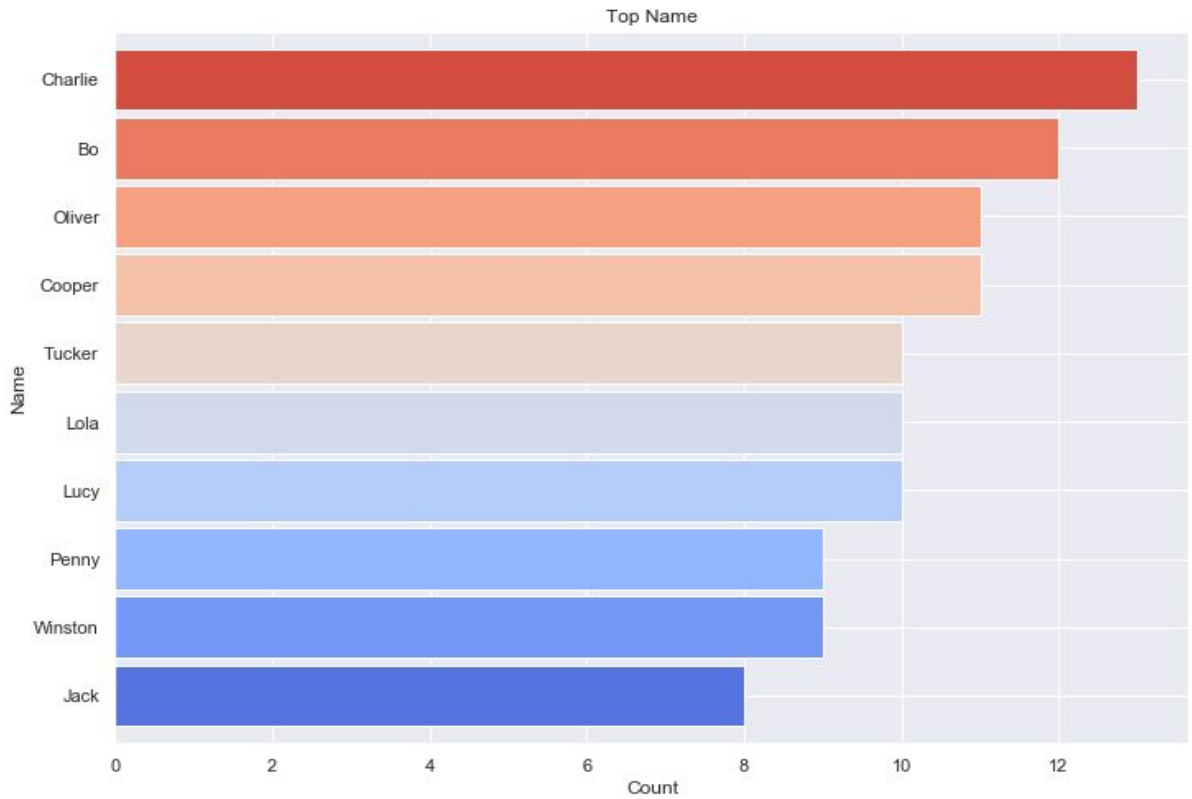
Analysis and Visualization of the Dog Rating Tweet data

