Are Youtube Videos Influencing Sentiment and Awareness Towards Climate Change?

Hemanth Allamaneni, Jaelon Braxton, Shivani Chowdhry

EPPS 6356.501 Data Visualization

Dr. Karl Ho

December 8, 2023

Abstract

Climate change is a matter slowly becoming a forefront concern of the planet. The industrial changes brought upon by human activities have slowly changed the equilibrium of Earth and has encouraged adverse weather factors that were not as prevalent in the past. While humanity has made progress in terms of surviving on this planet, we have also made great progress in terms of communication. Digital media has allowed for greater interconnectivity as well as documentation for discussing our surroundings. The power of recording is now within the hands of the people, as cameras have been integrated into many devices, including the mobile phones that have become our companions. People are more abundantly able to upload and both curate and create content that follows their narrative. While that may be the case, footage speaks louder than words as it is a medium of a plethora of documented stills. This research paper looks into the power of video narratives' effect on changing climate change sentiment. By creating a corpus of scraped data from the popular social media platform YouTube, we analyze and visualize how the platform is discussing the issue our entire planet faces. To comprehend the impact of climate change media, we selected various educational accounts to decipher how this digital landscape is evolving in terms of increasing sentiment on the topic. In terms of accomplishing an overall picture we made word clouds, bubble charts, and line graphs following 6 youtube channels.

Introduction

Earth's landscapes have vastly changed, with man made concrete infrastructures sprawling over the globe into every corner. These structures are the renovations behind settlements that add to humanity's overall comfort. While this may be a benefit for mankind, some of these industrial advances add to the planet's discomfort. The burning and utilization of

Nonrenewable resources like fossil fuels increase carbon levels and deteriorate our planet's ozone. Icecaps, melted from global warming have increased both our planet's sea levels and ocean born humidity. Our planet has slowly become less balanced, with global temperatures rising over 2° since the industrial revolutions of the1800s. As a result of the byproducts of human activity(Dahlman and LuAnn, 2023), we are experiencing more adverse weather effects at a greater frequency. While these industrial and technological advances have created these negatives, positives emerge in terms of how much our tech connects the 8 billion people on the planet. The advent of the internet and mobile devices have allowed for people across the globe to have access to humanity's pooled knowledge and thoughts. Physical distance has all but disappeared thanks to how our communication devices bridge these distances. Authorship and content creation has also become more accessible. As tech has slowly evolved to be more compact and affordable people are able to broadcast their thoughts on a magnitude unthinkable a century ago. Social media have now become the social hubs for trading thoughts, and technology has been further crafted to better document the things surrounding everyone's life. Some of these surroundings however, include the state of our planet, which fortunately has been deserving of more screen time. While our surroundings are getting better documentation, they also are being discussed in different ways. Some have turned a blind eye to the research being done to make a plea for our planet, some denounce our planet is changing all together, while others have taken measures to better educate the public on the how and why behind climate change. These voices emerge sometimes to political interests, and sometimes out of concern from the implications of humanity not altering their path and caring more for the planet. The advent of mobile technologies and further documentation of the planet has led to the inspiration behind our argument that video media are accelerating positive sentiment towards discussing climate

change. While both opposing ideologies over this planet have what may be assumed equal voice in discussing their opinion, they are essentially reporting over the same curated content and events. Our research argues that the recent increased amount of documentation and greater amount of authorship are helping to legitimize the adverse weather phenomena and emergence of more frequent natural disasters such as floods, hurricanes, wildfires, and unseasonal weather(Rebecca, 2023). Video, of the media documentation has increasingly taken, triumphs storytelling of the past, as it provides little for imagination against textual heavy mediums like newspapers and tweets. Studies have also suggested that film based is more effective in terms of persuasion than text(Wittenberg, et al., 2021) We aimed our focus at educational videos on Youtube to analyze how effective their online content is at discussing and delivering climate concerns.

Research Question: To what extent have climate change related videos impacted climate change sentiment through awareness and how are channels evolving to better convey their narratives?

Null Hypothesis (H0): YouTube channels do not impact climate change sentiment or awareness with their videos and are not finding alternate methods to increase their following.

Alternative Hypothesis (Ha): YouTube channels do increase positive climate change sentiment and awareness with their videos and are finding alternate methods to increase their following.

Tools Used

Our technical choices were driven by the need for versatility, efficiency, and specialized capabilities. Jupyter Labs and RStudio catered to the diverse language requirements, while Pandas, time libraries, and SQL addressed data manipulation and database operations. Visualization libraries and sentiment analysis tools enhanced our ability to communicate findings

effectively. The below provided and described comprehensive toolset empowered us to conduct a thorough analysis of YouTube data, from statistical insights to sentiment exploration.

1. Environments/IDEs: Jupyter Labs and RStudio

- **Decision Reasoning:**
    - **Jupyter Labs:** Jupyter Labs provides an interactive and flexible environment for data exploration and analysis. Its support for multiple languages, including Python, makes it suitable for our diverse data processing needs. The notebook format allows for documentation alongside code, enhancing collaboration and reproducibility.
    - **RStudio:** RStudio is a powerful integrated development environment specifically designed for R. We utilized RStudio for tasks involving R, such as sentiment analysis. Its features like R script editor, data viewer, and integrated plotting tools streamline R-based analysis.

2. Data Manipulation and Analysis: Pandas, Time Libraries

- **Decision Reasoning:**
    - **Pandas:** Pandas is a fundamental library for data manipulation and analysis in Python. Its Data Frame structure aligns well with tabular data, enabling us to efficiently handle and process information about YouTube channels and videos.

- **Time Libraries:** Time libraries in Python (e.g., datetime) are crucial for handling temporal aspects of our data. They allow us to convert and manipulate timestamps, which is essential when analyzing trends over time.

3. Data Visualization: Visualization Libraries (e.g., HBR Themes, Virdis)

- **Decision Reasoning:**
    - **HBR Themes and Virdis:** Visualization is key to understanding patterns and trends in our data. HBR themes provide aesthetically pleasing and publication-ready visualizations. Virdis, known for perceptually uniform colormaps, ensures accurate representation of data in plots and enhances interpretability.

4. Sentiment Analysis: R with Word cloud, Topic models

- **Decision Reasoning:**
    - **R for Sentiment Analysis:** R is well-suited for sentiment analysis tasks, offering specialized libraries such as word cloud and topic models. These tools allow us to extract insights from user comments on YouTube videos, helping us gauge audience sentiment and preferences.

5. Database Operations: SQL

- **Decision Reasoning:**
    - **SQL:** SQL plays a pivotal role in handling structured data. We used SQL liberally for creating, querying, and manipulating tabular data. Leveraging the relational

database model ensures efficient organization and retrieval of information, supporting the seamless integration of diverse datasets.

