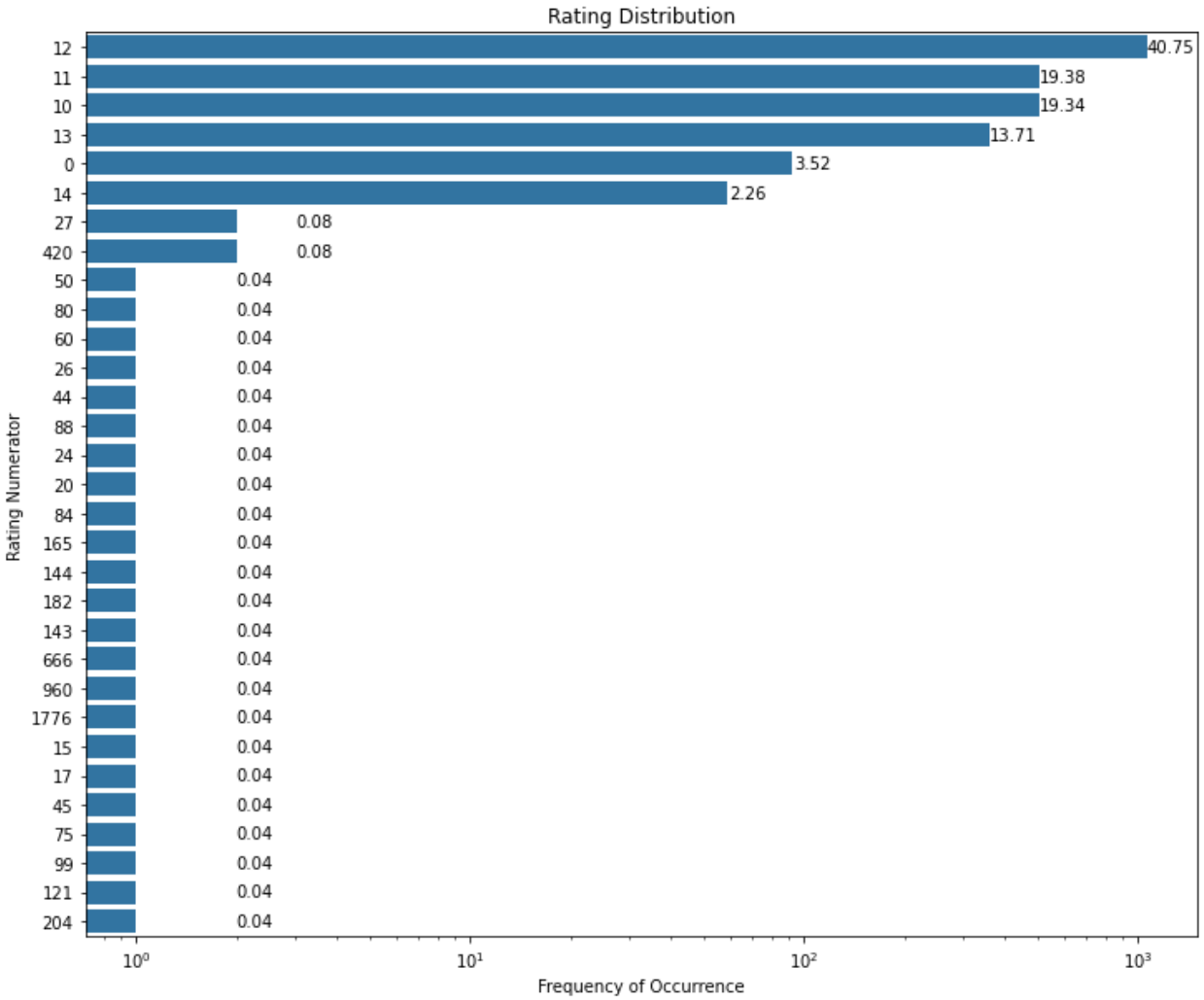
INSIGHTS FROM VISUALIZATIONS MADE FOR WERATEDOGS TWITTER ARCHIVE

WeRateDogs rate people's dogs with a humorous comment about the dog. The rating comes with a denominator of 10 and the numerator is always greater than 10. The project investigated three datasets. One is the WeRateDogs twitter archive, second is the image predictions and third is the data gathered from twitter API (each tweets favorite count and retweet count).

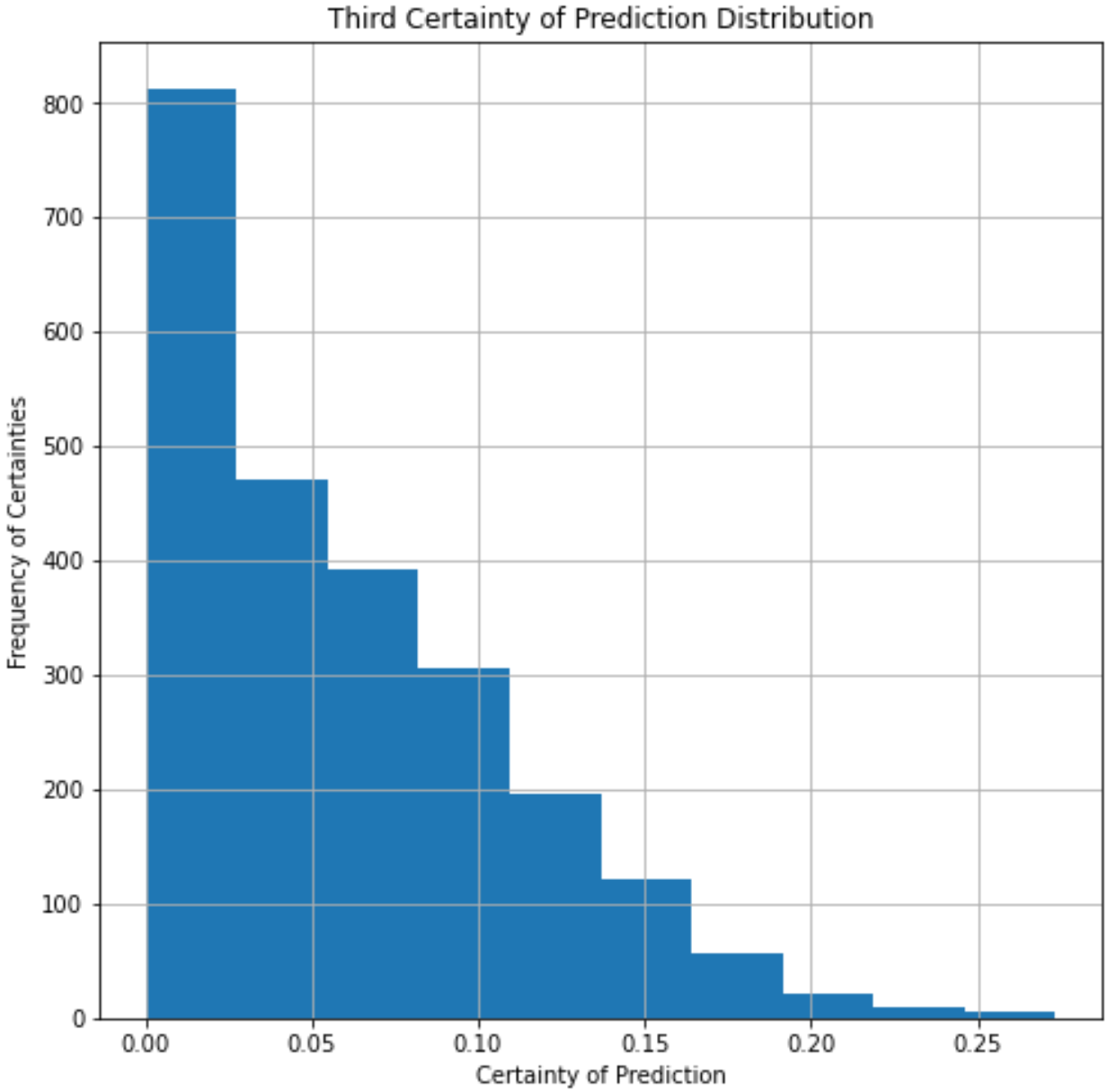
The second and third data were used to augment the twitter archive into a master dataframe and a fourth dataframe was created from it - retweets. The elementary insights drawn from the data are:

1. Rating Numerator has 12 as the modal value.



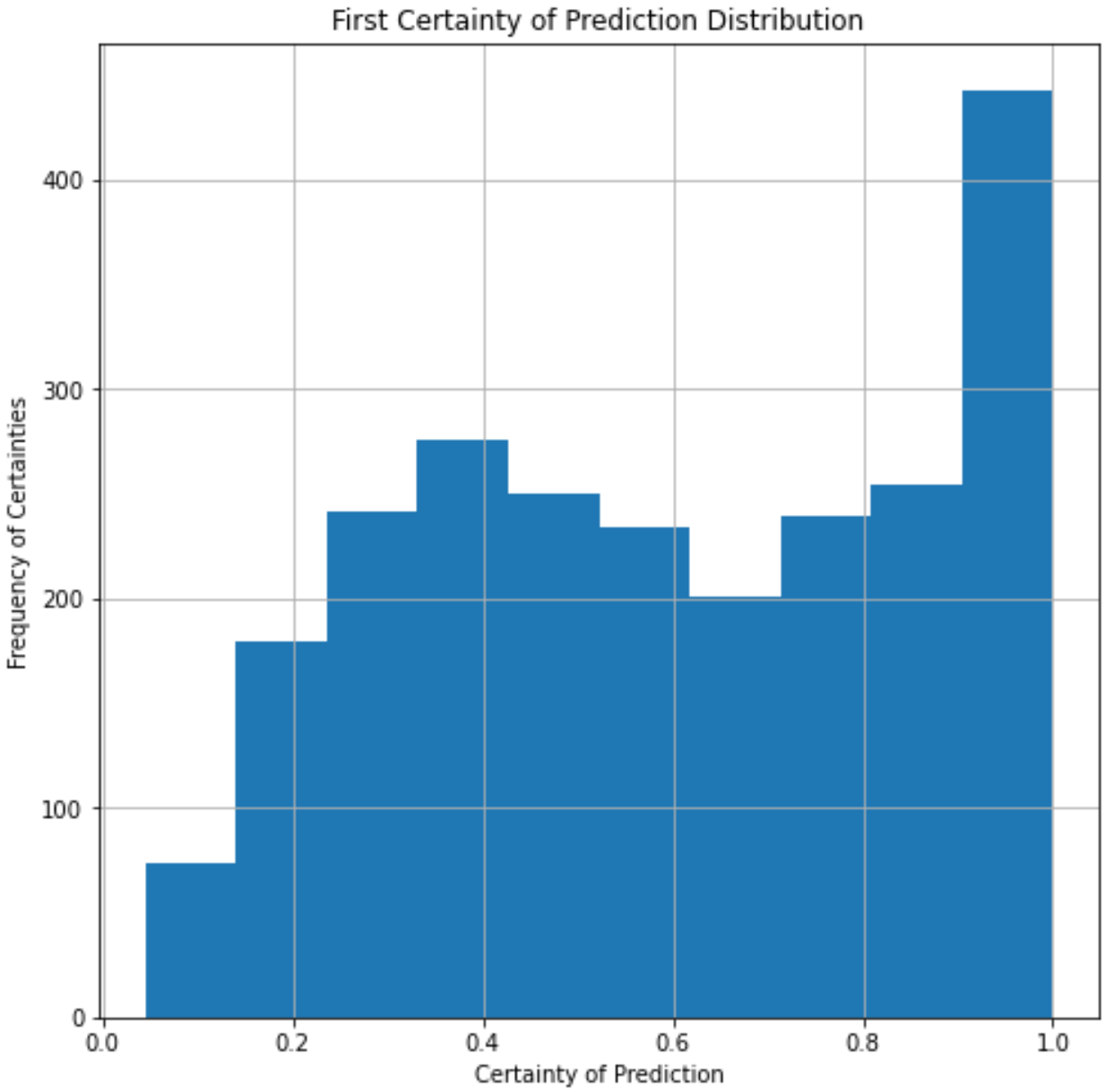
12 has a percentage of 40.75, followed by 11 which is 19.38%, and majority of numerator have a percentage of 0.04.

