

# WeRateDogs Twitter Archive Analysis

## Analysis

The data from the WeRateDogs twitter archive contains the following information – The tweet id, the favourite count, retweet count, the tweet text, image URLs etc. The dog stages and ratings were extracted from the tweet text. The predictions data consists of the image URL, 3 predictions, the confidence of each prediction as a percentage and a Boolean whether each prediction predicts a dog breed or not. Below are a few questions analysed from the provided data.

### Which dog stage generated to the most retweets and likes on an average?

An average of the favourite and retweet counts were taken for each dog stage and plotted on bar charts.

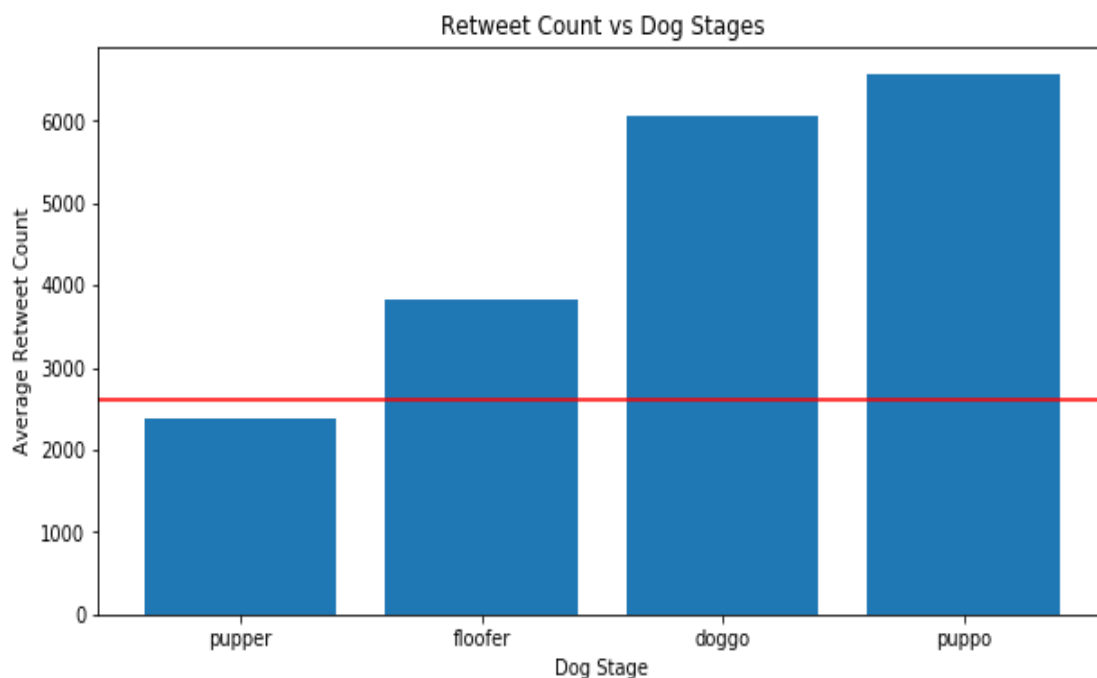


Figure 1

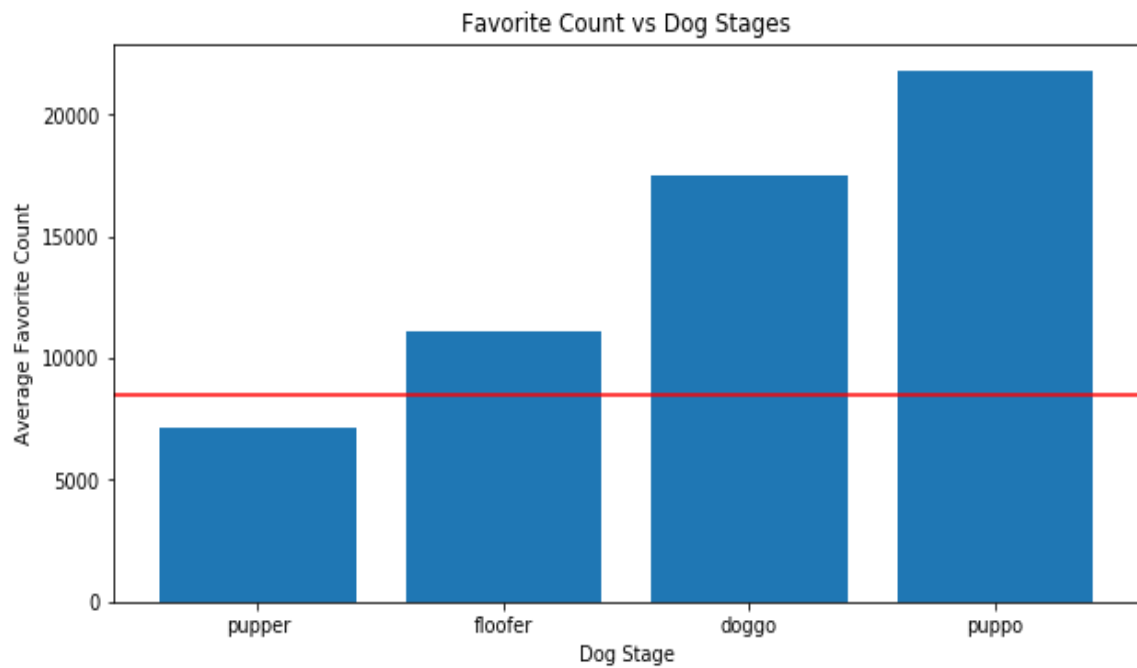


Figure 2

Based on the figures 1 and 2 above, the following conclusions can be made.

- 1) The retweet count and the favourite count follow a similar pattern with regards to the dog stages, with puppo having the highest likes and retweets followed by doggo, floofer and pupper.
- 2) puppo, doggo and floofer had above average likes and retweets while pupper had below average likes and retweets.
- 3) These can't be taken as absolute values as I believe the dog stage classifications are very subjective.

