

DS650001 - DATA VISUALIZATION

Calibrating the Credibility of Scientific Information on social media

The word cloud illustrates the frequency and color-coded credibility of various COVID-19 related terms. The size of each word represents its frequency, and the color indicates its credibility score. A legend at the top shows a gradient from red (low credibility) to green (high credibility). The most prominent words include 'new coronavirus' (large, pink), 'support' (medium, purple), 'reopen' (medium, purple), 'hydroxychloroquine' (medium, teal), 'vaccine' (medium, purple), 'distancing' (medium, orange), 'testing' (medium, light blue), 'lockdown' (medium, orange), 'thanks' (medium, green), 'mask' (medium, orange), 'world' (medium, orange), 'covid' (medium, pink), and 'test' (medium, light blue).

black china tracing hand
people school death thank pandemic
new coronavirus health
support reopening hydroxychloroquine vaccine
distancing testing lockdown thanks mask world
covid test

PROJECT SUMMARY

Batch details	DS650001 FALL 2023
Team members	Tanmay Agarwal (UCID: ta424) Tejaswini Rao Akula (UCID: ta58) Harshad Ravindra Hatte (UCID: hh349) Balaji Kolusu (UCID: bk423) Shalu Saroj (UCID: ss4676)
Project Domain	Data Visualization
Proposed project title	Calibrating Credibility of Scientific Information on Social Media

Contribution Statement

It was a collective effort by each of the project members in our group where we discussed, ideated and considered the feedback from the professor and the mentors to deliver the project. Below are the individual contributions by each project member of our group:

1. **Tanmay Agarwal:** In the context of project deliverable 3, I collaborated in the formulation of visualization tasks for the dashboard, encompassing the incorporation of actions, filters, and animations to seamlessly integrate all visualization components. My focus was directed towards specific visualization tasks, notably Task 1: Analyzing sentiments expressed by both experts and non-experts, Task 2: Conducting subjective sentiment analysis across months, and Task 3: Examining sentiment score distribution within distinct topic categories. Additionally, I helped in co-designing the project deliverable-3 report, with a particular emphasis on articulating the dashboard's purpose and elucidating the adaptations made to our original visualization tasks in alignment with the overall dashboard design.

2. **Shalu Saroj:** Analyzed the data and crafted the problem statement; Figured out the target audience for the project. Worked on the tasks to carry and documented more than 10 potential visualizations along with their objectives. Provided visualization graphs for Tasks 2 and 3; Ideated and offered inputs for the rest of the tasks. Made a presentation report for Deliverable 3, and contributed to Dashboard. For deliverable 4, prepare the document.

3. **Tejaswini Rao Akula:** Individual contribution on the Data E/A/P (Exploration, Analysis and Preprocessing) to brainstorm ideas. Drafted Project presentation as part of Deliverable-3. Collaborated with the team to figure out the Visualization tasks that aligns with the goal and objectives of the project. Prepared Visualization Designs for Task 5 . Depicted the best alternative designs for Task 1 and Task 5 as part of the Project report, drafted the task definitions and descriptions, as part of Visual encoding for the same.

4. Balaji Kolus: For Deliverable 3, I developed a dashboard employing subjective sentiment analysis, depicting sentiment distribution across different topic categories, and illustrating the correlation between sentiment and expert-labeled tweets. I crafted visualizations for both Task 2 and Task 3, offering a comprehensive analysis of a dataset containing social media posts. The goal was to extract meaningful insights into public sentiment, subjectivity levels, dominant topic patterns, and underlying intents. Throughout the project, I assumed responsibility for executing diverse tasks and diligently documented the objectives of four potential visualizations.

