

# act\_report

April 29, 2019

*****	Udacity
Nanodegree:	<i>Data Analyst</i>
Student:	<i>Theresa Kocher</i>
Date:	<i>29th April 2019</i>

## 1 Introduction

In this study, three different views to the tweets of the famous twitter account *WeRateDogs* were examined. The first view to the data comes from the [WeRateDogs tweets archive](#). It contains the rates every dog got from the *WeRateDogs* authors. The second view to this tweets contains the [dog's breed](#), which was classified by a neural network in the Udacity's *Machine Learning* Nanodegree program. And the last view to the tweets is accessed from twitter itself by using the [Tweepy API](#). This data contains how often a tweet was retweeted and labeled as favorite from other twitter users.

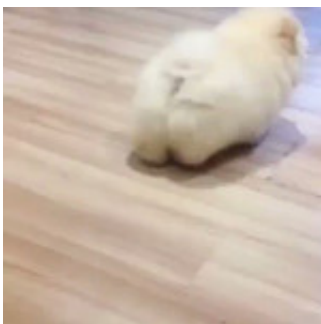
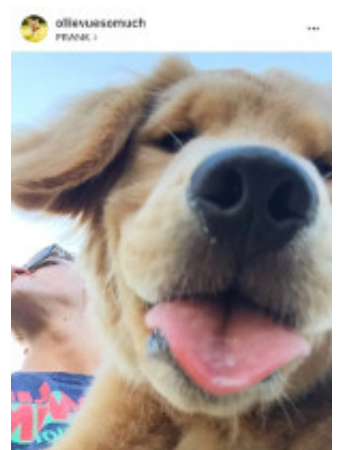
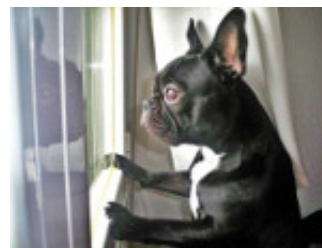
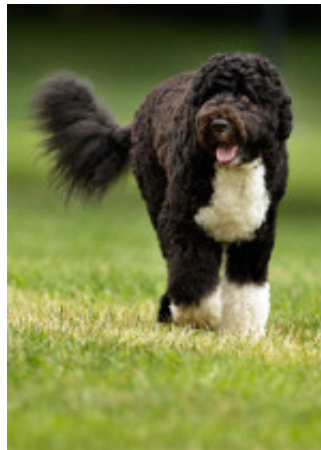
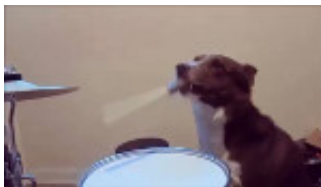
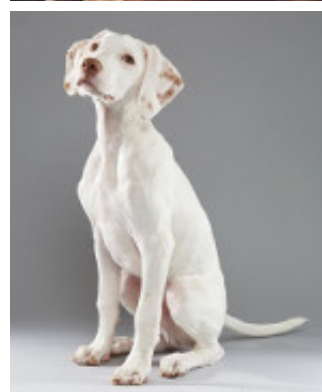
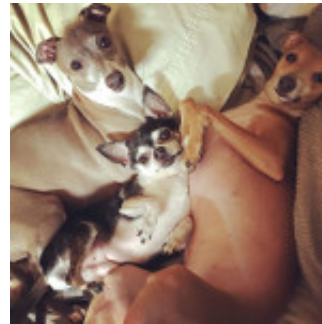
## 2 Which one is the best dog?

All sources give us a different option to estimate the value of the dogs or at least of the corresponding tweets. The first data set gives us the rates the authors published in the tweet. This tells us which dogs gave the most pleasure to the authors. The second source gives us a predictability confidence for each tweet, which tells us how easy it is to predict the dog's breed. And the last source gives us a rating from the twitter community with every tweet's favorite and retweet counts.

Therefore, we have three different measures to rate the dogs: (1) the authors rating, (2) the neural network's breed prediction and (3) the twitter community's rating. Let's see which dogs are the best under each measurement. And afterwards, let's see if there are coherences between those measurements.

### 2.1 WeRateDogs author's best rating

The highest rate the *weRateDogs* authors gave were 14/10 credits. In the history of *WeRateDogs* tweets, only 13 dogs achieved this rate. In the following these most outstanding dogs are shown.



## 2.2 Udacity's neural network best breed prediction

But if we ask the Udacity's neural network, which dog was the easiest to classify, it will answer different. This dog's breed *komondor* was the easiest to predict with a prediction confidence of 0.999956.



Best predicted dog	* * *
<b>Dog name:</b>	an
<b>Dog breed:</b>	komondor
<b>prediction:</b>	0.999956
<b>rate:</b>	10 / 10
<b>retweet:</b>	496
<b>favorite:</b>	1040

## 2.3 Twitter's community most retweeted dog

The last measurement is obvious and it's defined by the twitter's community. Which dog is the most retweeted and most favored tweet from the twitter account of *WeRateDogs*. This is obvious. Following dog is retweeted and favored the most.



