

WeRateDogs Insights

1. Introduction

WeRateDogs is a popular Twitter account that rates people's dogs and adds funny comments about them. The ratings they give usually have a denominator of 10, but the numerators are usually higher than 10, like 11/10, 12/10, 13/10, and so on.

WeRateDogs has a large following of over 4 million people and has received attention from media worldwide. Luckily, Twitter has created a special tool called the developer's API that allows us to access the tweets about dogs. Additionally, we can use a neural network algorithm to predict different dog breeds based on the images in the tweets. By combining these sources of data and using data wrangling techniques, we can gather, examine, clean, and merge this large amount of data. This process helps us uncover connections and patterns within the data.

2. Analysis of We Rate Dogs Twitter Data

After I collected, examined, and cleaned the WeRateDogs Twitter data, I performed some analysis on the dataset to discover useful information.

Among all the information in the dataset, I focused on following specific features:

- retweet count
- favorite count
- dog type
- rate ratio
- dog age category
- dog breed
- image prediction algorithm.

This led me to the following insights and visualizations:

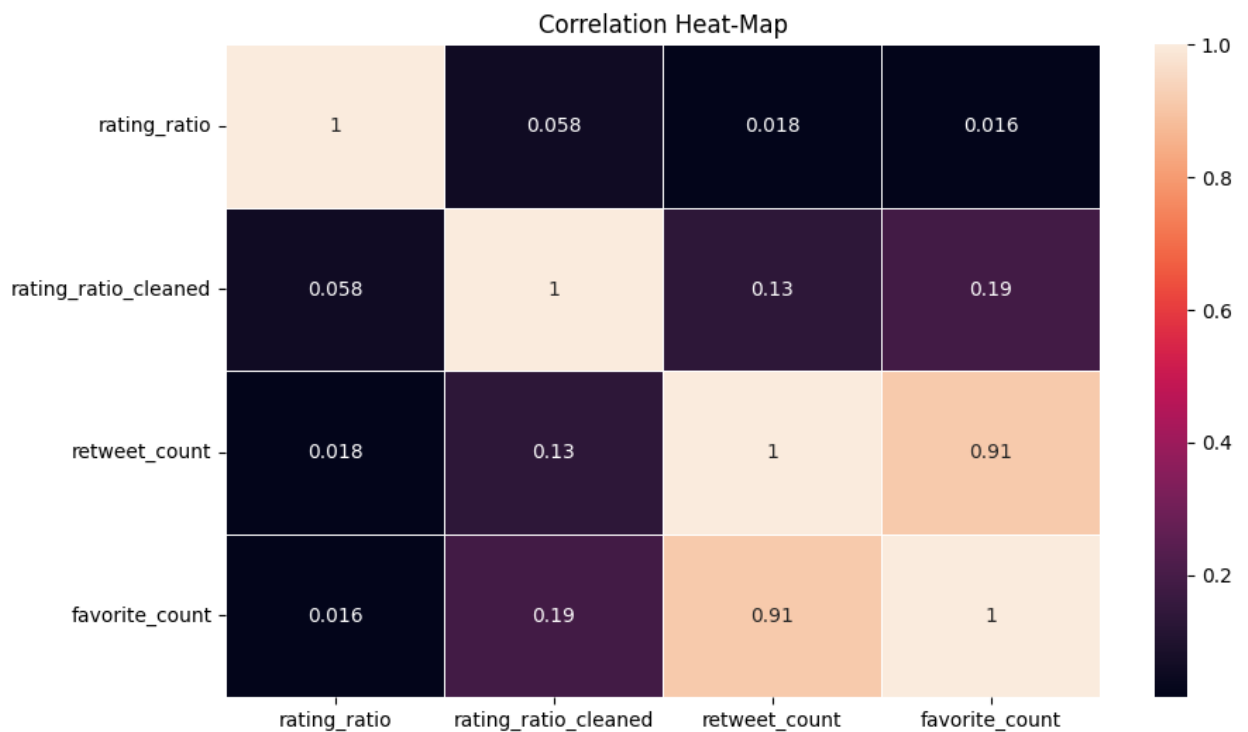
3. Insights

1. Is there any correlation between rating ratio, rating ratio cleaned and retweet or favorite?
2. What is the data partition based on Dog Age Category?
3. What is the most common Dog breed for p1 algorithm?
4. What is the rate of true dog picture upload?
5. Is the rate affecting by the picture is related to dog or not?

