“Hey, it's been a while.”

The Spectacle of Wrongdoing and Recovery on YouTube

Neluka Ameresekere, Susane Dang, Jingyi Long

CTS*3000

Dr. Susan Brown

April 2023

Introduction

What is an influencer? Currently, “influencer” refers to individuals who cultivate audiences online, resulting in loyalty that can be measured in metrics such as likes and comments (Hund, 2023). This quantification of engagement can then be leveraged by influencers to promote themselves for brand sponsorships. In her book, *The Influencer Industry: The Quest for Authenticity on Social Media*, Emily Hund traces the history of influence, and how the idea of the internet influencer came to be. Throughout this paper, we use “influencer” to refer to internet influencers.

Current influencers exist because of increased access to the internet and the development of social media. Early influencers used newsletters and blogging sites, but as technology progressed, it became easier for people to connect with audiences through social networks like Twitter and Facebook (Hund, 2023). YouTube is the world’s largest video platform with over 2.5 billion active monthly users and is the home of the internet’s most famous influencers (Asarch, 2021; WeAreSocial et al., 2023).

The influencer is similar to a celebrity, but unlike celebrities, the influencer bonds with audiences through content that directly addresses the viewer. In “The Youtube Apology,” Gabriella Karlsson likens this bond to the formation of a one-sided parasocial relationship (2020). Influencers refer to their followers and fans as “friends,” and they perform an intimate version of their “authentic” self through their content such as showcasing their feelings on personal topics or major life events. Intimacy builds trust, and the trust from the audience to the influencer builds a stronger parasocial bond (Karlsson, 2020). This bond transforms viewers into subscribers, who become invested in the lives of the influencer.

When an Influencer Makes a “Mistake”

Because of the personal nature of the influencer’s content, their followers often feel personally affronted when their favourite creator has seemingly committed a transgression. This can range from a number of offenses, such as using racial slurs, lying, and being insincere, to assault, abuse or even faking one’s death.

As a result of this “bad behaviour,” followers may demand an apology or acknowledgement of wrongdoing. Whether the influencer believes they have actually done something wrong is often irrelevant, and they must perform an apology to regain their audience’s trust.

The Apology Video

The apology video has become common among YouTubers, regardless of their branding. These videos are so similar in style and content that they have become their own genre subject to parody (Makalintal, 2019).

But what does it mean to apologize in a public space? These creators, whose entire brand is commonly about them being their "authentic" selves, are performing a persona when they create internet content. This performance extends to their apology—it is a performance of shame.

It is not enough for the fans that a creator is sorry for what they have done. The creator is expected to have made demonstrable changes in their following content. Again: another performance.

People who "expose" the creator are usually doing so with the intention of holding the influencer accountable for their behaviour. However, since this often takes place on a public forum, the exposure is transformed into a public spectacle. The spectators can extend beyond the creator's followers to other online communities and even mainstream media outlets.

In *Updating to Remain the Same: Habitual New Media*, Wendy Chun describes this act of exposure as "violent" and likens it to Sedgwick's epistemology of "the closet," as it seeks to shame the individual (2008). Although Chun specifically relates it to slut-shaming, and how victims of slut-shaming can protest against this public humiliation, the forced accountability is similar to the exposure of the influencer. However, the influencer has financial incentive to apologize, whether they agree or disagree that what they did was wrong.

Goals

Originally, we wanted to examine public personas and the performance of an apology. Our main guiding questions were:

- What makes an apology work?
- How was the YouTuber's popularity affected by their controversy and subsequent apology?
- Why do viewers care so much about the intentions of a stranger online?

We thought that data on this topic would be readily available because influencers are public figures, and these apologies have become milestones in their careers. However, the information we found was scattered and incomplete.

As a result, we pivoted our focus and started compiling this data ourselves, using transcription software as well as internet archival tools to collect metadata. Our new goal was to create a dataset of YouTube apologies that can be used for the type of analysis that we had originally wanted to do. This would also be useful for others who are researching this topic, as it would not be necessary to collect all the data from scratch. Recording how we gathered this information was also crucial, so that it may be replicated by others who wish to do similar work. Unlike other projects that have analyzed YouTuber apologies, our dataset also includes videos that were unlisted, privated, deleted or were from a channel that was terminated entirely by YouTube. In this way, our dataset works as an archive for YouTube apologies. We did not want to exclude these videos, as if they had never been made. Furthermore, deleting a video or otherwise changing its status could be a factor used in examining the authenticity of a YouTuber's apology.

In "Remains", Tonia Sutherland (2021) discusses the right to be forgotten, and how it is often a privilege—one that is not afforded to people of colour. We aimed to level this inequality in our dataset by including the actions of those who enjoy this privilege such as white people, so that they cannot pretend as if their wrongdoings never happened.

The interactive visualization "The Aftermath of a YouTube Apology" by Arjun Kakkar and Russell Samora included information such as video length and genre (2020). However, by recording other information in our dataset, namely the creator's race and age, we could potentially identify trends in apology reception and the biases that contribute to these trends.

While we had wanted to do more content analysis than was actually possible within the given timeframe, we still used textual analysis tools on a smaller scale to test that data’s usability and consider how it can be incorporated into future projects.

Considerations

The main ethical element to this project was the consideration if it was okay for us to possess this information in the first place. The dataset will likely outlast the videos that it was sourced from, and there are videos within the dataset that were already private or deleted by the time of collection. What happens when those creators no longer wish for this data to be on the internet?

The purpose of this project was to archive the impact that these apologies had on image and subscriber retention from an outside perspective. The creators featured in our project were not affiliated with us in any way and had no impact on the data curation process. These creators have established themselves in positions of power where their words and actions have meaningful impact that should be recorded.

We do not aim to criticize YouTubers based on their actions years ago, but believe that removing the history and backlash of those actions is akin to changing history which we seek to prevent. There is an argument to be made on whether we are "replicating harmful data regimes'' according to the Feminist Data Manifest-no (Cifor et al., 2019). However, we do not believe this is the case as our project does not emphasize race in any way or target marginalized groups. Additionally, all of the YouTubers that we are analyzing come from positions of power or possess racial privilege, so we do not believe that this dataset will replicate nor perpetuate systemic injustice. Our project is strictly for archival purposes and data retention of what we believe to be key events.

Methods

The Process

The original methodology included selecting and examining the YouTuber's controversy, a close reading of the subsequent apology and the changes to the YouTuber's following throughout the event. These were going to be done through five case studies of YouTubers who fall into different types of apologies. The original case studies are as follows:

- An early example (Fine Brothers Entertainment)
- A recent example (The Try Guys)
- An example of a "good" apology (Jenna Marbles)
- An example of a "bad" apology (James Charles and Tati Westbrook)
- An example of a repetitive apologist (Logan Paul)

Our original process was linear: we would start by doing an environmental scan of past apologies done by YouTubers through informal sources such as articles and video essays, given the pop culture nature of the project. After choosing a number of case studies to focus on, we would delve deeper into the context surrounding the YouTuber's apology, social media metrics, and subsequent reception by their audience. We also planned on performing sentiment analysis on the comments of the apology videos through Python.

However, as we started transcribing these apologies, we realized that we were limiting ourselves to a single controversy within the creator's career. In addition, some of the chosen video case studies (ex. The Fine Bros) were deleted or inaccessible from the YouTuber's channel. This limited our ability to perform sentiment analysis since we were unable to obtain the comments from the published video.

We then decided to do an in-depth case study of a single creator who has been associated with multiple controversies so that we could better identify the effects of their apology. Focusing on a creator like Logan Paul who has been perceived as both hero and villain (multiple times) would act like a control—it would be possible to see what specifically about the apologies and scandals were different each time.

This may not be especially indicative of any larger trends because Logan Paul is young, white, and a millionaire. These factors may affect how he is perceived compared to other creators who are less successful or privileged. These are qualities of people that are generally favoured by society and are not held as accountable for problematic behaviour.

Finally, we settled on a distant reading of a wide variety of apologies. By analyzing the contents of the videos, it would be possible to find emerging topics and similarities between creators. This could then be compared to the public reception of the apology, helping us identify whether other factors such as race or gender affected the creator's decrease or increase in popularity (via subscribers, video views, and likes).

Initial Data Collection

Prior to the creation of our YouTuber apology dataset, there were no comprehensive datasets of apologies available to the public. There has been some analysis done on this topic, which we consulted when developing our project. Kakkar and Samora's visualization, "The Aftermath of a YouTube Apology", was a primary source of inspiration for the type of project we wanted to create to present our findings. The visualization used a small sample size of 34 apologies, which was available through a YouTube playlist. Grace Choi and Anne Marie Mitchell's (2022) analysis on YouTube apologies used a sample size of 117 videos, which is substantially larger than Kakkar and Samora's visualization, but did not include information about which videos were included. As such, we wanted to create a dataset that could be accessible like Kakkar and Samora's, but as large (if not larger) than Choi and Mitchell's.

We started our data collection process by compiling apologies from different sources. As our team is all well-versed in the genre of the YouTube apology, we started our compilation by collecting apologies that we were personally familiar with. This knowledge was used in tandem with news articles about YouTube apologies as a way to cross-reference our data. After exhausting news articles and our own knowledge, we used YouTube's search feature to look for keywords that are typically included in apology video titles. This included, but was not limited to, the following: sorry, apology, response and truth. After that, we expanded our search beyond YouTube to other social media sites such as Reddit and Twitter. Apologies were also gathered through videos on YouTube that talk about apologies, with an example being "Youtuber Apology Tier List" by YouTuber Charles White (channel: penguinz0), which currently has just over 9 million views. All apologies collected were documented on a spreadsheet.

