**Task:** Amazon has several policies to reduce the bias of its Vine reviews: <https://www.amazon.com/gp/vine/help?ie=UTF8>.

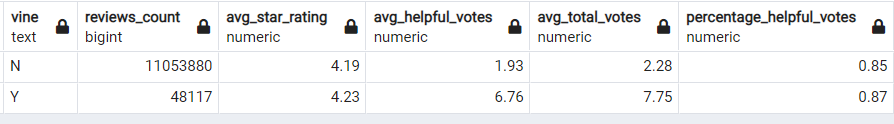
Most notably, Amazon believes that reviewers who provide more detailed and informative reviews are less biased, and that other customers appreciate this & will react with votes of “helpfulness” of the review.

**But are Vine reviews truly trustworthy? Your task is to investigate whether Vine reviews are free of bias.** While there are no hard requirements for the analysis, consider steps you can take to reduce noisy data, e.g., filtering for reviews that meet a certain number of helpful votes, total votes, or both. Submit a summary of your findings and analysis.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Analysis:**

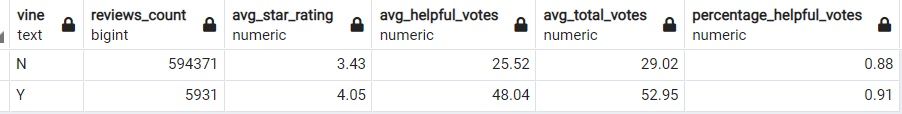
**The first couple of questions I wanted to examine were regarding comparison of reviews aggregated by Vine review or not-Vine review. On average, do products in Vine reviews receive more stars than non-Vine reviews? On average, do Vine reviews receive more total votes? On average, do Vine reviews receive more votes of helpfulness?**



Compared as a whole, Vine reviews do receive more stars on average than non-Vine reviews, but the difference is pretty negligible. Vine reviews do tend to receive significantly more ratings of helpfulness and total votes on helpfulness. This indicates that even though Vine reviews tend to be slightly higher in star rating, they are likely backing up those opinions with more evidence and detail, which is a possible indicator toward lack of bias in the reviews.

**After filtering the data for a certain # of reviewer votes, do these #s change?**

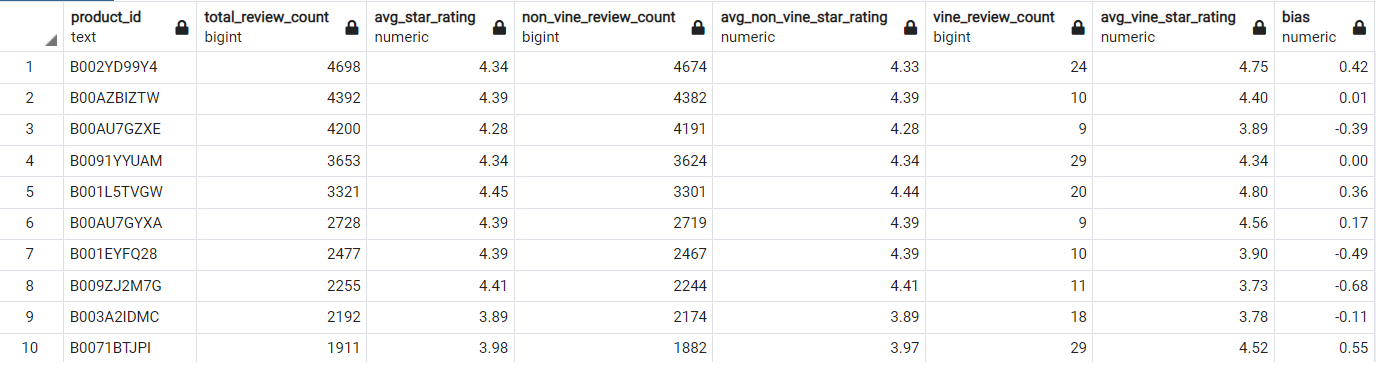
I filtered the table to include only reviews with more than 7 total votes since that was the average # of votes received by Vine reviews. At this point, there does appear to be a significant difference between the average star rating of Vine review vs non-Vine reviews. However, the Vine reviewers are on average still receiving many more helpful votes than non-Vine reviewers, indicating a lack of bias.



