

We Rate Dogs – Udacity’s Data Wrangling Project

Introduction:

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. WeRateDogs assigns a score out of 10 for the dog (like showed in Figure 1 below). The scope of this project is to gather the data, clean the data and analyze the data. The data is gathered and cleaned using standard Python libraries. The data is analyzed and visualized using standard Python libraries and Tableau.



Data Analysis and Visualization:

1. Distribution of Rating for the different dog stages

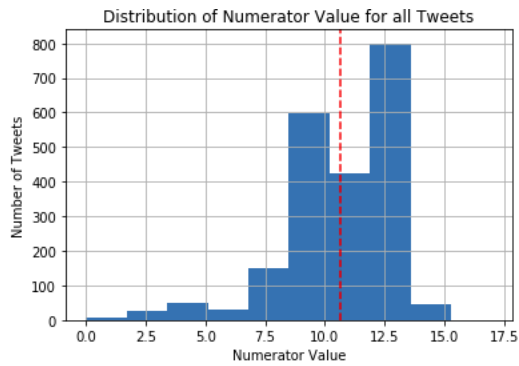
The dogs have been classified into the following dog stages.

- Pupper
- Floofer
- Puppo
- Doggo
- None

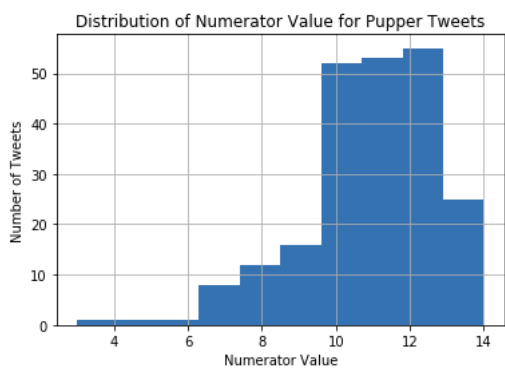
a) All dog stages

The dogs have been scored between 0 and 17. The average score is 10.63.

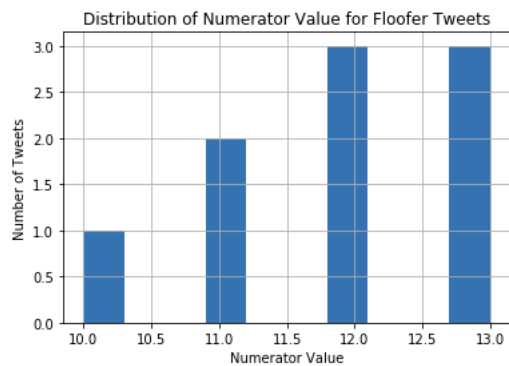
We Rate Dogs – Udacity’s Data Wrangling Project



