

Sentiment Analysis of Amazon review data in category of Pet Supplies and Grocery and Gourmet Food

Summary

- We are preparing to conduct a sentiment analysis on Amazon product reviews using two datasets: “Pet Supplies” and “Grocery and Gourmet Food.” The first dataset comprises a subset of 2,098,325 reviews, while the second dataset consists of 1,143,860 reviews. Both datasets encompass reviews collected over a time span from May 1996 to October 2018.
- We conduct a sentiment analysis at the word level, where we will exclude stopwords and generate word clouds to visualize the most commonly used words within each category. This approach will aid in comprehending the prevalent words associated with various categories, such as product, emotions, price, location, etc. By doing so, we aim to gain insights into the most popular terms within each category.
- We will use various lexicons, including “Bing,” “NRC,” and “AFINN.” Bing lexicon classifies words into positive and negative sentiments. We will find the top 10 most frequently used positive and negative words for each dataset and visualize the results.
- Using the NRC lexicon is helpful in identifying and categorizing the emotions expressed in text . We will visualize the percentage of each sentiment category for both datasets. Moreover, AFINN lexicon assigns a numeric score (sentiment score) to individual words based on their polarity or sentiment orientation.
- The reasons of using multiple sentiment lexicons, such as Bing, NRC, and AFINN are comprehensive sentiment analysis, handling ambiguity and reducing bias. Employing multiple lexicons can help mitigate such biases and provide a more balanced sentiment evaluation.

Overview and Results

Sentiment analysis at word level.

Conducting a basic sentiment analysis at the word level revealed the most frequently used words in each category. For grocery products, the top 10 commonly mentioned words include taste, flavor, love, coffee, product, tea, price, buy, time, and sugar. This suggests that taste is the primary factor consumers consider when it comes to grocery items, with coffee and tea being the most popular products mentioned. Price ranks 6th, indicating its influence on consumer buying decisions.

Similarly, for pet supplies, the top 10 frequently used words are dog, love, dogs, cat, food, product, cats, loves, time, and easy. This analysis indicates that the most popular pet supplies are related to dogs and cats.

These insights could provide valuable information for businesses to understand consumer preferences and tailor their strategies accordingly, emphasizing the importance of taste for grocery products and the popularity of dog and cat-related supplies in the pet industry.

Table 1: Pet Supplies

words	count	percentage
dog	717769	2.05
love	388019	1.11
dogs	384746	1.10
cat	355315	1.01
food	334474	0.95
product	319991	0.91
cats	291271	0.83
loves	253021	0.72
time	213732	0.61
easy	166424	0.48

Table 2: Grocery products

words	count	percentage
taste	240307	1.45
flavor	207661	1.25
love	202370	1.22
coffee	163258	0.99
product	162506	0.98
tea	161857	0.98
price	101662	0.61
buy	93492	0.56
delicious	85309	0.52
sugar	85450	0.52

```
## Found 500 records... Found 1000 records... Found 1500 records... Found 2000 records... Found 2500 r
## Found 500 records... Found 1000 records... Found 1500 records... Found 2000 records... Found 2500 r
```

Table 1: Top 10 Most Popular Sentiments Found in Each categories.



Figure 1 : As you can see, words like “taste,” “love,” “flavor”, “product” and “coffee” stand out since they were used more frequently in the grocery’s product review.



Figure 2: Dogs emerge as the most frequent words, followed by ‘love,’ ‘cat,’ ‘food,’ and ‘products,’ as evident from the data.

Sentiment Analysis using Bing Lexicon

Using the Bing lexicon, we categorize words into positive and negative sentiments for each dataset. This binary classification simplifies sentiment analysis by providing a straightforward approach to categorizing words based on their sentiment orientation. The analysis aims to identify the top 10 most frequently used positive and negative words in both datasets.

Table 2: shows the results of sentiment analysis using the Bing Lexicon for pet supplies and grocery products.

