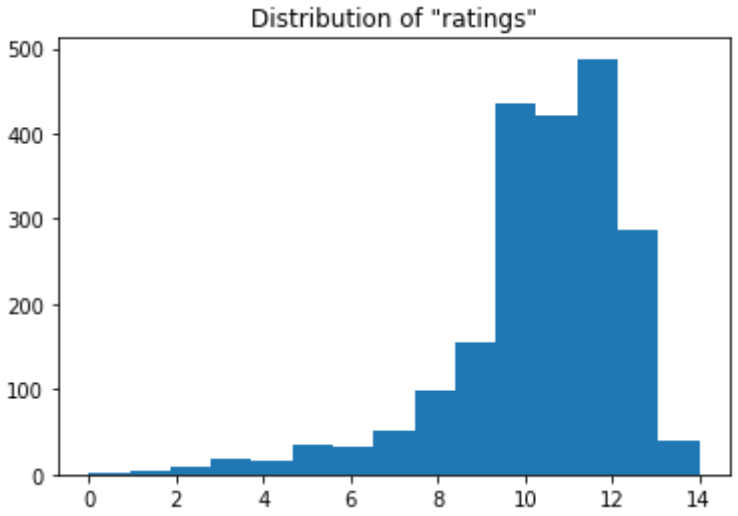
WeRateDogs Analysis Report

The visualizations made using Jupyter and Pandas is a lot limited compared to tools like R Studio. It was very helpful finding resources online helping with visualization techniques to effectively represent Bivariate and Multivariate analysis. The final, master data with the clean and wrangled data includes the factors that we investigated to see the effect on dog ratings.

The Rating distribution is focused around 10 and above. This is a confirmation of the 'novelty' rating system that gave rise to popular mime 'They're good dogs brent', is the most common approach for ratings dogs



'Dog stages' are also another invention of 'WeRateDogs' and it was interesting to see the occurrence of these in the dataset. The dog stages were not identified for most of the dogs but it was apparent which ones of the 'doggo', 'floofer', 'pupper', and 'puppo' dog stages was used the most frequently. There was also evidence that the dog stages represented another novelty factor as with the 10+ rating system. When looking at a subset of data with 10+ ratings, it was observed that dog stages were mostly used for dogs with 10+ ratings

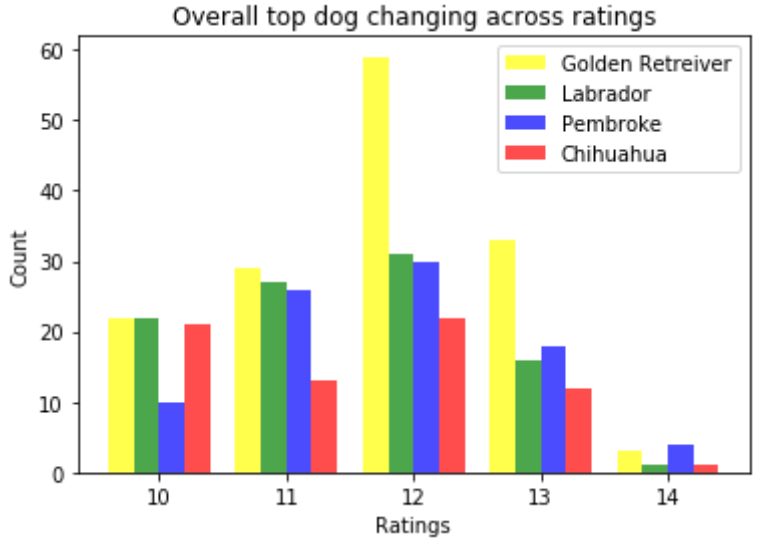
DogStage	Proportion with 10+ ratings
None	0.786838
pupper	0.820717
doggo	0.936709
puppo	0.966667
floofer	1.000000

