

# Action Report\_Project: Wrangle and Analyze Data

By. Shawn Gong

#Udacity/DAND/Term2

#Actionreport

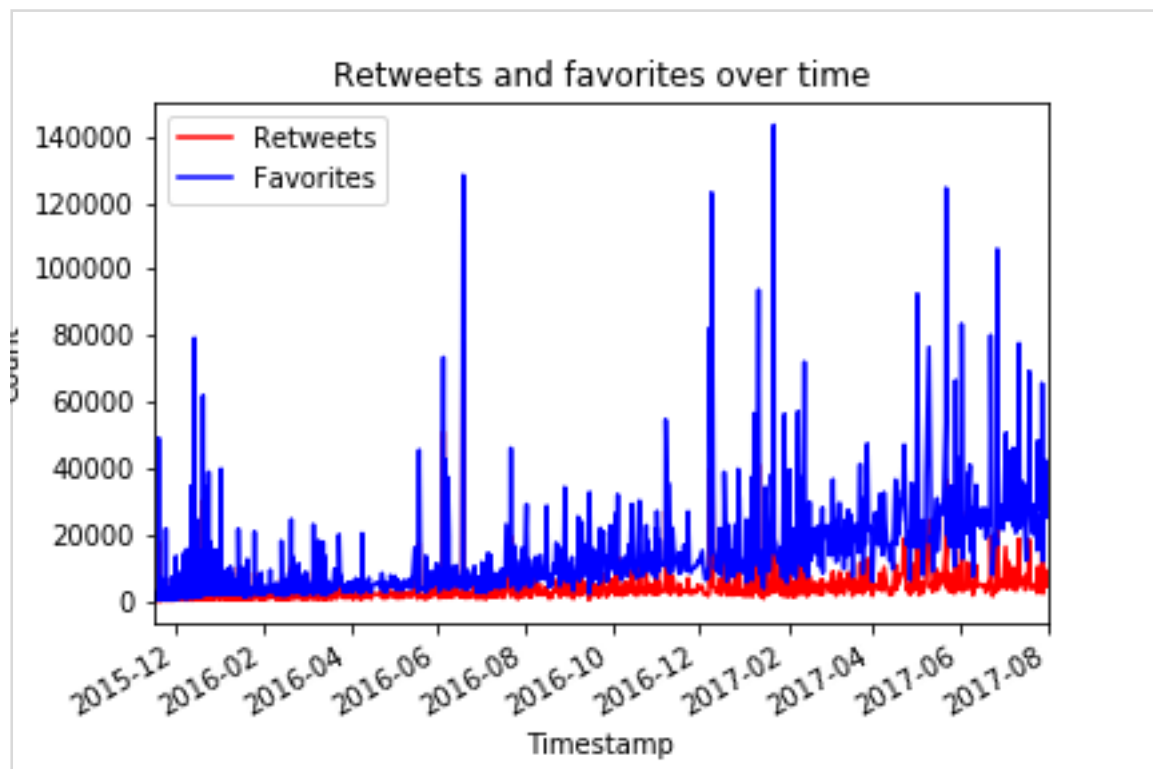
#Python

## Introduction

This report is of the brief analysis after wrangling the @WeRateDogs Twitter data. The analysis will focus on three variables: retweets, favourites and rating percentage.

