Twitter's WeRateDogs

Exploratory Data Analysis and Visualisations

Figure: Tweet of the dog with highest favourite count

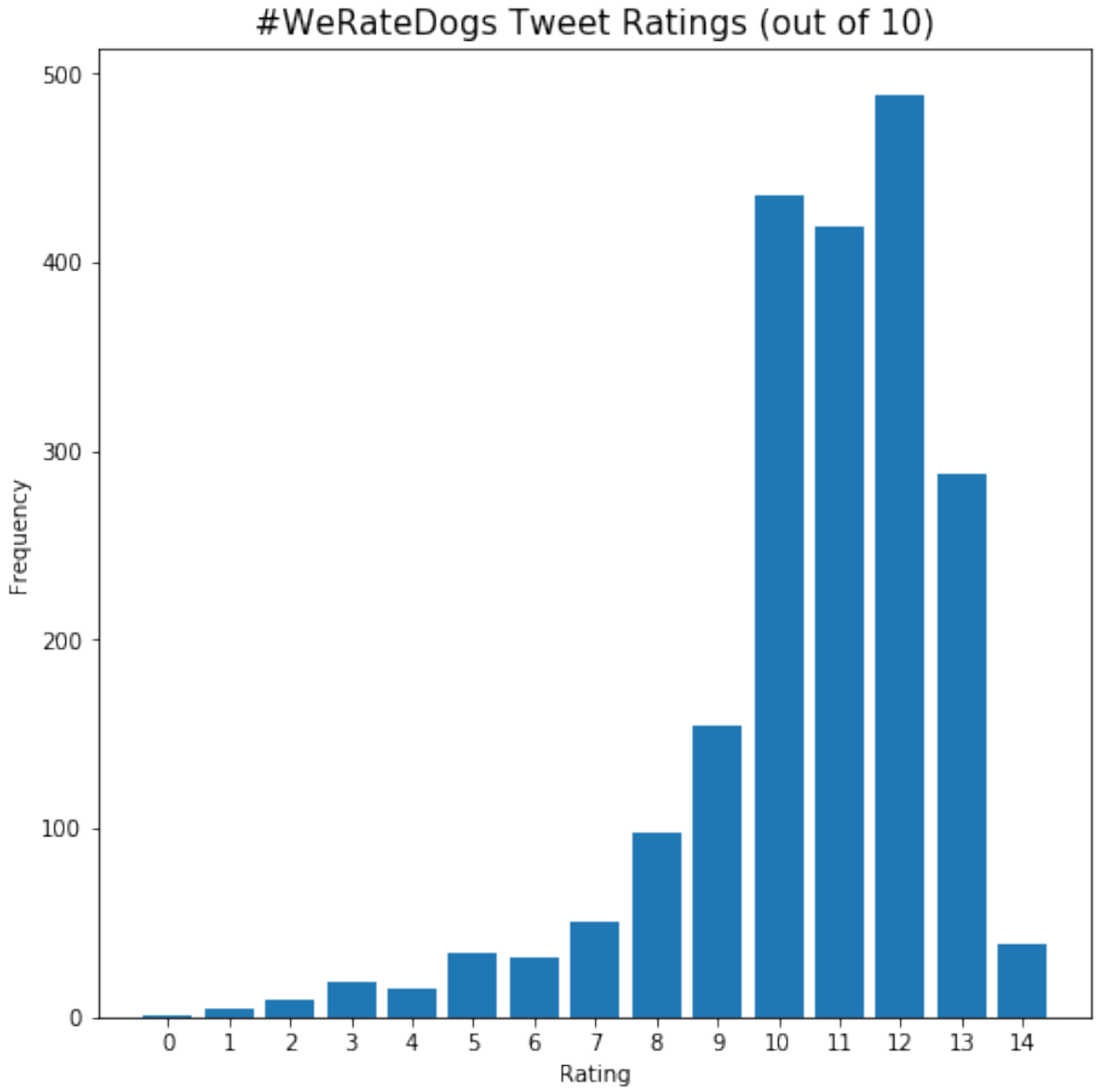
Duc Vu

August 2019

Exploratory Data Analysis and Visualisations

The data is cleaned as outlined by the data cleaning report. The following questions provide interesting insights and visualisations about Twitter's WeRateDogs from November 2015 to July 2017.

