# Twitter WeRateDogs®

By: Peter Schuld, Feb. 7th, 2020

Udacity Data Analyst Student

The popular WeRateDogs® (@dog_rates) Twitter account rates people's dogs with a humorous comment about the dog. Since launch in 2015 by college student Matt Nelson, the account has accumulated 5,000+ Tweets. A team of Data Scientists at Udacity have already used deep learning algorithms to analyse approx. 2,000 images from original Tweets (no Retweets). The machine learning (ML] algorithm identify the dog breed for 85% of them, while some images do not show dogs at all or disguise them wearing costumes. The WeRateDogs® Tweet with the most 'Likes' and most 'Retweets' shows the video clip of a 'doggo' or mature dog (see picture below). The ML algorithm correctly classified the dog breed **Labrador Retriever.**

In a next step, I have performed further Data Analyst tasks like data wrangling, cleaning and analysis to derive meaningful inside from the database.

The two most frequent dog breeds appearing on Twitter WeRateDogs® postings are **Golden Retriever** and **Labrador Retriever.** Another retriever breed **Chesapeake Bay Retriever** appear in the top 20 as well. Retrievers were originally bred to help hunters find birds or other game and bring them back undamaged. Yet, on WeRateDogs® they rather excel as adorable domestic animals than reliable hunting dogs.
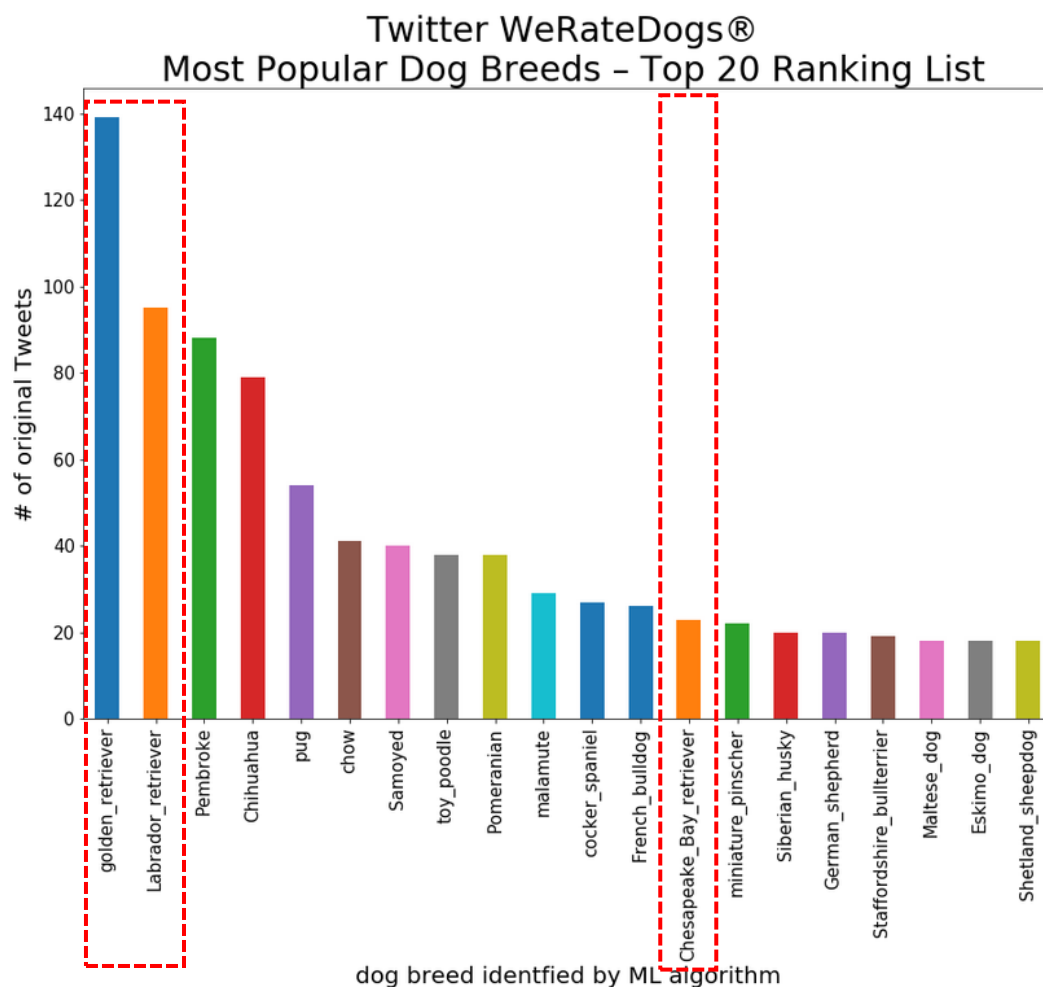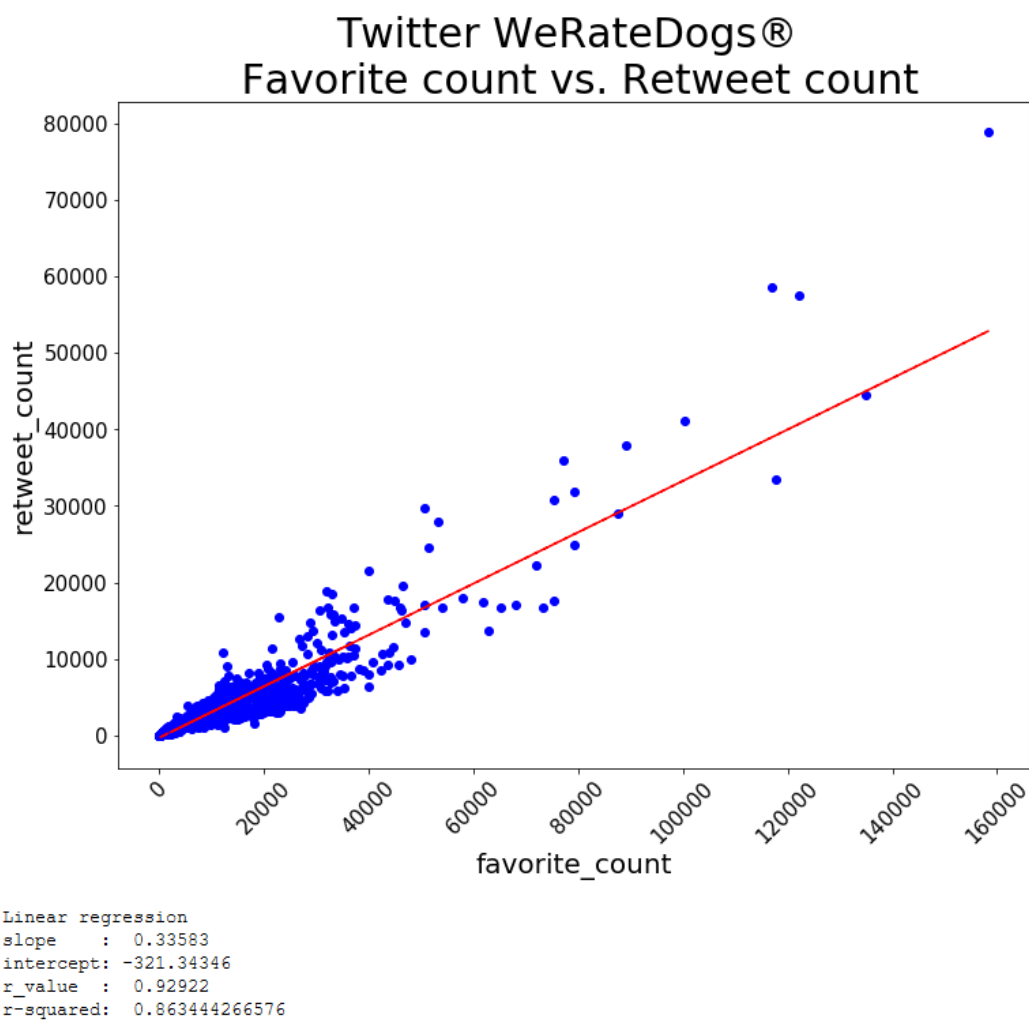

Traditional herding dog breeds like Pembroke and Samoyed also feature high on this Twitter account. Nevertheless, small breeds like Chihuahua, Pug and Toy Poodle are also frequently rated.

