Taylor Swift in the Media

Jinwon Lee (261048866), Minna Feng (261051043),
Yi Jia Tan (261051377), Yu Tong Hu (261051311)
McGill University
jinwon.lee@mail.mcgill.ca, minna.feng@mail.mcgill.ca, yi.tan@mail.mcgill.ca, yu.tong.hu@mail.mcgill.ca

Introduction

Taylor Swift, a prominent figure in the music industry, has been subject to extensive media study throughout her career. In this analysis of Taylor Swift's media coverage, the focus is on presenting an unbiased examination of how she is represented across various articles from the collection of 500 articles. The objective of this data collection and analysis is to provide a comprehensive report on how she is portrayed in the media in North America. We explore the various topics covered and determine whether the overall tone of the coverage is positive, negative, or neutral.

Our analysis on Taylor Swift's image in the media revealed six distinct categories: her dating life, mentions of her in the diverse industries, her interaction with other celebrities, her musical career, her personal success, and her presence in media buzz and controversies.

We obtained three main categories that stood out the most. The first category is her dating life which covers about 32.8% of the articles collected. It includes her current and past relationships as well as any speculation about her romantic involvements. Based on the TF-IDF score, her relationship with Travis Kelce is revealed to be the most discussed topic. The second most popular category, constituting 19.4% of the articles, revolves around her mention in media buzz and controversies. This category involves viral moments online, fan-related events and the considerations of her recognition and influence as a pop icon. The TF-IDF score reveals that Dave Portnoy's opinions were mostly the highlight of many articles as well as Halloween which was part of the time zone in which the articles were collected. The third most discussed category, covering 19.2% of the articles, is about her music career. This includes her tour, her concert or concert movie, and any new music releases. The topic discussed the most revolves around her recent Eras tour, especially in Argentina and Brazil.

The analysis of article polarity indicates that a majority, specifically 63.2%, maintained a neutral tone. This suggests that her coverage in the media mostly opted for an objective and balanced stance. This neutrality reflects that journalists were aiming to provide facts without taking sides or expressing strong opinions to allow readers to make up their own minds about Taylor Swift and the topic being discussed.

Data

We acquired our dataset of Taylor Swift articles through the Mediastack API[1]. We opted for this API because the suggested NewsAPI[2] did not produce enough articles related to Taylor Swift. In fact, our initial attempt at a broad query for Taylor Swift, without specifying the "searchIn" parameter, returned a surplus of 500 articles. However, upon closer inspection, it became evident that most of these articles veered into unrelated territory, discussing subjects like NFL Fantasy Football Predictions or F1, with only a fleeting mention of Taylor Swift on one line. To refine our dataset and focus on more relevant content, we strategically employed the "searchIn" parameter, zeroing in on articles where Taylor Swift's name was explicitly mentioned in the title and/or description. Despite this effort to enhance relevance, the number of qualifying articles dwindled to less than 100. This circumstance prompted us to seek other news sources to augment our article count, which prompted us to choose the Mediastack API.

Once in this API, we had access to a variety of tools and articles at our disposal. Since we wanted to understand how Taylor Swift is currently being portrayed in the media, we decided to select articles that were published in the past month, ensuring that our analysis would be representative of recent conversations. Focusing on articles from the past month ensures that the analysis is based on the most recent and up-to-date information.

Additionally, we strategically emphasized the sorting of results based on popularity, with the aim of incorporating articles discussing the most trending topics related to Taylor Swift. Popularity often indicates high reader engagement. By selecting the top articles based on popularity, we were

[1] https://mediastack.com/
[2] https://newsapi.org/

most likely to obtain content that has generated significant attention and impact among readers. Also, prioritizing popular articles allows for a more efficient analysis, focusing on content that has already demonstrated its significance in the eyes of the audience.

