

WE RATE DOGS - Data ANALYSIS

INTRODUCTION

Here, I am going to summarize my analysis done in the project notebook.

The dataset that I analysed is the tweet archive of Twitter user [@dog_rates](#), also known as [WeRateDogs](#). WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because ["they're good dogs Brent."](#) WeRateDogs has over 4 million followers and has received international media coverage.

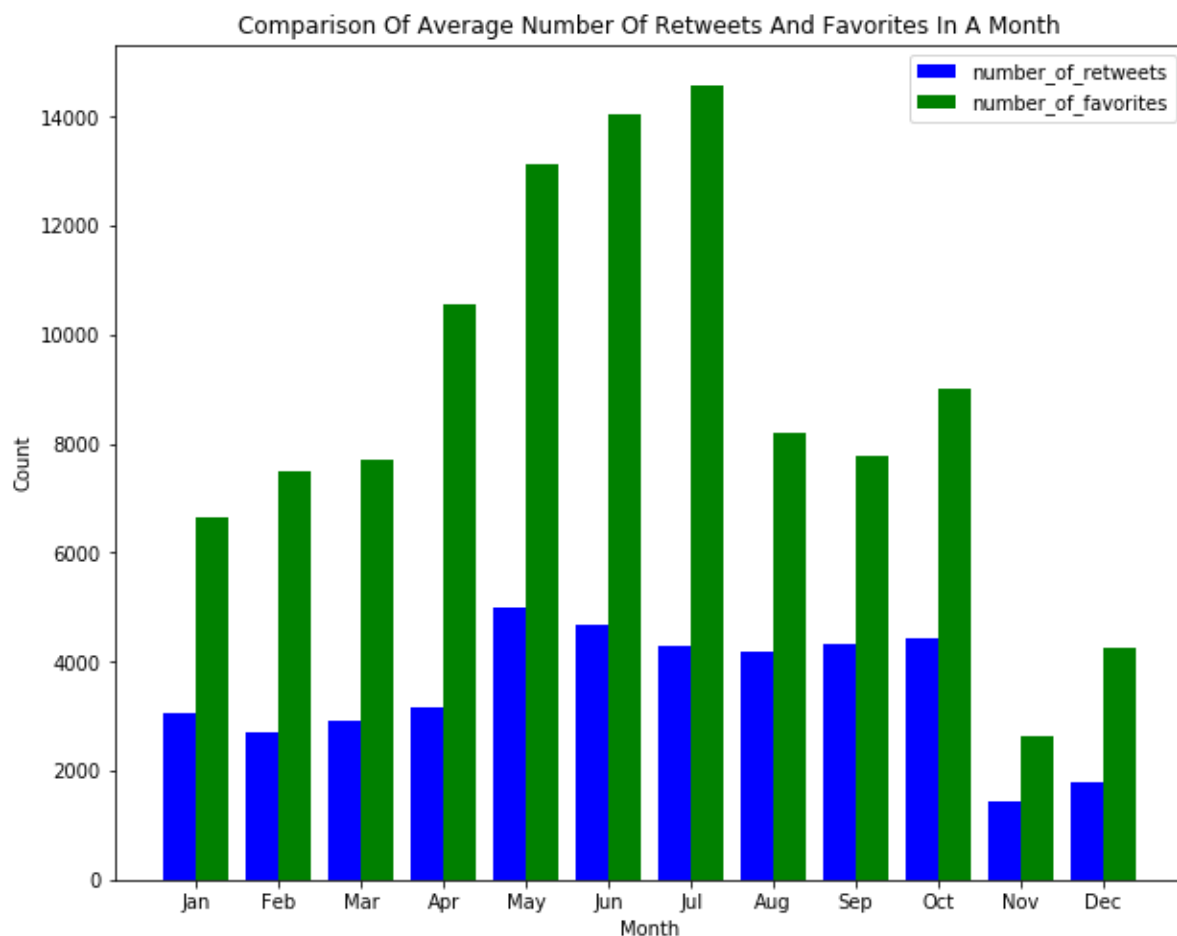
ANALYSING THE DATA

Here, I have analysed the dataframes `tweet_data_clean` and `img_pred_clean` from the report after storing them into two separate .csv

