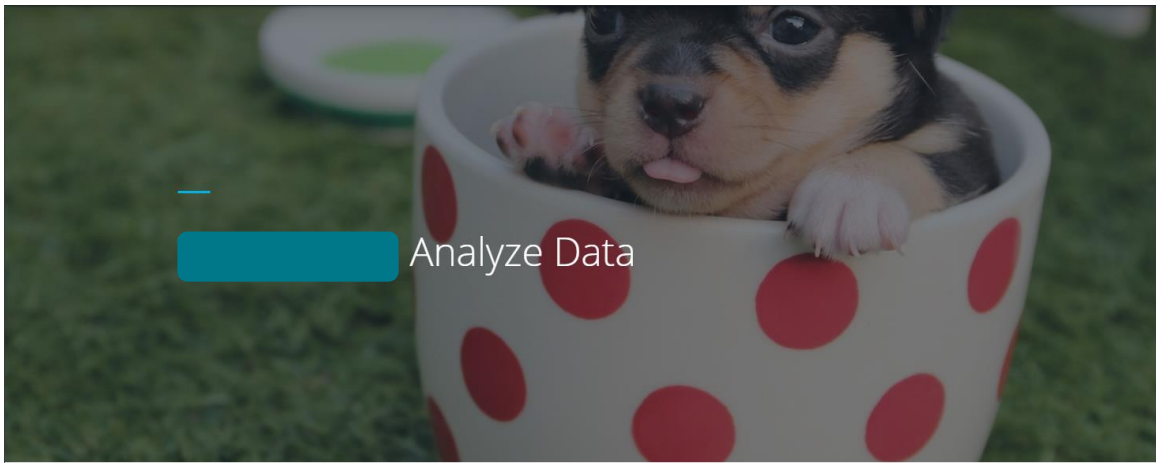


Act report

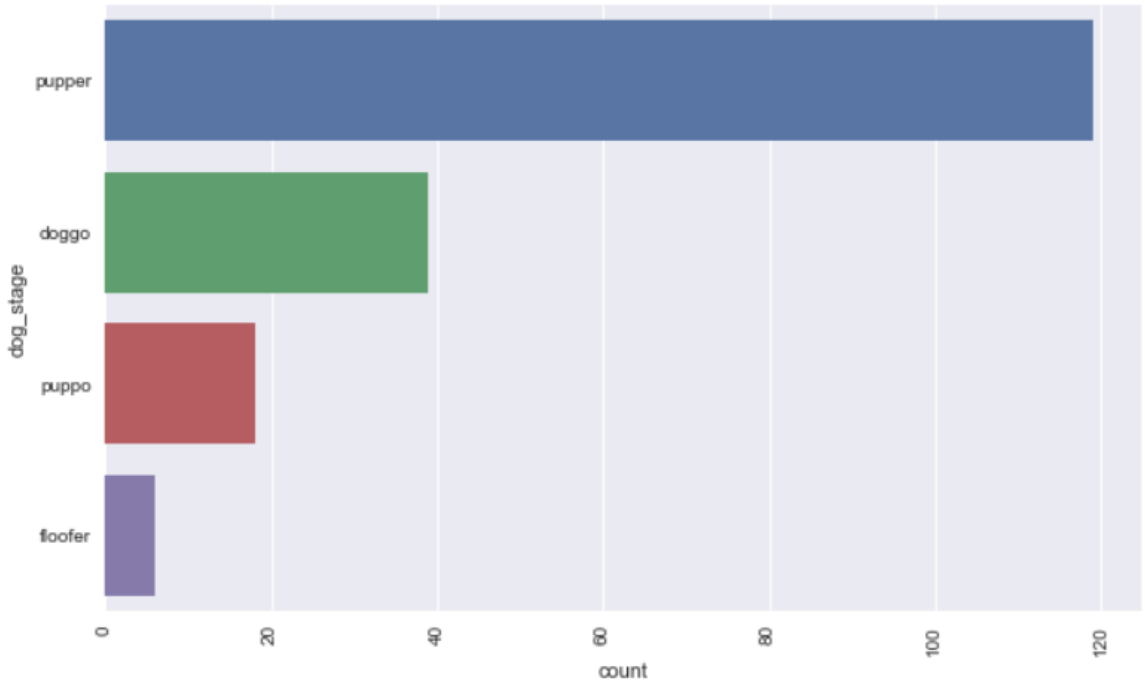
Kristijan Bakaric | Udacity – Analyse Data | 13th of February 2018

Data Analysis and Insights

As a part of the analysis of the The WeRateDogs Twitter archive and Tweet image predictions following variables were explored:

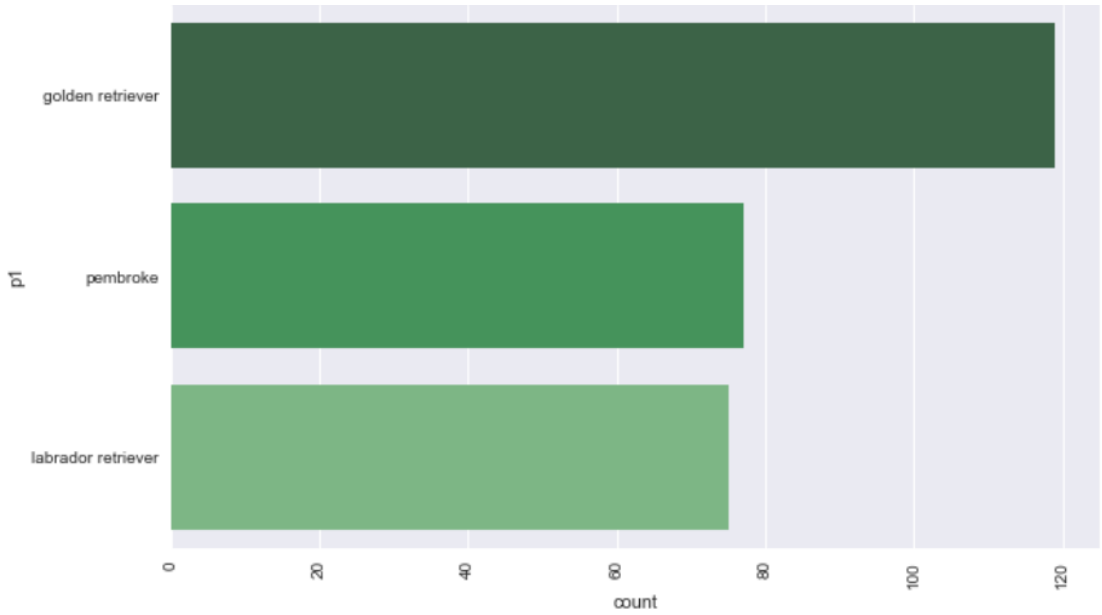
- **dog_stage**: The Dogtionary explains the various stages of dog: doggo, pupper, puppo, and floof(er) (via the #WeRateDogs book on Amazon)
- **p1** : image prediction (only first) of a dog breed from a tweet image as a product of neural network classification
- **name:** name of a dog extracted from the tweet
- **retweet_count:** count of retweets
- **rating_numerator**: ratings of the dogs (The fact that the rating numerators are greater than the denominators does not need to be cleaned. This unique rating system is a big part of the popularity of WeRateDogs)

Count of the Dog Stage



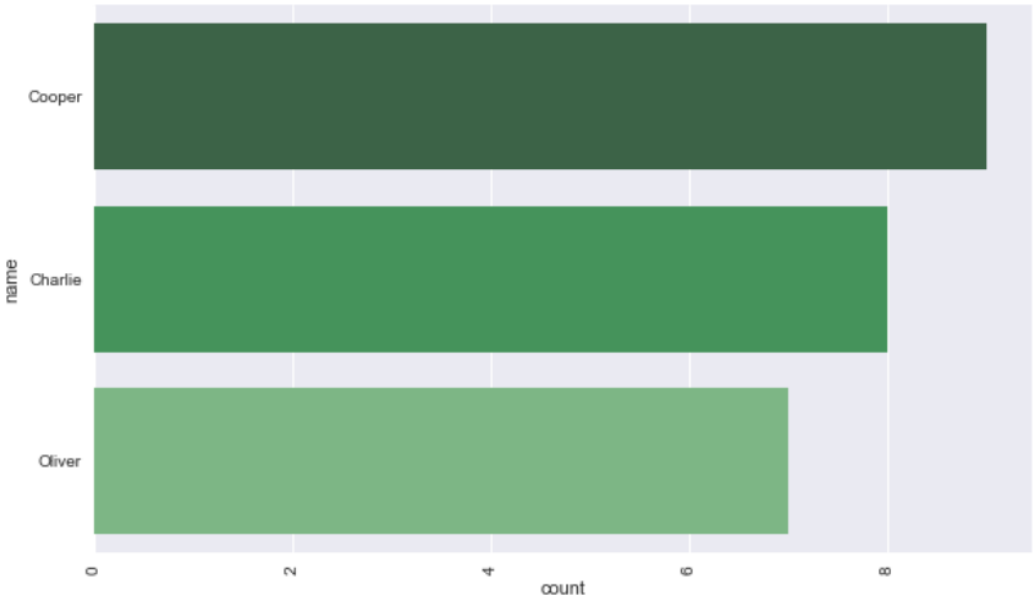
Pupper is most common dog stage amongst the tweets that have a dog stage present, which is followed by a doggo - age before wisdom :)

