

Visualizations and insights of WeRateDogs

Introduction

The analysis and visualization of the data within in this documents is based on the twitter account: WeRateDogs.

This is a twitter account rates people's dogs with a funny comment about the dog. These ratings(numerator) were out of 10(denominator), however, some of the dogs had unusual ratings whereby the numerator was far higher than the denominator and some in some cases the denominator was above 10. Data cleaning was performed on the data to provide a cleaner visualization of the data and to eliminate extreme outliers that may have skewed the data.

The main variable of interest in the insights and visualization of the data were:

- Dog breeds
- Ratings
- Retweet counts
- Favourite counts
- Dog stages(pupper, puppo, doggo, floofer)

Questions of interest:

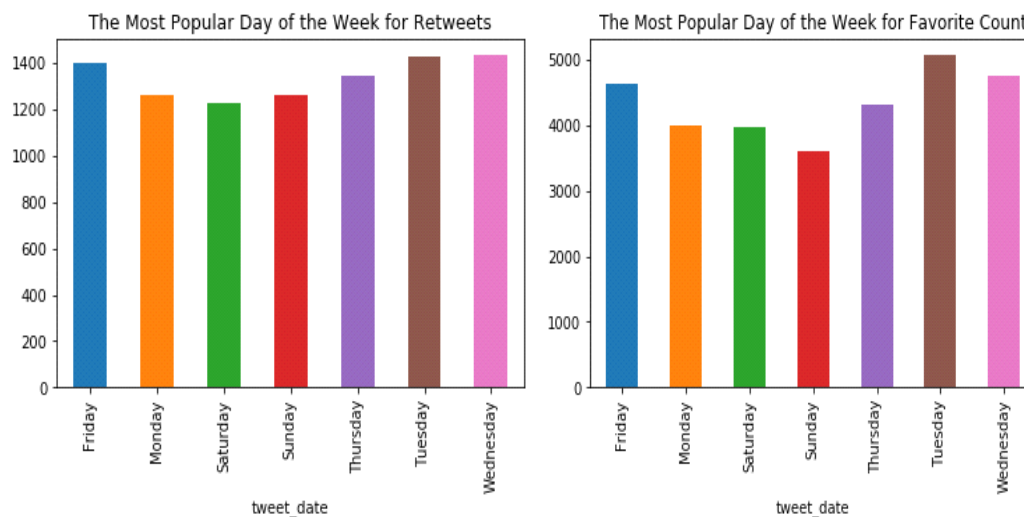
- What was the most common dog breed?
- What was the most common dog stage?

- Which dog breed had the highest ratings?
- Which dog stage had the highest ratings?
- What dog breed was the most popular?
- What dog stage was the most popular?
- Did the timing of tweets have any significance?
- Was there a correlation between Retweets and Favourites?

Activity of tweets over time

I've used a bar graph to display tweet activity over the days of the week.

