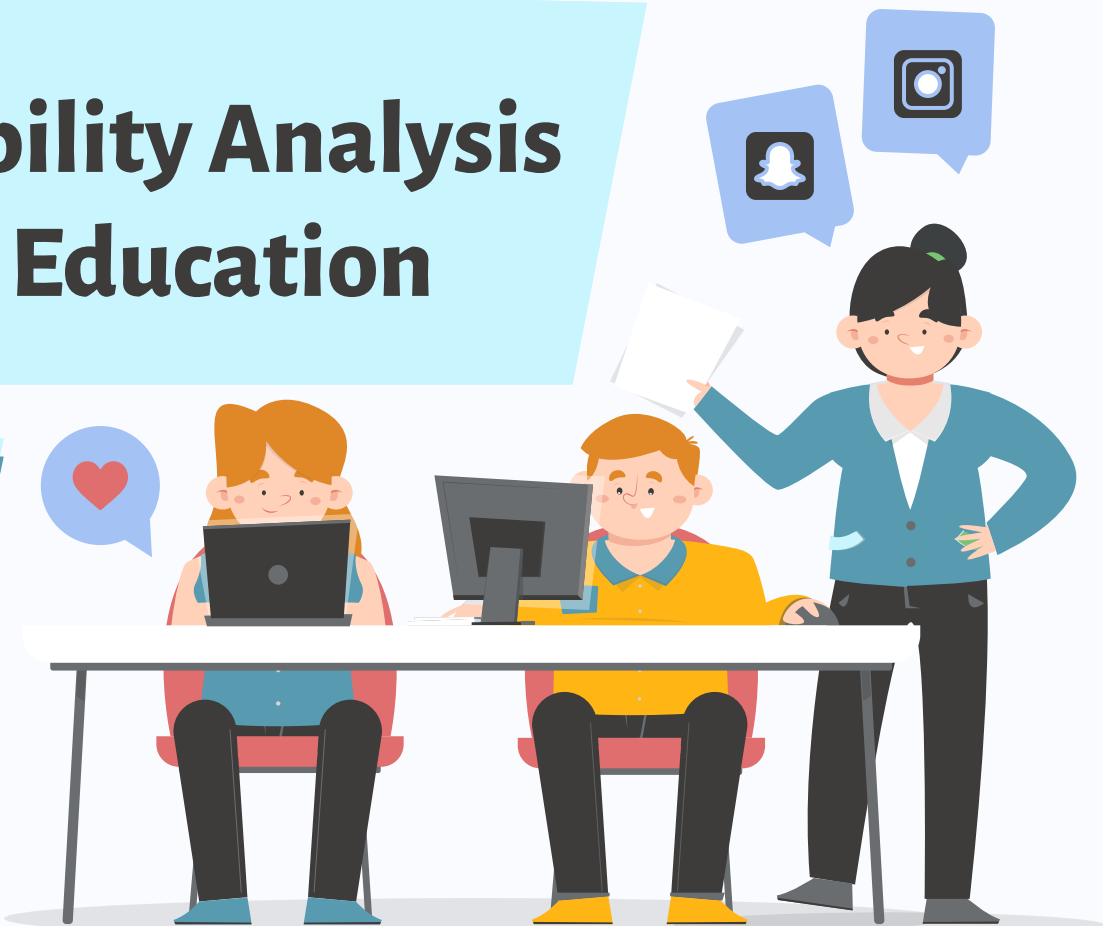


Twitter Credibility Analysis On Special Education



Big Data Platform Final Project
Maggie Chuang
March, 2023

Can Twitter be considered as a
credible source of information on
special education relevant topics ?



EXECUTIVE SUMMARY

CONCLUSION

1. 98.2 % of twitterers are not verified. Therefore, Twitter can't be a reliable source on special education.
2. Using solely the amount of tweets might not a good indicator of whether a twitterer is influential. Using engagement rate, influential score that considers twitterers followers, replies counts might be better.
3. North America has the highest tweeters count. However, ~ 98% of location data including location, country name, and country code are lost. Therefore, analysis regarding geological distribution might be inaccurate.
4. There are several peaks in July 2022, Oct 2022, and Feb 2023 yet they aren't all related to hot topics in special education.
5. Lots of twitter accounts are created in US In 2009, 2011 and 2019 while there is a peak in 2023 in Saudi Arabia. The reason might be due to the fact that special education has developed faster and earlier in US.
6. Most organizations exhibit a high percentage of unique tweets, indicating that they generate their original content rather than copying and pasting others' content. NGOs, in particular, are highly active in tweeting about special education topics, with a 98% unique tweet count.

RECOMMENDATION

1. Implement a verification process for users tweeting about special education-related topics.
2. Developed several metrics to evaluate influential scores considering tweet counts, followers, reply counts, tweet created time on the database so there is a credible way for all to measure the influence. Thus, users won't be biased by only the tweet counts and won't be confused by various measurements on the internet.
3. Developed strategies to encourage users to fill in location relevant information so that geological analysis can be conducted more accurately.
4. Twitters should develop algorithms that could automatically classify user's tweets by users profile and tweets using hashtags. This would help users to filter out relevant tweets in a more efficient and accurate way.

METHODOLOGY

Filter

Use keywords to filter out education relevant topics with a focus on special education.

EDA

Explore each variable to find insights.

Variable selection.

Change data type.

Influencer Analysis

Categorize twitterers into different categories.

Find out which category and who have the highest original tweet counts and retweet counts.

Location Analysis

Analyze the geological distribution of twitterers from global to local scale.

Time Analysis

Analyze time trends of daily tweet counts' from general to specific twitterers' categories.

