

Insights and Visualizations from Data Wrangling WeRateDogs

Introduction:

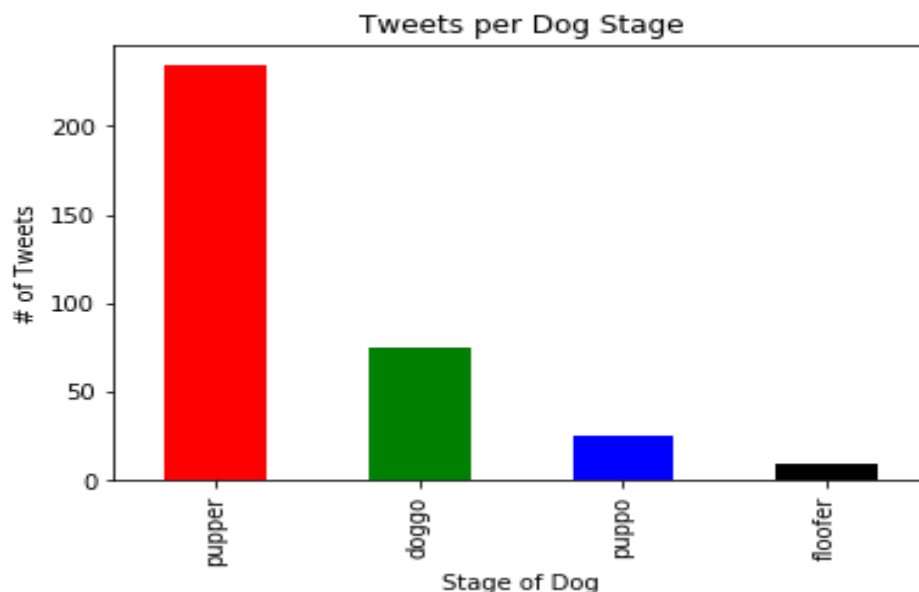
While gathering, assessing and cleaning data is imperative, analyzing and visualizing the cleaned data is just as important in understanding what your data is trying to tell you.

In the first portion of the analyzing the data, I wanted to get a high level view of the data by using the `.describe()` function on the new data set.

	tweet_id	rating_numerator	rating_denominator	max_conf	retweet_count	favorite_count
count	2.175000e+03	2175.000000	2175.0	1686.000000	2175.000000	2175.000000
mean	7.371205e+17	9.438621	10.0	0.549936	2757.987126	8746.022989
std	6.748668e+16	1.474649	0.0	0.299270	4646.599662	12052.503836
min	6.660209e+17	0.000000	10.0	0.000010	0.000000	52.000000
25%	6.768432e+17	10.000000	10.0	0.302892	608.000000	1911.500000
50%	7.098528e+17	10.000000	10.0	0.550604	1341.000000	4042.000000
75%	7.894226e+17	10.000000	10.0	0.821962	3220.000000	11107.000000
max	8.924206e+17	10.000000	10.0	0.999956	79515.000000	132810.000000

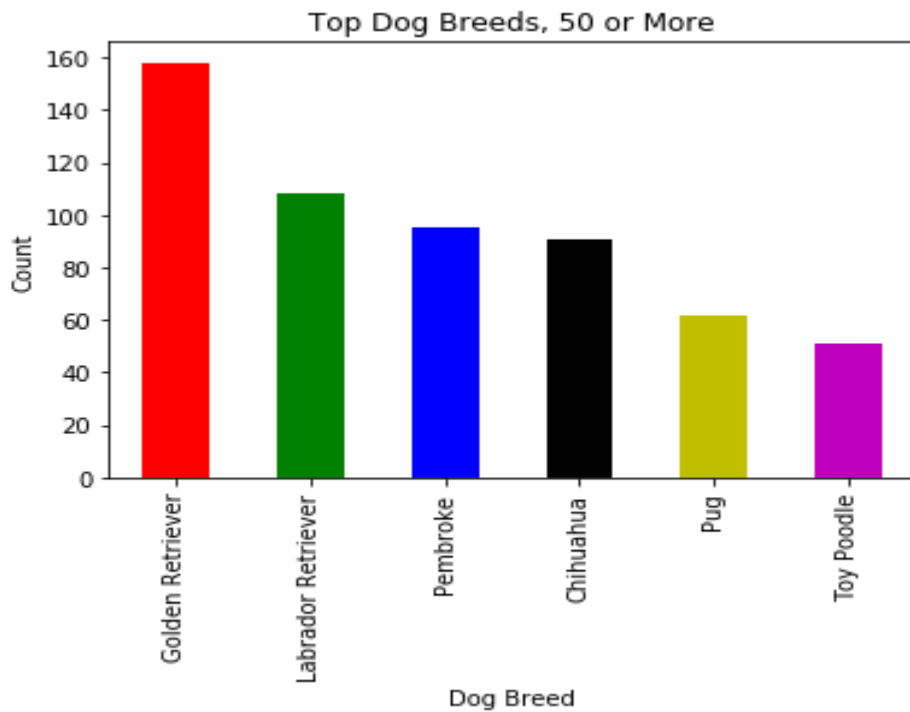
After viewing the combined data, I wanted graph out some comparisons to see which breed of dog was the most favored, top dog breed and stage of dog.

