WeRateDogs Twitter - Act Report

Vipin

july 2020

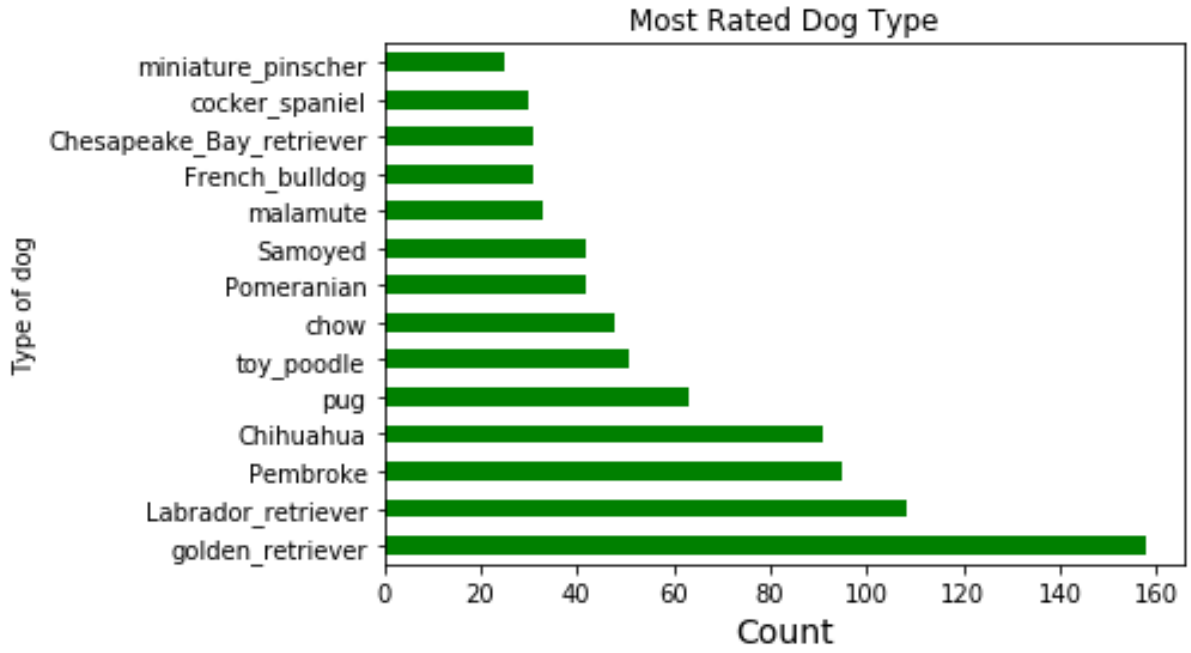
The WeRateDogs Enhanced Twitter archive contains data extracted from 2356 of the 5000+ tweets from the @dog_rates twitter account, posted between November 15, 2015 and August 1, 217. This data is comprised of dog ratings that were taken from the text of the tweet along with the dog name and dog stage if present. The retweet count and favourite count for each tweet were not included in the enhanced archive, and so I had to download this additional data from the twitter account using the tweet ID from the archive. Along with the Twitter data, I also downloaded an image predictions file from Udacity servers containing the top 3 predictions for dog breeds based on the images from the tweets (there can be up to 4).

Wrangling Data

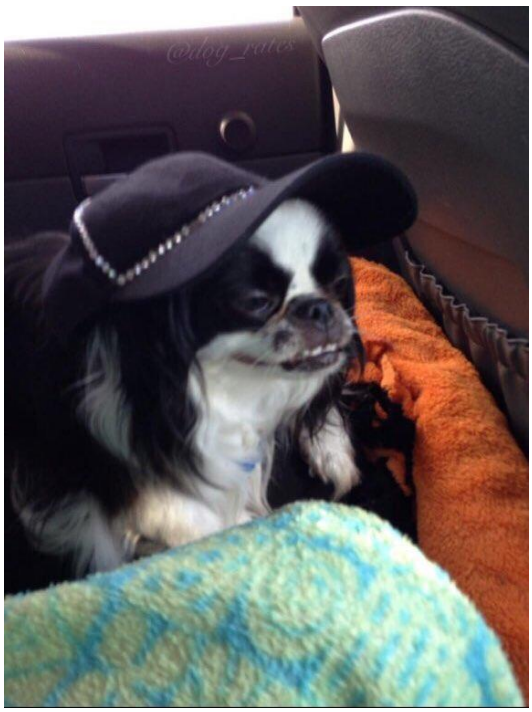
Before I could begin the analysis, the data had to be wrangled into shape to make it easier. I assessed the data both visually and programmatically for quality and tidiness; the quality of data is determined mainly by looking at several aspects or dimensions to ensure that it is complete, valid, accurate and consistent. After cleaning many of the issues found during the assessment, there were about 1950 tweets with good quality data.

Insights

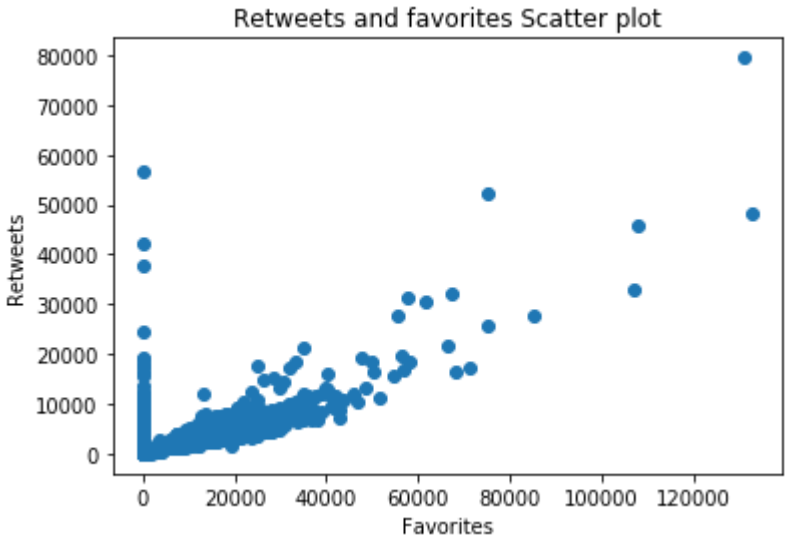
Golden retriever is the most common dog in this dataset. WeRateDogs has over 6000+ tweets. I was able to analyzed around 1500+ tweets. The most rated dog was golden retriever with more than 140 ratings.



Average rating among dog types. Almost all dogs deserve ratings above 10. This is a notable exception (for the curious ones). This Japanese spaniel has the lowest average rating. Bouvier_des_Flandres has the highest average rating



This plot is showing a few things here, but the overall picture is the strong positive correlation between retweet counts and favorite counts. It does make sense though, personally, if I 'like' a tweet (favourite it) I am more likely to retweet it as well.


Retweets and favorites Scatter plot
Retweets
Favorites
80000
70000
60000
50000
40000
30000
20000
10000
0
0
20000
40000
60000
80000
100000
120000