

Recommendations for Using Tik Tok as a Science Communication Tool by Utilizing the App's Analytics data and Creating Displays for that Data

Allinta Tadesse

University of California, Santa Barbara

Abstract:

The intention of this research paper is to supplement the university's knowledge of social media strategies with data analytics such as sentiment analysis and retention for certain science communication genres that can be tracked via a hashtag. In addition, it is also important to consider the constant updates and algorithms that come with using Tik Tok as a primary social media platform, and that fluctuations in user data is normal. For this analysis, we will be examining six different tik tok videos that encompass environmental issues and gather information regarding the audiences concern regarding the subject of the tik tok video. Utilizing sample data from Tik Tok's API and creating a framework for how content can be analyzed through my own videos, I was able to utilize Python to generate different ways to display data. Such as sentiment analysis to gauge how users interact with discourse, and retention to examine at which time stamp are people most engaged with the video In addition to these videos, a set of twenty science related hashtags were used to generate data regarding Tik Tok users' sentiments towards particular science topics; these Tik Tok hashtags were amongst the most popular science related Tik Tok hashtags on the app. The data collected on these tik toks can serve as a framework for how UCSB institutions would like for social media to be utilized in order to create engaging content for UCSB students and nonUCSB students that may come across the same content.

Literature Review:

The rise of social media usage during the COVID-19 pandemic came as a consequence of staying home due to social distancing; relying on utilizing technology to share information and remain connected with other people, according to *TikTok: An Emergent Opportunity for Teaching and Learning Science Communication Online*¹. Furthermore, the pandemic has also proved how significant of a role social media plays in how information and misinformation is spread through multiple groups on the social media platform. It has shown the importance and impact of social media for STEM professionals in explaining science and medicine in a widely accessible, accurate, and engaging way. Multiple studies across different approaches have shown that Tik

¹ <https://journals.asm.org/doi/10.1128/jmbe.00236-21>

Tok can be used as an effective tool for science communication by serving as a platform for science topics to be shared quickly and in language that is accessible to the general public.

Researchers that authored the article *Assessing the Content Topics of the Educational Videos on Tik Tok for Science communication*, utilize a clustering method and VOSviewer tool to examine the effectiveness of communicating various science topics such as climate change, space, human biology, environmental issues, and scientific experiments across 80 educational videos that specialize in science communication². The findings have shown to prove that Tik Tok has the potential to promote scientific literacy by encouraging engagement and retention of specialized topics such as sustainability and public health, in addition, the authors included that there it is necessary for continued research to show the effectiveness of the social media app's ability to effectively communicate these topics. The data collected for this research is from the app 'Douyin' essentially the Chinese version of Tik Tok that is made available to Chinese citizens and is virtually inaccessible to US residents due to the requirement of a Chinese phone number that is needed to register with the app. Although the structure and format of the two apps are virtually similar and can even be considered the same platform, essentially, the data is not necessarily usable or applicable to this analysis. However, what can be learned from the researchers that can be applied to the scope of this analysis is their methodology for data collection and display.

The methodology used to examine the effectiveness of science communication used keyword search, content type, and frequency distribution. The 80 videos examined were collected by searching for keywords such as "science" and "education" on Tik Tok (or in this circumstance, the app 'Douyin'). They were posted between July and September of 2020 from thirty different TikTok accounts that center their content around science communication. To identify the topics of the videos, qualitative coding was used to sort the videos into six main categories: earth science, biology, physics, chemistry, mathematics, and technology; subtopics were also included in the coding search. Analyzing the frequency distribution of these six core topics including subtopics allows the data to be described in percentage format and frequency counts that help describe the significance of a certain science topic based on outreach and user interest.

Similarly, the article *Static vs. dynamic methods of delivery for science communication: A critical analysis of user engagement with science on social media*, uses a methodology that compares the effectiveness of static and dynamic methods when communicating science topics via social media. Static posts are characterized by text and images, while dynamic posts are those that contain video or moving elements such as animations. The study uses both qualitative and quantitative methods to analyze the data that was collected via Facebook. Qualitative methods included identifying themes and patterns between user engagement and the post itself along with

² <https://www.atlantis-press.com/proceedings/isemss-22/125981960>

the topic of the post, in addition to a coding method used to classify comments based on their content. The quantitative approach included insights data from the Facebook page, along with statistical tests such as t-tests to determine if there was a correlation or a significant difference between the static and dynamic posts.³

