

WeRateDogs Twitter Archive - Act Report

WeRateDogs Data

The WeRateDogs Enhanced Twitter archive contains data extracted from 2356 of the 5000+ tweets from the @dog_rates twitter account, posted between November 15, 2015 and August 1, 2017. This data is comprised of dog ratings that were taken from the text of the tweet along with the dog name and dog stage if present.

The retweet count and favourite count for each tweet were not included in the enhanced archive, and so I had to download this additional data from the twitter account using the tweet ID from the archive.

Along with the Twitter data, I also downloaded an image predictions file from Udacity servers containing the top 3 predictions for dog breeds based on the images from the tweets (there can be up to 4).

Wrangling Data

Before I could begin the analysis, the data had to be wrangled into shape to make it easier. I assessed the data both visually and programmatically for quality and tidiness; the quality of data is determined mainly by looking at several aspects or dimensions to ensure that it is complete, valid, accurate and consistent.

After cleaning many of the issues found during the assessment, there were about 1950 tweets with good quality data.

Insights

the most favorite picture has 132810 favorite count & this it is

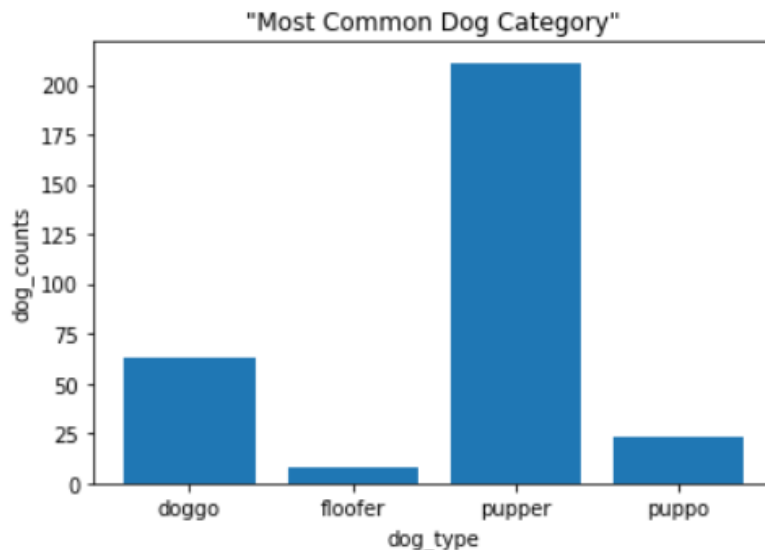


picture has the min retweet count is 16 & this it is



Most Common Dog Category is pupper

There is 4 dog categories ['pupper', 'doggo', 'puppo', 'floofer'] Their count is [211, 63 , 23,8] respectively so 'pupper' is the most common one



Strong Relationship Between Favourite Count And Retweet Count

There is a strong relationship between a tweet's retweet counts and favourite counts; in other words if a tweet has a high favourite count it will also have a high retweet count. Plotting these two values against each other produced the plot below, and clearly shows a high correlation by the very tight collection of points along the angled, straight line.

