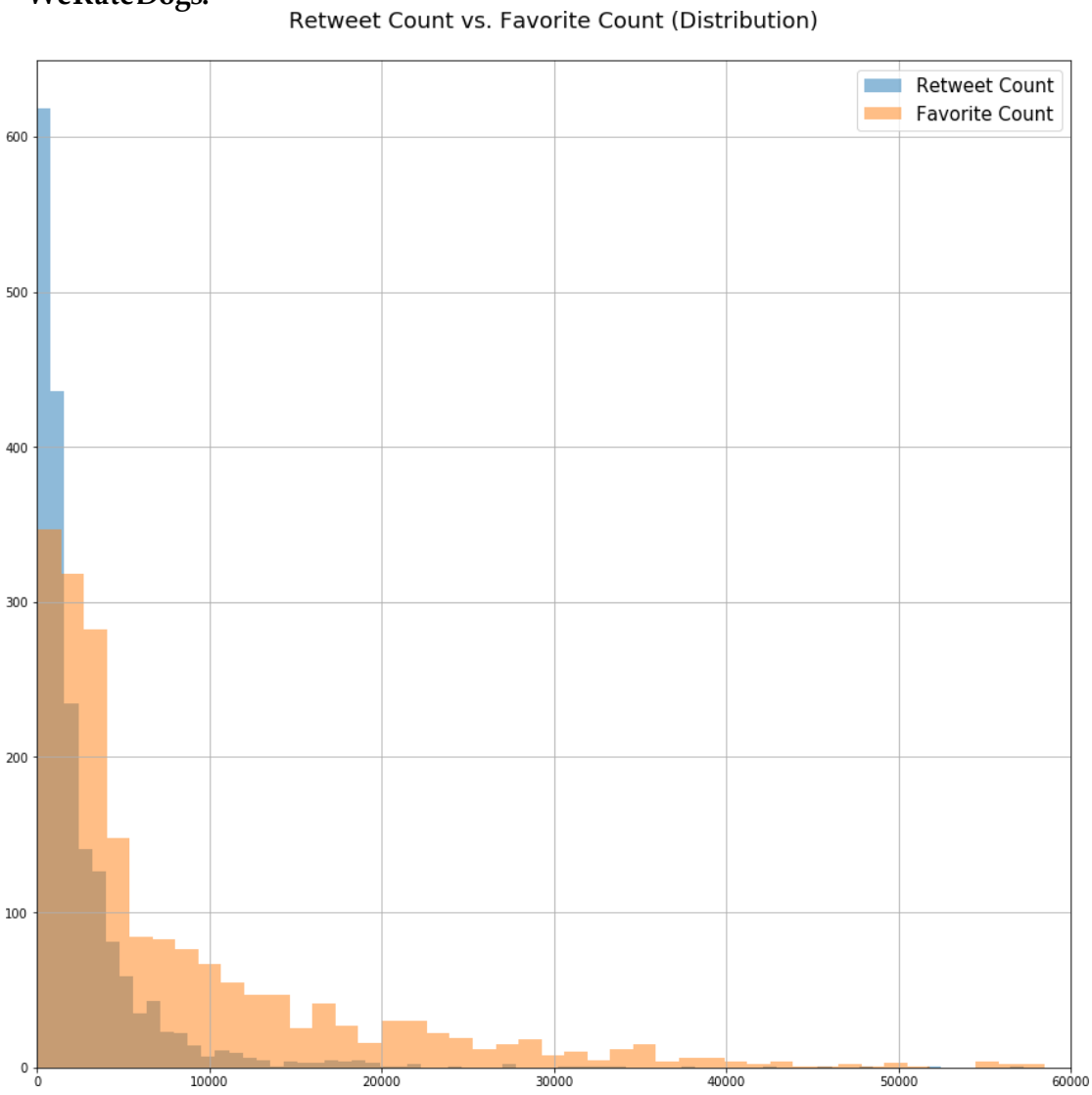
Act Report

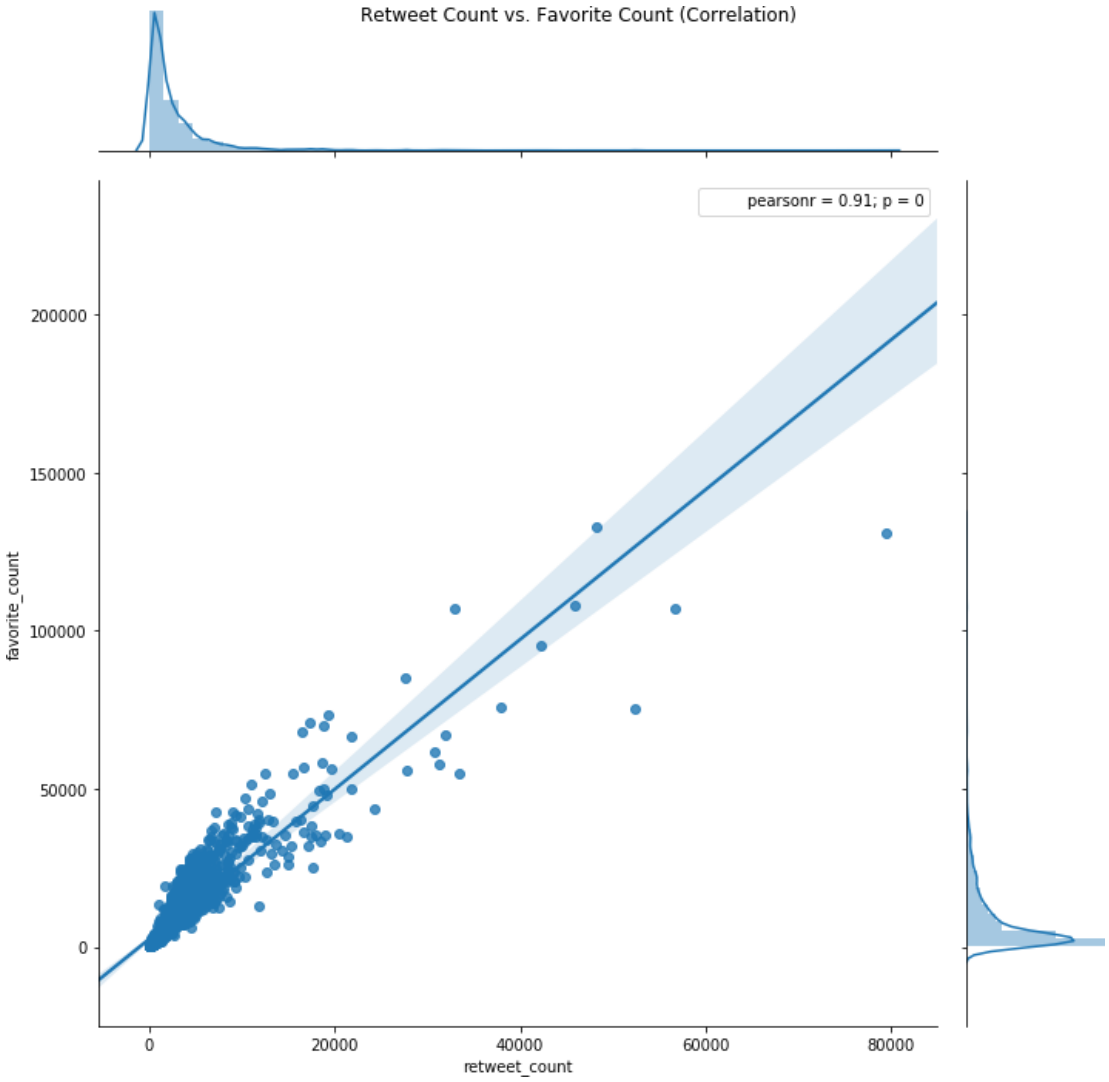
