar chart representing the highest scoring TF-IDF words for each category.



In our analysis, we highlight a couple of these graphs.

Behind-the-Scenes and Production Insights: This was dominated by *Mission Impossible*, which had the biggest production budget, a budget of $290 million, out of all the movies (Donnelly, Lang, & Donnelly, 2022)

Box Office Performance and Industry Impact: This category matched our expectations from our previous box office analysis. The movies came in order of gross income, with *Barbie* being greater than *Oppenheimer*, and *Oppenheimer* being greater than *Mission Impossible*.

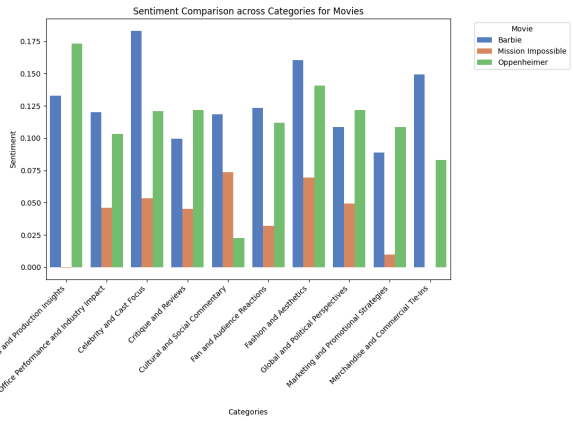
Fashion and Aesthetics: This was led heavily by *Barbie*. Not only was 'barbie' the top word in Figure 7, but the other movie titles weren't even mentioned in this category.

Global and Political Perspective: This category was overrun by *Barbie*. The movie's incredible outreach is seen in its gross income.

Sentiment Analysis

Sentiment analysis was performed using the TextBlob library (TextBlob, 2018), which facilitated the assessment of sentiment scores for each article. We looked specifically at the polarity attribute, a 0-to-1 score reflecting the sentiment's strength and direction. This allowed us to gain deeper insights into the underlying emotions and opinions expressed within the articles.

Figure 8: Bar chart representing the polarity for each category for *Barbie*, *Mission Impossible*, and *Oppenheimer*.



By comparing *Oppenheimer* and *Barbie* across various dimensions in Figure 8, we can see that they display relatively similar performance levels. *Barbie* notably leads 'merchandise and commercial ties' likely due to its association with the doll franchise. *Oppenheimer* excels in sentiment related to 'behind-the-scenes production', attributed to Christopher Nolan's renowned directorial focus on innovative and captivating production elements. However, *Mission Impossible 7* registers notably lower sentiment across the board, suggesting a weaker reception possibly tied to its perceived quality as a less favorable movie experience.

Discussion

The data from our analysis paints a nuanced picture of the media landscape for movies released in Summer 2023. Particularly, Barbie's dominance in categories such as fashion and aesthetics and the emergence of the term "Barbenheimer" underscore its unique cultural and market impact. The TF-IDF analysis in our study significantly contributes to understanding media narratives. For instance, the high TF-IDF scores for 'Barbie' and 'Oppenheimer' in each other's contexts suggest an intertwined media perception, influencing audience interest and discussion. This is exemplified by the creation of the term "Barbenheimer," indicating a cultural crossover and potentially boosting both movies' visibility. The dominance of 'Barbie' in categories like Fashion and Aesthetics aligns with its unique brand identity, resonating with audiences and setting it apart from other films. These insights reveal the power of thematic elements and branding in shaping media coverage and audience perception.

Moreover, our analysis underscores the significant impact of casting choices on media attention and audience reception. The high TF-IDF scores for both main actors in *Barbie* compared to *Mission Impossible 7* and *Oppenheimer* suggest that star power plays a crucial role in driving media narratives and, by extension, audience interest. This trend highlights the strategic importance of casting in film marketing and public perception, where well-chosen actors not only contribute to the film's thematic depth but also enhance its appeal and visibility in the crowded media landscape. Such insights from our data further demonstrate the multifaceted influences on a film's success, extending beyond mere content to include casting and branding.

The comparison with other films like *Mission Impossible*, which scored lower in sentiment analysis, suggests a potential shift in audience expectations, favoring thematic depth over traditional action elements.

Additionally, the data suggests that movies released in the summer performed better than those released in the fall, indicating a possible seasonal influence on audience interest and box office performance. This seasonal trend could be due to a combination of factors, including higher audience availability during summer holidays and the strategic release of high-profile films during this period. This pattern of performance highlights the importance of timing and seasonality in movie releases and their subsequent media coverage and audience reception.

