



Predicting HelloFresh Review Sentiment

Natural Language Processing & Machine Learning

Ruby Jang



HelloFresh

The most popular meal kit in the world

~1 billion

Meals delivered in 2021

7.5 million

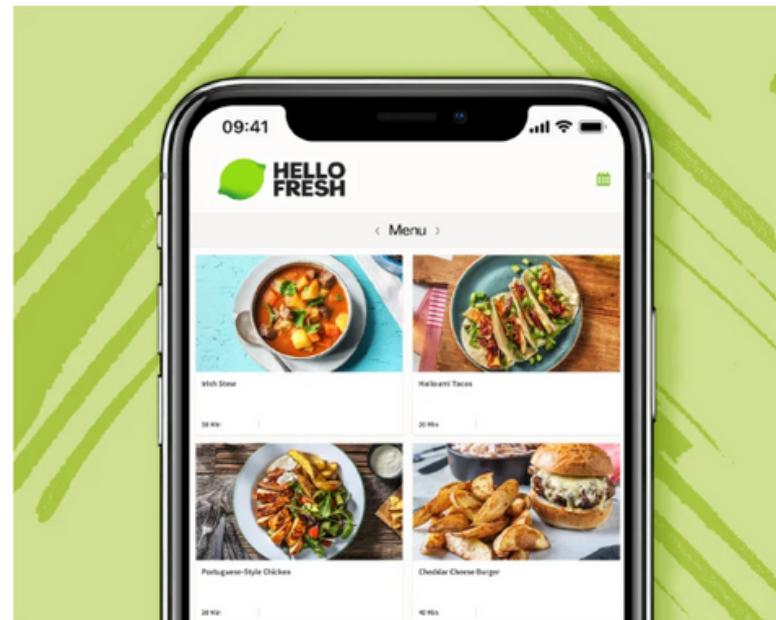
Active customers in Q3 2022

6.0 billion €

Net revenue in 2021

20,000+

Employees



Choose your meals



Create the perfect box



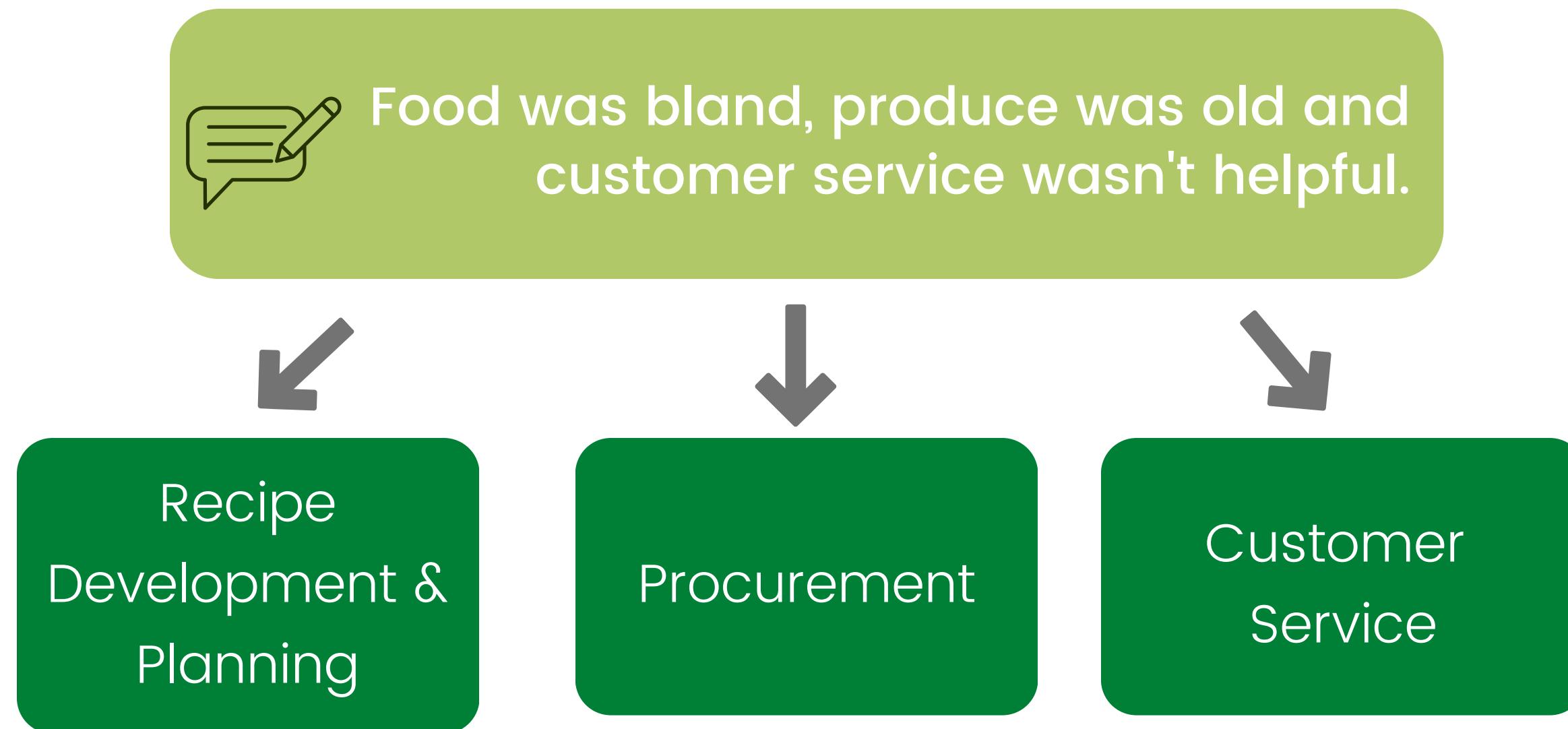
Get convenient weekly
deliveries



Cook seasonal, fresh
ingredients

Customer Reviews

- HelloFresh aims to continuously improve its product and operations by evaluating customer reviews and feedback.



Project goal - predict the sentiment of a review as positive or negative.

