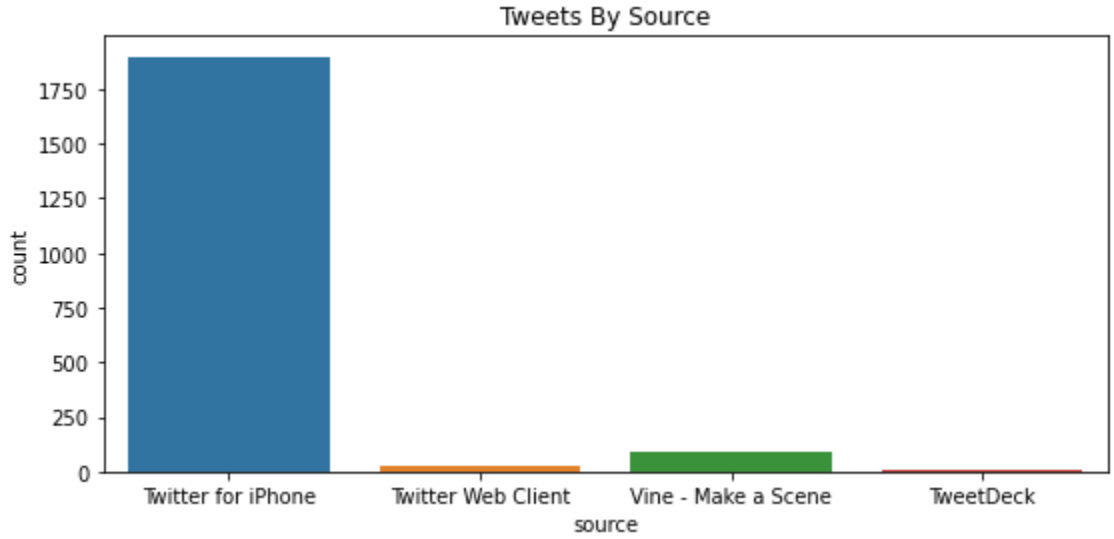
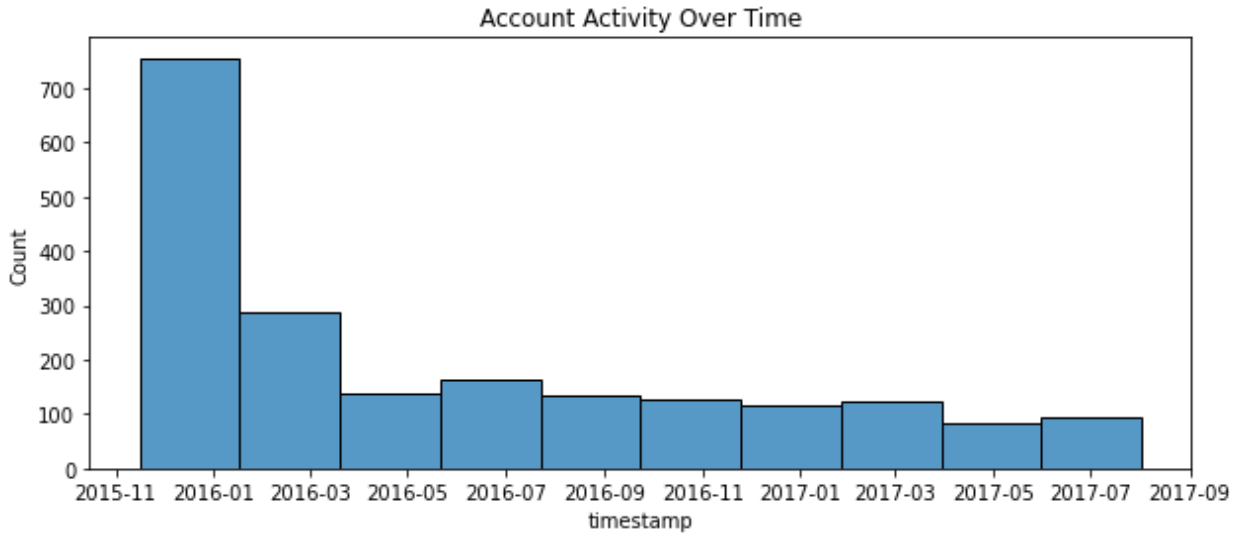
Findings Report

The following report is about our findings from analyzing the Twitter archive of the Twitter account: WeRateDogs.
It contains information about the dog's breeds, names, score given, etc.

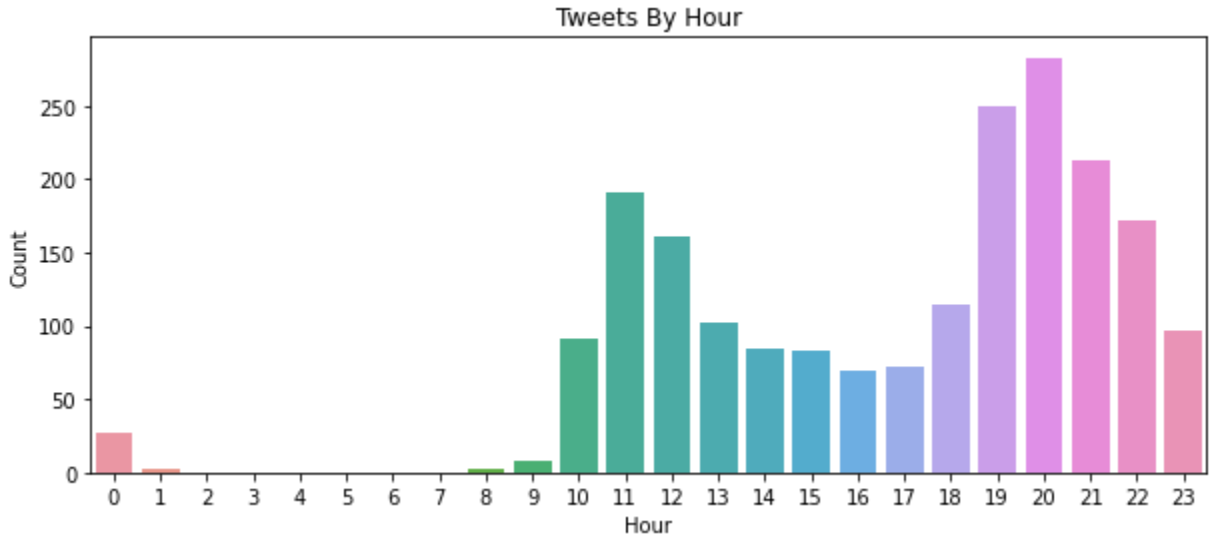
Before we start we need to analyze the source of all the tweets, and as we can see dog owners are mostly iPhone users:



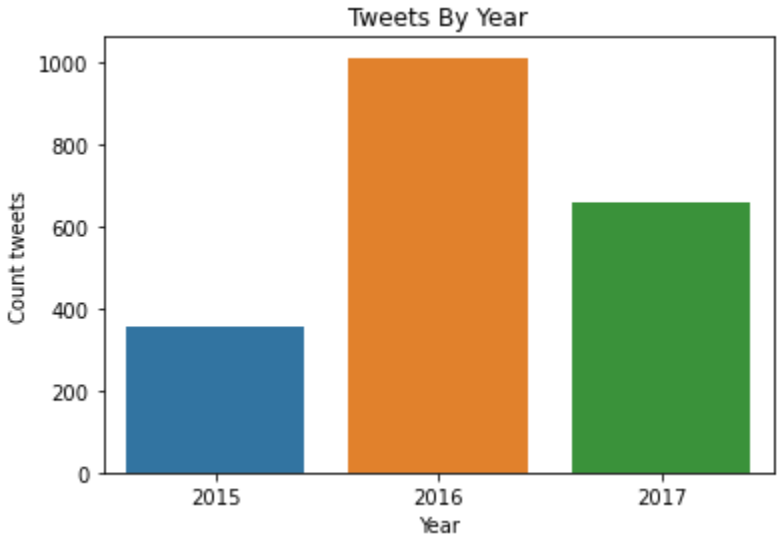
Looking at the timestamps of the dogs We can see that when WeRateDogs started out in 2015 the account was very active, but we can see that over time it started slowing down.



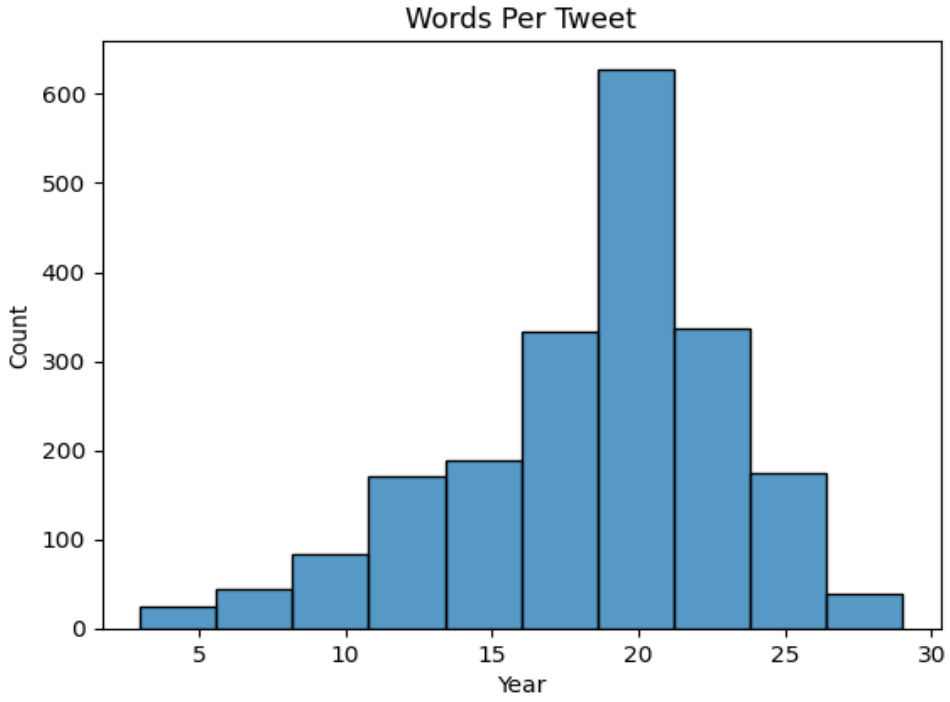
From analyzing the tweet's timestamp (that comes from the API) we can see that the account is most active in the evening betweeen 7.00 to 9.00 PM, and least active between 1 to 9 AM.



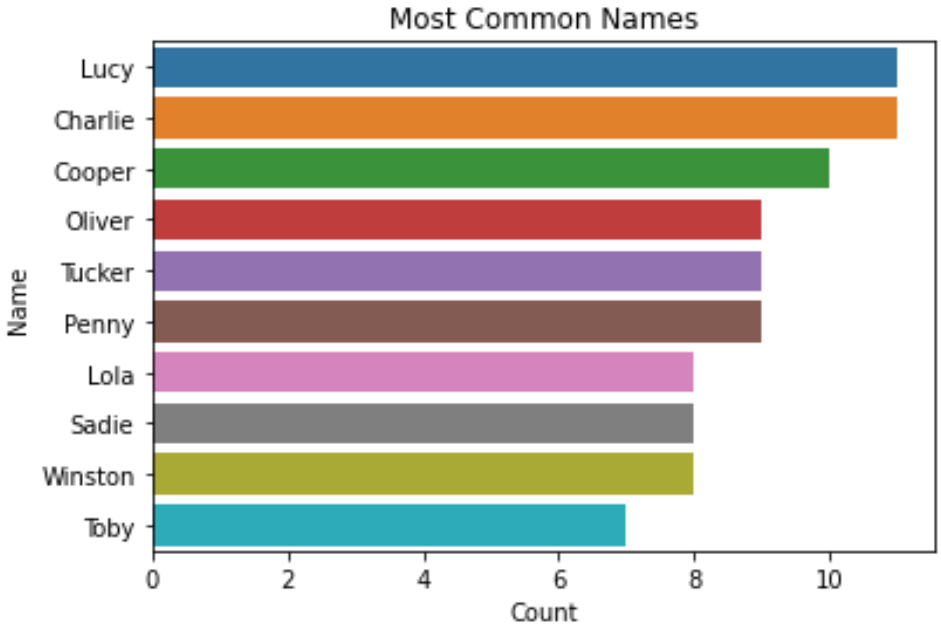
Also the data from the API we can see that the year which the account was most active (year that most tweets were posted) is 2016, with most than a thousand tweets posted!



By analyzing the text of all the tweets we can see that the median amount of words per tweet is exactly 20 (and the mean 18):



Interestingly enough it seems like Lucy and Charlie are the most common dog names in our dataset



Using the ‘image-predictions.tsv’ dataset we can get the dog breed for each tweet’s dog, This dataset was created using a machine-learning neural network that analyzed the dog pictures from all the tweets and created three predictions for each dog/tweet.

In the main dataframe (twitter_archive_master.csv) we have a column named ‘final_prediction’ which contains the number one dog-breed prediction for each tweet, these are the top 10 dog predictions: