We Rate Dogs Analysis

September 23, 2018

1 We Rate Dogs Tweet Analysis

This document is a brief summary of ratings done by WeRateDogs, a twitter account that rates dogs proessionally.

To start off lets get a basic statistical summary of the data and the numerical variables. It;s important to note that only official ratings were taken into account for this analysis, any retweets or replies were not taken into consideration.

	rating_numerator	rating_denominator	favorite_count	retweet_count \
count	1883.000000	1883.000000	1883.000000	1883.000000
mean	12.568773	10.499734	9134.133829	2774.637281
std	42.562836	7.010875	13088.984610	4876.332325
min	0.000000	2.000000	80.000000	12.000000
25%	10.000000	10.000000	2027.000000	607.500000
50%	11.000000	10.000000	4184.000000	1358.000000
75%	12.000000	10.000000	11491.000000	3168.500000
max	1776.000000	170.000000	164839.000000	84529.000000

The variable counts for rating numerator, retweets and favorites have greater means than medians. This suggests the distribution is skewed. A visualization of this distribution is provided below.

Distribution of Favorite Counts

Distribution of Retweet Counts
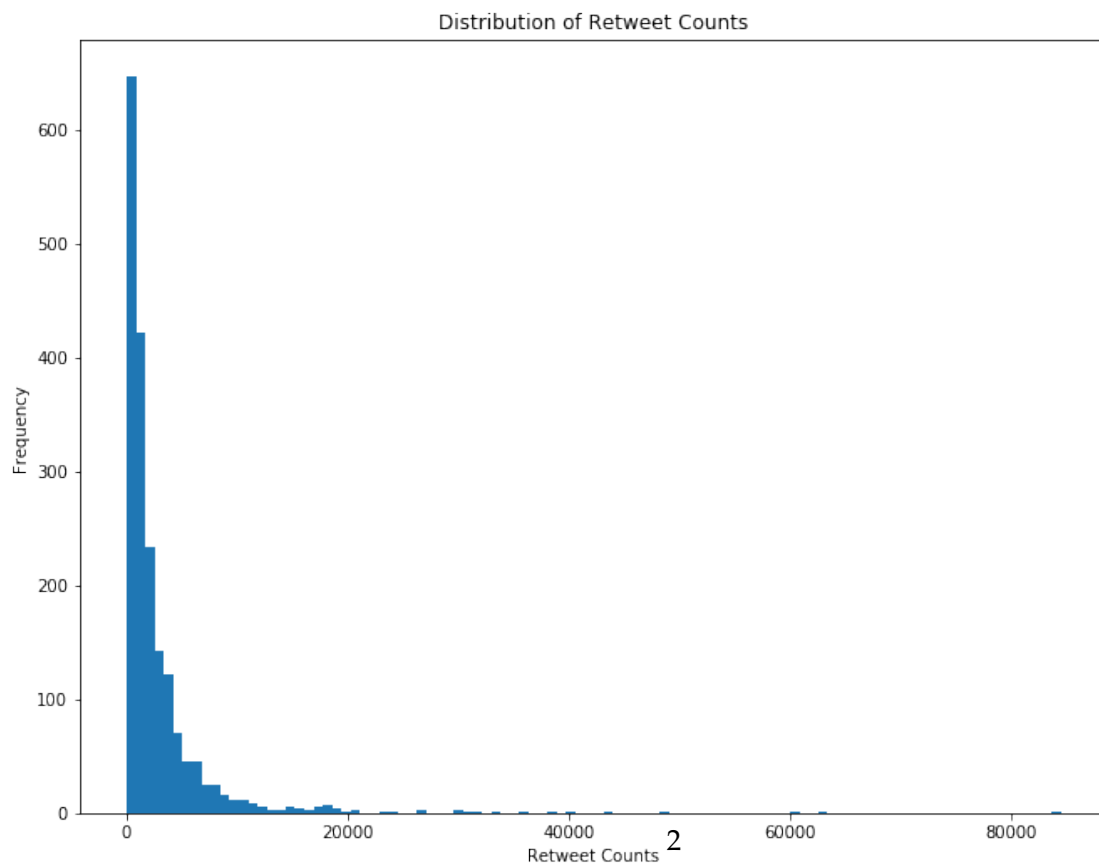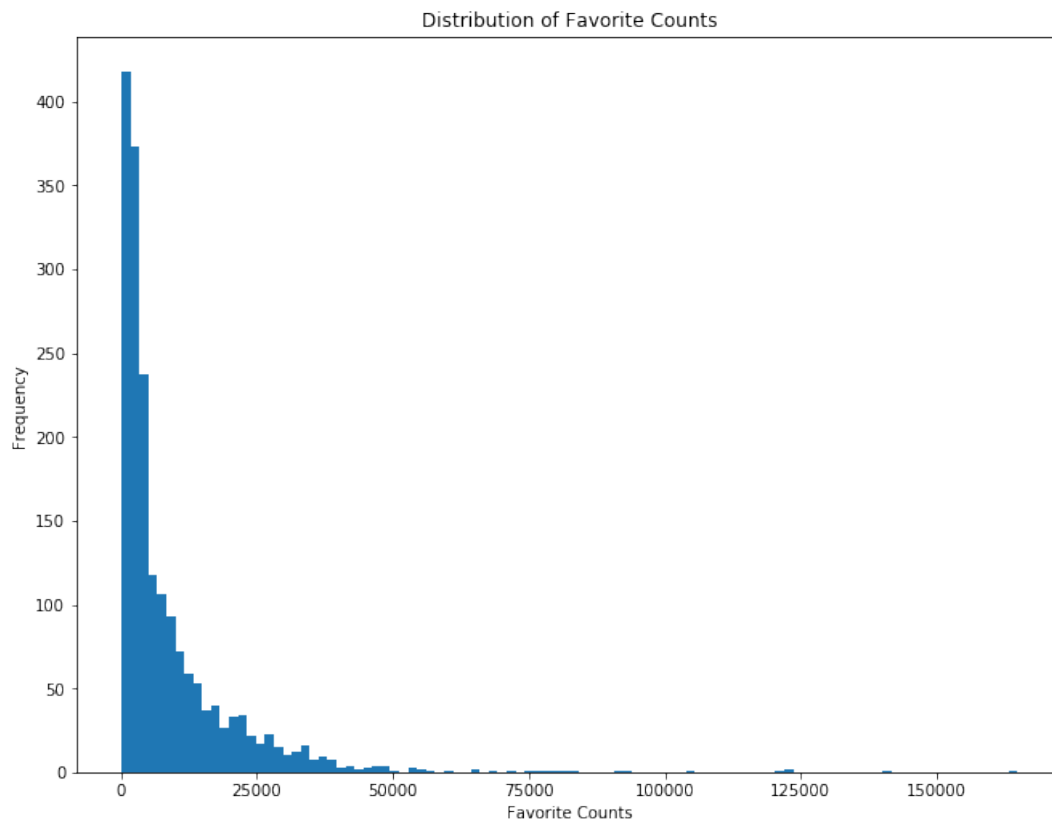
By doing a log transformation, to normalize any skew, we can observe the correlation between the two distributions. We would expect a pretty strong correlation. The log transformation is shown below.

