

Wrangle and Analyze Data Project

Act Report

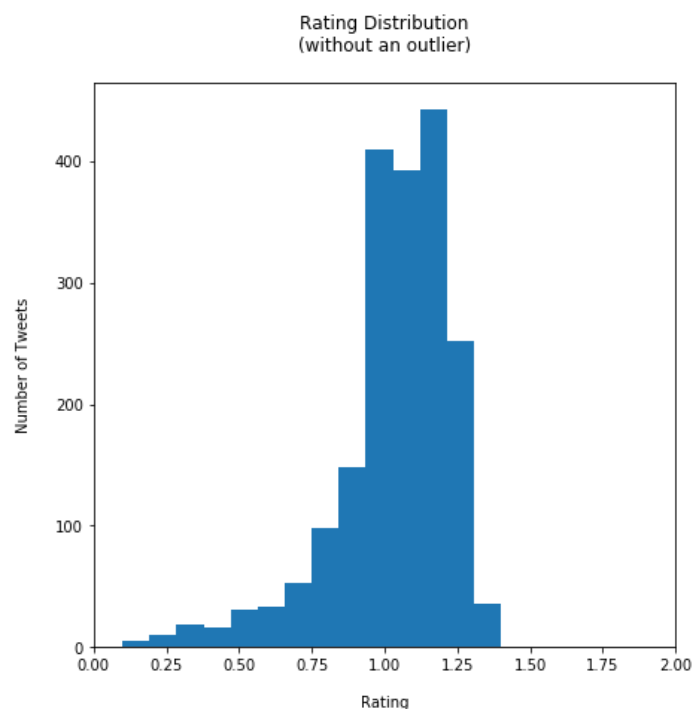
For analyzing and visualizing data I used two clean dataframes: `archive_tweets` (contain 2125 tweets) and `images_clean` (2073 tweets).

First of all, I check basic statistics for the numerical data in the `archive_tweets` dataframe: `favorite_count`, `retweet_count` and `rating`.

	<code>favorite_count</code>	<code>retweet_count</code>	<code>rating</code>
count	1948.000000	1948.000000	1948.000000
mean	8430.259754	2553.824949	1.143687
std	12495.974029	4610.488568	4.006036
min	75.000000	11.000000	0.100000
25%	1805.750000	565.750000	1.000000
50%	3792.000000	1223.000000	1.100000
75%	10470.000000	2889.250000	1.200000
max	161499.000000	81517.000000	177.600000

Rating Distribution

Secondly, I analyzed the rating distribution. Since there was an extreme value of the `rating` variable I considered it as an outlier and analyzed the rating equal or less than 20.



As we can see for the vast majority of dogs the rating is in the range from 0 to 1.5 and if we look closer for most of the dogs it is about 1 - 1.2 which means 10/10 - 12/10 in terms of the @WeRateDogs twitter account.

As for the outlier it is a [dog named Atikus](#) wearing festive attire to celebrate the Fourth of July. He is a cutie and definitely a patriotic dog and deserve that incredibly highest rating which in fact is a very important date for all Americans :)



WeRateDogs™ ✓
@dog_rates



This is Atticus. He's quite simply America af. 1776/10



11:00 AM · Jul 4, 2016 · [TweetDeck](#)

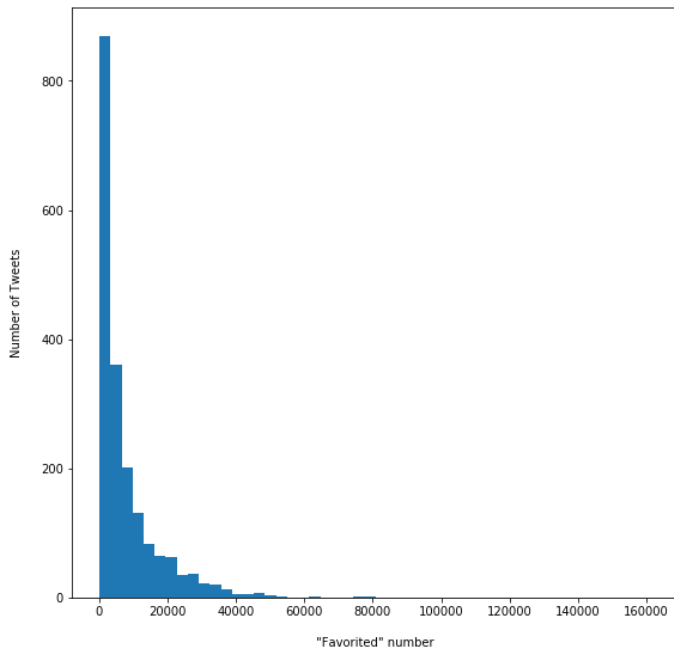
2.5K Retweets **5.3K** Likes



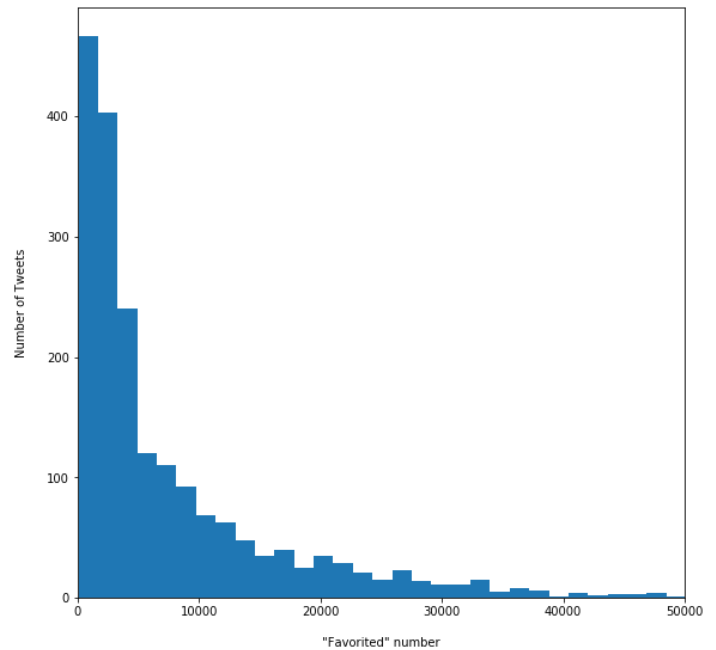
"Favorited" and "Retweeted" Distribution

Thirdly, I analyzed the distribution of the retweeted and favorited tweets using *favorite_count* and *retweet_count* variables.

