

Sahand Niasti

Analyzing and Visualizing Data

Introduction

A Twitter account, called WeRateDogs, rate dogs and users make funny comments about them based on their pictures. The dogs are “supposed” to be rated from 0 to 10 but a lot of ratings excess the maximum as an inner joke. The account rapidly became popular and after only two years, in October 2017, the account had about 3.75 million followers, which is truly outstanding for a rating page. We have already wrangled data and made it ready for further analysis. The process of data wrangling is explained in a different file (wrangle_act). Here, I will show some part of my analysis.

Popularity of dog breeds

First of all, I analyzed the relationship between the numbers of favorite and retweet and dog breeds to find out which dog breeds are the most beloved by the users. I used the most relevant results of the machine learning values, categorized the dog breeds based on pictures. According to my results, golden retrivers are the all time favorites among the users with more than 1,900,000 favorites, the second and third breeds are labrador retrievers (about 1,200,000 favorites) and pembrokes (about 1,000,000 favorites).

Table 1- Top 10 dog breed

breed	retweet_count	favorite_count
golden_retriever	586657	1943782
labrador_retriever	405404	1265355
pembroke	291771	1038797
chihuahua	250722	743689
samoyed	203446	583906
french_bulldog	156230	570921
cocker_spaniel	145743	407412
chow	131025	446933
pug	118682	383817
toy_poodle	115330	337270

I also analyzed which dog breed was mentioned the most in tweets and recognized by image prediction process. Golden retriever is still the most popular dog among WeRateDogs users throughout the period.

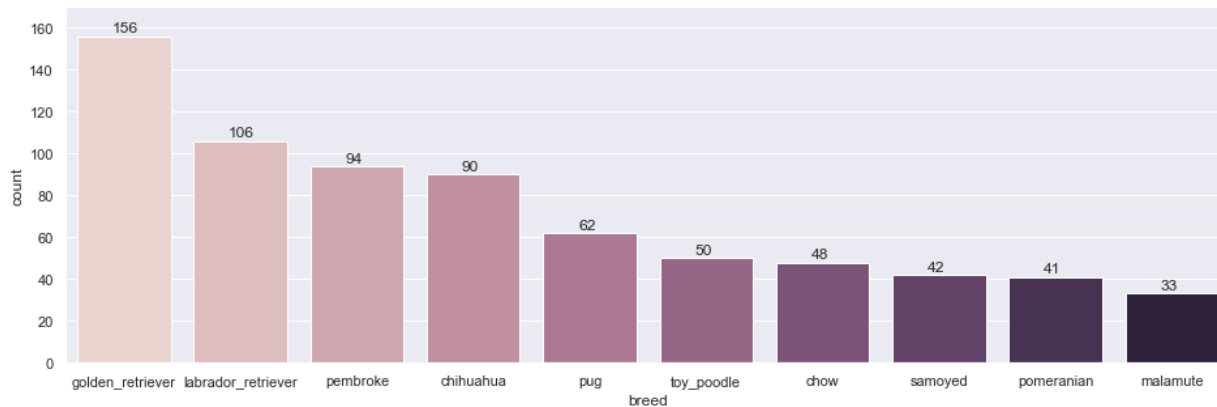


Figure 1 - Count of classified breeds in the dataset

Until now, we have calculated the sum value of retweets and favorite counts. but now, we want to compare single tweets.

Table 2 – the top 5 most favorite (on the left) and retweeted (on the right) breeds

retweet_count	favorite_count	breed	retweet_count	favorite_count	breed
48265	132810	lakeland_terrier	79515	131075	labrador_retriever
79515	131075	labrador_retriever	56625	107015	chihuahua
45849	107956	english_springer	52360	75163	eskimo_dog
56625	107015	chihuahua	48265	132810	lakeland_terrier
32883	106827	french_bulldog	45849	107956	english_springer

It can be seen that the most favorite tweet belonged to a Lakeland terrier breed with around 132,000, and the most retweeted dog breed as a single tweet belonged to a Labrador retriever. When we look at both tables, we cannot find our all-time favorite breed (golden retriever) among them. Let's sum up them and show aggregated retweet and favorite count in a single tweet and find out which breed is the most favorite one.

