Introduction

This report communicates findings from a cleaned dataset from WeRateDogs twitter archive, combined with two datasets; Tweet image predictions - a ML approach of finding the dog breed in the submitted images, and a dataset created via a Twitter API for further metadata gathering.

Insights and visualizations

This section presents three brief insights gained from the acquired data.

Rating ratio of dogs

To easier visualize and deduct an analysis, the dog rating has been calculated and processed to only show the data points which are within 3 standard deviations from the mean. The resulting distribution is easier to interpret.

The distribution of the rating ratio for the tweets are skewed and a majority of the data points are > 1. This indicates that either the dogs are overrated (not likely) or that people really do like dogs (most likely).
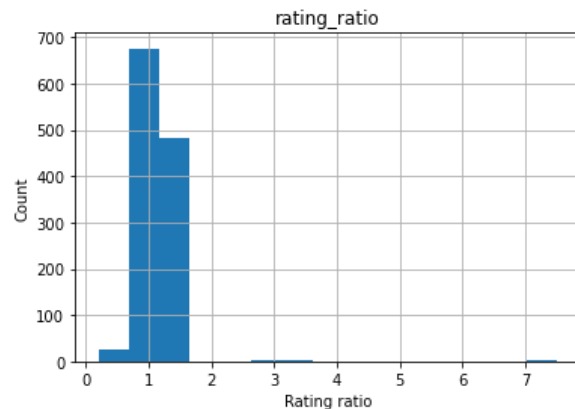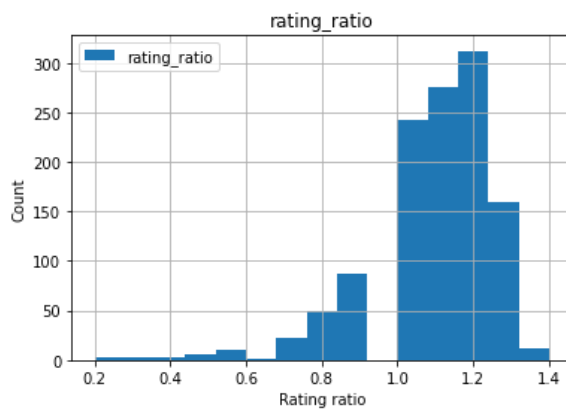
Most popular breed

Using the score given by the twitter user, we can find one of the most popular breed. The Pomeranian was ranked in the top 3 highest score.

Most frequent WeRateDogs user

Counting the sum of times a tweet have been sent either from an Iphone, TweetDeck or Web Client, we can see that the most frequence user is an iPhone user. We can also see that the average between the rating ratio for the categories is higher for TweetDeck, but lower in the retweet count and favorite count:

source	rating_ratio	retweet_count	favorite_count
TweetDeck	1.100000	1778.750000	5038.750000
Twitter Web Client	1.015385	2201.846154	5232.615385
Twitter for iPhone	1.098598	2817.238422	9750.385935

This could indicate that TweetDeck users overate their rating for the dogs, whereas the favorite count and retweet count could be a good indication for what the majority of people favorizes.