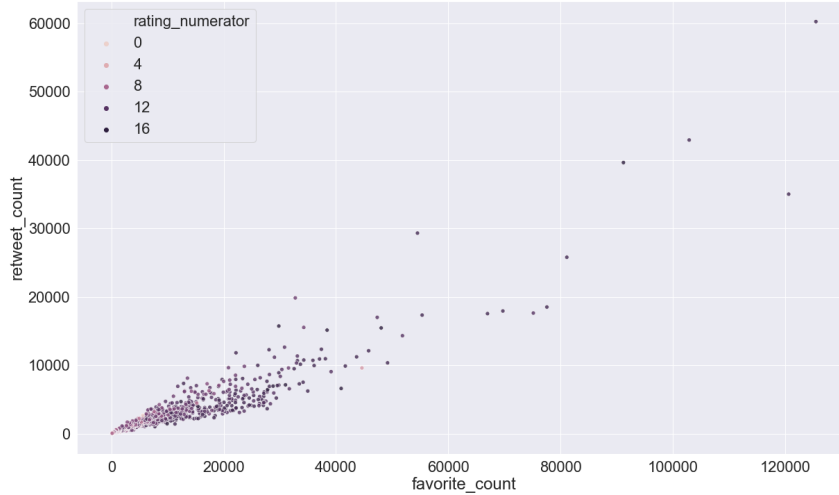
Most retweeted and favored dog		***
Dog name:	Stephan	
Dog breed:	Chihuahua	
prediction:	0.953442	
rate:	13 / 10	
retweet:	60236	
favorite:	125519	

### 3 Statistical evaluation of all measurements

It would be interesting to know if those used measurements to estimate the dogs do correlate.

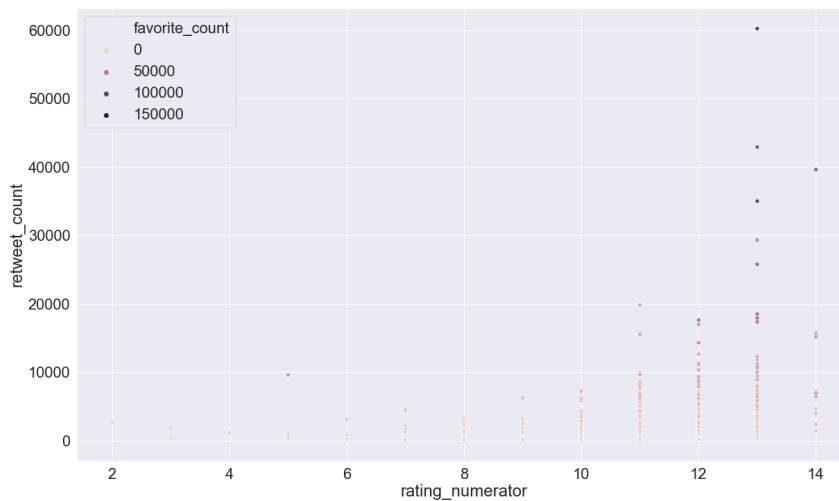
#### 3.1 Retweet and Favorit count

First of all, we will compare the retweet amount with the favorite amount. And in this scatter plot, we see that those measurements generated by the twitter's community correlate as was expected.



### 3.2 Retweet count and rating value

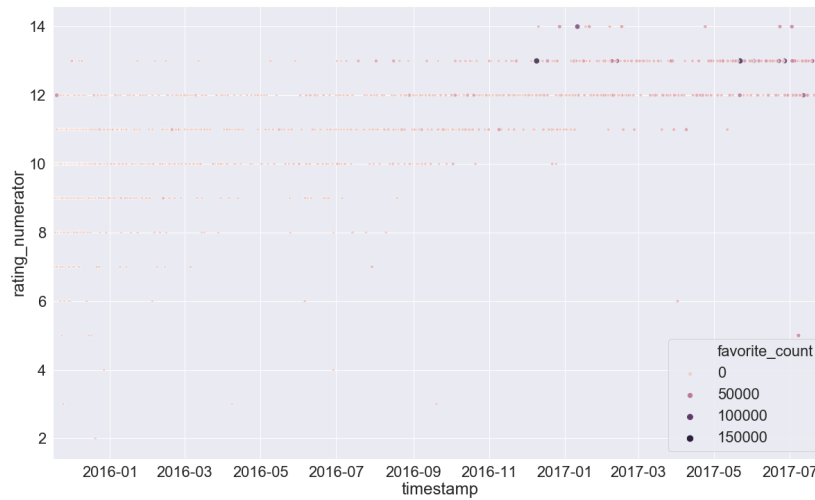
Another interesting correlation is visible in the scatter plot of the retweet amount over the author's rating. It seems like the higher the author's rating is, the more tweet's are retweeted from the community. Also the favorite count increases with the author's rate. But we have to be careful when interpreting these scatter plots. What we don't see here, is the change of the measurements over time.



### 3.3 Rating value over time

This is why an additional scatter plot is shown here. It shows that the given rate from the authors has changed over time. There are more great ratings (>10) and less low ratings (<10) in 2017 then there were in 2016. Also the favorite amount increased in the second half. This can be also

caused by the increase of popularity of the twitter account. This is another factor which was not considered yet in this data analysis.



## 4 Conclusion

The retweet and favorite counts correlate. And over time we can see that the author's rate increases. But so far, we can not tell if there is a coherence between the author's rate and the community's retweet or favorite count. Because there are factors not considered yet. Both, the time factor and the popularity factor can be biases-not considered factors-that may have influence to the measurements change and can lead to wrong interpretations.