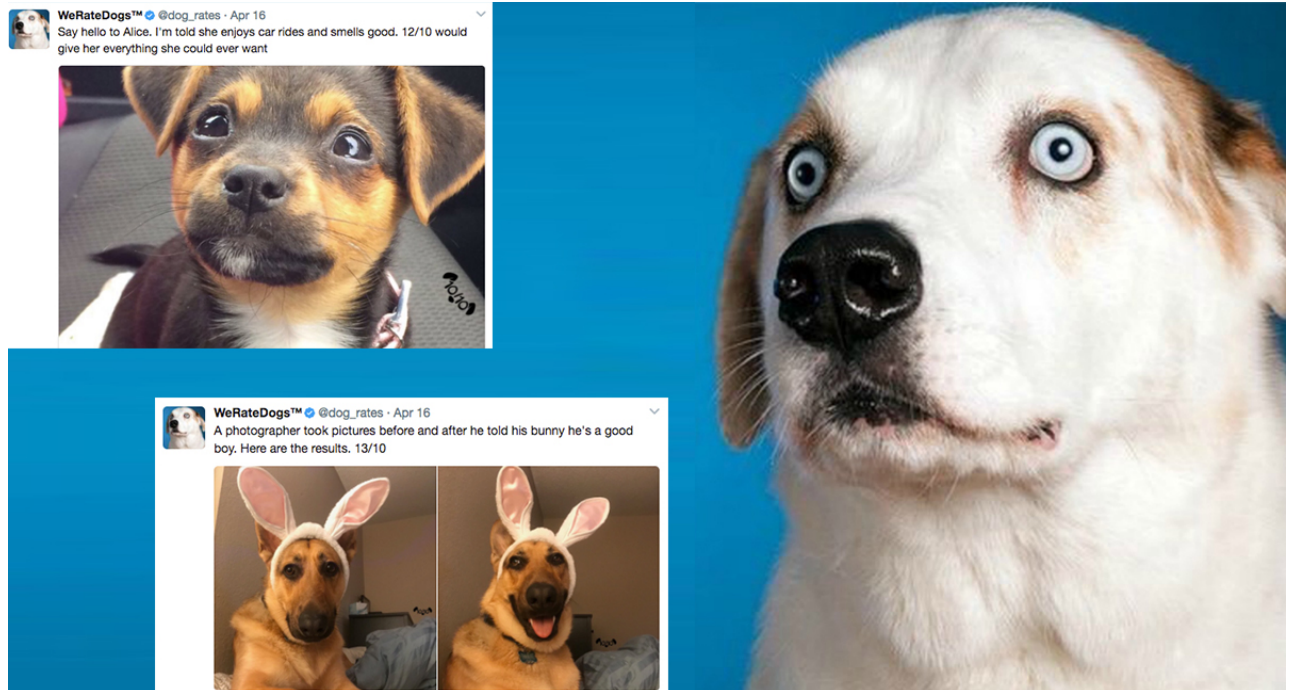


```
In [2]: # %pip install nbconvert
```

# Project: Wrangling WeRateDogs Twitter Archive Data

## Introduction



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 and a numerator of always greater than 10.

The Enhanced twitter archive file contains the basic tweet data extracted from 2356 of the 5000+ tweets.

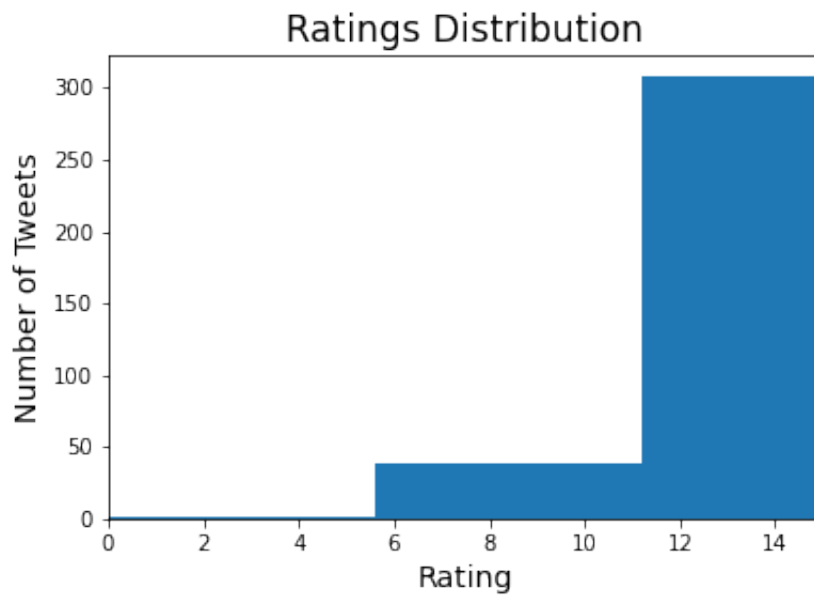
The Image prediction file provides information about the breed of dog present in each tweet.

The tweet json file contains information queried from the twitter with the help of the tweepy.

The datasets have been assessed and cleaned to later create a master dataframe called `twitter_archive_master` which was saved as a csv for later use. This dataset was also used to answer the following questions in the Exploratory Data Analysis.

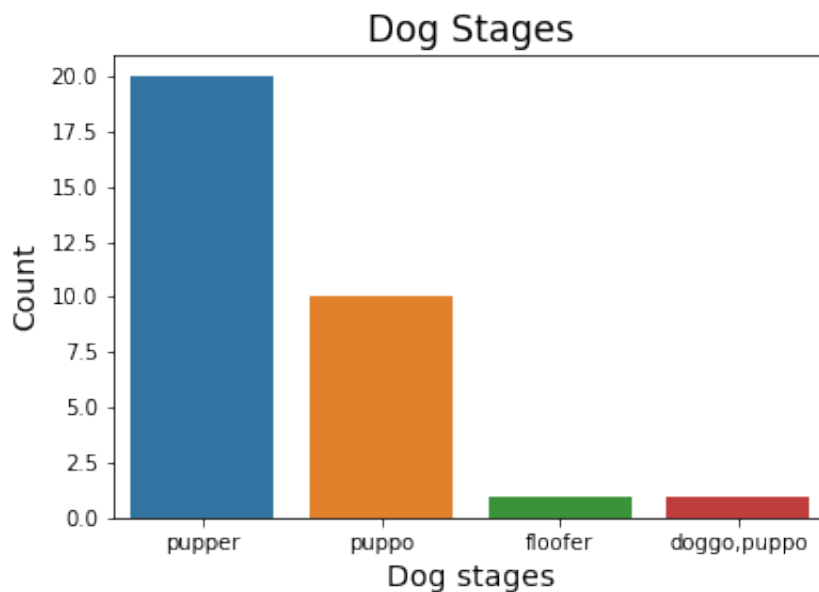
## Question 1: What is the rating distribution?

From the histogram below, 50% and 75% of the rating is at 13.0



## Question 2: What are the common stages?

From the below information 'doggo' stage is most classified dog stage



### Question 3: What is the correlation between Retweets and Favorites counts?

From the below visualization, it is evident that the more the retweets, the more it gets visible and hence the higher the favorites count

