Module 12 – Bachelor Thesis COM

“Public Opinion: The Secret to Celebrities Success?”

Varvara Paliura: s2861488

Supervisor: Dr. J.F. Gosselt

University of Twente, The Netherlands

1 July 2024

Table of Contents

Abstract **4**

1. Introduction **5**

1.1. Research Purpose 5

1.2. Research Objectives and Research Question 5

1.3. Academic and Practical Relevance 6

2. Theoretical Framework **7**

2.1. The Shift to Online Media and Its Impact on Public Opinion 7

2.2. Public Sentiment 9

2.3. User behavior in music consumption 10

3. Methodology **12**

3.1. Research Design 12

3.2. Corpus 13

Measures: Public Sentiment in News Media (Lexis Uni) *13*

Measures: Public Sentiment in Social Media (Reddit API) *14*

Measures: Playlists and Playlist Reach *14*

3.3. Procedure and Instruments 15

News media: BERT *15*

Social media: VADER *15*

3.4. Data Analysis Process 16

4. Results **17**

4.1 Media Sentiment 17

Social media sentiment per artist....................18

News Media Sentiment per Artist....................21

4.2. User Behaviour.................... 24

4.3. Correlation Analysis between Sentiments and User Behavior.................... 24

5. Discussion....................27

5.1. Discussion of main results....................27

5.2. Limitations of the Study.................... 28

5.3. Implications for Future Research....................28

5.4. Conclusion....................29

Reference list.................... 31

Abstract

Purpose: The purpose of this study was to investigate how public opinion expressed in the media might affect user behavior on streaming services. It aims to examine the relationship between sentiments expressed in social media and news media and behavioral outcomes such as music reception.

Methods: The study was operationalized by machine learning sentiment analysis models for News and Social media, and examined how sentiment affects user behavior on playlist metrics during the last five years. The playlist metrics, including the number of distinct Spotify playlists featuring the artists' music and the reach of these playlists, were analyzed. To determine specific comparative differences, post-hoc tests were performed after an ANOVA model was used to evaluate the artists' differences in sentiment scores. Correlation analysis was performed to investigate the connection between playlist metrics and sentiment ratings.

Findings: Despite the overall low correlation caused by significant variance in data per artist, the results revealed significant differences in sentiment scores among the artists. Male artists had a moderate to weak negative correlation between media sentiments and their playlist performance. In contrast, female artists did not show a consistent correlation between sentiment and playlist metrics.

Theoretical and practical contributions: This research contributes to the existing literature in media and user behavior studies with specifications on the hip-hop industry. It highlights different impacts on male and female artists, highlighting the nuanced interaction between media narratives and artist performance. Also, it opens new ways for building an artist's public image and suggests how an artist should be perceived in the media to achieve greater performance on streaming services.

Keywords: *Public Sentiment, Hip-Hop Artists, Sentiment Analysis, Media Influence, User Behaviour.*

1. Introduction

1.1. Research Purpose

People exchange opinions about celebrities on various channels, from traditional newspapers to social media, where opinions about them may quickly spread and shape public perception. As a consequence, public figures, especially in the music industry, are often subjects of intense public scrutiny. For example, Kim Kardashian and Kanye West's actions raised criticism from online users who felt that their daughter North West got the role in the Lion King anniversary concert not because of her talent, but because her parents are celebrities (Daily Express, 2024). Another example is Taylor Swift, who received severe criticism for having a significant carbon footprint, mostly as a result of using her own plane frequently (Our National Conversation, 2024). However, she mitigated this scrutiny after announcing her new album – "The Tortured Poets Department", which surprised her fans, and swayed the public attention to her career. These examples demonstrate that public figures' actions are closely monitored by the online community. But to what extent does public sentiment impact an artist’s career? Do celebrities get away with anything they say or how they act, or is there a lasting effect on their popularity and success? This is why it is crucial to understand what risks public opinion can pose for an artist’s career. This study examines how popular opinion toward celebrities in the media and on social networks affects behavioral outcomes, specifically streaming trends on Spotify.

1.2. Research Objectives and Research Question

The ratio between positive and negative perceptions and their effect on an artist's popularity will be examined. Doing so this study applies a two-step approach to analyze the potential connection between public opinion and celebrities' performance on Spotify. First, sentiment analysis will investigate how they drive public debate about artists in News and Social media. Secondly, the correlation between these sentiments in the media and an artist's popularity will be established to determine the connection between these two variables.

The central research question that guides this study is:

"How do public sentiments towards a celebrity, as expressed in news media and social media, relate to user behavior?"

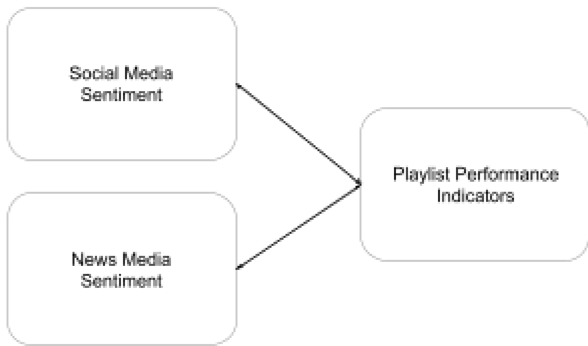

Figure 1

Graphical Representation of the Conceptual Model

1.3. Academic and Practical Relevance

The academic relevance of this research lies in exploring how users' behavior can be influenced by public sentiment online. There are already numerous papers investigating online media and how it affects user behavior; however, a very prominent industry – hip-hop music – has not been part of this research line. Furthermore, the findings of this study might offer significant contributions to sociocultural frameworks about popular culture and public opinion.

Nowadays public opinion may severely influence celebrity success, so it is critical to understand to what extent public discussion can affect user behavior. This study takes a mixed-methods approach, employing machine learning sentiment analysis techniques to assess public sentiment in both news and social media. It explores how different platforms, such as magazines and social media discussions, shape public opinion, influencing celebrity reputation and listener engagement.

Focusing on the hip-hop industry, this paper also explores a specific segment of popular culture. The insights from this research can serve as a base for marketing initiatives, partnership contracts, and PR campaigns, enabling the development of more successful tactics for increasing artists' audience reach on music streaming platforms. This involves creating targeted communication, developing strategic relationships with the public, and establishing PR efforts that increase positive media attention to improve streaming performance.

