WeRateDogs Analysis

After successfully cleaning my dataset, it was time to start analyzing the data. I was curious to see which dogs increased the Twitter account's engagement. In order to see and analyze this information, I had to use the Twitter API and the image prediction file to gather his information. My data is based off the data I gather and clean. This data is not all of the tweets from WeRateDogs, it is just a sample size. Also note that the breed is based off a neural network prediction and may not be 100% accurate. I created my visualizations in Tableau. I will put pictures of the visualization along with a link to the Tableau Public Worksheet.

To start off, I took a look at how many different breeds were represented. In our sample size, we have 111 different dog breeds. Pictures of Golden Retrievers were tweeted the most, 133 times. The next highest was Labrador Retrievers with 91 pictures. There were also 7 breeds who only were tweeted about once.

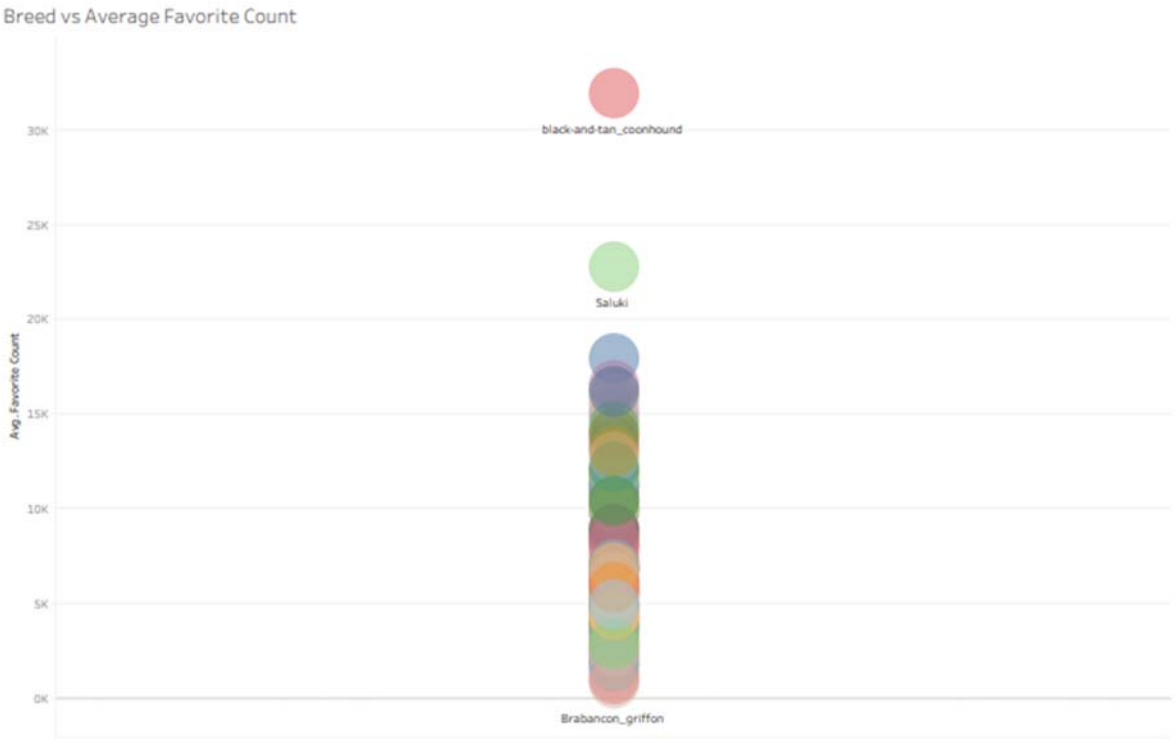
```
In [74]: df.Breed.value_counts()
Out[74]: golden_retriever                  133
         Labrador_retriever                 91
         Pembroke                           84
         Chihuahua                          76
         pug                                51
         Samoyed                            38
         chow                               38
         Pomeranian                         36
         toy_poodle                         33
         malamute                           28
         cocker_spaniel                     26
         French_bulldog                     26
         Chesapeake_Bay_retriever           22
         miniature_pinscher                 21
         Staffordshire_bullterrier          19
         Siberian_husky                     19
         Eskimo_dog                         18
         German_shepherd                    18
         Cardigan                           17
         Maltese_dog                        17
         beagle                             17
         Shih-Tzu                           17
         Rottweiler                         16
         Shetland_sheepdog                  16
         Italian_greyhound                  15
         Lakeland_terrier                   15
         Pekinese                           13
         American_Staffordshire_terrier     13
```

I then looked at the highest favorite and retweet counts for an individual tweet. They happened to be on the same tweet of a Labrador Retriever. This was WeRateDog's best tweet in this sample size. Impressive favorite and retweet counts for this doggo.