6. What is the most popular dog type?
 - i. By rating ratio
 - ii. By favorite count
 - iii. By retweet count
7. What is the most popular dog breed for p1 algorithm:
 - i. By rating ratio
 - ii. By favorite count
 - iii. By retweet count

4. Visualization

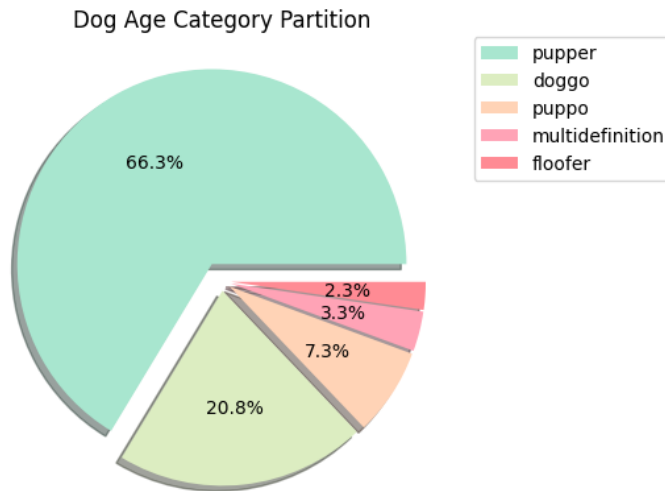
1. *Is there any correlation between rating ratio, rating ratio cleaned and retweet or favorite?*



Pic. 1: “Correlation Heat Map”

- There is a high correlation between retweet and favorite of tweet. It was expecting that.
- Also, there is a slight correlation between rating ratio and retweet and favorite. It was expected but low correlation was not expected.
- Cleaning of rating ratio affected the results.

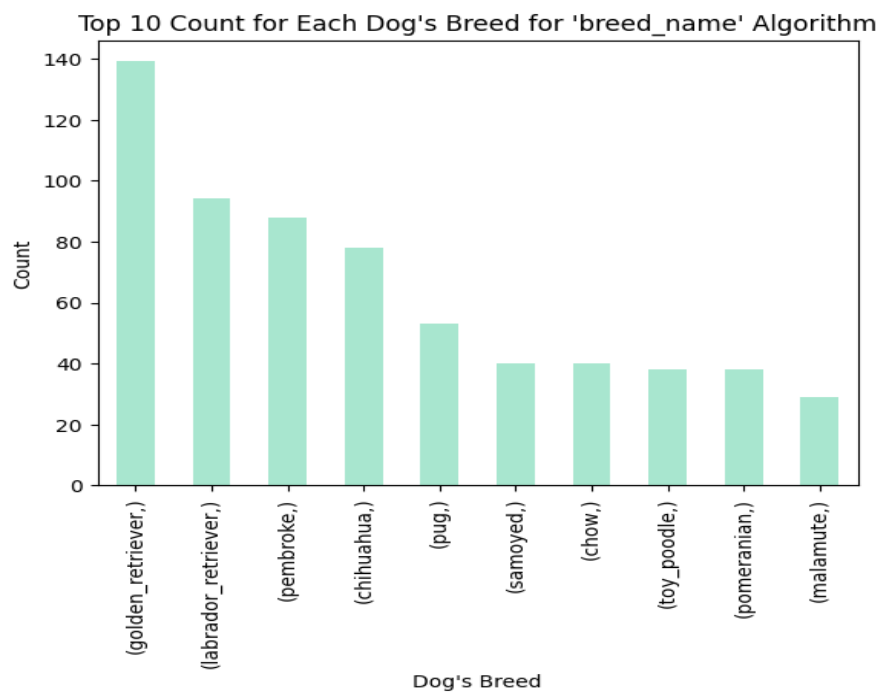
2. What is the data partition based on Dog Age Category?



Pic. 2: "Pie Chart – Dog Age Category Partition"

- 66.3% of dogs are in **pupper** stage.
- There is a much not categorized ("None") row. So, the result can be affected by these dropped rows.

