# Analysis & Visualizations Report

## Dog Rating

Let's take a look at the descriptive statistics on the rating ratio (rating numerator / rating denominators), we have the following result:
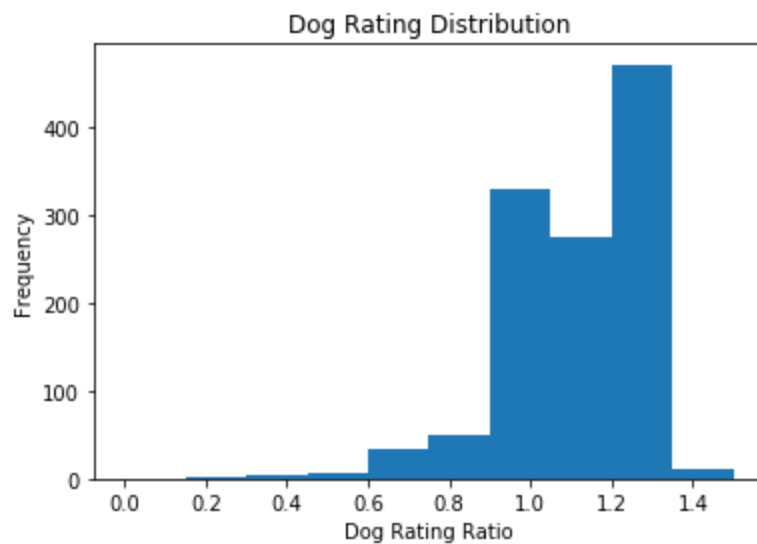
|         | rating_ratio |
|---------|--------------|
| count   | 1186.000000  |
| mean    | 1.097778     |
| std     | 0.264274     |
| min     | 0.200000     |
| 25%     | 1.000000     |
| 50%     | 1.100000     |
| 75%     | 1.200000     |
| max     | 7.500000     |

So we have the lowest dog rating at 0.2 and the highest dog rating at 7.5.

We can also expect the data distribution would be left-skewed distribution since the median is greater than the mean. Let's see how the rating ratio distribution looks like.

Dog Rating Distribution
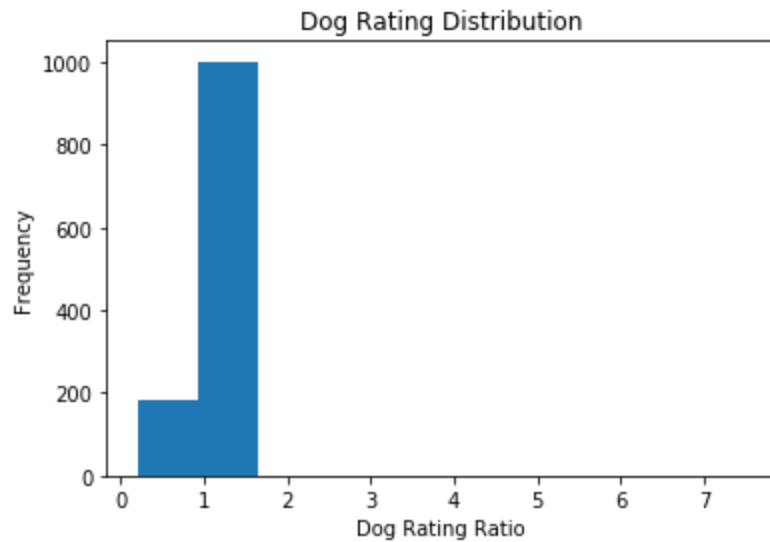
Since most of the dog ratings are in between 0 to 1.5, we have replotted the distribution with range 0 and 1.5.



Dog Rating Distribution

Most of the dogs are rated at between 0.9 and 1.3 and we can find the largest number of the dogs are at 1.2 and 1.3 rating range.

