

Analyzing, and Visualizing WeRateDogs Dataset

1. Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. It was started in 2015 by college student Matt Nelson, and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter.

The Twitter account often giving dogs scores above 10. 11/10, 12/10, 13/10, etc. Why? Because "**they're good dogs Brent.**"

It has acquired over 6.5 million followers since its debut.

For doing this analysis I collected data from three different sources.

Twitter archive data which contains basic tweet data. This data was send by to Udacity.

The next set of data is image prediction dataset which has table full of image predictions (the top three only) alongside each tweet ID, image URL, and the image number that corresponded to the most confident prediction.

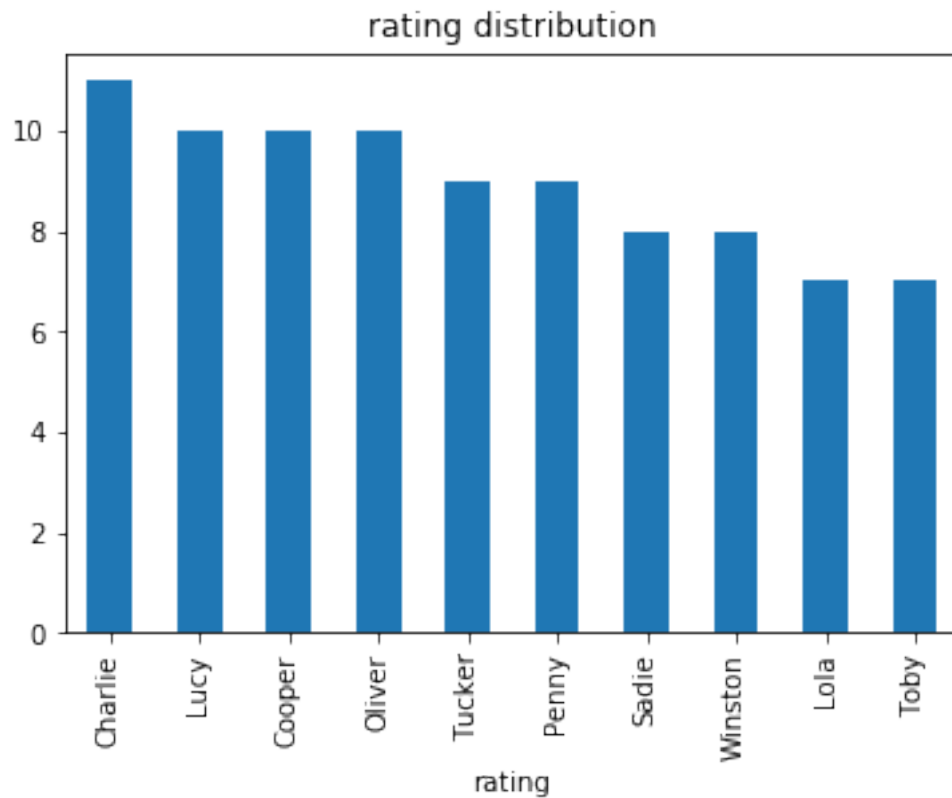
Finally, we also extracted Retweet count and favorite count data using twitter API to enrich our twitter archive data.

2. Analysis and Visualization

In order to gain insight using this data, I first thought of some basic questions . Let's start exploring these questions one by one

1. What are the top 5 dog's names?

The top 5 popular dog names are Charlie, Lucy, Cooper, Oliver and Tucker.



2. Why rating for one dog is 1776?

The dog name is Atticus and he is the highest rated dog in this dataset. This dog was rated on the 4th July which is Independence Day. So, the rating for this dog is 1776, the year when united states gained independence.



