Analyzing, and Visualizing Data for WeRateDogs

WeRateDogs is a popular account on Tweeter that rates people's dogs with a humorous comments. It has more than 8.8 million as of Sep 2020. The ratings usually have a denominator of 10, but the numerators most likely to be more than 10!!!

Data from three sources are gathered, assessed, and cleaned and ready for analysis. The result from the data wrangling phase is a one master dataset that we could use to get different insights regarding the data.

In the following sections will try to answer a couple questions which might be of interest for you, which are:

1- In general, how people tend to rate the dogs, is it right that many dogs are receiving rating with numerator higher than denominator?
2- Are highly rated dogs receive high favorite_count and retweet_count?
3- How the 'rating', 'favorite_count', 'retweet_count' and 'followers_count' changed over the years?
4- Which dogs' breeds are the most common? and how does this relate with the rating?
5- The last question, what is the most common name among dogs? and do the most common names receive high rating too?

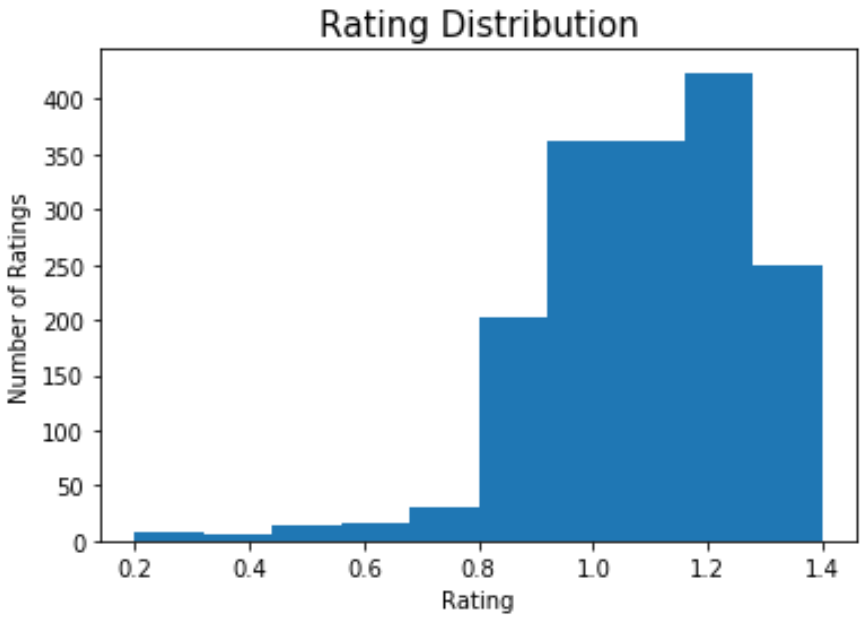
And here we start with answering these questions:

1- In general, how people tend to rate the dogs, is it right that many dogs are receiving rating with numerator higher than denominator?

Majority of dogs have rating above 1 (rating= numerator/ denominator) with mean equal 1.08 and more than 75% of dogs have rating 1 and above. this complies with the main feature of WeRateDogs!

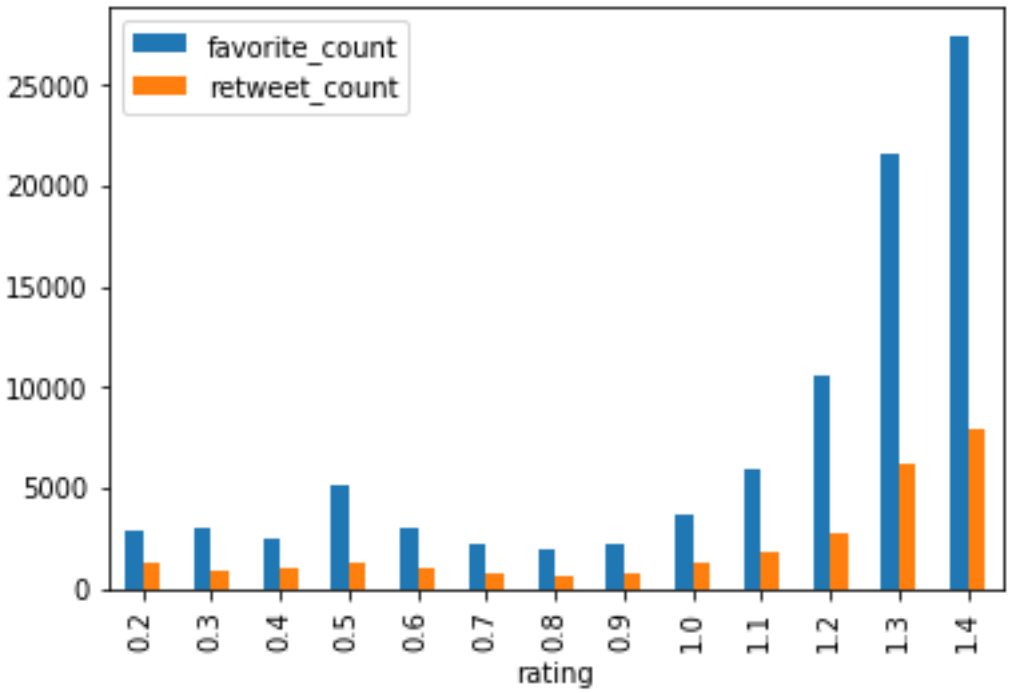
The following is illustrated in the following statistics and the plot:

count	1674.000000
mean	1.083692
std	0.176037
min	0.200000
25%	1.000000
50%	1.100000
75%	1.200000
max	1.400000

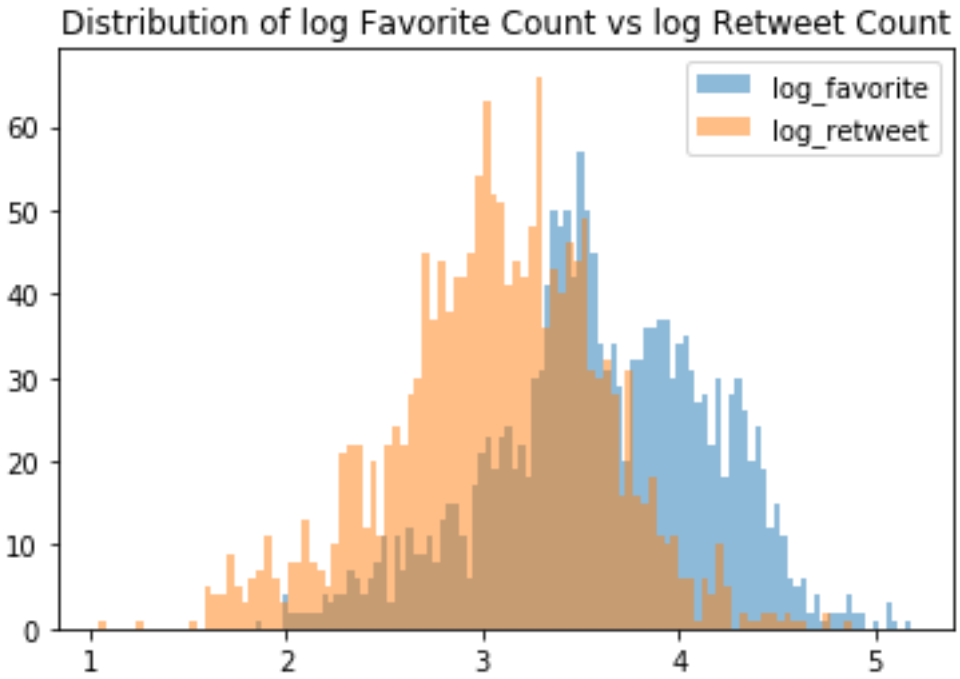


2- Are highly rated dogs receive high favorite_count or retweet_count?

The following plot shows the distribution of favorite_count and retweet_count based on the rating. High rating dogs receive high favorite count. The same with retweet count but the difference is more significant in case of favorite count.

