Act Report

- ## Introduction

The data-set that we will be analyzing and visualizing is the tweet archive of Twitter user [@dog_rates](https://twitter.com/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.

The data-set is consistent of two data-frame, twitter_archive, and image_predictions. Twitter_archive data-frame contains each tweet data such as: tweet id, tweet URL, tweet created date, followers count, and favorite count, etc.

Image_predictions data-frame, is output of neural network that classify breed of dogs on WeRateDogs Twitter archive. The results: 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). Here is a photo of their profile picture.

- Exploratory Data Analysis

Three questions were asked and answered:

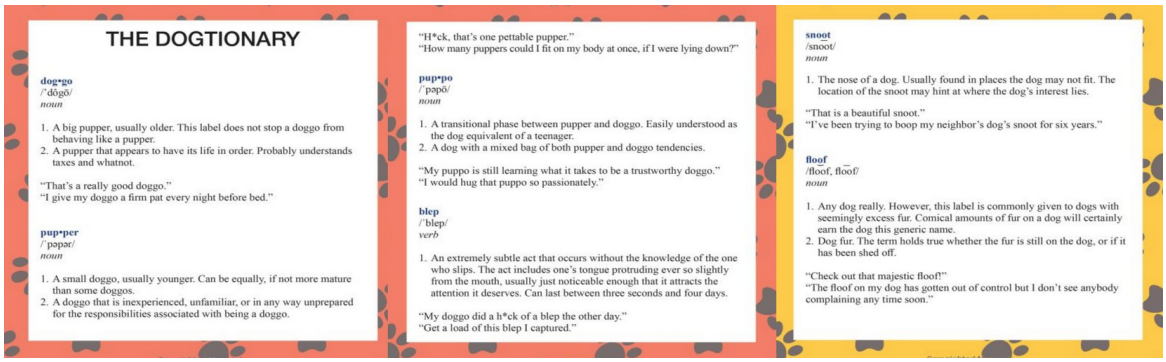
1- Which type of dog phase gets more favorite count?

2- Does the rating_numerator has effect on the favorite count?

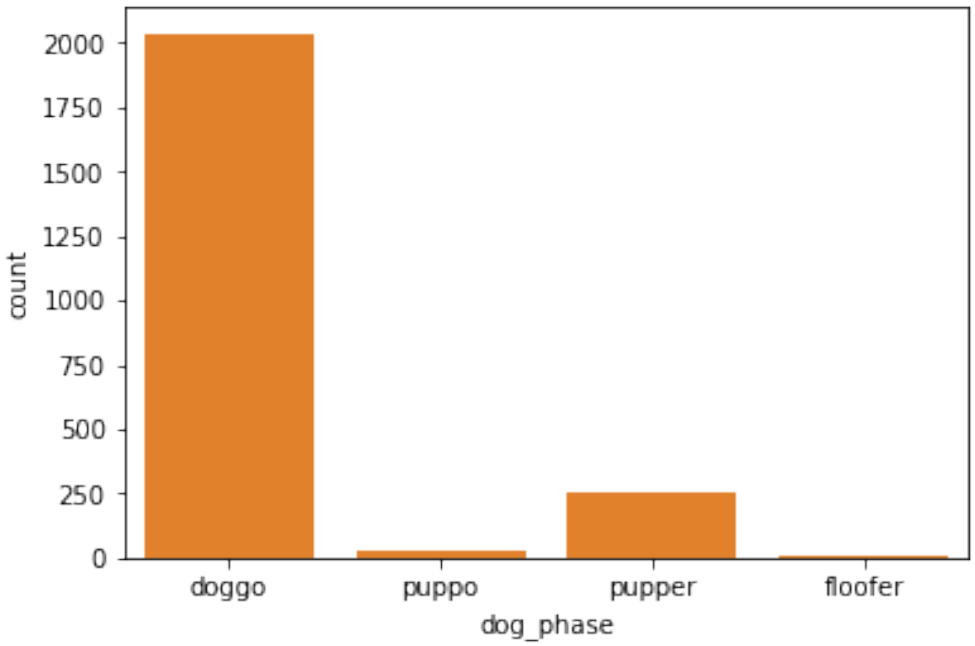
3- What is the most posted dog's breed?

First question: Which type of dog phase gets more favorite count?

