



NLP Analysis

EXAMINING REVIEWS ON MULTIPLE PRODUCTS AND PREDICTING
POSITIVE OR NEGATIVE REVIEWS ON UNLABELED REVIEWS

Background

Many businesses are taking great strides to identify what their customers think of their services and products. NLP is a great way to analyze reviews in order to extract actionable insights into their product lines.





The Process

This presentation aims to give insights into a product review dataset and evaluate a ML model and Sentiment Analysis to predict positive reviews.

- It contains two datasets (one has the star label and the other is unlabeled)
- The steps of this analysis are as follows:
 - *Text Cleaning*
 - *Stop Word Removal*
 - *Tokenization*
 - *Lemmatization and POS Tagging*
 - *Training Logistic Regression Models*
 - *Predicting on unlabeled data*
 - *Using NLTK Sentiment Intensity Analyzer*
 - *Evaluate the approaches and give insights into positive and negative reviews*



The Data

Here is a sample of the raw labeled data without any transformations

	ProductId	Score	Text	Summary
0	B004K30HO2	5	This product works exactly as it is advertised. No more rinsing wire baskets. The coffe is as good as the already filled k cups that you buy in the store. It's nice to just throw the whole thing in the trash when done and not clog the sink. Couldn't be happier as it gives me unlimited choices for coffee and it saves money too! Amazon is great!	Disposable k cups
1	B000E1BLS0	5	This Cheesecake is easy to make and easy to love! I love to bake but don't always have the time to start a cheesecake from scratch when I need one. It's easy to make this unique with adding swirled chocolate sauce or cherries to decorate...I like making the mini ones, people can pick one up and devour!	Love this stuff!
2	B001DIM8K8	5	I'll be honest, if you want a quick breakfast don't purchase this oatmeal. But, if you have time to spare or don't mind waiting for some "good eats", this is the oatmeal for you. Step 1. pour some milk into a sauce pan; Step 2. add a dash of salt [or salt to taste]; Step 3. add oatmeal, making sure it stays covered by milk; Step 4. turn on stove to low heat [not too low but definitely not too high, as you don't want the milk to evaporate/dry out leaving you with tough oats in pan] and cover pan; Step 5. do some little cleaning to built up appetite making sure to ck on progress periodically; Step 6. add more milk if oats are still too tough and let it simmer for a little while longer; Step 7. get your ladle/just pour directing into a bowl add a few raisins/cranberries/your favorite dried fruit and ENJOY!	Slow cooking
3	B004K30HO2	3	These cups work for the most part except for the fact that all most all of them leak out the top of the cap. All seem to wick some leakage through the paper filter at the cap and cup joint and approximately 10-15% overflow and leak around the cup and down the sides providing some grounds with the coffee.	Good Start but Could Be Better.
4	B0051LNWPI	5	Crazy-great condition, was doing great until my sister accidentally knocked it over while I was "sunning" it outside. Also, I think using filtered tap water was a bad idea because they're sensitive to salts. It's slowly recovering.	Pitcher perfect

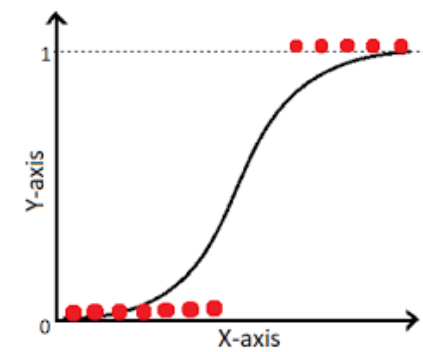
The Data



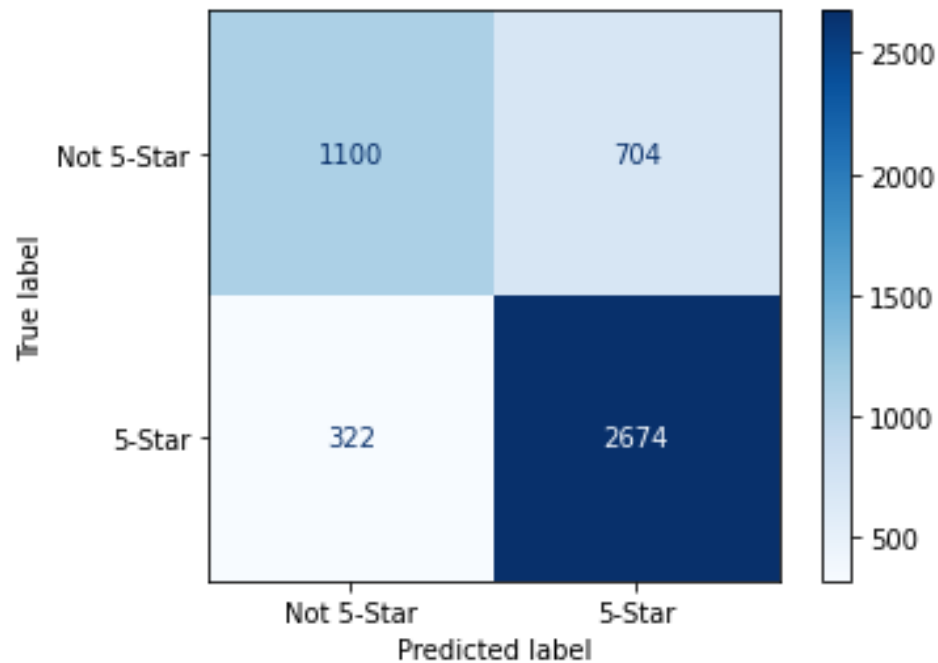
Here is a sample of the original and cleaned text after all data cleansing steps were taken.