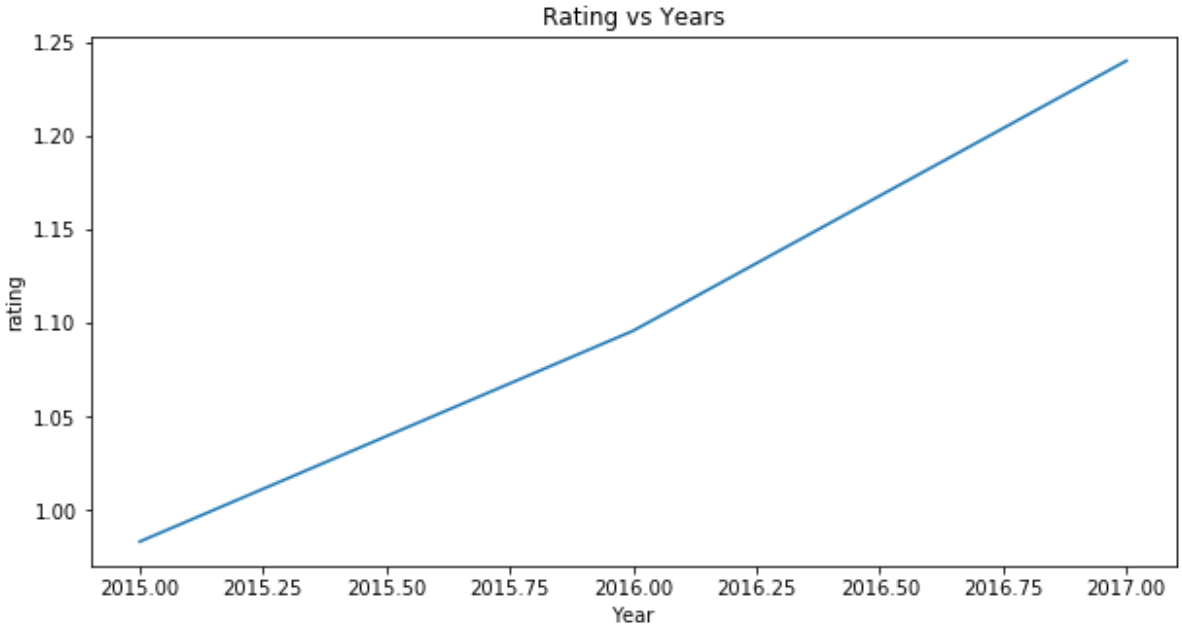


The following plot shows the distribution of the log of favorite count vs retweet count. the plot shows that favorite_count associated with higher ratings than the retweet_count.

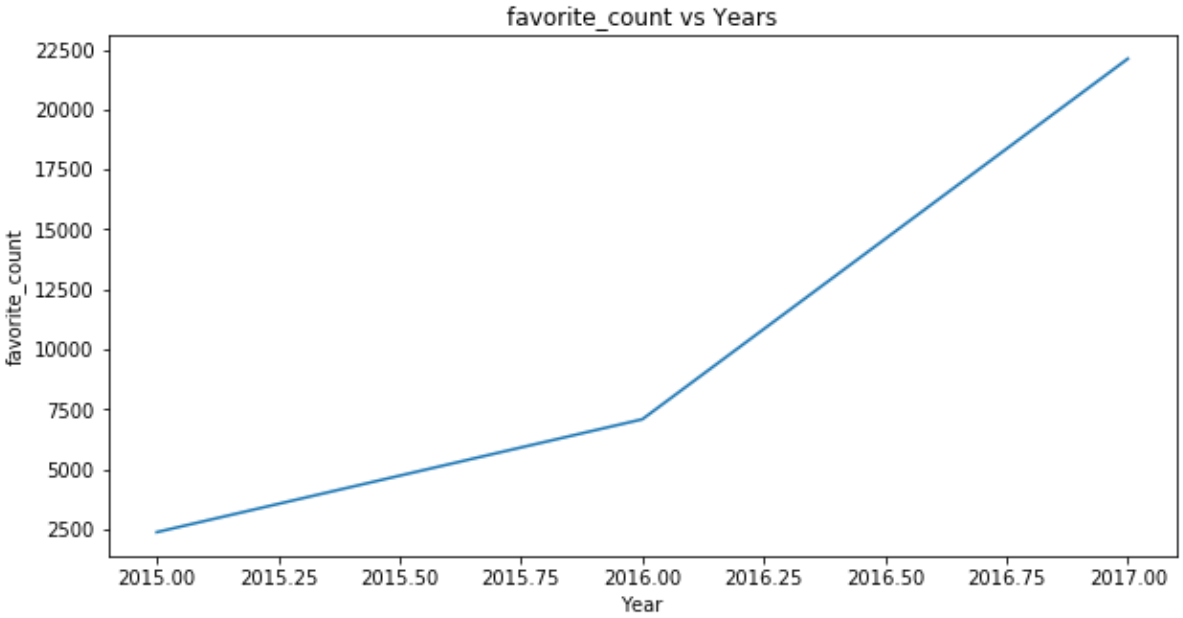


3- How the 'rating', 'favorite_count', 'retweet_count' and 'followers_count' changed over the years?
All of them increased over the years except for followers cunt which shows a slight drop in the number.