Table 3: Pet Supplies

words	sentiment	n
love	positive	388019
loves	positive	253021
easy	positive	166424
nice	positive	126215
recommend	positive	110392
perfect	positive	106452
clean	positive	91750
happy	positive	88573
tank	negative	85398
smell	negative	77814
top	positive	75660
pretty	positive	73528
hard	negative	71563
bad	negative	49232
issues	negative	42232
expensive	negative	41778
smells	negative	28467
break	negative	27730
cheap	negative	27652
issue	negative	27008

Table 4: Grocery products

words	sentiment	n
love	positive	202370
delicious	positive	85309
nice	positive	69098
sweet	positive	65635
fresh	positive	63194
favorite	positive	59424
free	positive	51991
perfect	positive	46520
excellent	positive	43306
hot	positive	42654
hard	negative	32950
bad	negative	31632
dark	negative	30432
expensive	negative	25067
fat	negative	20134
bitter	negative	19929
smell	negative	15613
cold	negative	15318
disappointed	negative	15240
lemon	negative	13452

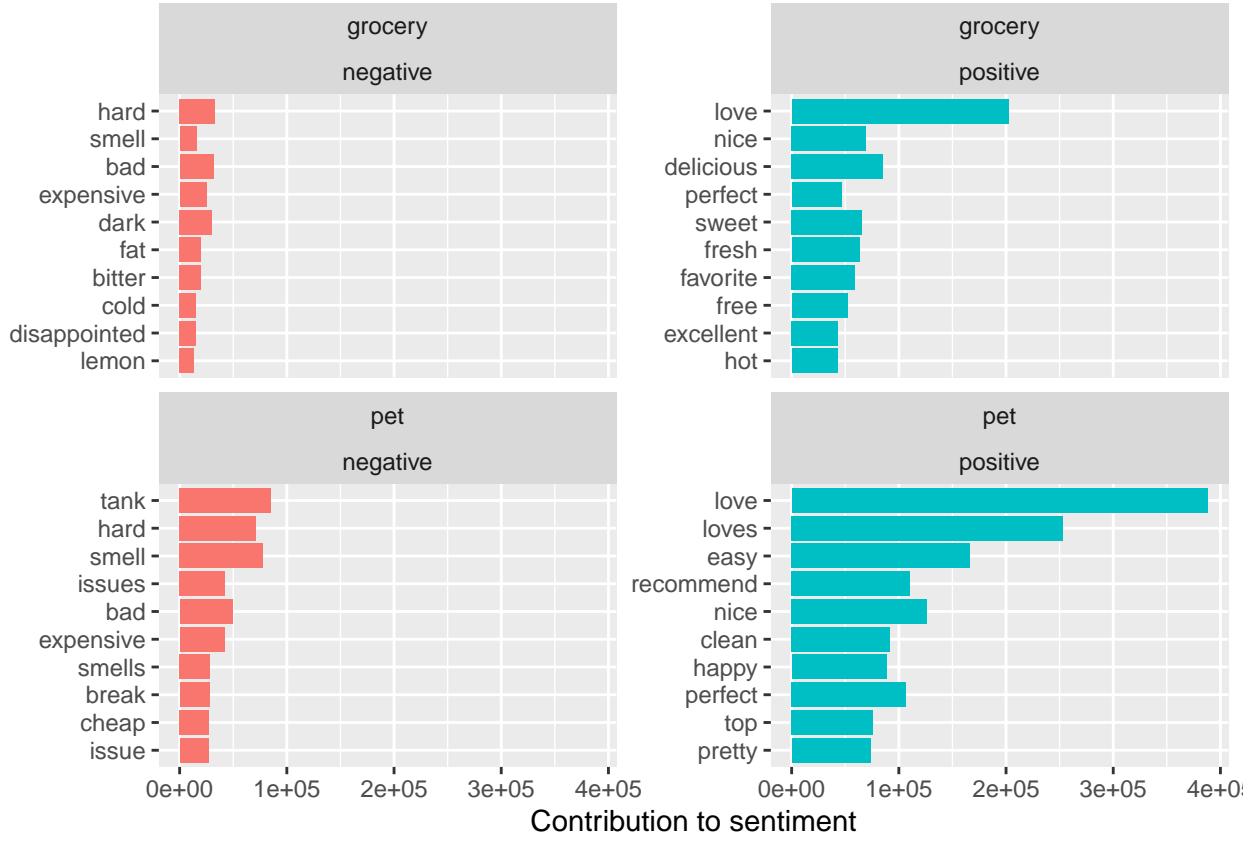


Figure 3: The bar chart illustrates the negative sentiments for pet supplies, where issues related to ‘tank,’ ‘smell,’ ‘hard,’ and ‘expensive’ were frequently mentioned. Similarly, for grocery products, negative feedback centered around attributes such as ‘hard,’ ‘bad,’ ‘dark,’ ‘expensive,’ ‘fat,’ and ‘bitter.’ Conversely, in the positive sentiment category for grocery products, customers expressed admiration for items described as ‘love,’ ‘delicious,’ ‘nice,’ ‘sweet,’ and ‘fresh.’

