Insights from the WeRateDogs Twitter Account

After cleaning the individual datasets from *wrangle_act.ipynb*, some visualizations were made using Plotly to gain some insights into the master dataframe. While many variables could've been considered for analysis, the ones that were chosen for this project pertain to dog popularity by breed and the relationship between rating and favorites count.

Figure 1 illustrates the favorites share of every tweet on the WeRateDogs Twitter page. While this makes it obvious that there's a wide range of species that are liked on this account, the biggest takeaway is that golden and lab retrievers own about 20% of favorites.

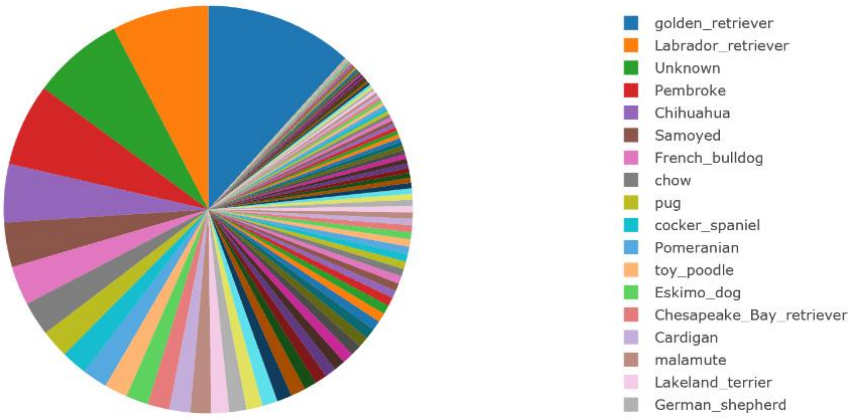

Figure 1. Percentage of Favorites by Breed

What this chart fails to consider, however, is that these two species make up about 15% of posts on the WeRateDogs account. Based on volume alone it makes sense that these two species garner a lot of favorites. When we look at favorites and retweets on a per tweet basis for the 10 most popular dogs (Figure 2), the rankings shift.

