The Effect of Twitter Cancel Culture on the Music Industry

Nikitha Gopal, Abigail Velasquez, Peter Wu

March 10, 2022

1 Abstract

Musicians often trend on social media for various reasons but in recent years, there has been a rise in musicians being "canceled" for committing offensive or socially unacceptable behavior. Due to the wide accessibility of social media, the masses are able to hold accountable musicians for their actions through "cancel culture", a form of modern ostracism. Twitter has become a well-known platform for "cancel culture" as users can easily spread hashtags and see what's trending, which also has the potential to facilitate the spread of toxicity. We analyze how public sentiment towards canceled musicians on Twitter changes in respect to the type of issue they were canceled for, their background, and the strength of their parasocial relationship with their fans. Through our research, we aim to determine whether cancel culture on Twitter could lead to an increase in toxicity towards a canceled individual.

2 Introduction

Every so often, a public figure or celebrity would trend on Twitter over behavior that the general public would consider offensive or socially unacceptable by today's standards. From politics to Hollywood to the music industry, all kinds of public figures have been called out and criticized for such behavior by the masses. This form of modern ostracism is known as "cancel culture", in which the masses attempt to call out and boycott an individual in response to problematic remarks or ideologies that they have expressed. This online phenomenon has resulted in celebrities facing real life repercussions. For example, world renown Harry Potter author, JK Rowling was 'cancelled' for making controversial remarks about transgender individuals. Due to public uproar, HBO did not extend an invite to Rowling for the Harry Potter reunion. The event was marking the 20th anniversary of the franchise she created. Another example is R. Kelly, an American singer and songwriter, who was accused for sexually assualting minors. In fact, R. Kelly was arrested for 10 counts for sexual asualting four women. The hashtag "#MuteRKelly", however, began trending on Twitter when Survingn R. Kelly, an investigative series began airing on the US network Lifetime. The hashtag was pushing to have R. Kelly songs removed from radio and music applications.

3 Problem Statement and Objectives

To investigate how cancel culture affects the overall public perception of musicians on Twitter, our main research question is the following: among English-speaking Twitter users, how does their sentiment towards canceled musicians change over time? As a result of getting canceled, we believe that there will be an overall increase in toxicity and negativity in terms of public sentiment on Twitter towards the canceled individual. To understand the nuances and quantify the effect of

cancel culture on public figures, we will be conducting case studies on six individuals from three different genres within the music industry, using Twitter as our main source and platform. To assist our main investigation, we ask three sub-research questions:

- **RQ1:** How does sentiment differ over time based on the type of issue the individual was canceled for?
- **RQ2:** How does sentiment differ based on the individual's gender and the genre they cater to?
- **RQ3:** How does sentiment differ for individuals with and without fans that have a strong parasocial connection to the individual?

4 Literature Review

5 Methods

5.1 Data Collection

We collected data from three different music genres: K-pop, Hip-Hop, and Pop. For each music genre, we chose one male and one female music artist who has gotten canceled as well as one male and female artist who have not had any controversies that they got canceled for to act as our control group. Each individual that we have chosen also has a Twitter account that is run by themselves and/or their social media team.

For K-Pop, our two canceled individuals were NCT and WayV's Lucas and Aespa's Giselle. After allegations from former girlfriends of gaslighting and cheating, Lucas was called out and apologized for such behavior two days after the allegations appeared [1]. As of February 2022, Lucas still remains on hiatus from participating in any activities. Giselle was called out on Twitter for lip-syncing a racial slur from American R&B artist SZA's track "Love Galore" in a behind-the-scenes video [2]. Giselle also apologized a few days after the incident occured. As our control individuals for the K-Pop genre, we chose NCT and NCT Dream's Jaemin and ITZY's Ryujin. In choosing our control individuals for this genre, we chose individuals that were part of groups that debuted or were formed within the same time frame as the groups that the canceled individuals are part of and their respective groups have at least over a million Twitter followers. It is also important to note that although Lucas and Jaemin are both members of boy group NCT, they work under different subunits or groups as Lucas is a member of WayV and Jaemin is a member of NCT Dream. This means that they are not directly associated with each other as they do not work together on a regular basis. Therefore, it is highly unlikely that the reputation of one would affect the other's reputation.

The next group of artists we analyzed were Western hip hop artists. In this group, the two canceled artists were DaBaby and Nicki Minaj. During a festival concert in July of 2021, DaBaby made some remarks to the crowd that were homophobic. After that concert, the controversy went mainstream while many festivals and brands dropped DaBaby [3]. Nicki Minaj faced backlash after she tweeted vaccine misinformation after being asked her vaccination status [4]. This controversy occurred in September of 2021 while the Covid-19 pandemic was still affecting many around the world. For the control artists in the hip hop group, we chose Lil Baby and Saweetie. The main criteria for these artists were if they had a similar following or presence as the canceled individuals and didn't experience cancellation from controversies. Lil Baby and DaBaby have a similar number of Twitter followers at 6.7 and 4.4 million. However, Nicki Minaj is the most followed female hip hop artist on

Twitter by a large margin at 23.7 million followers. For my choice for the other control artist, we chose Saweetie as we felt she has cultivated a similar following to Nicki Minaj through her music and online presence.

