Sentiment Analysis: Amazon Reviews

BY

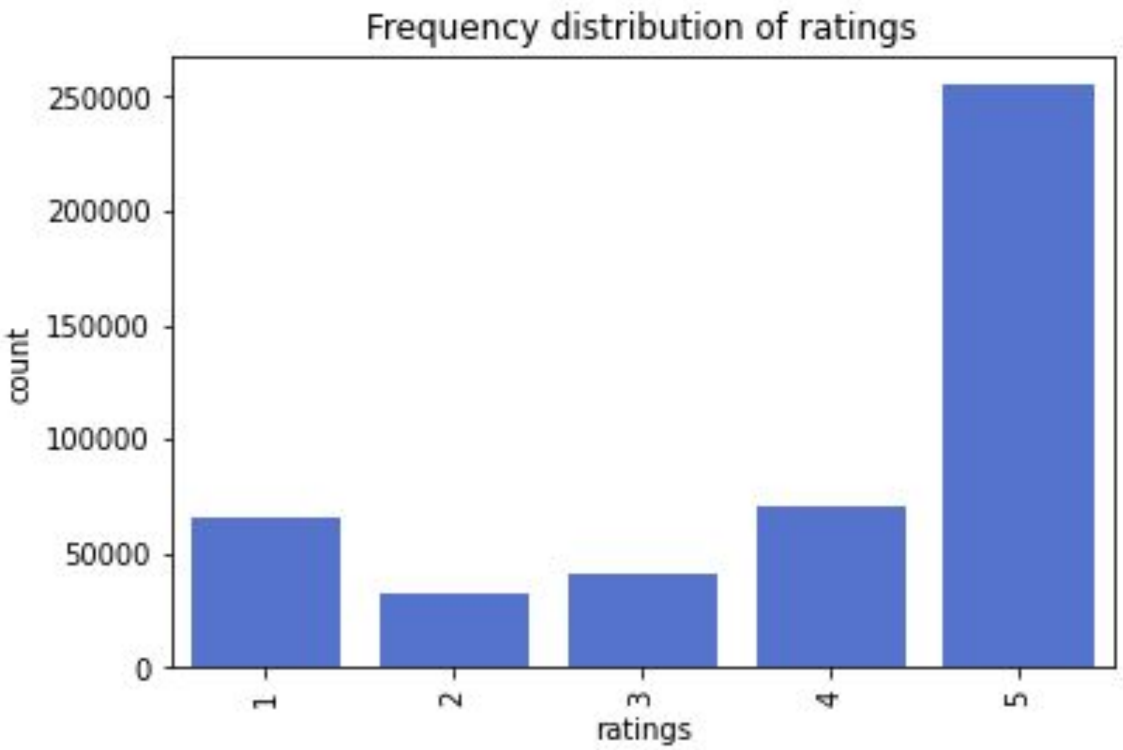
KETAKI NAGARKAR

Introduction

- There is a significant increase in online shopping and e-commerce
- Businesses can derive Important insights from reviews which will help to improve their product and profits
- Amazon reviews and social media feed form sites like Facebook and twitter are a good source of product reviews
- The project will use an open-source Amazon *cell phones and accessories products* reviews dataset form 2016
- Goals: Sentiment analysis and prediction of product sentiment/rating based on text reviews.

Data Extraction and Preprocessing:

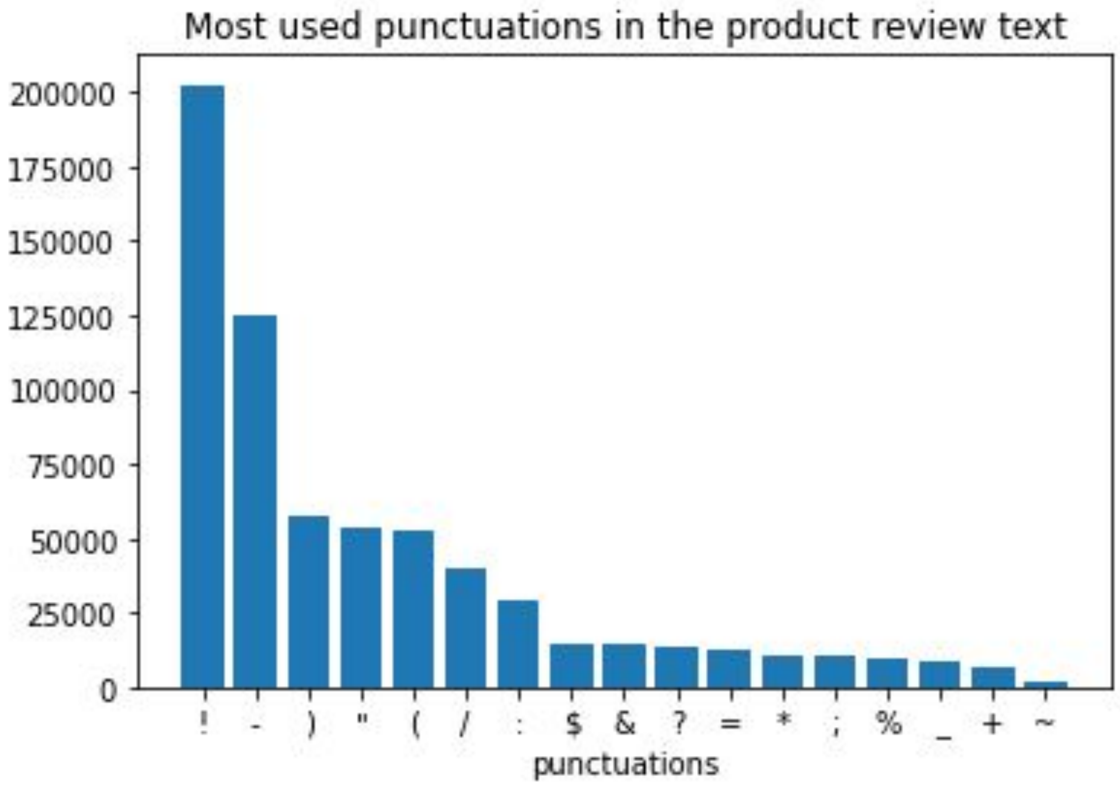
- Dataset was downloaded as a csv file and 10% random fraction was read into a pandas dataframe (500,000 rows)
- Some important columns were *review text, summary, reviewer name, time of review, overall rating* and *product asin*
- Majority of the review text were in English, these were kept, the rest removed
- Numbers and punctuations were extracted
- Text was converted into lower case, stop words were removed, text was lemmatized



