



# Data Analysis Writeup

UDACITY DAND PROJECT 7

James R. Small | September 7, 2018

## Data Analysis

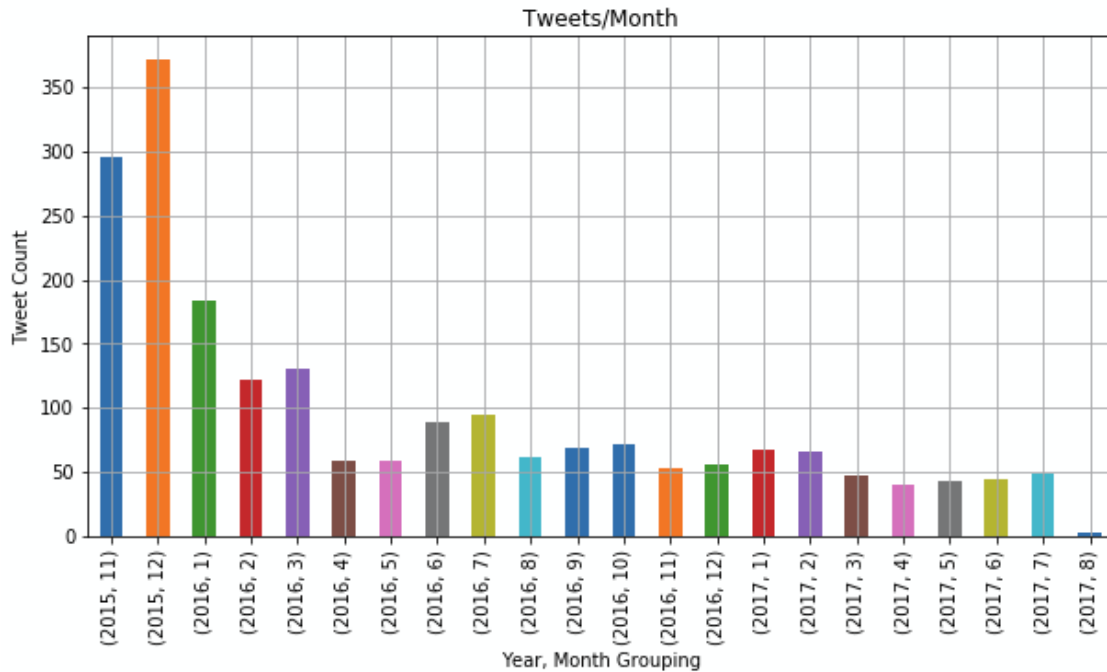
[WeRateDogs \(@dog\\_rates\)](#) tweets from November 15, 2015 – August 1, 2017 were examined for this report. While this span includes over 5,000 tweets, it was pruned down to 2,067 – only original tweets that included [a dog rating](#). The tweets averaged about 111 characters with a median of 115 and a standard deviation of about 26. The vast majority of the tweets (93.6%) were from Twitter’s iPhone client, with the rest from Vine, Twitter’s Web Client, and TweetDeck. Most tweets included an embedded photo, video, or animation (95.3%) with 24% including two or more. Those that did were run through an image classifier which attempted to determine the dog breed.

The examined tweets rated dogs using the unconventional system referenced above. The average score was 11.63/10. Within the tweets, over two-thirds (67.2%) included a dog name. Nearly one in five (18.4%) included a dog stage “category” (see [#WeRateDogs Book](#) for definitions of doggo, pupper, puppo, and floofer). All of the tweets were also retweeted and “liked” multiple times. Retweets averaged about a skewed 2,750 with a median of 1,335 and a huge standard deviation of about 4,808. “Likes” averaged about a skewed 8,805 with a median of 4,014 and another large standard deviation of about 12,742.

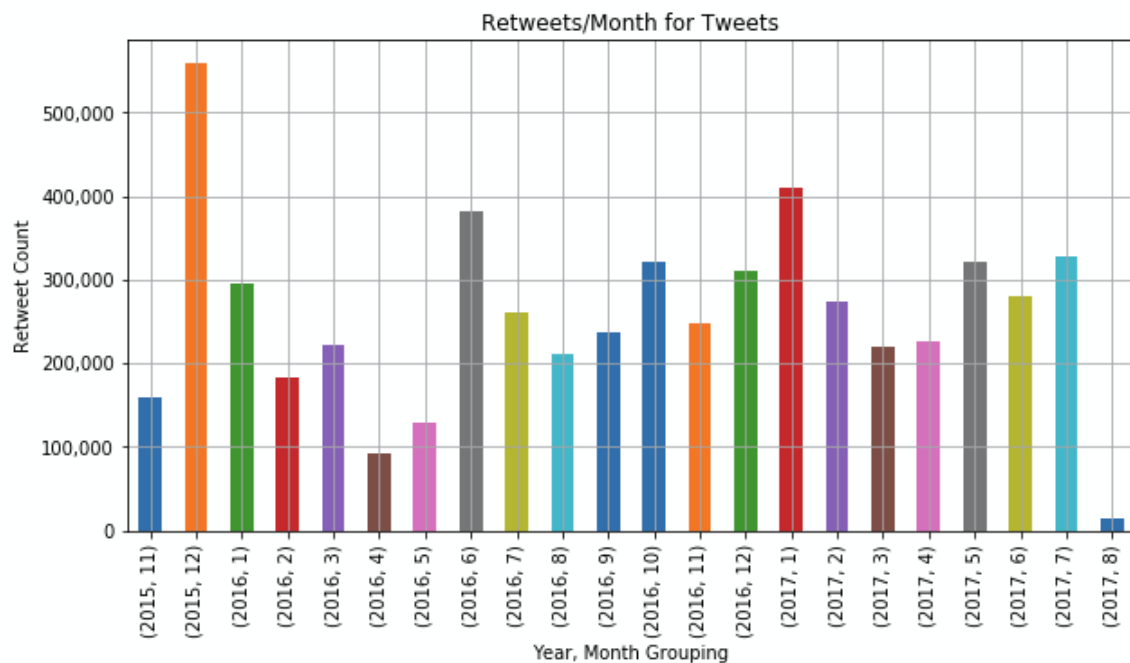
Looking at the image classification, the first embedded photo (out of up to 4) from the tweet was usually the one selected (85.8%). This makes sense as most of the tweets only contained one picture. The first guess had an average confidence of about 59.5% and predicted the picture was a dog about three-quarters (74.2%) of the time. The second and third guesses had much lower confidence ratings (13.5%, 6%) but still predicted the photo was a dog about three-quarters of the time (75.1%, 72.6%). The most popular breed predicted by the first guess was a golden retriever (6.6%), followed by a Labrador retriever (4.5%), and a pembroke (4.3%).

## Data Visualization

Looking at the tweets over time, we see a large initial spike peaking a month after the start at over 350/month. The tweet pace then drops off declining over half the following month and then trends downward. The decline slows over time with some ups and downs. The last month is a bit of an outlier though as it’s not a complete month of data. We can speculate that over time, @dog\_rates is going with quality (and perhaps novelty) over quantity.



Looking at retweets we see a different story. While the first month has a decent retweet count, the second month has a meteoric rise! This is followed by a large drop-off and continual declines for a few months. Around mid-2016, retweets take off again and then seem to bounce up and down after that.



Finally, examining the “likes” we see yet another trend. As with retweets, likes start out well and soar the second month. While there’s then a big decline like above, from about mid-2016 on the likes are generally trending upward. In fact, the best month to date was August of 2017. This shows a clear increase in popularity.