Prepared by Debarghya Biswas

Motivation

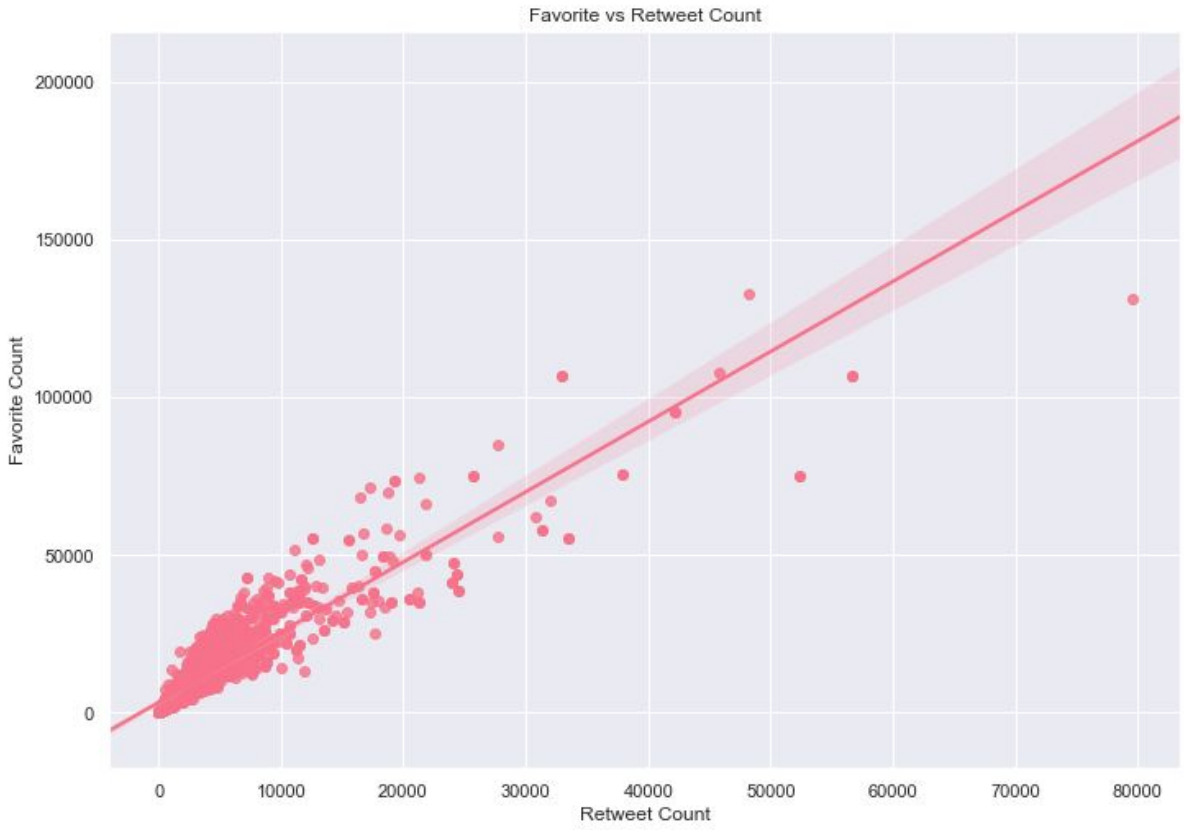
WeRateDogs is a twitter account dedicated mostly to rating a dog's picture on the internet. It has been a popular account, not only due to the dog picture it posted but also because of the unique rating system that they are using. In this analysis, we are trying to gain interesting facts among the dogs captured in the tweets and also the result of the image prediction made from the dogs, which generates 3 most popular tags for each dog.

What is the most popular name for the dog in the tweet data?