b) Dog stage: Pupper

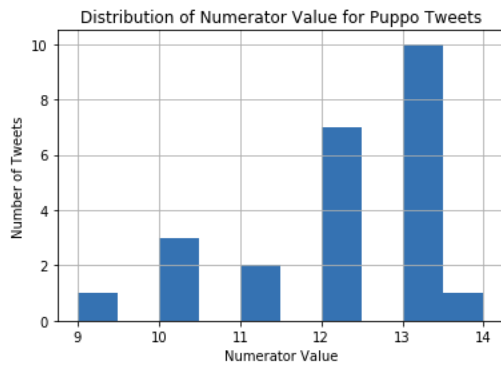


c) Dog Stage: Floofer

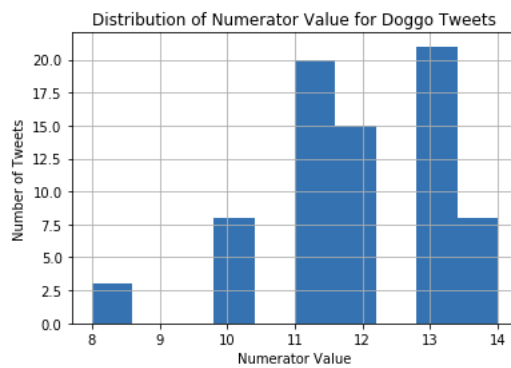


d) Dog Stage: Puppo

We Rate Dogs – Udacity’s Data Wrangling Project

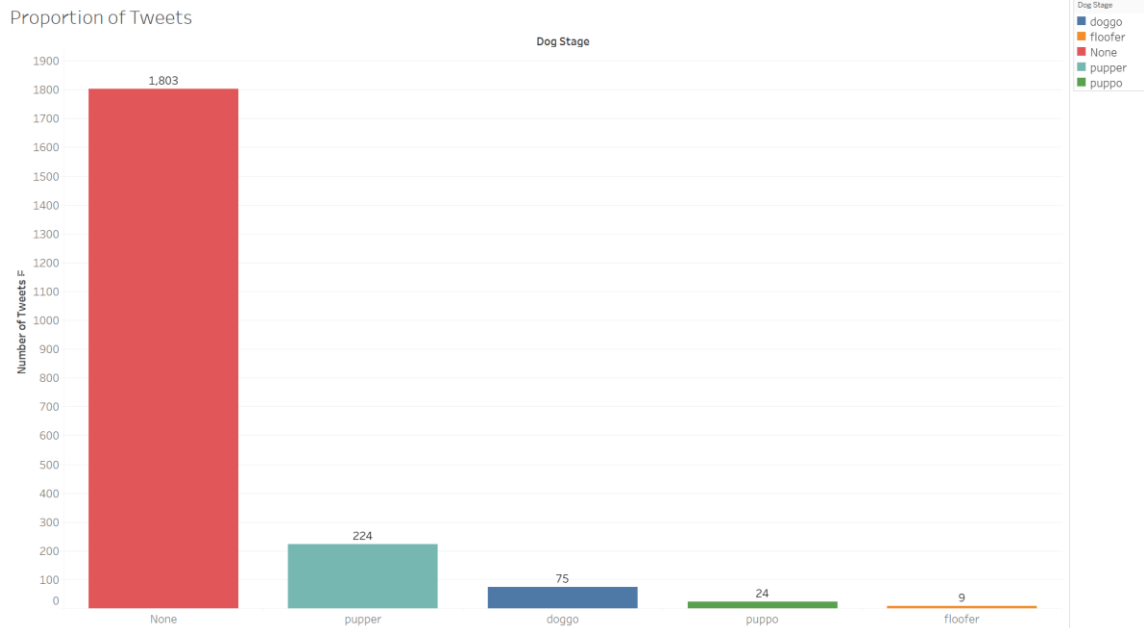


e) Dog Stage: Doggo



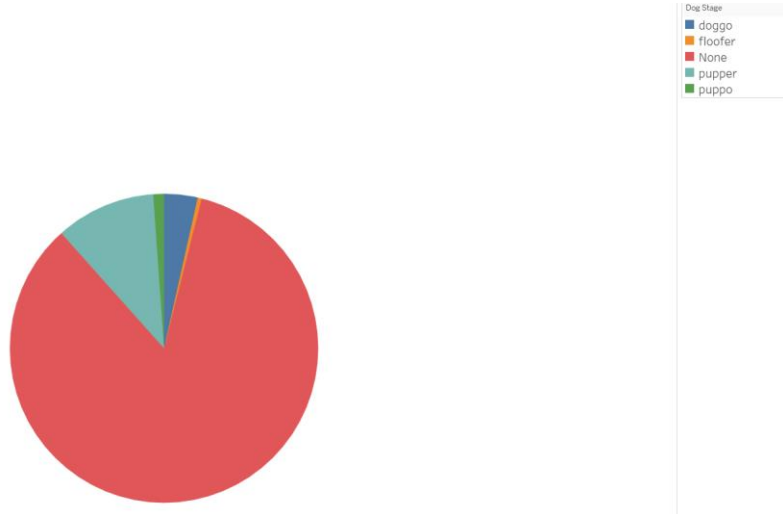
2. Proportion of tweets belong to each dog stage

Majority of the tweets have not been classified into a specific dog stage.



