

**SURVMETH 727**

**Final Project Report**

**Comparative Analysis of Audience Sentiment for Two Films  
with Different Box Office Performances**

**Group 10**

**Yiyang Shu**

**Yiyuan Li**

Github Repository for this project can be found at  
<https://github.com/yyshu917/727-final-project.git>

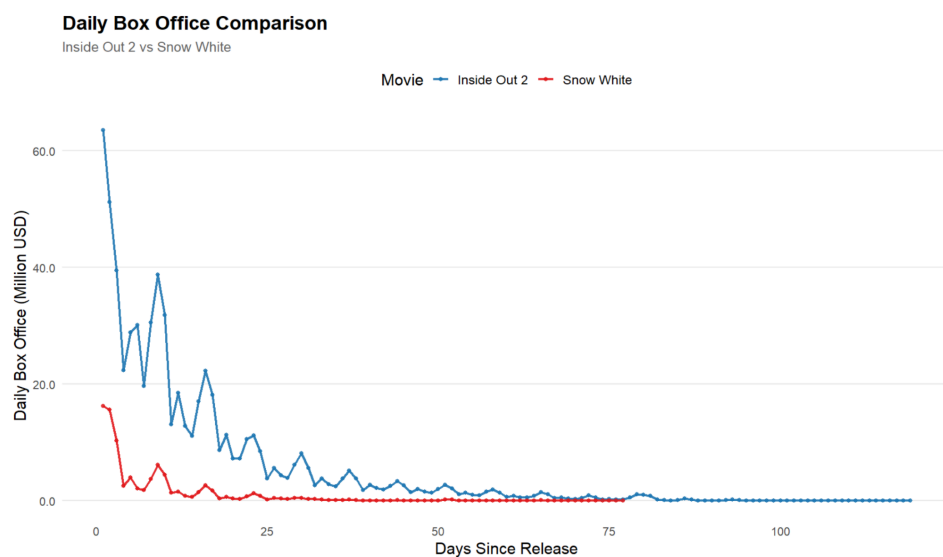
# 1.Introduction

## 1.1 Background

The film industry often presents stark contrasts in audience reception and commercial success, a divergence vividly illustrated by the box office performance of the animated sequel *Inside Out 2* (2024) and the live-action adaptation *Snow White* (2025).

*Inside Out 2*, released on June 14, 2024, became an undisputed blockbuster. Its tremendous success was founded upon a powerful combination of factors, such as exploring Riley's developmental stage of puberty and the exciting integration of new emotions alongside the beloved original cast. This enthusiasm fueled a historic theatrical run, resulting in a domestic box office gross of approximately \$652.98 million.

In stark contrast, *Snow White* (2025) experienced a substantially different outcome. Following its pre-release trailer and subsequent release, the film encountered a significant wave of negative public discourse and sustained controversy. This backlash, which centered on elements of character interpretation, narrative changes, and perceived production quality, likely contributed to its poor commercial showing. The film concluded its domestic theatrical run after 77 days with a gross revenue of approximately \$87.20 million.



In view of these two disparate box office performances, the precise difference in audience reaction and its correlation with commercial performances warrants close examination.

## 1.2 Question of Interest

- (1) What were the temporal trends in daily audience sentiment and comment volume for each film?

- (2) What correlations exist between audience sentiment, comment volume and subsequent daily box office revenue for each film?
- (3) What were the dominant emotional polarities and thematic subjects of audience discourse for *Inside Out 2* and *Snow White (2025)*?

## 2. Methodology

### 2.1 Data Source

Focusing on two films, *Inside Out 2 (2024)* and *Snow White (2025)*, we collected daily box office data from Box Office Mojo using web-scraping techniques. And obtained audience comments from the YouTube data API, sourced specifically from the discussion threads under each film's official trailer.

### 2.2 Research Subject

	Movies	Rows	Columns
Box Office Data	Inside Out 2 (From 2024-06-14 to 2024-10-10)	119	Date, Daily Gross, Gross to Date, Day Number
	Snow White (2025) (From 2025-03-21 to 2025-06-05)	77	
Comments Data	Inside Out 2	10980	Author, Date, Like Count, Comment, Video Id
	Snow White (2025)	25770	

### 2.3 Analysis Techniques

#### (1) Sentiment Analysis

After cleaning the raw comments data, a natural language processing (VADER) model was applied to assign a numerical sentiment score to each comment, ranging from -1.0 to +1.0. The daily average sentiment was calculated by averaging the scores of all comments posted each day. Alongside the comment volume, use time-series plot to track audience emotional trends over time.