Is there a correlation between the user provided ratings and the retweet count?

A regression test was done between user provided ratings and the retweet count. The following results were obtained.

OLS Regression Results						
Dep. Variable:	retweet_count		R-squared:	0.000		
Model:	OLS		Adj. R-squared:	-0.000		
Method:	Least Squares		F-statistic:	0.4950		
Date:	Fri, 22 Mar 2019		Prob (F-statistic):	0.482		
Time:	11:37:43		Log-Likelihood:	-21423.		
No. Observations:	2172		AIC:	4.285e+04		
Df Residuals:	2170		BIC:	4.286e+04		
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
intercept	2599.2641	103.813	25.038	0.000	2395.682	2802.847
rating_score	16.5319	23.498	0.704	0.482	-29.549	62.613
Omnibus:	2820.829	Durbin-Watson:	1.733			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	585549.492			
Skew:	7.040	Prob(JB):	0.00			
Kurtosis:	82.195	Cond. No.	4.62			

Figure 3

Since the rating score has a high p-value, there is not enough evidence to reject the null hypothesis that the slope is 0, that is there is no linear relationship between the rating score and retweet count.

Which dog stage has the highest proportion of dog predictions?

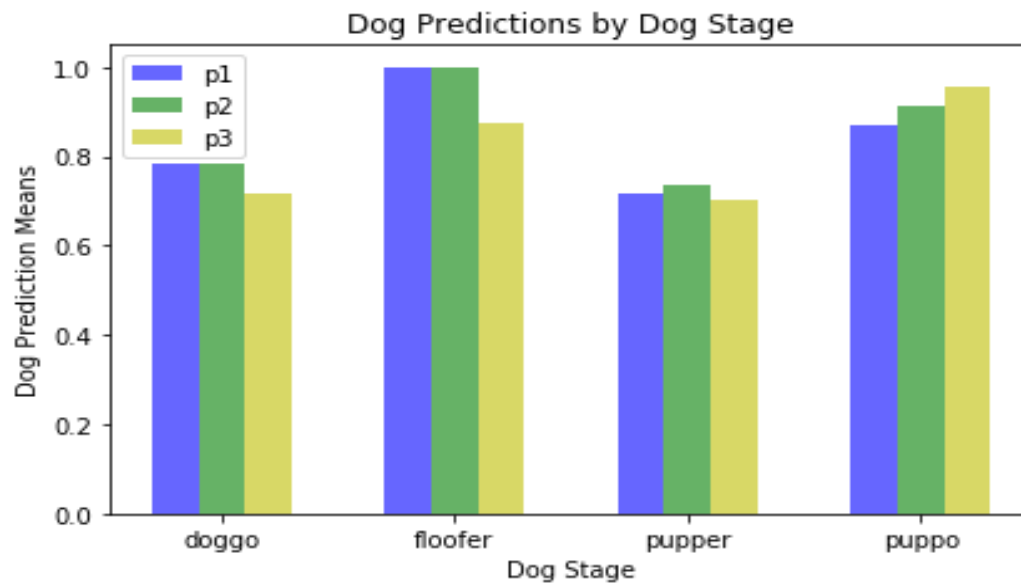


Figure 4

Based on the figure above, Floofer had the highest dog predictions percentage, followed by puppo, doggo and pupper respectively