An interesting trend to note is the drop in average star rating after filtering for votes of helpfulness of reviews. This suggests to me that perhaps reviewers find it most helpful or interesting when the reviewer mentions flaws or things they felt negatively about the product. Perhaps reviewers who are voting for the reviews are biased to find critical information more helpful than positive information or believe critical/negative information to be more truthful than positive information.

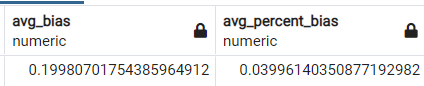
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Next, I felt it would be more useful to look at how vine and non-vine reviewers are rating the same product, as opposed to aggregating the data across all products being reviewed. I filtered the data to include only products with both vine & non-vine reviews. Then, created a column called “bias” indicating on average, how much higher or lower Vine reviews are than the non-Vine reviews for a given product.

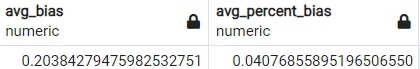


**Within products, how different on avg is the Vine vs non-Vine rating?**

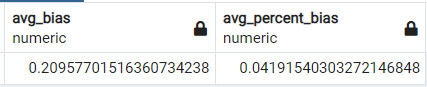
On average, Vine reviews for a product were only about .20 points higher than non-Vine reviews (4%). This trend held true for the most part, even after filtering for # of reviews.



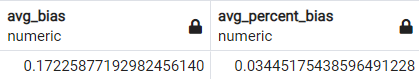
Filtering for 250 total reviews:



Filtering for at least 10 reviews each:



Filtering for at least 20 reviews by both vine and non-vine reviewers (over 450 products) actually drops the average difference in ratings to about .17 points (3.5%):



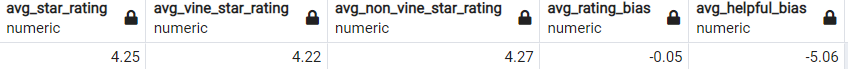
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Finally, I decided to examine how individual reviewers were reviewing products both for the Vine program and of their own volition. I created a new view grouping reviews by customer ID & containing their average review data for vine-and non-vine program reviews. I created “bias” columns to calculate on average, how much higher or lower the Vine reviews are compared to the non-Vine reviews and if they received higher or lower # of helpful reviews. I filtered this table to include only customers who had more than 5 of each type of review.

**Out of the nearly 1000 reviewers leaving both Vine and non-Vine reviews, is there a significant difference in how those reviewers rate products that are part of the program vs personally purchased? Is there a significant difference in how many helpful votes they are getting between reviews?**

The initial data indicates that on average, these reviewers are actually rating products that were part of Vine slightly lower than products that were not part of the Vine program. This could actually indicate a lack of a bias towards giving a higher review due to program participation. It could also indicate that reviewers feel less strongly about these products that they may not have personally chosen for themselves.

Interestingly, these reviewers also received more reviews of helpfulness on the products that they purchased & reviewed of their own volition than on the reviews they left for Vine. This could indicate that perhaps they feel less engaged with these products and are giving their reviews less thought. Or it could also indicate a lack of trust by other customers, who may be wary of believing reviews that were part of the Vine program.



These numbers are interesting in particular when viewed in context of the averages for the dataset as a whole: reviewers that participate in the Vine program tend to leave higher-rated reviews in general, compared to reviewers not a part of the Vine program as a whole: averaging a star rating of 4.27 vs the average 4.19 for non-Vine reviews. This could lend to the appearance of a bias by the program participants, but any bias is more likely a personal bias on the part of reviewers to think and feel more positively about the products or a personal want to leave positive reviews.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**In conclusion**, Vine reviews may have a slight positive-leaning bias when compared to non-Vine reviews. However, this bias is more likely personal bias on the part of the reviewer, not related to any implicit/explicit pressure from the Vine program to leave higher reviews in exchange for program perks. Additionally, the amount that this bias would be swaying a products overall star rating would be negligible.