Udacity Data Analyst Nanodegree Program
Project 3. Wrangle and Analyze Twitter Data

author: Andrew Nachtigal

WeRateDogs Exploratory Data Analyasis and Visualization

Comparing WeRateDogs twitter information for the following data elements:

1. Dog Names
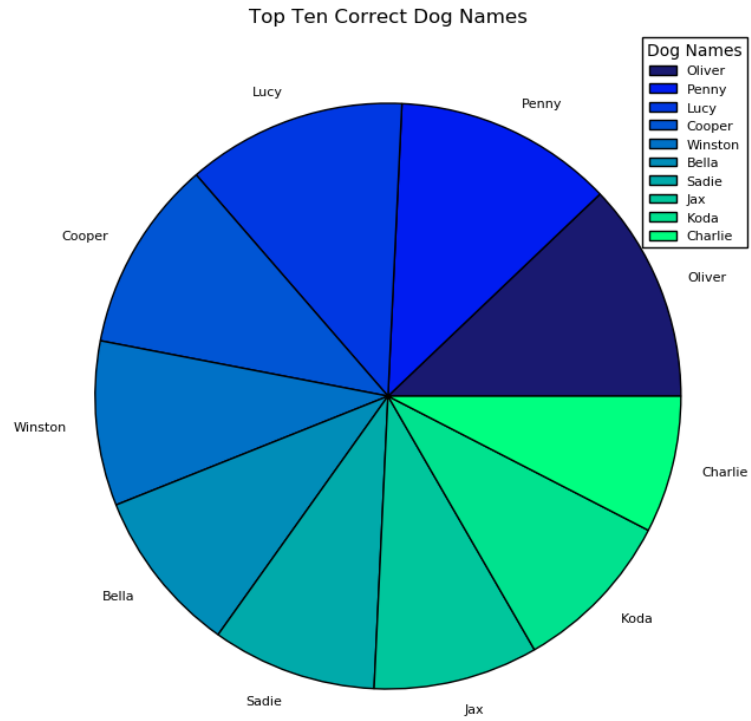2. Dog Ratings
3. Dog Stages
4. Dog Breeds
5. Tweet Timestamp

Original Dog Name Data

Grouping the most frequent dog names in a pie chart, we see incorrect names. The first two name are 'None' ands 'a', so there are clearly problems with the provided archived data. It may be useful to re-extract dog names from text.
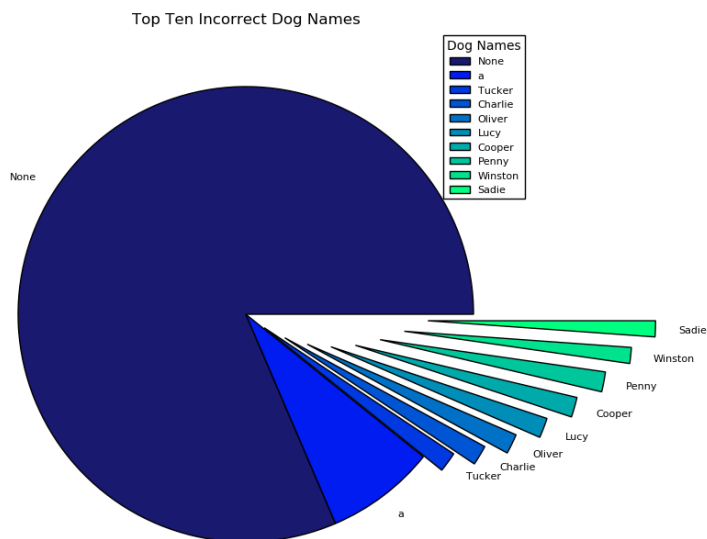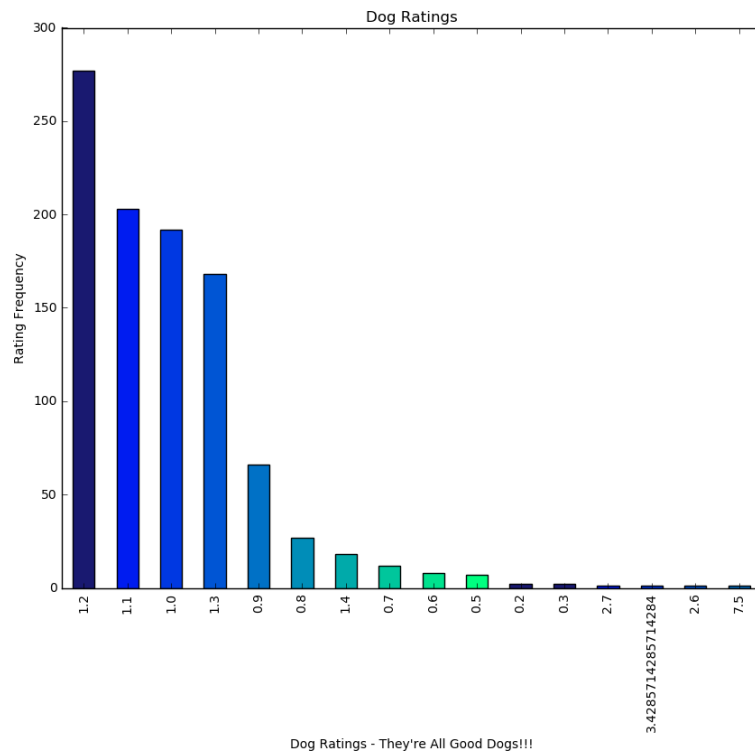
New Dog Name Data

After re-extracting dog names from the text, we see that the most frequent names are correct names.

Top Ten Correct Dog Names

Assigned Dog Ratings

This is a humorous twitter account and dog ratings are arbitrary so we don't try to fix them. The most frequent dog ratings can be seen in the following graph.

