



NLP Project: Automated Customer Reviews

Javier Dastas
Jarian Del Valle

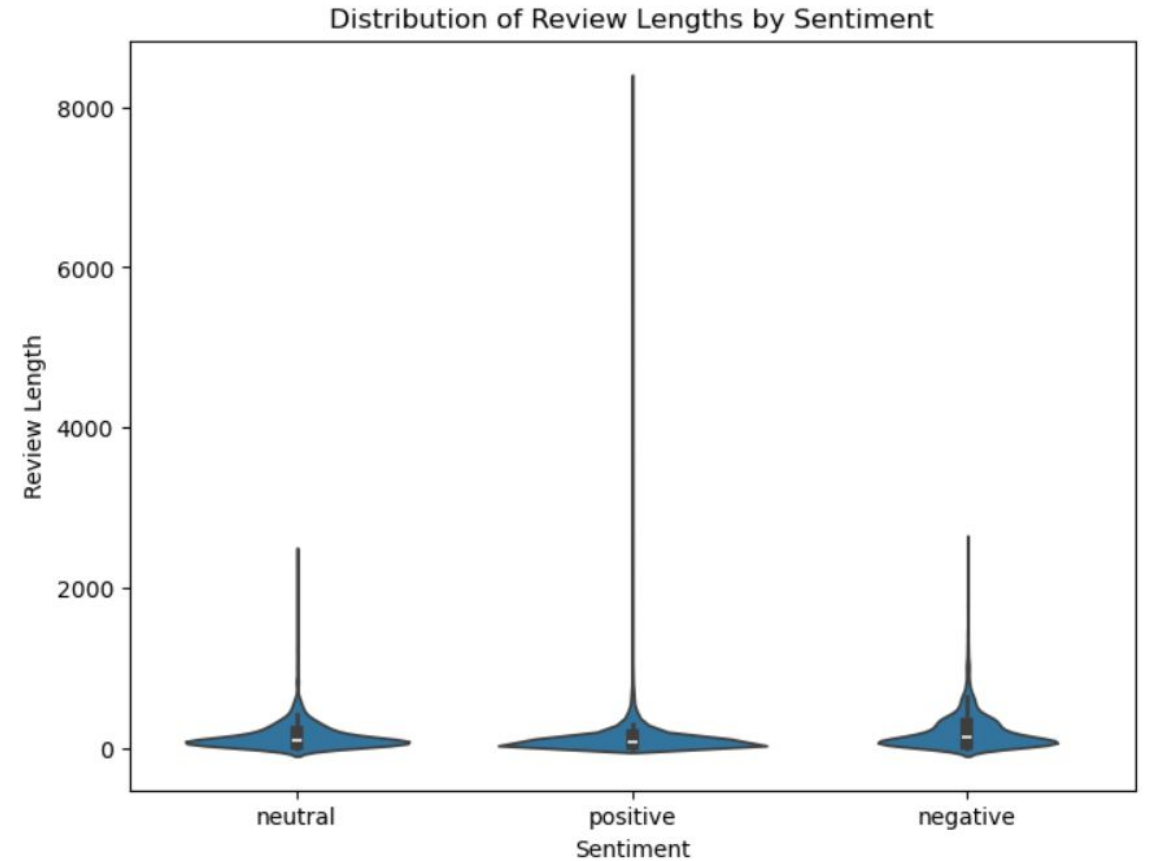
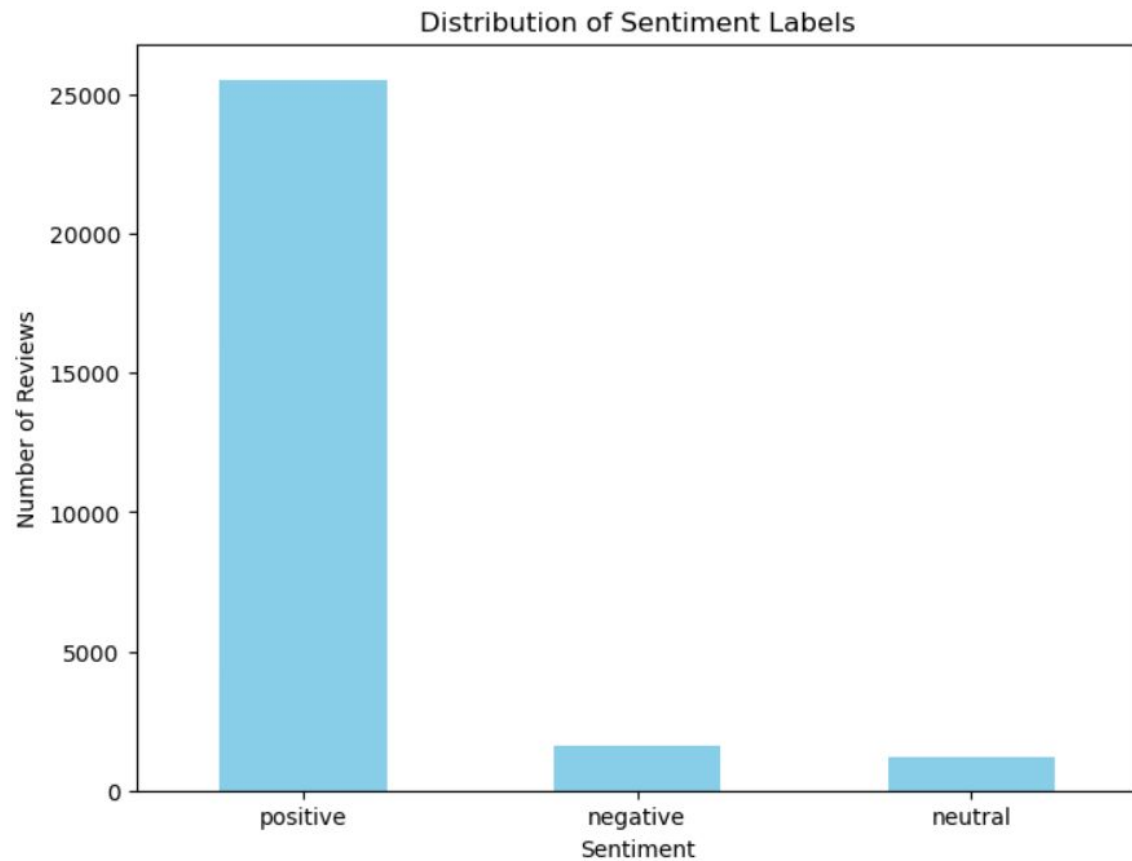
Project Objective

