

## Text Analytics Lessons Learned: When MAUDE Doesn't Talk to You

Grace Heyne Lybrand, Zencos Consulting, LLC; Michael Swanson, Boston Scientific

### ABSTRACT

Data scientists are trained in analyzing data, specifically unstructured text data. Often, they need to mine insights from unstructured text data in unfamiliar industries. This paper will articulate the lessons we learned as we analyzed data and successfully partnered with subject matter experts to inform and guide our dive into the medical device industry. Using SAS® Viya™, we text mine the narratives submitted with medical device failure reports to the FDA. As you embark on your next text analytics project use our lessons learned to strengthen your partnerships and insights.

### INTRODUCTION

Unstructured text data is only as useful as the problem it is being asked to solve. Often, defining the correct problem and methodology can be more difficult to nail down than the actual analytics results. Our Zencos team learned this as we enthusiastically dove head first into the Manufacturer and User Facility Device Experience (MAUDE) database that is collected by the FDA on medical device malfunctions and failures.

The MAUDE dataset contains data from the several hundred thousand medical device reports (MDRs) of suspected device-associated deaths, serious injuries and malfunctions. These reports are submitted to the FDA by mandatory reporters (manufacturers, importers and device user facilities) and voluntary reporters such as health care professionals, patients and consumers. We are pleased to partner with Boston Scientific, a major medical device manufacturing company, to mine actionable analytical insights from this rich data source.

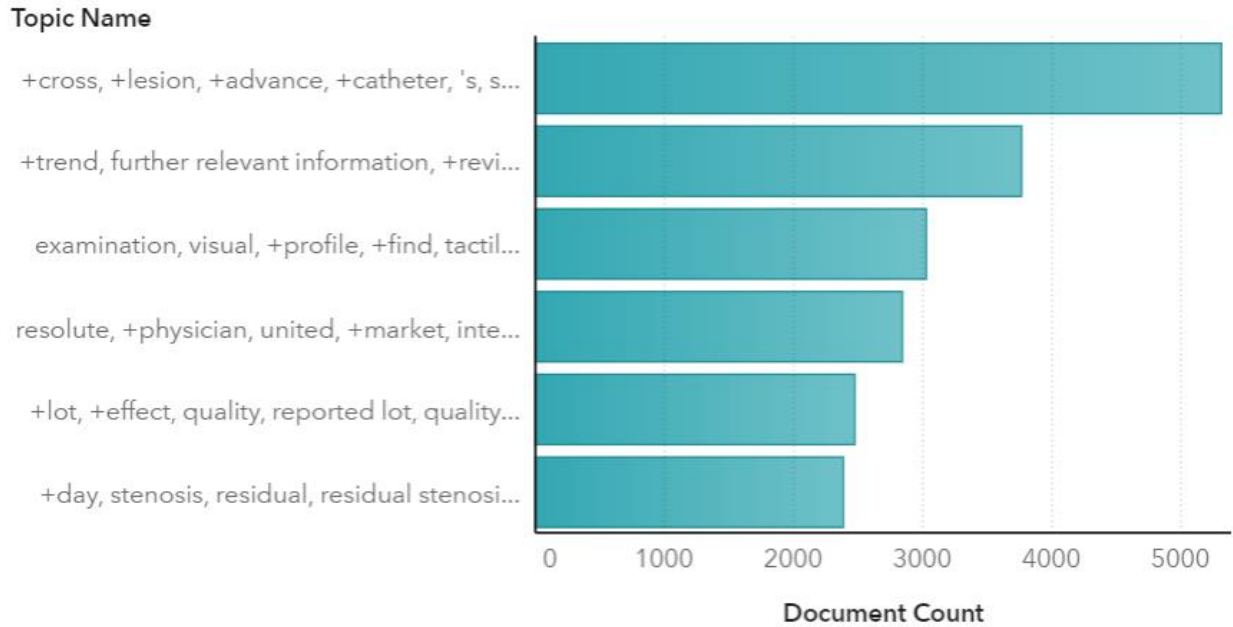
We learned the hard way that analytical knowledge is not all that is needed for success, and especially when working within such a specific subject, it is important to be discovery driven rather than data driven. Frequently, as data scientists we let the data speak for itself and reveal what is there to be found without bias. However, in this case the key to success was to pursue analytics in search of a specific answer to a question. Throughout this paper we outline the lessons learned that made us more successful analytical partners and how SAS® Viya™ enables that partnership.

