Udacity Data Wrangle Project – We Rate Dogs – Act Report

Wrangling Data

Before I could begin the analysis, I wrangled some of the data quality and tidiness issues to make it easier to analyze.

Insights and Visualizations

Favorite Counts Are Higher Than Retweet Counts

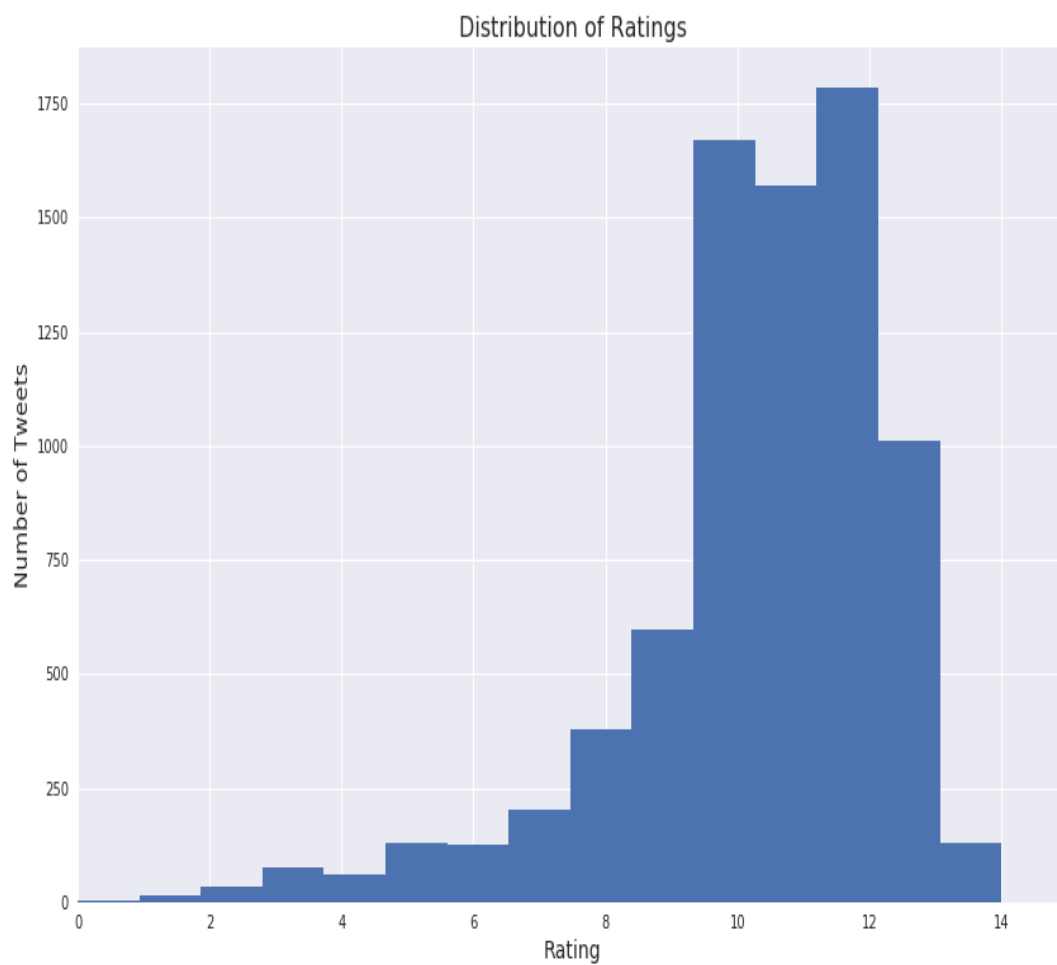ALL tweets have higher favorite counts than retweet counts.

