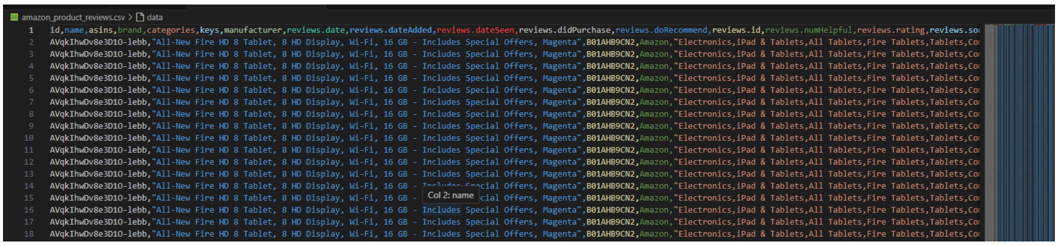
5.1. A description of the dataset used.

The dataset that I am using are the reviews of user feedback on a product on amazon. There are almost 3,000 reviews for one product and over 30,000 reviews in this dataset. As the image below shows how the header column splits each category. The dataset that I will be looking into is the review.text category.

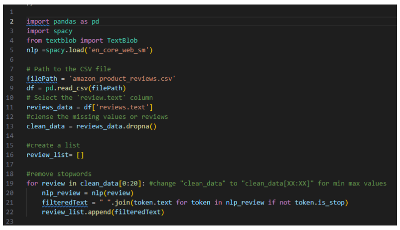

```
amazon_product_reviews.csv > data
id,name,asins,brand,categories,keys,manufacturer,reviews.date,reviews.dateAdded,reviews.dateSeen,reviews.didPurchase,reviews.doRecommend,reviews.id,reviews.numHelpful,reviews.rating,reviews.sou
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Col 2: name cial Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
AVqkIhwDv8e3D1O-lebb,"All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta",B01AHB9CN2,Amazon,"Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Co
```

5.2. Details of the preprocessing steps.

To facilitate the preprocessing steps, we'll start by importing the necessary tools. We'll use Pandas to read the CSV file, spaCy for natural language processing, and TextBlob for sentiment analysis. Since dealing with a large DataFrame could be cumbersome, we'll streamline the process by initially focusing on the 'reviews.text' column. This approach helps us efficiently manage the data we need to analyse. After importing the tools, the next step involves data cleansing. We'll use the '.dropna()' method to remove any missing information or values from the dataset, minimising the potential for errors during analysis.

Following data cleansing, we'll create a list specifically for storing the reviews. This list will undergo preprocessing steps such as removing stopwords and tokenization, which aids in sentiment analysis and polarity identification.

Additionally, we'll implement a function that enables users to input a review for analysis. This function will output the sentiment score along with a classification indicating whether the sentiment is positive, negative, or neutral.

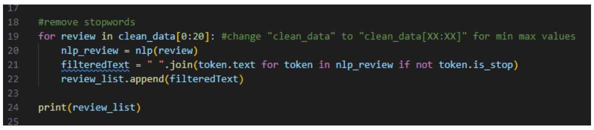

```python

import pandas as pd
import spacy
from textblob import TextBlob
nlp =spacy.load('en_core_web_sm')

# Path to the CSV file
filePath = 'amazon_product_reviews.csv'
df = pd.read_csv(filePath)
# Select the 'review.text' column
reviews_data = df['reviews.text']
#clense the missing values or reviews
clean_data = reviews_data.dropna()

#create a list
review_list= []

#remove stopwords
for review in clean_data[0:20]: #change "clean_data" to "clean_data[XX:XX]" for min max values
    nlp_review = nlp(review)
    filteredText = " ".join(token.text for token in nlp_review if not token.is_stop)
    review_list.append(filteredText)
```

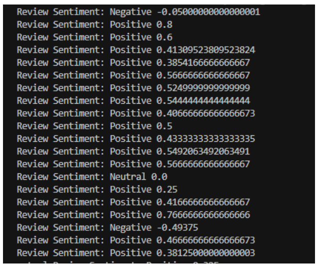

```python

#remove stopwords
for review in clean_data[0:20]: #change "clean_data" to "clean_data[XX:XX]" for min max values
    nlp_review = nlp(review)
    filteredText = " ".join(token.text for token in nlp_review if not token.is_stop)
    review_list.append(filteredText)

print(review_list)
```

5.3. Evaluation of results.

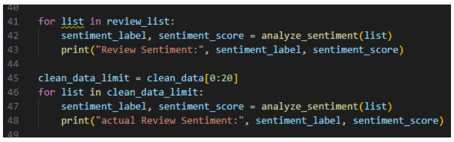
In conclusion of my results, my model works. It is able to produce positive, negative and neutral results with their corresponding polarity. I have tested it with the first 20 results and as the picture showed how it would look. Looking at these results the polarity seems to sit in the region of 0.3-0.5.



```
Review Sentiment: Negative -0.05000000000000001
Review Sentiment: Positive 0.8
Review Sentiment: Positive 0.6
Review Sentiment: Positive 0.41309523809523824
Review Sentiment: Positive 0.3854166666666667
Review Sentiment: Positive 0.5666666666666667
Review Sentiment: Positive 0.5249999999999999
Review Sentiment: Positive 0.5444444444444444
Review Sentiment: Positive 0.40666666666666673
Review Sentiment: Positive 0.5
Review Sentiment: Positive 0.43333333333333335
Review Sentiment: Positive 0.5492063492063491
Review Sentiment: Positive 0.5666666666666667
Review Sentiment: Neutral 0.0
Review Sentiment: Positive 0.25
Review Sentiment: Positive 0.4166666666666667
Review Sentiment: Positive 0.7666666666666666
Review Sentiment: Negative -0.49375
Review Sentiment: Positive 0.46666666666666673
Review Sentiment: Positive 0.38125000000000003
```

