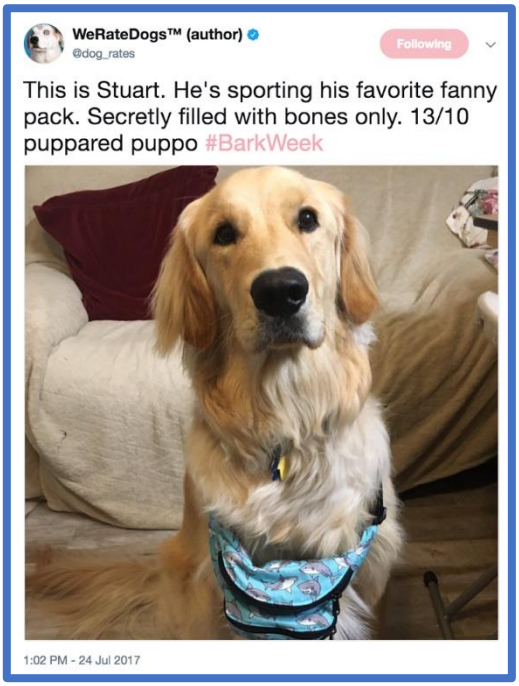
WeRateDogs Twitter Account Dataset Analysis

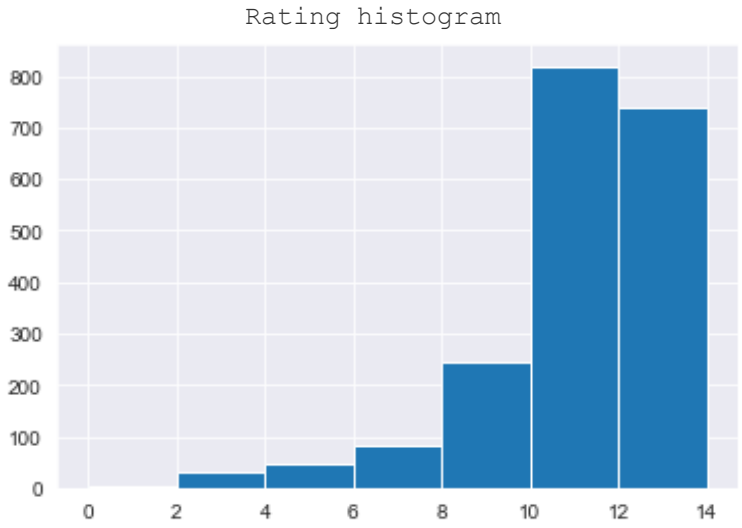


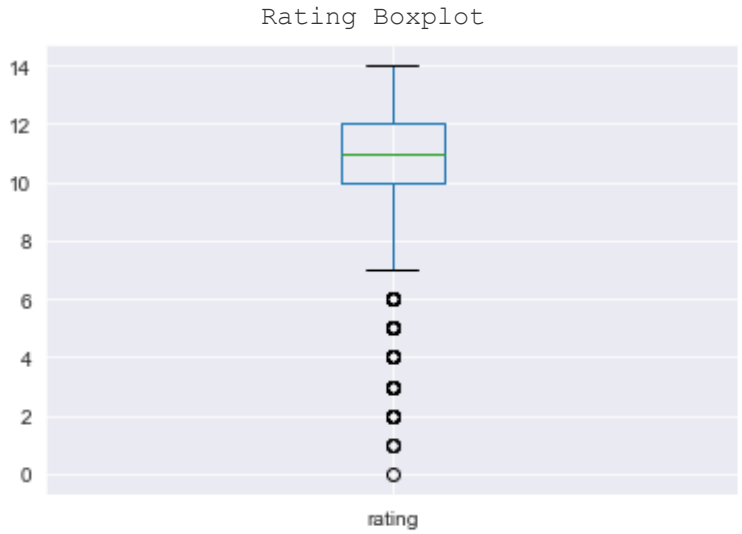
Exploratory Data Analysis

	tweet_id	rating	favorite_count	retweet_count	img_num	p1_conf	p2_conf	p3_conf
count	1.968000e+03	1968.000000	1968.000000	1968.000000	1968.000000	1968.000000	1.968000e+03	1.968000e+03
mean	7.360347e+17	10.539126	8945.185467	2780.971037	1.201220	0.594562	1.346723e-01	6.020101e-02
std	6.754238e+16	2.176618	12274.108524	4698.805472	0.557938	0.271988	1.010987e-01	5.096790e-02
min	6.660209e+17	0.000000	81.000000	16.000000	1.000000	0.044333	1.011300e-08	1.740170e-10
25%	6.758719e+17	10.000000	1984.000000	627.750000	1.000000	0.363174	5.345650e-02	1.605498e-02
50%	7.088226e+17	11.000000	4141.000000	1366.500000	1.000000	0.587635	1.174550e-01	4.945765e-02
75%	7.880674e+17	12.000000	11398.250000	3229.750000	1.000000	0.847560	1.956673e-01	9.157912e-02
max	8.924206e+17	14.000000	132810.000000	79515.000000	4.000000	1.000000	4.880140e-01	2.734190e-01

```
Index(['tweet_id', 'timestamp', 'source', 'text', 'expanded_urls', 'rating',
       'name', 'stage', 'favorite_count', 'retweet_count', 'jpg_url',
       'img_num', 'p1', 'p1_conf', 'p1_dog', 'p2', 'p2_conf', 'p2_dog', 'p3',
       'p3_conf', 'p3_dog'],
      dtype='object')
```

Q1: How are ratings distributed? what is the most frequent rating and the median rating?





Ratings mode = 12
Rating median = 11

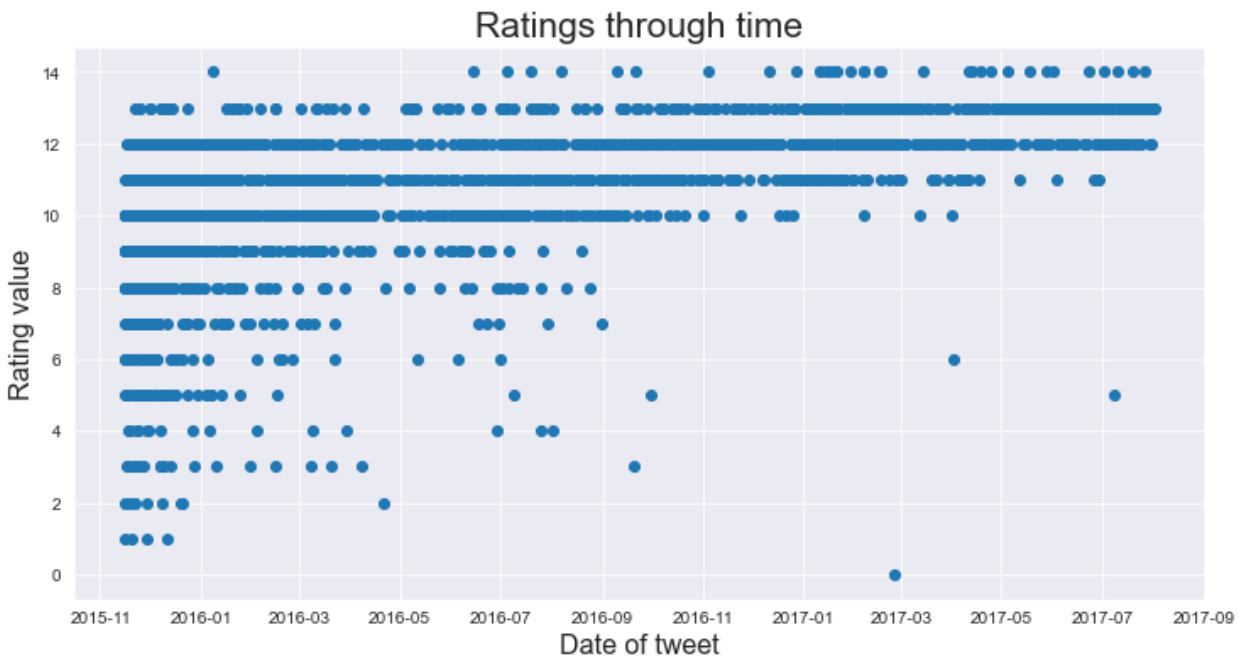
Insights

Rating distribution is left skewed, most of the ratings are in {10, 11, 12}.

Conclusions

@dog_rates account tends to give high ratings as for "they're good dogs brent".

Q2: Did the rating criteria change by time?



Insights

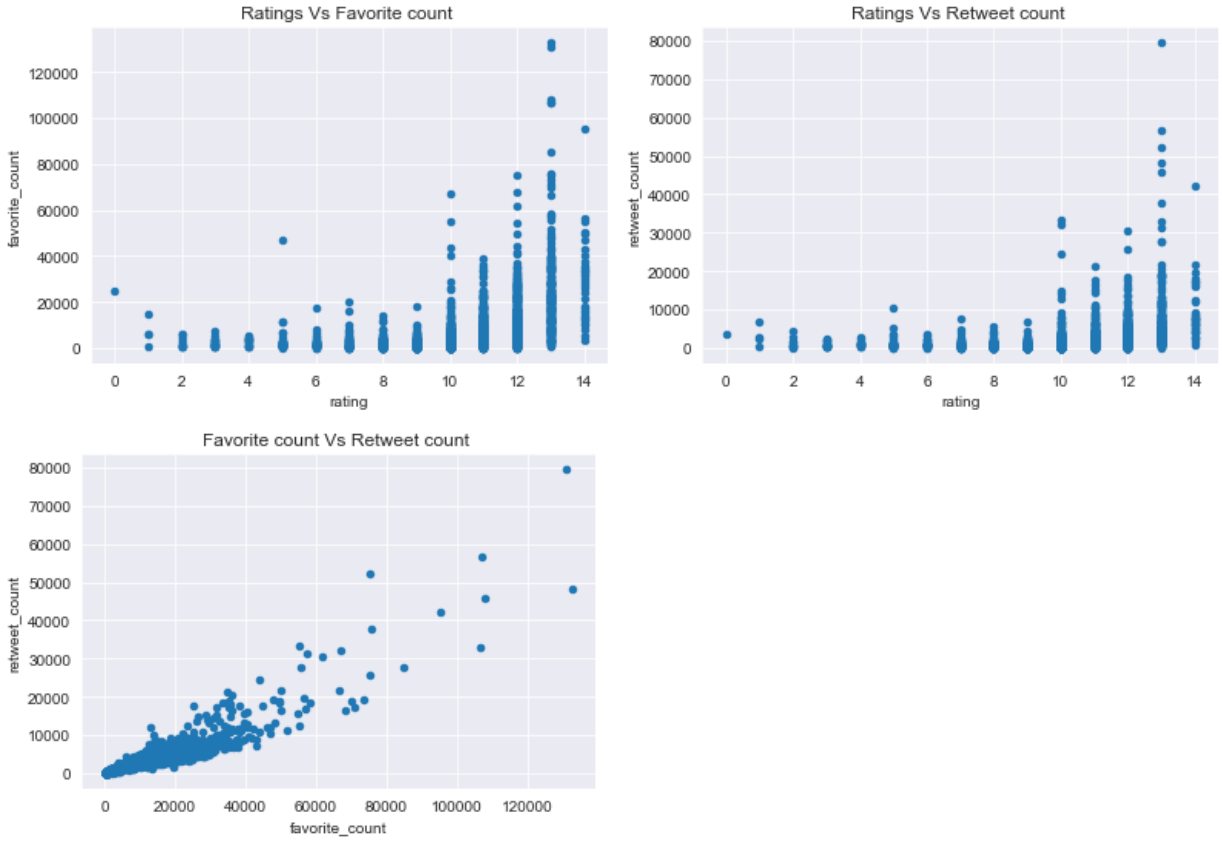
graph shows that @dog_rates account started w normal rating from 0 to 10 with some dogs exceeding 10.

by the time they tended to give higher ratings and nearly abandoning rates below 10 by the end of 2016.

Conclusions

@dog_rates didn't start with their unique rating system, instead it was evolving by time.

Q3: Is there a relation between each of ratings, favorite count and retweet count?



Insights

Graphs show that as a tweet gives a dog a higher rating it gets higher retweet counts and much favored.
Another interesting insight that retweets count is correlated with favorite count, and if we build a regression model trained on retweet counts it might result in a good approximation to favorite count.

Conclusions

People interact more with higher rating tweets.
Retweets count is strongly correlated with favorite count.