

We Rate Dogs Act Report

London, 28th of May of 2019
Author: Vivian Sacon

Content:

1. Overview
2. Dog Stage analysis
3. Dog predicted breed
4. Top rated

1. Overview

This report is about my efforts to analyze the final stage of data that I have worked in during the We Rate Dogs data wrangling project.

All the information contained in this report can also be found in the section Analyze of the file wrangle_act.ipynb that is the same file which contains all the efforts performed to conclude the We Rate Dogs data wrangling project.

After the three steps of data wrangling (Gather, assess, clean), I have created a final file called twitter_archive_master.csv with the final dataset properly treated.

An important thing to notice is that we don't have the complete information to all the tweets of the page WeRateDogs, this page has more than 5000 tweets. It is also important take into account that we are assessing just the main tweets and not considering retweets, tweets without images and tweets that are not related to dogs.

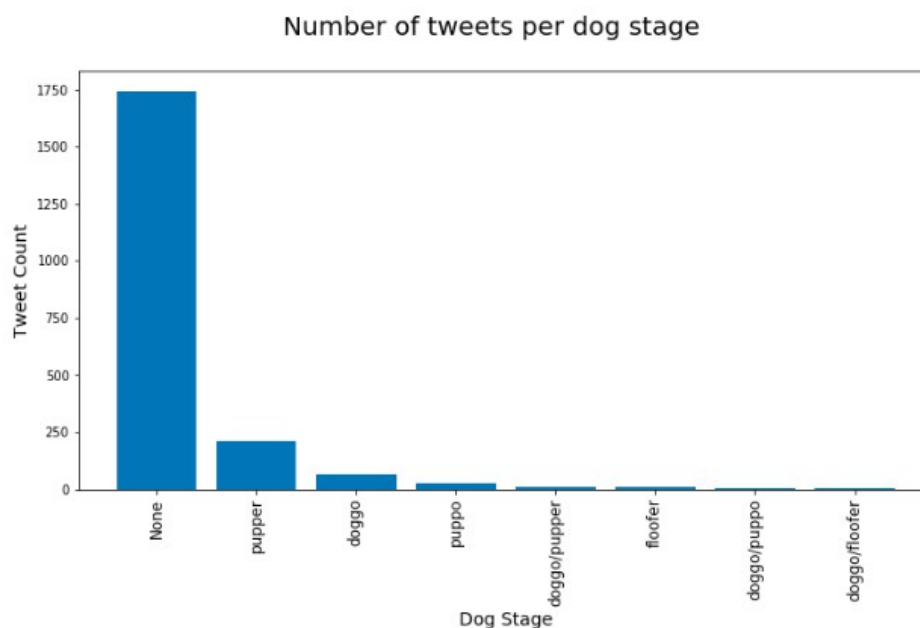
In a universe of more than 5000 tweets, we are analyzing 2064, which means that our analysis correspond to a sample of more or less 40% of the universe.

2. Dog Stage analysis

As we can observe in the two analysis below, the majority of the tweets didn't stage the dogs. 84% of the tweets are related to non staged dogs.

Looking to the rest of the analysis, the other majority of tweets staged the dogs as puppies, it is a relevant number of puppies because in a remaining 15% of the data, 10% is staged as pupper and other 0.4 as doggo/pupper.

The other categories represent less than 5% of the universe making the pupper the favorite stage classification.



```
#Percentage of tweets per dog stage
(WeRateDogsFinal['dog_stage'].value_counts()/WeRateDogsFinal['dog_stage'].count())*100
```

None	84.455206
pupper	10.217918
doggo	3.244552
puppo	1.113801
doggo/pupper	0.532688
floofer	0.338983
doggo/floofer	0.048426
doggo/puppo	0.048426

Name: dog_stage, dtype: float64

3. Dog predicted breed

I would like to start this analysis looking to the level of confidence in the predictions.

As we can see in the numbers below, less than 50% of the breed predictions have a confidence level equal or higher than 50%.

With this information in mind, I will proceed the breed analysis ignoring the confidence level.

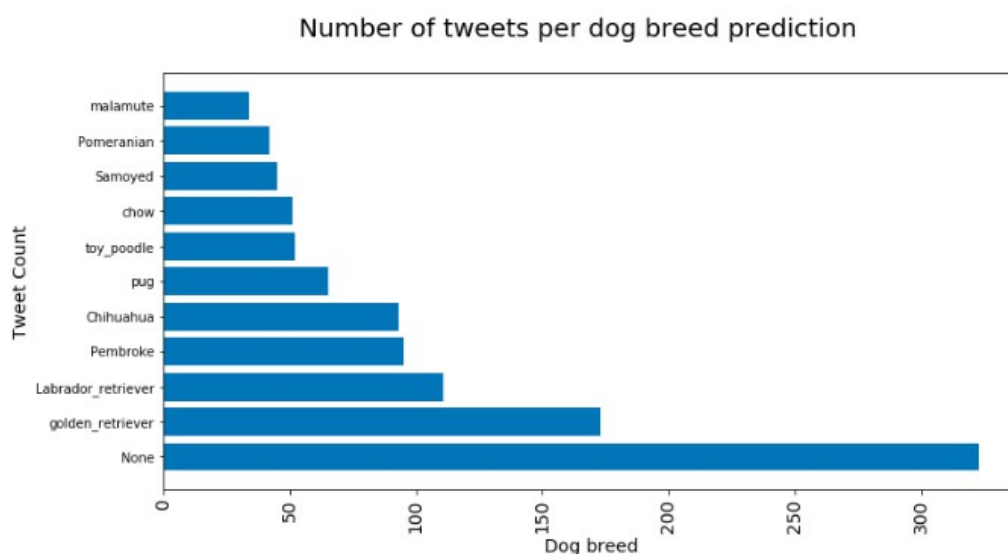
- Nr of tweets in the final file: 2065
- Nr of tweets with prediction confidence higher than 0.49: 966
- Percentage of tweets with prediction confidence higher than 0.49: 47%

Count of tweets per dog breed analysis:

Here we also have a considerable number of predictions defined as 'None', repeating what we saw in the dog stage analysis, however, at this time the number is less considerable, representing just 16% of the tweets.

The rest of the tweets are very well divided between different breeds, however, we have 8 breeds that may be considered relevant in this analysis, taking a special attention to the Golden Retriever breed that alone represents 8% of the tweets.

We definitely have a favorite breed along the users of the page, congratulations to all the cute Golden Retrievers out there.



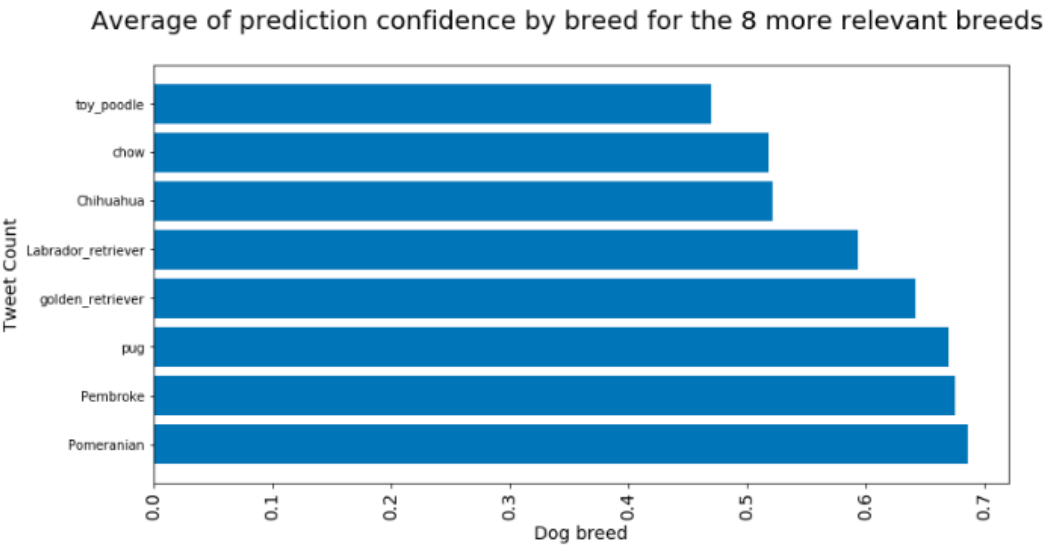
```
#Percentage of tweets per dog breed prediction
```

```
(WeRateDogsFinal['prediction'].value_counts()/WeRateDogsFinal['dog_stage'].count())*100
```

None	15.685253
golden_retriever	8.025013
Labrador_retriever	5.158937
Pembroke	4.742053
Chihuahua	4.637832
pug	3.074518
toy_poodle	2.605524
chow	2.344971
Pomeranian	2.136529
Samoyed	1.980198
malamute	1.667535
Chesapeake_Bay_retriever	1.511204
cocker_spaniel	1.511204
French_bulldog	1.459093
miniature_pinscher	1.250651
Eskimo_dog	1.146430
Cardigan	1.094320
German_shepherd	1.094320
beagle	1.042209
Staffordshire_bullterrier	1.042209
Siberian_husky	1.042209
Shih-Tzu	1.042209
Maltese_dog	0.990099
Shetland_sheepdog	0.937989
Rottweiler	0.937989
Lakeland_terrier	0.885878
basset	0.885878
kuvasz	0.885878
Italian_greyhound	0.833768
West_Highland_white_terrier	0.833768
...	

Average of confidence in the predictions per dog breed

We can see in the graph below that the most popular dog breeds also have a considerable high level of confidence in the predictions related to them which make the classification a bit more accurate.



Top rated breeds

Comparing the top rated breeds with the most tweeted breeds we can see that they are not strongly related.

The maximum mean for a dog breed is 2.7, scored by the breed clumber, folowed by more estable means that are stated between 1.6 and 1.12.

Our 8 most popular dog breeds have a range between 1.2 and 1.03 which is not far away from the more stable means.

If we look to the maximum score of our 8 most popular breeds we will see that Pomeranian has an impressive maximum score of 7.5, followed by our favorite golden_retriever that has 3.4.

```
#Mean of score by breed
WeRateDogsFinal.groupby('prediction')['dog_score'].mean().sort_values(ascending=False)
```

prediction	
clumber	2.700000
None	1.586378
Bouvier_des_Flandres	1.300000
Saluki	1.250000
Pomeranian	1.247619
Tibetan_mastiff	1.240000
briard	1.233333
Border_terrier	1.214286
silky_terrier	1.200000
standard_schnauzer	1.200000
Irish_setter	1.183333
Samoyed	1.175556
Gordon_setter	1.175000
golden_retriever	1.168478
kuvasz	1.163158
Australian_terrier	1.150000
giant_schnauzer	1.150000
wire-haired_fox_terrier	1.150000
Great_Pyrenees	1.146667
chow	1.145098
Norfolk_terrier	1.142857
cocker_spaniel	1.142424
Pembroke	1.141053
Eskimo_dog	1.140909
Doberman	1.133333
Irish_water_spaniel	1.133333
Greater_Swiss_Mountain_dog	1.133333
Leonberg	1.133333
kelpie	1.130769
Blenheim_spaniel	1.127273


```
#Mean of score by breed for the 8 most popular breeds
```

```
WeRateDogsFinal.loc[
    (WeRateDogsFinal['prediction'] == 'golden_retriever') | (WeRateDogsFinal['prediction'] == 'Labrador_retriever') |
    (WeRateDogsFinal['prediction'] == 'Pembroke') | (WeRateDogsFinal['prediction'] == 'Chihuahua') |
    (WeRateDogsFinal['prediction'] == 'pug') | (WeRateDogsFinal['prediction'] == 'toy_poodle') |
    (WeRateDogsFinal['prediction'] == 'chow') | (WeRateDogsFinal['prediction'] == 'Pomeranian')
].groupby('prediction')['dog_score'].mean().sort_values(ascending=False)
```

```
prediction
Pomeranian      1.247619
golden_retriever 1.168478
chow            1.145098
Pembroke        1.141053
Labrador_retriever 1.116216
toy_poodle      1.105769
Chihuahua       1.055230
pug             1.030769
Name: dog_score, dtype: float64
```

```
#Maximum score for the most popular breeds
```

```
WeRateDogsFinal.loc[
    (WeRateDogsFinal['prediction'] == 'golden_retriever') | (WeRateDogsFinal['prediction'] == 'Labrador_retriever') |
    (WeRateDogsFinal['prediction'] == 'Pembroke') | (WeRateDogsFinal['prediction'] == 'Chihuahua') |
    (WeRateDogsFinal['prediction'] == 'pug') | (WeRateDogsFinal['prediction'] == 'toy_poodle') |
    (WeRateDogsFinal['prediction'] == 'chow') | (WeRateDogsFinal['prediction'] == 'Pomeranian')
].groupby('prediction')['dog_score'].max().sort_values(ascending=False)
```

```
prediction
Pomeranian      7.500000
golden_retriever 3.428571
Pembroke        1.400000
Labrador_retriever 1.400000
Chihuahua       1.400000
toy_poodle      1.300000
pug             1.300000
chow            1.300000
Name: dog_score, dtype: float64
```

4. Top Rated

To conclude, let's have a look at the top 15 rated tweets and for this analysis I have removed predictions with less than 0.5 of confidence and maintained dog breeds different than 'None'.

And there he is again, the tweet with higher score is related to our loved golden_retriever, making this breed not only the most tweeted but also the first one in our dog_score.

```
#Tweet with the higher score
WeRateDogsFinal.loc[
    (WeRateDogsFinal['prediction'] != 'None') &
    (WeRateDogsFinal['prediction_conf'] >= 0.5)].groupby(['tweet_id', 'prediction', 'dog_stage', 'prediction_conf']
    )['dog_score'].max().sort_values(ascending=False)[:15]
```

tweet_id	prediction	dog_stage	prediction_conf	dog_score
810984652412424192	golden_retriever	None	0.871342	3.428571
778027034220126208	clumber	pupper	0.946718	2.700000
813812741911748608	French_bulldog	None	0.709146	1.400000
884441805382717440	Pembroke	None	0.993225	1.400000
878057613040115712	French_bulldog	None	0.839097	1.400000
820314633777061888	Gordon_setter	None	0.940724	1.400000
856282028240666624	Chihuahua	None	0.876543	1.400000
822462944365645825	Pomeranian	pupper	0.960199	1.400000
807621403335917568	golden_retriever	pupper	0.873233	1.400000
821407182352777218	Irish_setter	doggo	0.505496	1.400000
831911600680497154	bloodhound	None	0.777562	1.400000
828650029636317184	golden_retriever	None	0.649209	1.400000
774314403806253056	Eskimo_dog	None	0.596045	1.400000
854120357044912130	black-and-tan_coonhound	pupper	0.854861	1.400000
890240255349198849	Pembroke	doggo	0.511319	1.400000

Name: dog_score, dtype: float64

The 5 Top rated images:

1 - Golden Retriever:



2 – Clumber:



3 – Eskimo dog:

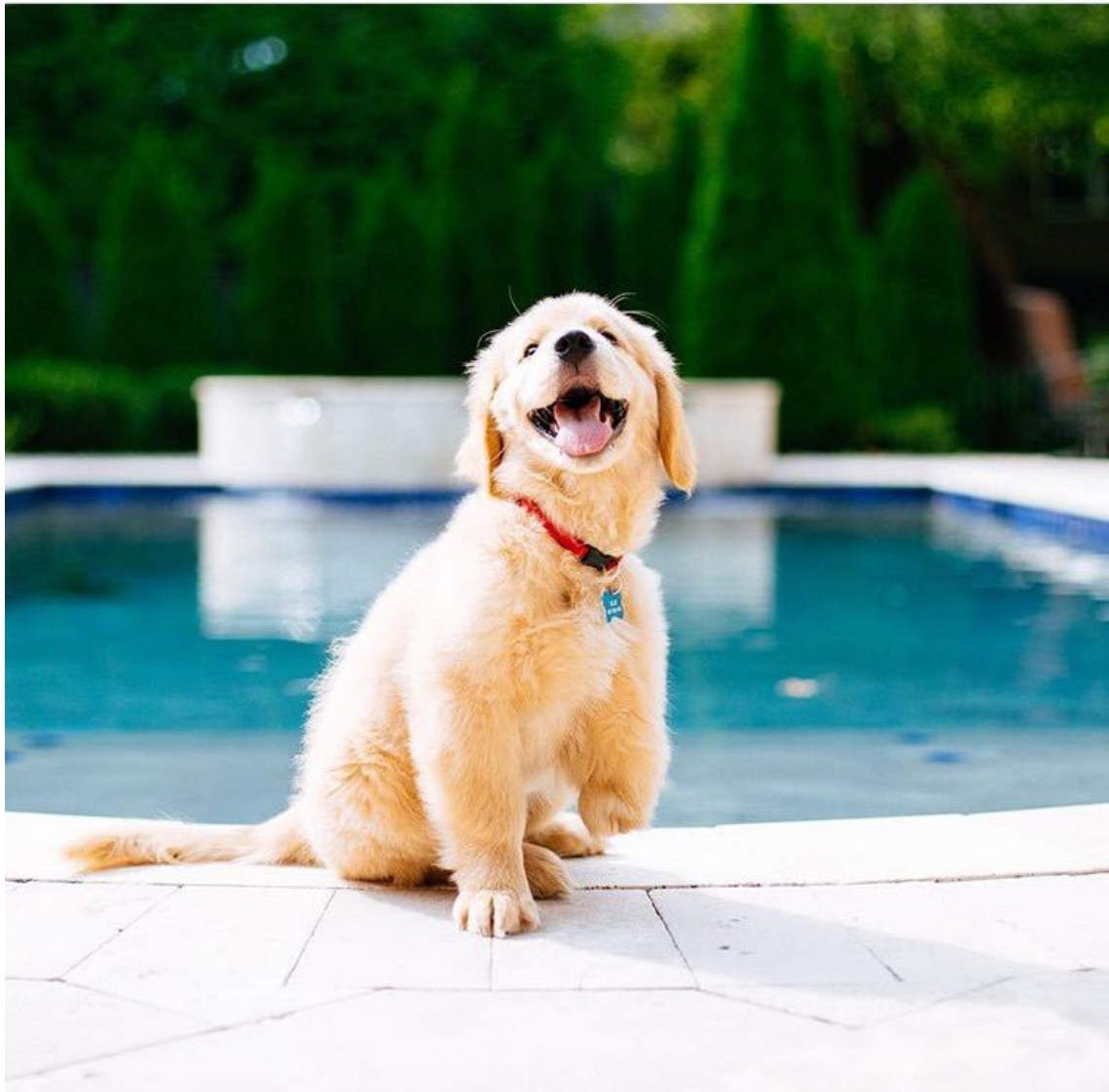


4 – Golden retriever:



ollievuesomuch

KillingIt >



This is Cooper. He was rescued when his previous owner ran from the cops and left him outside in subzero temperatures for three days. [REDACTED]

[REDACTED]

These days, Cooper really just enjoys being lazy, taking a good snooze, getting petted a lot, and taking in the smells, and we try to make sure he gets to do that as much as he wants. Everyone who meets him loves him and always comment on how sweet and calm he is. He's such a good dog.