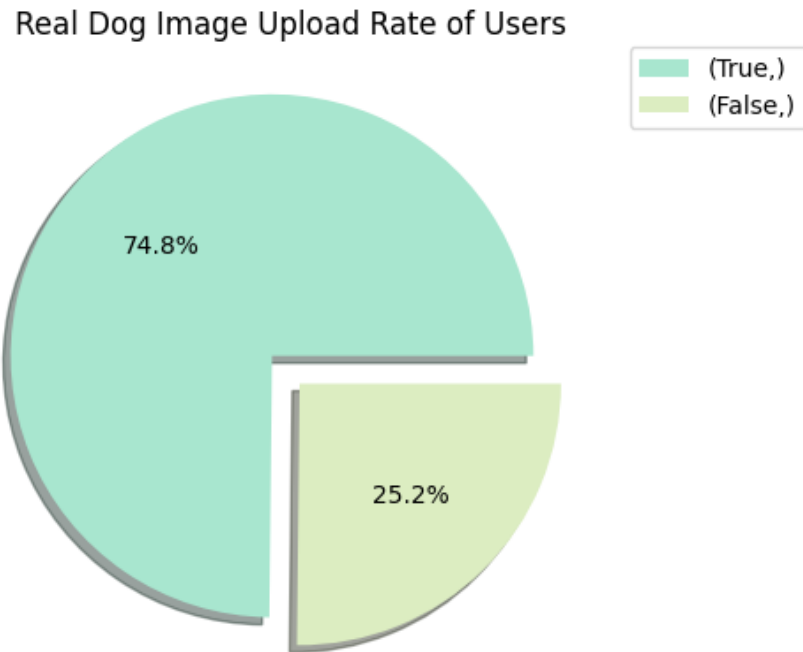
3. What is the most common Dog Breed for Breed Name?



Pic. 3: "Bar Chart – The most common Dog Breed Name"

- The most common dog breed is **Golden Retriever**.

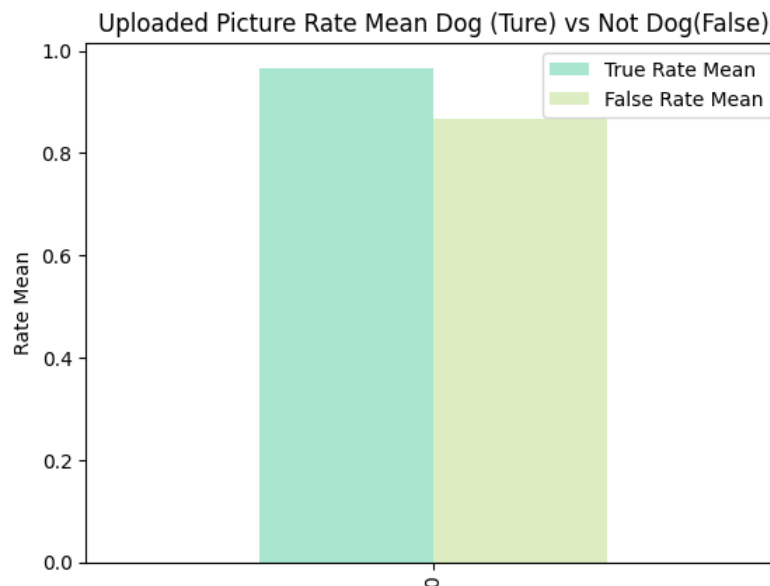
4. *What is the rate of true dog picture upload?*



Pic. 4: “Pie Chart – Real Dog Image Upload Rate of Users”

- 25.2% of uploaded pictures are not related to the dog breed.

5. *Is the rate affecting by the picture is related to dog or not?*



Pic. 5: “Bar Chart – Uploaded Picture Rate Mean Dog vs Not Dog”

- If users do not post dog pictures, they usually have a low score.

6. What is the most popular Dog Age Category Partition
- i. **By rating ratio?**

Table -1: “Describe Table for Dog Age Category by Rating Ratio”

	count	mean	std	min	25%	50%	75%	max
dog_category								
doggo	63.0	0.990476	0.042934	0.8	1.0	1.0	1.0	1.0
floofer	7.0	1.000000	0.000000	1.0	1.0	1.0	1.0	1.0
multidefinition	10.0	0.950000	0.158114	0.5	1.0	1.0	1.0	1.0
pupper	201.0	0.960199	0.098529	0.3	1.0	1.0	1.0	1.0
puppo	22.0	0.995455	0.021320	0.9	1.0	1.0	1.0	1.0

