Dani Corum

12/19/25

Final Report

1. Introduction

For this project, I explored four questions:

1. What are the most common words in each chapter of *Dune*?
2. What words are the most common across the entire novel?
3. How many words are used only once?
4. How does each chapter's sentiment change as the novel progresses?

I believe this topic matters as an experiment in how accurately a long and complex text, such as *Dune*, can be analyzed using word counts and sentiment analysis. The dataset I used was a simple txt file that was converted from [this pdf](#). The methods I used were function creation, tokenization, counting, sorting, data cleaning, and data visualization.

2. Methods

I cleaned up the data from the txt file by splitting it up by "= = = = = =", which was used to indicate a chapter break. I then tokenized the words, while removing any stop words. I also chose to remove any occurrences of the word "said," since Frank Herbert rarely used anything else in writing dialogue. With the data cleaned, I then used functions to find word counts, sort them, and perform sentient analysis. For both the word counts and the sentiments, I chose to use bar charts for visualization of the chapters' data. I evaluated the performance of my results by two sets of criteria. For the word counting, as long as the chapters were varied in their top words, I considered it a success. As for the sentiment analysis, I closely compared the changes between positive and negative sentiments to the plot of the novel itself.

3. Results

Based on my criteria for success, the program worked well. There was variety in the top words per chapter, which most often included the key characters in each, as well as other important subjects. With the sentiment analysis, as tedious as it was to assess each change, I believe it tracked the sentiments decently. Generally speaking, we can see that the novel starts positive, turns negative about a quarter through, has a mix of positive and negative, then drops back to negative for the ending chapters.

4. Discussion

My results showed that both word counts and sentiment analysis worked decently well in showing trends for each chapter of *Dune*. One limitation I saw was that the visualization of the word counts was difficult to follow. Since there were 77 different words to show, the color palette made it difficult to differentiate words from each other, and the legend was very long. Additionally, for the sentiment analysis, a couple chapters did not seem to align with their sentiments, but it is difficult to determine without rereading the novel. On that note, *Dune* is a very complex novel with many overlapping themes and a strange, yet detailed world. As a result, this project cannot supplement the original text.

5. Conclusion

When we analyze the major trends in word usage in *Dune*, we can see that the majority of the most common words are the names of the characters we follow in each chapter. Using sentiment analysis, we can attempt to track changes in sentiment across the novel, but the complexity of the narrative makes it difficult to be entirely accurate.