

# Analyzing Leatherneck Responses through Text Mining

**Coleman Strickland, Capt USMC**

---

## Purpose

The purpose of this project is to perform an in-depth analysis of responses gathered from 2019 Leatherneck participants. In total, there were 366 responses across seven questions totaling 18 pages of material. Simply reading through the document and gathering insights is difficult and time consuming. Utilizing current computer software and text mining practices, we provide an unbiased approach to further analyze participant opinion, cluster common themes, and output comments in a format for easier search and examination. These results are obtained in a fraction of the time, and provide feedback that is repeatable and reproducible across any text document.

This report is organized into three different sections:

- 1) Positive and Negative Word Counts: A naive look at the common positive and negative words found in the document.
- 2) Sentiment Network using Word Groups: A method in which frequently used words/phrases are divided into a network of reoccurring word groups with overlaying sentiment scoring.
- 3) Similarity Comparison: A clustering approach to identify common themes in the document and categorize those themes into an easily identifiable list.

## Positive and Negative Word Counts

We start by identifying the number of times words appear throughout the text. This procedure allows us to obtain a sense of what the document is saying without reading every comment. After calculating the word counts, we categorize the words as positive or negative based on how they are perceived in the English language. A word cloud of the output is depicted in Figure 1, with the color red identifying negative words and blue identifying positive words.

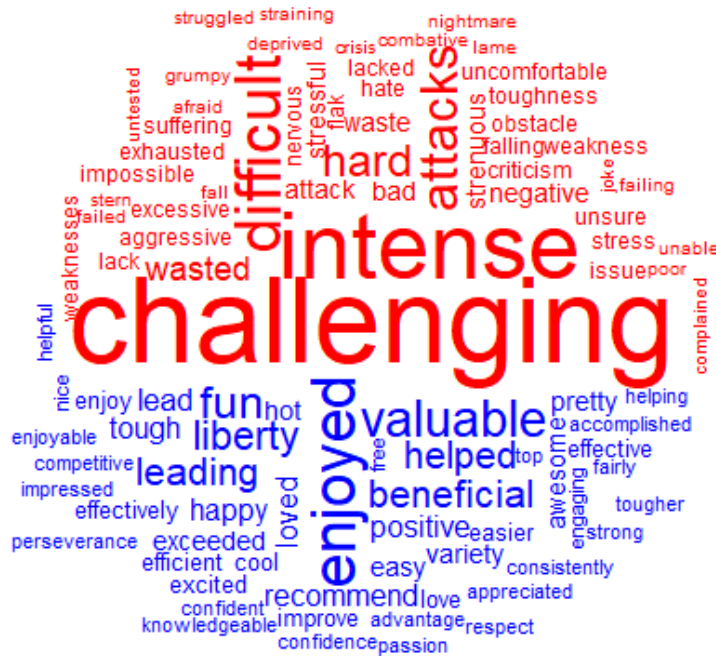
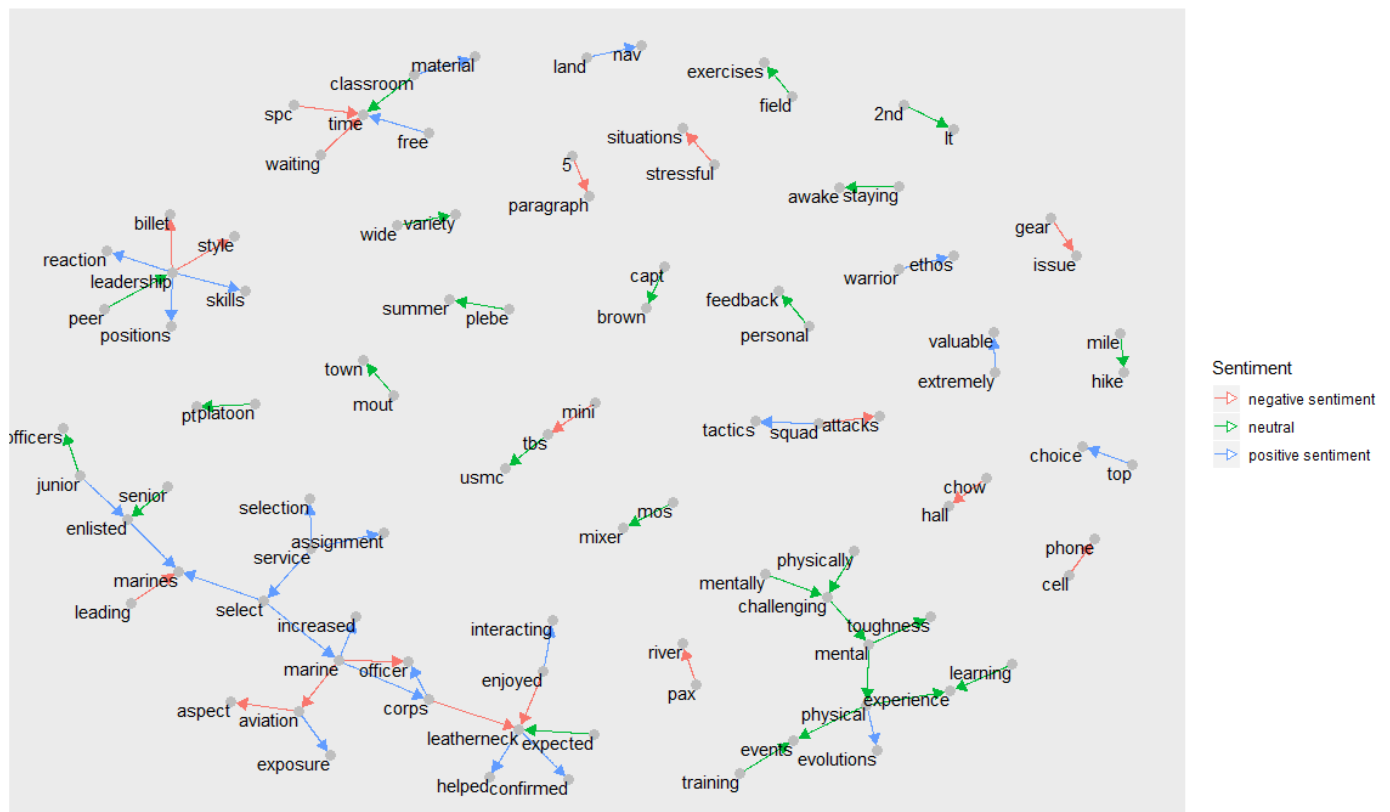


