Climate Change Awareness on the Twitter Domain

Prepared by Danil Meresenschi

Climate Change Education

- In August 2022, Stefan Ellerbeck, from the World Economic Forum, wrote an article covering a research initiative on climate change social distress and the United Nations outlook on educating the population on the issue.
- In a study done by University of Bath, it was revealed that across the 10000 young individuals surveyed in 10 countries → 75% were anxious about the climate state of the planet and believed that the “future is frightening.”
- In an effort to relief public concerns and educate society on the matter, the UN is calling for climate education to become compulsory in schools from 2025.
- Despite being part of the Paris Agreement, only certain countries are integrating the curriculum into their nation’s education system.

Executive Summary

Twitter is a more personable medium of information communication compared to traditional media outlets. Moreover, the platform is designed to elevate and represent voices and opinions from altering credentials. From institutionalized informers to independent investigators involved in activism, all parties are able to attempt educating the masses on a cause. Despite the controversies around the company's policies, the social media is widely considered a vehicle of free speech and expression that propels various outlooks. The objective of the project is to investigate what agents are spreading awareness on the issue of climate change and weather the content can be considered credible and impactful.

Assumptions

- Following and engagement are congruent with credibility and impact strength.
- Certain graphs, such as the location pie charts(more interpretable percentage frequency distribution across values), are modified to better outline a trend.
- The sampled data can be considered to be representative of the general population.
- The metrics that were established to be impactful in persuading the project's argument, are correlated to the argument.

Methodology & Insight

1. Data Filtering and Partitioning
    a. The original tweet dataset consisting of roughly 100 million records was filtered across rows and columns.
    b. Through sampling testing data frames and implementing filtering techniques the final data set included around 90000 records.
        - 99.912% of the initial dataset was filtered out.
    c. Rows were filtered on the premise of the tweet text including the string climate change.
    d. Columns were filtered down from the complex nested JSON structure to 12 columns in tabular format that were deemed to be important for the analysis.
    e. A 11% of total final dataset sample was taken, roughly 10000 records, for more efficient analytics to be done.
    f. For modularization purposes the sampled and final dataframe were saved in a parquet file to be utilized in the data analysis and data application notebooks, respectively.

Methodology & Insight

2. Data Analysis Test
    a. The analysis was done and tested on the previously mentioned sample dataset, all the steps are re-executed in the application notebook to provide insight, while this section is an in-depth overview of the procedural data engineering and transformations.
    b. At first certain columns needed adequate casting to be converted into proper data types. Those mainly involved converting columns into integers, to ensure proper aggregation in future functions.
    c. The **First part** of the analysis was designated to identifying top tweeters based on retweet and reply counts of a tweet, along with a profile's friends and followers.
        - After some feature engineering, a new column was created as a result of name and source name concatenation for more informative results.
        - Using the groupby function the 4 metrics were the summed output in descending order and yielded the statistic per profile with added description of the profile.
        - Bar Charts of the distributions was then created with adjusted scales for clearer interpretation.

