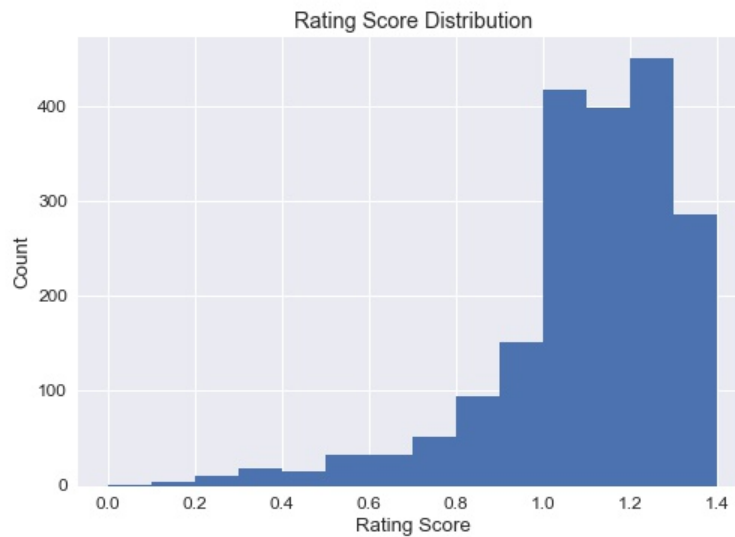
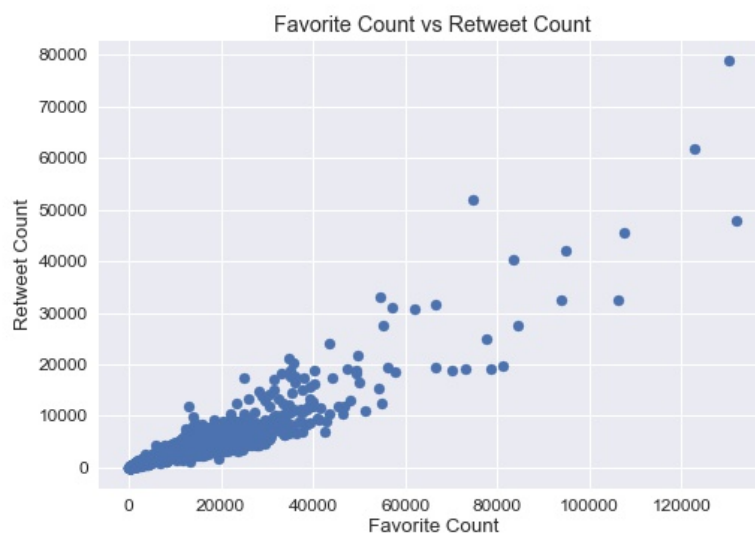


