Analysing We Rate Dogs Twitter Data

In this post, we analyse the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs 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 "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

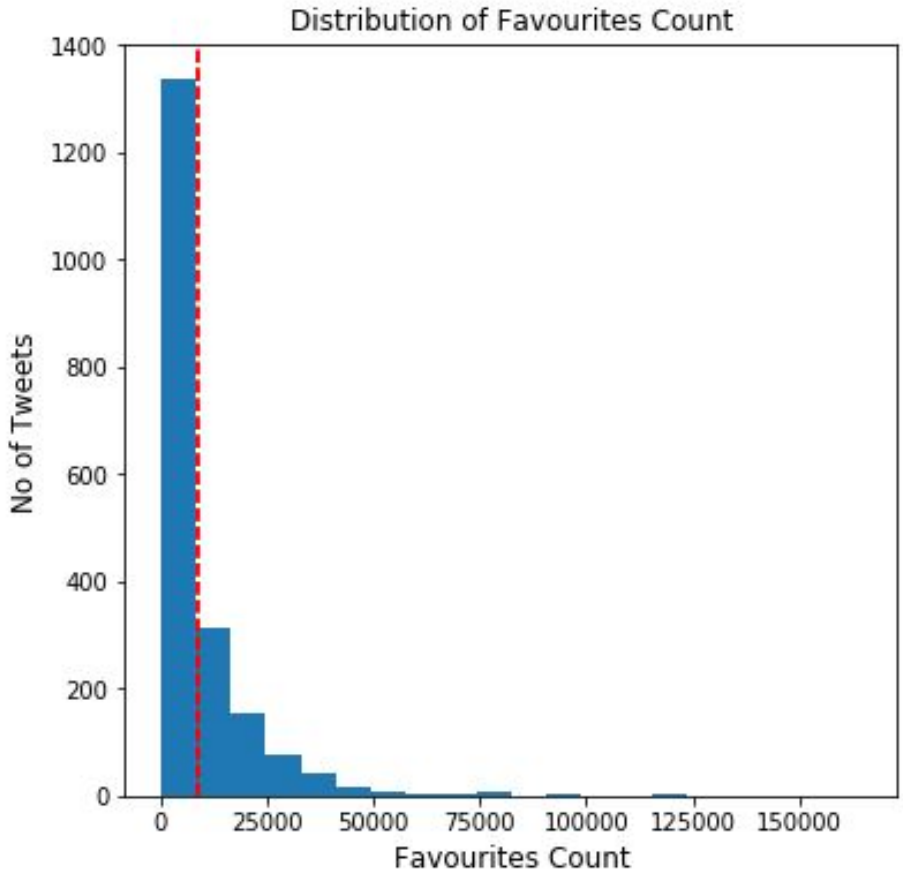
The data gathered from WeRateDogs tweet archive contains information about the dog like dog name, dog stage, favourite count, retweet count and dog breed prediction by a neural network.

Image via Boston Magazine

Analysing the distribution of data:

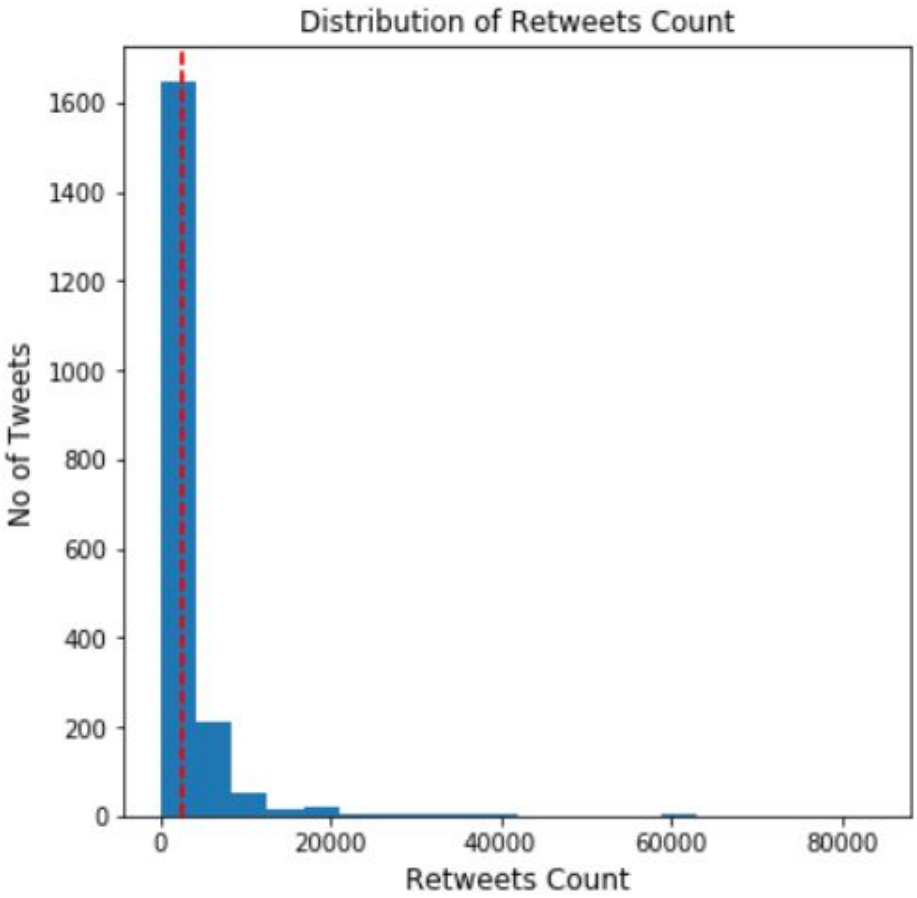
We analyse the distribution of data in each parameter using histograms.

