By Raoul Batcho

August 8, 2022

Introduction

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 however are almost always greater than 10 (11/10, 12/10, 13/10, etc.). In theory the rates should be 1 to 10. However, WeRateDogs disagree, they admit almost all dogs deserve a 10 and sometimes more than that.
WeRateDogs has over 6 million followers and has received international media coverage. One of those coverages were about the quote *"they're good dogs Brent",* this was an exchange in which WeRateDogs shut down a person having an issue with its rating system in humorous ways.

In the framework of this project, we wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering, then assessing and cleaning is required for worthy analyses and visualizations.

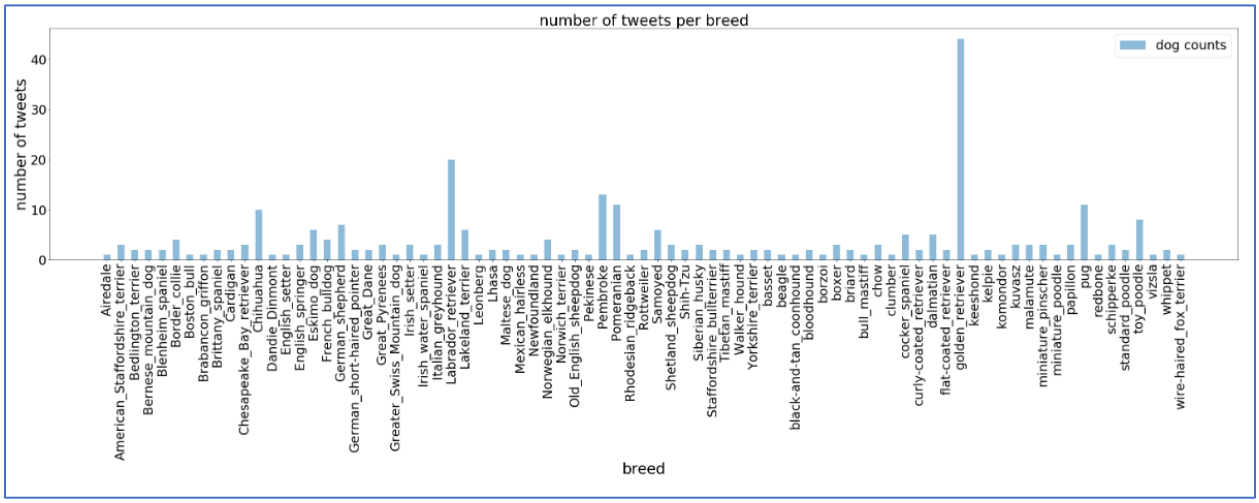
Analyzing and Visualizing WeRateDogs

After having gathered, assessed and cleaned WeRateDogs twitter data I conducted some analysis on the dataset to establish some insights. The final twitter_archives_master dataset has features like tweet_id, source, text, expanded_urls, name, jpg_url, img_num, favorites, user_followers, user_favourites, date_time, day, month, year, rating, stage, breed, and confidence.

WeRateDogs has over 5000+ tweets. I have wrangled 2356 tweets and I finally come up with a with a clean dataset twitter_archive_master of 280 tweets in order to accurately get the following insights.

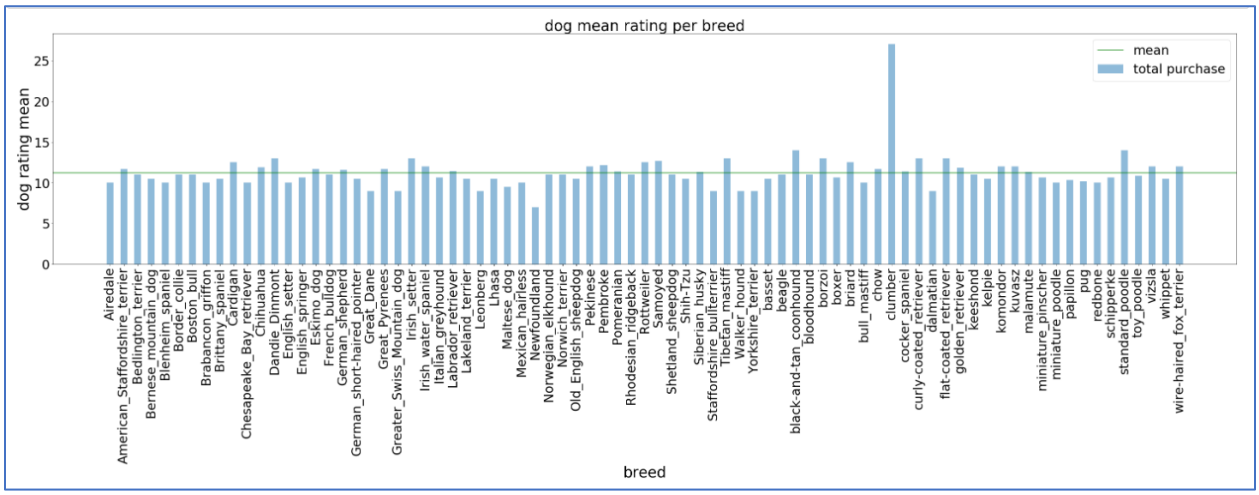
1. Most common dog type

The most common dog type was golden retriever with 44 ratings as shown in the following graph.