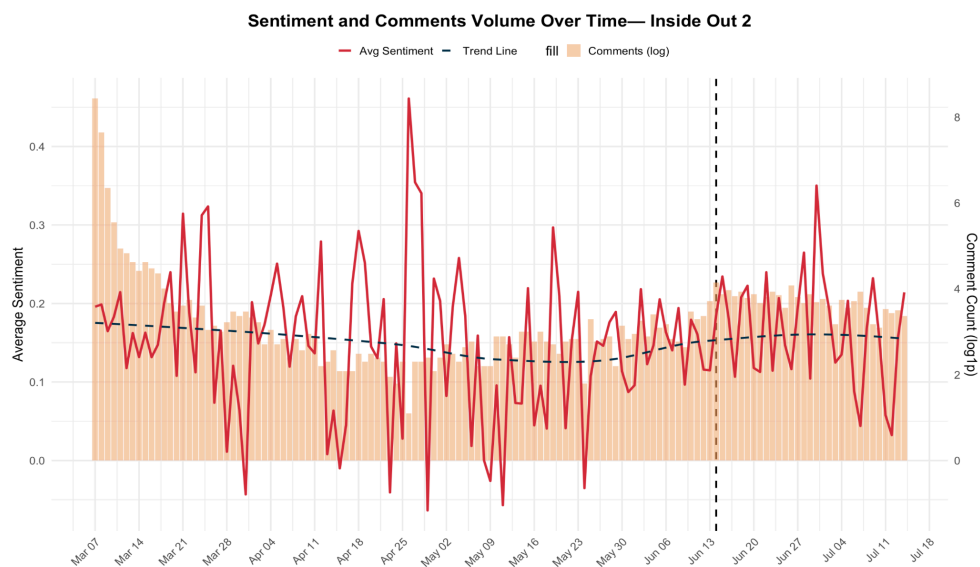
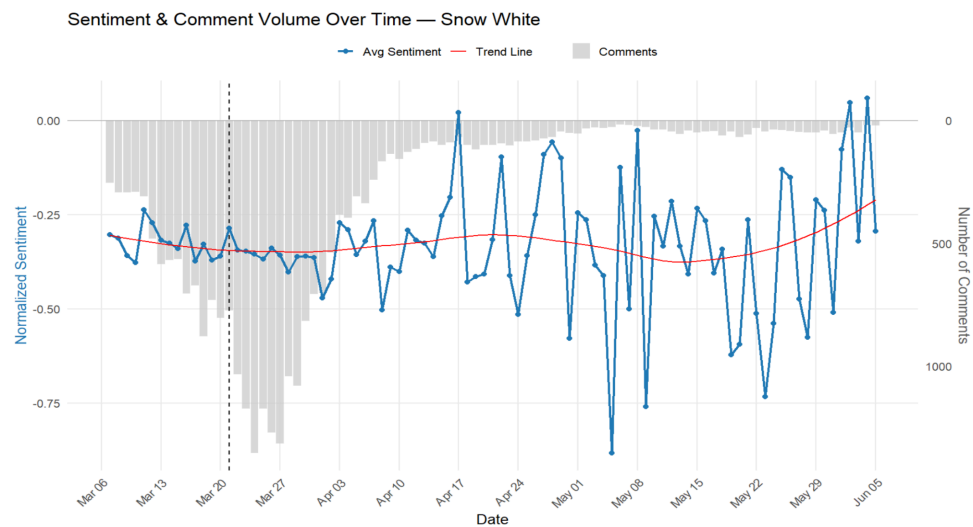
#### (2) Correlation Analysis

Separately define daily comment volume and daily average sentiment score as the predictors, next-day box office revenue was defined as outcome. Pearson correlation coefficient was used to qualify the relationship between variables, and a scatter plot with regression line was used for visualization.

#### (3) Word Frequency and Topic Modelling

For word frequency, the cleaned tokens were tallied to determine the frequency of each unique word. The top 30 most frequent words were then visualized using two horizontal bar charts. For topic modeling, Latent Dirichlet Allocation (LDA) was used to identify underlying topics within the text corpus by grouping words that frequently co-occur.

### 3.Sentiment Analysis



The trend of comment volume showed how people's engagement looks like. For *Snow White (2025)*, in the pre-release stage, the comment volume dramatically increased leading up to the release date marked by the dashed vertical line on March 21st. And then, the volume peaked immediately following the release date (March 22nd onward). This massive volume exceeded 1,000 comments per day and was sustained for about one week. This suggests the release was a highly polarizing or widely discussed event. After the initial post-release week, the comment volume decreased rapidly day by day as the immediate public discourse subsided.

For *Inside Out 2*, the highest volume of comments occurred on March 7th—the date the official trailer was released. This initial burst of volume is usually a common pattern for high-profile movies, indicating strong immediate interest. Following this initial spike, the comment volume settled into a lower, fluctuating level throughout the pre-release marketing campaign. And then, there is a noticeable, sustained increase in comment volume around the movie's release (June 14th). The volume appears to remain relatively stable and consistent at

this higher level until the end of the plotted period, suggesting continuous discussion and positive word-of-mouth well after the initial release.