- Automate customer reviews using NLP techniques
- Gain defining data insights
- Test traditional ML models for task
- Transfer learn pre-trained models from Hugging Face

Dataset:

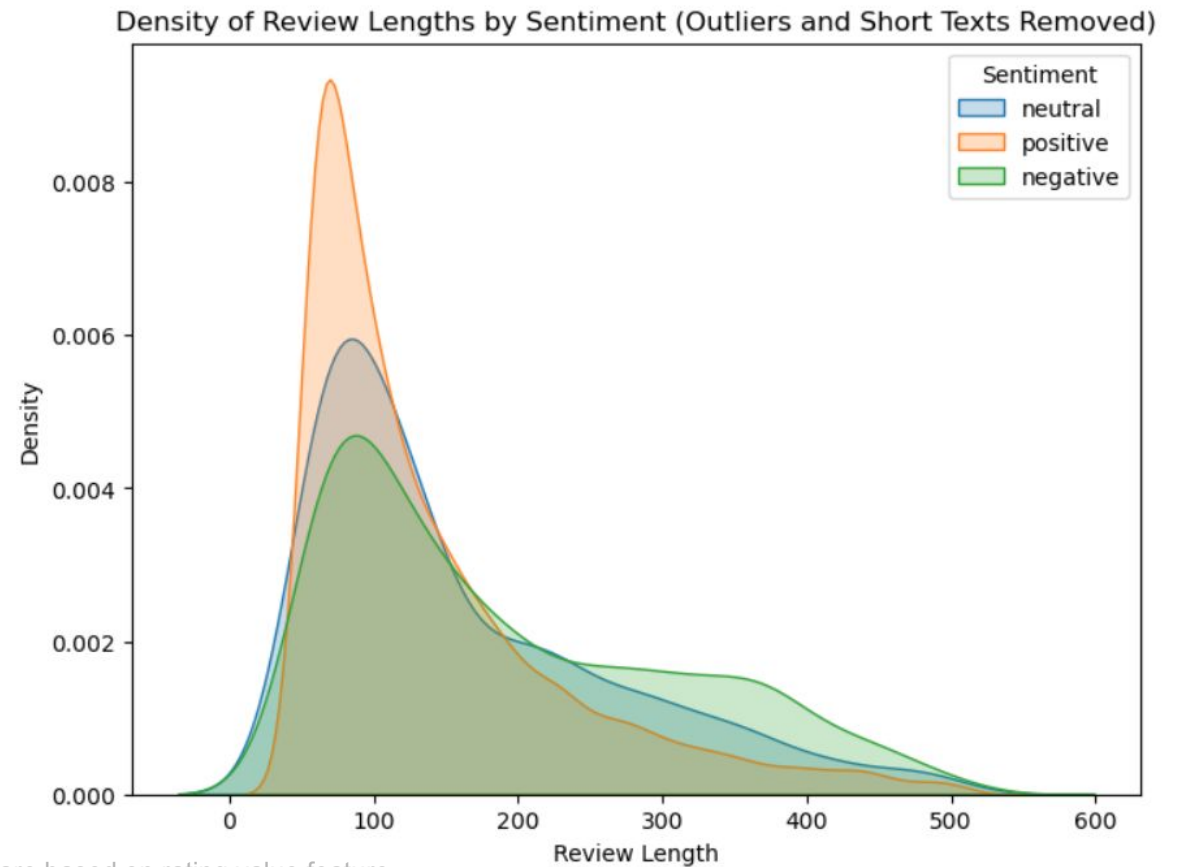
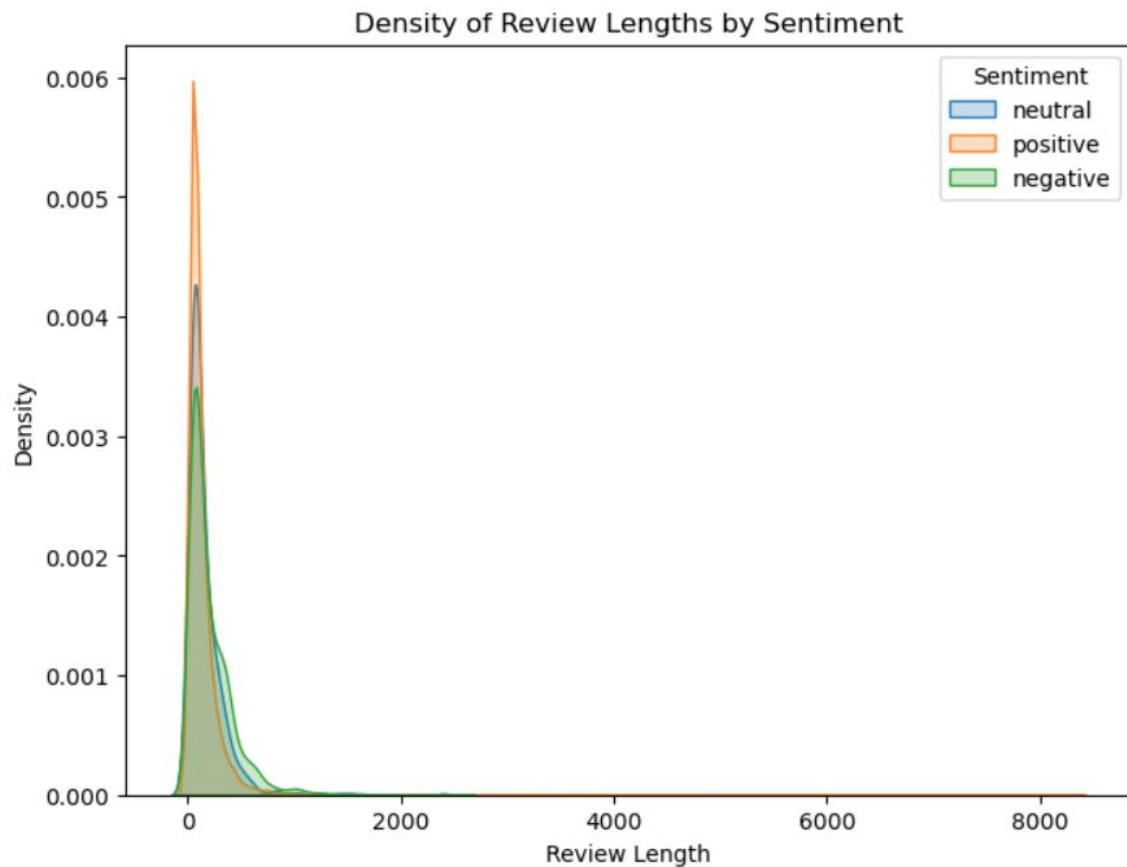
<https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products/data>

