Amazon Sentiments: Analyzing Book Review Sentiments using Machine Learning

Your four years of hard work in the data science building have finally paid off. You have graduated from Uva and now work for the multinational corporation Amazon as a data analyst. You are assigned to the Book sector, Amazon's initial focus as a third-party retailer. It is your first week on the job and your boss approaches you with your first assignment.

Your boss has noticed what she believes to be several trends in the preferences of book buyers on Amazon. Among other trends, your boss specifically thinks that overall review sentiments are unproportionately low (harsher, less positive) compared to star ratings and that specific book categories and price categories are highly associated with and indicative of customer sentiments. Your boss asks you to investigate the recent haul of data to make an assessment as to whether their intuition is correct.

The only problem... you have never worked with string data or sentiment analysis packages before.

It is your first assignment. Your reputation with your boss is important, and you want to derive important insights regarding this question they have posed. Luckily they have already provided you with the data.

All you need to figure out is how to preprocess the data, implement Vader scores, perform sentiment analysis, and use predictive modeling to highlight the importances and predictive power of each variable.

Your task:
1. Clean/preprocess the data to prepare a final cleaned dataset for analysis.
2. Assign Vader sentiment scores as a new variable, and create other possibly impactful feature-engineered variables (text word count, helpfulness breakdown)
3. Create visualizations and perform sentiment analysis
4. Experiment with predictive modeling to further support variable importances

Deliverables:
1. Cleaned, finalized csv dataset – ready for analysis
2. 1-page document evaluating results/conclusions
3. One file containing all source code

4. README summary/contextualization of project