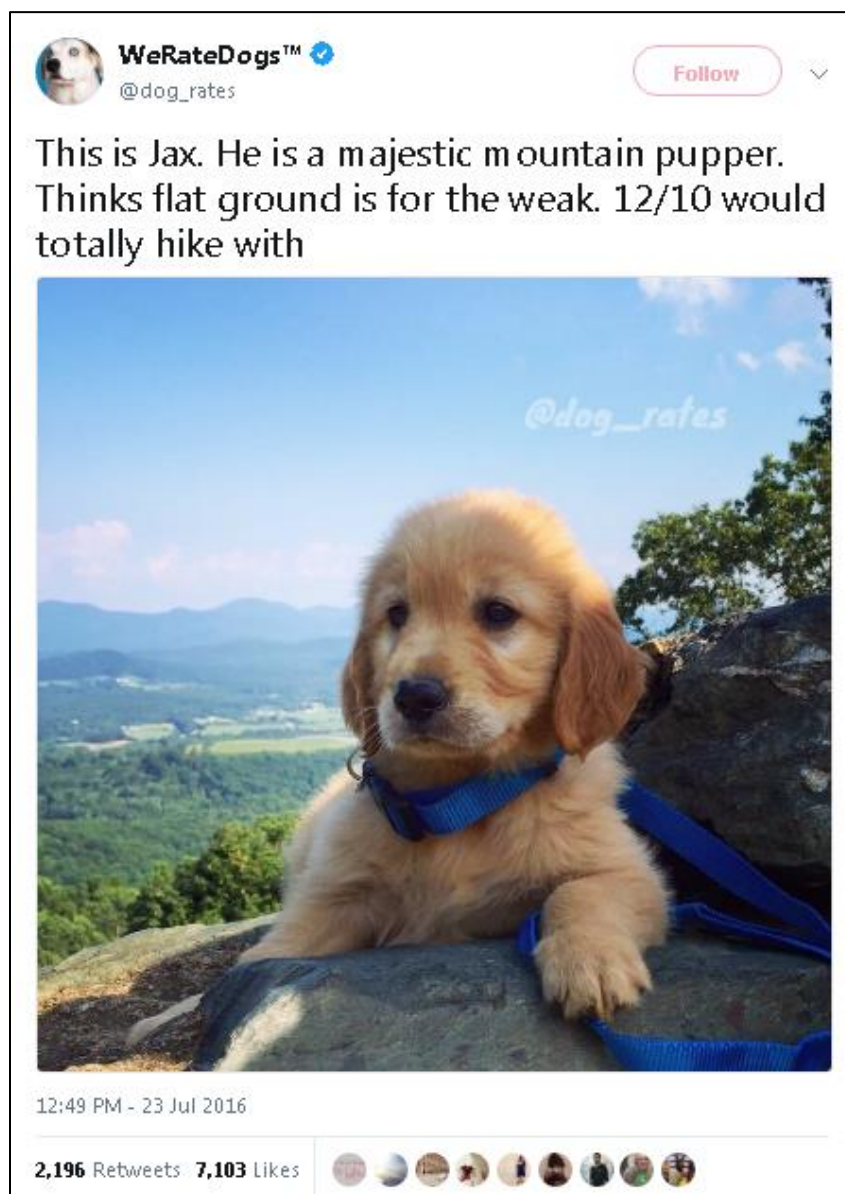


## Raymond Atherley

### ACT REPORT

This report will communicate some of the insights found during this project and display the corresponding visualizations. This project consists of three different datasets containing various tweet information from the WeRateDogs twitter website. An example tweet is shown below. The information includes ratings of the dogs, the dog 'maturity' stages, and the tweet metadata. It also consists of predictions of the dog breeds from the photos within the tweets as well as retweet and favorite count data. The final and master file "twitter\_archive\_master" had 1974 records.

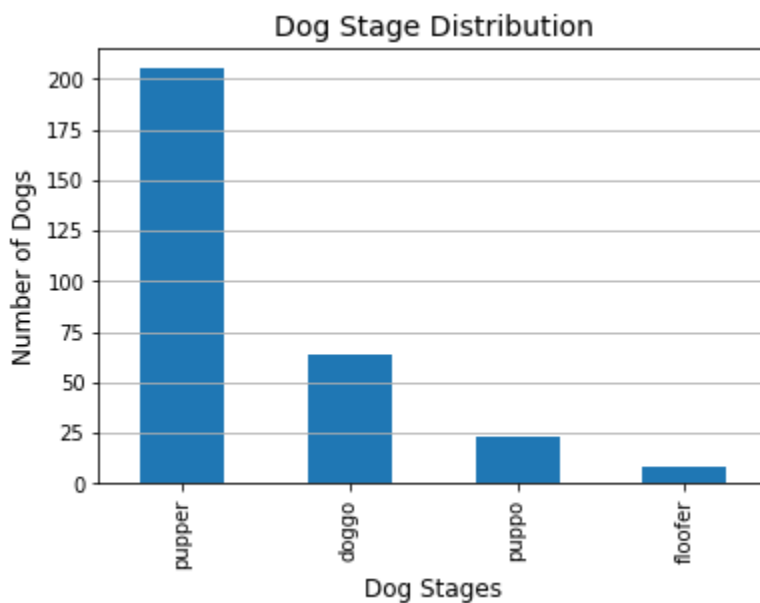


There were three interesting findings that we will discuss.

