Act Report

After assessing and cleaning the tweeter archive data, the image predication data, and the data obtained from the API, several questions needed to be addressed with respect to the data.

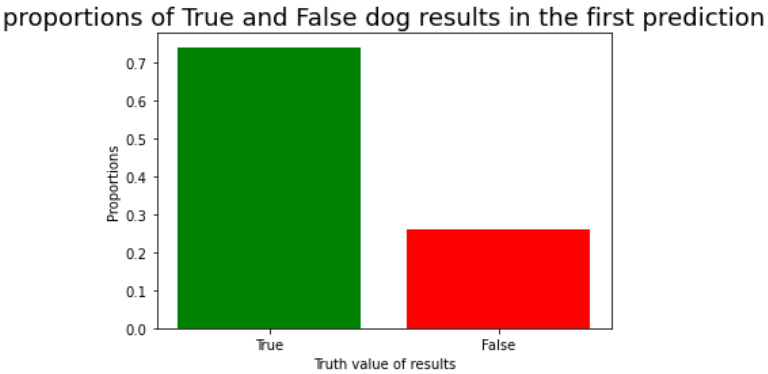
Questions that needed to be responded to:

1. What is the proportion of image preditions that gave a dog image in the first, second, and third prediction? How do these proportions compare against each other?
2. Compare the population of dogs in different stages.
3. Is there a relationship between Retweet count and favourite (like) count?

To effectively respond to these questions the tweeter archive dataset was analysed using python programming language. Below are the insights obtained from the data.

Question 1A: What is the proportion of image predictions that gave a dog image in the first prediction?

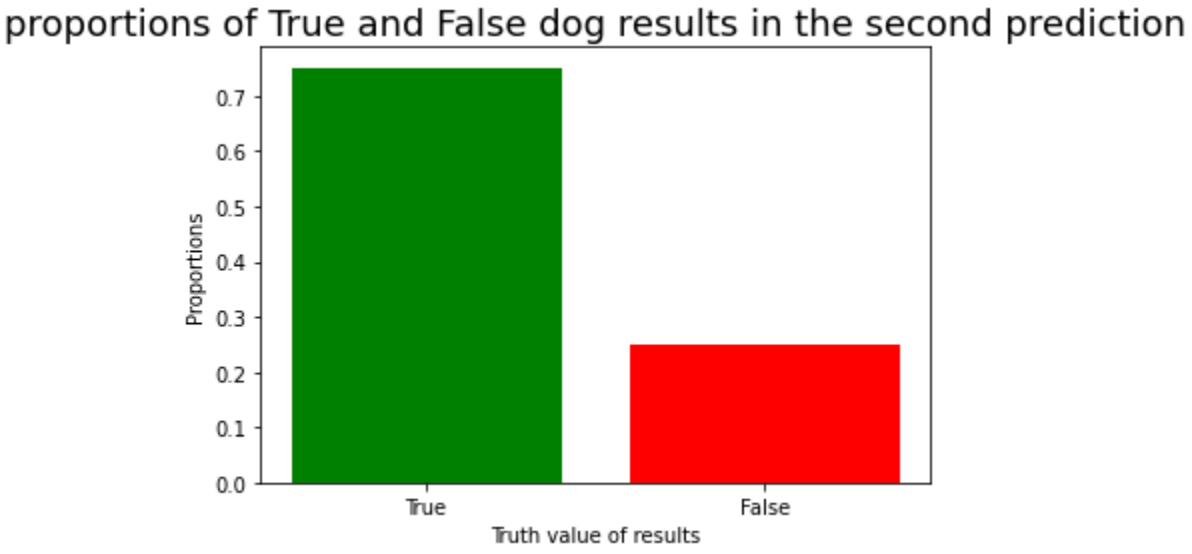
To obtain the proportion of true dog images that were obtained in the first prediction, we had to obtain the total number of true prediction and divide the number by the total number of predictions. The results indicated that in the first prediction, a total of 1490 images were detected as dogs while 522 were not dogs. The bar graph below shows a visual representation of how the proportion of true dog and the false dog results of the first prediction compare.



This affirms that in the first prediction, about 74% of the predicted images were dogs while about 26% were not.

