Members: Anjali Nuggehalli, Angie Zhou, Hugh Chapin

Summary:

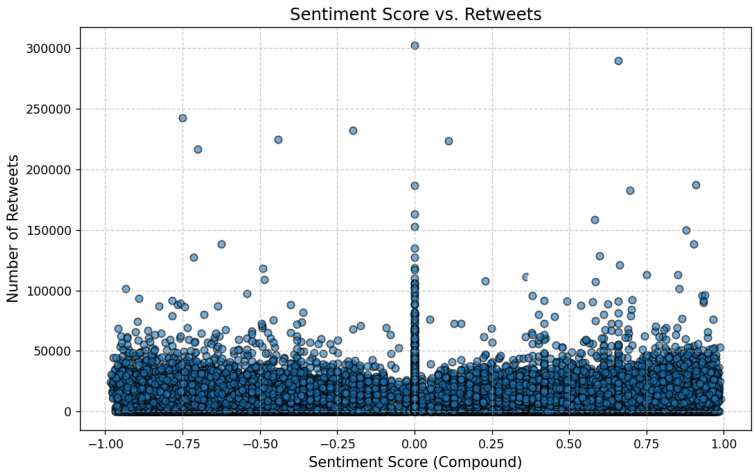
- Picked our project idea, dataset, ethical focus
- Submitted project proposal

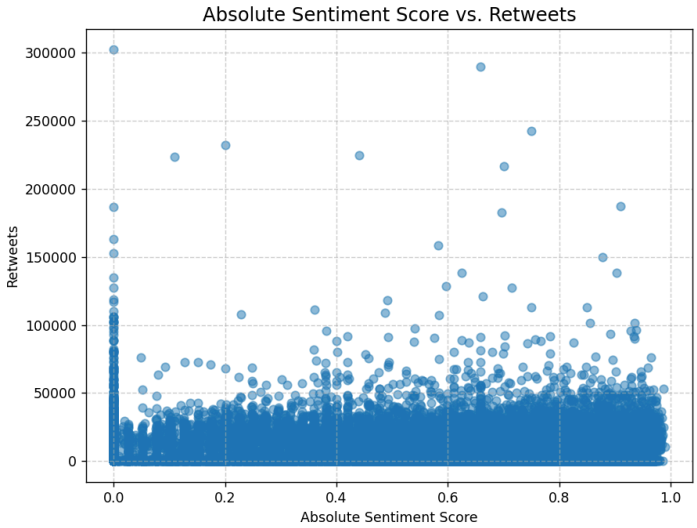
Results:

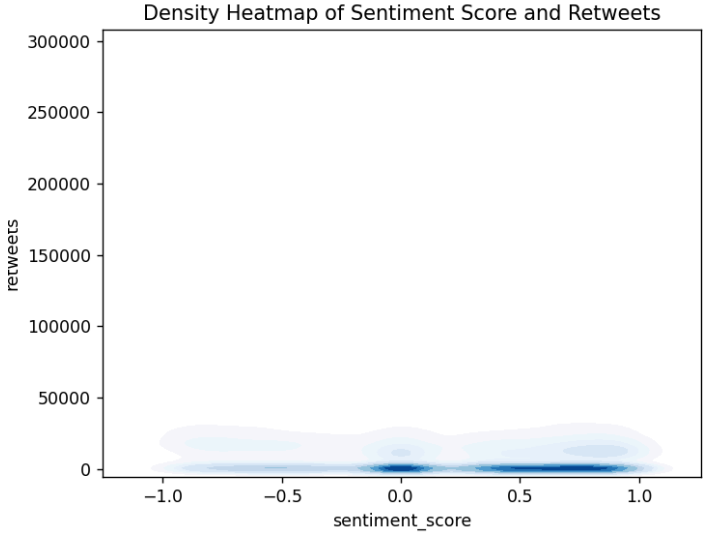
- Example sentence: ". @ RichLowry is truly one of the dumbest of the talking heads - he doesn't have a clue!" (# retweets: 767 #favorites: 1530)
    - VADER output: {'neg': 0.153, 'neu': 0.724, 'pos': 0.123, 'compound': -0.1759}
    - Spacey output:
        - . -> PUNCT
        - @ -> X
        - RichLowry -> PROPN
        - is -> AUX
        - truly -> ADV
        - one -> NUM
        - of -> ADP
        - the -> DET
        - dumbest -> ADJ
        - of -> ADP
        - the -> DET
        - talking -> VERB
        - heads -> NOUN
        - - -> PUNCT

        - he -> PRON
        - does -> AUX
        - n't -> PART
        - have -> VERB
        - a -> DET
        - clue -> NOUN
        - ! -> PUNCT
    - Hugging Face transformers: [{'label': 'NEGATIVE', 'score': 0.9998005032539368}]
- Dataset Analysis:
    - The mean compound Vader sentiment score for all of Trump's tweets is 0.22, but when looking at his tweets after 2015, the average sentiment score lowers to 0.20.
    - The dataset contains mentions of Trump in addition to just his own tweets. When we removed any tweets that mentioned @realDonaldTrump, the average sentiment post 2015 goes to 0.18.
    - With these changes to the dataset, our size moved from 43352 tweets to 14790 tweets (post 2015 and no mentions @ trump).
    - Given these results, we made several plots. We charted the sentiment score of his tweets on the x-axis, and the number of retweets as measure of engagement on the y-axis. We also charted absolute sentiment value, to see if higher positive or negative sentiment in general had a relationship with retweets. Finally, because it can be difficult to see all of the data points, we also made a heatmap for the

regular sentiment analysis. For a preliminary analysis, it is difficult to make a conclusion about their relationship.







Ethics: We plan to address a research question while evaluating our results, focusing on whether more emotive tweets tend to capture a larger audience. We will then discuss the implications of this in the political sphere, emphasizing the need for greater awareness and improved detection of fake news that manipulates emotions to gain credibility. Ultimately, our project is deeply

rooted in ethics, maintaining a consistent emphasis on the ethical implications throughout its execution.

Problems:

- Beyond running VADER, we are thinking about other ways we can analyze the corpus to better understand it. We are still in the deliberation phase of this, and are looking for more tools to utilize. We tried doing sentiment analysis with Spacy Text Blob, but it we weren't able to get it to work. We also want to figure out better ways of extracting the results from what we currently have.
- Another issue is that if the tweet doesn't have any explicitly negative words, the VADER model will indicate it as a neutral sentence even if the context behind it is negative (ex: "I wonder when we will be able to see @ BarackObama's college and law school applications and transcripts. Why the long wait?"
- The data we're using is not fully cleaned, so we will have to do some more preprocessing in order to get accurate results.
    - Some of the tweets in the dataset are mentions @realDonaldTrump, and there are many tweets that happened before Trump ran for president and got into politics.

Hours:

- Anjali - 2 hours
- Angie - 2 hours
- Hugh - 2 hours

Code:

Github repo: https://github.com/anjalinugg4/NLP_FinalProject