

Act Report

January 12, 2022

Mikhail Gorbunov

	Table of Contents	
Data Gathering		3
Cleaning Data		4
Storing Data		5

1. Basic statistics

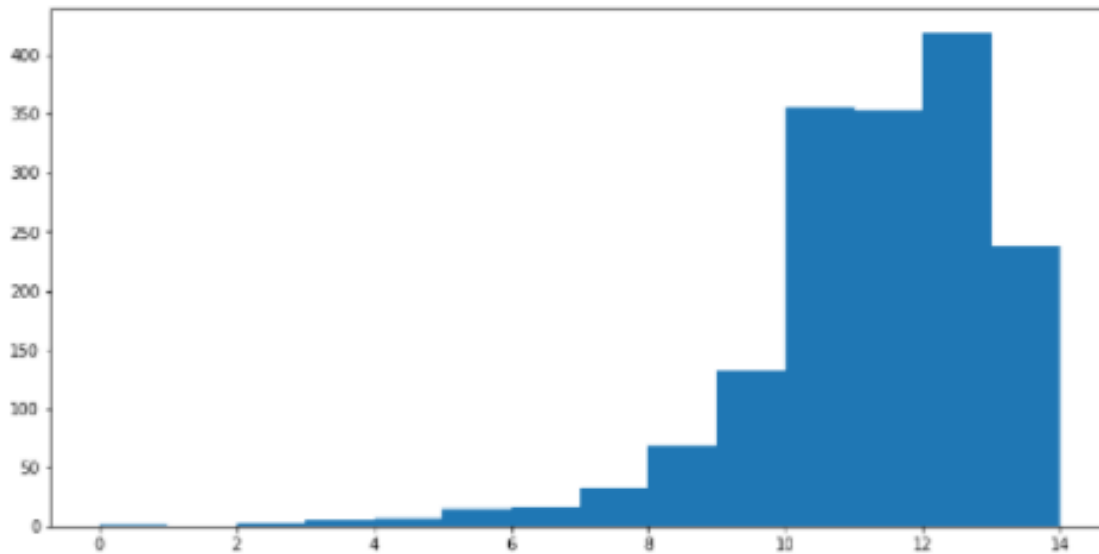
	retweet_count	reply_count	like_count	quote_count	rating_numerator
count	1643.000000	1643.000000	1643.000000	1643.000000	1643.000000
mean	1598.417529	32.171029	5545.990262	55.37675	10.806452
std	3880.221878	72.892923	11093.588921	216.46260	1.791925
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	10.000000
50%	530.000000	14.000000	1883.000000	0.000000	11.000000
75%	1696.500000	35.000000	6244.000000	44.000000	12.000000
max	71860.000000	1145.000000	146691.000000	4353.000000	14.000000

2. Visualizations

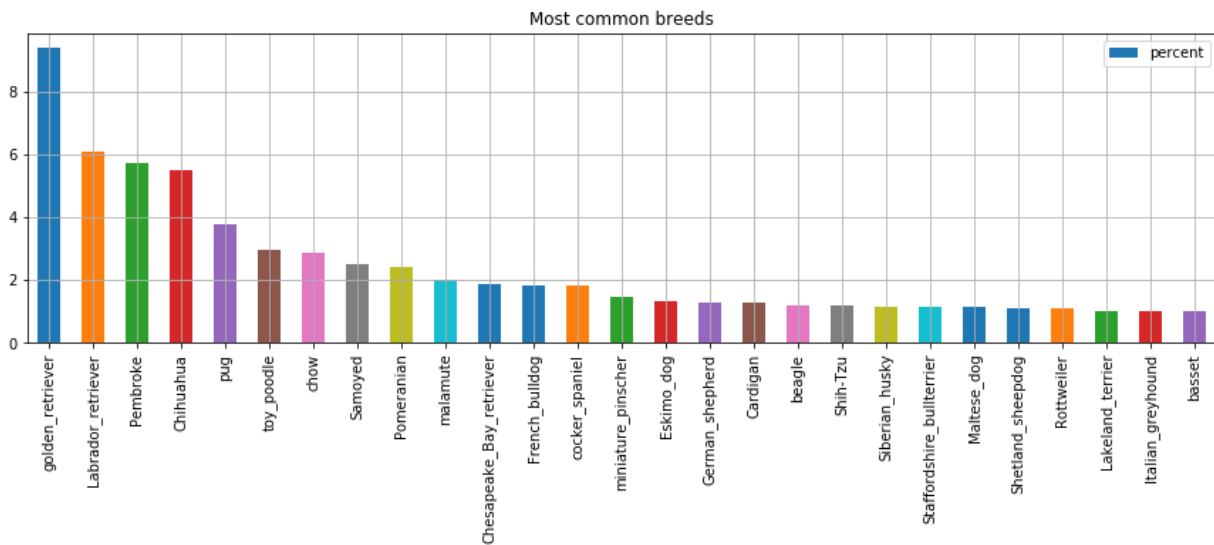
The most popular tweet image:



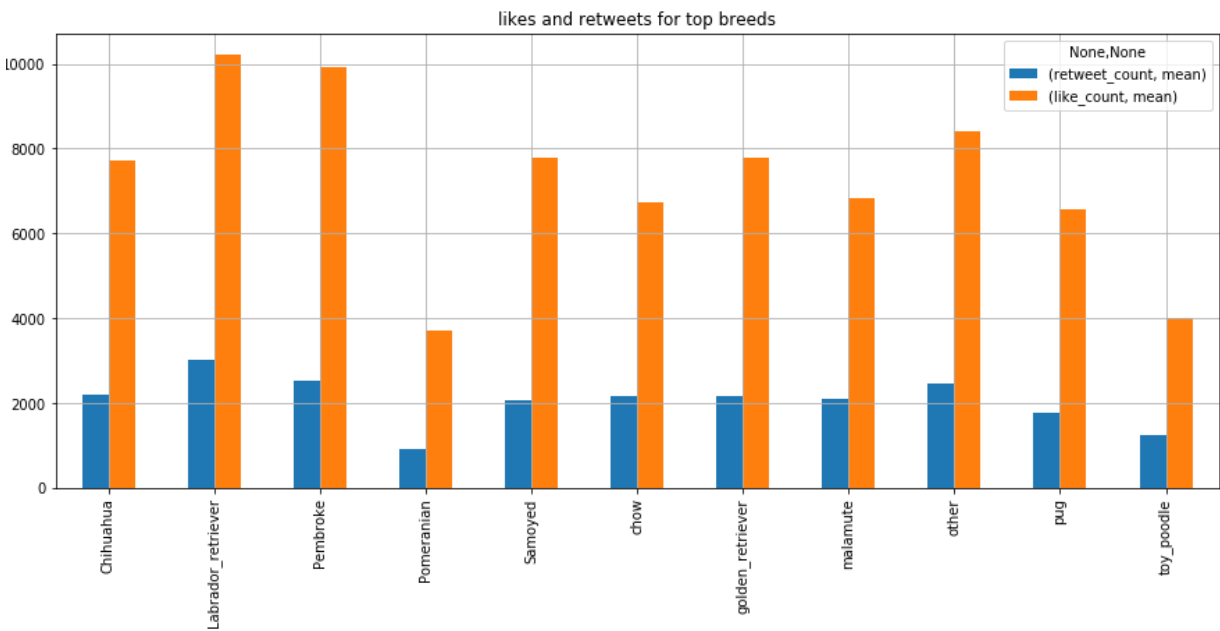
Rating distribution:



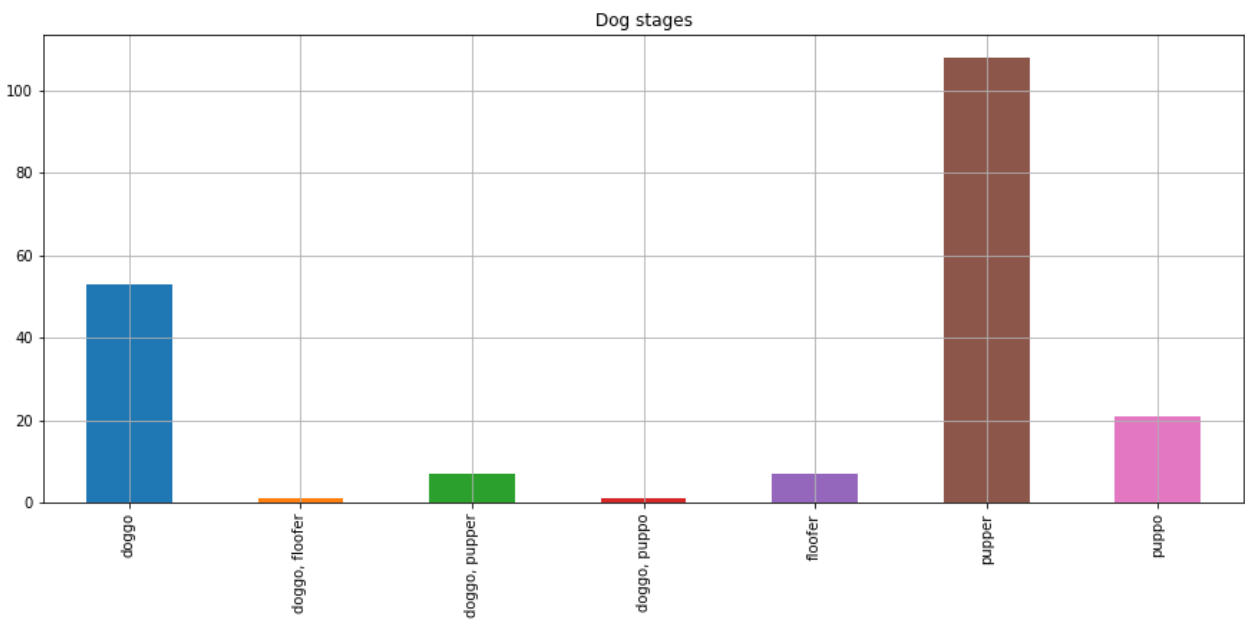
The most common breeds:



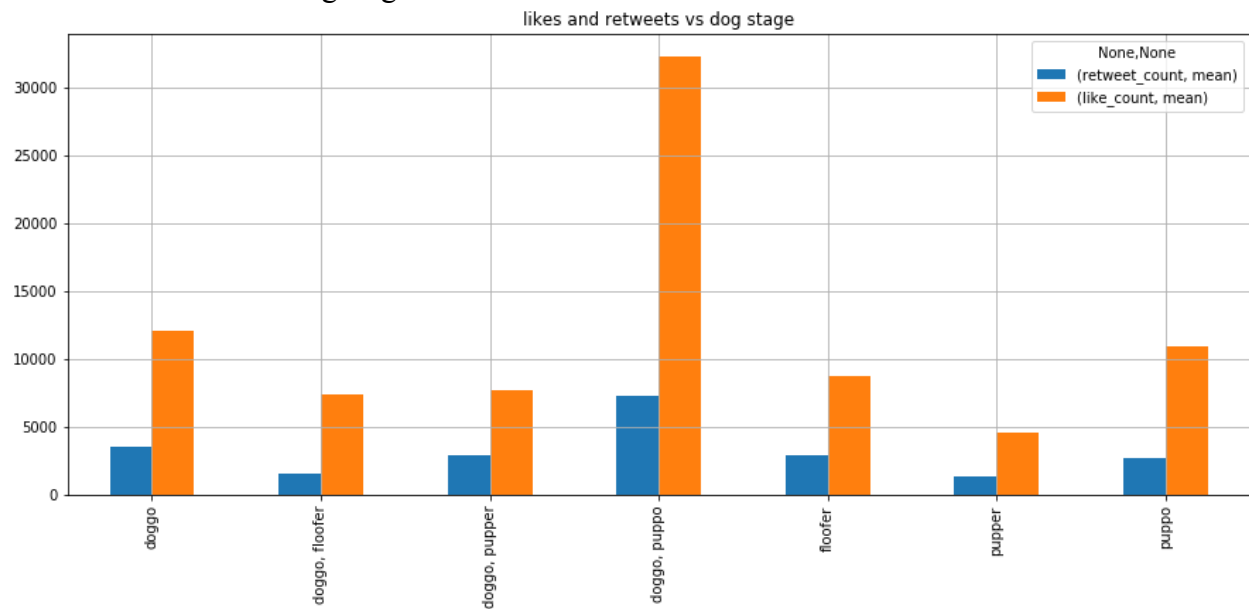
Likes, retweets vs breeds:



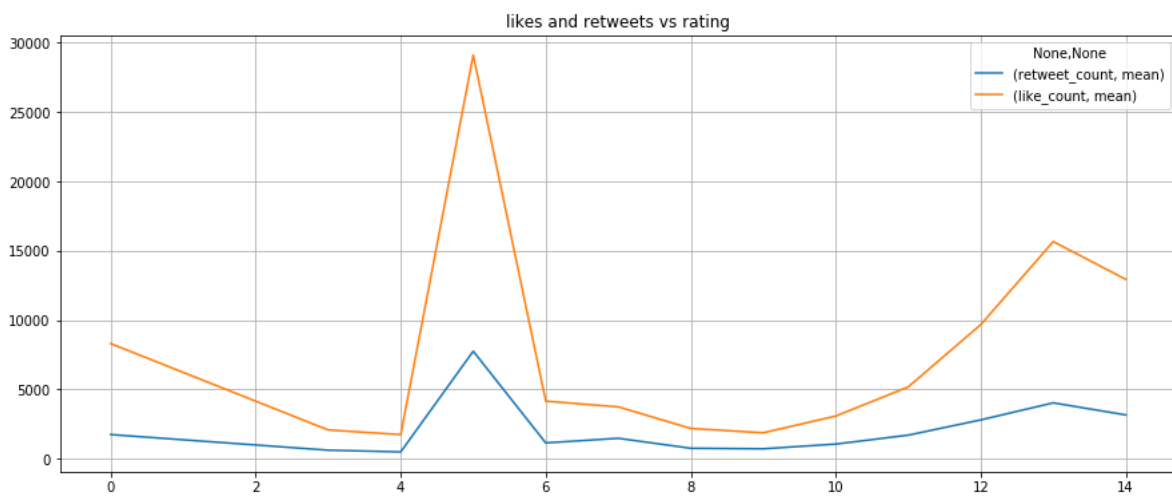
Dog stages prevalence:



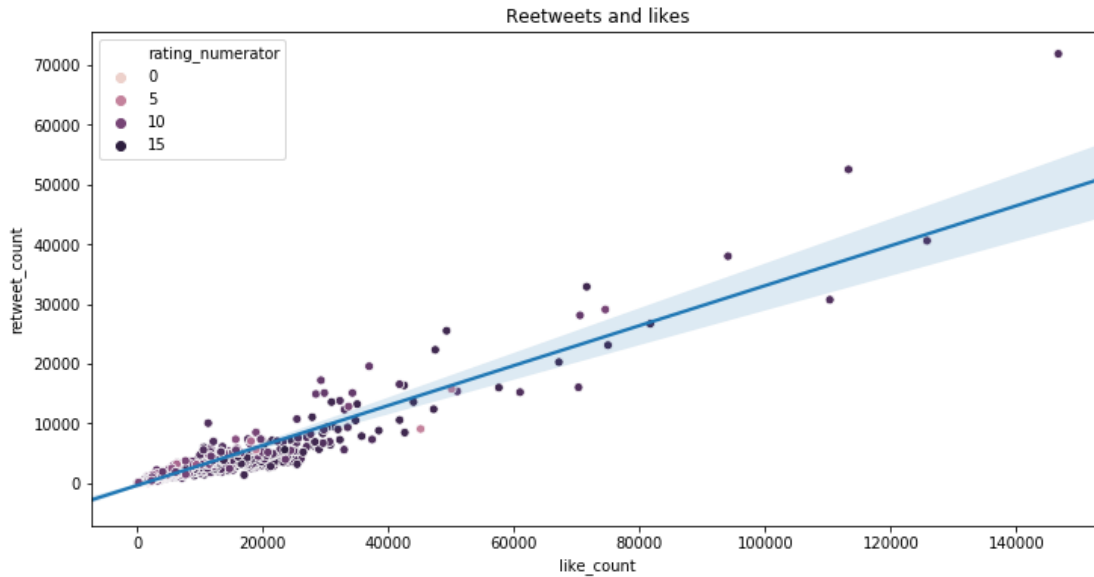
Retweets and likes vs dog stages



Likes and retweets and rating distribution:



Retweets vs Likes:



3. Insights

1. The mean of rating_numerator is 10.8 whereas the three most frequent values are 12, 10 and 11. The distribution is skewed left.
2. Tweets with ratings around 5 and 13 have higher average likes count and retweet count.
3. Looking at the dog stages, pupper is the most frequent stage, whereas doggo, puppo combined category or doggo single category have maximum likes.
4. The golden retriever is the most common breed, whereas labrador retriever has the maximum of likes
5. If the tweet has a higher likes quantity, it has more retweets.
6. The most frequent names are Cooper, Oliver, and Lucy.