Through the collection of these apologies, we created criteria to help filter which apologies would stay in the dataset. Our criteria was developed through a mix of our own ideas, as well as criteria from *The YouTube Apology* by Gabriella Karlsson.

To meet our criteria and be archived as part of our apology dataset, the content must have been created by an independent YouTuber (not on behalf of an agency) and be uploaded directly to YouTube. These criteria were adapted from Karlsson's work, as they applied to our project needs. We also required that the link to the apology video was accessible on the original YouTube channel it was posted in or if necessary, through the Wayback machine for videos unavailable on YouTube. If videos did not have sufficient metadata, they were not included. Lastly, we narrowed our focus by only including one apology per Youtuber. In the case a YouTuber had created multiple apology videos, we chose the one that had the most views.

By applying the criteria mentioned above, we ended with 80 YouTubers in our spreadsheet from an original list of over 100 apologies.

Tools

The main tools used throughout the apologies project were YouTube, 4K Video Downloader, Social Blade, the Internet Archive's Wayback Machine, WikiTubia, and Python libraries including NLTK, Pandas, Whisper, and Gensim.

YouTube

Youtube is the original hosting platform of all the apology videos we looked at, as we focused on influencers for whom YouTube is their primary platform and source of income. Although some apology videos were inaccessible due to channel deletion, video deletion or privating, captures from the Wayback Machine were used to view archived versions of video and channel pages. YouTube playlists were used to compile apology videos for easy access and organization purposes.

4K Video Downloader

4K Video Downloader is a YouTube downloader application that allows users to download videos in full resolution. It has the ability to extract audio from videos and download videos in different formats and in batches. 4K Video Downloader was used to download YouTube playlists of the apology videos as audio files for audio transcription.

Social Blade

Due to the nature of our project, the historical data we used was not created by ourselves, but rather, gathered from third party platforms.

Social Blade is a site dedicated to tracking changes in social media engagement metrics. The curation of YouTube subscriber and viewership data largely depended on Social Blade, as the YouTube API could only gather current data.

Prior to 2017, Social Blade hosted all of its channels' statistics on its website. However, in November 2017, Social Blade was required by YouTube to only show statistics up to three years back due to EU data protection regulations (Social Blade, n.d.). As a result, any data we wanted to obtain on Social Blade from 2020 and before had to be accessed via the Wayback machine.

Additionally, in May 2019, YouTube announced that channels with over 1,000 subscribers would have their public subscriber counts abbreviated (e.g., 432,930 as 432K and 51,389,232 as 51M) (Team Youtube, 2019). This change took place in September 2019, and greatly affected our methods in gathering subscriber data. Prior to the change, Social Blade was able to track daily changes in subscriber counts. However afterwards, these figures were only updated in set increments.

USER STATISTICS TABLE FOR JAMESCHARLES (APR 12TH, 2019 - MAY 11TH, 2019)

DATE		SUBSCRIBERS		VIDEO VIEWS	
2019-04-12	Fri	+28,103	16,086,297	+3,779,158	1,396,026,630
2019-04-13	Sat	+31,528	16,117,825	+5,693,385	1,401,720,015
2019-04-14	Sun	+31,377	16,149,202	+3,447,497	1,405,167,512
2019-04-15	Mon	+27,908	16,177,110	+2,537,798	1,407,705,310
2019-04-16	Tue	+27,572	16,204,682	+3,648,234	1,411,353,544
2019-04-17	Wed	+32,951	16,237,633	+6,637,796	1,417,991,340
2019-04-18	Thu	+30,385	16,268,018	+3,597,548	1,421,588,888
2019-04-19	Fri	+30,922	16,298,940	+4,788,207	1,426,377,095
2019-04-20	Sat	+31,152	16,330,092	+5,337,410	1,431,714,505
2019-04-21	Sun	+28,306	16,358,398	+3,271,920	1,434,986,425
2019-04-22	Mon	+23,873	16,382,271	+2,392,900	1,437,379,325
2019-04-23	Tue	+24,933	16,407,204	+4,590,524	1,441,969,849
2019-04-24	Wed	+32,159	16,439,363	+6,631,980	1,448,601,829
2019-04-25	Thu	+17,569	16,456,932	+3,935,582	1,452,537,411
2019-04-26	Fri	+14,363	16,471,295	+2,650,153	1,455,187,564
2019-04-27	Sat	+17,037	16,488,332	+2,303,295	1,457,490,859

YOUTUBE USER ANALYTICS / STATISTICS FOR JAMES CHARLES (2023-03-22 - 2023-04-20)

DATE		SUBSCRIBERS		VIDEO VIEWS	
2023-03-22	Wed	–	23.8M	+786,335	3,871,991,021
2023-03-23	Thu	–	23.8M	+779,946	3,872,770,967
2023-03-24	Fri	–	23.8M	+911,548	3,873,682,515
2023-03-25	Sat	–	23.8M	+1,806,174	3,875,488,689
2023-03-26	Sun	–	23.8M	–	3,875,488,689
2023-03-27	Mon	–	23.8M	+4,312,381	3,879,801,070
2023-03-28	Tue	–	23.8M	+1,961,555	3,881,762,625
2023-03-29	Wed	–	23.8M	–	3,881,762,625
2023-03-30	Thu	–	23.8M	+3,838,761	3,885,601,386
2023-03-31	Fri	–	23.8M	+1,761,733	3,887,363,119
2023-04-01	Sat	–	23.8M	–	3,887,363,119
2023-04-02	Sun	–	23.8M	+1,922,789	3,889,285,908
2023-04-03	Mon	–	23.8M	+2,305,122	3,891,591,030
2023-04-04	Tue	–	23.8M	+1,039,336	3,892,630,366
2023-04-05	Wed	–	23.8M	–	3,892,630,366
2023-04-06	Thu	–	23.8M	+2,372,855	3,895,003,221

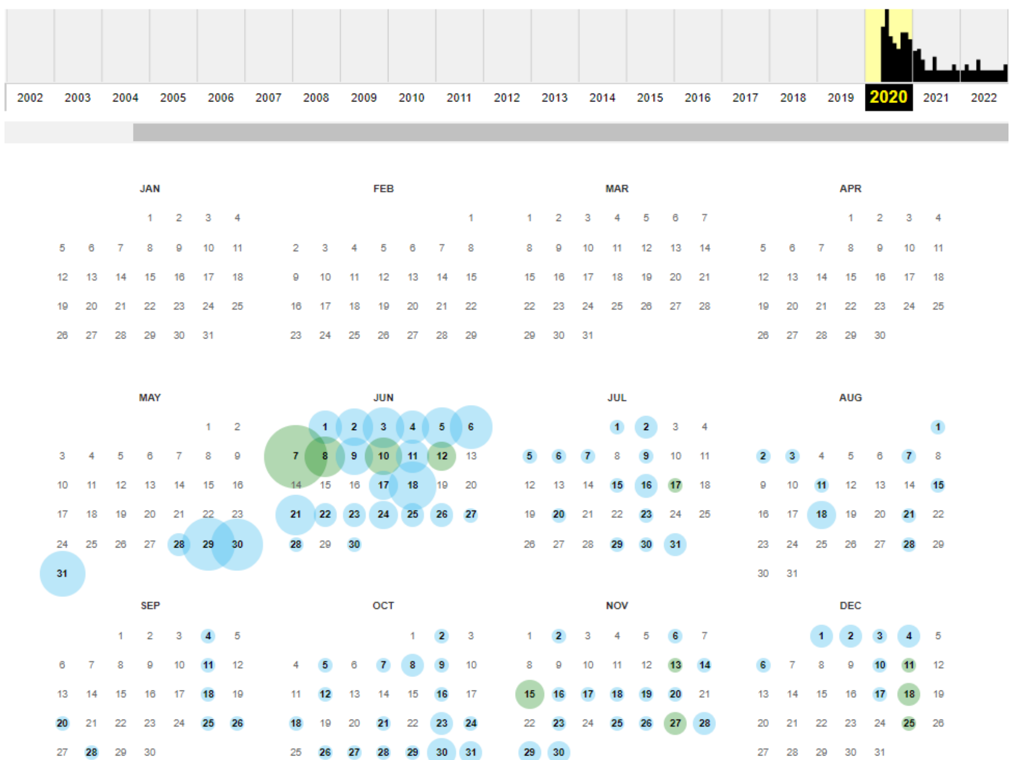
Screenshots of James Charles' monthly subscriber counts from April 2019 captured from The Internet Archive (left) and April 2023 (right).

Given Social Blade's inability to provide daily subscriber counts after August 2019, we had to manually look for subscriber counts using The Internet Archive for apologies posted after that date – this accounted for approximately half the apologies we curated. Any apologies that came prior were automatically scraped from Social Blade with a Python script by Anjali Shrivastava (2020).

The Internet Archive's Wayback Machine

The Wayback Machine is a digital archive of web pages that allow visitors to view "Internet sites and other culture artifacts in digital form" (Internet Archive, n.d.). It preserves web pages through periodic site crawls, which can take years to complete. As a result, many captures are contributed by individuals who have chosen to preserve a webpage on a specific date.

Many apologies we looked at were deleted or made private due to reasons such as backlash from the apology or the controversy being resolved. Although a fair number of apologies could still be viewed on YouTube in the form of reuploads, the re-uploaded apology did not have the original subscriber or likes data we required. To gather this data, we used the Internet Archive's Wayback Machine to look at archived captures of the original video on various dates.



Myka Stauffer's Internet Archive Wayback Machine page for her now private apology video.

Due to the crowdsourced nature of the Wayback Machine, the availability of pages is entirely dependent on the number of crawls performed and the interest of individuals who have chosen to capture the pages.

Wikitubia

Wikitubia is an unofficial, crowd-sourced wikia dedicated to YouTube creator channels with a minimum of over one thousand subscribers (Wikitubia, n.d.). Wikitubia has been active since February 2007, with almost 21,000 content articles created during its lifespan (Wikitubia, n.d.). During our apology data collection, Wikitubia was used to gather the original links to YouTuber channels. After the introduction of channel handles by YouTube, the main channel URLs were changed to personalized handles that the creator could choose (YouTube Team, 2022). Through this URL change, the Wayback machine was not able to retrieve the appropriate data from the current YouTube channel pages as it recognized the handle as an entirely new URL. Using the original channel links provided by Wikitubia circumvented this, and allowed us to obtain important channel information such as subscriber counts.

Python

We used various Python libraries including Whisper, Pandas, Natural Language Toolkit, and GenSim.

Whisper

Whisper is a free transcription script developed by OpenAI. This helped us automate the transcription process, but sometimes resulted in inaccurate text.

```
import whisper

# transcribe FineBros apology
model = whisper.load_model("base")
result = model.transcribe("fine_bros.mp3")
print(result["text"])
```

Sample code used to transcribe text.

Pandas

Pandas is a data analysis library that we used to manipulate the data. This allowed for sorting by different categories such as genre, race, and gender.

Natural Language Toolkit

Natural Language Toolkit (NLTK) allows users to work with and analyze natural language. Using NLTK, we were able to clean up the text and find the frequency of certain words and phrases. Pandas helped organize the text for use with NLTK.

```python
# making lemmatized list
lemma = WordNetLemmatizer()
Tati_lemma = []
Tati_filter = []

from nltk.corpus import stopwords
stopwords = stopwords.words('english')

for word in Tati_filter:
    Tati_lemma.append(lemma.lemmatize(word))

for word in Tati_token:
    if word not in stopwords:
        Tati_filter.append(word)

# lemmatized version of tags
Tati_text = nltk.Text(Tati_filter)

from nltk.probability import FreqDist
fdist = FreqDist(Tati_text)
fdist.most_common(20)
```

Code using NLTK to clean up and filter a corpus, and obtain most frequent words.

Gensim

Gensim is a library used for topic modelling. As we shifted the focus of our project, we were unable to do as much as originally planned. However, we used topic modelling on smaller batches of data to test out the tool.

```python
NUM_TOPICS = 10

ldamodel = gensim.models.ldamodel.LdaModel(corpus=scam_corp, num_topics = NUM_TOPICS, id2word=scam_dict, passes= 50)
ldamodel.save('model5.gensim')

# store topics
topics = ldamodel.print_topics(num_words=10)
counter = 0

# print the topics
for topic in topics:
    counter = counter + 1
    print("Topic #", counter)
    print("--------------------------")
    print(topic)
```

```
Topic # 1
--------------------------
(0, '0.090*"thing" + 0.057*"take" + 0.041*"someone" + 0.039*"years" + 0.031*"creators" + 0.030*"else" + 0.029*"worl
d" + 0.029*"shit" + 0.029*"two" + 0.025*"hope"')

Topic # 2
--------------------------
(1, '0.059*"also" + 0.057*"saying" + 0.053*"got" + 0.052*"go" + 0.043*"videos" + 0.041*"fucking" + 0.039*"first" +
0.037*"many" + 0.036*"part" + 0.027*"point"')

Topic # 3
--------------------------
(2, '0.204*"know" + 0.081*"say" + 0.062*"said" + 0.033*"mean" + 0.033*"much" + 0.032*"ever" + 0.031*"okay" + 0.029*
"yeah" + 0.028*"love" + 0.027*"happened"')

Topic # 4
--------------------------
(3, '0.083*"get" + 0.075*"never" + 0.055*"feel" + 0.053*"back" + 0.047*"person" + 0.036*"good" + 0.033*"friends" +
0.030*"last" + 0.029*"wanted" + 0.028*"play"')

Topic # 5
--------------------------
(4, '0.254*"people" + 0.064*"see" + 0.044*"every" + 0.037*"us" + 0.034*"twitter" + 0.031*"came" + 0.024*"saw" + 0.0
21*"around" + 0.020*"best" + 0.020*"watching"')

Topic # 6
--------------------------
(5, '0.094*"one" + 0.092*"guys" + 0.087*"lot" + 0.063*"something" + 0.042*"stuff" + 0.036*"done" + 0.033*"thought"
+ 0.031*"bad" + 0.025*"maybe" + 0.023*"thank"')

Topic # 7
--------------------------
(6, '0.312*"like" + 0.077*"things" + 0.070*"even" + 0.040*"wrong" + 0.033*"still" + 0.032*"game" + 0.031*"better" +
0.023*"call" + 0.023*"today" + 0.022*"long"')

Topic # 8
--------------------------
(7, '0.126*"really" + 0.086*"would" + 0.063*"way" + 0.054*"right" + 0.043*"could" + 0.041*"day" + 0.030*"anything"
+ 0.027*"making" + 0.026*"started" + 0.025*"getting"')
```

Sample code for topic modelling to determine ten topics with ten keywords each, and its output.

Due to time constraints and our lack of experience with topic modelling, this can be explored further in the future.

Other Tools

We were previously interested in using the YouTube Data API for accessing video information. However, the API could only access current data–historical data and data for deleted videos were unavailable, so Social Blade and Python were used in place of it.

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a sentiment analysis tool that specialized in social media sentiment. We successfully used this tool, but were unable to perform it on a larger scale and obtain much meaningful data due to time constraints.

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
sentiment = SentimentIntensityAnalyzer()

text_001 = Austin
sent_001 = sentiment.polarity_scores(Austin)
print("Sentiment of text 1:", sent_001['neg']*100, "% negative,",
sent_001['neu']*100, "% neutral", sent_001['pos']*100, "% positive")
```

```
Sentiment of text 1: 13.0 % negative, 77.9 % neutral 9.1 % positive.
```

Code using VADER to perform sentiment analysis for a single apology.

Data Collected

ID

An identifier was created for each creator to better locate specific apologies across multiple lists. The collected apologies were numbered chronologically and combined with a shorter version of the YouTuber's name that was still easily recognizable. (e.g. Beni and Rafi Fine, or The Fine Brothers, were the fourth apology on the list. Their name was shortened to "Fine", resulting in the ID "004_Fine").

Video Information

This included the video's title, duration, and the date it was posted. Since some videos have since been deleted by the creator and then reuploaded by other individuals, we collected the URLs of the original upload. This ensured accuracy when we collected information about the video from around the time it was released.

We also recorded information regarding the video's number of "likes" and "dislikes". In November 2021, YouTube announced that dislike counts would no longer be publicly available (The YouTube Team, 2021). Due to this, dislike counts for videos still available on YouTube were obtained through the "Return Youtube Dislike" Chrome extension (Anarios, 2013). Dislike counts for apology videos not available on YouTube, but captured on the Wayback Machine prior to December 2021 when the change was implemented, were able to be obtained. If an apology video was no longer available on YouTube and was posted after December 2021, their dislike count was unable to be gathered.

A video's likes percentage was calculated with the following formula:

$$Video\ Likes\ Percentage = \frac{Video\ Likes\ Count}{Video\ Likes\ Count + Video\ Dislikes\ Count}$$

Additionally, in May 2019, YouTube announced that channels with over 1,000 subscribers would have their public subscriber counts abbreviated (e.g., 432,930 as 432K and 51,389,232 as 51M) (Team Youtube, 2019). This change took place in September 2019, and greatly affected our methods in gathering subscriber data. Prior to the change, Social Blade was able to track daily changes in subscriber counts. However afterwards, these figures were only updated in set increments.

Since some videos are unavailable, we created a section for video status, which include:

Video Statuses	Description
Original	The video is on the creator's channel and is completely public to view.
Unlisted	The video is on the creator's channel but is hidden. It is accessible through a specific URL.
Private	The video is on the creator's channel but it is hidden. It is inaccessible to the public.
Deleted	The video does not exist on the channel anymore.
Terminated	The YouTuber's channel itself was taken down by Youtube, and as a result, all their videos have been completely removed.

Different statuses have different effects on a creator's channel statistics. When a video is unlisted or private, the views from the video are still included in the channel's all-time view count. However, if a video has been deleted, its views are subtracted from the channel's all-time views.

Visiting the video's URL meant that we could access the video if it was still uploaded or unlisted. Anything that was private, deleted, or terminated, led to a page indicating their status. For these videos, we obtained their metadata through captures on the Internet Archive's Wayback Machine.

Youtube Creator

Basic Information

Some basic information about the video's creator was recorded, including their channel name, legal name, and other online aliases. For example, Daniel Keem is the individual who runs the DramaAlert channel, but is also known as "KeemStar".

Gender

Our only two categories are man and woman, but that is because of all the Youtubers we found, all of them identified as one or the other. This category may evolve to include more gender identities if we grow this dataset in the future.

This was recorded so that we could identify any potential trends in the treatment of different genders online.

Race

We recorded race to see if there was any correlation between racial groups and the reception of their apologies. These categories were determined by consulting data collection guides from Public Health Ontario, which contained the following:

Race Categories	Description/Examples
Black	African, Afro-Caribbean, African-Canadian descent.
East Asian	Chinese, Korean, Japanese, Taiwanese descent.
Latino	Latin American, Hispanic descent.
Middle Eastern	Arab, Persian, West Asian descent, e.g. Afghan, Egyptian, Iranian, etc.
South Asian	South Asian descent, e.g. East Indian, Pakistani, Sri Lankan, Indo-Caribbean, etc.
Southeast Asian	Filipino, Vietnamese, Cambodian, Thai, other Southeast Asian descent.
White	European descent.
Another race category	Another race category (write-in response).

(Public Health Ontario, 2021)

We also looked at the race categories used by Statistics Canada (2017), which were more specific and included groups such as Chinese and Filipino. However, we decided to use the list from Public Health to make it easier to sort by race when analyzing the text, and for our purposes, the specificity of Statistics Canada's categories were not necessary.

Age (at the time of video)

The age of the creator was calculated with their birthday and the date that the apology was posted.

Country of Origin

This is the country in which the creator lives and works.

Channel Type

Many YouTubers have secondary channels that are dedicated to content that is not appropriate for their main channel. For example, someone who makes comedy videos may have a second channel for behind-the-scenes content, outtakes, or vlogs. These secondary channels almost always have lower viewership than a main channel, and posting a video to a second channel may indicate the YouTuber's intent to publicize their scandal and/or their apology.

Channel Genre

We determined several channel genres for categorizing the YouTubers:

Genres	Description/Examples
Beauty	Makeup and fashion.
Gaming	Commentary, reviews, playthroughs.
Lifestyle	Vlogs, storytimes.
Family	Content focused on the family's daily life or raising children.
Commentary	Covering current events in pop culture.
