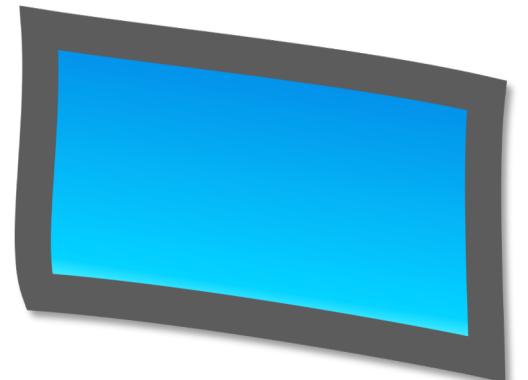
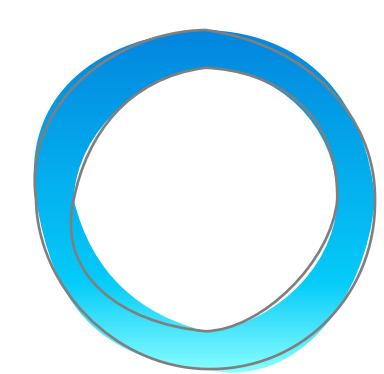


Is talking to Alexa enough? Or do we need more?

**A study to explore whether Alexa-enabled Echo devices are better
with a screen based on user experience**

Siegfried Leung



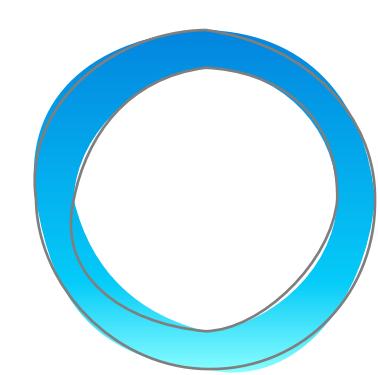


How do you interact with Alexa?



amazon alexa





Study user reviews on Echo devices with and without a screen

Echo Show 1st Gen



Echo Show 2nd Gen



Echo Spot 1st Gen

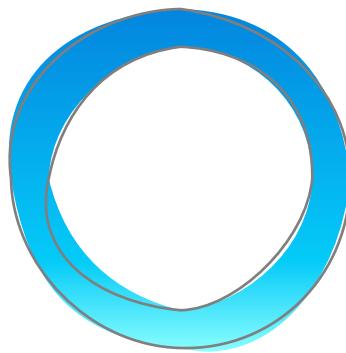


Echo Plus 2nd Gen



Echo Plus 1st Gen





Outline

- Exploratory data analysis on reviews/star ratings/length
- Word cloud analysis to examine best and worse reviews
- Vader sentiment analysis to compare reviews against star ratings
- Sentiment polarity ratio analysis to profile reviewers
- Conclusion: is Echo better with a screen?

Echo Show 1st Gen



Echo Show 2nd Gen



Echo Plus 1st Gen

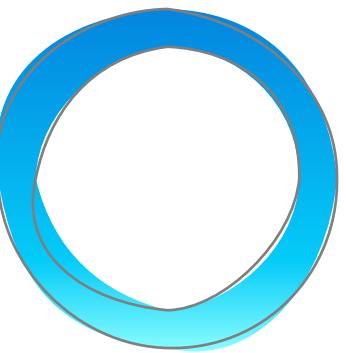


Echo Spot 1st Gen



Echo Plus 2nd Gen



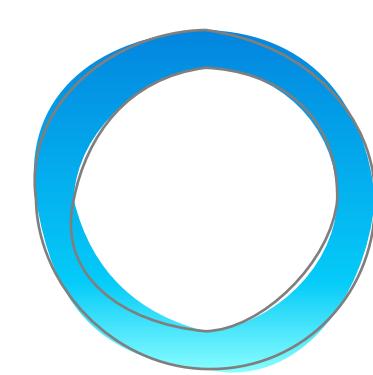


User experience based on Amazon reviews

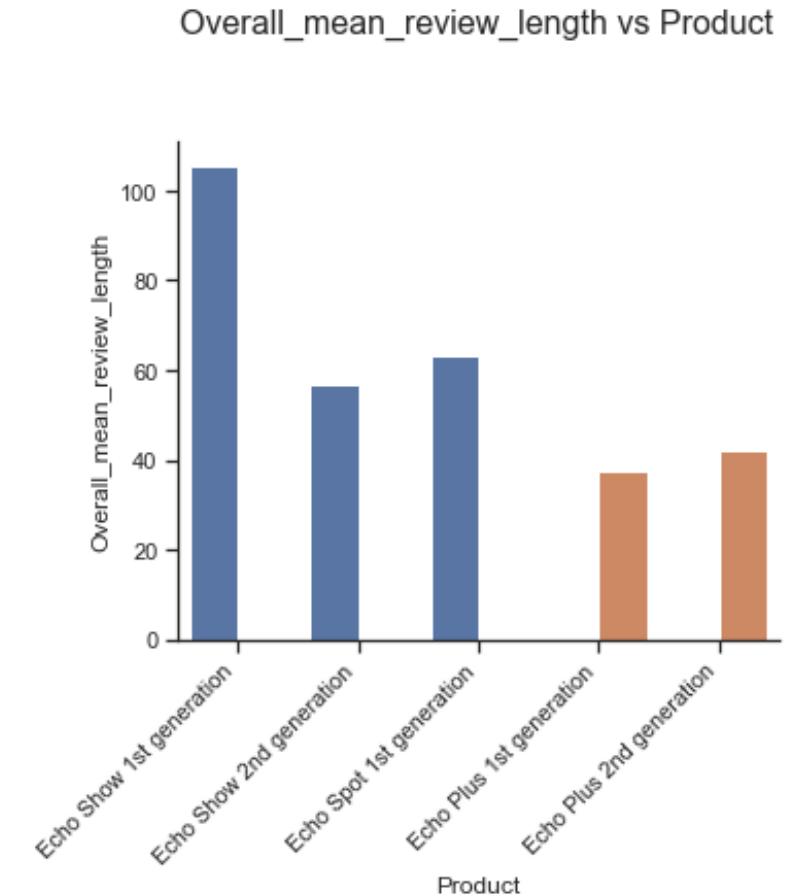
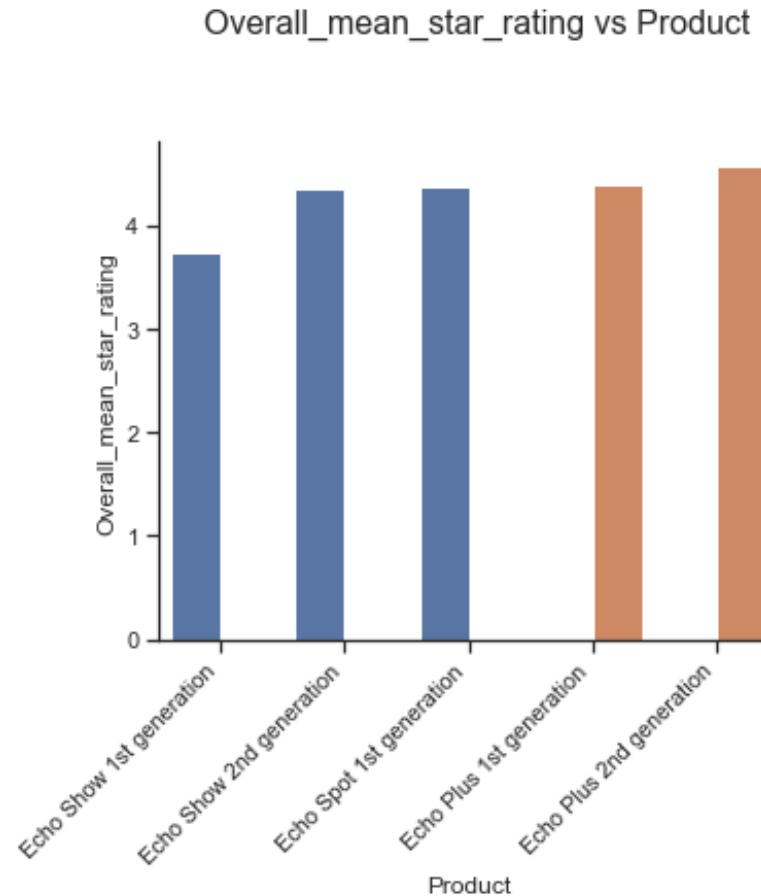
- Data: reviews of 5 Echo devices by 2019/02

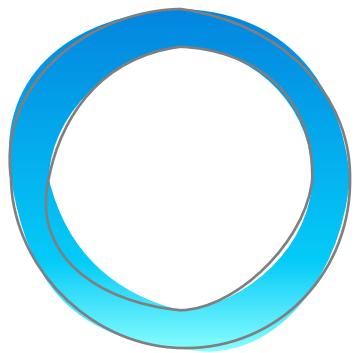


- Reviews extracted from Amazon using the web scrapper in conjunction with Chrome (<https://www.scrapehero.com/amazon-review-scraper/>)
- Extracted data include review title, review content, star rating, author, and date
- Up to 4000 reviews per products
- Text data (review title & content) was cleaned for tokenization

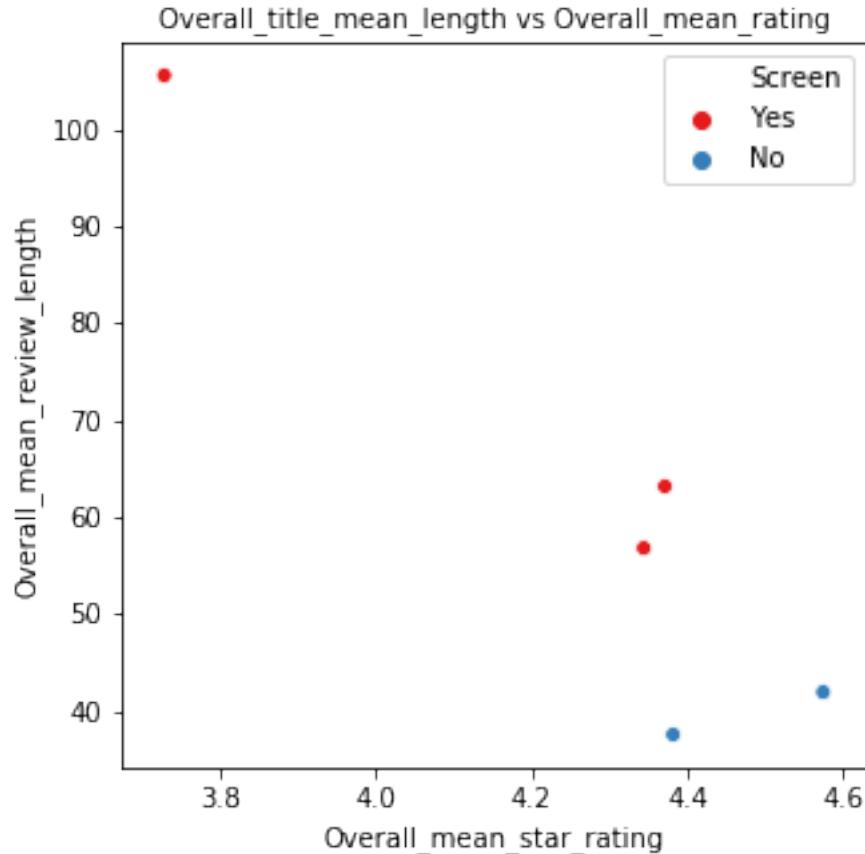


Echo Show Gen 1 has a slightly lower rating than other products, which have >4-star in average

