Analysis and Insights

Udacity Data Analyst Nanodegree, Term 2
Created by: Matthias Rieke

Introduction

This project is part of the nanodgree "Data Analyst". This project is about wrangling, assessing, cleaning and analyzing data associated with the twitter user WeRateDogs (@dog_rates).

The twitter user WeRateDogs rates pictures of dogs in a very entertaining way. What's special about the rating is, that WeRateDogs ratings almost always have a numerator greater 10, while the denominator is 10. The argument for the numerator being greater than 10 and the denominator being 10 is, that all of the dogs are good dogs, therefore deserving at least a ten, but in most cases deserve better.

Below, you can find the insights and visualizations, which have been taken from the cleaned dataset.

1. What is the most rated dog stage and what is their mean standardized rating?

Taking a look at the data, we can see that pupper is the most rated dog stage, while floofer being the less rated one. But taking a look at the mean standardized rating for dog stages, we can see that pupper have the lowest mean, while floofer have the highest.
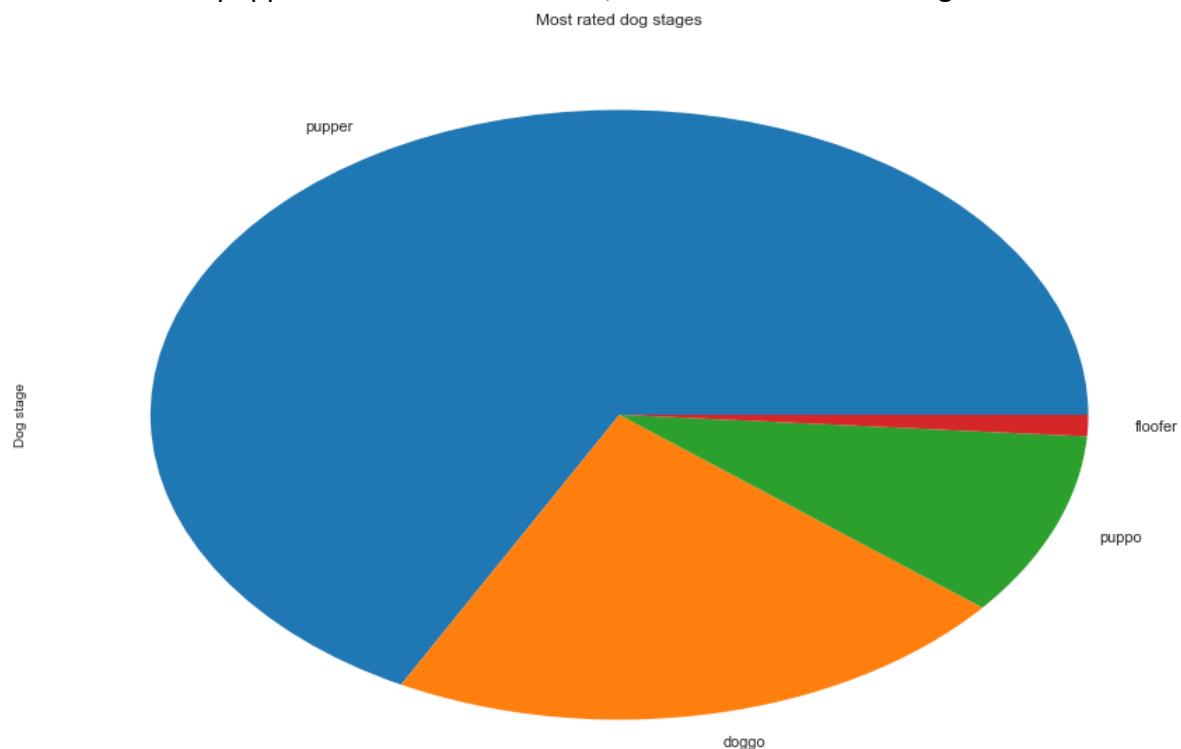

Figure 1: Number of ratings for each dog stage

Mean standardized rating for dog stages:
dog_stage
pupper   1.080592
doggo   1.167241

puppo   1.211538
floofer   1.266667

This makes sense, as the floofer only has few ratings, while pupper has many, therefore the average is more "smoothed" than the floofer one.

