Central Garden & Pet Capstone Proposal

CENTRAL
GARDEN & PET

Agenda

1. Problem Statement
2. Project Plan
3. Data Understanding
4. Objective/Scope
5. Metric/Success Criteria
6. Methodology
7. Demo
8. Way Forward

CENTRAL
GARDEN & PET

Problem Statement

Background: The client, Central Garden & Pet (Sponsor), wants to develop a reproducible analytical procedure to utilize customer reviews to extract useful insights.

Problems to solve:

- How to identify the overall sentiment of customers for insights and turn them into more efficient and targeted strategies
- What factors/aspects drive those sentiments? - product, delivery, and so on.

CENTRAL
GARDEN & PET

Project Plan

- Exploratory Data Analysis
- Data Cleaning
- Feature Engineering

- Evaluate performance of initial Few-Shot text classification on validation set
- Experiment different predefined labels with the Few-Shot method
- Repeat Sentiment Analysis, and evaluate its performance on validation set
- Iterate these steps several times and pick the best resulting Few-Shot classification strategy

Data Preparation → **Modeling** → **Performance Monitoring/re-adjustment** → **Development** → **Final Deliverables**

- Cross-validation: Train, Test and Validation sets
  - Split to 80%:10%:10%
- Three Methodologies
  - Zero-Shot text classification (initial)
  - Few-Shot text classification
  - Sentiment Analysis

- Apply selected model to the test set and evaluate its performance

- Research docs, presentation on key findings, reproducible codes, etc.

(Iterate steps from EDA to Deployment for each brand and each retailer)

CENTRAL GARDEN & PET

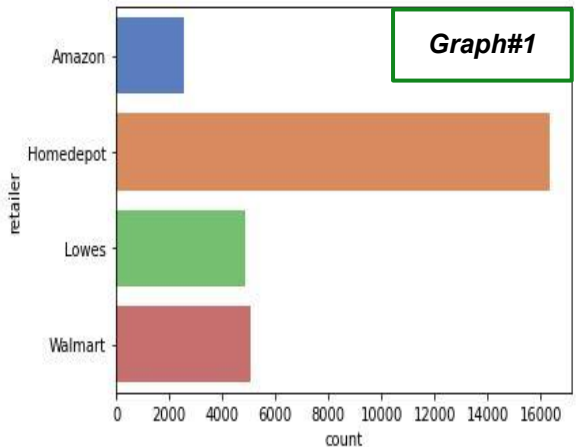
Data Understanding

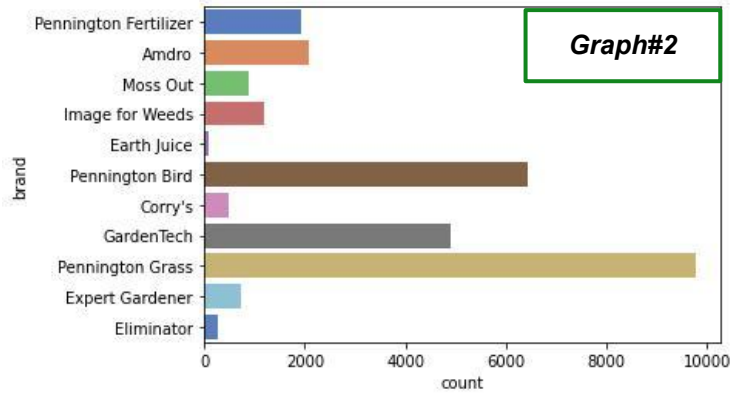
Exploratory Data Analysis:

The review data was collected from **4 retailers** and **11 brands** from **2021-06-01** to **2022-06-30**
Total of **28,786** reviews.

Home depot was the retailer with the most customer reviews. (See ***Graph#1***)

Pennington Grass was the brand with the most customer reviews.（See ***Graph#2***)





CENTRAL
GARDEN & PET

Objective/Scope

Overall Sentiment

Investigating the overall sentiment of customer satisfaction and identifying contributing factors for the sentiment

Key Factors

Factors could including **price**, **delivery speed**, **convenience**, **product effectiveness**, **quality**, **unique selling propositions**, **marketing outreach**, etc

Further Strategies

Investigating supporting insights for further strategies for **individual brands**

CENTRAL
GARDEN & PET

Metrics/Success Criteria

Benchmark Measurement Questions:

- Does **high price** link to negative reviews?
- Does **efficiency** of the product contribute to positive or negative reviews?
- Do products that customers find **easy to use** have positive reviews?
- Does the **effectiveness** of the products contribute to positive reviews?

Intrinsic Metrics- Focus on the performance of the models

- **Topic Coverage**
- **Prediction Error**

Extrinsic Metrics- Focus on the performance of the final outcomes

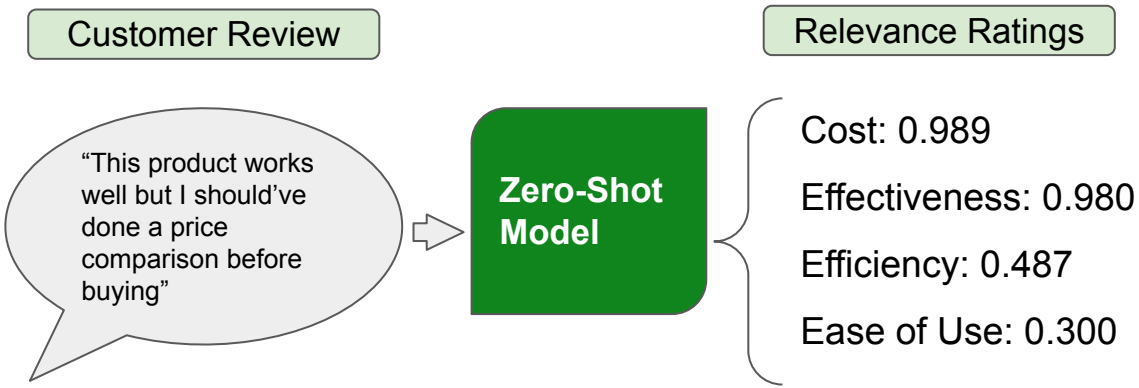
- **Goals Coverage**

CENTRAL
GARDEN & PET

Methodology: Zero-Shot Classification

What is Zero-Shot Classification?
Developed by Facebook software engineers, the Zero-Shot Classification is a Machine Learning technique with human-like flexibility and efficiency for customer review analysis. The model can classify any text with given labels without any prior data.

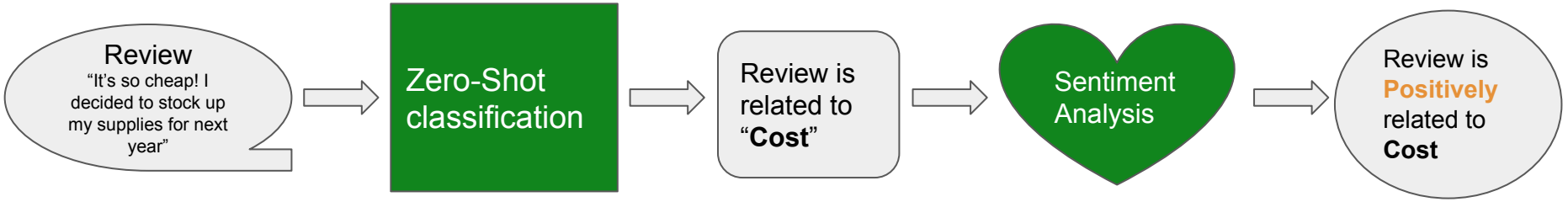
How does it work?
We created four basic category labels for classification - **cost, effectiveness, efficiency, and ease of use**. Each review will get a probability rating for each category depending on the semantics of the review.



CENTRAL GARDEN & PET

Methodology: Sentiment Analysis

Sentiment Analysis to determine the sentiments among the categories
Now that we have categorized the topics of each review under the predefined categories, we are interested in finding out the positive and negative sentiments associated with each topic.



Combining the Zero-Shot classification with Sentiment analysis we can intelligently label each of the reviews to be related to one or more of **price**, **delivery**, **effectiveness**, and/or **ease-of-use** with information on its sentiment.

review_lower	cost_m	efficient_m	effective_m	ease of use_m	sentiment
this works great for keeping the brass from gr...	0.251741	0.998199	0.999644	0.376084	Positive
waiting for it to work. will write more later	0.325730	0.027556	0.012378	0.095561	Positive
[this review was collected as part of a promot...	0.531525	0.998460	0.999477	0.999256	Positive
i received product reimbursement in exchange f...	0.716651	0.968388	0.996363	0.774749	Positive

CENTRAL GARDEN & PET

Demo: Zero-Shot Classification

Review 1:

```
grass grows fast and a lot of weeds. this works excellent in killing the
weeds out. it was easy to use just hook it to a hose and start spray.
```

cost_m	efficient_m	effective_m	ease of use_m
0.531525	0.99846	0.999477	0.999256

Review 2:

```
'waiting for it to work.  will write more later
```

cost_m	efficient_m	effective_m	ease of use_m
0.32573	0.027556	0.012378	0.095561

Observation:

Zero shot classification does well in identifying an effect and labeling the observations correctly.

CENTRAL GARDEN & PET

Demo: Zero-Shot Classification (shortcomings)

Review 3:

```
it killed a few bugs, but it rained the next day and we say
some bugs remained alive.
```

cost_m	efficient_m	effective_m	ease of use_m
0.653789	0.245608	0.780158	0.02614

Review 4:

```
great product - all the birds love it!
```

cost_m	efficient_m	effective_m	ease of use_m
0.039136	0.873667	0.993279	0.655259

Observation:

- Zero shot classification model sometimes gets confused, especially if the reviews are short.
- Need to set a cutoff threshold.

CENTRAL GARDEN & PET

Demo: Zero-Shot Classification (shortcomings)

Review 3:

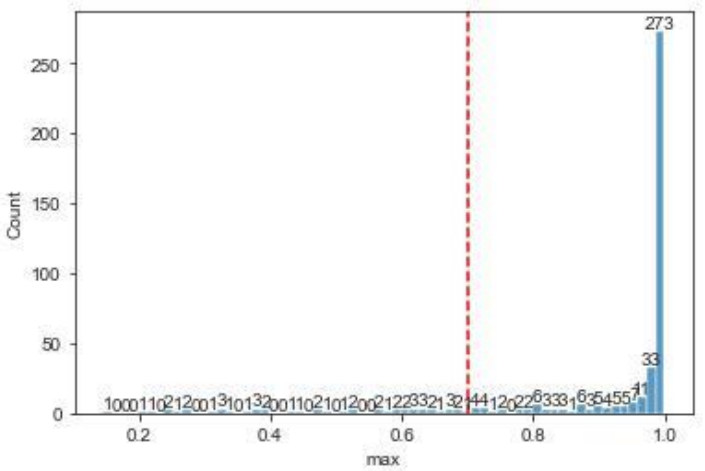
```
it killed a few bugs, but it rained the next day and we say
some bugs remained alive.
```

cost_m	efficient_m	effective_m	ease of use_m
0.653789	0.245608	0.780158	0.02614

Review 4:

```
great product - all the birds love it!
```

cost_m	efficient_m	effective_m	ease of use_m
0.039136	0.873667	0.993279	0.655259



Result:

Zero-Shot classification resulted in 89% of the data being classified into the labels with higher than 0.7 probability level.

Demo: Zero-Shot Classification (shortcomings)

Review 5:

```
nice quality sunflower seed, the cardinals and squirrels
love it.
```

cost_m	efficient_m	effective_m	ease of use_m
0.232189	0.891521	0.993074	0.289773

Observation:

- Zero shot classification mislabels some reviews.
- Fine-tuning the classification model.
- Setting a minimum word count.

