

We Rate Dogs

Deep dive analysis into Twitter Data

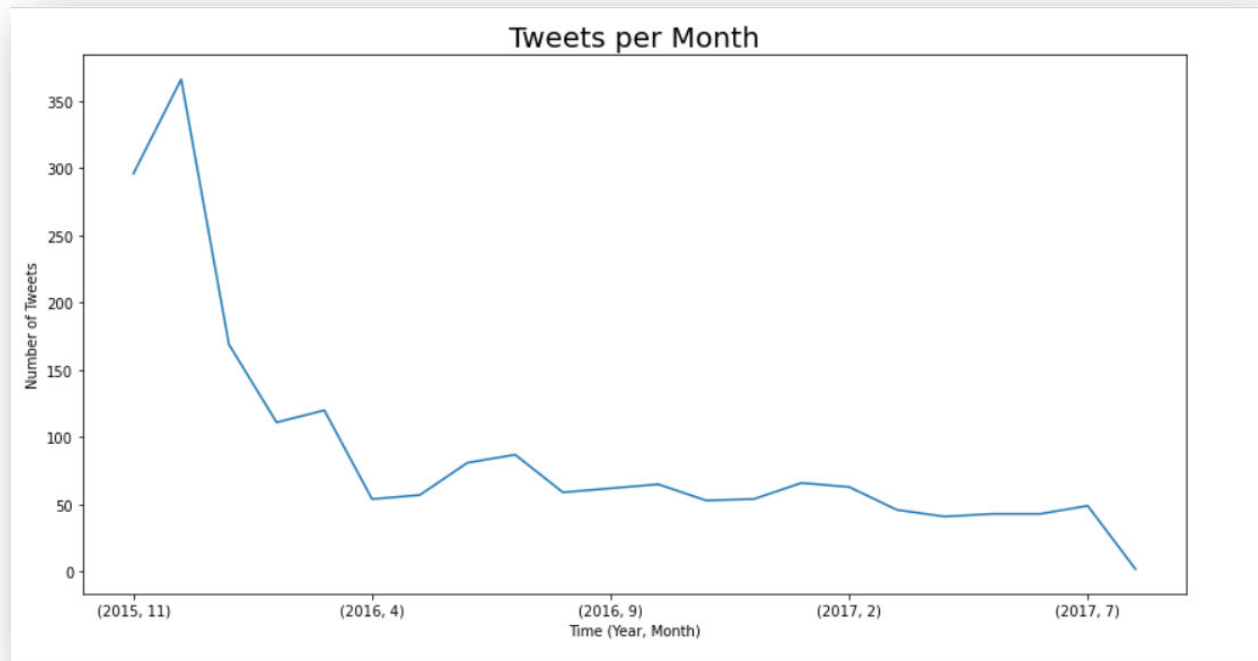
By LaShonda Dickson

WeRateDogs is a public twitter page where people can post pictures of their dogs and the audience can rate the pictures on a point scale from 0 to 10 (10 being the highest), although there were many people who would rate pictures above 10 causing the data to be skewed slightly. This is the purpose of defining, accessing and cleaning the data in effort to achieve the most reliable results.

