

Act Report

Now that we have performed an assessment and cleaning of our WeRateDogs tweet data we can explore and analyze the data to find interesting insights.

Figure 1. Shows the descriptive statistics from the dataset summarizing basic details such as the means, percentile, min, max, etc. from our numeric values.

For example, in the retweet_count column we can see that the tweet with the most retweets had 79515 retweets.

	tweet_id	rating_numerator	rating_denominator	rating	retweet_count	favorite_count	user_count	img_num	p1_conf	p2_conf
count	2.091000e+03	2091.000000	2091.000000	2091.000000	2091.000000	2091.000000	2.091000e+03	1965.000000	1965.000000	1.965000e+03
mean	7.365318e+17	11.101387	10.451459	1.059854	2841.095170	8983.248207	3.200945e+06	1.202545	0.594666	1.347608e-01
std	6.713600e+16	7.910217	6.654129	0.216125	4714.588405	12213.156088	4.409045e+01	0.559762	0.272081	1.011342e-01
min	6.660209e+17	0.000000	2.000000	0.000000	16.000000	81.000000	3.200799e+06	1.000000	0.044333	1.011300e-08
25%	6.768157e+17	10.000000	10.000000	1.000000	643.500000	2045.500000	3.200901e+06	1.000000	0.362925	5.351500e-02
50%	7.095662e+17	11.000000	10.000000	1.100000	1403.000000	4188.000000	3.200947e+06	1.000000	0.587830	1.175080e-01
75%	7.877641e+17	12.000000	10.000000	1.200000	3285.000000	11434.000000	3.201001e+06	1.000000	0.847292	1.959500e-01
max	8.924206e+17	204.000000	170.000000	1.400000	79515.000000	132810.000000	3.201018e+06	4.000000	1.000000	4.880140e-01

Figure 1. Dataset's Descriptive Statistic

Figure 2. Shows the tweet of the dog with the most retweets and yes he deserves his popularity and rating! Unfortunately his name was not specified but we know he is a doggo that learns something new every day. On the other hand, our dataset also shows the results of the breed predictions made by the algorithm. The first prediction of the algorithm was indeed "Labrador retriever" which perfectly matches the real breed of this dog and had 83% confidence. The second prediction of the algorithm was "ice bear" with 4.5% confidence probably because it is also very common to see a polar bear swimming in the water! The third prediction was a "whippet" with 1.8% confidence.

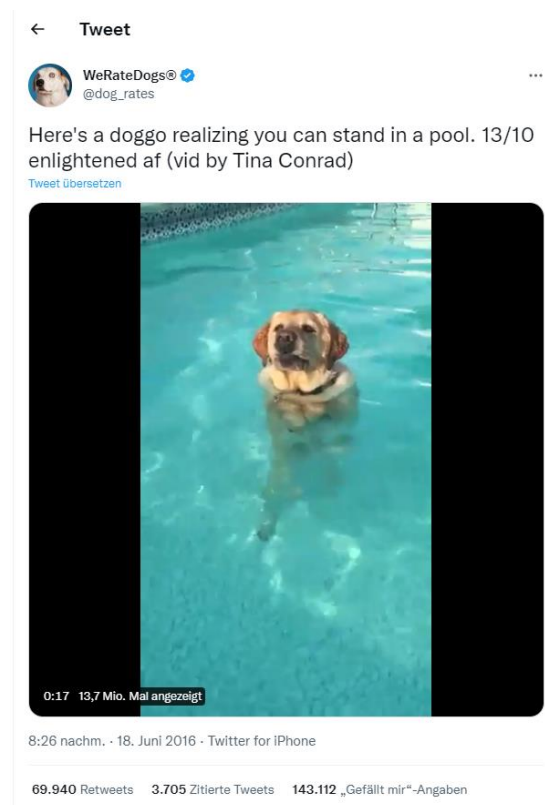


Figure 2. Most Retweetet Tweet

Now let's check the ratings on the high level. Figure 3. shows the histogram of the ratings which shows a left skewed distribution indicating that most of the dogs were over rated as the mean is 1.06. This was expected as most dogs were over rated. This was to be expected since we knew from the beginning that WeRateDogs runs a unique rating system and dogs can get a rating higher than 1.