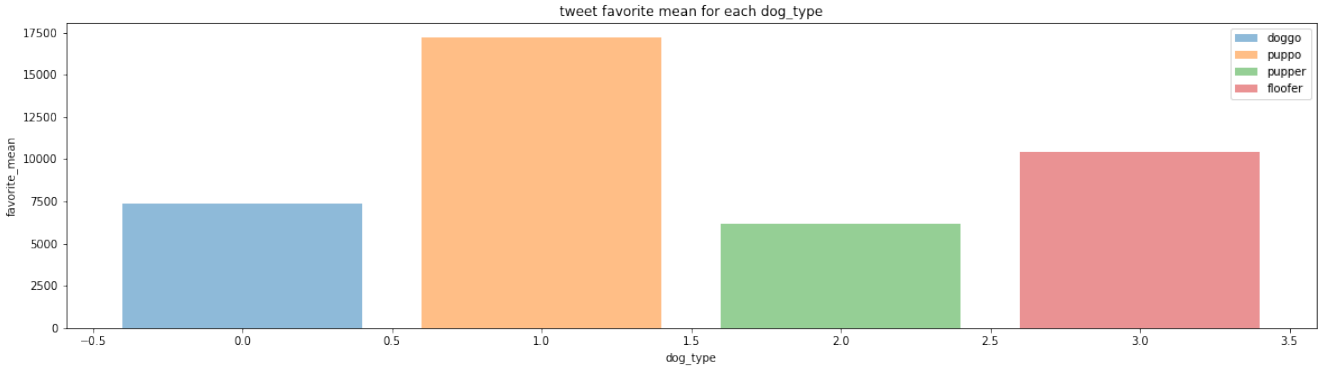
By dog phase we mean the age stage of the dog, and our data had four stages; doggo, puppo, pupper, and floofer. You can use the following dictionary (Dogtionary) as reference for you:



plotting each dog stage with it's value count, we find that doggo age stage has the highest value count = 2000 approximately, see figure bellow.

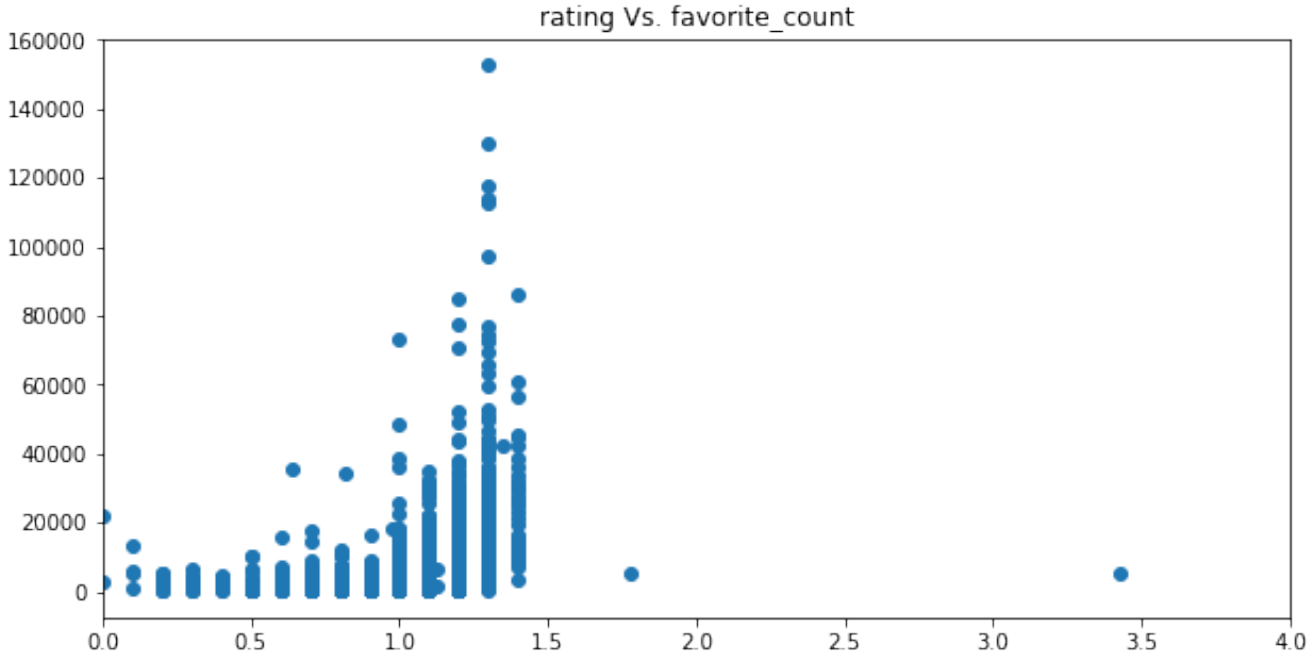


Surprisingly, when we calculated the mean value for favorite count for each dog stage, puppo had the highest mean (each tweet with puppo have higher favorites than other dog stages).



