ACT Report: WeRateDogs Twitter

An analysis of the @dog\_rates twitter from 2017

Finding patterns in retweet, favorite, and ratings

# Abstract

## Method

For each set of observations and variables a graph was set up and visually analyzed to see if there was any evidence of an impact. If there was sufficient visual evidence, or in cases where no graph could provide such a visual, hypothesis testing was done. When a hypothesis test was run, the null and alternative are provided as the first statement in the section, with detailed analysis following. Please note, the questions, and observations, are ordered here by level of impact, and not in the order that they were tested in the system. With the most impactful being first, and the least impactful being last. Each observation will include the name of the dataframe, where applicable, and any graphs provided.

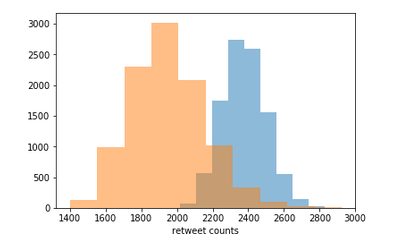
# Analysis

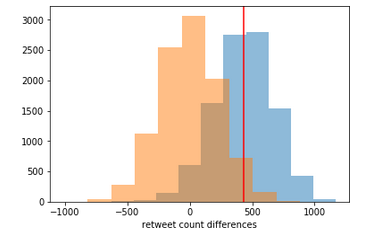
## Does having a name improve retweet?

The null in this case is that not having a name gives at least as many retweets as having a name, while our alternative is That dogs with names get more retweets than dogs without names.

At this point, I had to return to the wrangling part of the analysis. I noticed that there were dogs named ‘a’ , ‘an’ …etc. in the list, and I needed to verify that these dogs have names or don’t and properly handle the columns. While dogs do get, from some owners, rather absurd names, such as ‘Hey You’… a simple conjunction is unlikely.

I started by creating dfa2, consisting of tweet\_id and name from df1\_c and retweet\_count from dfa3\_c. then altering the names column from strings to 0/1 int. This allowed me to run a more meaningful analysis on the subject. I then calculated the difference in means for retweet counts for these two categories. These had a mean difference of 431 retweets. I then generated a series of 10000 choices from the dfa2 dataframe in order to generate a significant sample of test values, and generated the standard deviations for named vs. unnamed dogs. In addition I generated a sample of the differences and the standard deviation for the differences. Plotting the distributions seemed to show two normal distributions. I then proceeded to calculate a p-value for this set of data. The p-value of indicates a strong influence in retweet counts just based on the dog having a name attached to it.

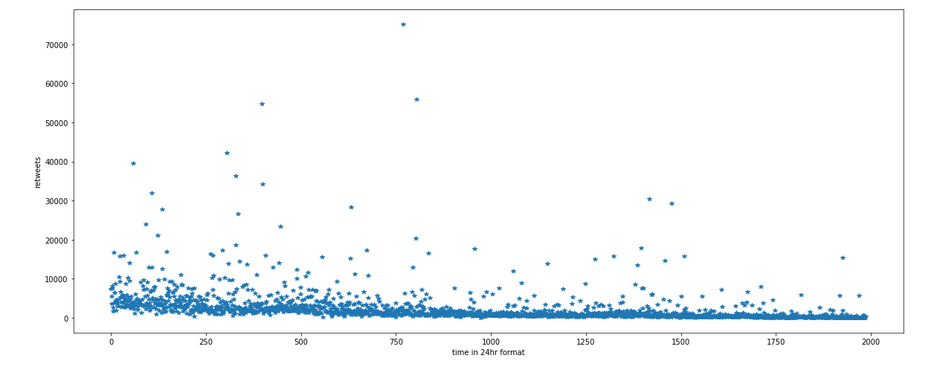




The red line represents the actual tested data difference. In the second graph the null hypothesis in presented in orange, and the actual tested data in blue. While it is visually clear there is a difference, good practice would be to mathematically prove it, as we did above and showed that there is much less than even a 0.001% chance that this metric is not having an impact, for this reason we reject the null hypothesis and state that it is highly likely that simply naming a dog in a tweet improves the probability of it being retweeted.

## Does tweet time improve retweet?

This was pretty simple to answer, as the scatterplot below shows there is really no effect of tweet time on the number of retweets provided. This is however perhaps a little surprising as well, as the retweets are very narrowly spread out over the day. The scatterplot was created with dfa1, and is provided below.



## Does type of dog influence retweet?

The null in this case is that the type of dog, using our predictive data, had an impact on retweets

## Does rating impact retweets?

Though ratings are arbitrary, perhaps they are having an impact on the retweet rate

## Does tweet time improve favorites?

As with retweets, we will compare favorites to timestamp

## Does having a name improve favorite?

This is a similar comparison to the above for retweets

## Does type of dog influence favorites?

Again as above for retweets

## Does rating alter favorites?

Again as above for retweets

## Does retweet impact favorites?

Compare retweet to favorites

# Insights

It is clear from the data provided that a dog photo with a name attached to it is more likely to be retweeted than a dog without a name attached to it.

### External Resources:

https://stackoverflow.com/questions/55092403/how-to-extract-hourminute-from-a-datetime-stamp-in-python