Analyze Amazon Store Products Reviews Data

Created by: Ahmed Miri
Last Update: August 24th, 2022

General Information about the Dataset

- We have 10,261 rows with 9 features.

- There are 900 different products in the dataset.

- The number of reviews given by a certain customer ranges between 5 and 42 reviews.

- There are 1,429 reviewers writing 10,261 reviews.

- There is no reviewer who reviews a product multiple times.

- The average amount of reviews is 11 per product.

- The highest amount of reviews is 163 per product.

- The lowest amount of reviews is 5 per product.

The Words Statistics of the Review Text

- The average word count of one review is 90 words.
- The minimum count is 1 word.
- The maximum count used is 2,043 words.

count	10261.00
mean	90.19
std	111.23
min	1.00
25%	31.00
50%	54.00
75%	104.00
max	2043.00

The Words Statistics of the Summary

- The average word count of one summary is 4 words.
- The minimum count is 1 word.
- The maximum count used is 25 words.

count	10261.00
mean	4.37
std	2.85
min	1.00
25%	2.00
50%	4.00
75%	6.00
max	25.00

Overall Score Distribution


Distribution of Overall in the Store
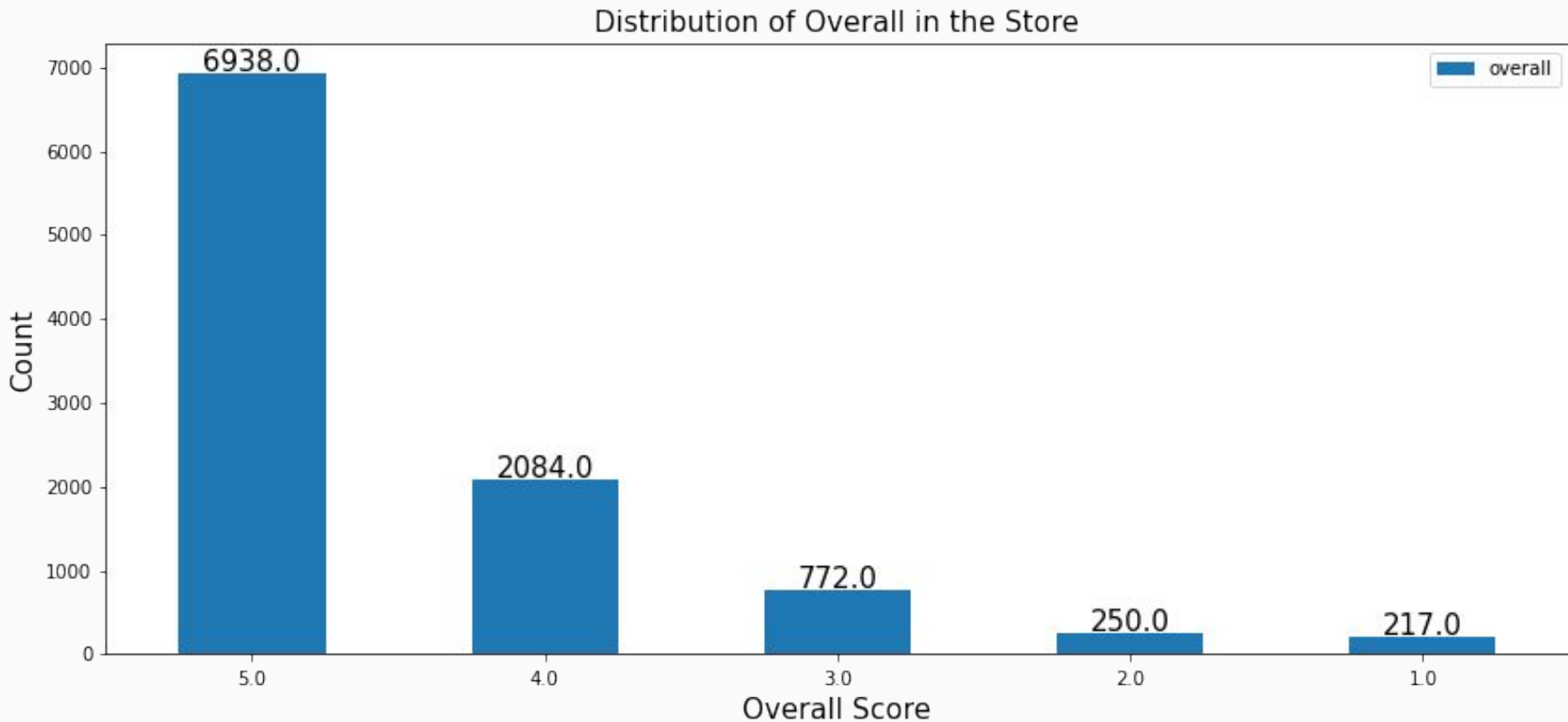
Overall Score Percentage

- We have 6,938 reviews with 5 overall score, that equal to 67.6% of the total reviews.
- We have 2,084 reviews with 4 overall score, that equal to 20.3% of the total reviews.
- We have 772 reviews with 3 overall score, that equal to 7.5% of the total reviews.
- Also, we have 250 reviews with 2 overall score, that equal to 2.4% of the total reviews.
- And we have 217 reviews with 1 overall score, that equal to 2.1% of the total reviews.
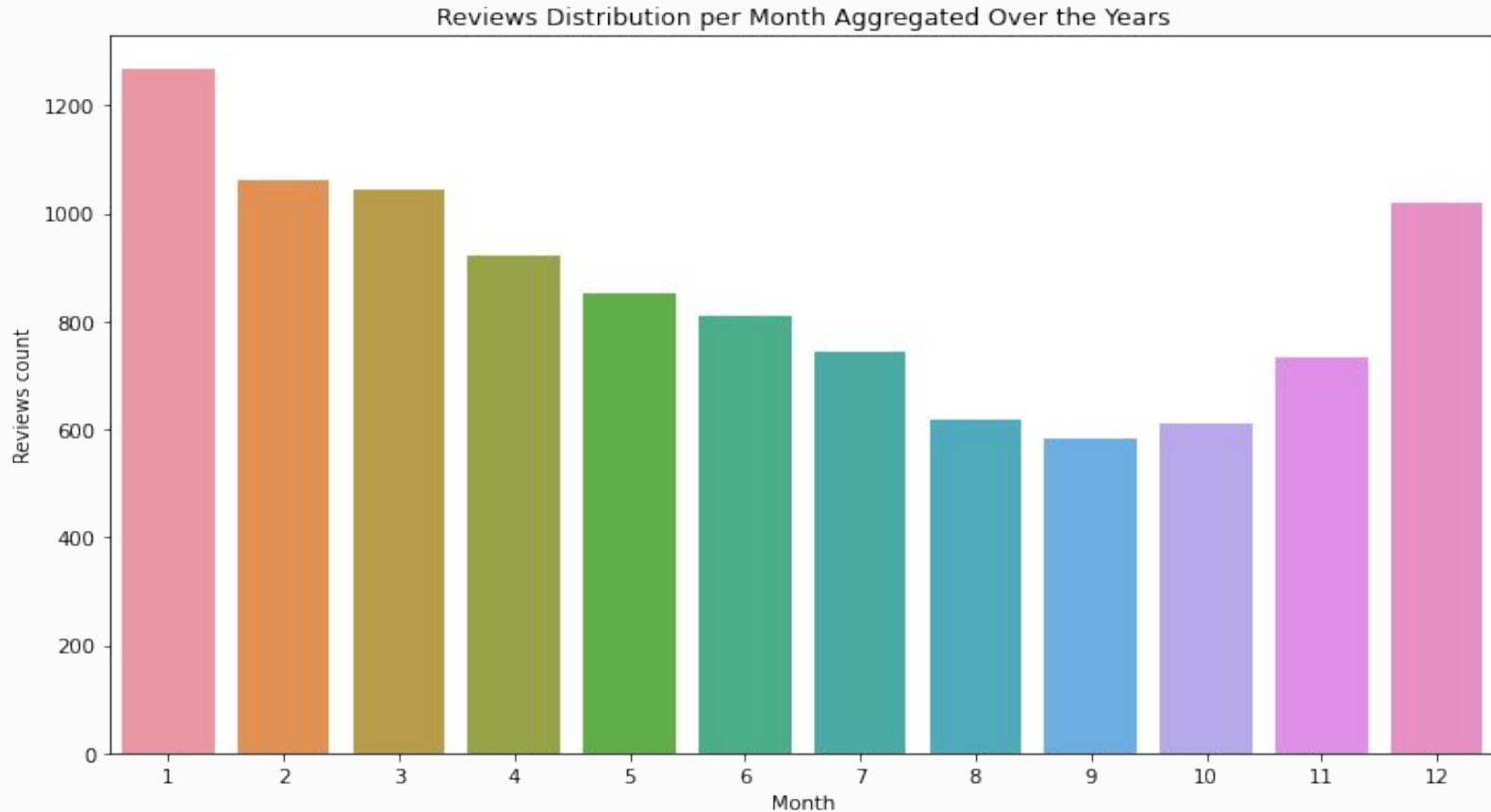