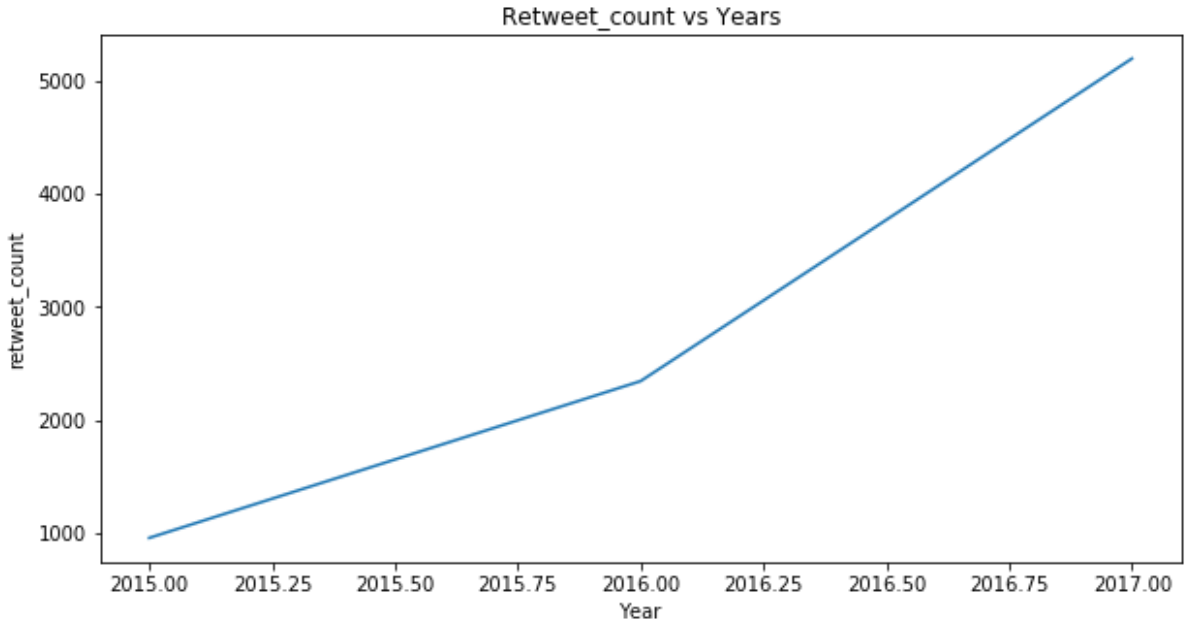
a- Rating: increased over the years. More dogs are receiving higher ratings over the years.



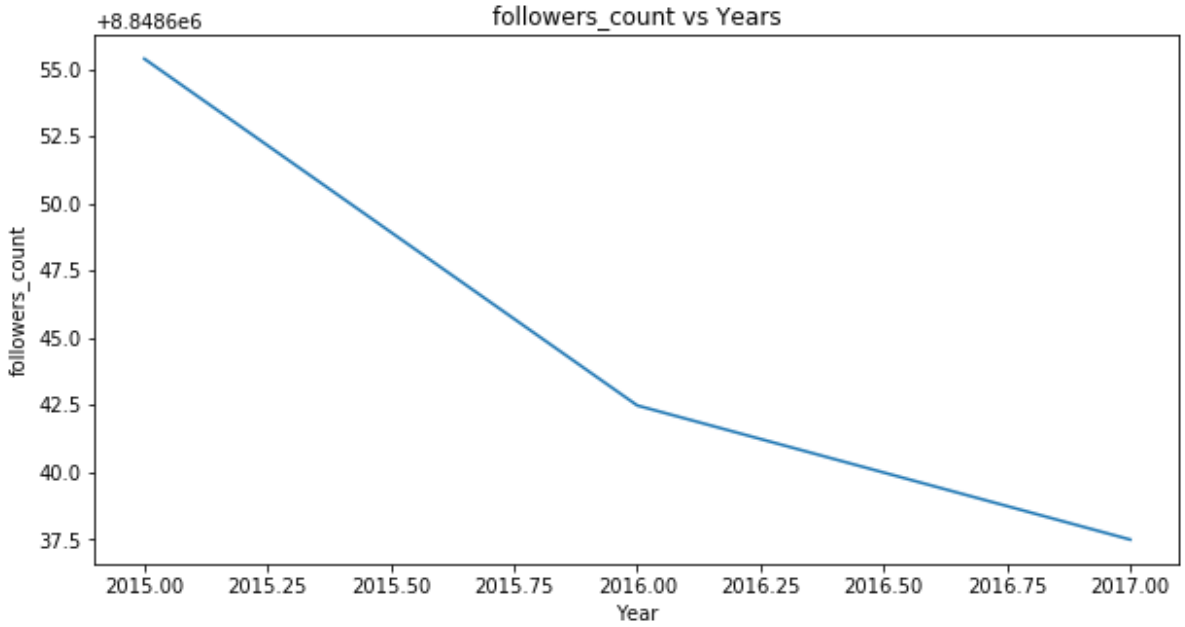
b- Favorite_count: increased over the years:



c- retweet_count also increased over years:

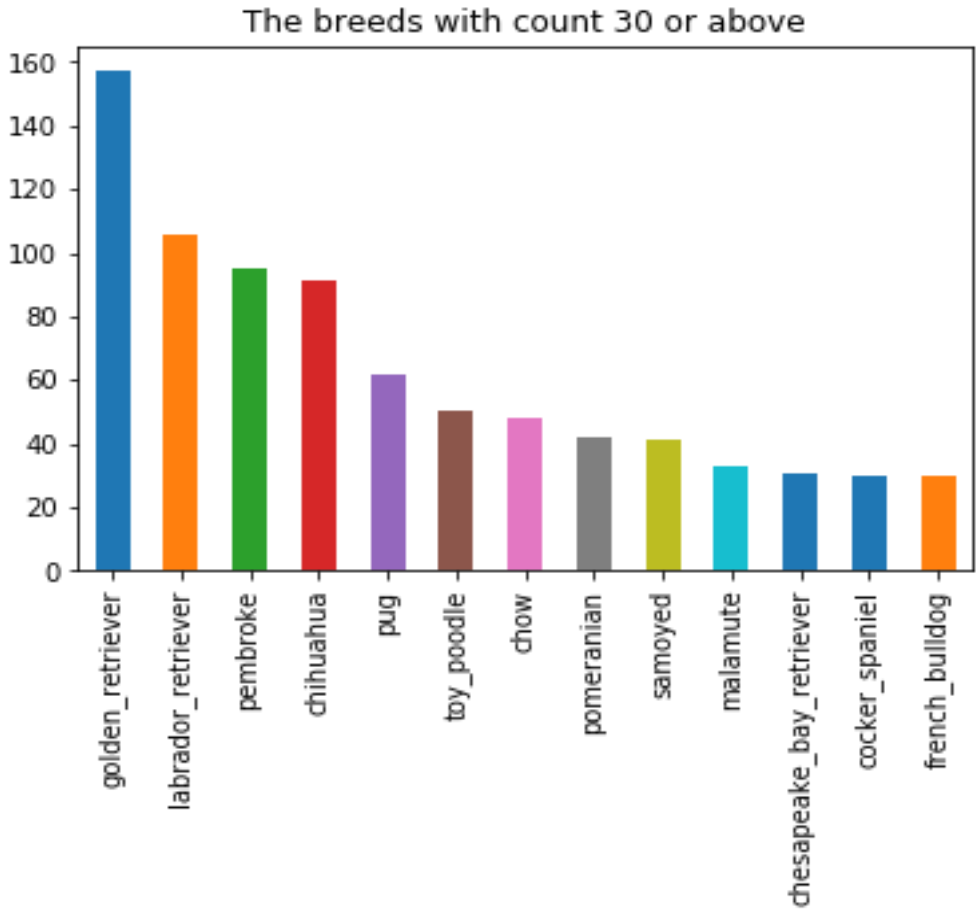


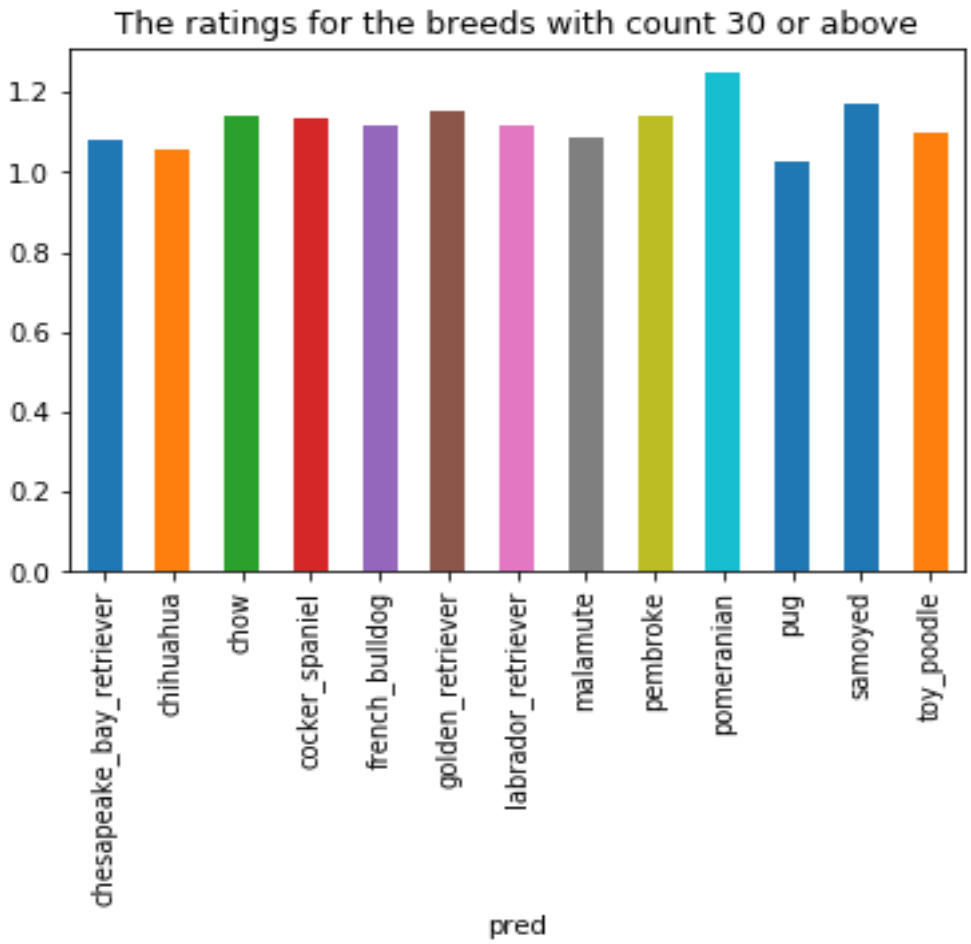
d- followers_count: only this parameter shows decrease over the years from 2015 to 2017, however the decrease is not significant given the large number of followers.