Message Uniqueness

Identify whether the message is duplicate or unique for twitterers of each category using LSH text similarity.

OVERVIEW

DATA SOURCE

BEFORE

Filter

AFTER

Volume

~ 100 million

(99,994,342)

~ 33 million

(33,490,185)

Variable

39 variables

15 variables

Topics

General Education

Special education

Time Range: 2022-04 ~ 2023-02

BACKGROUND

Objective: Identify whether Twitter is a credible source of information with a focus on special education topics.

Motivation: Special education helps individuals with disabilities achieve their full potential and ensures they have equal access to education. Even though Twitter might not include sufficient information on special education currently, it can still be useful for sharing information and raising awareness. By identifying whether Twitter is a credible source of information on this topic could help us better understand current trend and provide recommendations for Twitter to optimize its platform and increase its credibility.

FILTERING PROCESS ⇒ KEYWORD SELECTION

Assumptions: The tweets were collected on the topics of general education, schools, universities, learning, and knowledge sharing. Thus, I didn't include general words such as education, school, universities, learning. Also, I chose to put more weights on special education. Therefore, I contained more keywords for this topic. I classified education topics into below categories(demonstrate some of them only):

- **K-12 relevant:** 'preschool','kindergarten', 'primary education', 'secondary education', 'higher education', 'k12', 'K-12', 'k12online','teacher', 'parenting', 'tuition', 'college', 'highschool', 'middle school', 'professor', 'academics', 'academy',,, 'public school', 'private school', 'ACT', 'SAT', 'scholarship', 'tuition',
- **Special Education:** 'ece', 'special needs','dyslexia','tck', 'autism', 'bilingual', 'aspergers', 'special education', 'special-needs', 'aided', 'exceptional', 'alternative provision', 'special ed', 'SDC', 'SPED', 'individual differences', 'disabilities' self-sufficiency', 'learning disabilities', 'communication disorders', 'emotional disorders', 'behavioral disorders', ' physical disabilities', 'therapy', 'osteogenesis imperfecta', 'cerebral palsy', 'lissencephaly', 'muscular dystrophy', 'developmental disabilities', 'autism spectrum disorder', 'intellectual disabilities'

VARIABLE SELECTION

How variables are related to credibility analysis

Basic Tweet Information

- tweet_id → aggregate use
- tweet_likes → influential analysis
- tweet_created: time analysis
- tweet_reply_count: influential analysis
- language → select English
- tweet_text → text analysis

Place Information

- location:
country name + state name
- country_name
- country_code

Retweet(rt) Information

- rt_status.created_at → when the retweet is created.
- rt_status.favorite_count → how many ppl like this retweet
- rt_status.retweet_count → number of times this Tweet has been retweeted
- rt_status.reply_count → how many replies are in this retweet
- rt_status.entities →
hashtags.text: We can see the hashtags of this text.

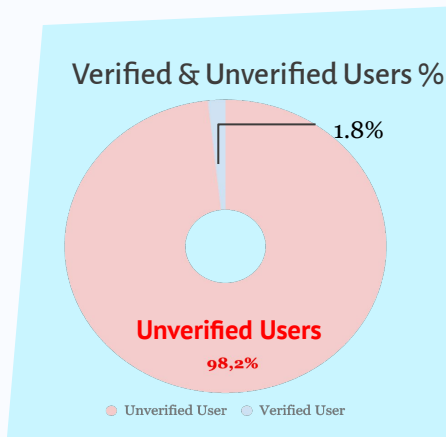
User Information

- user_id → unique identifier for each user, great for aggregation use (ex: user_count)
- user_name → after aggregation, can use to identify who the twitter is (would be useful if he's someone famous)
- user_description → can be used to classify twitter's profile / identify whether he's an advocate/supporter related to special education
- followers_count → can be used to measure whether this user is popular/ influential
- favourites_count → The number of Tweets this user has liked in the account's lifetime. Can be used to measure this user's engagement to Twitter.
- verified → know whether this user's account is verified or not.
- account_created → the time when this account is created
→ Can be used to observe the trend users create twitter account.

coordinates and geo have more data loss than place. Therefore, I used place to do geographical analysis.

Text V.S. tweet_text V.S. extended_tweet.full_text → tweet_text appears more frequently than extended_tweet.full_text and contains the full text of text without urls → chose tweet_text to conduct text analysis.

EXPLORATORY DATA ANALYSIS



98.2% of Twitter users of special education topics are not verified.

Original Tweets %	36
Retweets %	64

Retweets % on special education is 2 times more than original tweets

Categorize Twitterers into different groups

Categorize Twitterers into different categories by searching keywords in user_name & user_description column