Art/music	Visual arts, musicians.
Health	Fitness, diet, gym.
Comedy	Comedy sketches.
Prank	Dedicated to pranking on friends, family, or the public.
Technology	Reviews, demos.
Food	Cooking, mukbangs.

We included YouTubers from across a wide variety of genres because we were interested in seeing if there were any trends within specific communities.

While some of the genres may seem similar, they are actually distinct types of content. Pranks may be a part of the comedy genre, but there are Youtube channels that are entirely dedicated to pranking.

Date Joined

This is the date that the main channel was created, and helped us gauge the length of the apologist's YouTube career prior to the scandal.

Wayback Machine URLs

Due to limitations stemming from YouTube updates, the YouTube Data API, and EU data regulations on Social Blade, historical subscriber data was not easily accessible. Using code adapted from Anjali Shrivastava (2020), we were able to obtain accurate daily subscriber data using Wayback Machine captures of Social Blade's monthly channel statistic pages. In order to automate the data collection process, we added a column in the spreadsheet with Wayback Machine URLs of the channel's monthly statistics page. The spreadsheet was parsed through Python and the subscriber count was fetched from the Wayback Machine captures.

However, given YouTube's update surrounding the abbreviation of public subscriber counts in September 2019, the code could no longer properly parse Social Blade pages for accurate data (Team YouTube, 2019). As a result, Social Blade URLs were not used in combination with the Wayback Machine if an apology video was uploaded after September 2019.

Instead, the URLs for YouTube channels were added to the column. Captures from the channel URLs in the Wayback Machine were used to determine subscriber counts for various points in time (e.g. Jake Paul had 4061 captures of his channel on the Wayback Machine from 2015 to 2023). By using an archived YouTube channel URL, its front page displayed the channel's subscriber count.

If a channel URL was not sufficient enough to yield subscription results for the periods of time we required, video URLs from both the original apology video, and other channel videos from the YouTuber were put into the Wayback Machine to find the closest matching date. When watching a YouTube video, the page capture displays information such as the number of views, comments, and the subscription count we needed. YouTubers for whom we were able to get subscription information through YouTube videos were indicated in this section with "Videos from Channel put onto Wayback".

Channels with apologies before/during 2019 that did not have a proper archive on the Wayback Machine were labelled "Not archived" as there was no other way to obtain the information.

Reason

The reasons for which viewers demanded the YouTuber's apology video were categorized into the following:

Reason	**Description/Examples**
Beef	Conflict between the creator and another influencer.
Scamming	Scams involving fans losing money and/or receiving low-quality products.
Animal abuse	Physical harm, improper care.
Child abuse	Abuse including emotional and psychological harm.
Lying/misinformation	General lying, plagiarism.
Exploitative content	Faking a death, recording the deceased, claiming to be part of a marginalized community for fame.
Assault/abuse	Towards a spouse or romantic partner, including physical and sexual abuse.
Racism	Blackface, racial slurs.
Infidelity	Towards a spouse or romantic partner.
Grooming	Includes in-person and online.
Harassment	Doxxing, causing fans to target individuals.
Insensitive content	Complaints that viewers considered entitled, mockery, or ignorant.

These were determined throughout our data collection process, adding more categories as needed.

Subscribers

Current Subscribers

All current subscriber counts were taken directly from Social Blade. Within a YouTuber's page on Social Blade, those visiting the page are able to access a section labelled "Live Subscriber Count". This provided a real-time subscriber count for any YouTube channel, which updated every second. Even for channels that were terminated, the live subscriber count would show the subscriber count that they had before termination.

A screenshot of YouTuber PewDiePie's live subscriber count page.

Subscriber Count (main channel)

The subscriber count of the YouTuber's main channel was recorded at different intervals, ranging from one day before the apology was uploaded, to six months later. This was obtained from Wayback captures of Social Blade as well as captures of the YouTuber's channel page.

Results and Discussion

After the data was collected from YouTube, Social Blade, Wikitubia, and the Wayback Machine, the results were visualized through a number of graphs.

Subscriber Counts

Through Python, the available daily subscriber count and total subscriber count were plotted with the Python library matplotlib.pyplot (see Figures 1 and 2). Due to YouTube's changes, these graphs only show the statistics of apologies posted prior to September 2019, which include 36 of the total 80 apologies. Labels were added to highlight the apologies with the most drastic increase and decrease in subscriber counts.

Figure 1.

The Impact of an Apology on Daily Subscriber Changes (36 YouTubers)

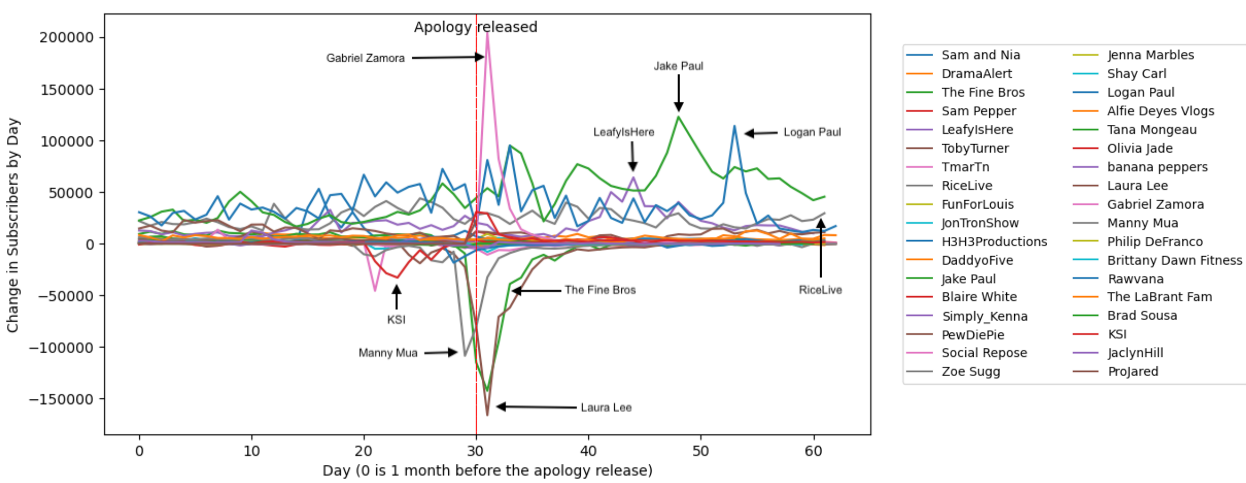
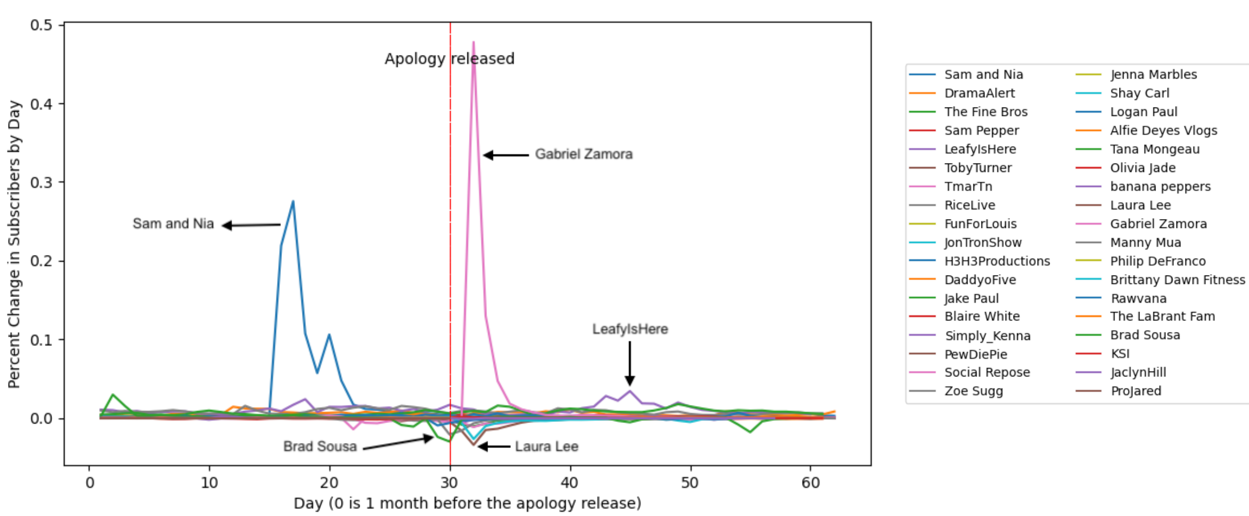


Figure 2.

The Impact of an Apology on Total Subscriber Counts (36 YouTubers)

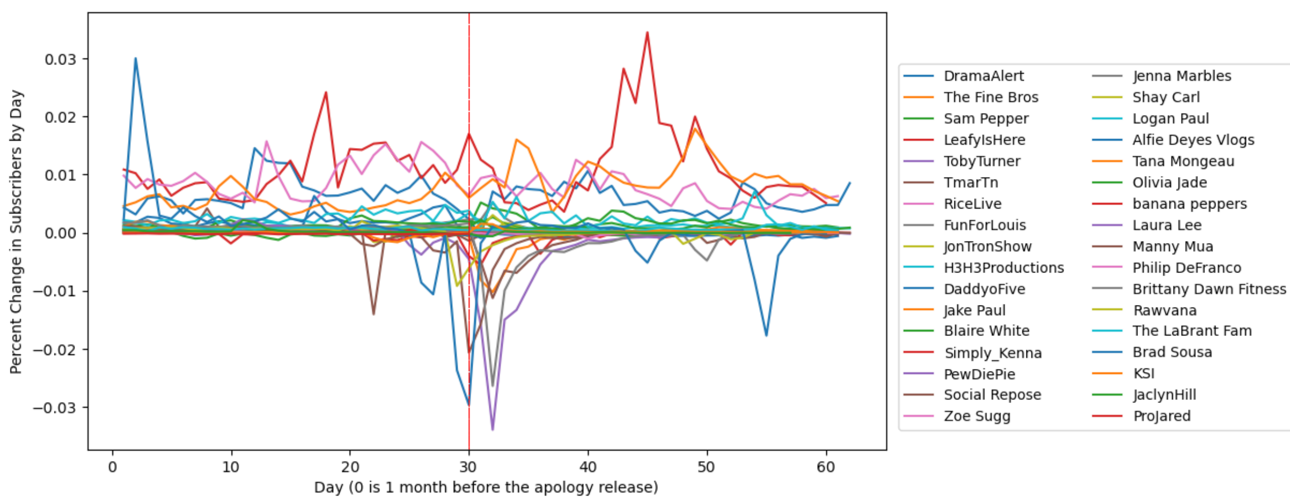


Although the data in Figure 1 were generally in the same range, Figure 2, which plotted the percent change in channel subscribers with respect to the channel's total subscriber count, had two main outliers: Sam and Nia, and Gabriel Zamora. Zamora nearly doubled his total subscriber count in one day while Sam and Nia saw close to a 30% increase in their subscriber count a few weeks before their apology. This increase was due to a video posted on August 5 announcing Nia's pregnancy which went viral, along with another video posted on August 8 sharing news of Nia's miscarriage (McNeal & Zarrell, 2015). Although Sam and Nia had the second largest gain in subscribers in percent change, this accomplishment is diminished in Figure 1, where the amount of subscribers they gained in one day (53,080) pales in comparison to Gabriel Zamora or Jake Paul's highest daily subscriber changes (204,116 and 122,845, respectively).

To have a clearer visual of the other 34 apologies, which were compressed in Figure 2 due to the vertical scale's consideration for the graph's outliers, the outliers were excluded in Figure 3. In comparison to Figure 2, the edited graph was able to visualize data that had been obscured due to the previous vertical scale.

Figure 3.

The Impact of an Apology on Total Subscriber Counts (34 YouTubers)



While the differences between Figures 2 and 3 led us to believe that percent change would be a greater indicator of the apology's reception, it was greatly hindered by the size of the channel. In the case of smaller YouTube channels, percentage change had the positive effect of showing the impact a small number could have on the channel's overall percentage change. For example, in Figure 4, Brad Sousa lost -3,403 subscribers on the day of his apology, which was nearly a 3% decrease in total subscribers and the worst percent change among lifestyle YouTubers. This change was not properly highlighted in the graph representing daily subscriber change (see Figure 5) as the amount he lost was measly in comparison to fellow lifestyle YouTuber KSI who lost the most subscribers (33,000) among his peers. However, where percent change spotlights Sousa, it undermines the number of subscribers KSI lost, as 33,000 subscribers represented less than 0.16% of his twenty million subscribers at the time.

Figure 4.

The Impact of an Apology from Lifestyle YouTubers on Daily Subscriber Changes

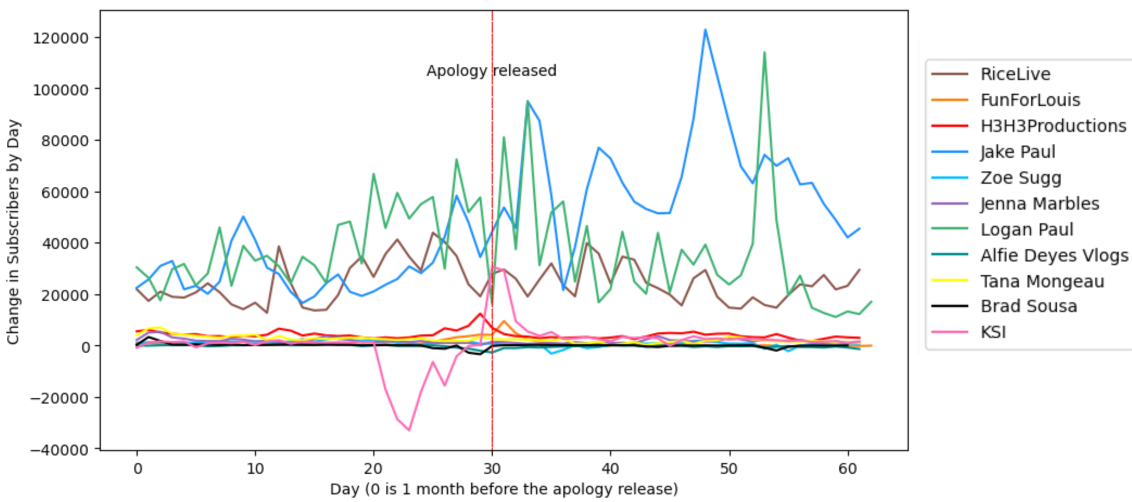
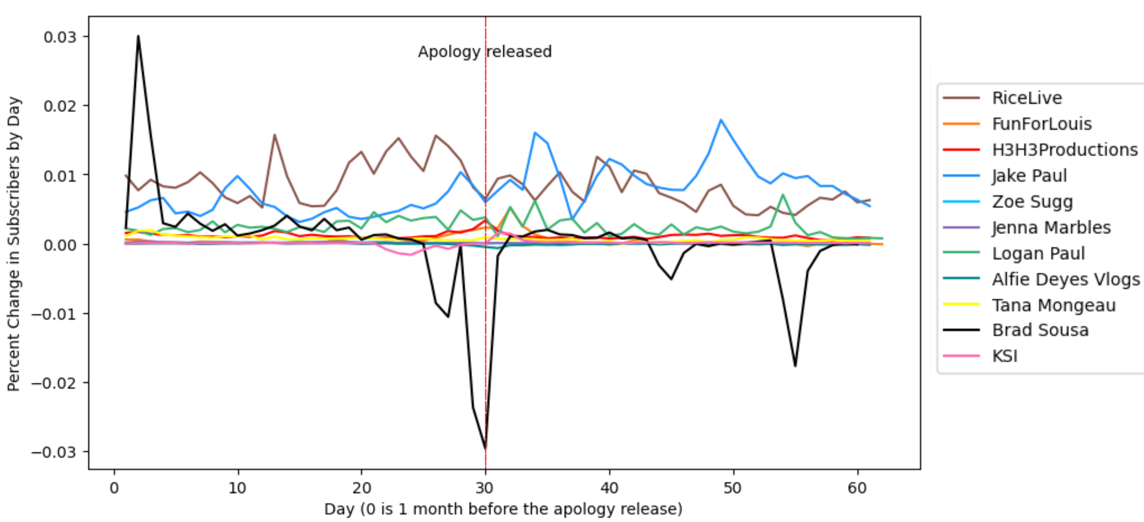


Figure 5.

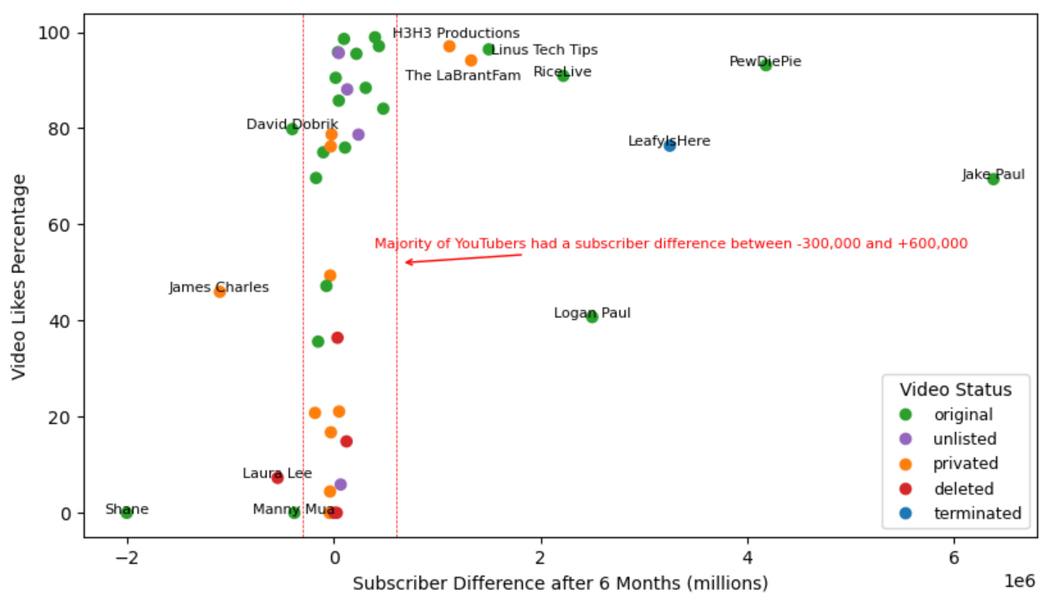
The Impact of an Apology from Lifestyle YouTubers on Total Subscriber Counts



Although graphs using percent change and daily subscriber counts presented limitations in cohesive data visualization, plotting data of different apologies along a longer timeframe yielded interesting results. From the sample of percent change of 34 channels in Figure 3, many YouTubers sharply lost subscribers in the days following their apology. However, on average, channels had monthly growth in subscribers despite their controversies. For apologies posted after September 2019, subscriber counts at set intervals were gathered, including subscriber count a day before the apology and six months afterwards. All apologies with a subscriber count for those two dates were plotted in Figure 6, which includes apologies posted before and after 2019.

Figure 6.

Apology Reception and Subscriber Difference after 6 Months (43 Apologies)



Despite the temporal difference, the majority of the YouTubers fell between the range of losing 300,0000 subscribers to gaining 600,000 subscribers six months after the apology. There was, however, a significant difference in the apology date for subscriber changes outside of that range. With the exception of Linus Tech Tips, all YouTubers who gained over a million subscribers six months after the apology posted their apologies before September 2019. Notably, both Jake Paul and PewdiePie posted their apologies in 2017 and were the most successful channels following their apology, gaining 6.4 million and 4.2 million subscribers, respectively. On the other hand, three of the five YouTubers who lost more than 300,000 subscribers, were posted after 2019. This could be attributed to the rise of cancel culture in 2019, where influencers would quickly become de-platformed by the online community if they made any mistakes that influencers from the past may have been able to get away with. An article from Insider published in September of 2019 detailed this rise in cancel culture, and how it quickly destroyed careers for influencers who were not prepared for it (Dodgson).

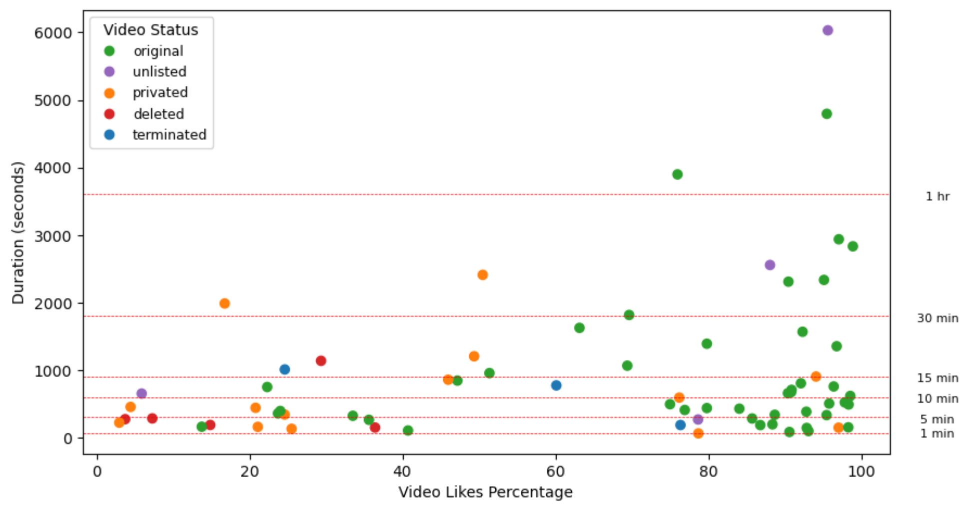
Regardless of the year, there is a strong correlation between the percentage of video likes and overall subscriber impact. This speaks to the importance of addressing controversy as the impact of an apology with more dislikes can be interpreted as negatively affecting brand perception and audience trust. Viewers that do not feel that a YouTuber's apology is genuine are more likely to leave them.

Likes Percentage and Apology Duration

On average, the majority of apologies that were well-liked by viewers were still publicly available on YouTube (see Figure 7). Of the 46 apologies that exceeded 50% in video likes, 78% (36 apologies) were still publicly available in their original form. When taking into account unlisted videos, which are available on a YouTuber's channel but only accessible through the original video link, the percentage of public apologies increases to 85%. For apologies with more dislikes than likes, only 33% (8 apologies) were publicly available or 38% when including unlisted videos.

Figure 7.

The Impact of Duration on Apology Reception by Video Status (70 Apologies)

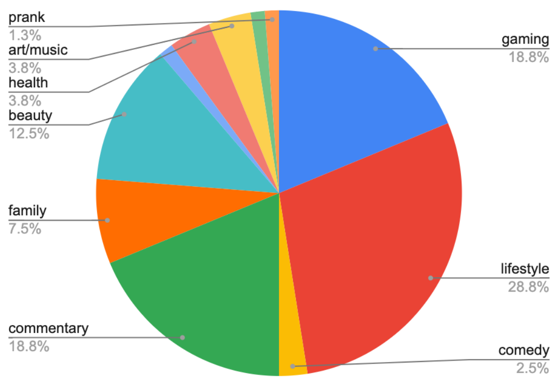


The length of an apology was also found to be indicative of its overall reception. Apologies with less than 50% in video likes had an average length of 8:47 minutes, whereas the average length of well-liked videos was double that, at 19 minutes. Even after accounting for outliers, such as Lindsay Ellis' nearly two-hour-long apology, the median for disliked apologies was 5:49 minutes compared to 10:36 minutes for apologies with over 50% in video likes. In both cases, well-liked apologies were approximately double the length of their counterparts. In general, apology videos that exceeded 30 minutes drastically increased the chances of the apology being well-liked: of the 11 apologies above that threshold, 10 had more likes than dislikes and half of those apologies had over 95% likes compared to dislikes.

Channel Types and Reasons for Apology

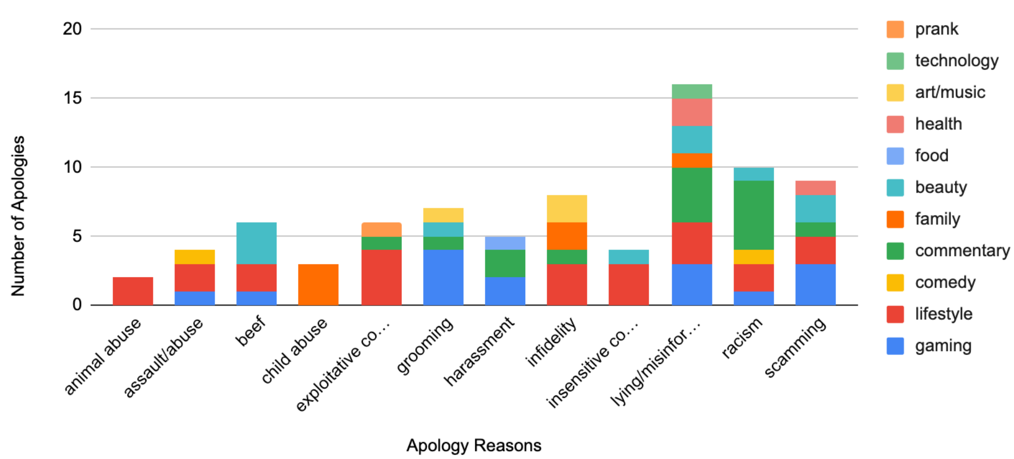