We opted to narrow down our search to articles specifically published in the United States. This design choice was made because we wanted to focus our attention on a single country, given that there was such an abundance of relevant articles available. Furthermore, we only wanted to investigate how Taylor Swift is covered in North America for the project. Given that Taylor Swift is an American celebrity, we felt that it would be appropriate and would neither hinder nor skew the proportion topics that arose from the dataset. Limiting the analysis to U.S. publications ensures a more targeted examination of how she is perceived within her home country.

In addition, we also limited the language to English, given that we needed to be able to interpret and annotate the results ourselves, this facilitated a more accurate and nuanced interpretation of the content. English-language content is widely available on the internet, and many major news outlets publish in English. It also simplifies the preprocessing steps and reduces the complexity of language-related challenges. Lastly, considering that it is the most widely spoken language in the U.S., this language filter would help narrow down the results to those that are most widely accessed by Americans. All the above query parameters had the potential to introduce bias to our data collection, which will be discussed later.

Given that we needed 500 articles, we queried about 600 articles. The extra 100 articles acted as a buffer, as we expected certain entries to be unusable or removed. To clean the data, we conducted a thorough cleaning process, eliminating articles with duplicate titles. By exploring the data, we found that these articles were most likely reposts, and had identical content to each other. Additionally, we removed any articles with absent descriptions, since this field was crucial for our data annotation. This step included removing any articles where the description was a copy of the title. These steps resulted in a refined set of slightly over 500 articles. Subsequently, to focus on the top 500 most popular articles, we pruned the surplus entries from the bottom.

The starting code was found in the documentation page on Mediastack. Within the parameters, we selected "English" as the "languages," "us" as the "country," "Taylor Swift" as the keyword, "popularity" for sorting and "2023-10-28, 2023-11-28" for date. Since each page presented only a limited number of articles, we were able to expand the data, by using the "limit" and "offset" parameters which function as pages for articles. "limit" allows you to specify a pagination limit, which the maximum is 100. Meanwhile, "offset" allows to set a pagination offset value, determining where the results begin. For instance, an offset value of 100 combined with a limit value of 10 would display results 100 to 110. In the process of retrieving the data, we implemented a for-loop to navigate through the articles, specifically extracting the “title” and “description” of each article. Notably, we observed the presence of HTML entities within the article descriptions, necessitating the use of the html library in Python to decode them and ensure accurate representation. Following this step, the result was stored in a JSON file to facilitate organization. To clean the data, we executed an additional loop on the JSON file, checking for and eliminating any potential duplicate articles or any empty description articles.

Method

To create our typology, we began by performing an open coding on 200 of the articles by equally assigning each member 50 articles. For our first pass, we simply each annotate as we each saw fit, without prior discussion or influence. Our categories after the first pass included a wide range: dating life, controversy, NFL, song, party, collaboration, finance, music, concert, event, memes, criticism, politics, and success among others. However, many of those included very few articles or were too specific. We reviewed those categories, and discussed things like granularity, overlapping topics, gaps in the topics, and reviewed each other’s work. Categories like politics, parties, and memes only contained one or two articles, so we merged them to form more general categories. We agreed upon six topics that we had and used those to perform a second coding pass: dating life, NFL, media buzz and controversy, celebrity interactions, music, and accomplishments.

After the second coding, we realized we needed to refine and clarify topic definitions, as there were many edge cases. We realized there was an overlap between the topics of music and accomplishments, and explored the possibility of merging them, or splitting them in different ways. We settled on keeping them separate but making their definitions more explicit. We also renamed the NFL category to "mentions in other industries," as we wanted to make sure this typology would be more comprehensive to other industries. She is currently predominantly mentioned in NFL articles due to her relationship with her boyfriend, Travis Kelce, an NFL player for the Kansas City Chiefs (US Weekly, 2023). This change also made it more relevant and applicable to future industries she could be indirectly related to and made it clearer what this category was in relation to her. We also

clarified our understanding of a couple other topics, by reviewing tricky cases we had noted down for each of our sections. In case of disagreement, we discussed our separate points of view and were able to come to an agreement, which often required clarifying the typology as well. We finally performed a last pass to review the coding with the refined definitions. Once our topics were set, we coded the other 300 articles by separating them evenly among team members.

