

WeRateDogs Data Analysis

The [WeRateDogs](#) twitter user joined in November 2015 and has 9.3 Million followers. The user posts pictures and videos of dogs giving them ratings out of 10. Part of the uniqueness of the posts is that the ratings are often much larger than 10 or out of other numbers than 10. Some dogs are also given special stages such as 'floofer' or 'pupper' based on appearance.

Data was gathered on some of the WeRateDogs statuses using 3 different sources and methods:

1. Directly from a csv file containing archived data
2. Using the python requests library to download machine learning image predictions for dog breeds performed on the status images
3. Directly querying the twitter API for retweet and favourite counts

The data was cleaned to create a tidy master dataset. This included removing retweets and replies, converting the rating system to be out of 10 to make comparisons easier and putting the dog stages into a single column.

From the tidy dataset it was possible to derive insights about the WeRateDogs statuses. The status with the greatest retweet and favourite counts was determined to be the same dog. This was a dog that was predicted to be a labrador retriever in the dog stage doggo with a rating of 13 out of 10. The name was not given. The status was made on the 18th of July 2016. The picture came from a video thumbnail and can be seen below.



Picture of dog with the greatest retweet and favourite counts

The second insight related to tweet source it was found that the Twitter for Iphone source had the greatest average favourite and retweet counts beating the TweetDeck and Twitter Web Client sources. It must be noted that Twitter for Iphone had the vast majority of the statuses.

Finally the data was grouped by the dog breed predictions and the largest average retweet and favourite counts for these were calculated.

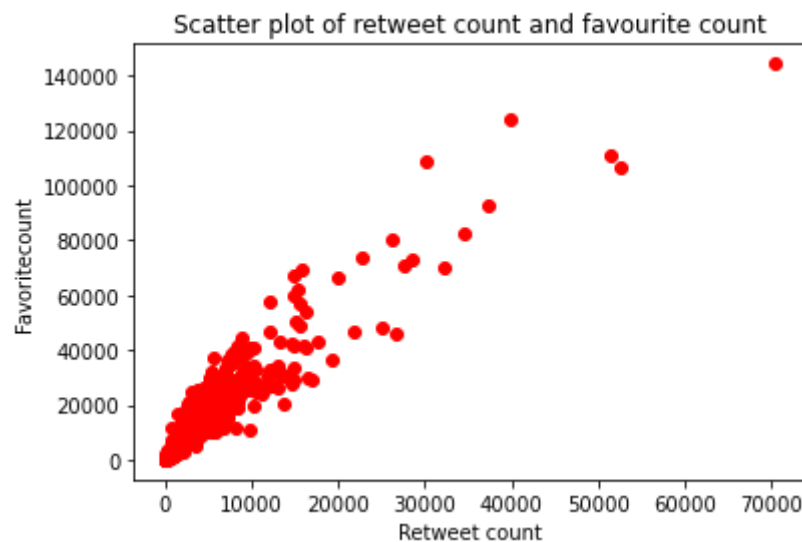
The top 3 for favourite count were found to be:

1. Bedlington terrier
2. Saluki
3. Fench bulldog

The top 3 for retweet count were found to be:

1. Bedlington terrier
2. Afghan hound
3. Standard poodle

For a visualisation a scatter plot was plotted to see how the retweet count and favourite count relate to each other. It was found that there is positive correlation between these variables which makes sense as these are similar measures of engaging positively with statuses. The plot is displayed below:



Overall WeRateDogs has some interesting data to analyse for those who have an interest in dogs.