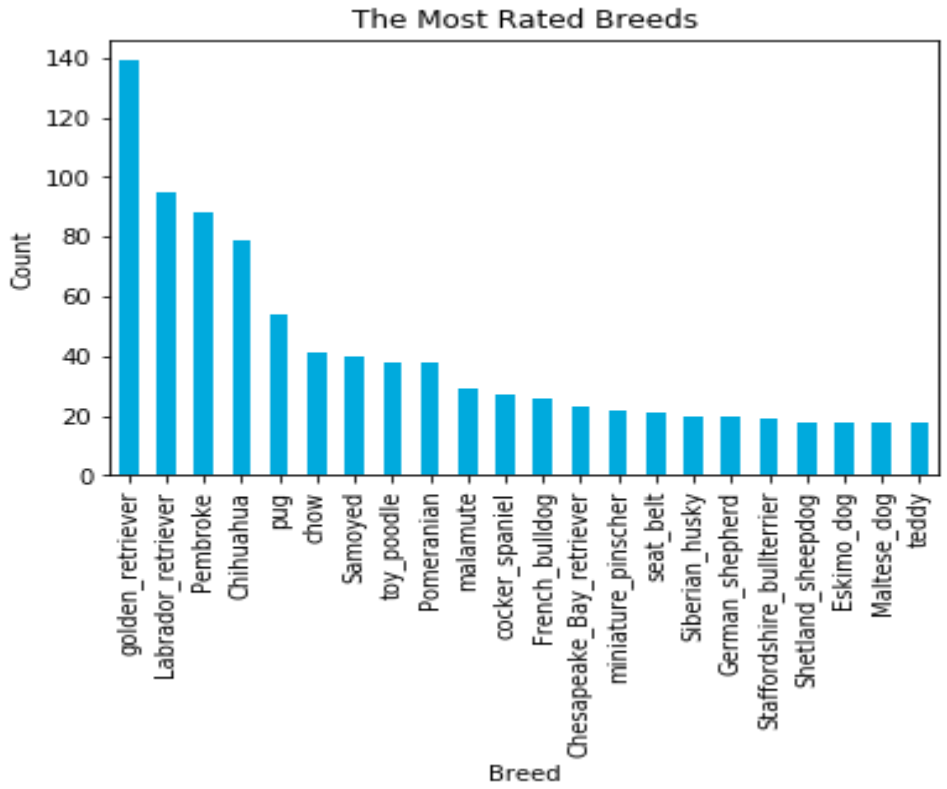
## Analyzing and Visualizing Data

## Introduction

Today we age going to analyze a dataset that contains information of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. WeRateDogs has over 4 million followers and has received international media coverage.


## Breeds

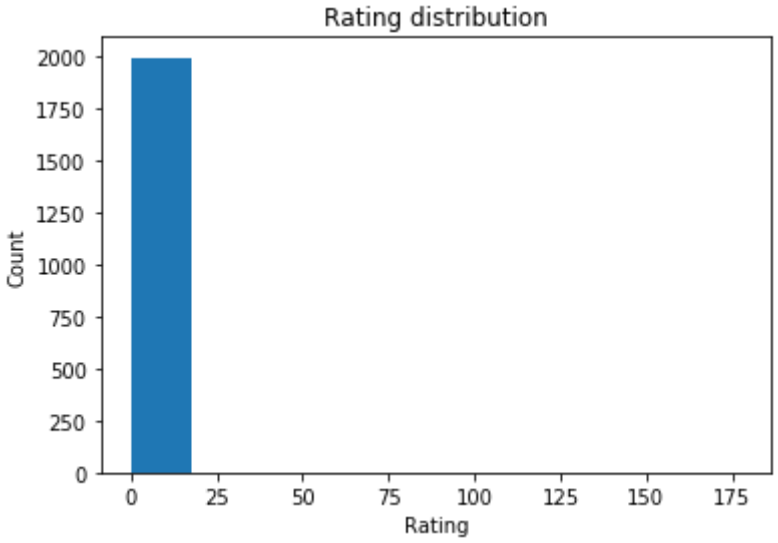
I analyzed which breeds were the most commonly ranked at@dog_rates.



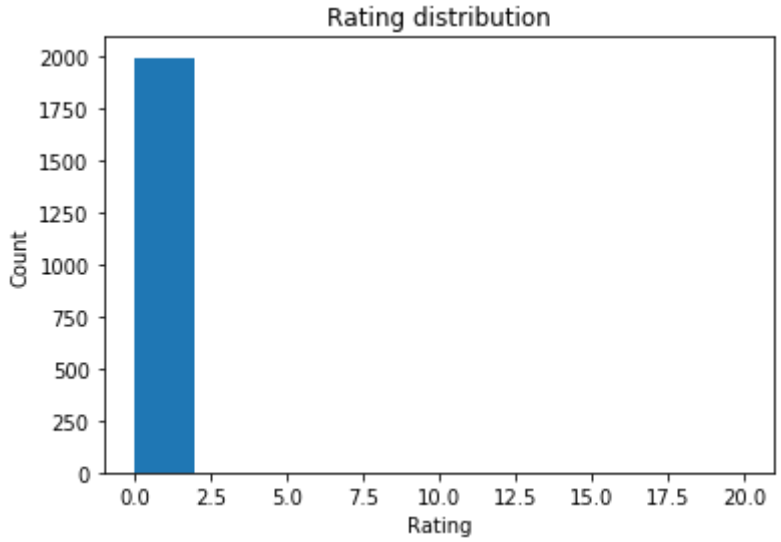

As we can see above, the golden retriever is the most frequently rated breed at WeRateDogs. This breed was rated about 140 times. The next most frequent breeds are Labrador retriever and Pembroke (Welsh Corgi). These breeds were rated about 90 times.