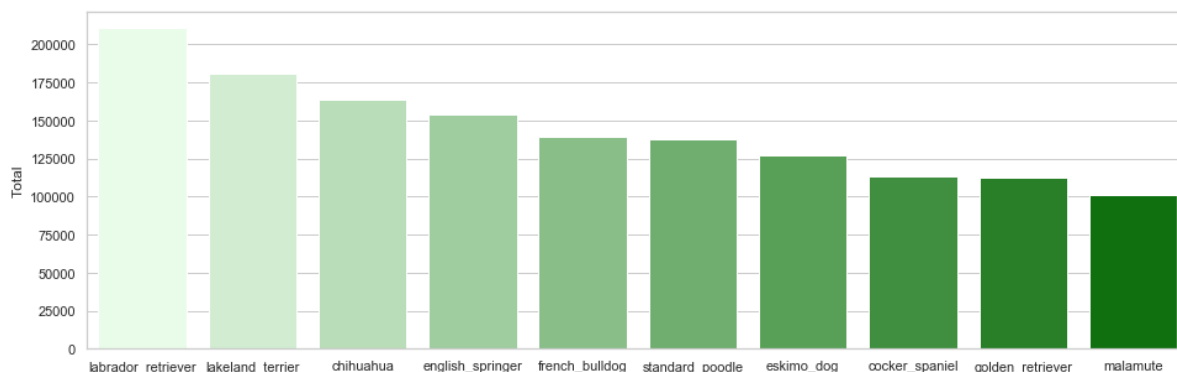


Figure 2 – Total retweets and favorites grouped by dog breeds

Interestingly, golden retriever is the 9th most popular breed among single most popular tweets. In this category, Labrador retriever with more than 200,000 favorite and retweet is the most popular one. Lakeland terrier and chihuahua are in second and third place respectively.

Tweets per day

Now is the time to see which days of the week users of WeRateDogs make more comments and retweets. By looking at bar chart below, it is clear that first day of the week is the most common day which users tweet during a week. As we arrive the weekend, tweets per day reduced until Sunday.

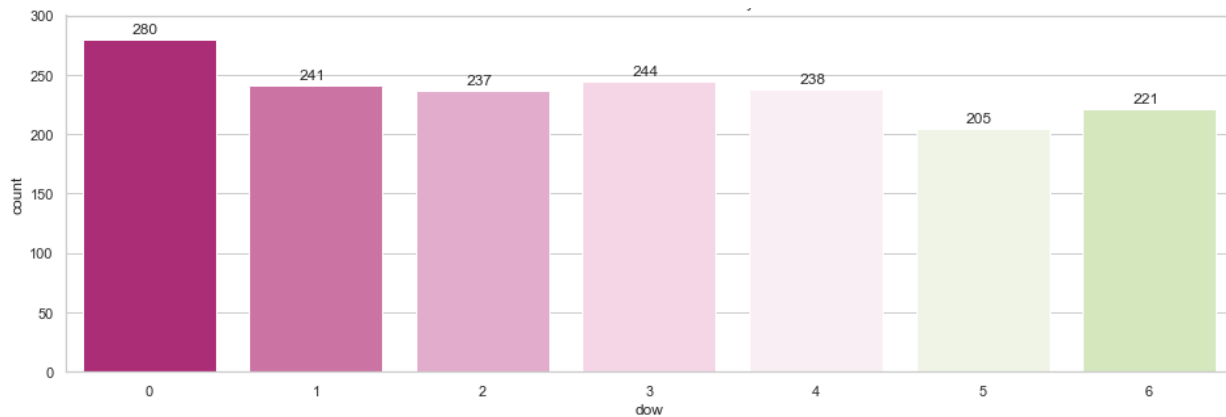


Figure 3 - Count of tweets in the dataset for each day of the week

Dog Stages

I analyzed the specific "dog stages", excluding the missing values to make a pie chart about the proportions of the categories. The most often used stage is "pupper" with more than 65%, the second one is "doggo" (about 22%). The account owners use the words puppo and floofer in 10% of the cases all-together.