EDA: Ratings

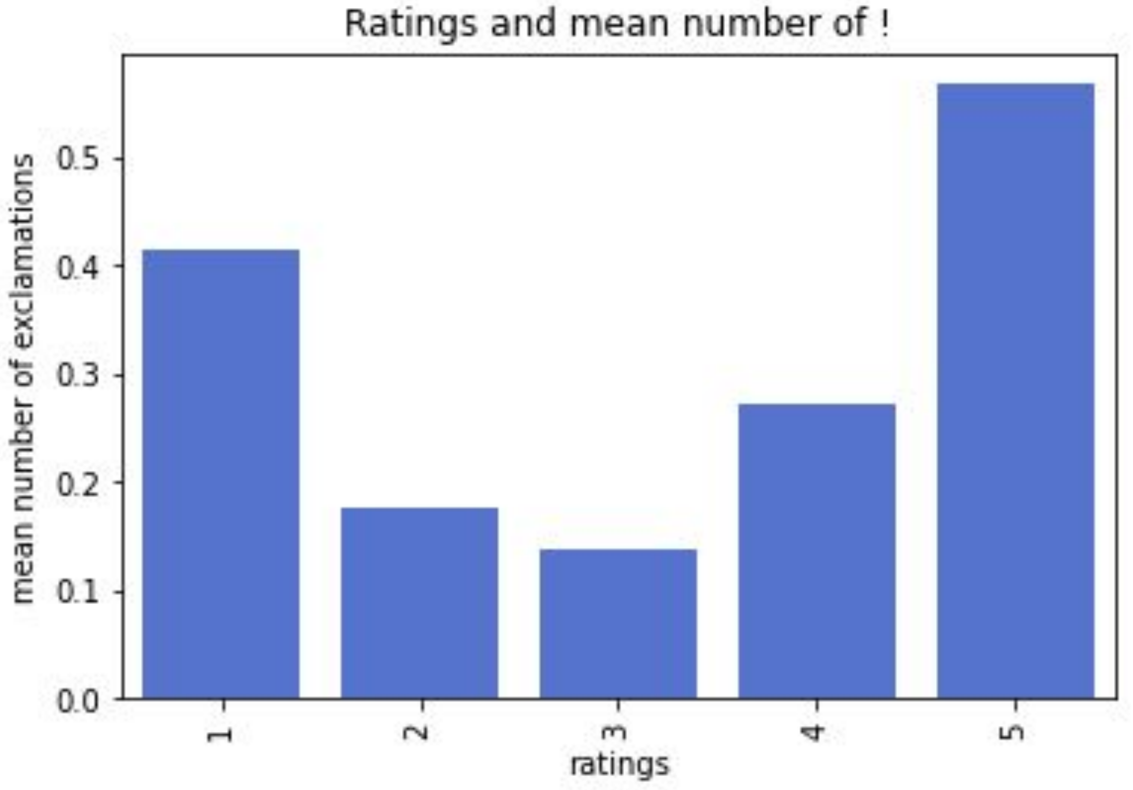
- Ratings were numbered 1 through 5
- 1 being the lowest, 5 being the highest
- 5 was the most common rating compared to 1,2,3 and 4

EDA: Punctuations



- Punctuations were extracted
- Periods, commas were removed (like stop words)
- Exclamations stood out as a commonly used punctuation

EDA: Exclamations



- Ratings of 5 and 1 had the highest mean number of exclamations in a review text
- One way ANOVA test amongst the five rating groups for number of exclamations yields a very low p value ($<.001$)

EDA: Examples of tweets with !!!!!

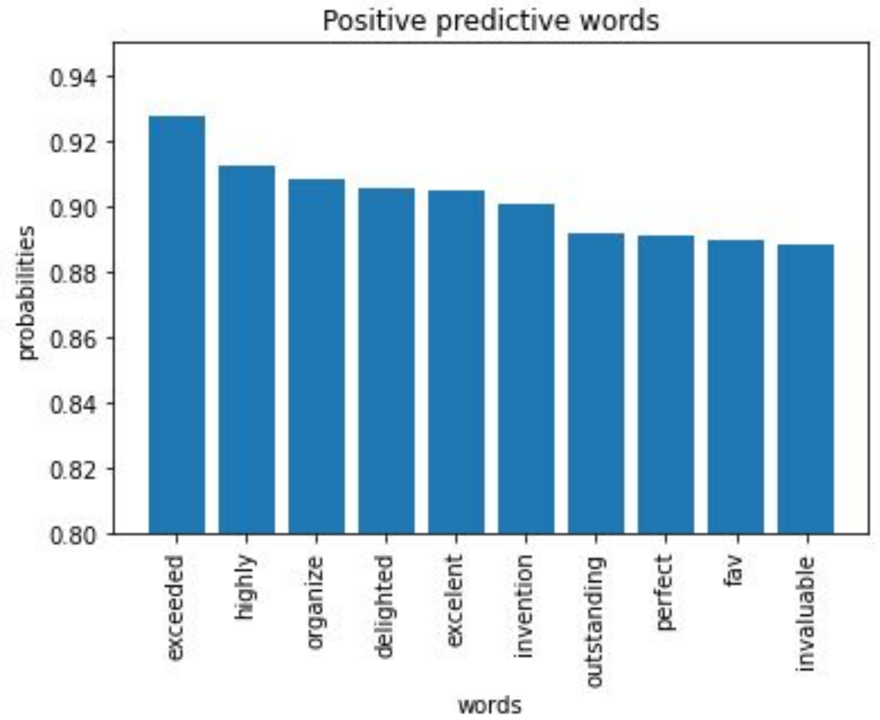
Example of a review with a positive sentiment and high number of exclamations:

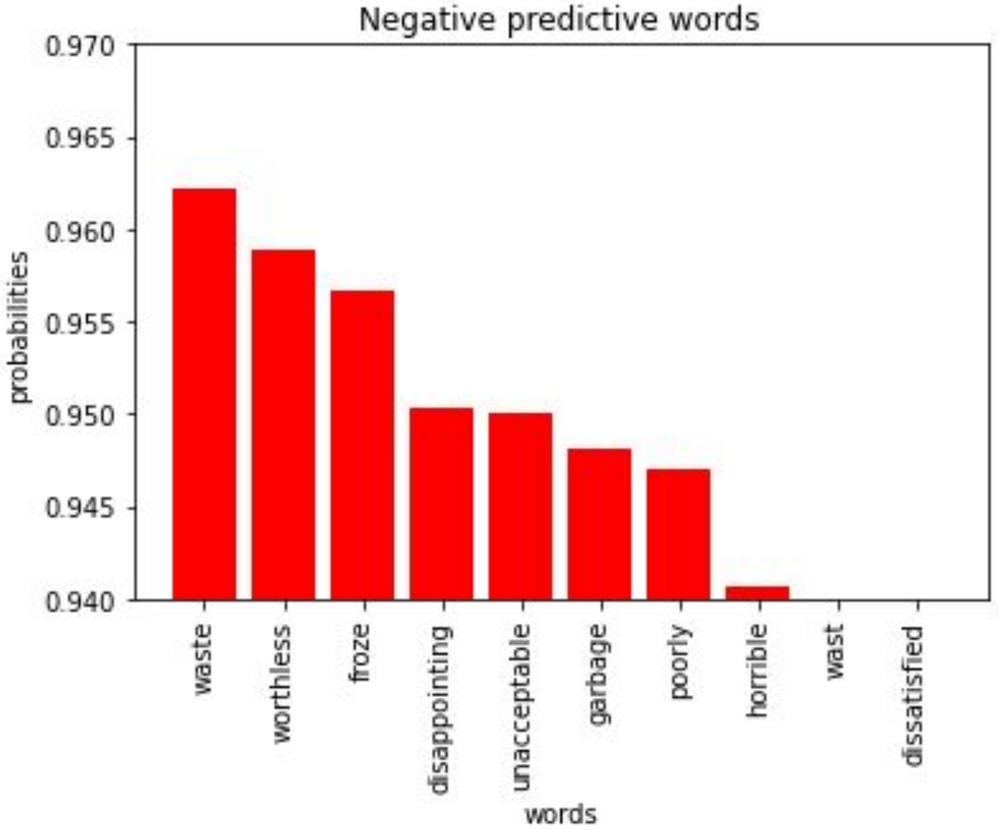
It's a very cute phone case but, it doesn't have a lip. It so far is protective,the picture doesn't come off or fade. I LOVE THIS PHONE CAESE SO MUCH!!!