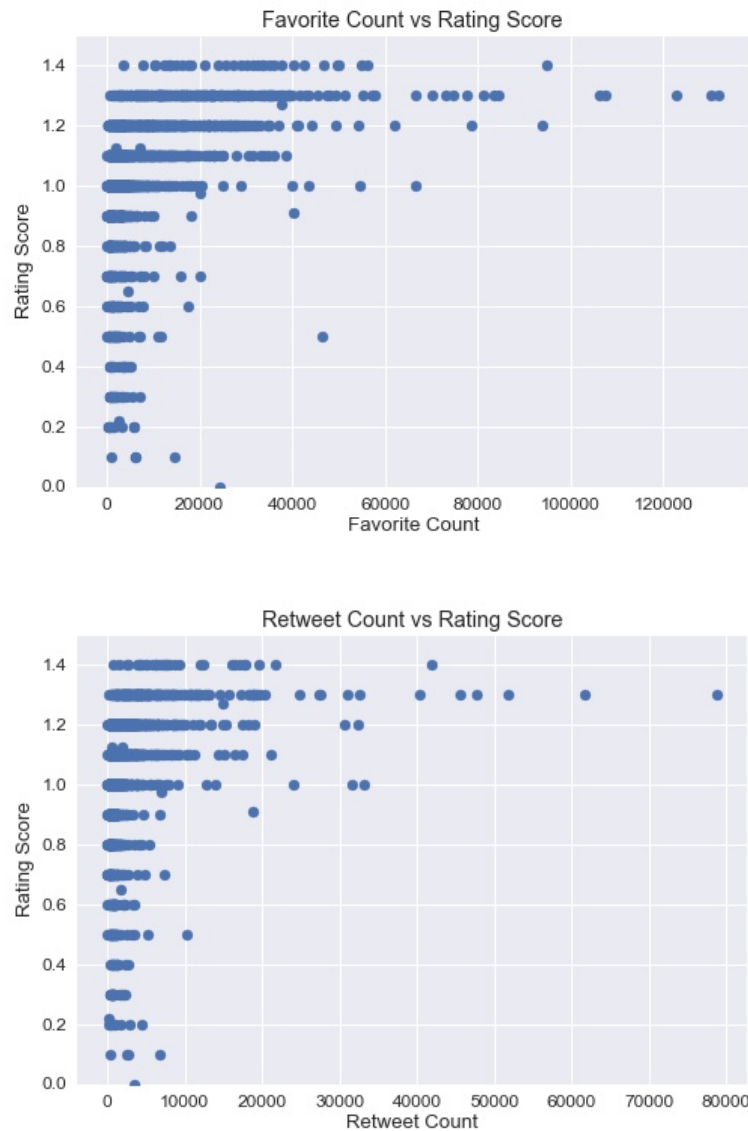
Act Report

A number of insights were garnered upon evaluation of the master data set of the WeRateDogs tweets. Considering that WeRateDogs does not abide to a one out of ten scale, but a fraction normally above one, a proportion rating score was created. Plotting a histogram of the tweet rating scores shows that they do indeed like to rate the pups above 1.0 or over 10 in their fractional scoring system (11/10, 12/10, etc.).



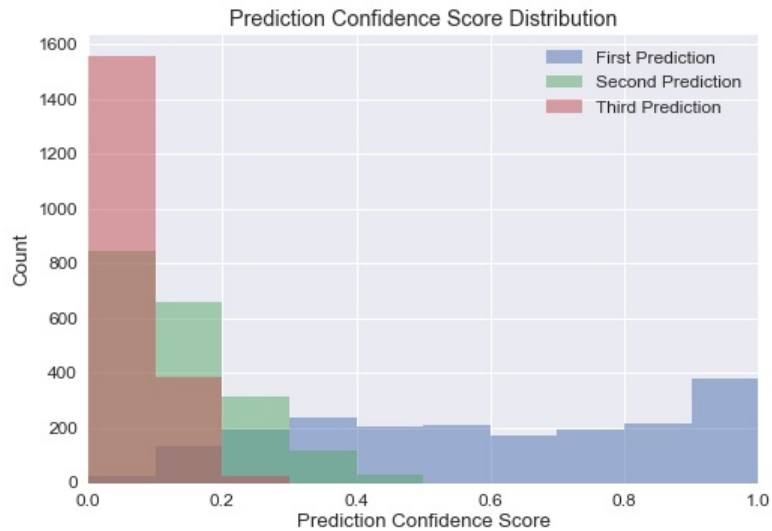
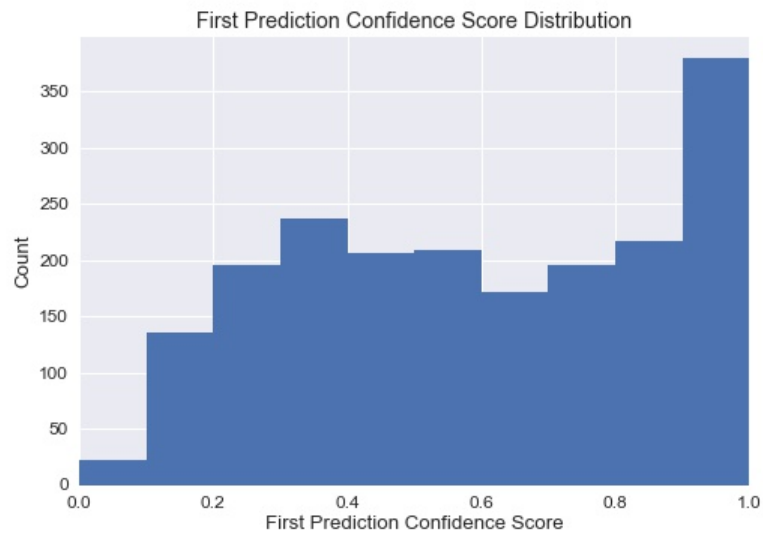
Considering the popularity of WeRateDogs tweets, it seemed reasonable to look at the favorites and retweet counts. There seems to be a strong positive correlation between the two, which makes sense since we tend to share things we like with others. I became interested in seeing if favorite count and retweet count positively correlated with the rating score of the tweet. And indeed tweets with higher rating scores tended to have higher favorite and retweet counts.





In addition to acquiring data regarding the WeRateDogs tweets, a dataset with predictions on the type of animal pictured in the tweet via a neural network was provided for analysis. Looking at the distribution of percent confidence ($100\% \times \text{confidence score}$; how confident the algorithm is with its prediction) for the first prediction it shows that the majority of predictions is below 80% confidence. When looking at the count distribution of the top three predictions there is a significant drop off in confidence after the first prediction. The second and third predictions are only confident up to 40% and 20% respectively.

Joshua Demeo
171220



Getting together a master data set was no easy tasks and I definitely can understand and appreciate why data scientists spend so much time gathering, assessing, and cleaning data. The creation of visualizations seem to be so far off from the initial process, but well worth the sweat. I look forward to building my own predictive models in the future. Cheers!