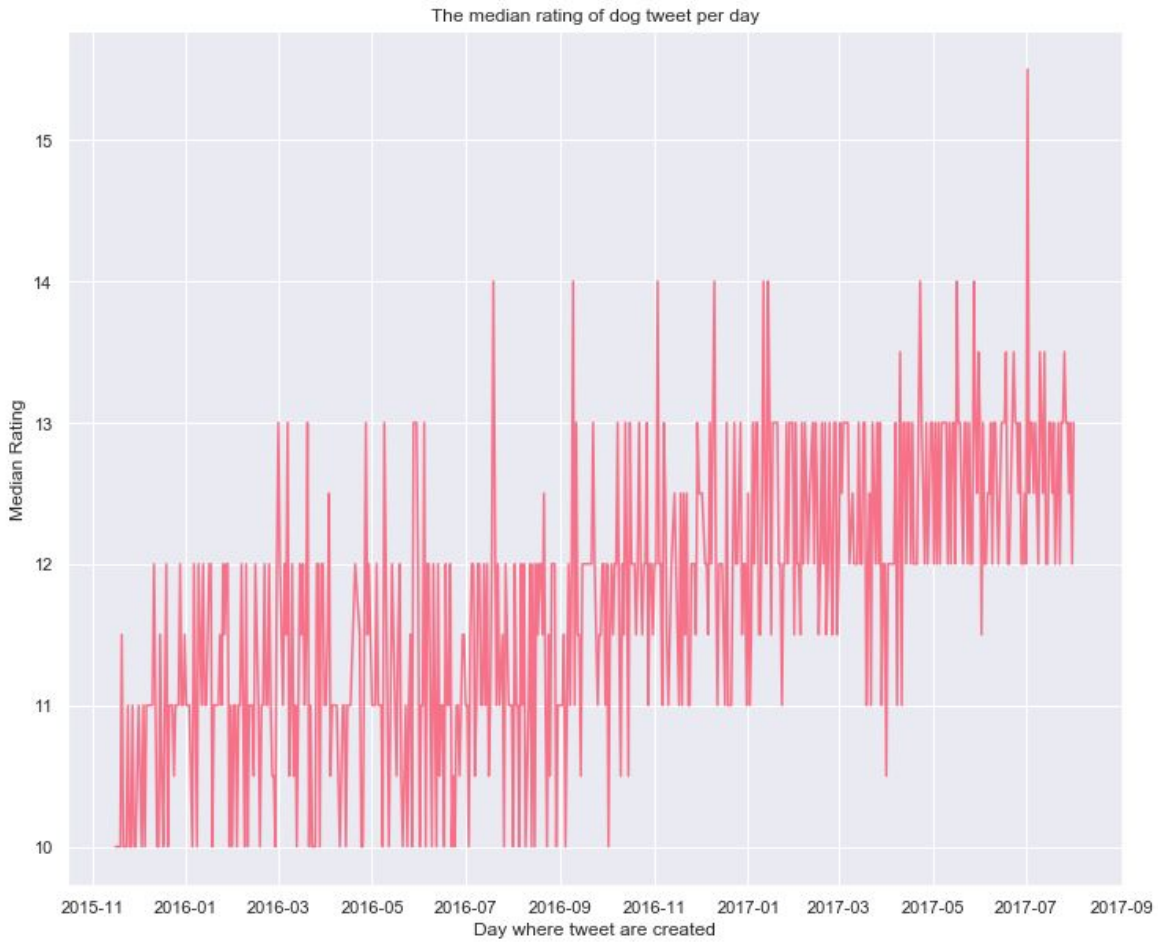
As shown in the figure above, the most popular name seems to be Charlie followed by Bo. It could be because of the fame of Charlie the Dog, an animated cartoon fictional character in the Warner Brothers Looney Tunes series of cartoons.

Is there a relation between the favorite count and retweet count?