CENTRAL GARDEN & PET

Sample Deliverables

Homedepot

	Cost	Efficient	Effective	Ease of Use
Positive	34.38	84.38	76.02	87.91
Neutral	17.19	8.75	13.27	2.2
Negative	48.44	6.88	10.71	2.2

Amazon

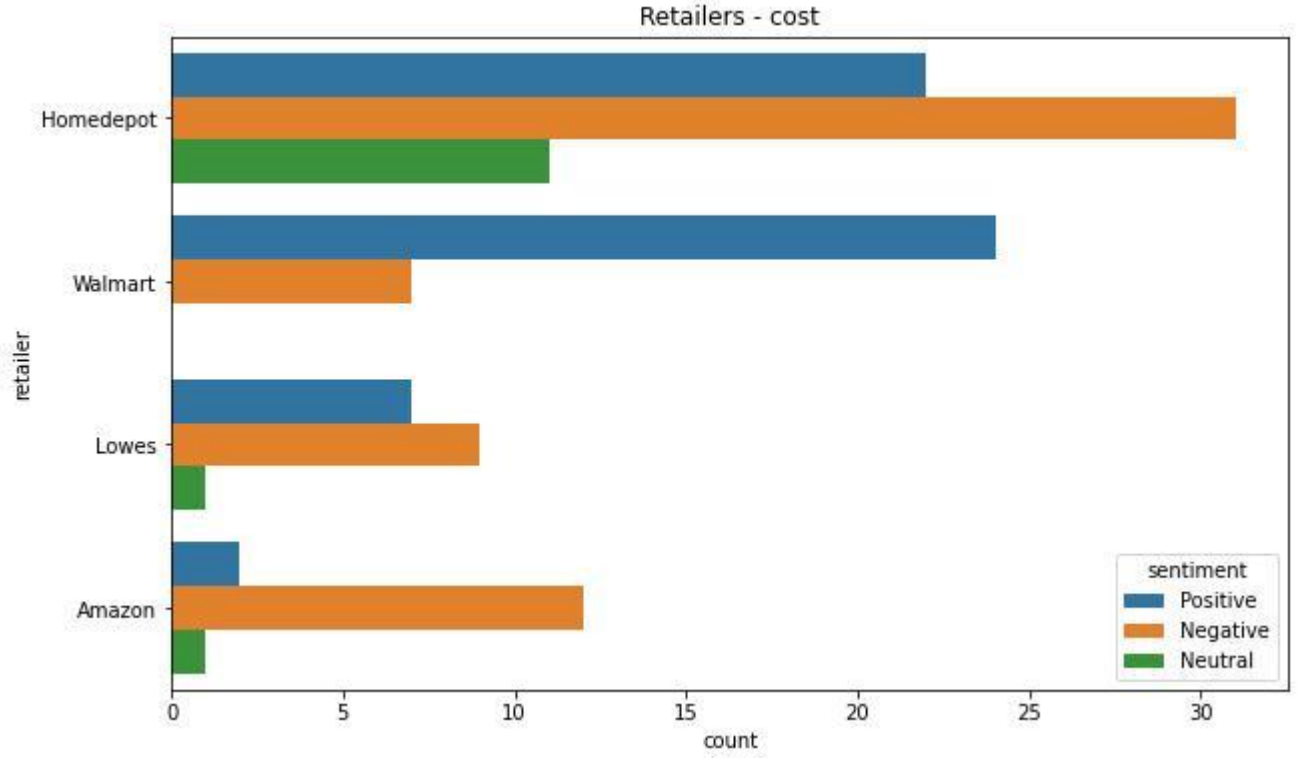
	Cost	Efficient	Effective	Ease of Use
Positive	13.33	50	45.45	62.5
Neutral	6.67	22.22	22.73	37.5
Negative	80	27.78	31.82	0

Output:

Dashboard view of the heat-map of the types of reviews and sentiment based on the retailers.

CENTRAL
GARDEN & PET

Sample Deliverables

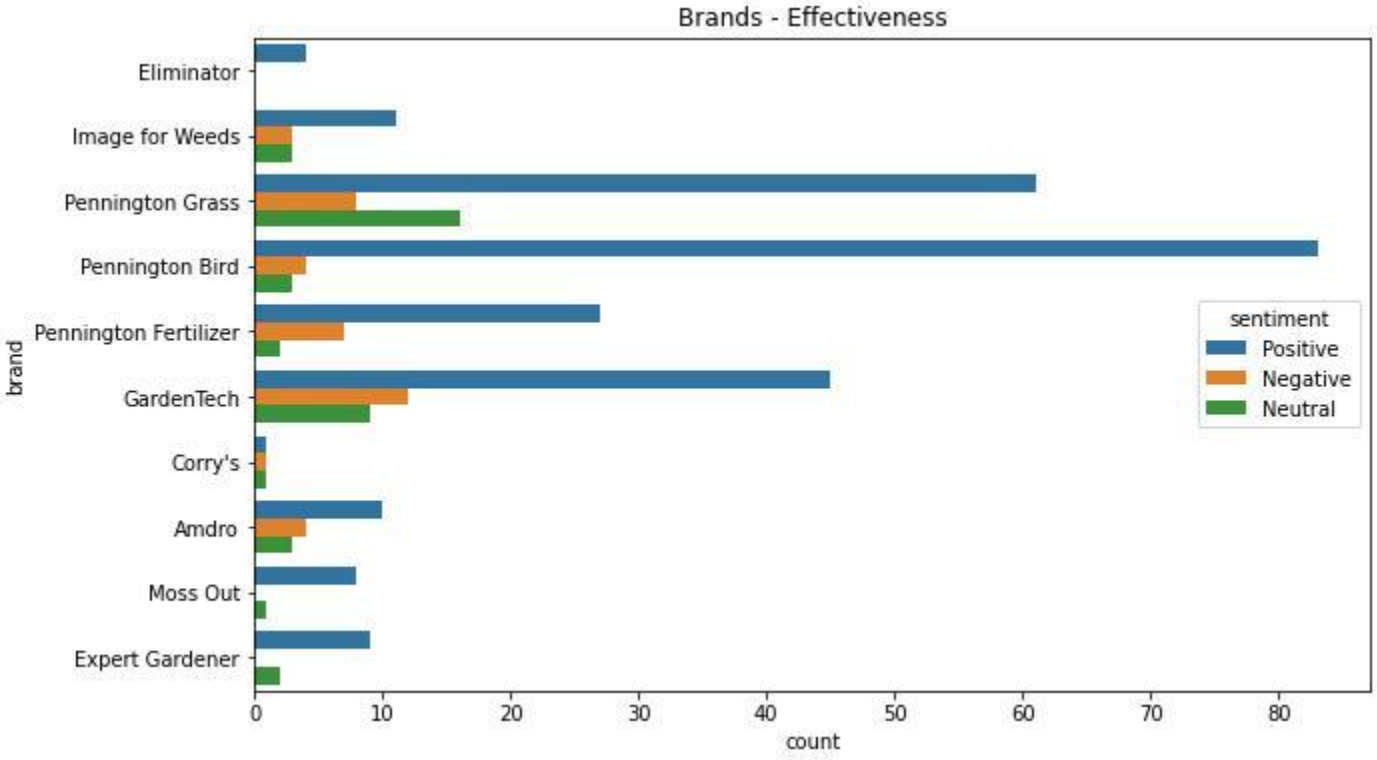


Output:

Distribution of the reviews for each **retailers** and their sentiments.

CENTRAL
GARDEN & PET

Sample Deliverables

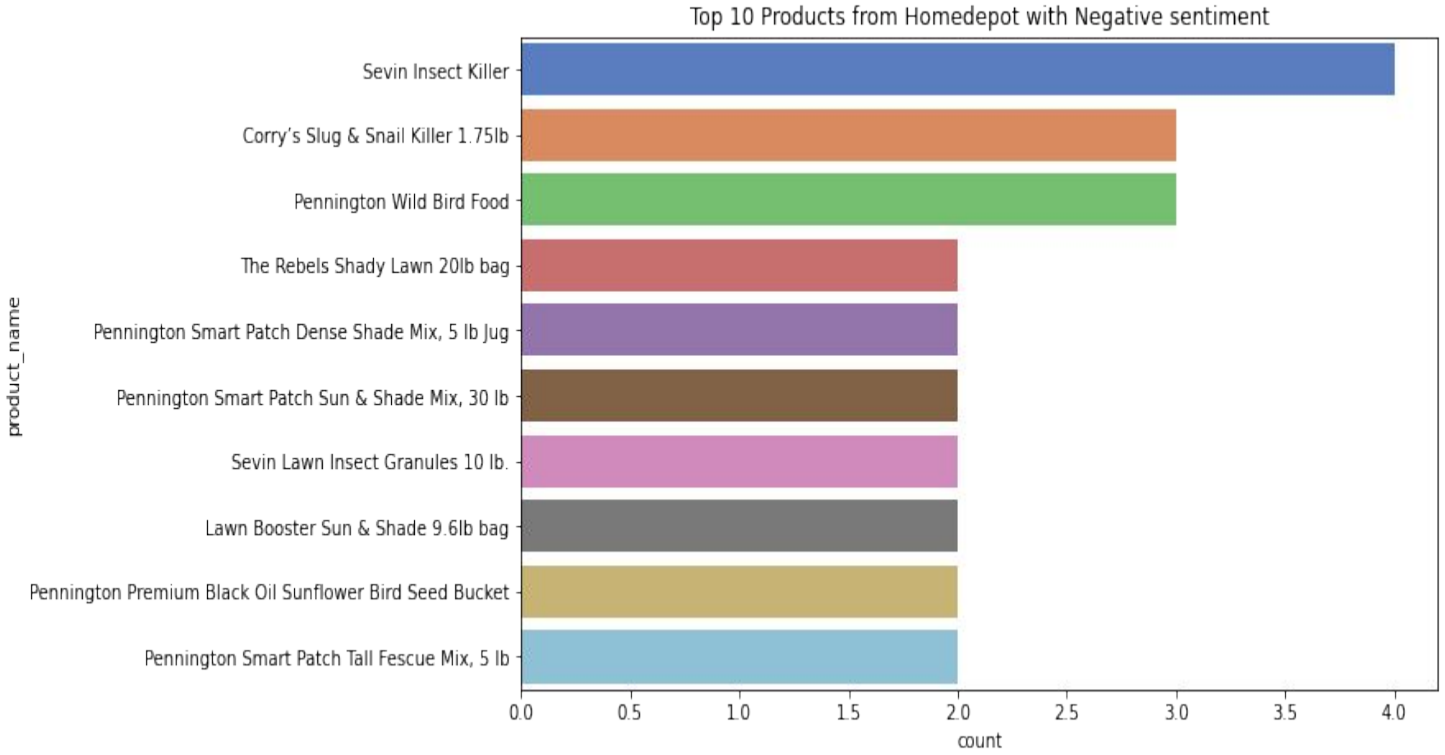


Output:

Distribution of the reviews for each **brands** and their sentiments.

CENTRAL
GARDEN & PET

Sample Deliverables



Output:

List of **products** with the most negative reviews.

CENTRAL GARDEN & PET

Way Forward

- Fine tuning the zero shot model to improve classification.
(currently 88% of the data were classified with higher than 0.7 confidence level)
- Additional analysis to produce intelligent information and visualizations
- Deep dive analysis on products with negative reviews to identify gaps.

CENTRAL
GARDEN & PET

Appendix

CENTRAL
GARDEN & PET

Heatmap of all the retailers

Homedepot

	Cost	Efficient	Effective	Ease of Use
Positive	34.38	84.38	76.02	87.91
Neutral	17.19	8.75	13.27	2.2
Negative	48.44	6.88	10.71	2.2

Lowes

	Cost	Efficient	Effective	Ease of Use
Positive	41.18	81.25	74.07	87.88
Neutral	5.88	8.33	9.26	6.06
Negative	52.94	10.42	16.67	6.06

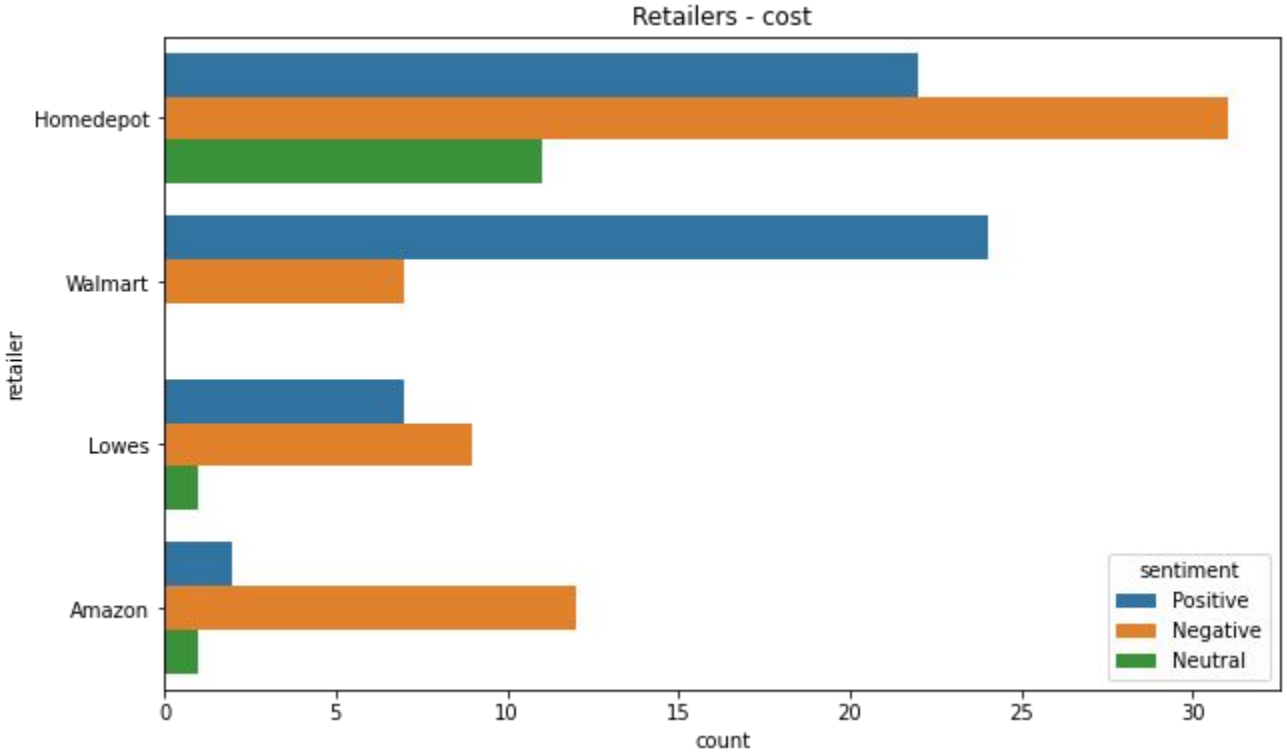
Walmart

	Cost	Efficient	Effective	Ease of Use
Positive	77.42	93.1	90.91	96.88
Neutral	0	5.17	6.06	3.125
Negative	22.58	1.72	3.03	0

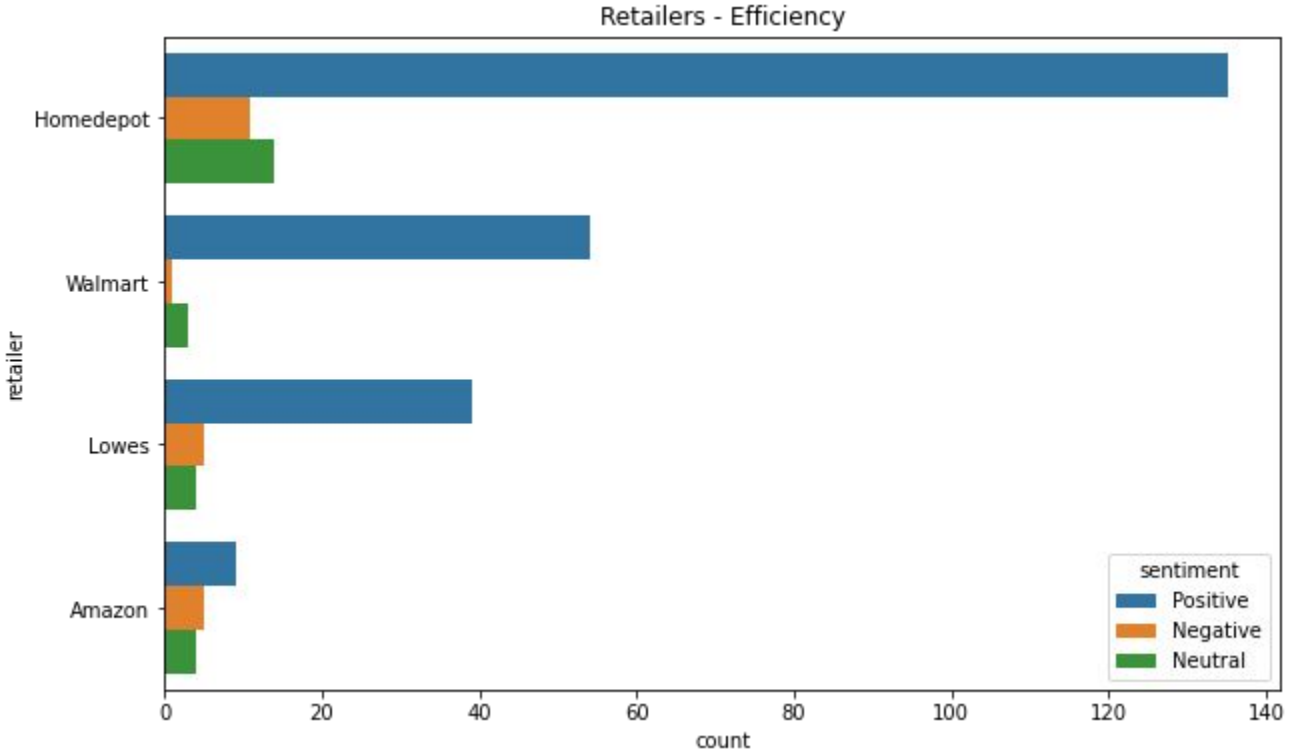
Amazon

	Cost	Efficient	Effective	Ease of Use
Positive	13.33	50	45.45	62.5
Neutral	6.67	22.22	22.73	37.5
Negative	80	27.78	31.82	0

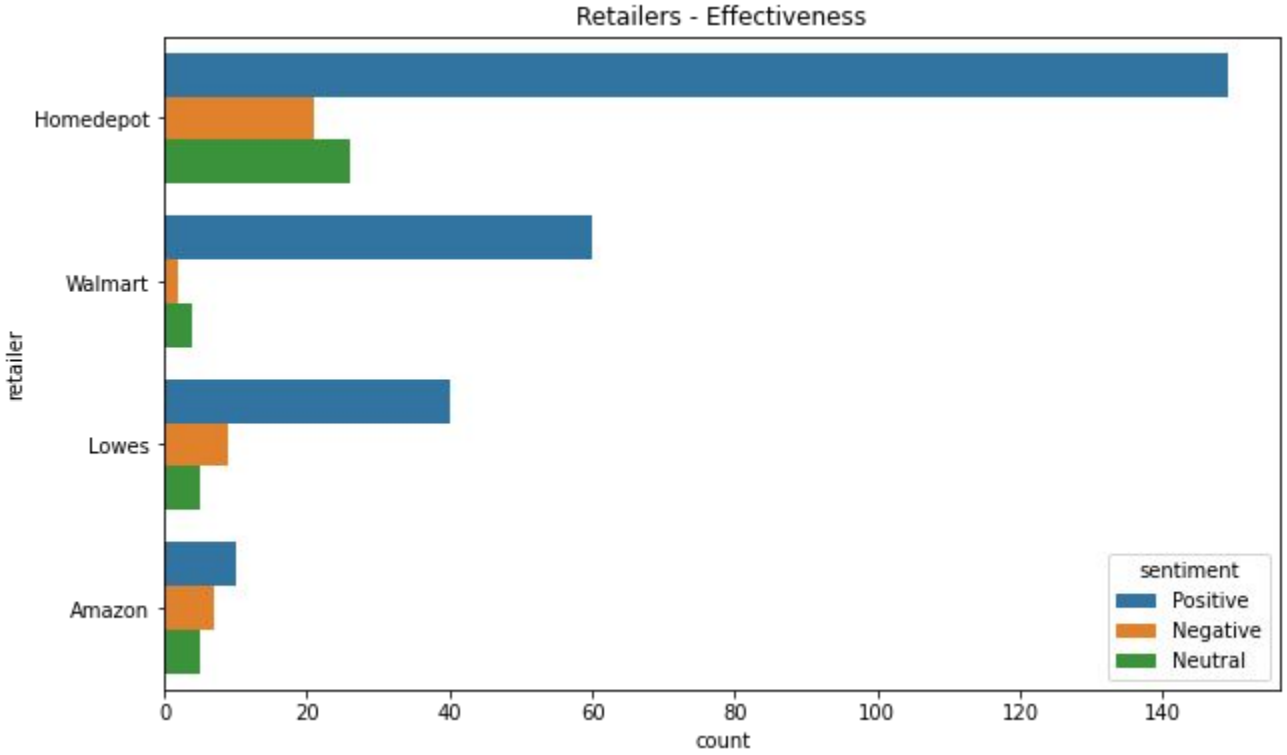
Sentiment Histogram of the Retailers by labels



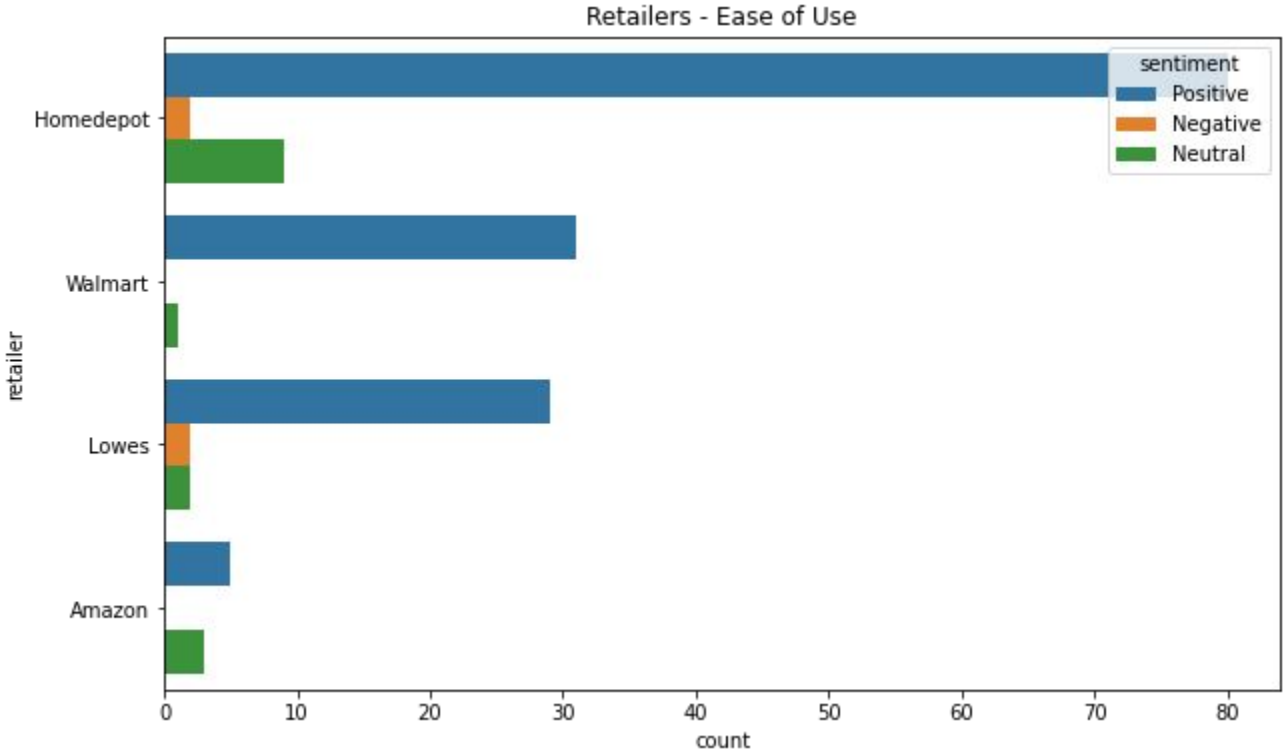
Sentiment Histogram of the Retailers by labels



