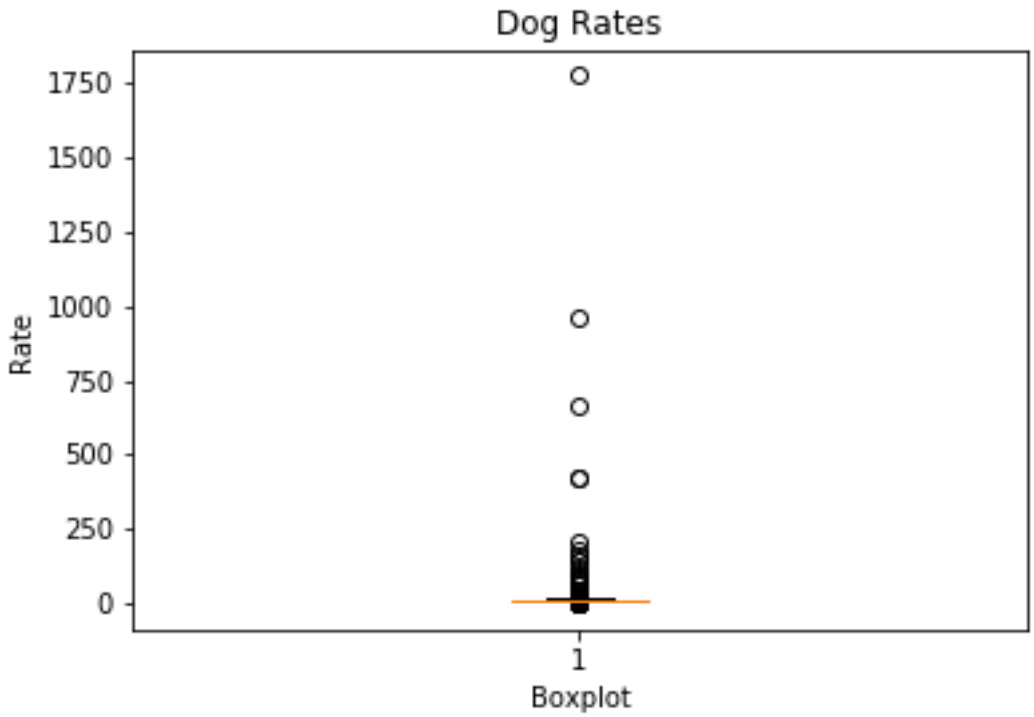
External Document

Analyze

Outlier dog rates



	rating_numerator	rating_denominator	retweet_count	favorite_count
count	2175.000000	2175.000000	2175.000000	2175.000000
mean	13.221149	10.492874	2724.261609	8732.824368
standard deviation	47.725112	7.019084	4686.940457	12430.526513
min	0.000000	0.000000	0.000000	51.000000
25%	10.000000	10.000000	593.000000	1875.500000
50%	11.000000	10.000000	1311.000000	3967.000000
75%	12.000000	10.000000	3136.000000	10925.500000
max	1776.000000	170.000000	77746.000000	143985.000000

