Data Analysis on WeRateDogs Tweets

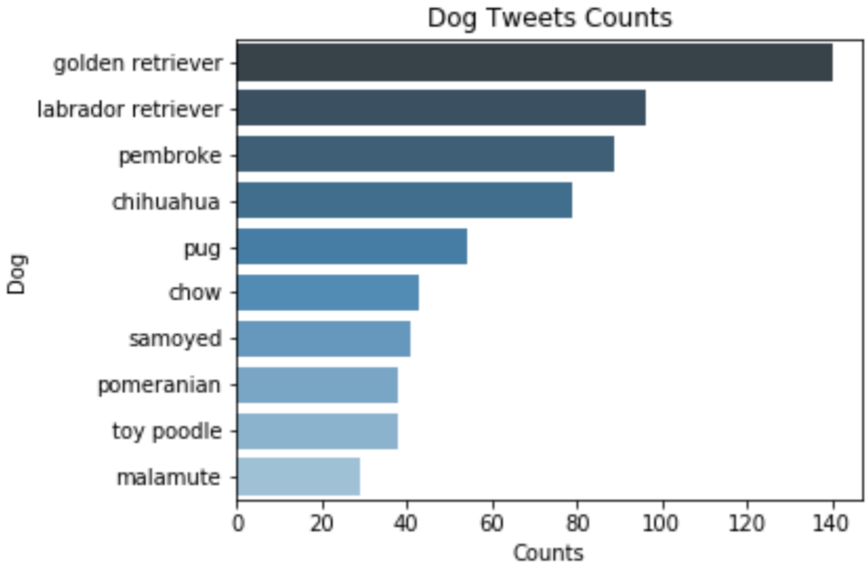
Po-Ching Yang

2018/7/15

This report summarizes the findings on WeRateDogs' tweet archive (up to 5000*+ tweets), with additional data of retweet counts and favorite counts gathered via Twitter API, as well as the image prediction results from Udacity. The four findings and the visualizations are summarized in the following sections.

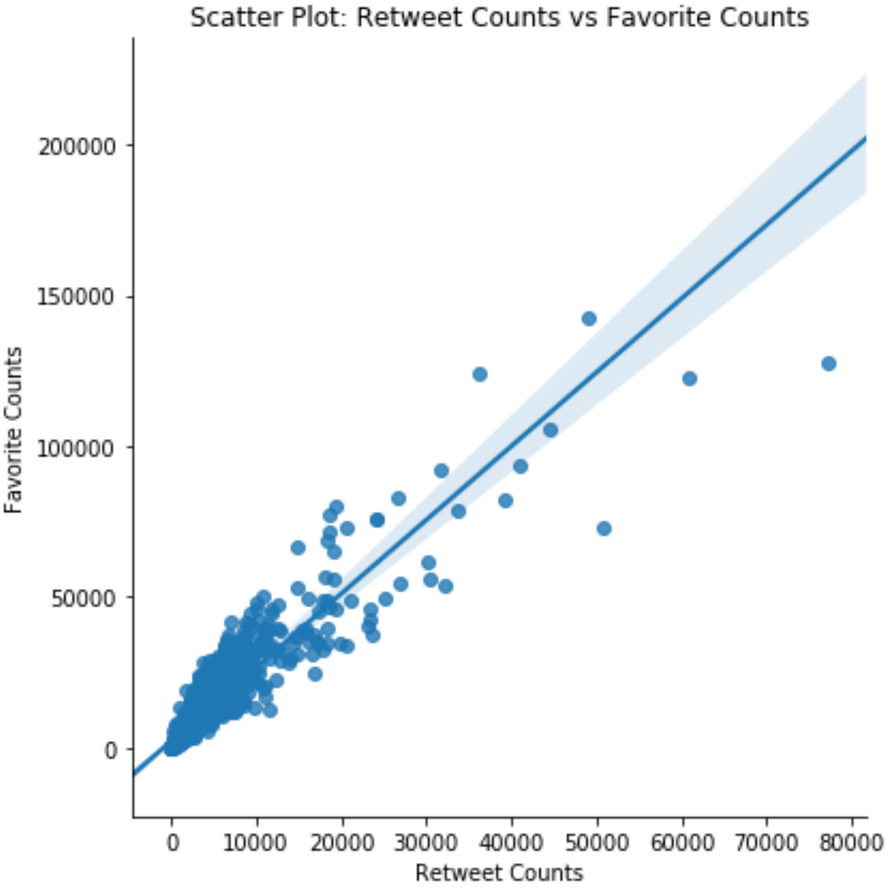
1. Insight 1: The dog breed with most tweets is "Golden Retriever"