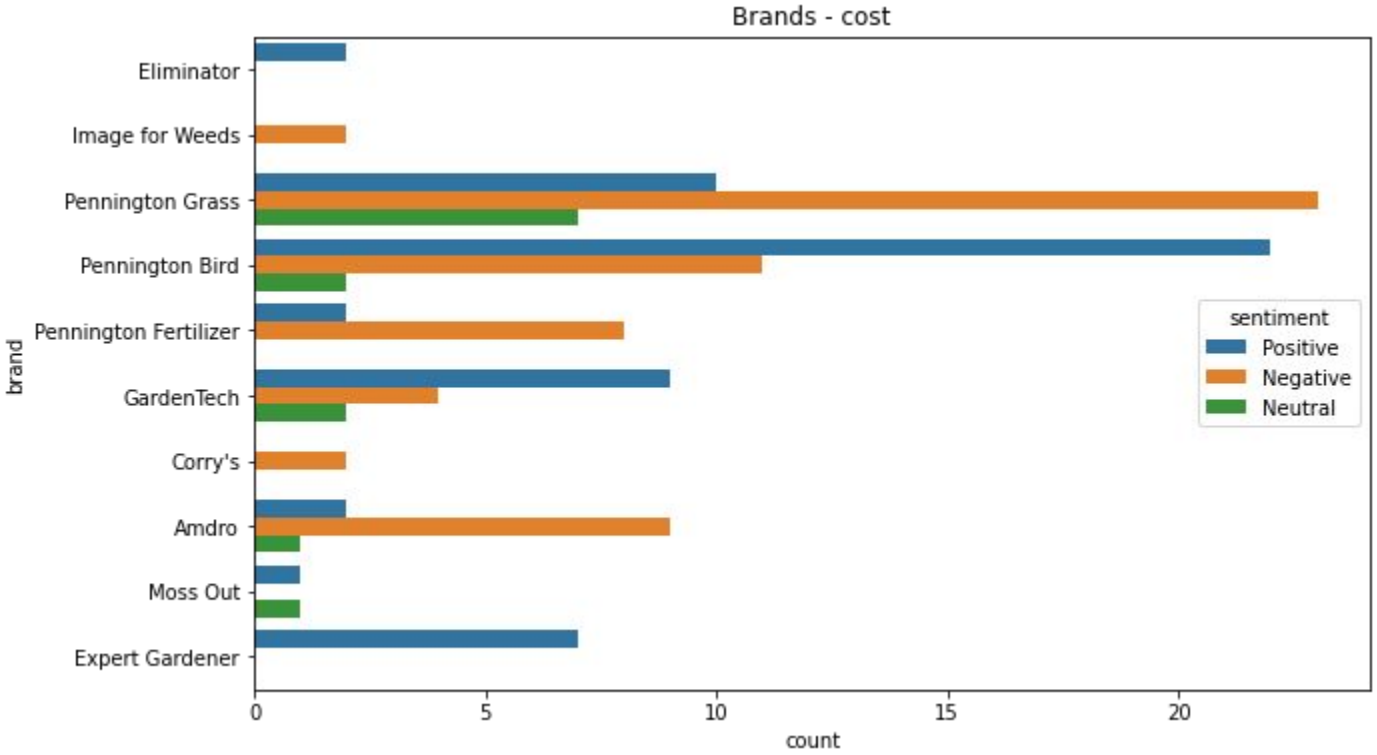
Sentiment Histogram of the Retailers by labels



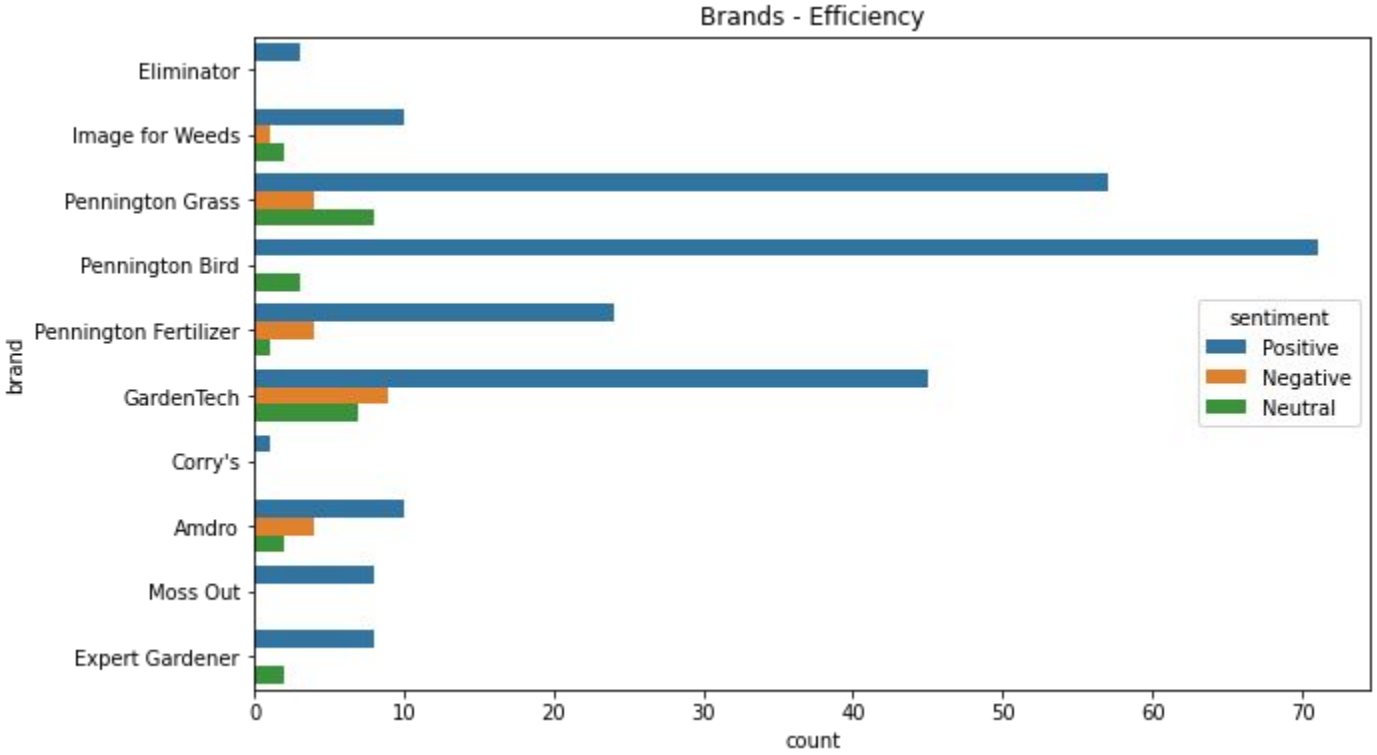
Sentiment Histogram of the Retailers by labels



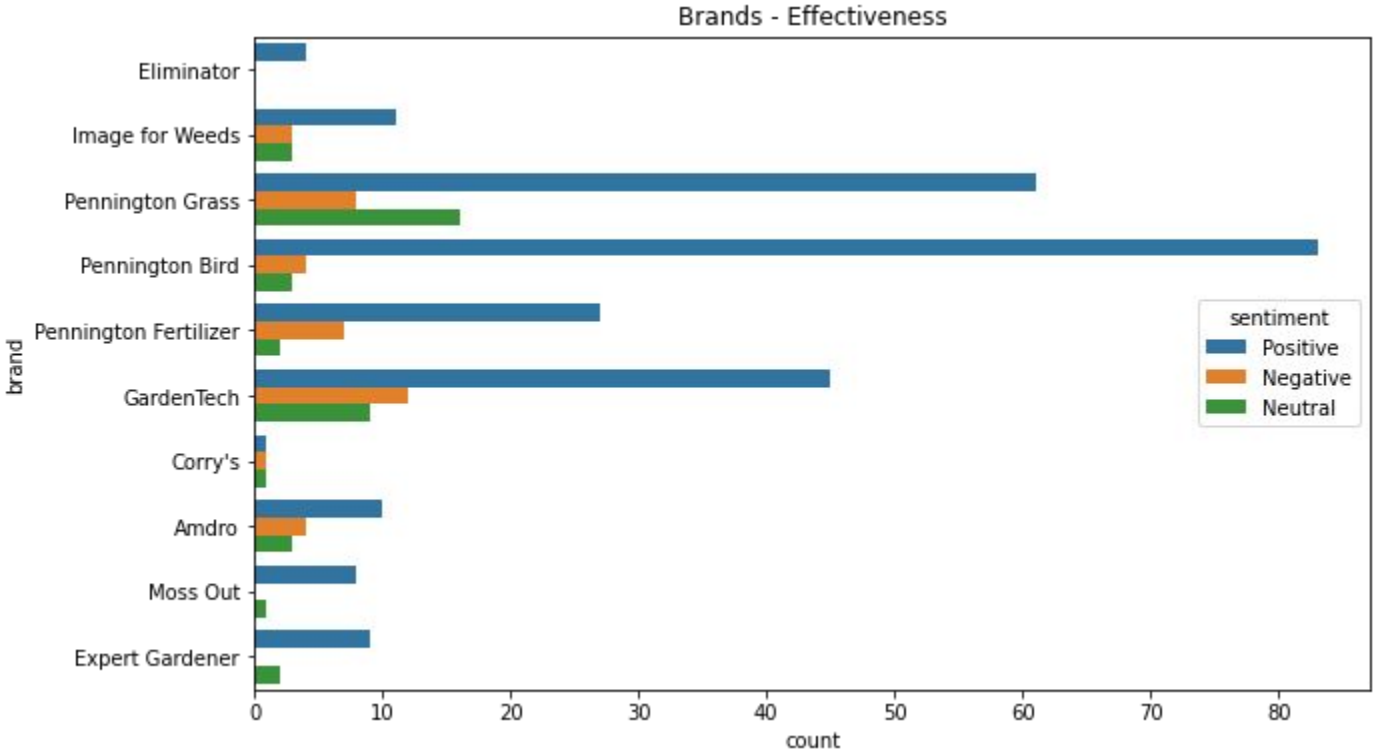
Sentiment Histogram of the Brands by labels



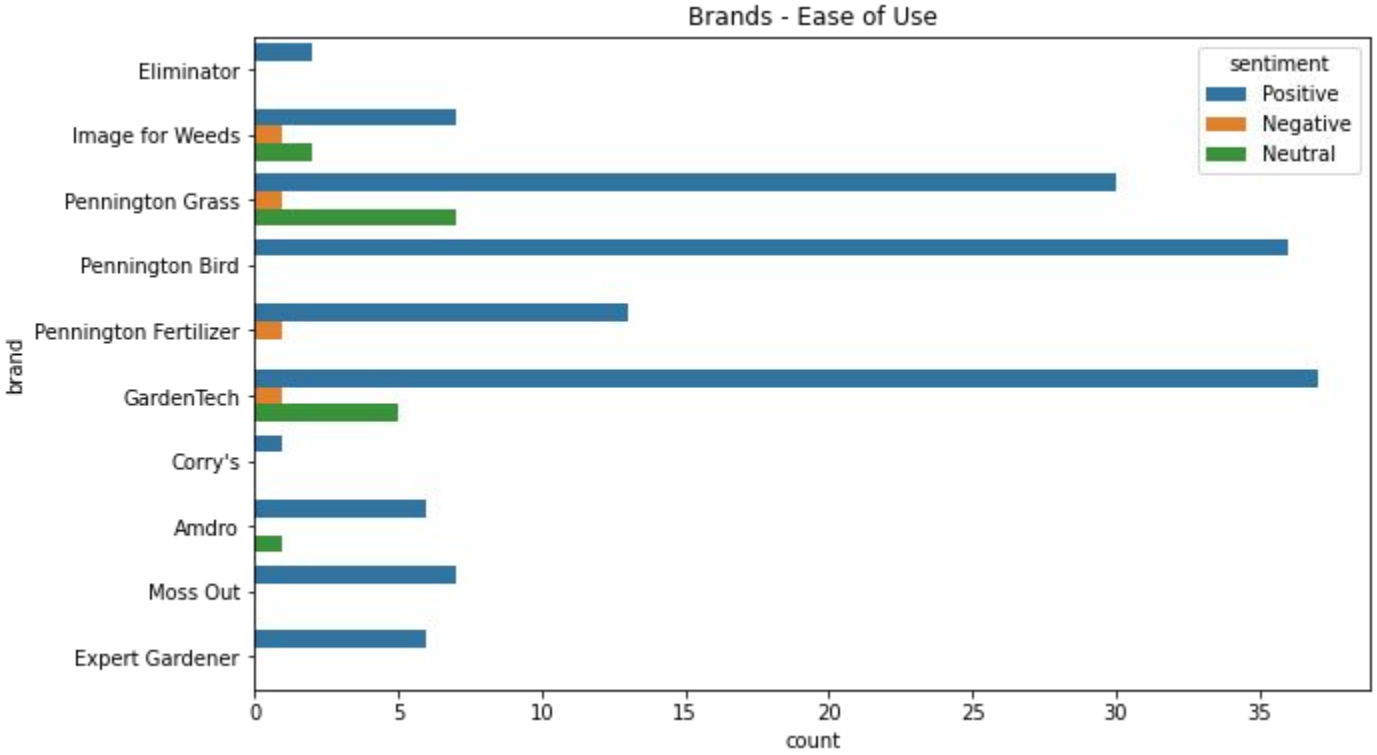
Sentiment Histogram of the Brands by labels