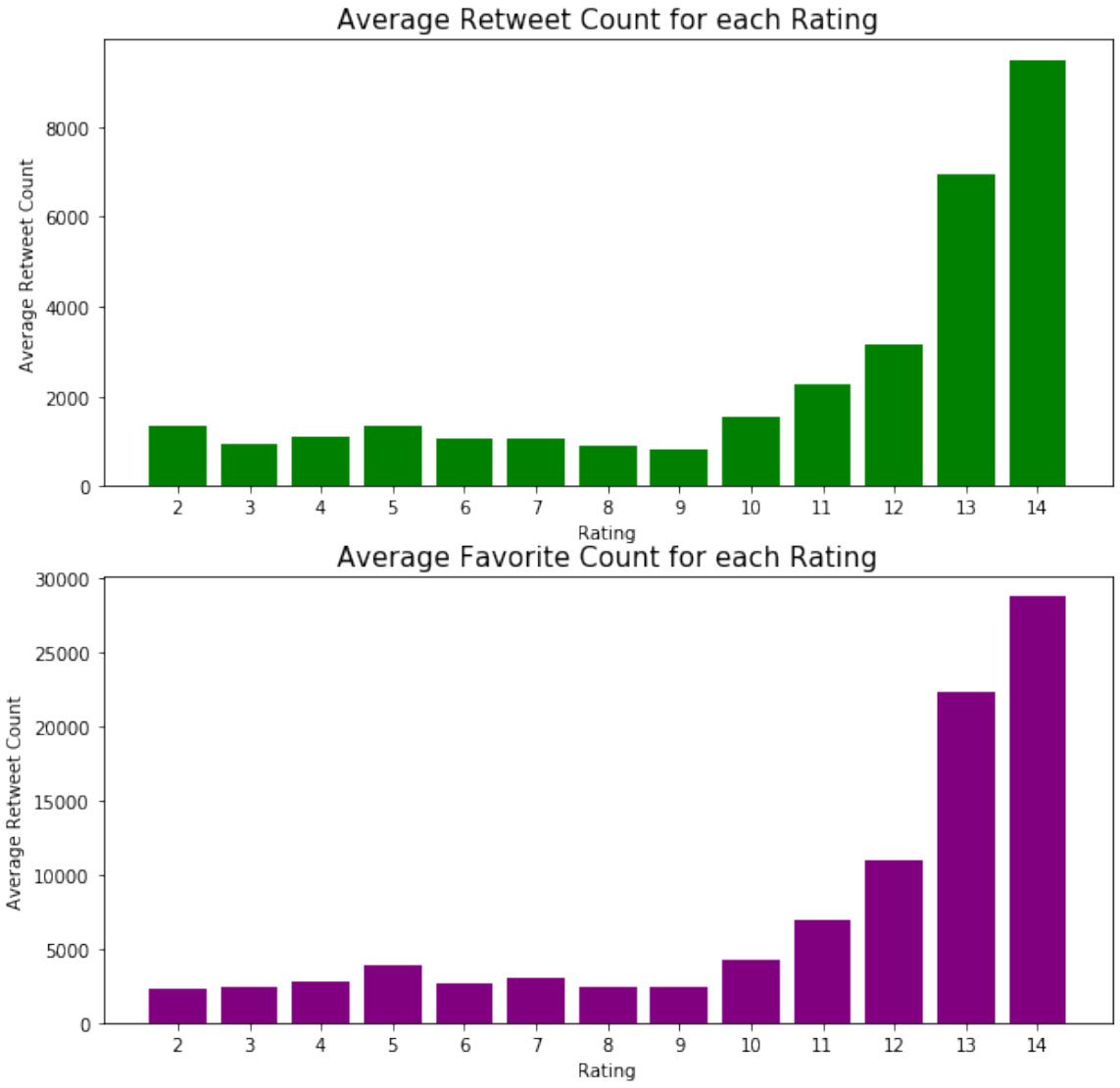
What is the distribution of the #WeRateDogs Tweet Ratings?



