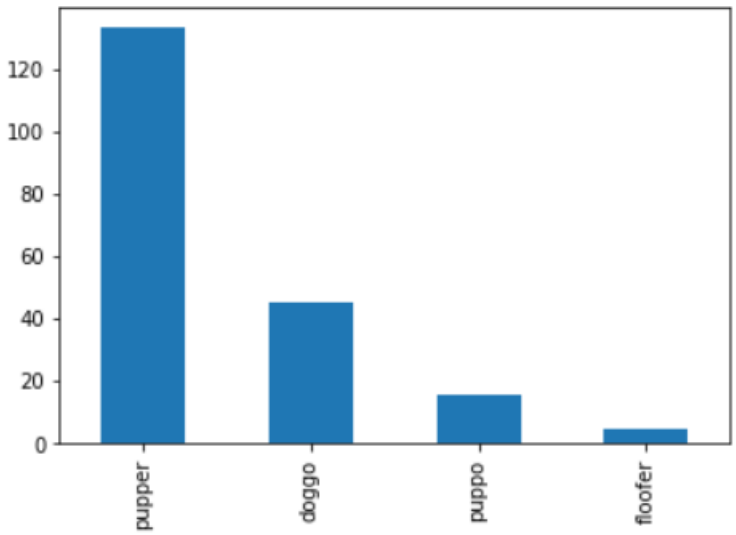
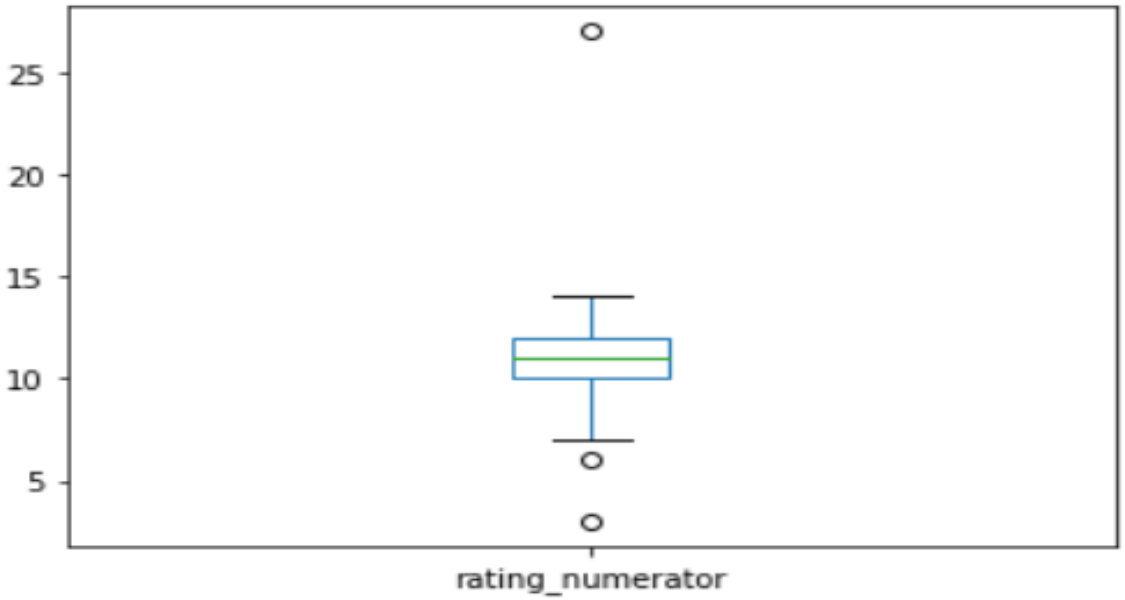
# Act Report

The dataset that we will be wrangling (and analyzing and visualizing) is the tweet archive 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. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10

