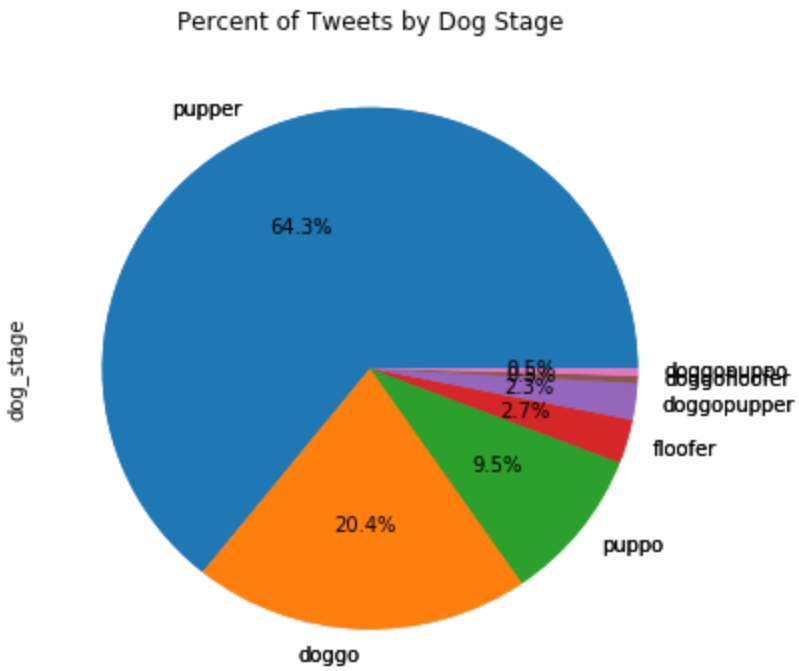
After gathering, assessing, and cleaning the data for the rating dogs twitter page, I have analyzed it in various ways. Through my analysis, I have come upon three key takeaways. First, I calculated the minimum, maximum, and mean of the rating(described in the data as rating_numerator) as grouped by the stage of the dog, disregarding the denominator and only focusing on the numerator. Although a significant portion of the tweets are not grouped into a stage (aka ‘None’ in the data), separating it by dog stage provides some meaningful results. Particularly, the mean of all the stages are relatively the same, ranging from 11 to 13. The minimums and maximums range significantly more, but here we are just looking at the means. The average, or mean, of all the ratings is relatively the same across dog stages, meaning that the owner of the twitter handle does not have a particular favorite dog stage and does not skew the ratings favoring one stage over another.

The second particularly interesting analysis I made was aggregating the number of favorite tweets and retweets when grouped by dog stage. We see that when grouped by dog stage, there is a difference between average number of favorites and retweets. Specifically, we see that the stage of ‘pupper’ has the lowest average, with ‘None’ being the second lowest in both favorites and retweets. There could be a couple of reasons for this. First, the stage ‘pupper’ and ‘none’ account for the majority of the tweets, giving them more opportunity to be correctly averaged, whereas the other categories have significantly fewer tweets, so they could be skewed incorrectly. Or, the categories provide significant data on what will achieve higher favorite and retweet counts. There would need to be more data, specifically in the other categories for a true decision to be made here.

The third analysis I made was in regards to the amount of images associated with each tweet. I calculated the minimum, maximum, and average amount of favorites the tweets got and grouped by amount of images associated. The result allows us to see that the minimum number of favorites will actually go up with the amount of images. However, a similar caveat to the preceding point needs to be made in that there is far less data for tweets with 2, 3, and 4 images.

In the visualization below, I have produced a pie chart which shows the percentage of tweets by the dog stage, not including the ‘none’ category. As is evident, the pupper category has the largest share, follower by doggo, puppo, floofer, and some combination categories.


Percent of Tweets by Dog Stage
dog_stage
pupper
64.3%
doggo
20.4%
puppo
9.5%
floofer
2.7%
doggopupper
2.3%
doggofloofer
0.5%
doggopuppo