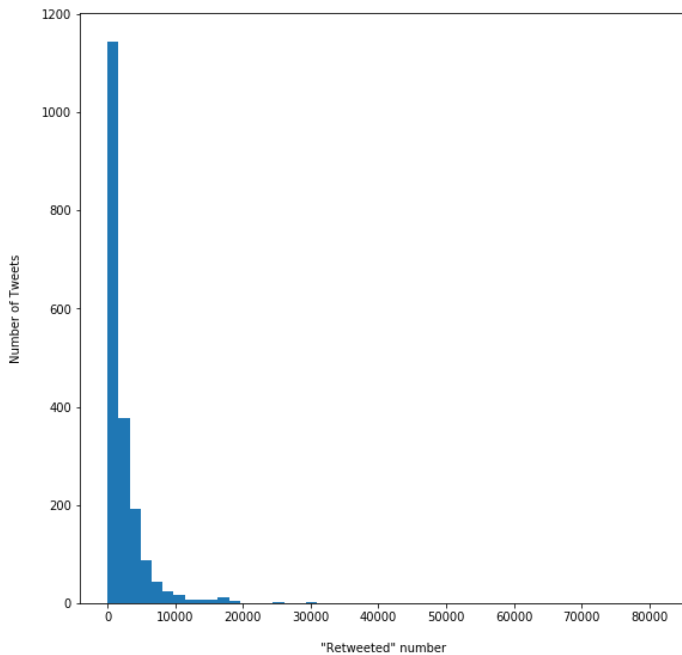
"Favorited" distribution



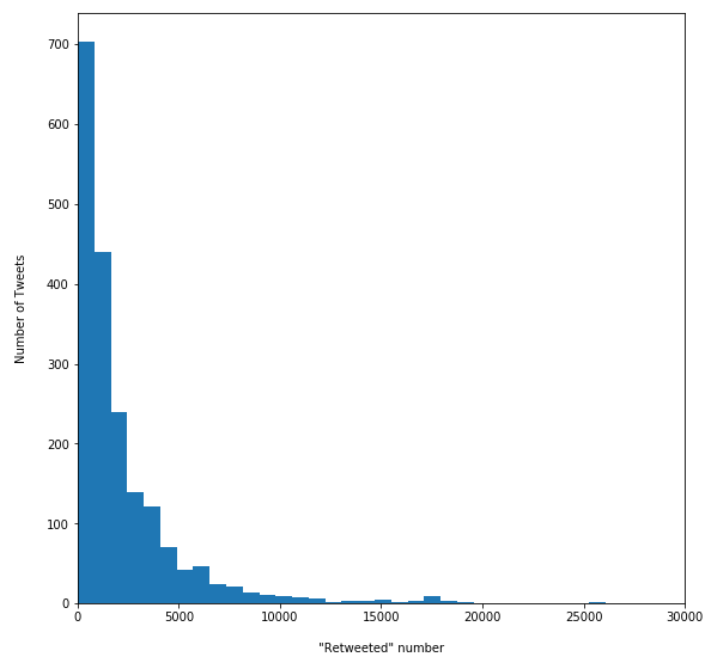
"Favorited" distribution, look closer



"Retweeted" distribution



"Retweeted" distribution, look closer



Both distributions are left-skewed which means that most tweets got less likes and retweets than smaller part of them and only a very small number of tweets got more than 20,000 likes and 10,000 retweets.

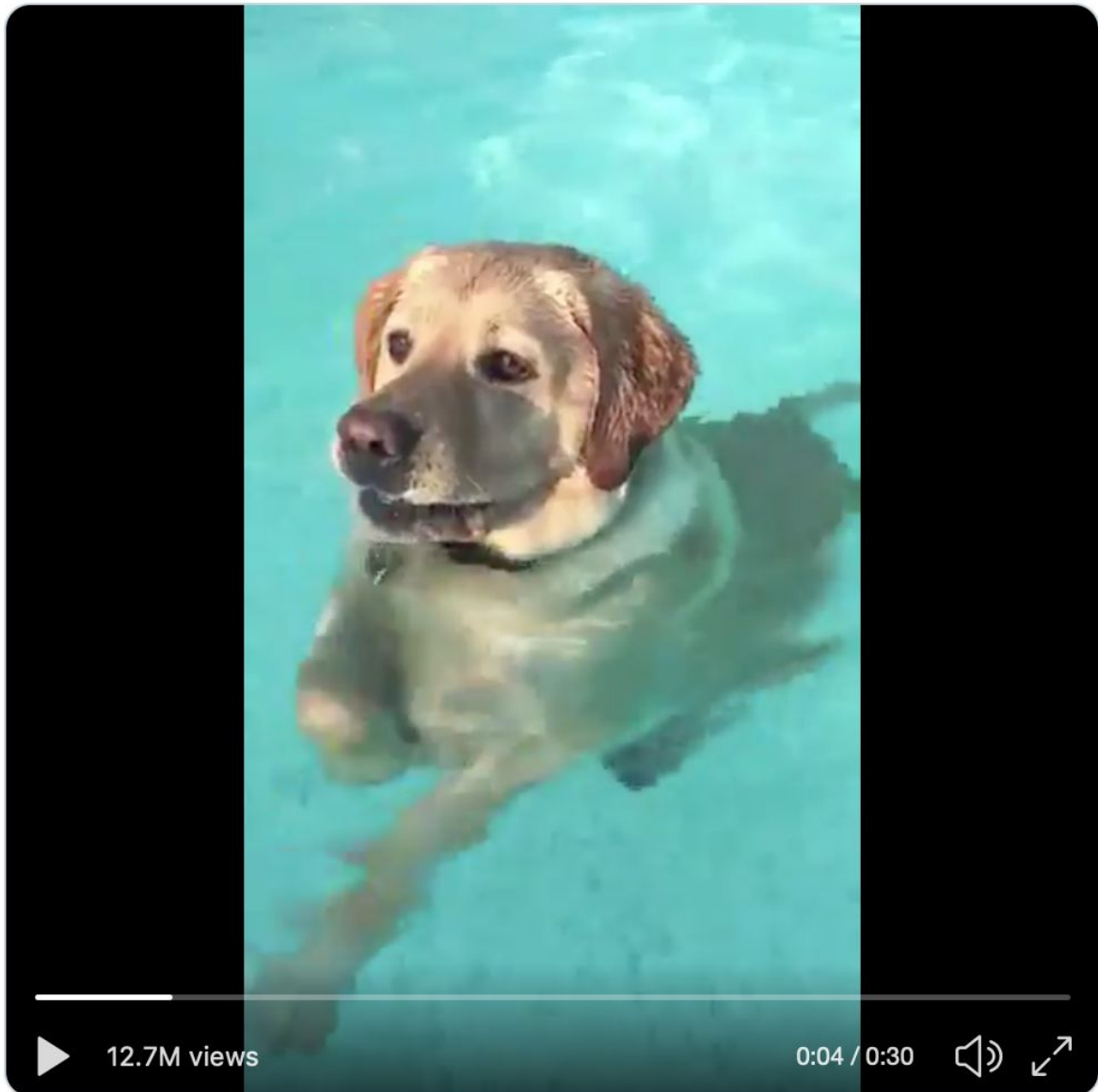
And the [most favorited and retweeted tweet](#) got more than 161,100 likes and 80,600 retweets. The screenshot is below but since it is a video click the link above to get some fun and take a look at a very cute swimmer.



