A COMPUTATIONAL ANALYSIS OF THE BUSINESS MEDIA'S RESPONSE TO COVID-19

Caryn Biebuyck (20715978) & Lauren Downing (19202113)

Introduction

The severity of the coronavirus pandemic has arguably been reflected in the reporting of various media agencies in South Africa. Accordingly, this report intends to deliver a holistic computational analysis of the response by various business-oriented media agencies, over the period of the first 100 days since the first confirmed case in South Africa (from 5 March 2020 to 13 June 2020). To achieve this, various methods of text analysis were applied to the text of the articles published during the allocated period.

After researching the available alternatives, the following five media agencies were chosen as the subject of this study, based on being prominently positioned in the industry, mostly subscription-free, and technologically-accommodating. These agencies are Fin24, Moneyweb, BizNews, EWN Business and BusinessTech.

Using various R libraries, the articles of each media outlet were able to be scraped, cleaned and formatted so that mechanisms of numeric and semantic analyses could be applied for each and across all the articles. Thereafter, using structural topic modelling (STM) and Latent Dirichlet allocation (LDA) methods, topics were able to be extracted from the various texts and linked back to the sentiments of the articles. In doing so, a gauge of the media's response to the spread of the coronavirus can be achieved.

A GitHub repository houses all the relevant R Notebooks, data, and image files, and can be accessed at https://github.com/laurenldowning/DataAnalysisAssignment2.

Data Collection

By way of the `RSelenium` and `rvest` libraries, the text, title and date of the articles of each media outlet for the allocated period (5 March 2020 to 13 June 2020) were iteratively scraped from the various web pages of the chosen media houses, and thereafter stored in data frame structures. Articles were either explicitly chosen from the web page's coronavirus section or implicitly chosen by filtering for the following key words in the article's title, namely "active case", "alcohol ban", "cigarette ban", "coronavirus", "covid-19", "epidemic", "infection", "lockdown", "pandemic", "self-isolation", "solidarity fund", "vaccine", and "virus". The process of which is documented in the file '1_data_collection.Rmd'.

Data Cleaning and Pre-Processing

The text that is scraped is often filled with non-textual artefacts, such as escape characters, numbers, punctuation, and white spaces. As such, the text was cleaned and wrangled into the correct, one-word-per-row format as to comply with the prerequisites of the analysis procedures. Wrangling included tokenisation and the removal of stop-words and senseless abbreviations from the text. At this stage, six data frames were created, one for each media house and one for the combined works. As it stood, the tidied Fin24 data structure contained, 369 documents with 56,036 words; the Moneyweb data structure contained 420 documents with 80,657 words; the BizNews data structure contained 242 documents with 82,061 words, the EWN Business data structure with 1,462 documents and 225,374 words, the BusinessTech data structure with 559 documents and 79,852 words, and lastly, the all-encompassing compiled media data structure with 3,052 documents and 523,980 words. The cleaning and pre-processing procedure can be found in file '2_data_tidying.Rmd'.

Text Analysis

With the data appropriately cleaned and pre-processed, each data frame was analysed for three things, namely quantitative numeric metrics, sentiment, and topics. These analyses can be followed in the corresponding notebooks.

Quantitative Numeric Analysis

Each media house was subject to a word counting function to determine the 25 most frequented words across all articles. This was also applied across all media houses. These statistics were visualised using both bar charts and word-cloud visualisation techniques.

Next, using the most populated data frame as a basis, each media house was compared to the standard, namely EWN Business, to determine the correlation between the words used in the sample of articles.

Thereafter, the term frequency-inverse document frequency (TF-IDF) was determined for each and across all media outlets to clarify the context of work usage. By determining the TF-IDF, one is determining the most frequented words (term frequency, i.e. TF), relative to their use in other documents (inverse document frequency, i.e. IDF). Put simply, the higher the TF-IDF score, the rarer the term, and vice versa.

Finally, correlating pairs of words were identified and counted for each and across media outlets using the `widyr` package. The progress of the above can be traced in the file '3_

quantitative_numeric.Rmd'

Sentiment Analysis

To determine the sentiment of the media outlet articles, the terms of each document were compared to the terms of the Loughran-McDonald sentiment lexicon. This lexicon was chosen due to it being financially- and economically-oriented which made it a good match considering the financial nature of the documents requiring analysis. Overall sentiment of the media outlets was determined and visualised, followed by individual analyses of the top contributing words towards each sentiment category for each media outlet. This was repeated across all media outlets, in addition to calculating and visualising the word-sentiment proportions across all articles and sentiment categories. The progress of the above can be traced in the file '4_sentiment_analysis.Rmd'

Topic Modelling

At this stage, the tidied media data structures were subjected to transformation for the purpose of topic modelling. The data structured were firstly cast into Document-Term Matrices so that Martin Ponweiser's method of using the Harmonic Mean to determine the number of topics to model could be used. Details of how to implement this approach can be found at https://knowledger.rbind.io/post/topic-modeling-using-r/. After finding the appropriate values for the number of topics to consider, two methods of topic modelling were applied. By implementing two methods of topic modelling, a systematic exploration of the results can be considered.

Firstly, prompted by the growing popularity and ease of implementation, structural topic modelling (STM) was applied using both the beta and gamma approach. Next, Latent Dirichlet allocation (LDA) was used given its reputation for returning reasonable and interpretable results, especially in a context where long documents are involved. LDA has the added benefit of being easily traceable to specific documents so that overall topic sentiment can be determined. The beta implementation of LDA weighs the relevance of a particular term, which in this case was a word, to a topic. This was also used as the basis for the topic sentiment modelling. One of the limitations of LDA is that it is a soft clustering method, meaning it is difficult to gauge the topic quality and coherence. The progress of the above can be traced in the file '5_topic_modelling.Rmd'

Results and Discussion

Quantitative Numeric Analysis

Word Counts & Word Clouds

Below are the 25 most frequented words across the sample articles of each media outlet. As seen below, there is an expected commonality between these words as they all pertain to words referring to Covid-19, lockdown, location, timespan, and various business, healthcare, and economic topics.

Through inspection of Figures 1 through 12, it is apparent that Fin24 (Figure 1) and Moneyweb (Figure 3) cover more topics that are economic in nature due to the prevalence of the stemmed *econom* and *busi*, compared to the other media outlets. Moreover, given the stems *govern*, *minist* and *presid*, it is revealed that EWN Business (Figure 7) is more inclined towards politically charged topics.

Figure 11 illustrates the overall word usage across the sample of media outlets. The high usage of the word *lockdown* indicates that the topic is of concern, which is to be expected given the economic impacts and societal reaction of a lockdown situation. It is interesting to observe that the word *covid* is preferred over the synonymic *coronavirus*, and that *south* is heavily used without the expected accompaniment of *africa*.

These frequented words can be intuitively visualised using word clouds, as seen in the figures below each graph.

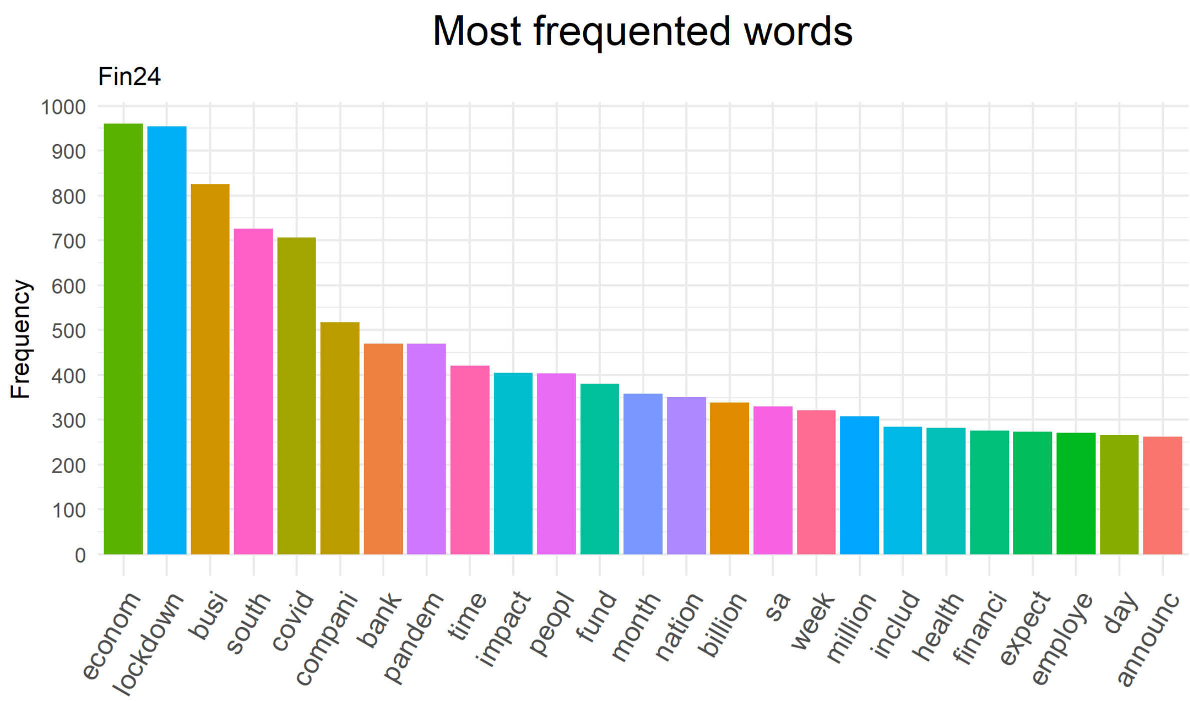

FIGURE 1: MOST FREQUENTED WORDS: FIN24

FIGURE 2: WORD CLOUD: FIN24

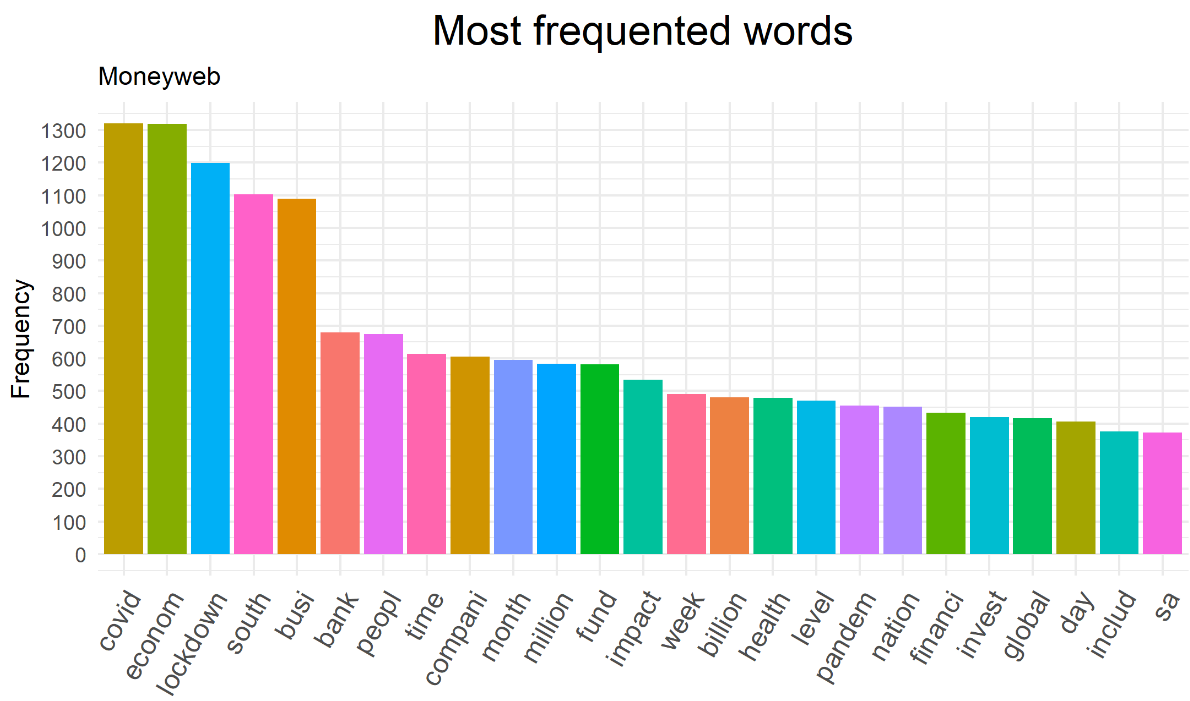

FIGURE 3: MOST FREQUENTED WORDS: MONEYWEB

FIGURE 4: WORD CLOUD: MONEYWEB

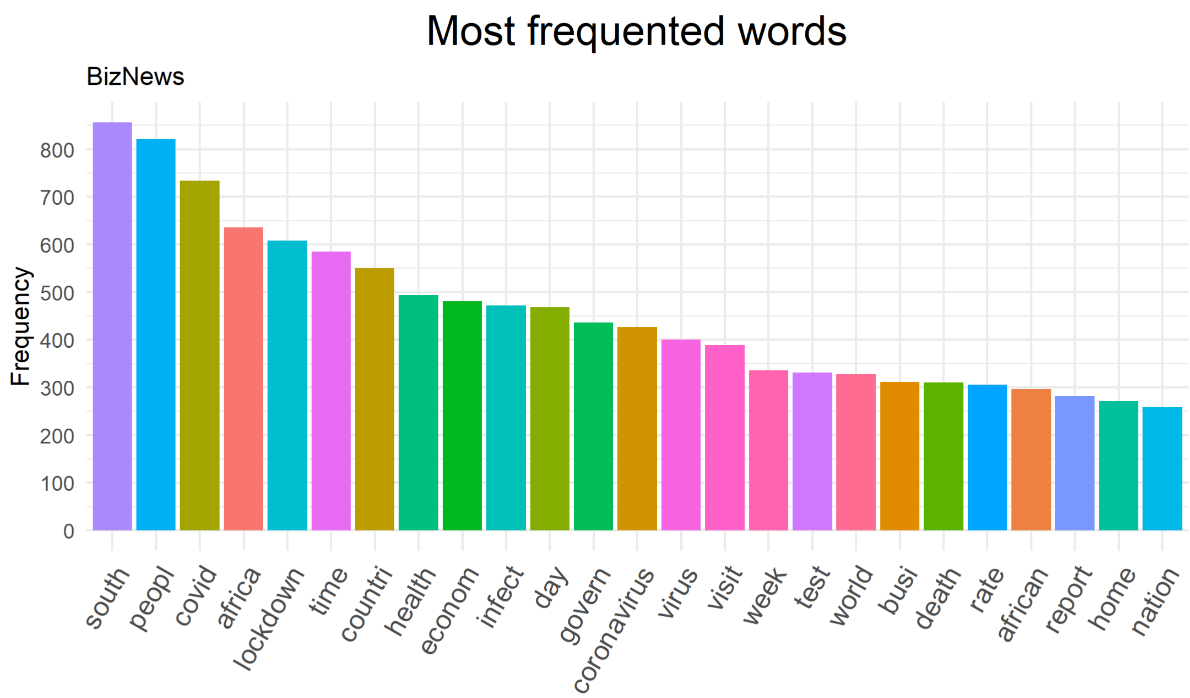

FIGURE 5: MOST FREQUENTED WORDS: BIZNEWS

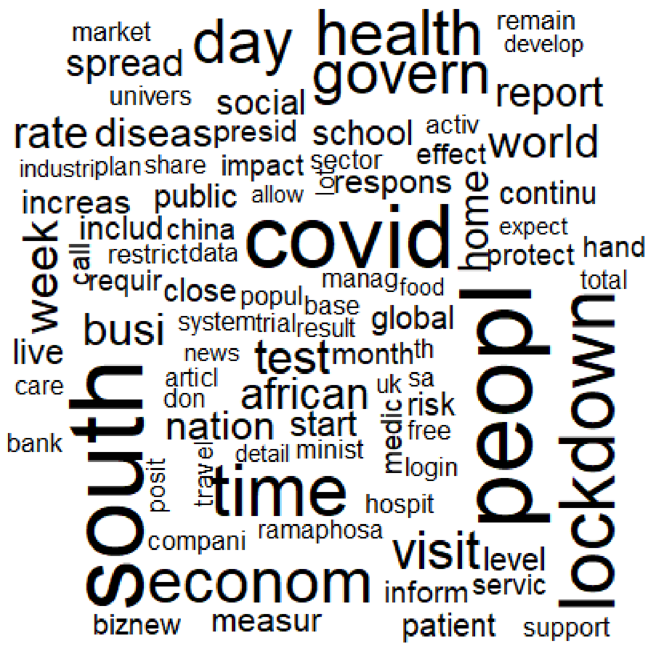

FIGURE 6: WORD CLOUD: BIZNEWS

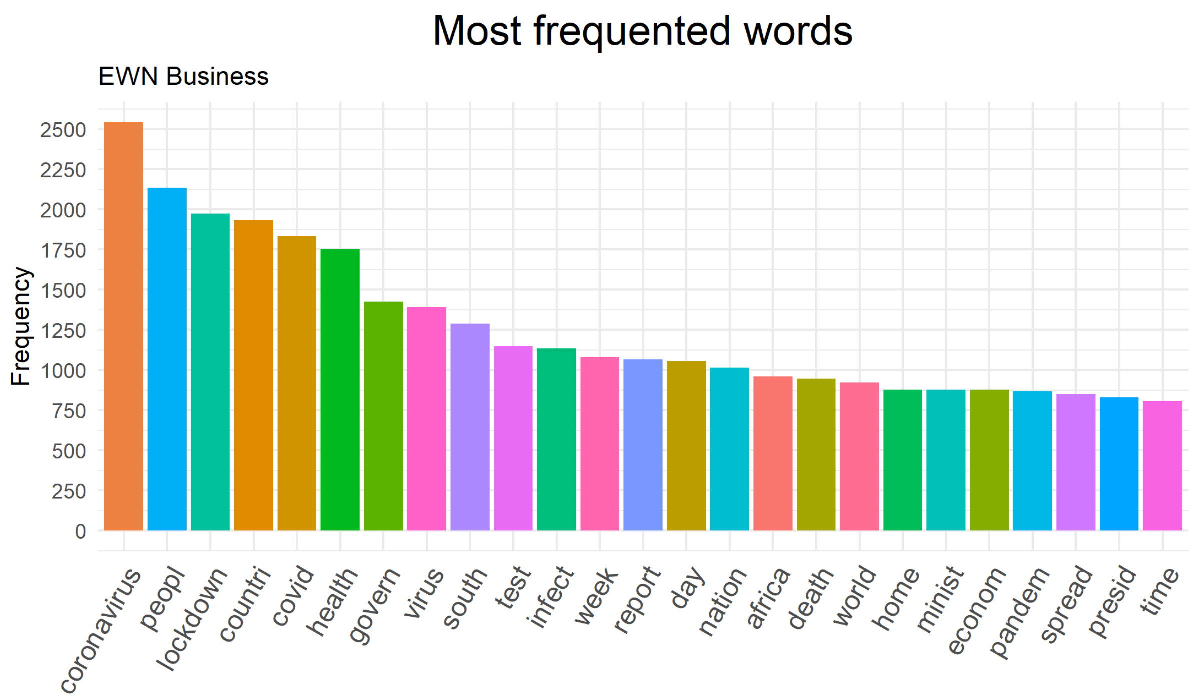

FIGURE 7: MOST FREQUENTED WORDS: EWN BUSINESS

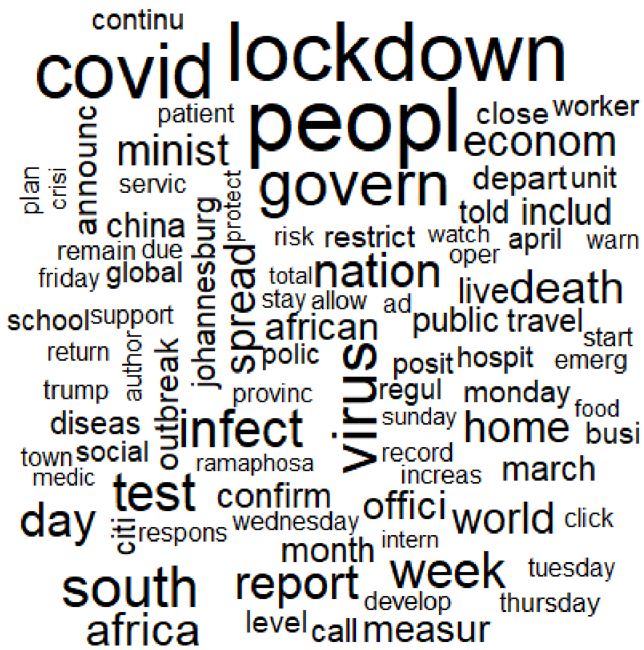

FIGURE 8: WORD CLOUD: EWN BUSINESS

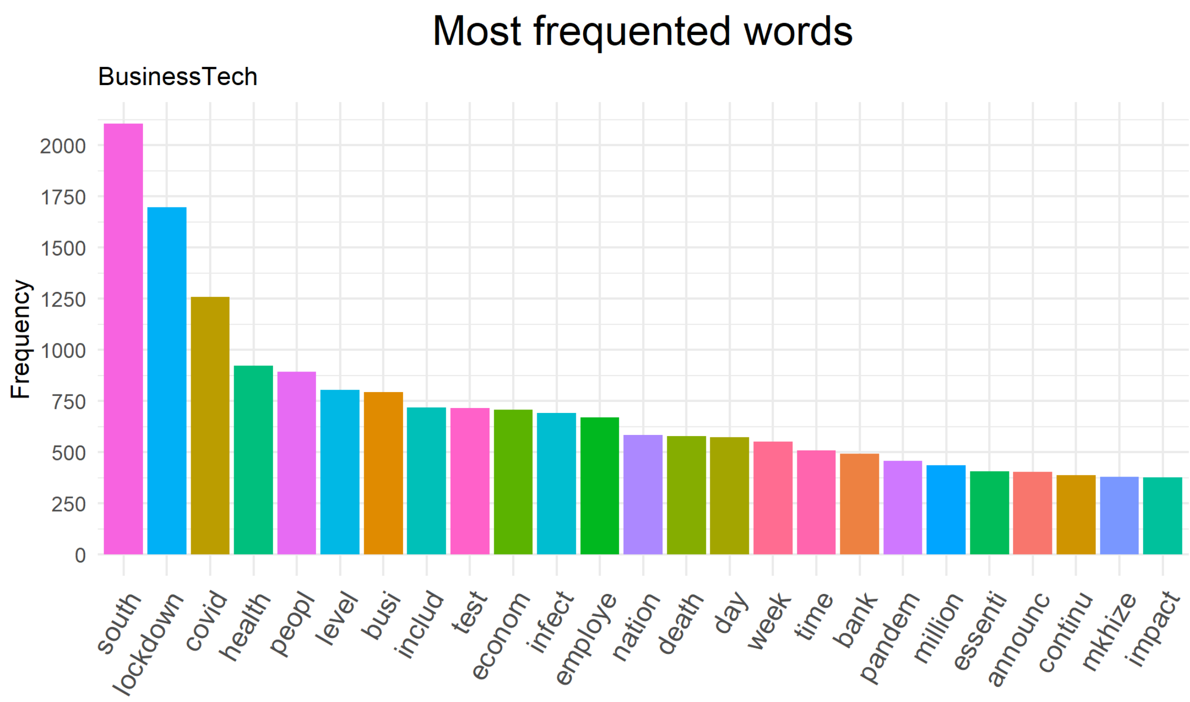

FIGURE 9: MOST FREQUENTED WORDS: BUSINESSTECH

FIGURE 10: WORD CLOUD: BUSINESSTECH

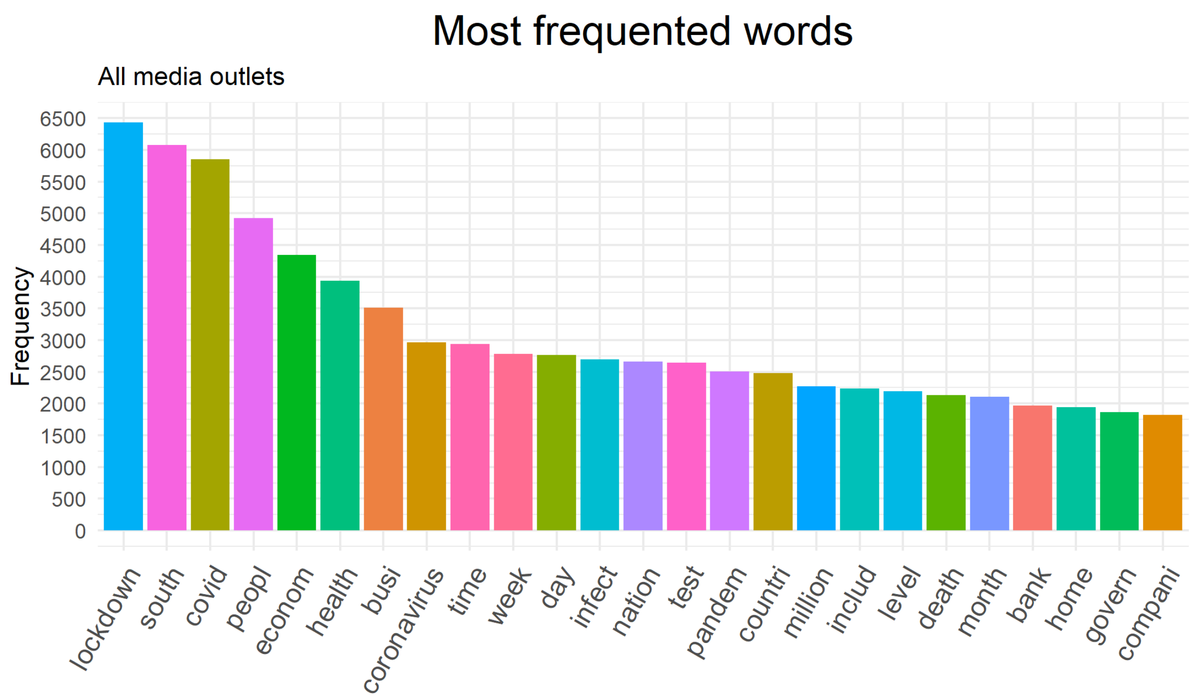

FIGURE 11: MOST FREQUENTED WORDS: ALL MEDIA OUTLETS

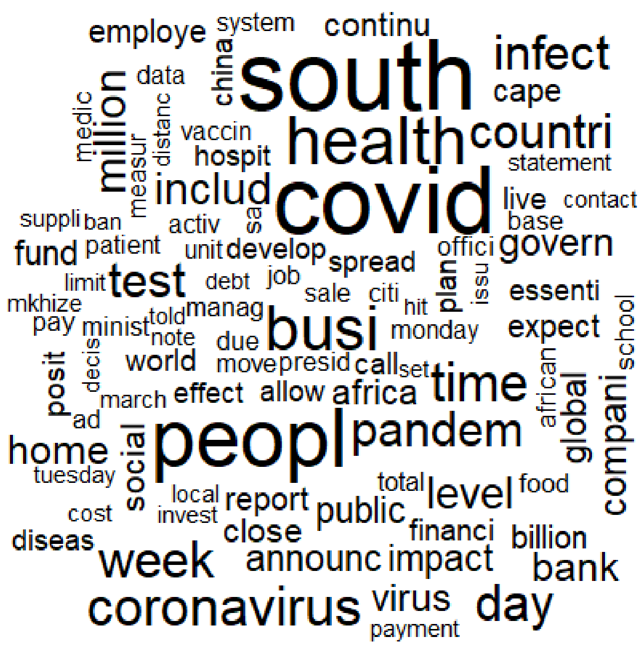

FIGURE 12: WORD CLOUD: ALL MEDIA OUTLETS

Comparing Media Outlets, using EWN Business as a base

Using the complimentary functions `spread` and `gather` from the `tidyr` package, the ratio of words to total words for each media house as compared to EWN Business was able to be gauged, as visualised in Figures 13 to 16 below. Using EWN as a standard for comparison was motivated by the fact that the data frame contained the most documents and it was therefore assumed that this would result in the largest variety of words. Subsequently, using a correlation test of method `pearson`, the similarity and differences in the sets of word frequencies were able to be quantified. When interpreting the visualisations, the further away a word is from the origin, the more reciprocal it is between articles.

Fin24 vs EWN Business

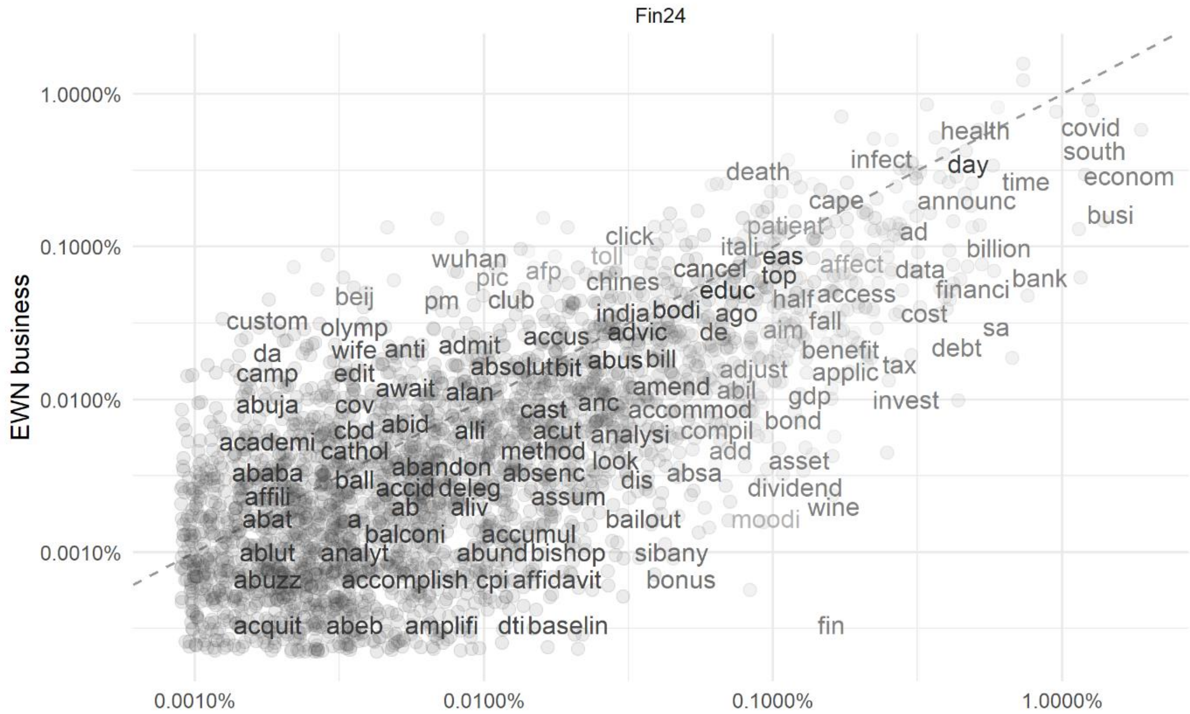

FIGURE 13: WORD CORRELATION: EWN BUSINESS VS FIN24

Using a 95 percent confidence interval, the sample estimates correlation between EWN Business and Fin24 is 0.7499742, which indicates a strong positive linear relationship.

Moneyweb vs EWN Business

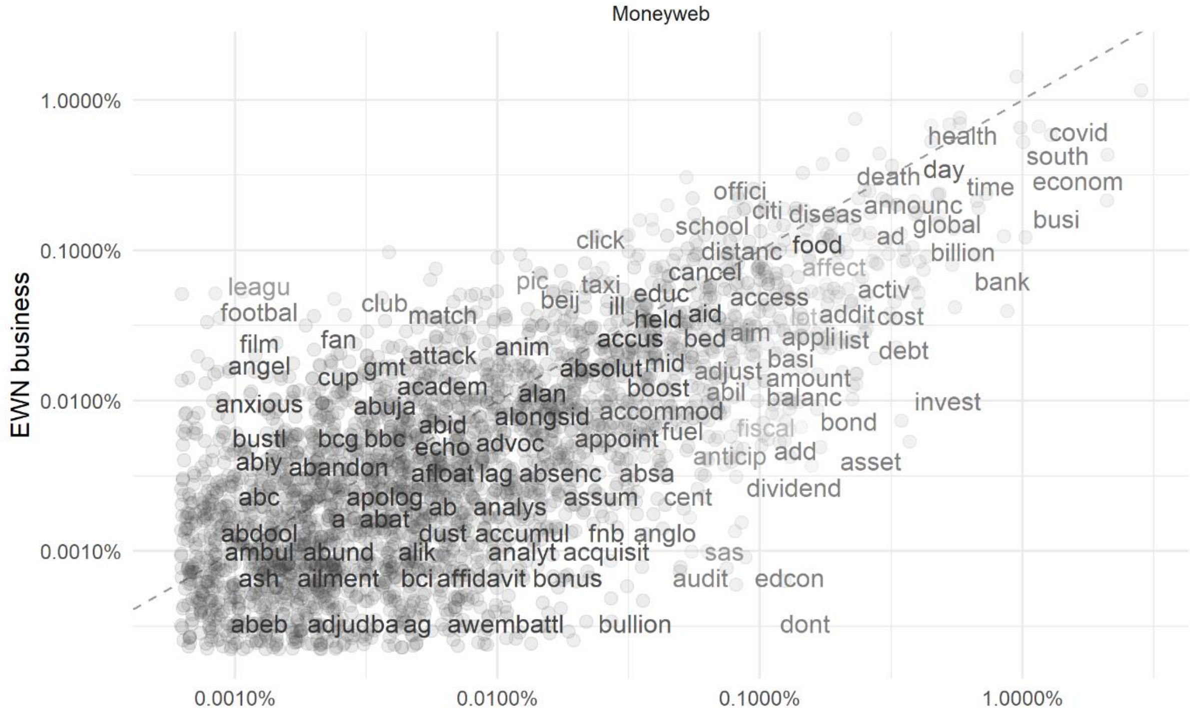

FIGURE 14: WORD CORRELATION: EWN BUSINESS VS MONEYWEB

Using a 95 percent confidence interval, the sample estimates correlation between EWN Business and Moneyweb is 0.777429, which indicates a strong positive linear relationship.

BizNews vs EWN Business

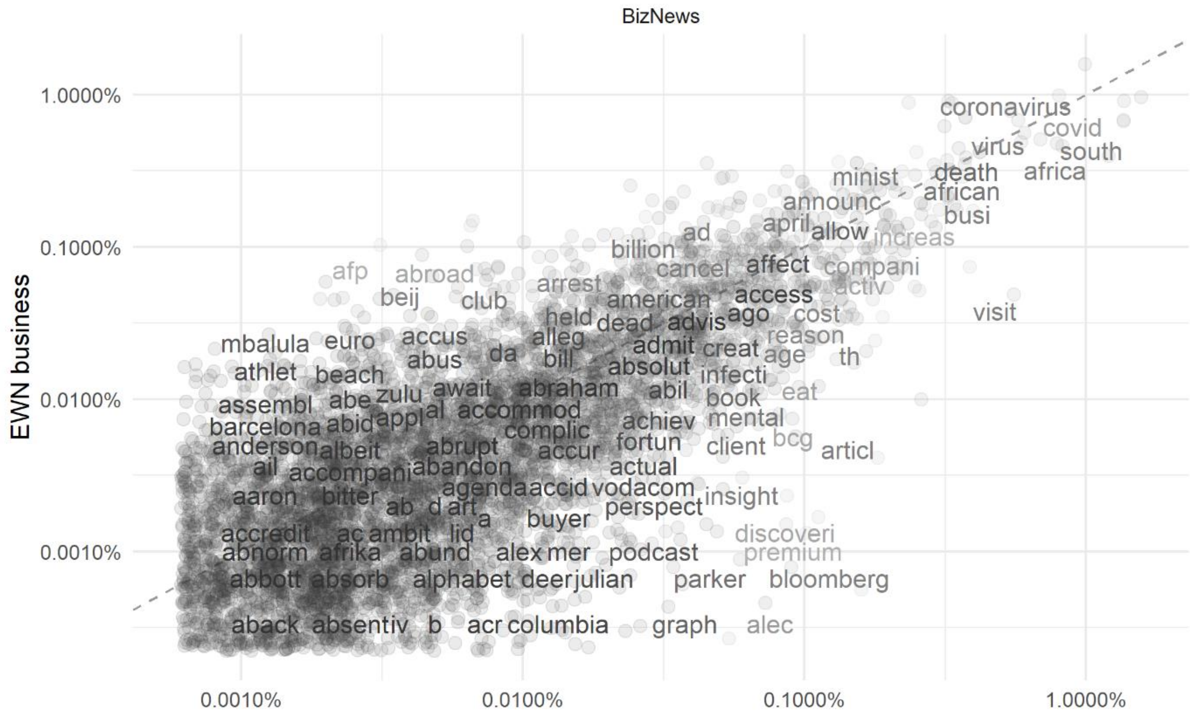

FIGURE 15: WORD CORRELATION: EWN BUSINESS VS BIZNEWS

Using a 95 percent confidence interval, the sample estimates correlation between EWN Business and BizNews is 0.8791541, which indicates a strong positive linear relationship.