As shown in the bar chart above, the distribution follows a left skew (negative skew). Here, there is a gradually increasing frequency from ratings 0-7, followed by a rapid increase peaking at rating 12, and then fall sharply, with zero frequency at rating 15.

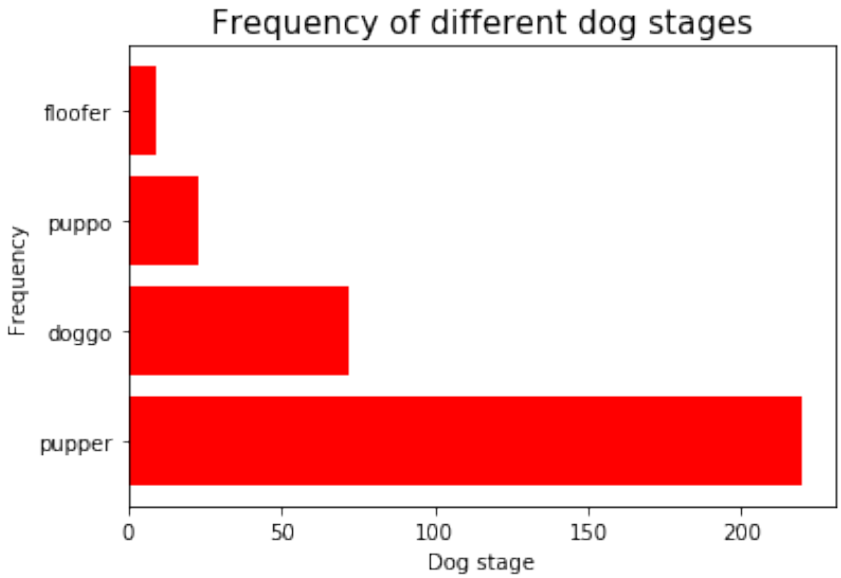
How does Ratings affect the Average Retweet Count and the Average Favorite Count?

From the previous bar chart, it can be seen that there were very few observations for ratings 0 (1 row) and 1 (4 rows). These will not give a fair representation of the average count, and therefore omitted in the following plots.



The above graphs show that both average retweet counts and favorite counts follow a very similar trend with ratings. From the ratings 2-9, the count is low and does not follow any pattern. From ratings 10 onwards, the counts appear to increase exponentially up to rating 14.

What is the most common stages of dogs (ignoring multiple stage dogs)?



Only about 15% of all tweets contain information about the stages of the dogs. With this horizontal bar chart, it can be shown that the most common type of dog is "pupper", and least common type is "floofer".