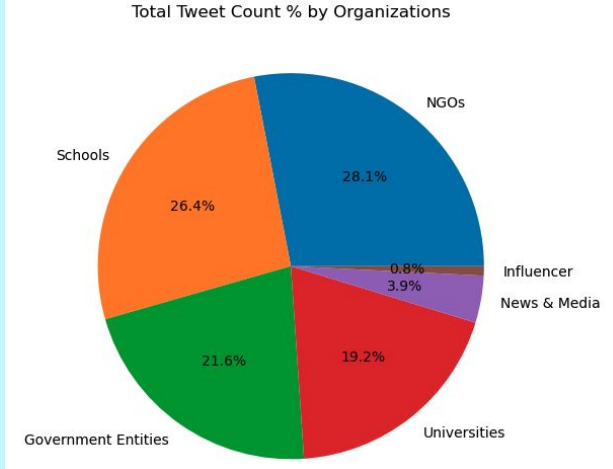
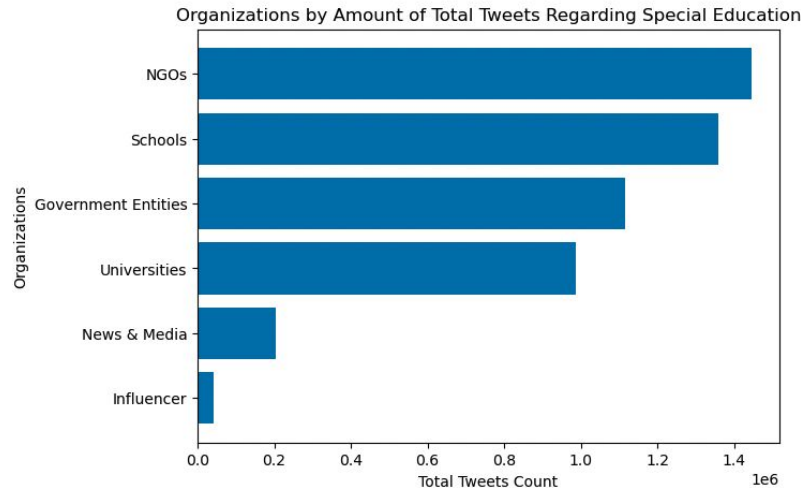
- ◆ Categories:
 - Government Entities
 - Universities
 - Schools
 - Nonprofit organizations(NGOs)
 - News & Medias
 - Social Media Influencers
 - Others

Null Values Summary

- Over 50% of place relevant information is lost. Therefore, analysis regarding geological distribution might be inaccurate.
- 17% of user_description information is lost. Therefore, there might be some inaccuracy to use keywords to map with user_description

EXPLORATORY DATA ANALYSIS

Which organizations do most special education relevant tweets belong to (by total tweets counts)



	twitterer_category	count	%
1	NGOs	1444876	28.058460
2	Schools	1358307	26.377352
3	Government Entities	1114838	21.649358
4	Universities	987698	19.180390
5	News & Media	203318	3.948290
6	Influencer	40483	0.786151

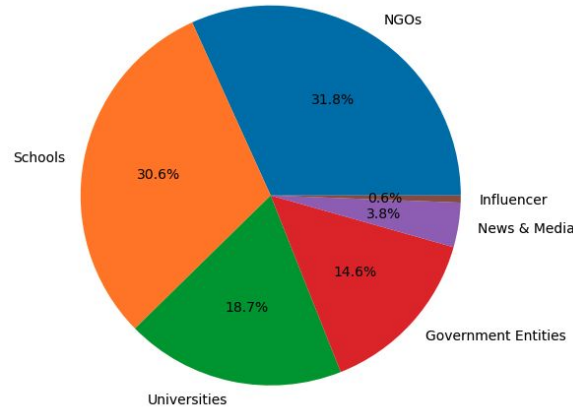
NGOs have the highest percentage (28.1%) of special education relevant tweets, followed by schools. This is not surprising considering the vital role NGOs play in promoting and improving special education, as recognized by research on the "Role of NGO's in Improving Education Facilities of Specially Abled." NGOs are likely to share informative messages and promote their efforts in enhancing the quality of life for people with disabilities, which explains their high percentage of tweets related to special education.

Research link: <https://psychosocial.com/article/PR201164/11863/#:~:text=It%20is%20a%20recognized%20fact,institutional%2C%20charity%2Dbased%20approach.>

EXPLORATORY DATA ANALYSIS

Which organizations have the highest original tweets/ retweets on special education topics?

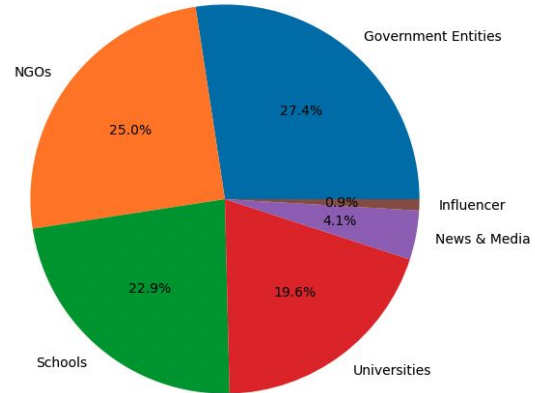
Original Tweet Count % by Organizations



Original Tweet Volume %

NGOs have the highest percentage of original tweets (31.8)% followed by Schools (30.6%). This is reasonable since NGOs likely produce unique content to raise awareness about special education topics, which will be discussed in later sections. (Message Uniqueness Analysis)

Retweet Count % by Organizations



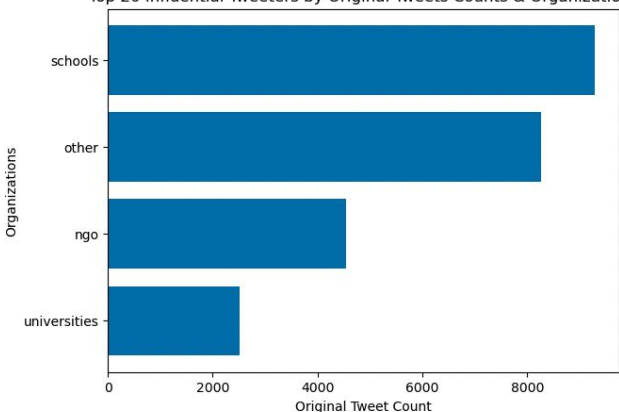
Retweet Volume %

Government Entities have the highest percentage of retweets (27.4)%. The reason might be because they collaborate more frequently with NGOs and schools on special education initiatives, which could lead to higher retweet counts for government entities as they share information about these collaborative efforts.

