

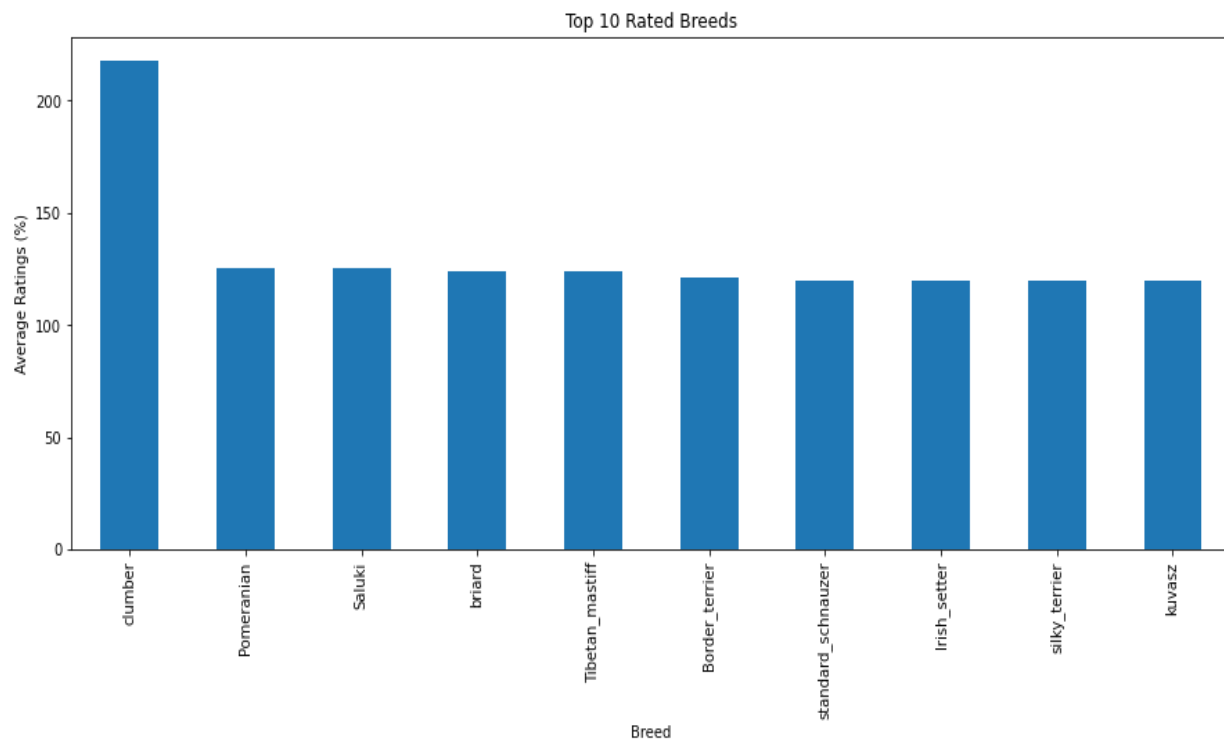
Act Report: Analyzing the WeRateDogs Twitter Archive Data

