Having access to data from over 2000 tweets from WeRateDogs twitter account I wanted to see if an algorithm designed to recognize dog breed from a photo attached to the tweet was able to correctly name the dog breed. The algorithm took a photo present in a tweet and gave 3 predictions, information if the prediction is actually a dog breed and the confidence on that guess.
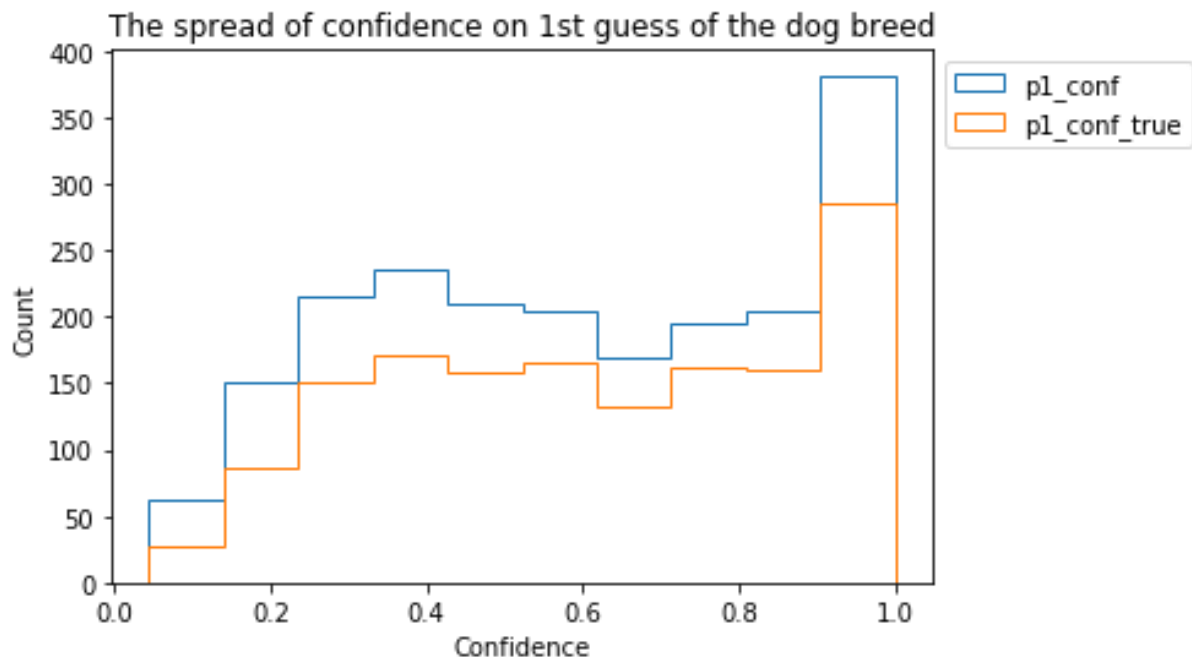
The histogram represents the count of confidence on 1st guesses for all dog tweets. It represents both all of the guesses and the guesses where the answer it gave was actually a dog breed (not a food type or a flower). Using describe function I was able to look into it with more detail – over 500 of the 2028 guesses didn't return a dog breed – that's over 25% of the guesses. As a follower of WeRateDogs I wanted to defend the algorithm – some of the photos are hilarious and represent dogs dressed up as hot dogs or covered in food – hence wrong guesses. Removing wrong guesses doesn't seem to improve the score – the mean goes up by only 0.02.

Another part of the tweets I wanted to investigate was the retweet count and favourite count in the function of time. I created this graph:
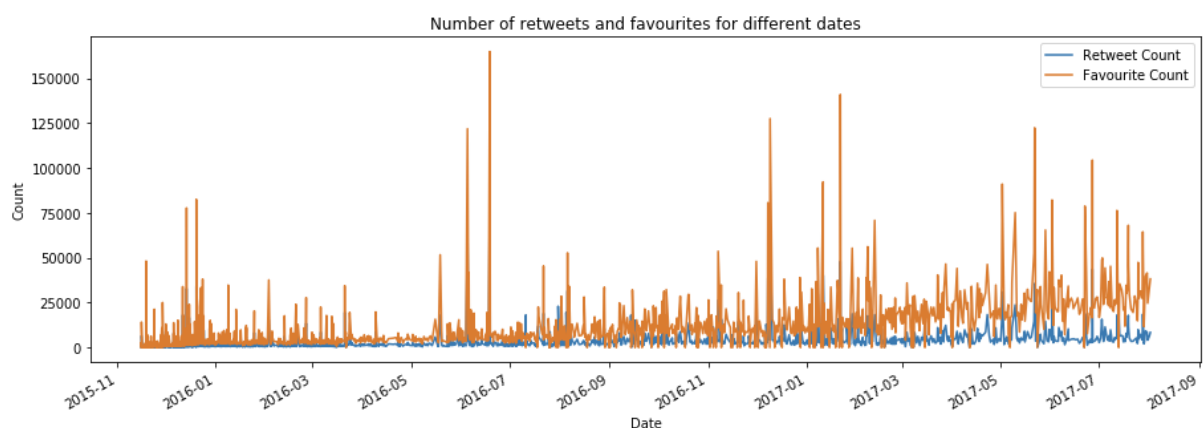
It shows that with time the average number of both retweets and favourites increase – the account gets more popular and receives more online traffic. There are dates where the favourite and retweet count are really impressive. Looking into 5 most popular tweets when it comes to retweets and

favourites we discover that 4 tweets are listed in both of the rankings. They also all have a rating of 13/10 which is quite good, considering that most dogs receive rating of 11 or 12/10. There was, however, a dog that got way higher rating of 1776:

This photo was posted on the 4th of July which is the Independence Day in United States of America. Considering the fact that American nation is always very happy and proud on that day, I'm not surprised that seeing a dog in such a costume got the twitter author excited enough to give him such rating.

The third, and final part, of my analysis was checking if the tweets with most likes were guessed correctly by the algorithm. Unfortunately, only the first and top rated dog was a right guess – it was Labrador retriever. Rest of the guesses had a very low confidence rate. One of the dogs was guessed correctly but the photo represented 2 dogs and French bulldog it had guessed wasn't the dog the tweet was about. It seems that the algorithm has a problem with the focus and prioritizes the high confidence guesses over lower confidence guesses on dog breeds. This way it will recognize an apple on the photo with a high confidence over a chihuahua with a lower certainty.