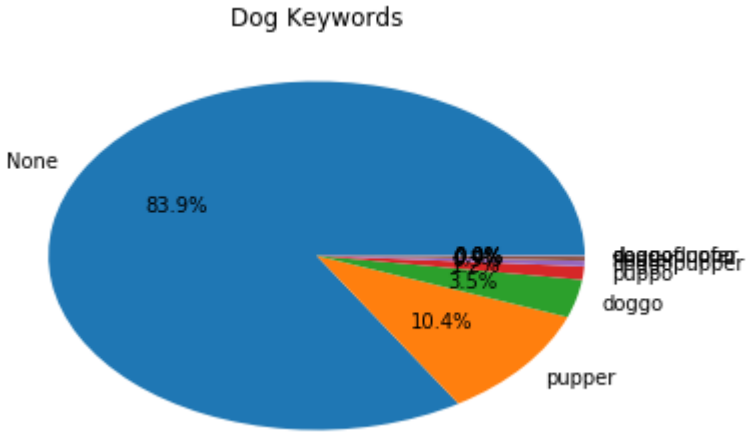
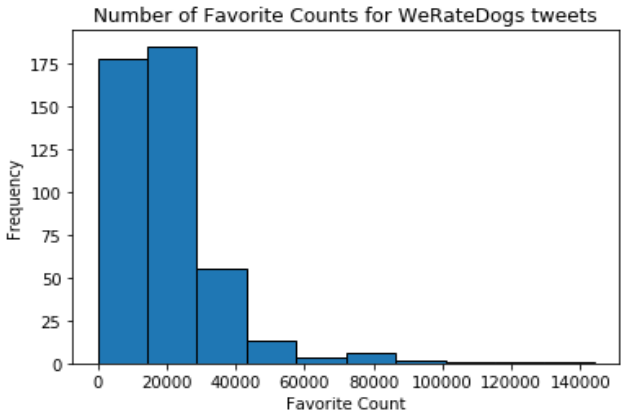
Act Report

Having wrangled the WeRateDogs twitter data, I moved to make assessments and conclusions about the data using visualizations. I first wanted to look at the dog keywords for the tweets (pupper, floofer, doggo, puppo) and see the distribution.



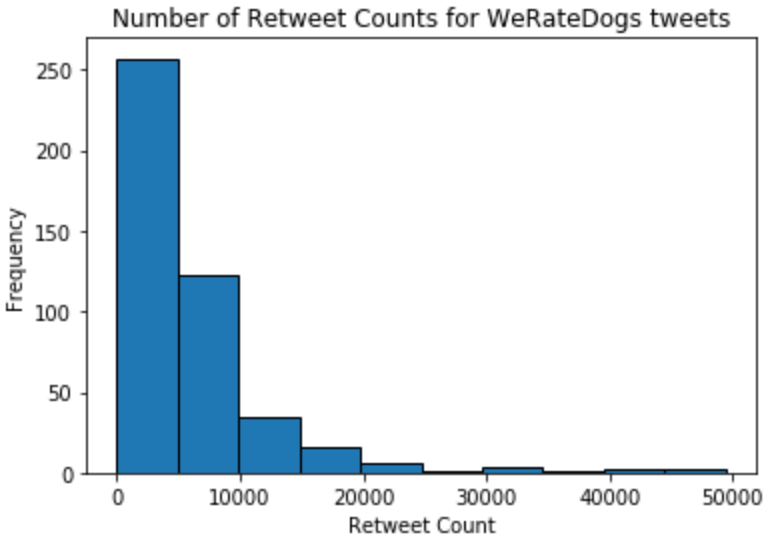
We can see from this that most tweets don't have one of those keywords attached to them at all. Of the ones that do have keywords, pupper is the most popular, followed by doggo, puppo and floofer. I certainly find it perplexing as to why so many of the tweets in this data set have no dog keywords associated with them.

I then took interest in looking at the number of favorite counts for the WeRateDogs tweets. I chose a simple histogram to show the frequency and distribution of the amount of favorites each tweet would get.