3. Why some dogs have very low rating?

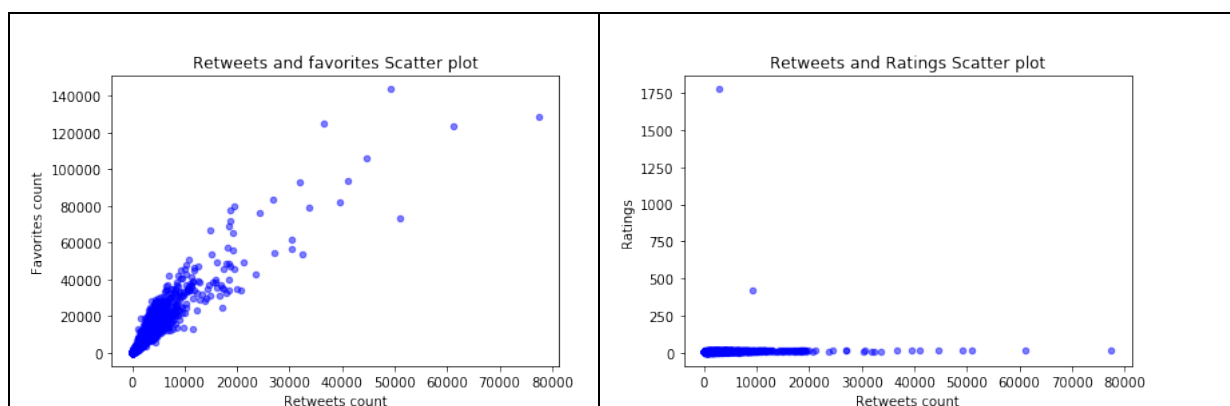
After examining some pics, it seems that low ratings are given to tweets which have no dog images in it. For example, the ones above have a picture of fan , chicken and a caterpillar. So, it looks like that pics with dogs images usually get high rating.



4. Is there any correlation exists between dog rating, favourite count and retweets count?

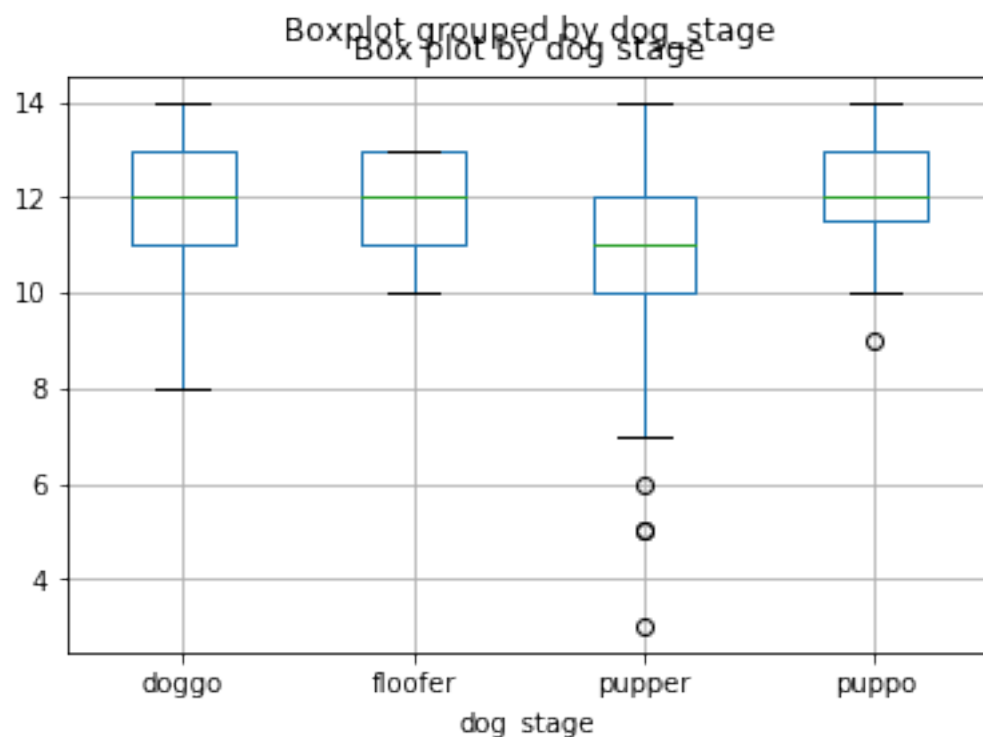
There seems to be a strong correlation between retweets and favorite counts. So, a tweet with more retweets would likely to get more likes in comparison to tweets with less number of retweets.

