

# Analyzing WeRateDogs Tweets

By Hussain Balhareth

This is a brief non-technical report on data wrangling project at udacity.com for data given to them by WeRateDogs, a twitter account that **professionally** rates dogs. Pictured below is their twitter bio.

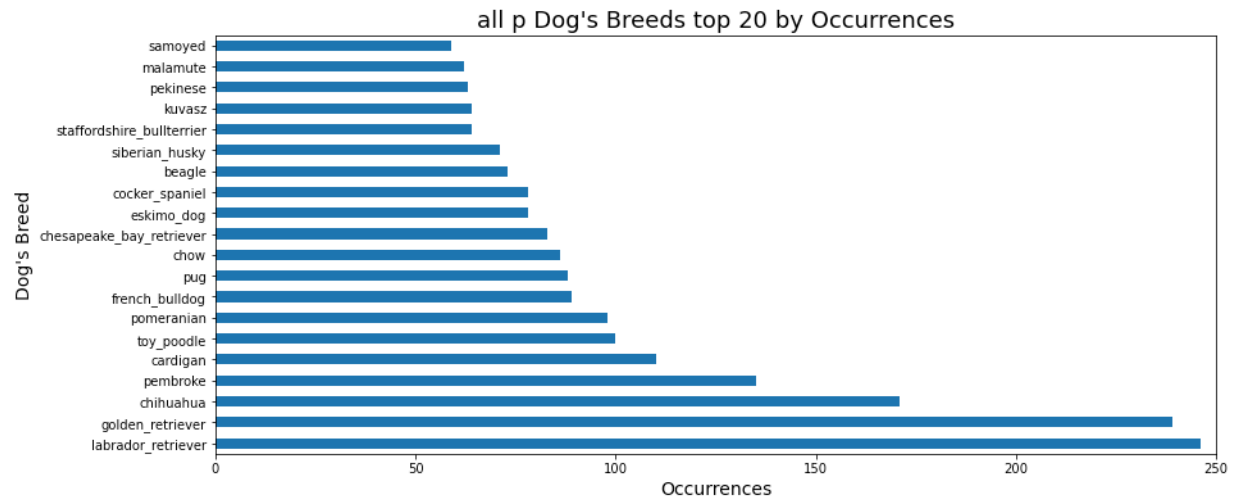


At the start of any analysis it is important to get some basic summary statistics of our numerical variables.

	rating_numerator	img_num	p1_conf	p2_conf	p3_conf	retweet_count	favorite_count
<b>count</b>	1919.00	1919.00	1919.00	1919.00	1919.00	1905.00	1905.00
<b>mean</b>	10.51	1.20	0.59	0.13	0.06	2512.75	8316.08
<b>std</b>	2.20	0.56	0.27	0.10	0.05	4195.74	11414.51
<b>min</b>	0.00	1.00	0.04	0.00	0.00	16.00	81.00
<b>25%</b>	10.00	1.00	0.36	0.05	0.02	611.00	1860.00
<b>50%</b>	11.00	1.00	0.59	0.12	0.05	1294.00	3904.00
<b>75%</b>	12.00	1.00	0.85	0.20	0.09	2915.00	10314.00
<b>max</b>	14.00	4.00	1.00	0.49	0.27	79515.00	132810.00

Below is the number of count a breed was predicted in three predictor algorithms:

No	breed	count
1	labrador_retriever	246
2	golden_retriever	239
3	chihuahua	171
4	pembroke	135
5	cardigan	110
6	toy_poodle	100
7	pomeranian	98
8	french_bulldog	89
9	pug	88
10	chow	86
11	chesapeake_bay_retriever	83
12	eskimo_dog	78
13	cocker_spaniel	78
14	beagle	73
15	siberian_husky	71
16	staffordshire_bullterrier	64
17	kuvasz	64
18	pekinese	63
19	malamute	62
20	samoyed	59



Below is the top dog stages by count:

No	Dog stage	count
1	pupper	202
2	doggo	59
3	puppo	21
4	multiclass	9
5	floofer	7
6	pupper	202

Below is the top 20 dog names by count:

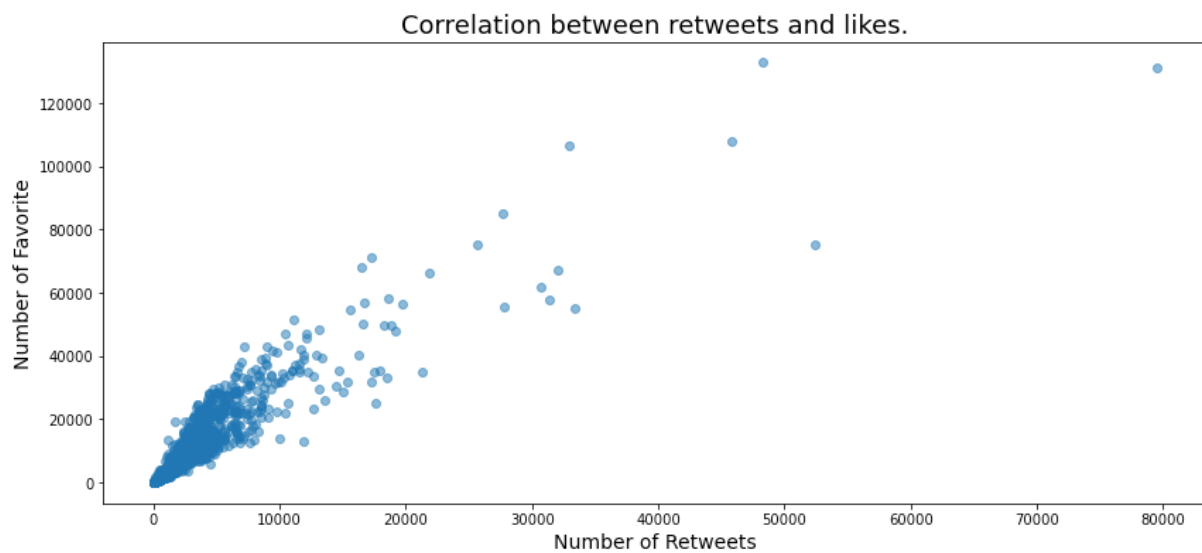
No	Dog name	count
1	Charlie	11
2	Oliver	10
3	Lucy	10
4	Cooper	10
5	Sadie	8
6	Winston	8
7	Penny	8
8	Tucker	8
9	Daisy	7
10	Toby	7
11	Lola	6
12	Jax	6
13	Stanley	6
14	Koda	6
15	Bella	6
16	Oscar	5
17	Louis	5
18	Chester	5
19	Buddy	5
20	Leo	5

The mean success rate of the three predictors ranges from 0.72 to 0.74.

Below is a table showing the tweet counts, retweets and likes sum, and average ratings viewed by stages of dogs:

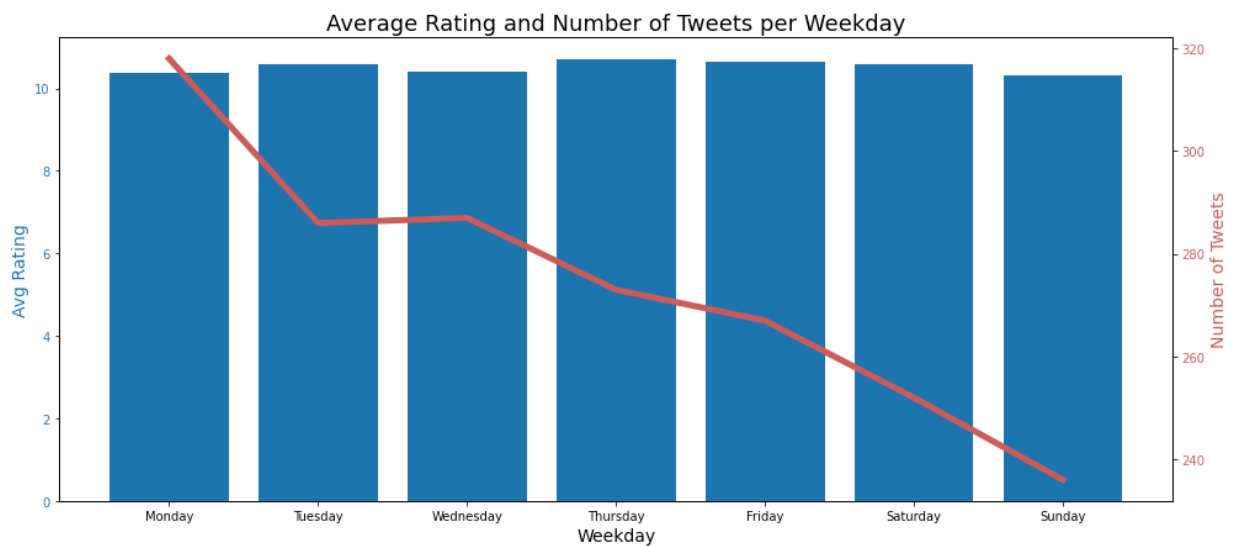
	tweets_count	retweet_sum	favorite_sum	mean_rating
dogtionary				
pupper	202	439982.0	1351383.0	10.628713
doggo	59	368434.0	1012201.0	11.796610
puppo	21	136931.0	456205.0	12.000000
multiclass	9	38239.0	122850.0	11.000000
floofer	7	34781.0	92442.0	12.000000

I found a strong positive correlation between retweets and likes as visualized below:



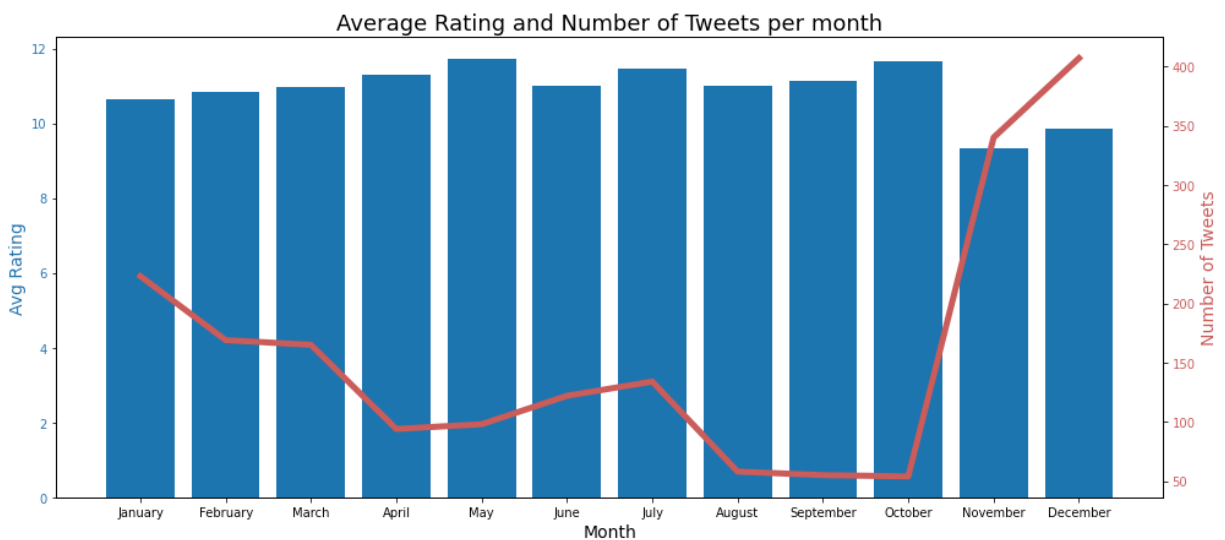
Below table and graph summarizes the pattern of tweets count, retweets/likes total, and average rating on weekly basis (notice that group by and aggregate over all year long weekdays of 52 weeks and hence the value is higher than monthly):

	tweets_count	retweet_sum	favorite_sum	avg_rating
day_of_week				
Monday	318	802713.0	2621742.0	10.364780
Tuesday	286	737289.0	2493569.0	10.576923
Wednesday	287	717125.0	2455358.0	10.414634
Thursday	273	641465.0	2153590.0	10.695971
Friday	267	620610.0	2088959.0	10.640449
Saturday	252	735574.0	2225295.0	10.575397
Sunday	236	532022.0	1803616.0	10.309322



Similarly, below table and graph summarizes the pattern of tweets count, retweets/likes total, and average rating on monthly basis (notice that group by and aggregate over all year long 12 months and hence the value is less than weekly):

	tweets_count	retweet_sum	favorite_sum	avg_rating
month				
January	223	535265.0	1626863.0	10.654709
February	169	402209.0	1487683.0	10.852071
March	165	391667.0	1452731.0	10.981818
April	94	317971.0	1175988.0	11.297872
May	98	404105.0	1452020.0	11.714286
June	122	577504.0	1957158.0	10.991803
July	134	525983.0	2077240.0	11.455224
August	58	208152.0	607052.0	11.017241
September	55	188082.0	581028.0	11.127273
October	54	207123.0	653639.0	11.666667
November	340	358156.0	976323.0	9.317647
December	407	670581.0	1794404.0	9.857494



## Conclusion

1. labrador\_retriever and golden\_retriever are the top two predicted dog breeds
2. There is (as expected) a strong positive correlation between likes and retweets
3. There is a proof to weekly and monthly seasonality towards ratings and notably the number of tweets.