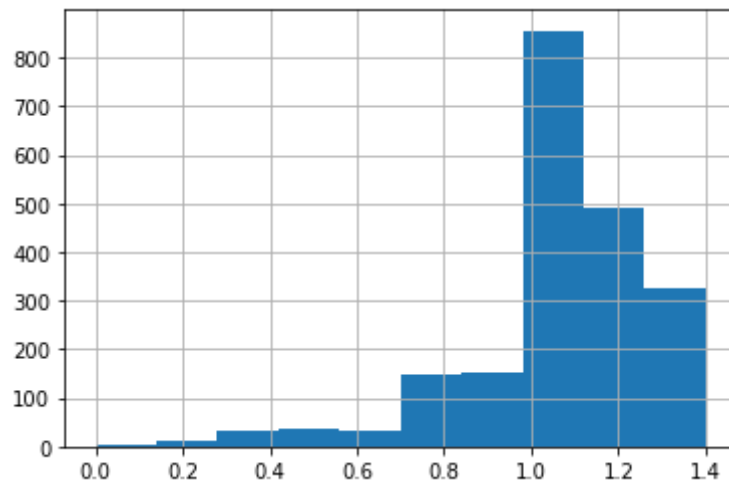


Figure 3. Rating Histogram

Finally, let's see what was the top most frequent dog breeds and the most repeated dog names. In Figure 4 we can see that the most frequent dog breeds were Golden retriever, Labrador retriever, pembroke, chihuahua and pug. These results indicate that the algorithm is doing well in identifying dogs. In Figure 5. We can see that the most repeated names were Charlie, Lucy, Cooper, Oliver and Tucker, all of them nice names for good dogs.

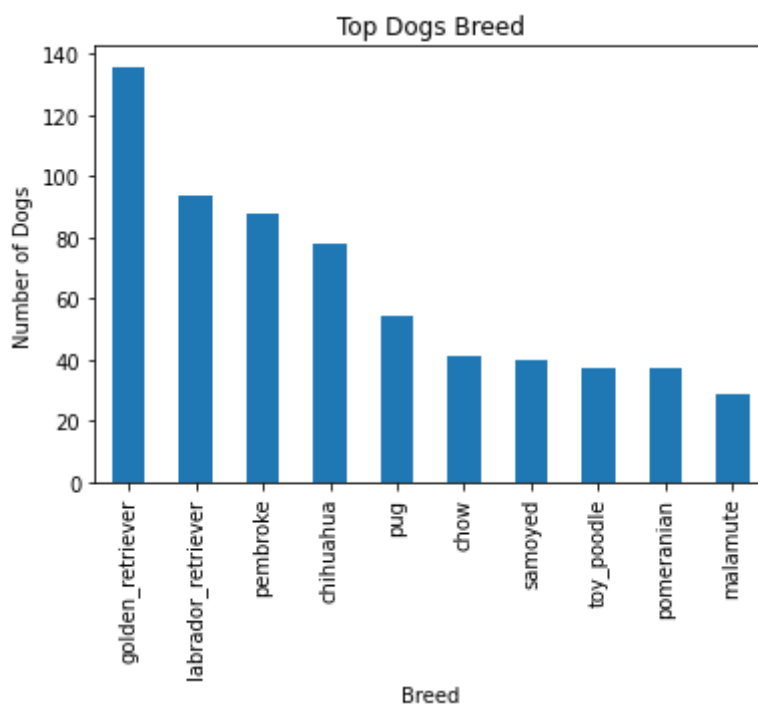


Figure 4. Top Dogs Breed

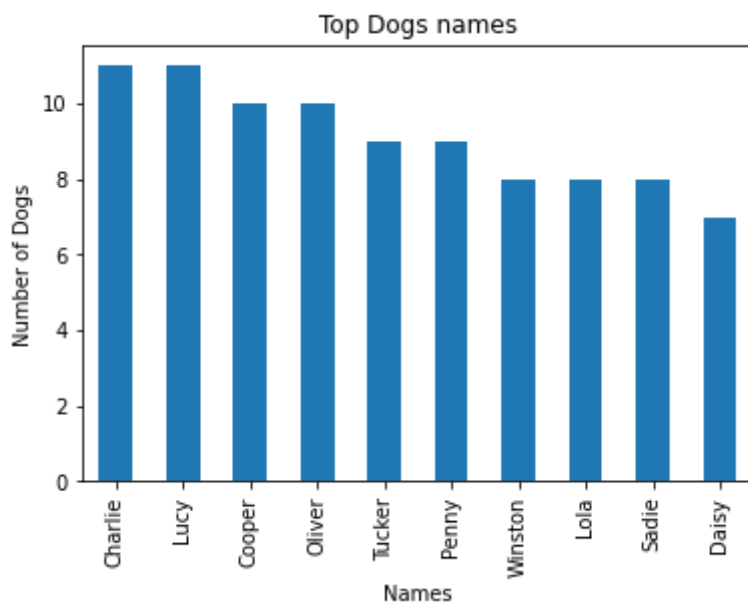


Figure 5. Top Dogs Names