

Udacity Project:

Wrangling and Analyzing Data

by Stefan Cornelißen, 2018-01-23

Act Report - WeRateDogs Analysis

This project aimed to analyze data from the WeRateDogs Twitter account. On this account pictures of funny or cute dogs are published and rated. The analysis seeks to answer questions like “Which dog breed has achieved the highest rating, or which dog breed earned most favorites or retweets?” The analysis is based on three main elements:

1. The rating from WeRateDogs. A cute dog in a funny situation may achieve 12 out of 10 points. The unique rating system is a special characteristic of this Twitter account.
2. The second element represents the dog breed. Each picture was analyzed regarding the breed of the dog and other elements or things within the picture.
3. The third element is the performance data of a tweet. The metrics “retweet and favorite count” indicate how well the audience liked a certain tweet respectively the dog picture. The higher the value, the more popular the Tweet has become.

Analysis and Insights

The following questions were analyzed within this act.

Question 1: What were the 10 highest rated dog breeds on average, according to the prediction with the highest confidence level?

Breed	Avg. Rating
Saluki	1.25
Briard	1.23
Tibetan Mastiff	1.23
Border Terrier	1.21
Silky Terrier	1.20
Standard Schnauzer	1.20
Eskimo Dog	1.18
Gordon Setter	1.18
Irish Setter	1.18
Samoyed	1.17

The highest rated dog breed is the Saluki. Its average rating is 1.25 or 12.5/10. Furthermore, the distribution of the average ratings per dog breed is close to each other. There is no breed which is rated far better than the rest.

Question 2: Which are the top 10 predicated dog breeds in terms of the highest confidence-levels on average?

Breed	Avg. Prediction Confidence Level
Bernese Mountain Dog	0.65
Komondor	0.52
Samoyed	0.52
Pembroke	0.50

Pug	0.48
Blenheim Spaniel	0.48
Golden Retriever	0.45
Dalmatian	0.40
German Shepherd	0.37
Vizsla	0.36

The highest average confidence level in predicting the dog breed experienced the Bernese Mountain Dog. Second is the Komondor. Both breeds have a unique appearance, which probably made the prediction easier. The highest confidence level is 65% on average, which should be kept in mind. In my opinion, the level is too low, to really rely on these results. For deeper analysis, you should consider running the images through a model which is either dedicated to dogs only or which is trained better.

Question 3: Which dog breeds have the highest average retweet and favorite-count, according to the prediction with the highest confidence level?

Breed	Avg. Retweet Count
Standard Poodle	6284
English Springer	5658
Afghan Hound	5635
Eskimo Dog	5182
Giant Schnauzer	4845
Saluki	4841
Great Pyrenees	4703
French Bulldog	4607
Lakeland Terrier	4577
Flat-Coated Retriever	4278