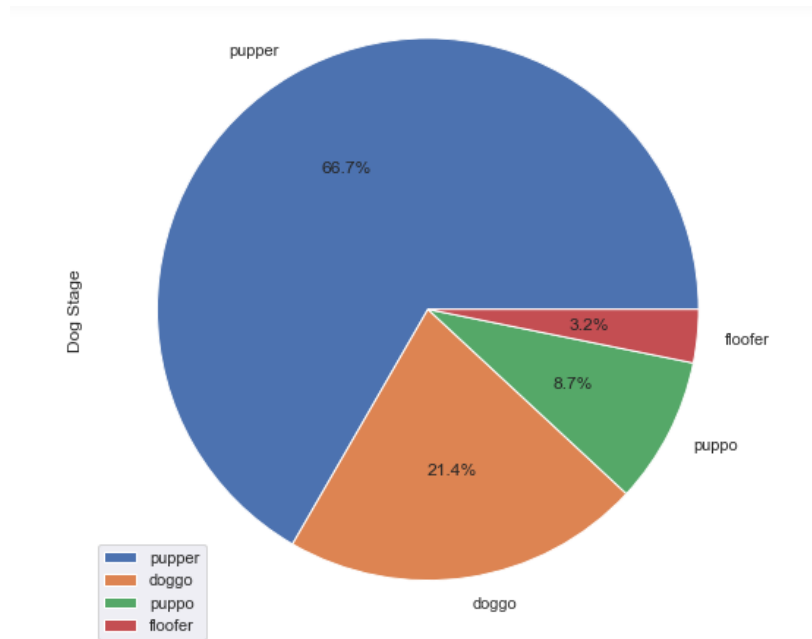


Figure 4 – Proportion of dog stages

Most Popular Dog Names on the Account

I would like to find the most popular dog names. As rating has not anything to do with popularity, in this case, I will rate popularity by the number of tweets, assuming that the most popular dog names occur more times on the site.

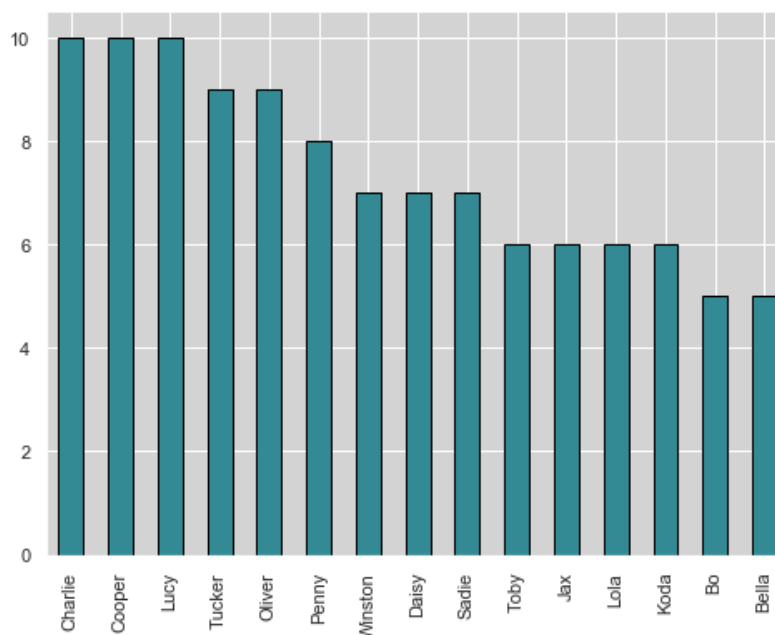


Figure 5 – Popularity of dog names