Tweets per Dog Stage



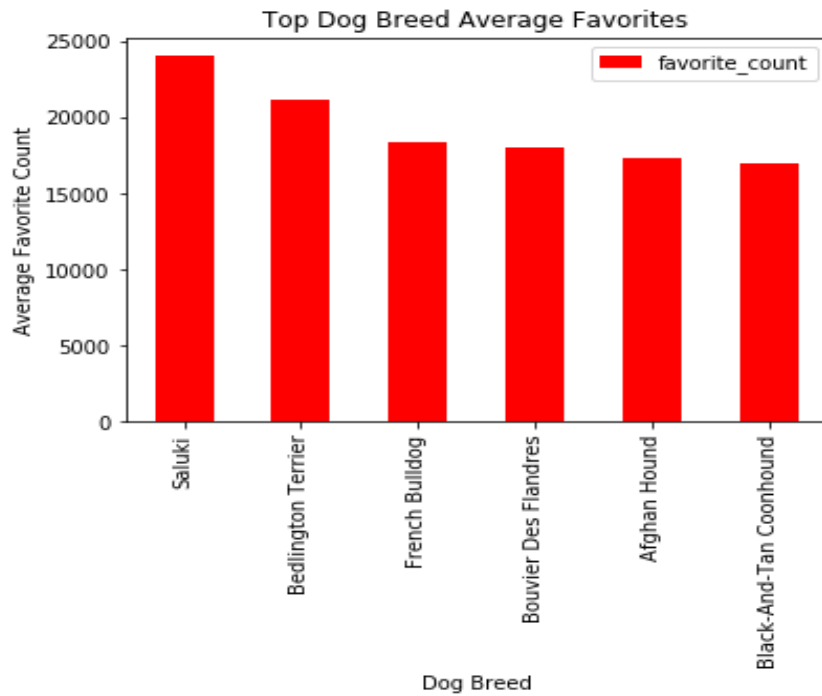
I filtered out the stages that did not have data 'NONE' and looked at the most tweets between all 4 stages, Pupper, Doggo, Puppero, Floofer. The Pupper stage was the most tweeted dog stage among the dataset.

Top Dog Stage with 50 or More Tweets



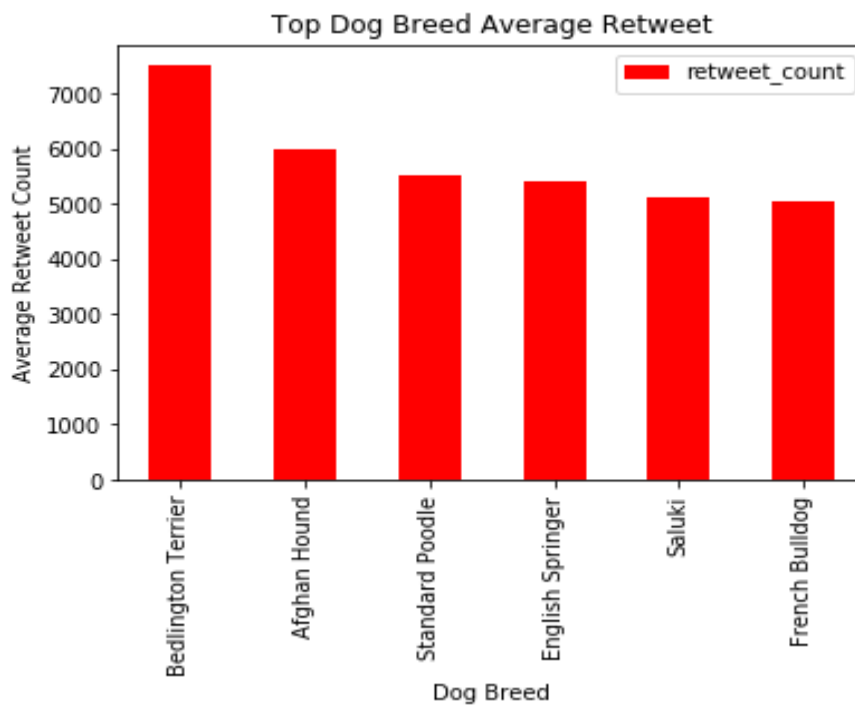
Next, I wanted to view which dog breeds were the most tweeted among the dataset, to get a good starting point for the rest of my analysis. I chose 50 or more tweets due to the amount of dog breeds and wanted to get a quick, easy view. The Golden Retriever was the most tweeted dog breed.