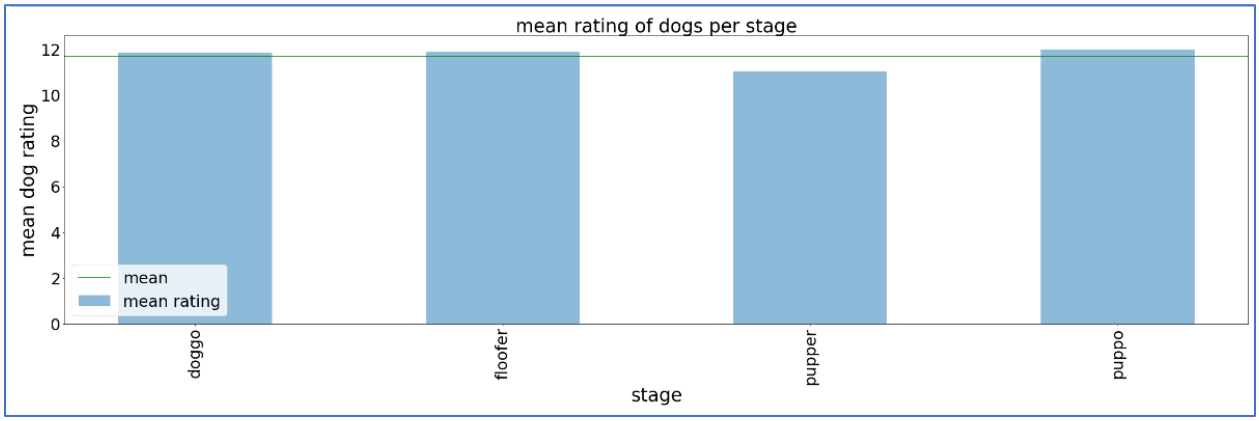
2. The highest and lowest mean of rating per breed

The next graph visualization shows that Clumber is the breed of dog that has got the highest mean of rating according to the mean of rating per breed and the lowest mean of rating is shown for Newfoundland. The overall mean of rating being ~12.



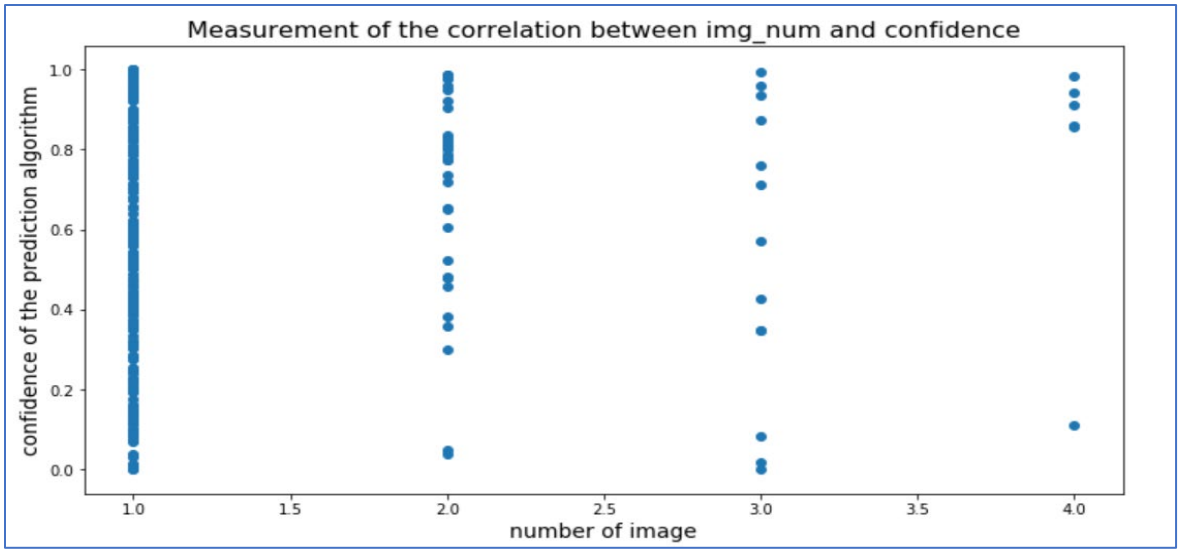
3. Mean rating of dogs per stage

The following graph shows that puppo has got the highest rating followed by floofer and doggo and then pupper that has got the lowest rate.



1. Correlation between the number of image and the confidence of the prediction algorithm.

- The following graph shows that the img_num and confidence are very weakly correlated and the correlation coefficient 0.12 tells so. Hence the number of images doesn't impact how confident the algorithm is in its prediction.
- The graph tells also that most of the tweets have 1 image


Measurement of the correlation between img_num and confidence
confidence of the prediction algorithm
1.0
0.8
0.6
0.4
0.2
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
number of image