**DS 710 Final Project -- Executive Summary**

**Introduction**

Racism and sexual harassment are taking center stage in the news today and in social media due to the current racial tension and a deluge of sexual harassment allegations against prominent celebrities and politicians. As such, I wanted to find out whether there’s a difference between two current Twitter movements involving these issues -- #metoo and #blacklivesmatter – as far as engagement and whether there’s a more passionate group. Kicked off in October 2017, #metoo calls for women to share their stories about sexual harassment and #blacklivesmatter was sparked in 2013 by the acquittal of George Zimmerman in the shooting death of African-American teen Trayvon Martin.

This information could be interesting to the social science community, politicians and businesses to understand how active and passionate people are on social media when talking about race vs. sexual harassment. It could also help prioritize future legislation to combat racism and sexual harassment and identify how and when to engage with people of these two movements.

**Data Collection & Analysis**

Leveraging the REST API for Twitter, I gathered 1,072 original tweets (no replies or retweets) from #metoo and 1,053 tweets from #blacklivesmatter (so 2,125 tweets total) between Dec. 8-10, an arbitrary timeframe. To ensure engagement could be maxed, I waited to gather tweets until Dec. 15. To understand which movement had the most engagement, I added the number of favorites and retweets for each tweet. To measure how passionate people are who participate in each movement, I used three different elements:

* Passion – Here, I assumed that passionate people use more exclamation points, question marks and capital letters to express themselves about an issue. These three values were added together for each tweet to measure Passion.
* Sentiment – I also assumed that passionate people express emotion through certain words, which can be tagged as positive or negative as a measure of Sentiment. To measure Sentiment, I used the TextBlob package in Python to provide a polarity score (between -1 and 1 with 0 being neutral).
* Tweet Length – Finally, I assumed that people who write longer tweets are more passionate about the subject because they have a lot to say. Since #blacklivesmatter would have the upper advantage of tweet length because it’s 11 characters longer than #metoo, I removed each hashtag from each tweet in both samples.

While #metoo is a newer movement than #blacklivesmatter and has been bolstered by recent news of sexual assault charges against prominent politicians and celebrities, racism is also very much at the top of the news. As such, I hypothesized that no significant difference exists between the two movements, as far as Engagement, Passion, Sentiment and Tweet Length. In my hypothesis, one movement would have to dominate each of the three categories (Passion, Sentiment and Tweet Length) to be determined as a more passionate group of people.

From a statistical standpoint, my null and alternative hypotheses for all four categories are HO: µmetoo = µblm; and HA: µmetoo ≠ µblm. All hypotheses were evaluated against a significance level of α=0.01.

**Results & Conclusion**

**Engagement:** After applying my two-sided t-tests, I found that there is not strong evidence of a difference in engagement between both movements (see means and p-value in Table 1). As you can see in Figures 1.1 and 1.2, about 75% of tweets for both movements received 0-3 engagements (a very low amount), and both boxplots show the two movements have equal medians and an almost exact distribution of values each with a few extreme outliers from popular Twitter accounts. This low engagement for both groups indicates followers might not want to engage because they don’t agree or don’t want to publicly agree to avoid conflict. Adding the number of replies for each tweet might’ve helped to provide additional insight into the engagement score difference between the two groups, but Twitter doesn’t currently support developers with this capability. Future analysis using this factor will be valuable to see if it changes this result.

**Passion/Sentiment/Tweet Length:** Overall, neither movement dominated all three categories. Two-sided t-tests (Table 1) revealed strong evidence of a difference in Passion and Sentiment between the two movements with p-values significantly less than α=0.01. However, #blacklivesmatter had a higher overall mean (Table 1) and median for Passion (Figure 3) and #metoo demonstrated more polarity, especially in positive Sentiment (see Figures 4.1 and 4.2). The Tweet Length t-test demonstrated evidence of a difference in both means and medians (Figure 2) but it wasn’t significant enough.

In conclusion, the results do not indicate that one movement has more engagement or more passionate people than the other, which aligns with my original hypothesis. This could indicate both movements should be prioritized similarly by politicians and businesses, if they choose to engage with and discuss these movements. Future in-depth analysis over time, especially during profound events involving racism and sexual harassment, might reveal more distinct differences between the two groups, as this study was merely a sample taken during one short time period. Additionally, analysis into the specific issues being discussed and language being used might also provide more actionable takeaways.

**Figures and Table**

|  |  |
| --- | --- |
|  |  |
| Figure 1.1 – Side-by-side comparison of engagement score distributions for #metoo and #blacklivesmatters. | Figure 1.1 – Side-by-side comparison of engagement score distributions for #metoo and #blacklivesmatters (magnified to show similar medians and distributions). |
|  |  |
| Figure 2: Side-by-side boxplots of Tweet Length distribution for #metoo and #blacklivesmatter tweets. | Figure 3: Side-by-side boxplots of Passion distribution for #metoo and #blacklivesmatter tweets. |
|  |  |
| Figure 4.1: Sentiment analysis for #metoo tweets, segmented into the frequency of negative (red), neutral (yellow) and positive (green) tweets. | Figure 4.2: Sentiment analysis for #blacklivesmatter tweets, segmented into the frequency of negative (red), neutral (yellow) and positive (green) tweets. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1** | #metoo  (means for each category) | #blacklivesmatter  (means for each category) | p-value |
| Engagement | 8.371269 | 7.305793 | 0.6473 |
| Passion | 11.03265 | 15.11491 | 6.61E-11 |
| Sentiment | 0.10296882 | 0.01684822 | 4.30E-13 |
|
| Tweet Length | 129.6241 | 137.8452 | 0.01041 |
|