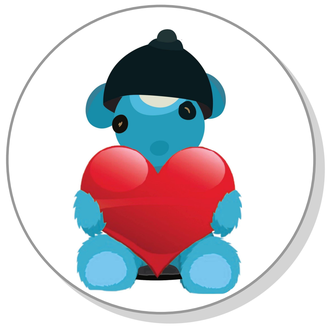
Signed, Blueno: Stress and Anonymous Forums

A temporal analysis of Dear Blueno and Blueno Bears Admirers

Authors

Jed Fox, Zahid Hasan, Nick Keirstead, Zuhal Saljooki

jfox10 zhasan1 nkeirste zsaljook



Introduction

- We categorized Dear Blueno (DB) and Blueno Bears Admirers (BBA) posts by week
- We identified some weeks as "stressful": shopping period, reading period, and finals week.
- We used sentiment and frequency analysis to see if we could find a statistically significant change in posting patterns during these stressful times.

Hypothesis

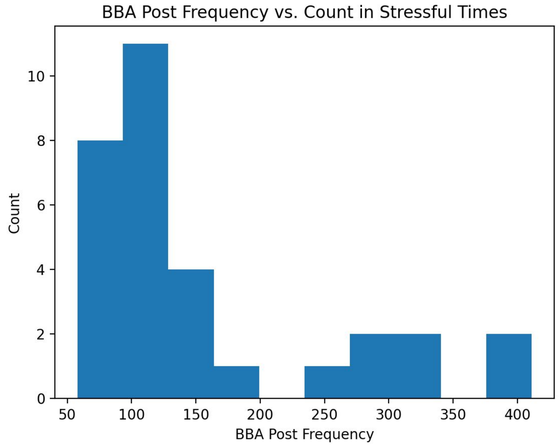
We believe that BBA post frequency and Dear Blueno post sentiment change depending on whether the time period is stressful or not.

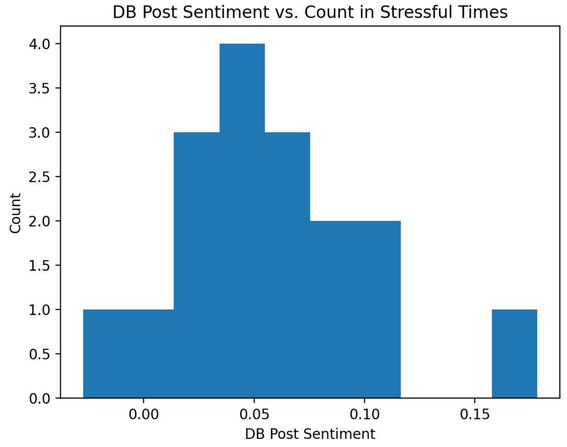
Data and Methods

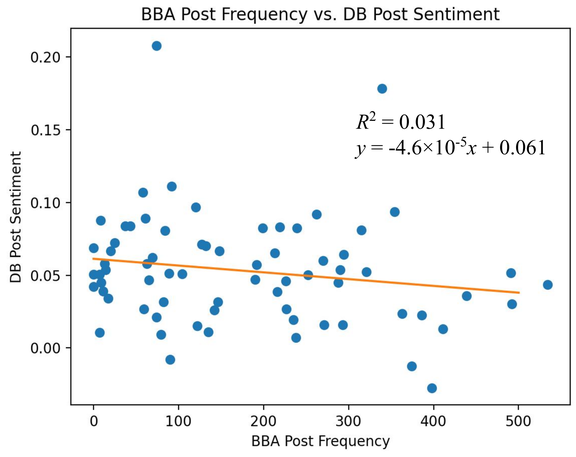
- We obtained exported post data from the BBA and DB moderation teams
- The DB data contains all non-deleted posts created between November 17, 2018 (when DB was created) and March 19, 2020 (when the post data was exported)
- The BBA data contains all non-deleted posts created between September 9, 2018 (when BBA was created) and February 2, 2021 (when the post data was exported)
- The summer months were excluded because of low engagement
- We used the VADER module in Python's NLTK library to classify post sentiment. For each post, we break it into sentences, score each sentence, and obtain the average. We then take the average sentiment over all posts in a time period in this way.
- We started weeks at midnight Eastern Time.

Acknowledgments

We thank the Dear Blueno and BBA teams for their tireless moderation work and for providing us the raw post data we needed to perform this analysis.







Conclusions

The R^2 value of 0.031 for the simple linear regression gives us a higher confidence that we picked the right stress periods..

With a p-value of 0.66 for the BBA data, and a p-value of 0.37 for the DB data, we can conclude that stress periods do **not** affect either BBA post frequency or DB post sentiment.

Analysis

- We were thinking of using a *t*-test, but we realized our data was not normal, so a *t*-test would not work for us (see histograms to the left).
- We decided to use Mann-Whitney *U* test.
- Its a non-parametric test, meaning it does not depend on the distribution of the underlying data.
- We used simple linear regression as a sort of check on the validity of our stress periods.
- Hypothetically, there could still be a relation between stress periods and our data, but we just happened to pick the wrong weeks for stress.
- Using linear regression on the BBA post frequency vs. DB post sentiment allows us to do a sanity check on the stress periods.
- If there is a correlation between the two, then it would indicate that they are related temporally, and our stress periods are off.
- However, since linear regression produces an R^2 value of 0.031, we can say with confidence that they are not correlated, giving us higher confidence in our stress periods