4- Which dogs' breeds are the most common? and how does this relate with the rating?

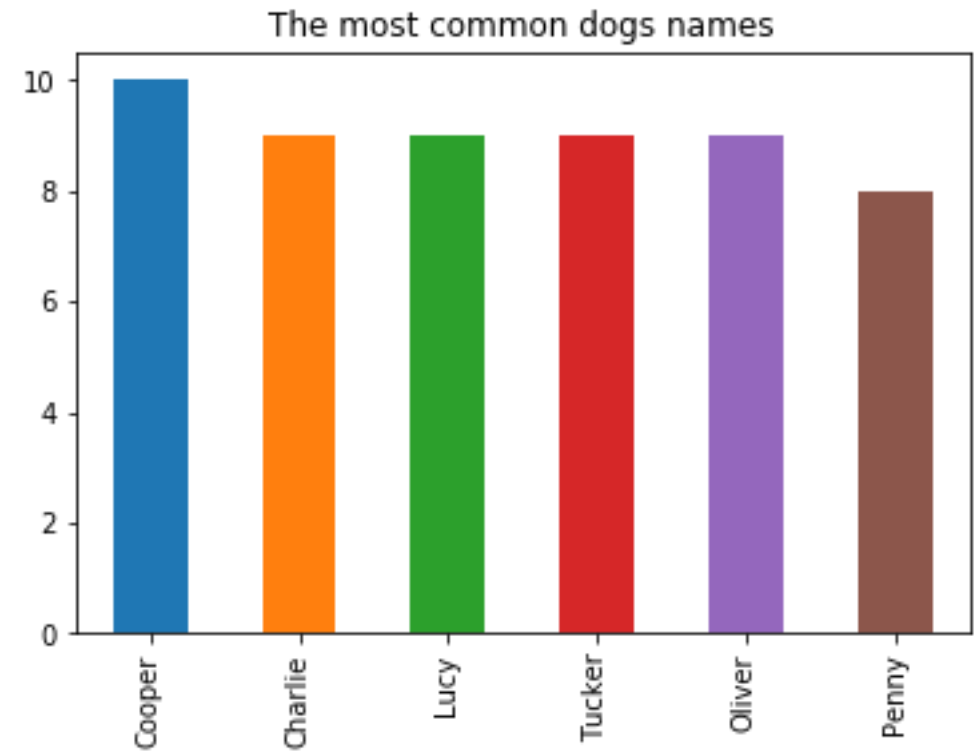
The most common dogs' breed is golden_retriever. The first plot shows the top dogs' breeds with counts higher than 30, and the second plot shows the rating of these breeds. All the common breeds have rating above 1.

