

## WeRateDogs Twitter Data Analysis

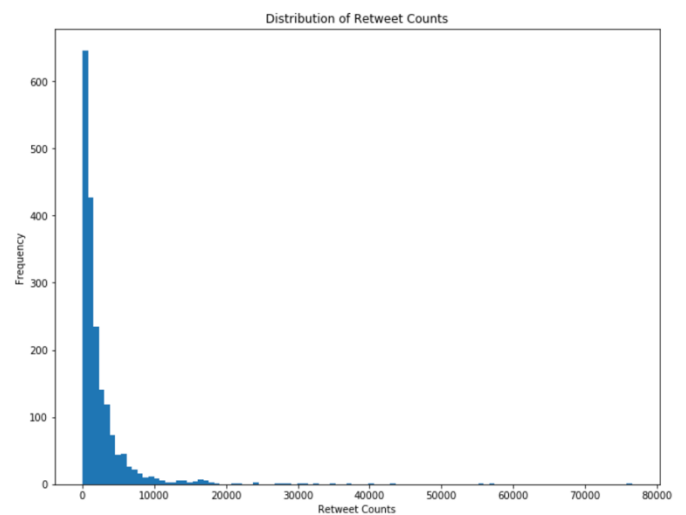
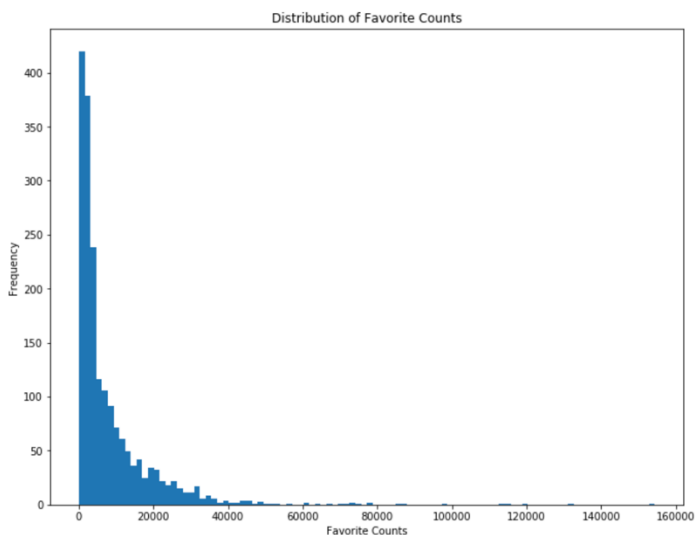
Nick Smith

WeRateDogs is a lighthearted Twitter account that posts cute pictures of dogs sent by followers and gives each an official rating. This analysis focuses on account engagements and official ratings, not amended ratings that are provided in replies to original posts, to assess how popular this account is. The account has received many engagements over time and might provide insight into user interactions with posts based on each dog's rating.