files `twitter_archive_master.csv` and `image_predictions_master.csv`

I have analysed the data on 3 different questions of interest

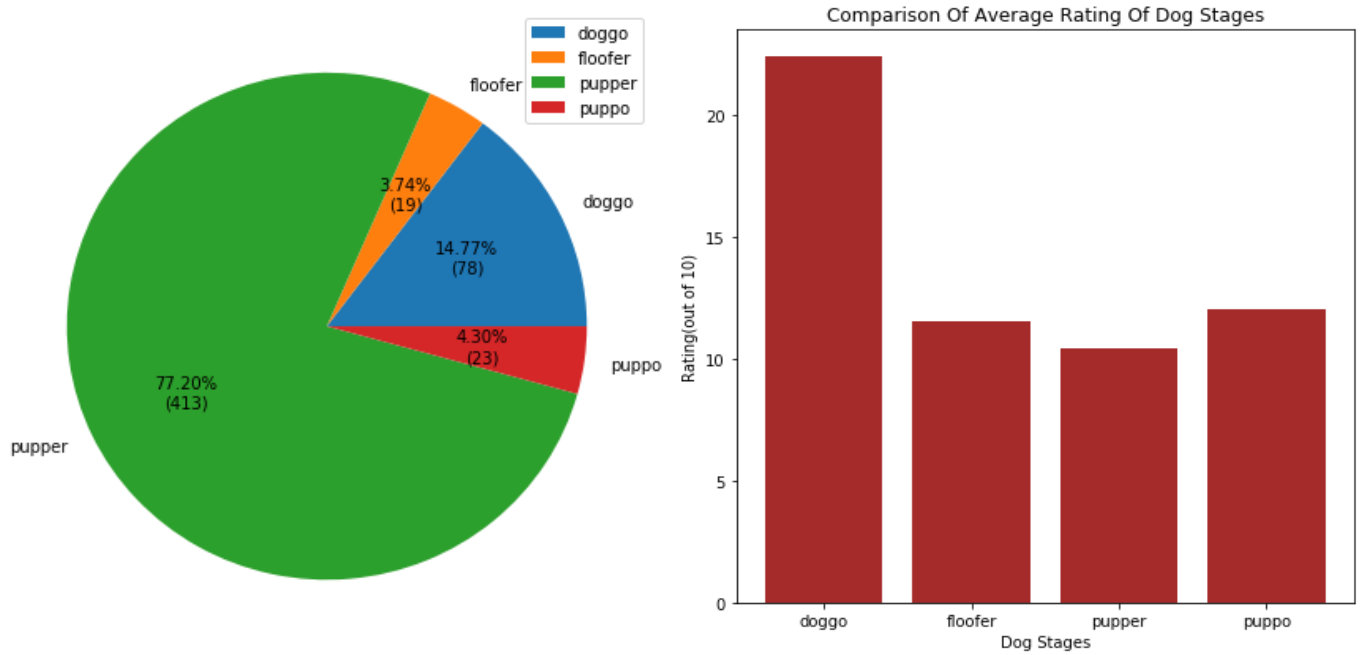
1. Comparison Of Average Number Of Retweets And Favorites In A Month



Observations:

1. As we can see that on an average, number of favorites are quite higher than number of retweets.
2. July has the highest average number of favorites and November has the lowest average number of favorites.
3. May has the highest average number of retweets and November has the lowest average number of retweets.
4. June has the second highest both average number of favorites and retweets.

2. Comparison Of Average Rating Of Dog Stages



Observations:

1. From the bar graph, it can be clearly seen that doggo has the highest average rating while pupper has the lowest.
2. From the pie chart, we can see that there are 77.2% puppers in the overall population of the dataset followed by doggo, then floofer and at the last, puppo. This indicates the possible reason of lowest average rating of pupper and highest average rating of doggo because of huge differences in their proportion in the dataset

3. Algorithm's First Prediction Efficiency for top 10 most frequent predictions

	prediction_name	prediction_total	prediction_correct	prediction_efficiency(in %)
0	Golden retriever	139	116	83.453237
1	Labrador retriever	95	65	68.421053
2	Pembroke	88	70	79.545455
3	Chihuahua	79	47	59.493671
4	Pug	55	44	80.000000
5	Chow	41	26	63.414634
6	Samoyed	40	30	75.000000
7	Pomeranian	38	29	76.315789
8	Toy poodle	38	24	63.157895
9	Malamute	29	18	62.068966

Observations for Algorithm's First Prediction Efficiency for top 10 most frequent predictions:

1. In the above table we can clearly see that the algorithm has proved to be most efficient for Golden Retrievers.
2. Algorithm has proved to be least efficient for Chihuahuas.
3. The most predictions are recorded for Golden Retrievers and least for Malamutes

Therefore, we can say that, this algorithm favors the Golden Retrievers in this dataset