Tweets
9,687

Following
12

Followers
7.7M

Likes
141K

Moments
4

Follow

WeRateDogs™ ✔

@dog_rates

Your Only Source For Professional Dog Ratings   IG, FB, Snapchat ⇨ WeRateDogs partnerships@weratedogs.com

⌖ DM YOUR DOGS

🔗 weratedogs.com

🗓 Joined November 2015

Tweet to   Message

Introduction

WeRateDogs is a <u>Twitter</u> account that rates people's <u>dogs</u> with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "<u>they're good dogs Brent</u>." WeRateDogs was started in 2015 by college student Matt Nelson and has received international media coverage. Now, it has 7.7 million followers.

Visualization and Analysis

The clean and tidy DataFrame on which the analyses and visualizations are based has been assessed and cleaned. The resulting Dataframe contains 1666 observations. The key information included in my analysis is favorite counts, retweet counts , source and rating out of 10 .

((First Insight))

Twitterers Do "Favorite" Three Times As Much As they do "Retweet"

	rating_out_of_10	favorite_count	retweet_count
count	1666.000000	1666.000000	1666.000000
mean	10.838535	9288.184274	2834.429772
std	1.962955	12641.538890	4831.960790
min	0.000000	81.000000	16.000000
25%	10.000000	2176.000000	650.500000
50%	11.000000	4464.000000	1440.500000
75%	12.000000	11734.750000	3261.000000
max	34.000000	132810.000000	79515.000000

```python
# total favourites
sum(master_dataset.favorite_count)
```
15474115

```python
# total retweets
sum(master_dataset.retweet_count)
```
4722160

As shown above , the favorite_count mean is around 9288 favorites while the median is 4464 favorites. On the other hand, retweet_count has a mean of almost 2834 retweets and a median of 1440.5 . These two variables, favorite_count and retweet_count, are having significantly greater means than medians. Additionally , their standard deviation is also greater than the mean.In my opinion, this might be an indication that the favorite_count and retweet count are heavily skewed.

Distribution of Retweet Counts

Distribution of Favorite Counts
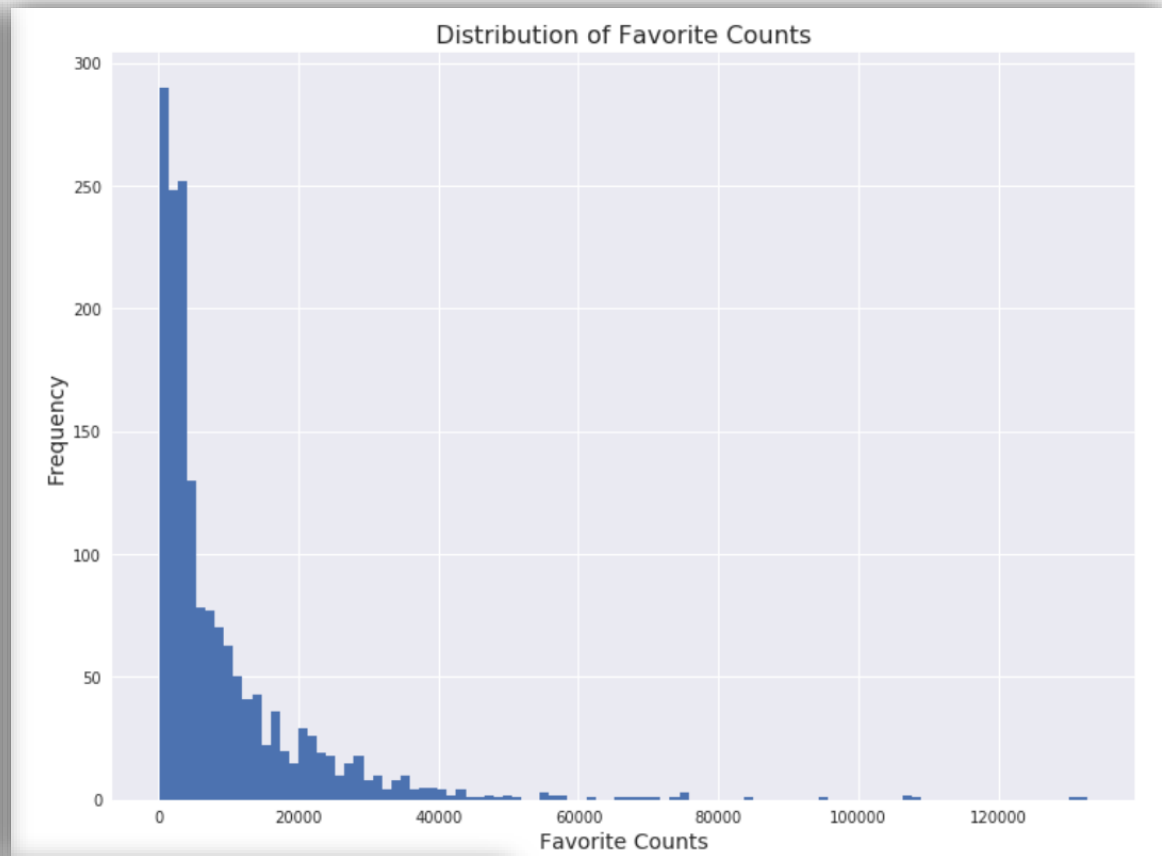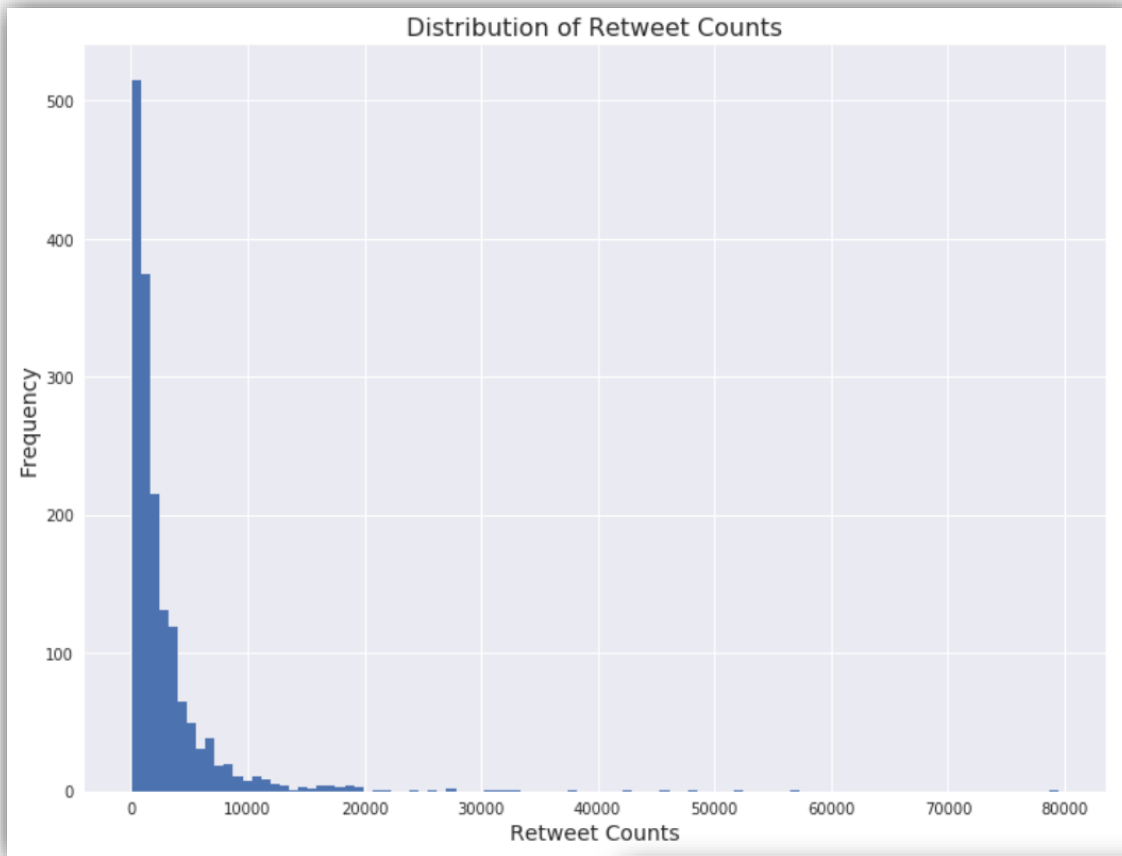
```
# checking the possibility of having the Retweet count greater than
master_dataset[master_dataset['retweet_count'] > master_dataset['favorite_count']]
```

tweet_id	timestamp	source	text	expanded_urls	rating_out_of_10	dog_name	dog_stage	favorite_c

