

Analysis of WeRateDogs Dataset

Jie Gu
03/02/2019

Introduction

The dataset is the tweet archive of Twitter user [@dog_rates](#), also known as WeRateDogs](<https://en.wikipedia.org/wiki/WeRateDogs>). WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "[they're good dogs Brent](#)." WeRateDogs has over 4 million followers and has received international media coverage.

In the following, I will answer three questions that I'm interested in:

1. Is there a correlation between the rating and number of retweets and favorites?
2. Which breeds of dogs is the most popular (appear the most in the dataset)?
3. Which breeds of dogs are the most favored (get the highest ratings, retweets or favorites)?

Data Analysis and Conclusions

Question 1

The pair plot of the three variables is given in Fig. 1.

Along with the correlation matrix (Fig. 2), it can be seen that the number of retweets and favorites are strongly positively correlated. However, the rating is only moderately positively correlated with the other two variables.

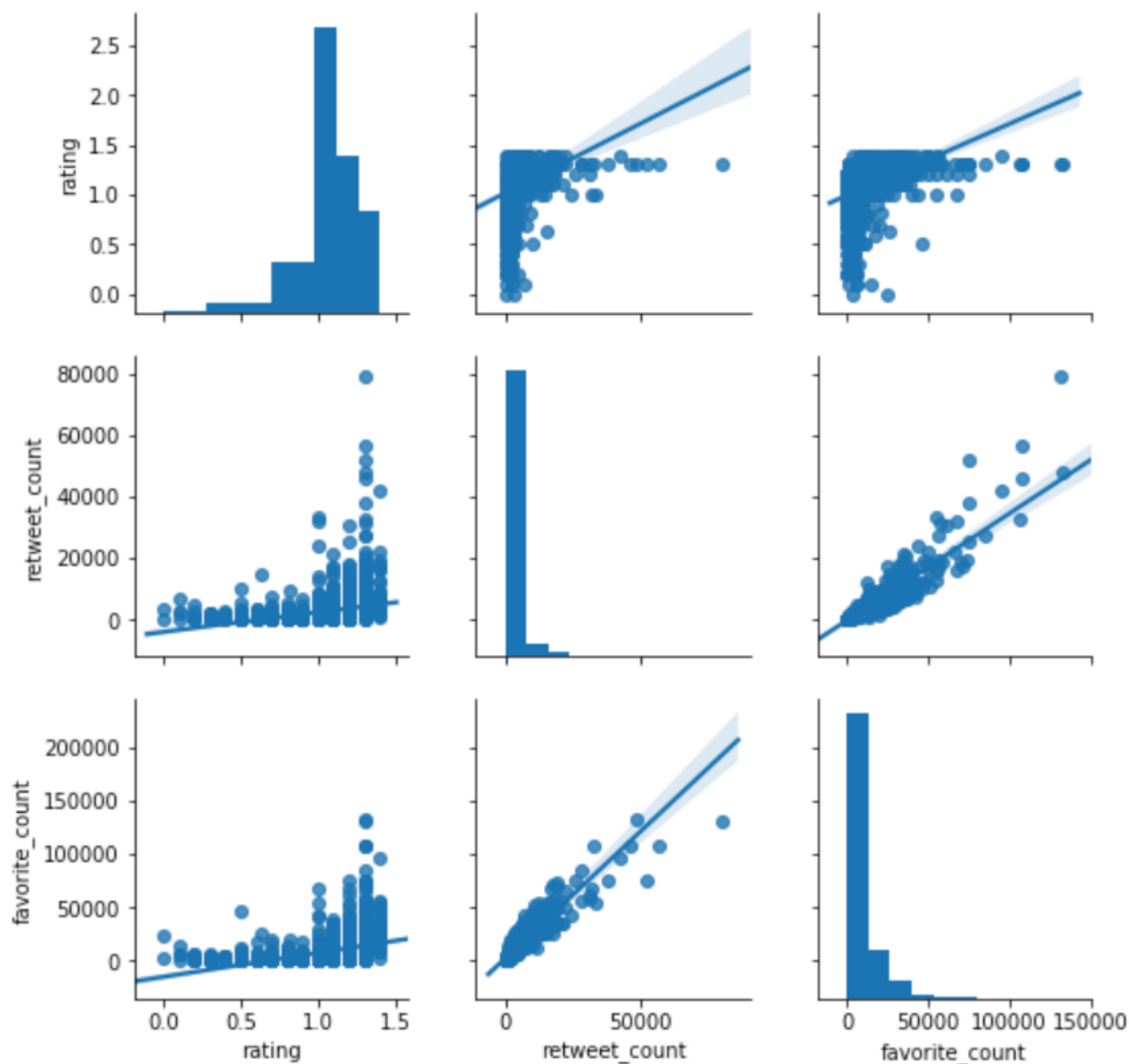


Fig. 1: Pair plot of rating, number of retweets and number of favorites

	rating	retweet_count	favorite_count
rating	1	0.299741	0.402516
retweet_count	0.299741	1	0.912953
favorite_count	0.402516	0.912953	1

Fig. 2: Correlation matrix of rating, number of retweets and number of favorites

Question 2

Golden retriever, labrador retriever, pembroke, chihuahua and pug are the most popular breeds.

Question 3

The most liked breeds, with different metrics used, are:

- rating: pomeranian, kuvasz, eskimo dog
- number of retweets: eskimo dog, great_pyrenees, french_bulldog
- number of favorites: french_bulldog, cardigan, basset

French bulldogs are the most liked breed if all the three metrics are considered.