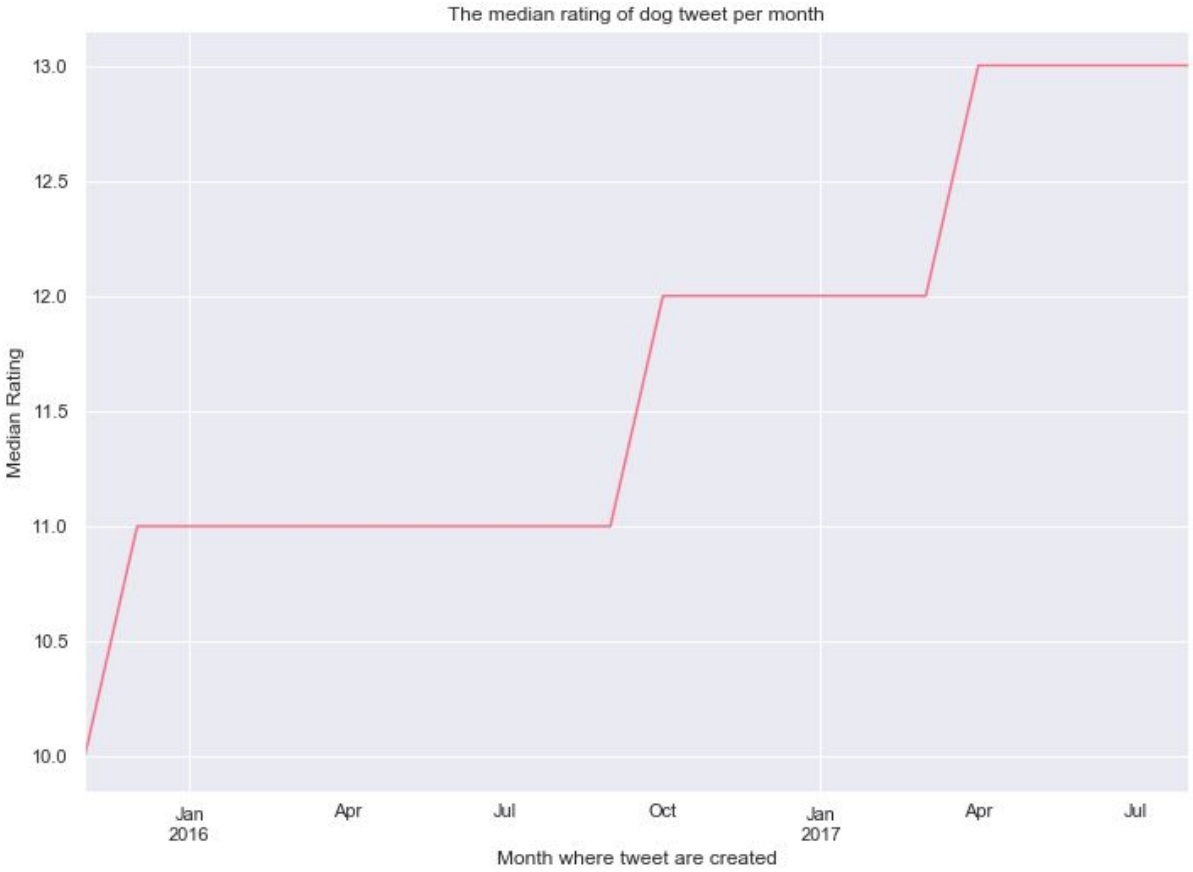


It seems like there is a linear relationship between the favorite count and the retweet count. This makes sense since once the tweet gets retweeted, it gets more exposure, thus have a higher chance to get favorited and retweeted.

Is there any pattern for the rating given in the tweet?

