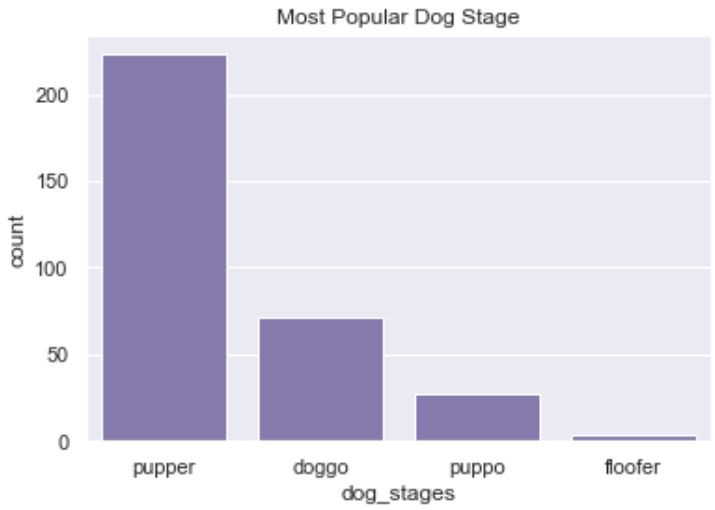
WeRateDogs Twitter Analysis

Dogs are often known to be man's best friend. Dog's are no stranger to social media. Owners will often place pictures of their beloved pet on various social media platforms. One platform, Tweeter, has a Twitter Account named WeRateDogs. This account consists of humorous comments about the about a person's dog. There is also a rating that is given to each dog or groups of dogs. The rating system is a bit peculiar in that it has a numerator/denominator rating system: 10/10, 11/10, 12/10, 13/10, 5/10. The numerators are mostly greater than 10, with some exceptions.

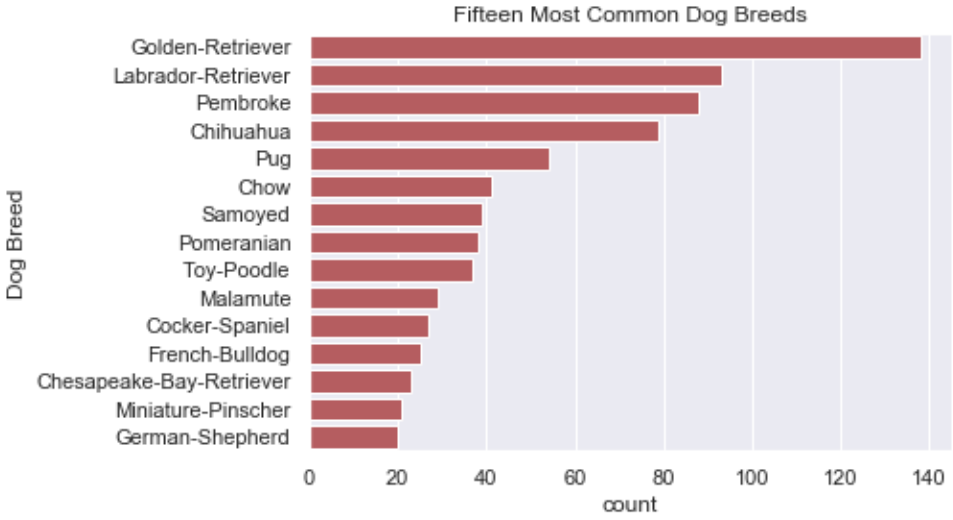
This project consisted of data wrangling, data analyzing and visualization using data from three different sources including a tweet archive from WeRateDogs. Some of the following insights consists of determining any correlations between the number of likes and the number of shares a particular tweet about a dog got, what were the most common dog breeds, most common dog names, and average ratings for certain dog breeds among other insights gathered.

There are four general dog stages that are depicted in the following chart:



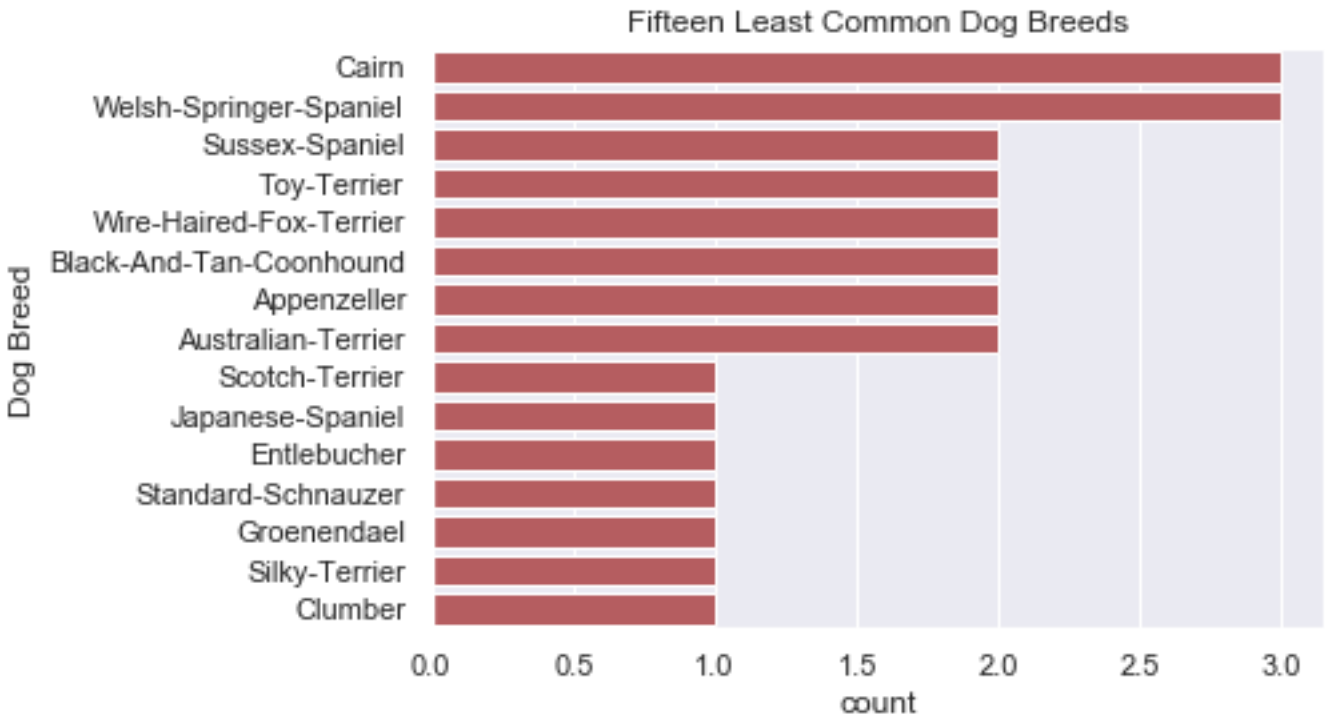
Based on dog stages that were documented in this Twitter archive, the figure above depicts pupper being the most frequent with a count of over 200. Based on the WeRateDogs book on Amazon, the Dogtionary refers to a pupper as a small doggo or younger and can be equally or more mature than some doggos. A doggo is usually an older dog that is like a big pupper. A puppo can be thought of the equivalent of a teenager, a transitional phase between pupper and doggo. The last category is floofer, which can be at any of the other three stages, but generally has an excessive amount of fur.

When it comes to choosing which breed of dog to adopt, the figure below me be an indication of which breeds of dogs are more popular.



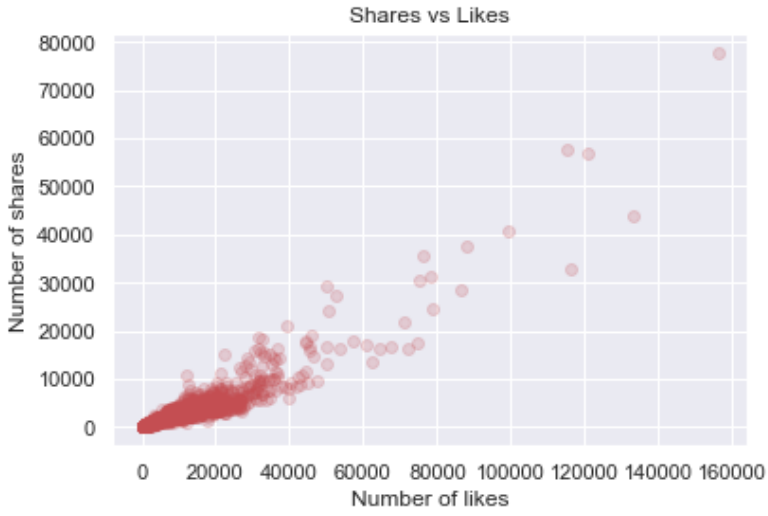
The Golden-Retriever along with the Labrador-Retriever breeds are top-dog based on this dataset. Smaller dog breeds such as the Chihuahua and Toy-Poodle also are also represented among 15 most common dog breeds. Some may find it surprising that German-Shepherds are not higher on the list and that Pit Bulls were not among the top 15 most common dog breeds.