October 2, 2022

0.0.1 Introduction:

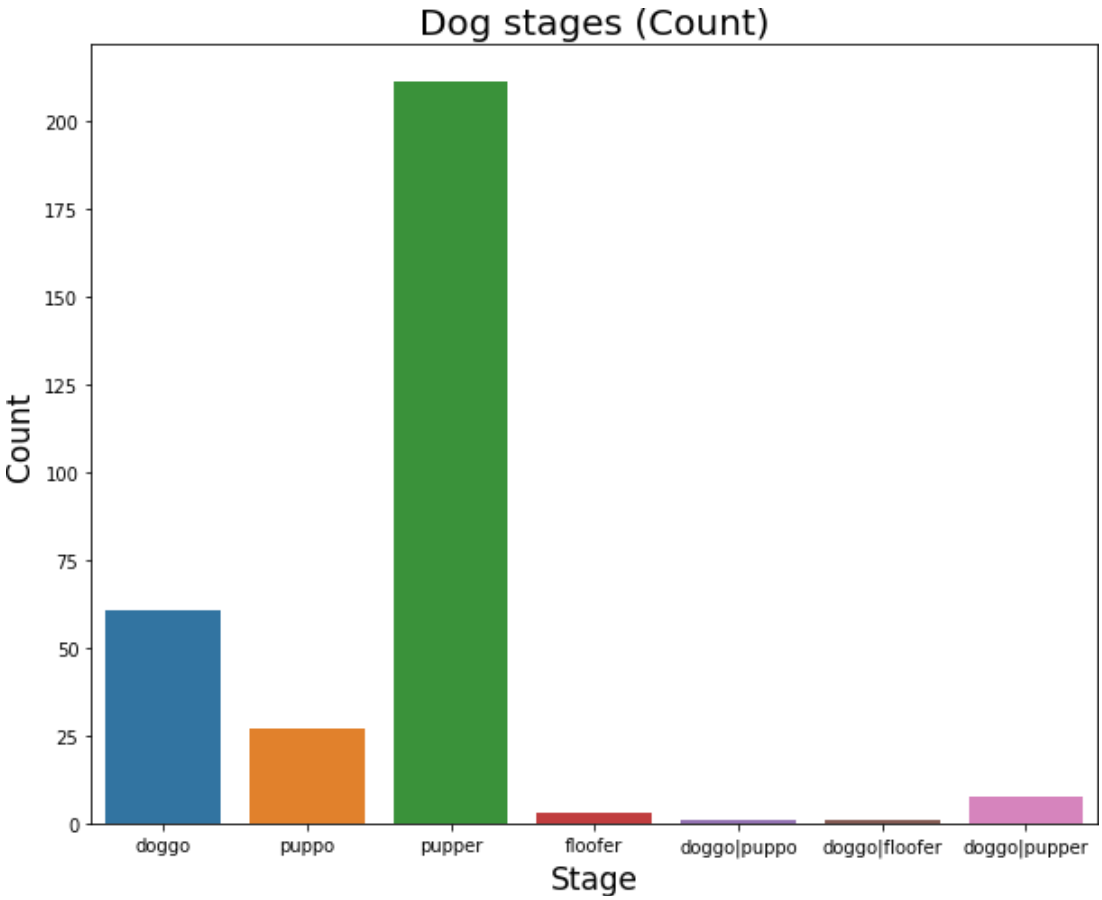
This report aims to show some insights from analyzing data of a Twitter account named WeRateDogs.

