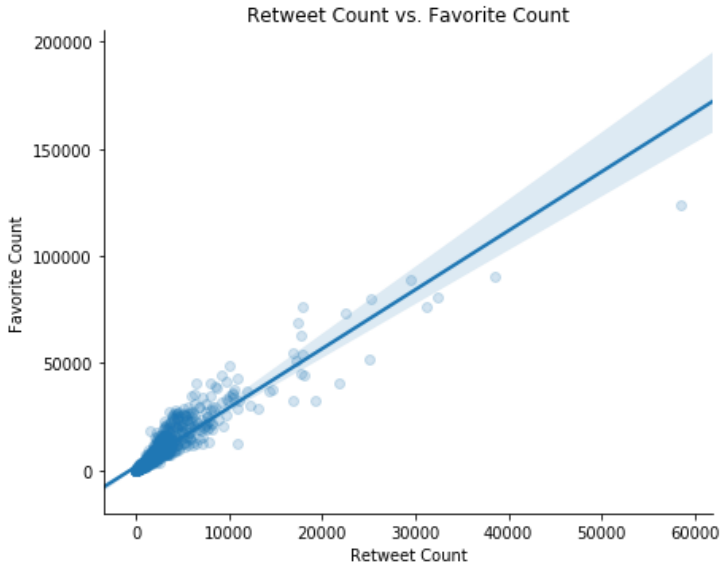
Analysis of WeRateDogs

About the data

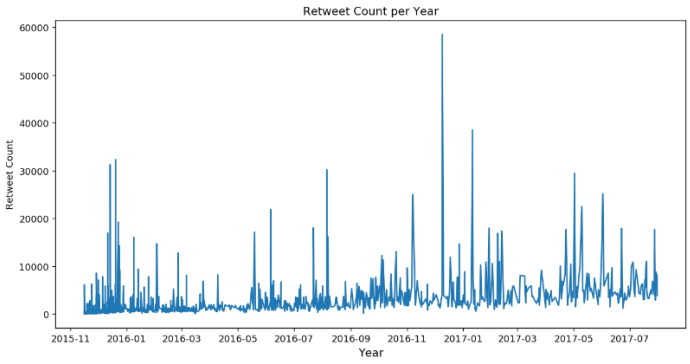
We have used three different sources for this project. The archive of the tweets provided via a CSV from Udacity, the Twitter API used to retrieve more information about the tweets with number of each was retweeted, and the predicted dog breed in each tweet's image programmatically determined from a neural network also provided by Udacity. Combining this data from these sources, we able to analyze a better data correlation.

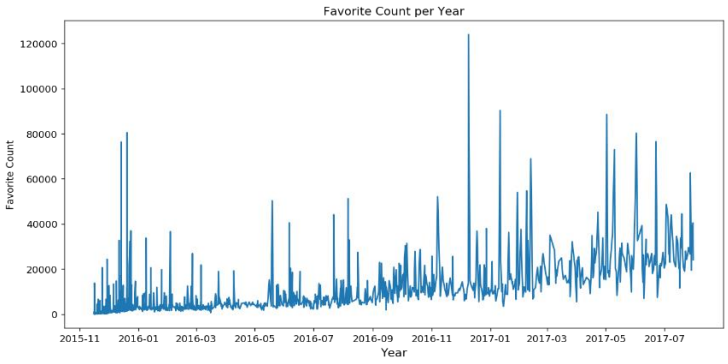
These tweets were from year 2015 to 2017 and were analyzed the correlation between retweets and favorites between the different dog stage, the distribution of dog breeds in the tweets, and the popularity of the dog breeds based on retweets and favorites.

First, I wanted to know how the retweets and favorites correlate. A plot was made (see below chart). Majority of the data commence below 35,000 favorites and below 10,000 retweets. Both counts are highly positively correlated. The most dog tweets had more favorites than retweets. This is more likely expectation since more people will favorite instead of taking ownership of the retweet. I noticed this in many Twitter account profile says, "My tweets are my own" so, they don't retweet.

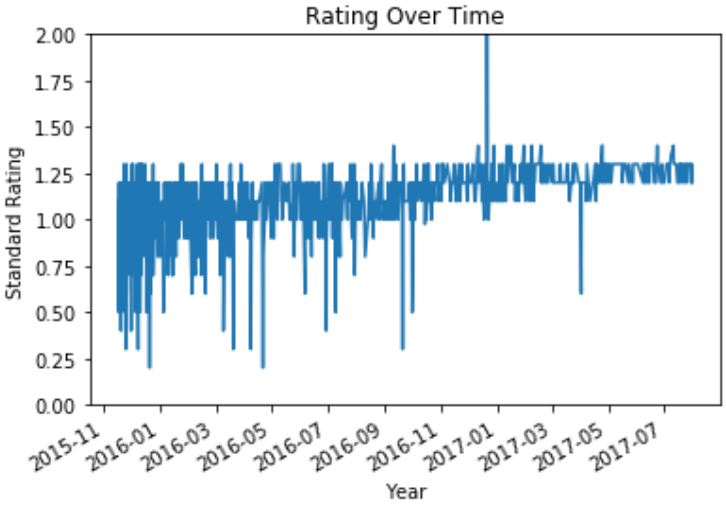


On below charts, I wanted to see the distribution of counts for each retweet and favorite in overtime.
WeRateDogs, a Twitter account started in 2015, the frequency of retweet and favorite were in its starting days until about 2016. After that, the frequency decreases. But the most dog tweets had more favorites than retweets in overtime.





This summarize the frequency rating overtime decreases. Good data distribution from year 2015 to 2016 then reduced dramatically after that.



Golden retriever is the number one most rated dog type. Pembroke as second, and Labrador retriever as third.