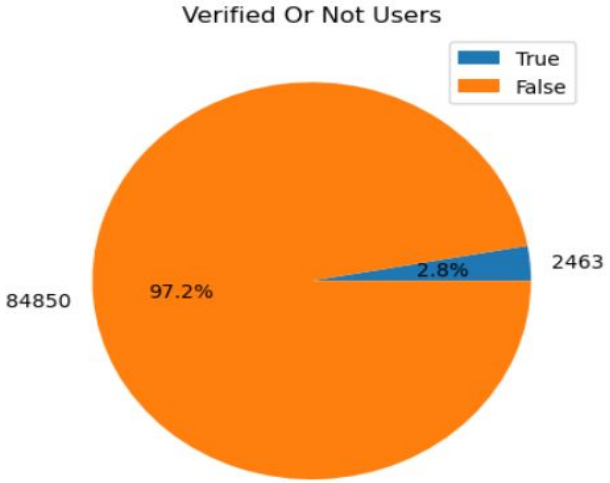
Methodology & Insight

2. Data Analysis Test
    d. In the **Second Part** the goal was to find the percentage frequency distribution of verified and not-verified users. Using the groupby count aggregation query, the outputs were visualized in a pie chart as percents of the total dataset.
    e. The **Third Part** involved aggregating replies and retweets based on location and identify the locations with the highest number of those statistics.
        - After grouping by designated locations and removing groups that had no analytical capacity, a pie chart with percentage partitions was created to interpret the location of the Twitterers from their profile specifications.
    f. The **Fourth Part** included analysis done on the premise of time. The largest number of retweets and replies were investigated across months and years.
        - The month and year columns were created after using extract appropriate values from the date column.
        - The time-based columns were aggregated and the results were ordered in decreasing order to be visualized with a bar chart.

Methodology & Insight

3. Data Application & Insight
    a. Upon testing the functionality of the analysis on a smaller sample, the used methodologies and queries were applied on the full final dataset to gain insight on the trend of climate change awareness and education.
    b. **Verified Accounts**
        - A small fraction of the tweeters involved in the discussion of climate change are actually twitter verified, which may signify independent awareness spread.

verified	count(verified)	percent_of_total
True	2463	2.820886
False	84850	97.179114