Favourite Count:



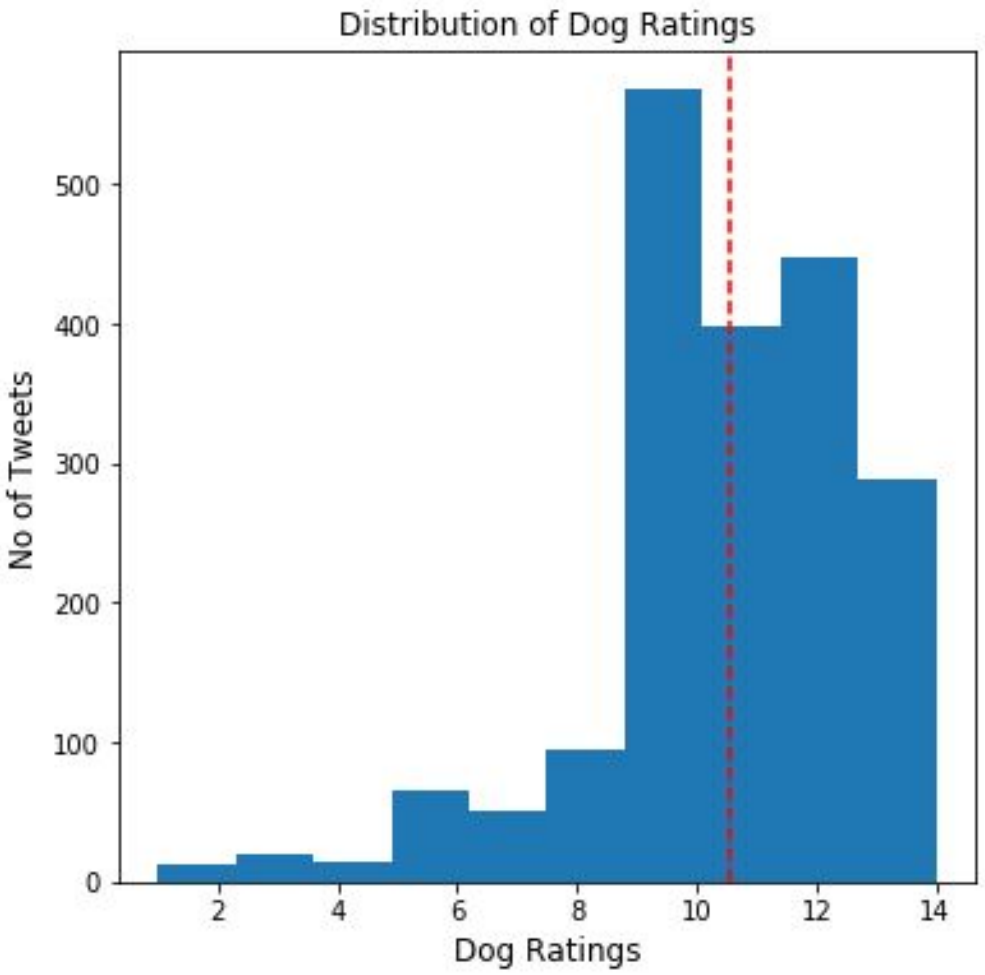
The distribution of favourites count appear to be right skewed with a mean of 8801.1

