# Analyzing WeRateDogs Tweets

*By Moh A T Elgaser*
*Date: Feb 28th, 2018*

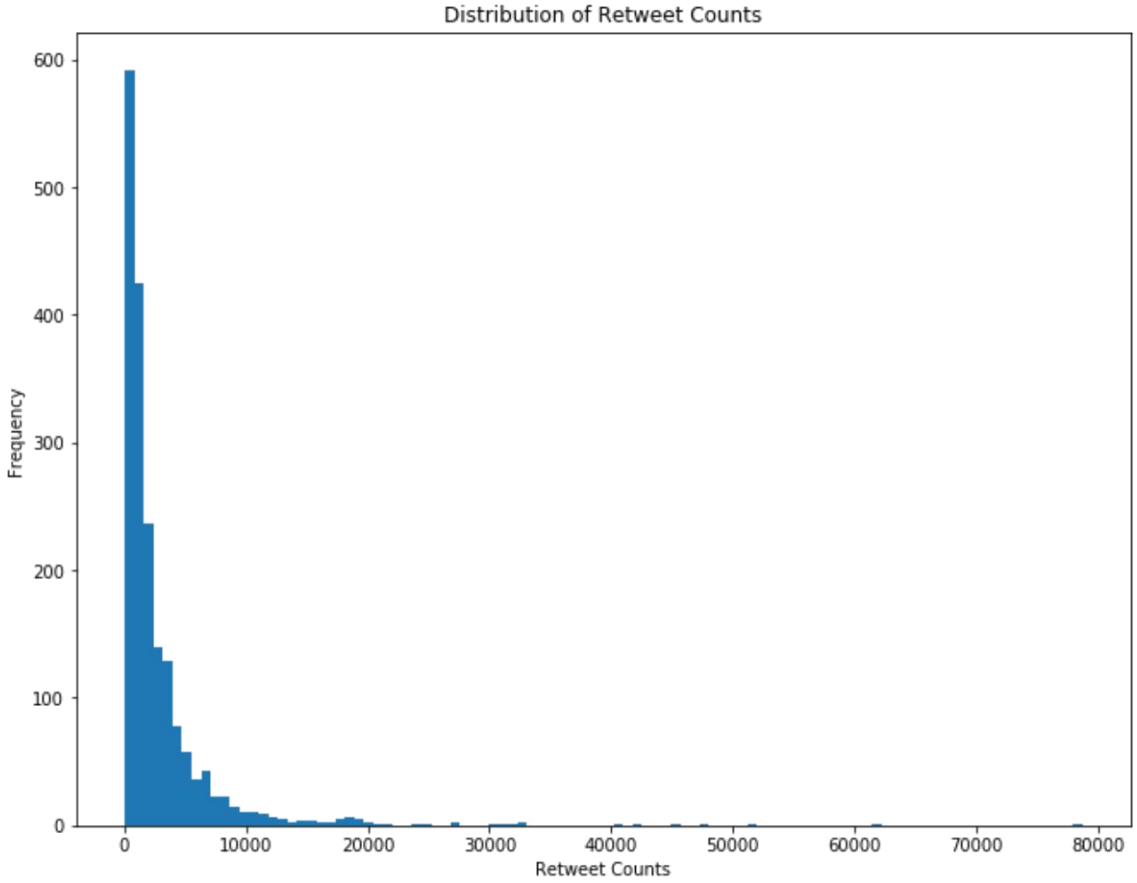
This is a brief analysis of ratings done by WeRateDogs. In case you didn't know, WeRateDogs™ is a twitter account that **professionally**rates dogs. Pictured below is their twitter bio.