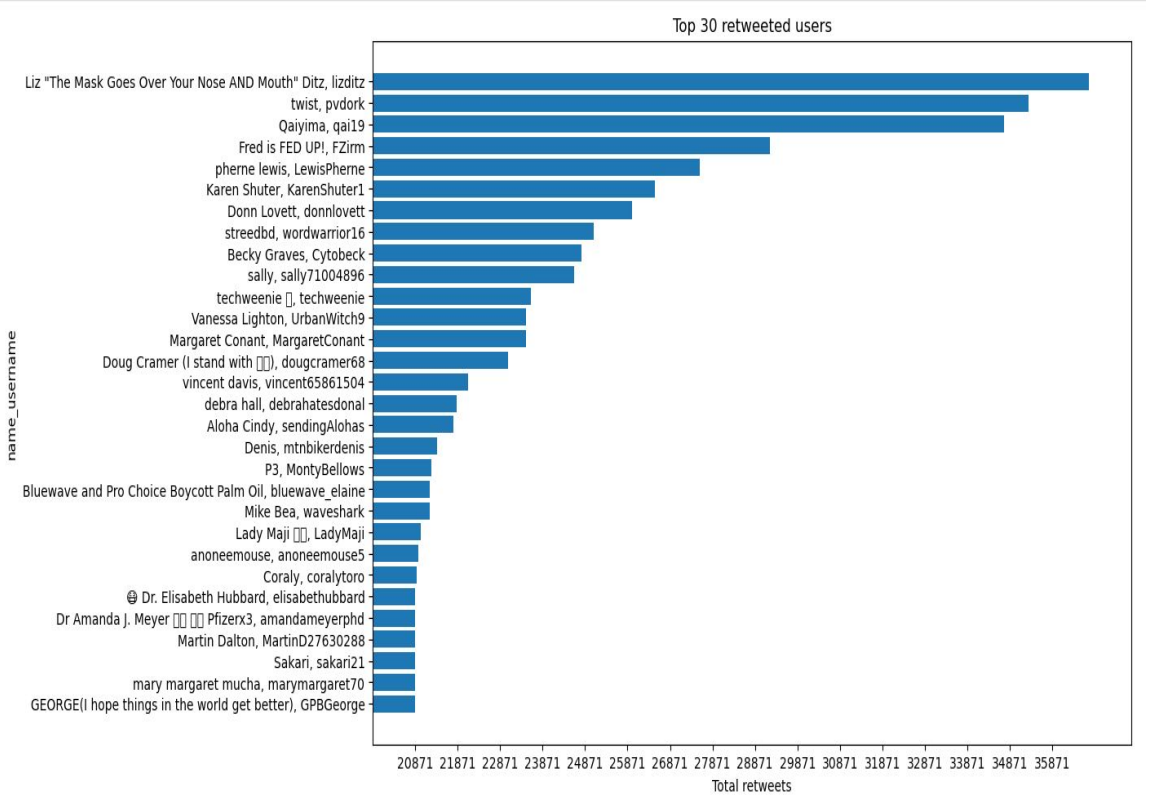


Methodology & Insight

3. Data Application & Insight

c. Top Twitterers

- Both top retweeting and replying twitterers seem to have little to no institutional affiliation, and share activist content to promote the issue.
- Number of retweets is significantly greater than replies, which is adequate since more engagement occurs this way.
- Check Appendix 1 for the top 30 replied to users graph.

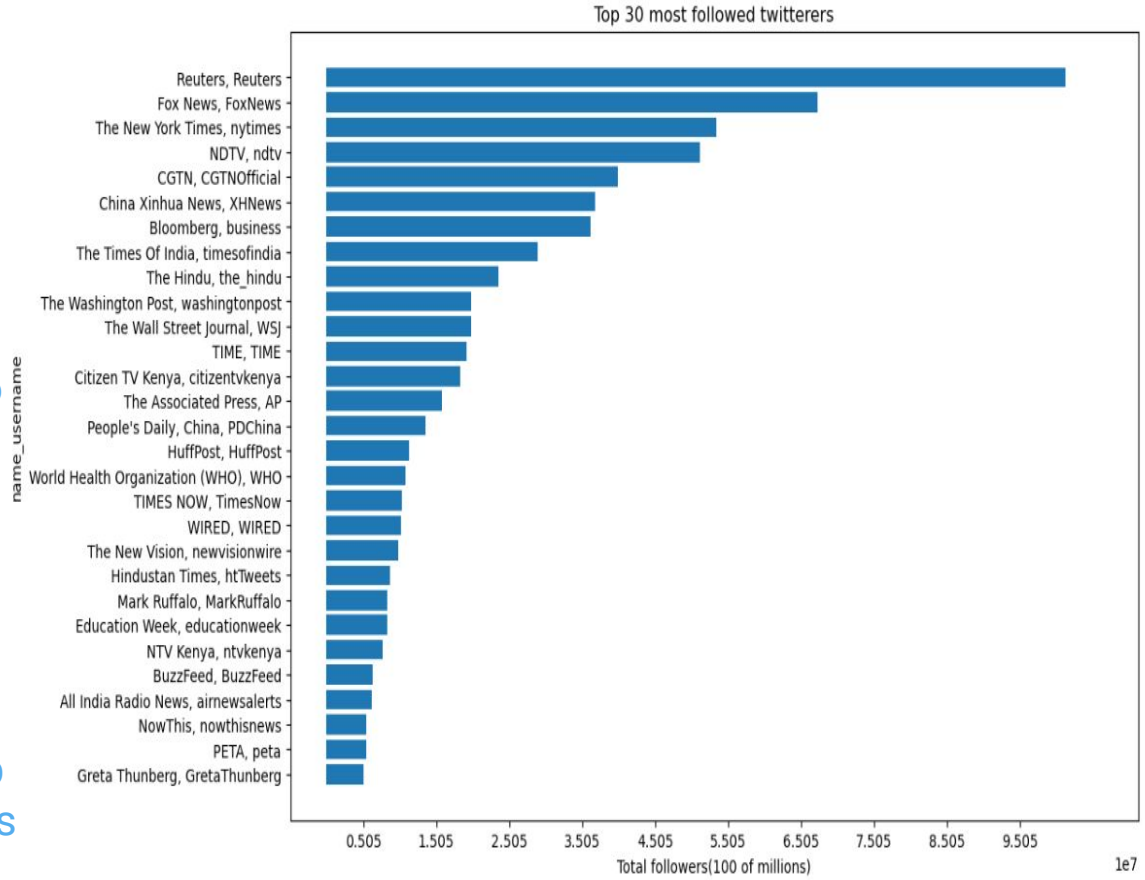


Methodology & Insight

3. Data Application & Insight

c. Top Twitterers

- The top of the most followed user base, are mostly mainstream news portals and high-profile celebrity/activists working to communicate the issue of climate change.
- The top most befriended twitterers, however, are a list of unaffiliated users promoting the cause.
- Check Appendix 2 for the top 30 most befriended twitterers graph.

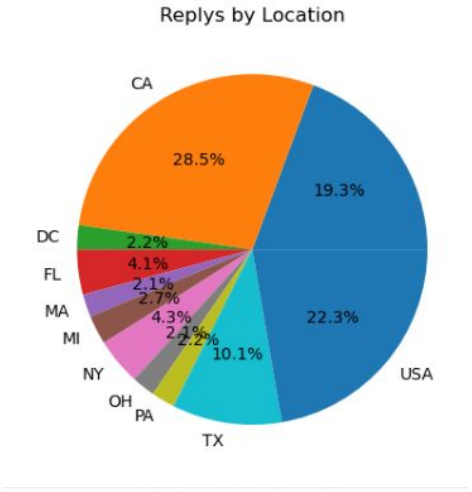


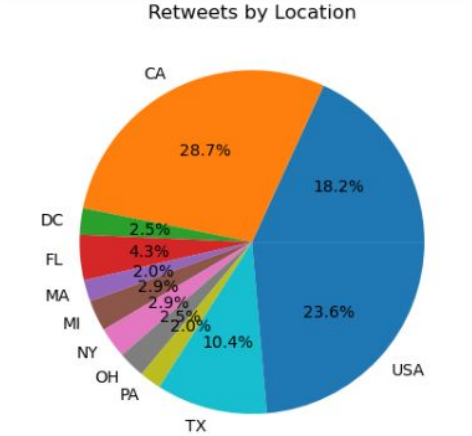
Methodology & Insight

3. Data Application & Insight

d. Top Tweeting Locations

- There is no significant percent distributed difference between the Replies and Retweets across locations.
- The unlabeled piece of the pie chart represents the remaining locations not included in the initial location choice.
- Considering Texas is a Republican state, and California is a Democratic state, the usual republican stance on climate change tends to be less strict as opposed to the democratic point of view.
- This might suggest that there since when investigating defined location, Texas and California have the largest individual percent share → could be indicative that there is great variety of though on Twitter when it pertains to climate change.
- Check Appendix 3 for University of Texas/Texas Tribune Poll, which displays the sentiment division on the issue across political party affiliations.



