UDACITY

Act Report

Yousef Majeed

This report is for the wrangling data project (fourth project) in the Data Analyst Nanodegree Program that running by Udacity corporate.

Data Analyst Nanodegree

May-29, 2019

Overview

This report aims to provide insights and visual displays based on the analysis made in the project called Data Wrangling Project – Enhanced Twitter Archive for WeRateDogs account.

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. WeRateDogs asks people to send photos of their dogs which are rated on a scale of one to ten but are invariably given ratings in excess of the maximum, such as "13/10". WeRateDogs has over 4 million followers and has received international media coverage.
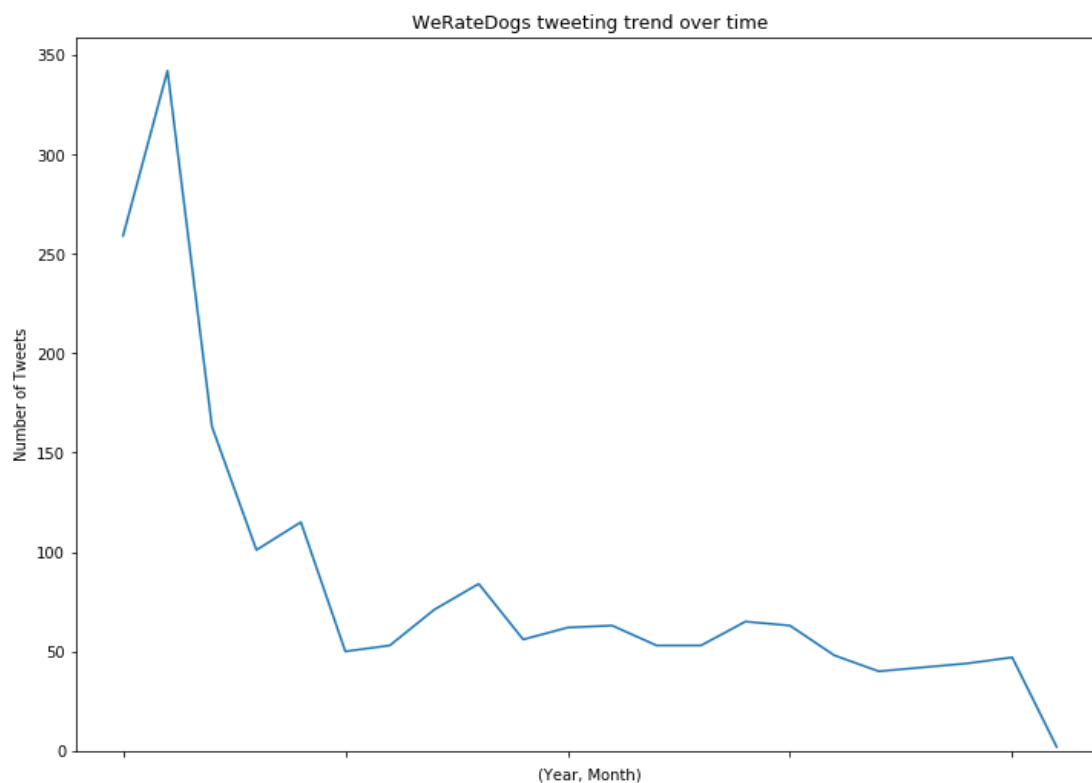
Analysis & Visualizing Data

1. The dataset contains 1876 observations and 24 variables.

2. Most of Most of the dogs are rated 12/10. Excluding the 2 rating outliers (420 and 1776), the highest rating received by any dog is 14/10. However, only 2% (i.e. 37) dogs got this rating. So, we can say that WeRateDogs is probably very conservative when it comes to rating a dog the highest.12.
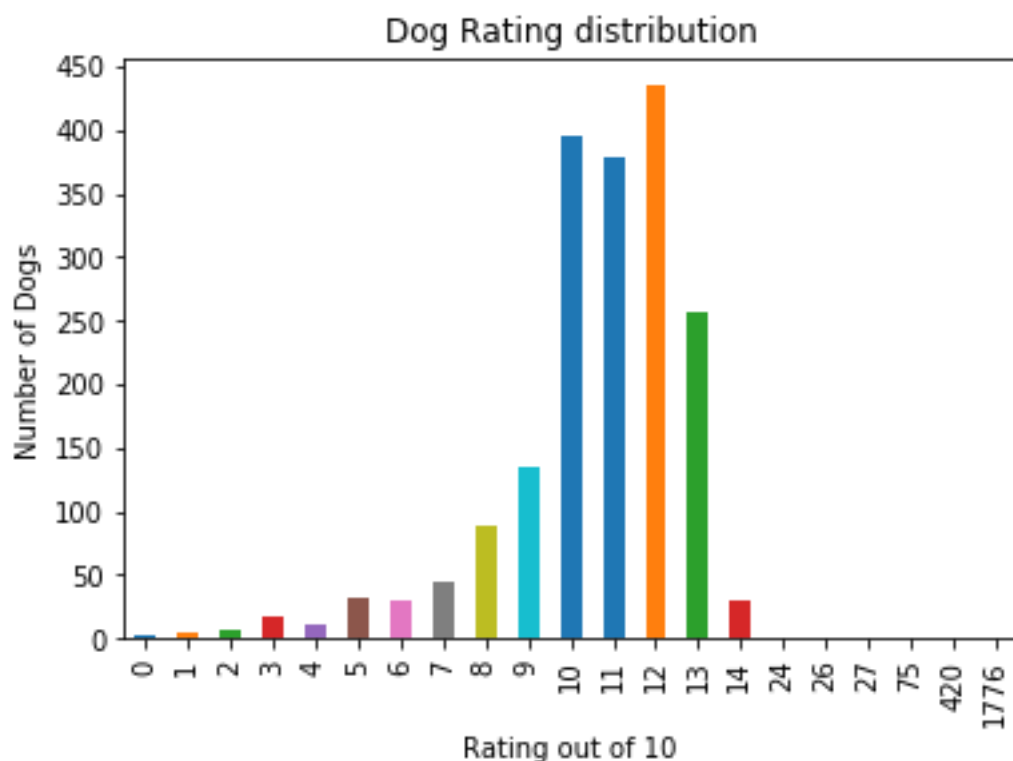
3. WeRateDogs, a Twitter account started in 2015, tweeted a lot in its starting days. For example, it posted around 300 tweets in the month of November 2015, which increased to around 350 tweets in exactly the next month. However, its tweeting activity reduced drastically in the subsequent months, averaging between 50 to100 since April, 2016.

4. The most common dog stage is pupper, which is a small dog, ussualy younger, as presented by the @dog_rates account. We also have many dogs that are not classified.

5. WeRateDogs has posted 98% of the tweets from iPhone, out of the 1994 tweets, 1955 were posted from iPhone. Just a confirmation that the only mobile device WeRateDogs uses to post tweets is an iPhone.

Most used Twitter source
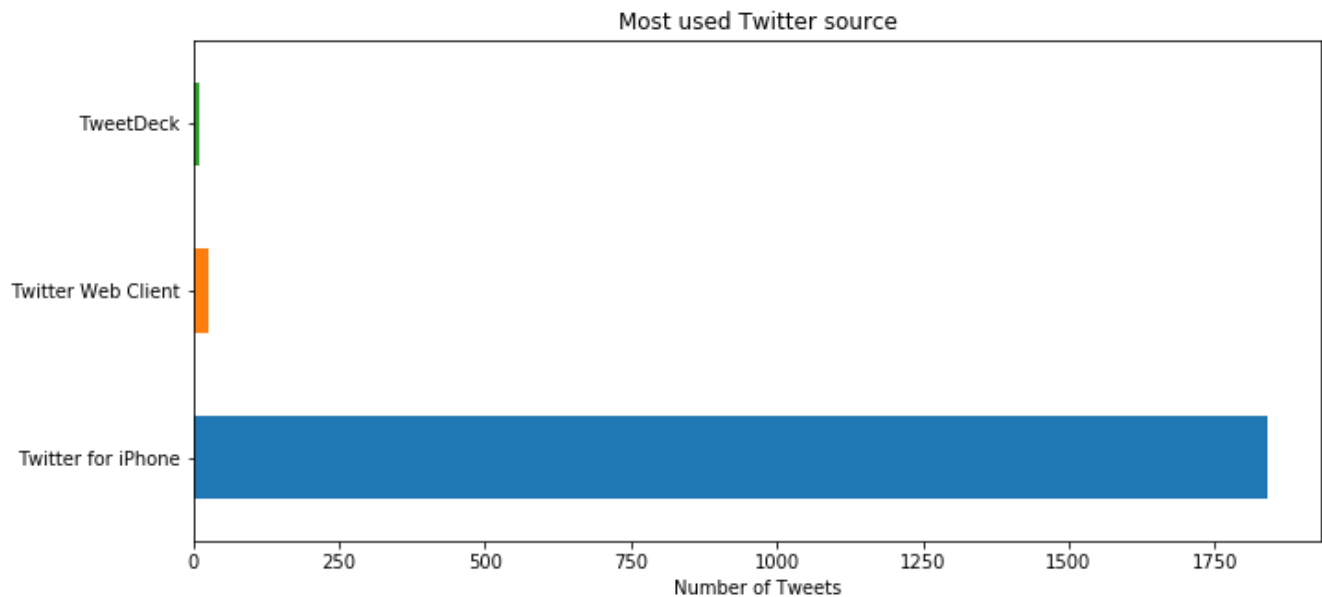
6. Charlie is the most common dog name, 11 of the dogs were named Charlie and it's the most we have. A close second will be Lucy, Cooper and Oliver, with 10 dogs being named as these.

Top 6 common dog names
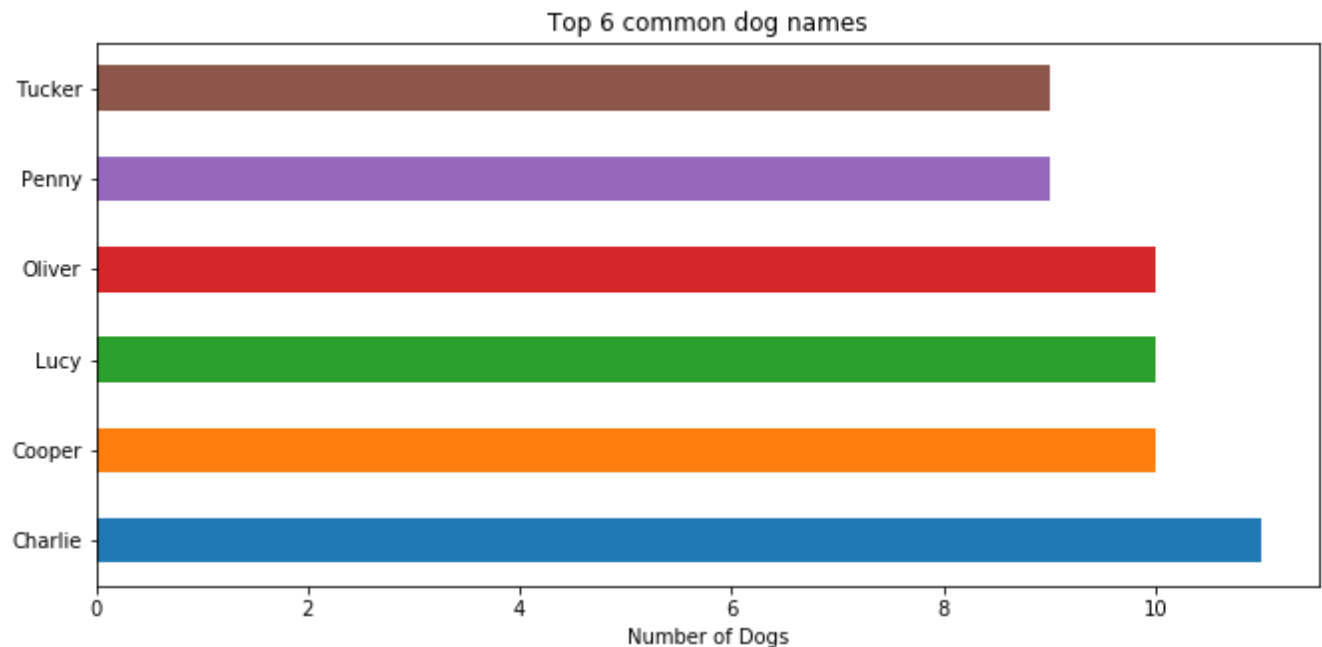
7. The highest number of likes is 132810. The median number of likes is 4216.0 and the mean is 9073. Here we have a distribution of likes which clearly states that we also have outliers, thus, a higher mean.

Distribution of Likes
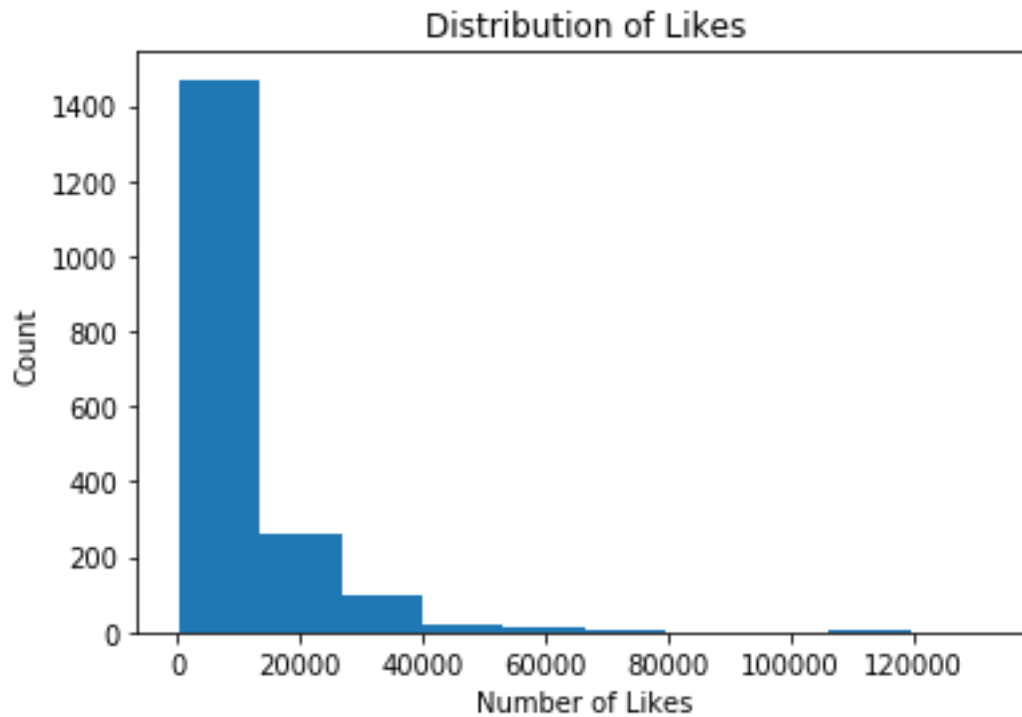
8. We have a strong positive relationship between likes and retweets. As the number of likes increases the retweets increase also.

Likes vs. Retweets
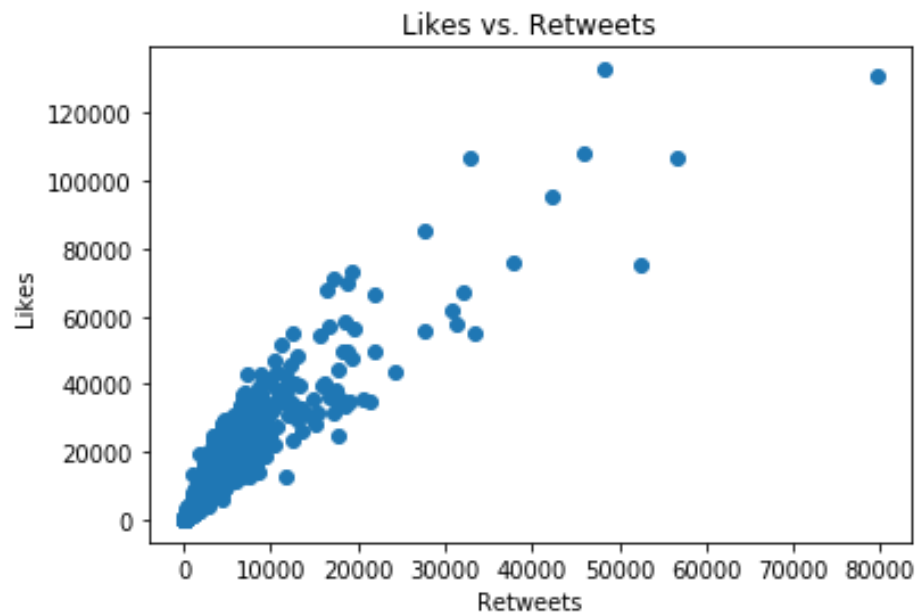
Insights

An interesting insight that I found is the fact that the ratings are awarded upfront, so they are not based on likes and retweets. The ratings are subjective, based on the picture of the dog. From 1876 observations I noticed that the average numerator is 12, so is still near 10. Surprisingly I discovered that in the pictures are not only dogs, as we just saw.