original_text	summary	clean_text	clean_summary
This product works exactly as it is advertised. No more rinsing wire baskets. The coffe is as good as the already filled k cups that you buy in the store. It's nice to just throw the whole thing in the trash when done and not clog the sink. Couldn't be happier as it gives me unlimited choices for coffee and it saves money too! Amazon is great!	Disposable k cups	[product, work, exactly, advertised, rinse, wire, basket, coffe, good, already, fill, k, cup, buy, store, nice, throw, whole, thing, trash, do, clog, sink, happier, give, unlimited, choice, coffee, save, money, amazon, great]	[disposable, k, cup]
This Cheesecake is easy to make and easy to love! I love to bake but don't always have the time to start a cheesecake from scratch when I need one. It's easy to make this unique with adding swirled chocolate sauce or cherries to decorate...I like making the mini ones, people can pick one up and devour!	Love this stuff!	[cheesecake, easy, make, easy, love, love, bake, always, time, start, cheesecake, scratch, need, one, easy, make, unique, add, swirl, chocolate, sauce, cherry, decorate, like, make, mini, one, people, pick, one, devour]	[love, stuff]
I'll be honest, if you want a quick breakfast don't purchase this oatmeal. But, if you have time to spare or don't mind waiting for some "good eats", this is the oatmeal for you. Step 1. pour some milk into a sauce pan; Step 2. add a dash of salt [or salt to taste]; Step 3. add oatmeal, making sure it stays covered by milk; Step 4. turn on stove to low heat [not too low but definitely not too high, as you don't want the milk to evaporate/dry out leaving you with tough oats in pan] and cover pan; Step 5. do some little cleaning to built up appetite making sure to ck on progress periodically; Step 6. add more milk if oats are still too tough and let it simmer for a little while longer; Step 7. get your ladle/just pour directing into a bowl add a few raisins/cranberries/your favorite dried fruit and ENJOY!	Slow cooking	[honest, want, quick, breakfast, purchase, oatmeal, time, spare, mind, wait, good, eats, oatmeal, step, 1, pour, milk, sauce, pan, step, 2, add, dash, salt, salt, taste, step, 3, add, oatmeal, make, sure, stay, cover, milk, step, 4, turn, stove, low, heat, low, definitely, high, want, milk, evaporate, dry, leave, tough, oat, pan, cover, pan, step, 5, little, cleaning, built, appetite, make, sure, ck, progress, periodically, step, 6, add, milk, oat, still, tough, let, simmer, little, longer, step, 7, get, ladle, pour, direct, bowl, add, raisin, cranberry, favorite, dry, fruit, enjoy]	[slow, cooking]
These cups work for the most part except for the fact that all most all of them leak out the top of the cap. All seem to wick some leakage through the paper filter at the cap and cup joint and approximately 10-15% overflow and leak around the cup and down the sides providing some grounds with the coffee.	Good Start but Could Be Better.	[cup, work, part, except, fact, leak, top, cap, seem, wick, leakage, paper, filter, cap, cup, joint, approximately, 10, 15, overflow, leak, around, cup, side, provide, ground, coffee]	[good, start, could, well]
Crazy-great condition, was doing great until my sister accidentally knocked it over while I was "sunning" it outside. Also, I think using filtered tap water was a bad idea because they're sensitive to salts. It's slowly recovering.	Pitcher perfect	[crazy, great, condition, great, sister, accidentally, knock, sun, outside, also, think, use, filter, tap, water, bad, idea, sensitive, salt, slowly, recover]	[pitcher, perfect]