2. Theoretical Framework

2.1. The Shift to Online Media and Its Impact on Public Opinion

Over the last decade, we have seen a widespread transition from offline to online sources of information. Nowadays, almost every part of our lives can be done online, and offline sources are increasingly taking a secondary role. For instance, in recent years, there has been a huge drop in offline media consumption from newspapers to television viewership (Liu, 2016). This fact raises questions like why people choose online newspapers to print ones, how social media changed the way we receive and interact with information, and what is the relationship between news media and social media sentiments.

People tend to choose online news sources over printed newspapers due to higher accessibility, immediacy, and interactivity. Sixty-seven percent of the worldwide population has access to the internet and, consequently, access to information online (Statista, 2024). New generations prefer online sources for convenience and accessibility (Sung & Kim, 2020).

Immediacy is another feature of online sources, that allows a fast content reception, which helps users to always stay informed. Also, online news consumers feel present at events as if it was their own experience to a much greater extent than those who consume printed versions (Omar et al, 2020). In addition, the ability to interact with the news content, for example leaving a reaction or comment, increases user engagement with news and other online content (Oeldorf-Hirsch, & Sundar, 2015). In sum, the multitude of features offered by online media sources has completely changed the field of news consumption.

Social media offers an infinite amount of content across various platforms, covering all kinds of topics. Users can not only access this content but also discuss it and even create it themselves. The huge amount of user-generated content has led to new ways for people to interact with creators (Simon, 2016). This change has boosted the variety of information sources and has allowed users to be more involved with the content they find online. In addition, online platforms have expanded the type of content we distribute and consume by integrating not only text but also audio and visual content. This makes content more engaging and easy to comprehend.

The shift from offline to online media consumption changed not only the way information is received but also users' responses. This transition highlights the nature of news and social media, where journalists and opinion leaders play distinct, but complementary roles in creating the news agenda. Despite online journalists practicing the more audience-driven approach, they still feature more professional reports with in-depth analysis, in comparison to opinion leaders, who use innovative approaches to engage their audience with their thoughts and ideas (Steensen, 2009; Kelling & Thomas, 2018; Bamakan et al, 2019). While journalists report news and provide information often influenced by the audience's demands, opinion leaders actively shape and propagate narratives inside digital networks (Boatwright, 2022). Due to the online journalists' approach and the nature of opinion leaders, the public discourse is always shifting.

This interaction is always expressed with a certain sentiment in both news and social media. And because they often reflect and nudge each other to public discussion, there is a reason to assume the intricate relation between news media sentiment and social media reactions.

2.2. Public Sentiment

Public opinion is the aggregated result of individual opinions on certain topics. In this research, public sentiment and public opinion will be used as synonyms because both are referred to as a set of people's collective attitudes and feelings regarding a certain situation or a particular topic (Davison, 2024; Cambridge University Press, n.d.). These sentiments can set a public agenda, and shape narratives in the media and public discourse.

Recent studies have identified social media as a dual force that can both generate new sentiments and modify sentiments in news media (Ren et al., 2021). Social media generates new sentiments that news media report later. Another way around social media reacts to news articles changing the sentiment to a more positive or negative extent over time. This process indicates a two-way influence where each medium potentially affects the sentiment of the other. Complementary to this, the sentiment expressed by user comments in the news content often aligns with the sentiment expressed in the news media, suggesting that user engagement is affected by the tone set in the news media. (Kumar et al., 2018). Sentiments in news and social media are interrelated, often feeding into and off one another. By tracking these sentiments across platforms, one can affect the overall public mood and its potential impacts on a celebrity's reputation and success. This understanding is crucial for those managing public relations, brand strategy, and reputation management for public figures.

Sentiment analysis is an effective method for measuring public opinion quantitatively while highlighting emotional undertones in text. It can be applied to social media, product reviews, and public discourse, among other contexts (Cambria et al, 2022). Sentiment analysis methods can be divided into three categories: based on a pre-existing sentiment dictionary, traditional machine learning, and more complex deep learning techniques (Chaturvedi et al, 2018). With the rise of artificial intelligence, deep learning models are currently the most effective method to measure public sentiment (Rojas-Barahona, 2016). Machine Learning models automatically identify important features in sentences, removing the need for manual feature extraction.

2.3. User behavior in music consumption

Users' music consumption behavior can be shaped by many factors, including technological advances, personal preferences, and media influences. With the rise of digital platforms, the way people interact and share music has evolved a lot. Traditionally, ownership models, where users purchase individual tracks or albums in physical or digital formats, have dominated the music industry. However, as streaming platforms gained popularity, offering unlimited amounts of music for a fixed subscription fee, the user's music consumption behavior changed significantly. Nowadays users tend to consume more and more diverse music when they switch from ownership models to streaming services (Ferwerda & Tkalčič, 2019). This shift is driven by streaming services' accessibility, flexibility, and personalization features. The main concept of streaming services is to offer users a vast music library, which can accessed from almost any device. Also, the majority of platforms have the feature of downloading music inside an app, so users can listen to it offline.

Personal preferences also play a significant role in shaping users' music consumption behavior. According to Manolios et al (2019) personal values, such as "Openness to Change" and "Self-Enhancement," influence music preferences, leading users to explore diverse and complex music genres. For instance, individuals who value "Openness to Change" are more likely to enjoy diverse and complex music, while those who prioritize "Self-Enhancement" tend to select music that helps in achieving personal goals or relaxation. This indicates that personal values shape the type of music they prefer and consume. Furthermore, music is often used for self-expression and expansion of social connections (Kirk et al, 2016). The social influence encourages users to choose music that aligns with their social identity and the image they wish to project.

Additionally, the media can drastically affect music consumption through media platforms and content. For example, TikTok has become a major player in music promotion, where viral challenges and trends can bring songs to the top charts (Yang, 2023). An artist's popularity and audience engagement are heavily influenced by media coverage and public opinion, which can significantly impact their career. Positive media sentiment can boost streaming numbers and engagement, while negative sentiment can have the opposite effect. According to agenda-setting theory, the media plays a crucial role in shaping what audiences think about by highlighting certain topics or artists (Choi et al, 2016). This media coverage, whether positive or negative, can drive public interest and streaming activity (Verboord & Van Noord, 2016).

