

# **Exploratory Data Analysis Report**

## **Introduction**

The wrangled master dataset was explored and analyzed to gain insights about WeRateDog.

The following questions were answered:

1. Which tweet of WeRateDogs has the most retweets and favorites?
2. Which dog breed is the most popular in WeRateDogs overall based on retweet counts, favorite counts and rating?
3. How did WeRateDogs grow over the time?
4. Are there any correlation between rating, retweets and favorites and dog stage?

## **Exploratory Data Analysis**

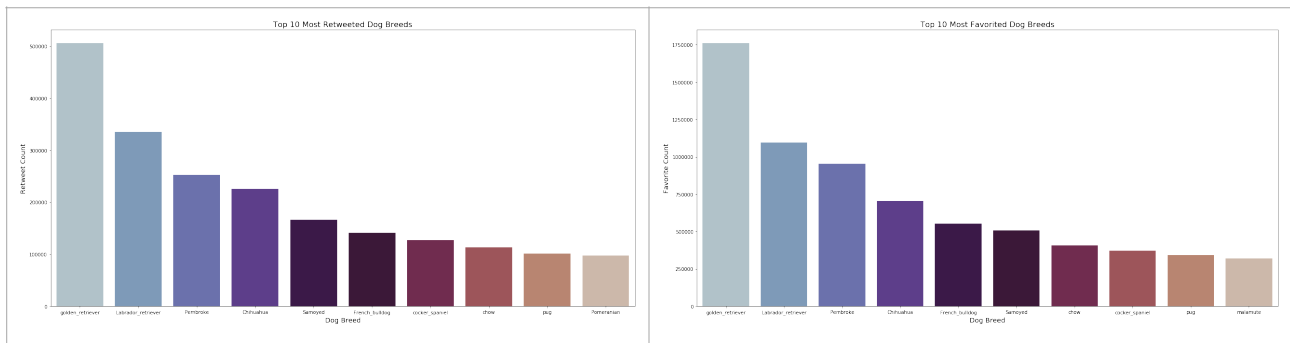
### **1. Which tweet of WeRateDogs has the most retweets and favorites?**

The tweet with the most number of retweets and favourites in the dataset is a 30 seconds video of a doggo labrador retriever swimming standing in a pool, which was retweeted more than 75,000 times, favorited more than 150,000 times and rated with the highest score of 13/10. This is a screenshot of the video.

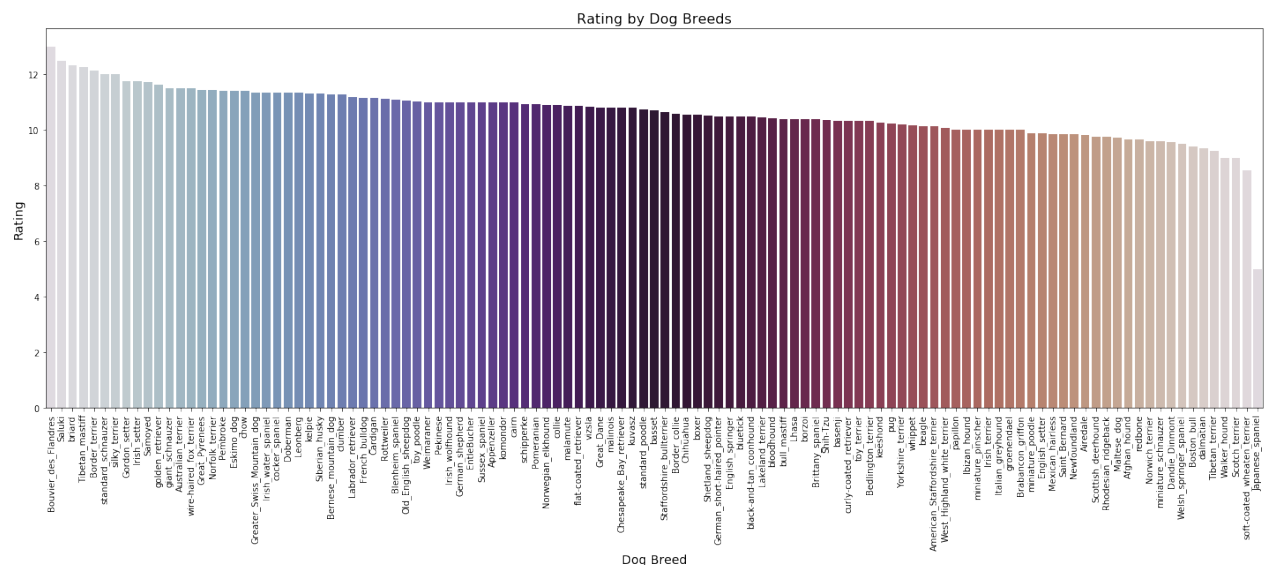


Source: [https://pbs.twimg.com/ext\\_tw\\_video\\_thumb/744234667679821824/pu/img/1GaWmtJtdqzZV7jy.jpg](https://pbs.twimg.com/ext_tw_video_thumb/744234667679821824/pu/img/1GaWmtJtdqzZV7jy.jpg)

