

# WeRateDogs Twitter Archive – Act Report

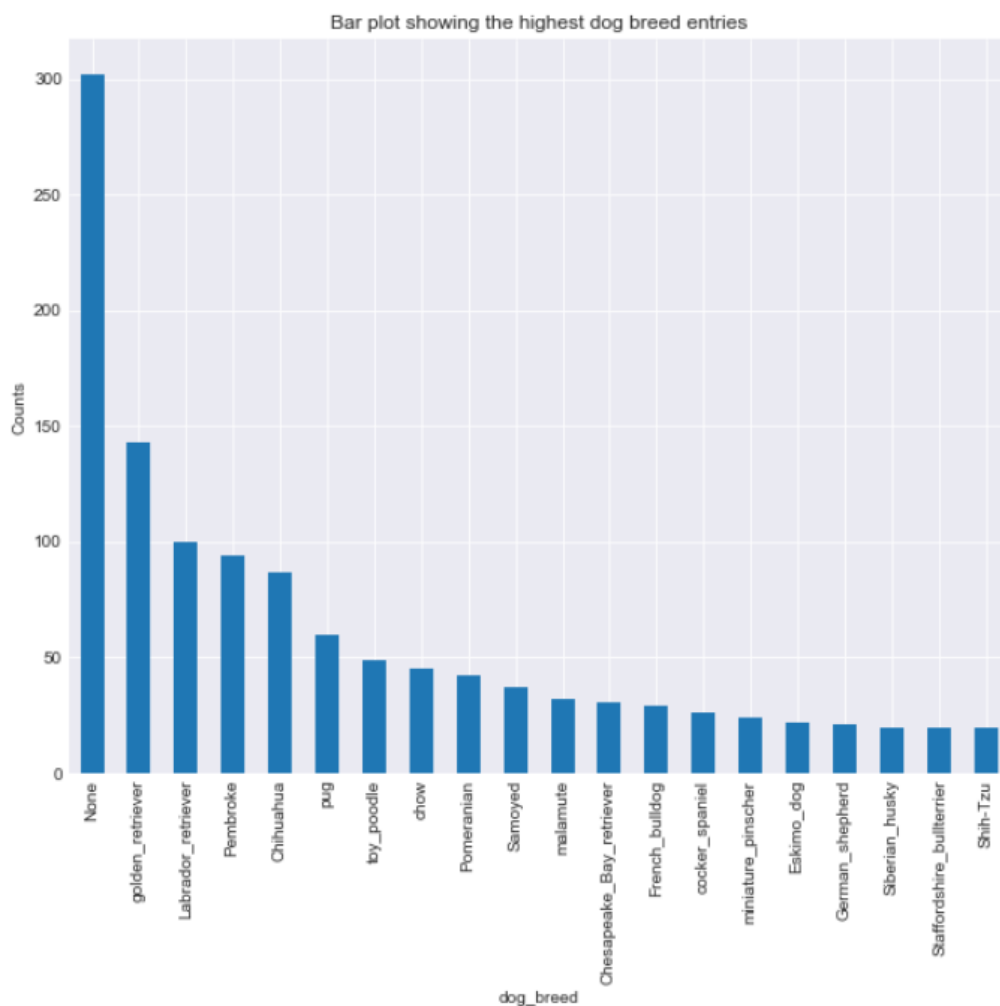
By Emmanuel Chisom Egwuonwu

## Data Analyzing and Visualization

The WeRateDogs Enhanced Twitter Archive contains about 5000+ tweets and over 2300+ tweets were successfully extracted from the twitter archive between November 15, 2015 and August 1, 2017. After assessing and cleaning the data, over 1900+ tweets were structured into one dataframe.

### Insight One – Most common dog type

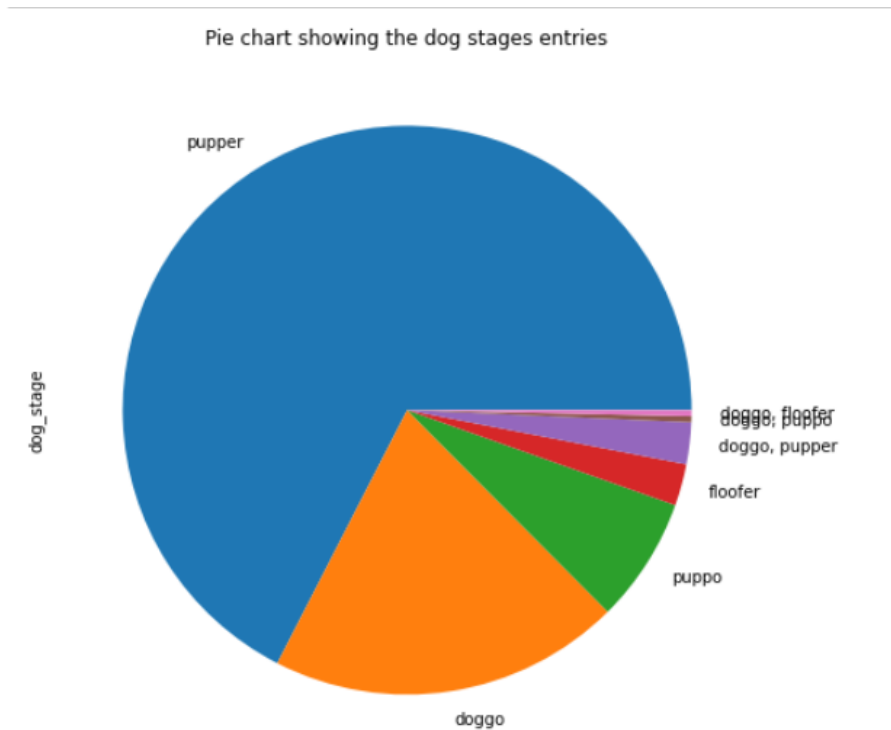
A group of dog breeds were created with value counts greater than or equal to 20. A bar plot was created to show the most common dog type.