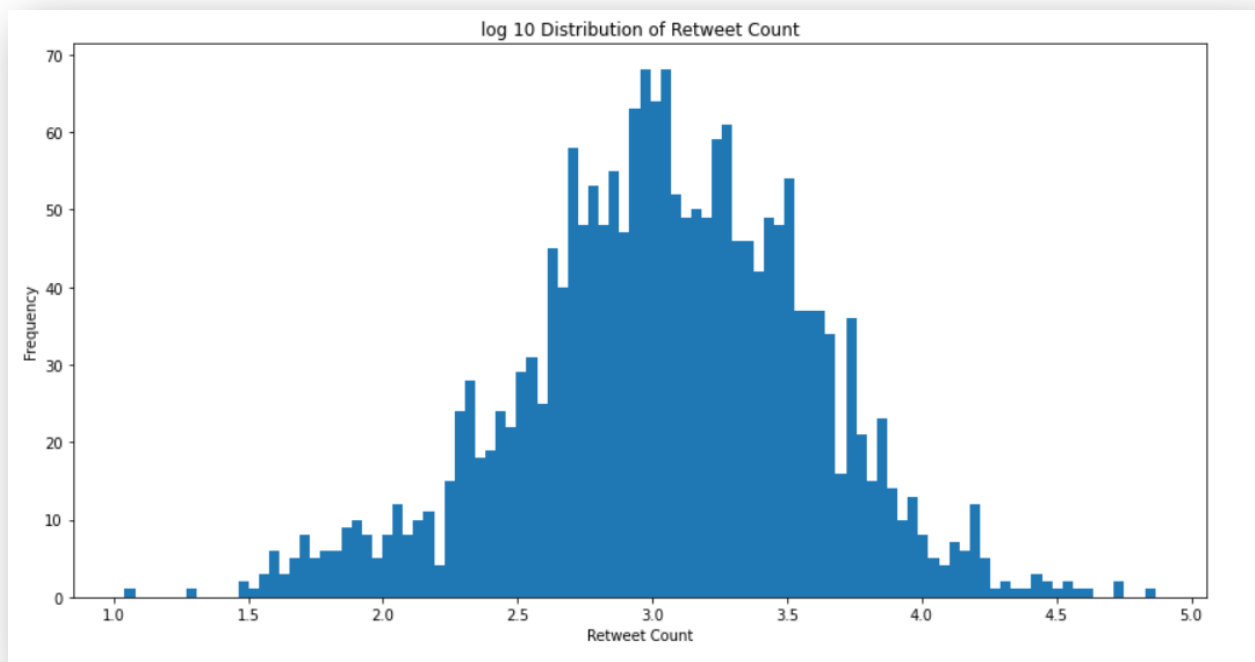
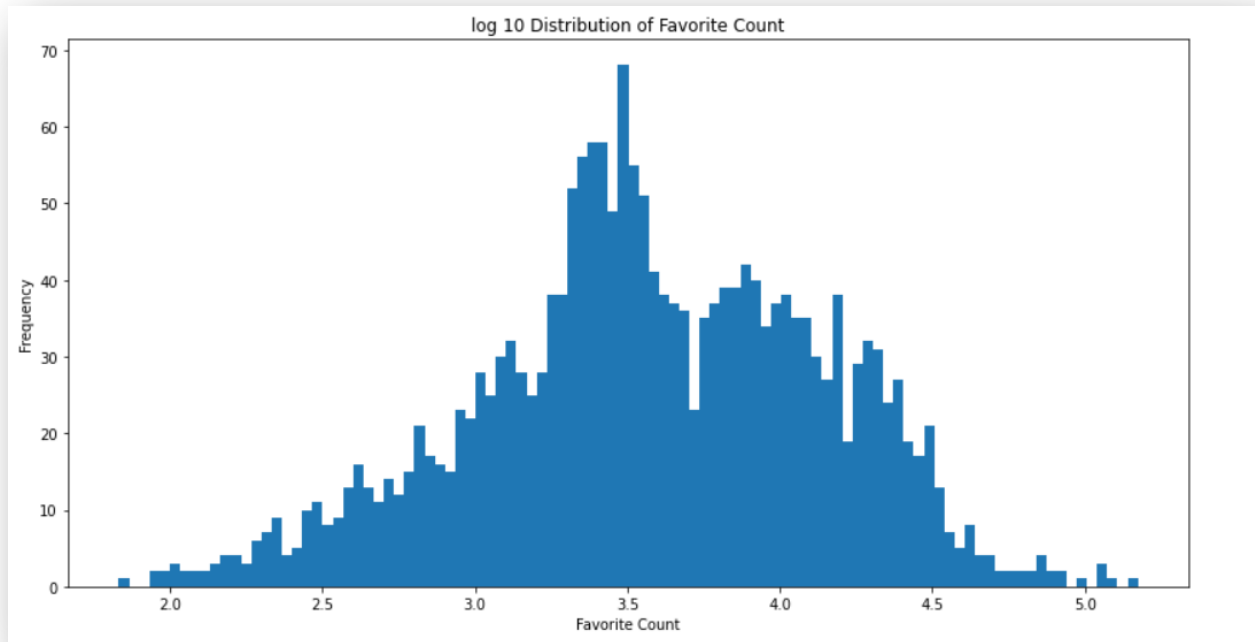
I was able to retrieve data from twitter archive via Twitter's API using tweepy along with some data provided by Udacity. Udacity was able to run every image in the WeRateDogs Twitter archive through a neural network which was able to classify dogs by their breeds. In result, the top three image predictions along with Tweet ID, image URL, and image that relates to the most confident predication.