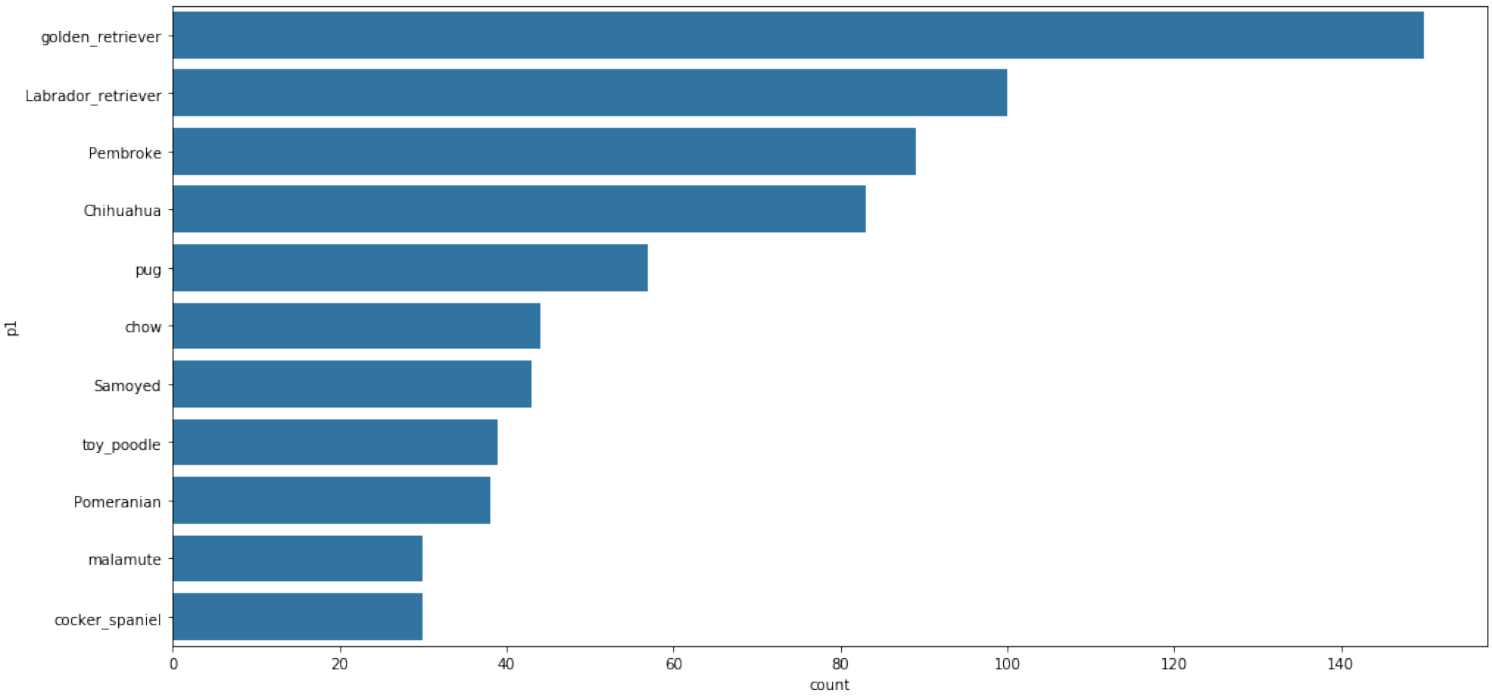
Second question: Does the rating has effect on the favorite count?

As mentioned above the rating have rating numerator > 10, because they are good dogs, so we thought to our self: does such rating have effect if a tweet gets high favorite count or not? Plotting the rating (rating numerator / rating denominator) against favorite count – removing the outliers. Most ratings falls under 1.5 and higher than 0.5, and within this range the tweets gets the highest favorite count.



Question three: What is the most posted dog's breed?

Here we used image-prediction data-set, 1st prediction because it’s the most accurate prediction for classifying the dogs breeds. Then we plotted the value count for most frequent dogs in the data-set. The golden-retrieve breed have the highest count = 140, and next in line the Labrador-retrieve with count = 100, they seem more popular then the rest, and they are the most posted dogs in tweets.



In The end, here are lovely 9 dogs from one of their tweets: