

act_report

September 15, 2022

0.1 Report: act_report

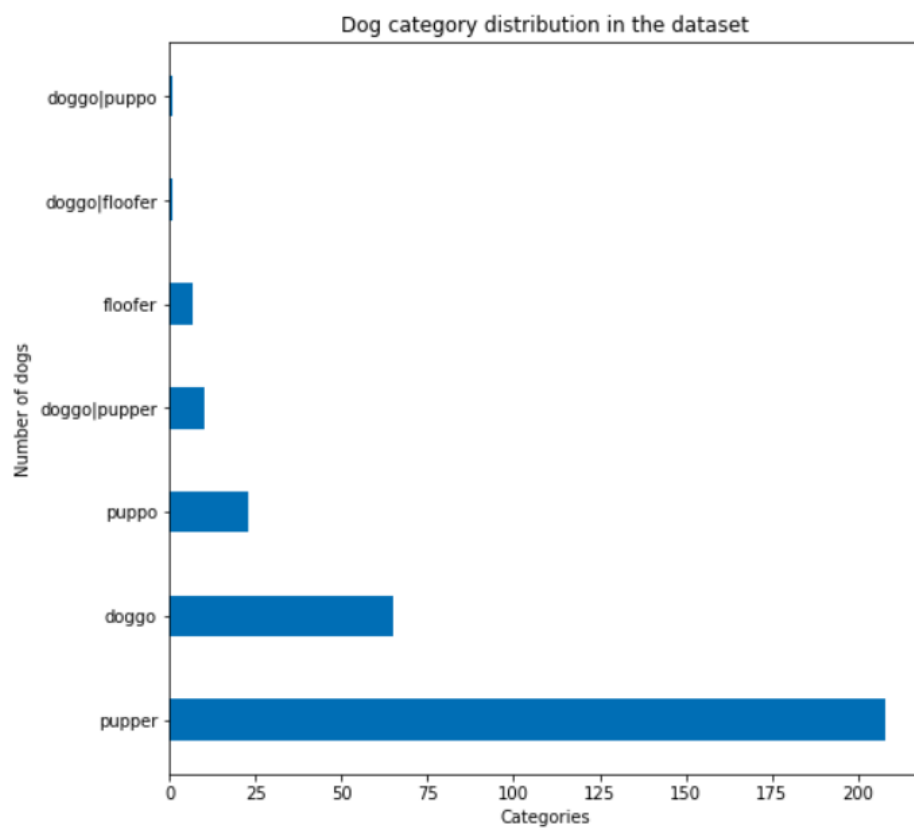
Once our dataset was clean, we moved on to observations and data visualization.

The first observation is about the dog tweets that have the highest ratings. For this I used the column of rate formed based on the numerator column and the denominator columns. Using the `describe` function, I got information about the score series such as its minimum value, its maximum value, its average and the value of its quantiles. I considered that the best rated tweet dogs are those with a rate higher than the 3rd quantile of the series. **Thus, 781 tweets of dogs were noted as the best.**

The second observation is about the dog tweets that people like the most. Here I used the `favorite_count` column. Using the `describe` function, I got information about the `favorite_count` series including its minimum value, its maximum value, its average and the value of its quantiles. I also took the third quantile to consider the most liked tweets. **This resulted in 509 tweets of dogs that people liked the most.**

The third observation is about the most represented category of dog in the dataset. Using the `value_counts` function on the category series, **we notice that the most represented dog category is pupper**. Not only does it far outstrip other categories it also associates with some dogs that have hybrid breeds such as doggo | pupper.

As for the visualization, I decided to visualize this third observation. As the number of categories is not numerous, I decided to use bar charts. So I used the functions of `matplotlib` to realize this visualization.



title