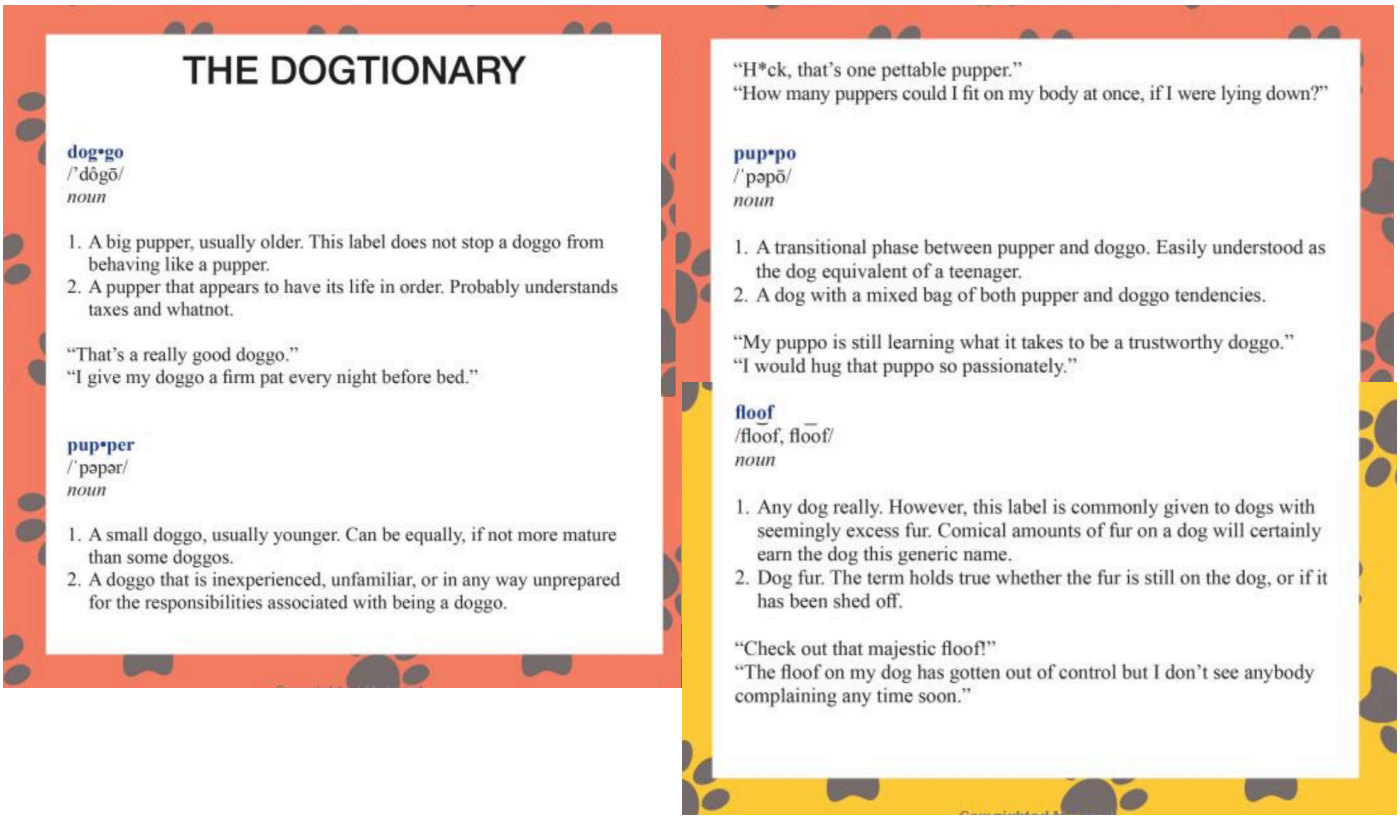
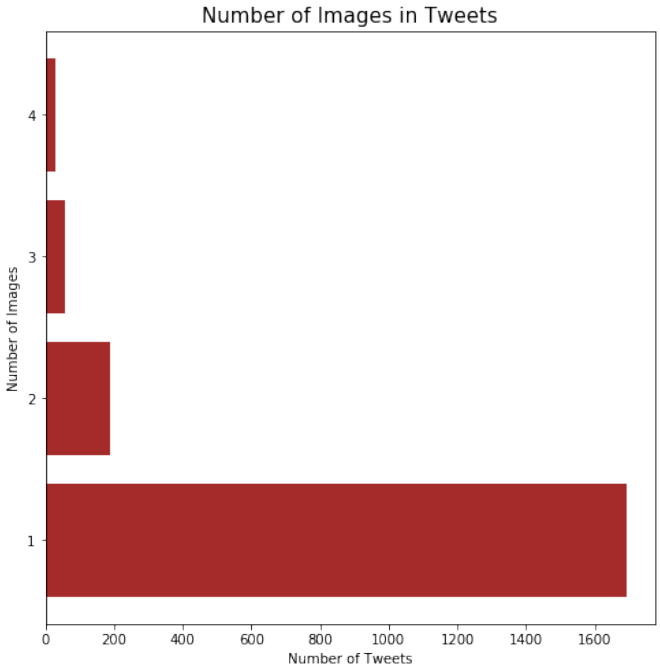


