

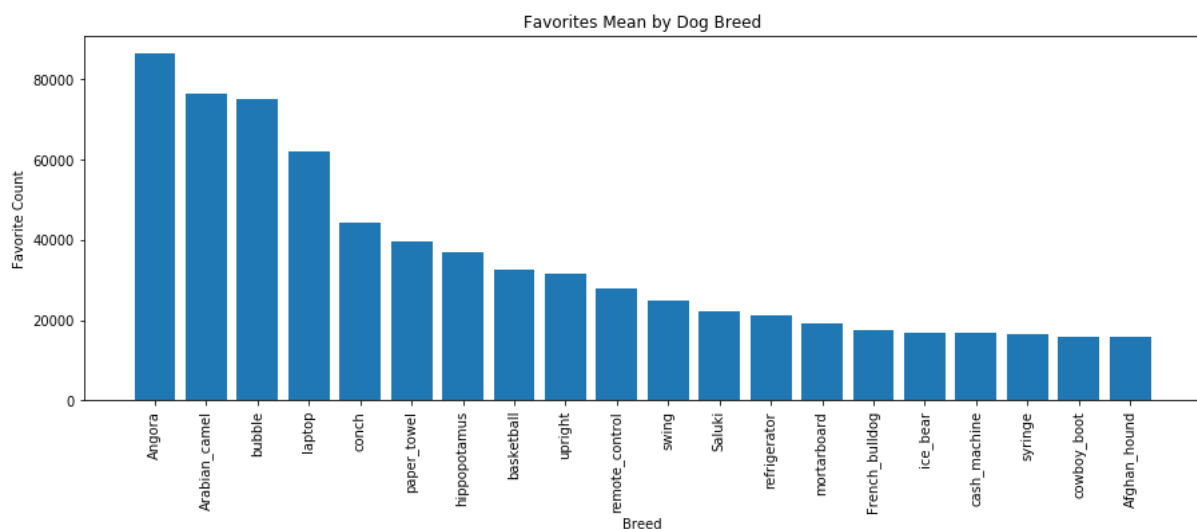
Exploring WeRateDog Tweets Dataset

We have a Tweets dataset from WeRateDog twitter account and a dataset of image predictions from the dogs that appears in those tweets. With this, we wanted to answer some questions by analyzing the mentioned data frames:

- **What breeds has higher favorite counts?**
- **What breeds are the most common?**
- **Has dog rating influence on number of favorites?**
- **Is there any influence on number of favorites to have better image predictions results?**

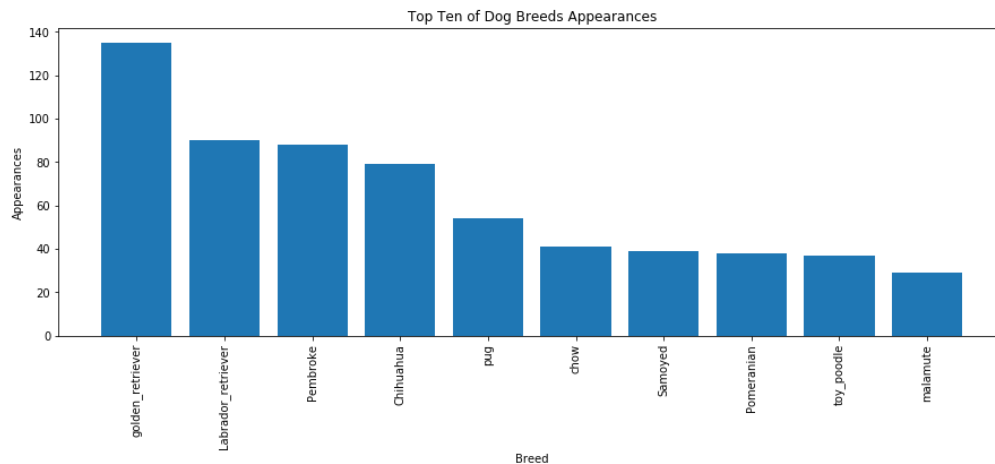
What breeds has higher favorite counts?