Percentage of Overall in the Store

20 Most Reviewed Products
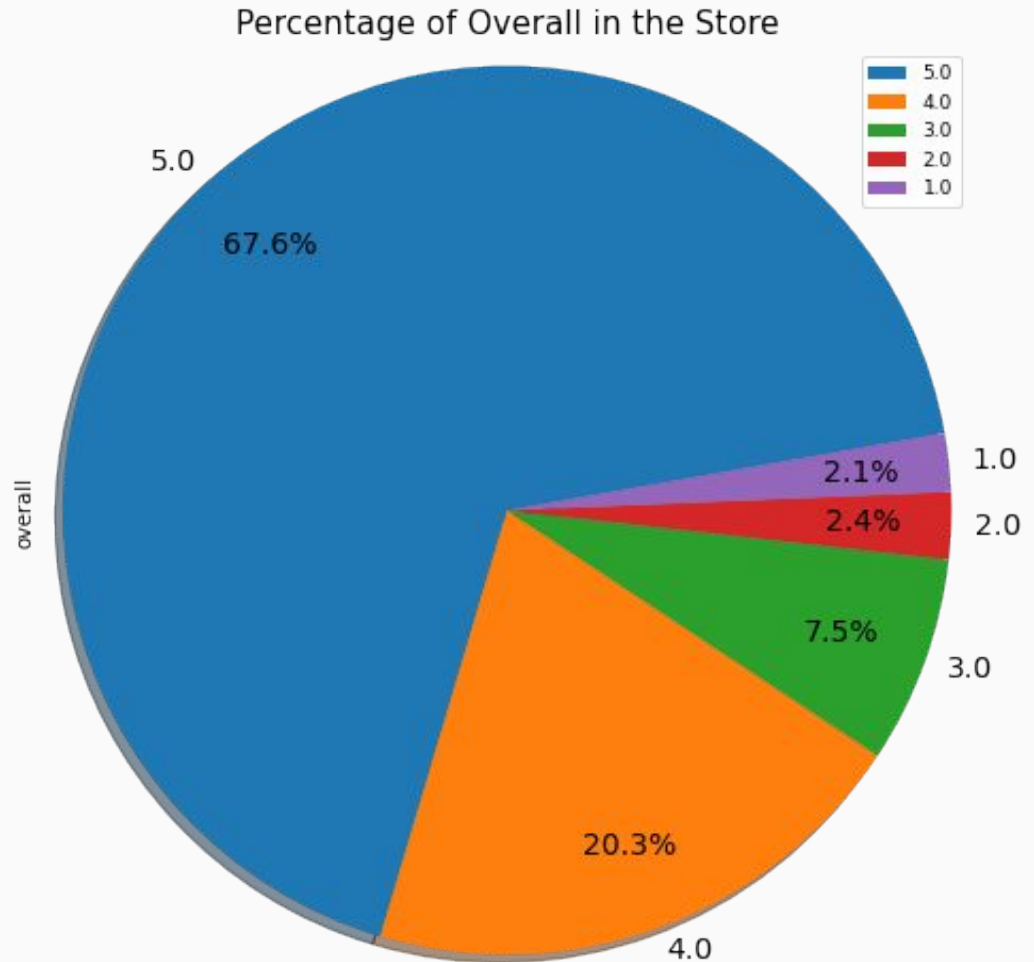
20 Most Reviewed Products

Reviews Count Distribution per Year
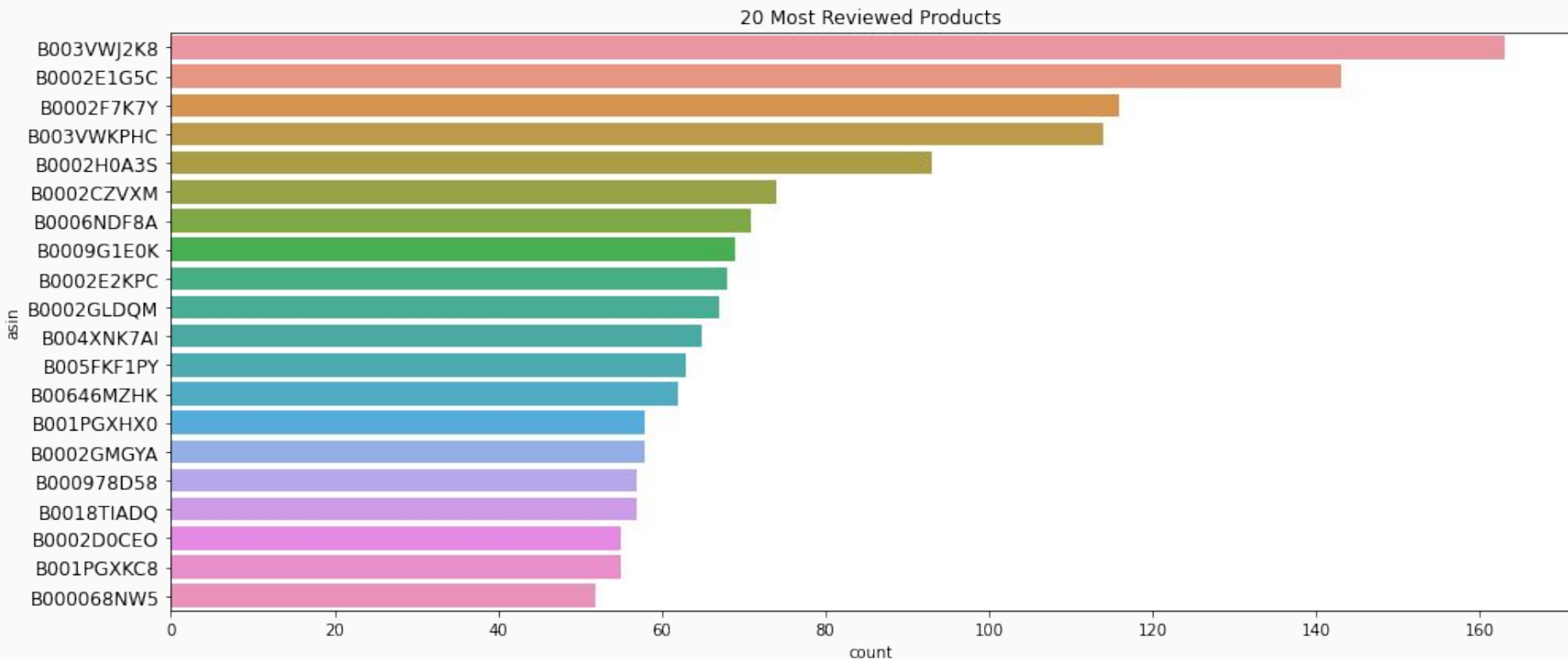
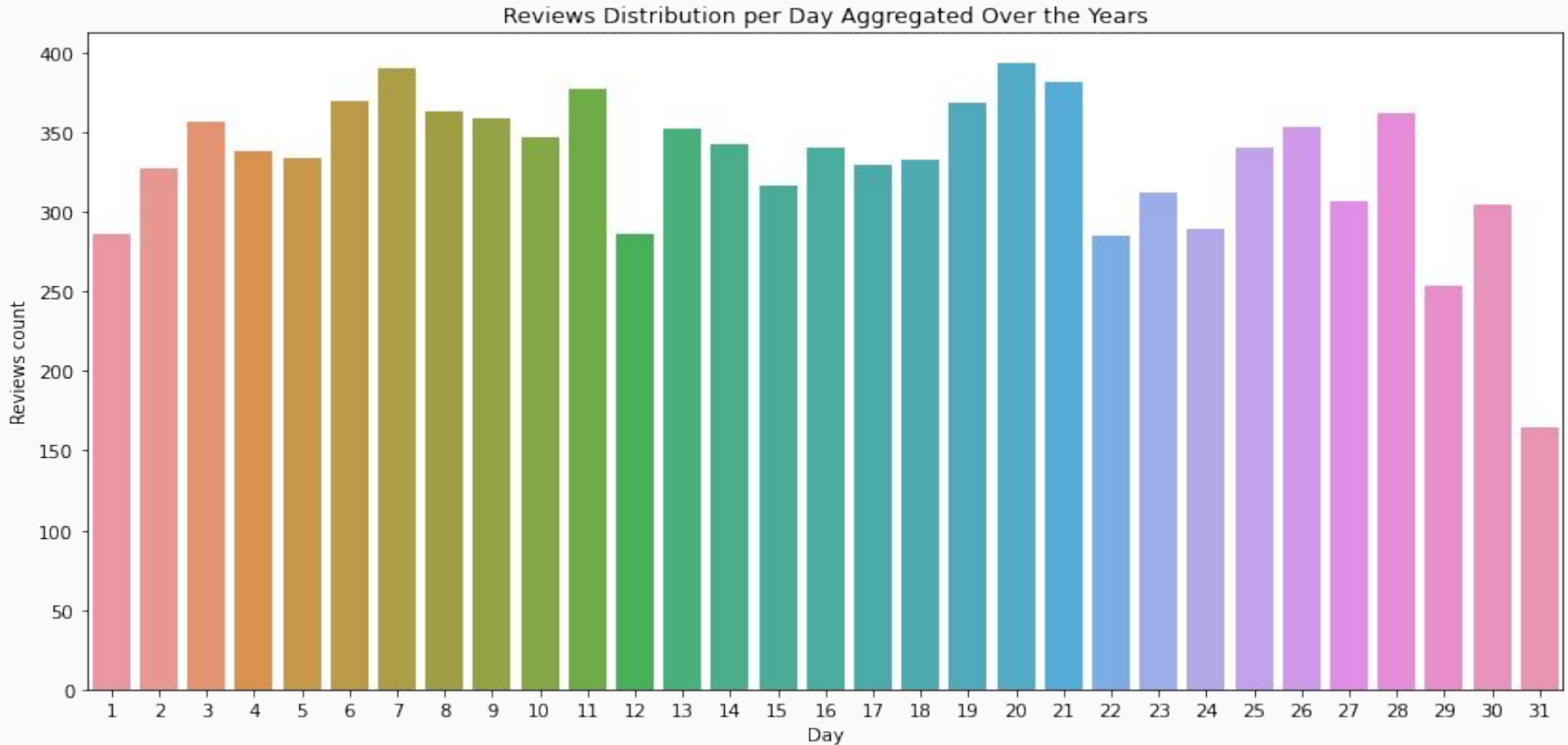Reviews Distribution Per Year

Reviews Count Distribution per Month Aggregated Over the Years
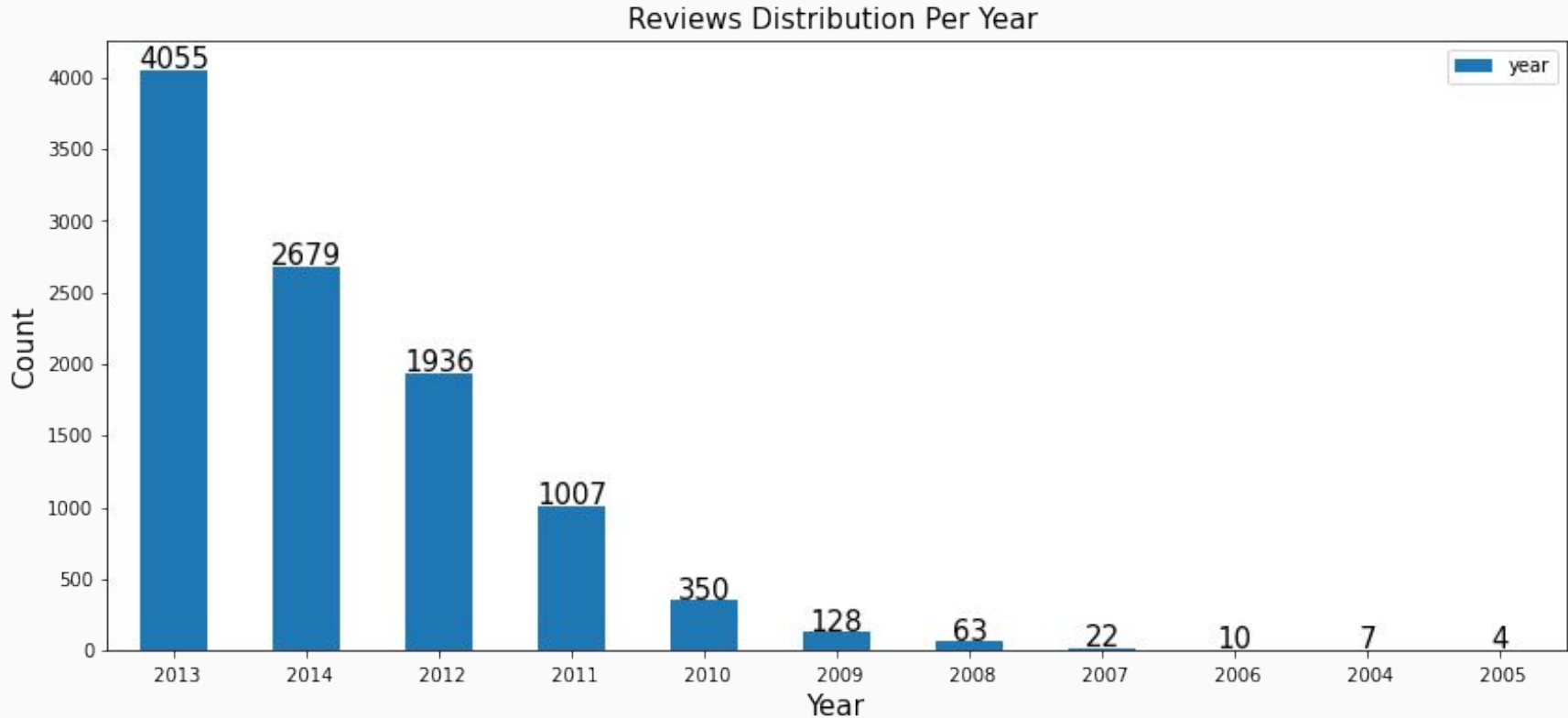
Reviews Distribution per Month Aggregated Over the Years

Reviews Count Distribution per Day Aggregated Over the Years



Reviews Distribution per Day Aggregated Over the Years

The Helpful Rate

The helpful feature in the dataset has values in list [x,y] format, which means that out of y people x people found that review helpful. But with this format, it will be difficult to add value to the dataset. A helpful_rate feature is created which returns the value of X divided by Y from [x,y].

- ❏ The helpful rate gives a value between 0.00 and 1.00.
- ❏ 0.00 refers to review is not helpful at all.
- ❏ 1.00 refers to review is very helpful.

Reviews Helpfulness Rate

- We can see that 73.9% of the reviews are not helpful at all.
- Only 20.9% of the reviews are very helpful.
- And the other 5.2% of the reviews are between mid and barely helpful.

Reviews Helpfulness Rate

0.0 73.9%

1.1% 0.75
1.4% 0.67
2.7% 0.5

20.9% 1.0

helpful_rate

Legend: 0.0, 1.0, 0.5, 0.67, 0.75
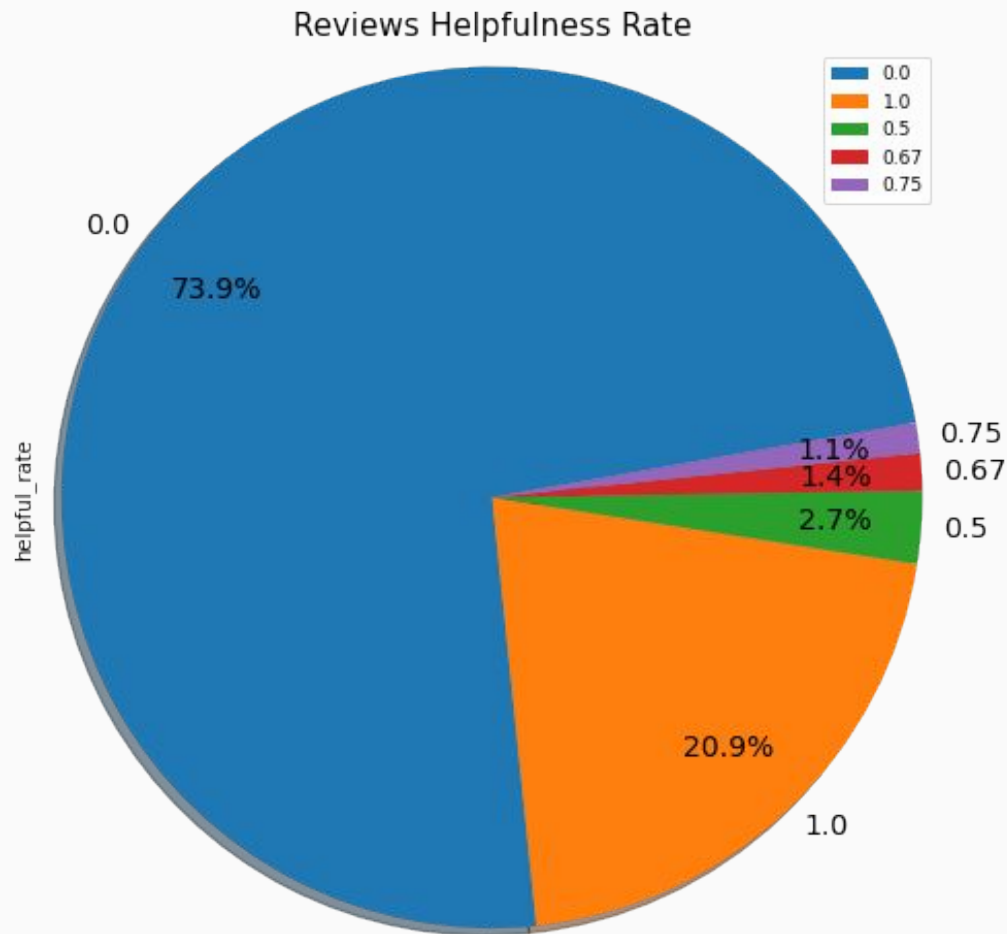
Percentage of Helpful Reviews By Overall Rating

- 67.2% of the helpful reviews are from 5 overall score reviews.
- 22.6% of the helpful reviews are from 4 overall score reviews.
- 6.7% of the helpful reviews are from 3 overall score reviews.
- Also, 1.7% of the helpful reviews are from 2 overall score reviews.
- And 1.7% of the helpful reviews are from 1 overall score reviews.