WeRateDogs™ ✓
@dog_rates



Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad)

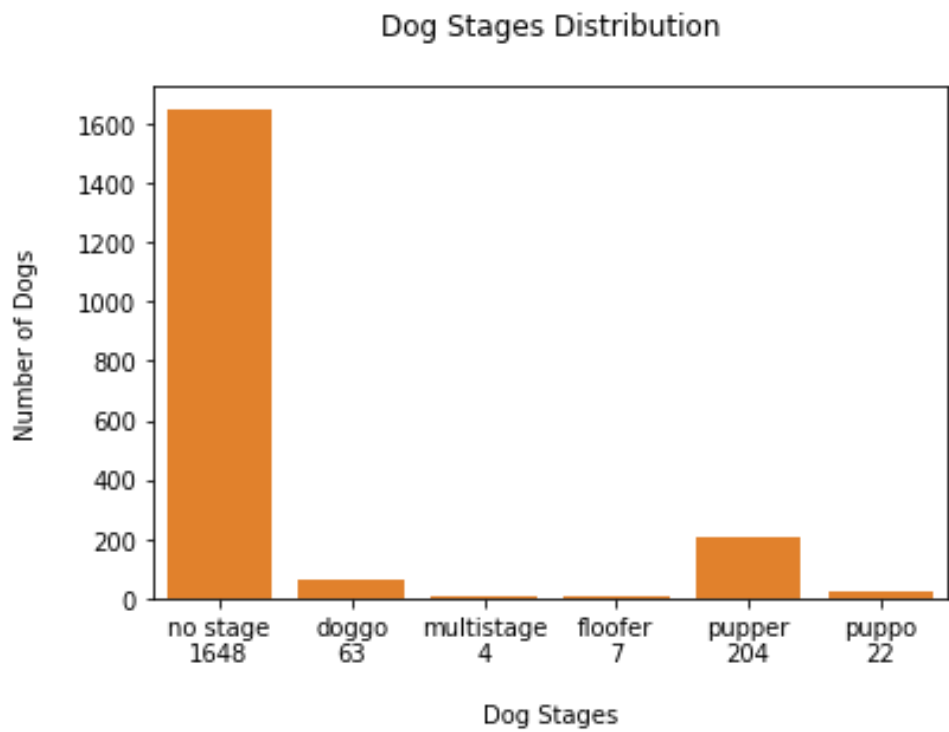


2:26 PM · Jun 18, 2016 · [Twitter for iPhone](#)

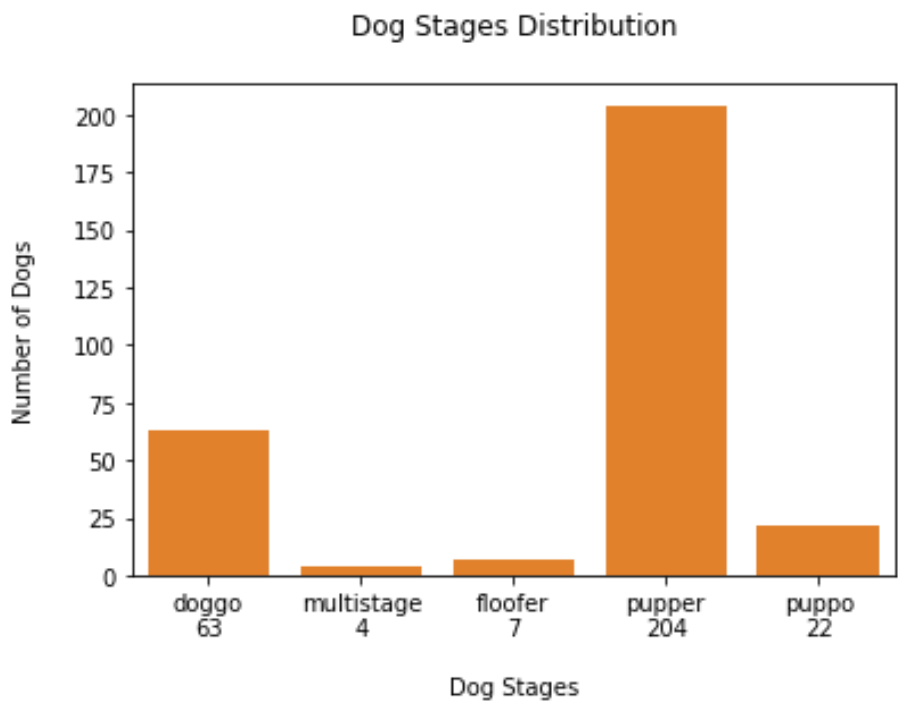
80.6K Retweets **161.1K** Likes

Dog Stages Distribution

Next, I analyzed dog stages distribution.



