

WeRateDogs - Analysis of Twitter Dataset

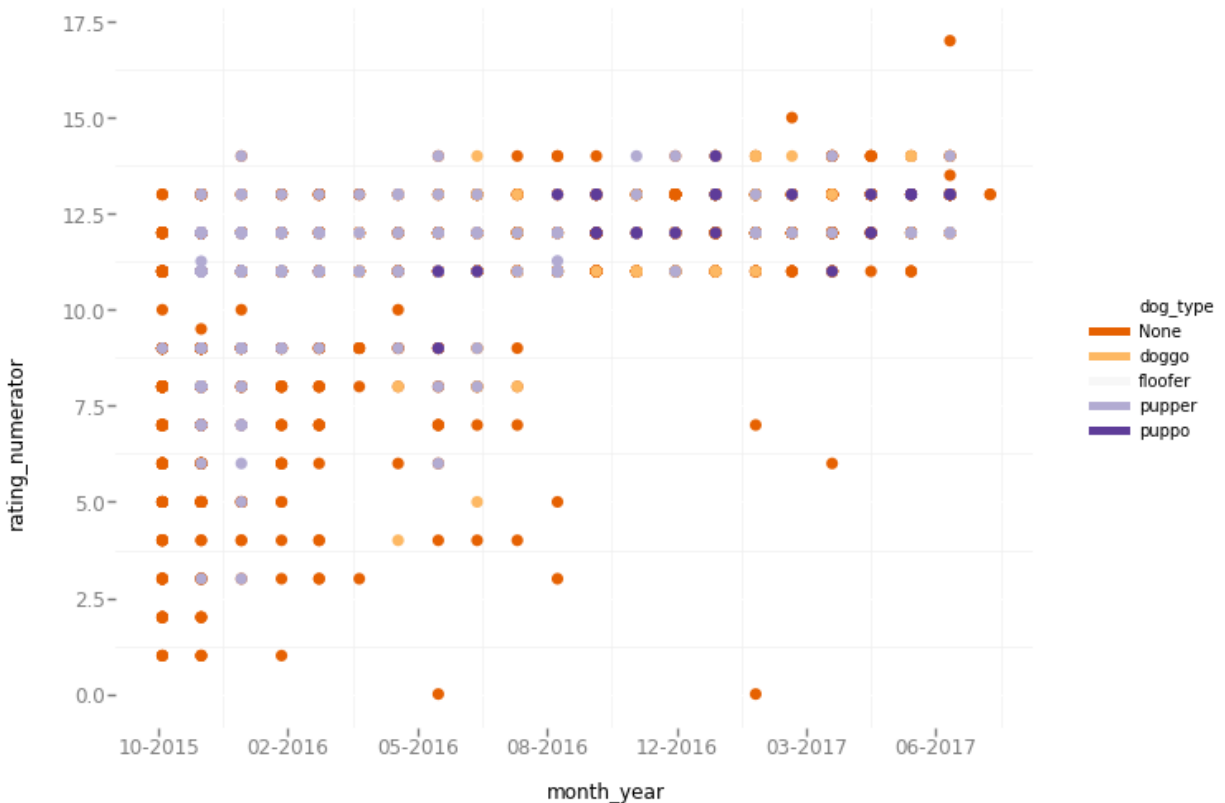
Introduction:

My analysis of the WeRateDogs dataset focused on identifying the trends in data and key features in the dog that affect popularity. I used metrics such as dog ratings, favorites count and retweet count. These metrics were on applied on dates on which the tweet was created, dog breeds data and dog category classifier labelled as dog type.