Data

- Reviews scraped from Trustpilot website
- 5820 records – title, body, date, rating
- 344 reviews had no review body

★|★|★|★|★ ✓ Invited

A day ago

Easy and convenient

I love the convenience of being able to order ahead and not having to think about what to make for dinner every night. Most meals are great portion sized but sometimes potatoes like for mash, and veg is small supplied and I have to top up with extra veg at separate cost.



	title	body	date	rating
0	Easy and convenient	I love the convenience of being able to order ...	2023-01-02 03:38:39+00:00	4
1	You can't skip 6 or more weeks	You can't skip 6 or more weeks. It has to be e...	2023-01-02 09:12:14+00:00	2
2	During the start of Covid I started to...hunker ...	During the start of Covid I started to order H...	2023-01-02 05:09:48+00:00	5
3	Soggy package when delivered	Package received was wet at the bottom of the ...	2022-12-31 09:45:46+00:00	1
4	Continues to dissapoint	Hello fresh has gotten increasingly disappoint...	2022-12-31 02:12:55+00:00	1

Pre-Processing

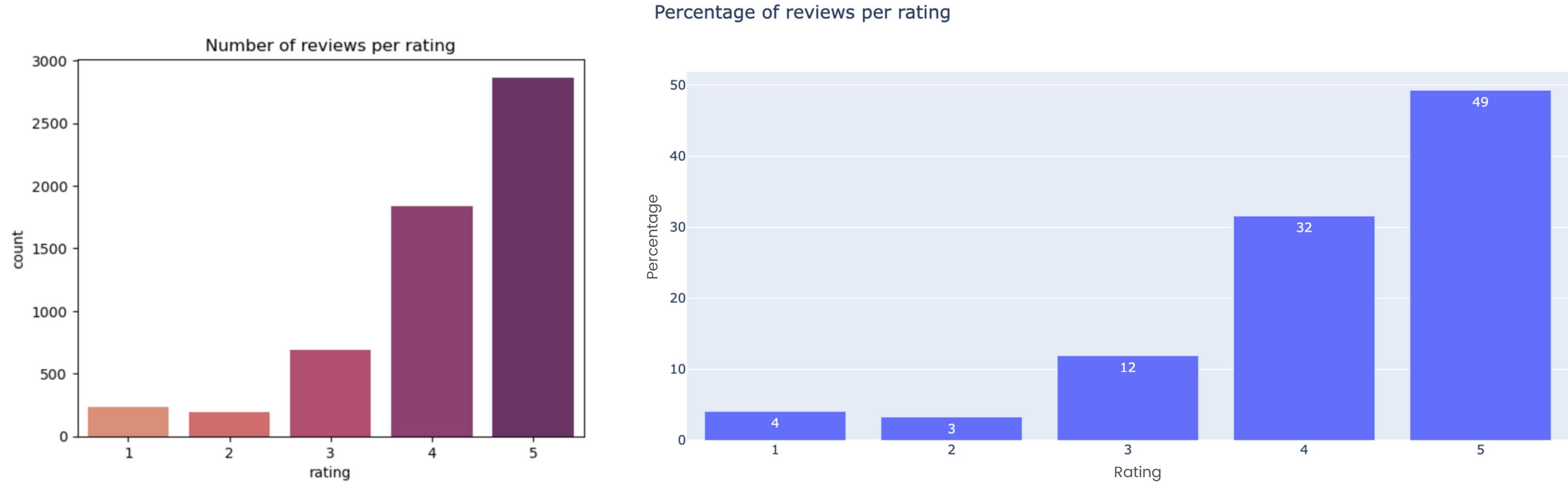
TEXT PRE-PROCESSING

- Lower case
- Remove numbers
- Filter stop words
- Filter punctuations
- Lemmatization

FEATURE ENGINEERING

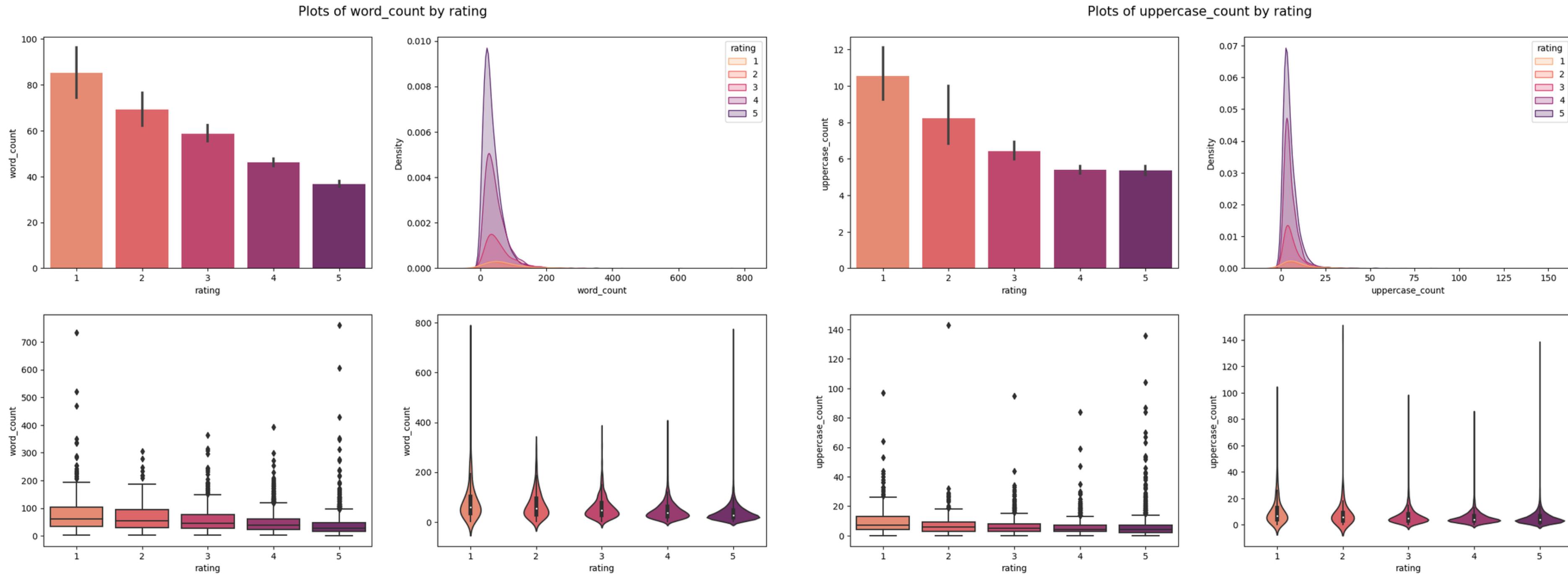
Feature	Description
Review	Combination of review title and body
Word count	Total number of words
Character count	Total number of characters
Word density	Average number of character per word
Uppercase count	Total number of uppercase characters
Polarity	Float between -1 (negative) to 1 (positive)

Exploratory Data Analysis (EDA)