Based on this data, I asked myself a few questions like: What is the most utilized source? How many dogs have a rating above 10? What is the difference of retweets and favorite count depending on the presence of a dog's name and the rating numerator? 10 most popular dog names? Just to name a few

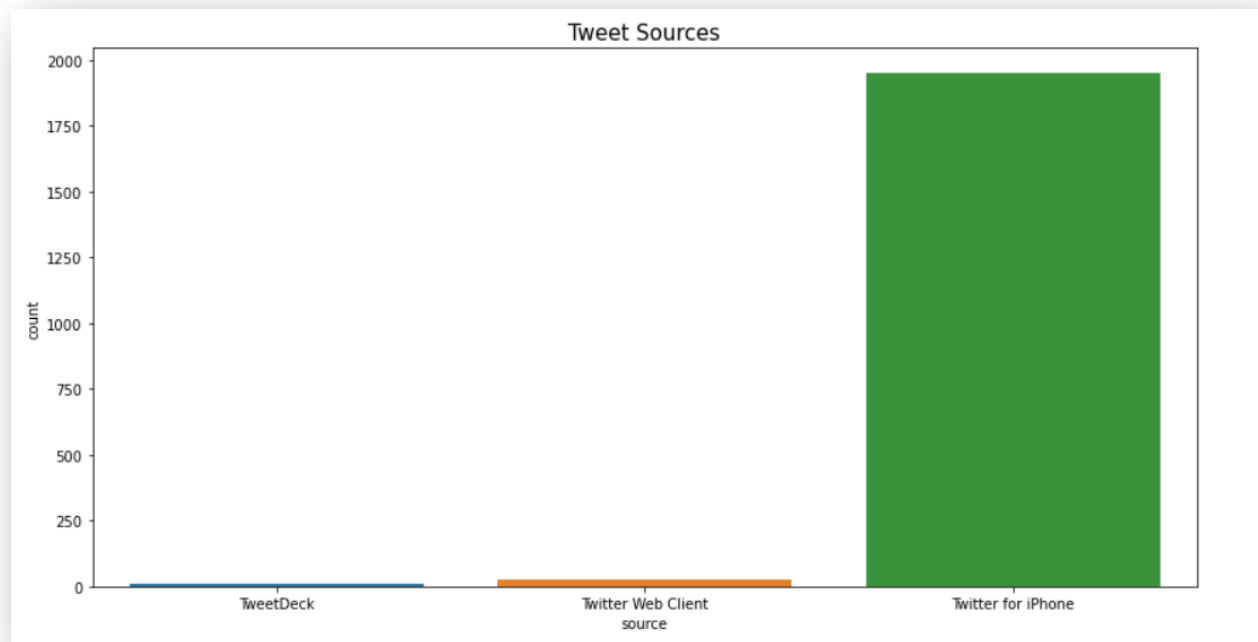
Before I could begin to answer these questions, I had to wrangle three datasets and create new variables and merge into new datasets for deeper analysis



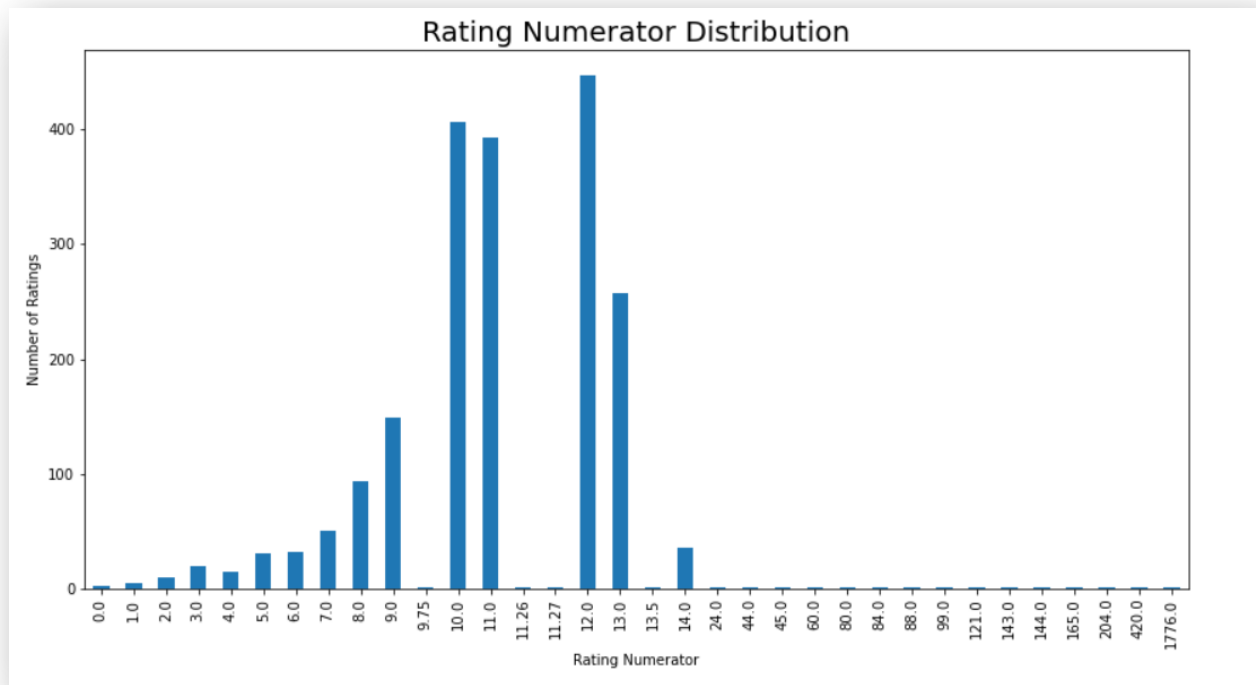
Number of tweets hit a peak December 2015 then fell dramatically through April 2016 and has remained constant until July 2017, which appears to me another drop afterwards.