Example of a review with a negative sentiment and high number of exclamations:

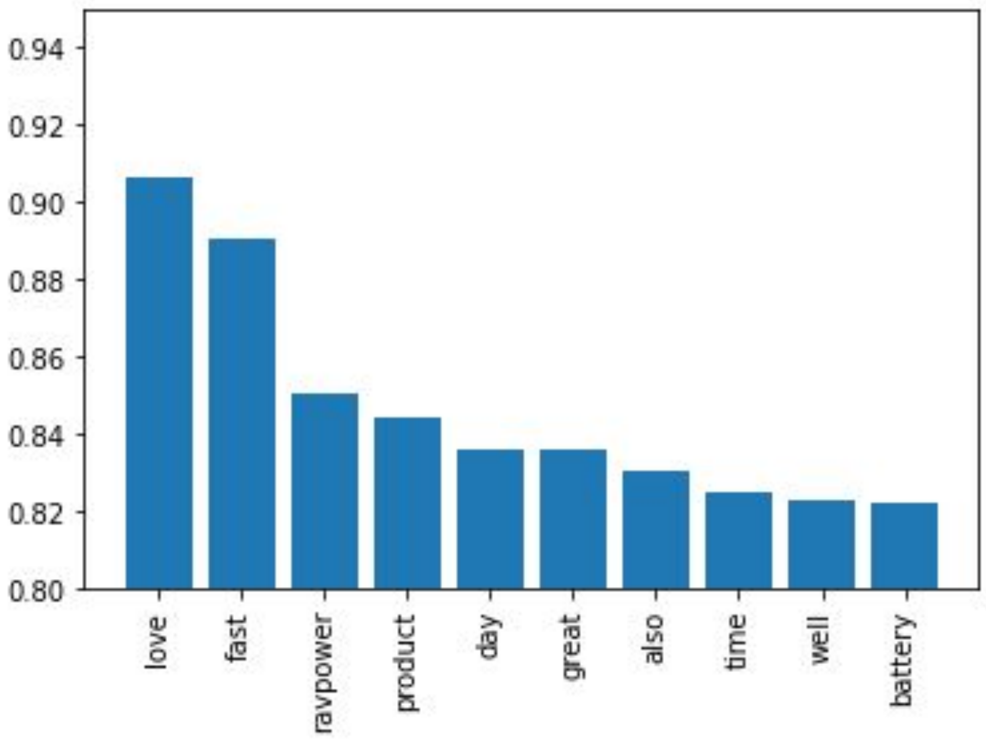
Don't know what they've done with the latest/greatest version of the Blue Parrot but it's junk ! Noise cancelling capability is good but the sound is so garbled it sound like people are talking to you with they're mouth full of oatmeal. My older model Blue Parrot was impressive in it's sound quality. DO NOT BUY THIS THING IT"S JUNK !!

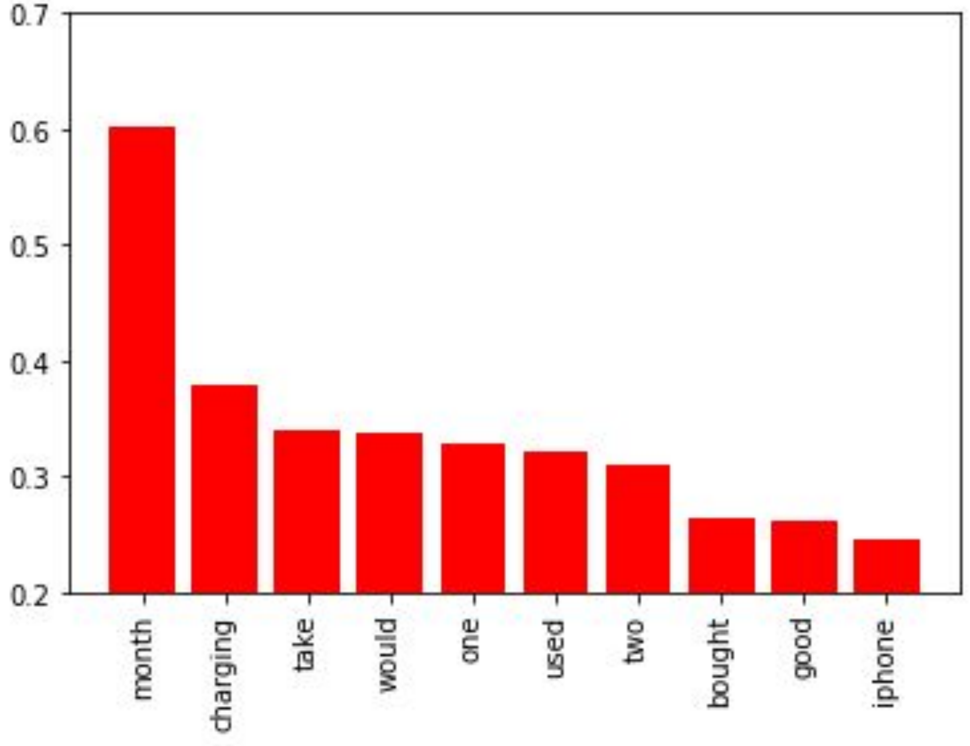
EDA: Predictive words for a sentiment (Cell Phones and Accessories)