Logistic Regression Evaluation

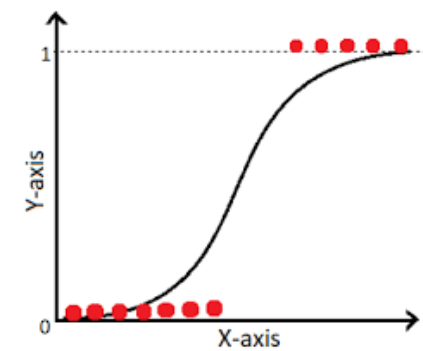


There were 3 different logistic regression models trained, however, this one had the best performance at an overall 79% total accuracy. It is much better at identifying 5-star reviews than reviews that are not 5 stars as indicated by F-1 scores



	precision	recall	f1-score	support
0	0.77	0.61	0.68	1804
1	0.79	0.89	0.84	2996
accuracy			0.79	4800
macro avg	0.78	0.75	0.76	4800
weighted avg	0.78	0.79	0.78	4800

Logistic Regression Evaluation



Now that the model has been trained, the next step is to call predict on the unlabeled data and then evaluate how it did. When the confidence is high, it is making good predictions!

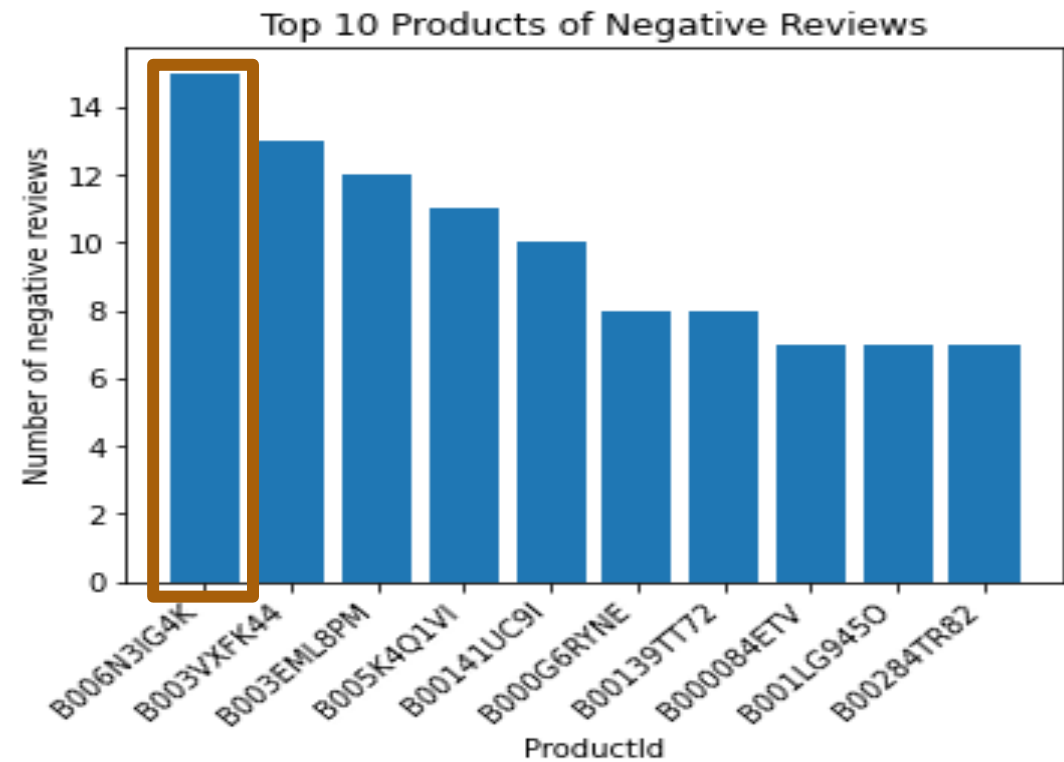
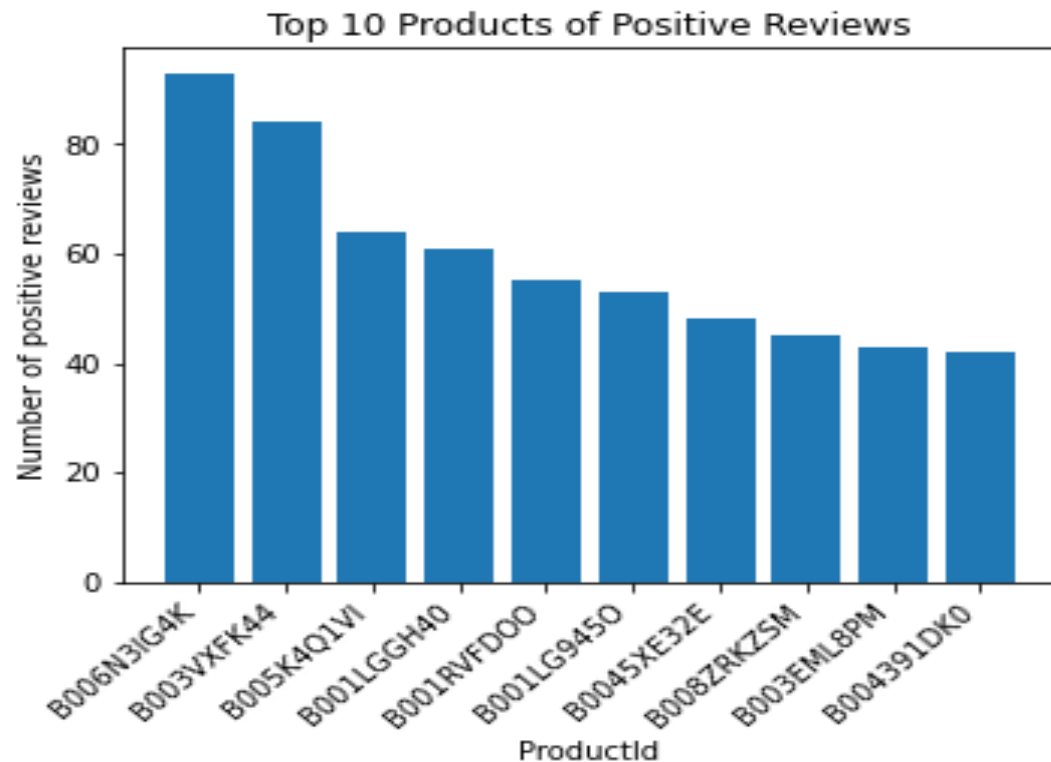
	original_unlabeled_text	predicted_ratings	prob_5_star_rating
	It is very very good sweets Amazon always is excellent Amazon always is excellent Amazon always is excellent	1	0.978402
	Fabulous gift. The blooming teas were lovely and unique. Packaging was beautiful and perfect for a gift.	1	0.976441
	My entire family loves these crackers. Great with cheese. Delicious alone too.	1	0.973840
	Love these was unable to find them in local stores, so I ordered them. Thanks great low carb snack.	1	0.968245
