UDACITY

Data Analysis Nanodegree

PROJECT 5: Documentation of analysis and insights into final data

Deepak Nandipati

Date Created: May 7, 2019

Introduction

"WeRateDogs" a twitter page that has been increasing in popularity with time, launched 3 years ago and now it holds a staggering 8.11 Million followers with over 10.3 thousand tweets. The page often posts cute dog pictures, in which a rating is often given, thus the name "WeRateDogs". Almost always the postings have a rating of greater than 10/10, but there have been exceptions. In this document, with the help of Udacity in providing twitter archive and image predictions of data, a dataset was analyzed and insights have been discovered about this page and they are as follows.

What are the most popular Dog breeds?

The image prediction dataset had quite a few errors that needed to be polished and cleaned. After isolating highest probability of image prediction, which pertain to dog images only. It was shown that Labrador retriever have the highest count of posts, followed by Golden retriever and Pembroke.
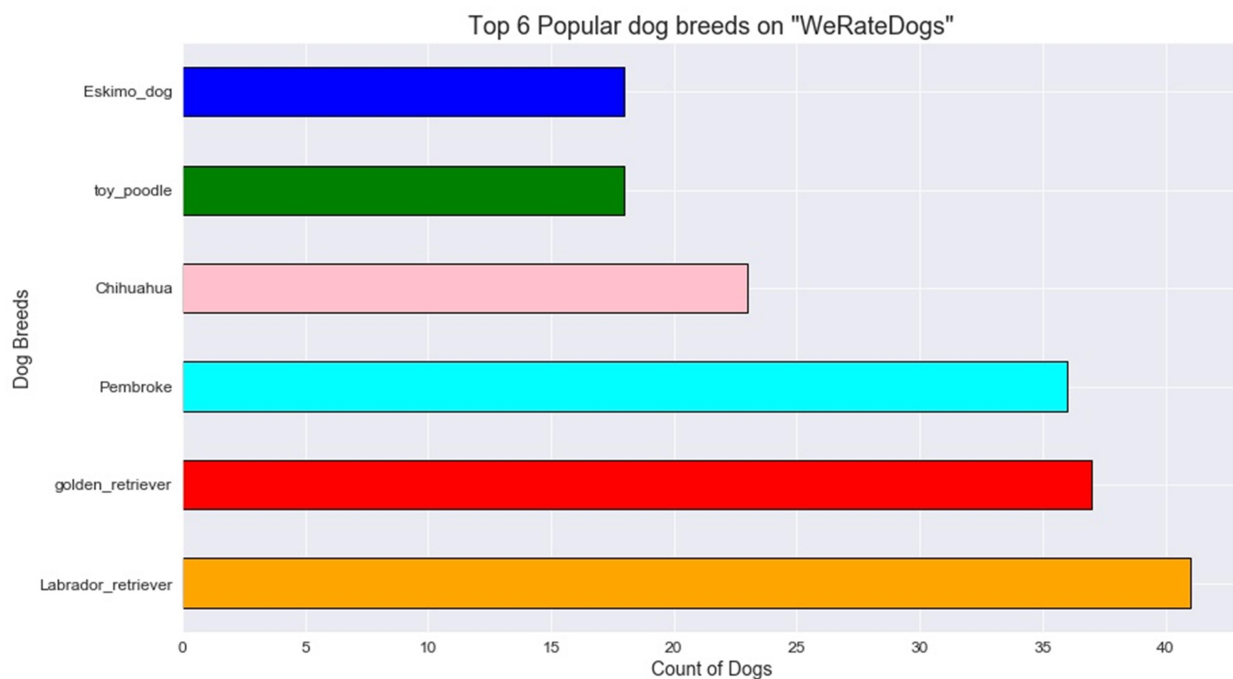
Fig1. It was found that Labrador retrievers are the highest in occurrence with 41 count, followed by Golden Retriever (37), Pembroke (36), Chihuahua (23) and Eskimo dog and Toy poodle with equal counts of (18).

