

Taylor Swift in the Media

Amy Huang, Selin Omur, Minh Tran, Ella Young

McGill University

amy.huang@mail.mcgill.ca, selin.omur@mail.mcgill.ca, minh.tran2@mail.mcgill.ca, ella.young@mail.mcgill.ca

Introduction

This study presents a comprehensive analysis of media coverage surrounding the iconic music artist, Taylor Swift, particularly focusing on North American news sources. In an era where media narratives can significantly influence public perception, understanding the nature of this coverage is crucial. Our analysis is rooted in a methodical examination of over 500 news articles, aiming to decipher the tone and thematic focus of the coverage.

By employing a rigorous open coding process on a subset of 500 articles, we meticulously categorized each article into eight distinct topics; (1) Taylor Swift: The Eras Tour Movie, (2) Travis Kelce & the NFL, (3) Albums & Songs, (4) Billionaire Status & Profits, (5) Awards, Award Shows & Records, (6) Performances & The Eras Tour, (7) Style & Appearance, and (8) Personal Life. The categories receiving the most attention from new media included Swift's many recent Awards, Award Shows and Records, and her relationship with Travis Kelce and the NFL. These categories each accounted for approximately 20% of the articles we examined. Our approach involved not only identifying these themes but also characterizing them through the calculation of the top ten terms with the highest Term Frequency-Inverse Document Frequency (tf-idf) scores.

Furthermore, we conducted a sentiment analysis to classify each article as positive, negative, or neutral towards Taylor Swift. This approach provided a nuanced understanding of the media's portrayal of Swift, offering insights into how she is represented in the public sphere. The sentiment analysis revealed that the news media's coverage was largely positive or neutral, with few articles focusing negatively on Swift. The study's findings are not just relevant to media professionals and Taylor Swift's fans, but also contribute to the broader discourse on celebrity media coverage in the digital age.

Data

Our project's dataset comprises a targeted collection of news articles centered on the popular music artist Taylor Swift. This dataset was meticulously gathered using NewsAPI.org, a reliable source for accessing a broad range of news content.

The dataset includes a total of 500 articles, carefully curated based on their publication dates ranging from October 20th to November 15th of this year. This specific time frame was selected to capture the most recent and relevant information about the artist. Additionally, to avoid potential biases associated with short-term trends or events, we ensured our sampling spanned over the entire month. This approach mitigated the risk of bias towards news that was trending on specific days, allowing for a more comprehensive and balanced representation of the media coverage during this period.

To ensure the dataset's relevance to our project's focus, we applied a strict filtering criterion. Only articles explicitly mentioning 'Taylor Swift' were included. Additionally, we employed an open coding methodology to 'remove' articles where Taylor Swift was not the primary focus of the coverage. This step was crucial to maintain the dataset's consistency and focus, ensuring that all content was directly related to the artist. Furthermore, to provide a specific cultural and geographical context, our study concentrated on North American sources. The origin of the news sources was examined to confirm this geographical focus, thereby enhancing the regional specificity of our analysis.

For each article, three key pieces of information were extracted: the title, the full content, and the publication date. This information provides a comprehensive view of each article, allowing for detailed analysis of the media coverage surrounding Taylor Swift during the specified period.

A significant challenge in the data collection process was the limitation imposed by NewsAPI.org, which only allowed us to retrieve 100 articles per API call. To circumvent this restriction and acquire the complete set of desired articles, we made multiple API calls, adjusting the date range with each request such that we collected articles from five different five day spans. This methodical approach enabled us to gather the full dataset effectively.

Considering the nature of the data, particularly the presence of alphanumeric quotes within the article content, we chose to store the dataset in a Tab-Separated Values (TSV) file. The TSV format was selected over more common alternatives like the CSV format due to its ability to handle data containing quotes without compromising the dataset's integrity by being interrupted by insignificant commas. This choice ensured a clear and unambiguous representation of the data, facilitating easier and more accurate analysis.

In summary, our dataset is a carefully curated collection of news articles that focus specifically on Taylor Swift, gathered over a nearly one-month period. The dataset's composition reflects deliberate choices in data collection, filtering, and storage, aimed at providing a comprehensive and unambiguous resource for our project.

Methods

To study the portrayal of Taylor Swift in news media, we developed a series of Python scripts to perform intricate data processing and analysis. These scripts, distinct from the dataset collection process, played a crucial role in extracting meaningful insights from the raw data.

Initially, we developed a Python script that used the NewsAPI platform to scrape a variety of news websites. We had to extract 100 articles from different benchmark dates to retrieve all 500 articles due to restrictions on developer use of NewsAPI. This script was instrumental in automating the extraction of article content, titles, and publication dates.

Another Python script was designed to extract all words from each article in their respective categories. This process involved parsing the content of each article and separating it into individual words, ignoring “stop words” such as any prepositions and articles. This step was crucial for the subsequent text analysis and ensured that our analysis was grounded in the actual content of the articles.

To identify the most significant words in our dataset, we implemented a Python script to calculate the Term Frequency-Inverse Document Frequency (TF-IDF) of the words. We intentionally excluded the aforementioned stop words to focus on the most meaningful terms. The TF-IDF algorithm helped us determine the importance of a word in a document set, which was essential in understanding the context in which Taylor Swift was discussed. The top 10 words for every category, along with their scores, were then saved in a JSON file for easy access and analysis.

A pivotal part of our methodology was sentiment analysis. We created a Python script to categorize the sentiment of each article as negative, positive, or neutral. This classification was not just about tallying sentiments; it was about understanding the overall media perspective on Taylor Swift. The results of this analysis were visually represented in a pie chart, providing a clear and immediate understanding of the sentiment distribution. Additionally, we prepared a separate graph that displays the sentiment distribution for each category, allowing us to visually represent and delve deeper into how sentiments varied across different categories of articles.

The decision to utilize Python scripts for each of these tasks was guided by Python's robust libraries and tools for data scraping, natural language processing, and data visualization. This choice enabled us to efficiently process large amounts of data and extract nuanced insights, which would have been challenging with less versatile programming environments. Our methods, therefore, reflect a combination of technical proficiency and strategic planning, aimed at comprehensively analyzing the media's portrayal of Taylor Swift.

Results

After our initial open coding on the first 200 articles, we established 8 categories, each offering insight into distinct aspects of Taylor Swift's media coverage. These topics were carefully chosen to offer a comprehensive view of the artist's portrayal in the news. The topics/categories and their definitions are as follows:

The Eras Tour Movie

Coverage mostly focusing on Taylor Swift's new concert movie, exploring the narrative surrounding this project, and how fans are responding to it.

Travis Kelce & the NFL

Articles whose primary focuses are Swift's relationship with Travis Kelce and/or her connection to the NFL.

Albums & Songs

Coverage centered around Taylor Swift's albums and songs, providing insights into the reception and discussion of her musical work.

Billionaire Status & Profits

Articles about Swift's financial standing and profits, exploring discussions on her billionaire status and financial success.

Awards, Award Shows & Records

Coverage related to Swift's achievements, participation in award shows, and the records she holds, highlighting her prominence in the music industry. This category includes award-winning and record-breaking songs/albums. This category does not include her outfits or physical appearance at awards shows.

Performances & The Eras Tour

Media discourse on Swift's on-stage performances and The Eras Tour (specifically the live performances, not the movie).

Style & Appearance

Articles focusing on Taylor Swift's fashion choices, style evolution, and overall appearance in the public eye.

Personal Life

Coverage of Swift's personal life and daily life, including aspects beyond her professional career, such as friendships/past relationships (excluding Travis Kelce), family, fans, and personal milestones.

Our analysis of the chosen topics revealed varying levels of media engagement. The categories that had the most media coverage out of the 500 articles analyzed were Awards, Award Shows & Records, Travis Kelce & the NFL, and Albums & Songs with the distribution of 21.9%, 18.8%, and 15.0%, respectively. This heightened focus emphasizes the media's keen interest in Swift's achievements and her personal relationships.

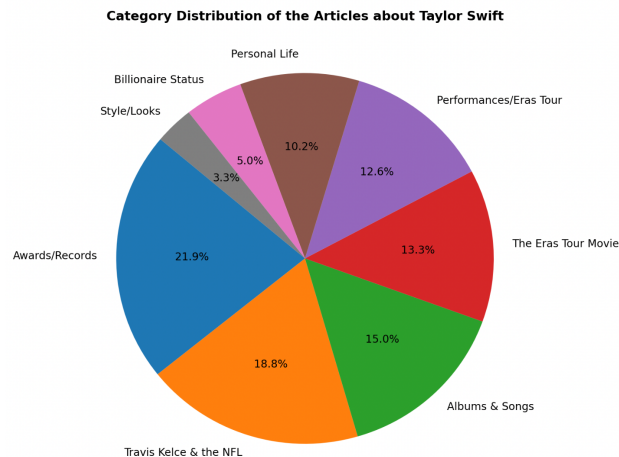


Figure 1: Category Distribution of the Articles about Taylor Swift

Our open coding process also revealed the nuanced sentiments expressed in the media coverage of Taylor Swift. The distribution of sentiments across the articles was predominantly positive or neutral, with a limited number having a negative stance. This analysis provides valuable insights into the overall tone and perception of Taylor Swift in the media landscape. The breakdown of the sentiments is as follows:

Sentiment Distribution of Articles about Taylor Swift

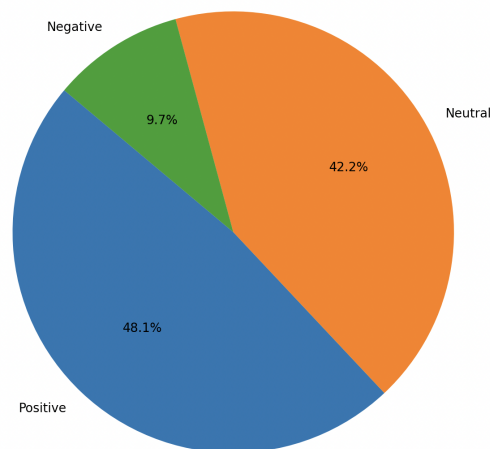


Figure 2: Sentiment Distribution of the Articles about Taylor Swift

Additionally, our analysis included a comprehensive examination of the overall sentiment within each category. The chart below illustrates the distribution of

sentiments –positive, neutral, and negative– across the different topics:

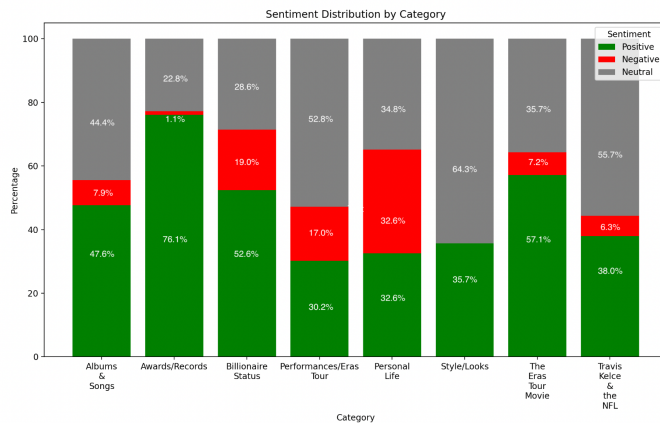


Figure 3: Sentiment Distribution by Category

To gain more insight about the language used in each category, we computed the top ten words with the highest Term Frequency-Inverse Document Frequency (TF-IDF) scores. The following highlights the distinctive words that play a significant role in shaping the narrative for each category:

The Eras Tour Movie

"box": 39.4 "office": 37.6 "million": 25.1
"movie": 17.6 "weekend": 17.6 "concert": 17.6
"film": 14.3 "theaters": 12.5, "amc": 12.5
"five": 12.5

Travis Kelce & the NFL

"chiefs": 73.5 "kansas": 68.1 "travis": 63.7
"kelce": 31.9 "tight": 30.5 "kelces": 17.9
"nfl": 17.6 "city": 15.8 "play": 14.3
"star": 14.3

Albums & Songs:

"music": 16.5 "album": 16.5 "1989": 15.3
"rerecording": 14.3 "version": 11.8 "vault": 10.8
"antonoff": 10.8 "lot": 8.9 "written": 8.9

"albums": 8.8

Billionaire Status & Profits:

"billionaire": 10.8 "worth": 8.9 "estimated": 7.2
"billion": 7.2 "33yearold": 7.2 "11": 5.5
"financial": 5.4 "business": 5.4 "bloomberg": 5.4
"matt": 5.4

Awards, Award Shows & Records:

"album": 32.9 "grammy": 31.9 "1": 26.9
"nominations": 25.1 "billboard": 19.8 "nominees": 17.9
"biggest": 17.9 "sza": 17.9 "music": 16.5
"10": 16.1

Performances & The Eras Tour:

"mexico": 14.3 "buenos": 10.9 "aires": 10.9
"leg": 10.8 "vancouver": 10.7 "tour": 10.7
"argentina": 9.9 "onstage": 9.9 "august": 8.8
"concert": 7.7

Style & Appearance:

"savvy": 5.4 "travis": 3.3 "halloween": 3.3
"kelce": 2.1 "images": 0.5 "tour": -0.5
"eras": -0.6 "swifts": -0.9 "2026": -1.2
"swift": -2.3

Personal Life:

"kelly": 10.8 "news": 8.9 "getty": 7.2
"gomez": 7.2 "brittany": 5.5 "mahomes": 5.5
"media": 5.4 "amy": 5.4 "sussmangetty": 5.4
"ul": 5.4

These TF-IDF results offer a deeper understanding of the distinctive terms that shape the narrative within each category, contributing to the specificity of Taylor Swift's media coverage.

Discussion

The overall sentiment of articles about Taylor Swift is mostly Positive (48.1%), followed by Neutral (42.2%) and a smaller proportion of Negative (9.7%) sentiment. This indicates that while almost half of the articles are positive, there's a significant amount of neutral reporting, with a relatively small amount of negative coverage.

In terms of categories, Awards, Award Shows & Records (21.9%) and Albums & Songs (18.8%) are the largest segments, indicating a significant amount of coverage is dedicated to her achievements and music.

From our TF-IDF analysis, "grammy", "nominations", and "nominees", are among the most frequently used words within Awards, Award Shows & Records, likely in correlation with the grammy nominations which were announced on November 10, 2023. Additionally, the media coverage of Taylor Swift's Awards, Award Shows & Records appeared to be the most overwhelmingly positive, with about 76% Positive articles, and only 1% negative. So, it's reasonable to assume that public opinion is in favor of Swift's success in the music industry, specifically at large awards shows like the annual Grammys. However, it's important to note that media coverage of Swift's Awards, Award Shows & Records is likely to be interpreted as positive due to the celebratory nature of the topic itself.

As for Albums & Songs, media coverage of this topic was dominated by words like "rerecording" and "version", likely alluding to Swift's rerecording of all her albums such that she owns the rights to her music, which are titled the same, followed by "(Taylor's Version)" (Bruner 2023). Specific references to "1989" and "antonoff" suggest discussions about specific albums and collaborations; Taylor Swift and Jack Antonoff are known to be a musical duo, and collaborated on her recent album "1989 (Taylor's Version)" (Gibson 2023a). The Albums & Songs category had almost an even number of Positive (47.6%) and Neutral (44.4%) articles, with minimal Negative (7.9%) articles. We can infer that recent media portrayals of Swift's music are generally not negative, but they also don't lean towards being exceptionally positive.

Travis Kelce & the NFL (15.0%) is also a substantial segment, given how specific it is, and how far removed

that industry seems to be from Taylor Swift. However, Swift and Kelce went public with their relationship in October 2023 (Gibson 2023b). Terms that dominate this category include "chiefs", "kansas", and "nfl", which align with Kelce's profession; a football player on the NFL team the Kansas City Chiefs. Travis Kelce & the NFL has a majority Neutral sentiment (55.7%), followed by Positive (38%), and Negative (6.3%). So, while there is seemingly substantial interest in the relationship, the media has not expressed many strong feelings.

Taylor Swift's Performances (12.6%) and the recently released The Eras Tour Movie (13.3%) have also received substantial media coverage. In terms of Performances, frequent words were geographically focused, with cities like "mexico", "buenos aires", "vancouver", and "argentina" indicating recent tour locations, as Swift embarks on her international leg of The Eras Tour, after first touring in the United States. Unlike prior categories, 17% of articles expressed negative views on the tour, perhaps in relation to her fans (known to be over the top), or her carbon emissions from traveling and putting on a massive performance. The Eras Tour Movie, however, only received 52% positive coverage, with only 7% of articles being negative. The TF-IDF analysis revealed frequent words to be "box", "office", "million", and "weekend" indicating strong box office performance in the first weekend. So, while the media may not be such a fan of The Eras Tour, they do like The Eras Tour Movie.

