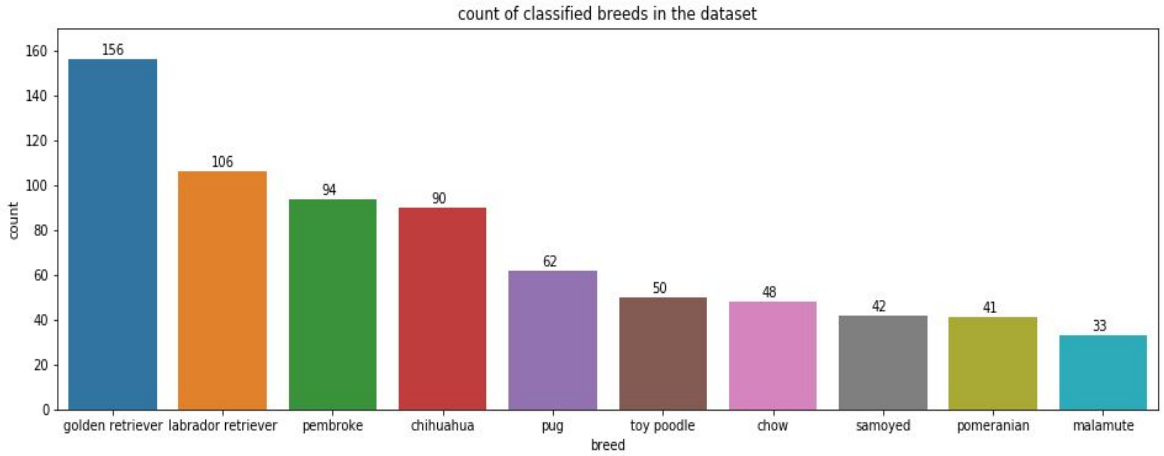
Analysis and Visualization of the Dog Rating Tweet data

Prepared by Akash Yadav

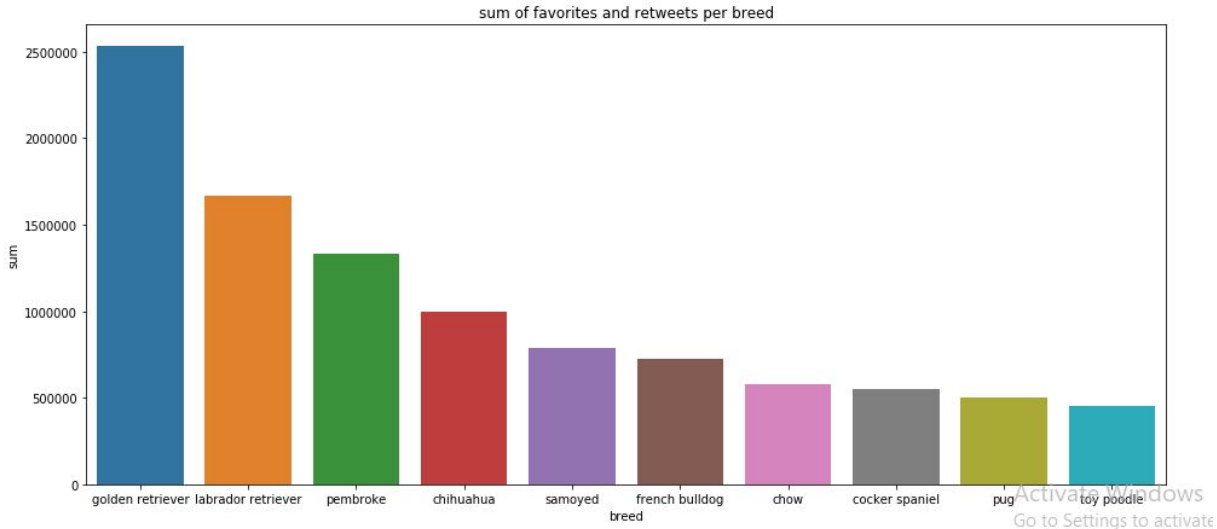
What is the most popular name for the dog in the tweet data?

As shown in the figure above, the most popular name seems to be golden retrievers .



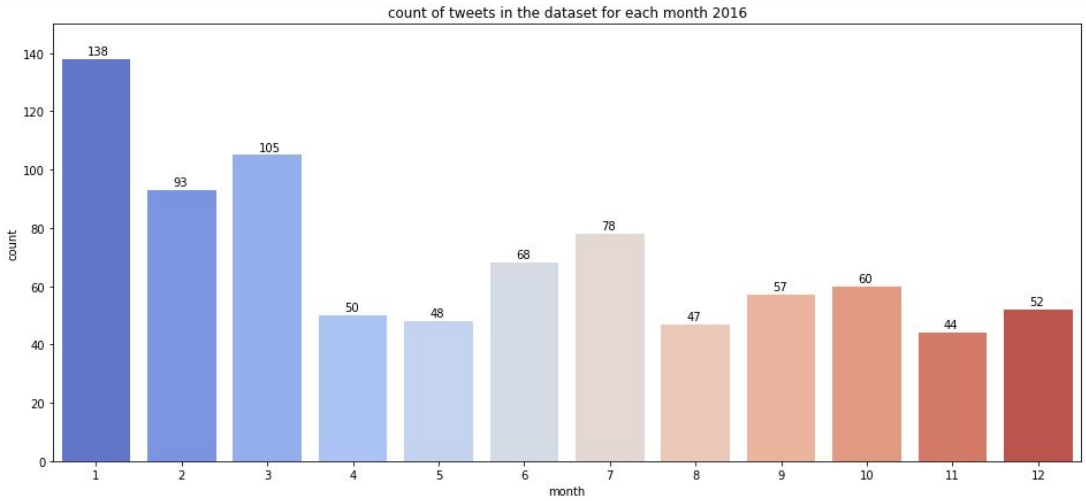
Is there a relation between favorite count and retweet count?

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



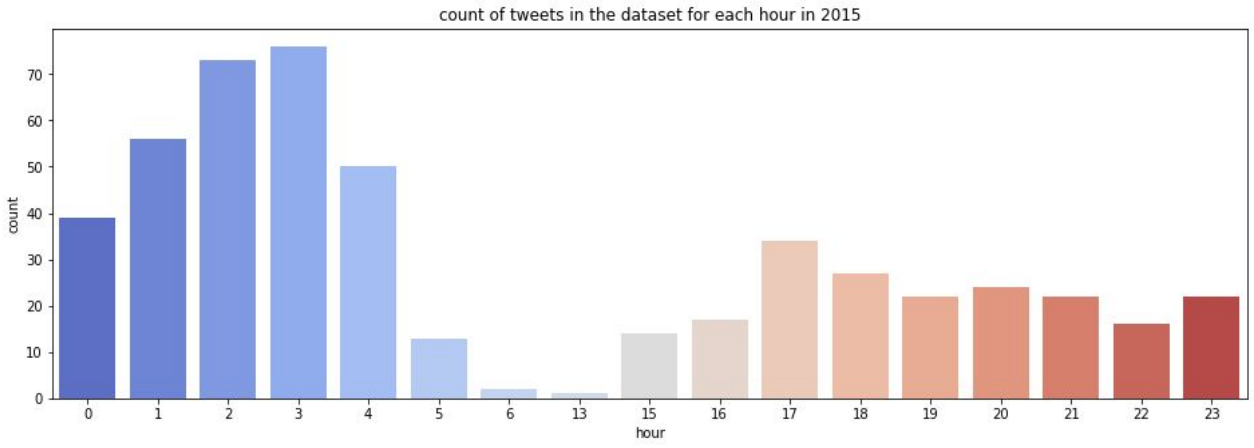
How about the retweet count and the favorite count per month?

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 tweets made is reduced, the retweet_count and favorite_count increases. We can conclude that reducing the number of tweets does not correlate directly to the retweet count and favorite count. Below the mention the image.



Is there a pattern visible in the timing of the tweets?

As we can see in this graphs of 2015 and 2016, the most posts are during the night between 0 (included) and 5 'o'clock. between 4 and 15 'o clock there are a very small number of tweets, in fact there is no single tweet for the hours 7 - 12 and 14 o' clock. There are some tweets after 14 'o clock, but not as much as between 0 (included) and 5 'o'clock. We can check the results.



How good is the image prediction?

For the images it seems pretty stable. There are months where there are more and months where are less posted images, but overall there is no clean up- or downtrend visible.