Act report of tweet archive 'WeRateDogs'

The cleaned the data and created a dataset that includes 22 columns and 1656 rows. We will conduct an analysis based on this data.

Most interesting observations

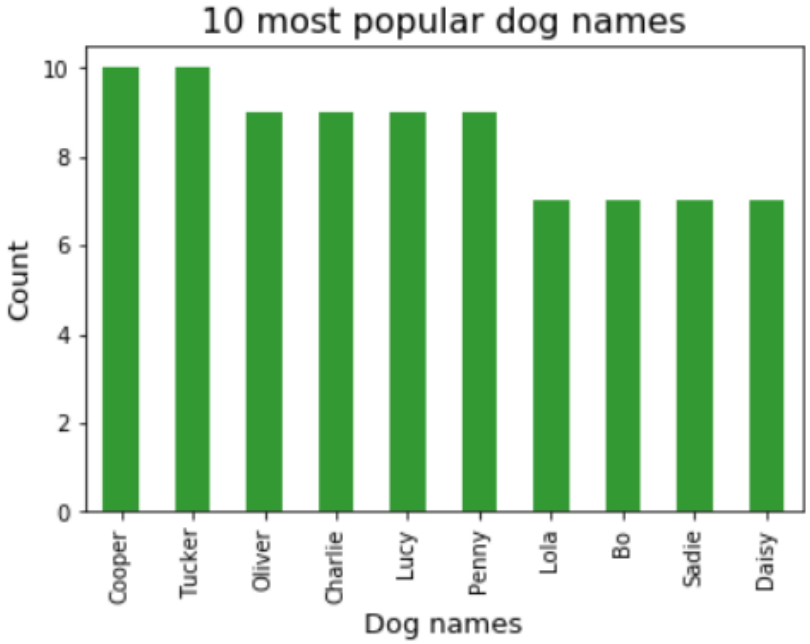
Observation 1

Of course, we were interested in finding the tweet that received the most retweets. Having done some simple manipulations using the *retweet count* column, we got a link to the desired tweet. It turned out t o be a video tweet with a nice dog:

Tweet description: “Here's a doggo realizing you can stand in a pool”
Link: https://twitter.com/dog_rates/status/744234799360020481

Observation 2

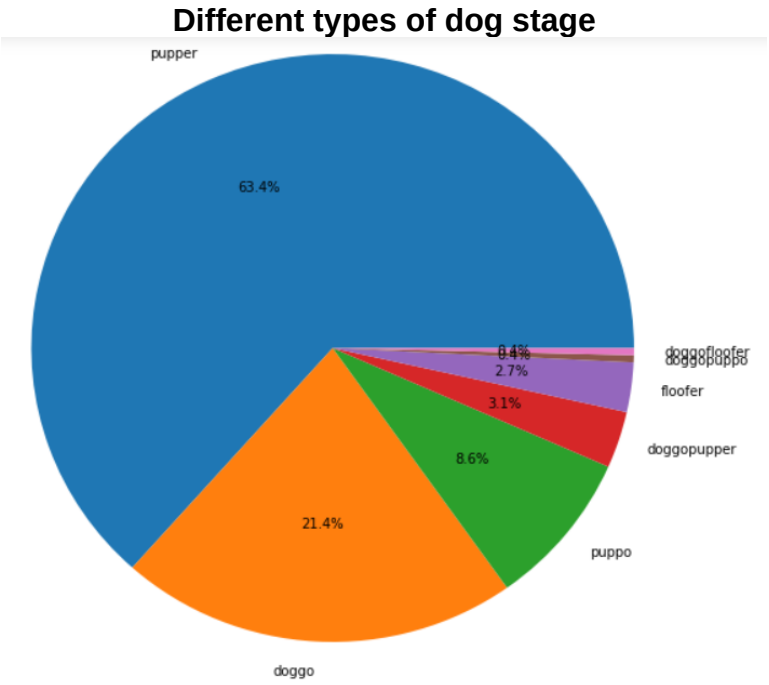
An equally interesting observation was to find out what are the most popular names for dogs. We hav e selected the top 10 for you:



And we see that the leading position in the **top 10 popular dog names** are shared by the names: Tucker and Cooper.

Observation 3

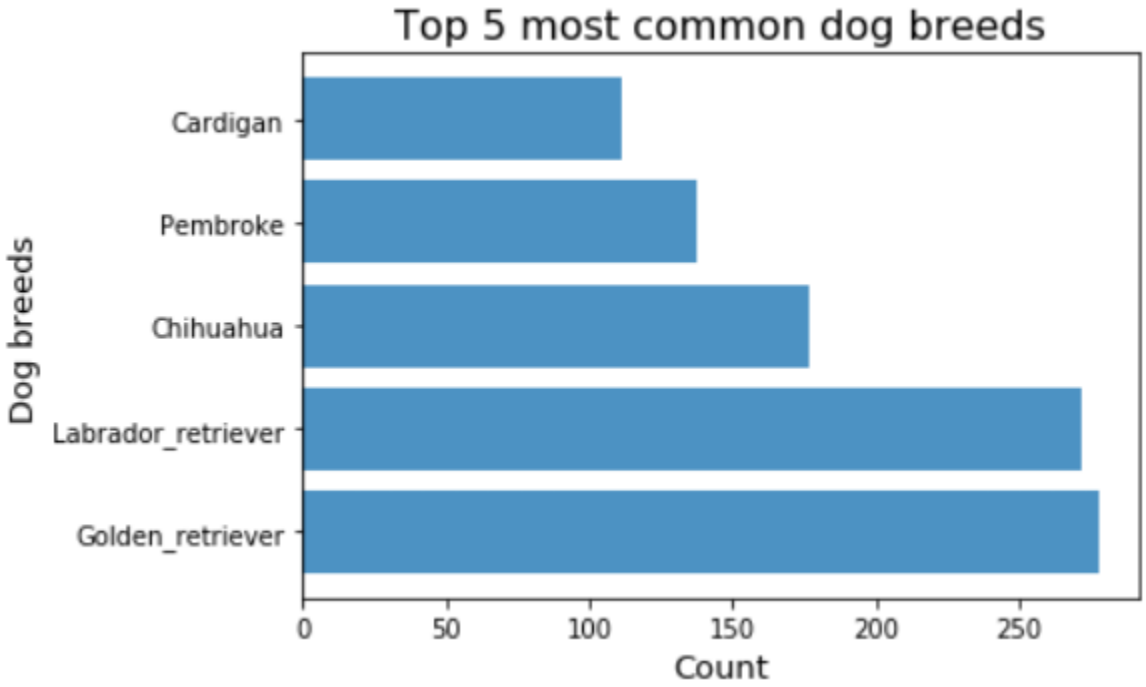
Let's take a look at the most common stages of dog development in **WeRateDogs** tweet archive. For this, we created a pie chart with different values of the dog's stage as a percentage.