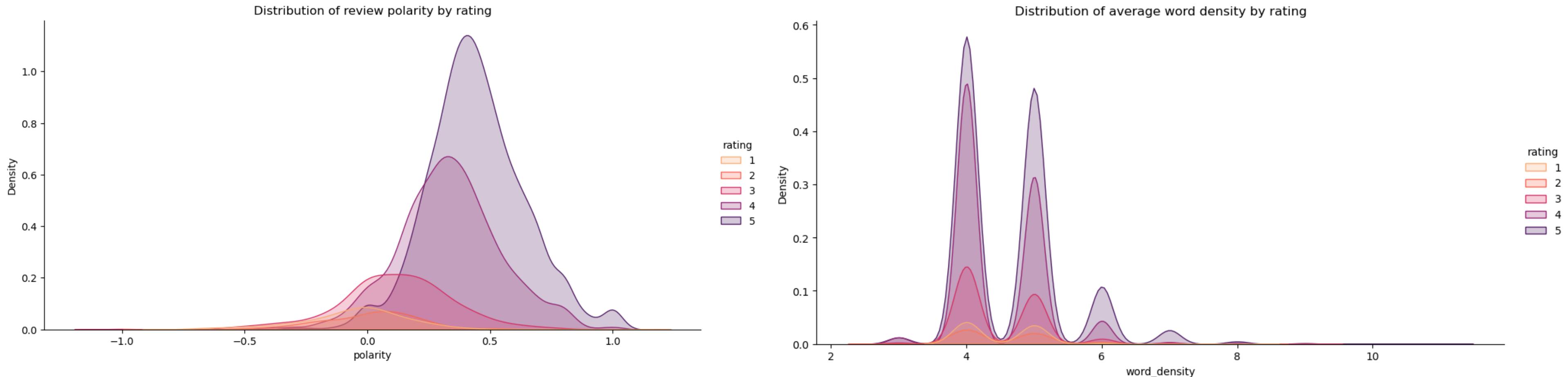
- Majority are rated 4 or 5, with **49%** rated 5.
- Reviews rated 1 or 2 are only **7%** of total reviews.

EDA - Continued



Number of words and uppercase characters tend to increase with lower ratings, with slightly more spread

EDA - Continued

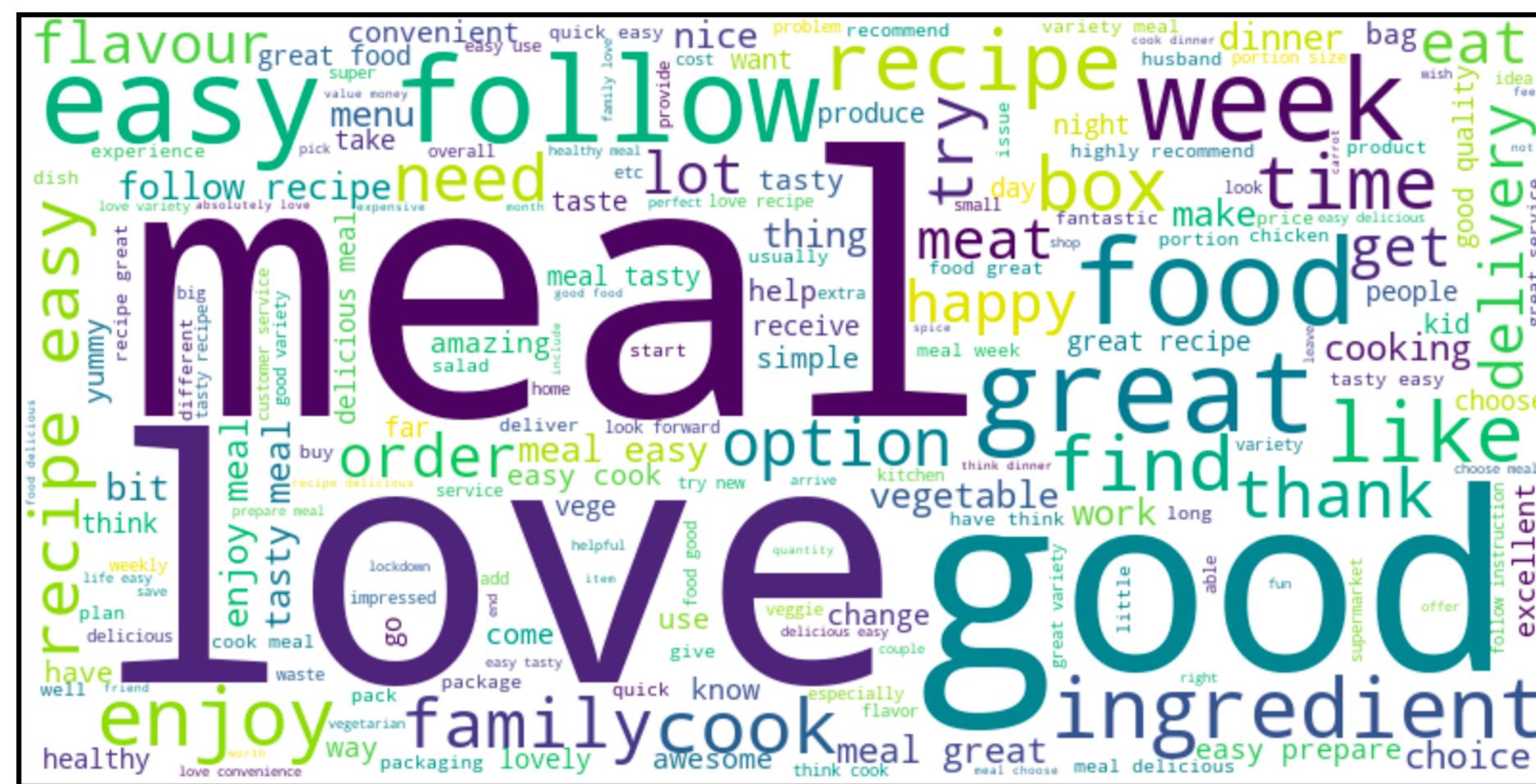


- Mostly positive polarity for high ratings
- Ratings 1 to 3 peak closer to 0 (neutral)

