Francisco Pineda

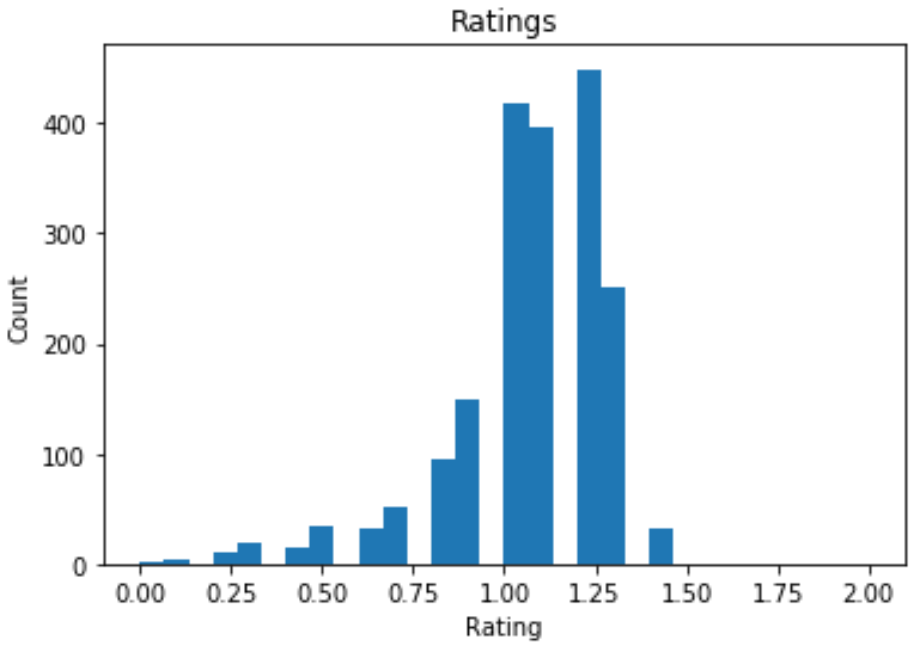
We Rate Dogs Report

Introduction: After cleaning and tidying up the data from the Weratedogs twitter account, some questions I wanted to ask were 'what was the average rating that was given to dogs by the account?', 'what was the relationship between ratings for each tweet?' and 'what was the most common dog name in the dataset?'. I have found insights to each of these questions.

What was the average rating for dogs in this dataset?

The "rating" of each tweet was created by dividing the numerator by the denominator given to each dog in a tweet. The denominator was often 10 but could range as low as 0 and as high as 177. It was common for the numerator to be larger than the denominator.

Below is a histogram of ratings within the dataset.



As you can see the majority of ratings fall between 1.00 and 1.30. The mean rating for the data set was 1.16. This is shown in the table below.

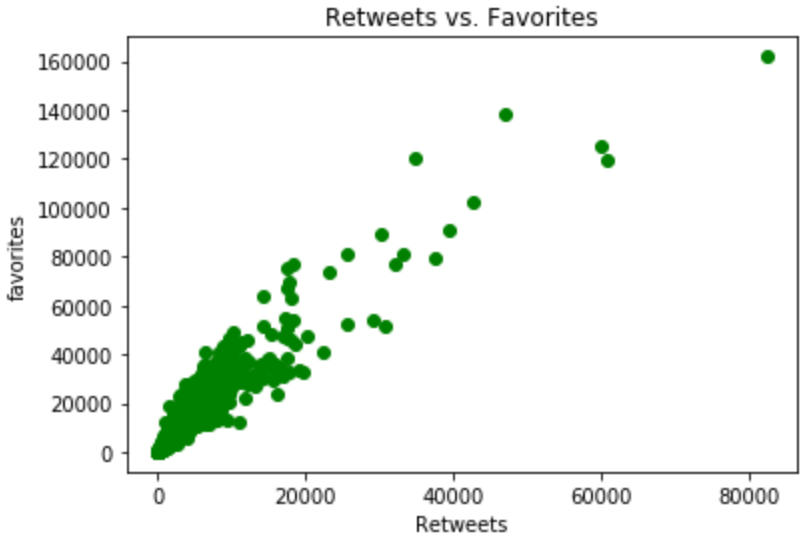
	rating	favorites	retweets
count	1967.000000	1967.000000	1967.000000
mean	1.169437	8630.089476	2616.364514
std	4.094789	12655.378254	4668.048291
min	0.000000	76.000000	11.000000
25%	1.000000	1851.500000	582.500000
50%	1.100000	3902.000000	1259.000000
75%	1.200000	10774.000000	2989.500000
max	177.600000	161887.000000	82260.000000

The standard deviation for ratings was 4.09 which is quite large. This likely because the Max value ratings shown in table below was 177. This outlier is causing the standard deviation to be so large.

The median rating for the dataset was 1.1.

The relationship between retweets and favorites?

Below is a scatterplot showing the relationship between retweets and favorites for each tweet in the dataset.



I found that the correlation coefficient between these two variables as .92. This indicates a strong positive correlation. You can absolutely see this correlation in the scatterplot.

What are the top names within the dataset?

As you can see by the histogram below there is a four way tie between the names Charlie, Lucy, Cooper and Oliver for the most common name in the dataset. Each one of these names has ten occurrences within the dataset. The next two most common names are Tucker and Penny with nine occurrences each.