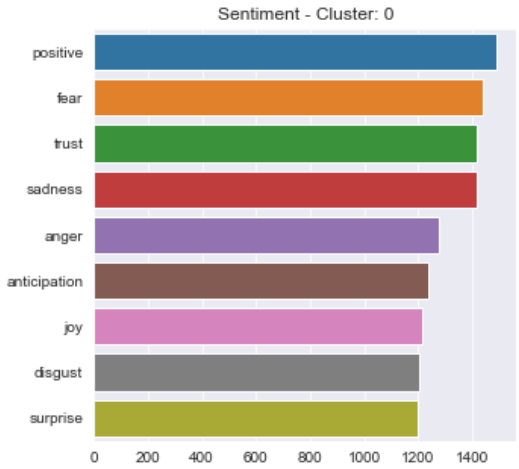
Analysis of Clusters - OPTICS 1

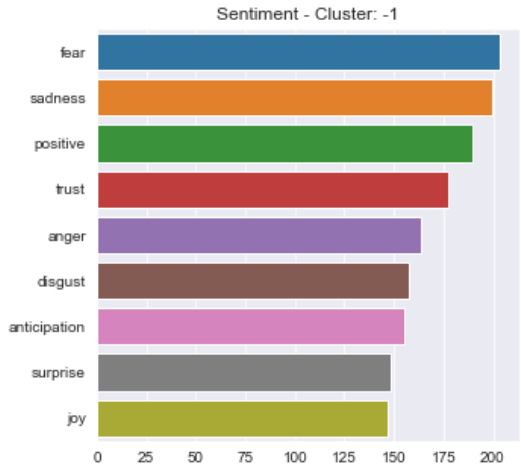
October 20, 2020

1 Analysis of Clusters - OPTICS 1

This run of OPTICS used default sklearn parameters with a max epsilon of 0.1 and min samples per cluster of 200. The algorithm used the cosine dissimilarity between document vectors as distance metric.Two clusters were produced, one relatively large and one relatively small. The growth of these wasnt tested with addition of more data.

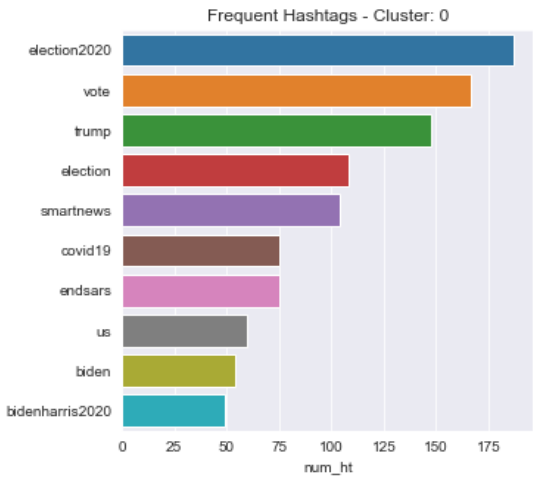
2 Cluster Sentiment

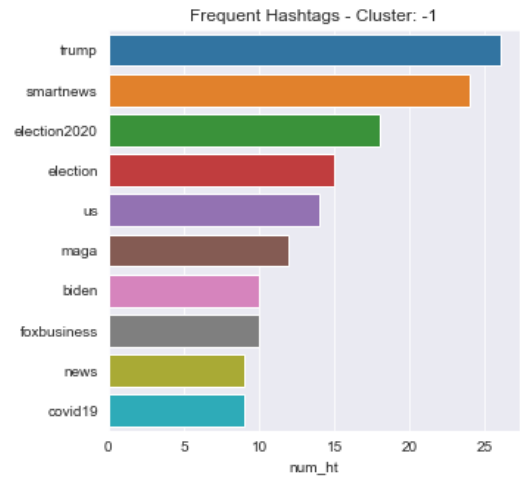




This feels slightly more telling than the HDBSCAN sentiments, although still it's challenging to draw conclusions with the differences in volumes.