Percentage of Helpful Reviews By Overall Rating

Top 10 Products with the Highest Helpful Reviews
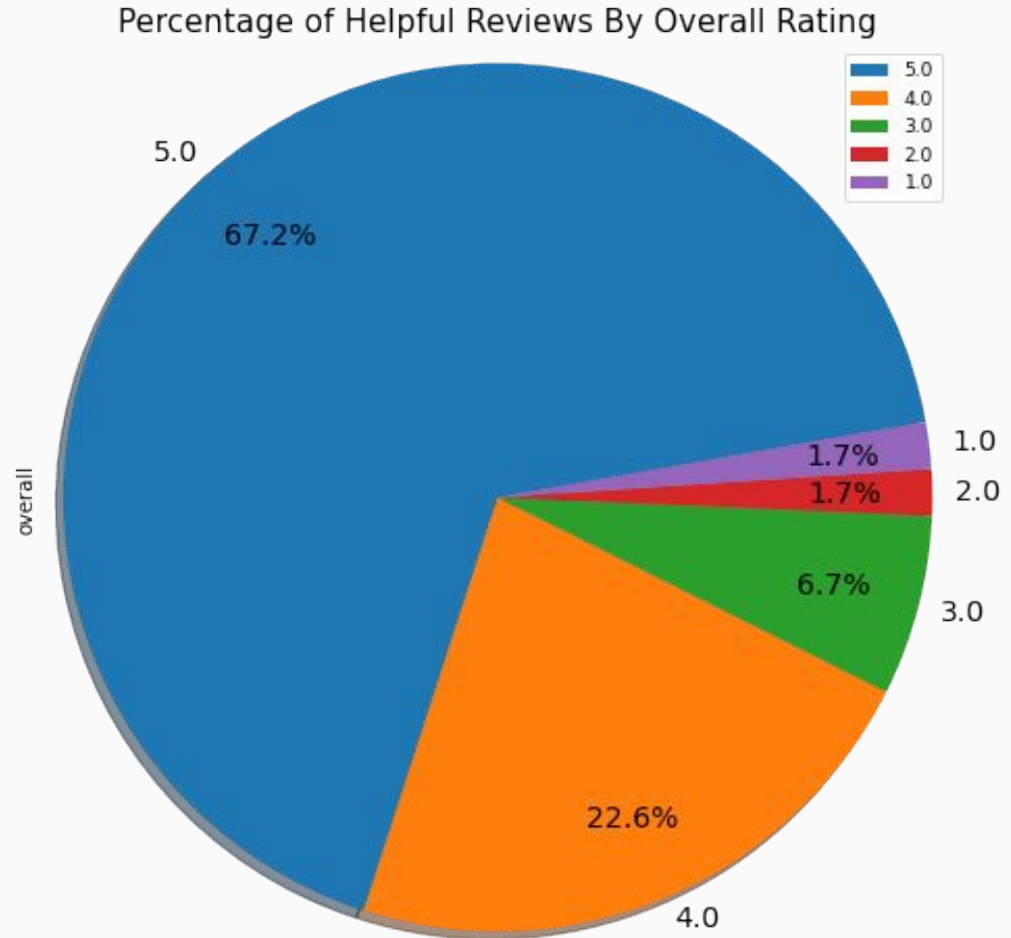
Top 10 Products with the Highest Helpful Reviews
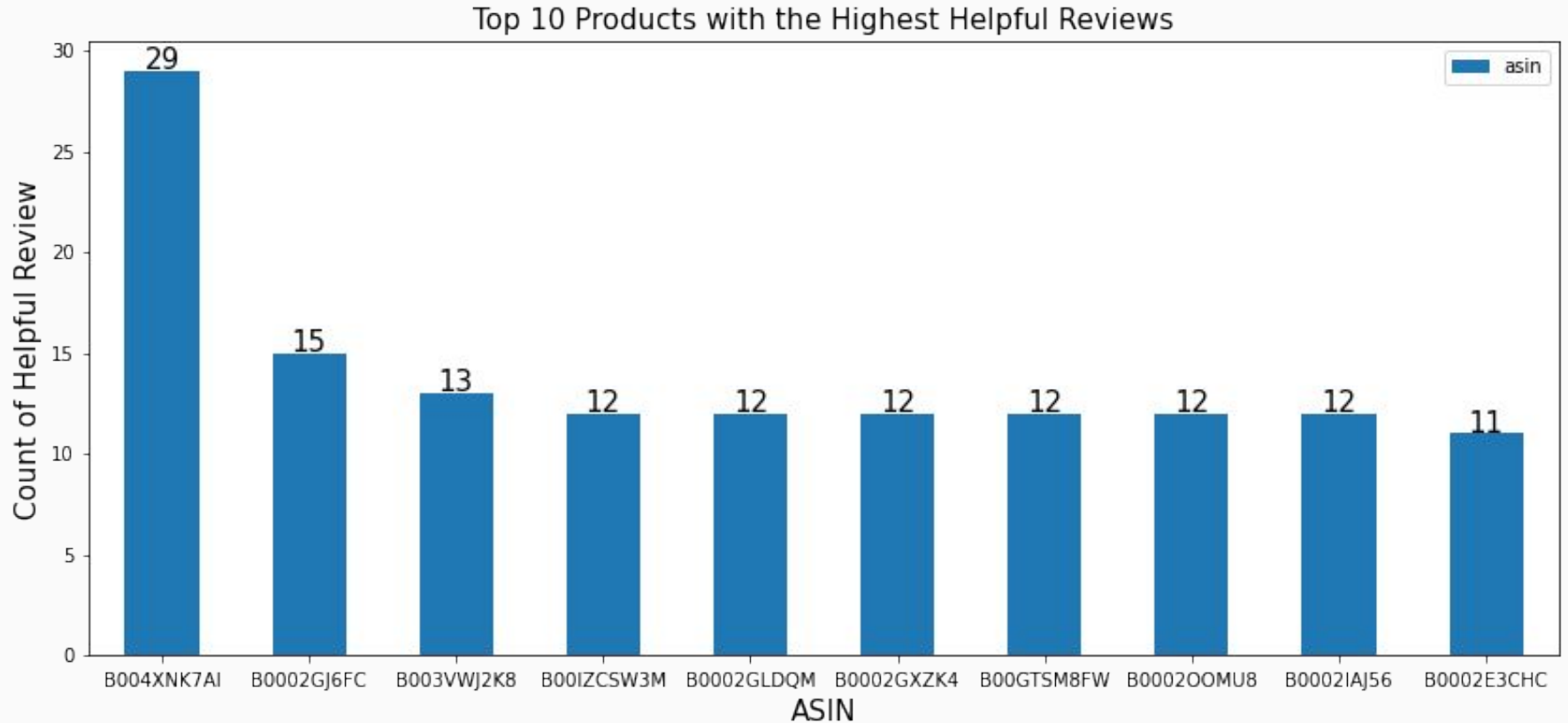
The Nature of the Sentiment

The sentiments of the review are decided on the overall score.

- ❏ If the overall score is greater than 3, the sentiment is Positive.
- ❏ If the overall value is less than 3, the sentiment is Negative.
- ❏ And if it is equal to 3, the sentiment is Neutral.

The Count of Sentiments

| Sentiment	Count
Positive	9022
Neutral	772
Negative	467

Percentage of Sentiments

- 87.9% of reviews are Positive sentiments.
- 7.5% of reviews are Neutral sentiments.
- 4.6% of reviews are Negative sentiments.

We can see that customers are happy with the store.

Sentiments of Reviews as a Percentage

Distribution of Sentiment over the Years
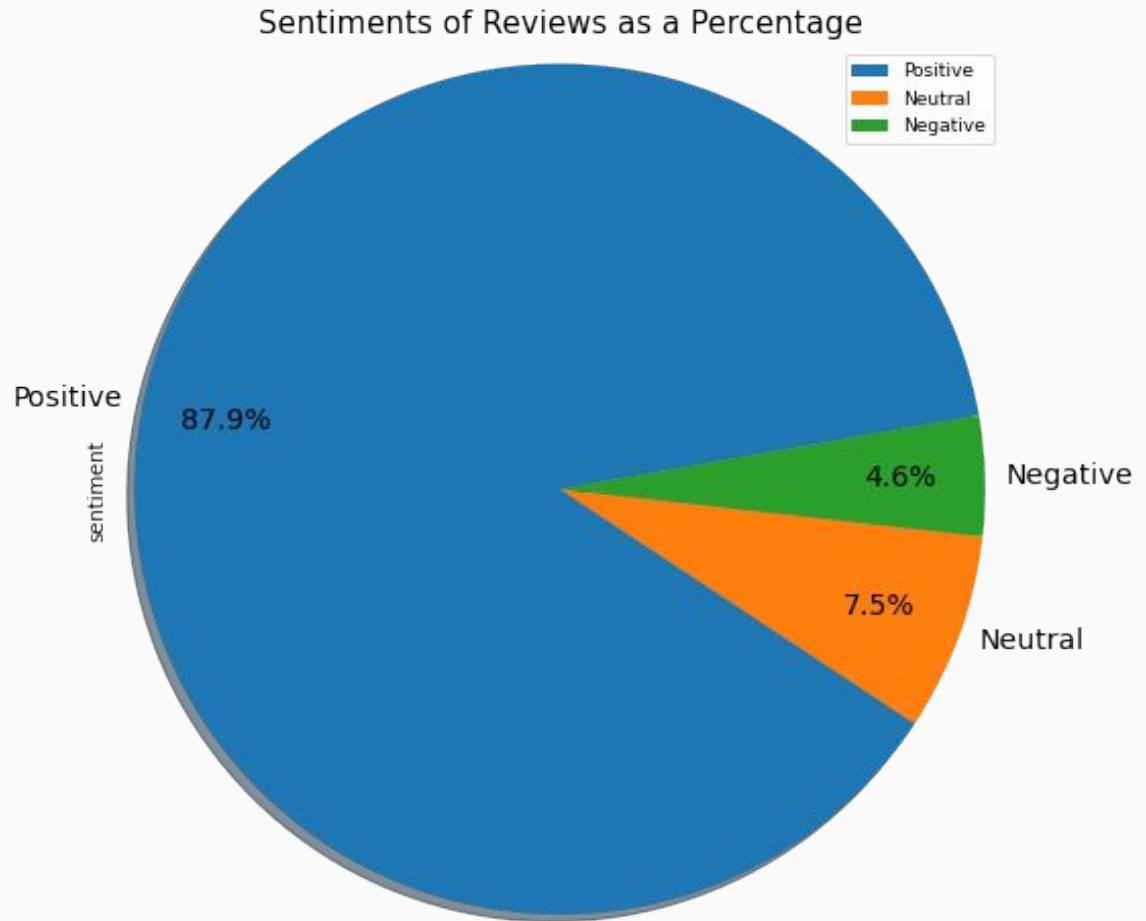
Distribution of Sentiment over the Years
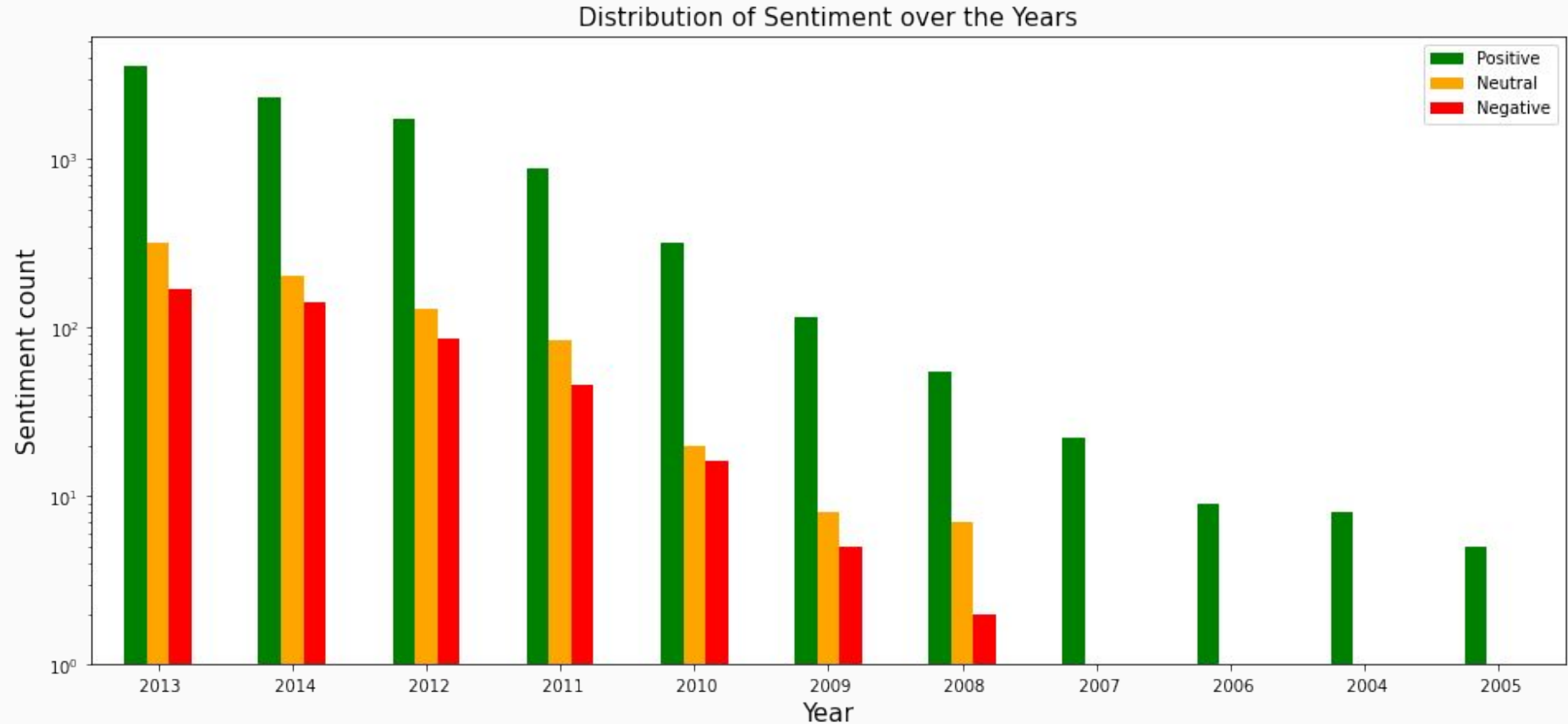
Line Distribution of Sentiment over the Years

