Breed and Dog Type Sentiment Analysis

In this work I'm going to analyse the data collected from WeRateDogs™ Twitter account located at https://twitter.com/dog_rates (further, Source) twitter account. I'm going to argue that the likes and retweets are not independent from the numeric ratings given to various breeds. Next, I'm going to show that the rating given to dogs is impartial and does not depend on dog stage.

It is known that the rating denominator is 10 in all cases and rating numerator can be bigger than 10, so I omitted rating denomenator in my calculations as it does not make a difference.

Breed Sentiment Analysis

Breed Ratings, Likes and Retweets

First we're going to take a look at what breeds get the highest average rating, likes and retweet counts. Breed predictions are extracted from the picture of the dog using machine learning. I've taken most probable breeds (of three variants given).

| Top 10 breeds by ratings		Top 10 breeds by likes		Top 10 breeds by retweets	
west_highland_white_terrier	29.75	english_springer	60528.5	english_springer	23441
labrador_retriever	15.33	lakeland_terrier	52115.67	standard_poodle	20708.5
siberian_husky	15.27	standard_poodle	48520.5	lakeland_terrier	17438
irish_terrier	13	italian_greyhound	47733.5	italian_greyhound	11041.5
bloodhound	13	french_bulldog	34462.9	giant_schnauzer	9653
bedlington_terrier	13	giant_schnauzer	33036	tibetan_mastiff	9337
yorkshire_terrier	13	irish_terrier	27393	eskimo_dog	8956.7
giant_schnauzer	13	eskimo_dog	24481	french_bulldog	8835.6
afghan_hound	13	irish_water_spaniel	24334	samoyed	6829.29
irish_setter	13	tibetan_mastiff	24026	chihuahua	6264.84

Breed Ratings Correlations

I built a regression model using likes and retweets as dependent variables and rating as a predictor. Here's the gist of it.

Likes:

	coef	std err	t	P>|t|	[0.025	0.975]
ratings	9513.8742	2461.118	3.866	0.000	4601.462	1.44e+04
intercept	-9.755e+04	2.91e+04	-3.349	0.001	-1.56e+05	-3.94e+04

R^2 is 0.182
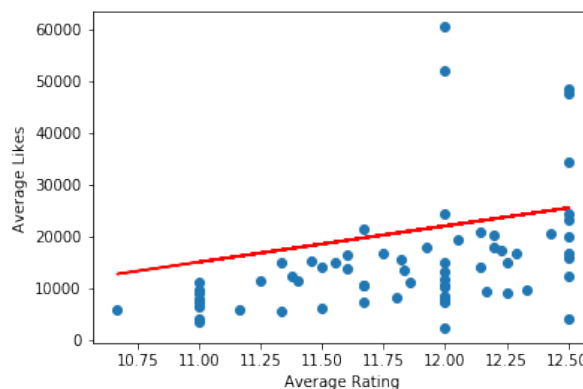
Retweets:

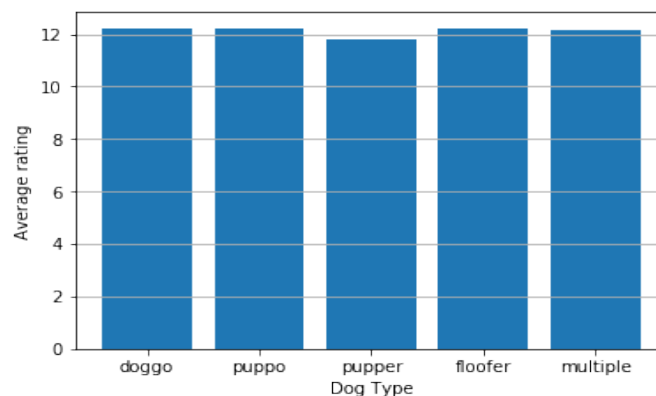|	coef	std err	t	P>|t|	[0.025	0.975]
ratings	2986.67	924.665	3.230	0.002	1141.035	4832.314
intercept	-3.112e+04	1.09e+04	-2.843	0.006	-5.3e+04	-9270.829

R² is 0.135
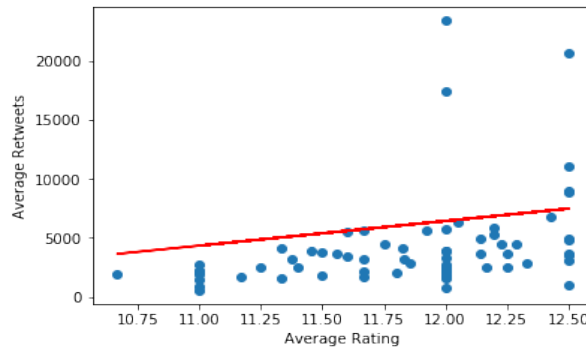
Data stated above shows that there is a weak but statistically significant correlation between average rating and likes and retweets an individual breed gets.

Dog Type Sentiment Analysis

Now let's look at the dog types. The dog type loosely represents a dog stage and has been extracted from tweet text into a categorical type with 5 possible values: doggo, pupper, puppo, floof and miltiple (which represents dog that has gotten multiple stage designations).

Dog type	Average rating
doggo	12.23
puppo	12.19
pupper	11.81
floofer	12.2
multiple	12.17

The ratings are almost the same. In order to find statistical significance of that I've build multiple regression model using dog type as a predictor and rating as a dependent variable. Here's the result:

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	12.1667	0.464	26.203	0.000	11.244	13.089

doggo	0.0591	0.507	0.117	0.907	-0.949	1.067
puppo	0.0208	0.544	0.038	0.970	-1.061	1.103
pupper	-0.3611	0.502	-0.720	0.473	-1.358	0.635
floofer	0.0333	0.689	0.048	0.962	-1.335	1.402

R² is 0.030

The calculations show that there's no statistically significant dependency between dog type and a rating it receives. I conclude that the rating is impartial in relation to dog type.