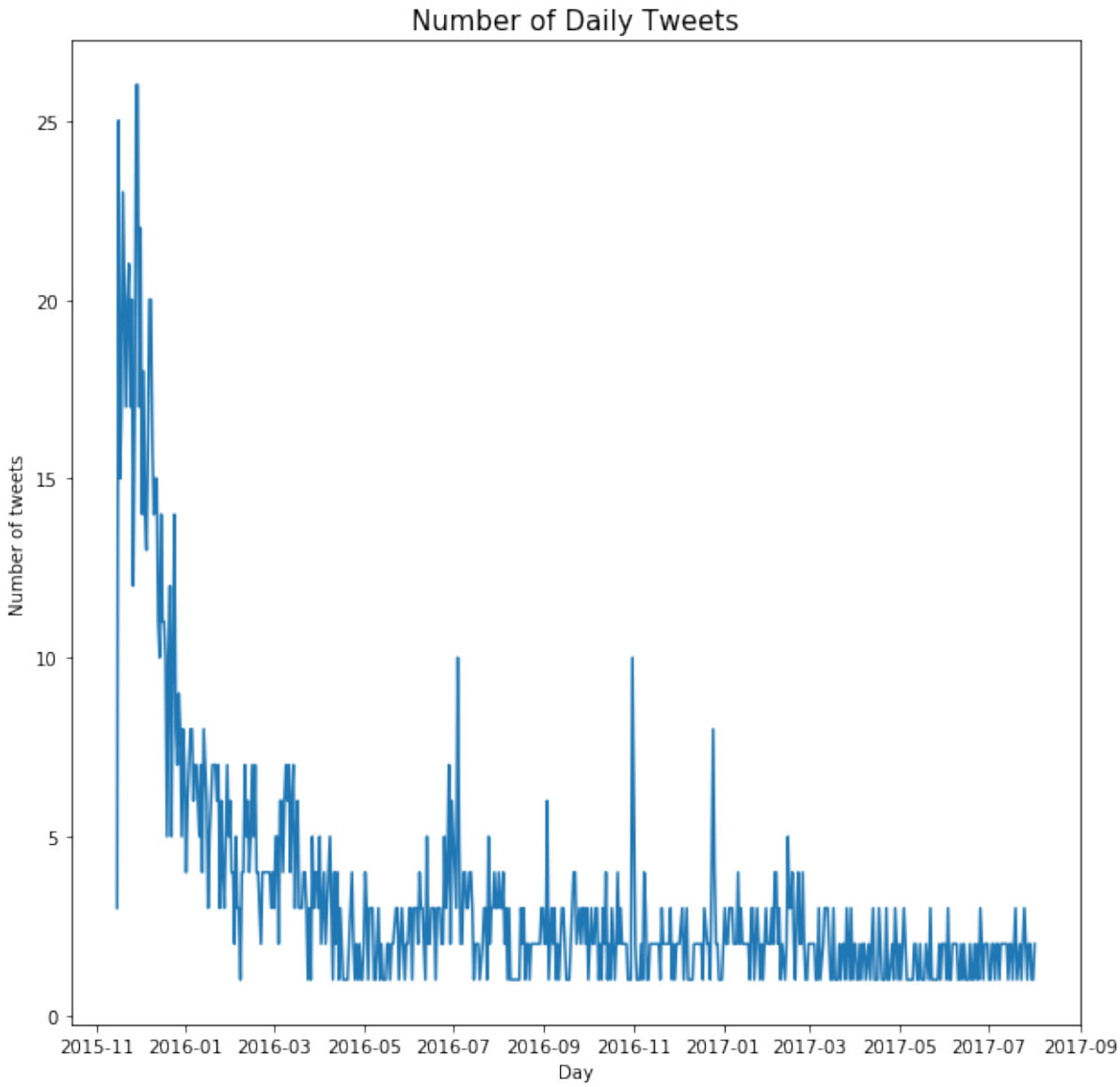
Figure: The Dogtionary explains the various stages of dog: doggo, pupper, puppo, and floof(er) (via the #WeRateDogs book on Amazon)

How many images are most commonly posted in each tweet?



