

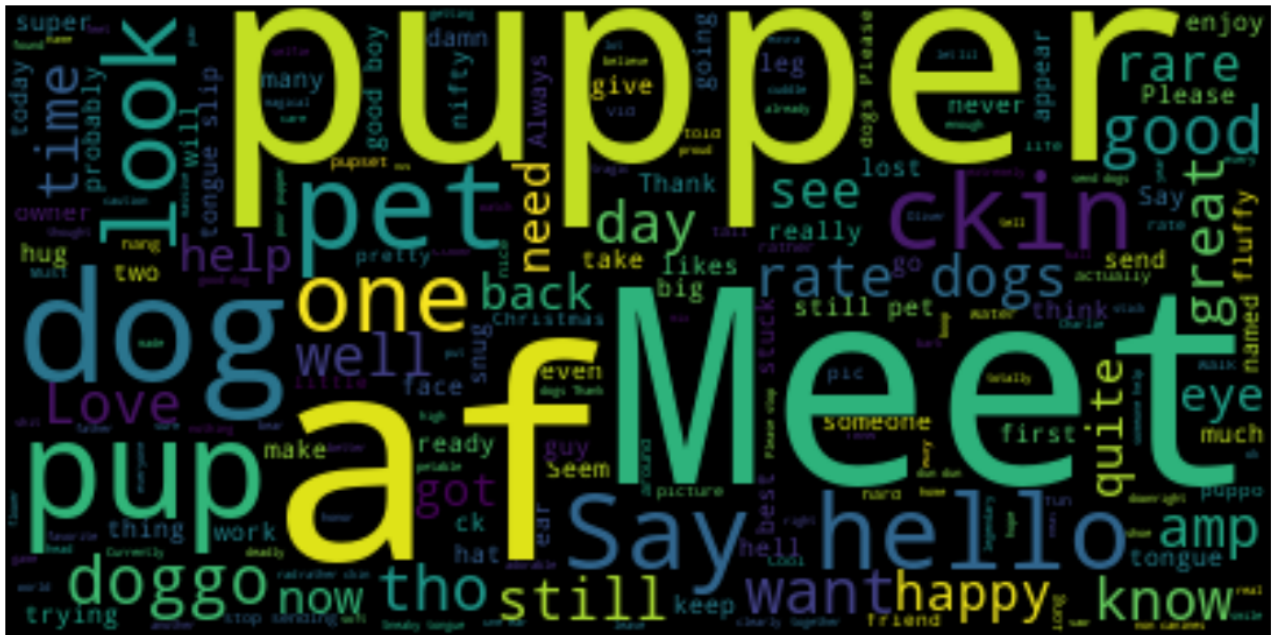
---

# WeRateDogs

## Analysis Report

Olga Timofeeva - 9 November 2020

---



## Analysis Foreword and Limitations

After the master data set was thoroughly assessed and cleaned as per tidy data criteria, the analysis can be started.

It is important to mention that some worthy data were removed during merging process from both image prediction data set and the twitter archive:

- 123 tweets in the twitter archive do not have image predictions, as mainly they contain videos. That was re-assessed manually.
- 112 tweets in image prediction data set are non-existent tweets that were removed from Twitter.
- Of course it could be another step of data cleaning process but this is limitation of my implementation for this project iteration due to the submission deadline.

---

As well as this, it would be interesting to analyse image prediction neural network but it will be mostly omitted due to the time constraint for this project.

Note: **all conclusions made in this report were based on descriptive statistics.**

Finally, there is a plotting function `plot_categorical_data` used in section below. Its parameters are well documented in the beginning of this notebook where the function is declared.