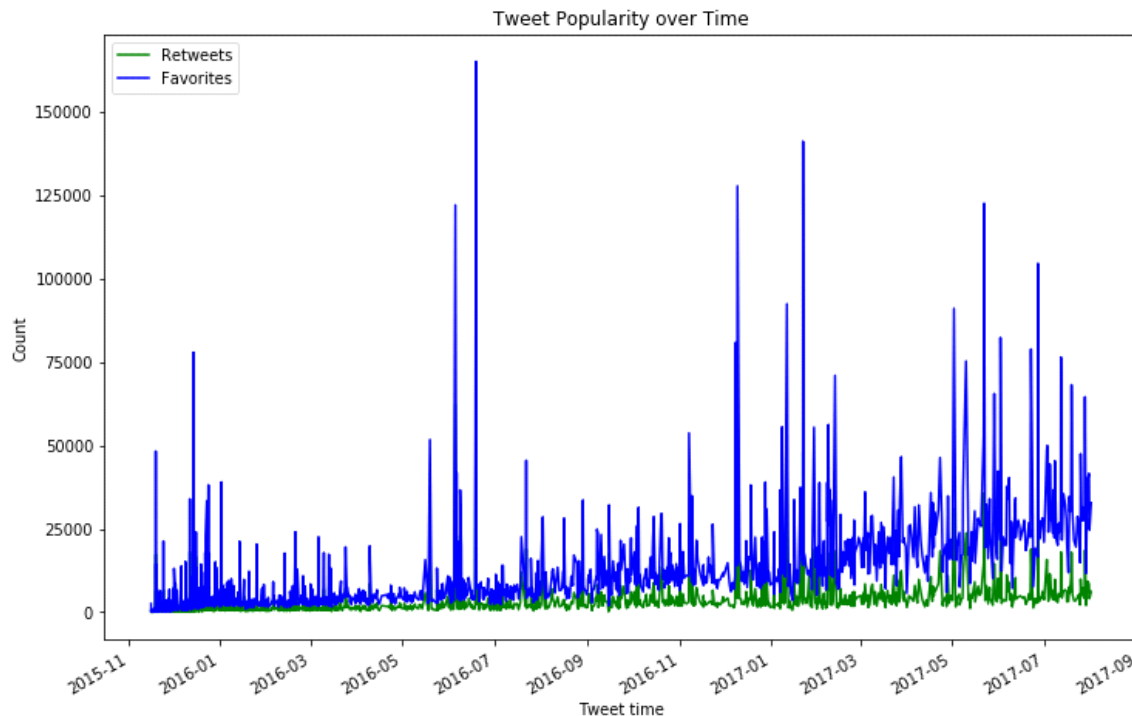
The most likely day of the week for retweets and favorite counts to occur:



It seems that people are more likely to retweet the posts about dogs on a Tuesday, wednesday and Friday. With the most popular day being Wednesday with a median of 1431.0.

The most popular day for favorite count is Tuesday with a median of 5063.0, but like the retweet the days people are more likely to favor a tweet is on Wednesday. Tuesday and Friday.

I've used a line graph to visualize tweet data over the years the account has been active.



The year 2016 saw the most tweets (841) of dogs for this twitter account. This may be due to the fact that the account and the rating of dogs was only initiated late in 2015, which had a tweet count of 511. It appears to have gained traction in 2016 and then people may have lost interest to post in 2017, which had a count of 511.

It is interesting though that the most retweets and favorites occurred in 2017 despite this being the year with the least tweet activity

Most common Dog Stage and Most Popular Dog Stage

The data included four dog stages: pupper, puppo, doggo and floofer.

dog_stage counts:

pupper - 180

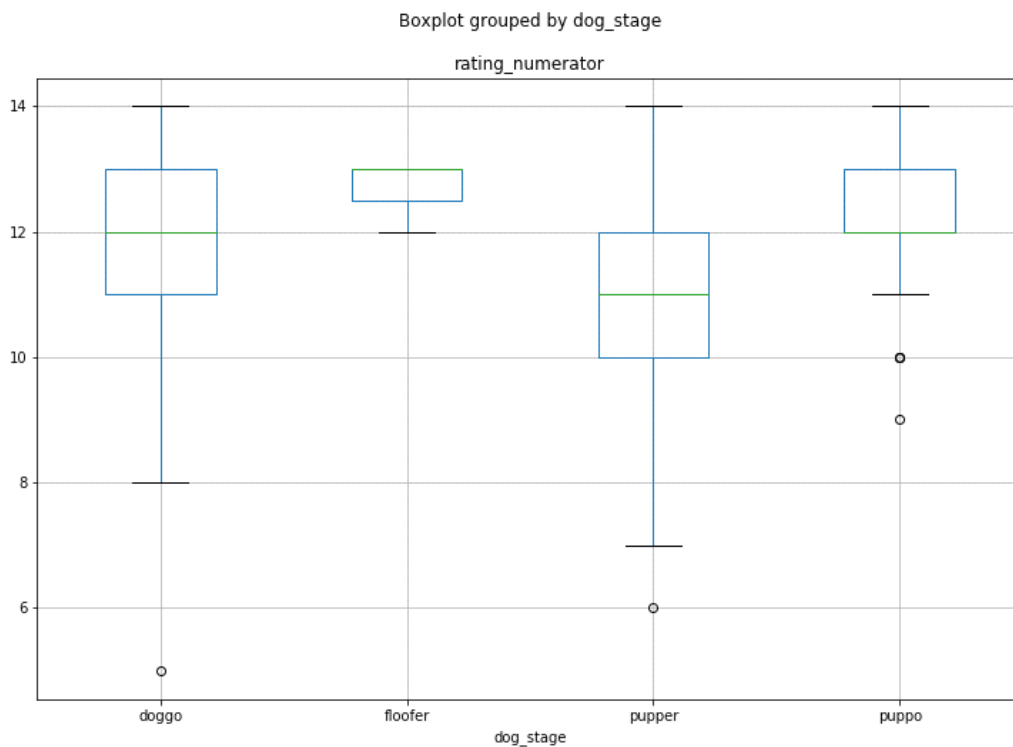
doggo - 61

puppo - 27

floofer - 3

The most common dog stage that was tweeted about is the pupper stage.

I took a closer look at the popularity of dog stages by plotting a boxplot, a boxplot was used to also gauge outliers in the data and how it would affect the results,

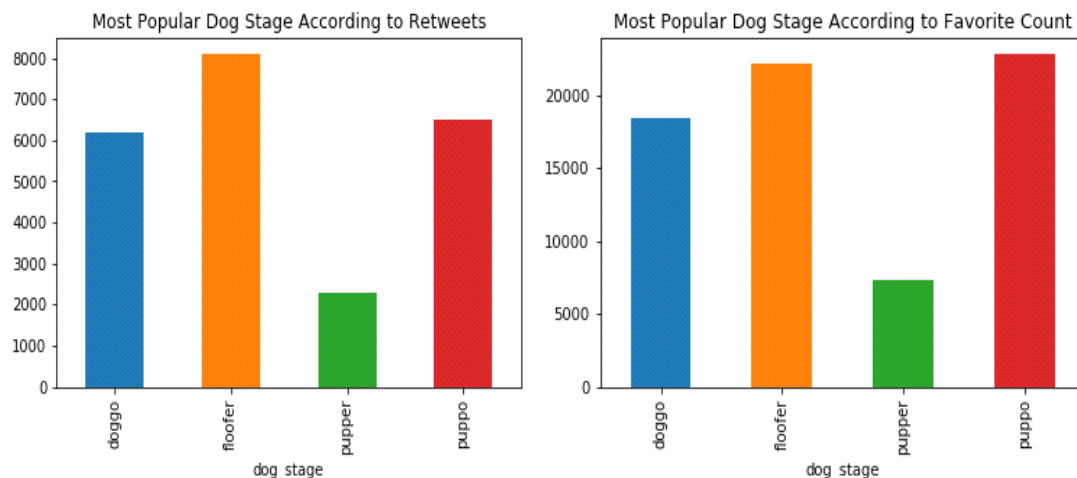


Puppies appear to get, on average, lower ratings than the other dog stages. It is interesting to note that the standard deviation for pupper is the second highest at 1.41. This std makes sense though, because we can see that there are some outliers on the boxplot, as well as for doggo and puppo.

Floofers, on average, have higher ratings than the other dog stages, with a mean of

12.66. Floofers appear to be consistently rated above 10. We also have to take into consideration that there are only 3 counts of floofers, a much lower count than the other stages

The most popular dog stage by rating seems to be floofer but i would argue that perhaps it is puppo or doggo because there is a much better sample size

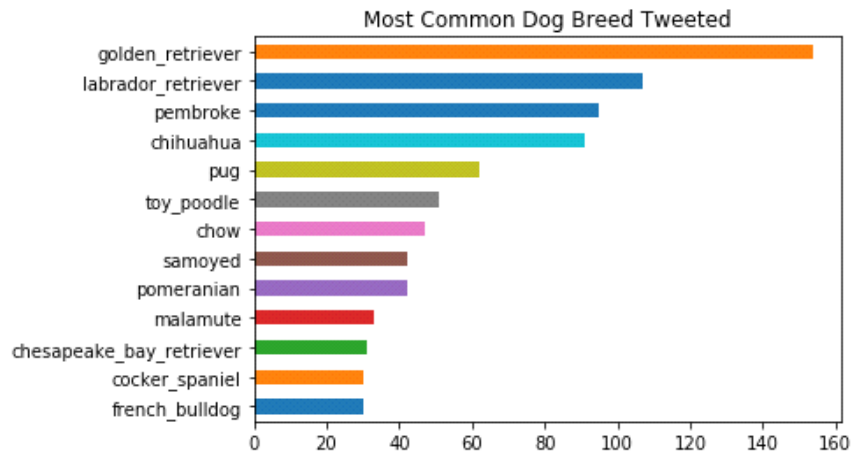


If we group the dog stages by favorite count then we see that puppo is the most popular by receiving the most 'likes'.

If we look at popularity by retweets, floofer is the most popular dog type, with this dog stage being retweeted the most.

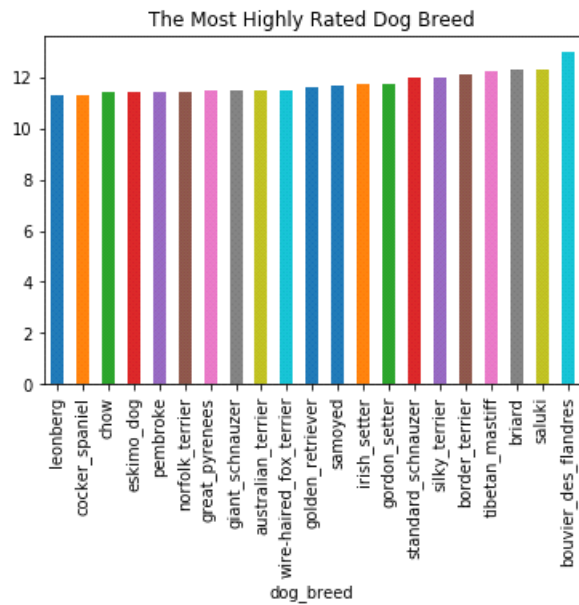
This appears to be consistent with my earlier observation that perhaps puppo or doggo is the most popular dog stage.