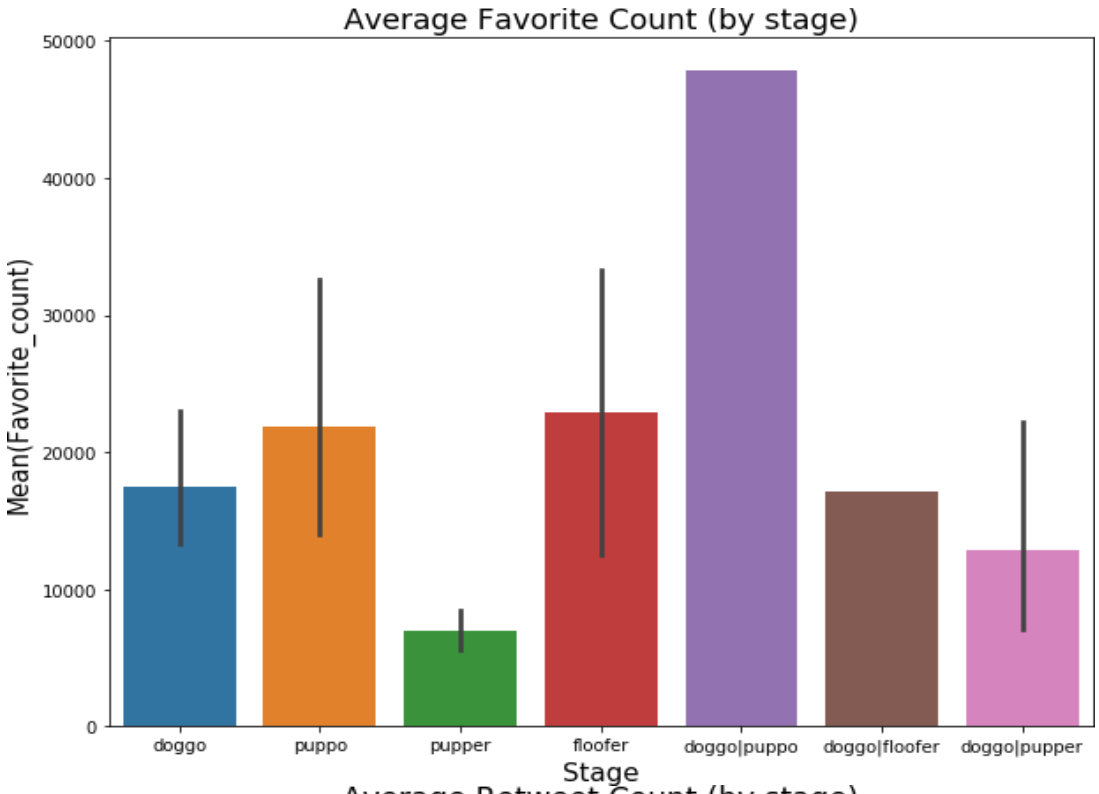


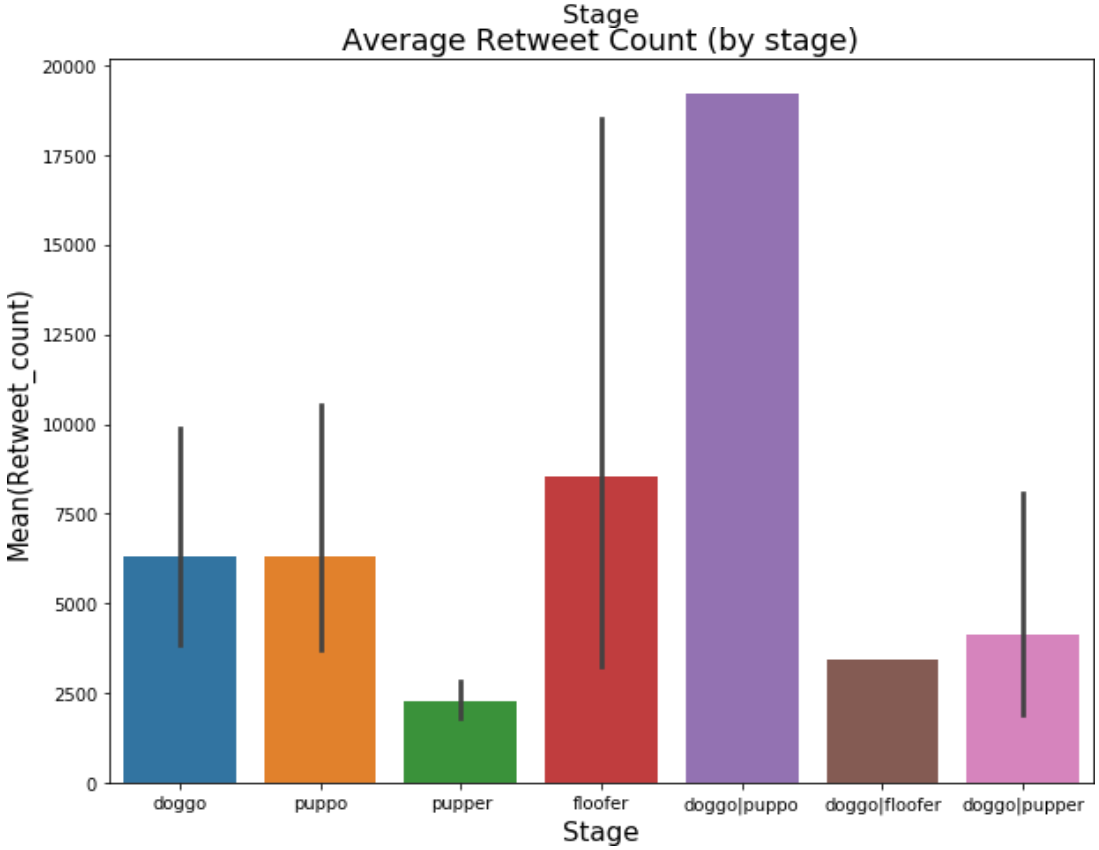
Looking at the histogram above, we note that the distribution of both retweets and favorite is significantly right-skewed, however, the mean of favorite count (8895.7) is greater than retweet count (2766.7). We also note that the maximum number of favorites a tweet received was 132,810, where the maximum number of retweets a tweet received was 79,515.



