

Jan 30th, 2019

Yi Su

## Act Report

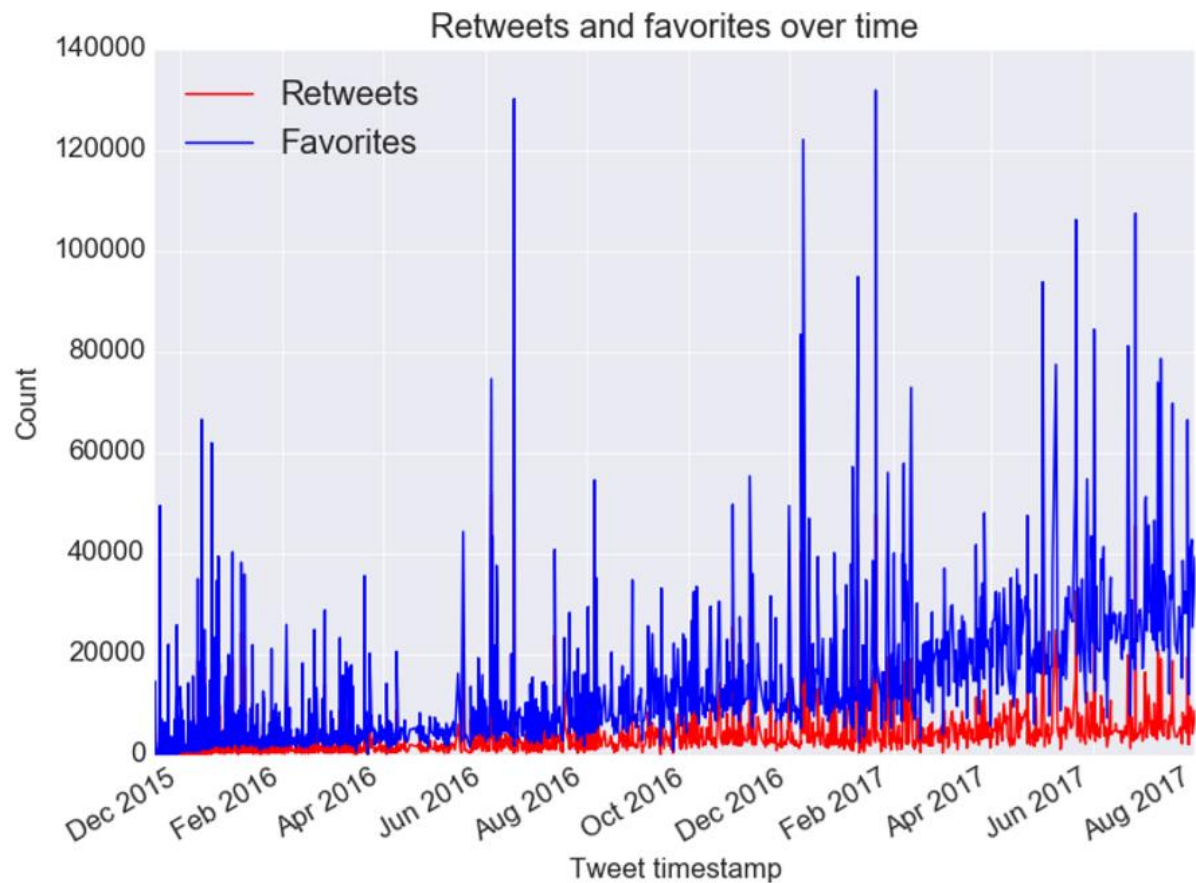
This analysis is from the WeRateDogs twitter account. My analysis focus on three aspects: rating ratio, number of retweet and number of favorite. Based on the number of retweets and favorites, and analysis of the rating scores. I wonder the trend in popularity over time of the twitter account. Some insights came out when I evaluated the master data set of the WeRateDogs tweets:

The mean for rating ratio is 1.165. According to chart below, We can see the 3 most common ratings are 1.2, 1.0, 1.1, and rating frequency becomes smaller as the rating becomes extreme. The mean for retweet is 2646. The mean for favorite is 8717.