The insight possible is limited from the data is available. We can form a certain level of intuition but more data is required to form something definitive. This was seen by looking at the data scraped from the 'expanded_url' field. The field identified if the post was a vine, an external twitter post, or a video. While we do see an increase in average ratings for vines, videos and external twitter post compared to the normal picture posts, there is a bias due to

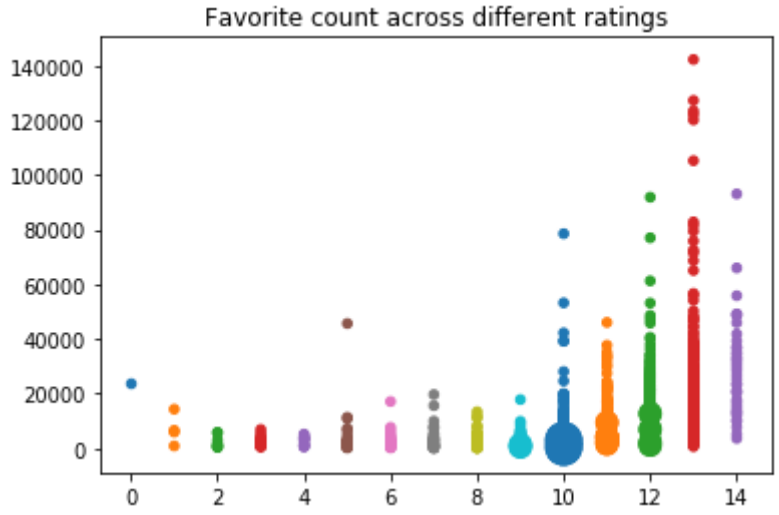
the smaller number of data collected for each: ‘photos’ had 1895 posts compared to just 29 external twitter posts

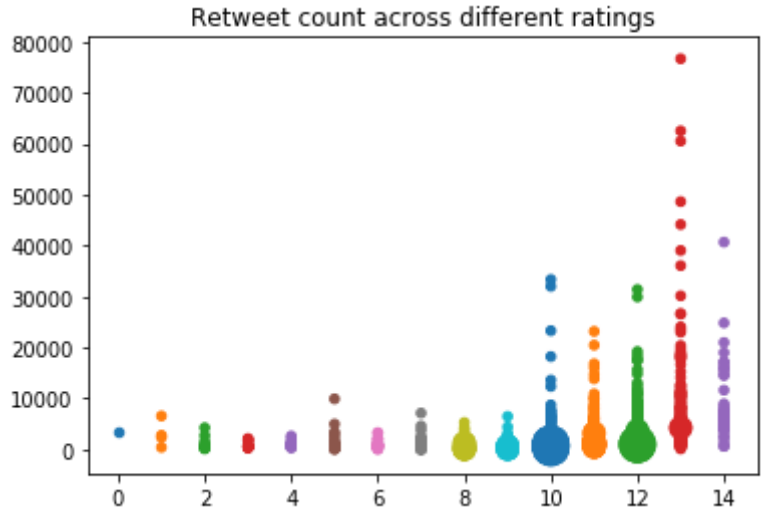
	expanded_urls	rating
0	external twitter post	12.655172
1	rateDogs photo	10.490765
2	rateDogs video	11.916667
3	rateDogs vine	11.395604

It was important to look at data that was more prevelant to see the different in patterns across the different ratings. The plots below show the distribution of ratings for several important factors. An attempt was made to see this for different dog breeds. Given the vast array of dog breeds that were collected, only the top 4 were used



It can be seen that the individual dog breeds also change across the different ratings. ‘Golden retrievers’ that were the most commonly rated breed, also saw a spike in the 12 factor rating which is most commonly used rating. It was the most interesting to view the reaction of the followers and viewers of the feeds and their reactions to the different ratings posted through WeRateDogs





It is evident that the novelty rating system has definitely caught on. Any rating above and including 10 has a significantly higher number of retweets and favorite count and it also happens to show an increase with the rating. It is great to see that this unique rating sytsem has caught on and seems to be something the public seems to agree with.