### TOP THREE LESSONS LEARNED

At Zencos we are fortunate to have partners that present us with challenging impactful problems, but also share their knowledge and skills to help us reach our mutual goals. Therefore, when we decided to dive into the MAUDE data to see what insights we could unearth and pass to our clients, we soon realized we needed help. Boston Scientific helped us utilize our expertise in partnership with theirs to achieve the best possible results. We learned a lot from our partnership with Boston Scientific, we hope the lessons we have learned will help inform future analytical endeavors.

#### 1. MOTIVATION FIRST

At Zencos, we wanted to uncover findings in the MAUDE data that would provide Boston Scientific some actionable analytics insights. We began by taking a sample of the MAUDE data and investigating the population for trends. This turned up topics seen in Figure 1, that while interesting, were far from being actionable analytics, due to how general they were. We needed Boston Scientific's help to drill down and mine actionable results.

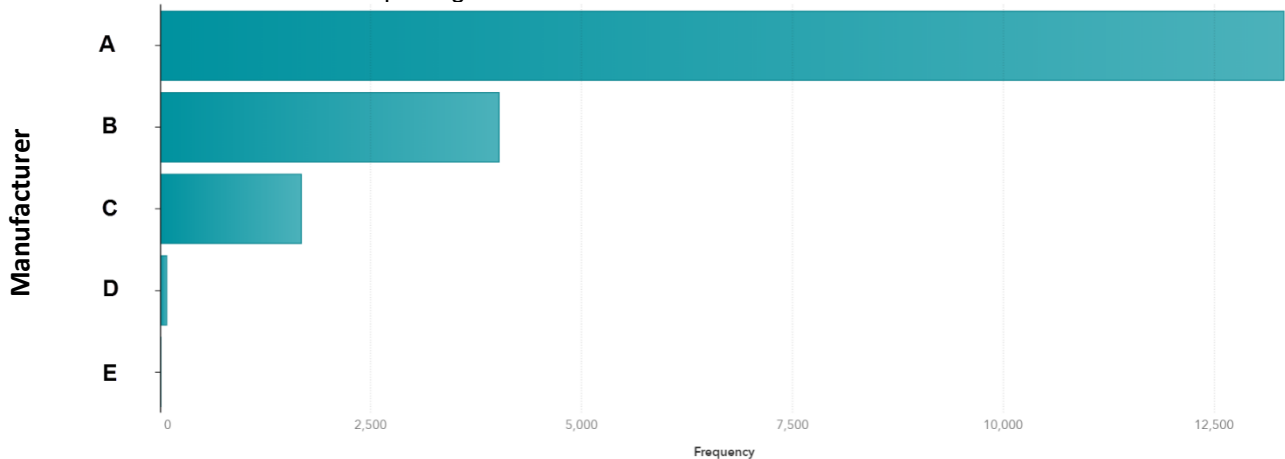


**Figure 1: Original Topics**

Boston Scientific helped us understand the ways we should consider the MAUDE data before diving in. For instance, we discussed the below questions:

1. Are we looking at a specific issue?
2. Are we looking for a characterization of the product issues?
3. Are we comparing our products to the competitors?

The difference in motivation led us to gather data differently, especially in terms of how we scope and scrub the data. For example, when scoping data, we consider things like: Which manufacturers are in scope? Which products? Which issues? Considering these different issues allowed us to focus on our motivation within our analysis. For example, in Figure 2 we see the population stent manufacturers with reported issues suggests that we should perform our analysis by manufacturer and focus only on the main three manufacturers with the bulk of the reports. It suggests a focus on the manufacturer cross sections because there three key players we observe and it will be useful to get insight into what and how each manufacturer is reporting their device failures.



**Figure 2: Manufacturer Spread**

## 2. FOCUS ON ACTIONABLE ANALYTICS

Boston Scientific is already producing a lot of actionable analytics. For our analysis to add maximum value we built on what is already occurring. By fitting text analytics results into what our customers are already doing we increase the credibility of the results and create more robust reasons to take specific actions. As we made progress we checked in to ensure our work was adding value to what Boston Scientific was already doing, not just confirming something or creating an insight for which there were no resources to pursue action.

### SELECTING TECHNIQUES

Text analytics can mean a lot of things. It can be as simple as word counts or as involved as a neural network, with a wide spectrum in between. When the Zencos team first dove in, we wanted to understand what topics were coming up in reports, how people felt when writing each report, and if that could give us any insights. However, once we began talking to our partners we realized taking a step back would allow us to move further forward than before.

### Techniques

Different techniques produce different results. It is an important first step for a partnership in text analytics is to be clear about methodology and what results can be expected from each technique. Below we walk through the techniques we discussed and some lessons we learned along the ways.