From this plot and previous plot, we can clearly see the rise in positive reviews from 2010. Reached its peak around 2013 and there is a dip in 2014, All the review rates dropped at this time. Negative and neutral reviews are very low as compared to positive reviews.
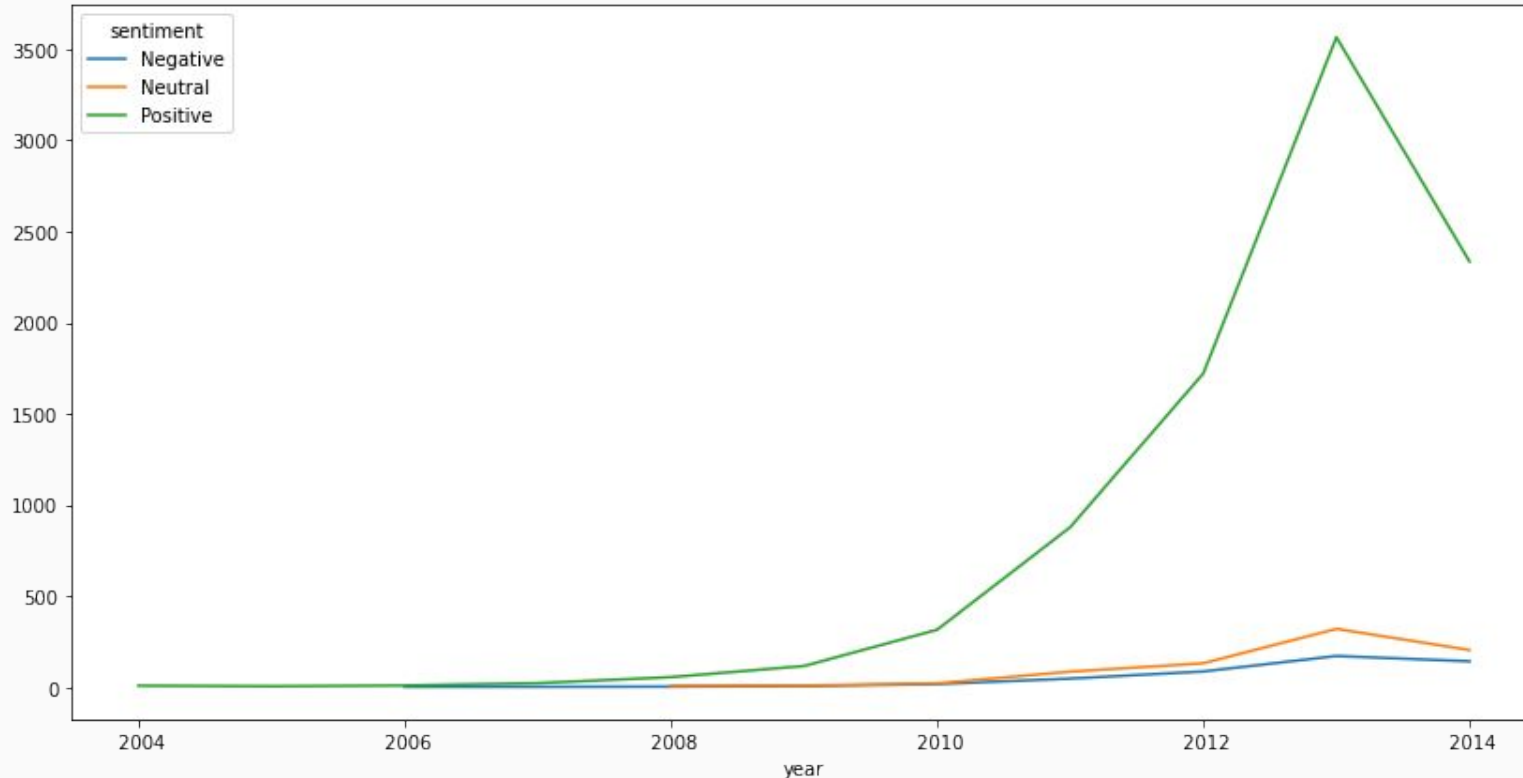
Top 10 Products with Positive Sentiment

Top 10 Products with Positive Sentiment

Top 10 Products with Neutral Sentiment

Top 10 Products with Neutral Sentiment
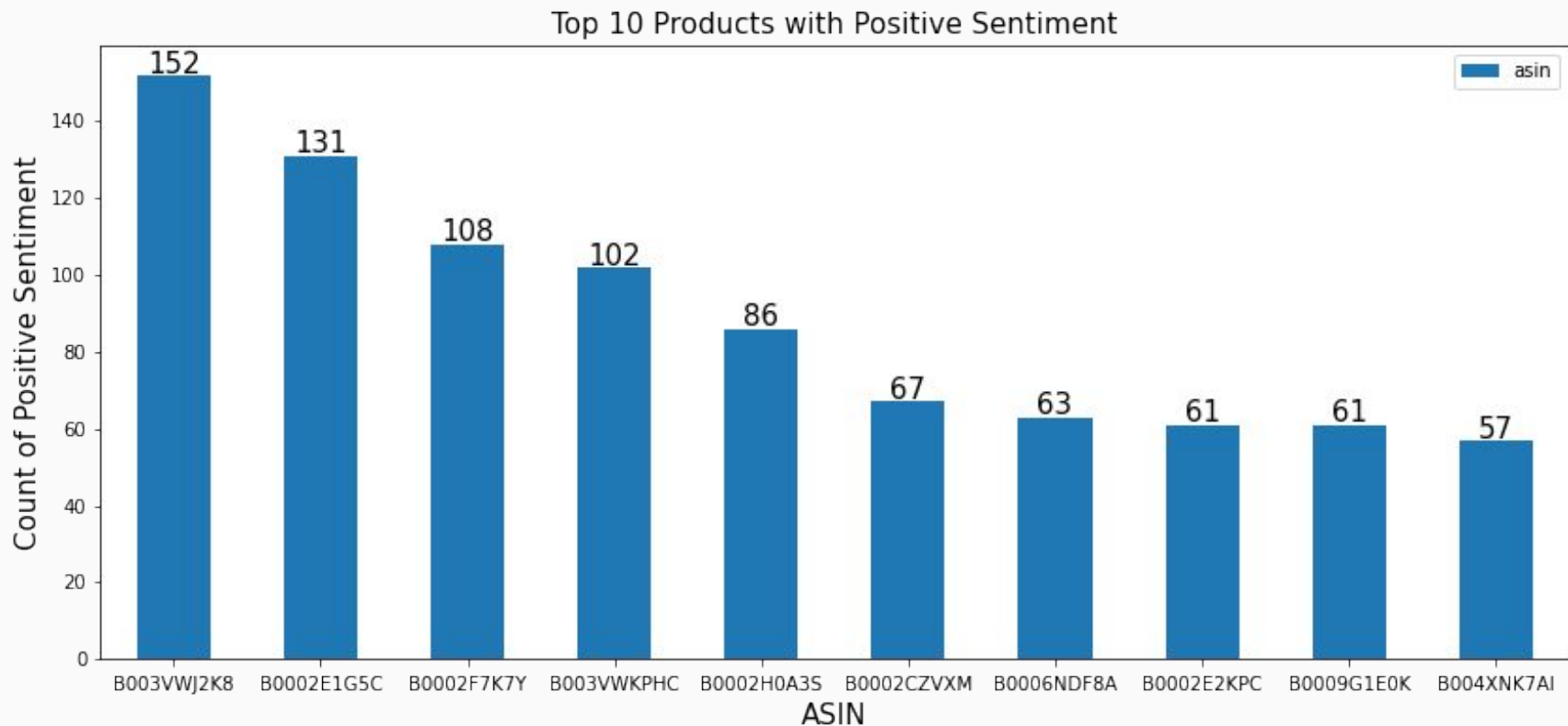
Top 10 Products with Negative Sentiment
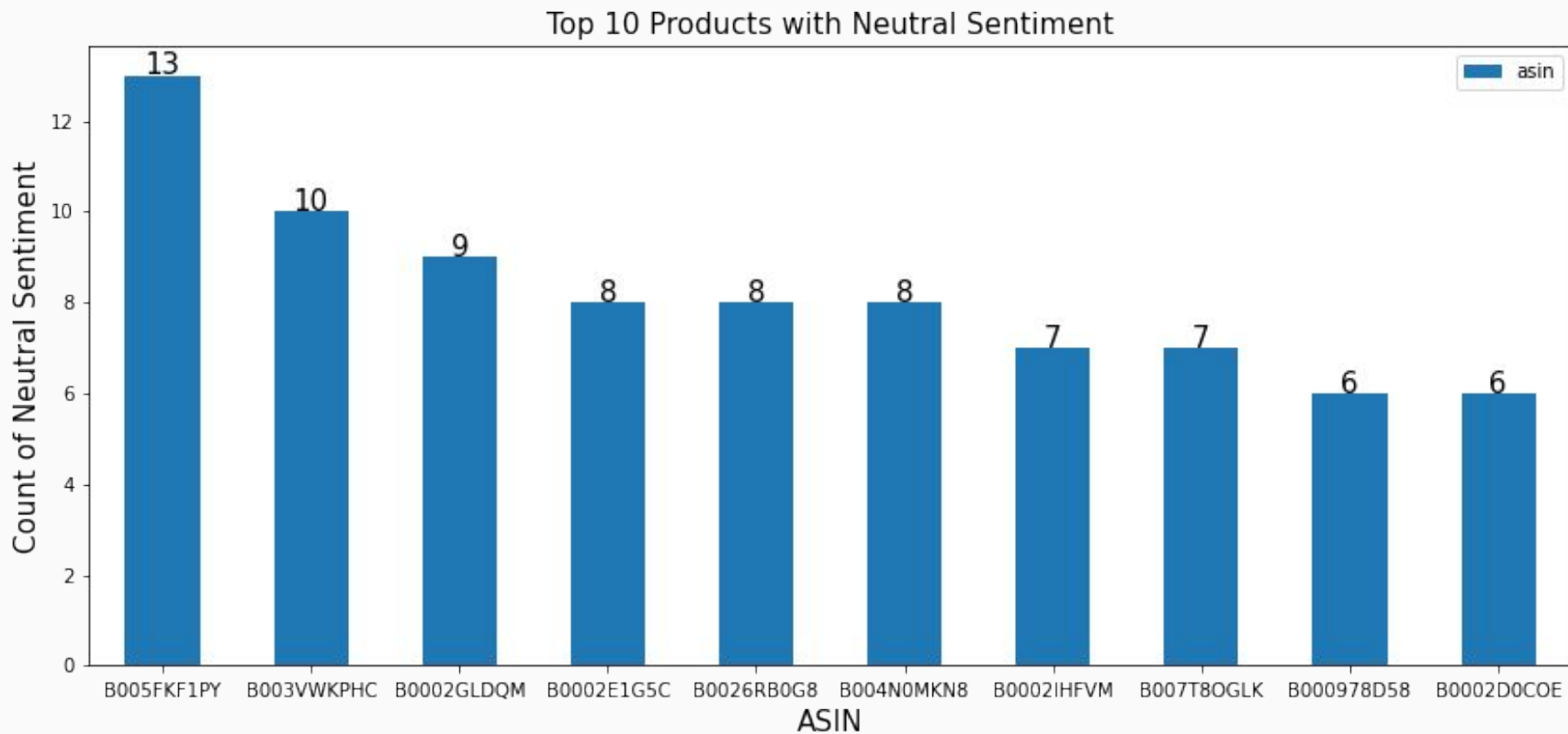
Top 10 Products with Negative Sentiment
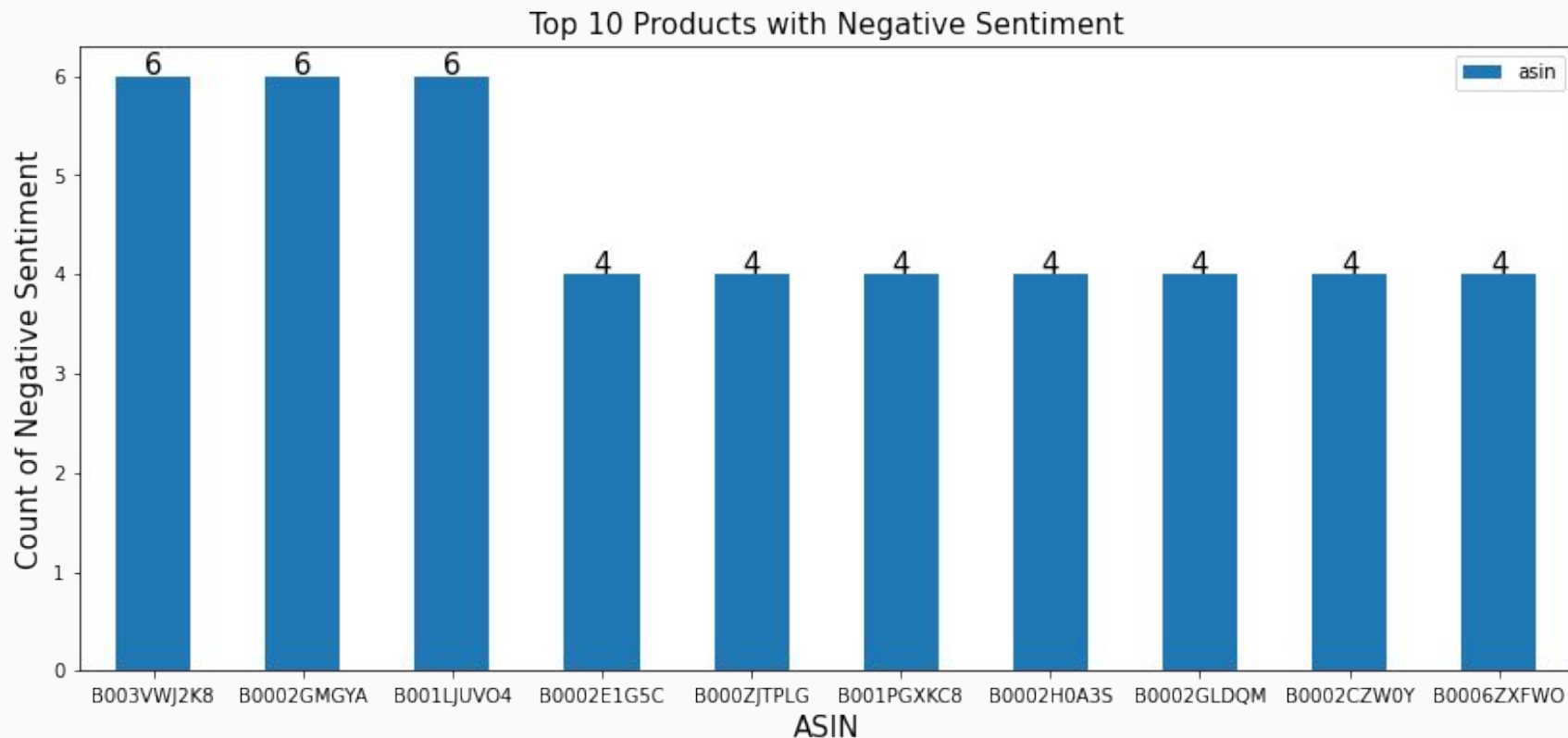
Correlation of Sentiment and Helpful Rate