Most Favored Top Dog Breeds



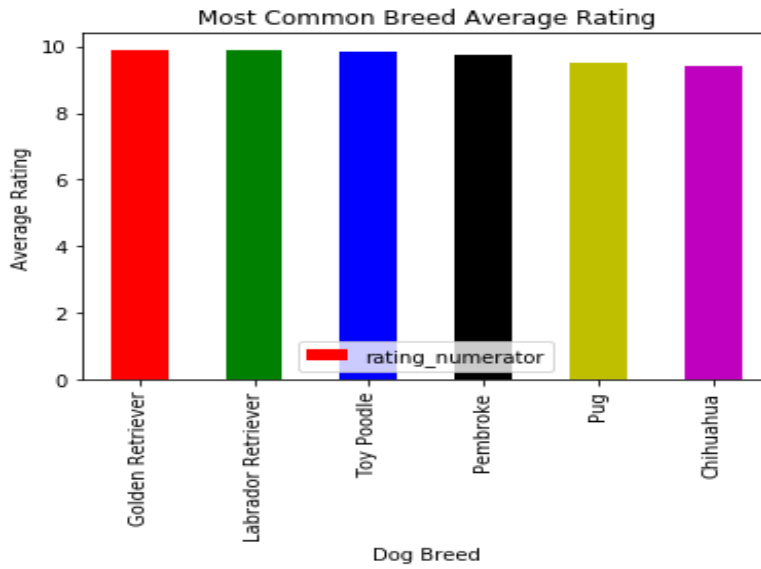
I took the same approach to the previous, with just looking at 6, easier view. The top dog breed that, on average, was favorited much more was the Saluki.

Most Retweeted Top Dog Breeds



Even though the Saluki was the most favorite dog breed, the Bedlington Terrier is most, on average, retweeted among all the dog breeds.

Average Rating for the Most Common Breeds

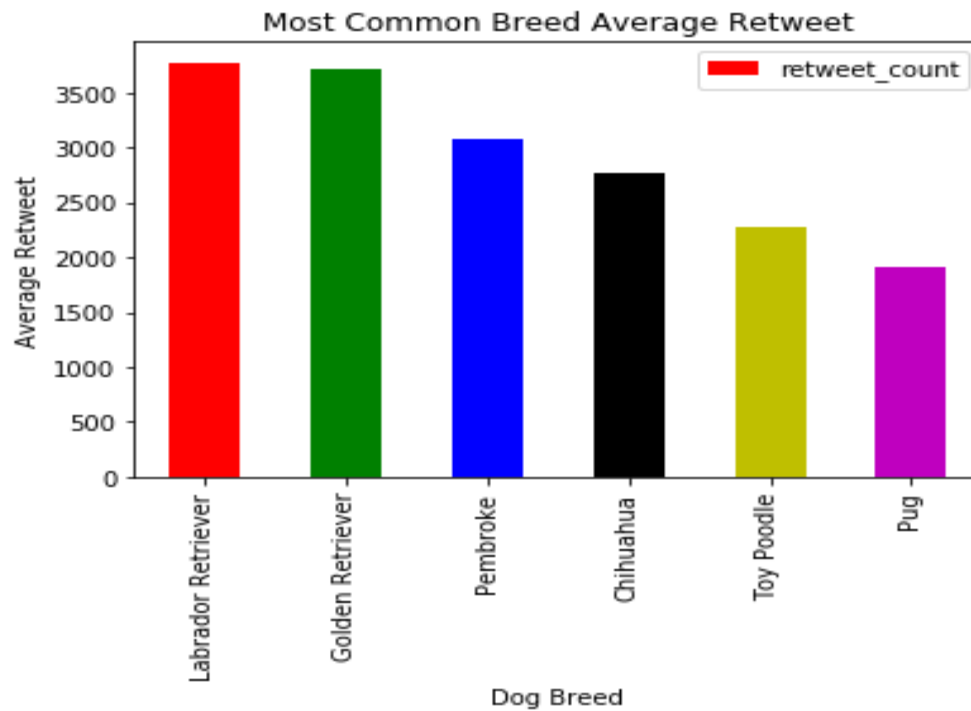


Looking at the most common breeds, we can see they receive very high ratings among the other dogs:

breed	rating_numerator
Golden Retriever	9.898734
Labrador Retriever	9.870370
Toy Poodle	9.823529
Pembroke	9.768421
Pug	9.500000
Chihuahua	9.428571