It appears both Favorite Count and Retweet Count are of normal distribution

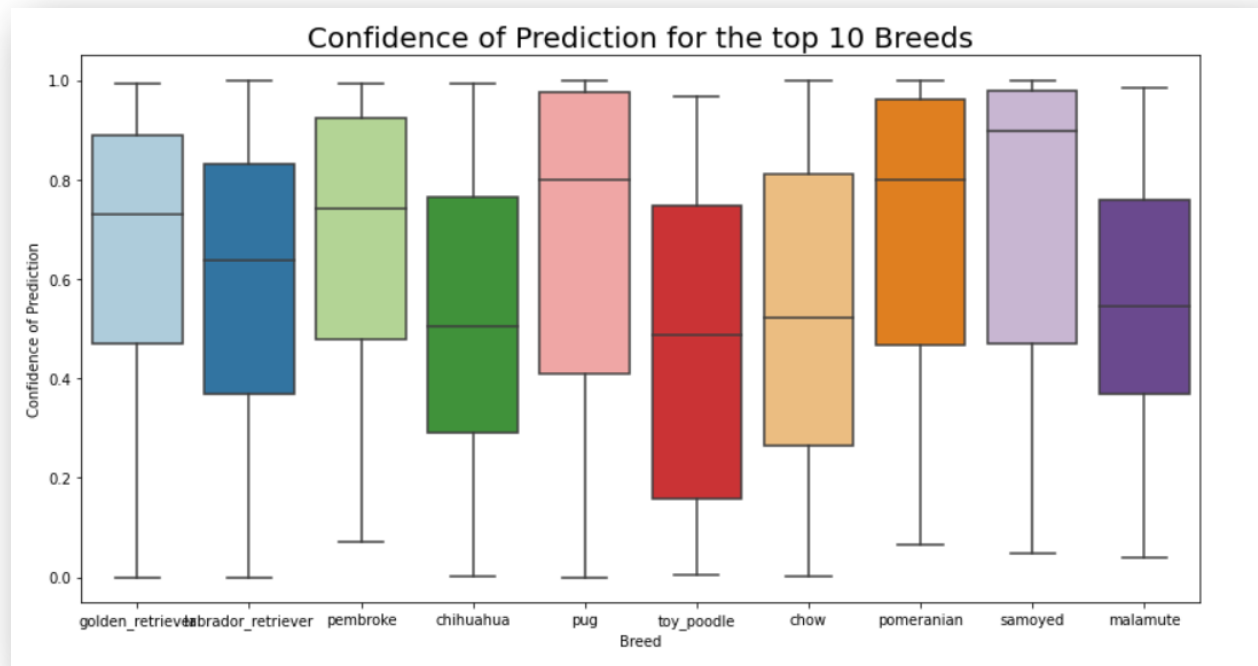


For the most part, Twitter for iPhone is the most utilized source followed by Twitter Web Client and TweetDeck



12 was used the most in numerator ratings

1150 dogs were rated above 10



The confidence varies between the top 10 predicted dog breeds and some breeds the algorithm was more confident (i.e., pug, pomeranian, samoyed), most likely due to their unique features, whereas some breeds the predictions were less (50% or less) like chihuahua and toy_poodle, probably because these breeds can be mistaken for other dog breeds were they share some features