Kellogg's Consultancy Report:

Assessing the Company's Reputation from Multiple News Data Sources

A Proposal

Authors:

Ilinca Curcudel (2649029) i.curcudel2@student.vu.nl

Iggy Nieuwenhuizen (2638457) i.nieuwenhuizen@student.vu.nl

Florent Brunet de Rochebrune (2619225) f.d.brunetderochebrune@student.vu.nl

January 23, 2023

Contents

Executive Summary

Kellogg's is a multinational food manufacturing company operating in more than 180 countries. Maintaining a good company reputation is important for: increased consumer trust and loyalty, increasing sales; employee morale, attracting high-quality staff to the company; better and easier partnerships with other companies; and easier access to capital such as loans and investments.

Kellogg's have been in multiple controversies: criticised for marketing unhealthy foods to children, using artificial ingredients and high levels of sugar, its treatment of workers in factories and environmental concerns. All affecting its reputation and the aforementioned benefits. Once the damage is done, it's significantly more difficult to regain trust. It is imperative to early assess the company's reputation once any news from any source regarding the company gets out, to investigate, take action, communicate with stakeholders, and possibly go into damage control.

1 Introduction

This report will argue for the implementation of a data science pipeline by Kellogg to track and analyze its reputation among the public. We will demonstrate the benefits of such a pipeline by presenting a preliminary pipeline and explaining how we plan to measure and interpret reputation for the Kellogg company. Furthermore, we will delve into the specific design choices made in the proposed pipeline.

1.1 The Significance of Reputation for the Kellogg Company

Kellogg's products are widely recognized and well-received by consumers worldwide, leading to repeat business and a strong reputation for the company and its brands. This reputation is crucial for Kellogg to maintain its competitive edge and continue generating returns for shareholders. The importance of building and preserving the Kellogg brand cannot be overstated. However, it is important to note that while a strong brand reputation may take years to establish, it can be quickly damaged if negative events are not handled appropriately.

Traditionally, marketing and communication teams have relied on monitoring news articles to address negative events. However, this approach is becoming increasingly difficult as the number of news sources continues to grow in the digital age. Fortunately, the internet also provides new opportunities for data collection and analysis. By using advanced data science tools, it is possible to gather and centralize news data about a company, and quickly turn it into actionable insights.

We propose that Kellogg should create a pipeline that uses these tools to monitor the sentiment and volume of news coverage about the company. By analyzing this data, the company can make informed decisions about when to launch media campaigns or take other actions in response to news coverage.

1.2 Defining and Operationalizing Reputation Measurement

The Agenda-setting theory is one of the most robust and empirically tested theories in any type of communication research. The first-level agenda-setting effect highlights that media can influence the public opinion by delimiting which issues are the most important ones. M.E. McCombs (1972) are amongst the first who aimed at researching the study of "agenda-setting". Their main finding was that there was a high comparability between the rank order of issues in the media and how the public perceived these issues. C.J. Fombrun (1990) study is one of the first elaborate quantitative research that tackles the relationship between media visibility and corporate reputation, and it seems that following their empirical analysis, they found the negative correlation between the two. Wartick (1992) did not have success in showing the positive influence of this relationship for most of the companies in his sample, however he emphasized that there was a significant and positive correlation between media visibility and corporate reputation on companies that had "good" or "average" reputation levels.

The second-level agenda-setting effect is about the evaluation and interpretation of the issues covered. This effect examines the influence of the properties or qualities that describe the attributes in the media and how the public perceives these. These attributes have two components: the affective component which describes the object (positive, negative or neutral) and the cognitive component that describes the characteristics of the object M.E. McCombs (2001). In various studies this theory was further used when researching the relationship between the attributes of corporations in the news media and corporate reputation. C.J. Fombrun (1990) only concluded that there is a significant correlation between media visibility and corporate reputation when it comes to high-diverse categories. Another research paper conducted by S. Einwiller (2010) has established that there was a positive correlation between media favorability and corporate reputation in the environmental and sustainable attribute.

We operationalize reputation monitoring by combining first-level agenda-setting and second-level agenda-setting, we identify that reputation monitoring has two important components. First-level agenda-setting is about how much the news speaks on a topic and hence how much people pay attention to it. Second-level agenda-setting goes one step further and also ascribed value to the tone of the news. Therefore our reputation monitoring pipeline will pay attention to both these factors, the amount of news on the company and the sentiment attributed to this news.