5. Harshad Hatte:

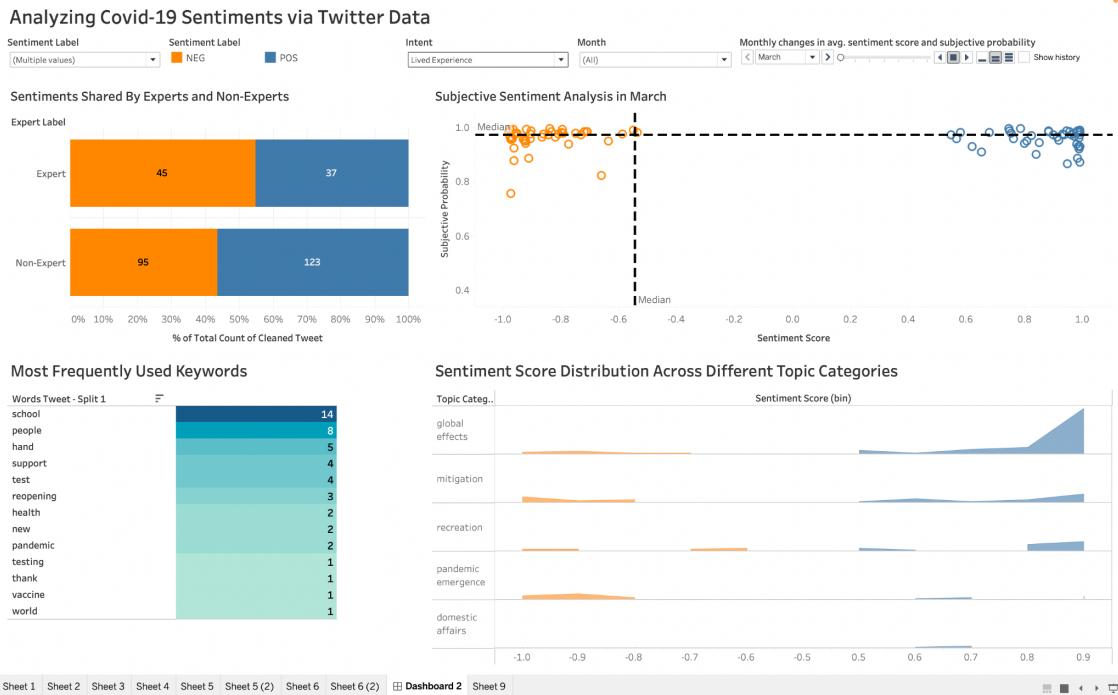
Conducted a comprehensive analysis of the dataset. Designed multiple visualizations to understand the dataset which helped me in formulating precise problem statements, Provided visualization graphs for Tasks 6 and 7. In Deliverable 3, my primary contribution involved preparing the second visualization on sentiment analysis over the months (4-9) in the dashboard. I actively participated in refining the dashboard's aesthetics to enhance its presentation. Additionally, I played a role in crafting the project report for both Deliverable 3 and Deliverable 4, ensuring a cohesive and comprehensive documentation of our work.

Explaining the dashboard interactions and how these interactions aid in drawing inferences.

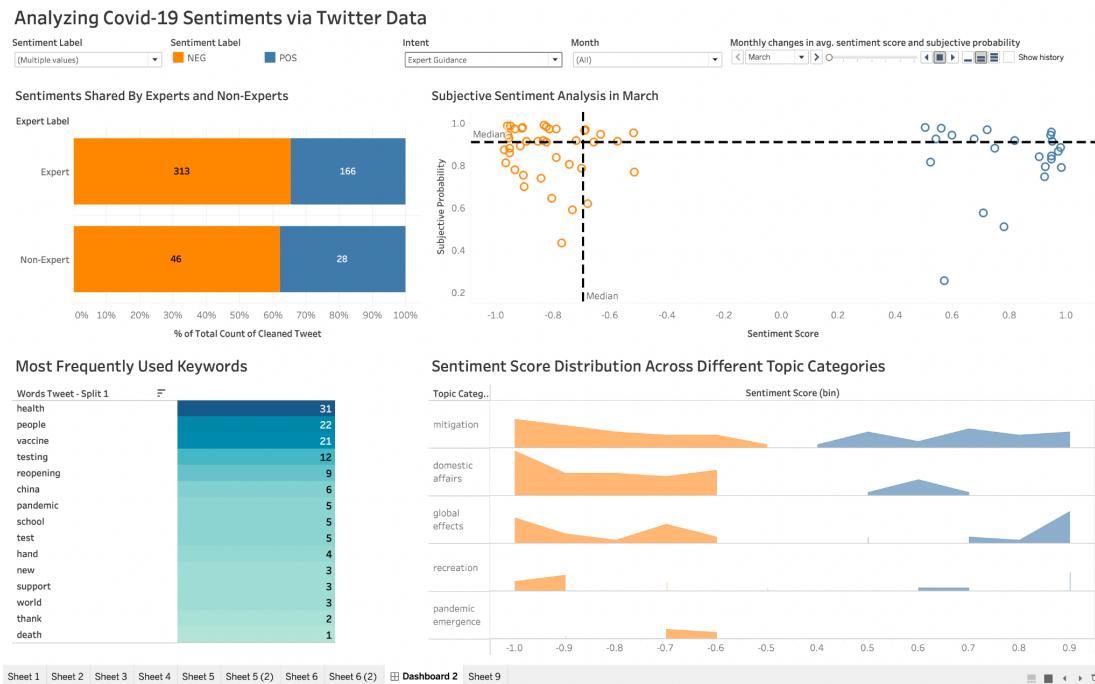
1. VIZ 1: Sentiment Distribution - Experts vs Non-Experts