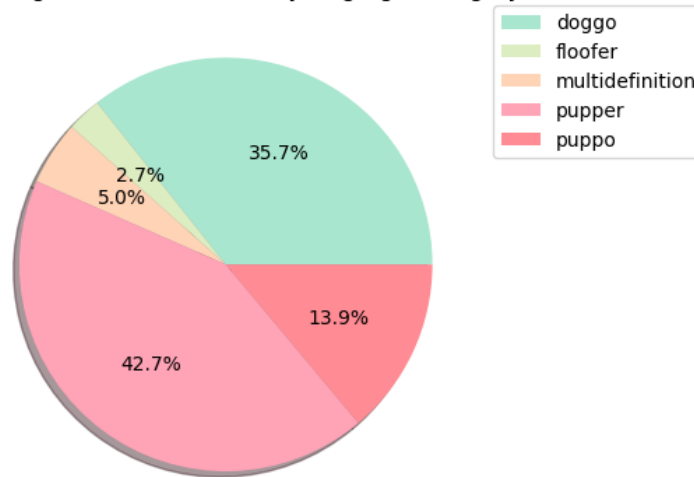
- All **floofers** have 10/10.
- Also, all dog age categories have high scores.

- ii. **By Favorite?**

Table -2: “Sum of Favorited Tweet by Dog Age Category”

	favorite_count
dog_category	
doggo	1219452
floofer	92442
multidefinition	170772
pupper	1457356
puppo	474806

Percentage of Favorite Count by Dog Age Category



Pic. 6: “Pie Chart – Percentage of Favorite Count by Dog Age Category”

Table -3: “Describe Table for Dog Age Category by Favorite”

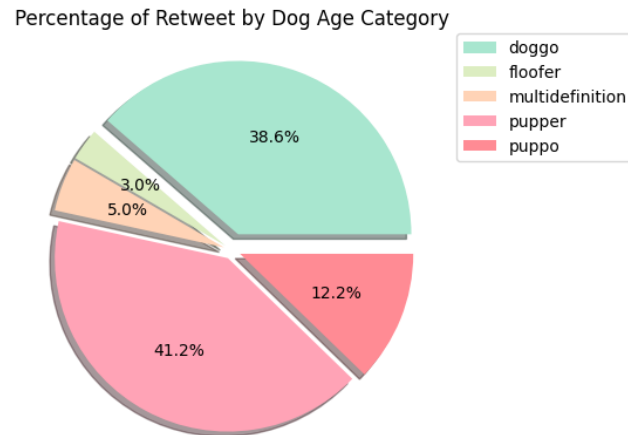
	count	mean	std	min	25%	50%	75%	max
dog_category								
doggo	63.0	19356.380952	22208.483825	2593.0	8268.00	12376.0	20502.50	131075.0
floofer	7.0	13206.000000	11258.315105	2262.0	5045.50	11302.0	17721.00	33345.0
multidefinition	10.0	17077.200000	15725.294965	4849.0	8799.25	10021.5	15852.00	47844.0
pupper	201.0	7250.527363	10743.931650	693.0	2449.00	3275.0	7842.00	106827.0
puppo	22.0	21582.090909	28034.339585	3277.0	7043.50	13387.5	19691.25	132810.0

- The most favored dog **pupper** age range. It got 42.7% of all favs. **Doggo** comes next with 35.7%.
- Each **doggo** gets an average of ~19.356 Favs while the **pupper** gets ~7250 favs. Therefore, each **puppo** gets an average of ~21.582 Favs.
- **Next Question:** Is there a significant difference between Doggo and Puppo?

iii. By Retweet?

Table -4: “Sum of Retweeted Dog Age Category”

	retweet_count
dog_category	
doggo	448919
floofer	34781
multidefinition	57807
pupper	478883
puppo	142427



Pic. 7: “Pie Chart – Percentage of Retweeted Count by Dog Age Category”

Table -5: “Describe Table for Dog Age Category by Retweet”

	count	mean	std	min	25%	50%	75%	max
dog_category								
doggo	63.0	7125.698413	12868.874516	725.0	2025.5	3327.0	5422.00	79515.0
floofer	7.0	4968.714286	6163.579715	496.0	1825.0	3265.0	4436.50	18497.0
multidefinition	10.0	5780.700000	6697.135524	1265.0	2453.0	2822.5	3498.25	19196.0
pupper	201.0	2382.502488	3584.033641	103.0	695.0	1219.0	2475.00	32883.0
puppo	22.0	6473.954545	10301.927187	716.0	1712.5	3130.0	6284.25	48265.0

- The most retweeted dog **doggo** age range (448.919 retweet). It got 41.2% of all retweets. **Doggo** comes next with 38.6%.
- Each **doggo** gets an average of ~7.125 retweets while the **puppo** gets ~6.473 retweet.
- **Next Question:** Is there a significant difference between **Doggo** and **Puppo**?