The correlation between dogs rating and the number of rewets or likes is very weak. So, number of retweets or likes are not associated with dog ratings, but with some other factors.



5. Does rating varies for different dog stages?

After analyzing the visual plot, it seems that several dogs belonging to pupper category received lower ratings. So, these outliers are pulling down puppers rating and therefore they have the minimum mean rating when we compare them to other dog stages. Floofers have minimum rating of 10 and they have consistently rated high. However, considering the number of data points we have for this dog type, we cannot draw out any conclusion that whether dog stage has any correlation with rating.



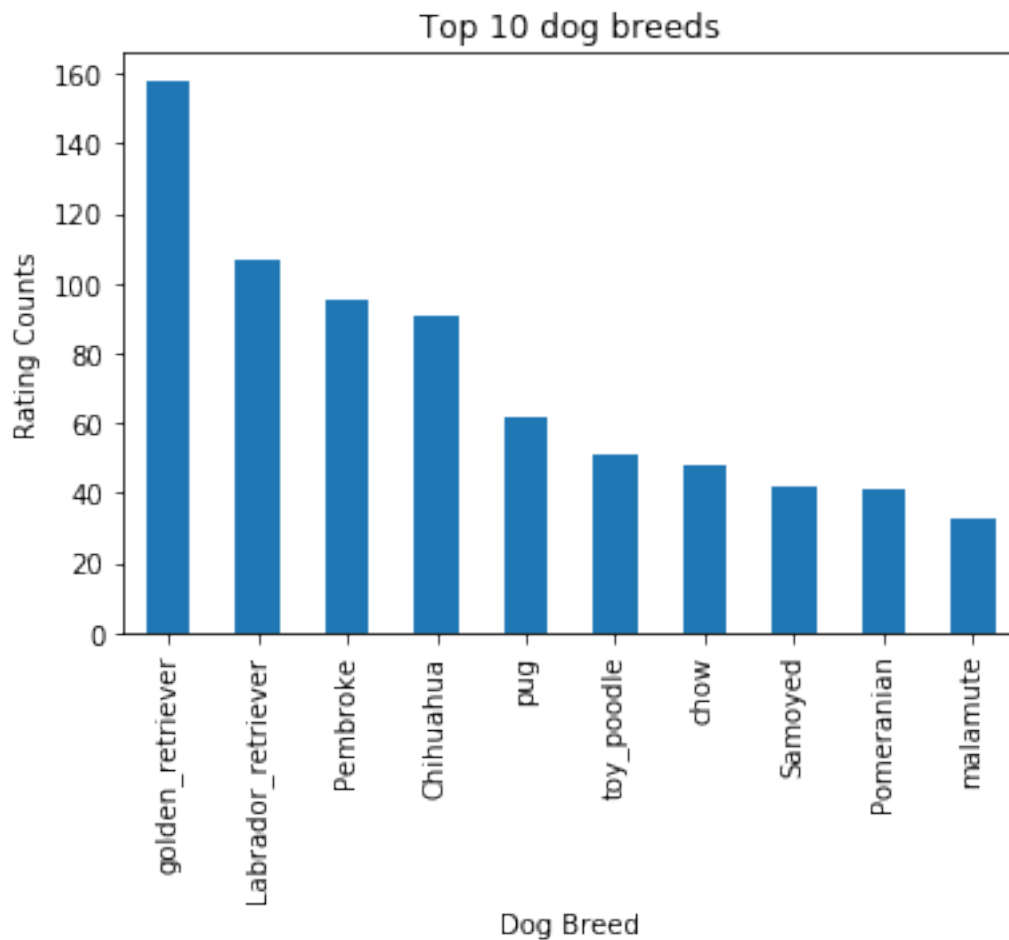
6. Which are the top 10 popular dog breeds?