Dr. Sarah, a seasoned analyst at the Public Health Sentiment Bureau, embarks on a journey through the realm of COVID sentiments on Twitter. Instead of sentiment scores, the visualization presents the relative proportions of Positive (POS) and Negative (NEG) tweets. It offers a clear, direct summary of positive and negative contributions from both experts and non-experts in the dataset, shaping an insightful snapshot of the COVID-19 public discourse.

In this interactive journey, she wields the 100% stacked bar chart to unravel the diverse sentiments within 'Expert Guidance' and 'Lived Experience,' two intent-filtered realms that define the public discourse on COVID-19.



Dr. Sarah, driven by empathy, first delves into the realm of 'Lived Experience.' As she analyzes the sentiment distribution, the vivid landscape unfolds. Non-experts, constituting the majority, contribute 56.42% positive sentiments and 43.58% negative sentiments. The chart vividly displays the narratives of 218 individuals, each tweet a poignant reflection of personal struggles and triumphs in the face of the pandemic.

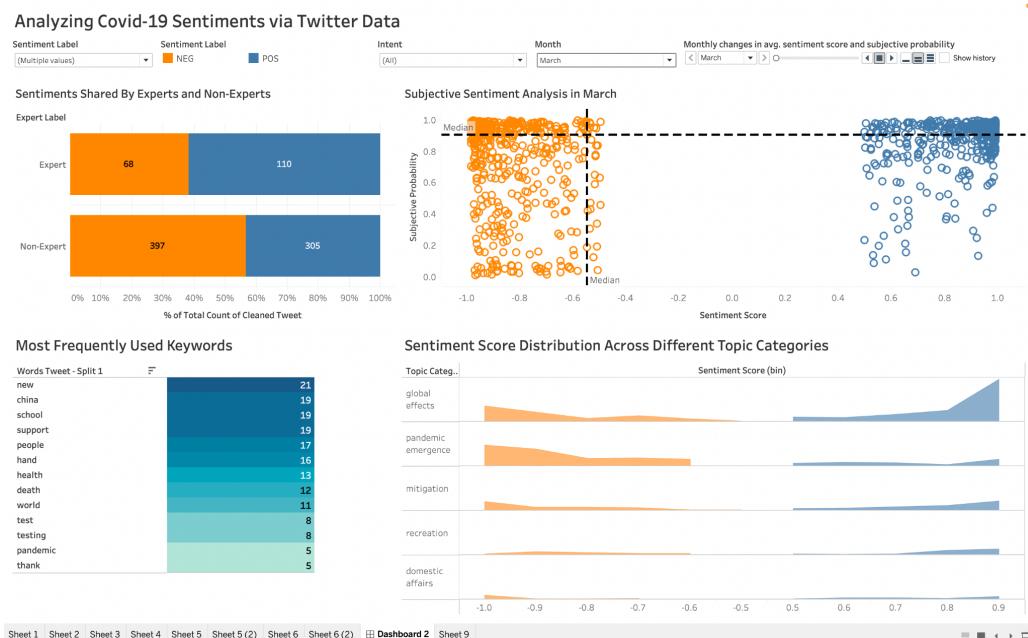


Switching gears, Dr. Sarah now sets her compass to 'Expert Guidance.' In this domain, experts take the stage, offering insights and counsel. The sentiment distribution, though nuanced, leans towards 45.12% positive sentiments and 54.88% negative sentiments from the 92 expert-driven tweets. Dr. Sarah interprets the authoritative tones, recognizing the pragmatic and objective nature characterizing these expert contributions.

In Public Healthcare, Dr. Sarah's interactive exploration becomes a valuable tool for officials. The juxtaposition of 'Expert Guidance' and 'Lived Experience' offers nuanced insights. It unveils the emotional pulse of the public—pragmatic and authoritative at times, deeply personal at others. This compass aids officials in tailoring responses that resonate with the dual nature of sentiments, aligning public health strategies with the diverse emotional landscapes unveiled by Dr. Sarah's journey.