0 rows × 21 columns

Plotting the distributions of `favorite_count` and `retweet_count` showed that both distributions are right skewed. At first glance , they look pretty similar but checking the breaks on the X-axis of both plots, clearly indicates that favorite_count is always higher than retweet_count .However , it is important to know that although there might be an exception to every rule, In our case THERE IS NO EXCEPTIONS .
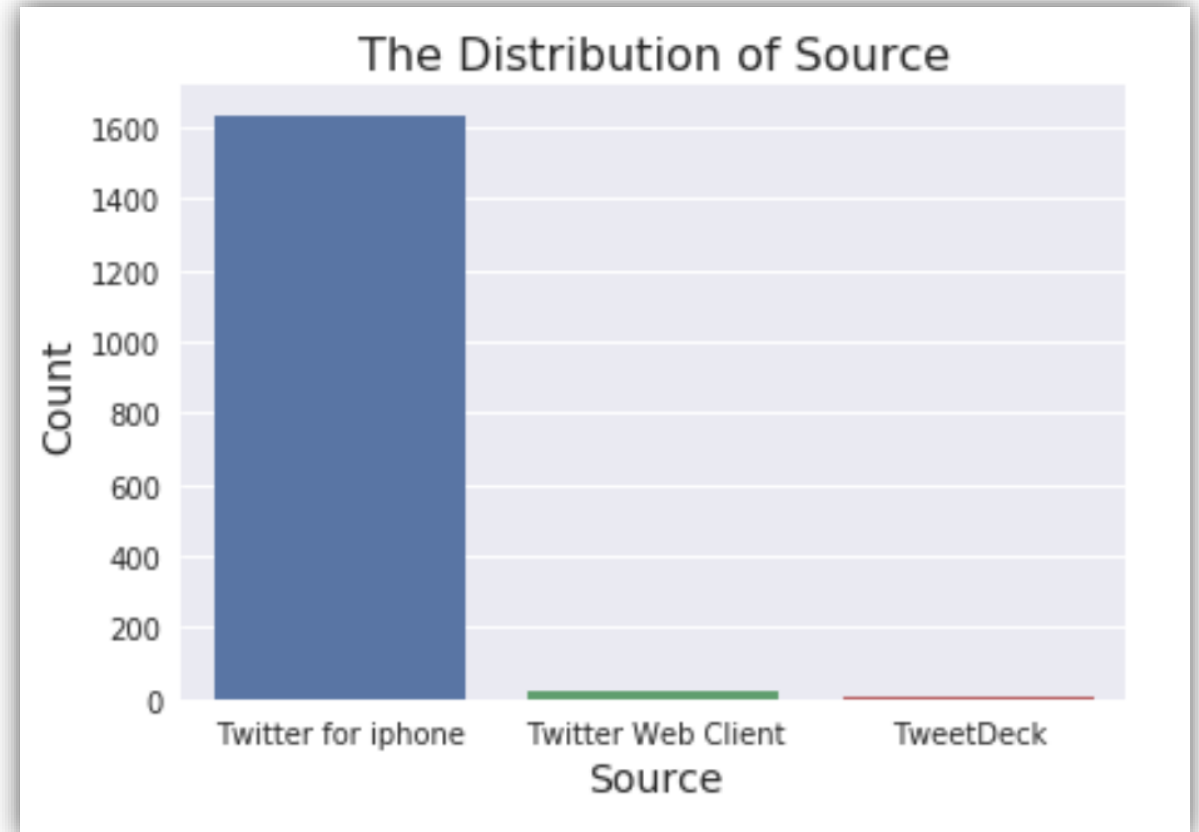
🐕 ((Second Insight))

The Team Managing WeRateDogs Account is Accessing Their Account, Mostly, Through Their iPhone Twitter App


```
Twitter for iphone    1635
Twitter Web Client      22
TweetDeck                9
Name: source, dtype: int64
```

The bar plot presents the distribution of source. We can see clearly the dominate source of `WeRateDogs` tweets have been tweeted through iPhone twitter app. We can see that the source of more than 98% of tweets is "Twitter for iphone". That means that the iphone twitter app is the main channel for people managing the Twitter user @dog_rates using to tweet, while the TweetDeck is pretty rare to be used (less than 1%).

The Distribution of Source

The Second insight leads me to imagine WeRateDogs team relaxing on their favorite couch while posting their great , funny and passionate tweets.

((Third Insight))

A Dog Rating of (13/10) was given to the Top 5 favorite tweets, and the Top 5 retweeted tweets. (Interesting ?? huh!!)

```
# Top 5 Favorite Tweets
master_dataset.sort_values('favorite_count', ascending=False)[['tweet_id','dog_n
```

	tweet_id	dog_name	dog_stage	favorite_count	retweet_count	rating_out_of_10
306	822872901745569793	NaN	puppo	132810	48265	13
769	744234799360020481	NaN	doggo	131075	79515	13
58	879415818425184262	Duddles		107956	45849	13
397	807106840509214720	Stephan		107015	56625	13
108	866450705531457537	Jamesy	pupper	106827	32883	13

```
# Top 5 retweeted Tweets
master_dataset.sort_values('retweet_count', ascending=False)[['tweet_id','dog_na
```

	tweet_id	dog_name	dog_stage	favorite_count	retweet_count	rating_out_of_10
769	744234799360020481	NaN	doggo	131075	79515	13
397	807106840509214720	Stephan		107015	56625	13
804	7392381577791694849	NaN	doggo	75163	52360	13
306	822872901745569793	NaN	puppo	132810	48265	13
58	879415818425184262	Duddles		107956	45849	13

```
master_dataset.rating_out_of_10.value_counts(sort=True, ascending=False).head(5)
12    424
10    360
11    357
13    221
9     132
```

What is more interesting is that: after checking the list of top 5 most frequent ratings, the rating (13/10) took the fourth place on that list , with only 221 dogs. This might indicate a weak relationship between the `rating_out_of_10` and both `favorite_count` and `retweet_count` .

# Let's check their Correlation by calculating the Correlation Coeffient

```
np.corrcoef(master_dataset.rating_out_of_10,master_dataset.favorite_count)
```
0.30908929420098485

```
np.corrcoef(master_dataset.rating_out_of_10,master_dataset.retweet_count)[
```
0.23325703626792099