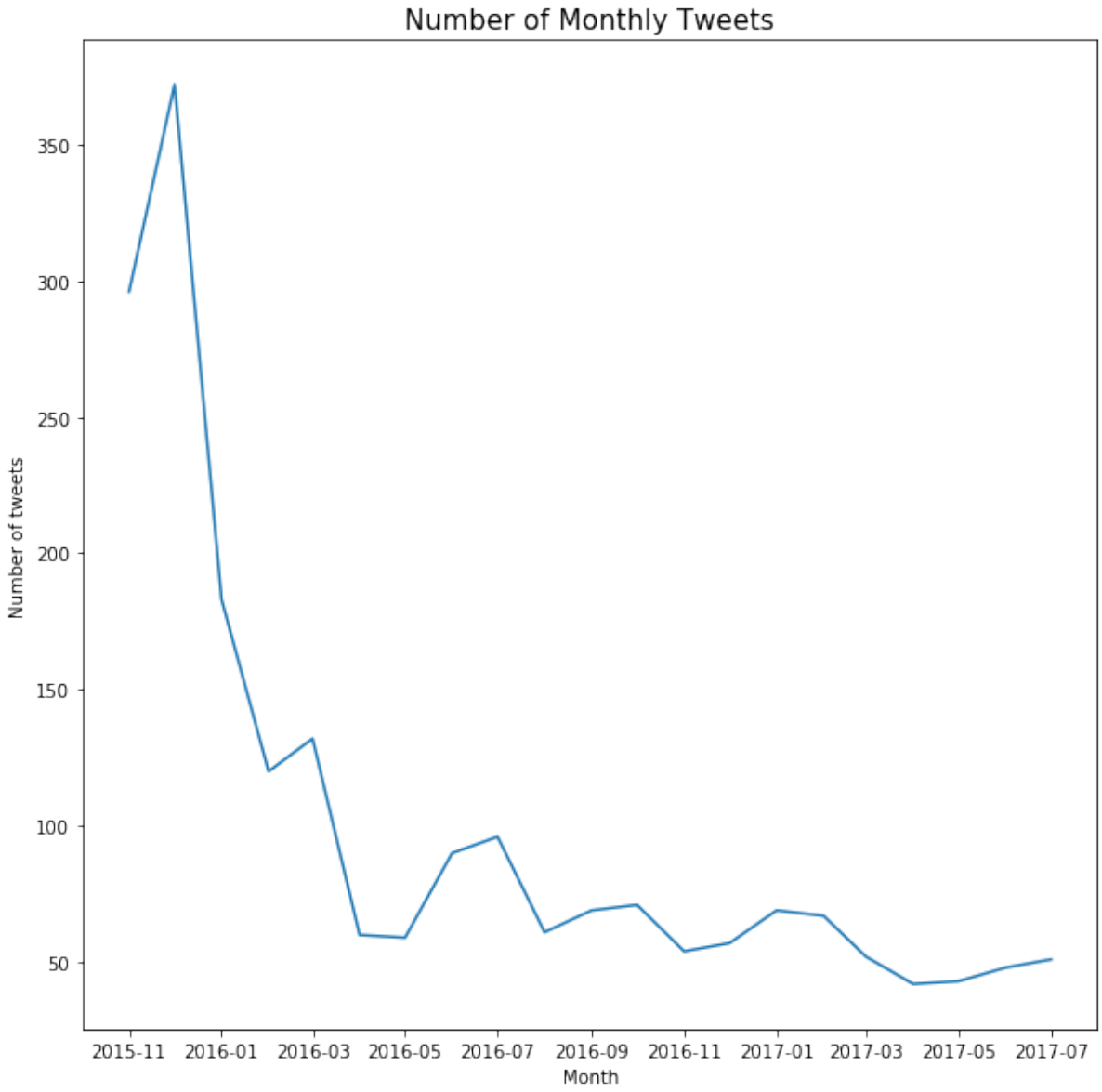
The most common number of images posted is one. This is as expected since one image per tweet tend to give the image more credit and recognition.

How does the number of WeRateDogs tweets vary daily and monthly?



As expected, the variations in the number of tweets posted daily fluctuate wildly and it is difficult to see a trend. Thus, the number of tweets posted month will be analysed.