Data Extraction

This project required us to extract YouTube data via their API functionality:

The below described suite of functions facilitates the extraction of valuable insights from YouTube channels, videos, and user comments, providing a comprehensive approach for data analysis and reporting.

1. **get_channel_stats(youtube, channel_ids)**
    - *Description:* This function retrieves comprehensive statistics for multiple YouTube channels.
    - *Parameters:*
        - **youtube**: A YouTube API client object.
        - **channel_ids**: A list of YouTube channel IDs.
    - *Returns:* A Pandas DataFrame containing channel statistics, including title, subscriber count, view count, video count, and upload playlist ID.
2. **get_video_ids(youtube, playlist_id)**
    - *Description:* Retrieves a list of video IDs from a specified playlist of a YouTube channel.
    - *Parameters:*
        - **youtube**: A YouTube API client object.

        - **playlist_id**: The ID of the playlist.

    - *Returns:* A list of video IDs from the specified playlist.

3. **get_video_details(youtube, video_ids)**

    - *Description:* Fetches detailed statistics for multiple YouTube videos.

    - *Parameters:*

        - **youtube**: A YouTube API client object.

        - **video_ids**: A list of YouTube video IDs.

    - *Returns:* A Pandas DataFrame containing various statistics for each video, such as channel title, title, description, tags, published date, view count, like count, favorite count, comment count, duration, definition, and caption.

4. **get_comments_in_videos(youtube, video_ids)**

    - *Description:* Retrieves top-level comments as text from multiple YouTube videos (up to the first 10 comments per video due to API limits).

    - *Parameters:*

        - **youtube**: A YouTube API client object.

        - **video_ids**: A list of YouTube video IDs.

    - *Returns:* A Pandas DataFrame containing video IDs and associated top-level comments in text.

It is important to note certain measurements had requirements as listed below:

- The code utilizes the Google API Python client (**googleapiclient.discovery**).
- Prior authentication and API key setup are assumed to be completed.
- Error handling is implemented, addressing scenarios where comments are disabled on a video.

YouTube Data Structure:

By combining real-time data generation through API endpoints and the construction of static tables with limiters, our multi-pronged approach provided a nuanced understanding of YouTube data. This methodology allowed us to explore the dynamic nature of YouTube content while facilitating in-depth analyses across predefined time intervals. The comprehensive dataset formed the foundation for our insights into the evolving landscape of YouTube channels and videos.

This approach involved the generation of data instances through API endpoints and the construction of static tables with specific limiters on YouTube data. The aim was to create a comprehensive snapshot of the YouTube ecosystem at specific points in time, enabling a detailed analysis over pre-established time intervals.

1. Static Data Generation Using API Endpoints:

- *Objective:* The utilization of YouTube Data API endpoints facilitated the dynamic extraction of real-time information about channels, videos, and comments.

- *Methodology:* By interfacing with the API, we programmatically retrieved up-to-date statistics, metadata, and comments. This real-time data generation allowed us to analyze the most current state of YouTube content and engagement.
- *Advantages:*
    - **Freshness of Data:** The use of API endpoints ensured that our datasets were reflective of the latest information available on YouTube.
    - **Dynamic Exploration:** Real-time data retrieval enabled us to adapt to YouTube's evolving landscape.

2. Construction of Static Tables with Limiters:

- *Objective:* To complement the dynamic data instances, we implemented a strategy to construct static tables with specific limiters, capturing the state of YouTube data at predefined intervals.
- *Methodology:* We applied limiters, such as date ranges and specific parameters, to restrict the scope of data collection. This approach allowed us to create static tables representing YouTube's landscape at distinct points in time.
- *Advantages:*
    - **Temporal Analysis:** Static tables served as snapshots, facilitating a time-based analysis by comparing data across predetermined time frames.
    - **Controlled Parameters:** Applying limiters allowed us to focus on specific aspects of interest, ensuring targeted insights.

Integration and Analysis:

- *Data Integration:* The dynamic data instances generated via API calls and the static tables constructed with limiters were integrated into a unified dataset, providing a holistic view of YouTube data over time.
- *Analysis:* This integrated dataset allowed us to perform multifaceted analyses, exploring trends, patterns, and changes in YouTube channel statistics, video metrics, and audience sentiments.

Possible Data Storage Optimization

While YouTube provides APIs for dynamic data retrieval, the platform does not inherently offer straightforward methods for seamless integration into proprietary databases without manual construction. To address this challenge, two standard solutions come to the forefront: Data Warehouses and Data Lakes.

1. Data Warehouses:

- **Definition:** A data warehouse is a centralized repository that allows for the storage and analysis of structured data from various sources.
- **Application to YouTube Data:**
    - *Structured Storage:* Data warehouses are well-suited for structured data, making them ideal for housing tabular data representing YouTube channel statistics, video metrics, and user interactions.
    - *Analytics-Focused:* Data warehouses are optimized for analytics and reporting, providing quick and efficient querying capabilities for extracting insights from YouTube data.

- **Advantages:**
    - *Performance:* Data warehouses are designed for high-performance query processing, enabling rapid analysis of large datasets.
    - *Structured Analysis:* Suited for structured data, facilitating straightforward analysis and reporting on YouTube metrics.

2. Data Lakes:

- **Definition:** A data lake is a storage repository that holds raw, unstructured, or semi-structured data, offering a more flexible approach to data storage.
- **Application to YouTube Data:**
    - *Diverse Data Types:* YouTube data often spans beyond structured formats, including unstructured data such as video content, comments, and images. Data lakes accommodate the diverse nature of YouTube data.
    - *Schema Flexibility:* Data lakes allow for schema-on-read, enabling the inclusion of data with varying structures, promoting flexibility in accommodating evolving data requirements.
- **Advantages:**
    - *Scalability:* Data lakes can handle vast amounts of raw data, scalable to the immense volume of content generated on YouTube daily.
    - *Adaptability:* Suited for handling both structured and unstructured data, accommodating the varied data types present on the YouTube platform.

Considerations and Integration:

- **Integration Challenges:** While data warehouses excel in structured data analysis, accommodating raw and unstructured data often requires additional preprocessing. Data lakes, on the other hand, provide flexibility but may require careful structuring for efficient analytics.
- **Hybrid Approach:** Combining the strengths of both data warehouses and data lakes in a hybrid architecture allows for comprehensive storage, processing, and analysis of YouTube data. Raw data can be ingested into a data lake, and structured, refined data can be migrated to a data warehouse for analytics.