2 Pipeline

To demonstrate the effectiveness of a data science approach for reputation monitoring, we have developed a preliminary pipeline for monitoring the reputation of the Kellogg company. In this section, we will provide an overview of the pipeline and explain the reasoning behind various decisions made throughout the pipeline.

Our pipeline begins by processing a large number of news articles and produces a graph that illustrates the reputation of the Kellogg company. This graph can be used to identify events that may have a negative impact on the company's reputation. The pipeline is outlined in Figure 1.

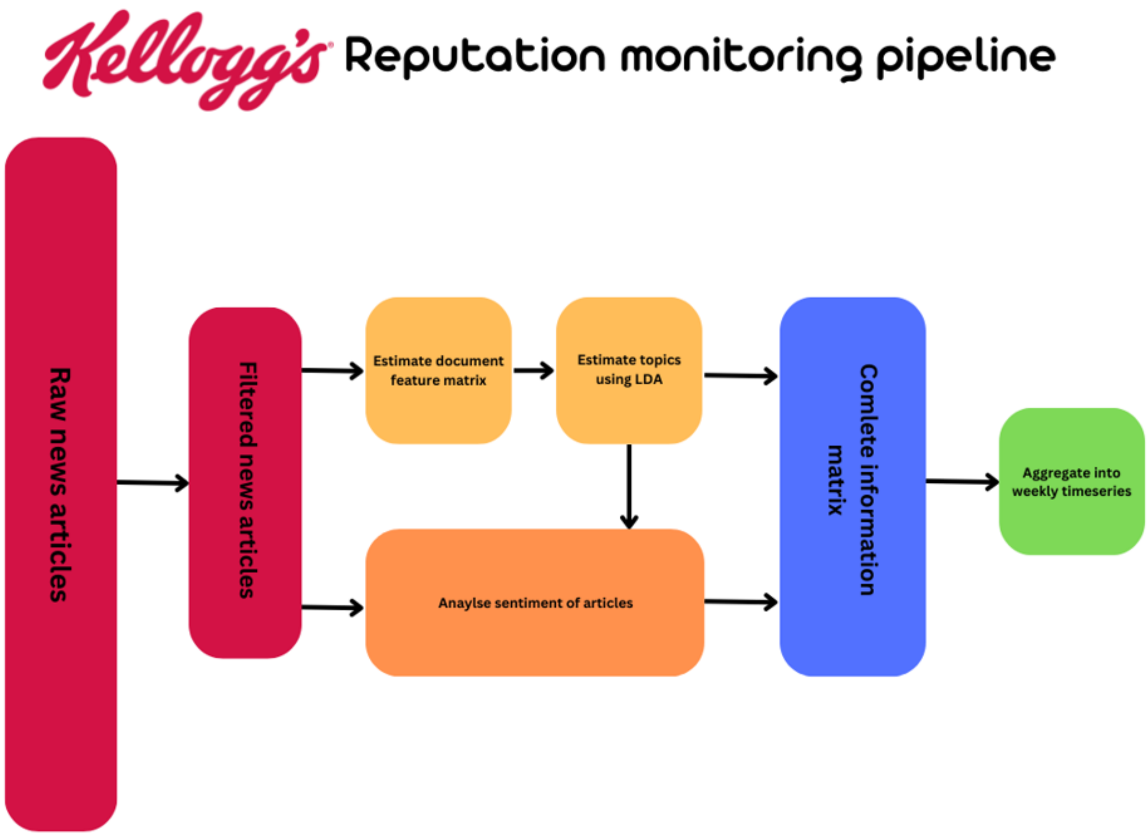


Figure 1: Pipeline for Reputation Monitoring

The overall goal of this pipeline was to extract insights from unstructured news data. We begin by identifying news articles that pertain to the Kellogg company and extracting two types of features from them: "topic scores" and "sentiment." These features are then used to create a weekly time series, which allows us to identify events that may have a negative impact on the company's reputation. This approach saves time for the marketing and communication department by automating the reputation monitoring process and reducing the potential for human errors.

2.1 Selecting Relevant Articles

Before analyzing the text data, a selection of relevant news articles was made from the initial 8,501 articles. The "LexisNexis Relevancy Scores" were used to determine the relevance of each article, with scores typically ranging from 50% to 99%. In this particular dataset, the lowest score was 85%, indicating a high level of relevance according to LexisNexis. The distribution of scores is shown in table 1 below.

