4.1 The dataset contains reviews from different users of different products sold by Amazon. Products range from Office supplies, health & beauty products, furniture and electronics. Each review has a different id, it also includes details as to when the review was done and reviewed. Also includes information such as the link to the review, what was written by the customer, the manufacturer of the product and the brand of product.

4.2 There are a few fields that are missing as can be seen above. What is important is the 5 missing usernames. Those are probably people that do not have accounts. Nevertheless I am not going to drop the username field because this assignment only requires me to do a sentiment analysis, a more high-level view rather than user specific. The following columns were dropped - "reviews.didPurchase", "reviews.doRecommend", "reviews.id", "reviews.numHelpful".

4.3 Best way to evaluate the model is to take the few test example used. In all three examples the model managed to correctly identify whether the feedback was positive or negative. Although this is a very basic the model is able to distinguish general emotional tone in the customer reviews.

4.4 In terms of the model's strengths, is that is it quick to implement. SpaCy does not have a built-in sentiment model but the NLP capability makes it possible to clean text in terms od lemmatization and tokenisation to ensure the consistent text processing.
However, the model may struggle with context because it can't distinguish sarcasm and or negation. The model also can't distinguish between the weighting of words.