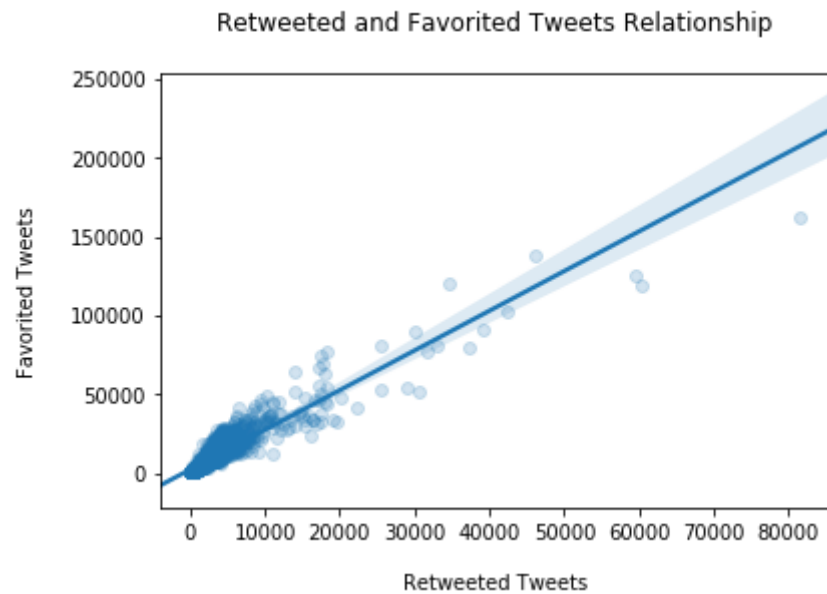
As we can see the vast majority of dogs do not have any dog stage so the next chart is a visualization for the distribution of only existing dog stages.



The most popular dog stage is a pupper which is a small doggo (a dog). The less popular stage is a floofer which means a fluffy dog. Obviously, there are not many tweets about fluffy dogs were posted by @WeRateDogs.

Retweeted and Favorited Tweets

I also found interesting to check if there are any relationships between tweets' retweets and likes.

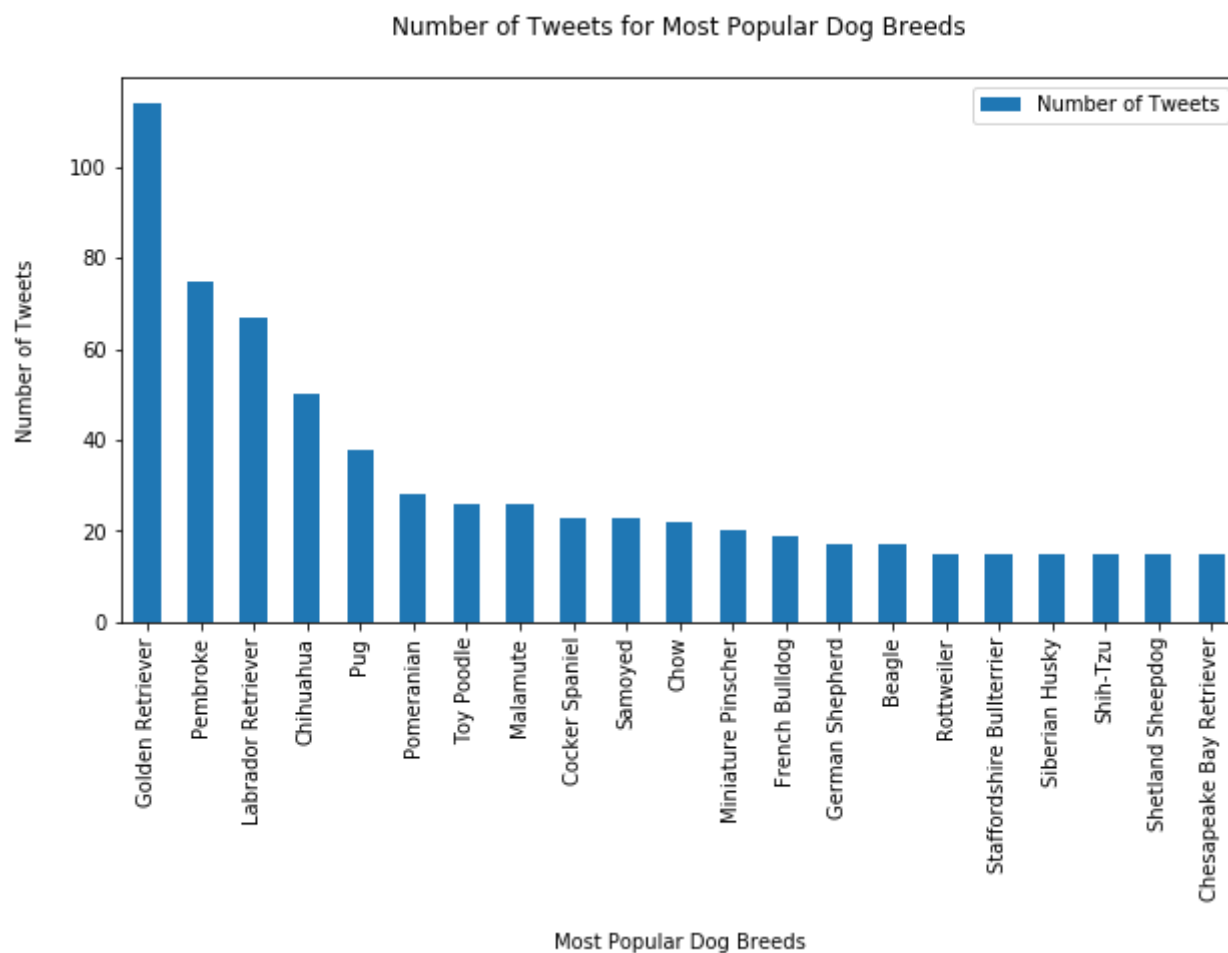


Clearly, people more often “favorite” tweets than retweet them. Besides that we can see that the more people “favorite” a tweet than often they retweet it.