As a dog lover, I was curious to see one of the dogs with the highest rating 7.5 so here it is.
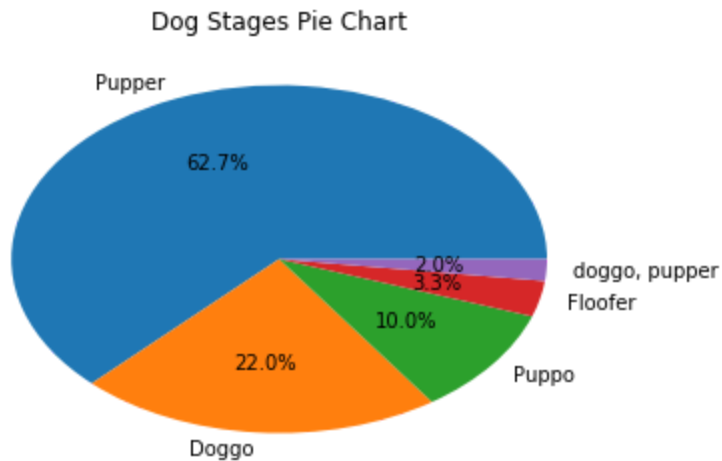
Tweet: This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*ckin magical af 9.75/10
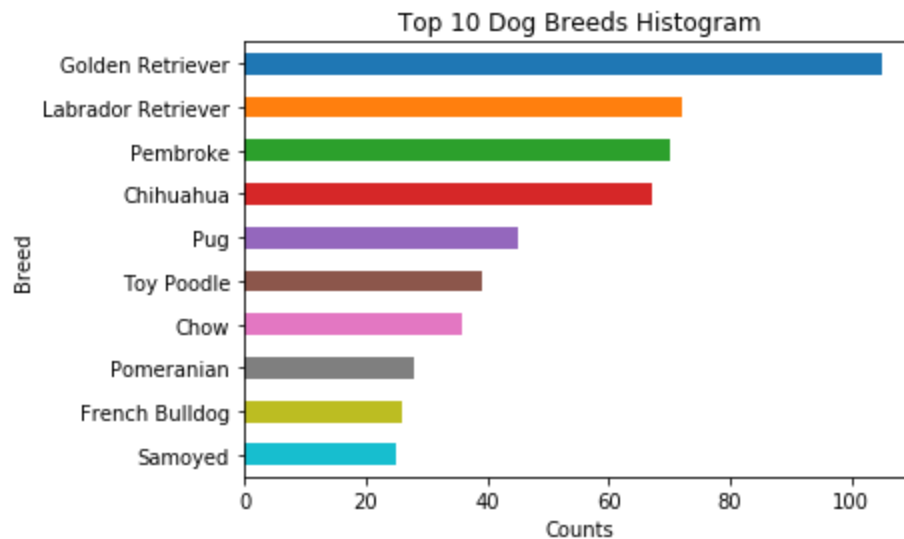
## Dog Stages

Let's have a look at the dog stages percentages.

According to the chart, the most common dog stage is "Pupper" with 62.7% followed by "Doggo" with 22.0%.

**Dog Stages Pie Chart**

Pupper
62.7%

2.0%
3.3%
10.0%

doggo, pupper
Floofer

22.0%

Puppo

Doggo

## Dog Breeds

Through a neural network, we were able to classify breeds of the dogs. The following image shows the top 10 frequent dog breeds that were rated.

**Top 10 Dog Breeds Histogram**

| Breed | Counts |
|---|---|
| Golden Retriever | |
| Labrador Retriever | |
| Pembroke | |
| Chihuahua | |
| Pug | |
| Toy Poodle | |
| Chow | |
| Pomeranian | |
| French Bulldog | |
| Samoyed | |

The most frequent dog breed is "Golden Retriever" followed by Labrador Retriever.
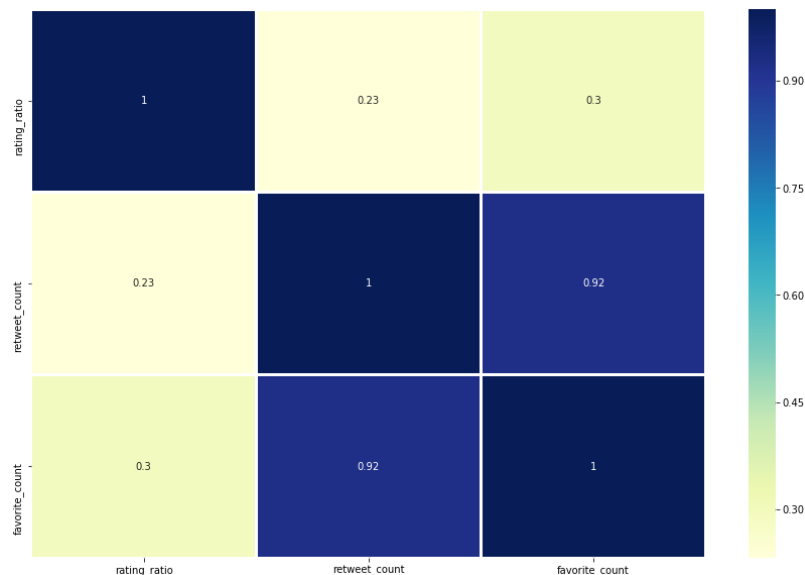
## Retweet vs Favorite