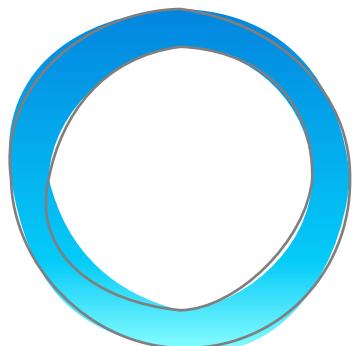




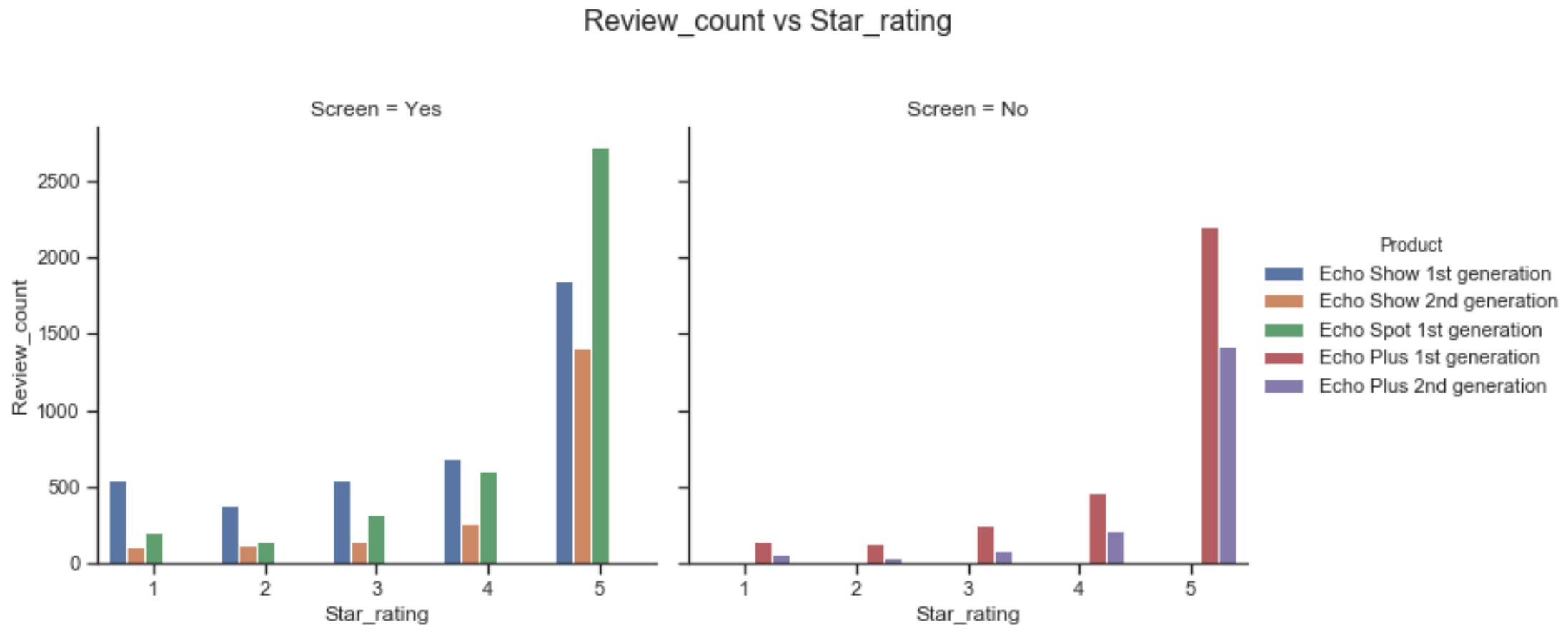
Reviewers appear to write more when satisfaction is low

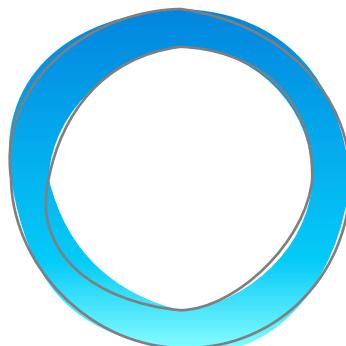


- At first glance, reviewers appeared to write more when they reviewed devices with less stars.



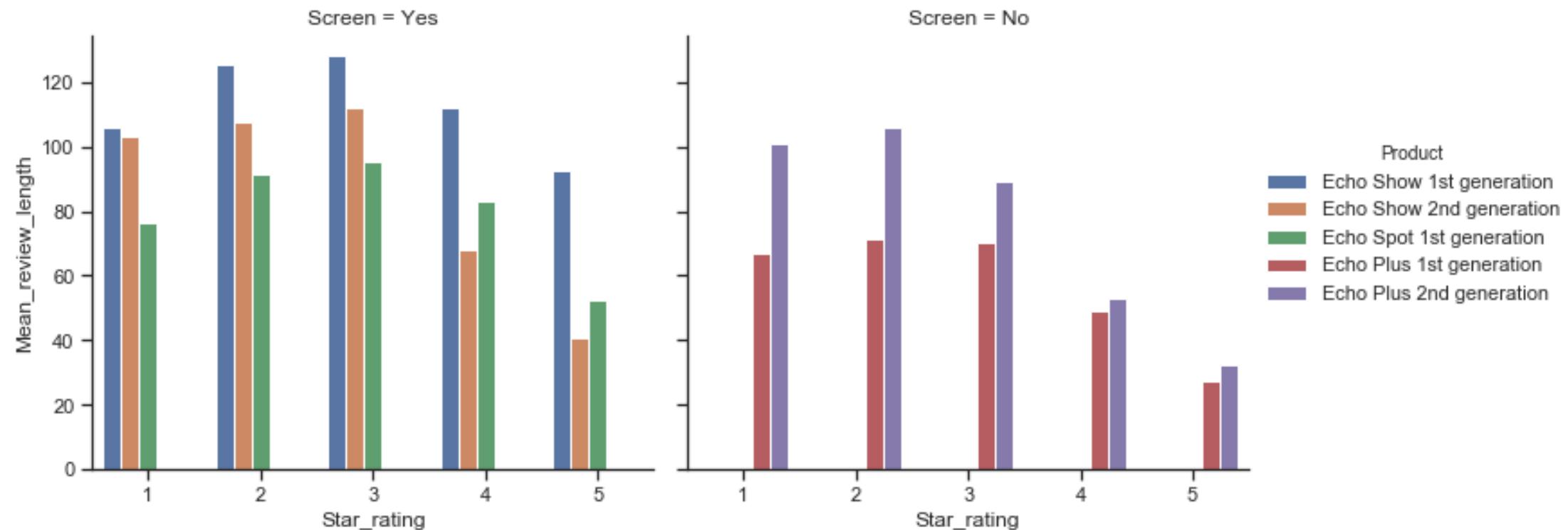
Data are skewed – most are 5-star reviews

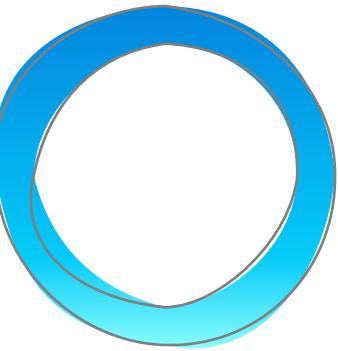




Again, the data suggests that the reviews tend to be longer if the star rating is lower

Mean review word count vs Star rating





What do people like about Echo devices?

Word clouds of 5-star reviews



Echo Show 1st Gen



Echo Show 2nd Gen



Echo Spot 1st Gen



Wordcloud of reviews



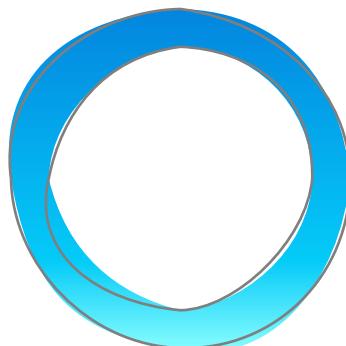
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people like about Echo devices?

Word clouds of 5-star reviews



Echo Show 1st Gen



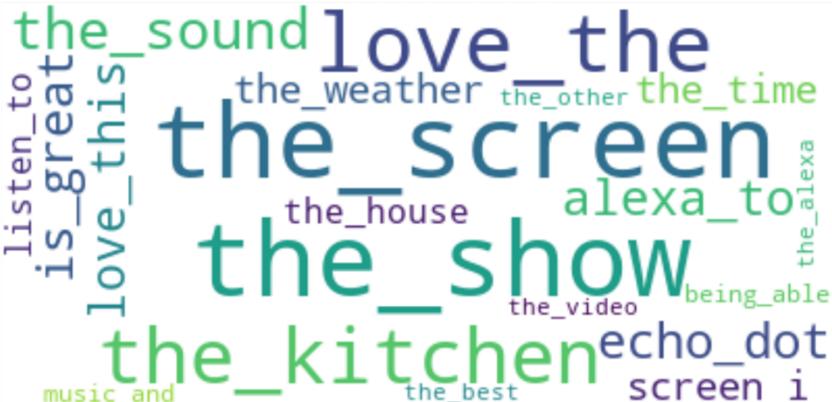
Echo Show 2nd Gen



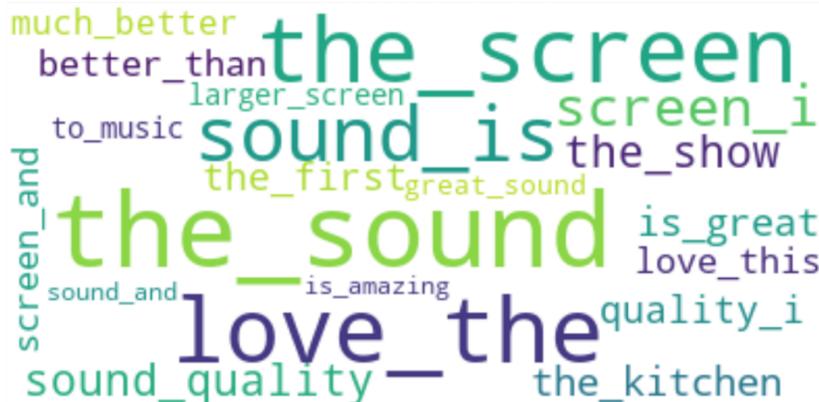
Echo Spot 1st Gen



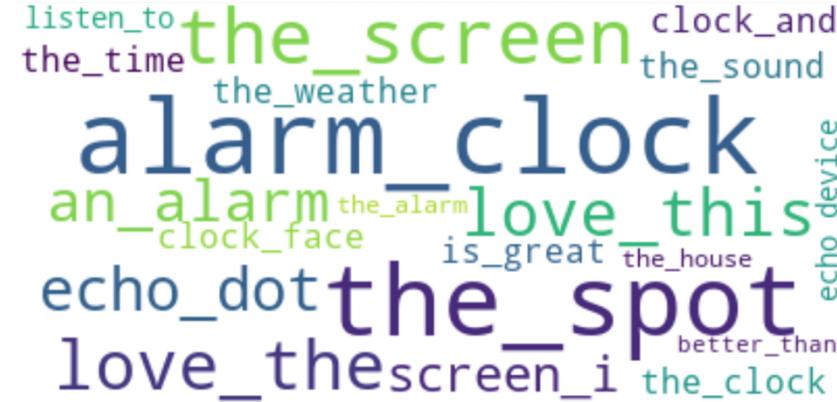
Wordcloud of reviews



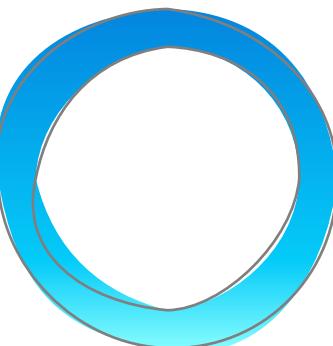
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people like about Echo devices? Word clouds of 5-star reviews



Echo Show 1st Gen



Echo Show 2nd Gen



Echo Spot 1st Gen



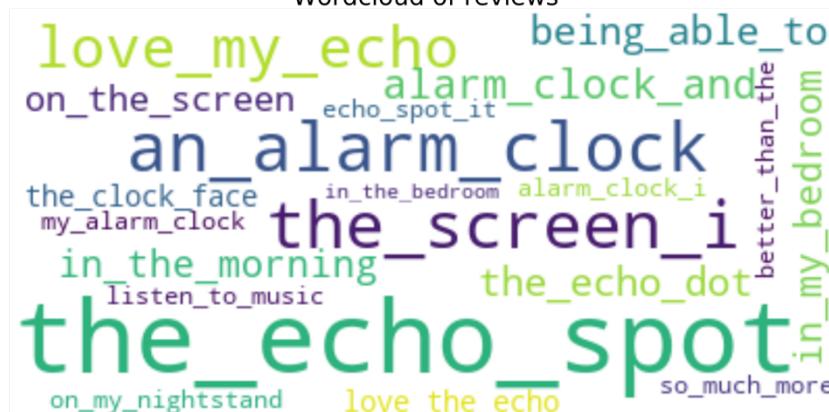
Wordcloud of reviews



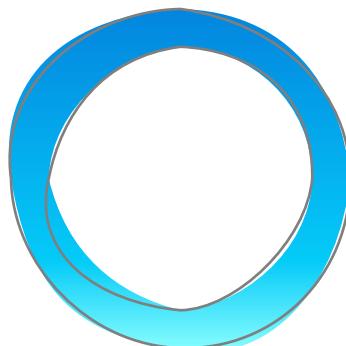
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people like about Echo devices?