The distribution of the number of favourite counts per WeRateDogs® posting shows an exponential shape. Postings easily receive a few thousand 'Likes'. However, only 5% of Tweets 'get viral' with more than 30,000 favourite counts. Postings with more than 60,000 'Likes' represent less than 1% of all Tweets.
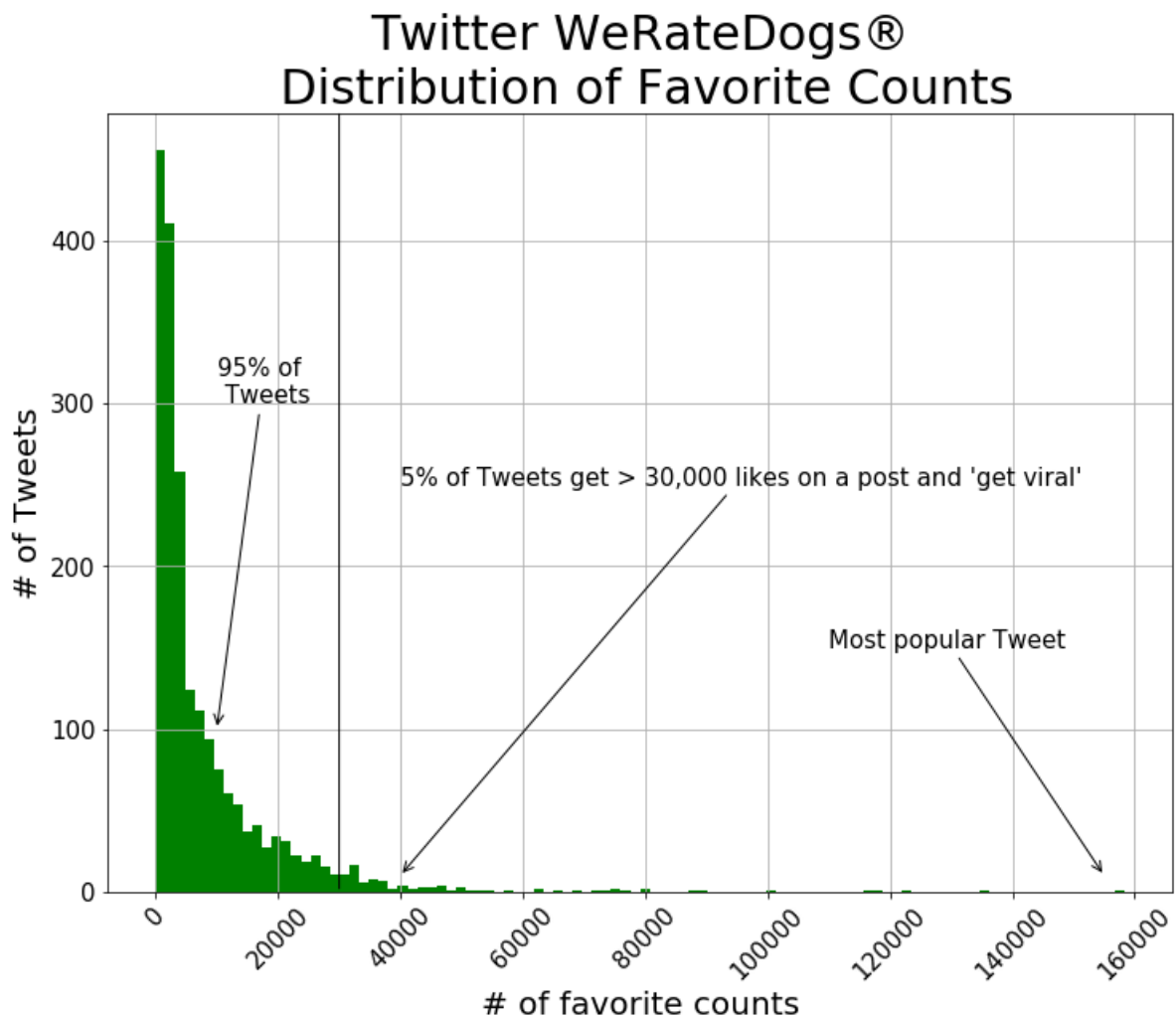
There is a strong statistical evidence for a positive correlation between favourite counts and retweet counts (($R^2 = 86\%$) . On average, postings receive 3-times more 'Likes' than 'Retweets' ($slope = 0.34$) . However, some of the most popular tweets receive proportionally more retweet counts.

