

Act Report

Yu Tao 2020/12/26

Introduction

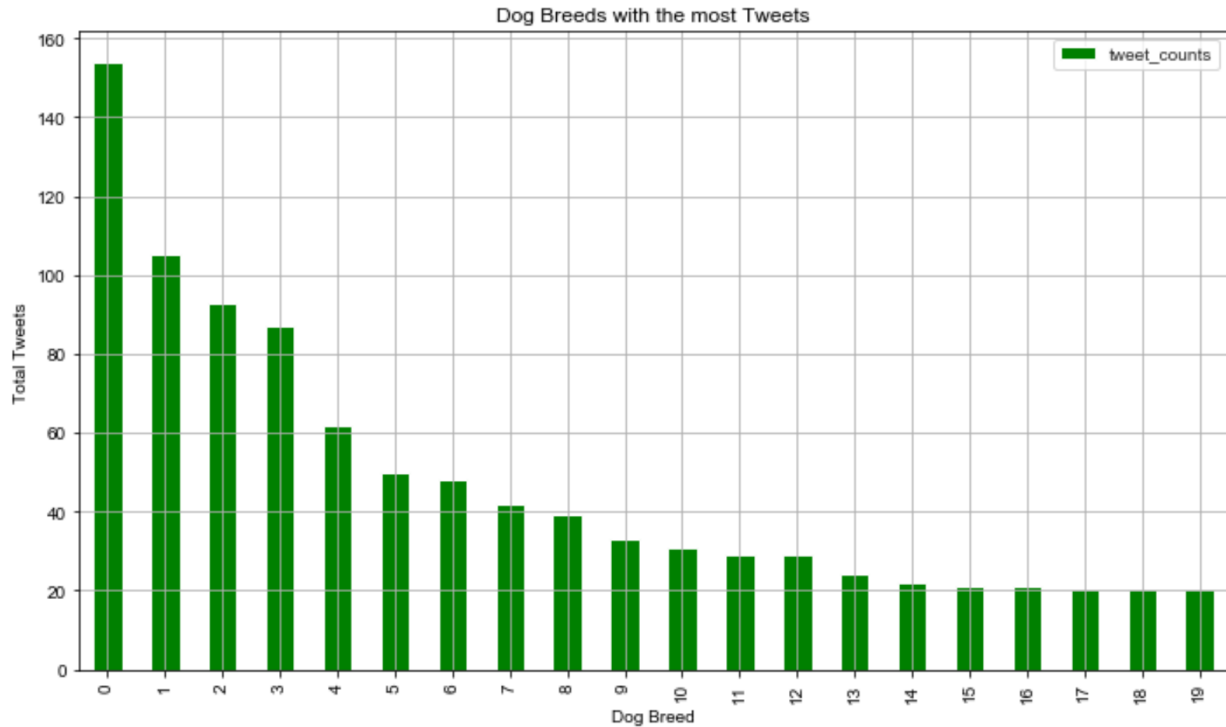
WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators are almost always greater than 10. 11/10, 12/10, 13/10, etc.

I wrangled WeRateDogs Twitter data to and created interesting and trustworthy analyses and visualizations, using three data frames I obtained from different sources: **(1) The WeRateDogs Twitter archive**, that contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of the tweets from WeRateDogs as they stood on August 1, 2017. **(2) The tweet image predictions**, that shows what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. **(3) Additional Twitter data**, that contains each tweet's favorite count and retweet count.

During the analysis of Twitter data from WeRateDogs, I provided various insights visualizations on the datasets, based on the following questions I came up with:

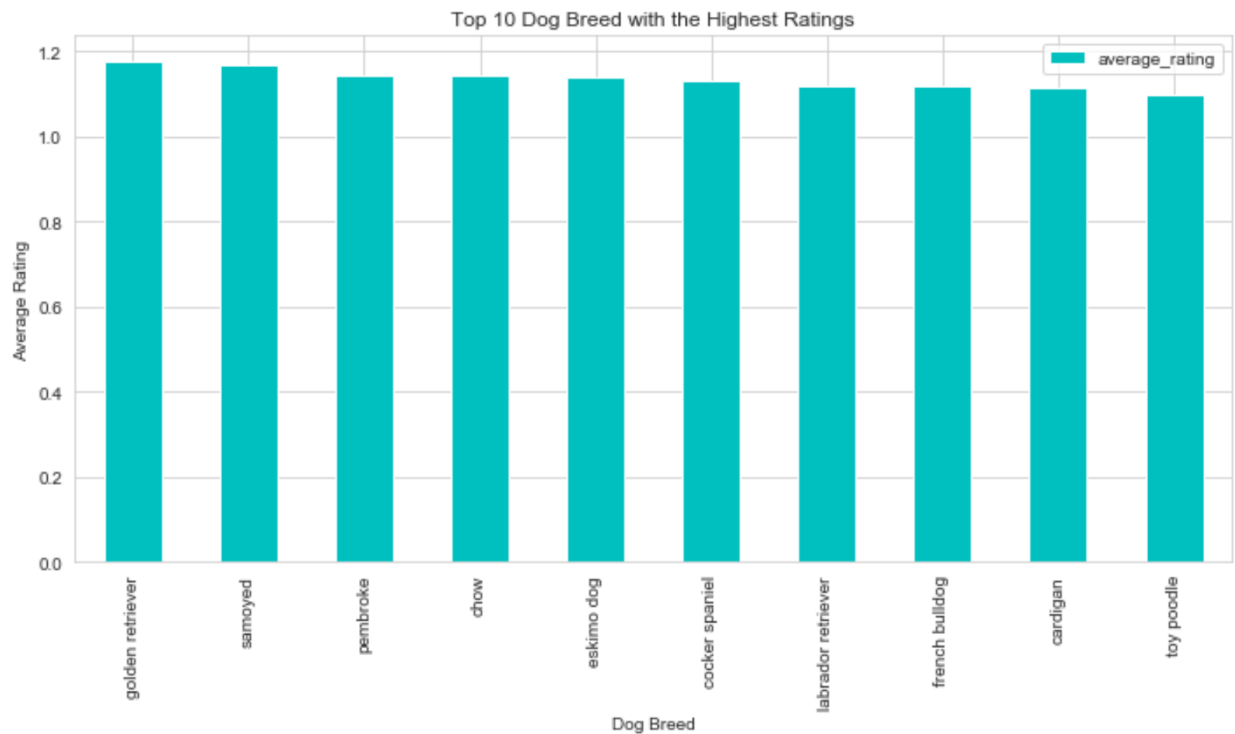
- What are the dog breeds with the most tweets?
- What are the dog breeds with the highest average ratings?
- What are the dog breeds with the most numbers of favorites or retweets?
- What are the distributions of the rating, favorite_count and retweet_count?
- What is the relation between the average rating and favorite/retweet counts for different dog breeds?
- What is the most liked/retweeted single tweet?
- What is the average favorite/retweet count on different dog_stage?
- What is the WeRateDogs account activity by year, month, day of the week, hour of the day?
- What is the distribution of favorite counts by month in each year?