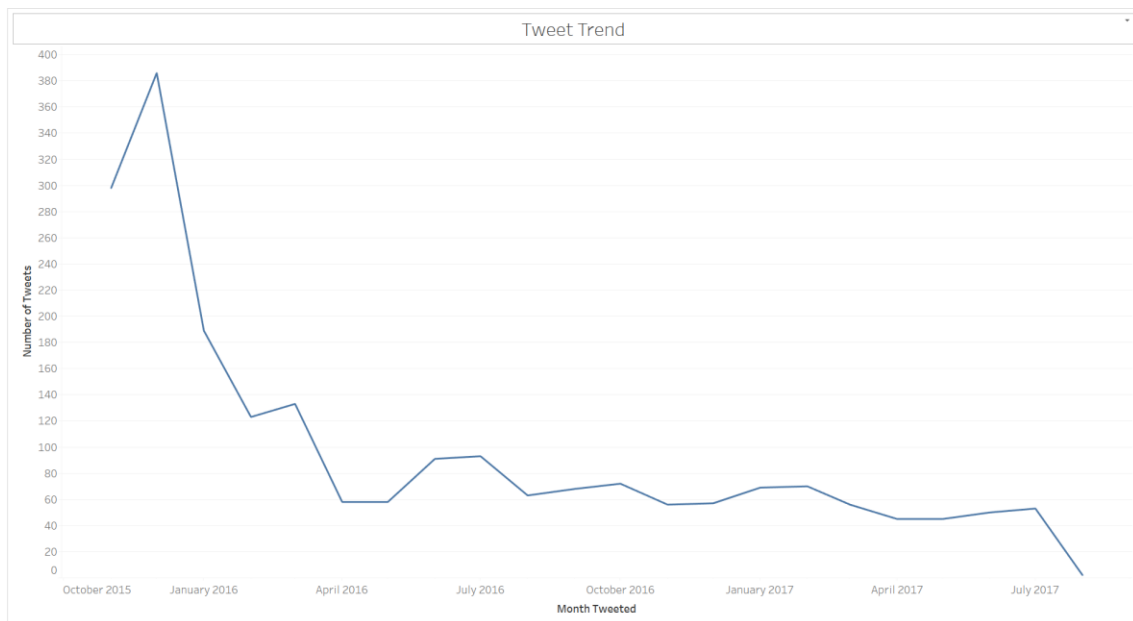
We Rate Dogs – Udacity’s Data Wrangling Project

Percentage of Tweets per Dog Stage



3. Tweet Trend

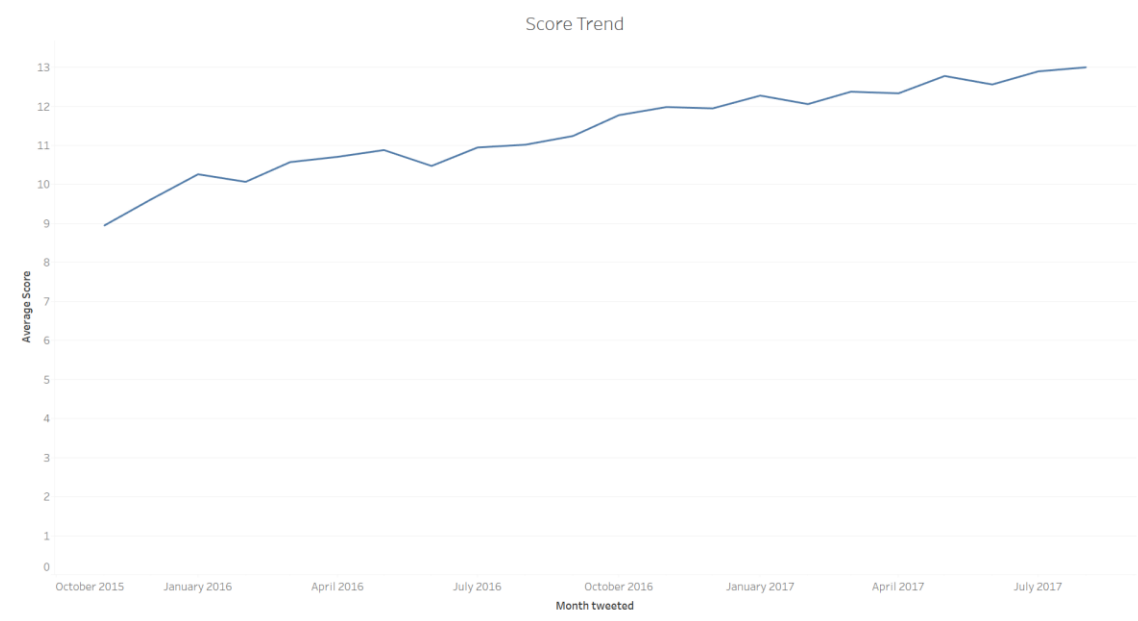
The number of Tweets coming in has reduced considerably from October 2015 to August 2017. The plot shows the number of tweets per month over the period.