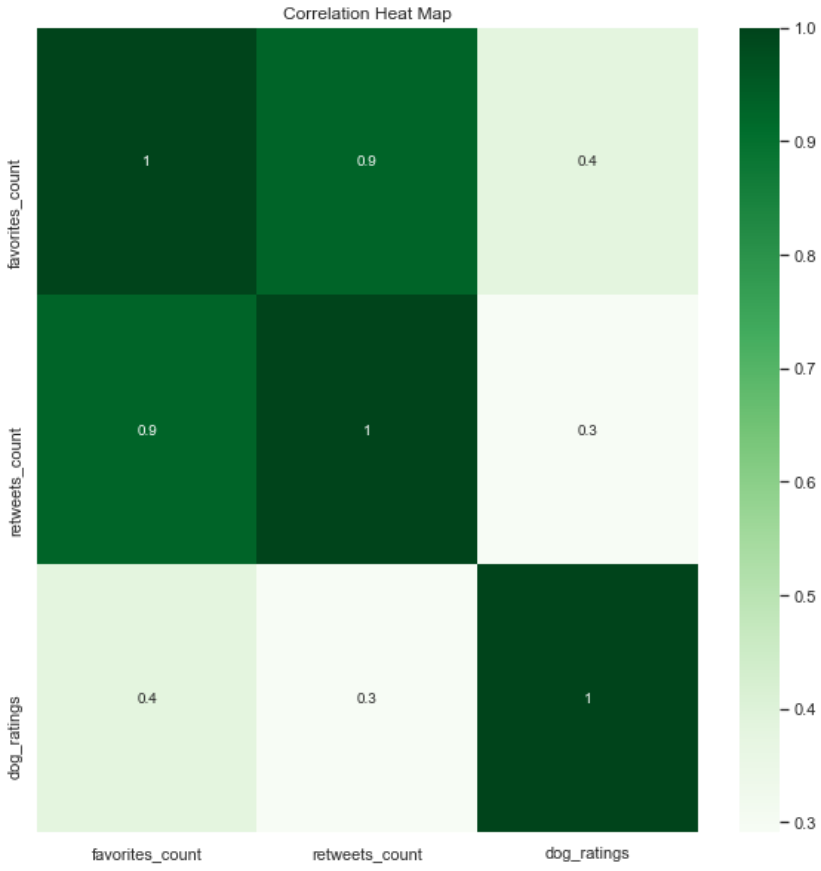
There are evidently several breeds of dogs to choose from. The following chart depicts some of the lesser common dog breeds in the dataset. It is possible that many of these breeds are unknown to even the most seasoned dog owners.



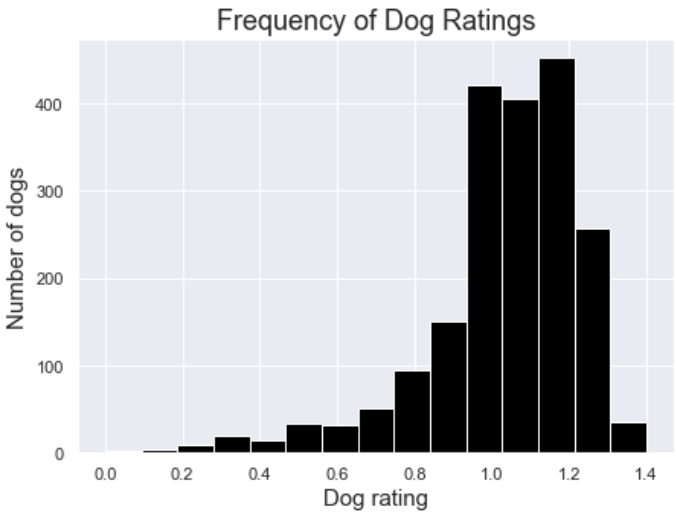
It is possible that dog breeds making it on to this list are not as widely known and thus may be more difficult to encounter.