What are the dog breeds with the most tweets?



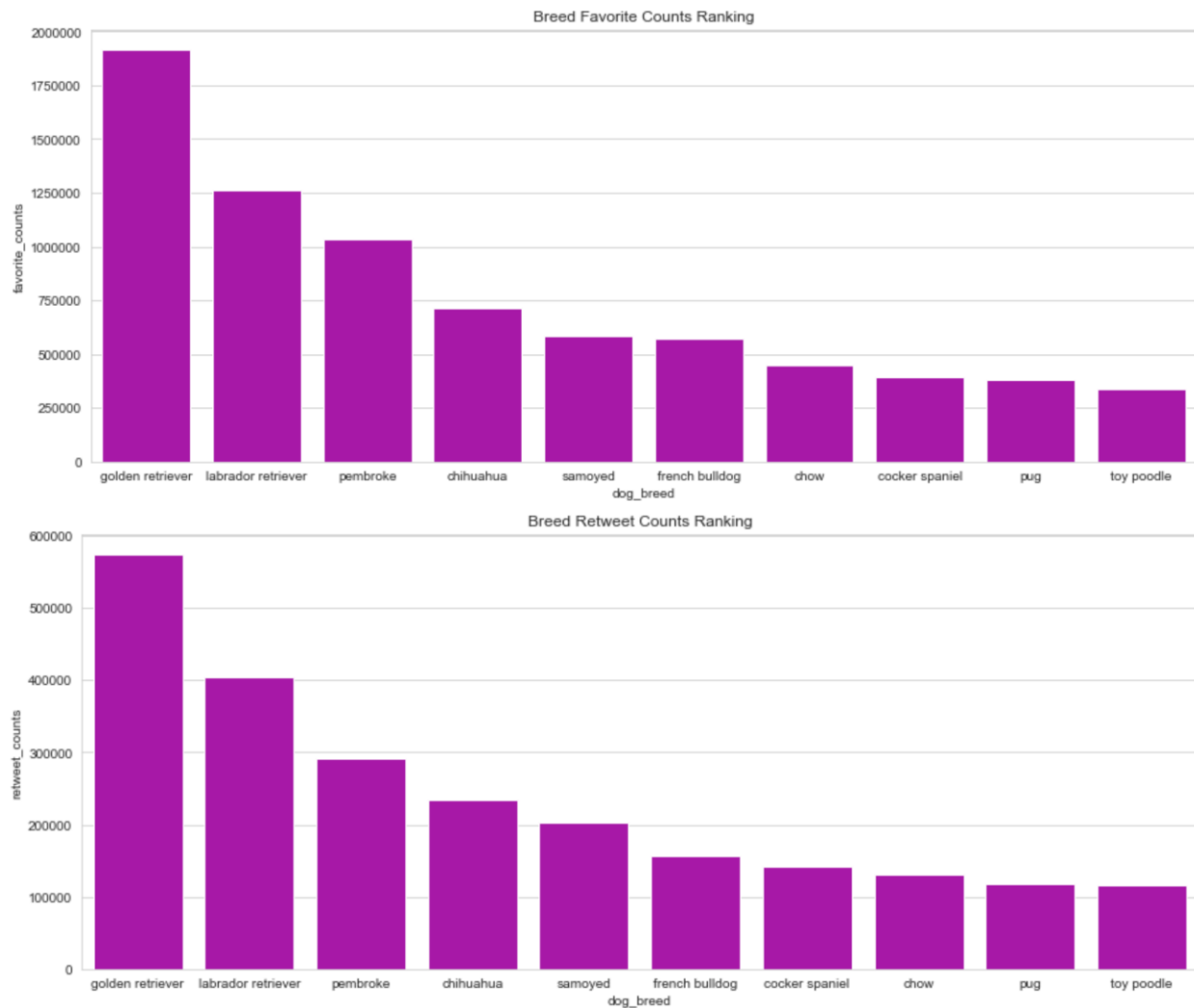
The above plot shows the 20 breeds with the most tweets, the top 3 of them are golden retriever (tweeted 154 times), labrador retriever (105) and pembroke (93).

What are the dog breeds with the highest average ratings?



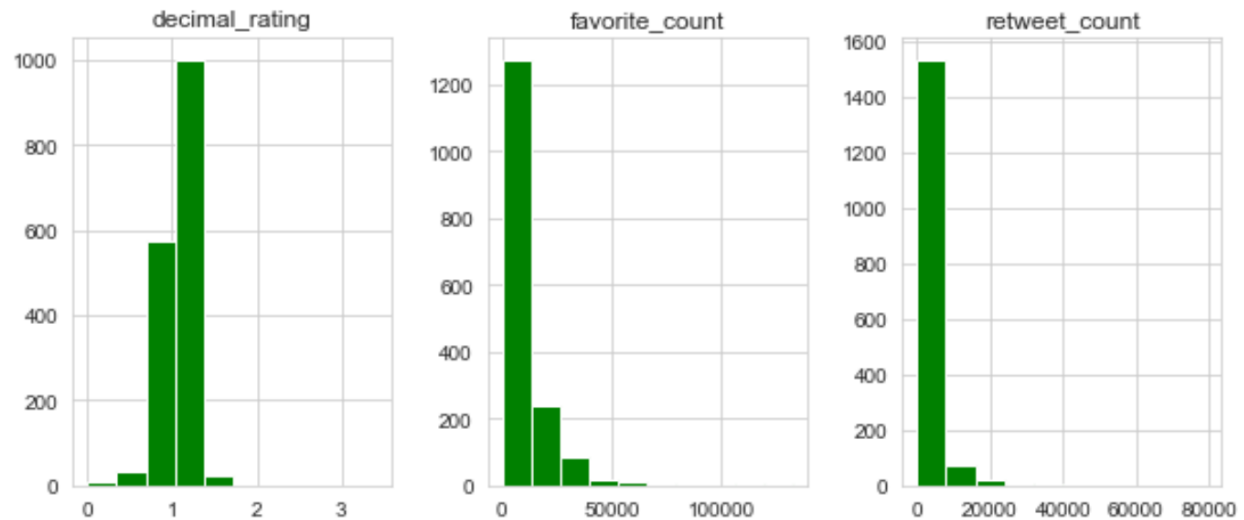
The plot above gives ten dog breeds with the highest average ratings, the top 3 breeds and ratings are: golden retriever (1.178), samoyed (1.169) and pembroke (1.144).

What are the dog breeds with the most numbers of favorites or retweets?



Golden retriever (1915008 favorites, 573953 retweets), labrador retriever (1264586 favorites, 405163 retweets), pembroke (1036756 favorites, 291262 retweets) are top 3 on both favorite and retweet lists.

What are the distributions of the rating, favorite_count and retweet_count?

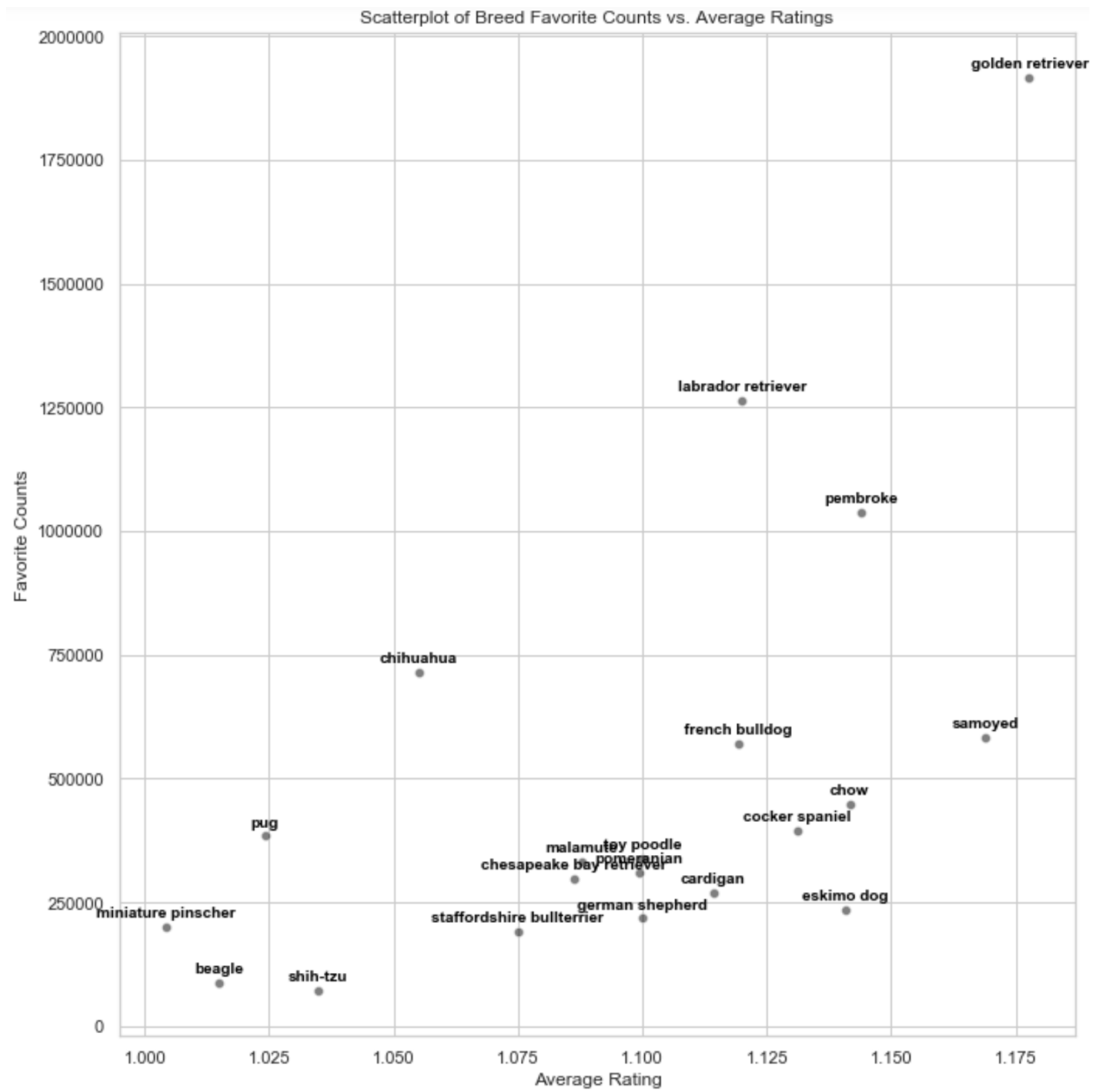


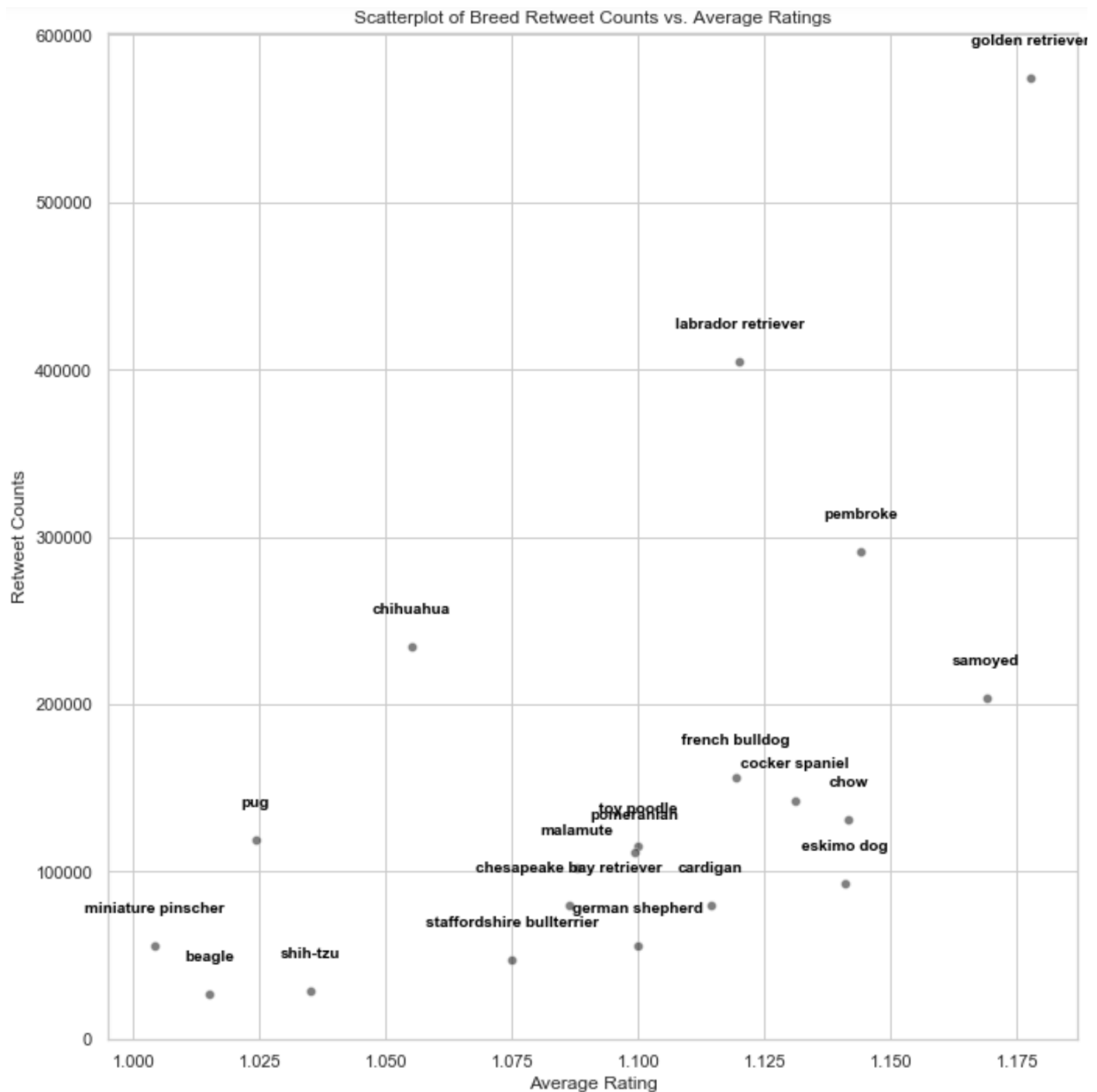
The average rating on dogs is 1.09, with a min rating of 0.00 and a max rating of 3.43. The interquartile range is 1.00 to 1.20.

