

Analyzing and Visualizing WeRateDogs

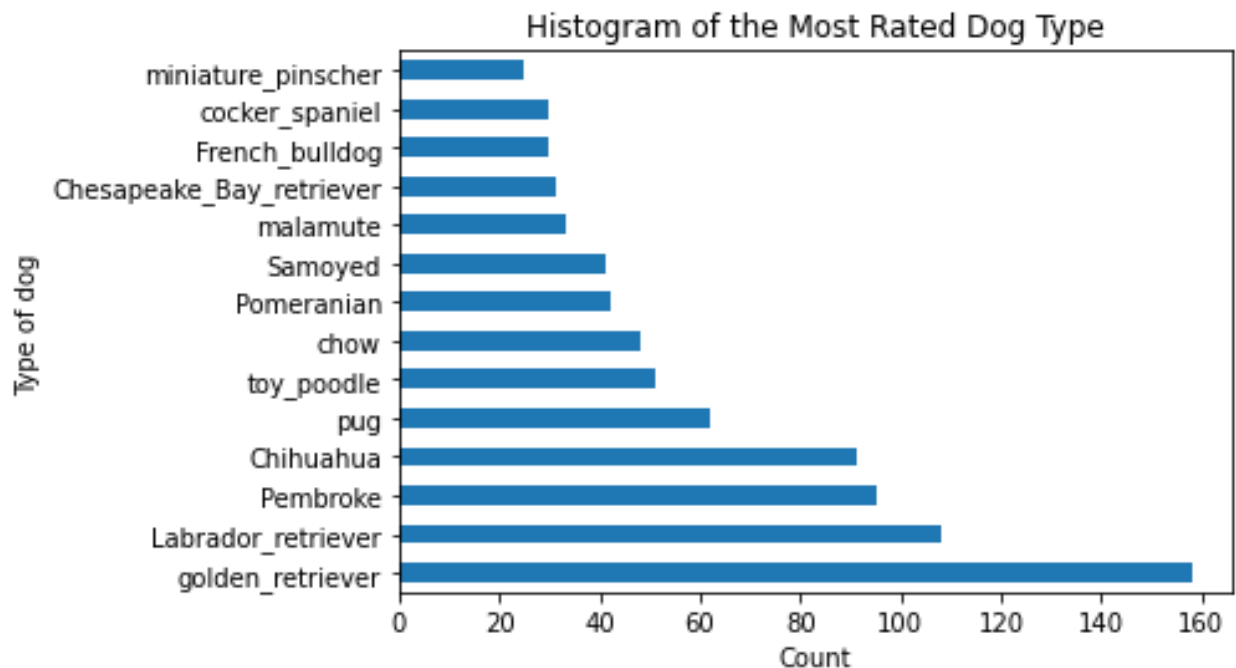
Introduction

This Project is based on Twitter account for @WeRateDogs, they seem to do a humorous wording to describe all rating and used unusual rating method where Numerator are mostly above 10 i.e. denominator. This Udacity project require to Gather data from three different sources, assess, clean and finally some visualization.

Initially upon gathering we had 2356 observation but after cleaning 1928 observations but doing this project I really understand why data wrangling is so important as lot of observation are repetitive, inappropriate, incomplete and messy but as project requirement only to clean up 8 quality issues and 2 tidiness issues so I can say I have complete somewhat but it's not perfectly cleaned.

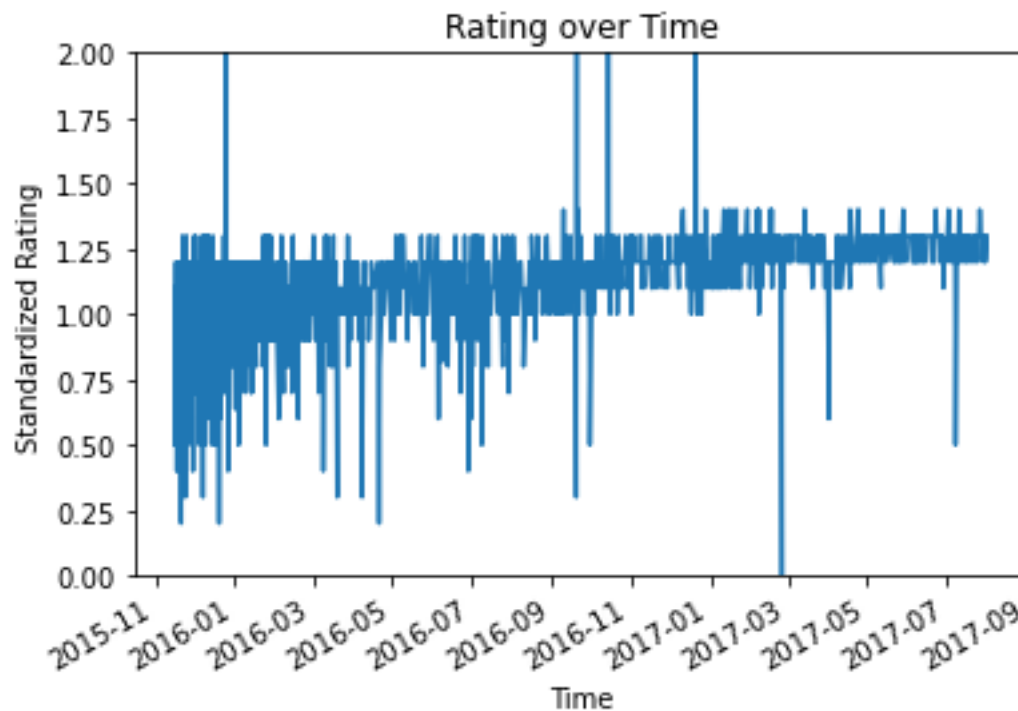
Most common dog type

WeRateDogs has over 6000+ tweets. I was able to analyzed around 1500+ tweets. The most rated dog was golden retriever with more than 140 ratings.



Standardized Rating over Time

The idea behind the WeRateDogs account is that they ask people to send them photos of their dogs, and they will rate them on a scale of 1-10 with humorous comments; however, they are often given ratings higher than 10. I assumed that almost all the dogs were given a rating higher than 10/10 but I was surprised to notice many with ratings lower than 10/10. In addition, many ratings did not have a denominator of 10. Therefore, to standardize the ratings I calculated a value of numerator divided by denominator. I was most curious to see if overtime, as the account became more popular and people associated the above 10/10 ratings with being funny, that the higher ratings would become more prevalent. Indeed, as shown in Figure 2, it appears that overtime the frequency of ratings below 1 decreases. Before 2016-11 there are many ratings below 1, while after that time there are barely any. The maximum standardized rating is about 1.3 except for three outliers including the joke ratings 1776/10 and 420/10 and 1 error that did not end up getting cleaned.



Retweet and Favorite Counts

