

Act Report

Clean DataFrame contains 1532 observations. The key information includes rating, dog stage, dog_breed, retweet counts and favorite counts.

	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count
count	1.532000e+03	1532.000000	1532.000000	1520.000000	1520.000000
mean	7.429474e+17	11.556136	10.494778	2452.927632	7961.617105
std	6.863141e+16	7.679495	6.646577	4339.772246	11891.012677
min	6.660209e+17	2.000000	7.000000	11.000000	0.000000
25%	6.787998e+17	10.000000	10.000000	556.000000	1648.500000
50%	7.200513e+17	11.000000	10.000000	1238.000000	3661.000000
75%	7.987548e+17	12.000000	10.000000	2870.500000	9904.000000
max	8.921774e+17	165.000000	150.000000	73286.000000	148901.000000

Analysis

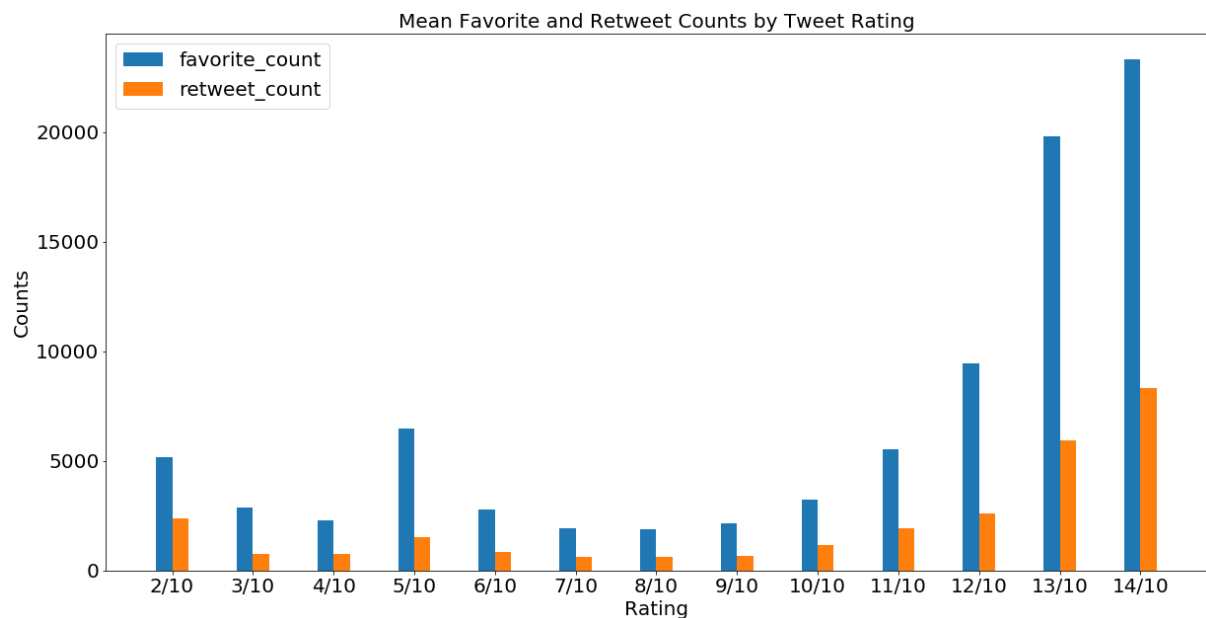


Fig1: These are the most frequent ratings on the WeRateDogs. We can see that with the rising numerator, also the number of tweets and retweets are rising. However there are some extremes with low ratings (<6).

ID	Rating	Most Frequent Breed
0	2/10	[Maltese_dog]
1	3/10	[Chihuahua]
2	4/10	[soft-coated_wheaten_terrier]
3	5/10	[flat-coated_retriever]
4	6/10	[Airedale]
5	7/10	[Chihuahua]
6	8/10	[Italian_greyhound]
7	9/10	[pug]
8	10/10	[Labrador_retriever]
9	11/10	[golden_retriever]
10	12/10	[golden_retriever]
11	13/10	[golden_retriever]
12	14/10	[Pembroke]

Table 1: Most common breeds for the most common ratings (2-14/10)

golden_retriever	150
Labrador_retriever	100
Pembroke	89
Chihuahua	83
pug	57
chow	44
Samoyed	43
toy_poodle	39
Pomeranian	38
cocker_spaniel	30

Table 2: Most common breeds of all of the whole dataset.

Insights

- The most common ratings have numerator 2-14 and denominator 10.
- I have identified mean favorite count and retweet count for the most common ratings. I found that both favorite counts and retweet counts are rising with ratings. For ratings 11/10 and higher mean favorite count is higher than 5000, and for 14/10 it is higher than 20000. The mean retweet count is higher than 5000 for ratings 13/10 and 14/10.
- I identified also that very low ratings, particularly 2/10 and 5/10 have favorite count higher than 5000 and retweet_count around 2000.
- I have identified that the tweet with the highest favorite count is as well as with the highest retweet count is the Labrador retriever (tweet_id=744234799360020481) with the rating 13/10. Tweet contains video how he is swimming and he is a doggo.
- I have identified most common breeds for the most common ratings (2-14/10) and save these data in the dataframe df_breeds.
- The most favourite breed of the whole dataset is the golden retriever.