Another approach that would be interesting to explore would be using a combination of mixed-methods and survey-based research utilized in the study found in the article *Assessing the Content Topics of the Educational Videos on Tik Tok for Science Communication*⁴. Survey-based research involves data collection from users themselves through insights on their demographics, behaviors, attitudes, and motivations. In the case of this capstone project, we would have access to the audiences' information via Tik Tok's user analytics. Furthermore, it is important to consider the other audience which would include campus staff and faculty of various departments in the sciences that may want to create their own Tik Tok account that covers details about their program in order to increase retention amongst students that are interested or fully pursuing a STEM degree at UCSB. Mixed-methods research includes combining survey-based research and content analysis to gain a better understanding of how the app engages with various audiences, but also for an audience that may not be as well versed in Tik Tok, but is looking for innovative ways to communicate science topics with the general public. In regards to having the ability to systematically receive and analyze data from Tik Tok, it may be difficult to standardize data collection according to *A Systematic Review of Early Approaches to Studying TikTok*, due to the centrality of Tik Tok's For You Page that constantly generates new content that the user may or may not be interested in.⁵

Methods for collecting data

Users that post content to the app Tik Tok also have access to their creator analytics, this serves creators by tracking the performance of their Tik Toks and gauging how audiences interact with their content. By opening a video that is already posted, the user is able to select the 'more data' tab in the bottom right corner of their video. This will take them to a page called 'video analysis' that allows users to see how one particular video performs under two major categories: Performance and Viewers. The Performance tab includes data such as total views, total likes, comments, shares and saves. In addition, the app generates two different graphs titled 'Video views' and 'Retention rate'. Video views allows for users to see the number of times views have watched their video within a week date range since the view has been posted. Retention rate serves to show the percentage of viewers that watch the video at a particular time stamp,

³ <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0248507>

⁴ <https://www.atlantis-press.com/proceedings/istemss-22/125981960>

⁵ <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/d2f247c1-0fe4-4c53-a956-0440f8d71346/content>

depending on the duration of the video. It is important to also note that the retention rate can only be calculated for the first three minutes of that video.

For my data collection and analysis, I examined data from videos posted across the span of a few months. I wanted to create a model for how social media analytics, specifically for tik tok, that can be utilized in the future by other groups interested in gauging what their audience on social media looks like and how they may be able to improve retention. By visiting the web version of tik tok, I was able to access my analytics and download it as a csv file to be later imported to excel. From that excel file, I transported that data into my Jupyter notebook to use python to get some more information about the performance of the videos through descriptive statistics, sentiment analysis and possibly a regression analysis.

Data analysis and visualization is essential for extracting insights and understanding the performance of the content posted from various data sets.

Sentiment analysis is also essential for content creators, companies, and essentially anyone looking to use social media for the purpose of business and user retention i.e. various UCSB institution groups in this case. When asked, sentiment analysis can provide a response to this question “How positively or negatively are users speaking about a particular topic?”, according to *Strategic Listening: A Guide to Python Social Media Analysis*⁶. In fact, this article provided some insight for how Python coding can be used for social media marketing and analysis, by using sentiment analysis and topic modeling in order to gauge how users gravitate towards particular topics.

In order to prepare for looking at my own data from videos that are centered around science, mainly environmental related topics, In the article, *Strategic Listening: A Guide to Python Social Media Analysis*, provides insight for how Python coding can be used for a social media marketing and analysis, by using sentiment analysis and topic modeling in order to gauge how users gravitate towards particular topics.

Methodology with Coding: Sampling from Popular Science Related Hashtags onTik Tok using Python

Collecting and Displaying Data on Python

I started with using a downloading an xls file of twenty of the most commonly used Tik Tok hashtags that are science related onto an excel sheet, which I later converted into csv file. I transferred that csv file from an excel data base onto Python, in order for me to generate a table

⁶ <https://www.toptal.com/python/python-social-media-analysis>

for all of the hashtags including quantity of posts, total views (views), and average view per post. I started on Python by importing pandas into my notebook, known as the code 'import pandas as pd', once the table with the data was transferred into the Python notebook, I further simplified the table by selecting the first five rows with three columns examining the categories: 'Hashtag', 'Views', and 'Average view per post'.

```
In [26]: df.loc[1:5, ['Hashtag', 'Views', 'Average view per post']]
```

Out[26]:

	Hashtag	Views	Average view per post
1	#scicomm	1,538,245	7056.0
2	#science	2,979,644,774	46079.0
3	#womeninstem	11,406,641	13051.0
4	#stem	97,444,008	3291.0
5	#womeninscience	588,349	1862.0

Figure 1: Top 5 hashtags with highest views and average view per post: #scicomm, #science, #womeninstem, #stem, #womeninscience

With the simplified chart above, I sought out to display the data visually through a bar graph, generated with this code:

```
plt.bar(x, y)
```

```
plt.xlabel('Hashtag')
```

```
plt.ylabel('Average view per post in percentage by x1000')
```

```
plt.title('Gauging which Hashtag is Most successful on tik Tok through Average view per post')
```

I also titled the graph: 'Gauging Which Hashtag is Most Successful on Tik Tok through Average view per Post'

Gauging which Hashtag is Most successful on tik Tok through Average view per post

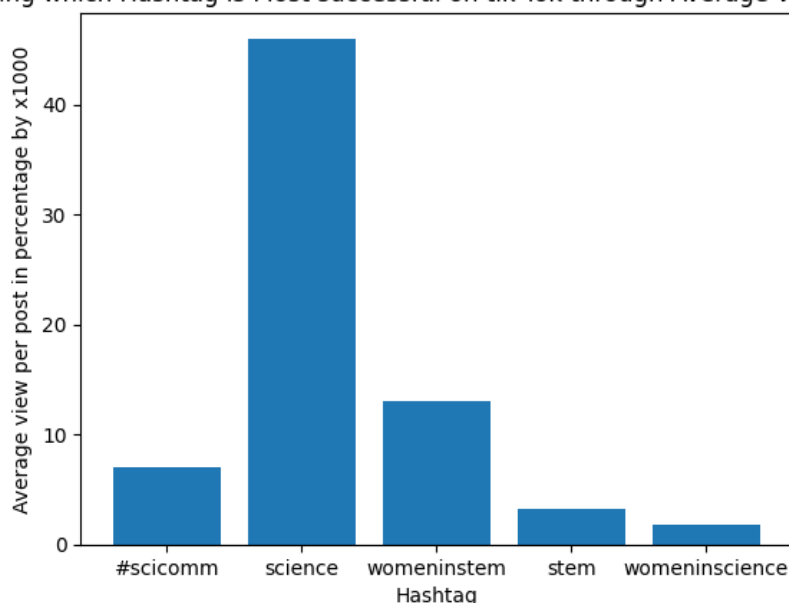


Figure 2: Gauging Which Hashtag is Most Successful on Tik Tok though Average View Per Post
 Note: Since the average view per post was in the thousands, I reduced these values to a decimal by multiplying by one thousand in order for the data to fit on the y-axis of the bar graph.

What does the ‘average view per post’ tell Tik Tok users? According to the app itself⁷, this and other categories such as total views, total likes, comments, shares, traffic source type, and percentage of viewers who watched the entire video can help content creators gain insight into what type of content are engaging for people who come across their page. Performance across various Tik Tok videos is prone to fluctuation which is normal according to the app. However, noticing differences in average watch time and these other analytics can help users learn what captures their audience’s attention the most.

Sentiment Analysis: Sample Tik Tok Data

Sentiment analysis is a rising metrics tool amongst users in order to gauge customers, or in this case audiences, emotional response to a particular keyword. Essentially, “...social media sentiment analysis provides context for your customers’ conversations around the social space,⁸”. Although sentiment analysis is most commonly used for marketing a product or a service. Arguably, sentiment analysis is also essential for science communicators to understand how their audiences react emotionally to a particular topic. Nowadays with highly politicized topics such as Climate Change and recently with the COVID-19 pandemic, it is imperative that science communications learn how their audience reacts to specific science topics when moving forward

⁷<https://www.tiktok.com/creators/creator-portal/en-us/tiktok-content-strategy/understanding-your-analytics/>
⁸<https://www.tiktok.com/creators/creator-portal/en-us/tiktok-content-strategy/understanding-your-analytics/>

with planning for their content. Following the coding guide provided by ‘Strategic Listening’ article, I was able to prompt Python into generating the average polarity and average subjectivity of the top five science communication Tik Tok hashtags. In addition, the article mentions the `polarity_scores` method, “...which takes a string as input