How is the distribution of dog stages?

It was shown from analysis that 'pupper' is the most common stage of dog that is posted with a count of 50, followed by doggo with 28, puppo with 6 and floofer with 3. Puppo is defined as a smaller puppy, whereas pupper is puppy and doggo is another word for dog. Most of the dataset did not fall into any of these categories due to image prediction not being accurate.
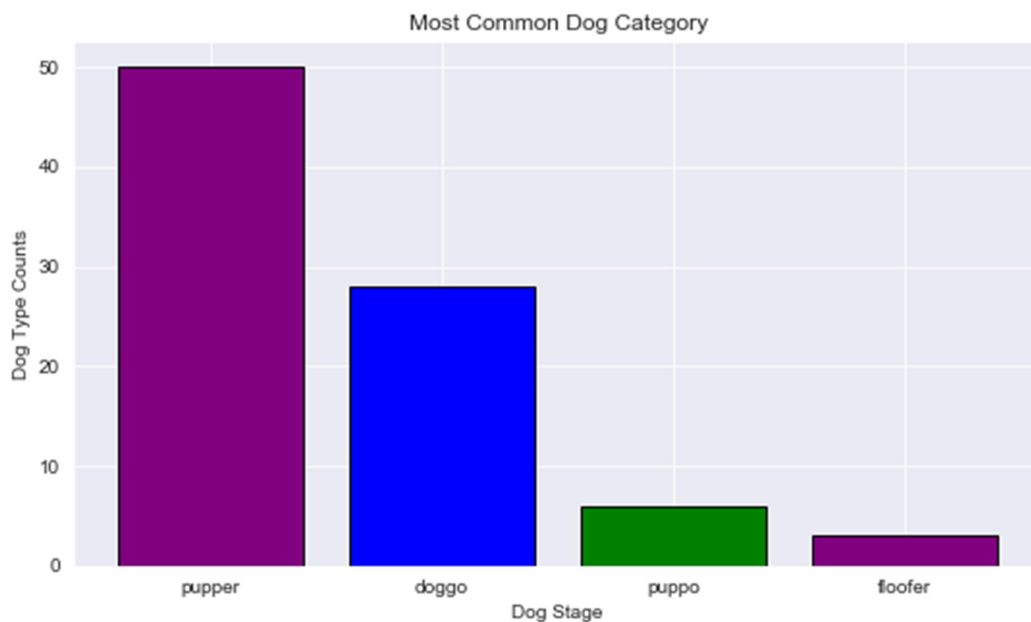

Fig2. Pupper has highest counts in twitter page, followed by doggo puppo and floofer

Dog ratings at "WeRateDogs" page

As mentioned earlier, the WeRateDogs page has always been about ratings that are not proportional to typical 100%. Below we can see the distribution of ratings out of 10 given to dogs. The most common rating received for dogs are 12/10,