In the context of YouTube data analysis, the choice between data warehouses and data lakes hinges on the nature of the data and the specific analytical requirements. A hybrid approach, integrating the strengths of both solutions, offers a robust strategy for storing, processing, and deriving valuable insights from the diverse and dynamic YouTube dataset and is something to be explored in future iterations of our research.

Considerations for future data structure optimization

In the ideal scenario, structuring YouTube data in a proprietary, tree-like format offers several advantages, providing the necessary flexibility for nuanced data manipulation and storage. A tree structure allows for tailored filtering through individual method calls, adaptable formatting based on specific needs, and storage in a non-centralized format. This approach also facilitates efficient and optimized data loads, with time complexity geared towards achieving optimal results. Unfortunately, the project scope was significantly constrained by the limitations imposed by YouTube's API access and analytics functionality in its free developer version.

1. Proprietary Tree-Like Data Structure:

- **Advantages:**
    - *Flexibility:* A tree-like structure enables dynamic data filtering, allowing for selective retrieval and processing based on specific criteria or use cases.
    - *Adaptability:* Formatting data on a needs basis becomes more intuitive with a tree structure, accommodating the evolving requirements of YouTube data analysis.
    - *Non-Centralized Storage:* A tree format allows for decentralized storage, distributing the data across nodes or branches, minimizing dependencies on a centralized structure.

2. Individual Method Calls and Time Complexity:

- **Advantages:**
    - *Granular Filtering:* Structuring data for individual method calls provides a granular approach to data retrieval, allowing for precise and targeted information extraction.
    - *Optimized Time Complexity:* By tailoring the data structure to specific needs, time complexity is optimized for efficient results, enhancing overall performance in data processing and analysis.

3. YouTube API Limitations:

- **Endpoint Call Restrictions:** YouTube imposes limitations on the number of endpoint calls allowed within a specified time frame, restricting the dynamic retrieval of real-time data.

- **Analytics Functionality Constraints:** The free developer version of YouTube's API may lack advanced analytics features, limiting the depth and breadth of insights that can be extracted from the platform.

Project Scope Constraints:

- **Scope Limitations:** The ambitious goal of implementing a proprietary, tree-like data structure was hindered by the constraints imposed by YouTube's API restrictions.
- **Analytical Limitations:** The free developer version of YouTube's API may lack certain analytics functionalities, restricting the depth and granularity of data that can be extracted and analyzed.

Considerations for Mitigation:

- **Hybrid Approach:** Considering the limitations, a hybrid approach integrating available data structuring options while respecting API restrictions can help achieve a balance between project goals and platform constraints.
- **Optimization Strategies:** Implementing optimization strategies, such as caching and strategic data loading, can mitigate the impact of API limitations on the overall project.

While the ideal scenario involves structuring YouTube data in a proprietary, tree-like format for optimal flexibility and performance, the practical implementation faces challenges due to YouTube's API limitations. Understanding these constraints is essential for refining project objectives, exploring alternative structuring options, and implementing mitigation strategies to achieve meaningful insights within the platform's constraints.

Text based analysis.

The objective of this analysis and primary focus of our research is to examine learner characteristics in user comments on the chosen climate change YouTube channels by utilizing sentiment analysis and topic modeling techniques. Scholars have been attempting to comprehend the crucial element of knowledge co-creation, which is purportedly taking place in the user comments on these videos. Studies indicate that people's learning and knowledge acquisition processes are increasingly influenced by interactive digital cultures on social media and informal learning.

Word Cloud Analysis

This section looks to analyze: What are the main keywords found in users' comments on climate change-related YouTube videos.

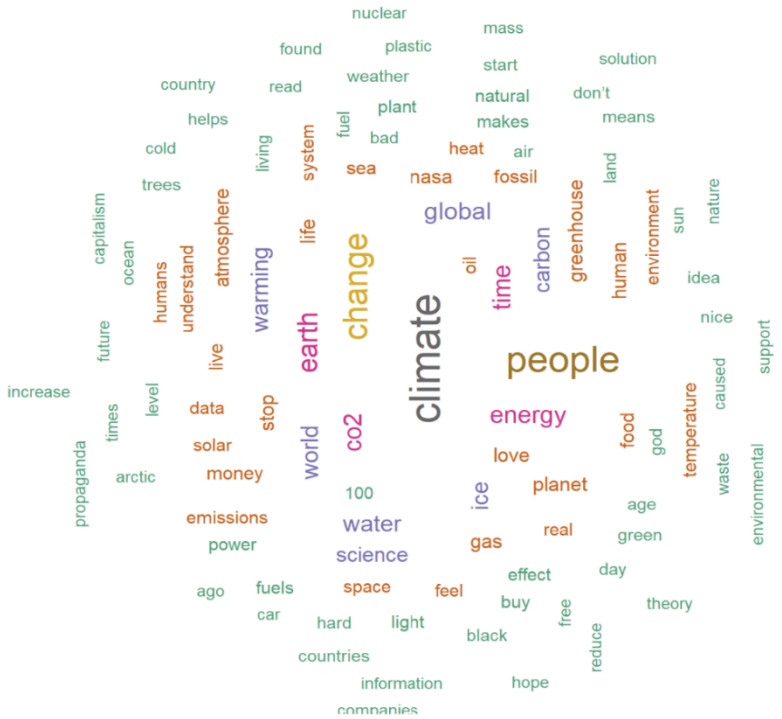
The color-coded word cloud above shows the key themes or most important words (keywords) within comments of the corpus. The keywords can be categorized into three important recurring themes in the order of their frequency of use in the comments:

The color-coded word cloud above shows the key themes or most important words (keywords) within comments of the corpus. The keywords can be categorized into three important recurring themes in the order of their frequency of use in the comments:

1. Words signifying the **scientific process of climate change** such as: 'earth', 'Co2', 'time', 'energy', 'global', 'carbon', 'warming', 'ice', 'water', 'science.'

2. Words related to the **causes and effects of climate change**: ‘people’, ‘love’, ‘planet’, ‘greenhouse’, ‘environment’, ‘human’, ‘fossil’, ‘solar’, ‘emissions’, ‘gas’, ‘food’, ‘air’, ‘heat’.