The average favorite count of all tweets is 9371, with a min count of 81 and a max count of 132810. The interquartile range is 2243 to 11852.

The average retweet count of all tweets is 2853, with a min count of 16 and a max count of 79515. The interquartile range is 665 to 3271.

What is the relation between the average rating and favorite/retweet counts for different dog breeds?



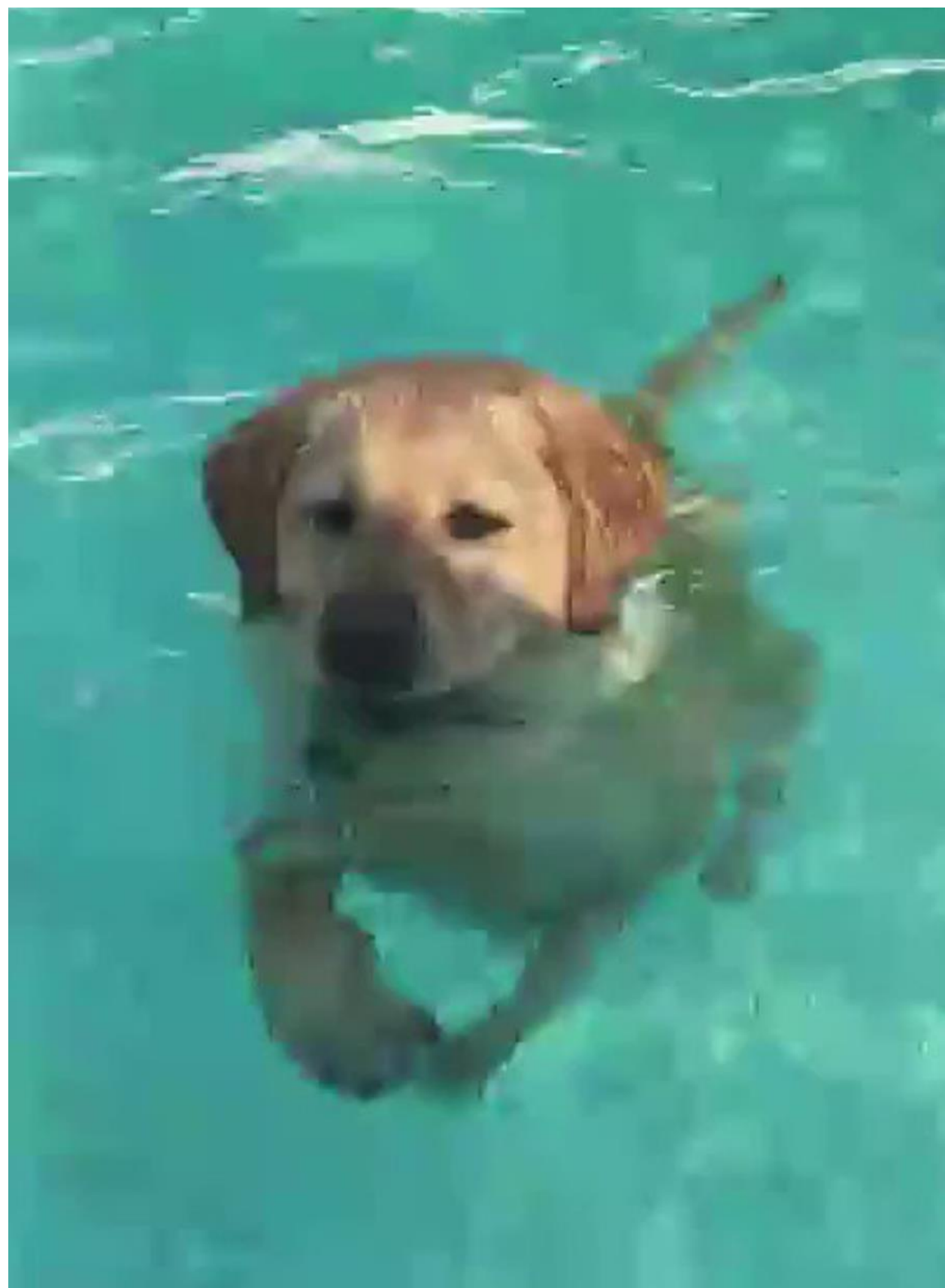


We can see from the two scatterplots above, the favorite/retweet count is roughly proportional to the average rating of the breed. the higher the rating, the more likely the post will get more likes and retweets. Among all the breeds, golden retriever is the most popular breed and gets far more favorites/retweets than expected.

What is the most liked/retweeted single tweet?