The correlation coeffient values, shown on the left, indicate WEAK relationships between `rating_out_of_10` and both `favorite_count` and `retweet_count`. However, plotting these relationships will surely confirm the existence of this WEAK relationship.
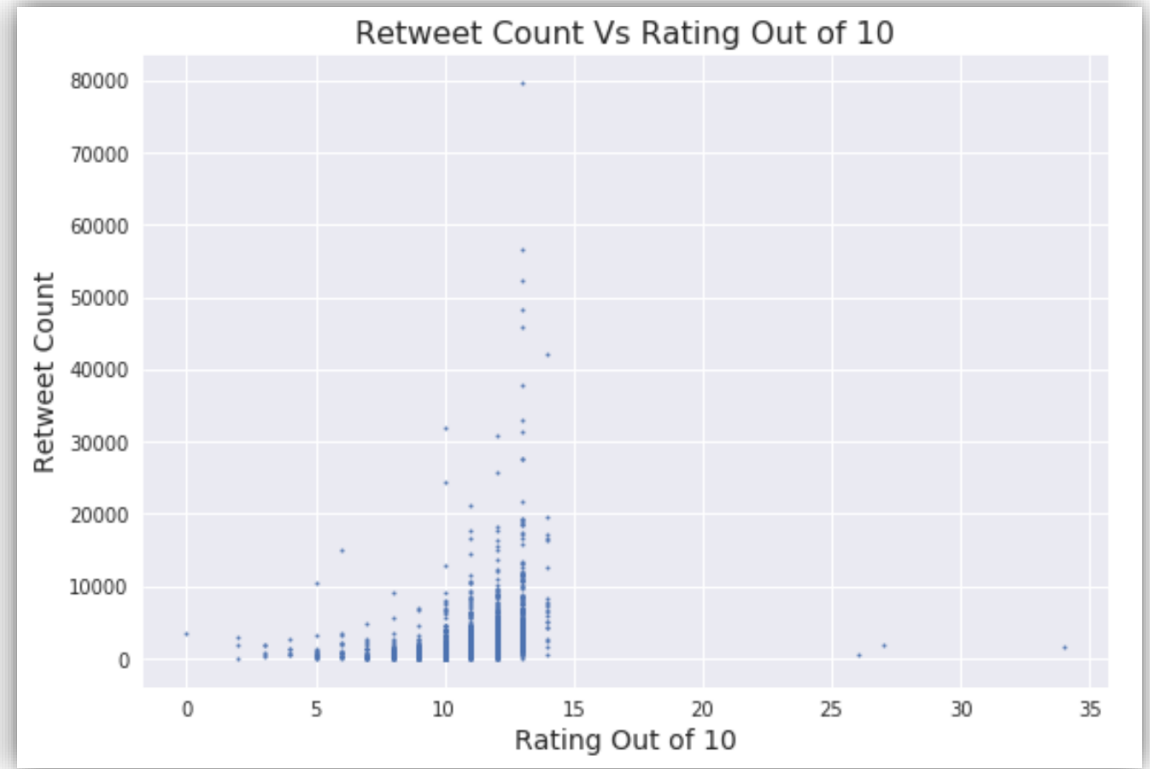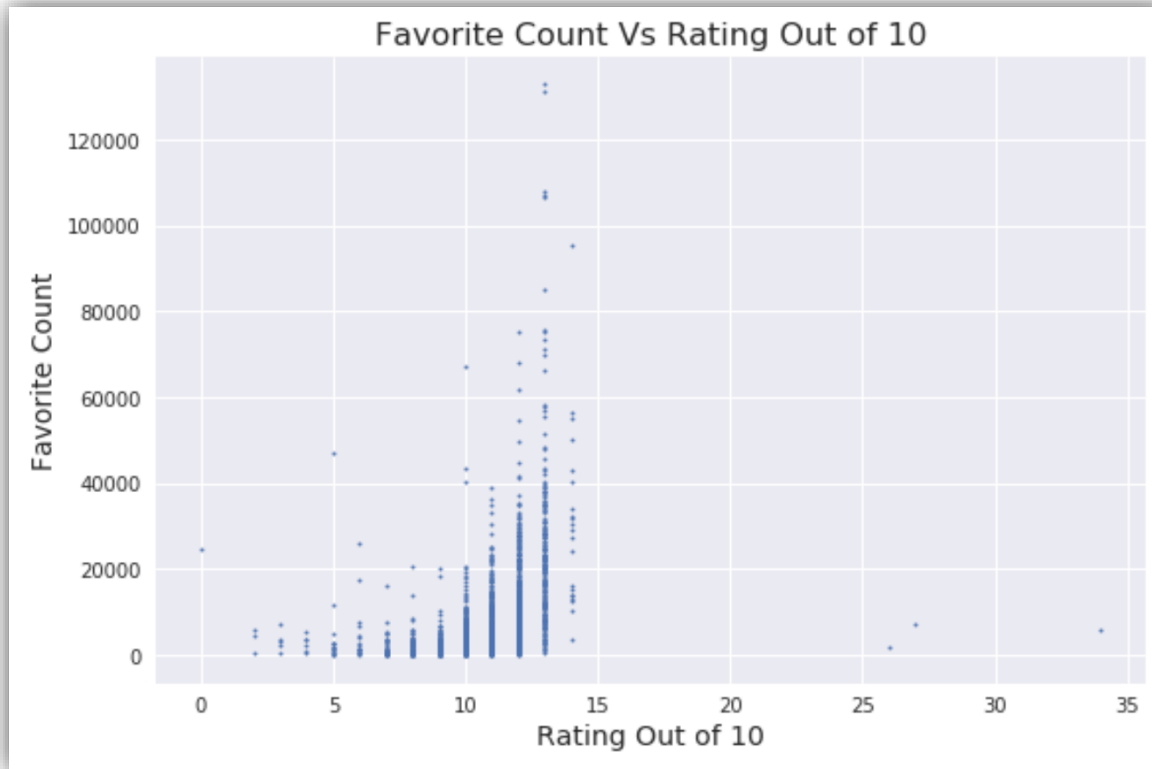
And here we are again, with two scatter plots showing Weak Relationships between the `rating_out_of_10` and both `favorite_count` and `retweet_count`. So as a conclusion , we have now a strong evident that having a high/low rating is now associated in any way with the tweet having high/low favorite/retweet count.