Word clouds of 5-star reviews



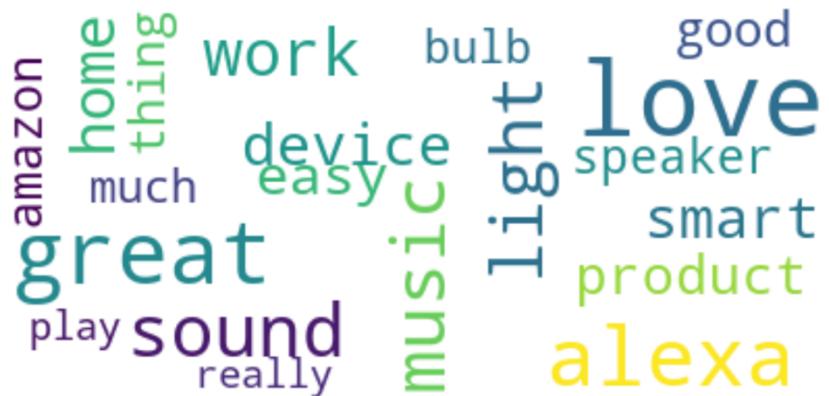
Echo Plus 1st Gen



Echo Plus 2nd Gen



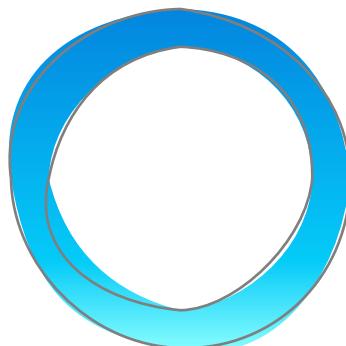
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people like about Echo devices?

Word clouds of 5-star reviews



Echo Plus 1st Gen



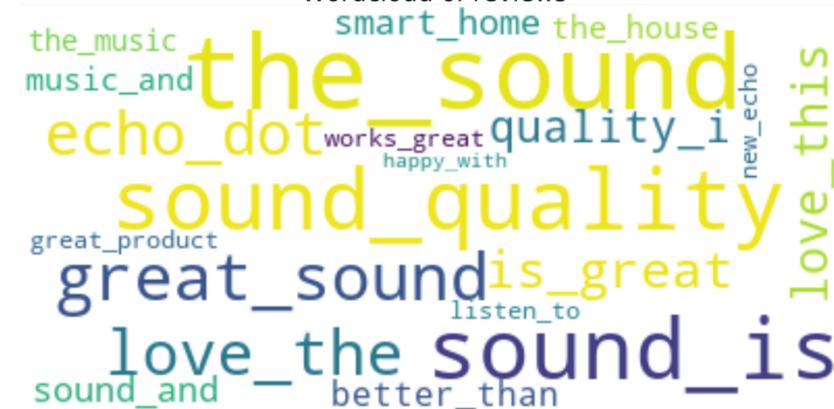
Echo Plus 2nd Gen



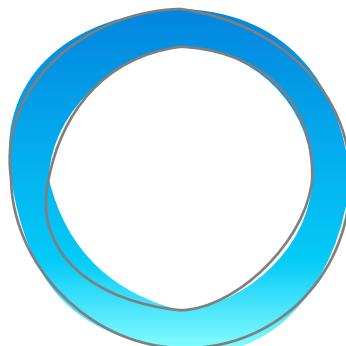
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people like about Echo devices?

Word clouds of 5-star reviews



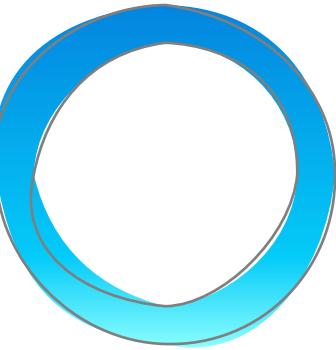
Echo Plus 1st Gen



Echo Plus 2nd Gen



The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices?

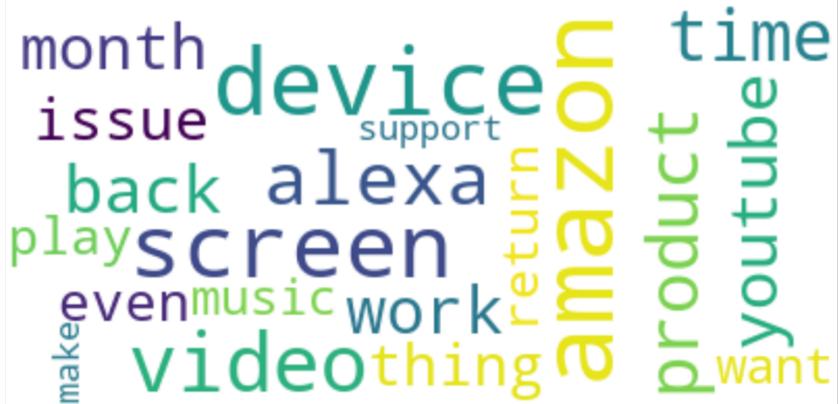
Word clouds of 1-star reviews



Echo Show 1st Gen



Wordcloud of reviews



Echo Show 2nd Gen



Wordcloud of reviews



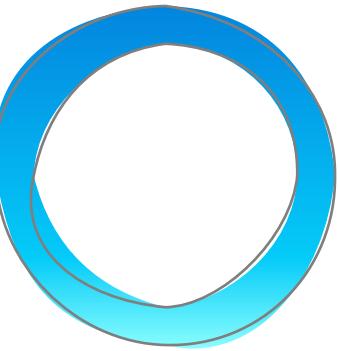
Echo Spot 1st Gen



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices?

Wordclouds of 1-star reviews



Echo Show 1st Gen



Word cloud of titles provides more insight for
Echo Show 1st Gen

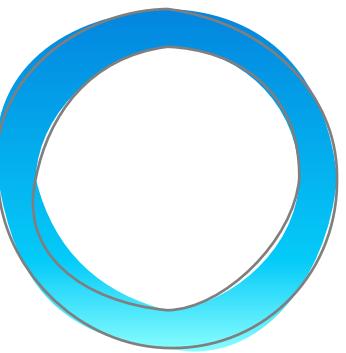
Echo Show 2nd Gen



Echo Spot 1st Gen



The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices?

Word clouds of 1-star reviews



Echo Show 1st Gen



Echo Show 2nd Gen



Echo Spot 1st Gen



Wordcloud of reviews



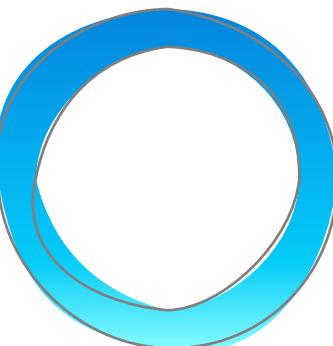
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency

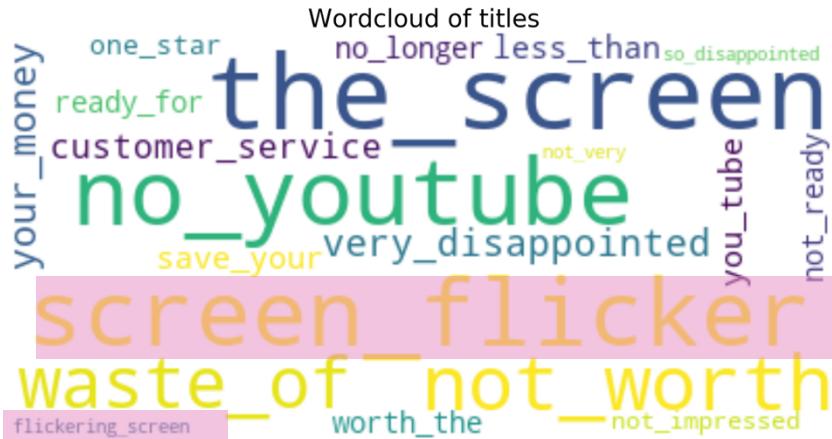


What do people dislike about Echo devices?

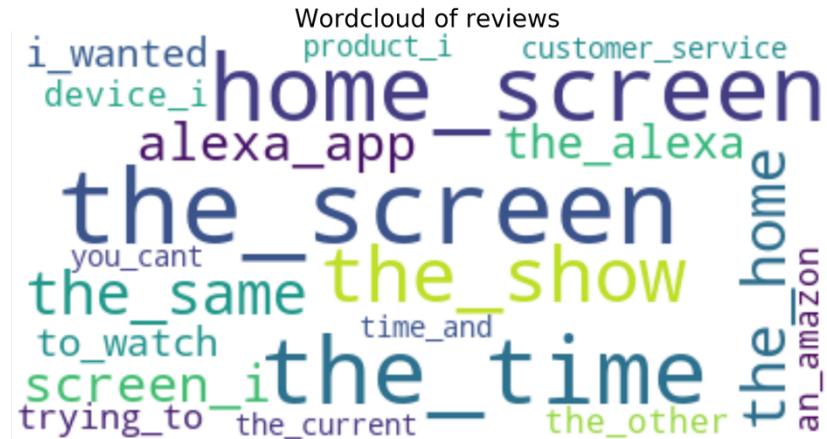
Word clouds of 1-star reviews



Echo Show 1st Gen



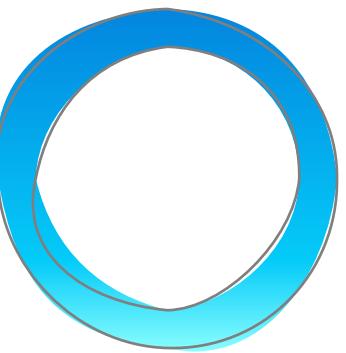
Echo Show 2nd Gen



Echo Spot 1st Gen



The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices?

Word clouds of 1-star reviews



Echo Show 1st Gen



Echo Show 2nd Gen



Echo Spot 1st Gen



Wordcloud of reviews

the_screen_started
the_device_i the_same_issue flash_player_i
intended_for_all on_the_screen
not_intended_for of_the_screen
the_screen_i what_date_were
to_return_it on_the_phone
the_alex_app i_would_have_a_few_months

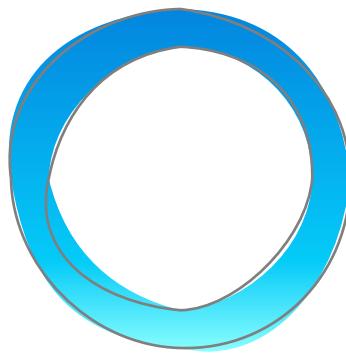
Wordcloud of reviews

this_product_i you_can_only
the_same_thing of_the_screen
want_to_return amazon_echo_show ask_alexa_to
the_home_screen
an_amazon_echo this_device_i the_screen_is
the_alex_app over_and_over
and_echo_show with_the_echo this_would_be
the_other_echo echo_show_and
to_do_everything this_thing_is

Wordcloud of reviews

bought_this_to on_the_phone on_the_screen
things_to_try _things
i_bought_this a_bedside_clock the_echo_spot
the_clock_face an_alarm_clock the_screen_i this_device_i
waste_of_money the_things_to does_not_work that_you_can

The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices?