We got the mean favorites grouped by dog breed predicted by the machine learning algorithm and these are the results:



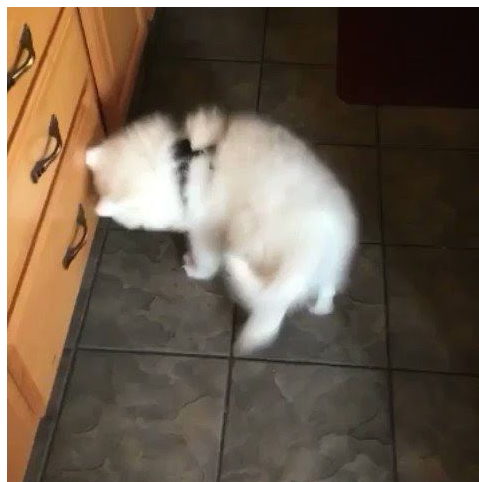
As you may see, some dog breeds seems to be preferred by the readers as Angora, Arabian Camel or Bubble. However, these do not seem to be dog breeds and do not sounds very common either.

What breeds are the most common?



In this case, we find breeds such as: Golden, Labrador or Pembroke. However, they do not appear in the top 20 of most favored dogs of WeRateDogs, why is this?

If we have a deeper look into the tweets we find out that Angora (most liked tweet) comes with this photo:



And what is more interesting, with this text: *'We only rate dogs. This is quite clearly a smol broken polar bear. We'd appreciate if you only send dogs. Thank you... 12/10'*.

It is in fact a dog but it seems that the text has a great influence of the number of favorites, despite of being a photo of low quality.

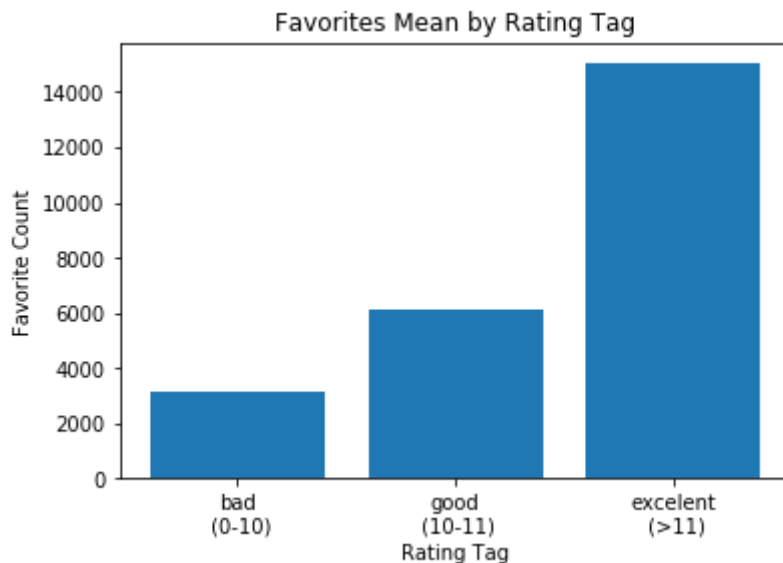
Has dog ratings influence on number of favorites?

We divided tweet ratings into three different bins:

- **bad:** Those with ratings below 10.
- **good:** 10 and 11

- **excellent:** those above 11.

Grouping the dataset by this rating tags and getting the mean value of each group for the number of favorites we get the results shown in the next chart:



We observe that it seems to be a positive correlation between good ratings and higher number of favorite counts.

Conclusion

As we've seen in the results, we may expect to get better broadcast from tweets that shows certain breeds. However, after looking at the most common identified breeds we noticed that they don't coincide at all.

The most favored dog breed is Angora but the picture itself it is low quality and has a good rating though. However, the tweet has an ironic message: *'We only rate dogs. This is quite clearly a smol broken polar bear. We'd appreciate if you only send dogs. Thank you... 12/10'*.

We also noticed that ratings seems to have an influence on the number of favorites as it appears to represent an indicator of the whole tweet (text + image).

To be able to build a model of prediction from this tweets database a further analysis should be done on the tweet's text content.