## Descriptive Statistics Analysis

### Descriptive Statistics using Pandas `describe()` Method

The basic statistics on quantitative columns were produced using pandas `describe()` method:

	tweet_id	rating_decimal	retweet_count	favorite_count	img_num	p1_conf	p2_conf	p3_conf
count	1.963000e+03	1963.000000	1963.000000	1963.000000	1963.000000	1963.000000	1.963000e+03	1.963000e+03
mean	7.357648e+17	1.165738	2406.321956	8144.520632	1.201223	0.593738	1.347014e-01	6.031649e-02
std	6.742179e+16	4.096079	4300.856047	11989.633678	0.558283	0.272073	1.009312e-01	5.096577e-02
min	6.660209e+17	0.100000	11.000000	69.000000	1.000000	0.044333	1.011300e-08	1.740170e-10
25%	6.758342e+17	1.000000	538.000000	1745.500000	1.000000	0.362716	5.370120e-02	1.622240e-02
50%	7.087111e+17	1.100000	1165.000000	3679.000000	1.000000	0.587342	1.174020e-01	4.948690e-02
75%	7.875578e+17	1.200000	2745.000000	10136.000000	1.000000	0.846807	1.955655e-01	9.164355e-02
max	8.924206e+17	177.600000	75647.000000	152777.000000	4.000000	1.000000	4.880140e-01	2.734190e-01

**Insight 1:** Following columns contain data that highly varies and statistics suggests data in those columns does not have normal distribution, and therefore WeRateDogs rating system may be inconsistent and skewed:

- rating (`rating_decimal` column)
- retweet\_count
- favorite\_count

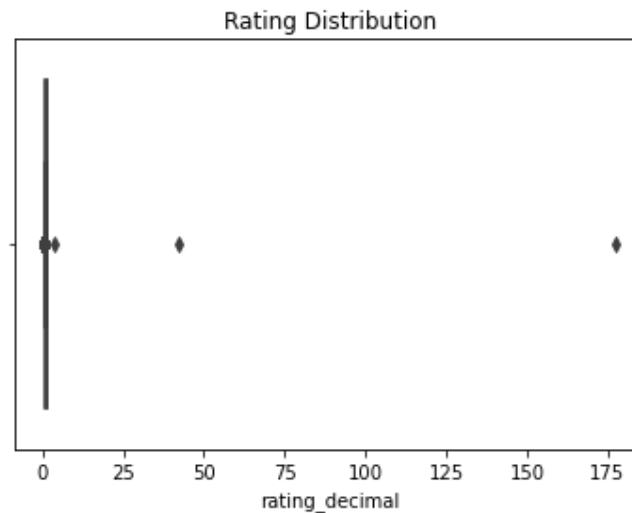
This insight will be explored further.

First, let's have a look at the box\_plot of `rating_decimal` column to gauge how skewed WeRateDogs rating system can be.

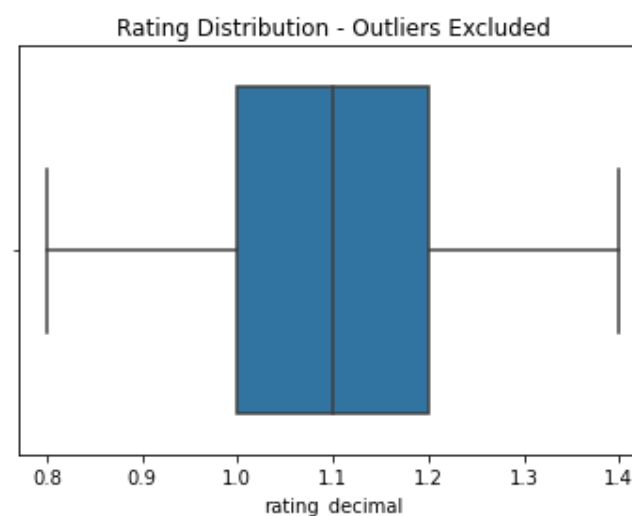
---

## Rating analysis and visualisation¶

The box plot below confirms that there are few outliers that are wildly exaggerated from the majority of WeRateDogs tweets. All dogs are good of course, however when raters get carried away with ratings, it makes rating system unreliable.



