PAWS FOR THOUGHT

By Janny Wong

WeRateDogs™ @dog_rates

This is outrageous. We do not rate Deadly Moroccan Road Lions, no matter how jubilant they may appear. We only rate dogs... 12/10

6:13 PM - 19 Mar 2018

Real world data rarely comes clean and as part of the Udacity Data Analyst Nanodegree program, I had the opportunity to test my newly acquired data wrangling and analysis skills.

The project brief was to wrangle WeRateDogs twitter archive data to create interesting and trustworthy analyses and visualizations.

For the uninitiated, WeRateDogs is an account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by a college student in the United States and has received international media attention. As of February 2021, it has almost 9 million followers and it's follower base includes climate activist Greta Thunberg and US congress representative Alexandria Ocasio-Cortez.

Dogs are rated on a scaled of one to ten, however are invariably given ratings in excess of 10, such as 12/10 (as illustrated in the screenshot on the top right). Why? Because "they're good dogs Brent".

In addition to the twitter data, an image predictions file was also analyzed. Every image in the twitter archive data was run through a neural network that can classify breeds of dogs. Twitter API was queried to obtain the number of retweet and favorite counts.

In my analysis that follows, I have analysed the following:
- Most popular dog breed
- Frequencies of photos that are not actually dogs
- The relationship between retweet count and favorite count and
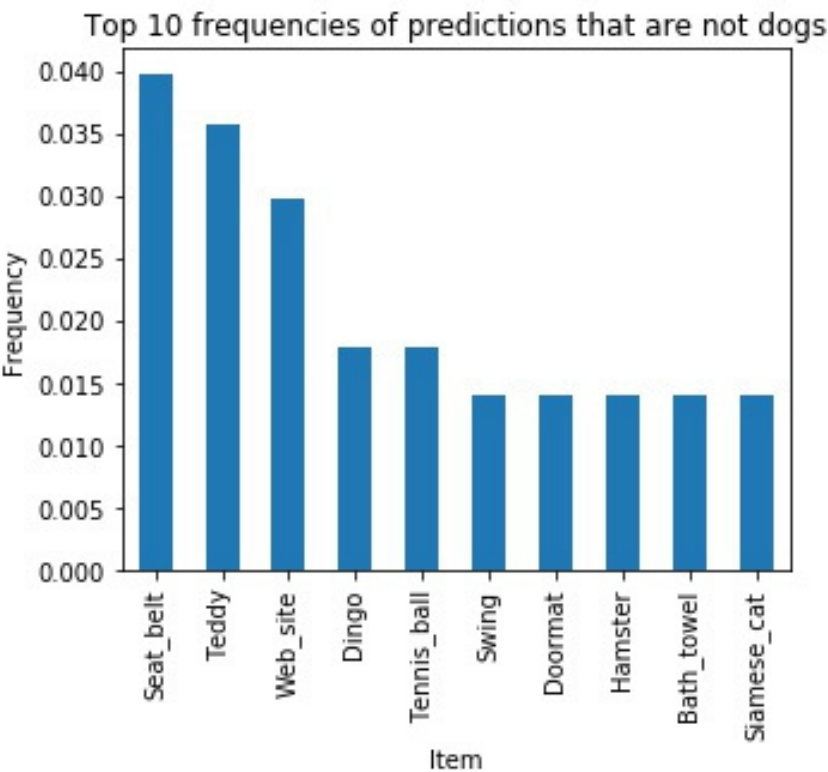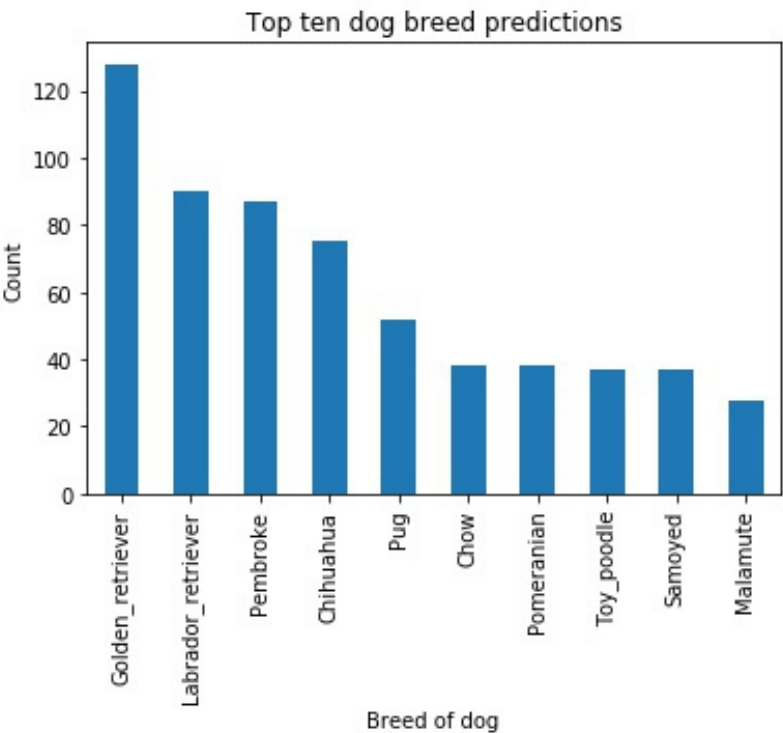- Number of tweets.