EDA: Predictive words for a sentiment (RavPower charger)





EDA results:

- Products had a rating of 5 more often than a rating of 4 or lower
- Use of punctuations such as exclamation marks are predictive of a strong negative or positive sentiment
- A higher use of ? is associated with lower ratings
- For cell phone accessories, some interesting words predictive of a positive sentiment are ‘invention’, ’invaluable’ and ‘organize’
- For the RavPower charger for a cell phone some interesting predictive words for negative sentiment were ‘take’, ’month’ and ‘iphone’

Modeling: Text and Numerical features

- A random fraction subset was used for modeling
- CountVectorizer and TfidfVectorizer were and compared to find the best vectorizer
- Training and Test data was split
- From the training and test sets, text and numerical features were extracted
- CountVectorizer was fit using Text features from the training set, then the text features from the test set was fitted and transformed, these were then converted into a dense array
- Numerical features from the training and test set were transformed used MinMaxScaler()
- Before fitting the models, transformed text and numerical features were then concatenated into the respective Training and Test sets.

Modeling: ML algorithms and Metrics

- Multinomial naive bayes, Random Forest Classifier, and Logistic Regression classifier were initially grid searched for best parameters
- The best parameters were then applied to the final models
- For model assessment, two metrics were taken into consideration, overall balanced accuracy and f1 score for positive sentiment.

