WeRateDogs – User Engagement Analysis

Tom Schonig, February 27th 2019

They Let Me Think I'm the Master…

…But dogs have taken over my life. My fiancé and I have two dogs and run a petcare business (the Cuddly Cottage) – and then the topic for my next Udacity project turned out to be wrangling and analyzing the WeRateDogs Twitter account! What an opportunity, to write code *about* dogs!

"A Doggo By Any Other Name…"

As an enthusiastic dog owner, I've been caught calling my pup innumerate, unintelligible nicknames. The phenomena of 'doggos', 'puppers', etc, have also led to countless hours of meme-entertainment. This made me curious whether there's any meaningful correlation between these silly, fun-loving names and social media engagement - or, if 'a dog is a dog is a dog.'

Thankfully, WeRateDogs offered a powerful dataset to examine this question, as they frequently use silly terms for their dogs. We identified tweets where the dog was described as a 'doggo', 'floofer', 'pupper', and 'puppo', and performed multiple linear regressions against each tweet's favorite / like and retweet activity.

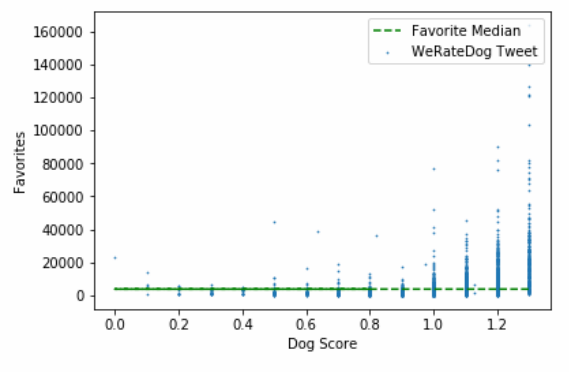
We had some surprising, statistically significant results:

- All-else equal, a dog described as a '***puppo'*** will have ***higher engagement*** by both metrics (on average: 13,940 more favorites / likes and 4,172 more retweets)
- All-else equal, a dog described as a '***pupper***' will have ***fewer favorites / likes*** (on average: 1,492 fewer favorites / likes)

Do the WeRateDogs Scores Predict User Engagement?

An observer of the account will easily remember the running gag – most of the dogs score above 10/10. But the account also makes plenty of jokes, sometimes causing the dogs to score *extremely* high or close to zero. So another natural question that the data could answer: do these scores have any use in predicting how 'successful' a tweet will be, in terms of user engagement metrics?

Visually, it looks like there might be – we can quickly see that their most-favorited tweets are all dogs with the highest scores.



But I want to know, not just assume based on a few data points. So, I passed each tweet's favorite / like count and score into a linear regression model. Perhaps-surprisingly, the model suggested there was no significant relationship. So, we conclude that **the rating has no predictive power**.

Do the Dogs Matter at All?

Answer: Of course they do! They're good dogs.

Or, at least, most of them are. Account followers may be familiar with another type of frequent joke: another type of animal is posted, with a low score for failing to behave like a dog. *Talk about being held to an unfair standard.*

So another question emerges: does their audience like these jokes? Is there any significant relationship here?

Thankfully, we also had access to prediction data from a neural network that predicted whether each photo was *actually a dog*. So we, once again, pass the favorite activity and the "is a dog?" prediction into a linear regression model…

…*And the audience doesn't like it*. With statistical significance, each tweet that doesn't contain a dog will, on average and with all-else equal, get 1,300 fewer favorites.