act_report

September 28, 2022

0.1 Report: act_report

0.2 Analysis and Insights

After I had adequately dealt with the quality and tidiness issues documentated in the accessing phase of the project, I then posed some questions whose answers I used in analysing the cleaned master dataset. These questions are;
* What is the relationship of the dog stages to favorite count and retweet count?
* Which are the most ranked and least ranked dog stages?
* What are the success rate of the algorithm predictions.

0.2.1 What is the Relationship of Dog stages to favorite count and retweet count.

The number of Favorites/Likes and retweets a tweet gets is an indication of the level of engagement the tweet had and to a large extent is a pointer to whether or not people like the tweet. Therefore, my first question seeks to establish relationship between the dog stages and number of likes and retweets every individual tweet gets.

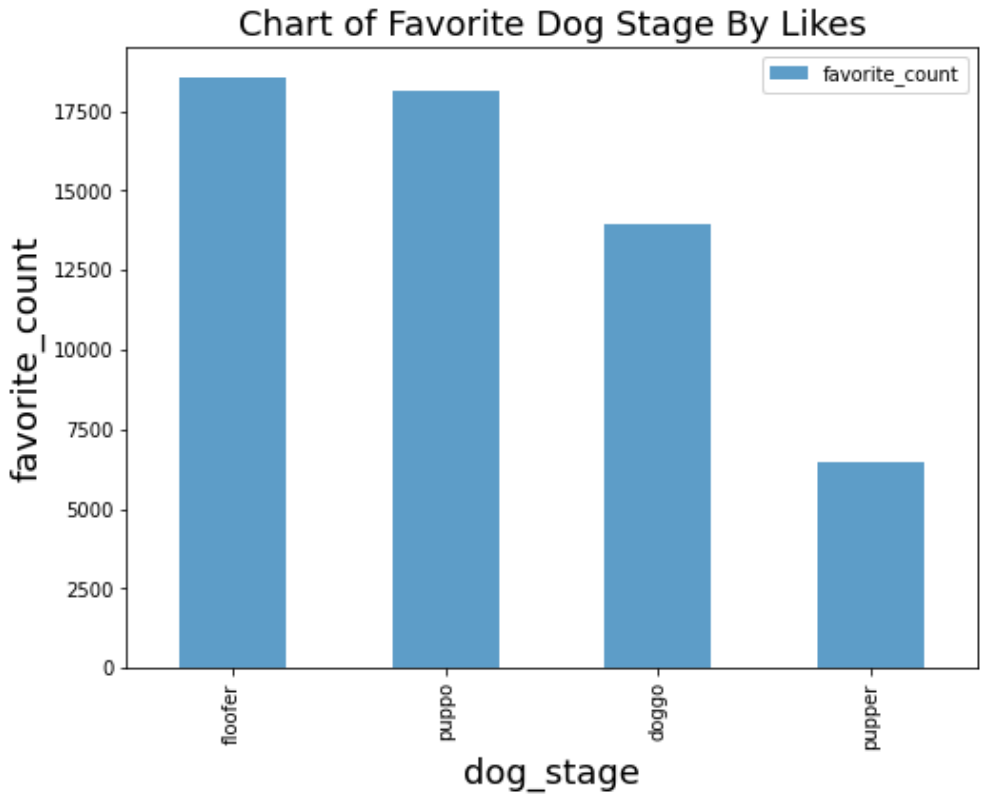


Chart of Favorite Dog Stage By Likes
favorite_count
17500
15000
12500
10000
7500
5000
2500
0
floofer
puppo
doggo
pupper
dog_stage