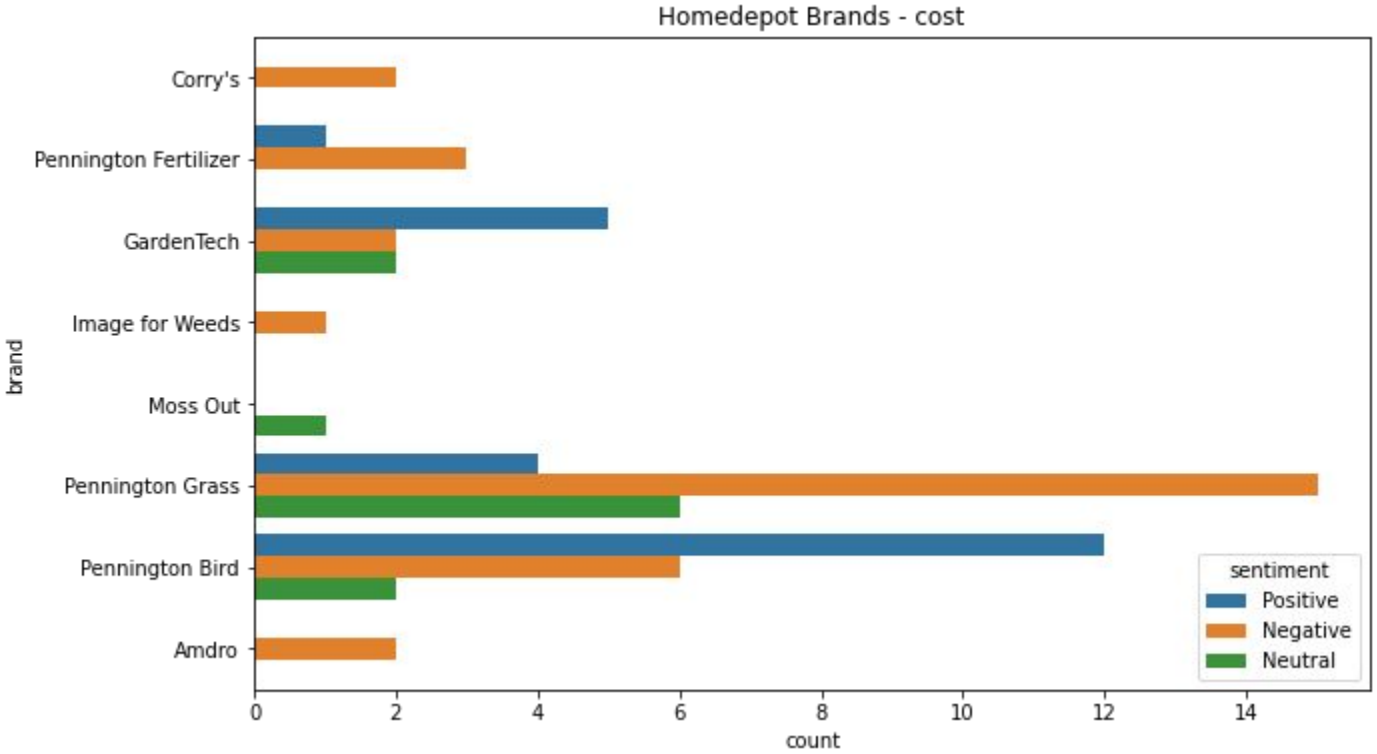
Sentiment Histogram of the Brands by labels



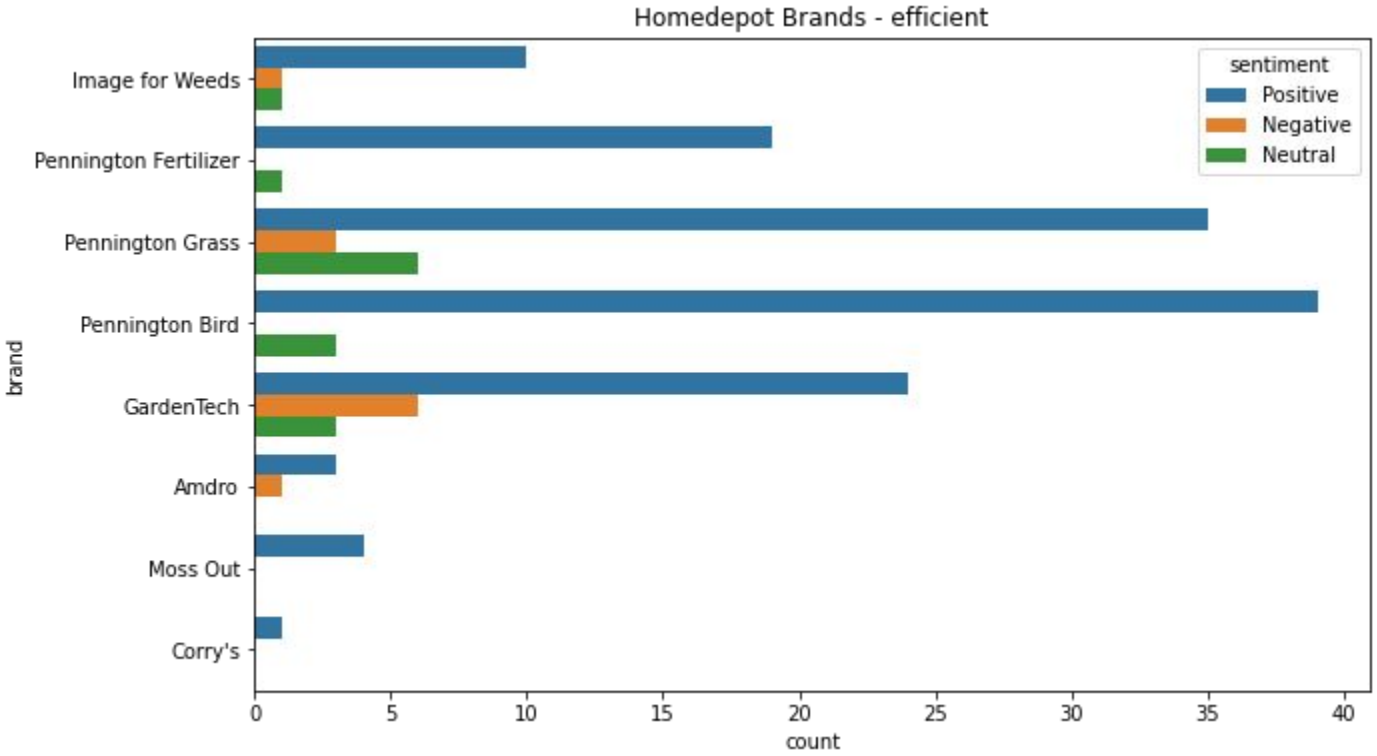
Sentiment Histogram of the Brands by labels



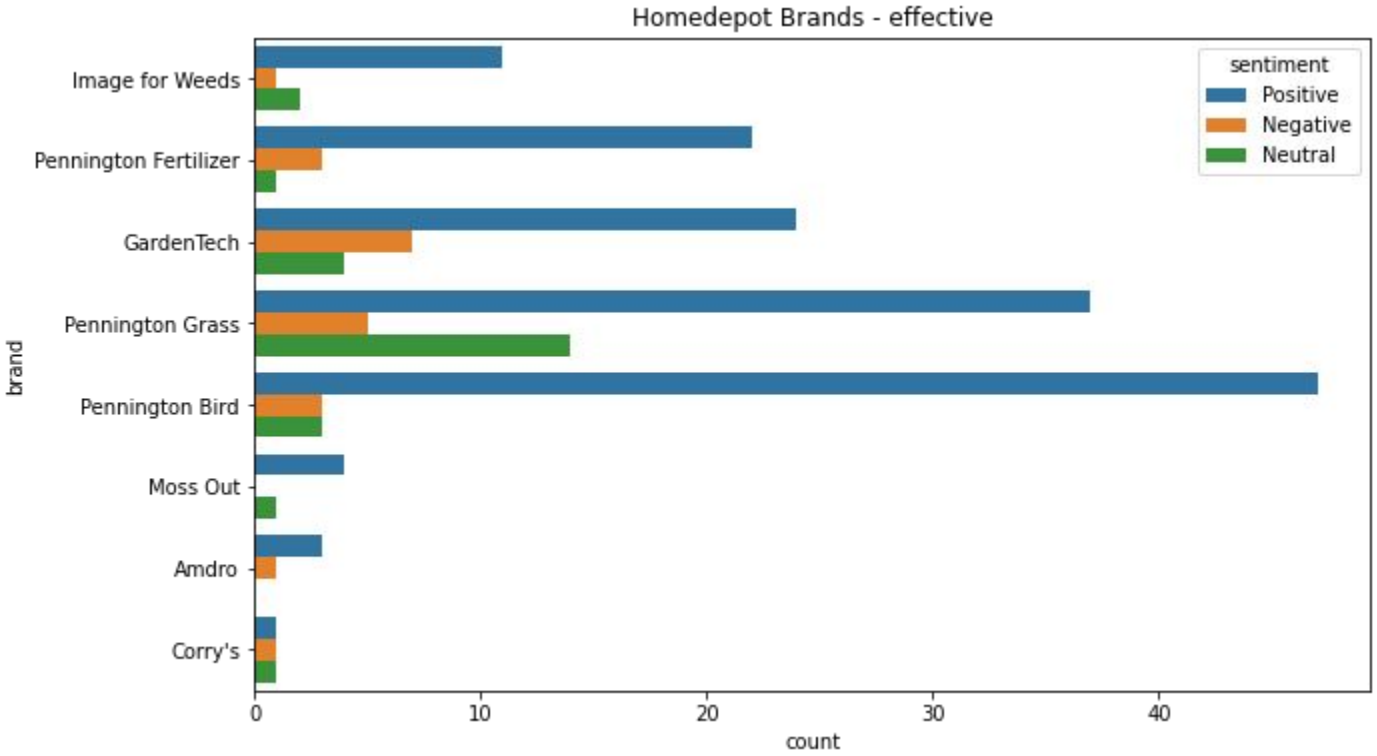
Sentiment Histogram of Homedepot by Brands - Cost



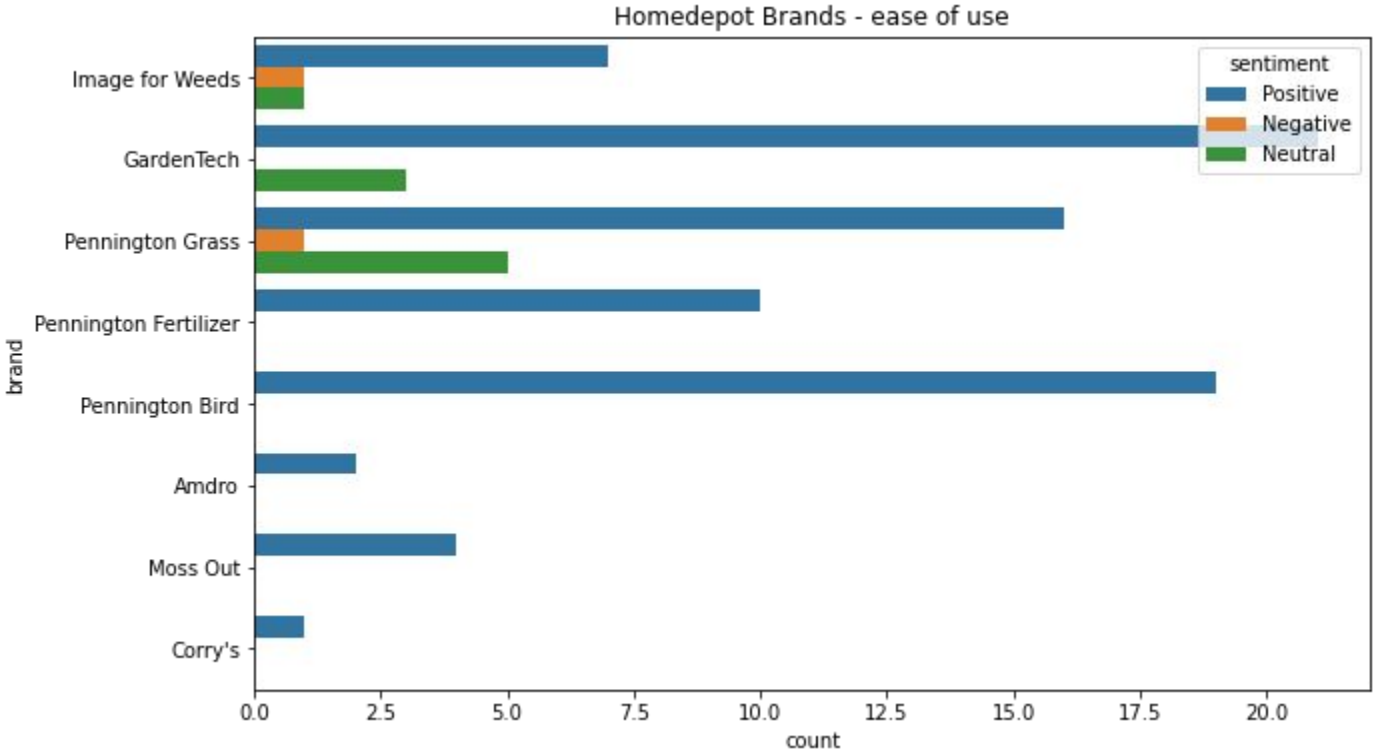
Sentiment Histogram of Homedepot by Brands - Efficient