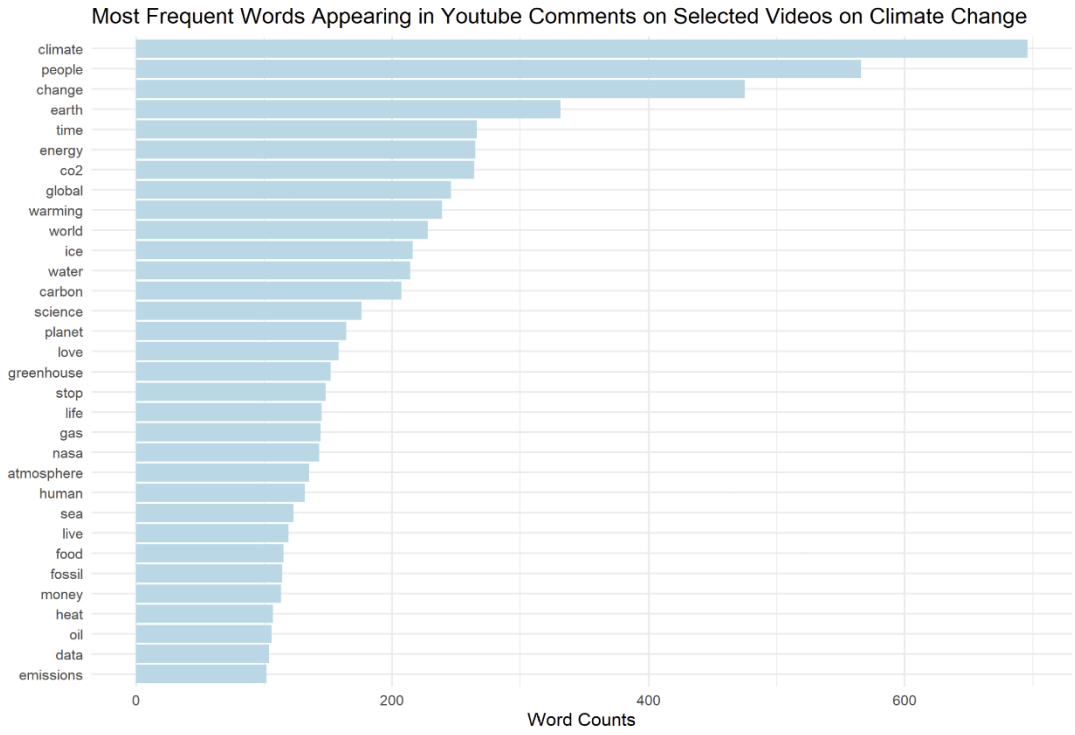
3. Words discussing the larger **political-geopolitical forces driving climate change**: ‘Capitalism’, ‘future’, ‘power’, ‘propaganda’, ‘companies’, ‘countries’, ‘waste’, ‘support’, ‘solution’, ‘money’, ‘information’, ‘fuels’, ‘reduce.’



Word Frequency Bar Chart

Below is a Word Frequency Bar Chart which shows the most frequent words appearing in YouTube comments and the frequency of their use in the comments. This chart was made after

filtering for the words which were repeated more than 100 times to show only the frequencies associated with the most frequently used words in the comments section.



Clearly, the most frequently used words other than ‘climate’, ‘change’ and ‘people’ are scientific-process words like ‘earth’, ‘time’, ‘energy’, ‘co2’, ‘global’, ‘warming’, ‘ice’, ‘water’, ‘carbon’, ‘greenhouse’. Then the second category of words are those that signify the effects or repercussions that climate change is having on the lives of people. These include ‘love’, ‘planet’, ‘life’, ‘stop’, ‘human’, ‘live’, ‘food’, ‘money’, ‘heat’. Lastly, some words can signify what the effects of human actions are in accelerating the process of climate change, for example, ‘energy’, ‘human’, ‘fossil’, ‘emissions’, ‘oil’.

The word ‘people’ which is a top keyword possibly indicates a focus on human involvement or impact in the context of climate change. This could include discussions on responsibility, behavior, or actions of individuals and communities. The word ‘energy’ which is also a top

keyword might relate to discussions about renewable energy sources and sustainable energy practices.

In summary, these keywords collectively paint a picture of a diverse set of concerns, perspectives, and calls to action related to climate change. The inclusion of emotional keywords like "love" alongside more technical terms like "CO2" suggests a multidimensional conversation that spans scientific, emotional, and practical aspects of addressing climate change. Analyzing the context in which these keywords are used, and their co-occurrences can provide even more nuanced insights into the discussions happening in the YouTube comments.

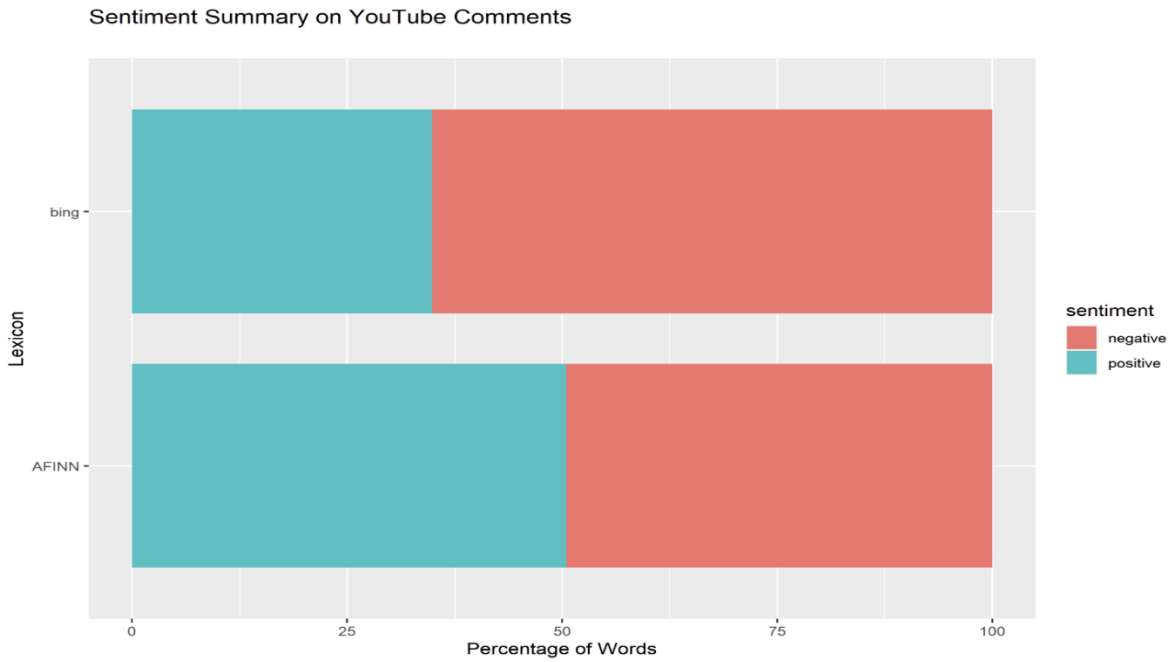
Sentiment analysis of user comments about climate change:

We used the AFINN and BING lexicons to analyze user sentiment on the videos in order to determine the polarity of the sampled comments. According to the BING lexicon, the results show as in figure below that, the general tone of the comments is negative in their discussions about climate change. However, according to AFINN lexicon data, there was almost a 50:50 split between the positive and negative sentiments in the users' comments about climate change. If the positive and negative sentiment words are in the same proportion in a sentiment analysis, it suggests that, on average, the sentiment of the YouTube video comments on climate change is neutral, according to the AFINN lexicon. In a neutral sentiment scenario, the total sentiment score is close to zero because the positive sentiment words cancel out the negative sentiment words, resulting in a balanced or neutral sentiment.