BusinessTech vs EWN Business

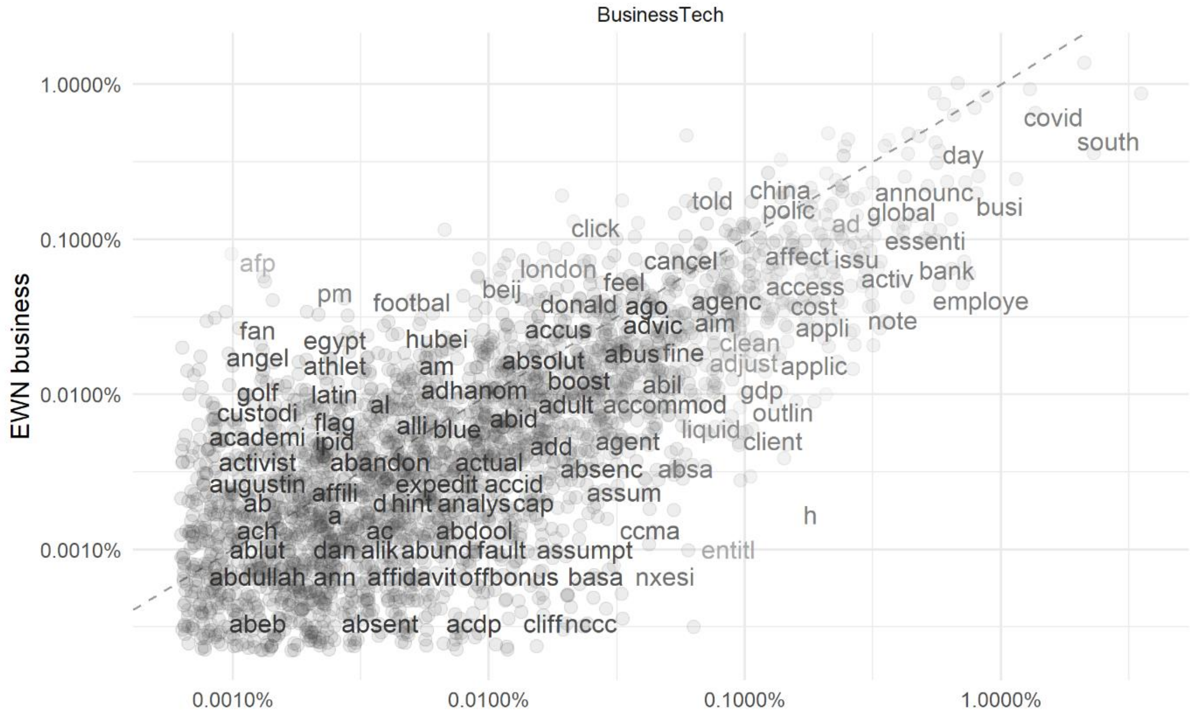

Figure 16: Word correlation: EWN Business vs BusinessTech

Using a 95 percent confidence interval, the sample estimates correlation between EWN Business and BusinessTech is 0.8560497, which indicates a strong positive linear relationship.

Term Frequency Inverse Document Frequency

In Figures 17 to 22 below, the 25 most unique words used by each media outlet in the 100-day period can be seen, with colour indicating the month in which these words were frequented. As anticipated, majority of the featured words are proper nouns or words relating to other news stories of the time which were covered by the particular media outlet. Moreover, there is a prevalence of acronyms, as well as words one would not typically associate with the Covid-19 pandemic, with the exception of *vaccin* Figure 21 and *vaccinolog* Figure 22.

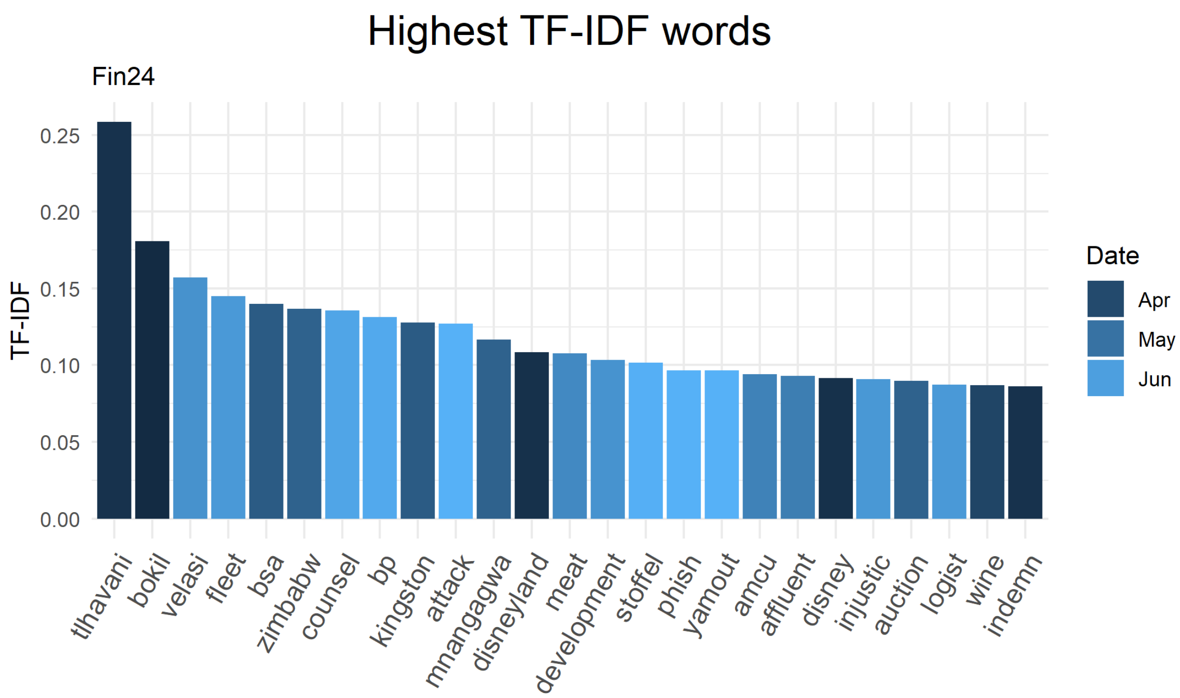

FIGURE 17: TF-IDF: FIN24

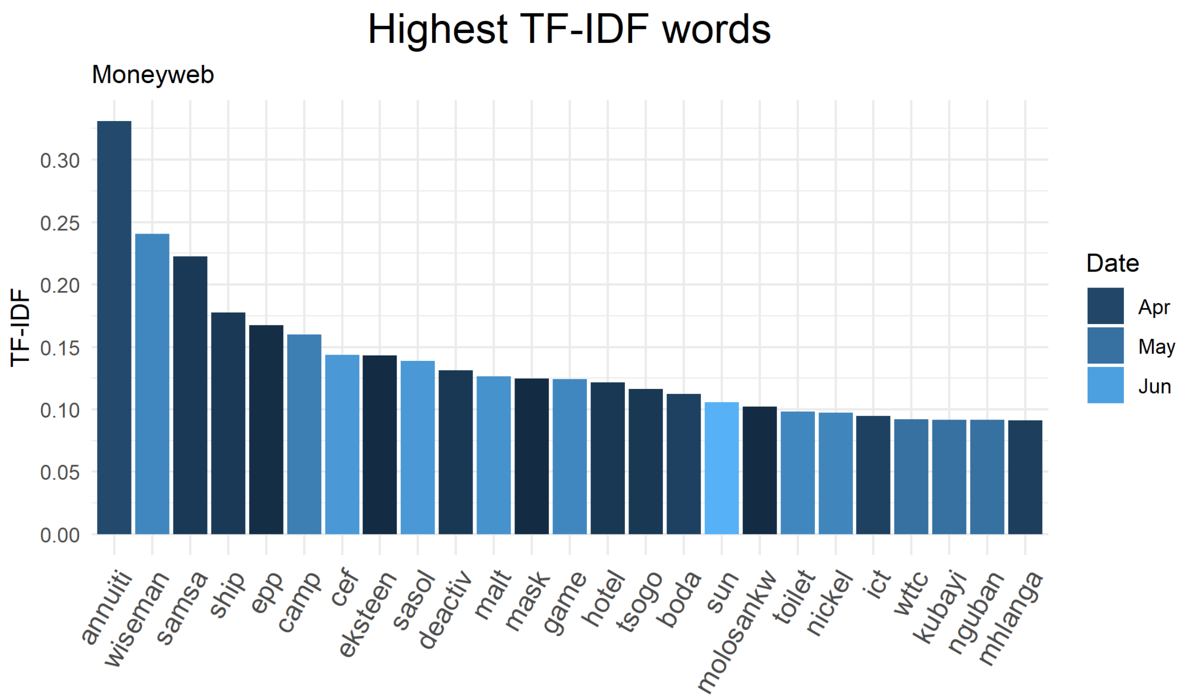

FIGURE 18: TF-IDF: MONEYWEB

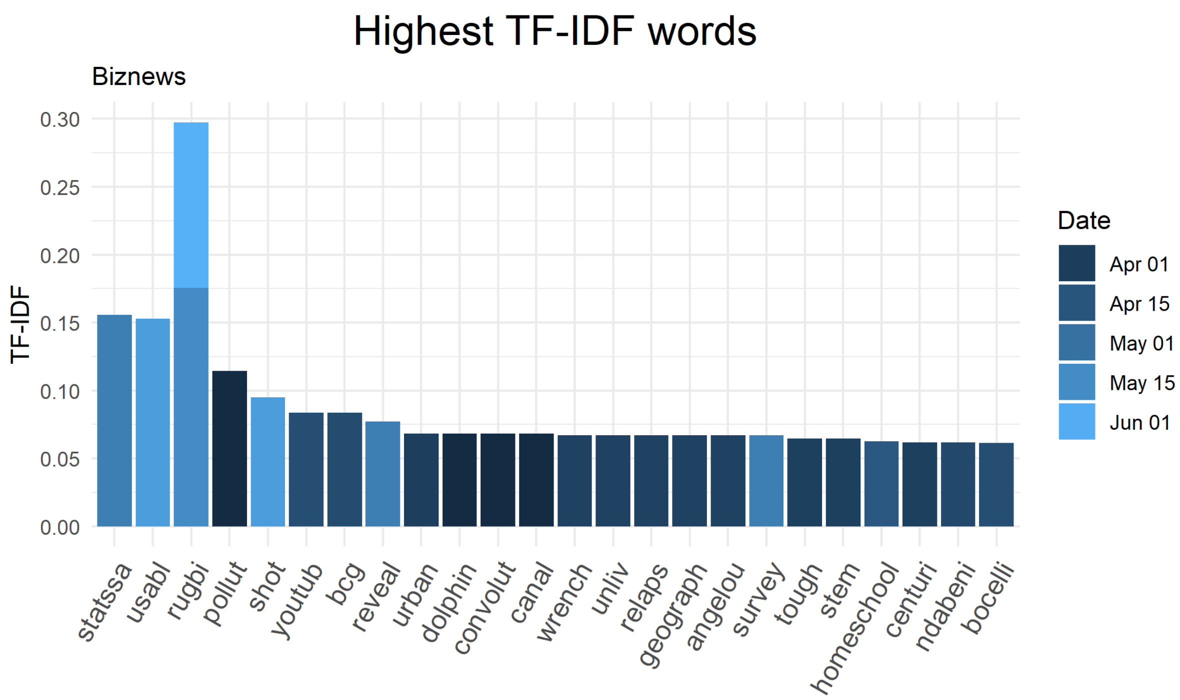

FIGURE 19: TF-IDF: BIZNEWS

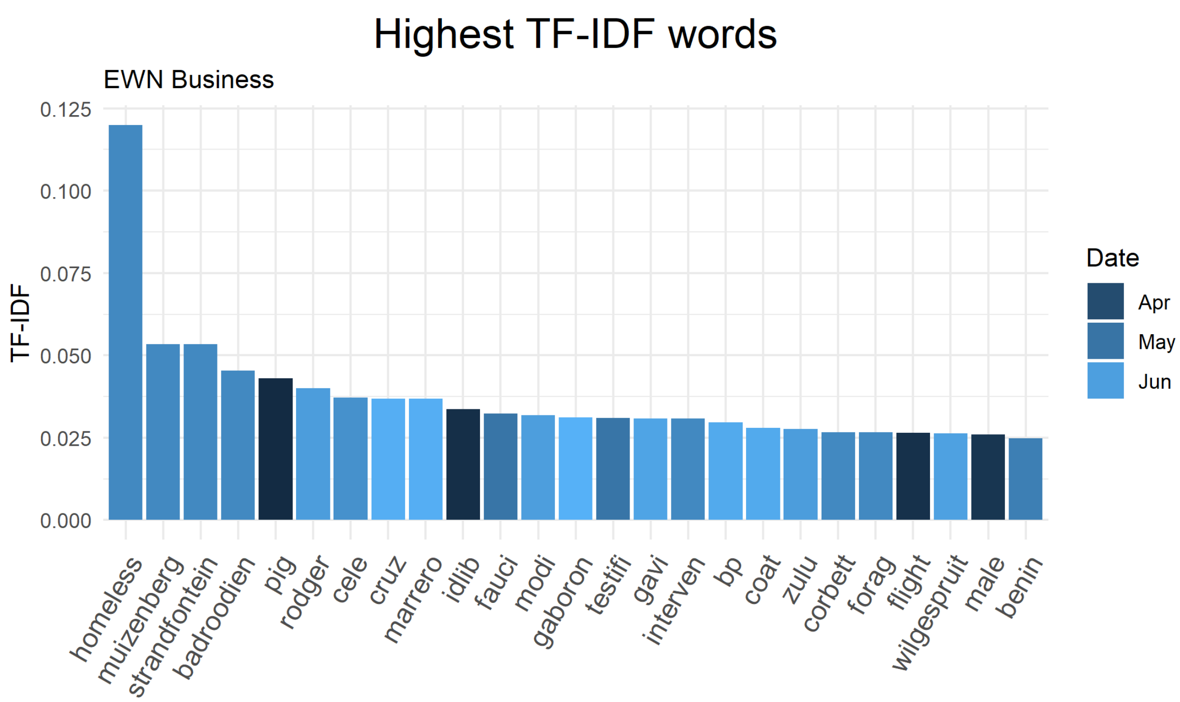

FIGURE 20: TF-IDF: EWN BUSINESS

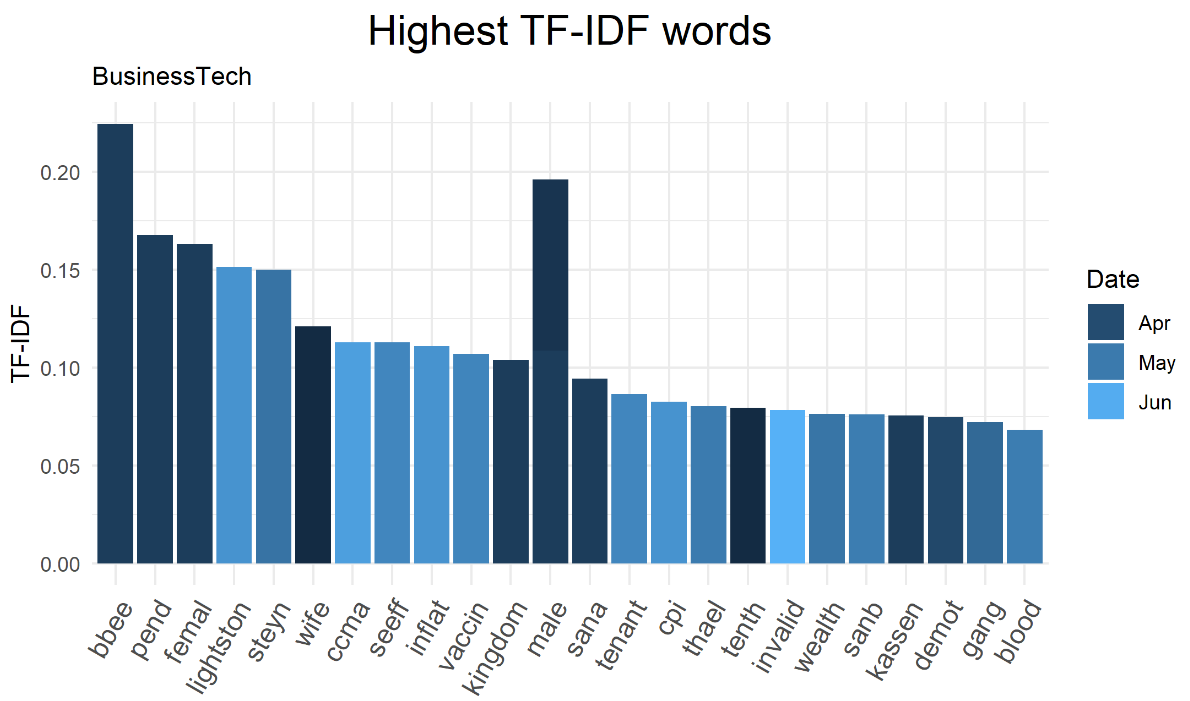

FIGURE 21: TF-IDF: BUSINESSTECH

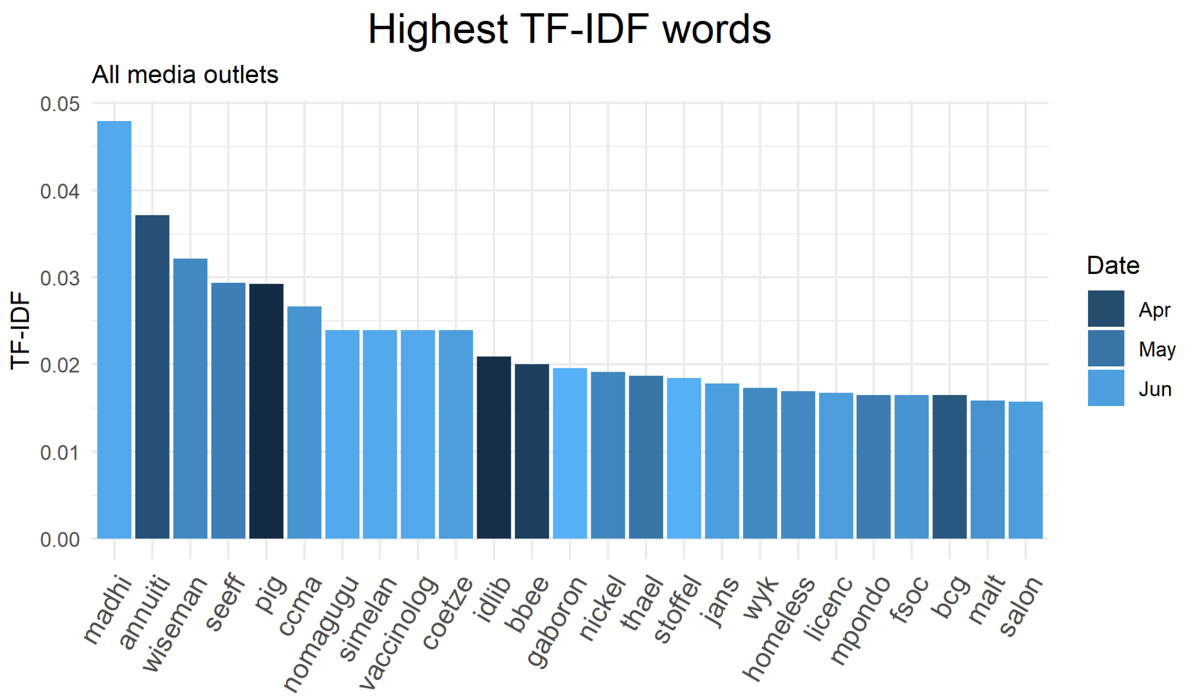

FIGURE 22: TF-IDF: ALL MEDIA OUTLETS

Counting and Correlating Pairs of Words

After filtering for the three most prominent Covid-19 words, namely *ban*, *covid* and *lockdown*, the correlations between these words and the top 10 word-pairings, as identified by the `widyr` package, are visualised below for each and across all media outlets (Figures 13 to 34). These visualisations are followed by `ggraph` visualisations which capture the specific word clusters that form between the top 20 Covid-19 oriented word pairs, with line thickness indicating the strength of the correlation.

Fin24

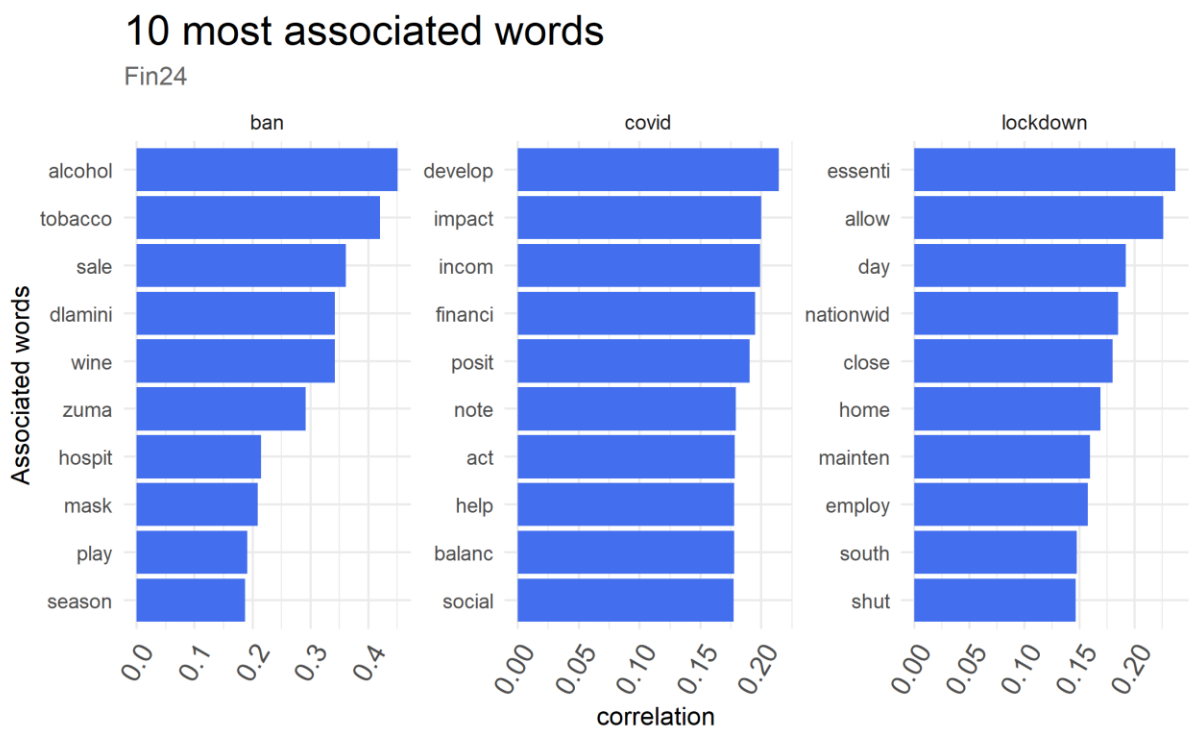

FIGURE 23: WORD PAIRS: FIN24

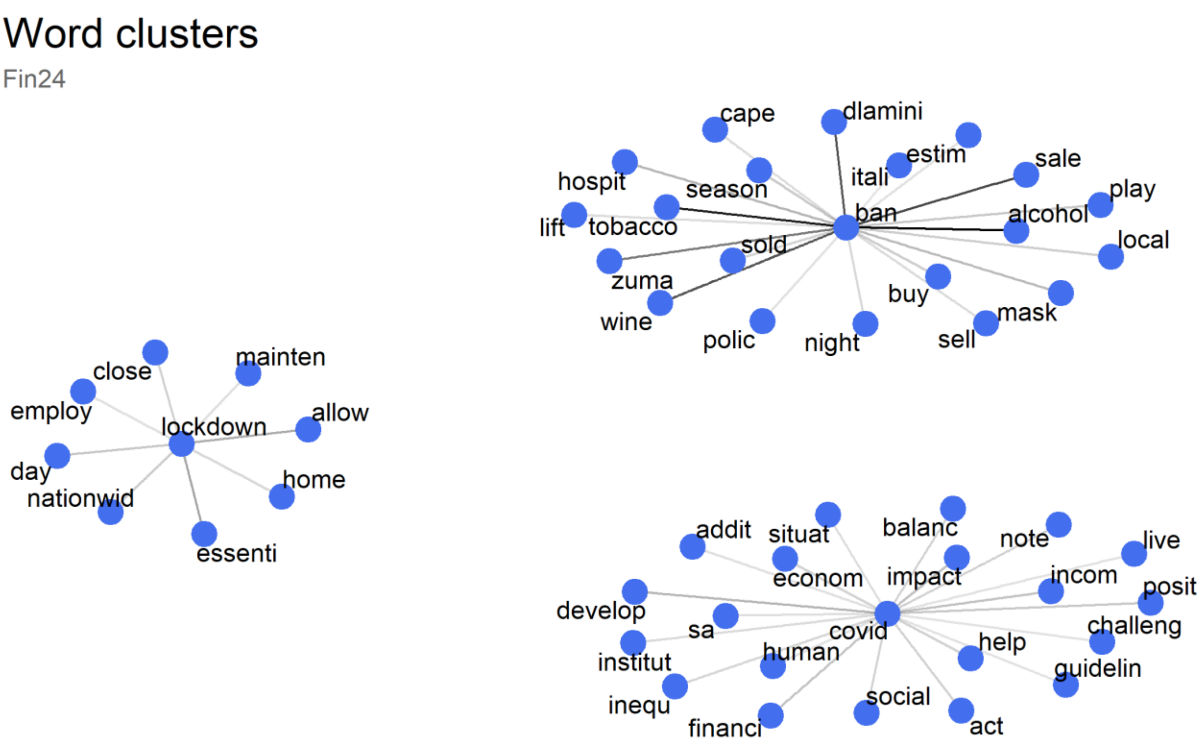

FIGURE 24: WORD CLUSTERS: FIN24

In Figure 23, the Fin24 word-pairings associated with *ban* are strongly correlated to those products and activities which were prohibited under the initial stages of the nationwide lockdown in South Africa, which occurred within the 100 days of data collected. Moreover, there is a strong correlation to the word *dlamini*, referring to Nkosazana Dlamini-Zuma, the minister in charge of the implementation of these bans.

Similarly, the word pairings associated with *covid* and *lockdown* can be correlated to the impact that each of these events has had on both the economy and citizens. Moreover, they are indicative of the priorities of South Africans during the observed period.

Surprisingly, the word clusters displayed in Figure 24 are completely separated. This could be due to the results being restricted to the top 20 results associated with each Covid-19-specified word. Moreover, the words that were initially associated with *lockdown* in Figure 23 have been reallocated to either of the other two specified words. However, this clustering could potentially indicate that Fin24 articles are relatively focused and use topic-specific terminology.

Moneyweb

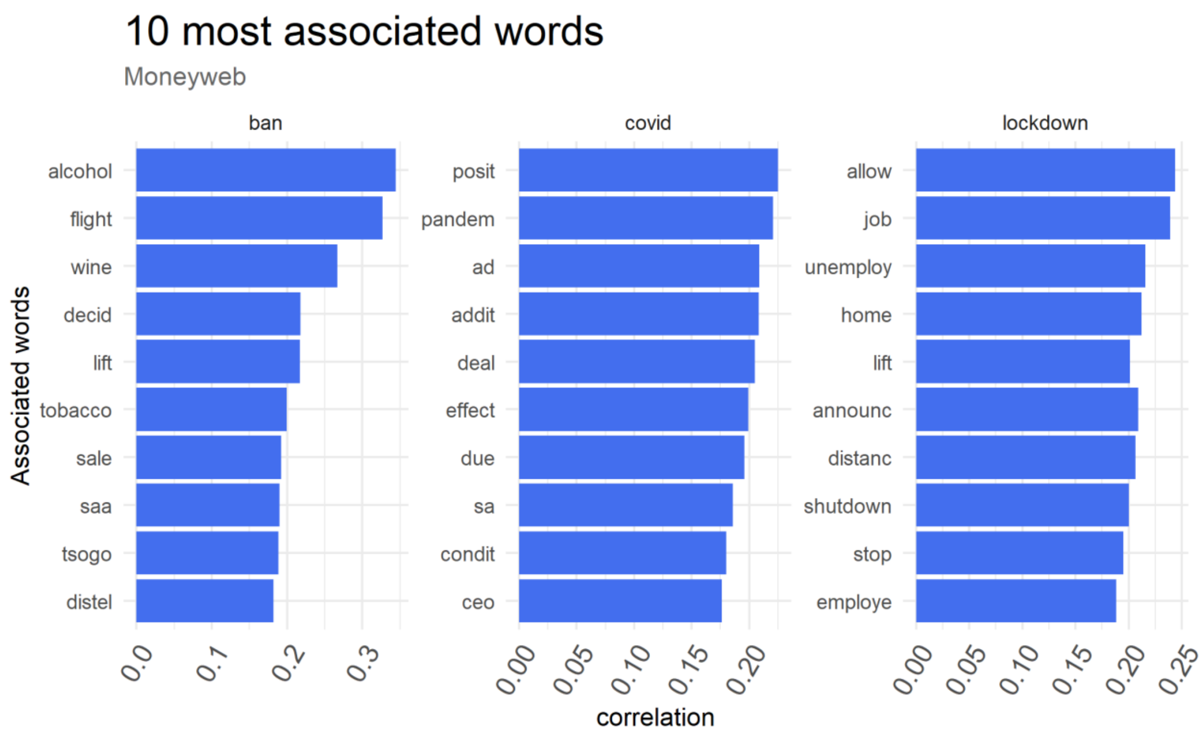

FIGURE 25: WORD PAIRS: MONEYWEB

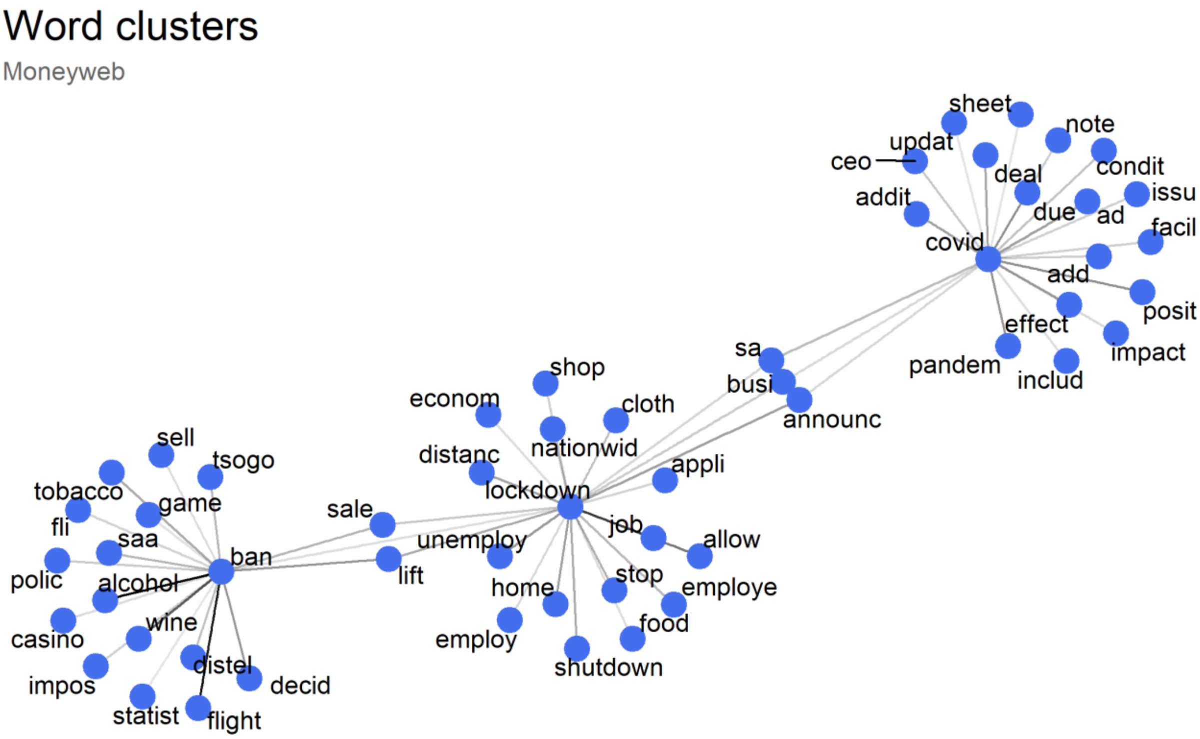

FIGURE 26: WORD CLUSTERS: MONEYWEB

The Moneyweb word-pairings depicted in Figure 25 show similar word associations and inferences to each of the specified words to that of Fin24. However, the presence of the word *lift* indicates that Moneyweb had a larger focus on the reporting of the removal of the bans and the potential dissatisfaction shown by South African citizens.

Although each Covid-19-specified word has unique sets of related words, there are evidently words that connect the three topics with *lockdown* being the intermediate entity, as can be seen in Figure 26.

BizNews

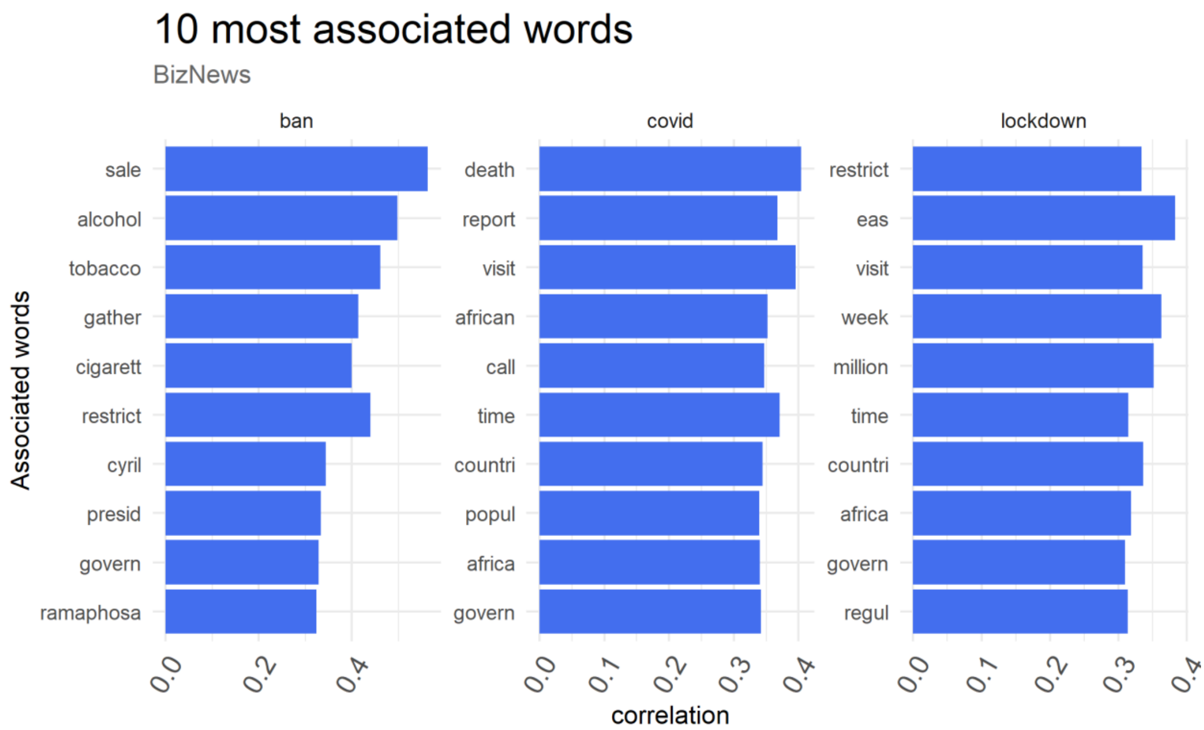

FIGURE 27: WORD PAIRS: BIZNEWS

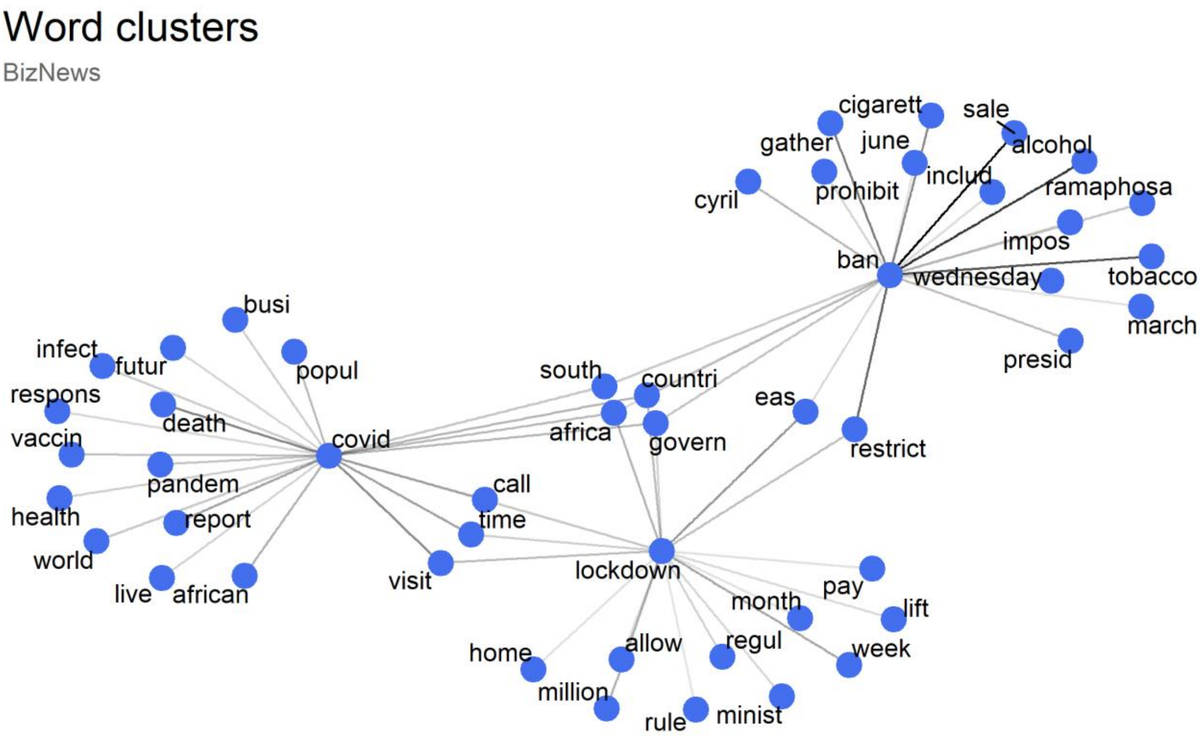

FIGURE 28: WORD CLUSTERS: BIZNEWS

The BizNews word-pairings in Figure 27 suggest a strong political focus of articles that contain the three key words.