The following visualizations give some insights into the number of shares and likes a dog gets.

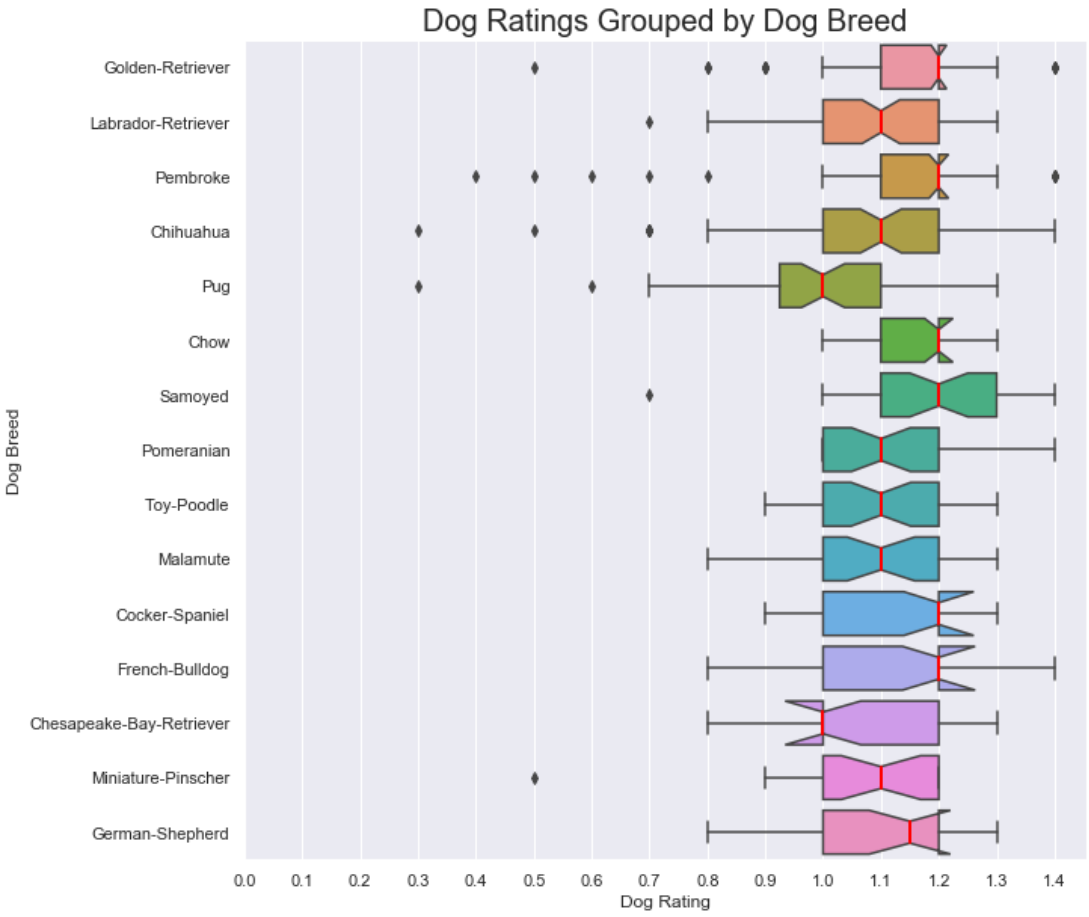




Both charts above show that the number of likes (favorites_counts) and the number of shares (retweets_count) have the strongest correlation at 0.9 as depicted more easily in the heat map above. The darker shadings of green indicate are stronger correlation. The same correlation is also depicted the scatterplot above. There is a moderate correlation between a dog's rating and the number of times it was liked. A weak correlation exists between the rating of a dog and the number of shares.