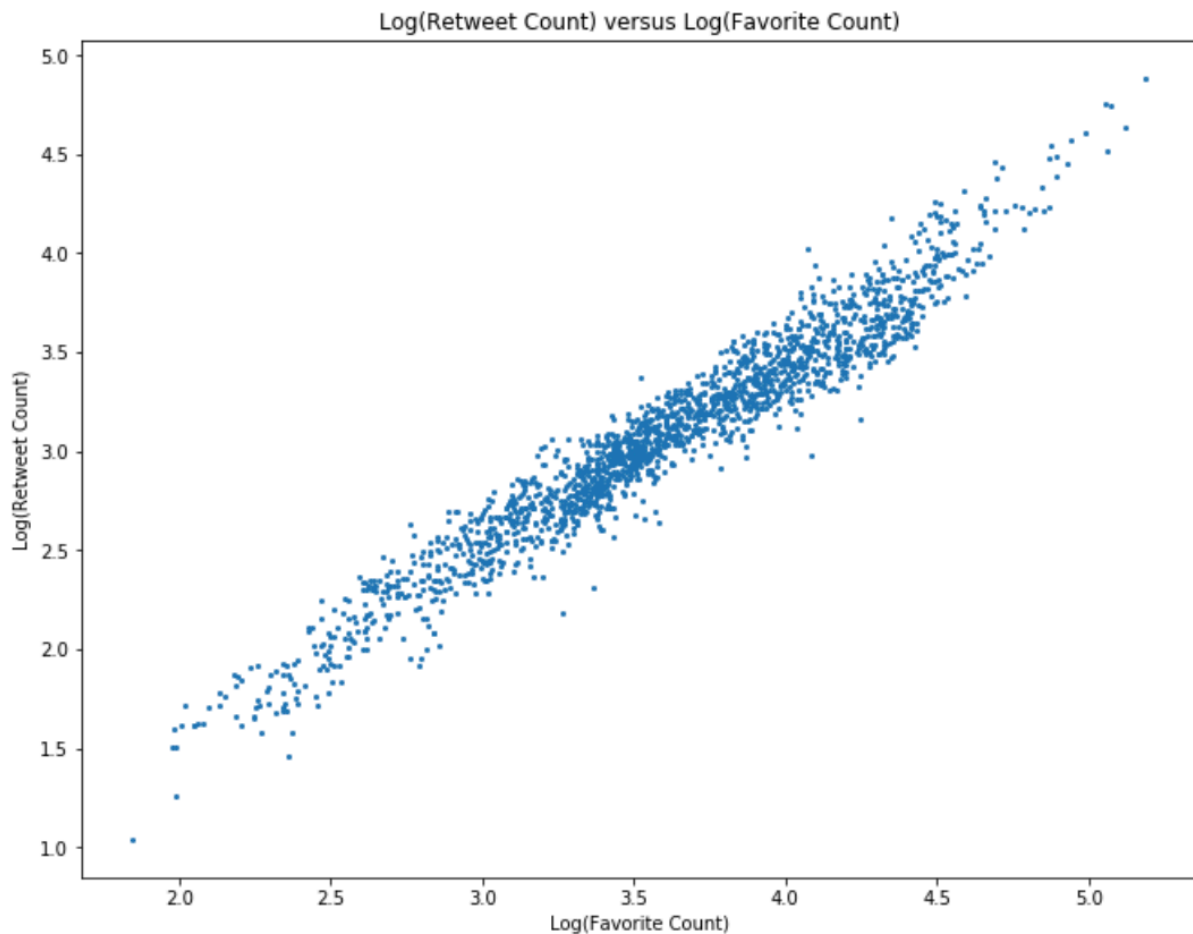
A summary of the variables we analyzed shows that `rating_numerator`, `favorite_count`, and `retweet_count` all have higher means than their medians. Since these are the variables we are most concerned with, it is important to note that these stats show the data is likely skewed.

	rating_numerator	rating_denominator	favorite_count	retweet_count	img_num	p1_conf	p2_conf	p3_conf
count	1881.000000	1881.000000	1881.000000	1881.000000	1881.000000	1881.000000	1.881000e+03	1.881000e+03
mean	12.557682	10.500266	8481.904306	2495.888357	1.211058	0.596755	1.346438e-01	6.035768e-02
std	42.586195	7.014584	12293.757213	4429.070301	0.570186	0.271116	1.007538e-01	5.110110e-02
min	0.000000	2.000000	70.000000	11.000000	1.000000	0.044333	1.011300e-08	1.740170e-10
25%	10.000000	10.000000	1869.000000	550.000000	1.000000	0.367368	5.351500e-02	1.619920e-02
50%	11.000000	10.000000	3825.000000	1216.000000	1.000000	0.594333	1.175660e-01	4.923690e-02
75%	12.000000	10.000000	10638.000000	2848.000000	1.000000	0.848514	1.954050e-01	9.272700e-02
max	1776.000000	170.000000	154293.000000	76564.000000	4.000000	1.000000	4.880140e-01	2.734190e-01