From the table, we can see that the mean of helpful rate is higher for negative reviews than for neutral and positive reviews. These mean values might have been influenced by the 0 values in helpful rates.

Sentiment	Mean of Helpful Rate
Negative	0.307
Neutral	0.275
Positive	0.260

N-Gram Analysis

Here we will be using N-Grams to analyze the text based on its sentiment for both review text and summary text.

❏   Unigram Analysis.

❏   Bigram Analysis.

❏   Trigram Analysis

Reviews Unigram Analysis

The most frequent one word in reviews based on sentiments.

Unigram Frequency of Positive Reviews

Frequent words of positive reviews

Unigram Frequency of Neutral Reviews
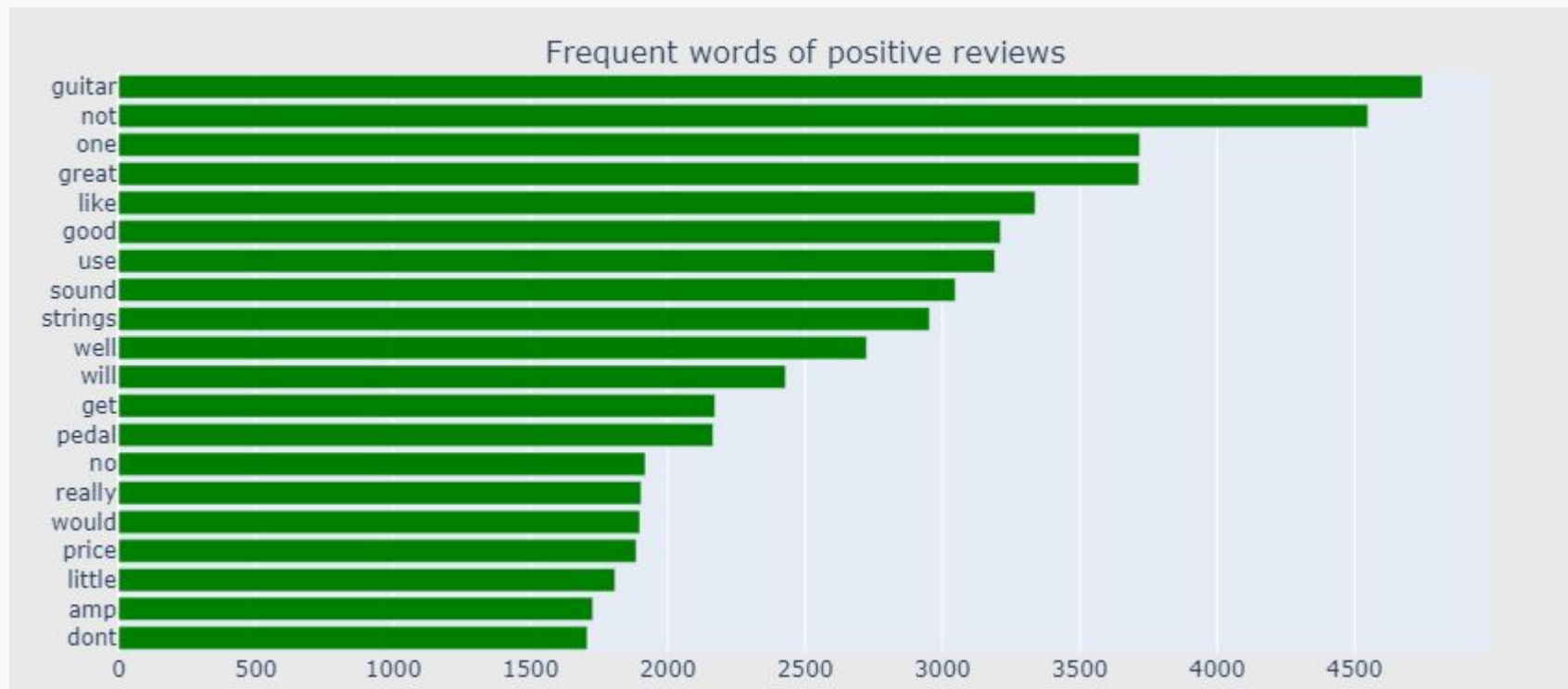
Frequent words of neutral reviews

Unigram Frequency of Negative Reviews

Frequent words of negative reviews
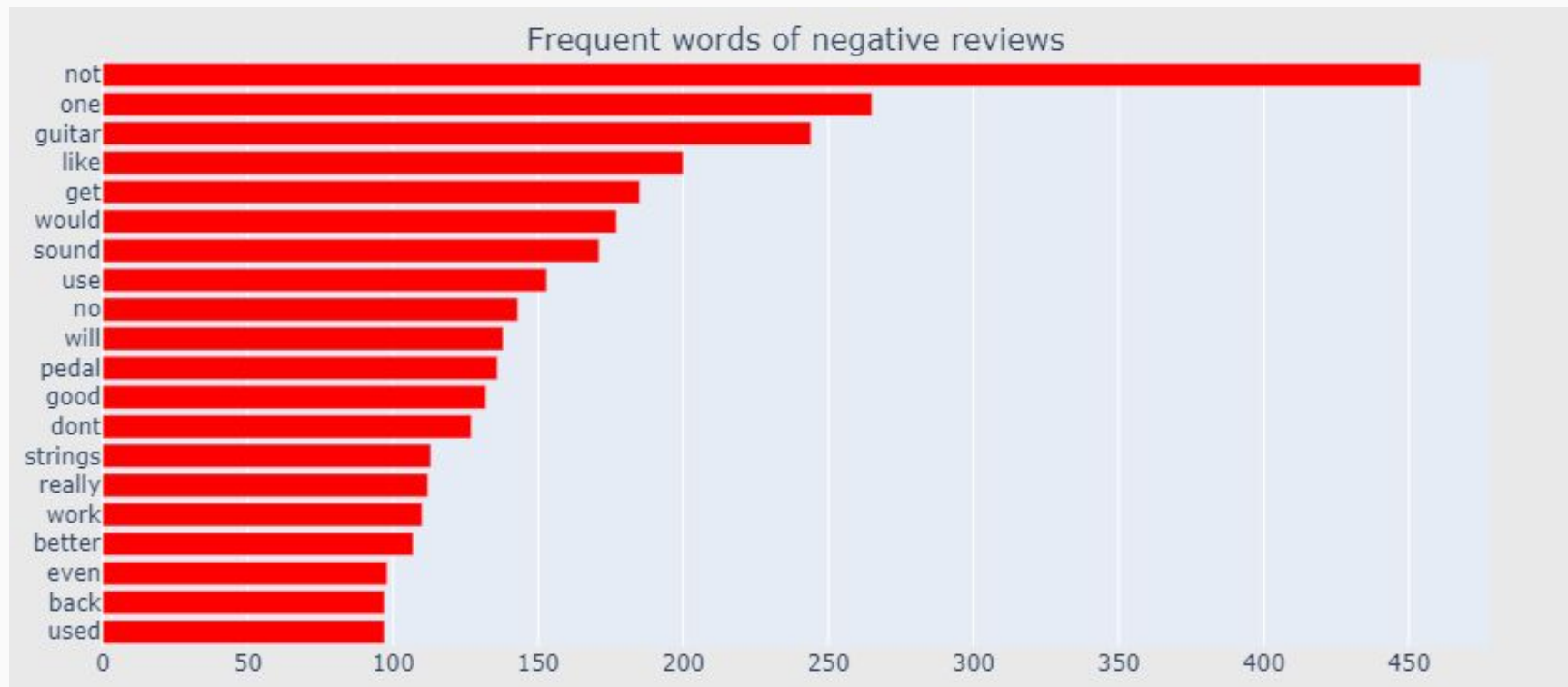
Summaries Unigram Analysis

The most frequent one word in Summaries based on sentiments.