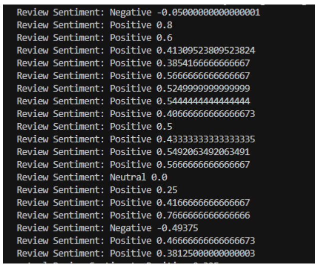
5.4. Insights into the model's strengths and limitations.

With my model I decided to check what would happen if I did not remove any stop words as you could see in the pictures below. I would be able to conclude that the strength of my model with no stop words helps the accuracy of the model. However , the limitation would be with the small library with spacy 'en_core_web_sm' rather than having 'en_core_web_md'. With a larger library this might be able to produce much more accurate results to my model.

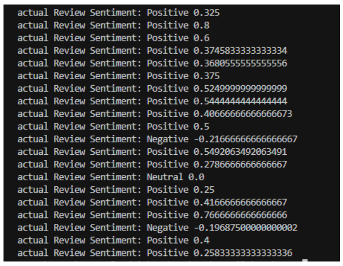


```python
for list in review_list:
    sentiment_label, sentiment_score = analyze_sentiment(list)
    print("Review Sentiment:", sentiment_label, sentiment_score)

clean_data_limit = clean_data[0:20]
for list in clean_data_limit:
    sentiment_label, sentiment_score = analyze_sentiment(list)
    print("actual Review Sentiment:", sentiment_label, sentiment_score)
```



```
Review Sentiment: Negative -0.05000000000000001
Review Sentiment: Positive 0.8
Review Sentiment: Positive 0.6
Review Sentiment: Positive 0.41309523809523824
Review Sentiment: Positive 0.3854166666666667
Review Sentiment: Positive 0.5666666666666667
Review Sentiment: Positive 0.5249999999999999
Review Sentiment: Positive 0.5444444444444444
Review Sentiment: Positive 0.40666666666666673
Review Sentiment: Positive 0.5
Review Sentiment: Positive 0.43333333333333335
Review Sentiment: Positive 0.5492063492063491
Review Sentiment: Positive 0.5666666666666667
Review Sentiment: Neutral 0.0
Review Sentiment: Positive 0.25
Review Sentiment: Positive 0.4166666666666667
Review Sentiment: Positive 0.7666666666666666
Review Sentiment: Negative -0.49375
Review Sentiment: Positive 0.46666666666666673
Review Sentiment: Positive 0.38125000000000003
```



```
actual Review Sentiment: Positive 0.325
actual Review Sentiment: Positive 0.8
actual Review Sentiment: Positive 0.6
actual Review Sentiment: Positive 0.3745833333333334
actual Review Sentiment: Positive 0.3680555555555556
actual Review Sentiment: Positive 0.375
actual Review Sentiment: Positive 0.5249999999999999
actual Review Sentiment: Positive 0.5444444444444444
actual Review Sentiment: Positive 0.40666666666666673
actual Review Sentiment: Positive 0.5
actual Review Sentiment: Negative -0.21666666666666667
actual Review Sentiment: Positive 0.5492063492063491
actual Review Sentiment: Positive 0.2786666666666667
actual Review Sentiment: Neutral 0.0
actual Review Sentiment: Positive 0.25
actual Review Sentiment: Positive 0.4166666666666667
actual Review Sentiment: Positive 0.7666666666666666
actual Review Sentiment: Negative -0.19687500000000002
actual Review Sentiment: Positive 0.4
actual Review Sentiment: Positive 0.25833333333333336
```