Quantile	0%	25%	50%	75%	100%
Relevancy score	85	91	92	96	99

Table 1: Quantiles of the Relevancy Scores

The decision was made to eliminate articles with less than 200 words in order to decrease the number of articles. These articles typically contain minimal information, are frequently generated by automation, and can cause fluctuations in the sentiment scores that will be calculated later. Additionally, all articles from newswire sources were removed as they are primarily brief press releases that do not impact public sentiment towards brands such as Kellogg.

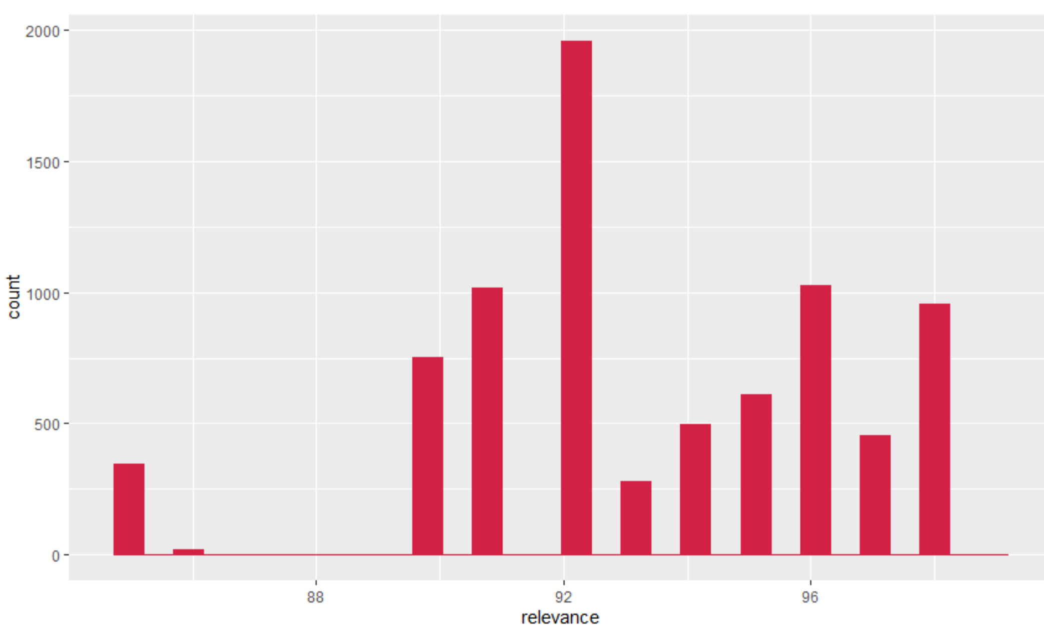


Figure 2: Histogram of Relevancy Scores

2.2 Preparing Data for LDA Estimation

The next step was to prepare the relevant articles for estimating a Latent Dirichlet Allocation (LDA) model. To estimate an LDA model, it is important to create a Document-Feature Matrix (DFM) that contains only relevant information. The following steps were taken to create a relevant DFM:

First using the package quanteda we tokenized the individual articles. In this step we also removed: punctuation, symbols, numbers, urls and uppercases.

The next step was to identify bi-grams, which are pairs of words that have a distinct meaning when used together. For example, "New York" has a different meaning than "new" and "York" when used separately. Once all bi-grams that appeared more than 500 times were identified, eight were selected as relevant. These were:

- per share (relevant because it differs from share/sharing in other context)
- kellogg company (might be relevant as news outlets that mention company might be more focused on business)
- kellogg co (Same as previous entry)
- market data (This indicates market news)
- nyse k (This is a reference to the stock ticker that would otherwise be lost)
- special k (This is an important brand of Kelogg's)
- rice krispies (This is an important product of Kelogg's)
- Battle creek (This is the town were Kelogg's is headquartered)

These bi-grams were merged into single tokens so their meaning would not be lost.

Lastly, all tokens were stemmed. Stemming is the process of reducing a word to its base form, or stem. For example, the stem of "eating," "eats," and "eaten" is "eat." This process merges words with similar meanings into single tokens and reduces dimensionality.

The first DFM was then created. As seen from the word cloud in Figure 3a, the DFM still contains many uninformative words.