Unigram Frequency of Positive Summaries

Frequent words of positive Summaries

Unigram Frequency of Neutral Summaries

Frequent words of neutral Summaries

Unigram Frequency of Negative Summaries
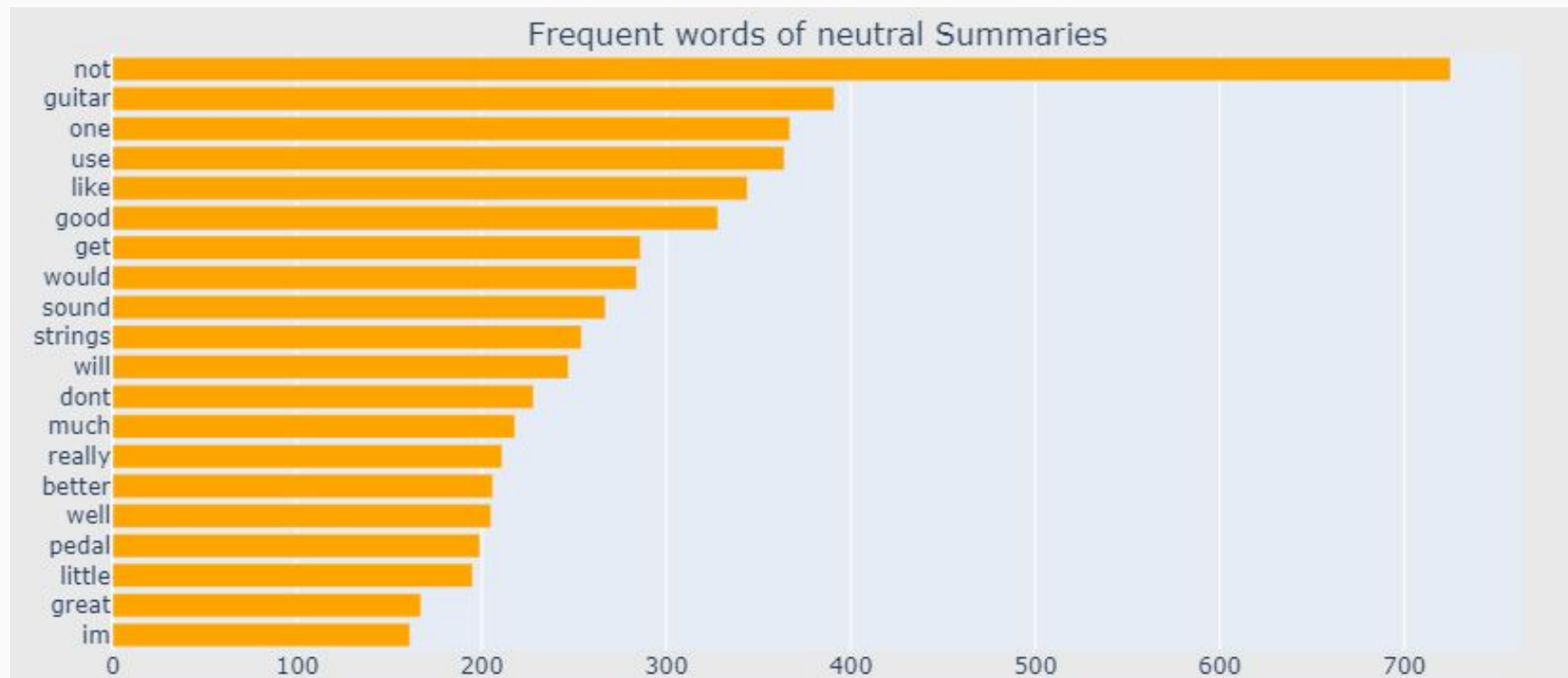
Frequent words of negative Summaries
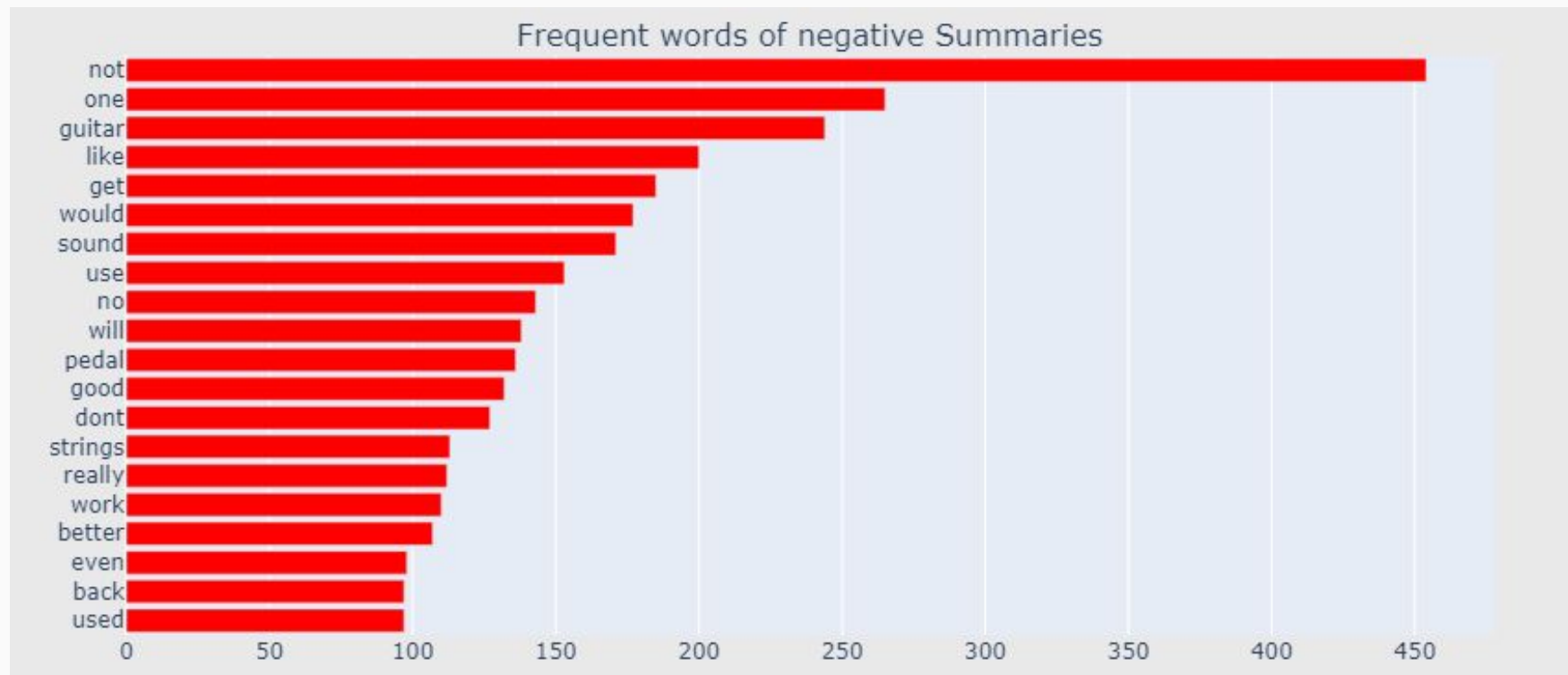
Unigram Conclusion

Unigram is not sufficient thus, we need to obtain the bigram and trigram to better understand the relationship between the sentiment on a side and the summary and review words on the other side.

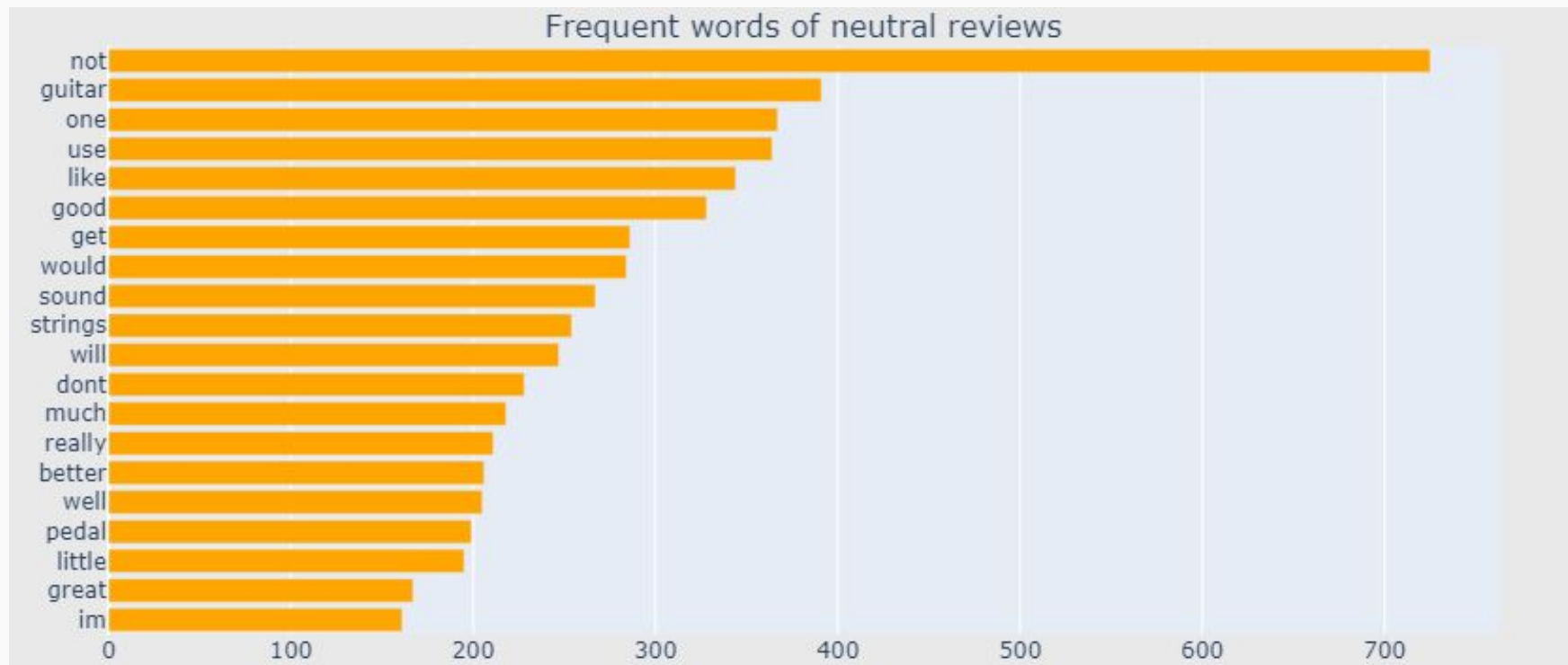# Reviews Bigram Analysis

The most frequent two words in reviews based on sentiments.

Bigram Frequency of Positive Reviews

Frequent two words of positive reviews

Bigram Frequency of Neutral Reviews
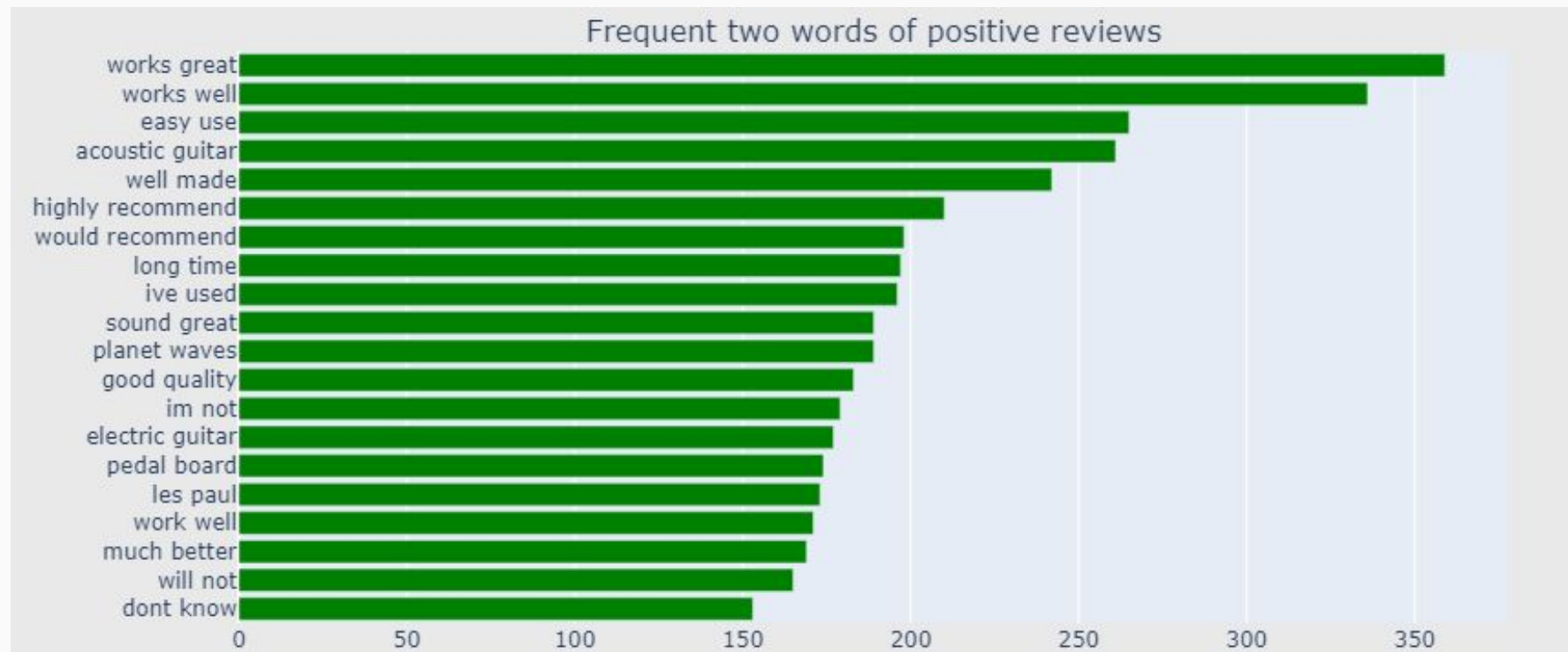
Frequent two words of neutral reviews