Dog Ratings - They're All Good Dogs!!!

Dog Stages

Dog stages are rather silly and the definitions don't have a lot of meaning. The frequency values for dog stages from the data are presented in the following plot.
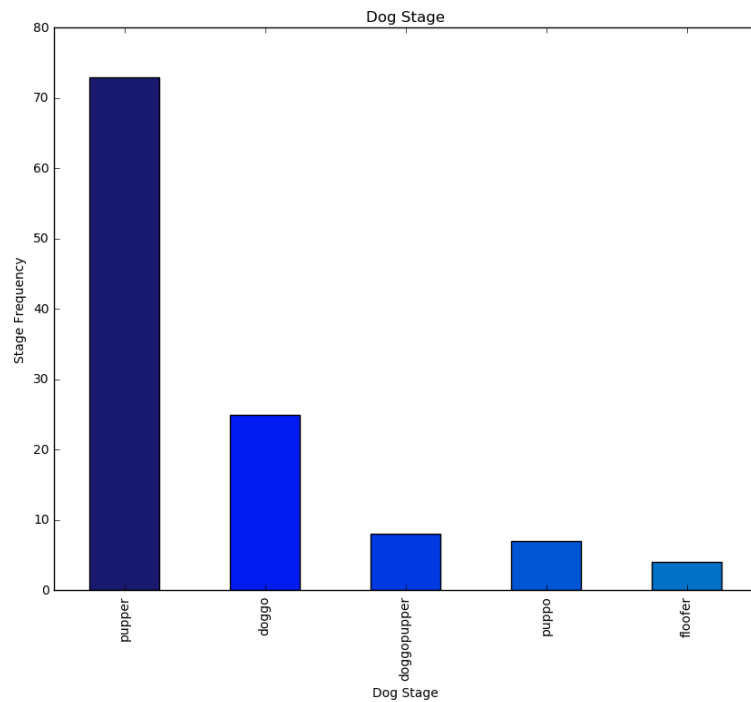

Dog Stages

We can compare dog stages with twitter likes (favorite counts). The doggo stage exhibits at least one significant outlier and the pupper stage appears to be the most variable for likes.
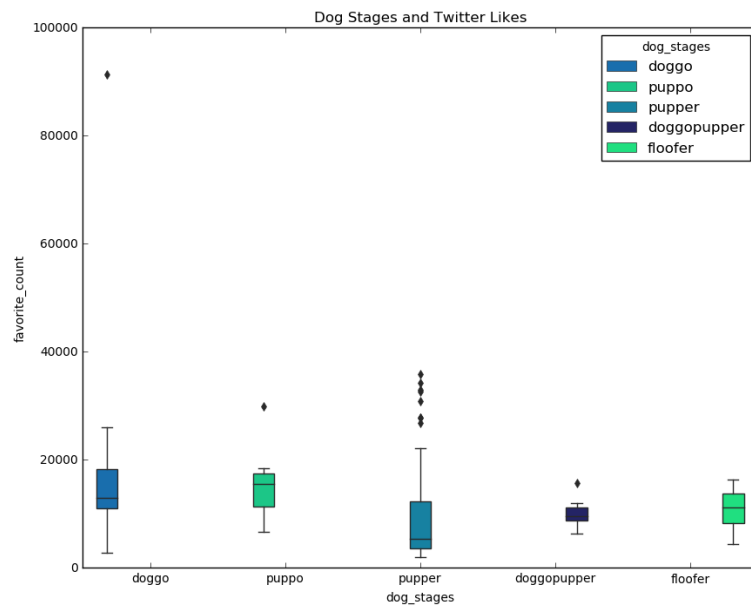
Twitter Like and Retweets

We can see that over time, the WeRateDogs Twitter account has increased in favorite count and retweet count. It seems that the popularity of the account is growing. Maybe people just like dog pictures.

Dog Like & Retweets
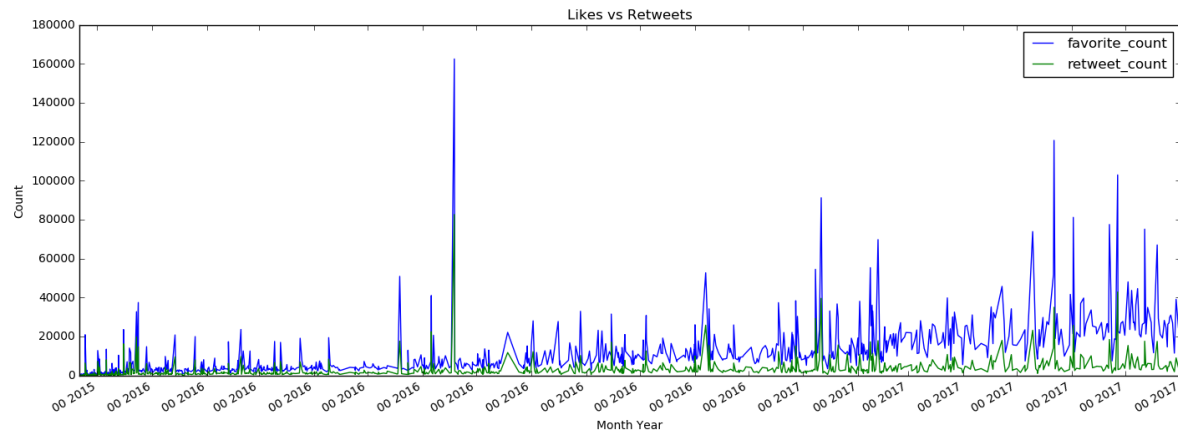
Dog Breeds

Grouping dogs by breed and plotting likes versus dog ratings, we see an upward trend. It would appear that a higher rated dog is correlated with more twitter likes.