(a) No Modifications (b) After Removing Stopwords

Figure 3: Wordclouds

There were a significant number of words that added little meaning, known as "stop words," which can be removed using pre-constructed dictionaries. The "marimo" dictionary was used for this purpose. Additionally, the words "also," "can," "inc," and "news" were removed. The results can be seen in the word cloud in Figure 3b.

The word cloud already appears much improved, but there are still issues to address. Many of these words may appear in more than 50% of the text and are not useful in identifying topics. Additionally, some words may appear so infrequently that they are not useful for statistical analysis. Therefore, all words that appeared less than 50 times, all words that had less than 3 characters, and all words that appeared in more than 50% of the text were removed. This resulted in the final wordcloud in Figure 4.

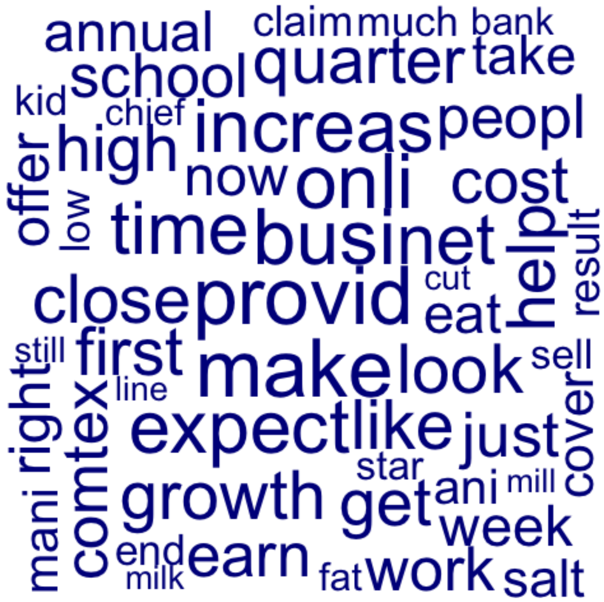


Figure 4: Wordcloud with Stopwords Removed

2.3 Estimating the LDA

The next step is to estimate a Latent Dirichlet Allocation (LDA) model. LDA is a statistical model that attempts to discover unobserved topics within a set of observations. Initially, an LDA model with 10 topics was chosen. The reasoning behind this was that if important topics were found to be missing, the model could be expanded by estimating $K(11, 12, 13, ...)$. If, on the other hand, it was deemed that topics should be merged, that would also be manageable. The top 10 words in each of the topics found in the LDA are presented in Figure 5.