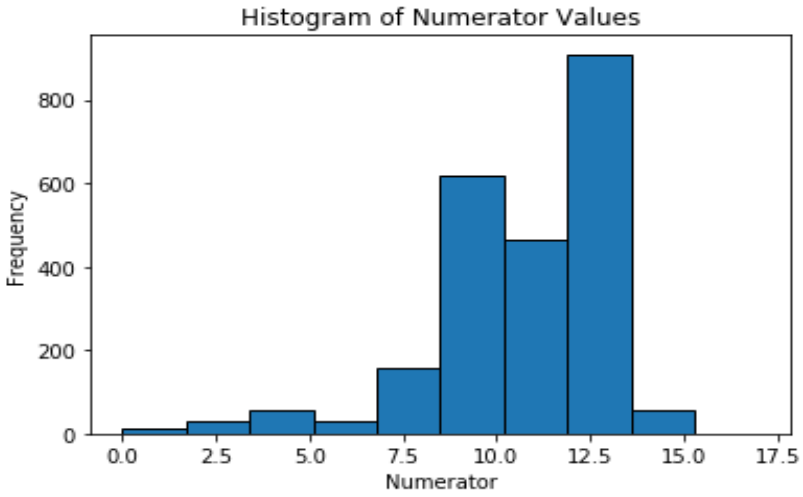
We see here that most tweets get between 10,000 and 30,000 favorites, with very few even reaching as high as ~140,000. The WeRateDogs twitter account is quite popular! The distribution of the favorite counts is very right skewed here, with very few tweets having more than 40,000 favorites.

Next, let's take a look at the retweet count. Logical next step after the favorite count, right? I followed the same process as well and created a histogram.



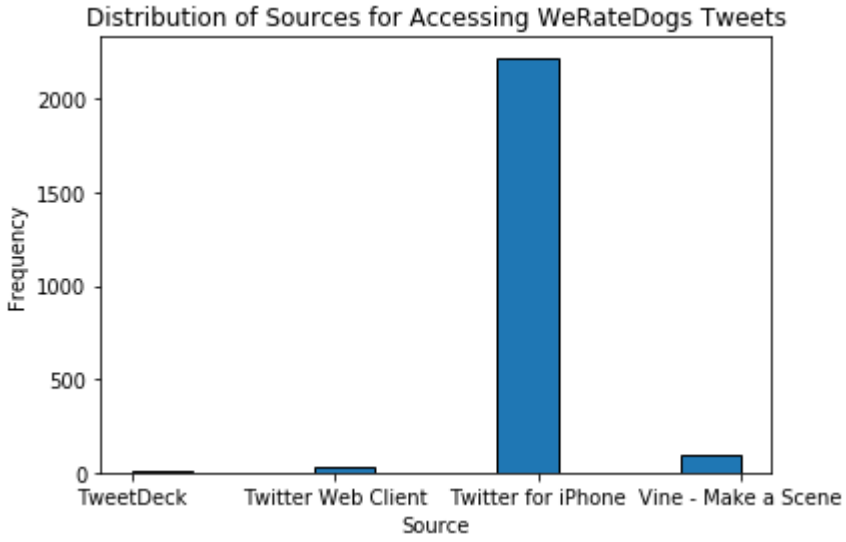
Unsurprisingly, the retweet histogram follows the same sort of distribution as the favorite histogram. This makes sense, as people who favorite the tweets are fairly likely to also retweet them. We see that most tweets get between 0 and 10,000 retweets, with few even going as high as 50,000.

I then wanted to look at the numerator values for the ratings for each WeRateDogs tweet. The denominators were almost always ten, but the numerators were pretty interesting:



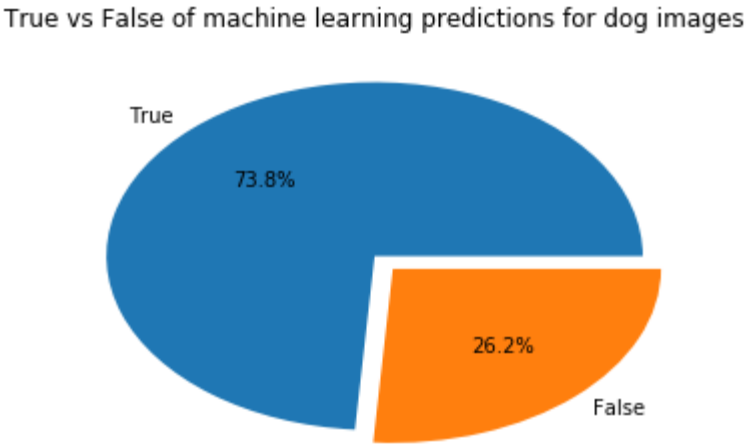
This distribution is left skewed, one we haven't seen yet. Funnily enough, most numerators for the ratings are above 10. WeRateDogs really likes dogs a lot.

Next, I took a look at the source each viewer of WeRateDogs accessed the tweets from.



An overwhelming majority of users access the WeRateDogs tweets from their mobile device. Even looking at the Twitter Web Client, there are hardly any users viewing WeRateDogs on their computer versus their mobile device.

Finally, I chose to look at a set of machine learning predictions about the breed of a dog in a given tweet.



The algorithm, we can see, is fairly accurate. It's not perfect, though; It's wrong about 26% of the time. However, we can rely on this algorithm to predict a dog's breed about 74% of the time.