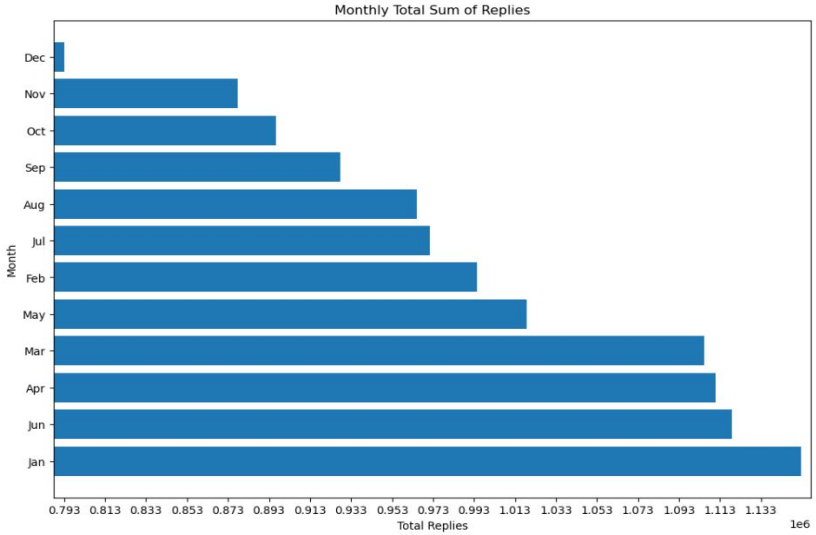


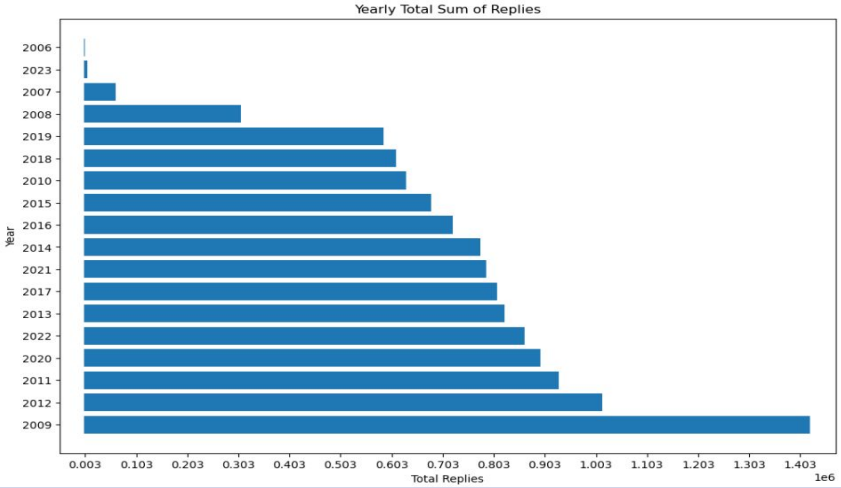
Methodology & Insight

3. Data Application & Insight

e. Top Tweeting Timelines

- The total sum trend for months and years is similar to the retweets metric.
- Jun and Jan are typically the hottest and coldest months of the year, respectively → The issue of climate change is most noticeable then, hence the increased tweet traffic.
- The immense tweet traffic in 2009 may be tied to The United Nations Climate Change Conference, Copenhagen 2009. The event fueled a lot of discussion around climate change.
- Check Appendix 4 for Yearly and Monthly Total Sum of Retweets.





Conclusion

The instantaneous mode of communicating information on the Twitter platform along with its liberacy in expression, serve as great foundations for addressing the ‘educatee’ of the 21st century. If one could filter through the content and reason rationally when interacting on the platform, the social media giant encapsulates practically every aspect of an adequate learning hub. In accordance with the completed analysis, it appears that there is a good mix of outlooks from various backgrounds on the same topic. The opportunity to assess an issue from many angles is crucial in forming an established and educated perception. Additionally, there is an element of virality present on Twitter, which allows user to immediately exchange contentions on freshly-occurring events. Moreover, the platform gives voice not only to established institutions, but to enthusiast who are looking for tools to fuel their desire to share.

Thank you for an amazing quarter!