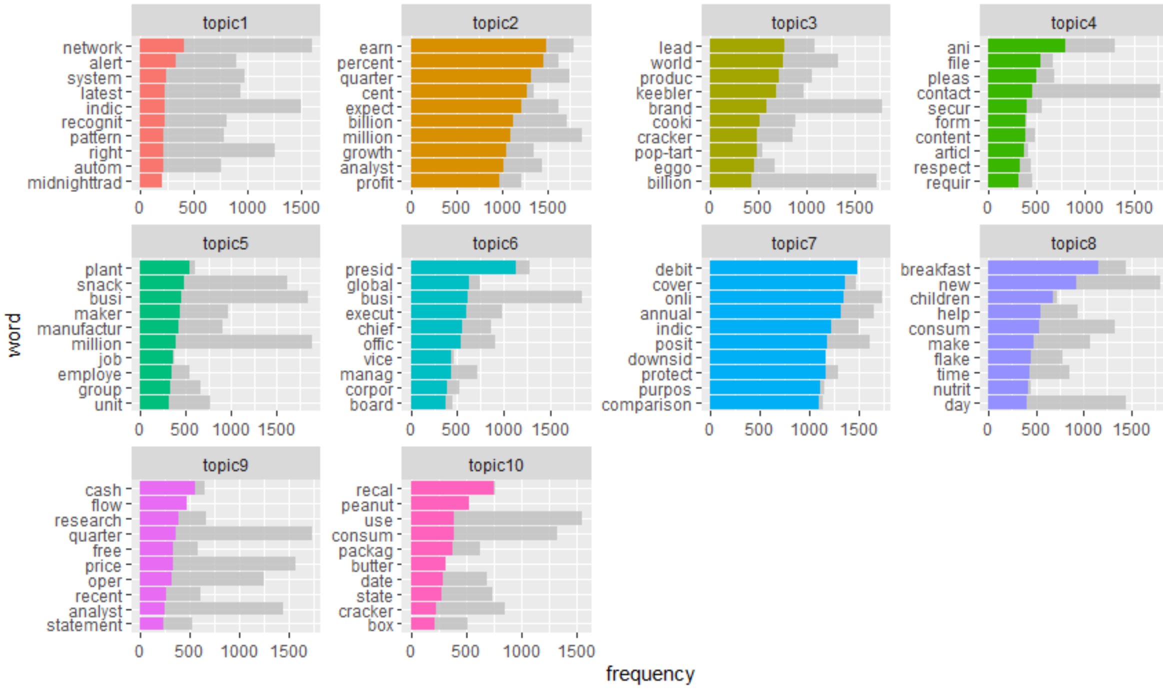


Figure 5: Most Frequent Words in LDA Topics

Now, the aim is to identify the topics. It is clear that some of the topics may need to be grouped together, therefore, there is no need to estimate a larger LDA model. The topics were grouped together as presented in Table 4 below.

Theme 1	Promotional activity	3, 4
Theme 2	Product launch	1
Theme 3	Product recall	10
Theme 4	Investment analysis	5, 6, 7, 8
Theme 5	Product claims	2, 9

Table 2: Grouping of LDA Topics

After merging these topics we end up with the following wordcloud for the five themes in Figure 6. With the top 10 words per theme found in Figure 7.