Bigram Frequency of Negative Reviews

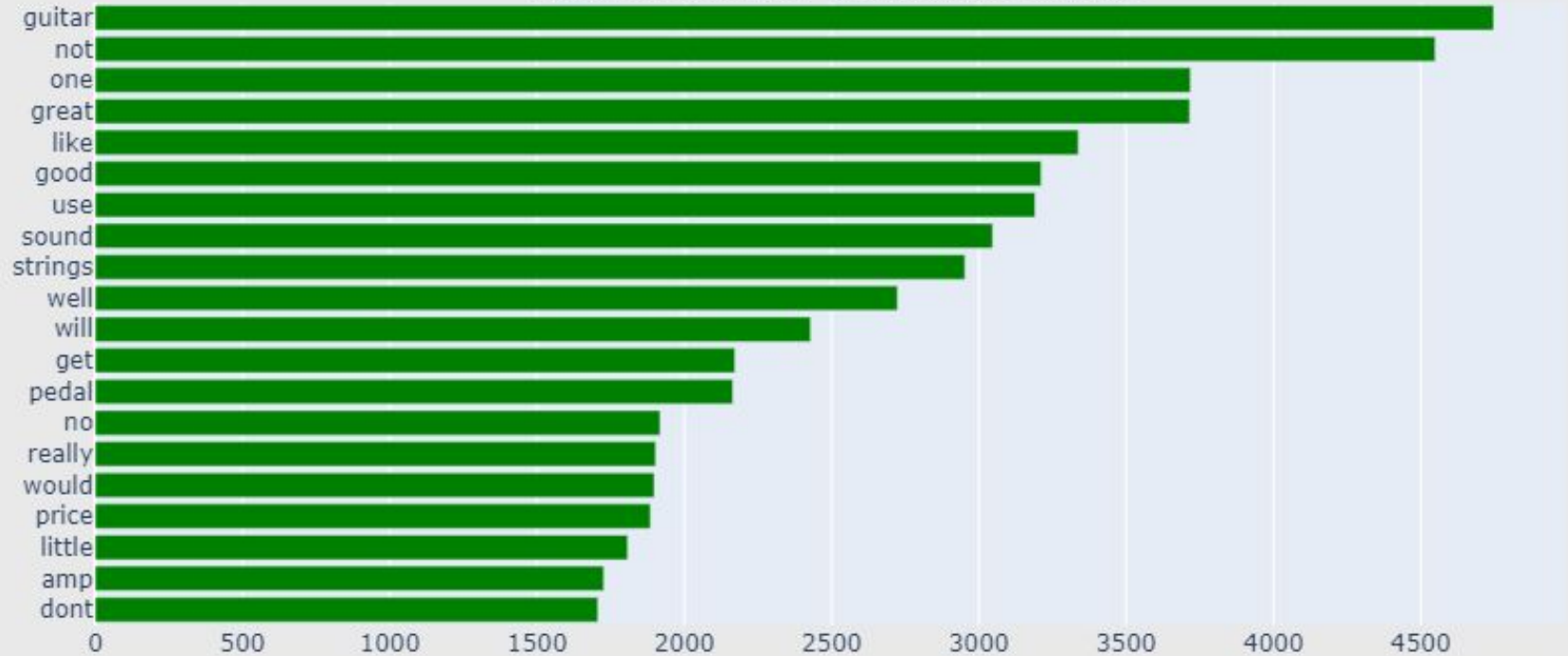Frequent two words of negative reviews
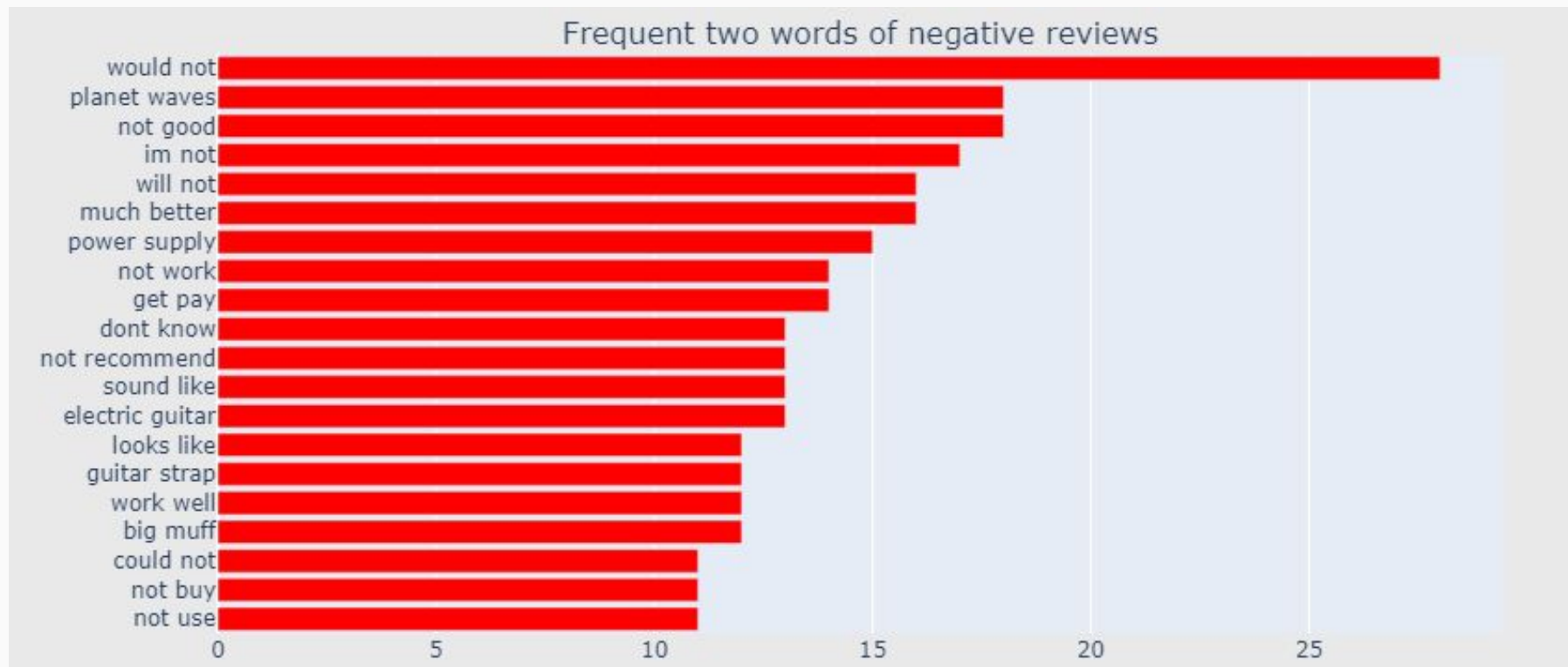
Summaries Bigram Analysis

The most frequent two words in summaries based on sentiments.

Bigram Frequency of Positive Summaries

Frequent two words of positive summaries

Bigram	
works great	
great price	
great product	
works well	
great strings	
good price	
great little	
good quality	
great tuner	
guitar strings	
great value	
easy use	
good strings	
guitar stand	
great pedal	
good product	
great guitar	
great stand	
acoustic guitar	
five stars	

Bigram Frequency of Neutral Summaries

Frequent words of neutral summaries

Bigram Frequency of Negative Summaries

Frequent two words of negative summaries

Bigram	Frequency
not good	9
not work	8
doesnt work	6
stopped working	5
poor quality	5
get pay	5
waste money	5
didnt work	4
not great	4
quality control	4
something else	4
not much	3
cheaply made	3
not worth	3
fell apart	3
dont like	3
not fit	3
not expected	3
not like	3
not thought	3

Summaries Bigram conclusion

- The bi-gram gave us better results than uni-gram.

Let us see if the tri-gram will show us better insights …

Reviews Trigram Analysis

The most frequent three words in reviews based on sentiments.

Trigram Frequency of Positive Reviews

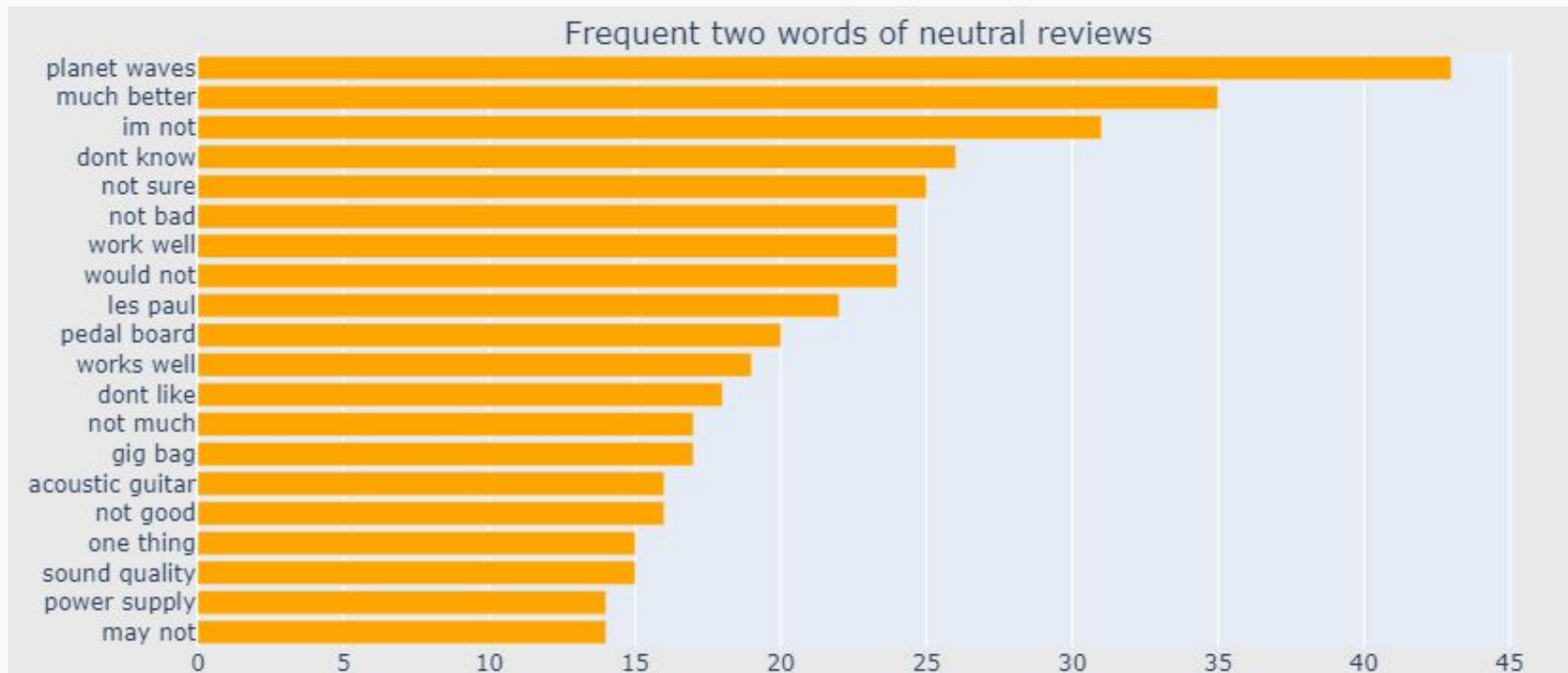Frequent three words of positive reviews

Trigram	Frequency
cant go wrong	~104
last long time	~75
would recommend anyone	~51
im not sure	~43
built like tank	~41
gets job done	~37
would highly recommend	~36
not much say	~34
ernie ball strings	~32
cant beat price	~32
buy another one	~29
gibson les paul	~27
local music store	~27
worth every penny	~27
ive ever used	~27
price cant beat	~27
time will tell	~27
solid state amp	~24
epiphone les paul	~23
acoustic electric guitars	~23

Trigram Frequency of Neutral Reviews

Frequent three words of neutral reviews

Trigram Frequency of Negative Reviews

Frequent three words of negative reviews

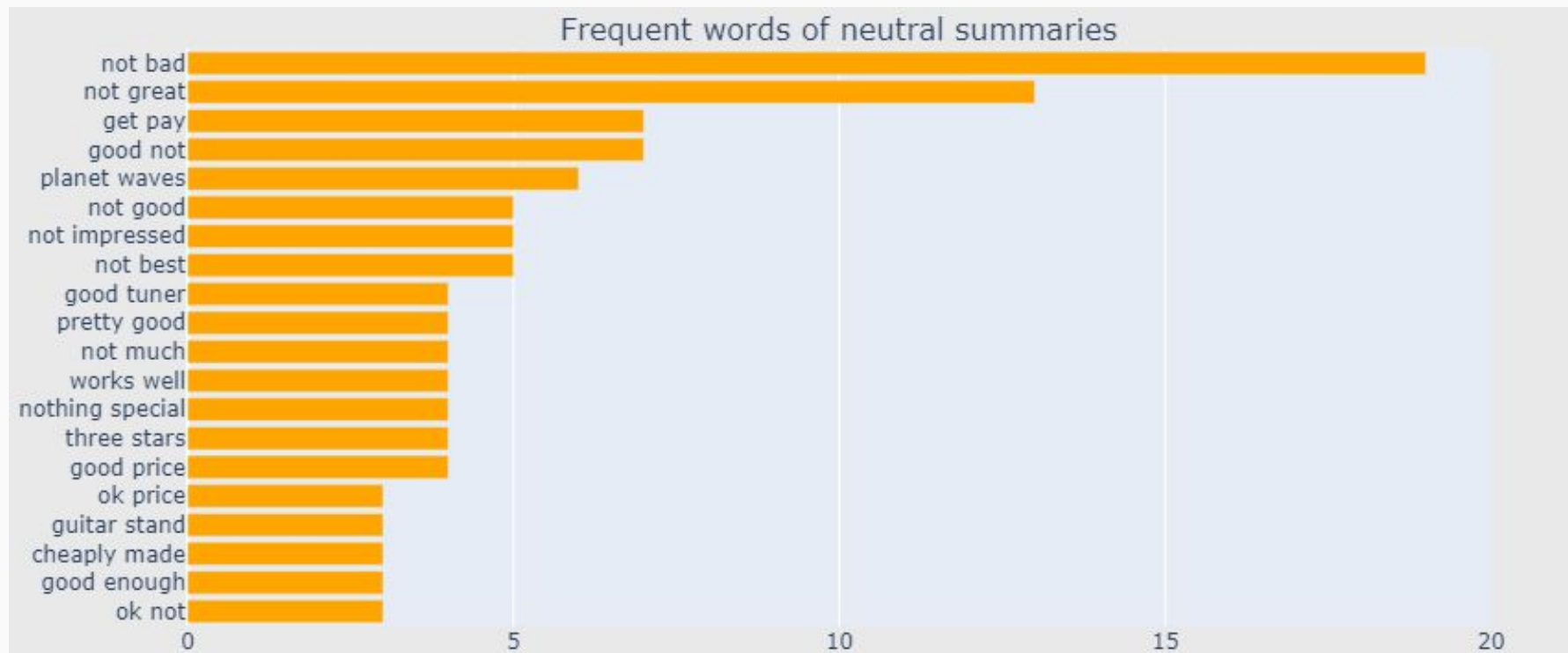| Trigram	Frequency
would not recommend	8
patches james limborg	8
boss patches james	8
ive ever used	5
im not sure	5
guitar sound like	5
pitch correction function	4
line level signal	4
low e string	4
not recommend product	4
didnt last long	4
doesnt work well	4
install boss patches	4
original spot plug	3
get something better	3
money get something	3
planet waves capo	3
ill never buy	3
high e string	3
not going work	3