Popular Dogs Breeds

To analyze dog breeds I created a new dataframe with only tweets for which all three prediction of what it is (dog or not, and dog breed if it is a dog) were a dog.

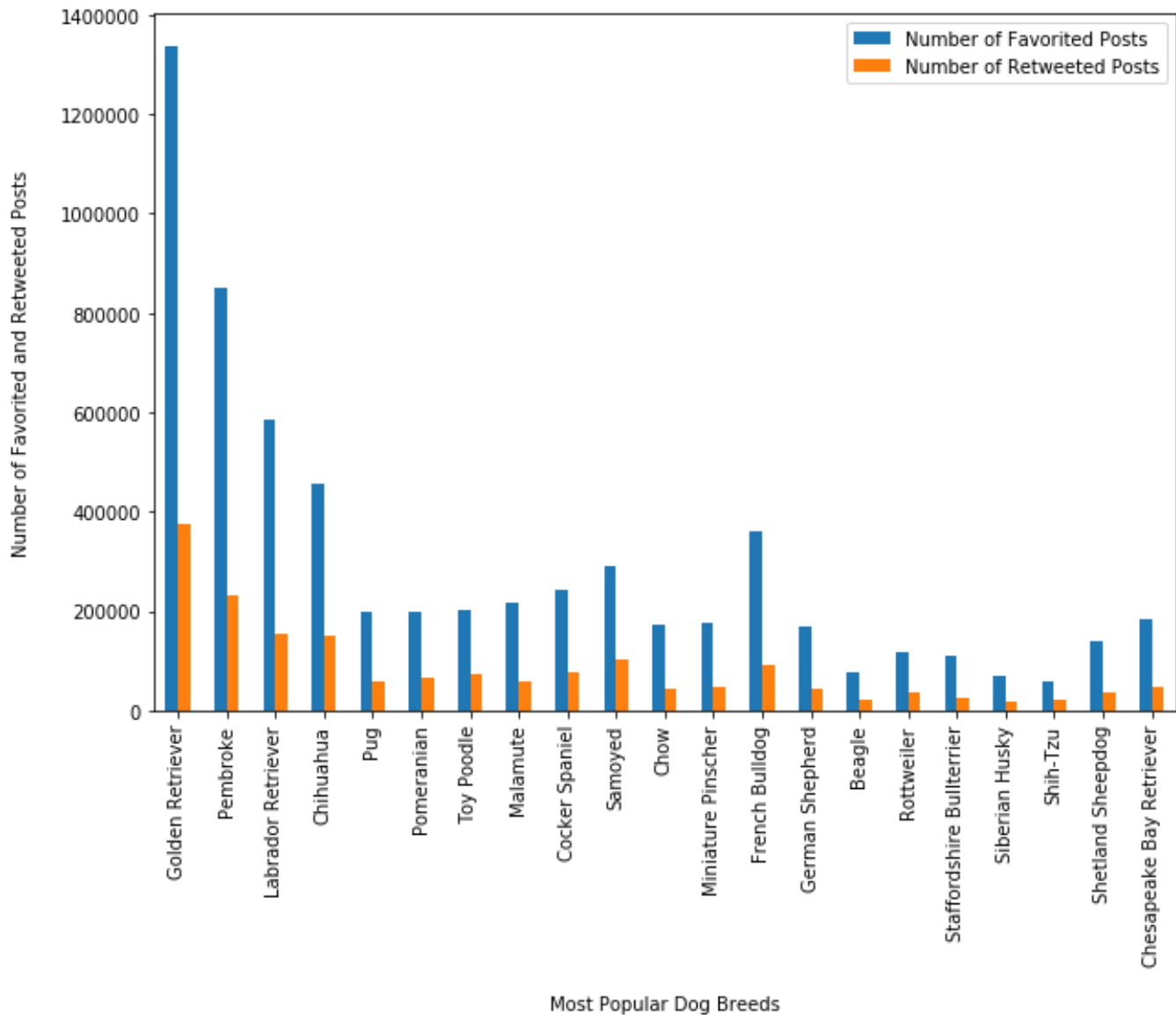
Then I selected only dog breeds which were mentioned in more than 15 tweets. So I got the most popular 21 dog breeds for further analysis.



As we can see on the chart above the most mentioned dog breed is a Golden Retriever. Since it is also one of the most popular dog breeds in the USA, Canada, UK and many other countries it is not surprising that this dog breed was mentioned in so many tweets.

On the following charts we can see if there is any relationship between the number of tweets about most popular dog breeds and the numbers how much these tweets were favorited and retweeted.