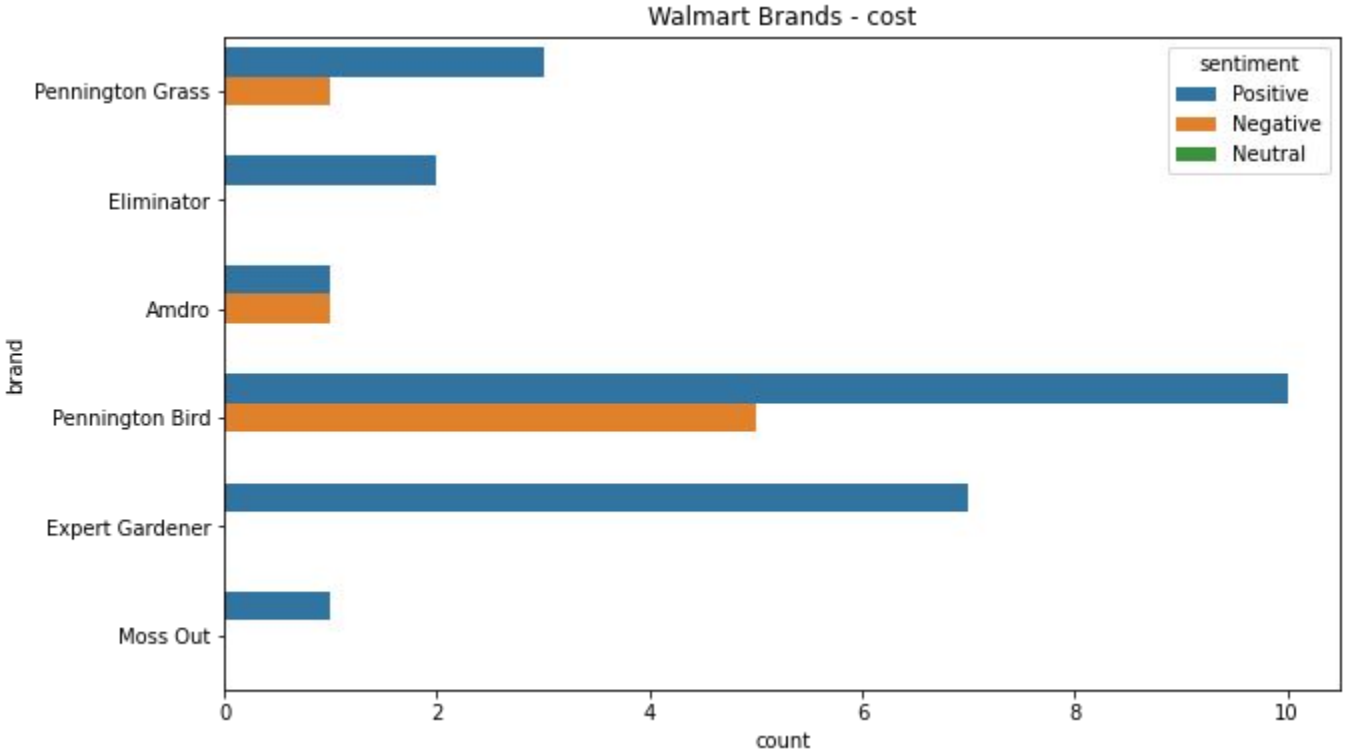
Sentiment Histogram of Homedepot by Brands - Effective



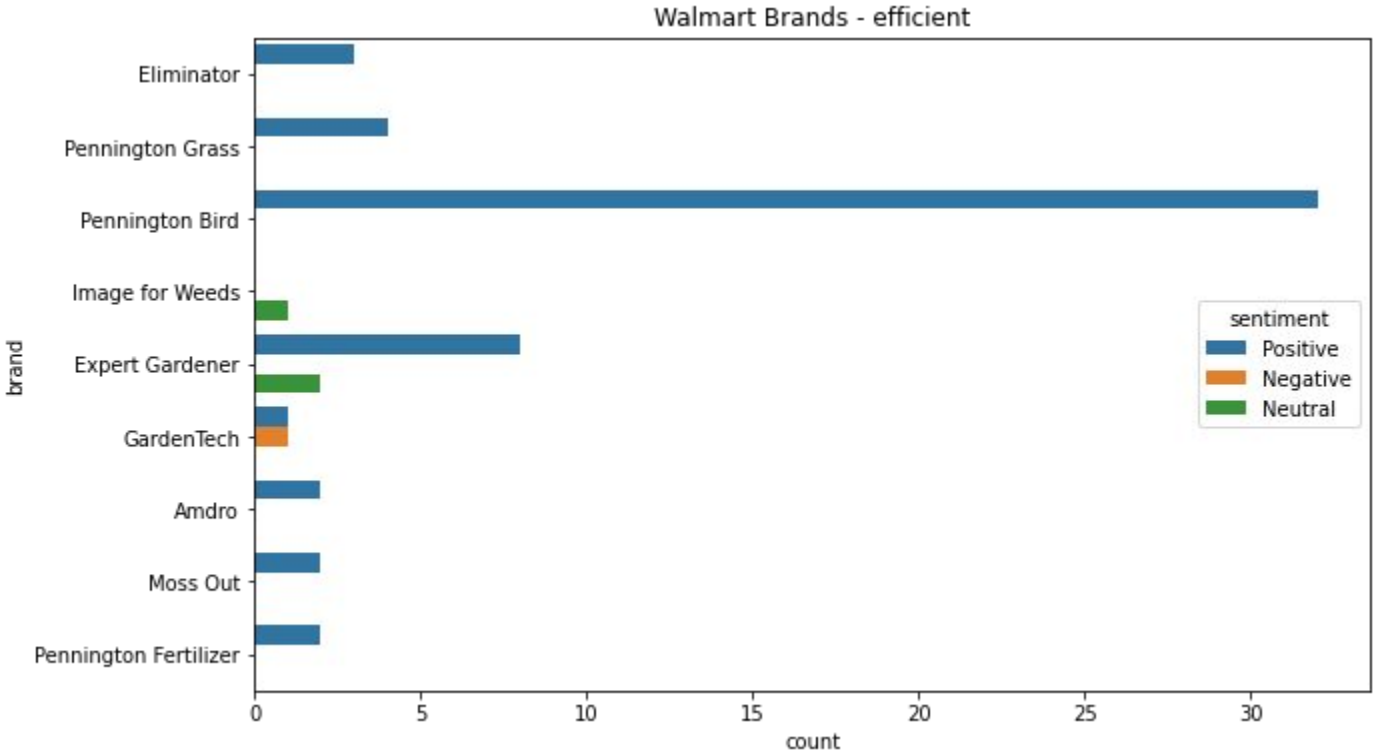
Sentiment Histogram of Homedepot by Brands - Ease of Use



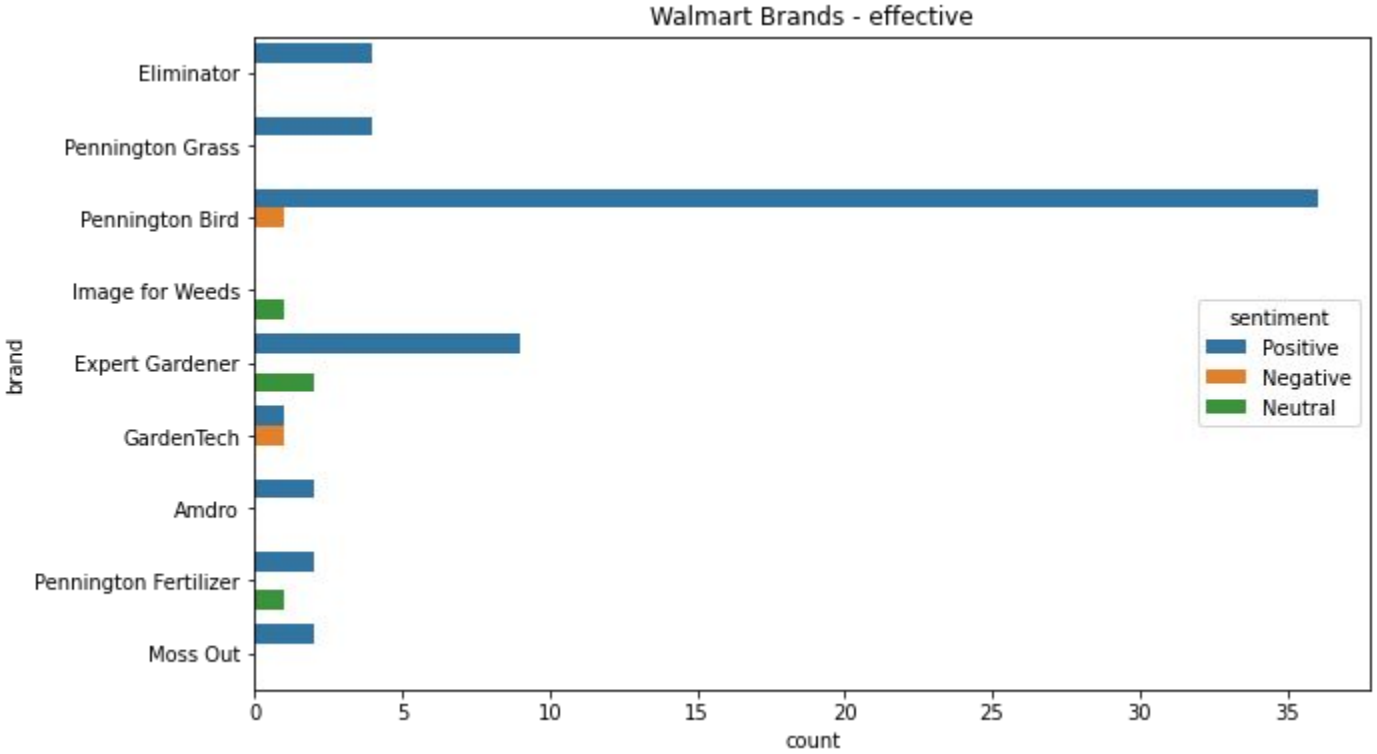
Sentiment Histogram of Walmart by Brands - Cost



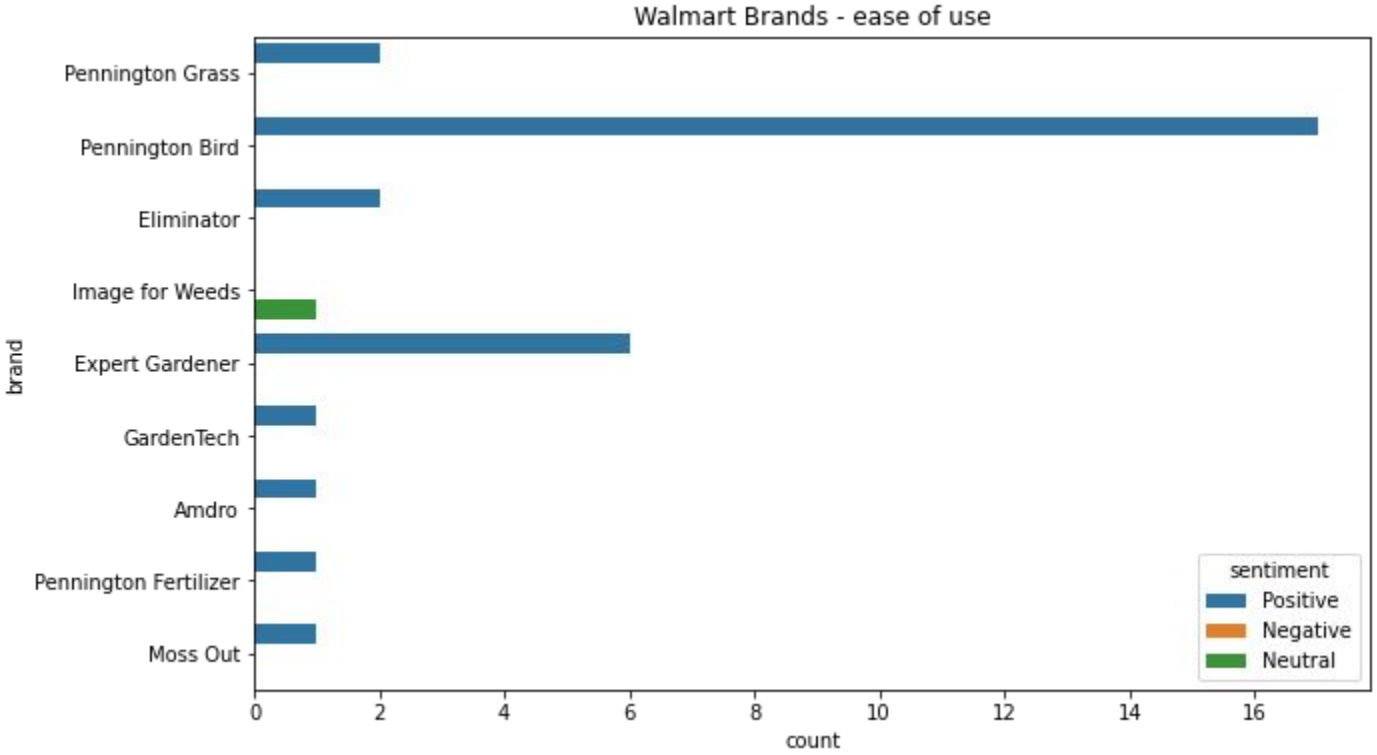
Sentiment Histogram of Walmart by Brands - Efficient