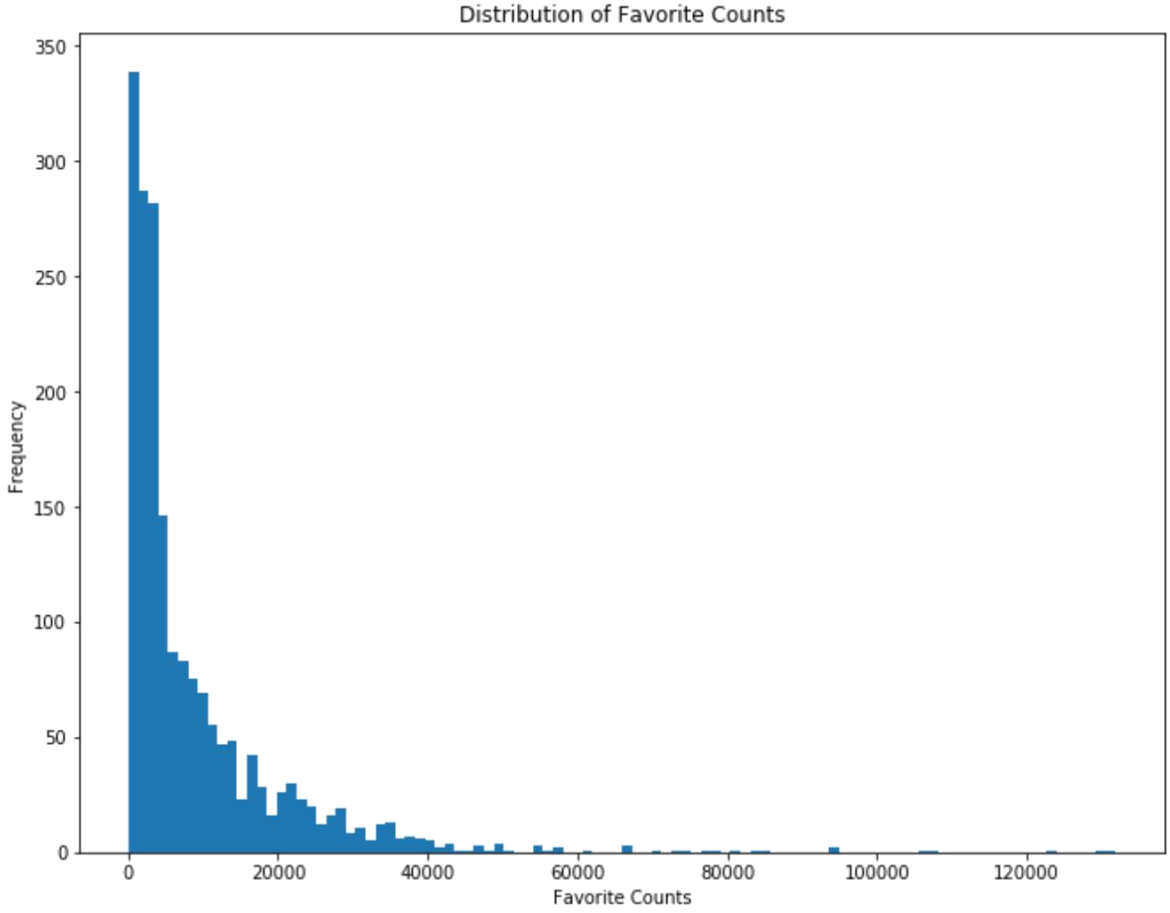

*For the sake of this analysis, official ratings were defined as pictures of dogs that were direct messaged to the account. As well as the first rating, any tweets that WeRateDogs™ had replied to in order to change the rating were discarded.*

At the start of any analysis it is important to get some basic summary statistics of our numerical variables.