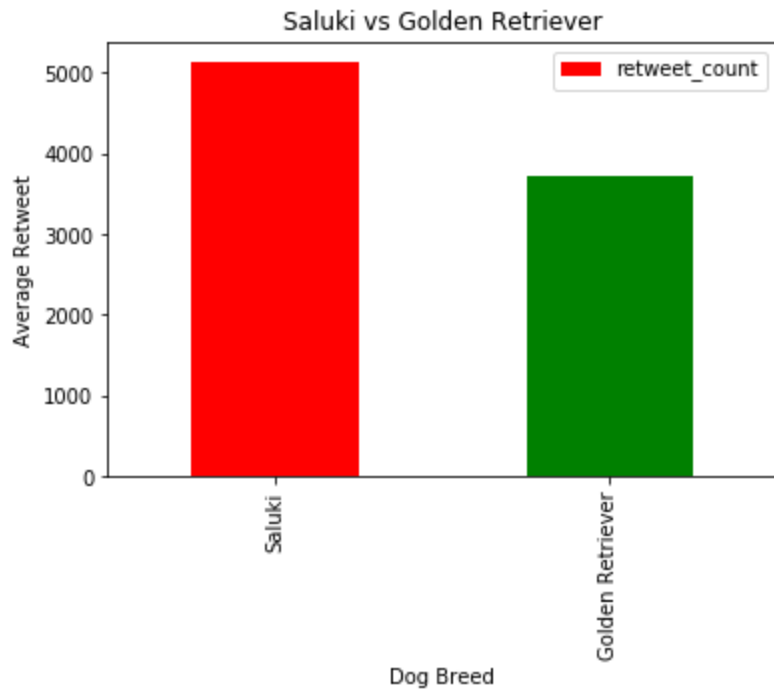
The Golden Retriever, among the most common, also has the highest rating.

Average Retweet for the Most Common Breeds



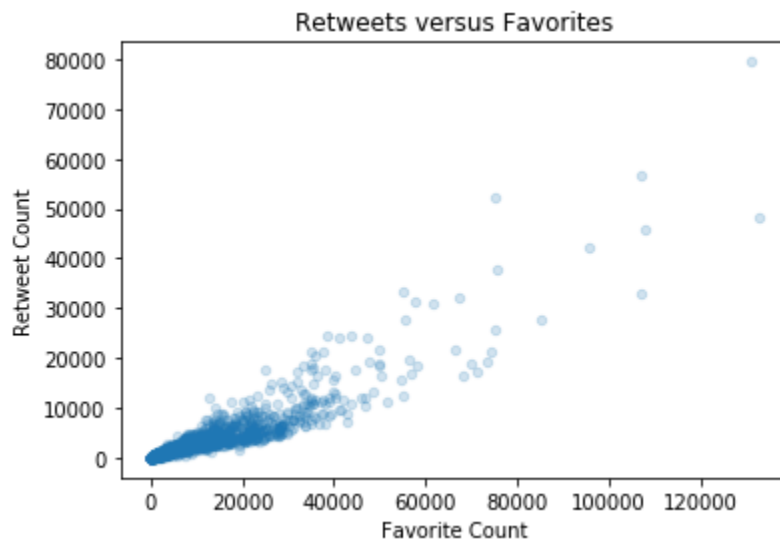
Looking at all 6 at the most common dog breeds among the twitter users, the Labrador Retriever, on average, was retweeted the most. Now that we have seen the most common breeds among the twitter users, I wanted to see if the most favored dog is retweeted more then the common dogs.

Saluki versus Golden Retriever



This graph shows even though the Golden is highest rated among the most common, it is not retweeted and favorited as much as the Saluki.

Relationship between Retweeting and Favoriting



I used a scatter plot to see the best way to visualize the relationship between if people are retweeting more or favoriting more. People seem to favor the feature option rather than retweeting.

In conclusion, the most common dog among twitter users were both types of retrievers, Golden and Labrador. The most popular(favorite) dog was the Saluki, even though it was the most retweeted, like the Bedlington Terrier. Most of the dogs, that were classed as a stage, were puppies. For future

analysis, I would like to see if the time and day of favoriting and tweeting makes a difference on how people look at their dogs and others.