Description of the Dataset

The dataset used contains customer reviews for a number of amazon products. The dataset includes information such as product, product ID, date of review, star rating, is the review helpful. For this project I have only used the item review (review.text column) as I am using spacy to analyse the text contained in the review. Therefore this is the only column needed.

Details of the Preprocessing Steps

To preprocess the dataset I first of all removed any rows that contained NaN values, then renamed the column so that it was easier to access. I then converted the column to string to be able to use string methods of which I used lower, strip and split. This meant that the reviews were now ready to be tokenized. Before tokenizing, I took a sample of the reviews to work with a smaller dataset.
I then created a function to iterate through the sample of the dataset and tokenize each review. The dataset was then ready to use the spacy text blob features polarity and sentiment.

Evaluation of Results

The sentiment and polarity scores are similar or the same every time I run the code, therefore I will only refer to the sentiment score, rather than both. The scores match the positivity of the reviews, I can see this because I printed the reviews that I chose so that I could compare them with the scores.

When the sentiment scores are 0.4+, the reviews have words such as enjoy, great, like, love. This matches the scores as they are closer to 1 on the scale of -1 to 1 (-1 being negative and 1 being positive).

The reviews which only have words relating to the product such as kindle, book, features or words that show that the product works or is convenient, such as prefer, working, ease have scores closer to 0, which shows that they are neutral.

I have found one negative review which includes the words bad and disappointing and the sentiment score is -0.05 (2dp), although a negative number it is closer to 0 so suggests a neutral review, this is because it also includes words such as quality, works, surprises.

The subjectivity score, which has a range of scores from 0 to 1, 0 being objective (factual) and 1 being subjective (opinion) also reflects the words in the reviews. The reviews that contain more abstract nouns such as love or adjectives such as easy, have a higher score. Reviews that include product details such as 4 years old, new version and upgrade have a lower score. The subjectivity scores tend to be higher than the sentiment scores.

Overall the reviews are neutral to positive

Insights into the Models Strengths and Limitations

Strengths
The model recognises positive and negative language and the scores generally reflect a positive, neutral or negative review. The subjectivity score is higher when emotive or descriptive language is used and is lower when the review contains words features of the product. This shows that the model works well in identifying language and assigning appropriate scores.

Limitations

Although I can see how the scores reflect the reviews, there are instances in the reviews where the language used contains features about the product such as the review below.

Love', 'control', 'lights', 'Phillips', 'bulbs', 'compatible', 'smart', 'plugs', 'control', 'electronics', 'house', '.', 'Definitely', 'recommend

This has a medium subjectivity score of 0.58 (2dp), the review is very positive as the features are used to show the positive functionality of the product. The sentiment score is 0.24 (2dp) which suggests a neutral review. This could be because the model is analysing individual words and 'definitely' and 'recommend' could be positive or negative depending on the other words in the phrase.

The model would work more effectively in giving accurate scores if it identified specific phrases relating to each product showing that it worked well. These are the type of reviews that are useful for the consumer as they give information to help them to decide whether the product is value for money. Reviews that use emotive language give more of a personal opinion which isn't as useful.

As I have already stated above, the subjectivity score tends to be higher than the sentiment score, therefore the model gives a more neutral review score as the language in the review is more factual or objective. This isn't necessarily true.

This would lead me to analyse the dataset in more detail and include the star rating to compare this with the sentiment score.