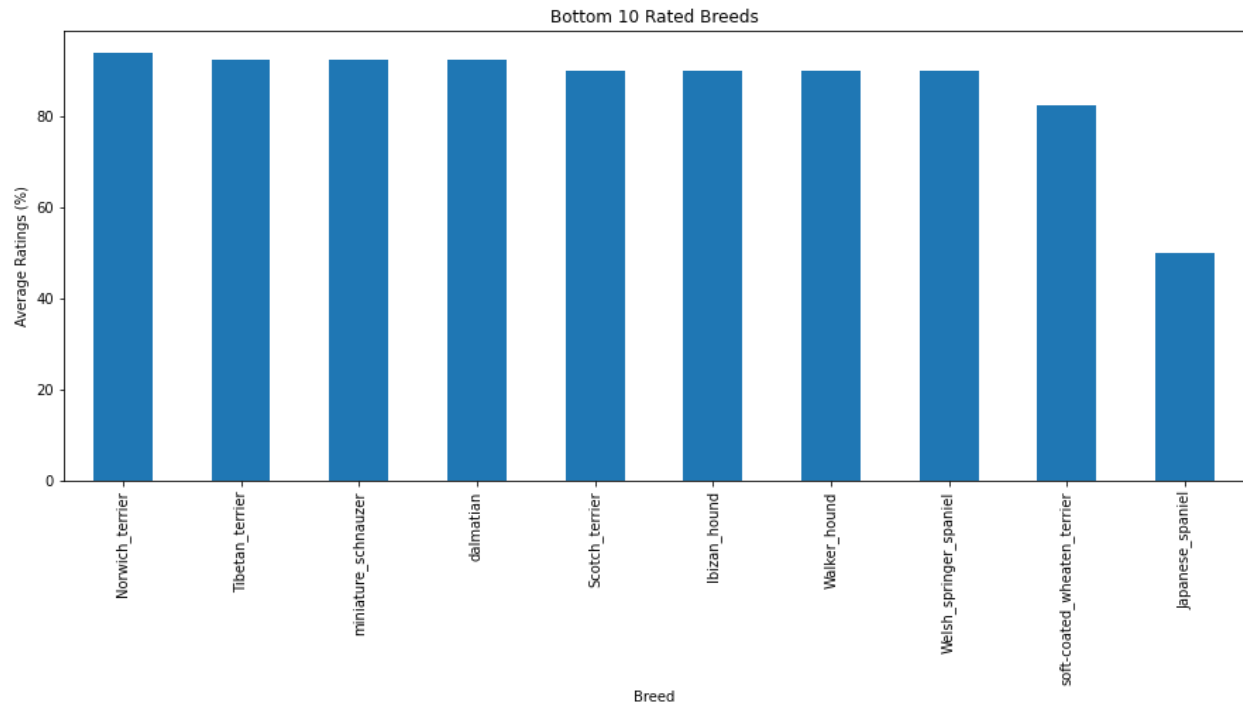
The objective of this section is to analyze and visualize the wrangled WeRateDogs twitter archive data to draw insights. For this analysis, I combine tools in pandas, matplotlib and seaborn to answer some questions about the data. To achieve the objective of this analysis, the following research questions are explored:

1. Which dog breed received the highest rating over the period?
2. What are the characteristics of highly rated dog breeds?
3. Based on rating, favorite counts and retweet counts, which breed is the best?

RQ #1: Which dog breed received the highest rating over the period?

We compare the average rating of each dog breed over the period.