Average word density is similar across all ratings

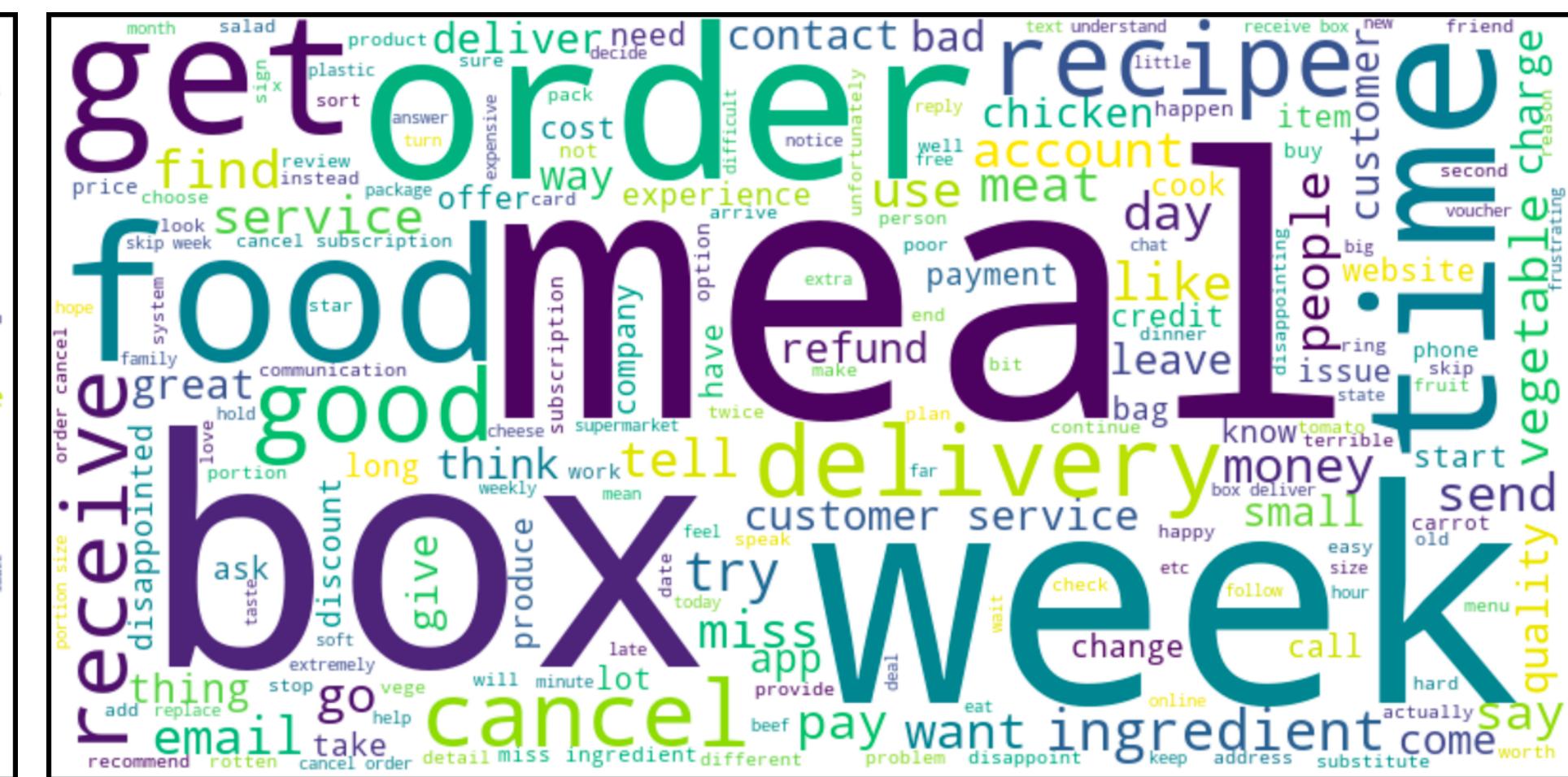
Word Cloud

Positive



