Project No. 4

WeRateDogs - Twitter Data

Act report - external document

Author: Robert HROMADA (SVK)

Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson, and has received international media attention both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter for breaking these aforementioned laws. [1]

Today, The WeRateDogs Twitter is "Your Only Source For Professional Dog Ratings Instagram and Facebook".[2]

Data were downloaded a followed:

1. Archive provided by WeRateDogs for Udacity learning pourpose,
2. Additional data downloaded from **WeRateDogs Twitter accout via API,**
3. Tweet image predictions data downloaded from provided web page.

Data Analysis and Visualisation

After gathering a cleaning phase my first look falls into numerical data. Interesting question is if data somehow correlated or not. Simply heatmap (Figure 1) constructed from followers_count', rating_numerator, rating_denominator, retweet_count and favourite_count showed that really strong correlation is between retweet_count and favourite_count (Figure 1, 2). We can say that who have retweeted have also done with favorites tag.
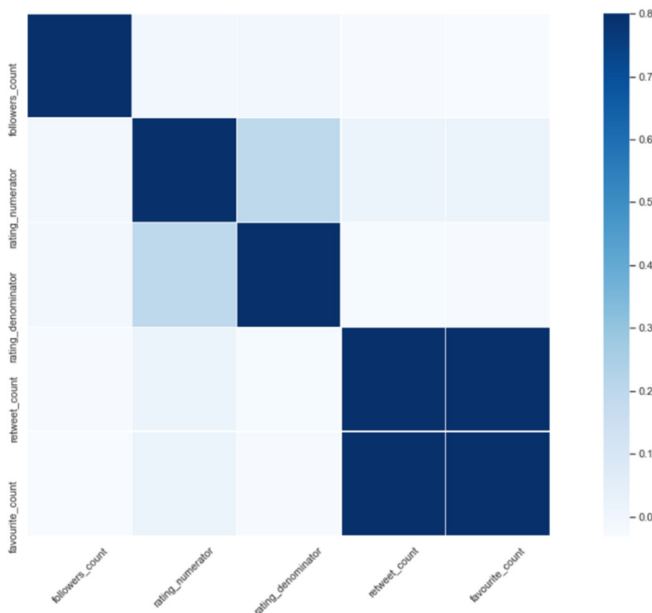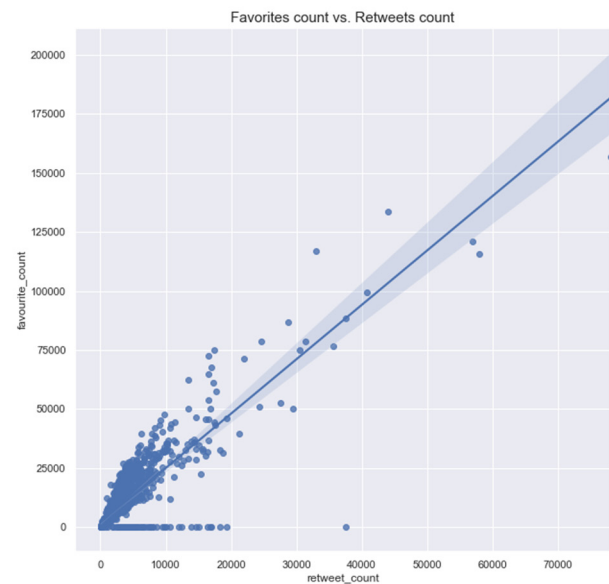

Figure 1: Heatmap

Figure 2: Favorites count vs. Retweet count

And dog breed is most common in by WeRateDogs? From the bar plot the favorite one is Golden retriever followed by Labrador retriever (Figure 3). It's hard to say why these breeds are at the top. Based on Wikipedia the breed's friendly, gentle temperament means it is unsuited to being a professional guard dog, but its temperament has also made it the third-most popular family dog breed in USA. [3]
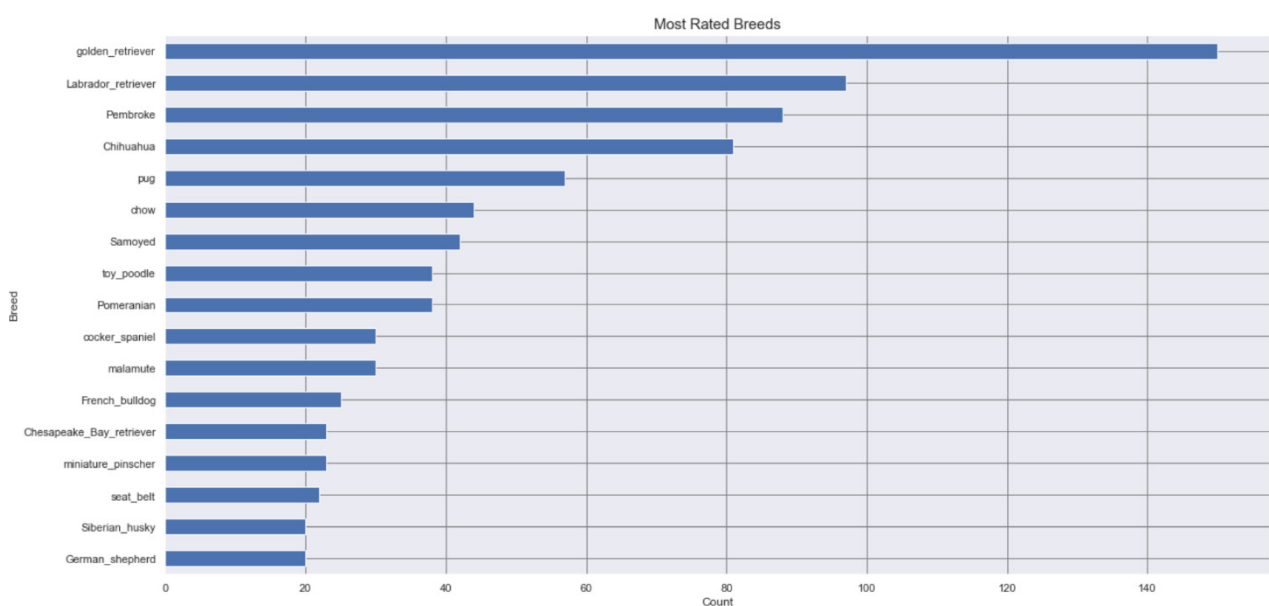

Figure 3: Most rated breeds

For the dog breed prediction for the image in the tweets were used three different algorithms. If we look little bit closer to result from these algorithms, we can see that results from all of them have almost same ratio between True vs. False prediction (Table 1).

Algorithm No.	TRUE	FALSE	% of True
1	1520	539	73,82
2	1540	519	74,79
3	1487	572	72,22

Table 1: Image prediction algorithm

And what about the favorite name? Does it exist between dog owners one top name for their beloved? From provided data it can be tricky question. Only 935 dogs in WeRateDogs provided archive have a name. More than 55% tweets are without name. But still we have some and top name (Figure 4) are Charlie, Penny, Tucker, Olivier and Cooper with ten votes. They are followed with Lucy (9) and Winston, Sadie and Lola (8).
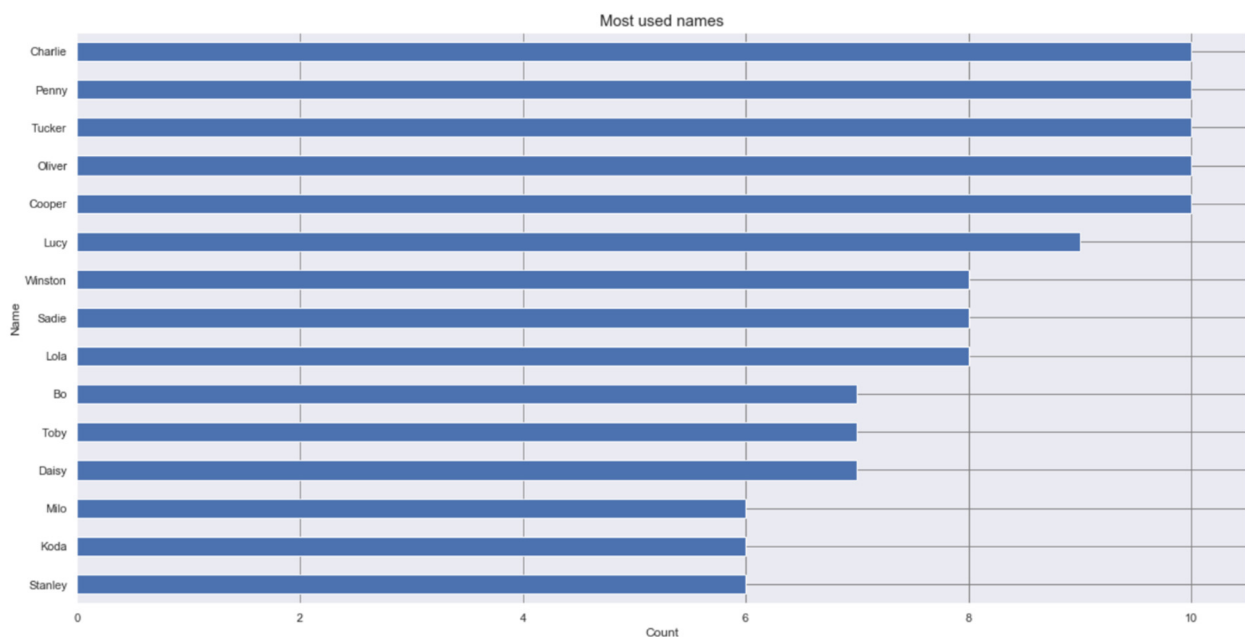

Figure 4: Most used names

WeRateDogs have wrote book "#WeRateDogs: The Most Hilarious and Adorable Pups You've Ever Seen" where the different dog stages are explained. Twitter archive contains dog stage as well (Table 2).

Dog stage	doggo, puppo	doggo, floofer	doggo, pupper	floofer	puppo	doggo	pupper
Count	1	1	11	7	23	65	210

Table 4: Dog stage

From this table is clearly visible that not all dogs were classified in some "dog stage" but we can find double stage as well. This is due to fact, that on a picture can be two dogs e.g. mother and her child.

Conclusion

As a conclusion for our first look into WeRateDogs archive we will use Amozon text [4]:

"Based on the Twitter and Instagram sensation, #WeRateDogs features the most heroic, over-the-top adorable, wildly successful, all-around entertaining dogs the world has ever seen! It combines extraordinary photos with ridiculous captions to expand the knowledge and overall lives of its readers. It is not only an exceptionally reliable source for dog greatness, it is also a one stop shop for happiness.

Labradors and poodles, beagles and dachshunds, chihuahuas, and great danes, bulldogs, collies, and shepherds. Here they are all cute and hilarious!"

References

[1] – https://en.wikipedia.org/wiki/WeRateDogs

[2] – https://twitter.com/dog_rates/

[3] – https://en.wikipedia.org/wiki/Golden_Retriever

[4] - https://www.amazon.com/WeRateDogs-Most-Hilarious-Adorable-Youve/dp/1510717145