#### *Data Cleaning*

Cleaning and parsing is how we take a jumbled, messy corpus and morph it into something that is analysis ready.

- Parsing: Breaking the documents into useable chunks, whether it be by words or phrases.
- Stemming: Process of reducing words to their root with the intention of capturing the meaning in the document, not the tense or other attributes of the word usage.
- Part of Speech Tagging: Tagging all words within each document with their corresponding part of speech.

**Stop Words**

A simple, but impactful tool during data cleaning is the creation of a custom stop word list. Stop words are common words that do not typically add value to text analytics, and they should be removed prior to analysis. Specifically, while looking at more specialized data atypical stop words often come into play. Partnering with Boston Scientific resulted in a more tailored stop word list that greatly improved our analysis. In Figure 3 are a few examples of words that were added to our stop word list:

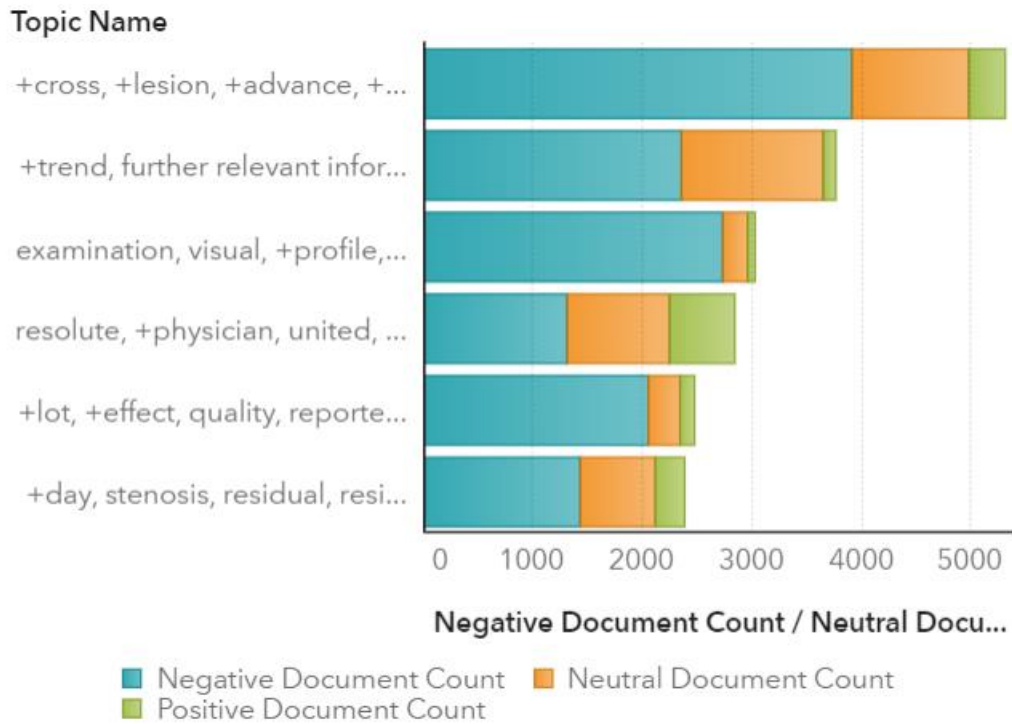
Term
additional information
alpine
applicable
applicable information
drug elluding
endeavor
faith
faith effort
information
integrity
per
physician
report
resolute
supplemental report
synergy
xience

**Figure 3: MAUDE Custom Stop word List**

**Term Frequency**

Word counts and frequencies are the basis for more robust text analytics methods. However, often it is useful to stop at the term frequency level and observe the insights that are there to be found. The word cloud in Figure 4 demonstrates that often the reports come from complications and are related to procedures. Observing term frequencies across difference subsections of the data gives a general idea of what is being discussed in each, and is often a great start to create actionable analytics.





**Figure 5: Sentiment Analysis on Original Topics**

### Tools

There are enough challenges when partnering to solve a subject matter specific analytical problem. We found SAS® Viya™ to be the best tool to collaborate and build on with SAS® Viya™ for a few reasons:

- Connectivity to open source: The freedom to use both tools without losing the common connection and performance found in SAS
- SAS Visual Analytics: Quick, easy to make, clear display of results.
- Tweaks are easy: The robust nature of SAS® Viya™'s text analytic features means that making changes to methodology or visualizations can be done quickly and with low cost, once again increasing the speed at which the project progresses

Throughout this process we learned which tools gave us to best abilities to quickly iterate and communicate results. One of our favorite ways to perform text analytics in SAS® Viya™ was in Visual Analytics using the Text Topics object, seen in Figure 6. This object provides quick insights in a no-code environment. This object is great for experimenting and developing quick insights to facilitate strategic discussions between partners.