In contrast to Moneyweb (Figure 26), the word clusters depicted in Figure 28 identify words that connect all three clusters rather than the clusters being connected by a single common cluster. This could potentially indicate that multiple topics are broached in articles relating to the coronavirus.

EWN Business

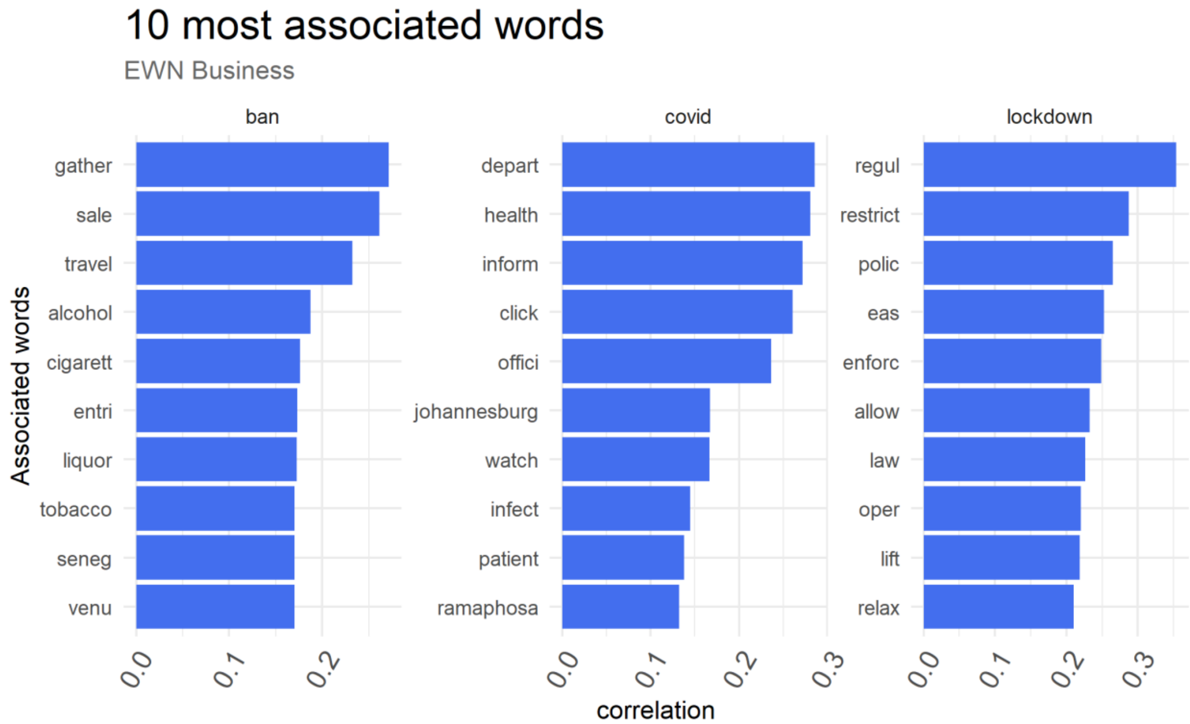

FIGURE 29: WORD PAIRS: EWN BUSINESS

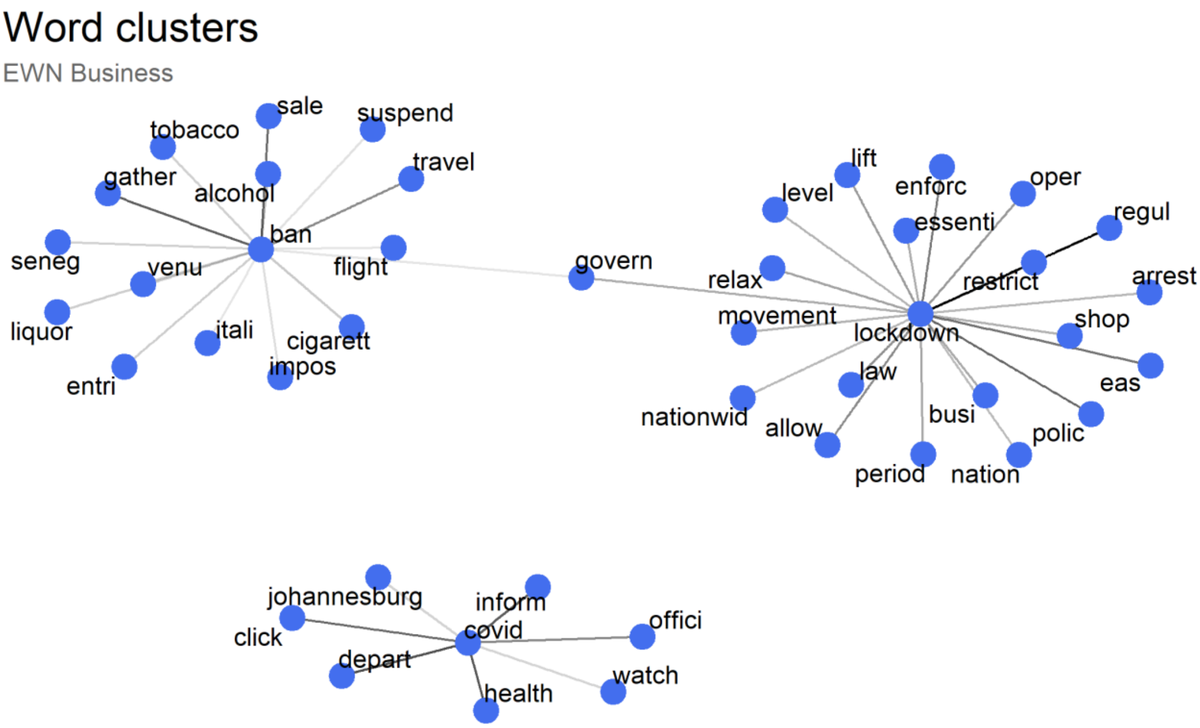

FIGURE 30: WORD CLUSTERS: EWN BUSINESS

The appearance of *seneg* (Senegal) under the *ban* facet of Figure 29 may be an indication that EWN Business reports more frequently on international matters. Moreover, the words associated with *ban* are more diverse, incorporating mention of the travel ban and the prohibition of gatherings.

The visualisation of EWN Business' word clusters in Figure 30 shows that the clusters are almost completely detached from one another, with only a single word (*govern*) connecting *lockdown* and *ban*. This could refer to the government's involvement in the implementation of both of these measures. As mentioned previously, this could be caused by the outcomes being restricted to the top 20 results associated with each Covid-19-specified word. Moreover, the same inference about topic-specificity in Fin24's articles can be made about EWN's articles.

BusinessTech

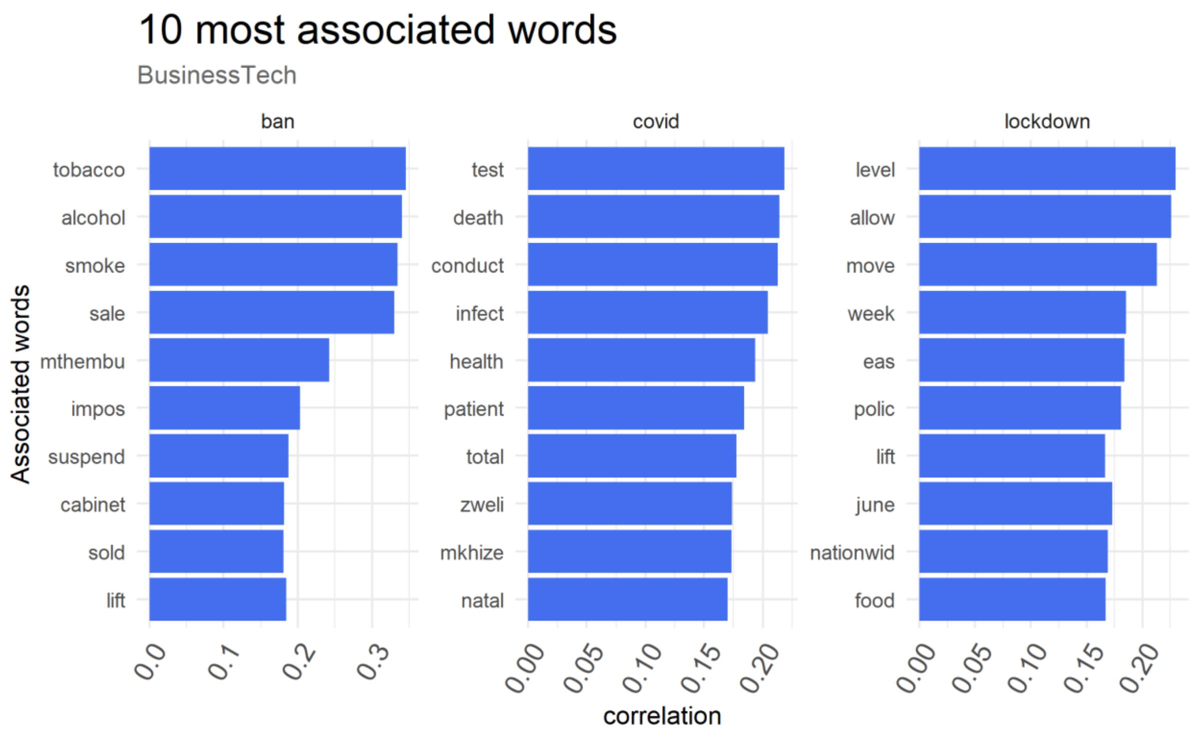

FIGURE 31: WORD PAIRS: BUSINESSTECH

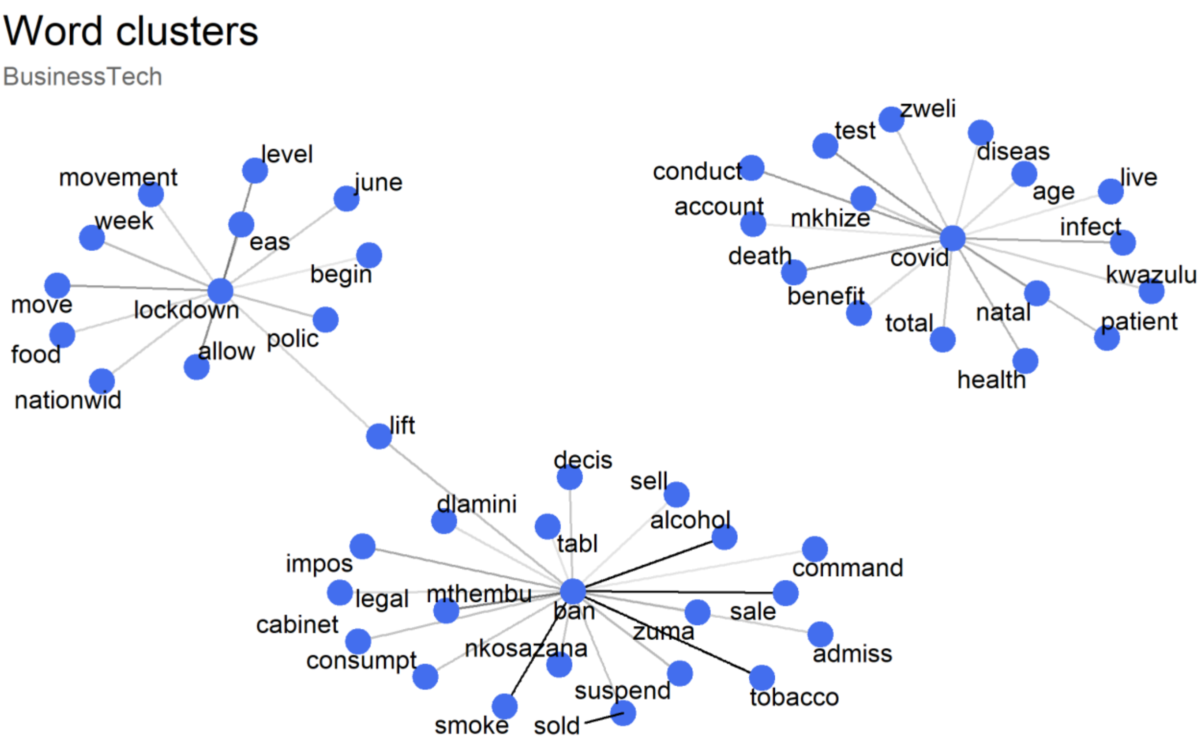

FIGURE 32: WORD CLUSTERS: BUSINESSTECH

BusinessTech's word-pairings and clusters that can be viewed in Figures 31 and 32, respectively, show that BusinessTech articles are more likely to name political players involved in topics surrounding bans and Covid-19. Aside from this, the similar general inferences can be made to those that have been made previously.

Across Media Outlets

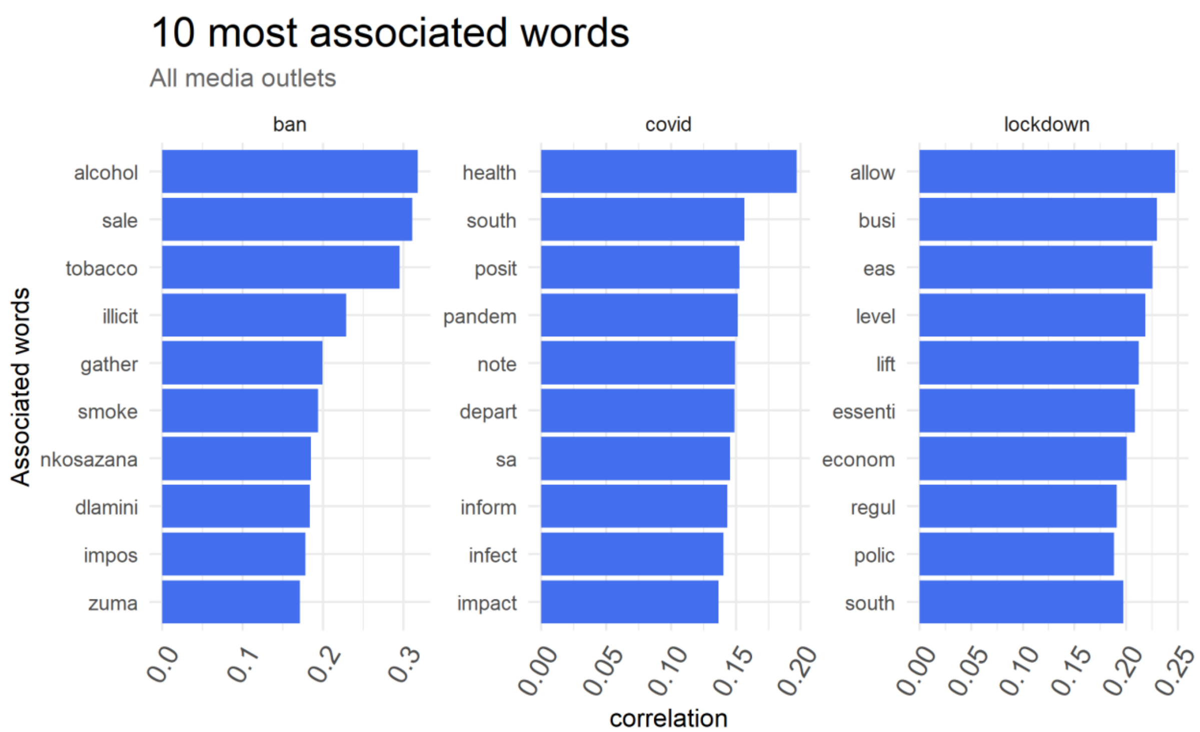

FIGURE 33: WORD PAIRS: ALL MEDIA OUTLETS

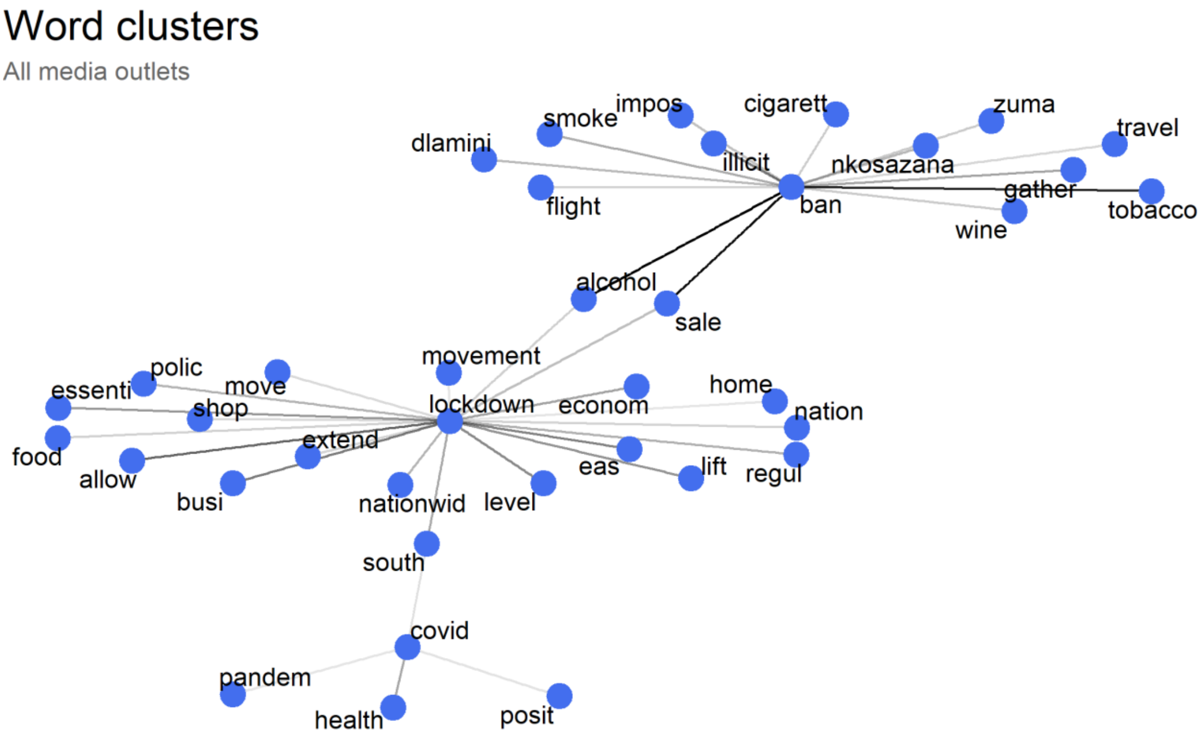

FIGURE 34: WORD CLUSTERS: ALL MEDIA OUTLETS

Through the aggregation of all the media outlet's data, the word-pairings identified in Figure 33 identify those words a news reader is most likely to see in the presence of the specified word. When considered alongside the word clusters in Figure 34, a more holistic view of business news outlets' word usage can be established.

Sentiment Analysis

Overall sentiment of each and across all media outlets was determined using the Loughran-McDonald sentiment lexicon, followed by the individual analyses of the top contributing words towards each sentiment category for each media outlet. Top words contributing towards each category was limited to the top 20 in specific cases.

Fin24

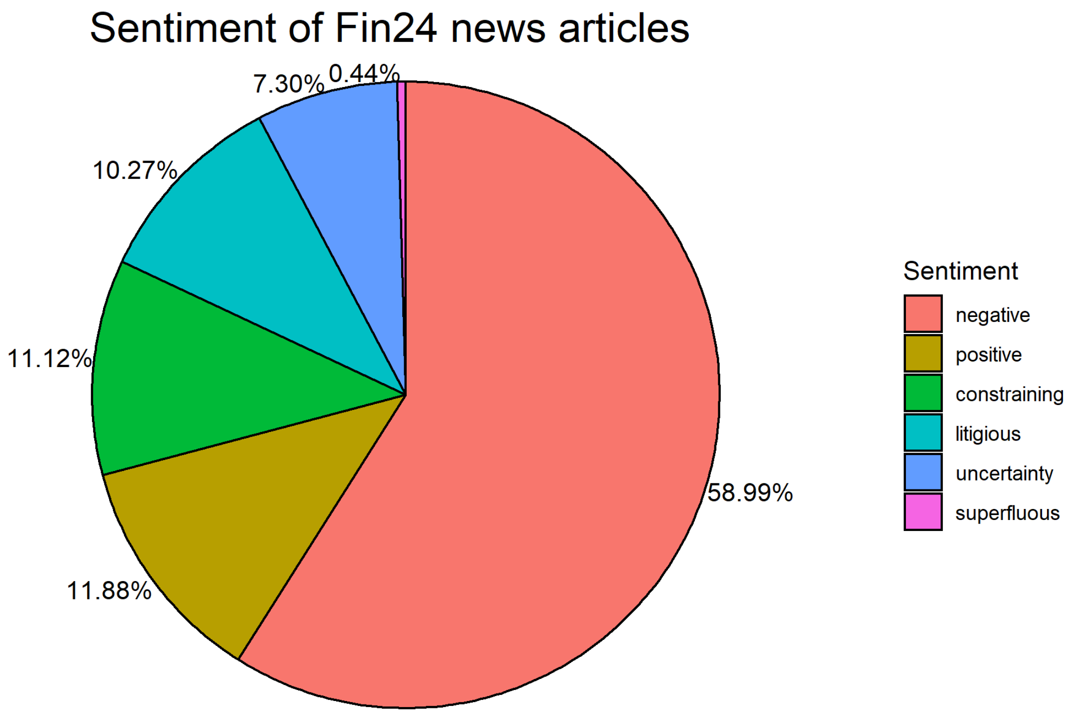

FIGURE 35: SENTIMENT ANALYSIS: FIN 24

As seen in Figure 35 above, the consensus of Fin24 articles during the allocated period of data collection portrays an overall negative sentiment of approximately 58.99%. Despite this, Fin24 had the highest positive sentiment percentage amongst the news outlets. The top contributing words for each sentiment category are visualised in the figures below (Figure 36 to Figure 41).

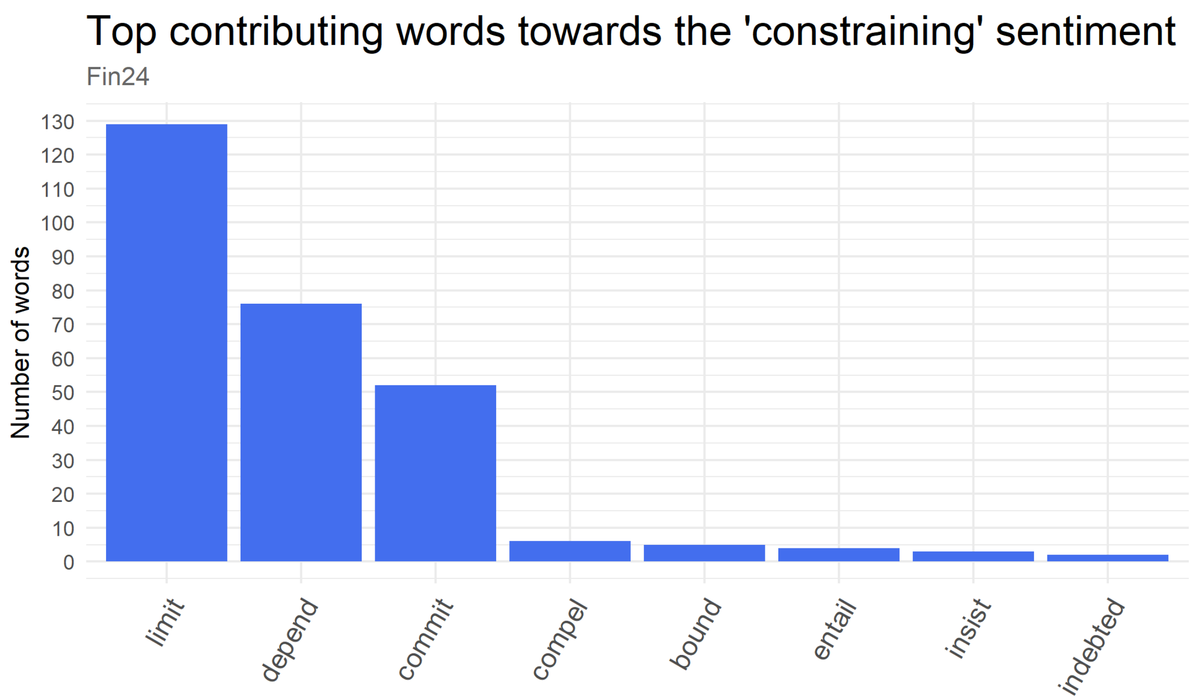

FIGURE 36: TOP CONTRIBUTING WORDS: CONSTRAINING: FIN24

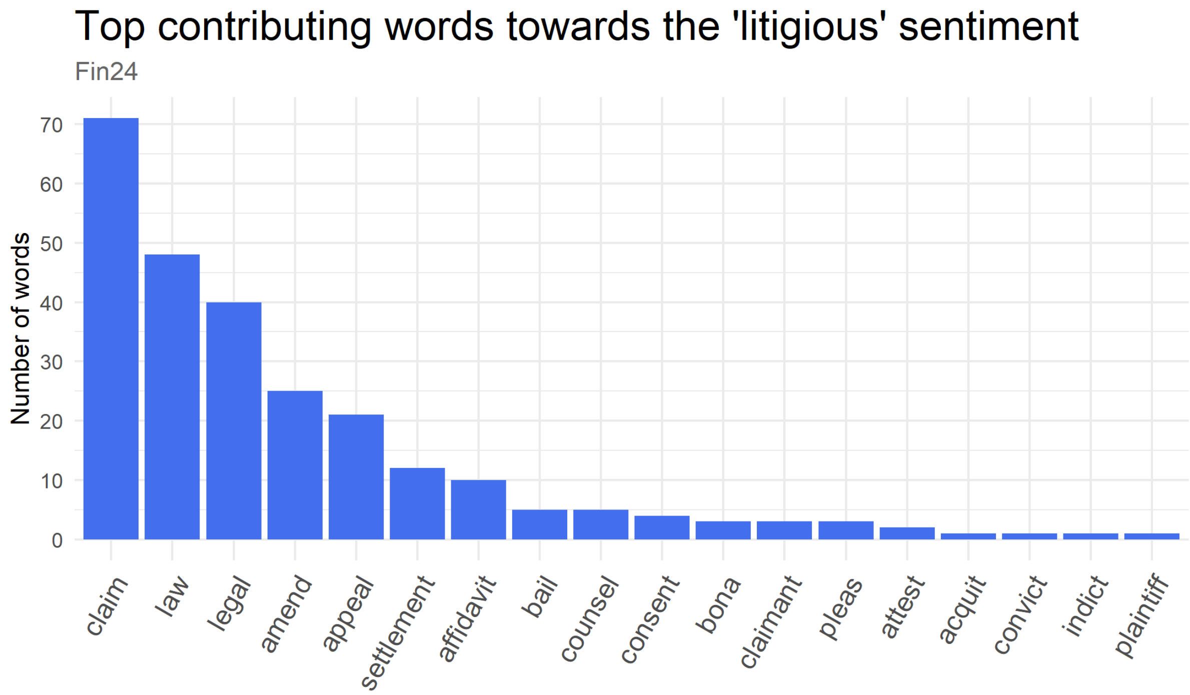

FIGURE 37: TOP CONTRIBUTING WORDS: LITIGIOUS: FIN24

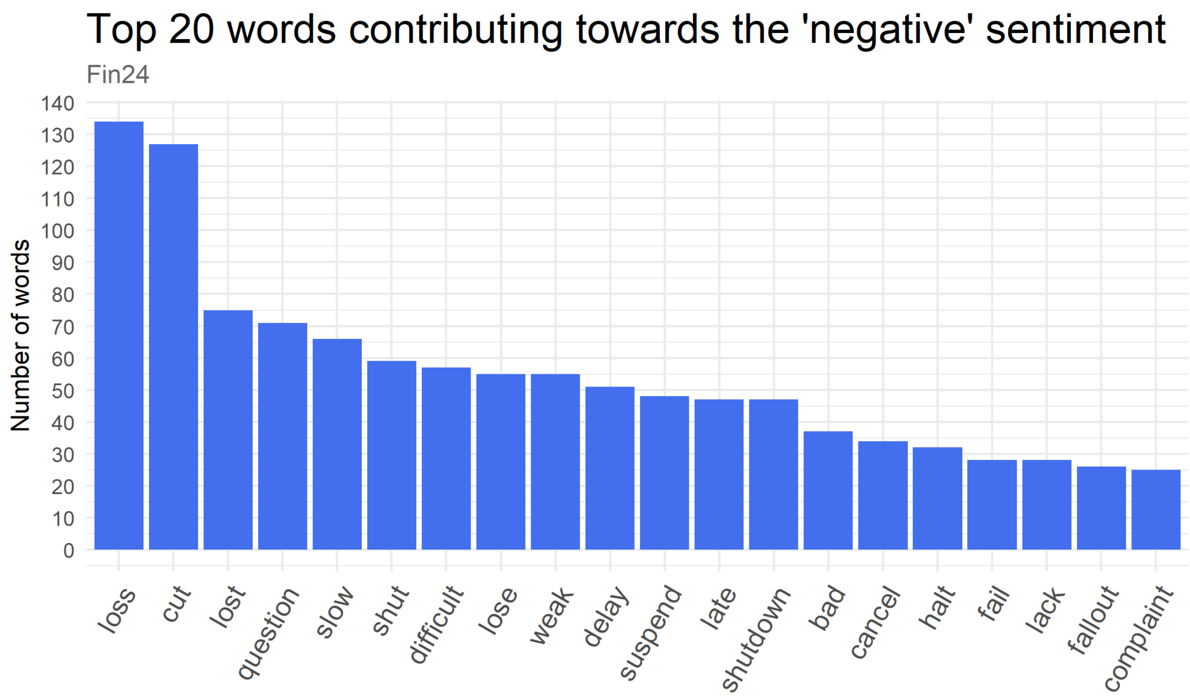

FIGURE 38: TOP CONTRIBUTING WORDS: NEGATIVE: FIN24

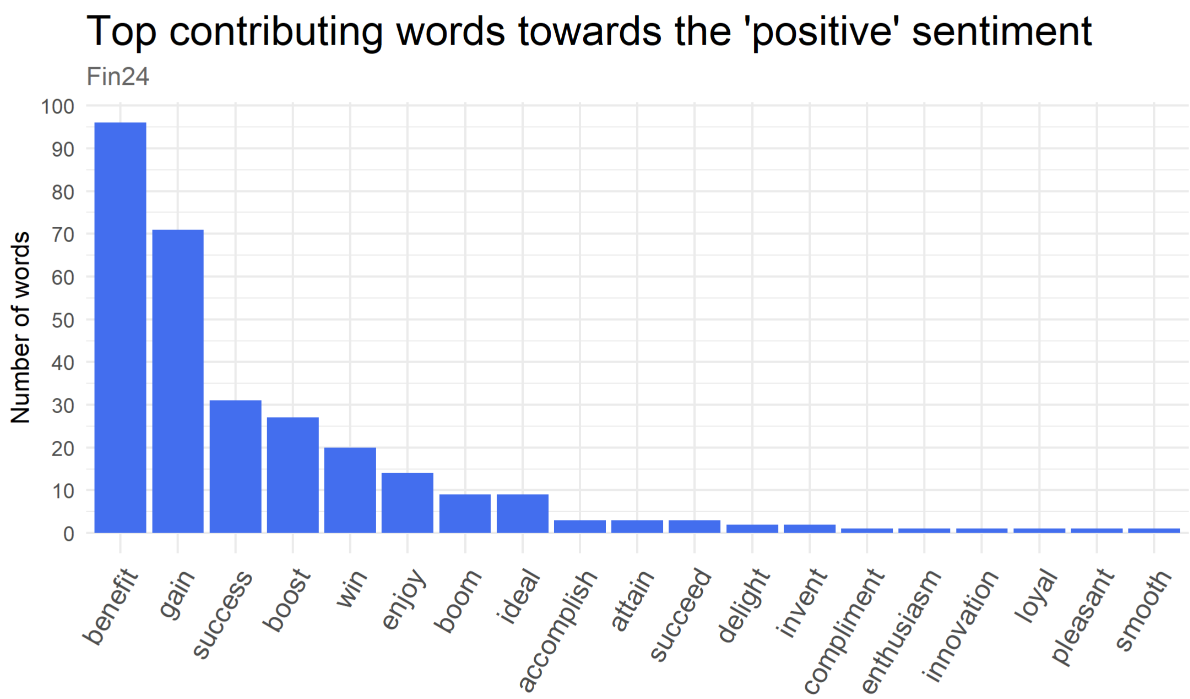

FIGURE 39: TOP CONTRIBUTING WORDS: POSITIVE: FIN24

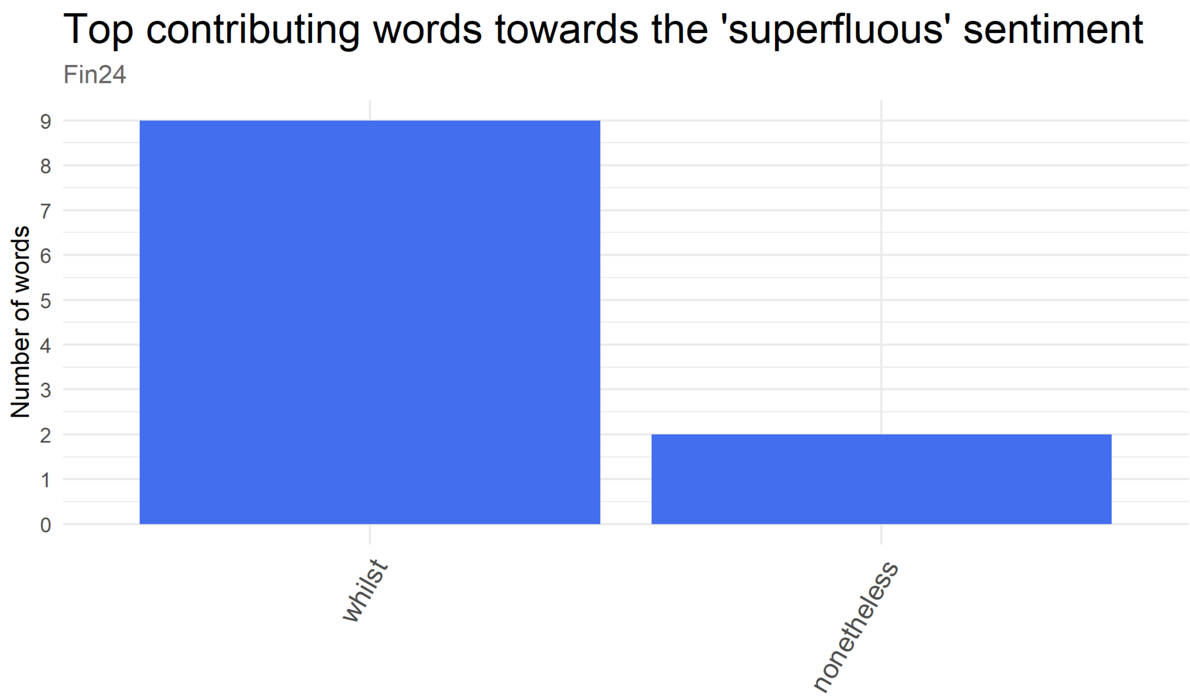

FIGURE 40: TOP CONTRIBUTING WORDS: SUPERFLUOUS: FIN24

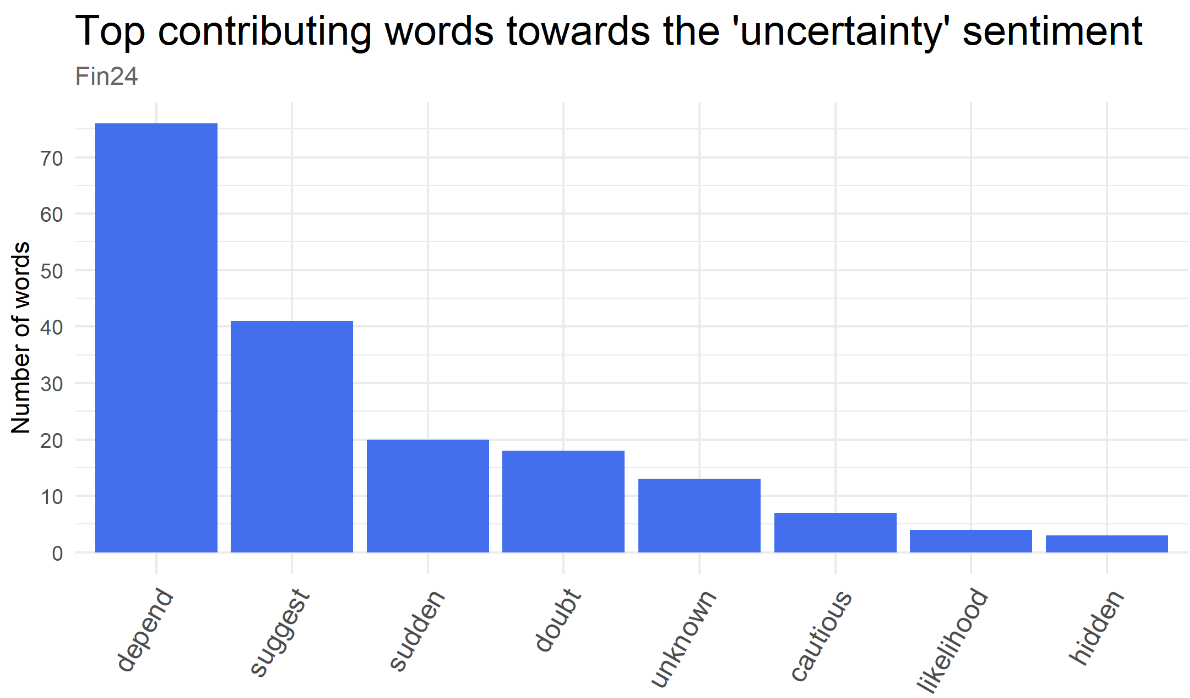

FIGURE 41: TOP CONTRIBUTING WORDS: UNCERTAINTY: FIN24

Moneyweb

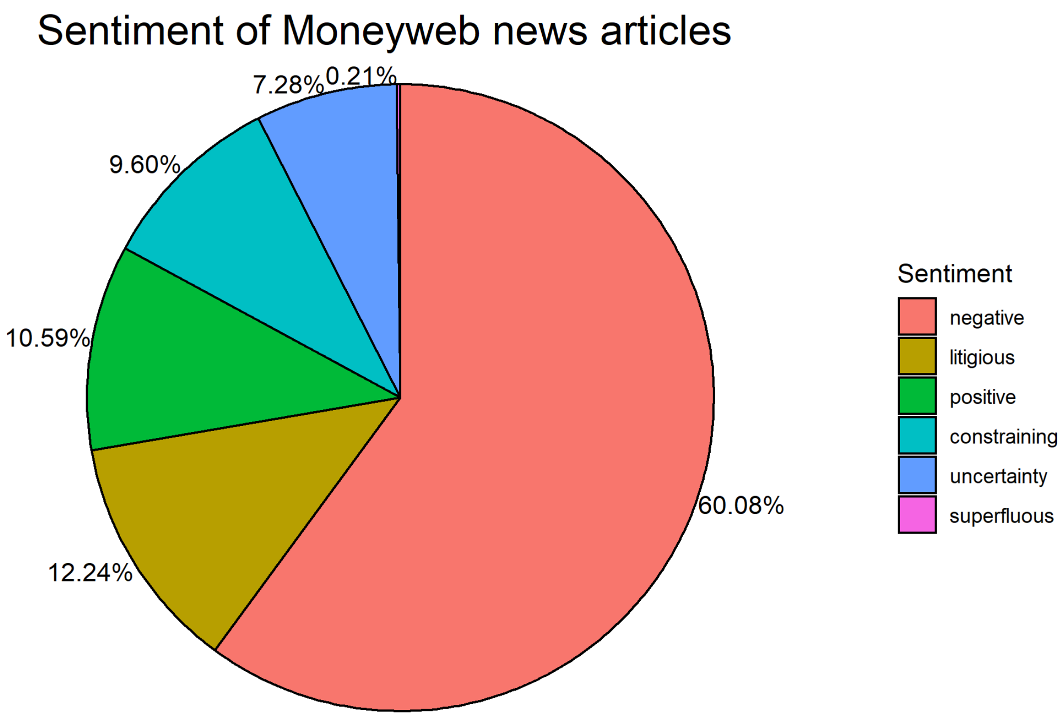

FIGURE 42: SENTIMENT ANALYSIS: MONEYWEB

As seen in Figure 42 above, the consensus of Moneyweb articles during the allocated period portrays an overall negative sentiment of approximately 60.08%, which is the highest negative sentiment percentage between the media outlets. The top contributing words for each sentiment category are visualised in Figures 43 to 48 which can be seen below.