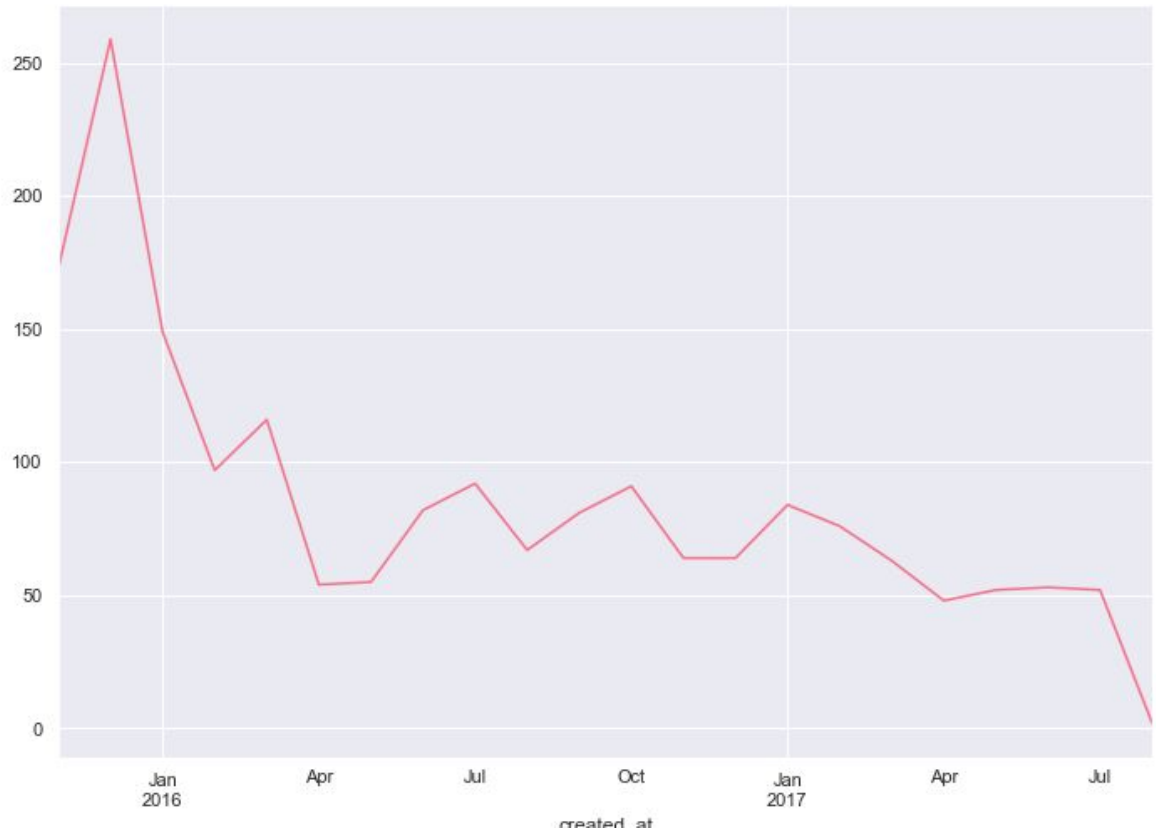


From the two plots above, it seems like the earlier tweet has a lower dog rating than the latter dog rating.



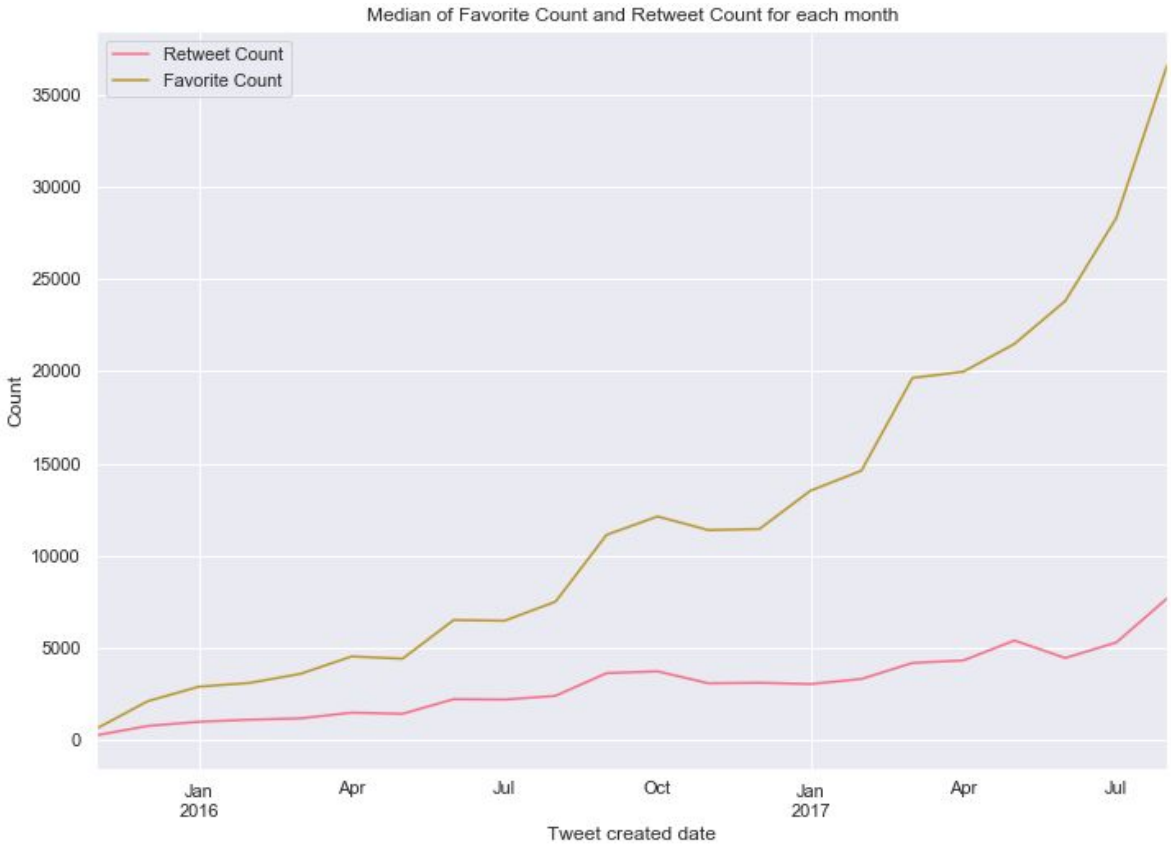
From the above, it seems like the rating given to dogs increases as time goes by.