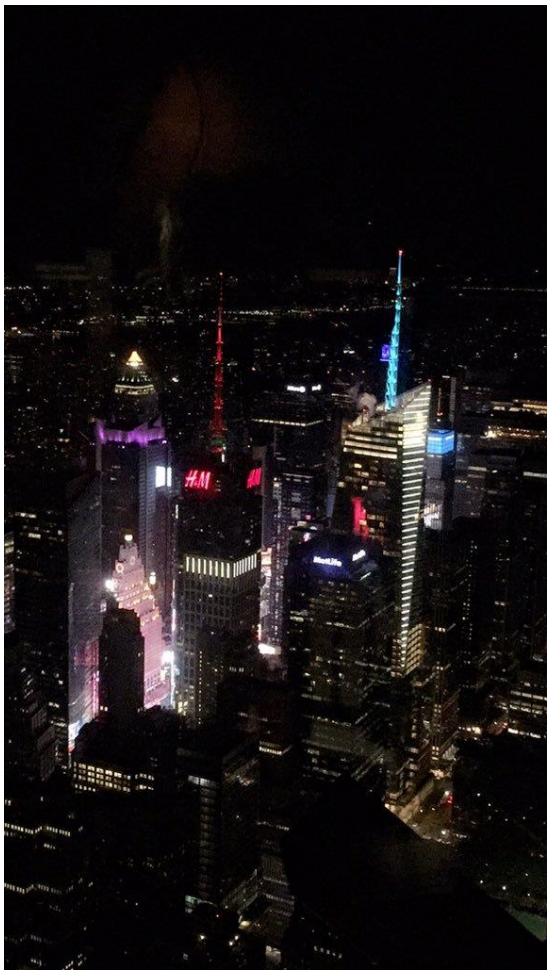
Most of the rates are between 10 and 12. There are some rates that are very far from the vast majority of the data. At least 50% have rates are in the [10,12] interval. And there are some strange rates, like 1,776. And what kind of picture gets a zero rate?

Rate : 1776

This is Atticus. He's quite simply America af.

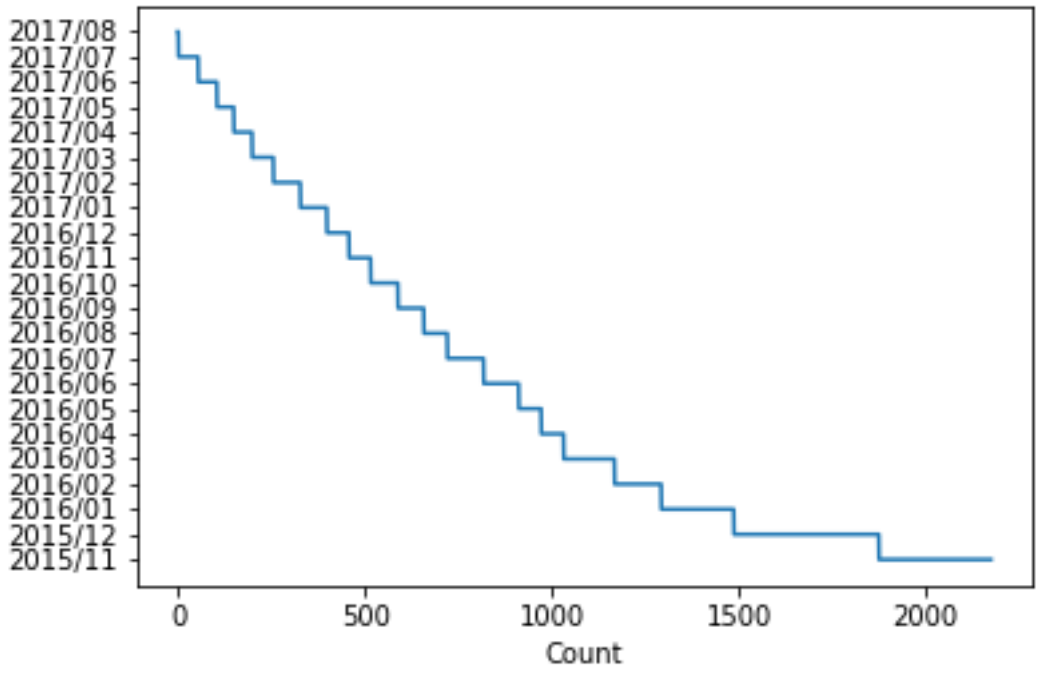
Looks like this dog has that much points because he shows what a big American he is. This is an American tweet site, so this is why they love this picture so much.