Retweet Count:



The distribution of retweets count appear to be right skewed with a mean of 2679.1

Dog Ratings:

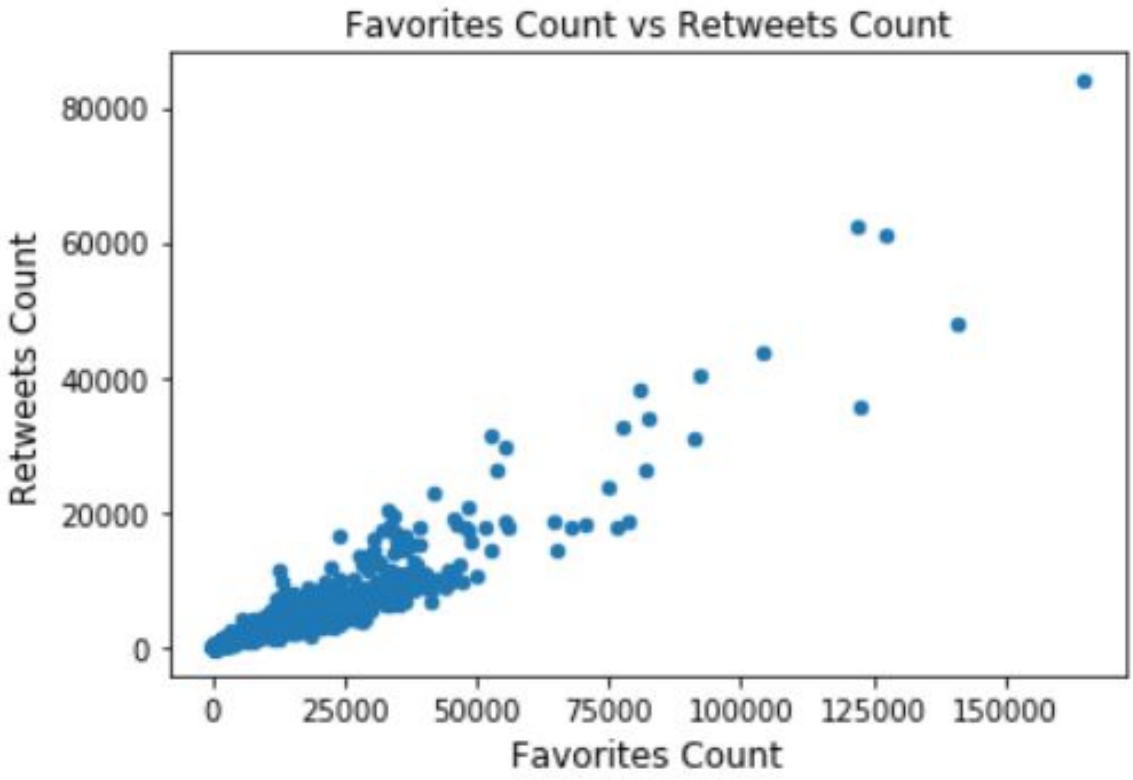


The distribution of Dog Ratings appear to be left skewed ranging from 1/10 to 14/10 with a mean of 10.6.

Analysing the relationship between parameters in the data:

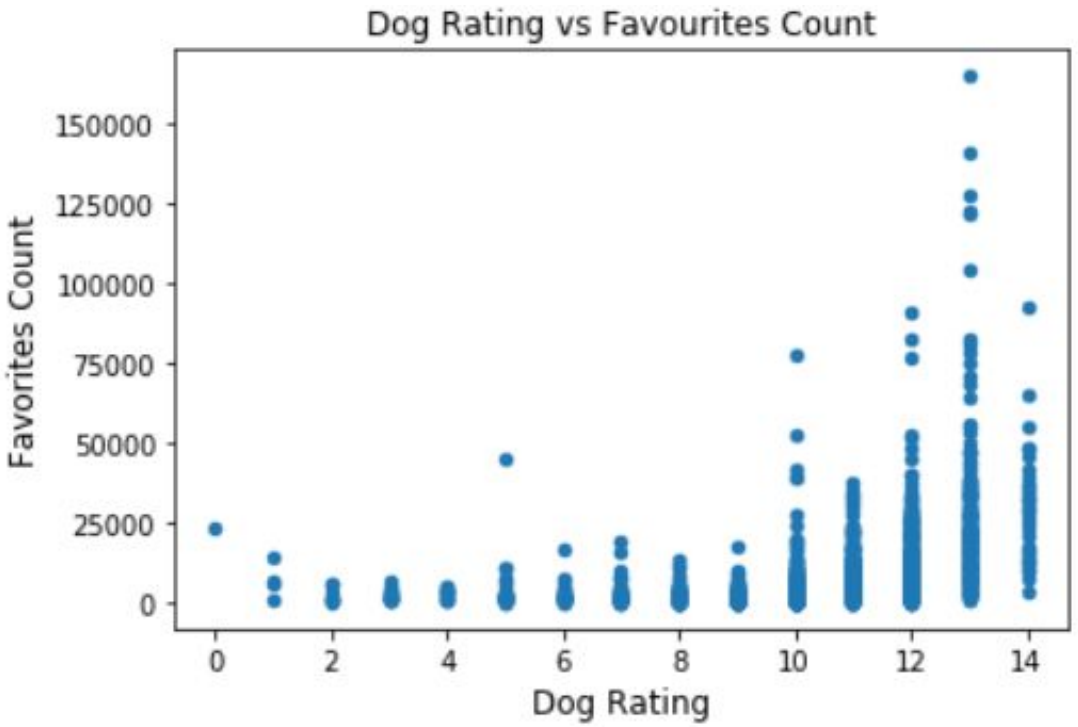
We use scatter plots and correlation to analyse the relationship between parameters in the data.

Are Favourited tweets, also been retweeted?



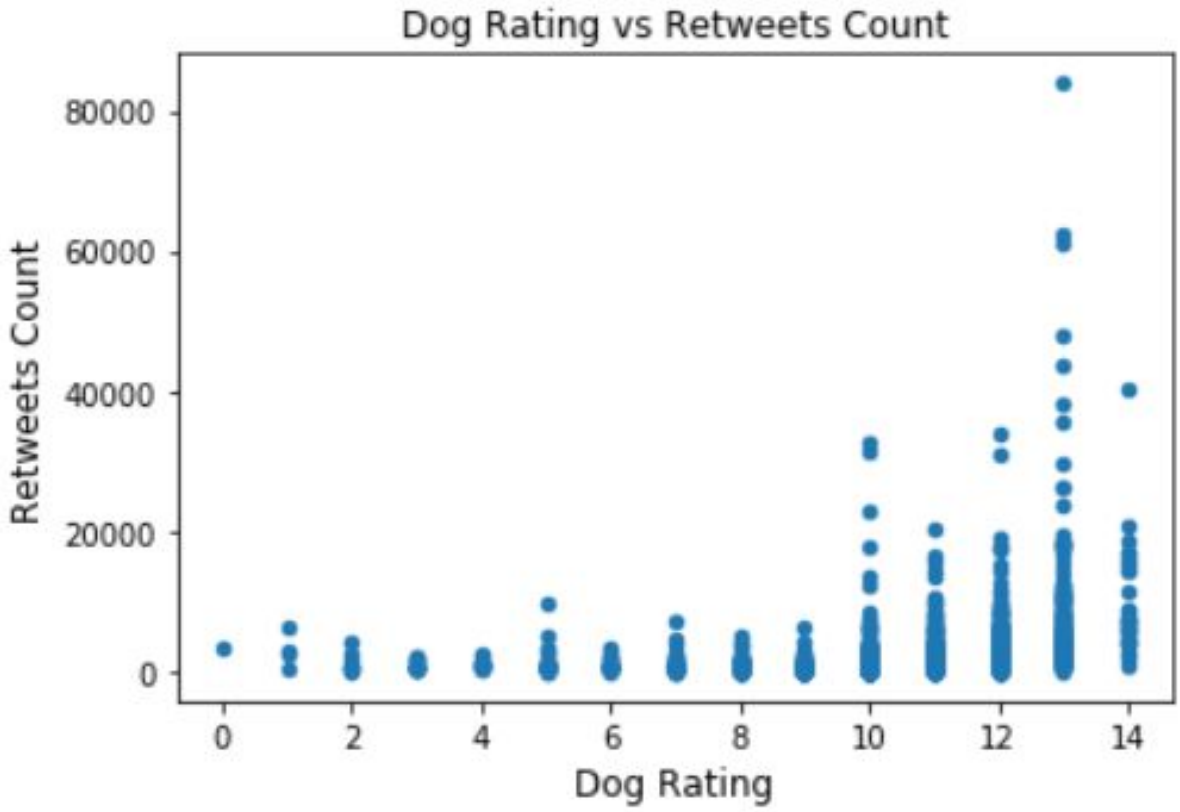
There seems to be a strong positive correlation(0.93) between favourites count and retweet count. Hence we can conclude that most of the dog rating tweets that are favourited were also retweeted.