Sentiment Analysis using NRC Lexicon

Using the NRC lexicon is helpful in identifying and categorizing the emotions expressed in text. Leveraging this lexicon enables the identification and categorization of emotions expressed in customer feedback and reviews for these specific domains. By analyzing a wide range of emotions, including joy, sadness, anger, trust, and more, we can gain valuable insights into customers’ emotional responses towards various products in both categories.

Table 3: displays the outcomes of sentiment analysis conducted with the NRC Lexicon for both pet supplies and grocery products.

Table 5: Pet Supplies

sentiment	n	percentage	dataset
positive	4317075	24.26	pet
trust	2549797	14.33	pet
negative	2425315	13.63	pet
joy	2187678	12.29	pet
anticipation	1719661	9.66	pet
sadness	1051603	5.91	pet
anger	976992	5.49	pet
fear	977329	5.49	pet
disgust	862872	4.85	pet
surprise	724988	4.07	pet

Table 6: Grocery products

sentiment	n	percentage	dataset
positive	2384015	29.00	grocery
joy	1237268	15.05	grocery
trust	1129705	13.74	grocery
anticipation	887257	10.79	grocery
negative	779110	9.48	grocery
surprise	397518	4.84	grocery
sadness	385635	4.69	grocery
anger	362395	4.41	grocery
disgust	342869	4.17	grocery
fear	314851	3.83	grocery

NRC Sentiment Analysis Comparison between Grocery Products and Pet Supplies' review