1. The first is the breakdown of the dog stages. These are as follows:

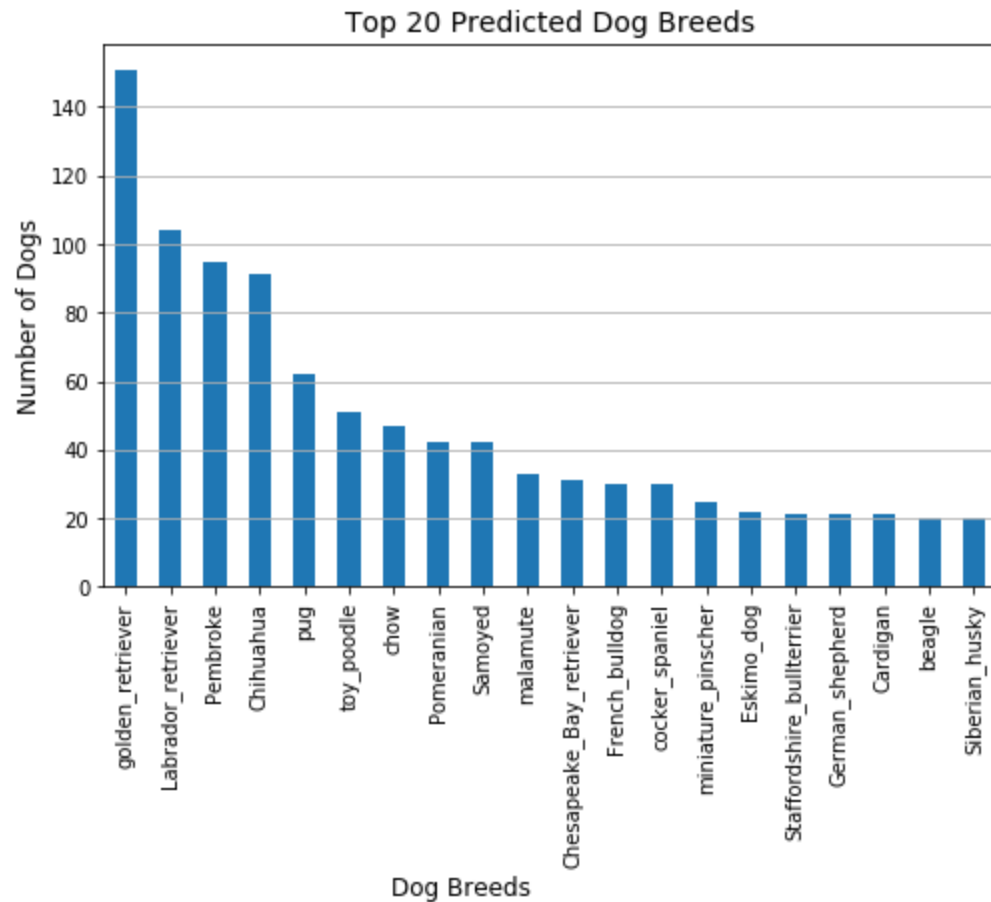
Dog Stage	Number	Percentage
pupper	205	68.3%
doggo	64	21.3%
puppo	23	7.7%
floofer	8	2.7%

It appears that most of the dogs with a dog 'maturity' stage in the dataset are categorized as 'pupper' with about 68.3% of the categorized data. 'Floofer' was categorized the least with only 2.7% of the stage categories.



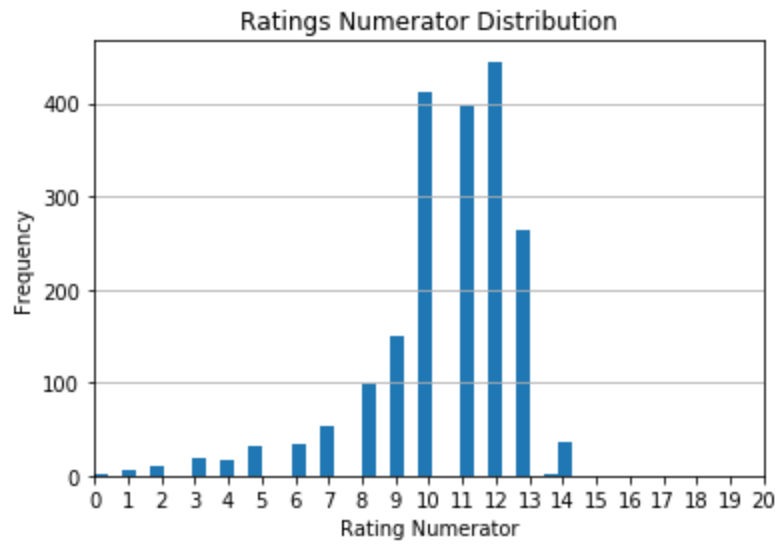
Above shows a plot of the distribution of the dog stages mentioned in the text. Note that for 1674 of the records, a dog stage was not extracted or available.

2. The second insight we will discuss is the predicted dog breeds. There were 113 predicted dog breeds. The plot below shows the top 20 breeds.



We can see in the plot above that the most predicted dog breed is the Golden Retriever with over 150 predictions followed by the Labrador Retriever with around 105. The Siberian Husky rounds out the top 20 dogs with around 20 predictions.

3. The third insight is the distribution of ratings. The statistical variables show that the average numerator is 11.7 and the median is 11. Since all the denominators in the cleaned dataset is 10, the median rating for the dogs is 11/10. The most frequent rating is 12/10. There is an large outlier of 1776/10 referring to a dog being 'really American' and only two entries above 14/10 in total.



We see above a distribution of the numerators and thus the ratings. The maximum x-axis scale was set to 20 as the great majority of numerators were below this value.