The types of YouTube channels behind the apologies we analyzed predominately represented gaming, lifestyle, commentary, beauty, and family channels (see Figure 8). These categories were the source of 86% of the apologies we looked at. Different genres of YouTube channels attract specialized audiences so we anticipated trends would appear amongst specific audiences and apology reasons.

Figure 8.
Apology by YouTuber Channel Type



There were several trends in reasons for apologies for each channel type (see Figure 9). Apologies for animal abuse and child abuse were exclusively done by lifestyle and family channels, respectively. Lifestyle YouTubers apologized for nearly every reason available and also made up the bulk of apologies for exploitative content, infidelity, and insensitive content. Gaming channels were disproportionately represented in apologies for grooming and scamming, whereas commentary channels had a large presence in apologies for misinformation and racism.

Figure 9.
Apology Reasons by YouTube Channel Type



Although we expected strong correlations between channel type, apology reason, and apology reception, the results were inconclusive. Several patterns emerged but they were not absolute, such as a favourable apology reception by commentary channels and for apologies addressing misinformation (see Figures 10 and 11). While some apologies by YouTubers for infidelity and exploitative content were able to regain the favour of their audiences, those were often ill-received. Apologies for animal abuse were divisive as they had the most polarizing video likes percentage for an apology reason: Brooke Houts received the second lowest likes percentage (3.73%) for her apology while Jenna Marbles received the highest percentage of likes in the dataset (98.92%).

Figure 10.

The Impact of Duration on Apology Reception by Channel Type (70 Apologies)

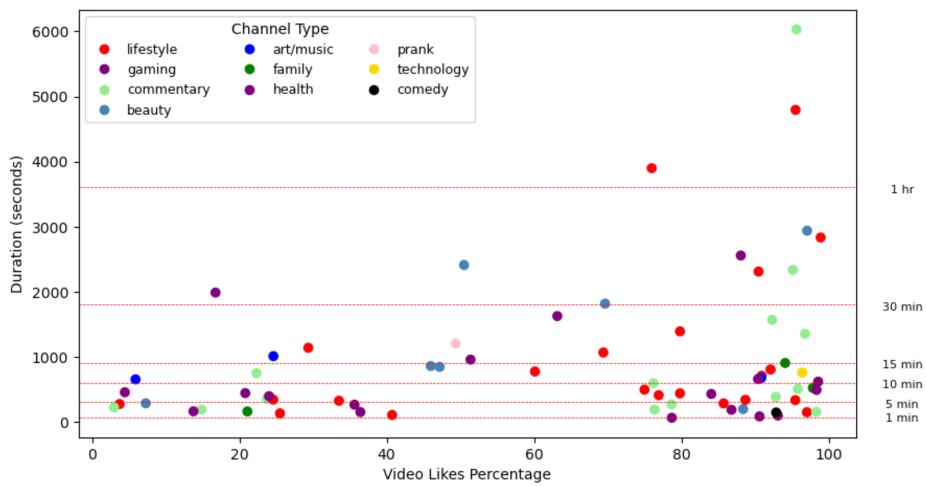
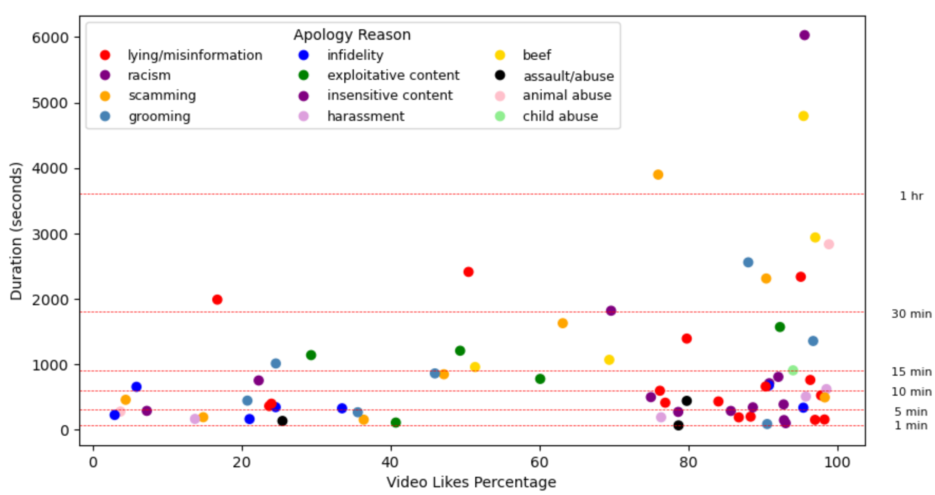


Figure 11.

The Impact of Duration on Apology Reception by Apology Reason (70 Apologies)



Topic Modelling

We used Gensim to experiment with topic modelling. The main variables to manipulate were the number of topics, the number of keywords associated with each topic, and the number of passes (times that the program read over the corpus).

Using NLTK's dictionary of stopwords, we cleaned up a corpus containing the text of all of the apologies. We started with around 20 topics, steadily decreasing it to 10 with each run. We also decreased the number of keywords and increased the number of passes.

After running it 50 times for 10 topic with 10 words each, we ended up with the following words:

Topic 1	***Topic 2***	***Topic 3***	***Topic 4***	***Topic 5***
want, made, could, videos, okay, part, wanted, started, hope, went	going, video, time, get, ever, understand, situation, came, two, everyone	would, even, way, make, see, wrong, years, let, long, else	things, never, something, feel, day, mean, much, friends, anything, us	like, one, sorry, right, everything, life, game, thought, little, always
Topic 6	***Topic 7***	***Topic 8***	***Topic 9***	***Topic 10***
say, think, better, twitter, maybe, shit, already, true, happen, conversation	thing, back, fucking, good, happened, last, talking, work, sure, give	know, lot, kind, also, take, trying, every, many, love, today	really, guys, said, saying, go, still, first, bad, call, thank	people, got, actually, person, stuff, done, yeah, someone, need, making

From these topics, very few actually contain apologetic words such as "sorry" or "wrong". However, from certain groupings of words we can gain a bit of an idea of the subject of the video—for example, in topic 6, there is some indication of regret with the word "better" appearing alongside words such as "say" or "think". In topic 5, "sorry" appears with the words "game" and "life", suggesting that the speaker did not take the subject seriously and "thought little" of their actions.

Future Development

Due to the size of this project and time constraints, there is much room for future development.

Analysis

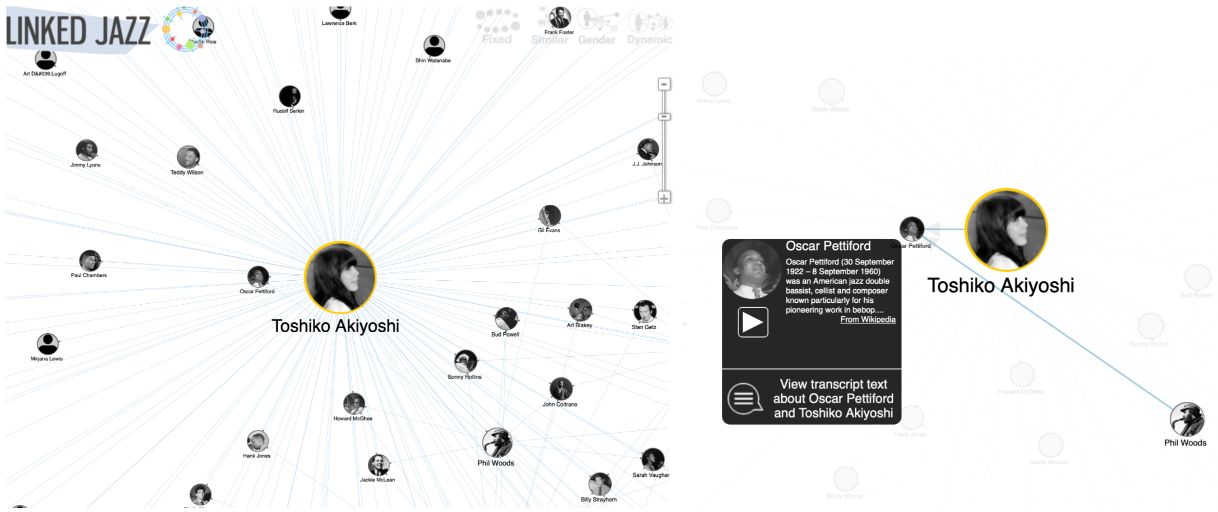
Since a majority of our time was dedicated to compiling the data, we would apply some tools and actually manipulate the data. This could go in several directions, including but not limited to:

- ***Sentiment analysis***: This could include analysis of the apologies themselves, or be extended to news coverage/videos about them.
- ***Topic modelling***: This can be used to see any emerging topics between different creators who have made apology videos, and examine relations to factors such as the reason for the apology or the influencer's genre.
- ***Textual analysis***: Tools such as frequency distribution could be analyzed for similar language and find characteristics of apology videos that are perceived as "good" or "bad".
- ***Visuals***: This can include different non-verbal cues that indicate sentiments such as tears or body movements. Furthermore, we could examine different aspects of the video's production, such as lighting, location, and the number of cuts, which were included in Choi and Mitchell's analysis (2020).
- ***Effects on creator growth***: This could involve a closer analysis of the increase or decrease of a creator's subscribers/views, which could potentially be related to different factors such as the reason for the apology, the creator's race, or gender.

Visualization

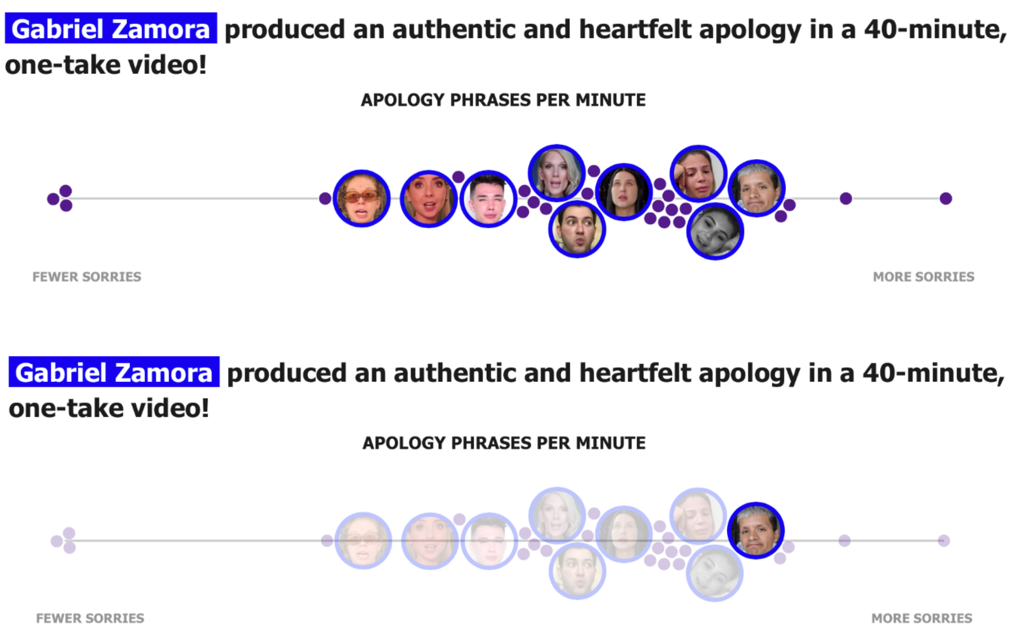
An interactive visualization of our analysis results would be engaging for viewers, as well as representing interpretations of the data in a way that is easier to approach.

We looked to other visualizations for features that we would ideally incorporate into our own work. For example, Linked Jazz (Semantic Lab at Pratt, n.d.) allows viewers to sort by specific individuals, and click on mentions of other artists to view them in context.



A screenshot of LinkedJazz allowing the user to view all of Toshiko Akiyoshi's connections, as well as view them in context.

The YouTube apology visualization created by Kakkar and Russell for Pudding (2020) compares apologies on a spectrum, with specific numbers viewable by hovering over the influencers' images.



Screenshots of the Pudding visualization.

A visualization could be created in a variety of ways, but we would most likely create a prototype with Figma.