4. Score Trend

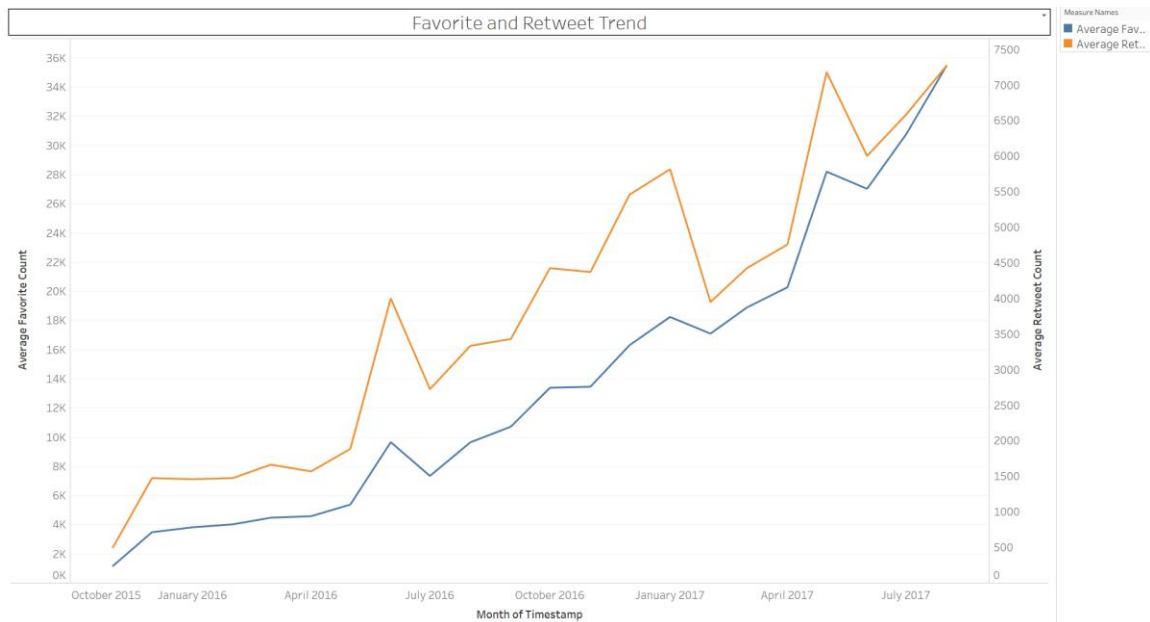
The average rating score shows as increasing trend over time. The plot shows the average score of the tweets per month.