Figure 1: Word Cloud

The word cloud in Figure 1 gives a quick overview of the most common words stated in the text. By organizing the word counts into those that are positive and negative, student preferences are more easily recognized.

## Sentiment Network using Word Groups

Next, we divide the text into word groupings called bigrams, which are two words that appear throughout a text numerous times in a specific order. For example, word groups such as “chow hall”, “Marine Corps”, and “MOUT Town” are common to a Marine’s vocabulary and have particular meanings when grouped together as a phrase. Figure 2 illustrates a network containing bigrams that occur two or more times throughout the document.



**Figure 2: Bigram Network**

The nodes of the network are the specific words related to each word group. The arrows, or arcs, in the network depict the order in which the words are used (i.e. “chow hall” vice “hall chow”). The colors of the arcs illustrate the average positive or negative sentiment of comments containing a particular bigram. After viewing Figure 2, a particular bigram can be analyzed further by searching the

LNK\_Comments\_Search.xlsx document. For example, when the term “chow hall” is searched, the following two results are found:

- “More coordination with the chow hall to expedite meal times and make sure there is meals everyday.”
- “I don't know how it would be done other than by building a new building but the chow hall failed to provide everyone with enough food in a short amount of time because so much time was wasted there.”

Because the comments about the “chow hall” take on a negative connotation, the network displays a red arc. Using the network from Figure 2, simple searches of the excel document using the associated bigram network can help readers easily navigate to word groupings with positive or negative sentiments.

---

## Similarity Comparison

The final procedure is to group all similar comments together in order to identify common themes. We start by taking an individual response and comparing the comment to all other comments in the document. If the comparison exceeds a specified threshold, the similar comment is set aside for further analysis. If there are three or more similar comments relating to an individual response, we assign the grouping a theme. This process is repeated until all comments have been analyzed. Finally, we output the result to visualize the cluster of common themes throughout the text.

After running the text through our algorithm, 11 common themes are identified.

- Leading Peers and Receiving Feedback on Leadership Skills
- TBS Exposure
- Aviation Exposure
- Liberty
- FEX Length
- Leatherneck’s Impact on Service Selection
- Evolving Timelines
- 7-Mile Hike/MOUT Intensity
- Leatherneck was Fun
- Leatherneck was a Challenge
- Marine Corps Exposure

Clustering the document into the above themes allows for easy interpretation of the text and provides the user an overview of the shared opinions commonly found in the document. Output from the clustering method can be found in Table 1.