2. The values of tp3_confidence tends to have a distribution that looks like an exponential distribution.

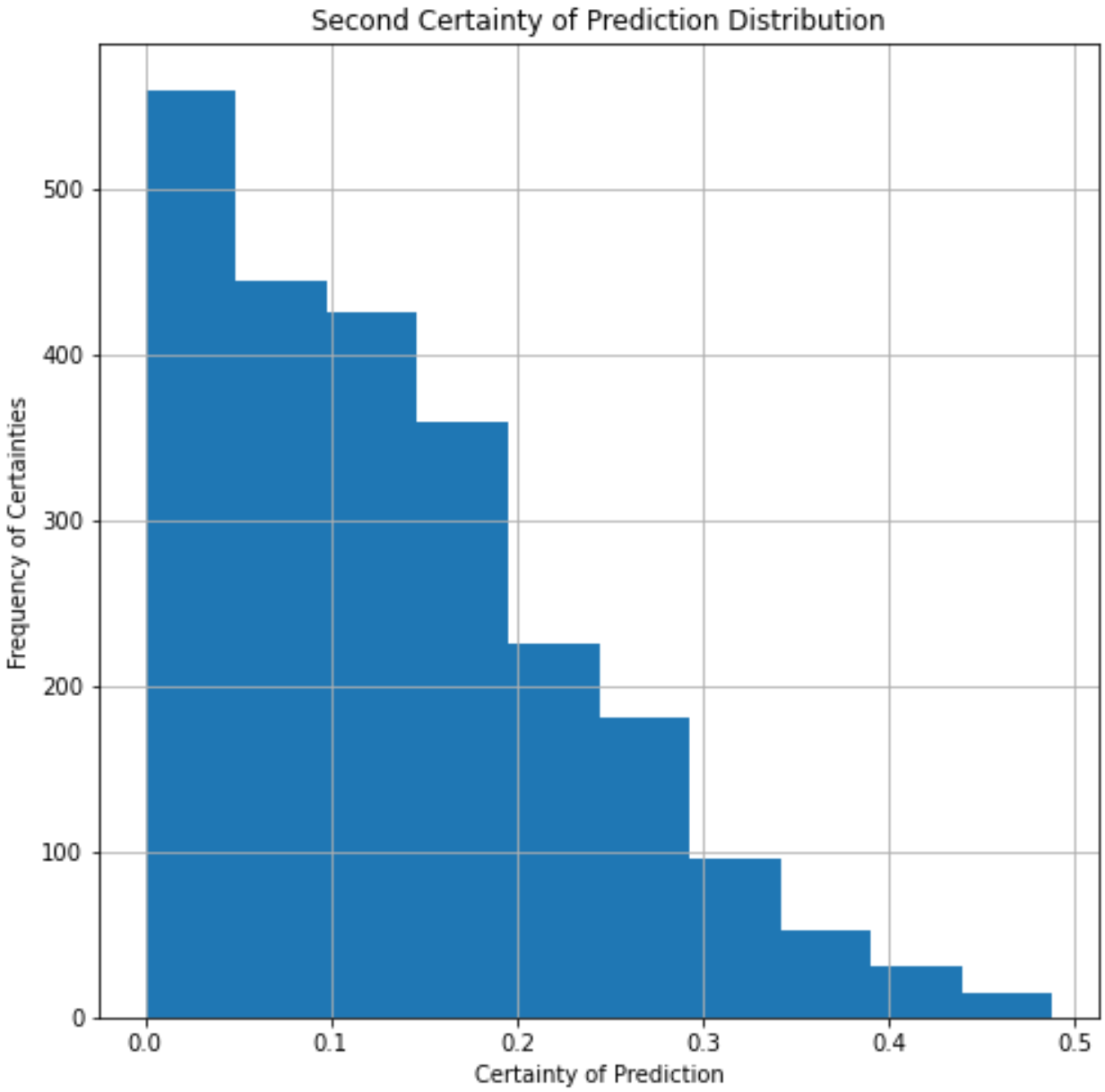


We can presume that the certainty of prediction of whether the image is a dog or not for tp3 is continuous and independent at a constant average rate.

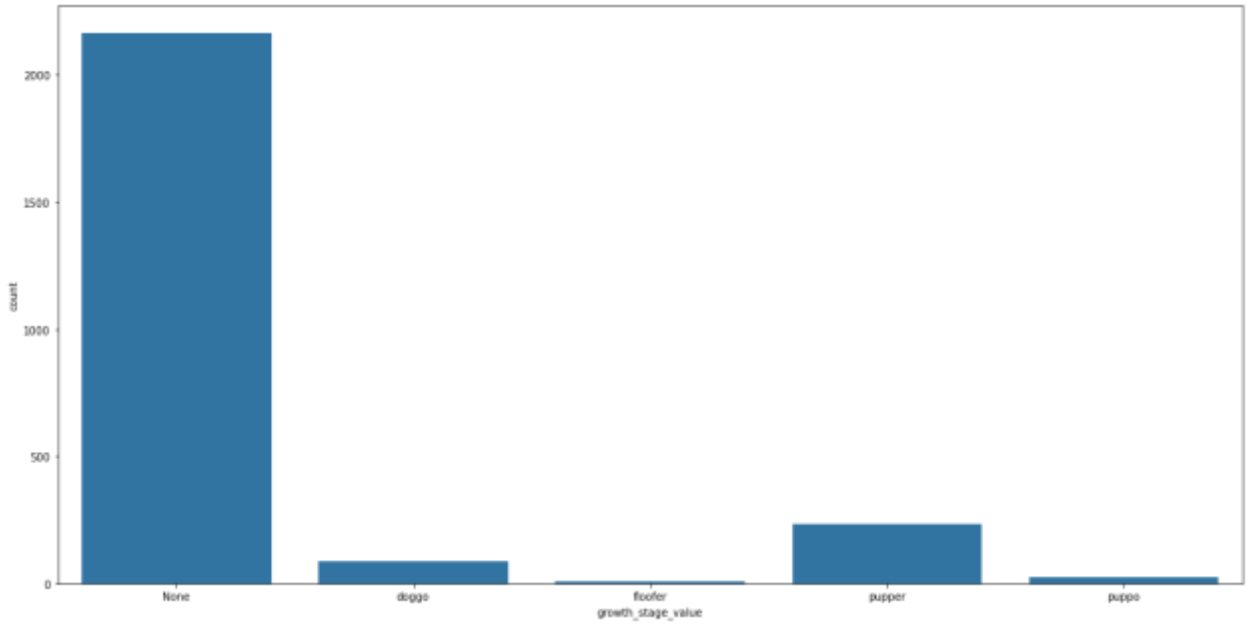
3. The values of tp1_confidence is bimodal. The highest mode being 0.945523.



