Sentiment analysis report

The dataset used were reviews of various Amazon products. In order to perform semantic analysis on a review inside the spreadsheet, I had to firstly remove all of the empty entries inside the spreadsheet. This was done by using the dropna() function of the pandas module. Next was removing the stop words such as 'the', 'is', etc as they have no meaning. The final step for preprocessing was converting all characters to lowercase for the NLP so that the analysis can be accurate and simple. I tested my sentiment function on 3 reviews:

The first review tested was the 4th row in the reviews column which reads: "This make an excellent ebook reader. Don't expect much from this device except to read basic ebooks. The good thing is it's cheap and good to read in the sun". My sentiment function gave it a polarity score of 0.599 and a subjectivity score of 0.605. This is a positive score and it appears very logical as this review's tone is very positive with words such as 'excellent'.

The second review tested was the 19th one. It reads "I was looking for a kindle whitepaper. I saw online for $80. What a deal. I ordered it on line and picked it up in the store. I got it home and couldn't adjust the brightness. After a lengthy time with online customer service I called customer service. After 20 minuets with speaking to a female Elmer Fud that doesn't speak english well I figured I would just return it. Although it looks Identical to the $120 model, you can not adjust the brightness. That would have been good information before I bought it." The overall tone is very negative with very little compliment given for the product. The polarity score returned was 0.2333 and subjectivity score of 0.255 which appears to be very accurate.

The final review tested was the 28th one. It reads " I have several Kindles so knew what to expect. It is a great reading device and I keep several in different locations." This review is quite ambiguous. The product may be described as 'great' but the language is definitely a lot more general. The polarity score was given a 0.4 and a pretty high subjectivity score was 0.675. I feel that the polarity score is low but the subjectivity score makes a lot of sense.

Overall I found that my sentiment function provided solid results. I feel as if though that results become slightly questionable when the language of the review isn't obvious and simple. When there is a sarcastic tone or the writer talks about other products, the model may struggle to pick up on it.