Looking at the plot above, we note there is a positive correlation between the favorite_count and the retweet_count. In other words, as the number of retweets increases and the number of favorites increases as well, which is expected as most people who would mark a tweet as favorite are likely to retweet that tweet.

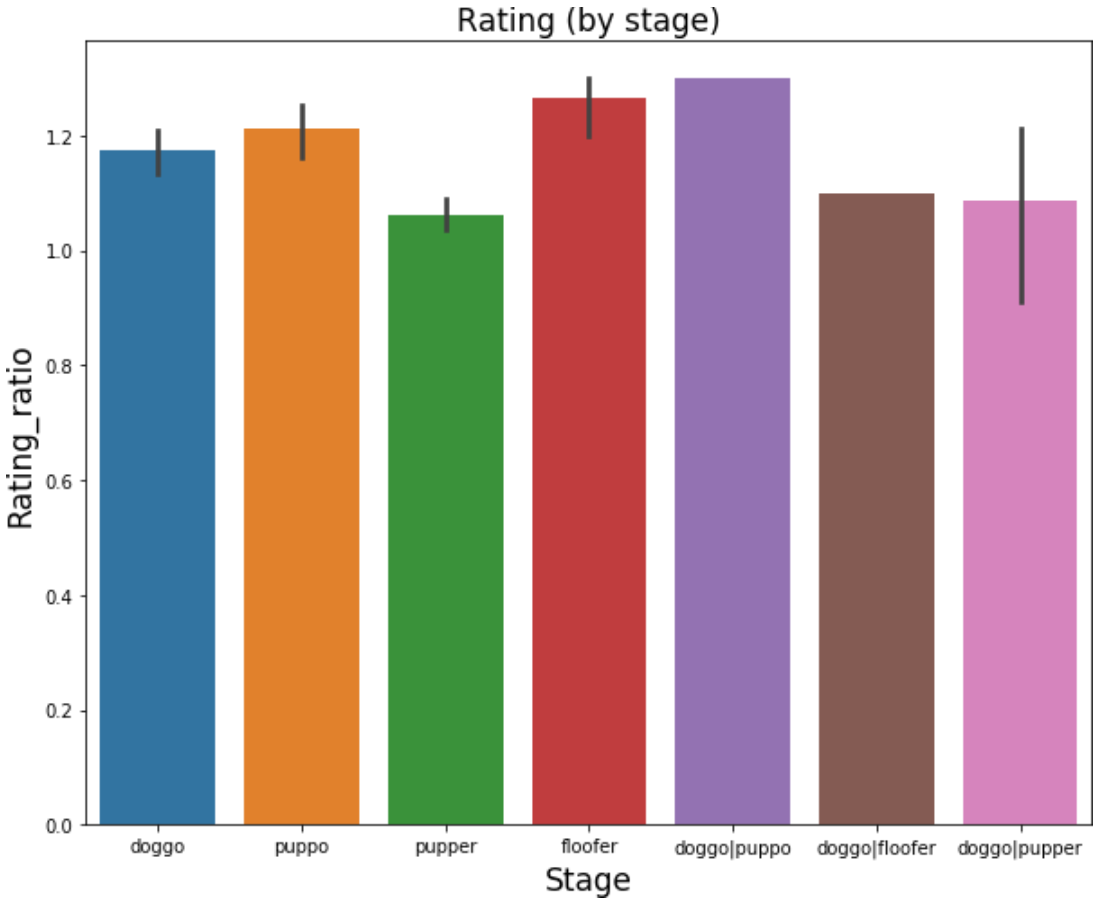



Average Favorite Count (by stage)
Mean(Favorite_count)
50000
40000
30000
20000
10000
0
doggo
puppo
pupper
floofer
doggo|puppo
doggo|floofer
doggo|pupper
Stage