The third group of artists we analyzed were Western pop artists. Two individuals who faced controversy in this genre were Doja Cat and Zayn Malik. Around the end of May 2020, the hashtag "#DojaCatIsOverParty" began circulating on Twitter since Doja Cat's 2015 song "Dindu Nuffin" resurfaced [5]. The term, "Dindu Nuffin", is recognized as 'a racial slur used to mock victims of police brutality' [5]. Following this, more racist allegations emerged suggesting that Doja Cat was contributing to racist conservations on online chat rooms years prior. In response to the controversy, Doja Cat apologized and explained the reasoning behind her actions in an instagram post, which is now deleted. Zayn Malik recently faced some backlash due to physical assault allegations made by his mother-in-law, Yolanda Hadid. On October 28 2021, TMZ reported that Yolanda Hadid filed a police report against Zayn for domestic violence [6]. As a result, Zayn was required to pay fines, complete 90 day probation, and attend anger management classes [6]. For the Western Pop control artists, we have chosen Adele and Harry Styles. As mentioned above, the main criteria for our control artists is to have a similar following or presence to our canceled artists. Adele is a strong female pop artist with 27.2 million followers on Twitter. Although her career has been longer and her following is significantly greater than Doja Cat's 4.2 M followers, both artists hold a similar presence in the media through their music today. Both Doja and Adele have given chart topper hits in this year, making them both equally relevant in the current pop industry. In regards to Harry Styles and Zayn Malik, both have a similar number of followers on Twitter (Harry Styles - 37.4 M and Zayn Malik - 31 M). In addition, their industry presence is also identical since they started their careers together in One Direction. Moreover, Harry and Zayn were both equally influential and popular artists through their independent music post One Direction, making them ideal complements.

We used the Twitter API to scrape tweets from Twitter relating to each individual. In order to see if there is a visible change in public sentiment, we decided to collect tweets six months before and after the date of cancellation. For each artist, we generated a list of related hashtags and keywords that can be used to collect tweets relating to the artist and the controversy. For each tweet, we gathered the date the tweet was created, author id, tweet id, and the text itself. Below we have presented a table displaying the hashtags and keywords we included for each query along with the date of cancellation and the number of initial tweets we collated.

Table 1. Canceled Artists' Dataset Overview

Artist Name	Date of Cancellation	Initial Number of Tweets	Hashtags + Keywords Used for Query
Lucas (NCT/WayV)	8/24/21	117,654	"lucas", "bubble","bbl", "scandal", "xuxi", "yukhei", "nct","czennie", "czennies", "nctzen", "lumis", "weishennie", "nctzens", "weishennies", "wayv", "weishenv", "ot23", "ot22", "ot6", "ot7", "#NCT", "#LUCAS", "#WayV", "WeiShenV", "#WELOVEY-OULUCAS", "#LUCAS_OUT", "#WAYV_is_7", "#LUCASBEST-BOY", "#WayVis7", "#LU-CAS_GETS_OUT_OF_NCT", "#LUCAS_GETS_OUT_OF_WayV"
Giselle (Aespa)	10/23/21	36,176	"giselle", "uchinaga", "kpop", "k-pop", "bbl", "bubble", "MYs", "MY", "aeri", "n word", "racist", "aespa", "#aespa", "#gisellenword", "giselleaespa", "#Giselle_OUT", "#gisellen-wordspace", "#giselleapologize", "#aespa_is_FOUR"
DaBaby	7/25/21	100,810	"dababy", #DaBaby-isOverParty, #DaBookings, #DaApology, #Dababy

Artist Name	Date of Cancellation	Initial Number of Tweets	Hashtags + Keywords Used for Query
Nicki Minaj	9/13/21	93,332	'"nicki minaj, #NickiMinaj, #Nicki, #MyCousinTookTheVaccine, #NickiMinajsCousinsFriendsBalls, #NickisCousinsFriendsBalls, #nickiminajdraggingparty, #iStandWithNicki, #istandwithnickiminaj, #barbz, barbz
Zayn Malik	10/28/2021	53,472	"zayn malik", "zayn", "zayn hits yolanda", "zayn and gigi", "zayn malik and gigi hadid", "zayn cheated", #WeAreWithZayn, #ZaynMalik, #FreeZayn, #welovezayn, #AlwaysWithYouZayn, #respectzayn, #WeSupportZayn, #cancelzayn, #WeHateZayn
Doja Cat	5/25/2020	387,572	"doja cat", "doja", "dindu nuffin song", #dojawasinnocent, #dojacatisNOToverparty, "WeAreSorryDoja", #dojacatisoverparty, #DojaIsOverParty, #DojaCat

