Act report

The purpose of the report is to convey my aim of this exploration, what is the best dog in the world. I start with the problems before I reach, then, my insights get through exploration with the photo of the best dog in the world from my point of view.

After my data wrangling activities, there are still many missing values on the columns such as names of dogs, stage of dogs, and breed of dogs. The breed may be possible to fill in by deep learning prediction that is, however, beyond my capability at this stage. Besides breed, there is no way to fill (guess) missing data such as name or dog stage. That means that this data is still far away from 'clean'.

Regarding the analysis of the data such as ranking of dogs in this dataframe, it is very difficult to carry on because of the availability and inconsistency of numerical data. Potential data to be used for the ranking of the dogs are rating, counts of favorites and retweet. But, the distribution of values of ranking, they range between o to 1766(mean=13.2, std=47.6)against their denominators are 10. It is hard to distinguish outliers. It was shown in Figure1(a box plot of the distribution of counts of dog rating).

The similar situation is for the data on both favorites counts and retweet counts as were shown below Figure2.

Having known that it is not reasonable to judge what is the best dog due to the above mentioned reason but , the following is a photo of the 2nd best dog judging from the favorite and retweet counts.

The most popular breed is Golden retriever, followed by Labrador retriever, Pembroke and Chihuahua and so on.

The most popular name is Charlie followed by Lucy, Cooper, Oliver, Tucker, Penny and so on. But, there are a variety of names including funny ones such as Meatball.
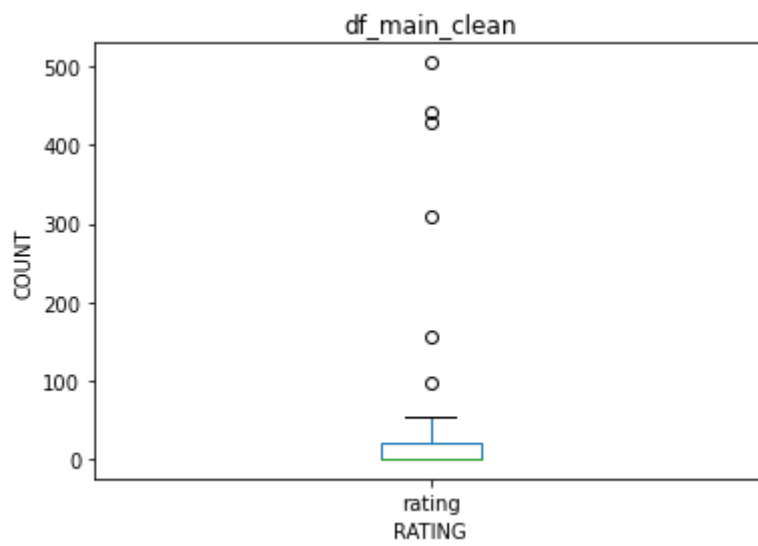
Figure.1

Figure 2.

number of counts of favorites and retweet