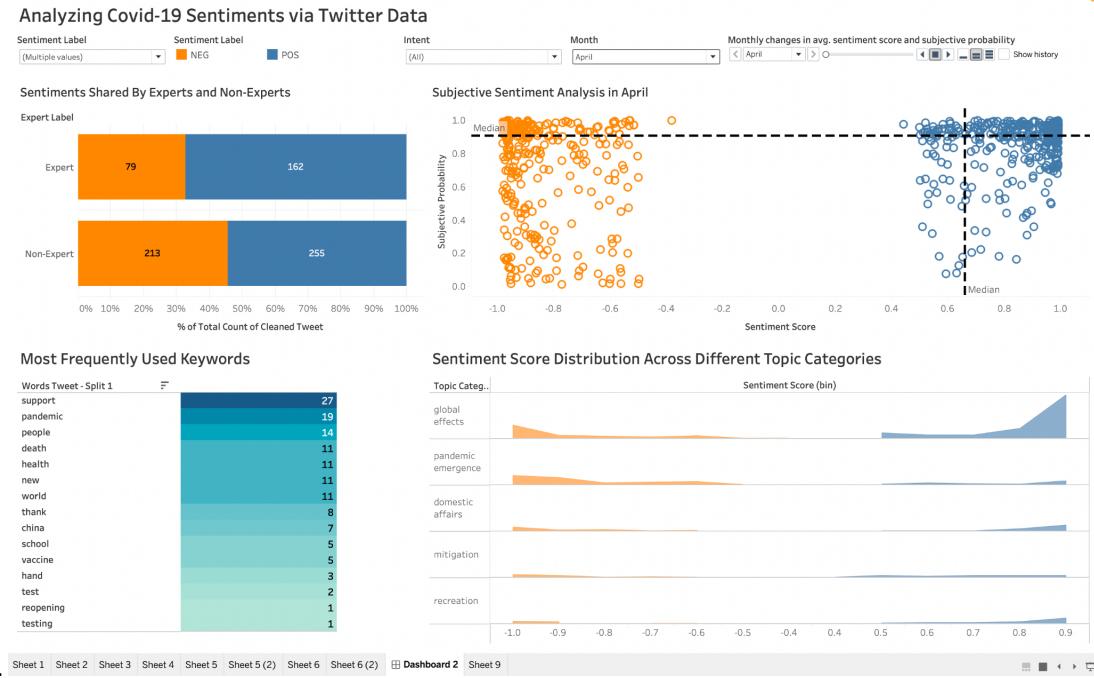
2. Viz 2 : Subjective Sentiment Analysis Across Months

As Dr. Sarah delves deeper into the spatial dynamics of sentiment analysis, she seamlessly transitions to the next chapter of her exploration—the Subjective Sentiment Analysis Across Months. With Intent as her guiding compass, she delves into the changing emotional landscape, enriched by reference lines and a temporal animation.



March unfolds with a hint of melancholy. Negative sentiments dominate, especially in non-expert contributions. Analyzing the "Subjectivity" metric reveals a peak during March, suggesting that tweets were predominantly subjective during this period. Dr. Sarah notes the expert touch,

injecting positivity amidst the negativity. Keywords like "china" and "new" cast shadows on this chapter



