1. **Exploratory Data Analysis (EDA):**
   * Your EDA section provides a good overview of the dataset, including the distribution of review lengths, rating distributions, and word clouds.
   * Consider adding more visualizations or analyses to gain deeper insights into the dataset, such as the distribution of review sentiments across different product categories or trends over time.
2. **Data Preprocessing:**
   * Your data preprocessing steps, such as removing null values and cleaning reviews, are essential for preparing the text data for analysis.
   * Consider exploring additional preprocessing techniques like lemmatization, stemming, or handling special characters to further refine the text data.
3. **Sentiment Analysis with VADER:**
   * VADER sentiment analysis provides a quick way to gauge the sentiment of reviews. Your visualization of sentiment scores and their correlation with ratings is insightful.
   * It might be helpful to analyze the distribution of VADER sentiment scores across different rating categories to see if there are any patterns or discrepancies.
4. **Machine Learning with DistilBERT:**
   * Fine-tuning DistilBERT for sentiment analysis is a powerful approach, and your implementation is well-documented.
   * Consider experimenting with hyperparameters or model architectures to potentially improve performance further. You could explore different learning rates, batch sizes, or fine-tuning strategies.
   * The evaluation metrics (accuracy, precision, recall, F1-score) provide a comprehensive view of the model's performance on the test set.
5. **Prediction on Unseen Data:**
   * The function to predict ratings for unseen text using the trained DistilBERT model is a valuable addition. It allows for real-world applications of the model.
   * Consider providing examples of predicted ratings for different types of reviews to showcase the model's versatility.
6. **Model Persistence:**
   * Pickling the trained model for future use is a good practice. Ensure that the pickled model can be loaded and used seamlessly in other environments or applications.

Overall, your code demonstrates a solid understanding of natural language processing techniques and machine learning models for sentiment analysis. Continuing to refine and experiment with different approaches will help improve the robustness and accuracy of your analyses.

Top of Form