"THEY'RE GOOD DOGS BRENT"

GOLDEN RETRIEVER WINS GOLD!

Golden retrievers appear to be the most popular dog breed photographed and posted on twitter with 128 posts. Labrador retrievers are not even a close second with 90 and Pembroke coming in third with 87 posts.

Top ten dog breed predictions

Top 10 frequencies of predictions that are not dogs

BARKING UP THE WRONG TREE

Of the image predictions, the neural network predicted seatbelts appear to be the highest non dog breed item. This may be due to photos of dogs in cars so the neural network initially predicts the dog as a seatbelt.

Teddy is second as there are many photos of dogs with soft toys.

Website is in third place, these tend to be chat screenshots of dogs.

I LIKE BIG MUTTS AND I CANNOT LIE

There appears to be a positive correlation between retweet count and favorite count.

This suggests that if a tweet gets many likes, it will become more popular leading to an increase in retweet count.

Retweet count vs Favorite count
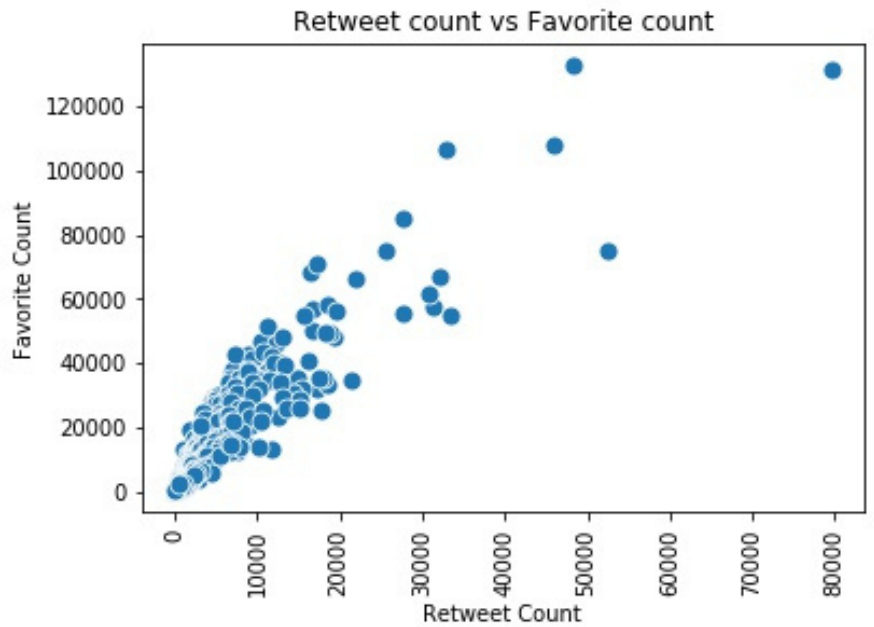
ONE DOG YEAR EQUALS SEVEN HUMAN YEARS

The number of tweets was analyzed during the duration of the dataset (15 November 2015 to 1 August 2017). The number of tweets peaked at 26 on 29 November 2017, when the account was first set up and was gaining in popularity. As time went by, the number of tweets fell sharply after the first couple of months, spiking in April 2016, October 2016 and December 2016.

Distribution of tweets over the duration of the dataset