	Highland Sugarworks syrup is the most delicious syrup ever. The entire family loves it. Rich and not too sweet. Perfect!	1	0.960654
	The vendor is fast and dependable. The tea is simply the best way to relax and great tasting too. Highly recommend !!	1	0.959117
	Our Baby loves it. And Amazon is very reliable and fast free super saver shipping is always a good deal. Thanks amazon.	1	0.955927
	You can fry, saute, and make wonderful salad dressing. Very creamy taste and a must for your well stocked kitchen.	1	0.951947
	The Brownie Mix is a great base for cheesecake, creamcheese, cool whip stuff, etc. The brownies are great, the whole family loves them, not just those who need a gluten-free diet. MAKES A GREAT GIFT!!!!	1	0.947155
	I love this coffee, the Nantucket Blend is very delicious. My boyfriend's favorite. The 2 pack is great value. I recommend this!	1	0.942906
	These pouches are so easy and convenient. They taste great and are perfect for busy moms.	1	0.940184
	It was a smooth transaction. Easy ordering, fast delivery, yummy breakfast birthday biscuits!! YaY! I will be ordering another jar very very soon.	1	0.938618
	My dog loves these treats! We started using them for training purposes. The treats of healthy too!	1	0.937690
	They taste so good. Easy on my stomach when I am running. Help keep my energy up on long runs. LOVE THEM!!!	1	0.934858

Sentiment Analysis



Let's see how NLTK's Sentiment Intensity Analyzer does on the unlabeled data. There are some low counts of negative reviews here which could be changed by adjusting the thresholds

Counts for Sentiment Analysis

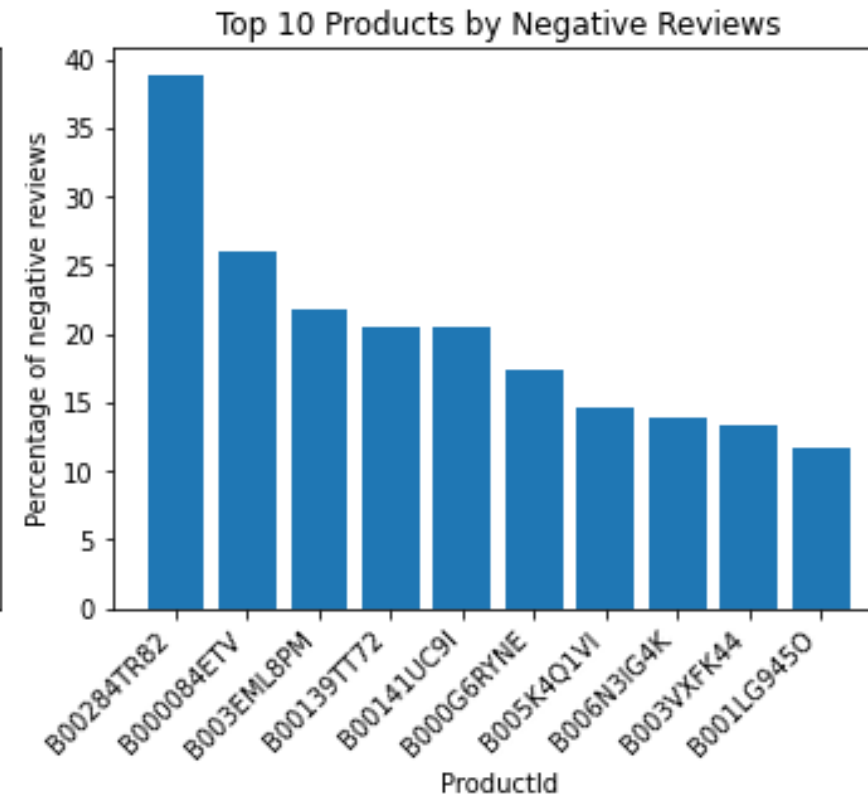
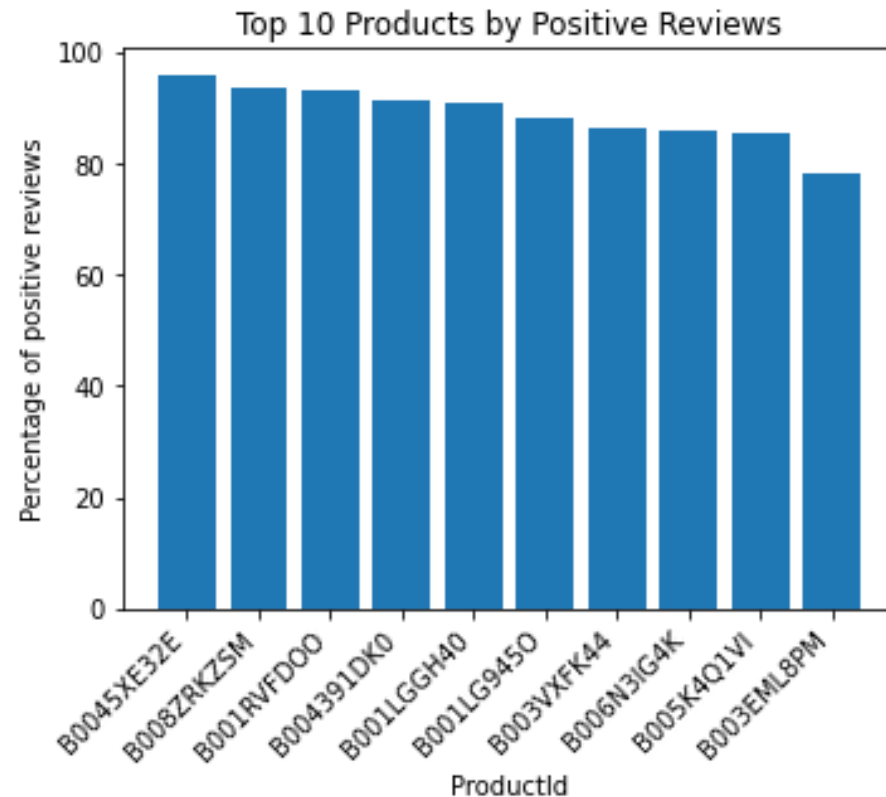