Log(Retweet Count) versus Log(Favorite Count)
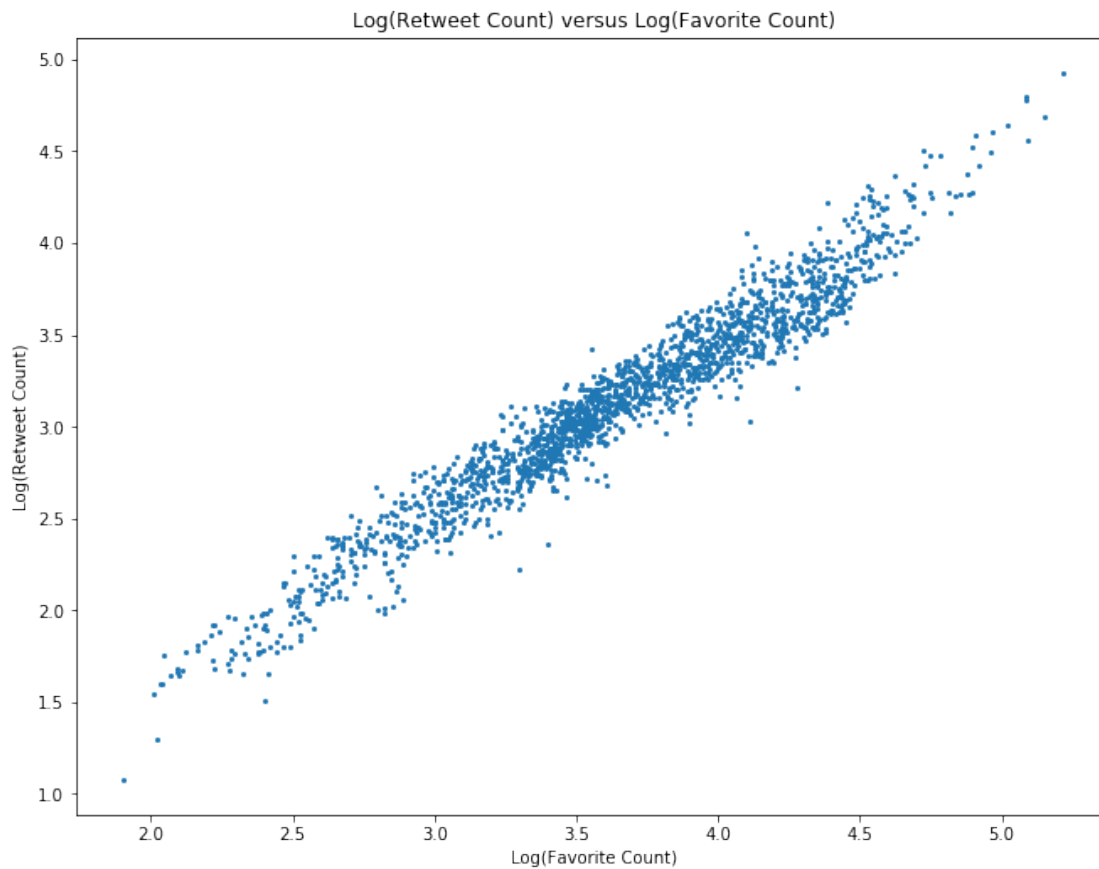
While a strong correlation is not surprising, 0.97 is remarkably high. It makes me wonder if the correlation is similar for all users or just the particular subset.

