WeRateDogs - A Journey into their Twitter Data
## by Khan Akbar

Insight #1: Top Rated and Lowest Rated Dog Twitter's WeRateDogs is increasing in popularity due to the rating culture of dogs. Followers rate out of 10 but none are under 10 because there just cannot be a dog that gets rated low. The ones that do get rated lowly are usually no pictures of dogs and or are dogs but disguised as other animals or objects. And even with this low rating, they do not go below a rating of 10.

Before performing my analyses and making my insights and visualizations, I had to wrangle the WeRateDogs Twitter archive. I took out the lowest ratings found within tweets – 0,1 and 2. These ratings should not include anything below a 10. This still left decimal ratings that I did not detect from my programmatic assessments such as 0.1, 0.11, and 0.12. Upon further investigation, these ratings were actually 10, 11, and 12 respectively due to periods from a slight pause (i.e. tweet text followed by the rating….10/10). In the end, the lowest-rated dogs were still rated at 10 or above. The highest-rated dog received 1776/10 which is the picture to the right. The lowest rated dog received a .1 but is really a 10. See the bottom of the page for tweet details

Insight #2: Most Retweeted Photo With a simple line of code (retweet_count.max) I have found out that the most retweeted photo is a dog realizing he can stand in the water.
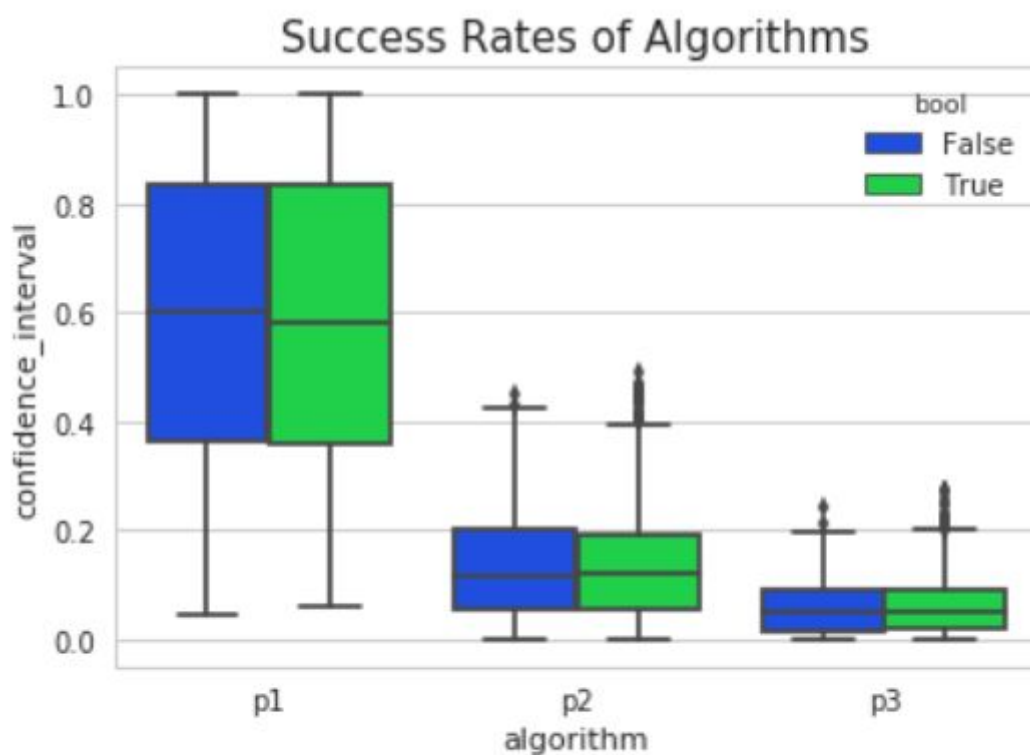
Insight #3: Source Type Counts The last analysis that I performed was a value_count of the different types of sources that tweets are coming from. The top number of tweets are coming from an iPhone (98%) followed by Twitter Web Client (1.5%) and lastly, TweetDeck (0.4%)

Visualizations:

Correlation Between Retweet and Favorite Count
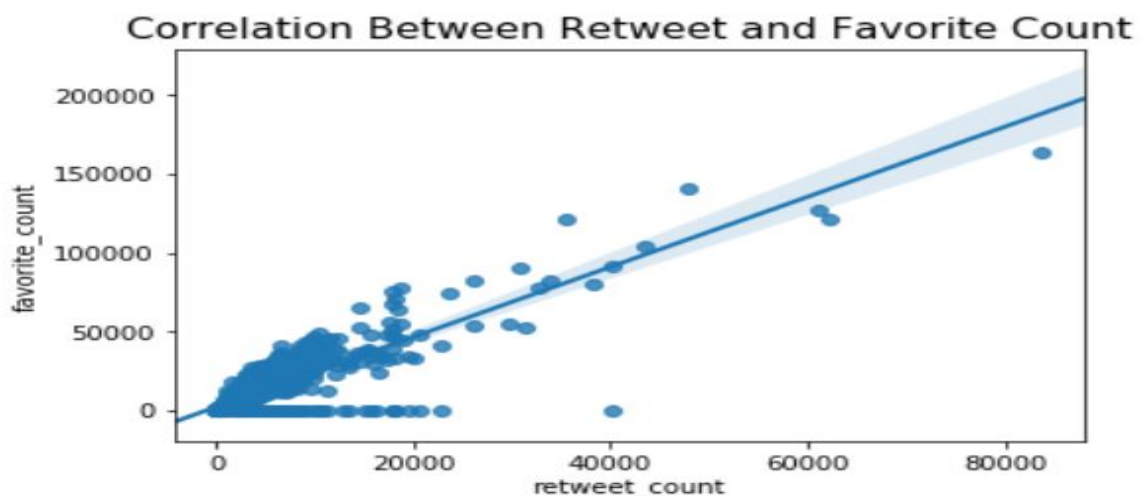
The first visualization analyzes the correlation between a retweet and a favorite count. By using regression to plot the best of fit line, we can see that for every retweet, there are multiple favorites of the tweet. Thus, when if we try to find the max count for both retweet count and favorite count, they will be the same tweet.

Based on confidence intervals, p1 had the highest percentage of confidence but there is a less amount of predictions that came out to be true. With p2, there is less confidence percentage (10-20%) but more predictions that came true. Lastly, p3 had the lowest confidence percentage(5-10%) and had about the same amount of true and false outcomes. The results show that there is approximately an equal amount of True and False outcomes for each of the algorithms, suggesting that each of the algorithms have a success rate of the same.

<matplotlib.figure.Figure at 0x7fbfee53e128>