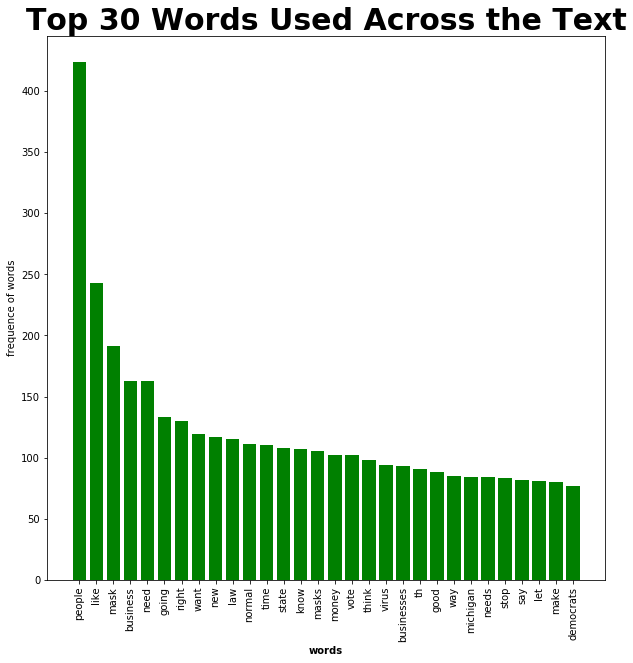
**Changing Minds and Saving Lives Summary**

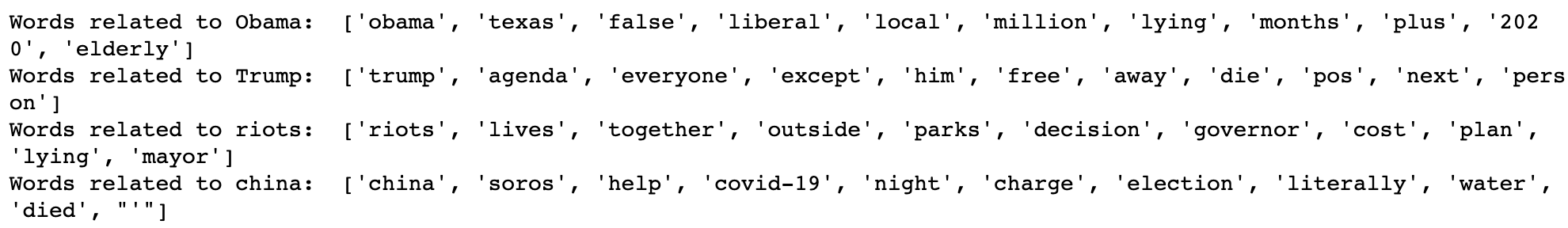
We were inspired to investigate the ReOpen USA protests happening in response to the quarantine lockdown restrictions. On motivation from Professor Jeremy Menchik at Boston University we started our project to study the language being used on these protest Facebook groups. We focused on one specific group: <https://www.facebook.com/groups/reopenusa2020/>

Our first challenge was in scraping the data, after exploring various pre-built Facebook scrapers we decided to build our custom one using Selenium Webdriver and Beautiful soup to scrape the data. We scraped the post text, all comments under the post, the permalink address to the post, and the timestamp. Next we cleaned our text by combining the comments and posts into one body of text in a .pkl file format, then we cleaned stop words, lemmatized the words, and tokenized them as appropriate for further analysis. Our analysis centered around the text of the comments and posts. We employed a range of tools including Word2Vec, LDA topic modelling, and Sentiment Analysis for positive or negative sentiment. Each had its own benefit and contribution in extracting intent and context for the body of text. Shown below are two representations of the most frequently used words:

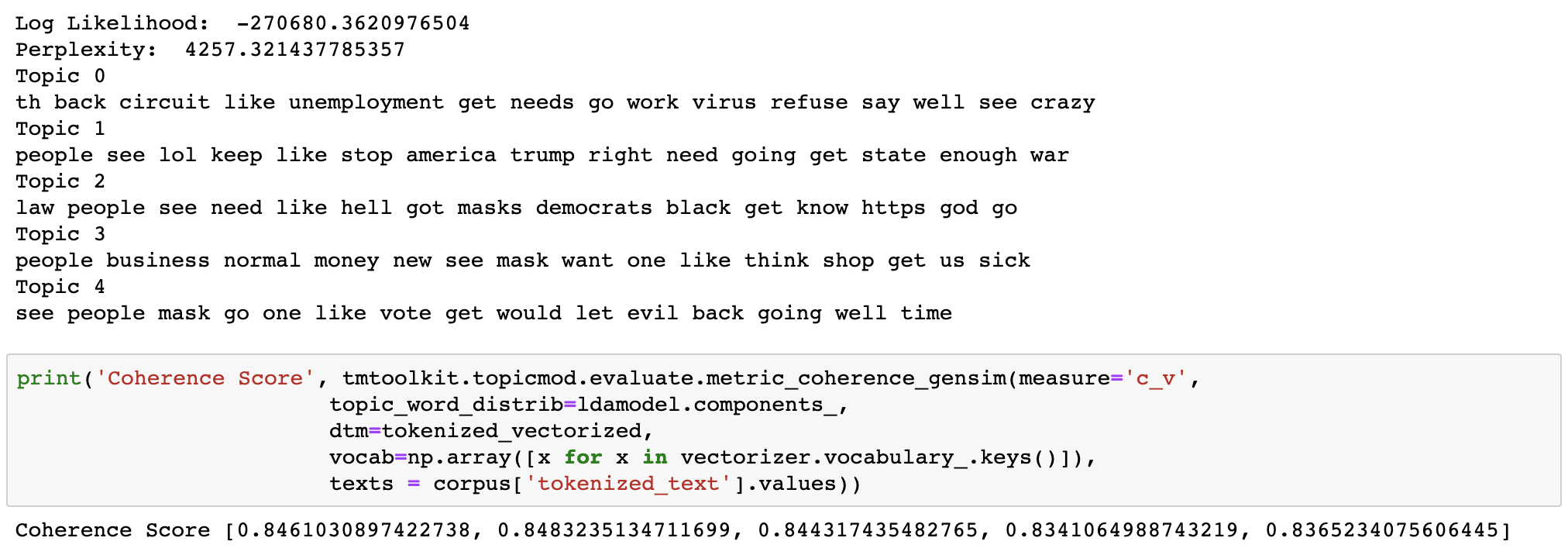




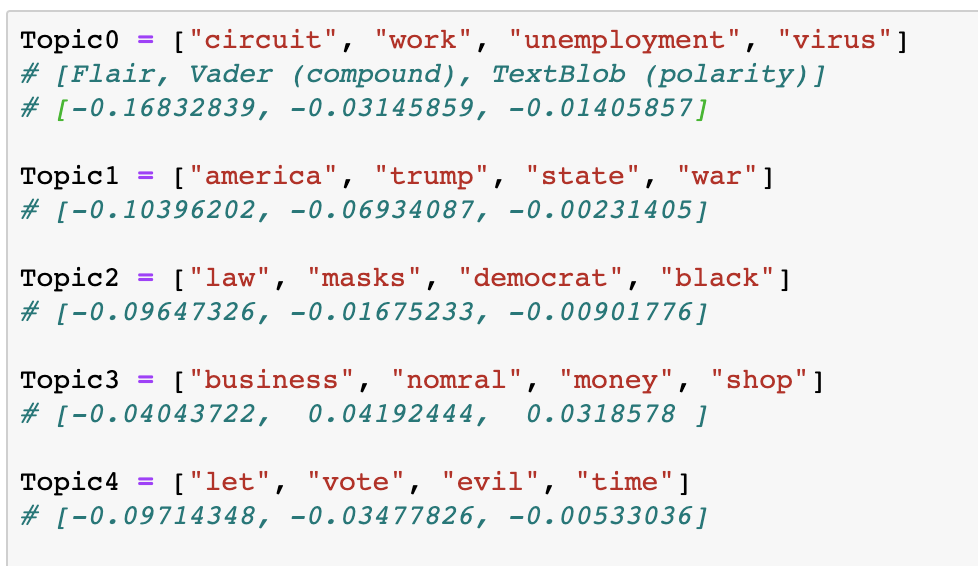
Word2Vec helped in showing us the context that some keywords were used, we primarily focused around Obama, Trump, riots, and China, where we expected Obama, China and riots to primarily have negative connotations and the word Trump to have positive connotations, this was largely confirmed in the Word2Vec model shown.



Next we employed topic modelling using Latent Dirichlet allocation to extract unique topics and get a bird’s eye view understanding of the topics occurring in the body of text. We tried the LDA method found in the Gensim and Sklearn libraries but ultimately found that the Sklearn model with 5 topics was the best due to having a high coherence score (model performance) and the least amount of clustering of topic words, showing that the topics were unique to one another. The results for the LDA topic model we settled on is shown below:



From the topic modelling we wanted to further investigate the sentiment behind each of these words occurring in the topics, so we decided to find the sentiment associated with each of the topics. We decided to use three models: Flair, Vader, and TextBlob and find the average sentiment score for each topic. We did this by finding all the sentences that each of the words in those topics were occurring in and finding the average sentiment of all those sentences. By capturing the sentiment for the sentence where the word was occurring we captured the context for the positive or negative sentiment. Out of the topics shown we further picked some words in it that were significant (i.e. not filler words) this could have been avoided with a better data cleaning approach, but we were pressed for time. We’ve shown the findings below for the following topics and their sentiment under each model.



Topic 0 refers to the virus, unemployment, work, and circuit so we can assume that it refers to some form of employment and jobs, this has a negative sentiment mostly driven by the frustration surround jobs. Topic 1 seems to center around the National political scenario and Trump, we also find a somewhat neutral leaning here, possibly empowered by praise for Trump and also by the negative state of the nation. Topic 2 clearly takes aim at Democrats and the rules that they seem to have imposed on this group, we find a surprisingly neutral stance here as we expected this group to be quite negative. Topic 3 is the most positive as it refers to businesses and returning to normal and this seems to be what the group is mostly hoping to achieve, businesses back to normal. Topic 4 seems to refer to voting of sorts and it falls neutral to negative leaning which makes sense because mail-in ballots have been viewed suspiciously in the right-wing media sphere.

Overall we mostly found our hypotheses to be true and found what we expected, especially regarding the topics and the sentiment they were returning. Given more time we would have liked to explore the topic modelling in more depth and also explore the posting patterns of users.

**Ethical Disclaimer**

We understand that there is a concern for these users’ privacy, we respect everyone’s right to their data and private information. However, this is a public Facebook group and hence a public forum where people are sharing and posting content, therefore we are only curating their data and opinions for academic purposes and not collecting any personal information or revealing anyone’s identity/personal information.