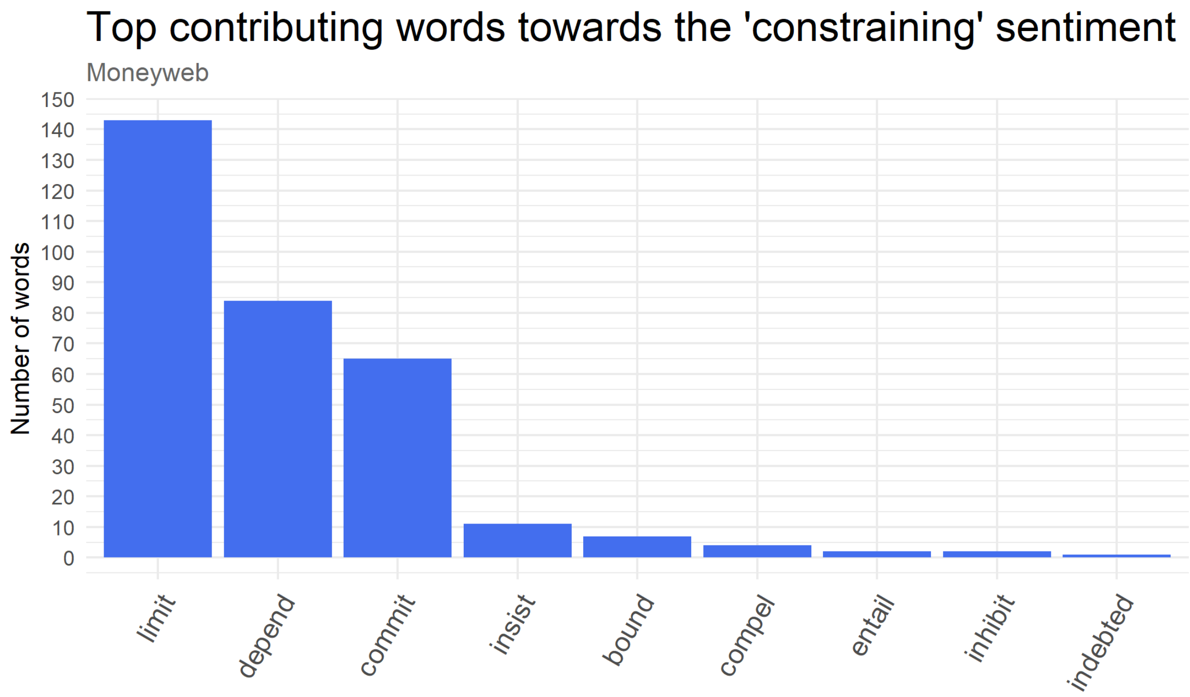

FIGURE 43: TOP CONTRIBUTING WORDS: CONSTRAINING: MONEYWEB

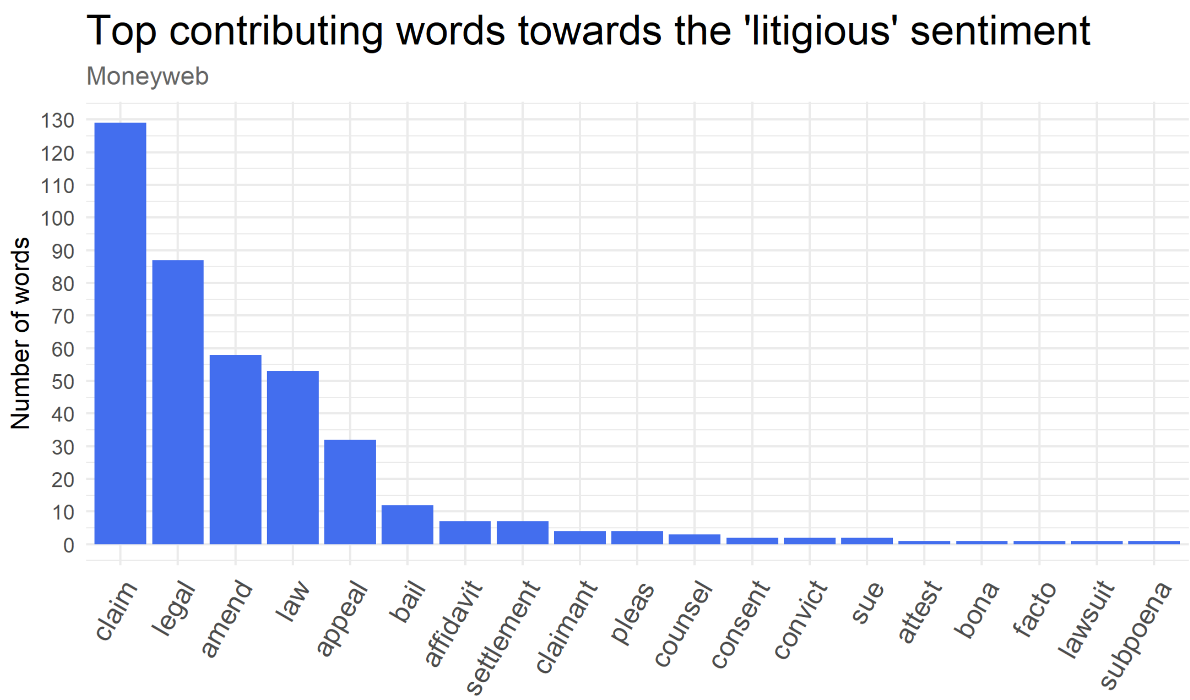

FIGURE 44: TOP CONTRIBUTING WORDS: LITIGIOUS: MONEYWEB

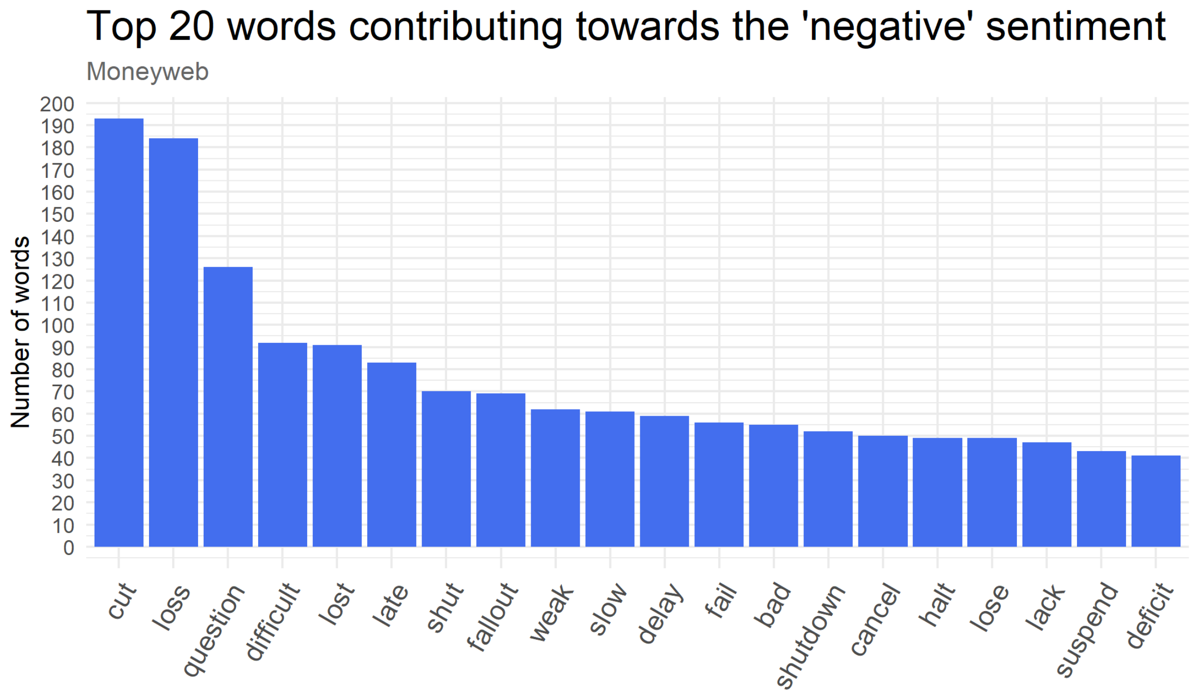

FIGURE 45: TOP CONTRIBUTING WORDS: NEGATIVE: MONEYWEB

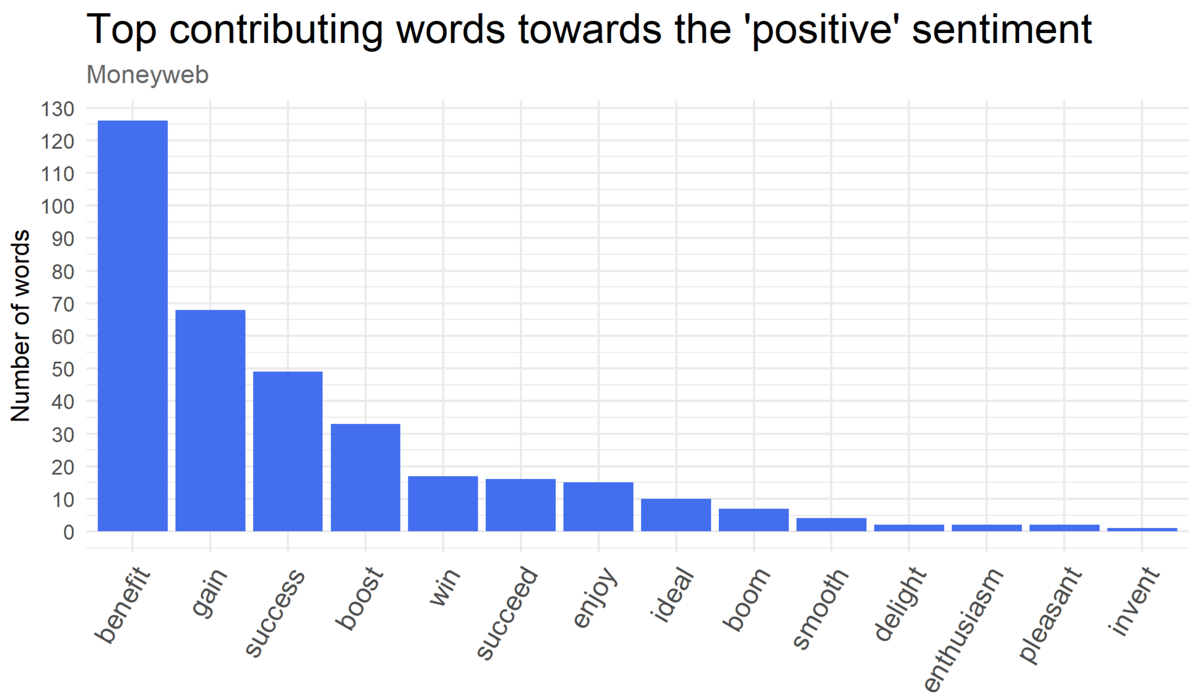

FIGURE 46: TOP CONTRIBUTING WORDS: POSITIVE: MONEYWEB

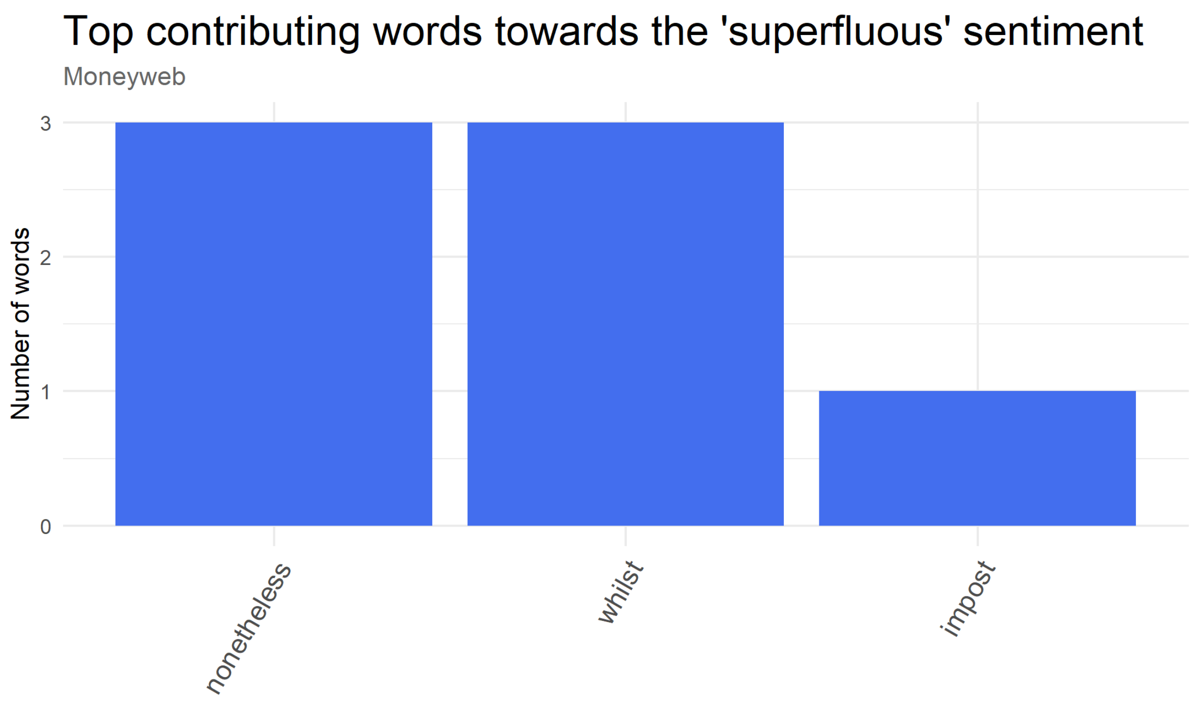

FIGURE 47: TOP CONTRIBUTING WORDS: SUPERFLUOUS: MONEYWEB

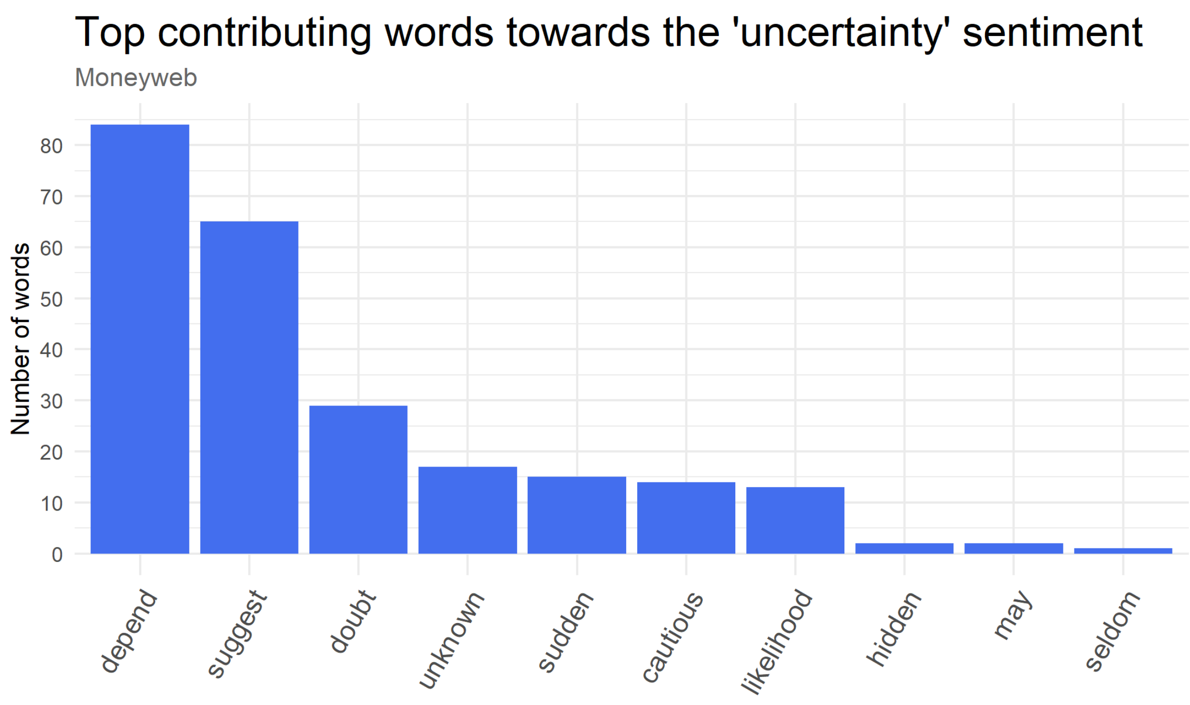

FIGURE 48: TOP CONTRIBUTING WORDS: UNCERTAINTY: MONEYWEB

BizNews

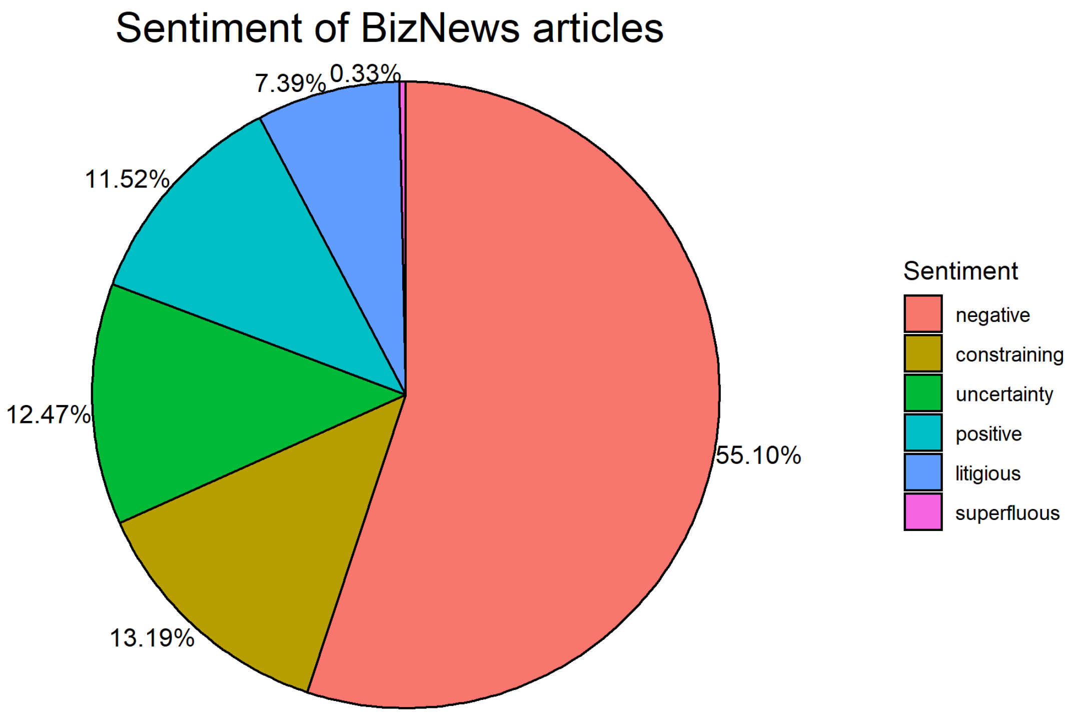

FIGURE 49: SENTIMENT ANALYSIS: BIZNEWS

Following the above trends, BizNews articles also followed a majority negative sentiment of approximately 55.10%, as seen in Figure 49 above. The top contributing words of each sentiment categories can be seen in the figures below (Figure 50 to Figure 55).

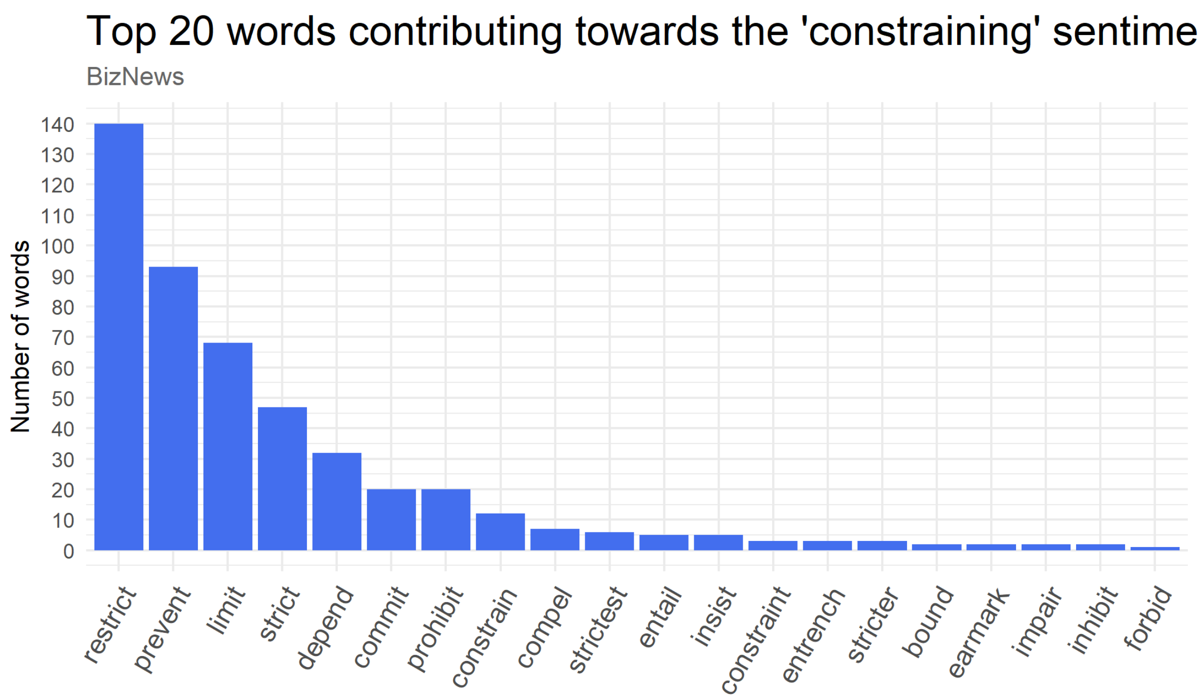

FIGURE 50: TOP CONTRIBUTING WORDS: CONSTRAINING: BIZNEWS

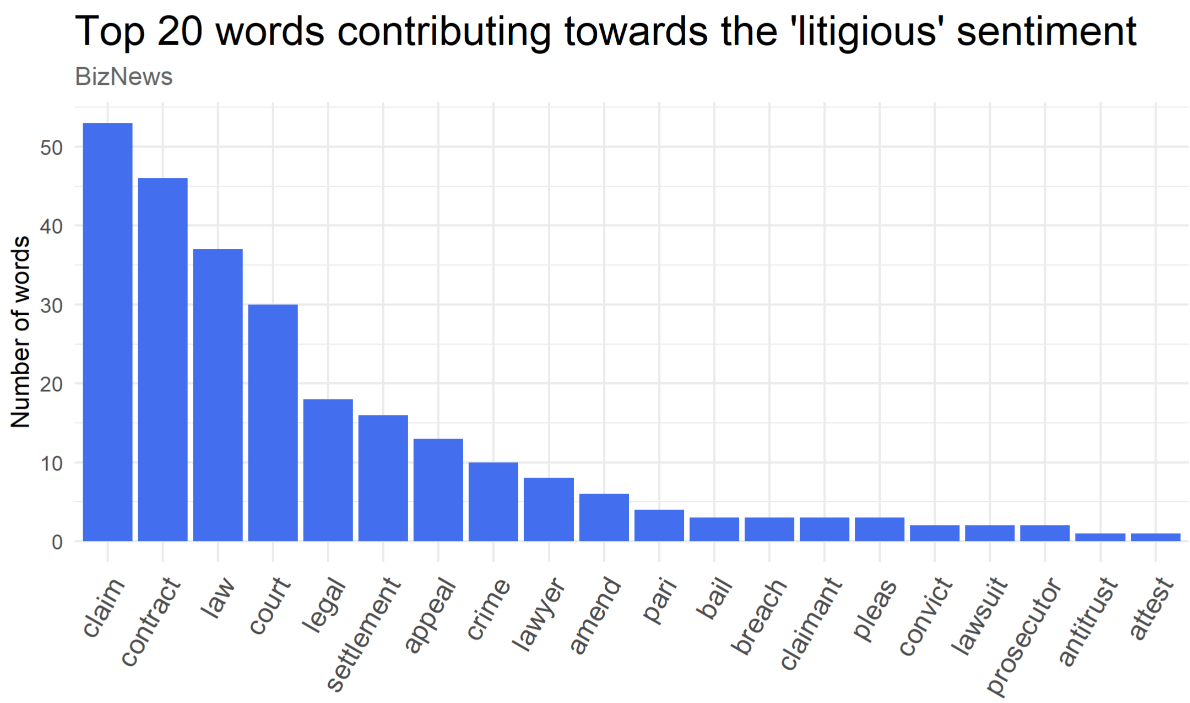

FIGURE 51: TOP CONTRIBUTING WORDS: LITIGIOUS: BIZNEWS

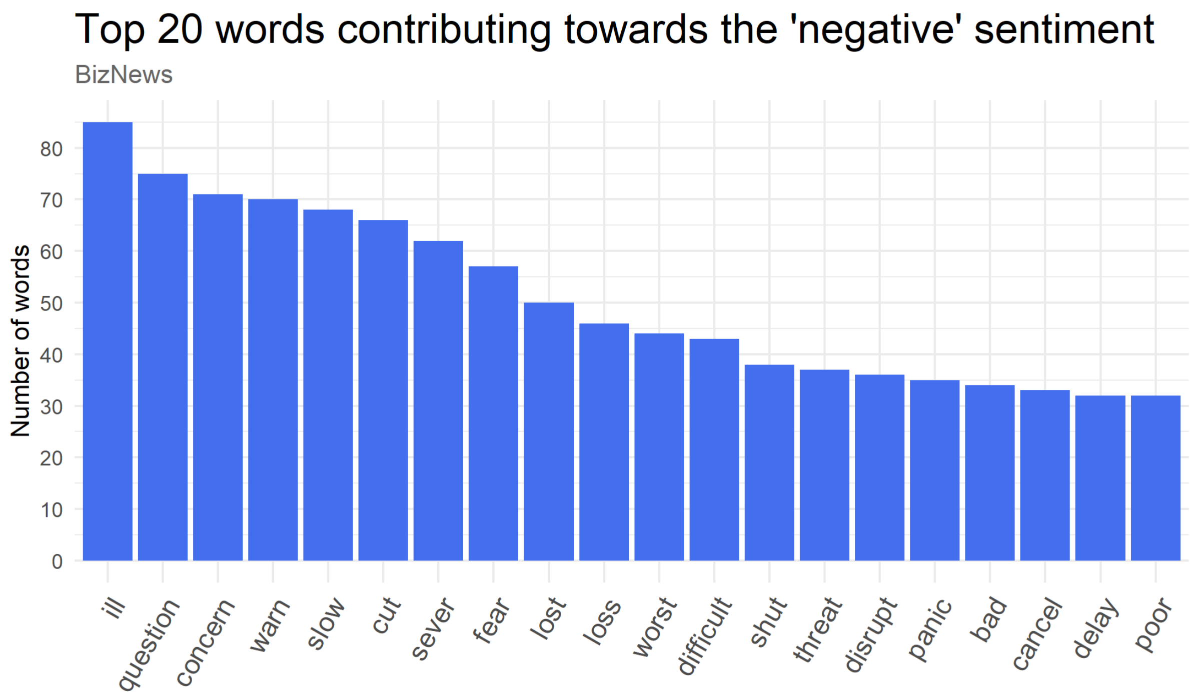

FIGURE 52: TOP CONTRIBUTING WORDS: NEGATIVE: BIZNEWS

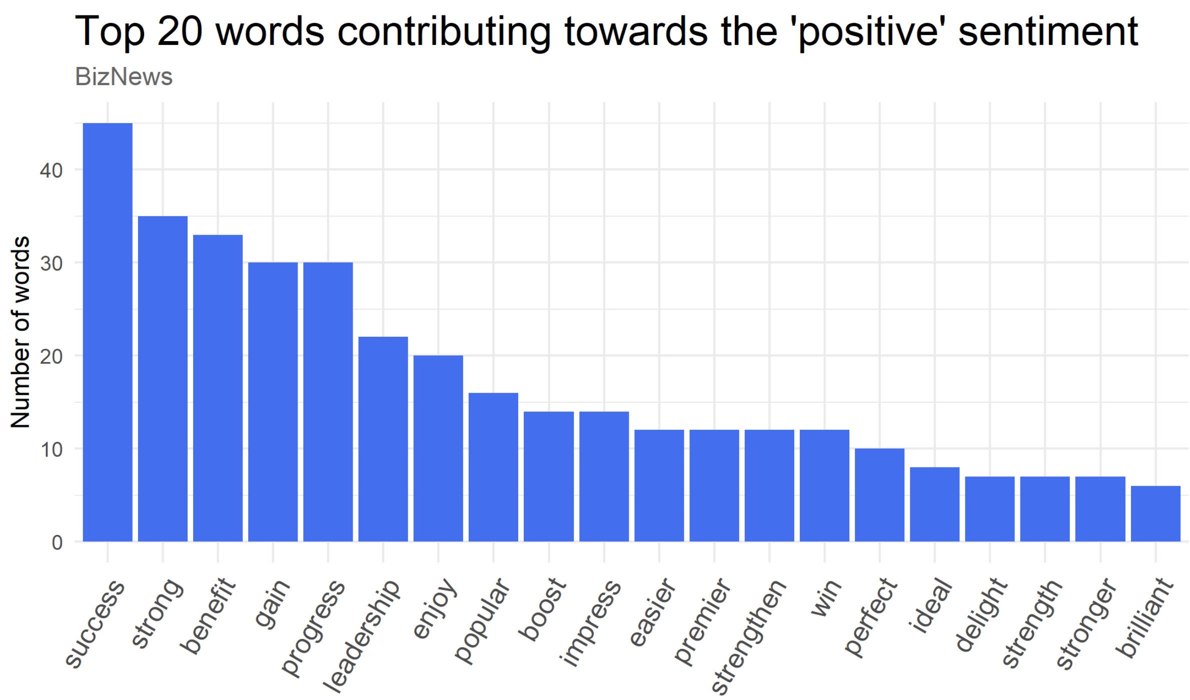

FIGURE 53: TOP CONTRIBUTING WORDS: POSITIVE: BIZNEWS

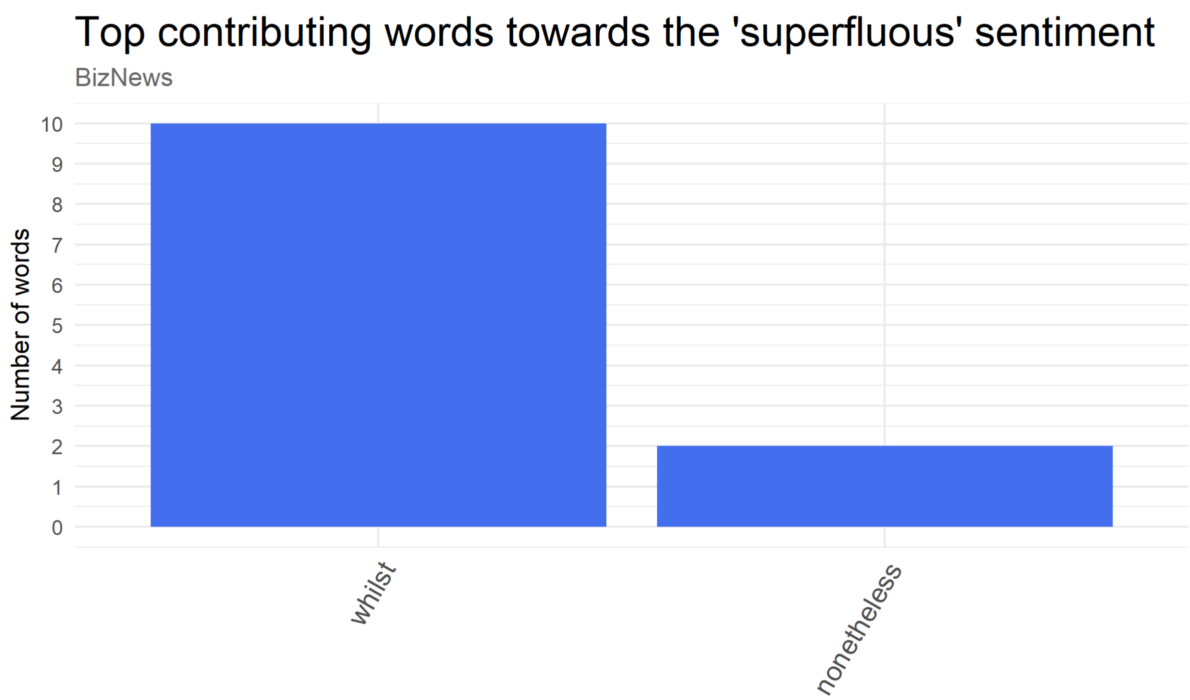

FIGURE 54: TOP CONTRIBUTING WORDS: SUPERFLUOUS: BIZNEWS

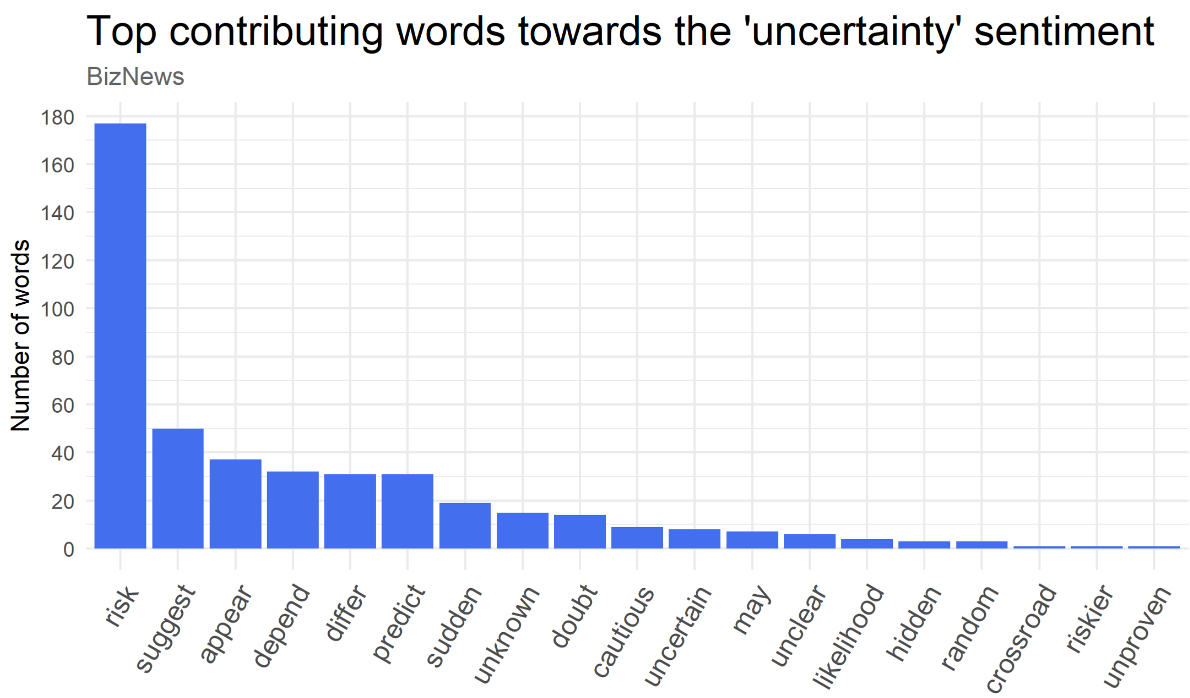

FIGURE 55: TOP CONTRIBUTING WORDS: UNCERTAINTY: BIZNEWS

EWN Business

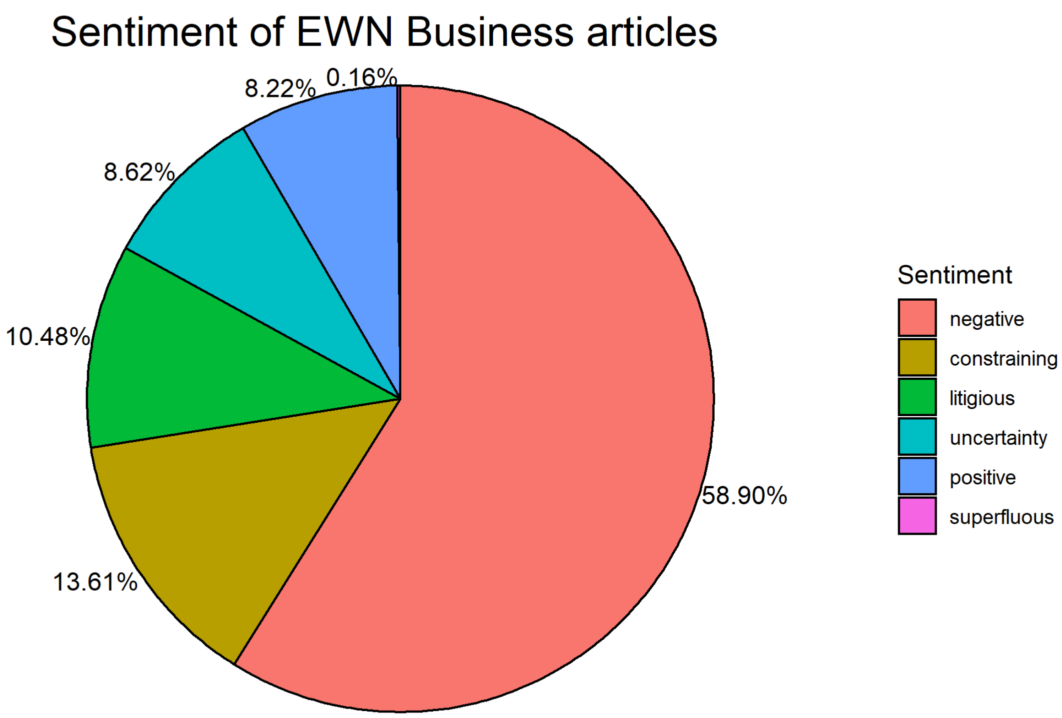

FIGURE 56: SENTIMENT ANALYSIS: EWN BUSINESS

In addition to having a exceedingly negative sentiment (58.9%) as seen in Figure 56 above, EWN Business also has the lowest positive sentiment amongst the media outlets. The breakdown of the top contributing words to each sentiment can be found in the graphs below (Figure 57 to Figure 61).

