

## Act Report On Analysing And Visualizing Weratedogs

---

### Introduction

The special bond between dogs and their human caregivers is strikingly similar to the attachment bond between human infants and their mothers; dogs depend on human care and are very adorable companions of man.



WeRateDogs is a Twitter account with millions of followers that rates images and humorous comments about man's best friend (dogs). These ratings almost always have a denominator of 10, and the numerators vary but are almost always greater than 10.

Udacity received a downloaded Twitter archive from WeRateDogs specifically for use in data wrangling projects. This archive provides basic twitter data (tweet ID, date, content, and others) for all of their 5000+ tweets as of August 1, 2017.

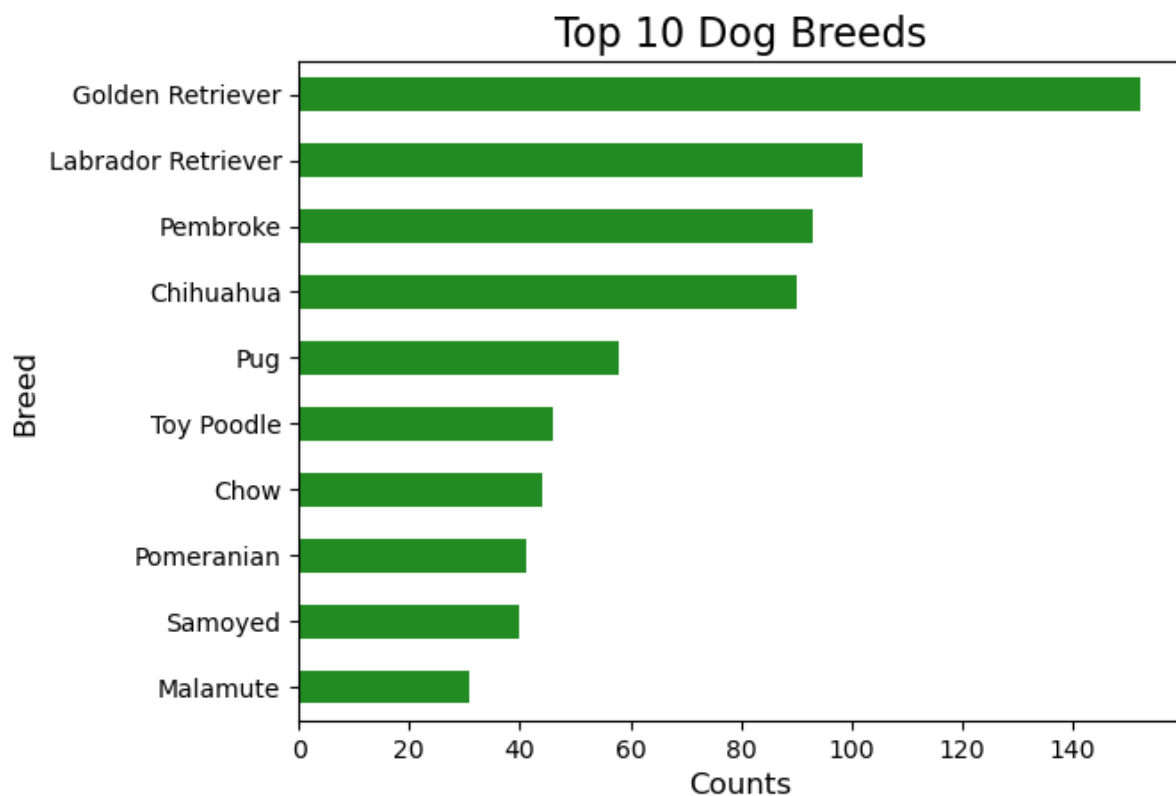
For this project, data was accessed from the WeRateDogs Twitter archive data, their tweet image prediction, and additional data from the Twitter API, which contained each tweet's retweet count and favorite ("like") count. The data was visually and programmatically assessed and cleaned.

### My analysis sought to answer the following questions:

1. What are the top 10 most popular dog breeds?
2. What is the common source of tweets?
3. Which year and month had the most tweet?
4. If there is a correlation between retweets and favorite posts (i.e., if the post is retweeted more times, it will get more visibility)?
5. If there is a correlation between Favorites counts and Ratings (i.e., whether higher rated dogs are more likely to get likes (favorites))?
6. If there is a correlation between Retweets counts and Ratings (i.e., whether higher-rated dogs are more likely to receive more retweets)?