Table 2. Control Artists' Dataset Overview

Artist Name	Initial Number of Tweets	Hashtags + Keywords Used for Query
Jaemin (NCT/NCT Dream)	117,654	"na jaemin", "jaemin", "bbl", "bubble", "nct", "nct dream", "nana", "nctzens", "czennie", "czennies", "nctzen", "#NCT", "#NCTDREAM"
Ryujin (ITZY)	36,176	"ryujin", "shin", "itzy", "#ITZY", "midzy", "midzys", "#RYUJIN"
Lil Baby	146,198	"lil baby", "#lilbaby"
Saweetie	70,904	'"saweetie" , "#saweetie"
Harry Styles	437,775	"harry styles", "harry", "styles", #harry, #harrystyles
Adele	34,282	"adele", #adele, #adele25, #adele21, #adelelove

5.2 Sentiment Analysis APIs

Initially, we looked at three sentiment analysis tools: Google Perspective, Vader, and TextBlob, in order to analyze the sentiment in our tweet datasets. However, upon conducting our preliminary analysis and EDA, we were not able to gather any meaningful results from Vader and Textblob that were not already evident through our results with the Google Perspective API. Therefore, we eventually decided to only use Google Perspective for our final analysis.

Using the Google Perspective API, we calculated the probability that each tweet in our datasets would be perceived as toxic, severely toxic, an insult, and containing profanity. The difference between toxic and severely toxic is that the severe toxicity attribute is less sensitive to milder forms of toxicity, such as text that contains profanity used in a positive case. Since we are focusing on sentiment among English-speaking Twitter users, we specified that we would be using the English language while using Perspective API. Therefore, any tweets that are not recognized as being primarily written in English by the API were not included in the final dataset containing the toxicity probability values for each tweet. In choosing which metrics we would use for our final project, we chose the severe toxicity and insult probabilities. Our preliminary results showed that toxicity and severe toxicity showed similar trends. Insult and profanity also shared similar trends with each other. Therefore, we narrowed down to only using severe toxicity and insult, which also were less susceptible to milder forms of toxicity such as positive uses of profanity.

5.3 RQ1: Type of Issue for Cancellation

The first research question we explored was how sentiment differed over time based on the type of subject the artist is "canceled" for. The six case studies, we present, fall into three different categories: discrimination against marginalized groups, abuse allegations, and misinformation. These categories were defined using our domain knowledge and research surrounding the detail of each controversy. In the discrimination against marginalized class, we have Doja Cat, Giselle, and Dababy. Both Doja Cat and Giselle received harsh criticism for singing terms which are frequently utilized

to degrade the African American community. Similarly, Dababy faced backlash for making homophobic remarks which disrespected the LGBTQ+ community. Next, in the abuse allegation class, we have Zayn Malik and Lucas. Zayn Malik made headlines for physically assaulting his mother-in-law and Lucas faced allegations for emotional abuse from former alleged girlfriends. In the final misinformation class, we have Nicki Minaj's controversy for COVID-19 vaccine misinformation.

Table 3. Categories of Cancellation for Canceled Artists

Group	Artist Names
Discrimination Towards Marginalized Groups	Giselle, DaBaby, Doja Cat
Abuse Allegations	Lucas, Zayn Malik
Misinformation	Nicki Minaj

To test the impact the type of conflict has on public sentiment, we conduct two types of analysis: qualitative and quantitative. For the qualitative analysis, we compared the words with the highest frequency before and after the controversy. We started by grouping the data from the artists in each category and compiled the text for every tweet in each group. To preprocess our text data, we removed all urls and standardized the text by converting it to lowercase. In addition, we removed any "stop words" from the tweets. Stop words are the common words, such as prepositions, conjunctions, and pronouns, that do not add much information or context to a piece of writing. After cleaning our text data, we were able to generate a table with the number of occurrences for every word within each category. Based on these results, we will be able to determine which category promoted more dialogue surrounding the conflict.

In regards to the quantitative analysis, for each category, we compared the severe toxicity and insult levels before and after the date of the controversy. For the preprocessing of our quantitative data, we initiated by removing any invalid severe toxicity and insult score that equaled 1000. A score of 1000 indicates that the tweet was not recognized by the API. We then converted all the values with the 'created_at' column to a datetime object and eliminated tweets that did not fall within the six months of the controversy date. Finally, we generated a 'days_cancel' column which presented the difference in the number of days the tweet was created to the controversy. To proceed with our analysis, we grouped the processed data by 'days_cancel' and computed the rolling average for the severe toxicity and insult score. In order to recognize any trends, we plotted the final data on a line plot which shows the change in sentiment for each category over time.

5.4 RQ2: Background of Musician

The background of the artists we researched was another aspect that we were interested in determining its influence on sentiment after cancellation. The differences between each artists' background, we believe, may influence the severity of sentiment, as well as the duration of sentiment during canceling. In order to determine whether or not the background of an artist has an effect on the sentiment they receive after being canceled, we created two different groups to look at. The first grouping of artists is based on the genre of music they're in. The genres that our artists fall into are K-pop, hip-hop, and Western pop. The second grouping of artists is based on sex, comparing female artists with male artists. The metric we used to compare artists in these groupings is determining the number of days after cancellation it takes for the median sentiment level to return to pre-controversy levels.

We dropped all of the control artists from these groups because the control artists weren't necessary for the analysis. We were only interested in how background affects the change in sentiment after getting canceled.

After sorting the artists for the groups in each category, we took the sentiment indicators (severe toxicity and insult levels) over time for each artist within the group and grouped by the number of days before and after controversy. The aggregate function we used on the sentiment indicators is median. In short, we calculated the median severe toxicity and insult levels for each group within genre and sex over the entire period of tweets we collected.

Next, we plot out the grouped data for the categories to visualize the median indicator levels for the groups against each other. In order to quantify the effect of the chart, we used the metric of determining the number of days after controversy until the median sentiment indicator levels returned to pre-controversy levels. We calculated this metric by taking the sentiment indicator level at day 0 or the day of the controversy and finding the first instance of the indicator level being below the day 0 level after the controversy has happened. For each group, we calculated this metric and compared it to other groups within the same category to determine if there is a difference for that background category.

5.5 RQ3: Parasocial Relationships

Parasocial relationships are one-sided relationships in which an individual develops illusions of intimacy with another individual, usually a media personality, as a result of repeated interactions between both individuals [7]. For example, a fan who has followed an artist for a period of time and has learnt enough about them to develop a sense of loyalty might feel that they have a personal connection with the artist. Depending on the strength of the relationship, media personalities can have strong influence over their audiences as evident with fandoms or fan bases like BTS's ARMYs or Taylor Swift's Swifties. To determine which artists have fan bases that share either a strong or weak parasocial relationship with the artist, we used two metrics: total mean engagement and fandom ratio. For the engagement metric, we first collected every tweet an artist posted 6 months before and after their cancellation date. We then calculated the total mean engagement by calculating the sum of the mean number of retweets, quote tweets, likes, and replies on an artist's tweets. To calculate the fandom ratio, we used our previous hashtag and keyword tweet datasets and so for each artist, we calculated the amount of tweets that contained at least one of the artist's fandom names divided by the total number of tweets in that artist's dataset.

Table 4. Cancelled Artists' Fandom Names

Artist Name	Fandom Name
Lucas (NCT/WayV)	"nctzen", "nctzens", "wayzennie", "wayzennies", "wayzenni", "weishennie", "czennie", "czennies", "#nctzen", "#weishennie", "#nctzens", "#weishennies", "lumi", "lumis"
Giselle (Aespa)	"MY", "MYs", "#MYs", "aerishine", "aerishines"
DaBaby	NONE or N/A
Nicki Minaj	"barb", "barbz", "barbs", "#barbz", "#barb"
Zayn Malik	"directioners", "directioner", "zquad", "#zquad"

Artist Name	Fandom Name
Doja Cat	"kittenz", "#kittenz"

Table 5. Control Artists' Fandom Names

Artist Name	Fandom Name
Jaemin (NCT/NCT Dream)	"nctzen", "nctzens", "czennie", "czennies", "#nctzen", "#nctzens", "nanadoongie","nanadoongies"
Ryujin (ITZY)	"midzy", "midzys", "#midzy"
Lil Baby	NONE or N/A
Saweetie	NONE or N/A
Harry Styles	"directioners", "directioner", "harries", "#harries"
Adele	"daydreamer", "daydreamers", "#daydreamer"

We then grouped them into either strong or weak parasocial relationships based on whether the artist reached above or below a certain threshold for each metric, in which our thresholds were determined based on our metric values and our own domain knowledge. For the total mean engagement, the cutoff threshold was 20,000 and for the fandom ratio, the cutoff threshold was 0.05. To be categorized as having a strong parasocial relationship, the artist's total mean engagement and fandom ratio would both have to be greater than or equal to its respective threshold. On the other hand, those who are unable to reach above both thresholds are then categorized as having a weak parasocial relationship. For example, Zayn Malik meets the threshold for total mean engagement as ~360,000 > 50,000 but does not meet the threshold for the fandom ratio metric as 0.003 < 0.05, so he was categorized as having a weak parasocial relationship.

Table 6. Engagement Metric Values Used to Determine Parasocial Strength

Artist Name	Total Mean Engagement	Fandom Ratio	Parasocial Strength
Lucas	112,107	0.123	Strong
Giselle	92,741	0.141	Strong
DaBaby	13,226	0.0	Weak
Nicki Minaj	81,179	0.052	Strong
Zayn Malik	359,275	0.003	Weak
Doja Cat	46,636	0.0001	Weak