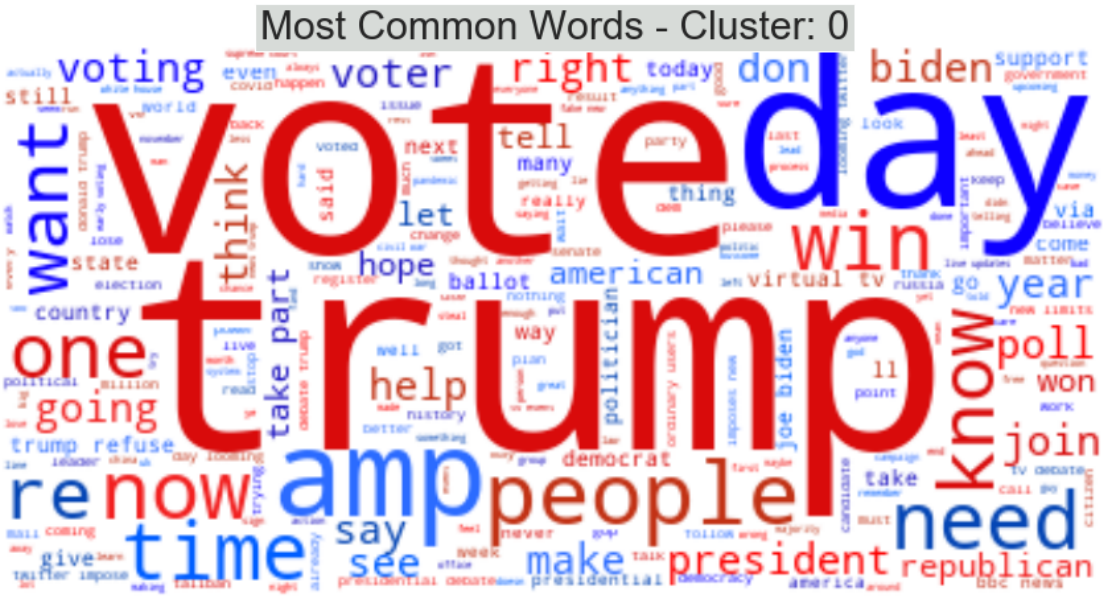
3 Cluster Hashtags

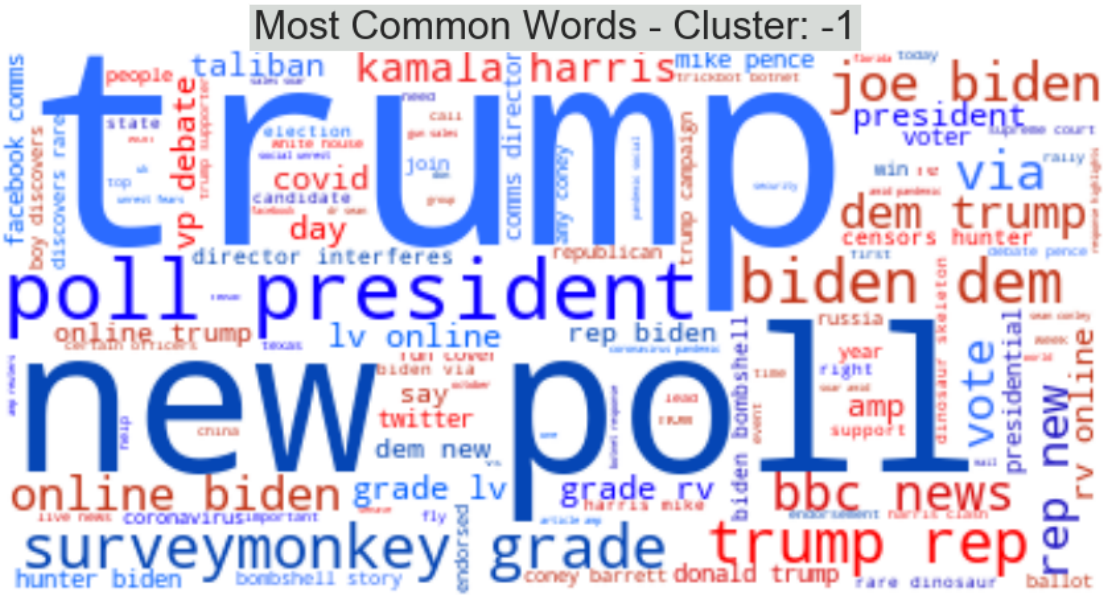




Cluster 0 doesn't seem to display anything pro-trump beyond 'trump' but cluster -1 does. Similarly for Cluster -1 and Biden. Cluster -1 also seems to feature more news hashtags.