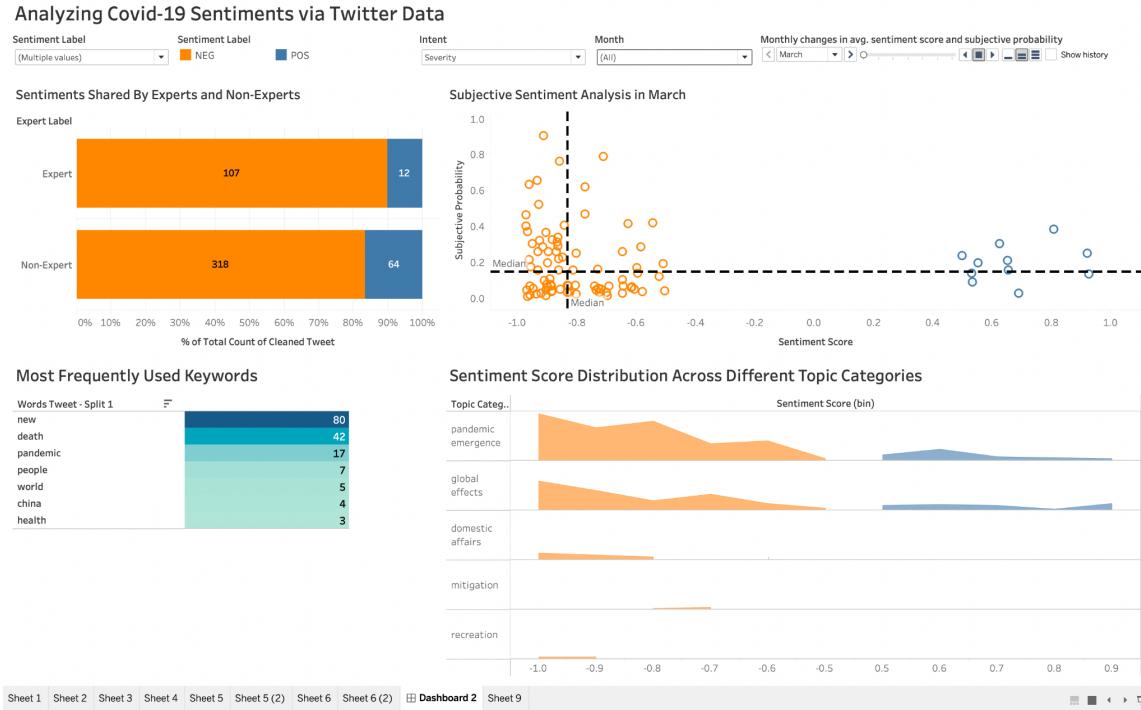
The animation shifts, painting April with brighter hues. Positive sentiments surge, mirroring societal support. Dr. Sarah identifies keywords like "support" and "pandemic" leading the optimistic chorus. The sentiment seesaw tilts favorably.

Variances in sentiment and subjectivity between categories, with a positive sentiment in "Global Effects" and a comparatively lower positive sentiment in "Pandemic Emergency," aligning with associated keywords like "death" and "new."

As the animation concludes, Dr. Sarah synthesizes these sentiment snapshots. This dynamic portrayal complements the initial visualization, offering a nuanced narrative. Dr. Sarah, armed with insights, stands ready to distill emotions into actionable guidance for those navigating the ever-shifting seas of public sentiment.

3. Viz 3: Most frequently used Keywords

Dr. Sarah turns her attention to the heatmap displaying the most frequently used keywords, offering a concise snapshot of global conversations. With Intent as the filter, she observes the dominance of non-experts in negative sentiment discussions, notably in the 'pandemic emergence' category.



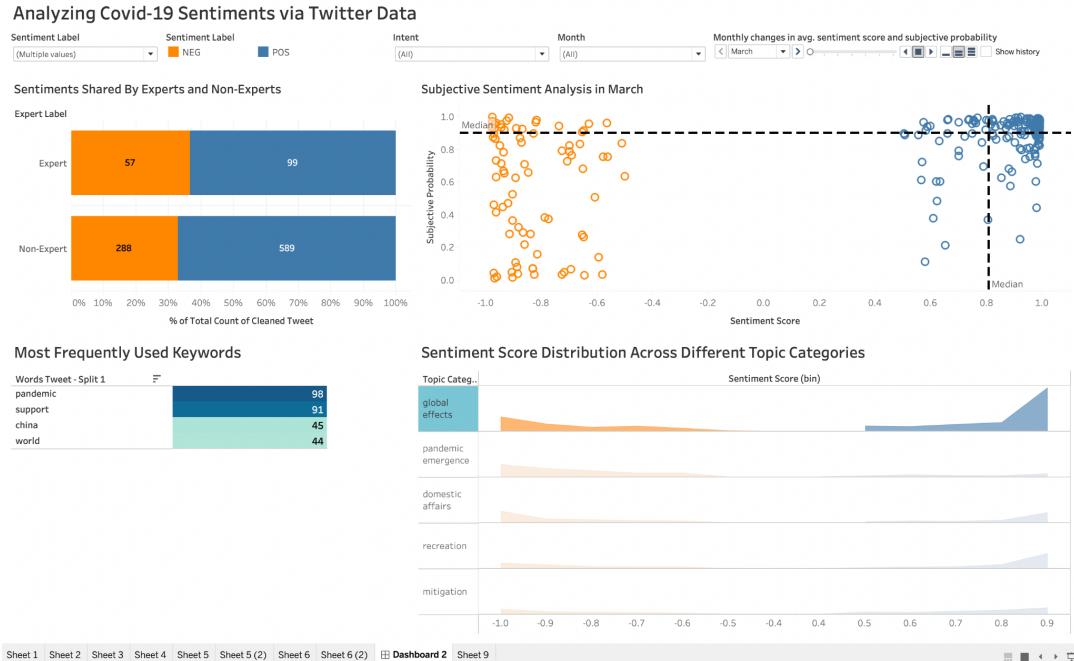
The heatmap, color-coded for intensity, aligns with overall sentiment analysis. The keywords 'new,' 'death,' and 'pandemic' resurface, with intent of severity, connecting objective tone with subjective concerns.