Framing theory further explains how the media's portrayal of an artist can influence public sentiment. Media outlets can shape the narrative by emphasizing specific aspects or sentiments, which affects how the audience perceives the artist (Watimin et al, 2023). For example, media coverage with a positive frame can enhance an artist’s reputation and audience engagement, while negative framing can lead to decreased interest in an artist’s work (Ruth, 2017). That is why, measuring the effect of public sentiment on user behavior can uncover how different types of media coverage influence streaming numbers. Understanding these factors allows artists and industry professionals to better navigate the media discourse, optimizing their communication strategies and engaging more effectively with their audiences.

3. Methodology

This research aimed to investigate the relationship between public sentiment toward celebrities and their success, reflected by their Spotify playlist reach per quarter. In this section, the research design, procedure, data collection process, and instruments will be discussed, that were used to investigate the research question.

3.1. Research Design

To answer the research question, firstly, a sentiment analysis for both News and Social media was conducted. Then, a user music consumption behavior analysis was employed. Lastly, the correlation analysis between media sentiments and user behavior was executed (see Figure 2). Given the nature of the research question, this study adopted a quantitative approach, using big data analysis techniques to examine media content.

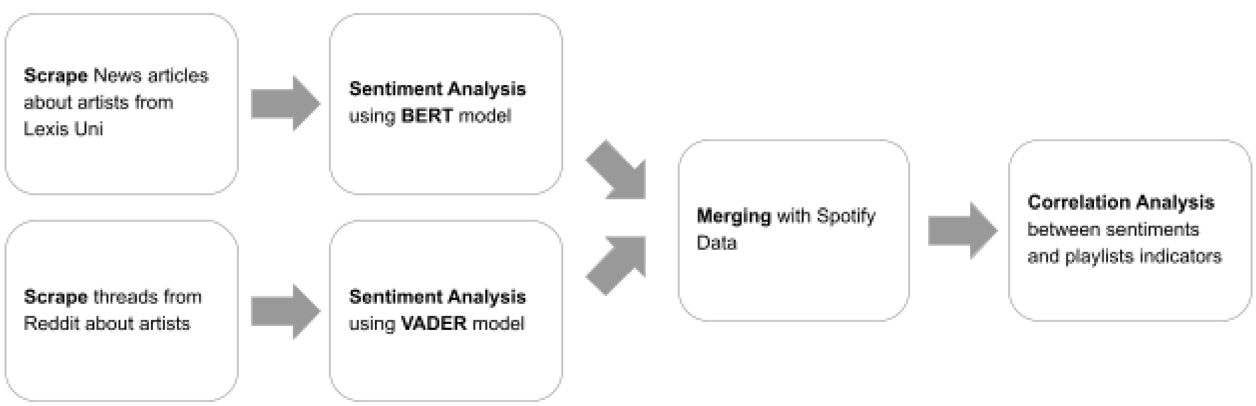

Figure 2 Study Design

The research was conducted in several phases: data collection and preparation, sentiment analysis, correlation analysis, and visualization of results. Firstly, data from news and social media sources was gathered within a specified time frame. Secondly, BERT for news media and VADER for social media were used to analyze the sentiment of the collected data. Then, correlations between public sentiment and Spotify data were identified.

3.2. Corpus

The corpus of this study consisted of public sentiment across *news media* and *social media* platforms. The news media source included major news outlets and reputable publications that frequently report on celebrities. As a main platform for the *News media*, the LexisUni was used. As the source for the *social media* data, Reddit was used. To guarantee a fair representation, this study focused on six celebrities from the American hip-hop industry, three of whom were male and three were female. The male chosen celebrities are Drake, Kanye West, and Travis Scott. The female celebrities include Cardi B, Nicki Minaj, and Doja Cat. These celebrities were picked because of their active social media presence and constant news mentions, which allowed to gather a considerable amount of data for the sentiment analysis.

Measures: Public Sentiment in News Media (Lexis Uni)

A purposive sampling approach was applied by selecting articles that mention one of the six specific celebrities. In addition, the articles were sorted by relevancy, and time filters were used to obtain the most relevant articles per quarter. The time range was from July 1, 2019, to March 31, 2024, based on the availability of Spotify playlist data in the open source. Ethical compliance was maintained by anonymizing user data and respecting platforms' terms of service. In total, 10,252 articles were scraped for six hip-hop artists. After the data was collected and cleaned from the duplicates and missing values, a new dataset was created with articles split into paragraphs, and the BERT model was applied to measure the sentiment for each paragraph. Then, for each article, the ratio of positive to negative sentiments was calculated. After that, the average ratio of positive and negative sentiments per quarter was identified before merging with Spotify data. This was made in order to understand the average sentiment per quarter for each artist in the News Media.

Measures: Public Sentiment in Social Media (Reddit API)

A combination of purposive and random sampling ensured a diverse range of social media content. The keywords for each artist were made to capture all events an artist participated in. For example, for Drake, the keywords were "Drake", "Drake new album", and "Drake OVO Sound", which captured his music career. For Travis Scott, the keywords were "Travis Scott", "Travis Scott Astroworld Festival", "Travis Scott Cactus Jack merchandise", and "Travis Scott new music". After establishing keywords, the time range was defined: from the year 2019 to 2024, in line with the news media time period. Then after setting these parameters, three to five Reddit threads were scraped for each keyword per quarter. Overall, 155,902 comments were collected from 314,144 Reddit threads. Then, the data was cleaned and the VADER model was applied for the sentiment analysis. The compound score is calculated automatically, which is an aggregated value of a comment's sentiment in a Reddit thread. The compound column ranged from -1 to 1 and was represented by the number with four decimal places.

Measures: Playlists and Playlist Reach

As a reflection of an artist's audience reach, two measures were included as the dependent variable in the data analysis: “Playlists” and “Playlist reach”. The “Playlists” represent the number of unique Spotify playlists with more than 200 subscribers, which include an artist's tracks on a certain day. The “Playlist reach” shows the number of followers of these “Playlists”. The main source of these measures was the Songstats app. This app aggregates data from streaming services and provides artists and music labels with valuable insights into performance. This paper used raw numbers provided by the service to perform a deeper analysis. All data collected for the analyzed celebrities was open-source.

