

WeRateDogs Twitter Archive - Act Report

WeRateDogs Data Description:

The WeRateDogs Enhanced Twitter archive contains data that was extracted from the text of the tweet. Key fields included dog ratings, dog name and dog stage if present.

The retweet count and favorite count for each tweet were not included in the enhanced archive, and so this additional data had to be downloaded from the twitter account using the tweet ID from the archive. However, I faced issues in creation of a Twitter developer account and hence used the "tweet-json.txt" file provided as part of the course material.

Data Wrangling:

Before I could begin the analysis, the data had to be wrangled so that it was in a form that could be conveniently analyzed. I assessed the data along the dimensions of completeness, validity, accuracy and consistency. This assessment was done both visually and programmatically.

After cleaning many of the issues found during the assessment, there were 1953 tweets with good quality data which were then analyzed.

Data Analysis and Insights:

Tweets are favorited more often than they are retweeted

All the tweets analyzed as part of the exercise had higher favorite counts than retweet counts. We observe that the average number of favorites per tweet is more than 3X the average no. of retweets per tweet.

- Mean no. of favorites per tweet: 8952
- Mean no. of retweets per tweet: 2784

This indicates that users following the WeRateDogs Twitter handle are more likely to favorite a tweet than retweet it. This could be a means of saving these tweets for quicker access in the future.

No impact of timestamp on no of retweets or no. of favorites received by tweet

No. of retweets or no. of favorites varies significantly with the day of week on which the tweet is posted. However, we notice a rise in both counts on Wednesday (day_of_week = 2)

	rating_numerator	retweet_count	favorite_count
day_of_week			
0	10.367284	2640.873457	8508.052469
1	10.600000	2740.020690	9168.551724
2	10.453925	3121.399317	10067.358362
3	10.670251	2524.913978	8339.487455
4	10.690647	2811.345324	8767.676259
5	10.552419	2995.145161	9064.612903
6	10.454545	2673.055336	8744.383399

We observe that whether a tweet is posted on a weekday or weekend does not impact the number of retweets or number of favorites received by the tweet.

- Mean no. of retweets per tweet when posted on weekends: 2833
- Mean no. of retweets per tweet when posted on weekdays: 2767
- Mean no. of favorites per tweet when posted on weekends: 8903
- Mean no. of favorites per tweet when posted on weekdays: 8968

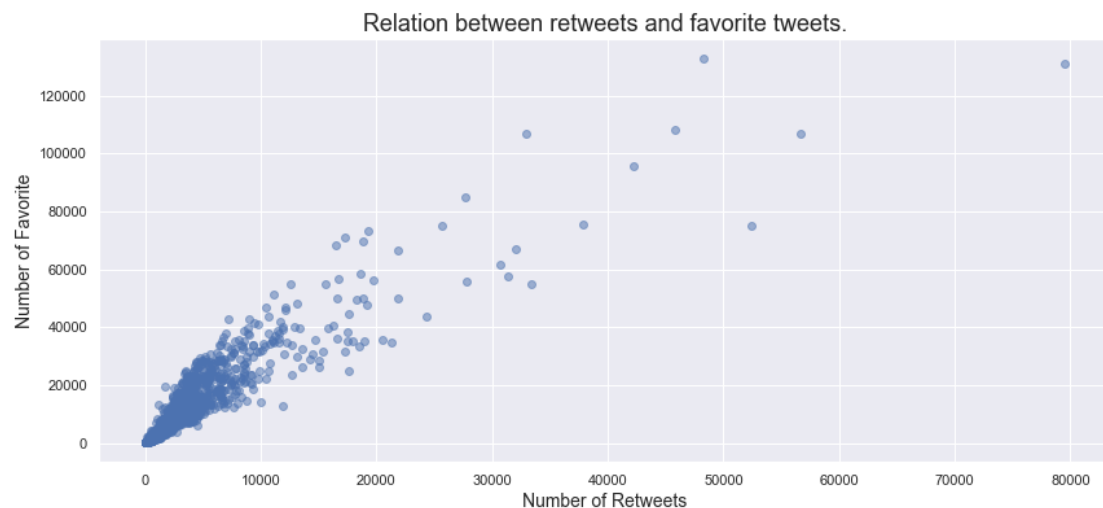
A lower average rating is received when a large no. of dogs have been tweeted in the month

