

INSIGHTS FROM WRANGLING WeRateDogs TWITTER DATA

JANE HARRY

# Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc

I used three data sets for this project:

* An enhanced twitter archive data from WeRateDogs
* A dog breed image prediction file, and
* Additional data gotten from twitter API

## INSIGHTS

1. The number of retweets for a rating increases with increasing likes
2. The least favorite dogs are dogs breed which are not accurately predicted by the neural network.

prediction1

desktop\_computer 99.000000

crash\_helmet 181.000000

hay 213.000000

trombone 247.000000

piggy\_bank 284.000000

platypus 288.000000

pitcher 313.000000

lorikeet 363.000000

coho 371.000000

fire\_engine 372.000000

cuirass 376.000000

ibex 423.000000

three-toed\_sloth 425.000000

bearskin 452.000000

ping-pong\_ball 484.000000

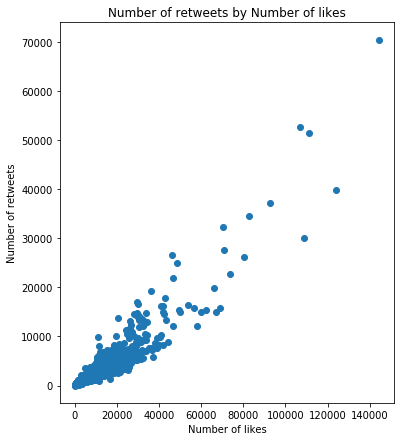
sea\_urchin 501.000000

pot 527.000000

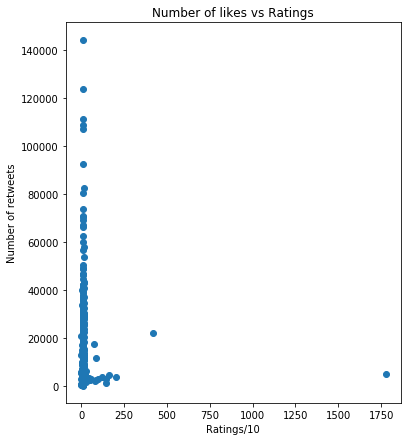
These are the least favorite Dog breeds, based on the number of likes

1. Dogs with high numerator ratings are typically the favorite of audiences, from the number of likes

## visualizations



A Positive correlation is seen between the number of retweets, and number of likes



The above graph shows the relationship between number of likes a rating tweet got, and the number of rating. The rating system, is a unique one, with numerator values typically over 10, and a denominator of 10.

## OBsERVATION AND CONCLusion

* The numerator rating of the dataset contained so many outliers that resulted in a skewed graph.
* Further cleaning needs to be done on the dataset to get valid predictions.