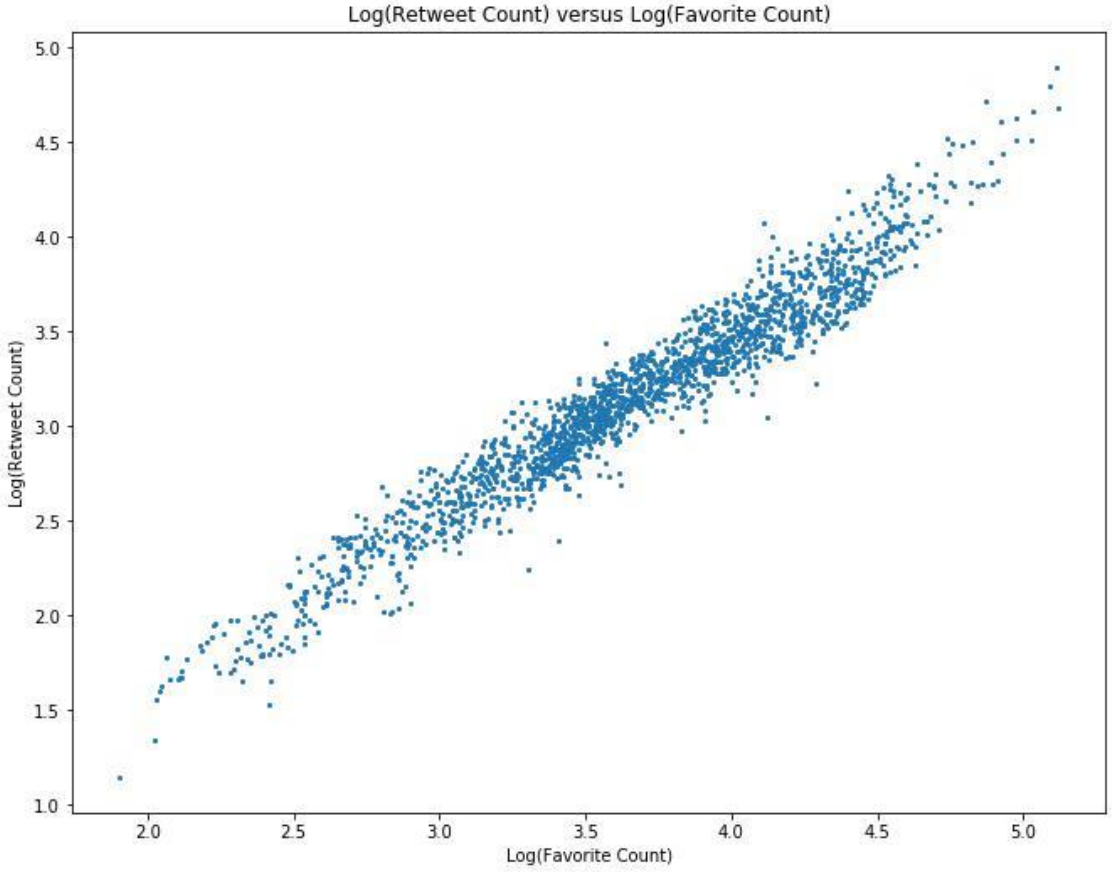
| | rating_numerator | rating_denominator | favorite_count | retweet_count | img_num | p1_conf | p2_conf | p3_conf |
|---|---|---|---|---|---|---|---|---|
| count | 1888.000000 | 1888.000000 | 1888.000000 | 1888.000000 | 1888.000000 | 1888.000000 | 1.888000e+03 | 1.888000e+03 |
| mean | 12.556674 | 10.498411 | 9234.475636 | 2847.300847 | 1.210805 | 0.597402 | 1.345871e-01 | 6.022818e-02 |
| std | 42.507183 | 7.001627 | 12707.559172 | 4841.181672 | 0.569541 | 0.271105 | 1.008693e-01 | 5.107432e-02 |
| min | 0.000000 | 2.000000 | 80.000000 | 14.000000 | 1.000000 | 0.044333 | 1.011300e-08 | 1.740170e-10 |
| 25% | 10.000000 | 10.000000 | 2085.750000 | 632.000000 | 1.000000 | 0.367736 | 5.321280e-02 | 1.609288e-02 |
| 50% | 11.000000 | 10.000000 | 4250.500000 | 1406.500000 | 1.000000 | 0.595183 | 1.175370e-01 | 4.919990e-02 |
| 75% | 12.000000 | 10.000000 | 11675.250000 | 3275.500000 | 1.000000 | 0.848712 | 1.954432e-01 | 9.254258e-02 |
| max | 1776.000000 | 170.000000 | 131844.000000 | 78763.000000 | 4.000000 | 1.000000 | 4.880140e-01 | 2.734190e-01 |

The variables rating numerator, retweet count, and favorite count all have greater means than medians. This hints that distribution is skewed. Favorite counts and retweets counts have means that are double that of their medians. The distributions of these are pictured below.