Now we will have a look what are the ratings for the tweets that are not outliers:



### Insight 2:

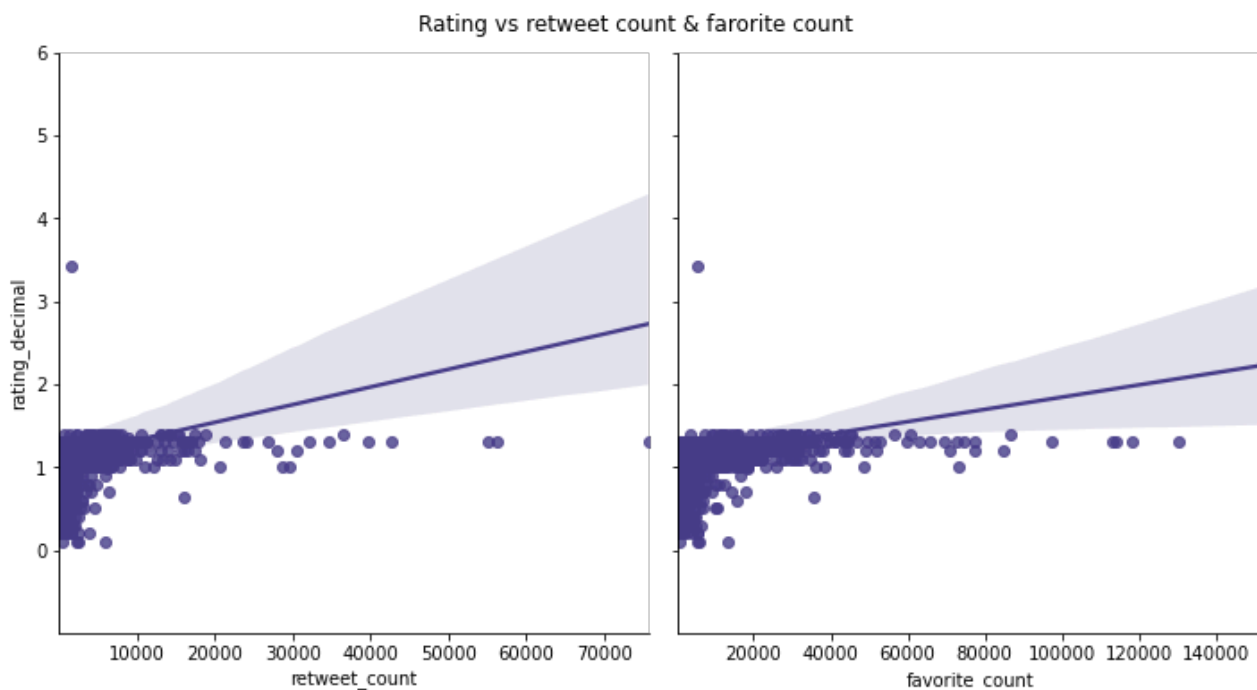
- WeRateDogs rating system is inconsistent and could be very subjective.
- The hypothetical rating normal distribution suggests that majority of tweets were rated with rating numerator to rating denominator ratio between 1.0 and 1.2.

Next we will assess some interesting correlations in the master data set.

---

## Can rating influence retweet and favourite count?

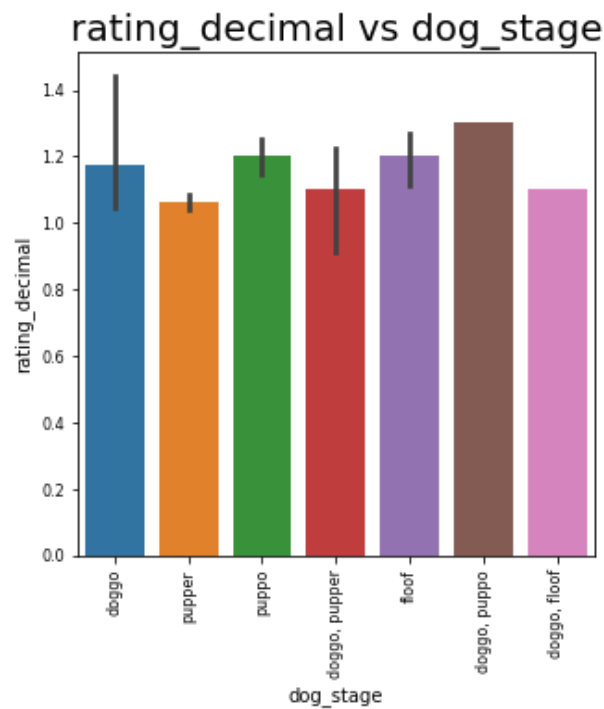
The below plot demonstrates correlation between rating and retweet count as well as correlation between rating and favorite count:



**Insight 3:** There is weak positive correlation between rating and retweet count as well as with favourite count. However, there is no strong evidence that ranking on a tweet can drive retweets and favourite counts.