WeRateDogs had a fairly unique rating system, ranking dogs out of 20. Because The vast majority of ratings were between 7 and 14, any rankings outside of this range was removed for expediency. A plot of the relationship between rating and favorite counts is shown below.

Numerator Rating versus Log(Favorite Count)
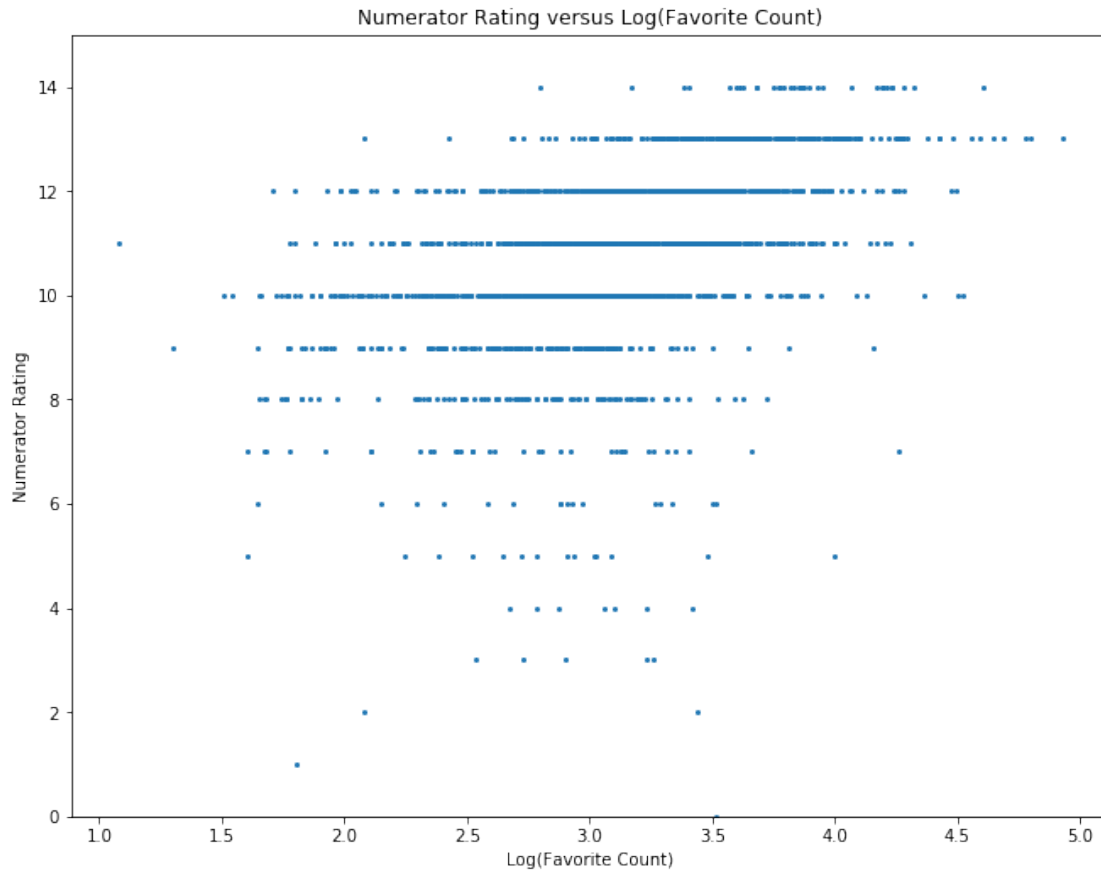
This plot shows tweets with lower ratings have a tendancy to not recieve as many likes. This could be because many of the tweets with lower favorite counts were during the early phase of @dog_rates.

Lastly, let's take a look at how many tweets @dog_rates posted per month.
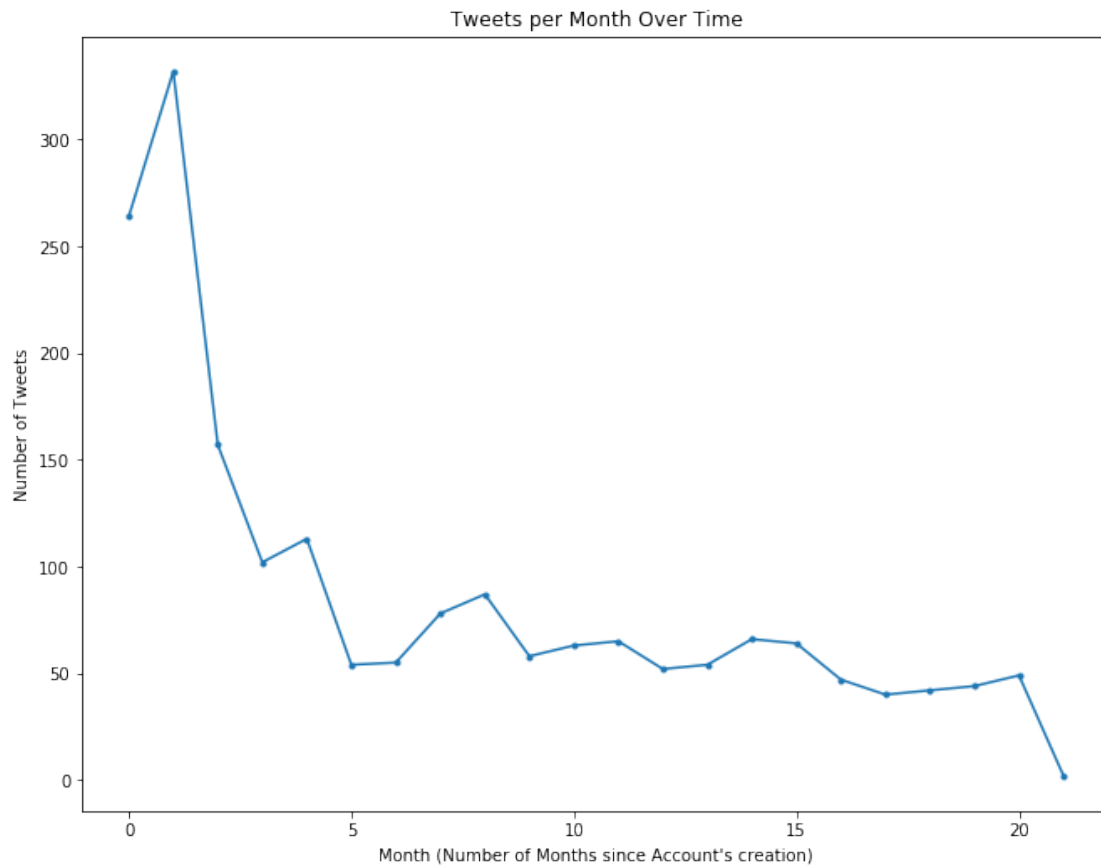
The graph clearly illustrate that @dog_rates post rate has declined significantly since it's creation. The slight dip at the end is less dramatic than it seeems, since the data for August only includes the first of the month. The increase around 7.5 months corrresponds with the summer months. Hinting the person running the twitter account was a student.