Exploratory Data Analysis



The sentiment classifications are based on rating value feature.

Exploratory Data Analysis



The sentiment classifications are based on rating value feature.

General Data Cleaning

Remove Punctuation

Clean from special syllables

Lower case text

reviews.text

I thought it would be as big as small paper bu...

This kindle is light and easy to use especiall...

Didnt know how much id use a kindle so went fo...

I am 100 happy with my purchase I caught it on...

Solid entry level Kindle Great for kids Gifted...

reviews.text

I thought it would be as big as small paper bu...

This kindle is light and easy to use especiall...

Didnt know how much id use a kindle so went fo...

I am 100 happy with my purchase I caught it on...

Solid entry level Kindle Great for kids Gifted...

reviews.text

i thought it would be as big as small paper bu...

this kindle is light and easy to use especiall...

didnt know how much id use a kindle so went fo...

i am 100 happy with my purchase i caught it on...

solid entry level kindle great for kids gifted...

Traditional ML Approach - Data Cleaning

Tokenize text

reviews.text
[i, thought, it, would, be, as, big, as, small...
[this, kindle, is, light, and, easy, to, use, ...
[didnt, know, how, much, id, use, a, kindle, s...
[i, am, 100, happy, with, my, purchase, i, cau...
[solid, entry, level, kindle, great, for, kids...

Remove stop words

reviews.text
[thought, would, big, small, paper, turn, like...
[kindle, light, easy, use, especially, beach]
[didnt, know, much, id, use, kindle, went, low...
[100, happy, purchase, caught, sale, really, g...
[solid, entry, level, kindle, great, kids, gif...

Lemmatize text

reviews.text
[think, would, big, small, paper, turn, like, ...
[kindle, light, easy, use, especially, beach]
[didnt, know, much, id, use, kindle, go, lower...
[100, happy, purchase, catch, sale, really, go...
[solid, entry, level, kindle, great, kid, gift...

[illegible]

Traditional ML Approach

- Pipeline and GridSearchCV- testing multiple traditional models with different hyperparameters
- Tested a MLPClassifier Neural Network

```
# Train and Test Neural Networks
from sklearn.neural_network import MLPClassifier

# Define and train the neural network- best model
nn_model = MLPClassifier(hidden_layer_sizes=(128, 64), max_iter=500)
nn_model.fit(X_train, y_train)

# Predict and evaluate
y_pred = nn_model.predict(X_test)
print(classification_report(y_test, y_pred))
```


Model Evaluation

Model	Best Score	Accuracy on Test Set
Naive-Bayes	0.75	0.76
Logistic Regression	0.77	0.78
Random Forest	0.85	0.87
Neural Network Pre-Trained (MLPClassifier)	0.86	0.86

Model Evaluation

Best parameters for Naive-Bayes: {'classifier__alpha': 0.5}
Best score for Naive-Bayes: 0.7511140525038607
Accuracy on test set for Naive-Bayes: 0.7586024351508734