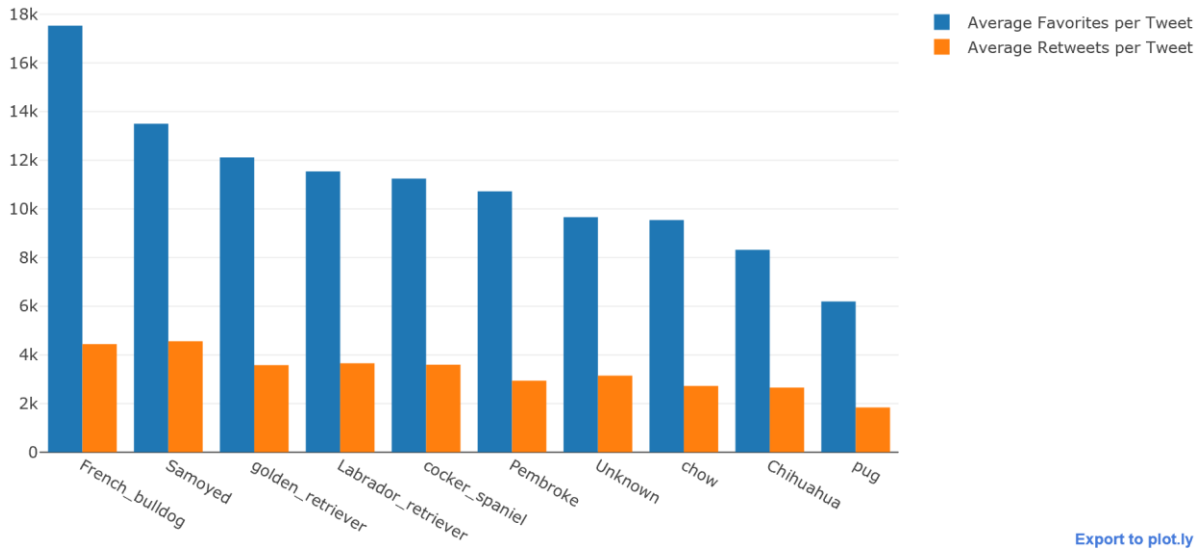

Figure 2. Tweet Metrics by Breed

French bulldogs lead the pack on a per tweet basis, getting an average of 17,500 favorites and 4,400 retweets per post (Samoyeds edge them out on retweets, at 4,500). The goldens and labs are still in the top five, with both right around 11,500 favorites and 3,600 retweets per tweet.

An additional observation in popularity had to do not with breed, but with rating. Rating factor is defined as a rating's numerator divided by the denominator, so a dog with a 12/10 rating corresponds with a rating factor of 1.2. Figure 3 shows this relationship. Each point represents a tweet, and the red markers show the median number of favorites for each rating factor from 0 to 1.4. A few cases had unique rating factors (0.975 or 1.127 for example) that were not considered in the median plot. There were some outliers with greater than 50k favorites or 1.4 rating factors, but the axes were scaled appropriately to better visualize the trend.

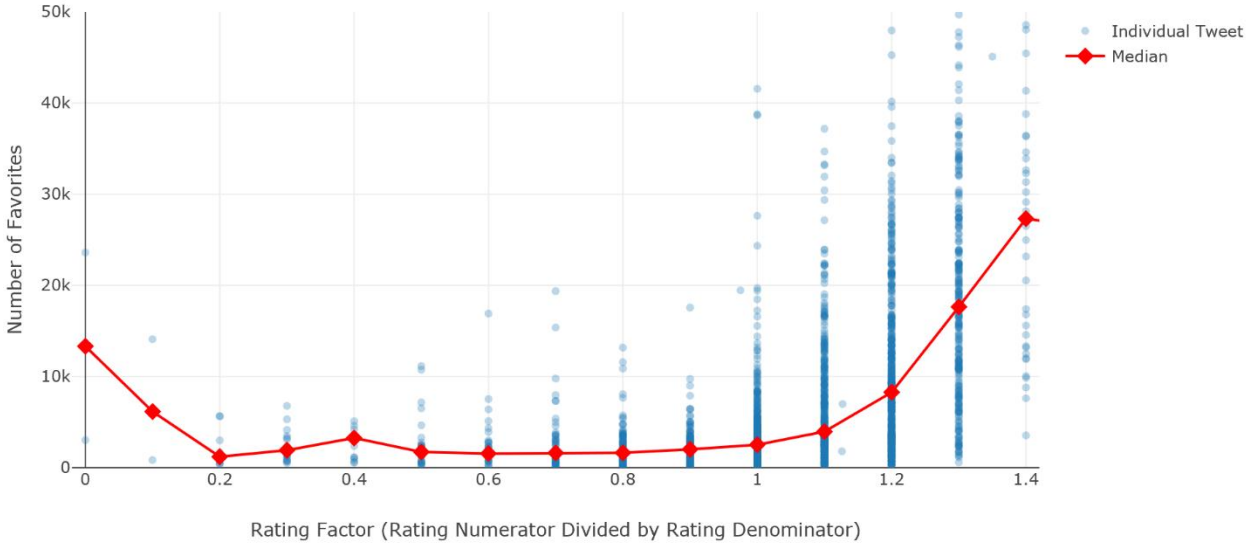

Figure 3. Rating Factor's Effect on Favorites

This plot illustrates a few things. First, it makes it easy to see that most of the tweets have ratings in the 1 to 1.3 range. Second, there seems to be an exponential increase from about 0.5 to 1.4 in number of favorites. However, seeing that there's significant variance, it would be foolish to automatically expect tens of thousands of favorites just for posting a tweet with a 12/10 rating. It seems clear that tweets with rating factors greater than one get more favorites than those with less than one, so WeRateDogs probably won't be straying from their unusual rating system anytime soon.