Although our study did not delve into a direct comparison of independent films and Hollywood blockbusters, it offers an intriguing perspective on their contrasting marketing methodologies. This aspect, while not the focus of our analysis, provides valuable context in understanding the broader dynamics of movie marketing and its impact on media representation. The contrast in marketing methodologies between independent and Hollywood films could wield a substantial influence on their portrayal within news articles, thereby potentially molding the outcomes observed in our analysis. For instance, independent films often adopt alternative marketing strategies, focusing on minimal budgets (e.g., around 5000 USD) and emphasizing word-of-mouth promotion and festival screenings rather than extensive marketing campaigns (Aronoff, 2022, Carter, 2019). This discrepancy extends across multiple facets: the focus of coverage might diverge, with Hollywood blockbusters often spotlighted for their star-studded casts and cutting-edge technology promoted through extensive marketing, while independent films may be lauded for their creative storytelling or grassroots support due to their reliance on minimal budgets and word-of-mouth promotion. Such differing approaches could engender varying tones and perspectives within articles, potentially emphasizing commercial success or failure tied to marketing investments for Hollywood productions, whereas independent films might be discussed in terms of their artistic merit, social impact, or innovative methods. Moreover, the visibility and exposure granted to these films might differ significantly, with Hollywood movies

often securing more prominent media coverage, thus influencing public perception. Critically, these contrasting marketing strategies may inadvertently impact how reviewers evaluate these films, potentially introducing biases that could influence the trends and conclusions drawn from our analytical findings.

Limitations

Despite its comprehensive approach, this study has inherent limitations and potential biases. Firstly, by using a list of keywords to do both data collection and sampling, we introduce a selection bias towards those keywords, which could potentially overlook nuanced discussions not directly tied to these terms. This also could risk confirmation bias, as the data may predominantly reflect views that align with the anticipated narrative of the films. For example, *Barbie* had the most articles in the 'Fashion & Aesthetics' category due to the movie being the only movie we looked at to do with fashion/aesthetics. Additionally, we have a risk of a bias towards publications. Since our data relies on news articles, we might not be capturing the full extent of minority opinions. At the same time, it is important to note that the results may have been different if our sample was composed of smaller production and independent movies. These biases, while acknowledged, indicate a need for understanding the limitations of our results.

Conclusion

This study provided a multifaceted examination of the media coverage for movies released in the summer of 2023, with a special focus on the *Barbie* movie. Through the analysis of over 500 news articles, we gained insights into the interplay between media representation and a movie's commercial success, as well as the broader cultural impact of these films.

Our findings indicate that *Barbie's* success can largely be attributed to its status as a cultural icon, its influence on fashion trends, its perfectly timed release date, and its active engagement in the social media landscape. The comparative analysis with other contemporary films such as *Oppenheimer* and *Mission Impossible 7* shed light on the unique factors that contribute to a movie's media dominance and market success. These factors include branding, narrative appeal, and societal trends.

In conclusion, the study underscores the significant role media coverage plays in shaping a movie's success and cultural footprint. It reveals the complex relationship between a movie's thematic elements, such as fashion and social commentary, and its gross income. The insights gathered from this analysis not only contribute to a deeper understanding of the factors influencing the success of major film releases but also point to areas for future research, particularly in the realm of independent cinema and the evolving role of social media in the film industry.

Group Member Contributions

Yianni Zavaliagkos: Data collector, creator of visual representations, and report writing and editing.

Marie-Elise Latorre: Main data collector, report writer, and editor

Taha Rhaouti: TF-IDF analytics, visualizations, and report editing.

Stephanie Fontaine: report writer, and editor.

References

Abdullah, M. (2023, December 11). ranahaani/GNews. Retrieved December 12, 2023, from GitHub website: https://github.com/ranahaani/GNews

Aronoff, A. (2022, April 4). How to: Market Your Independent Film. Retrieved from NYFA website: https://www.nyfa.org/blog/how-to-market-your-independent-film/

Carter, R. (2019, November 26). How much should indie films spend on marketing? Retrieved December 12, 2023, from robertbrucecarter.com website: https://robertbrucecarter.com/writing/2019/11/how-much-should-indie-films-spend-on-marketing/

Donnelly, B. L., Lang, B., & Donnelly, M. (2022, February 8). "Mission: Impossible 7": How COVID-19 Blew Up the Budget of Tom Cruise's Spy Sequel. Retrieved from Variety website: https://variety.com/2022/film/news/mission-impossible-7-budget-tom-cruise-1235173816/

New Statesman. (2023). A strategic analysis of the Barbie movie.Available: http://www.newstatesman.com/culture/2023/a-strategic-analysis-of-the-barbie-movie

Norton, K.I., Olds, T.S., Olive, S. et al. Ken and Barbie at life size. Sex Roles 34, 287–294 (1996). https://doi.org/10.1007/BF01544300. Accessed: 2023-11-28

Pantone. (2023). https://www.pantone.com/color-of-the-year/2023. Retrieved from Pantone website: https://www.pantone.com/color-of-the-year/2023

scikit-learn. (2019). scikit-learn: machine learning in Python. Retrieved from Scikit-learn.org website: https://scikit-learn.org/stable/

Shabo, V. (2023, May 19). Hollywood and Audiences Can Win with More Diverse and Inclusive Representation on TV and Film. Retrieved from New America website: https://www.newamerica.org/the-thread/diversity-inclusion-storytelling-tv-film/#:~:text=For%20viewers%2C%20representation%20that%20steers

Taste of Cinema. (2023). 10 Reasons Why “Barbie” Is The Movie of The Year. Available: http://www.tasteofcinema.com/2023/10-reasons-why-barbie-is-the-movie-of-the-year/

TextBlob. (2018). TextBlob: Simplified Text Processing — TextBlob 0.15.2 documentation. Retrieved from Readthedocs.io website: https://textblob.readthedocs.io/en/dev/