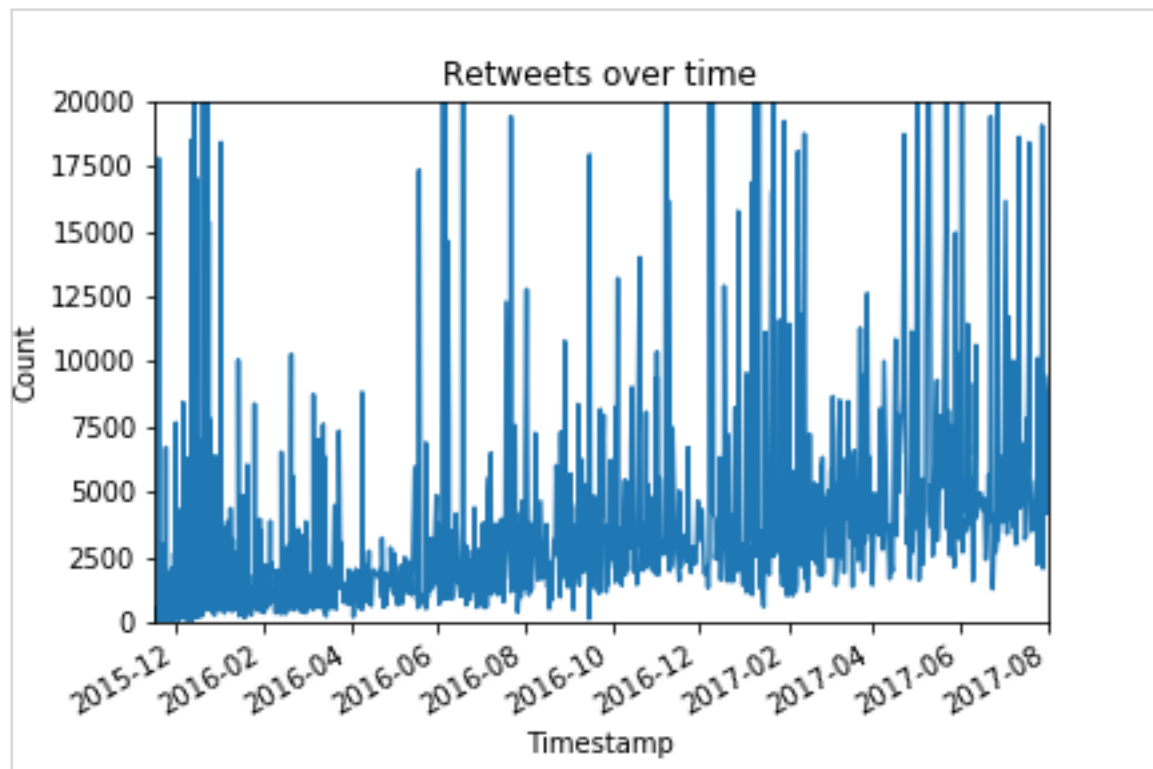
## Visualizations

### Retweets and favourites over time



As we can see from the plot, the retweets and favorites of the tweet are increasing over time, the data seems variant, but the bottom of distribution has a clear trend of increasing, which means the popularity of @WeRateDogs has been constantly increasing from the beginning.

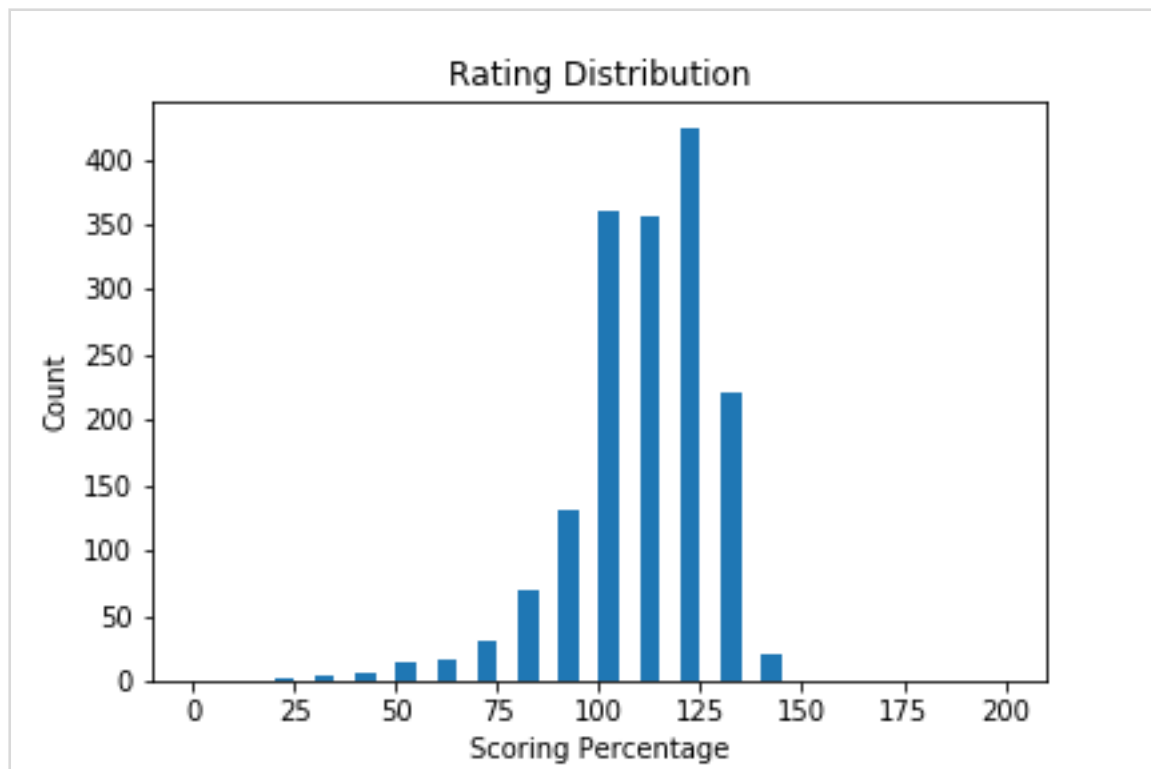
You might have noticed that the retweet count trend is not as clear as the favorite due to the gap of numbers, so we can zoom in to get a better sense of it:



I limited the y axis to 20000 maximum, and we got this plot. Although there's a bit fluctuation along the way, it clearly shows the increasing popularity trend of favorites is applicable to retweets as well.