Moving forward with the sentiment analysis on Python, I started by importing Vader (Valence Aware Dictionary and sEntiment Reasoner) into Python in order to proceed with the rest of the code. Once Vader was uploaded, I was able to re enter my csv file data to ensure that Python was looking at the data I was interested in which included the top five most popular science related hashtags: #scicomm, #science, #womeninstem, #stem, and, #womeninscience. I then prompted AI to generate prompts to model what Tik Tok comments would possibly look like under videos that fall along the above hashtags.

views = ['I love Scicomm!', 'Science is amazing!', 'Women in STEM rock!', 'STEM fields are fascinating', 'I am a scientist']. The sentiment analysis I generated in python generated average polarity and average sentiment for each keyword as a hashtag. The overall sentiment scores is based on a scale from -1 to 1, where -1 is the most negative response and 1 is the most positive. Below is the chart where the average polarity and average sentiment scores for each hashtag was calculated.

```
In [50]: print(sentiment_df)
```

	Keyword	Average Polarity	Average Subjectivity
0	Scicomm	0.625	0.60
1	Science	0.750	0.90
2	womeninstem	0.000	0.00
3	stem	0.700	0.85
4	womeninscience	0.000	0.00

Figure 3: Average Polarity and Average Subjectivity scores for each keyword or ‘hashtag’

Another method of viewing sentiment can be shown through the code below⁹, and prompting a sample phrase that may or may not have been mentioned in some tik toks that use the above hashtags. Polarity refers to the “output that lies between [-1,1], where -1 refers to negative sentiment and +1 refers to positive sentiment. Subjectivity is the output that lies within [0,1] and refers to personal opinions and judgments,”¹⁰

⁹ <https://stackabuse.com/sentiment-analysis-in-python-with-textblob/>

¹⁰ <https://www.topcoder.com/thrive/articles/getting-started-with-textblob-for-sentiment-analysis#:~:text=Polarity%20is%20the%20output%20that,to%20personal%20opinions%20and%20judgments.>


```
sentence = 'Climate Change is real'  
analysis = TextBlob(sentence).sentiment  
print(analysis)  
Sentiment(polarity=0.2, subjectivity=0.30000000000000004)
```

Python Analysis with my own Content: Using Sentiment Analysis for examining discourse

Being on social media, its inevitable that users and content creators both will encounter discourse and/or engage in some shape or form. Science communication spaces are not devoid of discourse, despite the usual objective nature of science, people still find a way to argue against some concepts that have been proven as fact due to preconceived notions or solely for the sake of participating in discourse online. Using sentiment analysis can help users gauge which comments to interact with discourse in the comments of their Tik Tok post, in addition to their overall feelings towards the content that was posted.

Discourse has been present in the internet space since its fruition, and arguably exacerbated by the lockdowns that occurred during the height of the COVID-19 pandemic. By using code generated by Aman Kharwal¹¹, I was able to examine sentiment that could be exchanged when interacting with discourse on social media. Since my data had already been imported into my Python notebook, I was able to start with the sentiment analysis for one of my videos where I discussed the trail derailment in East Palestine Ohio; the video features large bolded text with the phrase “we are living through the biggest environmental disaster in US history and barely anyone is talking about it???”. Immediately upon reading this sentence, it may seem like a comment that is asking for discourse, which was essentially my purpose with posting this Tik Tok video. Amongst all of my science related Tik Tok videos that I have made, this one recieved the highest overall engagement with 399 likes, 14 comments, three shares, and 1761 total views since the video has been posted. Overall, it also received the second highest retention rate with nearly over 20% of viewers watching the entire video. While majority of the comments left under my Tik Tok video shared the same sentiment as me regarding the environmental crisis that occured in Ohio that resulted in the annihilation of at nearly 43,00 aquatic animals¹². Seeing how other science communicators on the app were also discussing how there is a deficit in transparency from government agencies and the general public as to how this disaster would result in long term health implications, I brought it to myself to see how people who regularly or dont interact at all with my content would react to this information.

¹¹ <https://thecleverprogrammer.com/2022/04/12/tiktok-reviews-sentiment-analysis-using-python/>

¹²<https://abcnews.go.com/amp/Technology/environmental-disasters-affect-ecosystems-ohio-train-derailment-affect/story?id=97493398>

This sentiment analysis was calculated with using polarity and subjectivity. Below is the code and polarity and subjectivity scores for my caption ‘we are living through the biggest environmental disaster in US history and barely anyone is talking about it???’ , in addition to this comment that was left under my video ‘We are so tired that’s why something negative on the news everyday it’s so traumatizing our way of coping is normalizing’.

