Sentiment Analysis Report

1. Dataset Description

The dataset used in this analysis consists of Amazon product reviews. It features several columns, including product IDs, review texts, and ratings. For this project, our primary focus was on the 'reviews.text' column, which contains the textual content of customer reviews.

2. Preprocessing Steps

Preprocessing involved cleaning the dataset by removing missing values from the 'reviews.text' column to ensure the analysis was conducted on complete data. Furthermore, certain columns were cast to specific data types to avoid mixed-type data warnings and ensure consistency for analysis.

3. Evaluation of Results

The sentiment analysis was conducted using spaCy's 'en_core_web_sm' model with the SpacyTextBlob component. The analysis categorized reviews into Positive, Negative, or Neutral based on the polarity score. Initial tests on a sample of reviews showcased the model's capability to discern sentiment accurately.

4. Model Insights

The spaCy model with SpacyTextBlob provides a straightforward approach to sentiment analysis,

suitable for quickly gauging the sentiment of textual data. Its strengths lie in its ease of use and integration into Python workflows. However, the reliance on the 'en_core_web_sm' model may limit the depth of linguistic features analyzed. Upgrading to a larger spaCy model could enhance accuracy by leveraging more comprehensive word vectors.