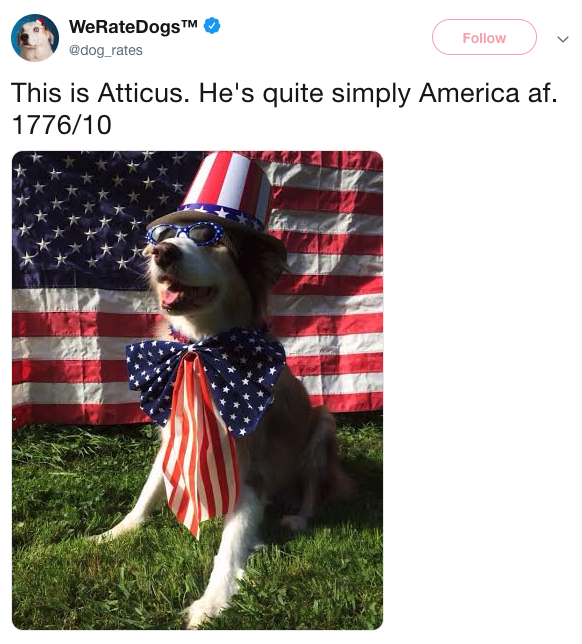
**Name:** Alaa Zarban

# Insights, analysis and visualizations

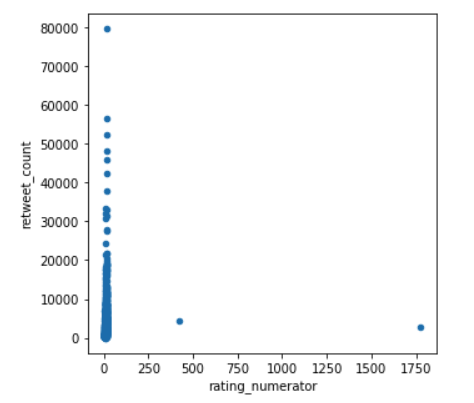
We started our analysis and visualizations by defining the scope of our analysis. This can be achieved by defining the column of interest that we need to focus on in order to reach our goal and clarity. I had couple of goal and concerns to be answered by the data provided. First, the relation between retweets and dog rating. I had a theory that people give high retweets to dogs with high rating more than the low dog rating. Moreover, a scatter plot has been plotted between retweets and rating\_numerator to show the relation. Surprisingly, two rating outliers has been found as a result of this scatter plot. The outliers tweets is shown below:



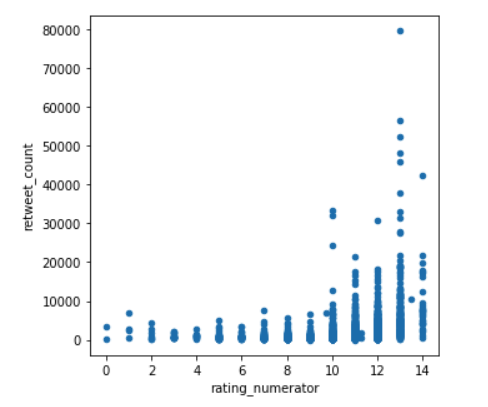


In order to make the scatter plot clearer, the two outliers has been dropped.

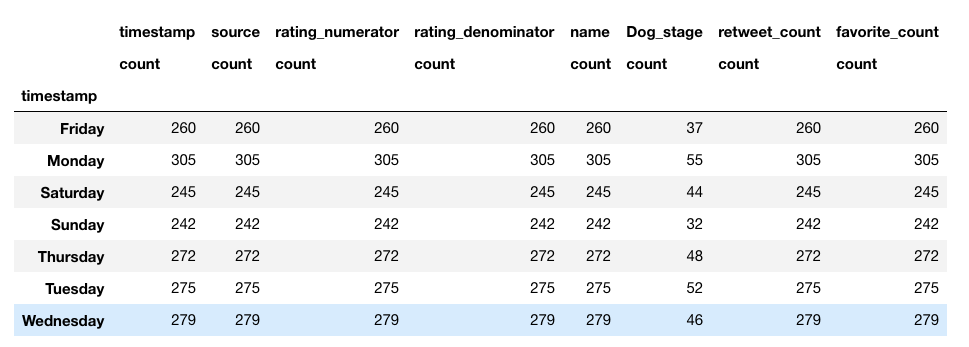
Before deleting outliers.



After cleaning



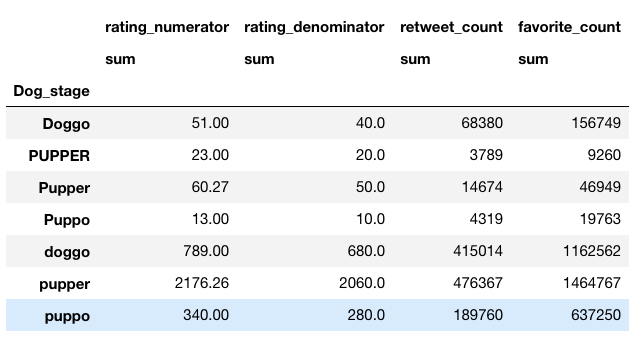
Then, I checked the max values of couple of columns. The highest retweet and favorite counts are 79,515.00 and 132,810.00 respectively. Moreover, I checked the tweet counts of WeRateDogs account broken down by day-of-week. And the result was as follows:



The most active day Monday.

I wanted to analyze the dog stage rating, retweet and favorite totals. As a result, we will see which dog stage is most popular and highest in rating across the data.

Unfortunately, we found another issue that needs to be cleaned. Upper and lower case duplicate records of Dog\_Stage column. Let's correct that and unify that to lower-case.



After lower-casing all dog stage values, the winner went to Puppers.