## Analyze ratings

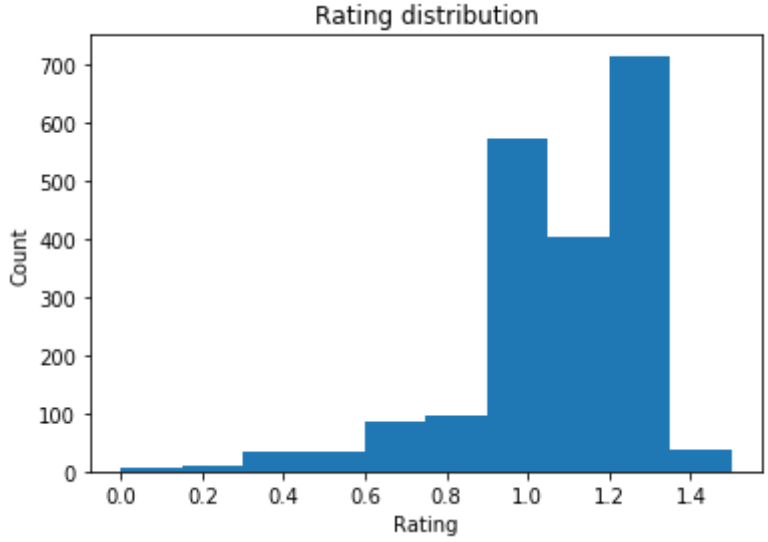
I analyzed the rating distribution:




As we can see above, the rating is generally below 20 and there are only a few rating that higher th, let's zoom in for better visualization.



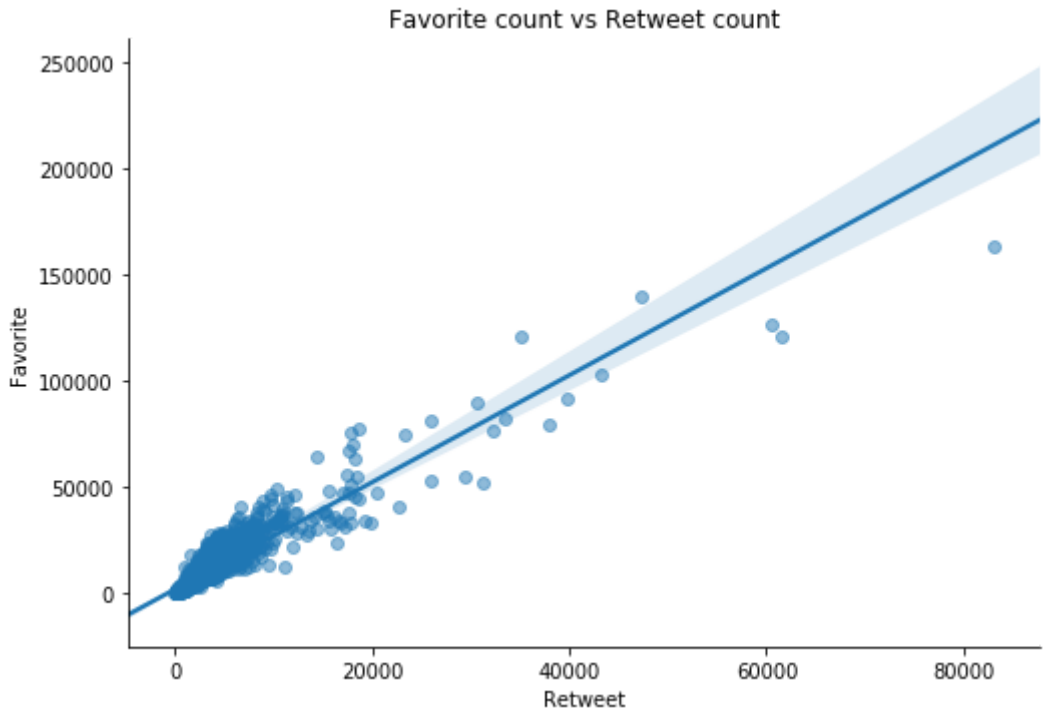

And again, as we can see above, the rating is generally below 2.5 and there are only a few rating that higher than 20. Let's zoom in again:




And finally, we can see on the histogram above, the most frequent rating is from 1.2 to 1.3. It means that the most dogs received rating 12/10 - 13/10.

## Retweet and favorite count

Here I analyze correlation between retweet and favorite count. Let’s plot a scatter plot to show correlation visually:




And let’s find correlation coefficient:

```
0.92959
```

The correlation coefficient between favorite_count and retweet_count is quite high - 0.929598. It means these variables have a strong correlation. We also can see it on the scatter plot above.