Act Report: Wrangling and analyzing data

Introduction

WeRateDogs is an official Twitter account that primarily rates people's dogs with humorous comments. The rating almost always has a denominator of 10 and the numerator is always greater than 10 and this is because they are good dogs. My task in this project was to gather, assess and clean data as well as analyze and create visualizations. Using the tweet archive and data I extracted

Investigation Overview

While analyzing the data, I was interested in answering the following questions:

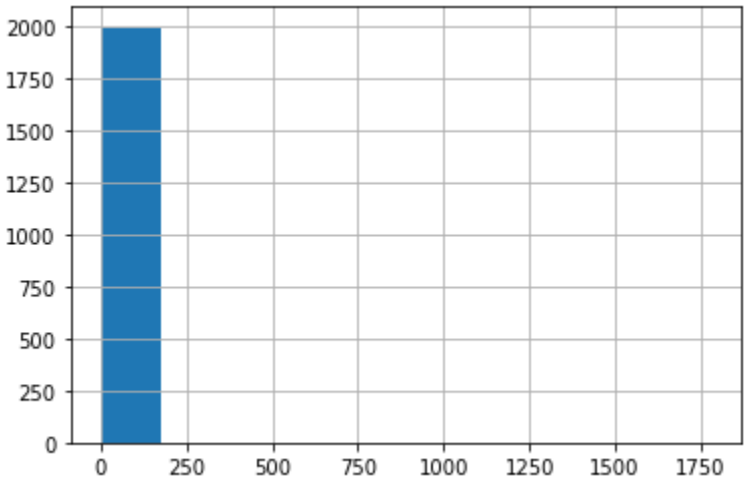
What was the average rating for the dogs?

Which image number was highly predicted?

Was there any correction between features?

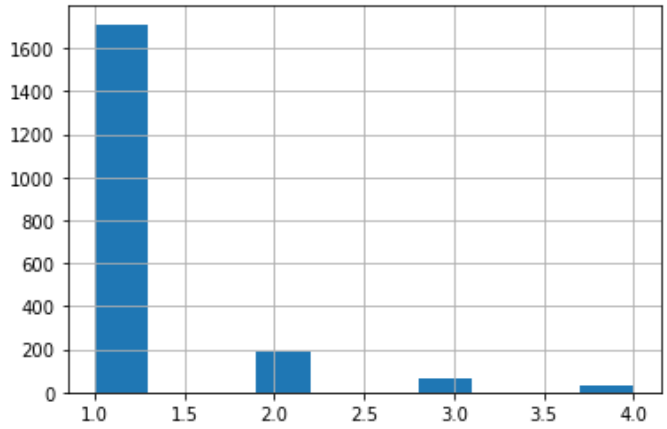
What is the rating numerator of the dogs?

The analysis shows that the majority of dogs received a rating from 10 to 50 and not more than 100



Which image number was highly predicted?

Also, I wanted to find out the most predicted image number and I found that number 1 was highly predicted.



Any correction?

I did not find any relation between features (at least the features I used)

	tweet_id	rating_numerator	rating_denominator	image_number
tweet_id	1.000000	0.023630	-0.022426	0.212854
rating_numerator	0.023630	1.000000	0.198240	-0.000684
rating_denominator	-0.022426	0.198240	1.000000	-0.004094
image_number	0.212854	-0.000684	-0.004094	1.000000