6   Results

6.1   RQ1: Type of Issue for Cancellation

To investigate how sentiment changes over time based on the type of issue an artist was canceled for, we preprocessed our discrimination, abuse, and misinformation grouped datasets by calculating the mean severe toxicity and insult score for each date. We then applied a rolling median over an

interval of 14 days on the processed data to smoothen the visualization in order to recognize the trend.
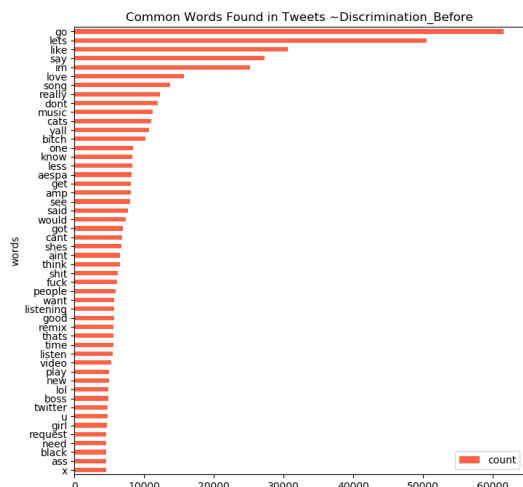

Figure 1: Most Common Words Before Cancellation For Discrimination Group
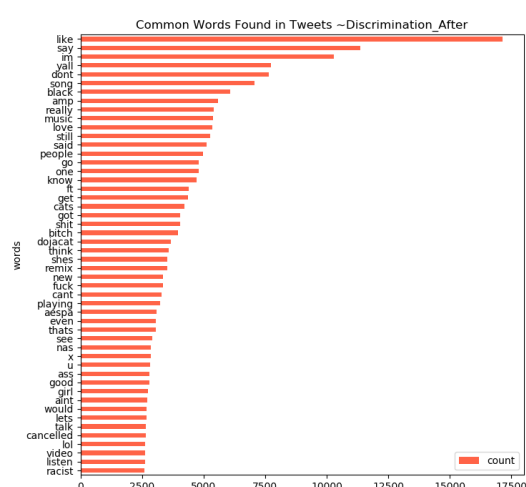

Figure 2: Most Common Words After Cancellation For Discrimination Group

The bar charts comparing the word frequency show that discrimination towards marginalized groups (specifically towards the African American community) promotes more conservation or dialogue surrounding the controversy in comparison to the abuse and misinformation category. This is evident since the "Most Common Words After Cancellation For Discrimination Issue Group" presents "black" at the 7th most frequent word and the word 'racist' also appears within the top 50 most common words post cancellation. However, both the abuse and misinformation graphs did not display any common words related to the conflict in question following the controversy.
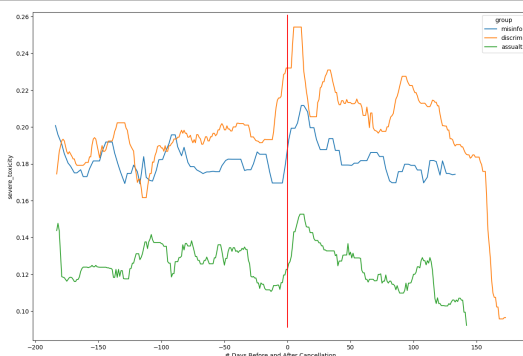

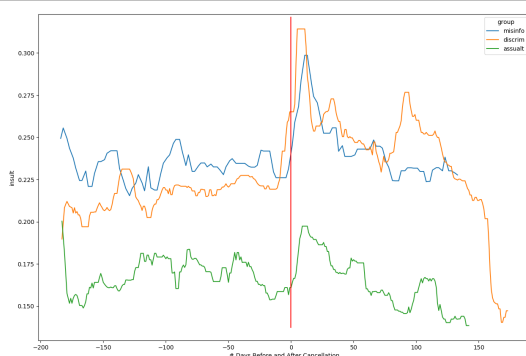**Figure 3:** Change in Median Severe Toxicity Score By Type of Issue

Figure 4: Change in Median Insult Score By Type of Issue

Table 7. Median Severe Toxicity Score By Type of Issue

	Misinformation	Discrimination	Abuse
Median Severe Toxicity Score Before Cancellation	0.1789	0.1486	0.1298

	Misinformation	Discrimination	Abuse
Median Severe Toxicity Score After Cancellation	0.1817	0.1717	0.1277

Table 8. Median Insult Score By Type of Issue

	Misinformation	Discrimination	Abuse
Median Insult Score Before Cancellation	0.2327	0.1781	0.1728
Median Insult Score After Cancellation	0.2390	0.204	0.1712

When looking at how the subject of the controversy affects sentiment, the artists we present who are canceled for discrimination towards marginalized groups promote higher negative sentiment compared to our artists with issues relating to abuse and misinformation. All groups have an initial spike immediately after the controversy. However, the discrimination category has a slightly large spike in severe toxicity and two more spikes are seen around the 40 and 90 day mark. In addition, when comparing the change in median severe toxicity score before and after the controversy, we observe the following: +1.6% increase for misinformation, +15.6% increase for discrimination, and -1.6% increase for abuse. The trend is not as distinct when comparing insult levels, however, it does provide a similar result.

When comparing the change in median insult score before and after the controversy, we observe the following: +2.7% increase for misinformation, +14.5% increase for discrimination, and -0.93% increase for abuse.

The trend is not as distinct when comparing insult levels, however, it does provide a similar result. Once again, discrimination exhibits the largest spike after the controversy. Although misinformation and discrimination follow a similar trend in insults levels post the controversy, discrimination still displays a slight increase in negative sentiment as another spike is recognized around the 100 day mark.

6.2 RQ2: Background of Musician

For the second subquestion measuring the effect of background on sentiment, we grouped artists into different genres and sex, and compared the differences in severity and length of sentiment by taking the median sentiment level for each group.

Table 9. Days After Canceling Until Severe Toxicity Scores Returns To Pre-Controversy Levels By Genre

Artist Name	K-Pop	Hip-Hop	Pop
# Days to Return to Pre-Controversy Levels	26	35	18
Severe Toxicity Levels at Day 0	0.1577	0.1764	0.2245

Table 10. Days After Canceling Until Insult Scores Returns To Pre-Controversy Levels By Genre

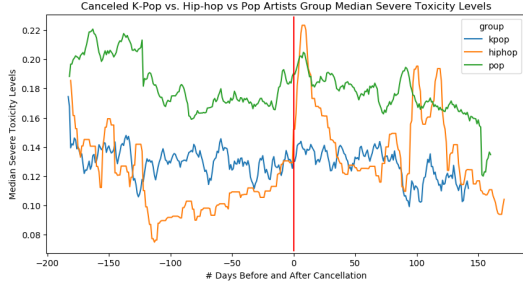| Artist Name	K-Pop	Hip-Hop	Pop
# Days to Return to Pre-Controversy Levels	18	30	1
Insult Levels at Day 0	0.1279	0.1449	0.1897

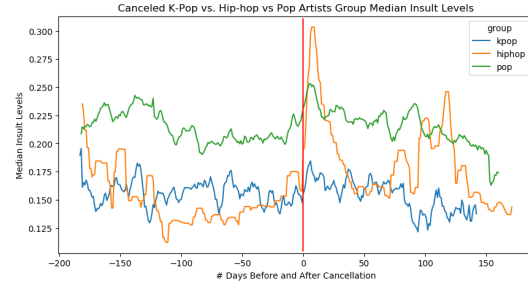**Figure 5:** Genre Comparison for Change in Median Severe Toxicity

Figure 6: Genre Comparison for Change in Median Insults

Comparing different genres to each other, hip-hop stands out. For median severe toxicity levels, hip hop artists took 30 days for the severe toxicity to return to the level it was at before controversy. K-pop artists took 18 days and pop artists took 1 day (Table 9). For median insult levels, hip hop artists took 35 days, k-pop artists took 26 days, and pop artists took 18 days to return to pre-controversy insult levels (Table 10). From the data, hip-hop artists seem to undergo a prolonged negative effect from cancellation compared to our other genres. Also, from our visualizations on genre severe toxicity and insult levels, canceled hip hop artists experience a more severe change in toxic sentiment compared to other genres as both plots have an extremely sharp peak in sentiment levels. Median hip-hop severe toxicity levels go from 0.176 to around 0.300 (Figure 5) while median hip-hop insult levels go from 0.145 to 0.220 immediately after controversy (Figure 6).

Table 11. Days After Canceling Until Severe Toxicity Scores Returns To Pre-Controversy Levels By Sex

Artist Name	Male	Female
# Days to Return to Pre-Controversy Levels	71	18
Severe Toxicity Levels at Day 0	0.111	0.1942

Table 12. Days After Canceling Until Insult Scores Returns To Pre-Controversy Levels By Sex

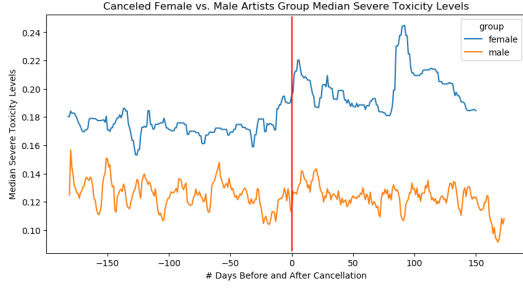| Artist Name	Male	Female
# Days to Return to Pre-Controversy Levels	71	13
Insult Levels at Day 0	0.1328	0.2436

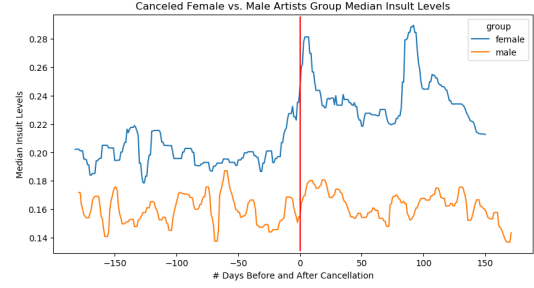**Figure 7:** Sex Comparison for Change in Median Severe Toxicity

Figure 8: Sex Comparison for Change in Median Insult

Our results for grouping by sex show that canceled female artists experience a greater backlash in response to controversy but for a much shorter period of time. Canceled male artists' sentiment levels returned to normal/pre-controversy levels 71 days after controversy while female artists took 13 days for severe toxicity and 18 days for insult levels to return to normal (Table 11, 12). Male artists experienced a much longer duration of the effect from cancellation compared to female artists. However, female artists sustain a greater level of negative sentiment for the entire time period of tweets we collected and have a sharper and more drastic peak after being canceled (Figure 7, 8).

To conclude, genre and sex seem to make a difference in the length and severity of the sentiment for canceled artists. Our canceled hip hop artists experienced longer and more drastic differences in sentiment over the period of being canceled compared to k-pop and pop artists. Our male artists who were canceled experienced the effects of being canceled for a much longer period of time than female artists, but female artists experience a more drastic effect and have higher toxic sentiment overall.

6.3 RQ3: Parasocial Relationships

To investigate how sentiment changes over time based on the strength of a canceled artist's parasocial relationship with their fans, we preprocessed our strong and weak grouped datasets by first calculating the mean severe toxicity and insult score for each date. We then applied a rolling median over an interval of 14 days on the transformed data in order to facilitate the observation of long-term trends.
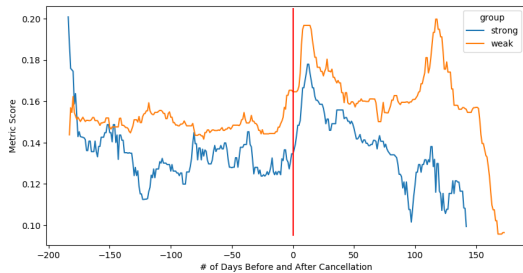

Figure 9: Change in Median Severe Toxicity Score Over Time
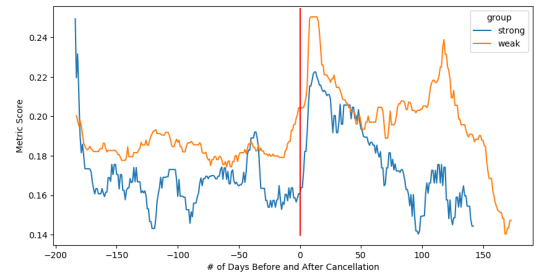

Figure 10: Change in Median Insult Score Over Time

Based on both figures, we can see that artists that have a strong parasocial relationship with their

fans tend to have overall lower negative sentiment compared to artists with a weak parasocial relationship. Figure 9 and Figure 10 both show a dramatic increase in median severe toxicity and insult score directly after the artist's cancellation. Both strong and weak parasocial groups also experience an increase in negative sentiment after 100 days, which is possibly due to the controversy resurfacing. However, the weak parasocial group shows a more significant increase that is comparable to its initial spike seen directly after cancellation.

Table 13: Median Severe Toxicity and Insult Before Cancellation

	Strong	Weak
Median Severe Toxicity Score	0.1323	0.1492
Median Insult Score	0.1658	0.1826

Table 14: Median Severe Toxicity and Insult After Cancellation

	Strong	Weak
Median Severe Toxicity Score	0.1388	0.1606
Median Insult Score	0.1744	0.2033

In terms of how sentiment has changed compared to before cancellation, the median severe toxicity score increased by +4.5% for the strong group and +7.4% for the weak group after cancellation. The median insult score increased by +5.5% for the strong group and +11.5% for the weak group (Table 14). Therefore, we can see that there is a slight percent increase in severe toxicity and insult scores for both groups after cancellation. When comparing our weak parasocial group against our strong parasocial group, there is also a slight positive difference in percent increase with a +2.9% difference for severe toxicity and +6% for insults. This supports that artists with a weak parasocial relationship experience slightly increased toxicity overall after cancellation compared to those with a strong parasocial relationship.

7 Discussion

In analyzing how sentiment towards canceled musicians changes over time, we can conclude that there is an overall increase in negative and toxic sentiment after cancellation. However, the duration of such negative sentiment towards a canceled artist appears to be short-term.

Depending on the severity of the issue, we believe this could affect the intensity of negative sentiment towards the canceled artist. Our research shows that artists canceled for discrimination against a marginalized group experienced greater backlash compared to artists canceled for abuse or misinformation. It was unexpected that artists canceled for abuse allegations did not demonstrate higher levels of negative sentiment. However, due to the sensitivity of the topic as well as other factors such as the strength of the artist's parasocial relationship with their fans, it is possible these reasons are why conversations about these individuals were less toxic than expected.

Given that canceled artists' backgrounds also could affect sentiment, we analyze how the genre of music they cater to and their sex affects sentiment after cancellation. We found that hip-hop artists experienced a higher rise in negative sentiment and took the longest to return to pre-controversy

median severe toxicity and insult levels compared to K-pop and pop. The huge spike in negative sentiment for hip-hop artists was unexpected, but is likely influenced by issues they were canceled for, their target audience, and their general reputation in the eyes of the public. In terms of sex, male artists took longer to return to pre-controversy median sentiment levels, but female artists had a more drastic rise and a greater overall average in negative sentiment after cancellation. This suggests that there might be a gender imbalance with Twitter cancel culture in which an artist's sex could affect the intensity of negative sentiment they feel towards a canceled artist.

