

## WeRateDogs – Analysis & Visualizations

### Part 1: Analyzing Data

- Analysis 1: Calculating the mean retweet count and mean favorite count for all 4 dog stages (doggo, floofer, pupper & puppo).

Here, I've calculated the mean retweet and mean favorite count for all the four dog stages.

Here, we can clearly see that the difference between the mean retweet count and the mean favorite count is vast.

The highest mean retweet count is of doggo: 7291 while the lowest mean retweet count is of pupper: 2436

The difference between the highest and lowest mean retweet count is also significant.

Also, the highest mean favorite count is of puppo: 23368 while the lowest mean favorite count is of pupper: 7998

Here, the difference is even greater as compared to the mean retweet count.

- Analysis 2: Analyzing number of tweets every year and every month of each year.

Here, I'm calculating the number of tweet count for every year and then every month of each year.

The year is extracted from the timestamp column. Year 2016 has the highest number of tweet count of 748.

Also, this can be further detailed by extracting number of tweets each month of every year. December 2015 has the highest number of tweet count of 245, followed by November of the same year.

- Analysis 3: Analyzing the source for various tweets.

Here, I've analyzed various sources of tweets.

Twitter for iPhone has more than 95% of the count in this case.

- Analysis 4: Analyzing dog breeds based on confidence intervals.

Here, I'm analyzing the count of different dog breeds at different confidence intervals.

At 90% confidence interval, dog breed Golden Retriever has highest count of 36 followed by Pembroke with a count of 26.

Here, at 95% confidence interval, dog breed pug has highest count of 20 followed by Pembroke with a count of 19.

We can clearly see that as the confidence interval increases, the count has gone down significantly.

- Analysis 5: At a particular confidence interval, the dog breed having maximum retweet & favorite count.

Here, I'm analyzing the dog breed having highest retweet and favorite count at a particular confidence interval.

At 99% confidence interval, dog breed Vizsla has the highest retweet and favorite count of 10964 & 31215 respectively.

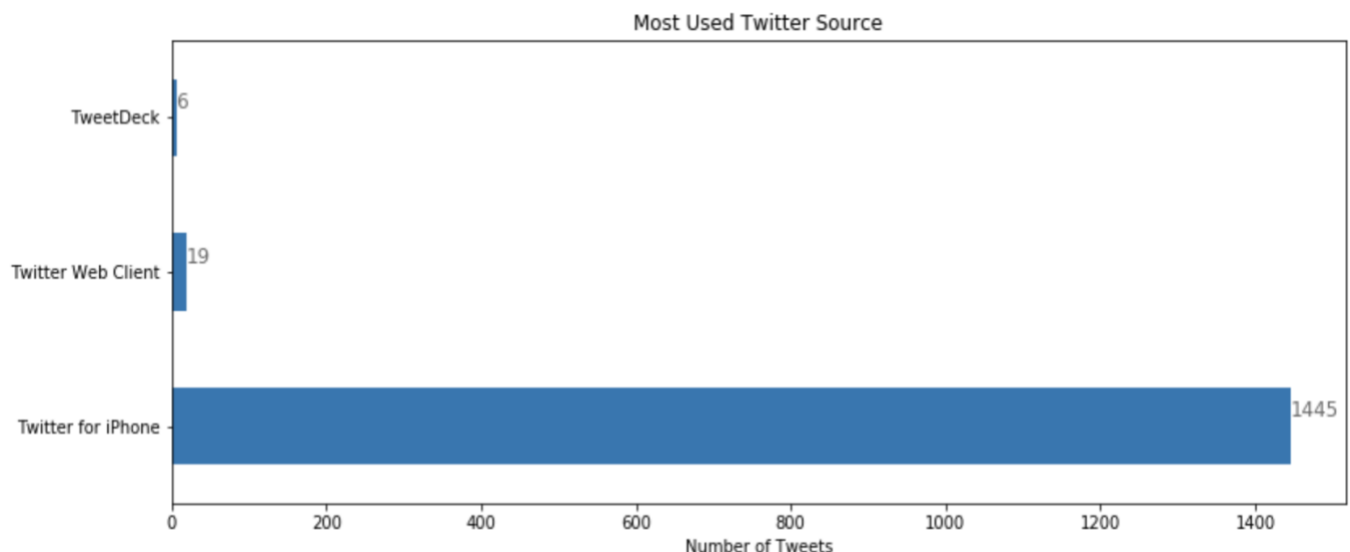
At 95% confidence interval, dog breed Weimaraner has the highest retweet and favorite count of 21983 & 51385 respectively.

At 99% confidence interval, dog breed Yorkshire Terrier has the highest retweet and favorite count of 34114 & 119650 respectively.