Best parameters for Logistic Regression: {'classifier__C': 1, 'classifier__solver': 'saga'}
Best score for Logistic Regression: 0.7673064195896757
Accuracy on test set for Logistic Regression: 0.7840127051349921

Best parameters for Random Forest: {'classifier__max_depth': None, 'classifier__min_samples_split': 2, 'classifier__n_estimators': 200}
Best score for Random Forest: 0.8502095742333996
Accuracy on test set for Random Forest: 0.8731251102876302

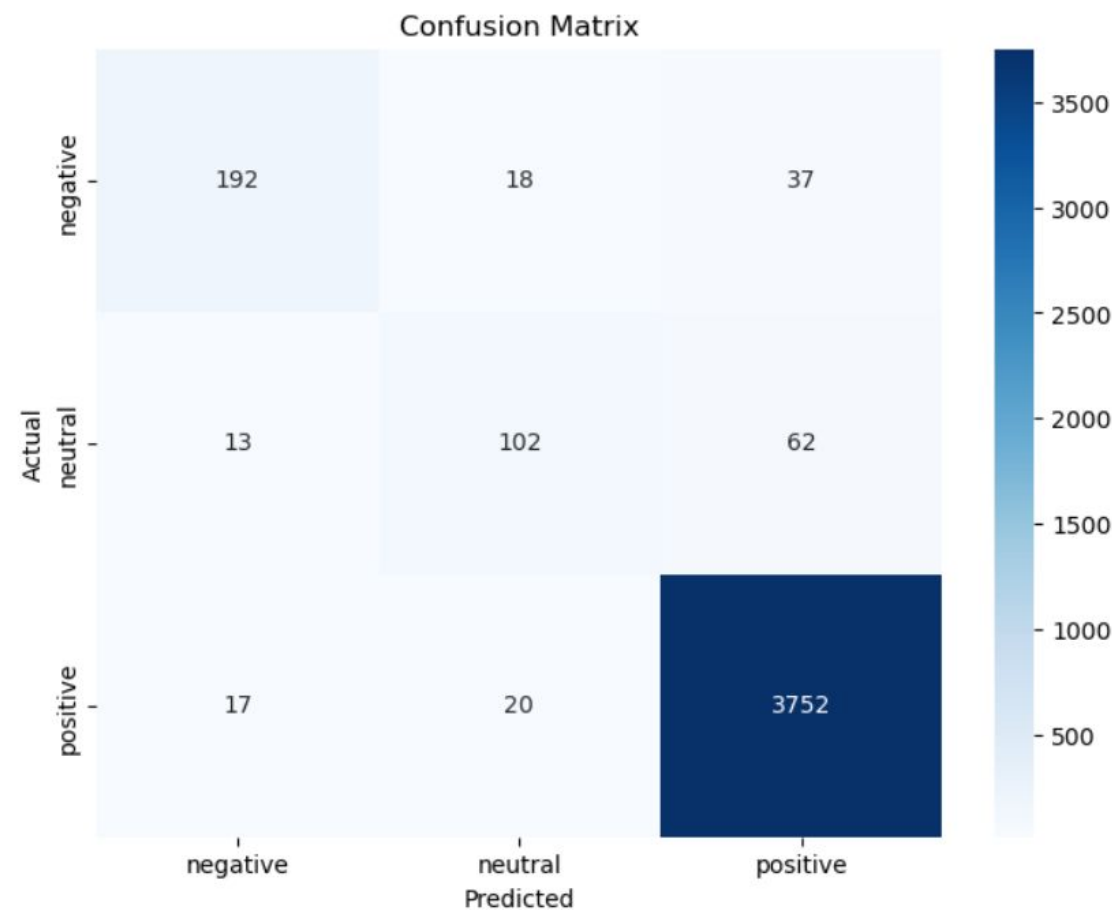
Transformer Approach(Pre-trained model)

- **Token Size is ≤ 512** (After the EDA/ETL Process)
- Selected Model: **DistilBertForSequenceClassification**
- Use the DistilBert specific Tokenizer: **DistilBertTokenizer**
- Pytorch Optimizer Version: **AdamW**(model.parameters(), lr=5e-5)
 - Keep learning-rate weight impact on pre-training model (lr=5e-5)
- Use **Epochs of 7** with **batch_size of 8** for the training.