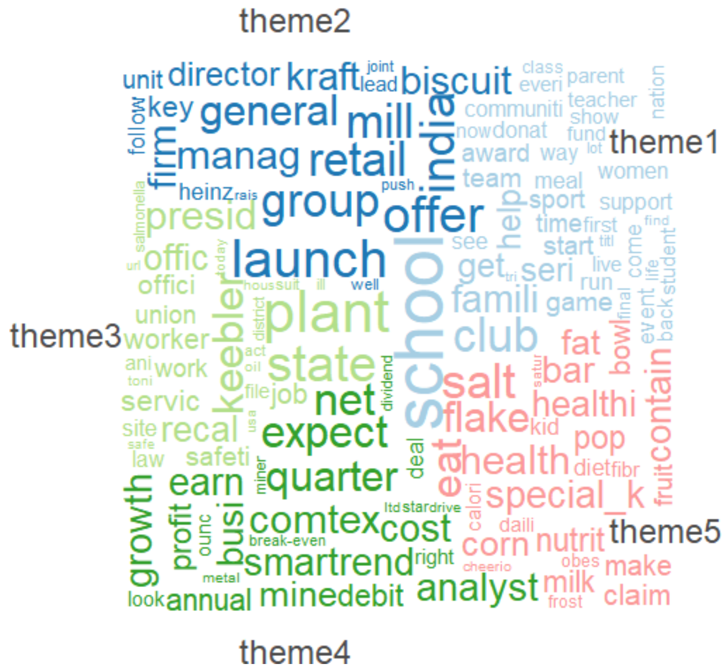


Figure 6: Wordcloud Based on Selected Themes

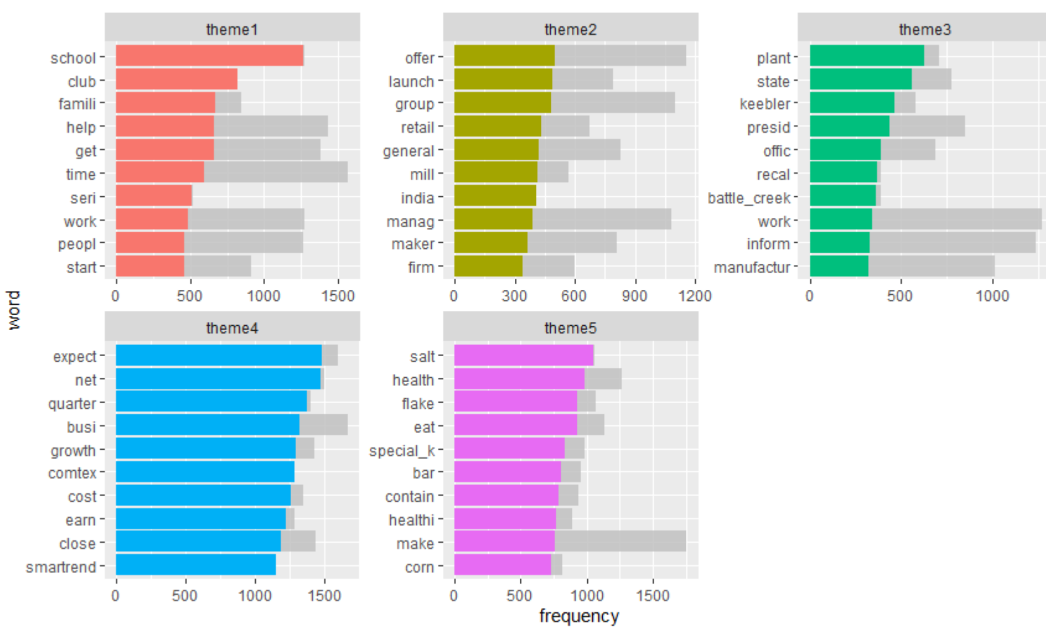


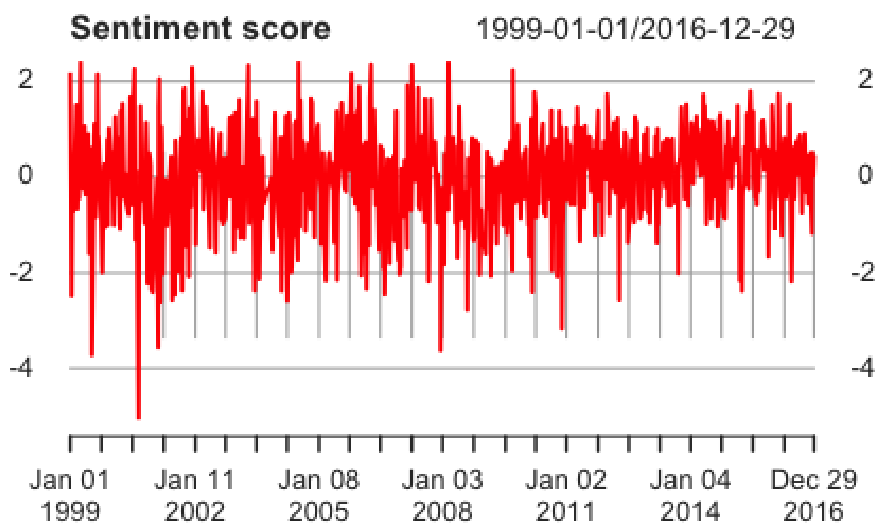
Figure 7: Most Frequent Words in Selected Themes

2.4 Text sentiment analysis

Textual sentiment analysis is the process of assigning scores to (parts of) texts on a scale from negative to positive, in order to determine the overall sentiment or attitude expressed in the text. This type of sentiment analysis is also known as "polarity-based" or "scaled" sentiment analysis.

The scores are usually represented as a numerical value, such as a decimal or percentage, indicating the degree of positivity or negativity. For example, a score of *-1* would indicate a very negative sentiment, a score of *0* would indicate a neutral sentiment, and a score of *1* would indicate a very positive sentiment.

In our analysis, we determine the sentiment by utilizing a lexicon-based method, enhanced with valence shifters information. The sentiment score for an article is calculated by adding up the modified sentiment scores of all its unigrams/bigrams. The modification is done by assigning weights to each unigram based on its location in the document and taking into account the presence of valence shifting words. For more efficient results, we have scaled the sentiment scores via standardization.



3 Results

Further to be able to gain more insights about reputation related triggering articles, we have applied a new feature that looks over the amount of articles and the sentiment scores over time. The results can be found in Figure 8 and 9, respectively.

If there is the likelihood for one single negative article to make reference to Kellogg's, this does not necessarily imply that there is likely a negative event happening for the company. Combining the smoothed out amount of articles and the sentiment scores together leaves us with a normalized vector, giving more weight to when there are more negative articles written in a certain time frame. A threshold of *-0.6* was chosen and found to be the most optimal given the events previously discussed and given data.