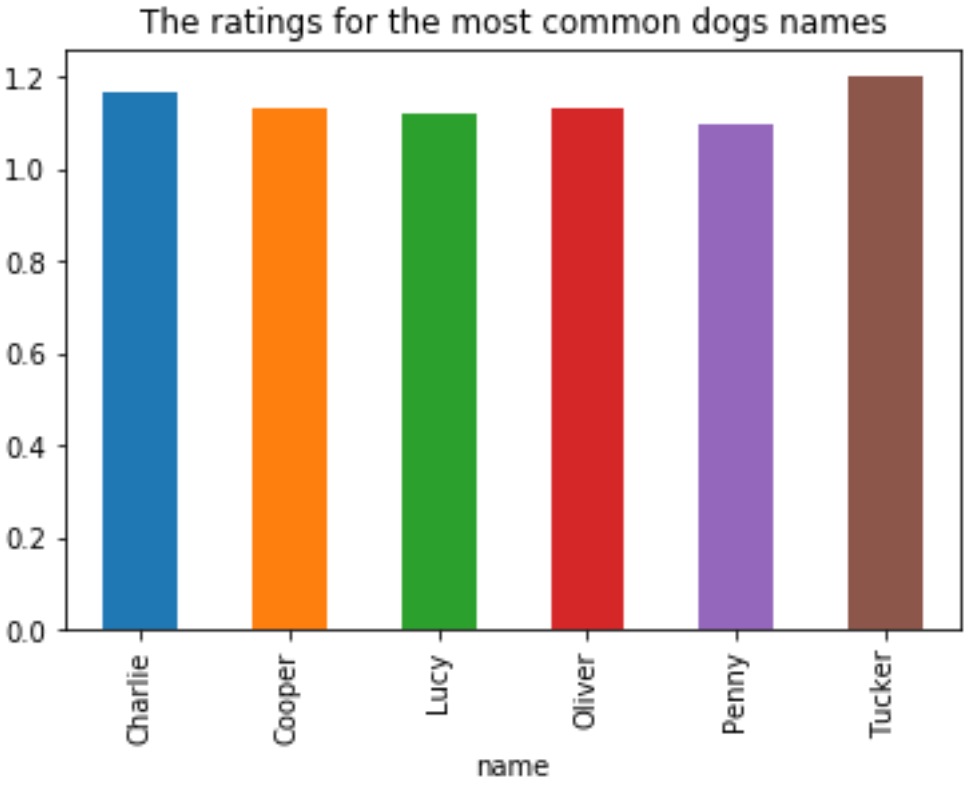




5- The last question is: what is the most common name among dogs? and how this is related with the rating?

The first plot shows the most common names among dogs. You could notice that the counts for names range from 8 to 10 for a population of 1678, which means we couldn't conclude that there is a popular name. However, the highest count goes for the name Cooper. The rating again is above 1 for the most common names.


The most common dogs names
10
8
6
4
2
0
Cooper
Charlie
Lucy
Tucker
Oliver
Penny


The ratings for the most common dogs names
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Charlie
Cooper
Lucy
Oliver
Penny
Tucker
name