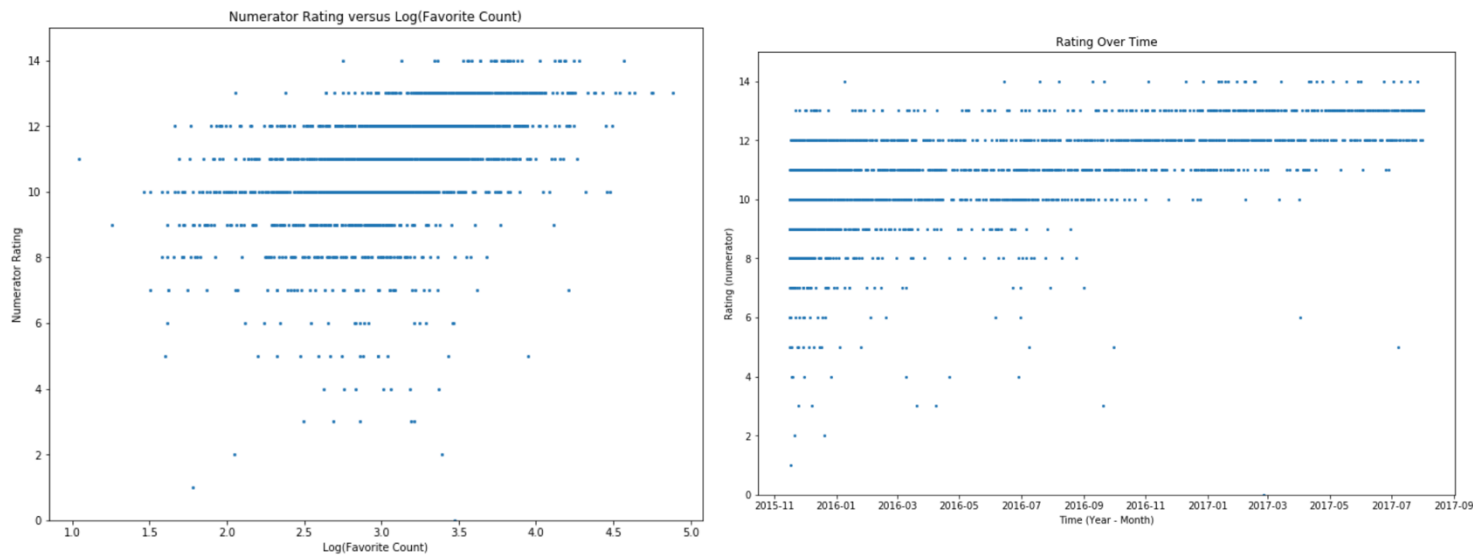
After graphing this data for retweet counts and we can see that this is true, and we can better understand what this data really shows us if we graph the same data using log transformations. We can likely expect posts that were retweeted to also have high favorite ratings. This would be fairly typical of a successful Twitter account, particularly one that serves to be a form of entertainment.



If we plot the tweets with high favorite and high retweet counts, we see that they are very strongly correlated. And when we run a correlation analysis, it shows these values to be correlated at 97%. This would indicate that not only does the account receive a lot of engagements, but those engagements result in relatively high levels of the account's content being redistributed through other users.