## RESULTS AND VERIFICATION

Thinking back to the beginning of our process at Zencos, we looked to find interesting insights, then planned to find ways to make our analytics actionable to our partners. We are often proponents of data driven methodology, but in the case of client needs it is useful to be discovery driven and to keep in mind the specific actions our results will lead our partners to take. For instance, Boston Scientific asked: How is stent damage displayed in our competitors? What factors impact the ability of a stent to function?

Once we've got the results, in the case of medical devices and many other fields, it takes communication in the partnership to ensure the results are robust in both analytics and subject matter expertise. Discussing whether results make sense is an important conversation to have early and often.

Specific questions and conversations lead to specific analytics, which in turn produces relevant results that answer the question and provide actionable analytical insights. Table 8 contains the results of specific analysis of the top three manufacturers of stents, one of which is Boston Scientific.

<b>Company A</b>	<b>Important Topic Terms</b>
1	profile, examination, visual, the, of, tactile examination, tactile, issue, find, and
2	lesion, guide, catheter, non-bsc, the, cross, advance, promus, artery, balloon catheter
<b>Company B</b>	
1	event, relate, that, lesion, united, state, it, malfunction
2	will, distal, there, supplemental, obtain, future, relevant, be
3	image, show, vessel, review, image review, confirm, deployment, delivery
4	onyx, resolute onyx, guide, patient, wire, unable, get
<b>Company C</b>	
1	of, lot, device, history, review, return, base, quality, this, identify
2	xpedition, lesion, sds, target, millimeter, resistance, target lesion, note, due
3	guiding catheter, into, guide, shall, unexpanded, back, may, catheter, retract, reintroduce
4	implant, left, pull, lad, main, proximal, guide catheter, unspecified, circumflex, left
5	guide, wire, non-abbott, sds, catheter, rx, pull, guide catheter, advance, new

**Table 8: Final Topics by Manufacturer**

Issues related to guiding a catheter are prevalent across manufacturers. Therefore, by investing heavily in the development of better guides they could certainly serve patients better by solving an issue that occurs across the industry. By observing the topics being reported amongst competitors Boston Scientific can begin to take positive action.

## KEYS TO PARTNERING WITH NON-SUBJECT MATTER EXPERTS

The medical device industry, as with many other highly specialized industries, has its own set of terms and nomenclature that describes the business and environments in which it operates. With the medical device industry, the mix of technology with patient-related terms makes understanding the business a challenge. When working with partners in analysis of data to answer key questions in the business, it is imperative that the analyst is grounded in the overall business, its products and customers.

Boston Scientific relies heavily on data to make many decisions within the business, the data that is used must be well understood, verified, and structured in a way that the analysis of the data produces reliable and accurate results. In the specific example of using MAUDE data, the initial dataset includes many pitfalls for those that don't have a good understanding of the structure and content of the data. It is important that those working within MAUDE data understand the key fields within the data that will help answer the questions that industry is attempting to answer, this understanding is achieved by working



with the industry SMEs during the process of data analysis to clarify and adjust the analysis to ensure appropriate interpretation of the data and the resulting output.

Another key element of the process is understanding the questions that the business is trying to answer. In understanding the question, the background knowledge of the industry and data is important to ground the analyst in getting the data and resulting output structured in a way that the answers to the question are clear and defensible.

In summary, the main keys to success when partnering with non-subject matter experts in analyzing industry specific data are communication and education, making sure all involved are grounded in the basics of the industry and lines of communication are open to ensure proper feedback in the iterative process of data analysis.

## CONCLUSION

As with any good partnership, it is important to be aligned on the goals and strategy, particularly when it comes to text mining a specialized corpus. Just like the relationship between a navigator and a pilot, the subject matter expert guides the data scientist and together they meet their goals.

## REFERENCES

Healey, Christopher "Text Analytics". NC State Institute for Advanced Analytics. March 8, 2018. Available at <https://www.csc2.ncsu.edu/faculty/healey/msa-17/text/index.html>.

MAUDE - Manufacturer and User Facility Device Experience. Food and Drug Administration. March 8, 2018. Available at <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm>.

## ACKNOWLEDGMENTS

Thank you to Tricia Aanderud for her support throughout this process and facilitating our partnership.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Grace Heyne Lybrand  
Zencos Consulting, LLC  
glybrand@zencos.com

Michael Swanson  
Boston Scientific  
Mike.Swanson@bsci.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.