

# WeRateDogs: Data Analysis and Visualization

## Introduction

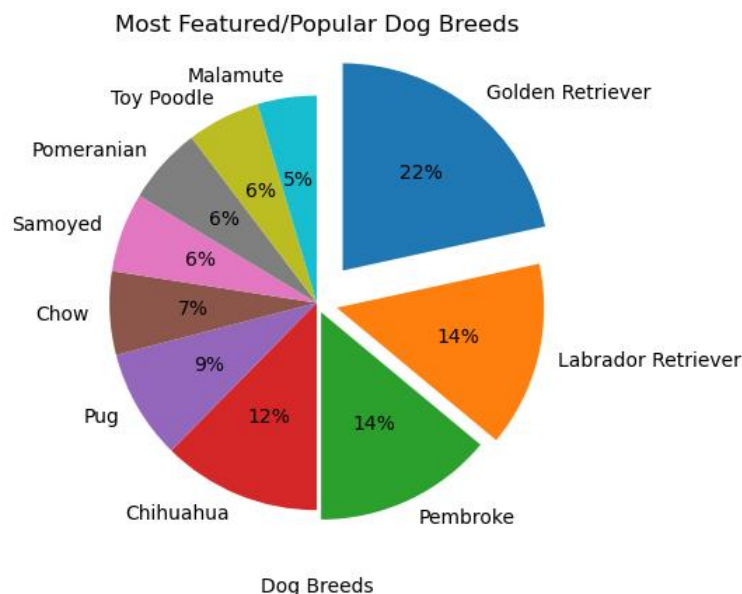
**WeRateDogs** is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10 and the numerators are almost always greater than 10.

In this report, we will be providing a summary of the analysis carried out on tweets from **WeRateDogs** Twitter account to gain insight into their contents. These analyses were based on three datasets provided which include a tweet archive of **WeRateDogs**, extracted retweet counts and favorite counts, and also an image prediction of dogs that predicts the dog breed of the dog images in the tweets.

## Analysis and Visualisations

### 1. Most featured/popular breed of dog

It was found that the most featured dog breed in tweets is the **Golden Retriever**. This is supported by a pie chart showing the number of tweets for each breed, where **Golden Retriever** has the highest count.



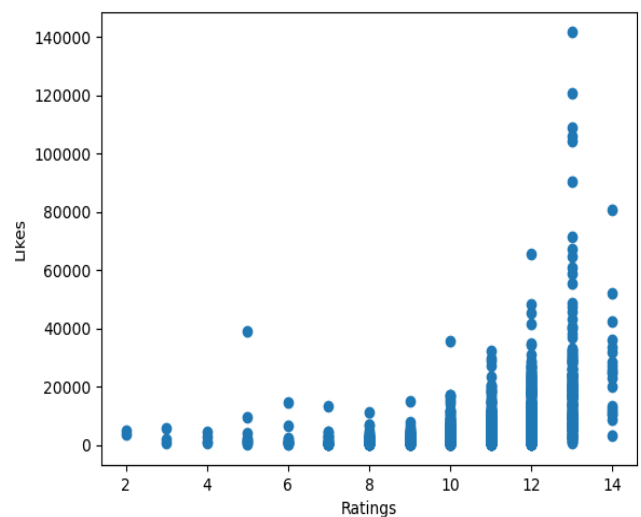
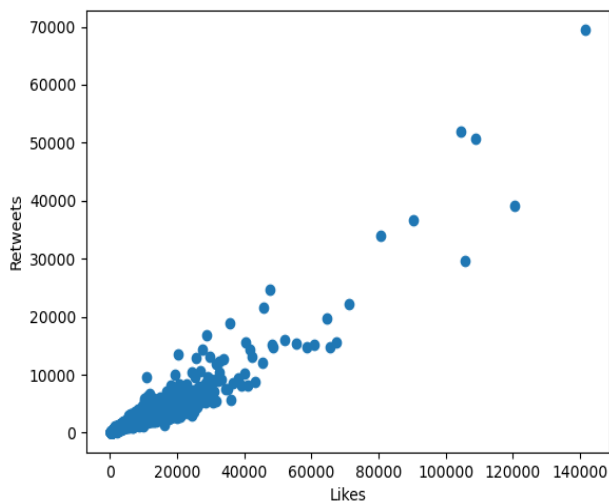
### 2. Dog breeds with the most tweet engagement (likes and retweets)

It was also observed that the dog breed with the most tweet engagements on average is **Saluki**. This is illustrated in a table showing the top 10 average number of likes and retweets for each breed, with **Saluki** having the highest average engagement.

Dog breeds with the most tweet engagement		
Dog Breeds	Favorite count (Likes)	Retweet Count (Retweets)
Saluki	20218	4051
French Bulldog	15878	3848
Afghan Hound	14289	4697
Giant Schnauzer	14282	4043
Black-And-Tan Coonhound	14229	3252
Flat-Coated Retriever	14040	3637
Irish Water Spaniel	13593	3539
Standard Poodle	13283	5314
English Springer	13092	4777
Cardigan	12690	3568

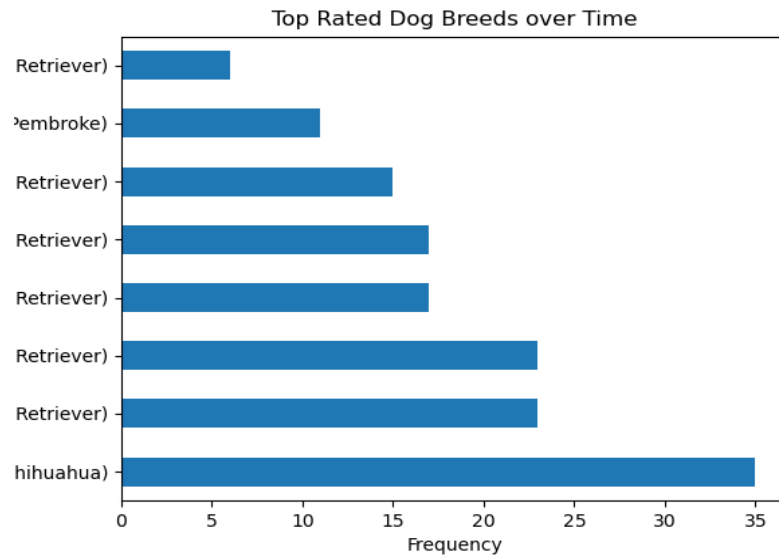
### 3. Correlation between likes and retweets and also between ratings and likes

We also found a fairly positive correlation between likes and retweets, as well as between ratings and likes, as shown by a scatter plot of these variables. This suggests that tweets with higher ratings and likes also tend to have more retweets and likes.



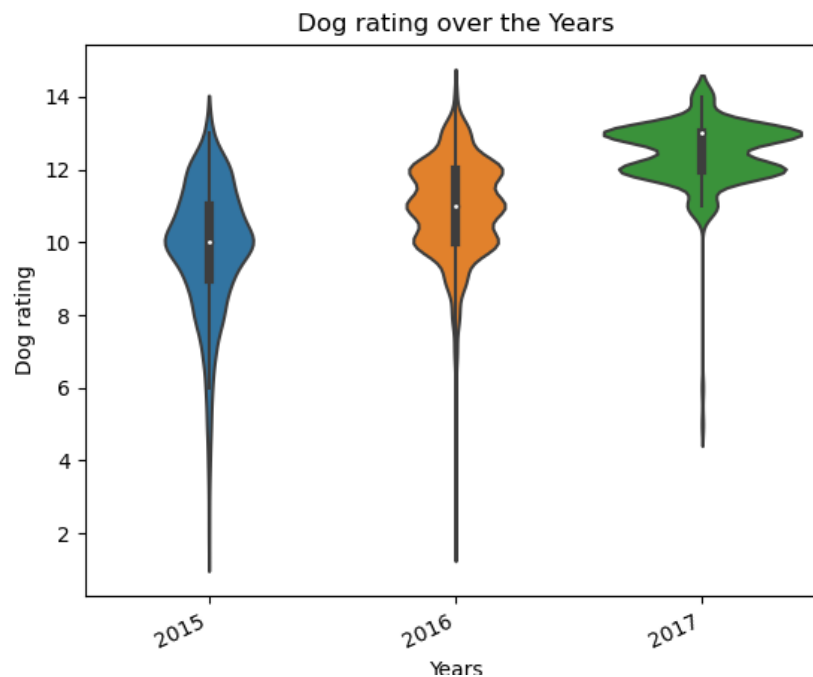
### 4. Most Popular dog breeds over time

Additionally, we observed that the most popular dog breed over time is the **Golden Retriever**, except in **2015Q4** and in **2017Q2** when **Chihuahua** and **Pembroke** were the most popular respectively. This is shown by a bar chart of the maximum number of tweets for each breed over time, where **Golden Retriever** has the highest, except for the mentioned outliers.



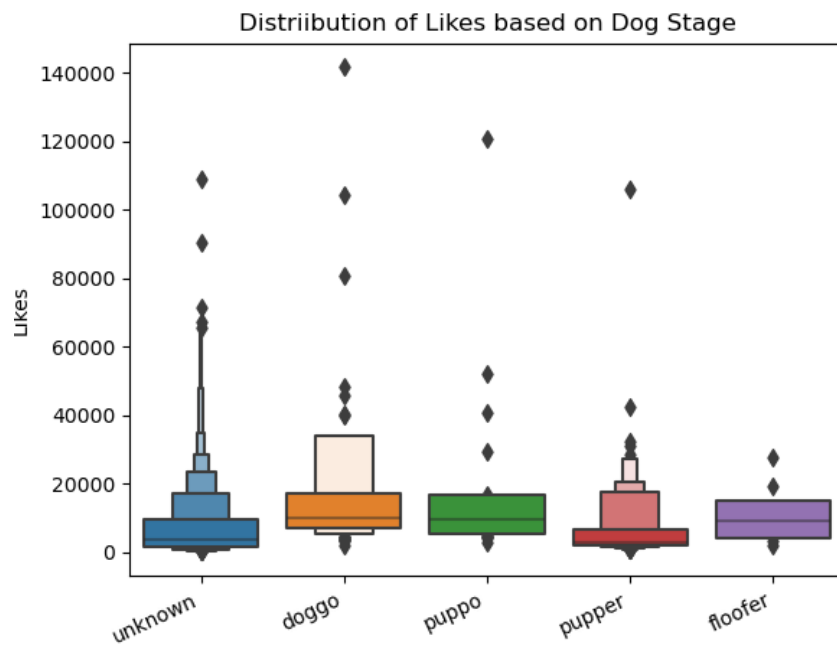
### 5. Dog Ratings over the years

Furthermore, it was found that the dog ratings were mostly between **6** and **14**, with a lower spike in **2015** and **2016**. However, this distribution of ratings tends to expand and increase in **2017**. A violin plot of the ratings over time shows this pattern.



### 6. Distribution of Likes based on Dog Stage

Lastly, it was observed that the dog stage of **Doggo** seems to be the most liked, while **Floofers** are few and least liked. This is supported by a boxen plot of the tweet likes for each dog stage, where **Doggo** has the highest, and **Floofer** has the lowest.



## Conclusion

In conclusion, the data analysis revealed that the Golden Retriever is the most featured dog breed in tweets and has the most engagement on average. There is a positive correlation between likes, retweets, and ratings. The popularity of dog breeds and ratings over time also showed interesting patterns. And, it also showed that Doggo is the most liked stage of a dog while Floofer is the least liked. The visualization of data supports all the insights and makes the analysis more comprehensive.