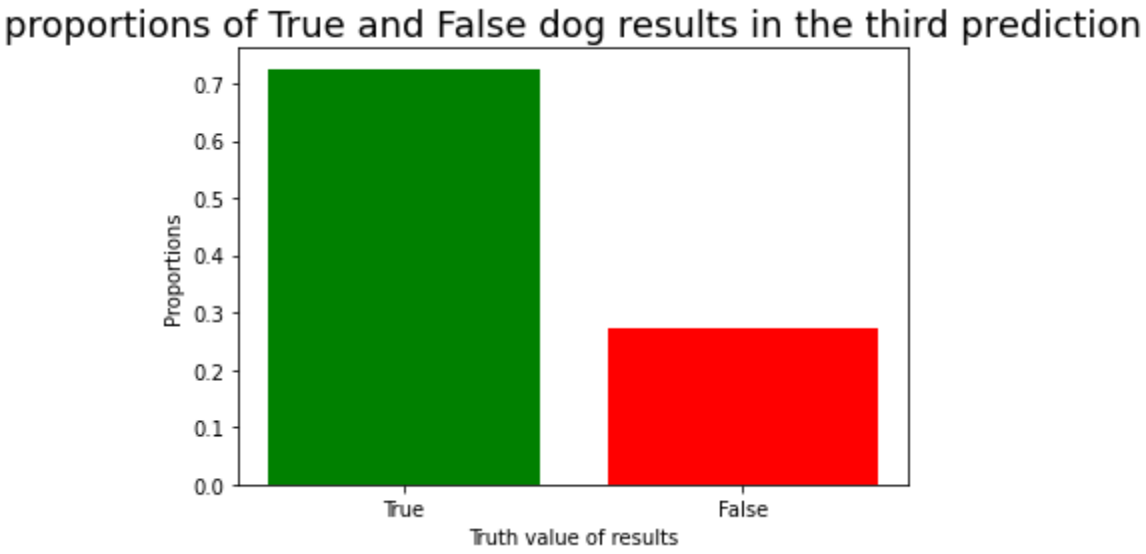
Question 1B: What is the proportion of image predictions that gave a dog image in the second prediction?

For the second prediction, the proportion of prediction that return true dog results was about 75%. The graph below shows the a comparison between true dog results and false dog results for the second prediction.



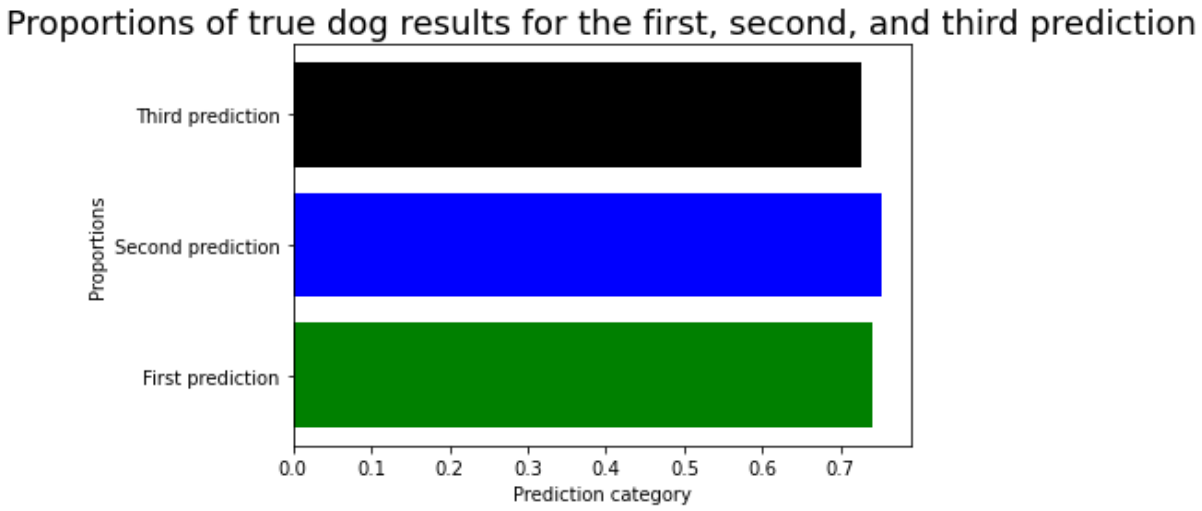
Question 1C: What is the proportion of image predictions that gave a dog image in the third prediction?

For the third prediction, the proportion of prediction that return true dog results was about 73%, while the proportion of false dog results was about 27%. The graph below shows a comparison between true dog results and false dog results for the second prediction.



Question 1D: How do the proportions of true dog results for the first, second, and third prediction compare against each other?