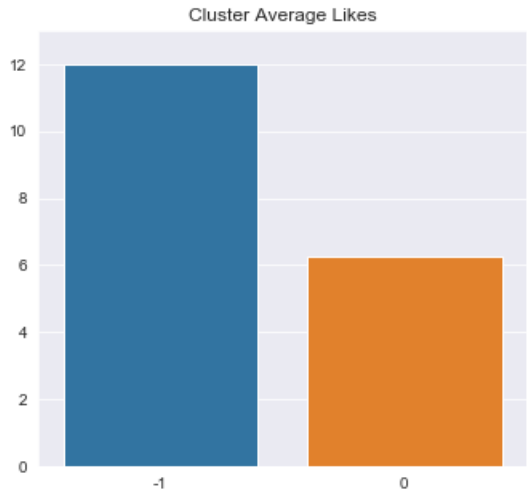
4 Cluster Wordclouds



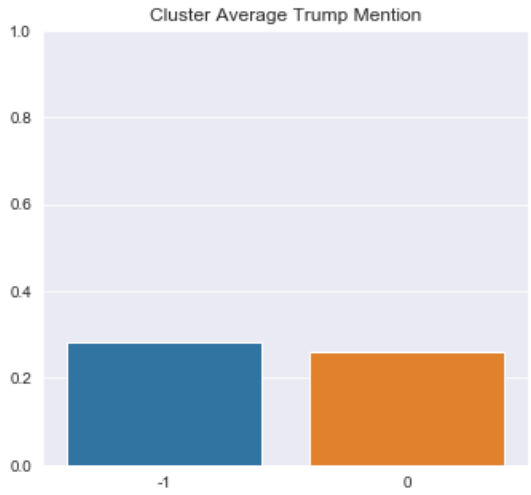


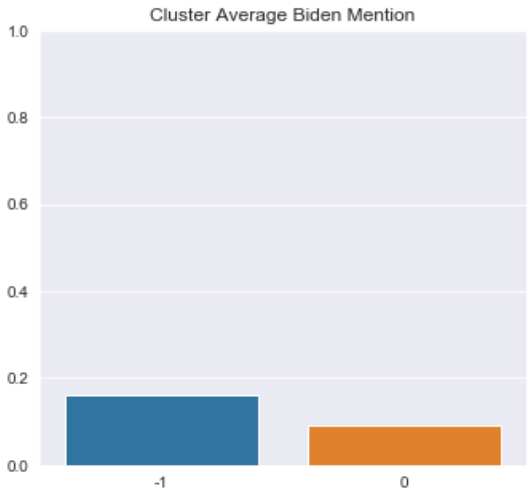
It definitely feels as though Cluster -1 has captured more news articles.

5 Cluster Likes and Retweets

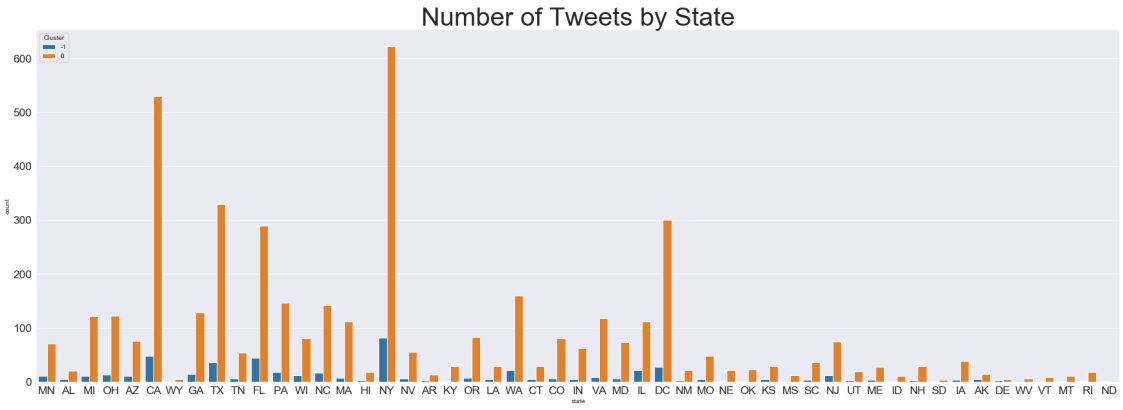


6 Cluster Candidate Mentions





7 Cluster Locations



8 Overall Conclusions

There were noticeable differences but not enough to conclude anything with such a difference in cluster sizes. While appearing slightly better than HDBSCAN, the clustering with Optics still did not prove very effective.