We Rate Dogs – Udacity’s Data Wrangling Project



5. Favorite and Retweet Count Trend

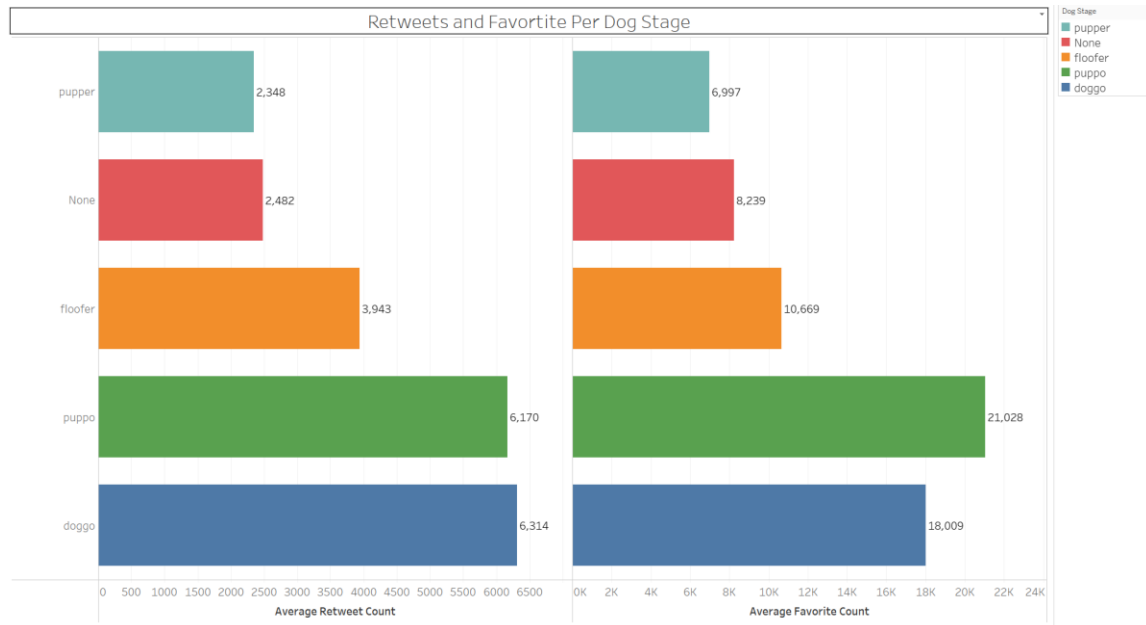
The number of Tweets that have been marked as favorite and retweeted has increased over time.



6. Favorite and Retweet Count per Dog Stage

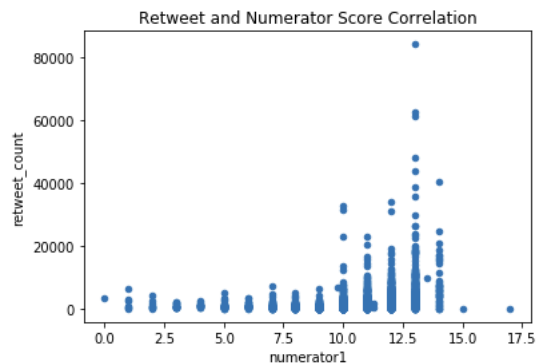
Doggo and Puppo dog stages have high retweet and favorite count. These seem to be the popular dog stages.

We Rate Dogs – Udacity’s Data Wrangling Project

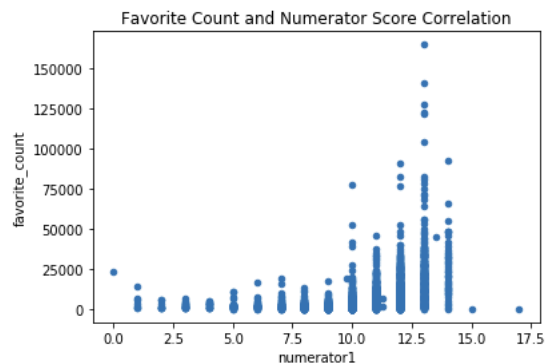


7. Correlation Analysis

a. There is a weak positive correlation between score received and the number of retweets.

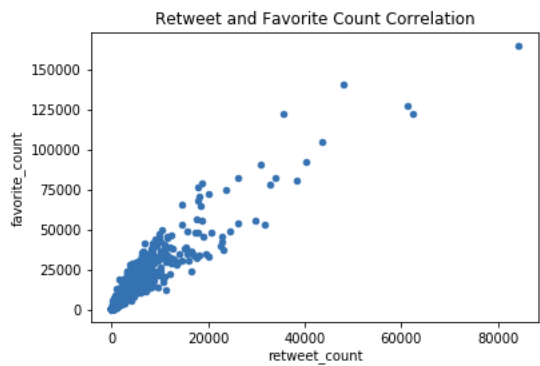


b. There is a weak positive correlation between score received and the favorite counts.



c. There is a strong positive correlation between favorite counts and retweets.

We Rate Dogs – Udacity’s Data Wrangling Project



d. Correlation matrix

	rating_denominator	favorite_count	retweet_count	numerator1
rating_denominator	NaN	NaN	NaN	NaN
favorite_count	NaN	1.000000	0.926698	0.376106
retweet_count	NaN	0.926698	1.000000	0.287346
numerator1	NaN	0.376106	0.287346	1.000000

Author: Ajith Sharma