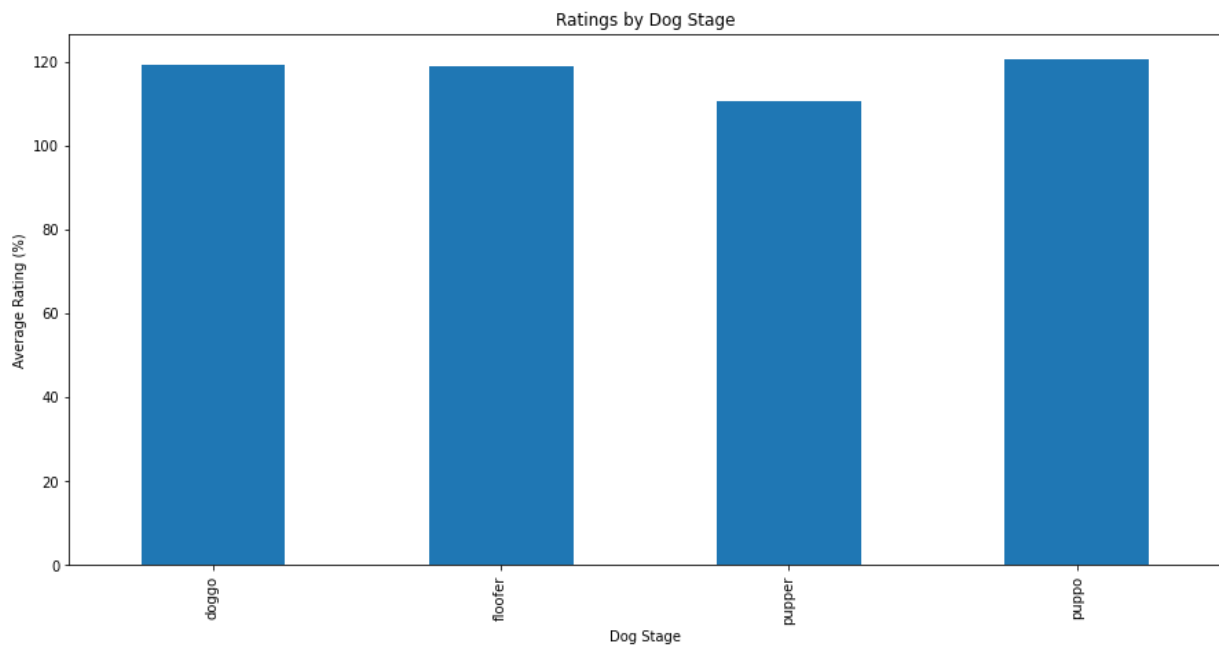
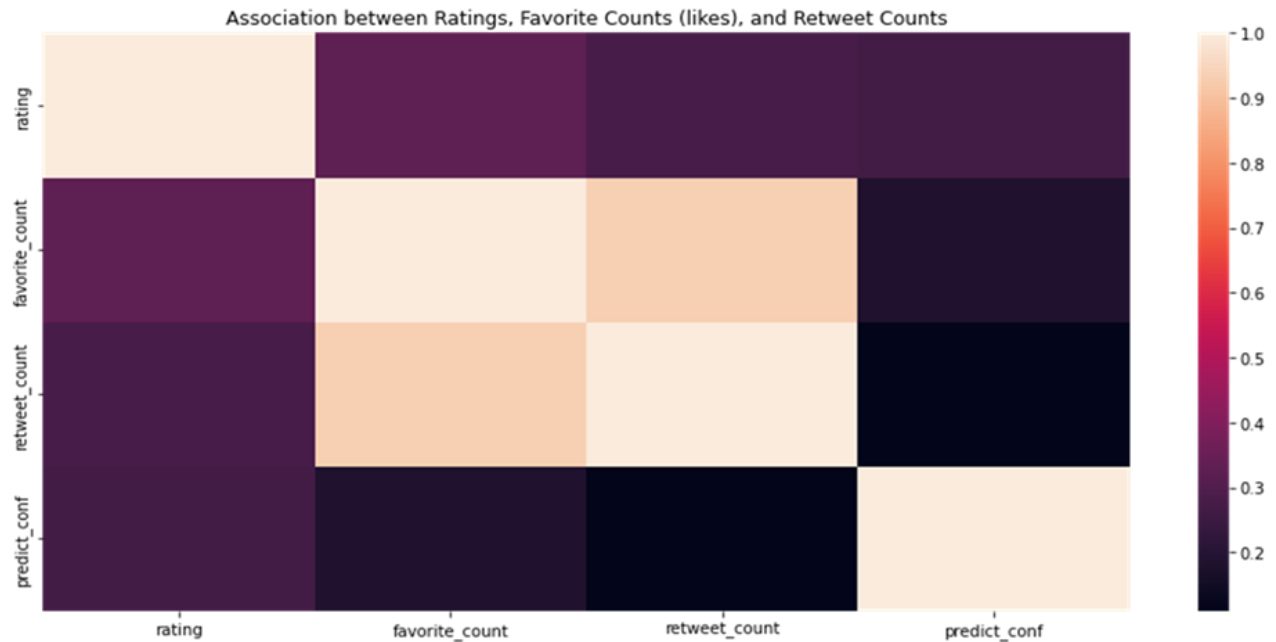


Observations:

- There are 111 different dog breeds in the data
- The average rating of dog breeds ranges from 50% to 217.57% with a mean of 108.13%
- On average, Clumber (rating 217.57%) is the highest rated breed followed by Pomeranian (125.36%) and Saluki (125%)
- Japanese-spaniel (rating 50%) has the lowest rating followed by Soft-Coated-Wheaten-Terrier (82.5%)

RQ #2: *What are the characteristics of highly rated dog breeds?*

Here we want to find the features of breeds with high ratings. Specifically, we want uncover the relationship between ratings and favorites counts, retweet counts, prediction confidence and dog stage.

