

act_report

October 29, 2019

Create a 250-word-minimum written report called `act_report.pdf` or `act_report.html` that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

Communications from the Insights and Visualization of the Dog ratings data

After the data cleaning process, I began with the Statistical summary of my data with heavy focus on the rating which seemed to be inappropriate as some values were above 100%.

From the statistical summary, the values for the rating begs the questions of 'how were the ratings assigned'? The difference between the minimum value of (0) and the maximum values of the rating (177) are overwhelmingly shocking. Probably, the data gathering process has some faults or the neural network algorithm had the wrong hyperparameters which lead to erroneous results. Also Because the mean and median values are far away from the maximum values. (this requires some form of extra investigation).

From a close look into the ratings that were above 200% (this was chosen in order to reduce the complexity of the analysis), it was discovered that 5 out of the 6 dog stages had values of 'None' which Which was a clear indication of data gathering mistakes. Better care should be taken in the future to avoid for bad data which would lead to bad visualization and analysis.

Furthermore, some data visualization were conducted to determine if there were any form of correlation between different data. From the graph of `favorite_counts` vs `retweet_counts`, it shows that the values of `retweet_counts` was directly proportional to the `favorite_counts` i.e. an increase or a decrease in the favorite counts was going to bring about a 'likely' equal increase or decrease in the retweet counts.

Last but not least, another visualization was conducted to see what the distribution of the 'dog stages' looked like. The dog stages signifies the 'maturity level' of the dogs. And from the visualization, it showed that majority of the dogs had a value of 'none' in the dog stages (around 1742 dogs). Only a handful of the dogs had their stages assigned of which pupper was the highest with a count of 211 followed by doggo with a count of 65, then puppo with a count of 23, and floofer with a count of 7.

Overall, in my opinion, more data will be needed to fully understand the dog ratings.