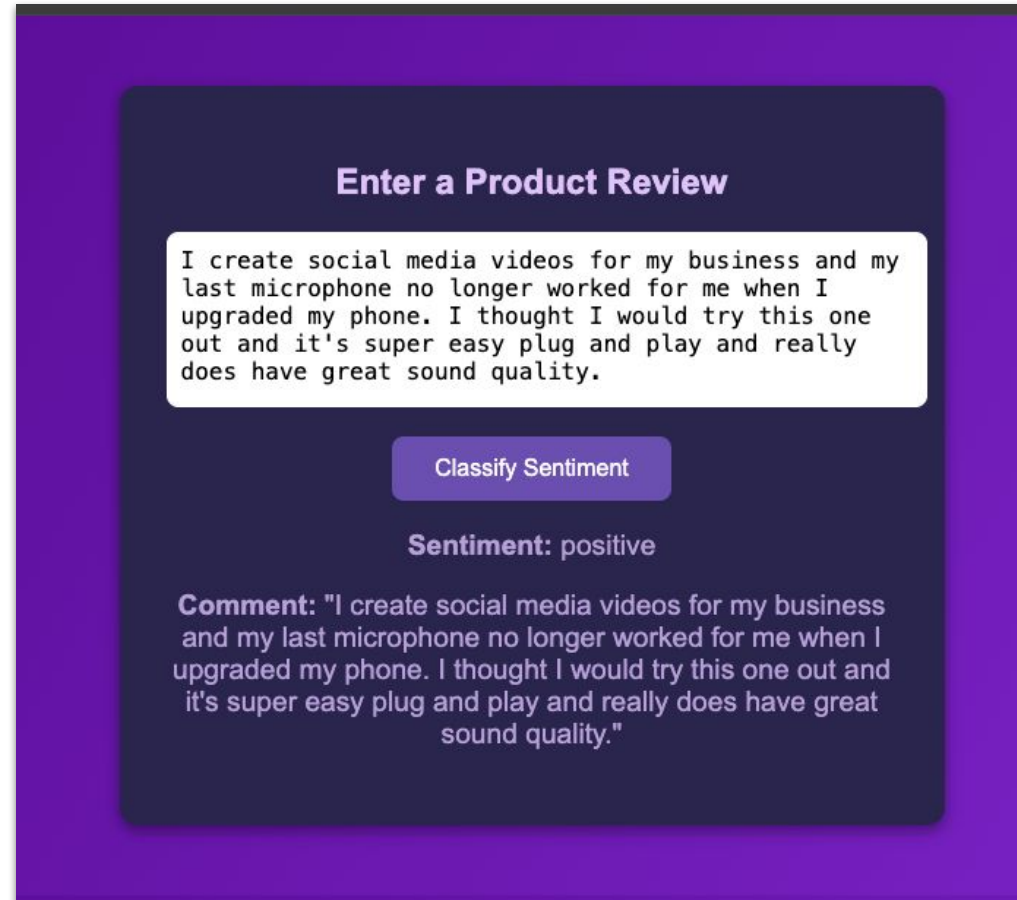
Transformer Approach Evaluation

Classification Report:

	precision	recall	f1-score	support
negative	0.86	0.78	0.82	247
neutral	0.73	0.58	0.64	177
positive	0.97	0.99	0.98	3789
accuracy			0.96	4213
macro avg	0.86	0.78	0.81	4213
weighted avg	0.96	0.96	0.96	4213



Model Deployment

A screenshot of a web application interface for sentiment classification. The interface has a purple background. At the top, there is a dark purple rounded rectangle containing the text "Enter a Product Review" in white. Below this, there is a white rounded rectangle containing a sample product review text. Underneath the text box is a purple button with the text "Classify Sentiment" in white. Below the button, the text "Sentiment: positive" is displayed in white. At the bottom of the dark purple rectangle, the original review text is repeated in a smaller font, labeled "Comment:".

Enter a Product Review

I create social media videos for my business and my last microphone no longer worked for me when I upgraded my phone. I thought I would try this one out and it's super easy plug and play and really does have great sound quality.

Classify Sentiment

Sentiment: positive

Comment: "I create social media videos for my business and my last microphone no longer worked for me when I upgraded my phone. I thought I would try this one out and it's super easy plug and play and really does have great sound quality."

Web Service Model Deployment

<http://dsmlbootcamp.org>

Model Deployment

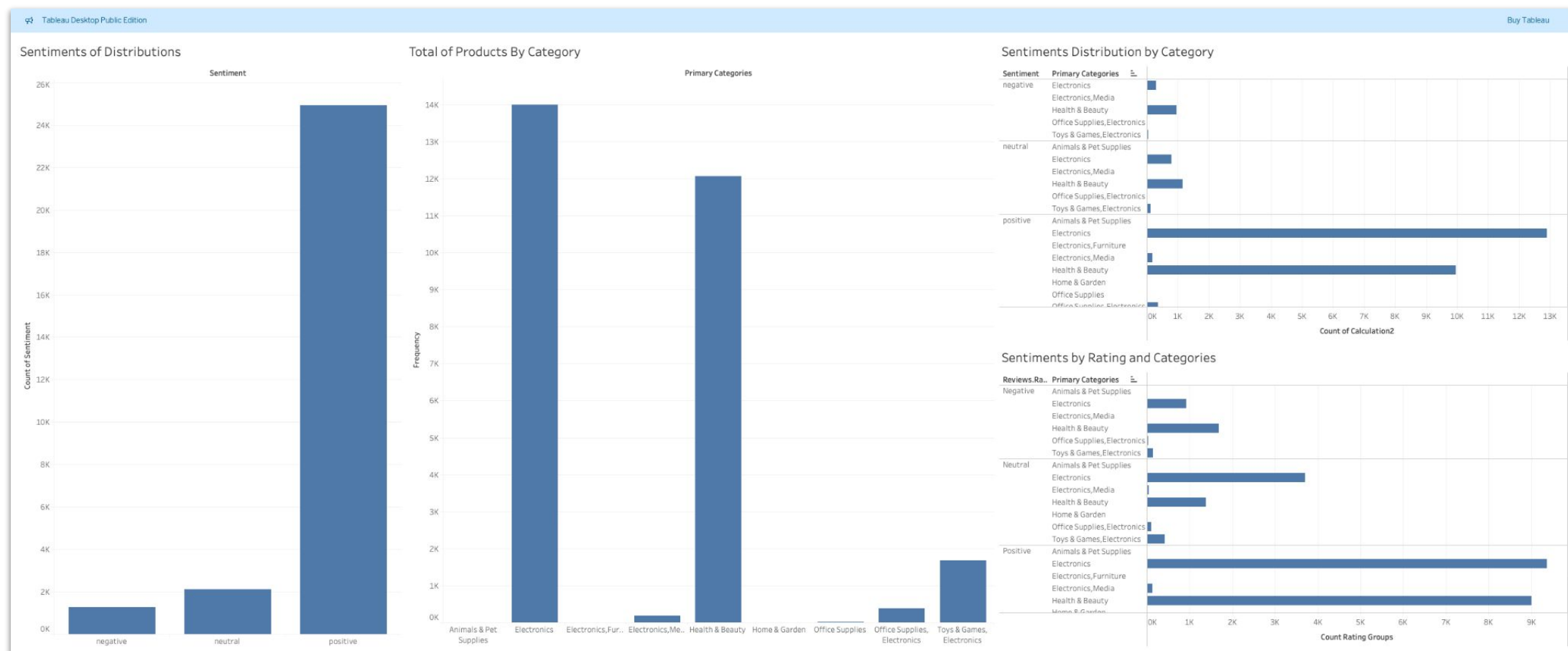


Tableau Dashboard

Model Deployment

	sentiment	category	combined_reviews	summary
1	negative	chargers,Health, Household & Baby Care,Health Personal Care	the batteries did not work well.	terrendous, which is surprising because usually Amazon does a great job with hassle free packaging.
2	negative	chargers,Health, Household & Baby Care,Health Personal Care	broken promises or guarantees.	to have gotten a dud batch I sure got it. We BLEW through all these batteries in less than a month.
3	negative	Speakers,MP3 Accessories,All Bluetooth & Wireless Speakers	y got frustrated and returned it.	ems to defeat the purpose. The dots are amazing though. hard to use, and does not support Guam.
4	negative	,Streaming Media Players,Electronics,Fire TV,Digital Device 4	d I have had no issues so far :).	boot' on screen. Cycled power several times and even tried to reset using remote buttons, no dice.
5	negative	ses & Bags,Laptop Accessories,Cases,computers,Backpacks	nd laptop,a very tight fit though	pointed with quality does the job it was only for keeping dust off second laptop,a very tight fit though.
6	negative	ccessory,Abis Electronics,Tablet Accessories,Power Adapters	ver their issue is). I'll bet I'll win!	ompany Statement about these defective, inferior, slave-labor built, imported, crap charging cables?
7	negative	s,Electronics, Tech Toys, Movies, Music,Computers & Tablets	it won't even turn on or charge.	ay to easy. The charger that comes with the kindle simply does not fit well and give it a good charge.
8	negative	ters & Tablets,Amazon Ereaders,Kindle E-readers,E-Readers	ses it will resolved in an update	etic connector is poorly designed and grows weak. Two pieces to keep up with; unsnaps constantly.
9	negative	ices,Back To College,Amazon Devices,Amazon,Holiday Shop	ess and get the ethernet addon	bill was 172.00 I get more choices options, and I got the components to run it off ethernet, so no lag.
