

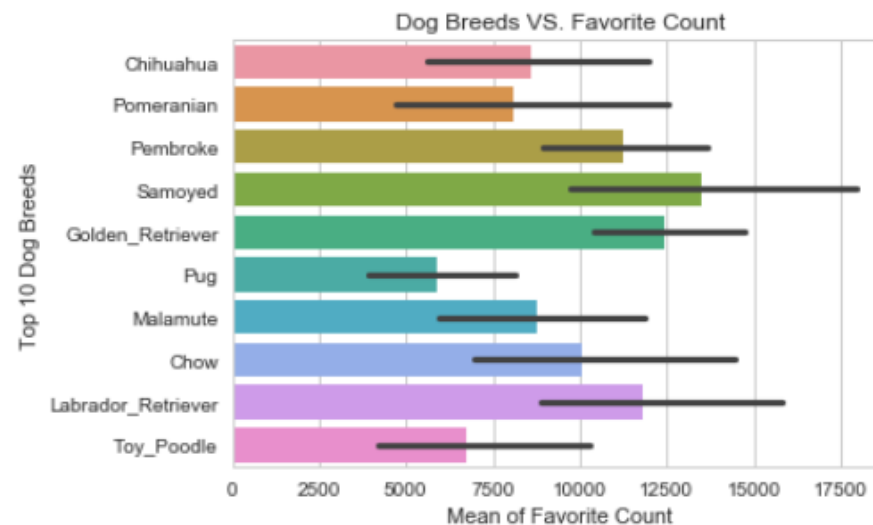
Analysis of *WeRateDogs*

Connie Hsiao

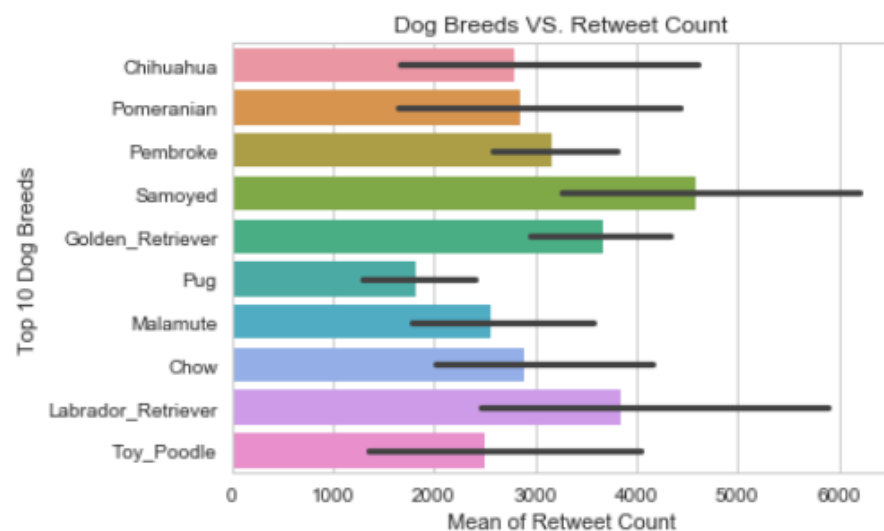
With the clean and complete dataset, I can explore and analyze the data in different ways. Thus, the analysis will be made up of three parts, people's preferences for dog breeds, Users' retweet habits and the relationship of ratings and the favorite counts.

Top 10 Dog Breeds' Popularity

I use the variable 'prediction_top1' to determine the dog breeds of each tweet and narrow down the analysis to the top 10 breeds with most tweets. With the data of retweet counts and favorite counts, I can find out which breed is the most popular one among the 10 breeds.



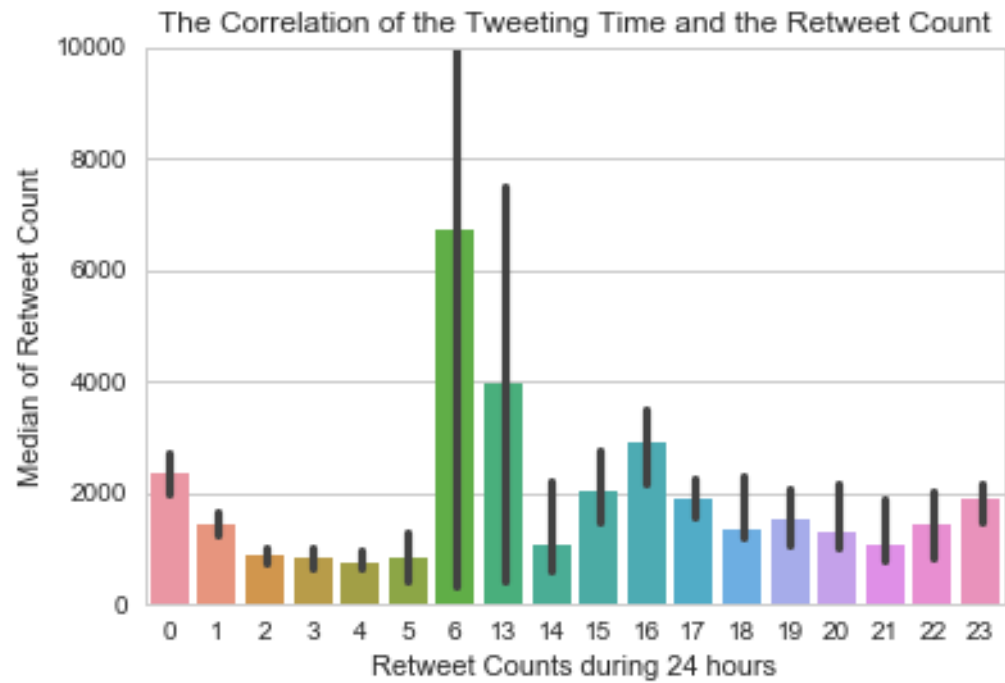
From the bar chart above, we can see that each Samoyed tweet can receive the most average favorites followed by Golden Retriever. How about retweet counts? Let's see another bar chart below.



Samoyed also received most average retweet count , and Golden Retriever is still the runner-up. We can say that Samoyed is the most popular dog!

When People Like to Retweet Most in a Day

Using the data of timestamp and the retweet_count, we can get the median retweet numbers of the tweets posted at different hours. By doing so, we can figure out when people like to retweet most in a day.

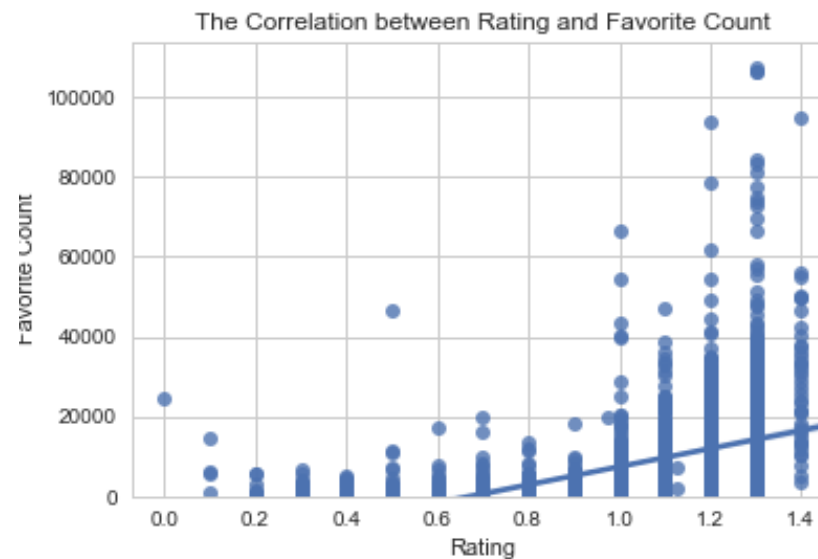


Hour (UTC)	0	1	2	3	4	5	6	13	14	15	16	17	18	19	20	21	22	23
Median Retweet Count	2323	1418	866	830	727	831	6709	3947	1270	2020	1434	1889	1352	1509	1270	1075	1434	1894

From the statistics and plot above, we can learn that @dog_rates loves to tweet at 1 AM UTC (5pm PST); there are 289 tweets posted at 1 AM. However, the tweets which posted at 6 AM UTC (10pm PST) have the largest median, 6,709 retweet. It seems that Weratedogs' s followers like to retweet before sleeping.

The Rating Matters or Not

I would like to know if there's any correlation between the ratings and the favorite counts, so I will use the data in 'rating_numerator','rating_denominator'and 'favorite_count' columns to analyze the relationship of the ratings and the followers' preferences.



The Pearson's r of rating and favorite count is 0.413, which means there is a moderately positive correlation between the variables. The ratings somewhat reflect people's preferences for dogs.