

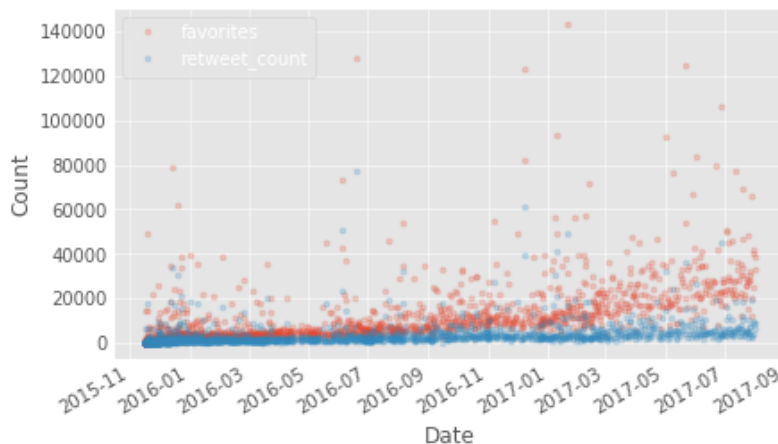
Data Analysis on Different Dog Breeds

By

Tamizharasu Gangatharan

For this analysis data was gathered from three different sources. And Using the tweet IDs from the WeRateDogs archive the Twitter API for each tweet's JSON data was queried using the Python's tweepy library.

Before diving into the statistical analysis, I began by answering some basic questions. What are the most common dog names in the dataset? What is the breed with lowest and highest rating of all. And then several plots and visualizations were created to have different insights on the data which lead to significant findings with different trends.



In this first plot, it can be seen that there is an increase in the page's overall 'favourites' and 'retweet' counts over the course of a single year. This clearly depicts that the page has grown popular over this duration. This increase also means that people started visiting the page much often, contributing to the increase in growth of the page.

Also, the difference in the data count when based upon gender classification, there were three times more male dogs than the female dogs which can be observed from the pie chart below. This is most likely due to more male dogs having their sex specified by either their owner or by WeRateDogs. But people also usually prefer to get male dog as a pet. So, there is nothing much to consider here despite these constraints.



The ratings for the dogs seems to increase over time irrespective of the type of breed which is evident from the below visualization.



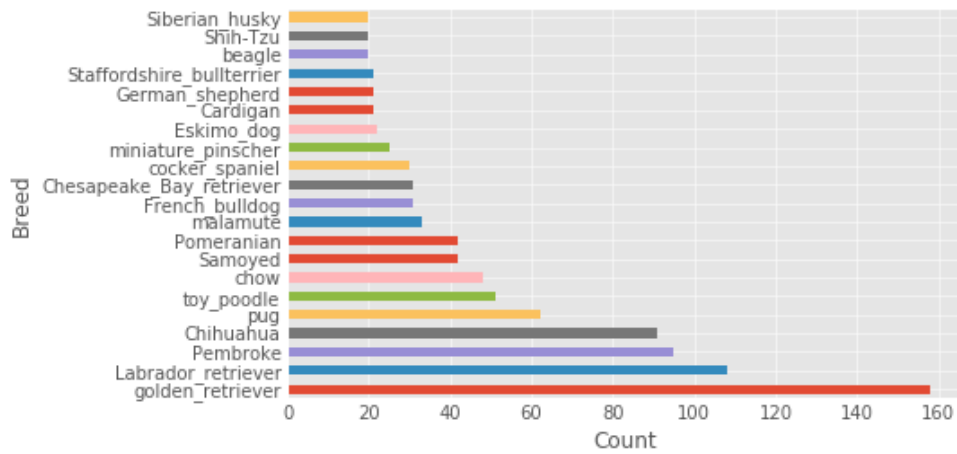
More High ratings were given to the dogs with time. This could mean either the dogs being rated were Good or the page started giving off good numbers and to improve the popularity.

A table representing various data elements based on different breeds of dogs is attached below to see how each breed of dogs differ from every other breed in terms of count within the given dataset and the mean rating that was given to each breed itself.

	count	mean	std	min	25%	50%	75%	max
breed								
Cardigan	21.0	11.142857	1.590148	7.0	10.00	11.00	12.0	13.0
Chesapeake_Bay_retriever	31.0	10.741935	1.510358	8.0	10.00	10.00	12.0	13.0
Chihuahua	91.0	10.516484	2.071568	3.0	9.50	11.00	12.0	14.0
Eskimo_dog	22.0	11.409091	1.402688	9.0	10.00	12.00	12.0	14.0
French_bulldog	31.0	11.193548	1.796652	8.0	10.00	12.00	12.0	14.0
German_shepherd	21.0	11.000000	1.449138	8.0	10.00	11.00	12.0	13.0
Labrador_retriever	108.0	11.180556	1.324567	7.0	10.00	11.00	12.0	14.0
Pembroke	95.0	11.389474	1.746088	4.0	11.00	12.00	12.0	14.0
Pomeranian	42.0	10.779762	1.619435	6.0	10.00	11.00	12.0	14.0
Samoyed	42.0	11.690476	1.352290	7.0	11.00	12.00	13.0	14.0
Shih-Tzu	20.0	10.350000	1.308877	8.0	9.75	10.50	11.0	13.0
Siberian_husky	20.0	11.025000	1.720427	5.5	10.00	11.00	12.0	13.0
Staffordshire_bullterrier	21.0	10.761905	1.374946	8.0	10.00	11.00	12.0	13.0
beagle	20.0	10.150000	1.531253	6.0	9.75	10.00	11.0	13.0
chow	48.0	11.416667	1.350072	7.0	11.00	12.00	12.0	13.0
cocker_spaniel	30.0	11.316667	1.177983	9.0	10.25	11.25	12.0	13.0
golden_retriever	157.0	11.624204	1.188427	8.0	11.00	12.00	12.0	14.0
malamute	33.0	10.878788	1.218544	8.0	10.00	11.00	12.0	13.0
miniature_pinscher	25.0	10.000000	2.565801	2.0	9.00	11.00	12.0	12.0
pug	62.0	10.241935	1.816910	3.0	9.25	10.00	11.0	13.0
toy_poodle	51.0	11.039216	1.264291	7.0	10.00	11.00	12.0	13.0

The Chow breed has the highest average rating followed by Labrador Retriever and by Siberian Husky. This can be confirmed as all these three breeds are friendly and easy to take care of in terms of feeding and grooming when compared to other breeds.

Finally, a perspective on most common breeds is represented with a visualization.



This histogram gives a perspective on the total number of dogs that were rated. The Golden Retriever is the dog to receive maximum number of ratings with a count of 160 while Beagle has the lowest count of all the dog breeds.