Analysis and Visualization:

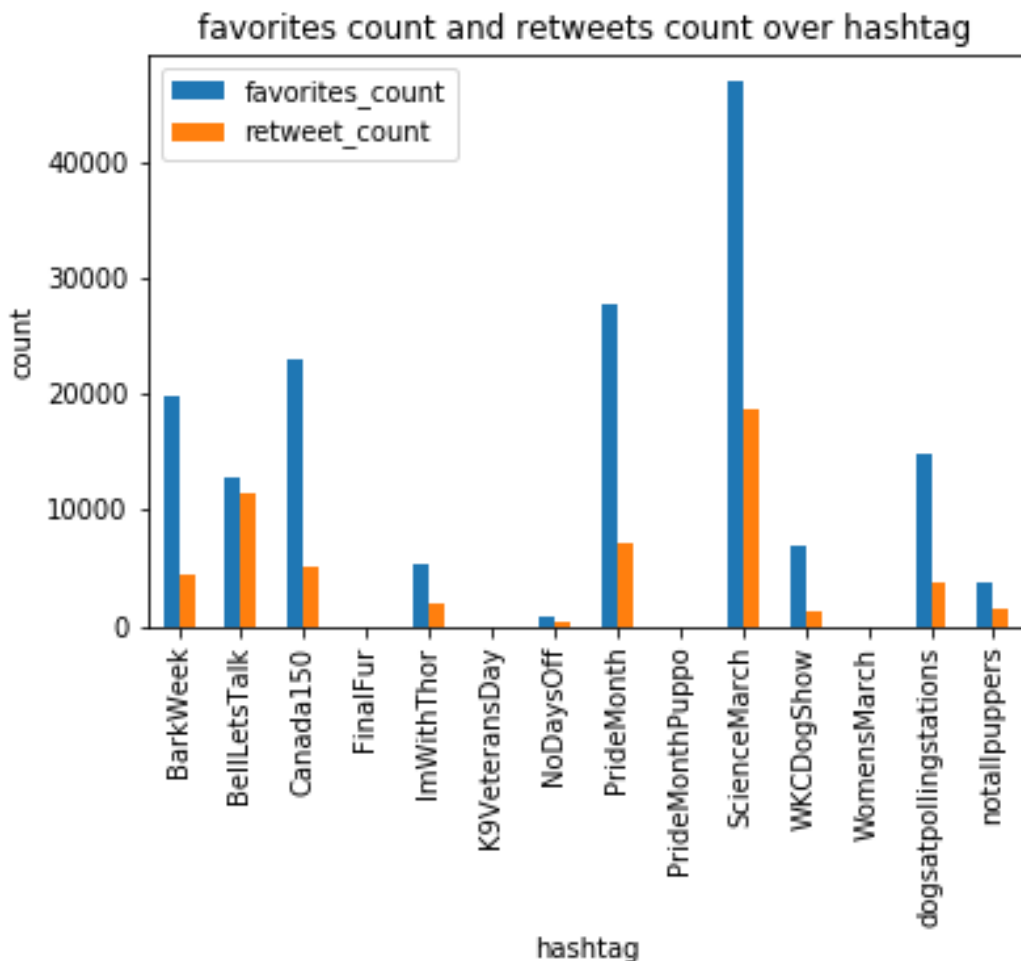
I have visualizations shown below. These show the various trends observed and key insights derived from data.

Dog type and rating analysis



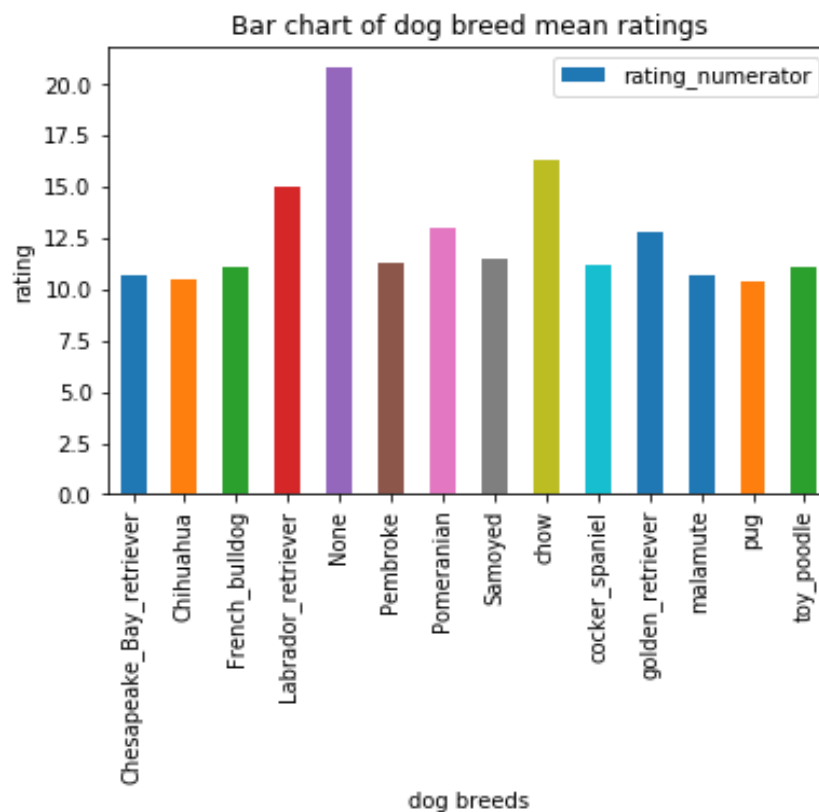
We can see from the above how the rating varies across months. We can also see the ratings for different dog types. We see that most of the dog types were none in the period of 10-2015. We also see that from the period of 08-2016 to 06-2017 there is an increase in the ratings. We can also see that most puppo's have a higher rating with the minimum being 8.75. We also see that very few doggos have low ratings.