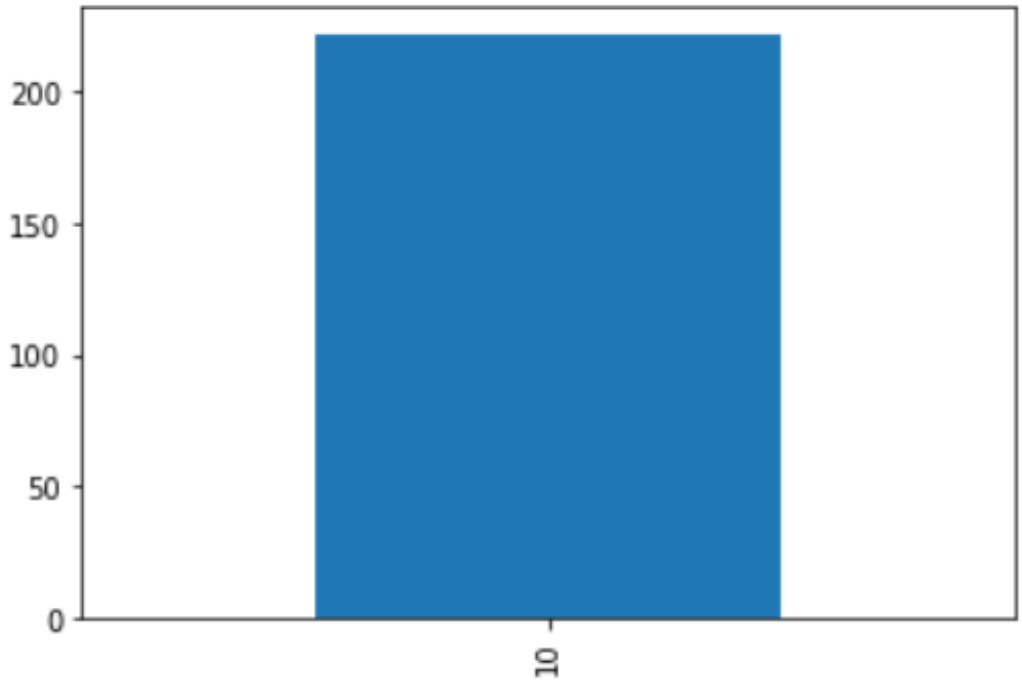
- After making the stages of the dogs in one column
  doing a bar plot to find which stage of the 4 stages has been repeated
  The most repeated stage : pupper
  The 2nd : doggo
  The 3rd : puppo
  The 4th : floffer



- After making a box plot to the rating_numerator column we find that there is outliers that we have to deal with it

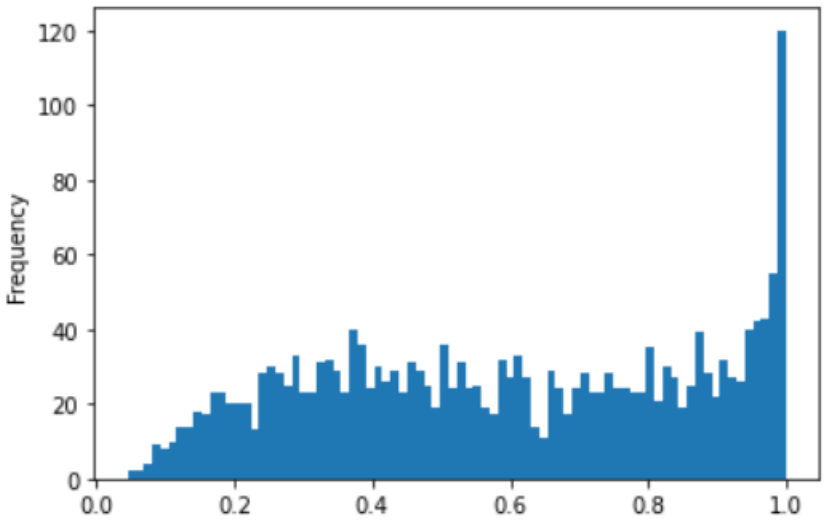



- After cleaning twitter archive enhanced all of the values of the “ rating denominator “ column become 10



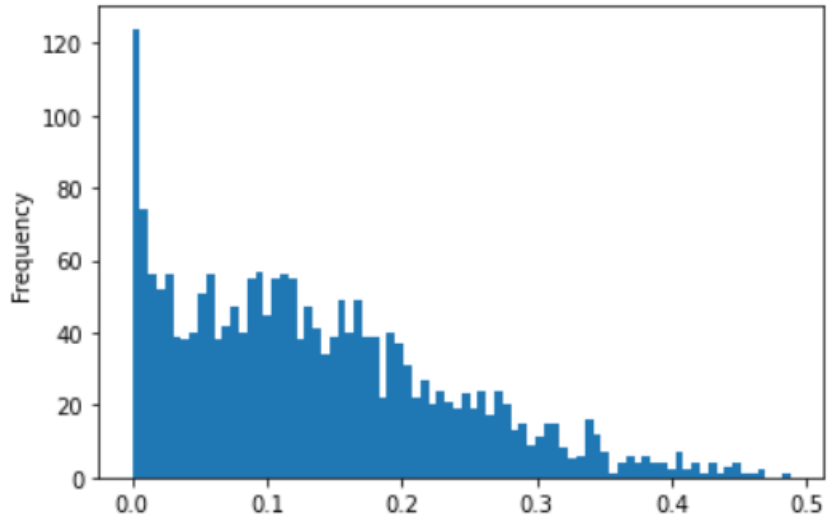
- After making a histogram to “ p1_conf “ column of the image prediction data frame we find that the most frequent values is that close to 1