Conclusion

This project ended up very different from where it began. Different obstacles such as the lack of a database and lack of available data forced us to change our approach several times, and we eventually spent most of our time collecting and organizing the data. However, from the data we were able to analyze, there were many interesting trends that revealed themselves. For instance, the length of an apology was indicative of its success, as audiences most likely viewed a longer apology video to be more comprehensive and sincere. The work done in this paper can be expanded upon by other researchers for deeper analysis, or an extension of the dataset.

References

Anarios. (2023). Return YouTube Dislike (Version 3.0.0.8) [Computer Software]. *Google Chrome Web Store.* https://chrome.google.com/webstore/detail/return-youtube-dislike/gebbhagfogifgggkldgodflihgfeippi

Asarch, S. (2021, February 21). POWER RANKING: the 10 most famous influencers on the internet. *Insider Inc.* https://www.insider.com/most-famous-influencers-insider-data-logan-paul-pewdiepie-2021-2.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. *Reilly Media, Inc*.

Cifor, M., Garcia, P., Cowan, T.L., Rault, J., Sutherland, T., Chan, A., Rode, J., Hoffmann, A.L., Salehi, N., Nakamura, L. (2019). Feminist Data Manifest-No. https://www.manifestno.com/.

Chun, W. H. K. (2008). *Updating to Remain the Same: Habitual New Media*. MIT University Press.

Dodgson, L. (2019, September 28). *YouTubers are calling out the platform's 'cancel culture' that subjects them to a rampant hate mob and sees them lose thousands of subscribers in a matter of hours.* Insider. Retrieved from https://www.insider.com/cancel-culture-what-it-means-creators-on-youtube-2019-9

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Hund, E. (2023). *The Influencer Industry: The Quest for Authenticity on Social Media*. Princeton University Press.

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, 2007.

Kakkar, A., Russell, S. (2020). The Aftermath of a YouTube Apology. [Visualization]. *The Pudding*. https://pudding.cool/2020/01/apology/.

Karlsson, G. (2020). The YouTube Apology. *Malmö universitet/Kultur och samhälle.* http://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-23129.

Makalintal, B. (2019). How YouTubers Turned the Apology Video Into a Genre. *Vice*, https://www.vice.com/en_us/article/ywykzb/how-youtubers-james-charles-jaclyn-hill-pewdiepie-turned-the-apology-video-into-a-genre.

McNeal, R., & Zarrell, R. (2015, August 12). Doctors Cast Doubt On Viral Video Stars Sam And Nia's Pregnancy Claims. *Buzzfeed News*. Retrieved from https://www.buzzfeednews.com/article/stephaniemcneal/people-are-doubting-vloggers-sam-nias-viral-pregnancy-announ

The Pandas Development Team (2020). pandas-dev/pandas: Pandas (Version 1.5) [Python library]. Zenodo. https://doi.org/10.5281/zenodo.7794821.

penguinz0. (2020, December 20). Youtuber Apology Tier List [Video]. *YouTube*. https://www.youtube.com/watch?v=Eq72iEjcU4w.

Public Health Ontario. (2021). Collecting Information on Ethnic Origin, Race, Income, Household Size, and Language Data: A Resource for Data Collectors. *Public Health Ontario*. https://www.publichealthontario.ca/-/media/documents/ncov/he/2021/03/aag-race-ethnicity-income-language-data-collection.pdf?la=en.

Python Software Foundation. Python Language Reference (Version 2.7). http://www.python.org.

Radford, A. & Kim, J. W. (2022). Robust speech recognition via large-scale weak supervision. *GitHub*. Retrieved from https://github.com/openai/whisper.

Rehurek, R., & Sojka, P. (2011). Gensim: Topic modelling for humans. [Python library]. Retrieved from https://radimrehurek.com/gensim/.

Semantic Lab Pratt, (n.d.). Linked Jazz [Network visualization]. https://linkedjazz.org/network/.

Statistics Canada. (2017). Visible Minority and Population Group Reference Guide, Census of Population, 2016. *Statistics Canada*. https://www12.statcan.gc.ca/census-recensement/2016/ref/guides/006/98-500-x2016006-eng.cfm.

Sutherland, T. (2021). "Remains". In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.

Social Blade. (n.d.). FREQUENTLY ASKED QUESTIONS (FAQ). *Social Blade*. https://socialblade.com/youtube/help/data-retention

Team YouTube (2019, May 21). Early heads up: abbreviated public subscriber counts across YouTube. *YouTube Help*. https://support.google.com/youtube/thread/6543166.

We Are Social, & DataReportal, & Meltwater (January 26, 2023). Most popular social networks worldwide as of January 2023, ranked by number of monthly active users (in millions) [Graph]. *Data Reportal*. https://datareportal.com/reports/digital-2023-global-overview-report.

Wikitubia. (n.d.). Home. *Fandom*. https://youtube.fandom.com/wiki/YouTube_Wiki.

The YouTube Team. (2022, October 10). Introducing handles: A new way to identify your YouTube Channel. *YouTube Official Blog*. https://blog.youtube/news-and-events/introducing-handles-a-new-way-to-identify-your-youtube-channel/.

The YouTube Team. (2021, November 10). An update to dislikes on YouTube. *YouTube Official Blog*. Retrieved from https://blog.youtube/news-and-events/update-to-youtube/.

Bibliography

Acker, A.. (2021). “Metadata”. In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.

Chun, W. H. K., & Barnett, A. (2021). *Discriminating data : correlation, neighborhoods, and the new politics of recognition*, MIT Press.

Chun, W. H. K. (2021). “Proxies”. In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.

Cifor, M. (2021). “Affect”. In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.

Drucker, J. (2021). “Visualization. In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.

Long, S. & Baker, J. (2019). The Elusive Digital / Critical Synthesis. *Debates in the Digital Humanities*. Retrieved from https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/72423b53-060d-4927-aa0b-36c90b03152c#node-fa619674283bc0e9aa2567f971c530ac87c852e9.

Roberts, S. (2021). “Abuse”. In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.

Wernimont, J.. (2021). “Quantification”. In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.

Zer-Aviv, M. (2021). “Obfuscation”. In N. Thylstrup (ed.) *Uncertain Archives*, MIT University Press.