Word clouds of 1-star reviews



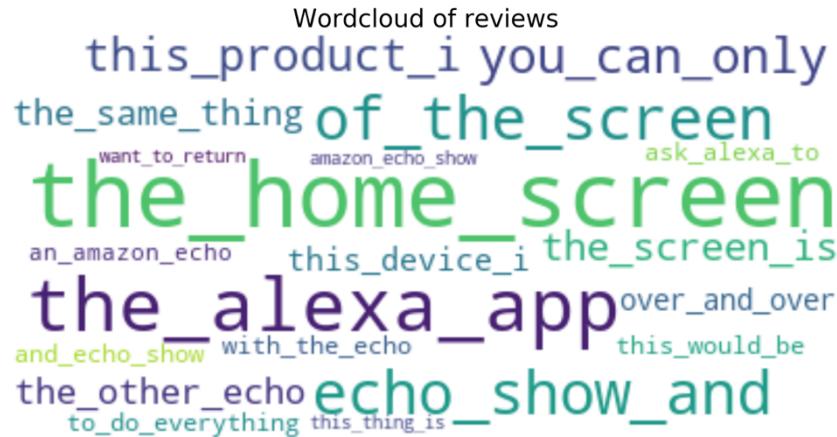
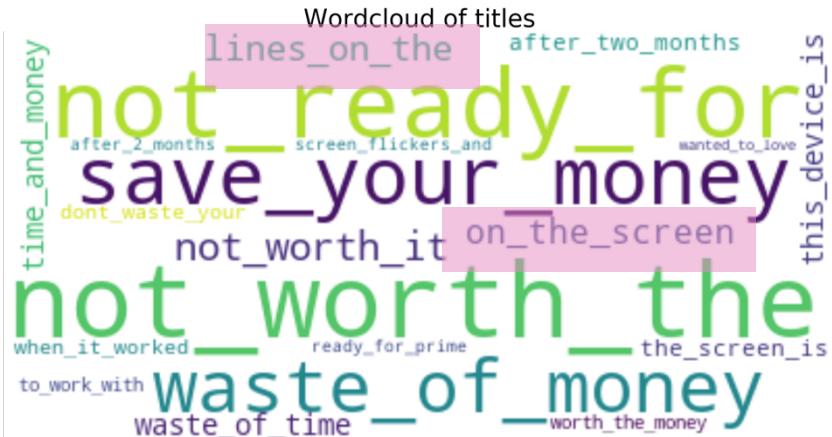
Echo Show 1st Gen



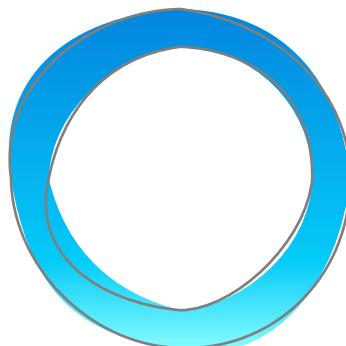
Echo Show 2nd Gen



Echo Spot 1st Gen



The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices?

Word clouds of 1-star reviews



Echo Plus 1st Gen



Echo Plus 2nd Gen



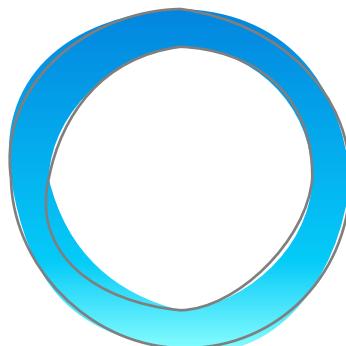
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices?

Word clouds of 1-star reviews



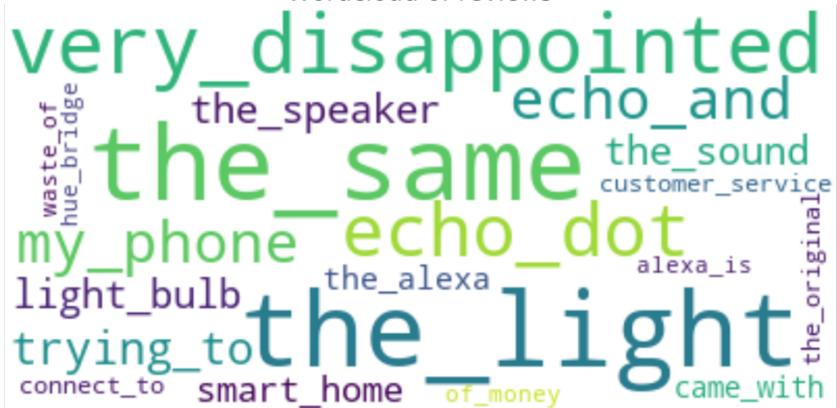
Echo Plus 1st Gen



Echo Plus 2nd Gen



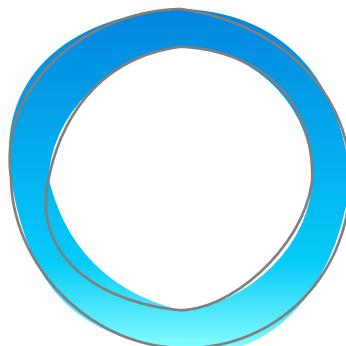
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency



What do people dislike about Echo devices? Word clouds of 1-star reviews



Echo Plus 1st Gen



Echo Plus 2nd Gen



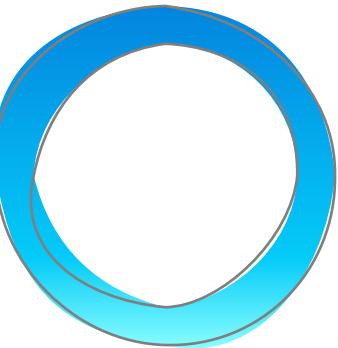
Wordcloud of reviews



Wordcloud of reviews



The top 20 most frequent words/phrase are shown; size based on frequency

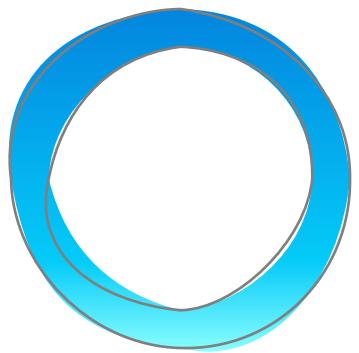


VADER Sentiment analysis

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is *specifically attuned to sentiments expressed in social media*

Outputs

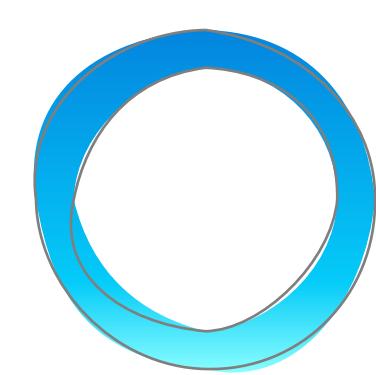
- Compound score - *single unidimensional measurement of sentiment*:
 - positive sentiment: compound score ≥ 0.05
 - neutral sentiment: $-0.05 < \text{compound score} < 0.05$
 - negative sentiment: compound score ≤ -0.05
- Pos, Neu, Neg scores - *multidimensional measurement of sentiment*:
 - These scores are ratios for proportions of text that fall in each category (Sum = 1)



VADER Sentiment analysis

Examples:

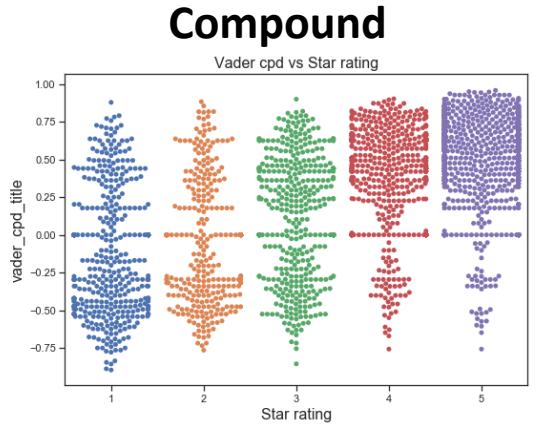
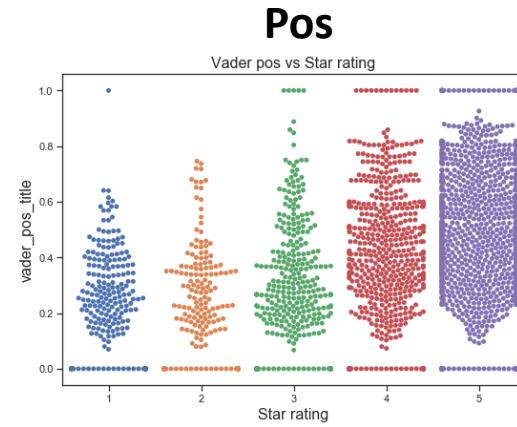
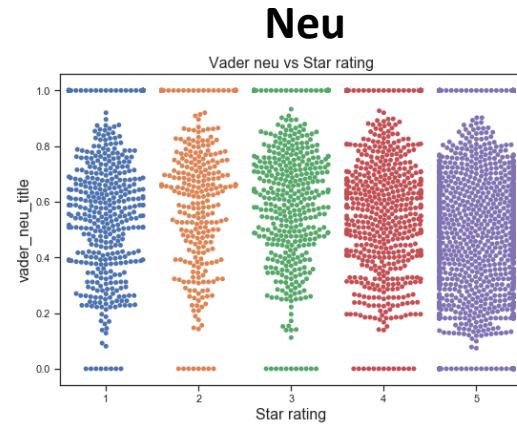
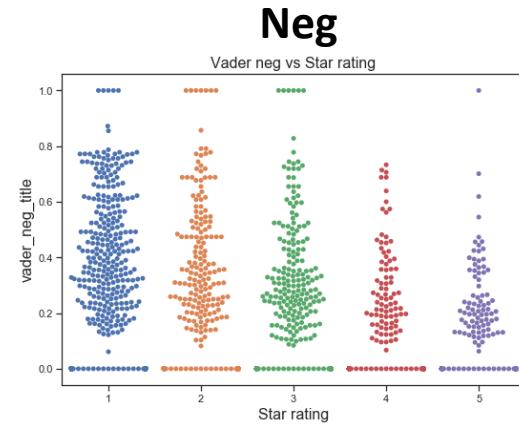
- VADER is smart, handsome, and funny.----- {'pos': 0.746, 'compound': 0.8316, 'neu': 0.254, 'neg': 0.0}
- VADER is smart, handsome, and funny!----- {'pos': 0.752, 'compound': 0.8439, 'neu': 0.248, 'neg': 0.0}
- VADER is very smart, handsome, and funny.----- {'pos': 0.701, 'compound': 0.8545, 'neu': 0.299, 'neg': 0.0}
- VADER is VERY SMART, handsome, and FUNNY.----- {'pos': 0.754, 'compound': 0.9227, 'neu': 0.246, 'neg': 0.0}
- VADER is VERY SMART, handsome, and FUNNY!!!----- {'pos': 0.767, 'compound': 0.9342, 'neu': 0.233, 'neg': 0.0}
- VADER is VERY SMART, uber handsome, and FRIGGIN FUNNY!!!-- {'pos': 0.706, 'compound': 0.9469, 'neu': 0.294, 'neg': 0.0}
- VADER is not smart, handsome, nor funny.----- {'pos': 0.0, 'compound': -0.7424, 'neu': 0.354, 'neg': 0.646}
- Today only kinda sux! But I'll get by, lol----- {'pos': 0.317, 'compound': 0.5249, 'neu': 0.556, 'neg': 0.127}
- Make sure you :) or :D today!----- {'pos': 0.706, 'compound': 0.8633, 'neu': 0.294, 'neg': 0.0}
- Catch utf-8 emoji such as ❤️ and 💋 and 😊----- {'pos': 0.279, 'compound': 0.7003, 'neu': 0.721, 'neg': 0.0}
- Not bad at all----- {'pos': 0.487, 'compound': 0.431, 'neu': 0.513, 'neg': 0.0}



