

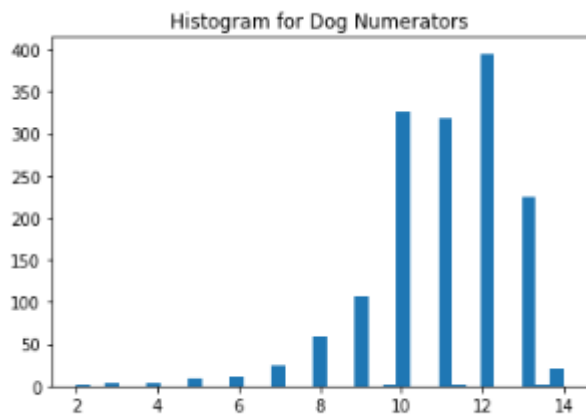
## Visualization and Analysis

The dataset we used to analyse contains the following data:

- tweet\_id
- timestamp This field contains the timestamp when the tweet was post.
- source
- text >> This field contains the tweet text.
- expanded\_urls
- rating\_numerator >> This field contains the numerator of the dog rating.
- rating\_denominator >> This field contains the denominator of the dog rating. It is always equal to 10.
- name >> This field contains the extracted dog name.
- jpg\_url
- img\_num >> Each tweet can contain 4 images. This field contains the image number of this image from the tweet.
- p1 >> This is the most likely dog breed from prediction.
- p1\_conf >> This field contains the probability of the prediction.
- p1\_dog >> This field predicts whether the image is a dog image.
- favorite\_count >> This field contains the number of tweet likes.
- retweet\_count >> This field contains the number of retweets of the tweet.

### Visualization

Since WeRateDogs has an intriguing method to rate dogs; that is the rating can exceed 10/10, I would like to see how the ratings are really distributed. From the plot below, I can see that most ratings are between 10/10 – 13/10.



### Insight #1 What breed of dogs are submitted the most?

From the dataset above, I would like to see which breed of dogs are most tweeted. My guess was that the most tweeted should be aligned with most popular dog breeds people kept as a pet.

Here is the result from the dataset. Golden Retriever and Labrador Retriever come is first and second. This is more or less aligned with the ranking from American Kennel Club. However, German Shepherd does not appear in the most-tweeted list. Maybe German Shepherd dogs are not kept as pet much by civilians, but are used as police or military dogs.

```
golden_retriever    145
Labrador_retriever  96
Pembroke            89
Chihuahua           82
pug                 57
chow                44
Samoyed             43
Pomeranian          38
toy_poodle          38
malamute            30
Name: p1, dtype: int64
```

Figure 1 Most Tweeted Dog Breed

Breed	2017 Rank	2016 Rank	2015 Rank	2014 Rank	2013 Rank
Retrievers (Labrador)	1	1	1	1	1
German Shepherd Dogs	2	2	2	2	2
Retrievers (Golden)	3	3	3	3	3
French Bulldogs	4	6	6	9	11
Bulldogs	5	4	4	4	5
Beagles	6	5	5	5	4
Poodles	7	7	8	7	8
Rottweilers	8	8	9	10	9
Yorkshire Terriers	9	9	7	6	6
Pointers (German Shorthaired)	10	11	11	12	13

Figure 2 Most Popular Dog Breed ( <https://www.akc.org/expert-advice/news/most-popular-dog-breeds-full-ranking-list/>)

## Insight #2 What breed of dogs received highest ratings?

p1	rating_numerator	
	mean	count
Saluki	12.500000	4
Tibetan_mastiff	12.400000	5
briard	12.333333	3
Border_terrier	12.142857	7
silky_terrier	12.000000	1
standard_schnauzer	12.000000	1
Irish_setter	11.833333	6
Eskimo_dog	11.777778	18
Gordon_setter	11.750000	4
Samoyed	11.697674	43

From our dataset, Saluki receives highest rating. However, we have to note that the number of tweets is only 4, so this may not be enough to conclude that Saluki generally will receive the highest ratings on WeRateDogs.



Figure 3 Saluki (<https://www.akc.org/dog-breeds/saluki/>)

### Insight #3 What breeds of dogs received highest retweets and favourites?

p1	favorite_count	
	mean	count
boxer	56439.000000	1
standard_poodle	48761.000000	2
Chihuahua	33410.444444	9
Mexican_hairless	33349.000000	1
Rottweiler	33267.000000	2
giant_schnauzer	33205.000000	1
Saluki	31575.000000	2
Blenheim_spaniel	31539.000000	2
Chesapeake_Bay_retriever	27652.333333	6
Irish_terrier	27546.000000	1

From our dataset, Boxer receives highest mean favourite counts. However, we have to note that the number of tweets is only 1, so this is likely not enough to conclude that Boxer generally will receive the highest number of likes on WeRateDogs.



Figure 4 Boxer (<https://www.akc.org/dog-breeds/boxer/>)

p1	retweet_count	
	mean	count
standard_poodle	27382.333333	3
boxer	17759.000000	1
Rottweiler	11145.000000	2
Chihuahua	10952.090909	11
Mexican_hairless	10303.000000	1
giant_schnauzer	9715.000000	1
Pomeranian	8763.250000	4
Lakeland_terrier	8262.000000	1
Tibetan_mastiff	7704.333333	3
Blenheim_spaniel	7670.500000	2

From our dataset, Standard Poodle receives highest mean retweet counts. However, we have to note that the number of tweets is only 3, so this is likely not enough to conclude that Standard Poodle generally will receive the highest number of retweets on WeRateDogs.



Figure 5 Standard Poodle (<https://www.akc.org/dog-breeds/poodle-miniature-standard/>)