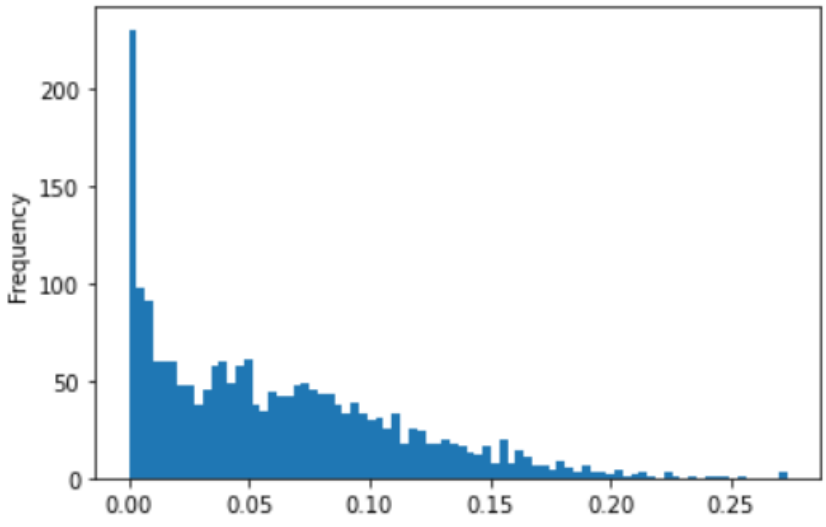
  so we conclude that the 1st prediction of neural network that can classify breed of dogs was accurate



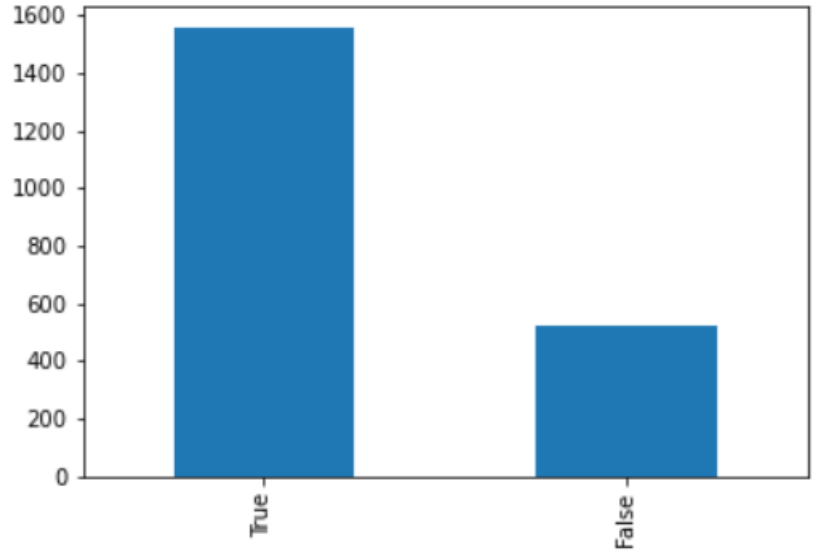
- After making a histogram to " p2_conf " and "p3_conf" columns of the image prediction data frame we find that the most frequent values is that close to 0

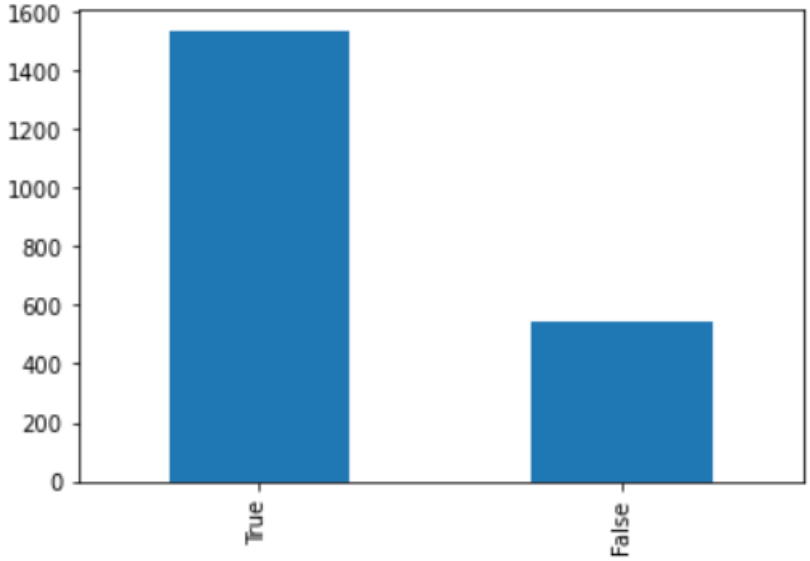
so we conclude that the 2nd and 3rd predictions of neural network that can classify breed of dogs was not accurate

