# We Rate Dogs Twitter Project

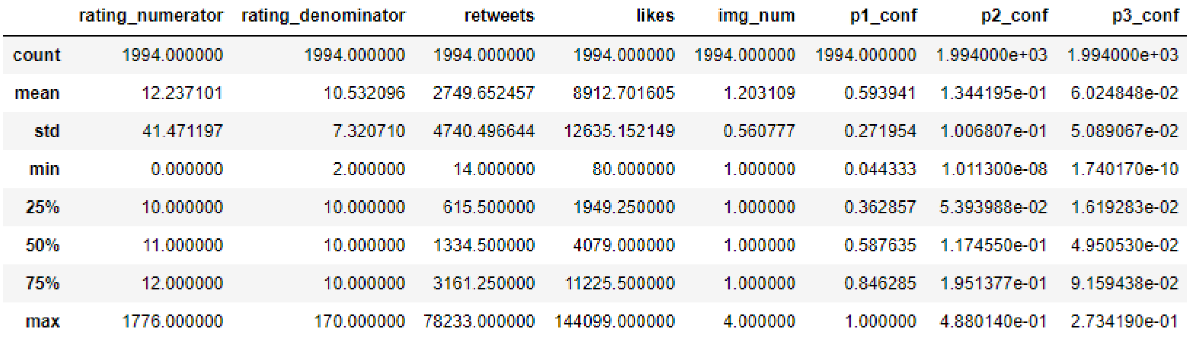
## Introduction

The analysis required gathering data from three sources. WeRateDogs supplied Udacity with their Twitter archive in CSV format to be used in this project. The archive contained the base information for their tweets up to August 1, 2017. The second source was a TSV file containing neural network data on each of the tweet images with the intent of identifying the dog breeds in the picture. This was obtained programmatically from the internet using the Requests library. Finally, the Twitter API was queried and JSON data was obtained for each tweet to be used for analysing likes and retweets.

## Analysis

In my Analysis, I initially looked at the descriptive statistics (Table 1). Average rating numerator was about 12.24 and denominator of 10.53. The mean retweets were 4,740 with a minimum of 14 and a maximum of 78,233. The mean likes were 8,912 with a minimum of 80 and a maximum of 144,099. Following the initial descriptive analysis, I looked at three key areas. These were which stage of dog had the highest rating, the most popular names and the relationship between likes and retweets.

**Table 1 – Descriptive statistics of dataset.**

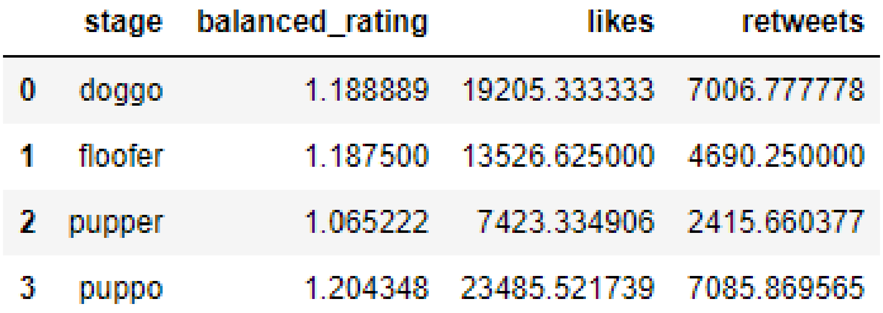


| | rating_numerator | rating_denominator | retweets | likes | img_num | p1_conf | p2_conf | p3_conf |
|---|---|---|---|---|---|---|---|---|
| **count** | 1994.000000 | 1994.000000 | 1994.000000 | 1994.000000 | 1994.000000 | 1994.000000 | 1.994000e+03 | 1.994000e+03 |
| **mean** | 12.237101 | 10.532096 | 2749.652457 | 8912.701605 | 1.203109 | 0.593941 | 1.344195e-01 | 6.024848e-02 |
| **std** | 41.471197 | 7.320710 | 4740.496644 | 12635.152149 | 0.560777 | 0.271954 | 1.006807e-01 | 5.089067e-02 |
| **min** | 0.000000 | 2.000000 | 14.000000 | 80.000000 | 1.000000 | 0.044333 | 1.011300e-08 | 1.740170e-10 |
| **25%** | 10.000000 | 10.000000 | 615.500000 | 1949.250000 | 1.000000 | 0.362857 | 5.393988e-02 | 1.619283e-02 |
| **50%** | 11.000000 | 10.000000 | 1334.500000 | 4079.000000 | 1.000000 | 0.587635 | 1.174550e-01 | 4.950530e-02 |
| **75%** | 12.000000 | 10.000000 | 3161.250000 | 11225.500000 | 1.000000 | 0.846285 | 1.951377e-01 | 9.159438e-02 |
| **max** | 1776.000000 | 170.000000 | 78233.000000 | 144099.000000 | 4.000000 | 1.000000 | 4.880140e-01 | 2.734190e-01 |

## Which stage of dog had the highest rating?

I wanted to explore which stage of dog appeared to be the more popular among viewers. While 10 was the most common denominator, a number were above or below this number. To overcome this issue, I created a balanced rating variable by dividing each dog's numerator by its denominator, then I grouped by dog stage and found the mean balanced rating, likes and retweets for each stage as seen in Table 2. Unsurprisingly the 'puppo' stage had the highest balanced rating, likes and retweets.