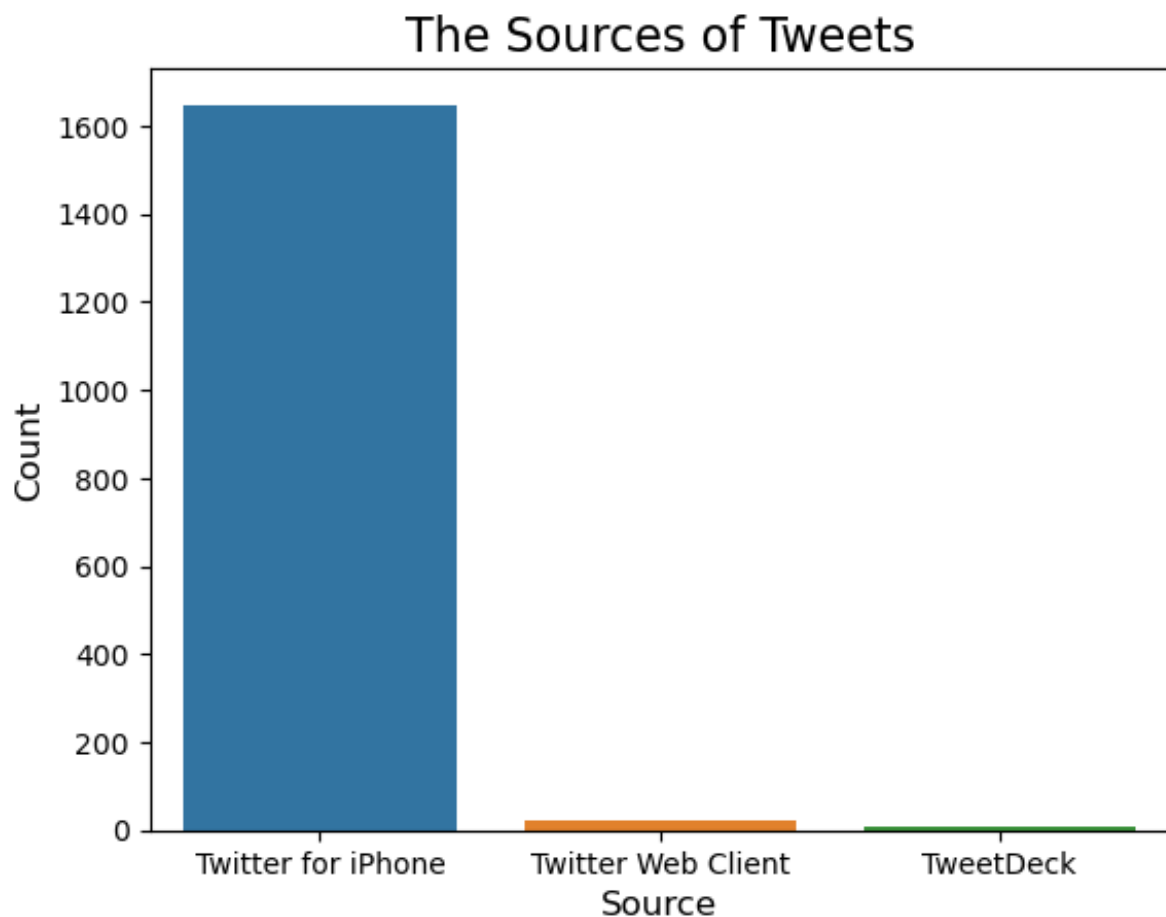
### Most popular dog breed

From the WeRateDogs datasets, I was able to analyse around 1743 tweets. The Golden Retriever emerged as the most popular dog breed with 152 counts, followed by the Labrador Retriever with 102 counts and the Malamute breed ended in the top 10 list with 31 counts.



