Act Report

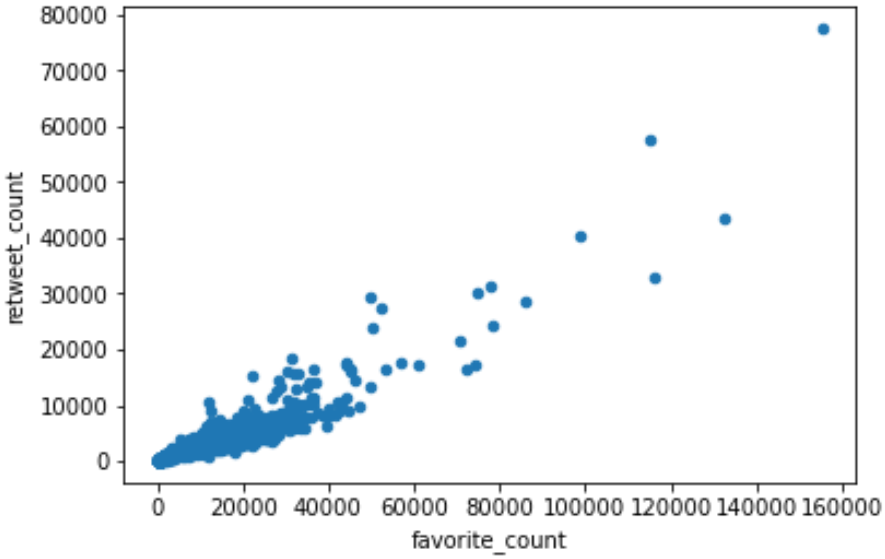
Introduction:

In this project I'll analyze and visualize the cleaned WeRateDogs Twitter Archive dataset (including image predictions and additional data -retweets and favorites count- from Twitter API), I'll focus in this analysis on 4 things: retweet and favorite count, dog stages, algorithm's prediction for the dog image and dog names.

Insight 1: retweets and favorites count

	retweet_count	favorite_count
count	1922.000000	1922.000000
mean	2223.373569	7685.811655
std	3932.532363	11315.504758
min	11.000000	70.000000
25%	529.250000	1692.750000
50%	1134.500000	3517.000000
75%	2535.000000	9289.000000
max	77464.000000	155507.000000

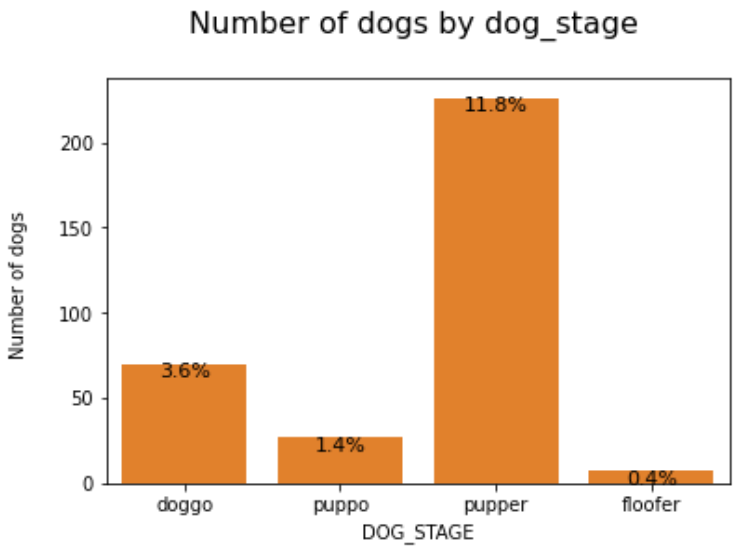
According to the summary statistics above we have the min values of retweet and favorite count respectively are (11, 70) and the max values are (77464, 155507) which may suggest a correlation between them, but let's confirm with a scatter plot:



Looking at the scatter plot above we can see clearly that the retweet count and favorite count are positively correlated, when favorite count increase retweet count increase too.

Insight 2: dog stages

In this insight I'll focus only on the tweets having the dog stage value available.