To implement our TF-IDF scoring mechanism, we used a Python script which processed a CSV file containing Taylor Swift-related articles, where we extracted the title and description, and their associated categories. The TF-IDF scores are calculated to gauge the importance of words within each category. The script begins by loading the CSV file, cleaning the text data by removing stopwords[3], punctuation, and non-alphabetic words, and then proceeds to compute TF-IDF scores for individual words. The scoring considers both term frequency (TF) and inverse document frequency (IDF), and the results are organized by our chosen typology. The top 10 words with the highest TF-IDF scores for each category are subsequently stored in a JSON file for further analysis.

When coding for polarity, we defined a positive, negative, or neutral article by its content. Initially, we wanted to define the polarity exclusively by the tone of the author of the article. This would include whether they used positive or negative connotation in the writing. However, by exploring the data through a first coding of polarity, we found that given Taylor Swift's favorable view in the media, most authors rarely used negative connotation to discuss her. We found that simply considering the tone of the article for polarity was not very insightful and representative of the actual conversations happening in the media. We then decided to consider the content of the article for the coding as well. This would mean that the articles where other public figures or celebrities were speaking negatively about Taylor Swift would get coded as negative. A neutral article was an article that reported information, without necessarily portraying Taylor Swift in a positive or negative point of view or expressing an opinion. A positive article used positive language and portrayed Taylor Swift in a very positive light.

We chose to compute the proportion of positive, negative, and neutral categories instead of scoring the articles from –1 to 1 where –1 represents negative, 0 represents neutral, and 1 represents positive. We aimed to distinguish whether the articles that we have collected were positive or negative or neutral towards Taylor Swift. Our objective did not require us to define the severity of the articles but only define how Taylor Swift is being covered in the media. Due to these facts, we only required the proportion of positive, negative, and neutral categories for the polarity of the articles.

Results

Our final chosen typology includes the following six categories: dating life, mentions in other industries, celebrity interactions, music, career success, and media buzz and controversies. The dating life category relates to her current or past dating life. This would include any reference to her current boyfriend, past boyfriends, or any speculation around who she is dating. This also included articles about Taylor Swift attending Travis Kelce's events and vice versa (i.e., Travis attending Taylor's Era Tour, Taylor attending Travis's NFL games) since it relates to their support to each other in their relationship. However, we did not include articles where she was merely mentioned in a broader context fully relating to her boyfriend. For instance, if her name was briefly mentioned in the title of an article about the NFL.

The next category – Taylor Swift's mentions in other industries – includes any articles where she is simply referred to in an industry separate from her own, and where the article mainly discusses events relating to that other industry. This category was important given that Taylor Swift's name is often mentioned in NFL-related headlines now that she is dating Travis Kelce. Many articles that fall into this category simply mention her in the headline and proceed to discuss NFL games for the remainder of the article, such as the scores, the plays, and future games. This category also includes how others have perceived Taylor and Travis's relationship to be affecting the games.

Next, the celebrity interactions category includes any in-person, or online direct interaction with Taylor Swift. This would include dinners with other celebrities, gifting between celebrities, and social media interaction between them. Often, these celebrities include her friends or other figures in the industry. It does not include other celebrities who talk about her and discuss her in the media.

The next category, music, is a big category which includes anything related to her songs, her concerts, her ongoing tour, or her concert movie. This does not include musical accomplishments, which would fall into the next category: career success. This category relates to accomplishments in her personal, musical, financial, or business life. As a very successful woman, her success in all aspects of her life is often discussed in the media. So, we chose to separate her

[3]https://gist.githubusercontent.com/larsyencken/1440509/raw/53273c6c202b35ef00194d06751d8ef630e53df2/stopwords.txt

musical success from her music when categorizing. This category often includes conversations about charts, breaking records, and Grammy nominations. In terms of financial success, it could include talk about her billionaire status for instance.