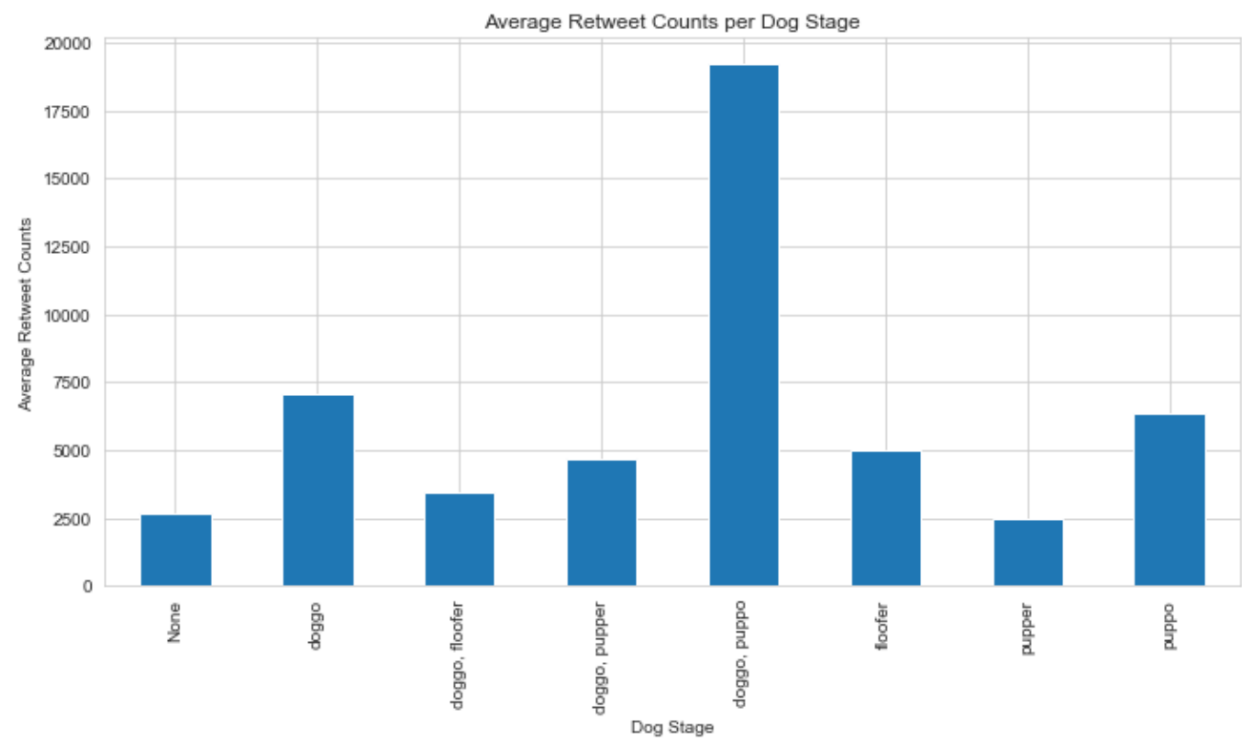
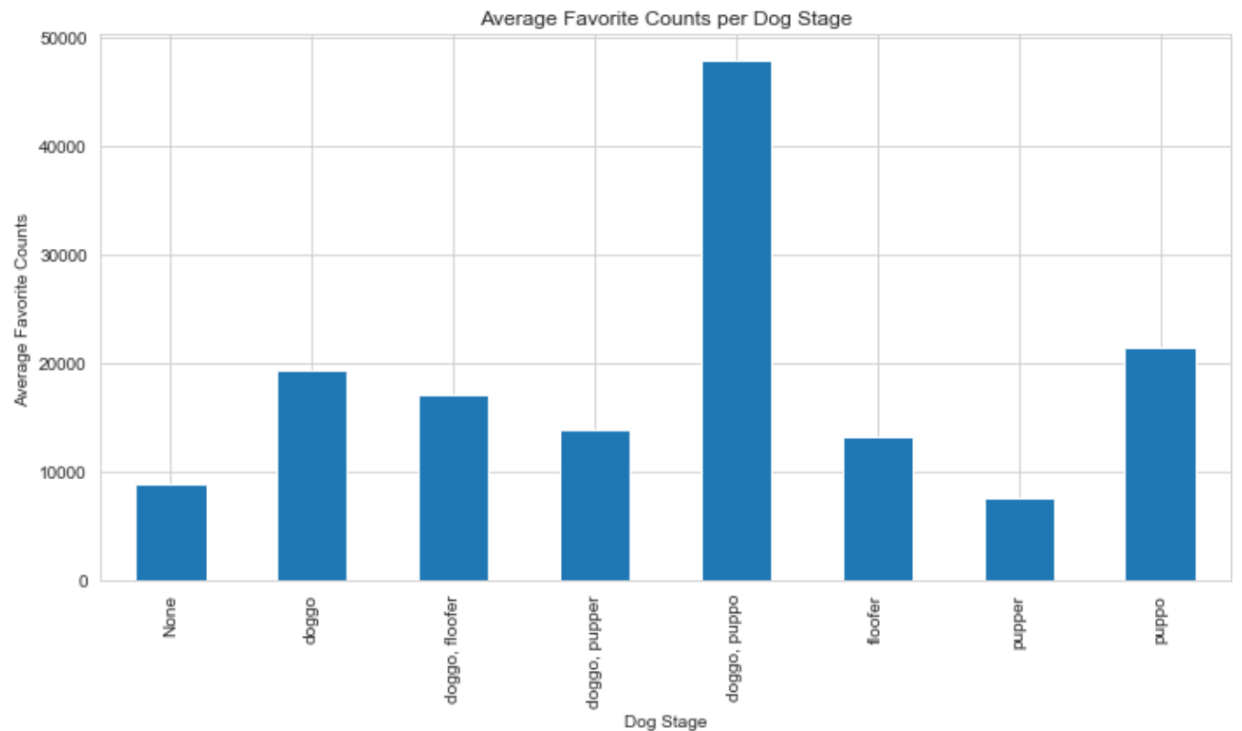


The most favorite picture is a a super supportive puppo participating in the Toronto #WomensMarch, and received 132810 favorites.



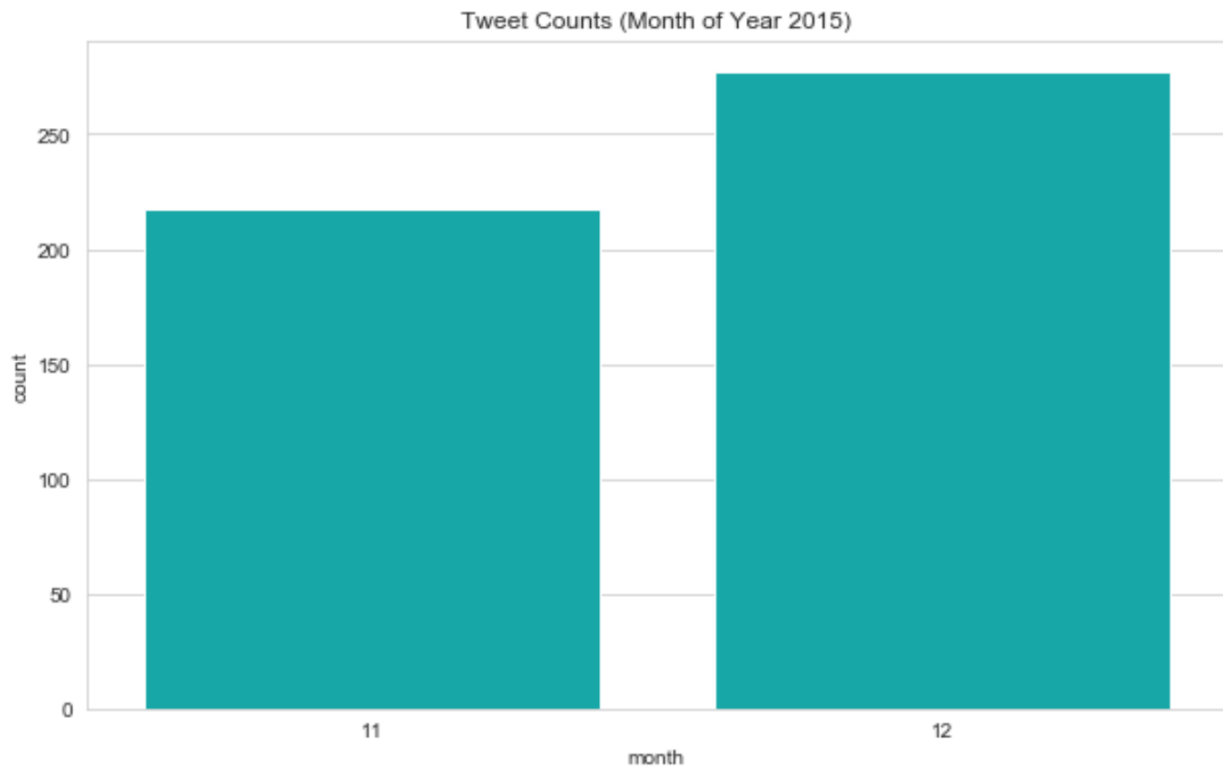
The most retweeted picture is a doggo that realizes it can stand in a pool (swimming actually), and it received 79515 retweets.

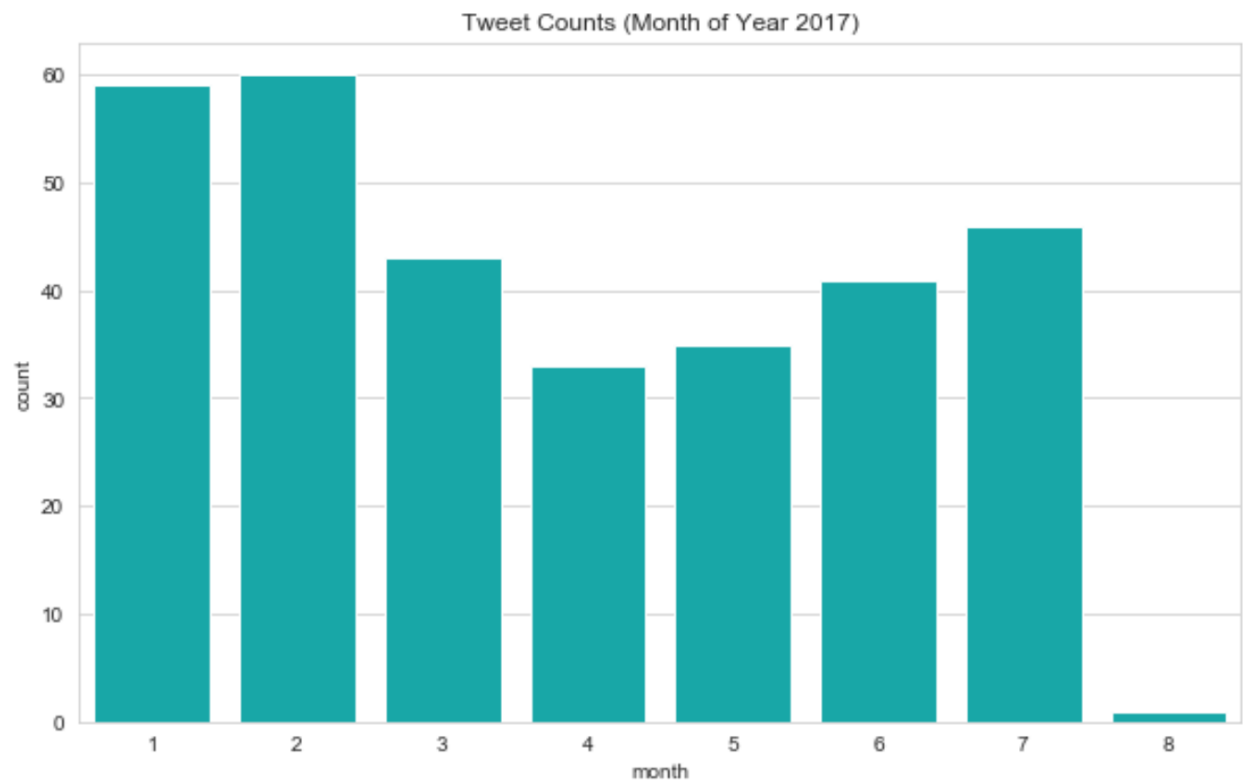
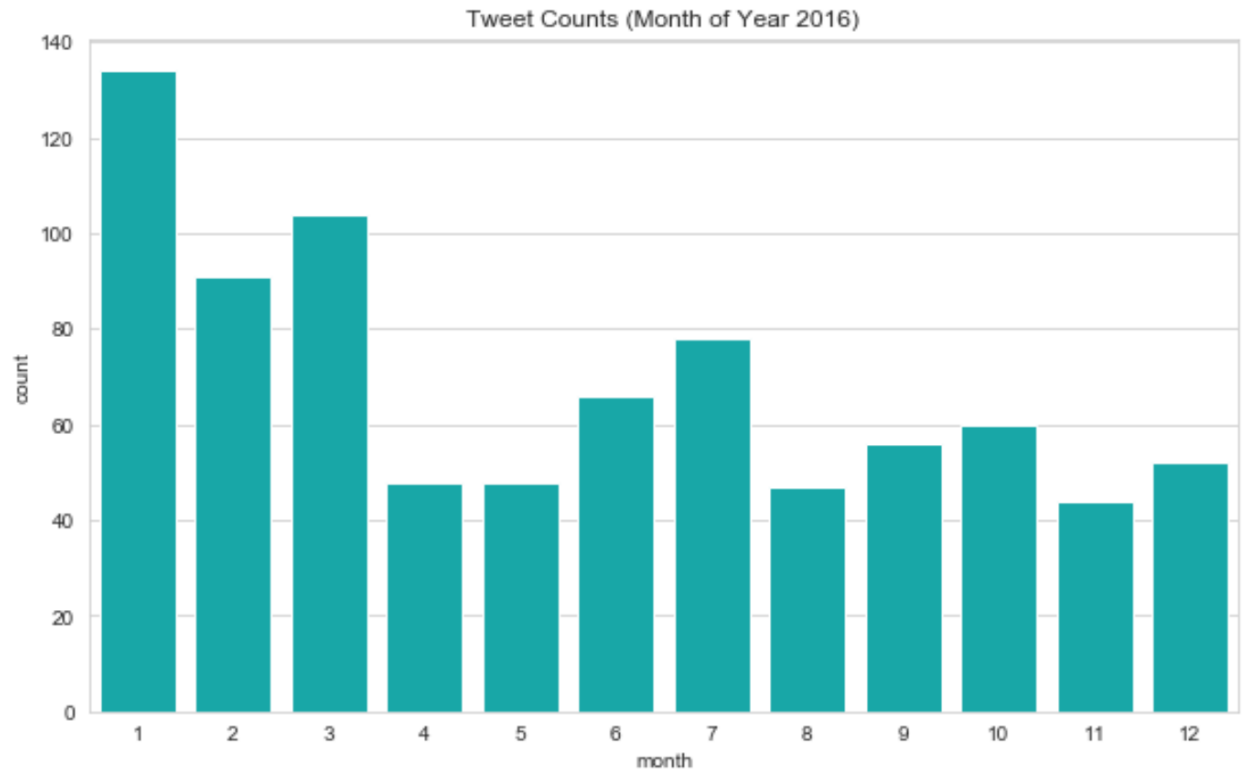
What is the average favorite/retweet count on different dog_stage?