followed by 11/10 and then 10/10. It is very rare that they rate dogs with something less than 10.
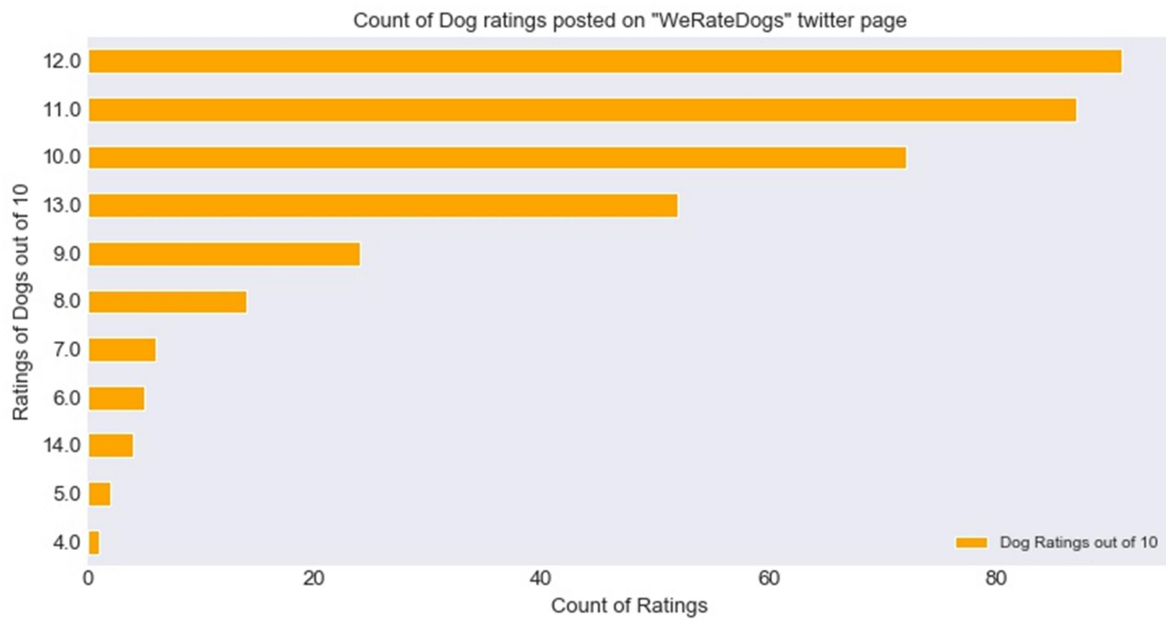

Fig3. The most common rating given at 'WeRateDogs' is 12/10 with a count of 152, followed by 11/10 with count of 138 and 10/10 with count of 127

Dog ratings since the start of page

Since the beginning of 'WeRateDogs' which was of November 15, 2015 it has been shown that the traffic of page increased rapidly. Along with the traffic, it also appears that the ratings of dogs have got better ever since. With the latest data in the archive of July 29, 2017, the graph below shows the slight increase in ratings. Y-axis being the percentage of Dog ratings, norm being above 100% and X-axis is timeframe of data
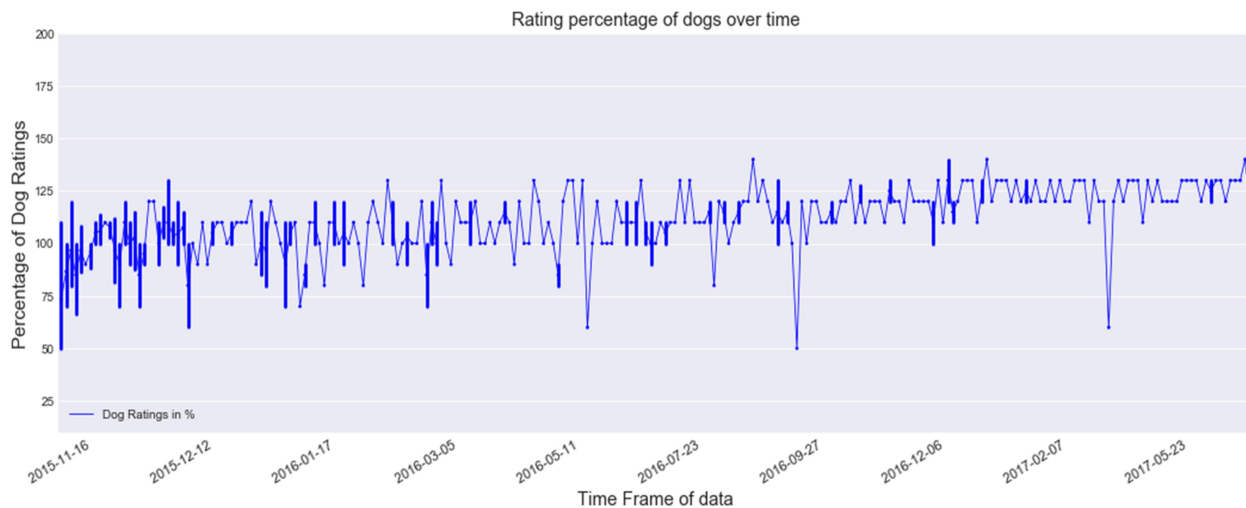
Fig4. Graph showing the increase dog rating over time. X-axis represents time and Y-axis the percentage of dog ratings

How are the Retweets and Favourites at 'WeRateDogs'?