3.3. Procedure and Instruments

The following section describes the specific steps and tools used to collect and analyze data, ensuring consistency and reliability.

News media: BERT

BERT is a deep learning model designed by Google. It uses a bidirectional transformer architecture to comprehend complicated text structures and context (Buche, 2020). A self-attention mechanism to capture allows BERT to understand the context from both the left and right directions in a sentence (Özçift et al, 2021). Because of this mechanism, BERT has an advantage in understanding the nuances and context of language. BERT is pre-trained on extensive texts like Wikipedia and books, enabling it to understand language patterns broadly (Devlin et al, 2018). After pre-training, BERT can be fine-tuned for specific tasks, such as sentiment analysis. As a result of its high adaptability, BERT works well when analyzing longer, more complex texts like newspapers, where context and syntax are essential.

Social media: VADER

In contrast, VADER assigns scores based on the emotional intensity of a predetermined list of words and phrases that are known as positive, negative, or neutral responses. Additionally, it takes into account elements that are typical of texts on social media, such as capitalization, punctuation, negations, and emoticons (Dhanalakshmi et al, 2023). VADER analyzes sentiment using predefined criteria. This particular model was selected because of its ability to manage the informal language commonly seen in social media posts.

3.4. Data Analysis Process

The data analysis was divided into several steps, from data cleaning to sentiment and correlation analyses. The first step was to clean up the datasets by removing errors, duplicates, and inconsistencies with missing or incomplete data. This assures that the analysis is performed on high-quality data. The timeframes were standardized. Dataset variables were parsed into correct data types suitable for correlation analysis, ensuring consistent formats across all datasets.

As stated above BERT was used for news media texts, and VADER was employed for social media texts. These tools extract the sentiment from the data, categorizing it into positive, negative, or neutral. After the initial analysis, sentiment scores were calculated to quantify the sentiment. Categorical data required required conversion to a format that can be fed to machine learning algorithms. One-Hot-Encoding has been chosen as it is a common solution. It assigns numerical values like -1, 0, and 1 to negative, neutral, and positive sentiments respectively, allowing quantitative analysis later. The sentiment scores were aggregated as an average sentiment score per quarter for each artist, to identify overarching trends in public sentiment. Analysis of Variance (ANOVA) was used to determine if there was a significant difference in Spotify data based on the type of sentiment. This helps identify patterns and trends in the data. Finally, the Pearson correlation analysis was used to assess the relationship between sentiment scores and playlist indicators. These methods helped identify the strength and direction of correlations.

4. Results

4.1 Media Sentiment

This section provides sentiment analysis data for both social and news media platforms. It begins with descriptive statistics to summarize the key sentiment measures, followed by an analysis of variance (ANOVA) to determine significant differences in sentiment among artists. Additionally, it includes post-hoc tests to explore specific differences in social media sentiment between artists. In addition, the sentiment distribution of *social* and *news media* is presented with insights into public opinion trends and variations across different artists.

Table 1

Descriptive Statistics for Sentiment Variables

Platform	Mean	Median	SD	Min	Max
News Media	0.53	0.53	0.12	0.27	0.86
Social Media	0.082	0	0.462	-1	1

Note Descriptive statistics include the mean, median, standard deviation (SD), minimum (Min), and maximum (Max) values for each variable

The *News Media Sentiment Average Proportion* was 0.53 (SD = 0.12), ranging from 0.27 to 0.86. The Social Media Average Compound Score was lower, 0.082 (SD = 0.462), ranging from -1.000 to 1.000.

Also, an analysis of variance (ANOVA) was conducted to examine if there are statistically significant differences in sentiment data among artists (see Table 2). It is crucial to know whether the means of the artists' sentiment proportions are significantly different because these differences may influence the interpretation of correlation results. The results of the ANOVA for *News Media* indicate a significant effect of artists on the *Sentiment Average*

Proportion: F(5, 89) = 19.49, p < .001. This suggests that the average positive proportion of sentiment scores varies significantly between artists. The results for *Social Media* sentiment indicate a significant effect of *artists* on the *Average Compound Score*, F(5, 89) = 4.59, p < .001. This shows that the average compound sentiment scores vary significantly between *artists*. As ANOVA analysis revealed, there are significant differences in sentiment scores among artists on both social and news media platforms. Thus, it was decided to execute correlation analysis separately per artist to prevent misleading results and prejudices.

Table 2

ANOVA for Average Positive Proportion and Average Compound

Source	Df	Sum of Squares		Mean Square		F Value		p-value	
		News	Social	News	Social	News	Social	News	Social
Artists	5	0.701	0.128	0.141	0.025	19.49	4.59	< .001	< .001
Residuals	89	0.64	0.495	0.007	0.006				

Note News = News Media Sentiment Average Proportion, Social = Social Media Average Compound Score, Df = Degrees of freedom.

Social media sentiment per artist

Tukey's HSD test analysis for the compound sentiment scores on social media revealed several significant differences among the artists (see Table 3). Notably, Doja Cat has a significantly higher compound sentiment score compared to Cardi B, with a mean difference of 0.12 (95% CI [0.04, 0.19], p < .001), indicating a more positive sentiment towards Doja Cat. Similarly, Drake has a significantly higher compound score compared to Cardi B, with a mean difference of 0.08 (95% CI [0.00, 0.16], p = .038), suggesting a more positive perception of Drake as well. In contrast, most other comparisons do not show statistically significant differences. For instance, the mean difference between Kanye West and Cardi B is 0.06 (95%

CI [-0.01, 0.13], p = .178), and the difference between Nicki Minaj and Cardi B is 0.04 (95% CI [-0.04, 0.12], p = .723), both of which are not significant. This pattern is consistent across other artist comparisons, such as Travis Scott versus Cardi B and Nicki Minaj versus Doja Cat, where the adjusted p-values indicate no significant differences. Overall, the data suggest that while Doja Cat and Drake are viewed more positively compared to Cardi B, sentiment scores among other artists do not differ significantly.

Table 3

Post-Hoc Tests for Compound Score for Social Media Sentiment

Comparison	Difference	Lower Bound	Upper Bound	Adjusted p-value
Doja Cat - Cardi B	.12	.04	0.19	< .001
Drake - Cardi B	.08	.00	0.16	.038
Kanye West - Cardi B	.06	-.01	0.13	.178
Nicki Minaj - Cardi B	.04	-.04	0.12	.723
Travis Scott - Cardi B	.06	-.01	0.14	.169
Drake - Doja Cat	-.04	-.12	0.04	.702
Kanye West - Doja Cat	-.06	-.13	0.02	.207
Nicki Minaj - Doja Cat	-.08	-.16	0.01	.092
Travis Scott - Doja Cat	-.06	-.13	0.02	.246
Kanye West - Drake	-.02	-.10	0.06	.974
Nicki Minaj - Drake	-.04	-.13	.05	.775
Travis Scott - Drake	-.02	-.10	.06	.982
Nicki Minaj - Kanye West	-.02	-.10	.06	.985
Travis Scott - Kanye West	.00	-.07	.07	.999
Travis Scott - Nicki Minaj	.02	-.06	.10	.980

Note Difference = mean difference between pairs of artists, Lower Bound and Upper Bound columns represent the 95% confidence intervals for these differences. The Adjusted p-value indicates the significance level after adjustment for multiple comparisons.

The Distribution of *Social Media* sentiment describes Reddit thread sentiment distribution in the period from 2019-07-01 to 2024-03-31 (see Figure 3). Doja Cat has the highest median across Social media content, however, the box is relatively short, indicating less variation in the sentiment distribution. Also, she has a few extremely high outliers, which can be described as a few highly positive comments on Reddit. Cardi B has the lowest compound scores, but the distribution is wide with the lowest median and outliers across all artists. Kanye West and Nicky Minaj have average scores, small and skewed distribution in comparison to other celebrities, indicating that the negative sentiment ratio is higher. Nicky Minaj also has a longer whisker pointing downward, confirming that there are more values below the median. Drake has a relatively wide distribution with an asymmetric box and low median, demonstrating that he has more positive comments on Reddit. Despite generally positive sentiment, one negative outlier indicates some threads with highly negative comments about him.

Summing up, Doja Cat has a generally positive perception among Reddit users, however, the relatively short box demonstrates less variation in sentiment. The public opinion about Cardi B on Reddit is highly polarized, with slightly positive and strongly negative sentiments. Kanye West and Nicki Minaj both show average scores but have skewed distributions, suggesting a higher proportion of negative sentiment. On the one hand, Doja Cat's sentiment includes a few highly positive outliers, suggesting occasional highly positive comments. On the other hand, Drake's sentiment distribution is generally positive but includes a significant negative outlier. This indicates that despite a generally positive public opinion, Drake also faces some criticism in the threads.

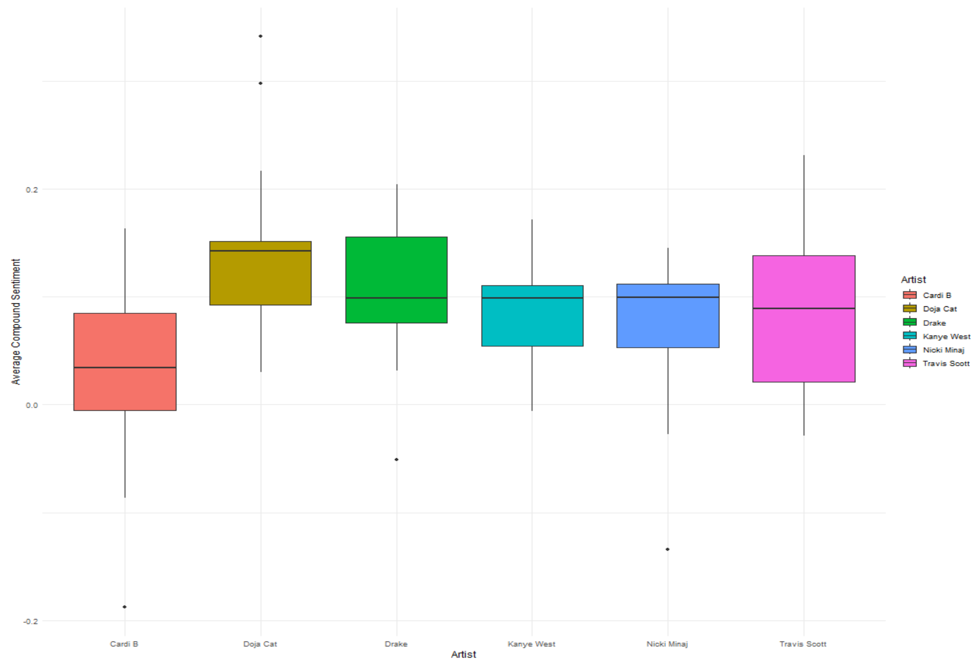

Figure 3. Box plot for Distribution of Compound Score of the sentiment for Reddit threads per Artist for the period from 2019-07-01 to 2024-03-31

News Media Sentiment per Artist

Tukey's HSD test analysis for the News Media Sentiment Average Proportion uncovered several significant differences among the artists (see Table 4). Doja Cat shows a notably higher positive proportion compared to Cardi B, with a mean difference of 0.22 (95% CI [0.13, 0.30], $p < .001$), indicating a more positive perception. Drake has a significantly lower positive proportion in comparison to Doja Cat (mean difference = -0.14, 95% CI [-0.22, -0.05], $p = 0.002$). Similarly, Kanye West shows a significantly lower positive proportion compared to Doja Cat (mean difference = -0.26, 95% CI [-0.34, -0.17], $p < .001$). Nicki Minaj also has a significantly lower positive proportion compared to Doja Cat (mean difference = -0.21, 95% CI [-0.30, -0.11], $p < .001$). Additionally, Travis Scott exhibits a significantly lower positive proportion compared to Doja Cat (mean difference = -0.15, 95% CI [-0.23, -0.06], $p < .001$). Kanye West has a significantly lower positive proportion than Drake (mean difference = -0.12, 95% CI [-0.21,

-0.03], p = 0.01). Furthermore, Travis Scott has a significantly higher positive proportion compared to Kanye West (mean difference = 0.11, 95% CI [0.02, 0.19], p = 0.01).

Table 4

Post-Hoc Tests for Positive Proportions for News Media Sentiment

Comparison	Difference	Lower Bound	Upper Bound	Adjusted p-value
Doja Cat - Cardi B	.22	.13	.30	< .001
Drake - Cardi B	.08	-.004	.17	.074
Kanye West - Cardi B	-.04	-.12	.05	.768
Nicki Minaj - Cardi B	.01	-.08	.11	.999
Travis Scott - Cardi B	.07	-.01	.15	.164
Drake - Doja Cat	-.14	-.22	-.05	.002
Kanye West - Doja Cat	-.26	-.34	-.17	< .001
Nicki Minaj - Doja Cat	-.21	-.30	-.11	< .001
Travis Scott - Doja Cat	-.15	-.23	-.06	< .001
Kanye West - Drake	-.12	-.21	-.03	.001
Nicki Minaj - Drake	-.07	-.17	.03	.306
Travis Scott - Drake	-.01	-.10	.07	.998
Nicki Minaj - Kanye West	.05	-.04	.15	.622
Travis Scott - Kanye West	.11	.02	.19	.004
Travis Scott - Nicki Minaj	.06	-.04	.15	.504

Note Difference = mean difference between pairs of artists, Lower Bound and Upper Bound columns represent the 95% confidence intervals for these differences. The Adjusted p-value indicates the significance level after adjustment for multiple comparisons.

Figure 4 shows the *News media sentiment* distribution for news media mentions of all six artists from 2019-07-01 to 2024-03-31. The boxplot shows that Doja Cat has the highest proportion of positive mentions in news articles, in contrast to other celebrities. She also has a symmetric boxplot, which means the positive sentiment proportion is normally distributed. The whiskers of the Cardi B box are quite long, indicating a wider range of data. Cardi B and Nicky Minaj have approximately the same distributed sentiments with only one visible difference –

Nicky Minaj's median of the sentiment proportion is 0.05 higher, which is not a big difference, but still, it is visible that Cardi B is talked about more negatively in the news than Nicky Minaj. Kanye West has the smallest positive sentiment proportion among all six celebrities. There are two artists with some outliers – Drake and Travis Scott. However, for Drake, the outlier is above 0.7, whereas Travis Scott's outlier is lower – around 0.35. Also, Travis's Scott box upper whisker is the shortest indicating that the positive sentiment proportion distribution has values above the median than expected.

In conclusion, Doja Cat is the most positively viewed artist in the news media, while Kanye West has the most negative sentiment. Cardi B has the widest range of sentiment, suggesting polarizing public opinion in the news articles. Nicki Minaj and Drake have relatively stable positive sentiment proportions, indicating that they are portrayed in the news consistently. Both Drake and Travis Scott have notable outliers, indicating occasional extreme sentiments in the news.

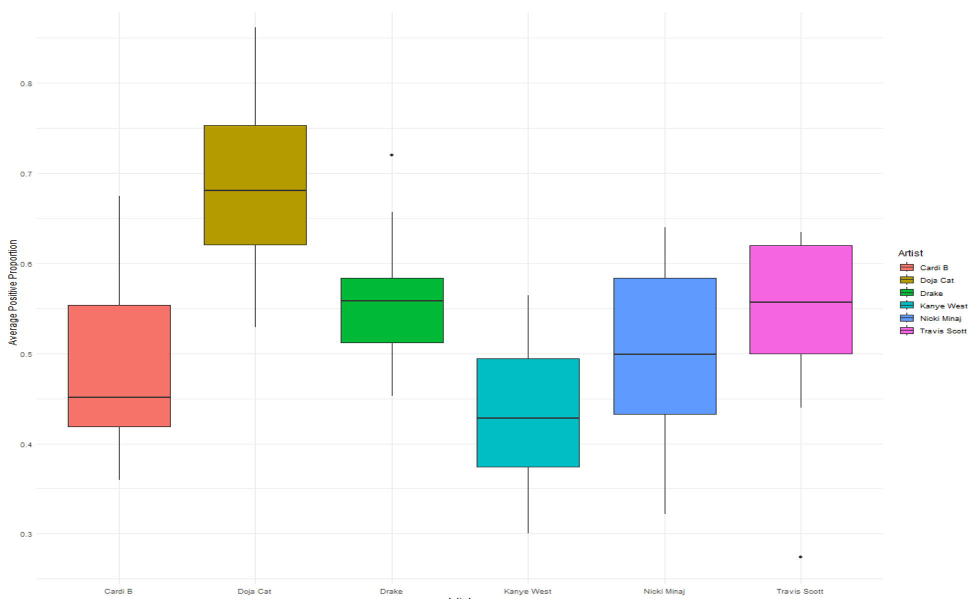

Figure 4. Box plot for Distribution of Positive sentiment proportion for News Media per Artist for the period from 2019-07-01 to 2024-03-31

4.2. User Behaviour

User behavior is reflected in Spotify playlist indicators, namely the number of playlists and their reach measured in the number of followers. This section will describe the key measures to give a fair representation of the data frame used for the analysis.

Table 5

Descriptive Statistics for Playlist Variables

Variable	Mean	Median	SD	Min	Max
Playlists	46,763	43,879	25,188	4,476	106,623
Playlist Reach	458,196,100	443,983,100	127,323,000	230,780,600	721,383,800

Note Descriptive statistics include the mean, median, standard deviation (SD), minimum (Min), and maximum (Max) values for each variable.

Table 5 shows that the number of playlists between 2019-07-01 and 2024-03-31 was 46,763 (SD = 25,188), ranging from 4,476 to 106,623. The average number of followers of these playlists was 458,196,100 (SD = 127,323,000), with a minimum reach of 230,780,600 and a maximum reach of 721,383,800. The descriptive statistics for playlist indicators illustrate significant variability in the number of playlists and their reach, reflecting diverse user engagement on Spotify.

4.3. Correlation Analysis between Sentiments and User Behavior

Due to the significant differences in artists' sentiments revealed by Tukey's HSD test analysis, a correlation analysis was performed for each artist individually to ensure accurate results. The aggregated sentiment data from news and social media sources and the Spotify playlist performance were used for the correlation analysis. The correlation matrix visually represents the relationship between four variables: Playlists, Playlist Reach, and average sentiment score of social media and news media (see Table 6).