References

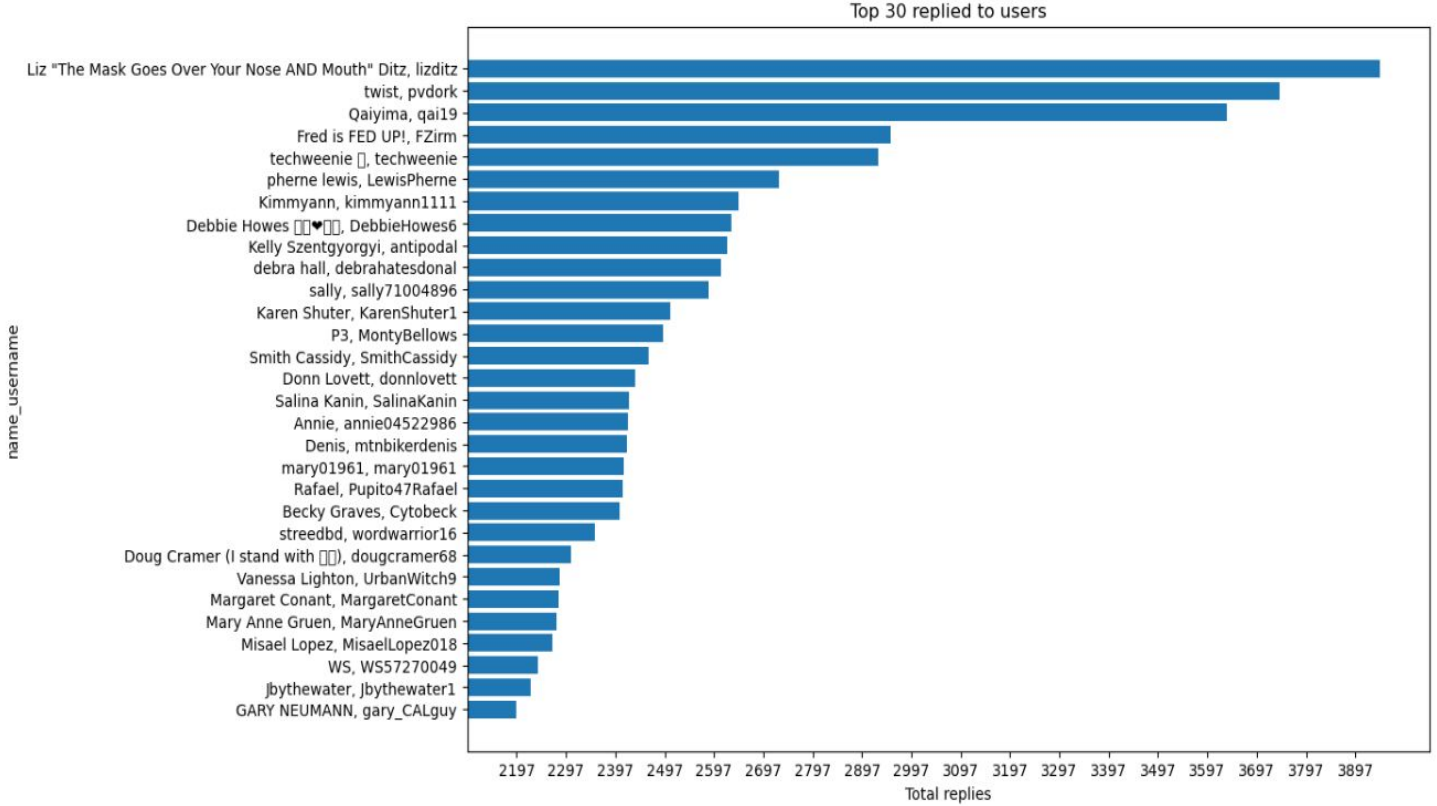
COP 15 and CMP 5. Unfccc.int. (n.d.). Retrieved March 5, 2023, from https://unfccc.int/process/conferences/pastconferences/copenhagen-climate-change-conference-december-2009/statements-and-resources/outcome-of-the-copenhagen-conference

Ellerbeck, S. (2022, August 24). Should schools teach climate change studies? these countries think so. World Economic Forum. Retrieved March 2, 2023, from https://www.weforum.org/agenda/2022/08/climate-change-schools-education/#:~:text=The%20UN%20is%20calling%20for,objective%20in%20the%20Paris%20Agreement.