Twitter WeRateDogs®
Favorite count vs. Retweet count

```
Linear regression
slope    :  0.33583
intercept: -321.34346
r_value  :  0.92922
r-squared:  0.863444266576
```

Decipher the rating scale

Twitter WeRateDogs® has a unique rating scale where most ratings have a denominator of 10 and a numerator greater than 10 (e.g. 11/10, 12/10, 13/10 or more than 100%). Apparently, this rating scale with numerators exceeding denominators was not originally intended and its validity was questioned by the early user Brent. Nevertheless, he only received the explanation: "because they're good dogs Brent." and the unique rating scale continued ever since.
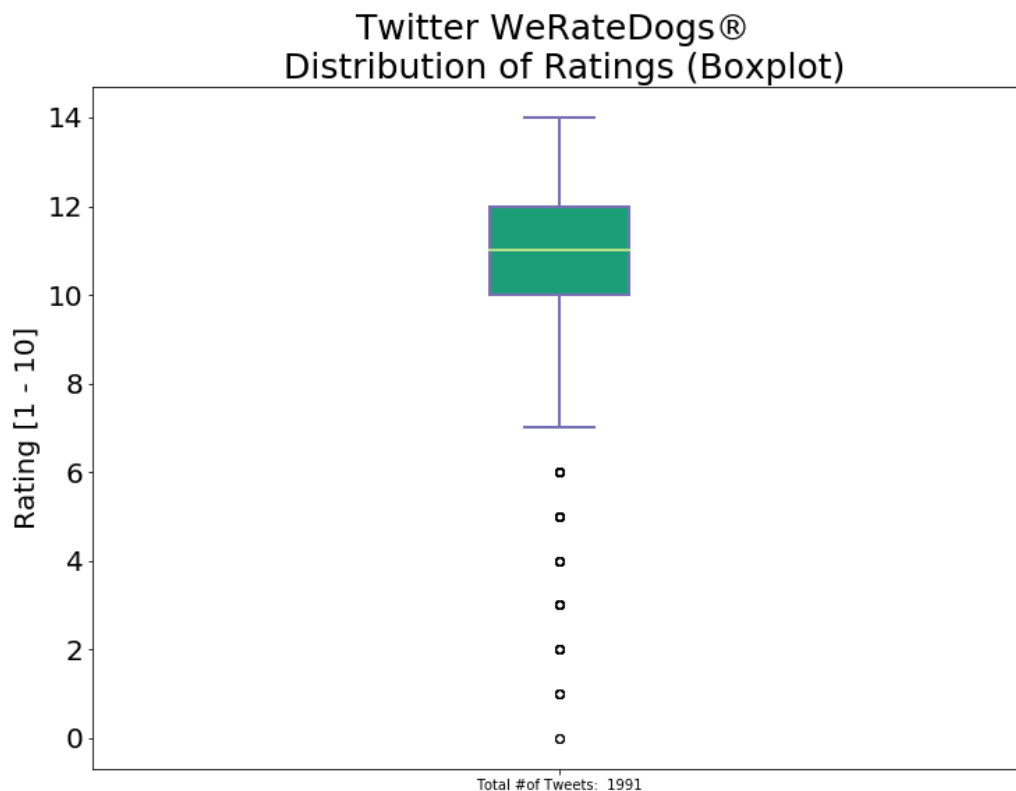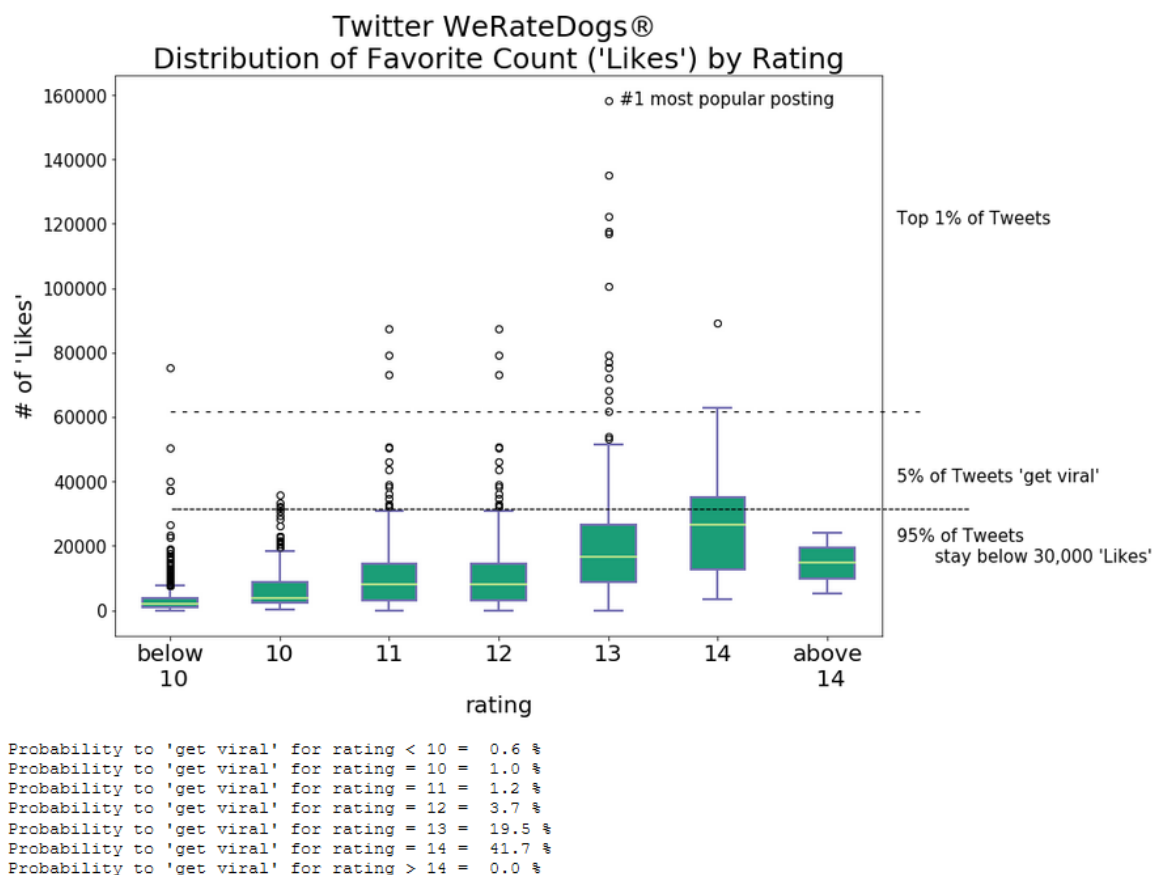