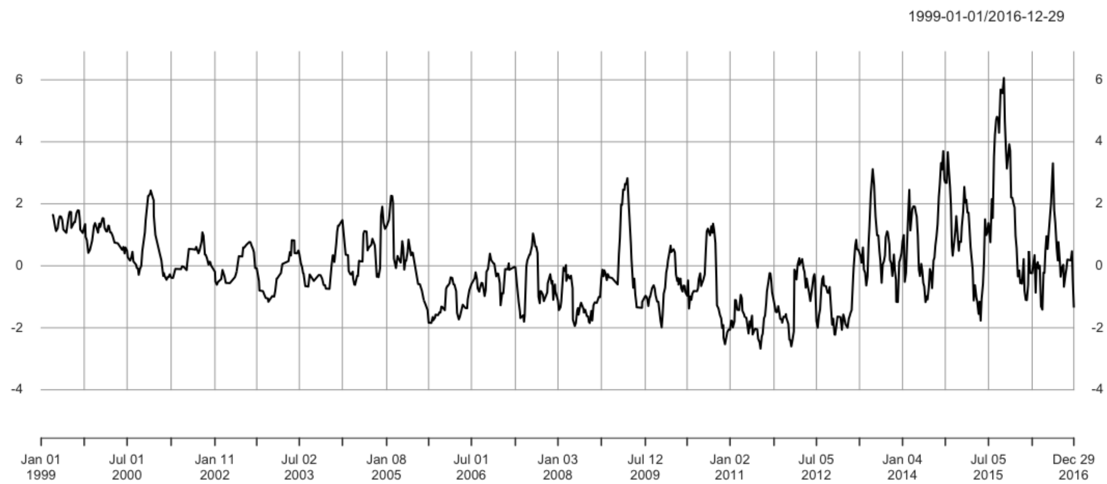


Figure 8: Smoothed Amount of Articles

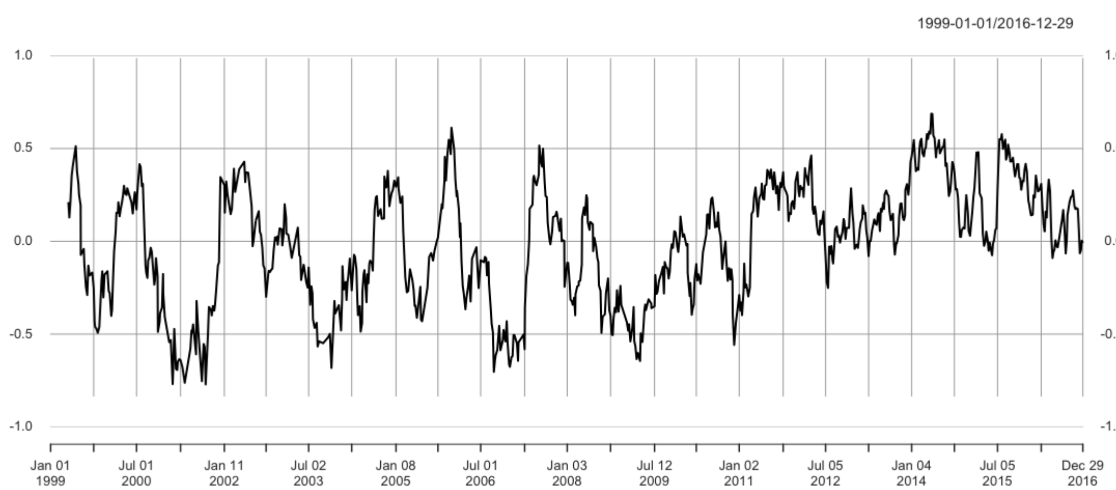


Figure 9: Smoothed Sentiment Score

In *Figure 8* it's noticeable that there is a spike increase in the total number of articles which make reference to Kellogg's. Also, looking at *Figure 9* it's noticeable that these articles have an associated positive sentiment score, which highlights that the reputation of Kellogg's company was merely affected. It seems that in 2015, following the annual RepTrak survey, held by the Reputation Institute, the Kellogg Company was placed as America's most reputable consumer company. One of the reasons attributed to its success were the company's "impressive sustainability and product-based initiatives in 2014 Culliney (2015). This event explains the increase in the number of articles between 2014-2015, and hence the associated positive sentiment score about the reputation of the company.

To further assess the negative likely reputation events of Kellogg's, we have plotted the above described feature. The results can be seen in Figure 10.