<b>Theme</b>	<b>Similar Comments</b>
Leading peers and receiving feedback on leadership skills	Leading my peers and seeing how I felt and acted as their leader. Also receiving feedback on said leadership
	Learning more of yourself as a leader and how you would act in certain scenarios
	Squad leader for attacks. Loved it.
	Peer evaluations and SPC debrief.
	Understanding where your leadership traits stand through SPC counseling and peer reviews.
TBS Exposure	Getting some insight on what TBS would be like
	What TBS is like.
	Being surrounded by Marines who cared about our service assignment. Classroom time and exposure to TBS
	I got to lead a squad in a simulated attack and see what I was like under pressure. Invaluable insight into what I need to improve on.
Aviation Exposure	Spend more than one day in aviation exposure (Pax River).
	More aviation stuff one day is not enough when nearly 100 mids will be Marine pilots
	More exposure to what it is like being in the fleet versus just what it is like for TBS. And A LOT more exposure to Marine aviation one day at pax river is not sufficient.
	I would have liked to see a more in depth aviation aspect. Many midshipmen are very interested in the aviation aspect and a few hours spent at a test squadron in Pax River wasn't the type of exposure that Marine Aviation deserved. I feel there should be time put in place to perhaps take an overnight trip to interact with a real squadron to better expose the midshipmen to the aviation aspect.
	Restructuring the interactions we had with Marine Aviators. The pax river visit felt like a rushed waste of time where we didn't get any real quality interaction with the pilots nor did we get time around the actual aircraft. Warrior night was also a really cool planned event that we didn't get to have due to rain but it was not rescheduled. I really wanted to talk with the enlisted marines and spend more time with them. Maybe including a NCO embedded in each platoon could have been valuable since on the yard we spend lots of time with Staff NCO's and never junior marines or NCOs.
Liberty	I heard other platoons didn't get as much liberty.
	A more uniform liberty schedule (some platoons consistently got much more liberty than others) builtin time to workout.
	I would give more time for liberty at the end of the night.
	Don't prioritize liberty. This gave many people the wrong focus and a huge amount of complaining came from people wanting liberty rather than wanting to train.
Longer FEXs	Make one of the field exercises longer in nature so we have a more realistic idea for what those are like at TBS. I was told the field ops at TBS are no shorter than a week and both of the ones we did were only for one night.
	Longer field exercises
	I think that more than 2 nights in the field would be beneficial. Either combine both FEXes into 1 or make one or both longer.
LNK's Impact on Service Selection	I already wanted to service select Marine Corps before the training.
	I thoroughly enjoyed my experience with the Marines at Quantico thus increasing my desire to service select Marine Corps.
	Leatherneck did not change my perspective on the Marine Corps so I came away feeling that becoming a Marine would be a desirable service assignment.

<b>Theme</b>	<b>Similar Comments</b>
Evolving Timelines	Rolling with the punches and constantly evolving timelines
	Peer leadership and managing shifting timelines
	The constant tasks.
7-Mile Hike/MOUT Intensity	I was expecting the 7 mile hike to be more intense but that was relatively easy. MOUT town was a bit much.
	I think the 7.8 mile hike should be more intense i.e. faster pace or more weight.
	I think that MOUT town could have been more intense. There were a few times during that evolution where it felt like more of a mess than an operation.
	I thought the 8 mile ruck was too easy.
LNK was fun	Difficult fun series of tests Yes because I had an enjoyable experience but was also challenged in ways I have never been before.
	I thought it was going to be challenging but fun. Yes there were definitely times where I was challenged and I never stopped having fun.
	That it was going to be difficult and generally stressful. No I had a lot more fun at leatherneck than I thought I would but it was difficult still due to the heat.
LNK was a challenge	I expected it to be a physically rigorous evolution with long days. Yes each evolution challenged me either mentally or physically to perform in a different way.
	It would be challenging and have long days ahead. Yes it did meet my expectations because I was able to get a good perspective on what the Marine Corps is about. It also pushed me mentally and physically.
	It was a physical challenge No it was more of a mental challenge of getting through the long days
	Push myself physically and mentally. Take an in depth look at the marine corps Yes I learned a lot about different aspects of the service and I was pushed physically
	I thought it would be a challenge but that I would enjoy it. I was also prepared for the hurry up and wait frustrations. Yes because I learned about perseverance and achieved my goals coming into it. It also really was a lot of hurry up and wait.
	A lot of waiting and being bored NO we were always kept busy and I enjoyed 95% of my time there.
	I expected it to be a hurry up and wait training No it was better organized then I thought it would be. We were treated as adults and I felt they did not waste our time.
Marine Corps Exposure	It would give me a good understanding of the Marine Corps. It did but I think it did a better job of showing what \ TBS will be like.
	Getting a better understanding of marine corps Yes the people were always willing to talk about the marine corps
	I expected it to give me a sense of what the marine corps is like. Yes it did. It showed me what the marine corps was like.

**Table 1: LNK Comment Themes**

---

## Conclusion

Although valuable insights can be gained using the three methods presented, additional analysis could be applied. Future work includes incorporating a more sophisticated sentiment-scoring algorithm. For example, negation oftentimes confuses sentiment-scoring algorithms as it indicates the opposite of what the algorithm reads. For example, the sentence “The Platoon was never late” has a positive sentiment; however, the algorithm reads key words such as “never” and “late” and scores the sentiment as negative. By incorporating negation identification into the algorithm, we can more accurately score sentiments across all comments.

Overall, we have provided an unbiased and easily repeatable analysis of the comments associated with Leatherneck 2019. The analysis provided in this document is meant to outline key findings and offer a framework to guide readers in easily searching Leatherneck comments to further analyze student opinions.