## Part 2: Visualizing Data

- Visualization 1:

Here, I'm visualizing the sources of various tweets. Below is the snapshot of the output.

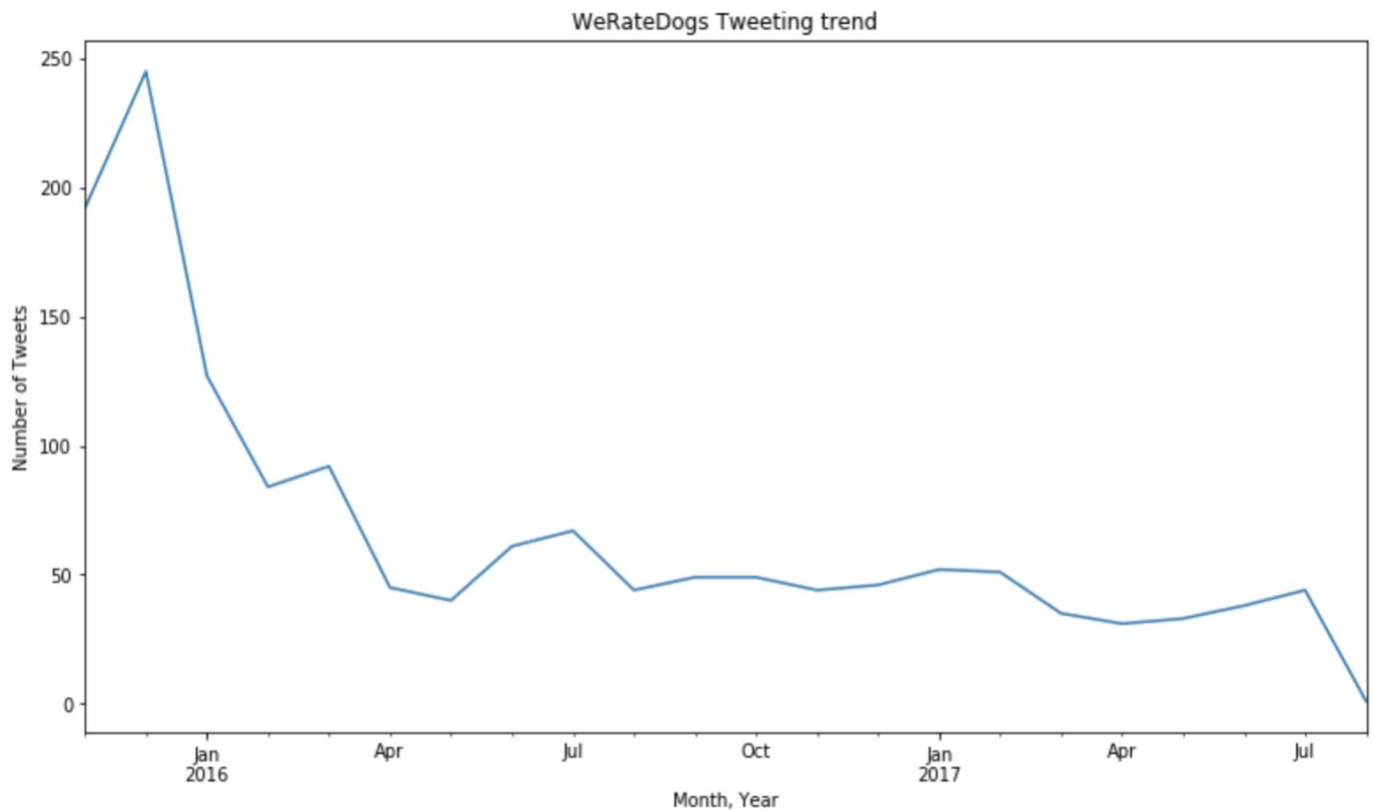


The above horizontal bar plot clearly indicates that Twitter for iPhone source has highest number of value counts of 1445. Both the other sources have hardly been used.

To visualize the horizontal bar plot better, I have also included the individual bar count to the right so that it gives much better insight.

- Visualization 2:

Here, I'm visualizing the tweeting trend of the WeRateDogs data. Below is the snapshot of the output.



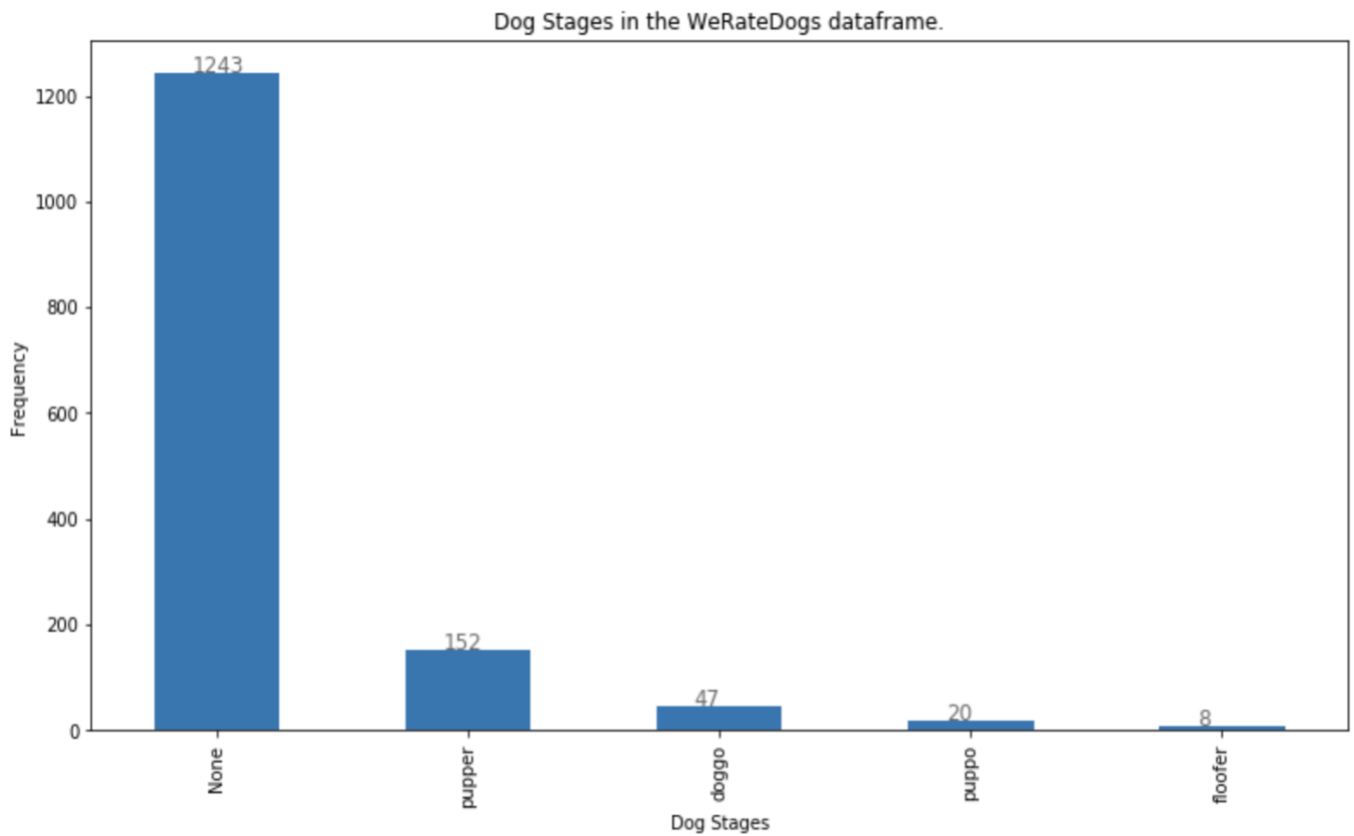
The above line plot represents the tweeting trend of WeRateDogs data for each the years 2016-2017. The trend is seen to be going down.

I have included month of each year in the above visualization to give a better insight. As you see that in the months from August 2016 to February 2017, the tweeting trend was constant but then dropped later on.

Similar was the case in December 2015 but then drastically dropped in January 2016.

- Visualization 3:

Here, I'm visualizing number of counts in each of the dog stages. Below is the snapshot of the output.

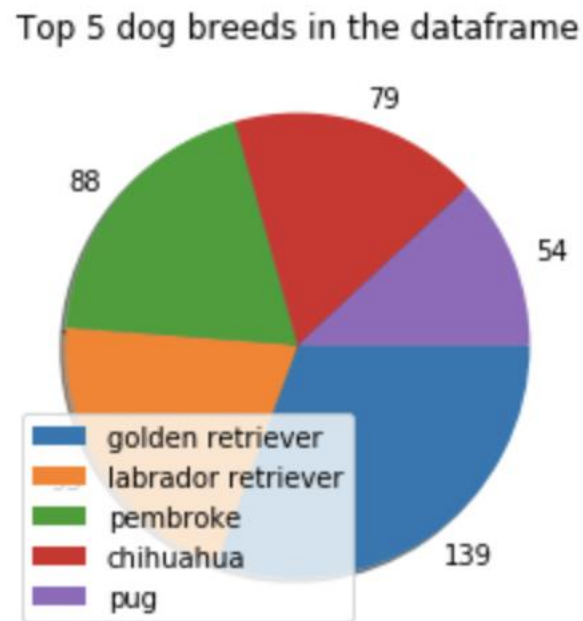


The above bar plot shows the value counts for each of the dog stages.