	Avg Rating	Tweets Number
Jan	10.711207	232
Feb	10.889535	172
Mar	10.939394	165
Apr	11.170213	94
May	11.683673	98
Jun	11.073171	123
Jul	11.470588	136
Aug	11.000000	61
Sep	11.193548	62
Oct	11.703125	64
Nov	9.378613	346
Dec	9.873786	412

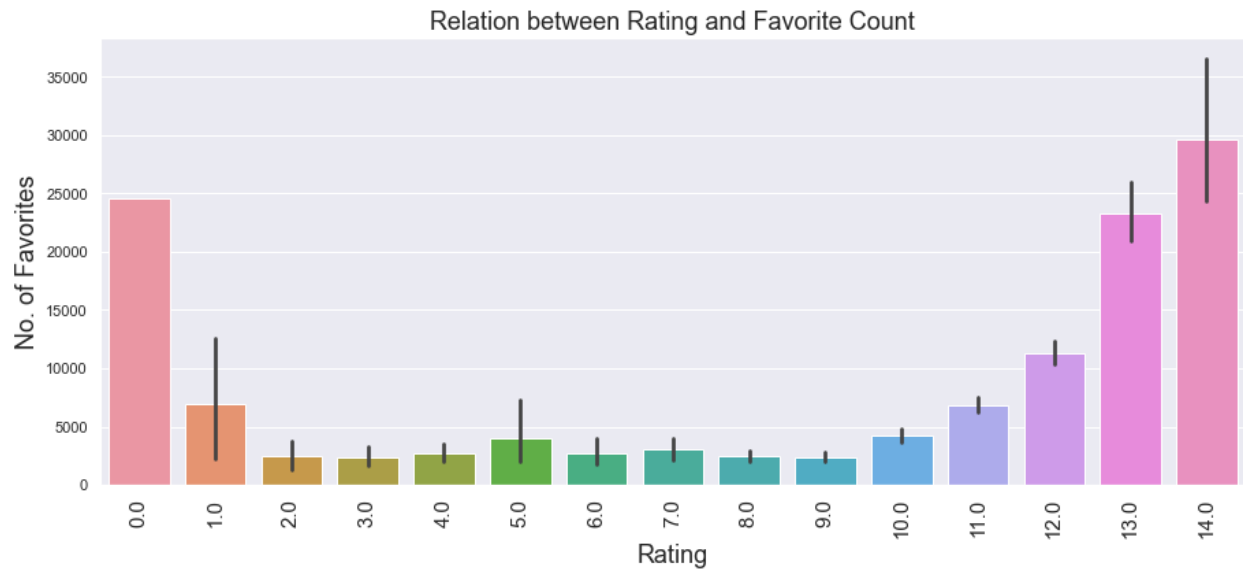
In the table indicated, it seems like the average rating seems to be move higher up when the no. of tweets in a month is low

Strong positive correlation between favorite count and retweet count

A plot to indicate how the favorite count and retweet count variables interact was analyzed as indicated below. Each point on the chart represents a tweet with the x-axis indicating the retweet count, while the y-axis indicates the retweet count. The chart suggests a strong positive correlation between a tweet's retweet counts and favorite counts. This makes sense since it indicates that if users like a tweet enough to be favorited, they are also likely to retweet it.



Impact of dog rating on favorite count and retweet count



The above chart indicates that tweets that receive a higher rating tend to be favorited more number of times by users

Golden retriever is the most popular breed amongst WeRateDogs followers

Analyzing the tweets at a dog breed level, we found that Golden Retrievers are most often tweeted about, with more than 150 tweets. Labrador Retrievers, Pembroke and Chihuahuas are closely spaced together in 2nd, 3rd, and 4th place as indicated in the chart below:

