

Act Report

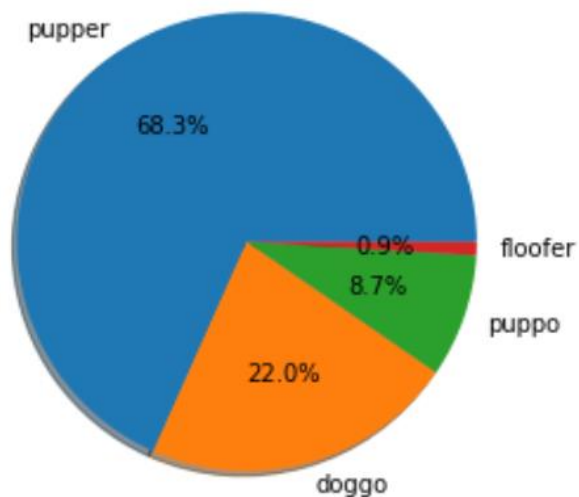
Dog Stages Popularity

What are the most popular dog stages ? However are they rated by tweeters ? Which types have the most favorite counts and retweet counts ?

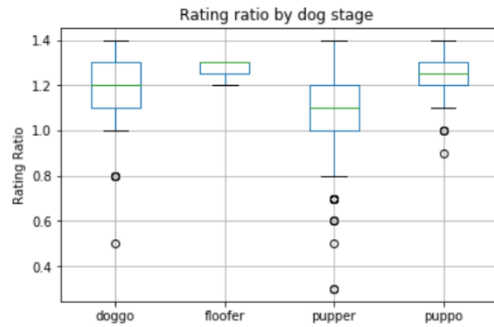
From the different analysis performed below, we can note when the dog stage description is available:

- Puppies are the most popular stage type among tweet poster (Pupper: 68%, Doggo: 22%, Puppo: 9%, Floofer:1%)
- Maybe due to a larger number of posts, puppies and doggos rating ratio means tend to be lower than puppo and floofer
- The observation is the same for favorite counts and retweet counts. The favorite counts and retweet counts means are lower for pupper, the boxplot is also smaller despite outliers
- Based on the descriptive statistics available, we would say that Puppo and Floofer would have relatively better / stable ratings from the tweeter author with a ratio around 1.2 / 1.3. And maybe because these dog stages are less frequent on dog_rates, they have in average more favorite and retweet counts

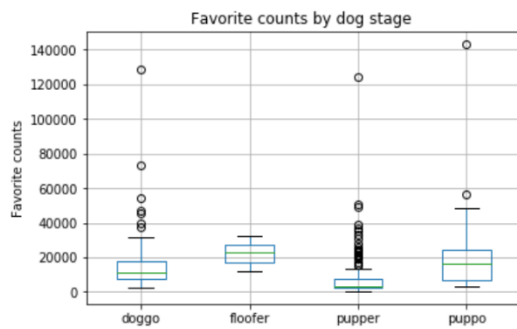
Tweet Post Breakdown by dog stage



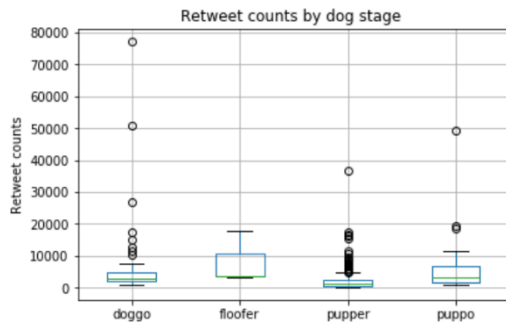
Descriptive statistics by dog stages based on rating ratio, favorite count and retweet counts



	count	mean	std	min	25%	50%	75%	max
doggo	71.0	1.164789	0.164836	0.5	1.10	1.20	1.3	1.4
floofer	3.0	1.266667	0.057735	1.2	1.25	1.30	1.3	1.3
pupper	220.0	1.057273	0.181802	0.3	1.00	1.10	1.2	1.4
puppo	28.0	1.214286	0.120844	0.9	1.20	1.25	1.3	1.4



	count	mean	std	min	25%	50%	75%	max
doggo	71.0	16579.676056	18616.926590	2533.0	7865.50	11465.0	18008.00	128366.0
floofer	3.0	22486.666667	10244.305361	12197.0	17387.50	22578.0	27631.50	32685.0
pupper	220.0	7078.859091	11475.414999	336.0	2326.75	3154.0	7301.75	124479.0
puppo	28.0	23178.750000	27598.889038	3209.0	7128.75	16204.5	24705.00	143493.0



	count	mean	std	min	25%	50%	75%	max
doggo	71.0	5830.154930	11059.886200	699.0	1926.00	2992.0	4785.5	77409.0
floofer	3.0	8266.666667	8373.308327	3157.0	3435.00	3713.0	10821.5	17930.0
pupper	220.0	2280.122727	3615.240845	89.0	671.25	1119.5	2365.5	36527.0
puppo	28.0	6726.392857	9646.891084	692.0	1710.25	3375.0	6826.0	49168.0

Dog Breed Prediction

We then analyze the dog breed prediction algorithm that determine the dog breed based on posted images.

We noted that on the prediction, the algorithm was not able to recognize the dog breed easily.

As per summary table below,

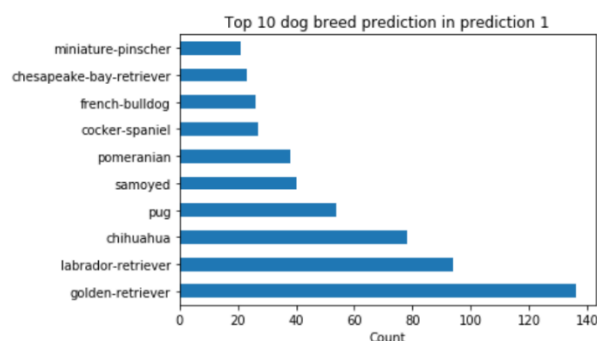
- for the dog breed recognition success rate for the first prediction is 50/50
- for the second and third predictions, the result are slightly lower This may be because the dog images are sometimes often mixed with other elements such as donut which make the image recognitions more difficult.

	Prediction_1	Prediction_2	Prediction_3
Dog Breed Recognized	0.499492	0.480203	0.457868
Dog Breed Not Recognized	0.500508	0.519797	0.542132

Then we checked the top 10 dog breed recognition for prediction 1, 2 and 3.

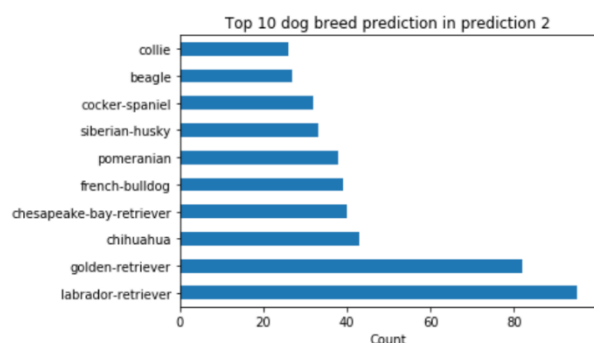
- This does not assure that the prediction match with the images but we can see 3 majors dog breeds that stands out for the three levels of predictions : Golden retriever, Labrador Retriever and Chihuahua.
- Is it because the algorithm is more trained to recognize this type of dog breed or because these are the most popular dog breed at the moment ? It is a question that can be interesting to continue to investigate.

The prediction confidence for top 10 breeds for prediction 1 is 66%, for prediction 2 is 18% and for prediction 3 is 8%.



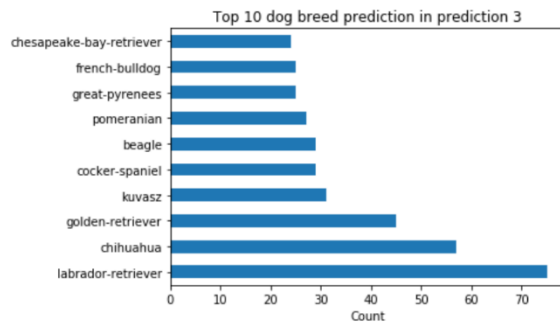
Descriptive statistic on prediction confidence level

count	477.000000
mean	0.665469
std	0.264478
min	0.044333
25%	0.447803
50%	0.697480
75%	0.918144
max	0.999956



Descriptive statistic on prediction confidence level

count	495.000000
mean	0.184136
std	0.106409
min	0.000046
25%	0.107930
50%	0.179918
75%	0.251811
max	0.467678



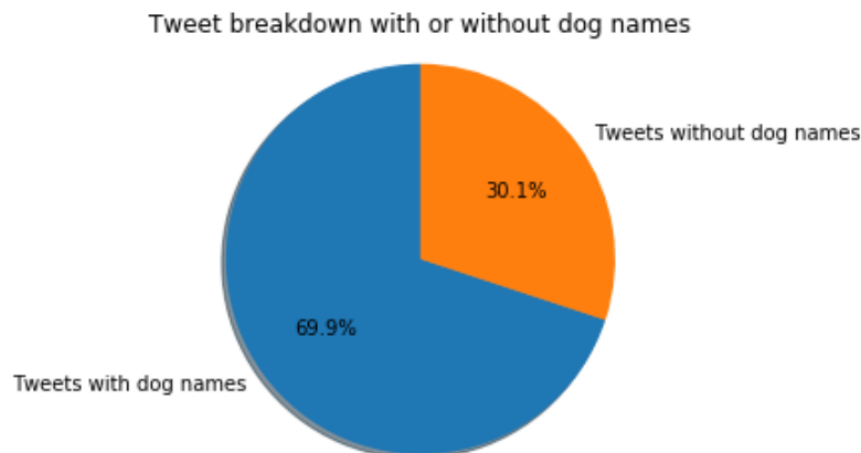
Descriptive statistic on prediction confidence level

count	557.000000
mean	0.079194
std	0.052657
min	0.000011
25%	0.037400
50%	0.073473
75%	0.113796
max	0.270673

Names, ratings & popularity

It would be interesting to see if the dog is named in the tweet can impact the ratings, favorites and retweet counts.

- 70% of the tweets have dog names. When the tweet author presents the dog by its name, it might that it is a picture of his dog or a dog that he/she knows.
- We can see that tweets with dog named have an average rating ratio better than tweet without dog names
- For favorite and retweet counts, the statistics are relatively better when the dog is named. Favorite and retweet counts third quartile for tweets with dog names are higher than tweets without dogs names.
- The most popular dog names are Charlie, Lucy, Oliver, Copper, Penny, Tucker



Descriptive statistics for tweet with dog names

	tweet_id	rating_numerator	rating_denominator	img_num	p1_conf	p2_conf	p3_conf	favorite_count	retweet_count	rating_ratio
count	1.378000e+03	1378.000000	1378.0	1378.000000	1378.000000	1.378000e+03	1.378000e+03	1378.000000	1378.000000	1378.000000
mean	7.433335e+17	11.978229	10.0	1.206096	0.600606	1.343891e-01	6.034780e-02	9186.414369	2649.381713	1.197823
std	6.893695e+16	47.592919	0.0	0.568777	0.273765	1.017331e-01	5.145831e-02	11912.090853	3970.316641	4.759292
min	6.664188e+17	2.000000	10.0	1.000000	0.044333	1.011300e-08	1.740170e-10	105.000000	22.000000	0.200000
25%	6.787808e+17	10.000000	10.0	1.000000	0.367573	5.142930e-02	1.543925e-02	2213.250000	662.000000	1.000000
50%	7.205826e+17	11.000000	10.0	1.000000	0.599557	1.183700e-01	4.929300e-02	4608.500000	1472.000000	1.100000
75%	7.996742e+17	12.000000	10.0	1.000000	0.868634	1.978602e-01	9.334248e-02	11932.750000	3224.250000	1.200000
max	8.924206e+17	1776.000000	10.0	4.000000	1.000000	4.880140e-01	2.734190e-01	124479.000000	61172.000000	177.600000

Descriptive statistics for tweet with no dog names

	tweet_id	rating_numerator	rating_denominator	img_num	p1_conf	p2_conf	p3_conf	favorite_count	retweet_count	rating_ratio
count	5.920000e+02	592.000000	592.000000	592.000000	592.000000	592.000000	592.000000	592.000000	592.000000	592.000000
mean	7.189424e+17	12.599662	11.520270	1.192568	0.580010	0.135206	0.059799	8271.445946	2942.324324	1.089020
std	6.083721e+16	22.129832	12.328811	0.536403	0.267942	0.099539	0.049783	14165.688666	6116.995360	1.704894
min	6.660209e+17	0.000000	10.000000	1.000000	0.063152	0.000010	0.000001	79.000000	13.000000	0.000000
25%	6.739443e+17	9.750000	10.000000	1.000000	0.351625	0.060076	0.017109	1497.250000	544.750000	0.900000
50%	6.906497e+17	11.000000	10.000000	1.000000	0.580437	0.115546	0.049527	3171.500000	1102.500000	1.100000
75%	7.484039e+17	12.000000	10.000000	1.000000	0.819369	0.190076	0.086856	8138.500000	2751.000000	1.200000
max	8.910880e+17	420.000000	170.000000	4.000000	0.999984	0.464816	0.255182	143493.000000	77409.000000	42.000000