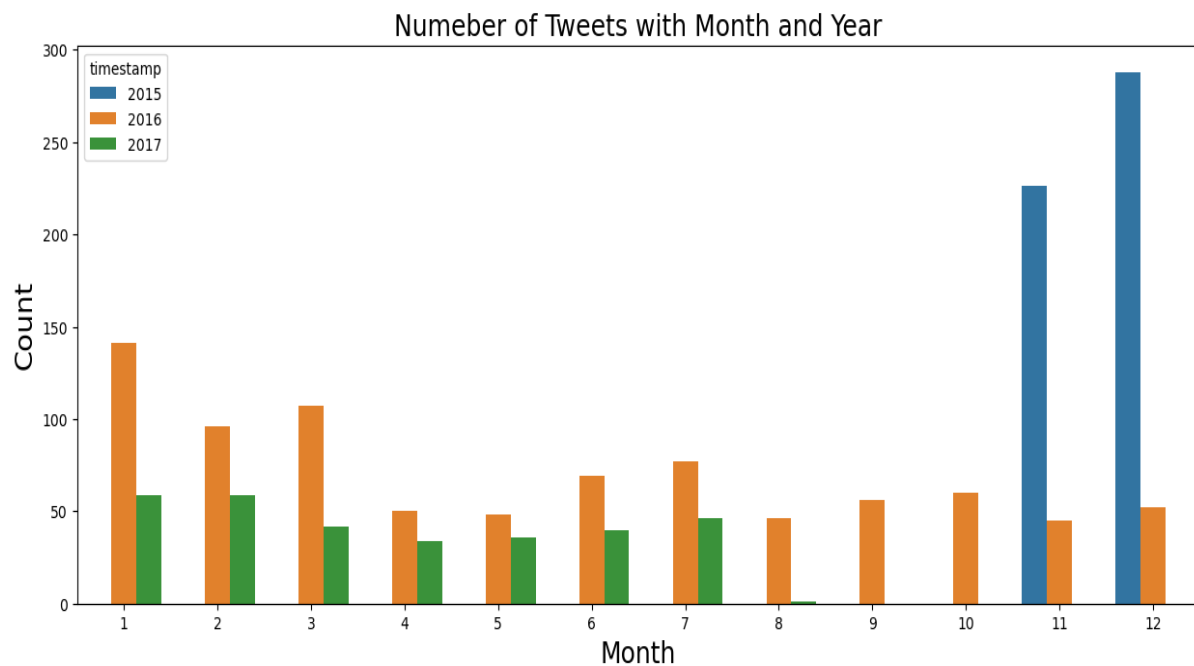
## The Common Source Of Tweets

The analysed tweets emanated from three main sources i.e., Twitter for iPhones, Twitter Web Clients and TweetDeck. However, Twitter for iPhones accounted for 98.2% (1,648) of all analysed tweets, followed by Twitter Web Client with 1.3% (22) and TweetDeck with the least of 0.5% (8).



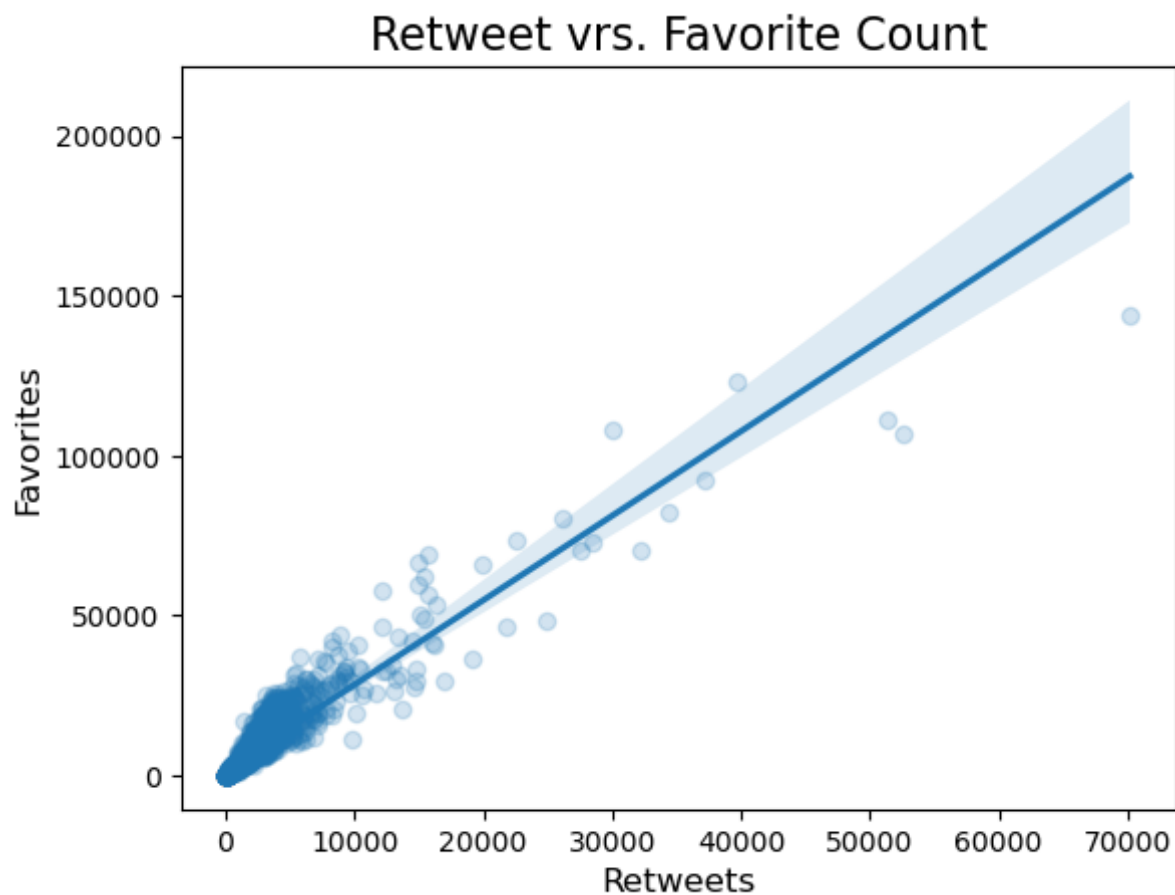
## The year and month with the most tweet

The year with the most was 2016, accounting for 887 of the analysed tweets. However, 2017 was the year with the least tweets, with 332, since 2017's tweets did not cover the entire year. On the month with the most tweets, December tops with 347 tweets. However, most tweets were recorded between November and March, and June and July, which happens to fall within Christmas and Summer holiday seasons.



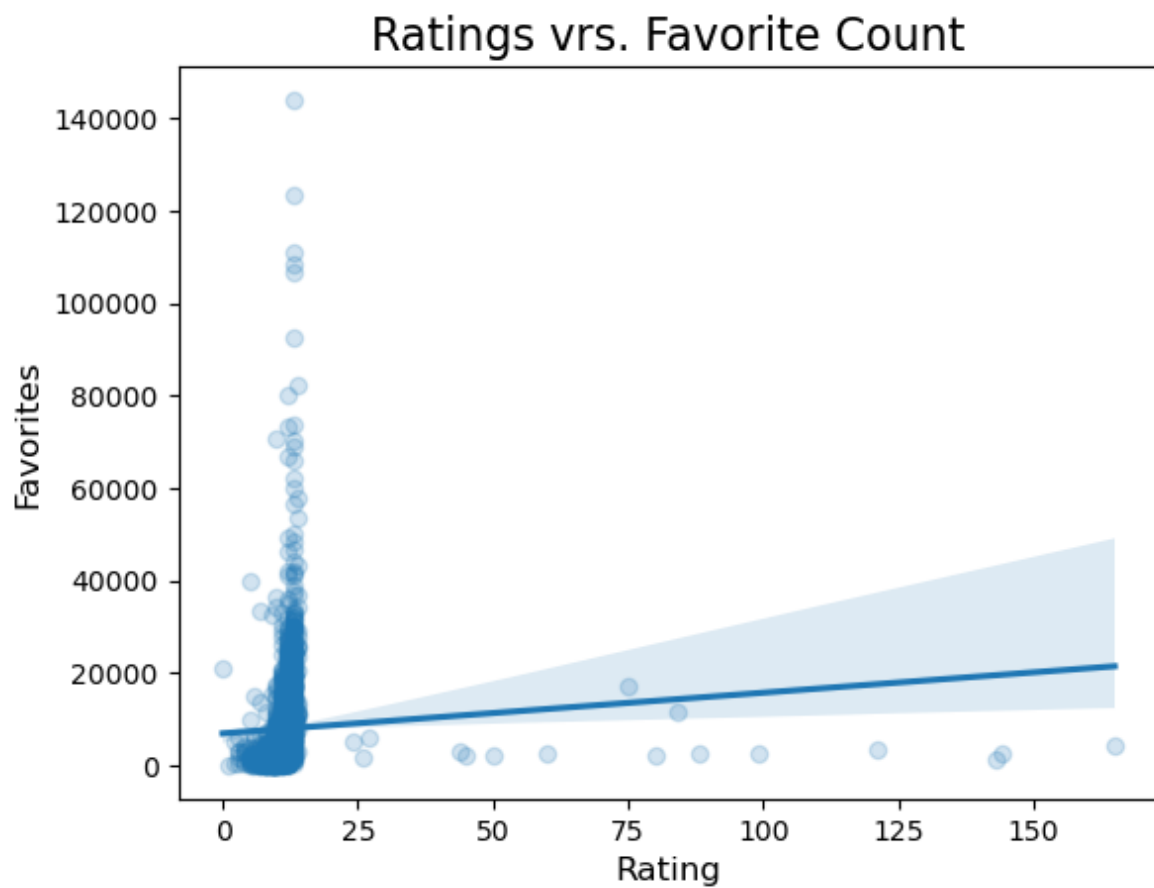
## The correlations between retweets and favorites posts

There is a very strong positive correlation with a coefficient of correlation value of 0.93 between retweet count and favorite counts of tweets. The trend line in the plot is positive which indicates posts which are retweeted more times are likely to get more visibility and receive more likes (favorites).



## The correlations between Favorite counts and Ratings

On whether higher-rated dogs are more likely to get favorites, there is a moderate positive correlation with a coefficient of correlation value of 0.06 between retweets counts and ratings. The plot's trend line is positive, indicating that posts with higher ratings are likely to be liked (favorite).



### The correlations between Retweets counts and Ratings

A slightly moderate positive correlation with a coefficient of correlation value of 0.05 between retweets counts and ratings. The plot's trend line is positive, indicating that posts with higher ratings are more likely to be retweeted.

