WeRateDogs

Report on Tweet
Archive **@dog_rates**

“This is Oliver. He was given a very little ice cream cone today, which is okay because he is also very little. 12/10”

What is ‘WeRateDogs’?

WeRateDogs is a Twitter account that **rates dog pictures**. The user tweets pictures from his followers and provides a rating and an ingenious, funny comment.

What data were explored?

The dataset included information regarding 2000+ tweets by this user. Each row displayed varied information, including timestamp, the name of the dog, and whether it was a *doggo, floofer, pupper*, or *puppo*, which are labels that refer to how old the dog seems.

Also, the dataset included information about the images used in each tweet. Specifically, it predicted what each dog’s breed was.

We worked on three questions regarding these data:

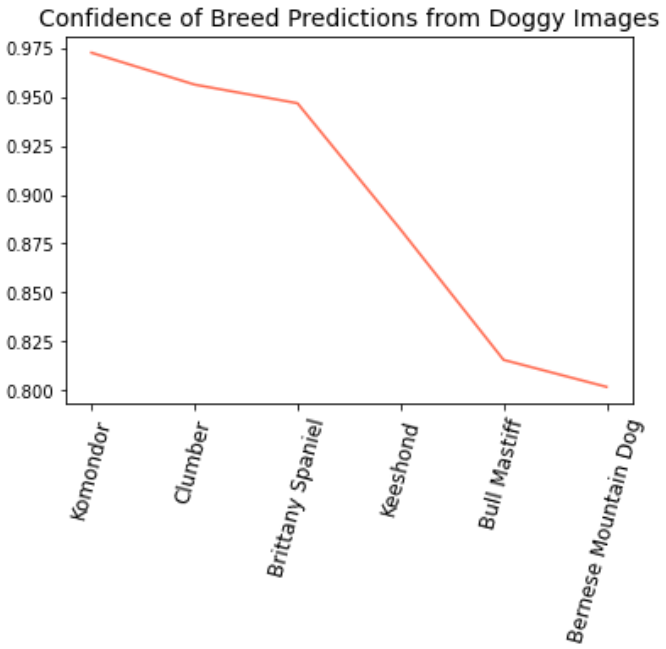
Do image ratings show any correlation with the time of the day in which these images are tweeted?

For this question, we classified each tweet according to their timestamp into *morning*, *afternoon*, *evening* and *night*. The results showed that WeRateDogs’ user tweets neither during the night (12 am to 6 am) nor in the morning (6 am to 12 pm), and that the evenings (6 pm to 12 am) are much more prolific than the afternoons (12 pm to 6 pm).

After the outliers were discarded, a t-test calculation showed that **the images tweeted in the afternoon were rated significantly higher than those tweeted in the evening**.

Which breeds have the highest confidence intervals for breed recognition?

After discarding all the images (rows) that the algorithm failed to recognize as dogs and isolating only the most likely prediction, the data showed that the breeds associated with predictions with the highest confidence included Komondor, Clumber, Brittany Spaniel, Keeshond, Bull Mastiff, and Bernese Mountain Dog.



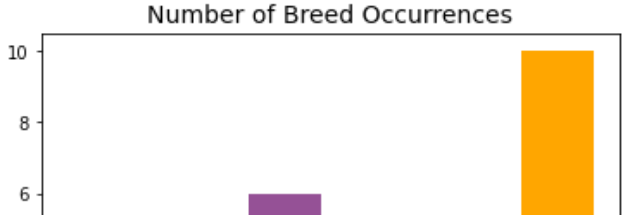
Incidentally, the breed associated with the lowest-confidence predictions was the Scottish Deerhound.

Are the breeds with highest confidence predictions also common among the images?

In other words, is the higher or lower confidence of the algorithm used to recognize breeds the result of having more or less images associated with each breed to train the algorithm? This question may seem silly to you, but since I have no idea where the algorithm came from, it seemed natural to a novice like me.

A brief summary analysis of these data showed that there were 109 breeds recognized, and that the range of occurrences of different breeds was 1 to 143. However, the mean of occurrences was 12.9, and the median was 7. This means that there were a few breeds that had a much higher count than the rest. In this case, these breeds were Golden Retriever (150), Labrador Retriever (100), Pembroke (89), and Chihuahua (83). None of which were in the group with highest confidence in prediction.

Indeed the group with the highest confidence in prediction had a relatively low number of occurrences, as shown in the bar plot below.



All this suggests, as it could have been predicted by a data analyst with more expertise regarding all things Twitter, that the recognition algorithm was not trained by this data set.

In sum, only three analyses cannot say much about the data. However, we do know now that ratings do seem to be affected by the time at which the images are tweeted, and also that higher recognition by the algorithm does not seem to be linked to the number of images sent to WeRateDogs for posting.

At a more existential and practical level, we would like you to take three comments home with you:

1. If you want your doggie to be rated as your heart believes it should, try sending the picture after lunch.
2. I need to express my deep admiration for the algorithm that recognized the breeds, because even after some Googling, I could not recognize them myself to save my life.
3. The fact that there are so many pictures of Retrievers - which are cute, do not get me wrong - and only three Komondors seems unfair and a missed opportunity. Therefore, I will leave you with a beautiful Komondor picture. You are welcome.