Are tweets with better dog rating been favourited more relatively?



There seems to be a weak positive linear relationship(correlation = 0.39) between dog ratings and favourites count. Hence we cannot infer that tweets with better dog ratings have relatively more favourite counts.

Are tweets with better dog rating been retweeted more relatively?



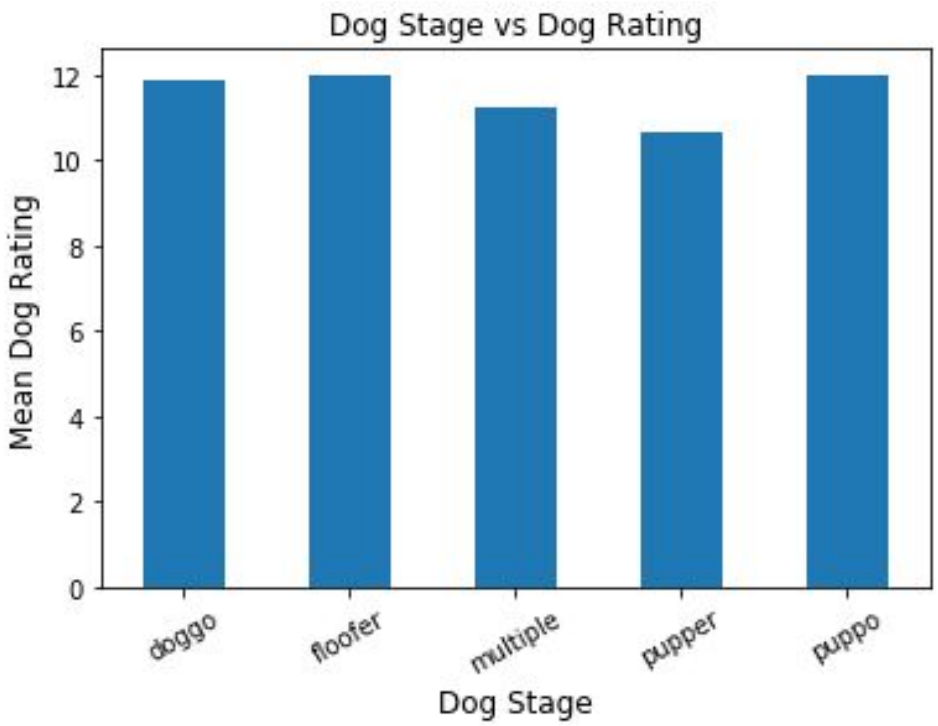
There seems to be a weak positive linear relationship(correlation = 0.30) between dog ratings and retweets count. Hence we cannot infer that tweets with better dog ratings have relatively more retweets counts.

Analysis based on dog stage:

For analysis based on dog stage, we are considering only those 303 tweets that have a dog stage entry out of the gathered 1963 tweets

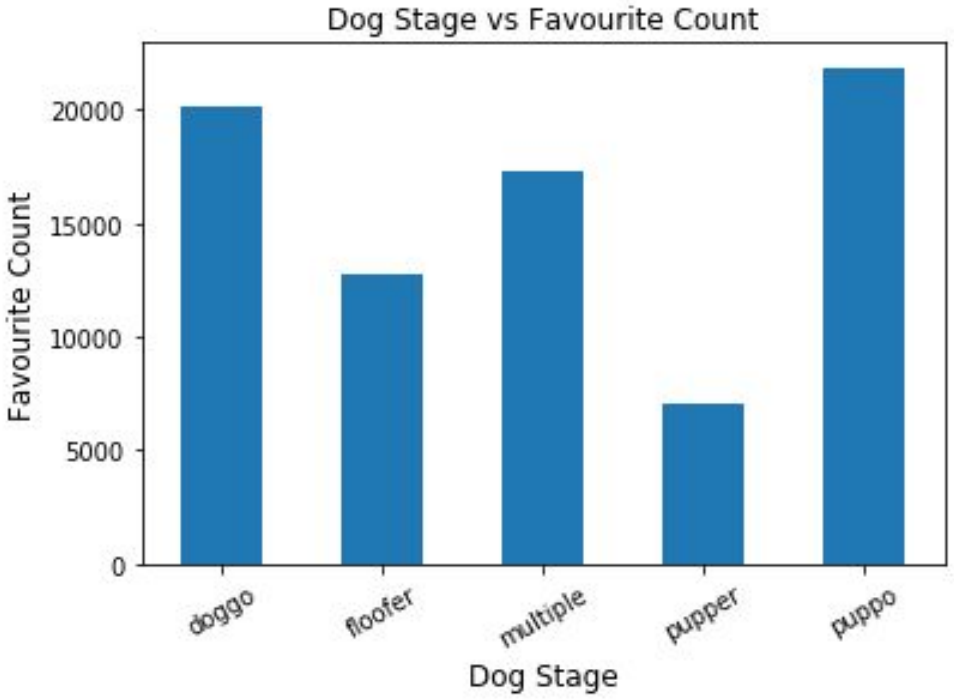
We use bar plots for analysis based on dog stage groups.