Sentiment Analysis

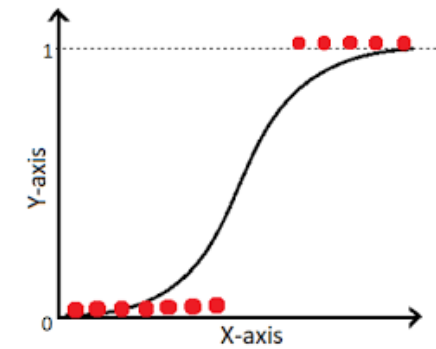


It's probably better to examine the proportions of positive and negative reviews

Proportions for Sentiment Analysis

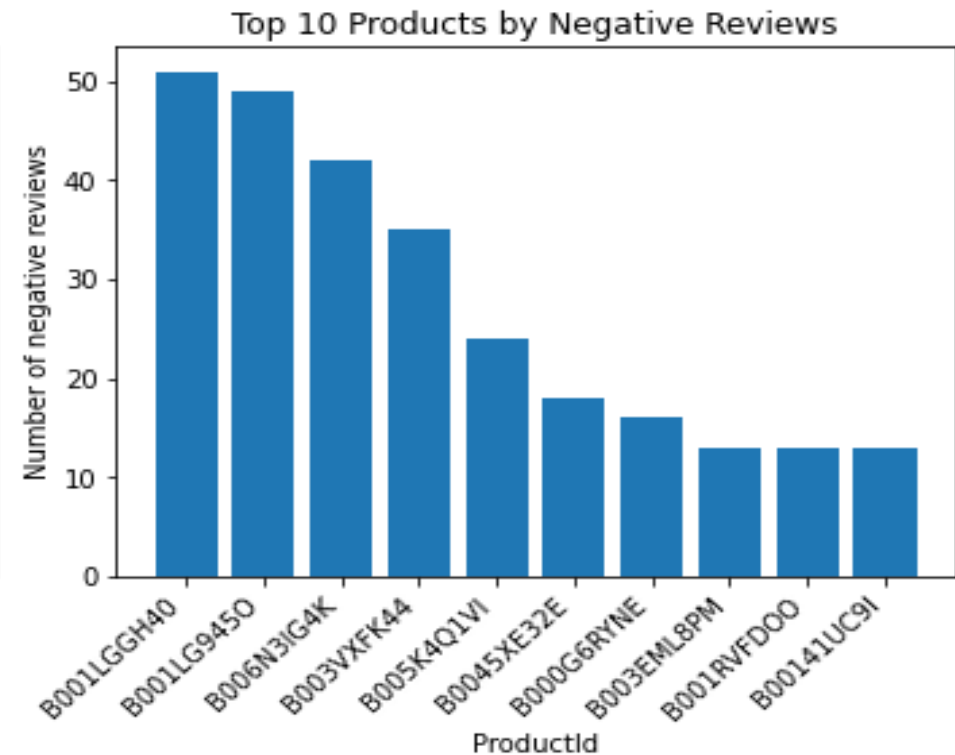
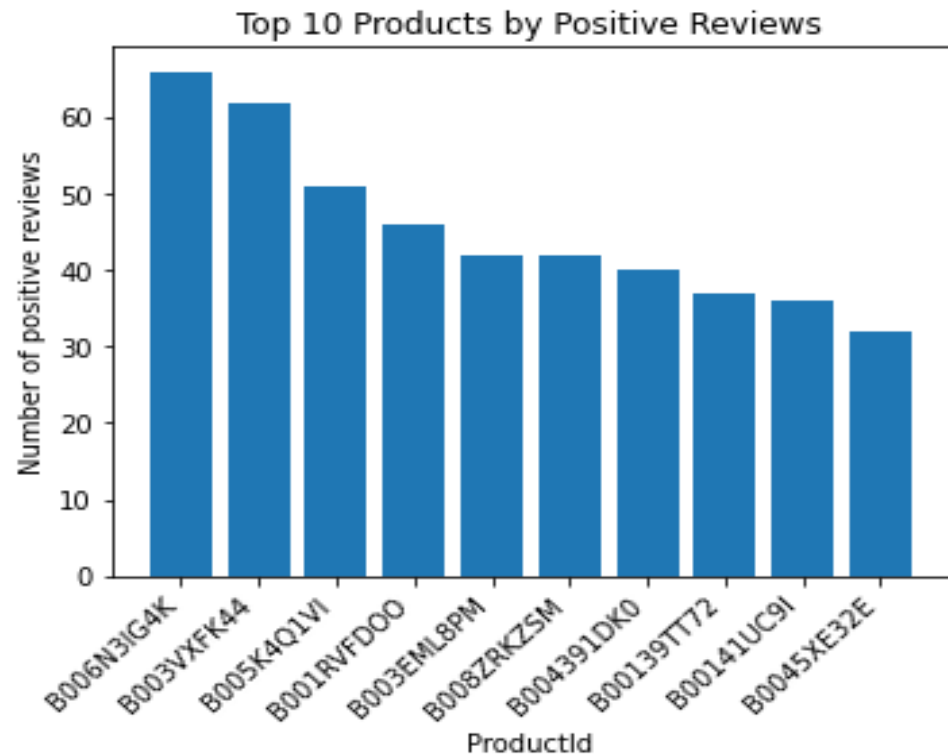


Logistic Regression

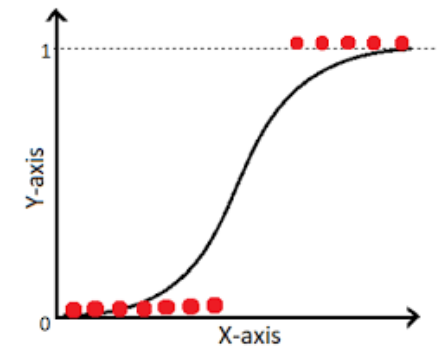


It appears that the Logistic Regression is counting more negative reviews than the Sentiment Analyzer. This could be adjusted for future analysis.

Counts for Logistic Regression Models

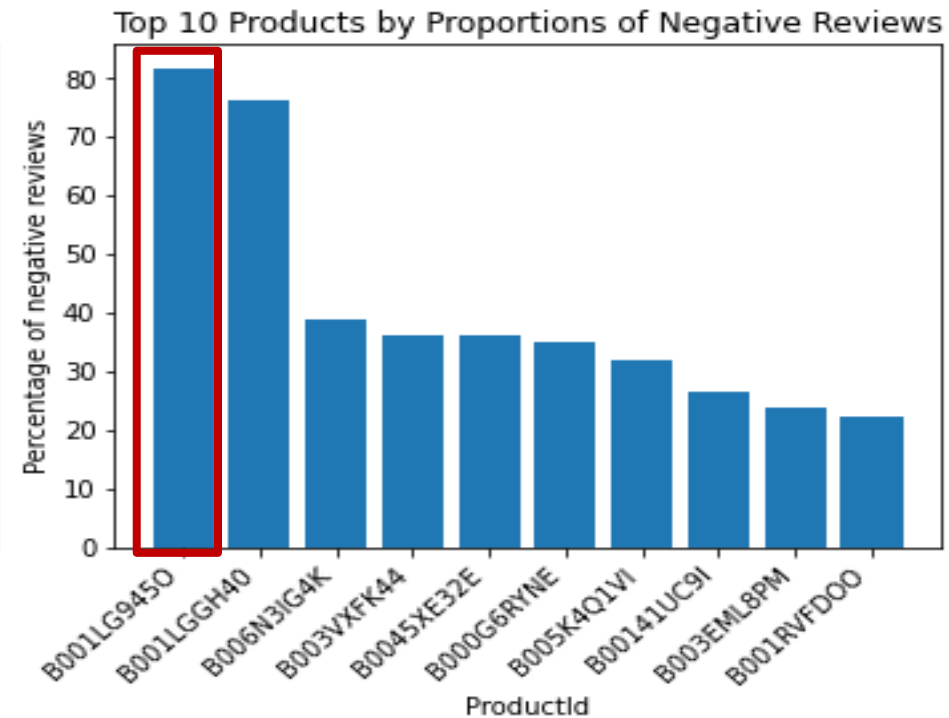
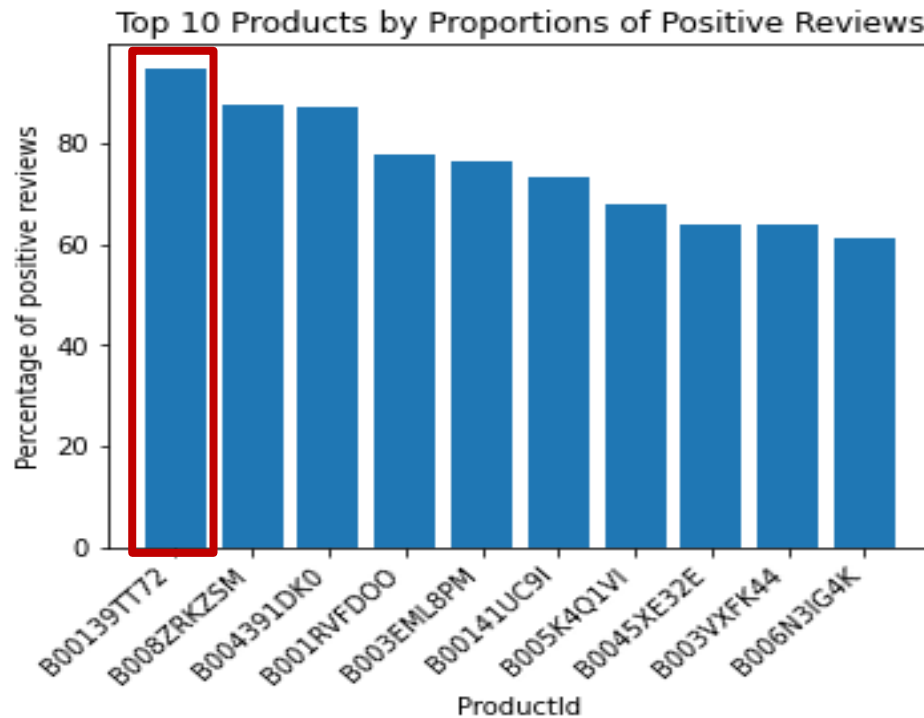