Finally, the last category is media buzz and controversies. This category includes other people discussing her in the media, any online or real-life controversy, and pop culture moments. Often, this includes various viral moments online, articles relating to fans, discussions about her from other public figures, or reference to her as a pop culture icon. It encompasses most discussion that does not fall into the other categories.

We have meticulously compiled articles across diverse categories, revealing a nuanced distribution of resources within each distinct subject area. In the realm of dating life, we amassed a total of 164 articles, constituting 32.8% of the overarching 500-article collection. Turning our attention to the mentions in other industries category, our efforts yielded approximately 51 articles, accounting for 10.2% of the total. Exploring the landscape of media buzz and controversies, we curated a compilation of 97 articles, representing a substantial 19.4%. For celebrity interactions, we secured 27 articles, comprising around 5.4% of the comprehensive dataset. Shifting our focus to the music category, our repository boasts 96 articles, making up 19.2% of the entire collection. Lastly, for career success, our compilation features 65 articles, constituting a notable 13%.

By performing most-frequent-word analysis using TF-IDF, we obtained the following top 10 most frequent words and their TF-IDF scores. For dating life, we obtained: kelce (264.53), travis (241.19), argentina (128.33), relationship (121.52), tour (115.11), chiefs (111.97), eras (110.13), kansas (106.51), city (106.38), and romance (90.71). For mentions in other industries, we obtained chiefs (132.06), eagles (122.31), broncos (78.66), kansas (68.35), city (67.84), vs (66.48), mahomes (59.14), nfl (55.09), patrick (53.1), and game (52.06). For the media buzz and controversies category, we obtained: portnoy (93.89), dave (82.15), halloween (50.86), world (49.54), ex (48.65), manning (44.23), viral (44.19), travis (43.72), simone (43.45), and post (42.1). For celebrity interactions, we obtained: brittany (86.22), mahomes (64.52), friendship (33.08), gift (30.7), kelly (30.7), clarkson (30.7), flowers (30.7), sends (28.97), friend (27.63), and gifts (25.61). For the music category, we obtained: tour (125.27), eras (123.08), concert (118.99), de (95.45), rio (79.72), fan (65.67), Janeiro (61.16), brazil (59.88), heat (57.25), and argentina (50.49). For the career success category, we obtained: album (95.73), version (90.71), top (90.58), hot (85.81), nominations (70.86), chart (69.34), awards (68.3), statue (64.47), grammy (64.25), and jung (61.92).

When analyzing the polarity of articles discussing her in the media, we found that 4.2% of articles were negative, 63.2% were neutral, and 32.6% were positive.

Discussion

The most popular words used in the dating life category all revolve to her current relationship or events surrounding her current relationship with Travis Kelce. The words "travis," "kelce," "chiefs," "kansas," and "city" most likely refer to the fact that she is currently dating a well-known NFL player for the Kansas City Chiefs. We infer that the presence of the words "argentina," "tour," and "eras" in the dating life category refer to her boyfriend's presence in the most recent leg of her Eras Tour. We can directly infer that a big part of the media coverage on Taylor Swift revolves around her boyfriend given that "kelce" and "travis" received the highest TF-IDF scores. Other words received lower scores given that they are also used in other contexts and in other categories. For instance, "chiefs," "kansas" and "city" were also seen in the NFL category. Given that this category was the most popular in terms of article count, we can also infer that her dating life is currently a very popular topic. If we had more data from a longer period, we would have been able to assess whether this was consistent throughout the years. However, we can conclude that Taylor Swift's current relationship is being very highly covered in the media.

The second most popular category that we observed in our data is media buzz and controversies. The most popular words are related to celebrity names, holidays, and social media. The words that we obtained the highest TF-IDF scores are "portnoy", "dave", and "halloween". "portnoy" and "dave" most likely refer to Dave Portnoy who is an American businessman (Barstool Sports, 2023). From this information, we can assume that he mentions Taylor Swift in various ways and discusses her in the media. Further, we assume the word "halloween" appears given the past month includes Halloween. Other words such as "world", "ex", "manning", "viral", "travis", "simone", and "post" most likely refer to certain posts are going viral in the media and references to other people discussing their relationship.