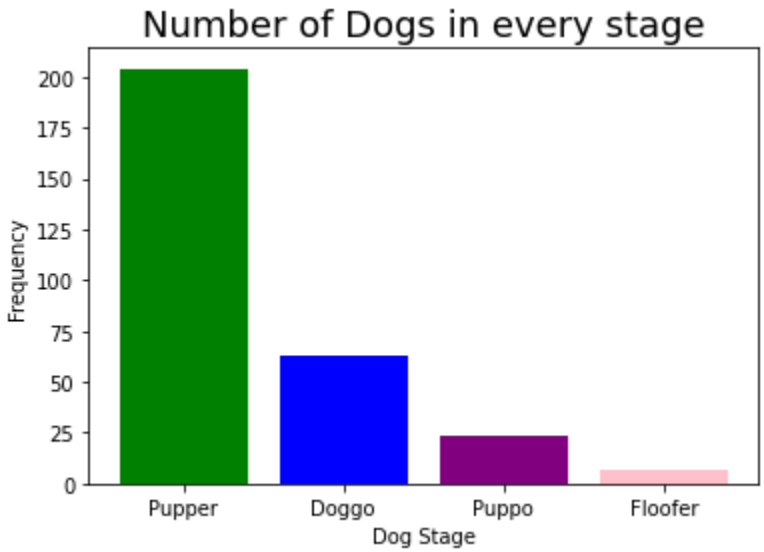
It is important to compare the true dog results from the first, second, and third prediction to establish which of the three prediction returned the highest proportion of true dog results.



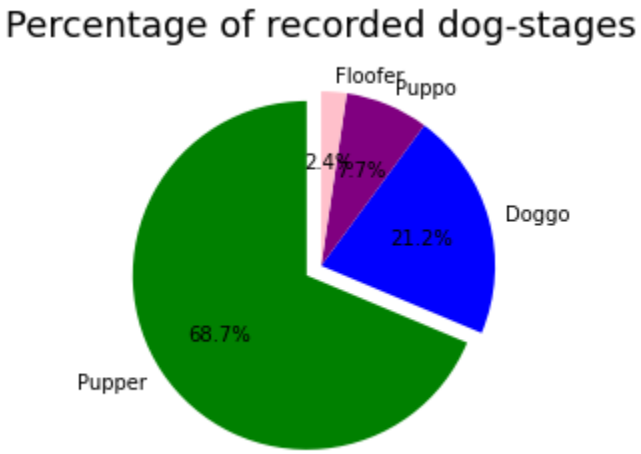
The bar graph above indicates that the second prediction had a slightly higher proportion of true dog predictions as compared to the first and the third prediction.

Question 2: How do the number of the dogs in different stages compare?

There are for dog stages that had been included in the tweeter archive data. They include ***Pupper***, ***Doggo***, ***Puppo***, and ***Floofer***. To show how their numbers compare, we generate a bar graph with the bars representing the number of dogs in each stage.

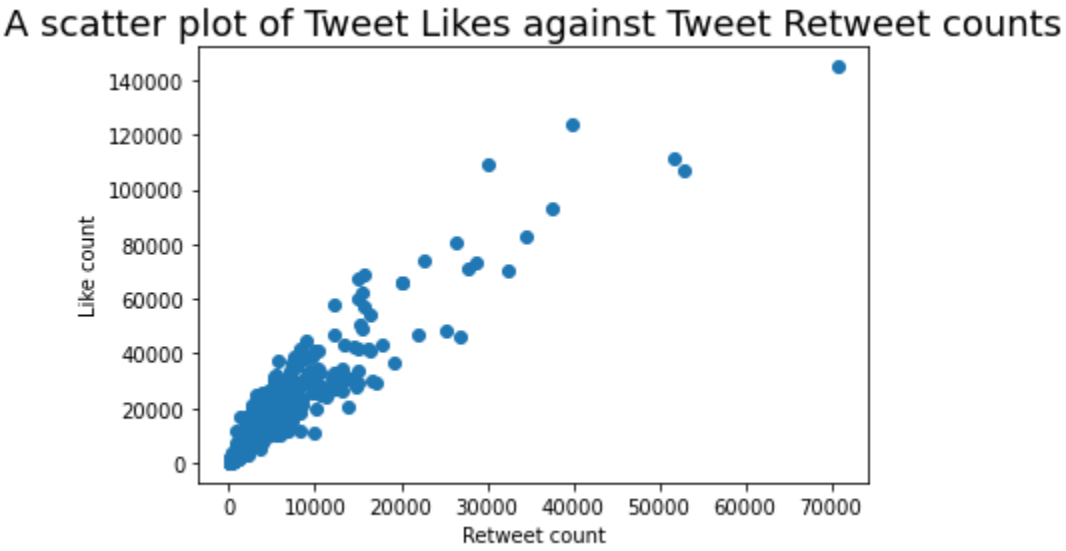


The graph above indicates that Pupper stage had the highest number of dogs, followed by Doggo, Puppo, and Floofer with the fewest number. In terms of percentages, the different dog stages compare as shown in the pie chart below.



Question 3: Is there a relationship between Retweet count and favourite (like) count?

To establish if there exists a relationship between retweet count and favourite/like count, we generate a scatter plot for the two variables.



The scatter plot above indicates that there exists a positive correlation between tweet likes and the retweet count. This means that a tweet with a high number of retweets is likely to have a high number of likes.