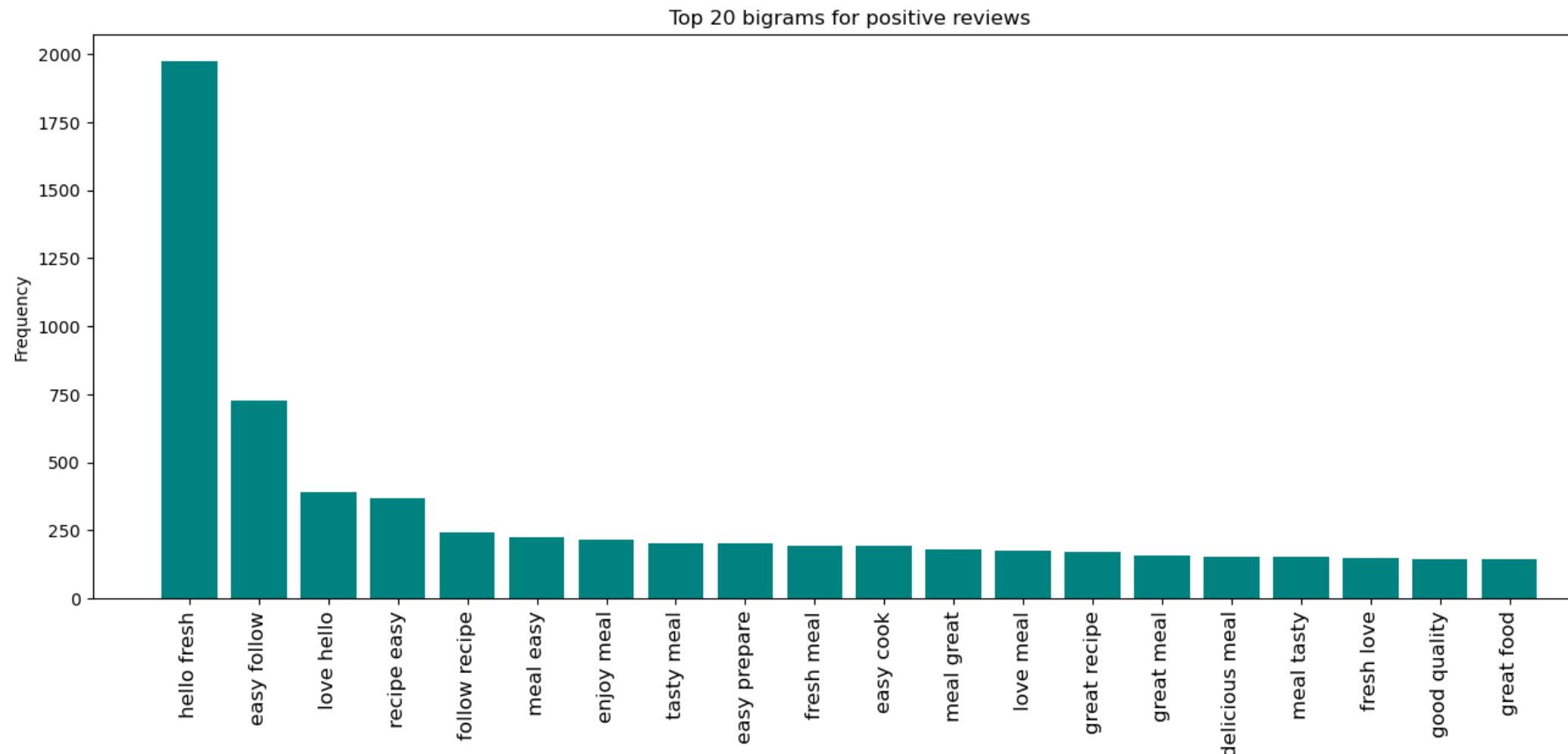
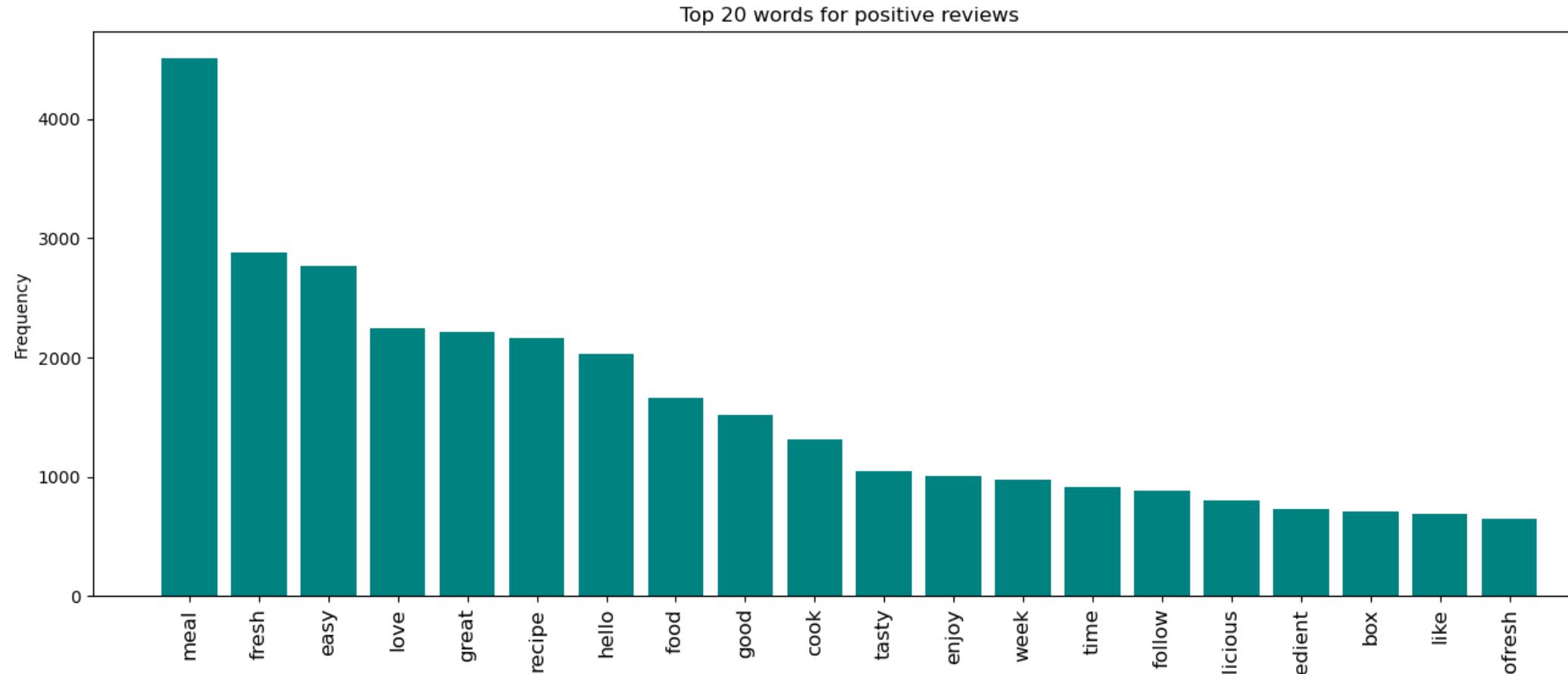
More emotional or evaluative in nature
e.g. love, good, great, easy, enjoy