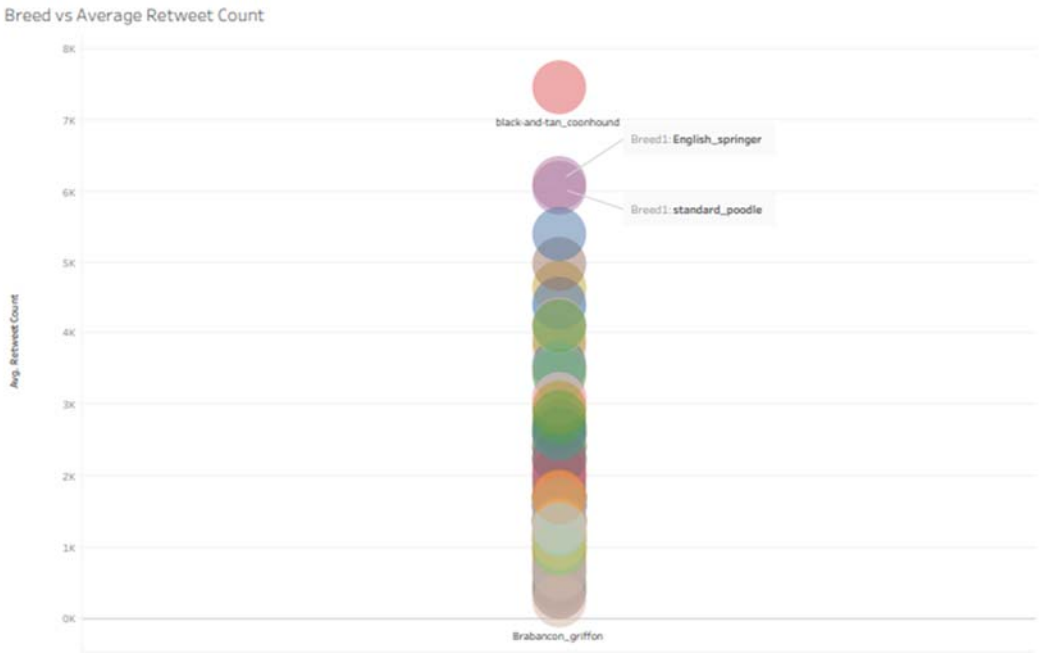
tweet_id	timestamp	source	text	rating_numerator	rating_denominator	dog_stage	Breed	Confidence	p1_dog	favorite_count	retweet_count
)360020481	2016-06-18 18:26:18	Twitter for iPhone	Here's a doggo realizing you can stand in a po...	13.0	10	doggo	Labrador_retriever	0.825333	True	160722.0	80421.0

* Tweets are only a sample size of WeRateDogs

Next, I used .astype to change the breed column to a category and then used a groupby function to find the mean favorite and retweet counts for each breed. This result was a little more interesting than the last one. I found that the highest average favorite count belonged to the black and tan Coonhound which has 31,963 favorites per tweet*. The highest average retweet count also belonged to the black and tan Coonhound which averaged 7,465 retweets per tweet*. I created a visualization, using Tableau, for this. The visualizations are below.



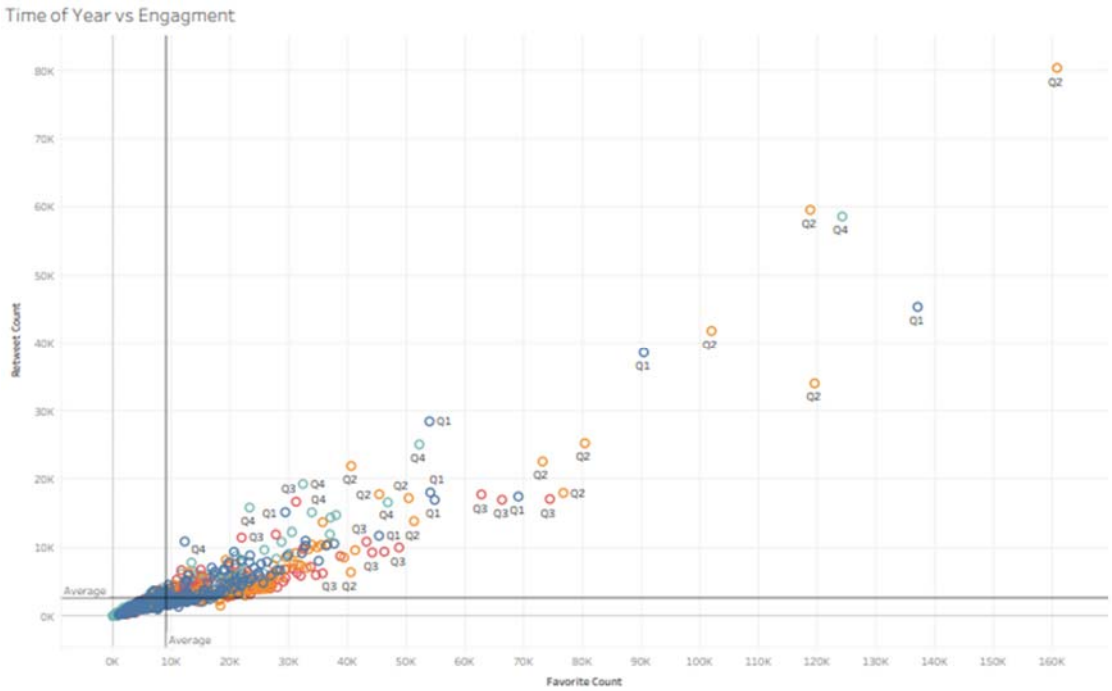
* Tweets are only a sample size of WeRateDogs



I also looked at whether the dog rating had an impact on the engagement. I took the numerator divided by the denominator to create the 'ratio'. I used Tableau to create this visualization and saw that typically the tweets with the ratio 1.3 were liked and retweeted more. I used colors to represent the ratio. Green is a smaller ratio and red is a higher ratio. The visualization is below.

* Tweets are only a sample size of WeRateDogs

Finally, I used Tableau Desktop to look at the quarters each tweet was posted in (Q1, Q2, Q3, and Q4) to see if tweets engaged the audience more or less depending on a certain time of the year. It seems like Q2 had the tweets with some of the highest favorite and retweet counts. This means it might be best to tweet the most during the spring and early summer months.



Here is the link to my Tableau Workbook. You can get more detailed views if you would like!

* Tweets are only a sample size of WeRateDogs