From the visuals, it is evident that golden retriever and Labrador retriever breeds are the most rated one. It could be because of the reason that these ones are among the most popular dog breeds in United States. Below is the dog breed ranking from American Kennel club.

<https://www.akc.org/expert-advice/news/most-popular-dog-breeds-full-ranking-list/>

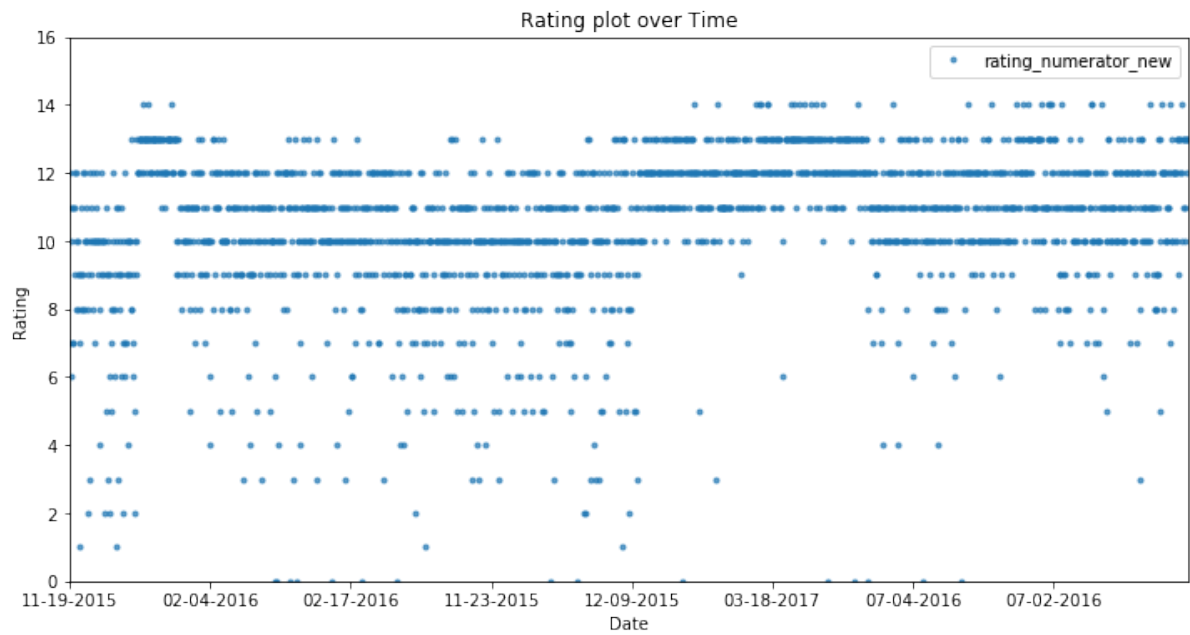
Also, it seems that the image prediction algorithm is not doing a very good job as it is not labeling all dogs with correct breed. For instance, 165 out of 306

tweets which were not classified as dogs are actually dogs. So, it is possible that the image prediction has misclassified many of the images as not dogs.



7. Is there any change in dog's ratings over a period of time?

It is quite evident from the plot that dog's ratings are generally high expect for few outliers in the beginning. If we look at the data most of the dogs are rated 8 and above and the lower rating is for those cases, when the image in the tweet is not of a dog.



3. Conclusion

In summary, the data analysis of the Twitter user WeRateDogs answers a lot of questions on dog ratings, breed, correlation between rating and retweets/favorite count etc. The insights from the data shows that WeRateDogs gives usually high rating to the pictures of dogs. So, they are true to their motto i.e "They're Good Dogs, Brent".