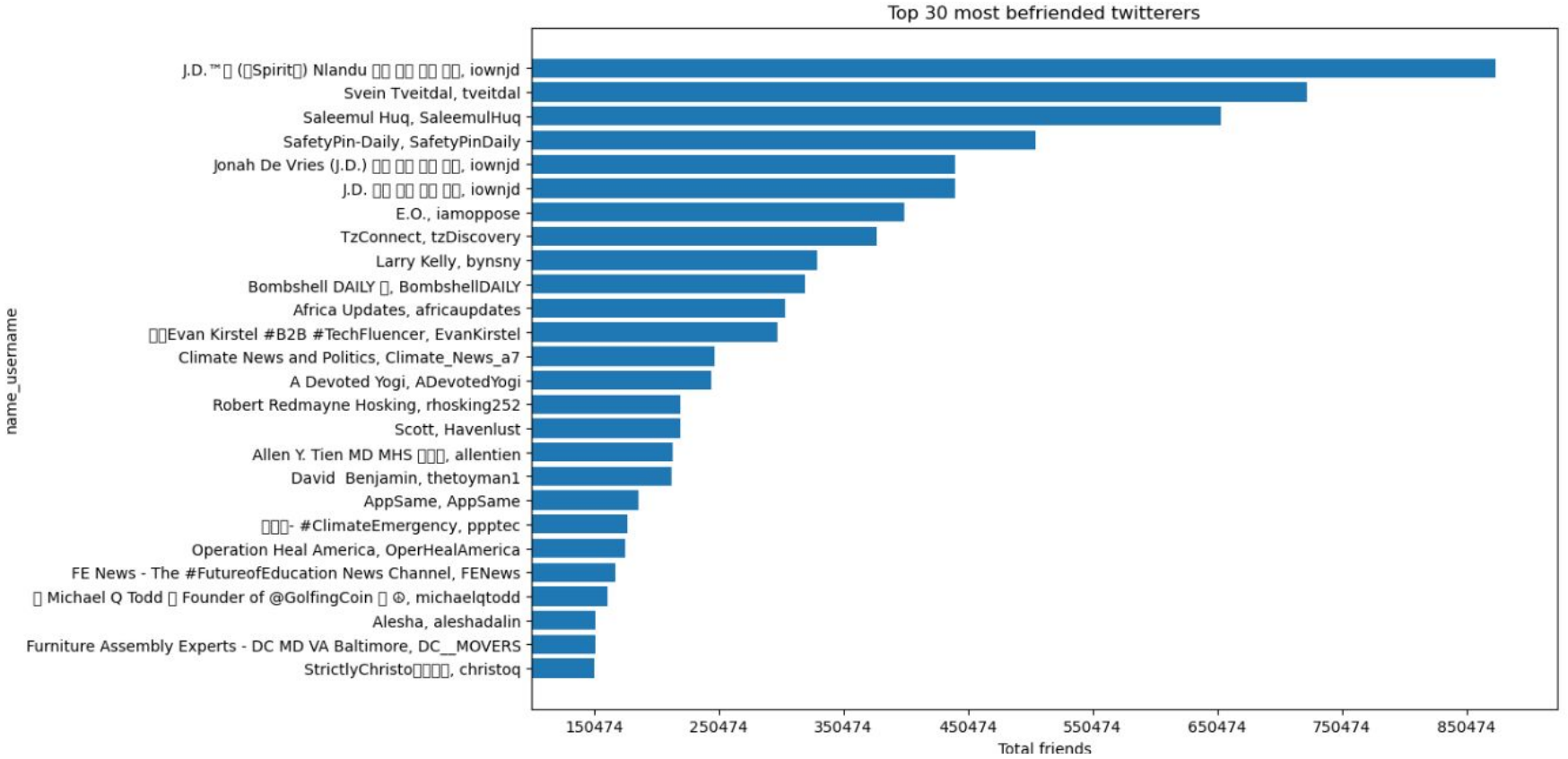
Hutchinson, A. (2021, November 16). New study shows that 25% of Twitter users produce 97% of all tweets. Social Media Today. Retrieved March 6, 2023, from https://www.socialmediatoday.com/news/new-study-shows-that-25-of-twitter-users-produce-97-of-all-tweets/610146/#:~:text=50%25%20of%20all%20tweets%2C%20based,use%20of%20Quote%20Tweets%20here).