Summaries Trigram Analysis

The most frequent three words in summaries based on sentiments.
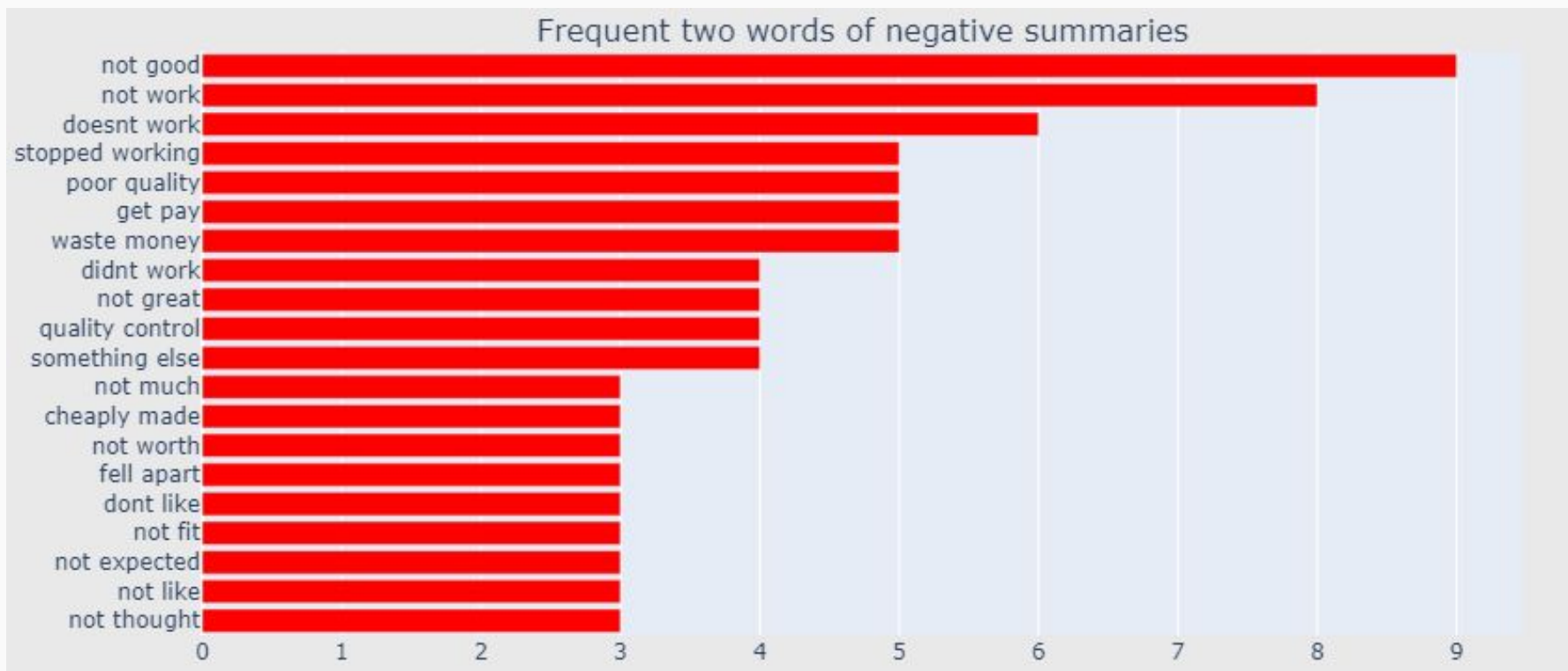
Trigram Frequency of Positive Summaries

Frequent three words of positive summaries

Trigram Frequency of Neutral Summaries
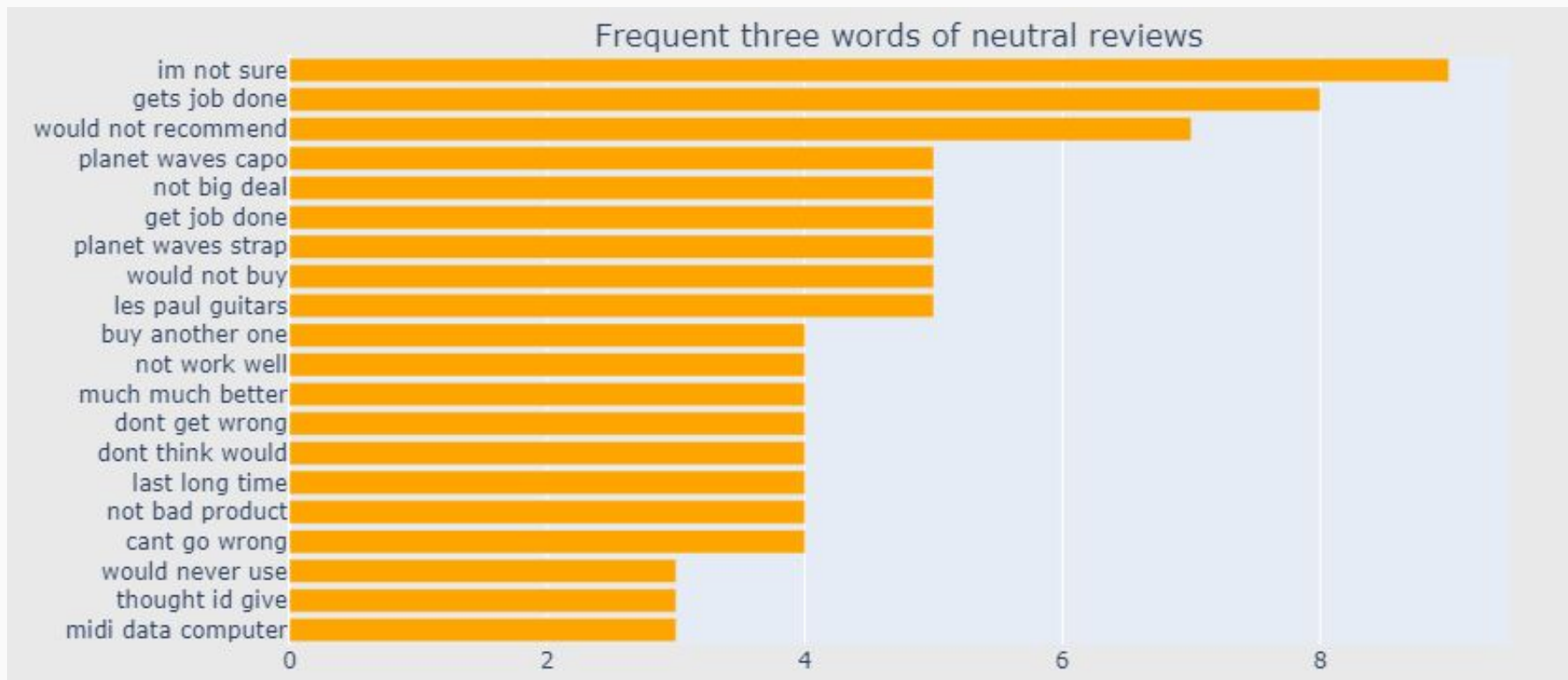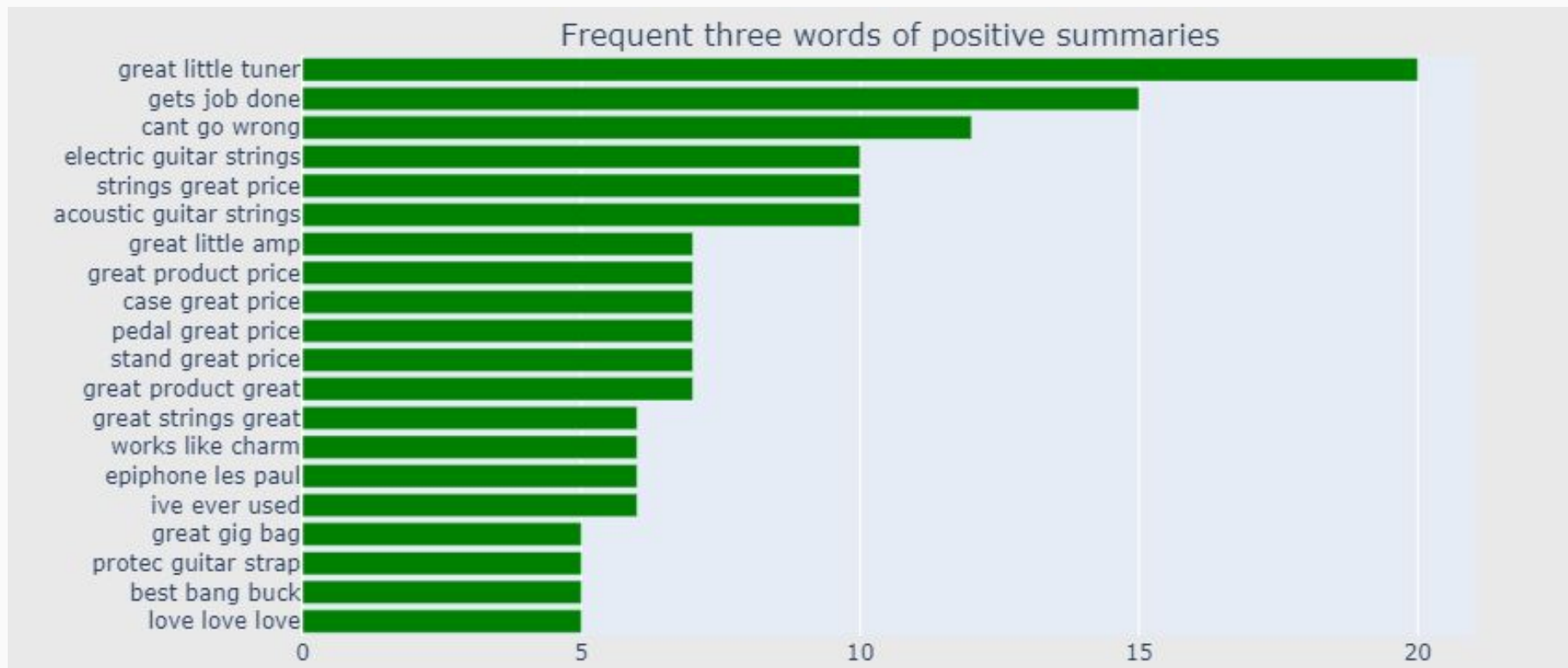

Frequent three words of neutral summaries

Trigram	Frequency
good not great	3
not bad price	2
not bad not	2
could much better	2
like lot reservations	1
hmmm like lot	1
noticeable improvement cost	1
hard see angle	1
bassheavy dull highs	1
range large fit	1
high range large	1
mid high range	1
lacking mid high	1
tuner little small	1
good tuner little	1
get something nicer	1
functional not great	1
garageband older macs	1
ipad garageband older	1
work ipad garageband	1

Trigram Frequency of Negative Summaries

Frequent three words of negative summaries

Trigram	Frequency
not worth money	2
poor quality control	2
low e string	2
ive ever used	2
fan elixir strings	1
big fan elixir	1
im big fan	1
waves triaction capo	1
planet waves triaction	1
older planet waves	1
variant older planet	1
bizarre variant older	1
everywhere except one	1
snarks everywhere except	1
use snarks everywhere	1
music reproduction quit	1
excellent music reproduction	1
taking much space	1
ends taking much	1
far ends taking	1

Trigram Analysis Conclusion

As we see, the trigram model gave us the best results as it relies not only on a word itself but the context and words around the target word.

Thank you 🎉 👌