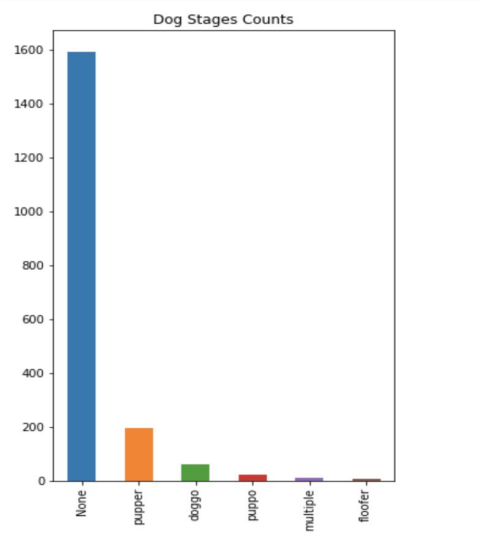
Data Visualization and Insights

By Ashwin Paramashivan

Introductions

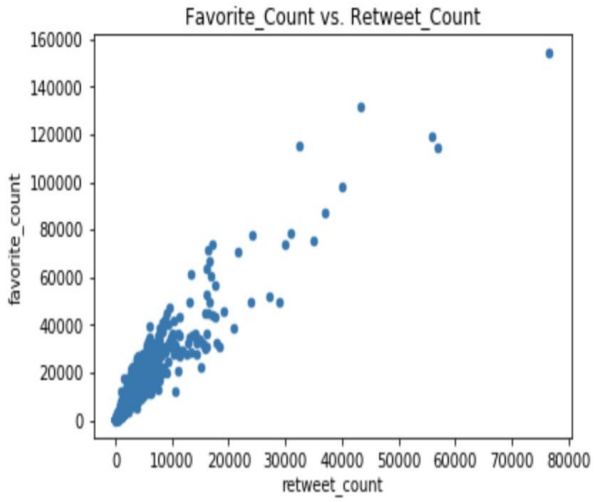
WeRateDogs is an online twitter blog which has ratings of all the dogs based on specific dog stages (pupper, doggo, puppo, floofer) and breeds. As you can see from one of the twitter post shown above a person rated the dog 13/10 or 1.3. The dogs which are liked the most get the higher rating. There are some dogs which have ratings greater than or less than 10/10 or 1. In my project, I gathered all the twitter data, wrote programs to clean the tables and came up with visualizations. Here I am going to show some statistics on the popularity of dog rates and accuracy of the predictions of dogs.

Popularity amongst dog based on stages



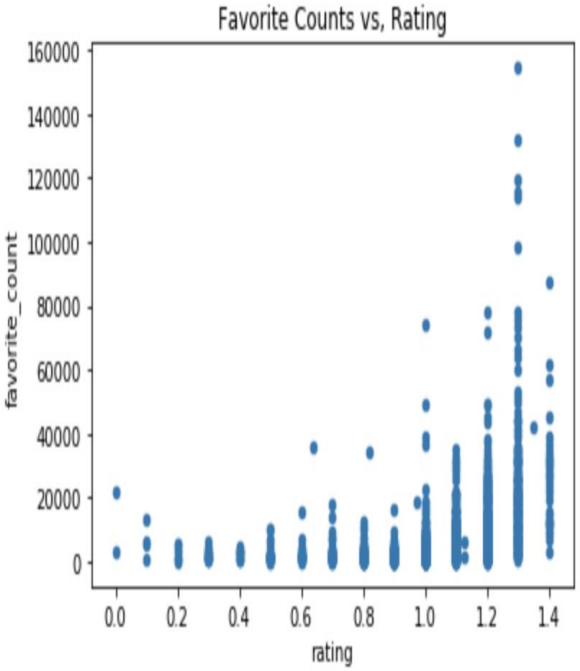
As seen in this bar chart, the dog with no stages are more popular than the dog that has 1 stage or more. The second most popular dog stage is pupper which is nearly 200. The least popular data is floofer which is less than 10.