More generally, Taylor Swift's Personal Life (10.2%) is of interest to the media. Unlike all the other categories, there is significant negative coverage here, at nearly 33%. Here there is a clear focus on personal associations with names like "kelly", "gomez", "brittany", and "mahomes" indicating a relationship with Brittany Mahomes (wife of Patrick Mahomes, who is on Travis Kelce's NFL team) (Gibson 2023b). Terms like "news", "media", and specific references to "getty" suggest that the majority of this coverage is the result of paparazzi, who sell their photos to the media company Getty Images. The paparazzi have a reputation for painting celebrities like Swift in a negative light, so this may be the cause of the majority negative sentiment.

Taylor Swift is now worth over \$1.1 billion, so naturally the media has a take (Pendleton 2023). The sentiment distribution over articles is more even in this category, while still being overwhelmingly positive, with 52% positive, 19% negative, and 28% neutral coverage. Because this category only makes up 5% of our sample, the TF-IDF analysis is less reliable; as such, many of the

terms relay surface level details. Words like "billionaire", "worth", "estimated", "billion", and "financial" indicate a focus on financial status and achievements.

Taylor Swift's Style & Appearance (3.3%) appears to be of least interest to the media currently. There is a divergent array of frequent words here, likely because so few articles from our sample were dedicated to this topic. Travis Kelce was mentioned frequently here ("travis" and "kelce"), with articles likely covering Swift's fashion when going on dates and attending NFL games. Additionally "halloween" was referred to, likely referencing a spike in fans dressing as Taylor Swift for Halloween (Kaufman 2023). Negative scores for "tour", "eras", "swifts", "2026", and "swift" could indicate these terms are common across categories, thus less distinctive in this context. Overall, it seems like our samples that focus on Style & Appearance did not contain enough articles to come to a consensus regarding sentiment or specific topics covered. However, it's important to note that no articles in this category expressed a negative sentiment.

In summary, the TF-IDF analysis provided a linguistic map of the distinctive terms shaping Taylor Swift's media narrative, while the sentiment and category distribution charts offer a quantitative view of how those narratives are received and the volume of coverage they receive. Together, they paint a picture of a media landscape that is heavily influenced by her professional achievements, with a keen interest in her personal life (particularly her relationships), and a positive reception overall, with pockets of negativity possibly relating to personal or critical aspects.

Group Member Contributions

Amy Huang completed the Data and Methods sections of the document. Huang also did an open coding on 125 articles.

Selin Omur completed the Results section of the document, as well as computing all graphs and numerical values. Omur also did an open coding on 125 articles.

Minh Tran was the primary contributor to the Python scripts that acquired, parsed, and filtered the data from the articles. Additionally, Tran completed the Python script that parsed the articles and calculated the Term Frequency-Inverse Document Frequency (TF-IDF). Tran also did an open coding on 125 articles.

Ella Young completed the Introduction and Discussion sections of the document, as well as the formatting and

general revisions for all sections. Young was the secondary contributor to the Python scripts that acquired, parsed, and filtered the data from the articles. Young also did an open coding on 125 articles.

References

- Bruner, R. 2023. Here's Why Taylor Swift Is Re-Releasing Her Old Albums, Time. <https://time.com/5949979/why-taylor-swift-is-rerecording-old-albums/>. Accessed: 2023-11-30.
- Enos, M. 2023. Air date for 2024 Grammys announced: Sunday, Feb. 4, live in Los Angeles; Grammy awards nominations to be announced Friday, Nov. 10, 2023, Grammy Awards. <https://www.grammy.com/news/2024-grammys-air-date-nominations-day-show-announcement-66th-grammy-awards>. Accessed: 2023-11-30.
- Gibson, K. 2023a. Taylor Swift and Jack Antonoff's friendship timeline, Peoplemag. <https://people.com/music/taylor-swift-and-jack-antonoff-friendship-timeline/>. Accessed: 2023-11-30.
- Gibson, K. 2023b. Taylor Swift and Travis Kelce's relationship timeline, Peoplemag. <https://people.com/taylor-swift-and-travis-kelce-relationship-timeline-7974203>. Accessed: 2023-11-30.
- Kaufman, G. 2023. Taylor Swift Halloween costumes are surging in 2023, Billboard. <https://www.billboard.com/music/pop/taylor-swift-eras-to-ur-costumes-halloween-1235455171/>. Accessed: 2023-11-30.
- Pendleton, D. et al. 2023. Taylor Swift vaults into billionaire ranks with blockbuster eras tour, Bloomberg.com. <https://www.bloomberg.com/graphics/2023-taylor-swift-net-worth-billionaire/>. Accessed: 2023-11-30.