Distribution of Retweet Counts
Frequency
Retweet Counts
0
100
200
300
400
500
600
0
10000
20000
30000
40000
50000
60000
70000
80000

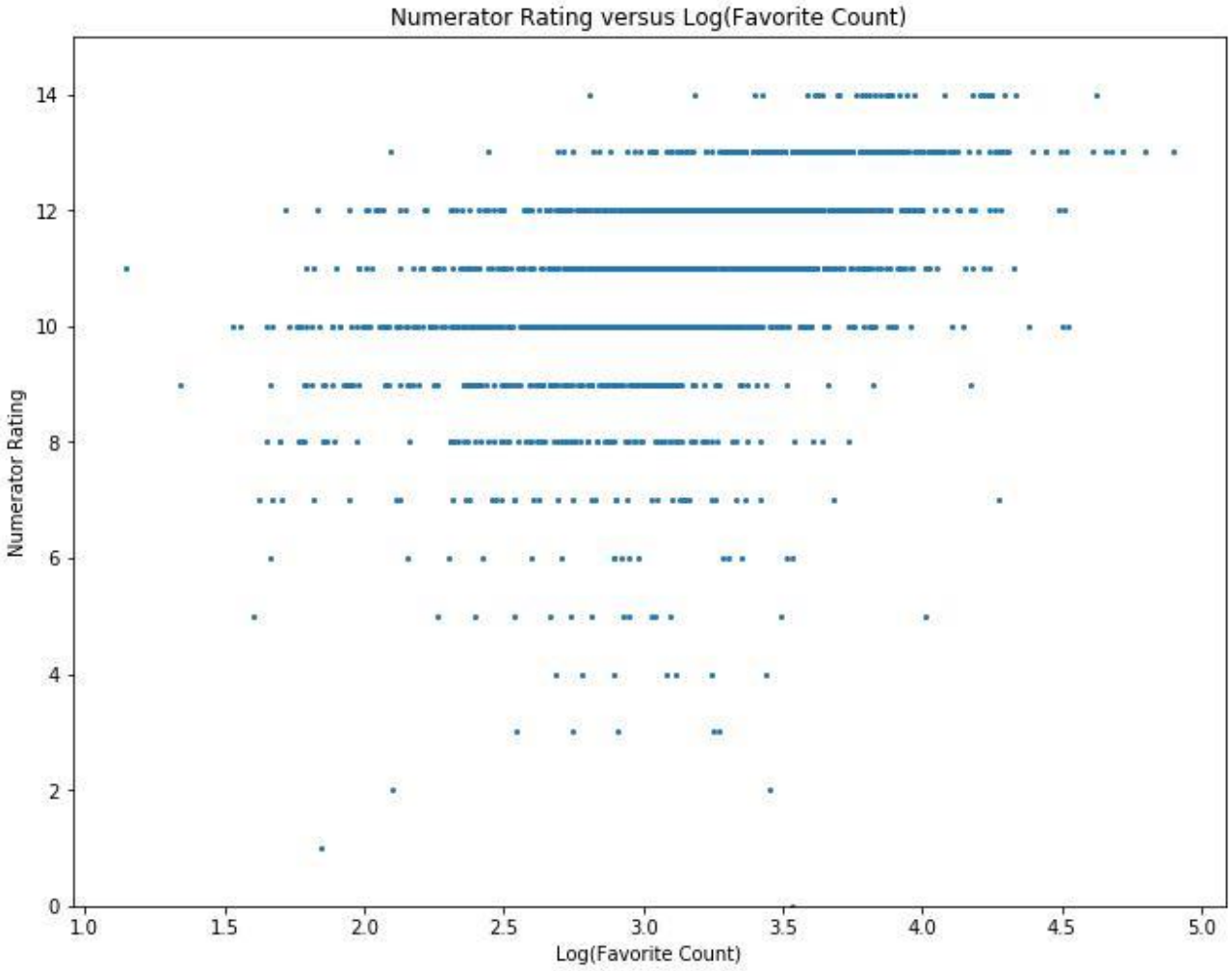

Distribution of Favorite Counts
Frequency
Favorite Counts
0
50
100
150
200
250
300
350
0
20000
40000
60000
80000
100000
120000

After doing a log transformation to account for these distributions skewness, we can see how these two distributions. I expect them to correlated pretty strongly. A scatterplot of the log transformation of both of these variables is pictured below.



