# Analysis

Once all the data was collected and cleaned, we were able to begin analyzing the data. I chose to try and tackle the following 3 questions:

1) Look at the number of tweets for each dog to see which dog was tweeted the most
2) Find out which dog get's the highest rating on average to find the "cutest" dog
3) Look at the average ratings given overtime (regardless of breed) to see how people's opinions on dog cuteness have changed over time

1) To answer the first question we first have to count how many times each dog came up according to the p1, p2, and p3 values. If p1_dog != True, meaning the algorithm did not detect a dog, I would use the p2 value instead of p1. If we then found the same was true for p2 (p2_dog != True) after going through p1, then we would use p3. If all 3 p values could not detect a dog from the image, we ignored the entry all together. Then I summed up the data and plotted the results in a bar graph shown below:


Tweets vs Dog Breed
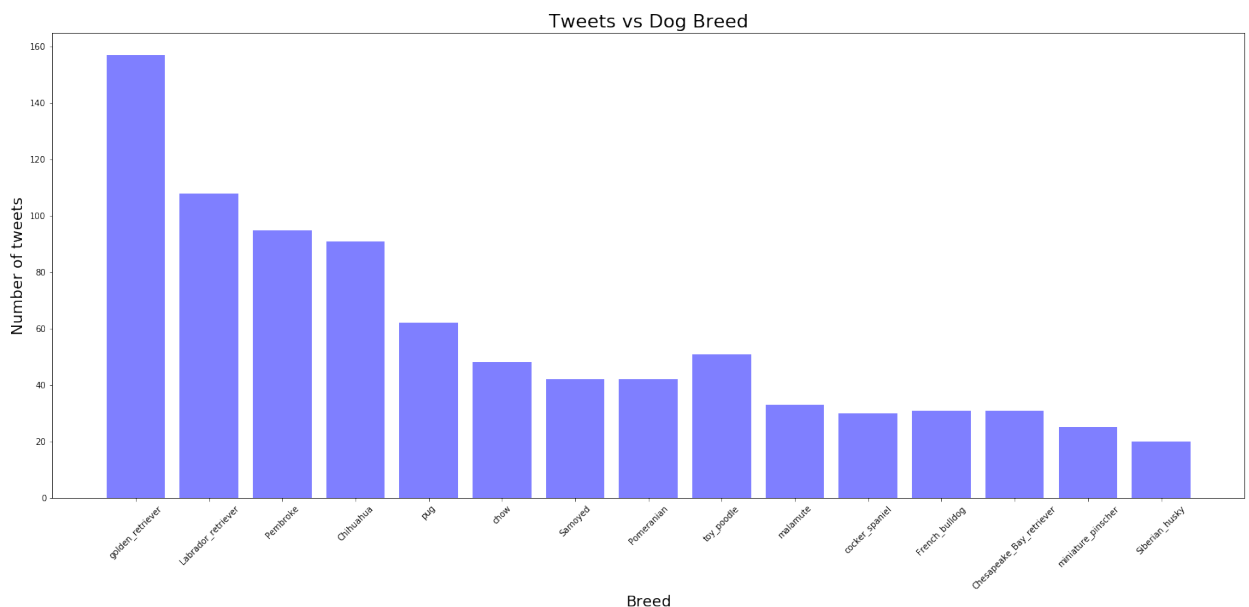
As the plot above shows, the golden retriever has the highest number of tweets by a substantial margin, sitting at a grand total of 157 tweets across our dataset

2) To answer the second question we would analyze the data in the same way as the first question, except this time we would need to calculate an 'overall_rating' for each entry. To do this, I divided rating_numerator by rating_denominator and then plotted the top 10 results in a bar graph shown below (2 entries had to be removed as they were very clearly outliers and would have skewed the dataset):