## 2. Which dog breed is the most popular in WeRateDogs overall based on retweet counts, favorite counts and rating?

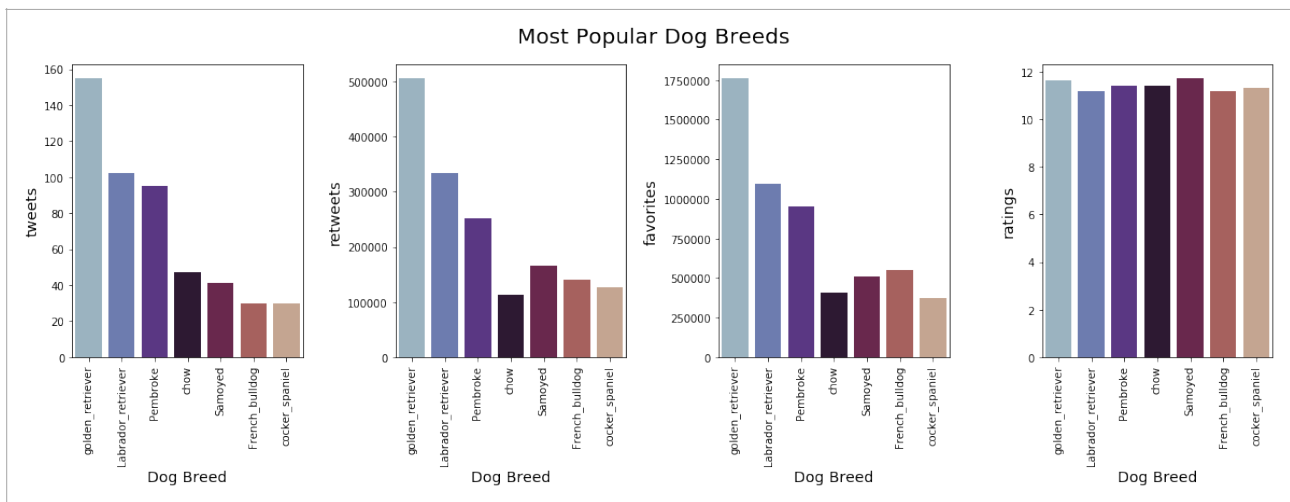


The two diagrams above show the Top 10 Most Retweeted Dog Breeds and the Top 10 Most Favorited Dog Breeds respectively. We can see that golden retriever is in the first place in both lists. However, it is not enough to conclude that golden retriever is the most popular dog breed because it could be due to the fact that the majority of the tweets are golden retriever images. Therefore, we need to consider the ratings as well.



This diagram shows the average rating by dog breeds. The ratings are pretty uniform with a 25th percentile of 10, a 75th percentile of 12 and an average of 11. To find the most popular dog breeds in WeRateDogs, dog breeds that have an above average rating as well as are in the both Top 10 Most Retweeted Dog Breeds and Top 10 Most Favorited Dog Breeds list will be considered.

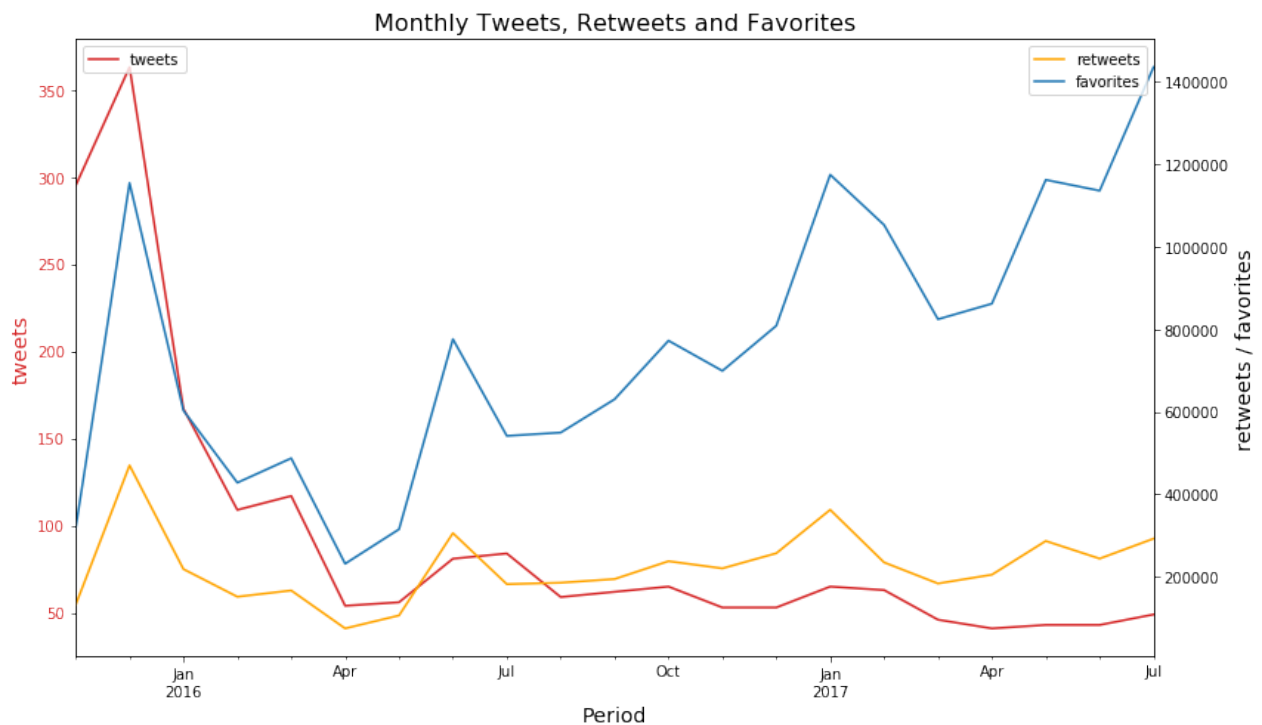
Based on the analysis, Pembroke, Labrador retriever, French bulldog, chow, golden retriever, Samoyed and cocker spaniel are the last seven standing dog breeds.