AUTHOR IDENTIFICATION

Who are the most influential Twitterers regarding special education (By Original Tweet Counts)

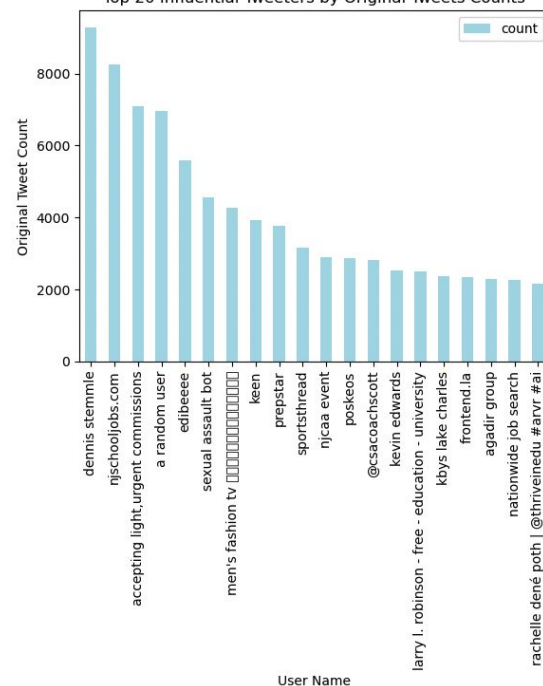
Top 20 Influential Twitterers by Original Tweets Counts & Organizations



- 12 out of 20 top 20 influential twitterers are non-verified and belongs to other category. They are most likely to be random users.
- Therefore, using original tweet counts to evaluate whether a twitterer is influential is not credible. Please refer to Appendix for other possible methods to measure a twitterer's influence.

user_name	twitterer_category	count
dennis stemmle	schools	9284
njschooljobs.com	other	8264
accepting light,urgent commissions	other	7091
a random user	other	6953
edibeeee	other	5597
sexual assault bot	ngo	4551
men's fashion tv ~メンズファッションテレビ~	other	4268
keen	other	3926
prepstar	other	3762
sportstread	schools	3157
njcaa event	schools	2909
poskeos	other	2870
@csacoachscott	other	2808
kevin edwards	universities	2515
larry l. robinson - free - education - university	schools	2502
kbys lake charles	other	2361
frontend.la	other	2350
agadir group	ngo	2302
nationwide job search	other	2253
rachelle dené poth @thriveinedu #arvr #ai	universities	2164

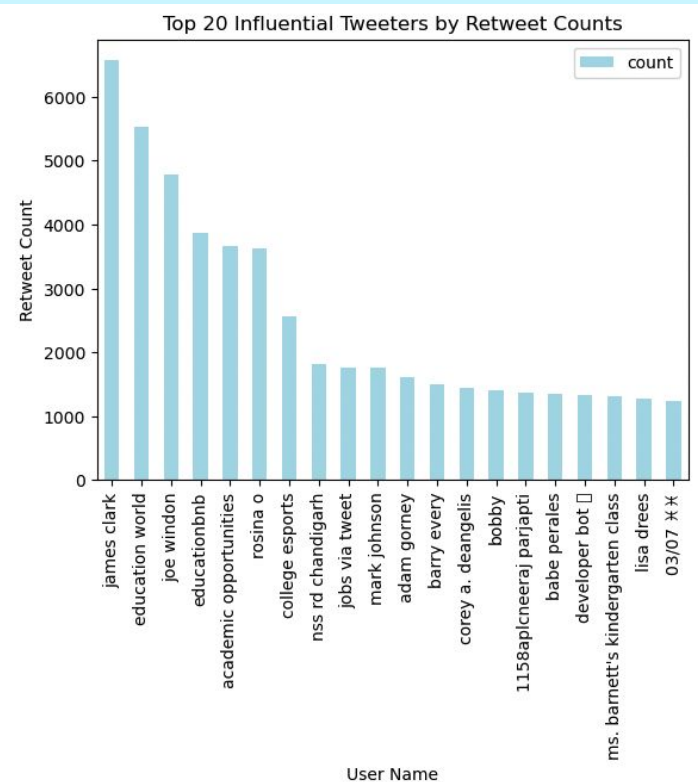
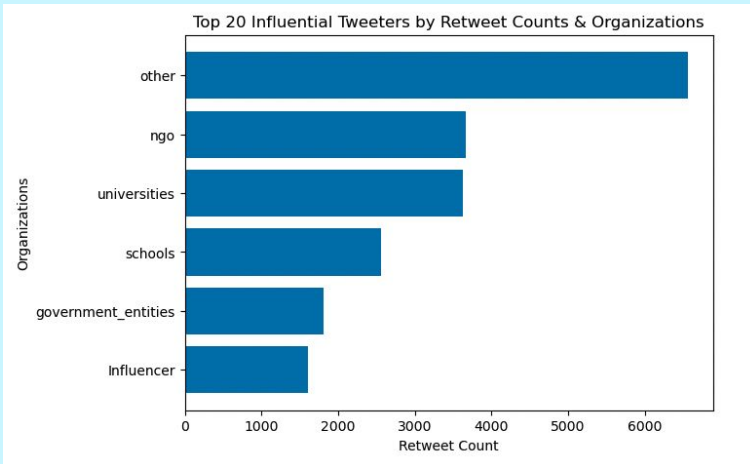
Top 20 Influential Twitterers by Original Tweets Counts



AUTHOR IDENTIFICATION

Who are the most influential Twitterers regarding special education (By Retweet Counts)

twitterer_category	
other	8
ngo	1
schools	4
universities	4
Influencer	2
government_entities	1