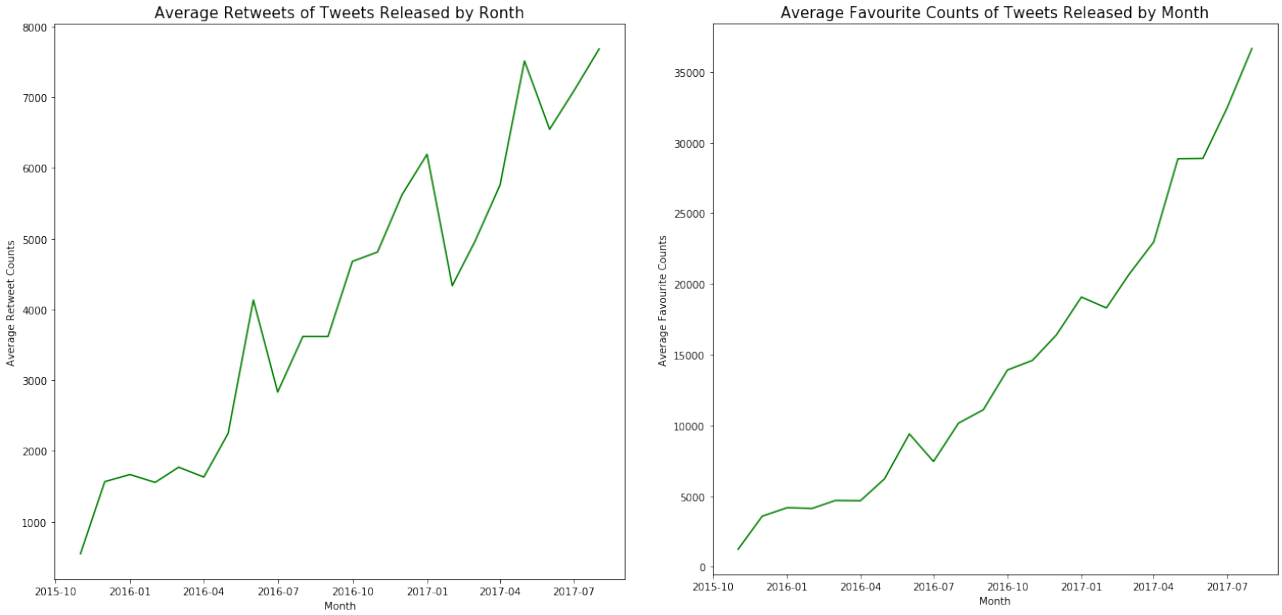
August 2017 does not contain many tweets because the data frame contains tweets from 1 August 2017 only. Thus, these tweets will be removed for further visualisations.



As shown in the graph above, the number of tweets is highest at the end of year 2015 (over 300). Then, the monthly number of tweets have decreased exponentially until around May 2016. The number of tweets experience a slight increase in June and July 2016 (80-100 tweets). Then it drops and fluctuate between around 40 and 70 tweets ever since. This plot show that the number of #WeRateDogs tweets have fallen, as it may have lost popularity since 2016.

Tweets from which month has the highest average WeRateDogs retweets and favorite counts?