From the diagrams above we can see that all these dog breeds have very similar rating. However, golden retriever has significantly higher amount of tweets, retweets and favourites compared to other dog breeds. So, golden retriever is the most popular dog breed in the WeRateDogs Twitter account in 2017. Second place goes to Labrador retriever, followed by Pembroke, which is in the third place in term of number of tweets, retweets and favorites.

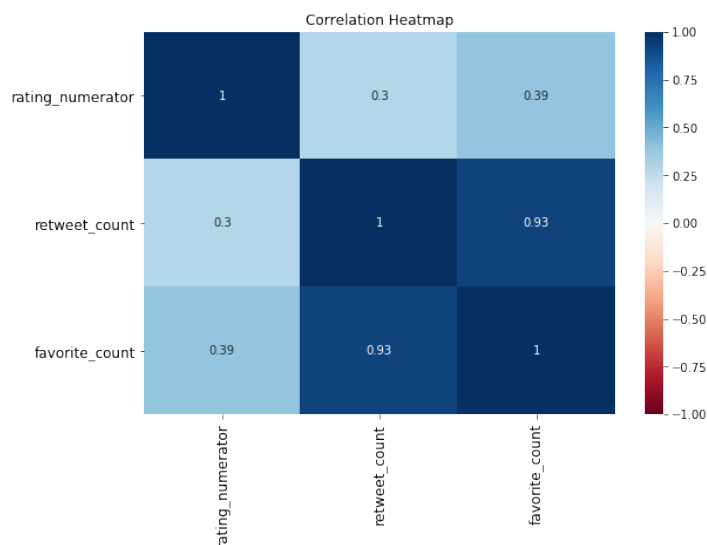
### 3. How did WeRateDogs grow over the time?

The monthly total number of tweets as well as the total number of retweets and favorite of the tweets were computed.

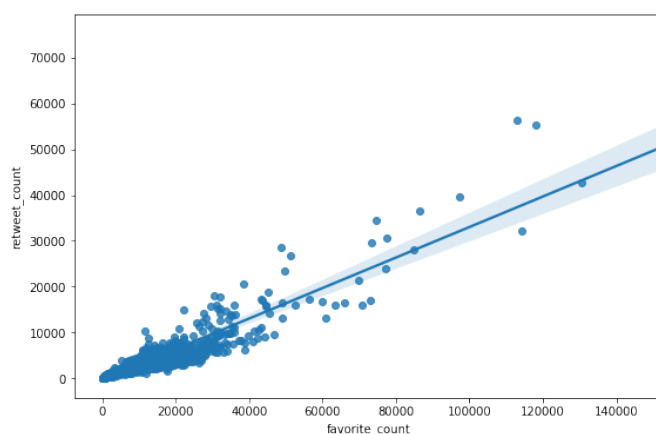


- The account was very active in the first two months when it was just started with more than 350 tweets in December 2015. However, there was a big decline in early 2016. The number of tweets per month was decreasing until it reached a relatively stable level with around 50-60 tweets per month.
- The number of retweets followed the similar trend as the number of tweets. It did not exhibit an obvious increase over a long period.
- Interestingly, in contrast to the two variables above, the number of favorites showed an upward trend even though the number of tweets has stopped increasing since 2016 July.

#### 4. Are there any correlation between rating, retweets and favorites and dog stage?



There is a very strong correlation between retweet and favorite count. The relationship between rating and favorite count is slight larger than the relationship between rating and retweet count.



Again, This regression plot shows there is a linear relationship between retweet and favorite count.

#### Linear Regression Models

To further explore the relationship between rating, retweets and favorites and dog stage, linear regression models were created.

Results:

- For every one unit increase in the rating, the number of favorite is predicted to increase by 2146.
- For every one unit increase in the rating, the number of retweet is predicted to increase by 581.
- A pupper is predicted to be less rated by 1.35 compared to a floofer, holding other variables constant, and p-value associated is statistically significant. However, this prediction is not really convincing because this categorical variable is imbalance, by having much more pupper than floofer in this dataset.