The third most popular category that we observed in our data is her music category. For her music, we could observe words that seemed to be related to her tour and how the weather affected her recent concert. The highest TF-IDF score words "tour," "eras," "concert," "de," "rio," "janeiro," "brazil," "argentina" most likely refer to the locations where the most recent leg of her tour was held for the past month. We could observe that it mostly mentions the countries in

South America. The words "fan" and "heat" most likely indicate that there was extreme heat during her concert and how her fans are affected by it. As we assumed that her concerts were in South America for the past month, we could observe a lot of countries in the area, which could certainly change if we had collected more data for a long period of time.

The fourth most popular category that we observed in our data is the career success category. For this category, we obtained the words "album," "version," "top," "hot," "chart," "nominations," "awards", and "grammy." These words likely refer to her songs and albums being on the charts or and nominations for Grammy awards. There were two outliers for this category which were "statue" and "jung". However, we speculated that "statue" could refer to her receiving a kind of statue, or being honored by a statue, given its presence in the career success category. We also speculated that "jung" likely refers to another artist in the industry who could be competing against Taylor Swift on the charts or for awards. Thus, we can imply that in terms of her career success, she is mostly discussed in terms of musical accomplishments.

The fifth most popular category is Taylor Swift's mentions in other industries. The most popular words were "chiefs", "eagles", "broncos", "kansas", "city", "vs", "nfl", and "games" which all refer to NFL teams or games. We can imply that she is currently highly relevant in the NFL sector, most likely due to her relationship with NFL player Travis Kelce. The bulk of the articles in this category referred to the NFL and simply mentioned Taylor Swift. It appears to mostly revolve around the Eagles, Chiefs, and Broncos NFL teams. These teams appear to be teams Travis Kelce has played against given the presence of the word "vs". We could also observe words mentioning NFL player Patrick Mahomes who plays in the same team as Travis Kelce. Thus, in terms of her mentions in other industries, Taylor Swift seems to be predominantly discussed in NFL-related articles.

The least popular category that was observed is celebrity interactions. For celebrity interactions category, we could observe different celebrity names. We obtained names like "brittany", "mahomes", "kelly", and "clarkson". The presence of these names implies that Taylor Swift likely had recent interactions with them both. Furthermore, this category contained words like "friendship", "friends", "gift", "gifts", "sends", and "flowers". It can imply that there were most likely gifts and flowers sent between Taylor Swift and other celebrities. The reason the words with highest TF-IDF scores are extremely specific and seem to revolve around a particular event or two is likely to since this is the least popular category, thus, the one with the lowest article count.

All in all, we could discover that most of the articles revolve around her dating life as it took 32.8% of our data. On top of that, given that the word "travis" was a high scoring word in other categories like the media buzz and controversies category, we can note that essentially more than 32.8% of our data was talking about their dating life directly and indirectly, making it an even more popular topic. Her career success and tour were followed as many of her songs were on the chart and recently held tour in South America. Most of these articles were positive or neutral in terms of polarity as most of the time they were delivering unbiased information about her music or concert or celebrating her relationship.

In terms of bias, the first place it could have been introduced is during data collection. By limiting our data collection to news articles published only in the USA, we might have introduced a form of regional bias. This bias arises from the fact that we focused exclusively on sources from a specific geographic location, in this case, the United States. In fact, it is entirely possible that the discourse in Canada, for example, is much more positive, or negative, depending on their view of Taylor Swift. The media landscape can vary significantly from one country to another, and U.S. media outlets may cater to a predominantly American audience. Consequently, the coverage may reflect the interests, values, and opinions of this specific audience, potentially leading to a biased representation of Taylor Swift's image. Additionally, Taylor Swift is a global figure, and her actions, events, or controversies may have different impacts and interpretations in various parts of the world. We are biased against global perspectives and reactions outside of North America.

