

Final Analysis o The WeRateDogs Twitter Account

Obed Makori

makoriobed95@mail.com

[Twitter](#)

[LinkedIn](#)

Final Analysis o The WeRateDogs Twitter Account

On successfull data wrangling process, a clead dataset was obtained reflecting scraped tweets from WeRateDogs Twitter Account. [WeRateDogs](#) is a Twitter account that rates people's dogs. The account was created in 2015 and has since amassed millions of followers. [WeRateDogs](#) is known for its humorous and often irreverent takes on dogs and their owners.

Results and discussion

I found that the distribution of favourite tweet counts lies between 0 and 10, 000. In addition, we can observe rom the histogram below that high tweets with high favourite counts (above 20000) are very minimal.

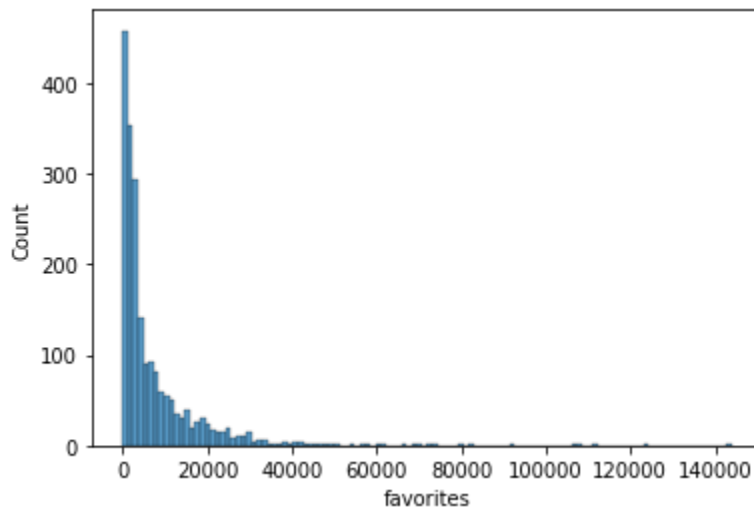


Fig 1: Distribution of Favourite counts

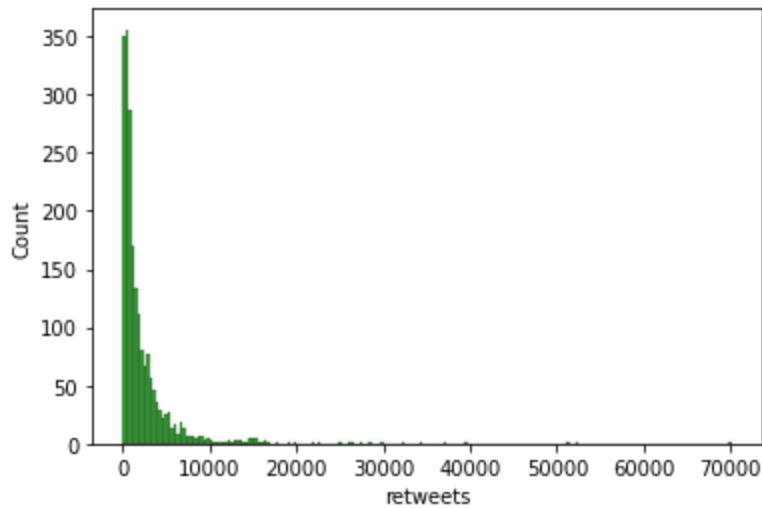


Fig 2: Distribution o Retweets

There is a positive correlation between the favourite count and retweets, meaning that tweets with higher favorite counts tend to also have higher retweet counts. This makes sense from a logical standpoint, as users who enjoy a tweet enough to favorite it are likely to also share it with their followers.

There are a few potential reasons for why this relationship exists. First, users who favorite a tweet are likely to be engaged with the content, and therefore more likely to retweet it. Second, favoriting a tweet is a public way of showing support for the author or message, which may encourage others to do the same.

Finally, users who favorite a tweet may be more likely to be influential in their social networks, and therefore their retweet can carry more weight. Of course, there are always exceptions to the rule and there are many factors that can influence whether or not a tweet is favorited or retweeted. However, in general, tweets with higher favorite counts tend to also have higher retweet counts.

	tweet_id	rating_numerator	retweets	favorites	img_num	first_confidence	first_dog	second_confidence	second_dog	third_confidence	third_dog
tweet_id	1.000000	0.024922	0.385572	0.612638	0.215664	0.102470	0.119945	-0.000821	0.124802	-0.044799	0.096319
rating_numerator	0.024922	1.000000	0.019155	0.016807	-0.003165	-0.006999	-0.030792	-0.019751	-0.036241	-0.004223	-0.030879
retweets	0.385572	0.019155	1.000000	0.928516	0.105084	0.047299	0.000895	-0.014108	0.012674	-0.038377	0.005517
favorites	0.612638	0.016807	0.928516	1.000000	0.130846	0.069632	0.042144	-0.016254	0.052219	-0.046275	0.033828
img_num	0.215664	-0.003165	0.105084	0.130846	1.000000	0.202714	0.032714	-0.156567	0.056999	-0.137806	0.059547
first_confidence	0.102470	-0.006999	0.047299	0.069632	0.202714	1.000000	0.127061	-0.511022	0.142213	-0.707314	0.118888
first_dog	0.119945	-0.030792	0.000895	0.042144	0.032714	0.127061	1.000000	0.110839	0.640160	0.049916	0.560674
second_confidence	-0.000821	-0.019751	-0.014108	-0.016254	-0.156567	-0.511022	0.110839	1.000000	0.096856	0.481933	0.064919
second_dog	0.124802	-0.036241	0.012674	0.052219	0.056999	0.142213	0.640160	0.096856	1.000000	0.033872	0.559345
third_confidence	-0.044799	-0.004223	-0.038377	-0.046275	-0.137806	-0.707314	0.049916	0.481933	0.033872	1.000000	0.035461
third_dog	0.096319	-0.030879	0.005517	0.033828	0.059547	0.118888	0.560674	0.064919	0.559345	0.035461	1.000000

Conclusions and Recommendations

Both the favorites and retweets original distributions had protracted positive tails. Tweets that go viral are relatively uncommon. The bottom two normalized graphs display similar distributions once more. With the exception of the surge in data for the normalized favorite count, they are essentially normal. That might be a result of my jitter work rather than a true characteristic of the data. Or perhaps a large number of tweets have a single liked count.

The number of favorites and retweets are strongly correlated (0.9285). Retweets should rise as a tweet receives more Favorites, and vice versa. The relationship appears to be nonlinear.