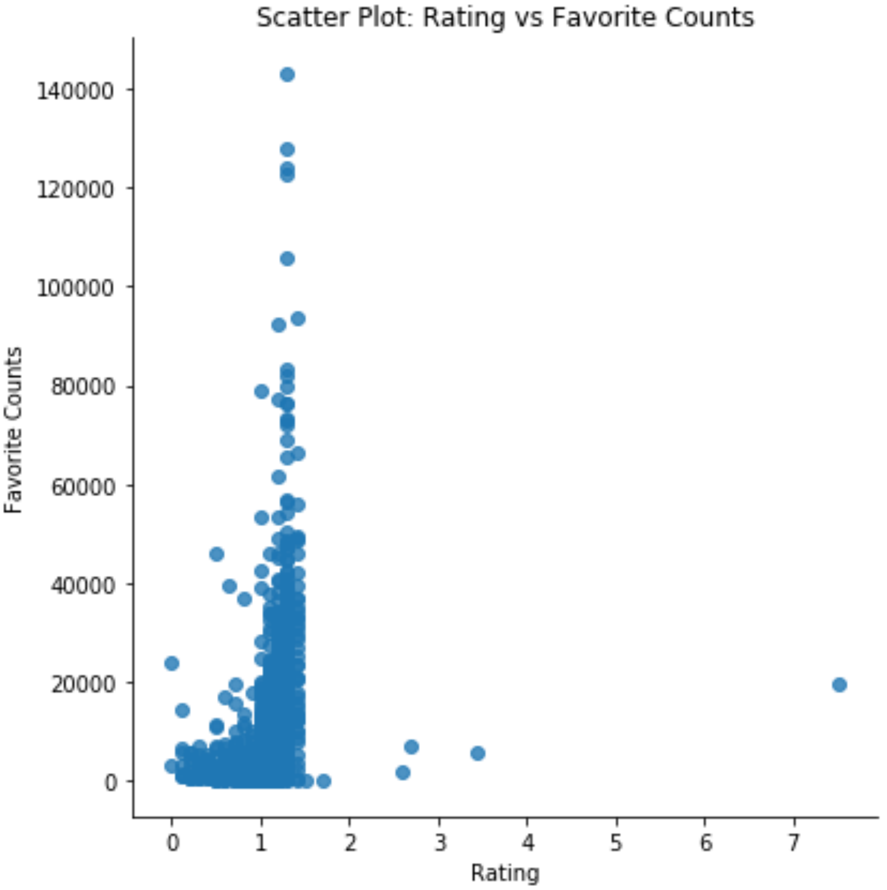
After combining the tweet archive of WeRateDogs and the image prediction results of each tweet, I was able to obtain the frequency counts of each breed in this dataset. I only used a subset of data where the prediction confidence level is above 50% and prediction results are thought to be dogs. The resulting plot below shows that the most frequent breed appeared in the tweets is golden retriever.



2. Insight 2: Rating Has Negligible Correlation with Retweets or Favorite Counts

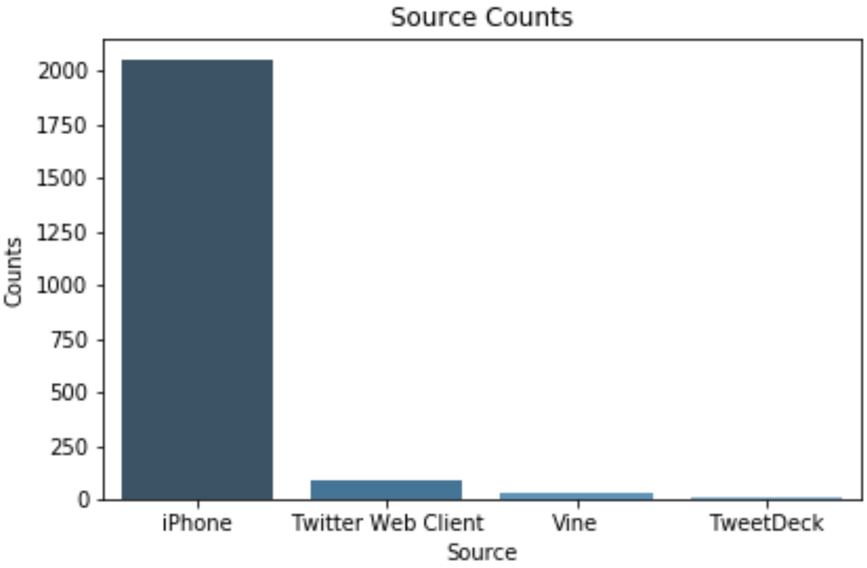
Next, I want to know the relationships among rating by WeRateDogs, retweet counts, and favorite counts. First, I calculated the correlation coefficients among the three variables. There is a very strong positive relationship between retweet counts and favorite counts, but these two variables have very weak or no linear relationships with rating. The visualizations for this finding are summarized as follow.


Scatter Plot: Retweet Counts vs Favorite Counts
Favorite Counts
200000
150000
100000
50000
0
0
10000
20000
30000
40000
50000
60000
70000
80000
Retweet Counts


Scatter Plot: Rating vs Favorite Counts
Favorite Counts
140000
120000
100000
80000
60000
40000
20000
0
0
1
2
3
4
5
6
7
Rating

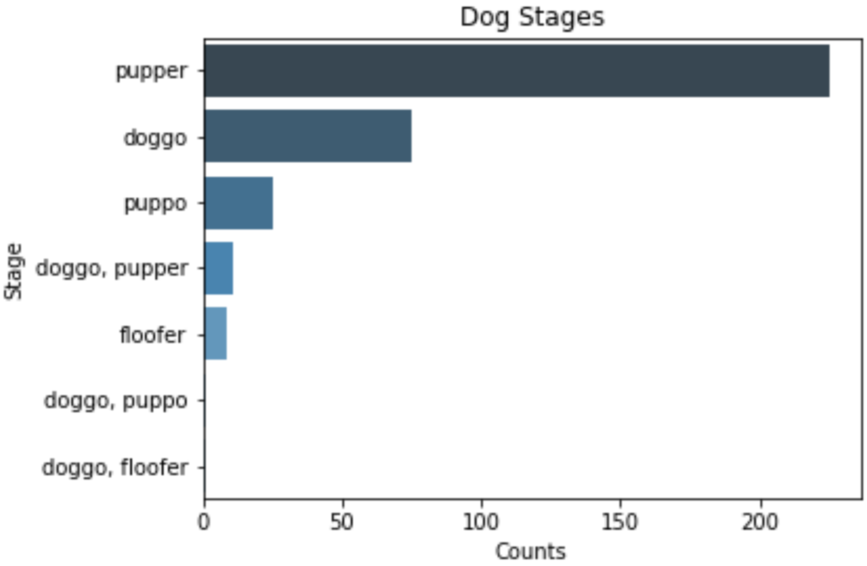
3. Insight 3: The Main Source of the Data Is iPhone

There are only four sources of data: iPhone, Twitter Web Client, Vine, and TweetDeck. Over 90% of data comes from iPhone, as the plot shown below.



4. Insight 4: The Most Common Dog Stage is "Pupper"

There are several dog stages described by the book "#WeRateDogs: The Most Hilarious and Adorable Pups You've Ever Seen", Matt Nelson. The most common stage found in this dataset is "pupper."


Dog Stages
Stage
pupper
doggo
puppo
doggo, pupper
floofer
doggo, puppo
doggo, floofer
Counts