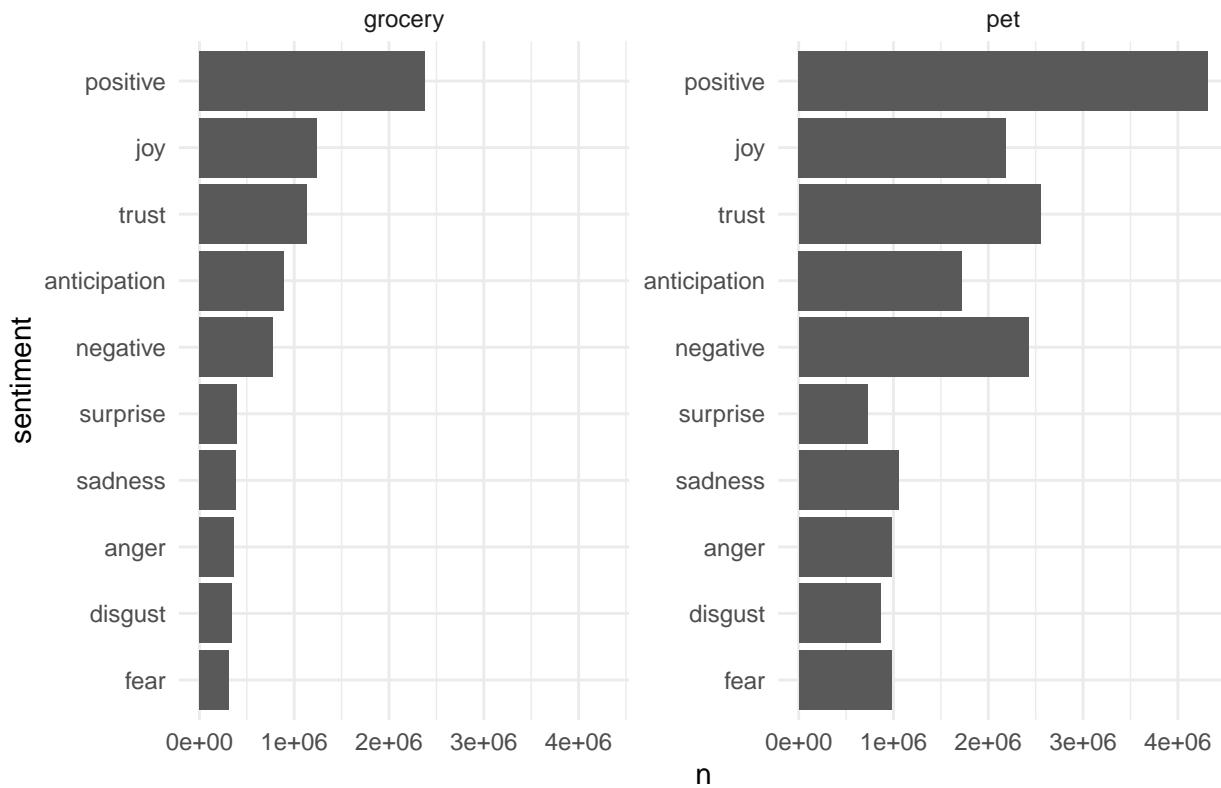


Figure 4: The bar chart illustrates a comparison of emotional responses towards various products in the NRC Sentiment Analysis for reviews of Grocery Products and Pet Supplies.

NRC Sentiment Analysis Comparison between Grocery and Pet product review

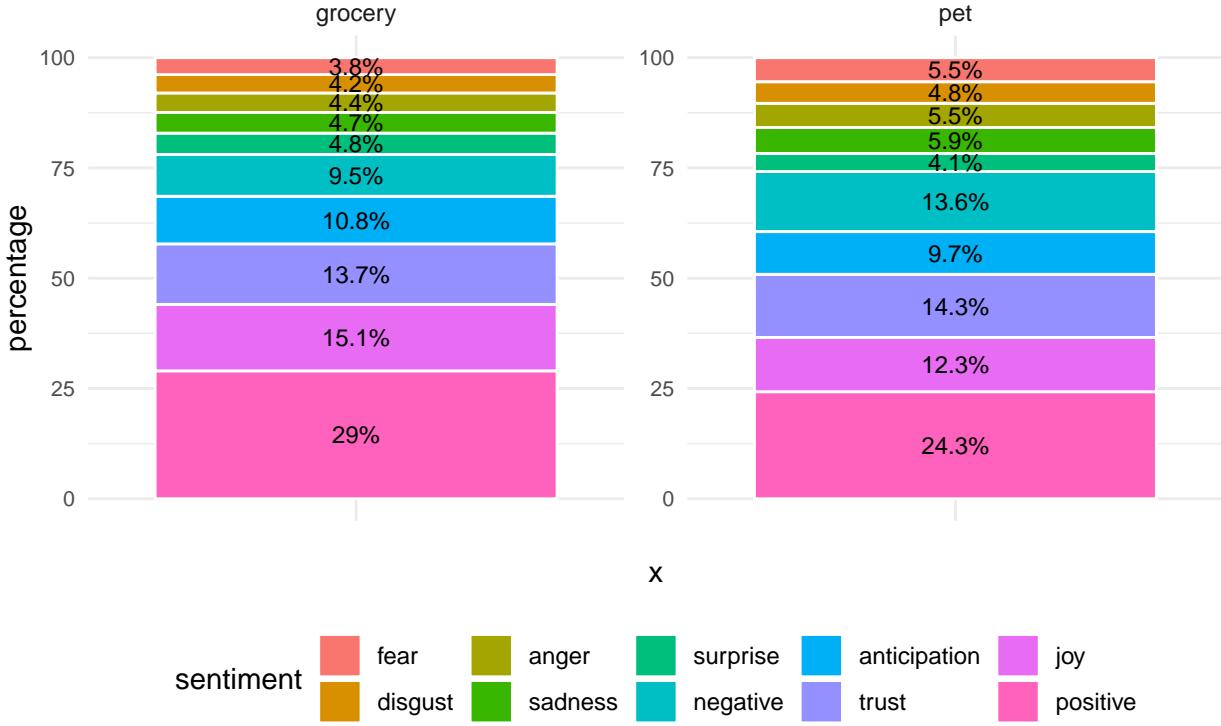


Figure 5: The stacked bar chart visually presents the proportions of the top 10 emotional expressions in each product category.