The strength of a canceled artists' parasocial relationship can also affect sentiment as we found that artists who have a strong parasocial relationship with their fans experience less toxicity overall. Fans who have a strong parasocial relationship with an artist might feel a sense of loyalty towards them. This suggests that because fans feel loyal towards an artist, they might be more compelled to defend them and suppress as much negativity as possible after the artist's cancellation. Especially considering how fans of WayV's Lucas were using mass tweeting hashtags such as #WELOVEYOULUCAS or #LUCASBESTBOY as signs of support after his controversy (Table 1), it appears that the strength of an artist's parasocial relationship can significantly influence public sentiment surrounding them.

After conducting our research, there are several limitations that we would like to point out. First of all, we only focused on sentiment among English-speaking Twitter users. Depending on the language and the social media platform, there is the possibility that our findings could differ by seeing more intense or reduced toxicity. In categorizing the strength of an artist's parasocial relationship, the thresholds we chose for our metrics were dependent on the data we collected as well as our domain knowledge. This would mean that using our thresholds to categorize other artists might not accurately represent the true strength of their parasocial relationship with their fans. Therefore, for future research, we would advise conducting network analysis on a canceled artist's Twitter followers to quantify the strength of their parasocial relationship. Also, since we only conducted case studies on six canceled individuals of varying follower counts, it is possible that our findings may differ if we have chosen different artists or included more canceled artists. Despite this, all of our canceled artists saw some sort of increase in toxicity post cancellation, suggesting that cancel culture may lead to an overall rise in negative sentiment over time. For future research, we would advise seeing if overall sentiment towards a canceled artist is affected by whether they publicly apologized or not. We suspect that if an artist had publicly apologized, they may show reduced toxicity and negative sentiment.

To summarize, it appears that there is a rise in toxicity towards canceled individuals over time. However, there are various factors that seem to contribute to the variation in negative sentiment. Cancel culture on Twitter is still an evolving phenomenon that has only begun to be recognized in recent years. Because of this and the many factors that can affect the amount of toxicity directed towards canceled individuals, it is difficult to determine whether cancel culture is effective in ostracizing them from society or to what extent would it affect their lives and reputations. Regardless, the intense backlash they face in the short term demonstrates that Twitter users can hold canceled individuals accountable for their actions to some extent. This could put pressure on canceled individuals to publicly apologize and acknowledge their actions, which is already evident with some of our case studies such as Aespa's Giselle (source). Cancel culture can emphasize that certain behavior is not acceptable in today's socio-political climate by holding accountable canceled individuals on a public stage. However, future research is needed to assess its effect on canceled individuals' reputations in the long-term and the extent to which cancel culture can go.

8   Conclusion

Our initial hypothesis was that online communities, specifically on the platform Twitter, respond to an artist or artists' controversy by effectively "canceling" them. We defined "canceling" as a rise in negative sentiment and toxicity towards an individual.

From our research into cancel culture, we determined that cancel culture does negatively affect music artists' sentiment on Twitter. Through our analysis, we found different factors and aspects of artists which have an effect on the severity and duration of negativity and toxicity received after being canceled. We identified these areas of interest as type of issue, background of artist, and parasocial relationships.

To conclude, our investigation through the three sub-question shows that cancel culture overall can impact public opinion and sentiment of controversial artists on Twitter. We captured the effect of cancel culture before and after artists' controversies, but long term effects like complete ostracization from the mainstream popular culture or continued toxicity didn't occur as sentiment eventually returned to normal within the time frame we looked at. Cancel culture is a topic that is heavily debated, but we believe that our analysis has provided some context to understand and quantify cancel culture as a whole.

9   References

1. https://www.soompi.com/article/1485476wpp/ncts-lucas-apologizes-for-past-behavior-involving-ex-girlfriends-sm-announces-he-will-halt-activities
2. https://www.soompi.com/article/1495215wpp/aespas-giselle-apologizes-for-mouthing-a-racial-slur-from-a-song
3. https://www.billboard.com/photos/dababy-rolling-loud-homophobic-comments-controversy-timeline
4. https://www.vox.com/22673528/nicki-minaj-vaccine-tweets-cousins-friend-met-gala
5. https://www.cnn.com/2020/05/25/entertainment/doja-cat-racial-slurs-trnd/index.html
6. https://stylecaster.com/why-zayn-malik-yolanda-hadid-fight/
7. https://en.wikipedia.org/wiki/Parasocial_interaction

10   Appendix

10.1   Figures of Most Common Words For Assault and Misinformation Groups

Figure 11: Most Common Words Before Cancellation For Assault Allegation Group

Figure 12: Most Common Words After Cancellation For Assault Allegation Group

Figure 13: Most Common Words Before Cancellation For Misinformation Group

Figure 14: Most Common Words After Cancellation For Misinformation Group