4. The values of tp2_confidence are seeming exponentially distributed as well, but not perfectly exponentially distributed.

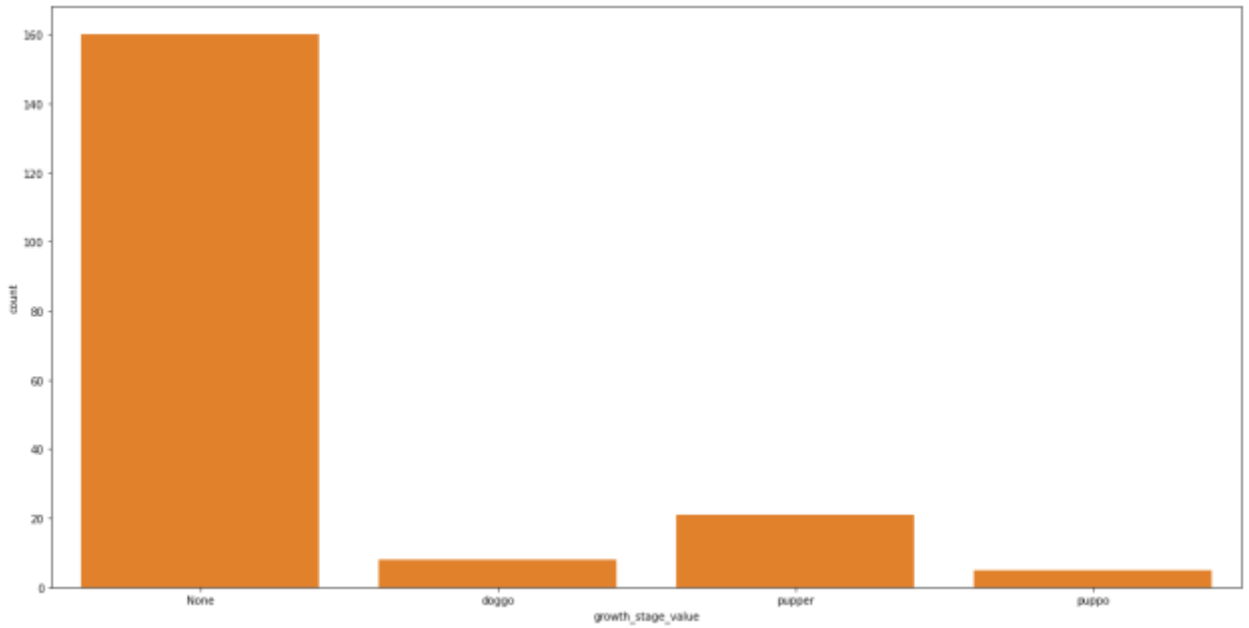


5. Floofer is the least classification of dogs at WeRateDogs archive.



Using the pandas value_counts function, we drew more insight that floofer is just 10, pupper 234, puppo 25 and doggo 86.

6. Many dogs were not classified to be under any growth category. From the above image a whooping sum of 2164 were not classified to any growth stage.
7. No floofer tweets were among the retweets data set.



160 dogs tweets with no growth classification were retweeted, 21 pupper were retweeted, 8 doggo were retweeted, and 5 puppo were retweeted.