Dog Breed vs Average Rating
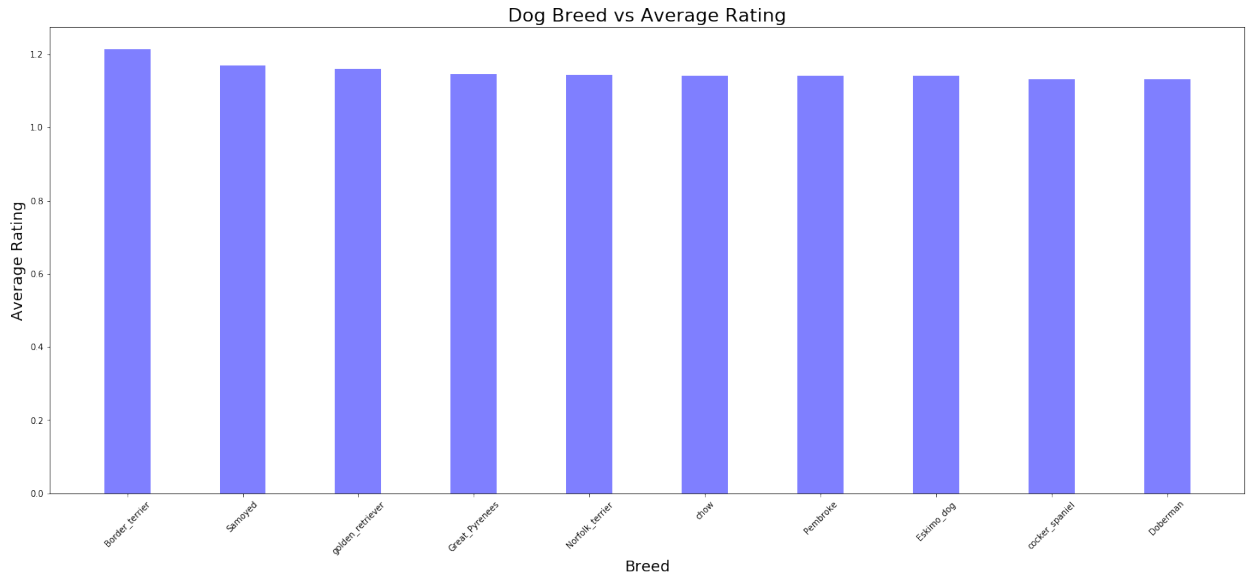
After plotting, we see that the highest rated dog according to our specifications was the Border terrier with a average rating of 1.214 from 7 entries

3) The last question could be answered by simply plotting the 'timestamp' against the 'overall_rating' calculated in the last question.
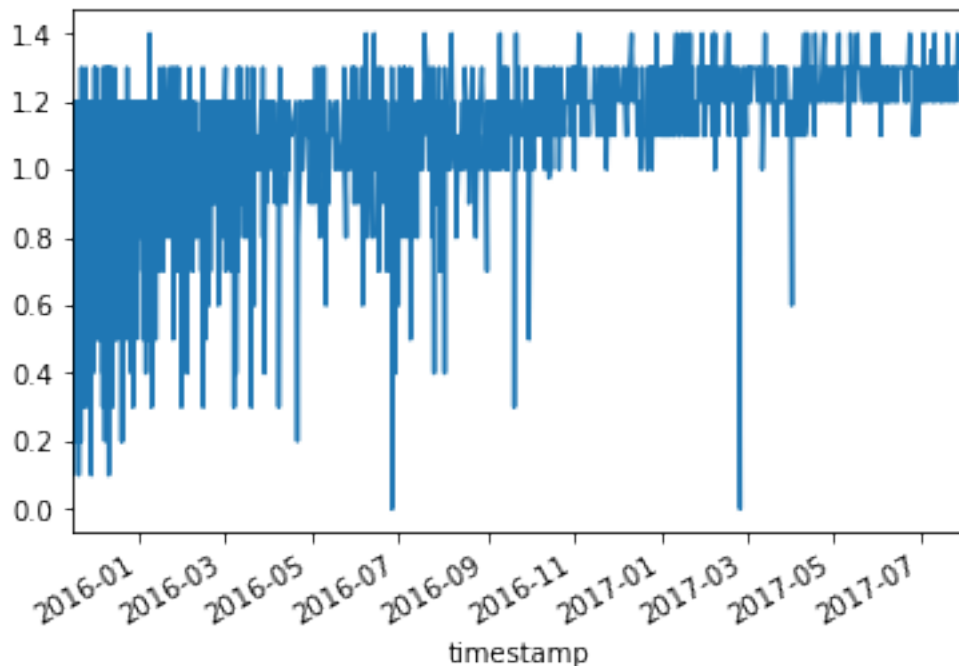
After removing the 2 outliers found in part 2, we see that there has been a general positive trend towards the ratings people give regarding their tweets