Sentiment from my ‘caption’ or the block of text I attached to my video:

sentence = 'we are living through the biggest environmental disaster in US history and barely anyone is talking about it???'

analysis = TextBlob(sentence).sentiment

print(analysis)

Sentiment(polarity= 0.05, subjectivity= 0.1)

Sentiment from comment I chose to respond to when the video was posted

(likes received: 4)

Code from my ‘caption’ or the block of text I attached to my video:

sentence = 'We are so tired that’s why something negative on the news everyday it’s so traumatizing our way of coping is normalizing’

analysis = TextBlob(sentence).sentiment

print(analysis)

Sentiment(polarity= -0.3, subjectivity= 0.5666666666666668)

Sentiment from my response to the comment (likes received: 3)

sentence = 'Believe me I know... but we as the public need to have enough critical thinking to properly consume news media cuz we have a right to know abt this'

analysis = TextBlob(sentence).sentiment

print(analysis)

Sentiment(polarity= 0.05714285714285714, subjectivity= 0.4004761904761905)

Although my video clearly expresses my own frustration regarding the lack of transparency behind this event, I understand that this sentiment comes from my positionality as an Environmental Studies major. However, this sentiment analysis result indicated a slightly positive polarity of 0.05, and an objectivity score of 0.1. Based on the definitions provided above, not only is the video’s caption relatively subjective, it is also mildly positive. Bringing to the conclusion that responses may be variable.

The user that left the comment “*We are so tired that’s why something negative on the news everyday it’s so traumatizing our way of coping is normalizing*”, received polarity score of -0.3 and a subjectivity score of 0.566 in the sentiment analysis, clearly holding negative connotations regarding the video’s subject matter i.e. trail derailment in East Palestine Ohio. The user’s

comment suggests a negative and more subjective response to the video that had been posted. Reading the comment also shows the visible frustration and exhaustion that the user holds when constantly encountering bad news. A common epidemic amongst Gen Z also exacerbated by the hyperconnectivity the COVID-19 lockdowns provided, the users response is understandable.

Science in nature is objective, it is how scientists and others with a strong scientific literacy often communicate science topics and news to the general public. However, the highly politicized nature of some science genres, most notably environmental centered topics, may pose complications for communicators in a virtual setting. The response comment I left under that users comment yielded a positive polarity of 0.057 and a subjectivity score of 0.4. The response left under the user's comment acknowledges their fatigue "*Believe me I know...*", and it also emphasises the importance of critical thinking when consuming news media. "*...but we as the public need to have enough critical thinking to properly consume news media cuz we have a right to know abt this,*". Ultimately, sentiment analysis is a beneficial tool for collecting data regarding how individuals react emotionally to content posted. However, it is also important to consider that prompting emotion based code in Python, it not the same as emotion that is communicated between people. Learning how to communicate science on an empathetic level is important for improving people's perceptions of science.

Video Retention in Short form Tik Tok Video

Referring to the *Static vs dynamic methods of delivery for science communication: A critical analysis of user engagement with science on social media*, their methodology characterises videos as dynamic methods of communicating science on social media i.e. video or moving elements. Following their methodology, I created a 15 second nonlecture video featuring dynamic shots of a gardening day for the AS Black Women's Health Collaborative at UCSB. By featuring a combination of text and b roll of organization members participating in gardening activities, the overall video received 41 likes and 309 views.

