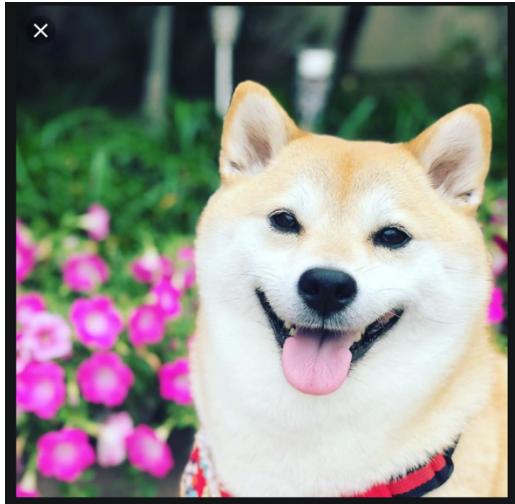Report

Shiba Inus are the best dog. The are my motivation for analyzing this dog dataset.


This study analyzes twitter data from the WeRateDogs account. The goal is to gain insight into tweets from this account – for example which tweets / content are the most popular or shared. Popularity in this case is measured through retweets and likes. An additional goal was to understand which factors were related to tweet popularity. Based on analysis of the clean dataset, the following insights were found :

1. The primary source for twitter is Twitter for iPhone. Most people tweet the WeRateDogs account from their iPhones as opposed to the Web Client, Vine, or TweetDeck. In the final dataset, twitter_archive_master.csv, 1185/1193 of the tweets came from Twitter for Iphone. The remaining came from Web Client and tweetdeck.
2. The average rating for dogs is 11.1/10. Since the average is higher than the total possible number of points, it showcases WeRateDog's unique rating system. This distribution of scores is heavily weighted in the middle. The majority of the ratings are between 10 and 15.
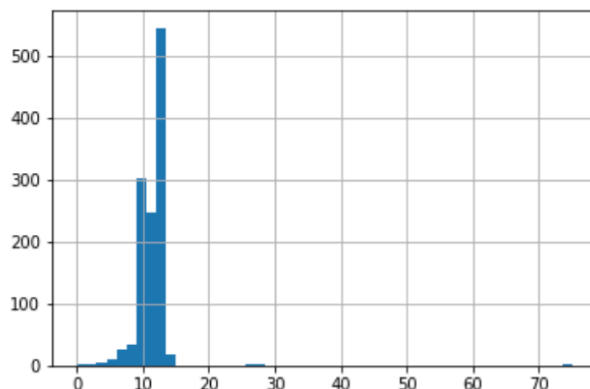
3. Stephen is the most retweeted dog. It is predicted with 50% confidence that he is a chihuahua. The next likely predictions are that he is a Pomeranian or a terrier. It was retweeted 60222 times.
4. The most liked dog was a puppo Lakeland terrier. The second and third predictions is that he is a Labrador retriever or an irish terrier. This photo was liked 138706 times.
5. Based on a linear regression, I found that for each unit change in retweet, likes increase by 2.66 likes. For each unit change in rating numerator, likes increase by 404 likes.
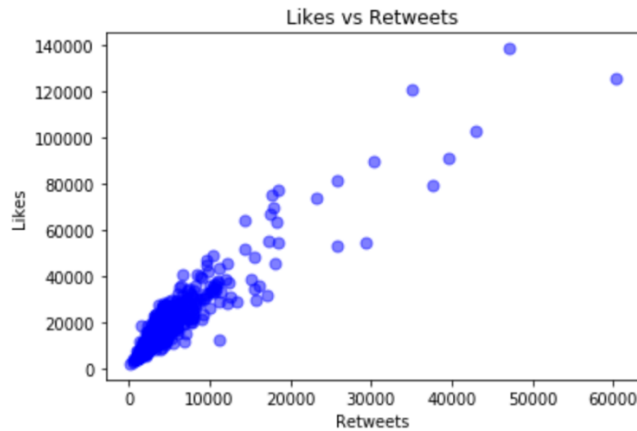
This visualization illustrates the positive association between likes and retweets. This is logical because the more often a message is retweeted, the more it is socialized and the more likely it is to be "liked".