Negative

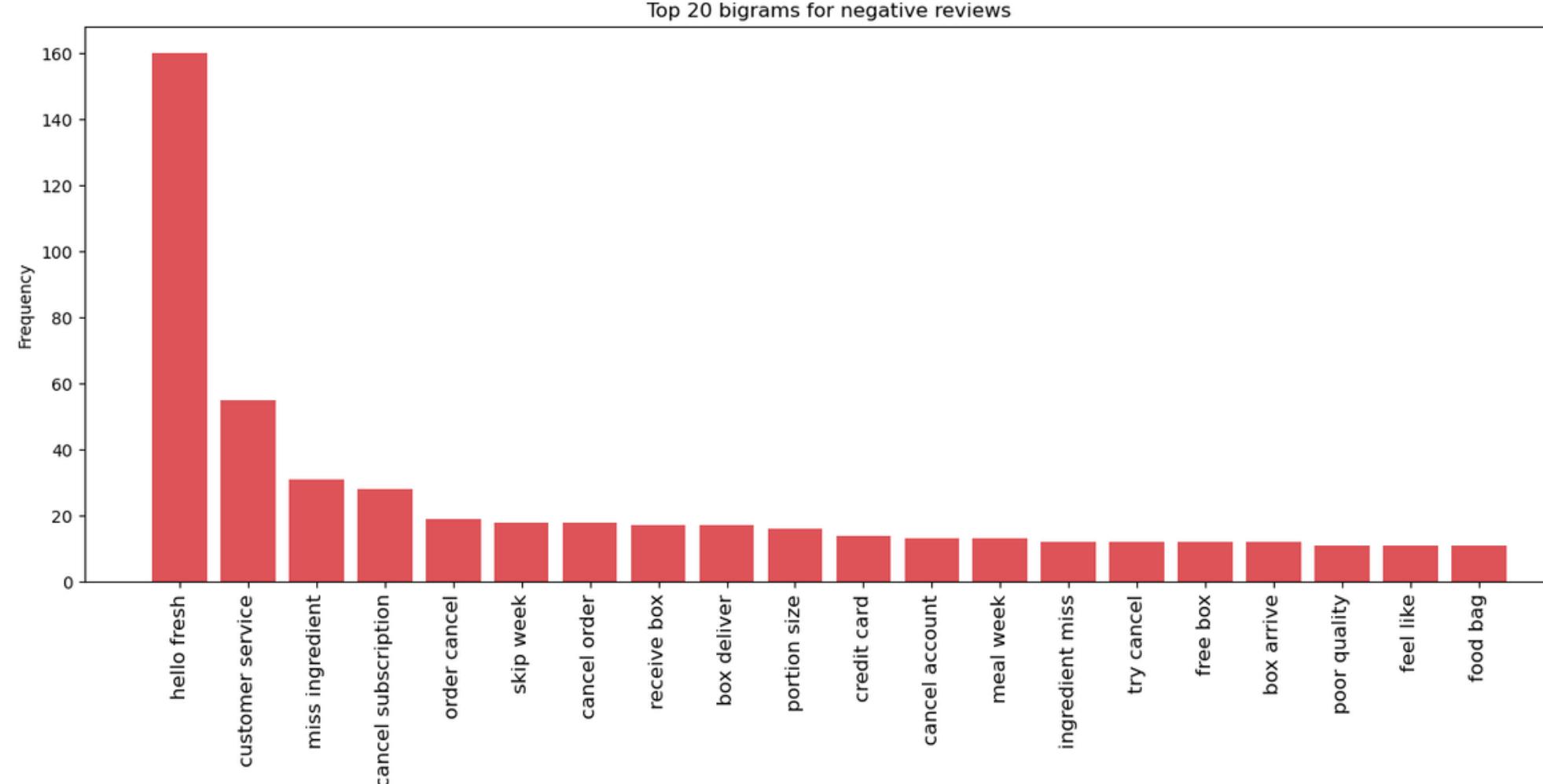
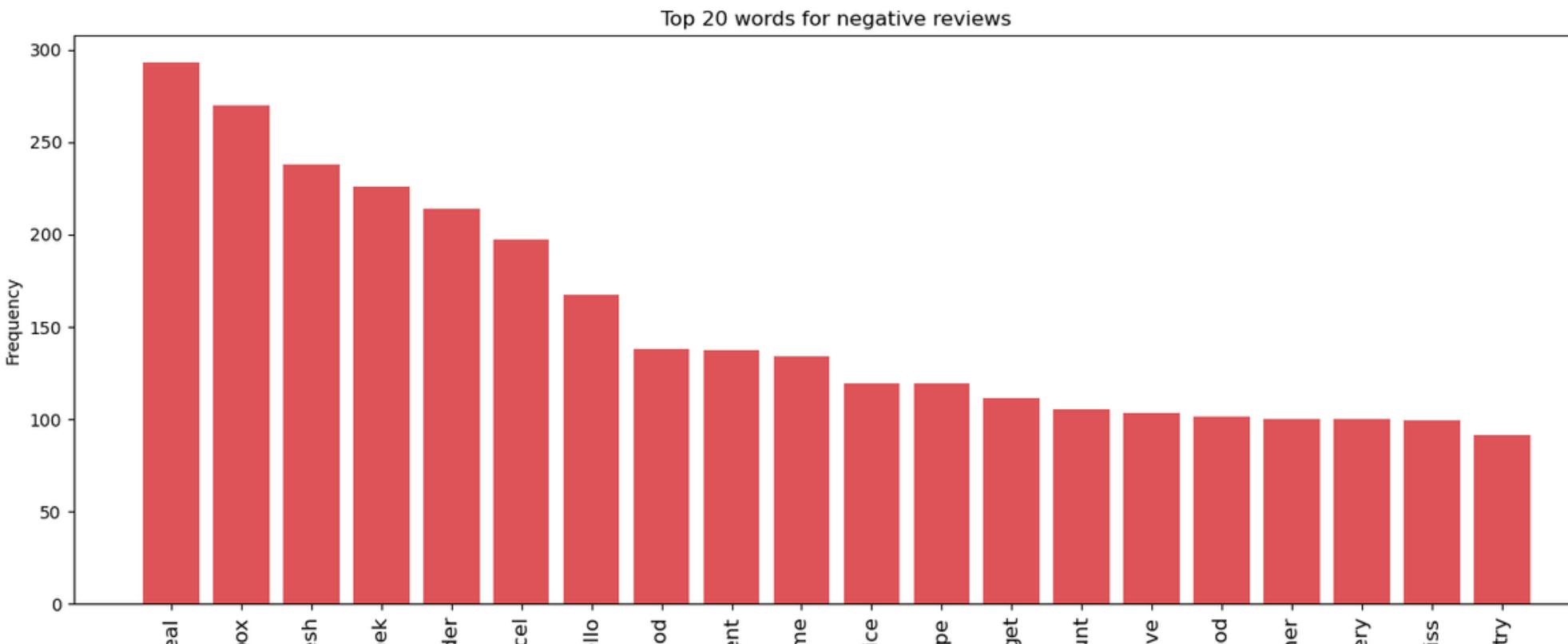


More practical or factual in nature
e.g. order, food, week, time, delivery, recipe

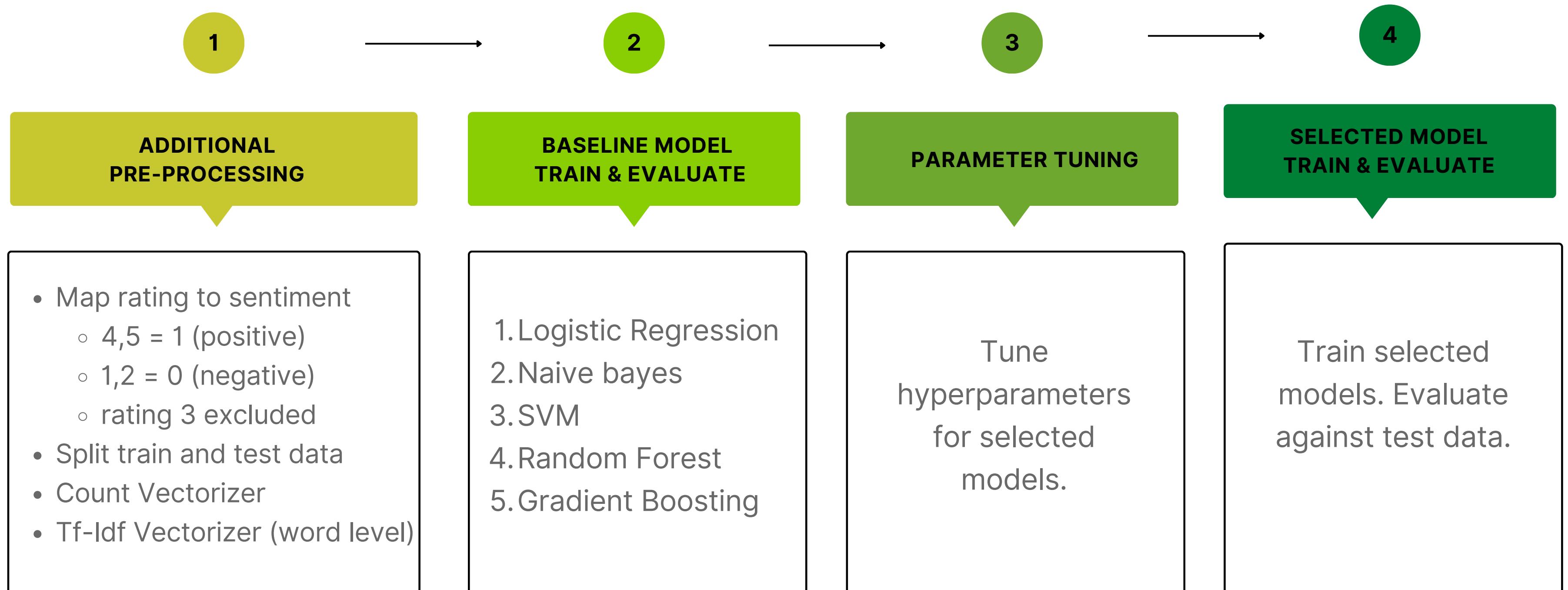
Word Frequency - Positive



Word Frequency - Negative



Machine Learning



Evaluation - Baseline

Count Vectorizer

	Logistic Regression	Naive Bayes	SVM	Random Forest	Gradient Boost
Accuracy (All classes)	96%	94%	95%	95%	95%
F1*	73%	70%	56%	64%	58%
Precision*	77%	63%	88%	83%	84%
Recall*	69%	78%	41%	52%	45%

*These metrics are scores for the minority class (negative reviews)

Based on highest F1 scores, **logistic regression** and **naive bayes** models will be selected for parameter tuning.