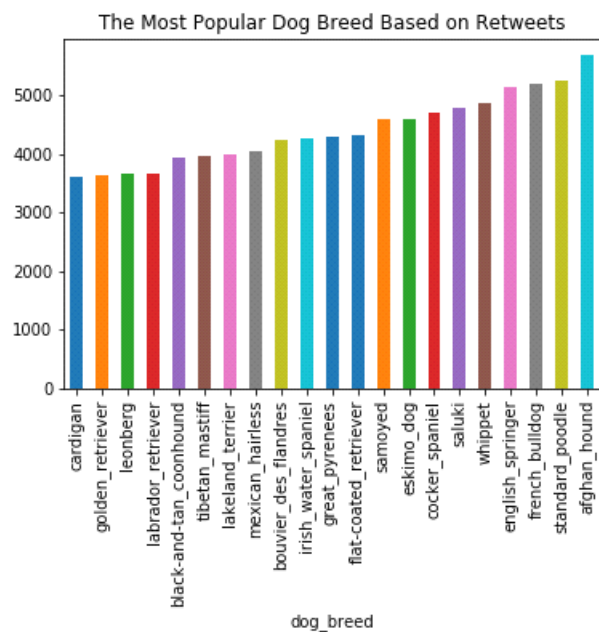
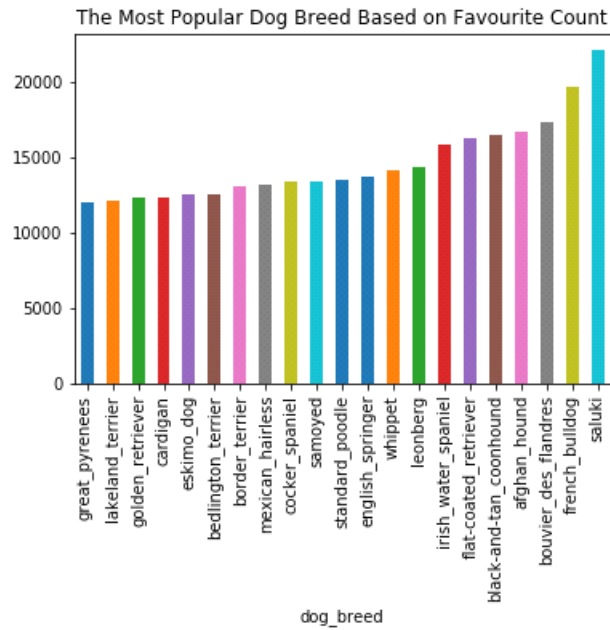
Most Common Dog Breeds



The bar graph shows us that the most common dog breed is the Golden Retriever with 154 tweets of this particular dog breed. The second most common breed is the Labrador Retriever with people tweeting this breed 107 times.

Most Popular Dog Breeds



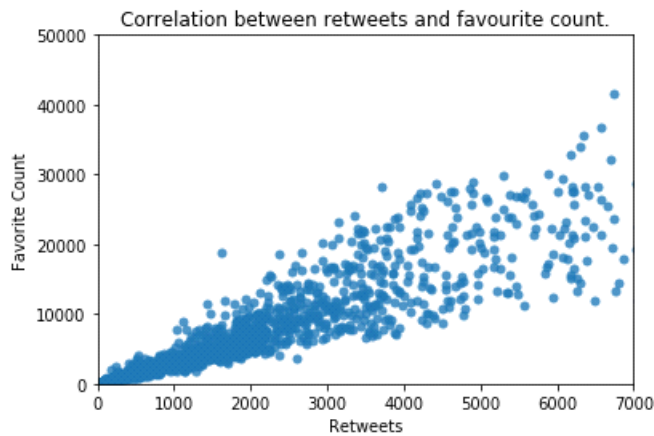
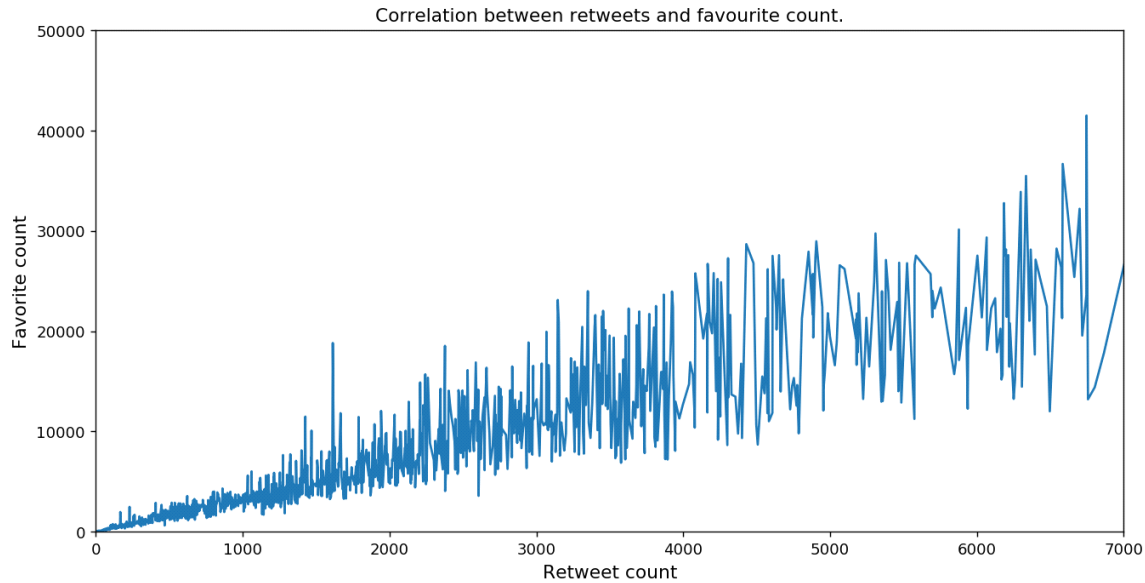