The distribution of the ratings confirms the unique scale. The first quartile (Q1 / 25th Percentile) is exactly 10.0, with the Median (Q2 / 50th Percentile) at 11.0, and Third quartile (Q3 / 75th Percentile) at 12.0. Hence, the Interquartile Range (IQR) of 2 is narrow and half the ratings are clustered between 10 and 12.

Apparently, most of the ratings below 10 are meant to be funny when read next to the image.

Example for a rating of 5/10:

Ratings do not explain Favourite Counts and the correlation between Rating and 'Likes' (favourite count) is not statistically significant ($R^2 = 14.7\%$). However, only a few postings with a rating below 10 did 'get viral', presumably because a low rating indicates some flaw. Incidentally, all postings with more than 100,000 'Likes' have a Rating of 13. Correspondingly, we can use the rating to forecast the chances of a Tweet 'getting viral'.


```
Probability to 'get viral' for rating < 10 =   0.6 %
Probability to 'get viral' for rating = 10 =   1.0 %
Probability to 'get viral' for rating = 11 =   1.2 %
Probability to 'get viral' for rating = 12 =   3.7 %
Probability to 'get viral' for rating = 13 =  19.5 %
Probability to 'get viral' for rating = 14 =  41.7 %
Probability to 'get viral' for rating > 14 =   0.0 %
```

In summary, the 5000+ Tweets in the Twitter WeRateDogs® account feature approx. 2000 original Tweets (no Retweets) showing images of dogs with ratings. While most Tweets get a few hundred or even thousands of 'Likes' only 5% manage to draw the attention of more than 30,000 users. Those 'viral' Tweets are not predicted by a high rating, but most of the highly successful Tweets have ratings above 10. Almost all the Top 1% Tweets have ratings of 12, 13 and 14. Therefore, the unique rating of WeRateDogs® operates on a scale of 0 to 14. Ratings below 10 indicating some shortcomings and less than 1% receive more than 30,000 'Likes'. Ratings of 10, 11 and 12 have probabilities of 'get viral' of 1.0%, 1.2% and 3.7%, respectively. Ratings of 13 have a 20% chance of 'getting 'viral and images with a rating of 14 have a 42% chance of becoming very successful Tweets.