



# Analysis and Insights into Final Data

16.03.2020

—BY

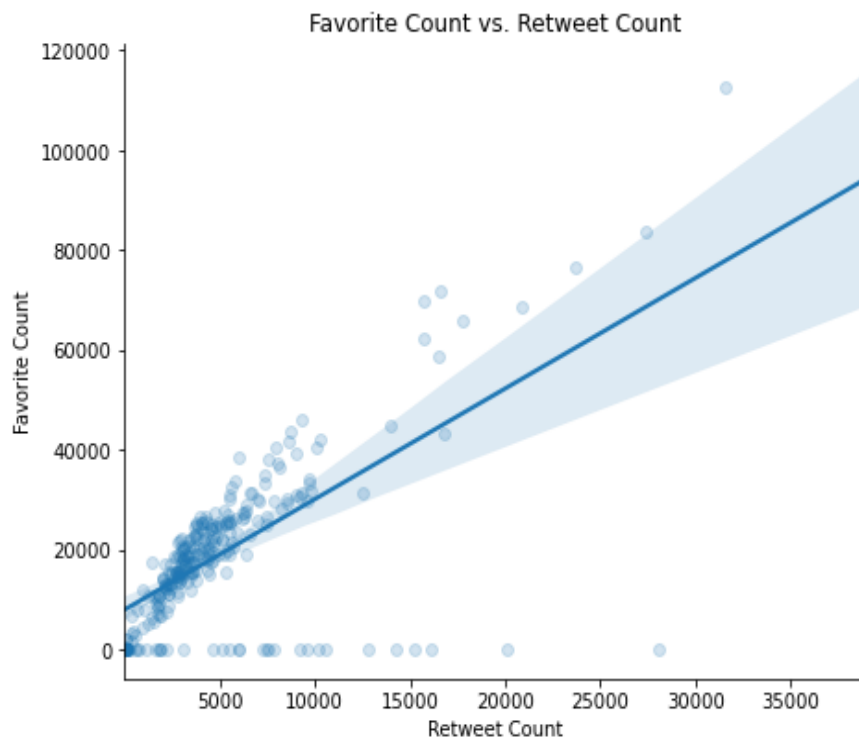
OKEBIORUN SARAH

## INTRODUCTION:

Real-world data sometimes doesn't come clean. The dataset that I would be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog\_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. This project works through the data wrangling process, focusing on the gathering, assessing, and cleaning of data. There are visualization and observation from the analysis provided as well.

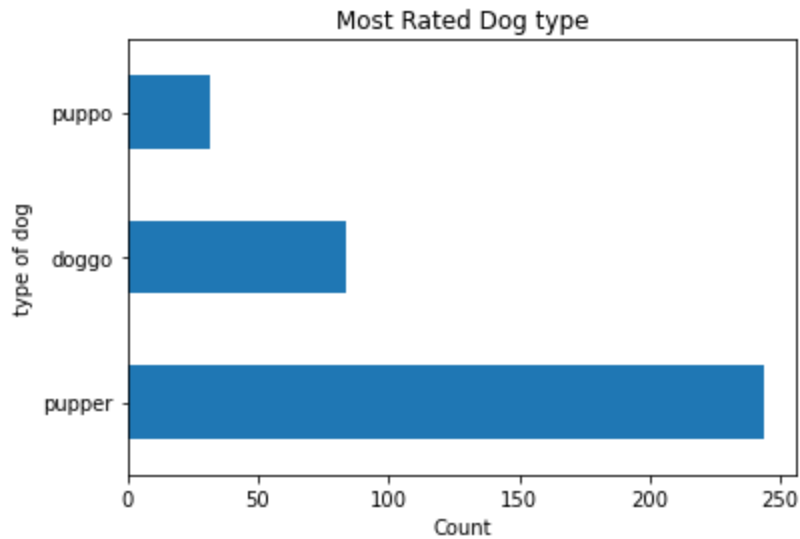
### Visualizing the retweet counts, and favorite counts comparison over time

There is a positive correlation between favorite ("like") counts, and how much a post was retweeted. This correlation is important for the owner of the WeRateDogs Twitter account to understand when determining the method to increase users' traffic on the page. A data analysis team could recommend previous posts with either a high retweet count or high favorite count so that page owners could model future posts off historically popular posts.