Rate : 0



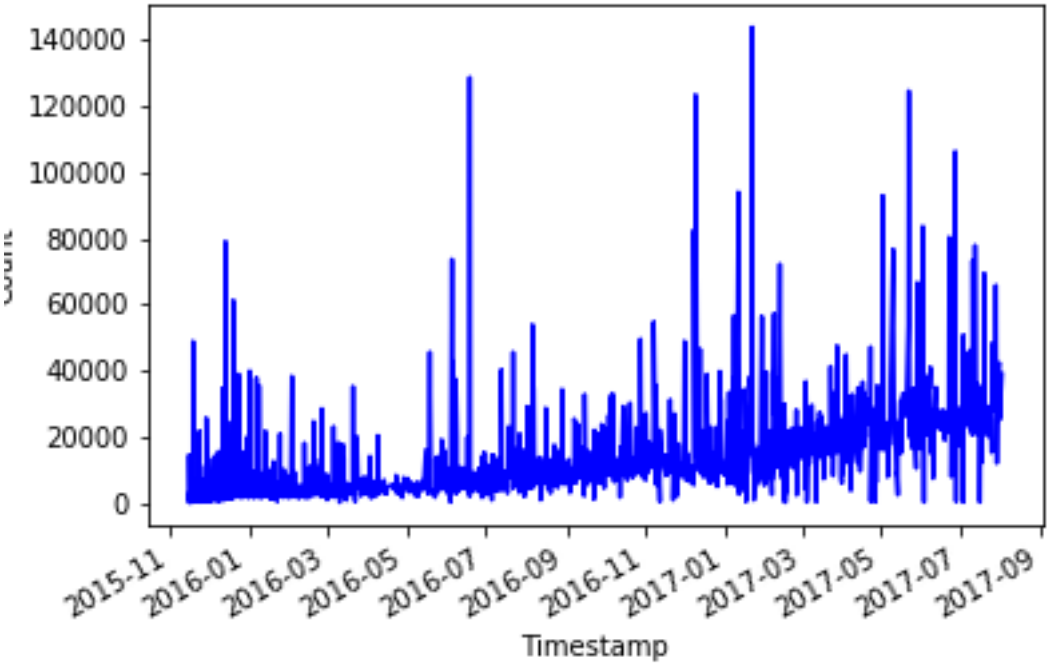
There have to be some mistakes with the rate of the first picture. It has been predicted as a “swing”, but it is a dog. Somehow they accidentally rated this cute dog as 0. The second picture has no place on this tweet page, it is understandable why it has this rate.

Tweet Count



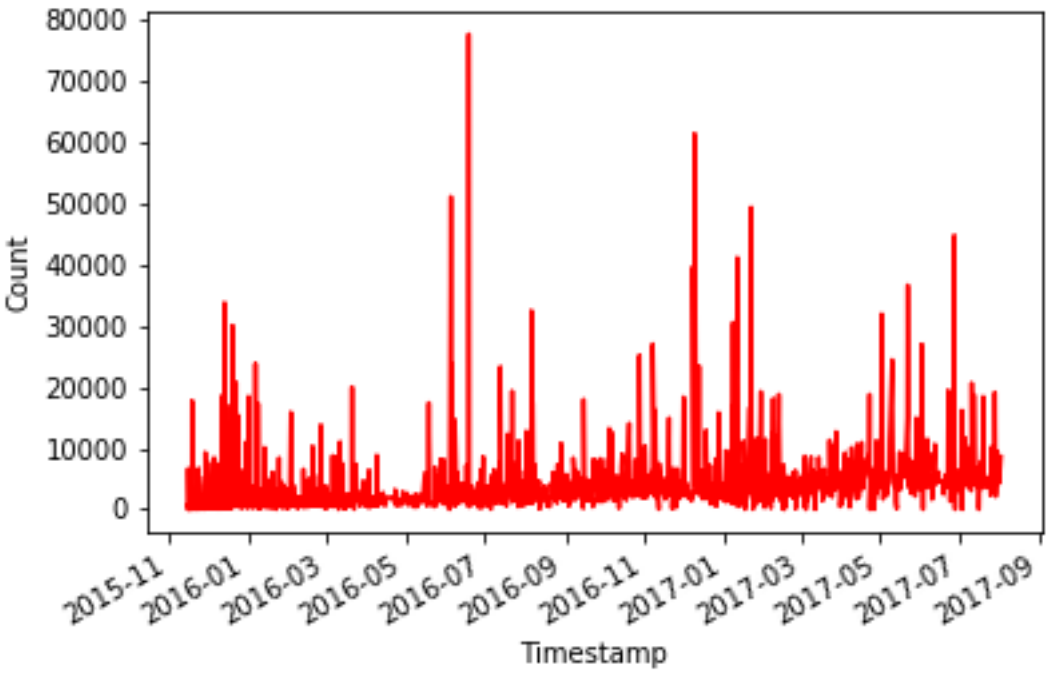
Tweet count is decreasing from 2015/11 to 2017/08. In the end of 2015 there were 2,000 tweets in one month. In 2017/08 there were around 0. Tweet counts are monotonically decreasing.

