

Summary Report

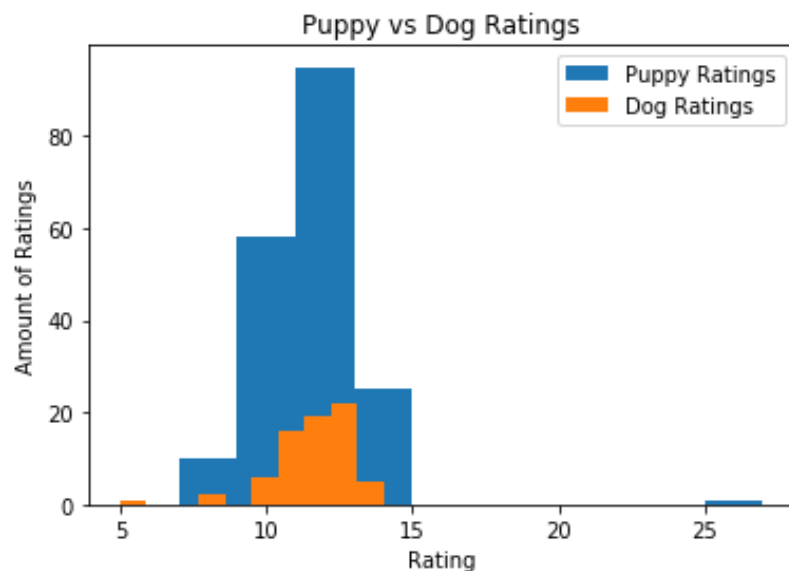
@WeRateDogs is a Twitter account that is known for its cute dog pictures, and its humorous way of rating dogs. Since they're usually always the best and cutest dogs, they often get ratings where the numerator is higher than 10. The site also categorizes their dogs as "puppo," "pupper," "floofer," and "doggo" based on their age group.

After gathering, assessing, and cleaning the data for this project, I had a few questions that I wanted to analyze.

What age group gets higher ratings?

It seems like puppies would get higher ratings than grown dogs, since they are so cute. So for my question, I wanted to see if that was actually true. I first used `pd.query` to query the dataset for "puppy" and "dog" categories. I did this to make my visualization simpler to view, since there are 4 dog stages. I then created a histogram called "Puppy vs. Dog Ratings," which plotted the puppies and dogs against the rating and number of ratings they go.

After plotting and viewing the histogram, it looks like puppies do indeed get a much higher number of overall ratings. However, the ratings themselves are pretty close to each other. While the dog ratings are more closely grouped, the puppy ratings have more spread. However, there are some outliers in this dataset as well.



What dog stage gets the best ratings?

Next, I wanted to see if there is a difference in the ratings for each individual dog stage, and find out if puppies still got the highest ratings. For this question, I decided to compare the mean numerator ratings for each dog stage, so that I could get a general overview of the average rating for that group.

I first queried my data frame to get the individual dog age groups. I then found the mean rating value for each group, and compared. Based on the results, it looks “floofer,” which I had put in the dog category, actually had the highest average ratings. “Puppo” came next, followed by ‘doggo.’ But what surprised me was the “pupper” category actually had the lowest average ratings of all the groups. I am not sure if this is due to the puppy category having a higher number of ratings than the dog category, from what I found earlier. But either way, that was surprising as I had expected the two puppy groups to have the highest average ratings.

What is the most popular dog breed?

I then wanted to see what was the most common breed on @WeRateDogs, since there are so many posts all about them. To do this, I got the value counts of all the dogs, using the ‘dog_breed’ category in my data frame.

After looking at the top results, it looks the golden retriever is definitely the most common breed, since it has the most results, followed by the labrador retriever and pembroke.

Do the most common breeds receive the highest number of retweets?

After finding out the most common breed, I wanted to see if those same dogs get the most number of retweets, since they seem to be popular. To find this out, I first grouped the dogs by breed, and then grouped them into “ratings” and “retweets” columns, and sorted them in descending order.

After comparing the two columns, it looks like the top retweeted dogs are still the golden retriever and labrador retriever. However, chihuahuas tied with Pembroke’s this time, for the third highest retweeted breed.