,

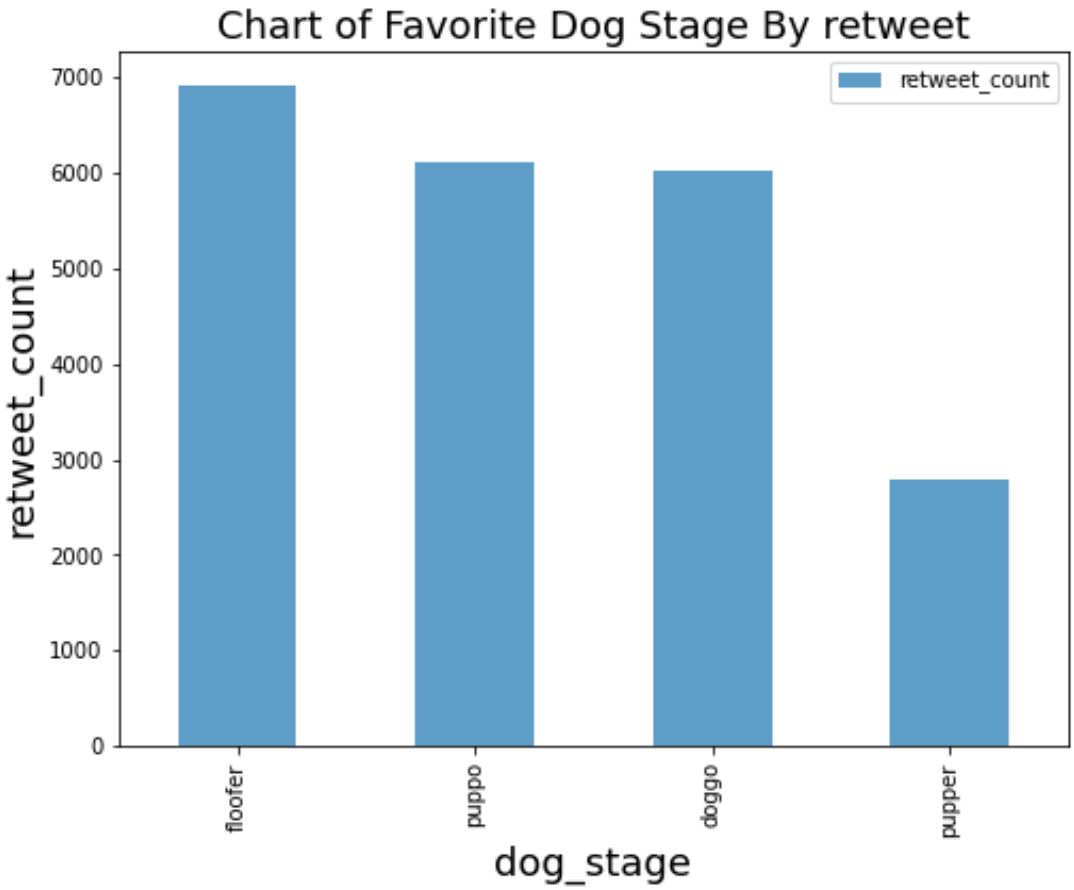
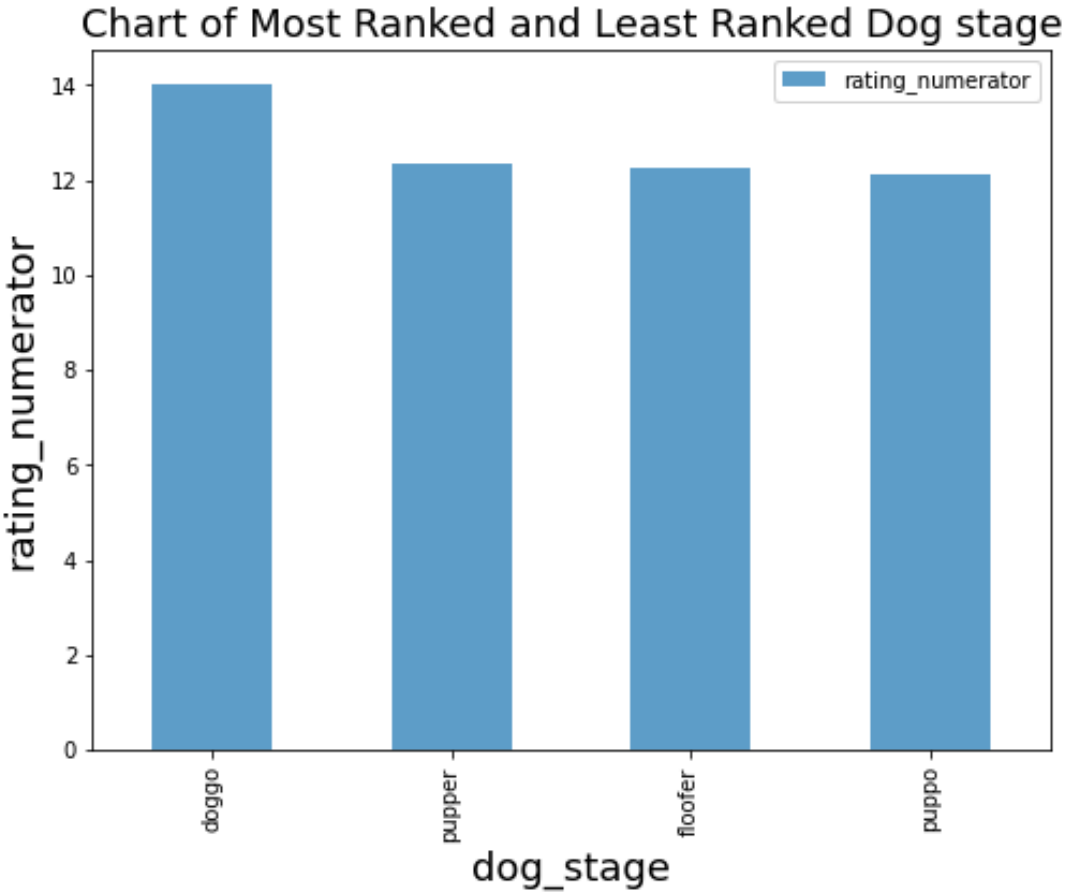