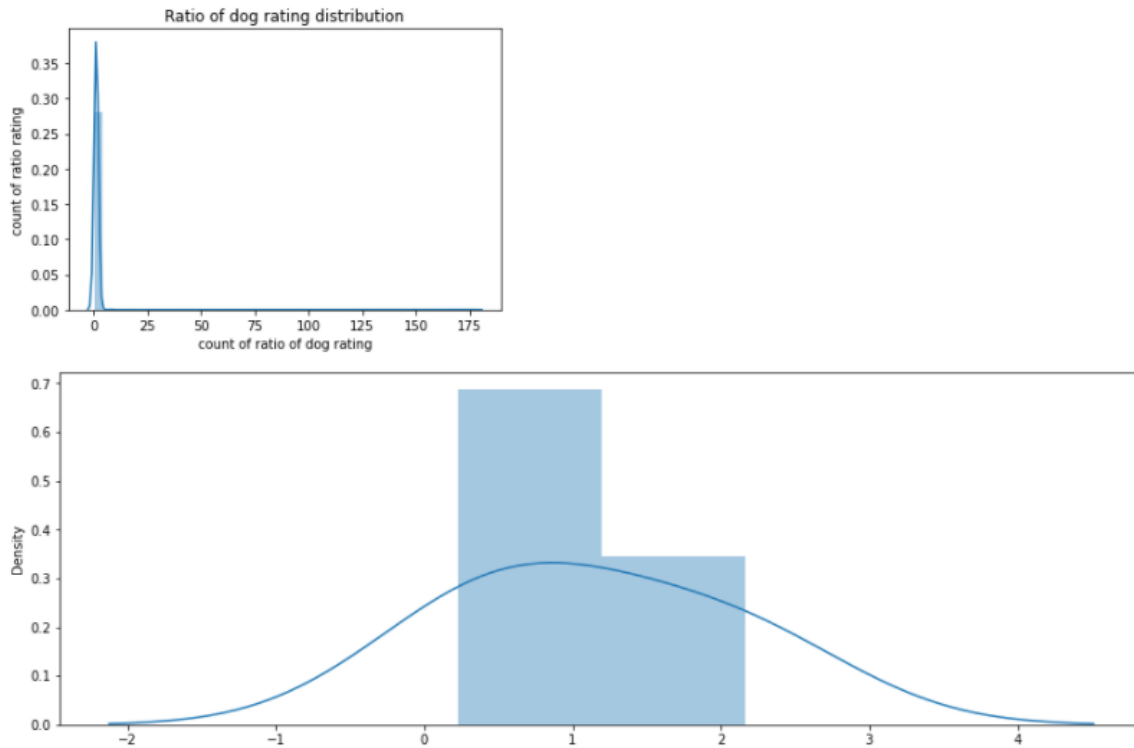
## Popular Dog type

The most popular dog type is pupper, with doggo coming in as the second most popular breed. Puppo isn't far behind. And lastly, floofer is the least dog type. The page owner could use this information to create targeted marketing efforts for a certain dog type that isn't popular to increase their popularity but also utilize the type that is proven to be popular to drive user traffic to the page.



## The ratio of dog rating distribution

Hereafter we calculated the ratio of dog rating from (rating numerator / rating denominator) rate, we can see that the chart takes normally distribution of dog rating.



## CONCLUSION:

I have encountered many obstacles in understanding these data so that I can analyze them, though, this report offers a straight look at the data wrangling process. There is so much more that can be done with this data set