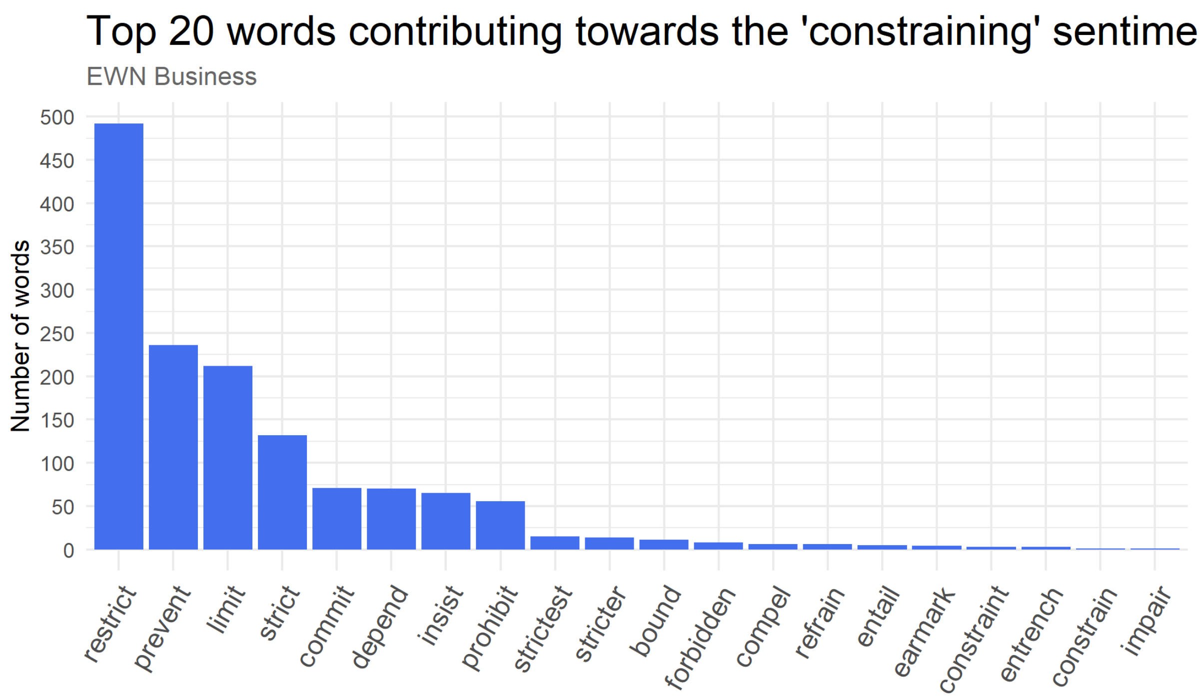

FIGURE 57: TOP CONTRIBUTING WORDS: CONSTRAINING: EWN BUSINESS

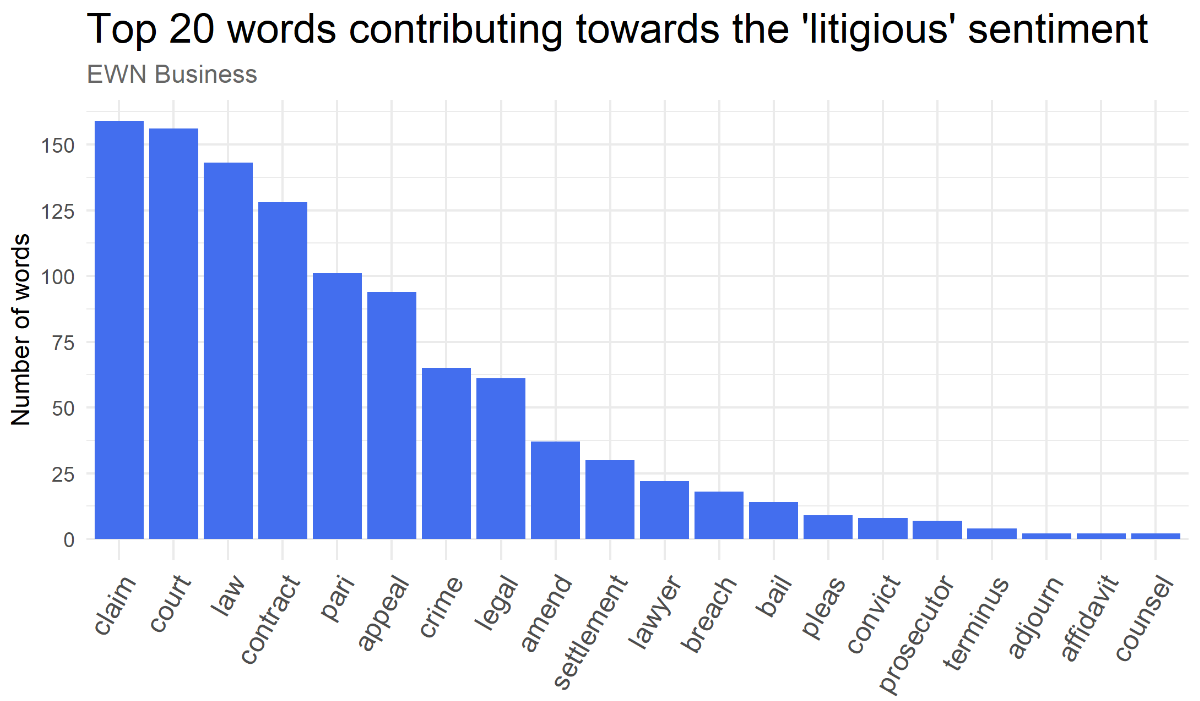

FIGURE 58: TOP CONTRIBUTING WORDS: LITIGIOUS: EWN BUSINESS

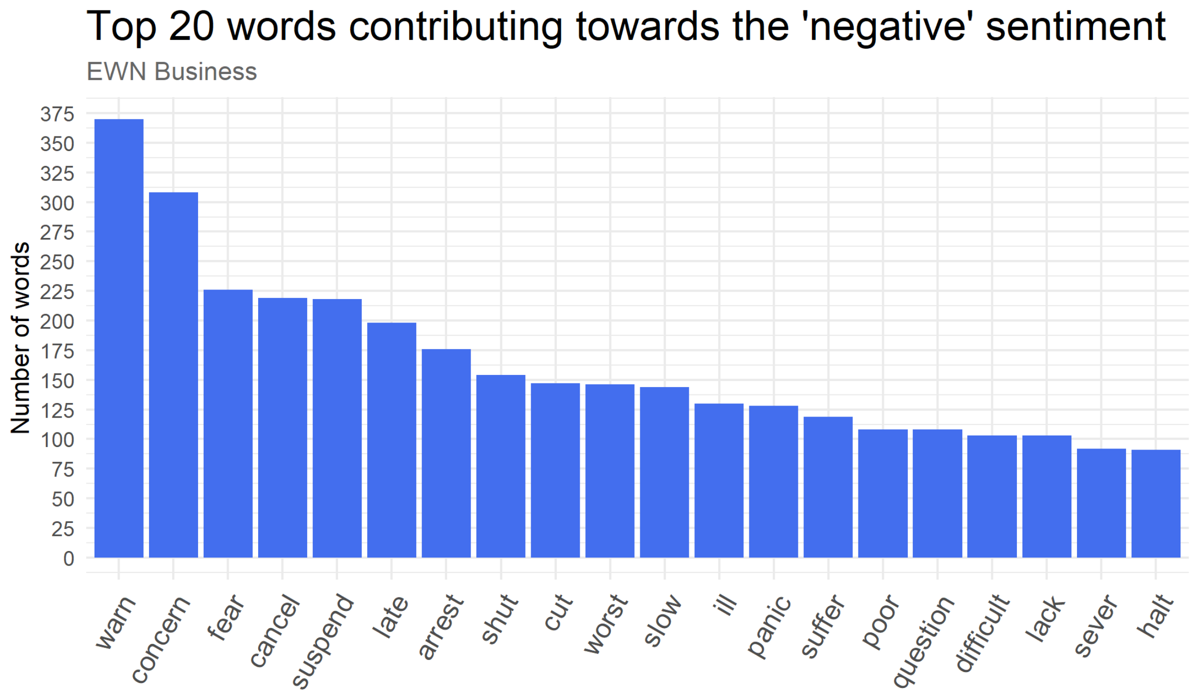

FIGURE 59: TOP CONTRIBUTING WORDS: NEGATIVE: EWN BUSINESS

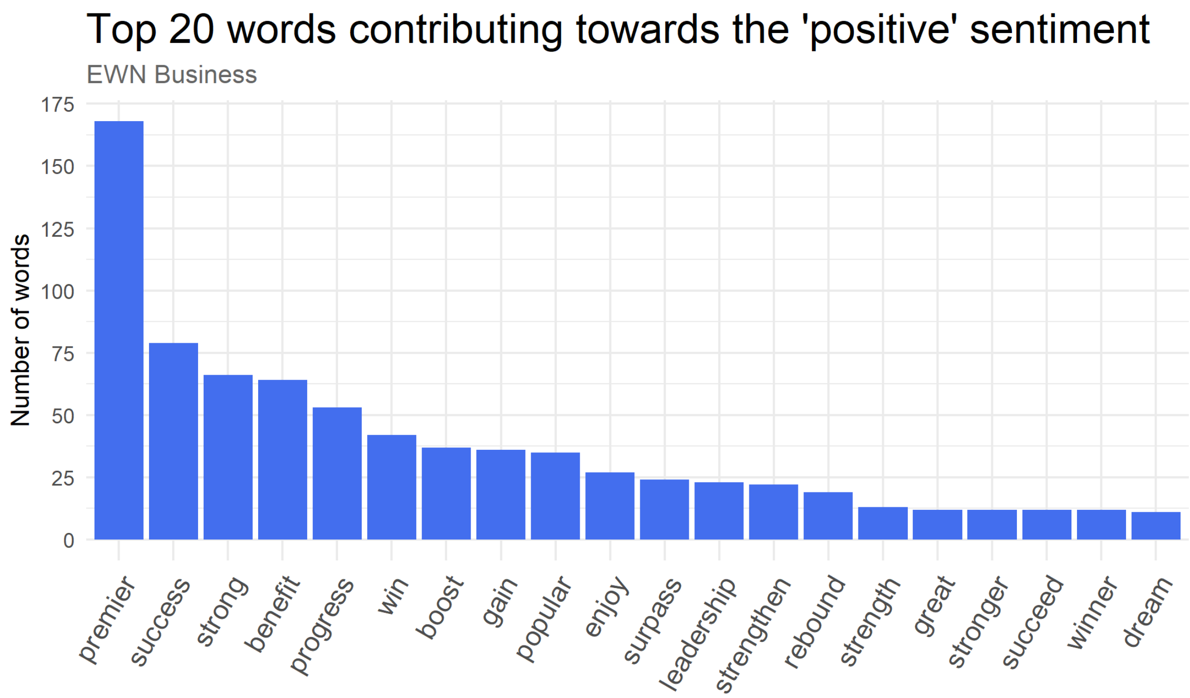

FIGURE 60: TOP CONTRIBUTING WORDS: POSITIVE: EWN BUSINESS

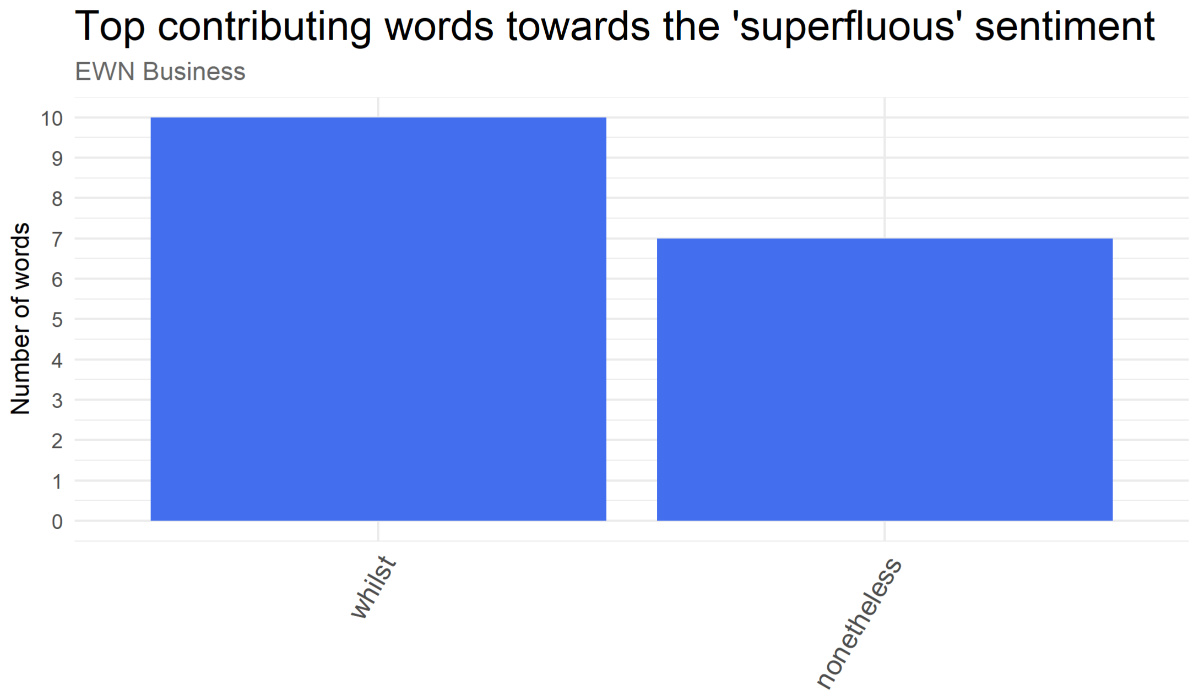

FIGURE 60: TOP CONTRIBUTING WORDS: SUPERFLUOUS: EWN BUSINESS

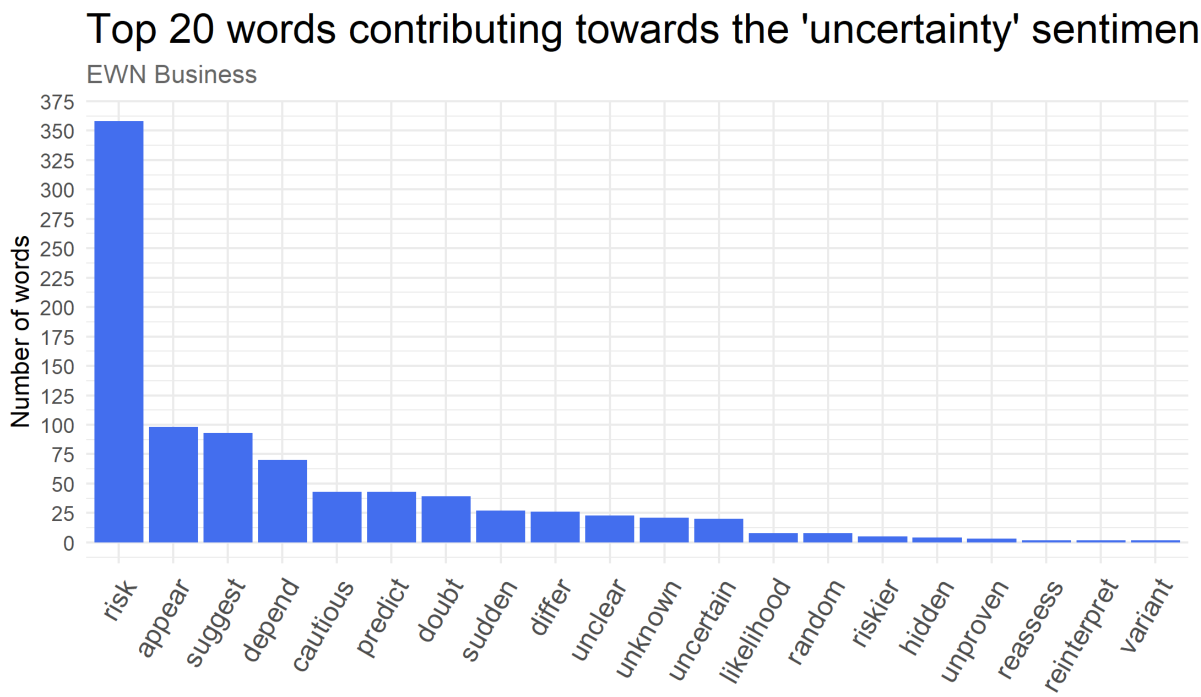

FIGURE 61: TOP CONTRIBUTING WORDS: UNCERTAINTY: EWN BUSINESS

BusinessTech

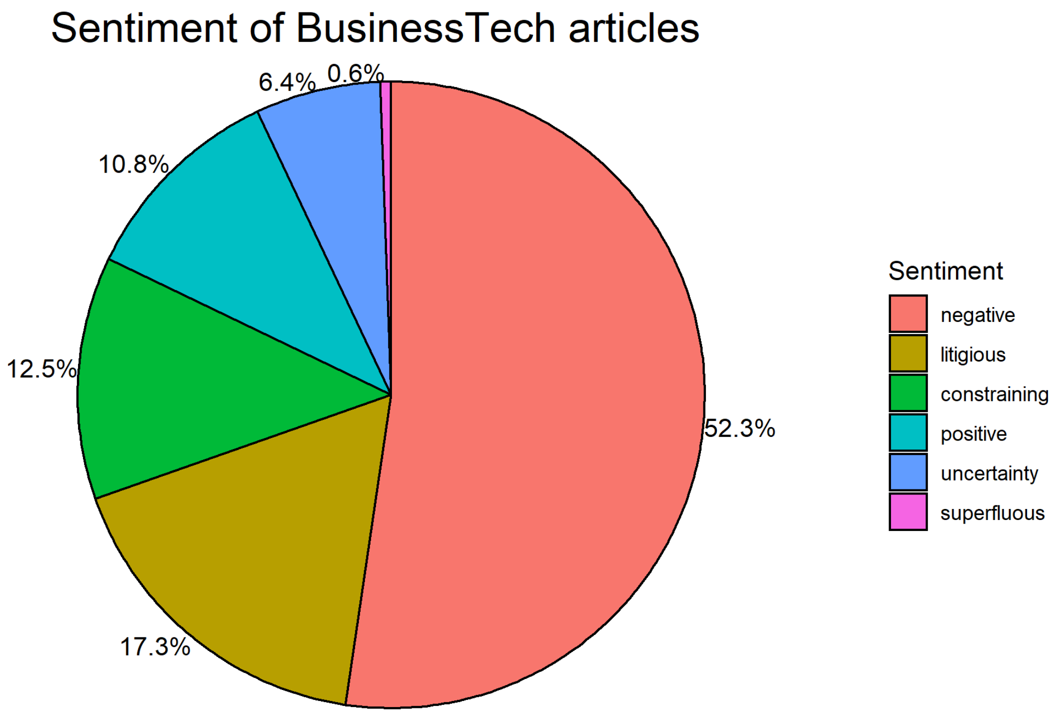

FIGURE 62: SENTIMENT ANALYSIS: BUSINESSTECH

BusinessTech had the lowest negative sentiment percentage, despite it still constituting the majority of the sentiment. Moreover, it had a significantly larger litigious sentiment percentage than that of the other media outlets, which could potentially be related to the earlier inference made of it producing more politically driven articles. The word contributions to each sentiment category can be seen in Figures 63 to 68 below.

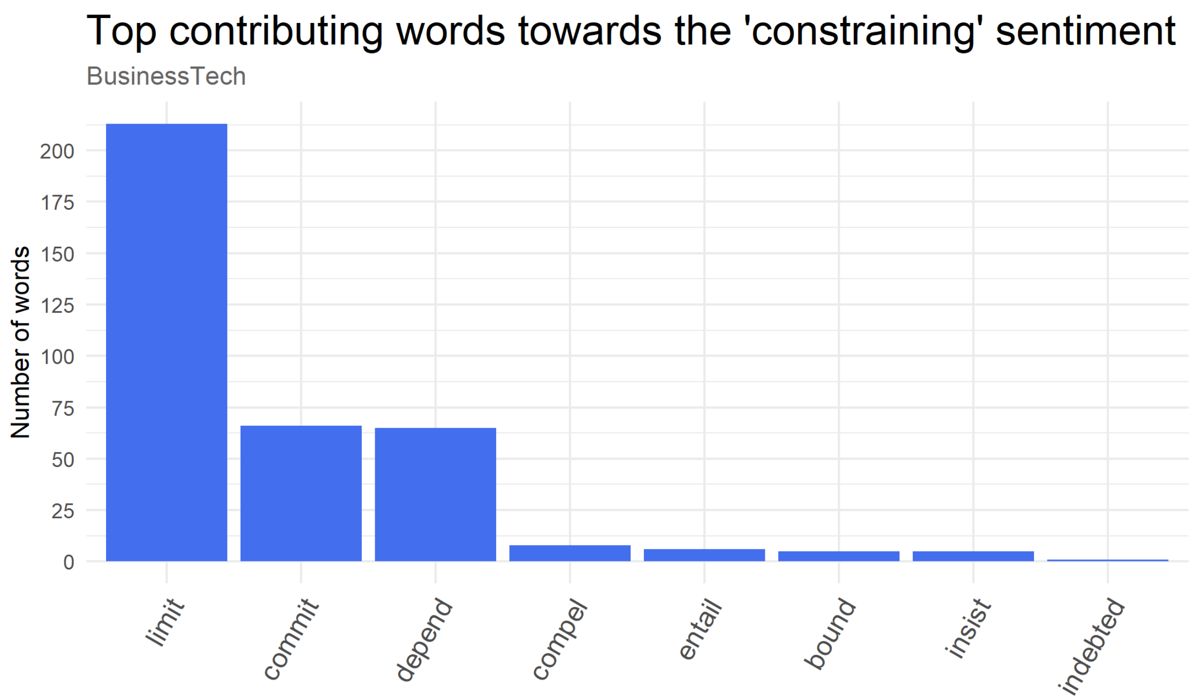

FIGURE 63: TOP CONTRIBUTING WORDS: CONSTRAINING: BUSINESSTECH

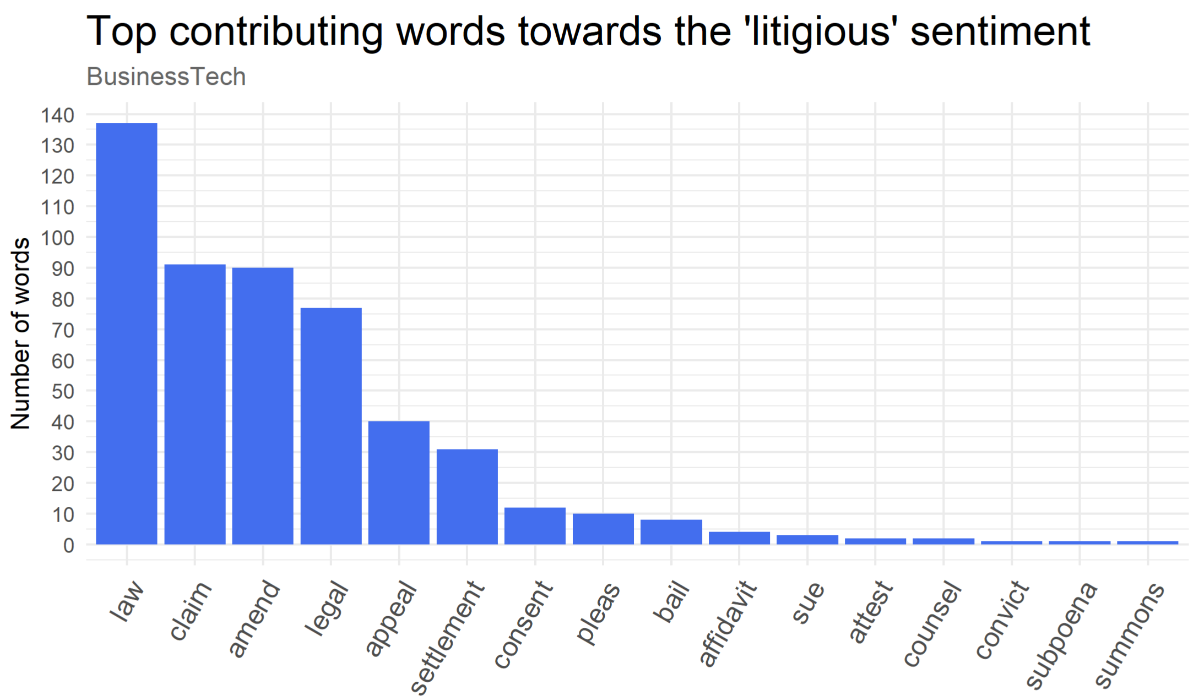

FIGURE 64: TOP CONTRIBUTING WORDS: LITIGIOUS: BUSINESSTECH

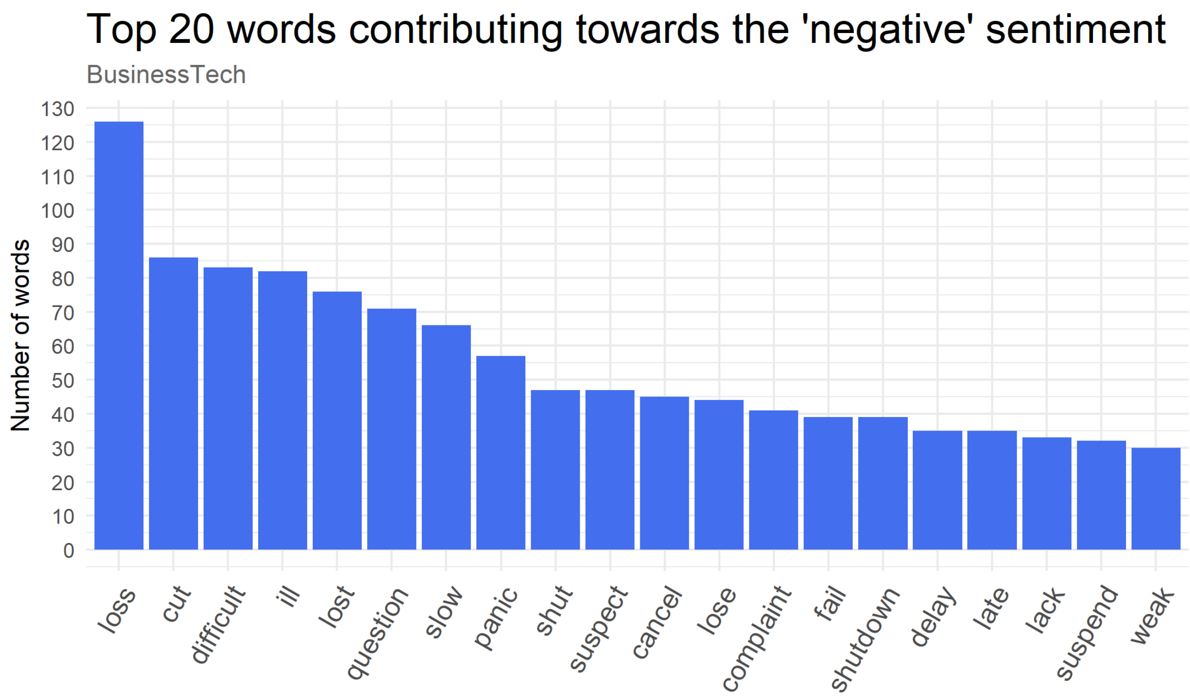

FIGURE 65: TOP CONTRIBUTING WORDS: NEGATIVE: BUSINESSTECH

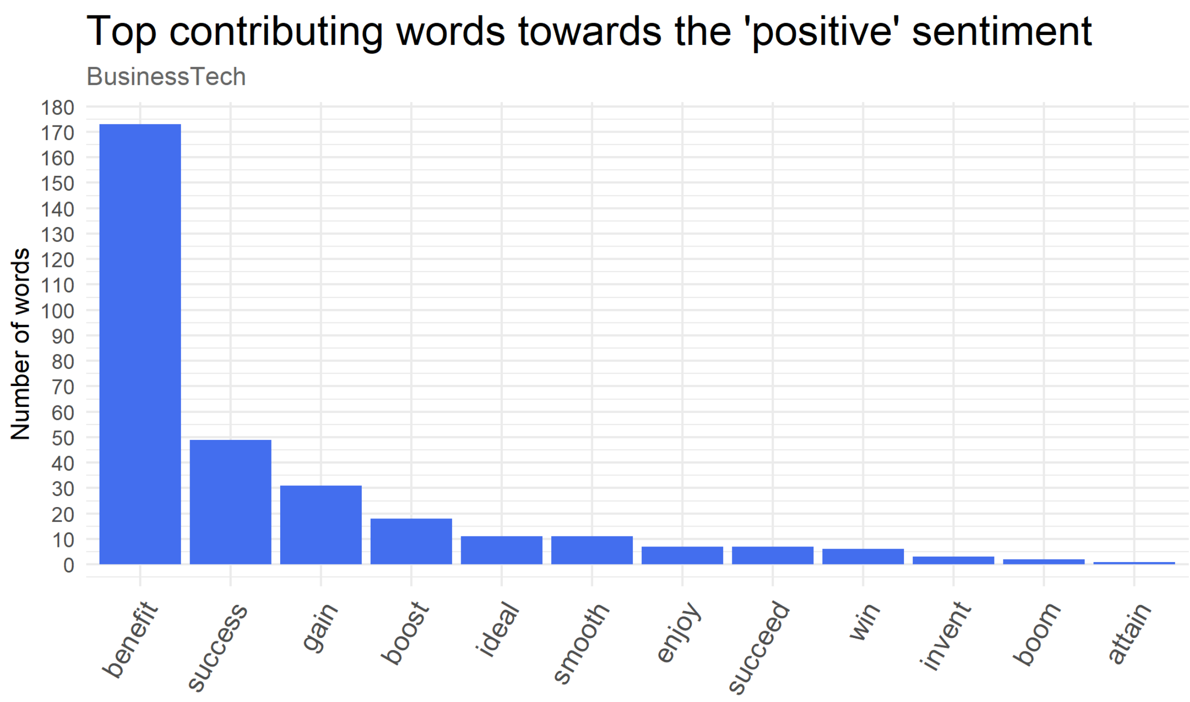

FIGURE 66: TOP CONTRIBUTING WORDS: POSITIVE: BUSINESSTECH

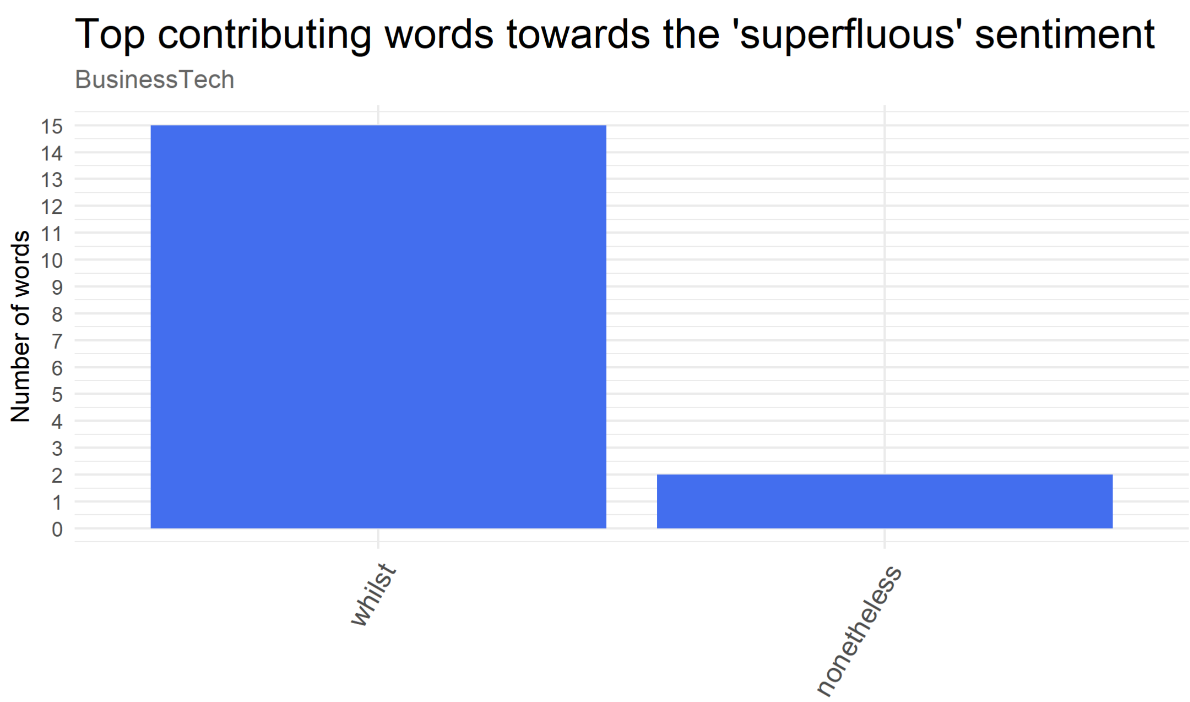

FIGURE 67: TOP CONTRIBUTING WORDS: SUPERFLUOUS: BUSINESSTECH

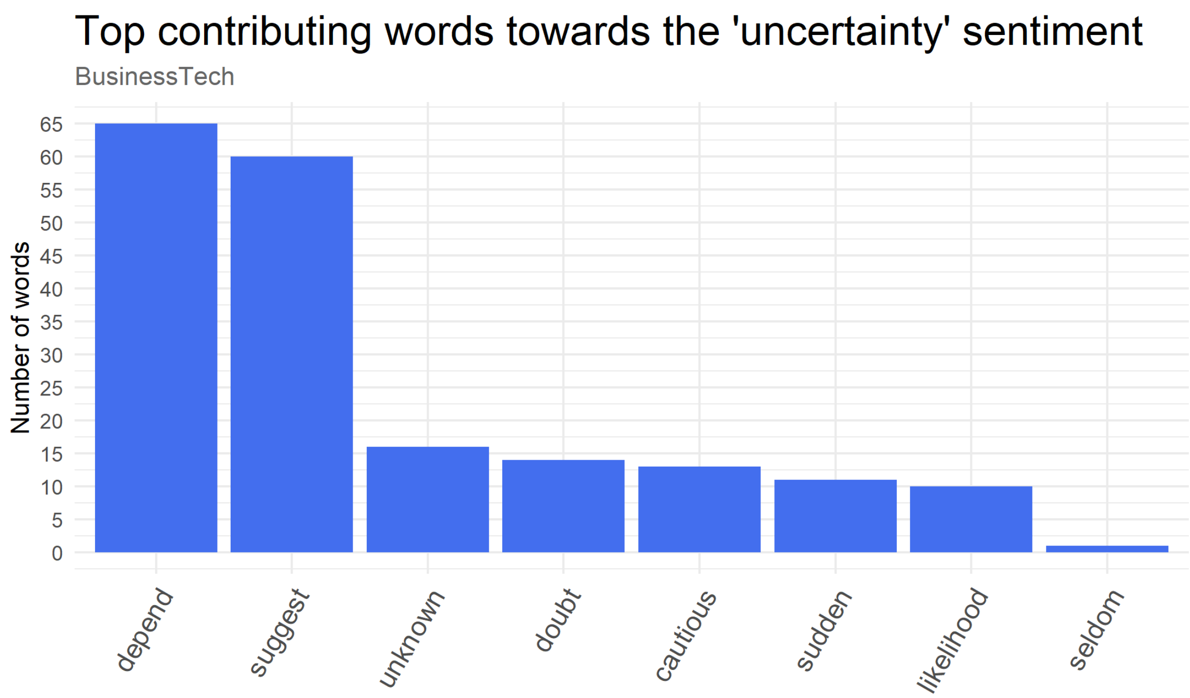

FIGURE 68: TOP CONTRIBUTING WORDS: UNCERTAINTY: BUSINESSTECH

Across All Media Outlets

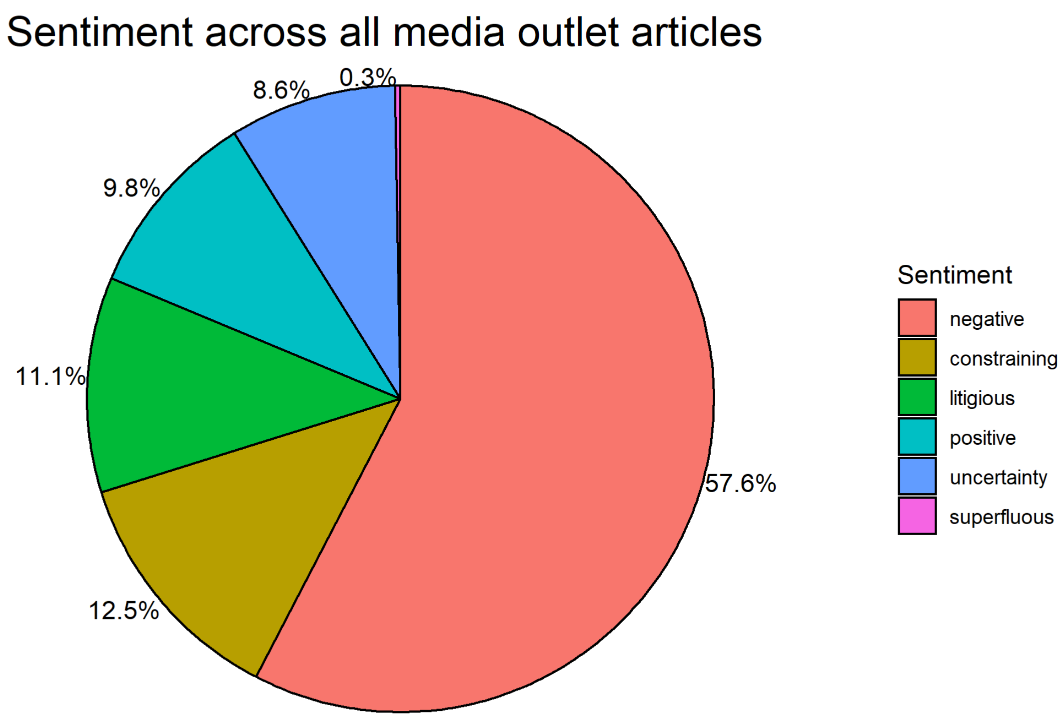

FIGURE 69: SENTIMENT ANALYSIS: ALL MEDIA OUTLETS

As seen in Figure 69 above, sentiment was generally found to be profoundly negative, which can be expected given the economic and societal hardships associated with the Covid-19 pandemic. Despite Covid-19 causing much uncertainty in South Africa and globally, the sentiment category pertaining to uncertainty positioned second to last. The top contributing words to each sentiment category can be seen below in Figures 70 to 75.