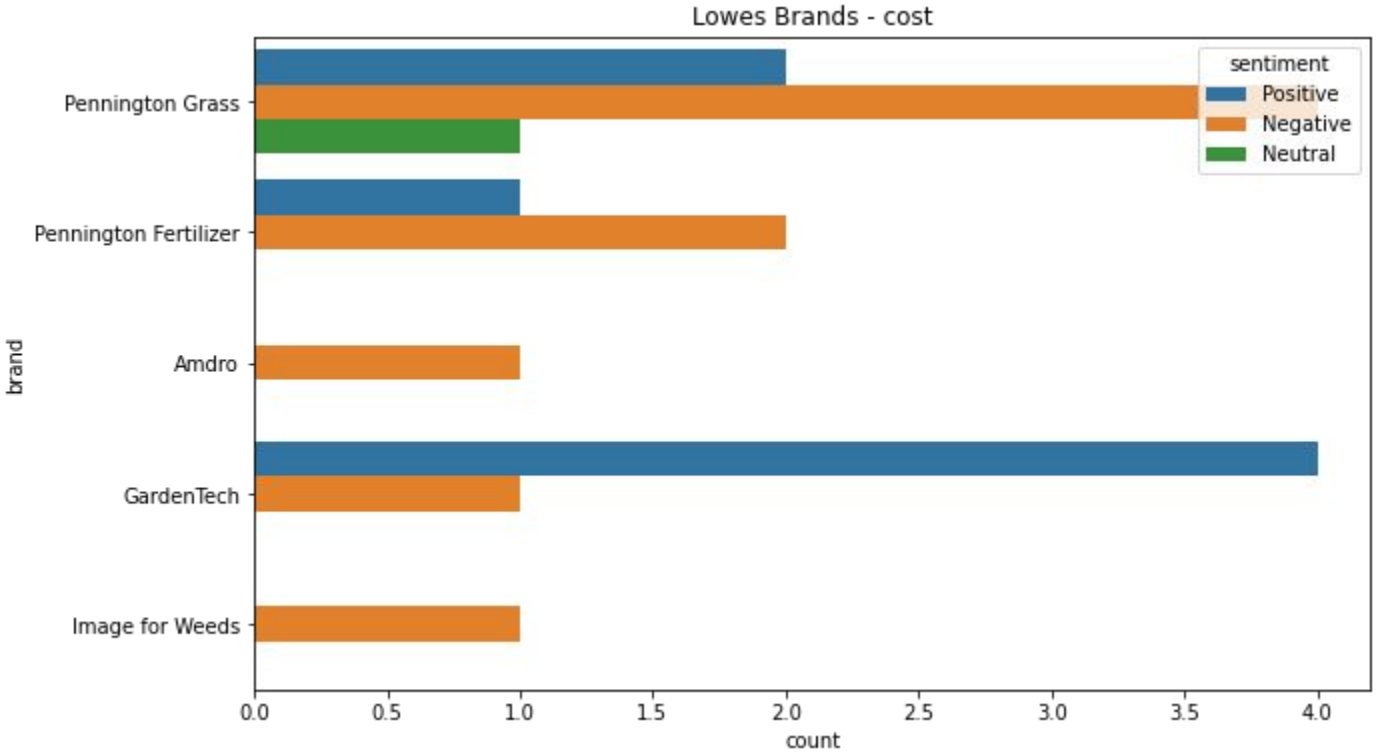
Sentiment Histogram of Walmart by Brands - Effective



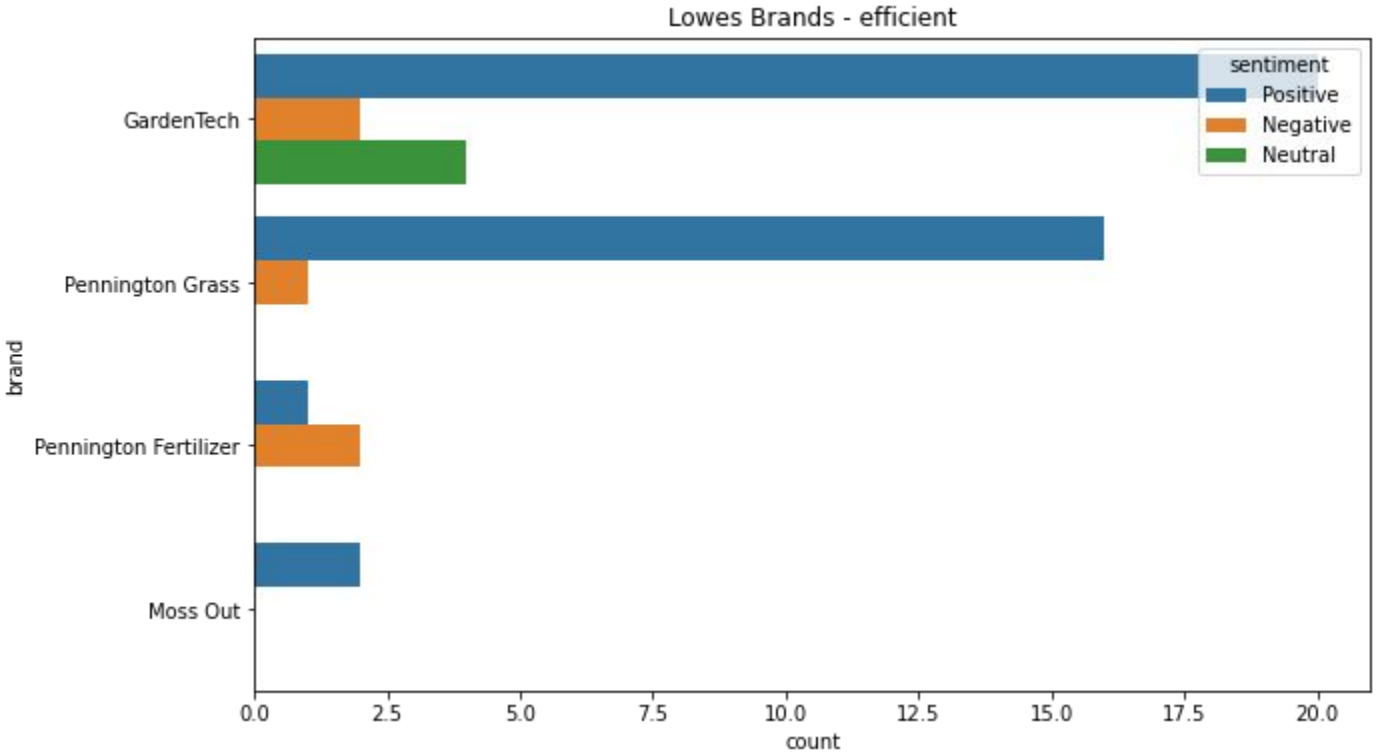
Sentiment Histogram of Walmart by Brands - Ease of Use



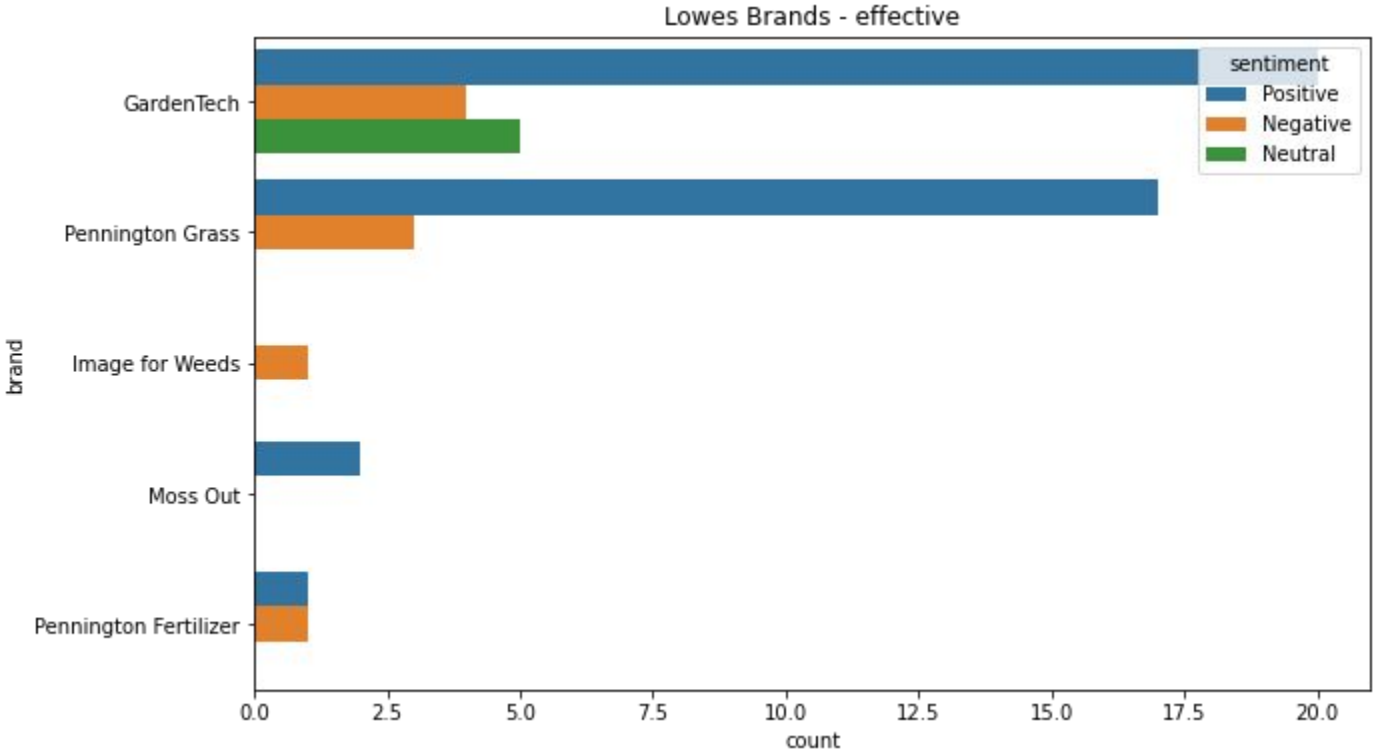
Sentiment Histogram of Lowes by Brands - Cost