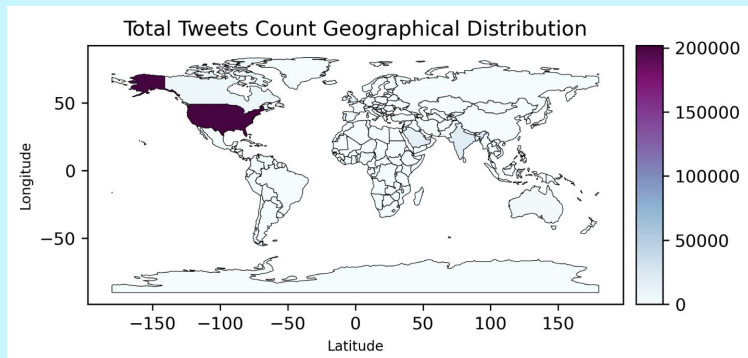


- Top 1 retweeters is James clark, a well-known American entrepreneur and computer scientist while Top 2 retweeters is education world is an online resource and publication who often shares articles and resources related to special needs education and retweets content from other accounts focused on special education and special needs topics.
- Nevertheless, there are still several random users mixed in between, indicating that using retweet counts to measure a twitterer's influence is not credible. We could use measurements such as engagement rate or influence score to evaluate twitterers influence. Please refer to Appendix for further information.

LOCATION ANALYSIS

Overview of the geographical distribution of special education relevant tweets

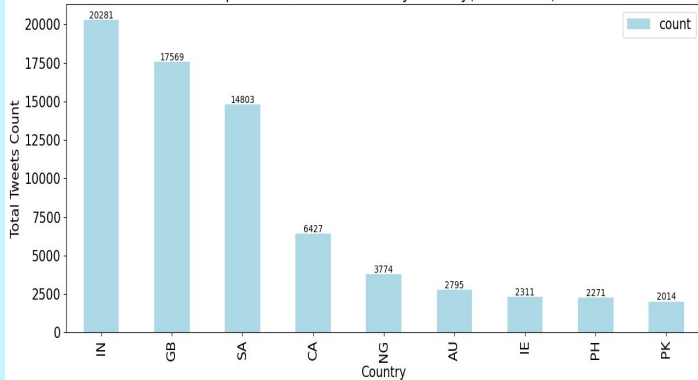
	country_code	count
0	US	201751
1	IN	20281
2	GB	17569
3	SA	14803
4	CA	6427
5	NG	3774
6	AU	2795
7	IE	2311
8	PH	2271
9	PK	2014



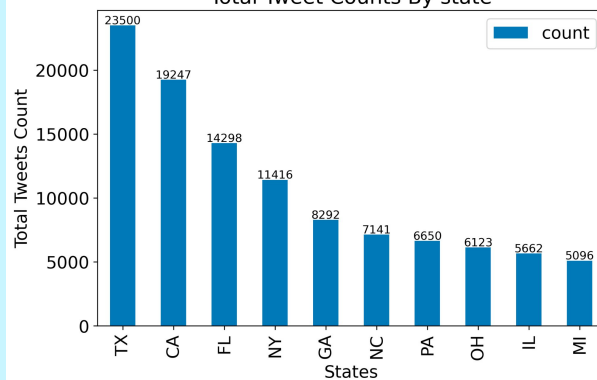
	continent	count
3	North America	227932
2	Europe	20583
1	Asia	16308
0	Africa	9719
4	Oceania	5962

- **North America** has the highest tweet counts.
- Top 3 countries having the highest tweet counts are US, India, UK.
- Among them, **US** has the highest tweets counts, which aligns with the country with most Twitterers.

Top 10 Total Tweet Counts By Country(Exclude US)



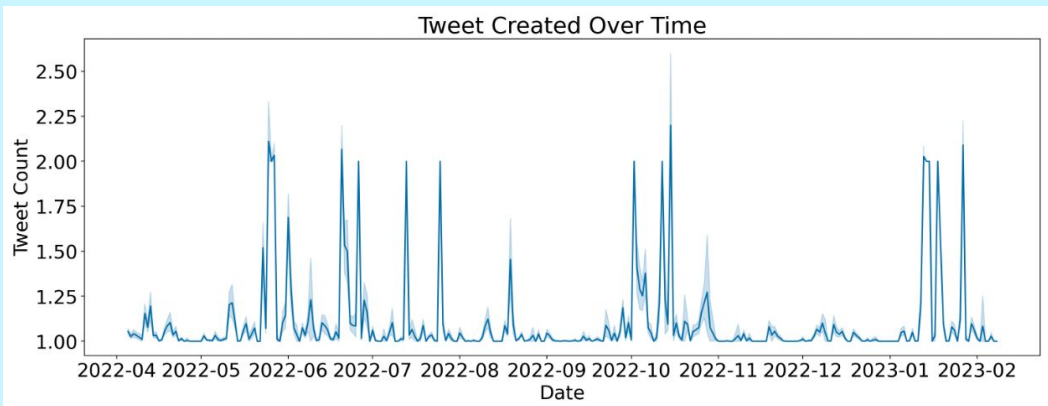
Total Tweet Counts By state



TX, CA, FL, NY, GA have the highest tweet counts regarding special education issues. These states have larger populations, which could result in a higher number of individuals with disabilities and a higher demand for special education services. Additionally, these states may have a higher number of special education organizations, schools, and advocacy groups, which could result in more tweets being shared about special education topics.