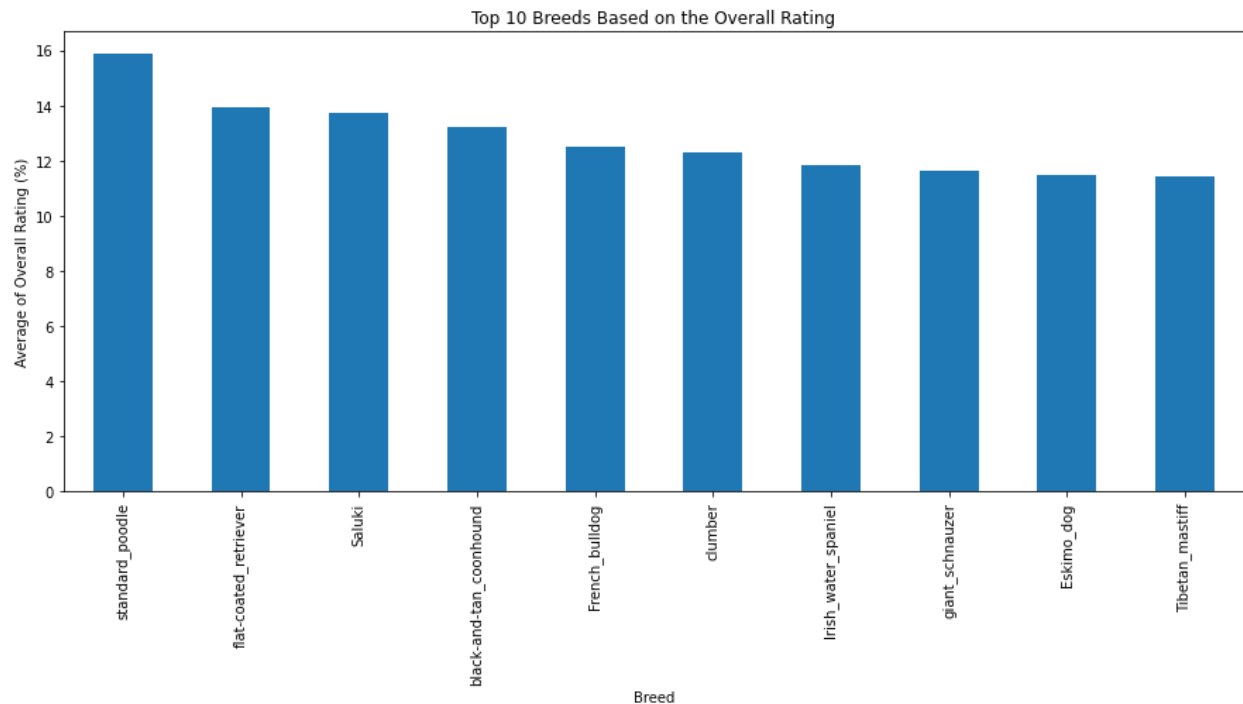


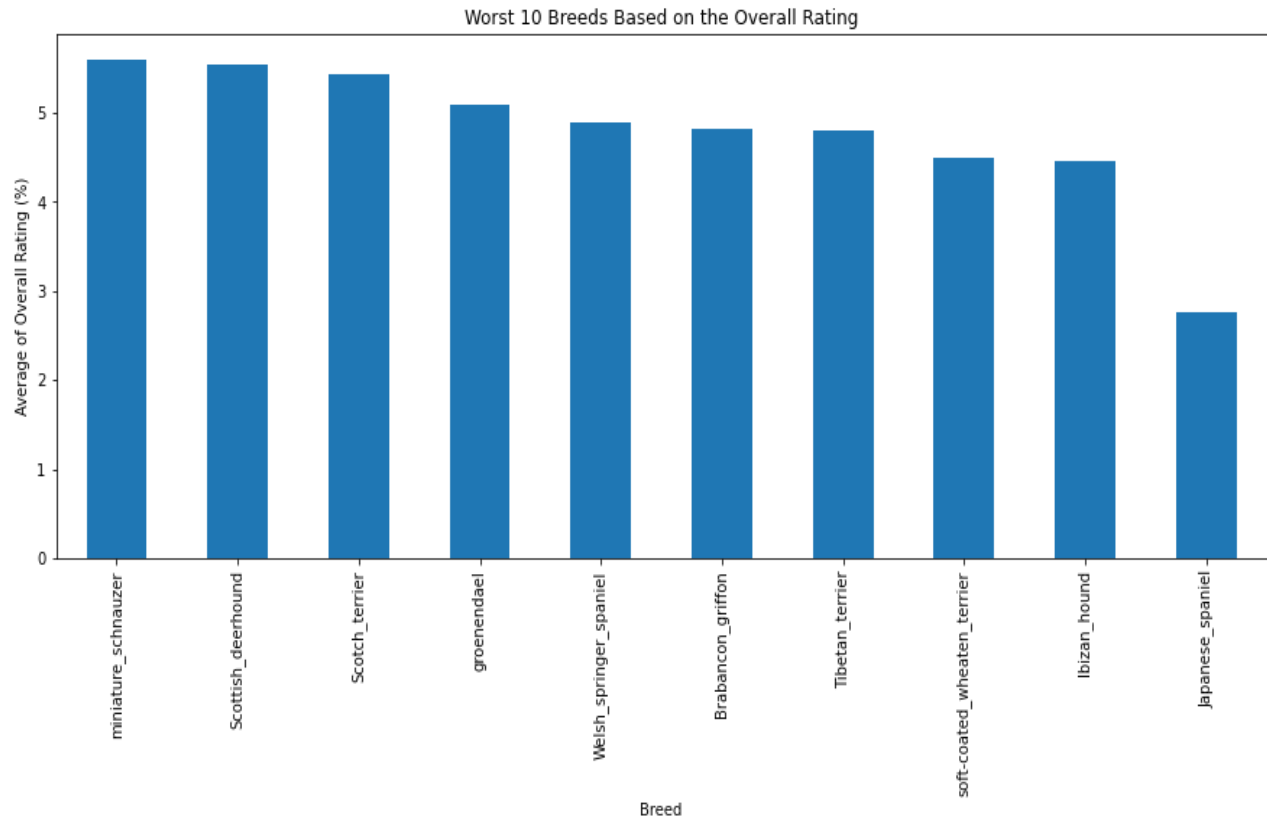
Observations:

- We observe that rating is positively associated with favorite_count, retweet_count, and predict_conf
- While rating is moderately correlated with favorite_count and retweet_count, it has a very weak correlation with predict_conf

- favorite_count and retweet_count are highly correlated.
- After grouping by dog breed, the correlation coefficient for all the variables have increased.
- Also, the correlation between rating and prediction confidence has increased drastically.
- puppo has the highest rating (120.5%) among the dog stages.
- Except pupper, notice that the difference in rating between the dog stages is almost negligible.

RQ #3: Based on rating, favorite counts and retweet counts, which breed is the best?





Observations:

- Standard_poodle tops the chart as the overall best breed.
- Flat-coated-retriever and Saluki are at second and third place respectively.
- The Japanese-spaniel is the worst breed based on the overall rating

The analysis reveals that:

- On average, based on only WeRateDogs rating, the Clumber breed (rating 217.57%) is the highest rated breed followed by Pomeranian (125.36%) and Saluki (125%). Japanese_spaniel (rating 50%) has the lowest rating followed by Soft-Coated-Wheaten-Terrier (82.5%)
- However, with the three metrics combined (i.e., rating, likes, and retweet count), Standard-poodle tops the chart as the overall best breed. The Japanese-spaniel was still the worst breed based on the overall rating.
- Furthermore, high WeRateDogs ratings was found to be associated with high levels of favorite counts (likes), retweet counts, and image prediction confidence.