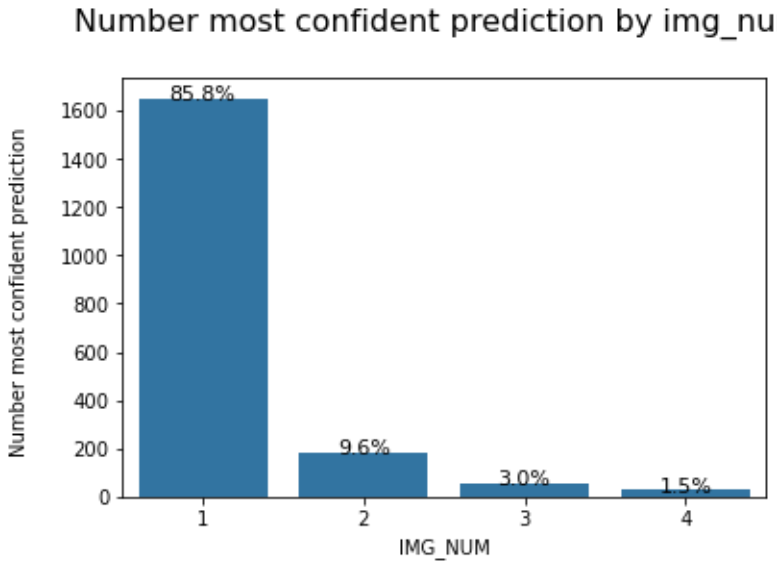


According to this bar count we can say that the dog more mentioned in the text tweet is in the Pupper stage having the highest proportion 11.8%.

```
dog_stage
doggo      1.158571
floofer    1.200000
pupper     1.059969
puppo      1.214815
```

After grouping the rating ratio's mean for each dog stage, we have the result above which shows that the dog stage having the highest rate is Puppo.

Insight 3: algorithm's prediction for the dog image

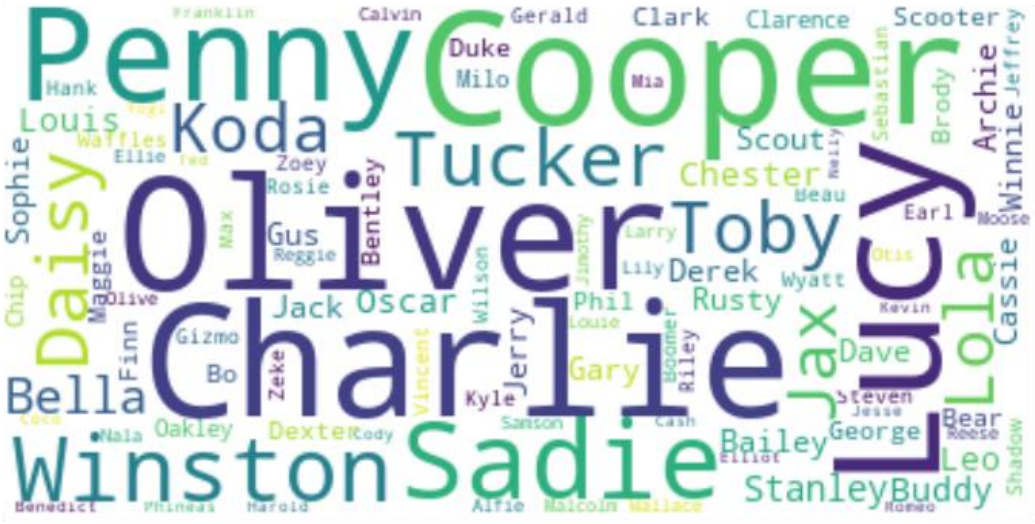


From the chart above always the 1st algorithm has the most confident prediction of dog breeds which is 85.8% of all tweets gives the most confident prediction of dog breeds.

Since the 1st algorithm is more confident I'll based on it to make a word cloud for see the different detected dog breeds.

According to this word cloud it's clear that the most 5 frequent dog breeds in our dataset are: Golden Retriever, Labrador Retriever, Pembroke, Chihuahua and Pug.

Insight 4: dog names



Again, using the word cloud of the available dog names, we find that the most 5 common dog names in our dataset are: Charlie, Cooper, Oliver, Lucy and Penny.

Conclusion: I know that my work may contain errors or mistakes, but I tried and worked hard to make something best, and I will be very happy to get your comments and recommendations.