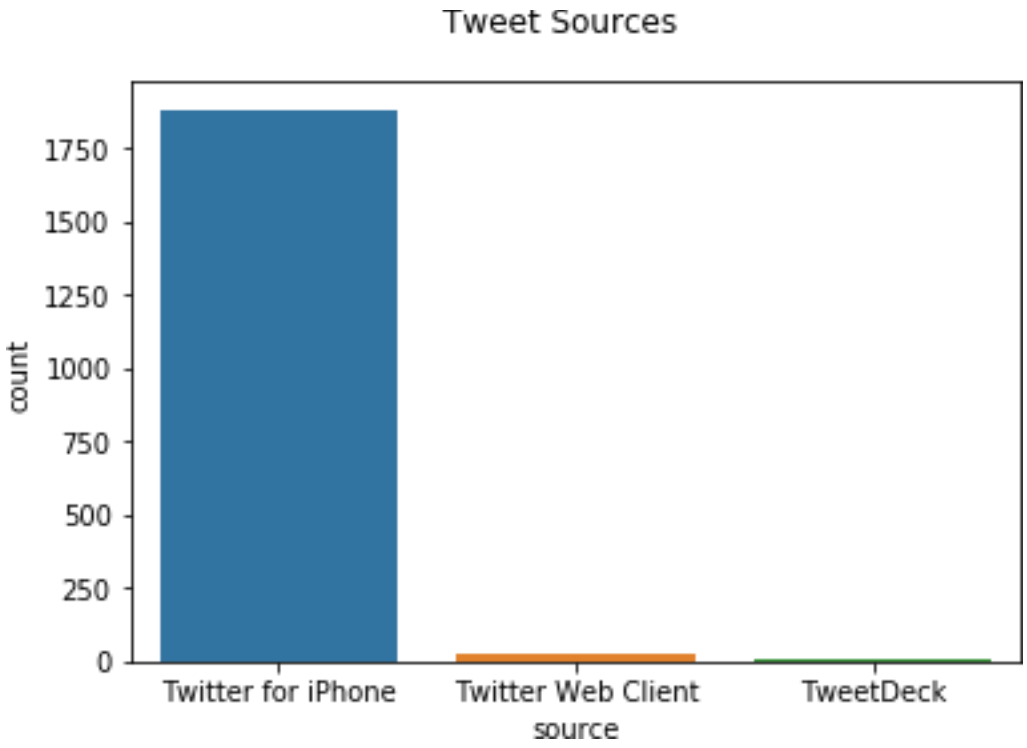

Average Retweet Count (by stage)
Mean(Retweet_count)
20000
17500
15000
12500
10000
7500
5000
2500
0
doggo
puppo
pupper
floofer
doggo|puppo
doggo|floofer
doggo|pupper
Stage