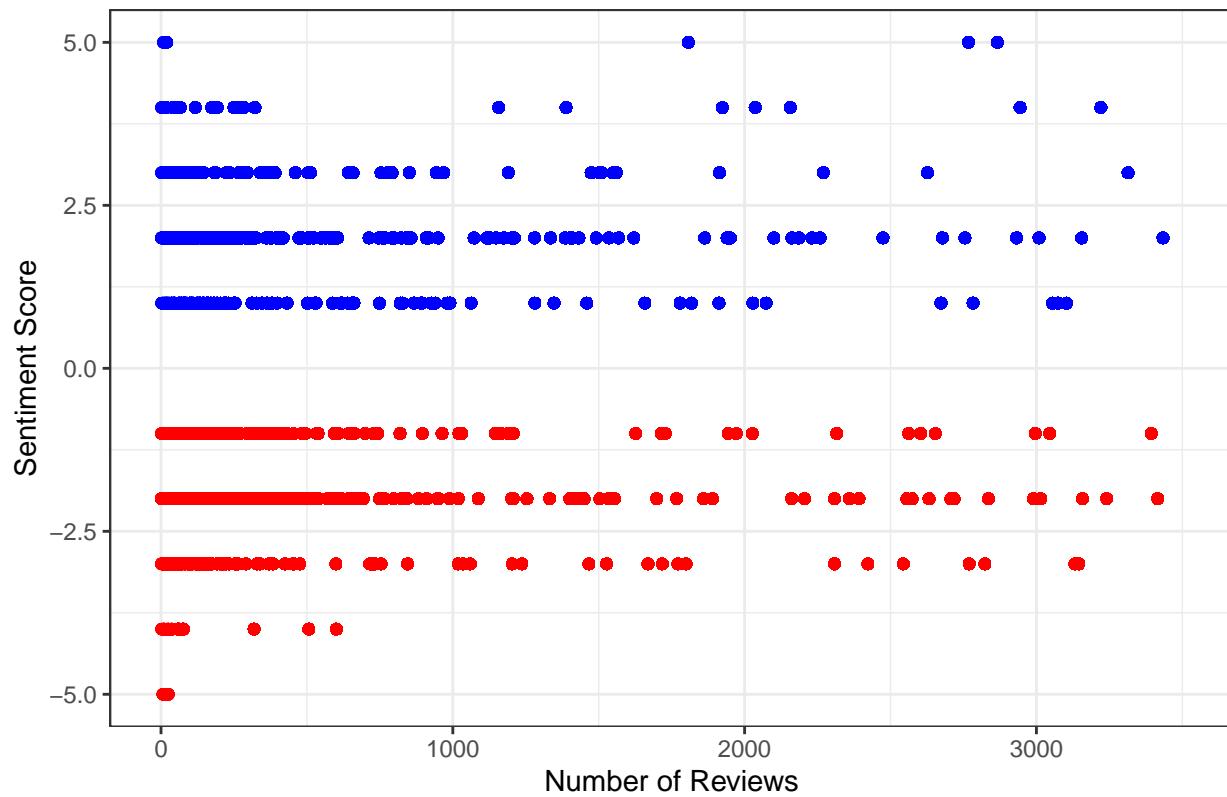
Based on the analysis using the NRC lexicon, it was observed that grocery products had a significant positive sentiment, accounting for 29.00% of the overall sentiment. Additionally, emotions of joy (15.05%), trust (13.74%), and anticipation (10.79%) were also prevalent in the context of grocery products. However, it is noteworthy that there was a noticeable negative sentiment of 9.47%.

On the other hand, pet supplies displayed a positive sentiment of 24.26%, along with trust (14.33%), joy (12.29%), and anticipation (9.66%). Surprisingly, pet supplies also exhibited a comparatively higher negative sentiment of 13.63%.

Sentiment Analysis using Afinn Lexicon

The AFINN lexicon utilizes a numeric score to represent the sentiment orientation of individual words, indicating their polarity. The simplicity, efficiency, and adaptability to various domains make AFINN a compelling choice for sentiment analysis. Although it may not encompass the full complexity of sentiment analysis as advanced techniques do, AFINN's user-friendliness and ability to yield valuable sentiment insights make it a valuable tool.