A strong positive correlation was found between the number of playlists and the number of followers (playlist reach) for the majority of artists except Cardi B and Travis Scott, who have a moderate correlation of 0.46 and 0.58 respectively. This could be interpreted as that other factors influence their playlist reach or their fanbase could interact with their music differently compared to other artists.

As for the correlation between the positive sentiment proportion of the *news media* and the number of playlists, the situation is twofold. All male artists have moderate to weak negative correlations, showing that for them, having a higher positive sentiment proportion does not necessarily mean their music will be on more playlists. Conversely, more playlists do not mean more positive news sentiment. At the same time for female celebrities, the correlations are highly differentiated. Nicki Minaj has a moderate positive correlation (0.65), demonstrating higher positive sentiment proportion correlates with a higher number of playlist appearances. Doja Cat, in contrast, has a moderate negative correlation (-0.60), indicating that with the rise in the number of playlists, her sentiment becomes more negative. At the same time, Cardi B does not have a correlation between positive sentiment proportion and the number of playlists at all.

The *Social media* sentiment correlation with the playlists number has less strong results in comparison with the News media sentiments because the compound score was more homogeneous (See Figure 3). For all male hip-hop artists, the correlation between Reddit sentiment and the number of playlists was again weak to moderately negative. Doja Cat has no correlation between playlists and Reddit sentiment. Cardi B and Nicky Minaj have a weak positive correlation for the same variables (0.30 and 0.33).

Table 6

Correlation Matrix of Playlist Metrics and Sentiment Scores for Artists

Gender	Artist	Playlists	Playlist Reach	News Media	Social Media
		1	.88	-.42	-.10
		.88	1	-.22	-.09
		-.42	-.22	1	.19
	Drake	-.10	-.09	.19	1
		1	.84	-.38	-.50
Males		.84	1	-.31	-.47
		-.38	-.31	1	.39
	Kanye West	-.50	-.47	.39	1
		1	.58	-.27	-.68
		.58	1	.12	-.12
		-.27	.12	1	.71
	Travis Scott	-.68	-.12	.71	1
		1	.46	0	.30
		.46	1	-.20	.30
		0	-.29	1	.58
	Cardi B	.30	.30	.54	1
		1	.81	.65	.33
Females		.81	1	.73	.33
		.65	.73	1	.54
	Nicki Minaj	.33	.33	.54	1
		1	.82	-.60	0
		.82	1	-.51	.08
		-.60	-.51	1	.28
	Doja Cat	0	.09	.28	1

Note News Media = Average Sentiment Proportion for News Media per artist, Social Media= Average Sentiment Proportion for Social Media per artist

5. Discussion

This section will focus on discussing the results of this study, highlighting the limitations of the study and recommendations for future research.

5.1. Discussion of main results

Due to the specifics of the corpus and the artists' sample size, the analysis did not show significant correlations between sentiments in the media and user behavior. The connection between media sentiment and user behavior may be more complex than initially assumed, demanding the implementation of advanced models or additional factors to be included in the analysis. The large variance in the artists' data variables led to a focus on clarifying patterns between media sentiment and public behaviors regarding the analyzed artists.

The correlation analysis shows some differences between male and female artists. While all male artists have negative correlations of sentiment scores with playlist indicators, female artists have different cases where the correlation can be positive, negative, or no correlation at all. That is to say, Travis Scott, Drake, and Kanye West demonstrate the low impact of public sentiment on their performance on streaming platforms. In comparison, female artists' results differ a lot. Despite Doja Cat's high positive sentiment across all media sources, she has a weak negative correlation between the playlist indicators and sentiment scores. Nicky Minaj has a moderate positive correlation, which could be explained by her neutral sentiment scores in News and Social Media. Cardi B is the only artist who showed no correlation between playlist indicators and sentiment scores.

Despite the generally low correlation between sentiments and user behavior, the study highlights significant gender differences, revealing distinct patterns between male and female celebrities in how media sentiment impacts user behavior. The male artists' audience is mostly indifferent to public sentiment, as indicated by the negative correlation between sentiment scores and playlist indicators. In contrast, female artists' fanbase seems to have a more varied impact, suggesting other potential variables that have a greater impact on user behavior besides media sentiment.

5.2. Limitations of the Study

In regards to the limitation of this study, there could be some inconsistencies in the data collection method and the type of analysis fit, that could make the results differ from the real picture. Big data methods were used to analyze the sentiments from News and Social media. A big array of news articles and comments from Reddit were included, which were scraped per quarter for each celebrity. Based on the data collection methods some texts may not have been included due to sorting and filtering specifics. Data from both media sources was scraped using keywords and sorted by relevance identified by Lexis Uni for news articles and by Reddit for comments. That is why some important pieces of information were possibly missed. For example, news articles that were not presented on Lexis Uni, but have strong sentiments expressed about the artists, or smaller Reddit threads that do not have a lot of comments.

In the context of correlation analysis, different artists showed different results, so it is impossible to say that there is a relation between public sentiment toward artists and their performance on streaming platforms. In this case, a wider artist sample size is required to show a bigger picture, and general trend across all hip-hop artists.

5.3. Implications for Future Research

For future research, it is important to consider other potential factors that could influence user behavior because outcomes can be greatly impacted by outside factors including public music taste shifts, marketing campaigns, or market trends. Also, incorporating more data sources for social media and streaming platforms could benefit the fair representation of the population, for example, TikTok or Twitter for social media and Apple or YouTube music for the streaming platforms. Additionally, it is crucial to see how these correlations change over time, so conducting long-term studies can offer more valuable insights into trends and patterns. Besides that, the impact of different events can be studied, and the extent to which they can affect both sentiment in the media and playlist indicators. To expand the research scope, comparing sentiment and performance across different genres of music can reveal genre-specific trends and factors that influence success.

5.4. Conclusion

The central research question focused on understanding the correlation between sentiment in News and Social media and playlist performance indicators for six hip-hop artists. The results of this research demonstrated distinct patterns between male and female artists.

For male artists, there is a negative correlation between positive sentiment and playlist performance, which indicates that public sentiment, whether positive or negative, does not significantly drive their playlist success. This tendency suggests that, regardless of sentiment, factors like media hype or other PR tactics may contribute to their continued visibility and interaction on streaming platforms. In contrast, female artists displayed more varied correlations, and it is not possible to define a clear trend. These results highlight how intricate the connection between streaming performance and public opinion could be.