On average, Dogs in which state got better ratings?



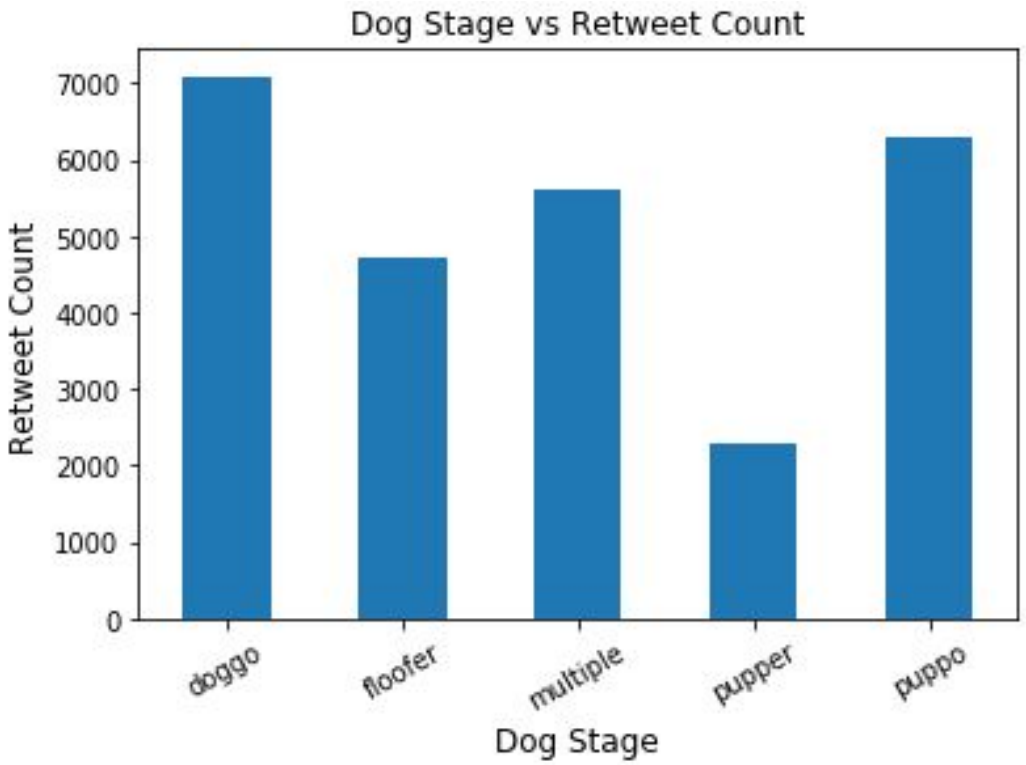
Dogs of stage 'pupper' have got the least average ratings and the average ratings of doggo, floofer and pupper aren't significantly different to arrive at a conclusion.

On average, Dogs in which state were favourited more?



Looks like the puppo's are the most favourited dogs while the pupper's are the least favourited.

On average, Dogs in which state were retweeted more?



Looks like the doggo's received the highest retweets while the pupper's are the least retweeted.

Conclusion:

From the limited dog stages data available and based on the dog ratings, retweet and favourite counts, it can be seen that people those who have responded to these tweets like puppo's and doggo's relatively more and pupper's were the least favourite.

Lets welcome the winner, **a Puppo.**

Source: WeRateDogs Twitter user