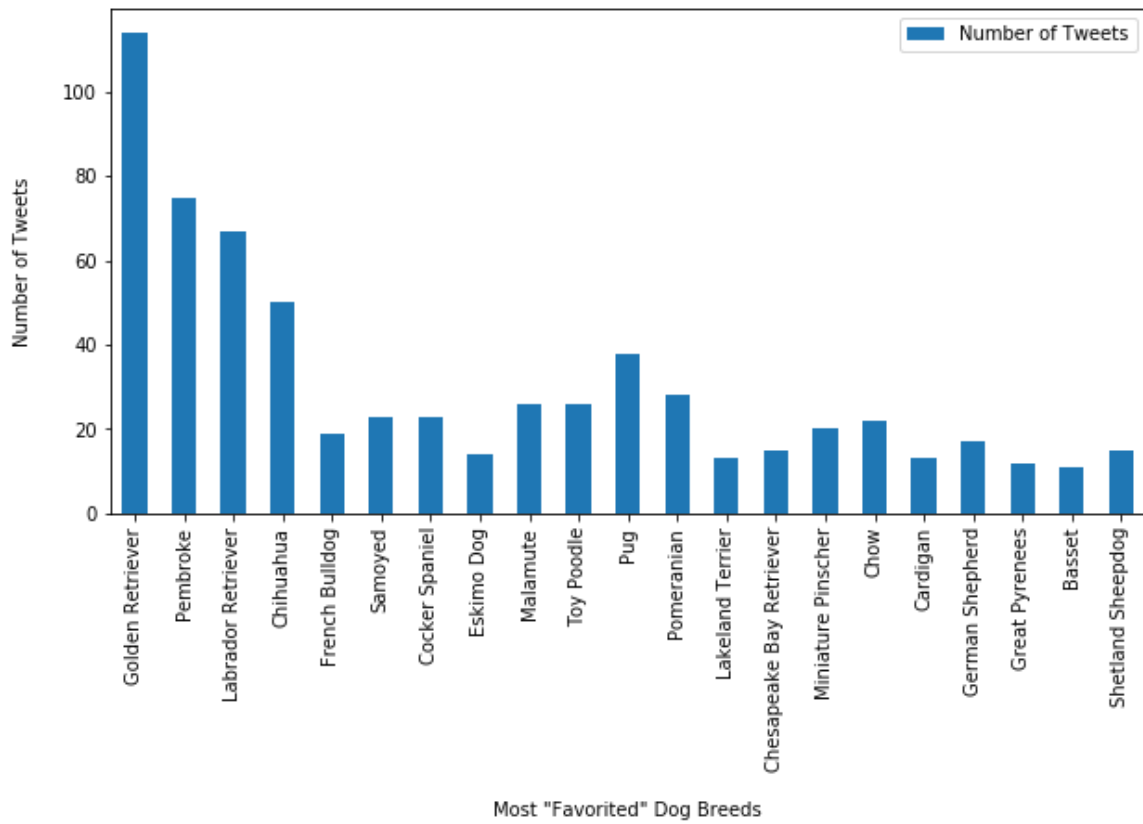
Numbers of Favorited and Retweeted Posts for Most Popular Dog Breeds



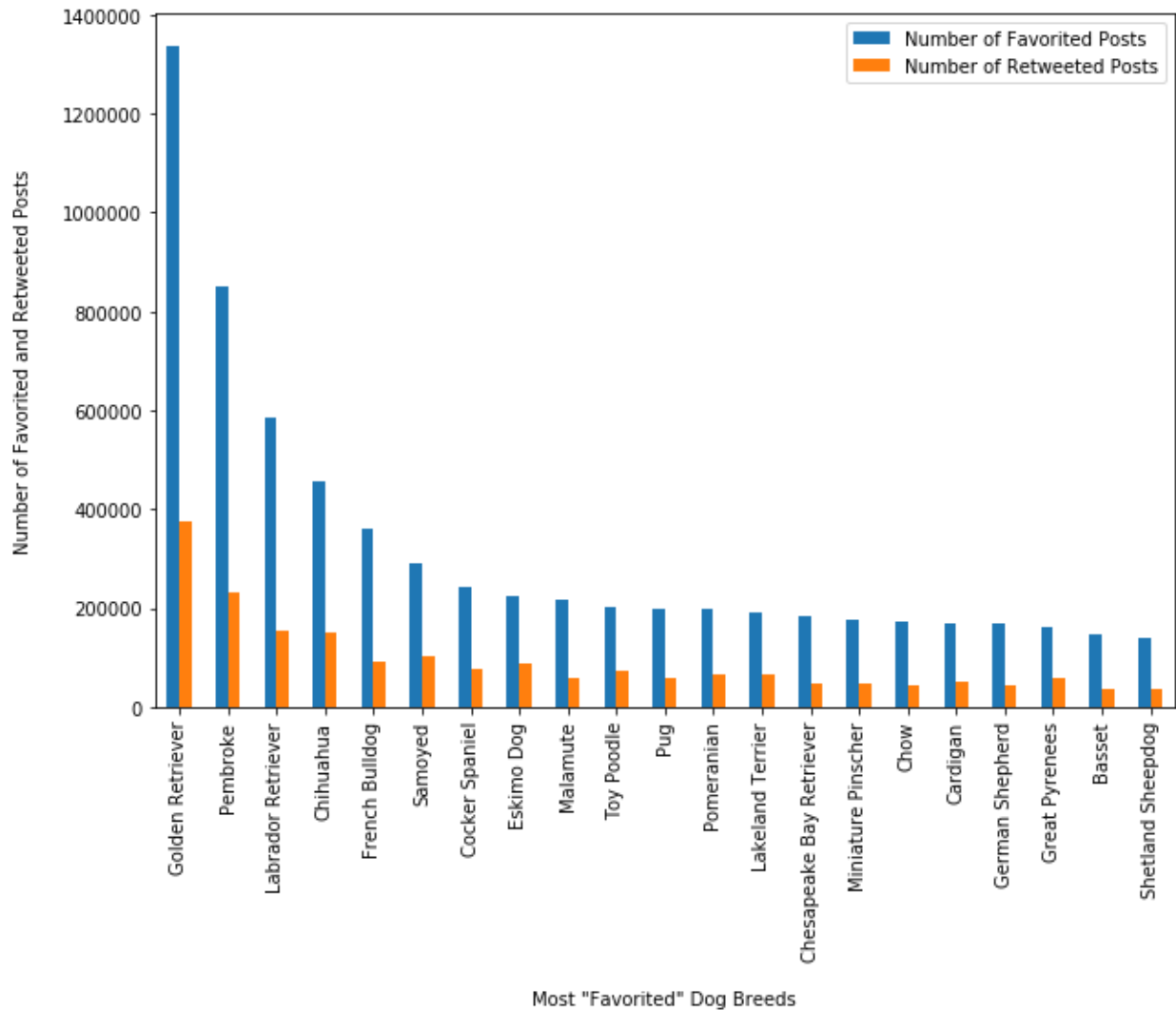
Although we could expect the more often the certain dog breeds were mentioned the more tweets about that dog breed were favorited and retweeted, we can see that it is true for only the first four dog breeds. For example, French Bulldog which was mentioned only 19 times got almost the same likes as Chihuahua which was mentioned in 50 tweets.

