# We Rate Dog Twitter Account Analysis

*Udacity Data Analysis Nanodegree Project*

*By Najwa Almulhem*

*February 3, 2019*

## Introduction



WeRateDogs is a popular Twitter account that rates people's dogs with a humorous comment about the dog. The account has more than 7 million followers as of today. People send their dog pictures to get the rating. Most of the dogs are rated normally but a lot are over rated because a lot of people really like dogs

## Getting the Data

In order to have a good analysis, I had to gather the data from three sources:

1- The WeRateDogs Twitter archive to get basic tweet basic information like tweet id, timestamp, text, source... etc. .In addition to the name of the dog, rating, and dog stage. The scope here is any tweet before August 2017.
2- Neural network image prediction for dog breeds which has 3 prediction with their confident percentage
3- Twitter API for additional tweet information like number of like and retweets.

## Assessing the Data

Data from real sources usually need a lot of cleaning before doing the analysis and the data in this article is no exception. After cleaning the data, I started asking questions.

## Let's dig deeper

I was curious about many things and I started to ask questions like:

## Which breed has most of the dogs in this account?

Popular Predicted Breeds
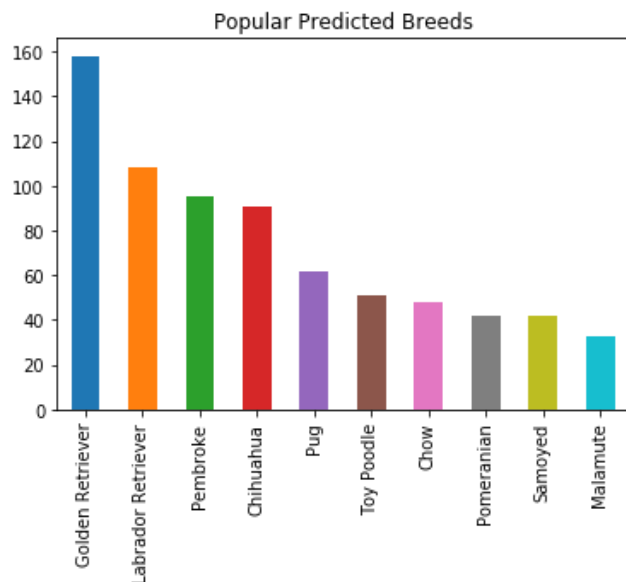
The most common breeds are Golden Retriever and Labrador Retriever with almost 150 for the Golden Retriever and 110 for Labrador Retriever. This is after choosing the dog predictions only and choosing the prediction with the highest confidence. Our analysis is also verified by this article in Wikipedia (List of most popular dog breeds).


Labrador Retriever


Golden Retriever

## What is the best time to post about your dog?

If I have a dog, I would like to have the highest number of likes and retweets. Because of course, my dog is the cutest!

So when should I do that, what day of the week? I always thought that people are more into social media in the weekend but our analysis showed something else. Wednesday on average has more likes, retweets .The difference is high between Wednesday, and the next day, around 1000 more likes and 800 more retweets.
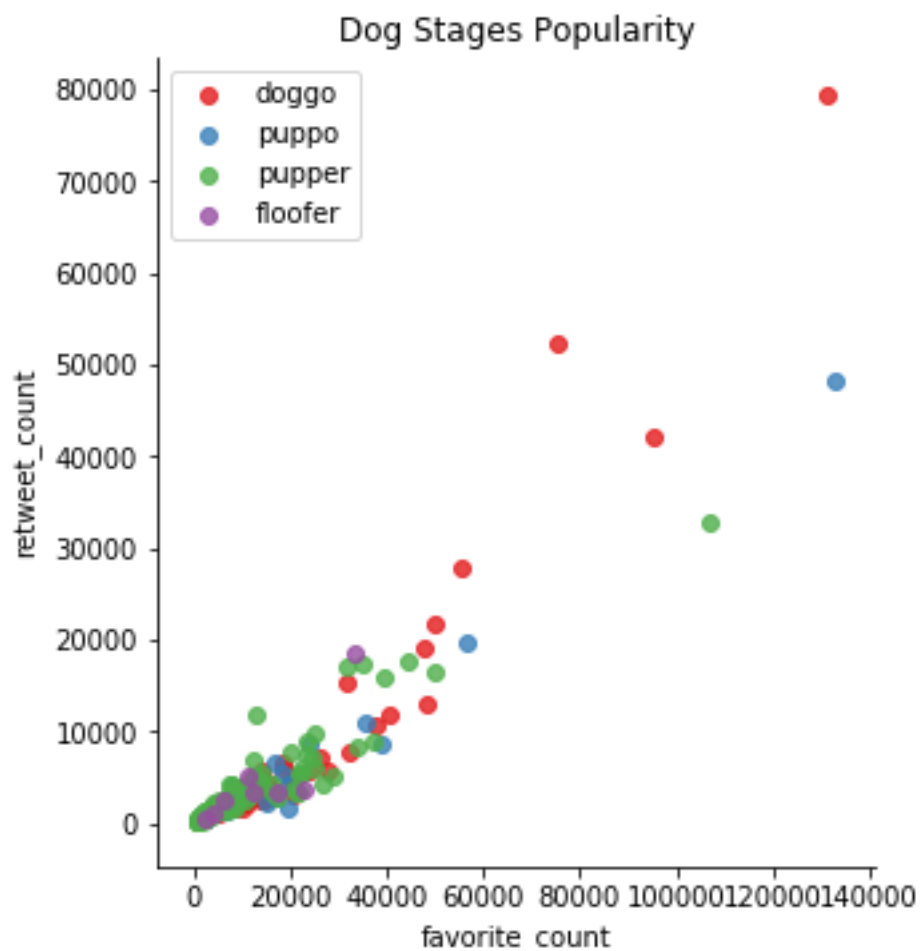
| day_of_week | favorite_count | retweet_count |
| --- | --- | --- |
| Wednesday | 9965.390572 | 3089.626263 |
| Tuesday | 9085.413559 | 2719.606780 |
| Saturday | 8977.011905 | 2967.952381 |
| Sunday | 8714.121569 | 2671.525490 |
| Friday | 8642.627660 | 2782.088652 |
| Monday | 8503.214067 | 2636.507645 |
| Thursday | 8339.214789 | 2525.964789 |

