Data Analyst Nanodegree

Wrangle and Analyze Data : WeRateDogs act report

Manon Philippot

March 17, 2019

Contents

1 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. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

Multiple datasets have been gathered, assessed, cleaned and merged together in the first part of this project in order to obtain a master dataset offering basic tweet data for all 5000+ of their tweets, such as the tweet's text, the rating, the dog name, the dog "stage, the prediction of the breed dog and additional data such as retweet count, favorite count and number of characters of the tweet.
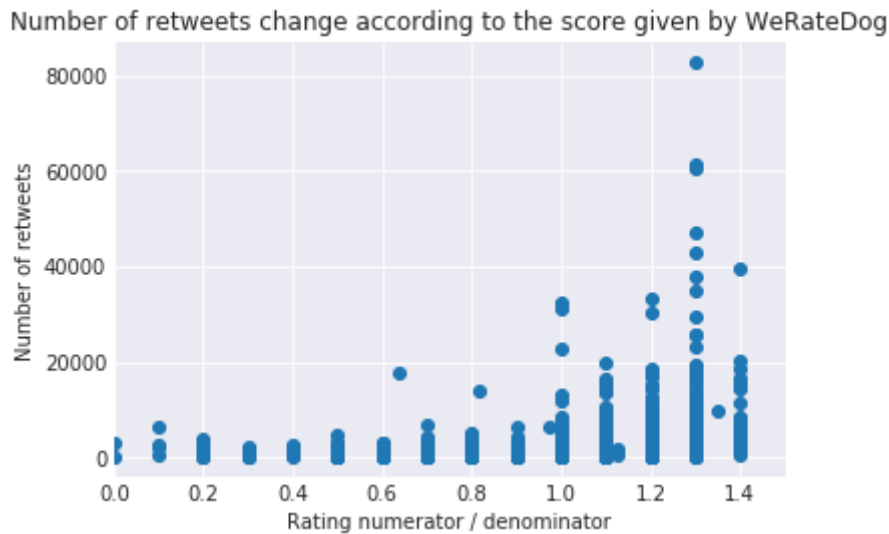
Let's take a look at what we discovered about WeRateDog with this dataset.

2 Insight

2.1 Insight 1 : number of retweets and score given by WeRateDog

How does the number of retweets change according to the score given by WeRateDog?

A scatter plot has been made, with the rating "numerator / denominator" on the x axis and the number of retweets on the y axis.

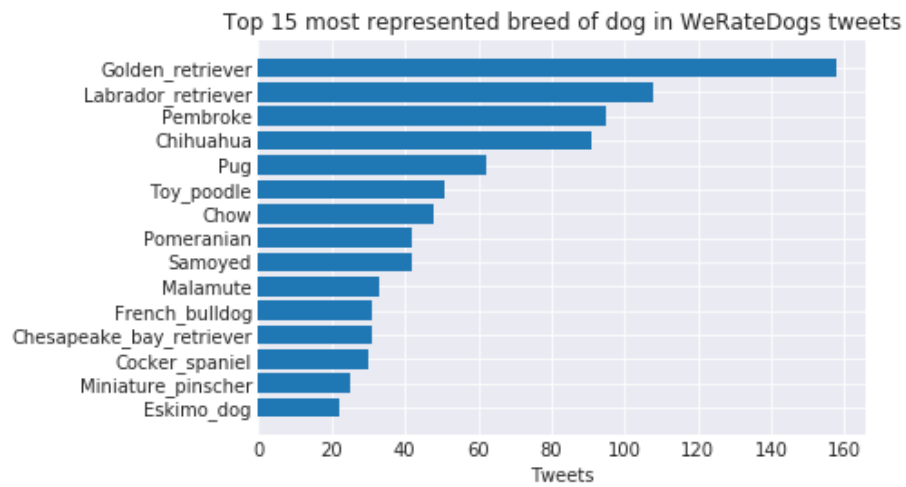Number of retweets change according to the score given by WeRateDog

The scatter plot seems to be left skewed. It seems that the better a dog picture is rated, the more likely it is to have a large number of retweets. However, this cannot be said with certainty. Maybe the disproportionate ratings (like 1776/10) win retweets also because of the funny aspect of the ratings ?

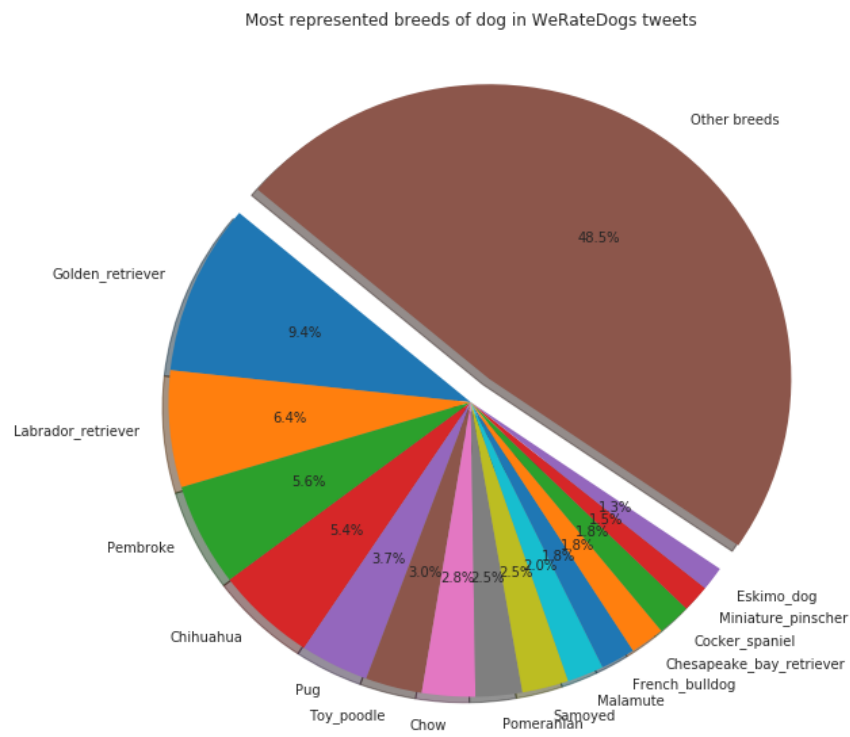## 2.2 Insight 2 : most common dog breeds

What are the most common dog breeds represented in WeRateDogs tweets?

We analysed a set of 1994 tweets and discovered that among all these tweets, there are 113 breeds of dog represented.

Among all these dog breeds, we found the top 15 of these dog breeds and we made a bar chart .

Top 15 most represented breed of dog in WeRateDogs tweets

The most represented dog breeds in WeRateDogs tweets are Golden retriever with 158 tweets, Labrador retriever with 108 tweets, Pembroke with 95 tweets, Chihuahua with 91 tweets and Pug 62 tweets.
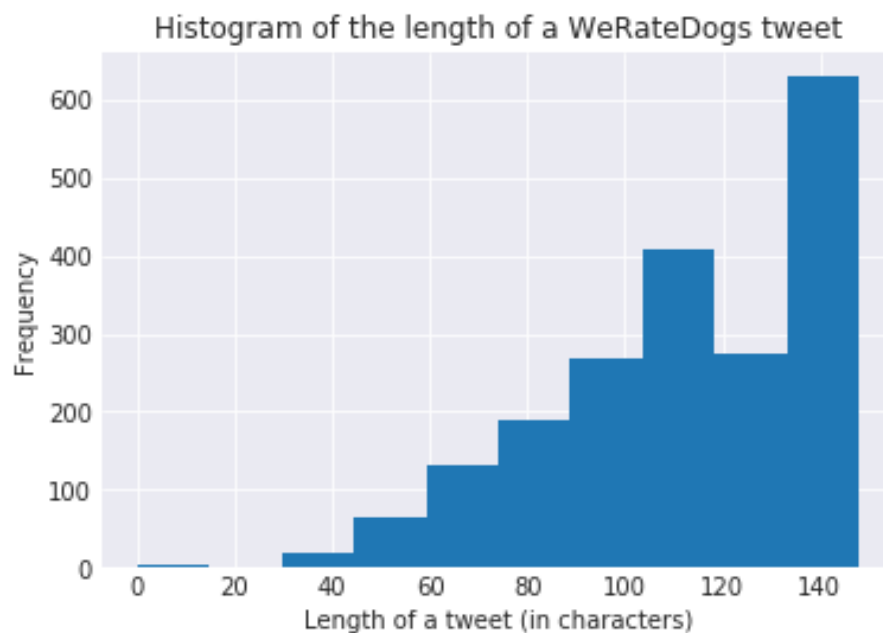

Most represented breeds of dog in WeRateDogs tweets

With this pie chart we notice that even though it's the 15 most represented dog breeds, it barely counts for 50% of all the tweets.

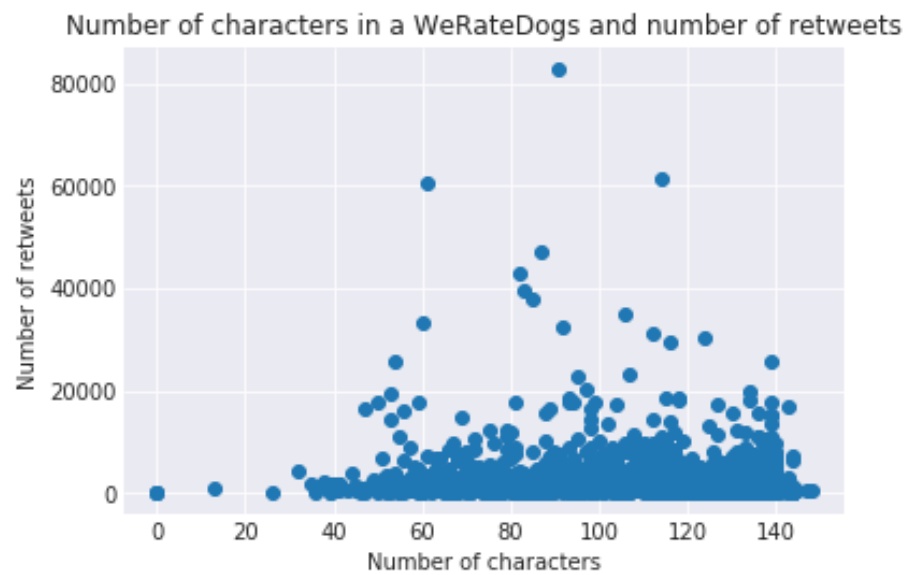## 2.3   Insight 3 : length of a WeRateDogs tweet

What is the average length of a WeRateDogs tweet? Does the length of a tweet affect the number of retweet received?

A histogram is a good way to see the distribution of a variable, here the length of a tweet (number of characters).

The distribution of the lengths of WeRateDogs tweets is left skewed, has an average length of 111 characters, and a median of 116 characters.

In order to see if the length of a tweet affect the number of retweet received, a scatter plot has been made, with the number of characters on the x axis and the number of retweets on the y axis.

Number of characters in a WeRateDogs and number of retweets

We can't conclude much looking at this scatter plot, although we can say that tweets with a number of characters betweet 0 and 40 characters seems to have less retweets in general (but we can't say it with certainty as we have a very small number of tweets observed with that specific amount of characters).