

Report on WeRateDogs Twitter account analysis

Aravind Gowthaman

The WeRate Dogs twitter analysis rates the owner's dogs cuteness with a unique rating system having the numerator more than the denominators if there the cuteness is overloaded.

Here I have done the analysis based on the knowledge learned from my Udacity's Data Analysis Nanodegree in which the data was gathered from three different sources:

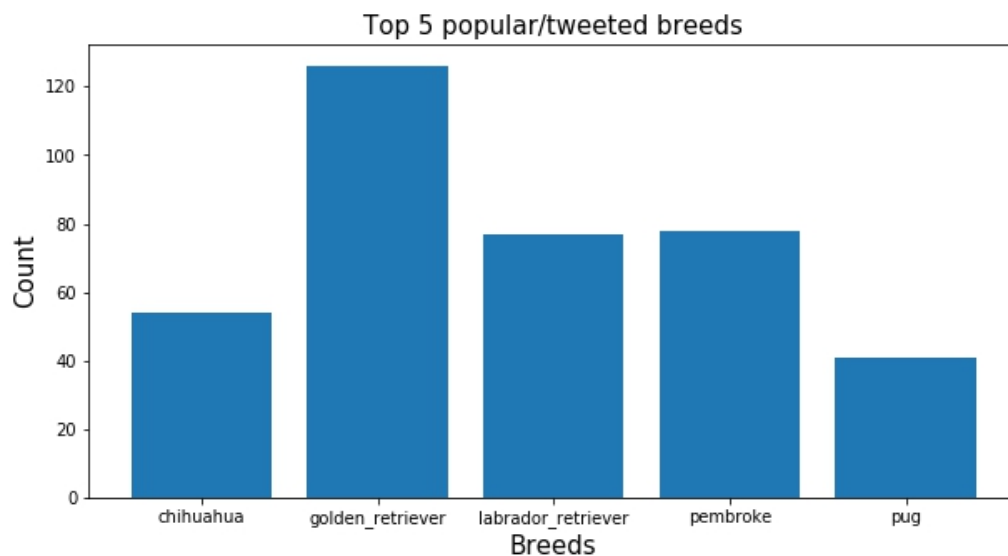
1. Twitter archive (which was provided as csv)
2. Through a http link.
3. Through Twitter API (tweepy).

Before I dive into the analysis I would like to project the limitations on which this analysis were performed

1. The dogstages which was extracted from the text column didn't have the stages defined properly for close to 1800 records of the data as the text either two classes defined or didn't have any.
2. The three Deep learning algorithm which predicted the images of dog breeds had 62% of confidence interval as the highest which meant that prediction could have been better.

Diving into the analysis.

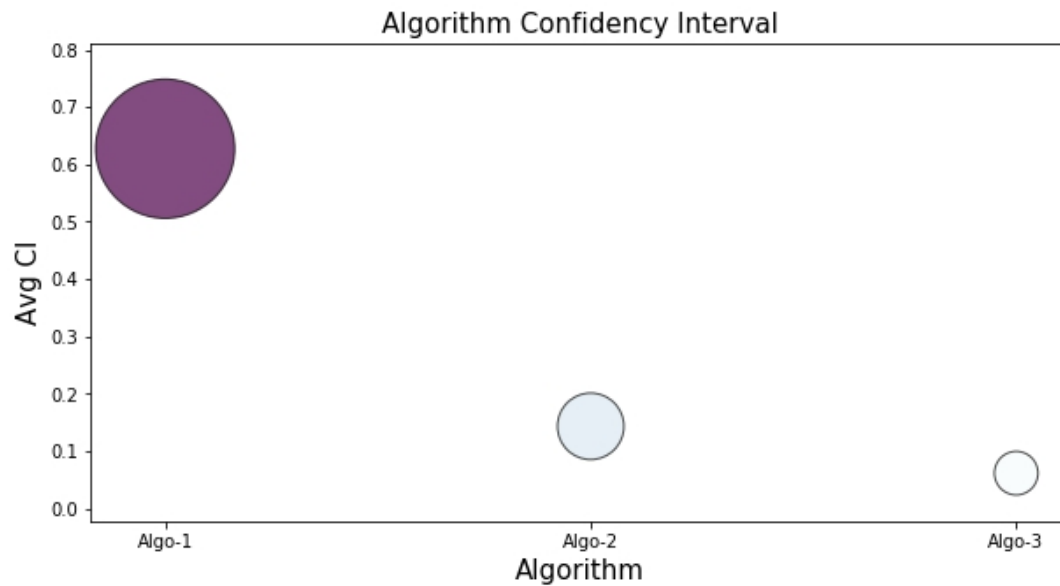
1. What was the most tweeted breed in the WeRateDogs account?



The golden retriever as expected, pug, labrador retriever, chihuahua and pembroke are among the top 5 most tweeted breeds in the account.

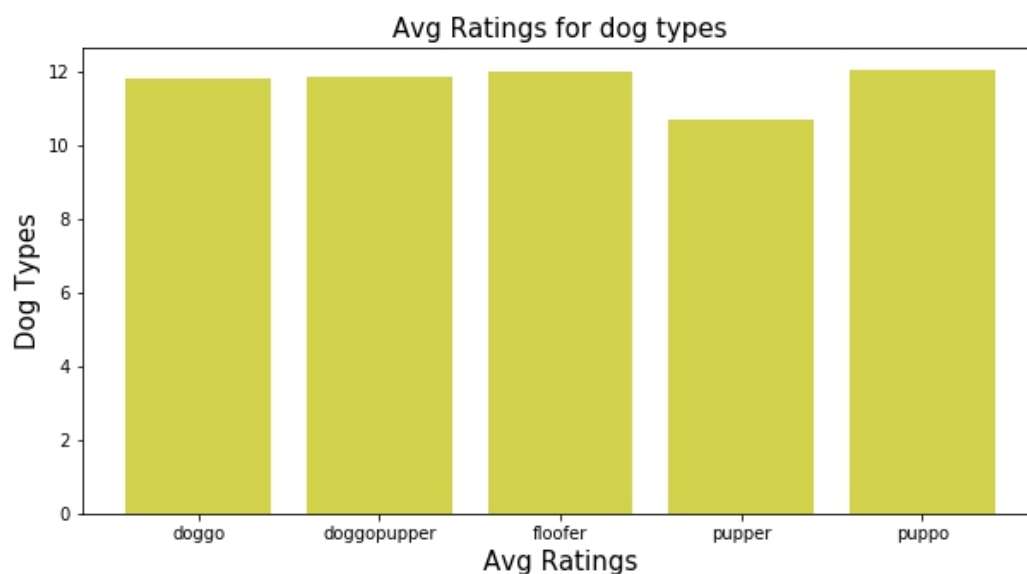
2. What was the average accuracy of the prediction?

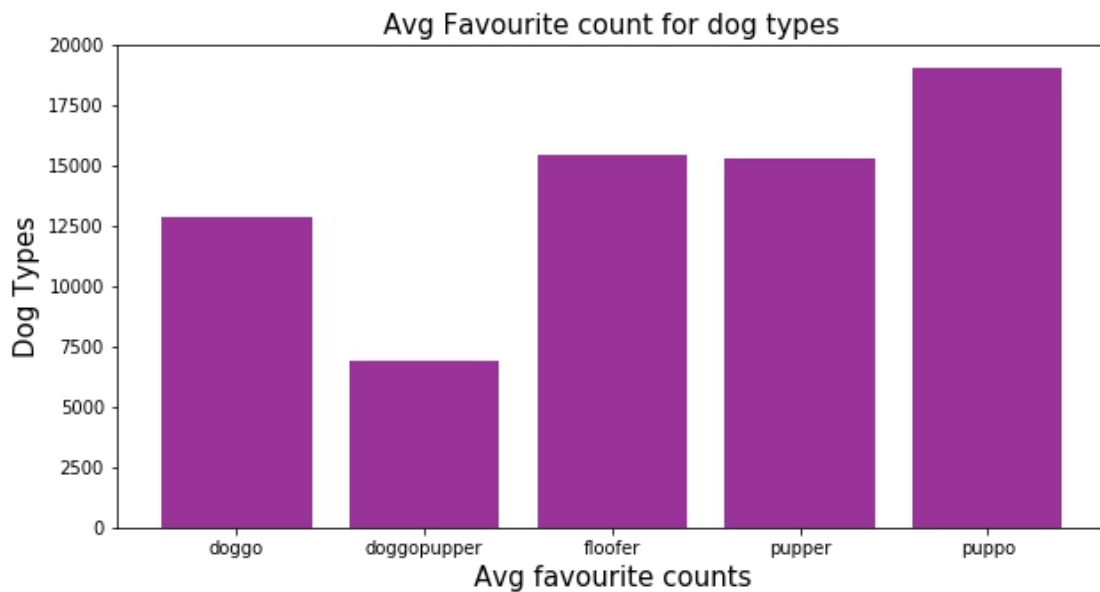
	Algorithms	Accuracy
0	Algo-1	0.627221
1	Algo-2	0.143309
2	Algo-3	0.061627



As it can be inferred from the above insight that algorithm-1 has highest average accuracy with higher confidence interval.

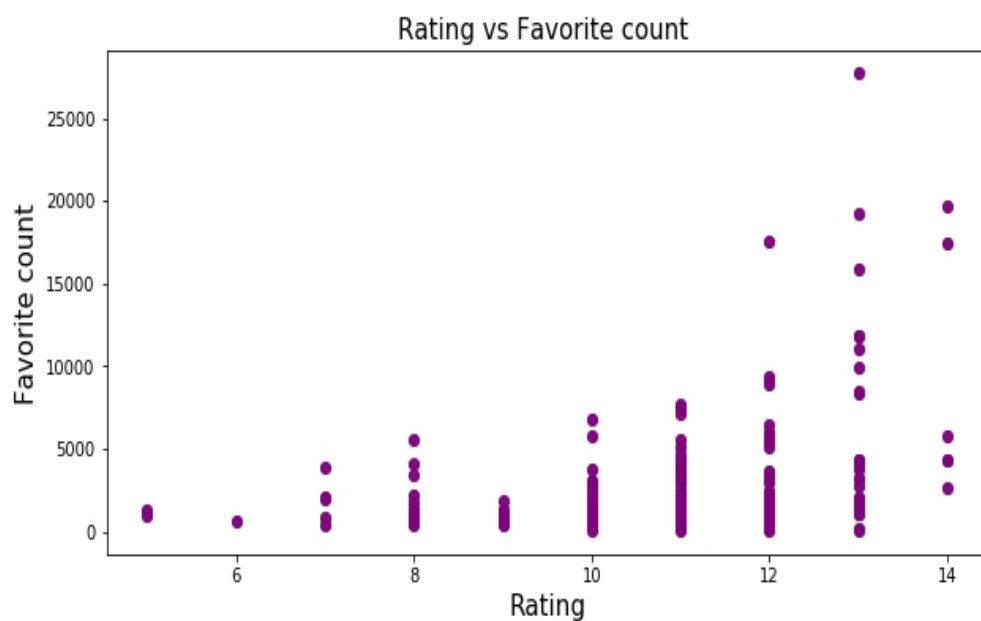
3. Which dog types received the highest rating and was the most liked?

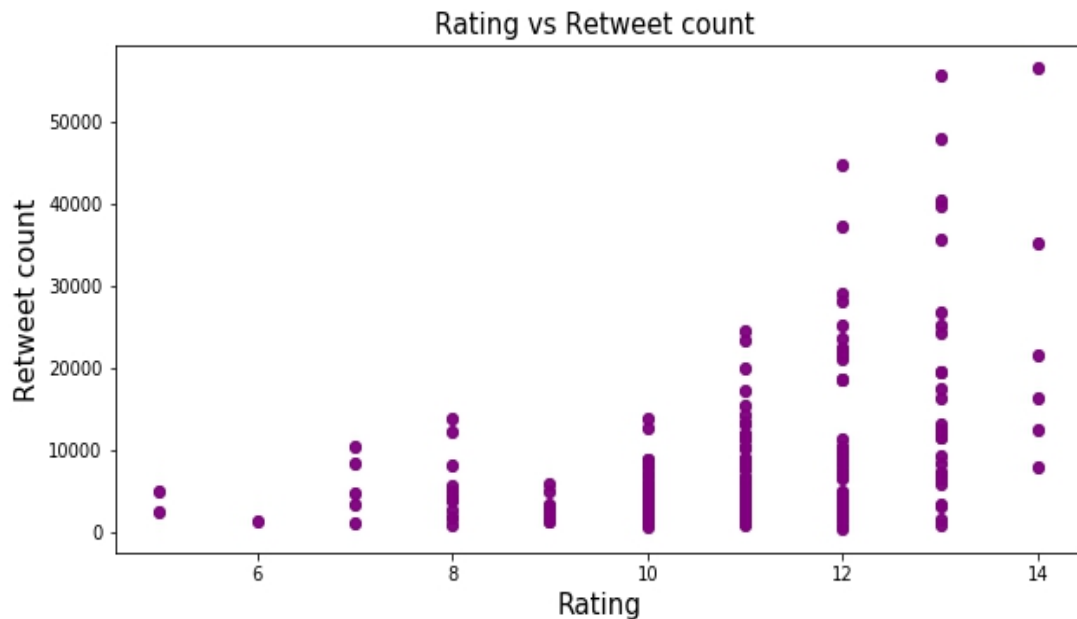




From the above insight on ratings with respect to dogStages all the breeds were more or less closer to each other. But with respect to Favories count, the puppo stage has been the most liked.

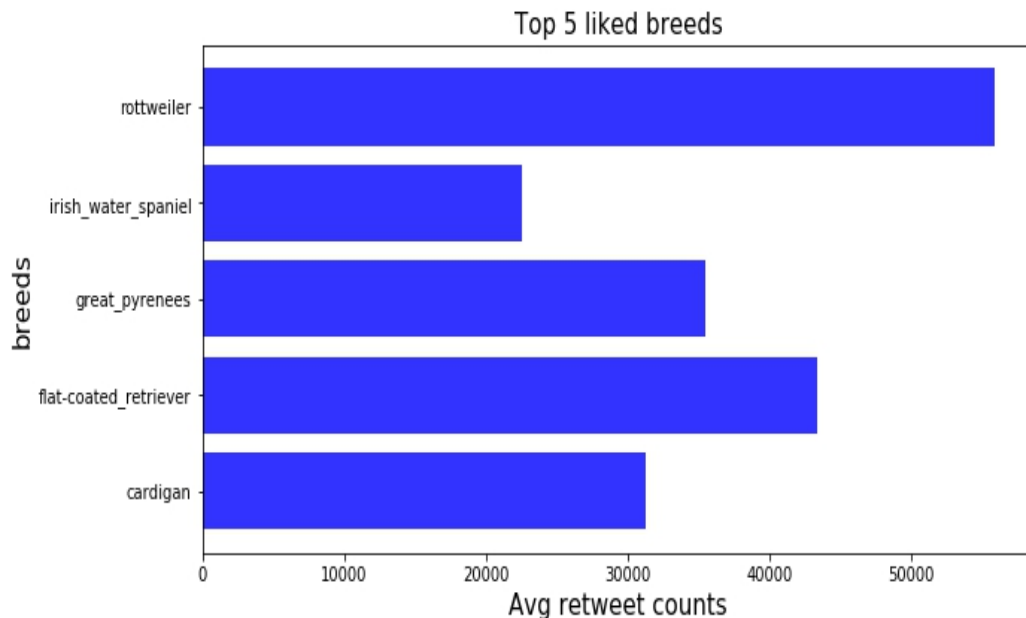
4. Does the Favorites count and Retweets count go hand in hand with numrator ratings?



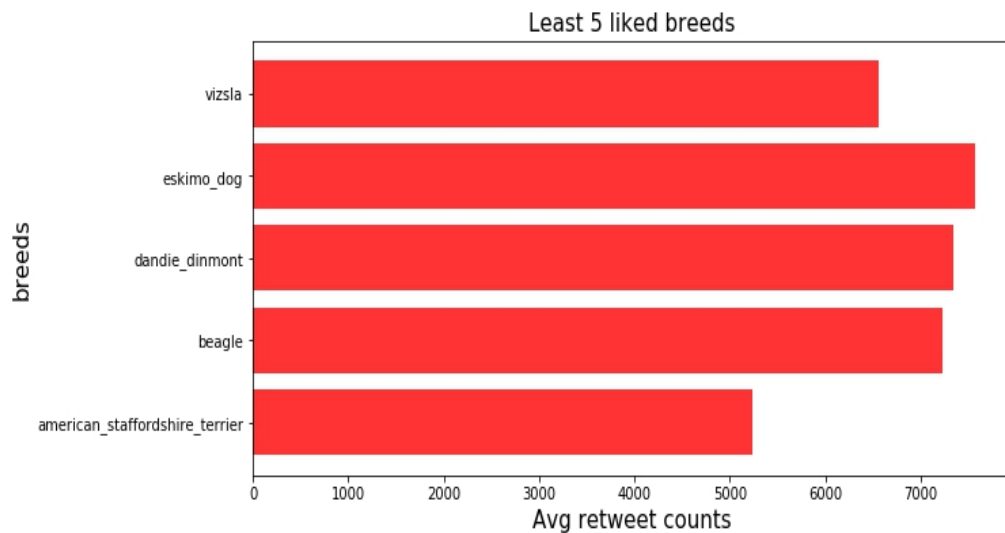


Both a retweet and favourite counts are relative to ratings, ratings having 13 received highest retweet count and favourite count.

5. Which was the most liked breed?



Rotweiller has been the most liked breed of them all in the WeRateDogs twitter account, while on the least liked breeds are visualized below.



6. Getting an insight on how the least and most liked breed for each dog stages

DogStage	p1	favorite_count	retweet_count
doggo	beagle	7234	2877
	briard	16390	5419
	chesapeake_bay_retriever	10919	3726
	cocker_spaniel	9106	1785
	dandie_dinmont	7340	1694
	golden_retriever	17974	6549.2
	great_dane	10426	3556
	great_pyrenees	35442	9849
	labrador_retriever	17601.5	3364
	norwegian_elkhound	11646	3053
	siberian_husky	9188	1824
doggopupper	dalmatian	7976	2070
	golden_retriever	5800	1848
floofer	english_springer	15414	2937
pupper	bloodhound	18917	2933
	briard	6380	1080
	cocker_spaniel	23865	3646
	dalmatian	7540	1246
	french_bulldog	19231	2774
	german_shepherd	8335	2574
	golden_retriever	18136	4396.666667
	irish_water_spaniel	22477	5105
	labrador_retriever	25723	4283

	norwegian_elkhound	21646	5977
	pembroke	12706.33333	5969.666667
	shetland_sheepdog	19514	4588
	siberian_husky	15598	2271
	tibetan_mastiff	22034	8243
	vizsla	6563	1401
puppo	american_staffordshire_terrier	5233	1141
	cardigan	31293	9023
	eskimo_dog	7568	1895
	flat-coated_retriever	43335	16795
	german_shepherd	10130	2725
	golden_retriever	17442	1416
	miniature_pinscher	9942	2695
	rottweiler	55865	17113

As it can be inferred that Beagle was the least liked breed suprisingly and great_pyrenees is the most liked among Doggo stage.

Among puppo it is rottweiler which is the most liked and eskimo is the least liked dog.

Among pupper it is labrador_retriever which is the most liked breed and vizsla is the least liked breed.

7. Getting a statistical insight on the retweet and favourite count

	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count	img_num	p1_conf	p2_conf	p3_conf
count	4.700000e+01	47.000000	47.0	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000
mean	8.216342e+17	12.191489	10.0	4511.276596	16397.446809	1.468085	0.666473	0.151602	0.050351
std	3.523956e+16	0.875718	0.0	4537.885898	11663.761206	0.830167	0.240319	0.117373	0.051275
min	7.588287e+17	11.000000	10.0	1080.000000	5233.000000	1.000000	0.128056	0.000068	0.000034
25%	7.982897e+17	11.500000	10.0	1836.000000	8106.000000	1.000000	0.481273	0.061287	0.012933
50%	8.199522e+17	12.000000	10.0	2933.000000	11588.000000	1.000000	0.699355	0.137187	0.032463
75%	8.477084e+17	13.000000	10.0	4944.000000	21145.000000	2.000000	0.865999	0.220300	0.087950
max	8.863661e+17	14.000000	10.0	23152.000000	55865.000000	4.000000	0.999828	0.467678	0.181985

The retweet count and favourites count has a mean of 4511 and 16397 with the favourite count having a larger spread. The Alogorithm-1 has the highest confidence interval with a larger spread than the other two

Conclusion

The inference is made up to my knowledge on data analysis and the above analysis is just a few of many insights that could have been done. As it turned out that golden retriever is the most tweeted breed and Rotweiller is the most liked breed of them all, while the American terrier was the least liked.