With analysis of data, it has been indicated that the Retweets and favourites of the page have been drastically increasing. Some of the posts have even gone viral such as the post on June 4, 2016 that was of a Siberian husky blowing bubbles under water. This 30-second video grasped the world's attention and has kick-started the surge of traffic on page. The original post at time of dataset consisted of 60,989 retweets and 119,787 favourites; here is a screenshot of the post:

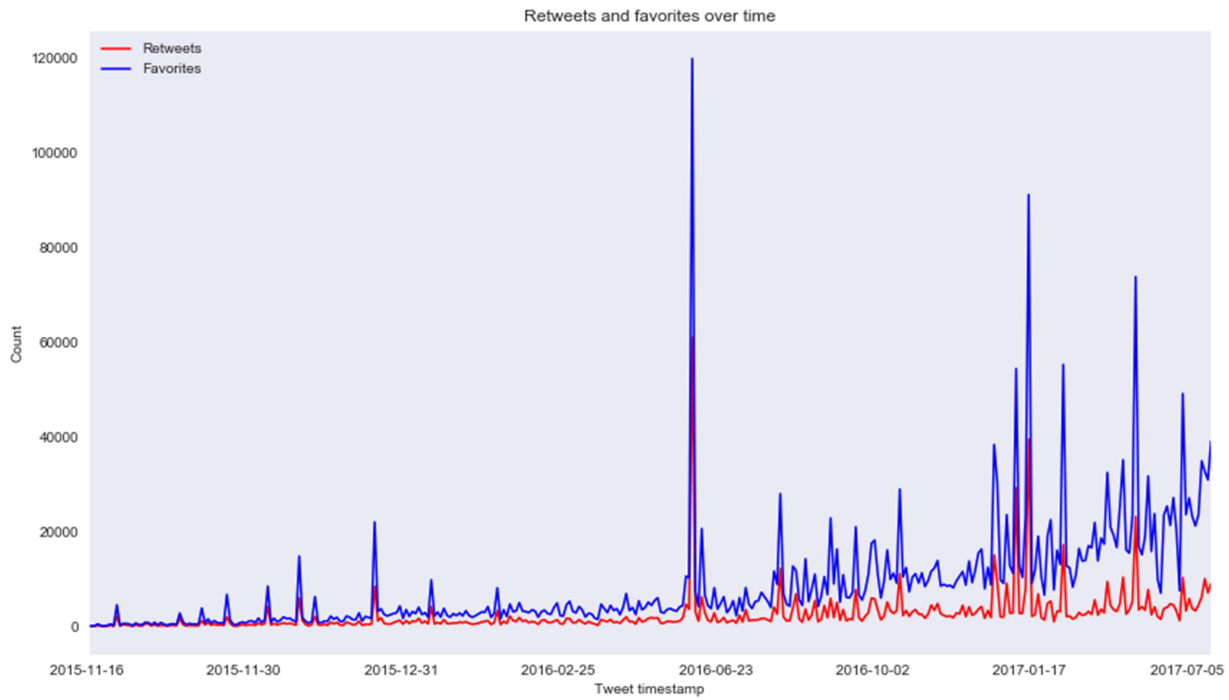URL: https://twitter.com/dog_rates/status/739238157791694849?lang=en

Fig5. Retweets are presented by color red and favourites by color blue. The graph shows increase of traffic on page after the viral post that was published on June 4, 2016.