

WeRateDogs Data Wrangling and Analysis

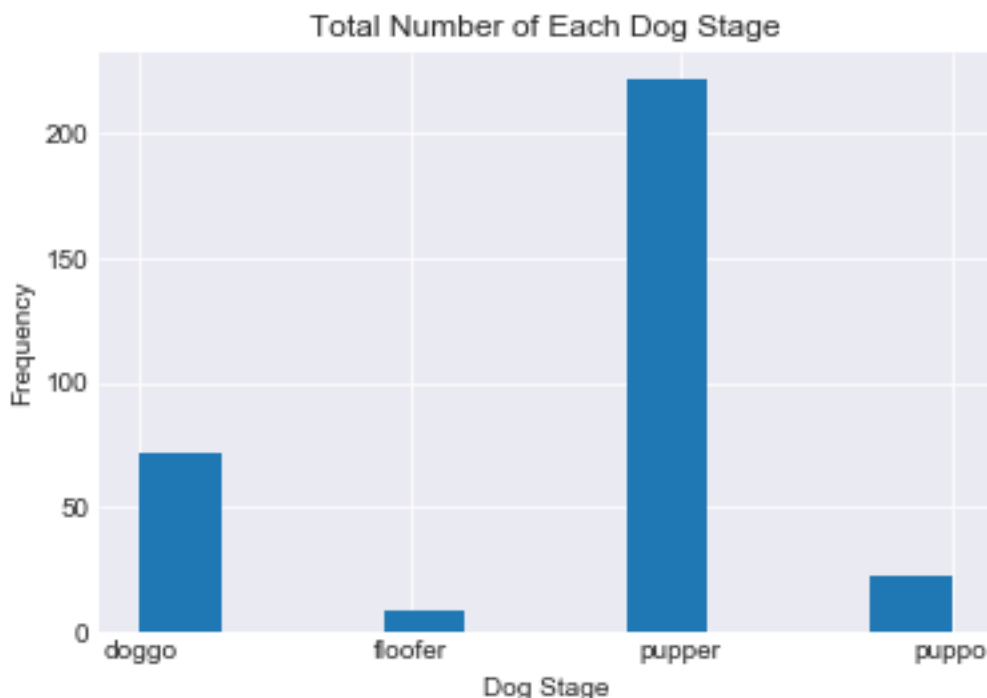
By Aaron Johnson

In this project I took tweet data from the popular WeRateDogs twitter account. WeRateDogs regularly posts images of dogs from all over and gives them a unique rating. The posts are meant to be humorous and the ratings are usually given on a scale from 1 to 10. The ratings are usually above 10. There are also various stages of dogs assigned to each dog. For example, one dog could have a dog stage of 'puppo'. Another could have a stage of 'doggo'.

Gathering the data for this project involved taking data from a few different sources. The bulk of the data was provided, but was a bit messy. Before this data could be used in analysis and visualization I had to clean it up a bit so that it all worked smoothly. Another source of data involved using Twitter's API to get even more information about the WeRateDog tweets.

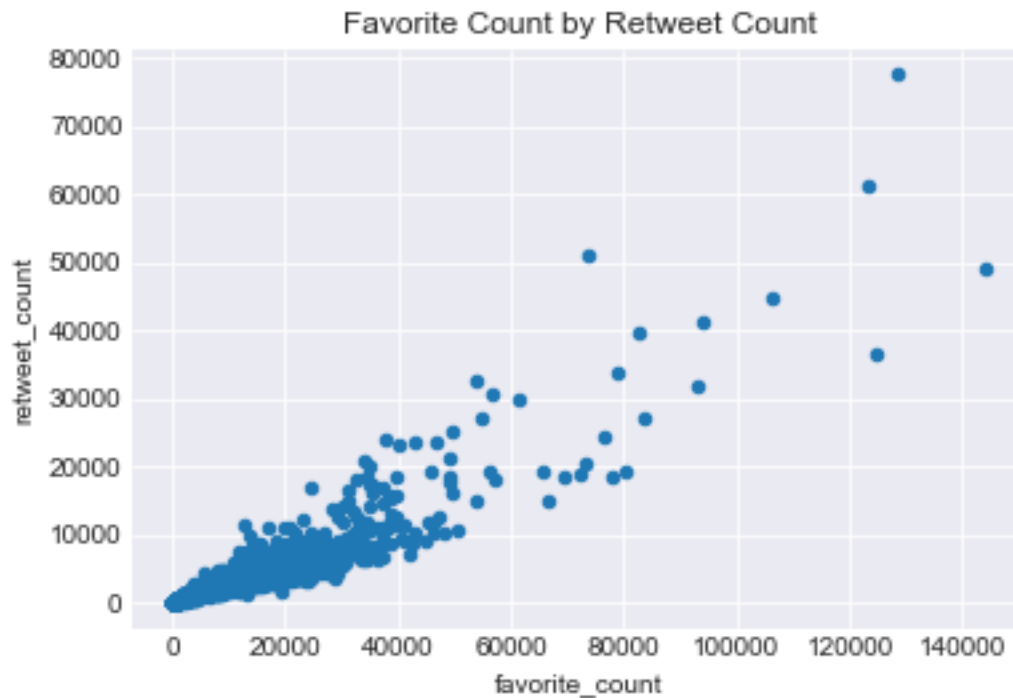
The final cleaned dataset had quite a few variables. Some of the more interesting ones that I was curious to explore was the rating of each dog. Out of the four different stages of dogs (doggo, pupper, floofer, and puppo) I was interested to see which dog stage was most frequent in this particular dataset. I also wanted to look at the relationship between the retweet count's and the favorite count's.

So...here's what I found:



As you can see here *puppers* were the most represented stage of dog in this dataset. For those interested, a pupper is defined as “A small doggo, usually younger. Can be equally, if not more mature than some doggos” (via the official #WeRateDogs book).

As for the relationship between retweet count and favorite count:



Here we can see what looks like a positive relationship between favorite count and retweet count. This makes perfect sense to me. The more a tweet gets retweet, the more likely it seems that it would be favorited.