Most of the dogs are in a **pupper** stage (63.4%)

Observation 4

Top 5 common dog breeds in tweet archive **WeRateDogs**
For this task, we created a single list of all available dog breeds from our dataset. Sorted the list of dog breeds by frequency of occurrence in the full list and created a plot with top five common dog breeds.



And our result for breeds into images:

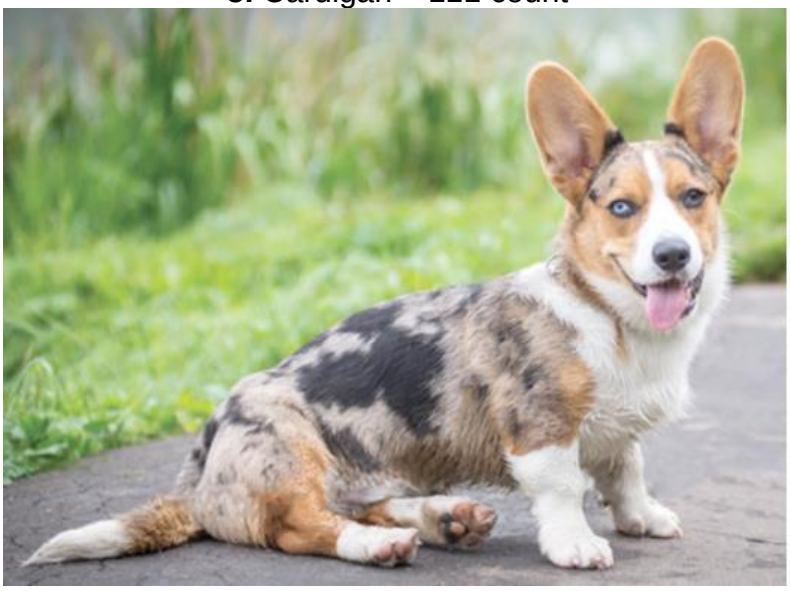
1. Golden Retriever – 278 count

2. Labrador Retriever – 272 count

3. Chihuahua – 177 count

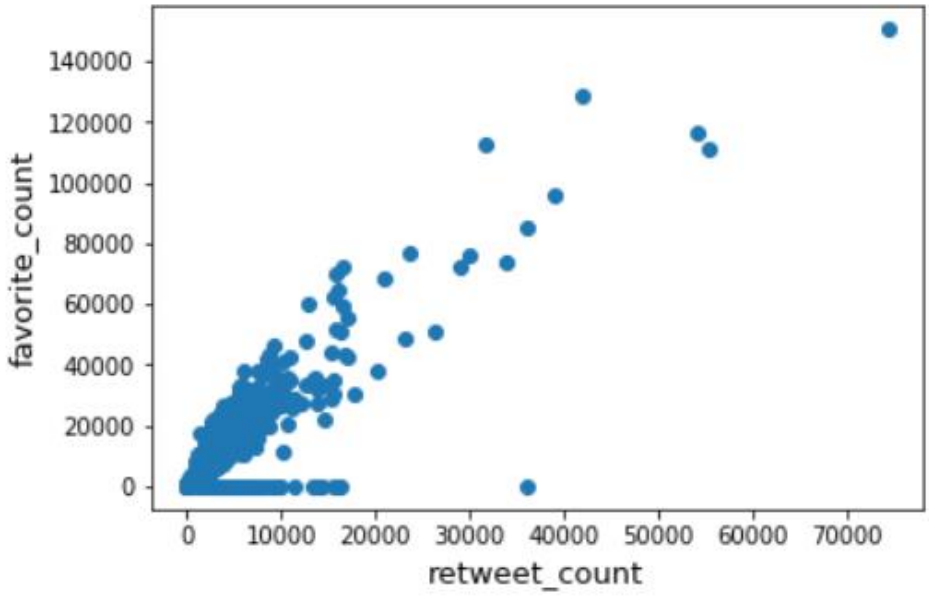
4. Pembroke – 138 count

5. Cardigan – 111 count



Observation 5

It is logical to assume that there is a correlation between *retweet_count* and *favorite_count*, let's take a look at it.
For this task, it is convenient to use a **scatter diagram**, and that's what we did!



As expected, we see a fairly pronounced positive correlation between *retweet_count* and *favorite_count*.
We also found the correlation coefficient (**r**) for these variables using the Pandas library function **corr()**. And our correlation coefficient = **0.86**, which indicates a **Strong positive relationship**.