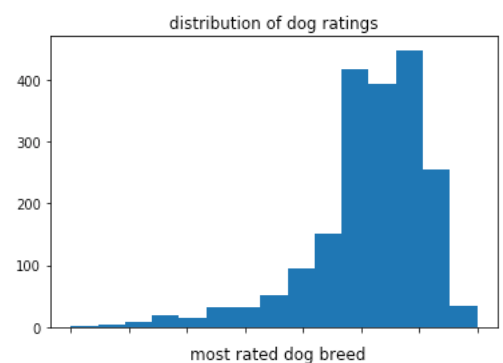# Wrangle and Analyze Data
Insights

by Markus Müller

This is a short analysis of the Dataset that was created from wrangling Datasets related to the Twitter page @WeRateDogs.

The analysis is possible since we cleaned the data to a point where we can use it to create insights. In the following I will answer questions related to the Dataset. The analysis is based on 1957 datapoints.
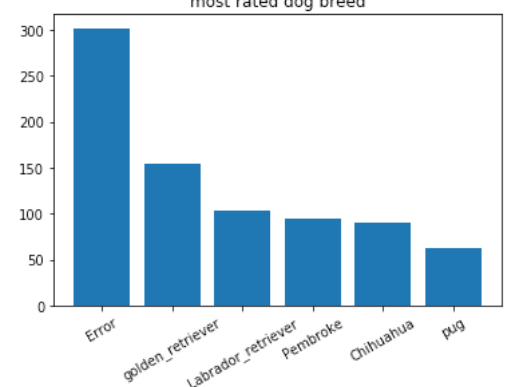

distribution of dog ratings

1. What is the distribution of the Rating?

  The distribution is skewed to the left with the most rated dogs in the area of 9 to 13, an average rating of 10.55 in the range from 0 to 14.

most rated dog breed

2. What is the most rated dog breed?

  The most rated dog breed is the Golden Retriever. Error shows that the neural net could't identify a dog breed.

3. What is the best and wors rated dog breed on average?

  The best rated dog breed is the Bouvier des Flandes (13.00) and the worst is the Japanese Spaniel (5.00).

4. What are the top 5 dog names?

  The top 5 dog names in this dataset are: Charlie, Cooper, Oliver, Lucy and Tucker

5. Are retweet count and favorite count related to each other?

relation retweets and favorite counts

  Yes, retweet and favourite count are related to each other. In fact they have a strong positive corrleation of r = 0.91, which means that retweet count gets higher with higher favorite count and vice versa.