	rating_ratio	retweet_count	favorite_count
<b>count</b>	1991.000000	1991.000000	1991.000000
<b>mean</b>	1.165082	2646.30437	8716.746359
<b>std</b>	4.066945	4718.08456	12785.932914
<b>min</b>	0.000000	11.00000	78.000000
<b>25%</b>	1.000000	591.00000	1869.000000
<b>50%</b>	1.100000	1269.00000	3936.000000
<b>75%</b>	1.200000	3032.00000	10878.000000
<b>max</b>	177.600000	83466.00000	164027.000000

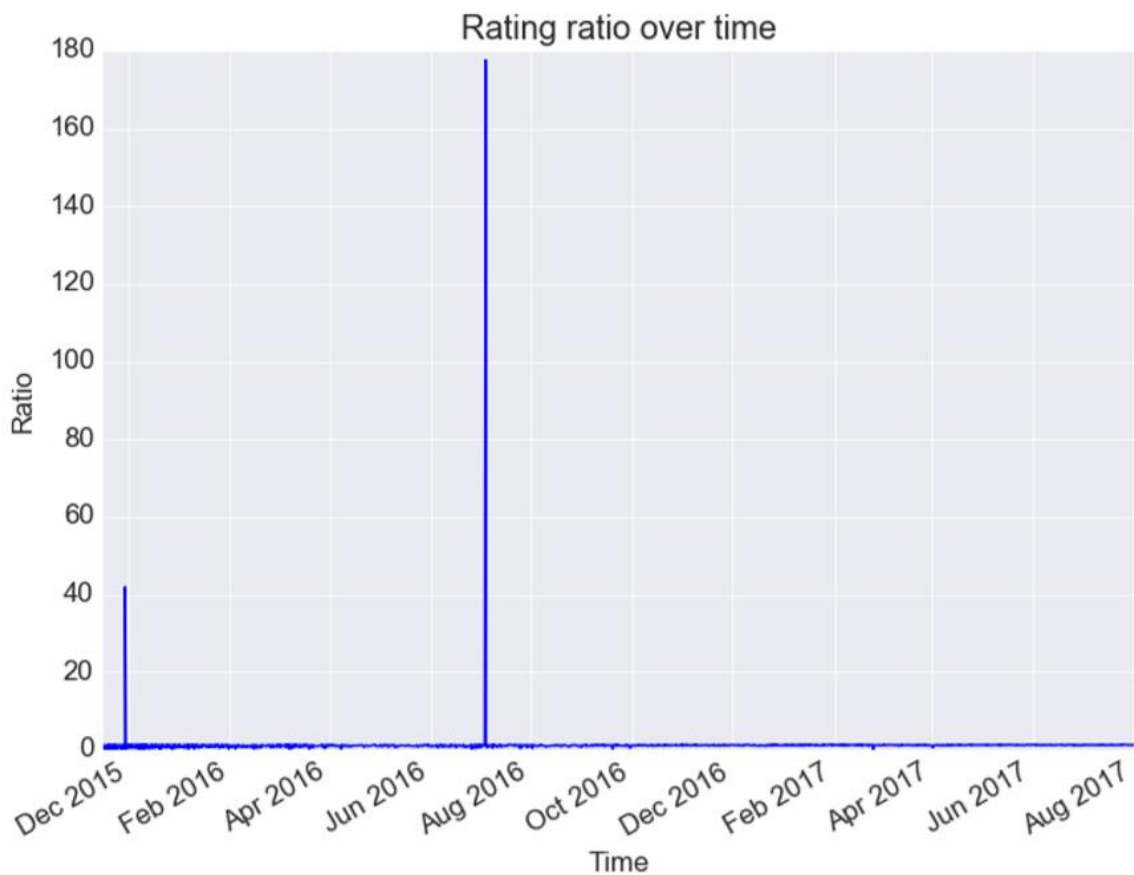
**Project 4**      **Udacity | School of Data Science**  
**Data Analyst Nanodegree**  
**Nov 13, 2018 Class**

The trend in the favorites and retweets is increased, presumably as the account became more popular. In the chart below, we see an upward trend for both retweets and favorites. Compared to retweets, The number of favorites has more increased. Also there are several large outliers in favorites for extremely popular tweets.



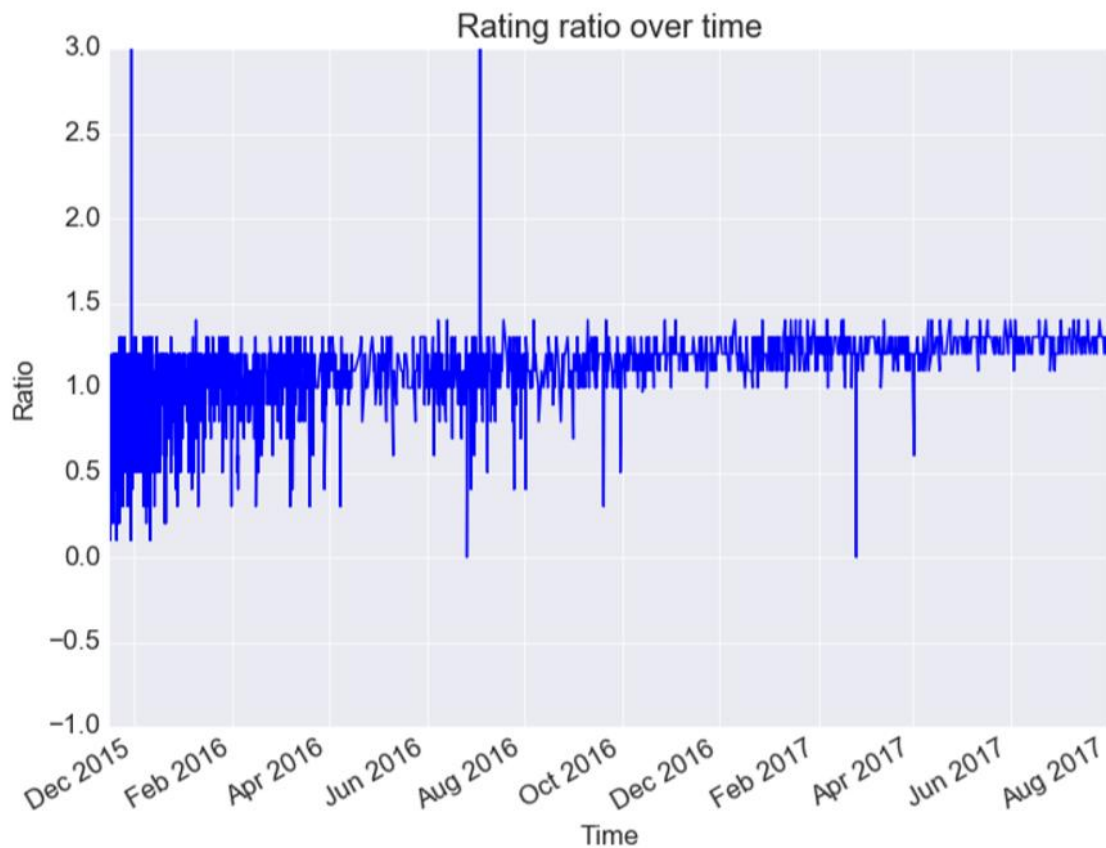
**Project 4**      **Udacity | School of Data Science**  
**Data Analyst Nanodegree**  
**Nov 13, 2018 Class**

The dog ratings does not abide to a one out of ten scale. There are a few ratings larger than 10. Considering the popularity of WeRateDogs tweets, it seemed reasonable to look at the favorites and retweet counts. In order to normalize the ratings, I created a ratio of the rating numerator divided by the denominator(a proportion rating score). When this is plotted, we see two extreme outliers.



**Project 4**      **Udacity | School of Data Science**  
**Data Analyst Nanodegree**  
**Nov 13, 2018 Class**

If we zoom in and ignore the outliers and view the the data, we can get a better idea of the rating ratio trend:



In the chart above, we can see that the scores trended towards higher than a 1:1 ratio as time goes by. They are far fewer outliers below 1. We can also see that a few dogs received zero scores, or scores close to zero. Lower scores are given in general earlier in the dataset.