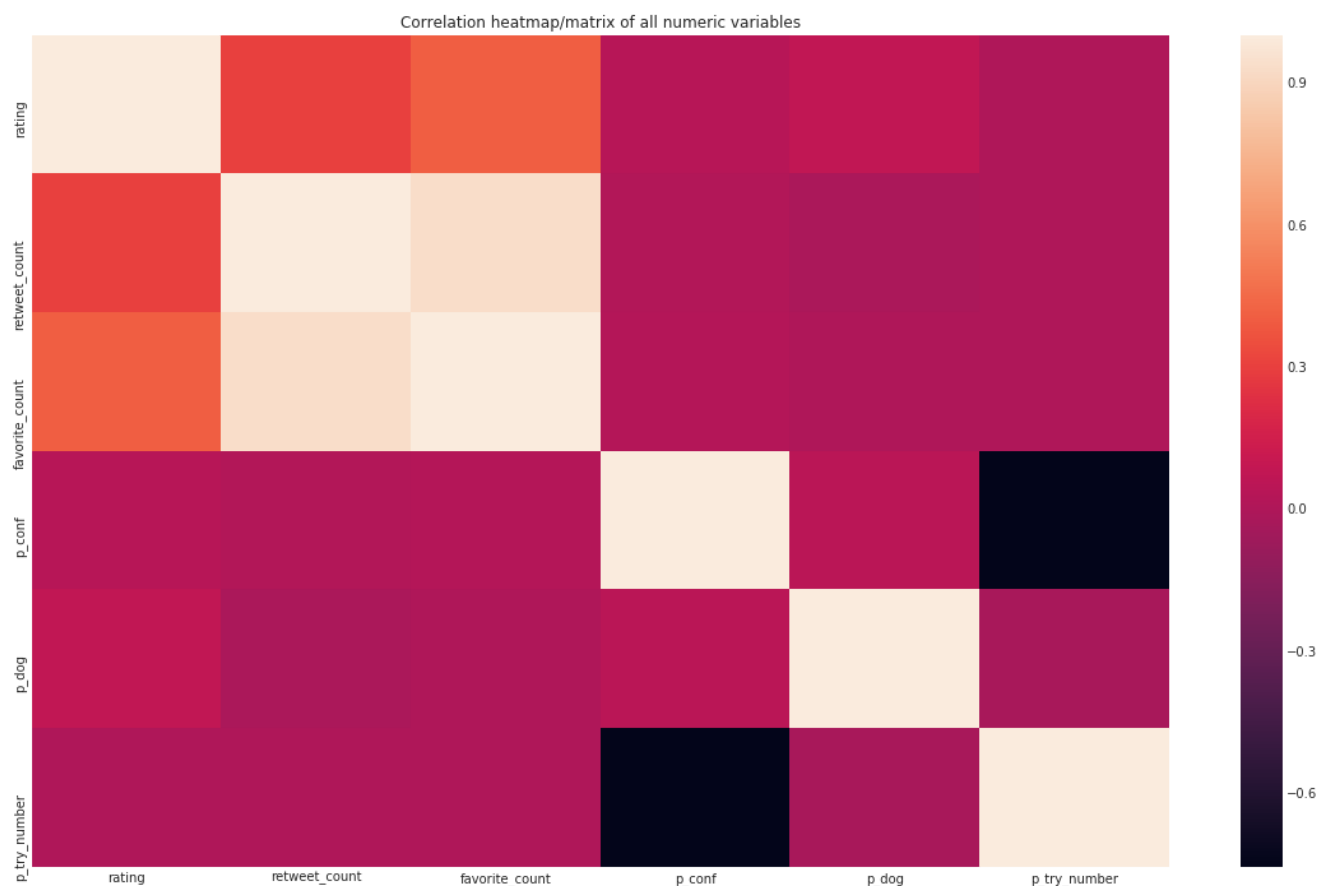
Breed	Avg. Favorite Count
Saluki	23210
French Bulldog	18312
Giant Schnauzer	16554
Afghan Hound	16549
Black-And-Tan Coonhound	16360
Flat-Coated Retriever	16155
Irish Water Spaniel	15703
Standard Poodle	15182
English Springer	15065
Cardigan	14673

On average the breed Standard Poodle appears to be retweeted most, whereas the Saluki appears to have earned the most favorites by far. However, you can see, that many breeds are in both top ten listings present. This might be a hint of a potential correlation between these variables.

Question 4: Does the rating correlate with the number of retweets or favorite-counts?

No, the variable rating does neither correlate strongly with the variable retweet_count (0.30), nor with the variable favorite_count (0.41). However, the variables retweet and favorite count correlate strongly with each other (0.93). It appears the audience of the WeRateDogs Twitter account express their liking using both ways.

The heatmap below shows the correlation coefficient ranges among all variables.



Question 5: Which items were also visible in the pictures and what are their frequencies?

The most frequent item within the dog pictures appears to be a "seat belt", followed by teddies. The second item is classified as a Dingo, which actually might come down to a misinterpreted dog breed.