In terms of language, limiting the articles to English could also be a potential cause of bias because we might miss out on cultural nuances and perspectives present in media outlets that publish in other languages. This could lead to a skewed understanding of how Taylor Swift is perceived globally. Limiting the analysis to English articles also excludes information and viewpoints from sources that publish in other languages.

Additionally, limiting articles to only those published in the last month might have biased us against more popular topics concerning events that happened earlier in the year. The analysis may be biased toward recent events and trends. If there are significant ongoing events, controversies, or positive developments related to Taylor Swift within the last month, the analysis might disproportionately emphasize these recent occurrences, overshadowing other aspects of her career. Furthermore, different media outlets have varying publication frequencies. Some may publish more frequently than others. By restricting the analysis to the last

month, we may inadvertently favor outlets with higher publication rates, potentially introducing bias.

By selecting the top 600 articles resulting from our API query sorted by popularity, we might have introduced bias in a couple ways. Popular content tends to get more visibility, leading to a potential echo chamber effect where widely shared or liked content becomes more popular simply due to its initial popularity. Also, content with sensational or provocative headlines may attract more clicks and shares, contributing to higher popularity metrics. This can lead to a bias toward sensationalized or clickbait content. Platforms use algorithms to determine and promote popular content. These algorithms may have their biases, favoring certain types of content or viewpoints, which can influence the popularity metrics. Popularity often reflects short-term trends. Sorting by popularity may bias the analysis toward content that is trendy or viral in the moment, potentially overlooking long-term topics.

During data annotation, bias could have been introduced by our own personal knowledge of Taylor Swift's career and personal life. Our own opinions, preferences, or preconceived notions about Taylor Swift could influence how we interpret and categorize content. Personal biases can result in selective attention to certain themes or topics that align with our interests or beliefs. This may lead to overemphasizing certain aspects of Taylor Swift's coverage while overlooking others. Moreover, determining whether an article is positive, negative, or neutral can be subjective. Personal biases might influence how we interpreted language, tone, or context, potentially leading to mischaracterizations of sentiment.

Group Member Contributions

In this collaborative project, each group member made substantial contributions, showcasing their unique skills. The entire team performed open coding on the 200 articles to determine the typology. We also completed coding for the rest of the articles, separating the workload evenly. We held frequent meetings with all members where we discussed, shared our thoughts, feedback, and planned the next steps.

In Yi Jia's case, he took the role in developing the core processing code, such as writing scripts to collect NewsAPI articles, to calculate the TF-IDF score, and the different percentages used throughout the project. Additionally, Yi Jia played a key role in cleaning the Mediastack articles obtained by Yu Tong using scripts, ensuring that the data was appropriately handled before annotation. In addition, Yi Jia contributed to the data and TF-IDF paragraph of the report.

Minna contributed many parts of the report. This mainly included the data section – discussing methods and reasoning behind data collection, the methods section – explaining the process behind building the typology and defining polarity, the results section – defining the chosen typology with examples, and finally the bias in the discussion section.

Jinwon contributed to the report in the discussion section where we had to analyze the TF-IDF scores in each category and discuss the popularity of categories in the data that we collected. Further, he contributed to how we defined polarity and the reasoning behind it.

Since we were unable to obtain sufficient articles through NewsAPI, Yu Tong contributed by going through the instructions and documentation of Mediastack API. Successfully implementing the code, Yu Tong managed to collect a dataset of over 8000 articles and stored over 500 articles for further analysis. In this report, she contributed to the introduction and the data section.

References

Barstool Sports. 2023. David Scott Portnoy "Dave". 2023. https://www.bloomberg.com/profile/person/21926633#xj4y7vzkg. Accessed: 2023-12-05.

US Weekly. 2023. Taylor Swift and Travis Kelce's Relationship Timeline: From Chiefs Games to Private Date Nights. https://www.usmagazine.com/celebrity-news/news/taylor-swift-and-travis-kelces-relationship-timeline/. Accessed: 2023-12-05.