The two movies showed a highly distinct engagement pattern. The total comment volume for *Snow White (2025)* was significantly higher than that of *Inside Out 2*. To be specific, *Inside Out 2* experienced its highest peak with the trailer release, followed by sustained, moderate volume post-release. *Snow White* saw a rapid buildup to a massive, short-lived peak immediately following the release, indicating a more intense, but less sustained, burst of public conversation.

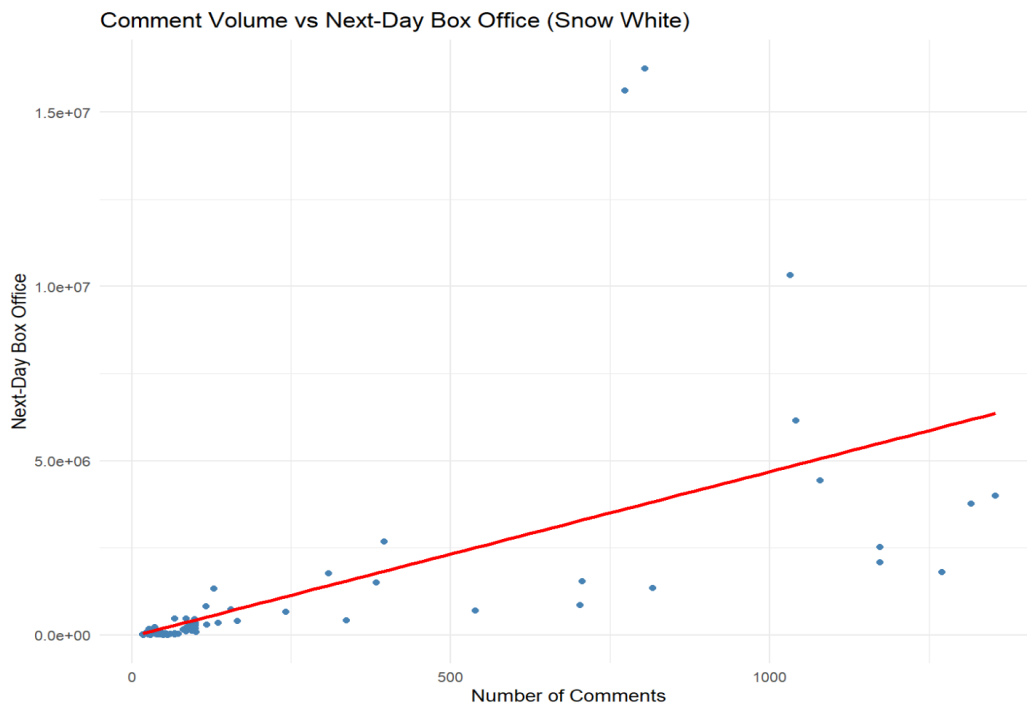
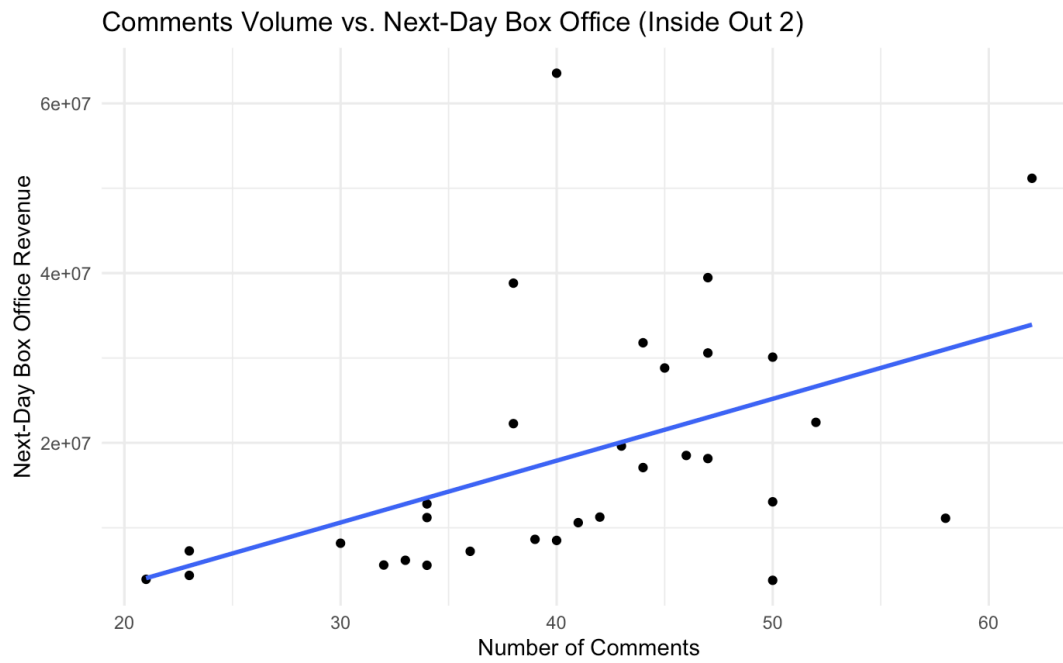
The trend of sentiment scores shows what audience reception looks like. For *Snow White (2025)*, the sentiment scores are predominantly negative, with the scores generally fluctuating below zero. The sentiment remained relatively stable and negative in the period leading up to the March 21st release, hovering around a score of -0.25 to -0.30. About 10 days after the release, the sentiment became extremely volatile. These extreme negative spikes suggest specific events or discussions drove intense periods of backlash.

