Project Overview:

My project analyzes sentiment from the four Sherlock Holmes novels written by Sir Arthur Conan Doyle and graphs the graphs each against each other.

Procedure:

After downloading and saving the text from the Internet (url: <https://sherlock-holm.es/>). My program calls the each of the four novels to be run through functions that break the novel into chapters, read the sentiment of each chapter, and then plots the sentiment from each chapter of the entire novel on a line graph. One graph for each novel.

I chose to separate a fair amount of my code into separate functions because to me, it appears more readable and also easier to debug as I go along. I can test individual functions as well as parts of functions for specific problems. I also chose to do sentiment analysis on a chapter by chapter basis rather than line by line of sentence by sentence because when I did smaller data sets, there seemed to be a greater number of outliers and it did seem to actually be comprehensive of the novel.

Results:

|  |  |
| --- | --- |
| *A Study in Scarlet* | *The Hound of Baskervilles* |
| *The Sign of Four* | *The Valley of Fear* |

Above are the graphs depicting sentiment from the four novels. The values along the x values are essentially arbitrary and the y values refer to the sentiment. There are some interesting conclusions that can be drawn from these graph as a whole. They all fall within a fairly neutral zone, which is to be expected when analyzing chapters in novels. There tends to be enough positive and negative language that the average ends up pretty neutral. In light of this, most of the novels tend to be mostly positive. They almost never dropped into the negative numbers.

There is also some interesting information that can be seen within each individual book. Like how both *The Sign of Four* and *The Valley of Fear* have a fairly noticeable positivity peak in the beginning that levels out, whereas *A Study in Scarlet* and *The Hound of Baskervilles* have spikes toward the end.

Conclusions:

I think this project went pretty well for me. I really enjoy data visualization and this is the most I can delve into that topic right now and I did enjoy it. I didn't unit test in my docstrings, I was not sure how to implement it with pickledumps and those things, so instead I went through and test each function individually. I think my project was appropriately scoped for myself. I didn't really know anything about these topics before beginning this project and I think I picked something challenging and interesting enough that it wasn't completely time-consuming, but it provided enough difficulty that I wasn't bored. I think I probably could have put more effort into the actual visualization piece.