Does @weRateDogs tweets more dog rating as it gains popularity?



Ignoring the sharp drop in July 2017, which might be due to the time we gather data, the plot above shows that @weRateDogs does not become more active in rating dogs as the month goes by. Actually, the tweet count has a tendency to decrease.

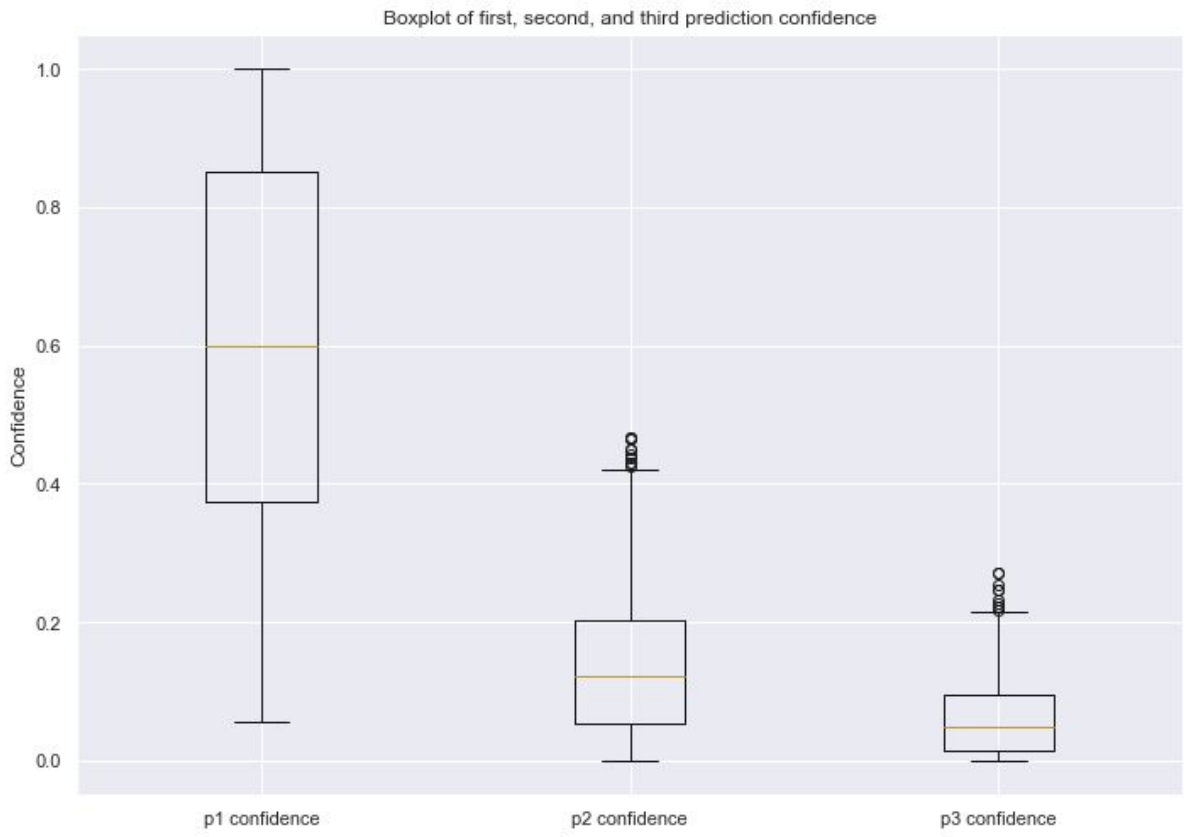
How about the retweet count and the favorite count per month?



To measure the account popularity, it is better to use the median of the retweet count and favorite count instead of the sum, as median is less affected by outlier and the sum may be affected by the number of tweets made per month, which differs each month.

From the plot above, we can see that even though the number of dog rating tweet made is reduced, the retweet_count and favorite_count increases. We can conclude that reducing the number of tweet does not correlate directly to the retweet count and favorite count.

How good is the image prediction?



From the above, we can see that first prediction tend to have higher confidence than the second and third prediction. We can also see that for the first prediction, more than 50% of the image has more than 50% confidence. It shows that generally, the first prediction is a good prediction.