The dog breed with the highest rating is the bouvier_des_flandres, but this does not provide a definitive conclusion to popularity of breed as this dog might be the only tweet for this breed.

The most popular dog breed based on Favourite count is the Saluki.

The most popular dog breed based on retweets is the Afghan Hound.

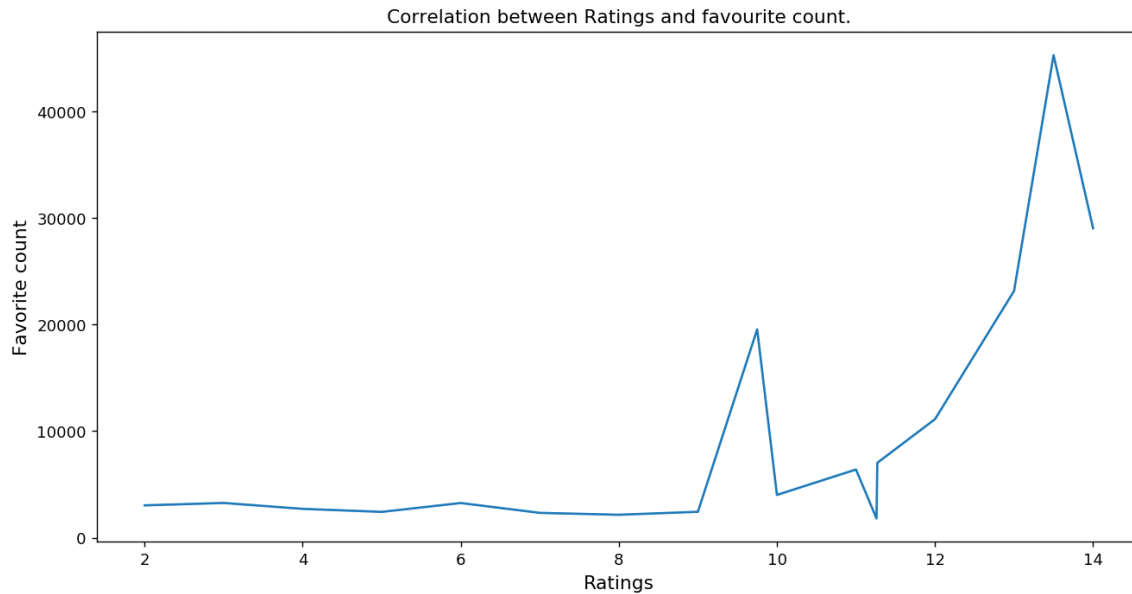
Relationship between favorite count and retweets



The same tweets has both the highest favorite counts and retweets, the same is true for the tweet that has both the lowest favorite counts, it also has the lowest retweet count.

The line graph and scatter plot seem to indicate that the higher the favorite count the higher the retweet count. This shows a strong positive linear relationship, with a correlation coefficient of 0.93.

There does not seem to be any strong correlation between retweets and ratings and favorite counts and ratings.



There appears to be a weak uphill (positive) linear relationship between ratings and favorite count. We can see in the line graph that the favorite counts are quite low and consistent until the ratings are over 10, with a weird spike and the dip at rating