(These images are from the code from the left, producing results to the right.)

```python
import pandas as pd
import spacy
from textblob import TextBlob
nlp =spacy.load('en_core_web_sm')

# Path to the CSV file
filePath = 'amazon_product_reviews.csv'
df = pd.read_csv(filePath)
# Select the 'review.text' column
reviews_data = df['reviews.text']
#clense the missing values or reviews
clean_data = reviews_data.dropna()

#create a list
review_list= []

#remove stopwords
for review in clean_data[0:20]: #change "clean_data" to "clean_data[XX:XX]" for min max values
    nlp_review = nlp(review)
    filteredText = " ".join(token.text for token in nlp_review if not token.is_stop)
    review_list.append(filteredText)

print(review_list)

def analyze_sentiment(review):
    # create a TextBlob object
    blob = TextBlob(review)
    # get the polarity score (where -1 is very negative, 0 is neutral, and 1 is very positive)
    polarity = blob.sentiment.polarity
    # Determine sentiment label based on the polarity score
    if polarity > 0:
        sentiment_label = "Positive"
    elif polarity < 0:
        sentiment_label = "Negative"
    else:
        sentiment_label = "Neutral"
    # Return both the sentiment label and the polarity score
    return sentiment_label, polarity

for list in review_list:
    sentiment_label, sentiment_score = analyze_sentiment(list)
    print("Review Sentiment:", sentiment_label, sentiment_score)

clean_data_limit = clean_data[0:20]
for list in clean_data_limit:
    sentiment_label, sentiment_score = analyze_sentiment(list)
    print("actual Review Sentiment:", sentiment_label, sentiment_score)
```

```
c:\Users\Fractal Admin\Desktop\python File\task 21\fixed.py:8: DtypeWarning: Columns (1,10) have mixed types. Specify dtype option on import or set low_memory=False.
  df = pd.read_csv(filePath)
['product far disappointed . children love use like ability monitor control content ease .', 'great beginner experienced person . Bought gift loves', 'Inexpensive tablet use learn , step NABI . thrilled , learn Skype ..
.', 'Fire HD 8 weeks love . tablet great value . Prime Members tablet SHINES . love able easily access Prime content movies download watch laterThis 1280/800 screen nice look nice crisp bright infact brighter ipad pro c
osting $ 900 base model . build fire INSANELY AWESOME running 7.7 mm thick smooth glossy feel amazing hold like futuristic tab ur hands .', 'bought grand daughter comes visit . set user , entered age Amazon makes sure a
ccesses sites content appropriate age . Simple loves capabilities . bought installed 64gig SD card gives little tablet plenty storage . price think tablet best . spend hundreds dollars additional speed capacity comes ba
sics tablets people need fraction cost .', 'amazon fire 8 inch tablet perfect size . purchased husband bigger screen phone . gotten years ago knew good purchase .', 'Great e - reading , nice light weight , price point g
iven , definitely worth purchase .', 'gave Christmas gift inlaws , husband uncle . loved easy use fantastic features !', 'Great device read books . like links borrowed library e - books . Switched popular tablet brand h
appy choice . took time books previous non - Kindle reader , finally figured way !', 'love ordering books reading reader .', 'easy elderly users cease ads pop .', 'Excellent product . Easy use , large screen makes watch
ing movies reading easier .', 'Wanted father tablet good value . watch movies play games . Easy use .', 'Simply need . Thank youAnd silk works wonders', 'Got present love size screen', 'kindle easiest use , graphics scr
een crisp , clear , brilliant colors .', 'nice reader . perfect want / need . good bargain', 'like tablet . given 5 stars push start times unlock screen little annoying .', 'Great video quality lots fun apps fun family'
, 'Love love love kindle fire 8 ..... 9 yr old granddaughter said bought Christmas .. purchased kindled past best .... love fact use memory card ... helps use books ... games ... music viedos ... thanks']
Review Sentiment: Negative -0.05000000000000001
Review Sentiment: Positive 0.8
Review Sentiment: Positive 0.6
Review Sentiment: Positive 0.41309523809523824
Review Sentiment: Positive 0.3854166666666667
Review Sentiment: Positive 0.5666666666666667
Review Sentiment: Positive 0.5249999999999999
Review Sentiment: Positive 0.5444444444444444
Review Sentiment: Positive 0.40666666666666673
Review Sentiment: Positive 0.5
Review Sentiment: Positive 0.43333333333333335
Review Sentiment: Positive 0.5492063492063491
Review Sentiment: Positive 0.5666666666666667
Review Sentiment: Neutral 0.0
Review Sentiment: Positive 0.25
Review Sentiment: Positive 0.4166666666666667
Review Sentiment: Positive 0.7666666666666666
Review Sentiment: Negative -0.49375
Review Sentiment: Positive 0.46666666666666673
Review Sentiment: Positive 0.38125000000000003
actual Review Sentiment: Positive 0.325
actual Review Sentiment: Positive 0.8
actual Review Sentiment: Positive 0.6
actual Review Sentiment: Positive 0.3745833333333334
actual Review Sentiment: Positive 0.3680555555555556
actual Review Sentiment: Positive 0.375
actual Review Sentiment: Positive 0.5249999999999999
actual Review Sentiment: Positive 0.5444444444444444
actual Review Sentiment: Positive 0.40666666666666673
actual Review Sentiment: Positive 0.5
actual Review Sentiment: Negative -0.21666666666666667
actual Review Sentiment: Positive 0.5492063492063491
actual Review Sentiment: Positive 0.2786666666666667
actual Review Sentiment: Neutral 0.0
actual Review Sentiment: Positive 0.25
actual Review Sentiment: Positive 0.4166666666666667
actual Review Sentiment: Positive 0.7666666666666666
actual Review Sentiment: Negative -0.19687500000000002
actual Review Sentiment: Positive 0.4
actual Review Sentiment: Positive 0.25833333333333336
PS C:\Users\Fractal Admin\Desktop\python File\task 21>
```

(Image 1 and 2, are the code and terminal results to help explain)