Sentiment Analysis Report

For this capstone project we had to develop a python program that performs sentiment analysis on a large dataset of product reviews.
The objective of this sentiment analysis was to evaluate the general sentiment people held towards amazon products, this was done using natural processing language (NPL) application.

Methods(s):

- I downloaded a dataset of Amazon product reviews.and saved it as a csv file so that it can be processed on the IDE (VS Code). This file contained text-base reviews along with metadata
- Installed and imported python libraries required to carry out the sentiment analysis such as spaCy, SpacyTextBlob, pandas. Also downloaded the spaCy english language model en_core_web_m
- Preprocessing stage - before passing the data through to the spaCy model it must go through a preprocessing stage. Preprocessing stage included handling the missing values inside of the data, lowercase conversion, stopword removal, punctuation removal, training and testing
- After the preprocessing stage, sentimental analysis was performed using SpacyTextBlob tool, which provided polarity scores for each and every review based on sentiment of the data. Reviews were classed as either positive, negative, or neutral depending on the polarity scores.

Results

Majority of the reviews expressed positive sentiment towards the amazon products with keywords relating to the quality, duration, value and customer service. Some reviews expressed negative sentiment, with keywords relating to bad quality, bad customer service, price etc. An even smaller number of reviews were classed as neutral, with no strong opinions on either side, resulting in neutral sentiment.

Conclusion

Overall, the model did what it needed to do, it performed the basics well and made it easier for me as a programmer to implement NLP techniques without extra custom coding or algorithm development. However, it does have its limitations. The model relies on pre-built libraries and sentiment analysis algorithms, limiting customisation options for fine-tuning or adapting the model to specific domains. It also had questionable accuracy especially when it came to capturing nuanced sentiments.