I also decided to test this idea for most “favorited” and most “retweeted” dog breeds and the results are on the following charts.

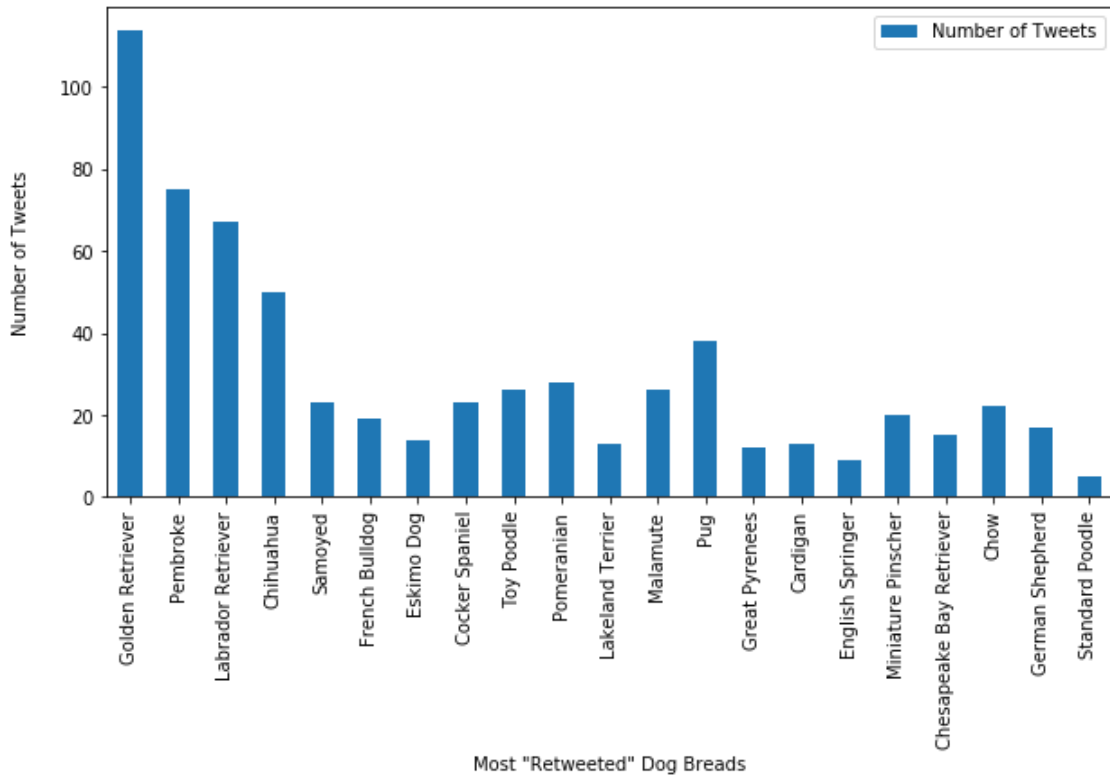
Number of Tweets for Most "Favorited" Dog Breeds



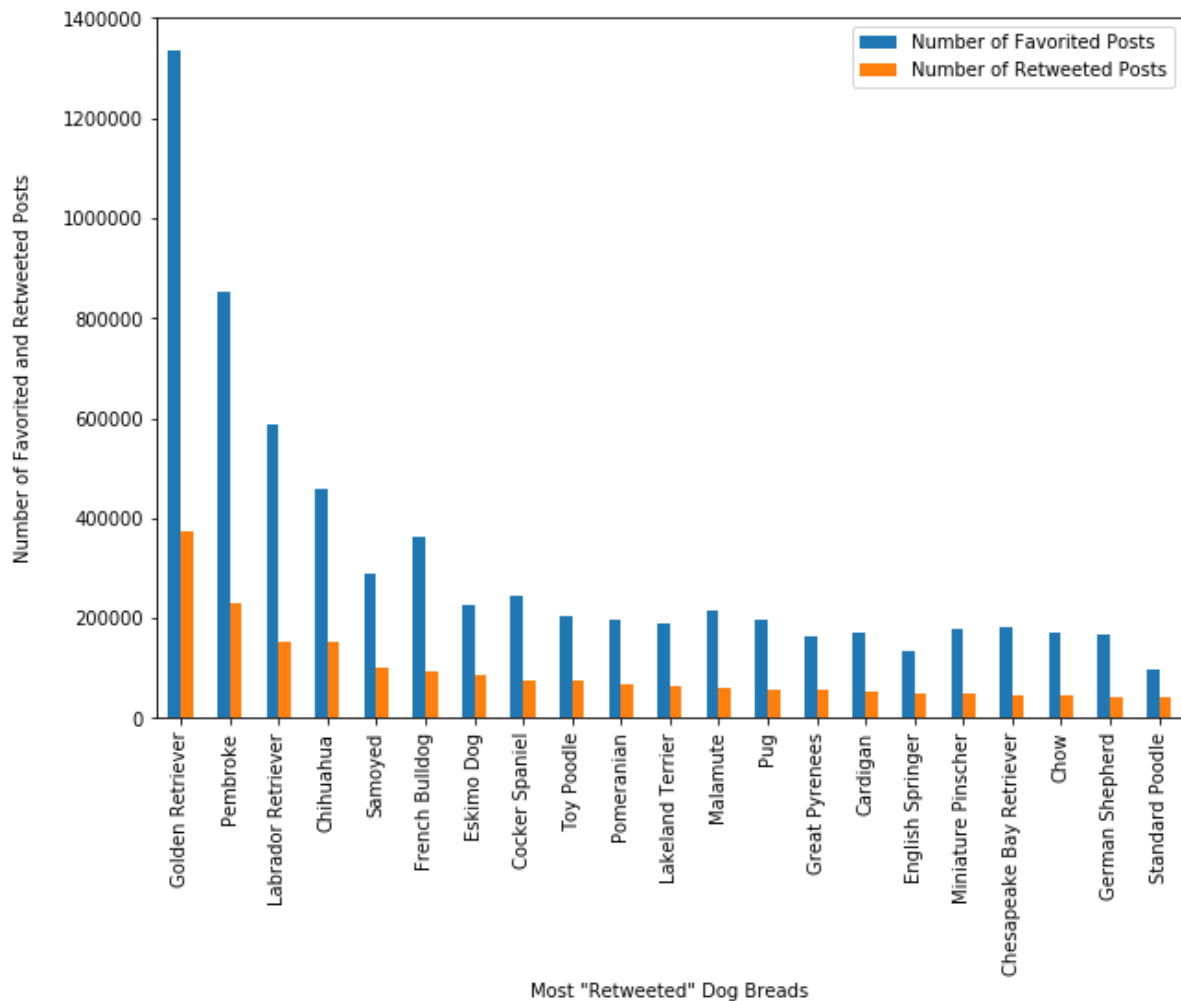
Numbers of Favorited and Retweeted Posts for Most "Favorited" Dog Breeds