This is the scatterplot between the retweet and the favorite counts per each tweet.
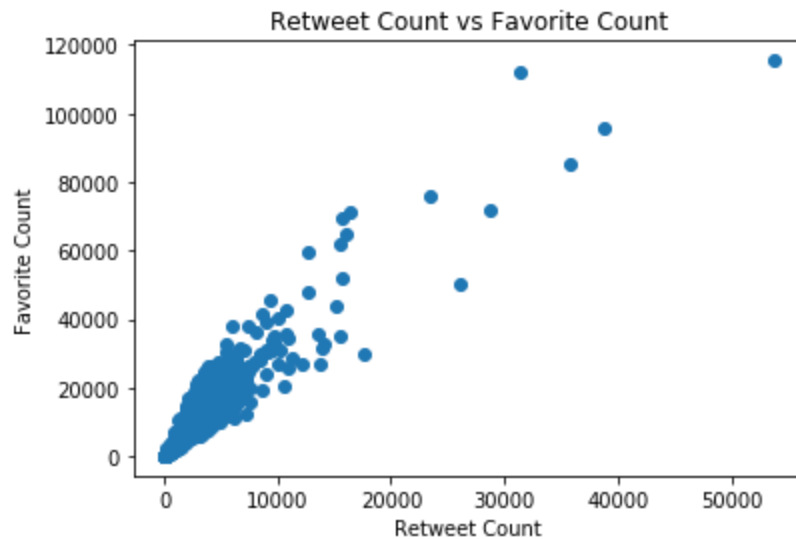


There is a positive linear relationship between Retweet and Favorite counts. As retweet count increases, the favorite count tends to increase too by a constant rate.

## Correlations
We also wanted to find out about the other correlations too such as between rating ratio and Retweet counts or rating ratio and favorite counts.



So there is definitely high correlation between Retweets and Favorites counts with 0.92. However, there are low correlations between the Rating vs Favorite (0.30) and the Rating vs Retweet (0.23)

## Rating Ratio Time Series
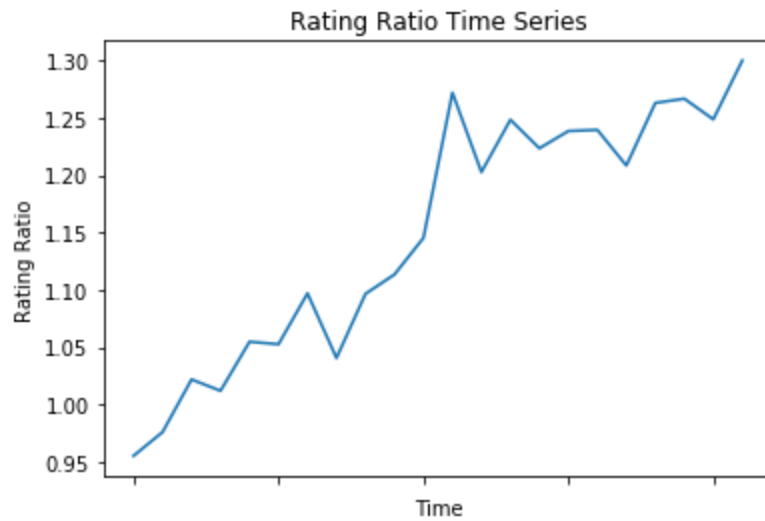


The tweet data were collected from 2015 November to 2017 August. We calculated the mean of the rating ratio for each month and plotted the data as time series.
As we can see the above, the ratings are increasing over time. This is totally expected since the account is getting more popular over time and people can send more dogs into this twitter account so people are likely to find new dogs they might really like.