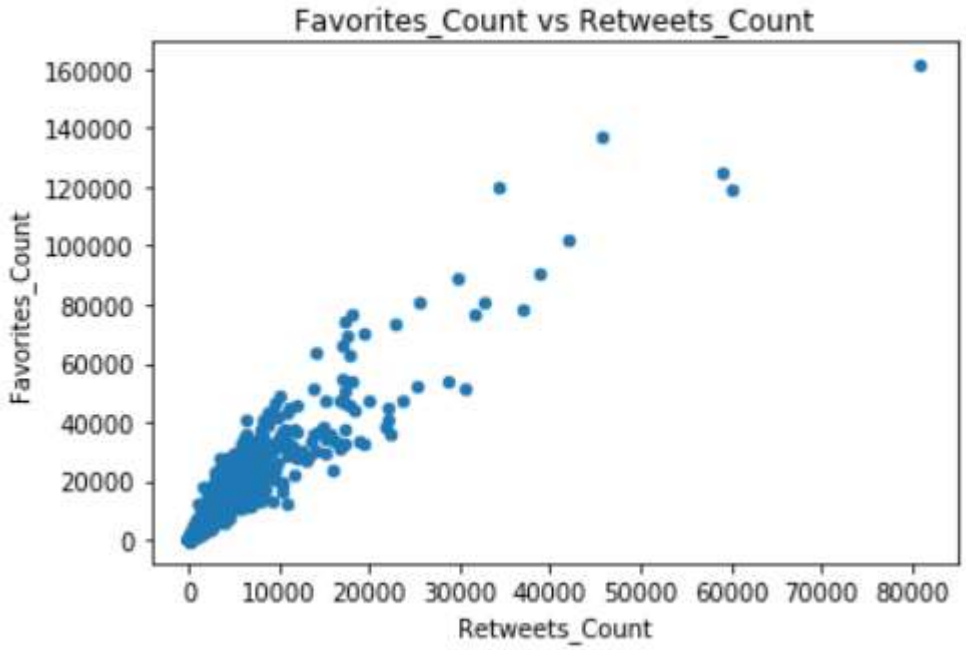
WeRateDogs Twitter Account Archive Analysis

Introduction

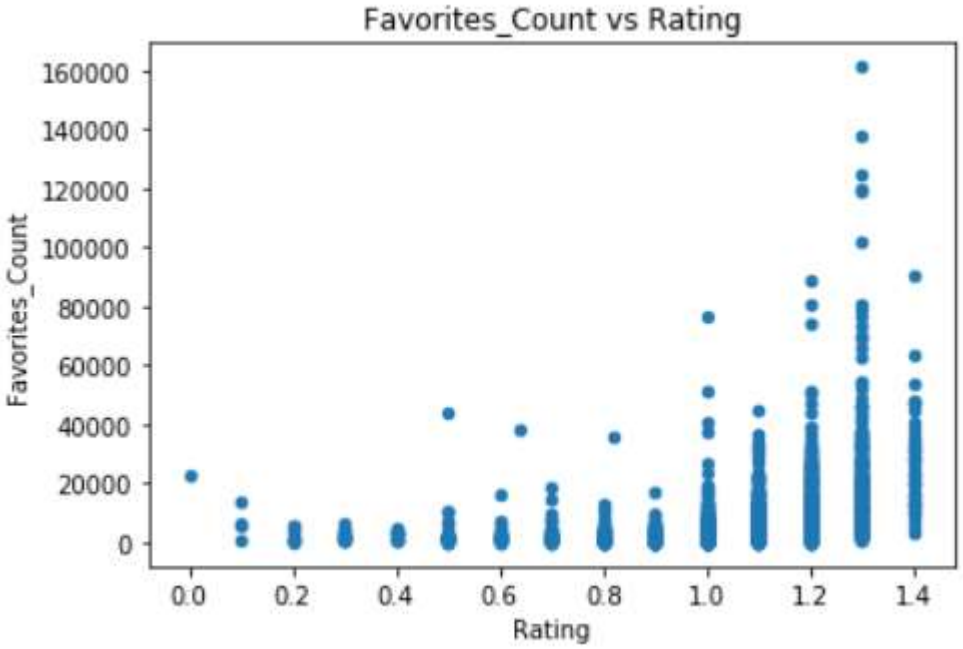
The WeRateDogs Twitter account attracted over 8 million followers by tweeting humorous comments about photos and videos submitted that typically contain one or more dogs. An archive of data about original messages (or "tweets") from this account from November 2015 through August 2017 was analyzed to gain insights into some aspects of the Twitter social media platform in general and into WeRateDogs. In particular, the archived tweets along with additional information added to enhance the dataset was analyzed for answers to the following questions:

Are favorite counts and retweet counts correlated? The primary metrics of the popularity of a tweet on twitter are the number of twitter users that mark the tweet as a favorite (favorite count) and the number of users that resend the tweet to their followers (retweet count). One way to gain some insight into the Twitter platform is to use the WeRateDogs archive to analyze the correlation between these two metrics of tweet popularity. We did this and the results are that, at least for this one successful Twitter account, favorite count and retweet count are highly correlated. This can be seen in the following scatter plot.



The positive correlation evident in this scatter plot is confirmed by calculating the correlation coefficient. This resulted in a value of 0.93 which is a high degree of correlation indeed.

Are the ratings assigned to dogs by the WeRateDogs authors correlated with popularity as measured by favorite counts? Each original tweet from this Twitter account contains a rating on a 0 to 10 scale (although the rating often exceeds this upper limit – for humorous affect). But does this rating assigned by the WeRateDogs authors correlate to the popularity of the tweet and accompanying photos or video of the dog? We analyzed this question and found no signification correlation as can be seen in the following scatter plot.

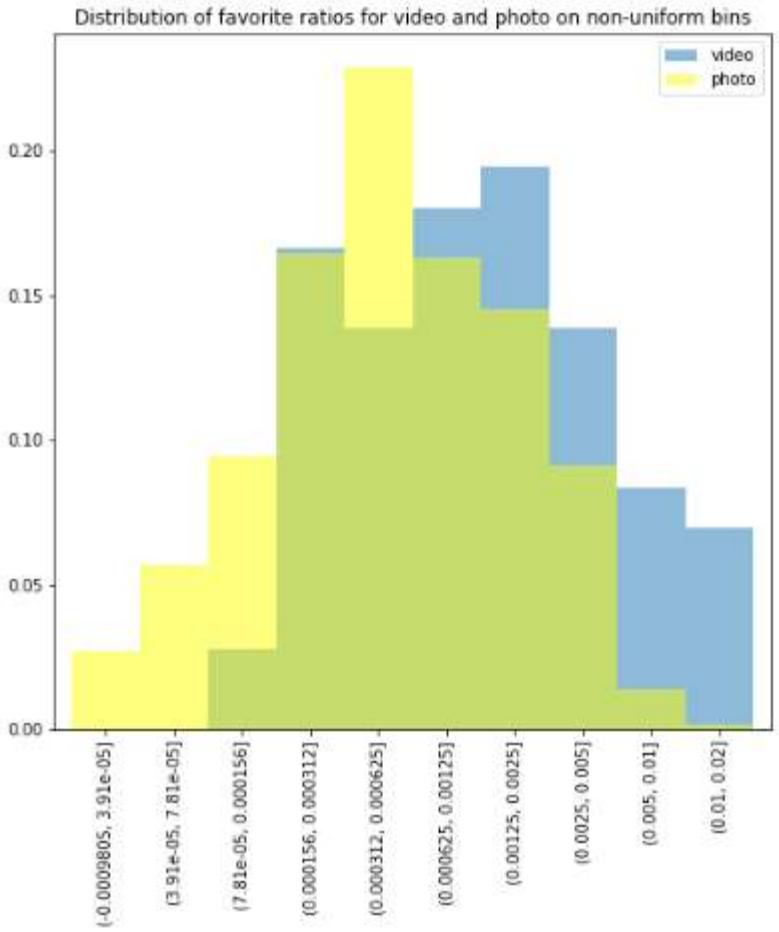


The computed Pearson correlation coefficient was 0.38 which confirms what is apparent in the scatter plot.

One of the humorous themes in tweets from the WeRateDogs account is the *protest tweet*. The author will complain that one again someone has submitted a photo or video that does not contain a dog. These tweets include phrases like “we only rate dogs” or “please only send dogs”. In reality, these submissions do indeed contain a dog; however, the dog is often disguised or is posed in an unusual position or blends in with the background. We thought that since this dataset included images that we could assess the accuracy of a neural network trained to detect the presence of a dog and predict its breed. This neural network delivers 3 predictions per photo. If it detects a dog it also predicts the breed. Here we only assess the accuracy of the neural network in detecting the dog in these more difficult photos. The results were as follows:

- In 16% of the cases the neural network failed to detect a dog at all
- However, when it did detect a dog in a photo, it did so with an average confidence of 0.77 which is quite high

The final question that we analyze using the WeRateDogs Twitter archive is: **Are videos submissions more popular than photos as measured by favorite counts?** In other words, do the WeRateDogs followers respond more favorably to tweets about videos than to tweets about photos? We performed a inferential statistical analysis of the archive dataset and found that indeed favorite counts for tweets about dogs in videos were significantly greater statistically than tweets about dogs submitted in still photo format. The difference in the distribution of favorite counts for videos and photos can be clearly seen in the following plot.



Note that the bin sizes in this plot is not uniform. The width of each successive bin from left to right increases by a factor of 2. This is done to provide more resolution at the lower end of the distribution.