In addition, if we zoom out a little and try to find the months that have more likes and retweets in average. As shown in the plot, late spring and early summer months on average have more likes and retweets. I tried to understand this more by looking at the distribution and my guess was correct, there are few outliers that are affecting this .One tweet with a high number of like or retweets can make the mean higher.

Mean of Favourites & Retweets by Month

Distribution of Tweets Over months by Favourites & Retweets
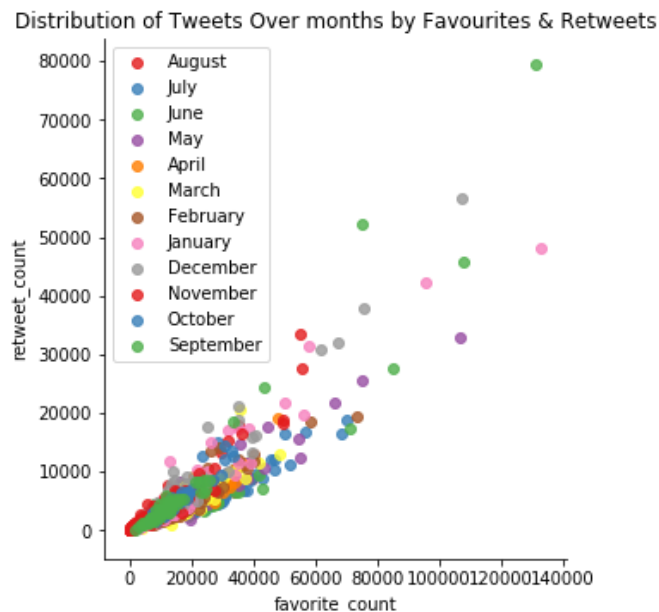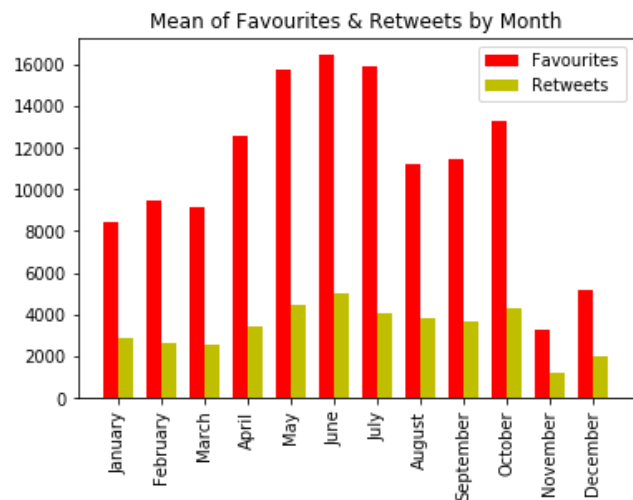
## When should I post about my dog?

In which dog stage should I post about my dog? As shown in the graph, doggo is the most popular stage sometimes but this is not always the case.



Dog Stages Popularity

## What breed usually get higher rates?

| breed_pred | rate | favorite_count | retweet_count |
|---|---|---|---|
| Clumber | 2.700000 | 7320.000000 | 1885.000000 |
| Bouvier Des Flandres | 1.300000 | 18032.000000 | 4479.000000 |
| Saluki | 1.250000 | 24060.000000 | 5133.750000 |
| Pomeranian | 1.247619 | 7451.357143 | 2670.738095 |
| Briard | 1.233333 | 9511.666667 | 2966.666667 |
| Tibetan Mastiff | 1.225000 | 12173.500000 | 4193.000000 |
| Border Terrier | 1.214286 | 13578.000000 | 3377.142857 |
| Silky Terrier | 1.200000 | 7317.000000 | 1891.000000 |
| Standard Schnauzer | 1.200000 | 2046.000000 | 928.000000 |
| Gordon Setter | 1.175000 | 5529.250000 | 2495.500000 |



Although Clumber gets high ratings, it does not get more likes or retweets.

## Most Retweeted and liked

Most Retweeted
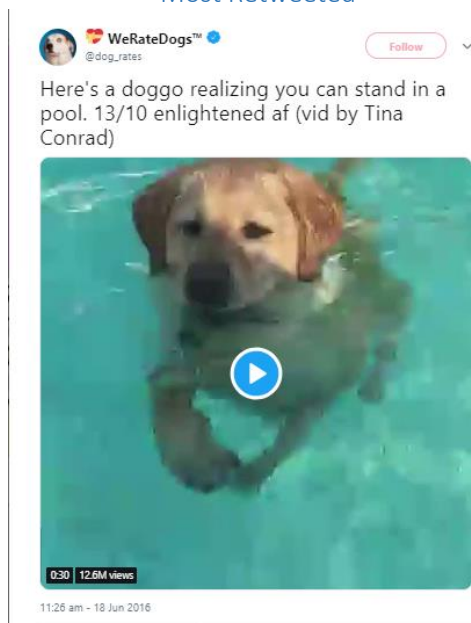


Most Liked



# Summary

This data is very interesting, but I would like to dig more into it in future and remove some outliers to have a better analysis. I would also like to get more details about the tweet and analysis it