The AFINN lexicon is a more nuanced quantitative lexicon which assigns a numerical sentiment intensity score to each word score. On the other hand, BING lexicon is qualitative in the sense it only categorizes words into positive and negative, without distinguishing them into

different levels of positivity or negativity. Therefore, we choose to go with the results of the AFINN lexicon rather than BING lexicon.

Therefore**,** we believe that the answer to the Research Question : “What key themes are prevalent in the user comments?” is that, because the positive and negative words are balanced, it indicates **a lack of strong positive or negative sentiment, or a neutral overall sentiment of people** toward climate change in the video comments. However, it is very important to note here that, while this analysis provides an overall sentiment score for the corpus, diving deeper into specific comments or subsets of data may reveal more nuanced sentiments or specific themes within the discussions on climate change on YouTube.



Topic Modelling of Comments Corpus

To model and investigate the possible themes in the comments, the topic modeling process was used. It began with the creation of a Document Term Matrix, which was utilized in conjunction with Latent Dirichlet Allocation (LDA).

Latent Dirichlet Allocation (LDA) is a probabilistic generative model commonly used for topic modeling. LDA assumes that each document is a mixture of topics and that each word in the document is attributable to one of the document's topics. The model learns two probability distributions: (i) the distribution of topics across all documents, and (ii) the distribution of words across all topics. Every comment was treated as a document in this instance, and each comment was treated as independent in the dataset and datasets are given unique IDs.

We preprocessed the text data by tokenizing it into words or phrases, removing stop words and custom stop words, and performing other necessary cleaning steps. The reason we had a list of custom stop words, such as "climate", "change", "videos" removed was because we did not want them to be included in the final topic models as these are the obvious keywords used in all the comments about the topic of climate change, and do not tell us specifically what themes or specific topics people are discussing about climate change.

After the cleaning, a document-term matrix was created, where each row represents a document or comment, each column represents a unique term (word), and the matrix entries contain the term frequencies or other weighting schemes.

Next, for model training, we choose several topics (K) for the model to identify. Since this is a hyperparameter, it needs some experimentation to get the optimal K. For this, we first used the K=7 and 6, and then studied the probabilities of words belonging to the identified themes. Further, we used the LDAvis topic browser to explore the distribution of the emergent

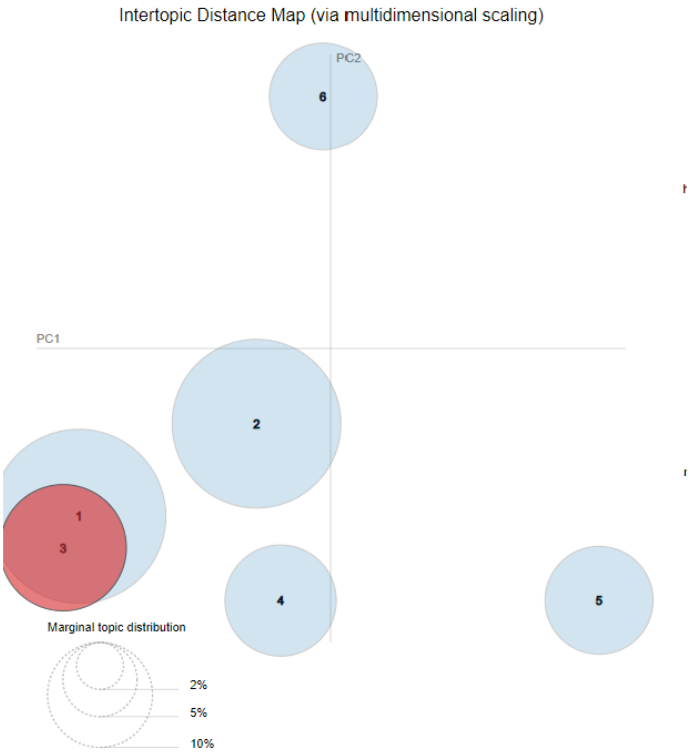
theme words, which showed that two topics had significant overlap in the words belonging to them as shown in the figures below. This was an indication that selecting K=7 or 6 was not an optimal choice, so in the end we decided to go with k=5 topics to model in our data.

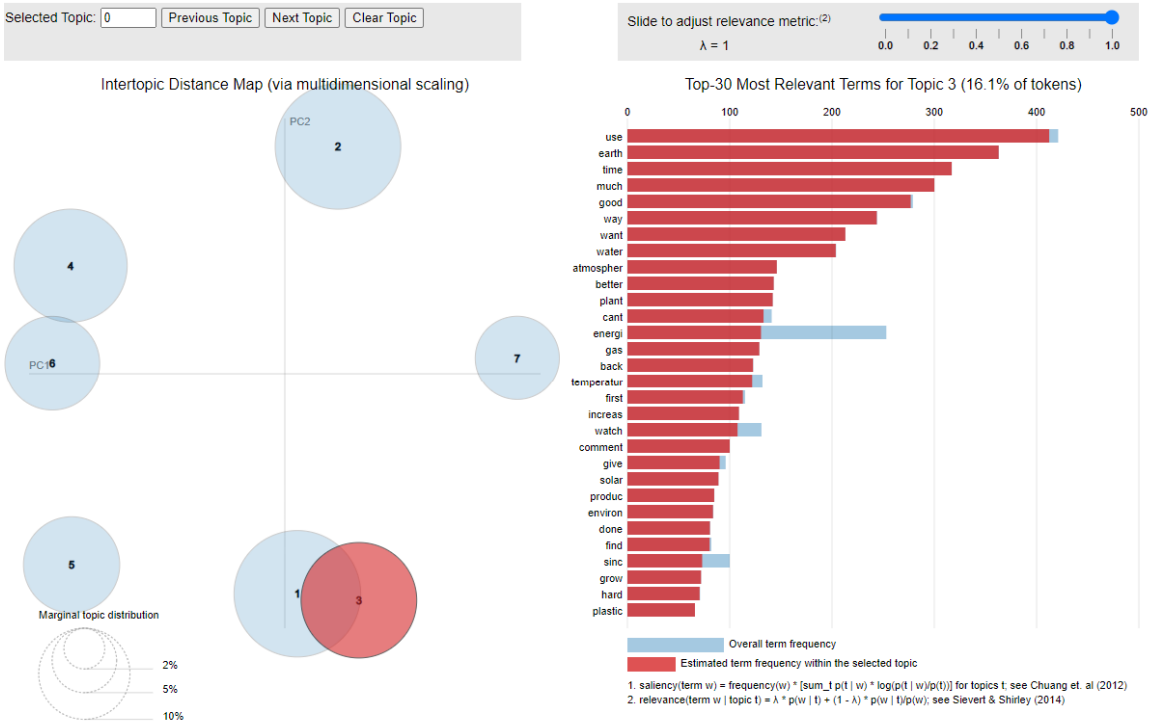
LDAvis topic browser showed significant overlap between topics 1 & 3 in case of both k=7 and k=6 topics:

All the above variables have a positively skewed histogram. Using the ln () function to take the logarithmic values of variables helps address problems like uneven variability or non-normal distribution of errors in a statistical model. This method aims to improve the accuracy and precision of the model's estimates by transforming the data, making it easier to handle and potentially enhancing the model's performance.

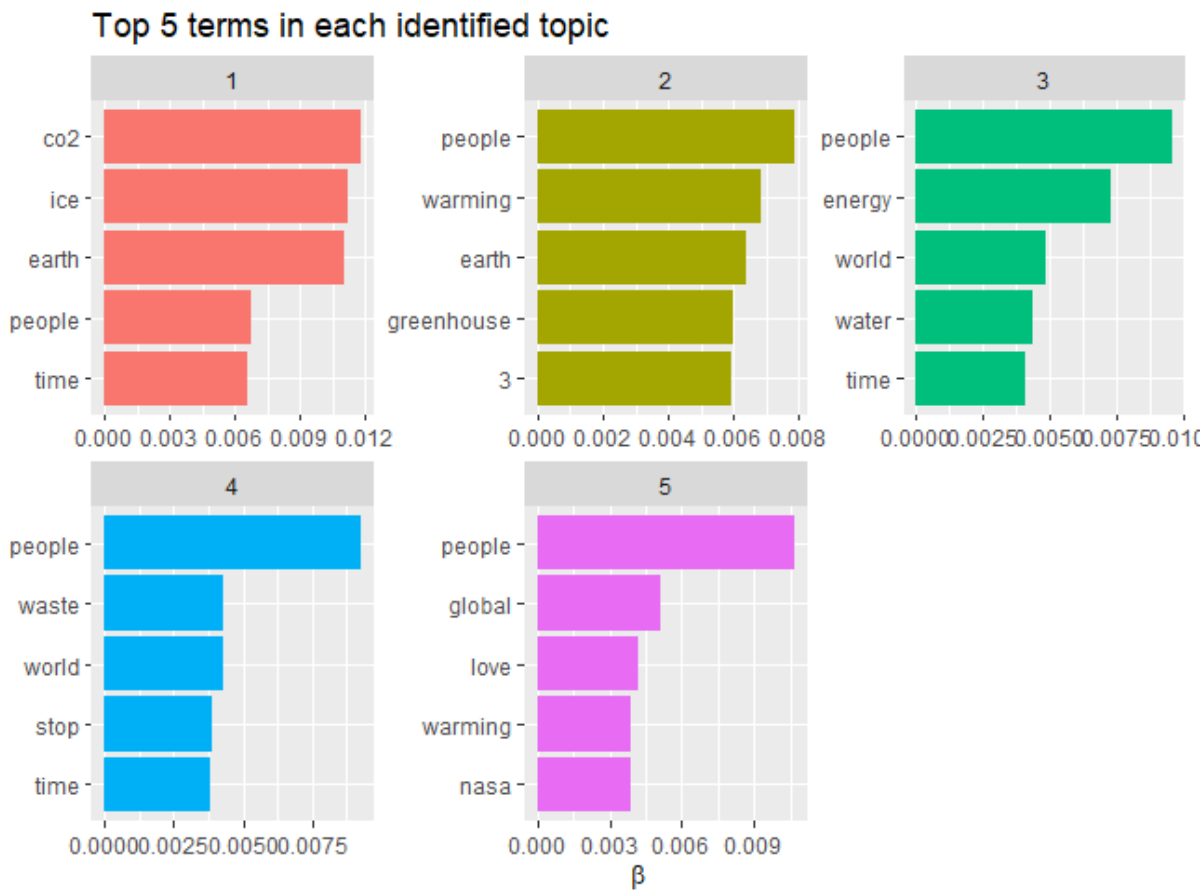
We also explored average rates of vio, mur, and rob over the years and found that the graphs are strikingly similar:

Selected Topic: 3 Previous Topic Next Topic Clear Topic





We further examined β values and looked at the probabilities of words belonging to the identified themes. We explored the top 5 words that were assigned to each topic and made further interpretations based on them. Here is what the Top 5 items identified in each of the 5 topics looks like: This trend is consistent with the data, as the two variables have a correlation coefficient of 0.53. A correlation coefficient of 0.53 indicates a moderate positive correlation, meaning that the two variables are moderately correlated, but not perfectly correlated.



To interpret the topic models, we looked at the terms with the highest beta values in each topic. These terms are most strongly associated with that topic. Then we consider the semantic meaning of these terms to infer the overarching theme or concept for that topic.

Topic 1:

Potential Theme: Earth and Environmental Changes

Reasoning: Terms like "co2," "ice," and "earth" suggest a focus on environmental factors and changes. The presence of "people" and "time" could imply discussions about the impact of these changes on human life over time.

Topic 2:

Potential Theme: Global Warming and Human Impact of Climate Change

Reasoning: Terms like "warming" and "greenhouse" strongly indicate a focus on global warming and the greenhouse effect. The recurrence of "people" might suggest discussions about the human impact on climate change.

Topic 3:

Potential Theme: Energy and Environmental Awareness

Reasoning: The presence of "energy" and "world" could suggest discussions about energy-related discussions of climate change on a global scale. "Water" might indicate awareness of water-related environmental issues like rising sea levels, floods, or droughts. The recurrent term "people" might indicate general discussions involving the effects on the larger public.

Topic 4:

Potential Theme: Environmental Actions and Waste Management

Reasoning: Terms like "stop" and "waste" suggest a focus on taking actions to address environmental issues and managing waste. The recurrence of "people" might indicate discussions about collective responsibility.

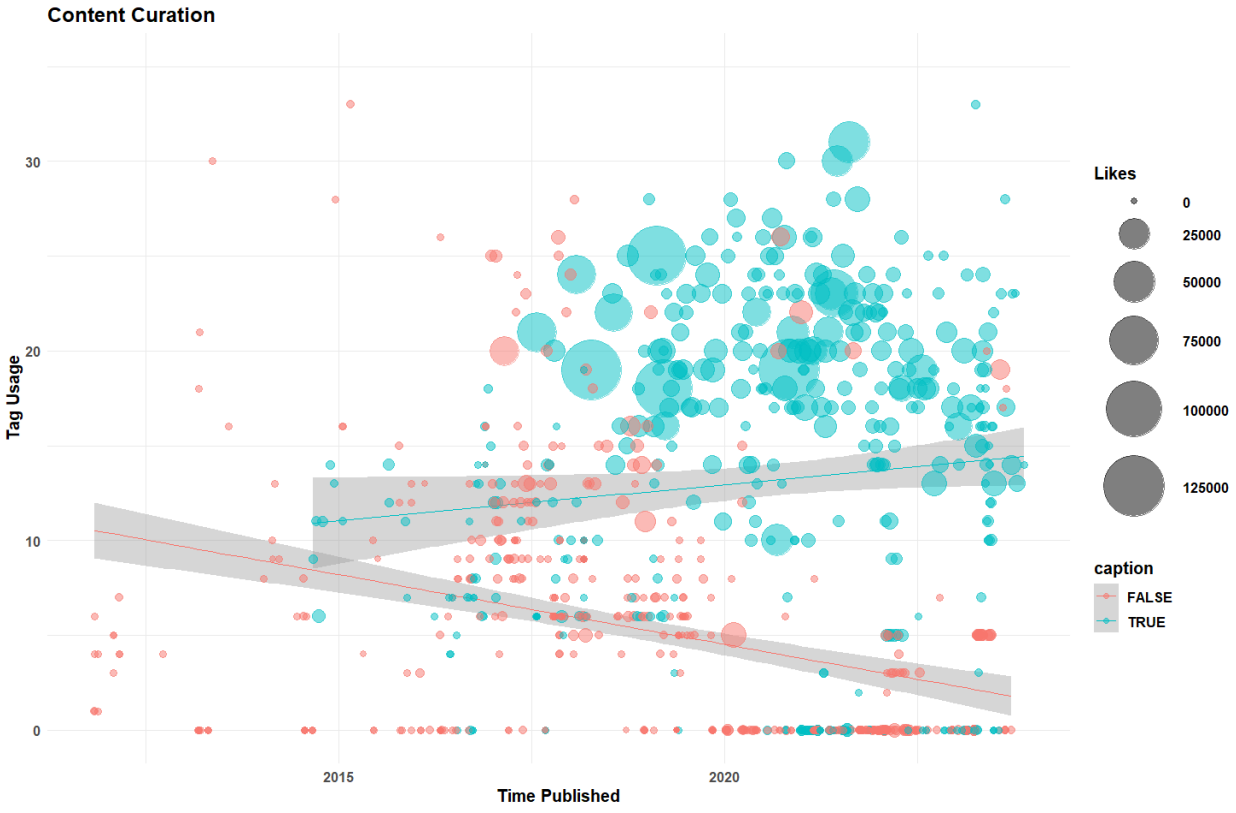
Topic 5:

Potential Theme: Global Issues and Positive Attitudes

Reasoning: Terms like "global" and "nasa" suggest a focus on global issues, possibly with a scientific or space-related perspective. The term "love" is an interesting word with an emotional connotation and might indicate positive attitudes or discussions about positive actions related to climate change or love for the planet.

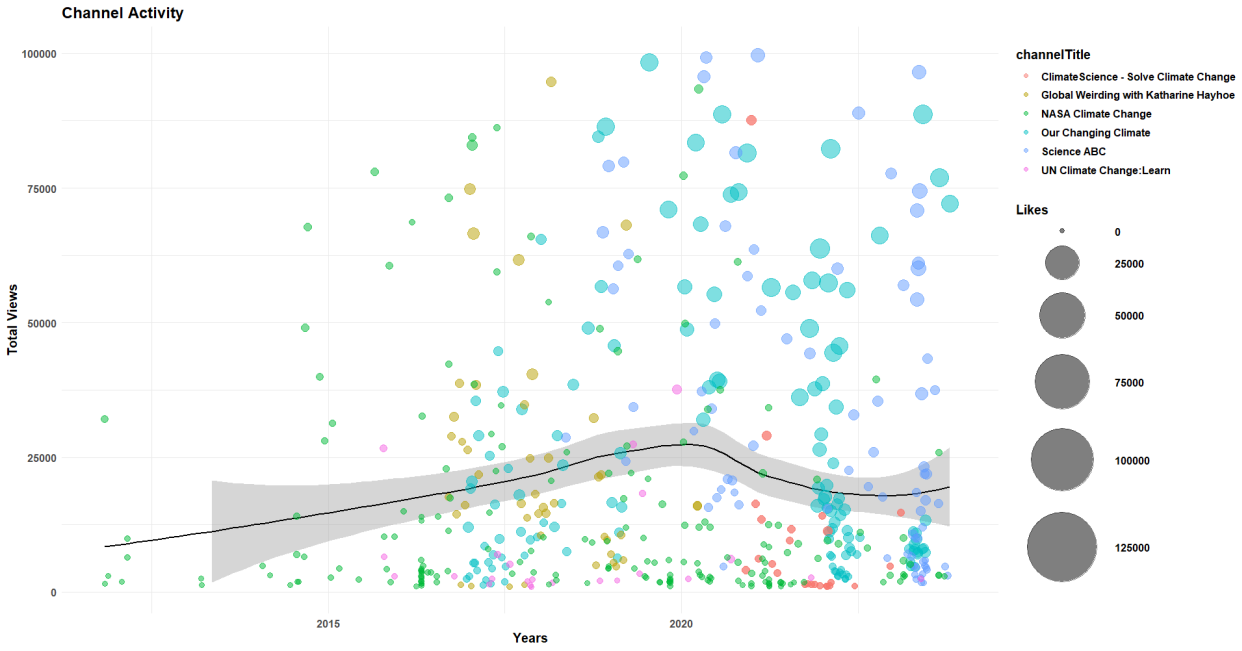
It is important to note that these interpretations are based on the terms extracted from the LDA model, and it is only possible to refine the themes further with a deeper understanding of the actual comments and the context in which they were made. Therefore, in future research on this topic, reviewing a sample of comments from each topic could provide additional insights into the discussions captured by our topic models.