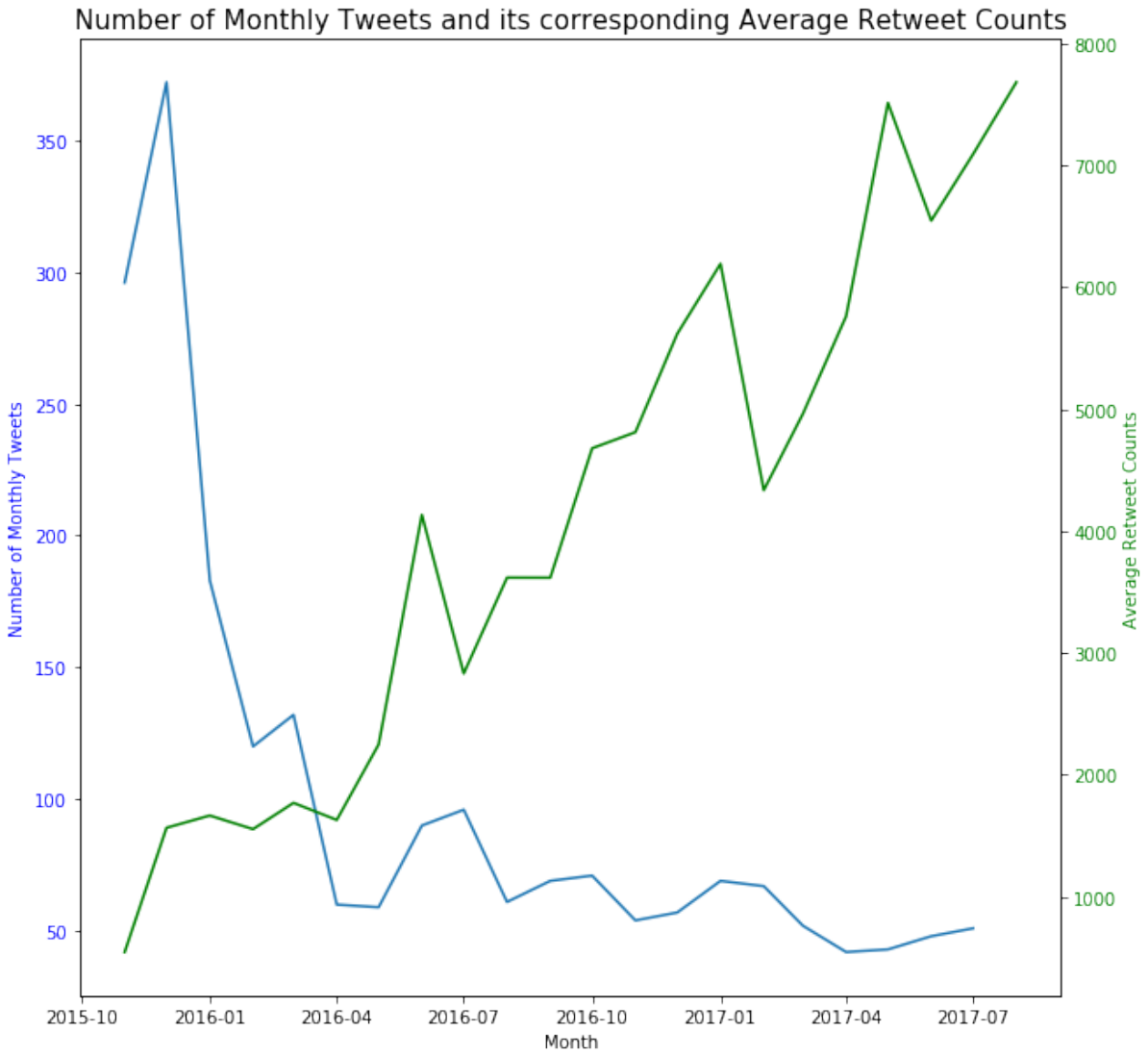
As daily number of tweets vary and would not provide a fair average number of retweets, the tweets are grouped by months. Then, the average number of retweets and favourites are calculated and visualised.



From both plots above, there appears to be clear trend that as tweets are released more recently, the average retweet and favourite counts increase. The months and year with the highest average retweet counts and favourite counts are May 2017 and July 2017, respectively.

Is there a correlation between the number of monthly tweets and its corresponding average retweet counts?

Below, the number of monthly tweets are compared with the average retweet counts for those tweets released in that month. It will be interesting to see if the number of monthly tweets affect the retweet counts.



As shown by the plot above, there appears to be an inverse relationship between the number monthly tweets and the corresponding retweet counts. As the number of monthly retweets decrease, the corresponding average retweet counts increases. This does not indicate that the WeRateDogs tweets lose popularity. Rather - with the number of monthly tweets decreasing, the quality of these tweets may have increased, evident by an increase in retweet counts. That being said, both monthly tweets and retweet counts may not influence each other directly. Further analysis may be required to establish whether one truly affects the other.