## Rating and dog stage correlation

Now we will create bar plot to analyse if dog development stage (dog\_stage) can drive higher ratings (rating\_decimal):

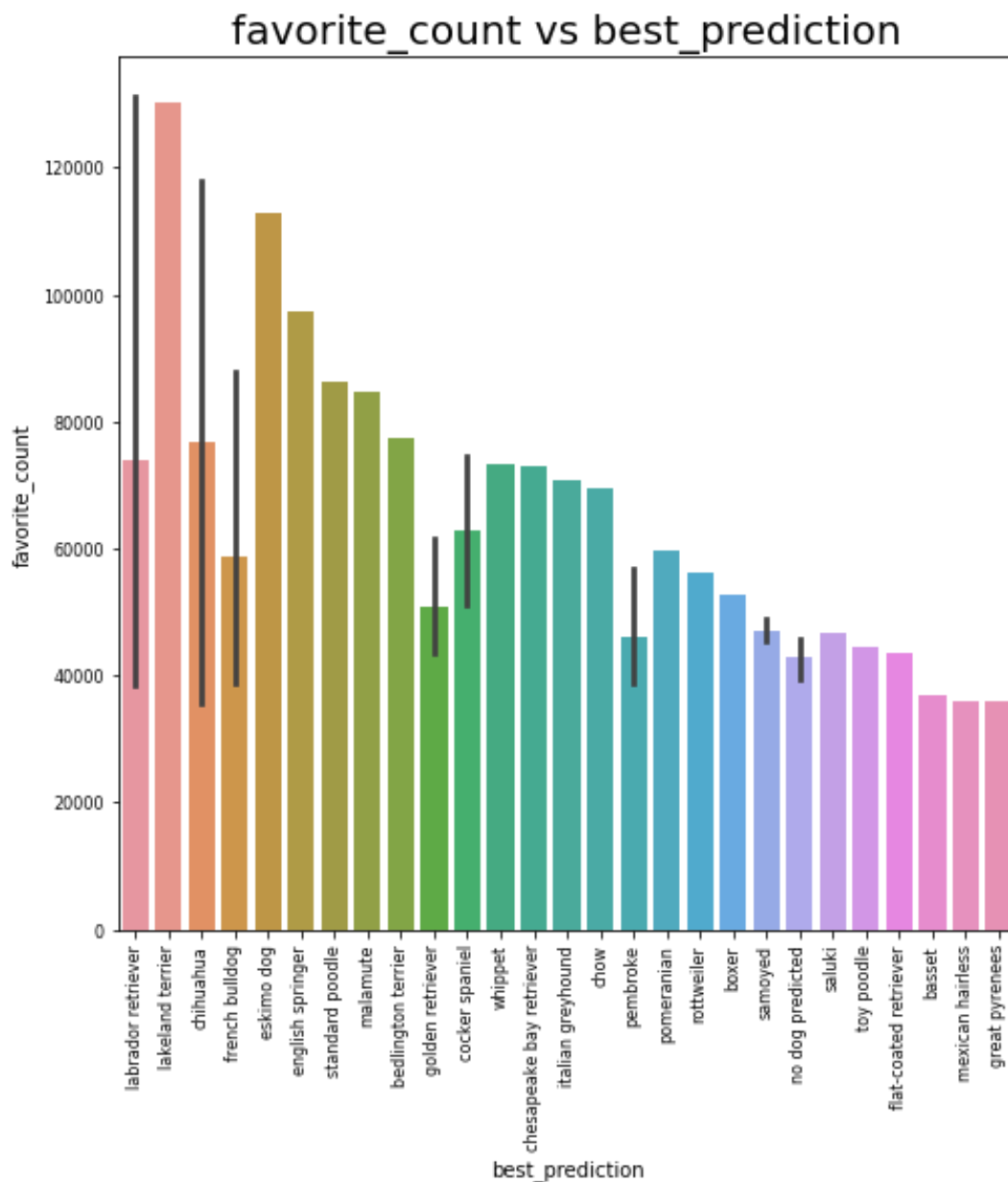


**Insight 4:** There is no strong evidence that dog development stage influences rating.

**What are the most retweeted and likeable dog breeds in WeRateDogs archive?**

#### Top 50 most likeable dog breeds

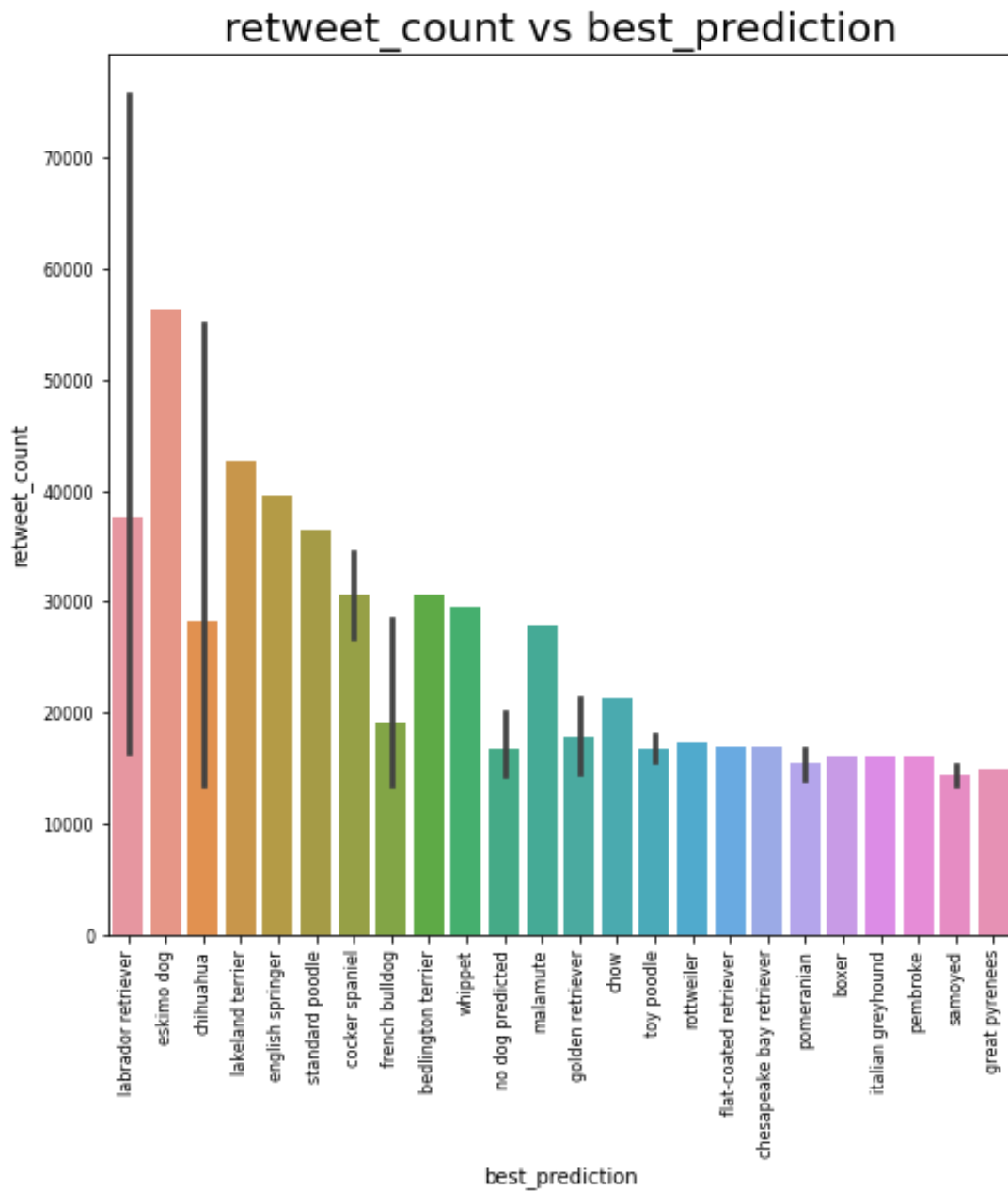
Next visualisation will be correlation between predicted dog breed (best\_prediction) and favourite count. Note that due to wide range of breeds, only top 50 by favourite\_count records were selected for plotting:



**Insight 5:** Interestingly, Labrador retriever is the most likeable dog breed, however the plot error bars suggest that this breed also has very high standard deviation for favorite count.