Evaluation of Channel Success

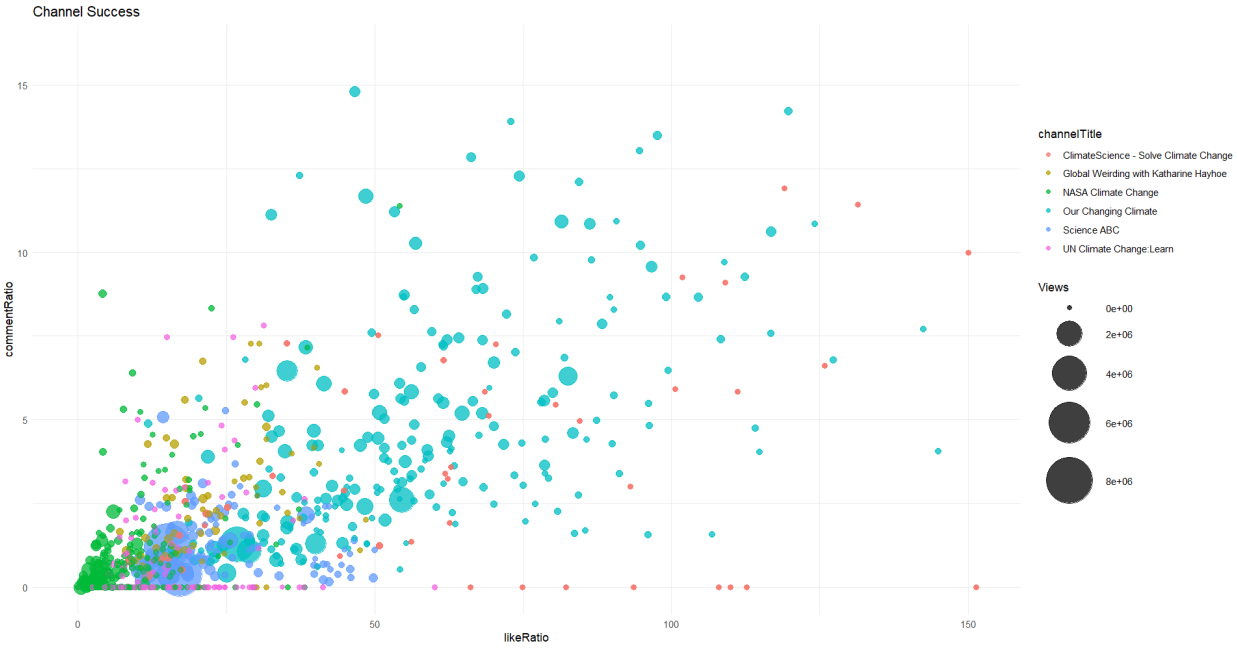


The following bubble charts evaluate the growth of climate change content on several channels. This chart spots trends in terms of how these accounts have optimized their accessibility and spotability, by comparing videos in terms of tag usage, captioning, and likes. Looking throughout the publishing of these climate change videos, overtime posted content is being captioned more and receiving better tags, with the most successful being the content on

that which combines both elements. These trends prove that overtime, creators are better understanding the platform as and continue to make more opportunities for viewers to find their content amongst similar topics. This is validated by how big the bubbles are gaining in terms of size over time, with 20 tags being around the average of best success in terms of tag amount on Youtube's platform. This chart confirms our hypothesis that climate change video content is spreading further awareness and is adapting their content to reach new levels of success.



This bubble chart analyzes the route in which channels have taken towards publishing their content, and how their views have been possibly attributed to timely trends. After applying a trend line to this chart, a peak in viewership can be observed during 2020. This and other vertical clustering can be seen on certain years where the data appears to stratify.



The chart contrasts each channel in terms of overall success by seeing which creator outperforms in terms of comments and likes. As shown by their clustering, *Our Changing Climate* is doing quite well in terms of engaging their audiences in a balanced manner, with even bigger channels such as *NASA Climate Change* and *UN CLimate Change* lagging behind. One curious point in the data surrounds the question of why viewership is low for videos with high commenting and likes. Typically social media with high engagement gets featured and promoted more than their competitors. The biggest takeaway from this chart is how it demonstrates which is the best channel in terms of modeling content and establishing success, aiding to pinpoint the potential flaws of other creators.

When overing all these bubble charts, it appears that accessibility is a factor in creating more likable content, there is better average success for videos that use a certain tag amount, and that audiences can be greater attracted in what could be potentially a seasonable trend. It can be concluded that these channels in particular are heading towards the right direction of reaching the masses.

Possible Future Additions

The proposed data strategy outlined in this plan addresses the need for efficiency and agility in handling data processing tasks. The preference for ELT (Extract, Load, Transform) over ETL (Extract, Transform, Load) pipelines acknowledges the importance of leveraging the power of modern cloud data warehouses. Furthermore, the envisioned interactive dashboard promises to provide a dynamic platform for real-time tracking of metrics related to ongoing discussions, offering valuable and actionable insights. Additionally, the account-based application introduced in the strategy aims to empower individuals and organizations by enabling them to securely store and load their specific data, facilitating the generation of insights through customizable scripts and visuals. This holistic approach not only streamlines data management but also fosters a collaborative environment, where data becomes a catalyst for informed decision-making.

Conclusion

In conclusion, our research unequivocally supports the hypothesis that visual media serves as a rich source of correlational data, enabling the identification of audience sentiment across diverse demographic segments. This comprehensive exploration has equipped us with valuable tools to measure the effectiveness of communication strategies within constrained environmental parameters. Beyond affirming our initial premise, our findings reveal substantial potential for expansion, particularly in fields such as social tipping analysis and large-scale Natural Language Processing (NLP) inferential models. The data extracted and analyzed during this study can be harnessed to track the evolution of target audience attention, establishing a correlated field that underscores the need for a generalized application. Envisioning a user-friendly platform for creating inferential sentiment models, we propose that such a tool would significantly benefit organizations and channels. This innovation has the potential to optimize communication

strategies in visual media formats objectively, ensuring that entities can tailor their approaches based on the dynamically shifting landscape of audience preferences and engagement.

References

Dahlman, Rebecca Lindsey AND LuAnn. “Climate Change: Global Temperature.” NOAA Climate.Gov, www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:~:text=Earth’s%20temperature%20has%20risen%20by,2%C2%B0%20F%20in%20total. Accessed 9 Oct. 2023.

Hersher, Rebecca. “Why Hurricanes Feel like They’re Getting More Frequent.” NPR, NPR, 27 Feb. 2023, www.npr.org/2023/02/27/1158969044/why-hurricanes-feel-like-theyre-getting-more-frequent#:~:text=Climate%20change%20makes%20hurricanes%20more%20dangerous%20Climate%20change%20is%20making,of%20storms%20isn’t%20changing.

Wittenberg, Chloe, et al. “The (Minimal) Persuasive Advantage of Political Video over Text - PNAS.” PNAS, www.pnas.org/doi/10.1073/pnas.2114388118. Accessed 9 Oct. 2023.