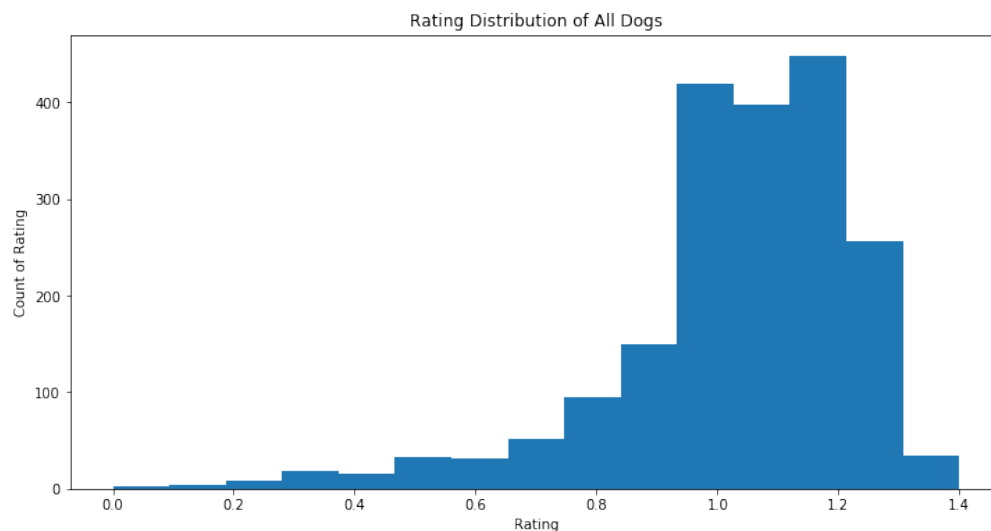# Analyze and Visualize Data: WeRateDogs

Introduction

This project is about data wrangling and visual analysis of data from the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. WeRateDogs has over 4 million followers and has received international media coverage.

Insights
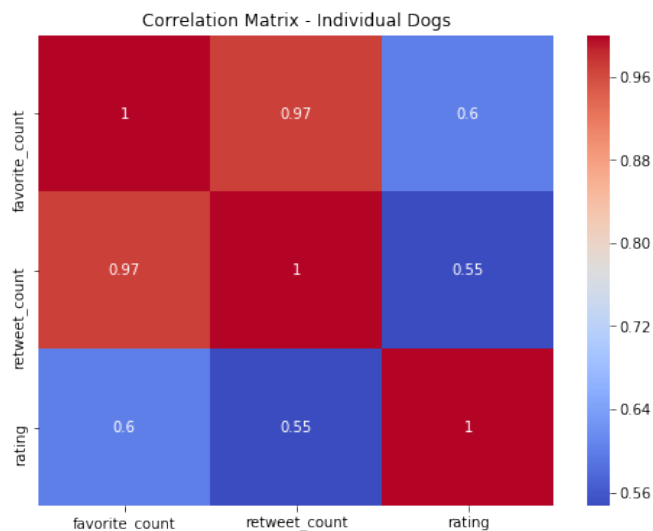
Rating Distribution

The rating distribution of all individual dogs showed that most of the ratings are between 10/10 and 12/10. Supposedly all dogs are good dogs, though there are a few cases with very low ratings. I clicked into these original tweets and realized the low ratings are really not the dogs' fault. For example, the person who sent the photo to @dog_rates forgot to include the dog in the picture. As a result, these types of tweets are given a rating as low as 0 or 1.
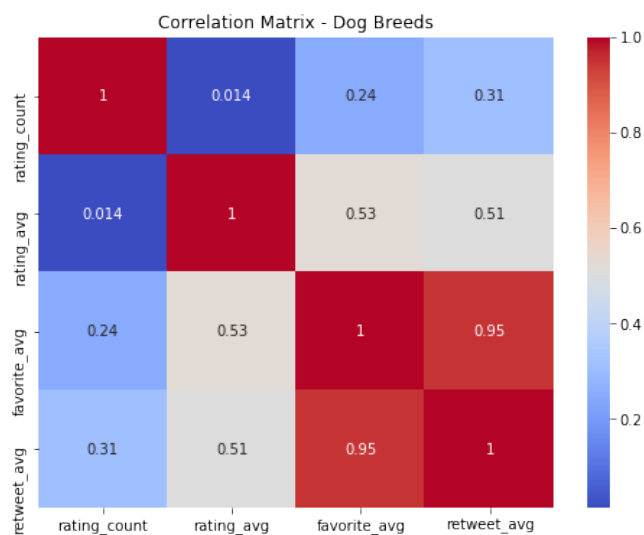

Correlation

I calculated the Spearman's correlation matrix as I'm not sure if the relationship is supposed to be linear. Let's first look at the correlation matrix based on individual dogs' data, we can see there is a strong positive relationship between favorite counts and retweet counts (0.97). This

indicates that when a tweet gains more favorites, it's likely to be retweeted more often and vice versa.