((Final and Very Interesting Insight))
The tweet of least rated dog (0/10), has been favorited more than the one of the top rated dog, who got (34/10) or as on the tweet (24/7)

```
# Least rated dog
master_dataset.sort_values('rating_out_of_10', ascending=True)[['tweet_id','dog_
```

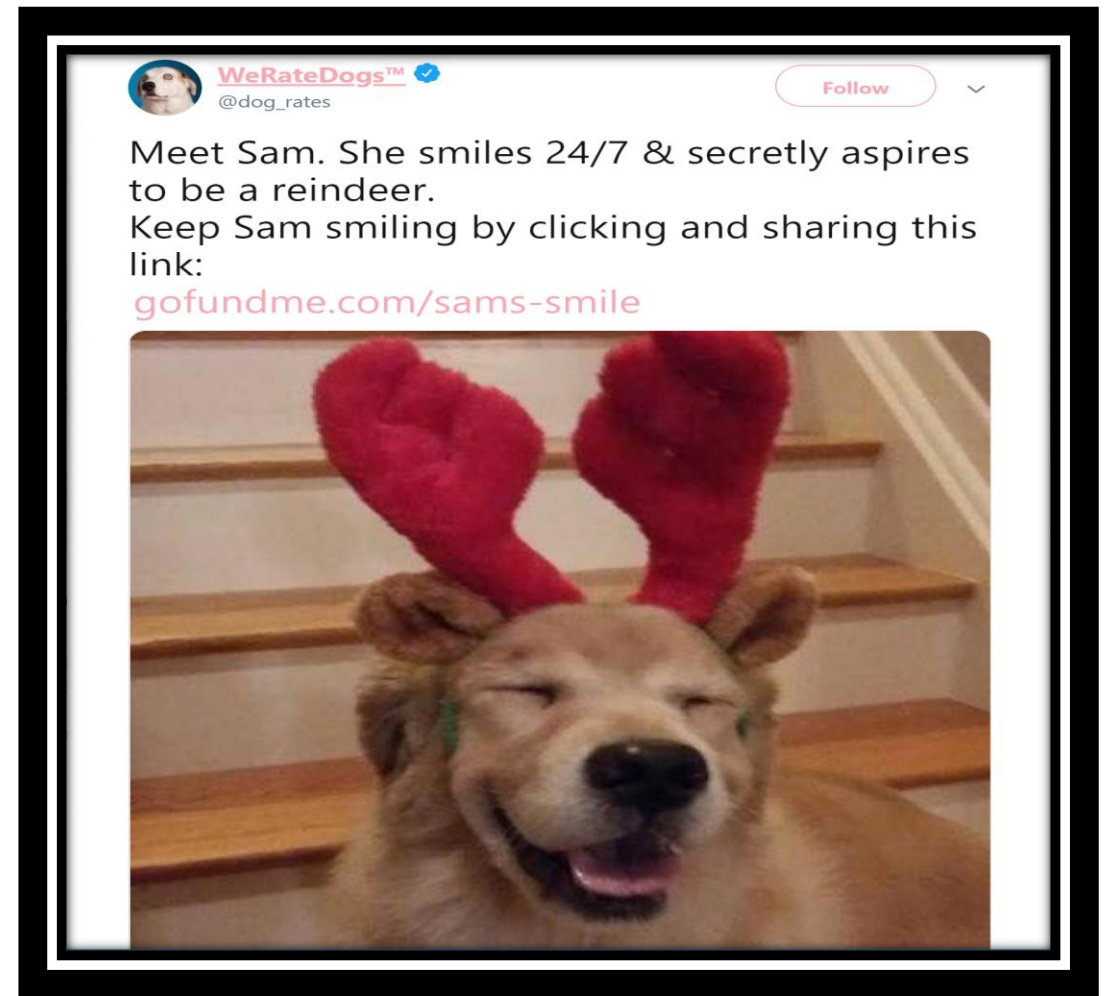|	tweet_id	dog_name	dog_stage	favorite_count	retweet_count	rating_out_of_10
232	835152434251116546	NaN		24574	3443	0

```
# Top rated dog
master_dataset.sort_values('rating_out_of_10', ascending=False)[['tweet_id','dog
```

	tweet_id	dog_name	dog_stage	favorite_count	retweet_count	rating_out_of_10
382	810984652412424192	Sam		5927	1655	34

I am not sure what could be the reason of that high favorite count for the tweet of that poor little dog.

I am wondering !!

One possible reason is : these likes might mean that people are agreeing with the rating (0/10) or they might be liking this tweet to send `WeRateDogs` team a hidden meassage that :

"Even if that POOR LITTLE Dog is UNDER RATED, we are still liking him". Because **"they're good dogs Brent"**