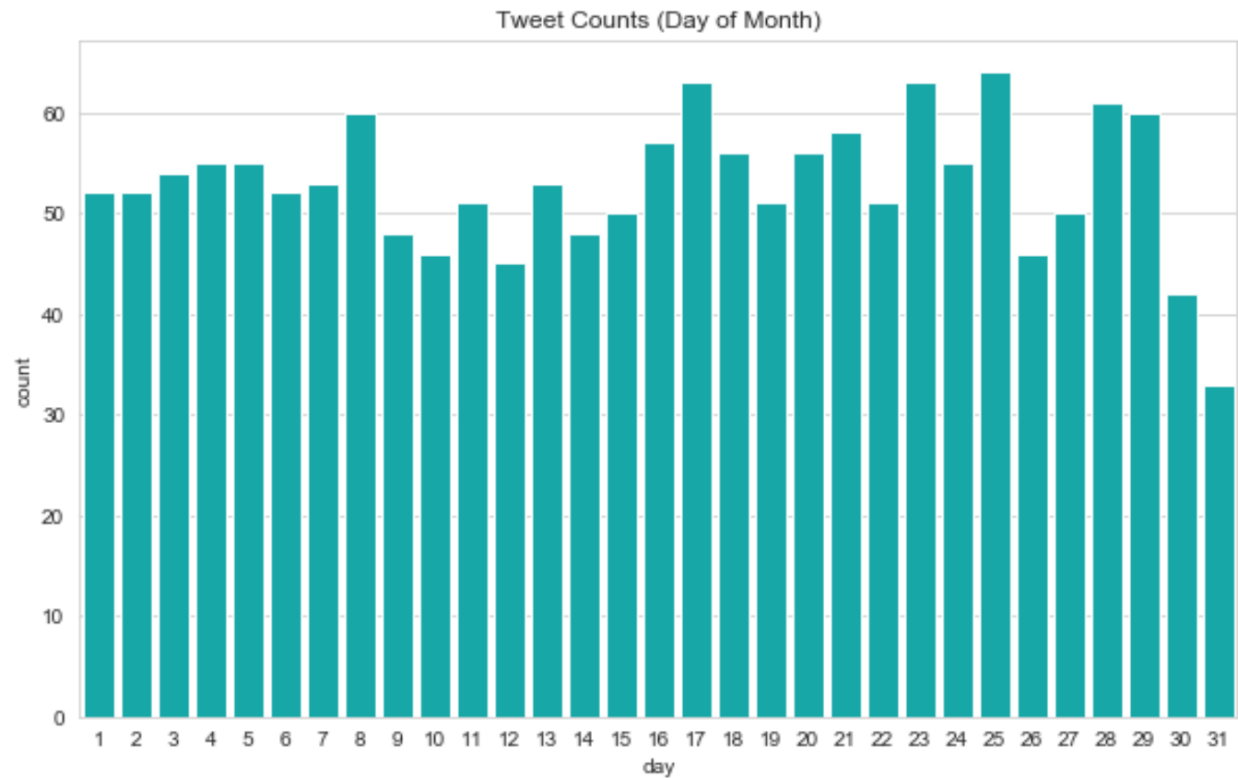
If we consider there are in total only 9 tweets where the dogs have multiple stages, and we exclude them in comparison because these posts might not contain general results, overall, puppo on average gets the most favorite count, doggo on average gets the most retweet count (slightly leading puppo).

What is the WeRateDogs account activity by year, month, day of the week, hour of the day?