From this visualization, it can be clearly seen that golden retriever is the most common dog breed with a value count of over 143, neglecting the 'None' entries.

### Insight Two – Most common dog stage

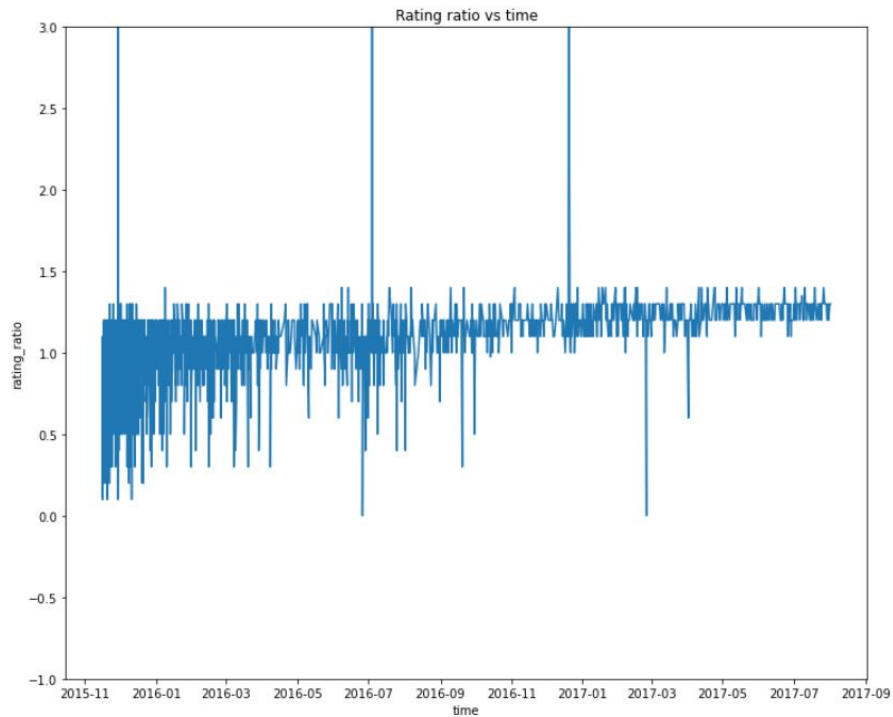
A group of dog stages were created with value counts less than 250 to eliminate the 'None' entries. A pie chart was created to depict the fractional distribution of dog stages.



From the visualization shown above, it can be clearly seen that 'pupper' has the highest fraction and therefore, it is the most common dog stage. It can also be seen from the chart that 'floofer' is the least common dog stage with a single entry. It can also be observed that some entries have double dog stages that were concatenated with a comma.

### Insight Three – Rating ratio over time

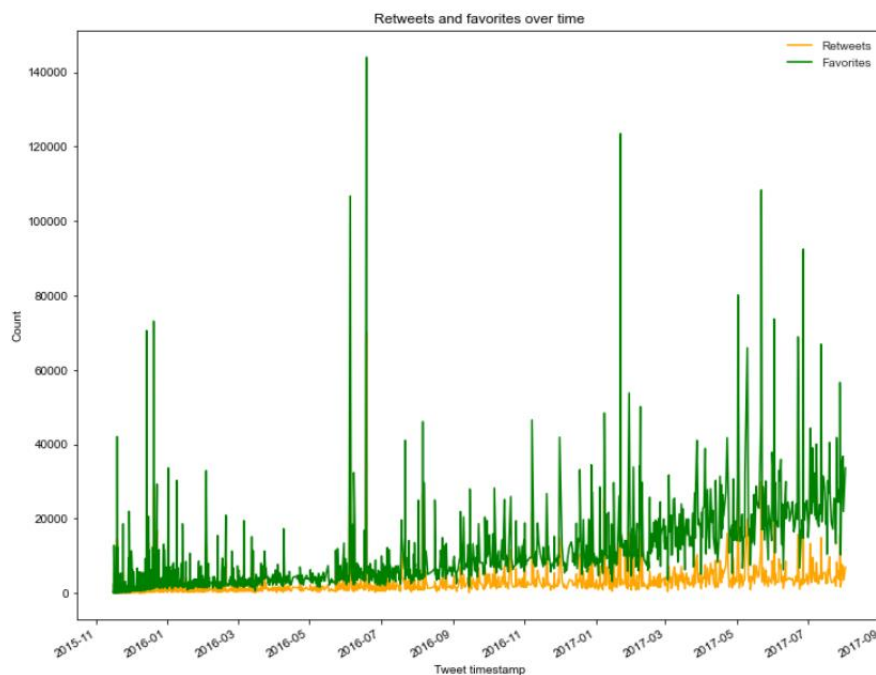
The rating ratio was gotten by dividing the rating numerator by the rating denominator. The rating ratio was plotted against time on a line plot to show the trends in rating ratio over time.



The line chart above shows the trend in the rating ratio over time. It can be observed that some tweets received 0 rating while some tweets received a very good rating.

#### **Insight Four – Favorite counts and Retweet counts over time**

The favorite count and retweet count was plotted over time to depict a trend.



From this plot, an upward trend was observed in the number of favorite and retweet counts over time. This trend indicated that the over the years, the account became more popular and therefore, an increase in the number of favorite and retweets was noticed.

These insights and visualizations include univariate plots, bivariate plots and multivariate plots.