

# Project: Wrangle and Analyze Data

By Utibe Edet Bassey

The dataset used for this project is the tweet archive of Twitter user @dog\_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. WeRateDogs Twitter account currently has 9.3 million followers and has received international media coverage. This project is limited to the tweets from 2015 to 2017.

From the 3 datasets (twitter-archive-enhanced.csv, image-predictions.tsv and tweet-json.txt) used in this project, the following insights were drawn:

1. Pupper dog stage has the highest frequency.
2. There is a positive correlation between favorite\_count and retweet\_count.
3. Of the 3 years in the dataset(2015, 2016, and 2017), 2016 has the highest frequency.
4. The favorite count increased as the year increased though their retweets dropped tremendously.

Below are the visualizations from which the insights were drawn.

## **Insight 1: Pupper dog stage has the highest frequency.**

A piechart was used to show the proportion of each dog stage. I used a function to calculate the percentage of each dog stage and appended it to the legend for visibility since the percentage values were overlapping because of the many dog stages in the chart.

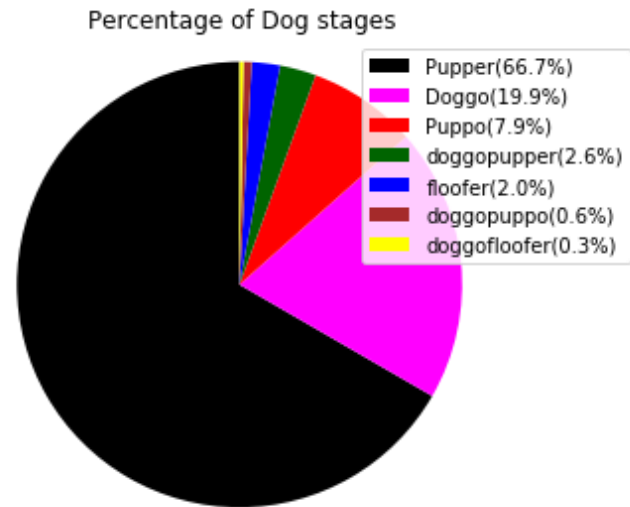


Figure 1: Piechart showing dog stages

For more visualization, a bar chart was used to show the dog stages

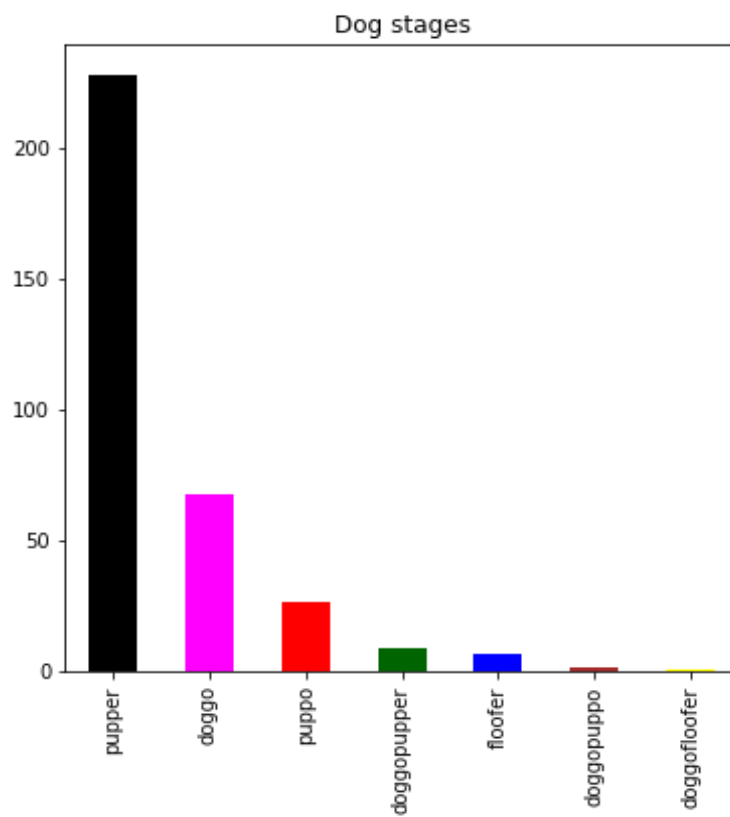


Figure 2: Barchart showing dog stages

**Insight 2: There is a positive correlation between favorite\_count and retweet\_count.** The scatter plot depicted below was used to determine the relationship between favorite\_count and retweet\_count.

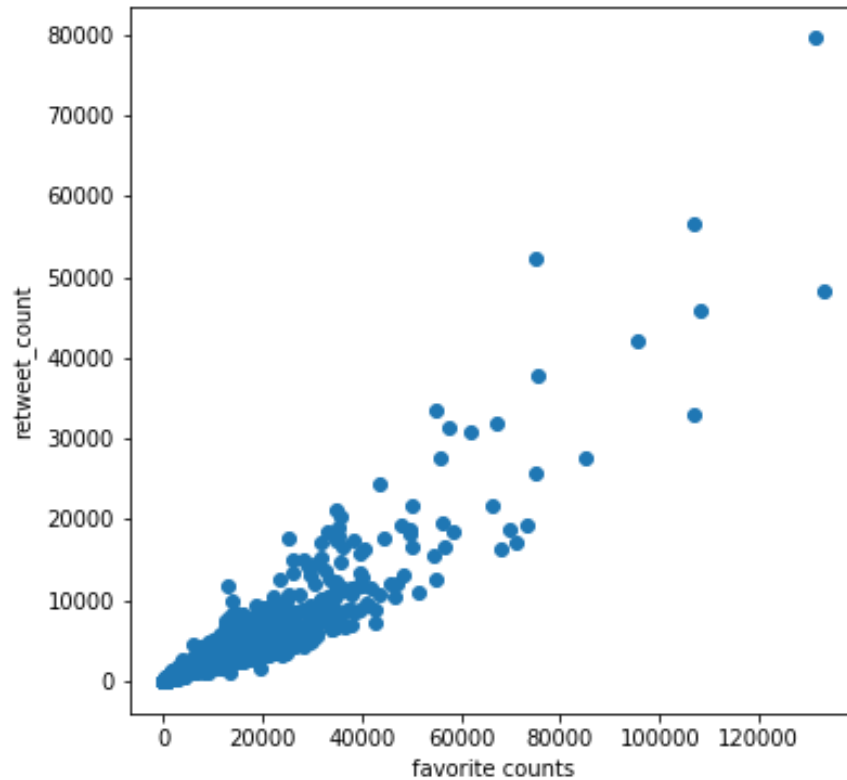


Figure 3: Scatterplot showing the relation between favorite count and retweet count

**Insight 3: Of the 3 years in the dataset (2015, 2016, and 2017), 2016 has the highest frequency.** A histogram was plotted to determine the highest frequency and is showed in figure 4 below.

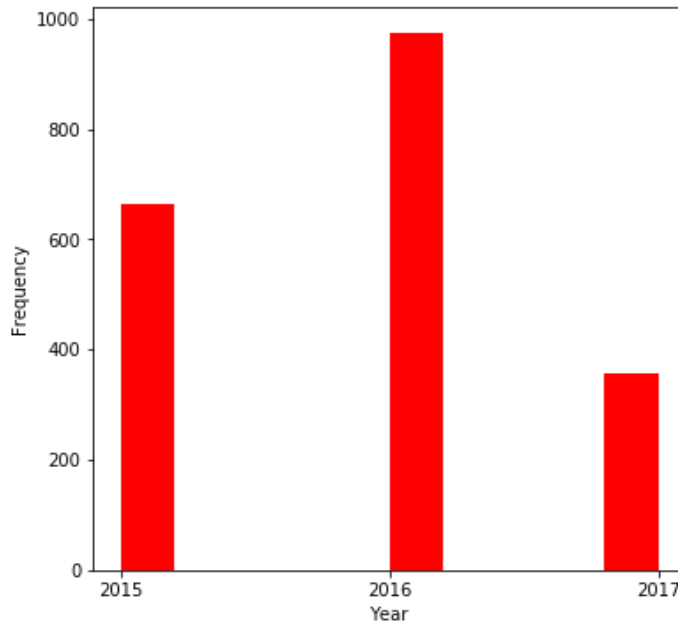


Figure 4: Histogram for year column

**Insight 4: The favorite count increased as the year increased though their retweets dropped tremendously from year 2016 to 2017.** Bar charts of year versus favorite count and year versus retweet count were plotted to determine this insight and are represented in figures 5 and 6 respectively.

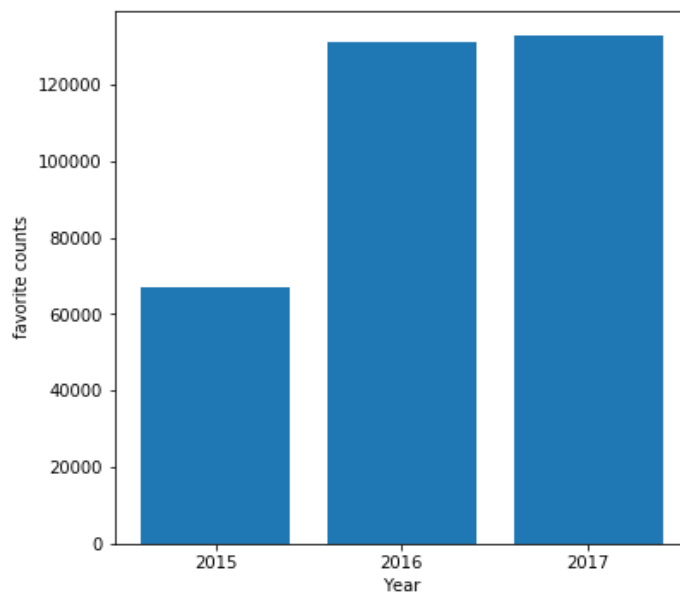


Figure 5: Bar chart of year versus favorite count

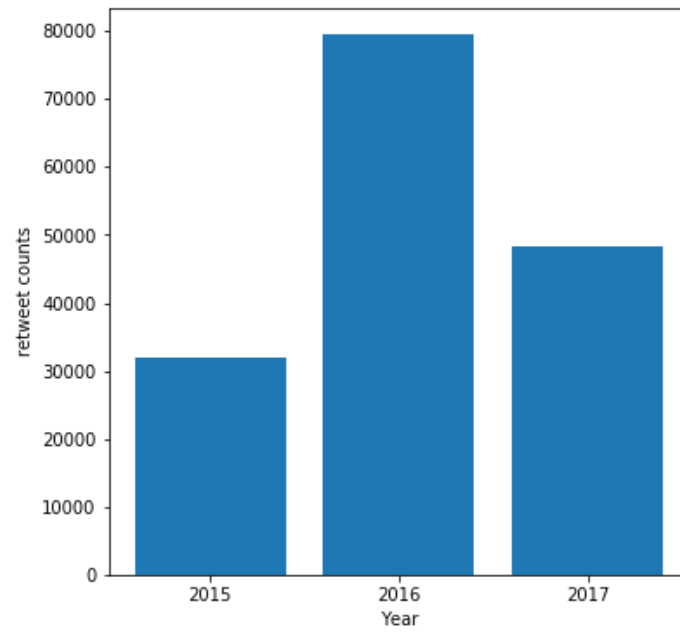


Figure 6: Bar chart of year versus retweet count