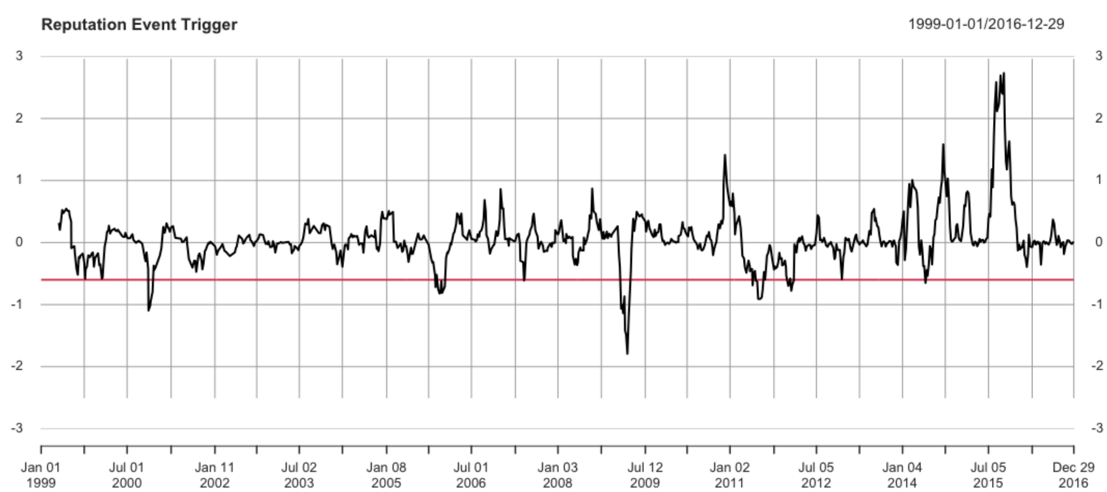


Figure 10: Combined Smoothed Article Amount and Sentiment Score with Reputation Event Trigger Threshold Line

It seems that considering the imposed threshold, there was a total of 6 negative reputation media encounters. As can be seen in Figure 8, there was an increase in the number of articles in early 2009, which triggered the larger spike in Figure 10. It seems that on April 20, 2009, the Federal Trade Commission (FTC) accused Kellogg's for making misleading cognitive health and deceptive claims about one of their products, and after the settlement Kellogg was not be able to misinterpret the results of different studies or tests anymore FTC (2009).

The other associated reputation events triggered were highlighted by the following news events:

- November 2000: glass found in cereal box, lobbying against keeping sugar amount from box ingredients list
- June 2005: Kellogg's Tony the Tiger voice actor died
- April 2009: FTC's complaint about deceptive claims
- June 2011: Kelloggs want to 'self-regulate' junk food advertising to children
- December 2012: 52-week low for the stock
- June 2014: Many job layoffs at the headquarters

4 Recommendations

A first step would be to conduct a thorough analysis of the current situation, including an evaluation of the news coverage and its impact on the company's reputation. Based on this analysis, developing a comprehensive communication strategy that addresses the key issues and stakeholders would be advised. This plan would include clear and transparent messaging to be communicated to employees, customers, investors, and the media.

In addition to communication, there should be necessary actions taken to address the issues, such as: making changes to company policies or procedures, providing additional training for employees, or launching a public relations campaign to counter negative coverage.

It's important to continuously monitor the news coverage and evaluate the effectiveness of the communication plan and any actions taken. Making sure that the transmitted message is transparent and authentic is vital. Additionally, working to continuously improve the reputation management strategy and be proactive in addressing potential risks to the company's reputation would be advised.

In summary, our overall approach would be to assess the situation, develop a communication strategy, take action, monitor and evaluate, ensure transparency and authenticity and continuously improve in order to restore and enhance the company's reputation.

An infographic made of the recommendations to share around the relevant departments:

5 References

C.J. Fombrun, M. S. (1990). What's in a name? reputation building and corporate strategy. *Academy of Management Journal*, 33:233–258.

Culliney, K. (2015). Kellogg 'most reputable' us consumer firm thanks to vocal sustainability efforts, says expert. *BackeryandSnacks* https://www.bakeryandsnacks.com/Article/2015/06/09/Kellogg-most-reputable-consumer-company-Reputation-Institute-RepTrack.

FTC (2009). Kellogg settles ftc charges that ads for frosted mini-wheats were false. *Federal Trade Commission - Press releases*.

M.E. McCombs, D. S. (1972). The agenda setting function of the mass media. *Public Opinion Quarterly*, 36:176–187.

M.E. McCombs, S. G. (2001). The convergence of agenda setting and framing. *Corporate Reputation Review*, S. Reese, O. Gandy, and A. Grant:67–8.

S. Einwiller, C.E. Carroll, K. K. (2010). Under what conditions do the news media influence corporate reputation? the roles of media systems dependency and need for orientation. *Corporate Reputation Review*, 12(4):299–315.

Wartick, S. (1992). The relationship between intense media exposure and change in corporate reputation. *Business and Society*, 31:33–42.