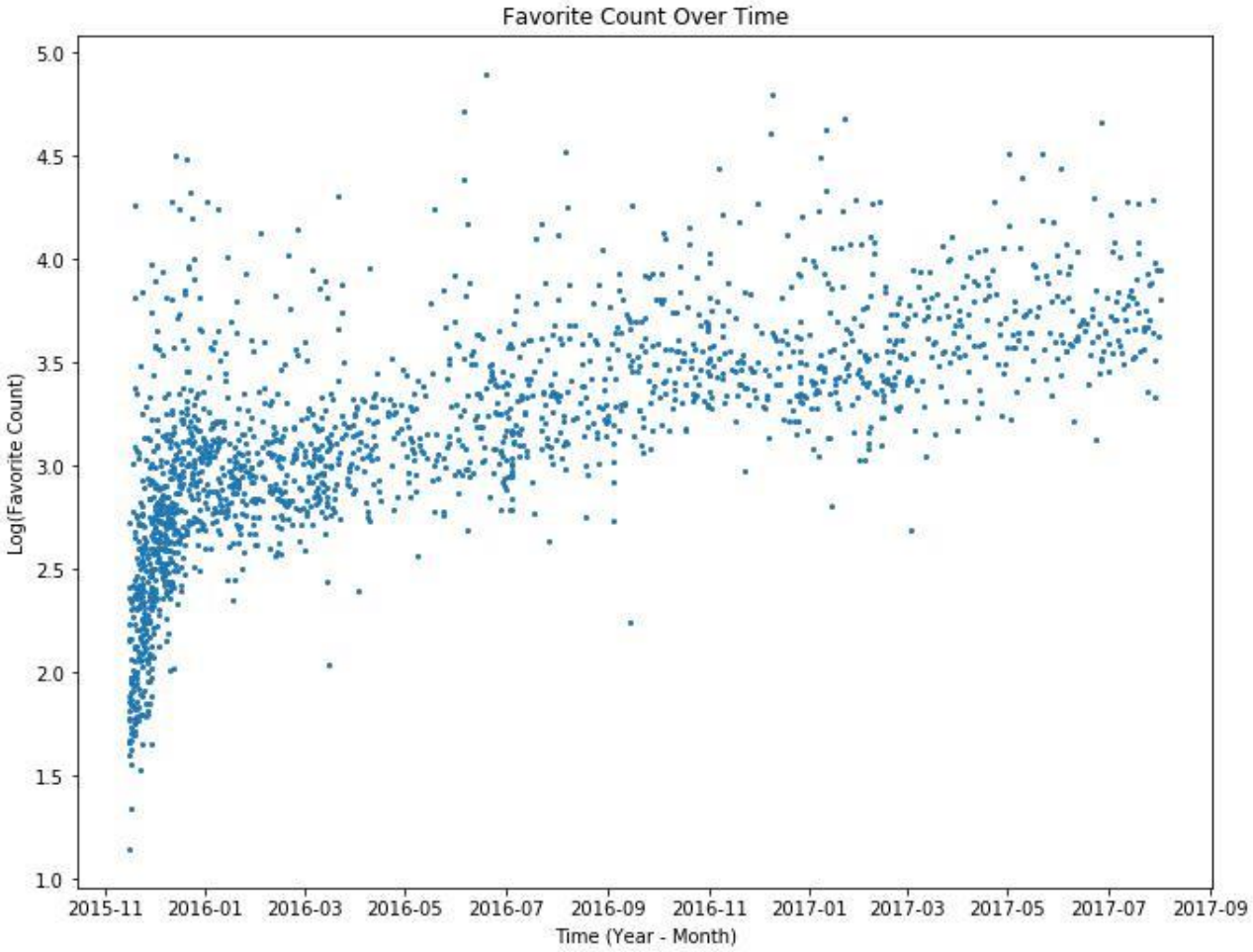

I wouldn't say that it is *super*surprising that these counts correlate so strongly. I would say that a correlation coefficient of 0.97 is *impressive*.This statistic makes me wonder if this correlation is similar for all users. Perhaps, different accounts have different correlations and slopes for these metrics.

A important factor of WeRateDogs is their unique rating system, the bulk of the ratings (numerator rating) are between 7 and 14. Anything outside of the range weren't taken into account for this next plot. Below is a plot of how the numerator rating relates to the tweets favorite count.



