# WeRateDogs Twitter Archive - Act Report
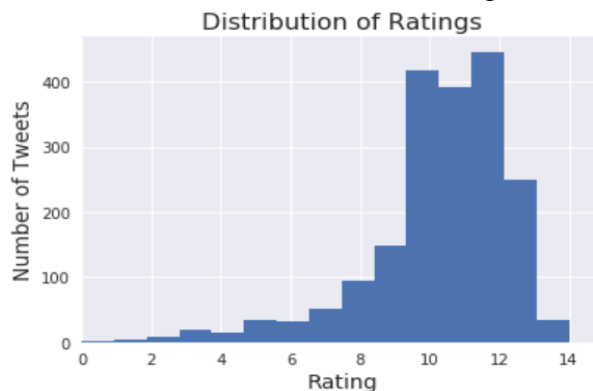
- **Pradeep Gurunathan**

## WeRateDogs Data

WeRateDogs Enhanced Twitter archive contains data extracted from 2356 of the 5000+ tweets from the @dog_rates twitter account, posted between November 15, 2015 and August 1, 217. This data is comprised of dog ratings that were taken from the text of the tweet along with the dog name and dog stage if present. The retweet count and favourite count for each tweet were not included in the enhanced archive, and so I had to download this additional data from the twitter account using the tweet ID from the archive. Along with the Twitter data, I also downloaded an image predictions file from Udacity servers containing the top 3 predictions for dog breeds based on the images from the tweets (there can be up to 4).
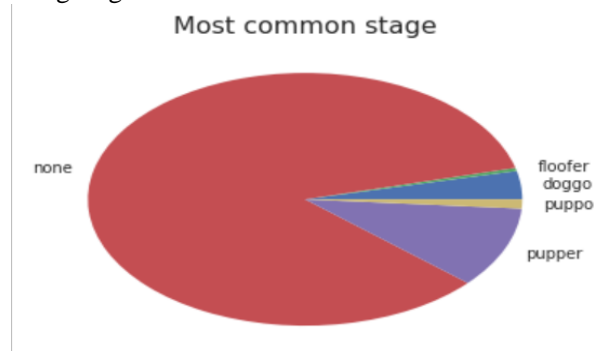
## Wrangling Data

Before I could begin the analysis, the data had to be wrangled into structure to make it easier. I assessed the data both visually and programmatically for quality and tidiness; the quality of data is established mainly by looking at several aspects or dimensions to ensure that it is complete, valid, accurate and consistent. After cleaning many of the issues found during the assessment, there were about 1950 tweets with good quality data.

## Insights

- From the initial analysis is it evident that favorite count more than retweet count.
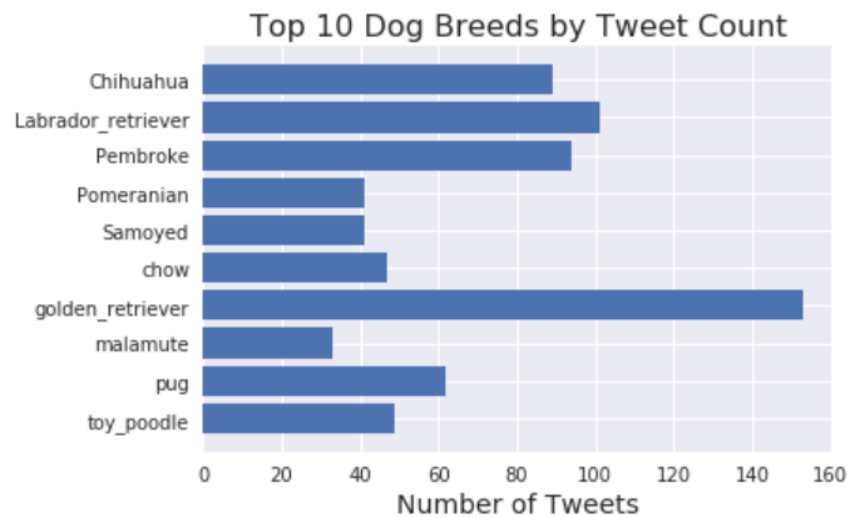- Ratings to the Dogs are skewed towards left with mean rating 10.53 and median rating as 11.



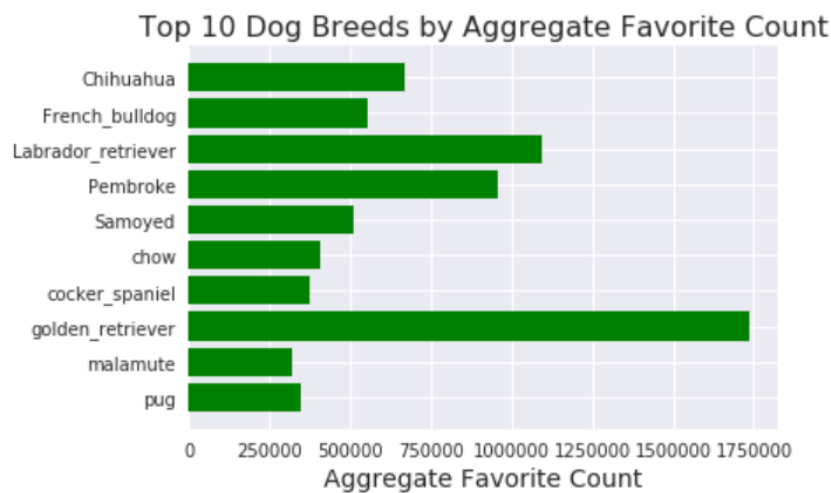- Only a 16% of tweets with dog stages identified.



- Most common dog that is tweeted about, with almost 160 tweets, is the Golden Retriever, as identified (predicted) by the neural network. Labrador Retrievers, Pembrokes and Chihuahuas are fairly close

together in 2nd, 3rd, and 4th place. Further down we'll take a look at what are the most favourited dog breeds.


Top 10 Dog Breeds by Tweet Count

- The plot below shows the top 10 dog breeds ranked by aggregate favourite counts. It looks very similar to the barplot below of top 10 dog breeds ranked by tweet, and in fact the top 4 dog breeds are the same. This also makes sense, as these dogs have the highest number of tweets.


Top 10 Dog Breeds by Aggregate Favorite Count

- The boxplot here shows the distribution of ratings within each of the top 10 dog breeds as ranked by aggregate favourite count. All the breeds range between 7 and 13, with a only a few outliers, while four of the breeds have the lower ratings at 10. Another four breeds have the highest ratings at 14.

Ratings for Top 10 Dog Breeds (by Aggregate Favorite Count)

Distribution within Breeds