Ramsey, R. (2019, November 6). Texans say climate change is happening, but it's a highly partisan issue, UT/TT Poll finds. The Texas Tribune. Retrieved March 5, 2023, from https://www.texastribune.org/2019/11/06/texans-say-climate-change-happening-highly-partisan-issue-uttt-poll/

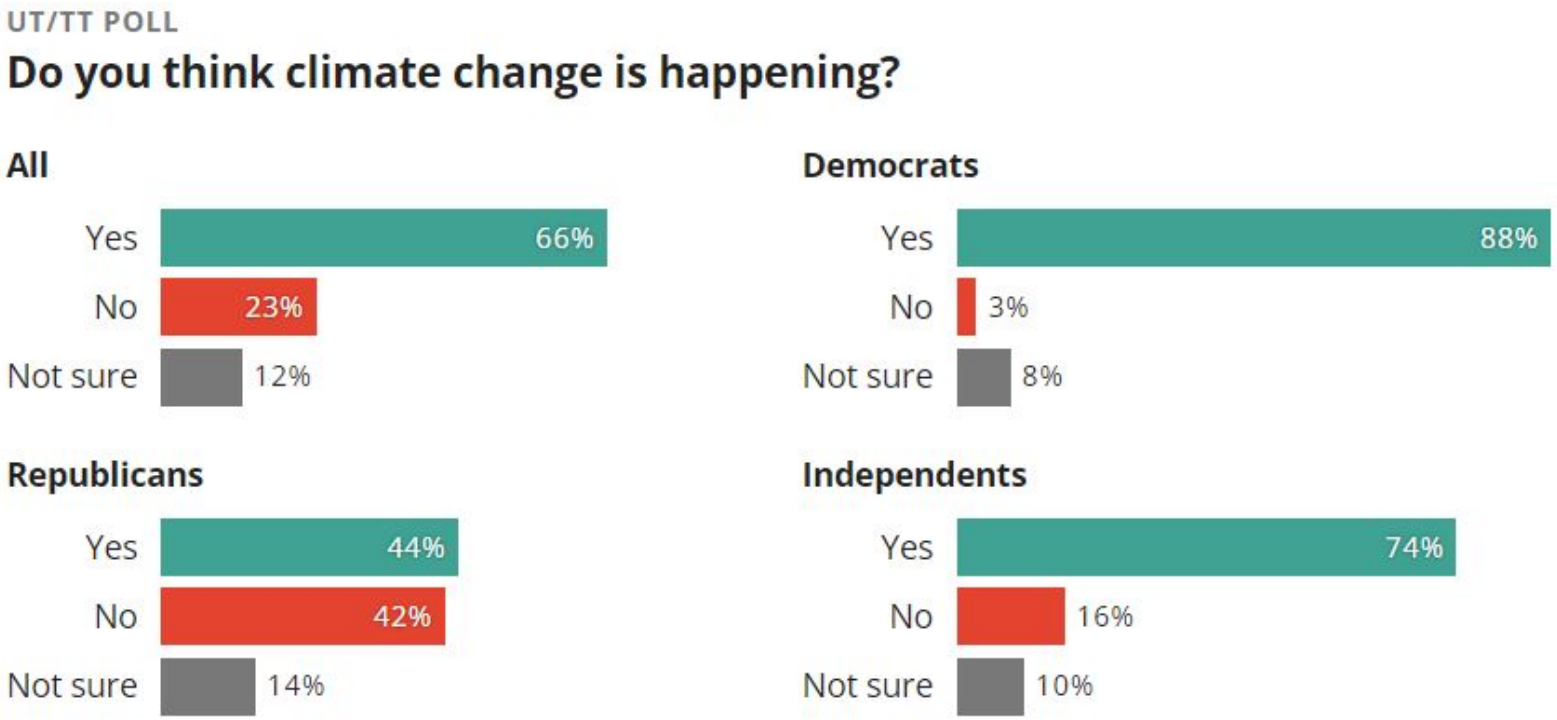
Appendix 1. Top 30 replied to users



Appendix 2. Top 30 most befriended twitterers



Appendix 3. University of Texas/Texas Tribune Poll



Appendix 4. Yearly and Monthly Total Sum of Retweets