Chart of Favorite Dog Stage By retweet
retweet_count
7000
6000
5000
4000
3000
2000
1000
0
floofer
puppo
doggo
pupper
dog_stage

Insight. From the two charts above there is a clear positive corelation between the Dog stage and the number of likes and retweets a certain tweet gets. Puppo got more likes and retweets and this would infer that people like the teenage version of dogs the most. I will suspect that is because at this stage they are not too needy and also do not cause too much trouble. while pupper got leat likes and retweets and this could be due to the fact that these are smaller dogs and are more difficult to deal with. Floofer understandly got highest retweets on the second chart, this however doesnt change the fact that Puppo is also the most retweeted dog stage. This is because floofer simply means furry dogs.

0.2.2 Which Are The Most Ranked and Least Ranked Dog Stage

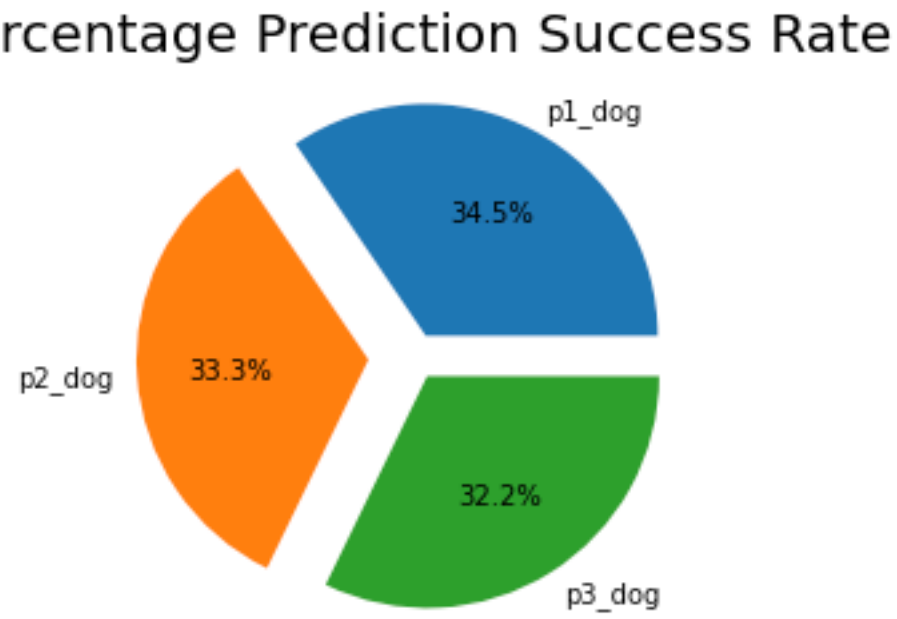
There are 4 dog stages according to WeRateDogs Twitter archive, doggo, pupper, floofer and puppo. I wanted to know which of the dog stages got higher ranking from the Twitter account. This I accomplished by plotting rating numerator against dog stage.



Insight The result of this chart shows that the people that engage with the WeRateDogs twitter page almost always have a contrary opinion of its rating of the dogs. This is because although puppo dog stage get more likes and retweets as seen on the previous analysis, doggo dog stage is the most ranked by the twitter handle, while puppo is the least ranked.

0.2.3 What are the success rate of the algorithm predictions?

Every image in the WeRateDogs Twitter archive was run through a neural network that can classify breeds of dogs. The resulted in a table full of image predictions (the top three only) alongside each tweet ID, image URL, and the image number that corresponded to the most confident prediction (numbered 1 to 4 since tweets can have up to four images). So I did an analysis to find out the accuracy of the predictions.



Insight From the analysis above, it is seen that the algorithm's predictions determining if image tweeted is a dog are mostly correct. However, the first predictions of the algorithm are more accurate with **77%** accuracy and this is shown in the pie chart with p1_dog taking the biggest pie.

0.2.4 Conclusion

The goal of this project was wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. This project exhaustively covers all the data wrangling processes; gathering, accessing and cleaning. A lot more insights can be drawn from the data provided, but that will involve more data cleaniing and manipulation. Therefore my findings are limited aand also tentative as it will require more in-depth analysis to give more concrete insights of the data.

References

- Stackoverflow

- Github
- Kaggle
- ALX-T Udacity Class lessons