

“WeRateDogs” Twitter Data Wrangling

The screenshot shows the Twitter profile page for the account "WeRateDogs®". The profile picture is a close-up of two dogs' faces. The header text "WeRateDogs®" includes a blue verified checkmark icon. Below the header, it says "15.7K Tweets". On the right side of the profile page, there are two advertisements: one for "JustFood For Dogs" with the tagline "More life. More years. More love." and another for "trupanion™" with the tagline "Medical insurance for the life of your pet.". At the bottom of the profile page, there is a "Follow" button and three circular icons with ellipses, a mail icon, and a person icon.

WeRateDogs®

@dog_rates

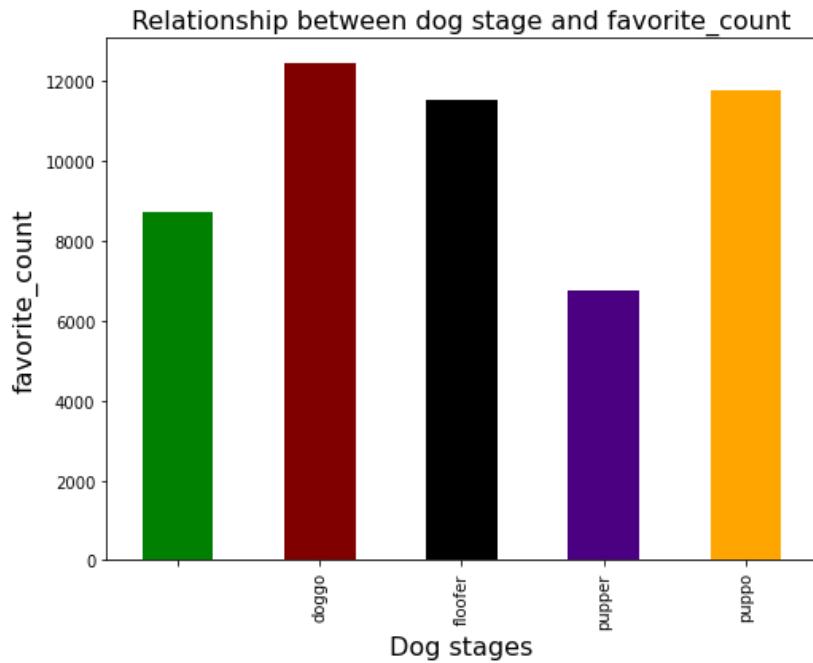
Your Only Source For Professional Dog Ratings Instagram and Facebook ➔
WeRateDogs partnerships@weratedogs.com | nonprofit: [@15outof10](#)

The aim of this project was to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations to test how well the data analysis students have mastered the data wrangling act taught in the Udacity classroom for the nanodegree program. The process involved the 3 steps always followed while wrangling datasets which include gathering, assessing and cleaning the acquired data.

I gathered 3 different datasets from different sources which included downloading a file manually and programmatically from the Udacity classroom and obtaining twitter a twitter JSON text file. The next step I did was assess the data programmatically for any abnormality and went straight to cleaning. I did a step by step cleaning process through defining the issues witnessed in the data during the assessment process, writing the right codes to help clean the data programmatically and finally testing to confirm the task performed was a success.

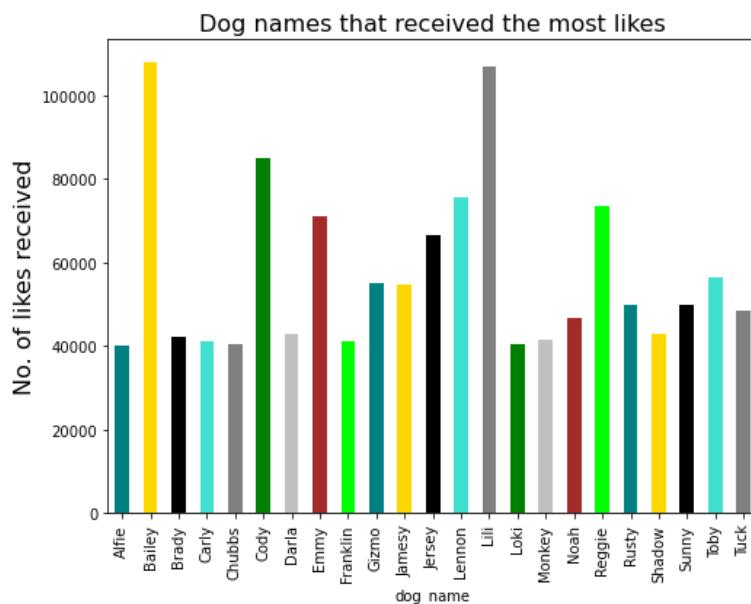
After the cleaning process I went into the exploratory data analysis and drew some insight and made conclusions on my findings. Below are some of the insights and visualizations I was able to come up with.

The image below shows a type of dog stage called puppo one of the most popular dog stages as shown through the analysis.



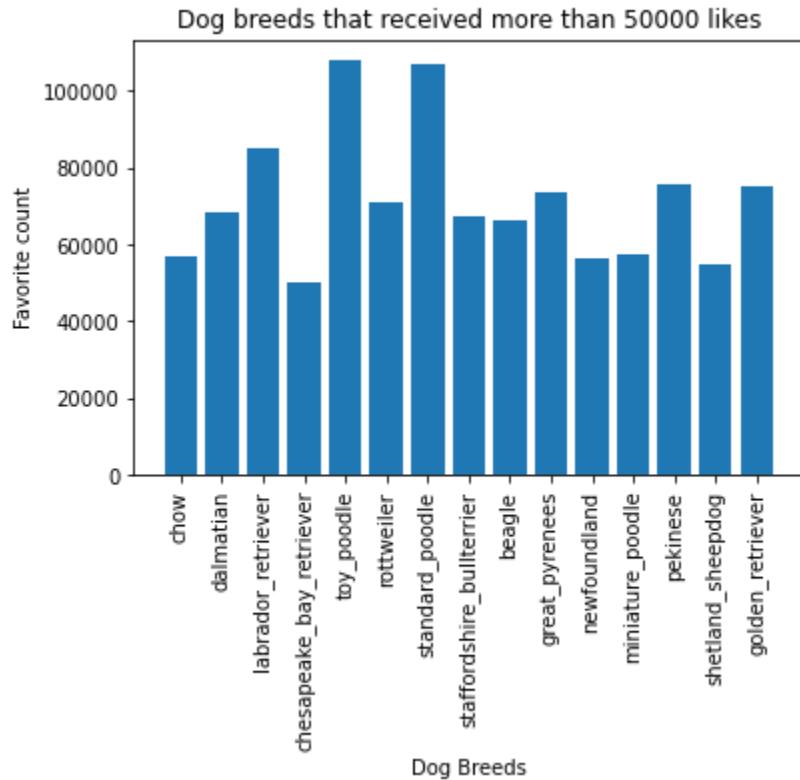
- The graph above clearly shows that the doggo and puppo are the most like dog stages since they received the highest likes in the "WeRateDogs" twitter report. The 1st bar shows dogs that were tweeted without listing.

I was able to look at the most liked dog names amongst the posts tweeted and the results are as shown below;



- b)The above graphs shows that the dogs posted on twitter and received the most likes, most of them were named bailey and lili, most people liked posts with dog names bailey and lili.

Doing a visualization on dog breeds gave a great insight into the type of breed that received the most likes.



- The visualization created in the above graphs shows the breeds of dogs that received the most likes all of which are above 50,000 likes. Toy poodles and standard poodles are the most liked followed by labrador retrievers, pekingese,golden retrievers and great pyrenees respectively.

a) Generally

The highest likes on the dog stage were received by doggo, puppo and fluffer respectively. While the most retweeted was pupper doggo and puppo respectively.

Hence Doggo seems to be the most popular dog stage since it appears to rate among the top in the most liked and most retweeted in the twitter posts despite pupper being the most tweeted.

Analysis on dog breeds indicate that chihuahua was the most retweeted breed followed by golden retriever and then samoyed, While the most posted breeds on twitter are the labrador retriever, golden retriever, samoyed, staffordshire bull terrier and toy poodle respectively. Toy poodles and standard poodles are the most liked breeds followed by labrador retrievers then pekingese, golden retriever and great pyrenees respectively.

From this observation it is safe to say the golden retriever is the most popular breed since it always finds itself among the top for the most liked and the most retweeted breed.

Below are some of the dog images that featured in the report, I have listed their names, with some of them having their breeds shown.

dog_name	dog_breeds	text_tweeted
Stuart	golden_retriever	This is Stuart. He's sporting his favorite fan...



dog_name	dog_breeds	text_tweeted
Kollin	chihuahua	Meet Kollin. He's a Parakeetian Badminton from...



dog_name	dog_breeds	text_tweeted
Gus	pupper	This is Gus. He's quite the cheeky pupper. Alr...



c) Conclusion.

From the above analysis we can make the following conclusions;

- The number of retweets tends to rise with the number of likes on a tweet.
- Doggo is the most popular dog stage since it ranks highest for likes and retweets.
- Golden retriever seems to be the most popular dog breeds since it always finds a way to make it to the top 5 of the list (likes and retweets).