Number of Tweets for Most "Retweeted" Dog Breeds



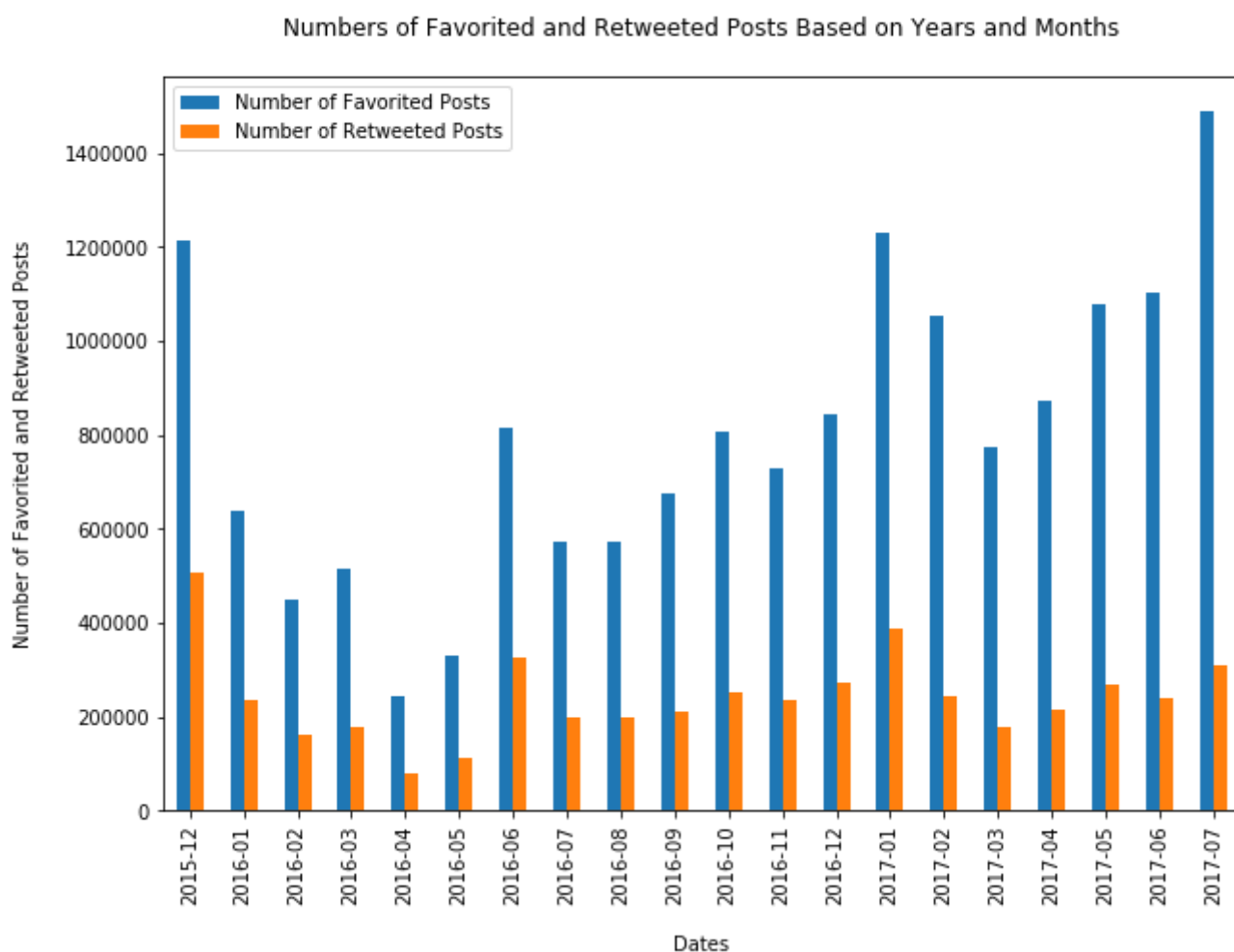
Numbers of Favorited and Retweeted Posts for Most "Retweeted" Dog Breeds



As we can see the only four dog breeds were most mentioned, favorited and retweeted in the same time. And from the fifth position we can see the differences. For example, French Bulldog is more favorited than Samoyed, yet the last one is more retweeted than French Bulldog while both of them got not so much tweets as the nine other dog breeds.

Favorited and Retweeted Posts Over Time

Finally, I tested the idea that numbers of retweeted and favorited posts increase over the time.



We can see that the number of the favorited tweets were gradually increasing within 20 months while the number of retweeted tweets just varied within certain limits. However, it is possible that 20 months are not enough to see the tendency clearly so this idea requires more data for further analysis.