

Act Report – Analysis of twitter data after data wrangling

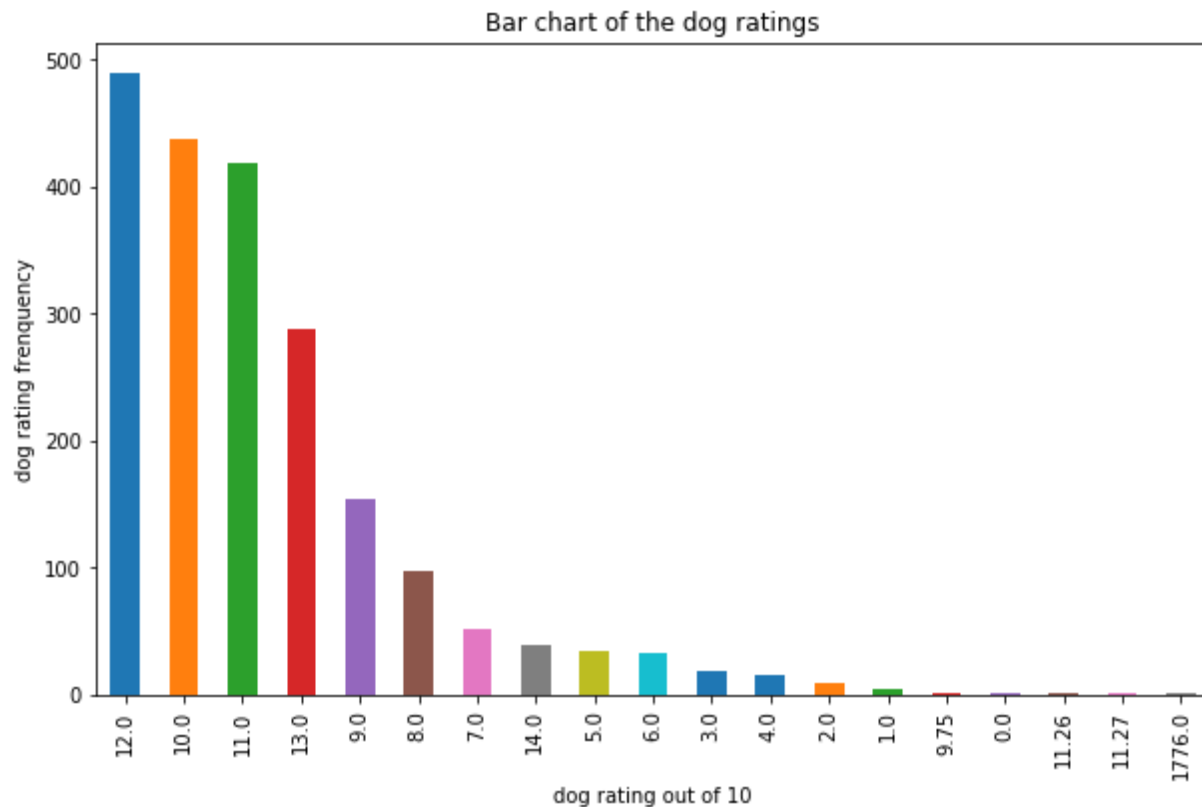
This report captures the summary of the analysis, the actual code for the analysis can be found in the `wrangle_act.ipynb` file

Insight 1: Basic statistics

	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count	img_num	p1_conf	p2_conf	p3_conf
count	2.092000e+03	2092.000000	2092.0	2092.000000	2092.000000	1665.000000	1665.000000	1665.000000	1.665000e+03
mean	7.365071e+17	11.454245	10.0	2843.900096	8989.270554	1.214414	0.605734	0.137438	6.129107e-02
std	6.710258e+16	38.657301	0.0	4714.185842	12212.636407	0.574108	0.266392	0.101625	5.193659e-02
min	6.660209e+17	0.000000	10.0	16.000000	81.000000	1.000000	0.044333	0.000010	2.160900e-07
25%	6.768177e+17	10.000000	10.0	643.750000	2044.750000	1.000000	0.379055	0.054790	1.593830e-02
50%	7.097095e+17	11.000000	10.0	1403.000000	4191.500000	1.000000	0.606747	0.121377	4.990060e-02
75%	7.874779e+17	12.000000	10.0	3286.500000	11455.750000	1.000000	0.853170	0.199396	9.466370e-02
max	8.924206e+17	1776.000000	10.0	79515.000000	132810.000000	4.000000	0.999984	0.467678	2.734190e-01

Some important numbers to be pointed out are: the maximum rating numerator is 1776, the maximum number of retweet is 79515 and the maximum number of likes are 132810 as indicated by the `favorite_count`. The standard deviation for the rating denominator is 0 which indicates that all records in the dataset has a denominator of 10, as intended.

Insight 2: What are the dog rating distributions look like and what's the most frequent rating?



As indicated in this bar chart, the most frequent rating given is 12/10 and the majority of the ratings are between 7-14.

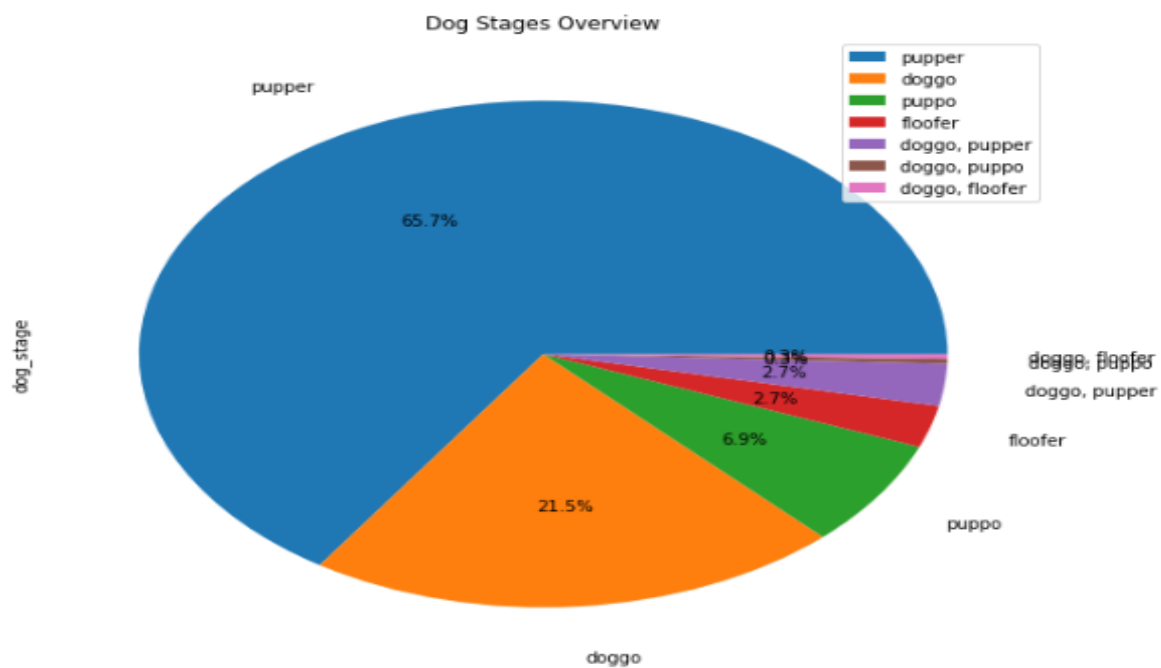
Insight 3: what is the highest rating of the dog out of 10 and what does the dog look like?

The highest rating given is 1776/10 and it is given to Atticus, a very cute puppy indeed 😊

This is Atticus. He's quite simply America af. 1776/10

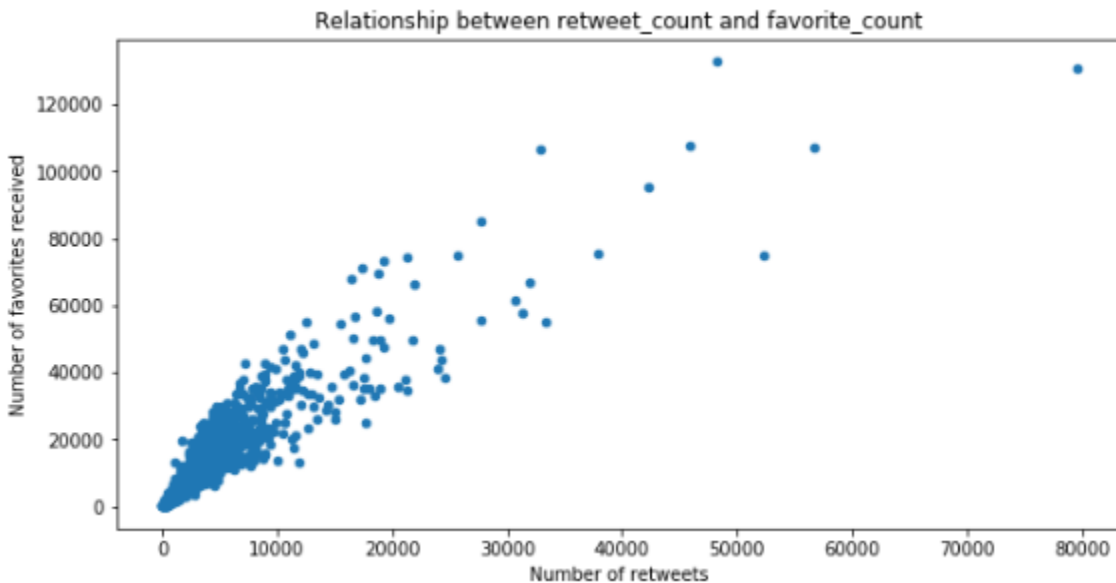


Insight 4: What is the most frequent dog stage being rated



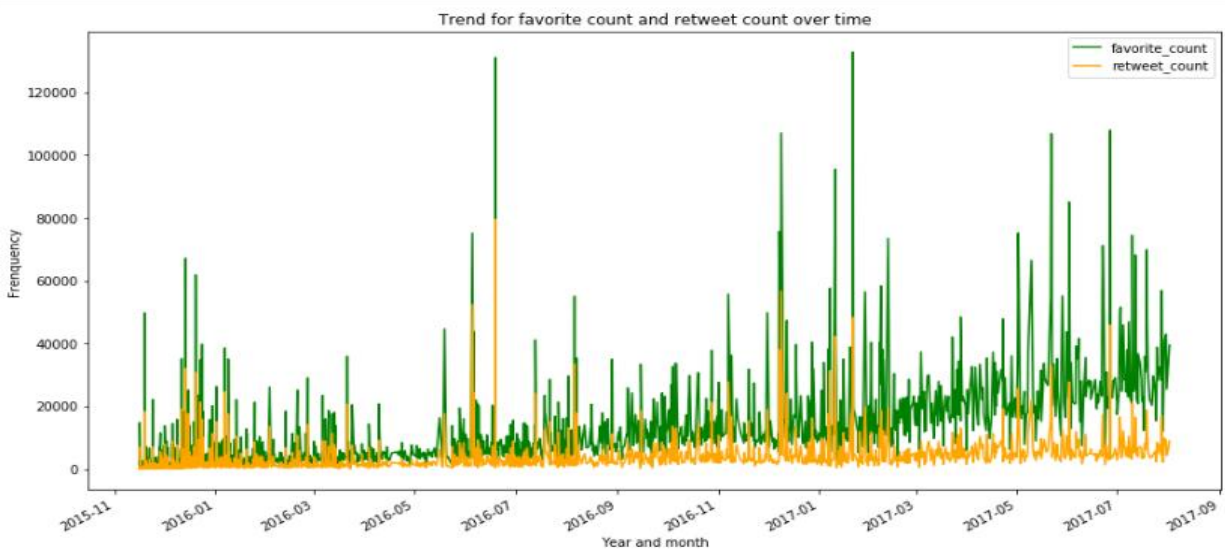
Evidently, the pie chart suggests that the most frequent dog stage that appeared in the twitter rating is pupper with a total of 65.7% followed by doggo 21.5% and puppo 6.9% and floofer being the lowest 2.7%. There are also dogs in transition stages and those have even lower percentage.

Insight 6 Is there a correlation between retweet_count and favorite_count?



The scatter plot shows a clear positive relationship between number of retweets and number of likes received and this observation is logical as well. As more people retweet the content, there is a higher number of audience accessing the tweet and thus a higher chance of being liked.

Insight 7: Is the twitter account more and more popular over time as indicated by the retweets and favorite counts?



There is a clear increasing trend for the number of likes this twitter accounts received over time. The trend for the number of retweets seems to be not so obvious but there is a slight increasing trend can be seen. There are notable overlaps for a few peaks suggesting that when we have high number of retweets we also have high number of likes, and this observation confirms the positive correlation found previously between retweet count and favorite count.