act_report

October 12, 2018

1 "We Rate Dogs" Data Analysis Report

1.1 About this report

Twitter account 'WeRateDogs (@dog_rates)" is famous for its unique tweets which evaluate dog pictures which are posted from readers. In this analysis, I got the tweets data from Twitter API and the additional data about pictures of those tweets and assess them. Specifically, I focused on 2 topics below

- How the ratings are related with stages and types of dogs.
- Relationship between favorite counts and the ratings.

1.2 Data analysis result

1.2.1 Topic1. How the ratings are related with stages and types of dogs.

It can be possible that a master of the account has personal preference of dog stages or dog types and the rating is somehow biased. I took a look at the realtionship between dog ratings, stages, and types.

Firstly, I visualized the relationship between dog stages and ratings. The dog stage "blep" has only 1 tweet, so we ignore this stage.

These 2 graphs below represent how the rating scores varies depending on dog styles and dog types.

Distribution of ratings of each stage of dogs

Distribution of ratings of each type of dogs
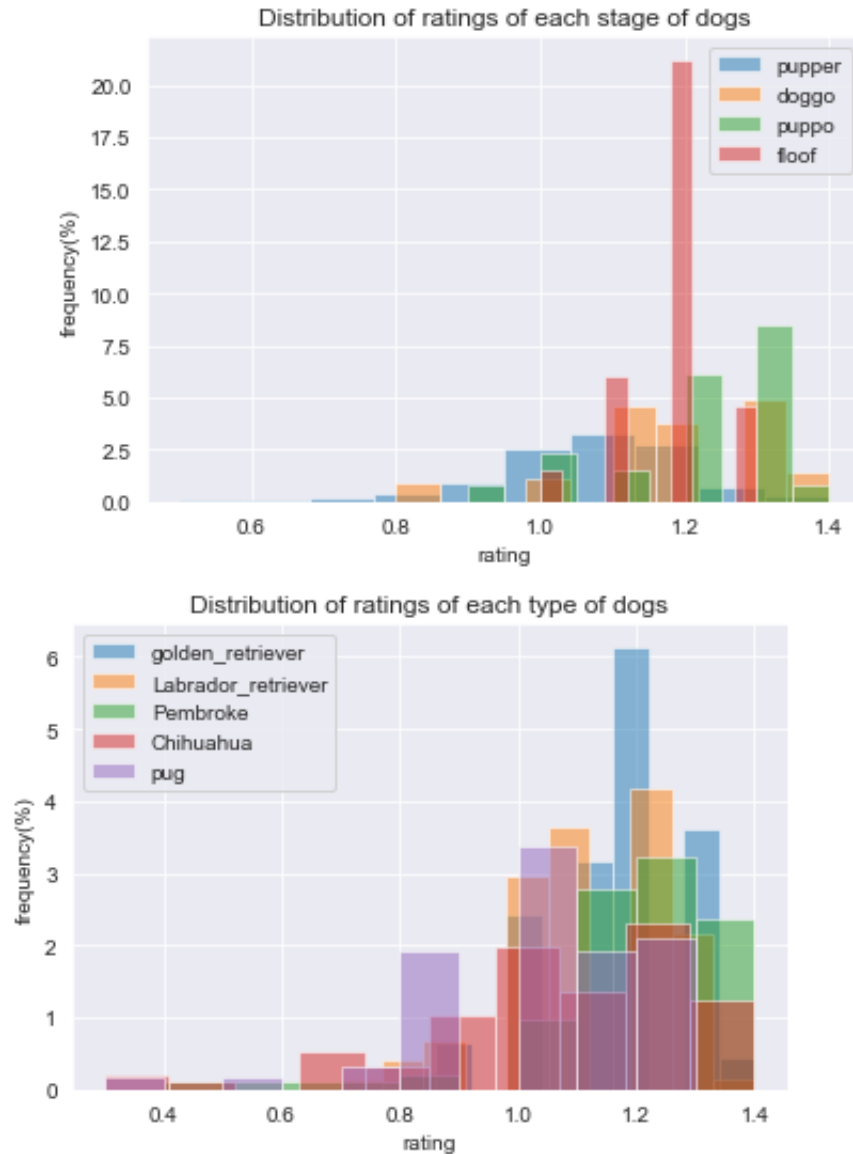
I applied multiple linear regression model to understand how these 2 factors are related with rating scores dog pictures get. I set "doggo" stage and "pupp" dog_type as baselines.

The result suggested that (in terms of dog stages) "doggo" has statistically significant difference from "pupper" and that (in terms of dog types) each of "Chihuahua","Labrador_retriever","Pembroke","golden_retriever" has statistically significant difference from "pug".

Also, it can be said that 27.4% of rating evaluation can be "explained" by these 2 factors.

1.2.2 Topic2. Relationship between favorite counts and the ratings

In this analysis, I have 2 indicators that show reader's preference on tweets. They are "favorite" counts and "retweet" counts. It is easily predictable that tweets with high
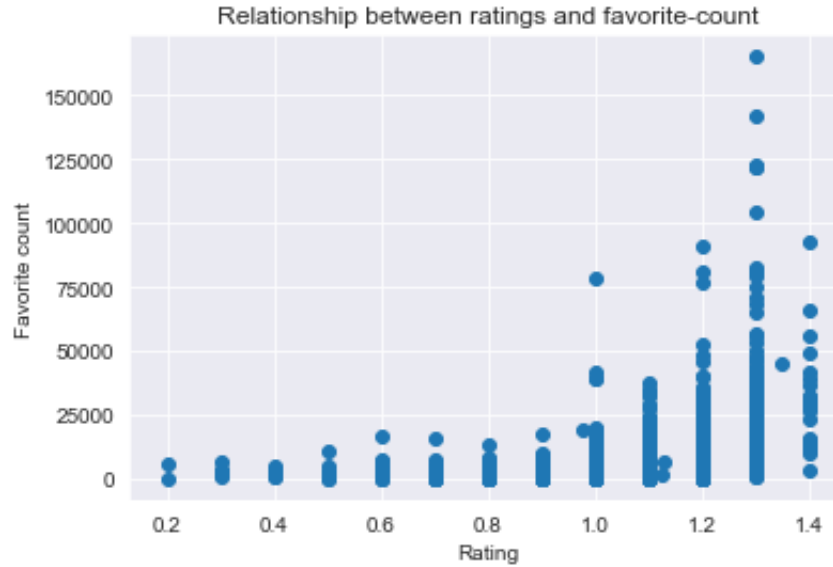
title

ratings are favored and also retweeted more and also that there is a strong positive correlationship between "favorite" counts and "retweet" counts. Actually it was true. I checked the relationship between these 2 counts and found that there is a positive correlationship.

So, in this project, I focused on "favite count" as one of the indicators of tweets' popularity among readers. The graph below represents the realationship between ratings written in the tweet and the tweet's favorite count. There can be found a positive correlationship between rating result and favorite accounts. This relationship is not completely explained by a linear model, as it clearly shows.

We have to note that it is not sure whether more readers favored because they found that the tweet had higher rating, or because pictures themselves are wonderful worth great rating.

Summary

In this analysis, I seeked the key to get high rating from 'WeRateDogs' from the perspective of readers, and also the key to get more favorites and retweets from the perspective of a master of the @dog_rates. It was found that some dog types and stages tend to get higher evaluation than others, and more highly evaluated tweets can be favored by more people.