AFINN Sentiment Analysis Grocery and Gourmet products



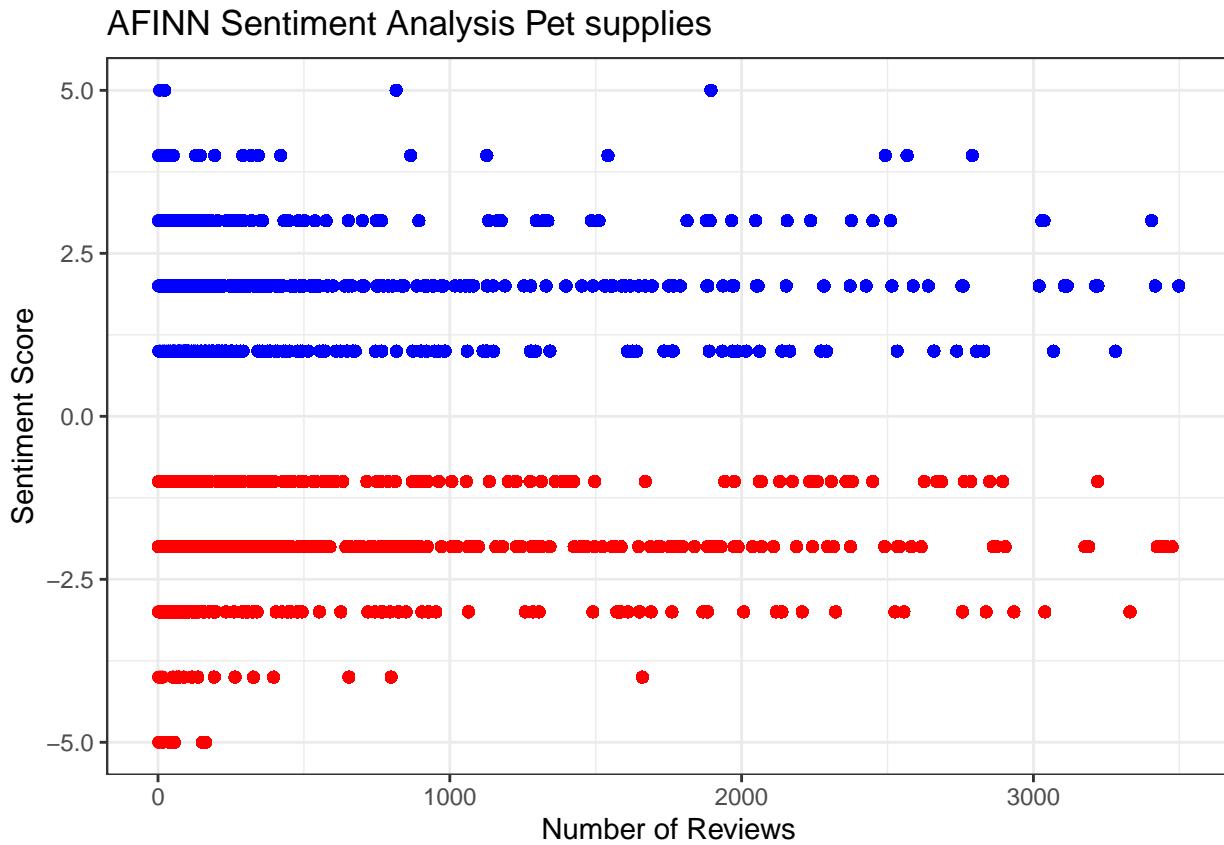


Figure 6: The AFINN lexicon analysis is visualized through a scatter plot, comparing the sentiment scores of the two datasets.

Conclusion and Suggestion

In this project, we performed sentiment analysis on product reviews for “Pet Supplies” and “Grocery and Gourmet Food” using different lexicons. The analysis provided insights into customer sentiments and the most frequently used words associated with each sentiment category.

These findings indicate that while both grocery and pet supplies evoke positive emotions, there is room for improvement in addressing the negative sentiments associated with pet supplies. It is advisable to conduct further analysis to identify the specific aspects contributing to the negative sentiment, which can subsequently lead to enhancing product offerings and service quality to better meet customer expectations.