2. How did the standardized ratings developed over time?
The rating of WeRateDogs doesn't follow the traditional rating of numerators from 1-10 and the denominator of 10. The idea of WeRateDogs is, that every dog deserves at least a 10 and therefore the numerators start at 10 and go up.
At the beginning, WeRateDogs had quite some standardized ratings which were lower than 1. Over time, the ratings below 1 are decreasing. I suppose that this kind of rating had to become popular and known, therefore being more and more used over time and the "false" rating become less used over time as well.
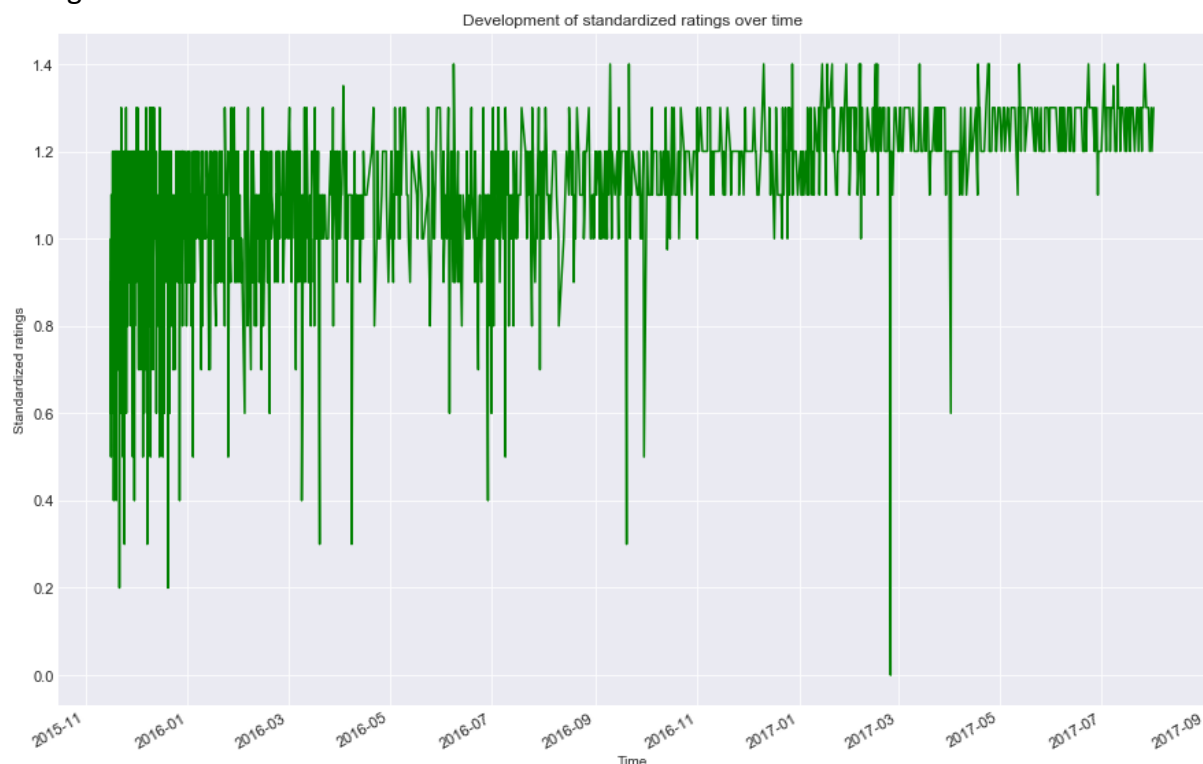

Figure: Development of standardized ratings over time

3. What are the Top20 dog types in terms of rating numbers?

Taking a look at the distribution below, we can see that golden retrievers and Labrador retrievers are the two dog types which have been rated the most. In fact, both of them belong to the breed "retriever" making the retriever breed the most rated dog type with a big gap to the third one, the Pembroke.
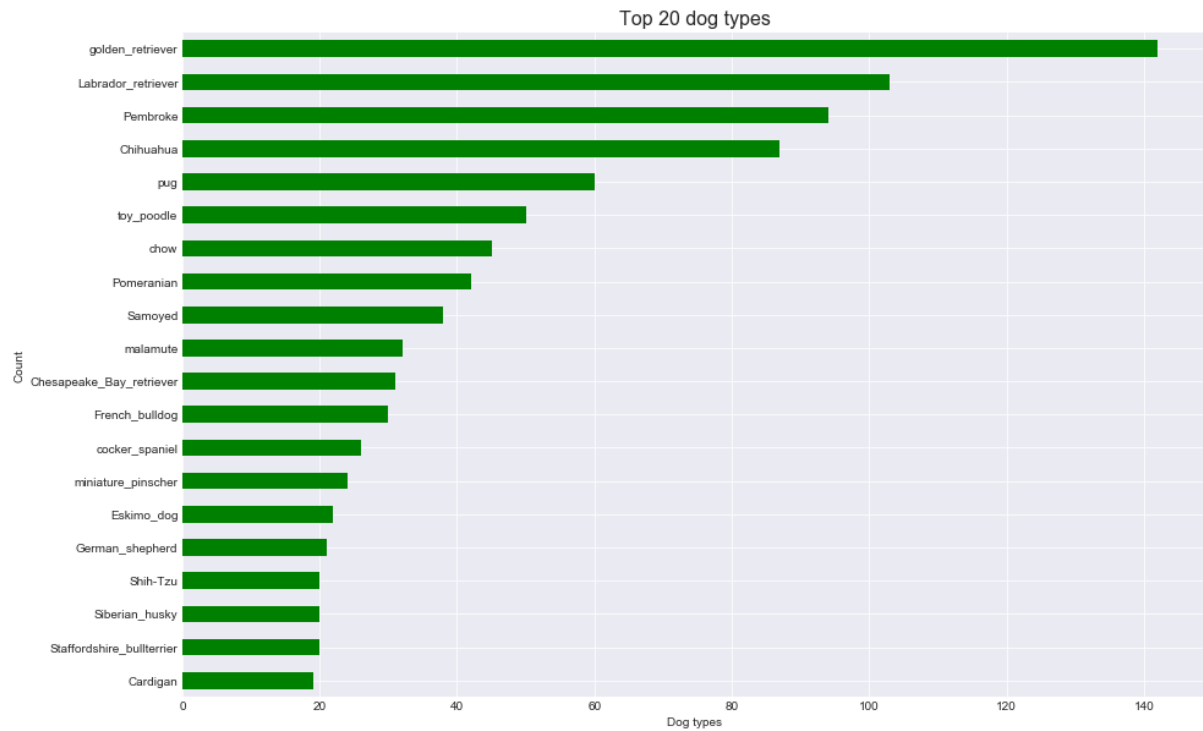

Figure: Top20 dogs in terms of number of ratings