Hashtag trend analysis:



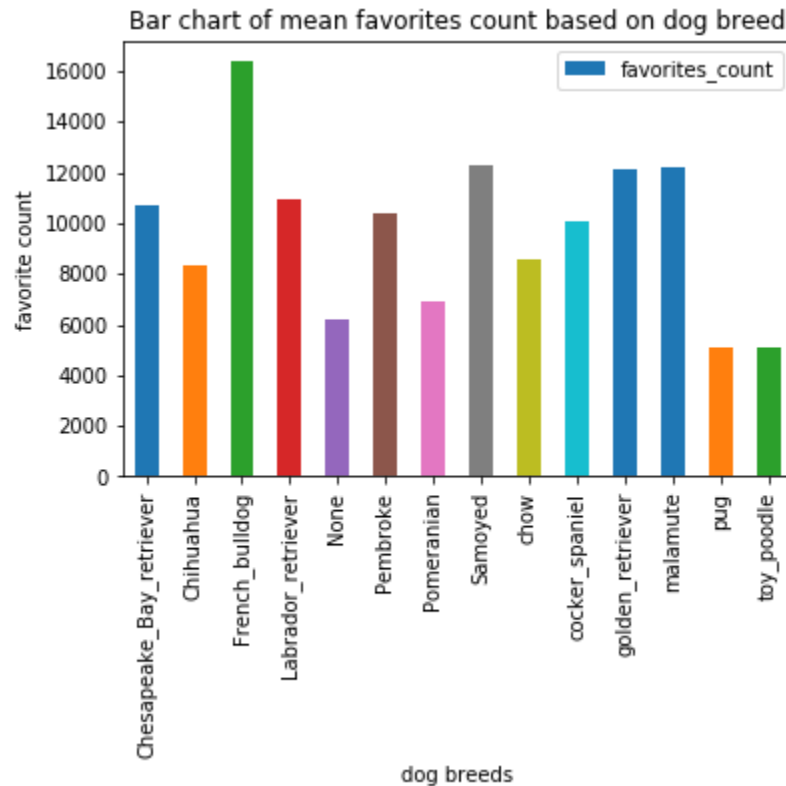
We can see how the retweet count and favorites count vary over the trends generated by we rate dogs. We see that science march has seen the highest popularity and retweets. While womens march, k9veteransday, PridemonthPuppo and final fur all see relatively small number of favorites and retweets. The person incharge of generating these trends should look at replacing the 3 lowest trending hashtags.

Dog breed rating analysis



We can see the mean ratings of the dog breeds above. We can see that the chow has the highest mean rating. We also see that atleast 9 dog breeds have close to the same mean rating. This should mean that there is no bias based on dog breeds.

Analysis of favorites count based on dog type

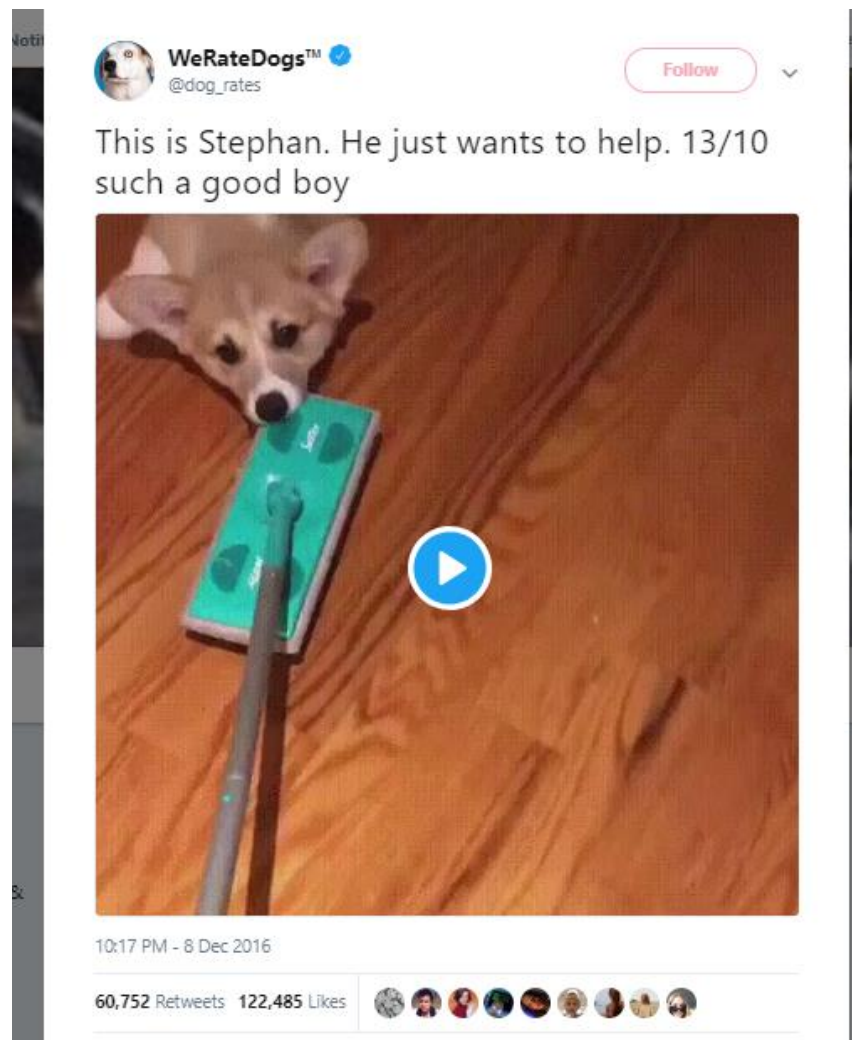


From the chart above, we see that French bulldogs have the highest mean favorite count. We can also see that the pug has the lowest mean favorite count. On comparing this chart with the tweet count for dog breeds. We see that pug has slightly higher tweets than 6 other dog breeds, it is surprising that the pug continues to have the lowest favorite count.

Favorite count analysis based on dog type

	count	mean	std	min	25%	50%	75%	max
dog_type								
None	1153.0	8108.485690	11517.156953	51.0	1495.00	3599.0	10163.00	122441.0
doggo	55.0	13411.054545	15492.775010	739.0	6223.50	8593.0	12480.00	93308.0
floofer	18.0	7188.500000	6715.325223	1370.0	3792.25	4835.5	9052.75	30421.0
pupper	172.0	6407.656977	7888.437527	260.0	2224.50	3200.5	7125.25	50383.0
puppo	21.0	14775.333333	12929.786716	1561.0	6841.00	9868.0	19575.00	55971.0

We see in the table above the distribution of favorites_count based on dog type. We see a high variation between minimum and the maximum. It will be an interesting analysis to see how the maximum value of 122441 compares with the maximum favorites received in the twitterverse. We see that puppo has the highest mean so we can assume that puppo receives more favorites than other dog breeds.



We can see that the tweet above has the highest favorites. This tweet is hilarious and it could be that comedic tweets are likely to receive more favorites. This tweet also has the highest number of retweets.



The tweet with the second highest favorite count is shown above. This tweet is a picture taken with a former president and that explains the popularity of the tweet.