### Rating distribution

It will be interesting to see the rating distribution. Because as David documented in the project motivation, @WeRateDogs focuses on giving funny comments for cute dogs and the rating scale is rather casual. The typical rating is over the 10 rating scale. Additionally, during our exploration, we found there are rating scales that are larger than 10.

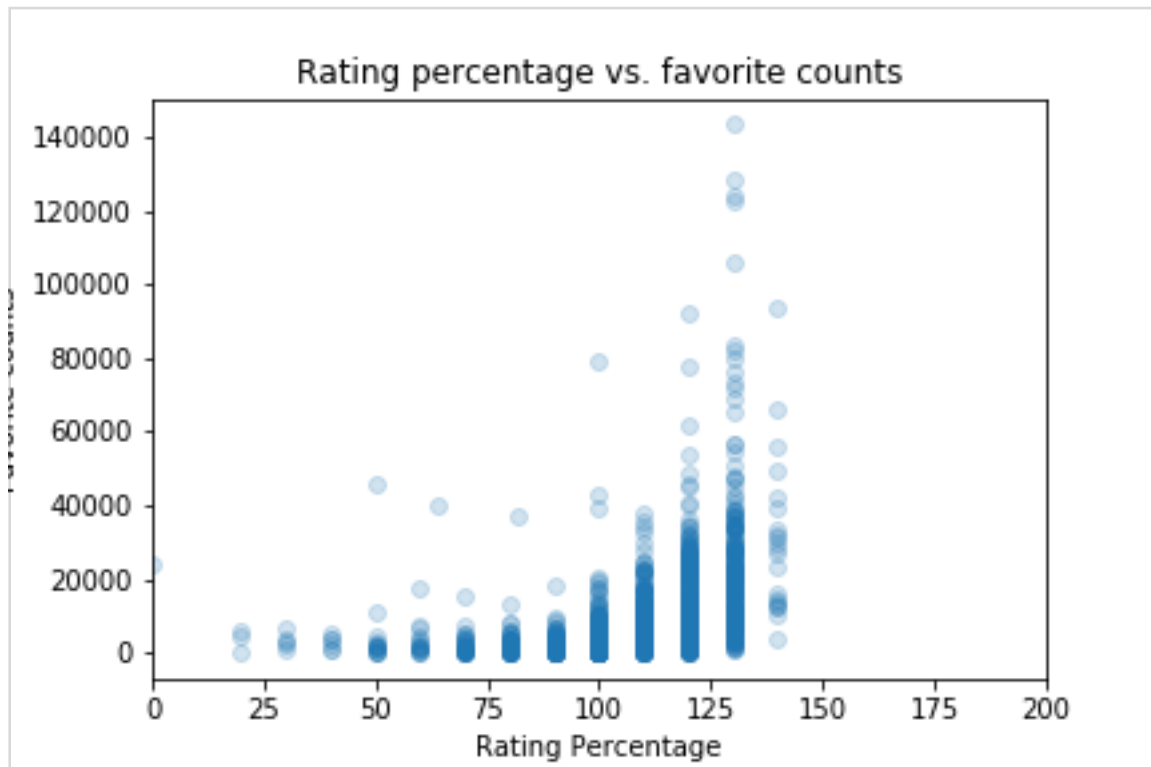


Interestingly, The rating percentage follows a left skewed distribution as we can see on the histogram. To get a better sense of the distribution, here's some descriptive statistics about these three variables:

	retweet_count	favorite_count	rating_percentage
count	1666.000000	1666.000000	1666.000000
mean	2782.738896	9232.478391	108.786502
std	4859.682151	13034.745627	25.154334
min	13.000000	80.000000	0.000000
25%	628.250000	2122.500000	100.000000
50%	1399.500000	4364.500000	110.000000
75%	3172.000000	11503.500000	120.000000
max	77362.000000	143411.000000	750.000000

The mean and the median of the rating percentage are both greater than 100%, we can also see that 75% of the observations have rating more than 100%. This fact reflects the rating logic of @WeRateDogs.

## Rating and favorite counts



This scatter plot reveals even more interesting facts about this dataset. There's big differences in favorite counts resulting from the rating percentage. Despite 75% of observations are greater than 100%, the dogs rated at 130% are most likely to get the highest favorite count than any other group. As we can see in the histogram in the previous section, the amount rating at 100% and 110% is quite similar to the dogs rating at 125%, but there can also be such a big difference in favorites. I don't know if there is a rule of "13/10", but this coincidence might reveal a "hidden fact" that both the followers and account owners have never seen before: The dogs that are rated 13/10 are often the most liked on @WeRateDogs.