

### Insight and visualization 1

How is retweet\_count and favorite count spread out?

The mean retweet count is 2663.2212.

Quantile spread:

0.25 576.25

0.50 1274.50

0.75 3049.50

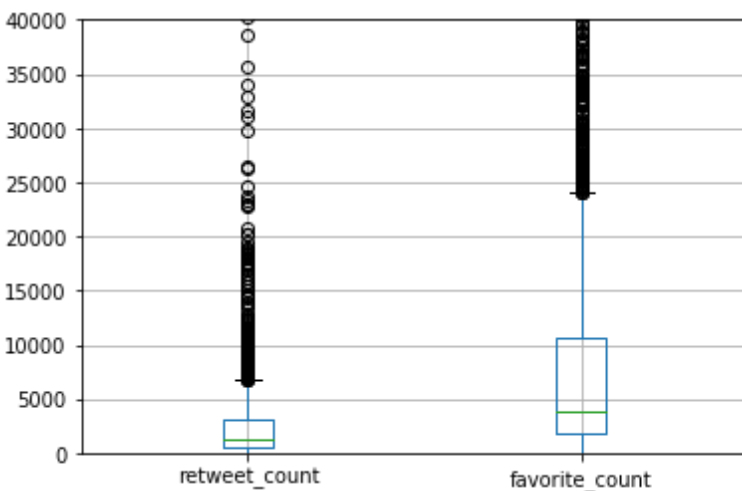
For Favorite count, it looks like this

0.25 1835.75

0.50 3890.00

0.75 10724.00

The boxplot shows an uneven distribution with many outliers. Favorites and Retweets are not evenly spread out.

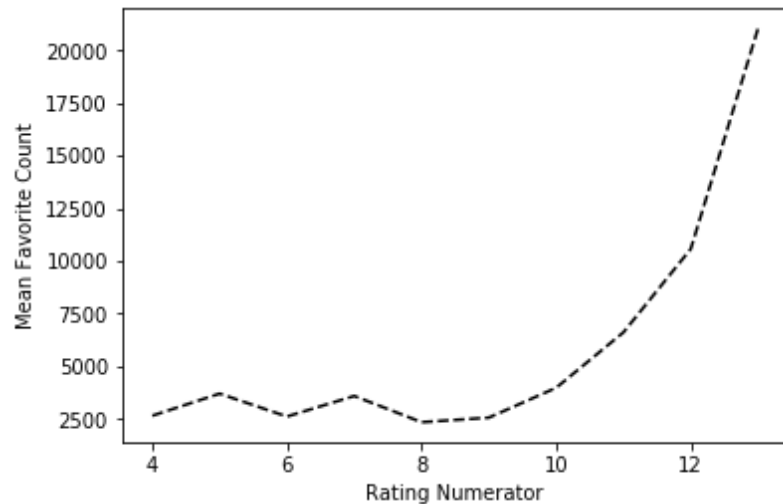


### Insight and visualization 2:

How are ratings (from weratedogs) and tweet favorites related?

The data for rating numerator seems to have errors, the highest rating is 1776

For the report, only checking ratings between 4 and 13, which excludes 45 dogs from our analysis which I view as outliers.



Comparing the Numerator and the mean favorites, there is a clear upward trend. Once the ratings reach 10, the favorites go up as well. This might be a direct correlation, users are more inclined to favor a tweet of a dog if it was highly rated, to begin with. Most dogs are rated favorable, the ratings clearly increase at rating 10. 80% of all dogs are rated higher than 10

rating\_numerator

4	17
5	36
6	32
7	54
8	98
9	156
10	442
11	426
12	499
13	307

### Insights and Visualization 3: Most popular dog breeds

Using the image predictions, which dogs get most favorited and retweeted?

Checking out the three most favorited and retweeted dog breeds, Golden Retriever is clearly in the lead, followed by the Pembroke Corgi and the Labrador.

It is however not clear how many dogs are actually mixed breeds, as the image recognition seems to recognize only pure breeds.