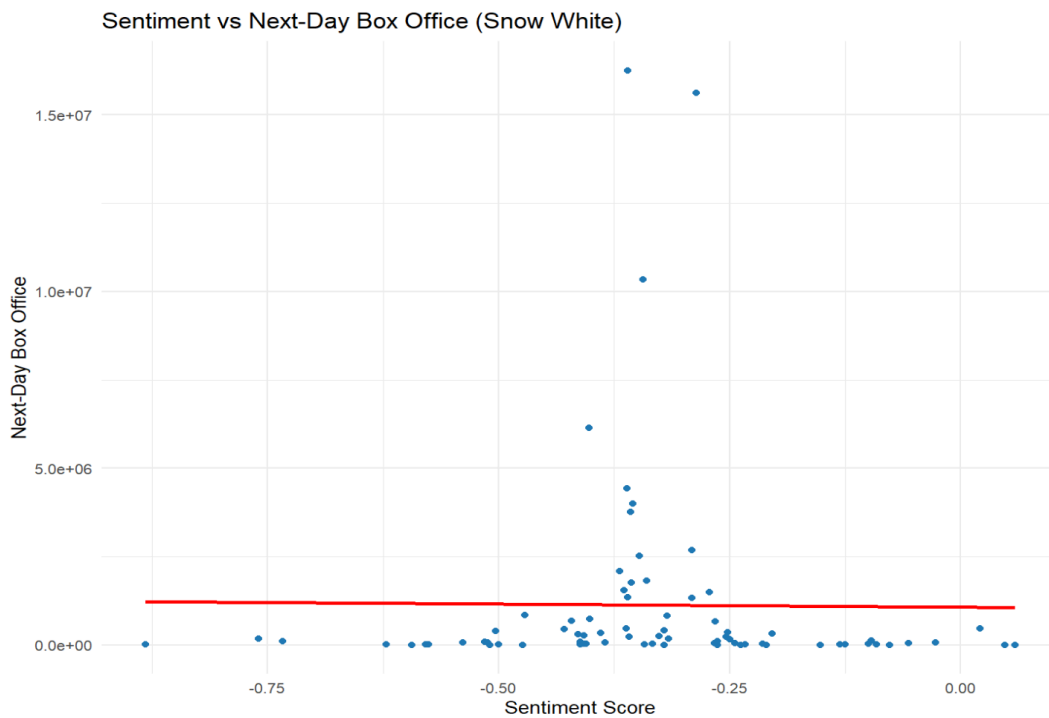
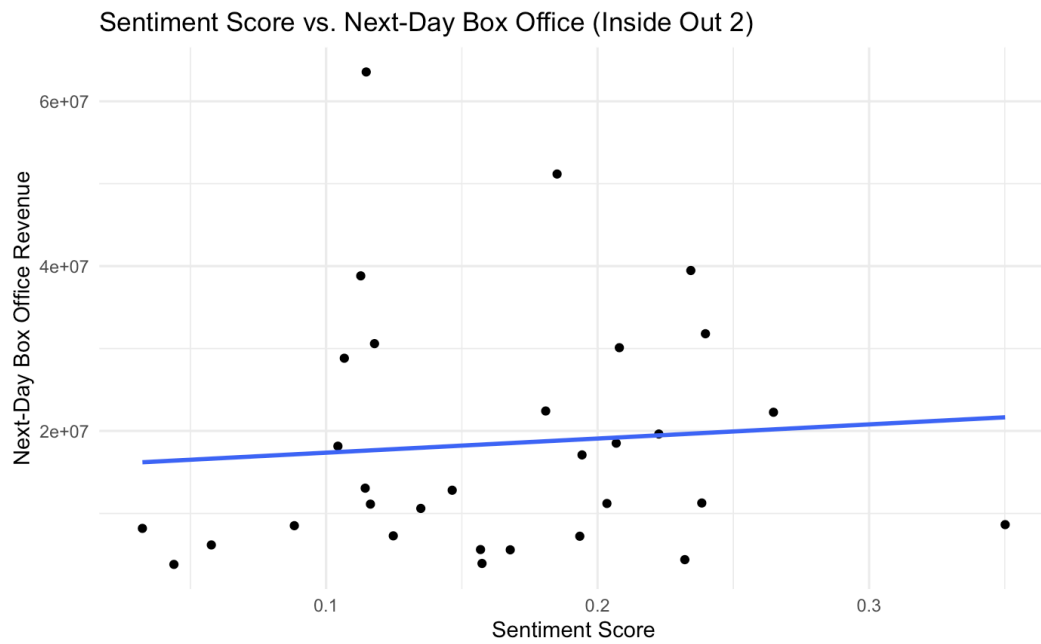
For *Inside Out 2*, the sentiment scores are overwhelmingly positive, with the average sentiment consistently remaining above zero throughout the entire observation period. Moreover, the daily sentiment score shows significant fluctuation during the long pre-release period (March through May). However, in the days leading up to and following the assumed release period, the sentiment scores tend to remain in a tighter, positive band (approximately between 0.1 and 0.2), indicating a more consistent and strong favorable impression from the audience once the movie was available.

The sentiment for the two movies is a stark contrast. The strength and intensity of the negative public reaction to *Snow White* was far greater than the positive public reaction to *Inside Out 2*, and the sentiment fluctuation was much higher as well.

Overall, the plots reflect some rationales: (1) Issues of controversy, politics, and social commentary are far more likely to generate massive, intense, and immediate comment volume than simple movie anticipation. (2) Negative, angry, or outraged comments are often more readily shared.

## 4. Correlation Analysis





To examine whether sentiment and discussion volume have an effect on box office, our correlation analysis examined the relationship between two audience engagement measures—Daily Comment Volume and Sentiment Score—and the Next-Day Box Office Revenue for both *Inside Out 2* and *Snow White*. This next-day offset was used to account for the temporal influence of online discussion on immediate consumer behavior.

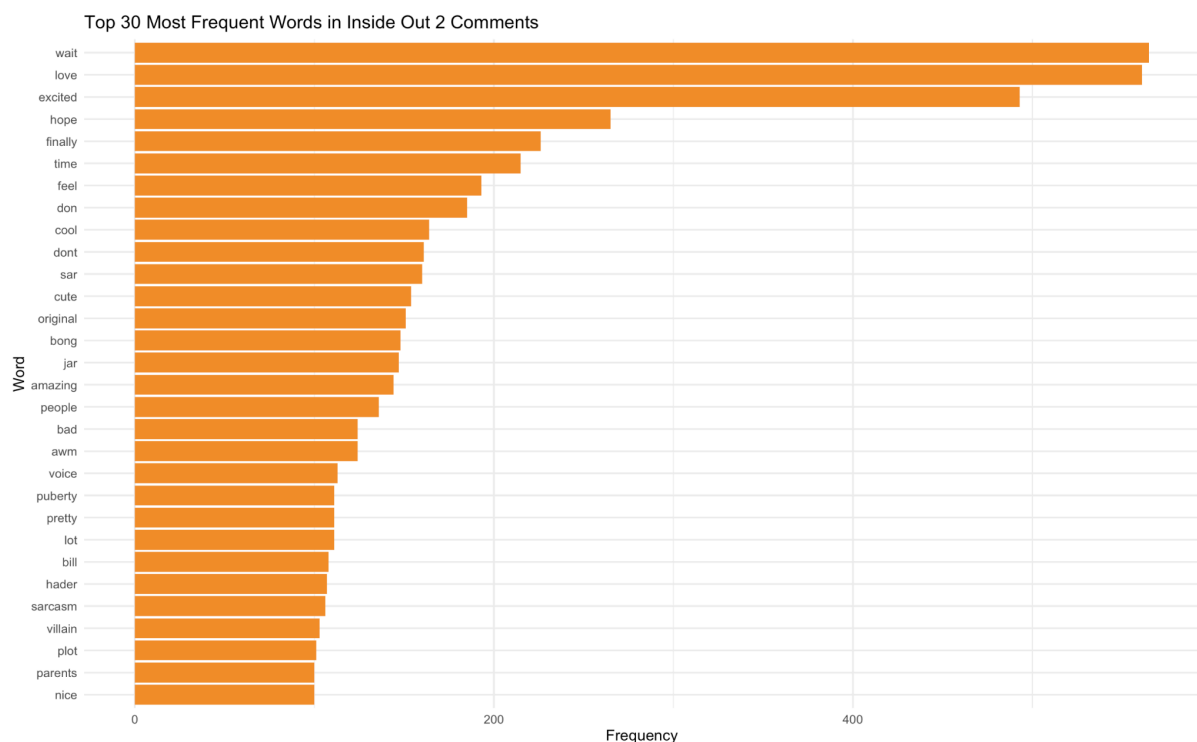
(1) Comment Volume vs Next-Day Box Office:

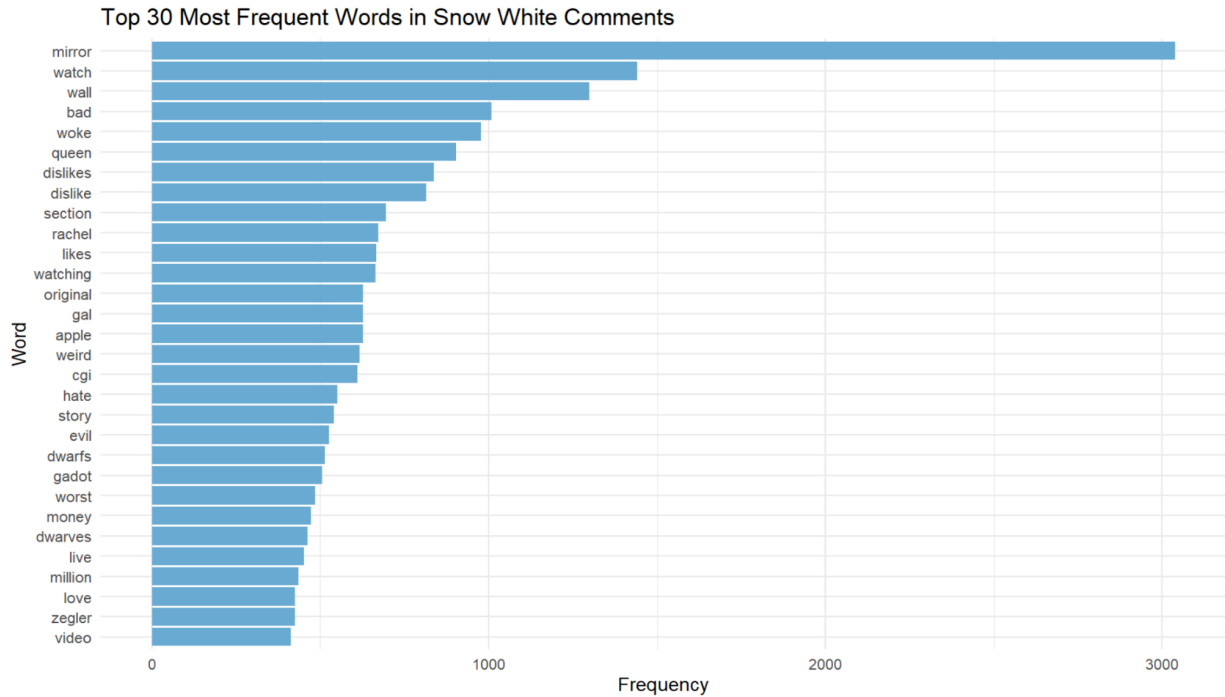
Both films show a clear positive correlation between daily comment volume and next-day box-office revenue. For *Inside Out 2*, the relationship is moderate ( $r = 0.476$ ), with an upward-sloping regression line suggesting that greater discussion volume generally aligns with higher revenue, despite a few high-comment, low-revenue points. *Snow White* also shows a positive slope ( $r = 0.607$ ), but most data cluster at very low comment and revenue levels, with a few extreme outliers driving the correlation. This pattern indicates a less consistent relationship than that observed for *Inside Out 2*.

## (2) Sentiment Score vs. Next-Day Box Office:

In contrast, sentiment scores exhibit a negligible or non-existent relationship with next-day box office revenue for both films. For *Inside Out 2*, the regression line is nearly flat, indicating that fluctuations in positive sentiment did not predict revenue changes. *Snow White* displays a similarly horizontal line near the x-axis, confirming that its predominantly negative sentiment was also uncorrelated with short-term box-office outcomes.