Evaluation - Tuned

Logistic Regression

	Baseline		Tuned	
	Count Vectorizer	TF-IDF	Count Vectorizer	TF-IDF
F1*	72%	67%	73%	72%
Precision*	77%	88%	78%	75%
Recall*	69%	54%	69%	68%

Naive Bayes

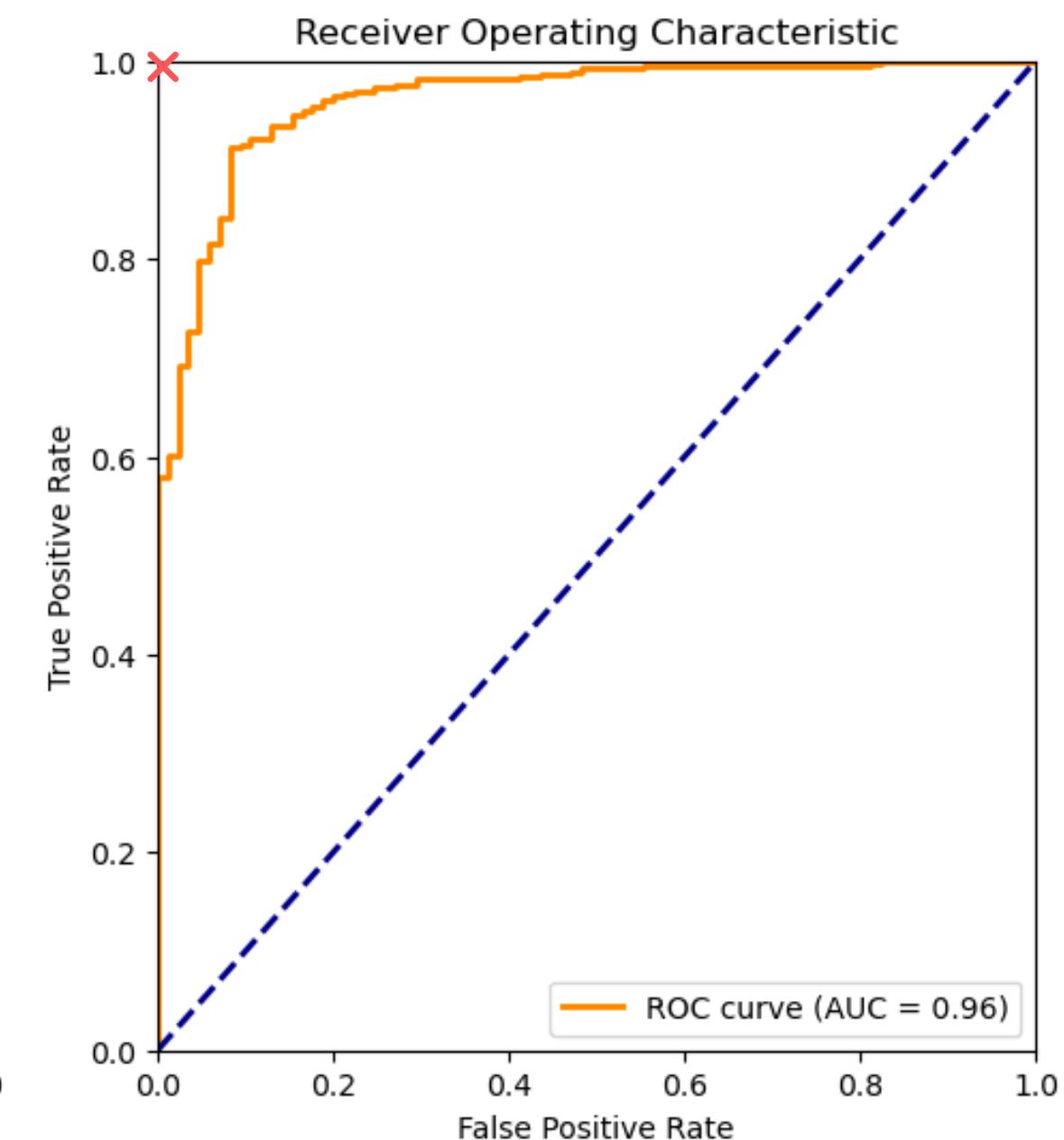
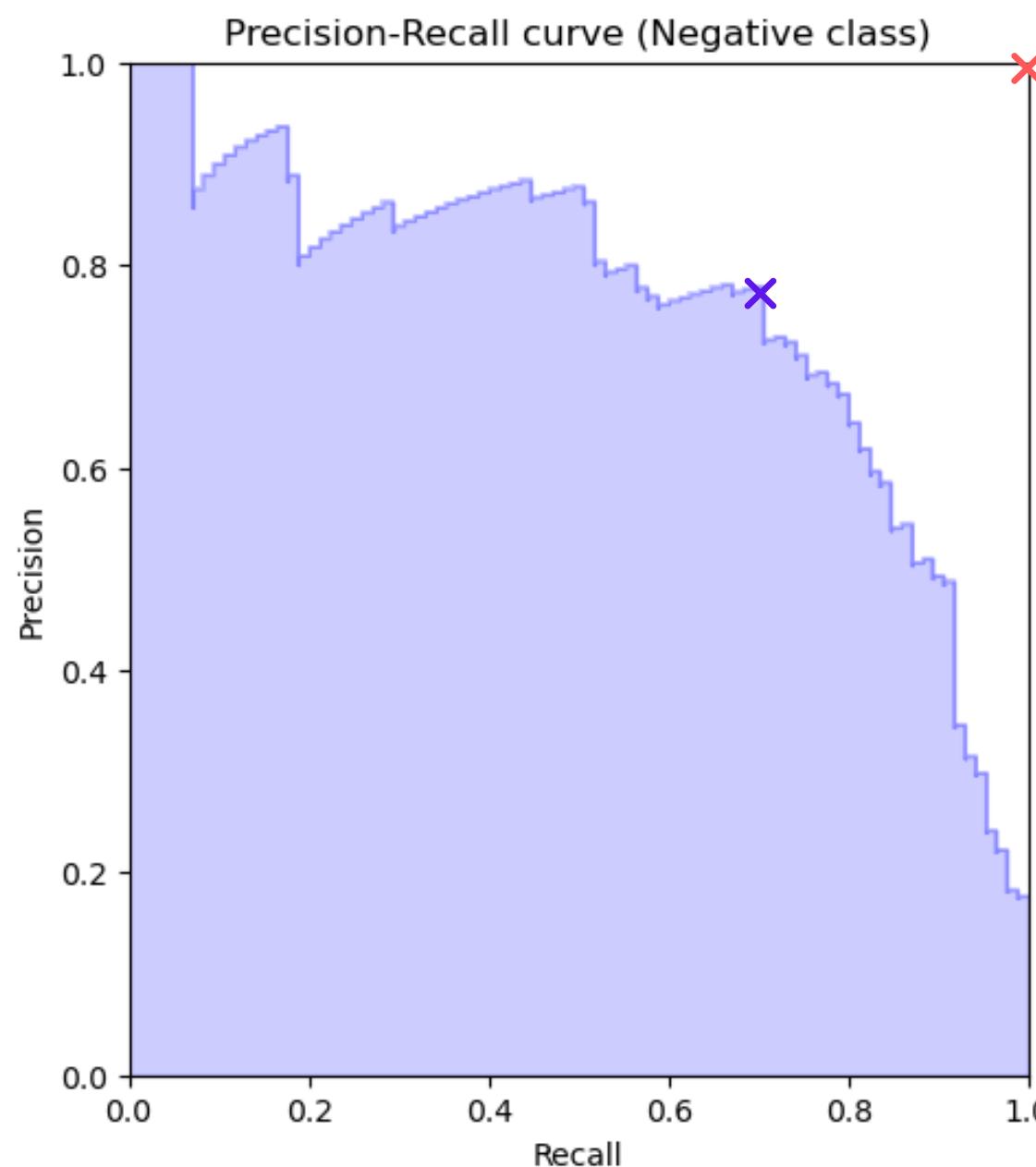
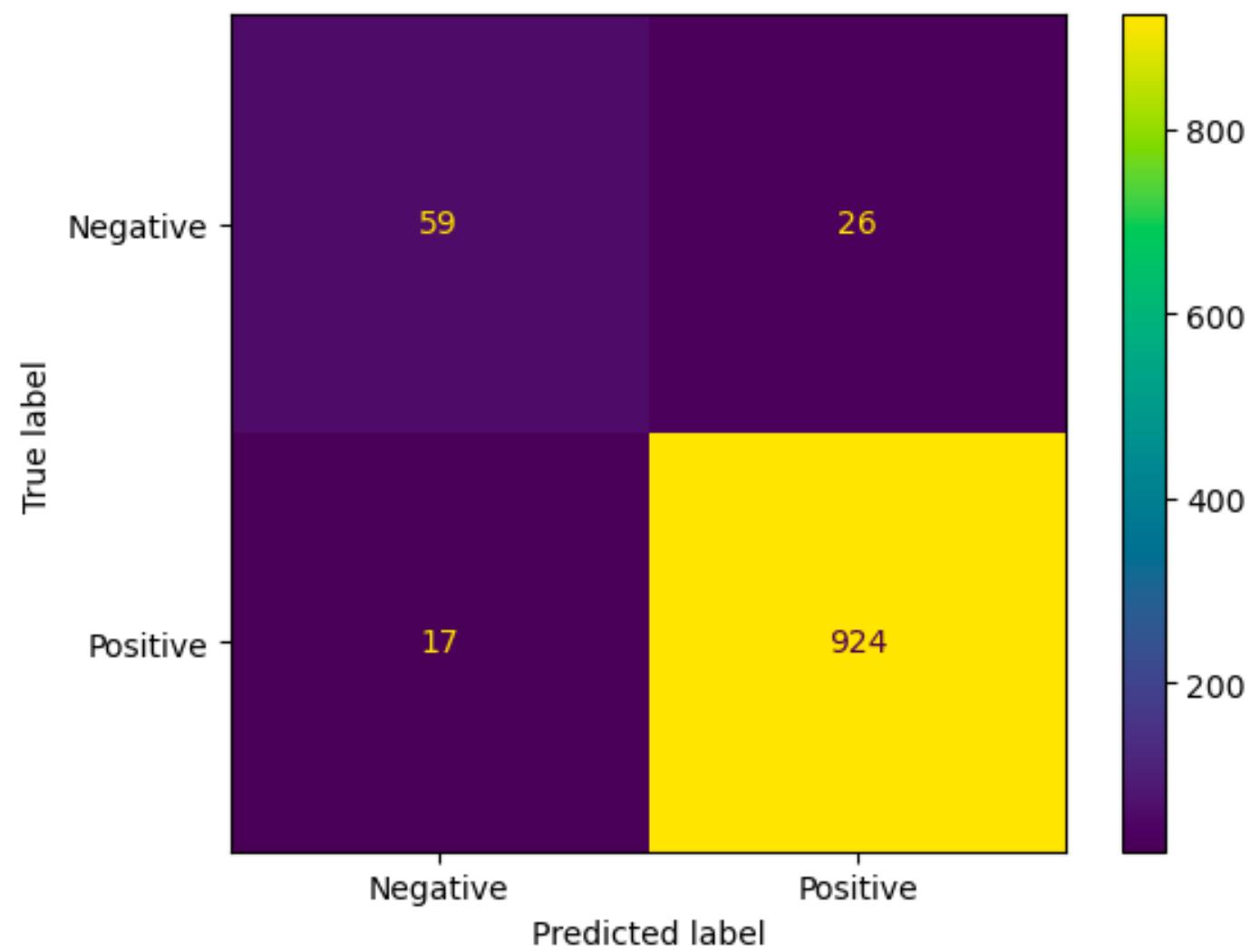
	Baseline		Tuned	
	Count Vectorizer	TF-IDF	Count Vectorizer	TF-IDF
F1*	70%	57%	67%	70%
Precision*	63%	86%	64%	73%
Recall*	78%	42%	71%	67%

*All metrics are scores for the minority class (negative reviews)

Logistic Regression fit on count vectorized features with parameter tuning had the highest F1 score of 73%.

Final Model

× Best Model
× Model



F1 = 73%
Precision = 78%
Recall = 69%

Conclusion

- Logistic regression was the best model for predicting sentiment of HelloFresh reviews from Trustpilot.
- Using F1-score as key metric, model achieved 73% for negative review classification.
- Additional work required to resolve class imbalance - SMOTE
- Neutral reviews were excluded - consider multi class classification.
- Collecting data from multiple sources would be valuable.