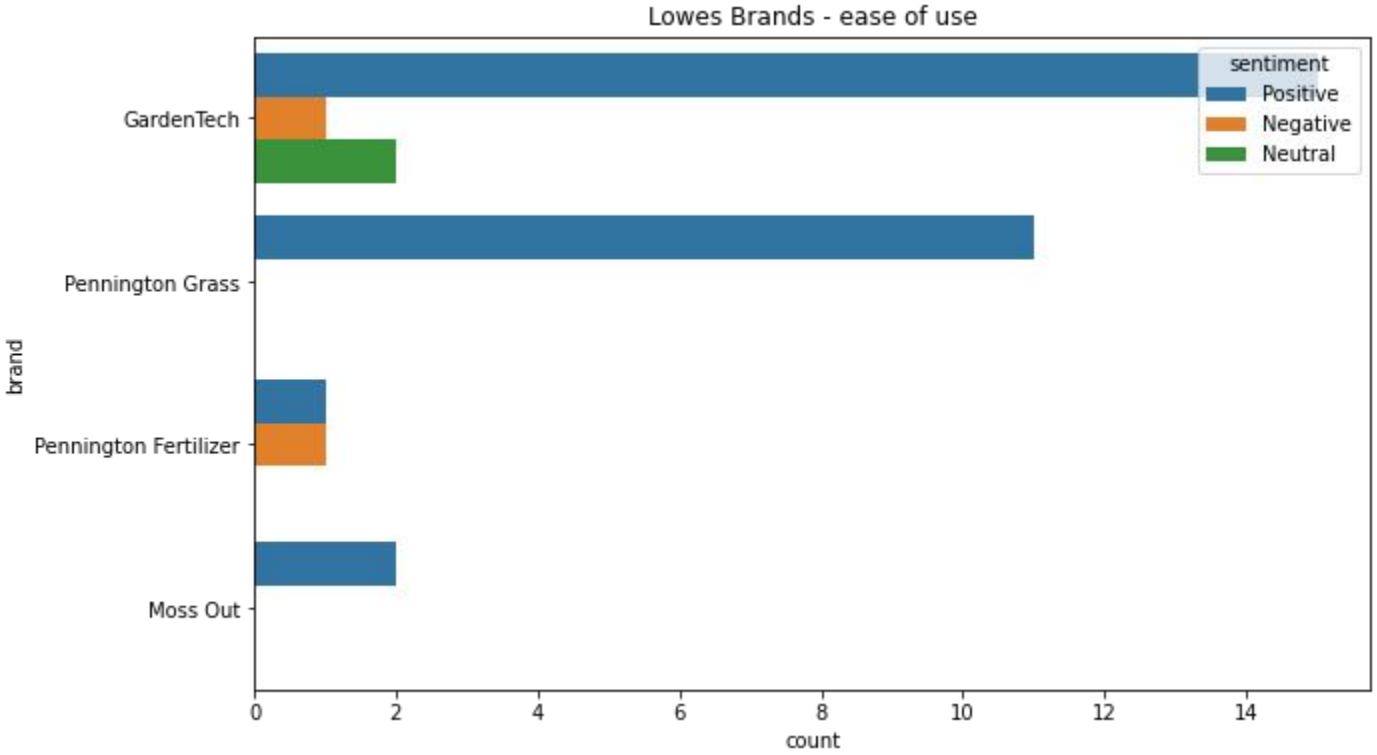
Sentiment Histogram of Lowes by Brands - Efficient



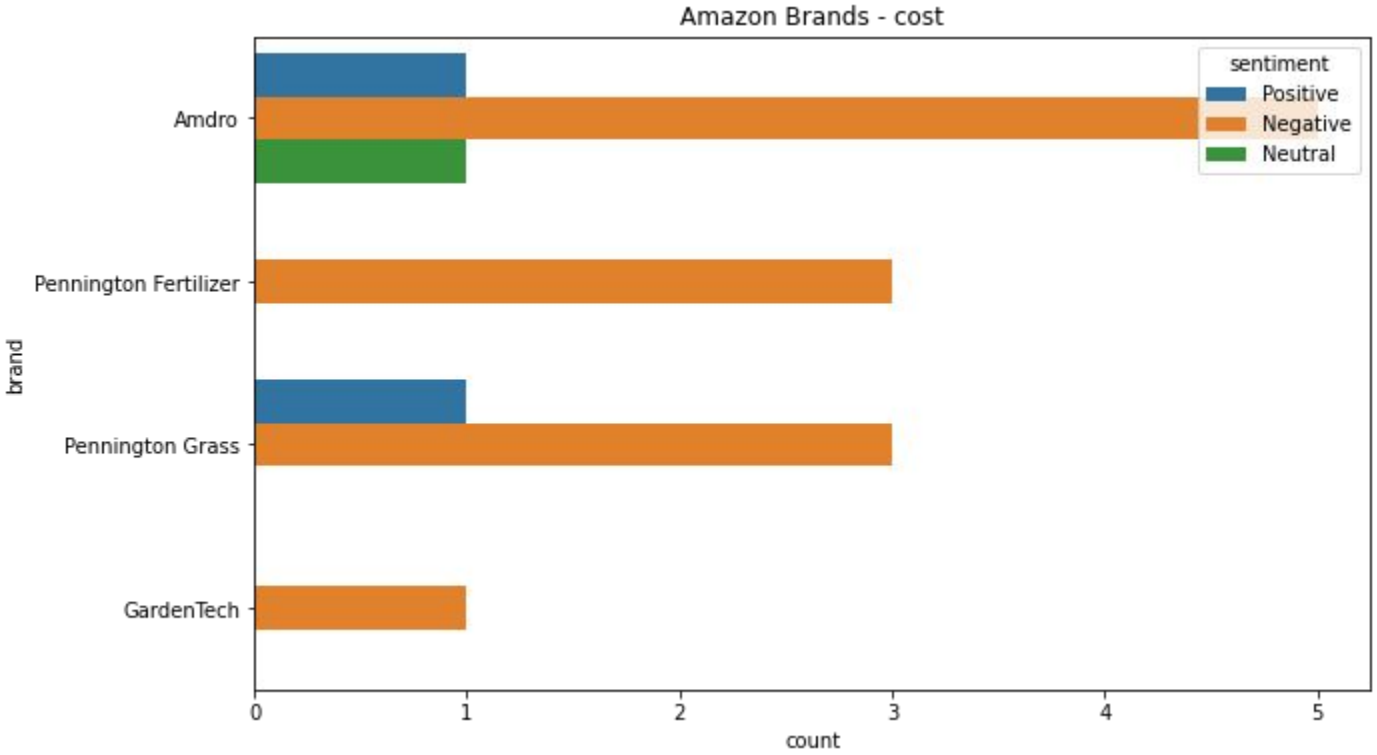
Sentiment Histogram of Lowes by Brands - Effective



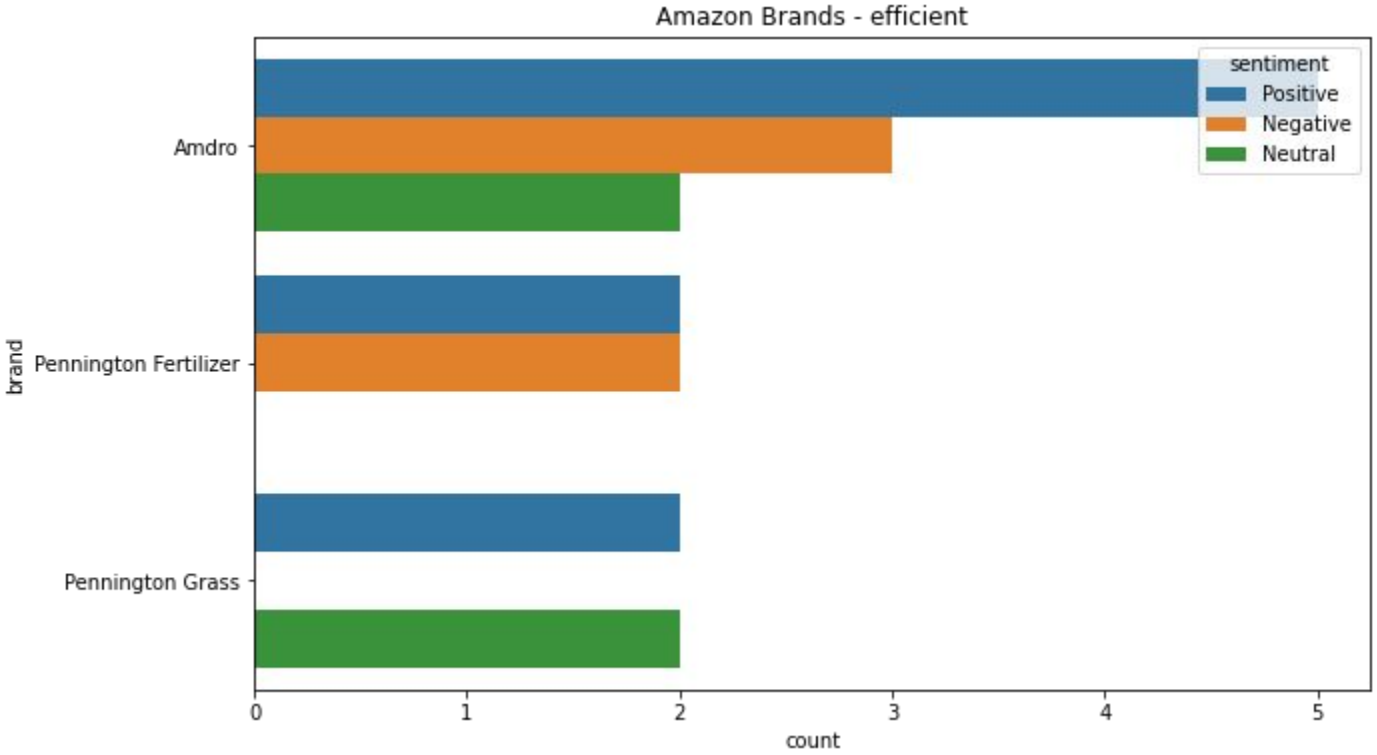
Sentiment Histogram of Lowes by Brands - Ease of Use



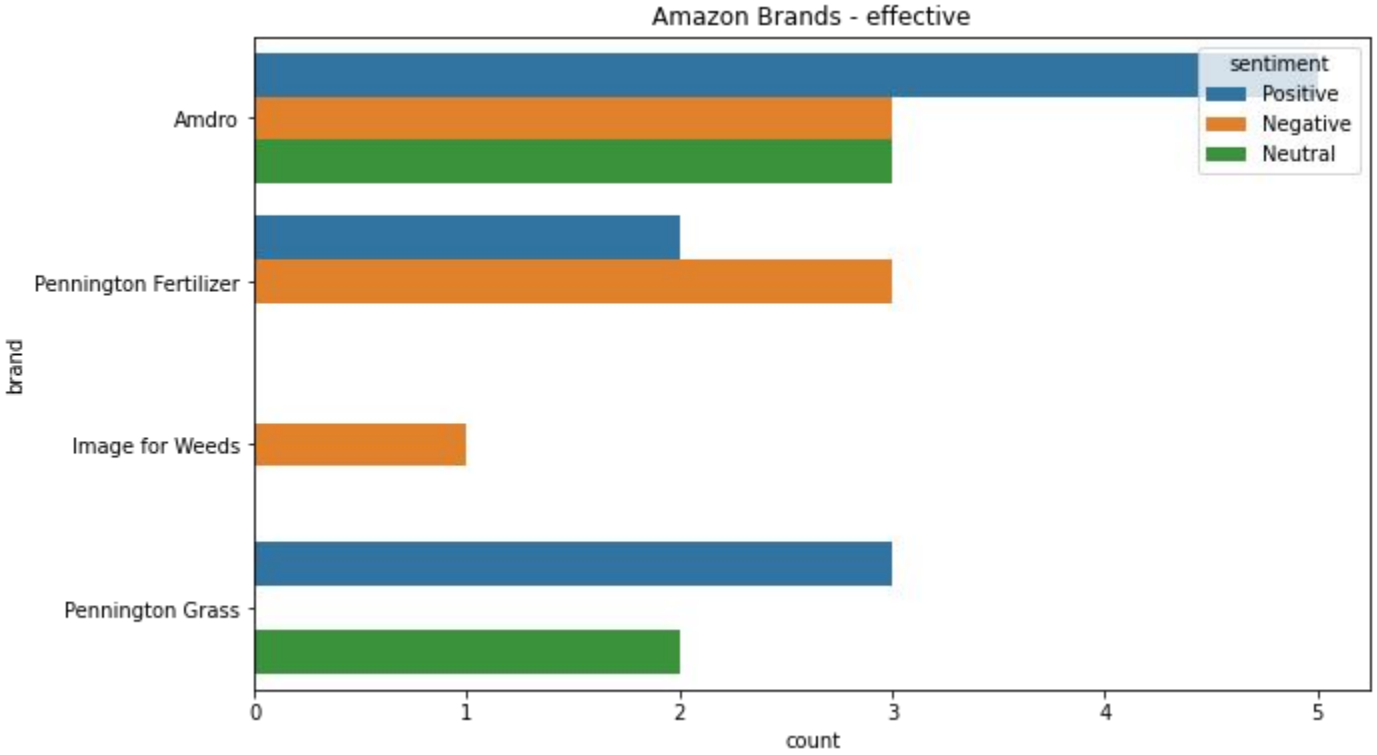
Sentiment Histogram of Amazon by Brands - Cost



Sentiment Histogram of Amazon by Brands - Efficient



Sentiment Histogram of Amazon by Brands - Effective



Sentiment Histogram of Amazon by Brands - Ease of Use