7. What is the most popular Dog Breed for Breed Name
i. **By Rating Ratio**

Table -6: “Describe Table for *Breed Name* by Rating for Dog False Categorized by Prediction”

	count	mean	std	min	25%	50%	75%	max
p1								
lacewing	1.0	0.1	NaN	0.1	0.1	0.1	0.1	0.1
electric_fan	1.0	0.1	NaN	0.1	0.1	0.1	0.1	0.1
coho	1.0	0.2	NaN	0.2	0.2	0.2	0.2	0.2
maze	1.0	0.2	NaN	0.2	0.2	0.2	0.2	0.2
scorpion	1.0	0.2	NaN	0.2	0.2	0.2	0.2	0.2

Table -7: “Describe Table for *Breed Name* by Rating for Dog True Categorized by Prediction”

	count	mean	std	min	25%	50%	75%	max
breed_name								
japanese_spaniel	1.0	0.500000	NaN	0.5	0.500	0.50	0.500	0.5
afghan_hound	3.0	0.766667	0.404145	0.3	0.650	1.00	1.000	1.0
soft_coated_wheaten_terrier	10.0	0.850000	0.246080	0.4	0.825	1.00	1.000	1.0
black_and_tan_coonhound	2.0	0.850000	0.212132	0.7	0.775	0.85	0.925	1.0
miniature_schnauzer	4.0	0.850000	0.238048	0.5	0.800	0.95	1.000	1.0

- As I mentioned before, when the pictures are not related to dogs, they get a low score.
- Example Second picture have 0.5 rate and it is not dog.
- Rate Ratio does not give specific information for dog breeds.



Picture -8: Prediction Breed: “Car Mirror”

ii. **By Favorite**

Table -8: “Sum of Favorited Dog Breed”

	favorite_count
breed_name	
golden_retriever	1729926
labrador_retriever	1113836
pembroke	994755
chihuahua	674072
samoyed	531052

Table -9: “Maximum of Favorited Dog Breed”

breed_name	favorite_count
lakeland_terrier	132810

Table -10: “Describe Table for Dog Breed by Favorite”

	count	mean	std	min	25%	50%	75%	max
breed_name								
saluki	4.0	24060.000000	20883.171518	3444.0	10886.25	20637.0	33810.75	51522.0
french_bulldog	27.0	17468.370370	21899.214990	341.0	3356.50	9726.0	24385.00	106827.0
afghan_hound	3.0	17326.666667	10259.517159	7078.0	12191.50	17305.0	22451.00	27597.0
giant_schnauzer	2.0	17314.500000	24154.060539	235.0	8774.75	17314.5	25854.25	34394.0
black and tan coonhound	2.0	17012.000000	23898.794991	113.0	8562.50	17012.0	25461.50	33911.0

- The most favorited dog is **golden retriever** (1.729.926). Therefore, the rate of meet a **golden retriever** picture on twitter is very high.
- The average favorite dog breed is the **saluki** (24.060 Fav). But there are only 4 **saluki** breed tweeted.
- The **lakeland terrier** is the most favorite dog breed 132.810 times. (In one tweet)

iii. By Retweet

Table -11: “Sum of Retweeted Dog Breed”

	retweet_count
breed_name	
golden_retriever	520591
labrador_retriever	364288
pembroke	280367
chihuahua	219851
samoyed	185856

Table -12: “Maximum of Retweeted Dog Breed”

breed_name	retweet_count
labrador_retriever	79515

Table -13: “Describe Table for Dog Breed by Retweet”

	count	mean	std	min	25%	50%	75%	max
breed_name								
standard_poodle	7.0	6631.857143	15703.847760	63.0	338.50	888.0	1283.50	42228.0
afghan_hound	3.0	5976.000000	4452.291208	1893.0	3602.50	5312.0	8017.50	10723.0
english_springer	10.0	5401.600000	14251.566613	106.0	284.50	485.0	1783.50	45849.0
giant_schnauzer	2.0	5161.500000	7162.284587	97.0	2629.25	5161.5	7693.75	10226.0
saluki	4.0	5133.750000	4365.237021	1011.0	2511.00	4212.5	6835.25	11099.0

- The most retweeted dog is **golden retriever** (520.591). There is a very high probability of encountering a **golden retriever** image on twitter.
- The average retweeted dog breed is the **standard poodle** (~6631 Retweeted).
- The **labrador retriever** is the most retweeted dog breed 79.515 times. (In one tweet)

5. Further Analysis

- Analysis can be conducted by date column.
- Also the comments can be perform by statically.