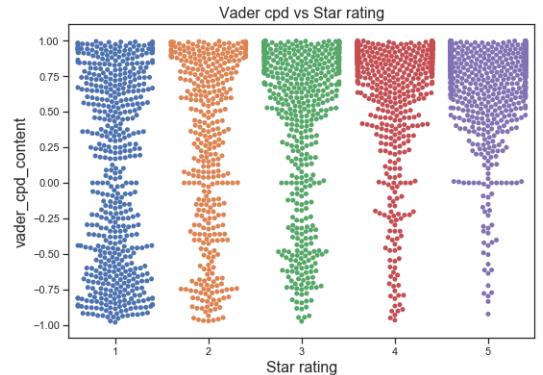
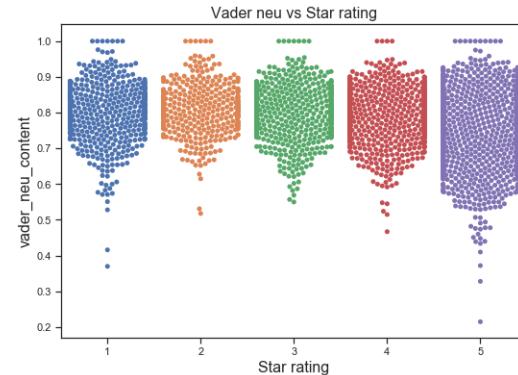
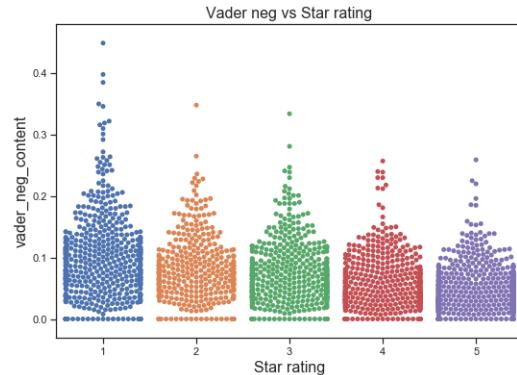
VADER sentiment analysis: Echo Show Gen 1

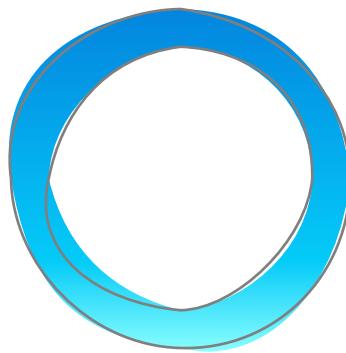
Compound, Pos and Neg scores agree with star ratings

Title



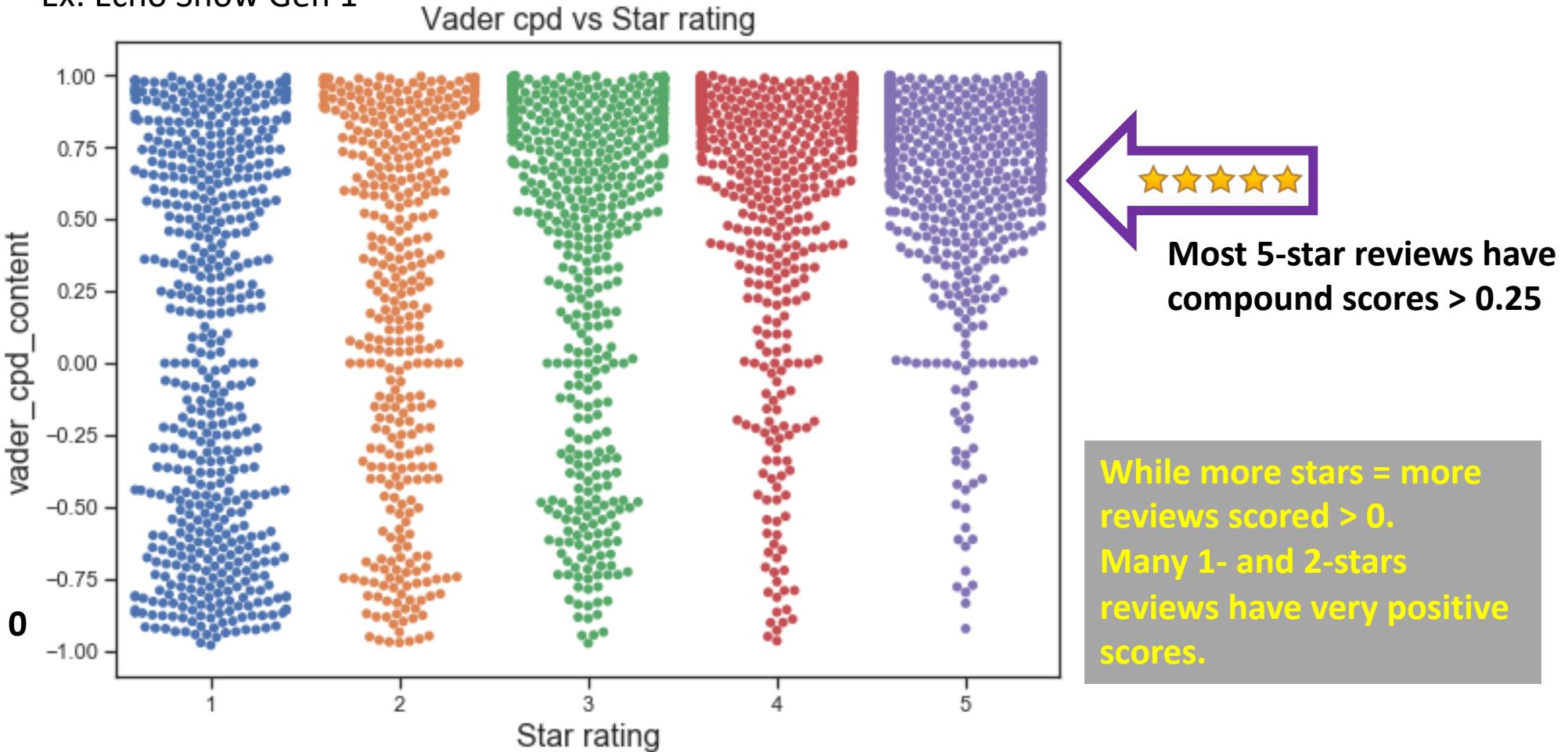
Review

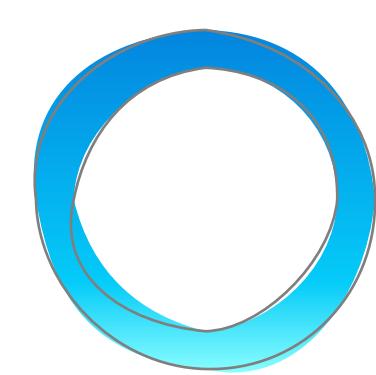




Distribution of VADER compound scores basically agrees with the star rating

Ex: Echo Show Gen 1



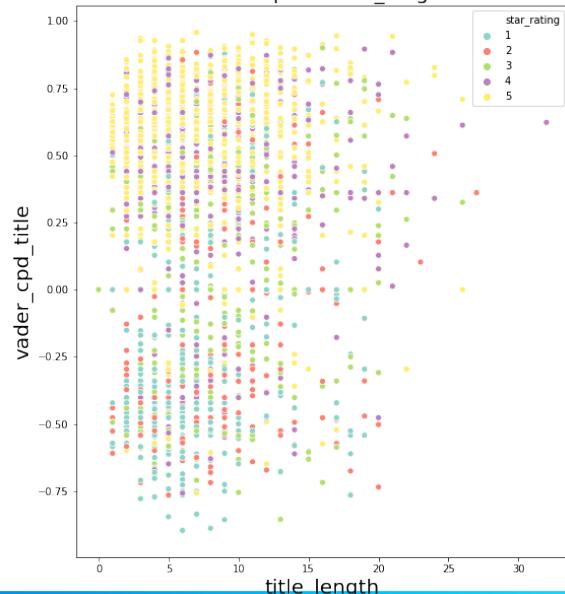


The compound scores do not seem to be affected by review's length

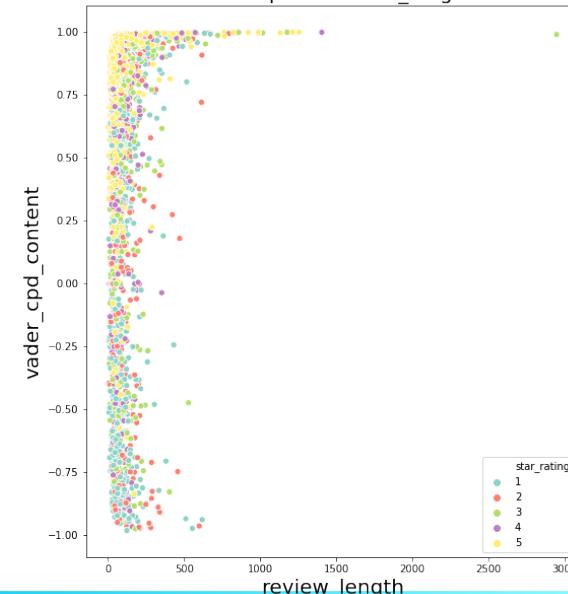
Echo Show 1st Gen



Cpd score vs Title length



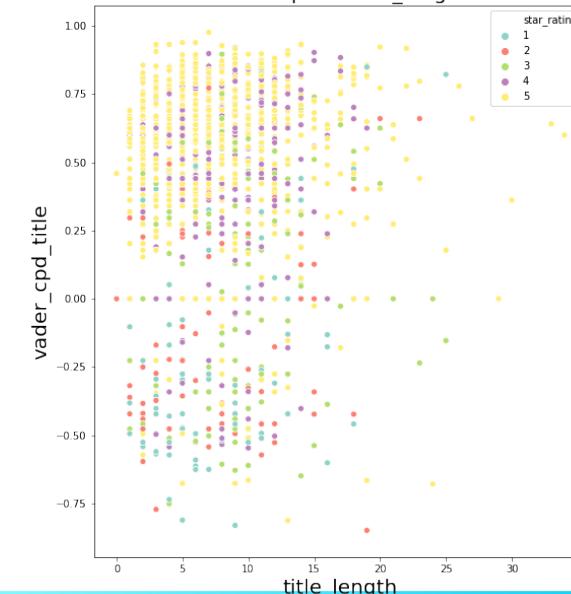
Cpd score vs Review length



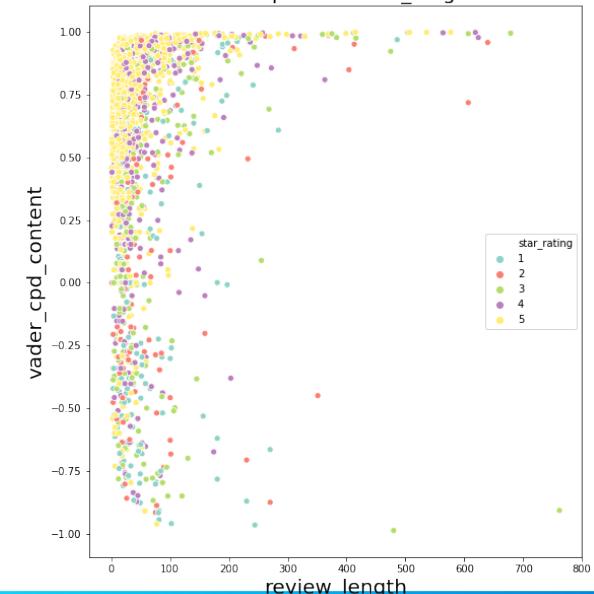
Echo Plus 1st Gen

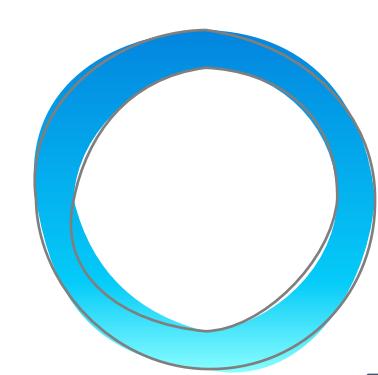


Cpd score vs Title length



Cpd score vs Review length



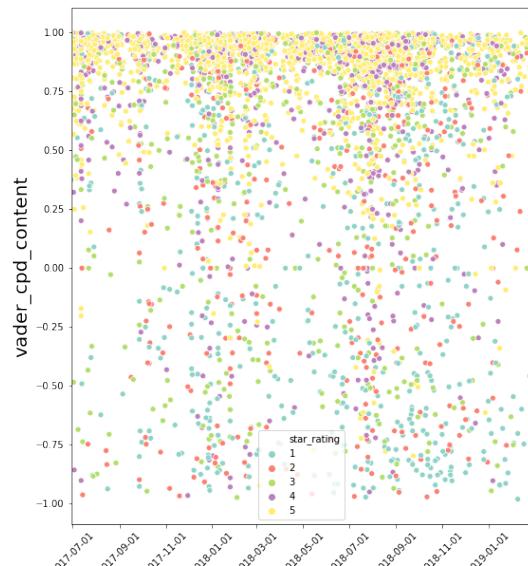


Do user's opinions change over time? Not really. (#reviews do peak after holiday/sales event)

Echo Show 1st Gen



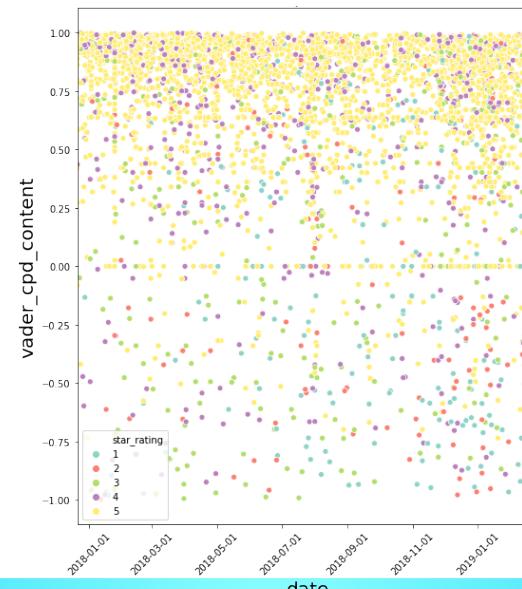
Cpd score vs Review date



Echo Spot 1st Gen



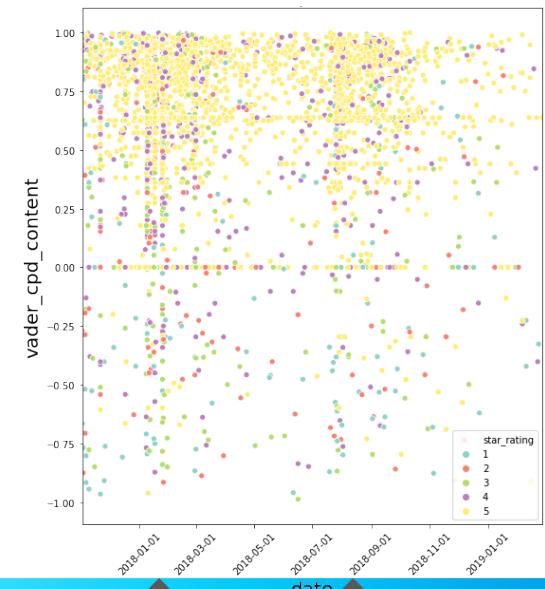
Cpd score vs Review date



Echo Plus 1st Gen



Cpd score vs Review date

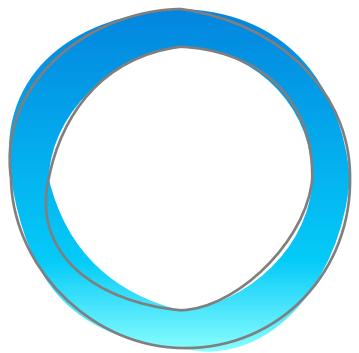


After X'mas

Prime Day

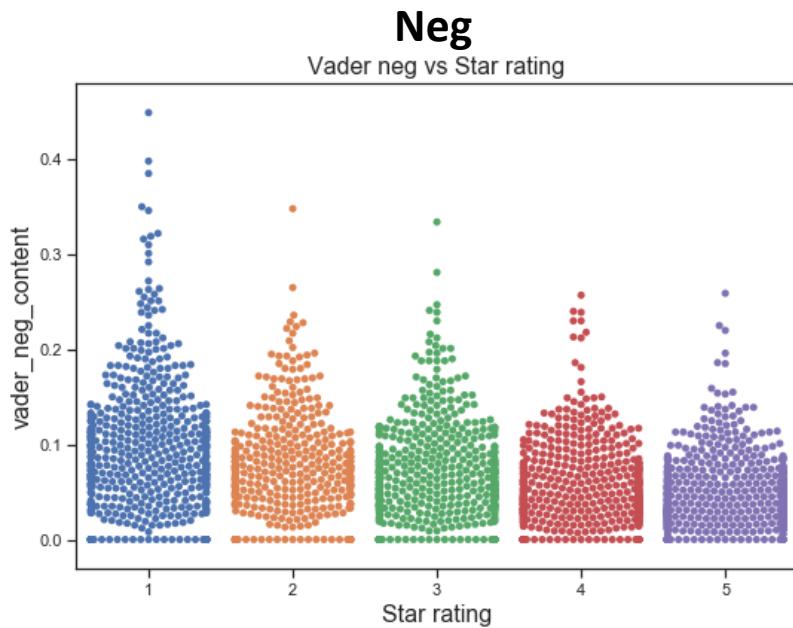
After X'mas

Prime Day

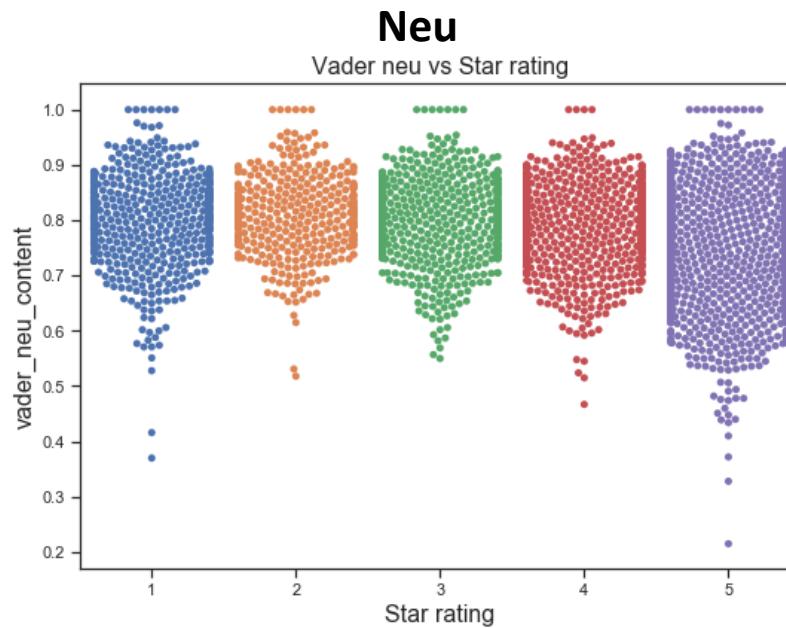


Look deeper into the positive, negative, and neutral portions of the reviews

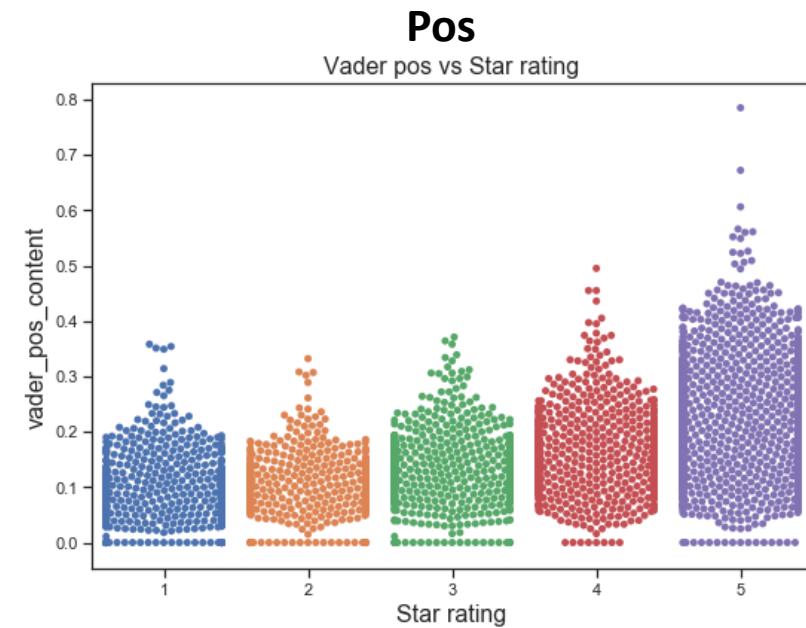
Ex: Echo Show Gen 1



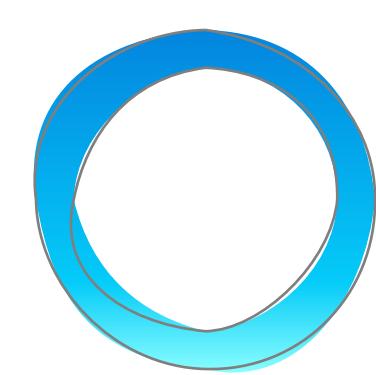
Less stars, more negative portion; but not more than half of the reviews.
(Max Neg < 0.5)



Most with Neu > 0.5.
Neutral portion appears to be similar for different star ratings.



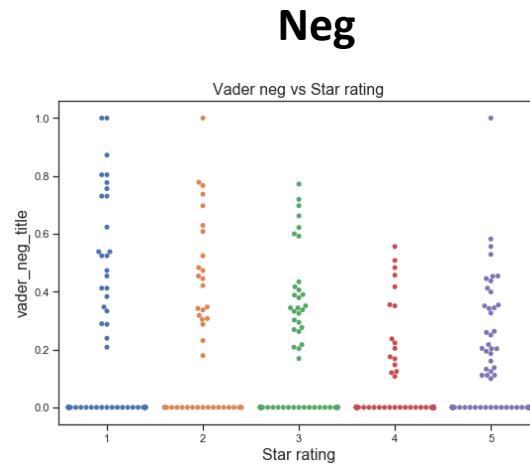
More stars, more positive portion. The range of Pos is wider than that of Neg.



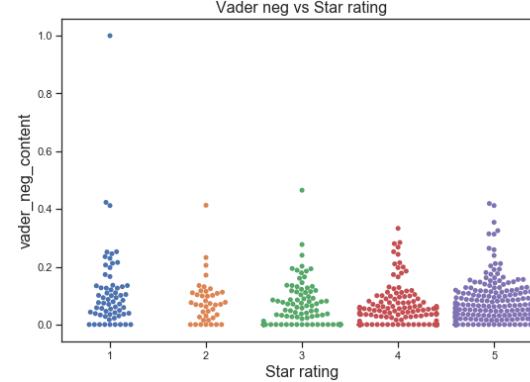
Titles yielded less information, so focus on the review content

Ex: Echo Plus Gen 2

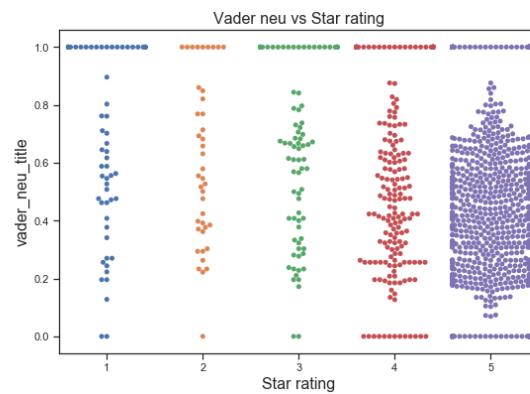
Title



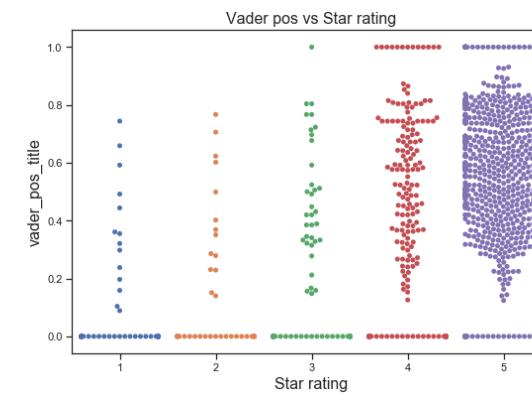
Review



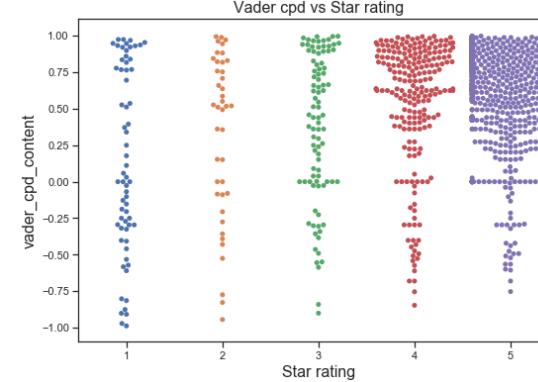
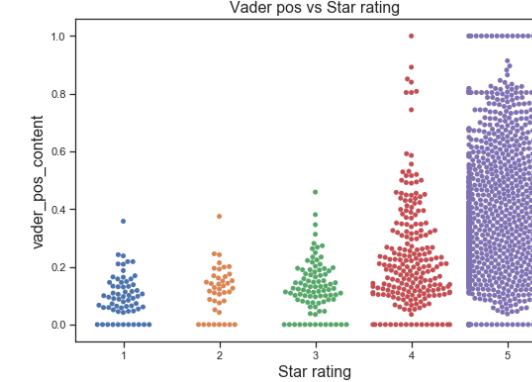
Neu

