

1. TOPIC & KEYWORDS

Our group's chosen topic for analysis is on **Donald Trump's second impeachment**. Given such a recent event that has garnered much attention from around the world, we were interested in finding out the sentiments of people regarding this event, especially since the public is known to be distinctly split among Trump supporters or the opposition. This was also the first time in US history that a president was to be impeached by the House twice, and to be tried even after leaving office. The relevant keywords we used were: a) Trump Impeachment and b) Impeach Trump.

While it is possible to search for particular keywords for Twitter, only hashtags can be used to retrieve relevant posts from Instagram. As such, our keywords are in the format of hashtags when scraping for data on Instagram (i.e. #trumpimpeachment, #impeachtrump). One limitation about the keywords we used is that there is no way to distinguish between Trump's first impeachment from his second one. This was unavoidable, as there was no hashtag or keyword unique to his second impeachment. The following is a timeline of events:



2. DATASET

Data was scraped starting from 6 January 2021 (when the Capitol Hill riot occurred and calls began to be made for a second impeachment) until 17 January (the day we conducted scraping). Python packages used were **Twint for Twitter** and **InstaLoader for Instagram**. The number of comments collected was 175,749 for Twitter and 6,660 for Instagram, which adds up to a total of 182,409.

3. DATA CLEANING & EDA

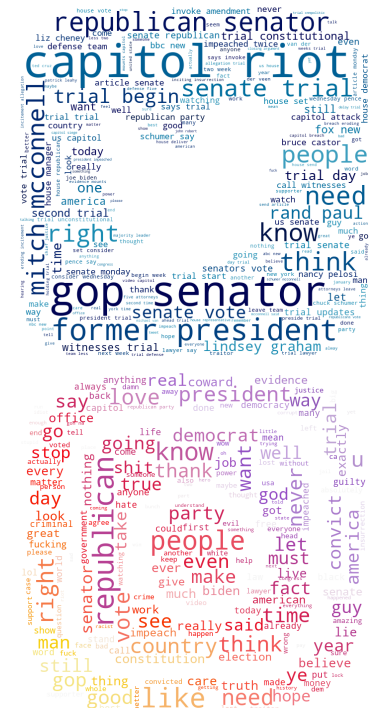
We conducted basic EDA to check if the comments that we collected were useful for analysis. Rows with no comments, non-english comments, or comments that were auto-generated were dropped. Text preprocessing was then performed to clean the data, by tokenising, changing to lowercase, keeping only words and removing stopwords. Additional stopwords specific to each platform were added after their initial WordClouds were generated. Emojis were kept because Vader can analyse them and be incorporated into the compound sentiment score, which we will be analysing further on.

4. ANALYSIS & RESULTS

4.1 Difference in content

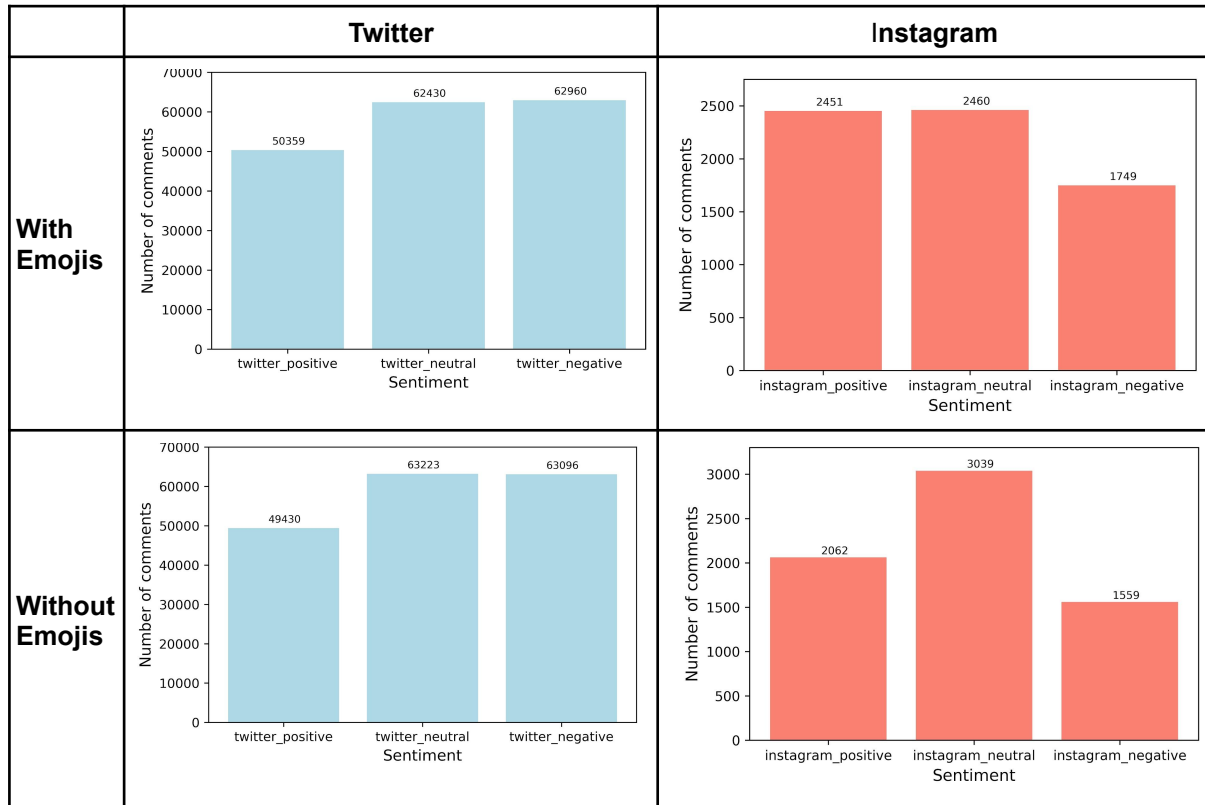
Word clouds were used to analyse the difference in content of comments on the platforms. We can see that the word cloud for Twitter has brought up many **topics directly related** to Trump's impeachment such as the Capitol riots that Trump incited. 'Mitch McConnell' was also a top related topic because of his feud with Trump in which he blamed Trump for the riots despite voting against Trump's impeachment, thus creating much online chatter. Other top keywords are related to the senators ("republican senator" and "gop senator"). GOP stands for Grand Old Party, which also refers to the Republicans. People often use Twitter as a news source and retweet news they see, which seems to result in **top topics on Twitter being mostly factual and related to news about the trial.**

On the other hand, we can see that the word cloud for Instagram is quite different. Despite “republican” being one of the top keywords, the **majority of other top keywords are less directly related** to the Impeachment, and **more on expression of emotions** such as ‘like’, ‘hope’, ‘love’, ‘good’, ‘truth’, ‘coward’ and ‘free’. Perhaps this is because Instagram has no retweet function, and users are likely to comment their thoughts on the post directly.



4.2 Difference in sentiment

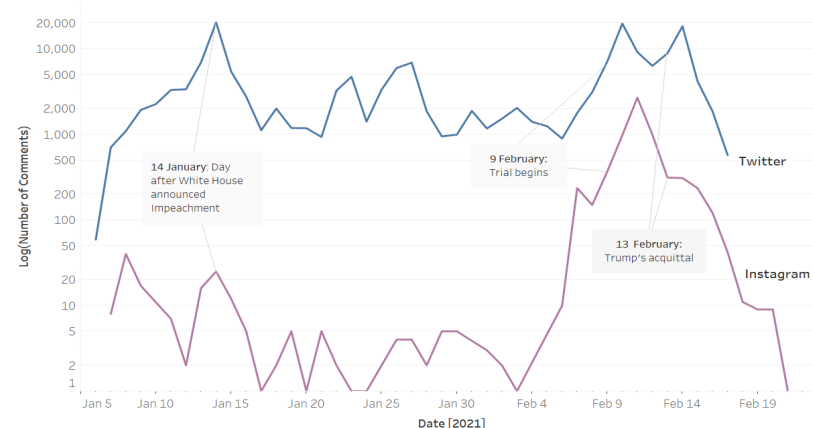
Sentiment analysis was done on preprocessed text with emojis, and with emojis removed. While there was **little difference in the results for Twitter**, there was a **clear increase in proportion of comments that were analysed as having positive sentiment on Instagram when emojis were included**. This is perhaps due to the much greater use of emojis by Instagram users to express their sentiment. As a whole, there is an obvious distinction between the two platforms: generally, Instagram users were more positively inclined regarding Trump's impeachment while Twitter users were more negative.



4.3 Time-series Comparison

A time-series comparison was generated using Tableau. We can see that for both Instagram and Twitter, there is generally a similar trend in the number of comments that have been retrieved across time. Corresponding to the timeline mentioned earlier, the number of comments posted per day peaked around the period when 1) Trump's impeachment article was released, 2) the trial began and 3) Trump was acquitted.

Timeseries comparison of number of comments between Twitter and Instagram



5. AUTHOR CONTRIBUTION

Chen Jian Yu: Merged datasets, created sentiment charts, wrote 'Analysis & Results' (50% of writing). **Joshua Wong:** Collected/Cleaned Twitter data, created WordClouds, wrote 'Keywords' (25% of writing). **Ow Ling Jia:** Collected/Cleaned Instagram data, created time-series chart, wrote 'Datasets' and 'EDA' (25% of writing).

REFERENCES

NPR. (2021). Trump Acquitted In Impeachment Trial On Capitol Insurrection: Trump Impeachment Aftermath: Updates. Retrieved 21 February 2021, from <https://www.npr.org/sections/trump-impeachment-trial-live-updates/2021/02/13/967098840/senate-acquits-trump-in-impeachment-trial-again>

The Straits Times. (Feb, 2021). Trump urges Republicans to dump Senate leader Mitch McConnell. (2021). Retrieved 21 February 2021, from <https://www.straitstimes.com/world/united-states/trump-urges-republicans-to-dump-senate-leader-mitch-mcconnell>