Correlation Matrix - Individual Dogs

Then let's look at the correlation matrix based on aggregated data at dog breed level. There is also a strong positive relationship between average favorite counts and average retweet counts per each dog breed. The relationships between rating counts and average rating/favorite/retweet are relatively weak, which indicates a commonly seen dog breed doesn't necessarily gain a lot of excitement from the crowd on the internet.

Correlation Matrix - Dog Breeds

Most Rated Dogs vs Most Liked Dogs

The most commonly seen and rated dog breed is golden retriever with more than 150 ratings. The runner-ups are breeds like Labrador retriever, Pembroke and Chihuahua.

dog_breed	rating_count	rating_avg	favorite_avg	retweet_avg
golden_retriever	154	1.1552	11211.0	3192.0
Labrador_retriever	103	1.1165	10630.0	3243.0
Pembroke	95	1.1411	9990.0	2645.0
Chihuahua	90	1.0556	7421.0	2316.0
pug	62	1.0242	5514.0	1620.0
toy_poodle	50	1.1020	5800.0	1903.0
chow	47	1.1404	8635.0	2403.0
Pomeranian	42	1.0929	6904.0	2320.0
Samoyed	41	1.1732	12315.0	4033.0
malamute	33	1.0879	9689.0	2845.0
Chesapeake_Bay_retriever	31	1.0839	9005.0	2313.0
cocker_spaniel	30	1.1333	12338.0	4205.0
French_bulldog	30	1.1167	18360.0	4676.0
miniature_pinscher	25	1.0000	7245.0	1961.0
Eskimo_dog	22	1.1409	11610.0	4120.0

The most liked dog breed is, surprisingly, saluki, although this breed only has 4 ratings with an average rating of 12.5/10, which is also the highest average rating among all dog breeds. The runner-ups include French bulldog and bouvier des flandres.

dog_breed	rating_count	rating_avg	favorite_avg	retweet_avg
Saluki	4	1.2500	21754.0	4384.0
French_bulldog	30	1.1167	18360.0	4676.0
Bouvier_des_Flandres	1	1.3000	16101.0	3802.0
Afghan_hound	3	0.9667	15426.0	5068.0
black-and-tan_coonhound	2	1.0500	15332.0	3522.0
flat-coated_retriever	8	1.0875	15165.0	3928.0
Irish_water_spaniel	3	1.1333	14644.0	3843.0
Leonberg	3	1.1333	13282.0	3283.0
whippet	11	1.0182	13213.0	4362.0
English_springer	10	1.0500	12712.0	4661.0
standard_poodle	11	1.0727	12518.0	4742.0
cocker_spaniel	30	1.1333	12338.0	4205.0
Samoyed	41	1.1732	12315.0	4033.0
Border_terrier	7	1.2143	12209.0	2899.0
Mexican_hairless	7	0.9857	12171.0	3636.0

The highest rated dog breed, as already mentioned above, is saluki. And then, runner-ups include briard, Tibetan mastiff and border terrier.

dog_breed	rating_count	rating_avg	favorite_avg	retweet_avg
Saluki	4	1.2500	21751.0	4382.0
briard	3	1.2333	8475.0	2528.0
Tibetan_mastiff	4	1.2250	10820.0	3558.0
Border_terrier	7	1.2143	12204.0	2899.0
silky_terrier	1	1.2000	6516.0	1601.0
standard_schnauzer	1	1.2000	1792.0	769.0
Gordon_setter	4	1.1750	4880.0	2124.0
Irish_setter	4	1.1750	9425.0	2874.0
Samoyed	41	1.1732	12313.0	4032.0
golden_retriever	154	1.1552	11209.0	3191.0
wire-haired_fox_terrier	2	1.1500	7588.0	2525.0
giant_schnauzer	4	1.1500	8369.0	2467.0
Australian_terrier	2	1.1500	9972.0	2654.0
Norfolk_terrier	7	1.1429	5016.0	1653.0
Great_Pyrenees	14	1.1429	11628.0	4040.0

It's interesting to learn that the most popular and cutest breeds (at least in my mind) like golden retriever and corgi are not the highest rated breeds by @dog_rates, nor the most favorite breeds by the broad audience of @dog_rates. As shown below, the top 15 most rated dog breeds don't necessarily have a corresponding high average rating or favorite/retweet count.