Act report

The wrangle project was done by gathering, assessing, cleaning and analyzing the data from WeRateDogs Twitter. In this section the analysis is presented.

Dog information

First of all, it was part of the analysis to get some understanding of the dogs in the tweets. Figure 1 shows that from the dogs that are classified, 68,3 % of the dogs are classified as pupper dogs follwed by 24,2 % of doggo dogs. It seems that, per definition of a pupper, small and younger dogs are widely spread in the tweets. Per definition it also needs to be said, that the dog classes cannot be delineated sharply, so this information only can give a tendency about the dogs stage. Also the majority of the dogs was not classified. Nevertheless there is data that supports the widespread distribution of pupper dogs in the tweets.
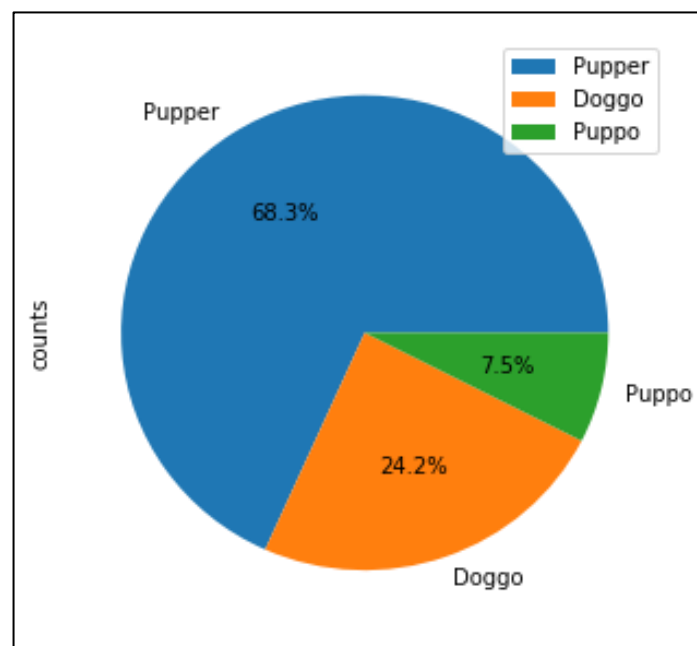

Figure 1. Percentage of the dog stages in the dataset

Data about the dogs breed shows that golden retrievers are the ones that appear the most in the tweet images before labradors, pembrokesand and chihuahas. With having used a prediction algorithm this data needs to be used carefully but can give a first indication about which dog breed is most common to share pictures in twitter and therefore is widly spread and popular in social media.

Table 1. Top 5 most predicted dog breeds in the dataset

Breed	Count
Golden retriever	137
Labrador retriever	94
Pembroke	88
Chihuahua	78
Pug	54

Tweet information

By analyzing the tweet data it can be clearly said, that there is a high correlation between retweet count and favorite count (r = 0.79) as shown in Figure 2.
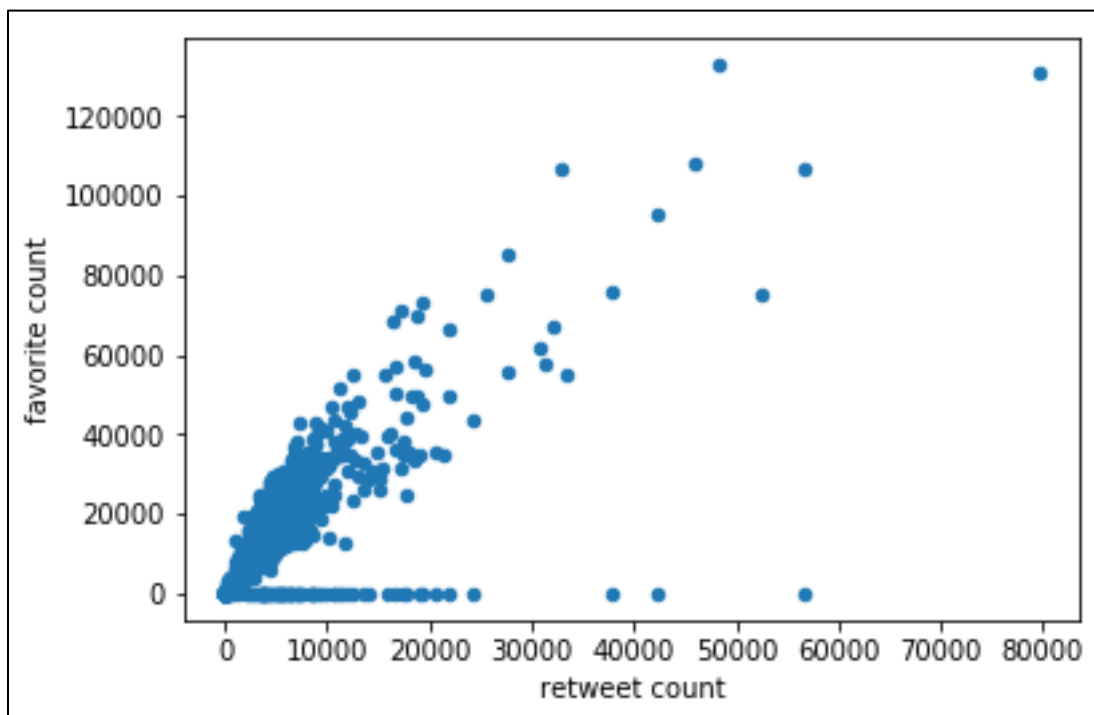

Figure 2. Scatter plot of the retweet and favorite count of single tweets

Finally a analysis of the tweet ratings grouped by the dog stage shows that there is a difference in the mean ratings. According to this the single stage puppo receives the highest mean ratings before doggo. This shows that there could be a difference in perception of dog tweets and ratings based on the dog stage, with the tendency that tweets with younger dogs, as puppo, receive a higher rating.

Figure 3 also shows that there is a small difference in the mean rating depending on the dogs breed.
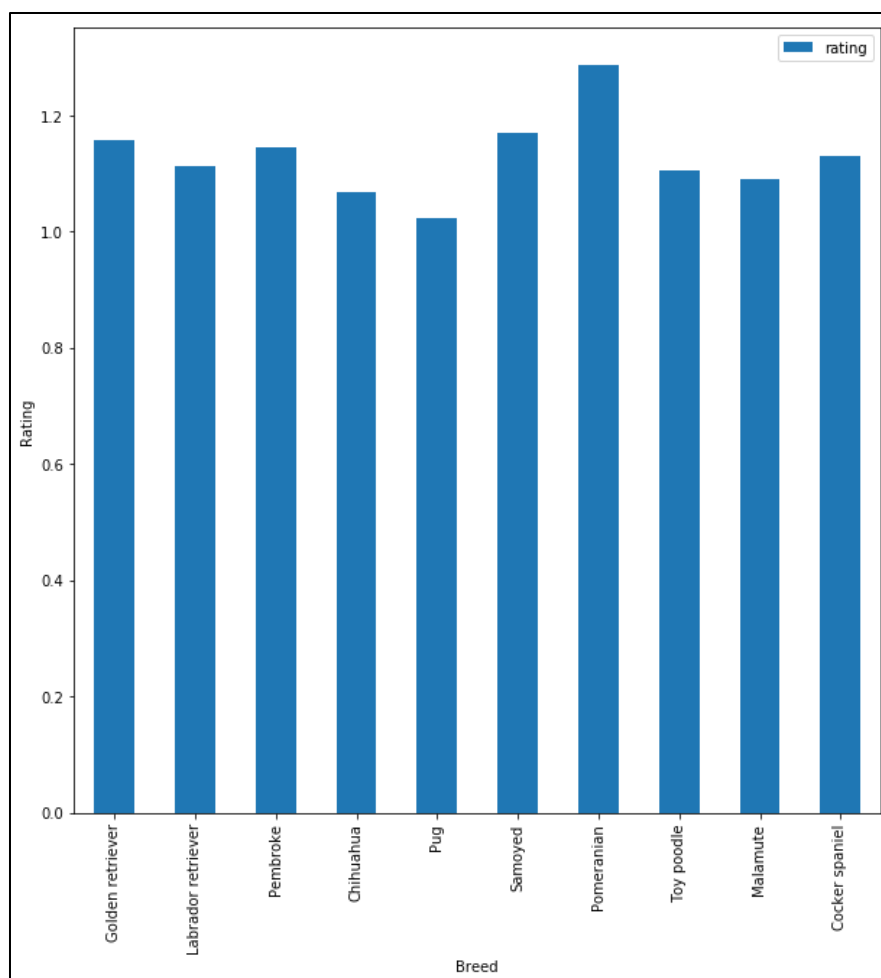

Figure 3. Mean rating of the 10 most common dog breeds

With this in mind, tweeter users could use this information to receive a high rating by respecting the dog's stage and breed. Nevertheless a detailed statistical analysis is needed to confirm this hypotheses.