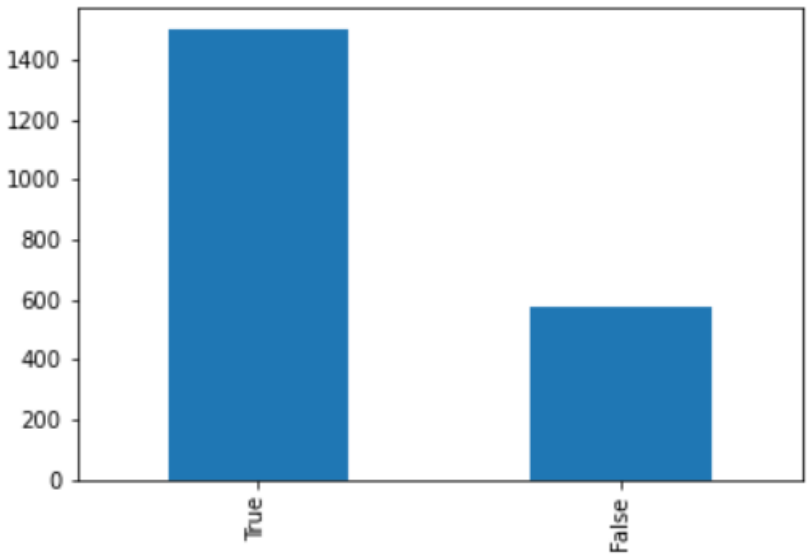


- After making a bar plot to “p1_dog” , “p2_dog” and “p3_dog” we find that the true prediction of the neural network in the neural network more than the false prediction

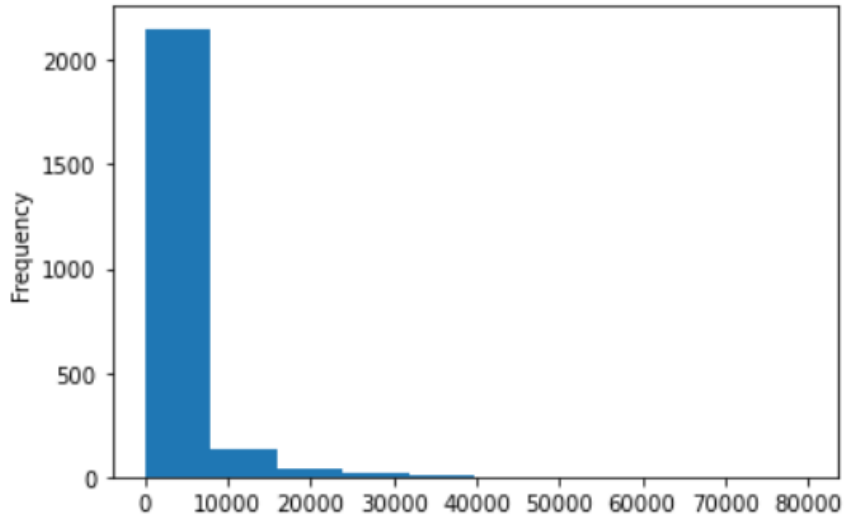


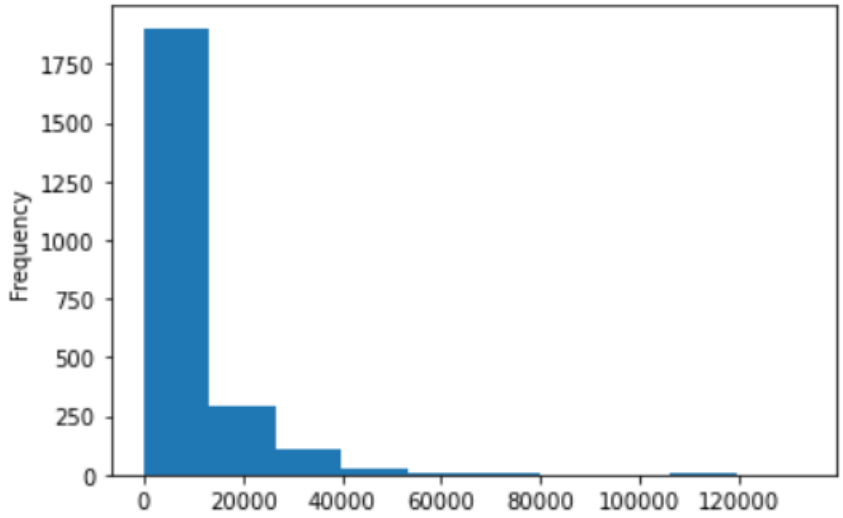






so we conclude from the last 2 visualization that the neural network can predict if it is breed of dog and predict correctly the most likely prediction of the dog in the first prediction only

- Making a histogram to the retweet count and value count to find the frequencies



- Making a box plot to the retweet count and value count to find if these numbers are rational or not and finding outliers