10	negative	amazon Tablets,Electronics,Computers & Tablets,iPad & Tablets	r this is the actual performance	tes. Screen is attractive but it even with a good signal in public places, hanging on to Wifi is difficult.
11	negative	nents,Electronics,Hard Drives & Storage,Computers & Tablets	he schools IT department help.	asted 4 or 5 years with more use. it had stereo speakers, lasted a week on a charge not 2 or 3 days.
12	negative	blets,Amazon,Tablets & eBook Readers,Tablets & E-Readers	Amazon Fire TV or Smart TV's.	e of the amazon Fire. Got extended warranty. So basically getting the next one at 50% off next year.
13	negative	& Tablets,Electronics,iPad & Tablets,Tablets & eBook Readers	nded up returning that one too.	adds when I do ... very very stupid . It has cracked . And I have never dropped it . It isn't made well.
14	negative	uters & Tablets,Electronics,Amazon,Tablets & eBook Readers	comes with very little warranty.	e 7" Fire had just as good of picture quality and for half the amount, ended up being the better deal.
15	negative	onics,Amazon,Tablets & eBook Readers,Tablets & E-Readers	not something you can change.	hange the tablet to run like as you wish. It,Ãs pre installed menu is not something you can change.
16	negative	Tablets,Computers & Tablets,Other Computers & Networking	he schools IT department help.	buying apps once you authorized as allow its done. Not as many free good apps Like apple have it.
17	negative	nazon Tablets,Computers & Tablets,Tablets & eBook Readers	r this is the actual performance	ne protector was not made for the Fire 7 as it was to large and had no directions that were coherent.
18	negative	n Tablets,Frys,Computers & Tablets,Tablets & eBook Readers	r this is the actual performance	ne protector was not made for the Fire 7 as it was to large and had no directions that were coherent.
19	negative	& Tablets,Office Electronics,Amazon,Tablets & eBook Readers	d ones and sending these back	dit card stored in the account, and to turn on parental controls it blocks other features too like Alexa.
20	negative	ablets,Kids' Tablets,Electronics,iPad & Tablets,Tablets for Kids	it won't even turn on or charge.	er tablets. The charger that comes with the kindle simply does not fit well and give it a good charge.

Generate summaries for each product category using a GenAI
'facebook/bart-large-cnn'

Insights

- EDA and preprocessing unlocks meaning
- Transformer tokenization impacts performance
- More data, better results
- Model choice isn't a one size fits all



THANK YOU

Javier Dastas
Jarian Del Valle