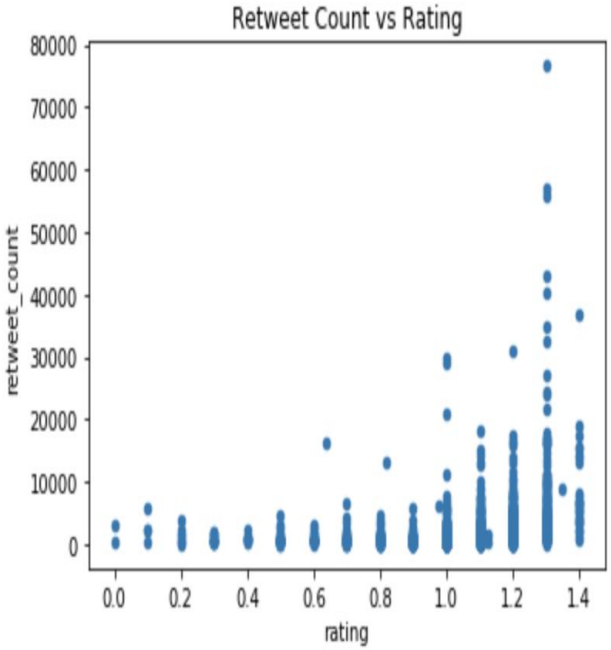
Relationship between favorite count and retweet count



There is a strong relationship between retweet count and favorite count. The dog which has 75000 retweets has about 150000 favorite counts from the user. For the dog which has 10 retweet counts has 10 favorite counts. So there is a positive trend in this graph. For more retweet counts then there are more favorite counts.

Relationship between rating, retweet count and favorite count

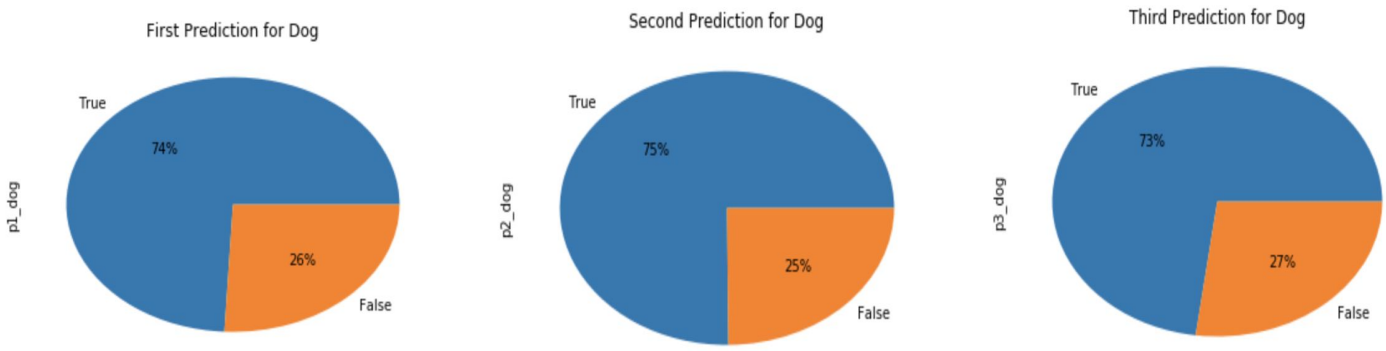




There is a weak positive relationship between favorite counts vs ratings graph. This goes the same for retweet counts vs. rating graph. There are some dogs with high

ratings which have low favorite counts and retweet counts. But there is not a dog with a low rating that has high favorite counts or retweet counts.

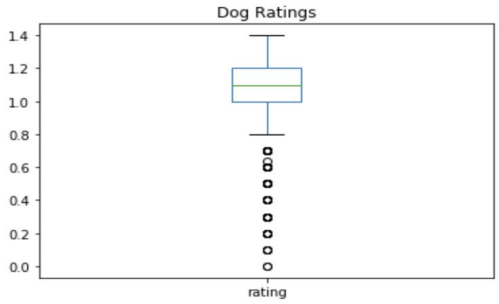
Dog breed prediction accuracy



Looking at these three pie charts above show the accuracy of prediction for a specific breed of dog. True represents whatever the user guessed is a breed of dog. For instance if a user A for prediction 1 says golden retriever and user B for prediction B says fox. The user A's prediction for dog breed is true but the user B's prediction is false. So there are three predictions listed above. On average for the three pie charts above, there is a 3/4 chance that the prediction is accurate and 1/4 chance that the prediction is inaccurate.

Brief statistics for each dog ratings

```
count    1886.000000
mean        1.058448
std         0.215989
min         0.000000
25%         1.000000
50%         1.100000
75%         1.200000
max         1.400000
```



Conclusion

The graphs above states a summary for the ratings of each dog. Their average rating is 1.05. The main limitation I faced is that I got a sample of 1886 datas. On twitter there are over 5000+ posts. I could have gathered the recent data but it would be difficult to get all those image predictions for a later date.