TIMELINE ANALYSIS

Relationship between emergence of special education issues overtime



Although there were several peaks in Twitter activity related to special education in July 2022, October 2022, and February 2023, not all were necessarily related to hot topics in the field. The peak in July 2022 may be related to the start of school year. On the other hand, the peak in October 2022, during which government entities had the highest tweet count, was likely due to the US election.

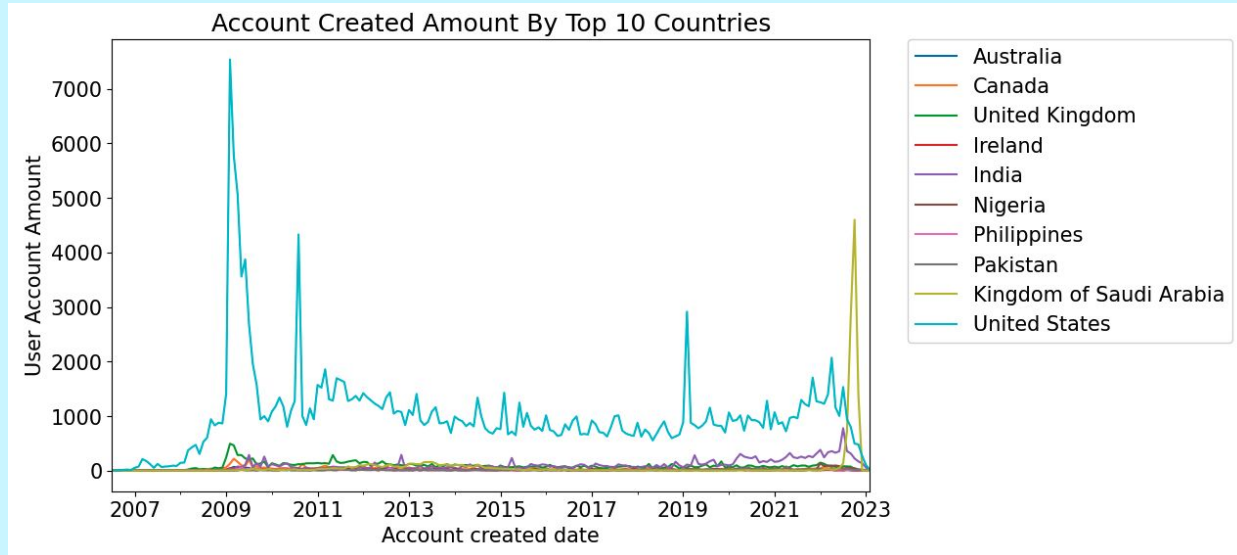
July 2022: This peak may have been due to the observance of Disability Pride Month, which takes place in July. It is a month-long observance that celebrates and recognizes the contributions of people with disabilities. It may also relate to the start of school year, a time when many educational issues are being discussed.

October 2022: This peak may have been due to the US midterm elections, which were held in November 2022. Politicians and political parties may have been using social media, including Twitter, to reach out to their constituents and promote their platforms, including those related to special education.

February 2023: This peak may have been due to the annual observance of International Day of Persons with Disabilities, which takes place on December 3rd. The lead up to this observance may have generated increased awareness and discussion on social media, including Twitter, about issues related to special education and disability rights.

TIMELINE & COUNTRY ANALYSIS

Relationship between emergence of special education issues overtime by countries

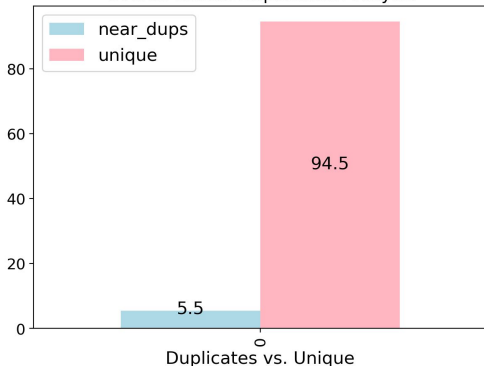


Twitter account creation peaked in the US in 2009, 2011, and 2019, while a peak was observed in Saudi Arabia in 2023. This may be due to the fact that special education has been developing earlier and more quickly in the US than in Saudi Arabia, and it has taken longer for the topic to emerge on Twitter in the latter country.

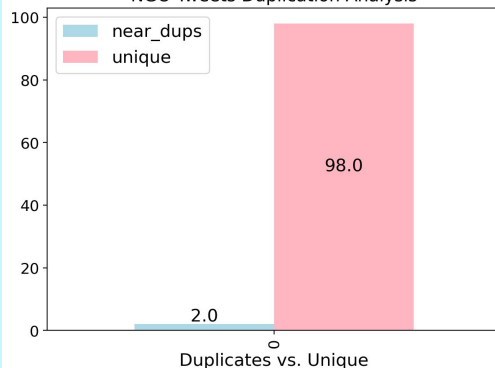
MESSAGE UNIQUENESS ANALYSIS

Whether organizations create their own tweets or copy paste others text (exclude retweets)