Modeling: Scenario 1

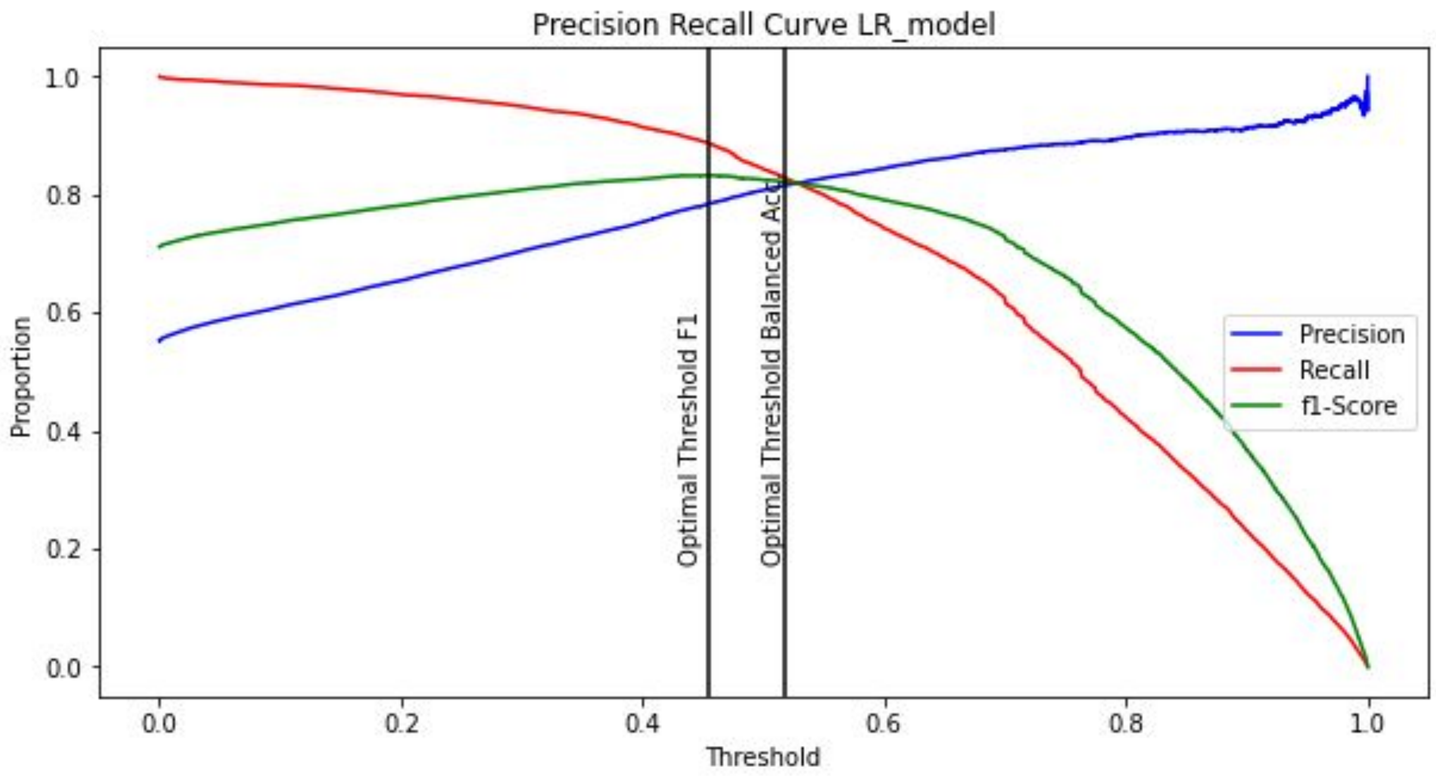
Comparing Balanced Accuracy score at best threshold (Business perspective)

model	balanced accuracy	best threshold	parameters	best model for balanced accuracy
Multinomial Naive Bayes	0.79819	0.565	alpha=0.9	
Random Forest Classifier	0.79766	0.499	bootstrap=False,min_samples_leaf=2,min_samples_split=10,n_estimators=300	
Logistic Regression Classifier	0.79933	0.518	penalty='l2',C=0.1	Logistic Regression Classifier

Modeling: Scenario 2

Comparing f1 score for positive class (Consumer perspective)

model	f1 score	best threshold	parameters	best model for f1 score
Multinomial Naive Bayes	0.82959	0.515	alpha=0.9	
Random Forest Classifier	0.83159	0.432	bootstrap=False, min_samples_leaf=2,min_samples_split=10,n_estimators=300	
Logistic Regression Classifier	0.83269	0.454	penalty='l2',C=0.1	Logistic Regression Classifier



Results:

- Logistic Regression Classifier performs better compared to the other two models
- Higher Balanced Accuracy score of 0.79933 at an optimal threshold of 0.518
- higher f1 score of 0.83269 for the positive class at an optimal threshold of 0.454

Discussion:

- Using the above EDA techniques, by examining use of the exclamations, highly satisfied or dissatisfied consumer reviews can be extracted in larger datasets.
- RavPower charger product, ‘iphone’ came up as one of the words indicating a negative sentiment
- RavPower review example with the word ‘iphone’ ***‘IPhone adapter not fit’***
- Based on the review, RavPower can make changes to their chargers to make them more compatible with iPhone adapters.
- Similar to the RavPower charger product, the modeling can be used to discover predictive words and review insights whereby product feedback can be given to the business.
- Modeling can be used to derive consumer sentiment based on social media feed and reviews for products

Next Steps:

1. For further study, additional kinds of punctuations, and special characters such as emojis can be explored to gain insights into their influence on sentiments.
2. Other product type amazon review datasets, or social media feed datasets should be studied.
3. The findings across groups of products should be compared for similarities and differences of how certain features influence sentiment.