**Table 2 – Rating, likes and retweets by dog stage**



|  | stage | balanced_rating | likes | retweets |
|---|---|---|---|---|
| 0 | doggo | 1.188889 | 19205.333333 | 7006.777778 |
| 1 | floofer | 1.187500 | 13526.625000 | 4690.250000 |
| 2 | pupper | 1.065222 | 7423.334906 | 2415.660377 |
| 3 | puppo | 1.204348 | 23485.521739 | 7085.869565 |

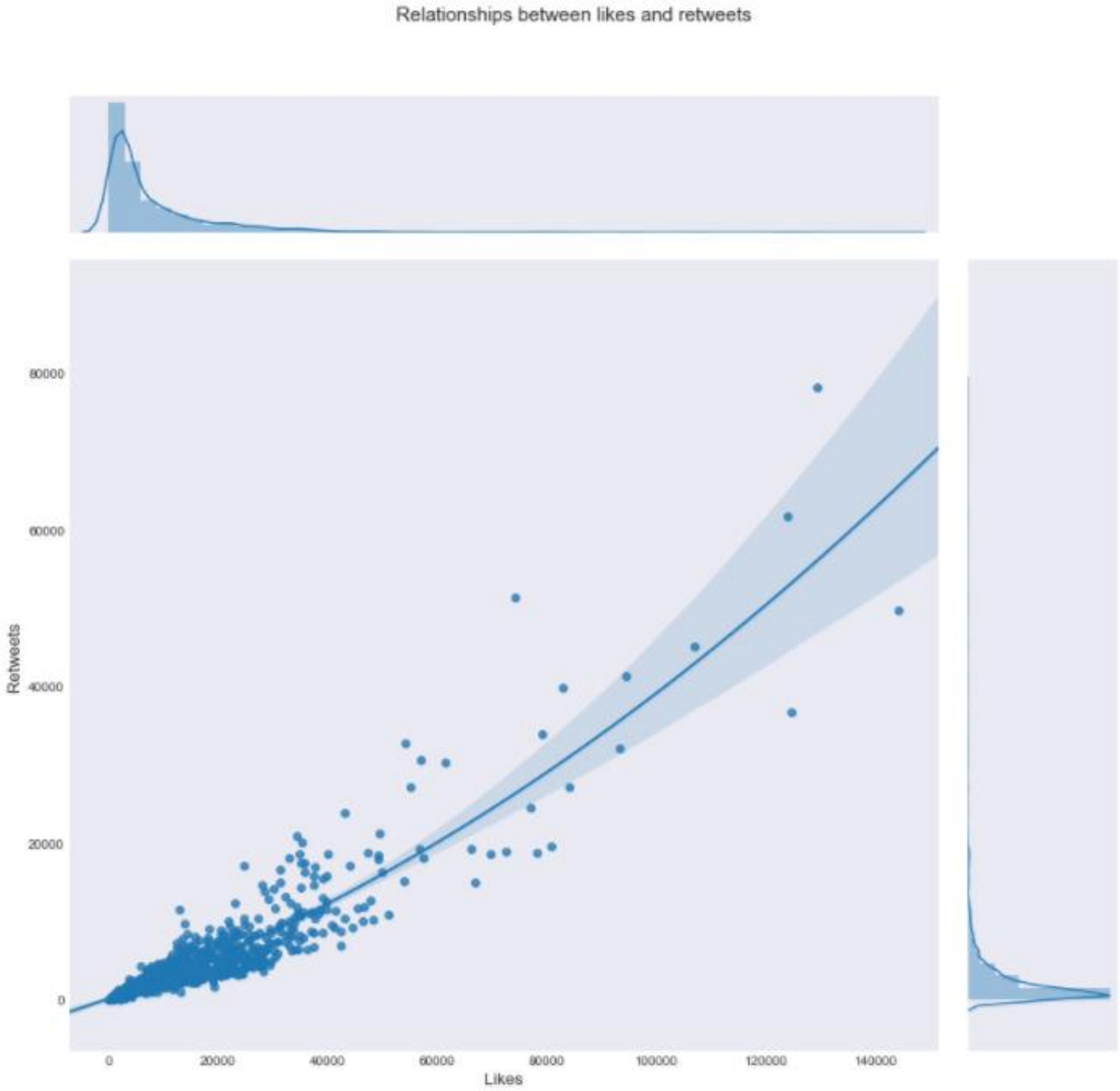
## Most popular dog names

I wanted to see which dog names were the most common in the dataset. To examine this, I performed a value count on the name variable. The most popular dog name (not including NaNs) was 'Charlie' with a count of 11. This was followed closely by 'Oliver', 'Lucy' and 'Cooper' all with counts of 10.

## Relationship between likes and retweets.

The final analysis I performed was to examine the relationship between the likes and retweets of a post. From Figure 1 we can see that both likes and retweets are heavily skewed right. People are also more inclined to like something than retweet it. The trend line shows that there is a significant positive correlation between the two variables which makes sense given that popular tweets will be both liked and retweeted.

**Figure 1 – Relationship between likes and retweets**



## Conclusion

We saw that the most highly rated, liked and retweeted dog stage was a ‘puppo’. The most popular name for a dog was Charlie and that there is a high positive correlation between liked and retweets.