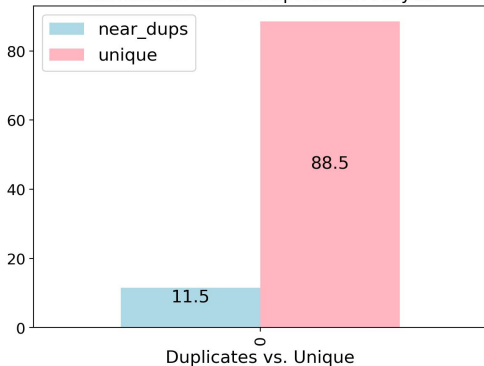
Overall Tweets Duplication Analysis



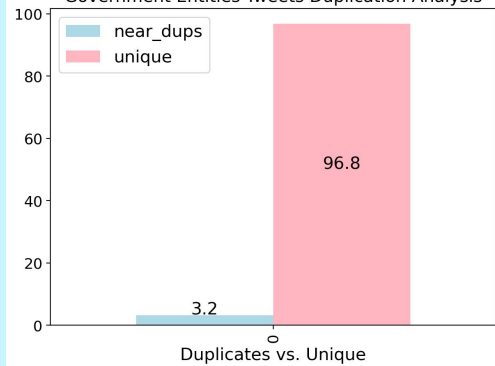
NGO Tweets Duplication Analysis



Universities Tweets Duplication Analysis



Government Entities Tweets Duplication Analysis

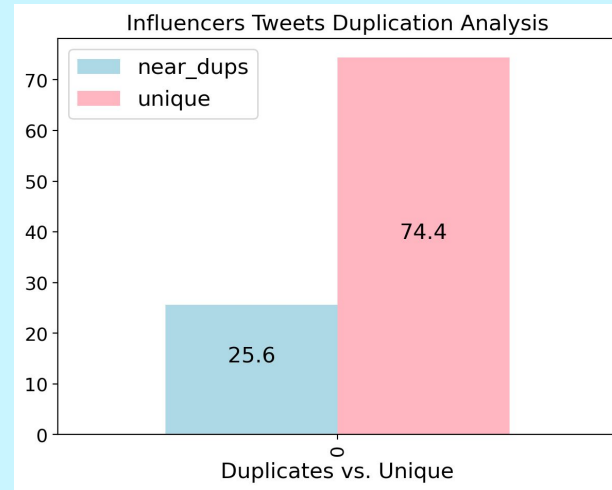
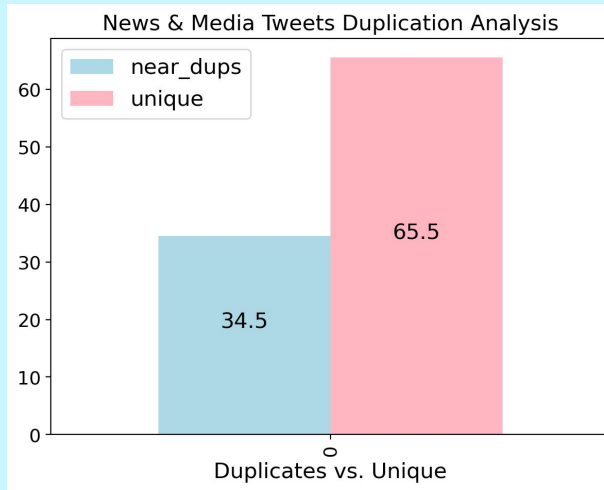
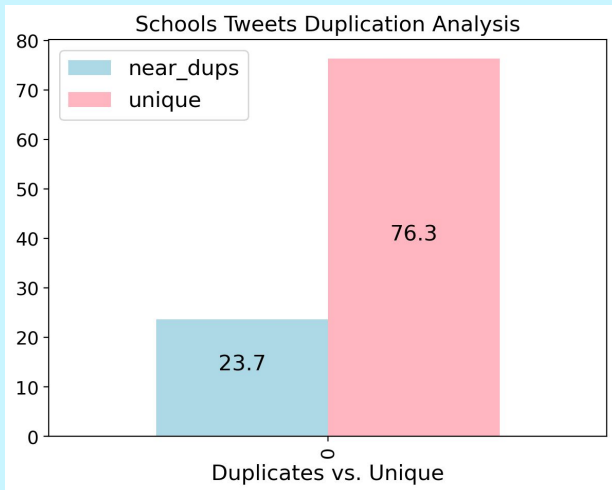


- Most organizations exhibit a high percentage of unique tweets, indicating that they generate their original content rather than copying and pasting others' content. Among them, **NGOs, Universities,** and **Government Entities** stand out with a lower percentage of duplicate tweets.
- **NGOs**, in particular, are highly active in tweeting about special education topics, with a 98% unique tweet count. This is not surprising, given their significant contributions to the advancement of special education for individuals with disabilities worldwide. As a result, they are likely to produce their unique content to disseminate information and raise awareness about special education topics.

Note: Values in above bar charts are by percentage. Ex: 98 means 98%

MESSAGE UNIQUENESS ANALYSIS

Whether organizations create their own tweets or copy paste others text (exclude retweets)



The tweets from **Schools**, **News & Media**, and **Influencers** on special education topics tend to be relatively similar and lack originality. This is understandable as these categories may not have extensive knowledge or familiarity with special education topics, resulting in less diverse messaging. Moreover, compared to other categories, News & Media and Influencers have a lower count of original tweets on special education topics, indicating that they do not tweet as frequently on this subject, which may explain why they have fewer unique tweets.

Note: Values in above bar charts are by percentage. Ex: 76.3 means 76.3%

CONCLUSION

1. It may not be accurate to rely solely on Twitter as a reliable source of information on special education, as most users are not verified.
2. Using solely the amount of tweets might not a good indicator of whether a twitterer is influential. Using engagement rate or influential score that considers a user's followers and reply counts may provide a better indication of a user's impact.
3. It's important to note that the high number of English-speaking Twitter users does not necessarily reflect the global distribution of opinions on special education. Additionally, a significant amount of location data, including location, country name, and country code, is lost, which may result in inaccurate geographical analyses, further proving that Twitter might not be a reliable source of information on special education.
4. Although there were several peaks in Twitter activity related to special education in July 2022, October 2022, and February 2023, not all were necessarily related to hot topics in the field. The peak in July 2022 may be related to the start of school year. On the other hand, the peak in October 2022, during which government entities had the highest tweet count, was likely due to the US election. Therefore, the time trend of tweet counts doesn't necessarily reflect topics in special education.
5. Twitter account creation peaked in the US in 2009, 2011, and 2019, while a peak was observed in Saudi Arabia in 2023. This may be due to the fact that special education has been developing earlier and more quickly in the US than in Saudi Arabia, and it has taken longer for the topic to emerge on Twitter in the latter country.
6. NGOs are the most active Twitter users in the special education field and their tweets are 98% unique. This is reasonable as NGOs have played a critical role in developing special education worldwide and creating their own content to spread informative messages. News & Media and Influencers have a lower count of original tweets on special education topics, indicating that they do not tweet as frequently on this subject, which may explain why they have fewer unique tweets.

Overall, it's important to consider the limitations of using Twitter as a source of information and to use multiple sources to ensure accuracy and completeness.

RECOMMENDATION

- **For verification issues:**
 - Twitter should consider implementing a verification process for users who tweet about special education-related topics. This could involve requiring users to provide some evidence of their expertise or involvement in the field of special education, such as a professional certification, membership in a relevant organization, or a record of published work. By verifying users who have a genuine interest or expertise in special education, Twitter can help ensure that users are able to find and trust accurate and reliable information on the platform.
 - Twitter could also consider partnering with special education-related organizations and experts to curate and promote high-quality content related to special education. By highlighting verified accounts and trusted sources of information, Twitter can help users identify reliable information and avoid misinformation or incomplete information on special education topics.
 - Twitter could explore ways to increase the visibility and reach of verified accounts and trusted sources of information on special education. This could involve promoting such accounts in search results, recommendations, or other areas of the platform, as well as providing tools or resources for users to filter or sort content related to special education based on credibility or other factors.
- **For influence measurements:** Twitter should develop several metrics that consider additional factors such as followers, reply counts, and tweet creation time on the database. By using these metrics, users can more accurately measure the influence of a given Twitter user without being biased by tweet counts or confused by various measures found on the internet.
- **For location data loss:** Twitter should develop strategies to encourage users to fill in location relevant information so that geological analysis can be conducted more accurately. In addition, this could help special education related NGOs to capture the trend by countries.
- **For target search:** To help users filter out relevant tweets more efficiently and accurately, Twitter should develop algorithms that automatically classify tweets based on user profiles and tweet text, using hashtags instead of relying on users to add them manually. This could save time and effort for users and enhance the overall quality of Twitter conversations on special education and other topics.

APPENDIX

- **Twitter engagement rate**

- engagement rate = $(\text{Retweet} + \text{likes} \setminus \text{number of tweets}) \setminus \text{Total number of followers}) * 100$
 - reference: <https://getmashhor.com/en/twitter-engagement-calculator>
- Engagement rate on Twitter is calculated as the sum of: (Likes + Retweets + Quotes + Replies) divided by the number of tweets, then by the total number of followers, then multiplied by 100.
 - reference: <https://mention.com/en/twitter-engagement-calculator/#:~:text=Engagement%20rate%20on%20Twitter%20is,followers%2C%20then%20multiplied%20by%20100.>

- **Influence score:**

- Kred Influence = $(\text{Retweets} + \text{Replies}) * (\text{Followers} + \text{Following})$
- Retweet Reach (RTR) score = $\text{RTR} = (\text{Number of Retweets} + \text{Number of Mentions}) / \text{Number of Followers}$
- Twitter Authority Score (TAS): $\text{TAS} = \log_{10}(\text{Followers}) \times \log_{10}(\text{Statuses})$
- Followerwonk Social Authority score = Social Authority = $(\text{Retweets} \times 0.5) + (\text{Mentions} \times 0.5)$
- $(\text{retweet_count} / \text{total tweet}) * \text{follower}$

There are various methods to measure twitterers' influence, however, as stated in recommendations, Twitter should develop several metrics that consider additional factors such as followers, reply counts, and tweet creation time on the database. By using these metrics, users can more accurately measure the influence of a given Twitter user without being biased by tweet counts or confused by various measures found on the internet.

REFERENCE

- **Keyword for education:**

(1) The Complete Guide To Twitter Hashtags For Education: <https://www.teachthought.com/twitter-hashtags-for-teacher/>

- **Keyword for special education:**

(1) https://en.wikipedia.org/wiki/Special_education

(2) THE EDVOCATE'S LIST OF 123 TWITTER FEEDS FOR SPECIAL EDUCATORS:

<https://www.theedadvocate.org/edvocates-list-123-twitter-feeds-special-educators/>

- **Variable definition:** <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet>

- **Research on Role of NGO's in Improving Education Facilities of Specially Abled:**

<https://psychosocial.com/article/PR201164/11863/#:~:text=It%20is%20a%20recognized%20fact,institutional%2C%20charity%2Dbased%20approach.>

- **Influencer's definition:**

- <https://blog.hootsuite.com/how-much-do-influencers-make/#:~:text=10%2C000%E2%80%93350%2C000%20followers%20%3D%20Micro%2D,1%2C000%2C000%2B%20followers%20%3D%20Mega%2Dinfluencers>
- <https://influencermarketinghub.com/what-is-an-influencer/>
- <https://getgist.com/influencer-synonyms/>

- **Government department & agencies keywords:** <https://www.usa.gov/federal-agencies>

- **NGO keywords:** <https://www.gdrc.org/ngo/ngo-keywords.html>

- **University/ Schools/ News Keywords:** ChatGPT