Surprisingly, None is the most common dog stage value throughout the data followed by pupper. I have also included the individual count above the bar to provide a much clear insight.

- Visualization 4:

Here, I'm visualizing top 5 dog breeds in the WeRateDogs data. Below is the snapshot of the output.

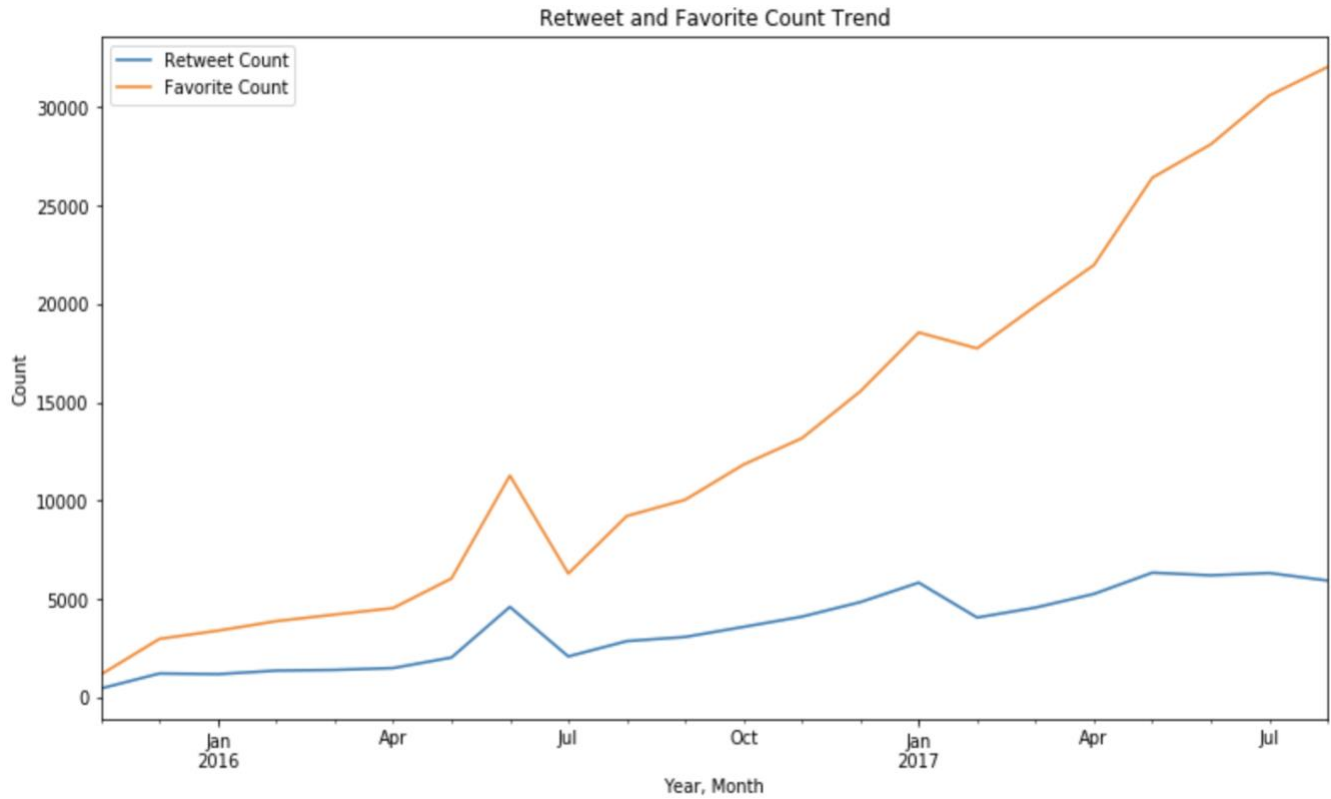


The above Pie chart represents the value counts of top 5 dog breeds in the data. Golden Retriever has the highest number of value counts (139) followed by Labrador Retriever (93).

The color-coded legend pie chart provides a great insight to see which dog breeds repeats most in the data.

- Visualization 5:

Here, I'm visualizing the mean retweet & favorite count trend throughout the WeRatedogs data. Below is the snapshot of the output.



The above line plot for mean retweet and favorite count clearly indicates that the favorite count has an upward trend as compared to the retweet count.

Also, the difference between the two is also significant. I have included month of each year in the above visualization to give a better insight.