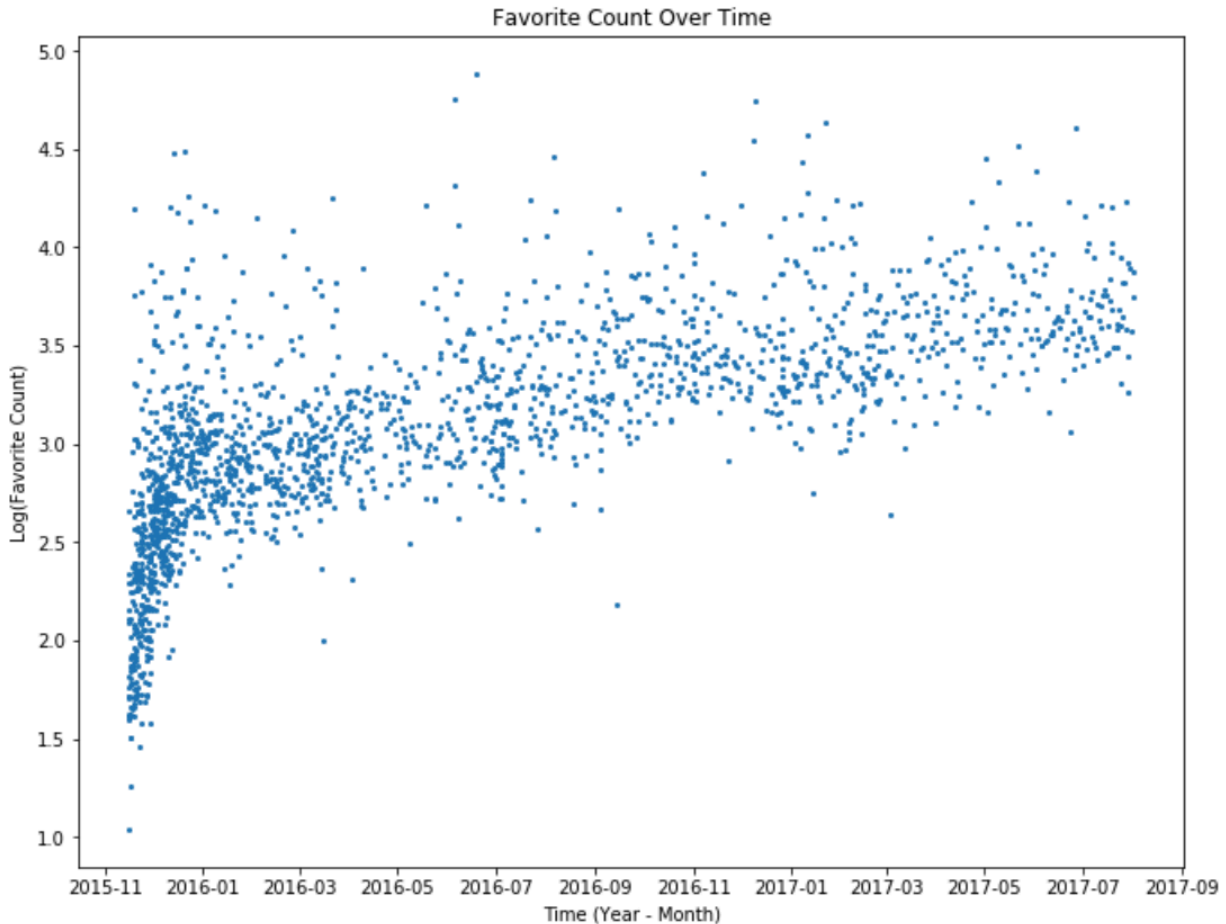


Another interesting metric we can look into that could provide clarity on the account's popularity is its rating system. The account rates dogs on a scale of 1-10; however, there are a significant portion of ratings that go as high as 14. If we compare the ratings included in a tweet to that tweet's favorite count, we may be able to see if this system is one of the reasons people enjoy the content of the account. It would also be helpful to see how the rating scores changed over time as well. We can see these trends in the graphs below:

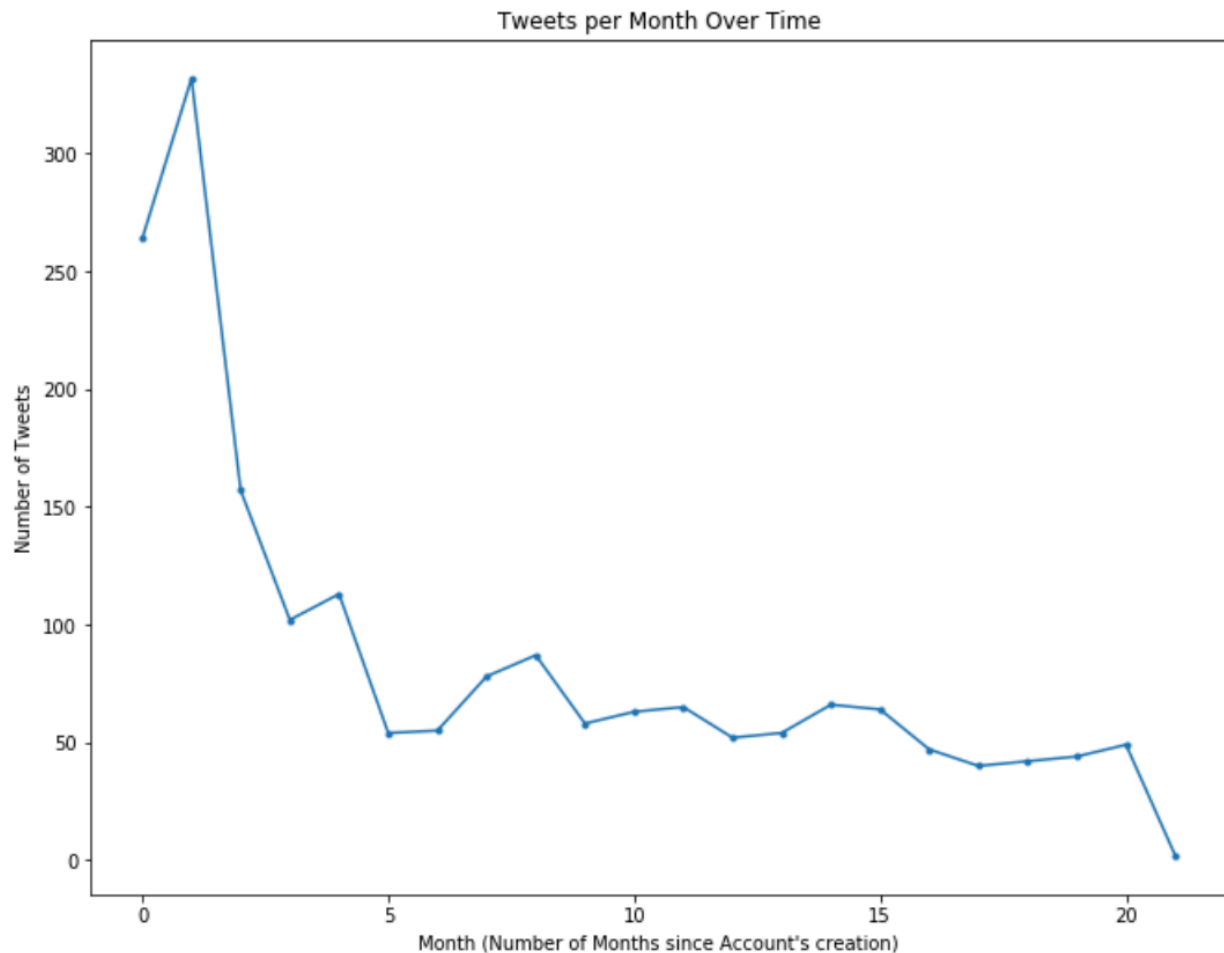


It appears that higher dog ratings do correlate to higher scores for the dogs. The graph even shows that the tweets with ratings above the 1-10 scale usually receive even more favorites than the tweets below the 10-point threshold. When we look at the ratings over time, we also see that there was almost a clean shift from to only 10+ rating around August 2019. This might indicate that the abnormal rating system is part of the appeal of the account and that the owner of the account realized that people enjoyed the hyperbolic ratings more.

Now that we have examined two different way of looking at the account's popularity, it might make sense to see if that account maintained its popularity over time. We can look at this by plotting the accounts favorite counts for the time period included in our data (Nov. 2015 – Aug 2017):



It would seem that the account's favorite counts over time saw a steady increase likely indicating that the account gained increasingly popular. One way to check and see if that is true though would be to see the account's overall activity by plotting its tweet numbers for the same time period (measured in 5-month increments):



Interestingly, the account's activity dropped significantly after its creation. There could have been a recovery in the years following those we have the data for, but it's hard to say without acquiring more data.

So can we determine whether or not the account brought significant joy to peoples' lives and success and fame to the owner of the account? I think both could be true at least for a period of time. We can see that the account's content continued becoming popular even though the amount the account posted significantly decreased. People continued engaging with the account content, with increasingly higher likes, especially after the rating system became more absurd and fun. It would be interesting to see if additional data might provide some insight into what may have caused this change.