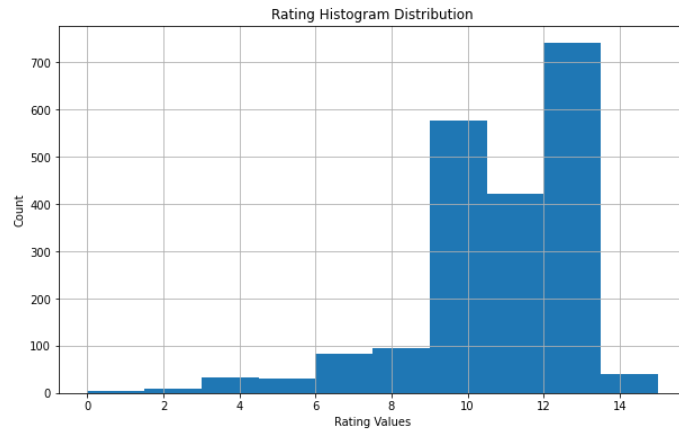# Act Report

In this file, I will share the insights I got from analyzing the WeRateDogs tweeter account tweets collected from three different sources, assessed, and cleaned. The dataset's columns look like this:

- tweet_id: the tweet id
- timestamp: tweeted time
- source: a device tweeted from
- text: tweet text
- expanded_urls: tweet URL
- rating_numerator: rating number ranging 0 - 15
- rating_denominator: max rating number
- name: dog's name
- dog_stage: dog's aging stage
- jpg_url: dog picture
- dog_breed: dog bread
- retweet_count: number of tweet's retweet
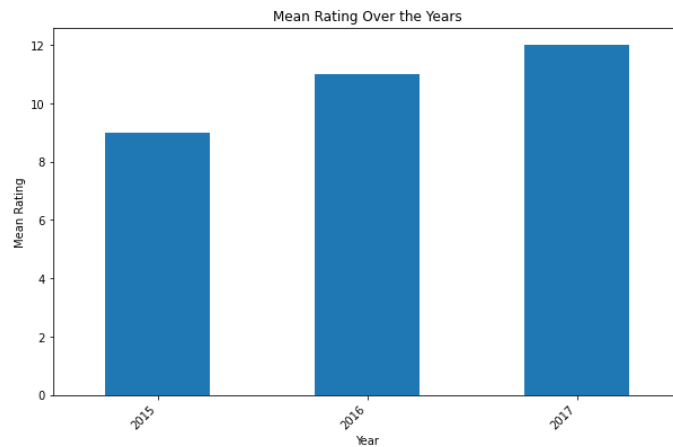- favorite_count: number of likes the tweet received

A clean or tidy dataset alone won't do any job unless we approach the dataset through curiosity and questions. So, I came up with the under-listed questions:

1. What is the average rate value?
   a. For ratings higher than 10
   b. For ratings lower than 10
   c. For each year
2. What is the ratio for the source of tweets?
3. What is the most popular dog name?
4. What is the most popular dog breed?
5. What is the ratio of dog stages?
6. What is the most retweeted tweet?
   a. Yearly
7. Is the retweet counts increasing over time?
8. Most favorite dog

I will next share the insights and visualizations I extracted by analyzing the data using the above questions. Since the tweeter account and my analysis all revolve around the rating system, let's look at the histogram distribution of rating points out of 10. Because "they're good dogs, Brent, " most dogs are rated either 10 or more out of 10.