## 5. Word frequency & Topic modelling





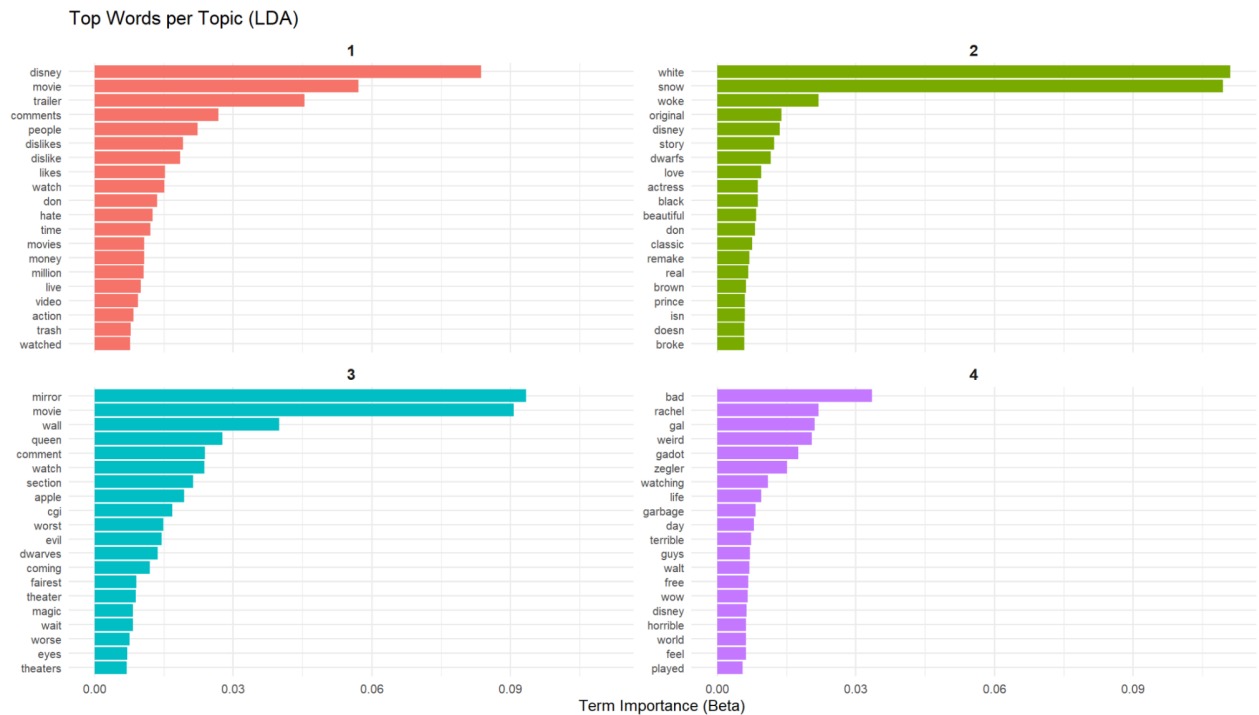
After examining the sentiment trends and the relationship between audience engagement and box-office performance, we next aim to understand the thematic focus of online discussions surrounding the two films. To do this, we conducted a word frequency analysis of YouTube trailer comments, removing stop words and other overly common terms such as the movie titles themselves. The results reveal a stark contrast in the emotional tone and narrative emphasis between *Inside Out 2* and *Snow White*. Comments for *Inside Out 2* are dominated by highly positive and anticipation-driven words such as “wait,” “love,” “excited,” “hope,” and “finally.” This pattern suggests that viewers engaged with the trailer primarily through expressions of emotional enthusiasm and personal connection to the original film. Additional frequently used terms like “cute,” “amazing,” and “cool” further reinforce the broadly supportive and celebratory sentiment surrounding the sequel.

In contrast, the word frequency distribution for *Snow White* points to a more polarized and critical discourse. The most common terms, “mirror,” “watch,” and “wall”, reflect a sarcastic online meme phrase that audiences use to criticize the film. People often begin with “mirror, mirror on the wall” and follow it with a mocking line, such as “*Mirror, mirror on the wall, who is the bravest of them all? Disney, for not locking the comment section.*”

Words like “bad,” “dislike,” “hate,” “weird,” and “evil” indicate substantial negativity in viewer responses, while the prominence of actor-related terms such as “rachel,” “gal,” and “zegler” suggests that casting choices played a major role in shaping audience reactions. Compared to the overwhelmingly positive vocabulary surrounding *Inside Out 2*, the language used for *Snow White* centers on criticism, skepticism, and controversy.

Due to the controversy reflected in the *Snow White* comment section, we further applied topic modelling (LDA) to uncover the underlying themes driving these discussions. The

model identifies **four distinct topics**, each representing a different dimension of public criticism or concern.



**Topic 1** centers on general reactions to the film and its promotional materials, with frequent references to **“Disney,” “trailer,” “comments,” “dislikes,”** and **“likes.”** These terms suggest a broad wave of dissatisfaction with both the marketing campaign and the film’s perceived creative direction. Many commenters explicitly summarized the disproportionate ratio of dislikes to likes as evidence of widespread public rejection. A typical example from this topic reads: *“For people who don’t use the extension currently: Likes: 71,454 — Dislikes: 2,167,344.”* Such comments frame audience sentiment as overwhelmingly negative even before the film’s release.

**Topic 2** captures debates surrounding the film’s reinterpretation of the original Snow White story, reflected in terms such as **“white,” “snow,” “black,” “brown,” “original,” “classic,”** and **“remake.”** Notably, **“woke”** emerges as one of the most prominent words. Historically, *woke* originated in African American English as an adjective referring to awareness of racial injustice and structural discrimination. In contemporary political discourse, however, the term has been co-opted, often pejoratively, to describe media or corporate decisions perceived as overly politically correct or driven by diversity initiatives rather than artistic intent. Within this topic, **“woke”** signals audience frustration with the film’s casting of a Latino actress in a traditionally fair-skinned role and broader concerns that the remake is motivated by modern ideological agendas rather than fidelity to the original narrative. Typical comments include: *“Now I’m going to see Snow Black,”* and *“GO WOKE, GO BROKE!”* highlighting how race, representation, and anti-“wokeness” rhetoric became central to the discourse.

**Topic 3** highlights the viral meme phrase identified earlier, with “**mirror**,” “**movie**,” “**wall**,” and “**queen**” dominating the topic. These terms reveal how humor and sarcasm became key mechanisms for expressing disapproval and amplifying criticism. Additionally, the presence of words such as “**cgi**” and “**dwarves**” indicates widespread criticism with the film’s visual effects, particularly the digitally rendered dwarves. Representative comments include: “*Evil Queen: Did I frighten you? The general public: No, Snow White and her seven CGI did,*” and “*Evil Queen: Mirror, mirror on the wall, who’s the greatest of them all? Mirror: obviously you and the apple.*” Such remarks show how meme culture shaped the tone and spread of negative sentiment.

Finally, **Topic 4** focuses heavily on casting-related backlash. Words such as “**bad**,” “**rachel**,” “**gal**,” “**zegler**,” and “**gadot**” point clearly to audience dissatisfaction with Rachel Zegler, the actress playing Snow White. Much of the criticism is intertwined with her public comments about the original 1937 film as well as perceptions that she does not match the traditional depiction of the character. At the same time, comparisons with Gal Gadot, who plays the Evil Queen and is far more positively received, reinforced the polarized reaction. Typical comments include: “*Rachel Zegler: Snow White is not gonna be saved by the Prince. The Prince: Good, because I’m here for your stepmother,*” and “*I will only watch it for Gal Gadot.*” These responses illustrate how casting decisions, actor popularity, and public controversies contributed significantly to the negative discourse surrounding the film.

In summary, the four topics reveal that criticism of *Snow White* was not driven by a single issue but by a series of interconnected controversies. Viewers expressed dissatisfaction with the film’s marketing, objected to perceived ideological or political motivations behind the remake, mocked visual and technical choices such as CGI, and reacted strongly to casting decisions, particularly those involving Rachel Zegler. The prominence of terms related to race, representation, and “wokeness” demonstrates that the discourse extended beyond artistic preferences into broader cultural and political debates. Meanwhile, the widespread use of memes and sarcastic commentary illustrates how online humor amplified the negativity and helped it spread rapidly across platforms.

## 6. Discussion

### 6.1 Conclusion

Across all analyses, the two films exhibit fundamentally different patterns of online engagement and audience perception, reflecting both their unique production contexts and the cultural narratives surrounding them. *Inside Out 2* generated a sustained stream of positive sentiment and steady discussion volume, consistent with the behavior typically observed for highly anticipated sequels with strong franchise loyalty. Although its comment volume showed only a moderate next-day correlation with box-office performance, this stable engagement suggests that positive sentiment functioned more as general goodwill than as a direct predictor of revenue spikes. Its audience conversation concentrated on excitement,

emotional resonance, and nostalgia. These elements tend to support long-term box-office strength .

In contrast, *Snow White* attracted an extraordinarily high but short-lived comment volume, dominated by intense negative sentiment. The release triggered a surge of engagement far exceeding that of *Inside Out 2*, yet this engagement was primarily driven by controversy rather than enthusiasm. The thematic and topical analyses further illuminate why *Snow White* generated such a strong but unstable reaction. The discourse was fragmented across multiple, overlapping controversies—ranging from frustration with the remake’s departure from the original story to broader political debates about race, “wokeness,” and representation. The prevalence of meme culture and sarcastic commentary amplified the spread of criticism, contributing to a highly viral but diffuse conversation. Though negative virality drove engagement but did not translate into box-office gains.

Overall, the findings suggest that comment volume reflects the intensity of public attention, whereas sentiment reflects its emotional direction, but neither alone reliably predicts short-term box-office performance.

## 6.2 Limitation

First, All audience discussions were collected from YouTube trailer comments. YouTube demographics skew younger and more online-active, and may not represent the broader moviegoing population. Sentiment patterns on other platforms like Reddit, TikTok, or X (Twitter) might differ substantially. Also we only gain the comments from the trailer. Further discussion about comments from evaluation videos might also be important.

Second, the correlation between comment volume and box office is likely a spurious relationship, heavily influenced by shared, time-dependent patterns rather than direct causality. Comment volume usually is highest when the film is just released, peaking during the initial phase of public interest. Concurrently, the box office was also highest when it just released, followed by a natural decline in revenue. Because both variables are simultaneously influenced by the confounding factor of time relative to the release date and external market factors like time (weekend or not), this cannot strictly imply a causal relationship. The strong volume correlation simply confirms that the periods of maximum public attention (volume) align with periods of maximum commercial activity (revenue).

Third, there is a limitation in the measurement itself. The single numerical sentiment score may fail to capture the nuanced emotional drivers that translate into ticket sales. The lack of daily correlation between sentiment score and box office suggests that the cumulative, long-term sentiment established prior to release, which shapes initial public expectation, is likely more critical to general box-office success than minor daily changes in audience mood.

A film's overall performance is largely determined by pre-release discussion and expectation, making day-to-day fluctuations in sentiment less impactful once the film is in theaters.