Logistic Regression



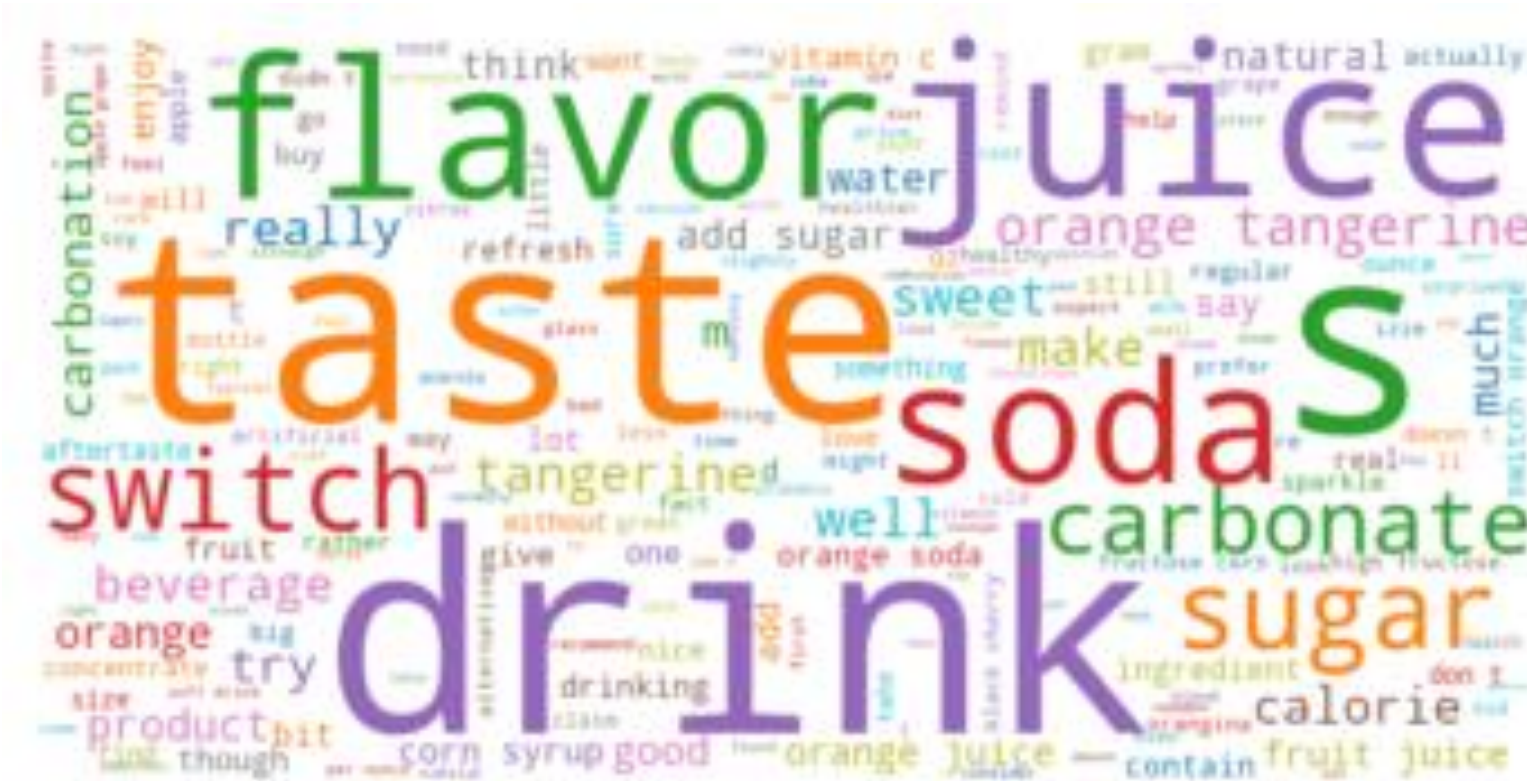
The proportions are showing some obvious winners in terms of most negative and most positive reviews.

Proportions for Logistic Regression Analysis



[illegible]

Conversely, the product with the most negative reviews was a drink that consisted of 82% negative reviews!





Conclusion

- In this case, the final recommendation would be to utilize the logistic regression model as it was doing a better job at predicting positive and negative reviews.
 - This was an arbitrary decision, and more testing could be done to further test this theory
- For the sports drink, the dissatisfaction with the taste could be an opportunity for the marketing team to focus on other aspects of the product. For example, they could highlight the health benefits of the drink, the quality of the ingredients, or the sustainability practices used in the production process.
- For the dog treats, the marketing team could focus on the natural and wholesome qualities of the treats, highlighting the fact that they are made with high-quality, organic ingredients that are beneficial for dogs as this is what the reviewers liked.