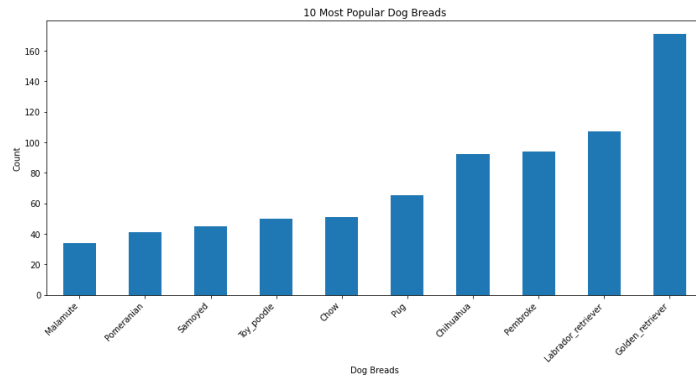Rating Histogram Distribution

The average rate value for all the data points is 11; for data points with more value than 10 is 12, and for data points with a smaller rate value than 10 is 7. After calculating the average for each year, I came up with the following graph:
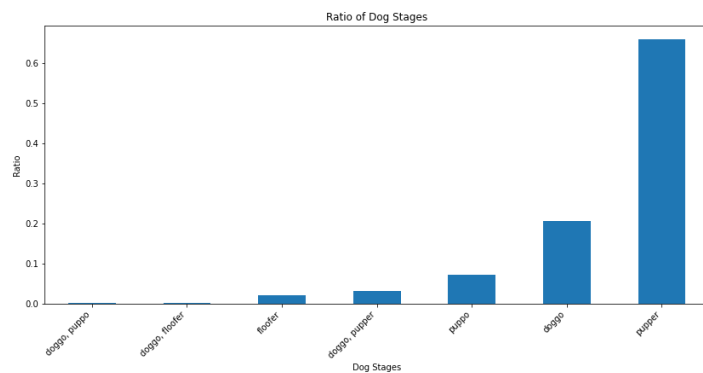

Mean Rating Over the Years

Twitter for iPhone controlled the first position by having almost 96% of the tweets tweeted from it; next is Twitter Web Client, and finally, Tweet Deck.

Five dog names came first with a value of 10 each when finding the most popular dog names; the names are Cooper, Tucker, Oliver, Penny, and Charlie.
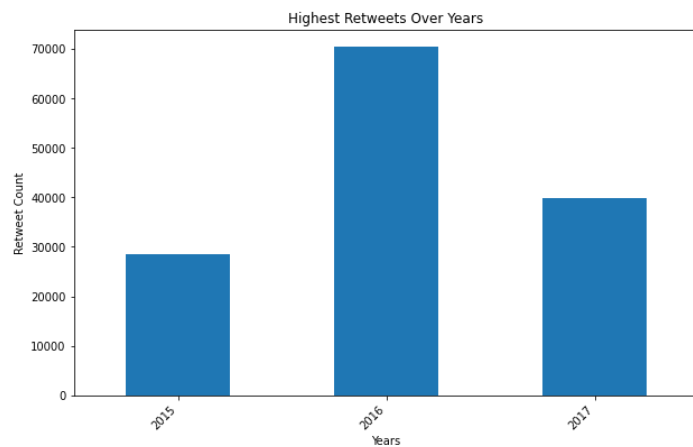
The most frequent dog breed is the Golden Retriever. The graph for the 10 most frequent dog breeds looks like this:
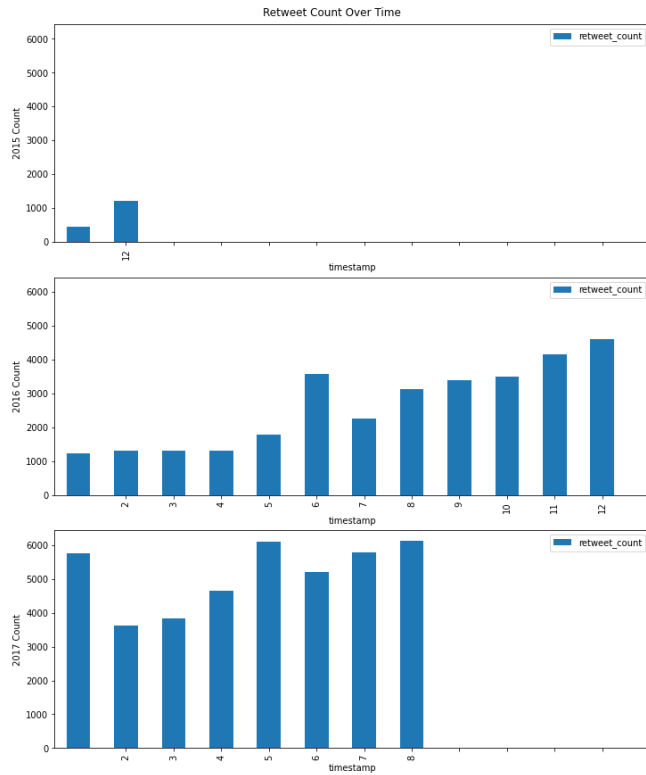
10 Most Popular Dog Breads

The graph looks like the following when visualizing the ratio of dog stages:

Ratio of Dog Stages

The tweet with tweet id 744234799360020481 is the first and the only to be retweeted 70385 times. The following graph shows the maximum tweet numbers for each year.

Highest Retweets Over Years

To analyze whether the retweet counts are increasing over time, I grouped the data points using year and month as multi-index and calculated the average of retweet counts. Since it is so hard to visualize such a wide insight, I categorized the data points yearly and drew the graph for each with the same scale. The graph looks:

Retweet Count Over Time

And finally, the most favorite dog looks:

Isn't he cute?