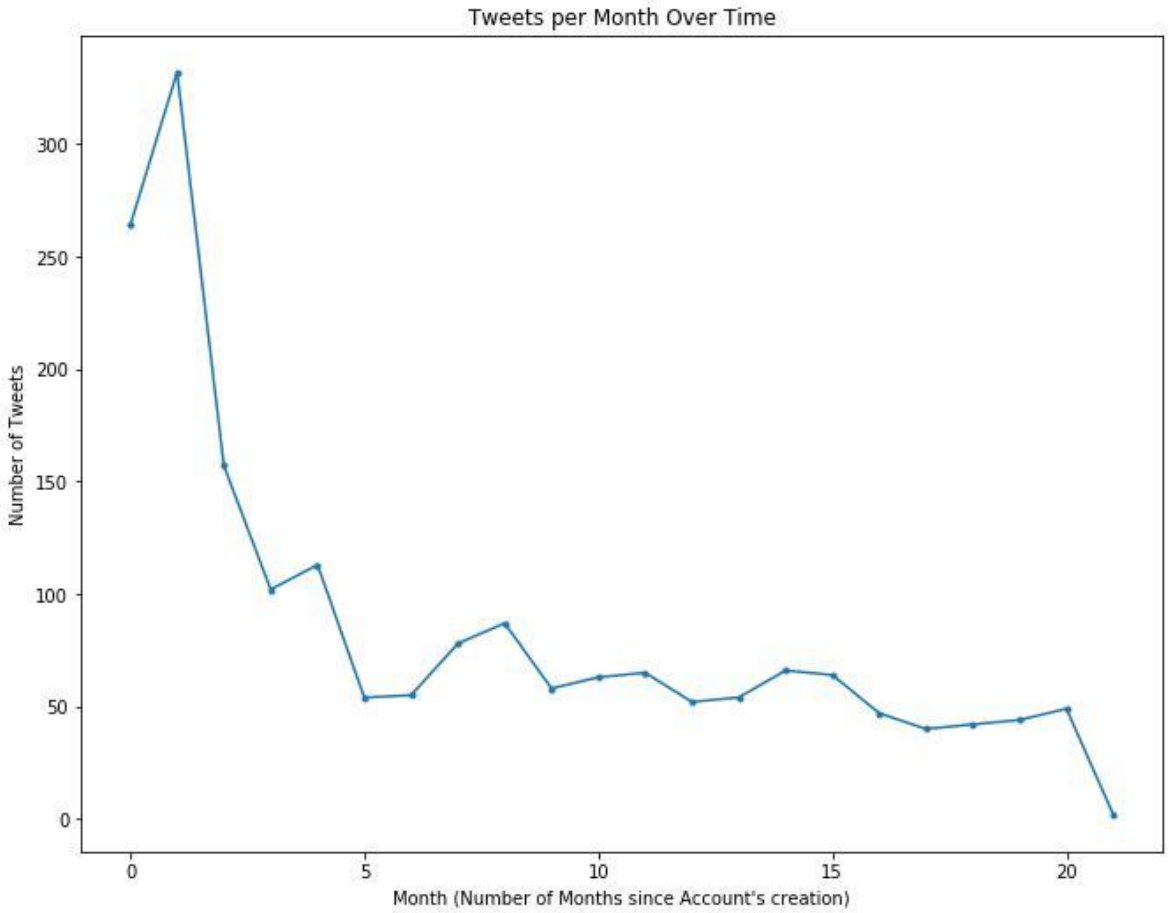

This shows that tweets that have lower ratings tend to not have as many likes. Perhaps this is distorted by the fact that some tweets that have lower favorite counts were tweets in the early days of @dog_rates.

Let's look into how the favorite count of tweets changes with time, to measure look at dog_rates popularity over time.