Most common stages value counts

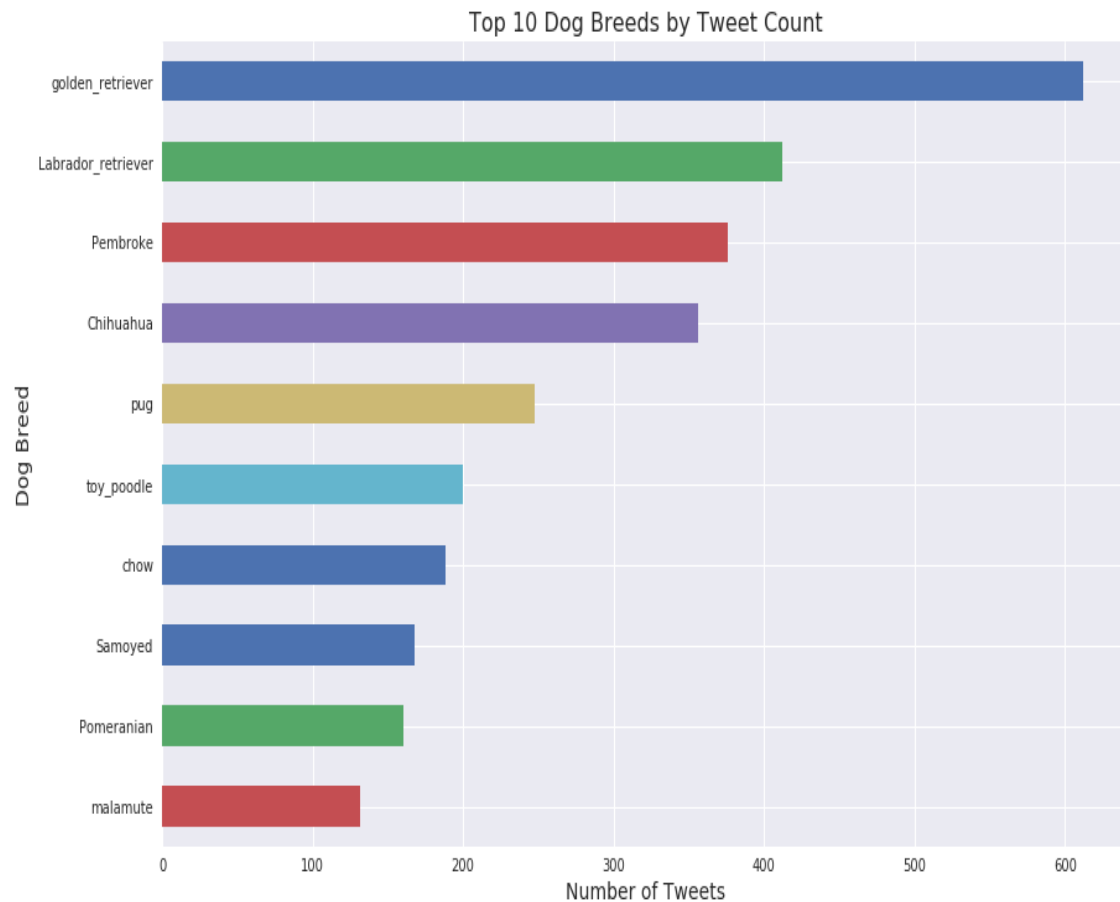|	
None	7484
pupper	208
doggo	73
puppo	23
floofer	8

Number of Distinct Breeds: 113

Many High Ratings

The distribution here shows just how many tweets there are with really high ratings; in fact, there are over 1500 tweets with ratings of 10 and up. That's 79% of all tweets!

Distribution of Ratings

Top 10 Dog Breeds by Tweet Count

The bar chart below shows that the most common dog that is tweeted about,

with almost over 600 tweets, is the Golden Retriever. Labrador Retrievers,

Pembrokes, and Chihuahuas are fairly close together in 2nd, 3rd, and 4th place.

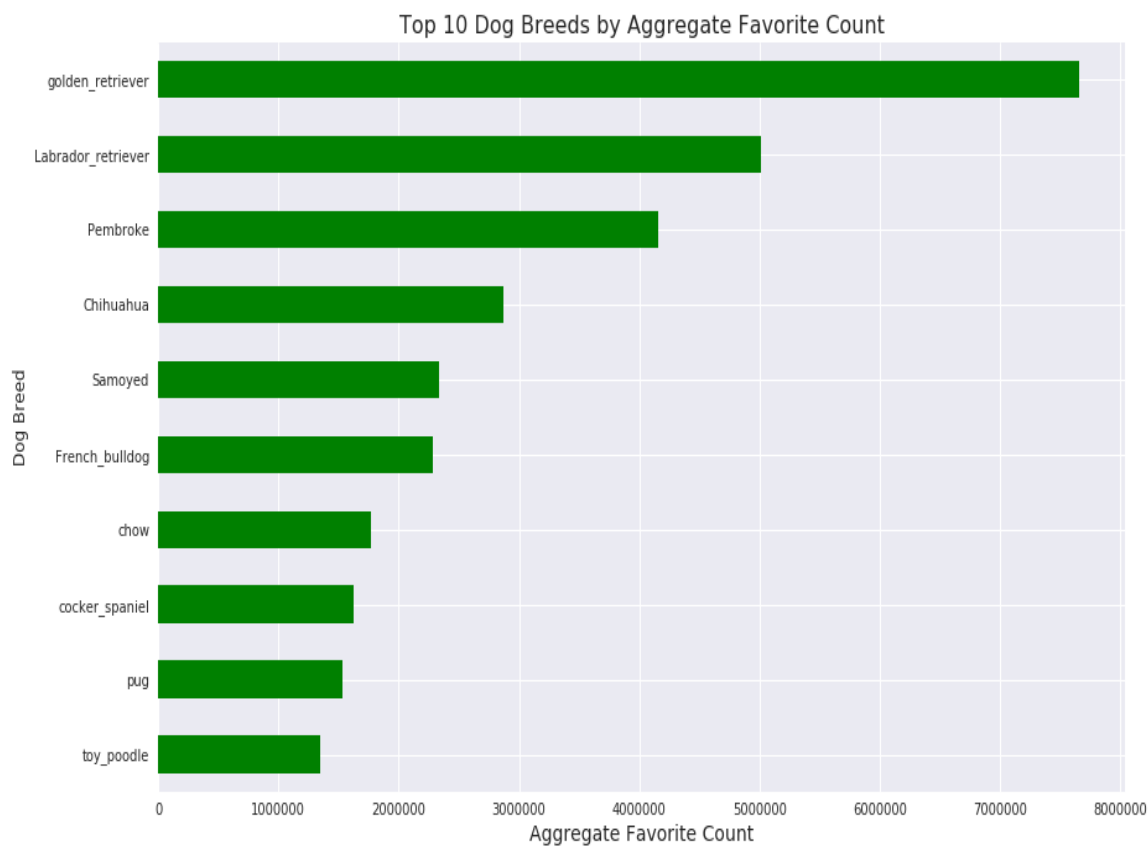Further down we'll take a look at what are the most favorited dog breeds.

Top 10 Dog Breeds By Aggregate Favorite Count

The plot below shows the top 10 dog breeds ranked by aggregate favorite counts.
It looks very similar to the barplot earlier of the top 10 dog breeds ranked by
tweet, and in fact, the top 4 dog breeds are the same. This also makes sense, as
these dogs have the highest number of tweets.

Both groups have these 8 dog breeds in common:

[

'Samoyed',   'Chihuahua',   'chow',   'Labrador_retriever',

'golden_retriever', 'Pembroke',   'pug',   'malamute'

]

Top 10 Dog Breeds by Aggregate Favorite Count

Top 10 Dog Breeds By Aggregate Favorite Count (Distribution Within Breeds)

The boxplot below shows the distribution of ratings within each of the top 10 dog breeds as ranked by aggregate favorite count. All the breeds range between 7 and 13, with only a few outliers, while four of the breeds have the lower ratings at 10. Another four breeds have the highest ratings at 14.

Ratings for Top 10 Dog Breeds (by Aggregate Favorite Count)