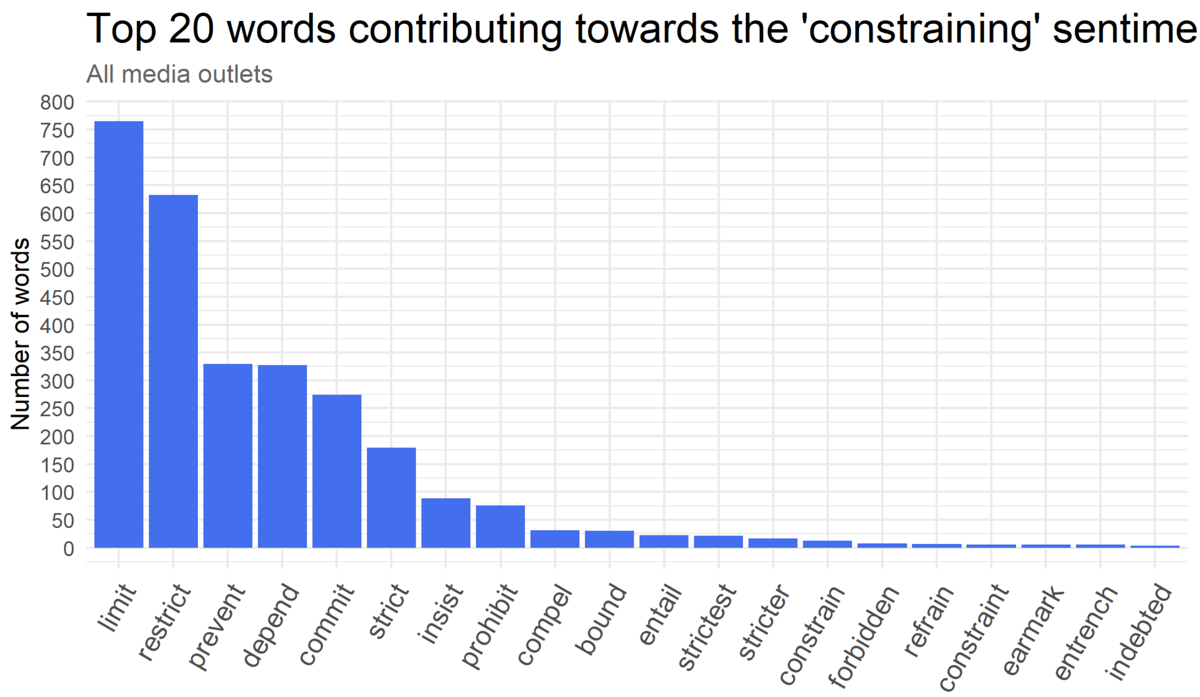

FIGURE 70: TOP CONTRIBUTING WORDS: CONSTRAINING: ALL MEDIA OUTLETS

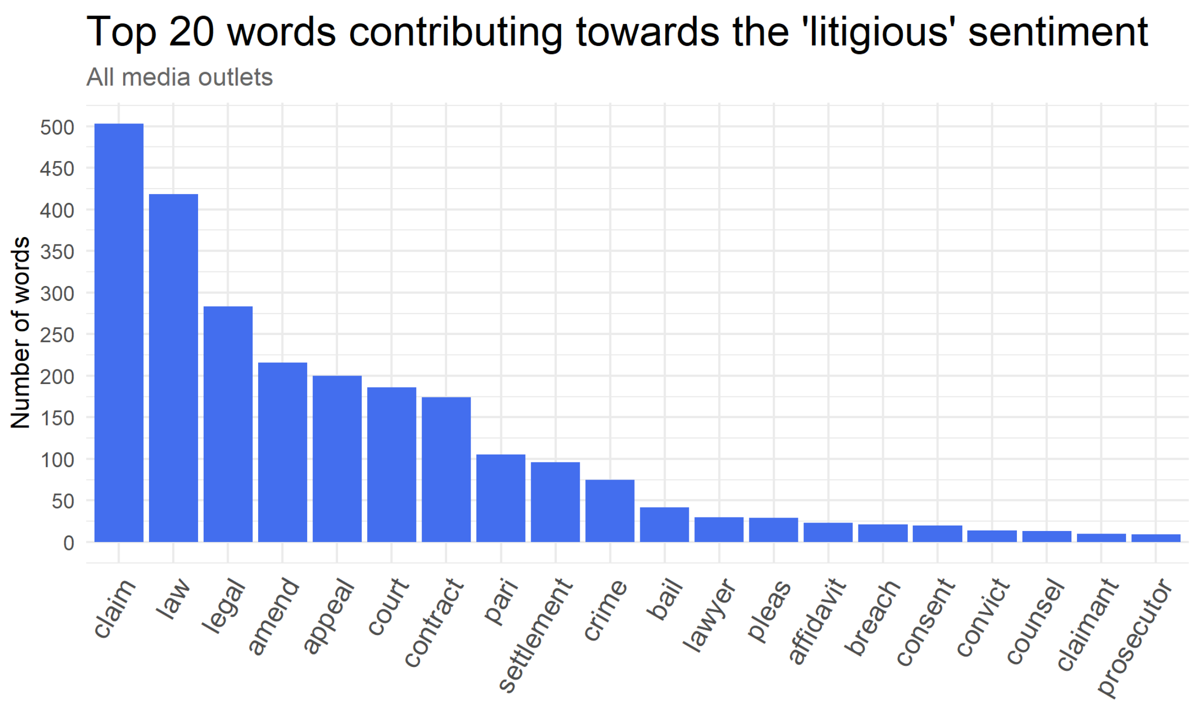

FIGURE 71: TOP CONTRIBUTING WORDS: LITIGIOUS: ALL MEDIA OUTLETS

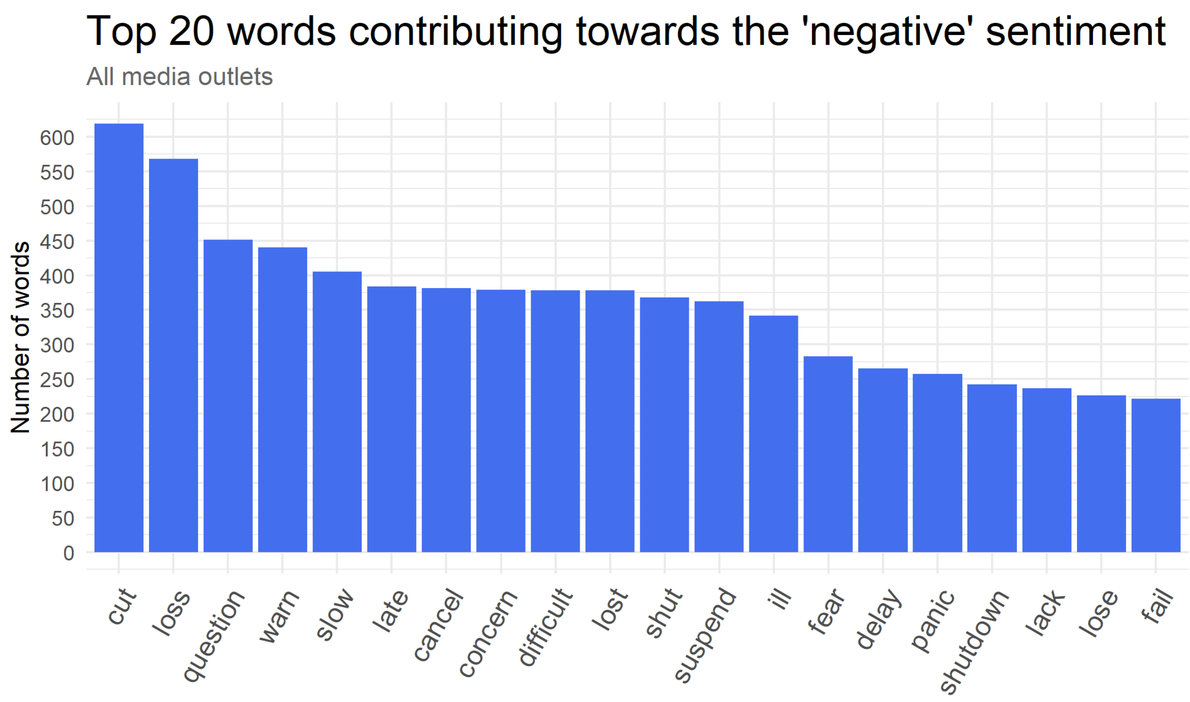

FIGURE 72: TOP CONTRIBUTING WORDS: NEGATIVE: ALL MEDIA OUTLETS

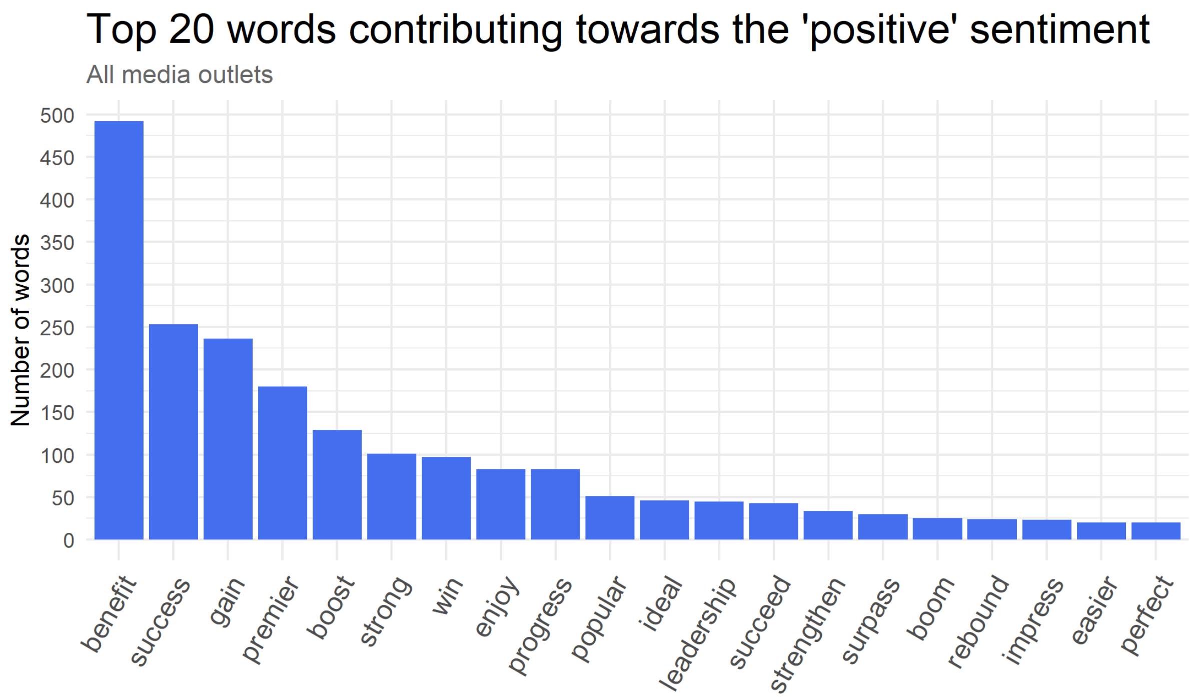

FIGURE 73: TOP CONTRIBUTING WORDS: POSITIVE: ALL MEDIA OUTLETS

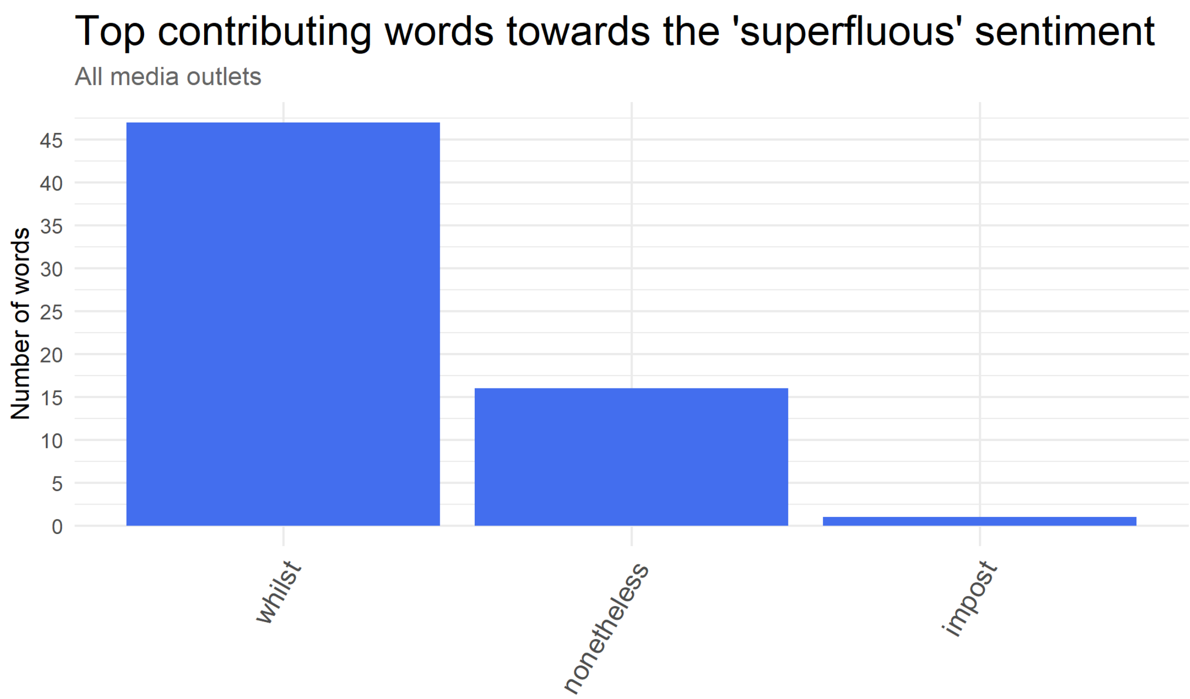

FIGURE 74: TOP CONTRIBUTING WORDS: SUPERFLUOUS: ALL MEDIA OUTLETS

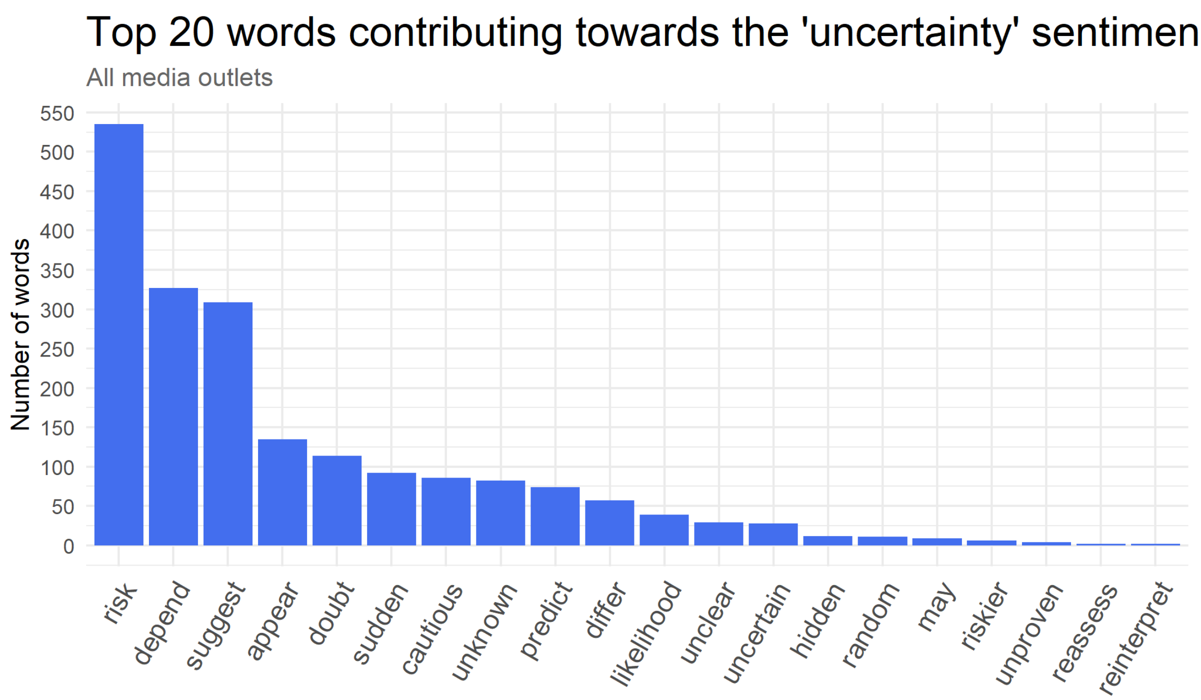

FIGURE 75: TOP CONTRIBUTING WORDS: UNCERTAINTY: ALL MEDIA OUTLETS

Word Sentiment Proportions Across All Articles

Figures 76 to 81 below present an explicit analysis of the word sentiment ratios across the surveyed media outlets. The ratio is calculated as the total proportion of words belonging to the sentiment category per media outlet over the total number of words per media outlet.

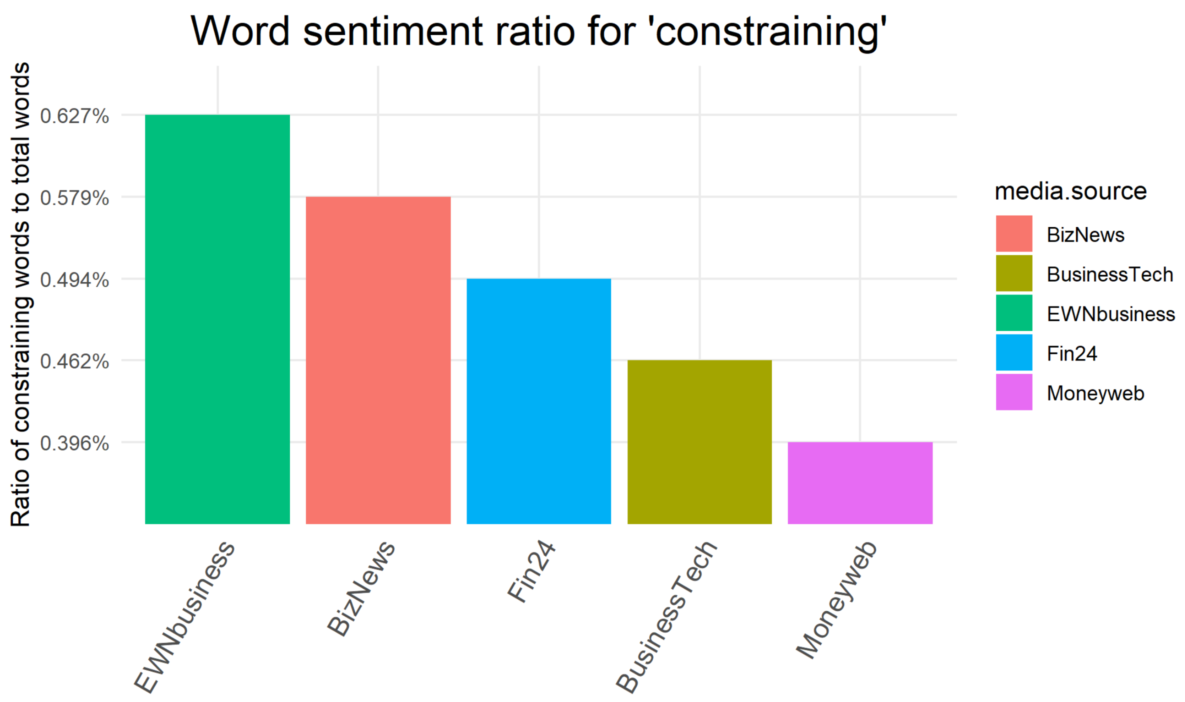

FIGURE 76: WORD SENTIMENT RATIO: CONSTRAINING

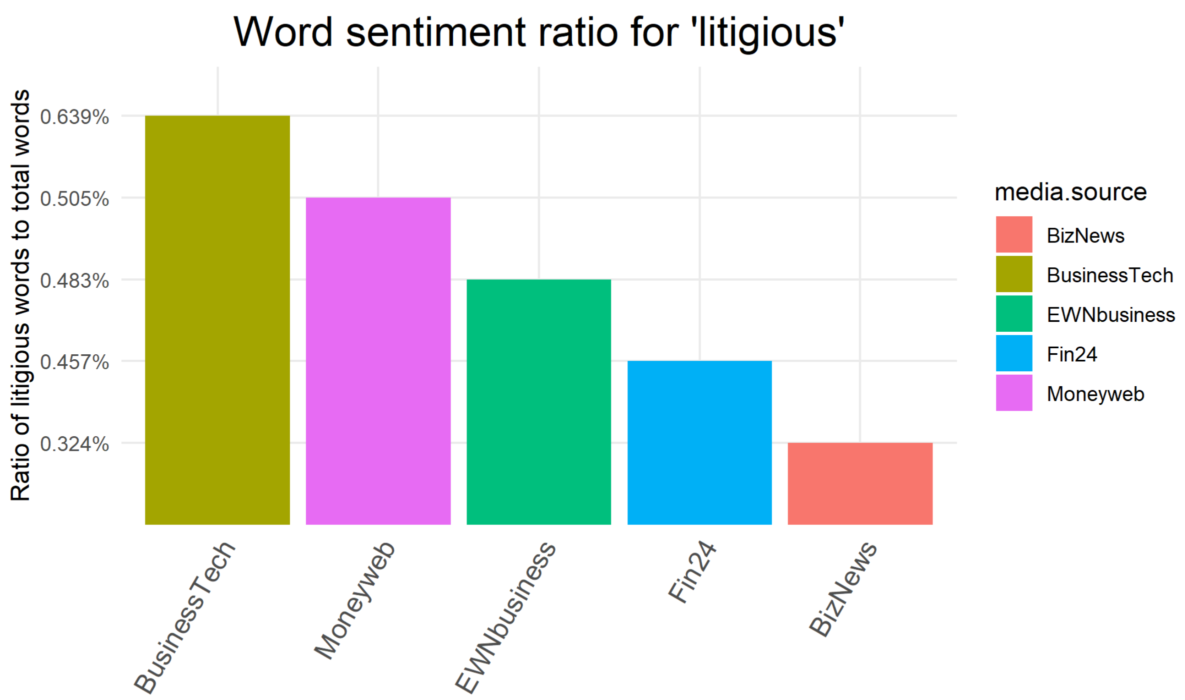

FIGURE 77: WORD SENTIMENT RATIO: LITIGIOUS

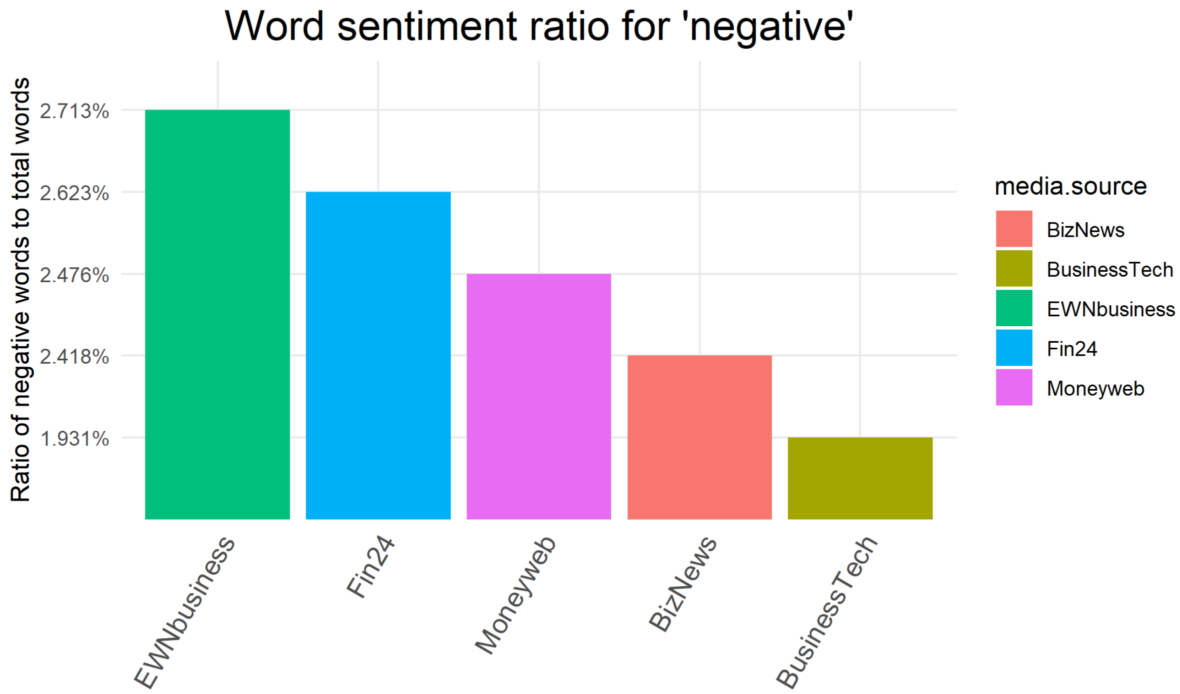

FIGURE 78: WORD SENTIMENT RATIO: NEGATIVE

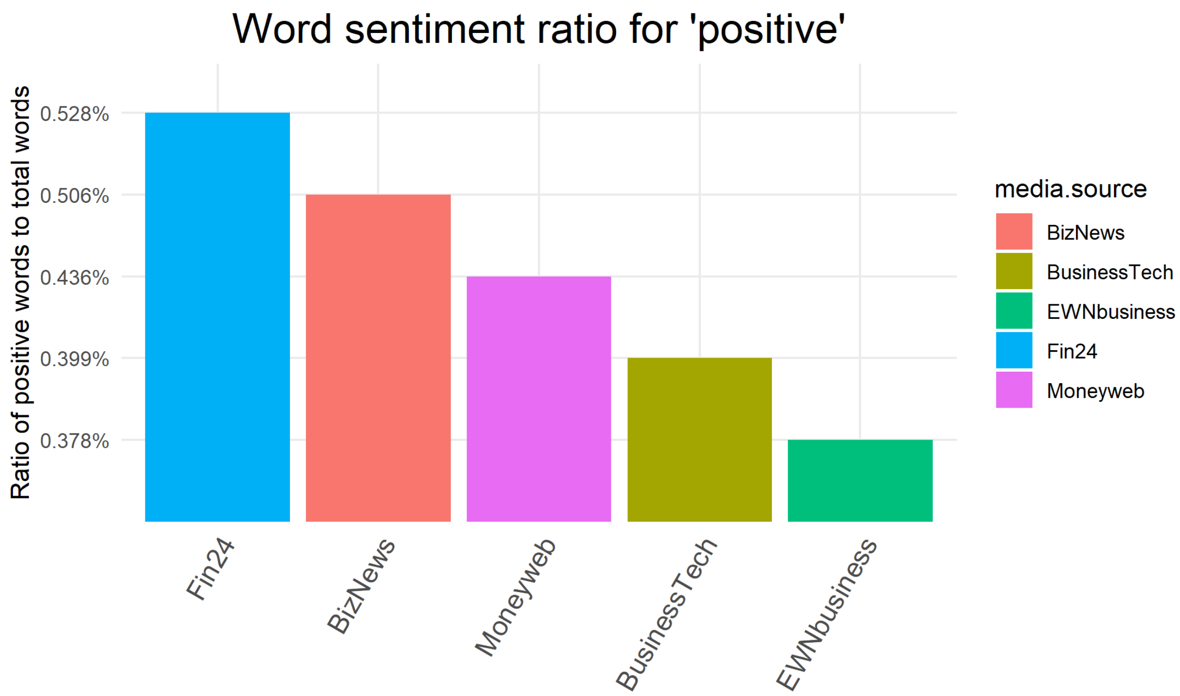

FIGURE 79: WORD SENTIMENT RATIO: POSITIVE

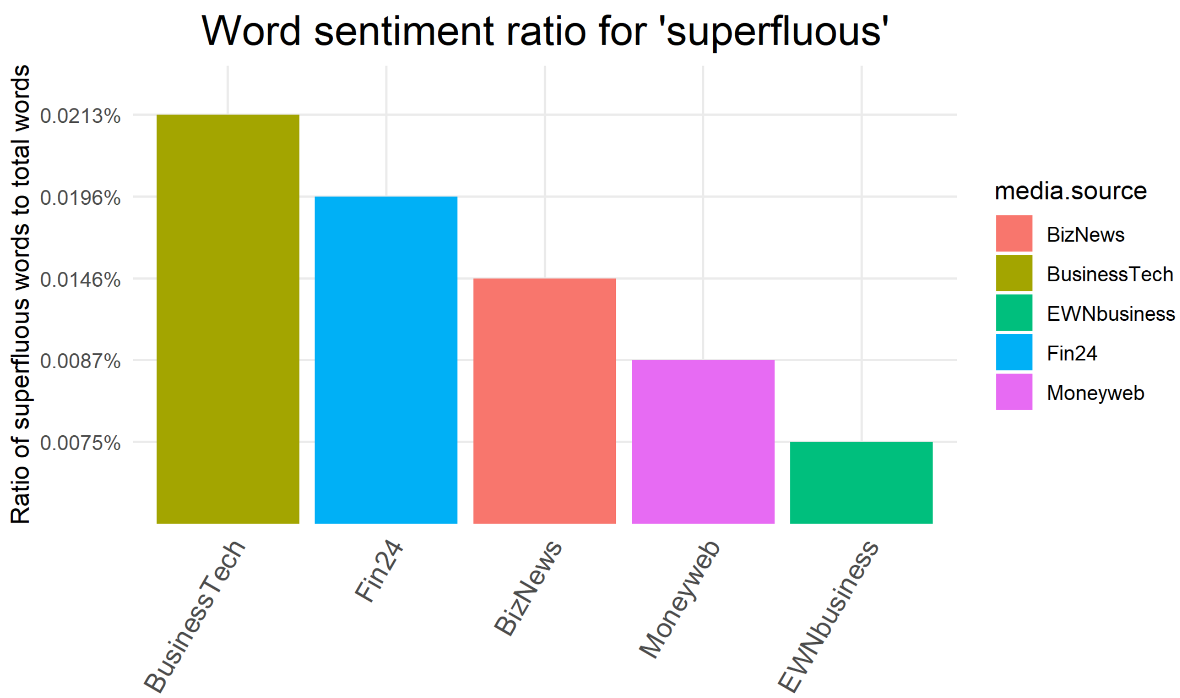

FIGURE 80: WORD SENTIMENT RATIO: SUPERFLUOUS

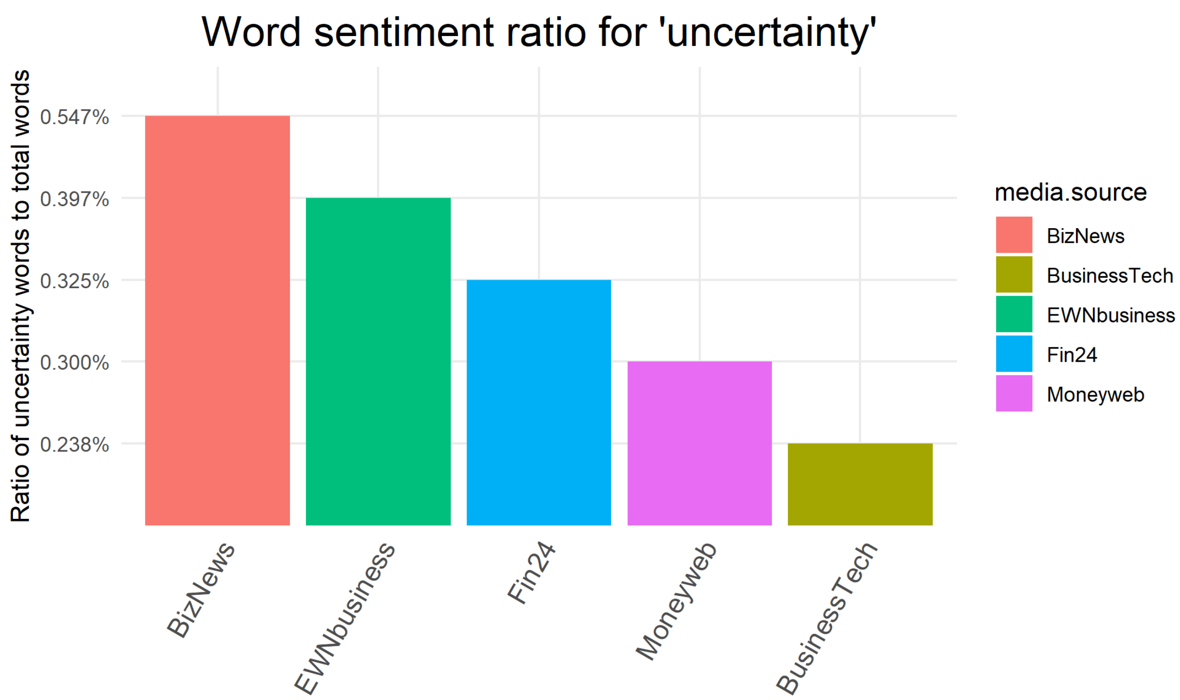

FIGURE 81: WORD SENTIMENT RATIO: UNCERTAINTY

Topic Modelling

Two methods of topic modelling were conducted on the news article datasets, namely structural topic modelling (STM) and Latent Dirichlet allocation (LDA). In order to conduct the topic modelling, the number of topics (K) for each dataset needed to be specified. These values were determined by using Ponweiser's Harmonic Mean method. Using the graphs outputted by the function, the estimated K-values for the datasets ranged from 6 to 15. These graphs were documented and can be viewed in the GitHub repository through the following file path: /images/topic_modelling/determining_k.