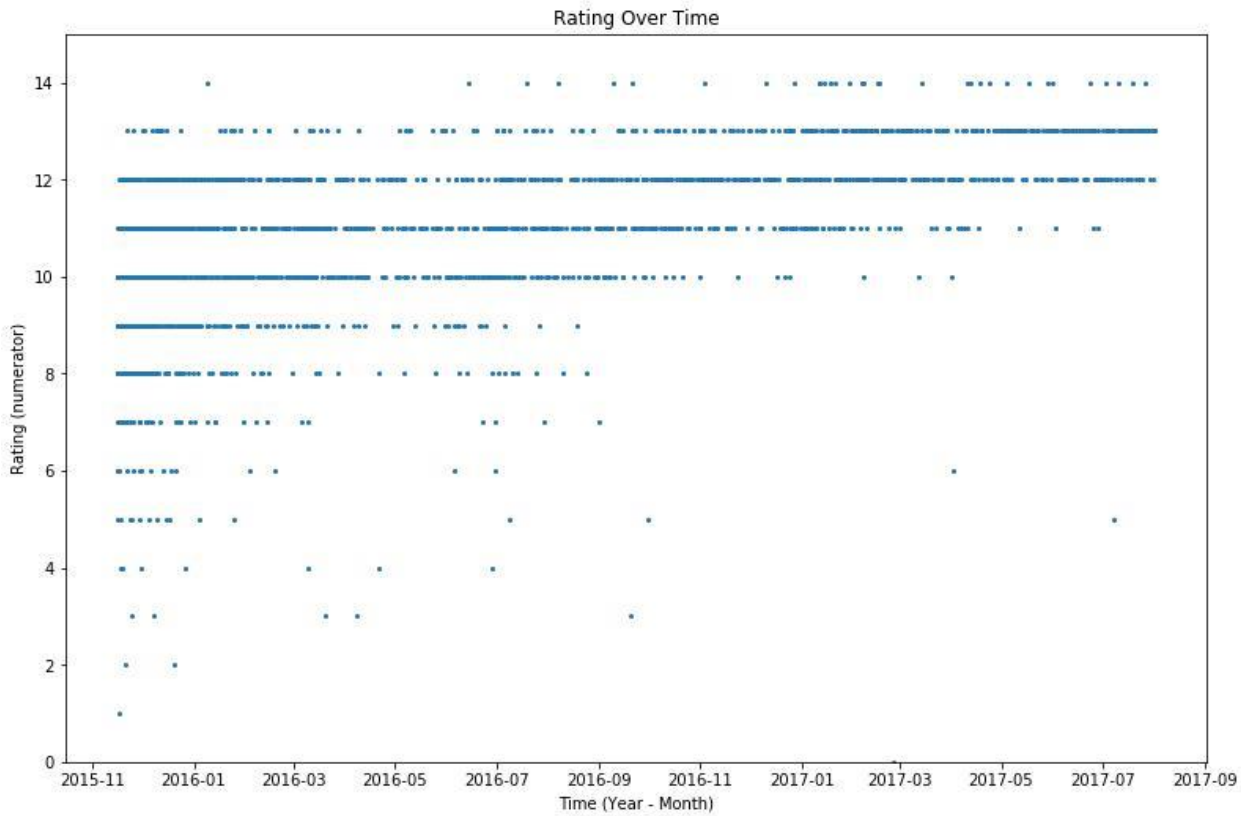

In the above graph, we can see that @dog_rates started gaining followers and popularity quickly from the start. It appears that there is a intense growth of favorite count in the first three or four months. While after those three or four months, the favorite growth rate is still positive, but not as great. It also appears that dog_rates started tweeting less, after the second month.

Let's look at how many tweets per month have been tweeted for all of dog_rates existence.




We can definitely see that the dog_rates tweeting rate has decreased since its creation. The dip towards the end is because the dataset only contains tweets from the first of that month (August 1st 2017). The slight increase around 7.5 months since creation corresponds to the summer months. This hints that the person running the account is probably a student.

Below is how the rating system (numerator rating) has changed over time.




We can see that over time that the amount of ratings below 10 has decreased. The*"they're good dogs Brent"* meme was tweeted out on September 12th 2016, and we can see that after September 2016 there have only been two tweets that have with ratings less than 10. So it looks like WeRateDogs™ purposely rated dogs 10+ to spite Brent. Below is a picture of the meme I am referring to.

Unfortunately, I cannot delve deeper into this analysis, but there are some things that I would like to do if I had more time.

Looking into if the ratings and favorite counts are higher for certain dog breeds according to the neural network. Perhaps the person running the account is biased towards a certain breed.

Due to my cleaning tasks, I realize that there are quite a few pictures of dogs that the neural network didn't think were dogs. In most of these cases, the dog wasn't posing, and the dog was doing something, that confused the neural network. I think it would interesting to see if dogs that were doing something silly/unconventional (for a dog) got higher ratings or favorite counts.