The understanding of how public perception can affect music artists and its causes on behavioral outcomes such as music reception may contribute to the academic discourse on media influence. Also, this research can be informative for sociocultural theories in the context of public opinion and popular culture. As for the practical side, this paper can be a basis for the predictive model development for consumer behavior, especially for predicting cultural trends. Whatever the sentiment, male artists may find it essential to take advantage of media hype and keep themselves visible. Strategies focused on improving positive public opinion could be beneficial for female artists. Still, positive sentiment across media may be leveraged, and negative sentiment can be minimized by using many platforms to interact with followers.

Reference list

Bamakan, S. M. H., Nurgaliev, I., & Qu, Q. (2019). Opinion leader detection: A methodological review. *Expert Systems With Applications, 115*, 200-222. https://doi.org/10.1016/j.eswa.2018.07.069

Boatwright, B. C. (2022). Exploring online opinion leadership in the network paradigm: An analysis of influential users on Twitter shaping conversations around anthem protests by prominent athletes. *Public Relations Review, 48(4)*, 102229. https://doi.org/10.1016/j.pubrev.2022.102229

Buche, A. (2020). BERT for opinion mining and sentiment farming. *Bioscience Biotechnology Research Communications, 13(14)*, 35-39. https://doi.org/10.21786/BBRC/13.14/9

Cambria, E., Kumar, A., Al-Ayyoub, M., & Howard, N. (2022). Guest editorial: Explainable artificial intelligence for sentiment analysis. *Knowledge-Based Systems, 238, 107920.* https://doi.org/10.1016/j.knosys.2021.107920

Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion, 44, 65-77.* https://doi.org/10.1016/j.inffus.2017.12.006

Choi, J., Han, Y., & Kim, Y. (2016). A research for finding relationship between mass media and social media based on agenda setting theory. In R. Lee (Ed.), *Software Engineering Research, Management and Applications 654, 102-113.* Springer. https://doi.org/10.1007/978-3-319-33903-0_8

Daily Express. (2024). Kim Kardashian faces backlash over North West's role as Simba amid nepotism claims. Retrieved from https://www.the-express.com/entertainment/celebrity-news/138842/Kim-Kardashian-North-West-role-simba-nepo-baby-claims

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

Dhanalakshmi, P., Kumar, G., Satwik, B., Sreeranga, K., Sai, A., & Jashwanth, G. (2023). Sentiment analysis using VADER and logistic regression techniques. *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), 139-144.* https://doi.org/10.1109/ICISCoIS56541.2023.10100565

Kelling, K., & Thomas, R. (2018). The roles and functions of opinion journalists. Newspaper Research Journal, 39, 398 - 419. https://doi.org/10.1177/0739532918806899

Kumar, N., Nagalla, R., Marwah, T., & Singh, M. (2018). Sentiment dynamics in social media news channels. *Online Social Networks and Media, 8, 42-54.* https://doi.org/10.1016/j.osnem.2018.10.004

Liu, F. (2016). Declining News Media Viewership and the Survival of Political Disagreement. *International Journal of Public Opinion Research, 29, 240-268.* https://doi.org/10.1093/IJPOR/EDW004

Oeldorf-Hirsch, A., & Sundar, S. S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior, 44, 240-249.* https://doi.org/10.1016/j.chb.2014.11.024

Omar, B., Al-Samarraie, H., & Wright, B. (2020). Immediacy as news experience: exploring its multiple dimensions in print and online contexts. *Online Inf. Rev., 45, 461-480.* https://doi.org/10.1108/oir-12-2019-0388/v2/response1

Our National Conversation. (2024). Taylor Swift and her private jet: Celebrity carbon footprint and privilege. Retrieved from https://www.ournationalconversation.org/post/taylor-swift-and-her-private-jet-celebrity-carbon-footprint-and-privilege

Ren, J., Dong, H., Padmanabhan, B., & Nickerson, J. V. (2021). How does social media sentiment impact mass media sentiment? A study of news in the financial markets. *Journal of the Association for Information Science and Technology, 72(10), 1213–1229.* https://doi.org/10.1002/asi.24477

Rojas-Barahona, L. M. (2016). Deep learning for sentiment analysis. *Language and Linguistics Compass, 10(12), 701-719.* https://doi.org/10.1111/lnc3.12228

Ruth, N. (2017). They don't really care: Effects of music with prosocial content and corresponding media coverage on prosocial behavior. *Musicae Scientiae, 22(3),* 415-433. https://doi.org/10.1177/1029864917716735

Simon, J. (2016). User generated content – users, community of users and firms: toward new sources of co-innovation?. *Info, 18, 4-25.* https://doi.org/10.1108/INFO-04-2016-0015

Songstats. (n.d.). Spotify analytics for artists & labels. Retrieved June 9, 2024, from https://songstats.com/platforms/spotify

Statista. (2024). Number of internet and social media users worldwide as of April 2024 (in billions). Retrieved from https://www.statista.com/statistics/617136/digital-population-worldwide/#:~:text=As%20of%20April%202024%2C%20there,population%2C%20were%20social%20media%20users.

Steensen, S. (2009). The shaping of an online feature journalist. *Journalism, 10, 702 - 718.* https://doi.org/10.1177/1464884909106540

Sung, N., & Kim, J. (2020). Does the internet kill newspapers? The case of South Korea. *Telecommunications Policy, 44, 101955*. https://doi.org/10.1016/j.telpol.2020.101955

Verboord, M., & Noord, S. (2016). The online place of popular music: Exploring the impact of geography and social media on pop artists' mainstream media attention. *Popular Communication, 14, 59 - 72.* https://doi.org/10.1080/15405702.2015.1019073.

Watimin, N. H., Zanuddin, H., Rahamad, M. S., & Yadegaridehkordi, E. (2023). Content framing role on public sentiment formation for pre-crisis detection on sensitive issue via sentiment analysis and content analysis. *PLOS ONE, 18(10)*. https://doi.org/10.1371/journal.pone.0287367

Yang, H. (2023). Research on Music Advertising in TikTok under a Systematic View. *Advances in Education, Humanities and Social Science Research*, *4*(1), 381. https://doi.org/10.56028/aehssr.4.1.381.2023