

Act report

Out[89]:

	tweet_id	retweet_count	favorite_count	score	img_num	confidence
count	1.638000e+03	1638.000000	1638.000000	1638.000000	1638.000000	1638.000000
mean	7.399139e+17	2421.684371	8408.680708	10.893966	1.215507	0.550225
std	6.803793e+16	4380.829674	12298.723025	2.441771	0.576958	0.299202
min	6.660209e+17	11.000000	69.000000	0.000000	1.000000	0.000010
25%	6.776988e+17	547.500000	1951.000000	10.000000	1.000000	0.303879
50%	7.149960e+17	1208.000000	3944.500000	11.000000	1.000000	0.550604
75%	7.932375e+17	2729.500000	10390.750000	12.000000	1.000000	0.821962
max	8.921774e+17	74012.000000	150021.000000	75.000000	4.000000	0.999956

Each variable initially seems to follow a logical pattern without abnormal outliers sending red flags. For example, the rate_denom the minimum is 7 which makes sense because 0 cannot be the lowest since it's part of a ratio and you cannot divide by zero.

Even, though some of the numerators in the rate_num column are unnecessarily high and above 10 for the denominator, the rating system for the WeRateDogs feed seems to follow a random rating system. Therefore, this should not draw attention as an issue.

Also, for the confidence column, the numbers are in the bounds from 0 to 1, which is good because they are confidence intervals and are bound between 0 and 1.

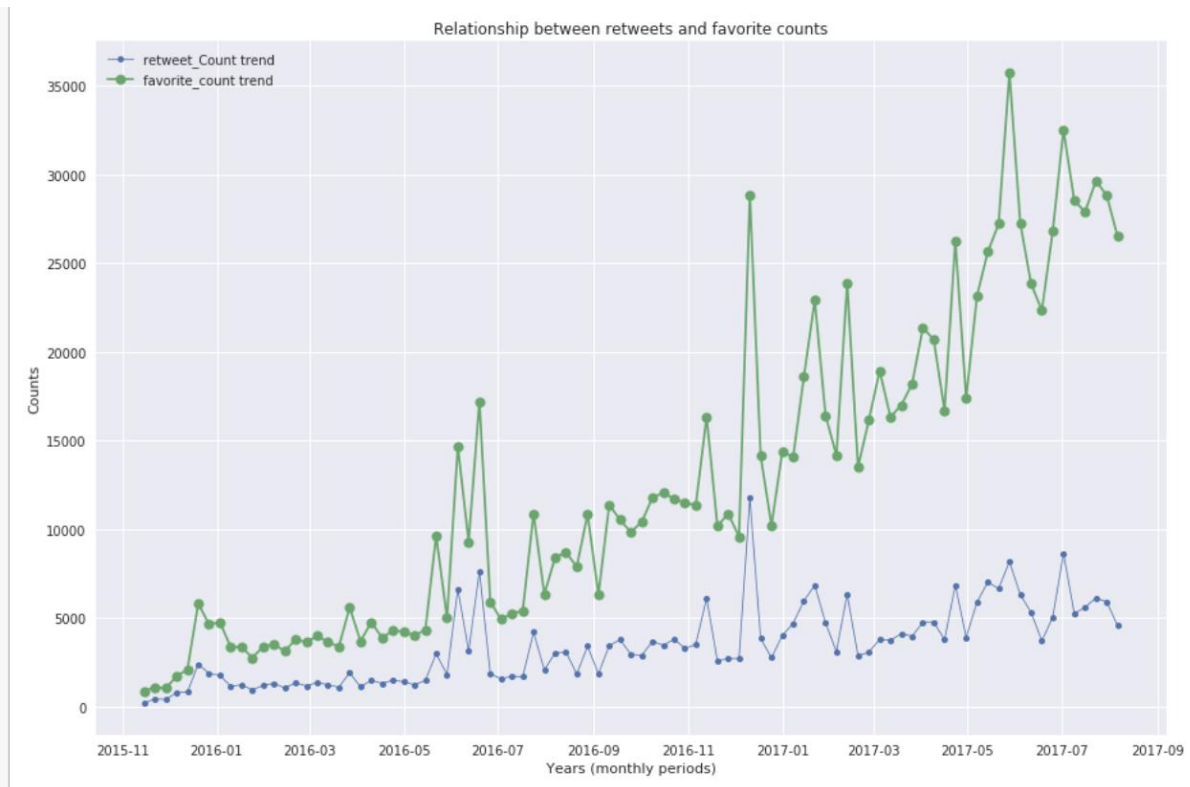
Out[90]:

	tweet_id	retweet_count	favorite_count	score	img_num	confidence
tweet_id	1.000000	0.371289	0.598027	0.411765	0.204043	0.102957
retweet_count	0.371289	1.000000	0.930349	0.235866	0.091590	0.018801
favorite_count	0.598027	0.930349	1.000000	0.307315	0.118822	0.048497
score	0.411765	0.235866	0.307315	1.000000	0.124200	0.117648
img_num	0.204043	0.091590	0.118822	0.124200	1.000000	0.145801
confidence	0.102957	0.018801	0.048497	0.117648	0.145801	1.000000

The correlation chart is useful for finding connections between variables, especially with hypothesis testing or an A/B test. The numbers range from 0 to 1, and a positive number is a positive correlation and vice versa for a negative number.

The correlation coefficient between retweet_count and favorite_count is 0.930349, which is close to 1 and positive demonstrating a strong positive correlation between those two metrics.

In order to find the relationship between retweet_counts and favorite_counts, a time series analysis was used, and the following results were obtained:



The overall trend of the retweet and favorite counts shows the popularity of this twitter page gradually growing over time.

As mentioned before, with this timeseries chart above, the favorite_count and retweet_count are positively correlated with one another. This is due to the fact that most people retweet 'tweets' that they like in order for others to see it. It is like free advertising for the tweet and shows people on your feed what you're interested in. It's a sharing feature when you want others to see something you've read/seen.

As shown in the chart, there are 3 peaks, those being, **the middle of spring into summer(twice), and then the Christmas/Holiday time**. This is probably due to the fact that people with dogs are more active in the warmer months, posting cute things their dog is doing outside. Also, during the holidays, they are more likely to share pics/etc about things they care about during this time i.e. their dogs.