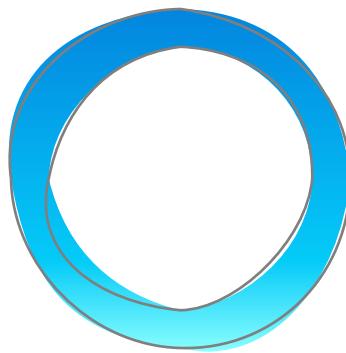


Pos



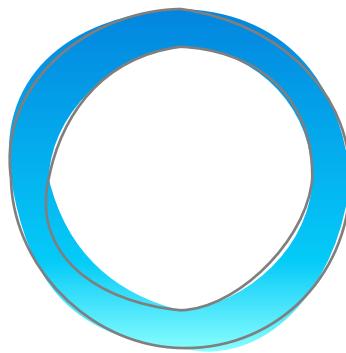
Compound





Focus on both the positives and negatives from the reviews

- The distribution of the compound scores look right (more stars = more reviews with compound score > 0)
 - But we do see 1-/2-star reviews got high scores, why?
- The Pos and Neg scores alone appear to make sense
 - But there are reviews scoring 0 or close to 0 for Pos and Neg while many have Neu score ~ 1, regardless of star rating, what does that mean?
- To better understand the reviewers:
 - Focus on the positive and negative things they said
 - The relative proportion between the two which would be independent of the review's length



Focus on both the positives and negatives from the reviews

We can calculate the *polarity* by computing the relative difference of the positive and negative portions by the Pos and Neg scores:

$$\text{Sentiment polarity ratio} = \text{Pos} / \text{Neg}$$

Examples:

Review 1 scored Pos = 0.50, Neg=0.05

$$\text{Pos/Neg} = 0.50/0.05 = 10 \rightarrow \text{Log Pos/Neg} = 1$$

Review 4 scored Pos = 0.20, Neg=0.02

$$\text{Pos/Neg} = 0.20/0.02 = 10 \rightarrow \text{Log Pos/Neg} = 1$$

Review 2 scored Pos = 0.05, Neg = 0.50

$$\text{Pos/Neg} = 0.09/0.90 = 0.1 \rightarrow \text{Log Pos/Neg} = -1$$

Review 5 scored Pos = 0.09, Neg = 0.90

$$\text{Pos/Neg} = 0.09/0.90 = 0.1 \rightarrow \text{Log Pos/Neg} = -1$$

Review 3 scored Pos = 0.25, Neg = 0.25

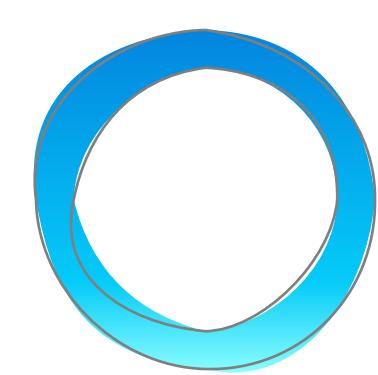
$$\text{Pos/Neg} = 0.50/0.50 = 1 \rightarrow \text{Log Pos/Neg} = 0$$

Review 6 scored Pos = 0.01, Neg = 0.01

$$\text{Pos/Neg} = 0.50/0.50 = 1 \rightarrow \text{Log Pos/Neg} = 0$$

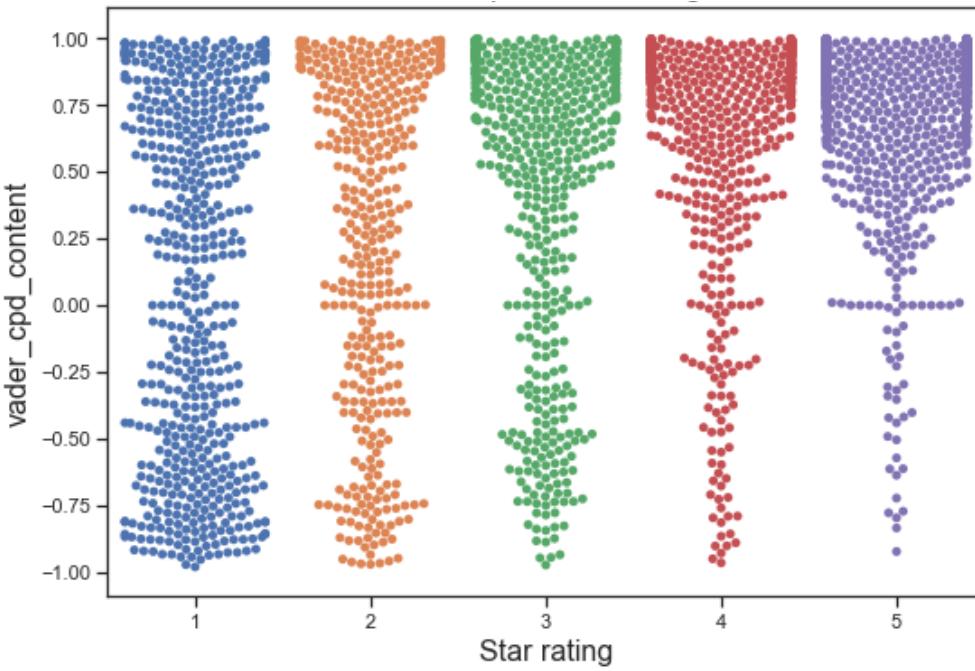
If Pos or Neg = 0.000, set value = 0.001

Ratio is independent of content length

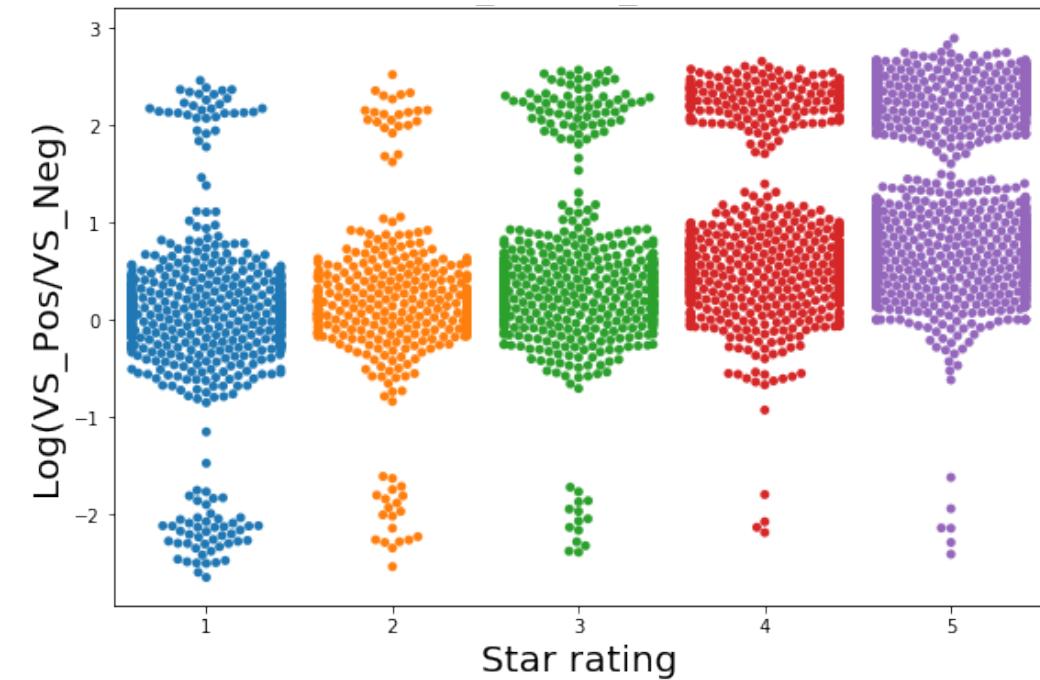


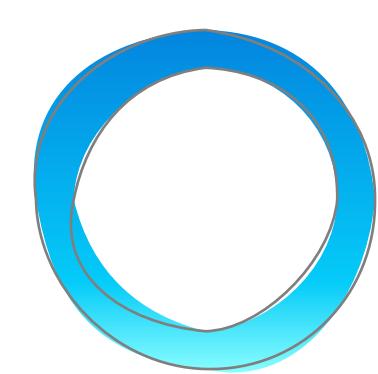
Sentiment polarity ratio:
the relative proportion between the positive
and negative text (*Pos/Neg*)

Vader Compound scores



Log(Pos/Neg)



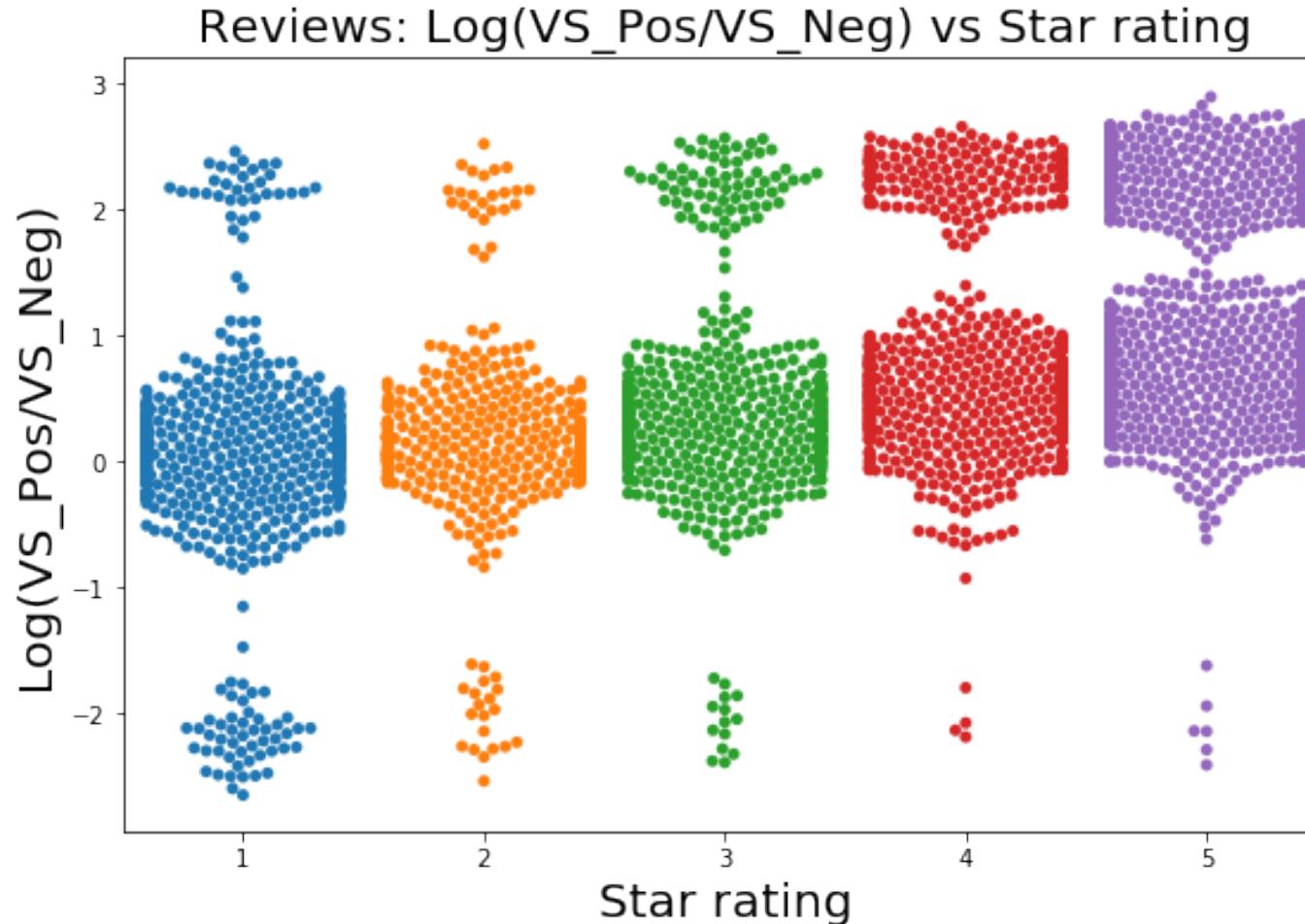


Now we can see, relative to the amount of negative text, how positive the reviews are