Likes



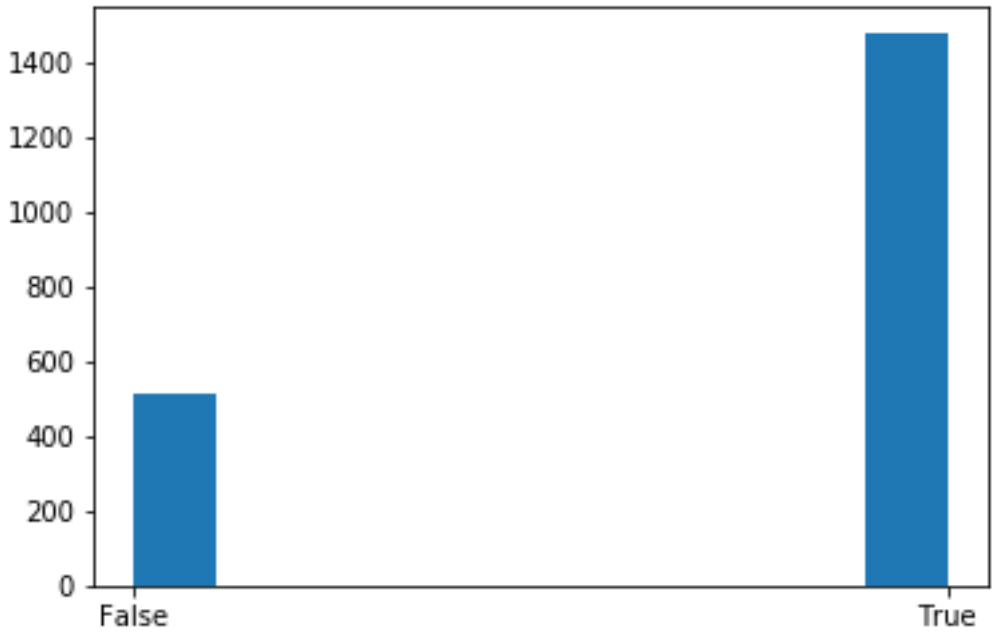
Tweets are decreasing but likes are growing, so I can say that the audience is growing too. It does not mean that there are more likes because they tweet less. I think there would be more likes if there would be more tweets.

Retweets



There is a growing tendency in retweets too. The big breaks at the Y axis maybe a little misleading, because it seems like there is just a little growing. If we examine this tendency in the knowledge of this, we can see that at the beginning the minimum retweet counts are around 0, and at the end of the chart it is around 2-3,000. It is a huge difference.

Is the tweet about a dog?



This chart shows us what is the first prediction about the picture. If it is a dog, it is true. If not, it is False. The y axis is the count.

Around 75% of the pictures are not about dogs in the eye of the machine learning software.