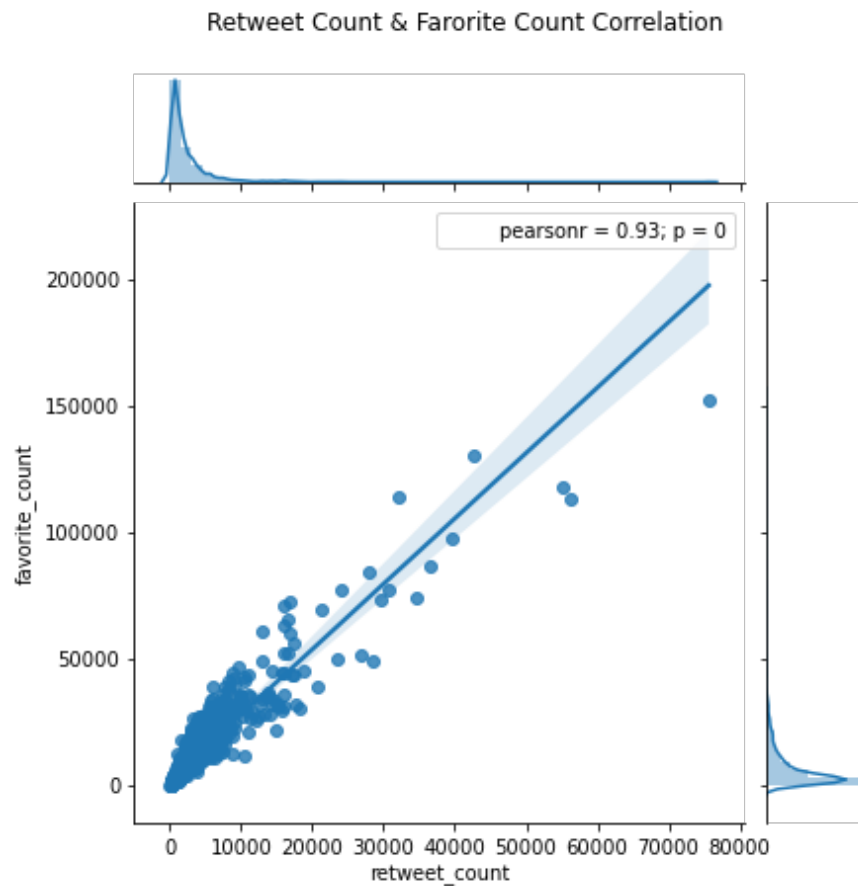
### Top 50 most retweeted tweets

Now we will see what are the predict dog breeds on tweets that were retweeted the most. We only take 50 tweets with highest retweet count for plotting:



We see very similar correlation between dog breed and retweets number as between dog breed and favorite count.

Let's see if favorite count and retweet count are correlated:

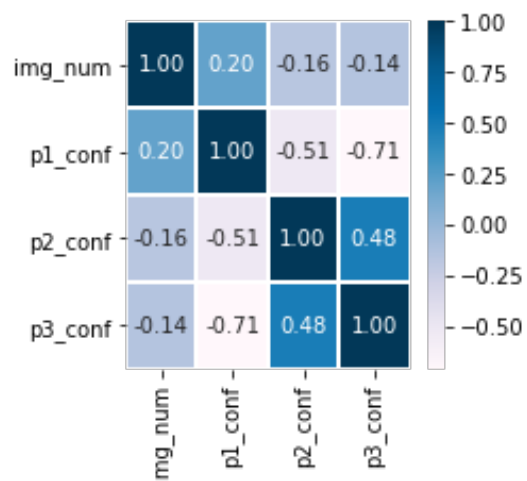


**Insight 6:** Favorite count and retweet count are highly positively correlated with 0.93 Pearson coefficient.

## Analysis of image predictions

### Correlation between prediction coefficients and number of images (img\_num)

Prediction Coefficients & Number of Images Correlation



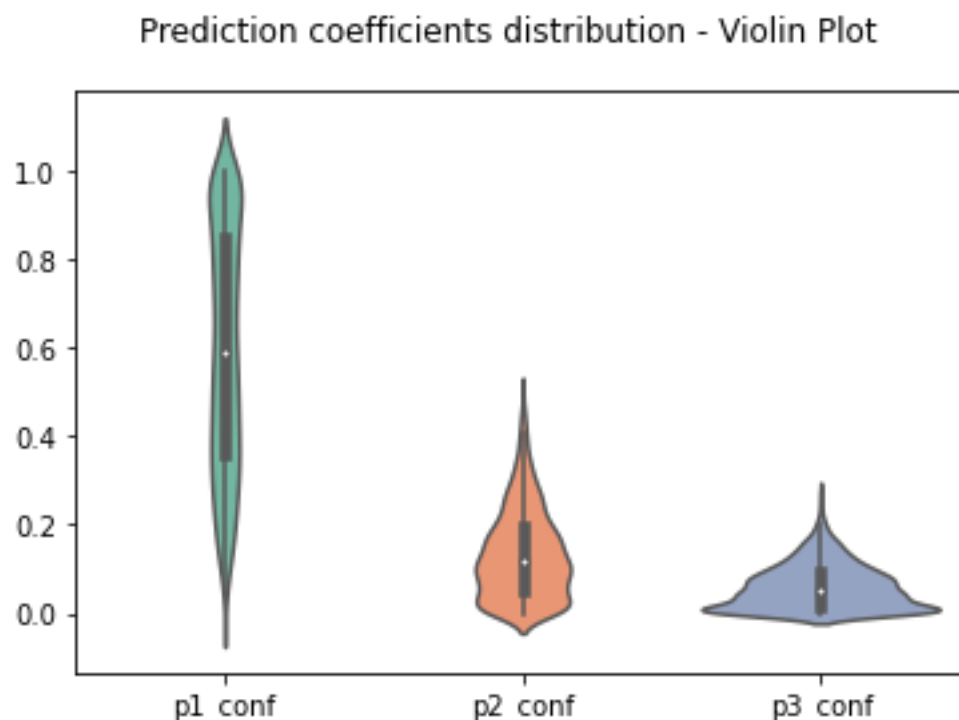


---

**Insight 7:** There is no strong correlation between `img_num` and prediction confidence coefficients. Assumption: the image prediction neural network might need more training.

### Is the image prediction neural network getting more stable with increasing number of trials?

Next visualisation will help to see distribution of three prediction coefficients. This will help to understand whether predictions get more stable with number of predictions.



#### Insight 8:

- first prediction: has much higher prediction range, although probability of these predictions are much narrower than p2 and p3 predictions
- second prediction: has higher probability that images in tweets will be predicted with p2 value
- third prediction has the highest probability that images will take on the given value in p3

This means that neural network is getting more stable with increasing number of trials. However, this analysis did not take quality of the prediction into account. As seen during assessment stage, some predictions with high confidence close to 1 predicted not a dog but other object on the supplied image.

---

## Is number of trials improve prediction of a dog breed?

Let's assess if false dog predictions are reduced on third trial compared to second and first trial. Data in the table below was gathered in the supplied Jupiter notebook.

Count	Prediction 1	Prediction 2	Prediction 3
Dog Predicted (True)	1456	1472	1423
Dog Not Predicted (False)	507	491	540

*This statement is based purely on descriptive statistics and shall be verified with a reliable model:*

### Insight 9:

No, as per descriptive statistics above there is no evidence that the neural network is improving on recognising dogs when first prediction could not recognise dogs on a given image.

## Conclusion

- WeRateDogs indeed uses a very unique rating system. This rating system however is not driven by specific factors in the data set but rather subjective to:
  - how cute dog on the tweet is, or
  - is the picture emotionally moving or
  - personal opinion of a person who gives rating
- Interestingly, a Labrador retriever got the most likes and retweets
- Neural network that was used for predicting dog breeds might need more training as number of images where dog was not determined did not decrease with number of trials.
- A more sophisticated image recognising neural network would need to predict not only what dog breed is on a picture but also what makes the picture so likeable. This could help to predict what kind of pictures can get better engagement with the audience.