In synthesis, Dr. Sarah recognizes the heatmap's power to distill global discussions. Keywords, orchestrated by intent and sentiment, present a comprehensive tapestry of human concerns. This formal exploration equips her with insights to navigate the complex seas of public discourse, ensuring a nuanced understanding of collective sentiments.

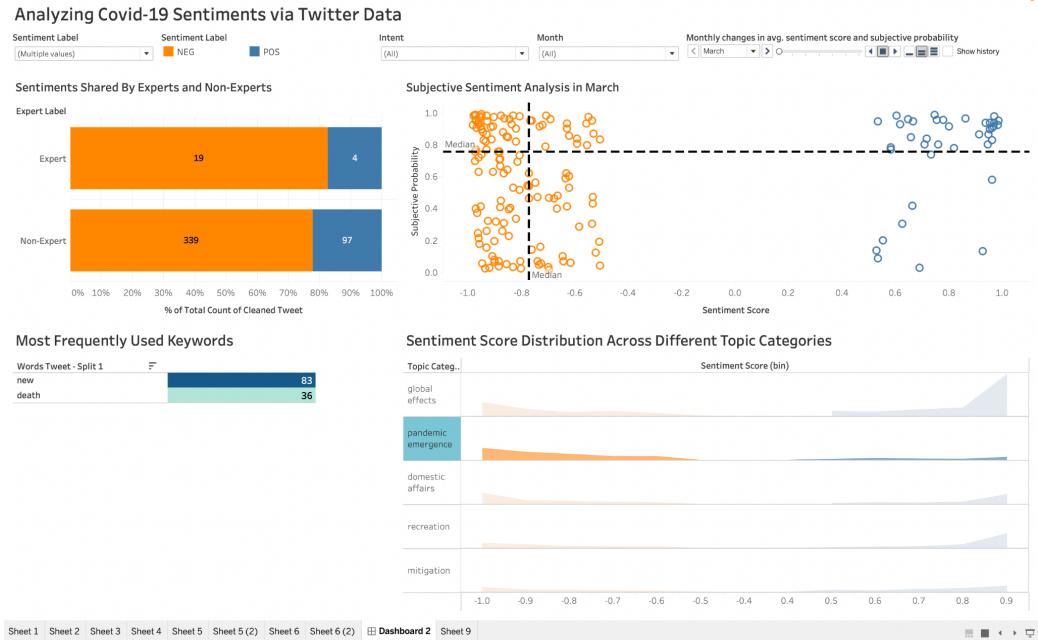
Dr. Sarah concludes, acknowledging the significance of this visualization as a compass, guiding her through intricate layers of global sentiment in an interconnected digital world.

4. Viz 4: sentiment vs topic categories

Armed with insights from this visualization, Dr. Sarah sees beyond the surface. The AUC plot, coupled with the refined focus on top categories, equips her to unravel the complex interplay between sentiment and topic categories. This exploration, she realizes, is a step closer to deciphering the intricate language of collective sentiments in the digital landscape.



In her exploration, she starts by selecting the "Global Effects" category to understand the overall sentiment patterns related to the global impact of the pandemic. The sentiment score distribution reveals predominantly positive sentiments within the "Global Effects" category. This encourages her, as she observes keywords like "support," indicating widespread global cooperation in addressing health challenges. Further, she filters keywords and hashtags to explore the specific factors contributing to positive sentiments. They find keywords like "pandemic" and "China," suggesting positive sentiments associated with international collaboration and support.



In contrast, she switched to the "Pandemic Emergence" category to investigate any negative sentiments related to the initial emergence of the pandemic. The sentiment distribution in this category reveals an overall negative trend. Keywords like "new" and "death" dominate the negative sentiment, indicating concerns and challenges associated with the early stages of the pandemic.

In the end, she infers that the global community's response to health issues in the "Global Effects" category is positive, fostering collaboration and support. The negative sentiments in the "Pandemic Emergence" category lead to the inference that challenges and concerns were more prevalent during the initial stages of the pandemic.