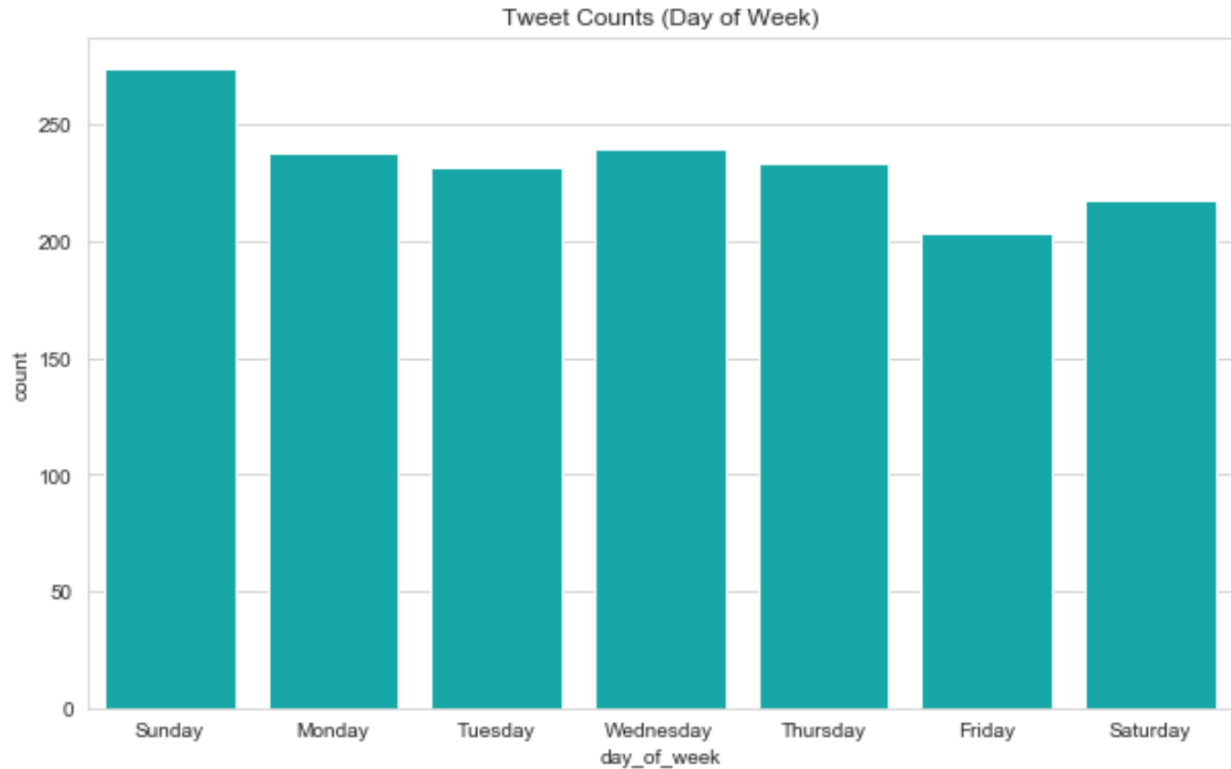




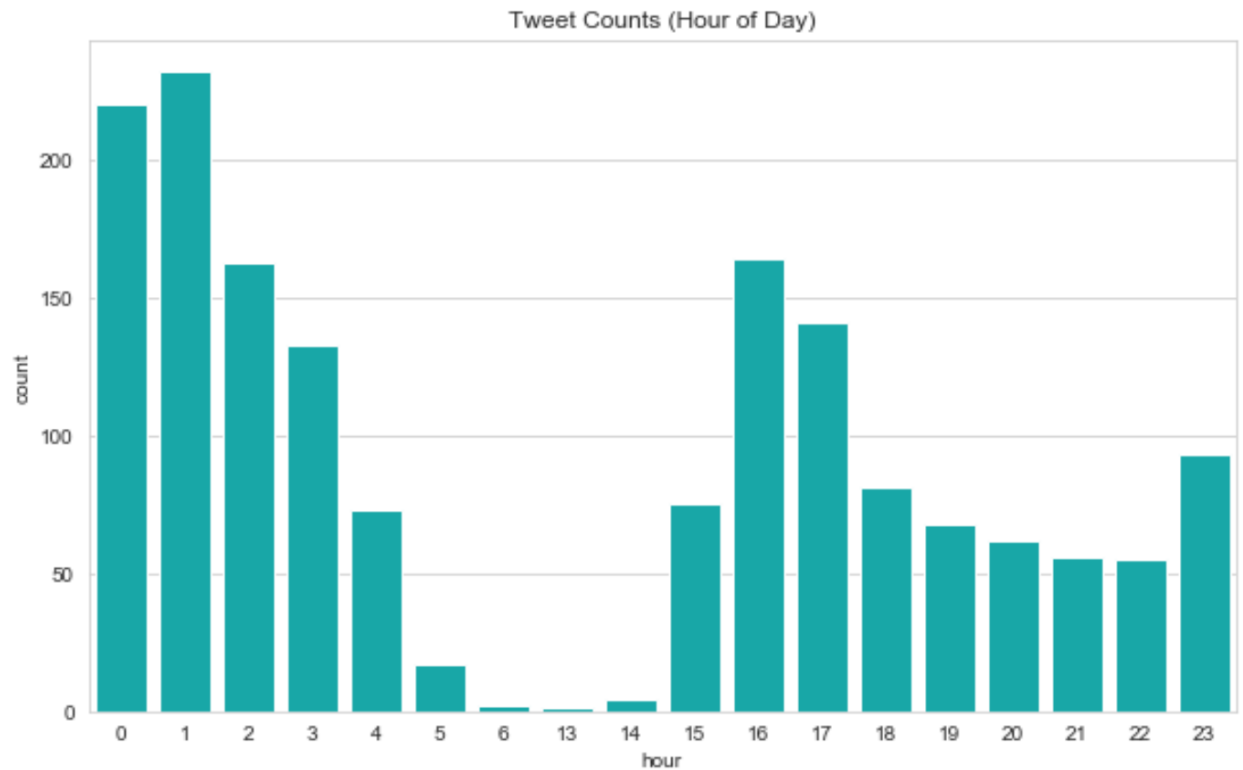
WeRateDogs on average posts more than 200 tweets during November and December in 2015, this number drops for the next 3 months (Jan, Feb, Mar of 2016) with ~100 tweets/month, then drops again and stabilizes from April 2016 until July 2017 (averages 40 to 50 tweets a month).



The total tweet counts in each day of the month are pretty stable (fewer tweets on 30 or 31 is because some months don't have the 30th or 31st day).

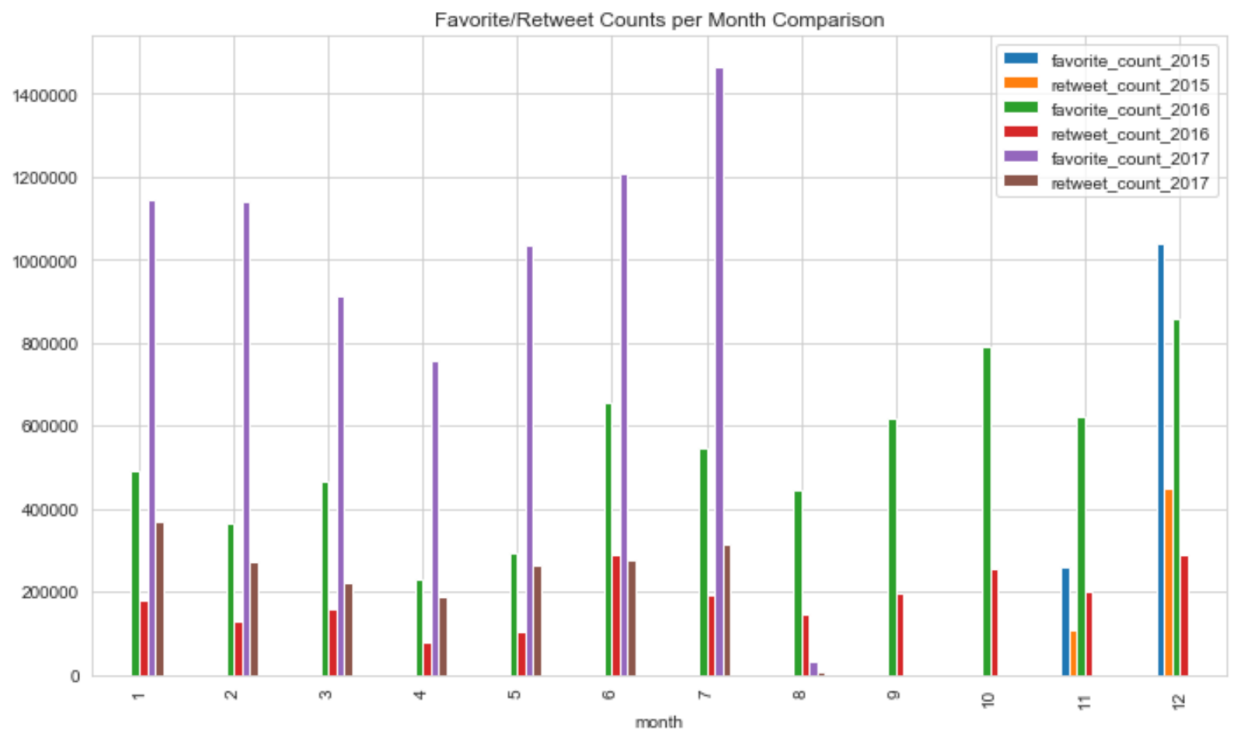


From the day of week total posts, WeRateDogs seems to post more tweets on Sunday (with a total post number of ~270) than any other day of the week.



More tweets actually come from hours of 0-3 and 16-17, There is no single post for the hours of 7-12 (maybe it's the sleep time for the people who run the account).

What is the distribution of favorite counts by month in each year?



Overall, the account is gaining popularity in receiving favorites and retweets over the years (we can see increases in height from the green bar to the purple bar showing trend in favorites, red bar to the brown bar showing trend in retweets). The decrease from blue bar to green bar is actually due to different numbers of tweets in the respective months (Dec), we have many more tweets in Dec 2015 than Dec 2016).