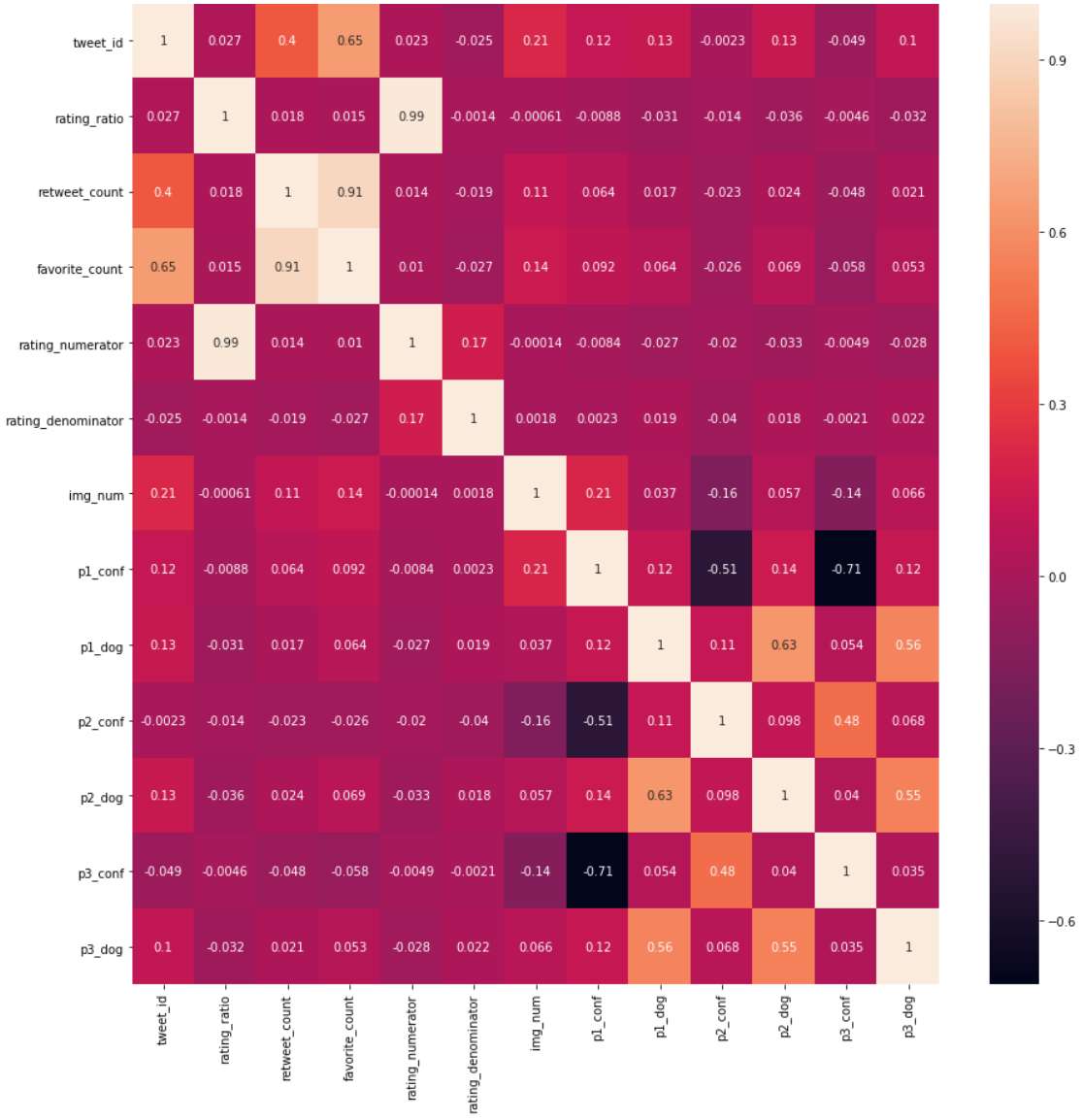
Looking at the plots above, we note the following:

1. The pupper stage has the highest count where the (doggo|floofer) and (doggo|puppo) has the lowest count.
2. The average retweet_count for (doggo|puppo) is the highest (greater than 17,500 retweets), where the average retweet_count for the pupper is the lowest (less than 2,500).
3. Similarly, the average favorite_count for (doggo|puppo) stage has the highest average favorit_count, where the pupper has the lowest average favorite count.



Looking at the plot above, we note the doggo|puppo stage has the highest average rating where the pupper stage has the lowest rating.


Tweet Sources
1750
1500
1250
1000
750
500
250
0
count
Twitter for iPhone
Twitter Web Client
TweetDeck
source



Looking at the heatmap above, we note the following:

1. There is a strong positive correlation (0.92) between `favorite_count` and `retweet_count` which support our observation above.
2. There is a strong positive correlation (0.92) between `rating_ratio` and `rating_numerator` which is given (rating_ratio is based on the rating_numerator and rating denominator)
3. There is a relatively strong negative correlation (-0.71) between `p3_conf` and `p1_conf` which is interesting as different predictions may have opposing confidence level