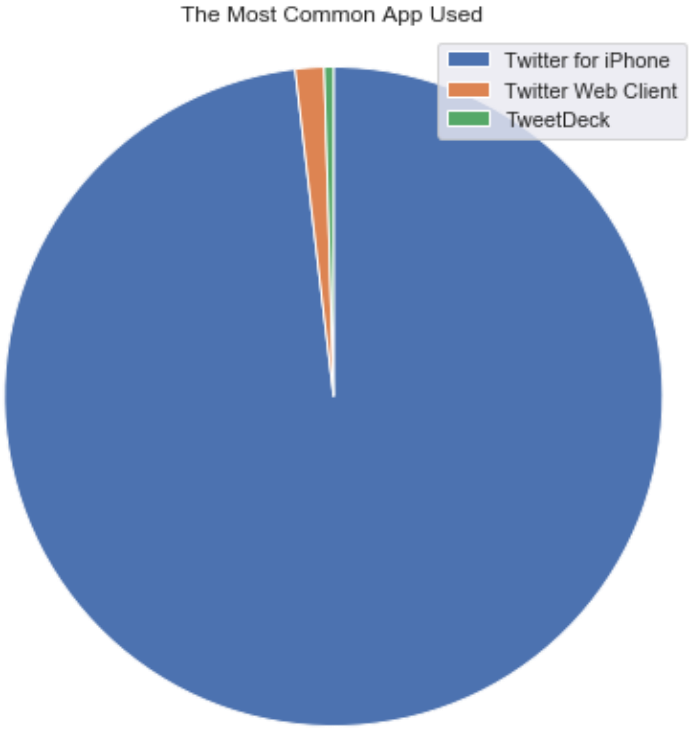
The distribution dog ratings are negatively skewed. Most of the ratings fall towards the right side. Fifty percent of all ratings are from 1.0 to 1.2 which falls within the interquartile range (IQR).



The chart above gives some insights on dog ratings based on the 15 most common breeds Two thirds of the breeds in the list above have ratings that have an IQR of 1.0 1.2. The red lines within the boxplot demonstrate that some of the dog breeds have rating distributions that are positively or negatively skewed. The rating distribution Golden-Retrievers is negatively skewed whereas the ratings for Chesapeake-Bay-Retrievers are positively skewed. The top five dog breeds so have some outliers towards lower ratings. For the most

part, the top 15 most common dog do appear to have favorable dog ratings with some outliers leaning towards less favorable ratings.

When it comes to tweeting, sharing, and liking which app was the most common?



The iPhone app is clearly the most common way people interacted on WeRateDogs

This project is part of Udacity's Data Analyst Nanodegree program. Below is a list of some of the resources outside of Udacity that were used while coding to come up with the analysis and visualizations presented here.

Resources

- https://stackoverflow.com/questions/51898826/converting-object-column-in-pandas-dataframe-to-datetime
- https://stackoverflow.com/questions/29902714/print-very-long-string-completely-in-pandas-dataframe
- https://kanoki.org/2019/11/12/how-to-use-regex-in-pandas/
- https://stackoverflow.com/questions/24775648/element-wise-logical-or-in-pandas
- https://stackoverflow.com/questions/32072076/find-the-unique-values-in-a-column-and-then-sort-them
- https://thispointer.com/pandas-apply-a-function-to-single-or-selected-columns-or-rows-in-dataframe/
- https://pandas.pydata.org/pandas-docs/stable/user_guide/options.html
- https://stackoverflow.com/questions/43269548/pandas-how-to-remove-rows-from-a-dataframe-based-on-a-list
- https://stackoverflow.com/questions/49291740/delete-rows-if-there-are-null-values-in-a-specific-column-in-pandas-dataframe
- https://stackoverflow.com/questions/11854847/how-can-i-display-an-image-from-a-file-in-jupyter-notebook
- https://pandas.pydata.org/docs/getting_started/intro_tutorials/05_add_columns.html
- https://stackoverflow.com/questions/11346283/renaming-columns-in-pandas
- https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.str.isupper.html
- https://stackoverflow.com/questions/44703945/pandas-trouble-stripping-html-tags-from-dataframe-column
- https://stackoverflow.com/questions/49733918/how-to-keep-the-same-datatype-after-saving-data-to-csv-file
- https://heartbeat.fritz.ai/seaborn-heatmaps-13-ways-to-customize-correlation-matrix-visualizations-f1c49c816f07
- https://github.com/mGalarnyk/Python_Tutorials/blob/master/Statistics/boxplot/Box_plot_interpretation.ipynb