Count of the p1 Prediction of the Dog Breed from Tweet Images



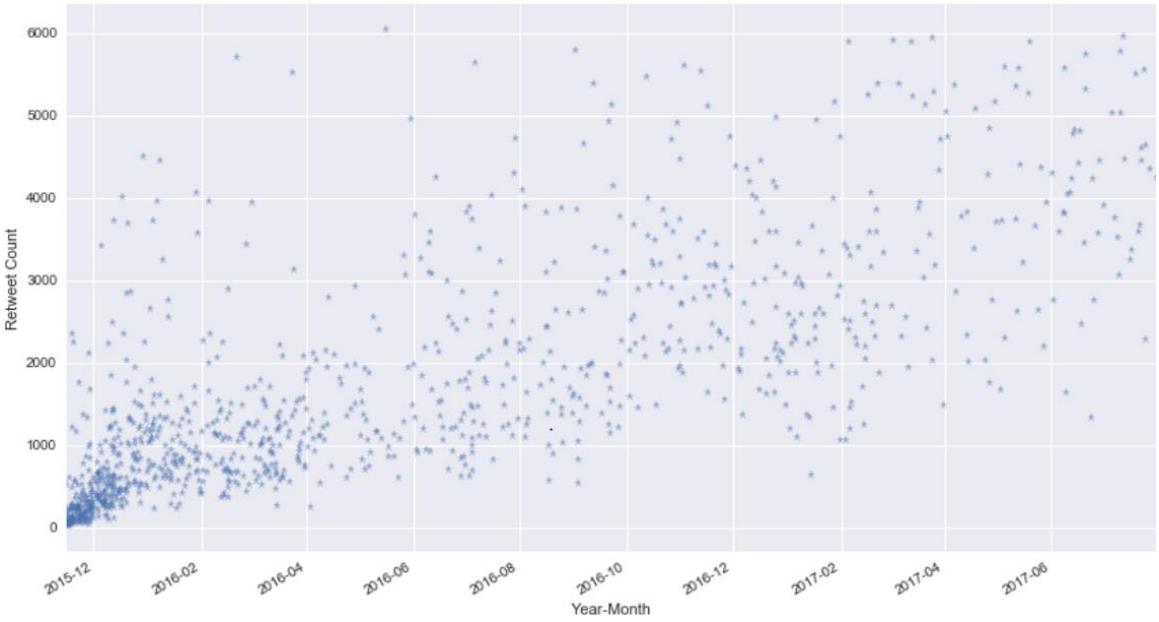
Since every image in the WeRateDogs Twitter archive was ran through a neural network that can classify breeds of dogs, in the chart above golden retriever has the highest count of predicted breed for p1 prediction, followed by pembroke and labrador retriever. Chart is only visualising top 3 breed counts.

Top 3 Names from the Tweets



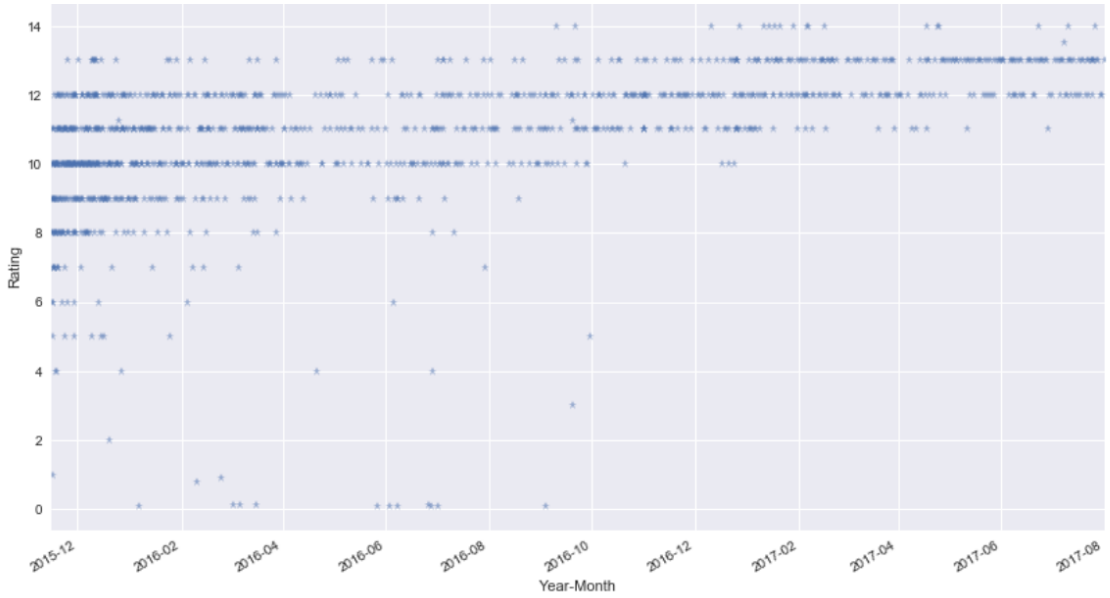
...and the top name has a winner - Cooper :) Chart is visualizing only top 3 names.

Retweet Count over Time



There is a trend of increase of retweets from 2015 to 2017. Data was queried where retwwet count is smaller than 90th percentile (6065).

Rating Numerator over Time



There is a trend of increase of unusual numerator from 2015 to 2017 where in 2017 it is only restricted to 10 and above. Data was queried where rating_numerator < 20.