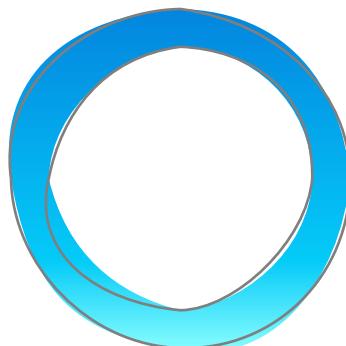
Ex: Echo Show Gen 1



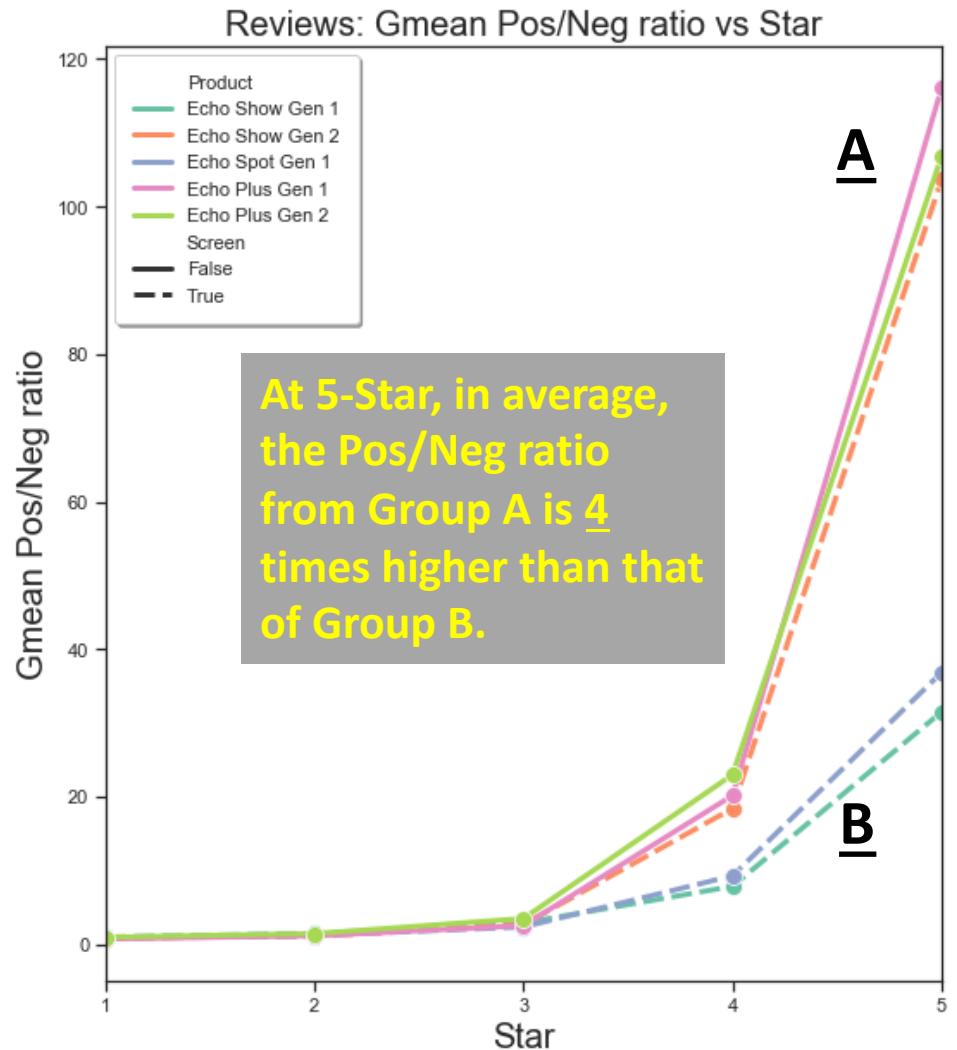
Most have
 $\text{Log}(\text{Pos}/\text{Neg}) = +/- 1$:
Pos vs Bad within 10
fold



Almost all with
 $\text{Log}(\text{Pos}/\text{Neg}) > 0$:
Pos portion \geq Neg portion
and many with
 $\text{Log}(\text{Pos}/\text{Neg}) > 2$, i.e. 100-fold more positive



Reviewers can be profiled by calculating the geomean of Pos/Neg for each star rating



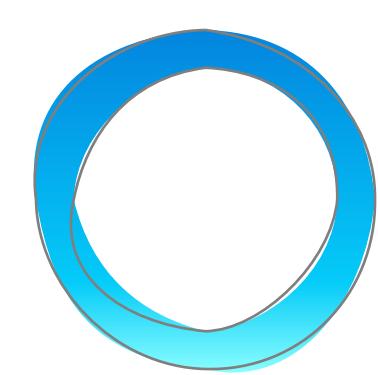
Group A - the more satisfying group:



Group B - the less satisfying group:



The results indicate 2 distinct classes of reviewers, which differ at giving 4-/5-star reviews.



Echo Show 1st Gen vs 2nd Gen

Sound matters



Echo Show 1st Gen

Wordcloud of reviews

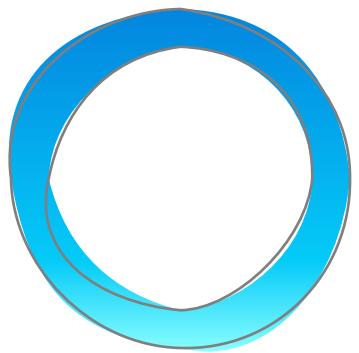
the_sound_quality echo_show_andwith_the_echo on_the_screen drop_in_feature the_sound_is in_the_kitchen the_ability_to_the_screen_i being_able_to love_my_echo sound_quality_i all_the_time listen_to_music the_original_echo ask_alexa_to the_fact_that i_love_the_echo echo_show_it



Echo Show 2nd Gen

Wordcloud of reviews

gen_echo_show on_the_screen the_bigger_screen the_new_echo in_the_kitchen the_sound_is in_my_kitchen this_echo_show listening_to_music in_the_sound sound_quality_i echo_show_2nd better_than_the the_larger_screen new_echo_show much_better_than



The sound quality is important among Echo Show 2nd Gen buyers, just like Echo Plus



Echo Plus 2nd Gen

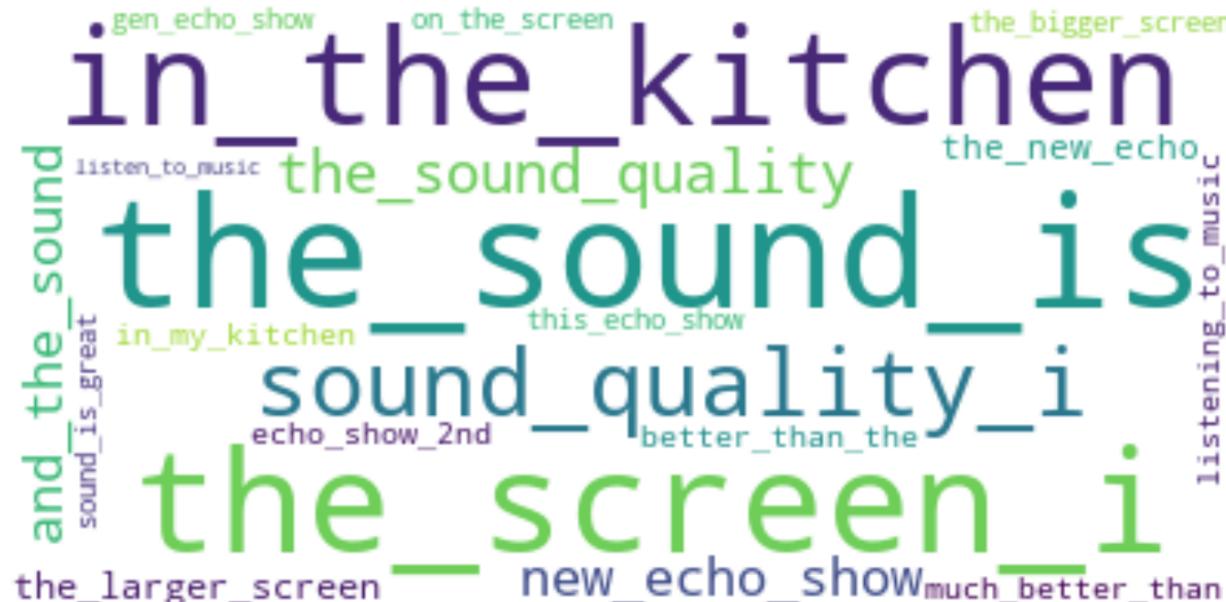
Wordcloud of reviews

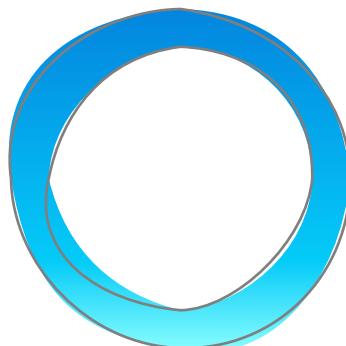


Echo Show 2nd Gen



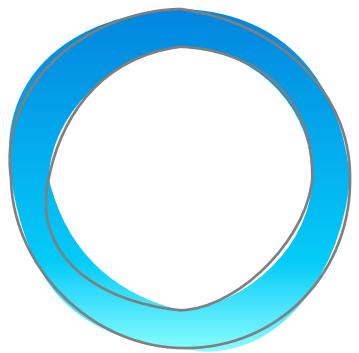
Wordcloud of reviews





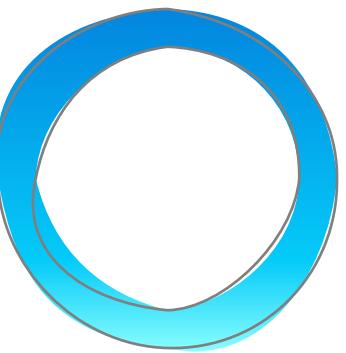
Next Steps

- Improve the sentiment model - replace VADER sentiment tools with a new model based on sentiment polarity ratio (Pos/Neg)
- Improve word cloud analysis – improve bigrams and trigrams word clouds with better clean-up procedures
- Expand analysis to compare products of different generations or from brands



Conclusions

- Examined user experience with Echo devices with and without a screen based on reviews from Amazon
- EDA:
 - The mean ratings are similar among the 5 products examined (the worst being Echo Show Gen 1)
 - The review's length appears to correlate with the mean rating
 - The data is shown to be heavily skewed however with dominantly good reviews
- Word clouds:
 - Identified good features from 5-star reviews and bad features/issues from 1-star reviews
- VADER sentiment analysis:
 - The distributions of the computed sentiment scores appear to make sense, although some are inconsistent with the star ratings
 - The results are difficult to interpret
 - To better understand the reviews, Pos/Neg scoring was developed to specifically examine the good and bad portion of the reviews



Conclusions

- The reviewers for each products were furthered profiled based on the the sentiment polarity ratio (Pos/Neg)
- The product's reviews could be classified into 2, differing at the 4- and 5-star reviews
- While reviewers seem to like the bigger screen in the Echo Show 2nd Gen, the pivoting feature appears to be related to the device's sound quality, very much like the screenless Echo Plus speakers