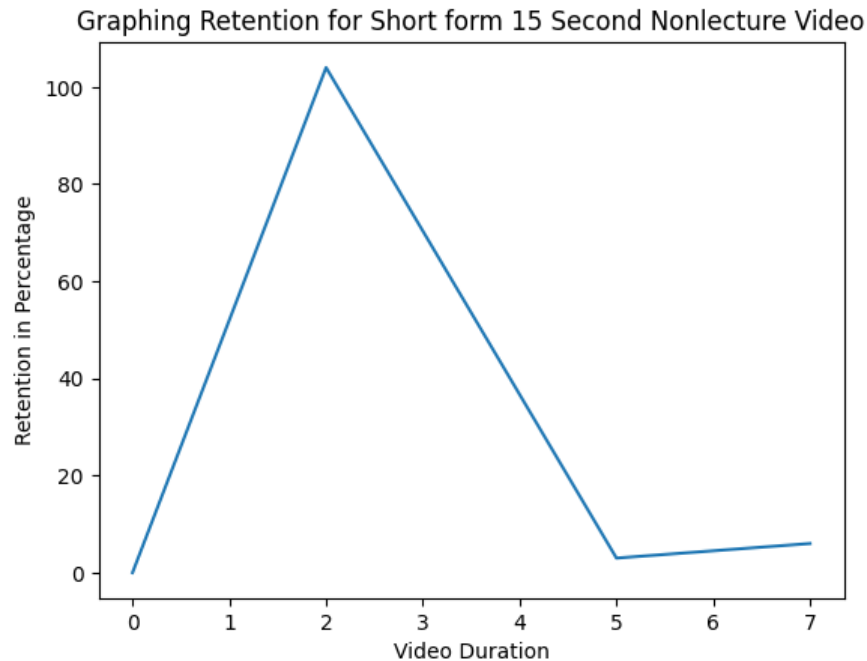


Figure 4: Graphing Retention for Short form Dynamic method for communicating science, 15 second nonlecture video utilizing visuals and text to display information. The x-axis and y-axis was taken from the graph provided by the app's analytics section that is automatically calculated after every post.

This video was created in shifting the common narrative of science. It highlights the transformative power of science as a means of self-care, enabling individuals to find solace and tranquility in nature's embrace. Delving deeper into the narrative, I wanted to weave together multiple elements to create an intersectional perspective by centering a marginalized group engaging in science as a community. Engaging in scientific practices, such as cultivating plants and understanding their needs, can extend beyond the realm of mere knowledge acquisition. It highlights the transformative power of science as a means of self-care, enabling individuals to find solace and tranquility in nature's embrace.

Reflection and Comments:

As we are shifting in strategies for communicating various science topics to the general public, the social media Tik Tok can provide users with sufficient data to learn what content results in higher user traffic. By utilizing Tik Tok's video and user performance analytics in order to gauge how Tik Tok users engage with certain science topics, and analytics tools such as retention rate and sentiment analysis can assist in how content can be shifted in order to increase user engagement and retention. By utilizing research methodologies such as clustering methods,

qualitative coding, and mixed-methods approaches, we are able to further assess the effectiveness of science communication on Tik Tok.

This research paper aims to enhance the understanding of social media strategies through data analytics, specifically sentiment analysis and retention, for science communication genres on TikTok. The analysis focuses on six TikTok videos covering broad and specific science themes with the purpose of engaging various audiences' perspectives towards particular science topics. By leveraging sample data from TikTok's API that was downloadable; In addition to Python programming, data was collected and various visualization techniques were employed to analyze the performance of the videos.

Additionally, a set of popular science-related hashtags on TikTok was used to gather insights into user sentiments regarding specific science topics. Methodologies employed in these studies from the literature review including keyword search, clustering, and mixed-methods research can also provide valuable insights for data collection and analysis. For data collection, TikTok's creator analytics were utilized, providing performance metrics such as views, likes, comments, shares, and retention rate. Python programming was employed to further analyze the collected data, including descriptive statistics, sentiment analysis, and potential regression analysis. In regards to implications, it is important to note that data analysts are required to apply Tik Tok API in order to have access to their entire catalogue of data.

Data analysis and visualization play a crucial role in extracting insights and understanding content performance. Sentiment analysis, a valuable tool for understanding audience response, was employed to gauge users' emotional reactions to specific science-related topics. By using the VADER tool, the average polarity and subjectivity scores were calculated for the top five science communication hashtags on TikTok. This analysis helps science communicators understand audience sentiments and preferences, facilitating content planning and engagement.

Furthermore, the study explores the role of sentiment analysis in assessing discourse within the science communication space on social media. Discourse is inevitable on social media platforms, and sentiment analysis assists in identifying relevant comments for engagement and understanding overall audience perceptions of the content.

Conclusion:

In conclusion, this research provides insights into social media strategies and data analytics for science communication on TikTok. By utilizing sentiment analysis and retention rate analysis, content creators can better understand audience sentiments, improve engagement, and effectively communicate science topics. The findings serve as a framework for universities and institutions,

including UCSB, to create engaging content for students and the general public through social media platforms. Future research can build upon these methodologies to further explore the effectiveness of TikTok as a science communication tool and enhance audience engagement.