Data Insight Report

Udacity nanodegree „Data Analyst“

Verena Dietrich - 22.05.2021

“What dogs are the most popular dogs on twitter?”, to answer this question I had a closer look on the twitter account WeRateDogs. This is more a fun account that does not want to discriminate specific dogs. Therefore, it is common to use a rating that is above 100% (Figure 1: Frequency of ratings), since every dog is beautiful in its own way. 75% of the ratings are 100% or greater than 100%. Nevertheless, the rating still represents which dogs can be considered as more popular than others.

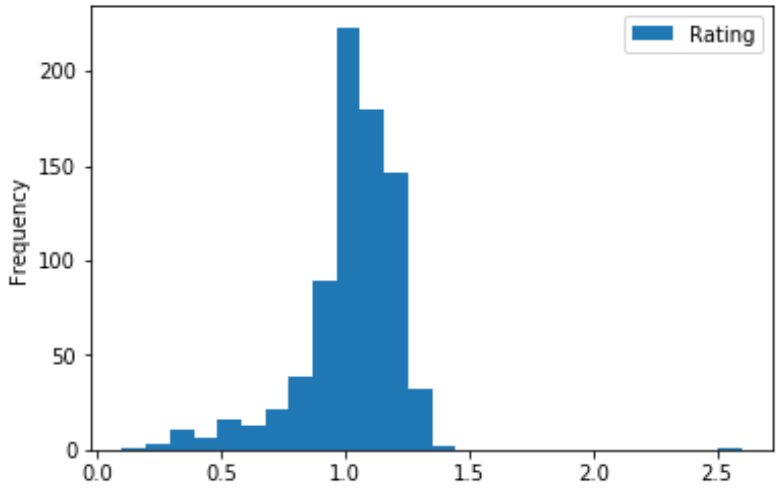

Figure 1: Frequency of ratings

To evaluate the popularity of dogs on twitter, I saw three parameters which should be considered:

1. The rating
2. The number of likes
3. The number of retweets of the tweet of interest.

Taking these parameters equally in account, the Eskimo dog is the most popular dog (Table 1: Popularity of dog breads; The higher the scoring the higher the popularity considering the parameters rating, number of likes, number of retweet).

Table 1: Popularity of dog breads; The higher the scoring the higher the popularity considering the parameters rating, number of likes, number of retweets

bread	scoring
eskimo_dog	168.0
great_pyrenees	165.0
kelpie	164.0
norfolk_terrier	160.0
golden_retriever	157.0
toy_poodle	151.0
shih-tzu	144.0
pembroke	143.0
labrador_retriever	133.0
cardigan	130.0

Separately considered the Eskimo dog is also the bread with the highest number of retweets and likes. The dog bread with the highest average rating is the Kuvasz. The Eskimo dog is at position 7 for the rating (Figure 2: The on average highest rated, most liked and most retweeted breads).

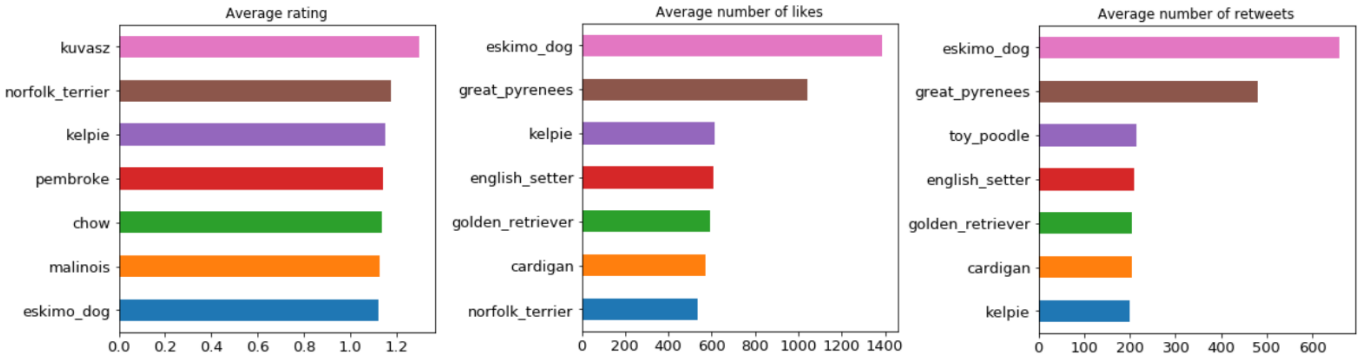

Figure 2: The on average highest rated, most liked and most retweeted breads

This already suggests a positive correlation between the number of retweets and the number of likes. This assumption can be conformed by the correlation coefficient. Correlation coefficient for number of likes and number of retweets is with 0.98 almost 1. This is a clear correlation especially compared to the correlation coefficients of 0.14 for the scoring and number of likes and 0.13 for scoring and number of retweets (Figure 3: Correlation und distribution for rating, number of counts and number if retweets). Here are only breads considered that have at least three tweets with an image with a dog of this bread in the foreground.

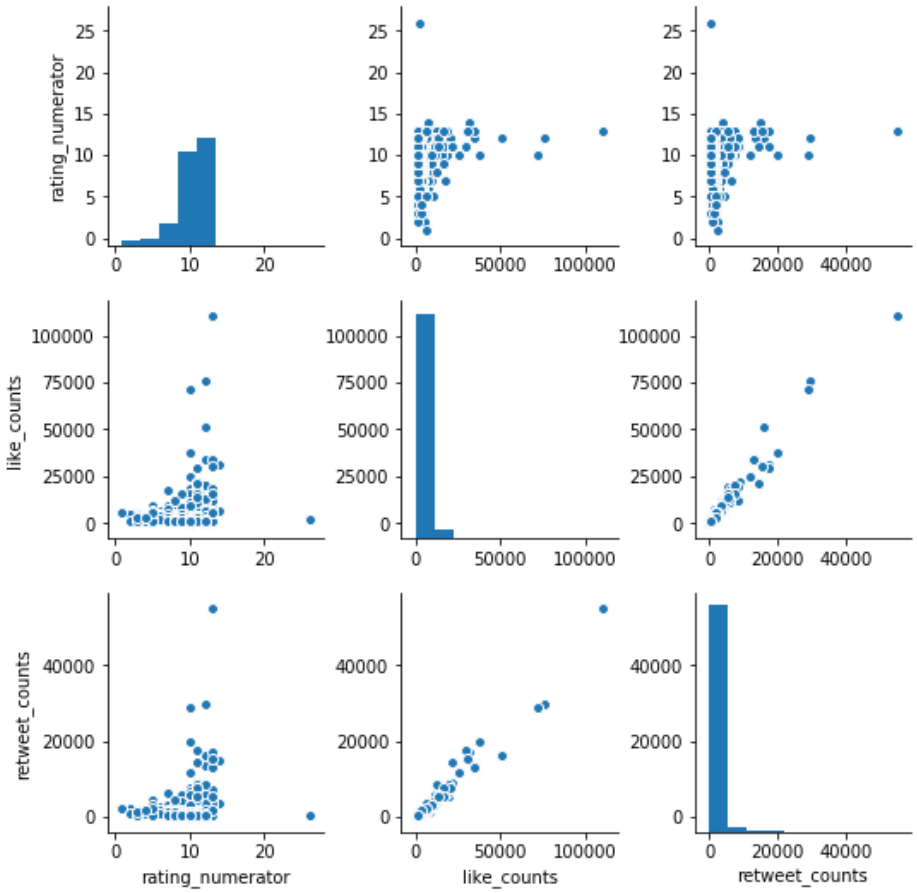

Figure 3: Correlation und distribution for rating, number of counts and number if retweets

Beside the bread, another factor that could influence the popularity of a tweet could be the stage of the dog. It can be assumed that in general younger dogs like puppers are more popular than older dogs. To evaluate this hypothesis, I performed three hypothesis testing for the three parameters against the significance level of 5%. The Hypothesis to accept or reject is in all three cases:

$$H_0 : \mu_{all} - \mu_{pupper} >= 0$$

$$H_1 : \mu_{all} - \mu_{pupper} < 0$$

This hypothesis can not be confirmed for all three parameters.

For the rating we can assume puppers get on average a higher rating than other dogs. With a significance level of 5% and an observed p-value of larger than 2.1%, the H_0 hypothesis can be rejected. Therefore, I assume the difference in the observed ratings for the puppers and all dogs is significant. I assume puppers get on average a higher rating than other dogs.

For the number of likes and the number of retweets, we have to accept the H_0 hypothesis with a p-value of 76.6% for the number of likes and 82.4% for the number of retweets. Therefore, we assume that the number of likes and the number of retweets for the puppers are on average not higher than the number of likes for other dogs.