Structural Topic Modelling

Fin24

Across Fin24 articles, K was found to be 6. Using the beta method in Figure 82 below, these topics seems to centre around the following themes, namely (1) economic consequence, (2) prohibited activities, (3) general Covid-19 implications, (4) business impact, (5) societal effects, and (6) government involved industries.

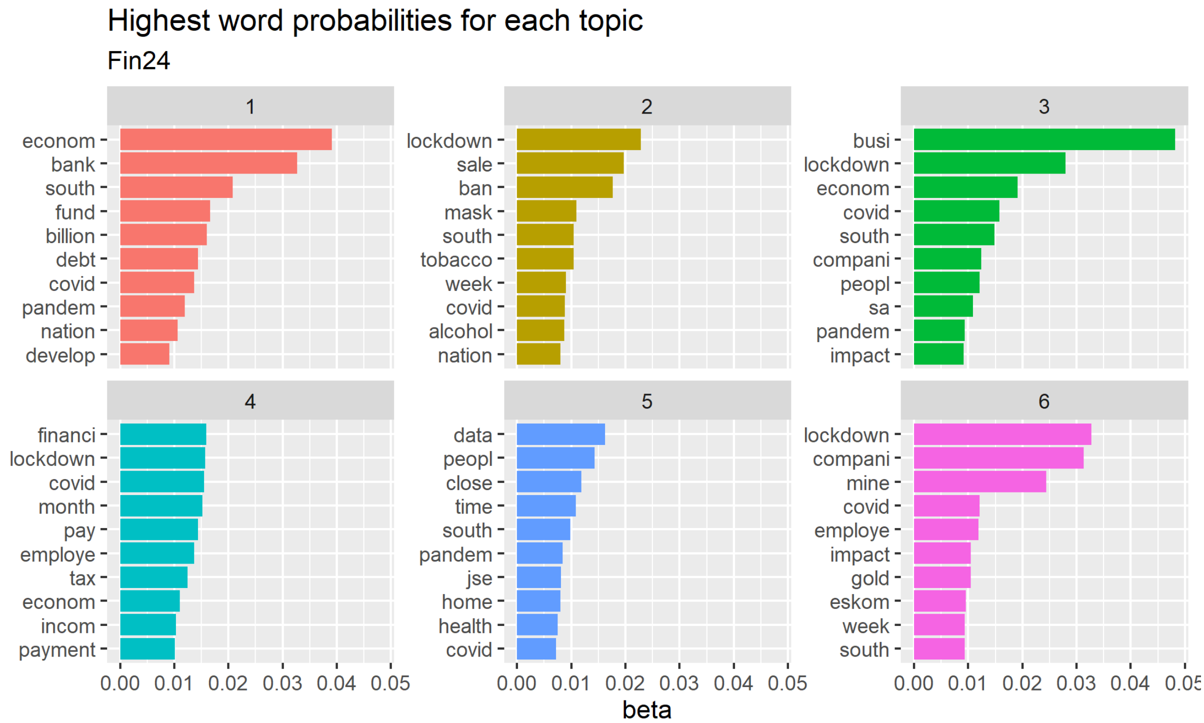

FIGURE 82: STM: BETA: FIN24

Figure 83, which depicts the topics modelled on a gamma matrix, indicates that each article in the dataset is strongly related to a single topic, however, there is a small probability that an article may cross-reference another topic. This is to be expected since the articles in question all relate to topics within the scope of the coronavirus pandemic and are predominantly economically oriented.

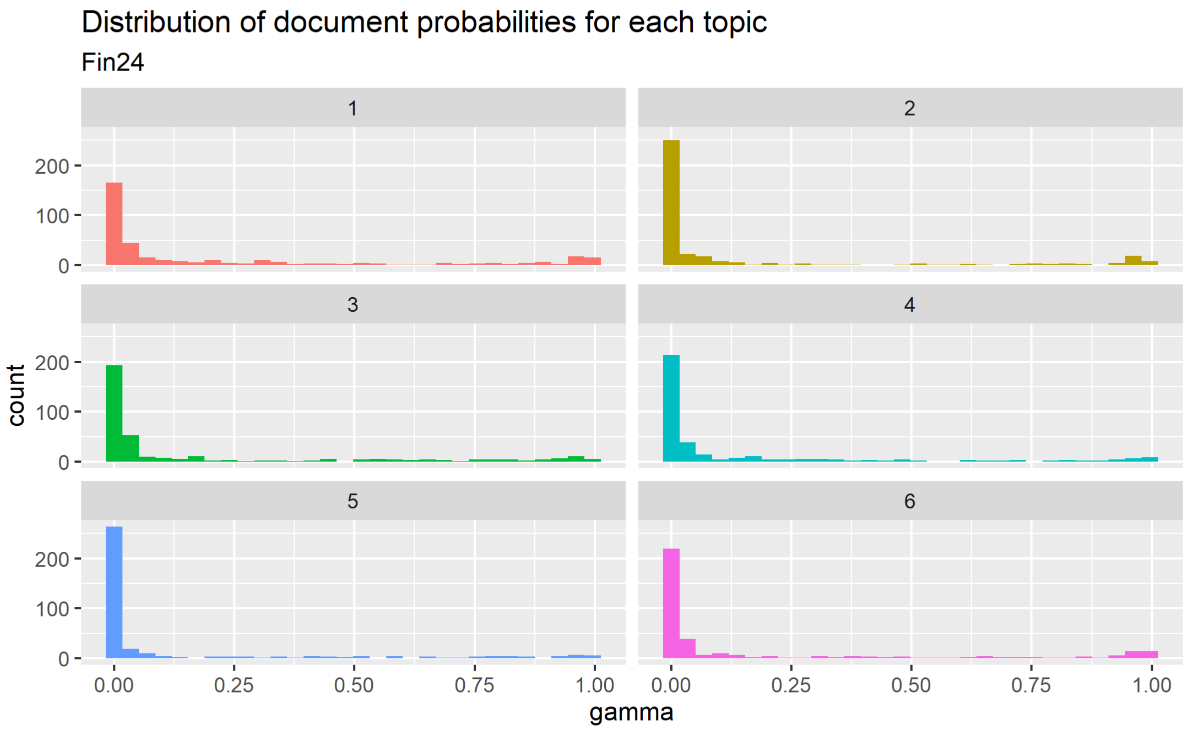

FIGURE 83: STM: GAMMA: FIN24

Moneyweb

Across Moneyweb articles, K was found to be 15. Using beta in figure 84 below, some of these emerging topics focus of themes pertaining to (3) general Covid-19 implications, (4) financial consequence, (7) stock markets, (9) healthcare, (10) locations and permissions, (8, 11) salaries and payments, (13) economic consequence, (14) and government-associated industries.

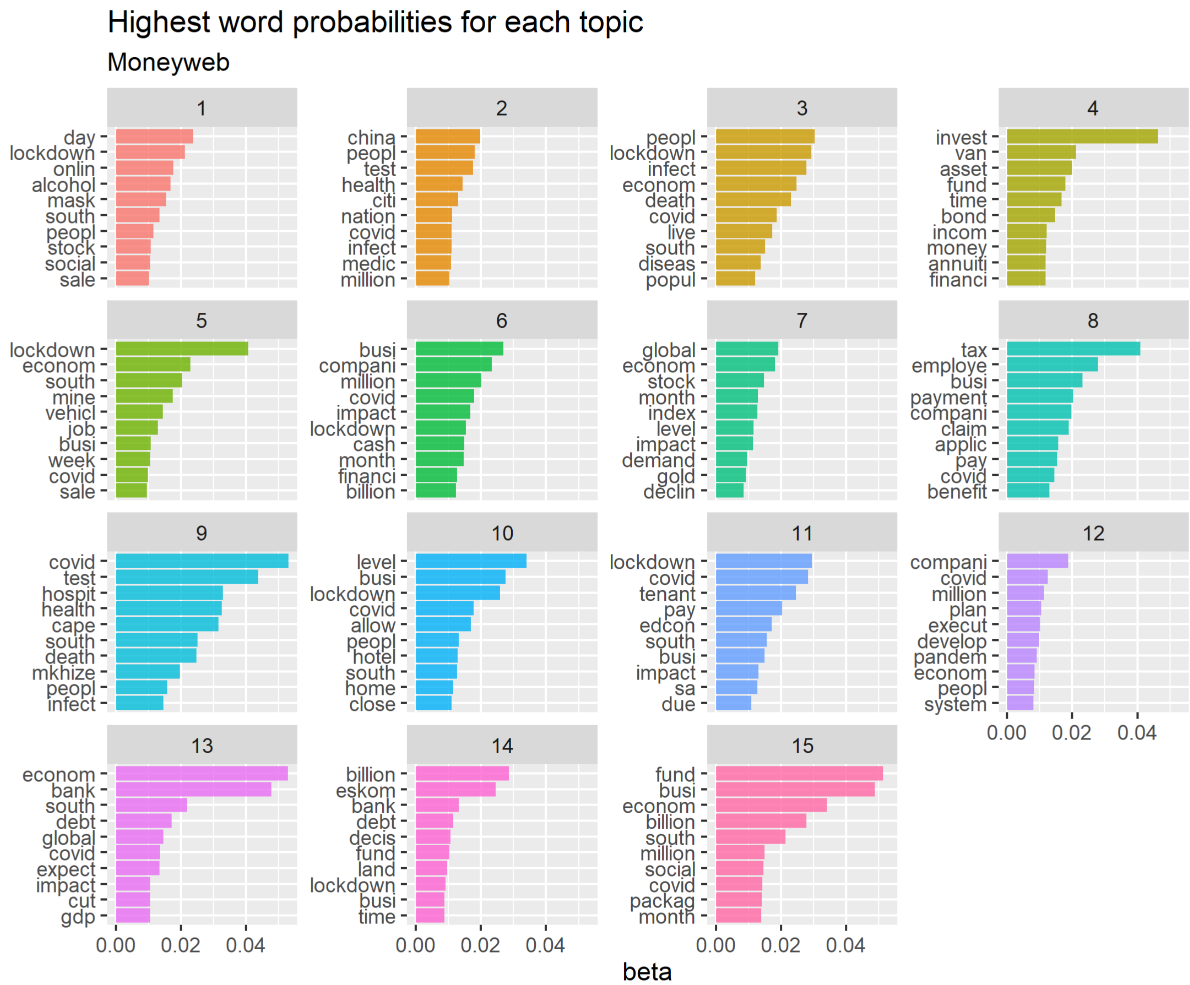

FIGURE 84: STM: BETA: MONEYWEB

Figure 85 below, the topics modelled in a gamma matrix, indicate that each article in the dataset is strongly related to a single topic, which was to be expected given the underlying topic of Covid-19.

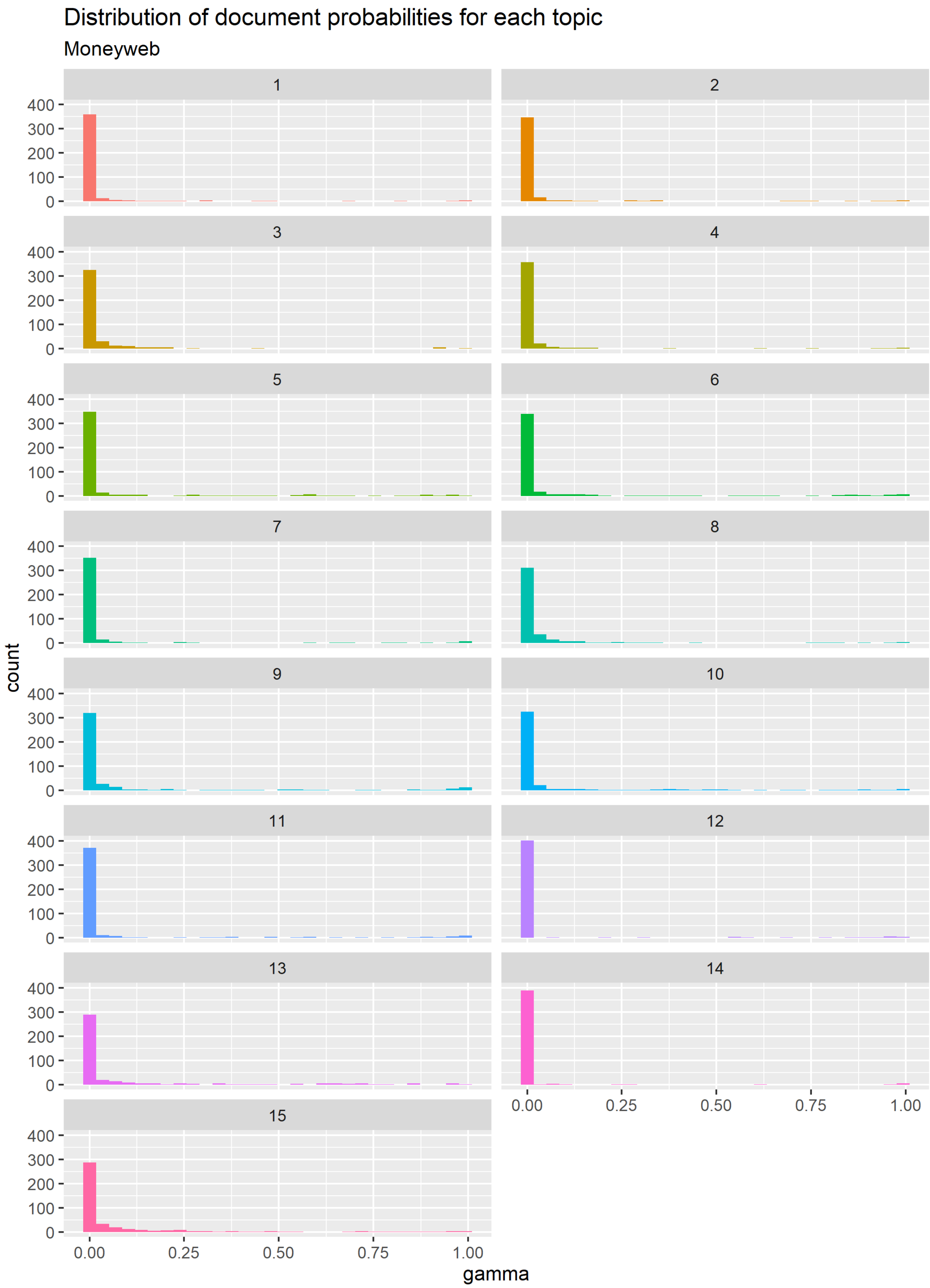

FIGURE 85: STM: GAMMA: MONEYWEB

BizNews

Across BizNews articles, K was found to be 8. Using beta in Figure 86 below, the emerging topics seem broader in nature and primarily pertain to themes of (10) health care, (2) general reporting, (3) recreational activities, (4) the economy, (5) society, (6) public institutions, (7) the virus's spread, and (8) political impacts.

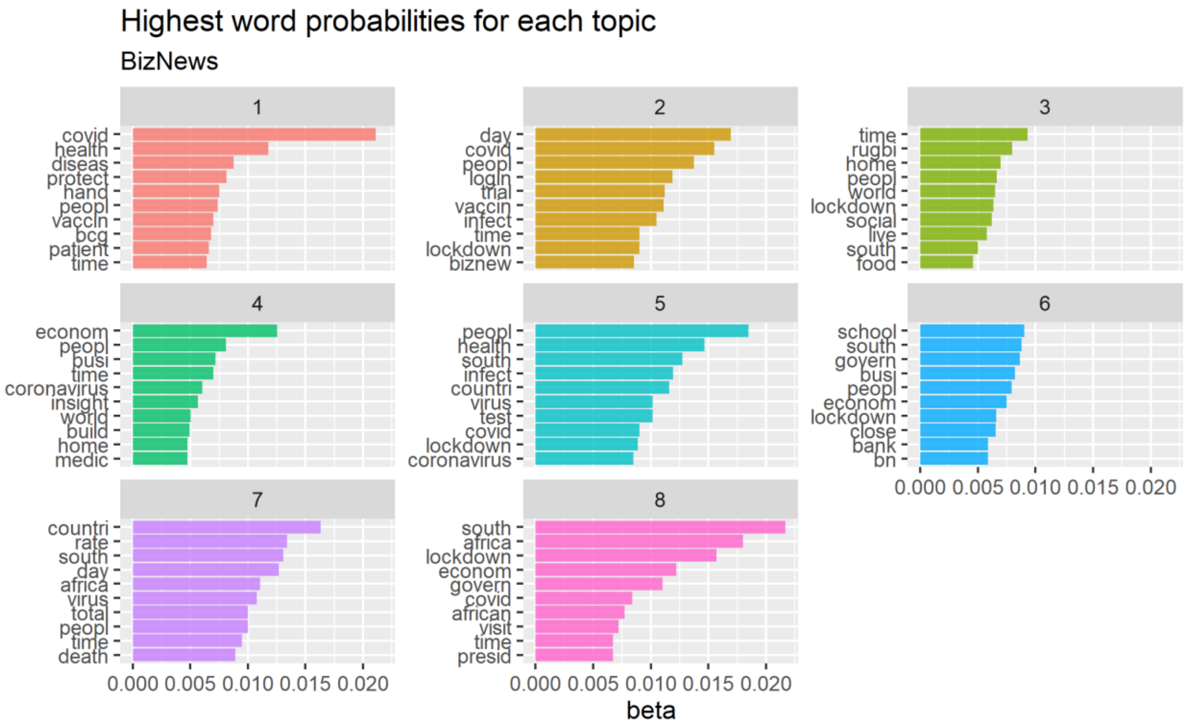

FIGURE 86: STM: BETA: BIZNEWS

The visualisation of the distribution of document probabilities for each topic for BizNews can be seen in Figure 87 below. BizNews articles are also strongly associated with a single topic, however, topic 2 has an increased likelihood to have articles that span over multiple topics.

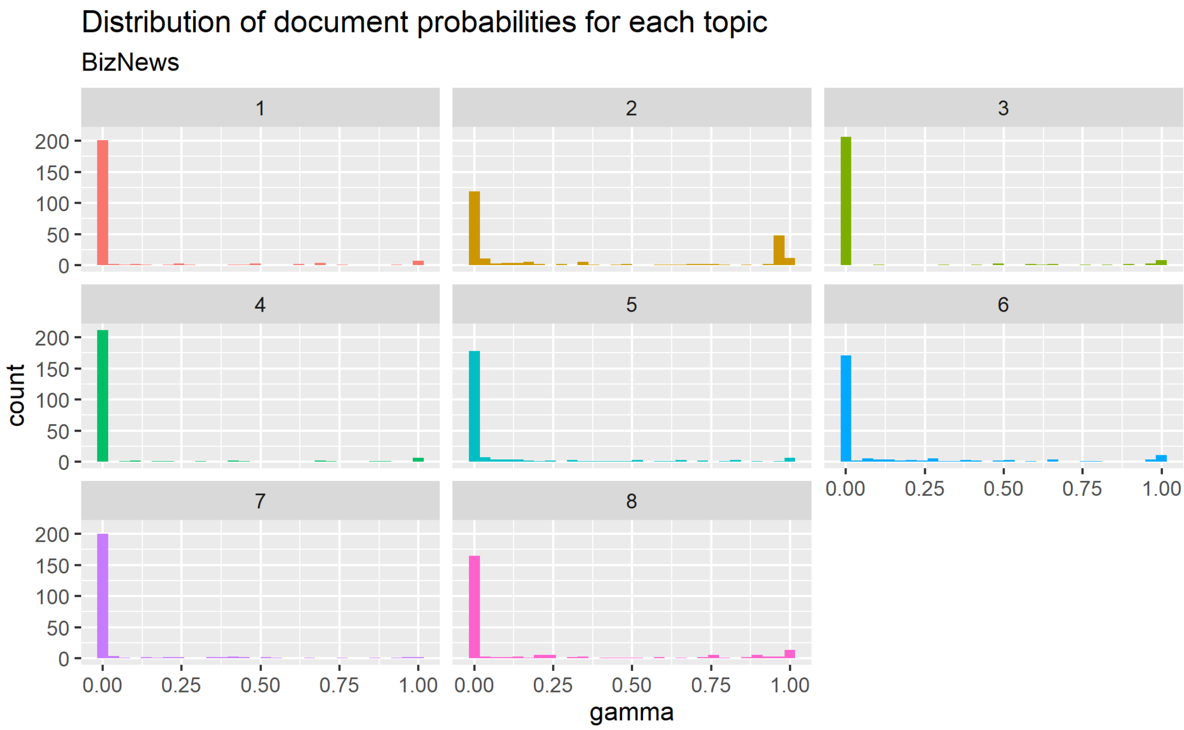

FIGURE 87: STM: GAMMA: BIZNEWS

EWN Business

It should be noted that instead of using K = 13 to model topics for the EWN Business dataset, it was made that K = 10 due to having insufficient onboard memory and processing capacity for the dataset.

Some of the topics identified include (3) health, (4) education, (7) events, (9) spread, and (10) international affairs, which can be seen in Figure 87 below. The heavy focus on international events relates to the previous inference that EWN Business is geared more towards international affairs in comparison to the other outlets.

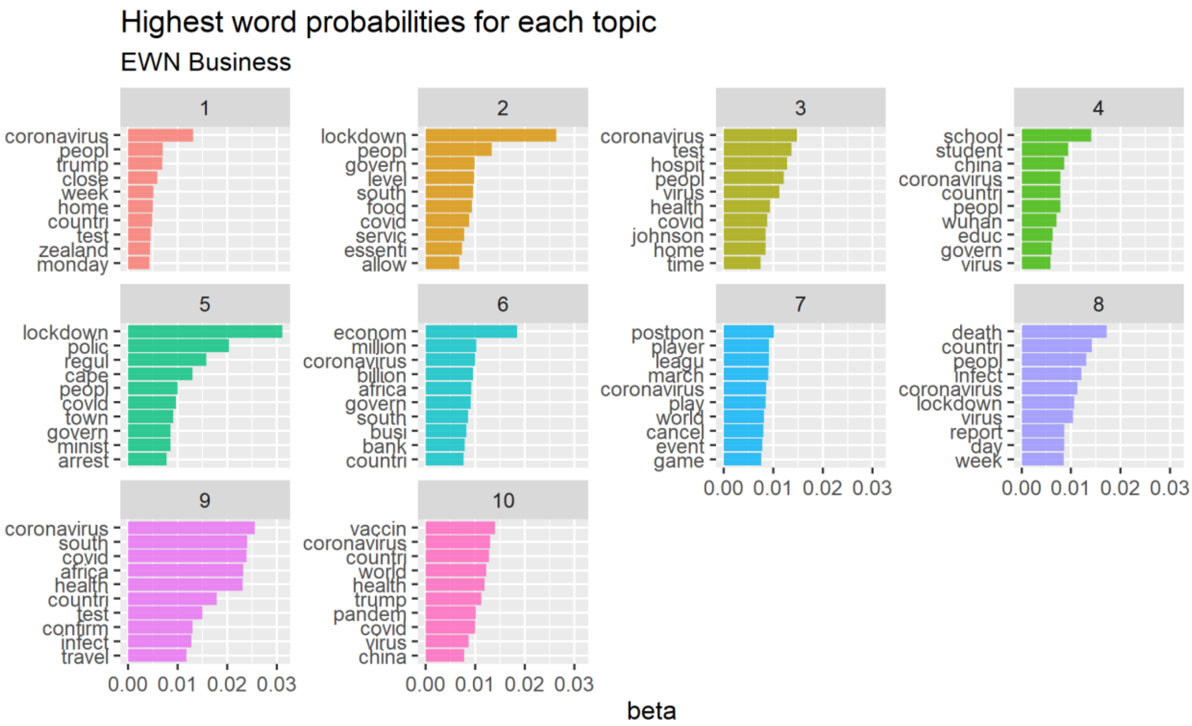

FIGURE 88: STM: BETA: EWN BUSINESS

As depicted in Figure 89 below, the articles published by EWN Business show a strong association with a specific topic, particularly that of topics 1, 3, 4, and 7. Each of these topics makes reference to either international affairs or event cancellations, and one can infer that articles written about these topics are very topic-specific.

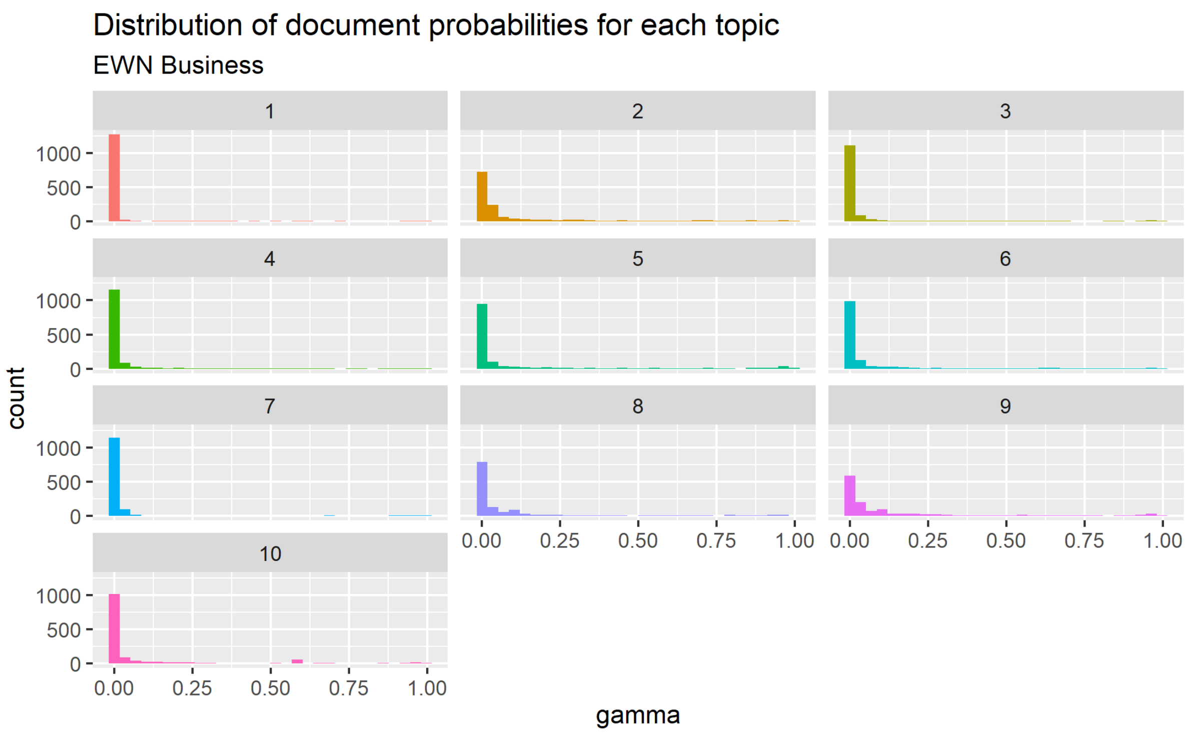

FIGURE 89: STM: GAMMA: EWN BUSINESS

BusinessTech

Across BusinessTech articles, K was found to be 13. Using beta in Figure 90 below, topics pertaining to the following themes were identified, namely, (2) legislation, (3) finances, (7) health care, (8) education, (9) employment, (10) restrictions, (12) political action, and (13) the economy.

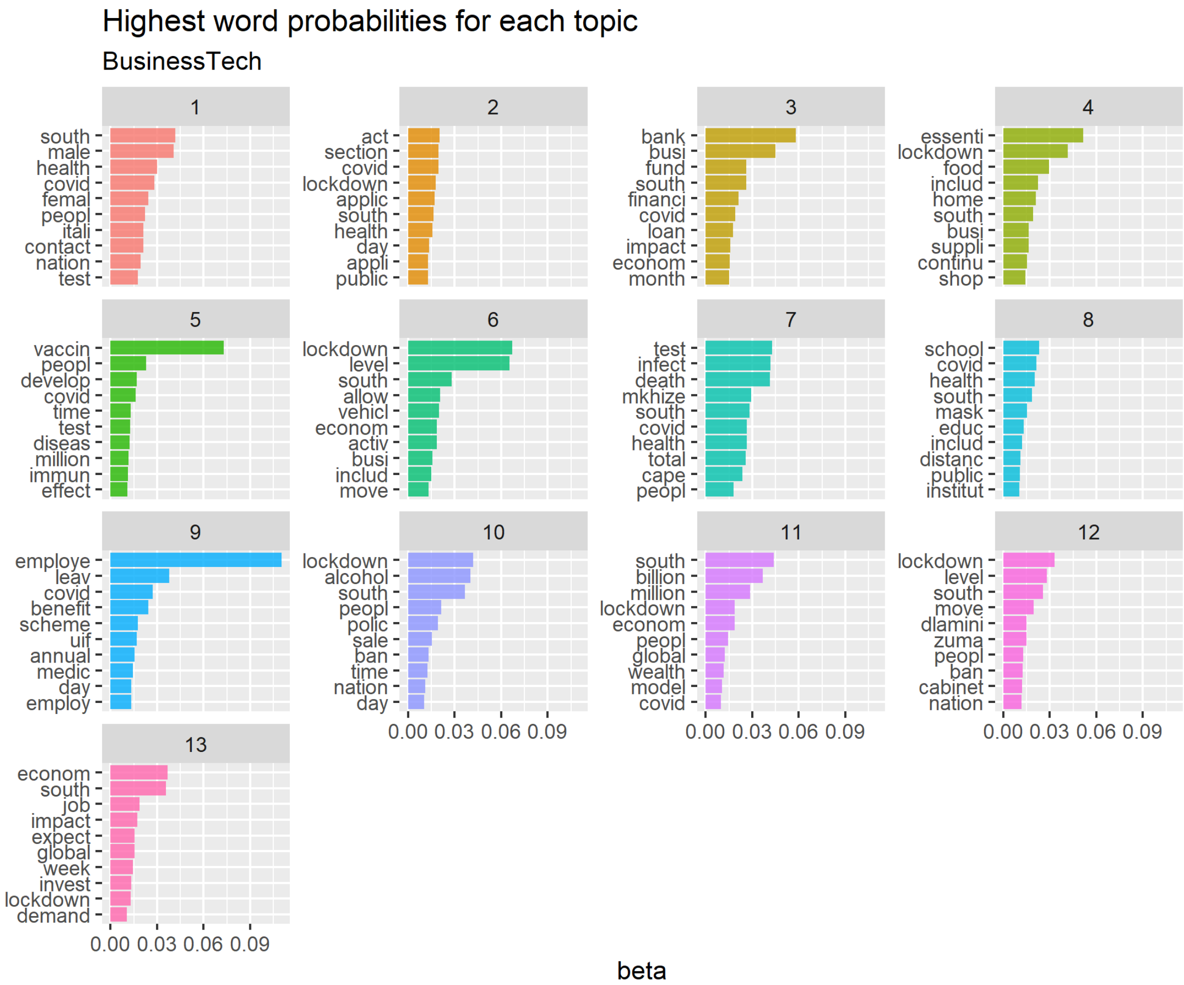

FIGURE 90: STM: BETA: BUSINESSTECH

BusinessTech's distribution of document probabilities for each topic is similar to those of the previously observed news outlets in that each of its published articles can be strongly associated with a single topic. This can be viewed in Figure 91 below.

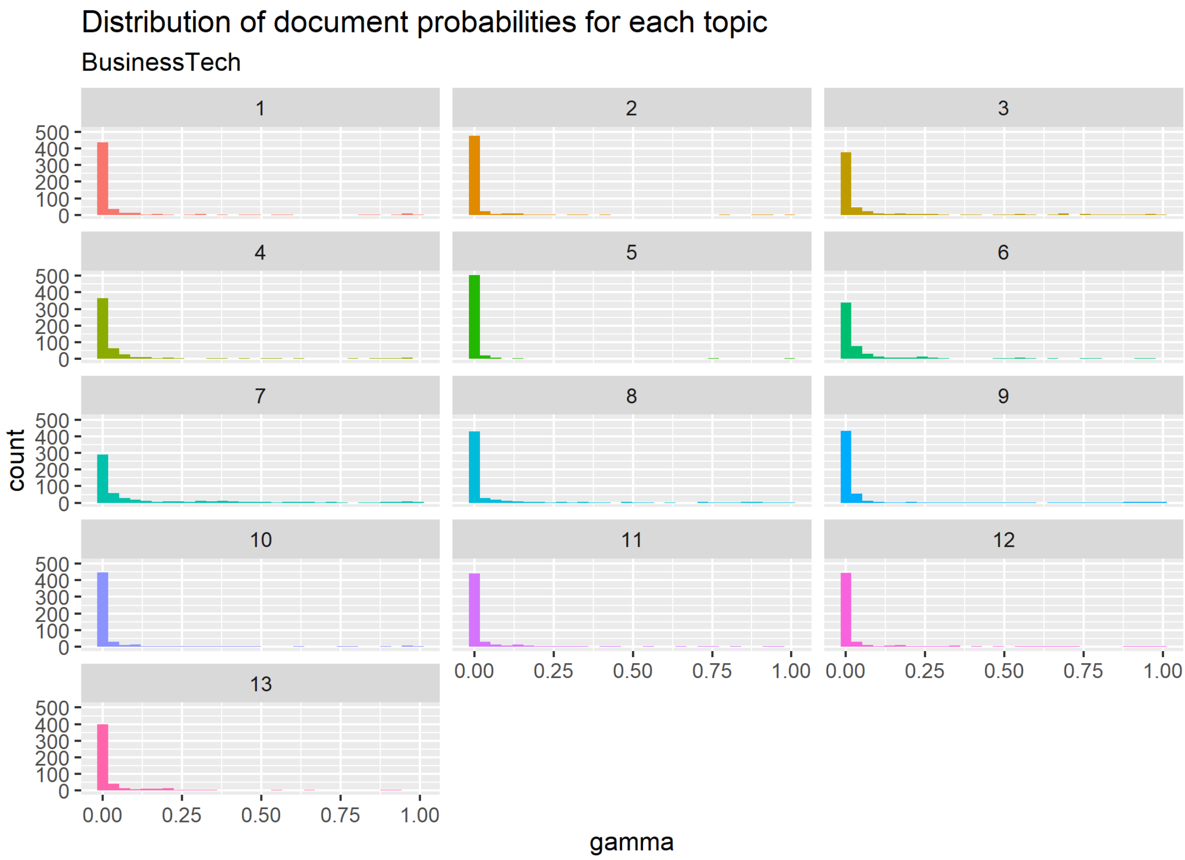

FIGURE 91: STM: GAMMA: BUSINESSTECH

Across All Media Outlets

Across all media outlets, 11 topics were identified overall. These topics focus on (6, 11) the economy, (8) education, (4) employment, (2, 3) health care, and (5) regulation.

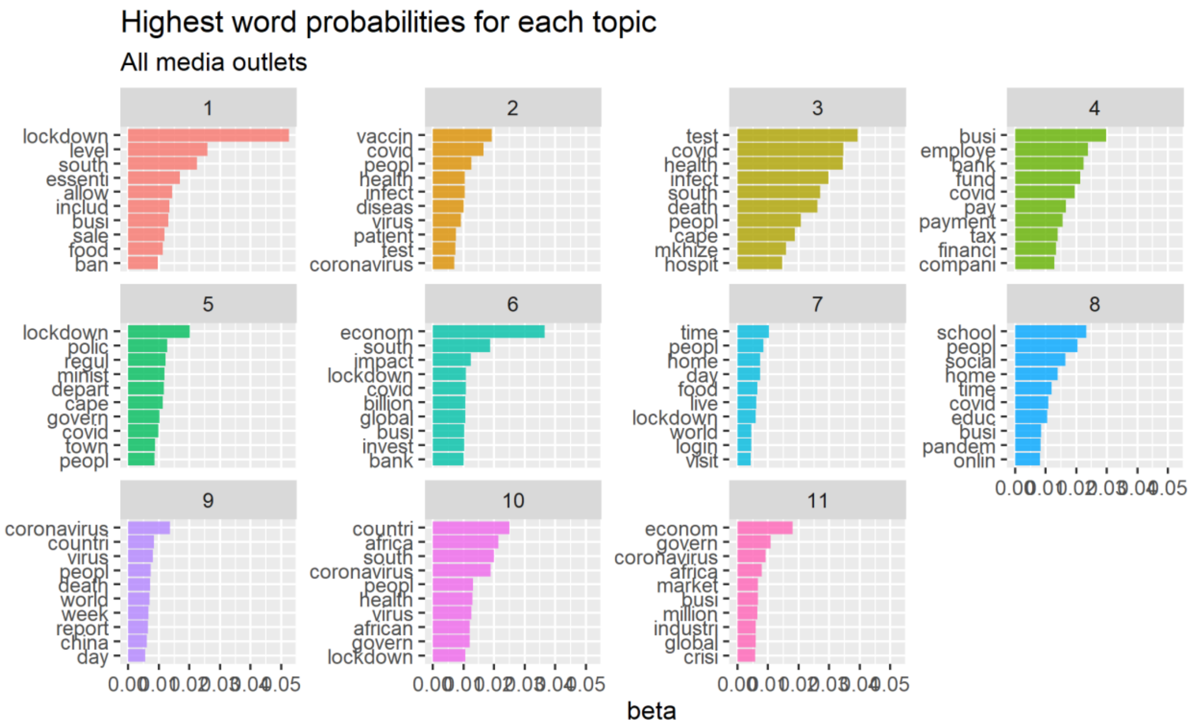

FIGURE 92: STM: BETA: ALL MEDIA OUTLETS

When studying the identified topics for all articles across the media outlets on a gamma matrix, it is evident that topics 2, 5, 7, 9, and 11 contain a larger number of topic-specific documents. The remaining topics have fewer articles addressing them, however, these articles are still relatively topic-specific.

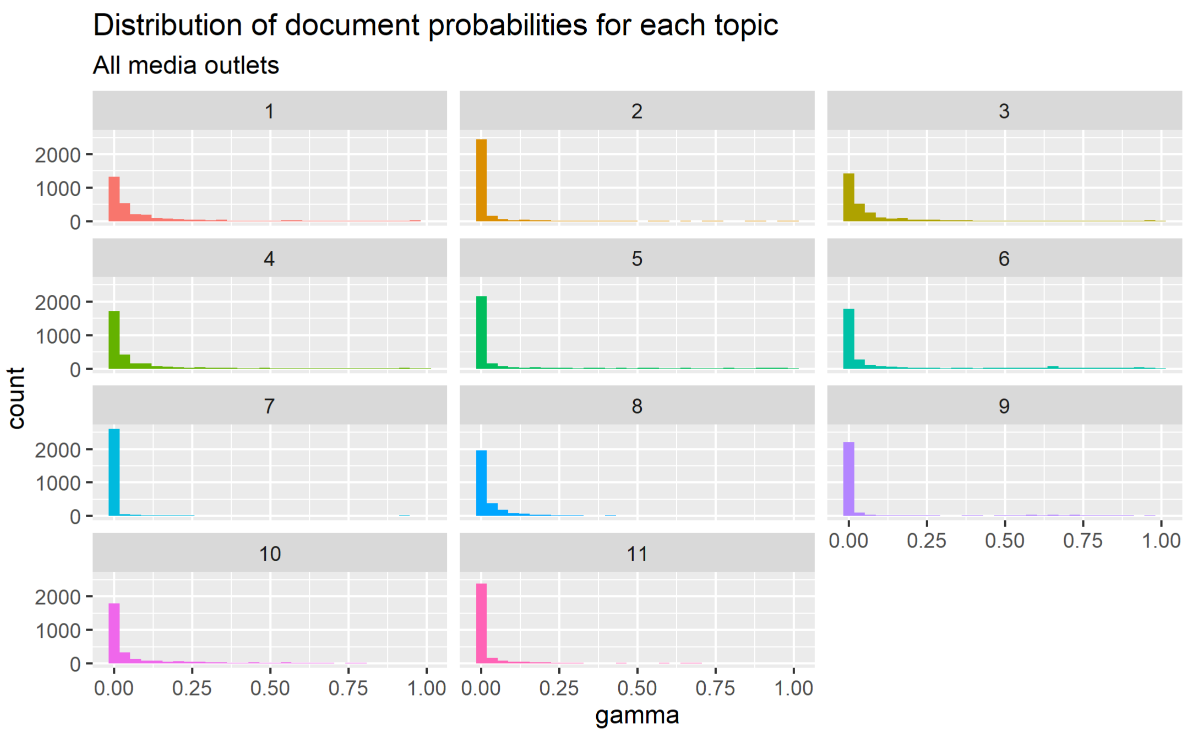

FIGURE 93: STM: GAMMA: ALL MEDIA OUTLETS

Latent Dirichlet Allocation

The LDA topic modelling was conducted using the same K-values, except in the case of EWN Business where the suggested K = 13 was used. The most prevalent words for each topic can be seen in Figures 94 to 99 below.

Through inspection, the topics produced by LDA seem to be more general compared to those produced by STM, although consistent with the themes represented.

Fin24

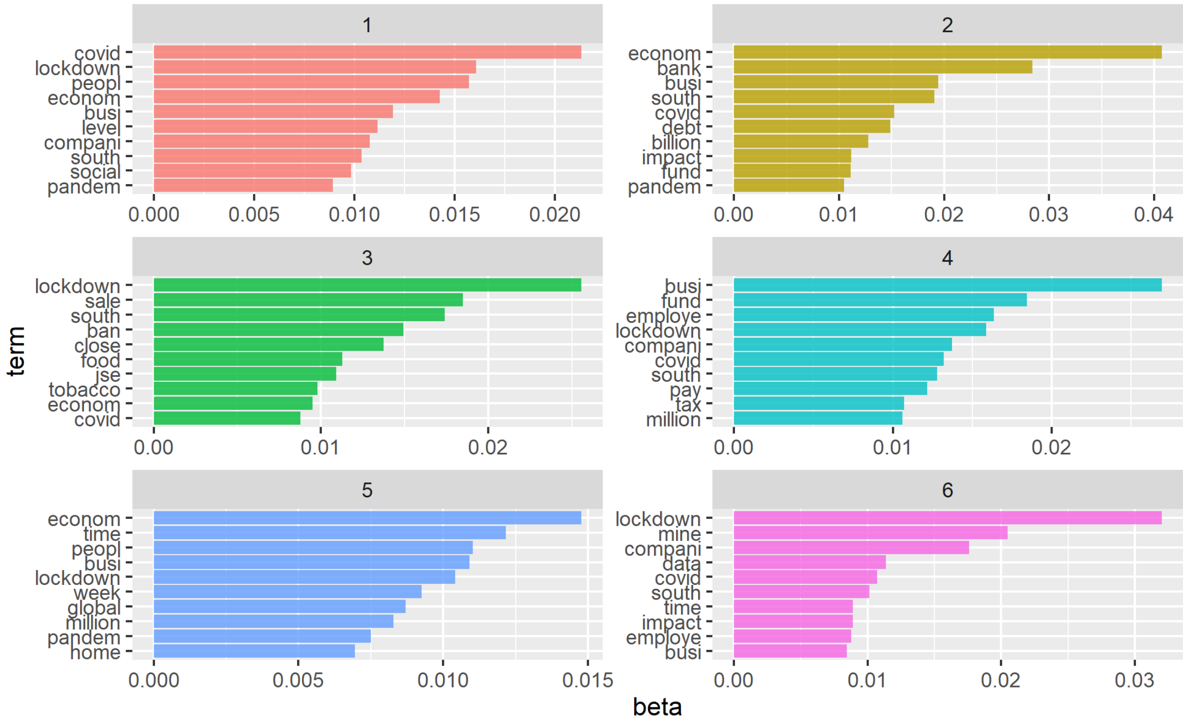

FIGURE 94: LDA: FIN24

Moneyweb

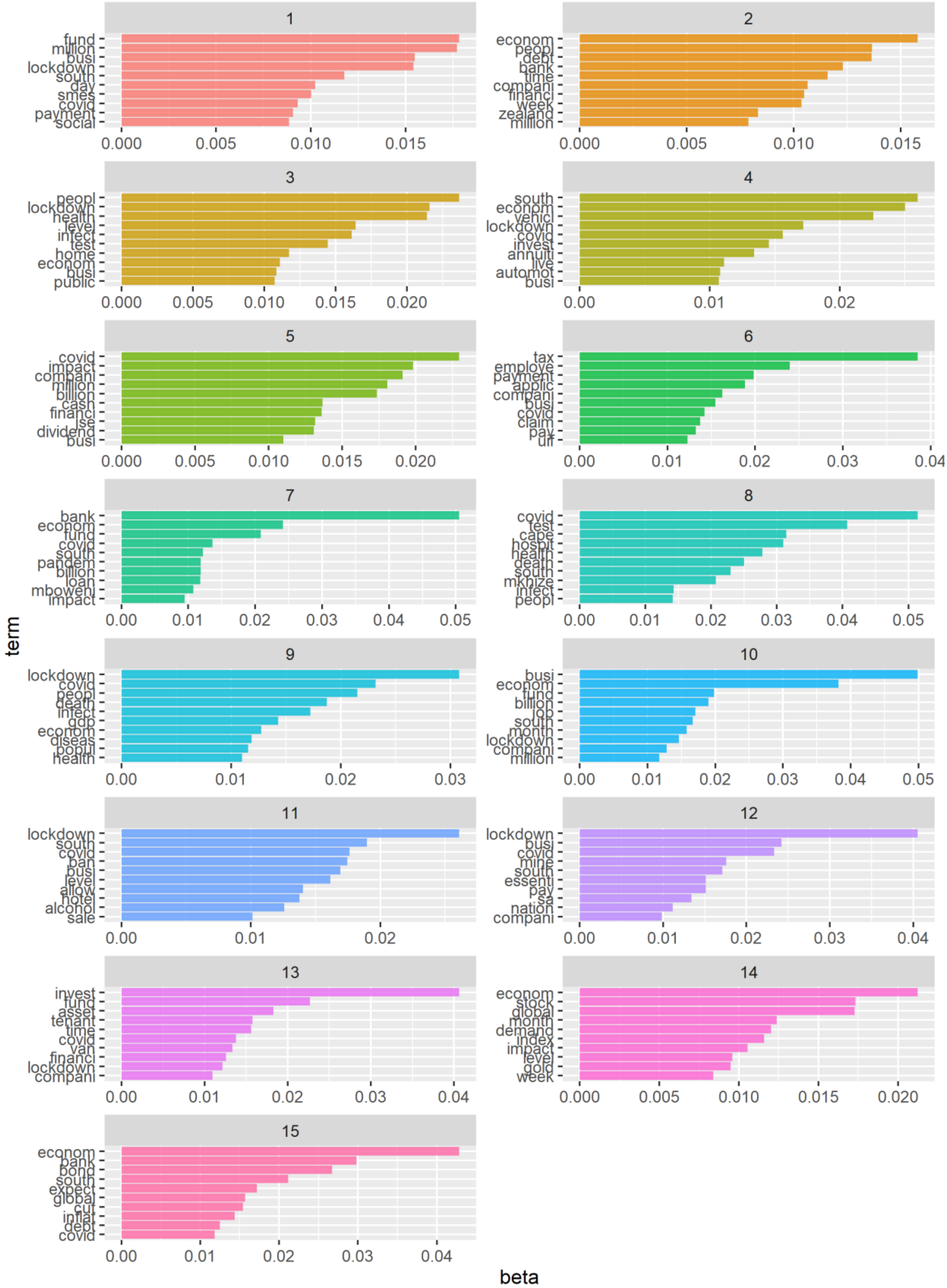

FIGURE 95: LDA: MONEYWEB

BizNews

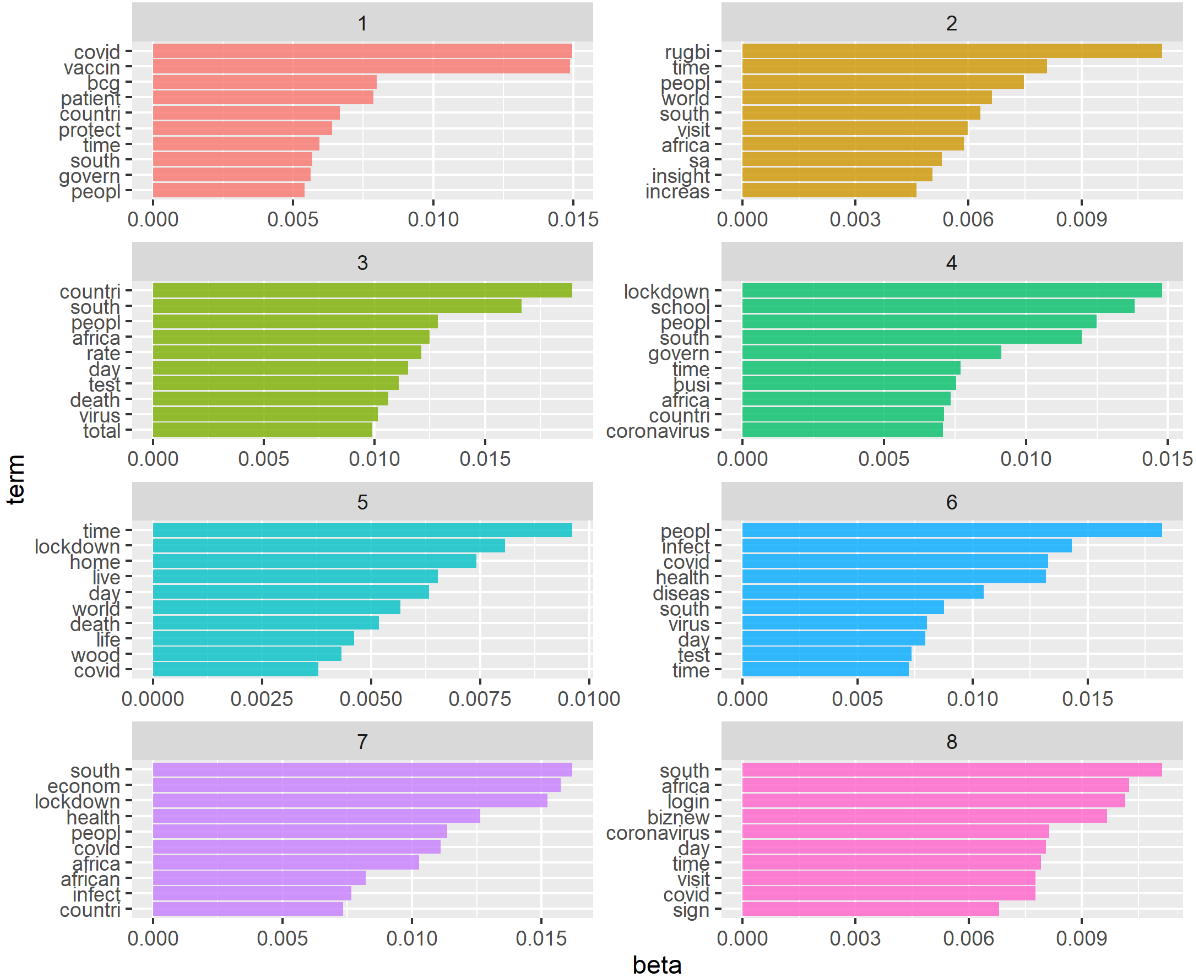

FIGURE 96: LDA: BIZNEWS

EWN Business

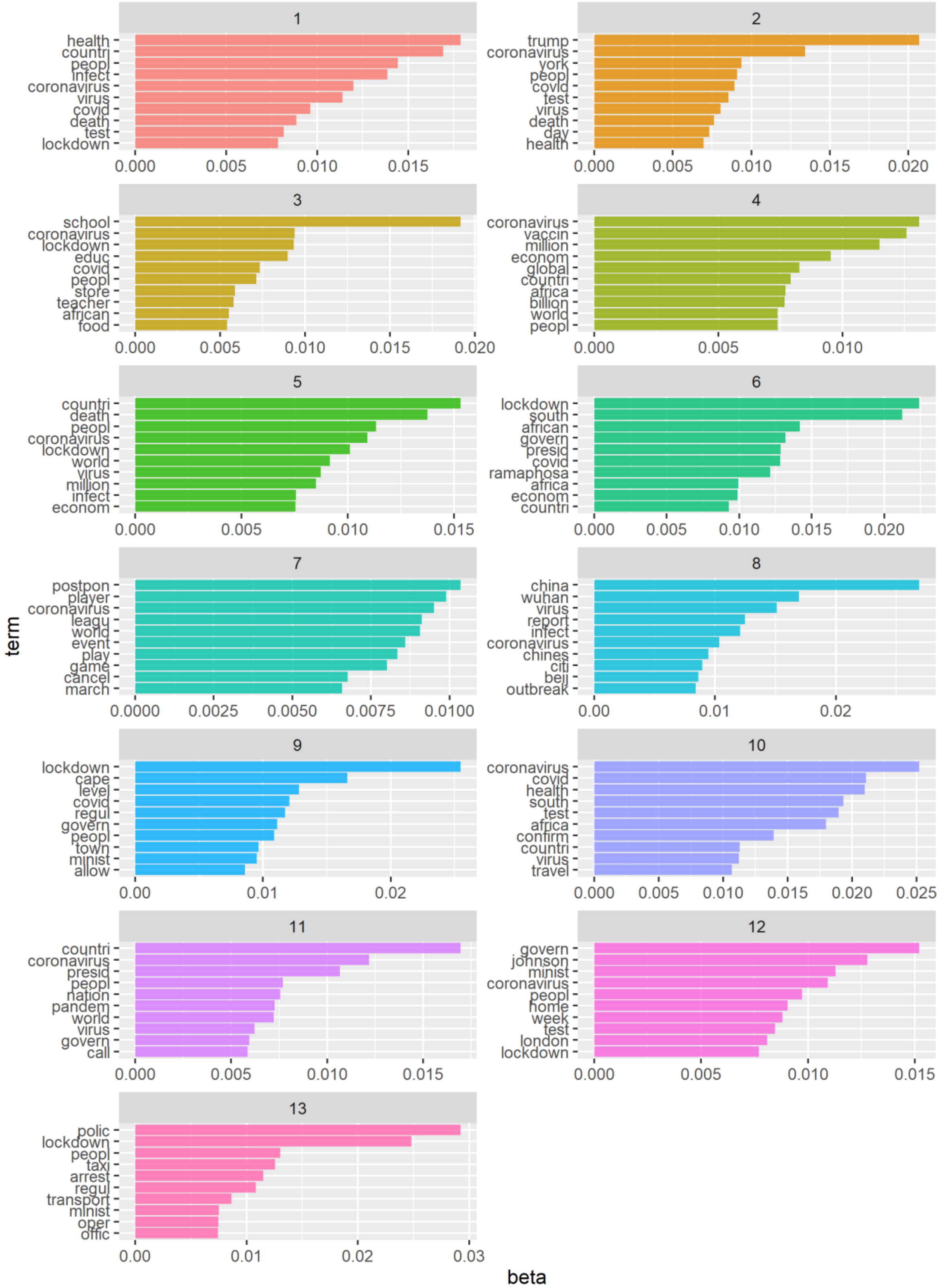

FIGURE 97: LDA EWN BUSINESS

BusinessTech

FIGURE 98: LDA BUSINESSTECH

Across All Media Outlets

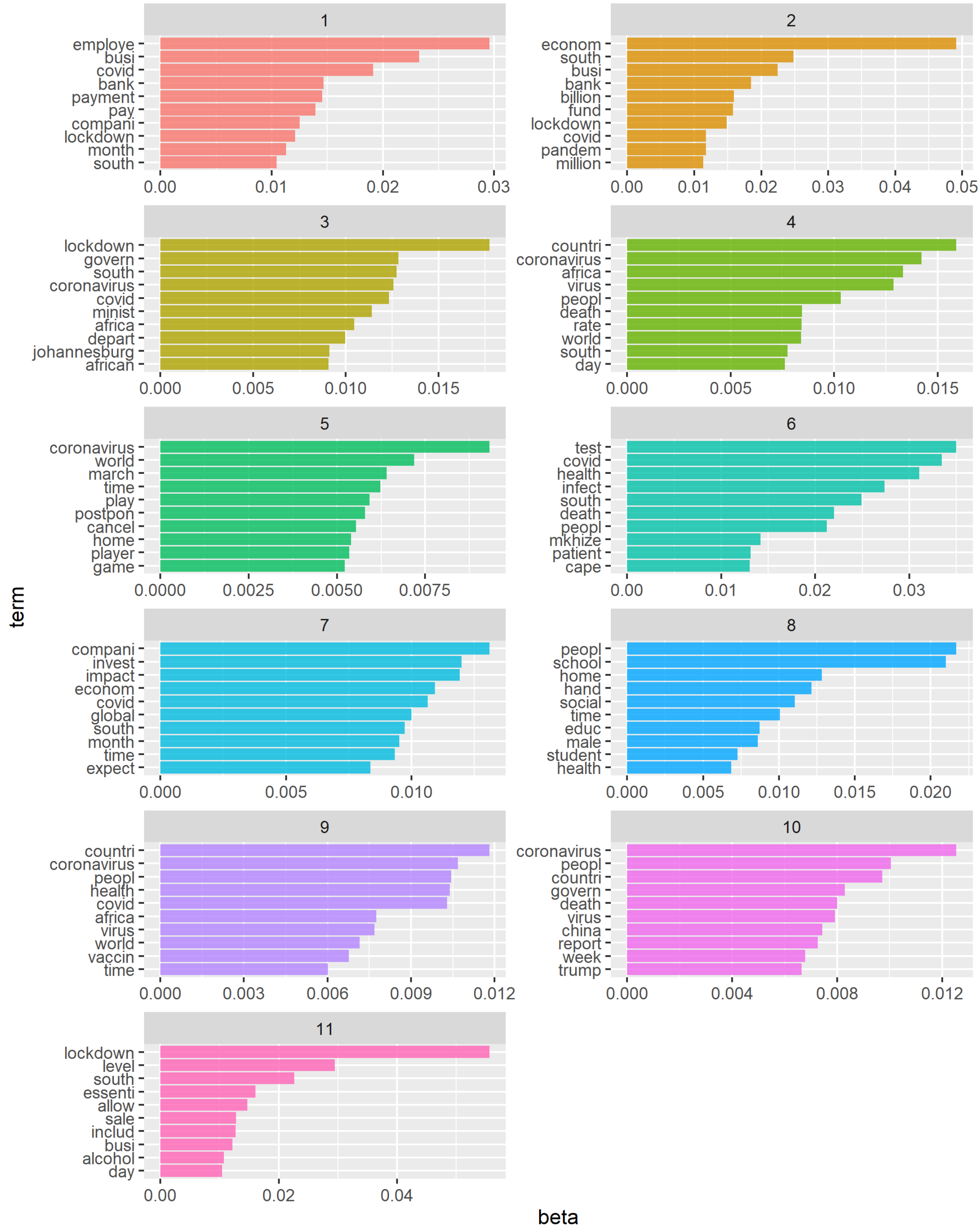

FIGURE 99: LDA: ALL MEDIA OUTLETS

Topic Sentiment Modelling

Using the beta generated by the LDA topic models, sentiment analysis was conducted on each set of topics for the media outlets and the combined dataset. The results of this analysis can be seen in Figures 100 to 105 below. Similar to the results observed in the sentiment analysis that was conducted on the datasets, the sentiment of the topics is decidedly negative. Moreover, similarities can be observed when comparing the other sentiments across the datasets and the topic models for the appropriate media outlet.

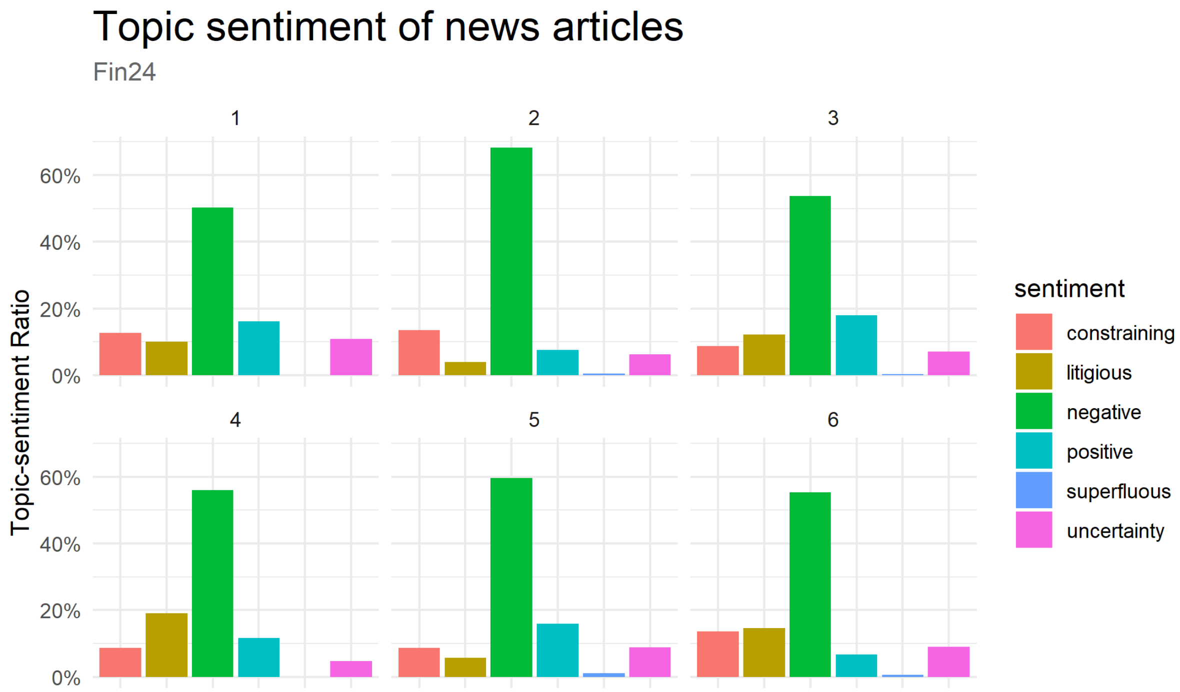

FIGURE 100: TOPIC SENTIMENT: FIN24

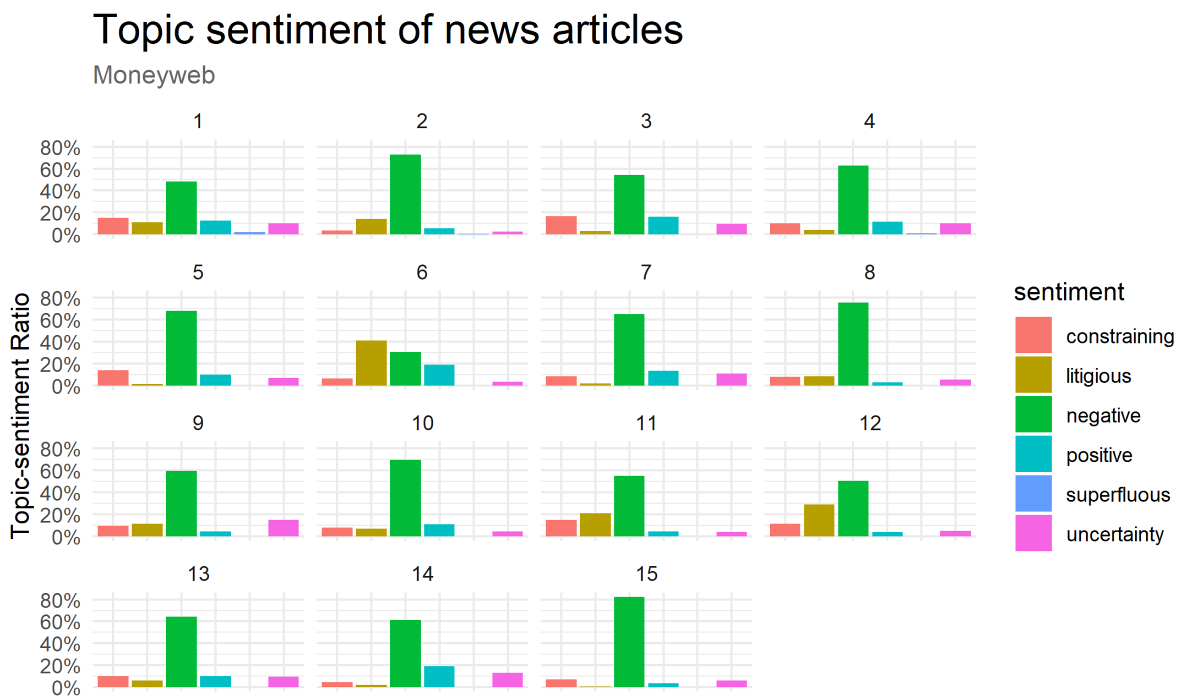

FIGURE 101: TOPIC SENTIMENT: MONEYWEB

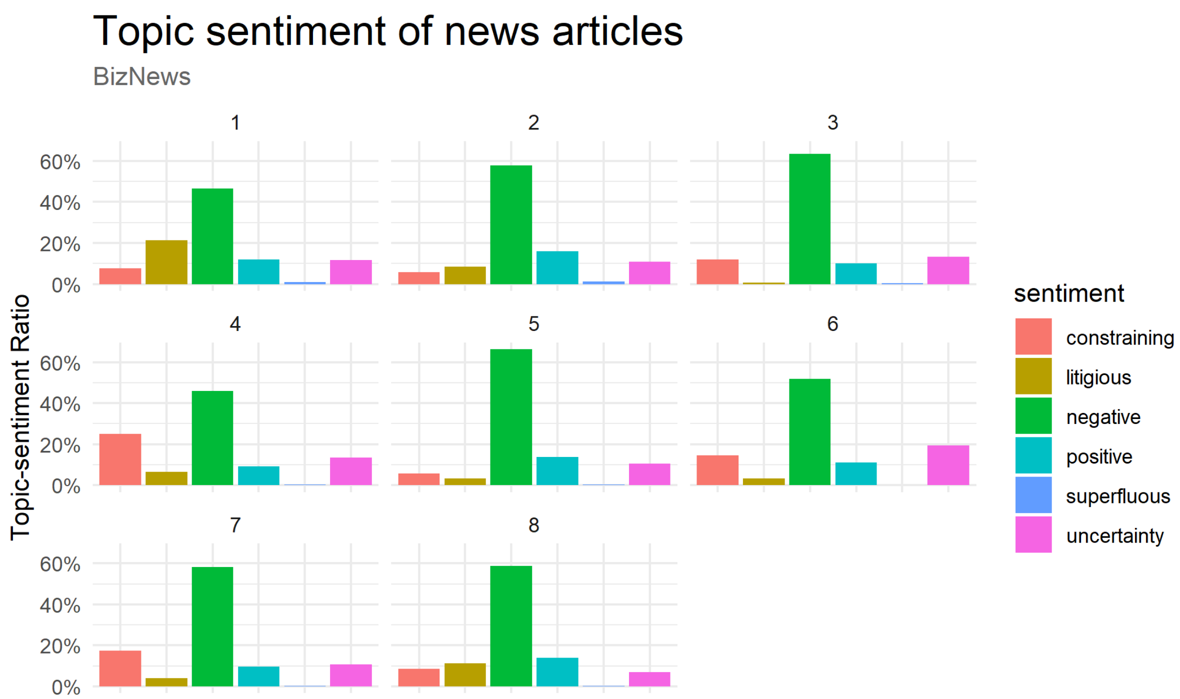

FIGURE 102: TOPIC SENTIMENT: BIZNEWS

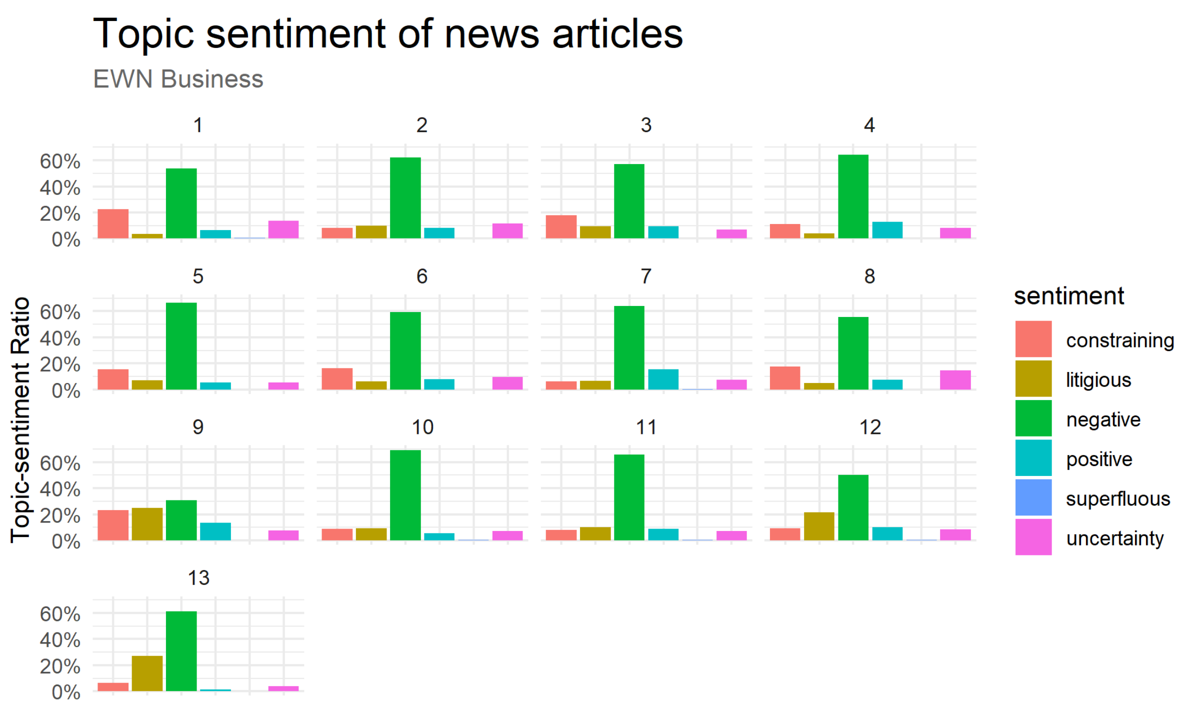

FIGURE 103: TOPIC SENTIMENT: EWN BUSINESS

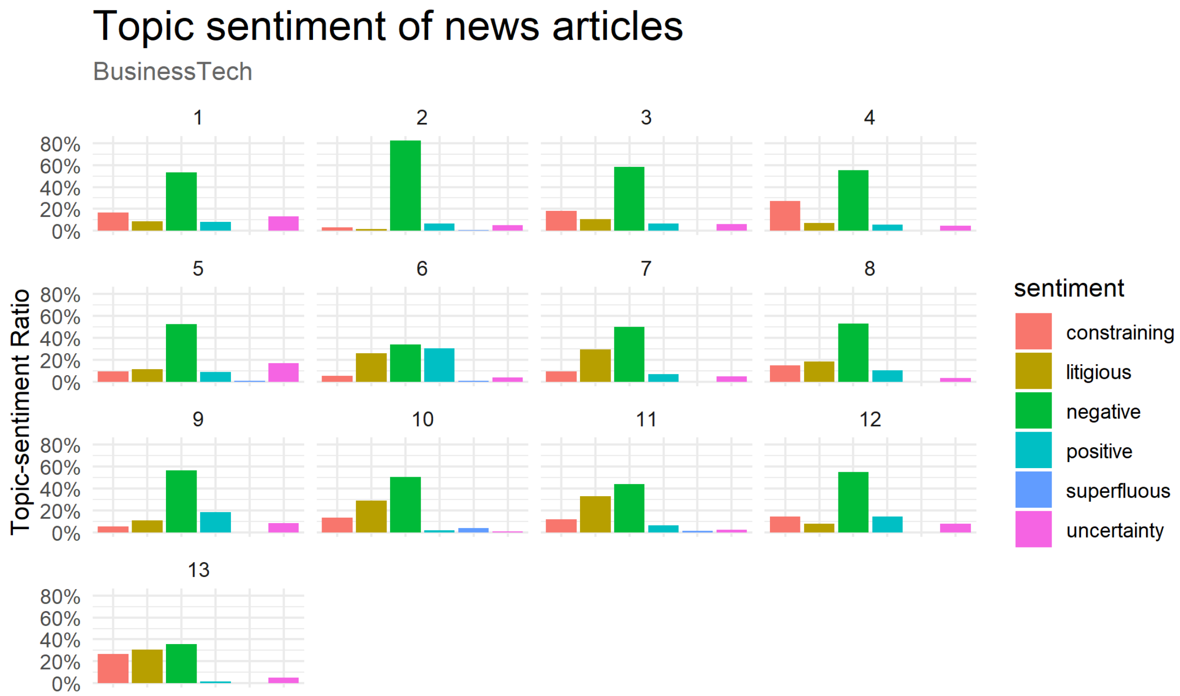

FIGURE 104: TOPIC SENTIMENT: BUSINESSTECH

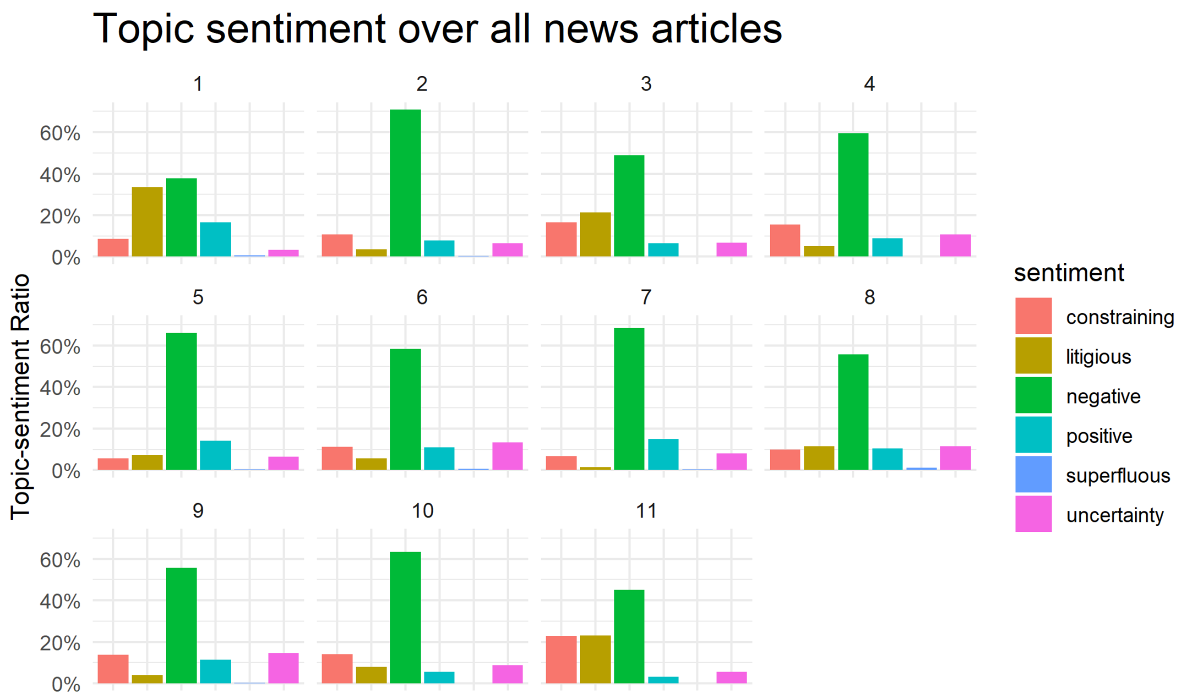

FIGURE 105: TOPIC SENTIMENT: ALL MEDIA OUTLETS

Conclusion

The analysis of the five media outlets and their combined datasets provided insights into the most frequented words used and the general sentiments of the articles published, in addition to the topics of the articles that are published in Covid-19-related business journalism.